

Supplementary Figures

Chronic exposure to polycyclic aromatic hydrocarbons is associated with alterations in skin virome composition and an increased virus–host coevolution

Shicong Du,¹ Xinzhao Tong,^{1,2} Marcus H. Y. Leung,¹ Richard J. Betts,³ Anthony C. Woo,⁴ Philippe Bastien,⁴ Namita Misra,⁴ Luc Aguilar,⁴ Cécile Clavaud,⁴ and Patrick K. H. Lee^{5*}

¹School of Energy and Environment, City University of Hong Kong, Kowloon, Hong Kong SAR, China

²Department of Biological Sciences, School of Science, Xi'an Jiaotong-Liverpool University, Suzhou, P. R. China

³L'Oréal Research and Innovation, Singapore

⁴L'Oréal Research and Innovation, Aulnay-Sous-Bois, France

⁵School of Energy and Environment and State Key Laboratory of Marine Pollution, City University of Hong Kong, Hong Kong SAR, China

Correspondence: *B5423, Yeung Kin Man Academic Building, School of Energy and Environment, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China; E-mail: patrick.kh.lee@cityu.edu.hk; Tel: (852) 3442-4625; Fax: (852) 3442-0688.

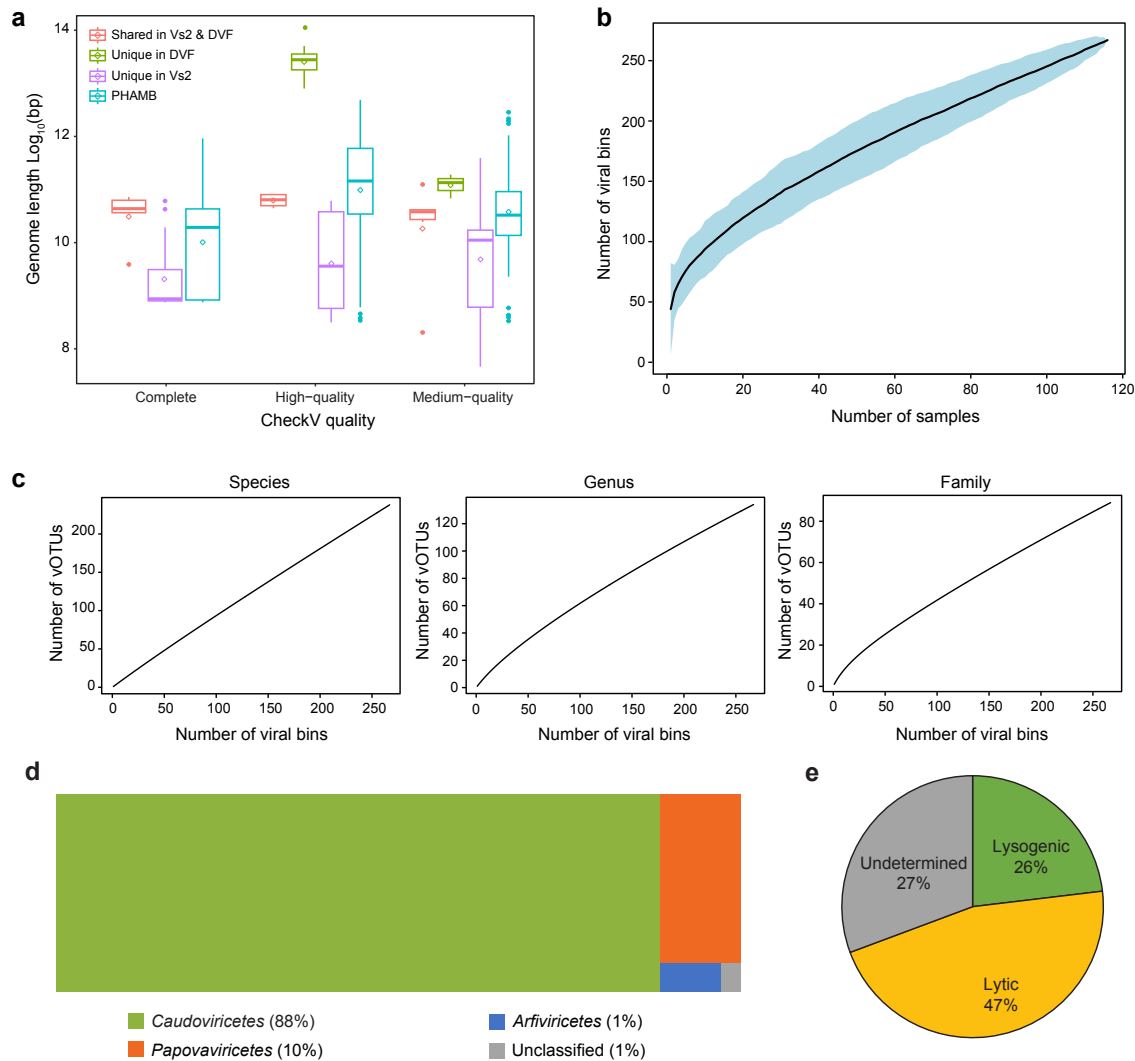


Fig. S1 Overview of the skin viromes. (a) Comparison of the numbers and genome lengths of complete, medium- and high-quality viruses identified by Virsorter2 (Vs2), DeepVirFinder (DVF), and PHAMB. The different colors indicate shared viruses identified by both Vs2 and DVF (red), unique viruses identified by DVF (green) or Vs2 (purple), and viruses identified by PHAMB (blue). The combination of Vs2 and DVF identified 30 complete, 57 medium- and 49 high-quality viruses, whereas PHAMB identified 19 complete, 125 medium- and 120 high-quality viruses. **(b)** Accumulation curves of the viral bins in the 124 cheek skin samples. The confidence intervals derived from the standard deviations are shown in light blue. **(c)** Accumulation curves of the viral bins clustered at the species, genus, and family levels. **(d)** Tree map of the taxonomic distribution of viral operational taxonomic units (vOTUs) in the skin. **(e)** Proportions of viruses with lysogenic lifestyles and lytic lifestyles, respectively, as identified by analysis of the vOTUs in the samples. In each box-and-whisker plot, the box indicates the median, the first quartile, and the third quartile; the whiskers span 1.5 times the interquartile range; and the diamond indicates the mean. Points that lie beyond the whiskers indicate outliers.

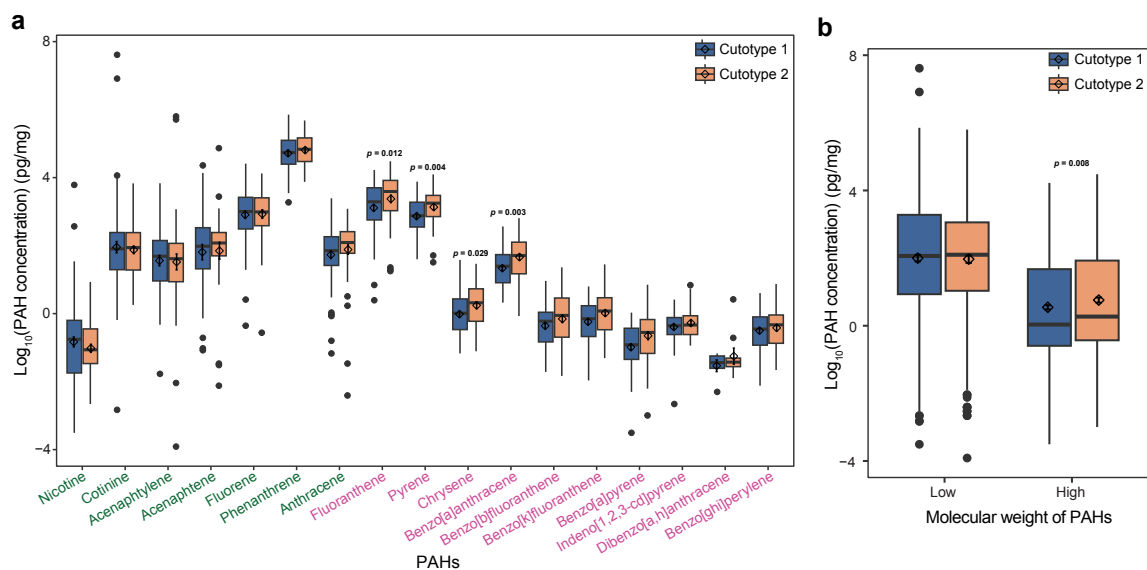


Fig. S2 Measured concentrations of polycyclic aromatic hydrocarbons (PAHs) in hair samples from the participants. (a) The concentrations ($\log_{10}$ scale) of high- and low-molecular-weight PAHs are shown in purple text and green text, respectively, for cutotypes 1 and 2. A molecular weight cut-off of 200 g/mol was used to distinguish the two groups of PAHs. (b) Total average concentrations ($\log_{10}$ scale) of high- and low-molecular-weight PAHs in cutotypes 1 and 2. Only those with $p_{\text{adj}} < 0.05$ (Wilcoxon rank sum test, two-sided) are indicated on the figure. In each box-and-whisker plot, the box indicates the median, the first quartile, and the third quartile; the whiskers span 1.5 times the interquartile range; and the diamond indicates the mean. Points that lie beyond the whiskers are outliers.

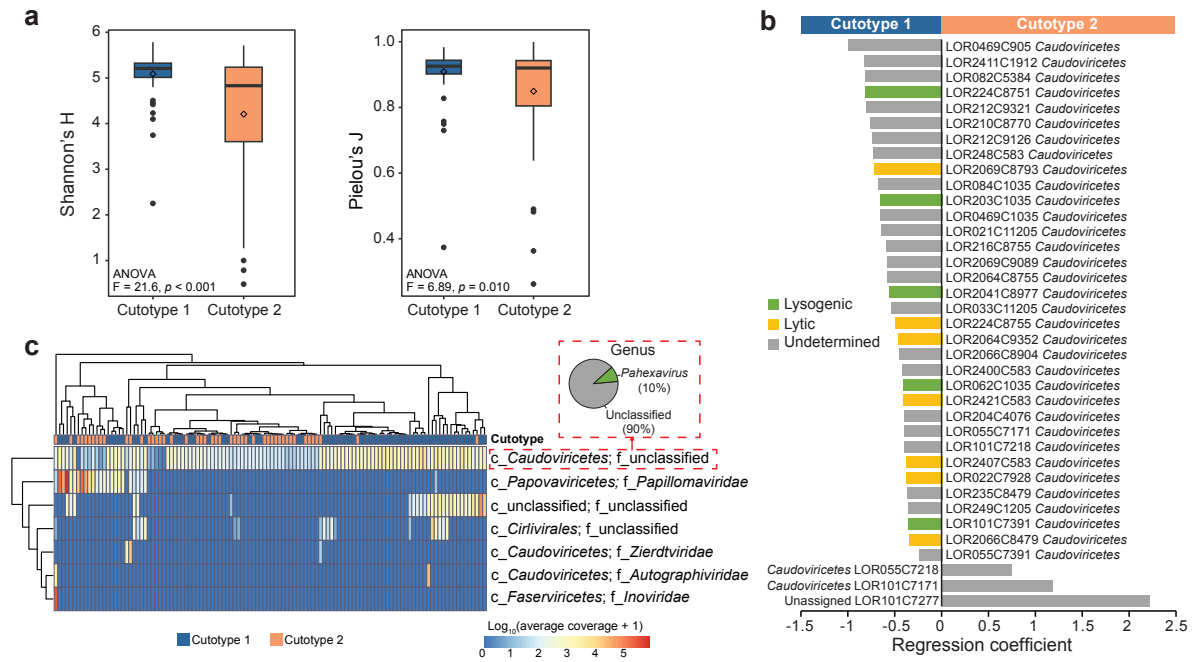


Fig. S3 Diversity and taxonomy of viral operational taxonomic units (vOTUs). (a) Alpha-diversity indexes (i.e., Shannon's H and Pielou's J evenness) of the vOTUs between cutotypes 1 and 2. (b) Enrichment of vOTUs between cutotypes according to MaAsLin2. The viral lifestyle and taxonomy of each vOTU are indicated. (c) Heatmap of the average coverage ($\log_{10}$ scale) of the vOTUs organized by cutotype across the samples. The pie chart shows the proportion of genera within the unclassified family of *Caudoviricetes*. In the label text, the prefixes "c" and "f" indicate the lowest possible assigned taxonomic ranks at the class and family levels, respectively. In each box-and-whisker plot, the box indicates the median, the first quartile, and the third quartile; the whiskers each span 1.5 times the interquartile range; and the diamond indicates the mean. Points beyond the whiskers indicate outliers.

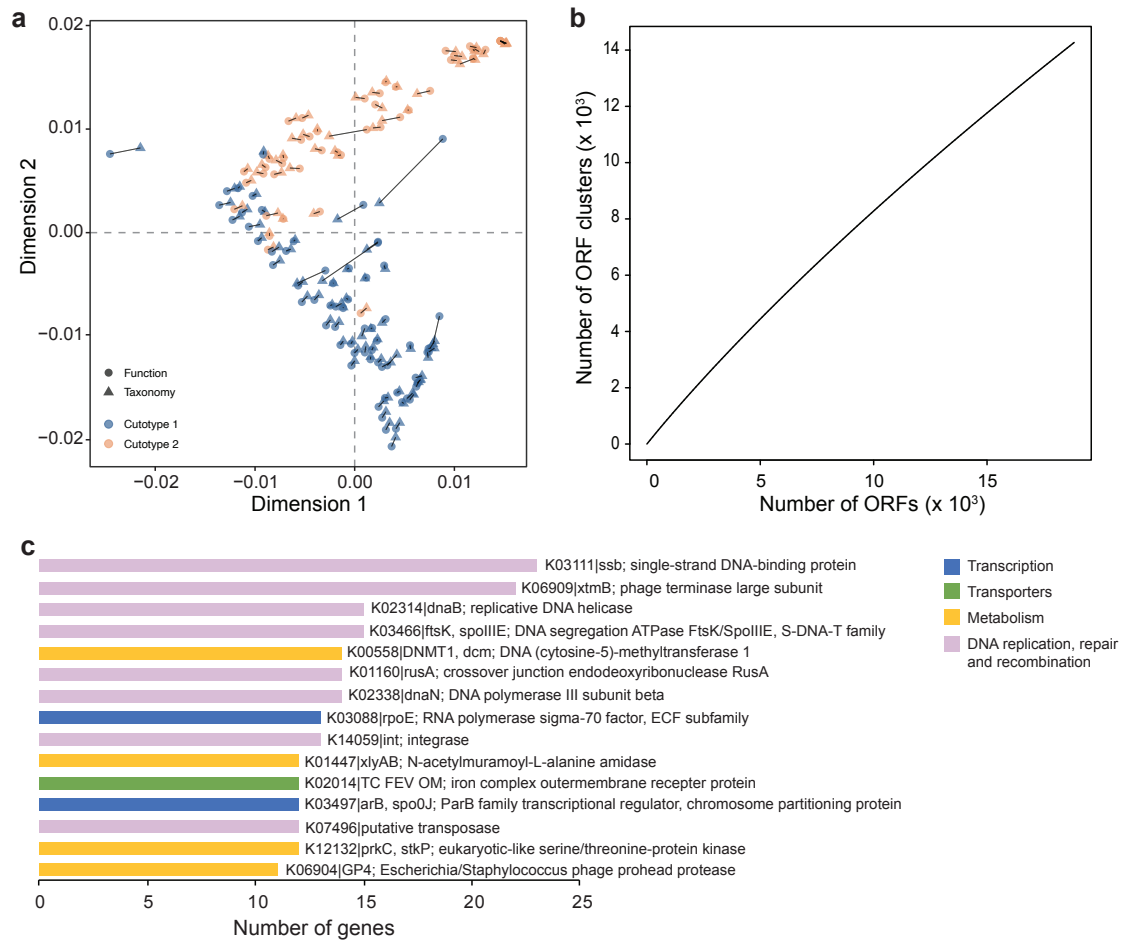


Fig. S4 Functional compositions of viral operational taxonomic units (vOTUs). (a) Procrustes analysis showing the congruence between the taxonomic and functional compositions of the vOTUs. (b) Accumulation curve of the open reading frames (ORFs) of vOTUs. (c) Functional annotation of the ORF clusters based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) (only clusters with > 10 genes are shown).

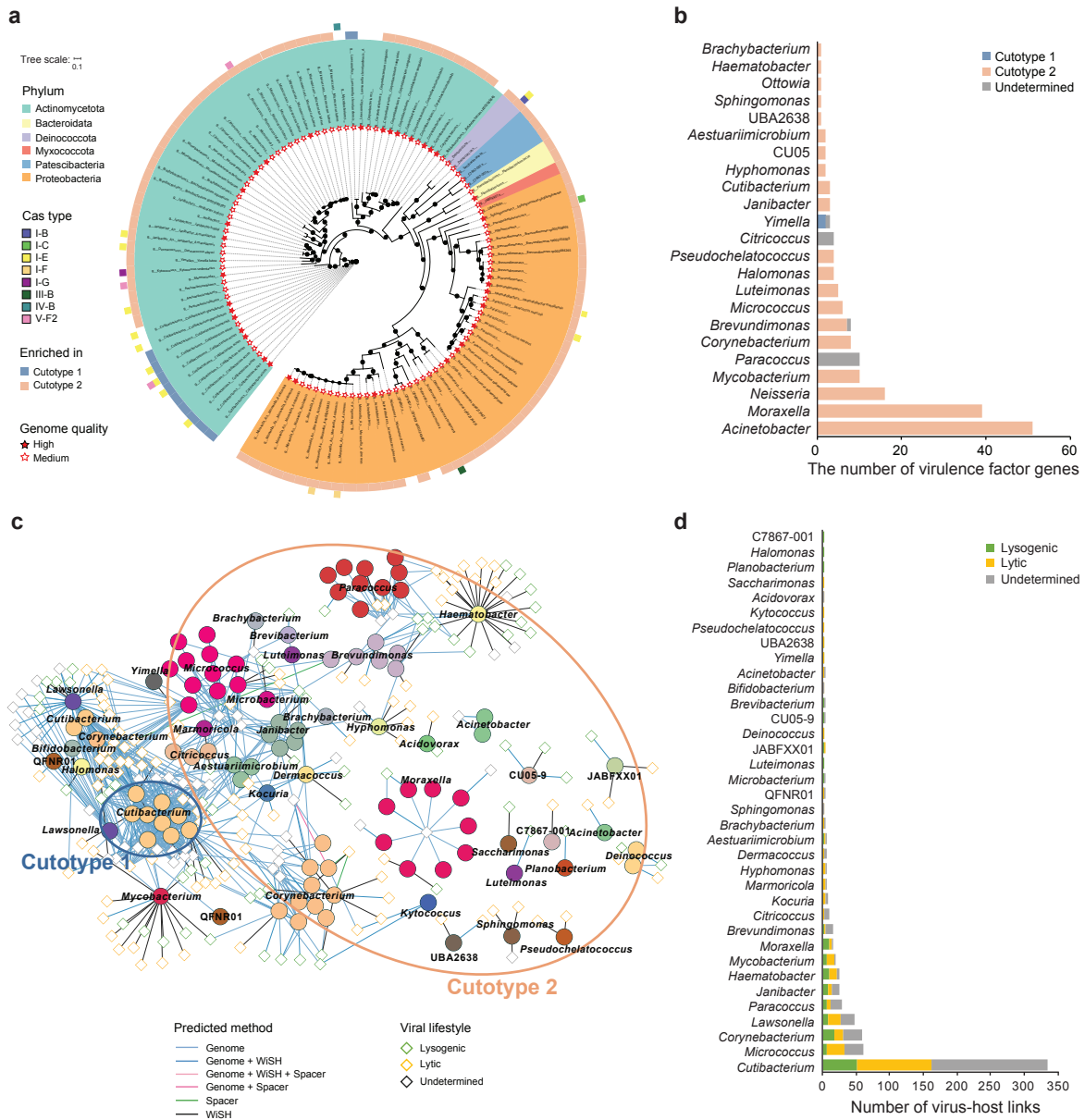


Fig. S5 Representative metagenome-assembled genomes (rMAGs) reconstructed from 124 cheek skin samples. (a) Phylogenetic tree depicting 112 rMAGs. In the labels, the prefix “s” indicates a known species, while the prefixes “g” and “f” indicate the lowest possible assigned taxonomic ranks at the genus and family levels, respectively. rMAGs that could not be assigned to a known species are marked with a red star in the inner ring. Enrichment of rMAGs between cutotypes, based on MaAsLin2 analysis, is shown in the second outer ring. The type of clustered regularly interspaced short palindromic repeat-associated protein identified in each rMAG is indicated in the outermost ring. **(b)** Number of virulence factor genes identified in the rMAGs at the genus level. **(c)** Network illustrating viruses and their predicted bacterial hosts at the genus level. Circles and diamonds represent predicted bacterial and viral hosts, respectively, and the edges of shapes are color-coded based on the prediction method used. **(d)** The frequency of virus–host links in the samples at the genus level of hosts.

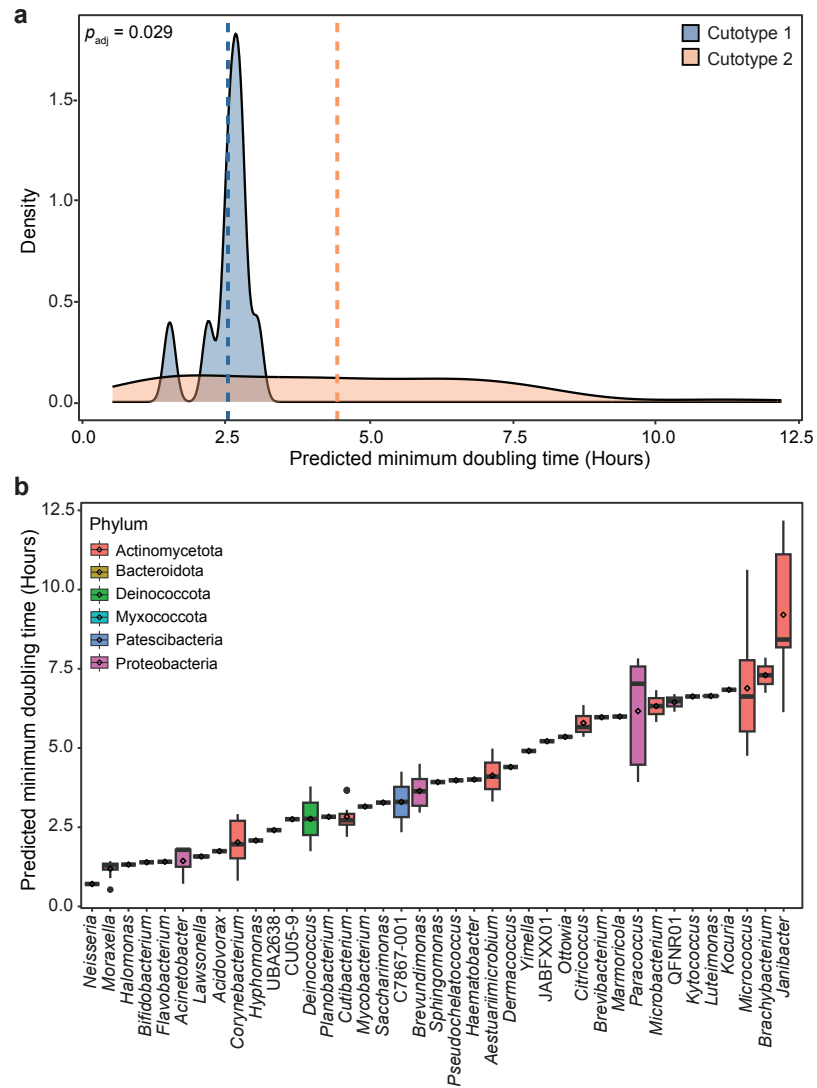


Fig. S6 Estimated maximum growth rate of the representative metagenome-assembled genomes according to (a) cutotype and (b) genus. The difference between cutotypes 1 and 2 was statistically significant ($p_{\text{adj}} < 0.05$, Wilcoxon rank sum test, two-sided). In each box-and-whisker plot, the box indicates the median, the first quartile, and the third quartile; the whiskers each span 1.5 times the interquartile range; and the diamonds indicate the mean value. Points that lie beyond the whiskers indicate outliers.

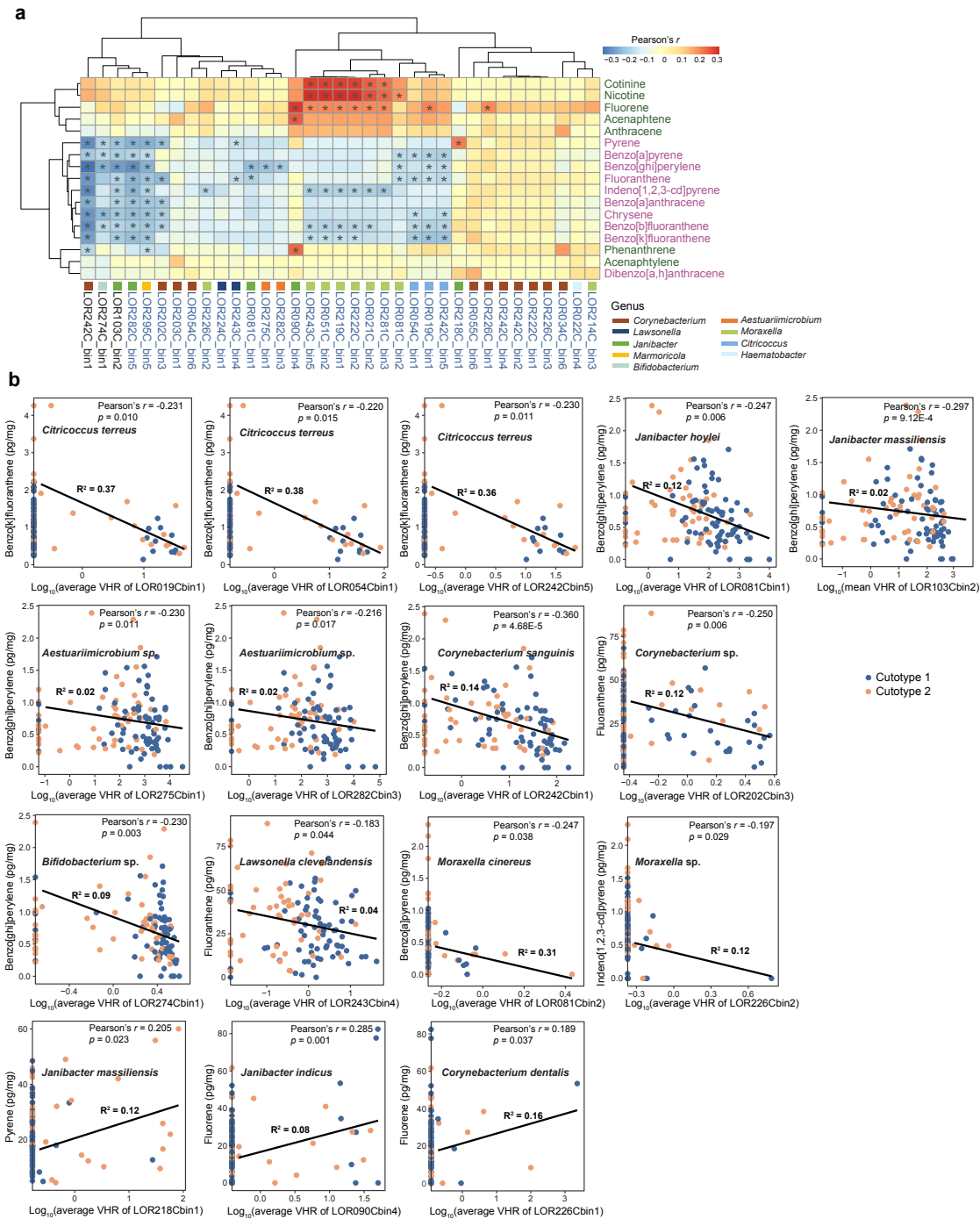
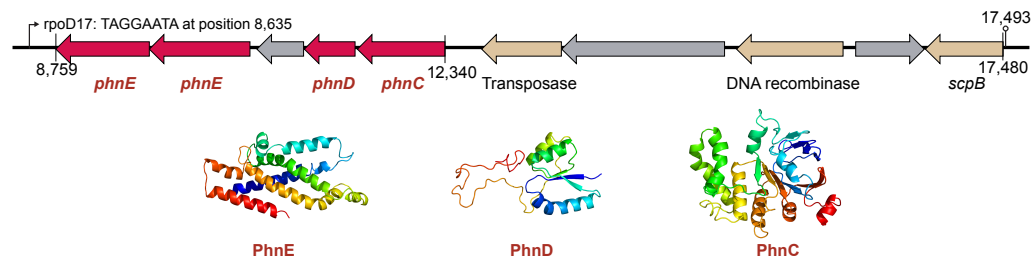


Fig. S7 Relationships between the concentrations of polycyclic aromatic hydrocarbons (PAHs) and virus-to-host abundance ratios (VHRs). (a) Heatmap showing Pearson's correlations between the concentrations of PAHs and selected representative metagenome-assembled genomes (rMAGs) (i.e., genera whose VHRs differed significantly between cutotypes). High- and low-molecular-weight PAHs are represented by purple- and green-colored text, respectively. The rMAGs enriched in cutotype 2 are colored dark blue. An asterisk (*) denotes $p < 0.05$, based on the corresponding Pearson's coefficient. (b) Largest Pearson's correlation between the average VHR of the rMAGs shown in panel (a) and the measured concentrations of PAHs (excluding nicotine and cotinine). The results of LOR295Cbin.5 and LOR282Cbin.5 are shown in Fig. 3e.

S4C6930_fragment (8.7 kb / 152.6 kb)



S14C924_fragment (8.6 kb / 39.64 kb)

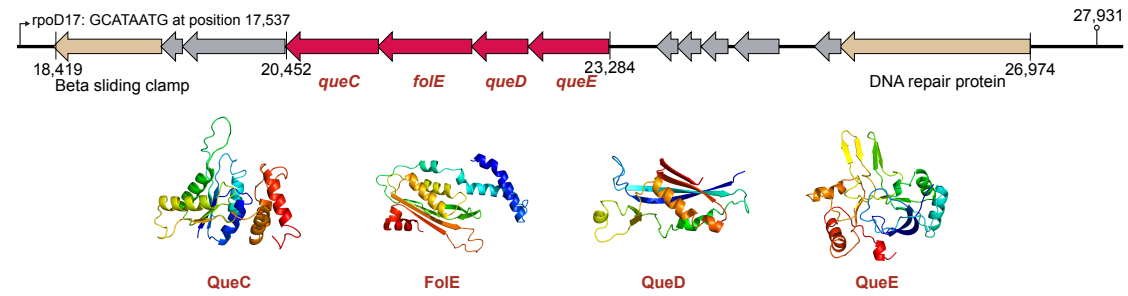


Fig. S8 Genomic context and protein structure of the seven auxiliary metabolic genes carried by two bacteriophages in cutotype 2. An arrow and a hollow circle indicate the promoter and terminator, respectively. The structures are color-coded from the N- to the C-terminus.