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

Shicong Du
City University of Hong Kong
Xinzhao Tong
Xi'an Jiaotong-Liverpool University
Marcus H. Y. Leung
City University of Hong Kong
Richard J. Betts
L'Oréal Research and Innovation
Anthony C. Woo
L'Oréal Research and Innovation
Philippe Bastien
L'Oréal Research and Innovation
Namita Misra
L'Oréal Research and Innovation
Luc Aguilar
L'Oréal Research and Innovation
Cécile Clavaud
L'Oréal Research and Innovation
Patrick K. H. Lee
patrick.kh.lee@cityu.edu.hk

City University of Hong Kong

Research Article

Keywords: skin, viromes, virus–host interactions, polycyclic aromatic hydrocarbons, auxiliary metabolic genes, viral lifestyle

Posted Date: April 8th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4205907/v1

Additional Declarations: Competing interest reported. The authors RJB, ACW, PB, NM, LA, and CC are employees of L'Oréal Research and Innovation.

Chronic exposure to polycyclic aromatic hydrocarbons is associated with alterations in skin virome composition and 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]*

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

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

[3]L'Oréal Research and Innovation, Singapore

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

[5]School 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.

Abstract

Background: Viruses play an essential role in the human skin microbiome under healthy and diseased conditions. We previously reported that exposure to polycyclic aromatic hydrocarbons (PAHs) led to the differentiation of bacteria samples from the cheeks of 124 Chinese women into two cutotypes with distinct taxonomic and functional compositions. In addition, we found that women with a high level of PAH exposure exhibited skin dryness and hyperpigmentation. However, the characteristics of viruses associated with these cutotypes are not well understood. In this study, we employed bulk metagenomic sequencing to investigate the diversity and functions of skin viromes and the virus–host interactions of the two cutotypes.

Results: The viruses associated with the two cutotypes had distinct diversities, compositions, functions, and lifestyles. Moreover, virus–host interactions were substantially increased by high levels of PAH exposure, especially when the hosts were potential pathogens. Under high-PAH exposure conditions, the viruses were enriched in xenobiotic degradation functions, and there was evidence suggesting that the insertion of bacteriophage-encoded auxiliary metabolic genes into hosts aids biodegradation. Furthermore, under low-PAH exposure conditions, virus–host interactions followed the "Piggyback-the-Winner" model, with *Cutibacterium* strains being "winners," whereas under high-PAH exposure conditions, the interactions followed the "Piggyback-the-Persistent" model, with the biodegradation bacteria being "persistent."

Conclusion: Our findings demonstrate that PAHs not only influence skin bacteria but also impact the composition of skin viruses and their interactions with skin bacteria, which may influence skin health. A detailed understanding of the roles played by viruses on skin exposed to air pollutants would aid the development of strategies to harness the potential of viruses to prevent undesirable skin conditions.

Keywords: skin, viromes, virus–host interactions, polycyclic aromatic hydrocarbons, auxiliary metabolic genes, viral lifestyle

Introduction

The skin is a physical barrier between the body and the environment and is inhabited by a microbial consortium consisting of bacteria[1,2], fungi[1,2,3], and viruses, especially bacteriophages[2,4,5]. This consortium plays an indispensable role in modulating immune processes that are essential for maintaining skin health[6], and its composition varies depending on external environmental factors (e.g., season and pollution) and host physiological factors (e.g., age and genetics)[6,7,8].

Air pollution appears to play a significant role in various skin diseases, particularly in people living in urban environments[9]. For example, polycyclic aromatic hydrocarbons (PAHs) are ubiquitous air pollutants that substantially affect the skin[10]. PAH exposure has been implicated in the development of inflammatory skin diseases and cutaneous malignancies[11,12], disorders in lipid metabolism in the stratum corneum and in skin protein composition[12], and skin cancer[13]. PAH exposure can also alter the diversity, taxonomic composition, and metabolic profiles of skin bacterial communities[14,15], consequently affecting skin phenotype[15]. For example, high levels of PAH exposure have been linked to skin dryness and hyperpigmentation[15] and to a skin microbiome enriched in bacterial species capable of degrading xenobiotics[16]. Recent studies on skin metagenomes have revealed the composition and functional potentials of the skin microbiome[2,8,15,17] and the microbial taxonomy, functionality, and antimicrobial resistance profiles of skin-identified microbial sub-community clusters, denoted "cutotypes."[17] Cutotype studies have clarified specific ecosystemic interactions between bacteria, their functions, and their human hosts[15,17].

The interactions between microbes and the skin's immune systems are diverse and range from mutualism to pathogenicity[18]. Given that bacteriophages can modify the abundance and metabolism of their hosts, it is crucial to determine how bacteriophage–host interactions can impact healthy and inflammatory skin conditions[19]. For example, in patients with acne, there

is a reduced abundance of *Propionibacterium acnes* bacteriophages and their hosts in hair follicles, and a positive correlation was found between the abundance of *P. acnes* bacteriophages and the severity of acne, which suggests that *P. acnes* bacteriophages play a significant role in maintaining skin health[20]. Moreover, *Corynebacterium* and *Staphylococcus* bacteriophages were identified as key organisms affecting virus–host dynamics in healthy skin[4]. When exposed to environmental stresses, bacteriophages can alter their lifestyles and increase their replication and survival by adjusting their strategies for infecting bacteria, leading to increased spread of bacteriophage progeny[21,22]. However, the characteristics of viruses associated with cutotypes and the interactions between viruses and bacterial hosts on skin exposed to PAHs remain largely unknown.

We previously demonstrated that skin bacteria from the cheeks of 124 Chinese women residing in two cities in China and with various PAH exposure levels could be differentiated into two distinct cutotypes based on PAH exposure levels[15]. In the present study, we used the same metagenomic samples to investigate the viruses associated with these two cutotypes in terms of diversity, composition, function, and lifestyle. In addition, we studied the effects of PAH exposure on the interactions between the viruses and their hosts to understand the interplay between the viruses, pollution, and bacterial metabolism modification due to pollution exposure. Moreover, we examined whether virus–host relationships followed the "Piggyback-the-Winner" (PtW) or "Piggyback-the-Persistent" (PtP)[23] ecological models under low and high levels of PAH exposure, respectively. Additionally, we identified the insertion of bacteriophage-encoded auxiliary metabolic genes (AMGs) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems to assess the potential for virus–host coevolution. Overall, our findings revealed that there was strong coupling between the viruses and their bacterial hosts, highlighting that viruses assist skin microbial metabolism through progeny spread and gene insertion under high-PAH exposure conditions.

Results

High levels of PAH exposure altered skin virome compositions

Examination of the 124 cheek skin samples enabled the recovery of 11,781 putative viral bins of > 5 kb (**Supplementary Table 1**). After quality filtering, 264 viral bins with > 50% completeness were obtained, comprising medium-quality (50%–90% completeness; n = 125), high-quality (> 90% completeness; n = 120), and complete (n = 19) bins (**Supplementary Table 1**). Additionally, 30 complete, 57 medium-, and 49 high-quality single viral contigs were identified, comprising only about half of the number of viral bins identified (**Fig. S1a**). However, rarefaction analysis indicated that the 264 viral bins did not fully represent the skin viromes (**Fig. S1b**). Subsequently, clustering of the 264 viral bins revealed 238 species-level viral operational taxonomic units (vOTUs), 88 of which were in at least two of the samples. The viral bins were also clustered into 134 and 89 genus- and family-level vOTUs, respectively. The rarefaction curves of the vOTUs were unsaturated at all taxonomic ranks, suggesting that additional samples would be required to capture skin virome diversity (**Fig. S1c**). Most of the vOTUs (88%) were unclassified members of the class *Caudoviricetes* (tailed bacteriophages), demonstrating that bacteriophages comprise a significant proportion of skin viruses[4,5] (**Fig. S1d**). Moreover, most of the viruses were unclassifiable at lower levels, similar to the viromes that have been found at other body sites[2,4]. Additionally, bacteriophages with a lytic lifestyle were predominant, comprising 47% of the virome (**Fig. S1e**).

Previously, metagenomic analysis of the 124 samples revealed that the skin bacteria differentiated into cutotypes 1 and 2, which differed substantially in terms of diversity, and taxonomic and functional compositions[15]. Cutotype 2 was associated with high levels of PAH exposure, featured high bacterial diversity and a low abundance of *Cutibacterium acnes*, and was found in those with a dry and dull skin phenotype, whereas cutotype 1 was associated with low levels of PAH exposure, featured a low bacterial diversity and a high abundance of *C.*

122 *acnes*, and was found in those with a normal skin phenotype[15] (**Fig. S2a**). The average
123 concentration of the 10 high-molecular-weight PAHs (i.e., > 200 g/mol) was significantly
124 greater in the hair of those with cutotype 2 (55.9 ± 28.5 pg/mg hair) than of those with cutotype
125 1 (75.2 ± 28.4 pg/mg hair) (Wilcoxon rank sum test [WRST], two-sided, $p_{adj} = 0.008$), whereas
126 there was no difference in the average concentration of the seven low-molecular-weight PAHs
127 (i.e., < 200 g/mol) between the cutotypes (WRST, two-sided, $p_{adj} = 0.48$) (**Fig. S2b**). Canonical
128 correspondence analysis showed that the concentrations of most PAHs were significantly
129 associated with vOTUs (**Supplementary Table 2**). Partial least squares path modeling of the
130 relationships between the bacteria, viruses, and PAHs showed that PAHs were significantly and
131 positively associated with both bacteria (0.235) and viruses (0.444) (direct (0.245) + indirect
132 (0.199) path coefficients) (**Fig. 1a**). The indirect path coefficient between PAHs and viruses,
133 mediated by bacteria, was 0.199 (direct path coefficients between PAHs and bacteria (0.235) ×
134 those between bacteria and viruses (0.849)) (**Fig. 1a**). This is similar in magnitude to the direct
135 path coefficient (0.245), indicating the crucial role played by bacteria in the relationship
136 between PAHs and viruses.

137 Given that bacteria and viruses may interact when exposed to PAHs, the viruses associated
138 with the two cutotypes were further explored. Principal coordinate analysis confirmed that the
139 viruses could be separated into two distinct clusters that corresponded with the bacterial
140 cutotypes (permutational multivariate analysis of variance (PERMANOVA), $R^2 = 0.222$, $p <$
141 0.001; **Fig. 1b**). This was supported by the result of the permutation test for homogeneity of
142 multivariate dispersions ($F = 32.22$, $p < 0.001$; **Fig. 1c**). The viruses associated with cutotype
143 1 had a significantly higher Shannon diversity (analysis of variance [ANOVA], $p < 0.001$; **Fig.**
144 **S3a**) than those associated with cutotype 2. Similar numbers of viral bins were identified from
145 samples of cutotype 1 ($n = 130$) and cutotype 2 ($n = 134$), but a substantially greater number
146 of vOTUs were enriched in cutotype 1 ($n = 34$) than in cutotype 2 ($n = 3$) (WRST, two-sided,

147 p_{adj} < 0.05) (**Fig. S3b**). The genus *Pahexavirus* was marginally significantly higher in
148 abundance in cutotype 1 than in cutotype 2 (WRST, two-sided, p_{adj} = 0.055; **Fig. S3c**), whereas
149 the family *Zierdtviridae* was enriched in cutotype 2 (WRST, two-sided, p_{adj} = 0.038; **Fig. S3c**).
150 The virus species evenness varied significantly between the two cutotypes (ANOVA, $p < 0.05$),
151 but the average evenness of both cutotypes was > 0.9 (**Fig. S3a**), suggesting that no specific
152 viral populations were dominant. The abundances of lysogenic and lytic lifestyles, respectively,
153 were marginally different (WRST, two-sided, p_{adj} = 0.058) between the two cutotypes (**Fig.**
154 **1d**).
155
156 **High levels of PAH exposure decreased virome functional potentials**
157 The functional potential of viruses was examined to determine how differences in
158 taxonomic composition between cutotypes caused variations in their functional profiles.
159 Similar to skin bacteria[15], the taxonomy and functions of the viral samples were congruent in
160 cutotypes 1 and 2 (Pearson's r = 0.991, p = 0.001; **Fig. S4a**), suggesting that samples with
161 similar taxonomic compositions tended to have similar functions. Compared with factors such
162 as skin condition or geographic location, cutotype better accounted for taxonomic and
163 functional differences between viromes (**Supplementary Table 3**). Analysis of the vOTUs
164 identified 18,842 open reading frames (ORFs), which were clustered into 14,269 de novo ORF
165 clusters. Their accumulation curves were unsaturated, indicating that the viromes contained
166 highly diverse functions (**Fig. S4b**). The ORF clusters with > 10 genes mostly encoded proteins
167 with DNA replication, repair, and recombination functions (**Fig. S4c**).
168 Multivariate analysis of the Kyoto Encyclopedia of Genes and Genomes (KEGG)
169 Orthology (KO) gene families of viruses in cutotypes 1 and 2 revealed differential enrichment
170 of 755 KOs (675 in cutotype 1 and 80 in cutotype 2; **Supplementary Table 4**) and 56 KEGG
171 pathways (41 in cutotype 1 and 15 in cutotype 2; **Fig. 2a** and **Supplementary Table 4**). At the

pathway level, cutotype 1 was enriched in diverse metabolic functions associated with the biosynthesis of secondary metabolites (KO00521, KO00401, and KO00960), glycans (KO00541, KO00571, and KO00572), amino acids (KO00220 and KO00290), and cofactors and vitamins (KO00130, KO00790, and KO00770); cutotype 2 was enriched in fewer metabolic functions, and these were mostly associated with the degradation of amino acids (KO00280 and KO00310) and xenobiotics (KO00930, KO01220, and KO00362) (**Fig. 2a**). The arginine pathway was previously reported to be limited in the microbiome of cutotype 2[15]. Similarly, in this study, there was differential enrichment of the arginine biosynthesis pathway in cutotypes 1 and 2 (**Supplementary Table 4**). Specifically, three genes (i.e., *otc*, *argG* and *argH*) that encode proteins in the arginine biosynthesis pathway (M00844), and four genes (i.e., *argC*, *argJ*, *argB*, and *argD*) that encode proteins in the ornithine biosynthesis pathway (M00028), which produces the precursor of arginine biosynthesis, were found in one virus (LOR101C7391) in cutotype 2 (**Fig. 2b**). The repressor of the *argCJBD* operon (i.e., *argR*)[24] and the predicted promoter responsible for regulating transcription were also identified in LOR101C7391 (**Fig. 2b**), suggesting that it has a functional arginine biosynthesis operon.

As high-level PAH exposure enriches xenobiotic degradation functions in skin bacteria[15], we hypothesized that this also occurs in cutotype 2. To test this hypothesis, 25 KOs associated with the biodegradation and metabolism of aromatic xenobiotics were compared between cutotypes 1 and 2 (**Fig. 2c**). There was a greater abundance of KOs for the degradation of aminobenzoate, benzoate, and caprolactam (K01692; WRST, two-sided, p_{adj} = 0.0037), chloroalkane/alkene and naphthalene (K00121; WRST, two-sided p_{adj} = 0.0404), nitrotoluene (K10680; WRST, two-sided, p_{adj} = 0.0404), and atrazine (K01429; WRST, two-sided, p_{adj} = 0.022) in cutotype 2 viruses than in cutotype 1 viruses (Monte Carlo permutation test, n = 10,000) (**Fig. 2c**). This indicates that the viruses in cutotype 2 may confer adaptive advantages for xenobiotic biodegradation on their bacterial hosts when exposed to high concentrations of PAHs.

Numbers of virus–host interactions were increased under high-PAH exposure

To elucidate the responses of virus–host interactions under PAH exposure, the putative links between viruses and their bacterial hosts in all the cheek skin samples were examined. The *in-situ* bacterial hosts were represented by representative metagenome-assembled genomes (rMAGs) reconstructed from the samples, i.e., 77 medium-quality and 35 high-quality rMAGs (**Fig. S5a** and **Supplementary Table 5**). The predominant rMAGs were affiliated with the phyla Actinomycetota (n = 58) and Proteobacteria (n = 46). Only 10 rMAGs were enriched in cutotype 1, with *C. acnes* (n = 7) predominating, while 95 were enriched in cutotype 2, with *Micrococcus luteus* (n = 9) predominating (**Fig. S5a** and **Supplementary Table 5**). This result is consistent with the higher bacterial diversity and lower *C. acnes* abundance observed in samples from cutotype 2 versus cutotype 1, based on metagenomic short reads[15]. The average maximum growth rate of rMAGs in cutotype 1 was significantly higher than that of rMAGs in cutotype 2 (WRST, two-sided, p_{adj} = 0.029) (**Fig. S6a**). This was likely due to the dominant rMAGs in each cutotype, as *Cutibacterium* had a faster average predicted maximum growth rate (2.8 ± 0.4 hours) than *Micrococcus* (6.9 ± 1.6 hours) (WRST, two-sided, $p_{adj} = 1.191 \times 10^{-5}$) (**Fig. S6b**). Given that virulence within the skin microbiome can significantly impact the development and manifestation of various skin phenotypes[6], the global virulence in the rMAGs was examined. Virulence factor genes (VFGs) were found in 64 rMAGs, 62 of which belonged to those enriched in cutotype 2 and only two of which belonged to those enriched in cutotype 1 (each carrying only a single VFG) (**Supplementary Table 5**). The rMAGs with the highest number of VFGs (predominantly associated with the genus *Acinetobacter*) were found to be enriched in cutotype 2 (**Fig. S5b**). Almost all VFGs identified (97.7%) were found in human pathogens (e.g., *Acinetobacter baumannii* ACICU) (**Supplementary Table 5**).

Of the vOTUs analyzed, 88% (211/238) were linked to 91% (102/112) of the reconstructed rMAGs, with a significant number of links to *Cutibacterium* (n = 334), *Micrococcus* (n = 60),

and *Corynebacterium* (*n* = 59) genera (**Fig. S5c, S5d**, and **Supplementary Table 6**). At the rMAG level, 723 virus–host links were identified, with a similar number of links observed in rMAGs enriched in cutotypes 1 (*n* = 303) and 2 (*n* = 312), respectively. The complexity of virus–host links was reduced by analysis at the host genus level: this identified 375 virus–host links in the two cutotypes, with more pairings predicted in cutotype 2 (*n* = 204) than in cutotypes 1 (*n* = 84) (**Supplementary Table 6**). Based on the number of links identified in each host genus, *Cutibacterium* (*n* = 94) was most susceptible to viral infection, followed by *Lawsonella* (*n* = 47), *Corynebacterium* (*n* = 32), and *Micrococcus* (*n* = 31) (**Supplementary Table 6**). The only exact CRISPR spacer-match was between a temperate virus (S19C11122) and the human pathogen *Acidovorax temperans* (LOR022Cbin2) in a sample from cutotype 2, representing strong evidence of viral infection in cutotype 2. Nevertheless, this pathogen could be infected by another virus (S4C2191) with a lytic lifestyle (**Supplementary Table 6**).

To clarify the ecological role of virus–host interactions in the two cutotypes, the lineage-specific correlations between viruses and their host abundances and the virus-to-host abundance ratios (VHRs) were analyzed. Several overabundant (WRST, two-sided, $p_{adj} < 0.05$) viruses in cutotype 1 or 2 were linked to hosts such as *Cutibacterium*, *Aestuariimicrobium*, *Bifidobacterium,* and *Marmoricola* with a high abundance and prevalence (**Fig. 3a**). *Cutibacterium* and their linked viruses were enriched in cutotype 1, whereas *Micrococcus* and their linked viruses were enriched in cutotype 2 (**Fig. 3a**). The abundances of *C. acnes* strains and other *Cutibacterium* species and *M. luteus* strains and other *Micrococcus* species were positively correlated with their respective linked viruses in both cutotypes (**Fig. 3b** and **Supplementary Table 6**). The slopes of the linear regressions between the abundances of the eight *C.* acnes strains and their linked viruses were highly similar in cutotype 1 (0.963 ± 0.013) and cutotype 2 (0.961 ± 0.023), and significantly greater than the slopes for other *Cutibacterium* species (WRST, two-sided, $p_{adj} = 0.012$) (**Fig. 3b**). In contrast, the slopes of the

linear regressions between the abundances of the nine *M. luteus* strains and their linked viruses showed strain-specific differences in cutotype 1 (0.664 ± 0.437) and cutotype 2 (0.720 ± 0.335) (**Fig. 3b**). In both cutotypes, the average VHRs for most genera were > 1, indicating that the viruses tended to utilize lytic infection to target and lyse a majority of the bacteria. However, the average VHRs for a few genera, including those belonging to *Cutibacterium* and *Micrococcus*, were close to 1 (**Fig. 3c**). Most of the genera in cutotype 2, including some with biodegradation ability (e.g., *Janibacter*[25] and *Marmoricola*[26]) had significantly lower VHRs than those in cutotype 1 (WRST, two-sided, $p_{adj} < 0.05$) (**Fig. 3c**), but the correlations between the abundances of viruses and their hosts sometimes differed between cutotypes. For example, the abundance of a *Marmoricola* species (LOR295Cbin5) was significantly and negatively correlated with its linked viruses in cutotype 1 (Pearson's $r = -0.19$, $p = 0.027$), whereas it was positively correlated with its linked viruses in cutotype 2, albeit nonsignificantly (Pearson's $r = 0.059$, $p = 0.57$) (**Fig. 3d**).

Next, the influences of different PAHs on VHRs among species and strains were investigated. Low- and high-molecular-weight PAHs exerted differential effects on 24 rMAGs (**Fig. S7a**). The concentrations of many of the high-molecular-weight PAHs (e.g., pyrene, benzo[*a*]anthracene and benzo[*ghi*]perylene) were significantly and negatively correlated with the average VHRs of 21 rMAGs ($p < 0.05$; **Fig. S7a** and **b**), especially those for a *Marmoricola* species (**Fig. 3e**), suggesting that PAH exposure and virus–host interactions are partially linked.

Numbers of CRISPR-Acr interactions were increased under high-PAH exposure

The interactions between viruses and their hosts were further examined by investigating the relationship between the anti-bacteriophage immune system (i.e., CRISPR/Cas) of bacteria and the counter-defense mechanism (i.e., anti-CRISPR (Acr) proteins) of bacteriophages. Seven CRISPR/Cas operons were identified in rMAGs, five of which were in the rMAGs that

were enriched in cutotype 2 (**Fig. 4a**). This is consistent with a greater abundance of Cas operons in the contigs of cutotype 2 than in those of cutotype 1 (**Fig. 4b**). The rMAGs carrying a Cas operon were primarily associated with Actinomycetia and Alphaproteobacteria (**Fig. S5a**). The dominant Cas operon in the rMAGs was type IE (**Fig. S5a**), consistent with the results based on contigs (**Fig. 4b**).

Despite the hosts having CRISPR-Cas systems to protect against viral infection, they appeared to be infected with bacteriophages, possibly due to the presence of small protein inhibitors of CRISPR-Cas systems. Only 2 of the 357 predicted Acr homologs (182 in cutotype 1 and 175 in cutotype 2) could be matched to a known Acr protein. To explore the phylogeny of the 355 unclassified Acr homologs, a maximum likelihood tree was constructed. This revealed that 15 unclassified homologs clustered closely with known Acr proteins (five with AcrIF and others with diverse subtypes), and the rest were mostly distantly related to AcrIE and AcrIF (**Fig. 4c** and **Supplementary Table 7**). Anti-CRISPR-associated (aca) genes are often located adjacent to Acr-encoding genes to inhibit CRISPR-Cas immune system activity, thereby determining the success of viral infection[27]. Thirty-three potential Aca proteins (19 in cutotype 1 and 14 in cutotype 2) were predicted, and most encoded a conserved helix-turn-helix domain followed by a dimerization domain (**Supplementary Table 7**). Ten Aca-like proteins (Aca1–Aca7), which are similar in function and mechanism to Aca proteins[28], were also detected, and most (8/10) were carried by viruses in cutotype 2 (**Supplementary Table 7**).

Phage-encoded AMGs were associated with high-PAH exposure

To determine the potential role of bacteriophages in the metabolic responses of their hosts by cutotype, the diversity and functions of bacteriophage-encoded AMGs were investigated. This identified 153 AMGs encoding 61 types of metabolic functions, with four AMGs

297 encoding transporters (WRST, two-sided, p_{adj} = 0.016) and one AMG encoding a DNA
298 (cytosine-5)-methyltransferase (WRST, two-sided, p_{adj} = 0.003) enriched in cutotype 2 (**Fig.**
299 **5a** and **Supplementary Table 8**). The most common AMGs identified in bacteriophages with
300 lysogenic or lytic lifestyles were involved in the metabolism of some amino acids (**Fig. 5b**).
301 The AMGs involved in the biosynthesis of secondary metabolites, carbohydrate metabolism,
302 glycan biosynthesis and metabolism, and transporters were more abundant in bacteriophages
303 with a lysogenic lifestyle than in those with a lytic lifestyle (**Fig. 5b**). In contrast, AMGs
304 associated with the metabolism of cofactors and vitamins, other amino acids, and nucleotides
305 were more abundant in bacteriophages with a lytic lifestyle than in those with a lysogenic
306 lifestyle (**Fig. 5b**).
307 Several AMGs carried by bacteriophages with a higher average coverage in cutotype 2
308 than cutotype 1, such as *phnCDE* (which encode proteins for phosphonate/phosphate
309 transporters) and *queCDE* (which encode proteins for queuosine biosynthesis), were located
310 together in the same genomic region and shared the same promoter and terminator (**Fig. 5a** and
311 **Fig. S8**). Five AMGs associated with transporters were enriched in cutotype 2 over cutotype 1
312 (WRST, two-sided, p = 0.016; **Fig. 5a**). An AMG that encoded an aryl-alcohol dehydrogenase
313 (K00055; *aad*) involved in xenobiotic biodegradation was also identified in a bacteriophage
314 that infected a bacterium enriched in cutotype 2 (**Fig. 5c** and **Supplementary Table 8**).
315 Similarly, AMGs involved in the biosynthesis of arginine (K00619; *argA* encoding an amino-
316 acid N-acetyltransferase) and proline metabolism (K00286; *proC* encoding a pyrroline-5-
317 carboxylate reductase) were present in bacteriophages that infected bacteria (*Brevundimonas*
318 sp. [LOR281Cbin1] and *Haematobacter massiliensis* [LOR022Cbin4], respectively) enriched
319 in cutotype 2 (**Fig. 5c** and **Supplementary Table 8**). In addition, other AMGs associated with
320 the metabolism of cofactors and vitamins, such as the biosynthesis of vitamin B (K00643; *alaS*
321 encoding a 5-aminolevulinate synthase), were observed in bacteriophages that infected bacteria

enriched in cutotype 2 (**Fig. 5c** and **Supplementary Table 8**). The potential functions of all AMGs were further supported by high-confidence protein structure predictions (**Supplementary Table 8**). For example, the predicted structure of bacteriophage-encoded ArgA, ProC, AlaS, and Aad showed 99.7% confidence with 34% identity, 100% confidence with 36% identity, 100% confidence with 61% identity, and 100% confidence with 33% identity, respectively, to the reference sequences (**Fig. 5c**), suggesting that the predicted functions of these AMGs are similar to the references.

Discussion

Viruses, including bacteriophages, shape the diversity and composition of the human skin microbiome both in healthy individuals and those with diseases such as atopic dermatitis[19]. However, little is known about how cutotype influences skin viromes and virus–host interactions under PAH exposure. Bacteria in cheek skin samples of 124 female Chinese participants were previously categorized into two cutotypes, each displaying distinct taxonomic and functional compositions based on the level of PAH exposure and skin phenotype[15]. Specifically, participants of cutotype 2 who experienced high PAH exposure exhibited dry skin and hyperpigmentation[57], which are indicators of an accelerated skin aging process, whereas those belonging to cutotype 1, who experienced low PAH exposure, exhibited normal-to-high sebum concentrations and normal pigmentation. In the present study, we found that the viruses associated with the two cutotypes exhibited distinct diversities, compositions, functions, and lifestyles. Moreover, we found that increases in virus–host interactions under high PAH exposure were key modulators of bacterial populations and supplementers of their metabolism via bacteriophage-encoded AMGs.

Viruses affiliated with the class *Caudoviricetes* and the family *Papillomaviridae* were dominant on the cheeks of the participants who were long-term residents of two cities in China.

These viruses are also commonly found in different body sites and on the cheeks of individuals from North America[2], South Africa[5], and Australia[19], which suggests that they are important for maintaining skin health and homeostasis. However, there were taxonomic differences in the viruses between cutotypes, such as a high abundance of viruses from the family *Zierdtviridae* in cutotype 2 under high PAH exposure. Members of *Zierdtviridae* were also previously found to be enriched in metal-contaminated soils[29], suggesting that these viruses may be selected for in polluted environments. There have been no other investigations on the influence of pollutant exposure on the human virome, but a study on the earthworm gut showed a pattern of bacteria–virus diversity under high benzo[*a*]pyrene exposure[22] that contrasts with the pattern observed in the present study, i.e., high bacterial diversity[15] and low viral diversity in cutotype 2. This contrast may be attributable to differences in the concentrations and types of pollutant exposure: in the present study, under high-exposure conditions, those with cutotype 2 were differentially exposed to 10 different high-molecular-weight PAHs in a total concentration approximately four orders of magnitude lower than that of the single pollutant (benzo[*a*]pyrene) in the earthworm gut[22].

Prey–predator relationships between viruses and their bacterial hosts occur in many ecosystems, such as marine[30,31], terrestrial[32,33], and human gut[34] ecosystems, where viral infection can result in different outcomes (i.e., lysis, pseudolysogeny, or resistance) depending on host lineage[35]. Exposure to pollutants upregulates the expression of VFGs in some bacteria in the guts of mice[36], potentially increasing their pathogenicity[37]. In cutotype 2, the enriched rMAGs carried many VFGs, the interactions between viruses and potential bacterial pathogens carrying VFGs were frequent, and the viruses linked to these potential pathogenic hosts tended to have a lytic lifestyle. Viral-mediated lysis of potential pathogens may alter the balance and composition of microbiomes[38], potentially influencing skin health[20]. Moreover, identified viruses could be utilized as an alternative to antibiotics, e.g., in bacteriophage therapy to target

skin pathogens, especially antibiotic-resistant bacteria[39]. In addition, we found evidence that a pathogen in cutotype 2 could use the CRISPR-Cas immune system to defend itself against infection by a specific bacteriophage. Nonetheless, this bacterium remained susceptible to lytic infection by another bacteriophage, indicating the continuous war between bacteriophages and their potential hosts in skin and the role of CRISPR–Cas systems in generating bacteriophage genetic diversity[40]. Furthermore, although the abundance of CRISPR–Cas systems was higher in cutotype 2 than in cutotype 1, there was no substantial difference in the number of bacteriophage-carried Acr proteins between cutotypes, suggesting that bacteriophages may have also evolved diverse strategies (e.g., anti-restriction proteins) to evade or overcome bacterial defense mechanisms[41].

The interactions between viruses and bacteria often result in the modulation of bacterial host abundance, and these relationships can be described by several ecological models[23]. The "Piggyback-the-Winner" (PtW)[42] model explains the increased occurrence of lysogenic infection, which primarily targets fast-growing hosts (denoted "winners"), resulting in high abundances of both a virus and its host. Analogously, the PtW model suggests that increased occurrences of lytic infection primarily target less abundant hosts. In contrast, the "Piggyback-the-Persistent" (PtP)[43] or "Piggyback-the-Loser" (PtL)[44] model explains lysogenic infection that targets less abundant but continuously present hosts (denoted "persisters") or slow-growing hosts (denoted "losers"). In the present study, the virus–host interactions in cutotypes 1 and 2 could be described by the PtW and PtP or PtL models, respectively, indicating that environmental stresses and host ecological functions might have driven this differentiation. *Cutibacterium* were the most abundant and dominant taxa in cutotype 1, tended to have a fast growth rate, and thus could be considered the "winner." As *Cutibacterium* species' VHR was < 1, lysogenic infection was prevalent, with viruses piggybacking on *C. acne* strains and other *Cutibacterium* species. Although frequent interactions between viruses and *Cutibacterium*

have been observed on normal human skin[2,35], including samples of cutotype 1, further investigation is needed to understand the specific skin and/or environmental conditions under which viruses are likely to piggyback on linked *Cutibacterium* hosts. In contrast, many of the low-abundance genera in cutotype 1 had a slow growth rate and a VHR > 1, suggesting that lytic infection was prevalent and viral lysis played a role in modulating their abundances. This may explain why the skin microbiome is typically dominated by a few genera, especially in sebaceous body sites[35,45,46].

Similar to cutotype 1, in cutotype 2, lysogenic infection was observed in interactions between the dominant and abundant bacteria, including *Cutibacterium* and *Micrococcus*, and viruses. However, many of the low-abundance genera in cutotype 2 had a VHR close to 1, substantially lower than that of the low-abundance hosts in cutotype 1, suggesting that lysogenic infection was prevalent in cutotype 2 and these hosts could be considered "persisters" or "losers." Interestingly, many of these hosts, such as *Janibacter* spp.[25], *Marmoricola* spp.[26], and *Paracoccus* spp.,[47] could biodegrade PAHs. Similarly, it was reported that in a chlorinated-hydrocarbon-contaminated aquifer, virus–host interactions followed the PtP or PtL model[43]: the less abundant "persisters" exhibiting efficient transformation of contaminants were more prone than "winners" to lysogenic infection by viruses. These results suggest that "persisters" or "losers" may be involved in specialized ecological functions and that virus–host interactions could influence pollutant degradation by modulating host abundance.

Bacteriophages can insert AMGs into their hosts that help them adapt to their environment[22,48]. In the present study, we found evidence that an AMG encoding an aryl-alcohol dehydrogenase[49] was inserted into a bacterium in cutotype 2 under high-PAH exposure conditions, suggesting that this enzyme may help the bacterium to biodegrade xenobiotics. Pathways for xenobiotic degradation are enriched in many bacteria[15] and viruses under high-PAH exposure conditions, which suggests that bacteriophages endow bacteria with the ability

to metabolize xenobiotics. However, not all bacteriophage-encoded AMGs are functional or contribute to host metabolism[22]. This was the case with the presence of AMGs (i.e., *argA* and *proC*) for arginine biosynthesis in some bacteriophages in the present study, where arginine is a component of filaggrin-derived natural moisturizing factor (NMF) and helps maintain skin hydration[50]. Specifically, although *argA* and *proC* in bacteriophages were linked to their hosts in cutotype 2, there was no evidence that these genes were inserted into their hosts, and they were not enriched in any viruses. This suggests that viruses in cutotype 2 did not influence the metabolism of their hosts and thus did not alter the dry skin phenotype through the production of NMFs. Another example involves an AMG encoding a 5-aminolevulinate synthase for heme biosynthesis[51], which is associated with the production of vitamin B6, a compound essential for skin development and maintenance[52]. This AMG was linked to a bacterium in cutotype 2, but there was no evidence of its insertion into this bacterium. However, although the above-described AMGs appeared to not insert into bacteria, their metabolic functions likely serve as a reservoir for transfer between bacteriophages and hosts when suitable conditions emerge[53].

The present study has shed light on the nature of skin viruses under PAH exposure. Nevertheless, it has a few limitations. First, RNA viruses, which are known to influence skin health, were not considered[54]. Furthermore, owing to the low recovery of fungal sequences in the metagenomes, interactions between fungi and viruses were not considered. Second, the relatively low sequencing depth employed, together with potential biases in the genomic DNA extraction methods, might have led to underestimation of the diversity of viruses and their hosts. Nonetheless, the substantial number of interactions identified between most of the identified viruses and hosts shows that the results are representative. Third, the metagenomic sequencing results did not provide insights into the functionality or defectiveness of the identified viruses and their associated genes (e.g., Acrs and AMGs), and the ecological model of each cutotype was interpreted based on in silico results. Last, this study did not specifically address mobile

447 genetic elements such as plasmids and bacteriophage-plasmids, which play a pivotal role in the
448 co-evolutionary dynamic interactions between bacteriophages and bacteria[55]. Thus, future
449 research should include in vitro culturing experiments in a skin model under controlled settings
450 to verify virus–host interactions, determine the specific conditions under which AMGs are
451 inserted and expressed, and identify the bacteriophages that can be used to modulate bacterial
452 abundance under PAH exposure. This may lead to the development of new bacteriophage-
453 based treatments for bacterial infection.

454 In summary, this study provides the first insights into how the extent of PAH exposure
455 influences the diversity, composition, function, and lifestyle of skin viruses within cutotypes,
456 and into virus–host interactions, which can be described by two different ecological models.
457 The evidence suggests that viruses mitigate the effects of PAH on skin by inserting key
458 functional genes into their bacterial hosts. Overall, skin exposure to air pollutants was found to
459 not only alter the skin's microbiome but also its virome, highlighting the complex interplay
460 between environmental stresses and microbial consortia on the skin. A deeper understanding
461 of this interplay may ultimately lead to the development of strategies to harness the power of
462 viruses to improve skin health under pollution exposure.

463

464 **Materials and Methods**

465 **Characteristics of participants and processing of metagenomic sequences**

466 The metagenomic datasets consisted of 124 cheek skin samples collected from a cohort
467 of female participants aged 25–45 who had resided in Baoding (more polluted; n = 60) or
468 Dalian (less polluted; n = 64) for at least 15 years. These participants have been involved in a
469 series of studies that have examined differences between them by city and in terms of the
470 concentrations of PAHs in hair[56], the association of PAH exposure with facial aging signs[57], the
471 taxonomic compositions of skin bacteria and fungi based on 16S rRNA gene and internal

transcribed spacer region amplicon sequencing[14], metabolomic and proteomic profiles[58], and skin bacterial cutotypes based on metagenomic sequencing[15]. The sample collection and measurement of each participants' skin clinical parameters (i.e., acne onset, age group, sebum concentrations, and facial pigmentation frequency) have been detailed in previous studies[14,15,56,57,58].

The cheek swabs and four negative controls (i.e., new sterile swabs) were processed and sequenced as previously described[14,58]. Briefly, a PowerSoil DNA Isolation Kit (MO BIO Laboratories, CA, USA) was used according to the manufacturer's instructions[14] to extract genomic DNA from swabs that had been used to sample a cheek surface. Next, library preparation and sequencing were performed on an Illumina NovaSeq platform by SeqMatic (Fremont, CA, USA) to generate 150-bp paired-end reads. Quality control and assembly of reads into contigs were performed as previously described[14,15]. Briefly, adapters were removed from the raw sequences using AdapterRemoval (v.2.3.1)[59]. Quality filtering and trimming were performed using KneadData (v.0.7.4) with the default parameters, and human sequences were removed using the human genome hg38 as the reference.

To remove contaminant sequences, any reads in a sample that could be mapped to contigs in the negative controls were removed using an in-house script, and any unpaired reads were further removed from the paired-end FastQ files using fastq-pair (v.1.0)[60]. Subsequently, taxonomic classification of the reads was performed using Kraken2 (v.2.0.7-beta; *k*-mer length = 35, confidence score threshold = 0)[61], species-level abundance estimation was performed using Bracken (v.2.6.1; threshold for filter = 0)[62], and all contaminating species identified by the R package "decontam" (v.1.12) based on prevalence mode with 0.1 as the significance threshold were then removed, as previously described[15]. After quality control, ~1.1×10^9 paired-end clean reads were retained from all the samples. These reads were assembled into ~9.4×10^6 contigs using the "assembly" function in MetaWRAP (v.1.2.1)[63].

Recovery of viral bins and reconstruction of MAGs

Viral contigs were detected using VirSorter2[64] and DeepVirFinder[65], as previously described[48]. Variational autoencoders were employed for metagenomic binning: microbial species were binned using VAMB (v.3.1),[66] and bacteriophages were binned using PHAMB (v.1.0.1)[67]. VAMB was used to cluster contigs with a length > 1,000 bp into putative microbial taxa with a depth matrix and to generate the tetranucleotide frequencies of the contigs, which were then parsed with PHAMB to generate putative viral bins. The quality of these bins was assessed using CheckV (v.0.8.1; database v.1.0)[68] with the "end_to_end" pipeline[69]. Only the proviral regions of viral bins that were predicted as proviruses by CheckV were retained for further analysis. The lifestyle of viral bins was predicted using BACPHLIP[70] and VIBRANT[71].

Reconstruction of MAGs was performed as previously described[72]. Briefly, contigs with a length > 1,000 bp in each sample were binned into MAGs using the "binning" function in MetaWRAP[63]. The resulting MAGs were further refined using the "bin_refinement" function in MetaWRAP and then dereplicated into representative MAGs (rMAGs) using the "dereplicate" function with the default setting in dRep (v.3.2.2)[73]. The genomic quality of the rMAGs was assessed using the "lineage-wf" function in CheckM (v.1.1.2)[74]. Only medium-quality (completeness ≥ 50% and contamination ≤ 10%) and high-quality (completeness ≥ 75% and contamination ≤ 5%) rMAGs[75] were retained for downstream analysis. The taxonomies of rMAGs were assigned using the "gtdbtk classify_wf" function in GTDB-Tk (v.1.5.1)[76]. Different rMAGs that were assigned to the same species (95–99% average nucleotide identity [ANI]) were considered as separate strains. Phylogenetic analysis of rMAGs was performed using PhyloPhlAn3[77] and visualized using iTOL (https://itol.embl.de).

Viral bin clustering and taxonomic assignment of vOTUs

All viral bins with a completeness ≥ 50% were clustered into species-level vOTUs on the

basis of a 95% ANI in the > 85% alignment fraction, relative to the shorter sequences, using centroid-based clustering[69]. Genus- and family-level vOTUs were generated using a combination of shared genes and amino acid identity (AAI) based on Markov clustering[78], as described previously[69]. Briefly, viral bins with < 20% AAI or < 10% shared genes and an inflation factor of 1.2 were clustered into family-level vOTUs, while those with < 50% AAI or < 20% shared genes and an inflation factor of 2.0 were clustered into genus-level vOTUs.

The ORFs in the vOTUs were predicted using Prodigal (v.2.6.3)[79] with the parameter "-p meta." The protein-coding sequences of vOTUs were taxonomically assigned using the majority-rule approach, as previously described[69]. Briefly, the taxonomy (based on the International Committee on Taxonomy of Viruses) of the top hit in the IMG/VR (v4) database[80] obtained using DIAMOND (v.0.9.32; options: –query-cover 50 –subject-cover 50 –E-value 1e-5 –max-target-seqs 1000 –top 25) was transferred to each protein. Each vOTU was assigned to the lowest taxonomic rank of > 70% of the annotated proteins. For family- and genus-level taxonomy assignment, a vOTU must have at least two annotated proteins with > 30% average AAI or three annotated proteins with > 40% average AAI, respectively, aligned to a reference genome in the IMG/VR database[69].

Estimation of coverage of rMAGs and vOTUs

The clean reads from all the metagenomes were mapped to the vOTUs using Bowtie2 (v.2.4.4) with the default parameters. The "filter" function in BamM (v.1.7.3) was used to remove low-quality mappings, and reads that were aligned over ≥ 90% of their length at ≥ 95% ANI were retained. The vOTUs with ≥ 70% of their length covered by the reads were selected using a Python script in Read2RefMapper (v.1.1.0), and the average per-base-pair coverage of each vOTU in each sample was generated using the "parse" function in BamM with the parameter "tpmean" to remove the highest and lowest 10%-coverage regions. To normalize the

coverage of the vOTUs across all samples, the number of reads in each sample was divided by the average number of reads across all samples and multiplied by the coverage retained by BamM, as previously described[32].

To calculate the coverage of each rMAG in all samples, the clean reads from all metagenomes were mapped to each rMAG using the "make" function in BamM. Low-quality read mappings (< 75% of the aligned length of each read and < 95% ANI) were removed using the "filter" function in BamM. The coverage of each contig in each rMAG in a sample was calculated as the average number of reads aligned to each position in the contigs after removal of the highest and lowest 10%-coverage regions via the BamM "parse" function in the "tpmean" mode. Subsequently, the coverage of each rMAG in all samples was calculated as the average coverage of all its binned contigs, with each contig weighted by its length in base pairs. The coverage of rMAGs across all samples was normalized by read numbers following the method described above for the vOTUs.

Functional annotation of vOTUs and rMAGs

The ORFs of vOTUs were functionally annotated against the KO database using KOfamscan (https://www.genome.jp/tools/kofamkoala/)[81] with an E-value < 0.01. The coverage of genes in vOTUs was calculated by multiplying the corresponding coverage of vOTUs with the sequence coverage of genes that were aligned to vOTUs, using a BLASTn alignment threshold of identity ≥ 70% and an E-value ≤ 10^{-5}. Phage-encoded AMGs were annotated using the viral mode of DRAM (v.1.4.0)[82] and VIBRANT (v.1.2.1)[71] with the default parameters. The potential functions of AMGs were further assessed based on their predicted protein structures. The secondary and tertiary structures of AMGs were predicted by conducting a homology search against template models in the Protein Data Bank[83] and the AlphaFold Protein Structure Database[84] using Phyre2 (v.2.0)[85] with an alignment coverage of

572 > 65%. The BPROM (http://www.softberry.com/) and FindTerm (http://www.softberry.com)
573 servers were used to predict the promoters and terminators of AMGs, respectively.
574 The ORFs of rMAGs were predicted using Prokka (v.1.14.6)[86] and their functions were
575 annotated using EggNOG-mapper (v.2.1.11)[87] with the "Diamond" option and an E-value ≤ 10^{-3}.
576 VFGs were determined by aligning all ORFs to the Virulence Factor DataBase[88] using
577 BLASTp with an E-value ≤ 10^{-5}, an AAI of ≥ 80% and a query coverage of ≥ 80%[89]. The
578 maximal growth rate of each rMAG was estimated using the R package "gRodon" (v.2.3.0)[90]
579 based on the codon usage biases of the sequences.
580
581 **Determination of virus–host links**
582 Three complementary in silico methods were adopted to improve the accuracy of host
583 prediction, namely methods that find exact or close matches of CRISPR spacers and integrated
584 viral fragments in host genomes and a method that finds consistent *k*-mer signatures, as described
585 previously[48]. Briefly, CRISPR spacers in the rMAGs were identified using CRT (v.1.2)[91] and
586 PILER-CR (v.1.0.6)[92]. The identified spacers were compared against the vOTUs using BLASTn
587 with ≤ 1 mismatch, 100% identity, and 100% coverage to establish virus–host links. Genomic
588 regions shared between vOTUs and rMAGs were detected using BLASTn with an identity ≥
589 90%, a hit length ≥ 1,000 bp, and an E-value ≤ 0.001. The *k*-mer frequencies of sequences from
590 rMAGs and vOTUs were estimated using WIsH (v.1.1)[93], as previously described[48]. The host
591 predicted by WIsH that had the lowest *p*-value (≤ 10^{-5}) was retained for each vOTU. All the hosts
592 predicted by the above-mentioned three methods were retained for use in analyzing virus–host
593 interactions. The network of virus–host links was visualized using Cytoscape (v.3.9.0)[94].
594 The VHR of a vOTU and its linked host based on its rMAG was calculated by dividing
595 the respective coverages in a sample. The correlations between the abundances of vOTUs and
596 their predicted hosts, and between PAH concentrations and VHRs, were estimated by Pearson's

correlation analysis.

Detection of CRISPR/Cas and Anti-CRISPR/Cas systems

The CRISPR/Cas system in a rMAG was identified using CrisprCasTyper (v.1.2.3)[95]. The counter-defense systems in a vOTU, including Acr homologs, Acr-associated (Aca) genes, and Aca-like proteins, were identified using AcaFinder (version: Oct 15, 2022)[96]. To examine the phylogeny of Acr homologs, all the predicted Acr homologs were compared with the core dataset in anti-CRISPRdb (v.2.2)[97] using BLASTp with an identity ≥ 60% and an E-value ≤ 10^{-5}. Multiple sequence alignments between the predicted Acr homologs and the reference Acr proteins obtained from the core dataset were generated using FAMSA (v.1.5.12)[98] and trimmed using trimal (v.1.4)[99] to retain positions with < 50% gaps. A maximum-likelihood phylogenetic tree was constructed with the aligned sequences using FastTreeMP (v.2.1.11) with the auto model and then visualized using iTOL.

Measurement of PAHs and other compounds in hair samples

The participants' levels of exposure to pollutants were assessed by analyzing hair samples, as described previously[56,100]. Briefly, a hair was cut as close to the skin as possible from the occipital area of each participant and decontaminated, and then the first 12 cm (from the skin end) of the sample were analyzed. Specifically, this portion of the sample was pulverized, hydrolyzed, and extracted, and the extract was analyzed by gas or liquid chromatography–tandem mass spectrometry to determine its concentrations of 15 PAHs, nicotine, and cotinine.

Statistical analysis

All statistical analyses were conducted using R (v.4.2.2), and a *p*-value of < 0.05 was considered to indicate a statistically significant difference. α-diversity was assessed based on

Shannon's diversity index and calculated using the R package "vegan" (v.2.6-4)[101]. One-way ANOVAs with Tukey's honestly significant difference test were used to compare the differences between the α-diversities of cutotypes. Principal coordinate analysis based on Bray–Curtis dissimilarity was performed using the "vegdist" function in the R package "vegan," and the multivariate homogeneity of variances was analyzed to test for differences in multivariate dispersions between cutotypes using the "betadisper" function in the same R package. The influences of factors (cutotype, city, acne onset, age group, sebum concentrations, and facial pigmentation frequency) on skin viral communities were assessed using a PERMANOVA through the "adonis2" function in the R package "vegan." Canonical correspondence analysis between PAH concentrations and viral taxonomic compositions was performed using the R package "vegan." The Procrustes test was utilized to determine congruency between taxonomic and functional compositions using the "protest" function in the R package "vegan." The associations of rMAGs, vOTUs, and functions (i.e., KOs and KEGG pathways) with cutotypes were determined via a generalized linear model through MaAsLin2[102], with an adjusted *p*-value (i.e., *q*-value) of < 0.05 as the threshold for statistical significance. PLS-PM was performed using the R package "plspm" (v.0.5.0), and the Euclidean distance matrix of PAHs and the Bray–Curtis dissimilarity matrix of bacterial and viral compositions were used to account for the effects of PAHs on bacteria and viruses. Two-tailed WRST with Benjamini–Hochberg (BH) adjustment was used to determine statistical differences between the two groups.

Data availability

The raw paired-end metagenomics sequences and rMAG sequences have been deposited in the National Center of Biotechnology Information's BioProject under accession number PRJNA730653. The following publicly available databases were used in this study: IMG/VR

(https://genome.jgi.doe.gov/portal/IMG_VR/), the Virulence Factor DataBase (http://www.mgc.ac.cn/VFs/), and anti-CRISPRdb (http://guolab.whu.edu.cn/anti-CRISPRdb/). The high-confidence structures and sequences of the AMGs, and the sequences of Acr homologs, are available at https://github.com/Dorothydu12/skin_virome_pah.

Code availability

Custom scripts and input and output files generated for analysis and figures are available at https://github.com/Dorothydu12/skin_virome_pah.

References

1. Findley K, *et al.* Topographic diversity of fungal and bacterial communities in human skin. *Nature* **498**, 367-370 (2013).
2. Saheb Kashaf S, *et al.* Integrating cultivation and metagenomics for a multi-kingdom view of skin microbiome diversity and functions. *Nat Microbiol* **7**, 169-179 (2022).
3. Hurabielle C, *et al.* Immunity to commensal skin fungi promotes psoriasiform skin inflammation. *Proc Natl Acad Sci U S A* **117**, 16465-16474 (2020).
4. Hannigan GD, *et al.* The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *mBio* **6**, e01578-01515 (2015).
5. van Zyl LJ, *et al.* Novel phages of healthy skin metaviromes from South Africa. *Sci Rep* **8**, 12265 (2018).
6. Belkaid Y, Segre JA. Dialogue between skin microbiota and immunity. *Science* **346**, 954-959 (2014).
7. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal stability of the human skin microbiome. *Cell* **165**, 854-866 (2016).
8. Wei Q, *et al.* Shotgun metagenomic sequencing reveals skin microbial variability from different facial sites. *Front Microbiol* **13**, 933189 (2022).
9. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: A review. *Front Public Health* **8**, 14 (2020).
10. Mallah MA, *et al.* Polycyclic aromatic hydrocarbon and its effects on human health: An overeview. *Chemosphere* **296**, 133948 (2022).

11. Isler MF, Coates SJ, Boos MD. Climate change, the cutaneous microbiome and skin disease: implications for a warming world. *Int J Dermatol* **62**, 337-345 (2023).

12. Abolhasani R, Araghi F, Tabary M, Aryannejad A, Mashinchi B, Robati RM. The impact of air pollution on skin and related disorders: A comprehensive review. *Dermatol Ther* **34**, e14840 (2021).

13. Chakravarti D*, et al.* The role of polycyclic aromatic hydrocarbon–DNA adducts in inducing mutations in mouse skin. *Mutat Res Genet Toxicol Environ Mutagen* **649**, 161-178 (2008).

14. Leung MHY*, et al.* Changes of the human skin microbiota upon chronic exposure to polycyclic aromatic hydrocarbon pollutants. *Microbiome* **8**, 100 (2020).

15. Leung MHY*, et al.* Skin microbiome differentiates into distinct cutotypes with unique metabolic functions upon exposure to polycyclic aromatic hydrocarbons. *Microbiome* **11**, 124 (2023).

16. Sowada J, Schmalenberger A, Ebner I, Luch A, Tralau T. Degradation of benzo[*a*]pyrene by bacterial isolates from human skin. *FEMS Microbiol Ecol* **88**, 129-139 (2014).

17. Li Z*, et al.* Characterization of the human skin resistome and identification of two microbiota cutotypes. *Microbiome* **9**, 47 (2021).

18. Chen YE, Fischbach MA, Belkaid Y. Skin microbiota–host interactions. *Nature* **553**, 427-436 (2018).

19. Wielscher M*, et al.* The phageome in normal and inflamed human skin. *Sci Adv* **9**, eadg4015 (2023).

20. Barnard E, Shi B, Kang D, Craft N, Li H. The balance of metagenomic elements shapes the skin microbiome in acne and health. *Sci Rep* **6**, 39491 (2016).

21. Pamela KC, Christina AK, John HP. Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. *Mar Ecol Prog Ser* **164**, 125-133 (1998).

22. Xia R, Sun M, Balcázar JL, Yu P, Hu F, Alvarez PJJ. Benzo[*a*]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. *ISME J* **17**, 1004-1014 (2023).

23. Brown TL, Charity OJ, Adriaenssens EM. Ecological and functional roles of bacteriophages in contrasting environments: marine, terrestrial and human gut. *Curr Opin Microbiol* **70**, 102229 (2022).

24. Kim SY, Lee J, Lee SY. Metabolic engineering of *Corynebacterium glutamicum* for the production of L-ornithine. *Biotechnol Bioeng* **112**, 416-421 (2015).

25. Tikilili PV, Nkhalambayausi-Chirwa EM. Characterization and biodegradation of polycyclic aromatic hydrocarbons in radioactive wastewater. *J Hazard Mater* **192**, 1589-1596 (2011).

26. Li J*, et al.* Autochthonous bioaugmentation with non-direct degraders: A new strategy to enhance wastewater bioremediation performance. *Environ Int* **136**, 105473 (2020).

27. Stanley SY*, et al.* Anti-CRISPR-associated proteins are crucial repressors of anti-CRISPR transcription. *Cell* **178**, 1452-1464. e1413 (2019).

28. Shehreen S, Birkholz N, Fineran Peter C, Brown Chris M. Widespread repression of anti-CRISPR production by anti-CRISPR-associated proteins. *Nucleic Acids Res* **50**, 8615-8625 (2022).

29. Zhang H*, et al.* Dissecting the metal resistance genes contributed by virome from mining-affected metal contaminated soils. *Front Environ Sci* **11**, (2023).

30. Coutinho FH*, et al.* Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun* **8**, 15955 (2017).

31. Bartlau N*, et al.* Highly diverse flavobacterial phages isolated from North Sea spring blooms. *ISME J* **16**, 555-568 (2022).

32. Emerson JB*, et al.* Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol* **3**, 870-880 (2018).

33. Albright MB*, et al.* Experimental evidence for the impact of soil viruses on carbon cycling during surface plant litter decomposition. *ISME Commun* **2**, 24 (2022).

34. Liang G*, et al.* The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature* **581**, 470-474 (2020).

35. Liu J*, et al.* The diversity and host interactions of *Propionibacterium acnes* bacteriophages on human skin. *ISME J* **9**, 2078-2093 (2015).

36. Viennois E*, et al.* Dietary emulsifiers directly impact adherent-invasive *E. coli* gene expression to drive chronic intestinal inflammation. *Cell Rep* **33**, (2020).

37. Al Haj Ishak Al Ali R, Mondamert L, Berjeaud JM, Jandry J, Crépin A, Labanowski J. Application of QSAR Approach to Assess the Effects of Organic Pollutants on Bacterial Virulence Factors. *Microorganisms* **11**, 1375 (2023).

38. Chatterjee A, Duerkop BA. Beyond Bacteria: Bacteriophage-Eukaryotic Host Interactions Reveal Emerging Paradigms of Health and Disease. *Front Microbiol* **9**, 1394 (2018).

39. Kortright KE, Chan BK, Koff JL, Turner PE. Phage therapy: A renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**, 219-232 (2019).

40. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their

phage foes. *Nature* **577**, 327-336 (2020).

41. Hobbs SJ*, et al.* Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522-526 (2022).

42. Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms Microbiomes* **2**, 16010 (2016).

43. Paterson JS*, et al.* A hydrocarbon-contaminated aquifer reveals a Piggyback-the-Persistent viral strategy. *FEMS Microbiol Ecol* **95**, fiz116 (2019).

44. Knowles B*, et al.* Lytic to temperate switching of viral communities. *Nature* **531**, 466-470 (2016).

45. Almoughrabie S*, et al.* Commensal *Cutibacterium acnes* induce epidermal lipid synthesis important for skin barrier function. *Sci Adv* **9**, eadg6262 (2023).

46. Christensen GJM, Brüggemann H. Bacterial skin commensals and their role as host guardians. *Benef Microbes* **5**, 201-215 (2014).

47. Phale PS, Malhotra H, Shah BA. Degradation strategies and associated regulatory mechanisms/features for aromatic compound metabolism in bacteria. *Adv Appl Microbiol* **112**, 1-65 (2020).

48. Du S, Tong X, Lai ACK, Chan CK, Mason CE, Lee PKH. Highly host-linked viromes in the built environment possess habitat-dependent diversity and functions for potential virus-host coevolution. *Nat Commun* **14**, 2676 (2023).

49. Lopez ES, Elufisan TO, Bustos P, Charles CPM, Mendoza-Herrera A, Guo X. Complete Genome report of a hydrocarbon-degrading *Sphingobium yanoikuyae* S72. *Appl Sci* **12**, 6201 (2022).

50. Kubo A*, et al.* The stratum corneum comprises three layers with distinct metal-ion barrier properties. *Sci Rep* **3**, 1731 (2013).

51. Heinemann IU, Jahn M, Jahn D. The biochemistry of heme biosynthesis. *Arch Biochem Biophys* **474**, 238-251 (2008).

52. Kato, N. Role of vitamin b6 in skin health and diseases. In Handbook of Diet, Nutrition and the Skin; Preedy, V.R., Ed.; Human Health Handbooks; Wageningen Academic Publishers: Wageningen, The Netherlands, 2012.

53. Tan Y*, et al.* Enhanced bacterium–phage symbiosis in attached microbial aggregates on a membrane surface facing elevated hydraulic stress. *Environ Sci Technol* **57**, 17324-17337 (2023).

54. Kawamura T, Ogawa Y, Aoki R, Shimada S. Innate and intrinsic antiviral immunity in skin. *J Dermatol Sci* **75**, 159-166 (2014).

55. Pfeifer E, Rocha EPC. Phage-plasmids promote recombination and emergence of

phages and plasmids. *Nat Commun* **15**, 1545 (2024).

56. Palazzi P*, et al.* Exposure to polycyclic aromatic hydrocarbons in women living in the Chinese cities of BaoDing and Dalian revealed by hair analysis. *Environ Int* **121**, 1341-1354 (2018).

57. Flament F, Bourokba N, Nouveau S, Li J, Charbonneau A. A severe chronic outdoor urban pollution alters some facial aging signs in Chinese women. A tale of two cities. *Int J Cosmet Sci* **40**, 467-481 (2018).

58. Misra N*, et al.* Multi-omics analysis to decipher the molecular link between chronic exposure to pollution and human skin dysfunction. *Sci Rep* **11**, 18302 (2021).

59. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes* **9**, 88 (2016).

60. Edwards JA, Edwards RA. Fastq-pair: efficient synchronization of paired-end fastq files. *bioRxiv*, (2019).

61. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).

62. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput Sci* **3**, e104 (2017).

63. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).

64. Guo J*, et al.* VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).

65. Ren J*, et al.* Identifying viruses from metagenomic data using deep learning. *Quant Biol* **8**, 64-77 (2020).

66. Nissen JN*, et al.* Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol* **39**, 555-560 (2021).

67. Johansen J*, et al.* Genome binning of viral entities from bulk metagenomics data. *Nat Commun* **13**, 965 (2022).

68. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* **39**, 578-585 (2021).

69. Nayfach S*, et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol* **6**, 960-970 (2021).

70. Hockenberry AJ, Wilke CO. BACPHLIP: Predicting bacteriophage lifestyle from conserved protein domains. *PeerJ* **9**, e11396 (2021).

71. Kieft K, Zhou Z, Anantharaman K. VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).

72. Tong X, Leung MHY, Shen Z, Lee JYY, Mason CE, Lee PKH. Metagenomic insights into the microbial communities of inert and oligotrophic outdoor pier surfaces of a coastal city. *Microbiome* **9**, 213 (2021).

73. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868 (2017).

74. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).

75. Bowers RM*, et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725-731 (2017).

76. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925-1927 (2020).

77. Asnicar F*, et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* **11**, 2500 (2020).

78. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).

79. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

80. Camargo AP*, et al.* IMG/VR v4: An expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res* **51**, D733-D743 (2023).

81. Aramaki T*, et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251-2252 (2019).

82. Shaffer M*, et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* **48**, 8883-8900 (2020).

83. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods Mol Biol* **1607**, 627-641 (2017).

84. Varadi M*, et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-D444 (2022).

85. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).

86. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).

87. Huerta-Cepas J*, et al.* eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-D314 (2018).

88. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Res* **50**, D912-D917 (2022).

89. Chen T*, et al.* Virus–pathogen interactions improve water quality along the Middle Route of the South-to-North Water Diversion Canal. *ISME J* **17**, 1719-1732 (2023).

90. Weissman JL, Hou S, Fuhrman JA. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc Natl Acad Sci U S A* **118**, e2016810118 (2021).

91. Bland C*, et al.* CRISPR Recognition Tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).

92. Edgar RC. PILER-CR: Fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18 (2007).

93. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: Who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113-3114 (2017).

94. Shannon P*, et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

95. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: Automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J* **3**, 462-469 (2020).

96. Yang B, Zheng J, Yin Y. AcaFinder: Genome mining for anti-CRISPR-associated genes. *mSystems* **7**, e00817-00822 (2022).

97. Dong C*, et al.* Anti-CRISPRdb v2.2: An online repository of anti-CRISPR proteins including information on inhibitory mechanisms, activities and neighbors of curated anti-CRISPR proteins. *Database* **2022**, baac010 (2022).

98. Deorowicz S, Debudaj-Grabysz A, Gudyś A. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci Rep* **6**, 1-13 (2016).

99. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-

1973 (2009).

100. Duca R-C, Hardy E, Salquèbre G, Appenzeller BMR. Hair decontamination procedure prior to multi-class pesticide analysis. *Drug Test Anal* **6**, 55-66 (2014).

101. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**, 927-930 (2003).

102. Mallick H*, et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* **17**, e1009442 (2021).

Acknowledgments

This research was funded by L'Oréal Research & Innovation, Pudong, China, and the City University of Hong Kong (7020049). We thank the participants for their time and the staff of SeqMatic for their technical advice on sequencing.

Author contributions

SD performed data analysis, data interpretation, and wrote the manuscript. XT, MHYL, and PB performed the data analysis. RJB, ACW, NM, and LA assisted with data collection and interpretation. CC conceived the study and performed data interpretation. PKHL conceived and supervised the study, and wrote the manuscript. All authors read and approved the final version of the manuscript.

Corresponding author

Correspondence to Patrick K. H. Lee (patrick.kh.lee@cityu.edu.hk).

Ethics approval and consent to participate

This research was approved by the Sino-German Cosmetics Institute Ethics Review Board (Protocol 2015–033-DY-024) and conducted in accordance with the principles of the Declaration of Helsinki developed by the World Medical Association.

Competing interests

The authors RJB, ACW, PB, NM, LA, and CC are employees of L'Oréal Research and Innovation.

Figures

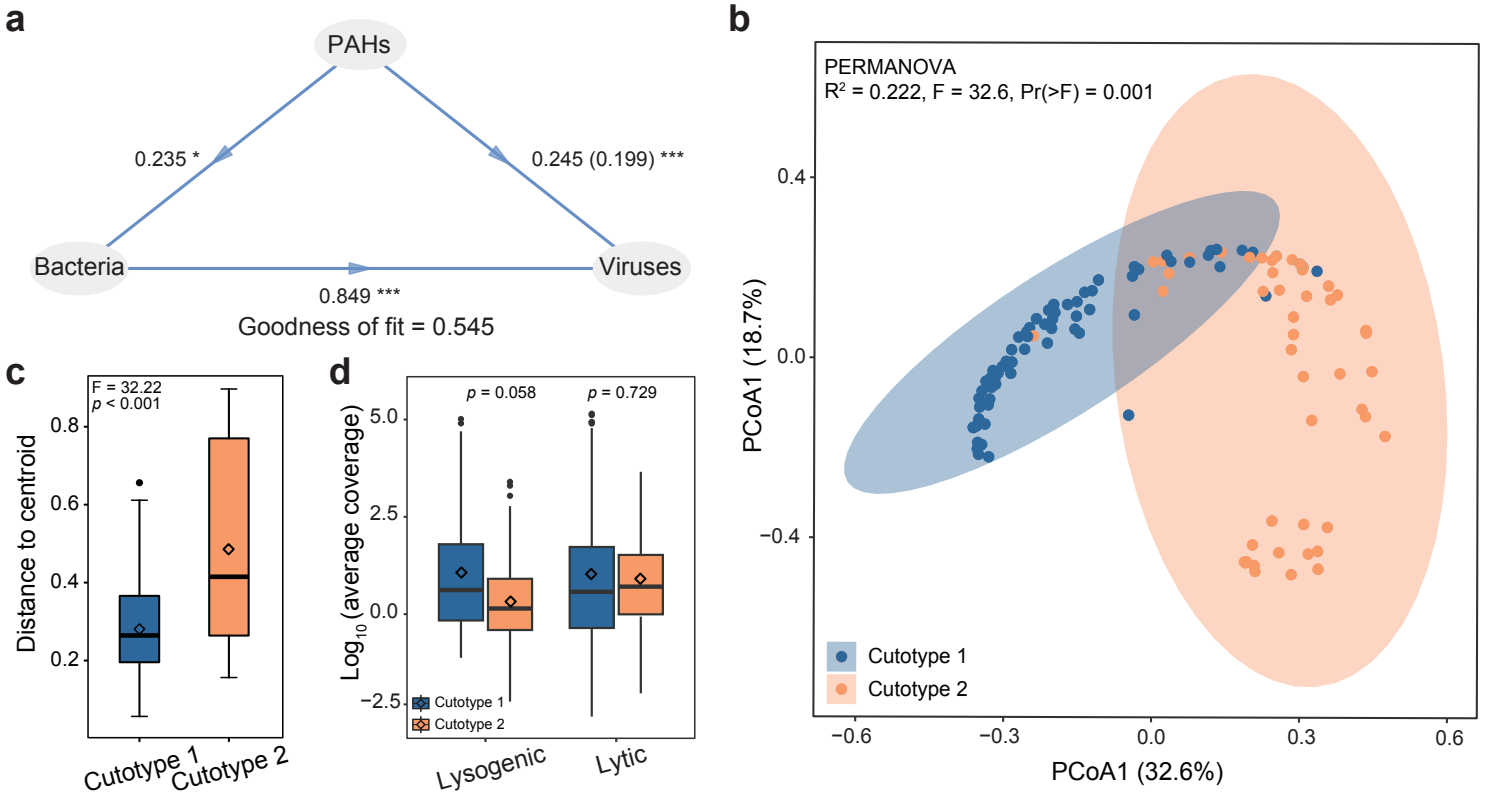


Fig. 1 Data of two clusters of skin viromes, with each cluster corresponding to bacterial cutotypes with low and high exposure to polycyclic aromatic hydrocarbons (PAHs), respectively. (a) Relationships between PAHs, bacteria, and viruses based on partial least squares path modeling (PLS-PM). The numbers next to the lines and in the brackets denote statistically significant direct and indirect path coefficients, respectively (*$p < 0.05$ and ***$p < 0.001$). **(b)** Principal coordinate analysis of the Bray–Curtis dissimilarity matrix of viruses in all cheek skin samples. Cutotypes 1 and 2 were defined based on the bacteria in the samples. Ellipses represent 95% confidence intervals. **(c)** Differences between the beta dispersions of the two cutotypes. **(d)** Lifestyles of viruses in the two cutotypes (i.e., lysogenic ($n = 62$) or lytic ($n = 110$)). In each box-and-whisker plot, the box indicates the median, the first quartile, and 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.

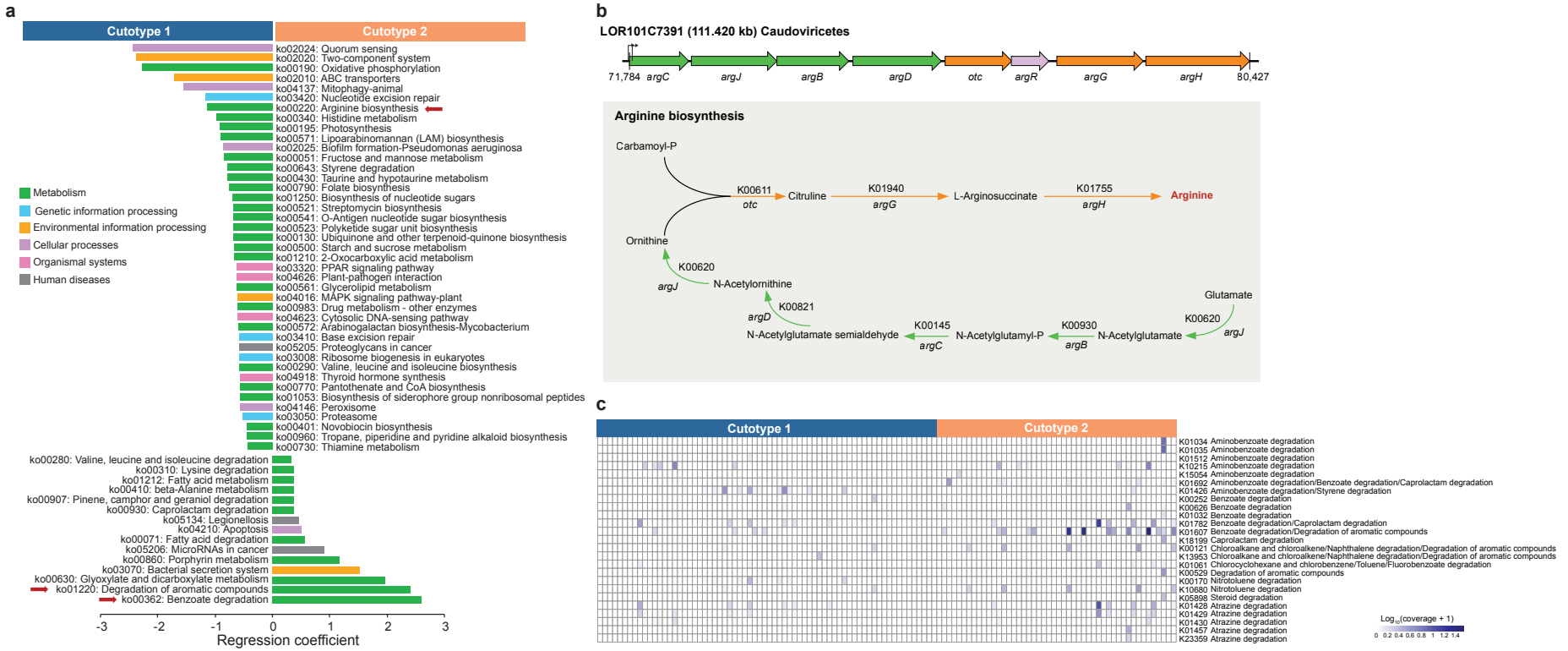
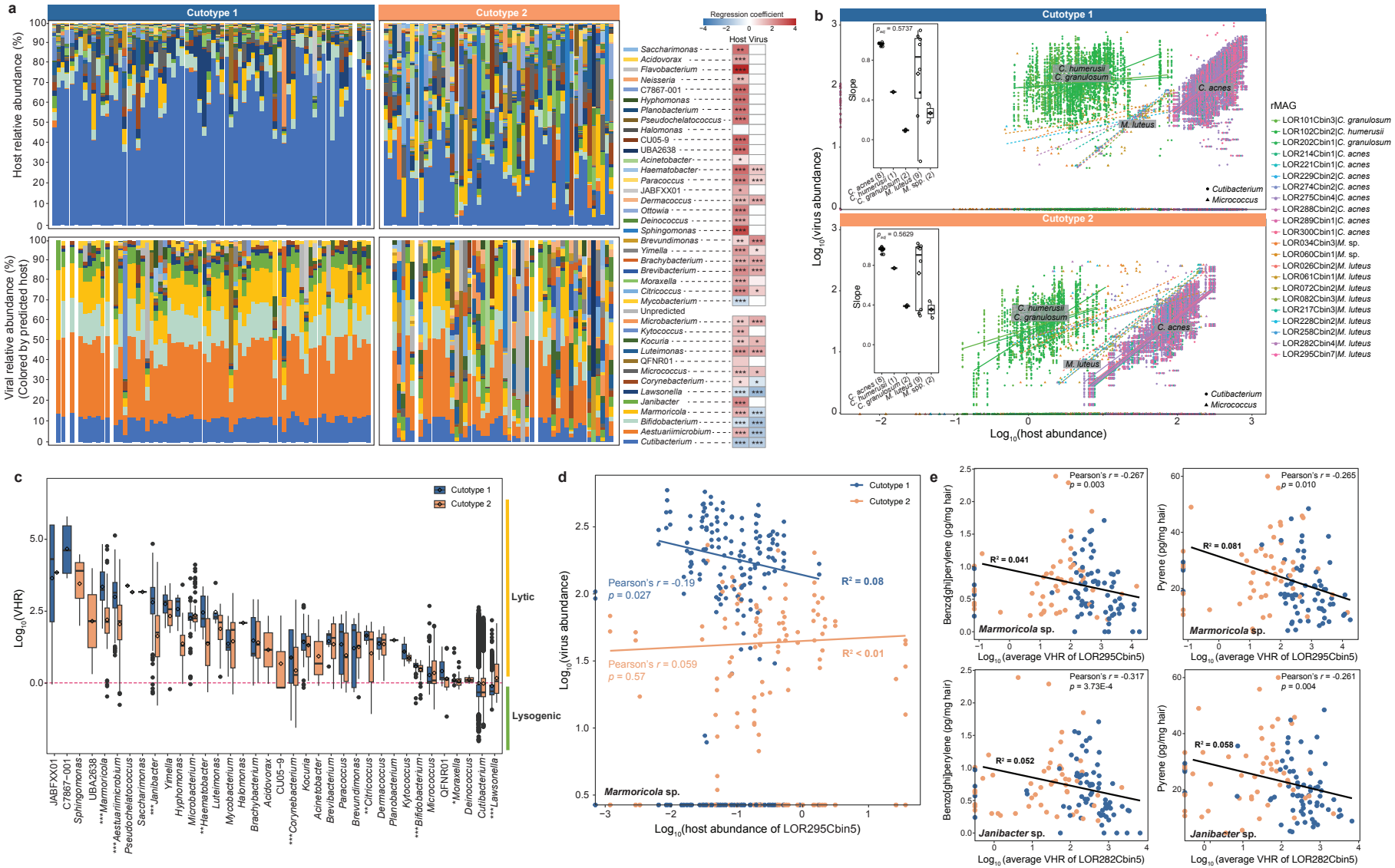


Fig. 2 Functional profile of skin viromes. **(a)** Enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways between the two cutotypes according to MaAsLin2. All statistically significant pathways (Wilcoxon rank sum test, two-sided, $*p_{adj} < 0.05$) are shown. Red arrows indicate the key pathways enriched in the respective cutotypes[10]. **(b)** Schematic illustration of genes in the arginine biosynthesis pathway carried by a virus in cutotype 1. The color represents the different functional modules, and the arrows indicate the predicted terminators. The genes and pathways involved in arginine biosynthesis are shown in the gray box. **(c)** Heatmap showing the coverage of KEGG Orthology (KO) for xenobiotic degradation in the two cutotypes in all the samples.


a
Cutotype 1
Cutotype 2
Host relative abundance (%)
Viral relative abundance (%)
(Colored by predicted host)
Regression coefficient
Host Virus
Saccharimonas
Acidovorax
Flavobacterium
Neisseria
C7867-001
Hyphomonas
Planobacterium
Pseudochelatococcus
Halomonas
CU05-9
UBA2638
Acinetobacter
Haematobacter
Paracoccus
JABFXX01
Dermacoccus
Ottowia
Deinococcus
Sphingomonas
Brevundimonas
Yimella
Brachybacterium
Brevibacterium
Moraxella
Citricoccus
Mycobacterium
Unpredicted
Microbacterium
Kytococcus
Kocuria
Luteimonas
QFNR01
Micrococcus
Corynebacterium
Lawsonella
Janibacter
Marmoricola
Bifidobacterium
Aestuariimicrobium
Cutibacterium
b
Cutotype 1
Cutotype 2
Slope
C. humerusii
C. granulosum
C. acnes
M. luteus
Cutibacterium
Micrococcus
Log10(virus abundance)
Log10(host abundance)
rMAG
LOR101Cbin3|C. granulosum
LOR102Cbin2|C. humerusii
LOR202Cbin1|C. granulosum
LOR214Cbin1|C. acnes
LOR221Cbin1|C. acnes
LOR229Cbin2|C. acnes
LOR274Cbin2|C. acnes
LOR275Cbin4|C. acnes
LOR288Cbin2|C. acnes
LOR289Cbin1|C. acnes
LOR300Cbin1|C. acnes
LOR034Cbin3|M. sp.
LOR060Cbin1|M. sp.
LOR026Cbin2|M. luteus
LOR061Cbin1|M. luteus
LOR072Cbin2|M. luteus
LOR082Cbin3|M. luteus
LOR217Cbin3|M. luteus
LOR228Cbin2|M. luteus
LOR258Cbin2|M. luteus
LOR282Cbin4|M. luteus
LOR295Cbin7|M. luteus
c
Cutotype 1
Cutotype 2
Log10(VHR)
Lytic
Lysogenic
d
Cutotype 1
Cutotype 2
Pearson's r = -0.19
p = 0.027
R² = 0.08
Pearson's r = 0.059
p = 0.57
R² < 0.01
Marmoricola sp.
Log10(virus abundance)
Log10(host abundance of LOR295Cbin5)
e
Pearson's r = -0.267
p = 0.003
R² = 0.041
Pearson's r = -0.265
p = 0.010
R² = 0.081
Pearson's r = -0.317
p = 3.73E-4
R² = 0.052
Pearson's r = -0.261
p = 0.004
R² = 0.058
Marmoricola sp.
Janibacter sp.
Benzo[ghi]perylene (pg/mg hair)
Pyrene (pg/mg hair)
Log10(average VHR of LOR295Cbin5)
Log10(average VHR of LOR282Cbin5)

1037

Fig. 3 Numbers of virus–host interactions in the two cutotypes. (a) Left panel: relative abundances of viruses and their linked bacterial hosts (colored to indicate the taxonomy of linked hosts). Right panel: heatmap of the regression coefficient from the general linear model of MaAsLin2 showing differentially abundant hosts and viruses (colored based on the linked host) between the two cutotypes (Wilcoxon rank sum test (WRST), two-sided, $*p_{adj} < 0.05$, $**p_{adj} < 0.01$, $***p_{adj} < 0.001$). Only genera enriched or depleted in cutotypes 1 and 2 are shown. The color indicates the degree of enrichment (red) or depletion (blue) in cutotype 2. **(b)** Pearson's correlations of abundances ($\log_{10}$ scale) between viruses and their linked hosts of various *Cutibacterium* species (including eight *C. acnes* strains) and *Micrococcus* species (including nine *M. luteus* strains) in cutotypes 1 and 2. The inset boxplots show the slopes of linear regression for strains of each species. **(c)** Virus-to-host ratio (VHR) at the bacterial genus level for each virus–host pair. The statistical significance of the difference in VHRs between cutotypes 1 and 2 is indicated on the *x*-axis (WRST, two-sided, $*p_{adj} < 0.05$, $**p_{adj} < 0.01$, $***p_{adj} < 0.001$). A VHR > 1 indicates that the viruses tended to have a lytic lifestyle, while a VHR < 1 indicates that the viruses tended to have a lysogenic lifestyle. **(d)** Pearson's correlations of the abundances ($\log_{10}$ scale) between viruses and their linked host (i.e., a *Marmoricola* sp.) in the two cutotypes. **(e)** Pearson's correlations between the average VHR of selected *Marmoricola* and *Janibacter* species and the measured concentrations of polyaromatic hydrocarbons (PAHs; only the VHR–PAH pairs with the two largest Pearson's correlation coefficients (*r*) are shown). Linear regression lines and coefficients of determination (R^2) are shown in panels d and e. In each box-and-whisker plot, the box indicates the median, first quartile, and 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.

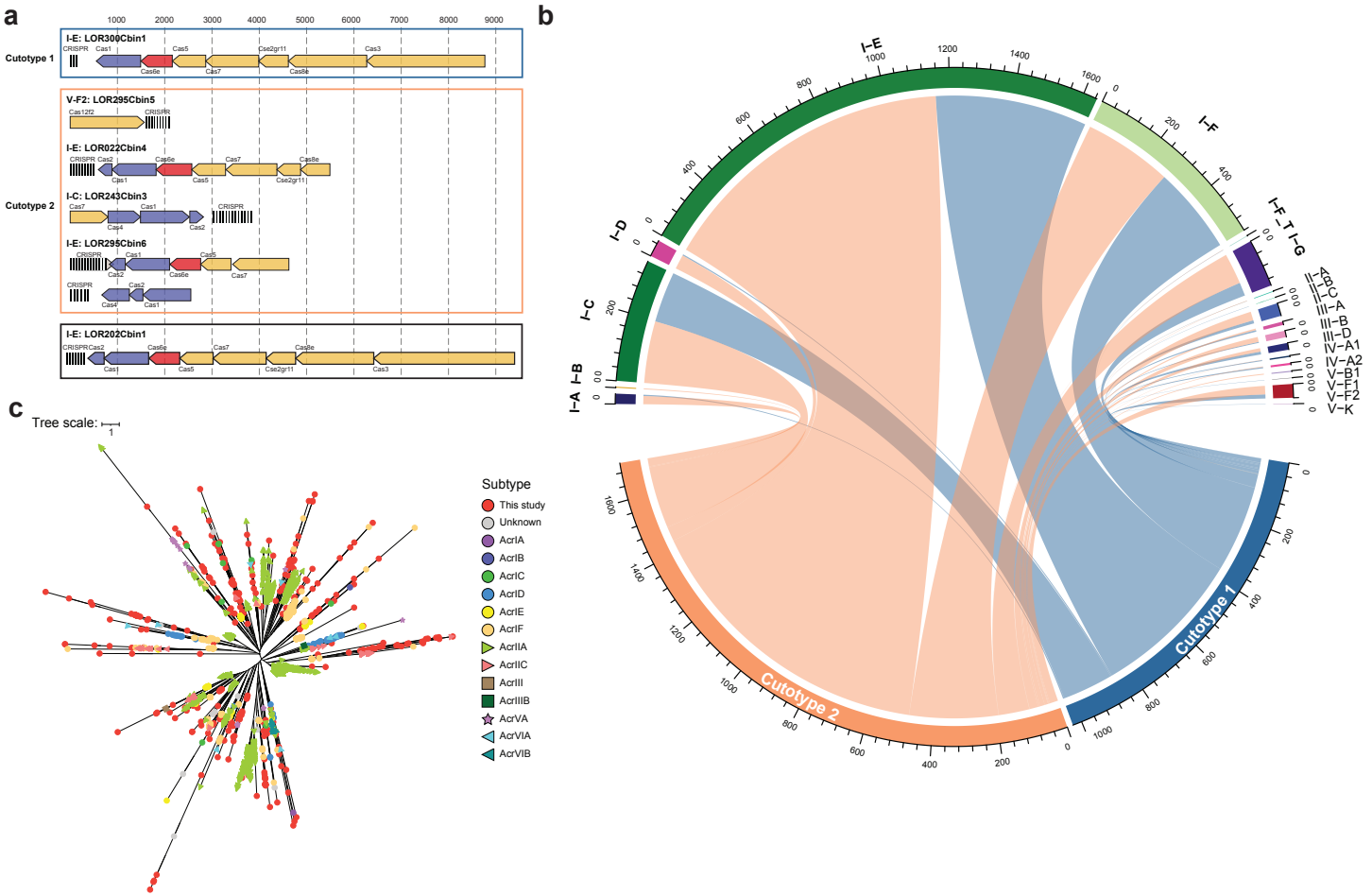


Fig. 4 Clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein (Cas) systems and predicted anti-CRISPR (Acr) proteins in the skin viromes. (a) The six representative metagenome-assembled genomes (rMAGs) in cutotypes 1 and 2 carried complete CRISPR/Cas systems, with one rMAG enriched in cutotype 1 (blue box) and four rMAGs enriched in cutotype 2 (orange box). **(b)** Circos diagram illustrating the type and number of Cas operons identified in the contigs of the two cutotypes. **(c)** Maximum-likelihood phylogenetic tree of the 357 Acr homologs identified in this study and the 3,690 Acr proteins obtained from a curated anti-CRISPR database[98]. The red branches represent predicted Acr proteins that are most distantly related to known Acr proteins.

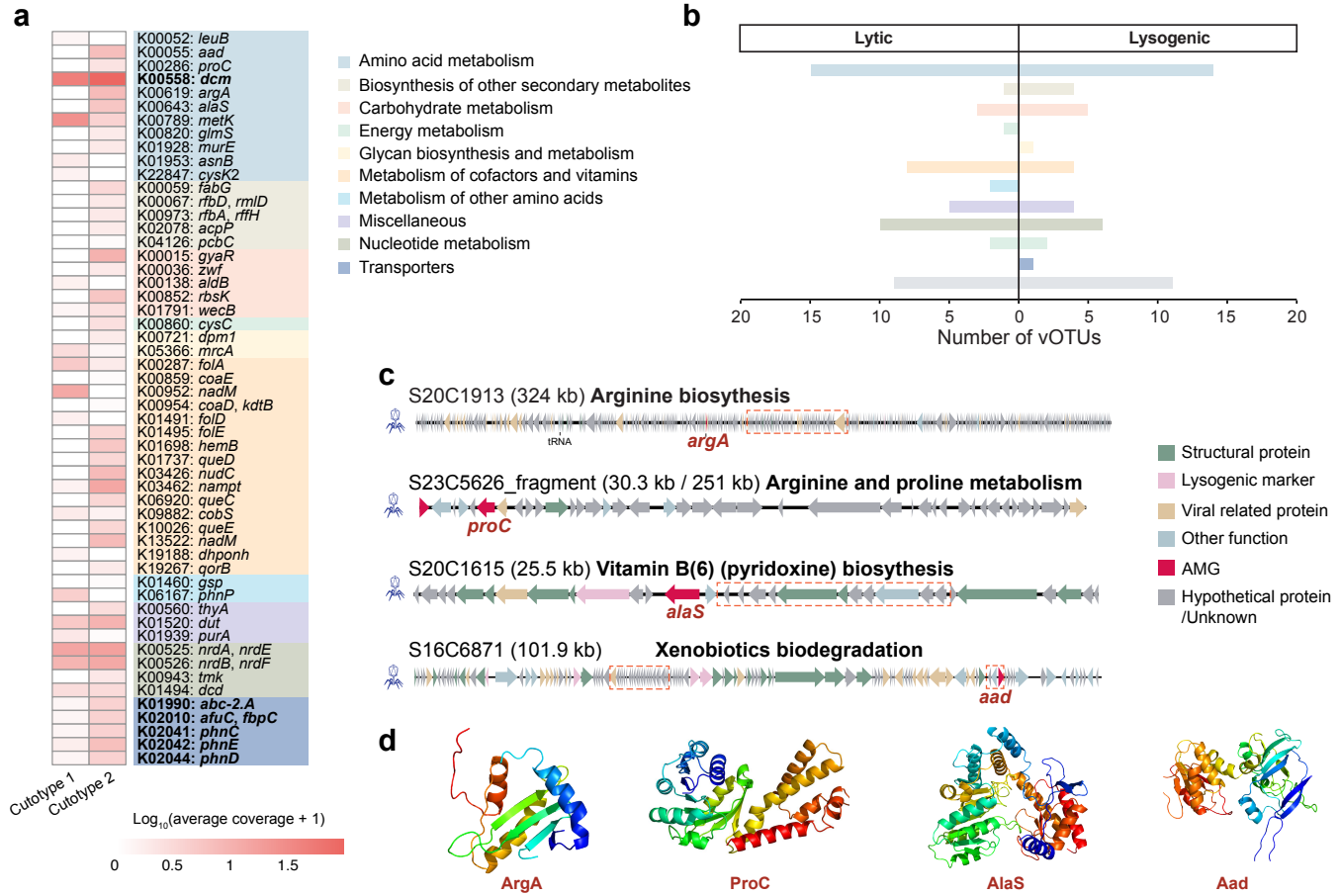


Fig. 5 Phage-encoded auxiliary metabolic genes (AMGs) in the two cutotypes. (a) Heatmap of the average coverage of each AMG and its Kyoto Encyclopedia of Genes and Genomes (KEGG) functional annotation. The color indicates the KEGG pathway level. Bold font indicates KEGG orthology (KO) with significant differences between cutotypes (Wilcoxon rank sum test, two-sided, $p_{adj} < 0.05$). **(b)** Number of viral operational taxonomic units carrying AMGs with a lysogenic or lytic lifestyle, with the KEGG function indicated. **(c)** Genomic context of four bacteriophage-encoded AMGs in viruses from cutotype 2 that may aid the bacterial host in adapting to high exposure to polyaromatic hydrocarbons. The AMGs are highlighted in red, and the matched fragment (with an identity ≥ 90%) between a virus and its linked host is indicated by a red dashed box. **(d)** Protein structures of the four bacteriophage-encoded AMGs. The structures are color-coded from the N- to the C- terminus.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable5.xlsx
- SupplementaryTable6.xlsx
- SupplementaryTable7.xlsx
- SupplementaryTable8.xlsx
- SupplementaryFigures19Mar.pdf