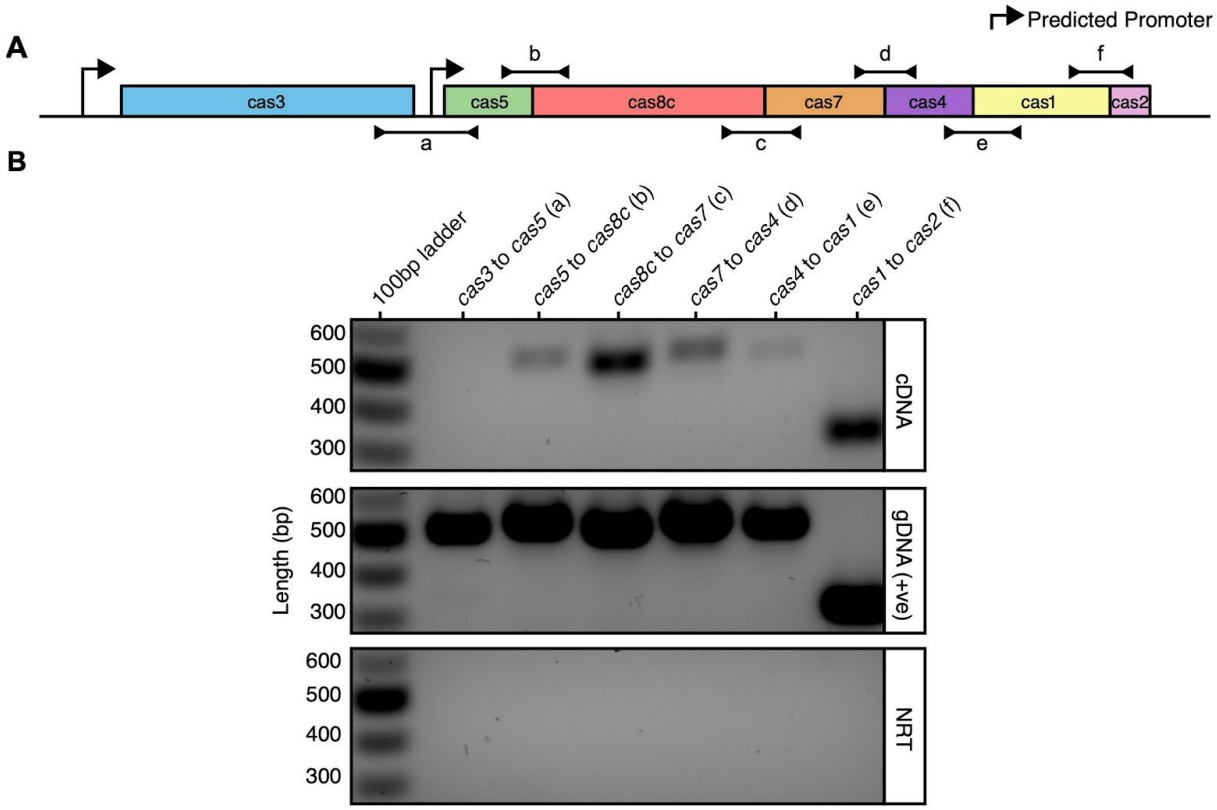

Figure S1. Mapping of the *cas* operon related to Figure 1. (**A**) Schematic of the computationally predicted promoters and primers designed to amplify regions that overlap each pair of neighboring genes. (**B**) The genes *cas3* and *cas5* are regulated by distinct promoter sequences. All of the other tested gene pairs were detected on the same transcript. A positive genomic DNA (gDNA) control is shown in addition to a negative control: NRT (no reverse transcriptase).

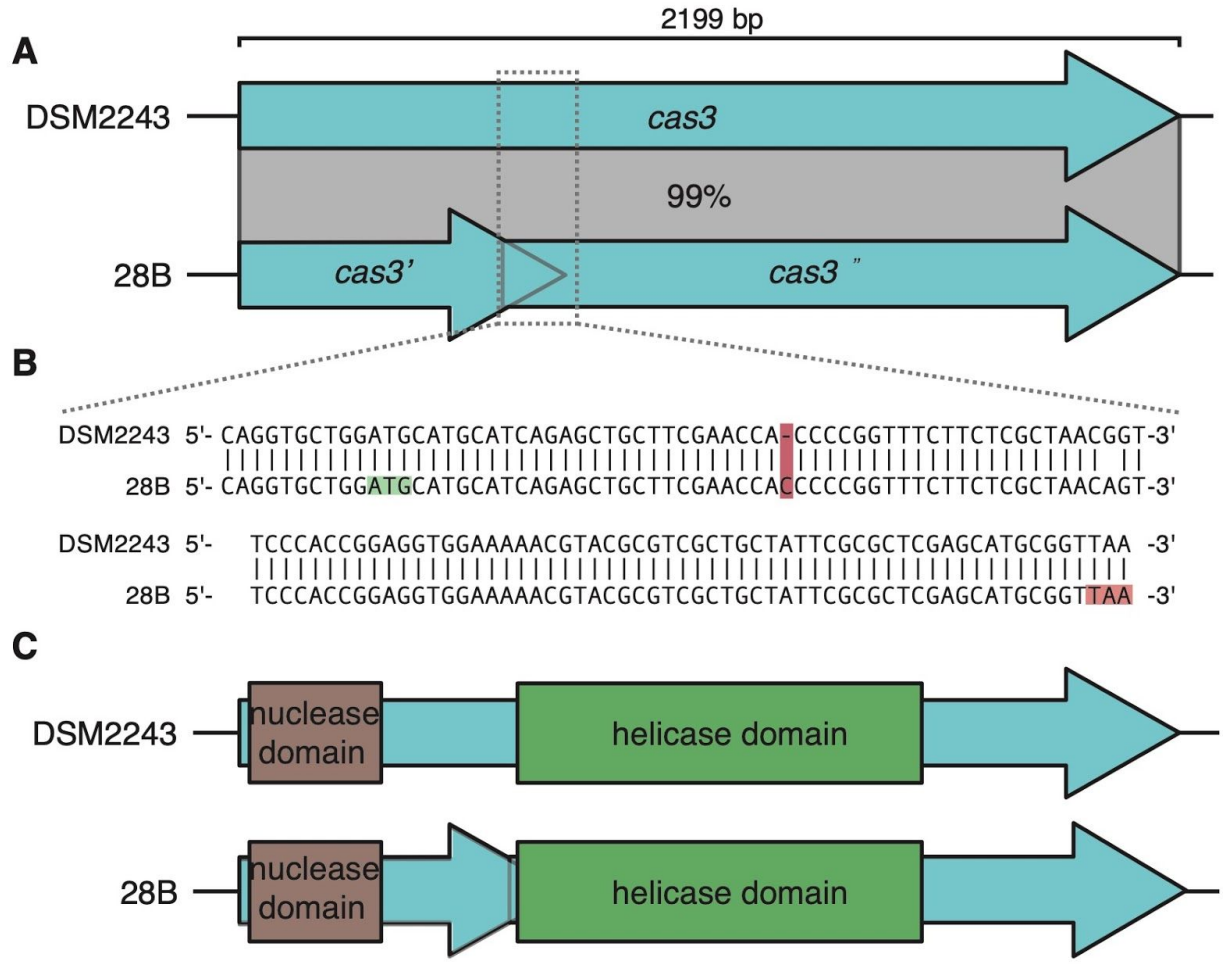

Figure S2. Truncated *cas3* in *E. lenta* 28B related to Figure 3. (**A**) The annotation of *cas3* in *E. lenta* DSM 2243 denotes a 2,199 sequence that comprises a single protein while in strain 28B, two distinct proteins are predicted: cas3' and cas3''. Nucleotide identity between these genes and the *cas3* of DSM 2243 is 99%. (**B**) Sequence alignment reveals an insertion near the start codon of the *cas3''* gene of the 28B strain. The ATG marked in green denotes the start codon of *cas3''* in 28B. The TAA marked in red denotes the stop codon of the *cas3'* gene of 28B. Sanger sequencing was used to verify this insertion. (**C**) Schematic representation of conserved domains in *cas3* reveals that nuclease and helicase domains are divided between *cas3'* and *cas3''* respectively in the strain 28B.

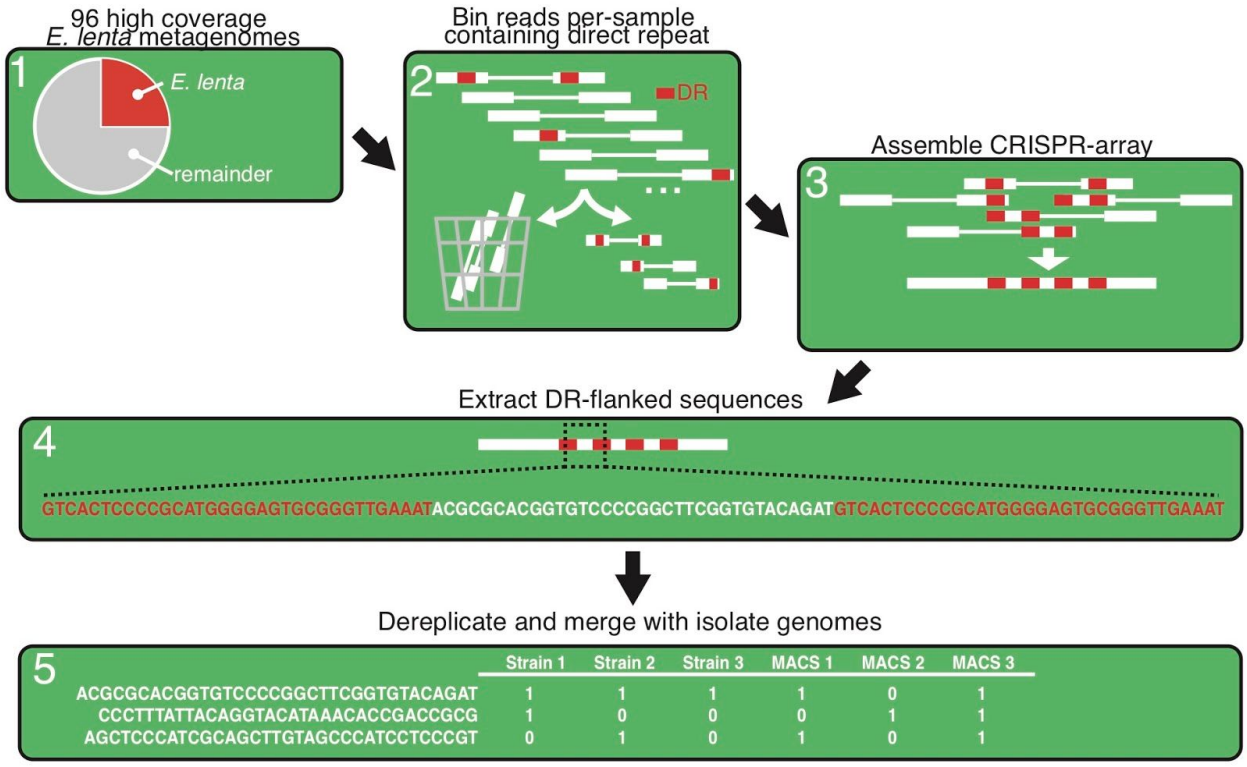

Figure S3. Assembly Strategy for uncovering metagenome-assembled CRISPR spacers relating to Figure 4. (**1**) A set of metagenomes is identified with elevated abundance of organism of interest to increase the probability of array recovery. (**2**) Paired-end reads are screened for at least one instance of the direct repeat (vsearch --usearch_global). (**3**) Resulting binned reads are assembled on a per-sample basis to recover the spacer array. (**4**) Sequences between 25 to 40 nucleotides flanked by an approximation of the direct repeat (≤3 mismatches), are extracted and (**5**) tabulated to examine spacer occurrence across isolates and metagenomes.

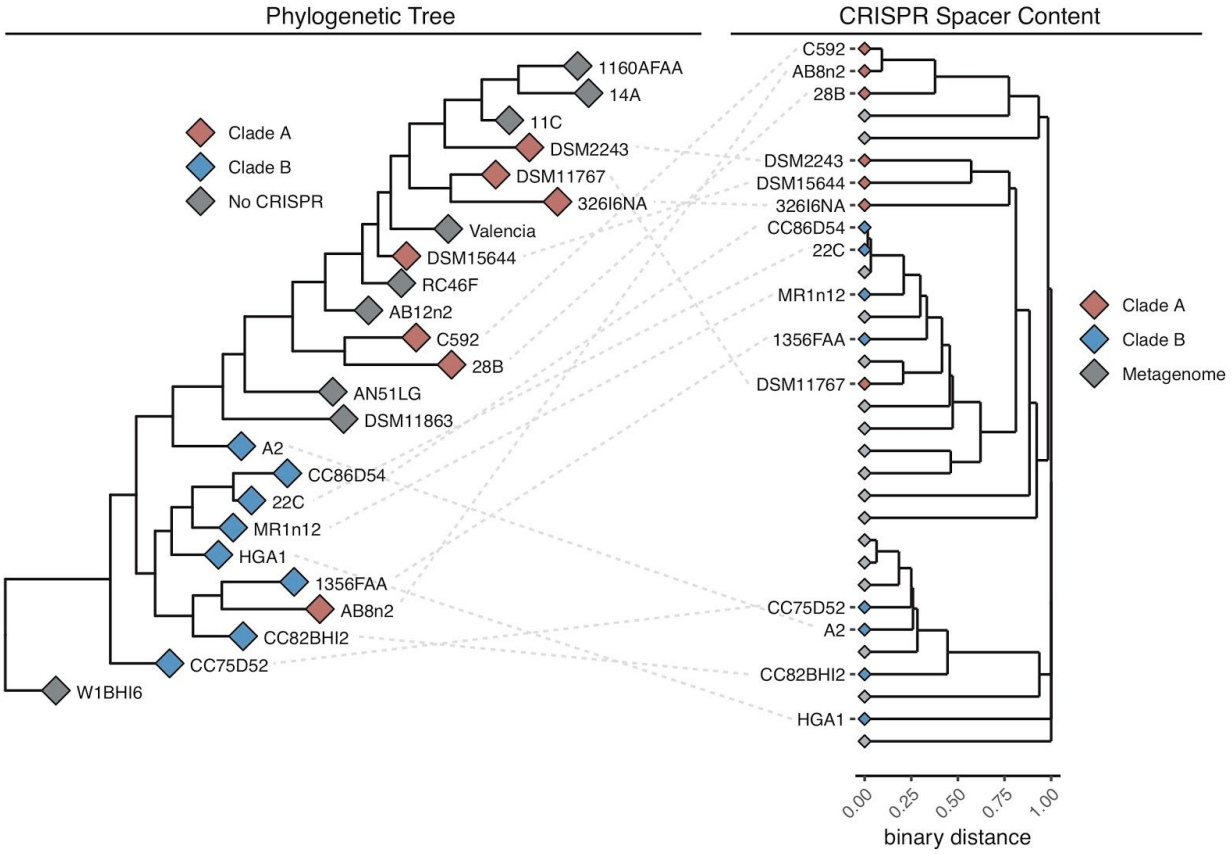

Figure S4. CRISPR spacers correlate with phylogeny related to Figure 4. With the exception of 1 strain for each metric, *cas* gene phylogeny and spacer content are consistent with the whole-genome derived phylogeny.

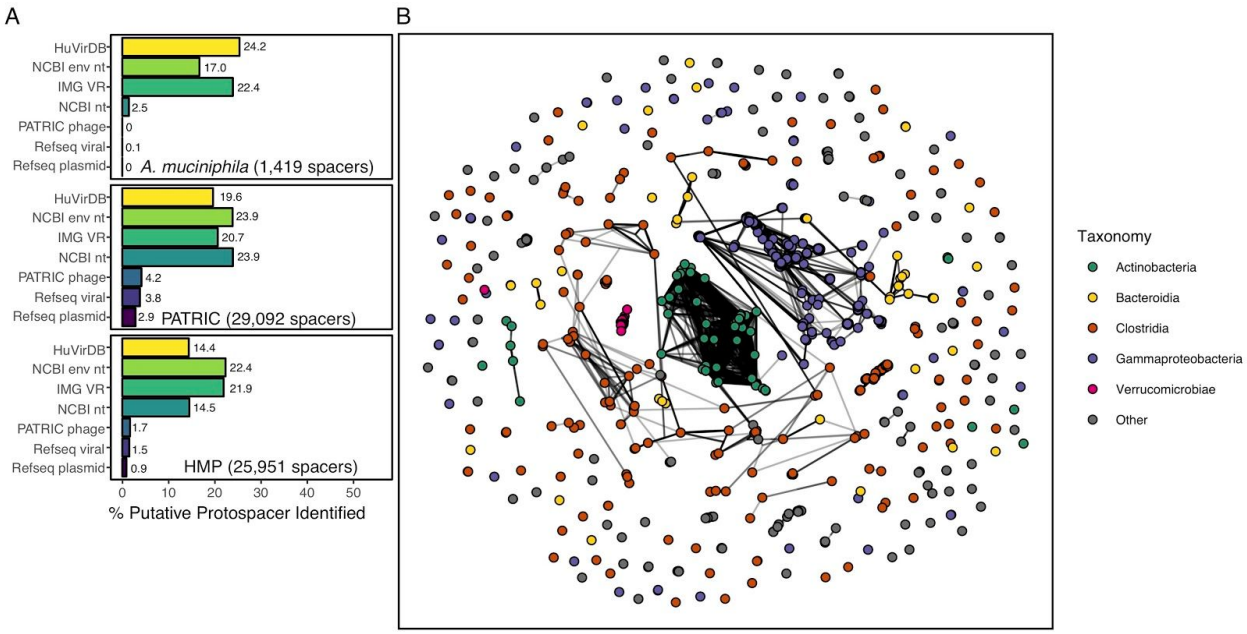

Figure S5. Protospacers across databases relating to Figure 5. (A) Number of identified protospacers as a function of input database demonstrates that highly prevalent, but under-characterized, gut bacteria have improved identification of CRISPR targets in human viromes; for example, *A. muciniphila* (shown here) and *E. lenta* (shown in **Figure 5A**). The numbers in the brackets correspond to the number of spacers extracted from the sum of all genomes in the dataset. **(B)** Network analysis of genomes (linked by shared protospacer targets) reveals that CRISPR targets are conserved within bacterial classes.