Table S1. Geochemical characteristics of samples collected from EPL, I-EPL and BMMRE.

Table S2. Information of published genomes of marine SAR11 (Pelagibacter) infecting phages.

Table S3. The distribution of protein families in the HTVC019Pvirus phage genome.

Table S4. The distribution of protein families in each HTVC010P-related phage genome.

Table S5. Codon usage frequency of phage, bacterial and archaeal genomes reconstructed from EPL and I-EPL samples.

Table S6. The detection of phages similar to those reported in this study in the published Lake Mendota data. The detected ones were sorted by sampling date and colored. Only those with a minimum identity of 90% are shown. These TerL were clustered based on 99% aa similarity using cd-hit before conducting the phylogenetic analyses, the representative sequences were indicated by "#", the ones belonging to same cluster are in the same background color.

Table S7. Coverage and relative abundance of Fonsibacter and phages in Mendota Lake.

Table S8. General information of related TerL sequences obtained from IMG platform.

Fig. S1. The complete genome of *Fonsibacter*_30_26 with a prophage. (a) The location of the prophage on the chromosome of its host. The location of 5S, and 23S/16S rRNA genes and the HVR2 are also shown. (b) The GC skew of the complete genome. The position of prophage is indicated by two purple lines.

Fig. S2. The absence of rpL30 gene in SAR11 genomes. (a) Phylogenetic analyses of the rpL30 identified in SAR11 genomes of AAA795-P11 (in red and bold). The rpL30 were determined based on TIGRFAMs HMM database search using HMMsearch within all published SAR11 genomes. The bootstraps value ≥ 70 is indicated by a black dot. (b) Detailed information of rpL30. The location of rpL30 on chromosome of E.coli (as reference; red circle). The absence of rpL30 in *Fonsibacter*_30_26 is shown in the middle. The scaffold with a rpL30 in AAA795-P11 is shown at the bottom. However, all four genes are most close to those from Marinimicrobia bacterium spp..

Fig. S3. Mapping of paired-end reads from EPL samples to the complete genome of *Fonsibacter*_30_26, using (a) Geneious and (b) bamcov (https://github.com/fbreitwieser/bamcov). The hypervariable and prophage regions in the genome are highlighted. The hypervariable region was documented by mapped paired-end reads from the EPL_05172017_7.5m sample, and similar (pro)phage may be in this sample.

Fig. S4. The hypervariable region of *Fonsibacter*_30_26. (a) Annotation of protein-coding genes in the hypervariable region. The genes were assigned to different functional categories and shown in different colors. The transmembrane domain and signal peptide were predicted and shown if identified, by a star and an open circle, respectively. (b) The four incomplete transketolase proteins in the hypervariable region. The full-length transketolase detected in the genome of *Fonsibacter*_30_26 but outside the hypervariable region is shown for comparison. Two of the transketolase proteins from the hypervariable region only have the N-terminal domain, and the other two have the transket_pyr and C-terminal domains.

Fig. S5. The alignment of phage genomes for (a) uv-Fonsiphage-EPL, HTVC025P and HTVC011P, and (b) HTVC010P and the three complete and one draft HTVC010P-related phage. The alignment was generated using viptree (https://www.genome.jp/viptree/) with default parameters.

Fig. S6. Relative abundance of *Fonsibacter* (accumulated) and infecting phages with complete genomes reconstructed from EPL and I-EPL samples. See Table 1 in the main text for information of complete genomes, the sample from where the complete genome was reconstructed is indicated by a "X". The sequencing coverage of the phage genome is given above the bar. In a given sample, the calculation of relative abundance of *Fonsibacter* was performed by firstly summing the total coverage of all *Fonsibacter* rpS3 genes (referred to as "a"), and also the total coverage of all bacterial and archaeal rpS3 genes (referred to as "b"), then the accumulated relative abundance of *Fonsibacter* in this sample was calculated as "a/b X 100%". The relative abundance of a given *Fonsibacter*-infecting phage was determined by firstly obtaining the coverage of its TerL gene (referred to as "c"), then calculated as "c/b X 100%".

Fig. S7. Related phages in Lake Mendota samples. (a) Mapping of paired-end reads from Lake Mendota samples to genomes of phages reported in this study. The mapped reads have been filtered with a minimum similarity of 98% to the reference genomes. Check Supplementary Table 5 for more details of the samples from Lake Mendota. (b) The reconstruction of a complete phage highly similar to EPL_08022017_1.5m_HTVC010P-related_32_16. Paired-end reads spanning the ends of the scaffolds are shown, and none local assembly error or Ns (gap) was reported by ra2.py, suggesting the 100% completeness of this phage genome.

Fig. S8. Phylogenetic analyses of HNH endonuclease identified in *Fonsibacter* phages. The HNH endonucleases of EPL_32_16 and Lake_Mendota were as queries for NCBI BLASTp search, the hits with >= 80 alignments and >= 50% similarity were retained for analyses, with the one from the prophage included as well. Note that one of the HNH endonucleases was from Candidatus Pelagibacter sp. TMED203, the corresponding scaffold ID is NHJA01000005 at NCBI (https://www.ncbi.nlm.nih.gov/nuccore/NHJA01000005.1/). This scaffold contained many genes encoding hypothetical proteins and had a peptidase M15 gene, which was identified in most HTVC010P-related phages and thought to perform host cell lysis (see above). We suspect this is a misbinning of a phage scaffold into the genome bin of its bacterial host, or this scaffold represented a fragment of prophage in the host genome. We tried to assemble all the scaffolds (46 in total) in the genome bin TMED203 but failed to link NHJA01000005 to any of the others.

Fig. S9. The detection of phages related to the ones reported in this study in global freshwater and marine/saline habitats. This analysis was performed by BLASTp search at IMG metagenomic datasets, with the TerL of phages reported in this study as queries, and only those with a minimum amino acid similarity of 80% were retained for analyses. (a) Box plots show the similarity of TerL detected in IMG datasets to those reported in this study, the numbers of total detections are shown above the boxes. For the outliers, their sampling sites (colored triangles for freshwater, and colored squares for marine/saline) and related phages (indexed number) are shown on the right. (b) Cooccurrence of phage (based on TerL) and *Fonsibacter* (based on rpS3) in the sample. For each sampling site, the total number of samples with phage-*Fonsibacter* cooccurrence were summed and shown. Sampling sites were grouped and colored by continental plates. (c) Samples with phage TerL detected while without *Fonsibacter* in the corresponding sample. Data was prepared as described in (b).

Fig. S10. Phylogenetic analyses showed a single lineage of uv-Fonsiphage-EPL related phages detected in freshwater ecosystems. The tree was built based on the phage large terminase (TerL) proteins. Note that the TerL identified in the hydraulically fracturing related samples in Ohio was almost identical to some of those found in the Lake Erie of Ohio. The *HTVC019Pvirus* groups defined recently are shown. The Pelagiphages in HTVC019Pvirus group III are indicated in blue and bold, uv-Fonsiphage-EPL is indicated in red and bold. Only one related phage TerL was from Africa (Lake Kivu), all other from North America, and it is interesting that the Africa one showed a high similarity to the one identified in Lake Mendota (98.8%), given that these two sites are quite distant from each other.

Fig. S11. Phylogenetic analyses of *Fonsibacter* based on the ribosomal protein S3 (rpS3) nucleotide sequences. All those *Fonsibacter* rpS3 from freshwater-related samples with TerL detected, and also those from marine/saline samples with TerL similarity outliers (blue triangles), were included for analyses. *Fonsibacter* from freshwater-related samples with TerL similarity outliers are indicated by red triangles. The three groups within *Fonsibacter* are shown based on phylogeny. The nodes representing *Fonsibacter* genomes are indicated by stars. We found two lineages only detected in sampling site-specific samples, related information is shown in the left upper corner. The sampling continents are shown in color strips (the outer ring). The rpS3 sequences from marine and brackish SAR11 were used as references, and the two subclades are collapsed. The bootstraps value ≥ 90 is indicated by a black dot.