RESULTS

Sequencing and annotation of newly sequenced phage genomes

Details of phage isolation and sequencing are described in supplementary file, phate_sequencing.pdf. In brief, *Yersinia* phages LYP215 and LYP264 were isolated from raw sewage from the Livermore, California Water Reclamation Plant. Genomic sequences were generated using MiSeq (Illumina, San Diego) and assembled using CLC Genomics Workbench (Qiagen, Redwood City).

Because the P2 genome was the closest well-annotated homolog to LYP215 and LYP264, we tested gene finding in P2 using PhATE, and compared the calls to those reported for P2 in Christie and Calendar, 2016. PHANOTATE predicted 49 genes corresponding to all of the 44 reported genes (4 with start sites different from all 3 bacterial gene callers), plus five additional gene calls not previously reported. On default settings, the three bacterial gene finders each reported 42 or fewer gene calls (suppl file: phate_P2_CGC.pdf). Because PHANOTATE tends to over- (as opposed to under-) predict genes, and because it performed adequately in calling genes on P2 (missed no reported gene calls), we proceeded with PhATE annotation of the two newly sequenced phage genomes (LYP215, LYP264) by selecting PHANOTATE calls for sequence annotations, and by invoking all currently supported homology searches (suppl. file: phate_LYP264_sequenceAnnotation_main.out.txt). Annotations were assigned based on manual inspection of the summary results for each genome. Genes were assigned putative functions based on closest homologs detected using BLAST+ and HMM searches. Genes with clear homology (>50% identity with >50% coverage) to hits from multiple databases were assigned functional annotations; others were deemed “unknown hypothetical” (suppl. file: phate_genome_annotations.pdf).

Comparison of LYP264 putative holin with related sequences

The putative holin gene of LYP264 (PHANOTATE call #45) did not closely resemble any known holin from other phage included in the current virus-centric databases used here (26-39% identity over >=50% of sequence). To further investigate this sequence, we used Clustal Omega (McWilliam et al., 2013) to align the predicted LYP264 holin gene peptide sequence with its closest pVOG groups (0765, 5078) and its closest blastp hit (WP_087902590.1, *E. coli* “primosomal protein” or “bacteriophage holin family HP1 ” ; suppl. file: phate_LYP264_45_clustalo.pdf). Clearly the LYP264 putative holin gene could be reasonably clustered with either PVOG (annotated in some genes as “holin”), yet its sequence was nearly identical to the *E. coli* protein (93% identity), suggesting that this phage gene and corresponding *E. coli* gene may have common origin, and that the *E. coli* gene is more correctly annotated as a phage holin.

REFERENCES

Christie GE and Calendar R. 2016. Bacteriophage P2. *Bacteriophage* 6:e1145782. doi: 1080/21597081.2016.1145782.

McWilliam H, Li W, et al. 2013. Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 41:W597-600.