

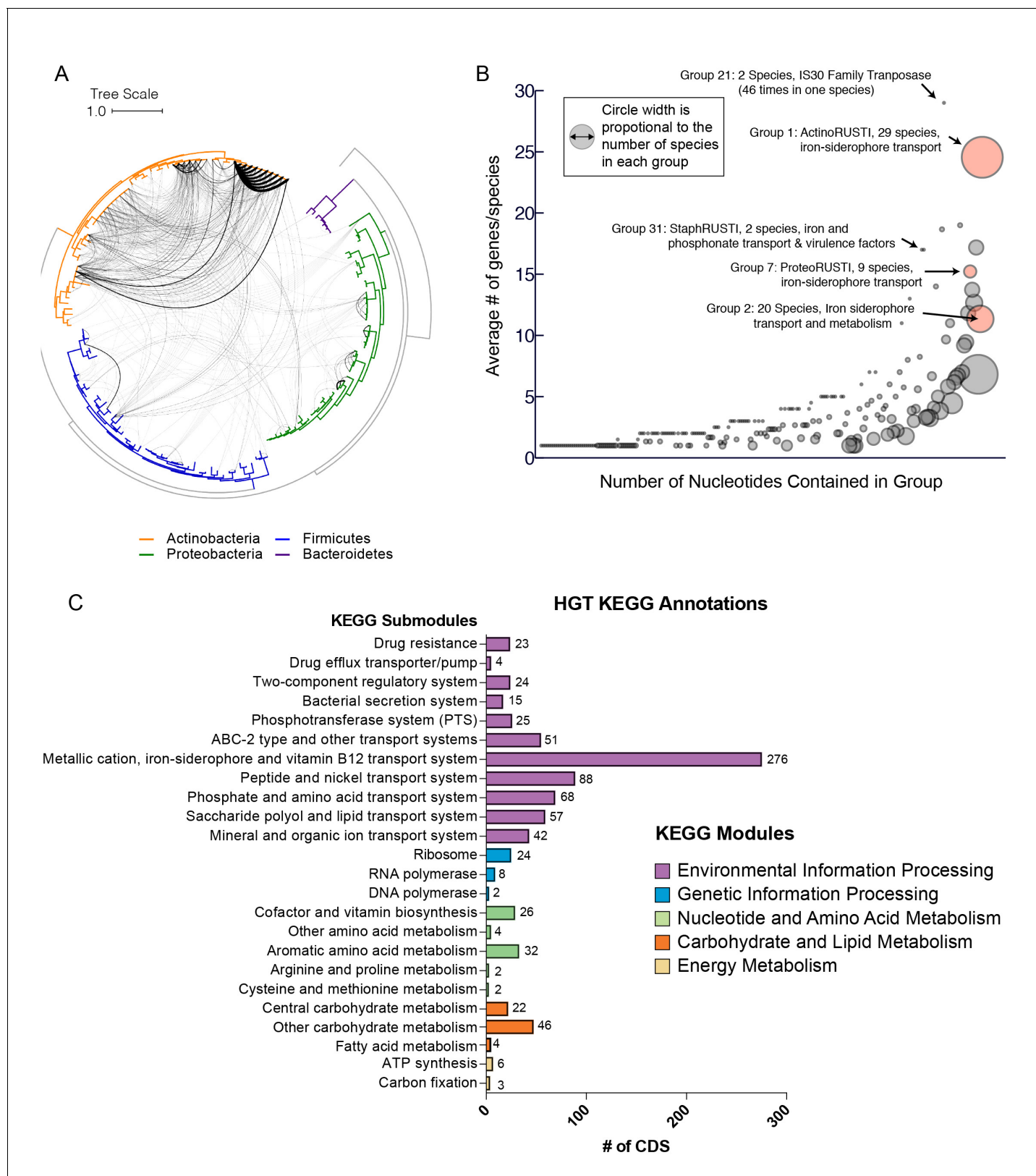


---

## Figures and figure supplements

Extensive horizontal gene transfer in cheese-associated bacteria

**Kevin S Bonham *et al***

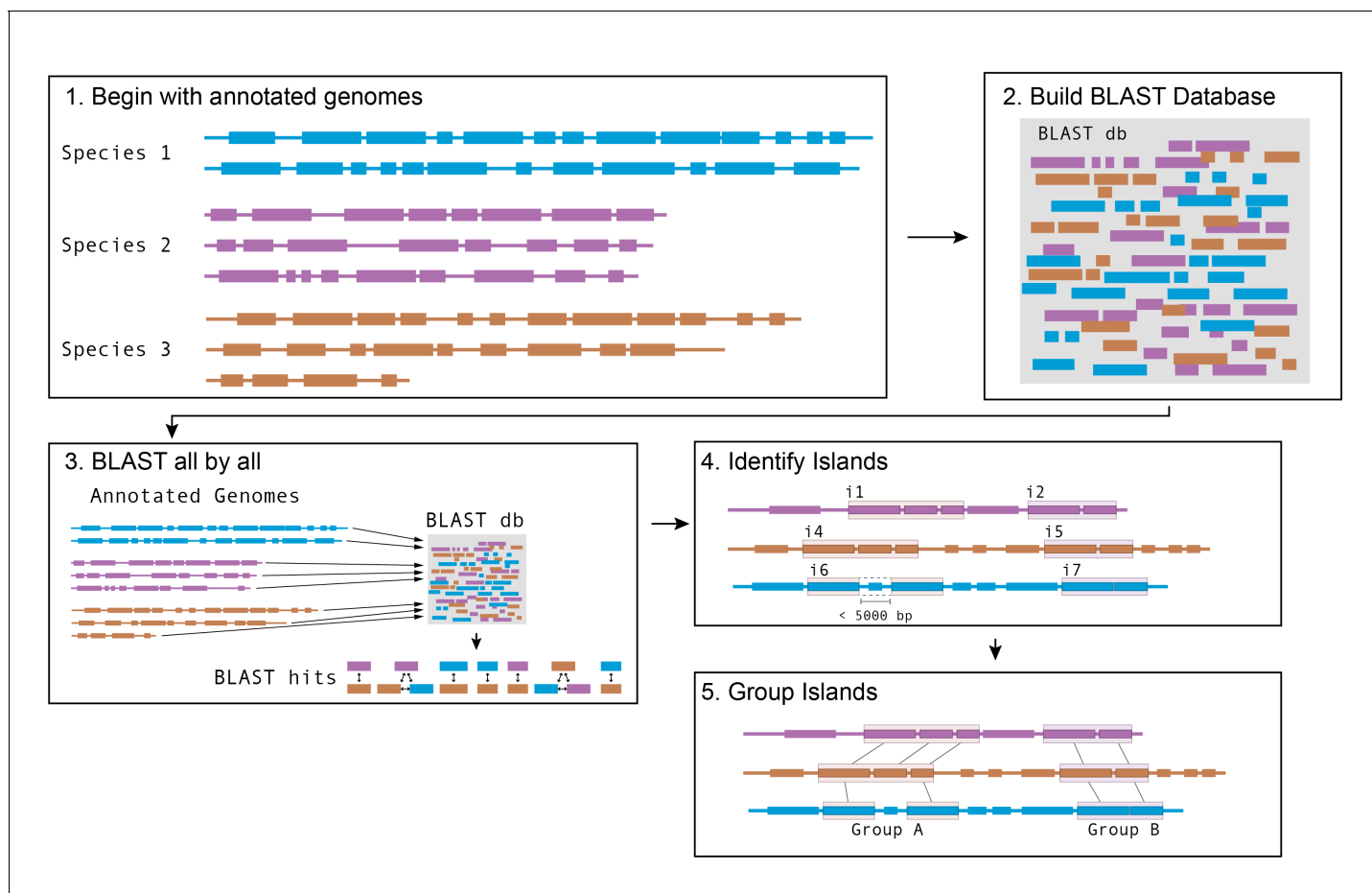


**Figure 1.** Extensive horizontal gene transfer in the cheese microbiome. (A) All HGT events in analyzed cheese-associated bacteria. Connection thickness is scaled to number of shared protein coding sequences. Maximum likelihood tree based on 16S RNA alignment using Ribosomal Database Project (RDP). (B) HGT events clustered into 264 'groups' based on genomic proximity. Groups are plotted based on total nucleotide content (x-axis, Figure 1 continued on next page

*Figure 1 continued*

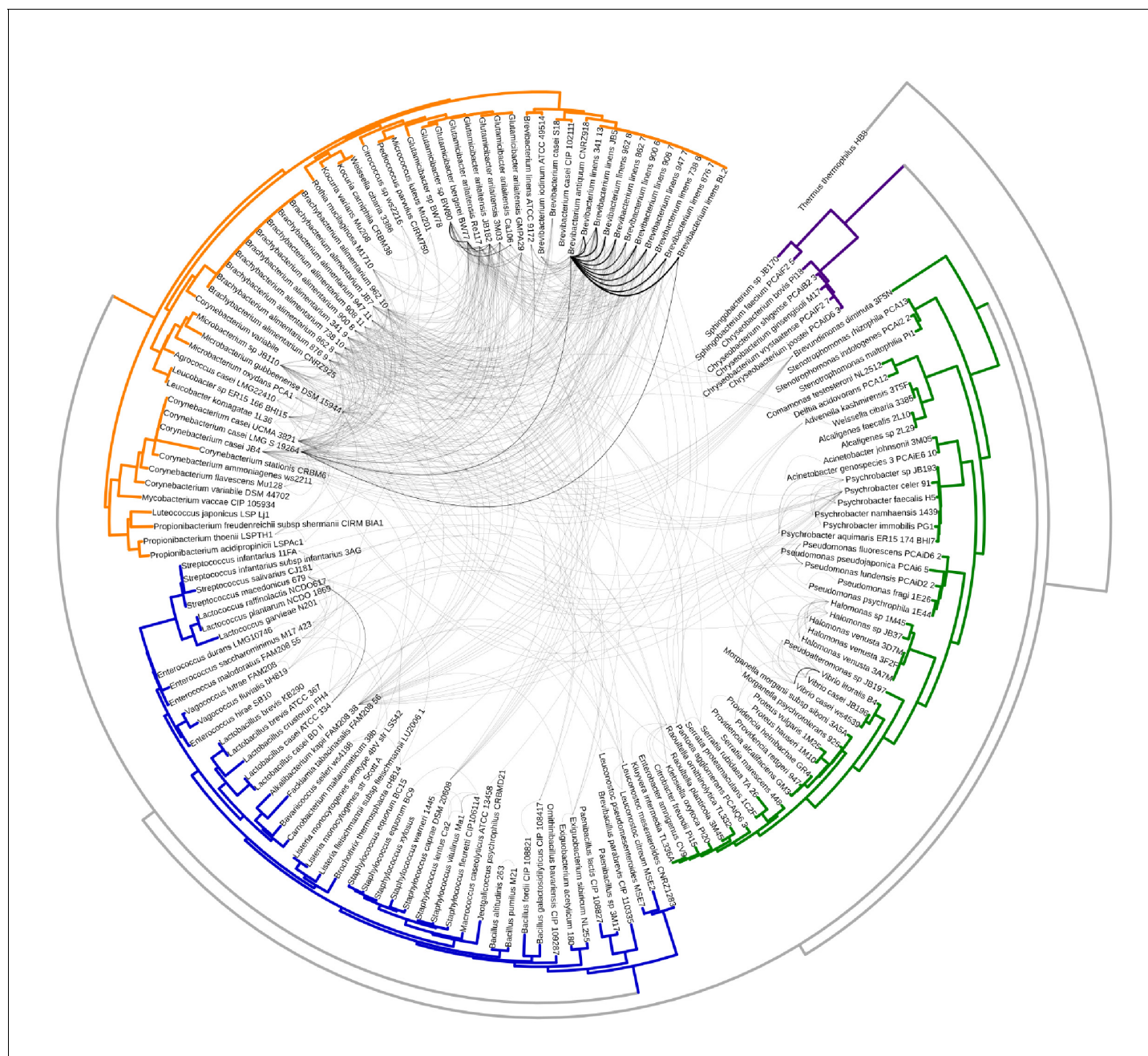
from low to high), and the mean number of genes per species (y-axis). Diameter of each circle is proportional to the total number of species in the group. Groups highlighted in red are described further in the text. **(C)** Quantification of KEGG modules and submodules for protein coding genes (CDS) identified as horizontally transferred. Annotations were generated by BLAST-Koala. Genes without function prediction are not depicted.

DOI: [10.7554/eLife.22144.003](https://doi.org/10.7554/eLife.22144.003)



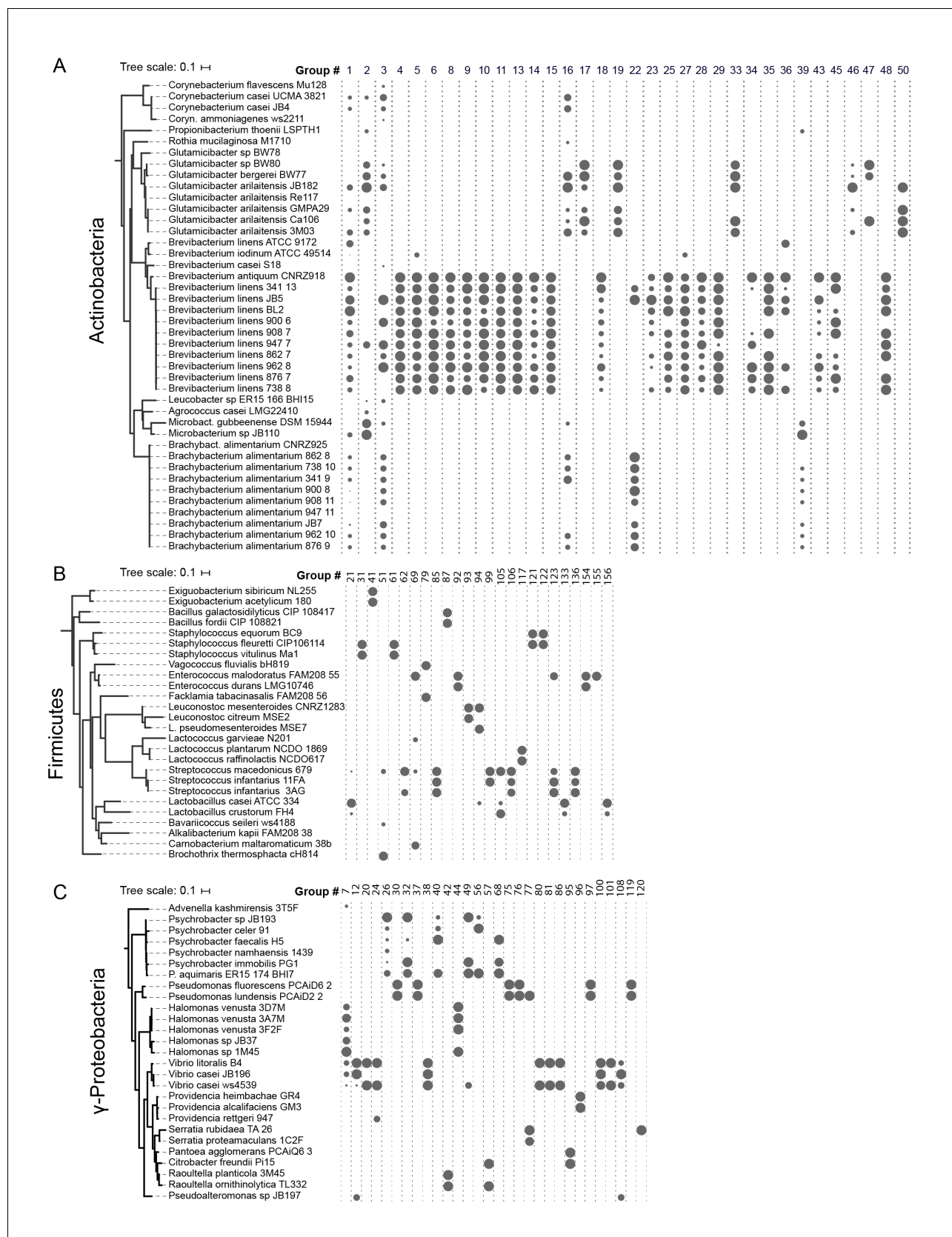
**Figure 1—figure supplement 1.** Schematic of software pipeline to identify HGT. (1) Sequenced genomes are annotated with IMG/ER and downloaded in Genbank format. (2) All annotated genes in all genomes are used to assemble a BLAST database using BLAST+ command-line tools. (3) All protein coding genes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

DOI: [10.7554/eLife.22144.004](https://doi.org/10.7554/eLife.22144.004)



**Figure 1—figure supplement 2.** Same as **Figure 1A** with branch labels. 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).

DOI: 10.7554/eLife.22144.005



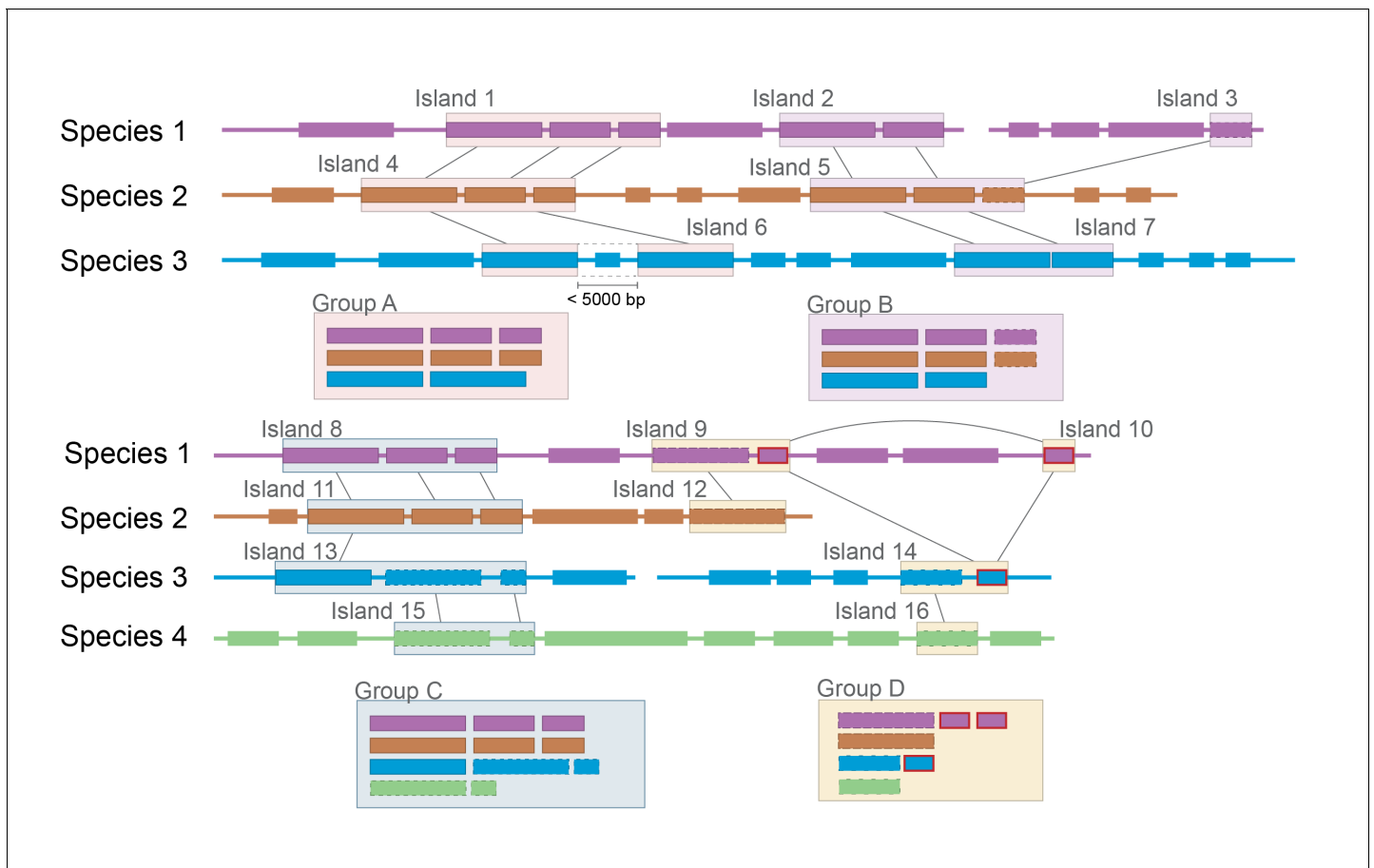
**Figure 2.** HGT Groups in Actinobacteria, Firmicutes, and  $\gamma$ -Proteobacteria groups. (A) The 31 largest HGT groups that contain predominantly Actinobacteria. The areas of circles are scaled to  $\log_2(n)$ , where  $n$  is the total number of nucleotides in that group for each species. The largest circle

Figure 2 continued on next page

*Figure 2 continued*

size represents the largest HGT group in that phylum. 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  $\gamma$ -Proteobacteria.

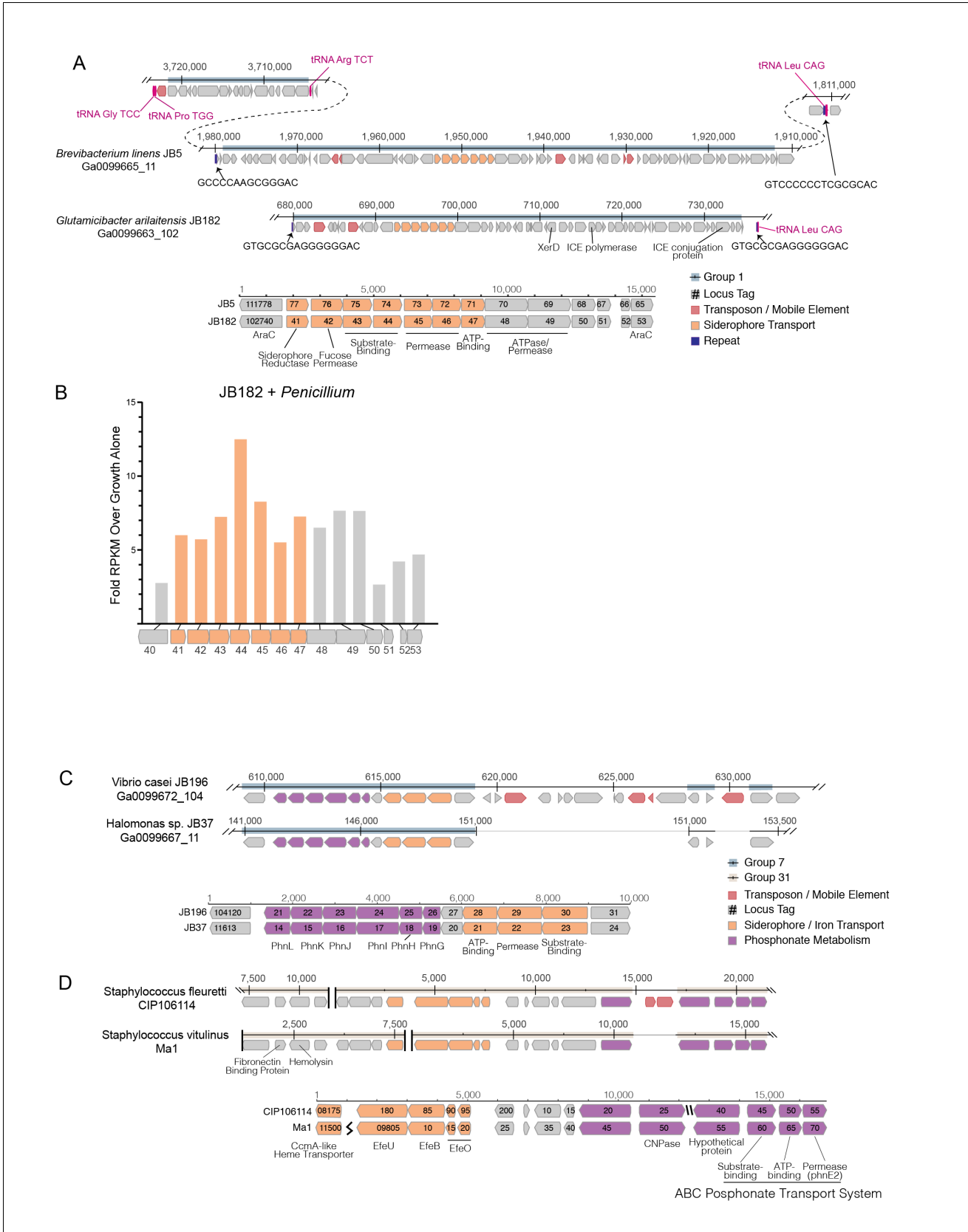
DOI: [10.7554/eLife.22144.006](https://doi.org/10.7554/eLife.22144.006)



**Figure 2—figure supplement 1.** Group A: Expected clustering: contiguous genes in multiple species are in a single group. Although island 6 (i6) lacks one gene present in i1 and i4, (possibly because of a transposon insertion), it is still considered related. Group 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. Group C: Possible mis-grouping of two HGT events in a single group: although species 4 does not share any genes with species 1 and 2, these islands are nevertheless clustered because of 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. Group D: Mis-grouping because of 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.

DOI: [10.7554/eLife.22144.007](https://doi.org/10.7554/eLife.22144.007)

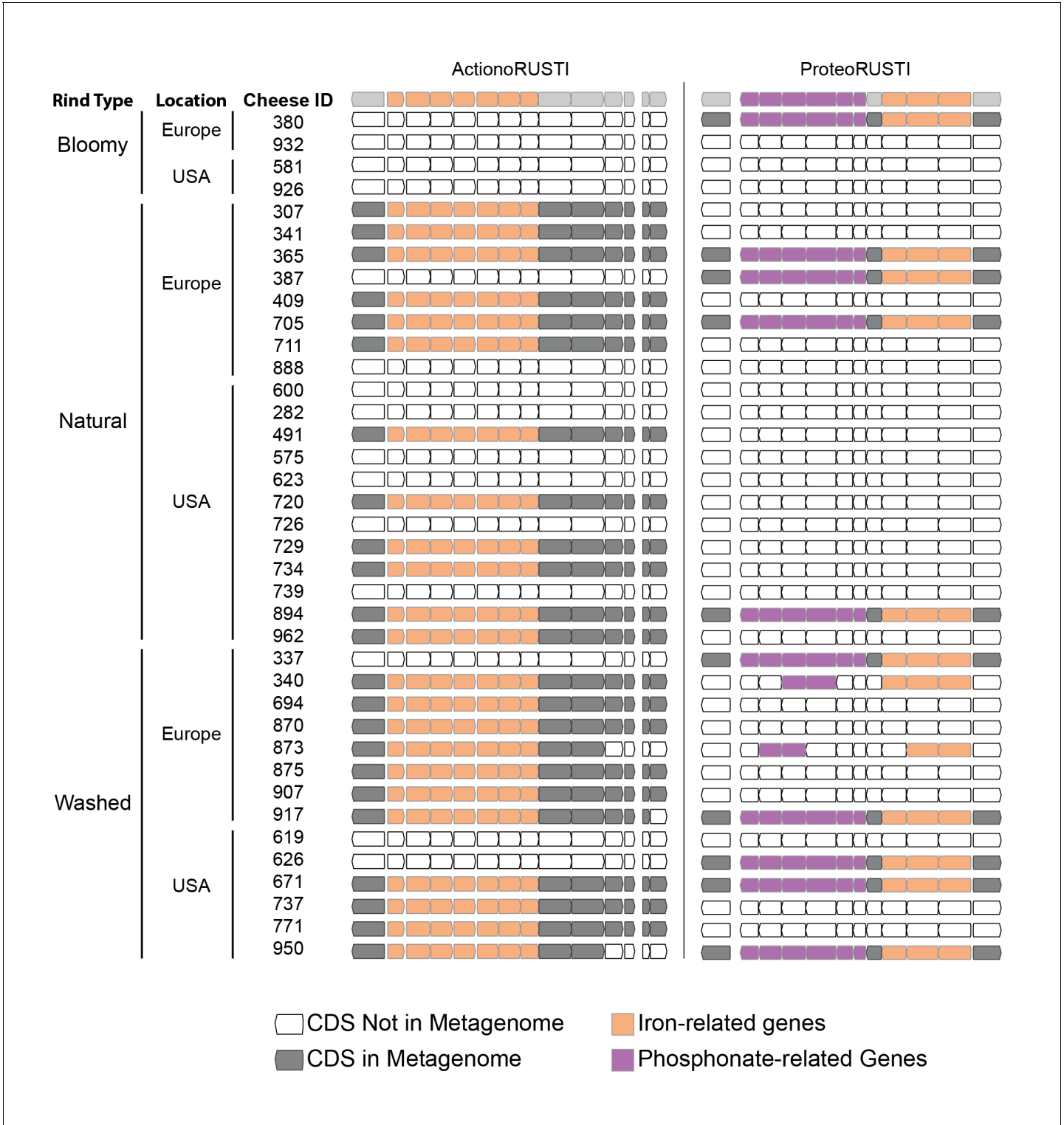




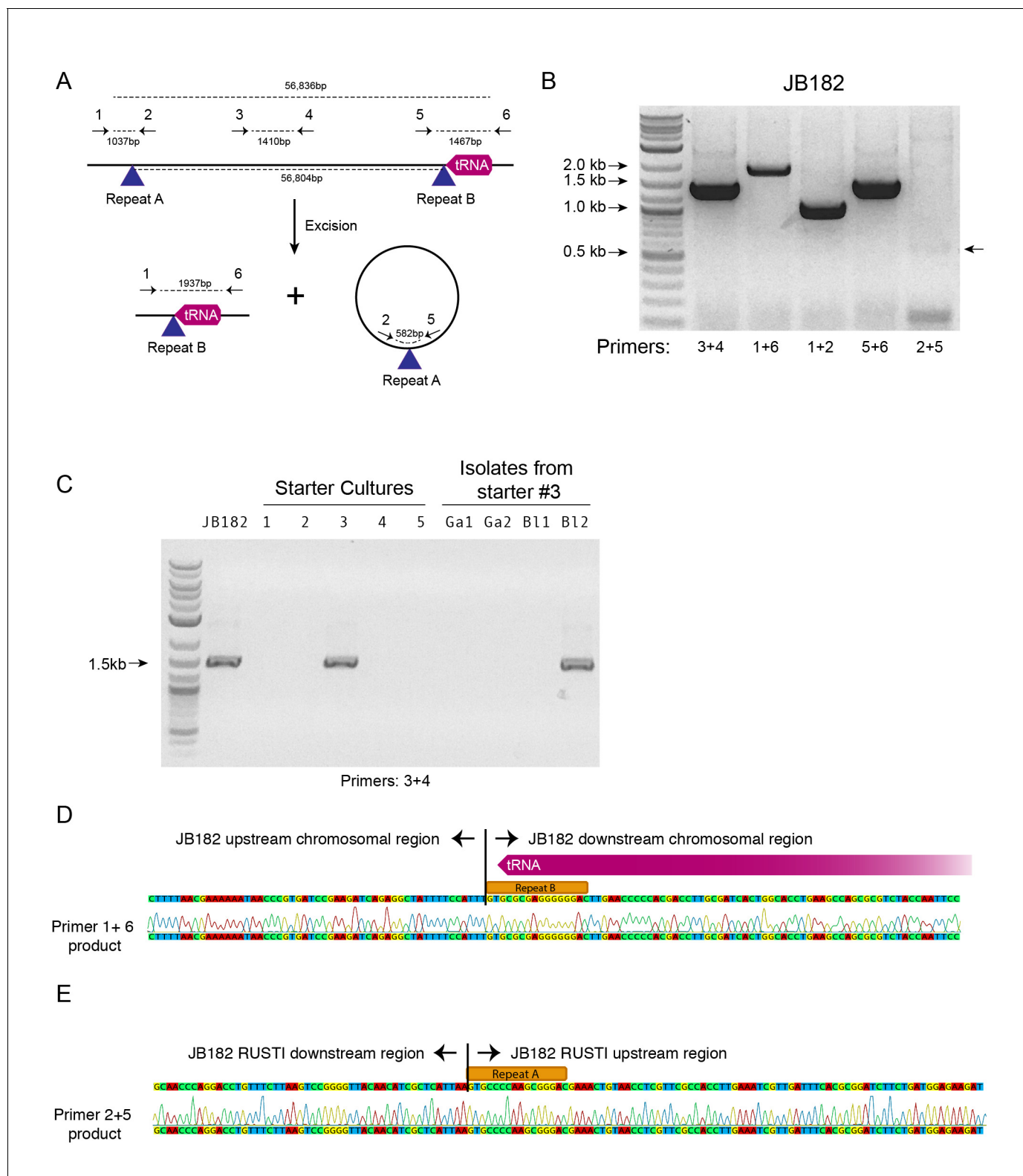
## Figure 3 continued

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 depict gaps in the alignment resulting from 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 two 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).

DOI: [10.7554/eLife.22144.008](https://doi.org/10.7554/eLife.22144.008)



**Figure 4.** Presence of RUSTI in cheese metagenomes. Genes in ActinoRUSTI (*G. arilaitensis* JB182) and ProteoRUSTI (*V. casei* JB196) regions were compared with 32 assembled metagenomes from the US and Europe. Filled CDS represents a positive (>97% identical nucleotides) hit in that metagenome.  
DOI: 10.7554/eLife.22144.009



**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 five 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 four isolates selected based on colony morphology were also tested. The expected size for the amplicon is ~1.4 kb. Sequencing of the 16S ribosomal RNA genes for these isolates suggested that two isolates are *Glutamicibacter arilaitensis* and two are *Brevibacterium linens*. *G. arilaitensis*. Figure 5 continued on next page

*Figure 5 continued*

JB182 was used as a positive control. (D) The ~2000 bp band from the PCR amplification using primers 1 and 6 and (E) the ~500 bp band from amplification using primers 2 and 5 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.

DOI: [10.7554/eLife.22144.010](https://doi.org/10.7554/eLife.22144.010)