Title

Extensive Horizontal Gene Transfer in Cheese-Associated Bacteria

Author Names and Affiliations

Kevin S. Bonham[1], Benjamin E. Wolfe[2], Rachel J. Dutton[1, 3].

(1) Division of Biological Sciences, University of California, San Diego, La Jolla, CA,

(2) Department of Biology, Tufts University, Medford, MA

(3) Center for Microbiome Innovation, Jacobs School of Engineering, University of California, San Diego, La Jolla, CA

Abstract

Acquisition of genes through horizontal gene transfer (HGT) allows microbes to rapidly gain new capabilities and adapt to new or changing environments. Identifying widespread HGT regions within multispecies microbiomes can pinpoint the molecular mechanisms that play key roles in microbiome assembly. We sought to identify horizontally transferred genes within a model microbiome, the cheese rind. Comparing 31 newly-sequenced and 134 previously sequenced bacterial isolates from cheese rinds, we identified over 200 putative horizontally transferred genomic regions containing 4,844 protein coding genes. The largest of these regions are enriched for genes involved in siderophore acquisition, and are widely distributed in cheese rinds in both Europe and the US. These results suggest that horizontal gene transfer (HGT) is prevalent in cheese rind microbiomes, and the identification of genes that are frequently transferred in a particular environment may provide insight into the selective forces shaping microbial communities.

Introduction

Great strides have been made in characterizing the composition of microbiomes, and in understanding their importance in the ecology of natural systems, in agriculture and in human health. However, despite these advances, the forces that shape the diversity, structure, and function of microbiomes remain poorly understood [1]. Investigating these underlying mechanisms *in situ* is difficult, since observational and sequenced-based analysis rarely enables causal conclusions [2]. Replicating microbial communities *in vitro* is also an enormous challenge, due to high levels of diversity and the difficulties in establishing pure cultures of most bacterial species. These obstacles significantly hamper our ability to move from observations of microbial diversity to the molecular mechanisms shaping key processes such as species interactions and microbial evolution.

Horizontal gene transfer (HGT) is a major force in microbial evolution and can lead to the wholesale acquisition of novel functions. In some cases, these novel functions can have significant adaptive consequences, such as in the transfer of antibiotic resistance genes [3]. HGT also allows rapid adaptation to new niches [4], since ecologically relevant genes may be acquired by species not previously adapted to a particular niche [5,6]. The movement of microbes to new environments has been shown to increase both the rate and impact of HGT, and HGT is most frequent for genes under positive selection [7]. In moving to a new environment, microbes can face novel abiotic conditions (temperature, moisture, salinity, pH, and nutrients) and novel biotic challenges and opportunities due to the presence of microbial neighbors.

Evaluating HGT within the context of microbial communities has the potential to uncover new insights concerning the extent, mechanisms, and ecological impact of this important process. Advances in genome sequencing have begun to provide a glimpse into HGT within environmentally, medically and economically important microbiomes [8,9]. For example, extensive gene sharing has been observed throughout the commensal human microbiome [6,10]), including genes that enable nutrient acquisition from novel food sources [6,10], and pathogenicity islands and antibiotic resistance genes in pathogenic microbes [11–13]. While these studies offer valuable insights into the frequency and potential impact of genes that can be transferred in microbial communities, the complexity of these systems makes further examination of the effects of these HGT events on their evolution and ecology difficult.

The microbial communities of fermented foods experience strong selection as a result of growing in novel, human-made environments. Previous work has demonstrated that HGT can be a major driver of adaptation in food systems and other human-managed environments [9,14]. Prior analysis of microbial species from cheese have revealed several instances of HGT in this environment. Lactic acid bacteria (LAB) such as *Lactobacillus* and *Lactococcus*, which are used in the initial fermentation of milk, are known to harbor antibiotic resistance genes and may be reservoirs for transfer to pathogenic enterococci [15,16] and other pathogenic microbes. Other food-associated bacteria may also contribute to antibiotic resistance gene transfer [17–19]. In yogurt, another dairy ferment utilizing LAB, HGT of metabolic genes between protocooperative species *L. bulgaricus* and *S. thermophilus* has been observed [20,21]. Sequencing of *Penicillium* species isolated from the cheese-making environment identified HGT of large genomic islands between these key fungal inhabitants of cheese [22–24].

During the aging of cheese, bacteria and fungi form a multi-species biofilm called the rind [25]. We have previously shown that these communities can be used to examine community-based processes,such as succession and interspecies interactions, within an experimentally tractable system [26,27]. Given that biofilms such as these are densely populated, and microbes in cheese rinds are under strong selection to obtain limited nutrients (e.g. free amino acids, iron) as well as tolerate cell stress [28], we predicted that HGT might be widespread in cheese rind microbiomes and therefore might provide a useful experimental model for HGT within microbial communities.

We sought to determine the diversity, distribution, and functional content of HGT in bacterial species isolated from cheese rinds. Specifically, we predicted that 1) HGT would be widespread, 2) that HGT genes would be enriched for functions related to survival in cheese environment, and 3) that there would be uneven distribution of HGT events across taxa. We analyzed the genomes of newly isolated and sequenced cheese-associated bacterial species (31 genomes) and those available in public databases (134 additional genomes). We present data which suggests that there has been extensive HGT in cheese-associated bacteria. The regions of DNA identified appear to encode a number of functions which would be expected to provide adaptive advantages within the cheese environment. In particular, we identified three large multi-gene islands that are shared within multiple Actinobacteria, Proteobacteria and Firmicutes species respectively. These genomic regions are not related, but appear to have analogous functions involving iron acquisition, and are widely distributed in geographically distant cheeses. This work provides foundational knowledge in an experimentally-tractable system in which future work can help to provide insight on the role of HGT within microbiomes.

Results

Identification of putative horizontally transferred regions

To establish a diverse database of cheese-associated bacterial genomes, we isolated species from cheese samples collected as part of previous work [26]. A total of 31 isolates, representing 4 bacterial phyla and 11 genera, were selected for genome sequencing using Illumina and PacBio (Supplementary Table 1). Recently, a large collection of cheese-associated bacterial genomes were sequenced [29], which allowed us to include additional genomes in our analysis. Our isolates were from cheeses produced in the United States, Spain, Italy and France, while the Almeida et al. collection was almost exclusively from France. We also included genomes from the NCBI reference sequence (RefSeq) database that are associated with cheese, for a total of 165 bacterial genomes.

We next developed a computational pipeline for the identification of putative horizontally transferred genes adapted from work on the human microbiome [10]. We built a central BLAST database containing all ORFs from all cheese-associated genomes. For each gene in each genome, we performed BLAST against this database, and compiled a list of hits (Supplementary Figure 1, Materials and Methods). For each hit, we examined the length and percent identity of aligned regions. Closely related species will have many nearly identical genes due to vertical inheritance. To avoid capturing these genes in our analysis, we determined the pairwise average nucleotide identity (ANI) between species within the same genus [30,31]. ANI provides a measure of the overall similarity of two genomes. We tested varying thresholds for length and ANI in order to examine the effects of these parameters on the results (Supplementary Figure 2). Higher maximum ANI cutoffs and shorter lengths are more likely to yield false positives, since closely related species are more likely to share short stretches of high nucleotide identity. At the same time, a lower maximum ANI cutoff may exclude legitimate HGT events, especially considering that closely related species are also more likely to engage in HGT. Based on our most conservative gene identity parameters

(minimum 99% identity over 500 nucleotides), we identified at least one putative horizontally transferred gene in 130/165 cheese-associated species in the analysis, for a total of 4,844 genes (Figure 1A, Supplementary Table 3). At least one putative HGT protein coding gene was found in 130 out of 165 species (78.8%). Because this analysis depends on the species included for comparison, this list of HGT is almost certainly an underestimate.

Since multiple genes can be transferred in a single HGT event, we next assembled the putative HGT genes into groups based on genomic proximity. Individual coding sequences (CDS) for each species were grouped into islands if they were found within 5000 nucleotides of one another on the same contig. These islands were then clustered with islands in other species if they shared at least one CDS in common. The 4,844 genes clustered into 259 individual groups (Figure 1B, Supplementary Table 3). Mobile elements such as transposons complicate our method of group clustering, since non-contiguous islands may be grouped together if they share a common transposon. Indeed, this appears to have occurred with Group 1, which contains genes from several disparate genomic regions. In other cases, a single species may have genes in a single group spread across multiple contigs (Supplementary Figure 3B), but this may accurately represent a single HGT event.

Most HGT groups we identified (234, 90%) contain only members of the same phylum, or even a single genus (180 or 69%, Supplementary Table 4). This supports previous studies that suggest that HGT is most prevalent among closely related species [32]. However, we uncovered several notable exceptions. For example, *Alkaibacterium kapii* FAM208.38, a Firmicute, has a substantial (~8kb) fraction of Group 1, which is predominantly found in *Actinobacteria* species. Groups 2 and 3 each have hits found in 3 different phyla (though both are predominantly found in Actinobacteria).

Functions encoded in HGT regions

HGT can enable rapid evolution of microbes entering a new environment, and genes that are under positive selection are observed more frequently [4]. Identifying the functions of genes that are frequently transferred may provide a window into the selective forces that are most important for adapting to the cheese rind environment. Across all of the genes identified in our analysis, the most abundant gene functions are transposases, conjugal transfer genes, phage-related proteins and other mobile elements (673/4844 or 14% of all protein coding sequences). Nearly a third (75/259) of all HGT groups contain mobile elements. These genetic elements are likely involved in either directly or indirectly in the mobilization and transfer of DNA.

In order to determine if gene functions other than mobile elements are enriched in identified HGT regions, we used BlastKOALA [33] to assign KEGG functional annotations (Figure 1C, Figure 2, Supplemental Table 3). Approximately half (52.4%) of genes could not be assigned KEGG annotations. Of the KEGG-annotated genes, the most frequent module (281/2264 or 11%) was “metal ion, iron siderophore and vitamin B12 transport systems”. Five of the ten largest HGT groups as measured by total number of genes (Groups 1, 2, 3, 7 and 8) contain siderophore transport systems (K02013-K02016). Low availability of iron in cheese is known to limit the growth of several bacterial species [34,35]. Previous work has also shown that genes

Figure 1

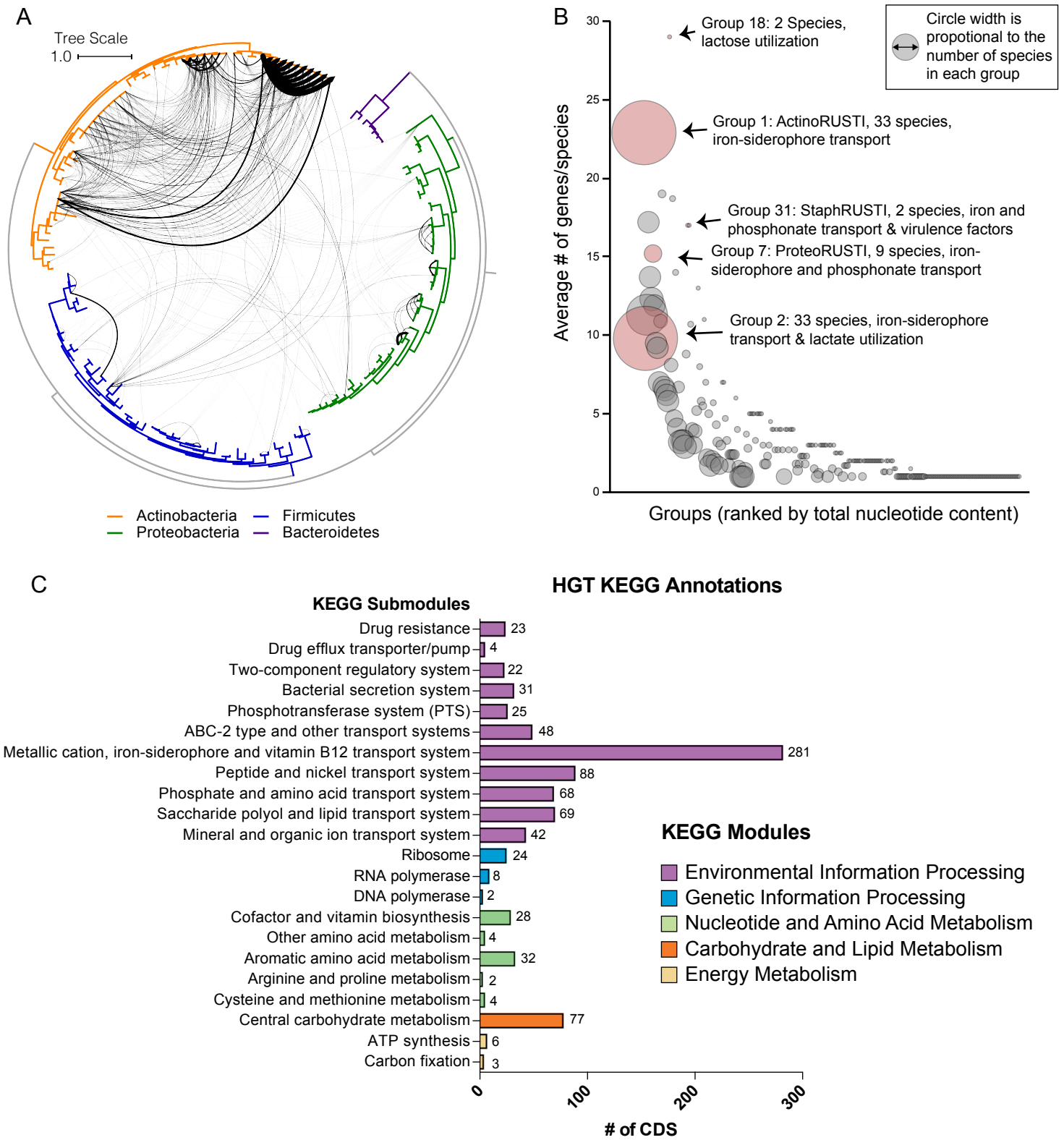


Figure 1: Extensive Horizontal Gene Transfer in the Cheese Microbiome
(A) All HGT events in analyzed cheese-associated bacteria. Connection thickness is scaled to # of shared protein coding sequences. Phylogenetic tree based on 16S RNA alignment using Ribosomal Database Project (RDP). (B) Mean number of genes per species in HGT groups. Diameter of each circle is proportional to the total number of species in the group. Groups highlighted in red are described further in the text. (C) Quantification of KEGG modules and submodules for protein coding genes (CDS) identified as horizontally transferred. Annotations were generated by BLAST-Koala. Genes without function prediction are not depicted.

involved in iron acquisition are present in higher numbers in cheese-associated species compared to closely related species from other environment [35,36].

Many other horizontally transferred genes (267/2264 or 12% of KEGG annotated genes) are also involved in the transport of nutrients relevant for growth in the cheese environment. Lactate is an abundant carbon source in freshly-made cheese, since the initial stages of cheesemaking involve the fermentation of lactose to lactate by lactic acid bacteria [25]. We observe a large number of genes (63/2264 or 2.8% of KEGG annotated genes) involved in lactate import (lactate permease - K03303) or lactate metabolism. Lactate dehydrogenase (K00101), which reduces lactate to pyruvate, represented nearly 1% of all horizontally transferred protein coding sequences.

Apart from lactate, the primary source of energy for microbial growth in cheese would be derived from metabolism of the abundant lipids and proteins, particularly casein [28]. Glutamate importers (43/2264 or 1.9%, eg. K12942, K10005-K10008) and short peptide/nickel transporters (88/2264 or 3.9% eg. K03305) were identified, suggesting pathways for utilization of casein degradation products. Transporters for micronutrients, including phosphonate (K05781, K06163-K06165), molybdate (K02017, K02019, K02020, K03750, K03750, K03639), and metal ions like zinc and manganese were also identified.

HGT of drug resistance genes is of particular concern, since mobile resistance genes from food-associated microbes may be transferred to animal- and human-associated microbes [14]. Cheese rind communities frequently contain filamentous fungi including *Penicillium* species and other microbes that could potentially produce antimicrobial compounds and thus select for antibiotic resistance in co-occurring species. However, less than 1% of KEGG-annotated genes in this dataset are related to drug resistance. A tetracycline resistance gene was identified in 8 *Brevibacterium* species (group 10) and a tripartite multidrug resistance system (K03446, K03543) in 3 *Pseudomonas* species (group 37).

We also noticed a small number of genes that should be part of the core genome and not expected to be horizontally transferred. For example, group 27 is found in all 10 strains of *B. linens* in this dataset, as well as the closely related *B. antiquum* CNRZ918, and contains the SSU ribosomal protein S1p, as well as DNA Polymerase 1. It is possible that these results are false positives since *B. linens* and *B. antiquum* have an ANI ~88%, and these genes are typically more highly conserved than average. At the same time, other ribosomal genes that should also be highly-conserved protein coding genes have substantially lower homology between these species than S1p (Supplementary Table 5). Further, another gene within this HGT group (SAM-dependent methyltransferase) is not typically highly conserved, but nevertheless is >99% identical between these *Brevibacterium* species. We cannot exclude the possibility that this is a false positive, but this may be an example of homologous recombination facilitated by the high sequence identity of the ribosomal protein gene. Several other groups also contain ribosomal proteins (42 - L5p and S3p, 180 - L4p, 219 - S3p), but these groups do not contain any other protein coding genes, and they are clustered with other ribosomal protein coding genes which is a more typical arrangement.

Figure 2

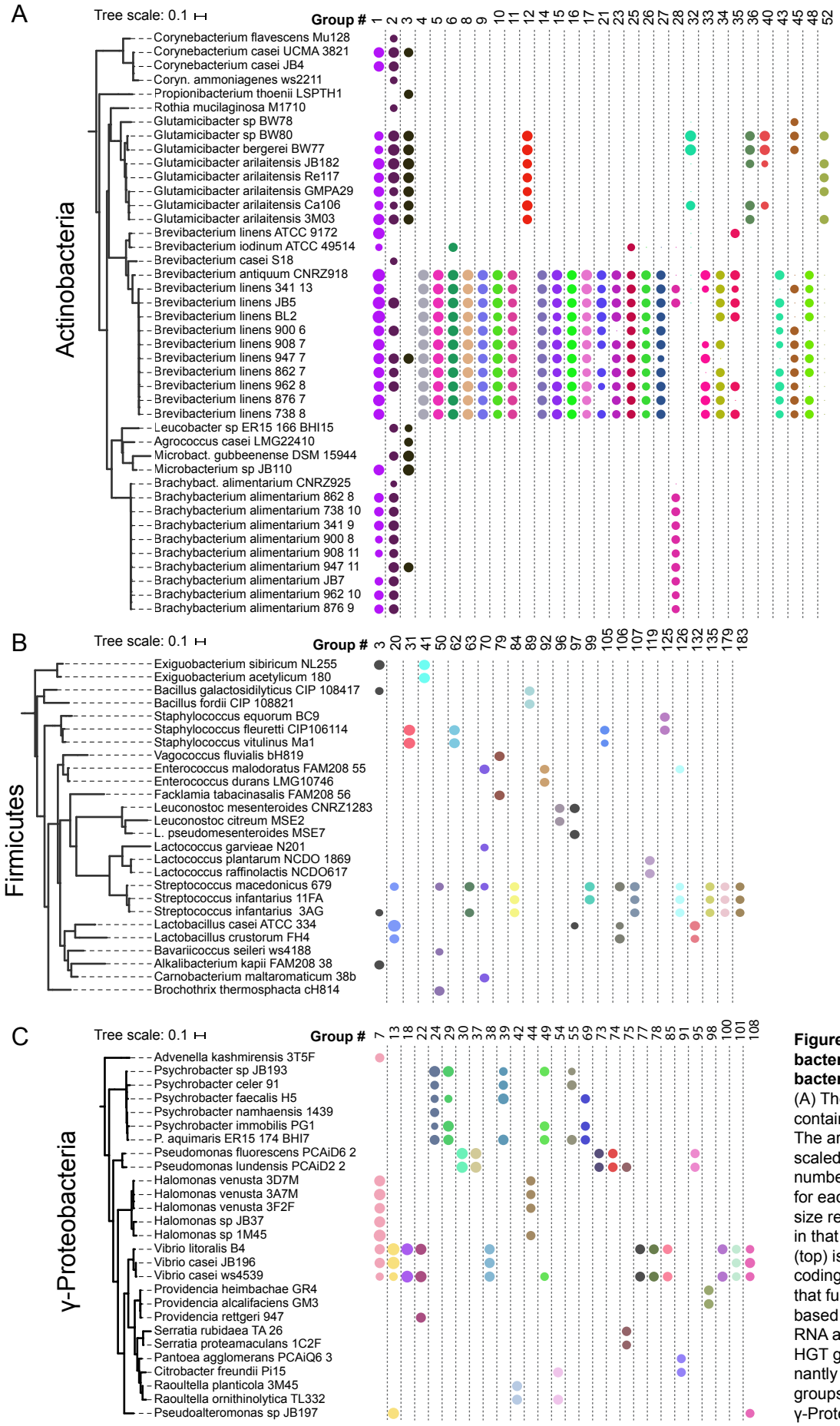


Figure 2: HGT Groups in Actinobacteria, Firmicutes, and γ-Proteobacteria Groups
(A) The 31 largest HGT groups that contain predominantly Actinobacteria. The areas of colored circles are scaled to log2(n), where n is the total number of nucleotides in that group for each species. The largest circle size represents the largest HGT group in that phylum. Protein function matrix (top) is shaded if at least one protein coding sequence in the group has that function. Phylogenies (left) are based on small subunit ribosomal RNA alignment. (B) The 25 largest HGT groups that contain predominantly Firmicutes. (C) The 28 Largest groups that contain predominantly γ-Proteobacteria.

Iron Acquisition HGT

The abundance of iron acquisition genes identified as HGT suggests that iron is a driving force in the adaptation to growth on cheese. The largest HGT region we identified in cheese-associated bacteria, Group 1, includes an island of ~47 kbp (~1% of the genome of *B. linens* JB5) and 34 genes. This island is found in whole or in part in 15 different species in 5 different Actinobacterial genera (*Brachybacterium, Brevibacterium, Corynebacterium, Microbacterium,* and *Glutamicibacter,* formerly *Arthrobacter*), and one Firmicute (*Alkalibacterium*). The core of this region, flanked by AraC-like transcriptional regulators (eg Ga0099663_102740 and Ga0099663_102753 from JB182), contains several genes predicted to form a siderophore import complex, including two cell-surface associated substrate binding protein genes (Ga0099663_102743-44), two membrane permease genes (Ga0099663_102745-46), and an ATPase subunit (Ga0099663_102747). A siderophore reductase (Ga0099663_102741) is present immediately downstream of the AraC regulator, but has less than 99% identityl between the species we analyzed (Figure 3A, Supplementary Table 3). We named this region i**R**on **U**ptake/**S**iderophore **T**ransport **I**sland (RUSTI).

Horizontally transferred genes are not always expressed in the recipient genome, due to possible incompatibilities in promoter sequence [3]. Since iron is a limiting resource in cheese [27,34], we reasoned that if RUSTI is a functional operon, it would likely have increased expression in the presence of additional competition for iron. In order to assess whether RUSTI genes are regulated in the presence of competition, we grew *G. arilaitensis* JB182 alone or in the presence of *Penicillium* and performed RNA sequencing (RNA-seq) to monitor gene expression. In addition, we cultured *B. linens* JB5 alone or in the presence of a competing bacterium, *Staphylococcus succinus* BC15*.* In both cases, the genes in RUSTI were significantly upregulated in the presence of a competing microbe relative to growth alone (Figure 3B), suggesting that this horizontally transferred region is transcriptionally active and may be responding to competition for limited iron in cheese.

Hundreds of different siderophores have been identified belonging to three major classes: hydroxamate, catecholate and α-hydroxycarboxylate [37]. In order to predict the function of the RUSTI transporters, we compared their protein sequences to the Transporter Classification Database (TCDB) [38] using BLAST (Supplementary Table 6). The two genes annotated as permease subunits and one gene annotated as an ATP binding subunit each share substantial homology to the catechol ferric enterobactin transport system (FepD, FepG and FepC respectively) in *E. coli [39–41]*. Two genes annotated as substrate binding proteins have weak homology to vibriobactin and iron(III) dicitrate binding proteins from *Vibrio cholerae* and *E. coli* respectively.

Siderophore-related genes are also well-represented in γ-Proteobacterial HGT groups. Like Group 32 in Actinobacteria, Group 39 contains both siderophore acquisition and siderophore biosynthesis genes and is found in 3 species of *Psychrobacter*. The HGT group with the most protein coding genes that we identified in γ-Proteobacteria (group 7) is found in several *Vibrio* and *Halomonas* species, and like ActinoRUSTI contains an ABC siderophore transport system with an individual substrate-binding, permease and ATP-binding domains (Figure 3C). Though this group appears to have analogous function in the acquisition of iron with RUSTI from

Figure 3

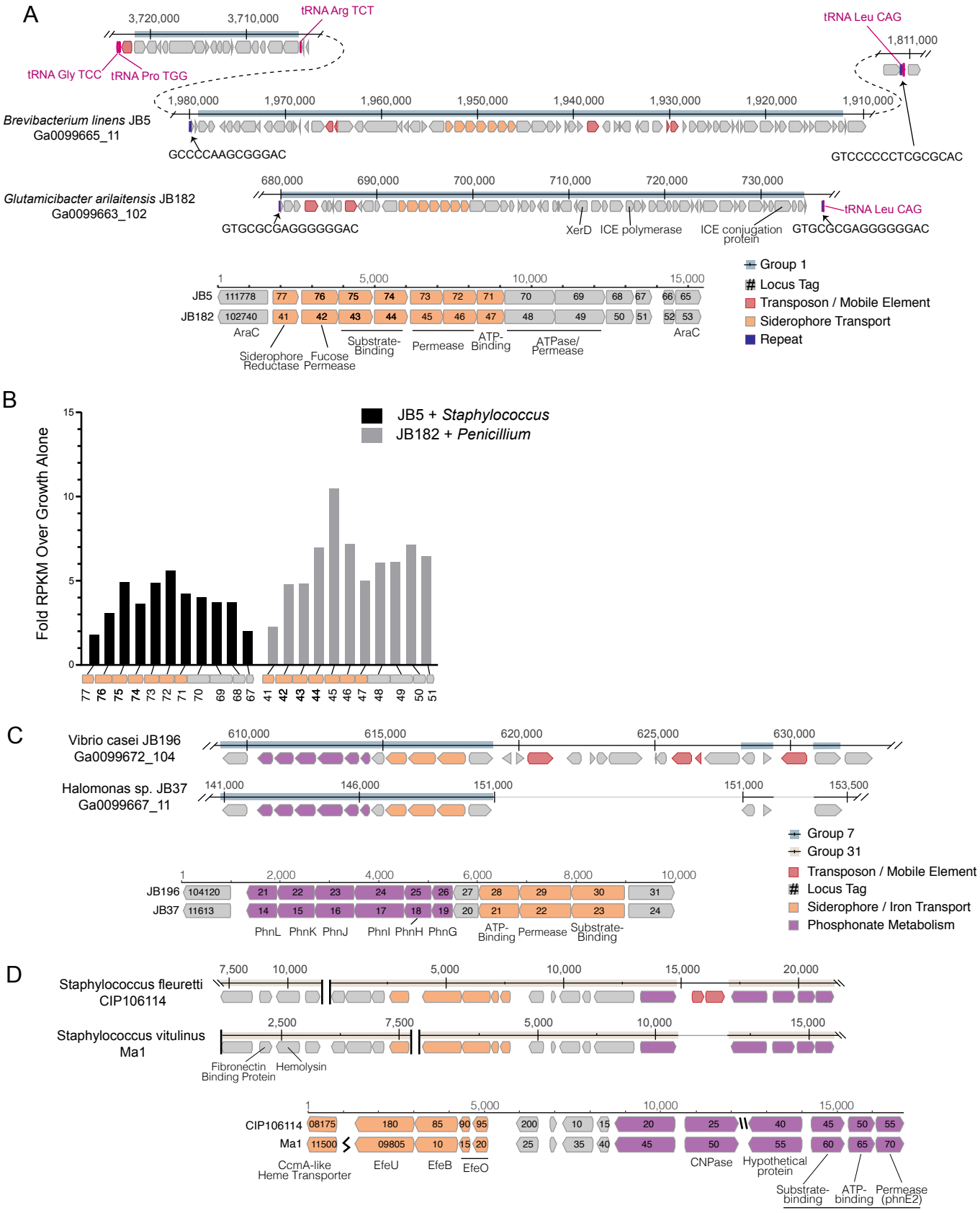


Figure 3: Structure of RUSTI Islands

(A) At-scale schematics for genomic context of HGT Group 1 (top) for B. linens JB5 and G. arilaitensis JB182 and alignment of RUSTI operon (bottom). Nucleotide position values (top) refer to contigs Ga0099665_11 and Ga0099663_102 respectively. Dotted line for JB5 depicts regions of the contig that are not shown. Nucleotide position values (bottom) refer to operon starting from stop codon of leading AraC coding sequence. (B) At scale schematics for genomic context of HGT Group 7 for Halomonas sp. JB37 and V. casei JB196 (top) and alignment of iron and phosphonate metabolism genes (bottom). Nucleotide position values (top) refer to contigs Ga0099667_11 and Ga0099672_104 respectively. Grey lines for JB196 depicts gaps in the alignment due to insertions in JB37. Nucleotide position values (bottom) refer to **operon starting from stop codon of leading protein coding**

Actinobacteria, this ProteoRUSTI does not appear to be related. TCDB analysis suggests homology to hemin transporters in *Yersinia pestis* and *Bordetella pertussis* (Supplementary Table 6).

The same gene island also contains genes related to the Phn family involved in phosphonate import and metabolism [42]. Phosphonate metabolism genes have previously been associated with iron siderophore acquisition in acidic environments [43]), and cheese is typically close to pH5 during the initial periods of rind community growth. Interestingly, BLAST of this region against the NCBI RefSeq database reveals that several uropathogenic *E. coli* strains share identical DNA sequences (Supplementary Table 7). Highly similar sequences are found in *Oligella urethralis*, another gram negative pathogen of the urogenital tract, and *Vibrio harveyi*, a bioluminescent ocean-dwelling microbe. Iron sequestration by animals is a common defense against pathogens [44] and enhanced iron acquisition is commonly associated with virulence. Mammals produce lactoferrin in milk for the same reason [45], and these data suggest that the same genes would be adaptive in both pathogenesis and growth on cheese.

The convergence of strategies for both pathogenesis and growth on cheese is also demonstrated in the Firmicutes (Figure 3D). Two species of cheese-associated *Staphylococcus* (*S. fleuretti* CIP106114 and *S. vitulinus* Ma1) share a large (~20kb) cluster of genes (Group 31) that includes hemolysin and fibronectin binding protein (FnBP), which are involved in virulence in *S. aureus [46,47]*. Hemolysin (also known as alpha toxin) forms pores in cell membranes, and is so-named due to its ability to lyse red blood cells. FnBP enables binding to and invasion of cells, and has been implicated in the formation of biofilms in methicillin resistant *S. aureus [48]*. It is unclear if these genes could provide a selective advantage to cheese-associated Staphylococci, but Group 31 also contains genes for iron acquisition. These genes are homologous to the EfeUOB systems in *E. coli* and *Bacillus subtilis* and FepABC system in *S. aureus,* which are active in low-pH conditions [49–51]. These iron acquisition genes are also found in association with hemolysin and FnBP *S. aureus* and the animal-associated *S. sciuri*, though at only ~80% nucleotide identity (supplementary Table 7) [52].

The genomes sequenced are from a limited number of cheeses from Europe and the United States. In order to determine the distribution of RUSTI across a more expansive sampling of cheese microbiomes, we used BLAST searches against assembled metagenomic data from 38 different cheeses using representative Proteo-, Actino-, and Staph-RUSTI sequences (Figure 4). Gene islands at least 97% identical to ActinoRUSTI were readily identified in 23 (61%) of these metagenomes, in both natural and washed rind cheeses from the United States and Europe. Though less common (26% of metagenomes), ProteoRUSTI was also identified in diverse cheeses in both the US and Europe. StaphRUSTI could not be found in any of the metagenomes we analyzed. These data demonstrate that siderophore-associated HGT islands are widespread in cheese rind microbiomes. Whether independent HGT events are happening within each cheese production and aging facility, or if they happened before the widespread distribution of these microbes across cheese production regions is unknown.

Figure 4

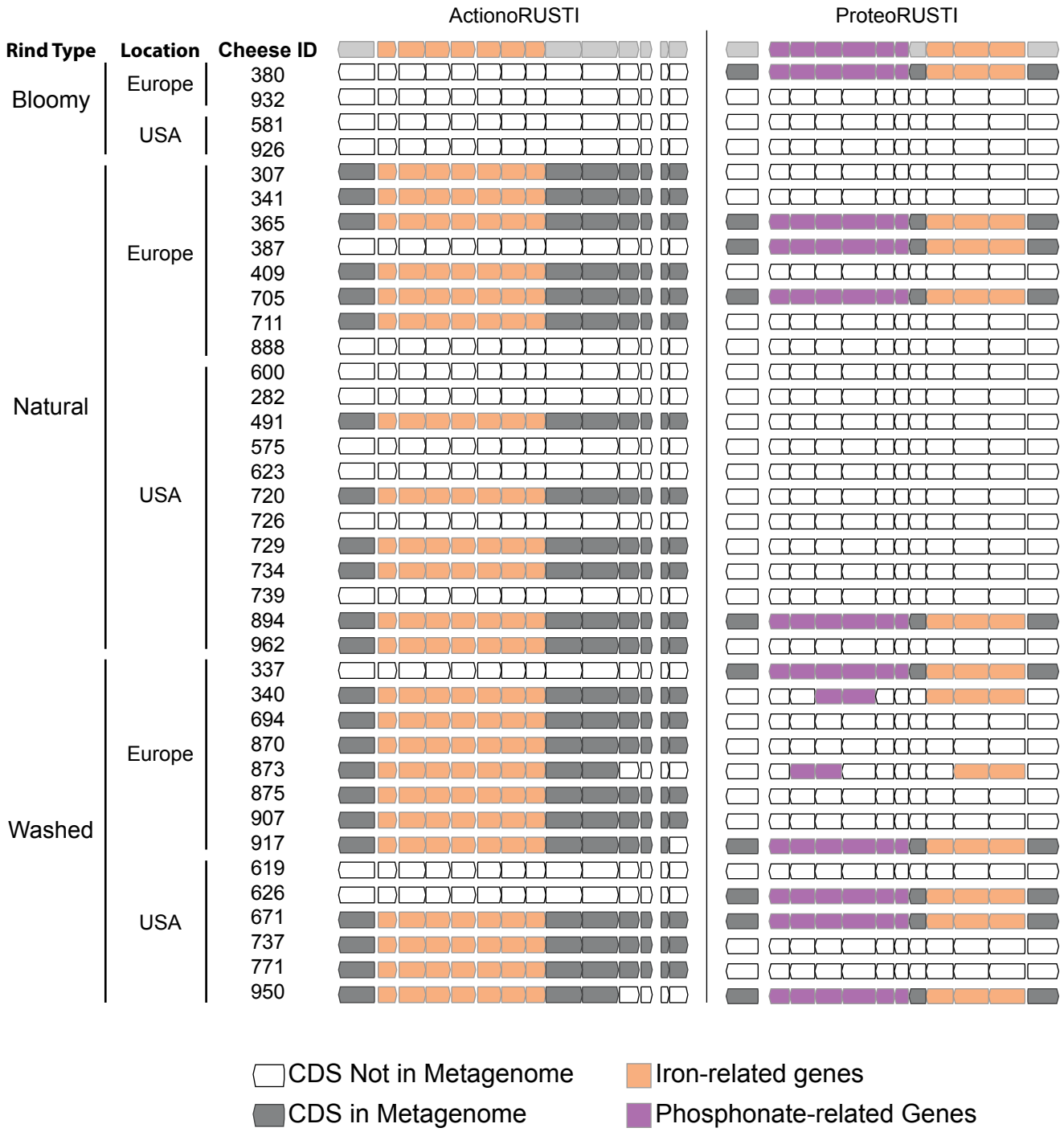


Figure 4: Presence of RUSTI in Cheese Metagenomes
Genes in ActinoRUSTI (G. arilaitensis JB182) and ProteoRUSTI (V. casei JB196) regions were compared to 32 assembled metagenomes from the US and Europe. Filled CDS represents positive (>97% identical nucleotides) hit in that metagenome.

A potential mechanism of transfer and source of the Actinobacterial RUSTI

To begin to understand potential mechanisms which could mediate HGT in cheese-associated bacteria, we analyzed the sequences surrounding the RUSTI region of *Glutamicibacter* JB182. Conjugative elements are a common way for HGT to occur [53]. Integrative and conjugative elements (ICEs) can in part be identified by the presence of signature proteins associated with core functions of integration into and excision from the host genome (recombinase), replication as an extrachromosomal element (polymerase), and conjugation from the host to recipient cell (conjugation) [54]. Analysis of the *Glutamicibacter* JB182 RUSTI region reveal homologs of each of these protein classes (Figure 3A). A recombinase of the site-specific tyrosine recombinase XerD family (Ga0099663_102762) [55], a hexameric ATPase conjugation protein of the VirD4/TraG/TraD family (Ga0099663_102784) [56], and a homolog of the bi-functional primase-polymerase DNA replication protein family (Ga0099663_102766). Interestingly, Actinobacterial ICE systems typically utilize a conjugation apparatus belonging to the SpoIIE/FtsK family, which allows transfer of double-stranded DNA [57,58]. However, the conjugation machinery here is more reminiscent of Gram-negative and Firmicute systems of single-stranded transfer [59].

ICE integration is site-specific, and frequently occurs at the 3' end of tRNA genes [54]. Immediately downstream of the RUSTI region in *Glutamicibacter* is a Leucine tRNA. The 3' end of the tRNA forms an imperfect repeat with the region immediately upstream of the RUSTI region, which strongly suggests that the tRNA-Leu is used at the integration site (*att* site) for this ICE. In order to determine whether this ICE is still active, we performed PCR using primers within and flanking the putative integration site (Figure 5A). We were able to detect PCR products which suggest that at least a portion of the cells within the population have lost the RUSTI ICE from their chromosome, and it is present as an extrachromosomal circular form (Figure 5B). Sequencing of the PCR product (primers 1+6) that spans the predicted excision site matched the predicted remaining sequence, containing Repeat element B (Supplementary Figure 4A). Sequencing of the PCR product (primers 2+5) that spans the predicted circularization site matched the predicted sequence, containing Repeat element A (Supplementary Figure 4B).

There are several possible explanations for the widespread distribution of of nearly identical ActinoRUSTI. Initial transfer events may have occurred in a single location, on the surface of a cheese or in livestock, and subsequently been dispersed to many separate cheesemaking facilities. The continued mobility of the ICE in JB182 raises the alternate possibility that it may be continually introduced to many cheeses from a common source. Cheese producers often use commercially-available "starter" cultures that contain desirable species, including many Actinobacteria [60]. We tested 5 common starter cultures for the presence of ActinoRUSTI by PCR, and positively identified it in 1 of them (Figure 5C). This culture is known to contain 2 species of Actinobacteria, *Brevibacterium linens* and *Glutamicibacter nicotinae*. In order to identify which species may be the RUSTI donor, we plated the starter culture and isolated 4 distinct strains based on colony morphology. SSU sequencing revealed that two isolates were *Glutamicibacter* and two were *Brevibacterium*. One of the *Brevibacterium* isolates tested

Figure 5

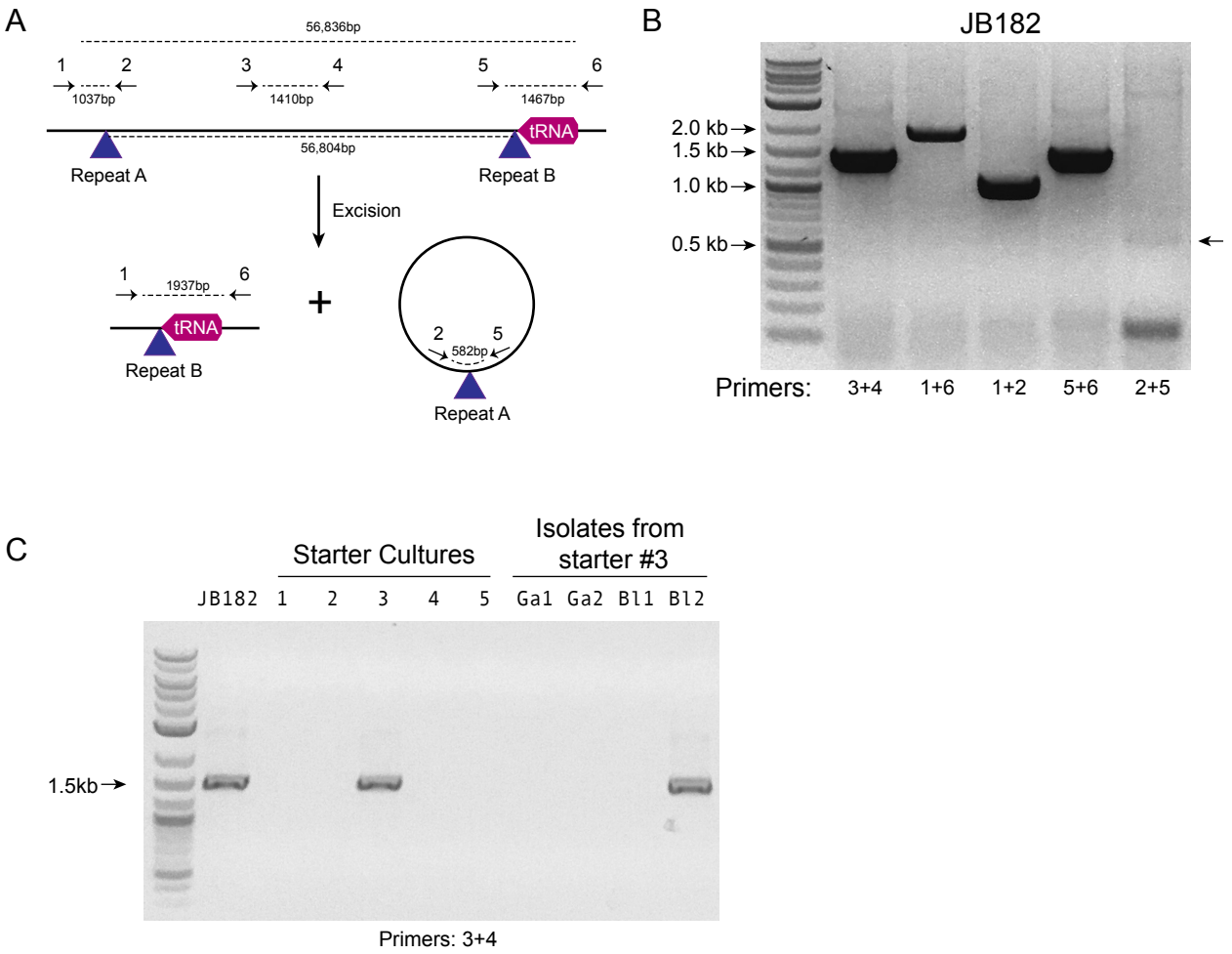


Figure 5: Mobility of RUSTI
(A) Schematic for PCR primer design - see materials and methods for details. (B) PCR testing for the presence of RUSTI and for the excision of the ICE in an overnight culture of G. arilaitensis JB182. (C) DNA was extracted from 5 commercially available starter cultures and tested for the presence of RUSTI using PCR with primers specific for the HGT region (Materials and Methods). Starter culture 3 was plated on PCAMS media, and 4 isolates selected based on colony morphology were also tested. The expected size for the amplicon is ~1.4kb. Sequencing of the 16S ribosomal RNA genes for these isolates suggested that two isolates are Glutamicibacter arilaitensis and two are Brevibacterium linens. G. arilaitensis. JB182 was used as a positive control.

positive for ActinoRUSTI by PCR (Figure 5C). These data are consistent with the hypothesis that HGT from a starter culture could explain some of the dissemination of ActinoRUSTI.

Discussion

In this paper, we provide evidence of extensive horizontal gene transfer in cheese-associated bacteria. Many of the transferred regions are large multi-gene islands, and are shared by numerous species. Genes involved in nutrient acquisition, especially iron and lactate, are particularly abundant, suggesting that HGT may provide a selective advantage in the iron and sugar limited environment of cheese. The largest HGT we identified appears to be an active ICE and is found in a starter culture, raising the possibility that we are observing contemporary processes that may have ongoing importance. These data support previous studies that show HGT is an important factor in the evolution of microbial communities [4] and suggest that cheese rind communities may be a useful model for studying this process in greater detail.

For this study, we focused on bacterial members of the cheese microbiome, but many cheeses also contain numerous fungal species. Indeed, HGT has been previously documented in cheese-associated *Penicillium* species [22]. HGT between bacteria and fungi has also been documented in other environments, though is thought to be rare [61]. Evaluating bacterial-fungal gene transfer in cheese could provide additional insights into the extent and importance of gene exchange in microbial communities. Further sequencing of bacterial genomes from cheese could also continue to reveal HGT.

We were able to show that ActinoRUSTI in *G. arilaitensis* JB182, is likely contained within an integrative and conjugative element, although many other mechanisms for transfer are likely at play. Indeed, the appearance of phage-related genes, transposases and other mobile elements in many HGT groups suggests that we are observing the results of multiple methods for mobilizing DNA.

Our method for identifying HGT does not permit us to determine the direction of gene flow, and indeed it seems likely that the original sources of many of these genetic elements are not present in our dataset. In some cases, we can infer a possible origin, such as the *Brevibacterium* strain in a starter culture that may be the source of RUSTI in multiple cheese species around the world. However, this does not enable us to identify where this species acquired them. Further characterization of cheese-associated microbes, as well as those found in dairy farms or in cheese caves may provide a more complete picture, but the evidence that at least some of these genetic elements are found in human pathogens and ocean dwelling bacteria suggests that genes are shared across diverse environments.

Though previous studies demonstrated that iron is limiting for *Glutamicibacter, Brevibacterium,* and *Corynebacterium* species growing on cheese [34,35], the preponderance of siderophore and other iron acquisition genes we observe being horizontally transferred suggests that the same is true across bacterial phyla. Limiting iron is a deliberate strategy on the part of mammalian hosts to block the growth of infectious microbes, and this strategy influences the composition of milk due to the presence of lactoferrin [45]. Interestingly, convergent strategies

for acquiring iron are utilized by pathogens and by cheese-associated microbes and we observe that in some cases these disparate species appear to have shared genes through horizontal transfer. The presence of these same genes in a microbe found in ocean habitats suggests that these genes have broad utility for the common challenge of iron limitation.

We have yet to demonstrate the functional consequences of these genes on individual species or on the community as a whole. Given that iron is limiting in cheese, and that ActinoRUSTI genes are upregulated in response to other species (Figure 4C), it is likely that these genes are functional and may play a role in competition. The prevalence of siderophore import, but not siderophore synthesis pathways may suggest that species may cheat by scavenging the biosynthetic products of others [62].

The identification of widespread sharing of genes in cheese microbial communities could have important implications. In particular, the possibility that a starter culture is the source of mobile gene elements suggests that the genomic content, rather than just specific species must be considered when designing microbial supplements. In addition to starter cultures used for fermented foods, living microbial supplements ("probiotics") are increasingly being adopted in agriculture [63,64] and for a wide range of human health conditions [65–67] and even as cosmetics [68]. The need to screen for clinically relevant elements such as antibiotic resistance genes is widely recognized [69], but other mobile gene elements from these organisms may also enter native microbial populations with unknown consequences.

Though we and others observe a large number of HGT events in microbial species [32], the biotic and abiotic conditions that affect this frequency, and what effects HGT may have on the community remain unclear. A model system to study HGT in a community context is particularly important, since sequence-based characterization of complex communities has particular limitations when it comes to HGT. Further, even if complete characterization *in situ* were possible, many microbial communities are difficult to experimentally manipulate *in vitro*.

By contrast, cheese rind-associated bacteria are readily isolated and cultured, and model communities may enable identification of features of microbial communities and their environments that alter frequency and extent of HGT. The cheese rind model system may provide an opportunity to observe HGT as it happens and to investigate how community composition affects the frequency of transfer and the persistence of genes. The *in vitro* cheese system enables experimental manipulation to investigate the role of community composition in driving HGT. Further, since many gene products may only have survival benefits in the context of community competition and cooperation, investigating the role of RUSTI and other horizontally transferred genes on microbial growth in the context of their natural community is critical.

Many species of cheese-associated bacteria have close relatives in the soil, on skin, in the ocean, making insight from this system potentially applicable to diverse environments. In addition, horizontal acquisition of iron-uptake genes has been documented in numerous environments including the oceans and in human pathogens [70,71], suggesting that the specific processes occurring in cheese may also be generally informative across systems. Understanding the extent of HGT in the cheese microbiome is the first step towards addressing

how the movement of genes shapes and is shaped by a microbial community. Using cheese rinds as a model system can help elucidate the factors that influence the frequency of HGT, how it impacts competition and cooperation, and helps shape a microbiome.

Methods

Sequencing and Genome Assembly

Bacterial strains JB4, 5, 7, 37, 110, 182, 196, and 197 were isolated from cheeses in a single geographic region and sequenced using a combination of Illumina short-read (100bp, paired end) and Pacbio long-read sequencing. DNA was extracted using Genomic Tip 100/G (Qiagen, USA) or Power Soil (MoBio, USA). Illumina library preparation and sequencing were performed at Harvard University by the Bauer Core facility. Pacbio library preparation and sequencing were performed by the University of Massachusetts Medical School Deep Sequencing Core. De novo hybrid assembly was performed using SPAdes (v3.5.0) [72]. Genomes were annotated using the Integrated Microbial Genomes Expert Review (IMG/ER) annotation pipeline [73]. In addition, we also sequenced 8 additional rind isolates of *Brachybacterium* (strains 341.9, 738.10, 862.8, 876.9, 900.8, 908.11, 947.1, 962.10) and *Brevibacterium* (strains 341.13, 738.8, 862.7, 876.7, 900.6, 908.7, 947.7, 962.8) and 3 additional isolates of *Glutamicibacter* (strains BW77, 78, 80) from different cheeses in a broad geographic distribution. For these isolates, we prepared draft genomes using Illumina short-read sequencing and assembled with CLC genomics workbench.

Phylogenetic trees

16S sequences were retrieved from the sequenced genomes and aligned using the structure-based aligner, Infernal v1.1rc4[74], as implemented in the Ribosomal Database Project release 11 [75]. The alignment was imported into Geneious v9 (Biomatters, LTD), and a tree was calculated using the maximum likelihood method PHYML (GTR model)[76]. The tree was rooted using *Thermus thermophilus*. The tree was then uploaded to Interactive Tree of Life (iTOLv3)[77] to enable mapping of HGT data (connections and group abundance profiles).

RNAseq

Four replicate transcriptomes from two treatments were sequenced: 1) *G. arilaitensis* alone and 2) *G. arilaitensis* + *Penicillium*. We used a strain of *Penicillium solitum* that was isolated from a natural rind cheese and was used for experiments in Wolfe et al. 2014. For each experimental unit, approximately 80,000 CFUs of *Glutimicibacter arilaitensis* were spread across the surface of a 100mm Petri dish containing 20 mL of cheese curd agar (10% freeze-dried fresh cheese, 3% NaCl, 1.7% agar, 0.5% xanthan gum) [26]. For the + *Penicillium* treatment, approximately 100,000 CFUs were co-inoculated onto the plates with the Arthrobacter. Plates were incubated in a bin with moist paper towel (> 90% relative humidity) at 24 °C for 5 days.

Rind biofilms were then harvested by scraping the cheese curd surface (*G. arilaitensis*) or from liquid cheese culture (2% freeze-dried fresh cheese, 3% NaCl) (*B. linens*) and stored in

RNAProtect Reagent (Qiagen) to stabilize mRNA frozen at -80°C. RNA was extracted using a standard phenol-chloroform protocol used for many different fungal and bacterial species and has been adopted from transcriptomics work in gut microbiomes, (see[78]). This protocol uses a standard bead-beating step in a lysis buffer to release cell contents from biofilms stored in RNAProtect. DNA was removed from the samples using a TURBO DNA-free kit (Life Technologies), and 5S, tRNA and large rRNA was depleted using MEGAClear (Life Technologies) and RiboZero (Illumina) kits, respectively. To remove both fungal and bacterial large rRNA, we used an equal mixture of Ribo-Zero Yeast and Bacteria rRNA Removal Solution. To confirm that the samples are free of DNA contaminants, a PCR of the 16S rRNA gene was with standard primers (27f and 1492r). Overall quantity and quality of the RNA preps were confirmed by Nanodrop and Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit.

RNA-seq libraries were constructed from purified mRNA using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs) where each library received a unique 6 base pair barcode for demultiplexing after the sequencing run. Each library was run on an Agilent 2100 Bioanalyzer High Sensitivity DNA chip to confirm that primer dimers and adapter dimers are not present in the sample and to determine the size of the library. Final libraries were standardized to 10 nM each after quantification with a Qubit dsDNA HS Assay Kit (Life Technologies) and the pooled in equal amounts to get similar sequencing coverage across all libraries. The pooled library samples were be sequenced using paired-end 100bp reads on an Illumina HiSeq Rapid Run by the Harvard Bauer Core Sequencing Core Facility.

To quantify gene expression and determine if genes within RUSTI were differentially expressed when grown with the competitor *Penicillium*, we used Rockhopper [79]. Only forward reads were used for this analysis. The assembled and annotated *Glutamicibacter arilaitensis* strain JB182 or *Brevibacterium linens* JB5 (described above) genomes were used as reference genomes for mapping. We considered genes that had a greater than 4-fold difference in expression when grown with *Penicillium* or *Staphylococcus* and were significantly different (based on Rockhopper's *q*-values, which control for false discovery rate using the Benjamini-Hochberg procedure) to be differentially expressed genes.

PCR

PCR reactions were performed using Q5 Hot Start Mastermix (New England Biolabs). Where JB182 RUSTI is integrated in the chromosome, PCR using primer 1 (CAACTGTGCCACGCAATTCA) and primer 2 (CGGCTACTTCTCGGATGGTC) are expected to produce a 1037bp product that includes the 5' ICE repeat. Primer 3 (CGCAATCGTGGTGTATCTGC) and primer 4 (GACGGGATCAGGAACGACG) should produce a 1410bp product, while primer 5 (GCCGCATCTACCTCGATGAA), and primer 6 (CCAAATCGCGACGCATTGAT) are expected to form a 1467bp product. Primers 1 and 6 are separated by approximately 59kb when RUSTI is present and are not expected to form a PCR product, but should form a 1937bp product if RUSTI is excised. Primers 2 and 5 should not form a PCR product when RUSTI is integrated, but would form a 500bp product if the excision circle is present.

Additional Software

Annotated genomes were compared using blastn from BLAST+ (v2.3.0) [80]. Protein coding genes were considered potential HGT if their sequence was at least 99% identical for at least 500 nucleotides. Neighboring candidate HGT were identified as part of the same island if they were separated by no more than 5000 nucleotides. Scripts to import and store genome information and blast results and to analyze results are available on github [81].

Genomic average nucleotide identity (ANI) was calculated using the “ani.rb” script from the enveomics collection (commit “e8faed01ff848222afb5820595cccc4e50c89992”) with default settings [31].

Metagenomes

Shotgun metagenomic data from [26] and [27] were assembled with CLC Genomic Workbench 8.0. Representative sequences for ActinoRUSTI or ProteoRUSTI were compared to assembled metagenomes by BLAST. Hits with >97% similarity were considered positive hits for target regions.

Acknowledgements

The authors would like to thank Rajashree Mishra for assisting with initial stages of project development, Dana Boyd for assistance with ICE identification, Brian Tsu for assistance with the Transporter Classification Database, and members of the Dutton lab for helpful conversations.

Conflicts of Interest

The authors declare that they have no competing interests.

Funding

KSB, BEW, and RJD received support from National Institutes of Health grant P50 GM068763 to Harvard University.

Figures

Figure 1: Extensive Horizontal Gene Transfer in the Cheese Microbiome

(A) All HGT events in analyzed cheese-associated bacteria. Connection thickness is scaled to # of shared protein coding sequences. Phylogenetic tree based on 16S RNA alignment using Ribosomal Database Project (RDP). (B) Mean number of genes per species in HGT groups. Diameter of each circle is proportional to the total number of species in the group. Groups highlighted in red are described further in the text. (C) Quantification of KEGG modules and

submodules for protein coding genes (CDS) identified as horizontally transferred. Annotations were generated by BLAST-Koala. Genes without function prediction are not depicted.

Figure 2: HGT Groups in Actinobacteria, Firmicutes, and γ-Proteobacteria Groups

(A) The 31 largest HGT groups that contain predominantly Actinobacteria. The areas of colored circles are scaled to log2(n), where n is the total number of nucleotides in that group for each species. The largest circle size represents the largest HGT group in that phylum. Protein function matrix (top) is shaded if at least one protein coding sequence in the group has that function. Phylogenies (left) are based on small subunit ribosomal RNA alignment. (B) The 25 largest HGT groups that contain predominantly Firmicutes. (C) The 28 Largest groups that contain predominantly γ-Proteobacteria.

Figure 3: Structure of RUSTI Islands

(A) At-scale schematics for genomic context of HGT Group 1 (top) for *B. linens* JB5 and *G. arilaitensis* JB182 and alignment of RUSTI operon (bottom). Nucleotide position values (top) refer to contigs Ga0099665_11 and Ga0099663_102 respectively. Dotted line for JB5 depicts regions of the contig that are not shown. Nucleotide position values (bottom) refer to operon starting from stop codon of leading AraC coding sequence. (B) At scale schematics for genomic context of HGT Group 7 for *Halomonas sp.* JB37 and *V. casei* JB196 (top) and alignment of iron and phosphonate metabolism genes (bottom). Nucleotide position values (top) refer to contigs Ga0099667_11 and Ga0099672_104 respectively. Grey lines for JB196 depicts gaps in the alignment due to insertions in JB37. Nucleotide position values (bottom) refer to operon starting from stop codon of leading protein coding sequence. (C) At-scale schematics for genomic context of HGT Group 31 for *S. fleuretti.* CIP106114 and *S. vitulinus* Ma1 (top). For both species, the group is split across 2 different contigs and nucleotide position values (top) refer to the relative position for that contig. Alignment of iron and phosphonate metabolism genes from Group 31 (bottom).

Figure 4: Presence of RUSTI in Cheese Metagenomes

Genes in ActinoRUSTI (*G. arilaitensis* JB182) and ProteoRUSTI (*V. casei* JB196) regions were compared to 32 assembled metagenomes from the US and Europe. Filled CDS represents positive (>97% identical nucleotides) hit in that metagenome.

Figure 5: Mobility of RUSTI

(A) Schematic for PCR primer design - see materials and methods for details. (B) PCR testing for the presence of RUSTI and for the excision of the ICE in an overnight culture of *G. arilaitensis* JB182. (C) DNA was extracted from 5 commercially available starter cultures and tested for the presence of RUSTI using PCR with primers specific for the HGT region (Materials and Methods). Starter culture 3 was plated on PCAMS media, and 4 isolates selected based on colony morphology were also tested. The expected size for the amplicon is ~1.4kb. Sequencing

of the 16S ribosomal RNA genes for these isolates suggested that two isolates are *Glutamicibacter arilaitensis* and two are *Brevibacterium linens*. *G. arilaitensis.* JB182 was used as a positive control.

Supplementary Table 1: Genome Information

Genome statistics for newly sequenced genomes, determined by IMG/ER. Gene IDs refer to IMG bioproject or RefSeq Accession. Genomes from Almeida et. al. do not yet have Accession numbers.

Supplementary Table 2: Pairwise Species Comparison Summary

Total protein coding sequences (CDS) and nucleotides determined to be horizontally transferred For every pair of species that were compared by the HGT detection pipeline. Also shows calculated ANI and 16S distance (see materials and methods). Species pairs that have ANI > 0.89 were not compared and are not shown.

Supplementary Table 3: Full Group Annotations

All protein coding sequences identified as HGT, sorted by group # (ranked by total nucleotide content), species and genome location within species. Certain functional annotations are identified by color (eg orange for iron) based on text in annotation. Locus tags and contig IDs beginning with lower-case letters were assigned by kvasir, and do not correspond to any published database.

Supplementary Table 4: Group Summary Statistics

Summary statistics for each HGT group.

Supplementary Table 5: Highly Conserved Genes in *Brevibacterium* species

Protein coding sequences from Group 27, as well as selected highly conserved genes from *Brevibacterium antiquum* CNRZ918 were compared to other *Brevibacterium* strains by BLAST. *B. linens* 947.7 has substantially lower nucleotide identity for the 4 genes found in Group 27 than other *B. linens* strains, despite similar nt distance for other highly conserved genes. This suggests Group 27 is a true example of HGT between CNRZ918 and other *B. linens* strains, rather than a false-positive.

Supplementary Table 6: TCBD Hits for Transporters in RUSTI

Representative CDS of Actino- and ProteoRUSTI from *G. arilaitensis* JB182 and *V. casei* JB196 respectively were compared to the Transporter Classification Database (TCDB). Results colored according to type of siderophore transported according to annotation.

Supplementary Table 7: RefSeq BLAST

Actino- and ProteoRUSTI from *G. arilaitensis* JB182 and *V. casei* JB196 respectively, as well as the consensus sequence for StaphRUSTI (see figures 3 and 4) were compared to the NCBI RefSeq database using BLAST.

Supplementary Figure 1: Kvasir Schematic

Schematic of software pipeline to identify HGT. (1) Sequenced genomes are annotated with IMG/ER and downloaded in Genbank format. (2) All annotated genes in all genomes are used to assemble a BLAST database using BLAST+ command-line tools. (3) All protein-coding genes (CDS) from all species are queried against the BLAST database. Hits from the same species are discarded; hits from species with an ANI >89% were discarded; other hits are saved. (4) For each species, coding sequences that have at least one BLAST hit are grouped into islands based on proximity. Genes that are within 5kb of each other on the same contig are considered part of the same island. (5) Islands in each species are compared to islands in each other species to form groups. Islands that share at least 1 gene in common according to BLAST parameters in step 3 are placed in the same group.

Supplementary Figure 2: HGT Identification Parameters

Different parameters for minimum length of gene match for HGT, maximum % ANI identity for related species, and maximum distance between genes in an island were compared. (A) Number of positive HGT hits identified when varying the minimum protein coding gene length. (B) Number of HGT groups constructed when varying the maximum separation between hits that are classified as belonging to the same group. (C) Number of nucleotides or (D) number of protein coding sequences in HGT regions by 16S similarity. Note - since 500bp is the minimum length for protein coding sequences in this analysis.

Supplementary Figure 3: Group Clustering

(A) Expected clustering: contiguous genes in multiple species are in a single group. Though island 6 (i6) lacks one gene present in i1 and i4, (possibly because of a transposon insertion), it is still correctly considered related. (B) Ambiguous grouping: Islands 2 and 3 from species 1 are found on different contigs, but are grouped together. They may be found in close proximity in the genome, but on different sides of a gap in the assembly, or they may be quite distant from each other. The grouping of related genes in species 2 into a single island suggests that they may have been transferred in a single event, but the possibility of two unrelated HGT events landing in the same spot cannot be excluded. (C) Possible mis-grouping of two HGT events in a single group: Though species 4 does not share any genes with species 1 and 2, these islands are nevertheless clustered due to the proximity of coding sequences in species 3. This may correctly represent a single gene cluster that subsequently diverged in each species, or unrelated HGT that happened to insert in close proximity. (D) Mis-grouping due to mobile element: Mobile elements (outlined in red) found in multiple locations in multiple genomes may insert next to unrelated HGT islands, causing spurious grouping by the algorithm.

Supplementary Figure 4: ICE Excision Sequencing

The ~2000bp band from the PCR amplification using primers 1 and 6 (A) and the ~500bp band from amplification using primers 2 and 5 (B) were extracted and sequenced. Alignment with the JB182 genome reveals 100% alignment with expected and the spliced chromosomal region containing the 3' repeat and the excision circle containing the 3' repeat respectively.

1. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J. 2016; doi:10.1038/ismej.2016.45

2. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev. 2013;77: 342–356.

3. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. nature.com; 2000;405: 299–304.

4. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev. 2011;35: 957–976.

5. Tasse L, Bercovici J, Pizzut-Serin S, Robe P, Tap J, Klopp C, et al. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 2010;20: 1605–1612.

6. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010;464: 908–912.

7. Niehus R, Mitri S, Fletcher AG, Foster KR. Migration and horizontal gene transfer divide microbial genomes into multiple niches. Nat Commun. 2015;6: 8924.

8. McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. Science. 2010;330: 50.

9. Andam CP, Carver SM, Berthrong ST. Horizontal Gene Flow in Managed Ecosystems. Annu Rev Ecol Evol Syst. 2015;46: 121–143.

10. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011;480: 241–244.

11. McCarthy AJ, Loeffler A, Witney AA, Gould KA, Lloyd DH, Lindsay JA. Extensive horizontal gene transfer during Staphylococcus aureus co-colonization in vivo. Genome Biol Evol. 2014;6: 2697–2708.

12. Hiramatsu K, Cui L, Kuroda M, Ito T. The emergence and evolution of methicillin-resistant Staphylococcus aureus. Trends Microbiol. 2001;9: 486–493.

13. Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. Science. science.sciencemag.org; 2012;337: 1107–1111.

14. Rossi F, Rizzotti L, Felis GE, Torriani S. Horizontal gene transfer among microorganisms in food: current knowledge and future perspectives. Food Microbiol. 2014;42: 232–243.

15. Wang HH, Manuzon M, Lehman M, Wan K, Luo H, Wittum TE, et al. Food commensal microbes as a potentially important avenue in transmitting antibiotic resistance genes. FEMS Microbiol Lett. femsle.oxfordjournals.org; 2006;254: 226–231.

16. Mathur S, Singh R. Antibiotic resistance in food lactic acid bacteria—a review. Int J Food Microbiol. Elsevier; 2005;105: 281–295.

17. Cocconcelli PS, Cattivelli D, Gazzola S. Gene transfer of vancomycin and tetracycline resistances among Enterococcus faecalis during cheese and sausage fermentations. Int J Food Microbiol. Elsevier; 2003;88: 315–323.

18. Delorme C. Safety assessment of dairy microorganisms: Streptococcus thermophilus. Int J Food Microbiol. Elsevier; 2008;126: 274–277.

19. Flórez AB, Delgado S, Mayo B. Antimicrobial susceptibility of lactic acid bacteria isolated from a cheese environment. Can J Microbiol. NRC Research Press; 2005;51: 51–58.

20. Li X, Xing J, Li B, Yu F, Lan X, Liu J. Phylogenetic analysis reveals the coexistence of interfamily and interspecies horizontal gene transfer in Streptococcus thermophilus strains isolated from the same yoghurt. Mol Phylogenet Evol. 2013;69: 286–292.

21. Liu M, Siezen RJ, Nauta A. In silico prediction of horizontal gene transfer events in Lactobacillus bulgaricus and Streptococcus thermophilus reveals protocooperation in yogurt manufacturing. Appl Environ Microbiol. Am Soc Microbiol; 2009;75: 4120–4129.

22. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun. 2014;5: 2876.

23. Ropars J, Rodríguez de la Vega RC, López-Villavicencio M, Gouzy J, Sallet E, Dumas É, et al. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. Curr Biol. 2015;25: 2562–2569.

24. Gibbons JG, Rinker DC. The genomics of microbial domestication in the fermented food environment. Curr Opin Genet Dev. 2015;35: 1–8.

25. Button JE, Dutton RJ. Cheese microbes. Curr Biol. 2012;22: R587–9.

26. Wolfe BE, Button JE, Santarelli M, Dutton RJ. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell. 2014;158: 422–433.

27. Kastman EK, Kamelamela N, Norville JW, Cosetta CM, Dutton RJ, Wolfe BE. Biotic interactions shape the ecological distributions of Staphylococcus species. MBio. 2016; doi:10.1128/mBio.01157-16

28. Monnet C, Landaud S, Bonnarme P, Swennen D. Growth and adaptation of microorganisms on the cheese surface. FEMS Microbiol Lett. 2015;362: 1–9.

29. Almeida M, Hébert A, Abraham A-L, Rasmussen S, Monnet C, Pons N, et al. Construction of a dairy microbial genome catalog opens new perspectives for the metagenomic analysis

of dairy fermented products. BMC Genomics. 2014;15: 1101.

30. Chan JZ-M, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ. Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol. 2012;12: 302.

31. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes [Internet]. PeerJ Preprints; 2016 Mar. Report No.: e1900v1. doi:10.7287/peerj.preprints.1900v1

32. Ravenhall M, Škunca N, Lassalle F, Dessimoz C. Inferring horizontal gene transfer. PLoS Comput Biol. journals.plos.org; 2015;11: e1004095.

33. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428: 726–731.

34. Monnet C, Back A, Irlinger F. Growth of aerobic ripening bacteria at the cheese surface is limited by the availability of iron. Appl Environ Microbiol. 2012;78: 3185–3192.

35. Monnet C, Loux V, Gibrat J-F, Spinnler E, Barbe V, Vacherie B, et al. The arthrobacter arilaitensis Re117 genome sequence reveals its genetic adaptation to the surface of cheese. PLoS One. 2010;5: e15489.

36. Walter F, Albersmeier A, Kalinowski J, Rückert C. Complete genome sequence of Corynebacterium casei LMG S-19264T (=DSM 44701T), isolated from a smear-ripened cheese. J Biotechnol. 2014;189: 76–77.

37. Hider RC, Kong X. Chemistry and biology of siderophores. Nat Prod Rep. 2010;27: 637.

38. Saier MH Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res. 2016;44: D372–9.

39. Elkins MF, Earhart CF. Nucleotide sequence and regulation of the Escherichia coli gene for ferrienterobactin transport protein FepB. J Bacteriol. 1989;171: 5443–5451.

40. Shea CM, McIntosh MA. Nucleotide sequence and genetic organization of the ferric enterobactin transport system: homology to other periplasmic binding protein-dependent systems in Escherichia coli. Mol Microbiol. 1991;5: 1415–1428.

41. Chenault SS, Earhart CF. Organization of genes encoding membrane proteins of the Escherichia coli ferrienterobactin permease. Mol Microbiol. 1991;5: 1405–1413.

42. Jochimsen B, Lolle S, McSorley FR, Nabi M, Stougaard J, Zechel DL, et al. Five phosphonate operon gene products as components of a multi-subunit complex of the carbon-phosphorus lyase pathway. Proc Natl Acad Sci U S A. 2011;108: 11393–11398.

43. Osorio H, Martínez V, Nieto PA, Holmes DS, Quatrini R. Microbial iron management mechanisms in extremely acidic environments: comparative genomics evidence for diversity and versatility. BMC Microbiol. 2008;8: 203.

44. Parrow NL, Fleming RE, Minnick MF. Sequestration and scavenging of iron in infection. Infect Immun. 2013;81: 3503–3514.

45. Ellison RT 3rd. The effects of lactoferrin on gram-negative bacteria. Adv Exp Med Biol. 1994;357: 71–90.

46. Sinha B, François PP, Nüße O, Foti M. Fibronectin- binding protein acts as Staphylococcus aureus invasin via fibronectin bridging to integrin α5β1. Cellular. Wiley Online Library; 1999; Available: http://onlinelibrary.wiley.com/doi/10.1046/j.1462-5822.1999.00011.x/full

47. Cheung AL, Ying P. Regulation of alpha- and beta-hemolysins by the sar locus of Staphylococcus aureus. J Bacteriol. Am Soc Microbiol; 1994;176: 580–585.

48. McCourt J, O'Halloran DP, McCarthy H, O'Gara JP, Geoghegan JA. Fibronectin-binding proteins are required for biofilm formation by community-associated methicillin-resistant Staphylococcus aureus strain LAC. FEMS Microbiol Lett. 2014;353: 157–164.

49. Cao J, Woodhall MR, Alvarez J, Cartron ML, Andrews SC. EfeUOB (YcdNOB) is a tripartite, acid-induced and CpxAR-regulated, low-pH Fe2+ transporter that is cryptic in Escherichia coli K-12 but functional in E. coli O157: H7. Mol Microbiol. Wiley Online Library; 2007;65: 857–875.

50. Miethke M, Monteferrante CG, Marahiel MA, van Dijl JM. The Bacillus subtilis EfeUOB transporter is essential for high-affinity acquisition of ferrous and ferric iron. Biochim Biophys Acta. 2013;1833: 2267–2278.

51. Turlin E, Débarbouillé M, Augustyniak K, Gilles A-M, Wandersman C. Staphylococcus aureus FepA and FepB proteins drive heme iron utilization in Escherichia coli. PLoS One. 2013;8: e56529.

52. Kloos WE, Schleifer KH. Characterization of Staphylococcus sciuri sp. nov. and Its Subspecies1. of Systematic and …. ijs.microbiologyresearch.org; 1976; Available: http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-26-1-22

53. Wozniak RAF, Waldor MK. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. 2010;8: 552–563.

54. Ghinet MG, Bordeleau E, Beaudin J, Brzezinski R, Roy S, Burrus V. Uncovering the prevalence and diversity of integrating conjugative elements in actinobacteria. PLoS One. 2011;6: e27846.

55. Subramanya HS, Arciszewska LK, Baker RA, Bird LE, Sherratt DJ, Wigley DB. Crystal structure of the site-specific recombinase, XerD. EMBO J. 1997;16: 5178–5187.

56. Hamilton CM, Lee H, Li PL, Cook DM, Piper KR, von Bodman SB, et al. TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. J Bacteriol. 2000;182: 1541–1548.

57. te Poele EM, Bolhuis H, Dijkhuizen L. Actinomycete integrative and conjugative elements. Antonie Van Leeuwenhoek. 2008;94: 127–143.

58. Bordeleau E, Ghinet MG, Burrus V. Diversity of integrating conjugative elements in actinobacteria: Coexistence of two mechanistically different DNA-translocation systems. Mob Genet Elements. 2012;2: 119–124.

59. Burrus V, Waldor MK. Shaping bacterial genomes with integrative and conjugative

elements. Res Microbiol. 2004;155: 376–386.

60. Robinson RK. Dairy Microbiology Handbook: The Microbiology of Milk and Milk Products. Wiley; 2005.

61. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008;9: 605–618.

62. Cordero OX, Ventouras L-A, DeLong EF, Polz MF. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. Proc Natl Acad Sci U S A. 2012;109: 20059–20064.

63. Verschuere L, Rombaut G, Sorgeloos P, Verstraete W. Probiotic bacteria as biological control agents in aquaculture. Microbiol Mol Biol Rev. Am Soc Microbiol; 2000;64: 655–671.

64. Chaucheyras-Durand F, Durand H. Probiotics in animal nutrition and health. Benef Microbes. 2010;1: 3–9.

65. Cuello-Garcia CA, Brożek JL, Fiocchi A, Pawankar R, Yepes-Nuñez JJ, Terracciano L, et al. Probiotics for the prevention of allergy: A systematic review and meta-analysis of randomized controlled trials. J Allergy Clin Immunol. Elsevier; 2015;136: 952–961.

66. Onubi OJ, Poobalan AS, Dineen B, Marais D, McNeill G. Effects of probiotics on child growth: a systematic review. J Health Popul Nutr. jhpn.biomedcentral.com; 2015;34: 8.

67. McKenzie YA, Thompson J, Gulia P, Lomer M. British Dietetic Association systematic review of systematic reviews and evidence-based practice guidelines for the use of probiotics in the management of irritable bowel syndrome in adults (2016 update). J Hum Nutr Diet. Wiley Online Library; 2016; Available: http://onlinelibrary.wiley.com/doi/10.1111/jhn.12386/full

68. Whitlock DR, Jamas S, Weiss L. Ammonia oxidizing bacteria for treatment of acne [Internet]. US Patent. 20160151427:A1, 2016. Available: https://www.google.com/patents/US20160151427

69. Sanders ME, Akkermans LMA, Haller D, Hammerman C, Heimbach J, Hörmannsperger G, et al. Safety assessment of probiotics for human use. Gut Microbes. 2010;1: 164–185.

70. Gyles C, Boerlin P. Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease. Vet Pathol. 2014;51: 328–340.

71. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell. 2009;21: 1897–1911.

72. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19: 455–477.

73. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res. 2012;40: D115–22.

74. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29: 2933–2935.

75. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37: D141–5.

76. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003;52: 696–704.

77. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44: W242–5.

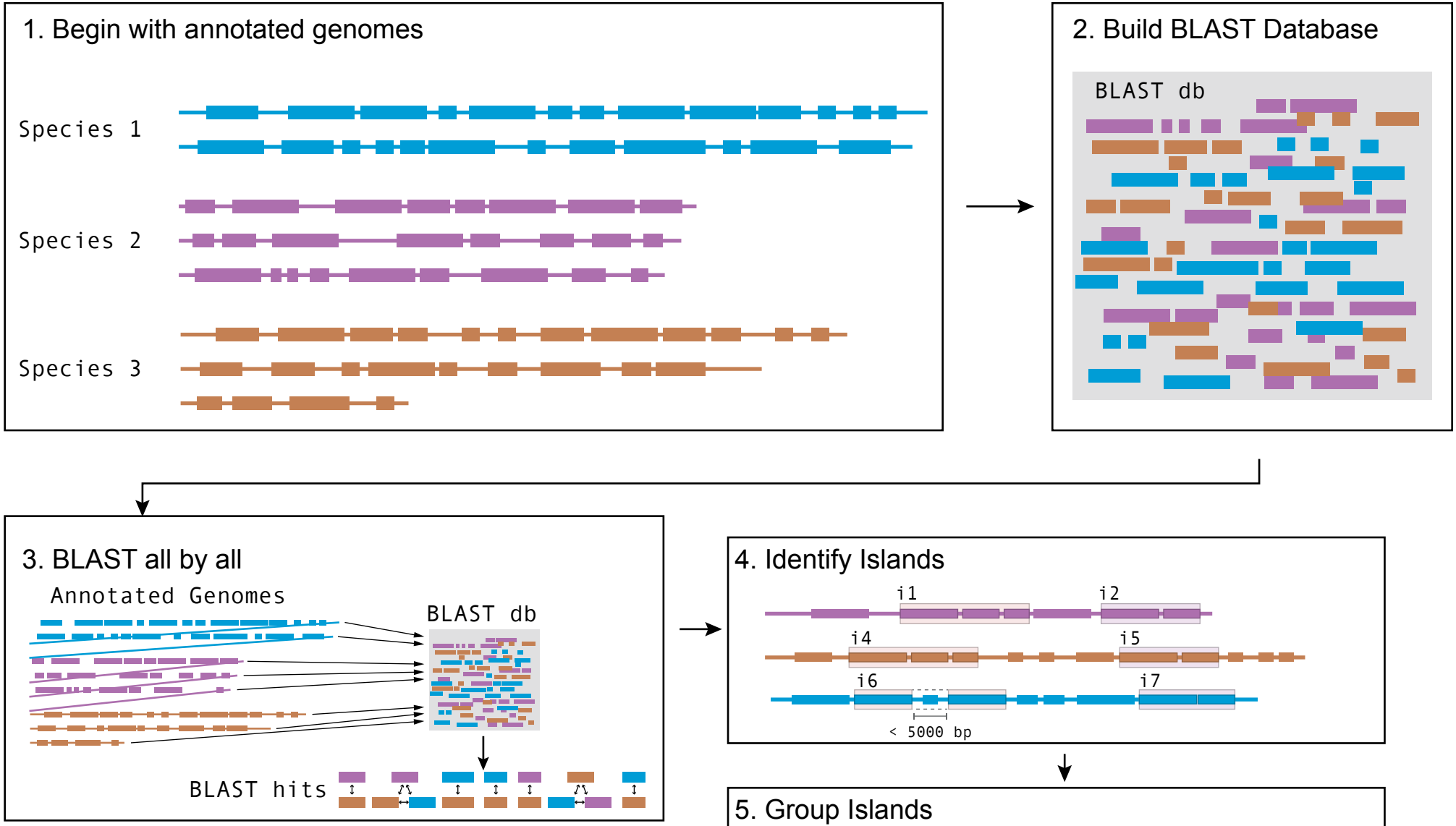
78. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505: 559–563.

79. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, et al. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res. 2013;41: e140–e140.

80. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.

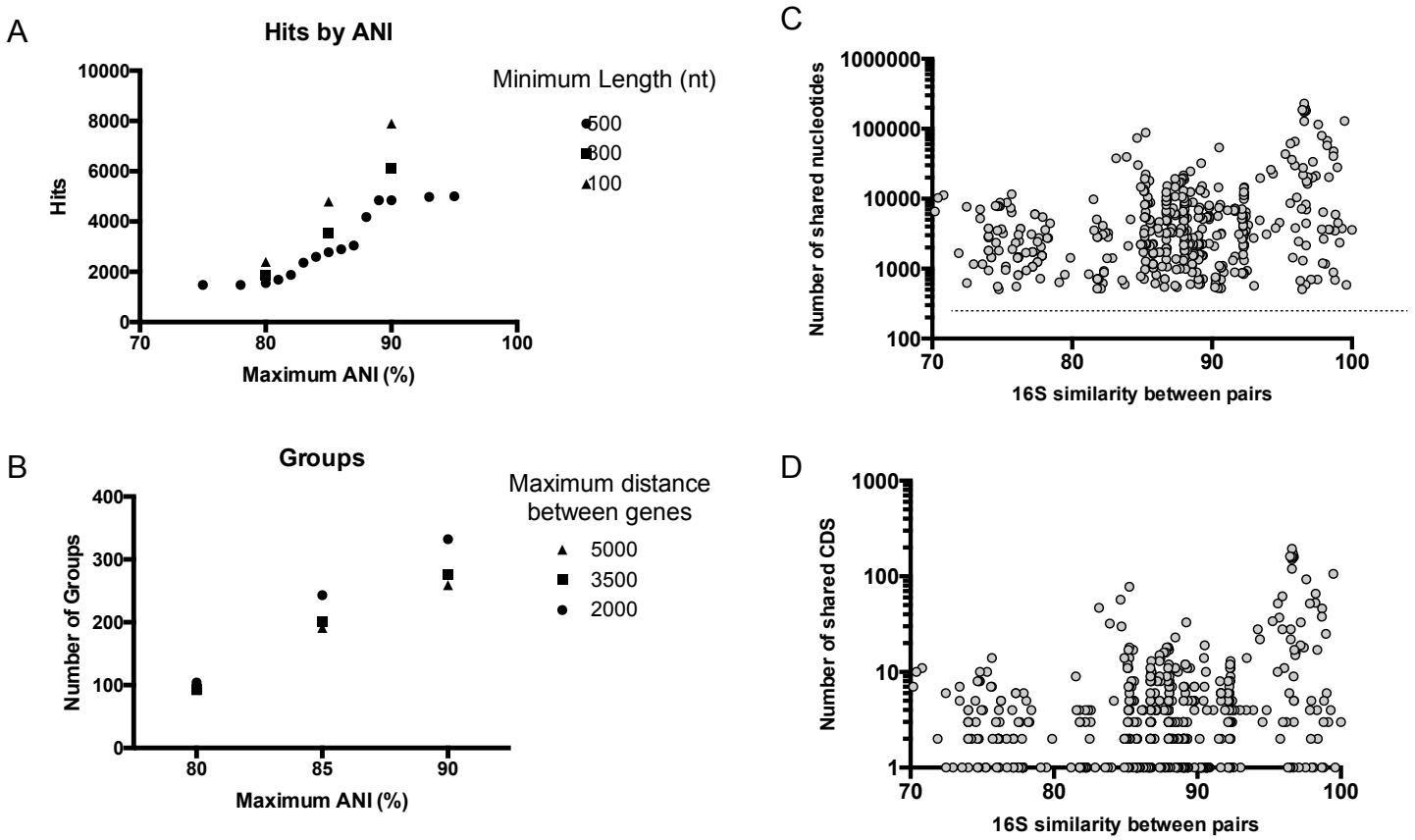
81. Bonham KS. Kvasir: Release for Publication [Internet]. Zenodo; 2016. doi:10.5281/zenodo.53899

Supplementary Figure 1



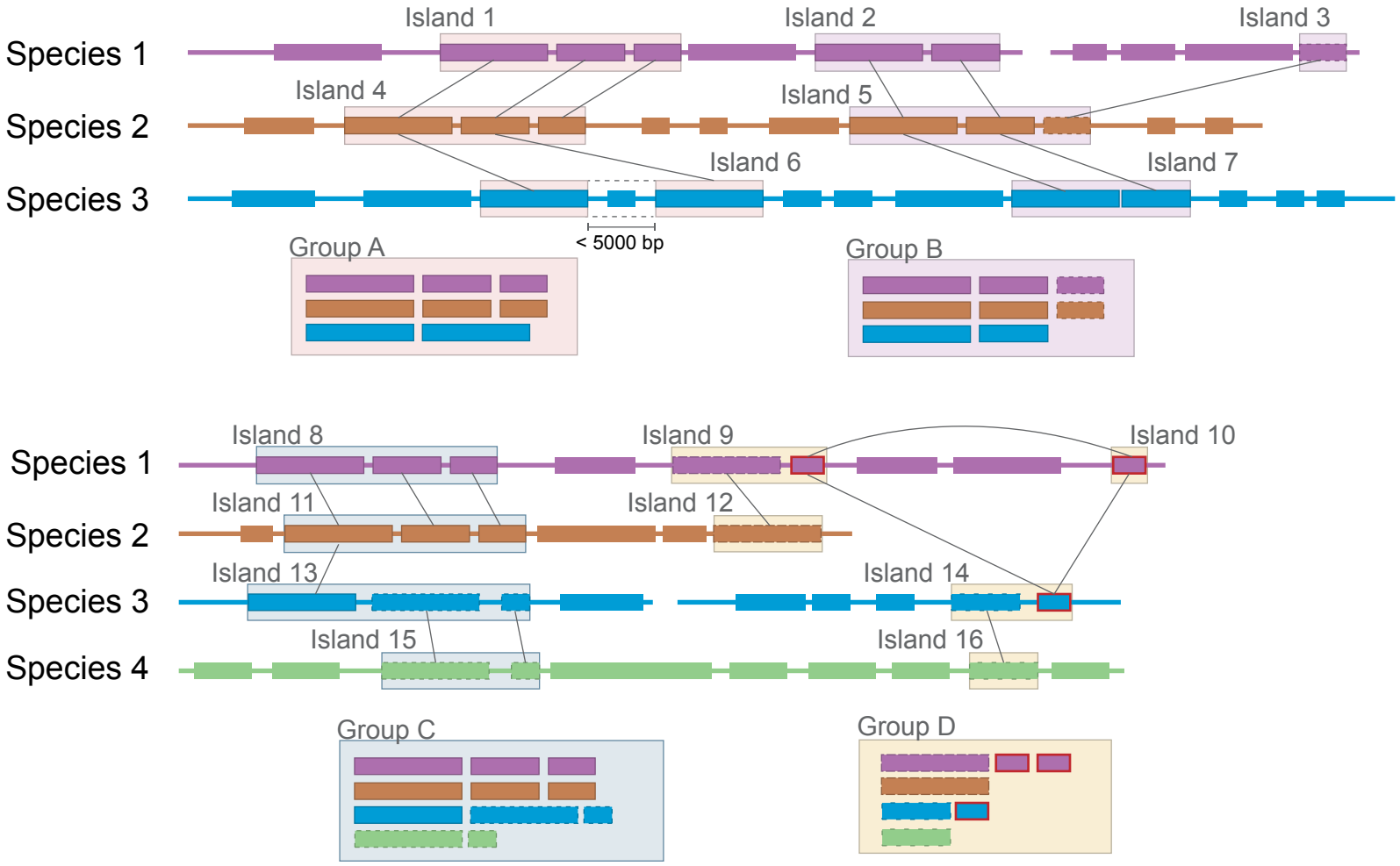
Supplementary Figure 1: Kvasir Schematic
Schematic of software pipeline to identify HGT. (1) Sequenced genomes are annotated with IMG/ER and downloaded in Genbank format. (2) All annotated genes in all genomes are used to assemble a BLAST database using BLAST+ command-line tools. (3) All protein-coding genes (CDS) from all species are queried against the BLAST database. Hits from the same species are discarded; hits from species with an ANI >89% were discarded; other hits are saved. (4) For each species, coding sequences that have at least one BLAST hit are grouped into islands based on proximity. Genes that are within 5kb of each other on the same contig are considered part of the same island. (5) Islands in each species are compared to islands in each other species to form groups. Islands that share at least 1 gene in common according to BLAST parameters in step 3 are placed in the same group.

Supplementary Figure 2



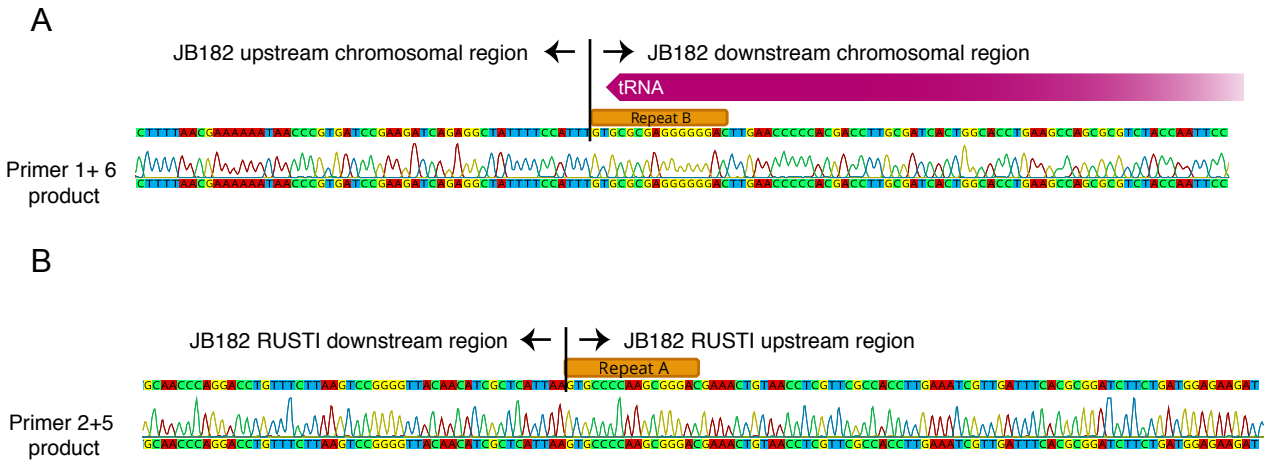
Supplementary Figure 2: HGT Identification Parameters
Different parameters for minimum length of gene match for HGT, maximum % ANI identity for related species, and maximum distance between genes in an island were compared. (A) Number of positive HGT hits identified when varying the minimum protein coding gene length. (B) Number of HGT groups constructed when varying the maximum separation between hits that are classified as belonging to the same group. (C) Number of nucleotides or (D) number of protein coding sequences in HGT regions by 16S similarity. Note - since 500bp is the minimum length for protein coding sequences in this analysis.

Supplementary Figure 3



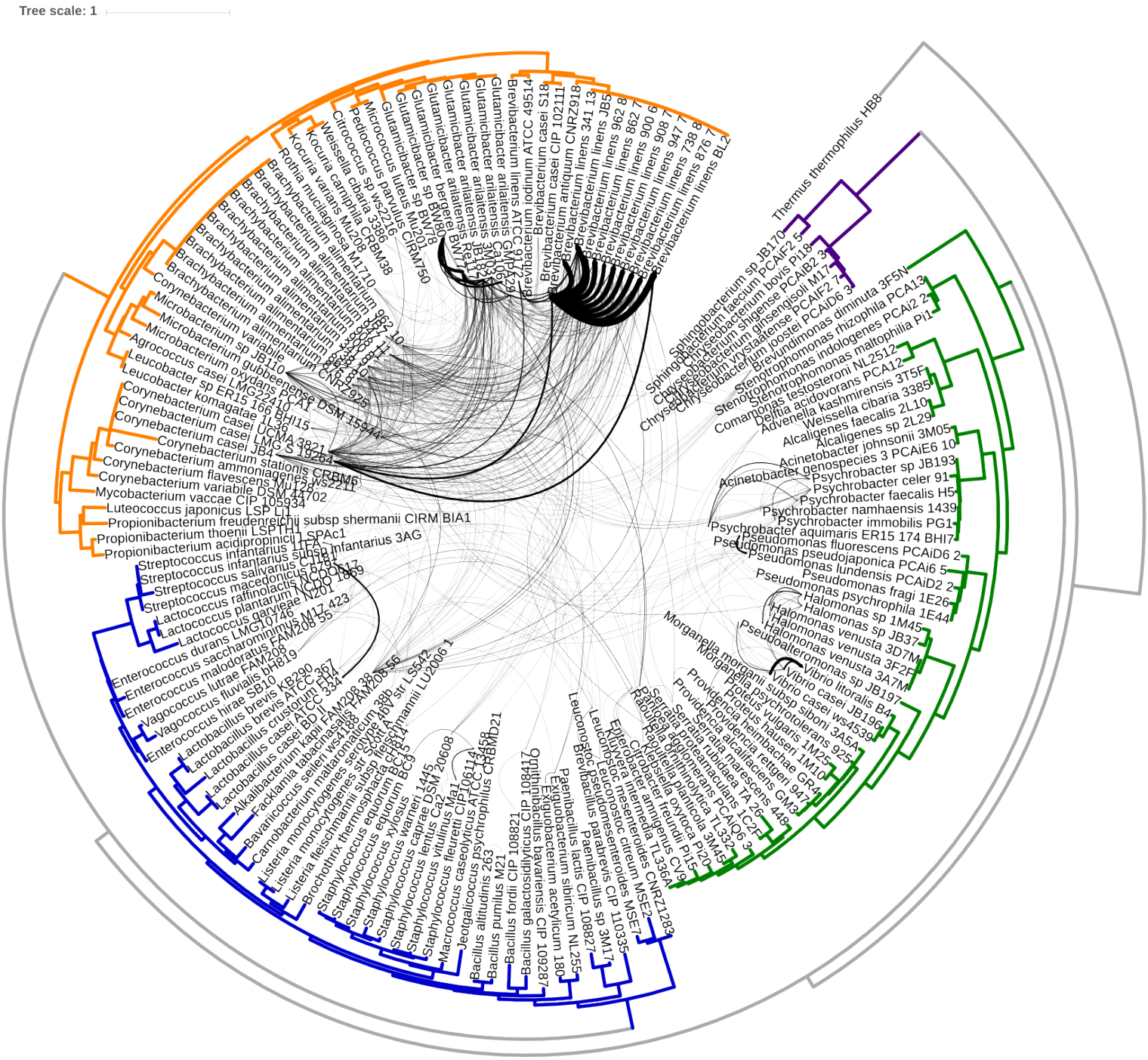
Supplementary Figure 3: Group Clustering
(A) Expected clustering: contiguous genes in multiple species are in a single group. Though island 6 (i6) lacks one gene present in i1 and i4, (possibly because of a transposon insertion), it is still correctly considered related. (B) Ambiguous grouping: Islands 2 and 3 from species 1 are found on different contigs, but are grouped together. They may be found in close proximity in the genome, but on different sides of a gap in the assembly, or they may be quite distant from each other. The grouping of related genes in species 2 into a single island suggests that they may have been transferred in a single event, but the possibility of two unrelated HGT events landing in the same spot cannot be excluded. (C) Possible mis-grouping of two HGT events in a single group: Though species 4 does not share any genes with species 1 and 2, these islands are nevertheless clustered due to the proximity of coding sequences in species 3. This may correctly represent a single gene cluster that subsequently diverged in each species, or unrelated HGT that happened to insert in close proximity. (D) Mis-grouping due to mobile element: Mobile elements (outlined in red) found in multiple locations in multiple genomes may insert next to unrelated HGT islands, causing spurious grouping by the algorithm.

Supplementary Figure 4



Supplementary Figure 4: ICE Excision Sequencing
The ~2000bp band from the PCR amplification using primers 1 and 6 (A) and the ~500bp band from amplification using primers 2 and 5 (B) were extracted and sequenced. Alignment with the JB182 genome reveals 100% alignment with expected and the spliced chromosomal region containing the 3' repeat and the excision circle containing the 3' repeat respectively.

Supplementary Figure 5


Tree scale: 1
Thermus thermophilus HB8
Sphingobacterium sp JB170
Sphingobacterium faecium PCAiF2 5
Chryseobacterium bovis Pi18
Chryseobacterium shigense PCAiB2 3
Chryseobacterium ginsengisoli M17
Chryseobacterium wystaatense PCAiF2 7
Chryseobacterium joostei PCAiD6 3
Brevundimonas diminuta 3F5N
Stenotrophomonas rhizophila PCAi2 2
Stenotrophomonas indologenes PCAi1
Stenotrophomonas maltophilia Pi1
Comamonas testosteroni NL2512
Delftia acidovorans PCA12
Weissella kashmirensis 3T5F
Advenella cibaria 3385
Alcaligenes faecalis 2L10
Alcaligenes sp 2L29
Acinetobacter johnsonii 3M05
Acinetobacter genospecies 3 PCAiE6 10
Psychrobacter sp JB193
Psychrobacter celer 91
Psychrobacter faecalis H5
Psychrobacter namhaensis 1439
Psychrobacter immobilis PG1
Psychrobacter aquimaris ER15 174 BHI7
Pseudomonas fluorescens PCAiD6 2
Pseudomonas pseudojaponica PCAi6 5
Pseudomonas lundensis PCAiD2 2
Pseudomonas fragi 1E26
Pseudomonas psychrophila 1E44
Halomonas sp 1M45
Halomonas sp JB37
Halomonas venusta 3D7M
Halomonas venusta 3F2F
Halomonas venusta 3A7M
Pseudoalteromonas sp JB197
Vibrio litoralis B4
Vibrio casei JB196
Vibrio casei ws4539
Morganella morganii subsp siboni 3A5A
Morganella psychrotolerans 925
Proteus vulgaris 1M25
Proteus hauseri 1M10
Providencia heimbachae GR4
Providencia alcalifaciens 445
Providencia rettgeri 947
Serratia marescens GM3
Serratia rubidaea TA 26
Serratia plymuthica 1C2F
Serratia proteamaculans PCAiQ6 3
Serratia grimesii 3M45
Serratia liquefaciens PI20
Serratia fonticola CV9
Serratia ficaria PI15
Rahnella aquatilis 1L336A
Klebsiella planticola
Klebsiella oxytoca
Klebsiella pneumoniae
Klebsiella ornithinolytica TL332
Citrobacter freundii PI1
Citrobacter amalonaticus
Citrobacter intermedia
Enterobacter mesenteroides CNRZ1283
Enterobacter cloacae MSE2
Enterobacter aerogenes MSE7
Leuconostoc citreum 110335
Leuconostoc mesenteroides
Leuconostoc pseudomesenteroides
Leuconostoc gelidum
Leuconostoc carnosum
Paenibacillus sp 3M17
Paenibacillus parabrevis CIP 108827
Paenibacillus lactis CIP 108827
Brevibacillus sibiricum NL255
Exiguobacterium acetylicum 180
Exiguobacterium sibiricum
Oceanobacillus iheyensis CIP 109287
Oceanobacillus kimchi CIP 108458
Oceanobacillus caseolyticus ATCC 13
Bacillus fordii CIP 108821
Bacillus galactosidilyticus CIP 108417
Bacillus pumilus M21
Bacillus altitudinis 263
Jeotgalicoccus psychrophilus CRBMDD21
Macrococcus caseolyticus
Macrococcus fleurettii CIP106114
Staphylococcus vitulinus Ma1
Staphylococcus lentus Ca2
Staphylococcus caprae DSM 20608
Staphylococcus warneri 1445
Staphylococcus xylosus
Staphylococcus equorum BC9
Staphylococcus equorum
Staphylococcus saprophyticus
Staphylococcus hominis
Staphylococcus cohnii
Staphylococcus sciuri
Staphylococcus aureus str LS42
Staphylococcus aureus str LS42
Staphylococcus aureus str LS42