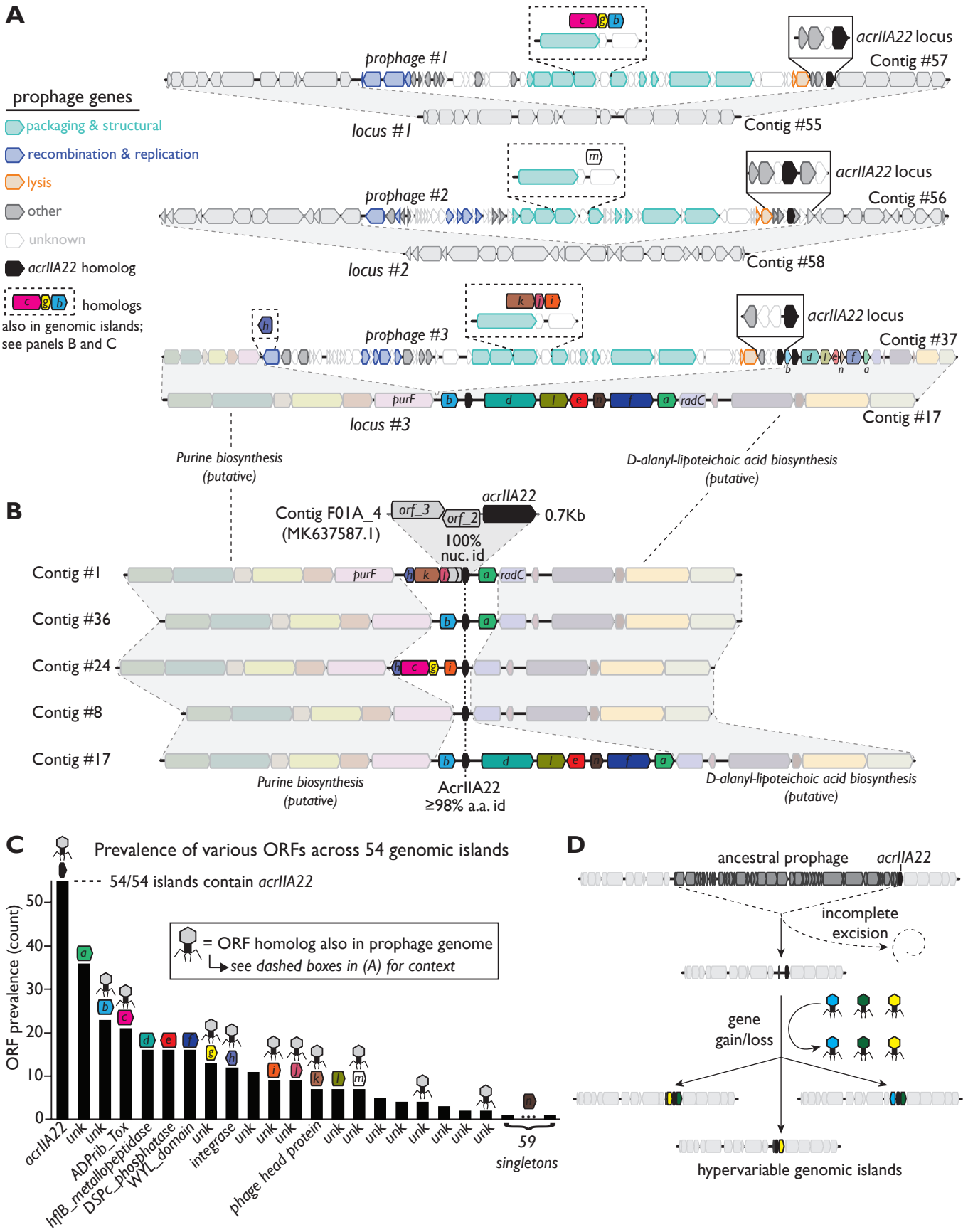
A
prophage genes
packaging & structural
recombination & replication
lysis
other
unknown
acrIIA22 homolog
homologs also in genomic islands; see panels B and C
prophage #1
acrIIA22 locus
Contig #57
locus #1
Contig #55
prophage #2
acrIIA22 locus
Contig #56
locus #2
Contig #58
prophage #3
acrIIA22 locus
Contig #37
purF
radC
locus #3
Contig #17
Purine biosynthesis (putative)
D-alanyl-lipoteichoic acid biosynthesis (putative)
B
Contig F01A_4 (MK637587.1)
orf_3
orf_2
acrIIA22
0.7Kb
100% nuc. id
Contig #1
Contig #36
Contig #24
Contig #8
Contig #17
Purine biosynthesis (putative)
AcrIIA22 ≥98% a.a. id
D-alanyl-lipoteichoic acid biosynthesis (putative)
C
Prevalence of various ORFs across 54 genomic islands
54/54 islands contain acrIIA22
= ORF homolog also in prophage genome
see dashed boxes in (A) for context
ORF prevalence (count)
0
10
20
30
40
50
acrIIA22
unk
ADPrib_Tox
hflB_metallopeptidase
DSPc_phosphatase
WYL_domain
integrase
phage head protein
59 singletons
D
ancestral prophage
acrIIA22
incomplete excision
gene gain/loss
hypervariable genomic islands

S4 Fig. .

S4 Fig. ***AcrIIA22* homologs are found in hypervariable regions of prophage and bacterial genomes in the CAG-217clostridial genus.** (**A**) Homologs of *acrIIA22* are depicted in three related prophage genomes, integrated at three different genomic loci, revealed by a comparison of prophage-bearing contigs (#57, #56, #37) relative to unintegrated contigs (#55, #58, #17 respectively), which are otherwise nearly identical. Prophage genes are colored by functional category, according to the legend at the left of panel A. Genes immediately adjacent to *acrIIA22* (solid boxes) vary across phages, despite strong relatedness across much of the prophage genomes. Bacterial genes are colored gray, except for contig #17, which is also depicted in panel B, below. (**B**) Homologs of *acrIIA22* are depicted in diverse genomic islands, including Contig #1, whose sequence includes a portion that is identical to the original metagenomic contig we recovered (F01A_4). All *acrIIA22* homologs in these loci are closely related but differ in their adjacent genes, which often have homologs in the prophages depicted in panel A (dashed boxes). Bacterial genomic regions flanking these hypervariable islands are nearly identical to one another and to prophage integration locus #3, as shown by homology to contig #17 from panel A. Contigs are numbered to indicate their descriptions in S3 Table, which contains their metadata, taxonomy, and sequence retrieval information. All sequences and annotations can also be found in S2 Data and S3 Data. (**C**) We tabulate the prevalence of various protein families (clustered at 65% amino acid identity) in a set of 54 unique genomic islands. Each of these islands is flanked by the conserved genes *purF* and *radC* but contains a different arrangement of encoded genes. Domain-level annotations are indicated below each protein family (unk; unknown function). Gene symbols above each protein family are colored and lettered to indicate their counterparts or homologs in panels A and B. The phage capsid icon indicates sequences with homologs in prophage genomes. (**D**) An evolutionary model for the origin of the *acrIIA22*-encoding hypervariable genomic islands depicted in panel B is shown. This panel is reprinted from Fig 2C, for continuity. We propose that *acrIIA22* moved via a phage insertion into a bacterial genomic locus, remained following an incomplete prophage excision event, and its neighboring genes subsequently diversified via horizontal exchange with additional phage genomes. The individual numerical values that underlie the summary data in this figure may be found in S1 Data.