UNIVERSITY OF CALIFORNIA SAN DIEGO

Examining Extrachromosomal Elements in Fermented Food Microbiomes: Leveraging Long Reads to Characterize Plasmids and Viruses

A Dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Biology

by

Cong Ba Dinh

Committee in charge:

Professor Rachel J. Dutton, Chair
Professor Eric E. Allen
Professor Matthew D. Daugherty
Professor Terence Hwa
Professor Aspen T. Reese

2023

The Dissertation of Cong Ba Dinh is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my parents, who always prioritize my physical and mental wellbeing; and to Gary Heussler for encouraging and teaching everyone while sharing his constant smile.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Supplemental Tables 3.2.1-3.2.20 are not provided due to shared copyrights. Contact for use.

Supplemental Table 3.2.1. Collection schedule of sampled washed-rind cheeses.

Supplemental Table 3.2.2. Sequencing data

Supplemental Table 3.2.3. 16S amplicon read classification statistics.

Supplemental Table 3.2.4. ITS amplicon read classification statistics.

Supplemental Table 3.2.5. Long-read based taxonomic classification statistics.

Supplemental Table 3.2.6. Assembly statistics of assemblies, co-assemblies, isolates

Supplemental Table 3.2.7. Collection of high-quality bins from timepoint assemblies

Supplemental Table 3.2.8. Sequencing data of isolates from Cheese B.

Supplemental Table 3.2.9. High-quality MAGs recovered from Cheese A

Supplemental Table 3.2.10. High-quality MAGs and isolate genomes recovered from Cheese B

Supplemental Table 3.2.11. High-quality MAGs recovered from Cheese C

Supplemental Table 3.2.12. Percentage of Illumina short reads mapped to genomic catalog

Supplemental Table 3.2.13. Plasmid and virus predictions of mega-assemblies

Supplemental Table 3.2.14. Full results from the viralAssociationPipeline

Supplemental Table 3.2.15. Potential host changes of MGEs over time

Supplemental Table 3.2.16. Psychrobacter genomes used for pangenome analysis

Supplemental Table 3.2.17. Functional enrichment in *Psychrobacter* genomes from cheese

Supplemental Table 3.2.18. CFU counts of *in vitro* washed rind communities.

Supplemental Table 3.2.19. Isolate genome statistics and coverage at days 7 and 21

Supplemental Table 3.2.20. “Cloud” gene clusters unique to Cheese B *Psychrobacter* genomes

LIST OF SUPPLEMENTAL FILES

Supplemental File 1 Supplemental tables in Microsoft Excel format
"Dinh_dissertation_supplementalTables.xlsx"

Supplemental File 2 Supplemental Figure 2.6 in PDF format
"Dinh_dissertation_SupFig_2.6.pdf"

ACKNOWLEDGEMENTS

Our lab made coming in day after day worth the trek: Swole squad with Gary and Steven; statistics and jokes with Manon and Emily; Brooke and Christina tolerating our antics; then Tara and Rachel putting me in time out. No question was ever too difficult to be answered. Thank you to Lucas Patel, Ryan Hsu, Galilea Guerrero, and especially Carla Espinoza for enduring my mentorship.

Outside of lab, I relied heavily on Quynh for keeping me on track and maintaining a schedule as much as possible. Andy helped keep me sane by getting my jokes when no one else would. And Sam and Evelyn as half of team neighbors and being part of the pandemic pod, helping all of us stay sane without isolation.

Chapter 1, in full, is a reprint of the material as it appears in Federation of European Microbiological Societies Microbiology Reviews, 2020, Saak, Christina C.; Dinh, Cong B.; Dutton, Rachel J. The dissertation author was an investigator and author. Small edits were made in the reproduction for this dissertation. The authors would like to thank Brooke E. Anderson for useful feedback on the manuscript.

Chapter 2, in full, is a preprint of the material as it was prepared for to mSystems, 2022, Dinh, Cong B, Dutton, Rachel J. The dissertation author was the primary investigator and author of this paper. Nicholls, Samuel M, Loman, Nick J supported this work through data collection and discussions. Gary Heussler, Emily Pierce, and Manon Morin supported this work through discussions and teaching of a course that contributed data.

Chapter 3 includes material that is in preparation for publication as well as additional material that will be excluded from that publication. Contributors to this chapter include Saak,

Christina C.; Patel, Lucas, Dutton, Rachel J. The dissertation author was the primary researcher and author of this chapter. The publication that includes this data was with Saak, Christina C; Pierce, Emily C; Portik, Daniel; Hall, Richard; Ashby, Meredith; and Dutton, Rachel J.

VITA

2015 Bachelor of Science in Materials Science, Stanford University

2023 Doctor of Philosophy in Biology, University of California San Diego

2024 Doctor of Medicine, University of California San Diego

ABSTRACT OF THE DISSERTATION

Examining Extrachromosomal Elements in Fermented Food Microbiomes:

Leveraging Long Reads to Characterize Plasmids and Viruses

by

Cong Ba Dinh

Doctor of Philosophy in Biology

University of California San Diego, 2023

Professor Rachel J. Dutton, Chair

The physiological processes associated with microbiomes represent the biology of individual microbial cells and their interactions with other cells. Moreover, the interactions between microbes are not static, but represent dynamic processes that are subject to ecological

and evolutionary changes. One of the main drivers of rapid evolution in microbial species is the process of horizontal gene transfer, often mediated by mobile genetic elements. Few systems or tools exist that allow the study of this process in microbiomes. In this thesis, we review experimental methods to study horizontal gene transfer, as well as catalog computational methods to sequence and filter for mobile elements, including integrative and conjugative elements, plasmids, and viruses. We next leverage long read DNA sequencing to lay the groundwork for a complex in vitro washed cheese rind microbiome model. With our microbiome model and computational methods, we are able to 1) probe natural microbiomes in kefirs and cheese rinds to better understand the breadth and novelty of mobile elements and 2) begin to understand how different community members play a role in the evolution of species and mobility of plasmids within a genus. Our work in the identification, characterization, and tracking of mobile elements in microbiomes enhances our understanding of the unique and niche-adaptive genes carried by these fermenting microbes. Finally, we explore the potential clinical and industrial applications of long read and proximity ligation.

CHAPTER 1: Review of Experimental Methods to Study Horizontal Gene Transfer

Experimental approaches to tracking mobile genetic elements in microbial communities

Christina C. Saak, Cong B. Dinh and Rachel J. Dutton

Abstract

Horizontal gene transfer is an important mechanism of microbial evolution and is often driven by the movement of mobile genetic elements between cells. Due to the fact that microbes live within communities, various mechanisms of horizontal gene transfer and types of mobile elements can co-occur. However, the ways in which horizontal gene transfer impacts and is impacted by communities containing diverse mobile elements has been challenging to address. Thus, the field would benefit from incorporating community-level information and novel approaches alongside existing methods. Emerging technologies for tracking mobile elements and assigning them to host organisms provide promise for understanding the web of potential DNA transfers in diverse microbial communities more comprehensively. Compared to existing experimental approaches, chromosome conformation capture and methylome analyses have the potential to simultaneously study various types of mobile elements and their associated hosts. We also briefly discuss how fermented food microbiomes, given their experimental tractability and moderate species complexity, make ideal models to which to apply the techniques discussed herein and how they can be used to address outstanding questions in the field of horizontal gene transfer in microbial communities.

Introduction

Horizontal gene transfer (HGT) involves the intragenerational sharing of genetic material, as opposed to the intergenerational sharing characteristic of vertical inheritance. In bacteria, it is an effective evolutionary process that allows the rapid acquisition of new phenotypes and metabolic abilities (Wiedenbeck and Cohan 2011; Hall, Brockhurst and Harrison 2017), and has been hypothesized to significantly alter microbial communities, such as the one associated with the human gut, and its interactions with the body (Lerner, Matthias and Aminov 2017; Sitaraman 2018). For example, HGT can lead to rapid adaptation to new environmental pressures such as antibiotics (Lester et al. 2006). It can also turn non-virulent bacteria into pathogens (Avery, Macleod and McCarty 1944; Waldor and Mekalanos 1996; Faruque and Mekalanos 2012; Khalil et al. 2016) or allow bacteria to adapt to entirely new niches through the acquisition of metabolic pathways (Hehemann et al. 2010; Tasse et al. 2010). Genes associated with these phenomena are frequently found on mobile genetic elements (MGEs) or are moved alongside them. In brief, MGEs are pieces of DNA that can show mobility within their host's genome or between genomes of donor and recipient cells. Microbial MGEs include plasmids, bacteriophages and transposable elements. These elements differ in regards to their genetic structure as well as means of mobilization as outlined in the next section.

Due to the fact that microbes live together in microbial communities, or microbiomes, MGEs can form a web of potential concurrent DNA transfers. Furthermore, other types of interactions between species within a community may influence the process of HGT. Thus, it is crucial to understand not only the movement of individual MGEs, but to also understand

how they move collectively within communities, and how this process is impacted by interactions with other species and the environment. Therefore, we need global approaches that are capable of tracking as much of the diversity of MGEs as possible within a variety of communities and in as wide a range of hosts as possible.

In this review, we discuss open questions related to the process of HGT specifically within communities, and approaches that allow the tracking of genetic material within them. In discussing the breadth of techniques available, a particular focus is put on the transfer of plasmids between bacteria. Since plasmids often carry antibiotic resistance genes (ARGs) and the dissemination of ARGs via HGT has critical implications for public health, we will also incorporate discussions of techniques particularly aimed at these genes of interest. In the second half of the review, we highlight the emerging metagenomics techniques based on chromosome conformation capture (3C) and methylome analyses as they are opening new doors for studying the presence and transfer of MGEs in communities. Since the best techniques are not useful unless they can be applied to suitable model systems, we will end the review on discussing how the communities associated with fermented foods might constitute useful models to be studied with techniques discussed here.

Horizontal gene transfer in microbial communities: canonical and emerging mechanisms

Canonical HGT mechanisms

MGEs move within and between bacterial genomes through four canonical mechanisms: transformation, transposition, conjugation, and transduction. In transformations (Figure 1.1.1), environmental DNA is taken up via a proteinaceous DNA uptake machinery

(Chen and Dubnau 2004; Claverys, Martin and Polard 2009; Dubnau and Blokesch 2019). Bacteria that express this machinery are referred to as competent. Transposition (Figure 1.1.2) is mediated by transposable elements (TEs) that move, or transpose, between different locations in the genome (Siguier, Gourbeyre and Chandler 2017). In bacteria, it has long been known that antibiotic-encoding transposons can integrate into and transpose between MGEs such as plasmids, and then these transposons can be horizontally shared along with their host plasmids (Hedges and Jacob 1974). Integrons are another class of elements that have been found to capture genes associated with antibiotic resistance very efficiently. They employ site-specific recombination downstream of a resident promoter to enable expression of the integrated genes (Mazel 2006). These elements are not mobile by themselves, but have been found on transposons and plasmids, thereby contributing to HGT of their cargo genes via conjugation. Conjugation (Figure 1.1.3) is the transfer of DNA, such as extrachromosomal, circular plasmids, at points of physical contact (or mating junctions) between neighboring cells, often referred to as mating pairs (Chen, Christie and Dubnau 2005). In addition to circular plasmids, linear plasmids are known to exist in Gram-positive and Gram-negative bacteria (Hinnebusch and Tilly 1993) and to be able to undergo conjugation (Dib et al. 2015). In addition to plasmids (Figure 1.1.3b), conjugative and integrative elements (ICEs), sometimes referred to as conjugative transposons, can be shared by conjugation given suitable conditions (Figure 1.1.3a). Prior to conjugation, ICEs excise from the host genome and circularize; after transfer they can integrate into the new host's genome (Wozniak and Waldor 2010; Delavat et al. 2017). Prokaryotic viruses, called bacteriophages, mediate transduction of DNA (Figure 1.1.4). Transduction can happen due to faulty DNA packaging into viral

particles (general transduction), or due to imprecise excision of prophages from genomes (specialized transduction), or through emerging mechanisms (Chiang, Penadés and Chen 2019).

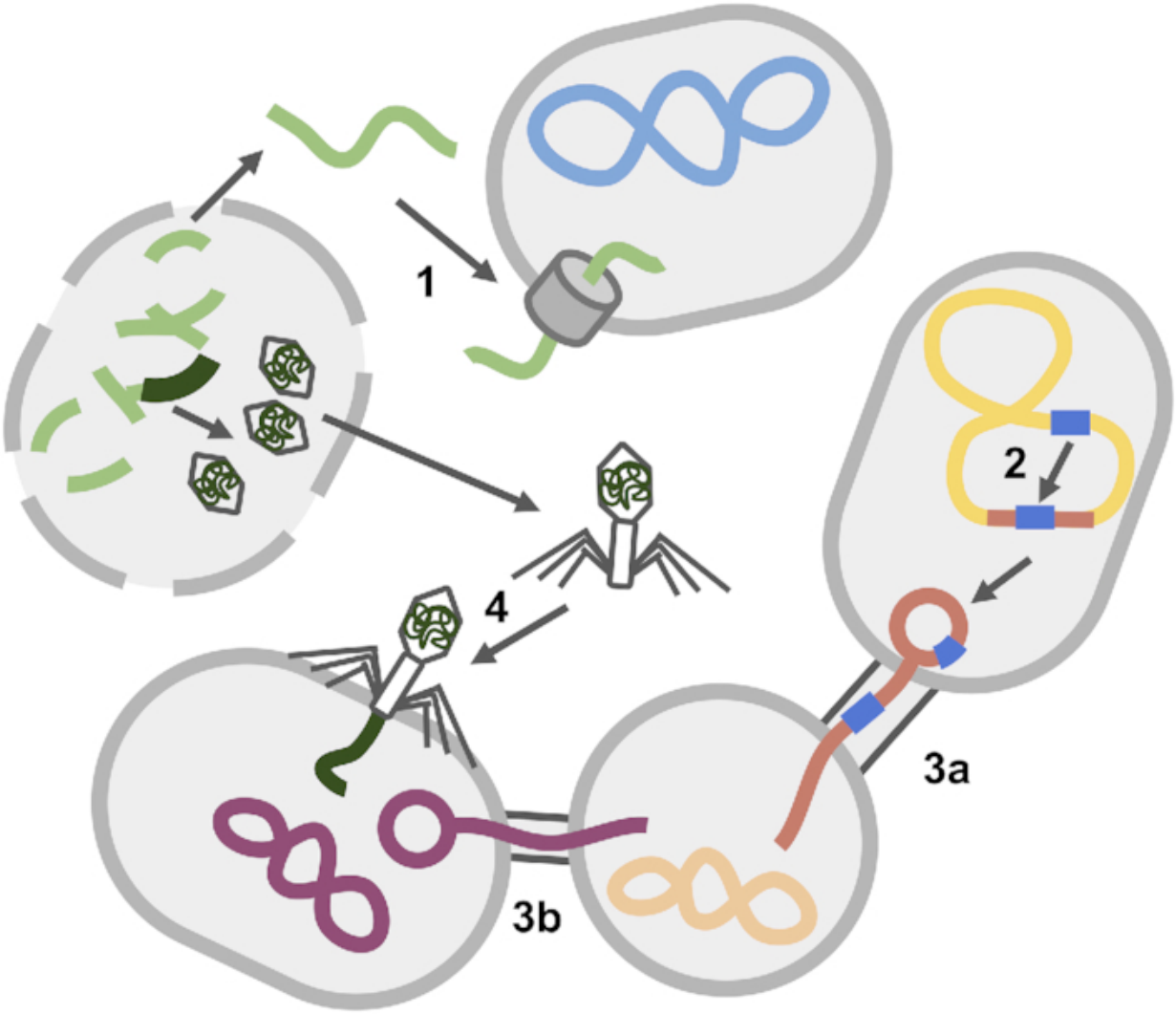


Figure 1.1 In microbial communities, a web of potential horizontal gene transfer mechanisms can co-occur. In transformation (1), free DNA (light green), for example released by lysed cells, is taken up from the environment. In transposition (2), DNA (blue) translocates from one location in the genome to another. If this location is on a conjugative element, such as an integrative and conjugative element (ICE, orange), it can then move to another cell by conjugation (3a). Other conjugative elements include plasmids (mauve, 3b). Transduction is mediated by prophages (dark green), which are integrated into the host genome. Given suitable conditions, they excise and induce production of infectious phage particles, which can carry host DNA (4).

Emerging HGT mechanisms

While these major classes of MGEs and their associated mechanisms of HGT have been studied for decades, new information about these elements is continuously coming to light. For example, ICEs are not always encoded in one contiguous sequence, but can be found in pieces that are distributed along chromosomes and that excise individually and then assemble into the complete ICE before transfer (Haskett et al. 2016, 2018; Greenlon et al. 2019). In addition, new HGT mechanisms and types of mobile DNA are still being discovered. Gene transfer agents, for example, combine elements of transduction and transformation (Westbye, Beatty and Lang 2017). Additionally, membrane vesicles can form from the donor's cell membrane or by phage-induced cell lysis and can carry chromosomal, plasmid or phage nucleic acids (DNA or RNA) (Grüll, Mulligan and Lang 2018). While the packaging mechanisms of nucleic acids into membrane vesicles are generally unknown, an archaeal plasmid has been found to cause its own transfer to plasmid-free related strains via membrane vesicles (Erdmann et al. 2017). The term “vesiduction” was recently proposed for this type of membrane vesicle-mediated HGT (Soler and Forterre 2020). It was further proposed that vesiduction is comprised of the extrusion of membrane vesicles into the environment, followed by their attachment to recipient cell surfaces, transport of DNA into the cytoplasm and finally the acquisition of genetic material. Distributive conjugal transfer (Gray and Derbyshire 2018; Clark, Gray and Derbyshire 2020) and mycoplasma chromosomal transfer (Dordet-Frisoni et al. 2019) are two recently described conjugative processes that can cause the exchange of large amounts of unrelated chromosomal loci. While having similar consequences, the two processes seem to rely on different machineries.

Transjugation, another emerging mechanism, is a hybrid of transformation and conjugation, but has so far only been described in Thermus thermophilus (Blesa et al. 2015, 2017). In addition, nanotubes are being investigated as possible conduits for non-conjugative plasmids (Dubey and Ben-Yehuda 2011). To make matters even more complicated, MGEs themselves can be contained within one another. As mentioned above, transposons can insert themselves into plasmids and even other transposons, for example (Sheppard et al. 2016).

Open questions about HGT in microbial communities

A multitude of open questions remains specifically about HGT in the context of communities. We believe that many of these questions can be addressed with emerging metagenomics techniques discussed in this review, since they aim at providing a comprehensive overview of the MGE pool and its movement through diverse microbial communities. In the future, this could allow us to predict how individual MGEs and their transfer might be affected by the overall pool of mobile DNA present. Gaining this birds-eye view is important both for understanding MGEs within an individual cell, as well as understanding HGT within communities. For example, conjugative elements have been found to silence transformation machinery in Legionella, showcasing that the different routes of HGT can affect each other's efficacy (Durieux et al. 2019). Conjugative elements are also known to affect each other's intercellular transfer and can be divided into three classes based on their transmissibility: self-transmissible, mobilizable and non-mobilizable (Smillie et al. 2010). Self-transmissible plasmids encode their own type IV secretion system in addition to other elements required for transmission including relaxases, which are proteins that

recognize the origin of transfer (oriT) and are crucial for the initiation and completion of conjugation. As such, they can be shared by conjugation between two bacterial cells in what is also referred to as a bi-parental mating. Mobilizable plasmids carry only the relaxases and an oriT, while relying on type IV secretion systems that are available in trans, for example from the chromosome or a co-resident plasmid. In tri-parental matings, the plasmid supplying the type IV secretion system is introduced into the donor cell by a third bacterial strain, often referred to as a helper strain. Last, non-mobilizable plasmids carry none of the factors required for conjugative transfer. Given the potential complexities of DNA transfer in microbial communities, techniques are required that enable the monitoring of many different MGEs at the same time.

Spatial heterogeneities within communities should also be considered when addressing HGT dynamics in diverse communities. For example, the concept of microbial hotspots and hot moments in soil was recently introduced and is characterized by increased activity and cellular interactions (Kuzyakov and Blagodatskaya 2015). It has been proposed that by contributing to the creation of these hotspots within soil, fungal hyphae change HGT dynamics within the soil community (Pratama and van Elsas 2019). To address the effect of spatial heterogeneities on HGT dynamics, tools are required that allow the tracking of MGEs at relevant spatial scales.

HGT in communities does not happen independently of other microbial interactions. For example, type six secretion (T6S) activity has been shown to be co-regulated with the competence machinery of V. cholerae; DNA released by susceptible, lysed cells was shown to be horizontally acquired by the predatory cells (Borgeaud et al. 2015). In a follow-up study, it

was shown that V. cholerae can acquire large genomics regions (larger than 150 kilobases) in a T6S-dependent manner (Matthey et al. 2019). Similarly, T6S-mediated killing of E. coli by Acinetobacter baylyi was shown to increase horizontal acquisition of genetic material by A. baylyi (Cooper, Tsimring and Hasty 2017). To study these microbial interactions in more complex systems or environments, it will be necessary to develop tools that allow simultaneous monitoring of a wide range of HGT events and other types of microbial interactions such as contact-dependent killing.

Together, these examples demonstrate that in order to understand global HGT dynamics in microbial communities, a variety of factors ranging from environmental and biotic interactions to the presence of other HGT machineries and MGEs should be considered. While more targeted approaches described in this review alone might not allow for this global approach, emerging metagenomics approaches are now opening the door to studying these aspects of HGT in microbial communities. Additionally, as better tools for bacterial single-cell genomics and RNA-seq become available, it should be possible to address HGT dynamics in the context of other processes happening in communities.

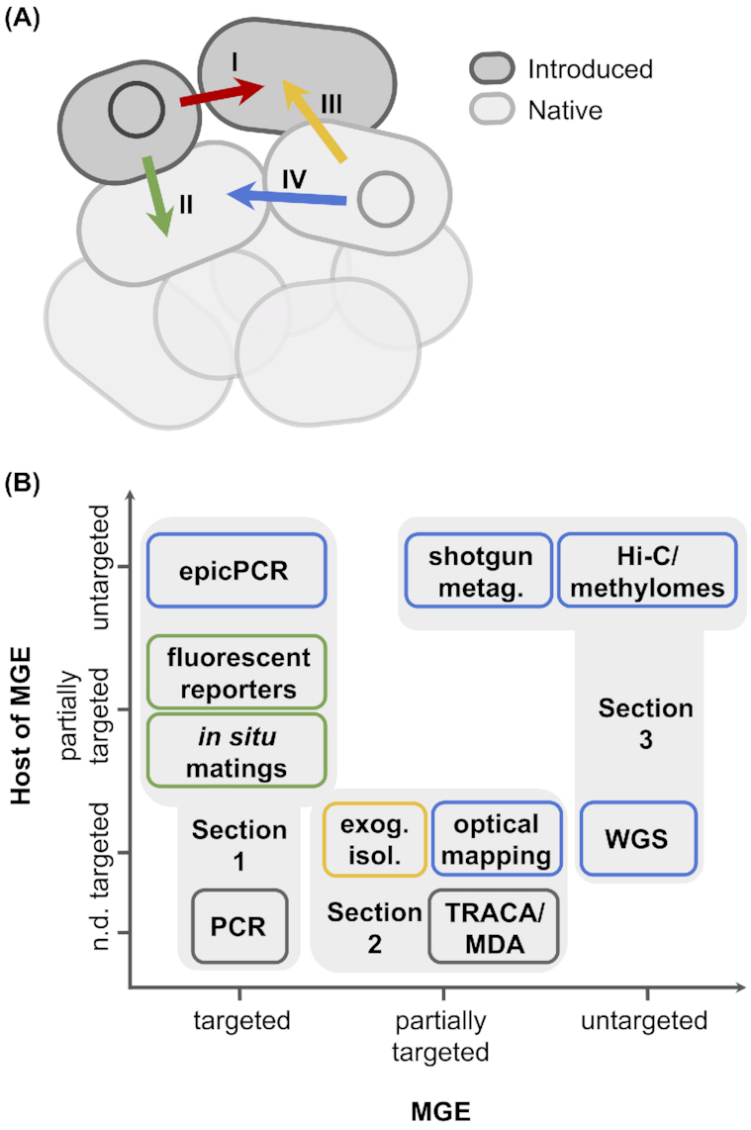


Figure 1.2 Considerations and approaches to studying HGT in microbial communities. (A) Potential transfer directions in microbial communities. MGEs (intracellular grey circles) can move between donor (D) and recipient (R) cells that have been introduced into a native community (transfer direction I, red arrow). Alternatively, the introduced donor can transfer MGEs to native recipients within the community (transfer direction II, green arrow). MGEs can also move from a native donor to an introduced recipient (transfer direction III, yellow arrow). Last, MGEs can be exchanged between native members of the community (transfer direction IV, blue arrow). (B) Experimental approaches detecting MGEs and/or identifying their host in microbial communities. Color outlines indicate the highest order transfer direction potentially detected based on Figure 1.2A; grey outlines indicate the inability to detect transfer direction. exog. isol. = exogenous plasmid isolation, TRACA = transposon-aided capture, MDA = multiple displacement amplification, WGS = whole genome sequencing, shotgun metag. = shotgun metagenomic sequencing. n.d. = not determined.

Experimental approaches to interrogate MGE transfer in microbial communities

Within the context of microbiomes, MGEs can be exchanged either between experimentally introduced organisms (transfer direction I), between introduced and native organisms (transfer directions II and III) or between native organisms (transfer direction IV) (Figure 1.2A). While transfer directions I-III might be of particular interest for assessing the risk of an introduced organism acquiring an undesirable gene from the native community or vice versa, transfer direction IV is particularly important when one wants to understand how microbes evolve within the context of natural microbial communities. A variety of techniques is available to researchers to track MGEs in microbial communities, some more suitable for certain research questions than others. For example, some techniques are appropriate for tracking plasmids, while others might be designed to track the overall population of mobile elements.

In this review, we give an overview of techniques available to study transfer of mobile DNA in microbial communities (Figure 1.2B). In Section 1, we discuss techniques that target pre-determined MGEs, i.e. that require pre-existing knowledge about specific elements. In Section 2, we discuss techniques that are partially targeted and allow the study of specific groups of MGEs, such as plasmids that can be maintained in a specific host, without requiring pre-existing knowledge about specific elements. However, certain research questions might require a broader view of MGEs in the community, such as those questions addressing the potential for microbial evolution within the context of microbial communities. To this end, in Section 3, we discuss various genomics and metagenomics techniques that have the potential to provide information about a wider range of mobile elements and their hosts and thus afford

a more untargeted approach to studying HGT in microbial communities. The novel metagenomics techniques described herein, 3C-related methods and the analysis of methylation profiles, are especially helpful for the study of DNA transfer in microbiomes since they have the potential to assign extrachromosomal elements such as plasmids to their host genomes, which remains challenging with traditional metagenomics approaches.

SECTION 1: TECHNIQUES TARGETING KNOWN MGEs

Detection of MGEs and Associated Genes of Interest in Microbiomes

In some cases, the potential HGT of specific genes or groups of genes is of interest. For example, the dissemination of ARGs is a world-wide problem and the presence of ARGs has been studied in diverse microbiomes ranging from those in wastewater (Karkman et al. 2018) and hospitals (Kamathewatta et al. 2019) to those associated with various types of food (Boehme et al. 2004; Devirgiliis et al. 2014; Fraqueza 2015; Abriouel et al. 2017; Quintieri, Fanelli and Caputo 2019). While experimental confirmation of antibiotic resistance is a common approach to confirming a suspected antibiotic resistance, this relies on the isolation and cultivation of specific microbial species, which is often not feasible. However, for clinical or public health surveillance, knowing the antibiotic resistance profile of unculturable or slowly growing community members can be critical. To avoid issues related to culturability, PCR-based approaches of candidate ARGs can be used to interrogate the presence of panels of pre-determined ARGs (Nawaz et al. 2011; Guo et al. 2017). In a variety of microbiomes, the PCR-based approach has also been directed at various MGEs including plasmids and integrons (Smalla et al. 2000b; Gillings et al. 2015) that are frequently associated with ARGs,

such as those associated with river waters (De la Cruz Barrón et al. 2018) and full-term infant guts (Ravi et al. 2018). A recent review of MGEs associated with ARGs showed that this is indeed a viable approach (Partridge et al. 2018).

The described PCR-based approaches rely on updated and stable databases of both MGEs and genes of interest, a myriad of which have been developed (Table 1.1). For example, databases allow researchers to search for common motifs of MGE types, therefore helping the detection of these elements by PCR. Automated screening of published genomes and manual curation of published phenotypic studies of isolated species and their antimicrobial resistance profiles further helps in generating these databases. Surveillance programs such as WHO AWaRe (Assessing Worldwide Antimicrobial Resistance Evaluation) and INFORM (International Network for Optimal Resistance Monitoring), CDC/FDA/USDA NARMS (National Antimicrobial Resistance Monitoring System), Pfizer ATLAS (Antimicrobial Testing Leadership and Surveillance) are not included in this table, but they use similar databases to track these elements epidemiologically. Many MGE databases overlap in their contents, but use different filtering and inclusion methods. We do not show specialized databases that only contain ARGs within a specific genus or against specific antibiotics. Additional databases combine and filter those found in databases listed in Table 1.1, such as ARG-ANNOT (uses NCBI) (Gupta et al. 2014), PATRIC (uses CARD, NCBI, manually curated) (Davis et al. 2016; Antonopoulos et al. 2019), DeepARG-db (based on CARD, ARDB (Liu and Pop 2009), UniProt (Uniprot Consortium 2019)) (Arango-Argoty et al. 2018), SARG (CARD, ARDB and NCBI-NR (Yin et al. 2018)) and BacMet (based on

NCBI Genbank, UniPortKB, Transporter Classification Database, and publications; separated into experimentally confirmed and predicted in NCBI-nr; manually curated) (Pal et al. 2014).

Table 1.1 Public databases of genes of concern and mobile genetic elements.

Genes or MGEs of Concern	Name	Description	Reference
Antibiotic Resistance (reviewed in Hendriksen *et al.* 2019)	CARD	Functional ARGs and mutations using Antibiotic Resistance Ontology. Manually curated	(Alcock *et al.* 2019)
	ResFinder	Acquired antibacterial resistance or chromosomal mutations	(Zankari *et al.* 2012)
	MEGARes	~8,000 antibiotic, metal, and biocide resistance sequences. Manually curated and formatted for bioinformatics	(Lakin *et al.* 2017)
	NCBI National Database of Antibiotic Resistant Organisms (NDARO)	Public collection of antibiotic resistance sequences, including Resfinder, AMRFinder, CARD, and individual contributions. Includes phenotypic antibiograms, as designed by EMBL AMR data hub	
	EMBL-ENA Antimicrobial Resistance data hub	Central infrastructure to share and analyze pathogen sequencing, including phenotypic antibiograms and metadata about source	(Matamoros *et al.* 2019)
	UniProt	Contains general protein sequences with ortholog categories. Manually curated and automated literature association	(UniProt Consortium 2019)

Table 1.1, continued Public databases of genes of concern and mobile genetic elements.

Genes or MGEs of Concern	Name	Description	Reference
Antibiotic Resistance (reviewed in Hendriksen *et al.* 2019)	ABRES, ResFams, SARGfams (from ARGs-OAP, an ARG database)	Hidden Markov Models for antibiotic resistance	(Gibson, Forsberg and Dantas 2015; Xavier *et al.* 2016; Yin *et al.* 2018)
	RED-DB	From EMBL, DDBJ, and Genbank. Genes clustered on 70% average nucleotide and associated with host organisms. Manually curated	(D'Andrea)
Virulence Factors (VFs)	NCBI Pathogen Detection	Includes NCBI NDARO above as well as pathogenic isolate sequences	
	Victors	VFs of human and animal pathogens. Manually curated	(Sayers *et al.* 2019)
	Virulence Factor Database	Medically significant bacterial pathogens and VFs	(Liu *et al.* 2019a)

Table 1.1, continued Public databases of genes of concern and mobile genetic elements.

Genes or MGEs of Concern	Name	Description	Reference
Plasmids	ACLAME	Gene ontology specific for MGEs, especially phage, prophage, and plasmids	(Leplae, Lima-Mendez and Toussaint 2010)
	NCBI plasmids, PLSDB, PlasmID	NCBI database contains public submissions. PLSDB and PlasmID contain curated plasmid sequences from NCBI-nt	(Brooks, Kaze and Sistrom 2019; Galata *et al.* 2019)
	PubMLST	Part of BIGSdb, containing multi-locus sequence typing associated with phenotype and incompatibility groups	(Jolley and Maiden 2010)
	PlasmidFinder and pMLST	Sequences and multilocus sequence typing of plasmids originally derived from PubMLST. Limited to enterobacteriaceae	(Carattoli *et al.* 2014; Jolley, Bray and Maiden 2018)
Genomic Islands (GIs)	IslandViewer 4 databases: IslandPick and Islander	Manually curated	(Bertelli *et al.* 2017)
	ICEberg	Curated and automated database of bacterial ICEs, integrative and mobilizable Elements, and *cis*-mobilizable elements	(Liu *et al.* 2019b)
	DarkHorse	DarkHorse results from isolate genomes in NCBI-nr, JGI, and JCVI	(Podell, Gaasterland and Allen 2008)
Genomic Islands (GIs) continued	PAIDB	Pathogenicity and antimicrobial resistance islands from literature, and those with similar GC bias	(Yoon, Park and Kim 2015)
	Predict GI database	Prokaryotic GIs. Part of Seqword with EuGI for eukaryotic GIs and hislands for human GIs	(Pierneef *et al.* 2015)

Table 1.1, continued Public databases of genes of concern and mobile genetic elements.

Genes or MGEs of Concern	Name	Description	Reference
Genomic Islands (GIs)	ImmeDB	MGEs associated with intestinal microbiome	(Jiang *et al.* 2019)
Transposons, integrons, and repetitive regions	GyDB	Phage and transposons, especially LTR-retroelements	(Llorens *et al.* 2011)
	The Transposon Registry	Nomenclature and database of transposons and mobile genomic islands. Public submissions	(Tansirichaiya, Rahman and Roberts 2019)
	Repbase	Automated database on recon and LTR-FINDER. Manually curated	(Bao, Kojima and Kohany 2015)
	repetDB	Automated database from REPET. Linked to Repbase and GyDB	(Amselem *et al.* 2019)
	SINEBase	Short interspersed elements eukaryotes	(Vassetzky and Kramerov 2013)
	ISfinder	Bacterial and archaeal insertion sequences. Requires written authorization to download	(Siguier *et al.* 2006)
	DFAM	HMM database of repetitive sequences, including eukaryotic	(Hubley *et al.* 2016)
	INTEGRALL	Integron and related sequences and gene cassettes, such as ARGs	(Moura *et al.* 2009)

However, despite these vast resources several problems still remain to be overcome. First, as indicated above, database-centered approaches to identifying ARGs rely on database accuracy and completeness (Boolchandani, D'Souza and Dantas 2019). Thus, unknown ARGs will be missed. Additionally, results from PCR-based approaches do not always correlate with

actual phenotypic antibiotic resistance profiles (Aristimuño Ficoseco et al. 2018). Various reasons for the discordance between genotypic and phenotypic antibiotic resistance profiles have been put forth (Hughes and Andersson 2017). For example, bacteria can be intrinsically resistant to antibiotics due to slow antibiotic uptake or efficient antibiotic efflux due to efflux pumps. Alternatively, the genomic context of a given resistance gene can affect its phenotypic expression. In addition, these approaches merely detect the presence of particular ARGs and MGEs without information about their genomic context or host species, making risk assessment of transfer to unwanted species impossible. Nevertheless, as our understanding of antibiotic resistance and HGT improves, these databases will continue to expand and inform computational tools that can identify these sequences in metagenomes for surveillance and tracking purposes.

Linking genes and MGEs of interest to their host

While these approaches have the potential to investigate ARGs and MGEs in communities, they fail to associate these elements with their hosts. One method to overcome this challenge, epicPCR (Emulsion, Paired Isolation and Concatenation PCR) was developed (Spencer et al. 2016). epicPCR is based on the premise of encapsulating individual cells in polyacrylamide beads before lysis. The DNA is released into the confinement of the bead and amplified by fusion PCR, which can be used to create a physical linkage of phylogenetic markers such as 16S rRNA genes with functional genes of interest. Control experiments showed that this technique was able to detect representatives of all major bacterial groups of a lake water community, with additional lysis steps improving detection of certain groups of

bacteria including Actinobacteria and Bacteroidetes. According to the authors, millions of cells can be screened this way for the cost of a single sequencing library preparation (Spencer et al. 2016). Given its ability to link genes or DNA sequences of interest to their bacterial host, this technique could hypothetically be used to investigate the linkage of specific MGEs of interest, including plasmids, with their host genomes. Since it does not require manipulation of the MGE or cultivation of its host, epicPCR has the potential to detect HGT events in all four transfer directions listed in Figure 2. Previously, this technique has been used to assign ARGs and a class I integron to their hosts in wastewater treatment plant communities (Hultman et al. 2018). Comparing influent and effluent water, this study detected putative HGT of resistance genes between operational taxonomic units. However, the authors concede that they were not able to definitely attribute the change of association between ARGs and 16S rRNA genes to HGT given the observed experimental variation. Further development of this technique or combination with other techniques discussed in this review might improve its power to detect HGT events.

Detecting transfer of plasmids to culturable members of microbiomes

Some of the initial experiments that were aimed at estimating the potential for ARGs to spread through a microbial community involved the introduction of donor species carrying ARGs on known mobile elements into microbiomes and selecting recipients on the respective antibiotic. In this type of experiment, it is important to be able to distinguish between donors and recipients. If both are co-introduced into the community (transfer direction I) and express different antibiotic/metal resistances, it is possible to enumerate recipients by selective plating

(Bale, Day and Fry 1988; Richaume et al. 1992). Alternatively, preventing expression of the MGE's resistance marker in the donor eliminates the need for additional, selectable markers in the potential recipients (Top et al. 1990). For these experiments it is necessary that the introduced donors and recipients can survive in the investigated environment long enough to allow horizontal transfer to take place.

In other cases, it is of interest to what degree native bacteria of a particular environment can act as hosts for a mobile element originating from an introduced donor (transfer direction II). In such experiments, it is also crucial to be able to distinguish original donors from potential recipients, especially if transfer efficiencies are low and the number of introduced donors is high. This can be done by ensuring that the donor has a limited lifespan within the probed community, while persisting long enough for transfer to occur. For example, an early study measured the transfer of a plasmid-encoded tobromycin resistance gene from E. coli SM10λpir, a strain capable of efficient plasmid mobilization, to a soil microcosm. This experiment was able to identify tobromycin-resistant recipients native to the soil microbiome after 25 days due to the disappearance of the introduced E. coli that was not adapted to the soil environment (Henschke and Schmidt 1990). On the other hand, it might be of interest to test transfer dynamics involving a donor that is able to survive in the community long-term. In these cases, an additional selection criterion is required to efficiently distinguish recipients from donors and non-recipients. Auxotrophic donors can be counter-selected for by plating on selection media lacking the compound the donor cannot self-produce. This approach was used to test the transfer of RP4 to the native microbiome present in a lake water microcosm (Sengeløv and Sørensen 1998). RP4 is a conjugative plasmid from the

incompatibility group Pα (IncPα) that has a long history of use as a genetic tool (Jacob and Grinter 1975; Babic, Guérout and Mazel 2008). As the authors of this study point out, there are several potential problems that could lead to an underestimation of the true extent of plasmid transfer: recipients that are native to the interrogated microbiome could be auxotrophic themselves; recipients could be prototrophic, but unable to express the transferred resistance gene or grow under laboratory conditions; the auxotrophic donor's ability to transfer plasmids in the tested environment could be impaired as compared to an otherwise isogenic non-auxotrophic strain. Another option to select against donors is to use bacteriophages that specifically eliminate donor cells from the population and allow the detection of native recipients (Smit et al. 1991; Richaume et al. 1992; Daane et al. 1996). Last, in a study comparing different donor counter-selection methods, counter-selection via donor-encoded suicide genes was found to be successful in filter-matings, but not in microcosm experiments (Sengeløv and Sørensen 1998).

The in situ mating approaches described above have not only been applied to detecting the transfer of plasmids to native members of microbiomes, but also of other genetic elements subject to HGT. One study tested the transfer efficiency of gene transfer agents in marine environments using antibiotic selection of the transconjugants (McDaniel et al. 2010). Elimination of the donor was not necessary in this study, since the authors subjected the communities directly to the MGE.

Detecting transfer of plasmids to unculturable members of microbiomes

As hinted at in the previous section, one common problem with the in situ mating approaches to tracking MGEs in communities is that they rely on the cultivation of recipients under laboratory conditions, which might lead to an underestimation of the true extent of HGT in communities. In experiments in which the interest is only in specific bacteria that are culturable under known conditions, this drawback might not be prohibitive. However, often all possible recipients within a community are of interest. In these instances, culture-independent approaches to identifying MGE recipients are required.

This class of approaches relies largely on the expression of fluorescent reporter genes in the cells that harbor a particular mobile element. These approaches have been extensively reviewed (Sørensen et al. 2005; Pinilla-Redondo et al. 2018), thus we will only give a brief overview. As with antibiotic resistance markers, lack of expression of fluorescent reporters in the recipient can be responsible for false negatives and lead to an underestimation of the true extent of HGT in a community. However, unlike with antibiotic resistance markers, cultivation under laboratory conditions is not required for many of these techniques, because reporter expression can be evaluated on a cellular level in a high-throughput manner. The next few paragraphs will discuss the various reporter methods that have been used to investigate HGT in microbial communities.

Generally, these approaches are targeted at MGEs that can be isolated, genetically manipulated and re-introduced into a donor. In addition, as with the in situ mating approaches, steps need to be taken to distinguish between donors and recipients. This was originally done by repressing reporter expression in the donor, for example by utilizing a lactose-repressible

promoter. One of the earliest studies employing this approach investigated how a plasmid encoding a GFP-reporter under the control of a lac promoter was transferred from an introduced Pseudomonas putida strain to native recipients of a marine bacterial community (transfer direction II) (Dahlberg, Bergström and Hermansson 1998). In this study, GFP expression was repressed in the donor cells, which chromosomally encoded the repressor LacI, and it was de-repressed in recipients, which did not encode LacI and could thus be enumerated by epifluorescence microscopy. While the single-reporter approach can be powerful, it does not distinguish donors and native non-recipients based on fluorescent signals. As such, dual-reporter approaches have been developed that distinguish the donors via a second, chromosomally encoded fluorescent reporter (Mølbak et al. 2003; Klümper et al. 2015). This creates three different pools of bacteria: donors expressing the chromosomally encoded fluorescence, non-fluorescent potential recipients and recipients expressing the plasmid-encoded fluorescence.

For higher throughput studies, the recipients can also be enumerated by flow cytometry and even isolated by fluorescence-activated cell sorting as was done for an experiment that introduced plasmid donors into a barley rhizosphere (Musovic et al. 2006). Subsequent 16S rRNA analysis of the sorted recipients showed broad range transfer, including from the Gram-negative donor to a Gram-positive recipient. One disadvantage of the flow cytometry-based approach of enumerating recipients is its failure to separate true recipients from cells that have vertically inherited the MGE. An alternative approach has thus been employed based on minimal-cultivation solid-surface matings followed by the enumeration of fluorescent microcolonies in a community, since it is presumed that each microcolony

originated from a single recipient cell (Musovic et al. 2010; Klümper et al. 2014). Stereomicroscopy coupled with micromanipulation can then be used to isolate recipients. The feasibility of this approach in natural environments depends on its complexity and ease of microscopic investigation. In addition, it was found that while stereomicroscopy made for a more efficient protocol, confocal microscopy detected higher rates of conjugation (Musovic et al. 2010). Last, mating approaches that rely on minimal cultivation re-introduce potential biases associated with growing complex microbial communities under laboratory conditions.

Another potential blind spot to consider is the transfer of plasmids from primary recipients to secondary recipients. For example, if the plasmid of interest can be efficiently shared between the potential recipients, high numbers of transconjugants sorted by fluorescence-activated cell sorting might lead to an overestimation of initial transfer rates between the donor and the primary recipients. This experimental gap is starting to be addressed through methods using barcoded plasmids that allow the distinction between primary and secondary transfer events (Bakkeren et al. 2019). Other approaches to exclusively estimating conjugation rates from donor cells to primary recipient cells exploit the different types of plasmid transmissibility. In one study, the transfer of a mobilizable, but not self-transmissible plasmid was induced by a second plasmid present in donor cells that itself is non-mobilizable due to lack of a suitable origin of transfer (del Campo et al. 2012).

In addition to the classification of the recipient pool, it might also be of interest to characterize effects of spatial organization on HGT in communities. Flow-cytometry approaches require the breaking apart of communities into individual cells causing a disruption of spatial organization. In contrast, combining the dual-reporter approach with

confocal microscopy allows the interrogation of spatial dynamics of MGE transfer. For example, confocal microscopy was used to investigate the spatial distribution of plasmid recipients on alfalfa sprouts and it was found that transfer to native community members was dominant on the roots (Mølbak et al. 2003).

To really understand HGT dynamics in communities, it is necessary to not only understand spatial aspects, but also temporal dynamics. Microfluidics can continuously supplement a growing community with nutrients over extended periods of time, while allowing the observation of cellular changes. To this end, microfluidics approaches have been applied to the question of HGT in microbial communities. For example, a study investigating the transfer of the broad host range plasmid RP4 from a Pseudomonas putida donor to a recipient community made up of activated sludge bacteria combined the dual-fluorescence approach with temporal imaging in a microfluidics device and 16S rRNA sequencing of the final recipient pool (Li et al. 2018a). It was found that more than 46% of the genera present in the activated sludge community were permissive to the RP4 plasmid. While these microfluidic devices allow a temporal study of the transfer dynamics in communities, the community needs to be grown in the microfluidic device under known conditions, which reintroduces the limitation imposed by culturability.

While most of the other reporter gene approaches discussed above allow the circumvention of problems associated with recipient culturability, they are still subject to shortcomings, which have also been recently reviewed (Pinilla-Redondo et al. 2018). One of such shortcomings is associated with expression and activity of the reporter gene in the new host. Similar to in situ mating approaches, this might cause an underestimation of the true

extent of HGT within a community. In addition, if detectable levels of the fluorescent reporter are not reached before the MGE is lost from a recipient cell, these transient transfers will be missed. The reporter gene approach also relies on the genetic manipulation of donors and MGEs limiting the kind of questions that can be answered with this approach. While alternative quantitative PCR approaches have been developed to address this shortcoming by detecting changes in donor-plasmid ratios without the need for genetic manipulation (Bonot and Merlin 2010; Merlin et al. 2011), they cannot be used to determine the identities of recipients. Another limitation of the reporter gene approaches is the reliance on fluorescent proteins, which might not be functional in all recipients or environmental conditions. For example, anaerobic or acidic environments often interfere with proper functioning of fluorescent proteins. This has been started to be addressed, for example by exposing cells to oxygen for aerobic fluorescence recovery before analysis by flow cytometry (Pinilla-Redondo, Riber and Sørensen 2018). Last, strong autofluorescence of environmental samples can impede the detection of native recipients (Bellanger et al. 2014). Overall, while the fluorescent reporter approaches have been powerful tools, their associated limitations cause them to provide a still partially targeted view on potential hosts of the introduced MGEs.

SECTION 2: PARTIALLY TARGETED TECHNIQUES THAT ALLOW THE STUDY OF SPECIFIC GROUPS OF MGES

Techniques Targeting Specific Groups of MGEs

While certain research questions are sufficiently addressed by focusing on known MGEs, other questions require a broader view. For example, it might be of interest to determine the frequency with which a species acquires ARG-containing MGEs from a given environment. In other cases, it might be of interest to characterize the diversity of mobile elements carrying a gene of interest within an environment. In cases such as these, techniques targeting specific groups of MGEs can be appropriate. These groupings can be based on host range, genetic cargo or element type.

Identifying plasmids that can be transferred to specific recipients

As mentioned above, sometimes it is important to know whether a specific bacterium, such as a pathogen, is at risk of acquiring MGEs from the surrounding microbial community. One area in which this knowledge is of particular importance is the emergence and risk management of antibiotic-resistant bacteria in the hospital environment, which while recognized as important for slowing the global increase of antibiotic resistance, has so far been understudied (Lerminiaux and Cameron 2019). However, one of the few studies addressing this question detected a diverse reservoir of carbapenemase-encoding plasmids in hospital plumbing and even found evidence for horizontal transfer of a carbapenemase-encoding plasmid between Enterobacter cloacae and Klebsiella pneumoniae in a hospital sink

(Weingarten et al. 2018). In this case, E. cloacae had been introduced into the sink by one patient and was able to establish itself in the surrounding community. Klebsiella pneumoniae was later introduced into the same sink community by another patient and engaged in conjugative plasmid exchange with E. cloacae.

Exogenous plasmid isolation has been employed to investigate the mobile plasmid fractions of various microbiomes that can be transferred to a specific, introduced recipient (transfer direction III). A variety of questions can be addressed by different variations of the exogenous plasmid isolation experimental design. For example, bi-parental matings between bacteria from porcine manure slurries as donors and Escherichia coli or Pseudomonas putida as recipients identified a variety of plasmids belonging to the incompatibility group Q (IncQ) in the porcine manure bacterial community (Smalla et al. 2000a). Curiously, IncQ plasmids are usually mobilizable, but not self-transmissible, indicating that mobilizing plasmids must have also been present in the donor community. The ability of microbial communities to mobilize an IncQ-derived plasmid, pD10, was previously demonstrated for the microbial community of a river epilithon (algae growing on rocks) (Hill, Weightman and Fry 1992). In this experiment, tri-parental matings including a pD10-carrying P. putida donor, a plasmid-free P. putida recipient and homogenized bacteria isolated from the epilithon were carried out to identify plasmids in the microbial community that could mobilize pD10. As with the in situ mating approaches described above one important consideration for exogenous plasmid isolation is the ability of the chosen recipient strain to colonize and persist in the community of interest. For example, a study involving a marked P. fluorescens strain, which is known to

be able to colonize and persist in the natural sugar beet phytosphere, isolated a handful of likely related plasmids (Lilley and Bailey 1997).

Exogenous plasmid isolation approaches rely on the presence of selectable markers on the transferred MGE which can be expressed in the chosen recipient strain. Mobile elements encoding for more complex phenotypes that might not be straightforward to select for can therefore be easily missed. In addition, this approach relies on favorable conditions for transfer and maintenance of the plasmid in the new host. For example, in a study employing exogenous isolation from porcine manure, more recipients were originally identified than the ones that were retained after three rounds of sub-culturing. This indicated that the pool of captured MGEs contained additional elements that could not be maintained in the E. coli host long-term (Binh et al. 2008). In contrast, a recently developed method employing an E. coli recording strain that stably integrates DNA sequences from acquired plasmids into a genomic CRISPR array does not rely on maintaining the MGE in the cell and can thus provide information on more transient transfer events (Munck et al. 2020).

Overall, using exogenous plasmid isolation to identify plasmids in microbial communities that can be transferred to a specific host is almost guaranteed to underestimate the range of MGEs present in a microbial community. In addition, this method fails to identify the original hosts of the isolated plasmids. However, as described above this technique is still of great interest to the scientific and medical community, for example if one wants to assess the risk that a particular species that has been introduced into an environment acquires specific resistance genes.

Culture-independent isolation of plasmids from microbiomes

In other cases, the overall pool of MGEs from a given community are of interest, rather than only those elements that can be transferred to a particular recipient of interest. For example, a bioinformatic algorithm designed to detect chromosomal mobile elements in human gut microbiomes identified many elements containing genes with predicted adaptive functions in this environment (Jiang et al. 2019). Thus, having insight into the overall pool of mobile elements in a given environment has the potential to shed light on the adaptive pressures it exerts on its resident bacteria. In the next few paragraphs, we will give an overview of culture-independent methods that have been developed to investigate the plasmidome of microbial communities, which represents one important constituent of the overall MGE pool. Since the described techniques specifically isolate circular episomal elements, such as circular plasmids, and each have technical limitations as outlined below that make the determination of the complete plasmid content of a sample unlikely, they fall into the category of partially targeted techniques (Figure 1.2B).

Transposon-aided capture (TRACA) of extrachromosomal, circular elements such as plasmids allows the capture of elements that might not encode a selectable marker or might not be stably maintained in a lab-adapted host. This method has been used to investigate various plasmidomes including those of the human gut (Jones and Marchesi 2007) and dental plaques (Warburton et al. 2011). To isolate the plasmidome, chromosomal DNA is sheared and removed with plasmid-safe DNAse. While it is possible to directly sequence the isolated DNA, contaminating chromosomal DNA that remains after exonuclease treatment can complicate assembly of sequencing reads (Zhang, Zhang and Ye 2011; Dib et al. 2015).

Instead, TRACA calls for tagging the remaining circular DNA with a Tn5 transposon containing a selectable marker and transforming it into E. coli (Figure 1.3). Since the transposon also contains its own origin of replication suitable for E. coli, the transformed plasmids can be maintained in the cells regardless of their native origin of replication. Downstream analyses include sequence determination of isolated plasmids using primer walking (Jones and Marchesi 2007; Warburton et al. 2011; Zhang, Zhang and Ye 2011) and phenotypic analyses, such as determination of antibiotic resistance profiles (Delaney, Murphy and Walsh 2018). This approach allows the isolation of plasmids lacking selectable markers and did not show a GC bias (Jones and Marchesi 2007). It thus has the potential to help fill the picture of which plasmids are present in a given microbial community. However, only relatively small plasmids (3 - 10 kb) were isolated by TRACA (Jones and Marchesi 2007; Warburton et al. 2011). The authors of both studies pointed out that it is still under investigation whether this points to a limitation of this method or whether the studied microbiomes contained predominantly small plasmids. Additionally, linear plasmids remain undetected with TRACA since they require a different type of replication machinery and can thus not be maintained in the E. coli host (Warburton et al. 2011; Dib et al. 2015).

Another approach to plasmid isolation from uncultured microbiome samples is the use of CsCl-ethidium bromide gradient ultracentrifugation. This technique allows the purification of plasmid DNA and its separation from proteins, RNA and chromosomal DNA (Garger, Griffith and Grill 1983). It has been used to enrich plasmid DNA in samples of activated sludge communities (Sentchilo et al. 2013). Since DNA is not amplified before analysis, large amounts of DNA are required, and thus low abundance plasmids are likely to be overlooked.

Multiple displacement amplification (MDA) was found to be another promising alternative (Delaney, Murphy and Walsh 2018). As with TRACA, exonuclease treatment is used in this technique to eliminate sheared chromosomal DNA and other linear DNA molecules. Random hexamer primers are then added to the sample along with phi29 polymerase. When the polymerase comes into contact with the primer of the adjacent amplified fragment, it displaces one of the strands and continues DNA synthesis. The displaced single strand of DNA is in turn bound by new primers and amplified (Figure 1.3). This process leads to a branching network of amplified DNA (Dean et al. 2001). The final amplified plasmidome can then be analyzed in various ways, including transformation into an E. coli host for further phenotypic characterization (Delaney, Murphy and Walsh 2018) or high-throughput sequencing for bioinformatic analyses (Kav et al. 2012, 2019; Li et al. 2012; Norman et al. 2014). MDA has been used to interrogate the plasmid contents of environmental (Li et al. 2012; Norman et al. 2014) or animal-associated communities (Kav et al. 2012, 2019). The amplification of DNA prior to downstream analyses can be viewed as a double-edged sword. On one hand, it can help generate enough DNA for high-throughput sequencing (Li et al. 2012). On the other hand, it can bias plasmid detection towards smaller plasmids (Norman et al. 2014). However, including an electroelution step in the MDA protocol allowed the recovery of larger plasmids (median size of 30 kb) from wastewater samples (Norman et al. 2014).

For both TRACA and MDA, the need for shearing chromosomal DNA reduces the chances of isolating and identifying very large plasmids, which are more likely to also be sheared than smaller plasmids. In addition, linear plasmids are excluded from detection in

both methods since they are degraded during treatment with circular plasmid-safe DNase (Dib et al. 2015). Last, since these techniques isolate plasmids from bulk extracted DNA, they fail to provide information about their hosts in the community context and thus cannot by themselves be used to interrogate actual transfer events in communities. A recent study compared six methods for extracting plasmidomes (exogenous plasmid isolation, alkaline lysis- and commercial kit-based methods as well as TRACA and MDA) from chicken cecal microbial communities consisting of hundreds of bacterial species (Delaney, Murphy and Walsh 2018). The authors found that exogenous plasmid isolation was the most consistent method for isolating a range of plasmids (at least four different plasmids), while MDA isolated a slightly larger number of diverse plasmids. The best plasmidome isolation method probably depends on the nature of the respective community and likely requires combining various approaches.

Comparing MGEs from species carrying a common gene of interest

From an epidemiological standpoint, one might only be interested in the specific MGEs associated with an antibiotic-resistant infection. This was the case in a study on extended-spectrum β-lactamase (ESBL)-producing Enterobacteriaceae isolated during an outbreak in a neonatal intensive care unit. Here, the authors investigated whether the presence of two ESBL-producing strains, Klebsiella pneumoniae and E. coli, in the same patient was due to horizontal transfer of the ESBL-carrying plasmid (transfer direction IV) (Bikkarolla et al. 2019). The study employed a method termed optical mapping, which uses fluorescent probes that differentially label AT- and GC-rich regions of the MGE of interest and

fluorescence microscopy to visualize individual DNA molecules in nanofluidic devices. The authors coupled this approach with a CRISPR-guided identification of the location of the ESBL gene on each plasmid. Comparison of the resulting plasmid bar codes and locations of the ESBL genes led to the conclusion that HGT of the ESBL-carrying plasmids had likely not occurred between the co-colonizing strains. Other methods for analyzing the isolated MGE DNA include sequencing, restriction fragment length polymorphism analysis as well as pulsed-field gel electrophoresis. However, the authors of the described study posit that their optical mapping approach constitutes a faster and more efficient way (Bikkarolla et al. 2019). As such, this method might also be of interest if one wants to analyze and compare MGEs isolated from microbiomes using previously described techniques.

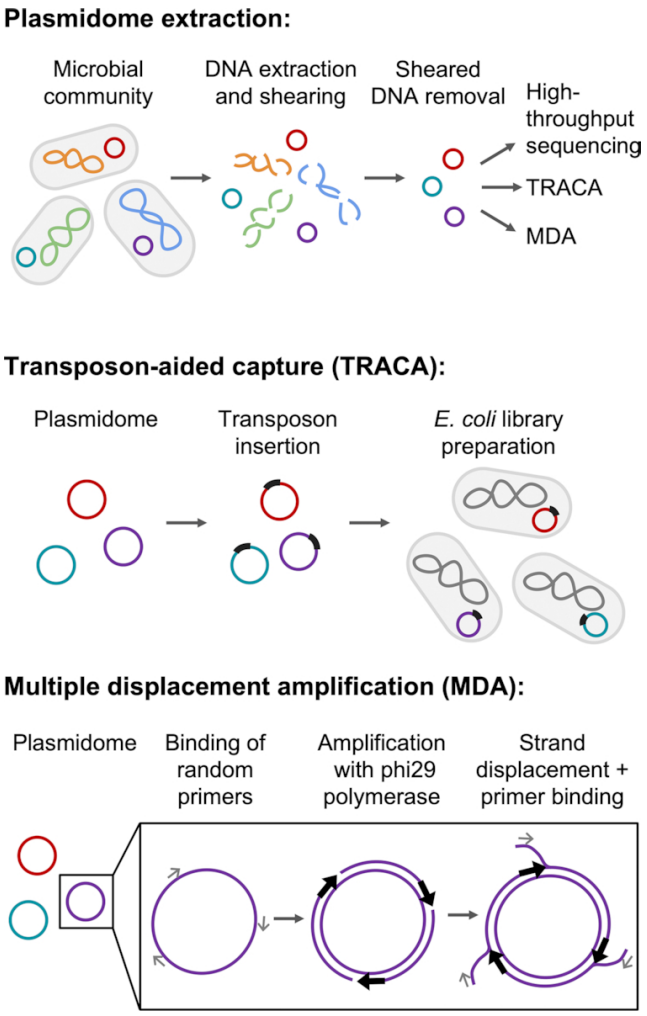


Figure 1.3. Transposon-aided capture (TRACA) and multiple displacement amplification (MDA) of plasmidomes rely on common pre-processing steps of metagenomic DNA. For both techniques, metagenomic DNA is extracted from the microbial community of interest. The extracted DNA is subjected to shearing and sheared DNA is removed with a circular plasmid-safe DNAse. Quantification of 16S rRNA gene sequences can be used to control for the successful removal of chromosomal DNA. The extracted plasmidome can then be subjected to high-throughput sequencing, TRACA and/or MDA. In TRACA, the plasmidome DNA is tagged with a Tn5 transposon containing a selectable marker (black insert) and transformed into E. coli. The isolated plasmids can then be analyzed in various ways including sequence determination by primer walking or phenotypic characterization. In MDA, the extracted plasmidome is amplified using random hexamer primers (short, grey arrows) and phi29 polymerase (thick, black arrows). When phi29 contacts the primer of the adjacent fragment, it displaces the newly synthesized strand and continues synthesis. The displaced strand is then bound by new random primers and further amplified, resulting in a branched network of amplified DNA, which can be analyzed by various techniques, such as transformation into E. coli for phenotypic characterization or high-throughput sequencing for bioinformatic analyses.

SECTION 3: UNTARGETED TECHNIQUES FOR INVESTIGATING MGE TRANSFER BETWEEN NATIVE MICROBIAL COMMUNITY MEMBERS

The techniques described in the previous sections target known or specific groups of MGEs. However, discovering novel mobile elements, genes of interest, and mechanisms of transfer requires techniques that provide less-biased overviews of the MGE pool and their hosts in communities. To this end, cultivation-independent techniques are needed that allow the direct linkage of native MGEs to their host genomes and determine their transfer between native community members (transfer direction IV). Ideally, these techniques should enable the simultaneous tracking of many different types of mobile elements, including chromosomally integrated elements such as prophages and ICEs, as well as non-integrated elements such as plasmids. In the following section, we will address different metagenomics approaches that allow the association of various types of MGEs with their host genomes. We will first discuss how different metagenomics approaches have been used to interrogate the presence of MGEs in microbiomes. We will then specifically highlight how high-throughput 3C-based techniques as well as DNA methylome analyses can be used to assign extrachromosomal elements to their host genomes in microbiomes of varying complexity.

Identifying genes of interest and MGEs in metagenomes

Whole genome sequencing can allow for sequence analysis of MGEs and genes of interest in isolated species. However, this requires the ability to cultivate the particular species in isolation and under lab conditions. Metagenomic sequencing does not rely on cultivation and can provide insights into the genomic content of microbial communities in an untargeted

manner. It thus allows for a more exhaustive characterization of genes of interest or MGEs present in a given microbiome. For example, Zhang et al. used shotgun metagenomic sequencing to analyze the occurrence of ARGs in communities from different environments: glacial soil, permafrost and sediment (Zhang et al. 2018a). Further combining shotgun metagenomic sequencing with functional metagenomic selections against predetermined panels of antibiotics allows not only the detection of known ARGs, but also the confirmation of their phenotype and the discovery of new ARGs. In brief, functional metagenomics selections involve the subcloning of DNA fragments from metagenome extractions into a lab-adapted plasmid and host strain, which can then be screened for a phenotype of interest, such as antibiotic resistance (Handelsman et al. 1998; Rondon et al. 2000; Moore et al. 2011). This approach was used, for example, to interrogate the resistomes of preterm infant gut microbiomes (Gibson et al. 2016). In this study, a total of 794 ARGs functional against 16 different antibiotics were detected. 79% of these genes were identified as ARGs for the first time. Similarly, functional selection of soil metagenomic libraries lead to the discovery of 2895 ARGs, most of which were previously unknown (Forsberg et al. 2014). Interestingly, this study also found that the genomic co-localization of ARGs and MGEs is lower in soil genomes as compared to pathogenic genomes, suggesting that HGT of ARGs might play different roles in different environments. For example, shotgun metagenomic sequencing of a microbiome isolated from a lake affected by industrial antibiotic pollution revealed the presence of a wide variety of ARGs and genes associated with MGEs (Bengtsson-Palme et al. 2014). What is more, in several instances their co-localization suggested the potential for HGT of ARGs in this environment. Despite these useful metagenomics approaches, the

content and genomic context of MGEs often remain elusive and the elements themselves can be difficult to assemble with short-read sequencing data, given the repetitive sequence structure often associated with these elements. This is starting to be addressed with microfluidics-based techniques (Lan et al. 2017; Bishara et al. 2018) and long-read sequencing technologies, such as PacBio SMRT sequencing (Eid et al. 2009) and Oxford Nanopore sequencing (Branton et al. 2008).

A recently developed microfluidics-based technique, single cell genomic sequencing (SiC-seq), distributes intact cells into individual droplets and tags each genome with unique barcodes (Lan et al. 2017). SiC-seq has been used to associate bacterial genera with ARGs as well as virulence factors and phage sequences within a coastal seawater community (Lan et al. 2017). While this technique can theoretically associate episomal sequences with their host genomes, the lack of genome amplification prior to sequencing yields low genome coverage (0.1%–1%) per barcode (i.e. per microbial cell), limiting the assembly of whole genomes from metagenomic data. Generally, for the study of extrachromosomal mobile elements, deep metagenomic sequencing and assembly, followed by single cell sequencing approaches may allow for association of physically unlinked sequences (Stepanauskas 2015). For example, combining metagenomic and single-cell sequencing has been used to explore the biology and genomes of uncultured cells from novel phyla (Dodsworth et al. 2013).

Just like microfluidics-based techniques, long-read technologies aim to preserve long-range sequence information. These technologies generate sequencing reads of many kilobases, which inherently preserve genomic context and improve assemblies of MGEs. For example, a study of metagenomic samples from a veterinary hospital comparing PCR-based

detection and Oxford Nanopore sequencing of ARGs found that the PCR-based approach could detect very low-abundance ARGs, while the long-read approach was able to assign chromosomal ARGs to bacterial genera (Kamathewatta et al. 2019). Another publication associated ARGs and their host mobile elements using long-read sequencing in wastewater treatment plant samples (Che et al. 2019). Open-access web tools are being developed to analyze Nanopore sequencing data of environmental samples and return information on ARGs, metal resistance genes, MGEs and associated taxa (Arango-Argoty et al. 2019). In another study, combining Illumina short-read with both PacBio SMRT and Oxford Nanopore long-read sequencing technologies allowed researchers to assemble full genomes from low-complexity natural whey starter cultures, distinguish between microbes on a strain level and assign prophages to their host genome (Somerville et al. 2019).

The described metagenomic sequencing generates reads that can then be mapped against the databases in Table 1.1 to screen for known resistance and mobile genes. The current standard in medical and research settings is unbiased, metagenomic Illumina short-read sequencing, but this remains time-consuming due to quality-controlled library preparation and bioinformatics analysis (Gu, Miller and Chiu 2019). Real-time sequencing and basecalling through Oxford Nanopore Technologies creates the opportunity to quickly screen communities, such as those associated with an infection, against a database of ARGs or MGEs. The Nanopore system involves parallel processing of DNA molecules through biological pores that measure displacement of the molecules (Branton et al. 2008). As sequence data for each piece of DNA becomes available in real time, each molecule can be screened for similarities against a known database with software by Metrichor called ARMA

(Real-time detection of antibiotic-resistance genes using Oxford Nanopore Technologies' MinION 2016; Schmidt et al. 2017) or a licensed University of California San Francisco pipeline called MetaPORE (Greninger et al. 2015) in just a few minutes after library preparation. Long-read technologies may further be able to provide genomic contexts to these resistance genes to allow for host association if these genes are chromosomally integrated and real-time analysis of reads may allow background human genomic reads to be removed to increase sequencing depth of the microbial community (Loose, Malla and Stout 2016). Web-based tools, such as nanoARG (Arango-Argoty et al. 2019) for Nanopore reads, and PATRIC (Wattam et al. 2017) or MG-RAST (Keegan, Glass and Meyer 2016) for Illumina short reads, as well as alignment against databases in Table 1.1, can be used for analysis of raw sequences. Together, these metagenomic sequencing techniques continue to expand databases and generate data sets that can be used to develop tools for computational analysis and cataloguing of MGEs within metagenomes (Table 1.2).

Table 1.2 Computational tools that identify mobile elements and genes of interest in metagenomes. Tools that allow for metagenomic input are marked with an asterisk (*), those that accept reads (metagenomic or isolate) are marked with a plus sign (+), and those that require annotated contigs are marked with a cross (†). Tools without a symbol require either isolate genomes or binned metagenome-assembled genomes. (rr) - request required for software. Hidden Markov Models (HMMs) are reviewed in (Eddy 2004).

Software	Output	Reference
Plasmid Identification		
Plasflow *	Phyla and plasmid annotation	(Krawczyk, Lipinski and Dziembowski 2018)
PPR-meta *	Phage and plasmid annotation	(Fang *et al.* 2019)
PLACENET	Assembly and contig linkages	(Lanza *et al.* 2014)
PlasmIdent *+	Circular plasmids and predicted genes, ARGs	github.com/imgag/plasmIDent
PlasmidFinder, pMLST *+	Assembly if input reads, hits to database	(Carattoli *et al.* 2014)
Cbar *	Plasmid annotation	(Zhou and Xu 2010)
PlasmidSeeker	Hits to database	(Roosaare *et al.* 2018)
Plasmid ATLAS +	Hits to database and related plasmids	(Jesus *et al.* 2019)
plasmidSPAdes *+	Plasmid contig assembler, includes PlasmidVerify	(Antipov *et al.* 2016)
Phage and Prophage Identification		
VirSorter *	Phage and prophage annotation	(Roux *et al.* 2015)
virMine *+	Viral assembly and annotation	(Garretto *et al.* 2019)
PHASTER *	Taxonomic and functional annotation	(Arndt *et al.* 2017)
PATRIC/RAST *+	Gene annotation, taxonomy. Assembly if input reads.	(McNair *et al.* 2018)
vConTACT 2, iVirus *	Gene, phage and archaeal virus annotation, taxonomy	(Merchant *et al.* 2016; Bolduc *et al.* 2017)
PPR-meta *	Plasmid and phage annotation	(Fang *et al.* 2019)
Viraminer *+	Viral annotation	(Tampuu *et al.* 2019)

Table1.2, continued Computational tools that identify MGEs and genes in metagenomes

Software	Output	Reference
Phage and Prophage Identification		
FastViromeExplorer +	Metagenomic viral abundance, hits to database	(Tithi *et al.* 2018)
MARVEL	Phage annotation	(Amgarten *et al.* 2018)
VirFinder, DeepVirFinder *	Viral annotation	(Ren *et al.* 2017), github.com/jessieren/DeepVirFinder
Phage_Finder †	Putative prophage regions	(Fouts 2006)
Phigaro *	Putative prophage regions and insertion sites	(Starikova *et al.* 2020)
PhiSpy *†	Putative prophage regions	(Akhter *et al.* 2012), github.com/linsalrob/PhiSpy
VIBRANT *	Viral metabolic pathway annotation	(Kieft, Zhou, and Anatharaman 2020)
metaviralSPAdes *+	Viral contig assembler. Includes viralVerify, viralComplete	(Antipov *et al.* 2020)
Genomic Island Identification		
PredictBias †	Putative genomic and pathogenicity islands	(Pundhir, Vijayvargiya and Kumar 2008)
Alien_Hunter, GIHunter, GIV, GIDectector (rr)	Putative GIs and abnormal windows	(Vernikos and Parkhill 2006; Che *et al.* 2010, 2014; Che and Wang 2013)
GI-SVM, Hgtident, MgFC, RVM, SVM-AGP	Putative GIs, HGTs, abnormal windows	(Tsirigos and Rigoutsos 2005; Vernikos and Parkhill 2008; Xiong *et al.* 2012; Lu and Leong 2016)
MJSD, GEMINI *	Segmentation to putative GIs	(Zhang *et al.* 2014; Arvey *et al.* 2009)
Centroid (rr)	Clusters, putative GIs	(Rajan, Aravamuthan and Mande 2007)
Design-Island	Putative GIs, functional annotation	(Chatterjee, Chaudhuri and Chaudhuri 2008)

Table1.2, continued Computational tools that identify MGEs and genes in metagenomes

Software	Output	Reference
Genomic Island Identification		
INDeGenIUS (rr)	Clusters, putative GIs	(Shrivastava, Reddy and Mande 2010)
IslandCAFE *†	Clusters, putative GIs, putative HGTs	(Jani and Azad 2019)
IGIPT	Putative GIs	(Jain, Ramineni and Parekh 2011)
Sighunt	Putative HGTs	(Jaron, Moravec and Martínková 2014)
GI-Cluster †	Putative GIs and functions	(Lu and Leong 2018)
GEMINI	Putative GIs, atypical windows	(Paila *et al.* 2013)
Shutter Island	Putative GIs	(Assaf, Xia and Stevens 2019)
SWGIS	Putative GIs, also eukaryotic	(Clasen *et al.* 2018)
GC-Profile * (rr)	Contigs segmented based on GC content	(Gao and Zhang 2006)
Cumulative GC-profile † (rr)	Contigs segmented based on GC content	(Zhang and Zhang 2004)
Zisland Explorer	Putative GIs and GC profile	(Wei *et al.* 2017)
LEMON +	Strain level assembly and putative HGT from reference	(Li, Jiang and Li 2019)
CONJscan/ MacSyFinder *	Relaxasosome, T4CP/VirB4, and T4SS annotation	(Cury *et al.* 2020)
DaisySuite (MicrobeGPS, DaisyGPS, Daisy) +	Putative donor and acceptor within reference	(Seiler, Trappe and Renard 2019)
Hgtident * (rr)	Putative HGT	(Xiong *et al.* 2012)
JS-CB	Putative HGT	(Jani *et al.* 2017)
MGTIpick	Putative GIs	(Dai *et al.* 2018)
SIGI-CRF, Colombo	Putative GIs	(Waack *et al.* 2006)
CGS † (rr)	Putative GIs	(Elhai, Liu and Taton 2012)
Kvasir	Comparative phylogenetics, unusual sequence identity	(Bonham, Wolfe and Dutton 2017)

Table1.2, continued Computational tools that identify MGEs and genes in metagenomes

Software	Output	Reference
Genomic Island Identification		
RecentHGT	Comparative phylogenetics, unusual sequence identity	(Li *et al.* 2018b)
MOSAIC	Comparative phylogenetics, unusual sequence identity	(Chiapello *et al.* 2005, 2008)
Near HGT	Comparative phylogenetics, putative HGT	(Adato *et al.* 2015)
Waafle +	Comparative phylogenetics, putative HGT	huttenhower.sph.harvard.edu/waafle
DLIGHT	Putative HGT orthologs, potential donor and recipient, and estimated time since HGT	(Christophe Dessimoz, Daniel Margadant and Gaston H. Gonnet 2008)
DarkHorse †	Phylogenetic discordance in hits to database, putative donors	(Podell and Gaasterland 2007)
RGPFinder	Comparative phylogenetics, genome segmentation	(Vallenet *et al.* 2009; Ogier *et al.* 2010)
MSGIP	Comparative phylogenetics, putative GIs, atypical windows	(de Brito *et al.* 2016)
Identification of Other Mobile Elements and Associated Sequences		
tRNAscan-SE *	tRNA genes annotation	(Lowe and Eddy 1997)
tRNAcc	tRNA-associated GIs	(Ou *et al.* 2006)
OriTFinder *	Virulence factor and ARG annotation, hits to database	(Li *et al.* 2018c)
ISsaga *	Insertion sequence annotation	(Varani *et al.* 2011)
ISMapper +	Insertion sequence annotation	(Hawkey *et al.* 2015)
ISEScan *	Insertion sequence annotation, by category	(Xie and Tang 2017)
INTEGRALL *	Hits to integron database	(Moura *et al.* 2009)
IntegronFinder *†	Hits to *attC* and *intI*	(Cury *et al.* 2016)
I-VIP *	Integron annotation	(Zhang *et al.* 2018)

Table1.2, continued Computational tools that identify MGEs and genes in metagenomes

Software	Output	Reference
Identification of Other Mobile Elements and Associated Sequences		
REPuter *	Exact repeats	(Kurtz and Schleiermacher 1999)
TEdenovo, TEannot, and PASTEC *	Repeated sequences, TE copies by cluster and transposon classification (Wicker *et al.* 2007)	(Flutre *et al.* 2011; Hoede *et al.* 2014)
deviaTE *+	TE identification and copy number quantification	(Weilguny and Kofler 2019)
Generic Repeat Finder *	Interspersed exact repetitive regions	(Shi and Liang 2019)
ProGeRF *	Imperfect repetitive regions	(Lopes *et al.* 2015)
SPADE *	Repetitive regions, periodicity, and putative function	(Mori *et al.* 2019)
RED *	HMM model, repetitive regions	(Girgis 2015)
iMGEins +	Putative transposon and retrotransposon insertions	(Bae *et al.* 2018)
Antibiotic Resistance Identification		
Resistance Gene Identifier *	ORF prediction and hits to database	(Alcock *et al.* 2019)
PATRIC	Gene and ARGs annotation	(Davis *et al.* 2016)
AMRFinderPlus *	Hits to database	(Feldgarden *et al.* 2019)
MetaCherchant *+	Hits to database, assembly graph	(Olekhnovich *et al.* 2018)
kmerResistance *+	Hits to database and abundance	(Clausen *et al.* 2016)
ResFinder, PointFinder *+	Hits to database and chromosomal mutations	(Zankari *et al.* 2012, 2017)
ARG-ANNOT *	Hits to database	(Gupta *et al.* 2014)
SRST2 *+	Hits to database	(Inouye *et al.* 2014)
NanoARG *+	Taxonomy and hits to database	(Arango-Argoty *et al.* 2019)
ARGpore *+	Hits to database	(Xia *et al.* 2017)

Table1.2, continued Computational tools that identify MGEs and genes in metagenomes

Software	Output	Reference
ABRicate *	Hits to database	github.com/tseemann/abricate
fARGene *+	Assembly, fragmented and complete ORFs, previously uncharacterized resistance genes	(Berglund *et al.* 2019)
Antibiotic Resistance Identification		
ABRESFinder *	Hits to database	http://scbt.sastra.edu/ABRES/hmmer.php
Ariba +	Assembly, hits to database	(Hunt *et al.* 2017)
Mykrobe +	Hits to database, assembly graph	(Bradley *et al.* 2015)
TypeWriter *+	Hits to database	(Mason *et al.* 2018)
GeneFinder *+	Hits to database	(Mason *et al.* 2018)
SSTAR *	Hits to database	(de Man and Limbago 2016)
MetaCompare +	Assembly, ARG, MGE abundance and hosts. Resistome hazard level	(Oh *et al.* 2018)
ARGs-OAP *	ARG annotation	(Yin *et al.* 2018)
VRprofile	Putative GI, virulence factors, ARGs, secretion system annotation	(Li *et al.* 2018a)
Virulence Factor Identification		
PAI-IDA	Pathogenicity Islands	(Tu and Ding 2003)
VFAnalyzer	Hits to database	(Liu *et al.* 2019a)
virulenceFinder +	Assembly if input reads, hits to database	(Joensen *et al.* 2014)
VirulentPred *†	Virulence Factor annotation	(Garg and Gupta 2008)
PAIfinder *	GIs, ARGs, Virulence Factor annotation	(Yoon *et al.* 2005)

Generally, metagenomic sequencing and assembly generates contiguous sequences, or contigs, that can then be analyzed by various computational techniques and tools to identify known and putative MGEs and suggest a host, as was recently reviewed in (Bertelli, Tilley and Brinkman 2018; da Silva Filho et al. 2018; Douglas and Langille 2019). Some groups have generated pipelines that combine multiple tools to annotate multiple classes of MGEs within metagenomes (Hasan et al. 2012; Lee et al. 2013; Soares et al. 2016; Bertelli et al. 2017; Gargis et al. 2019; Guliaev and Semyenova 2019; Song et al. 2019); lab-maintained servers and programs such as snakemake (Köster and Rahmann 2012) and docker (docker.com) increase the reproducibility of these pipelines. Generally, these computational tools differ in the type of input requirements, such as completeness of genomes to be analyzed, their tolerance of fragmented genes, whether they require contigs binned into genomes, as well as the generated output, e.g. the types of MGEs identified. Furthermore, these tools identify mobile elements either using comparative phylogenetic approaches or using parametric approaches considering factors such as GC content, gene functions or integration sites. For example, various computational methods have been developed to assemble and detect plasmid sequences from metagenomic data, such as those using sequencing coverage, codon-bias or the presence of plasmid-associated genes (Rozov et al. 2017; Arredondo-Alonso et al. 2018; Krawczyk, Lipinski and Dziembowski 2018; Antipov et al. 2019; Wood, Lu and Langmead 2019). However, while these tools are useful for assembling plasmids from metagenomic data, they often fall short in identifying their hosts. For example, recently transferred plasmids may not have been ameliorated to the host chromosome's codon bias and thus might show a different GC profile. In addition, plasmids

are often present in cells at a different copy number than chromosomes making the use of coverage information less useful for host assignment. Hence, despite the improvements of metagenomic assemblies, the association of extrachromosomal MGEs with their host genomes remains a challenge. The next two parts of this review will discuss 3C- and methylome-based analyses that are promising emerging strategies for overcoming these limitations.

Assigning MGEs to their hosts based on their interaction with host chromosomes

The 3D organization of chromosomes within cells has long been an interesting open question in the field of genetics. 3C-based techniques, also known as proximity ligation, have been used to address this question. These techniques rely on the cross-linking of DNA-binding proteins to tether pieces of physically proximal DNA within intact cells. After cross-linking and cell lysis, DNA is digested with restriction enzymes. The fragmented DNA-protein complexes are diluted prior to the subsequent ligation. This encourages ligation of DNA ends part of the same protein-DNA complex, rather than the ligation of DNA ends of separate complexes. The resulting chimeric ligation junctions combine DNA that was co-localized within the cell. Cross-links are then reversed and DNA is purified (Figure 1.4A). Ligation junctions can be analyzed in various ways, ranging from investigating the interactions between two specific loci by PCR to the investigation of all versus all genomic loci by next generation sequencing (Dekker, Marti-Renom and Mirny 2013).

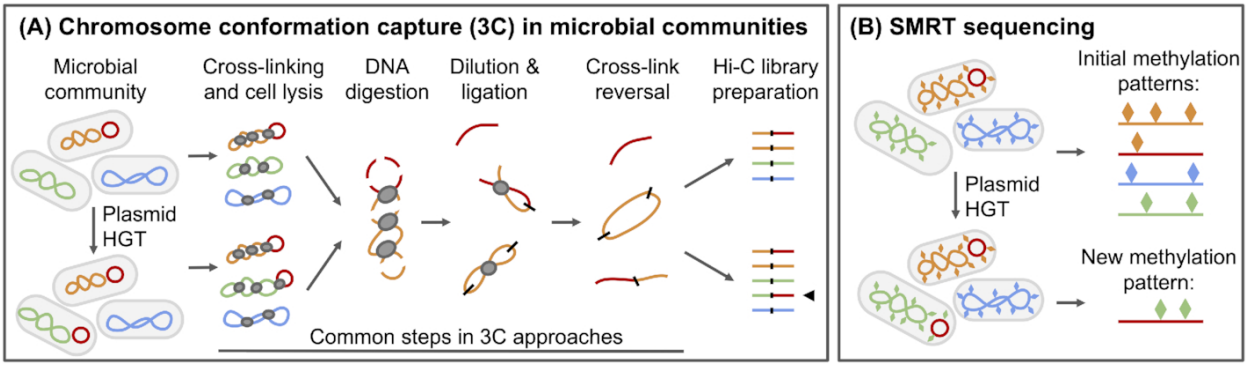


Figure 1.4 Novel metagenomics techniques that have the potential to assign extrachromosomal elements to their host genomes (A), Chromosome conformation capture (3C)-based metagenomics in microbial communities. Common steps of 3C-based techniques include the crosslinking of proteins (grey dots) associated with DNA molecules in intact cells followed by cell lysis. The DNA is digested with a restriction enzyme (or a combination of restriction enzymes) and diluted to encourage the formation of intramolecular ligation junctions (black bars) in the subsequent ligation reaction. Last, crosslinks are reversed. 3C techniques have the potential to detect contacts between chromosomal DNA (orange) and extrachromosomal DNA, such as plasmids (red circles). Hi-C, a high-throughput 3C-based technique, has the potential to detect HGT of MGEs, such as plasmids, in metagenomes. During Hi-C library preparation, the ligation junctions (black bars) are marked with biotin allowing for the specific enrichment of ligation junction-containing fragments, which can subsequently be sequenced using short-read sequencing. The black arrowhead indicates a ligation junction detected due to HGT of the plasmid from the original host (orange chromosome) to the new host (green chromosome). (B), Methylation profile-based assignment of MGEs to their hosts in microbial communities. Both chromosomally integrated and extrachromosomal elements carry their host cell's methylation profile (diamonds). Long-read sequencing that also detects these methylation patterns allows the assignment of MGEs to their host genomes. If DNA is shared with an organism that exhibits a measurably different methylation profile, it will acquire its new host's profile. New methylation patterns associated with MGEs are thus indicative of HGT in the community.

The earliest 3C technique detects interactions between two specific loci (Dekker et al. 2002). This technique requires the design of primer pairs that allow the detection of hybrid ligation junctions that would be created if two genomic loci of interest interact, such as a gene of interest with a putative regulatory region. This technique was then further developed into 4C (Simonis et al. 2006; Zhao et al. 2006), which detects interactions between one specific

loci and all other loci from the genome. This is achieved by circularizing the hybrid ligation product and using inverse PCR to amplify all genetic loci captured by the locus of interest. The resulting PCR fragments can then be identified by microarray (Simonis et al. 2006) or sequencing (Zhao et al. 2006). This technique could be useful, if one is interested in potential HGT events of a specific MGE that has a known sequence and thus allows the design of appropriate primers. While 4C investigates the interaction of one locus with all other genomic loci, another variation of 3C, called 5C, detects interactions between many specific genomic loci (Dostie et al. 2006). Finally, Hi-C tests the interactions between all possible pairs of loci in the genome and is the only unbiased 3C-based technique (Lieberman-Aiden et al. 2009). This is achieved by repairing restriction fragments with biotinylated nucleotides before ligation thus creating ligation junctions that can be affinity-purified. This ligation-junction enriched library can then be subjected to high-throughput, paired-end sequencing with each sequencing read from a pair representing one of the two interacting loci.

Together, 3C-based methods have helped develop 3D models of genomes of a variety of organisms and general principles are starting to be identified (Dekker, Marti-Renom and Mirny 2013). For example, the 3D structure of the Caulobacter crescentus genome was found to be organized at least partly by parS sequences (Umbarger et al. 2011). These centromere-like sequences have been shown to be essential for chromosome segregation in C. crescentus (Toro et al. 2008). Since 3C-based techniques exploit physical contacts between DNA in the confines of cells, it was posited that the resulting contact maps could be exploited to bin contigs, including those of extrachromosomal MGEs, for de novo assembly of genomes in the context of metagenomes. Table 1.3 gives an overview of studies that have used Hi-C-based

approaches to associate MGEs with their hosts in microbial communities of varying diversity. They all follow the general outline shown in Figure 1.4A. One of the first published studies, a synthetic community of five bacteria was parsed at strain level and DNA molecules residing in the same host cells, such as plasmids or multiple chromosomes, were associated with each other (Beitel et al. 2014). Similarly, the Hi-C-based MetaPhase workflow was shown to effectively parse genomes belonging to a synthetic community of 16 yeast strains and another community consisting of eight yeasts, nine bacteria and one archeon at species level (Burton et al. 2014). That same year, a workflow termed meta3C was applied to other microbial communities, including an environmental community of unknown composition that was collected from river sediments and enriched for bacterial microorganisms capable of growth in lysogeny broth (Marbouty et al. 2014). Contact data from this experiment revealed 184 significant clusters. The 11 largest clusters were further investigated and over 80% of genes identified by MG-RAST within in each cluster were attributed to the same taxonomic class. Further, the contact data enabled the authors to study the 3D genome organization of an unknown species within the community. More recently, 850 draft genomes were assembled from cow rumen metagenomes using traditional methods, but with the help of Hi-C contact information, an additional 63 draft genomes could be assembled (Stewart et al. 2018). In other studies employing Hi-C, plasmids and other MGEs were repeatedly associated with their presumed host genomes in microbiomes of varying levels of complexity (Marbouty et al. 2017; Press et al. 2017; Bickhart et al. 2019; Yaffe and Relman 2019). For example, it was shown that Hi-C can be used to assess host ranges for natural plasmids in a wastewater community without the need for cultivation of the hosts (Stalder et al. 2019). The authors of

this study report a detection limit of 0.01%, meaning that at least 1 in 10 000 cells in the sample must be of a particular host-plasmid association. Finally, it will be interesting to test to which degree Hi-C is able to detect recent or transient plasmid transfer. This likely depends on the level of noise in the Hi-C data and the frequency of plasmid transfer, i.e. how many recipients are to be expected within the community at any given time. Being able to detect transient transfers might also help illuminate transfer routes since the detection of a particular MGE in a new host might not be due to transfer to this organism from the initial host but might be due to secondary transfer via a third organism.

Table 1.3 Studies employing Hi-C techniques to assign MGEs to their hosts in microbial communities. All Hi-C libraries in these studies were sequenced by Illumina short-read approaches.

Reference	Metagenome sequencing	Community	Taxonomies reported	Interactions reported
(Beitel *et al.* 2014)	Simulation of Illumina short reads using reference genomes	synthetic (5 bacteria)	bacteria	plasmid-chromosome
(Burton *et al.* 2014)	Illumina short reads and mate-pair reads or simulation of 10 kb contigs from reference genomes	synthetic (16 yeast); synthetic (8 yeasts, 9 bacteria, 1 archaeon)	bacteria, archaea, yeast	plasmid-chromosome
(Marbouty *et al.* 2014)	from meta3C library	synthetic (3 bacteria), synthetic (11 yeast), environmental (river sediment)	bacteria, yeast	plasmid-chromosome
(Marbouty *et al.* 2017)	from meta3C library	host-associated (mouse feces)	mostly bacteria, few archaea or eukaryotes	Phage-chromosome
(Press MO *et al.* 2017)	Illumina short reads	host-associated (human feces)	bacteria	plasmid-chromosome
(Bickhart *et al.* 2019)	PacBio long-reads	host-associated (cow rumen)	mostly bacteria, few archaea or eukaryotes	phage/ARG-chromosome
(Stalder *et al.* 2019)	Illumina short reads	environmental (wastewater)	bacteria	ARG/plasmid/ integron-chromosome
(Yaffe and Relman 2019)	Illumina short reads	host-associated (human feces)	bacteria	various MGE-host association

Since DNA interactions can be studied directly through Hi-C, they do not rely on cultivation of host organisms or on the expression of reporter genes. It thus has the potential to detect the transfer of artificially introduced MGEs (transfer direction I and II) as well as the transfer of native mobile elements to introduced recipients (transfer direction III) or between native community members (transfer direction IV) regardless of the microbes' culturability. While these approaches are very promising toward gaining new insights into HGT dynamics within microbial communities, they are not without drawbacks. For example, strain-level genome deconvolution remains challenging (Burton et al. 2014). In addition, Hi-C data poses a computational challenge and requires expertise in its analysis. In regards to the latter barrier, open source computational pipelines, such as MetaTOR (Metagenomic Tridimensional Organization-based Reassembly), have been developed to aid in the analysis of data from metagenomic samples (Baudry et al. 2019). In this study, the authors collected a total of 20 mouse gut samples, constructed meta3C libraries, and compared assembly bins generated by MetaTOR to other binning softwares, MetaBAT (Kang et al. 2015) and CONCOCT (Alneberg et al. 2014). They report that MetaTOR compares favorably to these computational pipelines. When combining all 20 samples, MG-RAST (Keegan, Glass and Meyer 2016), a short-read metagenomics analysis pipeline, predicted the presence of 268 bacterial genomes; of these MetaTOR was able to provide 82 high-quality genomes, while MetaBAT and CONCOCT generated 22 and 12, respectively. A drawback to the MetaTOR pipeline is that it excludes MGEs that share physical contacts with more than one host genome, i.e. those that might have been transferred. However, the authors posit that this drawback can be overcome by incorporating additional analysis tools such as VirSorter (Roux et al. 2015) and PlasFlow

(Krawczyk, Lipinski and Dziembowski 2018). Other computational pipelines developed for the analysis of Hi-C data generated from metagenomes include the proprietary ProxiMeta (Press et al. 2017) and the open source bin3C (DeMaere and Darling 2019). Both studies tested their computational pipeline on the same data set generated from the microbiome isolated from a human fecal sample by (Press et al. 2017). bin3C showed an improved retrieval of high-quality, nearly complete genomes as compared to ProxiMeta for this data set. The best analysis pipeline to use for a given experiment probably depends on the sample type and complexity as well as the noisiness of the data.

Assigning MGEs to their hosts based on methylomes

3C-based techniques are not the only ones that have been used to address the largely unexplored transfer direction IV. An alternative approach is based on analyzing methylation profiles generated by sequencing technologies such as Illumina bisulfite sequencing (5mC/CpG, 5-hmC, 5-caC, 5-fC (Darst et al. 2010; Booth et al. 2012, 2014; Yu et al. 2012; Dalia, Lazinski and Camilli 2013; Lu et al. 2013; Song et al. 2013; Chen et al. 2017)), PacBio SMRT bisulfite sequencing (5mC/CpG, 5-hmC, m6A, m4C (Flusberg et al. 2010; Clark et al. 2012; Fang et al. 2012; Zhu et al. 2018; Chen et al. 2019; Hiraoka et al. 2019)), and Nanopore sequencing (5mC, 5hmC, 4mC, 5 mA; (Rand et al. 2017; Simpson et al. 2017; Stoiber et al. 2017; McIntyre et al. 2019; Ni et al. 2019; Liu et al. 2019b)) (Figure 1.4B). In brief, prokaryotes often encode restriction-modification (RM) defense systems, which employ restriction endonucleases that target specific DNA sequence motifs. To prevent self-targeting by the restriction endonuclease, these systems encode methyltransferases to mask the targeted

motifs within the host genome, creating characteristic methylation signatures. Known RM systems in prokaryotic genomes have been catalogued in REBASE (Roberts et al. 2010). In fact, a recent survey of 230 prokaryotic genomes found that 93% of genomes showed signs of methylation with 834 distinct methylation motifs (Blow et al. 2016). Since these systems are often associated with MGEs (Kobayashi et al. 1999), even closely related strains can display unique combinations of methylation motifs, also known as methylation profiles. Since all DNA molecules within a given host are exposed to the same RM systems, they are expected to receive the same methylation marks by their host cells, which has indeed been observed experimentally (Beaulaurier et al. 2018). In this study, the authors were not only able to use methylation profiles to identify 9 distinct contig bins in the fecal microbiome of low-to-medium complexity from an adult mouse, but they were also able to assign eight of the 19 detected MGEs to their respective hosts. This technique allowed for the association of both small and large MGEs to their hosts. What is more, five of the assigned MGEs were circular plasmids, showing the ability of this approach to assign extrachromosomal elements to their host genomes. Since this approach relies on the ability to detect unique methylation profiles for each member of the community, it follows that the ability to unambiguously assign MGEs to their hosts decreases with increasing community complexity. Indeed, in a simulation using methylomes available on REBASE, communities ranging from 20 to 200 members were investigated for their methylome diversity, and a sharp decrease of methylome uniqueness was observed beyond 20 community members (Beaulaurier et al. 2018). Since this study only took into account 6 mA methylation motifs detected by SMRT sequencing, the authors posit that future pipelines incorporating other methylation motifs, such as 5 and 4 mC, might allow

for the study of more complex microbiomes. In addition, combining this method with other computational binning methods such as CONCOCT or with Hi-C approaches might improve MGE detection and assignment to hosts.

Methylome analysis of SMRT sequencing data was also used to assign plasmids to their hosts in low-complexity natural whey starter cultures (Somerville et al. 2019). One of these cultures consisted of three species, one S. thermophilus, one L. delbrueckii and two L. helveticus strains, and contained three plasmids. One of the plasmids could not be assigned to its host. Low plasmid coverage as well as small plasmid sizes can both cause the failure to assign plasmids to host genomes. The other two plasmids showed methylation profiles consistent with a L. helveticus host. Since the methylation profiles of the two L. helveticus strains were found to be quite similar, the authors also considered coverage data to suggest specific hosts for these plasmids. This example demonstrates that combining different metagenomic analysis approaches usually generates more comprehensive results as compared to each individual tool by itself.

Together, the established and novel metagenomics techniques described in the above section constitute exciting tools for studying HGT in microbial communities. However, having the right tools is not sufficient without suitable model systems. In the last section of this review, we will discuss how the microbial communities in fermented foods might make ideal models for studying HGT dynamics within microbiomes.

SECTION 4: STUDYING HGT IN FERMENTED FOODS

Given the complexity of microbiomes, studying the movement of MGEs in simpler, experimentally tractable microbial communities could lead to the discovery of general principles of HGT that can be extrapolated to more complex communities. In addition, these model communities should allow for temporal sampling from their native environment. The native environmental conditions in turn should be easily replicable under laboratory conditions for further mechanistic studies. The microbial communities associated with fermented foods, which are made through the metabolic processes of microbial communities (Sieuwerts et al. 2008), could provide useful model systems for studying HGT in microbiomes. Cheese, yogurt, beer, wine, and chocolate are among the best known fermented products in the Western world, but cultures across the globe have developed their own fermented products with a variety of starting materials and microbes (Chilton, Burton and Reid 2015) providing diverse microbial communities within a range of environments to study and compare. Indeed, the diversity of a variety of these communities has been well characterized (Marsh et al. 2013; Meersman et al. 2013; Bokulich et al. 2014; Wolfe et al. 2014; Wolfe and Dutton 2015; Tamang, Watanabe and Holzapfel 2016; Walsh et al. 2017; De Mandal et al. 2018; Groenenboom et al. 2019). Depending on the type of fermented food, the associated microbial communities live in surface-associated biofilms, liquid-suspended biofilms, semi-solid substrates or in suspended cultures (Wolfe and Dutton 2015). Sometimes multiple types of these microbial habitats occur within the same fermented food. One example is kefir, in which part of the community lives associated with kefir grains, while the other part lives in the surrounding milk, with the kefir grains showing a more stable microbial

composition than the milk (Blasche et al. 2019). The variety of microbial lifestyles is a useful aspect of fermented food communities, since HGT dynamics are predicted to vary in different environments and studying HGT in these different communities could thus allow comparison of gene transfer dynamics in different types of environments and microbial community lifestyles. For example, HGT is proposed to be more efficient in biofilms because of more robust physical interactions between donors and recipients; furthermore, MGEs themselves often encode factors promoting biofilm formation (Madsen et al. 2012, 2018).

Another important aspect of finding suitable model communities for the study of HGT is the actual occurrence of HGT in these communities. From previous research, it is already known that bacteria and fungi associated with fermented foods often carry genes or genetic elements showing evidence of HGT (Rossi et al. 2014). This survey reported that niche adaptive genes, such as genes for the utilization of specific substrates, as well as ARGs and virulence factors have been hypothesized to be exchanged in food environments, especially in dairy products. Indeed, culture-dependent approaches that test isolated lactic acid bacteria against a panel of antibiotics revealed wide and varied resistance profiles (Nawaz et al. 2011; Guo et al. 2017). In addition to culture-dependent approaches, various computational and genomic approaches have been applied to study the potential for HGT in fermented food environments, especially the microbial communities associated with cheese (Cheeseman et al. 2014; Ropars et al. 2015; Bonham, Wolfe and Dutton 2017). Similar to what was reported by Rossi et al., bacteria isolated from cheese surfaces are often found to carry MGEs that harbor genes predicted to be adaptive in the cheese environments (Bonham, Wolfe and Dutton 2017; Pham et al. 2017; Anast et al. 2019; Levesque et al. 2019). For example, ActinoRUSTI

(Actinobacteria-associated iRon Uptake/Siderophore Transport Island) constitutes a conserved group of predicted siderophore uptake systems, which are encoded on various mobile elements in Actinobacteria including a plasmid and various ICEs. This uptake system is predicted to provide access to iron in the cheese environment and might thus provide a fitness advantage by providing access to a limited resource (Bonham, Wolfe and Dutton 2017). Understanding how RUSTI-encoding MGEs are shared within communities and what their consequences on hosts and surrounding communities are could help to identify metabolic approaches to modulating microbial communities.

While the described data hint at HGT happening between members of fermented food communities, it is unclear whether HGT happens within the relatively short time period of fermentation or whether it requires longer timescales. It is also unclear whether HGT happens only between community members that are regularly co-cultured during production processes or between microbes that co-occur in other environments. Since the individual community members of fermented foods can generally be cultured under standard conditions, whole genome sequencing allows for exact genomic characterization, including the presence of MGEs such as ICEs, plasmids and prophages. Given that in vitro systems containing defined communities based on fermented foods can be established, this opens up the possibility of temporal studies of these communities under laboratory-controlled conditions (Wolfe et al. 2014; Blasche et al. 2019). Applying the techniques discussed in this review, such as Hi-C and methylome analyses, to these defined communities could help elucidate how the identified elements are shared between members. Since abiotic conditions such as the pH or salinity of

the in vitro medium can be easily adjusted, it is possible to measure the effect of specific environmental parameters on HGT dynamics as well.

It is also important to consider the possibility that HGT happens between members of different microbiomes once they come into contact with one another. For example, a microbiome that is ingested might come in contact and exchange genetic material with the resident gut microbiome. Since fermented foods are meant for consumption often with the microbial community still intact and alive, the presence and mobility of MGEs has to be considered from a human health and food safety perspective. The degree to which genes can be transferred between ingested microbes and commensal microbiota is still being investigated and so far, conflicting results have emerged. For example, one study found no evidence of transfer of a tetracycline-resistance encoding plasmid from ingested Lactobacillus reuteri after transfer through the human gut microbiota (Egervärn et al. 2010). In contrast, a L. acidophilus strain was found to acquire vancomycin resistance from resident microbes in the mouse gut during transit (Mater et al. 2008). The true degree of HGT between ingested food and the resident gut microbiota is still unknown and should be further investigated in future experiments.

While this review has focused on MGE transfer between prokaryotes, HGT between eukaryotes and even interkingdom HGT between prokaryotes and eukaryotes should be considered further in future studies. For example, a genome analysis of a commercial yeast strain used for wine making was found to have horizontally acquired genes important for fermentation processes (Novo et al. 2009). Even more intriguing, some of these genes seem to have originated from another yeast species, which often contaminates wine fermentations and

suggests that invasion of a fermentation community may result in significant evolutionary effects. An intriguing example of the horizontal transfer of genetic material between prokaryotes and eukaryotes is an ancient transfer event of a siderophore synthesis operon from bacteria to yeast followed by operon amelioration (Kominek et al. 2019). Given the changes in gene structure required to switch from bacterial to eukaryotic gene expression, it is unclear whether and how such transfers might happen during relatively fast fermentation processes or whether these types of transfers are restricted to communities that establish themselves for long periods of time. Further, discovery of inter-kingdom HGT in microbiomes can be challenging due to the resistant cell walls of many eukaryotes, creating greater difficulty with DNA extractions and leading to biased DNA extractions, longer repetitive genomic regions that limit assembly, and misassemblies which produce false positive HGT hits (Pignatelli and Moya 2011; McLaren, Willis and Callahan 2019). In vitro cheese communities including yeast and filamentous fungi have been assembled in the context of other studies (Wolfe et al. 2014; Morin, Pierce and Dutton 2018; Bodinaku et al. 2019) and the increasing number of sequenced fungal genomes from these microbiomes might aid in overcoming the challenges in detecting HGT from and to these organisms. The presence of filamentous fungi in fermented food communities is also interesting from the perspective of microbial hotspots within microbiomes, to which fungi can contribute (Kuzyakov and Blagodatskaya 2015). A cheese rind-associated filamentous fungus, Mucor lanceolatus, was shown to contribute to the dispersal of various Proteobacteria, likely by allowing these bacteria to swim along the hyphae in the liquid layer surrounding it (Zhang et al. 2018b); hence it is conceivable that filamentous fungi in cheese rind communities

contribute to the creation of hotspots within which HGT is facilitated. Indeed, using a fluorescence-based approach similar to those described herein, the filamentous fungi Pythium ultimum was shown to facilitate plasmid conjugation between Pseudomonas putida donors and recipients, likely by promoting physical contacts between cells (Berthold et al. 2016). It remains to be seen whether this also applies to more complex microbial communities. Overall, the microbiomes of fermented foods have already been shown to make experimentally tractable models (Wolfe et al. 2014; Kastman et al. 2016, 2017; Morin, Pierce and Dutton 2018; Zhang et al. 2018b; Blasche et al. 2019; Bodinaku et al. 2019). Future research will show whether fermented food communities also make suitable models for studying HGT dynamics in microbial communities of varying diversity.

Concluding remarks

In this review, we have discussed techniques for associating MGEs, such as plasmids, with their host genomes in microbial communities. While emerging techniques such as Hi-C-based methods and methylome analyses open new experimental doors for studying HGT in communities, the most suited technique for a given study will depend on the specific microbial community and experimental questions. Given that different approaches have different advantages and disadvantages, it is likely that the most successful experimental designs will combine different approaches. Further combining them with other molecular approaches such as transcriptomics, proteomics and metabolomics could be a powerful way of investigating complex, multispecies HGT dynamics in the context of other forms of bacterial interactions, such as T6S-mediated competition. Applying these techniques to model

microbiomes such as those associated with fermented foods may provide powerful insights into microbial HGT dynamics and increase our mechanistic understanding of HGT in multispecies communities.

Acknowledgements

The authors would like to thank Brooke E. Anderson for useful feedback on the manuscript. Chapter 1, in full, is a reprint of the material as it appears in Federation of European Microbiological Societies Microbiology Reviews, 2020, Saak, Christina C.; Dinh, Cong B.; Dutton, Rachel J. The dissertation author was an investigator and author. of this paper. This work was supported by funds from the National Institutes of Health [DP2 AT010401], the Pew Charitable Trust and Packard Foundation to R.J.D.

Chapter 2: Novel Mobile Genetic Elements in Fermented Foods

Characterization of Mobile Elements in Fermented Food Microbiomes using Long Reads

Cong B. Dinh, Rachel J. Dutton

Abstract

Metagenomic short read sequencing of natural microbial communities has enabled extensive characterization of the taxonomic and functional composition of diverse environments. Although mobile elements are major drivers of microbial evolution, their recovery from short-read metagenomics can be challenging due to their sequence characteristics, such as the presence of repetitive regions. As a result, it can be difficult to characterize both the diversity of mobile elements within a given environment, and the complement of genes that are encoded on mobile elements. The application of long read sequencing, which can generate sequences of lengths over 100kb, to metagenomics can dramatically improve assembly of DNA sequences into long contiguous regions and complete genomes. Here, we use long read sequencing to characterize the microbial and predicted mobile genetic diversity within a set of fermented dairy products, cheese and kefir. Although fermented foods are typically less complex than other microbiomes, they contain viruses, fungi, and bacteria. Using nanopore long-read sequencing, we are able to recover a highly diverse set of complete phage genomes and plasmids, the majority of which have not previously been documented. Genes predicted to provide adaptive functions in the dairy environment are encoded in these elements. Our work demonstrates the utility of long-read sequencing for the detection and characterization of mobile elements within these microbiomes, and highlights the previously hidden plasmid and viral diversity in fermented foods.

Importance

Around the world, fermentation is used to preserve and enhance foods. Within the microbiomes of these fermented foods, microbial interactions and evolution can drive community function and stability, improving flavor and resisting invasion from possibly pathogenic microbes. However, the genetic capacity for key elements responsible for some of these interactions, such as phage and plasmids, has been challenging to document in a culture-independent way. In this work, we use recent advances in sequencing technology to examine the diversity of plasmids and phage in the cheese and kefir microbiomes and the functional potential encoded within these genetic elements.

Introduction

Diverse microbial communities drive dynamics and stability in natural environments, in various hosts, and in fermented food. The metabolic processes that underlie actions in the microbiomes are encoded in the genomes of fungi, bacteria, and viruses. The environments in which microbiomes exist are dynamic, and evolutionary processes can underlie survival in response to local biotic and abiotic changes. One of the major drivers of rapid adaptation in microbial species is horizontal gene transfer, in which entire genes and gene clusters can be incorporated into the genome of a new host, thus providing access to new metabolic pathways or gene products. Beyond changing composition of the microbiome, these gene transfer events can change the ecosystem's physiology, the resulting metabolites and products created by the community (Larkin & Martiny, 2017; Niccum et al., 2020). Two canonical methods of horizontal gene transfer are conjugation and transduction, largely conducted by plasmids and

viruses that carry these genes to new hosts. Examining the functions encoded on these elements can highlight the stressors and availability of various resources within the microbiome (Ochman et al., 2000).

Here, we apply long read sequencing to environments that have been controlled by humans for millennia: cheese rinds and kefir, both fermented dairy products. Many fermented food microbiomes are relatively simple, composed of fewer than 10 abundant genera of fungi and bacteria, and exhibit reproducible community assembly, making them useful as models for understanding the biological processes within microbiomes (Blasche et al., 2021; Kim & Chun, 2005; Landis et al., 2021; Marsh et al., 2013; May et al., 2019; Walsh et al., 2016; Wolfe et al., 2014). In light of the utility of fermented food microbiomes as models for community-based biological processes, we sought to expand our understanding of these systems through the application of long-read metagenomic sequencing, specifically by examining the plasmid and viral populations within these simple microbiomes. We identify a diverse set of plasmids and viruses associated with these fermented foods, many of which are not previously represented in public databases. We analyze the gene functions encoded on these plasmids and identify a number of genes that would be predicted to be beneficial in a dairy environment, or would be involved in microbial interactions with the fermented food microbiomes. Mobile elements within these various environments provide insight into possibly unknown stressors within their environment, and can reveal how long read sequencing can significantly contribute to our catalog of mobile elements from microbial communities.

Results

The five samples in this analysis included two kefirs and three cheese rinds: one each of natural rind, bloomy rind, and brine-washed rind. Long read sequencing generated reads that were up to 1.5 million basepairs (bp) long, with weighted medians (N50s) ranging from ~3 to 11kbp (Supplemental Figure 2.1). Long reads were assembled into contigs with metaFlye, corrected with five rounds of consensus-based methods (Kolmogorov et al., 2020; Vaser et al., 2017), and polished using a protein-annotation method to correct frameshifts resulting from nanopore-based sequencing errors, as described in (Arumugam et al., 2019). The resulting assemblies have N50s that range from 37kb to 1,161kb (Supplemental Table 2.1), correlating with genus level community complexity and long read N50s. Figure 2.1a shows a summary of the assemblies, with contigs clustered by tetramer frequency and coverage using anvi'o (Eren et al., 2015); their circularity from metaFlye assembly (Supplemental Figure 2.3); their genus taxonomic assignments by kraken2 (Wood et al., 2019); their annotation as plasmid or virus by viralVerify (Antipov et al., 2020) and VIBRANT (Kieft et al., 2019), respectively.

To estimate the relative abundances of taxa within these samples, we mapped raw long reads to the final assemblies (Figure 2.1b, Supplemental Figure 2.2). By using assembly taxonomy, rather than the relatively error-rich long longs, we improved the percentage of nucleotides remaining unclassified from 20% to 10% of compared to mapping the raw reads against RefSeq database alone using MetaMaps (Figure 2.1B, supplemental figure 2.2 (Dilthey et al., 2019)). This was not limited to reads shorter than 1kb, suggesting that both methods are not dependent on read length. We next sought to evaluate the quality of the

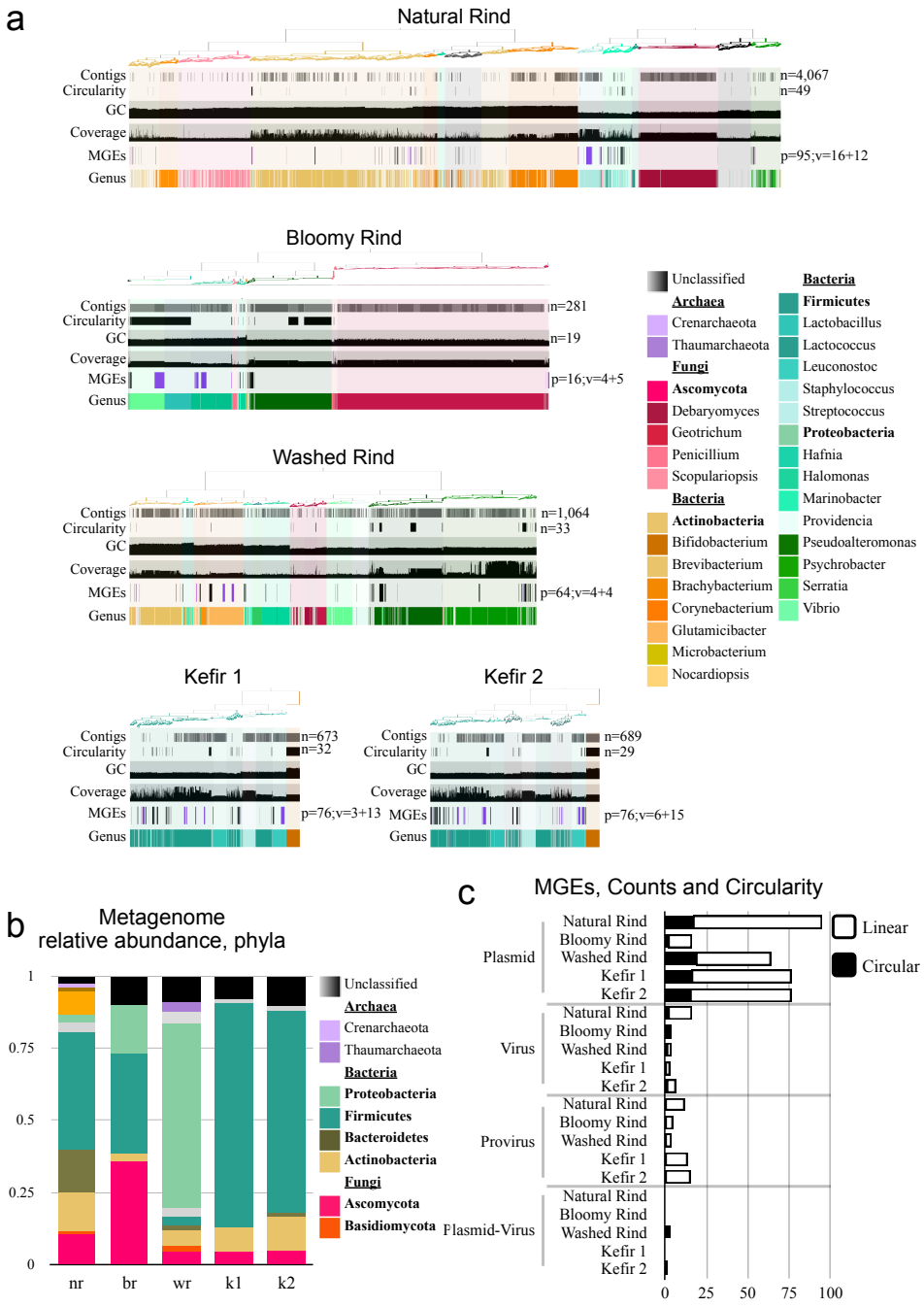


Figure 2.1 Summary of microbiomes. (a) anvi'o plots of each metagenome: tetramer and coverage based hierarchical clustering; Contigs: continuity of contigs with alternating colors for each separate contig with a count of total number of sequences on the right; Circularity: by metaFlye with the count of self-circular contigs on the right; GC: relative GC percentage; Coverage: of the contig by minimap2, MGE: plasmid (black, p = number of plasmids) and viral (purple, v=number of viruses+proviruses) annotations. (b) Relative abundances based on taxonomic assignment from manual binning of contigs at the phyla level. (c) Counts of the circular and linear contigs assigned as the various MGEs.

metagenome-assembled genomes (MAGs) resulting from the assembly and binning of each sample. We searched for conserved single copy genes as a measure of completeness and possible species and strain diversity (Cissé & Stajich, 2019; Parks et al., 2015) (Supplemental Table 2.9). For fungal bins, FGMP (Cissé and Stajich 2019) completeness scores of 96.1% and 95.6%, and similar length to genomes within the same genera suggest that the *Debaryomyces* bin from the natural rind and *Geotrichum* bin from the bloomy rind represent high-quality MAGs (Supplemental Table 2.4). In contrast, the *Candida* and *Scopulariopsis* in the natural rind, *Penicillium* in the bloomy rind, and the *Debaryomyces* and *Fusarium* in the washed rind were largely incomplete (Supplementary Table 2.2); the *Penicillium* bin within the bloomy rind is dominated by its mitochondrion. There were significant fungal species detected within these kefirs, with <1% reads mapping to *Kluyveromyces* or *Saccharomyces***.**

Exploring Plasmid Diversity

Plasmids were determined by using the program ViralVerify (Antipov et al., 2019), which relies on functional annotation of contigs in the metagenome assemblies. These results are shown in the Mobile Genetic Elements (MGEs) row (Figure 2.1a, Supplemental Table 2.4), with black lines indicating plasmid contigs. Out of 162 circular contigs found in the original assemblies, 73 are plasmids (Figure 2.1c, 2.2a). We found an additional 264 linear contigs annotated as plasmids, which may be examples of linear plasmids, unable to be circularized, or incomplete. To examine whether the plasmids from these samples had higher potential for horizontal gene transfer, we annotated 77 plasmids as mobilizable using mob-suite (Robertson & Nash, 2018). We identified 23 mobilizable plasmids in the cheese rinds, 7

of which were also circular. In the kefirs 1 and 2, there were 26 and 28 mobilizable plasmids, 10 and six of which were circular. One plasmid in kefir 1, two plasmids in kefir 2, and 1 plasmid in the washed rind coded for conjugation proteins.

We assigned a putative host to the plasmid contigs (Supplemental Figure 4a). In the natural rind cheese, the plurality of the 95 plasmid contigs were assigned to *Staphylococcus* (28, four of which are mobilizable), *Brevibacterium* (17), and *Lactococcus* (10). One plasmid contig was assigned to the Ascomycota *Fusarium*. The bloomy rind cheese's 16 plasmid contigs are annotated as coming from *Lactocaseibacillus* (3), *Halomonas* (2), *Pseudoalteromonas* (2), *Vibrio* (2), *Oceanisphera* (1), and *Stenotrophomonas* (1, also known as *Xanthomonas*). The washed rind cheese's 68 plasmid contigs are from the genera *Psychrobacter* (22), *Glutamicibacter* (8), *Pseudoalteromonas* (7, 1 mobilizable), *Brachybacterium* (6), *Halomonas* (6), *Vibrio* (3), *Marinobacter* (2, 1 mobilizable), and one each from *Corynebacterium*, *Euzebyella*, *Geminocystis*, *Microbacterium*, *Oceanisphaera*, along with 9 unclassified plasmid contigs. The majority of the 153 plasmid contigs from kefirs were assigned to *Lactococcus*, with additional plasmids in related lactobacilli *Leuconostoc* (14) and *Enterococcus* (1); there was 1 plasmid from *Cercospora* in Kefir 1 and there were 2 unclassified plasmid contigs in Kefir 2.

To better understand whether the collection of plasmids from these metagenomes represented previously-uncharacterized diversity, we used BLAST to compare the contigs to NCBI's nucleotide database (Figure 2.2b, Supplemental Figure 2.5b, Supplementary Table 2.4, 2.6); we permitted matches up to an e-value of 1 to capture distantly related sequences, we then split the matches with an arbitrary 40% coverage. As dairy products have been

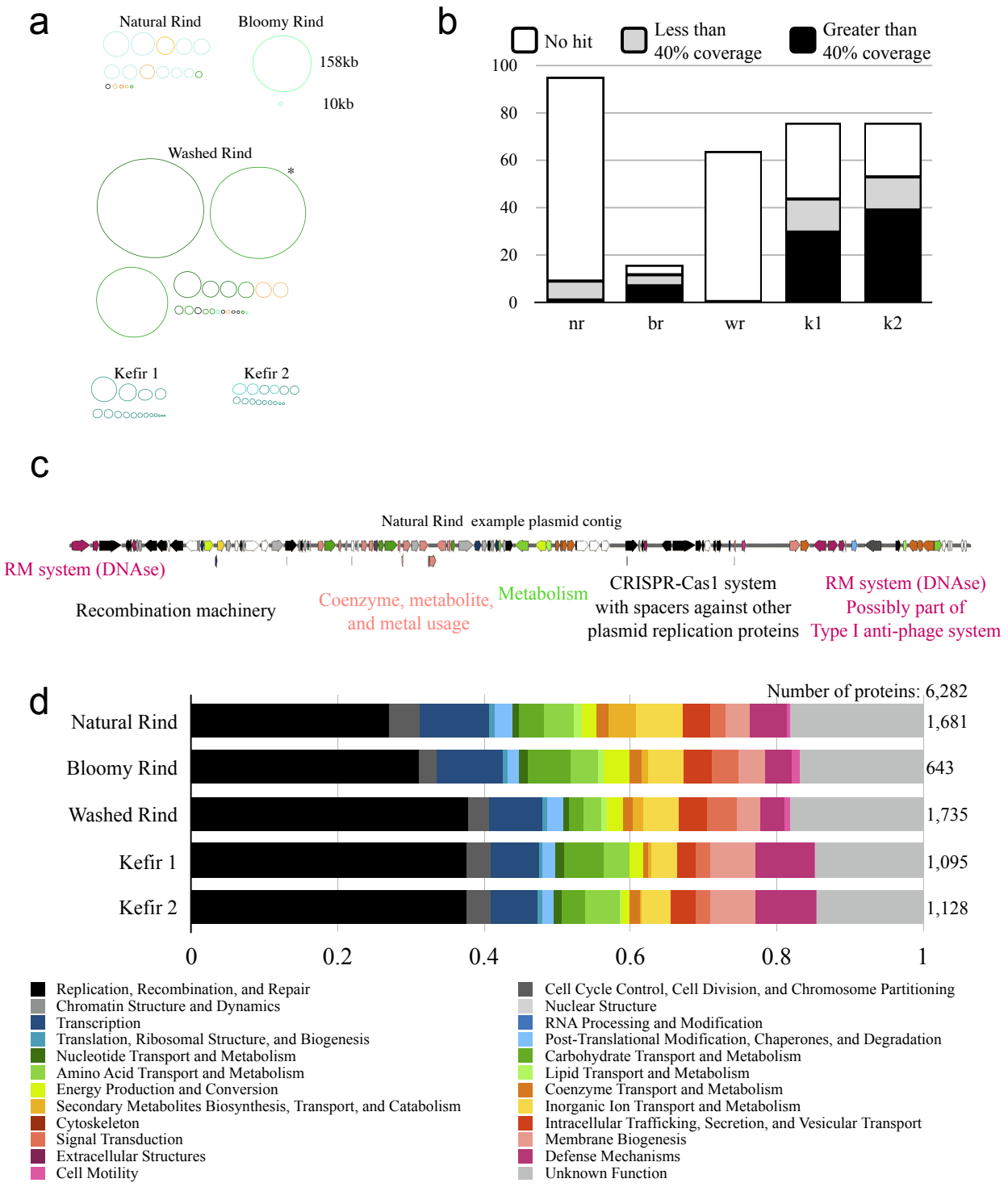


Figure 2.2 Plasmid analysis. (a) bandage plots of circular plasmid contigs, colored by putative genus. (b) BLAST results against ncbi-nt and plasmiDB. (c) example plasmid. (d) relative amounts of COG annotated proteins: black gradients: DNA structure and replication, blues: transcription and translation, greens: energy and macronutrient metabolism, oranges: micronutrients, reds and purples: structural pathways, trafficking, defense mechanisms

studied extensively, we expected many plasmids within these fermented products to match previously sequenced samples, especially *Lactococcal* plasmids (Kelleher et al., 2019). In kefir, half (50.3%) of the plasmids have a hit that covers more than 40% of the plasmid contig. In contrast, the majority of plasmid contigs from the cheese rinds did not have any BLAST matches, and of the 42 with a hit, only one from a natural rind *Brevibacterium*, and seven contigs from the bloomy rind had a coverage match above 40%. The plasmids that did not have any taxonomic assignment did not have a match in the database either, possibly because both the kraken and BLAST database uses NCBI. Within the bloomy rind, 3 of the contigs were a match to a *Lacticaseibacillus paracasei* plasmid.

We next analyzed the gene functions carried on the plasmid contigs using Eggnogger 2.0 and COG functions (Huerta-Cepas et al., 2017, 2019). 6282 genes were identified on the plasmids, with 80% falling into a functional category (Figure 2.2D). Principal component analysis (Supplemental Figure 2.8, Supplemental Table 2.8) of the relative proportion of each COG functional category indicates that there are more common distributions of functions within the mobile element portion, driven both by replication, transcription, translation, as well as metabolism of transport and metabolism of carbohydrates, lipids, and inorganic ions. Within the plasmids, the largest known functional categories are involved in replication, transcription, and translation (27-37.5% of all proteins, Supplemental Table 2.8), with 59 replication initiation proteins (51 repA, 8 repB), 49 plasmid partitioning proteins (24 parA, 25 parB), 84 type IV secretion systems, 36 transposons and 701 transposases, and 271 resolvases, 234 integrases (42 of which were similar to known phage integrases), 32 invertases, and 11 recombinases (Supplemental Table 2.8). An increased proportion of these replication and

partitioning genes are expected as these sequences are searched for by function-based plasmid annotators.

Carbohydrate, protein, and lipid metabolism and divalent cation uptake

We evaluated whether there were genes that had the potential to benefit the host by improving the accessibility and metabolism of nutrients on the cheese rind and milk environment. In Figure 2.2d, the genes involved in energy production are highlighted in green hues, and those involved in secondary metabolite and ion transport are in orange. As the main sugar in milk products, lactose, and its fermentation product lactate are major carbon sources, with trace amounts of glucose and galactose. Within the metagenomes, we identified 222 genes associated with carbohydrate metabolism, 44 of which were specifically lactate transport and dehydrogenases. In each of the kefirs, beta-galactosidase was found on a plasmid that has a high level homology to each other, and this plasmid is mobilizable in kefir 1. This enzyme was also found in bloomy and washed rind plasmids. Other carbon sources include milk proteins and triglycerides. In the cheeses, we find 228 genes associated with protein transport and metabolism. For example, we see peptidase that removes 5-oxoproline from the end of polypeptides; the main protein available in milk, casein, consists of a plurality of prolines and branched chain amino acids (leucine, isoleucine, valine). An additional 42 enzymes are predicted to be involved in fatty acid utilization (~1% of all proteins on plasmids). These were largely limited to the cheese samples, with only two found in kefir plasmids. In contrast, kefir plasmids had an average of 4.3% of their proteins annotated with carbohydrate transport and metabolism.

One major limiting component for microbial growth in dairy products is iron (Monnet et al., 2012); 2 iron-siderophore genes are found on plasmids in the natural rind, 4 in bloomy, and 5 in the washed. *FhuF*, a ferric iron reductase, was found on a Pseudoalteromonas plasmid (298) in the bloomy rind. As iron can be limited in milk products due to lactoferrin and other iron chelators(Kell et al., 2020), lactic acid bacteria are known to use other divalent cations as a replacement(Archibald, 1986). 8 divalent cation transporters are found on plasmids from the natural rind and both kefirs. In kefir, there is a manganese transporter, three explicitly manganese dependent phosphatases, and two transporters of other divalent cations: magnesium and cobalt.

Defense mechanisms, antimicrobial resistance, heavy metals resistance, CRISPR

Genes associated with microbial defense systems may indicate selection factors in these food environments as well, beyond the expansion of nutrient availability. There are 163 defense related genes on plasmids in cheese rinds, and 178 between both kefirs by eggnog-mapper; Resistance Gene Identifier (Alcock et al., 2019) found four antibiotic resistance genes using the strict filter: two genes associated with fosfomycin, and two genes associated with beta-lactam resistance, all in 4 different *Staphylococcus equorum* contigs of the Natural Rind. One of the beta-lactamase-containing plasmids contains the genes for antibiotic resistance regulation, heavy metal efflux and its regulation, as well as a Type I restriction modification system. In both of the beta-lactamase plasmids and one of the fosfomycin resistance plasmids, there is a toxin-antitoxin system, which can be associated with plasmid persistence. In one of the fosfomycin resistance plasmids, there is a cadmium resistance

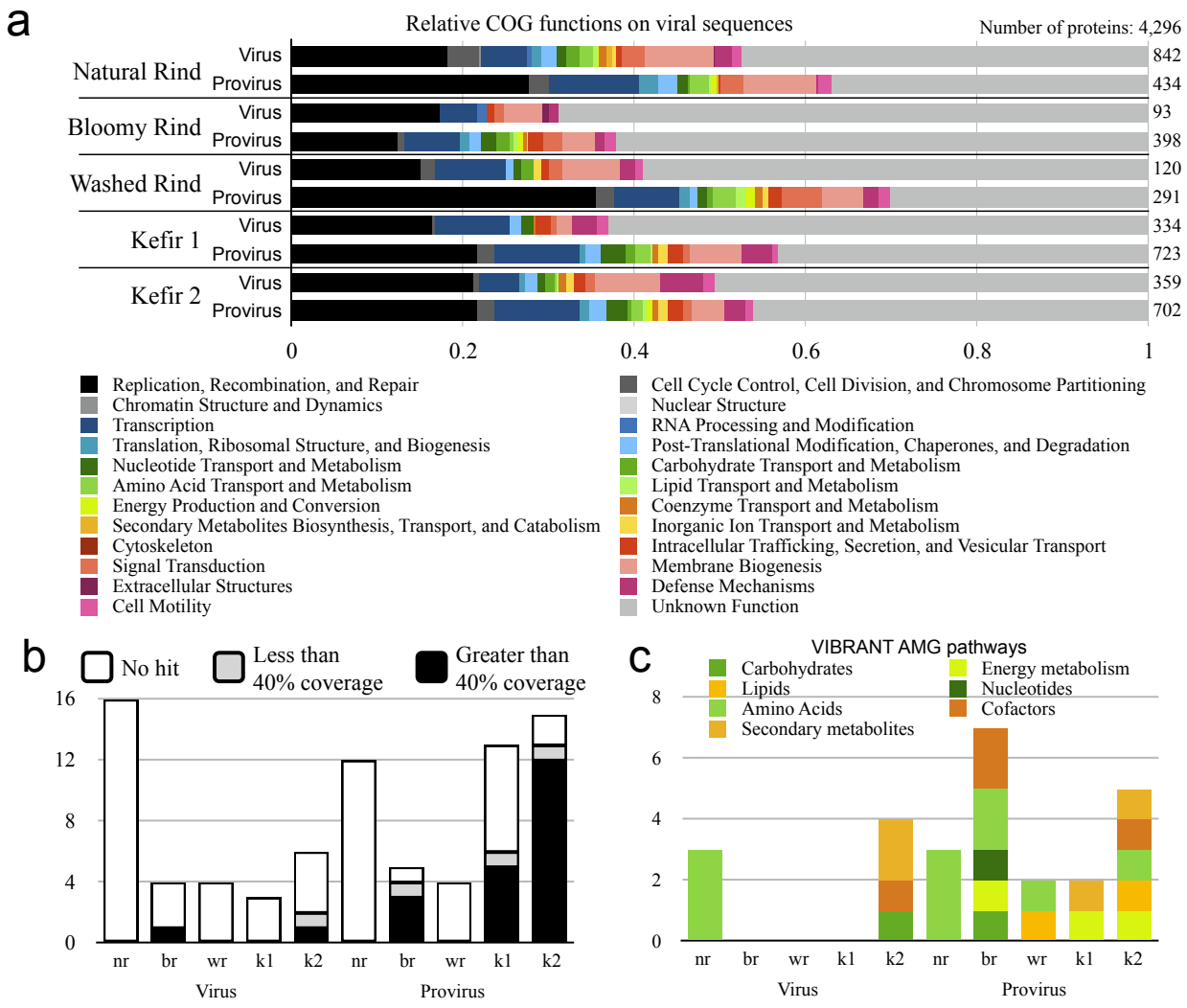


Figure 2.3 Virus analysis. (a) shows relative amounts of COG annotated proteins, as well as the total number of proteins on the right, with re-ordering for similar COG categories. (b) summarizes BLAST analysis of these contigs against ncbi-nt and viralVerify's database. (c) summary of auxiliary metabolic genes of viruses and proviruses of medium, high, and complete quality, with similar color legend as the COG functions.

genomic island. None of these plasmids are mobilizable on their own by mob-suite, though they contain mobilizing genes.

The majority of defense associated genes found on plasmids belong to functional categories involved in preventing the introduction of foreign DNA. We found 227 restriction nuclease-containing proteins (43nr+14br+37wr+65k1+79k2) which are classified in COG

category O: post-translational modification, protein turnover, chaperones; and category L: replication, recombination, and repair); and one CRISPR-Cas1 system on a washed rind plasmid. Using CRISPRCasFinder, we find spacers in the natural rind and washed rind (Couvin et al., 2018). CRISPR-Cas systems on plasmids appear to be biased against targeting other plasmids (Pinilla-Redondo et al., 2021), but may also confer their traditional immunity by enhancing restriction endonuclease activity against phage in DISARM systems (Ofir et al., 2018): we find examples of both in cheese rind plasmids. In one of these plasmids in the natural rind (Figure 2.2c), we find an intact CRISPR system with spacers targeting the *rep* gene of another plasmid. In addition, a putative CRISPR array on a *Moraxella* plasmid in the natural rind (5624) may target a *Psychrobacter* phage, as well as other uncharacterized phage. However, there were no Cas proteins identified within this plasmid, nor on contigs within the same genus. In the washed rind, there were 93 unique confident spacers found, with two *Psychrobacter* plasmids, each containing a Cas protein cluster along with 44 confident spacers on one, and 39 spacers on the other. Similarly, some of these spacers target replication initiation proteins on *Psychrobacter* plasmids and other viruses. In one case, the spacers target a RelB-like antitoxin on *Psychrobacter* plasmid, which could also limit the ability of a new plasmid from creating dependency, instead destroying poisoning the host with the relevant toxin; in another set of *Psychrobacter* spacers target *Pseudomonas* phage.

Exploring viral diversity

We next analyzed the metagenomes for viral-associated contigs using VIBRANT (Kieft et al., 2019), analyzing sequences for specific sets of proteins and protein families to

classify contigs as either plasmid or viral (Figure 2.1c). In total, 33 contigs were confidently assigned as viral and 49 chromosomal contigs contain prophages. We found higher numbers of viruses in the natural rind cheese sample (16 viral and 12 prophage contigs) and the kefir samples (9 viruses and 28 prophage), compared to the bloomy and washed rind cheese samples, which had a total of 4 viruses each and 5 prophages in the bloomy rind and 4 prophages in the washed rind.

Some of the viral contigs clustered with bacterial contigs, based on kmer similarity, indicating their possible host (Supplemental Figure 2.4b). For the natural rind, there are 12 contigs annotated as viral, and these are related to *Staphylococcus* (3), and two to *Brevibacterium* and one each to *Streptomyces* and *Sphingobium*. Of the 12 prophages, three are on *Staphylococcal* contigs, two on *Brevibacterium* contigs, and one each on *Streptomyces* and *Sphingobium*. Some of these prophages show both lysis and lysogeny, with long reads mapping across their circularized state and long reads that map across the genome into the *Brachybacterial* genome (contig 2619). Between the five bloomy rind prophages, three potentially target *Hafnia* (example contig 593 in Supplemental Figure 2.9a), and one each target *Staphylococcus* and *Vibrio*, and this rind's phages were all in lactobacilli. The washed rind contained two viruses with a significant kmer match: with one targeting *Glutamicibacter* and *Pseudoalteromonas*. One of the prophage contigs in the washed rind remained without a known host. However, the other prophages were found in *Glutamicibacter* (3) and *Psychrobacter* (1). One of the *Glutamicibacter* prophages has long reads that map across the circularized region, suggesting that a subset of these prophages were in a lytic cycle. This excised prophage appears to be present at half the abundance of its host, and we are able to

capture this phenomenon with a read depth of 5-10X (contig 337). The kefirs were dominated by *Lactococcus* phage and prophage, with one contig unassigned and one phage against *Streptococcus* in Kefir 2. Similarly for the prophages, most of these were found in *Lactococcal* (Kefir 1: 15; Kefir 2: 13) and *Leuconostoc* (Kefir 1: 3; Kefir 2: 2), with some evidence of active lysogeny as well (Example of Kefir 1 contig 595 in Supplemental Figure 2.9b).

In addition to these phages, there are some eukaryotic viruses in these environments. In the natural rind, there is a small 2kb gemycircularvirus, a very small virus that mainly infects fungi; in the bloomy rind, there are three contigs (157, 281, 423) that may be viral and cluster with *Geotrichum* and *Penicillium*. Even in the kefirs, there were contigs with kmer similarity and strong alignment to known phage, causing a kraken hit to the appropriate virus, but protein level matches were incomplete.

To better characterize the diversity within the viral population in these foods, we applied vcontact2 (Zablocki et al., 2019) to perform protein-based searches against previously-sequenced viruses (Supplemental Figure 2.6). When applying these clustering methods, we find that many contigs cluster together, suggesting either closely-related groups of viruses found in these samples or these contigs are fragments of the same viral genome. Within these clusters, we find that there are some contigs that share a large number of genes, but have different insertions. To further assess for viral novelty, we BLASTed these contigs against NCBI-nt (Figure 2.3b). Similar to the plasmids, we considered this a new virus if the best alignment covered less than 40% of the contig. Overall, all but one of the 24 cheese rind

viruses and 18 of the 22 prophages are novel. In kefir, 8 of 10 lactobacilli phage and 11 of 31 prophages were novel.

We then analyzed the genes encoded on these viral and proviral sequences. Some phage increase their host fitness to improve access to carbon and energy to maximize phage replication (Hurwitz et al., 2013). Of the 4,296 virally-encoded genes, 52% could be assigned to a known functional category, much less than proteins encoded on plasmids or chromosomes (Figure 2.3a). Similar to the plasmids, the viral contigs have a high proportion of replication, recombination, and repair proteins as these are often used to classify contigs as viral (Supplemental Figure 2.8, Supplemental Table 2.8). The other significant section includes membrane biogenesis, which includes many phage structural and tail proteins, and defense mechanisms, which includes restriction modification systems. Prophage may even encode fitness-improving metabolic pathways to increase their host abundance before activating their lytic cycle; as they do not directly benefit the prophage, and add more on the proviral genome, these added genes are sometimes called auxiliary metabolic genes. Of the medium and high quality viruses, we find 26 auxiliary metabolic genes, mostly associated with metabolism of carbohydrates, proteins, and lipids (Figure 2.3e), highlighting PurAMN proteins in the bloomy rind and an iron-sulfur binding protein in a bloomy rind provirus. These nucleotide metabolism genes have been found to be important in the cheese rind environment (Morin et al., 2018; Pierce et al., 2021), and viruses can potentially increase the production or availability of these nucleotides for their own replication. In this case, prophage encoding of purine related genes may also support the replication of the host before entering

its lytic cycle. Indeed, prophages are also more likely to contain genes associated with metabolism (Fig3a, Supplemental Table 2.8).

We also examined the contigs that were identified as both plasmid and phage (Brown-Jaque et al., 2015), that exemplify the lack of binary divisions between plasmids and phage, including Gene Transfer Agents (Lang et al., 2012). Currently, these are more rare in our communities, but no tools exist that specifically identify these hybrids. With our strict criteria, we find 1 *Lactococcus lactis* plasmid-phage in Kefir 2 contig 455 that encodes for folate metabolism and EcoRI Type I restriction modification system; and 1 plasmid-prophage in *Psychrobacter* in the washed rind contig 1206 that encodes for iron-sulfur binding domain, Phd/YefM antitoxin, DNA integration proteins, type IV secretion system.

Discussion

In this work, we use metagenomic long read sequencing of fermented food-associated communities to provide a detailed snapshot of the microbiome and their associated plasmids and viruses. Leveraging the relative ease of assembling mobile elements from long reads, we were able to extract a large, diverse set of plasmids and viruses. These genetic elements encode bacterial genes that may increase fitness within these controlled environments.

Mobile genetic elements can provide much insight into various environments, but finding them from natural communities is difficult. Genes and genomic islands from plasmids and viruses are more difficult to analyze due to higher levels of repetitive regions and conserved necessary proteins, such as those associated with replication and partitioning. With

short reads, it is possible to show adaptive functions on plasmids and other MGEs (Youngblut et al., 2020), but it is more difficult to associate them with their host (Maguire et al., 2020). As their assembly tends to be more broken, mobile elements may not be associated with their hosts as kmer based approaches are highly limited on short sequences. By using long reads in a natural environment, we find a trove of genetic elements that have not been previously sequenced, even in simple and often studied microbiomes and genera. Additionally, five rounds of consensus based polishing, followed by protein based frameshift corrections allowed for the use of conservative protein-based methods, which would often fail on the original assembly due to systemic errors in nanopore long reads. As suggested by others (Khot et al., 2020), we also suggest using a combination of tools to have greater confidence in the subset of plasmids or viruses, especially when attempting to complete a more comprehensive catalog.

Taxonomically, the cheese rind microbiomes were similar to those rinds of the same type. Surprisingly, the kefir samples did not contain many reads assigned to fungal taxa, despite previous reports of the Ascomycota Kazachstania, Kluyveromyces, Naumovozyma, and Saccharomyces in kefir (Marsh et al., 2013). Instead, the kefirs were dominated by several strains of lactobacilli (Fig 1c), similar to those grains and milks studied previously (Walsh et al., 2016). These lactobacilli dominate a large variety of liquid ferments, where microbial growth is primarily controlled by salt and low oxygen concentrations; they are widely studied due to their importance in food, both culturally and for preservation.

Despite using a conservative, homology-based, method to classify MGEs, we are able to find a surprisingly high number of plasmids and viruses within these fermented dairy

products. As many of these genera are important for food production, we expected many plasmids within these fermented products to match previously sequenced samples. In particular, lactic acid bacteria, and their plasmid and viruses, have been extensively studied and recently reviewed in (Ainsworth et al., 2014), (Davray et al., 2020; Kelleher et al., 2019), reporting that plasmids in lactobacilli are likely to improve fitness within a dairy environment. However, the majority of plasmids and viruses we found were novel. Overall, we found 327 plasmids, 33 viruses, and 49 prophage. We find that both plasmids and phage are less common on the bloomy and washed rinds compared to the natural rind and kefirs. Kefirs, containing many strains of *Lactococci* that have been reported to host up to 12 plasmids (van Mastrigt et al., 2018), can be expected to have more mobile elements, possibly coding for an average of 5 plasmids (Ainsworth et al., 2014);(Kelleher et al., 2019). Similar to previous studies (Davray et al., 2020), the genetic elements in all samples reveal many proteins related to genomic replication and repair, while also giving additional insight into the biotic and environmental stressors, such as heavy metals, including the efflux of excess toxic metals. Heavy metals resistance can be used in industrial strains of lactobacilli to ensure the maintenance of certain plasmids and these may be transferred. The putative hosts of viruses and plasmids are identified from kmer frequency similarities, though definitive assignment would require isolate genome sequencing, proximity ligation, or single-cell sequencing by isolating mobile elements with their hosts through physical means. By using 31-mers and inserting Ns using protein-based contig polishing, we increase our chances for accurately identifying relatively narrow-host range MGEs (Edwards et al., 2015).

Furthermore, these plasmids can encode for proteins that provide additional access to carbon from sugars, proteins, and fats. The existence of these plasmids may provide the host enough of a fitness advantage to outcompete their peers. Plasmids were much more likely to contain proteins related to defense mechanisms than chromosomal or viral sequences. We also find four antibiotic resistance genes in the natural rind that may be due to the constant antagonism between certain bacterial and fungal species that have been co-evolving for many years.

Viruses can drastically change the microbial community by decimating a specific species or acting as a mediator of horizontal gene transfer(Braga et al., 2020; Breitbart et al., 2018; Eggers et al., 2016). Some proviruses have even been suggested to actively gain genes within a strain of lysogenic Streptococcus(Haaber et al., 2016). We find 33 viruses and 49 proviruses, largely bacteriophages, in these simple communities. The majority of viruses were phage in the natural rind, and most prophages are in Lactococcal genomes (Supplemental Figure 2.4b). Though these are much less well annotated than plasmids, we are still able to find 165 genes related to bacterial metabolism and survival. As we are using long reads, it can be easier to assess for recircularization of integrated prophages by comparing the ratio of reads that map across the two ends of the prophage region against those that map from the prophage to the chromosomal region. There were 3 Actinobacterial and 1 Synechococcus prophage in the natural rind, 2 Actinobacterial prophage in the washed rind, and 1 Lactococcus lactis prophage in Kefir 1.

With few supplies, dynamics of microbial communities can be captured with a snapshot, showing viral activity and the presence of accessory gene agents that may be selfish and support microbes within the dairy niche.

Materials and Methods

Sample preparation and sequencing

DNA from cheese rinds were scraped and kefirs were extracted using liquid nitrogen grinding and phenyl-chloroform, with kefir 2 specifically extracted for high molecular weight DNA (Quick, 2018). These were then prepared with LSK-108 library prep kit and sequenced on ONT minION 9.4. Reads were basecalled with guppy 3.0 using the high accuracy model. Quality graphs were generated with nanoplot and available in supplementary materials.

Assembly and polishing

Long reads were assembled with metaFlye and polished using racon (v1.4.7,(Vaser et al., 2017)) four times, using reads remapped to each polished assembly using minimap2. This is followed by a final read-based polishing using medaka (v1.0.3) and frameshift correction with DIAMOND (Buchfink et al., 2015). A DIAMOND database was generated from all RefSeq (v 93) protein sequences from archaea, bacteria, fungi, plasmid, protozoa, and viruses(Buchfink et al., 2015). The DIAMOND blastx were completed with a frameshift allowance of 15 (-F 15 --range-culling --top 10 --outfmt 100), recommended for high error rate reads or organisms with significant divergence from previously sequenced organisms (Arumugam et al., 2019). Comparison of protein calls before and after frameshift correction is in Supplemental Figure 2.7. Final contigs used for further analysis were extracted from this diamond database using MEGANized with MEGAN Community Edition, version 6.19.8(Huson et al., 2016, 2018), with provided databases and default values with the long-reads option (--longReads --

lcaAlgorithm longReads --lcaCoveragePercent 51 --readAssignmentMode alignedBases --acc2taxa prot_acc2tax-Nov2018X1.abin).

Taxonomic Annotation and identification of plasmids and viruses

A custom kraken2 database (CustomKrakenDB, n.d.) was created from RefSeq nucleotide sequences (v93, archaea, bacteria, fungi, plasmid, protozoa, and viruses) with several added fungal genomes generated from within the Dutton Lab and in collaboration with the Wolfe Lab (Wood et al., 2019). Unclassified contigs were annotated using tetramer clustering that was integrated into Anvi'o and supported with Internal Transcribed Spacer (ITS) and 16S sequences, and MEGABLASTing of uncategorized sequences to larger databases.

Functional Annotation

Gene calls on all contigs were done using prodigal and these protein sequences were categorized by EGGNOG-MAPPER (v2,(Huerta-Cepas et al., 2017, 2019; Hyatt et al., 2010)). Plasmids were annotated with ViralVerify (-p, (Antipov et al., 2020)), and virus and prophage sequences were identified with VIBRANT (Kieft et al., 2020). CRISPR proteins were identified in plasmid sequences using CRISPRCasFinder (Couvin et al., 2018), with spacers of evidence levels 3 and 4 were submitted to CRISPRTarget (Biswas et al., 2013) to map these spacers against these databases: Genbank-Phage, RefSeq-Archaea, -Plasmid, and -Viral, as well as Genbank-Viral, and IMGvr. Antibiotic resistance genes were further identified using RGI against the CARD database (Alcock et al., 2019). Automated binning

was completed with DASTool (Sieber et al., 2018), combining the binning results from concoct (Alneberg et al., 2014), maxbin2 (Wu et al., 2016), and metabat2 (Kang et al., 2019).

Metagenome visualization

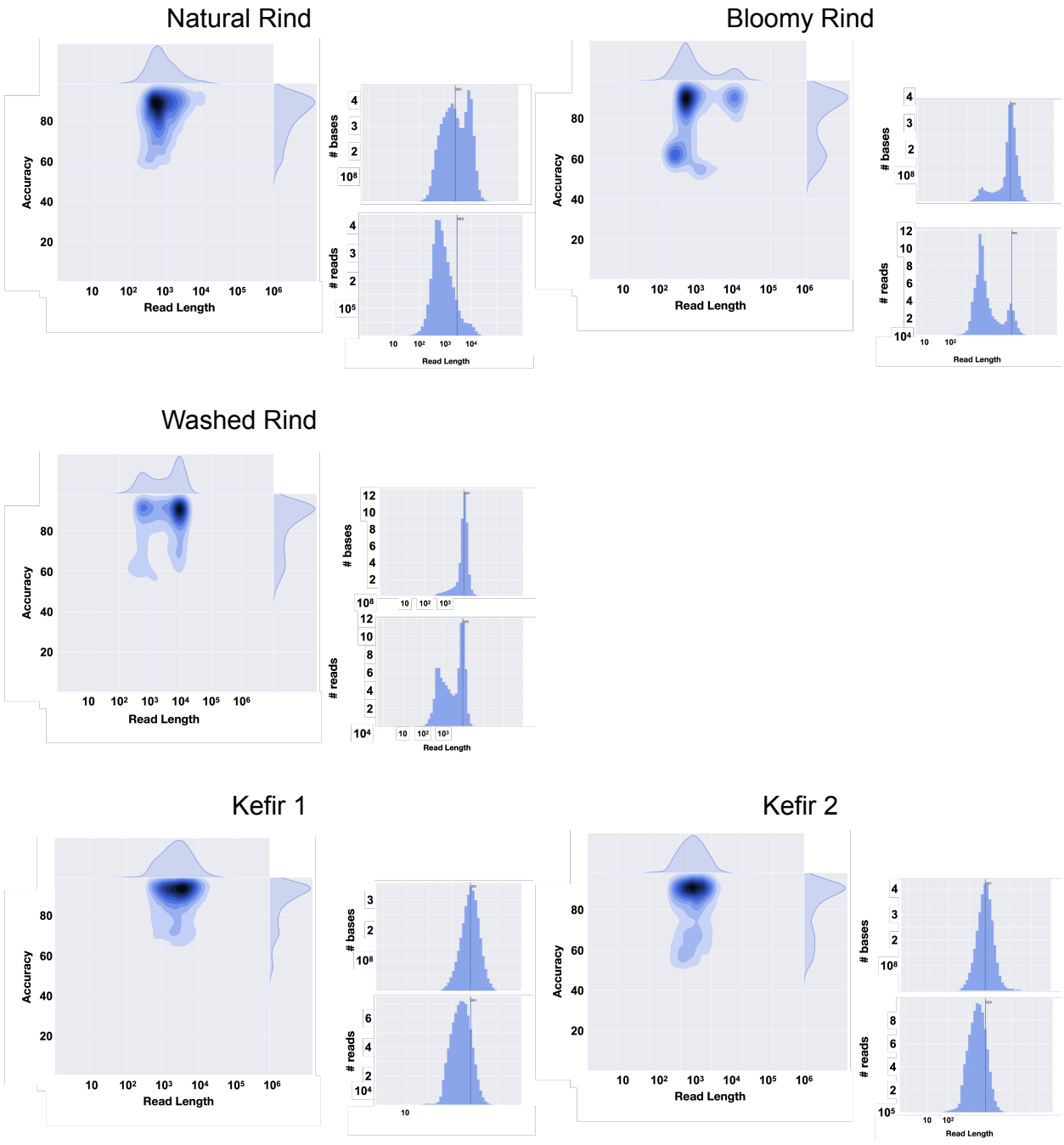
Anvi'o (v7.1 Hope) was used to visualize this data. Read alignment was completed with minimap2 (-ax map-ont), and these alignments were used for contig database construction and profiling (--cluster-contigs) for each sample. We then used a custom script to create an additional miscellaneous data table for anvi'o (anvi-import-misc-data). Within anvi'o interactive web browser, we manually binned the tetramer-and-coverage-clustered contigs by genus assignment by kraken and manual assignment from BLAST and ITS sequences. These bins were defined as metagenome-assembled genomes as high quality if over 90% complete and less than 10% redundant. Coverage by contig and bin were exported (anvi-get-split-coverages) and used to create the relative abundance plots. Unmapped reads were determined using samtools (v 1.10,); coverages are normalized by contig length, not by genome size.

Source dairy fermentations

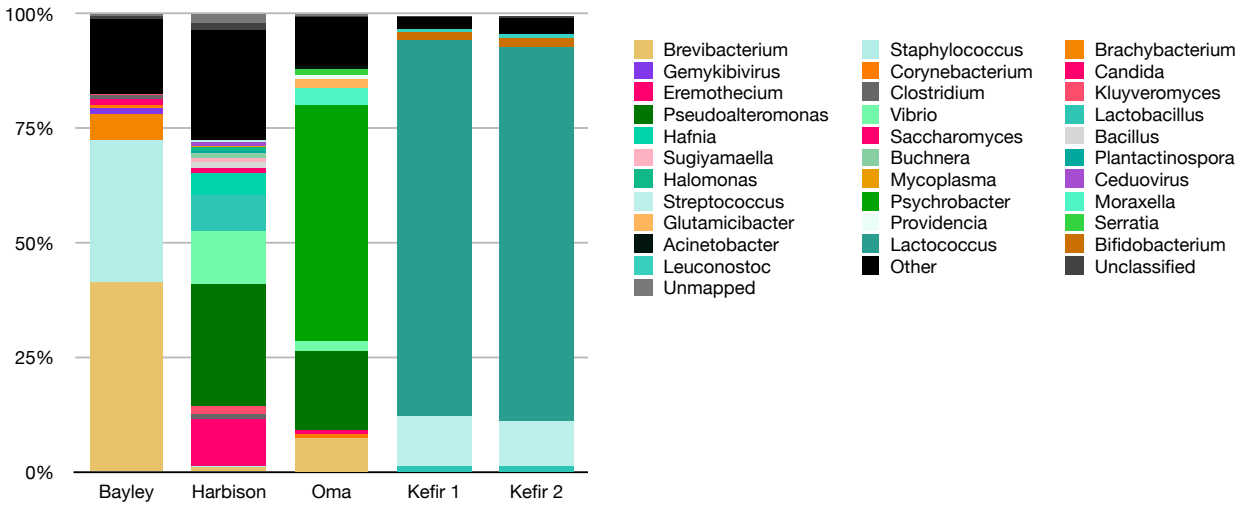
Kefirs are created by inoculating milk with kefir grains, which are matrices of proteins, lipids, and sugars containing a symbiotic community of yeasts and lactic acid bacteria. Over the next 12-24 hours, fermentation of the milk is carried out by the community, after which the grains can be strained and reused. Cheeses are also made from cows' milk, to which lactic acid bacteria and chymosin are added to create the initial curd, the body of the cheese. These cheese curds are then allowed to age within controlled environments to create specific and

reproducible biofilms on the surface, called cheese rinds. The three cheese rinds in this study are produced using three different methods. The natural rind cheese is formed from unpasteurized milk, with the curd inoculated with Penicillium roqueforti to create a blue cheese. The surface of this cheese is allowed to develop with minimal disturbance of the rind. The bloomy rind cheese is created from a pasteurized milk, the surface is inoculated with fungi, then wrapped with spruce bark; the characteristic white and fluffy bloom on the rind comes from the added fungi. The washed rind cheese is created from unpasteurized milk, but during the aging period the surface is regularly washed with a brine solution. All of these cheese rinds were sampled after about three months of aging within a damp cave environment.

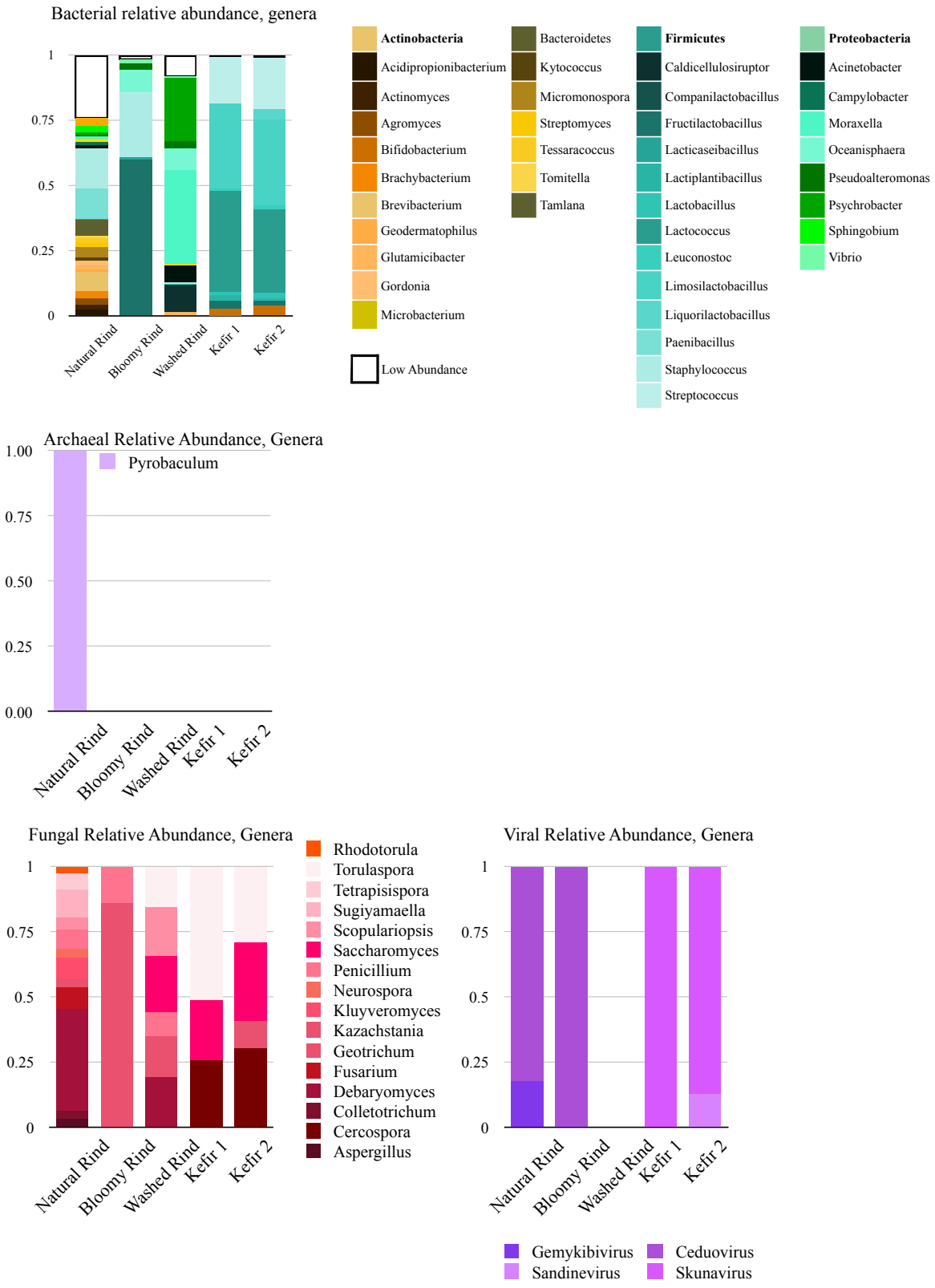
Supplemental Figures on next page



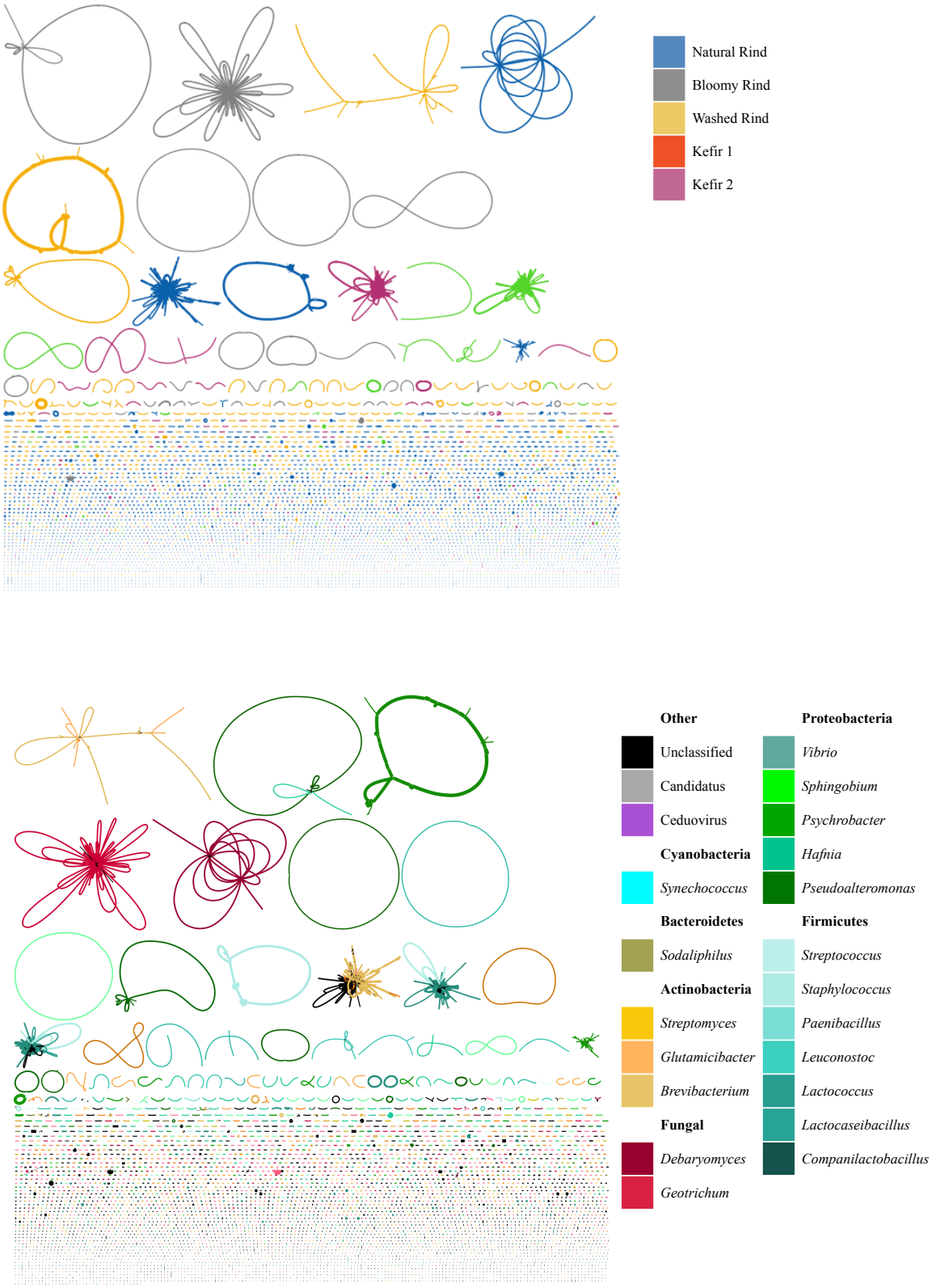
Supplementary Figure 2.1 Summary of sequencing. Within each sample, the left graph shows a heat map of log read lengths and their average accuracies, as determined by guppy v3.0.3. The right graphs show a linear-log graph of read length and number of reads (top) and bases (bottom) sequenced. N50s noted in the right graphs are as follows in bp; Natural rind: 2,938; Bloomy rind: 10,873; Washed Rind: 9,680; Kefir 1: 6,363; Kefir 2: 6,111.



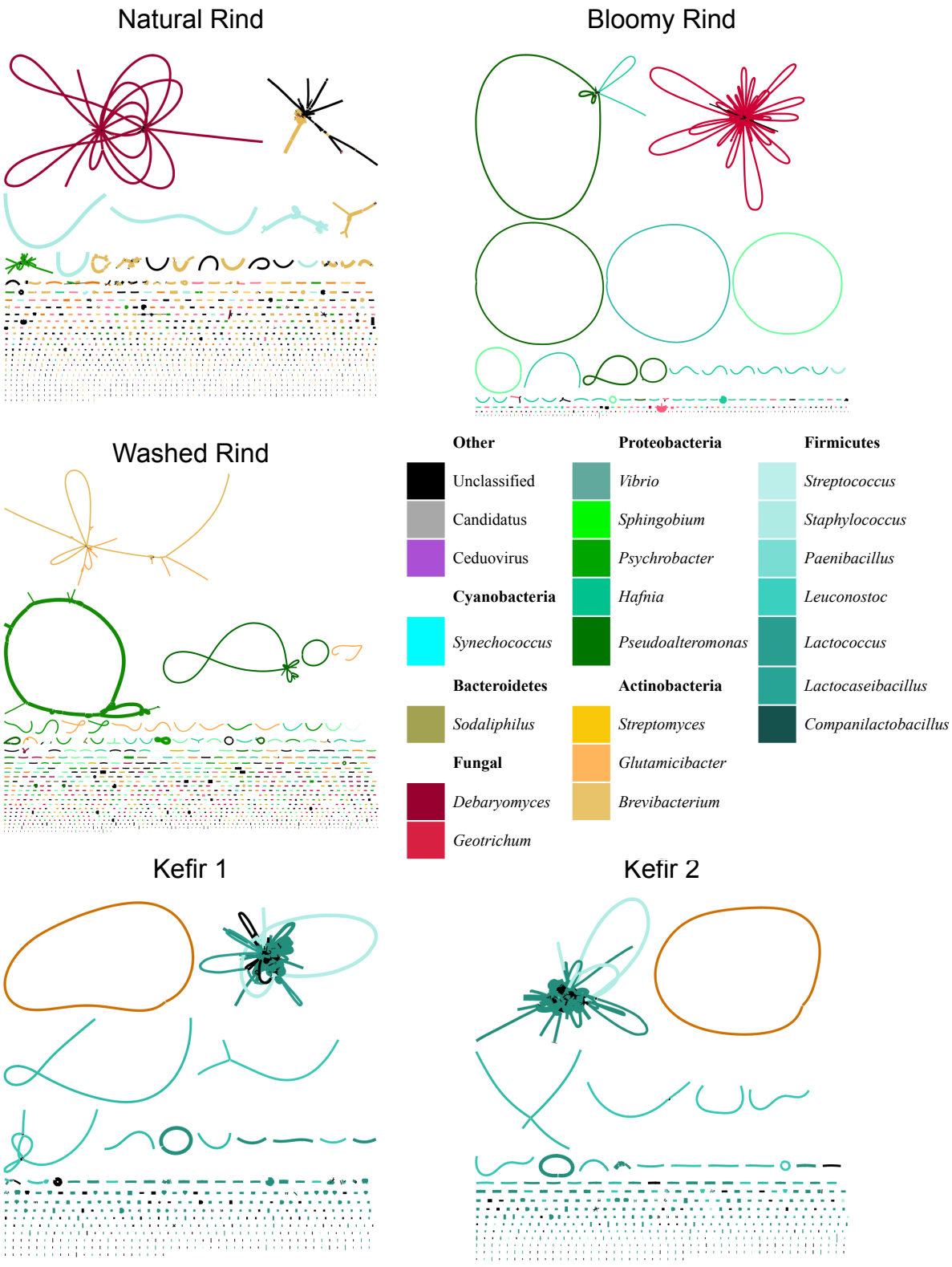
Supplemental Figure 2.2 Relative abundance. Metamaps results of raw reads against a standard database. Here, Unclassified denote the unmapped reads and those that may later map onto fungi that do not have representation within the database.



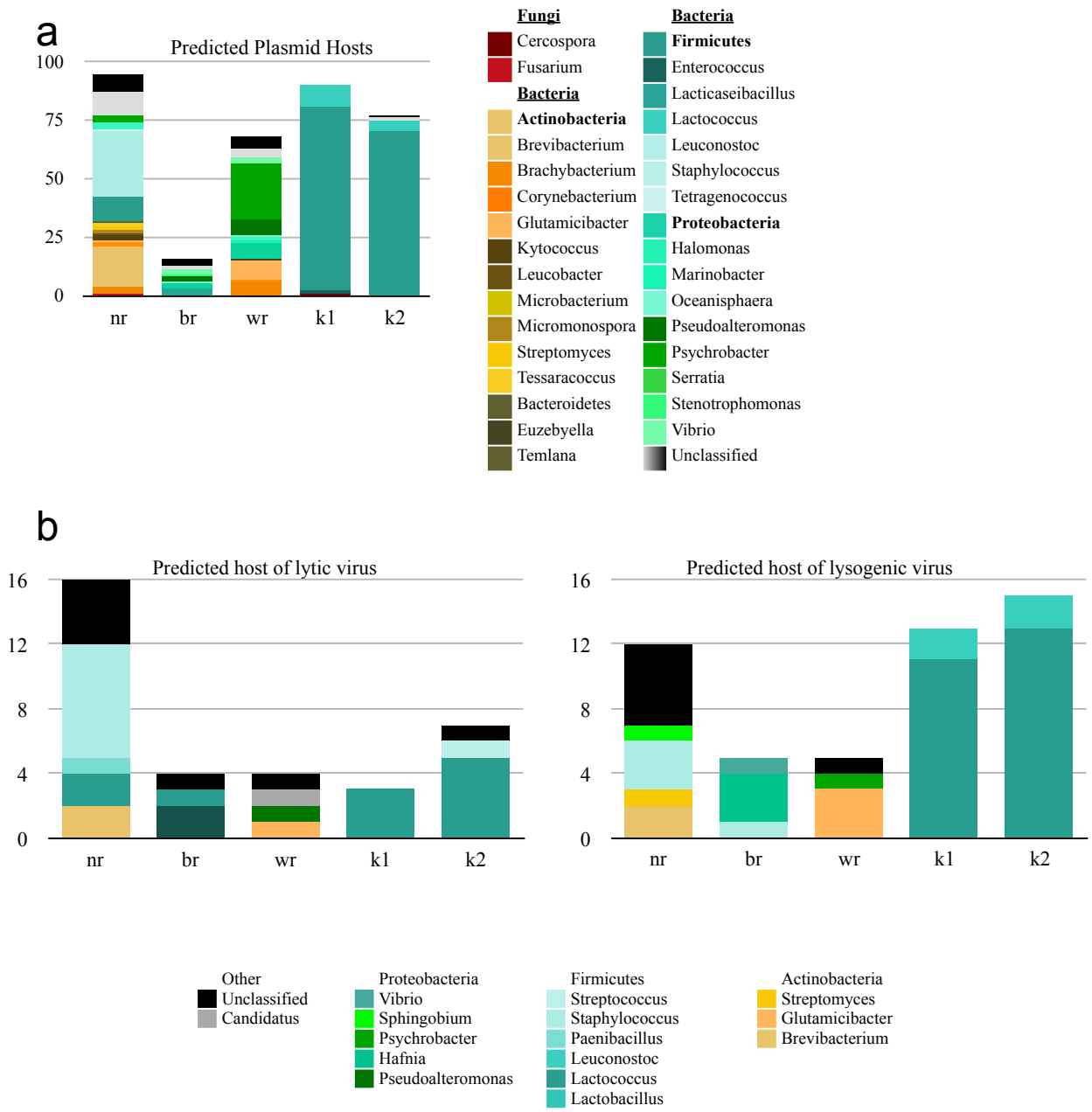
Supplemental Figure 2.2, continued Relative abundance. Minimap2 alignment to kraken 2 taxonomically called contigs.



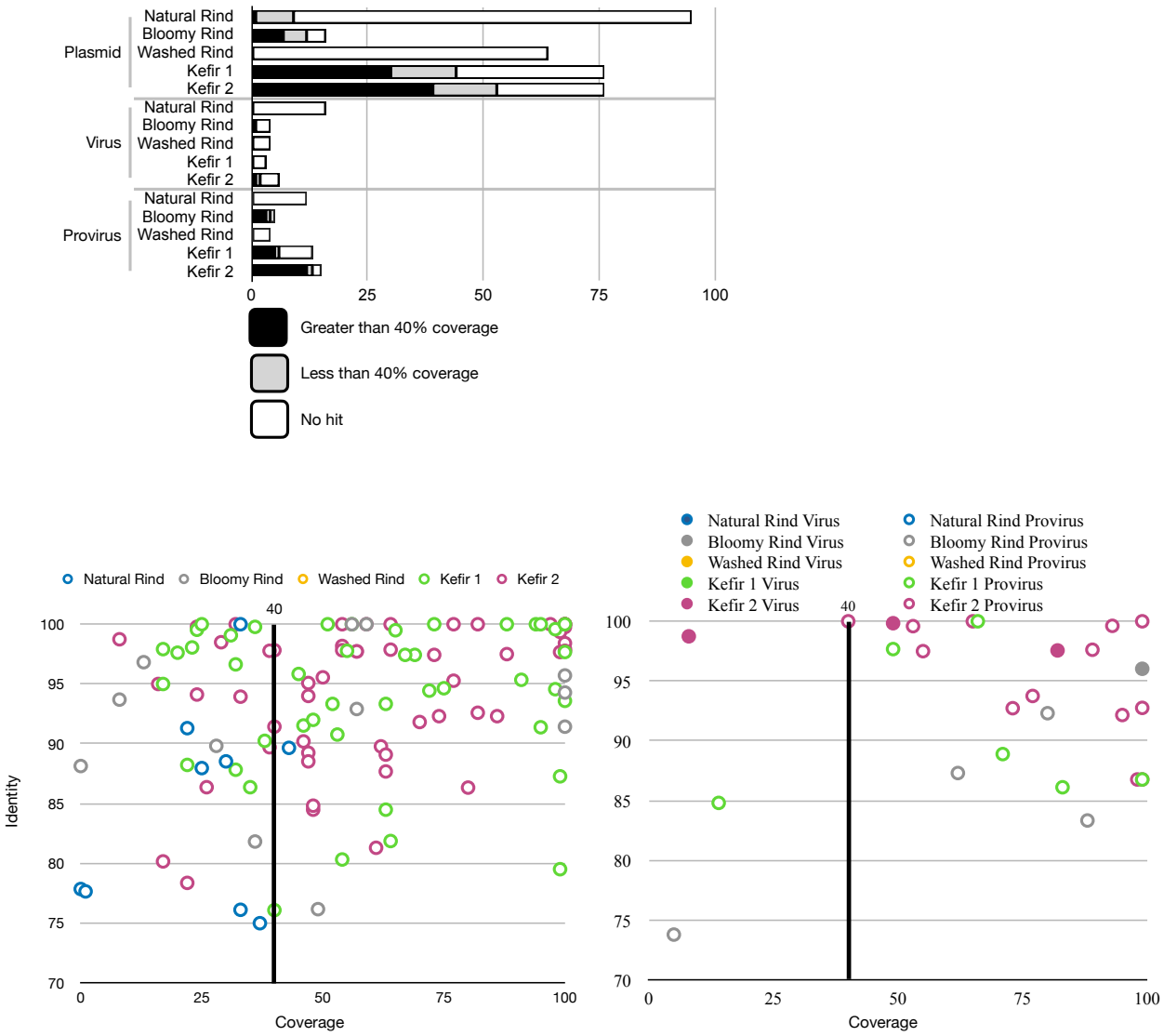
Supplemental Figure 2.3 Total metagenome assembly graph, colored by (top) sample and (bottom) genera.



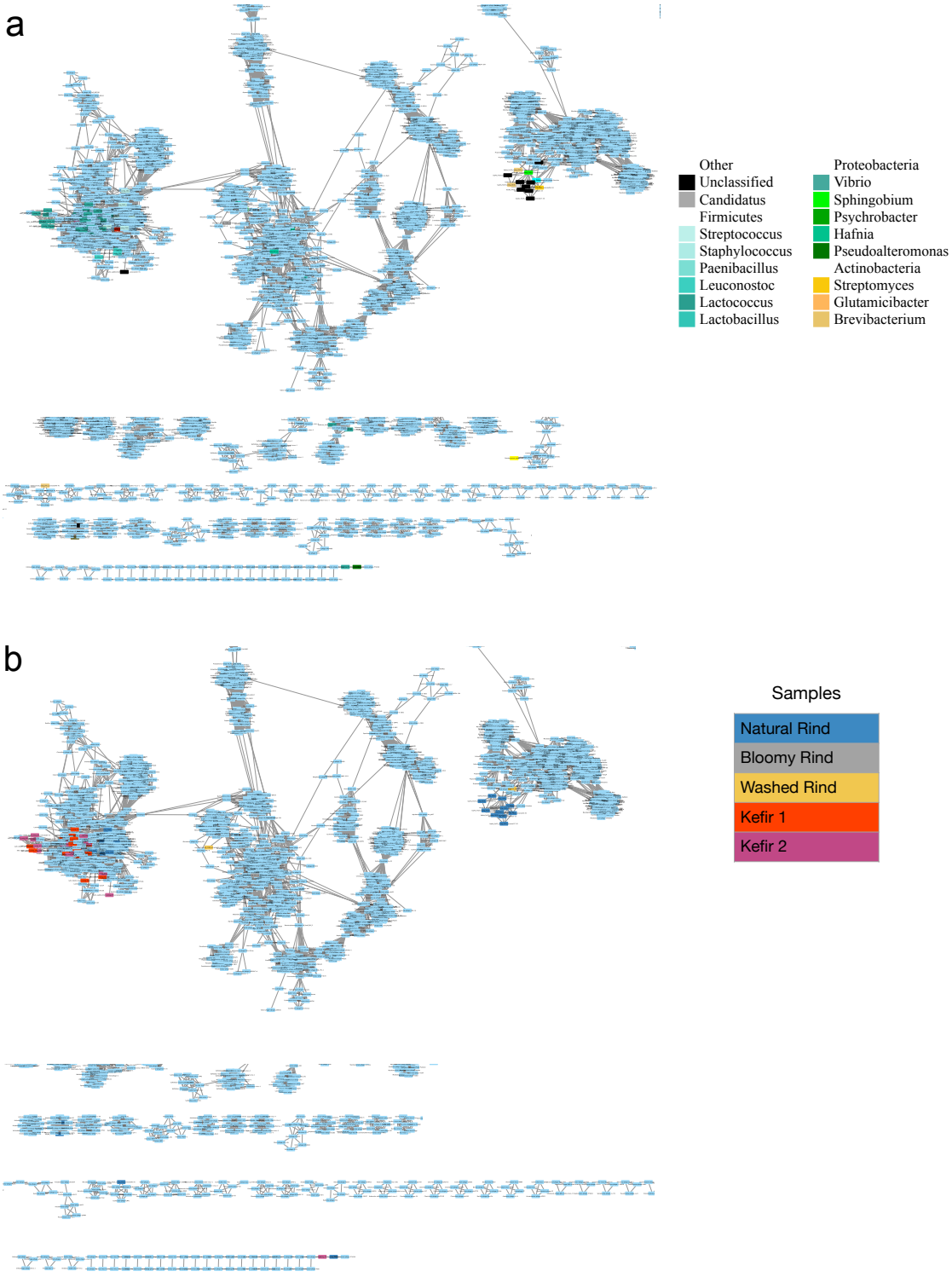
Supplemental Figure 2.3, continued Metagenome assembly graph, colored by by genera.



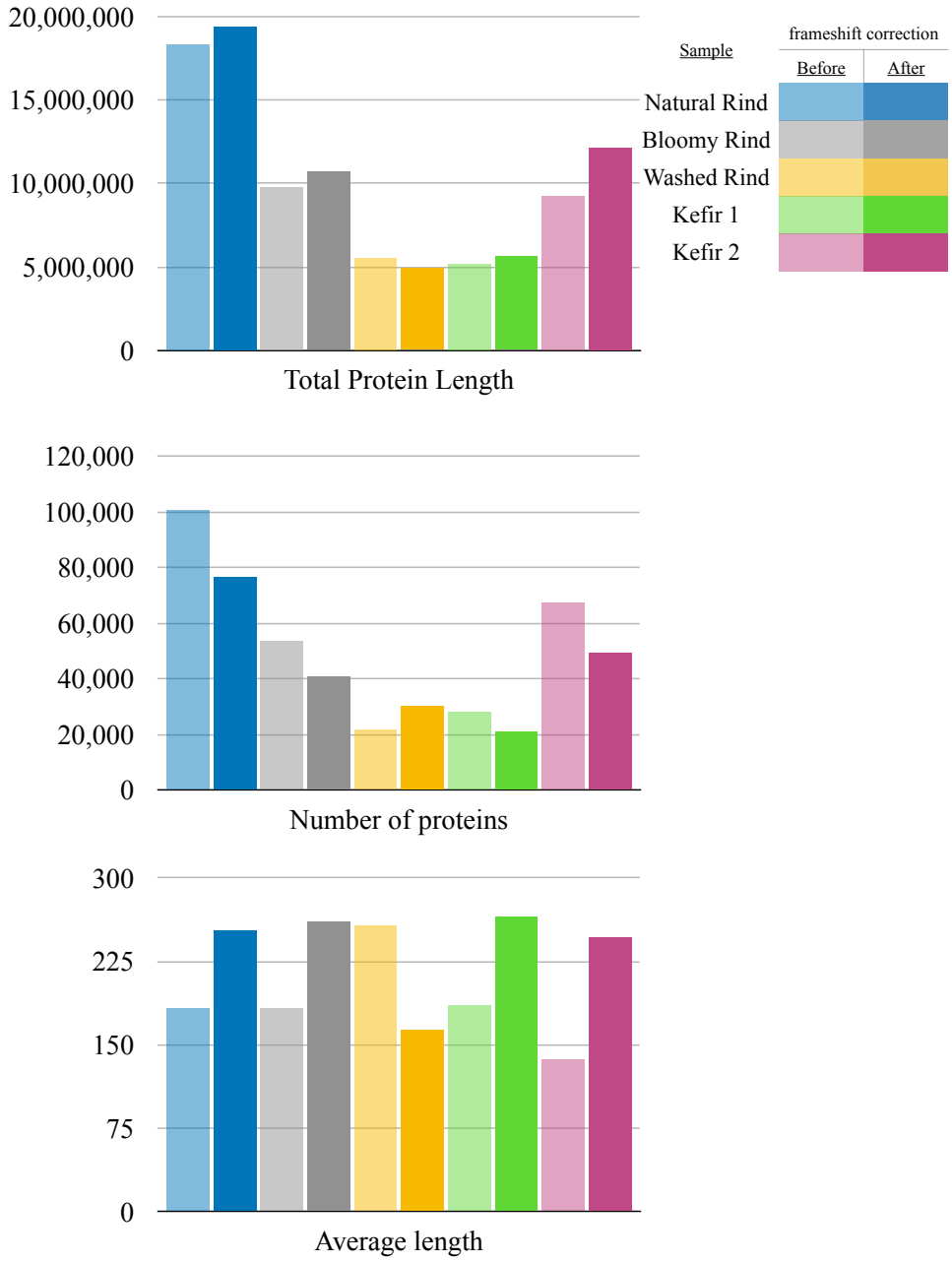
Supplemental Figure 2.4. Putative host assignment to (a) plasmid and (b) viruses by kraken2.



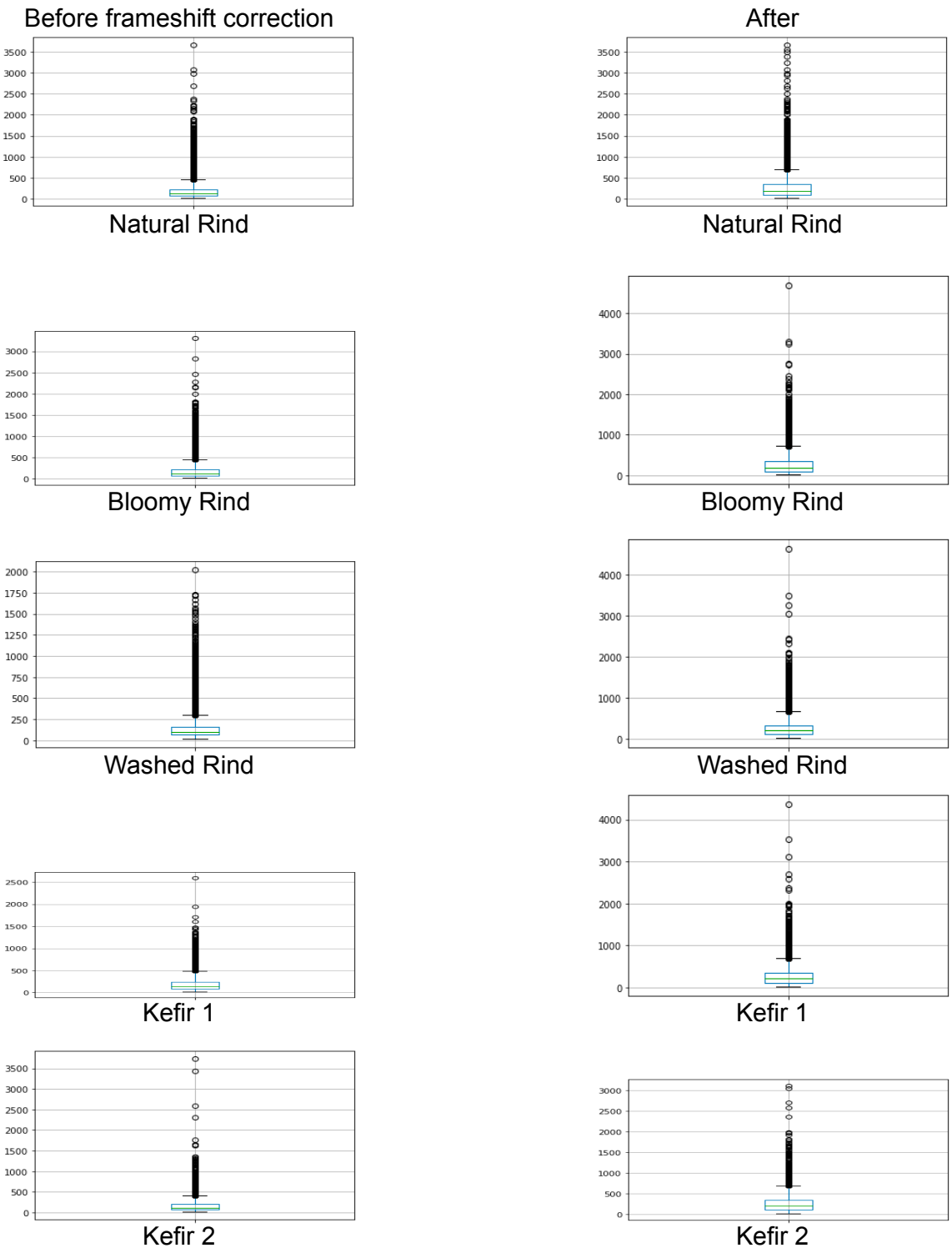
Supplemental Figure 2.5. Novelty of plasmid and viral contigs from BLAST alignment against NCBI-nt, PlasmiDB, and IMGvr.



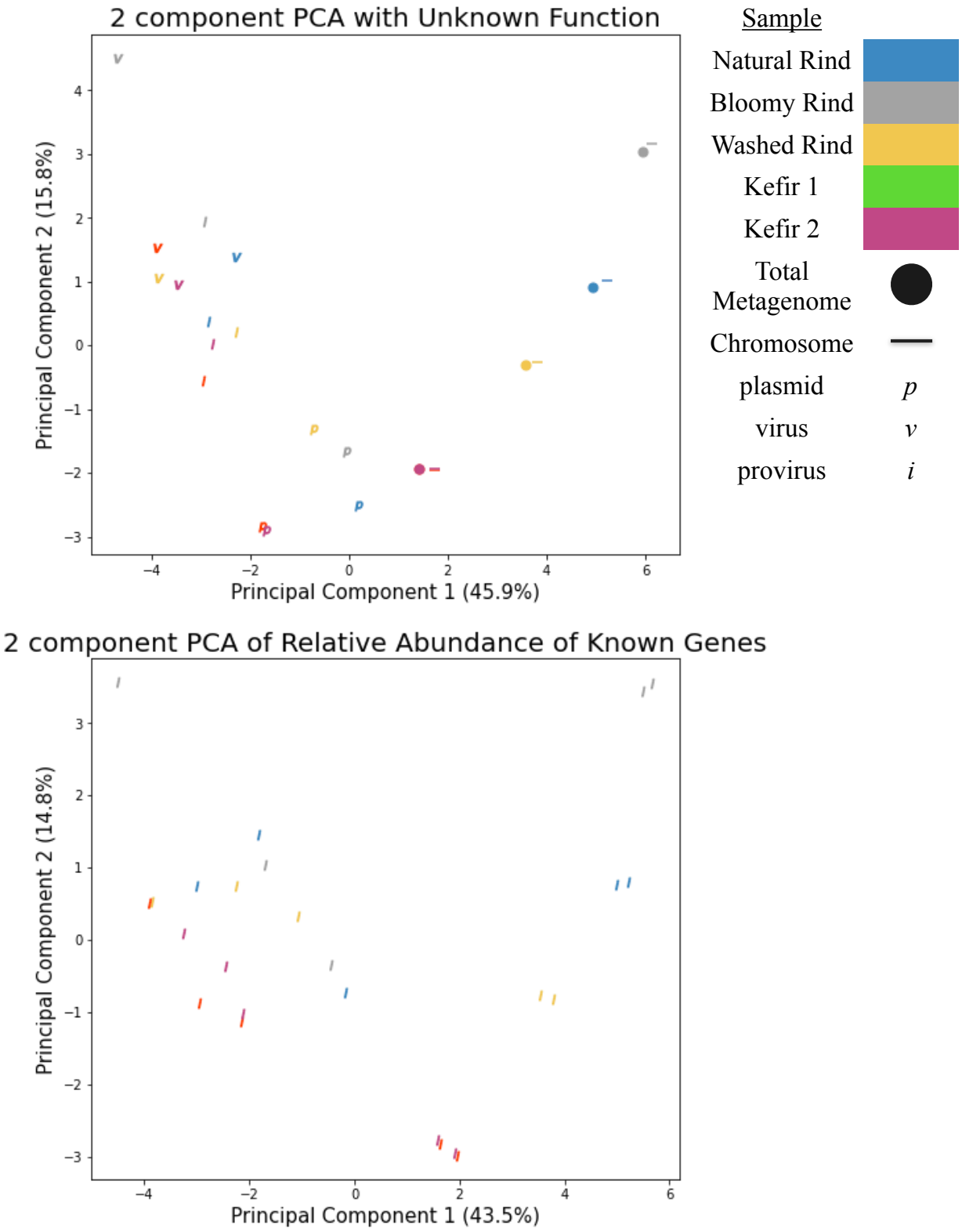
Supplemental Figure 2.6. vcontact2 colored by (a) sample and (b) phyla. Not shown: Psychrobacter prophage from the washed rind, possible Streptococcal lytic phage from kefir 2. Natural Rind=bayley; Bloomy Rind=harbison; Washed Rind=oma; Kefir 1=kefir_rbk; Kefir 2=kefir_rad. Available in PDF.



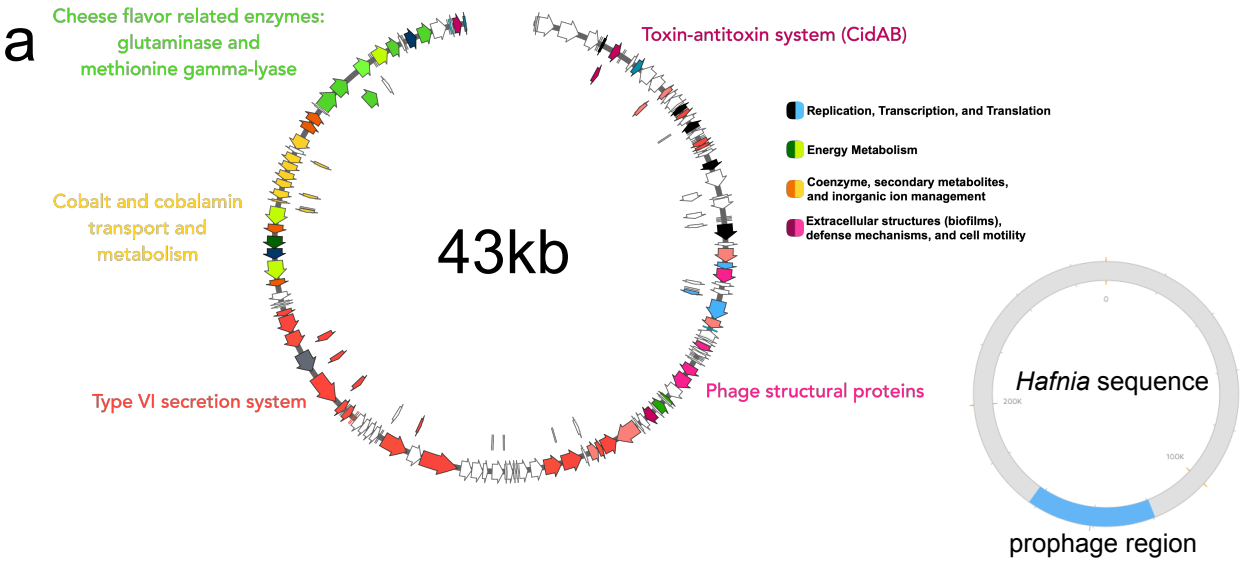
Supplemental Figure 2.7 The changes in protein lengths their counts before and after protein aware frameshift correction with DIAMOND. Page 1. Effects of frameshift correction with respect to (top) total protein length which increased, except for in the washed rind; (middle) number of proteins, which decreased, except for in the washed rind; and (bottom) the ratio of these two, indicating possibly longer and more complete protein lengths.

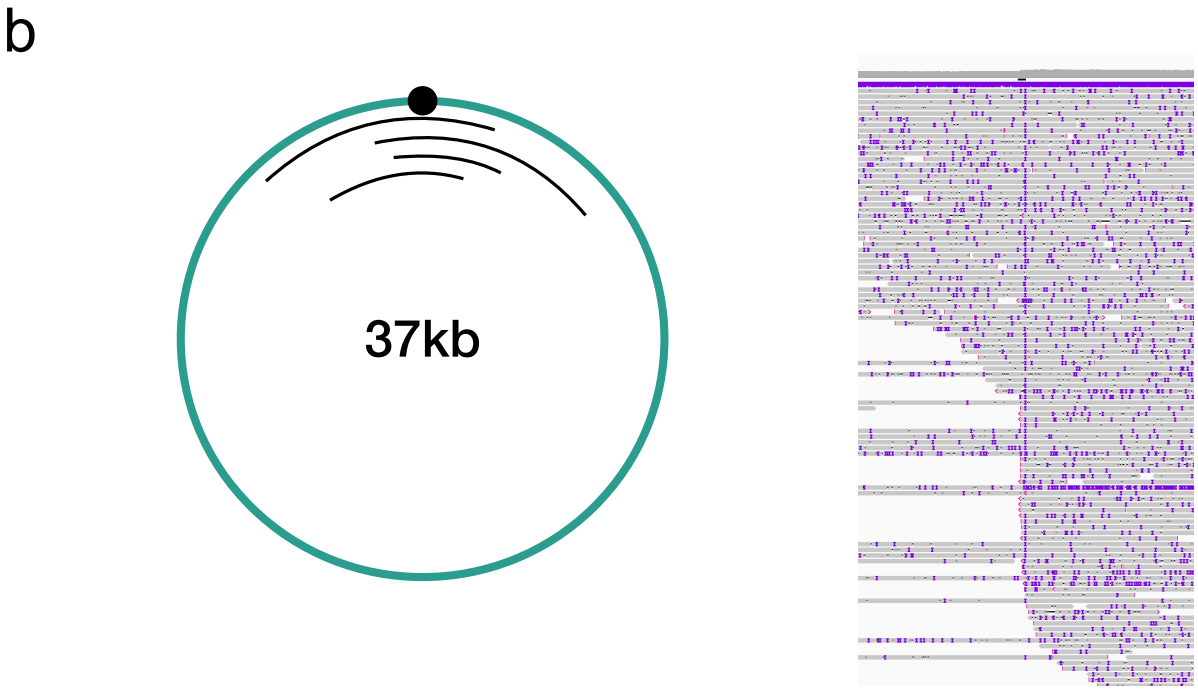


Supplemental Figure 2.7 continued The changes in protein lengths their counts before and after protein aware frameshift correction with DIAMOND. This page is the box plot of the protein lengths of each sample.



Supplemental Figure 2.8. Principle component analysis of relative amounts of COG categories.





Supplemental Figure 2.9. Examples of prophages: a) Bloomy rind *Hafnia* prophage encodes for genes potentially adaptive to the cheese rind niche environment; b) *lactobacilli* prophage example of long reads mapping across an identified *att* site, and on the right showing long reads mapping across region, as well as reads that map to both viral and host genome.

Acknowledgements

Chapter 2, in full, is a preprint of the material as it was prepared for to mSystems, 2022, Dinh, Cong B, Dutton, Rachel J. The dissertation author was the primary investigator and author of this paper. Nicholls, Samuel M, Loman, Nick J supported this work through data collection and discussions. Gary Heussler, Emily Pierce, and Manon Morin supported this work through discussions and teaching of a course that contributed data.

CHAPTER 3 in vitro Microbial Community Mimicking Washed Rind Cheese

Cong B. Dinh, Christina C. Saak, Lucas Patel, Rachel J. Dutton

Abstract

Models of moderately complex microbial communities permit the study of interspecific interactions in a more natural environment. Recently, fermented foods have been used as a starting point for these microbiomes, especially with plant-based and dairy-based systems (Kim and Chun 2005; Wolfe et al. 2014; Marsh et al. 2013; May et al. 2019; Walsh et al. 2016; Landis et al. 2021; Blasche et al. 2021). Often, these communities are still relatively simple, with less than 5 community members, which can make it difficult to study the interactions between closely related organisms, such as horizontal gene transfer. Here, we design a 16 member community that attempts to probe the assembly and evolution of these community members with various levels of relatedness. Additionally, we mimic the washed cheese rind biofilm to induce significant amounts of physical interactions at the onset of community formation. We examine the differences in evolution in this community and one without a key fungal member, and one without a major phyla. Finally, we attempt to apply a combination of proximity ligation and long read sequencing to these communities to look for evidence of plasmid transfer within a genus.

Introduction

Much of our knowledge of biology comes from the study of isolated species in well controlled environments. This is in contrast with natural ecology in which many different co-occur in dynamic conditions. Over the past few decades, synthetic biologists and

microbiologists have built up a strong foundation of two species systems, allowing for additional attempts at increased complexity. Indeed, these advancements were recently reviewed in (Deter and Lu 2022); in short, they distill the interspecific interactions as cooperation, competition, and predation, layered with additional community members or certain spatial organization. In particular, higher order interactions, which only occur in complex communities and cannot be described by pairwise interactions, may complicate the analysis of any synthetic community of greater than three members.

Previous work in the use of cheese model systems has focused on low-complexity communities, ranging from 3-7 species (Cosetta and Wolfe 2020; Wolfe and Dutton 2015; Morin et al. 2018). These studies have enabled a mechanistic understanding of certain interactions between species, such as cross-feeding of amino acids and competition for iron, and uncovered the importance of fungi in driving microbial interactions (Pierce et al. 2021; Cosetta et al. 2020; Bodinaku et al. 2019). However, due to a lack of intra-genus and intra-species diversity within these previous models, a variety of microbial interactions may not be represented. Washed rind cheeses, which have not previously been reconstructed in vitro, represent a potential system for studying a wider diversity of microbial interactions. These cheeses undergo a repeated washing process with a brine solution during the early stages of their aging. This washing creates a cheese surface that is high in salinity and moisture. Further, the physical nature of the washing process disrupts spatial structure of the community, potentially inducing greater opportunities for interactions. The combination of the salinity, moisture and physical disruption of the community leads to the formation of a unique

microbiome not found on other cheeses containing a diverse mix of bacteria, fungi (primarly yeast), and associated viruses.

Here, we attempt to establish a new in vitro model microbiome with higher complexity than previous efforts based on a washed rind cheese. The target microbiome contains species from diverse Phlya, which will allow the examination of interactions across diverse microbial groups, but also contains a cluster of species from the same genus, which will allow an examination of interactions within much more closely-related species. In this work, we established the conditions for reconstructing the specific environment of this community in an in vitro cheese system, showing that the community undergoes a reproducible succession following inoculation. We then use our ability to control the species inoculation profile to examine the effect of removing major taxonomic groups from the community. We combine culture-based approaches with shotgun metagenomic sequencing to evaluate changes in community composition under the different conditions, and to determine whether mutations over the course of community formation are enriched in certain species and genetic regions.

Results

Design of model microbiome

This model microbiome is based on a culture collection that was isolated from a specific washed rind cheese, all of which have genome sequences available. From this collection, we selected 16 species that would represent both the breadth and depth of diversity found in a typical washed rind cheese microbiome. Previous metagenomic studies of this

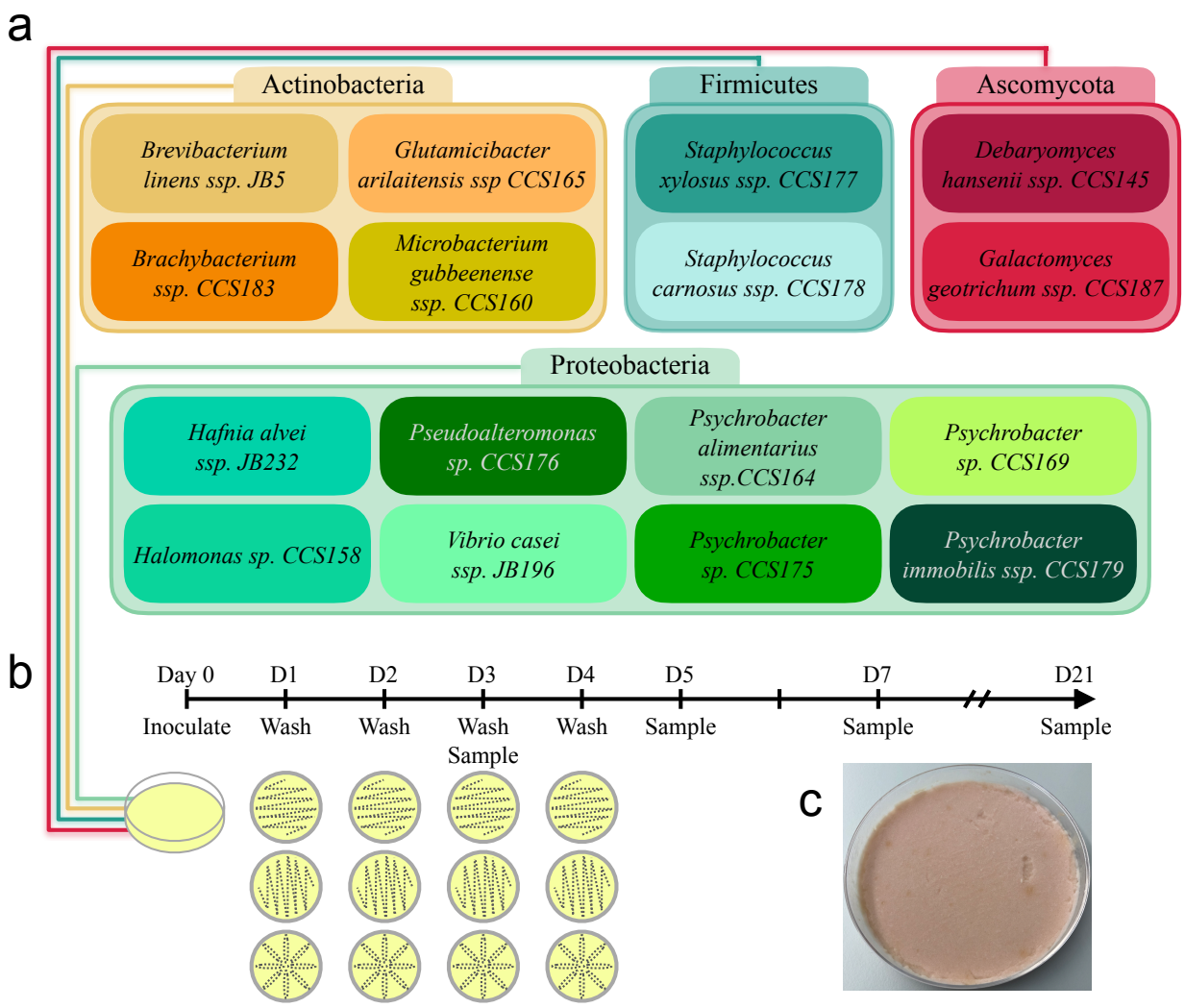


Figure 3.1.1. Overview of in vitro washed cheese rind community. (a) Full community members, divided and colored by Phyla. (b) Timeline of in vitro microbial community model as well as washing pattern. All three steps were done on each plate with a sterile cotton swab (c) example plate of full community at day 21 from the partially destructive sampling.

cheese revealed that multiple species and strains of the genus *Psychrobacter* co-occured, which could provide a unique opportunity to examine intra-species interactions (Supplemental Table 3.1.2). In particular, given that the horizontal transfer of genetic material such as plasmids is more likely to occur between more closely related species, we evaluated the genomic similarity and plasmid content of these *Psychrobacter* species.

To compare isolates from within the *Psychrobacter* genus, we aligned genomes with SPINE (Ozer et al. 2014, Supplemental Figure 3.1.1, Supplemental Table 3.1.2). Four *Psychrobacter* (P.) are further split into two groups: Group 1, containing *P. sp CCS164* and *P. alimentarius ssp. CCS175*, shares 83.8% of their chromosomal region, and *P. sp. CCS164* contains 4 plasmid contigs and *P. alimentarius ssp. CCS175* contains 5 plasmid contigs, as well as a lysogenic and lytic phage; Group 2, containing *P. cibarius ssp. CCS169* and *P. fozii ssp. CCS179*, shares 87.2% of their chromosomes, and *P. cibarius ssp. CCS169* contains 5 plasmid contigs while *P. fozii ssp. CCS179* has a lysogen and five plasmid contigs, two of which are nearly identical. The plasmids themselves are 36.6-71.5% unique, with those from *P. cibarius* sharing 16.5% with plasmids from *P. alimentarius* and 11.8% with plasmids from *P. fozii*. About 11kbp (11% of the plasmid contig lengths of *P. Sp. CCS164*), are shared among all *Psychrobacter* MGEs. The lytic phage and lysogens are largely unique.

The community also contains four other Proteobacteria: *Pseudoalteromonas*, *Halomonas*, *Vibrio casei*, and *Hafnia alvei*; two Firmicutes: *Staphylococcus xylosus* and *carnosus*; four Actinobacteria: *Brevibacterium linens*, *Brachybacterium aurantiacum*, *Glutamicibacter arilaitensis*, and *Microbacterium gubbeenense*; and two Ascomycota: *Debaryomyces hansenii* and *Galactomyces geotrichum*. All community members were sequenced as isolates and reassembled; all bacterial genomes were 100% complete based on a list of 139 bacterial single copy core genes, and less than 3% redundant except for the *Glutamicibacter* genome, which was 98.59% complete.

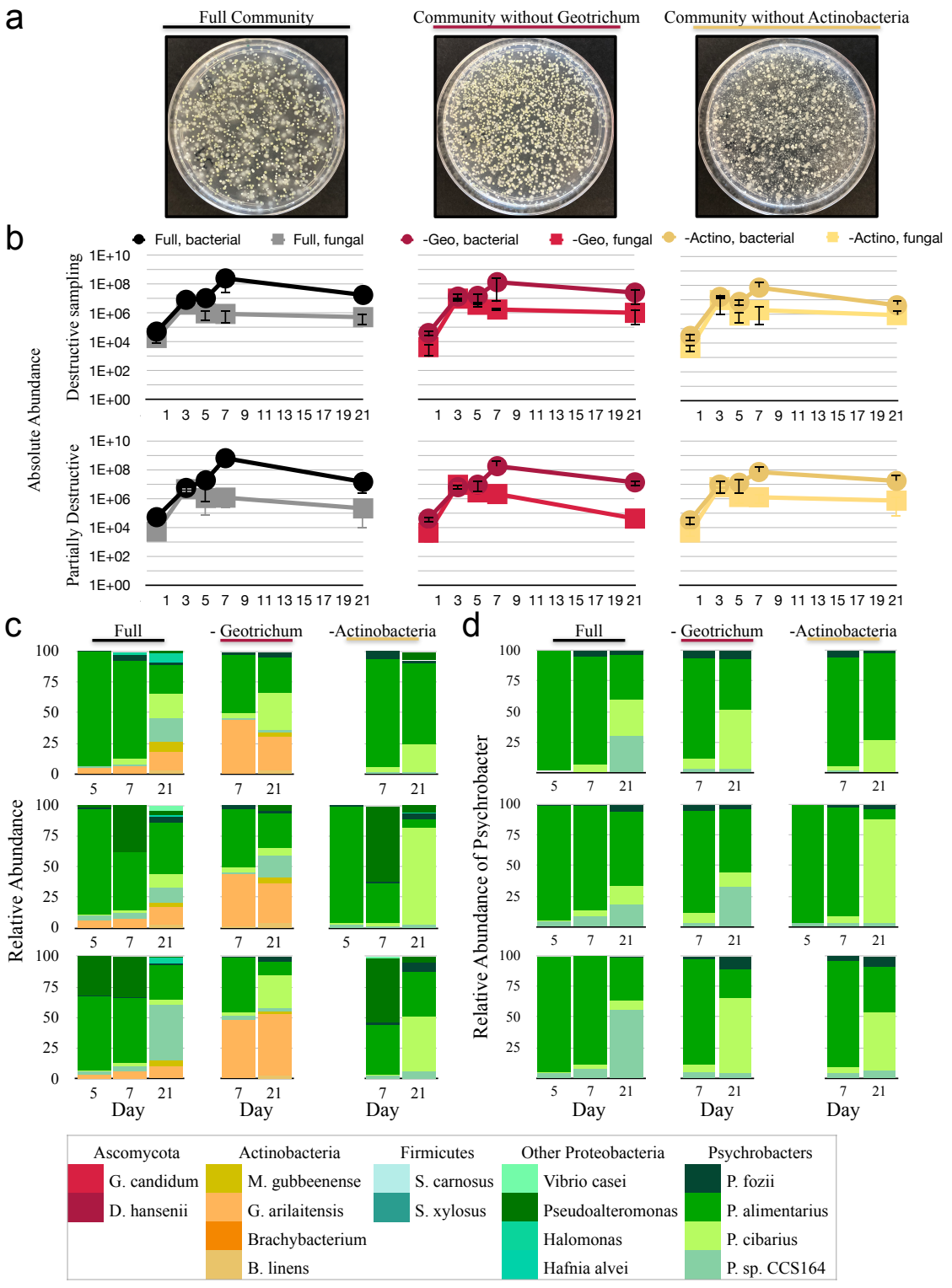


Figure 3.1.2. CFU counts and relative abundance of in vitro washed rind communities. (a) shows 10,000X diluted communities from Day 3 samples; the community without Geotrichum appropriately appears without the fuzzy white colonies and with additional yellow-orange toned Actinobacteria. (b) CFU counts, n=3 except for Day 0 Full community, n=2. (c-d) are relative abundance plots based on short read mapping to reference genomes.

To construct an in vitro washed rind microbiome, 100,000 CFUs of each member was added to the surface of a 10% Cheese Curd Agar petri dish (Cosetta and Wolfe 2020), then evenly distributed across the plate. The surface of the plate was washed with 20% NaCl brine every 24 hours in the first 96 hours for a total of four washes (Figure 3.1.1b, Supplemental Figure 3.1.2). The washing process was designed to mimic the production of washed rind cheeses, and was carried out by adding 1 mL of brine to the surface of the plate, then passing a sterile swab over the surface three times, in a vertical rastoring, then horizontal rastoring, and finally in a rosette pattern to gently mix the surface community into the brine. Communities were incubated in the dark at 15oC in a high humidity environment. Samples were taken by scraping communities from the surface of the agar plate with PBS+0.05%Tween and gently homogenized. In destructive sampling, the entire plates' biomass was split for glycerol stocks, CFU counting, proximity ligation, and genomic DNA extraction. All communities were also studied with partially destructive sampling, in which half of the sample was used to reinoculate the same plate. The washing steps and length of maturation mimic the commercial process on an accelerated time-scale that correlates with in vitro communities reaching maturity and stability at about half the time of commercial cheeses (Wolfe and Dutton 2015). All community experiments were performed in triplicate. To assess the reproducibility of the in vitro washed cheese rind model, CFU counts of each sampling day were done for both the bacterial and fungal members (Figure 1D). In addition, a portion of each sample was used for DNA extraction and metagenomic sequencing. Sequencing reads were then mapped back to the reference genomes to show relative abundance (Figure 3.1.2a).

In the full community, there was an initial even number of bacterial and fungal community members until day 3, after which the fungal population began to dwindle, largely from the loss of *Debaryomyces*. Due to the daily washing procedures, the communities are disturbed each day through day 4. The peak of bacterial abundance occurs before the final sampling at day 21. Overall, these communities are fairly reproducible with relatively small standard deviations at the Phyla level. In the partially destructive samples, the relative amount of Actinobacteria slowly increases over the course of the incubation period (Figure 3.1.2a). When examining the *Psychrobacter* species, all four species persist in each replicate, with *P. alimentarius ssp. CCS175* often the most abundant. Though *P. sp. CCS164* is closely related to this abundant species, it begins to overtake *CCS175* as the community ages. The in situ time series of the washed rind cheese upon which this model is based shows similar progression (Saak et al, 2023), with fungal members at similar abundance to bacterial species for the first two weeks, then being less abundant for the next 10 weeks. At the genus level, the in situ community also does not have significant amounts of Firmicutes at later weeks while Actinobacterial genera, especially *Brevibacterium* and *Glutamicibacter*, increase in abundance over time.

Previous studies revealed a positive correlation between *Geotrichum* and Proteobacterial species, and a negative correlation between *Geotrichum* and Actinobacterial species based on co-occurance patterns in a sequencing based survey of cheese rind microbiomes (Wolfe et al, 2014). In addition, pairwise experimental data showed stimulatory effects of *Geotrichum* on Proteobacterial growth, and inhibitory effects on Actinobacterial growth (Wolfe et al, 2014). We took advantage of the fact that this model contains all three of

these members (*Geotrichum*, Proteobacteria, and Actinobacteria) to examine whether these patterns hold in a community context. To do this, we reconstructed a community lacking *Geotrichum* and evaluated whether this resulted in any differences in the resulting community composition (Figure 3.1.3b,c in red, Supplemental Figures 3.1.3 and 3.1.4). In all replicates, the removal of *Geotrichum* favored the relative growth of the Actinobacteria over the Proteobacteria, even though overall bacterial absolute abundance was very similar between these two samples (Figure 3.1.1d). Absolute abundance based on read counts reveals that the decrease in Proteobacterial abundance is largely due to the poor growth of *Pseudoalteromonas*. In contrast, almost all of the Actinobacterial species reach higher absolute abundances, and do so by earlier timepoints, as compared to the full community (Supplemental Figure 3.1.3). Overall, these results are consistent with *Geotrichum* inhibiting Actinobacteria and stimulating Proteobacteria. In one of the replicates in the community without *Geotrichum*, *P. sp. CCS164* has increased in abundance in the last two weeks of incubation, similar to the complete community; the other two samples show P. *cibarius ssp. CCS169* as the dominant *Psychrobacter* strain.

We next tested the effect of removing all Actinobacterial members from the community. Across all conditions, *P. alimentarius ssp. CCS175* is dominant in the early stages of the community. Similar to the other conditions, *Psychrobacter* species dominate the communities, but are 93-95% of the community in the community without Actinobacteria. There is also a different pattern of which *Psychrobacter* is dominant. *P. sp. CCS164* does not increase in relative abundance, remaining below 5% of *Psychrobacter* in all replicates. Instead

P. cibarius ssp. CCS169 competes with *P. alimentarius ssp. CCS175* for the most abundant *Psychrobacter*.

Evolution in model microbial community

In addition to microbial dynamics and relative abundance, we examine the evolution of the 16 genomes between these different conditions over the course of the 21 day experiment. To do this, reads from sampled timepoints were mapped back to reference genomes of each isolate. Single nucleotide variants (SNVs) from discrepancies between the reference genomes and mapped reads can show adaptations to the different community pressures. We use a conservative method to only report variations that occur in a ratio greater than 5% + (1/3)coverage1/3-1.45, decreasing the threshold dynamically with the coverage of the contigs (Good et al. 2017). This conservative method will filter for confident variations, and all variations are reported in Supplementary Table 3.1.9. 10.4 million SNVs were initially reported, and these were filtered to 519,178 million SNVs across 140,815 nucleotide positions in 18,397 genes. Since we used a different reference strain's genome for this mapping, we removed any calls within the *Geotrichum* and *Debaryomyces* genomes. The Firmicutes and *Brachybacterium arcticum* had few mutations called, likely due to simply being at low abundance (<1%). Supplementary Table 3.1.7 shows the number of mutations in each of the samples sequenced for each community and time point, with Supplementary Table 3.1.8 listing the individual nucleotide positions and relative proportions of each nucleotide. Anvi'o plots cluster SNVs by co-occurance in different conditions and timepoints and present the relative occurrence of SNVs detected. In the full community and the community without

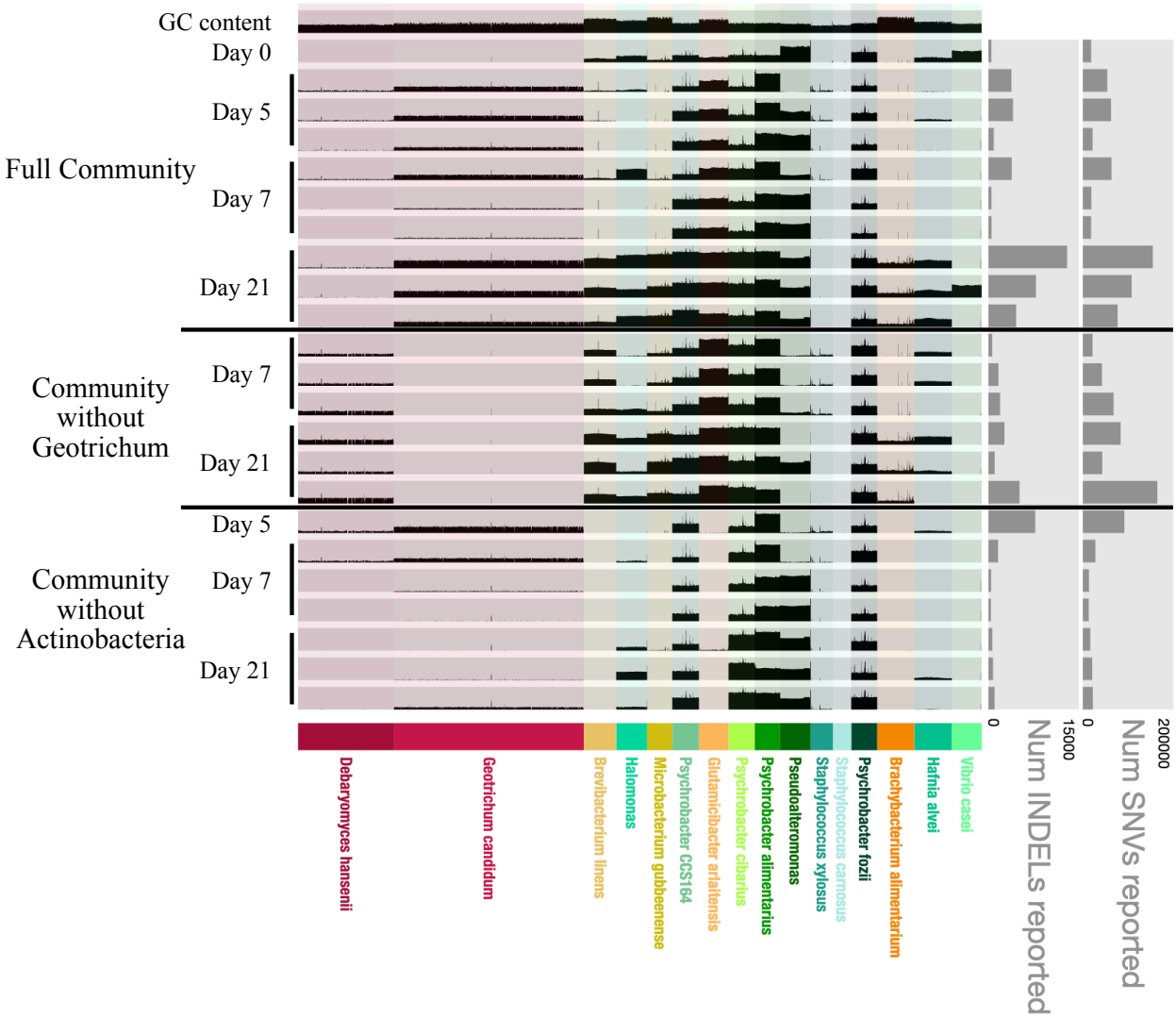


Figure 3.1.3. Overview of the communities and SNP analysis. Mapping of individual samples to the reference genomes in an anvi'o plot, with each lane showing the coverage of the genome. On the right, the additional graphs show the number of insertion/deletion mutations and single nucleotide variants.

Geotrichum, the total number of Actinobacterial mutations were similar, with unique mutations between replicates in *Glutamicibacter* and *Brachybacterium* in both conditions, and *Brevibacterium* and *Microbacterium* in the community without *Geotrichum*. The *Psychrobacter* dominated community composition in all conditions, and had similar number of SNV in the full community and the community without *Geotrichum*; the total number of

mutations within each phyla was least in the community without Actinobacteria (Figure 3.1.3a, Supplemental Table 3.1.7), suggesting that they contribute to additional biotic factors that encourage mutations or evolution. Across replicates, there were often unique mutations that persist within each replicate; for example, in the yeast *Debaryomyces hansenii*, replicate 1 in the full community, and replicate 1 and 3 in the community without *Geotrichum* shows the same mutations, though the frequency decreases over time in the full community.

Proximity Ligation and Long Read sequencing

Chromatin conformation capture techniques have been adapted for a wide range of applications, including to elucidate the 3-dimensional structure of chromosomes by highlighting regions that are physically close to each other. By applying this to a microbial genome and allowing significant time for ligation, we may generate concatemers of DNA within a microbial cell, potentially binding together chromosomal sequences with extrachromosomal elements. Pore-C is the use of long read sequencing to these preserved physical interactions, allowing for multiple contacts to be interrogated on one read. Within a community, we hoped to generate sequences between 2-5 kb in length using as few endonucleases as possible, which can be difficult within a community containing fungi and bacteria from three phyla. As we were most interested in possible gene transfer within the Psychrobacter, we also analyzed the plasmid contigs for cut sites as well. In silico digestion of the metagenome by EcoRI and BglII generated 40,035 fragments with an N50 of 4 kbp, average of 2.2 kbp, and a quarter of the fragments were between two and five kbp

(Supplemental Figure 3.1.8). These values assume an even distribution of each community member and the specific counts are simply of a restriction of a single copy of each genome.

We were able to sequence two of our samples: the full community at day 3 and day 21. There are 187,568 EcoRI sites and 111,730 BglII sites on Day 3 resulting in 209,866 reads from 41,165 concatemers after in silico restriction; and 2,291,016 EcoRI sites and 1,667,675 BglII sites on Day 21 resulting in 9,844,474 reads from. There are 102,235 fragments with both restriction sites on Day 3 and 882,151 on Day 21. Compared the *in silico* digestion of the metagenome which assumed an even distribution, the read N50 was 2,370 on Day 3 and 2,239 on Day 21 (Supplemental Figure 3.1.9).

Discussion

Model microbial communities allow for precise control of the input variables to better understand how different biotic and abiotic factors change how the different community members interact and grow. In particular, a community setting is required to study horizontal gene transfer, the transfer of genes between organisms rather than from parent to child or from mother to daughter cells. The rate of horizontal gene transfer and its impact on evolution has been debated at length, but it has played a major role in the spread of antibiotic resistances, especially in nosocomial infections. The widespread use of these life saving medications for human infections as well as in agriculture creates a strong fitness benefit to maintaining genomic islands that increase resistance to commonly used antibiotics, especially those derived from natural products. Genes associated with resistance to the natural products already exist in the environment from the persistent push for more resources and genetic replication. In some cases, the genes themselves are selfish and promote their own mobility in

the cases of viruses and genetic elements with toxic systems that kill hosts that lose the accessory element.

Methods

Strain source information

Source information for strains provided in Supplementary Table 3.1.1. Shortly, the majority of strains were isolated from washed rind cheeses and sequenced. Brevibacterium linens sp. JB5, Vibrio casei sp. JB196, and Hafnia alvei sp. JB232 were previously isolated and had their genomes assembled from a survey of cheeses (Wolfe et al. 2014).

Sequence analysis

Each genome was sequenced with illumina short reads and assembled independently. Assemblies that resulted in more than 20 contigs were sequenced again using a flongle (Oxford Nanopore Technologies) and reassembled with hybridSPAdes (Antipov et al. 2016). Genomes from the same genus were further analyzed by SPINE to separate core and accessory genomes for downstream analysis (Ozer et al. 2014). Further, these were analyzed by mob-suite and viralVerify for mobilizable plasmids and viral sequences (Robertson and Nash 2018; Antipov et al. 2020). Finally, Psychrobacter genomes and their plasmids, as well as the entire set of genomes, were analyzed for restriction sites for a minimal set of endonucleases that would result in many 2-10kbase sequence fragments.

in vitro washed rind community

Stock strains were grown in lysogeny broth and frozen in 20% glycerol. Each strain was inoculated with an initial even distribution, either 100,000 bacterial cells or fungal spores, on 10% Cheese Curd Agar (Cosetta and Wolfe 2020). Plates were incubated in the dark at 15C in high humidity environment. At 24, 48, 72, and 96 hours after inoculation, plates were scrubbed with a 20%wt NaCl brine solution using sterilized cotton swabs in a horizontal and vertical rastering pattern, followed by a rosette. Brine solution was also reserved. On days 3, 5, 7, and 21, microbial communities were collected using cell scrapers into 1000µL of PBS+0.05% Tween to assist in biomass recovery; for the day 3 sample, biomass collection was done before the brine wash. The samples were split for various analyses: spot plating, glycerol store, gDNA extraction for metagenomic short read sequencing, and proximity ligation for nanopore sequencing. In one set of samples, called nondestructive, we replated half of the scraped biomass back onto the dish. Spot plating was done on PCAMS, either with 50µg/L chloramphenicol or 100µg/L clyclohexamide + 21.6µg/L natamycin, with counting done 48 hours later. gDNA was collected using phenol-chloroform extraction and purified with 5 additional ethanol washes to remove residual phenol and chloroform. For short read sequencing, library preparation and metagenomic sequencing were performed by Novogene, Sacramento, CA.

Genomic DNA extraction

Add 200µL of glycerol reserve to 15mL conical tube with 3mL of Tris-Lysis Buffer (10mM Tris-Cl, 1% SDS) and incubate at 37 C for 1 hour, agitating every 15 minutes. Add 28

units of proteinase K (NED) and incubate at 50C for 1 hour. To this mixture, add an equal amount of phenol-chloroform-isoamyl alcohol (25:24:1) buffered at pH 8.0 with 10mM Tris-Cl, 1mM EDTA, and vortex for 2 seconds before rotating for 10 minutes at room temperature. Separate the phases with centrifugation at 3900 rpm, at 4C, for 15 minutes. Transfer aqueous phase to fresh 15 mL tube and add an equal volume of fresh buffered phenol-chloroform-isoamyl alcohol, and repeat vortex, rotation, centrifugation, and aqueous layer transfer to a new 5 mL tube or split into two 5 mL tubes. Add 1 volume of 4C isopropyl alcohol and 0.1 volumes of 3M sodium acetate and centrifuge at 17,000 G for 3 minutes. Discard the supernatant and wash the pellet with 70% ethanol 4 times, centrifuging to pellet the DNA between each wash. Dry the pellet and resuspend in 200 µL 10 mM Tris-Cl pH 8.5 at room temperature overnight. Samples with concentrations greater than 2.5 ng/µL were submitted to sequencing at Novogene and sequenced on NovoSeq 6000.

Proximity ligation and long read sequencing

Reserved biomass, approximately 750µL, was diluted in 10mL of PBS for two washes to remove environmental DNA. All centrifuge steps were done as follows: 4C, 3900g, 5 minutes. Cells were then pelleted and resuspended in 1mL PBS+1%formaldehyde, mixed with gentle inversions. After 10 minutes of crosslinking, the formaldehyde was quenched with 527µL of 2.5M glycine at room temp for 5 minutes, then on ice for 10 minutes with light agitation every three minutes. Cells were again pelleted and washed with 10mL 4C PBS, and resuspended in 1mL PBS.

All centrifuge steps during digestion and ligation were done at 4C, 300g for 10 minutes. At 4C, add 550µL of protease inhibitor cocktail (10mM Tris-HCl, pH8.0, 10mM NaCl, and 0.2% IGEPAL CA-630, 10% protease inhibitor). Pellet and remove supernatant. Resuspend in 200µL of 4C 1.5X digestion reaction buffer. Pellet and remove supernatant. Resuspend in 300µL of the digestion reaction buffer, and add SDS to the final concentration of 0.1% SDS. Incubate on a shaker at 65C and 300rpm for 10 minutes, then put on ice. Add Triton X-100 to a final concentration of 1%, incubate on ice for 10 minutes. Digest with EcoRI at 20U/µL and BglII at 10U/µL, invert to mix. Incubate at 37C for 18 hours. Mixing at 1000 RPM for 30 seconds every 15 minutes. Assess digestion efficacy ***. Add additional ligation buffer onto the suspension and allow to incubate at 16C for 6 hours, mixing at 1000 RPM for 30 seconds every 15 minutes. Finally, degrade the protein complexes with proteinase K at a final concentration of 1µg/µL. Incubate at 56C for 18 hours, mixing at 1000 RPM for 30 seconds every 15 minutes. Chill at 4C.

Transfer ligated DNA to 5mL centrifuge tube, and rinse the original tube with 200µL nuclease free water and combine in the same 5mL tube. Add an equal volume of phenol:chloroform:isoamyl alcohol in 10mM TrisHCl, pH 8.0, 1mM EDTA solution and gently invert for 5 minutes. These centrifuge steps are done at 4C, 16,000g for 5 minutes. Centrifuge and incubate on ice for 2 minutes, until organic phage is cloudy with degraded proteins and the interphase layer is thin. Transfer aqueous phase and split into two aliquots. Add 0.2 volumes of 5M NaCl and 0.1 volumes of 3M sodium acetate, pH 5.5. Add 3 volumes of 100% ethanol and gently invert. Precipitate overnight at -20C. Pellet the nucleotides, then wash with 4mL 80% ethanol, then 4mL 70% ethanol. Remove the supernatant and allow to

dry. Resuspend in 75μL TE buffer and combine two tubes. Quantify with Qubit at 10X dilution. These were then library prepped with SQK-LSK109 (ND, C1, D3 and D21) or -LSK110 (all other samples) and sequenced on R10 minION flow cells, Oxford Nanopore technologies.

Computational Analysis

Short reads

Genomic DNA was then collected from each replicate from each condition, and those with enough sample - Days 7 and 21 in from all conditions and Day 5 from the full community and in one sample of the community without Actinobacteria - were sequenced with Illumina technology, with at least 50 million paired 151bp reads sequenced from each sample. These were then mapped back to the reference genomes to show relative abundance (Figure 3.1.2a). The 1,606,847,570 reads were mapped against the reference genomes, both as whole genomes and modified for core and accessory regions. SNP analysis was performed on nondestructive samples, with each biomass collection acting as a passage.

Long Reads

Long reads were split on exact matches to restriction sites and analyzed with a modified version of pore-c tools (v3.0 Oxford Nanopore Technologies, bioRXiv), and MC4C (Vermeulen et al. 2020).

Metagenome analysis

Anvi'o v7 was used to visualize the metagenome and the coverage from short reads and cut-long reads.

This chapter's work was integrated with the following paper:

Longitudinal, multi-platform metagenomics yields a high-quality genomic catalog and guides an *in vitro* model for cheese communities

Christina C. Saak[1], Emily C. Pierce[1], Cong B. Dinh[1], Daniel Portik[2], Richard Hall[2], Meredith Ashby[2], Rachel J. Dutton[1]

[1] Division of Biological Sciences, Section of Molecular Biology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[2] Pacific Biosciences, 1305 O'Brien Dr, Menlo Park, CA 94025, USA

Abstract

Microbiomes are intricately intertwined with human health, geochemical cycles and food production. While many microbiomes of interest are highly complex and experimentally intractable, cheese rind microbiomes have proven powerful model systems for the study of microbial interactions. To provide a more comprehensive view of the genomic potential and temporal dynamics of cheese rind communities, we combine longitudinal, multi-platform metagenomics of three ripening washed-rind cheeses with whole genome sequencing of community isolates. Sequencing-based approaches revealed a highly reproducible microbial succession in each cheese, co-existence of closely related *Psychrobacter* species, and enabled the prediction of plasmid and phage diversity and their host associations. Combined with culture-based approaches, we established a genomic catalog and a paired 16-member *in vitro* washed rind cheese system. The combination of multi-platform metagenomic time-series data

and an *in vitro* model provides a rich resource for further investigation of cheese rind microbiomes both computationally and experimentally.

Importance

Metagenome sequencing can provide great insights into microbiome composition and function and help researchers develop testable hypotheses. Model microbiomes, such as those composed of cheese rind bacteria and fungi, then allow the testing of these hypotheses in a controlled manner. Here, we first generate an extensive longitudinal metagenomic dataset. This dataset reveals successional dynamics, yields a phyla-spanning bacterial genomic catalog, associates mobile genetic elements with their hosts and provides insights into functional enrichment of *Psychrobacter* in the cheese environment. Next, we show that members of the washed-rind cheese microbiome lend themselves to *in vitro* community reconstruction. This paired metagenomic data and *in vitro* system can thus be used as a platform for developing testable hypotheses for the dynamics within and functions associated with cheese rind microbiomes.

Introduction

Microbiomes play crucial roles in human health(Gilbert et al. 2018), geochemical cycles(Falkowski, Fenchel, and Delong 2008) and food production(Sieuwerts et al. 2008). While the characterization of community composition of diverse microbiomes has come a long way, our mechanistic understanding of community functioning, and thus our ability to predict and manipulate it, lags behind(Pierce and Dutton 2022). Given the complexity and

experimental intractability of many microbiomes of interest, model microbiomes consisting of a manageable number of community members that are experimentally tractable can help facilitate the generation and testing of hypotheses about microbiome function. Cheese rind communities have already provided valuable insights into the biology of microbiomes and microbial interactions within such as cross-feeding of amino acids and competition for iron, and they have helped uncover the importance of fungi in driving microbial interactions (Wolfe et al. 2014; Morin, Pierce, and Dutton 2018; Niccum et al. 2020; Kastman et al. 2016; Zhang et al. 2018; Kamelamela et al. 2018; Pierce et al. 2021; Cosetta et al. 2020; Bodinaku et al. 2019). The cheese rind model communities currently available for in vitro studies are based on natural and bloomy rind communities and represent low-complexity microbiomes ranging from 3-7 species (Cosetta and Wolfe 2020; Wolfe and Dutton 2015; Morin, Pierce, and Dutton 2018; Wolfe et al. 2014).

In contrast to natural and bloomy rind cheeses, washed-rind cheeses are produced by regular washing (or smearing) with a brine solution(Mounier et al. 2017). As such, the microbial communities on the surface of washed-rind cheeses experience homogenization throughout its development, which may facilitate intermicrobial interactions or evolutionary processes such as horizontal gene transfer. To date, several studies have examined the community composition of washed-rind cheeses using culture-dependent and culture-independent techniques(Delcenserie et al. 2014; Kothe et al. 2021; Irlinger and Monnet 2021; Quijada et al. 2020, 2018; Schmitz-Esser et al. 2018; Bokulich and Mills 2013; Wolfe et al. 2014; Bertuzzi et al. 2018; Quigley et al. 2012; Schornsteiner et al. 2014). These studies have shown, for example, that bacteria often outnumber fungi by orders of magnitude(Irlinger and

Monnet 2021; Quijada et al. 2020). Among the bacterial community members, Actinobacteria such as *Brevibacterium, Corynebacterium* and *Glutamicibacter*, and Proteobacteria, such as *Psychrobacter* and *Halomonas,* are usually detected in these communities(Quijada et al. 2020, 2018; Irlinger and Monnet 2021; Schmitz-Esser et al. 2018; Delcenserie et al. 2014; Kothe et al. 2021). Last, it has been shown that the communities associated with these cheeses show reproducible community succession(Irlinger and Monnet 2021). Emerging metagenomic techniques have the potential to enable a deeper characterization of the species- and strain-level diversity, functional potential, eco-evolutionary dynamics, and mobile genetic elements within these communities. In this study, we combine several metagenomics techniques (amplicon, short-read and long-read shotgun sequencing and metaHi-C) with a longitudinal dataset of three washed-rind cheese communities that were collected over the course of cheese ripening. This data provided insights into the reproducible successional trajectories of the studied washed-rind cheeses, provided a catalog of genomes of community members that can be used as references for future studies, identified plasmids and phages and associated them with their host genomes, and provided insights into the biology of these communities. For example, we investigated the striking diversity of *Psychrobacter* genomes recovered from the communities and present a functional enrichment study that highlights enrichment of genes involved in type six secretion as well as siderophore acquisition, two traits that are of high value in the densely populated, iron-limited environment of the cheese rind.Finally, we used the genomic catalog to establish a representative culture collection of washed rind community members and reconstituted an *in vitro* model communities based on the washed-rind cheese microbiome. This model microbiome contains 16 members from several microbial phyla,

including several *Psychrobacter* representatives, allowing the examination of interactions across diverse microbial groups and representing the most complex cheese rind-based model microbiome to date. We show that this *in vitro* community undergoes reproducible succession and, by removing certain taxonomic groups, we start to investigate the interaction dynamics in this community.

Results

Longitudinal sampling of three washed-rind cheeses from the same facility. Rind samples of three different types of washed-rind cheeses (Cheese A, B, and C) were collected from triplicate batches (each batch made approximately 1 week apart) at six timepoints throughout ripening (Figure 3.2.1A). Cheese types A, B, and C differ with respect to their milk source, use of pasteurized vs raw milk (pasteurized milk was used for cheeses A and B, raw milk is used for cheese C), production location, and production methods. However, during the aging process of each of the cheeses, similar aging practices were used, such as repeated washing with brine solution, and all were aged in the same facility. For all samples, DNA was extracted and 16S and ITS amplicon sequencing was performed. Subsequently, samples from batch 3 were characterized using in-depth metagenomic sequencing, including short-read metagenomic sequencing of all six timepoints, long-read metagenomic sequencing of rinds at weeks 2, 3, 4, 9 and 13, and metaHi-C of weeks 2, 4, and 13 in the case of Cheeses B and C, and weeks 2 and 13 in the case of Cheese A (Supplemental Figure 3.2.1, Supplemental Table 3.2.1, Supplemental Table 3.2.2).

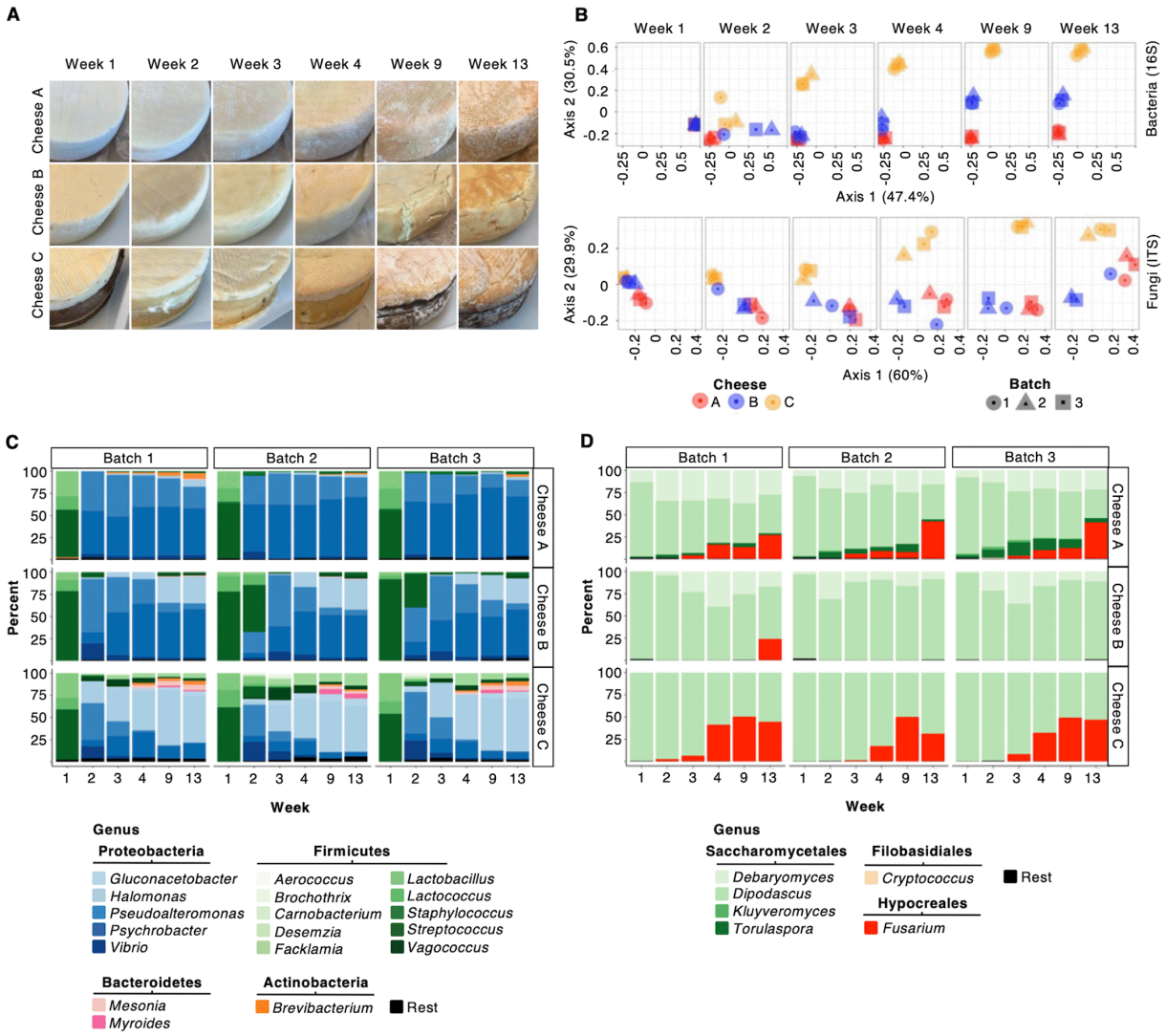


Figure 3.2.1. Three washed-rind cheeses from the same facility show reproducible succession patterns during ripening. (A) For each of the three different washed-rind cheeses (A, B and C) we followed the aging of three different batches produced one week apart. From each batch we collected rind from duplicate wheels at six timepoints. A detailed sequencing overview and collection schedule are found in Supplemental Figure 3.2.1 and Supplemental Table 3.2.1, respectively. Representative images of each of the three cheeses at different timepoints are shown. Cheese C is wrapped in spruce during ripening. In the pictures for weeks 2-4 the spruce has been removed. (B) The 16S and ITS data of the ripening communities was subjected to principal component analyses of the Bray-Curtis dissimilarities. Colors correspond to different cheeses and shapes correspond to different batches.(C, D) Relative abundance plots of (C) bacteria as determined by 16S amplicon sequencing and (D) fungi as determined by ITS sequencing. Shown are the relative abundances of amplicon sequence variants collapsed at the genus-level. Rest = Genera with <1% of the classified reads. Detailed information about 16S and ITS read statistics can be found in Supplemental Tables 3.2.3 and 3.2.4, respectively.

Successional dynamics of the rind communities throughout ripening. To gain a higher-level overview of community dynamics and reproducibility of succession patterns, we analyzed the relative abundance of bacterial and fungal populations using 16S and ITS sequencing (Supplemental Tables 3.2.3 and 3.2.4). The successional dynamics of each of the three cheeses was remarkably reproducible for both the bacterial and the fungal communities (Figure 3.2.1C, D). PCoA analysis based on Bray-Curtis indices indicates that while at the earliest sampled timepoints the different cheeses are highly similar, they diverge from each other along reproducible trajectories throughout aging (Figure 3.2.1B). One exception is one batch of cheese B, which clusters more closely to cheese A at the final fungal sequencing time point. This difference is mainly due to *Fusarium,* which is found in all 3 batches of cheese A and detected in the ITS sequences of that batch of cheese B, but not the other two batches of that same cheese (Figure 3.2.1D). Indeed, a wheel from this batch has a visibly different rind than cheese wheels from the other two batches (Supplemental Figure 3.2.2). Although at the genus level the three communities diverge over time, there are consistent successional patterns at the Phylum and Order level across cheeses. All three cheeses are dominated by Saccharomycetales throughout ripening with Hypocreales reproducibly establishing themselves in the communities of cheeses A and C, but not cheese B (Supplemental Figure 3.2.3B). Regarding the bacterial communities, all three cheeses are dominated by Firmicutes in Week 1, likely due to the lack of rind resulting in sampling of lactic acid bacteria in the cheese core. Proteobacteria quickly take over and dominate the bacterial communities of all three cheeses by the end of ripening. Cheese A also shows a reproducible establishment of

Actinobacteria and Cheese C additionally shows a reproducible establishment of Bacteroidetes in the communities (Supplemental Figure 3.2.3A).

While amplicon sequencing provides a high-level overview of community succession and its reproducibility, it does not address the relative abundance of bacteria and fungi to each other. To close this information gap, we performed taxonomic classification of long-read shotgun metagenomic sequencing data (Supplemental Table 3.2.5). The majority of reads were classified to at least Genus level (Supplemental Figure 3.2.4). Long-read-based taxonomic classifications revealed that all three cheeses were heavily dominated by bacteria and that the fungi only constituted a small proportion of the communities, especially at the end of ripening (Figure 3.2.2, Supplemental Figure 3.2.5). Consistent with the amplicon sequencing, long-read based taxonomy revealed genera that are shared between all three cheeses, such as *Psychrobacter* and *Pseudoalteromonas*, as well as genera that were specific to two or only one of the cheeses, such as *Alcaligenes* and *Sphingobacterium*, which are specific to Cheese C (Figure 3.2.2, Supplemental Figure 3.2.5).

Overall, Cheese C contained the largest number of unique taxa per rank at all sampled timepoints (Supplemental Figure 3.2.6), which is in accordance with the amplicon results. Cheese C also shows the highest degree of taxa turnover with an initial dominance of *Pseudoalteromonas*, which is then taken over by *Halomonas* and *Alcaligenes*. In addition, gram-positive community members such as *Brevibacterium* and *Sphingobacterium* become more abundant over time in Cheese C than they do in the other two communities. In contrast, Cheese B is initially dominated by the yeast *Debaryomyces*, before a bacterial community dominated by *Psychrobacter* and *Pseudoalteromonas* takes over. Eventually,

Pseudoalteromonas is largely displaced from the Cheese B community, while *Halomonas*, members of the family Moraxellaceae and other unidentified Proteobacteria become more abundant. In contrast to both Cheeses B and C, Cheese A shows very minimal community succession between weeks 2 and 13 and is dominated by *Psychrobacter* and other members of the family Moraxellaceae throughout this entire period of ripening.

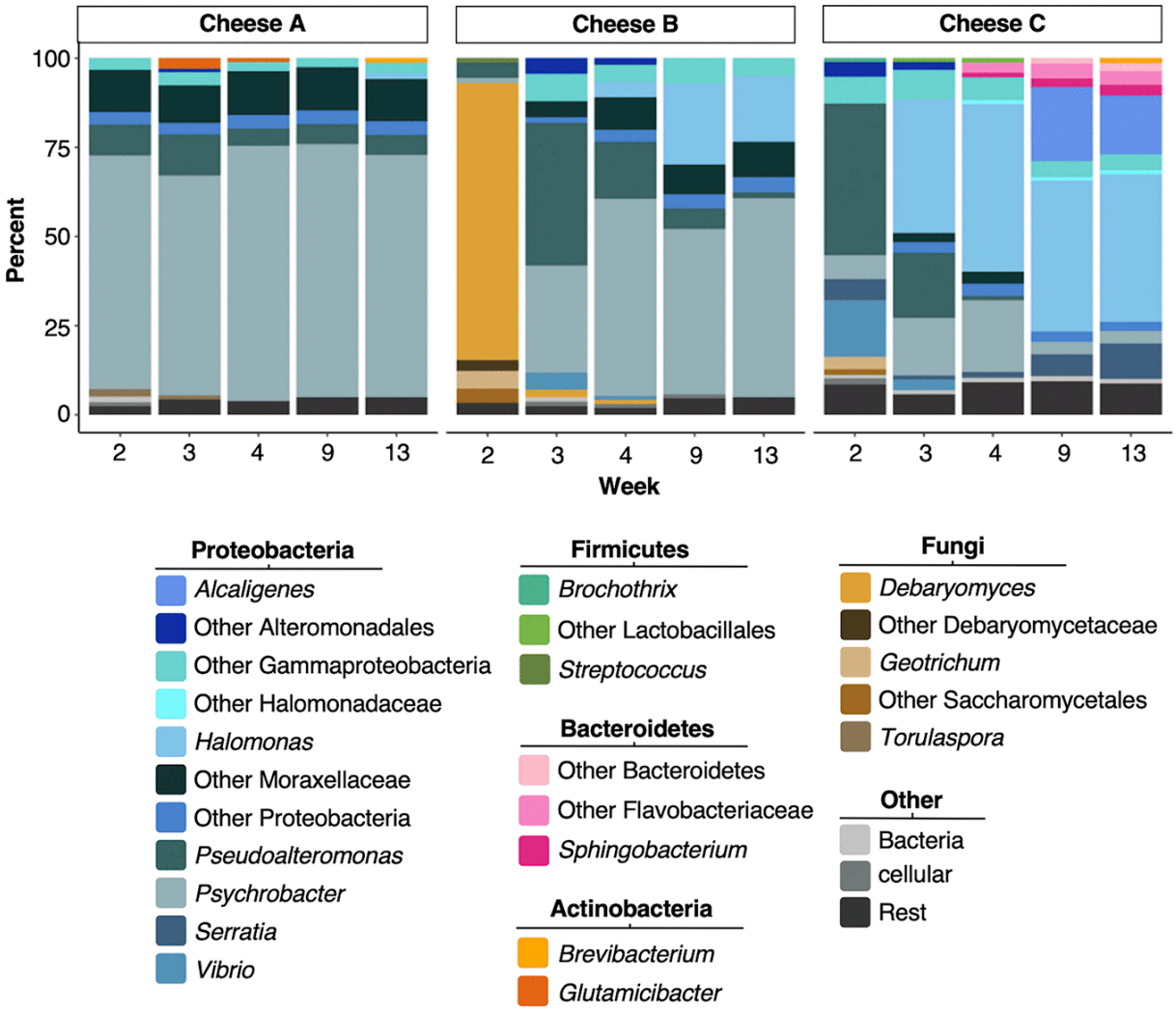


Figure 3.2.2. Long-read-based relative abundance reveals dominance of bacterial community members. Relative abundance plots of taxa (aggregated at the genus level or higher) detected by long-read shotgun metagenomic sequencing. Rest = taxa with <1% of the classified reads. Detailed information about read statistics can be found in Supplemental Table 3.2.5.

A phyla-spanning genomic catalog of washed-rind cheese communities. We next aimed to generate a genomic catalog for each of the three cheeses to help provide insights into species- and strain-level diversity as well as functional potential and horizontal gene transfer. To do so, the long-read shotgun data was assembled for each cheese either by timepoint or across all timepoints (co-assemblies) (Supplemental Table 3.2.6). The assemblies were then binned to generate metagenome-assembled genomes (MAGs) (Supplemental Table 3.2.7). In addition, community members from Cheese B were isolated and subjected to short- and long-read sequencing for *de novo* hybrid genome assembly (Supplemental Table 3.2.8). For each cheese, we combined all high-quality MAGs from the individual timepoint assemblies, the co-assemblies, and the isolate genomes (for cheese B). We then de-replicated each dataset and selected representative MAGs for each cheese. For selection of representative MAGs/ genomes within a given cluster, isolate genomes were given the highest priority, followed by circular MAGs from individual timepoint assemblies, then circular MAGs from co-assemblies, then complete, non-circular MAGs from individual timepoint assemblies and finally complete, non-circular MAGs from co-assemblies. If several bins per cluster fell into the same category the MAGs were prioritized based on quality as assessed by CheckM.

For cheeses A and C, we recovered 17 and 37 high-quality MAGs, respectively (Figure 3.2.3, Supplemental Tables 3.2.9, 3.2.11). In cheese A, 12 out of the 17 high-quality MAGs were both single-contig and circular. In the case of cheese C, 24 of the 37 high-quality MAGs were single-contig and 19 of those were circular. For cheese B, we recovered 11 high-quality MAGs and 16 isolate genomes (Supplemental Table 3.2.10). Four out of the 11 high-quality MAGs were both single-contig and circular. Consistent with our amplicon sequencing

data, comparing the genomic catalogs recovered from the three cheeses using ANI values reveals that the cheeses contain both common and unique genomes (Supplemental Figure 3.2.7). Specifically, when considering an ANI cut-off of 99, we observed that 6 MAGs were represented in the genomic catalogs from all three cheeses. Based on GTDB-Tk(Chaumeil et al. 2019), these MAGs were annotated as *Mesonia* sp., *Vibrio casei, Pseudoalteromonas nigrifaciens, Psychrobacter alimentarius, Vibrio litoralis,* and *Pseudoalteromonas pyrdzensis.* Additionally, cheeses A and B have 8 MAGs (and isolate genomes) in common, cheeses A and C share 2 MAGs and cheeses B and C share 5 MAGs (and isolate genomes) (Supplemental Table 3.2.12). Last, cheese A has 1 unique MAG, cheese B has 8 unique MAGs (and isolate genomes) and cheese C has 24 unique MAGs (Figure 3.2.3B).

All together, the genomic catalog covers the bacterial phyla Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria. For each cheese, Proteobacteria are the group with the most MAGs (and isolate genomes) with 14, 18 and 18 MAGs (and isolate genomes) representing this phylum in cheeses A, B and C, respectively. Further, we are able to recover multiple species and/or strain representatives for several genera. For cheeses A and B we observe multiple co-occurring *Psychrobacter* species and/or strains. 6/14 and 9/18 Proteobacteria are of the genus *Psychrobacter* for cheeses A and B, respectively. Both cheeses also contain a diversity of distinct *Pseudoalteromonas* representatives with 4 and 3 distinct MAGs (and isolate genomes) recovered from cheeses A and B, respectively. Cheese C also contains 2 and 3 distinct MAGs belonging to the genera *Psychrobacter* and *Pseudoalteromonas*, respectively. In addition, we observed species and/or strain-level diversity in the genera *Alcaligenes, Brevibacterium, Halomonas, Leucobacter,*

Microbacterium, Serratia, Sphingobacterium, Staphylococcus and *Vibrio.* We further note that 4/17, 8/27 and 15/37 MAGs (and isolate genomes) from cheeses A, B, and C respectively, were not classified at the species level by GTDB-Tk, which could indicate potentially new species in our genomic catalog.

To gain an overview of how abundant each of the recovered MAGs (and isolate genomes) was during community development, the communities were also subjected to short-read metagenomic sequencing and the reads were mapped to the genomes in the genomic catalog (Figure 3.2.3). For almost all timepoints, more than 50% of the reads were mapped to the genomic catalog (Supplemental Table 3.2.13). The only exception to this is Cheese B at week 2, in which only about 19% of the reads mapped to the genomic catalog. From the long-read based community composition analysis (Figure 3.2.2), we know that at this time point Cheese B is dominated by a fungus. Since the genomic catalog provided in Figure 3.2.3 only contains bacterial genomes, it is not surprising that a low amount of short reads from week 2 maps against the bacterial genomic catalog for Cheese B. In contrast, for some time points over 90% of the short reads from Cheeses B and C map to the genomic catalog indicating that the catalog provides a good cross-section of the bacterial communities at these time points.

The availability of species- and strain-resolved genomic catalogs next allowed us to interrogate the successional dynamics of each cheese more deeply. For example, amplicon and long read-based taxonomy suggested that there were relatively small changes in community composition over time in Cheese A. However, short read mapping back to the genomic catalog revealed species and strain-level temporal dynamics. Specifically, *Pseudoalteromonas nigrifaciens* initially slightly dominates over *Pseudoalteromonas*

prydzensis in Cheese A, while by the end of ripening in week 13 *Pseudoalteromonas prydzensis* slightly dominates over *Pseudoalteromonas nigrifaciens*. Similarly, *Vibrio casei* initially dominates over *Vibrio litoralis* in Cheese A, while already by week 3, *Vibrio litoralis* starts to dominate over *Vibrio casei.* Very similar dynamics are also observed in cheese B for *Pseudoalteromonas* and *Vibrio* intragenus temporal dynamics. *Pseudoalteromonas nigrifaciens* initially dominates over *Pseudoalteromonas prydzensis* before finally being taken over by the other. Similarly, *Vibrio litoralis* initially dominates over *Vibrio casei* before *Vibrio casei* catches up in terms of relative abundance towards the end of ripening. For both cheeses A and B, we saw little variation in abundances across the *Psychrobacter* MAGs (and isolate genomes), potentially due to a high similarity between the MAGs (and isolate genomes) or to stable coexistence of species/strains over time. For cheese C, the abundance of *Pseudoalteromonas* MAGs decreases over time, however their relative abundance to each other does not change significantly. For *Psychrobacter*, only one of the two representative MAGs decreases in abundance over time, while the other shows a stable abundance based on read mapping.

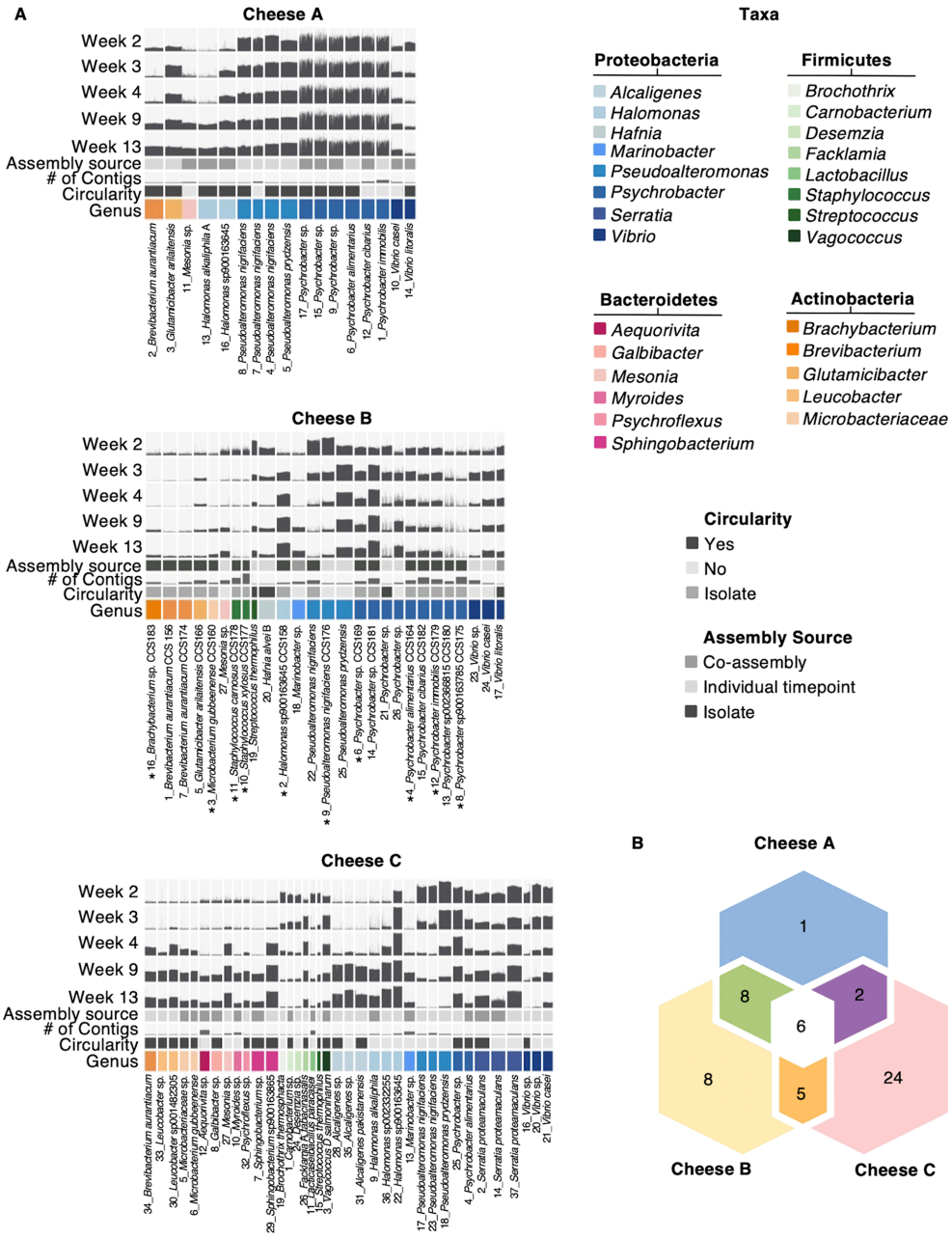


Figure 3.2.3. A phyla-spanning bacterial genomic catalog of three washed-rind cheeses. (A) Anvi'o plots show dereplicated high-quality metagenome-assembled genomes (MAGs) (and isolate genomes in cheese B). MAGs were generated either by assembling HiFi-reads from individual timepoints or by co-assembling reads from all timepoints of a cheese. For each MAG, the number of contigs is indicated. The bin names contain the genomic catalog bin number and the predicted taxonomy of the bin. To estimate relative abundances of MAGs over time, the short-reads from weeks 2, 3, 4, 9 and 13 were mapped to the genomic catalog for each cheese. Supplemental Table 3.2.7 shows all non-dereplicated bins. Genomes marked with a * were included in *in vitro* community reconstruction. (B) Overlap of high-quality MAGs (and isolate genomes) from cheeses A,B, and C, with MAGs having an ANI >99% considered to be the same.

Meta-HiC and long reads associate viruses and plasmids with their hosts. In addition to generating a genomic catalog for each cheese, we were interested in leveraging the existing long-read metagenomics data to characterize the diversity of plasmids and viruses within these microbiomes. To help capture any short and/or low abundance mobile genetic elements that might have been missed in the long-read assemblies, we first generated mega-assemblies of each cheese combining both long-read and high-depth short-read datasets. In brief, the short metagenomic shotgun reads were mapped to the co- and individual timepoint assemblies and any reads that did not map were assembled. The resulting short read-based contigs were combined with the de-replicated unbinned long read-based contigs and the contigs from the de-replicated MAGs to yield mega-assemblies for each of the three cheeses. These mega-assemblies were subjected to plasmid prediction and virus prediction (Supplemental Table 3.2.14, Supplemental Figure 3.2.8). We identified 419, 297 and 343 putative plasmid contigs in Cheeses A, B and C, respectively. Additionally, we identified 109, 116 and 214 lysogenic and 40, 32 and 67 lytic virus contigs for cheeses A, B and C, respectively. Of those, 4, 4 and 7 contigs were classified as complete, circular viruses, while 22, 32 and 57 contigs were classified as high quality draft viruses for cheeses A, B and C, respectively. We last considered the distributions of the predicted viruses (Supplemental Figure 3.2.9) and plasmids (Supplemental Fig 3.2.10). In this regard, cheese C stands out for the presence of several predicted viruses with sizes larger than 200,000 basepairs, which is a hallmark of jumbophages(Yuan and Gao 2017).

We next used the Hi-C data generated for various timepoints to identify putative hosts for the extrachromosomal MGEs (plasmids and lytic phages). MGE predictions were used as

inputs for the viralAssociationPipeline(Derek M. Bickhart et al. 2019; D. M. Bickhart et al. 2021) program together with the total mega-assemblies, the metaHi-C reads and the long reads (Figure 3.2.4, Supplemental Figure 3.2.11, Supplemental Table 3.2.15). Altogether, 70, 21 and 120 of the predicted extrachromosomal MGEs associate with MAGs, while an additional 209, 105 and 89 extrachromosomal MGEs associate with non-MAG (unbinned) contigs for cheeses A, B and C, respectively. 173, 199 and 195 of the extrachromosomal contigs remain unassociated (Supplemental Figure 3.2.11). Of the extrachromosomal MGEs associated with MAGs, we identify 40, 11 and 96 MGEs for Cheeses A, B and C, respectively, that associate with a MAGs (or isolate genome) at one timepoint only. 31, 10 and 23 MGEs from Cheeses A, B and C, respectively, associate with the same MAG (or isolate genome) at two timepoints and 1 MGE each for Cheese B and Cheese C associate with the same MAG at three timepoints. The vast majority of these putative MGEs was not associated with their putative hosts through the initial binning. For Cheeses A, B and C 7, 1, and 11 MGEs, respectively, associate with more than one MAG.d One MGE, a predicted plasmid (s673.ctg001008l_6, indicated by an asterisk in Figure 3.2.4), associates with two different MAGs at two different timepoints, indicating a putative HGT event. At Week 2 this MGE is associated with a *Psychrobacter sp.* host and at week 13 this same MGE is associated with a *Pseudoalteromonas prydzensis* host. Interestingly, this plasmid is predicted to encode genes involved in iron uptake, which have previously been associated with horizontally-transferred regions in cheese (Bonham, Wolfe, and Dutton 2017). The full results of the association pipeline are found in Supplemental Table 3.2.15. We next looked at MGEs that changed their genus association over time (Supplemental Table 3.2.16). Twelve, six, and three MGEs from

Cheeses A, B and C, respectively, changed hosts over the course of sampling based on taxonomic prediction of the host contig. We identified several predicted instances of interkingdom host changes, especially in Cheese B between *Geotrichum* and proteobacterial species. Again, we identified iron uptake genes associated with potentially transferred elements. Some of these elements also contain genes related to phosphonate transport, which has also previously been observed in association with horizontally-transferred iron uptake regions in cheese(Bonham, Wolfe, and Dutton 2017).

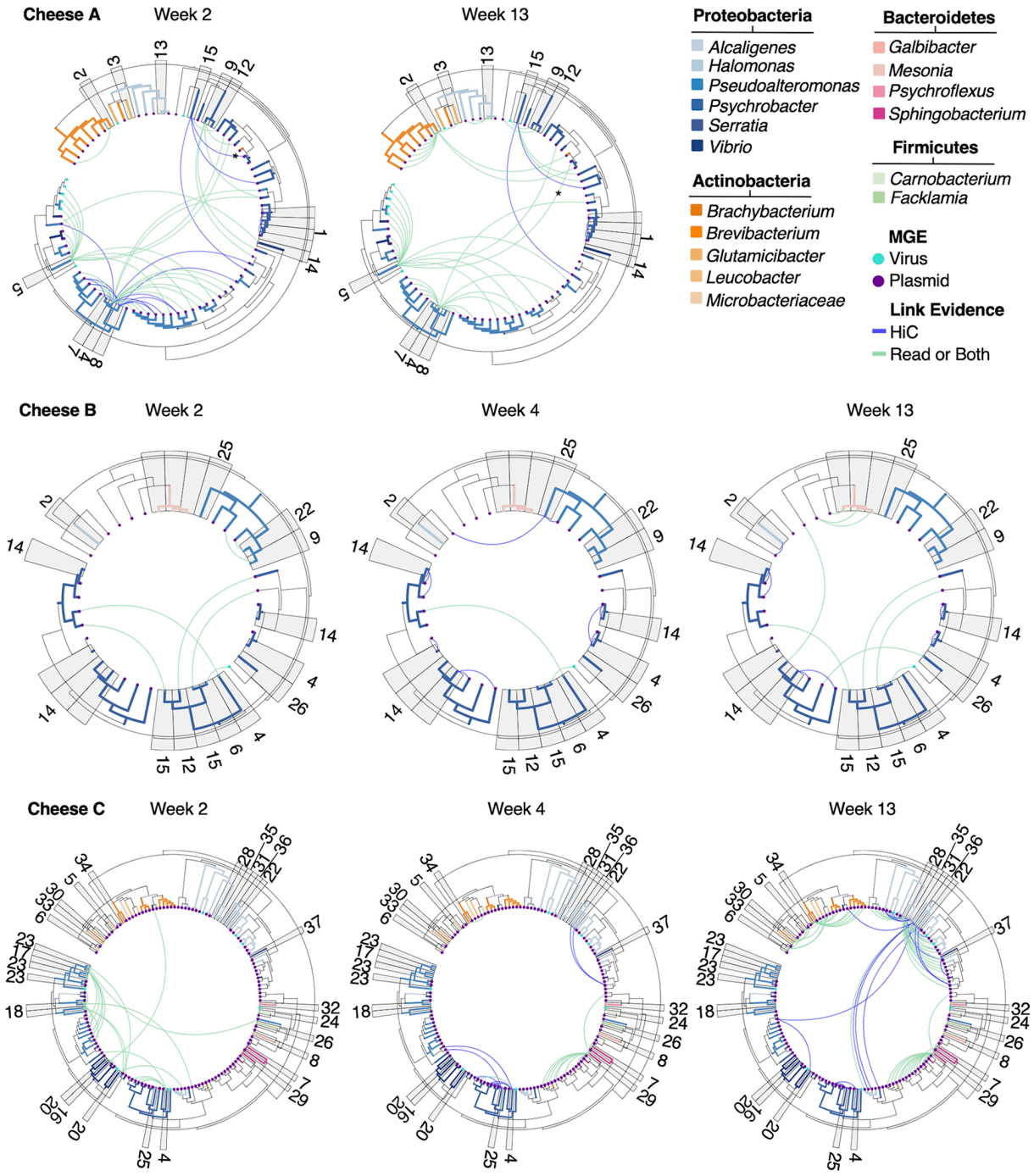
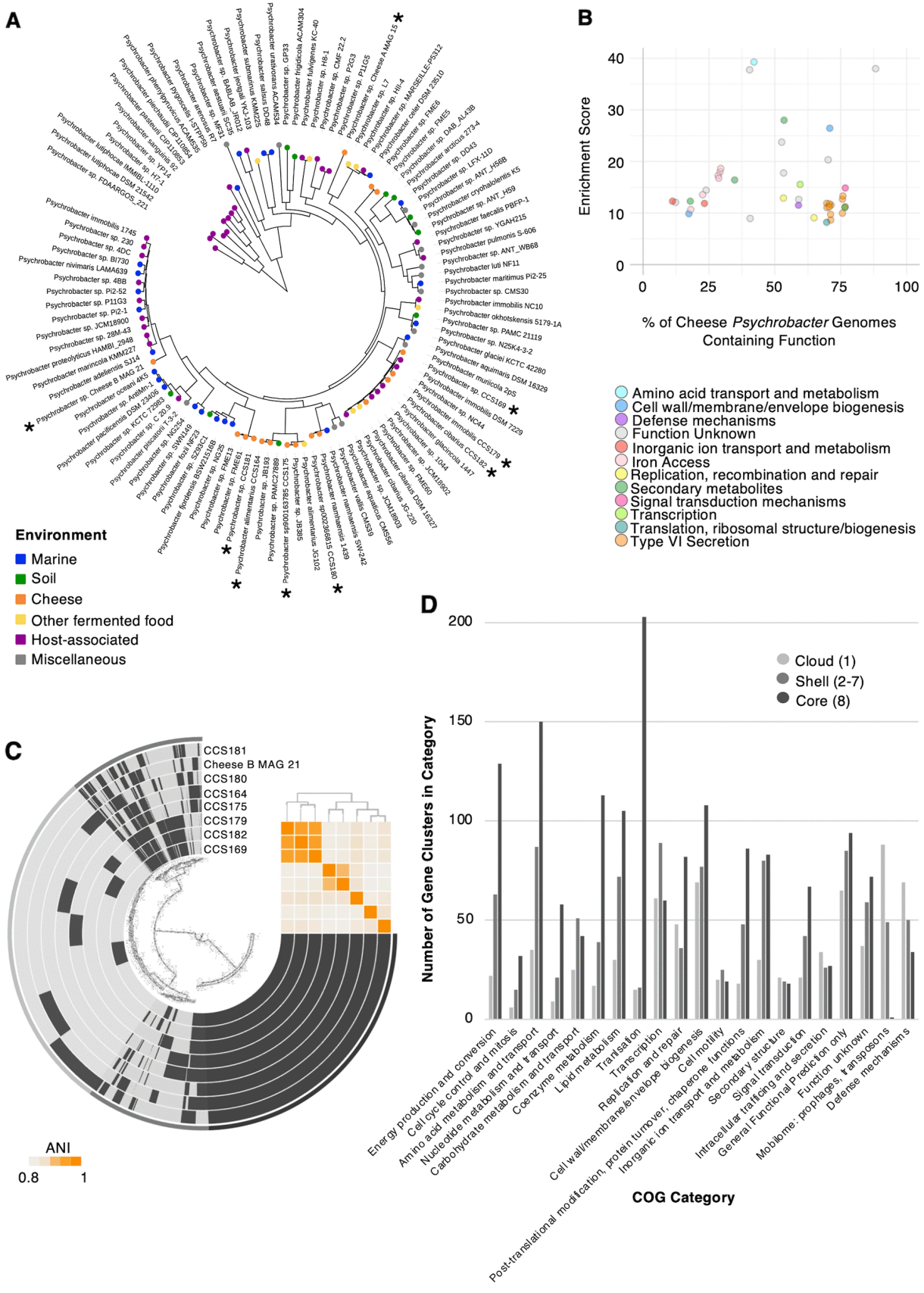


Figure 3.2.4. metaHi-C and long read-based evidence of MGE associations with hosts from the washed-rind cheese genomic catalog in iTOL graphs. The branches of the trees represent contigs and are colored by the contigs predicted taxa. Contigs that are classified as lytic viruses and plasmids are indicated with colored dots (teal = virus, purple = plasmid). Bin numbers for MAG host contigs are shown. Associations between MGEs and their hosts are indicated by the lines. Dotted lines indicate metaHiC-based evidence for association, while solid lines indicate long read-based evidence or both metaHiC and long read-based evidence. The asterisk indicates contig s673.ctg001008l_6.

Pangenome analysis of *Psychrobacter* and functional enrichment in cheese isolates. Our genomic catalog revealed high sub-genus level diversity within the *Psychrobacter* genus. To investigate conserved and unique gene sets across *Psychrobacter* species, we aimed to analyze the *Psychrobacter* pangenome. We first determined the non-redundant *Psychrobacter* isolate genomes and MAGs from the combined three-cheese dataset. This resulted in a total of seven isolate genomes and two MAGs. An additional 97 publicly available *Psychrobacter* genomes were combined with these nine genomes in a pangenomic analysis (Figure 3.2.5A, Supplemental Table 3.2.17). The 97 publicly available genomes were sourced from marine, soil, host-associated, cheese, other fermented food, and other miscellaneous environments . We attempted to include at least one representative of every *Psychrobacter* species possessing a publicly available genome. As expected, pangenomic analysis identified a core set of genes common to *Psychrobacter* from diverse environments. This core gene set, defined as present in 95% of genomes, consisted of 1045 genes and made up 9.4 percent of total gene clusters, with the remaining 90.6 percent of gene clusters classified as accessory by panX (Figure 3.2.5A). A core-genome SNP tree constructed from all variable positions of all single copy core genes showed some evidence of clustering of genomes by environment and, with the exception of *Psychrobacter immobilis*, by species (Figure 3.2.5B). These data suggest that *Psychrobacter* species may have environment-specific gene sets. Functional enrichment analysis of gene clusters was used to find functions (clusters of orthologous groups (COGs)) that were enriched in genomes of cheese isolates compared to genomes from other environments. Specifically, a group of genes related to iron access, particularly through the use of iron-chelating siderophores, was enriched in *Psychrobacter* genomes from cheese (adj.

q-value < 0.1, Figure 3.2.5C, Supplemental Table 3.2.18). In addition, genes related to type VI secretion systems, contractile defense systems that bacteria can use to transport effector proteins into target cells, were also enriched in cheese *Psychrobacter* genomes relative to other environments (adj. q-value < 0.1, Figure 3.2.5C). An alignment of type VI secretion gene regions identified in the cheese genomes revealed a consistent organization in genomes from cheeses made in geographically separated origins. Eleven of the 13 cheese genomes contained two distinct type VI clusters in separate genomic regions, one larger cluster (often with *impA* to *impH*) and one smaller cluster (often with *impI* to *impM*) (Figure 3.2.5D). The larger region was not identified in *Psychrobacter* sp. str. FME6. In *Psychrobacter* sp. str. CCS181, these two regions are adjacent to each other in the genome. These type VI clusters were often bounded by tRNA/tmRNA regions. In two strains, CCS179 and JB193, these clusters also contain phage-related gene sets within the tRNA/tmRNA boundary points in addition to the type VI genes. Both of these clusters were predicted by PHASTER to contain intact prophage(Zhou et al. 2011; Arndt et al. 2016).

Figure 3.2.5. Pangenomic analysis of *Psychrobacter* genomes from diverse environments. (A) Pangenomic analysis was performed on 106 genomes from marine, soil, cheese, other fermented food, and miscellaneous other environments. (B) Phylogenetic tree of *Psychrobacter* genomes based on alignment of single copy core genes. (C) Functional enrichment of COG categories in *Psychrobacter* genomes from cheese relative to other environments. (D) Alignment of the two type VI secretion gene clusters from *Psychrobacter* genomes from various cheeses.


A
Environment
Marine
Soil
Cheese
Other fermented food
Host-associated
Miscellaneous
B
Enrichment Score
% of Cheese Psychrobacter Genomes Containing Function
Amino acid transport and metabolism
Cell wall/membrane/envelope biogenesis
Defense mechanisms
Function Unknown
Inorganic ion transport and metabolism
Iron Access
Replication, recombination and repair
Secondary metabolites
Signal transduction mechanisms
Transcription
Translation, ribosomal structure/biogenesis
Type VI Secretion
C
CCS181
Cheese B MAG 21
CCS180
CCS164
CCS175
CCS179
CCS182
CCS169
ANI
0.8
1
D
Cloud (1)
Shell (2-7)
Core (8)
Number of Gene Clusters in Category
COG Category
Energy production and conversion
Cell cycle control and mitosis
Amino acid metabolism and transport
Nucleotide metabolism and transport
Carbohydrate metabolism and transport
Coenzyme metabolism
Lipid metabolism
Translation
Transcription
Replication and repair
Cell wall/membrane/envelope biogenesis
Cell motility
Post-translational modification, protein turnover, chaperone functions
Inorganic ion transport and metabolism
Secondary structure
Signal transduction
Intracellular trafficking and secretion
General Functional Prediction only
Function unknown
Mobilome: prophages, transposons
Defense mechanisms

***In vitro* community reconstruction.** Given the utility of previous model systems based on cheese rind microbiomes, we next aimed to establish whether the washed-rind cheese communities lend themselves to *in vitro* experimentation. To this end we selected 16 microbial species (13 isolates from Cheese B, 3 from previous isolation efforts), that would represent both the breadth and depth of diversity found in a typical washed-rind cheese microbiome. Given that the metagenomic analysis presented here revealed that multiple species and strains of the genus *Psychrobacter* can co-occur we chose to include 4 *Psychrobacter* isolates from Cheese B (*Psychrobacter alimentarius* CCS164, *Psychrobacter sp.* CCS169, *Psychrobacter sp*900163785 CCS175 and *Psychrobacter immobilis* CCS179). These isolates represent two different groups of *Psychrobacter* (Figure 3.2.5B). In addition to the *Psychrobacter* isolates, the community also contains four other Proteobacteria: *Pseudoalteromonas nigrifaciens* CCS176*, Halomonas sp900163645* CCS158*, Vibrio casei* JB196*, and Hafnia alvei* JB232; two Firmicutes: *Staphylococcus xylosus* CCS177 and *S. carnosus* CCS178; four Actinobacteria: *Brevibacterium linens* JB5, *Brachybacterium ssp.* CCS183, *Glutamicibacter arilaitensis* CCS165, and *Microbacterium gubbeenense* CCS160; and two fungi: *Debaryomyces hansenii* CCS145 and *Galactomyces geotrichum* CCS187. The selected cheese rind isolates were then combined into *in vitro* communities following established protocols. In brief, a total of 100,000 CFUs per community member were inoculated on the surface of a 10% Cheese Curd Agar petri dish (Cosetta and Wolfe 2020). To mimic the production process for washed-rind cheeses, the surfaces of the plates were washed with 20% NaCl using a sterile swab every 24 hours in the first 96 hours for a total of four washes. Communities were incubated in the dark at 15°C in a humidified environment

(Supplemental Figure 3.2.12). To assess the reproducibility of the *in vitro* washed cheese rind model, CFU counts of each sampling day (days 3, 5, 7, 21) were done for both the bacterial and fungal members (Supplemental Figure 3.2.13, Supplemental Table 3.2.19). In addition, a portion of each sample from sampling days 7 and 21 was used for DNA extraction and short-read metagenomic sequencing. Sequencing reads were then mapped back to reference genomes of the bacterial community members to track their relative abundance (Figure 3.2.6A, Supplemental Figure 3.2.15, Supplemental Table 3.2.20).

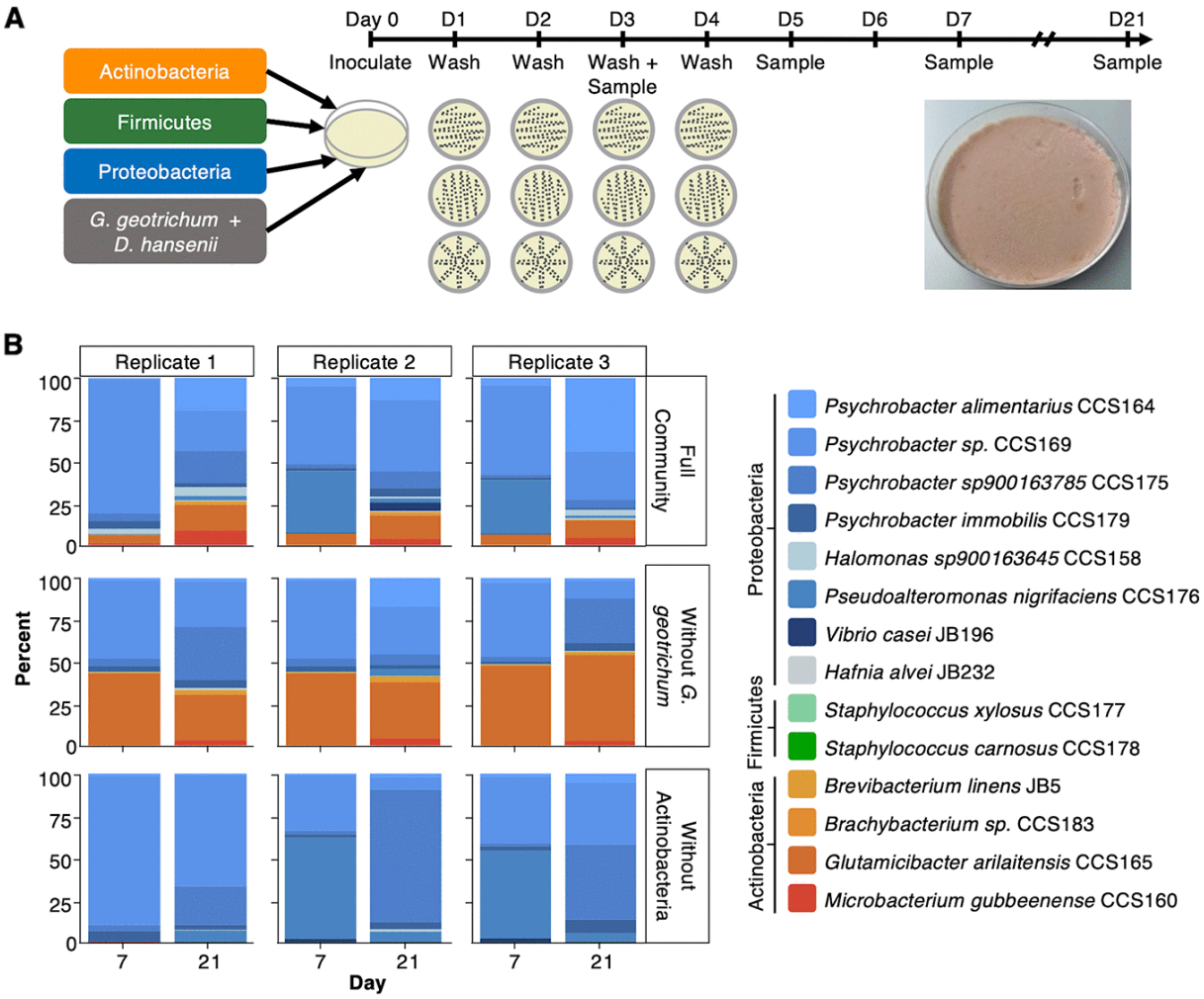


Figure 3.2.6. A 16-member *in vitro* model system based on the washed-rind cheese microbiome. (A) Relative abundance plots of the *in vitro* community after seven and 21 days of growth based on short read mapping to reference genomes. (B) Pangenomic analysis of eight *Psychrobacter* genomes from Cheese B. (C) Functional categories of core (present in all eight genomes), shell (in two-seven genomes), and cloud (unique to one genome) gene clusters from eight *Psychrobacter* genomes from Cheese B.

The additional sequencing and mapping allowed for analysis of *Psychrobacter* at the species level, finding that all four species persist within the full community. However, the relative proportion of each species varied over time, with *Psychrobacter CCS169* most abundant at day 7, and the other three strains, especially *Psychrobacter alimentarius CCS164,* increasing in relative abundance at day 21 (Figure 3.2.6A).

Previous studies revealed a positive correlation between *G. geotrichum* and gamma-proteobacterial species, and a negative correlation between *G. candidum* and actinobacterial species based on co-occurance patterns in a sequencing based survey of cheese rind microbiomes(Wolfe et al. 2014). In addition, pairwise experimental data showed stimulatory effects of *G. geotrichum* on gamma-proteobacterial growth, and inhibitory effects on actinobacterial growth(Wolfe et al. 2014). We took advantage of the fact that this model contains all three of these members (*G. geotrichum*, Proteobacteria, and Actinobacteria) to examine whether these patterns hold in a community context. To do this, we reconstructed a community lacking *G. geotrichum* and evaluated whether this resulted in any differences in the resulting community composition. The removal of *G. geotrichum* favored the relative growth of the Actinobacteria over the Proteobacteria, even though overall bacterial absolute abundance was very similar between these two samples (Figure 3.2.5a). Absolute abundance based on read counts reveals that the decrease in gamma-proteobacterial abundance is largely due to the poor growth of Pseudoalteromonas (Supplemental Table 3.2.19). In contrast, almost all of the actinobacterial species reach higher absolute abundances compared to the full community. Overall, these results are consistent with *G. geotrichum* inhibiting Actinobacteria and stimulating Proteobacteria.

We next tested the effect of removing all actinobacterial members from the community. Similar to the other conditions, *Psychrobacter* species dominate the communities, but increase to 93-95% in the community without Actinobacteria. These communities cluster separately from the full community and community without *G. geotrichum* in principal component analysis (Supplemental Figure 3.2.14). In contrast to the

other communities, *Pyschrobacter alimentarius CCS164* does not increase in relative abundance, remaining below 5% of *Psychrobacter* in all replicates. Instead *Psychrobacter CCS169* competes with *Psychrobacter CCS175* for the most abundant *Psychrobacter*.

To investigate the genetic basis of the coexistence of multiple *Psychrobacter* species within a single cheese, which was supported by our sequencing and *in vitro* growth data, we did an additional pangenomic analysis focused specifically on the eight *Psychrobacter* genomes from cheese B (MAG and isolate genomes) (Figure 3.2.6B). This analysis identified a set of 'cloud' gene clusters unique to each genome, 'shell' gene clusters found in at least 2 genomes but not all genomes, and 'core' gene clusters found in all eight genomes. As accessory gene sets may provide clues to the coexistence of multiple closely-related species or strains, we looked at the COG functional categories associated with the core, shell, and cloud gene sets. While central metabolic processes like amino acid metabolism and translation were most prevalent in the core set, the cloud set was enriched in defense mechanisms and the mobilome (phages and transposons)(Figure 3.2.6C, Supplemental Table 3.2.21). As these gene categories may be involved in interspecies interactions, this may be an interesting starting point for future experiments to investigate the mechanisms underlying species and/or strain co-occurrence.

Discussion

Washed rind cheeses harbor unique, moderate complexity microbiomes that may provide useful systems for the study of microbial succession, interactions, strain-level dynamics and horizontal gene transfer. This work represents a comprehensive, in-depth

analysis of a set of washed rind cheese microbiomes combining culture-independent and culture-dependent approaches. Using amplicon and metagenomic sequencing we showed that the three washed-rind cheeses that were the subject of this study show reproducible succession patterns. Both amplicon sequencing and long-read taxonomic classification show that the mature communities of the investigated cheeses are dominated by bacteria, in particular Proteobacteria. Overall, we detect very similar bacterial and fungal genera as previous studies investigating the composition of washed-rind cheeses(Delcenserie et al. 2014; Kothe et al. 2021; Irlinger and Monnet 2021; Quijada et al. 2020, 2018; Schmitz-Esser et al. 2018; Bokulich and Mills 2013; Wolfe et al. 2014; Bertuzzi et al. 2018; Quigley et al. 2012; Schornsteiner et al. 2014), but provide new insights into species and strain-level diversity and dynamics.

We also provide a catalog of high-quality MAGs (and isolate genomes) that spans the bacterial phyla Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria. In fact, two of the three cheeses yielded representative genomes from all four of those phyla despite low proportional abundance. It is likely that the long read sequencing helped not only recover those low abundance genomes, but also helped in the resolution of species and/or strain-level MAGs. For example, we were able to resolve several *Psychrobacter* genomes in the three cheeses which may have been difficult to resolve using short read sequencing alone.

We utilized the bacterial genomic catalog generated as part of this study to further investigate the biology within these microbiomes. First, we leveraged the long reads together with metaHi-C reads to associate putative mobile genetic elements with their hosts. Many of these mobile genetic elements were not binned with their putative hosts during the initial

binning and only became associated in this additional analysis. We were able to assign 15, six, and 29% of extrachromosomal MGEs to MAG hosts and 46, 32, and 22% of extrachromosomal MGEs to non-MAG contigs for Cheeses A, B, and C, respectively. Using this method we also detected several putative HGT events, including mobile genetic elements predicted to encode iron-uptake pathways. Previous examination of cheese bacterial genomes revealed a strong enrichment of iron-uptake pathways on horizontally-transferred regions. The changes in host association as suggested by metaHi-C in this study require further *in vitro* confirmation and follow-up to fully understand their exact nature and impact within the cheese microbiome.

Next, we investigated the striking diversity of *Psychrobacter* within the communities and investigated functional enrichment within cheese-associated *Psychrobacter* MAGs and isolate genomes as compared to *Psychrobacter* genomes associated with other environments. We identify an enrichment of iron acquisition genes as well as type VI secretion genes in cheese-associated *Psychrobacters*. Both of these functional enrichments are reasonable given what is known about the cheese rind environment. First, the cheese microbiome has evolved to thrive in iron-limiting conditions(Ellison 1994; Monnet et al. 2015; Kastman et al. 2016). Second, the dense microbial communities that make up cheese rind microbiomes may lead to contact-dependent microbial interactions. Type VI secretion systems enable the delivery of cargo proteins that modulate bacterial-bacterial and bacterial-eukaryotic interactions (reviewed in (Russell, Peterson, and Mougous 2014)). As such, contact-dependent interaction machinery enriched in cheese-associated *Psychrobacter* may play a role in species interactions in the densely populated cheese microbiome, as opposed to free-living, marine

Psychrobacter, for example. Future *in vitro* studies could aim to understand how type VI secretion-mediated dynamics contribute to community composition and function.

The comprehensive, longitudinal nature of this dataset makes it a potentially valuable resource for future investigations into the biology of cheese microbiomes or the study of microbiome dynamics in general. For example, the approaches applied here could be used to interrogate other types of MGEs, such as integrative and conjugative elements and prophages. Of particular interest in this dataset would be the >300,000 bp prophage predicted in cheese C, which might be a putative jumbophage. Additionally, one aspect of the long read data, which we did not explore in this paper, is the fact that it can detect methylation patterns on DNA, which can help associate extrachromosomal elements with their hosts and could theoretically be used to identify HGT events (Saak, Dinh, and Dutton 2020). Indeed, a recent study used this approach to recover MAGs from a marine microbial consortium and identify HGT events, phage infection and strain-level structural variation(Wilbanks et al. 2022).

Finally, we showed that, similar to previous work on cheese microbiomes(Wolfe et al. 2014; Morin, Pierce, and Dutton 2018), washed rind communities are amenable to *in vitro* reconstruction. The in vitro system established here represents a higher level of diversity than previous models, and includes species and strain-level diversity. The pairing of an experimental microbiome system with a corresponding in-depth metagenomic dataset should provide ample future opportunities for generating and testing hypotheses related to the many facets of biology present within these microbiomes.

Materials and Methods

Cheese rind sample collection and DNA extraction. Three production batches of three different washed-rind cheeses were sampled over the course of ripening starting from fresh cheese wheels. At weeks 1, 2, 3, 4, 8 and 12 of ripening the cheesemakers overnight shipped two wheels of each sampled batch. The cheeses were stored at 4 °C for up to 48 h prior to harvest. During the harvest, rind was scraped off each wheel using fresh razor blades as in(Wolfe et al. 2014). The rind samples were homogenized by gentle stirring with sterile pipette tips. A rice grain-sized portion of these samples was frozen in cryotubes at -80 °C in phosphate-buffered saline + 40% glycerol. Where applicable, 10 mg of each wheel were collected in sterile 2-ml tubes for cross-linking as part of the ProxiMeta Hi-C library preparation (see below). Finally, 500 mg of each sample were set aside in sterile 1.5-ml tubes at 4 °C for DNA extraction. DNA was extracted within 72h using a phenol-chloroform extraction protocol. If DNA was extracted at a later time, the rind samples were stored at -80 °C.

For the DNA extraction, the rind samples were ground into powder in liquid nitrogen (https://lab.loman.net/2018/05/25/dna-extraction-book-chapter/#__RefHeading___Toc505877552). The rind powder was incubated for 1 hour at 37 °C in 7 ml modified Tris-lysis buffer (10 mM Tris-Cl (pH 8), 100 mM EDTA (pH 8), 1% SDS, 20 µg/ml RNAse A, 20 mg/ml lysozyme). 87.5 µl of Proteinase K (800 units/ml) (New England Biolabs, Ipswich, MA, USA) were added, the samples were mixed by inversion and incubated at 50 °C for 1 hour. The samples were then subjected to two rounds of DNA extraction using equal volumes of phenol-chloroform isoamyl. The final aqueous phase was mixed with equal

volumes of ice cold isopropyl and 0.1 volume of 3M sodium acetate. The precipitated DNA was pelleted in 5 ml centrifuge tubes at 17,000 g for 3 minutes. When the volume of the final aqueous phase was larger than 5 ml, the supernatant was removed and the remaining sample was added to the same centrifuge tubes to repeat the pelleting step. The final pellets were washed in 1 ml ice cold 70% ethanol and air-dried for about 15 min. Last, the DNA was resuspended in 500 µl UltraPure™ DNase/RNase-Free Distilled Water (ThermoFisher Scientific, Waltham, MA, USA) and kept at room temperature overnight to allow the DNA to dissolve. DNA was then stored at -20 °C until further processing.

16S and ITS amplicon sequencing and analysis. For the 16S and ITS sequencing of the community DNA samples, we followed the Illumina-supplied "16S Metagenomic sequencing Library Preparation" protocol (Part # 15044223 Rev. B) and "Fungal Metagenomic Sequencing Demonstrated Protocol" (Document # 1000000064940 v01), respectively. The 16S-specific primers (16S forward: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-CCTACGGGNGGCWGCAG-3', 16S reverse: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GACTACHVGGGTATCTAATCC-3') targeted the 16S V3 and V4 regions. The ITS-specific primers (ITS forward: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-CTTGGTCATTTAGAGGAAGTAA-3', ITS reverse: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-GCTGCGTTCTTCATCGATGC-3') were the same as the ITS_fwd_1 and ITS_rev_1 primers listed in the "Fungal Metagenomic Sequencing Demonstrated Protocol". For each of the two

amplicon PCRs, 1.25 µl DNA (5 ng/µl) from the duplicate wheels for each sample were mixed and the resulting 2.5 µl were amplified either with the 16S- or the ITS-specific primers and Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). The PCR settings used were as follows: initial denaturation at 98 °C for 3 minutes, 25 rounds of: 98 °C - 10 seconds, 55 °C - 30 seconds, 72 °C - 15 seconds, final extension at 72 °C for 2 minutes. Amplicon PCRs were purified using AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA) and eluted with UltraPure™ DNase/RNase-Free Distilled Water (ThermoFisher Scientific, Waltham, MA, USA). For the index PCR, the 16S and ITS amplicon PCRs were subjected to amplification with IDT for Illumina Nextera DNA Unique Dual Indexes (Set A, now called IDT for Illumina DNA/RNA UD Indexes) (produced by: Integrated DNA Technologies, Coralville, Iowa, USA, sold by: Illumina, Inc., San Diego, California, USA) and Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA). Index PCRs were again purified using AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA) and eluted with UltraPure™ DNase/RNase-Free Distilled Water (ThermoFisher Scientific, Waltham, MA, USA). The final Index PCRs were quantified using the Qubit dsDNA HS kit (ThermoFisher Scientific, Waltham, MA, USA) with the Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA), diluted to 2 nM and pooled at equimolar ratios. The pools were diluted to the final loading concentration of 50 pM and spiked with 50 pM PhiX control v3 (Illumina, Inc., San Diego, California, USA) to a final concentration of 10% PhiX. 20 ul of the 16S(+PhiX) and of the ITS(+PhiX) pools were sequenced individually in-house on an iSeq 100 (paired-end, 150 bp) (Illumina, Inc., San Diego, California, USA).

The fastq files of the forward reads of the respective sequencing runs were imported into Qiime2 (version 2020.11.1) (Bolyen et al. 2019) and denoised with Dada2 (Bolyen et al. 2019; Callahan et al. 2016; McDonald, Clemente, et al. 2012) (qiime dada2 denoise-single --i-demultiplexed-seqs with flags --p-trim-left 0 --p-trunc-len 0 --p-n-threads 16). The denoised forward ITS reads were classified using classify-sklearn with a Naive Bayes classifier (qiime feature-classifier fit-classifier-naive-bayes) trained on a custom ITS database(Wolfe et al. 2014; Bolyen et al. 2019; Callahan et al. 2016; Bokulich et al. 2018; McKinney 2010; Pedregosa 2011). The denoised forward 16S reads were classified in the same way using a pre-trained Greengenes database (Greengenes 13_8 99% OTUs full-length sequences (MD5: 03078d15b265f3d2d73ce97661e370b1))(Bolyen et al. 2019; Callahan et al. 2016; Bokulich et al. 2018; McKinney 2010; Pedregosa 2011; Robeson et al. 2020; McDonald, Price, et al. 2012). Detailed read statistics can be found in Supplemental Tables 3.2.1 and 3.2.2. Ordination plots (method:PCoA,distance: Bray-Curtis) were generated using R (version 4.0.2) (https://www.r-project.org/) together with the qiime2R (version 0.99.4), tidyverse (version 1.3.1) and phyloseq (version 1.34.0) packages(Bisanz n.d.; Wickham et al. 2019; McMurdie and Holmes 2013) . The stacked bar plots were also created with R using the qiime2R, tidyverse and phyloseq packages. Taxonomies were collapsed at Genus level for both the ITS and the 16S data. For the stacked bar plots showing a higher taxonomic level, taxonomies were collapsed at phylum (16S) or order (ITS) level. In all cases, taxonomies with less than 1% abundance were aggregated under the category “Rest”.

metaHi-C. 10-15 mg of rind from duplicate wheels for each sample were fixed separately and then combined for further processing according to the ProxiMeta kit methodology from Phase Genomics (Seattle, WA). The multiplexed libraries were sequenced by Novogene (Sacramento, CA, USA) on a HiSeq 4000 with a run configuration of 2 x 150 bp.

Metagenomic shotgun sequencing. For the SMRT sequencing, DNA from duplicate cheese wheels of each sample was combined in equal ratios. 5 µg DNA of each sample were used as input for the library preparation, which was carried out by SNPSaurus (Eugene, Oregon, USA). The DNA was sheared to a modal size of 10 kb using a Megaruptor 2 (Diagenode Diagnostics, Seraing (Ougrée), Belgium). Libraries were prepared using Pacific Bioscience's Express Template Prep kit version 2.0 according to the manufacturer's protocol (https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-Multiplexed-Microbial-Libraries-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf, Menlo Park, CA, USA). Samples were pooled two at a time and size selected using a BluePippin (Sage Sciences, Beverly, MA, USA) with the 0.75% Agarose Dye-free 10-18kb cassette, U1 marker and a 10kb+ cutoff. The final libraries were sequenced by the Genomics & Cell Characterization Core Facility (University of Oregon, Eugene, Oregon, USA) on a Sequel II according to the SMRT Link Set Up (SMRT cell type = 8M, Sequencing kit = v2.0, Sequencing primer = v2, Binding kit = v2.0, Sequencing control = v1, Polymerase binding time = 4 hr, Movie time = 30 hr, Pre-extension time = 2 hr, Loading concentration = 100, 133, 150, or 200 pM (varied by cell), Loading method = Diffusion). HiFi reads were generated using the CCS tool (https://ccs.how/) version 6.2.0 with default settings (ccs in.subreads.bam out.subreads.bam \ --min-

passes 3 \ --min-snr 2.5 \ --min-length 10 \ --max-length 50000 \ --min-rq $MIN_RQ, $MIN_RQ is set accordingly: ccs.Q20 - $MIN_RQ = 0.99, ccs.Q30 - $MIN_RQ = 0.999, ccs.Q40 - $MIN_RQ = 0.9999).

Short-read sequencing libraries were prepared using the Nextera DNA Flex Library Prep workflow (Illumina, Inc., San Diego, California, USA). For each sample, input DNA consisted of DNA from two duplicate cheese wheels from the same sample that was mixed in equal proportions (ng). For multiplexing, IDT for Illumina Nextera DNA Unique Dual Indexes (Set A, now called IDT for Illumina DNA/RNA UD Indexes) (produced by: Integrated DNA Technologies, Coralville, Iowa, USA, sold by: Illumina, Inc., San Diego, California, USA) were used. Pooled libraries were sequenced by the IGM Genomics Center at the University of California San Diego (San Diego, CA, USA) on the NovaSeq 6000 System using both lanes of a NovaSeq SP flow cell and a run configuration of 2 x 250 bp.

Long-read based relative abundance estimations of washed-rind cheese communities. To investigate successional dynamics of the cheese communities based on HiFi reads we utilized the Pacific Biosciences supplied toolkit “PB-metagenomics-tools” (https://github.com/PacificBiosciences/pb-metagenomics-tools). First, HiFi reads were taxonomically classified with the “Taxonomic-Profiling-Nucleotide” pipeline, which uses minimap2(Li 2018) to align sequences to the NCBI nt database and MEGAN-LR(Huson et al. 2018) to interpret alignments and assign reads to taxa. The read counts by taxonomy were generated using the “MEGAN-RMA summary”. Outputs from the “MEGAN-RMA summary” were further processed in R (version 4.0.2) (https://www.r-project.org/) together with the janitor (version

2.1.0), tidyverse (version 1.3.1) and cowplot (version 1.1.1) packages to generate the stacked bar plot in Figure 3.2.2. Additionally, the outputs from the “MEGAN-RMA summary" were used as inputs for the “compare-kreport-taxonomic-profiles” script available in the PB-metagenomics-tools suite to generate Supplemental Figures 3.2.4-6.

Generation of metagenome-assembled genomes.

Metagenomic assemblies were performed with hifiasm-meta (v. 0.2-r053)(Feng et al. 2021) using the default settings. For the individual timepoint assemblies, reads were input separately by time point and cheese. For the co-assemblies, reads from all timepoints were combined by cheese. To evaluate the assemblies and identify high-quality MAGs, the PacBio HiFi-MAG-Pipeline was used (https://github.com/PacificBiosciences/pb-metagenomics-tools, part of the “PB-metagenomics-tools” suite). This pipeline uses minimap2(Li 2018) to align HiFi reads to the contigs to obtain coverage estimates, which are used with MetaBat2(Kang et al. 2019) to perform binning using all contigs. A separate bin set is also constructed from all circular contigs (e.g., one bin per circular contig), and the two binning strategies are compared and merged using DAS_Tool(Sieber et al. 2018). The dereplicated bins are evaluated using CheckM(Parks et al. 2015), and quality thresholds are applied to retain high-quality MAGs (defaults of >70% completeness, <10% contamination, <20 contigs). The high-quality MAGs are then analyzed using the Genome Taxonomy Database Toolkit (GTDB-Tk)(Chaumeil et al. 2019), which attempts to identify the closest reference genome and assign taxonomy for each MAG.

Isolation of bacterial community members. The glycerol stocks of one wheel of Cheese B at weeks 2 and 13 were thawed on ice, homogenized by vortexing and pipetting and aliquoted into smaller working stocks, which were frozen at -80 °C for at least 24 h before further processing. A dilution series of one working stock each of weeks 2 and 13 was plated on plate count agar supplemented with milk and salt (PCAMS, 5 g/L tryptone, 2.5 g/L yeast extract, 1g/L dextrose, 1 g/L whole milk powder, 10 g/L sodium chloride and 15 g/L agar) containing 100 μg/ml cycloheximide to select against the fungal community members. The plates were kept in a 15 °C incubator for 4 days and then in the light at room temperature for an additional 3 days to allow for pigment formation. Colonies with as many distinct colony morphologies as could be distinguished by eye were then purified by restreaking three times on plain PCAMS, each time with an incubation at 15 °C for 4 days and an additional 3-day incubation in the light at room temperature. Colonies from the final re-streak were patched onto plain PCAMS and incubated at 15 °C for 4 days. 2 ml overnight cultures (LB Miller broth) were inoculated from the patches and grown shaking at 240 rpm at 22 °C. After 24h, glycerol stocks were prepared using PBS + 20% glycerol and flash-frozen and stored at -80 °C. The only exception to this was CCS196, which was grown on PCAMS agar at 22 °C for 24 h, harvested in PBS+20% glycerol, flash-frozen and stored at -80 °C.

Short-read, whole genome sequencing of Cheese B isolates. Isolates (CCS156, 158, 160, 164, 165, 166, 169, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 196) were grown from glycerol stocks on plain PCAMS for 2 days at 15 °C and an additional 3 (all except CCS183) or 5 days (only CCS183) at room temperature on the benchtop. DNA was extracted

following the NexteraTM DNA Flex Microbial Colony Extraction protocol (Document #1000000035294v01) with the following modifications: AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA) were used instead of SPB beads and 2 ml cryotubes filled with approx. 250 µl acid-washed beads (1:1 ratio of 425-600 µm and 150-212 µm beads) were used instead of PowerBead Tubes. In addition, cells were collected from the primary streak using a sterile pipette tip. The extracted DNA was quantified using the Qubit dsDNA HS kit (ThermoFisher Scientific, Waltham, MA, USA) with the Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and stored at -20 °C. Steps 25 - 27 were skipped. Due to low amounts of cellular input, steps 7 and 8 were skipped for CCS183 and instead 50 µl were mixed with 20 µl AMPure beads. For the library preparation a maximum of 500 ng of DNA was mixed with UltraPure™ DNase/RNase-Free Distilled Water (ThermoFisher Scientific, Waltham, MA, USA) for a total volume of 30 µl. The libraries were prepared using the Illumina DNA Prep protocol (Document # 1000000025416 v09) and the IDT for Illumina DNA/RNA UD Indexes (Set A, produced by: Integrated DNA Technologies, Coralville, Iowa, USA, sold by: Illumina, Inc., San Diego, California, USA). Tagmented DNA was amplified for 5 cycles, except CCS191, which was amplified for 8 cycles due to low input amounts. Libraries were diluted to 30 nM using buffer RSB from the Illumina DNA Prep protocol, pooled at equimolar ratios and sequenced by Novogene (Sacramento, CA, USA) on a HiSeq 4000 with a run configuration of 2 x 150 bp.

Long-read whole genome sequencing of Cheese B isolates. A subset of isolates (CCS156, 158, 160, 164, 165, 166, 169, 174, 176, 179, 180, 181, 182, 183, 184 and 196) was

additionally sequenced using Nanopore technology (Oxford Nanopore Technologies, Oxford, United Kingdom) to improve isolate assemblies. Isolates were either grown shaking at room temperature in 2 ml LB Miller broth for at least 24 hours and until culture appeared cloudy (CC156, 158, 160, 164, 165, 166, 169, 174, 176, 179, 180, 182, 183, 184) or they were grown on PCAMS agar plates at room temperature for ~48 h (CCS181 and CCS196). Cells from liquid cultures were harvested by centrifugation at 10,000 rpm for 5 min, removal of supernatant by pipetting and freezing at -80 °C. Cells from the agar plates were recovered by adding 2.5 ml PBS onto the plate and dislodging cells with sterile cell scrapers. 1 ml of recovered cell suspension was centrifuged at 10,000 rpm for 5 min and frozen at -80 °C after supernatant removal. Cell pellets were thawed and DNA was extracted using the phenol-chloroform protocol detailed above without the liquid nitrogen grinding step (CCS156, CCS165, CCS166 and CCS176) or using the Qiagen Genomic-tip 20/g kit (Qiagen, Venlo, Netherlands) according to manufacturer instruction. DNA of CCS156, CCS165, CCS166 and CCS184 was extracted from the pellet resulting from the full 2 ml culture, while DNA from the rest of the isolates was extracted from pellets resulting from 1 ml of the overnight culture. DNA extractions were quantified using the Qubit dsDNA HS kit (ThermoFisher Scientific, Waltham, MA, USA) with the Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and their quality was assessed with the Tapestation gDNA assay (Agilent, Santa Clara, CA, USA). DNA was stored at -20 °C until the library preparation was carried out using the Nanopore kit SQK-LSK110 (the DNA control strand was included in the library preparation for CCS156, CCS165, CCS166 and CCS176, but not the others). Libraries were either sequenced right away using a flongle flow cell (FLO-FLG001) or stored at -80 °C until

sequencing. For most samples, base calling was done in real time during the sequencing run (CCS156, CCS158, CCS160, CCS164, CCS166, CCS169, CCS174, CCS179, CCS180, CCS181, CCS182, CCS183, CCS184, CCS196) using the fast basecalling model in the MinKnow software (version 21.02.1 running Gruppy version 4.3.4 for CCS156, CCS158, CCS166, CCS180, CCS181, CCS183 and CCS196 or version 21.06.0 running Guppy version 5.0.11 for CCS160, CCS164, CCS169, CCS174, CCS179, CCS182 and CCS184) with a minimum qscore of 7. For a couple of samples (CCS165 and CCS176) base calling was done post sequencing using Gyppy (version 5.0.16) on the command line with the fast base calling model (config file dna_r9.4.1_450bps_fast.cfg) and the minimum qscore set to 7. We recovered read N50s between 4 - 30 kb. In all cases, the fastq files in the "pass" folders were concatenated by isolate and used for further analysis.

Isolate genome assemblies. To generate the isolate assemblies, the Illumina reads were first trimmed using Trimmomatic(Bolger, Lohse, and Usadel 2014) (version 0.39) using the paired-end mode and arguments ILLUMINACLIP:NexteraPE-PE.fa:2:40:15:2:True SLIDINGWINDOW:4:20 MINLEN:20. The paired and unpaired, trimmed reads of CCS175, CCS177 and CCS178 were assembled using SPAdes(Prjibelski et al. 2020) (version v3.13.0, with arguments -t 14 -m 50 -k 33,55,77,99,127). For CCS156, CCS158, CCS160, CCS164, CCS165, CCS166, CCS169, CCS174, CCS176, CCS179, CCS180, CCS181, CCS182, CCS183, CCS184 and CCS196 hybrid assemblies were performed with SPAdes(Antipov et al. 2016) using both the paired and unpaired, trimmed Illumina reads as well as the Nanopore

reads as input (version v3.13.0 with arguments -t 14 -m 50 -k 33,55,77,99,127). Contigs below 1000 bp were removed.

Generation of genomic catalog. MAGs resulting from co-assemblies and individual timepoint assemblies (and isolate assemblies in case of cheese B) were de-replicated with dRep (version 3.2.2)(Olm et al. 2017). CheckM (version 1.1.3)(Parks et al. 2015) results were provided as an Excel sheet. Genomes for the genomic catalog were selected by hand from the resulting MASH clusters by prioritizing isolates (for CheeseB) over circular individual timepoint MAGs over circular co-assembly MAGs over complete, non-circular individual timepoint MAGs over complete, non-circular co-assembly MAGs. If several genomes per cluster fell into the same category, the genomes were prioritized based on low contamination and high completeness as determined by CheckM. Taxonomies of the final genomes in the genomic catalog were determined by GTDB-Tk (version 1.6.0, classify workflow)(Chaumeil et al. 2019). To determine the abundance of the genomes from the genomic catalog over time, the genomes in the catalog were indexed using bwa index and the Illumina reads from week 2 - 13 were mapped to the this indexed genomic catalog using bwa mem (Version: 0.7.17-r1188)(Li 2013). Before mapping, the short-read metagenomic reads were concatenated by samples across the two lanes of the sequencing run and trimmed using Trimmomatic(Bolger, Lohse, and Usadel 2014) (version 0.39) using the paired-end mode and arguments ILLUMINACLIP:NexteraPE-PE.fa:2:40:15:2:True SLIDINGWINDOW:4:20 MINLEN:20. Samtools(Danecek et al. 2021) (version 1.9 using htslib 1.9) was used to convert .sam files into .bam files with samtools view and .bam files from all timepoints of each cheese were

combined using samtools cat. Samtools flagstat was used to determine the percentage of reads mapped to the genomic catalog. Anvi'o (version 7.1 'hope')(Eren et al. 2021) was used to generate the plots in Figure 3.2.3.

Generation of mega-assemblies. For each cheese, contigs from the co- and individual timepoint assemblies were concatenated and contigs <1000 bp were removed. The concatenated assemblies were indexed using bwa index, and short-read metagenomic reads were mapped to this indexed, concatenated assembly using bwa mem (bwa version 0.7.17-r1188)(Li 2013). Before mapping, the short-read metagenomic reads were concatenated by samples across the two lanes of the sequencing run and trimmed using Trimmomatic(Bolger, Lohse, and Usadel 2014) (version 0.39) using the paired-end mode and arguments ILLUMINACLIP:NexteraPE-PE.fa:2:40:15:2:True SLIDINGWINDOW:4:20 MINLEN:20. Samtools(Danecek et al. 2021) (version 1.9 using htslib 1.9) was used to convert .sam files into .bam files with samtools view and .bam files from all timepoints of each cheese were combined using samtools cat. samtools view was then used to extract all short-read metagenomic reads that did not map to the indexed, concatenated assemblies. The resulting bam files were converted into fastq files using the bamToFastq command (version 0.5.3) from the BEDTools suite(Quinlan and Hall 2010) and fastq_pair from the fastq-pair package (version 1.0)(Edwards and Edwards 2019) was used to synchronize the newly generated fastq files. The reads in the resulting fastq files were assembled using SPAdes(Antipov et al. 2016) using both the paired and unpaired, trimmed Illumina reads as input (version v3.13.0 with arguments -t 14 -m 50 -k 33,55,77,99,127). To generate the mega-assembly for each cheese,

unbinned contigs from the co- and individual timepoint assemblies of the respective cheese were first de-replicated. To de-replicate these contigs, a custom nucleotide blast database was created consisting of the unbinned contigs. Unbinned contigs were then compared to this database using blastn(Camacho et al. 2009; Altschul et al. 1990) (version 2.10.1+, options: -outfmt "6 qseqid sseqid pident qcovs length qlen slen evalue score" -evalue 1e-6 -perc_identity 99 -word_size 20 -num_threads 12). A contig was considered redundant if there was an overlap of contigs of at least 90%; the smaller contig in this case was considered redundant. Seqkit(Shen et al. 2016) (version 2.2.0) was then used to extract the non-redundant unbinned contigs from the full set of unbinned contigs. These non-redundant unbinned contigs were then combined with the contigs resulting from the assembly of the unmapped Illumina reads and the contigs from the selected de-replicated MAGs in the genomic catalog. This mega-assembly should incorporate all non-redundant assembly information captured from the combination of Illumina and PacBio sequencing.

Identification of mobile genetic elements. Plasmid contigs were identified in each cheese metagenome by analyzing the mega-assemblies (contigs >5000 bp) with viralVerify (https://github.com/ablab/viralVerify, version 1.1) using the -p flag and Pfam database version Pfam33.1. Viruses were identified with VIBRANT(Kieft, Zhou, and Anantharaman 2020) (version 1.2.1). Numbers of predicted plasmid and virus contigs were plotted using R (version 4.0.2) (https://www.r-project.org/) and the readxl (version 1.3.1) and ggplot2 (version 3.3.5).

Assigning mobile genetic elements to hosts. To assign mobile genetic elements to hosts we used the viralAssociationPipeline.py script (https://github.com/njdbickhart/RumenLongReadASM/blob/master/viralAssociationPipeline.py)(D. M. Bickhart et al. 2021; Derek M. Bickhart et al. 2019). In brief, the contigs in the mega-assemblies for each cheese were classified taxonomically using Kraken2 and a custom Kraken database containing the default Kraken database supplemented with genomes of cheese-associated microbes. The outputs were reformatted in Excel to fit the input requirements for the viralAssociationPipeline.py script. metaHi-C reads were then aligned to the indexed mega-assemblies using bwa mem (Version: 0.7.17-r1188)(Li 2013) (bwa mem -v 1 -t 16 -5SP {mega-assembly} {forward_Hi-C} {reverse_Hi-C}). Reads that mapped to multiple locations in the assemblies were removed using grep -v -e 'XA:Z:' -e 'SA:Z:'. The resulting .sam files were then converted into .bam files using samtools view(Danecek et al. 2021). seqtk subseq (https://github.com/lh3/seqtk) was used to extract the predicted MGE contigs as a .fasta file. The lengths of the contigs in this file were determined using samtools faidx(Danecek et al. 2021) (version 1.9 using htslib 1.9). The viralAssociationPipeline was then run using the -a, -g, -b, -i, -v, -o, -s, -m, and -l flags. iTOL(Letunic and Bork 2021) was used to generate the diagram in Figure 3.2.4 showing the connections between MAGs (and isolate genomes in the case of cheese B) and MGEs (plasmids as predicted by viral verify -p and lytic phages as predicted by Vibrant). A summary graph showing the numbers of extrachomosomal MGEs associated with MAGs, with unbinned contigs as well as the numbers of unassociated extrachromosomal MGEs was plotted using R (version 4.0.2) (https://www.r-project.org/) and the readxl (version 1.3.1) and ggplot2 (version 3.3.5).

Psychrobacter pangenome analysis. To identify a non-redundant set of Psychrobacter genomes from our cheese isolate assemblies and full set of MAGs from the the three cheeses, genomes classified by GTDB-Tk (version 1.6.0, classify workflow)(Chaumeil et al. 2019) as belonging to the Psychrobacter genus were dereplicated with dRep (version 3.2.2)(Olm et al. 2017). CheckM (version 1.1.3)(Parks et al. 2015) results were provided as an Excel sheet. Genomes were selected by hand from the resulting MASH clusters by prioritizing isolates over circular individual timepoint MAGs over circular co-assembly MAGs over complete, non-circular individual timepoint MAGs over complete, non-circular co-assembly MAGs. This resulted in the selection of two MAGs (Cheese B MAG 21 and Cheese A MAG 15) and seven isolate genomes. An additional 97 publicly available Psychrobacter genomes were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly/) or from the JGI Integrated Microbial Genomes & Microbiomes(IMG/M) system (Supplemental Table 3.2.16). The full data set consisted of 17 genomes from cheese, eight genomes from other fermented foods, 35 host-associated genomes, ten genomes from soil, 19 genomes from marine environments, and 17 genomes from other miscellaneous environments. These 106 genomes were analyzed using the microbial pangenomics workflow in Anvi'o (Eren et al. 2021; Delmont and Eren 2018) (version 7.1 'hope'). Specifically, assemblies were run through gene prediction with Prodigal(Hyatt et al. 2010) (version 2.6.3), hits to bacterial single-copy gene collections were identified using HMMER (http://hmmer.org/), and genes were annotated with functions from the NCBI's Clusters of Orthologous Groups(Galperin et al. 2015). Identification of core and accessory gene sets and the construction of the phylogenetic tree based on single copy core

gene SNPs was done by panX(Ding, Baumdicker, and Neher, n.d.) (version 1.6.0) using FastTree 2(Price, Dehal, and Arkin 2010) and RaxML(Stamatakis 2014). The core gene set was defined here as present in at least 95 percent of genomes. Functional enrichment analysis was performed in Anvi'o with a comparison of genomes from cheese versus those not from cheese, a corrected q-value cutoff of 0.1, and COG20_FUNCTION as the annotation source(Shaiber et al. 2020). Regions associated with type VI secretion were extracted from cheese isolate genomes manually in Geneious Prime (www.geneious.com) and aligned and visualized using clinker and clustermap.js with default settings(Gilchrist and Chooi 2021). The separate pangenomic analysis of the 8 Psychrobacter genomes from cheese B was performed as described for the full data set, and the anvi-summarize function in Anvi'o was used to determine the core, shell, and cloud gene sets and associated COG functional categories. The summary chart was produced based on the Anvi'o output using Microsoft Excel for Mac Version 16.53.

***In vitro* washed-rind community reconstitution.** For the *in vitro* community reconstitution, 16 cheese isolates were chosen based on the metagenomic analyses. From this study, four *Psychrobacter* strains (CCS164, CCS169, CCS175 and CCS179), one *Halomonas* strain (CCS158), one *Pseudoalteromonas* strain (CCS176), one *Glutamicibacter* strain (CCS165), two *Staphylococcus* strains (CCS177, CCS178), one *Brachybacterium* strain (CCS183), one *Microbacterium* strain (CCS160) and two fungal strains (CCS145 - *Debaryomyces hansenii* and CCS187 - *Galactomyces geotrichum*) were chosen. In addition, three strains from a previous study were included in the community (JB5 - *Brevibacterium linens*, JB196 - *Vibrio*

casei and JB232 - *Hafnia ssp.*)(Wolfe et al. 2014). Strains of the 16-member community were inoculated at equal ratios (at 100,000 bacterial cells or fungal spores each) on 10% Cheese Curd Agar(Cosetta and Wolfe 2020). For each community and sampling conditions, triplicate communities were inoculated. Agar plates were incubated in the dark at 15°C in a humidified plastic bag. At 24, 48, 72, and 96 hours after inoculation, plates were scrubbed with a 20%wt NaCl solution using sterilized cotton swabs in a horizontal and vertical rastering pattern, followed by a rosette (Supplemental Figure 3.2.13). On days 3, 5, 7, and 21 microbial communities were collected into 1000µL of PBS+0.05% Tween using cell scrapers; for the day 3 sample, biomass collection was done before the brine wash. Half of each sample was split for various analyses: spot plating for CFU determination, glycerol stock preparation, and DNA extraction for metagenomic short-read sequencing in the case of days 7 and 21. Half of the scraped biomass was then replated onto the same petri dish. To calculate total bacterial and fungal CFUs, spot plating of serial dilutions of each sample was done on PCAMS, either with 50µg/L chloramphenicol to count the fungal community members or 100µg/L cycloheximide + 21.6µg/L natamycin to count the bacterial community members, with colony counting done 48 hours later (Supplemental Table 3.2.18). DNA was extracted from the *in vitro* community samples from days 7 and 21 using phenol-chloroform extraction (without a liquid nitrogen grinding step) and purified with 5 additional ethanol washes to remove residual phenol and chloroform. Library preparation and metagenomic sequencing of the multiplexed libraries were performed by Novogene (Sacramento, CA, USA) on a NovaSeq 6000 with a run configuration of 2 x 150 bp.

Metagenomic reads were then mapped back to the bacterial reference genomes using bwa mem to show relative abundance. For *Brevibacterium linens* (JB5, Accession number KF669529), *Vibrio casei* (JB196, Accession number Z_FUKS00000000), and *Hafnia alvei* (JB232, Accession number KF669544) previously published reference genomes were utilized. For the CCS strains the *de novo* assembled genomes from this study were utilized. 1,259,325,714 reads were sequenced, and over 99.7% of reads were aligned against the reference bacterial genomes (Supplemental Table 3.2.19). Anvi'o (version 7.1 'hope')(Eren et al. 2021) was used to visualize the metagenome and the coverage from short reads. The coverage values of the contigs were filtered for nucleotide positions that were within the interquartile range (25%-75%) of coverage values for each contig before being averaged. Across the contigs of each genome, these coverage values were also averaged, weighted by their length. To calculate the relative abundance for each community member, the coverage values were divided by the sum of all coverage values. The PCA plots of the coverages were prepared in R (version 4.0.2) (https://www.r-project.org/) and the tidyverse (version 1.3.1) and ggfortify (version 0.4.14)(Tang, Horikoshi, and Li 2016) packages. The principal component analysis was carried out using the prcomp function (center = TRUE, scale. = TRUE).

Data availability

Raw sequencing data. Fastq files of all metagenomic raw sequencing data are available on the Sequence Read Archive under Accession Number PRJNA778418. Subread files of the HiFi sequencing are available upon demand. Fastq files of all isolate sequencing data are available on the Sequence Read Archive under the Accession Numbers PRJNA837750 (CCS156),

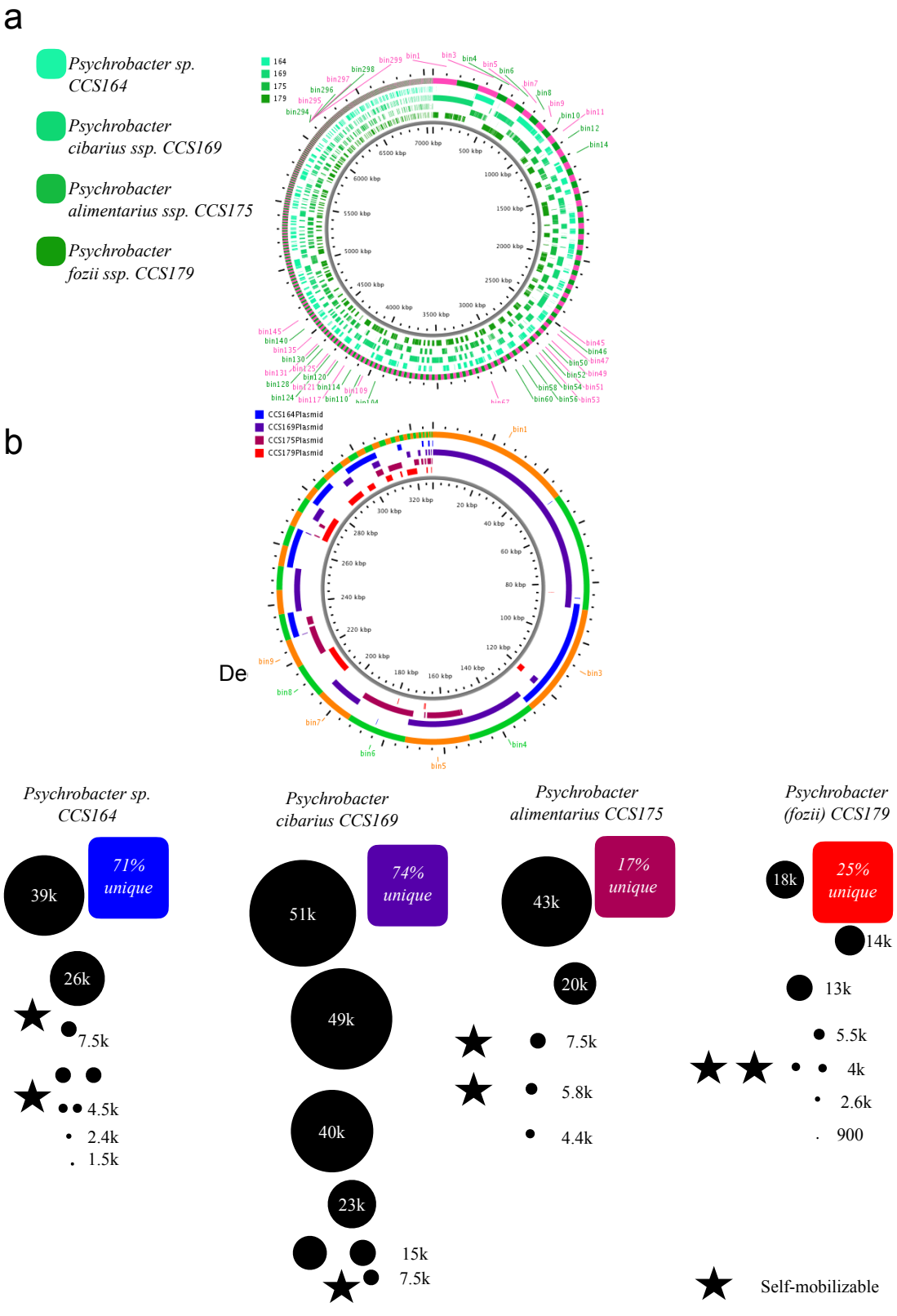
PRJNA837754 (CCS158), PRJNA837770 (CCS160), PRJNA837776 (CCS164), PRJNA838264 (CCS165), PRJNA837782 (CCS166), PRJNA837789 (CCS169), PRJNA838091 (CCS174), PRJNA838092 (CCS175), PRJNA838093 (CCS176), PRJNA838094 (CCS177), PRJNA838095 (CCS178), PRJNA838105 (CCS179), PRJNA838104 (CCS180), PRJNA838102 (CCS181), PRJNA838100 (CCS182), PRJNA838106 (CCS183), PRJNA838262 (CCS184), PRJNA838261 (CCS196). Fastq files of the in vitro community sequencing are available on the Sequence Read Archive under Accession Number PRJNA852571.

Assemblies, MAGs, supplemental files. Assemblies, MAGs and supplemental files have been deposited on Dryad (https://doi.org/10.5061/dryad.bg79cnpd8).

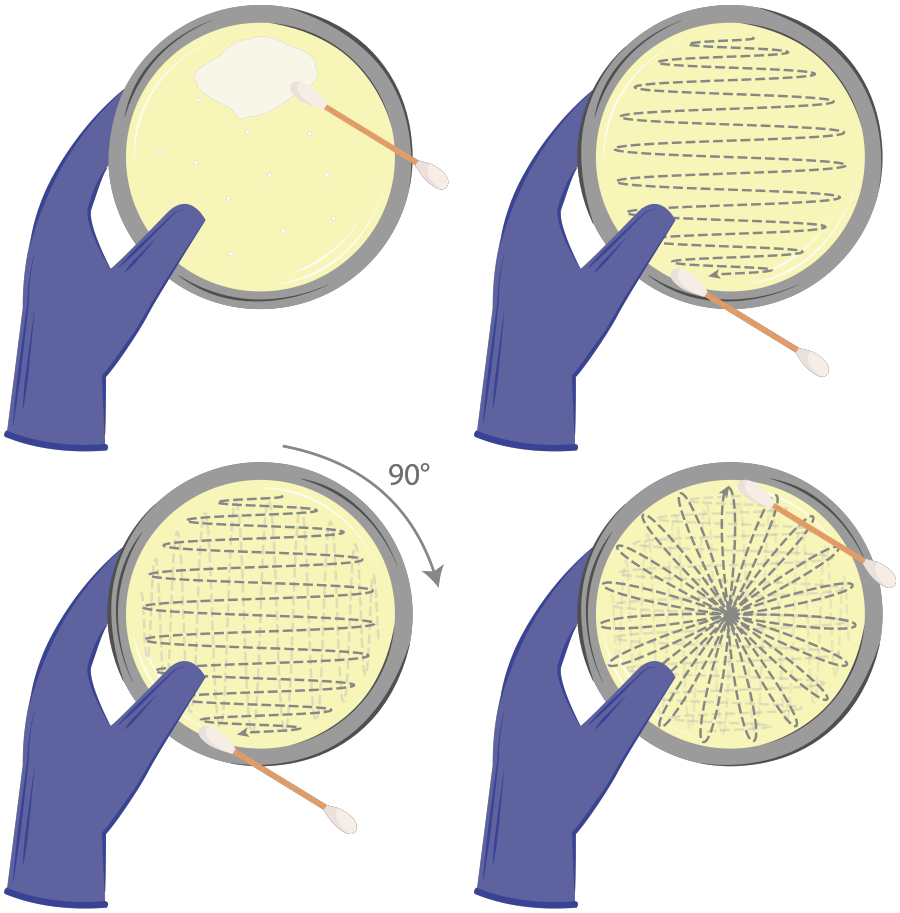
Acknowledgments

We thank Jasper Hill Farm for supplying the cheese samples, Pacific Biosciences for supplying SMRT sequencing reagents and project support as well as SNPSaurus and the Genomics & Cell Characterization Core Facility at the University of Oregon for providing SMRT sequencing services. In addition, we thank the Graduate Women in Science fellowship awarded to Christina C. Saak that allowed the generation of metaHi-C data and the IGM Genomics Center at the University of California San Diego for a NovaSeq sequencing grant to Emily C. Pierce. We also thank Steven Villareal for help with DNA extractions from Cheese B isolates for Nanopore sequencing, Mathieu Almeida for advice on the bioinformatic analyses and Brooke Johnson for help with exploratory data analysis.

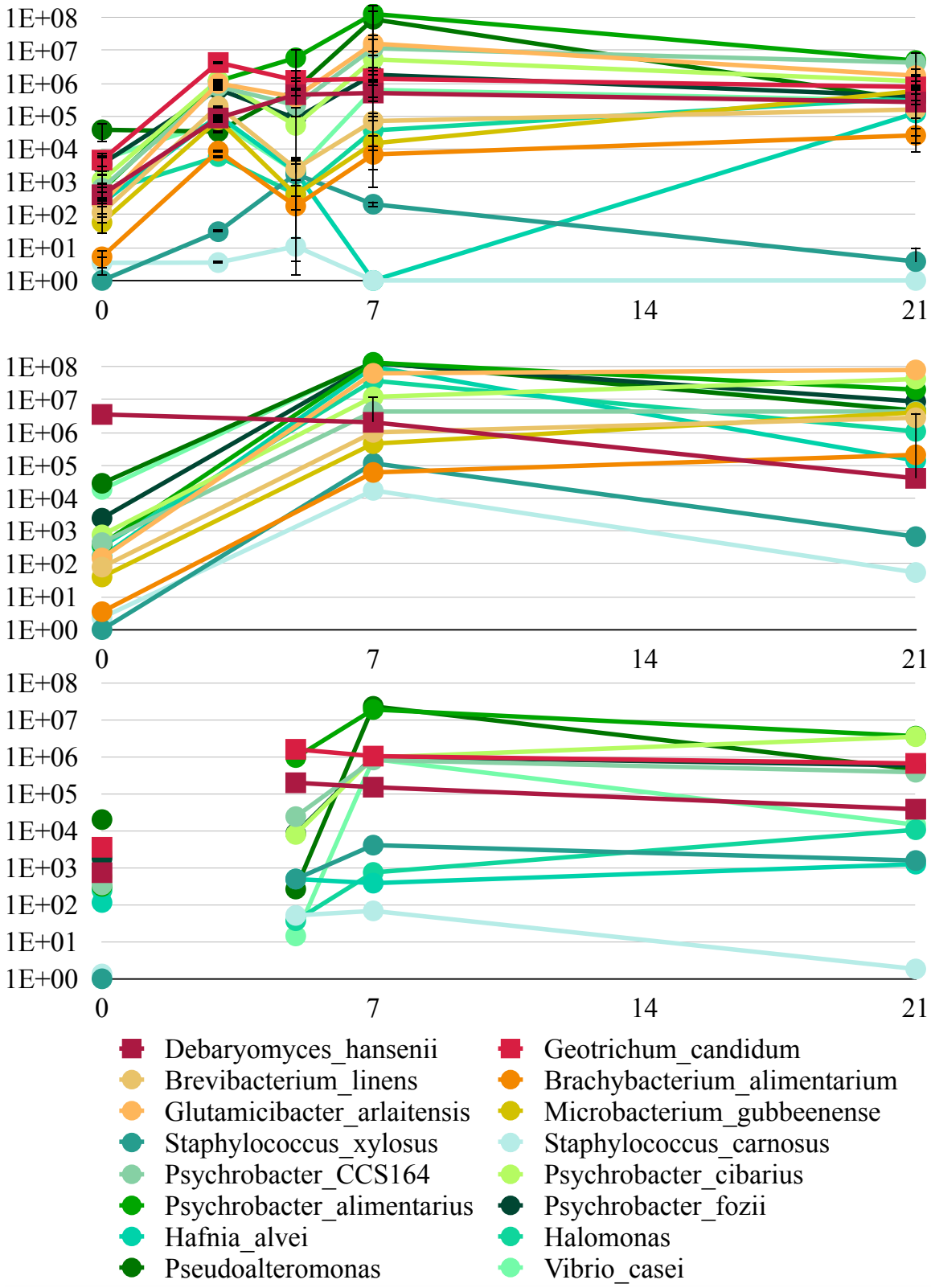
Supplemental Figures follow.



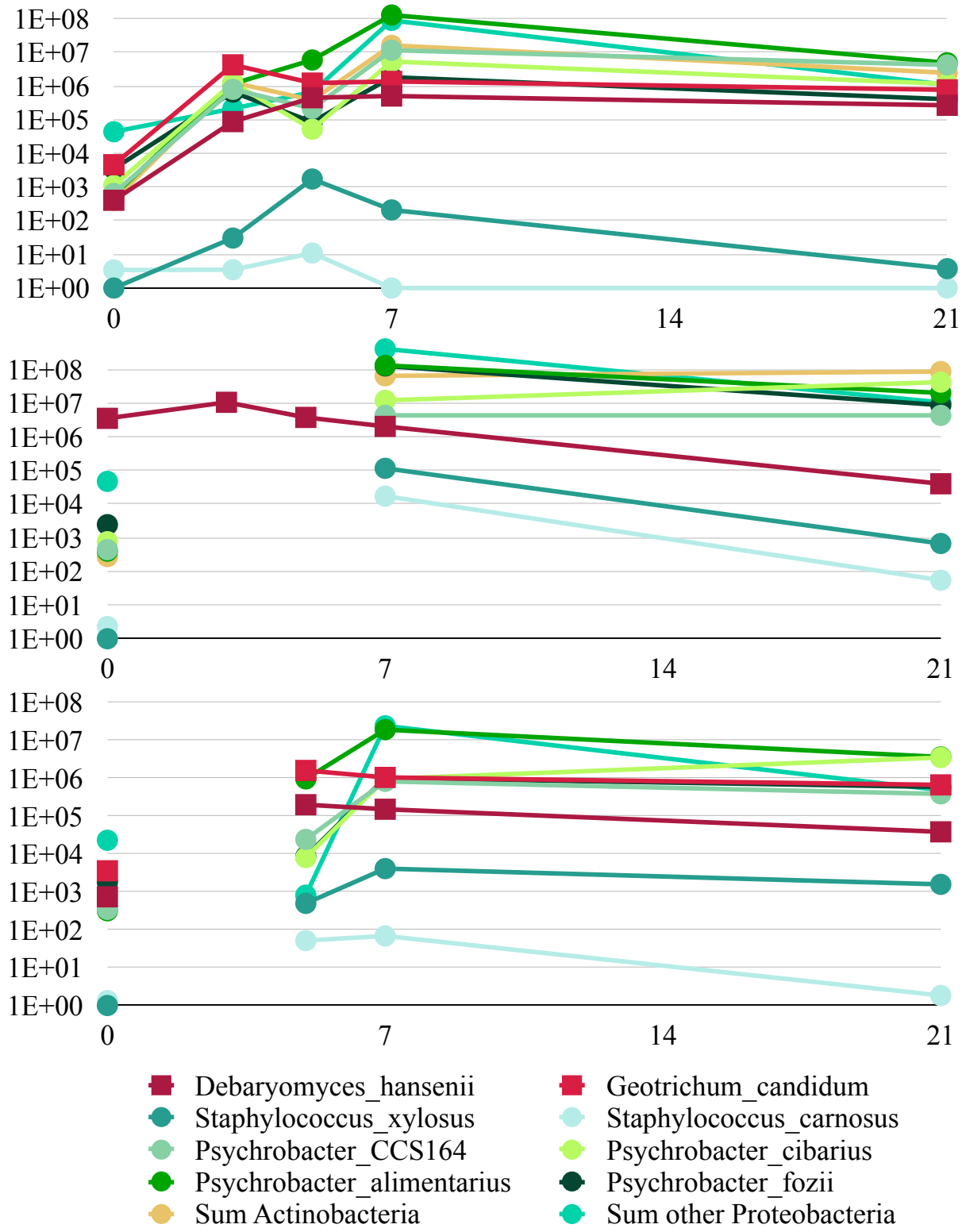
Supplemental Figure 3.1.1 Similarity of *Psychrobacterial* chromosomes and plasmids



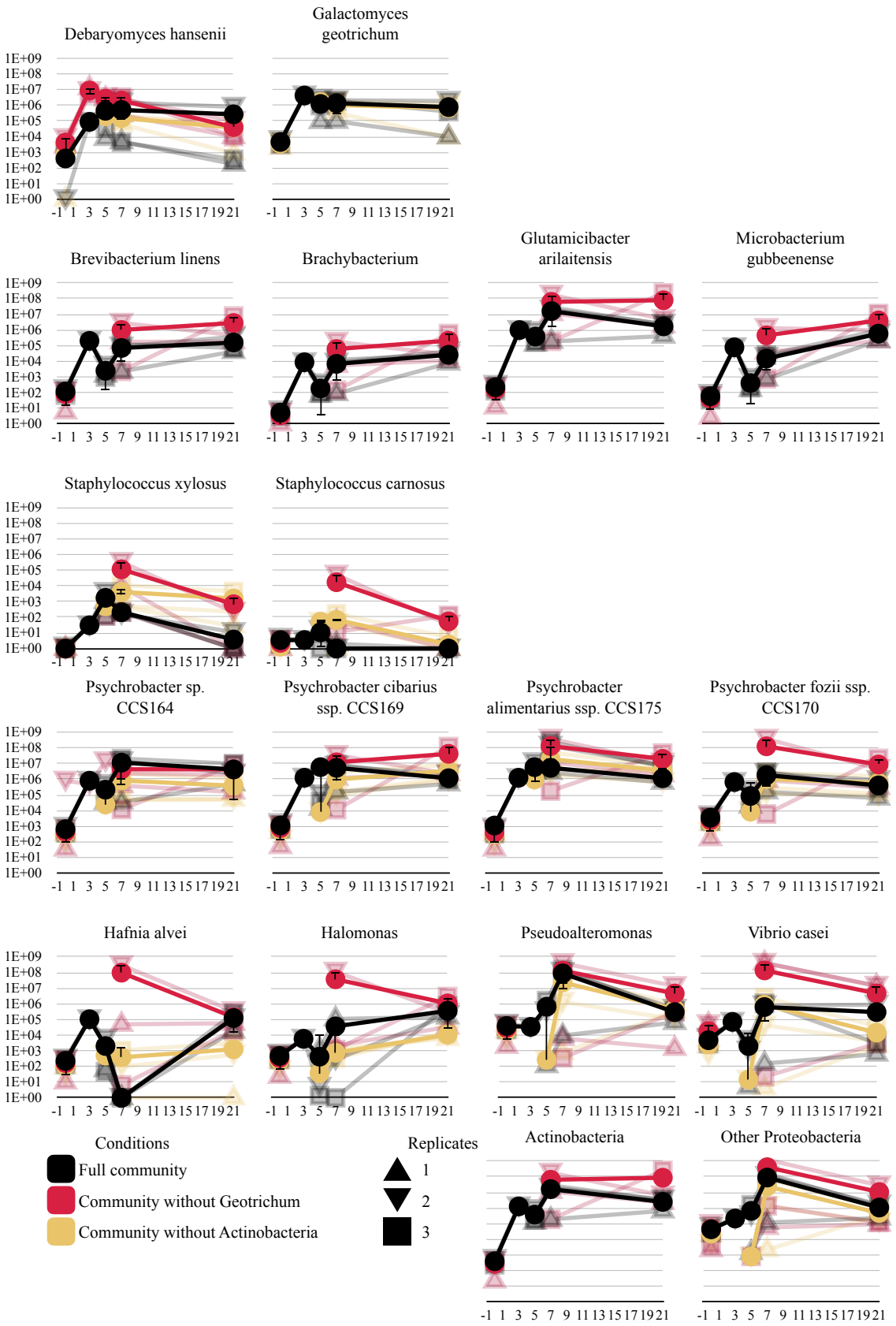
Supplemental Figure 3.1.2. Washing scheme done on days 1-4 of in vitro washed rind community.



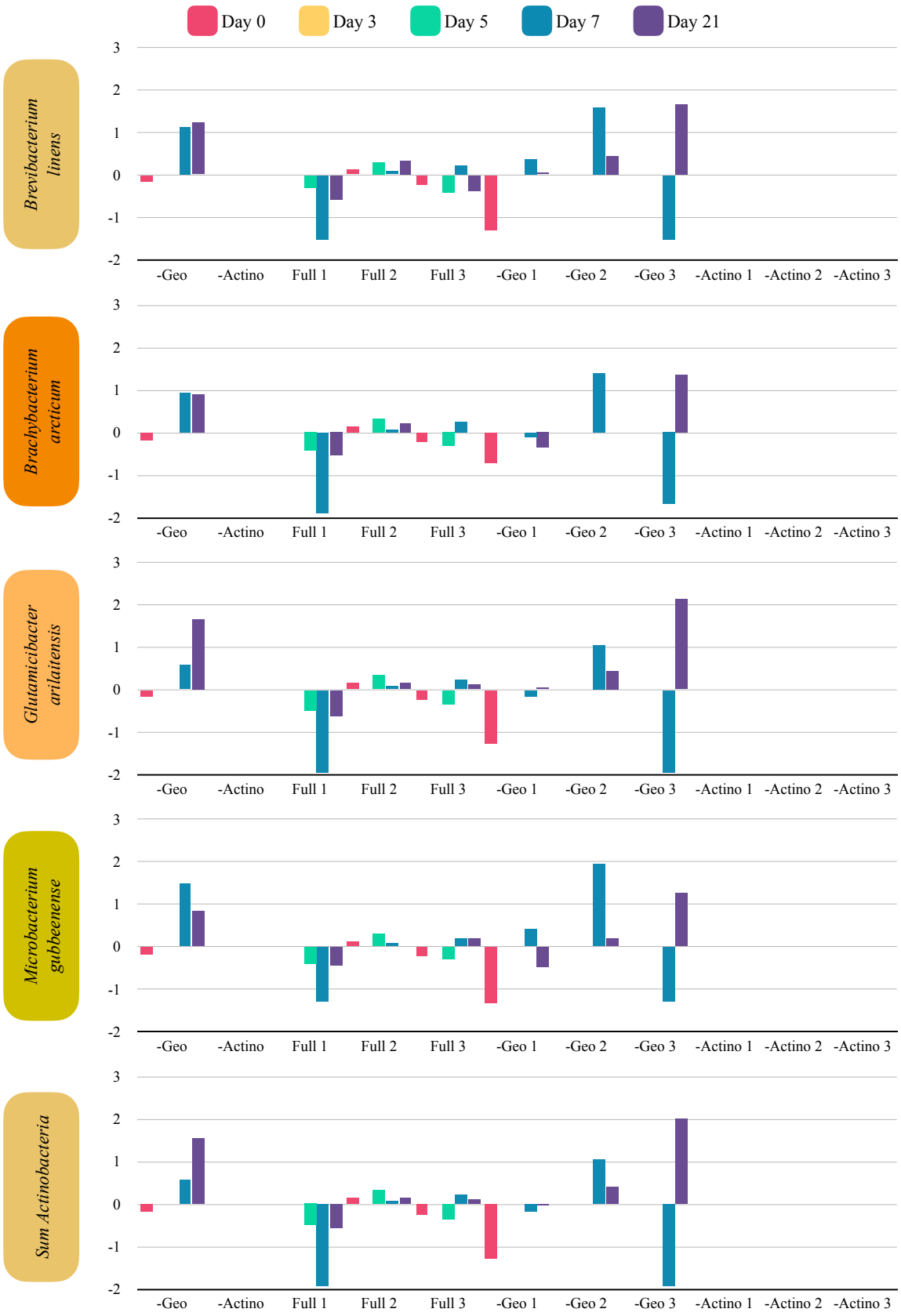
Supplemental Figure 3.1.3 Average absolute abundance within conditions.
Top: Full community. Middle: without *Geotrichum*. Bottom: without Actinobacteria



Supplemental Figure 3.1.4 Average absolute abundance within conditions, simplified
Top: Full community. Middle: without *Geotrichum*. Bottom: without Actinobacteria



Supplemental Figure 3.1.5 Absolute abundance by species

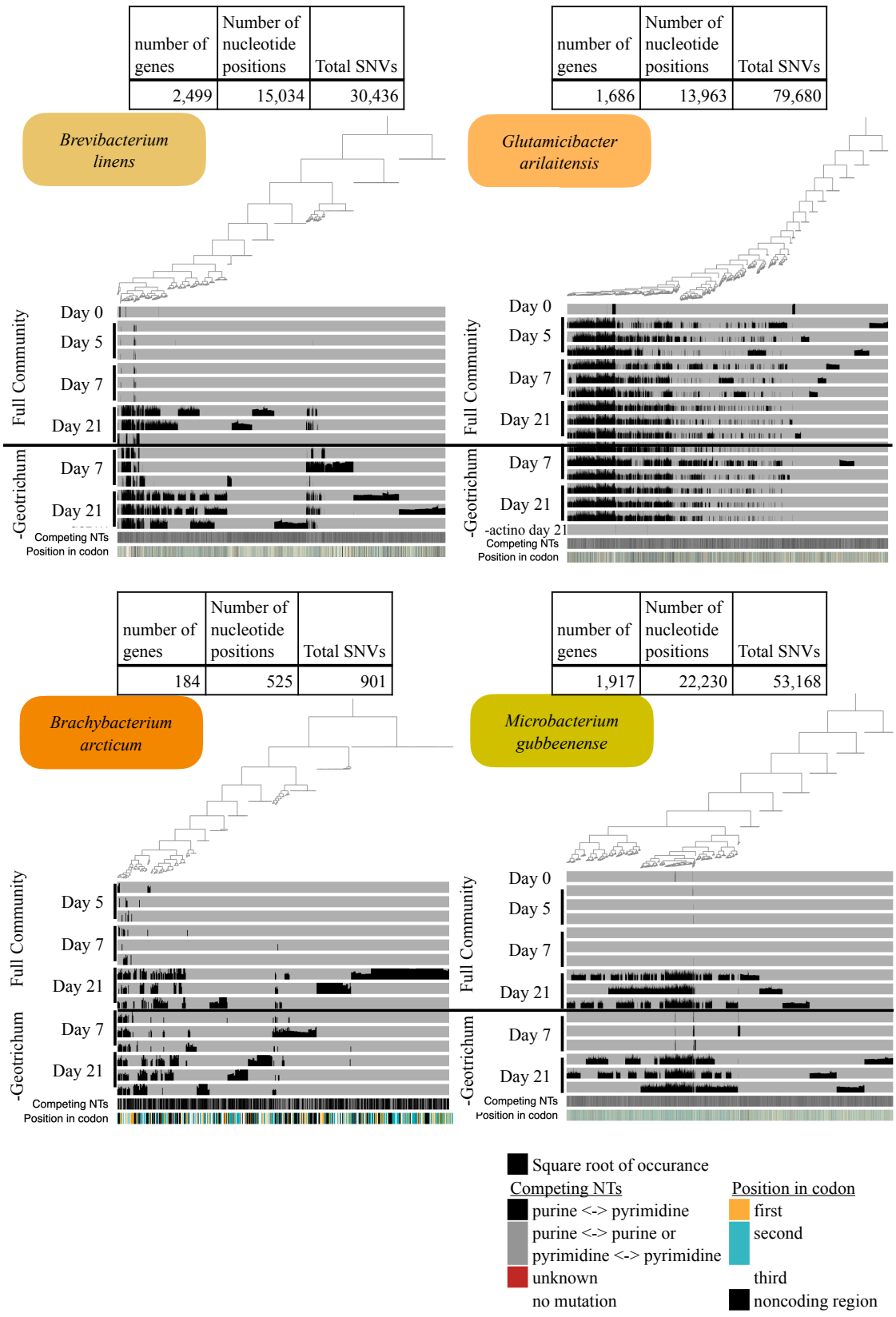


Supplemental Figure 3.1.6 Relative abundance to average abundance. Actinobacteria.

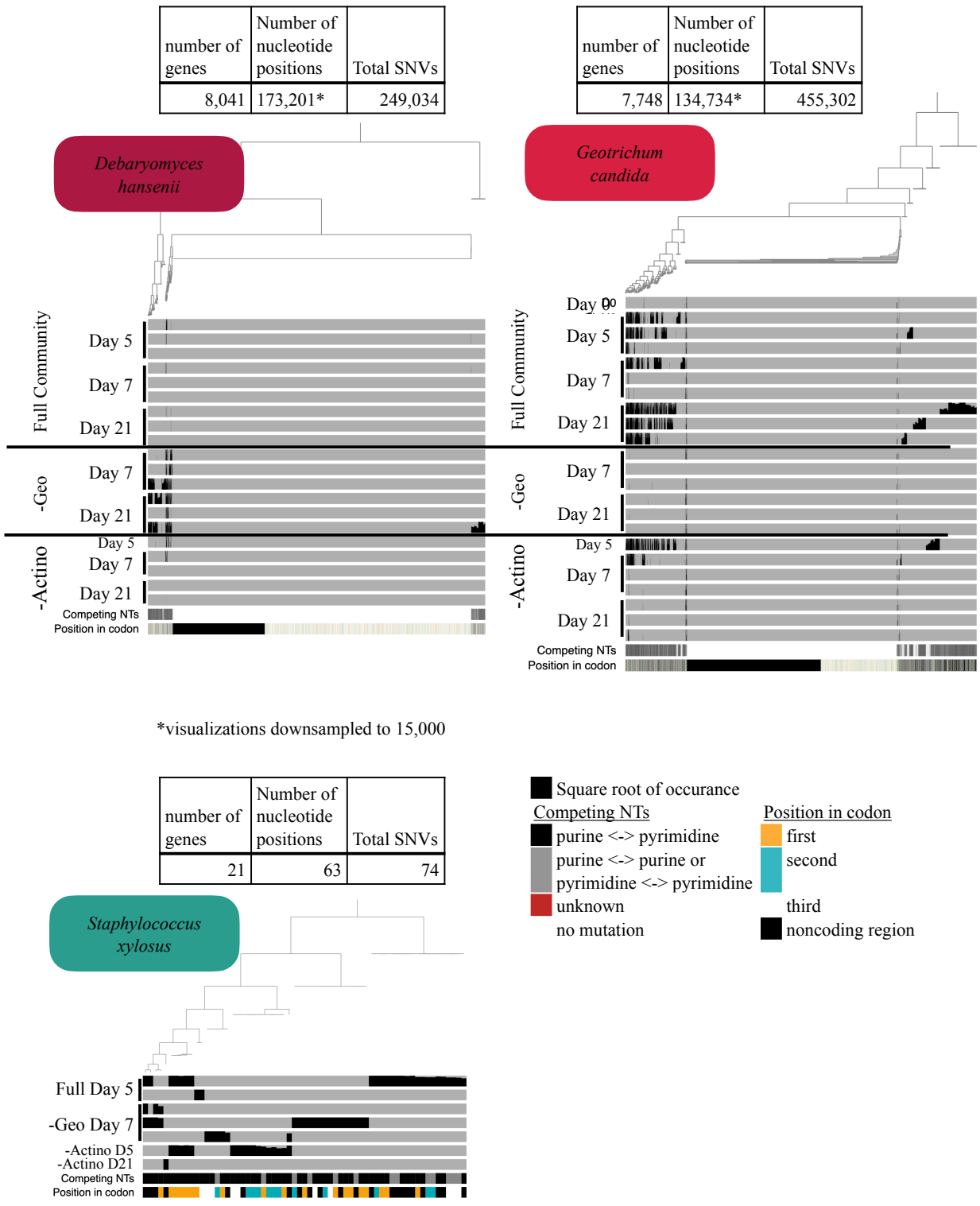
Supplemental Figure 3.1.6 continued Relative abundance to average abundance in the full community over time. Fungi and *Staphylococci*.

Supplemental Figure 3.1.6 continued Relative abundance to average abundance in the full community over time. *Psychrobacter*.

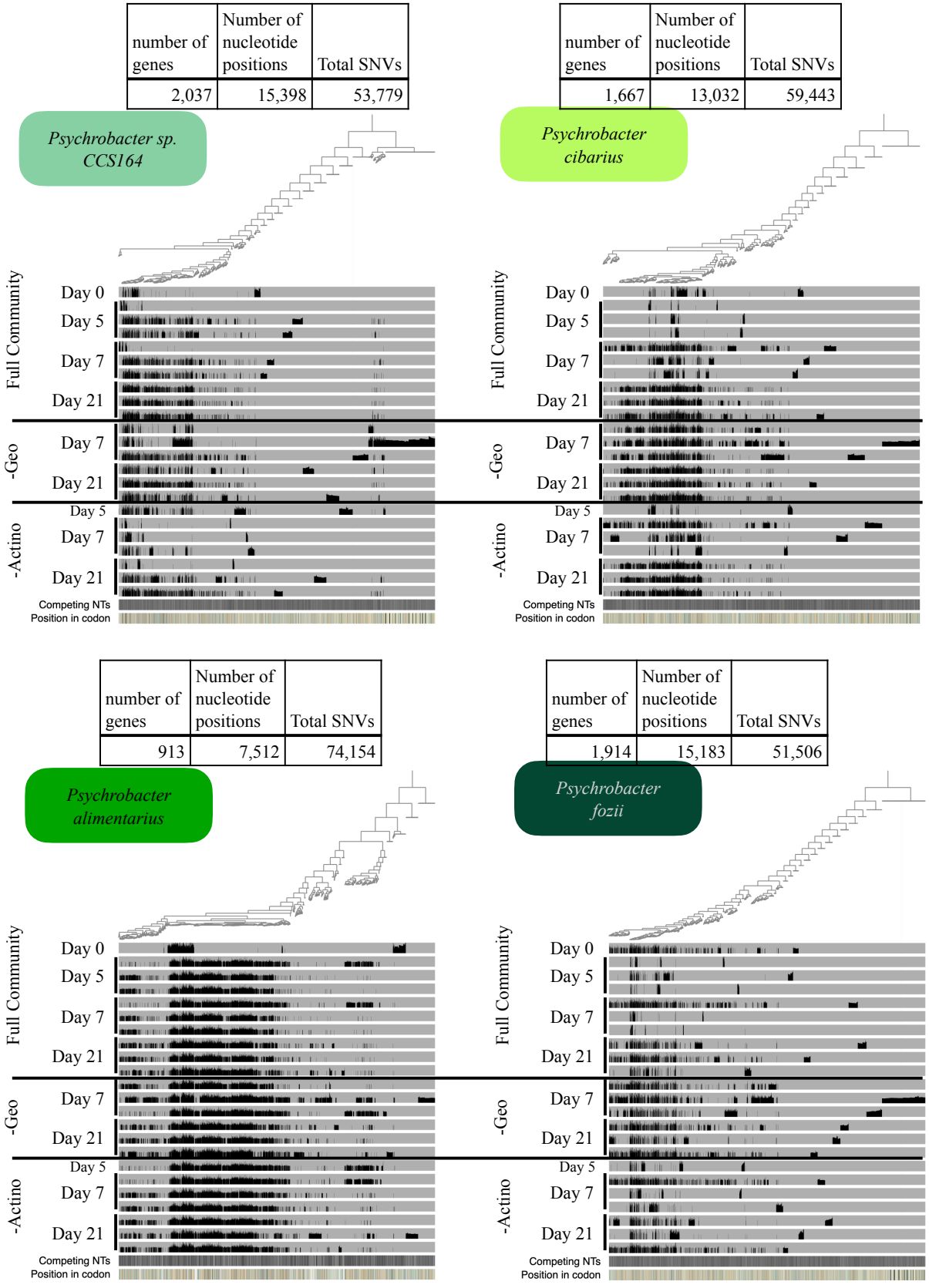
Supplemental Figure 3.1.6 continued Relative abundance to average abundance in the full community over time. non-*Psychrobacter* Proteobacteria.



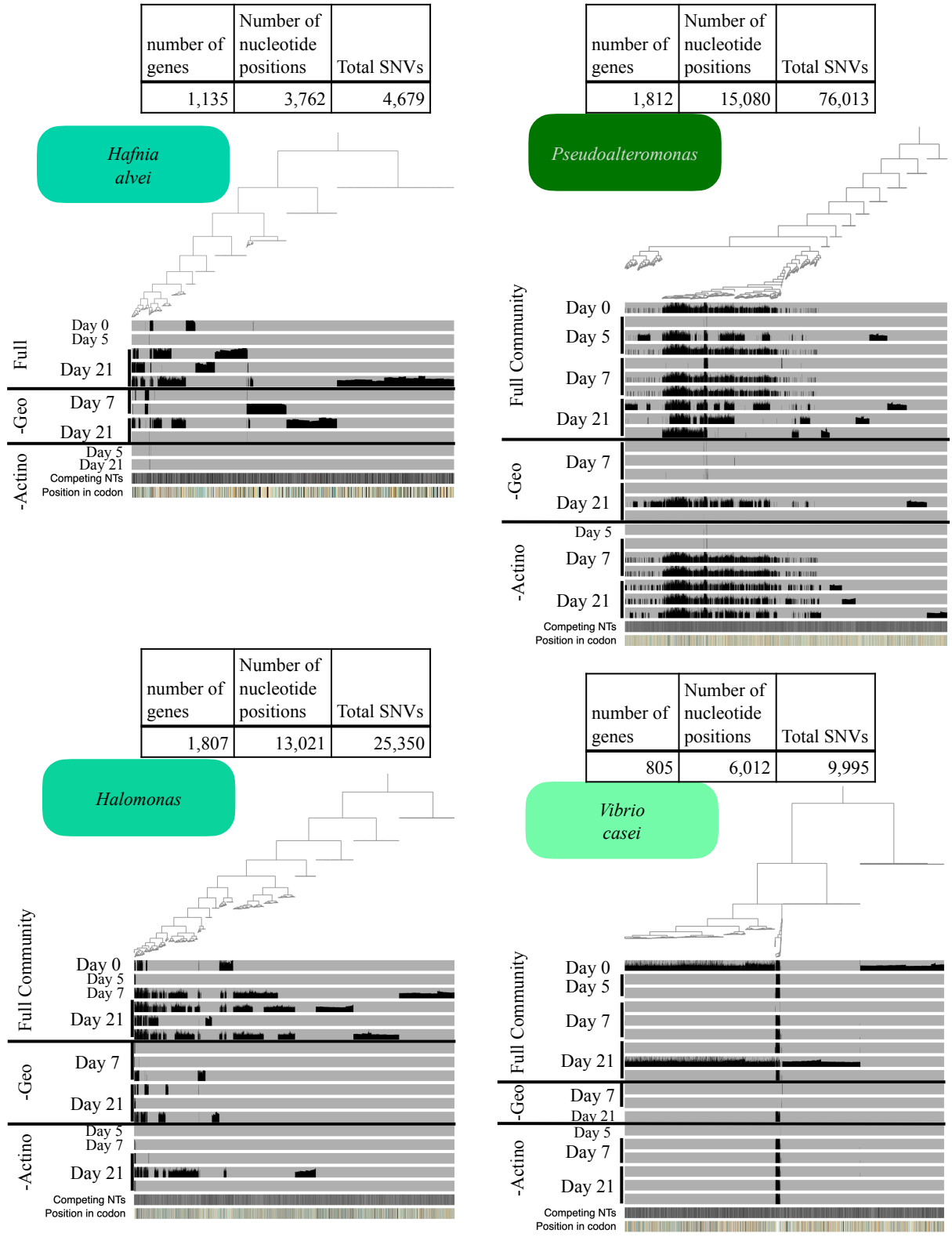
Supplemental Figure 3.1.7 Single nucleotide variant anvi'o graphs. Actinobacteria.



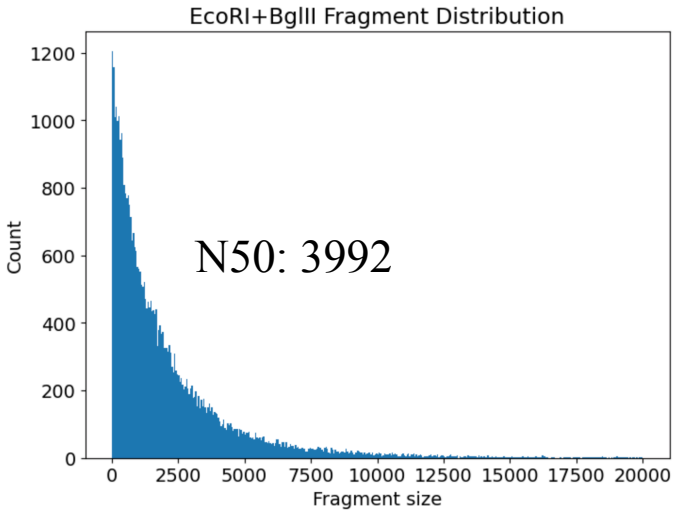
Supplemental Figure 3.1.7 Single nucleotide variant anvi'o graphs. Fungi and *Staph. xylosus*.

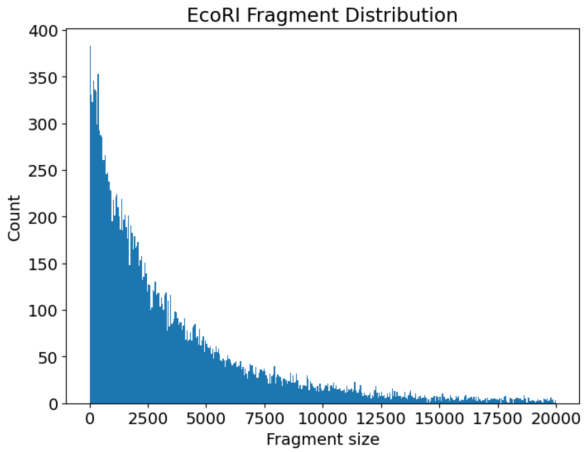


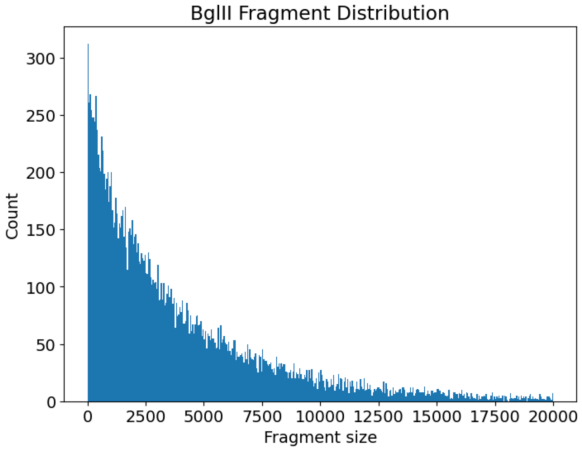
Supplemental Figure 3.1.7 Single nucleotide variant anvi'o graphs. *Psychrobacter*.



Supplemental Figure 3.1.7 Single nucleotide variant anvi'o graphs. non-*Psychrobacter* Proteobacteria.



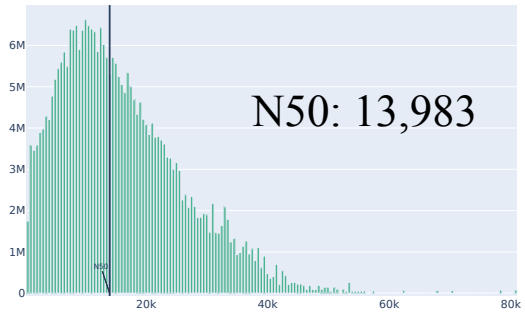


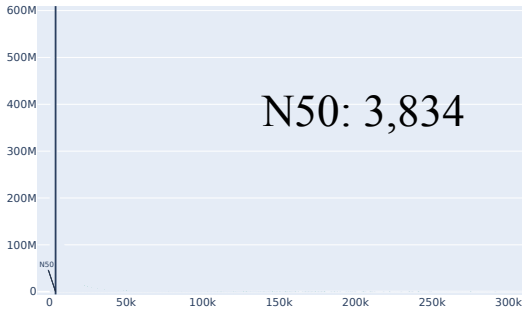


	Count	Maximum	Mean	% 2kb-5kb	% 1kb-8kb	N50
EcoRI+BglII	40035	55487	2188.46	25.24%	53.67%	3992
EcoRI	20888	99530	4056.48	28.93%	59.18%	7739
BglII	18871	86388	4544.2	29.68%	59.98%	8229

Supplemental Figure 3.1.8 in silico endonuclease analysis

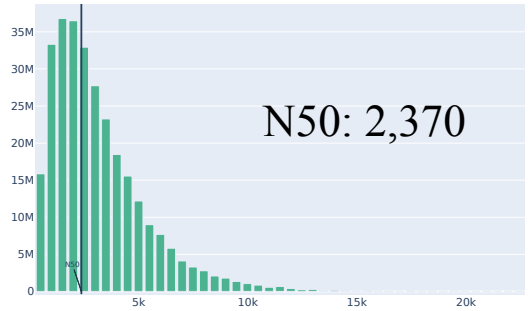
Concatemers

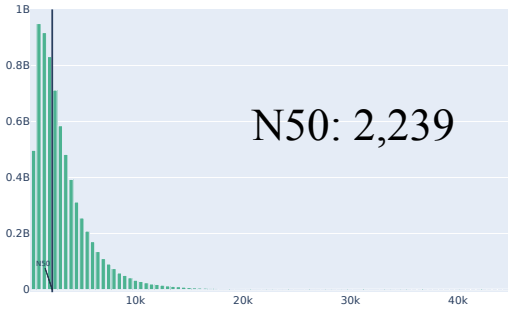




Day 3 **Day 21**

Cut Reads



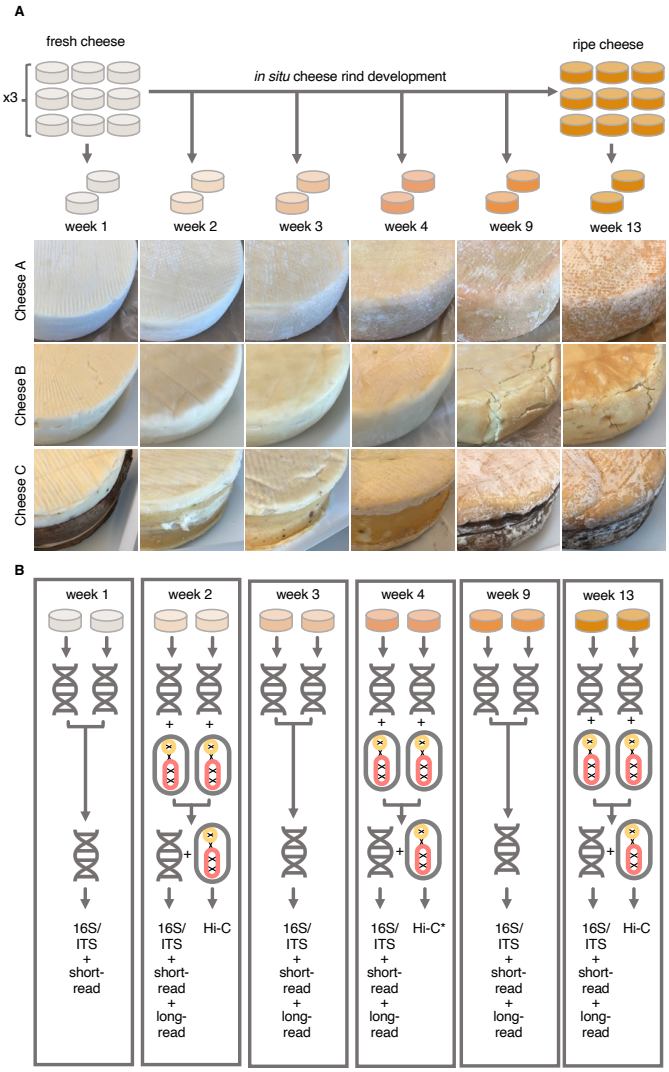


Day 3 **Day 21**

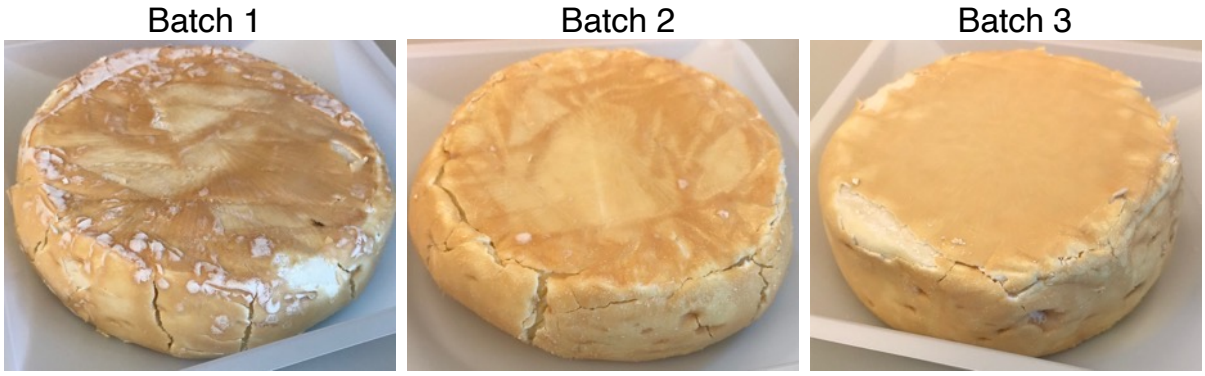
Restriction Endonuclease	Cut Sites
EcoRI	187,568
BglII	111,730

Restriction Endonuclease	Cut Sites
EcoRI	2,291,016
BglII	1,667,675

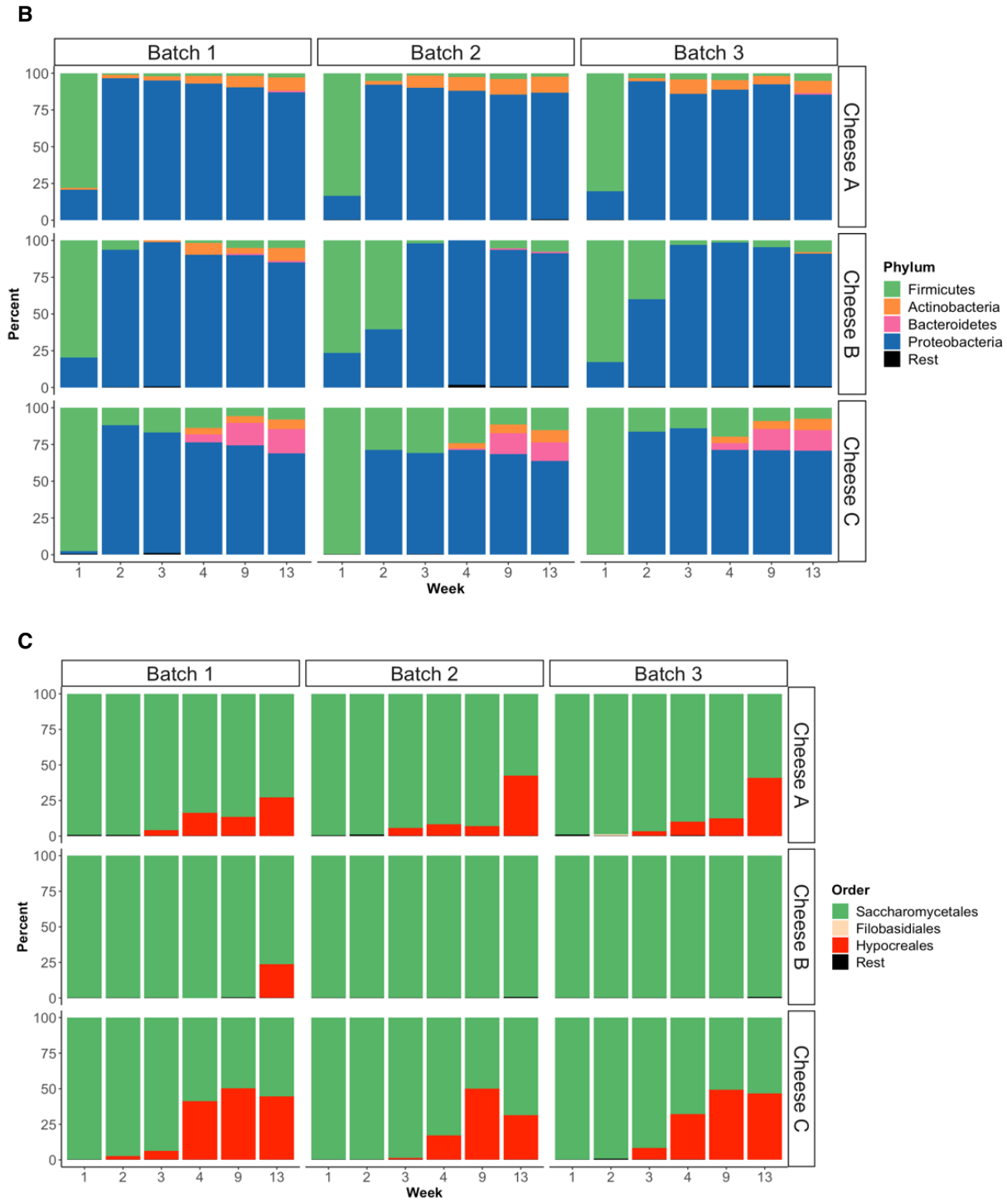
Supplemental Figure 3.1.9 Raw and digested long read sequencing summary



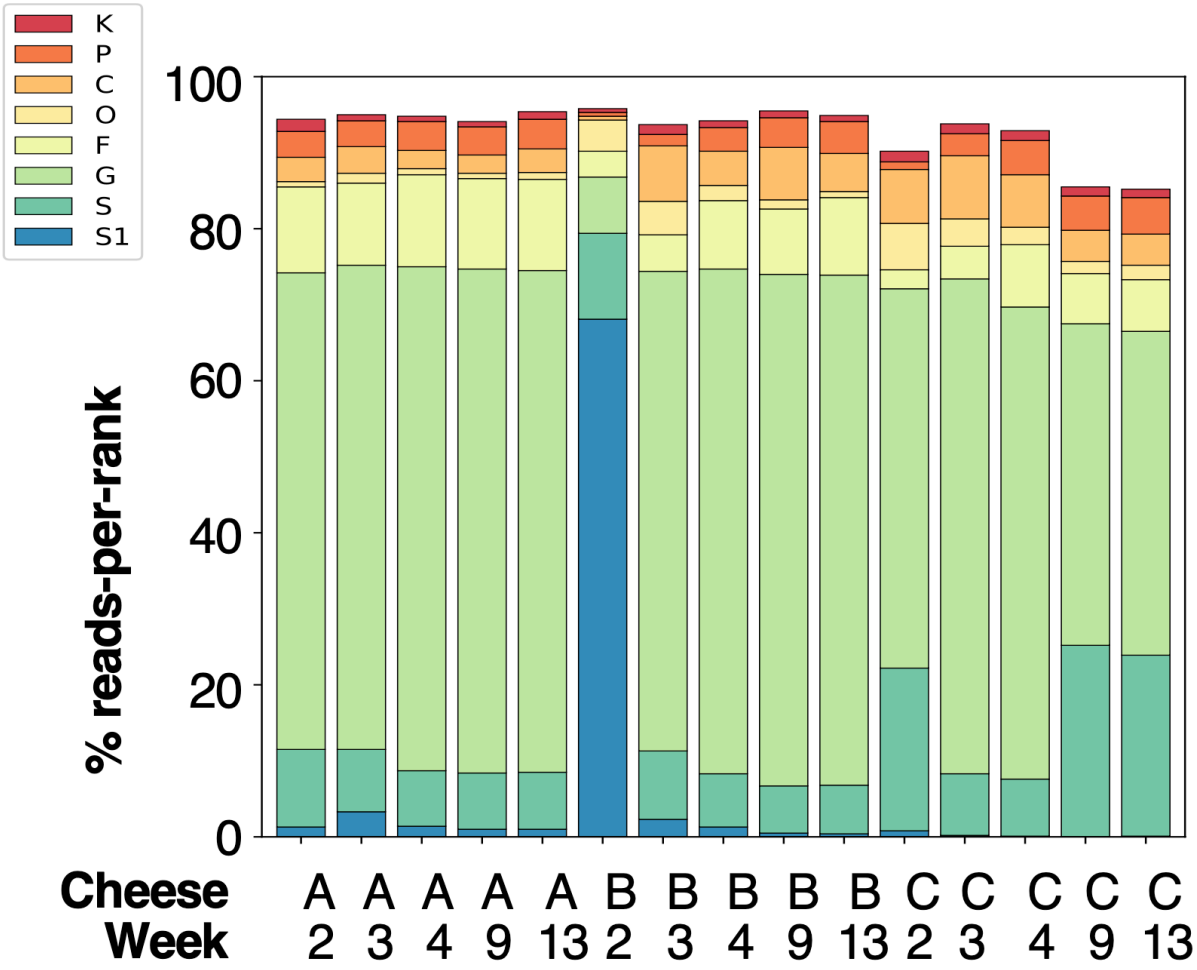
Supplemental Figure 3.2.1. Sampling (A) and metagenomic sequencing (B) of three ripening washed-rind cheeses from the same facility (A) For each of the three different washed-rind cheeses (A, B, and C) we followed the aging of three different batches produced one week apart. From each batch, we collected rind from duplicate wheels at six timepoints. A detailed collection schedule, including information and wash frequency and storage, can be found in Supplemental Table 3.2.1. Representative images of each of the three cheeses at different timepoints are shown. Cheese C is wrapped in spruce during ripening. In the pictures for weeks 2-4 the spruce has been removed. (B) DNA was extracted from each rind sample collected and DNA from duplicate wheels was then combined for downstream sequencing. All three batches of each cheese were analyzed at all of the six timepoints with 16S and ITS amplicon sequencing to estimate reproducibility of bacterial and fungal succession dynamics, respectively. One batch (batch 3) of each cheese was analyzed at all of the six timepoints with short-read shotgun sequencing. The same batch of each cheese was analyzed at weeks 2, 3, 4, 9 and 13 using long-read shotgun sequencing. Finally, rind samples of that batch from weeks 2 and 13 were fixed, combined and subjected to Hi-C sequencing. *Only cheeses B and C were analyzed by Hi-C at week 4.



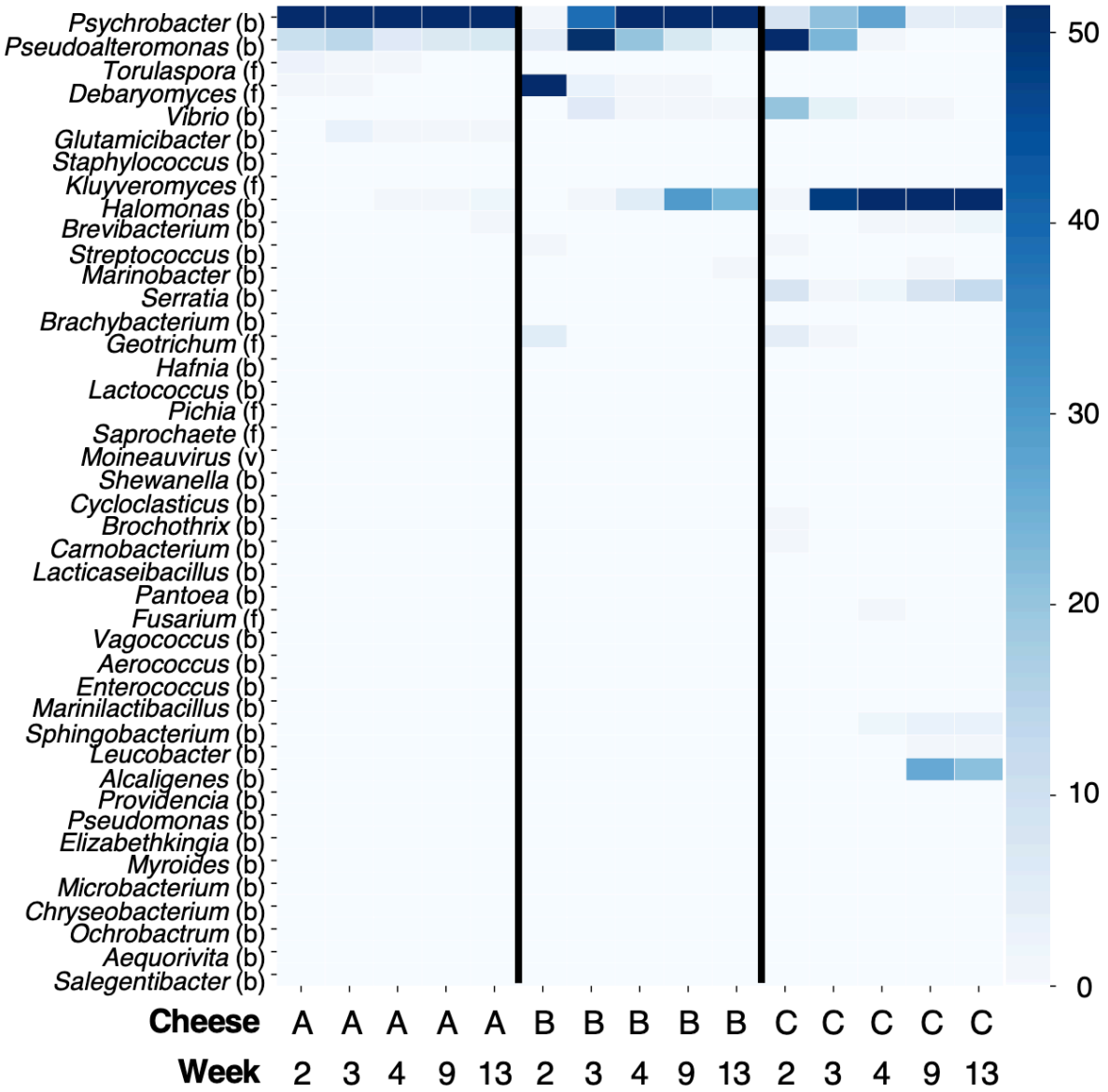
Supplemental Figure 3.2.2. Cheese wheels from batches 1, 2 and 3 of Cheese B at week 13.



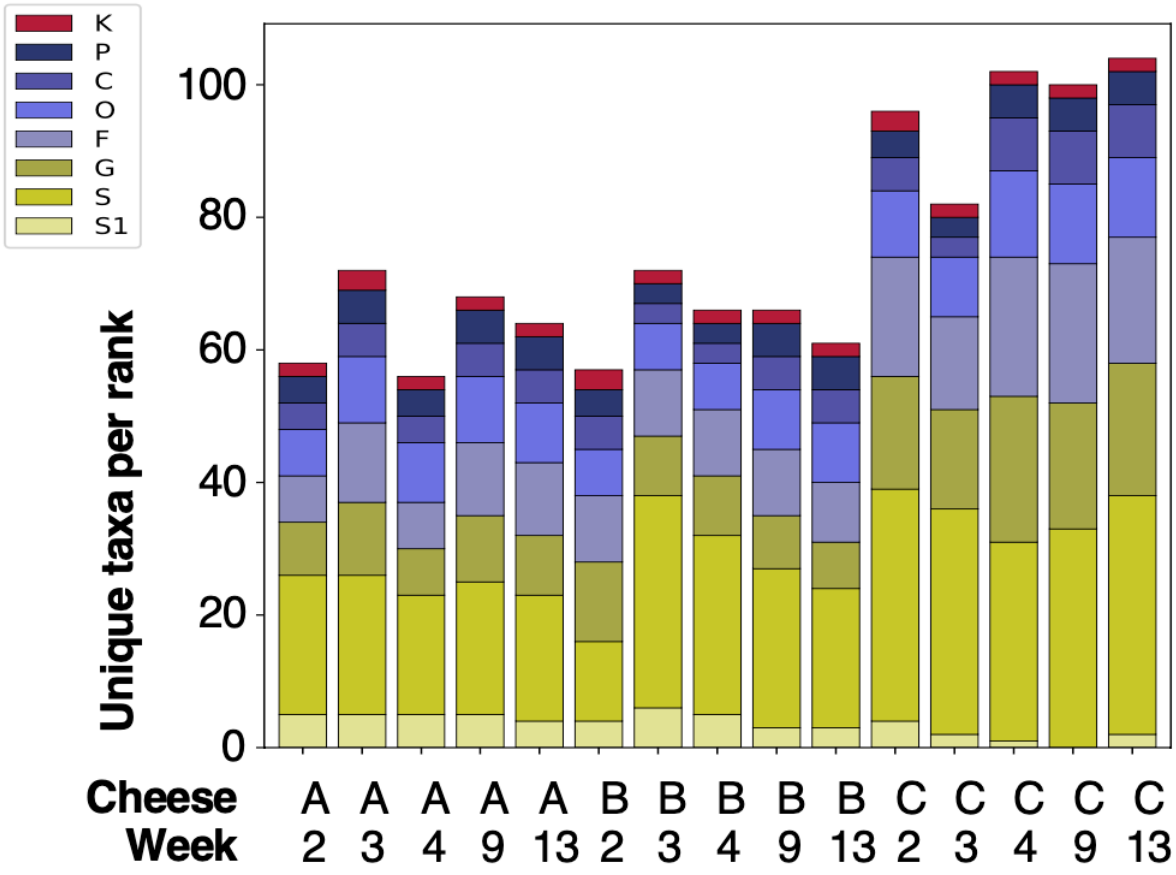
Supplemental Figure 3.2.3. *In situ* succession patterns of bacterial and fungal washed-rind cheese communities at Phylum- and Order-level, respectively. Relative abundance plots of (A) bacteria as determined by 16S amplicon sequencing and (B) fungi as determined by ITS sequencing. Shown are the relative abundances of amplicon sequence variants collapsed at the phylum-level (A) or order-level (B). Rest = Taxa with <1% of the classified reads.



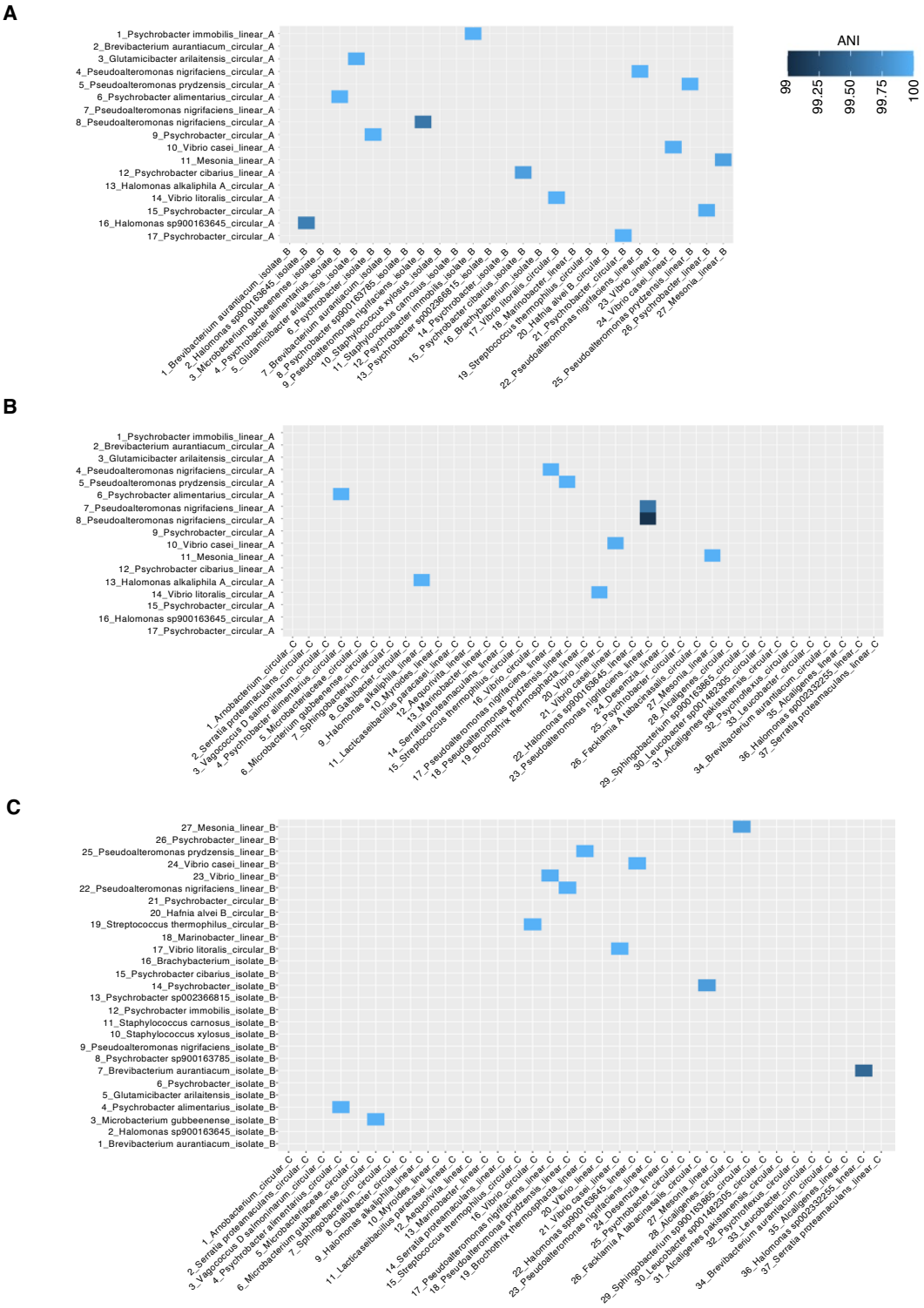
Supplemental Figure 3.2.4. The majority of long reads were classified to at least Genus level. HiFi reads were used as inputs for the "Taxonomic-Profiling-Nucleotide" pipeline followed by the "MEGAN-RMA" summary. The results were visualized with the "compare-kreport-taxonomic-profiles" tool. All pipelines are part of the "PB-metagenomics-tools" toolkit (https://github.com/PacificBiosciences/pb-metagenomics-tools). Axes labels were customized in Microsoft® PowerPoint for Mac (version 16.59).



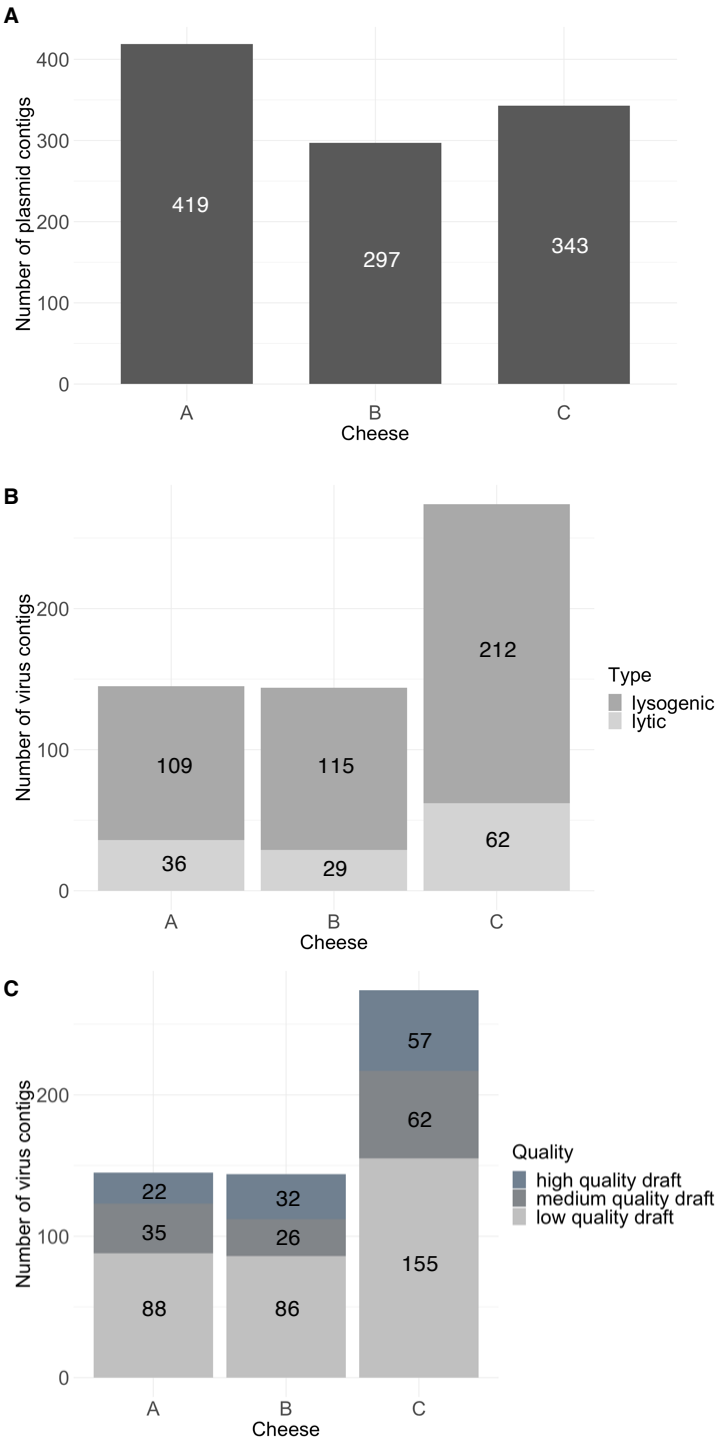
Supplemental Figure 3.2.5. Heatmap of long-read-based relative abundance estimation aggregated at Genus level. Bacterial taxa are indicated with a (b), fungal taxa are indicated with a (f) and viral taxa are indicated with a (v). HiFi reads were used as inputs for the "Taxonomic-Profiling-Nucleotide" pipeline followed by the "MEGAN-RMA" summary. The results were visualized with the "compare-kreport-taxonomic-profiles" tool. All pipelines are part of the "PB-metagenomics-tools" toolkit (https://github.com/PacificBiosciences/pb-metagenomics-tools). Axes labels were customized in Microsoft® PowerPoint for Mac (version 16.59).



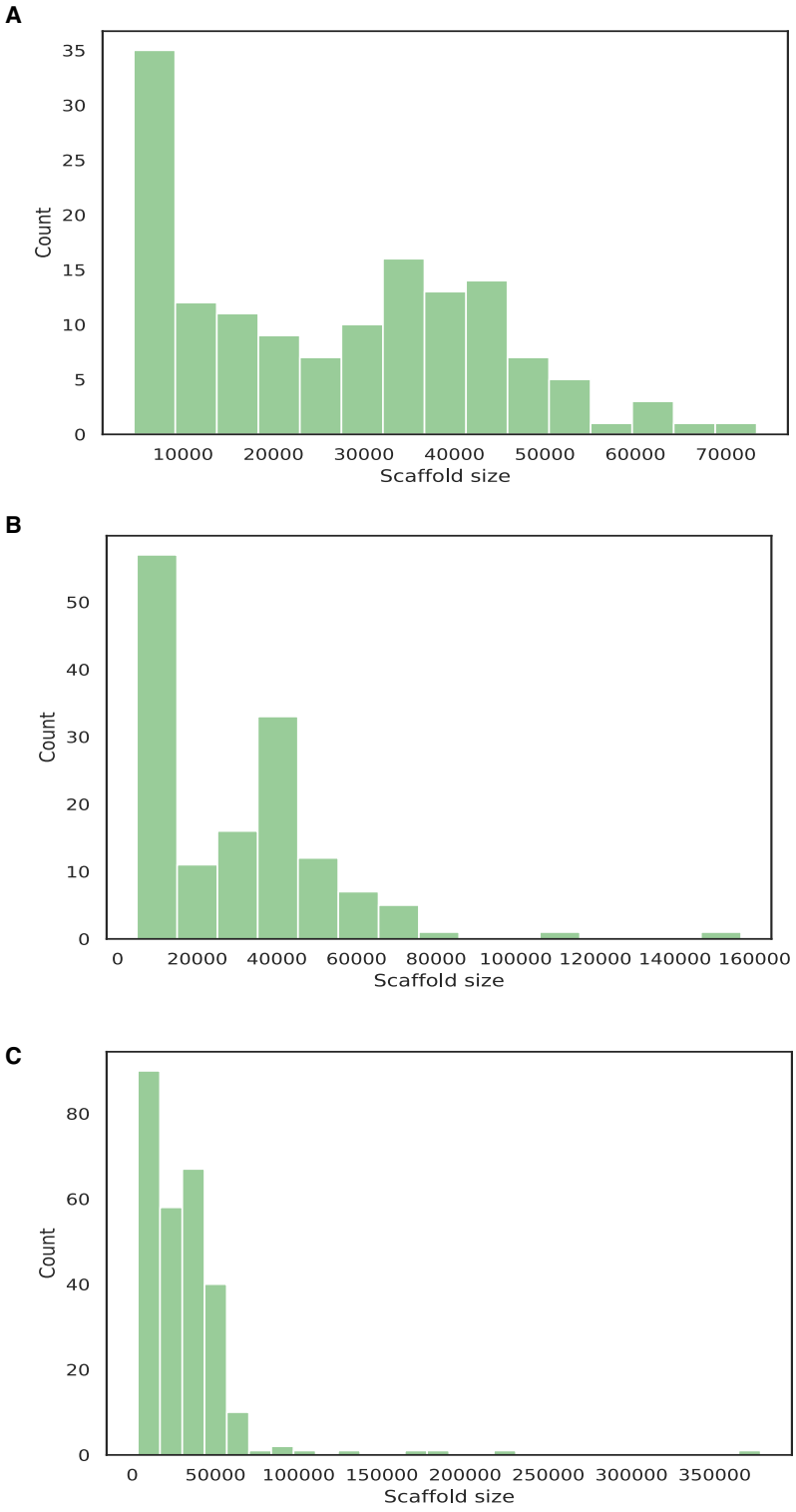
Supplemental Figure 3.2.6. Cheese C contained the largest number of unique taxa per rank at all sampled timepoints. HiFi reads were used as inputs for the "Taxonomic-Profiling-Nucleotide" pipeline followed by the "MEGAN-RMA" summary. The results were visualized with the "compare-kreport-taxonomic-profiles" tool. All pipelines are part of the "PB-metagenomics-tools" toolkit (https://github.com/PacificBiosciences/pb-metagenomics-tools). Axes labels were customized in Microsoft® PowerPoint for Mac (version 16.59).



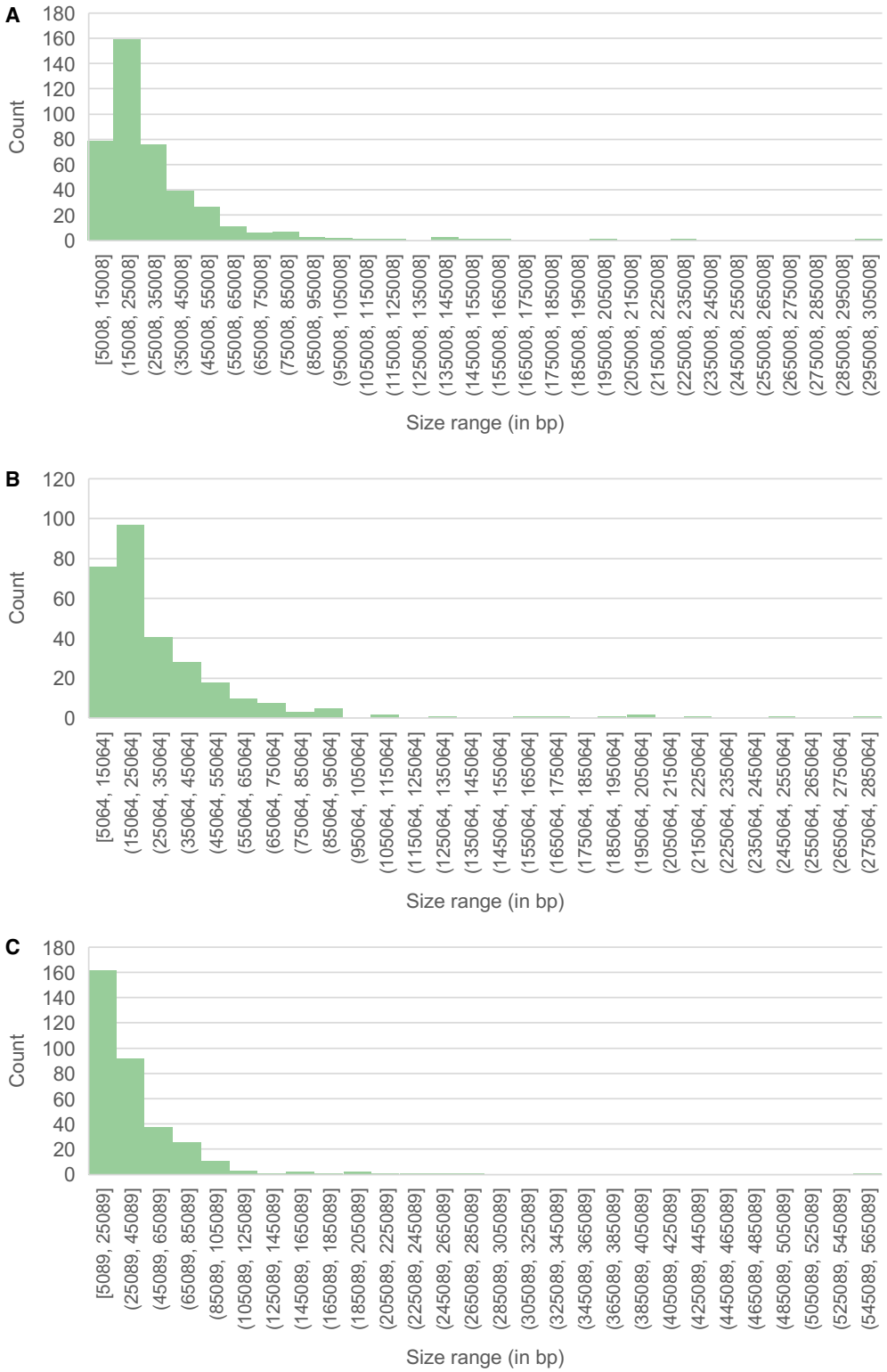
Supplemental Figure 3.2.7. ANI Comparison of high-quality MAGs (and isolate genomes) from Cheese A, B and C. (A) Comparison of high-quality MAGs (and isolate genomes) recovered from cheeses A and B. (B) Comparison of high-quality MAGs recovered from cheeses A and C. (C) Comparison of high-quality MAGs (and isolate genomes) from cheeses B and C.



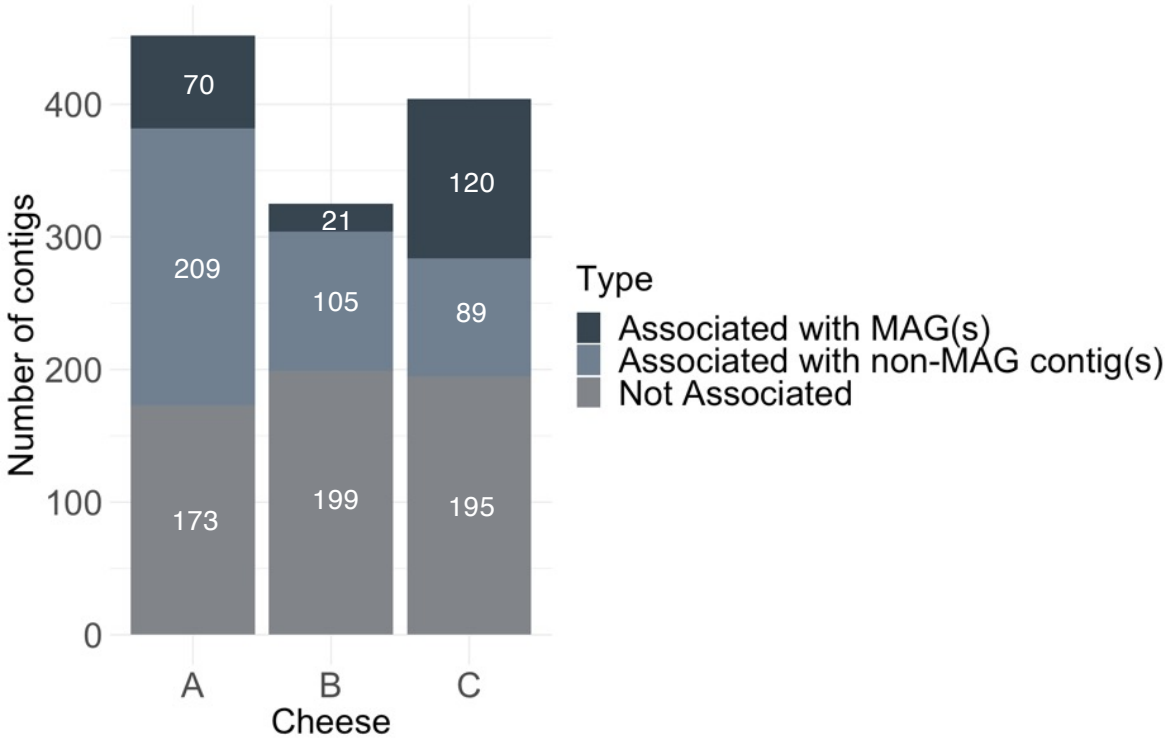
Supplemental Figure 3.2.8. Plasmids and viruses predicted in the mega-assemblies of cheeses A, B and C. (A) Number of contigs predicted to be plasmids by ViralVerify (with the -p flag). (B) Number of contigs predicted to be lytic and lysogenic viruses by VIBRANT. (C) Number of high-quality, medium-quality and low-quality viral contigs as determined by VIBRANT. Of the predicted viral contigs 4, 4 and 7 are predicted to be circular for Cheeses A, B and C, respectively. Full MGE prediction results can be found in Supplemental Table 3.2.13.



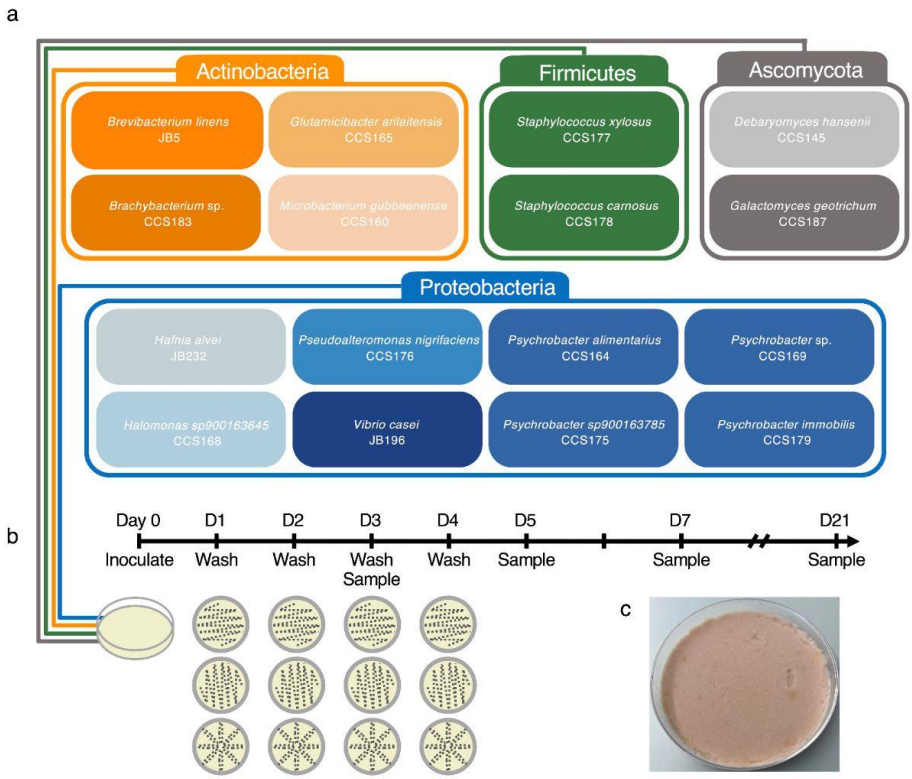
Supplemental Figure 3.2.9. Size distributions of predicted viral contigs. Size distributions of the predicted viral contigs are shown for (A) cheese A, (B) cheese B and (C) cheese C. Graphs are from the standard VIBRANT output files.



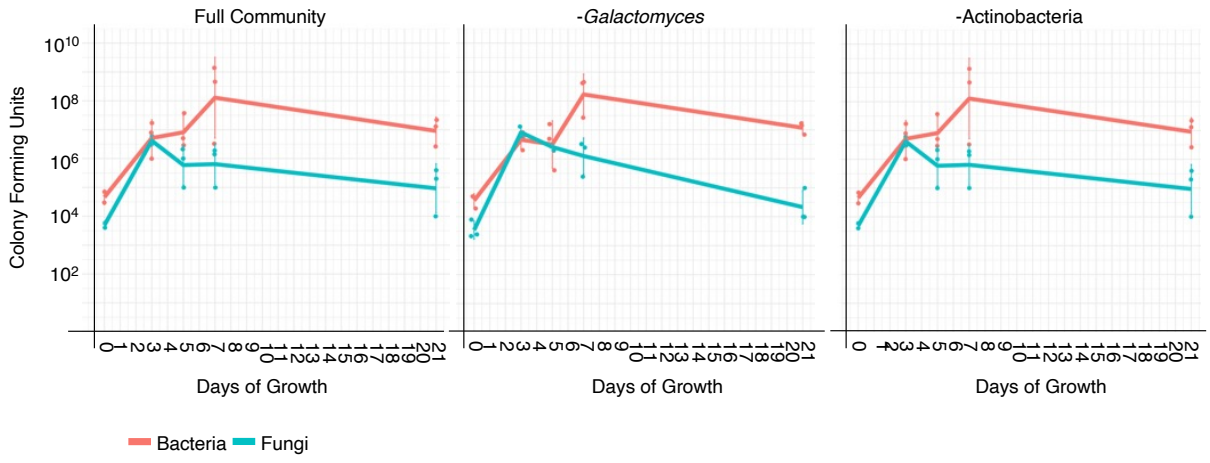
Supplemental Figure 3.2.10. Size distributions of predicted plasmid contigs. Size distributions of the predicted plasmid contigs are shown for (A) cheese A, (B) cheese B and (C) cheese C. Graphs were produced based on ViralVerify output using Microsoft® Excel for Mac Version 16.61.



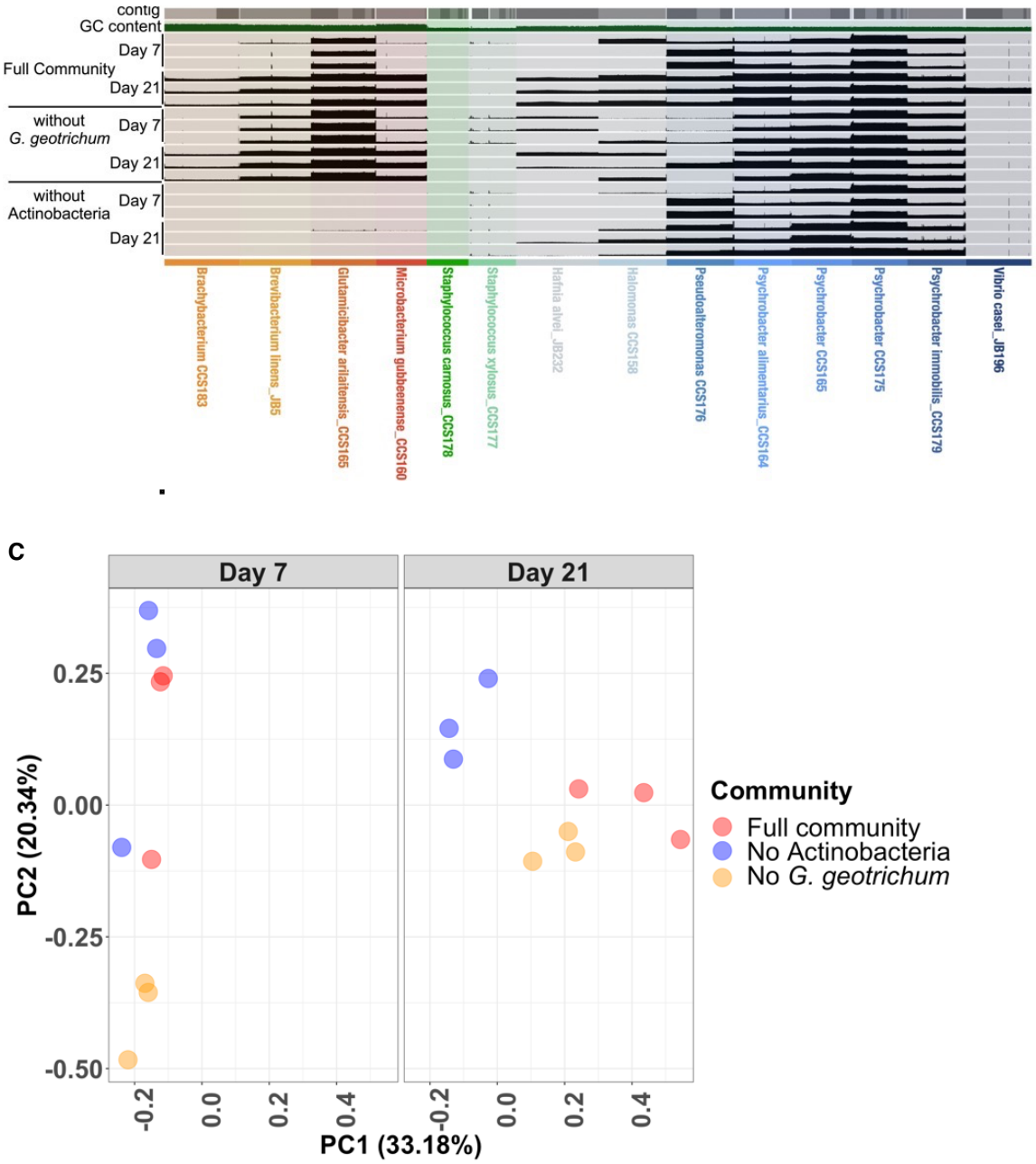
Supplemental Figure 3.2.11. Long reads and metaHiC combined associate large numbers of lytic viruses and plasmids with hosts and unbinned contigs. For each cheese the number of lytic virus and plasmid contigs (combined due to some overlap in the classification of contigs as both lytic phages and plasmids) that is associated with MAGs or with non-MAG contigs is shown in relation to the unassociated lytic virus/plasmid contigs.



Supplemental Figure 3.2.12. Overview of *in vitro* washed cheese rind community experiment. (a) Community members, colored by Phyla. (b) Timeline of *in vitro* microbial community model as well as washing pattern. All three steps were done on each plate with a sterile cotton swab (c) example plate of full community at day 21 from the partially destructive sampling.



Supplemental Figure 3.2.13. CFU counts of *in vitro* communities over time.



Supplemental Figure 3.2.14. Summary of read mapping against reference bacterial genomes of *in vitro* communities over time and PCA

Acknowledgements

Chapter 3 includes material that is in preparation for publication as well as additional material that will be excluded from that publication. Contributors to this chapter include Saak, Christina C.; Patel, Lucas, Dutton, Rachel J. The dissertation author was the primary researcher and author of this chapter. The publication that includes this data was with Saak, Christina C; Pierce, Emily C; Portik, Daniel; Hall, Richard; Ashby, Meredith; and Dutton, Rachel J.

Chapter 4 Reflections and Future Directions

How individuals behave in a group and how this group is more than the sum of its parts have remained a central thread through my academic and extracurricular careers. Can we further our understanding of how beings interact at a small scale and how this effects changes at a larger scale. From an undergraduate training in materials science, in which I learned about atomic interactions with a particular focus on polymer physics and tissue engineering, to a medical training to learn about how these tissues and organs interact within a human, to my graduate training in studying microbiomes in which many different organisms compete and cooperate. The interactions between different systems are complex, but need to be understood to better describe and influence them for our benefit.

Over the next decade, I will be starting my clinical training and hope to contribute to the understanding of how food, microbes, and community influence human health. But model systems, in which we can control variables and search for different outcomes, for these complex systems are elusive. The continued innovations and progress on large projects, such as body-on-a-chip, and standardization of multi-species engineered systems will raise the ceiling on how reductionist we have to be to answer basic science questions of causation instead of describing correlations. Once more causative variables are known, tests can be generated to quantify these variables in individuals and communities to personalize treatments and programs that can best resolve problems. Beyond biological questions, this framework can be applied to larger complex systems, such as a neighborhood or hospital. Leaning on skills learned from my extracurricular activities to lead large organizations, create new projects, and balance budgets with community-centered input, I can help integrate a view of a

patient that spans from the subcellular level, to their physical and mental health, to how they interact with public services and access basic needs. Our ability as physicians to treat maladies relies on the science and art of medicine and on the patients' self and environment to complete medication regimens or procedures, and follow up with additional visits.

In particular, the application of metagenomic sequencing in a clinical setting continues to be limited to academic centers, especially as these compete with the current gold standards of culturing blood or infected tissue and exudate. However, metagenomic sequencing may be able to provide more immediate answers on which species are in a certain community as well as identify genes associated with various antibiotic resistances such that these can be avoided. Especially as sequencings and proximity ligation technology continues to improve and decrease in cost, I hope to be at the forefront of implementing these techniques at the bedside.

References

"National Database of Antibiotic Resistant Organisms (NDARO) - Pathogen Detection - NCBI." n.d. Accessed December 16, 2019. https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/.

"Real-Time Detection of Antibiotic-Resistance Genes Using Oxford Nanopore Technologies' MinION." 2016. Oxford Nanopore Technologies. November 28, 2016. https://nanoporetech.com/resource-centre/real-time-detection-antibiotic-resistance-genes-using-oxford-nanopore-technologies.

Abriouel, H., C. W. Knapp, A. Gálvez, and N. Benomar. 2017. "Chapter 29 - Antibiotic Resistance Profile of Microbes From Traditional Fermented Foods." In Fermented Foods in Health and Disease Prevention, edited by Juana Frias, Cristina Martinez-Villaluenga, and Elena Peñas, 675–704. Boston: Academic Press.

Adato, Orit, Noga Ninyo, Uri Gophna, and Sagi Snir. 2015. "Detecting Horizontal Gene Transfer between Closely Related Taxa." PLoS Computational Biology 11 (10): e1004408.

Ainsworth, Stuart, Stephen Stockdale, Francesca Bottacini, Jennifer Mahony, and Douwe van Sinderen. 2014. "The Lactococcus Lactis Plasmidome: Much Learnt, yet Still Lots to Discover." FEMS Microbiology Reviews 38 (5): 1066–88.

Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, et al. 2019. "CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database." Nucleic Acids Research, October. https://doi.org/10.1093/nar/gkz935.

Allen, Heather K., Justin Donato, Helena Huimi Wang, Karen A. Cloud-Hansen, Julian Davies, and Jo Handelsman. 2010. "Call of the Wild: Antibiotic Resistance Genes in Natural Environments." Nature Reviews. Microbiology 8 (4): 251–59.

Alneberg, Johannes, Brynjar Smári Bjarnason, Ino de Bruijn, Melanie Schirmer, Joshua Quick, Umer Z. Ijaz, Leo Lahti, Nicholas J. Loman, Anders F. Andersson, and Christopher Quince. 2014. "Binning Metagenomic Contigs by Coverage and Composition." Nature Methods 11 (11): 1144–46.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." Journal of Molecular Biology 215 (3): 403–10.

Amgarten, Deyvid, Lucas P. P. Braga, Aline M. da Silva, and João C. Setubal. 2018. "MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins." Frontiers in Genetics 9 (August): 304.

Amselem, Joëlle, Guillaume Cornut, Nathalie Choisne, Michael Alaux, Françoise Alfama-Depauw, Véronique Jamilloux, Florian Maumus, et al. 2019. "RepetDB: A Unified Resource for Transposable Element References." Mobile DNA 10 (January): 6.

Anast, Justin M., Monika Dzieciol, Dylan L. Schultz, Martin Wagner, Evelyne Mann, and Stephan Schmitz-Esser. 2019. “Brevibacterium from Austrian Hard Cheese Harbor a Putative Histamine Catabolism Pathway and a Plasmid for Adaptation to the Cheese Environment.” Scientific Reports 9 (1): 6164.

Antipov, Dmitry, Anton Korobeynikov, Jeffrey S. McLean, and Pavel A. Pevzner. 2016. “hybridSPAdes: An Algorithm for Hybrid Assembly of Short and Long Reads.” Bioinformatics 32 (7): 1009–15.

Antipov, Dmitry, Mikhail Raiko Alla Lapidus, Pavel A Pevzner. 2020. “metaviralSPAdes: Assembly of Viruses from Metagenomic Data.” Bioinformatics , May. https://doi.org/10.1093/bioinformatics/btaa490.

Antipov, Dmitry, Mikhail Raiko, Alla Lapidus, and Pavel A. Pevzner. 2019. “Plasmid Detection and Assembly in Genomic and Metagenomic Data Sets.” Genome Research 29 (6): 961–68.

Antonopoulos, Dionysios A., Rida Assaf, Ramy Karam Aziz, Thomas Brettin, Christopher Bun, Neal Conrad, James J. Davis, et al. 2019. “PATRIC as a Unique Resource for Studying Antimicrobial Resistance.” Briefings in Bioinformatics 20 (4): 1094–1102.

Arango-Argoty, G. A., D. Dai, A. Pruden, P. Vikesland, L. S. Heath, and L. Zhang. 2019. “NanoARG: A Web Service for Detecting and Contextualizing Antimicrobial Resistance Genes from Nanopore-Derived Metagenomes.” Microbiome 7 (1): 88.

Arango-Argoty, Gustavo, Emily Garner, Amy Pruden, Lenwood S. Heath, Peter Vikesland, and Liqing Zhang. 2018. “DeepARG: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from Metagenomic Data.” Microbiome 6 (1): 23.

Archibald, F. 1986. “Manganese: Its Acquisition by and Function in the Lactic Acid Bacteria.” Critical Reviews in Microbiology 13 (1): 63–109.

Aristimuño Ficoseco, Cecilia, Flavia I. Mansilla, Natalia C. Maldonado, Hortencia Miranda, María E. Fátima Nader-Macias, and Graciela M. Vignolo. 2018. “Safety and Growth Optimization of Lactic Acid Bacteria Isolated From Feedlot Cattle for Probiotic Formula Design.” Frontiers in Microbiology 9 (September): 2220.

Arndt, David, Ana Marcu, Yongjie Liang, and David S. Wishart. 2017. “PHAST, PHASTER and PHASTEST: Tools for Finding Prophage in Bacterial Genomes.” Briefings in Bioinformatics, September. https://doi.org/10.1093/bib/bbx121.

Arndt, David, Jason R. Grant, Ana Marcu, Tanvir Sajed, Allison Pon, Yongjie Liang, and David S. Wishart. 2016. “PHASTER: A Better, Faster Version of the PHAST Phage Search Tool.” Nucleic Acids Research 44 (W1): W16–21.

Arredondo-Alonso, Sergio, Malbert R. C. Rogers, Johanna C. Braat, Tess D. Verschuuren, Janetta Top, Jukka Corander, Rob J. L. Willems, and Anita C. Schürch. 2018. “Mlplasmids: A

User-Friendly Tool to Predict Plasmid- and Chromosome-Derived Sequences for Single Species." Microbial Genomics 4 (11). https://doi.org/10.1099/mgen.0.000224.

Arumugam, Krithika, Caner Bağcı, Irina Bessarab, Sina Beier, Benjamin Buchfink, Anna Górska, Guanglei Qiu, Daniel H. Huson, and Rohan B. H. Williams. 2019. "Annotated Bacterial Chromosomes from Frame-Shift-Corrected Long-Read Metagenomic Data." Microbiome 7 (1): 61.

Arvey, Aaron J., Rajeev K. Azad, Alpan Raval, and Jeffrey G. Lawrence. 2009. "Detection of Genomic Islands via Segmental Genome Heterogeneity." Nucleic Acids Research 37 (16): 5255–66.

Assaf, Rida, Fangfang Xia, and Rick Stevens. 2019. "Identifying Genomic Islands with Deep Neural Networks." bioRxiv. https://doi.org/10.1101/525030.

Avery, O. T., C. M. Macleod, and M. McCarty. 1944. "STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III." The Journal of Experimental Medicine 79 (2): 137–58.

Bae, Junwoo, Kyeong Won Lee, Mohammad Nazrul Islam, Hyung-Soon Yim, Heejin Park, and Mina Rho. 2018. "iMGEins: Detecting Novel Mobile Genetic Elements Inserted in Individual Genomes." BMC Genomics 19 (1): 944.

Bakkeren, Erik, Jana S. Huisman, Stefan A. Fattinger, Annika Hausmann, Markus Furter, Adrian Egli, Emma Slack, et al. 2019. "Salmonella Persisters Promote the Spread of Antibiotic Resistance Plasmids in the Gut." Nature 573 (7773): 276–80.

Bale, M. J., M. J. Day, and J. C. Fry. 1988. "Novel Method for Studying Plasmid Transfer in Undisturbed River Epilithon." Applied and Environmental Microbiology 54 (11): 2756–58.

Bao, Weidong, Kenji K. Kojima, and Oleksiy Kohany. 2015. "Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes." Mobile DNA 6 (June): 11.

Baudry, Lyam, Théo Foutel-Rodier, Agnès Thierry, Romain Koszul, and Martial Marbouty. 2019. "MetaTOR: A Computational Pipeline to Recover High-Quality Metagenomic Bins From Mammalian Gut Proximity-Ligation (meta3C) Libraries." Frontiers in Genetics 10 (August): 753.

Beaulaurier, John, Shijia Zhu, Gintaras Deikus, Ilaria Mogno, Xue-Song Zhang, Austin Davis-Richardson, Ronald Canepa, et al. 2018. "Metagenomic Binning and Association of Plasmids with Bacterial Host Genomes Using DNA Methylation." Nature Biotechnology 36 (1): 61–69.

Beitel, Christopher W., Lutz Froenicke, Jenna M. Lang, Ian F. Korf, Richard W. Michelmore, Jonathan A. Eisen, and Aaron E. Darling. 2014. “Strain- and Plasmid-Level Deconvolution of a Synthetic Metagenome by Sequencing Proximity Ligation Products.” PeerJ 2 (May): e415.

Bellanger, Xavier, Hélène Guilloteau, Sébastien Bonot, and Christophe Merlin. 2014. “Demonstrating Plasmid-Based Horizontal Gene Transfer in Complex Environmental Matrices: A Practical Approach for a Critical Review.” The Science of the Total Environment 493 (September): 872–82.

Bengtsson-Palme, Johan, Fredrik Boulund, Jerker Fick, Erik Kristiansson, and D. G. Joakim Larsson. 2014. “Shotgun Metagenomics Reveals a Wide Array of Antibiotic Resistance Genes and Mobile Elements in a Polluted Lake in India.” Frontiers in Microbiology 5 (December): 648.

Berglund, Fanny, Tobias Österlund, Fredrik Boulund, Nachiket P. Marathe, D. G. Joakim Larsson, and Erik Kristiansson. 2019. “Identification and Reconstruction of Novel Antibiotic Resistance Genes from Metagenomes.” Microbiome 7 (1): 52.

Bertelli, Claire, Keith E. Tilley, and Fiona S. L. Brinkman. 2018. “Microbial Genomic Island Discovery, Visualization and Analysis.” Briefings in Bioinformatics, June. https://doi.org/10.1093/bib/bby042.

Bertelli, Claire, Matthew R. Laird, Kelly P. Williams, Simon Fraser University Research Computing Group, Britney Y. Lau, Gemma Hoad, Geoffrey L. Winsor, and Fiona S. L. Brinkman. 2017. “IslandViewer 4: Expanded Prediction of Genomic Islands for Larger-Scale Datasets.” Nucleic Acids Research 45 (W1): W30–35.

Bertuzzi, A. S., A. M. Walsh, J. J. Sheehan, P. D. Cotter, F. Crispie, P. L. H. McSweeney, K. N. Kilcawley, and M. C. Rea. 2018. “Omics-Based Insights into Flavor Development and Microbial Succession within Surface-Ripened Cheese.” mSystems 3 (1). https://doi.org/10.1128/mSystems.00211-17.

Bickhart, D. M., M. Kolmogorov, E. Tseng, and D. Portik. 2021. “Generation of Lineage-Resolved Complete Metagenome-Assembled Genomes by Precision Phasing.” bioRxiv. https://www.biorxiv.org/content/10.1101/2021.05.04.442591v1.abstract.

Bickhart, Derek M., Mick Watson, Sergey Koren, Kevin Panke-Buisse, Laura M. Cersosimo, Maximilian O. Press, Curtis P. Van Tassell, et al. 2019. “Assignment of Virus and Antimicrobial Resistance Genes to Microbial Hosts in a Complex Microbial Community by Combined Long-Read Assembly and Proximity Ligation.” Genome Biology 20 (1): 153.

Bikkarolla, Santosh K., Viveka Nordberg, Fredrika Rajer, Vilhelm Müller, Muhammad Humaun Kabir, Sriram Kk, Albertas Dvirnas, et al. 2019. “Optical DNA Mapping Combined with Cas9-Targeted Resistance Gene Identification for Rapid Tracking of Resistance Plasmids in a Neonatal Intensive Care Unit Outbreak.” mBio 10 (4). https://doi.org/10.1128/mBio.00347-19.

Binh, Chu Thi Thanh, Holger Heuer, Martin Kaupenjohann, and Kornelia Smalla. 2008. "Piggery Manure Used for Soil Fertilization Is a Reservoir for Transferable Antibiotic Resistance Plasmids." FEMS Microbiology Ecology 66 (1): 25–37.

Bisanz, Jordan. n.d. qiime2R: Import qiime2 Artifacts to R. Github. Accessed March 17, 2022. https://github.com/jbisanz/qiime2R.

Bishara, Alex, Eli L. Moss, Mikhail Kolmogorov, Alma E. Parada, Ziming Weng, Arend Sidow, Anne E. Dekas, Serafim Batzoglou, and Ami S. Bhatt. 2018. "High-Quality Genome Sequences of Uncultured Microbes by Assembly of Read Clouds." Nature Biotechnology, October. https://doi.org/10.1038/nbt.4266.

Biswas, Ambarish, Joshua N. Gagnon, Stan J. J. Brouns, Peter C. Fineran, and Chris M. Brown. 2013. "CRISPRTarget: Bioinformatic Prediction and Analysis of crRNA Targets." RNA Biology 10 (5): 817–27.

Blasche, S., Y. Kim, R. Mars, E. Kafkia, and M. Maansson. 2019. "Emergence of Stable Coexistence in a Complex Microbial Community through Metabolic Cooperation and Spatio-Temporal Niche Partitioning." bioRxiv. https://www.biorxiv.org/content/10.1101/541870v1.abstract.

Blasche, Sonja, Yongkyu Kim, Ruben A. T. Mars, Daniel Machado, Maria Maansson, Eleni Kafkia, Alessio Milanese, et al. 2021. "Metabolic Cooperation and Spatiotemporal Niche Partitioning in a Kefir Microbial Community." Nature Microbiology 6 (2): 196–208.

Blesa, Alba, Carolina Elvira César, Beate Averhoff, and José Berenguer. 2015. "Noncanonical Cell-to-Cell DNA Transfer in Thermus Spp. Is Insensitive to Argonaute-Mediated Interference." Journal of Bacteriology 197 (1): 138–46.

Blesa, Alba, Ignacio Baquedano, Nieves G. Quintáns, Carlos P. Mata, José R. Castón, and José Berenguer. 2017. "The Transjugation Machinery of Thermus Thermophilus: Identification of TdtA, an ATPase Involved in DNA Donation." PLoS Genetics 13 (3): e1006669.

Blow, Matthew J., Tyson A. Clark, Chris G. Daum, Adam M. Deutschbauer, Alexey Fomenkov, Roxanne Fries, Jeff Froula, et al. 2016. "The Epigenomic Landscape of Prokaryotes." PLoS Genetics 12 (2): e1005854.

Bodinaku, Ina, Jason Shaffer, Allison B. Connors, Jacob L. Steenwyk, Megan N. Biango-Daniels, Erik K. Kastman, Antonis Rokas, Albert Robbat, and Benjamin E. Wolfe. 2019a. "Rapid Phenotypic and Metabolomic Domestication of Wild Penicillium Molds on Cheese." mBio. https://doi.org/10.1128/mbio.02445-19.

Boehme, Sybille, Guido Werner, Ingo Klare, Rolf Reissbrodt, and Wolfgang Witte. 2004. "Occurrence of Antibiotic-Resistant Enterobacteria in Agricultural Foodstuffs." Molecular Nutrition & Food Research 48 (7): 522–31.

Bokulich, Nicholas A., and David A. Mills. 2013. “Facility-Specific ‘House’ Microbiome Drives Microbial Landscapes of Artisan Cheesemaking Plants.” Applied and Environmental Microbiology 79 (17): 5214–23.

Bokulich, Nicholas A., Benjamin D. Kaehler, Jai Ram Rideout, Matthew Dillon, Evan Bolyen, Rob Knight, Gavin A. Huttley, and J. Gregory Caporaso. 2018. “Optimizing Taxonomic Classification of Marker-Gene Amplicon Sequences with QIIME 2’s q2-Feature-Classifier Plugin.” Microbiome 6 (1): 90.

Bokulich, Nicholas A., John H. Thorngate, Paul M. Richardson, and David A. Mills. 2014. “Microbial Biogeography of Wine Grapes Is Conditioned by Cultivar, Vintage, and Climate.” Proceedings of the National Academy of Sciences of the United States of America 111 (1): E139–48.

Bolduc, Benjamin, Ho Bin Jang, Guilhem Doulcier, Zhi-Qiang You, Simon Roux, and Matthew B. Sullivan. 2017. “vConTACT: An iVirus Tool to Classify Double-Stranded DNA Viruses That Infect Archaea and Bacteria.” PeerJ 5 (May): e3243.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” Bioinformatics 30 (15): 2114–20.

Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. “Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2.” Nature Biotechnology 37 (8): 852–57.

Bonham, Kevin S., Benjamin E. Wolfe, and Rachel J. Dutton. 2017. “Extensive Horizontal Gene Transfer in Cheese-Associated Bacteria.” eLife 6 (June). https://doi.org/10.7554/eLife.22144.

Bonot, Sébastien, and Christophe Merlin. 2010. “Monitoring the Dissemination of the Broad-Host-Range Plasmid pB10 in Sediment Microcosms by Quantitative PCR.” Applied and Environmental Microbiology 76 (1): 378–82.

Booth, Michael J., Giovanni Marsico, Martin Bachman, Dario Beraldi, and Shankar Balasubramanian. 2014. “Quantitative Sequencing of 5-Formylcytosine in DNA at Single-Base Resolution.” Nature Chemistry 6 (5): 435–40.

Booth, Michael J., Miguel R. Branco, Gabriella Ficz, David Oxley, Felix Krueger, Wolf Reik, and Shankar Balasubramanian. 2012. “Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution.” Science 336 (6083): 934–37.

Borgeaud, Sandrine, Lisa C. Metzger, Tiziana Scrignari, and Melanie Blokesch. 2015. “The Type VI Secretion System of Vibrio Cholerae Fosters Horizontal Gene Transfer.” Science 347 (6217): 63–67.

Bourrie, Benjamin C. T., Benjamin P. Willing, and Paul D. Cotter. 2016. “The Microbiota and Health Promoting Characteristics of the Fermented Beverage Kefir.” Frontiers in Microbiology 7 (May): 647.

Bradley, Phelim, N. Claire Gordon, Timothy M. Walker, Laura Dunn, Simon Heys, Bill Huang, Sarah Earle, et al. 2015. “Rapid Antibiotic-Resistance Predictions from Genome Sequence Data for Staphylococcus Aureus and Mycobacterium Tuberculosis.” Nature Communications 6 (December): 10063.

Braga, Lucas P. P., Aymé Spor, Witold Kot, Marie-Christine Breuil, Lars H. Hansen, João C. Setubal, and Laurent Philippot. 2020. “Impact of Phages on Soil Bacterial Communities and Nitrogen Availability under Different Assembly Scenarios.” Microbiome 8 (1): 52.

Branton, Daniel, David W. Deamer, Andre Marziali, Hagan Bayley, Steven A. Benner, Thomas Butler, Massimiliano Di Ventra, et al. 2008. “The Potential and Challenges of Nanopore Sequencing.” Nature Biotechnology 26 (10): 1146–53.

Breitbart, Mya, Chelsea Bonnain, Kema Malki, and Natalie A. Sawaya. 2018. “Phage Puppet Masters of the Marine Microbial Realm.” Nature Microbiology 3 (7): 754–66.

Brito, Daniel M. de, Vinicius Maracaja-Coutinho, Savio T. de Farias, Leonardo V. Batista, and Thaís G. do Rêgo. 2016. “A Novel Method to Predict Genomic Islands Based on Mean Shift Clustering Algorithm.” PloS One 11 (1): e0146352.

Brooks, Lauren, Mo Kaze, and Mark Sistrom. 2019. “A Curated, Comprehensive Database of Plasmid Sequences.” Microbiology Resource Announcements 8 (1). https://doi.org/10.1128/MRA.01325-18.

Brown-Jaque, Maryury, William Calero-Cáceres, and Maite Muniesa. 2015. “Transfer of Antibiotic-Resistance Genes via Phage-Related Mobile Elements.” Plasmid. https://doi.org/10.1016/j.plasmid.2015.01.001.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. 2015. “Fast and Sensitive Protein Alignment Using DIAMOND.” Nature Methods 12 (1): 59–60.

Burton, Joshua N., Ivan Liachko, Maitreya J. Dunham, and Jay Shendure. 2014. “Species-Level Deconvolution of Metagenome Assemblies with Hi-C-Based Contact Probability Maps.” G3 4 (7): 1339–46.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” Nature Methods 13 (7): 581–83.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” BMC Bioinformatics 10 (December): 421.

Campo, Irene del, Raúl Ruiz, Ana Cuevas, Carlos Revilla, Luis Vielva, and Fernando de la Cruz. 2012. “Determination of Conjugation Rates on Solid Surfaces.” Plasmid 67 (2): 174–82.

Carattoli, Alessandra, Ea Zankari, Aurora García-Fernández, Mette Voldby Larsen, Ole Lund, Laura Villa, Frank Møller Aarestrup, and Henrik Hasman. 2014. “In Silico Detection and Typing of Plasmids Using PlasmidFinder and Plasmid Multilocus Sequence Typing.” Antimicrobial Agents and Chemotherapy 58 (7): 3895–3903.

Chatterjee, Raghunath, Keya Chaudhuri, and Probal Chaudhuri. 2008. “On Detection and Assessment of Statistical Significance of Genomic Islands.” BMC Genomics 9 (April): 150.

Chaumeil, Pierre-Alain, Aaron J. Mussig, Philip Hugenholtz, and Donovan H. Parks. 2019. “GTDB-Tk: A Toolkit to Classify Genomes with the Genome Taxonomy Database.” Bioinformatics , November. https://doi.org/10.1093/bioinformatics/btz848.

Che, Dongsheng, and Han Wang. 2013. “GIV: A Tool for Genomic Islands Visualization.” Bioinformation 9 (17): 879–82.

Che, Dongsheng, Cory Hockenbury, Robert Marmelstein, and Khaled Rasheed. 2010. “Classification of Genomic Islands Using Decision Trees and Their Ensemble Algorithms.” BMC Genomics 11 Suppl 2 (November): S1.

Che, Dongsheng, Han Wang, John Fazekas, and Bernard Chen. 2014. “An Accurate Genomic Island Prediction Method for Sequenced Bacterial and Archaeal Genomes.” Journal of Proteomics & Bioinformatics 7 (8): 214.

Che, You, Yu Xia, Lei Liu, An-Dong Li, Yu Yang, and Tong Zhang. 2019. “Mobile Antibiotic Resistome in Wastewater Treatment Plants Revealed by Nanopore Metagenomic Sequencing.” Microbiome 7 (1): 44.

Cheeseman, Kevin, Jeanne Ropars, Pierre Renault, Joëlle Dupont, Jérôme Gouzy, Antoine Branca, Anne-Laure Abraham, et al. 2014. “Multiple Recent Horizontal Transfers of a Large Genomic Region in Cheese Making Fungi.” Nature Communications 5: 2876.

Chen, Gary G., Jeffrey A. Gross, Pierre-Eric Lutz, Kathryn Vaillancourt, Gilles Maussion, Alexandre Bramoulle, Jean-François Théroux, et al. 2017. “Medium Throughput Bisulfite Sequencing for Accurate Detection of 5-Methylcytosine and 5-Hydroxymethylcytosine.” BMC Genomics 18 (1): 96.

Chen, Inês, and David Dubnau. 2004. “DNA Uptake during Bacterial Transformation.” Nature Reviews. Microbiology 2 (3): 241–49.

Chen, Inês, Peter J. Christie, and David Dubnau. 2005. “The Ins and Outs of DNA Transfer in Bacteria.” Science 310 (5753): 1456–60.

Chen, Yujiao, Yuqian Wu, Li Liu, Jianhua Feng, Tiancheng Zhang, Sheng Qin, Xingyu Zhao, et al. 2019. “Study of the Whole Genome, Methylome and Transcriptome of Cordyceps Militaris.” Scientific Reports 9 (1): 898.

Chiang, Yin Ning, José R. Penadés, and John Chen. 2019. “Genetic Transduction by Phages and Chromosomal Islands: The New and Noncanonical.” PLoS Pathogens 15 (8): e1007878.

Chiapello, H., I. Bourgait, F. Sourivong, G. Heuclin, A. Gendrault-Jacquemard, M-A Petit, and M. El Karoui. 2005. “Systematic Determination of the Mosaic Structure of Bacterial Genomes: Species Backbone versus Strain-Specific Loops.” BMC Bioinformatics 6 (July): 171.

Chiapello, Hélène, Annie Gendrault, Christophe Caron, Jérome Blum, Marie-Agnès Petit, and Meriem El Karoui. 2008. “MOSAIC: An Online Database Dedicated to the Comparative Genomics of Bacterial Strains at the Intra-Species Level.” BMC Bioinformatics 9 (November): 498.

Chilton, Stephanie N., Jeremy P. Burton, and Gregor Reid. 2015. “Inclusion of Fermented Foods in Food Guides around the World.” Nutrients 7 (1): 390–404.

Christophe Dessimoz, Daniel Margadant, and Gaston H. Gonnet. 2008. “DLIGHT.” March 2008. https://dl.acm.org/doi/10.5555/1804315.1804342.

Cissé, Ousmane H., and Jason E. Stajich. 2019. “FGMP: Assessing Fungal Genome Completeness.” BMC Bioinformatics 20 (1): 184.

Clark, Ryan R., Todd A. Gray, and Keith M. Derbyshire. 2020. “Quantifying and Characterizing Distributive Conjugal Transfer in Mycobacterium Smegmatis.” Methods in Molecular Biology 2075: 123–34.

Clark, Tyson A., Iain A. Murray, Richard D. Morgan, Andrey O. Kislyuk, Kristi E. Spittle, Matthew Boitano, Alexey Fomenkov, Richard J. Roberts, and Jonas Korlach. 2012. “Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing.” Nucleic Acids Research 40 (4): e29.

Clasen, Frederick Johannes, Rian Ewald Pierneef, Bernard Slippers, and Oleg Reva. 2018. “EuGI: A Novel Resource for Studying Genomic Islands to Facilitate Horizontal Gene Transfer Detection in Eukaryotes.” BMC Genomics 19 (1): 323.

Clausen, Philip T. L. C., Ea Zankari, Frank M. Aarestrup, and Ole Lund. 2016. “Benchmarking of Methods for Identification of Antimicrobial Resistance Genes in Bacterial Whole Genome Data.” The Journal of Antimicrobial Chemotherapy 71 (9): 2484–88.

Claverys, Jean-Pierre, Bernard Martin, and Patrice Polard. 2009. “The Genetic Transformation Machinery: Composition, Localization, and Mechanism.” FEMS Microbiology Reviews 33 (3): 643–56.

Cooper, Robert M., Lev Tsimring, and Jeff Hasty. 2017. “Inter-Species Population Dynamics Enhance Microbial Horizontal Gene Transfer and Spread of Antibiotic Resistance.” eLife 6 (November). https://doi.org/10.7554/eLife.25950.

Cosetta, Casey M., and Benjamin E. Wolfe. 2020. “Deconstructing and Reconstructing Cheese Rind Microbiomes for Experiments in Microbial Ecology and Evolution.” Current Protocols in Microbiology 56 (1): e95.

Cosetta, Casey M., Nicole Kfoury, Albert Robbat, and Benjamin E. Wolfe. 2020. “Fungal Volatiles Mediate Cheese Rind Microbiome Assembly.” Environmental Microbiology 22 (11): 4745–60.

Couvin, David, Aude Bernheim, Claire Toffano-Nioche, Marie Touchon, Juraj Michalik, Bertrand Néron, Eduardo P. C. Rocha, Gilles Vergnaud, Daniel Gautheret, and Christine Pourcel. 2018. “CRISPRCasFinder, an Update of CRISRFinder, Includes a Portable Version, Enhanced Performance and Integrates Search for Cas Proteins.” Nucleic Acids Research 46 (W1): W246–51.

Cury, Jean, Sophie S. Abby, Olivia Doppelt-Azeroual, Bertrand Néron, and Eduardo P. C. Rocha. 2020. “Identifying Conjugative Plasmids and Integrative Conjugative Elements with CONJscan.” Methods in Molecular Biology 2075: 265–83.

Cury, Jean, Thomas Jové, Marie Touchon, Bertrand Néron, and Eduardo Pc Rocha. 2016. “Identification and Analysis of Integrons and Cassette Arrays in Bacterial Genomes.” Nucleic Acids Research 44 (10): 4539–50.

CustomKrakenDB. n.d. Github. Accessed May 25, 2022. https://github.com/DuttonLab/CustomKrakenDB.

D’Andrea, Marco Maria. n.d. “RED-DB - Resistance Determinants DataBase.” Accessed January 8, 2020. http://www.fibim.unisi.it/REDDB/Default.asp.

Daane, L. L., J. A. Molina, E. C. Berry, and M. J. Sadowsky. 1996. “Influence of Earthworm Activity on Gene Transfer from Pseudomonas Fluorescens to Indigenous Soil Bacteria.” Applied and Environmental Microbiology 62 (2): 515–21.

Dahlberg, C., M. Bergström, and M. Hermansson. 1998. “In Situ Detection of High Levels of Horizontal Plasmid Transfer in Marine Bacterial Communities.” Applied and Environmental Microbiology 64 (7): 2670–75.

Dai, Qi, Chaohui Bao, Yabing Hai, Sheng Ma, Tao Zhou, Cong Wang, Yunfei Wang, et al. 2018. “MTGIpick Allows Robust Identification of Genomic Islands from a Single Genome.” Briefings in Bioinformatics 19 (3): 361–73.

Dalia, Ankur B., David W. Lazinski, and Andrew Camilli. 2013. “Characterization of Undermethylated Sites in Vibrio Cholerae.” Journal of Bacteriology 195 (10): 2389–99.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” GigaScience 10 (2). https://doi.org/10.1093/gigascience/giab008.

Darst, Russell P., Carolina E. Pardo, Lingbao Ai, Kevin D. Brown, and Michael P. Kladde. 2010. “Bisulfite Sequencing of DNA.” Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.] Chapter 7 (July): Unit 7.9.1–17.

Davis, James J., Sébastien Boisvert, Thomas Brettin, Ronald W. Kenyon, Chunhong Mao, Robert Olson, Ross Overbeek, et al. 2016. “Antimicrobial Resistance Prediction in PATRIC and RAST.” Scientific Reports 6 (June): 27930.

Davray, Dimple, Dipti Deo, and Ram Kulkarni. 2020. “Plasmids Encode Niche-Specific Traits in Lactobacillaceae.” Microbial Genomics, November. https://doi.org/10.1099/mgen.0.000472.

De la Cruz Barrón, Magali, Christophe Merlin, Hélène Guilloteau, Emmanuelle Montargès-Pelletier, and Xavier Bellanger. 2018. “Suspended Materials in River Waters Differentially Enrich Class 1 Integron- and IncP-1 Plasmid-Carrying Bacteria in Sediments.” Frontiers in Microbiology 9 (July): 1443.

De Mandal, Surajit, Sambanduram Samarjit Singh, Rajendra Bose Muthukumaran, Kawl Thanzami, Vinod Kumar, and Nachimuthu Senthil Kumar. 2018. “Metagenomic Analysis and the Functional Profiles of Traditional Fermented Pork Fat ‘Sa-Um’ of Northeast India.” AMB Express 8 (1): 163.

Dean, Frank B., John R. Nelson, Theresa L. Giesler, and Roger S. Lasken. 2001. “Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification.” Genome Research 11 (6): 1095–99.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. “Capturing Chromosome Conformation.” Science 295 (5558): 1306–11.

Dekker, Job, Marc A. Marti-Renom, and Leonid A. Mirny. 2013. “Exploring the Three-Dimensional Organization of Genomes: Interpreting Chromatin Interaction Data.” Nature Reviews. Genetics 14 (6): 390–403.

Delaney, Sarah, Richard Murphy, and Fiona Walsh. 2018. “A Comparison of Methods for the Extraction of Plasmids Capable of Conferring Antibiotic Resistance in a Human Pathogen From Complex Broiler Cecal Samples.” Frontiers in Microbiology 9 (August): 1731.

Delavat, François, Ryo Miyazaki, Nicolas Carraro, Nicolas Pradervand, and Jan Roelof van der Meer. 2017. “The Hidden Life of Integrative and Conjugative Elements.” FEMS Microbiology Reviews 41 (4): 512–37.

Delcenserie, V., B. Taminiau, L. Delhalle, C. Nezer, P. Doyen, S. Crevecoeur, D. Roussey, N. Korsak, and G. Daube. 2014. “Microbiota Characterization of a Belgian Protected

Designation of Origin Cheese, Herve Cheese, Using Metagenomic Analysis." Journal of Dairy Science 97 (10): 6046–56.

Delmont, Tom O., and A. Murat Eren. 2018. "Linking Pangenomes and Metagenomes: theProchlorococcusmetapangenome." PeerJ 6 (January): e4320.

DeMaere, Matthew Z., and Aaron E. Darling. 2019. "bin3C: Exploiting Hi-C Sequencing Data to Accurately Resolve Metagenome-Assembled Genomes." Genome Biology 20 (1): 46.

Deter, Heather S., and Ting Lu. 2022. "Engineering Microbial Consortia with Rationally Designed Cellular Interactions." Current Opinion in Biotechnology 76 (August): 102730.

Devirgiliis, Chiara, Paola Zinno, Mariarita Stirpe, Simona Barile, and Giuditta Perozzi. 2014. "Functional Screening of Antibiotic Resistance Genes from a Representative Metagenomic Library of Food Fermenting Microbiota." BioMed Research International 2014 (August): 290967.

Dib, Julián R., Martin Wagenknecht, María E. Farías, and Friedhelm Meinhardt. 2015. "Strategies and Approaches in Plasmidome Studies—uncovering Plasmid Diversity Disregarding of Linear Elements?" Frontiers in Microbiology 6: 463.

Dilthey, Alexander T., Chirag Jain, Sergey Koren, and Adam M. Phillippy. 2019. "Strain-Level Metagenomic Assignment and Compositional Estimation for Long Reads with MetaMaps." Nature Communications 10 (1): 3066.

Ding, Wei, Franz Baumdicker, and Richard A. Neher. n.d. "panX: Pan-Genome Analysis and Exploration." https://doi.org/10.1101/072082.

Dodsworth, Jeremy A., Paul C. Blainey, Senthil K. Murugapiran, Wesley D. Swingley, Christian A. Ross, Susannah G. Tringe, Patrick S. G. Chain, et al. 2013. "Single-Cell and Metagenomic Analyses Indicate a Fermentative and Saccharolytic Lifestyle for Members of the OP9 Lineage." Nature Communications 4: 1854.

Dordet-Frisoni, Emilie, Marion Faucher, Eveline Sagné, Eric Baranowski, Florence Tardy, Laurent Xavier Nouvel, and Christine Citti. 2019. "Mycoplasma Chromosomal Transfer: A Distributive, Conjugative Process Creating an Infinite Variety of Mosaic Genomes." Frontiers in Microbiology 10 (October): 2441.

Dostie, Josée, Todd A. Richmond, Ramy A. Arnaout, Rebecca R. Selzer, William L. Lee, Tracey A. Honan, Eric D. Rubio, et al. 2006. "Chromosome Conformation Capture Carbon Copy (5C): A Massively Parallel Solution for Mapping Interactions between Genomic Elements." Genome Research 16 (10): 1299–1309.

Douglas, Gavin M., and Morgan G. I. Langille. 2019. "Current and Promising Approaches to Identify Horizontal Gene Transfer Events in Metagenomes." Genome Biology and Evolution 11 (10): 2750–66.

Dubey, Gyanendra P., and Sigal Ben-Yehuda. 2011. “Intercellular Nanotubes Mediate Bacterial Communication.” Cell 144 (4): 590–600.

Dubnau, David, and Melanie Blokesch. 2019. “Mechanisms of DNA Uptake by Naturally Competent Bacteria.” Annual Review of Genetics, August. https://doi.org/10.1146/annurev-genet-112618-043641.

Durieux, Isabelle, Christophe Ginevra, Laetitia Attaiech, Kévin Picq, Pierre-Alexandre Juan, Sophie Jarraud, and Xavier Charpentier. 2019. “Diverse Conjugative Elements Silence Natural Transformation in Legionella Species.” Proceedings of the National Academy of Sciences of the United States of America 116 (37): 18613–18.

Edwards, John A., and Robert A. Edwards. 2019. “Fastq-Pair: Efficient Synchronization of Paired-End Fastq Files.” bioRxiv. https://doi.org/10.1101/552885.

Edwards, Robert A., Katelyn McNair, Karoline Faust, Jeroen Raes, and Bas E. Dutilh. 2015. “Computational Approaches to Predict Bacteriophage–host Relationships.” FEMS Microbiology Reviews 40 (2): 258–72.

Egervärn, Maria, Hans Lindmark, Johan Olsson, and Stefan Roos. 2010. “Transferability of a Tetracycline Resistance Gene from Probiotic Lactobacillus Reuteri to Bacteria in the Gastrointestinal Tract of Humans.” Antonie van Leeuwenhoek 97 (2): 189–200.

Eggers, Christian H., Carlie M. Gray, Alexander M. Preisig, Danielle M. Glenn, Jessica Pereira, Ryan W. Ayers, Mohammad Alshahrani, et al. 2016. “Phage-Mediated Horizontal Gene Transfer of Both Prophage and Heterologous DNA by φBB-1, a Bacteriophage of Borrelia Burgdorferi.” Pathogens and Disease 74 (9). https://doi.org/10.1093/femspd/ftw107.

Eid, John, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, et al. 2009. “Real-Time DNA Sequencing from Single Polymerase Molecules.” Science 323 (5910): 133–38.

Elhai, Jeff, Hailan Liu, and Arnaud Taton. 2012. “Detection of Horizontal Transfer of Individual Genes by Anomalous Oligomer Frequencies.” BMC Genomics 13 (June): 245.

Ellison, R. T., 3rd. 1994. “The Effects of Lactoferrin on Gram-Negative Bacteria.” Advances in Experimental Medicine and Biology 357: 71–90.

Erdmann, Susanne, Bernhard Tschitschko, Ling Zhong, Mark J. Raftery, and Ricardo Cavicchioli. 2017. “A Plasmid from an Antarctic Haloarchaeon Uses Specialized Membrane Vesicles to Disseminate and Infect Plasmid-Free Cells.” Nature Microbiology 2 (10): 1446–55.

Eren, A. Murat, Evan Kiefl, Alon Shaiber, Iva Veseli, Samuel E. Miller, Matthew S. Schechter, Isaac Fink, et al. 2021. “Community-Led, Integrated, Reproducible Multi-Omics with Anvi’o.” Nature Microbiology 6 (1): 3–6.

Eren, A. Murat, Özcan C. Esen, Christopher Quince, Joseph H. Vineis, Hilary G. Morrison, Mitchell L. Sogin, and Tom O. Delmont. 2015. “Anvi’o: An Advanced Analysis and Visualization Platform for 'Omics Data.” PeerJ 3 (October): e1319.

Falkowski, Paul G., Tom Fenchel, and Edward F. Delong. 2008. “The Microbial Engines That Drive Earth’s Biogeochemical Cycles.” Science 320 (5879): 1034–39.

Fang, Gang, Diana Munera, David I. Friedman, Anjali Mandlik, Michael C. Chao, Onureena Banerjee, Zhixing Feng, et al. 2012. “Genome-Wide Mapping of Methylated Adenine Residues in Pathogenic Escherichia Coli Using Single-Molecule Real-Time Sequencing.” Nature Biotechnology 30 (12): 1232–39.

Fang, Zhencheng, Jie Tan, Shufang Wu, Mo Li, Congmin Xu, Zhongjie Xie, and Huaiqiu Zhu. 2019. “PPR-Meta: A Tool for Identifying Phages and Plasmids from Metagenomic Fragments Using Deep Learning.” GigaScience 8 (6). https://doi.org/10.1093/gigascience/giz066.

Faruque, Shah M., and John J. Mekalanos. 2012. “Phage-Bacterial Interactions in the Evolution of Toxigenic Vibrio Cholerae.” Virulence 3 (7): 556–65.

Feldgarden, Michael, Vyacheslav Brover, Daniel H. Haft, Arjun B. Prasad, Douglas J. Slotta, Igor Tolstoy, Gregory H. Tyson, et al. 2019. “Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates.” Antimicrobial Agents and Chemotherapy 63 (11). https://doi.org/10.1128/AAC.00483-19.

Feng, Xiaowen, Haoyu Cheng, Daniel Portik, and Heng Li. 2021. “Metagenome Assembly of High-Fidelity Long Reads with Hifiasm-Meta.” arXiv [q-bio.GN]. arXiv. http://arxiv.org/abs/2110.08457.

Flusberg, Benjamin A., Dale R. Webster, Jessica H. Lee, Kevin J. Travers, Eric C. Olivares, Tyson A. Clark, Jonas Korlach, and Stephen W. Turner. 2010. “Direct Detection of DNA Methylation during Single-Molecule, Real-Time Sequencing.” Nature Methods 7 (6): 461–65.

Flutre, Timothée, Elodie Duprat, Catherine Feuillet, and Hadi Quesneville. 2011. “Considering Transposable Element Diversification in de Novo Annotation Approaches.” PloS One 6 (1): e16526.

Forsberg, Kevin J., Sanket Patel, Molly K. Gibson, Christian L. Lauber, Rob Knight, Noah Fierer, and Gautam Dantas. 2014. “Bacterial Phylogeny Structures Soil Resistomes across Habitats.” Nature 509 (7502): 612–16.

Fouts, Derrick E. 2006. “Phage_Finder: Automated Identification and Classification of Prophage Regions in Complete Bacterial Genome Sequences.” Nucleic Acids Research 34 (20): 5839–51.

Fraqueza, Maria João. 2015. “Antibiotic Resistance of Lactic Acid Bacteria Isolated from Dry-Fermented Sausages.” International Journal of Food Microbiology 212 (November): 76–88.

Galata, Valentina, Tobias Fehlmann, Christina Backes, and Andreas Keller. 2019. “PLSDB: A Resource of Complete Bacterial Plasmids.” Nucleic Acids Research 47 (D1): D195–202.

Galperin, Michael Y., Kira S. Makarova, Yuri I. Wolf, and Eugene V. Koonin. 2015. “Expanded Microbial Genome Coverage and Improved Protein Family Annotation in the COG Database.” Nucleic Acids Research 43 (Database issue): D261–69.

Gao, Feng, and Chun-Ting Zhang. 2006. “GC-Profile: A Web-Based Tool for Visualizing and Analyzing the Variation of GC Content in Genomic Sequences.” Nucleic Acids Research 34 (Web Server issue): W686–91.

Garg, Aarti, and Dinesh Gupta. 2008. “VirulentPred: A SVM Based Prediction Method for Virulent Proteins in Bacterial Pathogens.” BMC Bioinformatics 9 (January): 62.

Garger, S. J., O. M. Griffith, and L. K. Grill. 1983. “Rapid Purification of Plasmid DNA by a Single Centrifugation in a Two-Step Cesium Chloride-Ethidium Bromide Gradient.” Biochemical and Biophysical Research Communications 117 (3): 835–42.

Gargis, Amy S., Blake Cherney, Andrew B. Conley, Heather P. McLaughlin, and David Sue. 2019. “Rapid Detection of Genetic Engineering, Structural Variation, and Antimicrobial Resistance Markers in Bacterial Biothreat Pathogens by Nanopore Sequencing.” Scientific Reports 9 (1): 13501.

Gibson, Molly K., Bin Wang, Sara Ahmadi, Carey-Ann D. Burnham, Phillip I. Tarr, Barbara B. Warner, and Gautam Dantas. 2016. “Developmental Dynamics of the Preterm Infant Gut Microbiota and Antibiotic Resistome.” Nature Microbiology 1 (March): 16024.

Gibson, Molly K., Kevin J. Forsberg, and Gautam Dantas. 2015. “Improved Annotation of Antibiotic Resistance Determinants Reveals Microbial Resistomes Cluster by Ecology.” The ISME Journal 9 (1): 207–16.

Gilbert, Jack A., Martin J. Blaser, J. Gregory Caporaso, Janet K. Jansson, Susan V. Lynch, and Rob Knight. 2018. “Current Understanding of the Human Microbiome.” Nature Medicine 24 (4): 392–400.

Gilchrist, Cameron L. M., and Yit-Heng Chooi. 2021. “Clinker & Clustermap.js: Automatic Generation of Gene Cluster Comparison Figures.” Bioinformatics , January. https://doi.org/10.1093/bioinformatics/btab007.

Gillings, Michael R., William H. Gaze, Amy Pruden, Kornelia Smalla, James M. Tiedje, and Yong-Guan Zhu. 2015. “Using the Class 1 Integron-Integrase Gene as a Proxy for Anthropogenic Pollution.” The ISME Journal 9 (6): 1269–79.

Girgis, Hani Z. 2015. “Red: An Intelligent, Rapid, Accurate Tool for Detecting Repeats de-Novo on the Genomic Scale.” BMC Bioinformatics 16 (July): 227.

Good, Benjamin H., Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. 2017. “The Dynamics of Molecular Evolution over 60,000 Generations.” Nature 551 (7678): 45–50.

Gray, Todd A., and Keith M. Derbyshire. 2018. “Blending Genomes: Distributive Conjugal Transfer in Mycobacteria, a Sexier Form of HGT.” Molecular Microbiology 108 (6): 601–13.

Greenlon, Alex, Peter L. Chang, Zehara Mohammed Damtew, Atsede Muleta, Noelia Carrasquilla-Garcia, Donghyun Kim, Hien P. Nguyen, et al. 2019. “Global-Level Population Genomics Reveals Differential Effects of Geography and Phylogeny on Horizontal Gene Transfer in Soil Bacteria.” Proceedings of the National Academy of Sciences of the United States of America 116 (30): 15200–209.

Greninger, Alexander L., Samia N. Naccache, Scot Federman, Guixia Yu, Placide Mbala, Vanessa Bres, Doug Stryke, et al. 2015. “Rapid Metagenomic Identification of Viral Pathogens in Clinical Samples by Real-Time Nanopore Sequencing Analysis.” Genome Medicine 7 (September): 99.

Groenenboom, Anneloes E., Megan E. Parker, Anne de Vries, Suzette de Groot, Stephanie Zobrist, Kimberly Mansen, Peiman Milani, Remco Kort, Eddy J. Smid, and Sijmen E. Schoustra. 2019. “Bacterial Community Dynamics in Lait Caillé, a Traditional Product of Spontaneous Fermentation from Senegal.” PloS One 14 (5): e0215658.

Grüll, M. P., M. E. Mulligan, and A. S. Lang. 2018. “Small Extracellular Particles with Big Potential for Horizontal Gene Transfer: Membrane Vesicles and Gene Transfer Agents.” FEMS Microbiology Letters 365 (19). https://doi.org/10.1093/femsle/fny192.

Gu, Wei, Steve Miller, and Charles Y. Chiu. 2019. “Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection.” Annual Review of Pathology 14 (January): 319–38.

Guliaev, Andrei S., and Seraphima K. Semyenova. 2019. “MGERT: A Pipeline to Retrieve Coding Sequences of Mobile Genetic Elements from Genome Assemblies.” Mobile DNA 10 (May): 21.

Guo, Huiling, Lin Pan, Lina Li, Jie Lu, Laiyu Kwok, Bilige Menghe, Heping Zhang, and Wenyi Zhang. 2017. “Characterization of Antibiotic Resistance Genes from Lactobacillus Isolated from Traditional Dairy Products.” Journal of Food Science 82 (3): 724–30.

Gupta, Sushim Kumar, Babu Roshan Padmanabhan, Seydina M. Diene, Rafael Lopez-Rojas, Marie Kempf, Luce Landraud, and Jean-Marc Rolain. 2014. “ARG-ANNOT, a New Bioinformatic Tool to Discover Antibiotic Resistance Genes in Bacterial Genomes.” Antimicrobial Agents and Chemotherapy 58 (1): 212–20.

Haaber, Jakob, Jørgen J. Leisner, Marianne T. Cohn, Arancha Catalan-Moreno, Jesper B. Nielsen, Henrik Westh, José R. Penadés, and Hanne Ingmer. 2016. "Bacterial Viruses Enable Their Host to Acquire Antibiotic Resistance Genes from Neighbouring Cells." Nature Communications 7 (November): 13333.

Hall, James P. J., Michael A. Brockhurst, and Ellie Harrison. 2017. "Sampling the Mobile Gene Pool: Innovation via Horizontal Gene Transfer in Bacteria." Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 372 (1735). https://doi.org/10.1098/rstb.2016.0424.

Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. "Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products." Chemistry & Biology 5 (10): R245–49.

Hasan, Mohammad Shabbir, Qi Liu, Han Wang, John Fazekas, Bernard Chen, and Dongsheng Che. 2012. "GIST: Genomic Island Suite of Tools for Predicting Genomic Islands in Genomic Sequences." Bioinformation 8 (4): 203–5.

Haskett, Timothy L., Jason J. Terpolilli, Amanuel Bekuma, Graham W. O'Hara, John T. Sullivan, Penghao Wang, Clive W. Ronson, and Joshua P. Ramsay. 2016. "Assembly and Transfer of Tripartite Integrative and Conjugative Genetic Elements." Proceedings of the National Academy of Sciences of the United States of America 113 (43): 12268–73.

Haskett, Timothy L., Jason J. Terpolilli, Vinoy K. Ramachandran, Callum J. Verdonk, Phillip S. Poole, Graham W. O'Hara, and Joshua P. Ramsay. 2018. "Sequential Induction of Three Recombination Directionality Factors Directs Assembly of Tripartite Integrative and Conjugative Elements." PLoS Genetics 14 (3): e1007292.

Hawkey, Jane, Mohammad Hamidian, Ryan R. Wick, David J. Edwards, Helen Billman-Jacobe, Ruth M. Hall, and Kathryn E. Holt. 2015. "ISMapper: Identifying Transposase Insertion Sites in Bacterial Genomes from Short Read Sequence Data." BMC Genomics 16 (September): 667.

Hedges, R. W., and A. E. Jacob. 1974. "Transposition of Ampicillin Resistance from RP4 to Other Replicons." Molecular & General Genetics: MGG 132 (1): 31–40.

Hehemann, Jan-Hendrik, Gaëlle Correc, Tristan Barbeyron, William Helbert, Mirjam Czjzek, and Gurvan Michel. 2010. "Transfer of Carbohydrate-Active Enzymes from Marine Bacteria to Japanese Gut Microbiota." Nature 464 (7290): 908–12.

Hendriksen, Rene S., Valeria Bortolaia, Heather Tate, Gregory H. Tyson, Frank M. Aarestrup, and Patrick F. McDermott. 2019. "Using Genomics to Track Global Antimicrobial Resistance." Frontiers in Public Health 7 (September): 242.

Henschke, Rolf B., and Friedrich R. J. Schmidt. 1990. "Plasmid Mobilization from Genetically Engineered Bacteria to Members of the Indigenous Soil Microflora in Situ." Current Microbiology. https://doi.org/10.1007/bf02092881.

Hill, K. E., A. J. Weightman, and J. C. Fry. 1992. “Isolation and Screening of Plasmids from the Epilithon Which Mobilize Recombinant Plasmid pD10.” Applied and Environmental Microbiology 58 (4): 1292–1300.

Hinnebusch, J., and K. Tilly. 1993. “Linear Plasmids and Chromosomes in Bacteria.” Molecular Microbiology 10 (5): 917–22.

Hiraoka, Satoshi, Yusuke Okazaki, Mizue Anda, Atsushi Toyoda, Shin-Ichi Nakano, and Wataru Iwasaki. 2019. “Metaepigenomic Analysis Reveals the Unexplored Diversity of DNA Methylation in an Environmental Prokaryotic Community.” Nature Communications 10 (1): 159.

Hoede, Claire, Sandie Arnoux, Mark Moisset, Timothée Chaumier, Olivier Inizan, Véronique Jamilloux, and Hadi Quesneville. 2014. “PASTEC: An Automatic Transposable Element Classification Tool.” PloS One 9 (5): e91929.

Hubley, Robert, Robert D. Finn, Jody Clements, Sean R. Eddy, Thomas A. Jones, Weidong Bao, Arian F. A. Smit, and Travis J. Wheeler. 2016. “The Dfam Database of Repetitive DNA Families.” Nucleic Acids Research 44 (D1): D81–89.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. “eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” Nucleic Acids Research 47 (D1): D309–14.

Huerta-Cepas, Jaime, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2017. “Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.” Molecular Biology and Evolution 34 (8): 2115–22.

Hultman, Jenni, Manu Tamminen, Katariina Pärnänen, Johannes Cairns, Antti Karkman, and Marko Virta. 2018. “Host Range of Antibiotic Resistance Genes in Wastewater Treatment Plant Influent and Effluent.” FEMS Microbiology Ecology 94 (4). https://doi.org/10.1093/femsec/fiy038.

Hunt, Martin, Alison E. Mather, Leonor Sánchez-Busó, Andrew J. Page, Julian Parkhill, Jacqueline A. Keane, and Simon R. Harris. 2017. “ARIBA: Rapid Antimicrobial Resistance Genotyping Directly from Sequencing Reads.” Microbial Genomics 3 (10): e000131.

Hurwitz, Bonnie L., Steven J. Hallam, and Matthew B. Sullivan. 2013. “Metabolic Reprogramming by Viruses in the Sunlit and Dark Ocean.” Genome Biology 14 (11): R123.

Huson, Daniel H., Benjamin Albrecht, Caner Bağcı, Irina Bessarab, Anna Górska, Dino Jolic, and Rohan B. H. Williams. 2018. “MEGAN-LR: New Algorithms Allow Accurate Binning and Easy Interactive Exploration of Metagenomic Long Reads and Contigs.” Biology Direct 13 (1): 6.

Huson, Daniel H., Sina Beier, Isabell Flade, Anna Górska, Mohamed El-Hadidi, Suparna Mitra, Hans-Joachim Ruscheweyh, and Rewati Tappu. 2016. “MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data.” PLoS Computational Biology 12 (6): e1004957.

Hyatt, Doug, Gwo-Liang Chen, Philip F. Locascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” BMC Bioinformatics 11 (March): 119.

Inouye, Michael, Harriet Dashnow, Lesley-Ann Raven, Mark B. Schultz, Bernard J. Pope, Takehiro Tomita, Justin Zobel, and Kathryn E. Holt. 2014. “SRST2: Rapid Genomic Surveillance for Public Health and Hospital Microbiology Labs.” Genome Medicine 6 (11): 90.

Irlinger, F., and C. Monnet. 2021. “Temporal Differences in Microbial Composition of Époisses Cheese Rinds during Ripening and Storage.” Journal of Dairy Science, April. https://doi.org/10.3168/jds.2021-20123.

Jain, Ruchi, Sandeep Ramineni, and Nita Parekh. 2011. “IGIPT - Integrated Genomic Island Prediction Tool.” Bioinformation 7 (6): 307–10.

Jani, Mehul, and Rajeev K. Azad. 2019. “IslandCafe: Compositional Anomaly and Feature Enrichment Assessment for Delineation of Genomic Islands.” G3 , August. https://doi.org/10.1534/g3.119.400562.

Jani, Mehul, Soham Sengupta, Kelsey Hu, and Rajeev K. Azad. 2017. “Deciphering Pathogenicity and Antibiotic Resistance Islands in Methicillin-Resistant Staphylococcus Aureus Genomes.” Open Biology 7 (12). https://doi.org/10.1098/rsob.170094.

Jaron, Kamil S., Jiří C. Moravec, and Natália Martínková. 2014. “SigHunt: Horizontal Gene Transfer Finder Optimized for Eukaryotic Genomes.” Bioinformatics 30 (8): 1081–86.

Jesus, Tiago F., Bruno Ribeiro-Gonçalves, Diogo N. Silva, Valeria Bortolaia, Mário Ramirez, and João A. Carriço. 2019. “Plasmid ATLAS: Plasmid Visual Analytics and Identification in High-Throughput Sequencing Data.” Nucleic Acids Research 47 (D1): D188–94.

Joensen, Katrine Grimstrup, Flemming Scheutz, Ole Lund, Henrik Hasman, Rolf S. Kaas, Eva M. Nielsen, and Frank M. Aarestrup. 2014. “Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic Escherichia Coli.” Journal of Clinical Microbiology 52 (5): 1501–10.

Jolley, Keith A., and Martin C. J. Maiden. 2010. “BIGSdb: Scalable Analysis of Bacterial Genome Variation at the Population Level.” BMC Bioinformatics 11 (December): 595.

Jolley, Keith A., James E. Bray, and Martin C. J. Maiden. 2018. “Open-Access Bacterial Population Genomics: BIGSdb Software, the PubMLST.org Website and Their Applications.” Wellcome Open Research 3 (September): 124.

Jones, Brian V., and Julian R. Marchesi. 2007. “Transposon-Aided Capture (TRACA) of Plasmids Resident in the Human Gut Mobile Metagenome.” Nature Methods 4 (1): 55–61.

Kamathewatta, Kanishka Indiwari, Rhys Nathan Bushell, Neil David Young, Mark Anthony Stevenson, Helen Billman-Jacobe, Glenn Francis Browning, and Marc Serge Marenda. 2019. “Exploration of Antibiotic Resistance Risks in a Veterinary Teaching Hospital with Oxford Nanopore Long Read Sequencing.” PloS One 14 (5): e0217600.

Kamelamela, Noelani, Michael Zalesne, Joshua Morimoto, Albert Robbat, and Benjamin E. Wolfe. 2018. “Indigo- and Indirubin-Producing Strains of Proteus and Psychrobacter Are Associated with Purple Rind Defect in a Surface-Ripened Cheese.” Food Microbiology 76 (December): 543–52.

Kang, Dongwan D., Feng Li, Edward Kirton, Ashleigh Thomas, Rob Egan, Hong An, and Zhong Wang. 2019. “MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome Assemblies.” PeerJ 7 (July): e7359.

Kang, Dongwan D., Jeff Froula, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities.” PeerJ. https://doi.org/10.7717/peerj.1165.

Karkman, Antti, Thi Thuy Do, Fiona Walsh, and Marko P. J. Virta. 2018. “Antibiotic-Resistance Genes in Waste Water.” Trends in Microbiology 26 (3): 220–28.

Kastman et al. 2017. “Erratum for Kastman et al., ‘Biotic Interactions Shape the Ecological Distributions of Staphylococcus Species.’” mBio 8 (2). https://doi.org/10.1128/mBio.00329-17.

Kastman, Erik K., Noelani Kamelamela, Josh W. Norville, Casey M. Cosetta, Rachel J. Dutton, and Benjamin E. Wolfe. 2016. “Biotic Interactions Shape the Ecological Distributions of Staphylococcus Species.” mBio 7 (5). https://doi.org/10.1128/mBio.01157-16.

Kav, Aya Brown, Goor Sasson, Elie Jami, Adi Doron-Faigenboim, Itai Benhar, and Itzhak Mizrahi. 2012. “Insights into the Bovine Rumen Plasmidome.” Proceedings of the National Academy of Sciences of the United States of America 109 (14): 5452–57.

Kav, Aya Brown, Roye Rozov, David Bogumil, Søren Johannes Sørensen, Lars Hestbjerg Hansen, Itai Benhar, Eran Halperin, Ron Shamir, and Itzhak Mizrahi. 2019. “Unravelling Plasmidome Distribution and Interaction with Its Hosting Microbiome.” Environmental Microbiology. https://doi.org/10.1111/1462-2920.14813.

Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer. 2016. “MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function.” Methods in Molecular Biology 1399: 207–33.

Kell, Douglas B., Eugene L. Heyden, and Etheresia Pretorius. 2020. “The Biology of Lactoferrin, an Iron-Binding Protein That Can Help Defend Against Viruses and Bacteria.” Frontiers in Immunology 11 (May): 1221.

Kelleher, Philip, Jennifer Mahony, Francesca Bottacini, Gabriele A. Lugli, Marco Ventura, and Douwe van Sinderen. 2019. “The Lactococcus Lactis Pan-Plasmidome.” Frontiers in Microbiology 10 (April): 707.

Khalil, Rowaida K. S., Craig Skinner, Stephanie Patfield, and Xiaohua He. 2016. “Phage-Mediated Shiga Toxin (Stx) Horizontal Gene Transfer and Expression in Non-Shiga Toxigenic Enterobacter and Escherichia Coli Strains.” Pathogens and Disease 74 (5). https://doi.org/10.1093/femspd/ftw037.

Khot, Varada, Marc Strous, and Alyse K. Hawley. 2020. “Computational Approaches in Viral Ecology.” Computational and Structural Biotechnology Journal 18 (June): 1605–12.

Kieft, Kristopher, Zhichao Zhou, and Karthik Anantharaman. 2020. “VIBRANT: Automated Recovery, Annotation and Curation of Microbial Viruses, and Evaluation of Viral Community Function from Genomic Sequences.” Microbiome 8 (1): 90.

Kim, Myungjin, and Jongsik Chun. 2005. “Bacterial Community Structure in Kimchi, a Korean Fermented Vegetable Food, as Revealed by 16S rRNA Gene Analysis.” International Journal of Food Microbiology 103 (1): 91–96.

Klümper, Uli, Ariadni Droumpali, Arnaud Dechesne, and Barth F. Smets. 2014. “Novel Assay to Measure the Plasmid Mobilizing Potential of Mixed Microbial Communities.” Frontiers in Microbiology 5 (December): 730.

Klümper, Uli, Leise Riber, Arnaud Dechesne, Analia Sannazzarro, Lars H. Hansen, Søren J. Sørensen, and Barth F. Smets. 2015. “Broad Host Range Plasmids Can Invade an Unexpectedly Diverse Fraction of a Soil Bacterial Community.” The ISME Journal 9 (4): 934–45.

Kobayashi, Ichizo, Ayaka Nobusato, Noriko Kobayashi-Takahashi, and Ikuo Uchiyama. 1999. “Shaping the Genome--Restriction--Modification Systems as Mobile Genetic Elements.” Current Opinion in Genetics & Development 9 (6): 649–56.

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong Shin, Kristen Kuhn, et al. 2020. “metaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs.” Nature Methods 17 (11): 1103–10.

Kominek, Jacek, Drew T. Doering, Dana A. Opulente, Xing-Xing Shen, Xiaofan Zhou, Jeremy DeVirgilio, Amanda B. Hulfachor, et al. 2019. “Eukaryotic Acquisition of a Bacterial Operon.” Cell, February. https://doi.org/10.1016/j.cell.2019.01.034.

Köster, Johannes, and Sven Rahmann. 2012. “Snakemake--a Scalable Bioinformatics Workflow Engine.” Bioinformatics 28 (19): 2520–22.

Kothe, Caroline Isabel, Alexander Bolotin, Bochra-Farah Kraïem, Bedis Dridi, Food Microbiome Team, and Pierre Renault. 2021. “Unraveling the World of Halophilic and Halotolerant Bacteria in Cheese by Combining Cultural, Genomic and Metagenomic Approaches.” International Journal of Food Microbiology, June, 109312.

Krawczyk, Pawel S., Leszek Lipinski, and Andrzej Dziembowski. 2018. “PlasFlow: Predicting Plasmid Sequences in Metagenomic Data Using Genome Signatures.” Nucleic Acids Research 46 (6): e35.

Kurtz, S., and C. Schleiermacher. 1999. “REPuter: Fast Computation of Maximal Repeats in Complete Genomes.” Bioinformatics 15 (5): 426–27.

Kuzyakov, Yakov, and Evgenia Blagodatskaya. 2015. “Microbial Hotspots and Hot Moments in Soil: Concept & Review.” Soil Biology & Biochemistry 83 (April): 184–99.

Lakin, Steven M., Chris Dean, Noelle R. Noyes, Adam Dettenwanger, Anne Spencer Ross, Enrique Doster, Pablo Rovira, et al. 2017. “MEGARes: An Antimicrobial Resistance Database for High Throughput Sequencing.” Nucleic Acids Research 45 (D1): D574–80.

Lan, Freeman, Benjamin Demaree, Noorsher Ahmed, and Adam R. Abate. 2017. “Single-Cell Genome Sequencing at Ultra-High-Throughput with Microfluidic Droplet Barcoding.” Nature Biotechnology 35 (7): 640–46.

Landis, Elizabeth A., Angela M. Oliverio, Erin A. McKenney, Lauren M. Nichols, Nicole Kfoury, Megan Biango-Daniels, Leonora K. Shell, et al. 2021. “The Diversity and Function of Sourdough Starter Microbiomes.” eLife 10 (January). https://doi.org/10.7554/eLife.61644.

Lang, Andrew S., Olga Zhaxybayeva, and J. Thomas Beatty. 2012. “Gene Transfer Agents: Phage-like Elements of Genetic Exchange.” Nature Reviews. Microbiology 10 (7): 472–82.

Lanza, Val F., María de Toro, M. Pilar Garcillán-Barcia, Azucena Mora, Jorge Blanco, Teresa M. Coque, and Fernando de la Cruz. 2014. “Plasmid Flux in Escherichia Coli ST131 Sublineages, Analyzed by Plasmid Constellation Network (PLACNET), a New Method for Plasmid Reconstruction from Whole Genome Sequences.” PLoS Genetics 10 (12): e1004766.

Larkin, Alyse A., and Adam C. Martiny. 2017. “Microdiversity Shapes the Traits, Niche Space, and Biogeography of Microbial Taxa.” Environmental Microbiology Reports 9 (2): 55–70.

Lee, Chi-Ching, Yi-Ping Phoebe Chen, Tzu-Jung Yao, Cheng-Yu Ma, Wei-Cheng Lo, Ping-Chiang Lyu, and Chuan Yi Tang. 2013. “GI-POP: A Combinational Annotation and Genomic Island Prediction Pipeline for Ongoing Microbial Genome Projects.” Gene 518 (1): 114–23.

Leplae, Raphaël, Gipsi Lima-Mendez, and Ariane Toussaint. 2010. “ACLAME: A CLAssification of Mobile Genetic Elements, Update 2010.” Nucleic Acids Research 38 (Database issue): D57–61.

Lerner, Aaron, Torsten Matthias, and Rustam Aminov. 2017. “Potential Effects of Horizontal Gene Exchange in the Human Gut.” Frontiers in Immunology 8 (November): 1630.

Lester, C. H., N. Frimodt-Moller, T. L. Sorensen, D. L. Monnet, and A. M. Hammerum. 2006. “In Vivo Transfer of the vanA Resistance Gene from an Enterococcus Faecium Isolate of Animal Origin to an E. Faecium Isolate of Human Origin in the Intestines of Human Volunteers.” Antimicrobial Agents and Chemotherapy. https://doi.org/10.1128/aac.50.2.596-599.2006.

Letunic, Ivica, and Peer Bork. 2021. “Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation.” Nucleic Acids Research 49 (W1): W293–96.

Levesque, Sébastien, Alessandra G. de Melo, Simon J. Labrie, and Sylvain Moineau. 2019. “Mobilome of Brevibacterium Aurantiacum Sheds Light on Its Genetic Diversity and Its Adaptation to Smear-Ripened Cheeses.” Frontiers in Microbiology 10 (June): 1270.

Li, Bing, Yong Qiu, Jing Zhang, Xia Huang, Hanchang Shi, and Huabing Yin. 2018. “Real-Time Study of Rapid Spread of Antibiotic Resistance Plasmid in Biofilm Using Microfluidics.” Environmental Science & Technology 52 (19): 11132–41.

Li, Chen, Yiqi Jiang, and Shuaicheng Li. 2019. “LEMON: A Method to Construct the Local Strains at Horizontal Gene Transfer Sites in Gut Metagenomics.” BMC Bioinformatics 20 (Suppl 23): 702.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” arXiv [q-bio.GN]. arXiv. http://arxiv.org/abs/1303.3997.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” Bioinformatics 34 (18): 3094–3100.

Li, Jun, Cui Tai, Zixin Deng, Weihong Zhong, Yongqun He, and Hong-Yu Ou. 2018. “VRprofile: Gene-Cluster-Detection-Based Profiling of Virulence and Antibiotic Resistance Traits Encoded within Genome Sequences of Pathogenic Bacteria.” Briefings in Bioinformatics 19 (4): 566–74.

Li, L. L., A. Norman, L. H. Hansen, and S. J. Sorensen. 2012. “Metamobilomics – Expanding Our Knowledge on the Pool of Plasmid Encoded Traits in Natural Environments Using High-Throughput Sequencing.” Clinical Microbiology and Infection. https://doi.org/10.1111/j.1469-0691.2012.03862.x.

Li, Xiangchen, Wenjun Tong, Lina Wang, Siddiq Ur Rahman, Gehong Wei, and Shiheng Tao. 2018. “A Novel Strategy for Detecting Recent Horizontal Gene Transfer and Its Application to Rhizobium Strains.” Frontiers in Microbiology 9 (May): 973.

Li, Xiaobin, Yingzhou Xie, Meng Liu, Cui Tai, Jingyong Sun, Zixin Deng, and Hong-Yu Ou. 2018. “oriTfinder: A Web-Based Tool for the Identification of Origin of Transfers in DNA

Sequences of Bacterial Mobile Genetic Elements." Nucleic Acids Research 46 (W1): W229–34.

Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." Science 326 (5950): 289–93.

Lilley, A. K., and M. J. Bailey. 1997. "The Acquisition of Indigenous Plasmids by a Genetically Marked Pseudomonad Population Colonizing the Sugar Beet Phytosphere Is Related to Local Environmental Conditions." Applied and Environmental Microbiology 63 (4): 1577–83.

Liu, Bo, and Mihai Pop. 2009. "ARDB—Antibiotic Resistance Genes Database." Nucleic Acids Research 37 (suppl_1): D443–47.

Liu, Bo, Dandan Zheng, Qi Jin, Lihong Chen, and Jian Yang. 2019. "VFDB 2019: A Comparative Pathogenomic Platform with an Interactive Web Interface." Nucleic Acids Research 47 (D1): D687–92.

Liu, Meng, Xiaobin Li, Yingzhou Xie, Dexi Bi, Jingyong Sun, Jun Li, Cui Tai, Zixin Deng, and Hong-Yu Ou. 2019. "ICEberg 2.0: An Updated Database of Bacterial Integrative and Conjugative Elements." Nucleic Acids Research 47 (D1): D660–65.

Liu, Qian, Li Fang, Guoliang Yu, Depeng Wang, Chuan-Le Xiao, and Kai Wang. 2019. "Detection of DNA Base Modifications by Deep Recurrent Neural Network on Oxford Nanopore Sequencing Data." Nature Communications 10 (1): 2449.

Llorens, Carlos, Ricardo Futami, Laura Covelli, Laura Domínguez-Escribá, Jose M. Viu, Daniel Tamarit, Jose Aguilar-Rodríguez, et al. 2011. "The Gypsy Database (GyDB) of Mobile Genetic Elements: Release 2.0." Nucleic Acids Research 39 (Database issue): D70–74.

Loose, Matthew, Sunir Malla, and Michael Stout. 2016. "Real-Time Selective Sequencing Using Nanopore Technology." Nature Methods 13 (9): 751–54.

Lopes, Robson da Silva, Walas Jhony Lopes Moraes, Thiago de Souza Rodrigues, and Daniella Castanheira Bartholomeu. 2015. "ProGeRF: Proteome and Genome Repeat Finder Utilizing a Fast Parallel Hash Function." BioMed Research International 2015 (February): 394157.

Lowe, T. M., and S. R. Eddy. 1997. "tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence." Nucleic Acids Research 25 (5): 955–64.

Lu, Bingxin, and Hon Wai Leong. 2016. "GI-SVM: A Sensitive Method for Predicting Genomic Islands Based on Unannotated Sequence of a Single Genome." Journal of Bioinformatics and Computational Biology 14 (1): 1640003.

Lu, Bingxin, Hon Wai Leong. 2018. “GI-Cluster: Detecting Genomic Islands via Consensus Clustering on Multiple Features.” Journal of Bioinformatics and Computational Biology 16 (3): 1840010.

Lu, Xingyu, Chun-Xiao Song, Keith Szulwach, Zhipeng Wang, Payton Weidenbacher, Peng Jin, and Chuan He. 2013. “Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA.” Journal of the American Chemical Society 135 (25): 9315–17.

Madsen, Jonas S., Ole Hylling, Samuel Jacquiod, Sophia Pécastaings, Lars H. Hansen, Leise Riber, Gisle Vestergaard, and Søren J. Sørensen. 2018. “An Intriguing Relationship between the Cyclic Diguanylate Signaling System and Horizontal Gene Transfer.” The ISME Journal 12 (9): 2330–34.

Madsen, Jonas Stenløkke, Mette Burmølle, Lars Hestbjerg Hansen, and Søren Johannes Sørensen. 2012. “The Interconnection between Biofilm Formation and Horizontal Gene Transfer.” FEMS Immunology and Medical Microbiology 65 (2): 183–95.

Maguire, Finlay, Baofeng Jia, Kristen L. Gray, Wing Yin Venus Lau, Robert G. Beiko, and Fiona S. L. Brinkman. 2020. “Metagenome-Assembled Genome Binning Methods with Short Reads Disproportionately Fail for Plasmids and Genomic Islands.” Microbial Genomics 6 (10). https://doi.org/10.1099/mgen.0.000436.

Man, Tom J. B. de, and Brandi M. Limbago. 2016. “SSTAR, a Stand-Alone Easy-To-Use Antimicrobial Resistance Gene Predictor.” mSphere 1 (1). https://doi.org/10.1128/mSphere.00050-15.

Marbouty, Martial, Axel Cournac, Jean-François Flot, Hervé Marie-Nelly, Julien Mozziconacci, and Romain Koszul. 2014. “Metagenomic Chromosome Conformation Capture (meta3C) Unveils the Diversity of Chromosome Organization in Microorganisms.” eLife 3 (December): e03318.

Marbouty, Martial, Lyam Baudry, Axel Cournac, and Romain Koszul. 2017. “Scaffolding Bacterial Genomes and Probing Host-Virus Interactions in Gut Microbiome by Proximity Ligation (chromosome Capture) Assay.” Science Advances 3 (2): e1602105.

Marsh, Alan J., Orla O’Sullivan, Colin Hill, R. Paul Ross, and Paul D. Cotter. 2013. “Sequencing-Based Analysis of the Bacterial and Fungal Composition of Kefir Grains and Milks from Multiple Sources.” PloS One 8 (7): e69371.

Mason, Amy, Dona Foster, Phelim Bradley, Tanya Golubchik, Michel Doumith, N. Claire Gordon, Bruno Pichon, et al. 2018. “Accuracy of Different Bioinformatics Methods in Detecting Antibiotic Resistance and Virulence Factors from Staphylococcus Aureus Whole-Genome Sequences.” Journal of Clinical Microbiology 56 (9). https://doi.org/10.1128/JCM.01815-17.

Mastrigt, Oscar van, Elisa Di Stefano, Sylviani Hartono, Tjakko Abee, and Eddy J. Smid. 2018. “Large Plasmidome of Dairy Lactococcus Lactis Subsp. Lactis Biovar Diacetylactis FM03P Encodes Technological Functions and Appears Highly Unstable.” BMC Genomics 19 (1): 620.

Matamoros, Sébastien, Rene Hendriksen, Balint Pataki, Nima Pakseresht, Marc Rossello, Nicole Silvester, Clara Amid, et al. 2019. “Accelerating Surveillance and Research of Antimicrobial Resistance - an Online Repository for Sharing of Antimicrobial Susceptibility Data Associated with Whole Genome Sequences.” bioRxiv. https://doi.org/10.1101/532267.

Mater, Denis D. G., Philippe Langella, Gérard Corthier, and María-José Flores. 2008. “A Probiotic Lactobacillus Strain Can Acquire Vancomycin Resistance during Digestive Transit in Mice.” Journal of Molecular Microbiology and Biotechnology 14 (1-3): 123–27.

Matthey, Noémie, Sandrine Stutzmann, Candice Stoudmann, Nicolas Guex, Christian Iseli, and Melanie Blokesch. 2019. “Neighbor Predation Linked to Natural Competence Fosters the Transfer of Large Genomic Regions in Vibrio Cholerae.” https://doi.org/10.7554/eLife.48212.

May, Alexander, Shrinath Narayanan, Joe Alcock, Arvind Varsani, Carlo Maley, and Athena Aktipis. 2019. “Kombucha: A Novel Model System for Cooperation and Conflict in a Complex Multi-Species Microbial Ecosystem.” PeerJ 7 (September): e7565.

Mazel, Didier. 2006. “Integrons: Agents of Bacterial Evolution.” Nature Reviews. Microbiology 4 (8): 608–20.

McDaniel, Lauren D., Elizabeth Young, Jennifer Delaney, Fabian Ruhnau, Kim B. Ritchie, and John H. Paul. 2010. “High Frequency of Horizontal Gene Transfer in the Oceans.” Science 330 (6000): 50.

McDonald, Daniel, Jose C. Clemente, Justin Kuczynski, Jai Ram Rideout, Jesse Stombaugh, Doug Wendel, Andreas Wilke, et al. 2012. “The Biological Observation Matrix (BIOM) Format or: How I Learned to Stop Worrying and Love the Ome-Ome.” GigaScience 1 (1): 7.

McDonald, Daniel, Morgan N. Price, Julia Goodrich, Eric P. Nawrocki, Todd Z. DeSantis, Alexander Probst, Gary L. Andersen, Rob Knight, and Philip Hugenholtz. 2012. “An Improved Greengenes Taxonomy with Explicit Ranks for Ecological and Evolutionary Analyses of Bacteria and Archaea.” The ISME Journal 6 (3): 610–18.

McIntyre, Alexa B. R., Noah Alexander, Kirill Grigorev, Daniela Bezdan, Heike Sichtig, Charles Y. Chiu, and Christopher E. Mason. 2019. “Single-Molecule Sequencing Detection of N6-Methyladenine in Microbial Reference Materials.” Nature Communications 10 (1): 579.

McKinney, Wes. 2010. “Data Structures for Statistical Computing in Python.” In Proceedings of the 9th Python in Science Conference. SciPy. https://doi.org/10.25080/majora-92bf1922-00a.

McLaren, Michael R., Amy D. Willis, and Benjamin J. Callahan. 2019. “Consistent and Correctable Bias in Metagenomic Sequencing Experiments.” eLife 8 (September). https://doi.org/10.7554/eLife.46923.

McMurdie, Paul J., and Susan Holmes. 2013. “Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.” PloS One 8 (4): e61217.

McNair, Katelyn, Ramy Karam Aziz, Gordon D. Pusch, Ross Overbeek, Bas E. Dutilh, and Robert Edwards. 2018. “Phage Genome Annotation Using the RAST Pipeline.” In Bacteriophages: Methods and Protocols, Volume 3, edited by Martha R. J. Clokie, Andrew M. Kropinski, and Rob Lavigne, 231–38. New York, NY: Springer New York.

Meersman, Esther, Jan Steensels, Melissa Mathawan, Pieter-Jan Wittocx, Veerle Saels, Nore Struyf, Herwig Bernaert, Gino Vrancken, and Kevin J. Verstrepen. 2013. “Detailed Analysis of the Microbial Population in Malaysian Spontaneous Cocoa Pulp Fermentations Reveals a Core and Variable Microbiota.” PloS One 8 (12): e81559.

Merchant, Nirav, Eric Lyons, Stephen Goff, Matthew Vaughn, Doreen Ware, David Micklos, and Parker Antin. 2016. “The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences.” PLoS Biology 14 (1): e1002342.

Merlin, Christophe, Sébastien Bonot, Sophie Courtois, and Jean-Claude Block. 2011. “Persistence and Dissemination of the Multiple-Antibiotic-Resistance Plasmid pB10 in the Microbial Communities of Wastewater Sludge Microcosms.” Water Research 45 (9): 2897–2905.

Mølbak, Lars, Tine Rask Licht, Thomas Kvist, Niels Kroer, and Sigrid Rita Andersen. 2003. “Plasmid Transfer from Pseudomonas Putida to the Indigenous Bacteria on Alfalfa Sprouts: Characterization, Direct Quantification, and in Situ Location of Transconjugant Cells.” Applied and Environmental Microbiology 69 (9): 5536–42.

Monnet, Christophe, Alexandre Back, and Françoise Irlinger. 2012. “Growth of Aerobic Ripening Bacteria at the Cheese Surface Is Limited by the Availability of Iron.” Applied and Environmental Microbiology 78 (9): 3185–92.

Monnet, Christophe, Sophie Landaud, Pascal Bonnarme, and Dominique Swennen. 2015. “Growth and Adaptation of Microorganisms on the Cheese Surface.” FEMS Microbiology Letters 362 (1): 1–9.

Moore, Aimee M., Christian Munck, Morten O. A. Sommer, and Gautam Dantas. 2011. “Functional Metagenomic Investigations of the Human Intestinal Microbiota.” Frontiers in Microbiology 2 (October): 188.

Mori, Hideto, Daniel Evans-Yamamoto, Soh Ishiguro, Masaru Tomita, and Nozomu Yachie. 2019. “Fast and Global Detection of Periodic Sequence Repeats in Large Genomic Resources.” Nucleic Acids Research 47 (2): e8.

Morin, Manon, Emily C. Pierce, and Rachel J. Dutton. 2018. “Changes in the Genetic Requirements for Microbial Interactions with Increasing Community Complexity.” eLife 7 (September). https://doi.org/10.7554/eLife.37072.

Mounier, Jérôme, Monika Coton, Françoise Irlinger, Sophie Landaud, and Pascal Bonnarme. 2017. “Chapter 38 - Smear-Ripened Cheeses.” In Cheese (Fourth Edition), edited by Paul L. H. McSweeney, Patrick F. Fox, Paul D. Cotter, and David W. Everett, 955–96. San Diego: Academic Press.

Moura, Alexandra, Mário Soares, Carolina Pereira, Nuno Leitão, Isabel Henriques, and António Correia. 2009. “INTEGRALL: A Database and Search Engine for Integrons, Integrases and Gene Cassettes.” Bioinformatics 25 (8): 1096–98.

Musovic, S., A. Dechesne, J. Sorensen, and B. F. Smets. 2010. “Novel Assay To Assess Permissiveness of a Soil Microbial Community toward Receipt of Mobile Genetic Elements.” Applied and Environmental Microbiology. https://doi.org/10.1128/aem.02713-09.

Musovic, S., G. Oregaard, N. Kroer, and S. J. Sorensen. 2006. “Cultivation-Independent Examination of Horizontal Transfer and Host Range of an IncP-1 Plasmid among Gram-Positive and Gram-Negative Bacteria Indigenous to the Barley Rhizosphere.” Applied and Environmental Microbiology. https://doi.org/10.1128/aem.00013-06.

Nawaz, Muhammad, Juan Wang, Aiping Zhou, Chaofeng Ma, Xiaokang Wu, John E. Moore, B. Cherie Millar, and Jiru Xu. 2011. “Characterization and Transfer of Antibiotic Resistance in Lactic Acid Bacteria from Fermented Food Products.” Current Microbiology 62 (3): 1081–89.

Ni, Peng, Neng Huang, Zhi Zhang, De-Peng Wang, Fan Liang, Yu Miao, Chuan-Le Xiao, Feng Luo, and Jianxin Wang. 2019. “DeepSignal: Detecting DNA Methylation State from Nanopore Sequencing Reads Using Deep-Learning.” Bioinformatics 35 (22): 4586–95.

Niccum, Brittany A., Erik K. Kastman, Nicole Kfoury, Albert Robbat Jr, and Benjamin E. Wolfe. 2020. “Strain-Level Diversity Impacts Cheese Rind Microbiome Assembly and Function.” mSystems 5 (3). https://doi.org/10.1128/mSystems.00149-20.

Norman, Anders, Leise Riber, Wenting Luo, Li Li Li, Lars Hestbjerg Hansen, and Søren Johannes Sørensen. 2014. “An Improved Method for Including Upper Size Range Plasmids in Metamobilomes.” PloS One 9 (8): e104405.

Novo, Maite, Frédéric Bigey, Emmanuelle Beyne, Virginie Galeote, Frédérick Gavory, Sandrine Mallet, Brigitte Cambon, et al. 2009. “Eukaryote-to-Eukaryote Gene Transfer Events Revealed by the Genome Sequence of the Wine Yeast Saccharomyces Cerevisiae EC1118.” Proceedings of the National Academy of Sciences of the United States of America 106 (38): 16333–38.

Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. “Lateral Gene Transfer and the Nature of Bacterial Innovation.” Nature 405 (6784): 299–304.

Ofir, Gal, Sarah Melamed, Hila Sberro, Zohar Mukamel, Shahar Silverman, Gilad Yaakov, Shany Doron, and Rotem Sorek. 2018. “DISARM Is a Widespread Bacterial Defence System with Broad Anti-Phage Activities.” Nature Microbiology 3 (1): 90–98.

Ogier, Jean-Claude, Alexandra Calteau, Steve Forst, Heidi Goodrich-Blair, David Roche, Zoé Rouy, Garret Suen, et al. 2010. “Units of Plasticity in Bacterial Genomes: New Insight from the Comparative Genomics of Two Bacteria Interacting with Invertebrates, Photorhabdus and Xenorhabdus.” BMC Genomics 11 (October): 568.

Oh, Min, Amy Pruden, Chaoqi Chen, Lenwood S. Heath, Kang Xia, and Liqing Zhang. 2018. “MetaCompare: A Computational Pipeline for Prioritizing Environmental Resistome Risk.” FEMS Microbiology Ecology 94 (7). https://doi.org/10.1093/femsec/fiy079.

Olekhnovich, Evgenii I., Artem T. Vasilyev, Vladimir I. Ulyantsev, Elena S. Kostryukova, and Alexander V. Tyakht. 2018. “MetaCherchant: Analyzing Genomic Context of Antibiotic Resistance Genes in Gut Microbiota.” Bioinformatics 34 (3): 434–44.

Olm, Matthew R., Christopher T. Brown, Brandon Brooks, and Jillian F. Banfield. 2017. “dRep: A Tool for Fast and Accurate Genomic Comparisons That Enables Improved Genome Recovery from Metagenomes through de-Replication.” The ISME Journal 11 (12): 2864–68.

Ou, Hong-Yu, Ling-Ling Chen, James Lonnen, Roy R. Chaudhuri, Ali Bin Thani, Rebecca Smith, Natalie J. Garton, et al. 2006. “A Novel Strategy for the Identification of Genomic Islands by Comparative Analysis of the Contents and Contexts of tRNA Sites in Closely Related Bacteria.” Nucleic Acids Research 34 (1): e3.

Ozer, Egon A., Jonathan P. Allen, and Alan R. Hauser. 2014. “Characterization of the Core and Accessory Genomes of Pseudomonas Aeruginosa Using Bioinformatic Tools Spine and AGEnt.” BMC Genomics 15 (August): 737.

Paila, Umadevi, Brad A. Chapman, Rory Kirchner, and Aaron R. Quinlan. 2013. “GEMINI: Integrative Exploration of Genetic Variation and Genome Annotations.” PLoS Computational Biology 9 (7): e1003153.

Pal, Chandan, Johan Bengtsson-Palme, Christopher Rensing, Erik Kristiansson, and D. G. Joakim Larsson. 2014. “BacMet: Antibacterial Biocide and Metal Resistance Genes Database.” Nucleic Acids Research 42 (Database issue): D737–43.

Parks, Donovan H., Michael Imelfort, Connor T. Skennerton, Philip Hugenholtz, and Gene W. Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” Genome Research 25 (7): 1043–55.

Partridge, Sally R., Stephen M. Kwong, Neville Firth, and Slade O. Jensen. 2018. “Mobile Genetic Elements Associated with Antimicrobial Resistance.” Clinical Microbiology Reviews 31 (4). https://doi.org/10.1128/CMR.00088-17.

Pedregosa, Fabian. 2011. “Scikit-Learn: Machine Learning in Python.” Journal of Machine Learning Research: JMLR 12: 2825–30.

Pham, Nguyen-Phuong, Séverine Layec, Eric Dugat-Bony, Marie Vidal, Françoise Irlinger, and Christophe Monnet. 2017. “Comparative Genomic Analysis of Brevibacterium Strains: Insights into Key Genetic Determinants Involved in Adaptation to the Cheese Habitat.” BMC Genomics 18 (1): 955.

Pierce, Emily C., and Rachel J. Dutton. 2022. “Putting Microbial Interactions Back into Community Contexts.” Current Opinion in Microbiology 65 (February): 56–63.

Pierce, Emily C., Manon Morin, Jessica C. Little, Roland B. Liu, Joanna Tannous, Nancy P. Keller, Kit Pogliano, Benjamin E. Wolfe, Laura M. Sanchez, and Rachel J. Dutton. 2021. “Bacterial-Fungal Interactions Revealed by Genome-Wide Analysis of Bacterial Mutant Fitness.” Nature Microbiology 6 (1): 87–102.

Pierneef, Rian, Louis Cronje, Oliver Bezuidt, and Oleg N. Reva. 2015. “Pre_GI: A Global Map of Ontological Links between Horizontally Transferred Genomic Islands in Bacterial and Archaeal Genomes.” Database: The Journal of Biological Databases and Curation 2015 (June): bav058.

Pignatelli, Miguel, and Andrés Moya. 2011. “Evaluating the Fidelity of de Novo Short Read Metagenomic Assembly Using Simulated Data.” PloS One 6 (5): e19984.

Pinilla-Redondo, Rafael, Jakob Russel, David Mayo-Muñoz, Shiraz A. Shah, Roger A. Garrett, Joseph Nesme, Jonas S. Madsen, Peter C. Fineran, and Søren J. Sørensen. 2021. “CRISPR-Cas Systems Are Widespread Accessory Elements across Bacterial and Archaeal Plasmids.” Nucleic Acids Research, October. https://doi.org/10.1093/nar/gkab859.

Pinilla-Redondo, Rafael, Leise Riber, and Søren J. Sørensen. 2018. “Fluorescence Recovery Allows the Implementation of a Fluorescence Reporter Gene Platform Applicable for the Detection and Quantification of Horizontal Gene Transfer in Anoxic Environments.” Applied and Environmental Microbiology 84 (6). https://doi.org/10.1128/AEM.02507-17.

Pinilla-Redondo, Rafael, Valentine Cyriaque, Samuel Jacquiod, Søren J. Sørensen, and Leise Riber. 2018. “Monitoring Plasmid-Mediated Horizontal Gene Transfer in Microbiomes: Recent Advances and Future Perspectives.” Plasmid 99 (September): 56–67.

plasmIDent. n.d. Github. Accessed January 7, 2020. https://github.com/imgag/plasmIDent.

Podell, Sheila, and Terry Gaasterland. 2007. “DarkHorse: A Method for Genome-Wide Prediction of Horizontal Gene Transfer.” Genome Biology 8 (2): R16.

Podell, Sheila, Terry Gaasterland, and Eric E. Allen. 2008. “A Database of Phylogenetically Atypical Genes in Archaeal and Bacterial Genomes, Identified Using the DarkHorse Algorithm.” BMC Bioinformatics 9 (October): 419.

Pratama, Akbar Adjie, and Jan Dirk van Elsas. 2019. “Gene Mobility in Microbiomes of the Mycosphere and Mycorrhizosphere -Role of Plasmids and Bacteriophages.” FEMS Microbiology Ecology 95 (5). https://doi.org/10.1093/femsec/fiz053.

Press MO, Andrew H. Wiser, Zev N. Kronenberg, Kyle W. Langford, Migun Shakya, Chien-Chi Lo, Kathryn A. Mueller, Shawn T. Sullivan, Patrick S. G. Chain, and Ivan Liachko. 2017. “Hi-C Deconvolution of a Human Gut Microbiome Yields High-Quality Draft Genomes and Reveals Plasmid-Genome Interactions.” bioRxiv. https://doi.org/10.1101/198713.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments.” PloS One 5 (3): e9490.

Prjibelski, Andrey, Dmitry Antipov, Dmitry Meleshko, Alla Lapidus, and Anton Korobeynikov. 2020. “Using SPAdes De Novo Assembler.” Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.] 70 (1): e102.

Pundhir, Sachin, Hemant Vijayvargiya, and Anil Kumar. 2008. “PredictBias: A Server for the Identification of Genomic and Pathogenicity Islands in Prokaryotes.” In Silico Biology 8 (3-4): 223–34.

Quick, Josh. 2018. “Ultra-Long Read Sequencing Protocol for RAD004.” Protocols.io. January 22, 2018. https://doi.org/10.17504/protocols.io.mrxc57n.

Quigley, Lisa, Orla O’Sullivan, Tom P. Beresford, R. Paul Ross, Gerald F. Fitzgerald, and Paul D. Cotter. 2012. “High-Throughput Sequencing for Detection of Subpopulations of Bacteria Not Previously Associated with Artisanal Cheeses.” Applied and Environmental Microbiology 78 (16): 5717–23.

Quijada, Narciso M., Evelyne Mann, Martin Wagner, David Rodríguez-Lázaro, Marta Hernández, and Stephan Schmitz-Esser. 2018. “Autochthonous Facility-Specific Microbiota Dominates Washed-Rind Austrian Hard Cheese Surfaces and Its Production Environment.” International Journal of Food Microbiology 267 (February): 54–61.

Quijada, Narciso M., Stephan Schmitz-Esser, Benjamin Zwirzitz, Christian Guse, Cameron R. Strachan, Martin Wagner, Stefanie U. Wetzels, Evelyne Selberherr, and Monika Dzieciol. 2020. “Austrian Raw-Milk Hard-Cheese Ripening Involves Successional Dynamics of Non-Inoculated Bacteria and Fungi.” Foods (Basel, Switzerland) 9 (12). https://doi.org/10.3390/foods9121851.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” Bioinformatics 26 (6): 841–42.

Quintieri, Laura, Francesca Fanelli, and Leonardo Caputo. 2019. “Antibiotic Resistant Pseudomonas Spp. Spoilers in Fresh Dairy Products: An Underestimated Risk and the Control Strategies.” Foods (Basel, Switzerland) 8 (9). https://doi.org/10.3390/foods8090372.

Rajan, Issaac, Sarang Aravamuthan, and Sharmila S. Mande. 2007. “Identification of Compositionally Distinct Regions in Genomes Using the Centroid Method.” Bioinformatics 23 (20): 2672–77.

Rand, Arthur C., Miten Jain, Jordan M. Eizenga, Audrey Musselman-Brown, Hugh E. Olsen, Mark Akeson, and Benedict Paten. 2017. “Mapping DNA Methylation with High-Throughput Nanopore Sequencing.” Nature Methods 14 (4): 411–13.

Ravi, Anuradha, Lorena Valdés-Varela, Miguel Gueimonde, and Knut Rudi. 2018. “Transmission and Persistence of IncF Conjugative Plasmids in the Gut Microbiota of Full-Term Infants.” FEMS Microbiology Ecology 94 (1). https://doi.org/10.1093/femsec/fix158.

Ren, Jie, Nathan A. Ahlgren, Yang Young Lu, Jed A. Fuhrman, and Fengzhu Sun. 2017. “VirFinder: A Novel K-Mer Based Tool for Identifying Viral Sequences from Assembled Metagenomic Data.” Microbiome 5 (1): 69.

Richaume, Agnes, Eric Smit, Geneviève Faurie, and Jan D. van Elsas. 1992. “Influence of Soil Type on the Transfer of Plasmid RP4p from Pseudomonas Fluorescens to Introduced Recipient and to Indigenous Bacteria.” FEMS Microbiology Letters 101 (4): 281–91.

Roberts, Richard J., Tamas Vincze, Janos Posfai, and Dana Macelis. 2010. “REBASE—a Database for DNA Restriction and Modification: Enzymes, Genes and Genomes.” Nucleic Acids Research. https://doi.org/10.1093/nar/gkp874.

Robertson, James, and John H. E. Nash. 2018. “MOB-Suite: Software Tools for Clustering, Reconstruction and Typing of Plasmids from Draft Assemblies.” Microbial Genomics 4 (8). https://doi.org/10.1099/mgen.0.000206.

Robeson, Michael S., Devon R. O’Rourke, Benjamin D. Kaehler, Michal Ziemski, Matthew R. Dillon, Jeffrey T. Foster, and Nicholas A. Bokulich. 2020. “RESCRIPt: Reproducible Sequence Taxonomy Reference Database Management for the Masses.” bioRxiv. https://doi.org/10.1101/2020.10.05.326504.

Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, et al. 2000. “Cloning the Soil Metagenome: A Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms.” Applied and Environmental Microbiology 66 (6): 2541–47.

Roosaare, Märt, Mikk Puustusmaa, Märt Möls, Mihkel Vaher, and Maido Remm. 2018. “PlasmidSeeker: Identification of Known Plasmids from Bacterial Whole Genome Sequencing Reads.” PeerJ 6 (April): e4588.

Ropars, Jeanne, Ricardo C. Rodríguez de la Vega, Manuela López-Villavicencio, Jérôme Gouzy, Erika Sallet, Émilie Dumas, Sandrine Lacoste, et al. 2015. “Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi.” Current Biology: CB 25 (19): 2562–69.

Rossi, Franca, Lucia Rizzotti, Giovanna E. Felis, and Sandra Torriani. 2014. “Horizontal Gene Transfer among Microorganisms in Food: Current Knowledge and Future Perspectives.” Food Microbiology 42 (September): 232–43.

Roux, Simon, Francois Enault, Bonnie L. Hurwitz, and Matthew B. Sullivan. 2015. “VirSorter: Mining Viral Signal from Microbial Genomic Data.” PeerJ. https://doi.org/10.7717/peerj.985.

Rozov, Roye, Aya Brown Kav, David Bogumil, Naama Shterzer, Eran Halperin, Itzhak Mizrahi, and Ron Shamir. 2017. “Recycler: An Algorithm for Detecting Plasmids from de Novo Assembly Graphs.” Bioinformatics 33 (4): 475–82.

Russell, Alistair B., S. Brook Peterson, and Joseph D. Mougous. 2014. “Type VI Secretion System Effectors: Poisons with a Purpose.” Nature Reviews. Microbiology 12 (2): 137–48.

Saak, Christina C., Cong B. Dinh, and Rachel J. Dutton. 2020. “Experimental Approaches to Tracking Mobile Genetic Elements in Microbial Communities.” FEMS Microbiology Reviews 44 (5): 606–30.

Sayers, Samantha, Li Li, Edison Ong, Shunzhou Deng, Guanghua Fu, Yu Lin, Brian Yang, et al. 2019. “Victors: A Web-Based Knowledge Base of Virulence Factors in Human and Animal Pathogens.” Nucleic Acids Research 47 (D1): D693–700.

Schmidt, K., S. Mwaigwisya, L. C. Crossman, M. Doumith, D. Munroe, C. Pires, A. M. Khan, et al. 2017. “Identification of Bacterial Pathogens and Antimicrobial Resistance Directly from Clinical Urines by Nanopore-Based Metagenomic Sequencing.” The Journal of Antimicrobial Chemotherapy 72 (1): 104–14.

Schmitz-Esser, Stephan, Monika Dzieciol, Eva Nischler, Elisa Schornsteiner, Othmar Bereuter, Evelyne Mann, and Martin Wagner. 2018. “Abundance and Potential Contribution of Gram-Negative Cheese Rind Bacteria from Austrian Artisanal Hard Cheeses.” International Journal of Food Microbiology. https://doi.org/10.1016/j.ijfoodmicro.2017.11.013.

Schornsteiner, Elisa, Evelyne Mann, Othmar Bereuter, Martin Wagner, and Stephan Schmitz-Esser. 2014. “Cultivation-Independent Analysis of Microbial Communities on Austrian Raw Milk Hard Cheese Rinds.” International Journal of Food Microbiology 180 (June): 88–97.

Seemann, Torsten. n.d. Abricate. Github. Accessed December 16, 2019. https://github.com/tseemann/abricate.

Seiler, Enrico, Kathrin Trappe, and Bernhard Y. Renard. 2019. “Where Did You Come From, Where Did You Go: Refining Metagenomic Analysis Tools for Horizontal Gene Transfer Characterisation.” PLoS Computational Biology 15 (7): e1007208.

Sengeløv, G., and S. J. Sørensen. 1998. “Methods for Detection of Conjugative Plasmid Transfer in Aquatic Environments.” Current Microbiology 37 (4): 274–80.

Sentchilo, Vladimir, Antonia P. Mayer, Lionel Guy, Ryo Miyazaki, Susannah Green Tringe, Kerrie Barry, Stephanie Malfatti, Alexander Goessmann, Marc Robinson-Rechavi, and Jan R. van der Meer. 2013. “Community-Wide Plasmid Gene Mobilization and Selection.” The ISME Journal 7 (6): 1173–86.

Shaiber, Alon, Amy D. Willis, Tom O. Delmont, Simon Roux, Lin-Xing Chen, Abigail C. Schmid, Mahmoud Yousef, et al. 2020. “Functional and Genetic Markers of Niche Partitioning among Enigmatic Members of the Human Oral Microbiome.” Genome Biology 21 (1): 292.

Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. “SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.” PloS One 11 (10): e0163962.

Sheppard, Anna E., Nicole Stoesser, Daniel J. Wilson, Robert Sebra, Andrew Kasarskis, Luke W. Anson, Adam Giess, et al. 2016. “Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC.” Antimicrobial Agents and Chemotherapy 60 (6): 3767–78.

Shi, Jieming, and Chun Liang. 2019. “Generic Repeat Finder: A High-Sensitivity Tool for Genome-Wide De Novo Repeat Detection.” Plant Physiology 180 (4): 1803–15.

Shrivastava, Sakshi, Ch V. Siva Kumar Reddy, and Sharmila S. Mande. 2010. “INDeGenIUS, a New Method for High-Throughput Identification of Specialized Functional Islands in Completely Sequenced Organisms.” Journal of Biosciences 35 (3): 351–64.

Sieber, Christian M. K., Alexander J. Probst, Allison Sharrar, Brian C. Thomas, Matthias Hess, Susannah G. Tringe, and Jillian F. Banfield. 2018. “Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy.” Nature Microbiology 3 (7): 836–43.

Sieuwerts, Sander, Frank A. M. de Bok, Jeroen Hugenholtz, and Johan E. T. van Hylckama Vlieg. 2008. “Unraveling Microbial Interactions in Food Fermentations: From Classical to Genomics Approaches.” Applied and Environmental Microbiology 74 (16): 4997–5007.

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon, and M. Chandler. 2006. “ISfinder: The Reference Centre for Bacterial Insertion Sequences.” Nucleic Acids Research 34 (Database issue): D32–36.

Siguier, Patricia, Edith Gourbeyre, and Michael Chandler. 2017. “Known Knowns, Known Unknowns and Unknown Unknowns in Prokaryotic Transposition.” Current Opinion in Microbiology 38 (August): 171–80.

Silva Filho, Antonio Camilo da, Roberto Tadeu Raittz, Dieval Guizelini, Camilla Reginatto De Pierri, Diônata Willian Augusto, Izabella Castilhos Ribeiro Dos Santos-Weiss, and Jeroniza Nunes Marchaukoski. 2018. “Comparative Analysis of Genomic Island Prediction Tools.” Frontiers in Genetics 9 (December): 619.

Simonis, Marieke, Petra Klous, Erik Splinter, Yuri Moshkin, Rob Willemsen, Elzo de Wit, Bas van Steensel, and Wouter de Laat. 2006. “Nuclear Organization of Active and Inactive Chromatin Domains Uncovered by Chromosome Conformation Capture-on-Chip (4C).” Nature Genetics 38 (11): 1348–54.

Simpson, Jared T., Rachael E. Workman, P. C. Zuzarte, Matei David, L. J. Dursi, and Winston Timp. 2017. “Detecting DNA Cytosine Methylation Using Nanopore Sequencing.” Nature Methods 14 (4): 407–10.

Sitaraman, Ramakrishnan. 2018. “Prokaryotic Horizontal Gene Transfer within the Human Holobiont: Ecological-Evolutionary Inferences, Implications and Possibilities.” Microbiome 6 (1): 163.

Smalla, K., E. Krögerrecklenfort, H. Heuer, W. Dejonghe, E. Top, M. Osborn, J. Niewint, et al. 2000. “PCR-Based Detection of Mobile Genetic Elements in Total Community DNA.” Microbiology 146 (Pt 6) (June): 1256–57.

Smalla, K., H. Heuer, A. Götz, D. Niemeyer, E. Krögerrecklenfort, and E. Tietze. 2000. “Exogenous Isolation of Antibiotic Resistance Plasmids from Piggery Manure Slurries Reveals a High Prevalence and Diversity of IncQ-like Plasmids.” Applied and Environmental Microbiology 66 (11): 4854–62.

Smillie, Chris, M. Pilar Garcillán-Barcia, M. Victoria Francia, Eduardo P. C. Rocha, and Fernando de la Cruz. 2010. “Mobility of Plasmids.” Microbiology and Molecular Biology Reviews: MMBR 74 (3): 434–52.

Smit, E., J. D. van Elsas, J. A. van Veen, and W. M. de Vos. 1991. “Detection of Plasmid Transfer from Pseudomonas Fluorescens to Indigenous Bacteria in Soil by Using Bacteriophage phiR2f for Donor Counterselection.” Applied and Environmental Microbiology 57 (12): 3482–88.

Soares, Siomar C., Hakan Geyik, Rommel T. J. Ramos, Pablo H. C. G. de Sá, Eudes G. V. Barbosa, Jan Baumbach, Henrique C. P. Figueiredo, et al. 2016. “GIPSy: Genomic Island Prediction Software.” Journal of Biotechnology 232 (August): 2–11.

Somerville, Vincent, Stefanie Lutz, Michael Schmid, Daniel Frei, Aline Moser, Stefan Irmler, Jürg E. Frey, and Christian H. Ahrens. 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 Microbiology 19 (1): 143.

Song, Chun-Xiao, Keith E. Szulwach, Qing Dai, Ye Fu, Shi-Qing Mao, Li Lin, Craig Street, et al. 2013. “Genome-Wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming.” Cell 153 (3): 678–91.

Song, Weizhi, Bernd Wemheuer, Shan Zhang, Kerrin Steensen, and Torsten Thomas. 2019. “MetaCHIP: Community-Level Horizontal Gene Transfer Identification through the Combination of Best-Match and Phylogenetic Approaches.” Microbiome 7 (1): 36.

Sørensen, Søren J., Mark Bailey, Lars H. Hansen, Niels Kroer, and Stefan Wuertz. 2005. "Studying Plasmid Horizontal Transfer in Situ: A Critical Review." Nature Reviews. Microbiology 3 (9): 700–710.

Spencer, Sarah J., Manu V. Tamminen, Sarah P. Preheim, Mira T. Guo, Adrian W. Briggs, Ilana L. Brito, David A Weitz, et al. 2016. "Massively Parallel Sequencing of Single Cells by epicPCR Links Functional Genes with Phylogenetic Markers." The ISME Journal 10 (2): 427–36.

Stalder, Thibault, Maximilian O. Press, Shawn Sullivan, Ivan Liachko, and Eva M. Top. 2019. "Linking the Resistome and Plasmidome to the Microbiome." The ISME Journal 13 (10): 2437–46.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." Bioinformatics 30 (9): 1312–13.

Stepanauskas, Ramunas. 2015. "Wiretapping into Microbial Interactions by Single Cell Genomics." Frontiers in Microbiology 6 (April): 258.

Stewart, Robert D., Marc D. Auffret, Amanda Warr, Andrew H. Wiser, Maximilian O. Press, Kyle W. Langford, Ivan Liachko, et al. 2018. "Assembly of 913 Microbial Genomes from Metagenomic Sequencing of the Cow Rumen." Nature Communications 9 (1): 870.

Stoiber, Marcus, Joshua Quick, Rob Egan, Ji Eun Lee, Susan Celniker, Robert K. Neely, Nicholas Loman, Len A. Pennacchio, and James Brown. 2017. "De Novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing." bioRxiv. https://doi.org/10.1101/094672.

Tamang, Jyoti P., Koichi Watanabe, and Wilhelm H. Holzapfel. 2016. "Review: Diversity of Microorganisms in Global Fermented Foods and Beverages." Frontiers in Microbiology 7 (March): 377.

Tampuu, Ardi, Zurab Bzhalava, Joakim Dillner, and Raul Vicente. 2019. "ViraMiner: Deep Learning on Raw DNA Sequences for Identifying Viral Genomes in Human Samples." PloS One 14 (9): e0222271.

Tang, Yuan, Masaaki Horikoshi, and Wenxuan Li. 2016. "Ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages." The R Journal 8 (2): 474.

Tansirichaiya, Supathep, Md Ajijur Rahman, and Adam P. Roberts. 2019. "The Transposon Registry." Mobile DNA 10 (October): 40.

Tasse, Lena, Juliette Bercovici, Sandra Pizzut-Serin, Patrick Robe, Julien Tap, Christophe Klopp, Brandi L. Cantarel, et al. 2010. "Functional Metagenomics to Mine the Human Gut Microbiome for Dietary Fiber Catabolic Enzymes." Genome Research 20 (11): 1605–12.

Tithi, Saima Sultana, Frank O. Aylward, Roderick V. Jensen, and Liqing Zhang. 2018. "FastViromeExplorer: A Pipeline for Virus and Phage Identification and Abundance Profiling in Metagenomics Data." PeerJ 6 (January): e4227.

Top, E., M. Mergeay, D. Springael, and W. Verstraete. 1990. "Gene Escape Model: Transfer of Heavy Metal Resistance Genes from Escherichia Coli to Alcaligenes Eutrophus on Agar Plates and in Soil Samples." Applied and Environmental Microbiology 56 (8): 2471–79.

Toro, Esteban, Sun-Hae Hong, Harley H. McAdams, and Lucy Shapiro. 2008. "Caulobacter Requires a Dedicated Mechanism to Initiate Chromosome Segregation." Proceedings of the National Academy of Sciences of the United States of America 105 (40): 15435–40.

Tsirigos, Aristotelis, and Isidore Rigoutsos. 2005. "A Sensitive, Support-Vector-Machine Method for the Detection of Horizontal Gene Transfers in Viral, Archaeal and Bacterial Genomes." Nucleic Acids Research 33 (12): 3699–3707.

Tu, Qiang, and Dafu Ding. 2003. "Detecting Pathogenicity Islands and Anomalous Gene Clusters by Iterative Discriminant Analysis." FEMS Microbiology Letters 221 (2): 269–75.

Umbarger, Mark A., Esteban Toro, Matthew A. Wright, Gregory J. Porreca, Davide Baù, Sun-Hae Hong, Michael J. Fero, et al. 2011. "The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation." Molecular Cell 44 (2): 252–64.

UniProt Consortium. 2019. "UniProt: A Worldwide Hub of Protein Knowledge." Nucleic Acids Research 47 (D1): D506–15.

Vallenet, D., S. Engelen, D. Mornico, S. Cruveiller, L. Fleury, A. Lajus, Z. Rouy, et al. 2009. "MicroScope: A Platform for Microbial Genome Annotation and Comparative Genomics." Database: The Journal of Biological Databases and Curation 2009 (November): bap021.

Varani, Alessandro M., Patricia Siguier, Edith Gourbeyre, Vincent Charneau, and Mick Chandler. 2011. "ISsaga Is an Ensemble of Web-Based Methods for High Throughput Identification and Semi-Automatic Annotation of Insertion Sequences in Prokaryotic Genomes." Genome Biology 12 (3): R30.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 2017. "Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads." Genome Research 27 (5): 737–46.

Vassetzky, Nikita S., and Dmitri A. Kramerov. 2013. "SINEBase: A Database and Tool for SINE Analysis." Nucleic Acids Research 41 (Database issue): D83–89.

Vermeulen, Carlo, Amin Allahyar, Britta A. M. Bouwman, Peter H. L. Krijger, Marjon J. A. M. Verstegen, Geert Geeven, Christian Valdes-Quezada, et al. 2020. "Multi-Contact 4C: Long-Molecule Sequencing of Complex Proximity Ligation Products to Uncover Local Cooperative and Competitive Chromatin Topologies." Nature Protocols 15 (2): 364–97.

Vernikos, Georgios S, Juliuan Parkhill. 2008. "Resolving the Structural Features of Genomic Islands: A Machine Learning Approach." Genome Research 18 (2): 331–42.

Vernikos, Georgios S., and Julian Parkhill. 2006. “Interpolated Variable Order Motifs for Identification of Horizontally Acquired DNA: Revisiting the Salmonella Pathogenicity Islands.” Bioinformatics 22 (18): 2196–2203.

Waack, Stephan, Oliver Keller, Roman Asper, Thomas Brodag, Carsten Damm, Wolfgang Florian Fricke, Katharina Surovcik, Peter Meinicke, and Rainer Merkl. 2006. “Score-Based Prediction of Genomic Islands in Prokaryotic Genomes Using Hidden Markov Models.” BMC Bioinformatics 7 (March): 142.

Waldor, M. K., and J. J. Mekalanos. 1996. “Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin.” Science.

Walsh, Aaron M., Fiona Crispie, Kareem Daari, Orla O’Sullivan, Jennifer C. Martin, Cornelius T. Arthur, Marcus J. Claesson, Karen P. Scott, and Paul D. Cotter. 2017. “Strain-Level Metagenomic Analysis of the Fermented Dairy Beverage Nunu Highlights Potential Food Safety Risks.” Applied and Environmental Microbiology 83 (16). https://doi.org/10.1128/AEM.01144-17.

Walsh, Aaron M., Fiona Crispie, Kieran Kilcawley, Orla O’Sullivan, Maurice G. O’Sullivan, Marcus J. Claesson, and Paul D. Cotter. 2016. “Microbial Succession and Flavor Production in the Fermented Dairy Beverage Kefir.” mSystems 1 (5). https://doi.org/10.1128/mSystems.00052-16.

Warburton, Philip J., Elaine Allan, Stephanie Hunter, John Ward, Veronica Booth, William G. Wade, and Peter Mullany. 2011. “Isolation of Bacterial Extrachromosomal DNA from Human Dental Plaque Associated with Periodontal Disease, Using Transposon-Aided Capture (TRACA).” FEMS Microbiology Ecology 78 (2): 349–54.

Wattam, Alice R., James J. Davis, Rida Assaf, Sébastien Boisvert, Thomas Brettin, Christopher Bun, Neal Conrad, et al. 2017. “Improvements to PATRIC, the All-Bacterial Bioinformatics Database and Analysis Resource Center.” Nucleic Acids Research 45 (D1): D535–42.

Wei, Wen, Feng Gao, Meng-Ze Du, Hong-Li Hua, Ju Wang, and Feng-Biao Guo. 2017. “Zisland Explorer: Detect Genomic Islands by Combining Homogeneity and Heterogeneity Properties.” Briefings in Bioinformatics 18 (3): 357–66.

Weilguny, Lukas, and Robert Kofler. 2019. “DeviaTE: Assembly-free Analysis and Visualization of Mobile Genetic Element Composition.” Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13030.

Westbye, Alexander B., J. Thomas Beatty, and Andrew S. Lang. 2017. “Guaranteeing a Captive Audience: Coordinated Regulation of Gene Transfer Agent (GTA) Production and Recipient Capability by Cellular Regulators.” Current Opinion in Microbiology 38 (August): 122–29.

Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, et al. 2007. “A Unified Classification System for Eukaryotic Transposable Elements.” Nature Reviews. Genetics 8 (12): 973–82.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the Tidyverse.” Journal of Open Source Software 4 (43): 1686.

Wiedenbeck, Jane, and Frederick M. Cohan. 2011. “Origins of Bacterial Diversity through Horizontal Genetic Transfer and Adaptation to New Ecological Niches.” FEMS Microbiology Reviews 35 (5): 957–76.

Wilbanks, Elizabeth G., Hugo Doré, Meredith H. Ashby, Cheryl Heiner, Richard J. Roberts, and Jonathan A. Eisen. 2022. “Metagenomic Methylation Patterns Resolve Bacterial Genomes of Unusual Size and Structural Complexity.” The ISME Journal, April. https://doi.org/10.1038/s41396-022-01242-7.

Wolfe, Benjamin E., and Rachel J. Dutton. 2015. “Fermented Foods as Experimentally Tractable Microbial Ecosystems.” Cell 161 (1): 49–55.

Wolfe, Benjamin E., Julie E. Button, Marcela Santarelli, and Rachel J. Dutton. 2014. “Cheese Rind Communities Provide Tractable Systems for in Situ and in Vitro Studies of Microbial Diversity.” Cell 158 (2): 422–33.

Wolfe, Benjamin E., Julie E. Button, Marcela Santarelli, and Rachel J. Dutton. 2014. “Cheese Rind Communities Provide Tractable Systems for In Situ and In Vitro Studies of Microbial Diversity.” https://doi.org/10.1016/j.cell.2014.05.041.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. “Improved Metagenomic Analysis with Kraken 2.” Genome Biology 20 (1): 257.

Wozniak, Rachel A. F., and Matthew K. Waldor. 2010. “Integrative and Conjugative Elements: Mosaic Mobile Genetic Elements Enabling Dynamic Lateral Gene Flow.” Nature Reviews. Microbiology 8 (8): 552–63.

Wu, Yu-Wei, Blake A. Simmons, and Steven W. Singer. 2016. “MaxBin 2.0: An Automated Binning Algorithm to Recover Genomes from Multiple Metagenomic Datasets.” Bioinformatics. https://doi.org/10.1093/bioinformatics/btv638.

Xavier, Basil Britto, Anupam J. Das, Guy Cochrane, Sandra De Ganck, Samir Kumar-Singh, Frank Møller Aarestrup, Herman Goossens, and Surbhi Malhotra-Kumar. 2016. “Consolidating and Exploring Antibiotic Resistance Gene Data Resources.” Journal of Clinical Microbiology 54 (4): 851–59.

Xia, Yu, An-Dong Li, Yu Deng, Xiao-Tao Jiang, Li-Guan Li, and Tong Zhang. 2017. “MinION Nanopore Sequencing Enables Correlation between Resistome Phenotype and

Genotype of Coliform Bacteria in Municipal Sewage." Frontiers in Microbiology 8 (October): 2105.

Xie, Zhiqun, and Haixu Tang. 2017. "ISEScan: Automated Identification of Insertion Sequence Elements in Prokaryotic Genomes." Bioinformatics 33 (21): 3340–47.

Xiong, Dapeng, Fen Xiao, Li Liu, Kai Hu, Yanping Tan, Shunmin He, and Xieping Gao. 2012. "Towards a Better Detection of Horizontally Transferred Genes by Combining Unusual Properties Effectively." PloS One 7 (8): e43126.

Yaffe, Eitan, and David A. Relman. 2019. "Tracking Microbial Evolution in the Human Gut Using Hi-C Reveals Extensive Horizontal Gene Transfer, Persistence and Adaptation." Nature Microbiology, December. https://doi.org/10.1038/s41564-019-0625-0.

Yin, Xiaole, Xiao-Tao Jiang, Benli Chai, Liguan Li, Ying Yang, James R. Cole, James M. Tiedje, and Tong Zhang. 2018. "ARGs-OAP v2.0 with an Expanded SARG Database and Hidden Markov Models for Enhancement Characterization and Quantification of Antibiotic Resistance Genes in Environmental Metagenomes." Bioinformatics 34 (13): 2263–70.

Yoon, Sung Ho, Cheol-Goo Hur, Ho-Young Kang, Yeoun Hee Kim, Tae Kwang Oh, and Jihyun F. Kim. 2005. "A Computational Approach for Identifying Pathogenicity Islands in Prokaryotic Genomes." BMC Bioinformatics 6 (July): 184.

Yoon, Sung Ho, Young-Kyu Park, and Jihyun F. Kim. 2015. "PAIDB v2.0: Exploration and Analysis of Pathogenicity and Resistance Islands." Nucleic Acids Research 43 (Database issue): D624–30.

Youngblut, Nicholas D., Jacobo de la Cuesta-Zuluaga, Georg H. Reischer, Silke Dauser, Nathalie Schuster, Chris Walzer, Gabrielle Stalder, Andreas H. Farnleitner, and Ruth E. Ley. 2020. "Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity." mSystems 5 (6). https://doi.org/10.1128/mSystems.01045-20.

Yu, Miao, Gary C. Hon, Keith E. Szulwach, Chun-Xiao Song, Liang Zhang, Audrey Kim, Xuekun Li, et al. 2012. "Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome." Cell 149 (6): 1368–80.

Yuan, Yihui, and Meiying Gao. 2017. "Jumbo Bacteriophages: An Overview." Frontiers in Microbiology 8 (March): 403.

Zablocki, Olivier, Ho Bin Jang, Benjamin Bolduc, and Matthew B. Sullivan. 2019. "vConTACT 2: A Tool to Automate Genome-Based Prokaryotic Viral Taxonomy." In Plant and Animal Genome XXVII Conference (January 12-16, 2019). PAG. https://pag.confex.com/pag/xxvii/meetingapp.cgi/Paper/36846.

Zankari, Ea, Henrik Hasman, Salvatore Cosentino, Martin Vestergaard, Simon Rasmussen, Ole Lund, Frank M. Aarestrup, and Mette Voldby Larsen. 2012. "Identification of Acquired

Antimicrobial Resistance Genes." The Journal of Antimicrobial Chemotherapy 67 (11): 2640–44.

Zankari, Ea, Rosa Allesøe, Katrine G. Joensen, Lina M. Cavaco, Ole Lund, and Frank M. Aarestrup. 2017. "PointFinder: A Novel Web Tool for WGS-Based Detection of Antimicrobial Resistance Associated with Chromosomal Point Mutations in Bacterial Pathogens." The Journal of Antimicrobial Chemotherapy 72 (10): 2764–68.

Zhang, An Ni, Li-Guan Li, Liping Ma, Michael R. Gillings, James M. Tiedje, and Tong Zhang. 2018. "Conserved Phylogenetic Distribution and Limited Antibiotic Resistance of Class 1 Integrons Revealed by Assessing the Bacterial Genome and Plasmid Collection." Microbiome 6 (1): 130.

Zhang, Ren, and Chun-Ting Zhang. 2004. "A Systematic Method to Identify Genomic Islands and Its Applications in Analyzing the Genomes of Corynebacterium Glutamicum and Vibrio Vulnificus CMCP6 Chromosome I." Bioinformatics 20 (5): 612–22.

Zhang, Ren, Hong-Yu Ou, Feng Gao, and Hao Luo. 2014. "Identification of Horizontally-Transferred Genomic Islands and Genome Segmentation Points by Using the GC Profile Method." Current Genomics 15 (2): 113–21.

Zhang, Shuhong, Guangli Yang, Shugui Hou, Tingjun Zhang, Zhiguo Li, and Feng Liang. 2018. "Distribution of ARGs and MGEs among Glacial Soil, Permafrost, and Sediment Using Metagenomic Analysis." Environmental Pollution 234 (March): 339–46.

Zhang, Yuanchen, Erik K. Kastman, Jeffrey S. Guasto, and Benjamin E. Wolfe. 2018. "Fungal Networks Shape Dynamics of Bacterial Dispersal and Community Assembly in Cheese Rind Microbiomes." Nature Communications 9 (1): 336.

Zhao, Zhihu, Gholamreza Tavoosidana, Mikael Sjölinder, Anita Göndör, Piero Mariano, Sha Wang, Chandrasekhar Kanduri, et al. 2006. "Circular Chromosome Conformation Capture (4C) Uncovers Extensive Networks of Epigenetically Regulated Intra- and Interchromosomal Interactions." Nature Genetics 38 (11): 1341–47.

Zhou, Fengfeng, and Ying Xu. 2010. "cBar: A Computer Program to Distinguish Plasmid-Derived from Chromosome-Derived Sequence Fragments in Metagenomics Data." Bioinformatics 26 (16): 2051–52.

Zhou, You, Yongjie Liang, Karlene H. Lynch, Jonathan J. Dennis, and David S. Wishart. 2011. "PHAST: A Fast Phage Search Tool." Nucleic Acids Research 39 (Web Server issue): W347–52.

Zhu, Shijia, John Beaulaurier, Gintaras Deikus, Tao P. Wu, Maya Strahl, Ziyang Hao, Guanzheng Luo, et al. 2018. "Mapping and Characterizing N6-Methyladenine in Eukaryotic Genomes Using Single-Molecule Real-Time Sequencing." Genome Research 28 (7): 1067–78.