RESEARCH ARTICLE

The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome

Geoffrey D. Hannigan,[a] Jacquelyn S. Meisel,[a] Amanda S. Tyldsley,[a] Qi Zheng,[a] Brendan P. Hodkinson,[a] Adam J. SanMiguel,[a] Samuel Minot,[b] Frederic D. Bushman,[b] Elizabeth A. Grice[a]

Department of Dermatology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA[a]; Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA[b]

ABSTRACT **Viruses make up a major component of the human microbiota but are poorly understood in the skin, our primary barrier to the external environment. Viral communities have the potential to modulate states of cutaneous health and disease. Bacteriophages are known to influence the structure and function of microbial communities through predation and genetic exchange. Human viruses are associated with skin cancers and a multitude of cutaneous manifestations. Despite these important roles, little is known regarding the human skin virome and its interactions with the host microbiome. Here we evaluated the human cutaneous double-stranded DNA virome by metagenomic sequencing of DNA from purified virus-like particles (VLPs). In parallel, we employed metagenomic sequencing of the total skin microbiome to assess covariation and infer interactions with the virome. Samples were collected from 16 subjects at eight body sites over 1 month. In addition to the microenviroment, which is known to partition the bacterial and fungal microbiota, natural skin occlusion was strongly associated with skin virome community composition. Viral contigs were enriched for genes indicative of a temperate phage replication style and also maintained genes encoding potential antibiotic resistance and virulence factors. CRISPR spacers identified in the bacterial DNA sequences provided a record of phage predation and suggest a mechanism to explain spatial partitioning of skin phage communities. Finally, we modeled the structure of bacterial and phage communities together to reveal a complex microbial environment with a *Corynebacterium* hub. These results reveal the previously underappreciated diversity, encoded functions, and viral-microbial dynamic unique to the human skin virome.**

IMPORTANCE **To date, most cutaneous microbiome studies have focused on bacterial and fungal communities. Skin viral communities and their relationships with their hosts remain poorly understood despite their potential to modulate states of cutaneous health and disease. Previous studies employing whole-metagenome sequencing without purification for virus-like particles (VLPs) have provided some insight into the viral component of the skin microbiome but have not completely characterized these communities or analyzed interactions with the host microbiome. Here we present an optimized virus purification technique and corresponding analysis tools for gaining novel insights into the skin virome, including viral "dark matter," and its potential interactions with the host microbiome. The work presented here establishes a baseline of the healthy human skin virome and is a necessary foundation for future studies examining viral perturbations in skin health and disease.**

Received 17 September 2015 **Accepted** 23 September 2015 **Published** 20 October 2015

Citation Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodkinson BP, SanMiguel AJ, Minot S, Bushman FD, Grice EA. 2015. The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. mBio 6(5):e01578-15. doi:10.1128/mBio.01578-15.

Editor Joseph Heitman, Duke University

Address correspondence to Elizabeth A. Grice, egrice@upenn.edu.

The human skin is a barrier to the external environment and home to diverse and distinctive microbial communities. To date, most cutaneous microbiome studies have focused on bacterial and fungal communities, their modulation of cutaneous immune responses, and the association of these microorganisms with dermatological disorders (1). Recent metagenomic studies confirm the roles of the skin microenvironment and interpersonal variation in shaping the microbiome (2). Skin viral communities and their relationships with their hosts remain poorly understood, despite their potential to modulate states of cutaneous health and disease. Bacteriophages (phages, viruses that infect bacteria) can affect human health by altering the composition of their host bacterial communities through predation (3, 4). Evidence of such dynamism is provided by the acquisition and diversification of bacterial clustered regularly interspaced short palindromic repeat (CRISPR) elements (e.g., see reference 5) that target phage genomes for destruction by using nucleases guided by sequences encoded in the CRISPR arrays. Phages may also have long-term impacts on their hosts via lysogeny, in which phages integrate their genome into the host and adopt a quiescent state. New genes carried by lysogens can affect host metabolism, virulence, antibiotic resistance, and sensitivity to other phages (6–9). Phages may also serve as a genetic reservoir for bacterial adaptations during stress (i.e., antibiotic treatment) (10). Viruses that replicate on

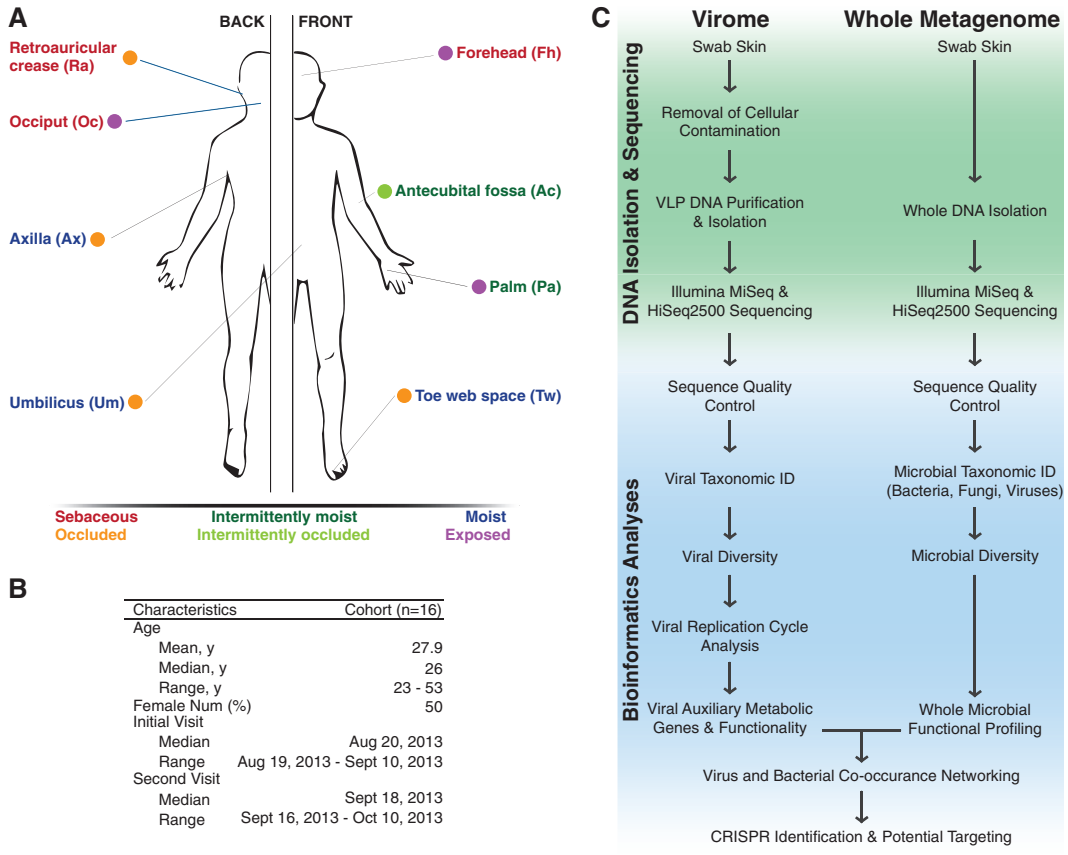


FIG 1 Study design for the analysis of cutaneous viral and whole metagenomic communities. (A) Eight skin sites of 16 subjects were sampled. Colored text indicates the microenvironment classification, and each colored ball represents the occlusion status of the anatomical site. (B) Characteristics of the cohort sampled. (C) Flowchart illustrating the procedures by which DNA was isolated from cutaneous swabs and sequenced for downstream bioinformatic analyses.

human cells are also present in the skin and can affect human health, including human papillomaviruses (HPVs), human polyomaviruses (HPyVs), and human herpesviruses, and can cause skin cancers and other dermatological disorders.

Previous studies employing whole-metagenome sequencing without purification for virus-like particles (VLPs) have provided some insight into the viral component of the skin microbiome but have not completely characterized these communities or analyzed interactions with the host microbiome (2, 11, 12). The study we present here employed techniques for the purification of viral DNA, thereby reducing contamination from human and bacterial cells, whose genomes are orders of magnitude longer than viral genomes. This allows for deeper viral sequencing and the use of reference-independent analyses to capture the impact of unknown or uncharacterized genomes, known as viral dark matter (13). We applied shotgun metagenomic analysis to purified VLPs, as well as unpurified whole skin microbial communities, conducting the first longitudinal, integrated analysis of the healthy human skin virome and the whole metagenome across diverse anatomical locations. The major questions we address with this novel data set are as follows. What are the biogeography and diversity of the human skin virome compared to those of the whole metagenome over time and across individuals? What genetic functions are encoded by the skin virome, including antibiotic resistance, virulence factors (VFs), and auxiliary metabolic genes (AMGs; host genes within phage genomes [14])? What can we infer about interactions between phages and their bacterial hosts, including the role of CRISPRs in maintaining virome community structure?

RESULTS

Sampling, sequencing, and quality control. Cutaneous skin swabs were collected from 16 healthy volunteers with no known skin conditions between the ages of 23 and 53 years (Fig. 1A and B). Anatomical skin sites were sampled bilaterally (the virome sample was collected at the site contralateral to the whole-metagenome sample) and consisted of multiple diverse microenvironments: sebaceous (retroauricular crease [Ra], occiput [Oc], and forehead [Fh]), moist (axilla [Ax], toe web [Tw], and umbilicus [Um]), and intermittently moist (antecubital fossa [Ac] and palm [Pa]) (Fig. 1A). Swab samples were collected at two time points separated by 4 weeks to assess the stability of the communities.

After swabbing each subject's skin, we used one sample of the contralateral pair to purify and extract the VLP DNA by using a protocol established for human and environmental viromes (15–17). We extracted the DNA from the contralateral sample to investigate the whole microbial community, including bacterial, fungal, and viral members. Samples were prepared for shotgun sequencing on the Illumina MiSeq and HiSeq2500 platforms by using the Illumina NexteraXT library preparation kit, which is

designed for double-stranded DNA (dsDNA). Therefore, our analysis focused on dsDNA viruses and replicative intermediates of single-stranded DNA (ssDNA) viruses. Sample collection, sequence processing, and bioinformatic analyses are outlined in Fig. 1C.

After quality filtering, the dsDNA virus data set contained a total of 260,714,906 high-quality sequence reads with a median of 650,506 sequence reads per sample. The whole metagenome data set contained a total of 368,341,329 high-quality sequence reads with a median of 981,031 sequence reads per sample (see Fig. S1A to D; see Table S1 in the supplemental material for sequence count statistics). Consistent with previous reports of similar human VLP preparations (16–19), a relaxed search against the entire NCBI nonredundant database revealed that 94.8% of the VLP reads did not significantly match a known genome (BLASTn E value, $<10^{-3}$), highlighting the importance of investigating viral dark matter. Similar classification identified 42.6% of the whole-metagenome reads as unknown. In this study, we used multiple reference-independent approaches to address this subset of unclassified dark matter. The viral and whole metagenome data sets were independently assembled into contigs, and contigs >500 bp in length were selected for further analysis (see Fig. S1E to H; see Table S2 for contig coverage, count, and length statistics). Of these phage contigs, 9.0% were taxonomically identifiable, highlighting the utility of using contigs in taxonomy instead of using unaligned reads (20).

During each sampling event, we collected a blank negative control that never came into contact with skin. DNA was extracted from the control and sequenced in parallel with the experimental samples. Sequences identified in the negative controls were subtracted *in silico* from the experimental samples (see Text S1 in the supplemental material for details). Using the Bray-Curtis dissimilarity metric, we found significant separation of the control samples from the skin samples (see Fig. S2A in the supplemental material), confirming minimal identity between the control and experimental samples and providing confidence that the viruses present are not the result of environmental or reagent contamination. As an additional control, we sequenced an even mock community sample. The observed community composition was significantly correlated to the expected community composition at the genus level (Spearman correlation rho = 0.6; $P = 0.01$), suggesting that our library preparation and sequencing techniques sufficiently depict the microbial community composition (see Fig. S2B).

Using methods previously outlined for quantifying virome contamination (21), we verified the reduction in cellular contamination within viromes by showing a significant reduction in normalized bacterial 16S rRNA gene levels in the purified viromes compared to the unpurified whole metagenomes (see Fig. S2C). We also supported virome purity by using a previously described method (16) to map significantly more sequences from the virome to the whole metagenome rather than the reverse (see Fig. S2D). Finally, we confirmed significantly less contamination from human cells in the virome than in the whole metagenome (see Fig. S2E). These analyses suggest that viral reads are in greater abundance after VLP purification and reinforce the utility of VLP purification techniques.

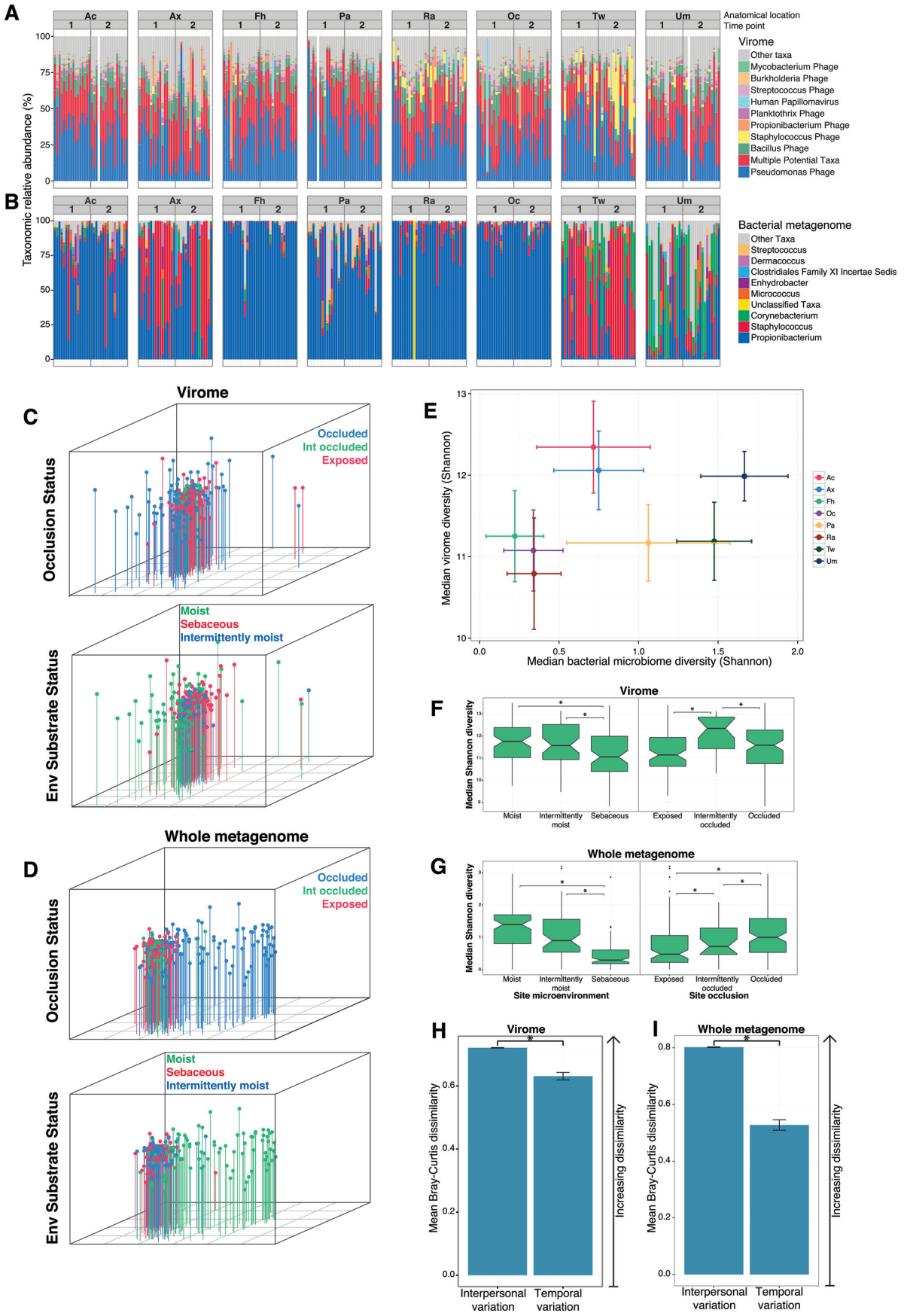
Skin virome composition. To examine the community membership of the skin virome, we used the viral UniProt TrEML reference database to annotate predicted open reading frames (ORFs) in the assembled viral contigs. Annotated ORFs were then subjected to a voting system that assigned taxonomy based on the most abundant ORF annotation within the contig, as described previously (22). Some contigs had ties in taxonomic votes, which were labeled as having "multiple hits" because they are not unknown, but we cannot assign a resolved viral taxonomy with confidence. The abundance of each taxonomically identified contig was quantified as the number of unassembled reads that aligned with the contig. Read counts were normalized in order to account for the differences in contig length, sequencing efficiency, and associated run variation of that overall sample by using methods previously described (22).

Most of the dsDNA viral contigs identified belonged to the *Caudovirales* order (tailed bacteriophages), suggesting a larger proportion of bacteriophages among skin dsDNA virus communities than previously suggested (11) (see Fig. S3A in the supplemental material). Most of the viruses were unclassifiable at the family level, but we could identify some phage families, including *Myoviridae* and *Siphoviridae* (see Fig. S3B). Interestingly, members of the family *Papillomaviridae* were most abundant on the palm, which is a region known to be afflicted by cutaneous warts. We also observed members of the family *Poxviridae*, which were observed as major virus taxa in a related skin metagenomic survey (23). It is important to note that while many viruses are not identifiable at the family level, they are often identifiable at the species level, as is the case with many orphan *Staphylococcus* phages (24).

At the species level, we observed bacteriophages of known skin inhabitants such as *Propionibacterium* phages and *Staphylococcus* phages (Fig. 2A), and their relative abundances were significantly variable across different skin microenvironments (see Fig. S3C and D; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests) and occlusion statuses (see Fig. S3F and G; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests). A large fraction of each virome contained contigs that maintained equal similarity to multiple phages, meaning they were not assignable to a single species and were therefore annotated as "multiple hits" (Fig. 2A). This is likely a reflection of the modular nature of bacteriophage genomes and highlights the need for more robust reference databases for a better understanding of phage genome architecture. There was also an abundant representation of environmental phages, including *Pseudomonas* and *Bacillus* phages.

The most abundant recognized metazoan virus was HPV, which was prominent in some individuals and generally present in significantly greater relative abundance at sebaceous and exposed sites (see Fig. S3E and H; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests). HPyVs were detected in very low abundance, where only six samples contained any sequence mapping to known HPyV genomes, and no sample had >100 putative HPyV sequences.

Skin total microbial community composition. In addition to examining the taxonomic composition of the virome, we further characterized the membership of the whole microbial skin community by using the corresponding sample set that was not subjected to VLP or microbial selection. Bacterial communities were classified from the unassembled sequences by using MetaPhlAn (25, 26), which annotates sequences on the basis of clade-specific markers from reference genomes. Additionally, bacterial, fungal, and viral species abundances were quantified from assembled contigs by using the lowest common ancestor algorithm in


A
Ac
Ax
Fh
Pa
Ra
Oc
Tw
Um
1
2
Taxonomic relative abundance (%)
100
75
50
25
0
Anatomical location
Time point
Virome
Other taxa
Mycobacterium Phage
Burkholderia Phage
Streptococcus Phage
Human Papillomavirus
Planktothrix Phage
Propionibacterium Phage
Staphylococcus Phage
Bacillus Phage
Multiple Potential Taxa
Pseudomonas Phage
B
Bacterial metagenome
Other Taxa
Streptococcus
Dermacoccus
Clostridiales Family XI Incertae Sedis
Enhydrobacter
Micrococcus
Unclassified Taxa
Corynebacterium
Staphylococcus
Propionibacterium
C
Virome
Occlusion Status
Occluded
Int occluded
Exposed
Env Substrate Status
Moist
Sebaceous
Intermittently moist
D
Whole metagenome
Occlusion Status
Occluded
Int occluded
Exposed
Env Substrate Status
Moist
Sebaceous
Intermittently moist
E
Median virome diversity (Shannon)
13
12
11
10
0.0
0.5
1.0
1.5
2.0
Median bacterial microbiome diversity (Shannon)
Ac
Ax
Fh
Oc
Pa
Ra
Tw
Um
F
Virome
Median Shannon diversity
Moist
Intermittently moist
Sebaceous
Exposed
Intermittently occluded
Occluded
G
Whole metagenome
Median Shannon diversity
Moist
Intermittently moist
Sebaceous
Site microenvironment
Exposed
Intermittently occluded
Occluded
Site occlusion
H
Virome
Mean Bray-Curtis dissimilarity
0.6
0.4
0.2
0.0
Increasing dissimilarity
Interpersonal variation
Temporal variation
I
Whole metagenome
Mean Bray-Curtis dissimilarity
0.8
0.6
0.4
0.2
0.0
Increasing dissimilarity
Interpersonal variation
Temporal variation

MEGAN5 (27). Consistent with previous whole-metagenome analyses of skin (2, 28), *Propionibacterium* (including *Propionibacterium acnes*), *Staphylococcus* (including *Staphylococcus epidermidis* and *S. hominis*), and *Corynebacterium* were the dominant bacterial genera (Fig. 2B; see Fig. S4A and B in the supplemental material) and *Malassezia* was the most abundant fungal genus (see Fig. S4A and C). Viruses were present in low abundance (average, 0.4% per sample), likely because of the relatively small genome size of viruses compared to that of prokaryotes and microeukaryotes, and this further highlights the utility of VLP isolation before sequencing (see Fig. S4A and D). The viruses recovered were primarily "unclassified" and *Staphylococcus* phages (see Fig. S4D).

Variation of the skin virome and total metagenome among anatomic sites. As demonstrated above and extensively in the previous literature (16–19), most of the viruses were taxonomically unidentifiable because of insufficient reference database information. In order to capture information from both characterized and uncharacterized genomes, we employed reference-independent approaches based on the relative abundance of each contig in our data set. To assess the beta diversity (diversity between samples) among anatomical sites, we calculated the Bray-Curtis dissimilarities between communities at the same and different anatomical sites. We identified significant differences in virome and whole-metagenome community structures based on microenvironment and occlusion status (Fig. 2C and D; $P < 0.001$, Adonis test). These findings parallel previous reports of the bacterial and fungal skin microbiomes (29, 30) and highlight an additional role for occlusion/exposure parameters in microbial community structure and function.

We further estimated and compared the alpha (within-sample) diversities of viral communities by using a reference-independent approach to calculate the Shannon diversity index. Here we estimated virome diversity, including the viral dark matter, by using the PHACCS toolkit (31), which calculates the degree of contig assembly to generate a "contig spectrum" that is compared to simulated communities varying in size and diversity until a suitable match is found. PHACCS predicts the virome size and diversity as if the entire community (both known and unknown viruses) were sequenced and annotated. The Shannon diversity of bacterial communities among anatomical sites was calculated on the basis of the reference-dependent taxonomic relative abundance information described above. We found that the virome and bacterial metagenome of sebaceous sites was less diverse than that of moist or intermittently moist sites (Fig. 2E to G; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests). While the virome was most diverse at intermittently occluded sites (e.g., Ac), the bacterial metagenome was most diverse at occluded sites (e.g., Tw and Um; Fig. 2E to G; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests), further highlighting the differences in viral and bacterial community diversity based on anatomic sites.

To assess the utility of reference-independent methods in determining differences in viral diversity, including that of the viral dark matter, we performed the above-described alpha and beta diversity analyses by using the reference-dependent taxonomic relative abundance information from Fig. 2A. The alpha diversity of the reference-dependent data set (see Fig. S5A and B in the supplemental material) was strikingly less than that predicted by the reference-independent methods employed by PHACCS (Fig. 2F). In contrast to the PHACCS-based analysis, there was no significant difference between the microenvironment or occlusion categories when using the reference-dependent data. Beta diversity between sites of different microenvironment and occlusion statuses mirrored the reference-independent findings (Fig. 2C; see Fig. S5C and D). Therefore, there is added value to using viral dark matter in some community analyses, but some metrics can be performed effectively by reference-based approaches.

Variation of the skin virome and whole metagenome over time. Previous studies suggest that temporal variation of the bacterial microbiome at a given skin site is minimal compared to interpersonal variability (30, 32, 33), so we examined both viral and whole microbial community changes over a 1-month period. There was a significant difference between the two time points in the shared diversity of the viromes, but not the whole metagenomes, as measured by Bray-Curtis dissimilarity (see Fig. S6A and B in the supplemental material; $P < 0.001$ and $P = 0.978$, respectively, Adonis test). These findings suggest that the whole metagenome is more stable than viral communities over time.

Using the same metric, virome temporal variability at a given skin site was significantly lower than interpersonal variability (Fig. 2H; $P = 1.26 \times 10^{-11}$, *t* test), similar to what we observed for the whole metagenome (Fig. 2I; $P = 3.50 \times 10^{-30}$, *t* test). Analogous to human fecal viromes, the largest source of skin virome variance appears to be interpersonal variation (16, 17). In contrast to the gut, which has been suggested to share >80% of the intrapersonal virome over time (16, 17), we found that less than 50% of the intrapersonal skin virome was shared over time (see Fig. S6C).

Evidence of a temperate replication style. Bacteriophages can exist as lytic or temperate phages. Lytic phages lyse the host soon after infection and do not exist in a latent, lysogenic state. Conversely, temperate phages are able to integrate their genomes into the bacterial host genome and exist as prophages, as well as excise and go through the lytic cycle. To examine the replication strategies of the phages residing on the skin, we used an established

FIG 2 Taxonomy and diversity of cutaneous viral and bacterial metagenomic communities. (A and B) Taxonomic relative abundance of the viral (A) and bacterial (B) communities by site over time. The viral relative abundance plots show the 10 most abundant taxa according to virus TrEMBL annotated contigs. The bacterial communities show the 10 most abundant taxa according to MetaPhlAn analysis. Each bar represents a single sample from a subject, and the bars are separated by time point and anatomical location as indicated at the top. (C and D) Nonmetric multidimensional scaling (NMDS) ordination plots of Bray-Curtis dissimilarities between virome (C) and whole-metagenome (D) samples, showing significant clustering ($P < 0.001$, Adonis test) by occlusion status and environmental substrate. (E) Alpha diversity (Shannon diversity metric) of the virome and bacterial metagenome for each anatomical site. The *x* axis represents median bacterial metagenome diversity, and the *y* axis represents median virome diversity. Each point is the median diversity of the two communities, and error bars indicate the population notch deviation of the median. (F and G) Viral (F) and microbial (G) Shannon diversity is presented by site microenvironment and occlusion, with asterisks indicating statistical significance ($P < 0.05$) by the Kruskal-Wallis and multiple-comparison *post hoc* tests. Box plots were calculated with the ggplot2 R package. (H and I) Intrapersonal variance compared to temporal variance of the virome (D) and the whole metagenome (E) as calculated by the mean (± the standard error of the mean) Bray-Curtis dissimilarity metric. A higher value indicates higher dissimilarity. An asterisk indicates statistical significance ($P < 1.0^{-10}$).

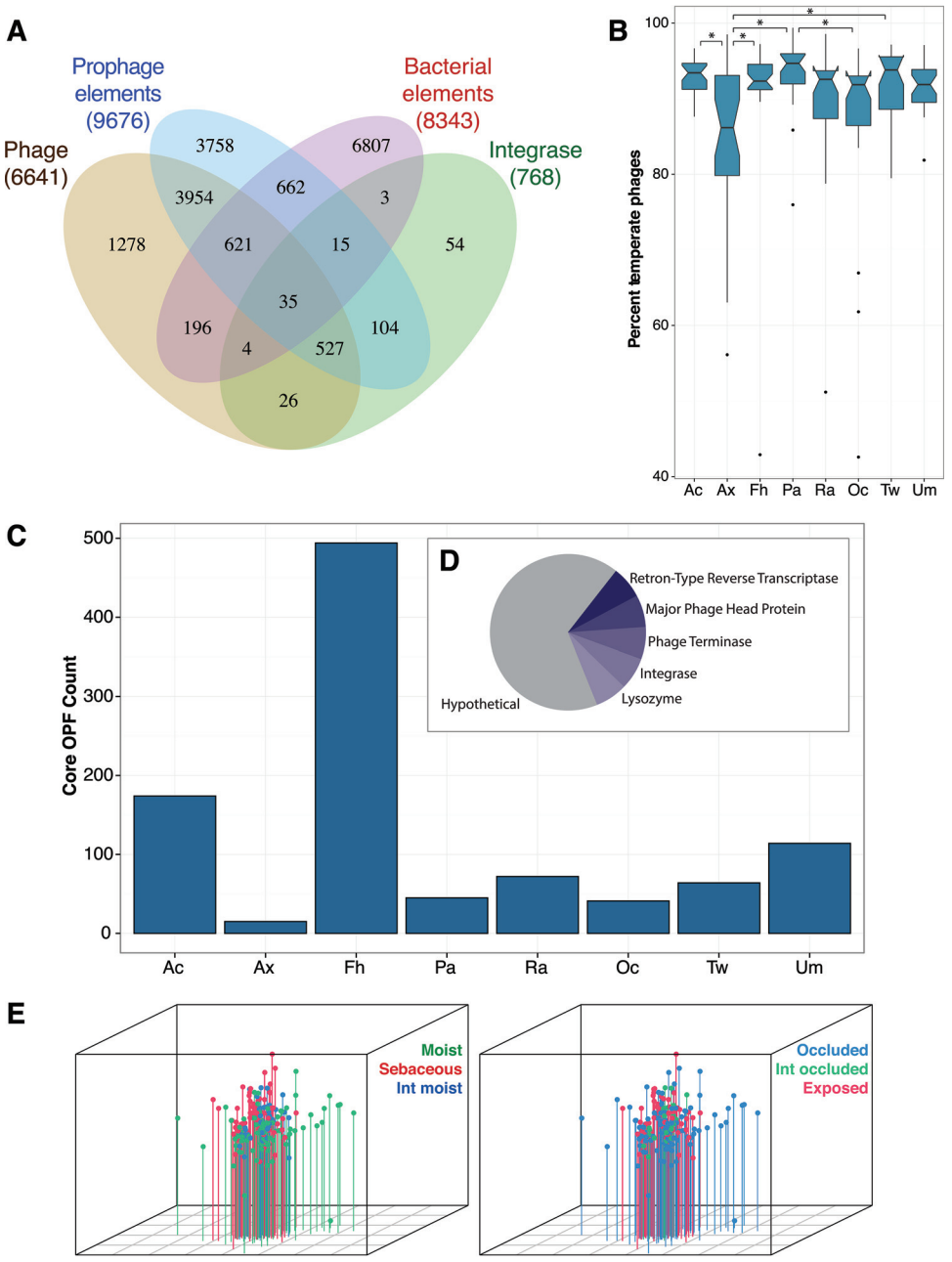


FIG 3 Replication cycle and functional enrichment of bacteriophages on the skin. (A) Euler diagram of the phage contigs (yellow) that also contain an integrase gene (green), at least one prophage element per 10 kb (blue), homology to a known bacterial genome (red), or a combination of these markers. (B) Box plot illustrating the percent relative abundances of predicted temperate phages per body site. Temperate phage contigs were defined as those that contained both a phage gene at least every 10 kb and one of the other three temperate markers. Relative abundance was calculated as the relative number of reads per kilobase of transcript per million mapped unassembled reads that mapped back to the assembled contigs. An asterisk indicates statistical significance at $P < 0.05$ by the Kruskal-Wallis and multiple-comparison *post hoc* tests. (C) The distribution of exclusive OPFs associated with each anatomical site. (D) The distribution and UniProt annotation of the 15 core OPFs found across the entire virome. (E) Bray-Curtis dissimilarity of the virome samples by OPF relative abundance. Clustering was statistically significant ($P < 0.001$) by the Adonis test for both environmental substrate and occlusion.

approach (17) of searching VLP contigs for temperate phage replication markers, including (i) the presence of integrase genes (including members of both the serine and tyrosine integrase families), (ii) the presence of temperate prophage genes, including *parABS* partitioning systems, and (iii) nucleotide sequence identity to bacterial genomes indicative of integration. Of the 6,661 contigs that were annotated as bacteriophages by our taxonomic criteria described above, 5,363 had at least one of these three temperate phage markers (Fig. 3A). More specifically, 592 (8.8%) contained at least a single integrase gene, as represented in the UniProt TrEMBL database; 856 (12.9%) aligned with known bacterial genomes, including *Actinobacteria*, *Firmicutes*, and *Proteobacteria*; and 5,137 (77.1%) contained ORFs similar to annotated prophage genes found in the ACLAME database of mobile genetic

elements (34). By these measures, each anatomical skin site had a median relative proportion of >85% temperate phages, with different relative abundances by site (Fig. 3B; $P < 0.05$, Kruskal-Wallis and multiple-comparison *post hoc* tests). This finding suggests that the majority of identifiable *Caudovirales* bacteriophages on the skin are temperate, consistent with studies of the human gut virome (16, 17).

Virome functional potential and AMGs. Though our data support a lesser role for host lysis in skin dsDNA bacteriophage populations, they likely influence bacterial communities via prophage integration and genetic exchange. We therefore investigated the genetic functional potential of skin viral communities compared to that of the whole metagenome. Functional pathways were interrogated by comparison to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (35) and analyzed by using the HUMAnN annotation and quantification program (36). Overall, the virome was enriched in information processing and peptide transport genes, while the whole metagenome was enriched for metabolic process genes (see Fig. S7A in the supplemental material). Gene ontology (GO) analysis revealed significant enrichment of genes for viral components and processes, DNA transcription, and RNA metabolic processes in the virome (see Fig. S7B), while the whole metagenome was enriched in genes for cellular nitrogen compound and carbohydrate derivative metabolic processes. Notably, the virome was significantly enriched in the GO term "establishment of viral latency" (see Fig. S7B), consistent with the observed dominance of temperate phages on the skin.

Some bacteriophages are known to encode AMGs (host genes within phage genomes) that promote viral infection by modulating host metabolic activity (reviewed in reference 14). We evaluated whether there were core AMGs conserved across the entire skin virome, thereby belonging to the overall core gene set. To accomplish this, we clustered the predicted virome contig ORFs into representative operational taxonomic unit-like sequences called operational protein families (OPFs) (9, 37). Core OPFs were defined as those OPFs that were present in all of the samples from a skin site. Core OPFs were differentially distributed across skin sites, with the greatest amount present on the forehead (Fig. 3C). Of the 15 core OPFs present in all of the virome samples, all were hypothetical or known phage genes and none were AMG candidates (Fig. 3D), suggesting a sparse population of core skin virome AMGs. As highlighted above, in comparison to the metagenome, the virome was enriched in genes for KEGG pathways related to transport (see Fig. S7B), as well as GO terms associated with regulation of RNA metabolic processes (GOEAST, $P < 0.05$). This indicates that potential AMGs, while not strictly belonging to a "core" set of genes, are present throughout the skin virome. We also investigated the distribution of OPFs with respect to the skin site microenvironment and occlusion and found significant differences (Bray-Curtis dissimilarity; $P < 0.001$, Adonis test), suggesting differential spatial distribution of virome functional potential (Fig. 3E).

Antibiotic resistance and VF enrichment. Because phages may alter the phenotypes of their hosts by conferring novel virulence and pathogenicity functions, we investigated the potential for antibiotic resistance and bacterial virulence encoded within the skin virome. Using blast algorithm parameters specified in previous foundational human virome studies (17, 38), we assessed antibiotic resistance potential by comparing ORFs from the assembled virome contigs to the Comprehensive Antibiotic Resistance Database (CARD) (39) (BLASTx E value, $<10^{-5}$). To further increase our confidence in the annotations beyond that of past studies, we filtered the BLASTx hits to keep only those with >75% identity. Viromes contained 29 unique antibiotic resistance gene (ARG) groups, which were related to antibiotic efflux and resistance to beta-lactamases, rifampin, tetracycline, and elfamycin (Fig. 4A). Tetracyclines are commonly used to treat dermatological conditions such as acne, and elfamycins are naturally occurring antibiotics with strong activity against *P. acnes* (40). To confirm that the ARGs identified are associated with the virome and not cellular contamination or artifacts, we demonstrated that ~50% of the ARGs are colocalized on contigs with other annotated phage genes or are themselves known phage-associated ARGs (Fig. 4B). ARGs were associated primarily with "multiple-hit," *Bacillus*, and *Streptococcus* phages (Fig. 4B). We also identified potential VFs associated with the skin virome by using the VF database (VFDB) (41) with the same BLASTx parameters and filtering as described for antibiotic resistance analysis above. We identified 122 unique VF genes, and >1/3 of the VF contigs were either known phage-associated genes or colocalized with phage genes (Fig. 4C). These findings together indicate that bacteriophages of the skin microbiome may be a significant source of transmissible genes associated with antibiotic resistance, virulence, and pathogenicity.

Inference of phage-bacterium interactions: co-occurrence network analysis. To predict phage-bacterium interactions of the skin, we constructed a correlation network from relative abundances of bacteria and known phages as previously described (42) (Fig. 5A). Positive interactions indicate that the bacteria and phage typically co-occur, while negative interactions suggest a mutually exclusive relationship between the bacterium and phage relative abundances. The resulting network of significant phage-bacterium interactions contained 21 nodes, 7 bacteria, and 14 phages. *Propionibacterium* and *Staphylococcus* bacteria were typically copresent with their phage counterparts, *Propionibacterium* and *Staphylococcus* phages, respectively (Fig. 5A). The overall co-occurrence structure suggests that the network is nonrandom, exhibiting scale-free properties such as short average path lengths (characteristic path length, 2.781) and a node degree distribution that approximately fits a power law ($R^2 = 0.781$) (43). Short average path lengths suggest that the skin phage-bacterium community network is able to respond rapidly to perturbations (44). The heterogeneity value (likelihood of uneven distribution of edges) of the network was 0.819, suggesting that there are fewer hubs and indicating the presence of potential "keystone" taxa in the network (45).

Hubs may be distinguished by identifying nodes of high degree. In the skin bacterium-phage network, *Corynebacterium*, with a degree of 10, had the greatest number of interactions, while all of the other nodes had degrees of ≤5. *Corynebacterium* positively associated with eight phages, including *Corynebacterium* and *Staphylococcus* phages, and negatively associated with two phages, including *Propionibacterium* phage (Fig. 5A). These features of the network topology suggest that the skin bacterium-phage network is able to rapidly respond to perturbations, and *Corynebacterium* may act as a key hub.

Inference of phage-bacterium interactions: CRISPRs. CRISPRs are a form of bacterial adaptive immunity against phage predators. Spacer sequences, generally 26 to 72 nucleotides in

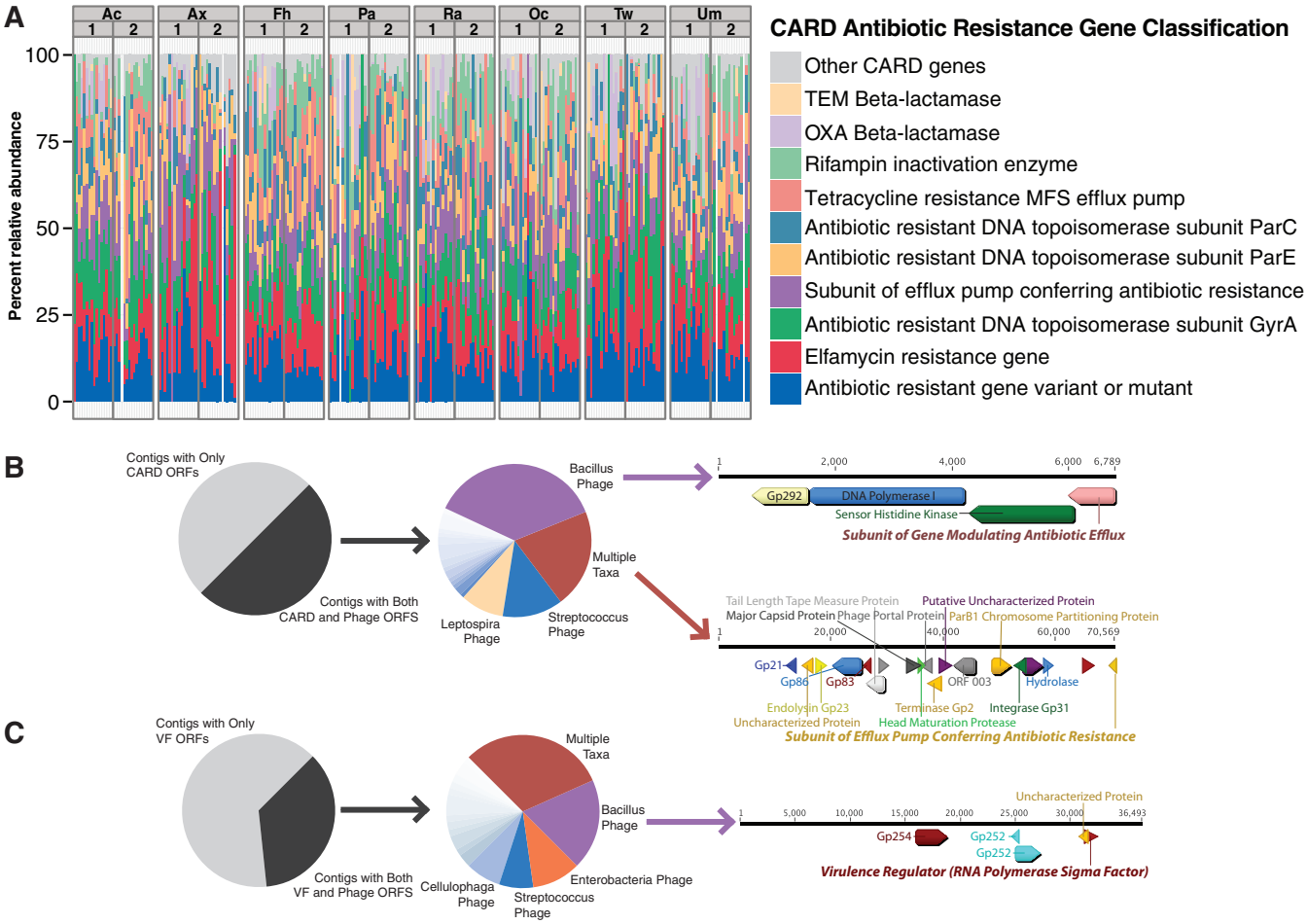


FIG 4 Antibiotic resistance and bacterial virulence in the skin virome. (A) Relative abundances of predicted ARGs according to the CARD. Each bar represents a subject, and the bars are separated by time point and anatomical location as indicated at the top. (B) Flow diagram of the ARGs associated with bacteriophage contigs. The leftmost part shows the proportions of ARGs that colocalize on contigs with other phage genes or are themselves known phage-associated genes. The middle part shows the distribution of phage taxa that contain predicted ARGs. The rightmost part shows two annotated examples of ARGs colocalized on phage contigs, with the CARD-predicted ARGs in bold italics. (C) Similar to panel B, a flow diagram of the VFs associated with phages. As in panel B, the leftmost part shows the distribution of predicted VFs associated with phages, the middle part shows the taxonomic distribution of those phages, and the rightmost part shows an annotated example.

length, are captured from invading phages and integrated into the bacterial chromosome. These spacer sequences provide a genomic record of phage predators encountered by the bacteria. We detected a total of 477 unique spacer sequences, identified by 68 unique CRISPR repeats in the whole metagenomic data set. Only 18 spacers aligned with VLP contigs (Fig. 5B). These spacers were found in 21 metagenomic contigs and mapped to 40 unique VLP contigs. Spacers found in the Um aligned only with Um VLP contigs. Two *Staphylococcus* spacers detected in the Ax aligned with 16 different VLP contigs that were found at every body site except the Pa (Fig. 5B). A *Propionibacterium* spacer found in both the Pa and Tw aligned with eight different VLP contigs from the Ax, Oc, Fh, and Ra (Fig. 5B). These findings indicate that phage-host dynamics may not be restricted by anatomical skin site, and spacers identified at one skin site may be restricting phage during invasions from other skin sites, which could, in part, explain the spatial partitioning of the skin virome. We further characterized the genomic CRISPR targets within the VLP contigs and found that the majority of the targets within coding regions belonged to phage portal proteins, which are genes involved in the packaging of DNA into phage particles (Fig. 5C). It is unclear whether this is an artifact of the low sampling of CRISPR spacers (approximately 12 spacers were annotated) or a biological phenomenon. Further work is required to understand this to a greater extent. Interestingly, the majority of CRISPR targets did not map to predicted ORFs, suggesting that there is not a targeting preference for genomic coding regions (Fig. 5C).

DISCUSSION

In summary, we present parallel analyses of the human skin virome (as determined from purified VLPs) and the whole metagenome. Purification of VLPs provides many advantages for virome-targeted analyses, including deeper sequencing of viruses and the ability to confidently assess viral dark matter by using reference-dependent and -independent approaches. However, this technique has previously been technically prohibitive for application to skin viruses because of the small microbial burdens in and on the skin. Advanced library preparation techniques utiliz-

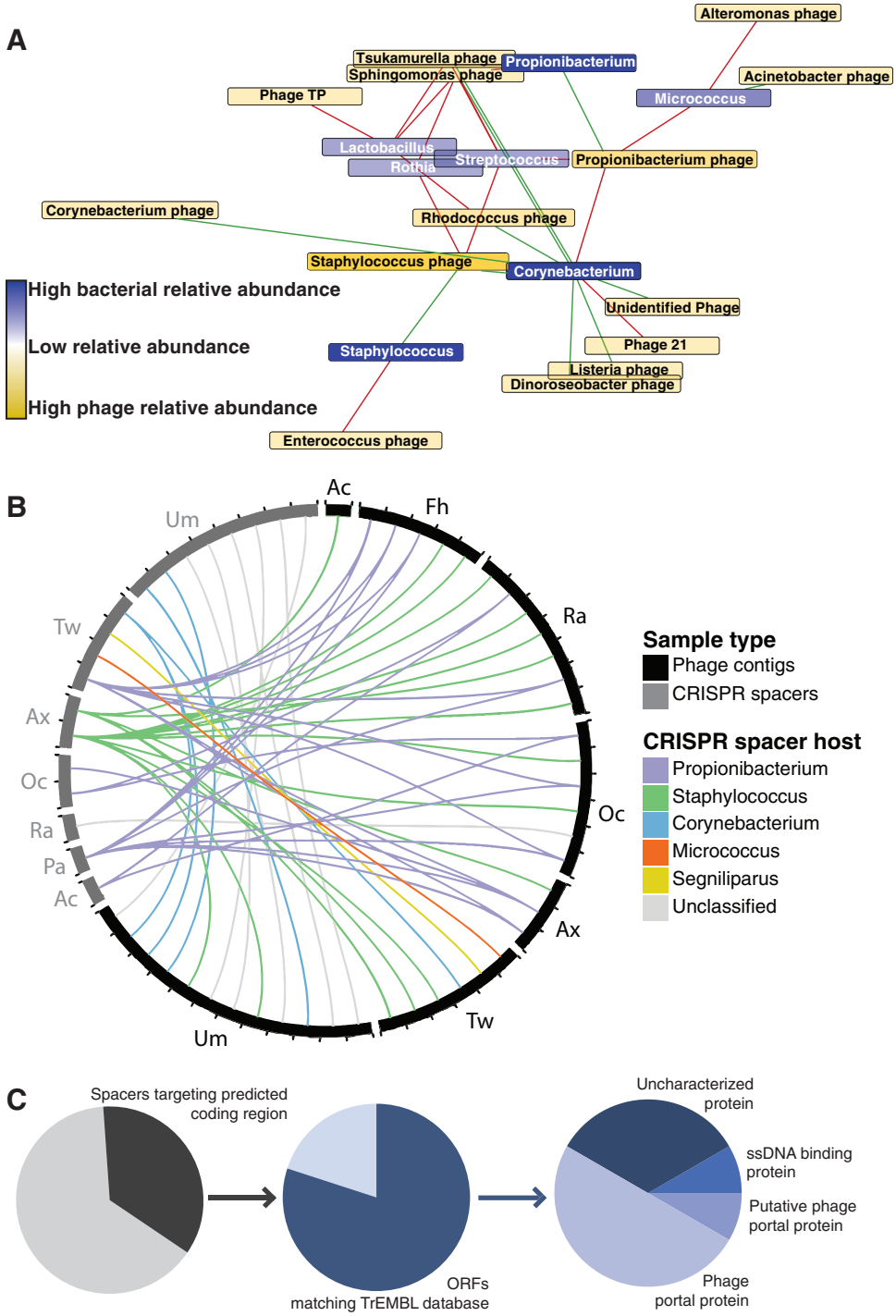


FIG 5 Modeled bacteriophage-host co-occurrence associations and CRISPR targets within the skin virome. (A) Network analysis of the correlations between bacteriophages of the virome and bacteria of the whole metagenome. Bacteriophages are represented by yellow boxes, while the bacterial genera are represented by blue boxes. The color intensity indicates the overall relative abundance of the taxon. The red lines represent a negative correlation, and the green lines represent a positive correlation. (B) Radial table showing bacterial CRISPR spacers (grey) that target viral phage contigs (black). The line colors represent the CRISPR spacer bacterial hosts. (C) Flow chart depicting the phage genome regions targeted by skin bacterial CRISPRs. The leftmost part shows the abundance of spacers that target a predicted coding region (ORF) within the phage genomes. The middle part is the distribution of ORFs matching a gene in the TrEMBL reference database. The rightmost part is the distribution of annotated coding region CRISPR targets.

ing ultrasmall amounts (<1 ng) of DNA have facilitated this study to characterize the human skin dsDNA virome in parallel with the whole metagenome in order to gain insight into multikingdom interactions of the skin microbiome.

Our results demonstrate that the skin virome is highly site specific and is modulated by occlusion and exposure, in addition to sebum and moisture. This significant effect of skin occlusion on viral and whole microbial communities has not yet been described

in previous skin whole microbial analyses and provides new insight into the variation of these communities across anatomical sites. Anatomical intrapersonal and interpersonal variations play a greater role in cutaneous viral community composition than intrapersonal temporal variation does, supporting the role for persistent commensal populations, rather than a dominance of new acquisition of different transient viruses from the environment.

The persistence of phage populations on the skin, and especially dsDNA phages, is possibly due in part to the temperate nature of their infections. While cutaneous phages that are primarily temperate may not exhibit a predator-prey dynamic with their hosts, they may give rise to novel bacterial strains via transfer of genes, including antibiotic resistance and VF genes, which were found in our samples. The dynamics of phage predator-prey relationships within communities is complex, and while our study provides a first look into these community dynamics in the skin, further studies are needed to more completely characterize these relationships.

Although we noted that the majority of the identifiable phages in the skin virome sampled were temperate, we were only able to predict the replication styles of the identifiable phages. This highlights the need for robust reference databases and the utility of reference-independent methods. Additionally, we were not able to detect ssDNA viruses or enveloped viruses. Because of our efforts to confirm a reduction of bacterial genomic DNA in our samples, we are confident that the majority of the sequences are, in fact, from free phages, and provide a valuable description of our identifiable virome library.

In addition to showing complex community dynamics within the skin viral communities, we also provided evidence of potential interactions between the virome and the other microbial communities by using co-occurrence network modeling and CRISPR identification techniques. Our network analysis allowed us to infer an extensive and multikingdom ecosystem structure. Understanding these ecological interactions and experimentally validating them will be critical for further developing targeted therapeutics such as phage therapy.

CRISPR analysis suggested differing degrees of ongoing phage infections at different sites or simply differential abundances of CRISPR arrays in the resident bacteria. CRISPRs not only targeted phages found at the same skin sites but also targeted phages at other skin sites, providing a record of successfully repelled attacks from phages now detected at other body sites. These findings suggest a potential mechanism for partitioning of the skin virome between different anatomical locations and warrants further investigation. While we focused on CRISPR mechanisms of interaction, there are other mechanisms of bacterium-phage interactions that are worth investigating in future studies such as restriction modifications.

A limitation to note in this study, and virome studies in general, is the bias within the reference databases used. We identified phages by their host bacteria, but the numbers of known phages that infect bacteria differ greatly between hosts and the genomic diversity within these phages also varies. For example, *Propionibacterium* phages exhibit limited genomic diversity (46) and thus there is a higher likelihood of identifying a sequence match to one of these phages if it is present. *Mycobacterium* and *Staphylococcus* phages are more diverse, and reference-dependent methods will miss phages that diverge heavily from known sequences. We attempted to minimize the impact of these variations on our relative abundance and diversity calculations by using both reference-dependent and reference-independent methods.

Phage genomes are highly mosaic, and when using *de novo* contig assembly methods, two (or more) genomic sections of different phages could assemble around a short, shared region, leading to taxonomic misidentification. In order to minimize this bias, we employed a voting system based on taxonomy assigned to all of the genes within the contig and required that contigs have at least one identified gene every 10 kb to ensure that enough genes are present for proper classification. *De novo* contig assembly could also affect the interpretation of specific gene colocalizations observed (i.e., Fig. 4B and C). Antibiotic genes could be adjacent to other genes but misrepresented because of contig assembly issues associated with mosaic genomes. More detailed molecular analyses are required to draw conclusions about genomic structure.

The skin microbiome is a low-biomass community in comparison with that of other body sites (i.e., the gastrointestinal tract), and care must be taken to control for potential contamination in reagents. Many shotgun metagenomic studies of low-biomass communities have not addressed background contamination by sequencing and analysis of appropriate negative controls, resulting in erroneous conclusions (47). Previous virome studies have attempted to increase biomass by pooling samples, but this design is not conducive to the identification of interindividual variability. In the present study, in addition to minimizing contamination during library preparation, blank background controls were collected and analyzed. *In silico* decontamination removed organisms from experimental samples that were also present in the background controls. We defined successful removal of background contamination as a significant difference between the background controls and the experimental samples based on the Bray-Curtis distance metric. A caveat to this approach is that it is reference dependent and potential uncharacterized contaminants would not be detected. Also, if the background controls were not sequenced to absolute saturation, some low-abundance contaminants would not be detected. While this could impact the lowest-abundance contaminants, it is unlikely that major contaminants remained in the samples.

Overall, the findings outlined here set the stage for future studies of (i) acquisition of viral communities, (ii) responses to perturbations such as antibiotic therapies and hygienic routines, (iii) factors impacting temperate versus lytic replication cycles (i.e., DNA-damaging UV radiation or antibiotics), and (iv) impacts on human health and disease. In the long term, this work may also inform potential therapeutic strategies for skin disorders based on phage therapy.

MATERIALS AND METHODS

See Text S1 in the supplemental material for a detailed description of our methods, as well as the source code and intermediate data files related to all of our experiments.

Sample collection. We recruited a cohort of 16 healthy individuals (ranging from 23 to 53 years old) in accordance with protocols approved by the University of Pennsylvania Internal Review Board. Sample collection was performed after informed consent was obtained from the subject. Exclusion criteria included self-reported antibiotic treatment (oral or systemic) 6 months prior to enrollment, observable dermatologic diseases, and significant comorbidities, including HIV infection and other immunocompromised states.

Sample sequencing and processing. Whole metagenome DNA was prepared from cutaneous swab samples by using techniques similar to

those previously described (33, 48). The VLP DNA extraction protocol was optimized from a previously described method (15). The DNA was prepared for sequencing by using an optimized protocol for the Illumina Nextera XT library preparation kit. Sequencing was performed on the Illumina MiSeq and HiSeq2500 rapid chemistry platforms. All community analyses were performed with custom Bash, R, and Perl scripts, building off of established concepts and utilizing existing algorithms and toolkits, including the BLAST+ (49) toolkit and bowtie2 (50).

Quality control was performed to remove sequencing adapters, low-quality sequences, and sequences with similarity to the human genome (51). Mock negative-control samples were also collected to control for background sequencing signals. We performed follow-up analyses of these control data to ensure a high-quality sequence set. Contigs were assembled by using the high-quality sequences in the Ray *de novo* assembly program (52).

Taxonomy and diversity. As previously described, virome taxonomy was assigned by annotating ORFs on the basis of the UniProt reference database (53) and assigning contig taxonomy on the basis of the most frequent ORF taxonomy similarity present (22). Alpha diversity was estimated by including both the known and unknown viruses with the PHACCS algorithm (54) and GAAS program (55). Beta diversity was assessed by using the Bray-Curtis dissimilarity metric within the vegan R package (CRAN) (56) and was based on normalized sequence counts (number of reads per kilobase of transcript per million mapped reads) for each contig by sample (56). Beta diversity information was also used for the intra- and interpersonal diversity calculations. Whole-metagenome taxonomy was assigned by using MetaPhlAn (25, 26) and MEGAN5 (27). Whole-metagenome diversity was calculated with the vegan R package. For comparison, the alpha diversity of each anatomic site from both the virome and the whole metagenome was calculated by the box plot notch calculation described in the ggplot2 R package (57), as well as by McGill et al. (58).

Prediction of bacteriophage replication cycle distribution. Virome replication cycle distribution was calculated by quantifying the presence of temperate marker genes, including integrase genes from both the serine and tyrosine families, prophage elements within the ACLAME database (34) including components of *parABS* partitioning systems, and bacterial reference genome elements. Sequences were mapped back to the temperate and lytic contigs to assess normalized relative abundance.

Functional annotation and comparison. Sequence functionality was predicted by mapping reads to a reduced KEGG reference database (35) and annotating them with the HUMAnN program (36). GO enrichment analysis was performed in GOEAST (59) with ORFs that were predicted with the Glimmer3 toolkit (60) and subjected to a BLAST search of the UniProt reference database. OPF and AMG analyses were performed similar to those in previous studies (9, 37) and utilized the UCLUST (61) algorithm in QIIME (62). The CARD (39) and VFDB (63) were used with predicted ORFs to estimate the potential for antibiotic resistance and virulence, respectively. Visualization of ARGs was performed in the Geneious program (64).

Inferred interactions between phages and bacteria. Inferred interactions between phages and bacteria were calculated with CoNet (65) within Cytoscape (66) as previously described (42). Only interactions supported by two of the five metrics tested (the Pearson and Spearman correlation metrics, the mutual information similarity metric, and the Bray, Curtis, and Kullback-Leibler distance metrics) were retained for analysis of potential interactions. *P* values from the multiple metrics were combined by the Simes method (67), and false-discovery rate correction was performed (68). Network analysis was performed with the Cytoscape NetworkAnalyzer plugin (69).

CRISPR identification and comparison to the virome. CRISPR targeting of the bacterial hosts against the viruses was performed with the PilerCR program for CRISPR identification within bacterial genomes (70). The CRISPR spacer sequences were mapped against the phage contigs from various locations to evaluate potential targeting with Circos (71). Phage ORFs targeted by spacers were identified by using the UniProt TrEMBL database and BLASTx (E value, $<10^{-10}$).

Nucleotide sequence accession numbers. The sequences determined in this study have been deposited in the NCBI Short Read Archive (SRA) under BioProject PRJNA266117 and SRA accession number SRP049645. The sequenced mock community has been deposited under BioProject PRJNA295605 as sample MG100410. The analysis scripts described in Materials and Methods and intermediate files have been archived at Figshare Digital Science, London, United Kingdom, and are available at doi: 10.6084/m9.figshare.1281248.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01578-15/-/DCSupplemental.

Text S1, DOCX file, 0.2 MB.
Figure S1, PDF file, 1.3 MB.
Figure S2, PDF file, 0.2 MB.
Figure S3, PDF file, 0.5 MB.
Figure S4, PDF file, 0.4 MB.
Figure S5, PDF file, 0.3 MB.
Figure S6, PDF file, 0.5 MB.
Figure S7, PDF file, 0.3 MB.
Table S1, XLSX file, 0.05 MB.
Table S2, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

We thank the volunteers for their participation in this study, the Penn Next Generation Sequencing Core for sequencing support, the Penn Medicine Academic Computing Services for computing resources, Brian Kim (Washington University) for discussions, and members of the Grice and Bushman laboratories for their underlying contributions.

This work was supported by a grant from the NIH (AR060873 to E.A.G.). G.D.H. and A.J.S. are supported by the Department of Defense National Defense Science and Engineering graduate fellowship program, J.S.M. is supported by NIH computational genomics training grant T32 HG000046, and B.P.H. was supported by NIH dermatology research training grant T32 AR007465. We have no conflict of interest to declare.

G.D.H. and E.A.G. conceived and designed the study. A.S.T. collected skin swabs from subjects. G.D.H., J.S.M., A.J.S., and A.S.T. prepared samples for sequencing. G.D.H., J.S.M., B.P.H., and Q.Z. analyzed sequence data. S.M. and F.D.B. contributed software and helped with analysis. G.D.H., J.S.M., and E.A.G. drafted the manuscript.

REFERENCES

1. **Hannigan GD, Grice EA.** 2013. Microbial ecology of the skin in the era of metagenomics and molecular microbiology. Cold Spring Harb Perspect Med **3:**a015362. http://dx.doi.org/10.1101/cshperspect.a015362.
2. **Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, Segre JA.** 2014. Biogeography and individuality shape function in the human skin metagenome. Nature **514:**59–64. http://dx.doi.org/10.1038/nature13786.
3. **Bohannan BJM, Lenski RE.** 1997. Effect of resource enrichment on a chemostat community of bacteria and bacteriophage. Ecology **78:**2303–2315. http://dx.doi.org/10.1890/0012-9658(1997)078[2303:EOREOA]2.0.CO;2.
4. **Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pašić L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F.** 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4:**739–751. http://dx.doi.org/10.1038/ismej.2010.1.
5. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol **10:**200–207.
6. **Oliver KM, Degnan PH, Hunter MS, Moran NA.** 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science **325:**992–994. http://dx.doi.org/10.1126/science.1174463.

7. **Tyler JS, Beeri K, Reynolds JL, Alteri CJ, Skinner KG, Friedman JH, Eaton KA, Friedman DI.** 2013. Prophage induction is enhanced and required for renal disease and lethality in an EHEC mouse model. PLoS Pathog **9:**e1003236. http://dx.doi.org/10.1371/journal.ppat.1003236.
8. **Brussow H, Canchaya C, Hardt W-.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev **68:**560–602. http://dx.doi.org/10.1128/MMBR.68.3.560-602.2004.
9. **Hurwitz BL, Brum JR, Sullivan MB.** 2015. Depth-stratified functional and taxonomic niche specialization in the "core" and "flexible" Pacific Ocean virome. ISME J **9:**472–484 http://dx.doi.org/10.1038/ismej.2014.143.
10. **Modi SR, Lee HH, Spina CS, Collins JJ.** 2013. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. Nature **499:**219–222. http://dx.doi.org/10.1038/nature12212.
11. **Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA, Pariente K, Segondy M, Burguière A, Manuguerra J, Caro V, Eloit M.** 2012. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. PLoS One **7:**e38499. http://dx.doi.org/10.1371/journal.pone.0038499.
12. **Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM.** 2014. Metagenomic analysis of double-stranded DNA viruses in healthy adults. BMC Biol **12:**71. http://dx.doi.org/10.1186/s12915-014-0071-7.
13. **Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR, Brucker W, Kumar V, Kandasamy J, Keenan L, Bardarov S, Kriakov J, Lawrence JG, Jacobs WR, Jr., Hendrix RW, Hatfull GF.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:**171–182. http://dx.doi.org/10.1016/S0092-8674(03)00233-2.
14. **Breitbart M.** 2012. Marine viruses: truth or dare. Annu Rev Mar Sci **4:**425–448. http://dx.doi.org/10.1146/annurev-marine-120709-142805.
15. **Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F.** 2009. Laboratory procedures to generate viral metagenomes. Nat Protoc **4:**470–483. http://dx.doi.org/10.1038/nprot.2009.10.
16. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466:**334–338. http://dx.doi.org/10.1038/nature09199.
17. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.** 2011. The human gut virome: inter-individual variation and dynamic response to diet. Genome Res **21:**1616–1625. http://dx.doi.org/10.1101/gr.122705.111.
18. **Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S, Rodriguez-Brito B, Salamon P, Rohwer F.** 2008. Viral diversity and dynamics in an infant gut. Res Microbiol **159:**367–373. http://dx.doi.org/10.1016/j.resmic.2008.04.006.
19. **Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F.** 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185:**6220–6223. http://dx.doi.org/10.1128/JB.185.20.6220-6223.2003.
20. **Wommack KE, Bhavsar J, Ravel J.** 2008. Metagenomics: read length matters. Appl Environ Microbiol **74:**1453–1463. http://dx.doi.org/10.1128/AEM.02181-07.
21. **Roux S, Krupovic M, Debroas D, Forterre P, Enault F.** 2013. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. Open Biol **3:**130160. http://dx.doi.org/10.1098/rsob.130160.
22. **Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD.** 2013. Rapid evolution of the human gut virome. Proc Natl Acad Sci U S A **110:**12450–12455. http://dx.doi.org/10.1073/pnas.1300833110.
23. **Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, Segre JA.** 2014. Biogeography and individuality shape function in the human skin metagenome. Nature **514:**59–64. http://dx.doi.org/10.1038/nature13786.
24. **Gutiérrez D, Adriaenssens EM, Martínez B, Rodríguez A, Lavigne R, Kropinski AM, García P.** 2014. Three proposed new bacteriophage genera of staphylococcal phages: "3alikevirus," "77likevirus" and "Phietalikevirus." Arch Virol **159:**389–398. http://dx.doi.org/10.1007/s00705-013-1833-1.
25. **Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C.** 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods **9:**811–814. http://dx.doi.org/10.1038/nmeth.2066.
26. **Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C.** 2013. Computational meta'omics for microbial community studies. Mol Syst Biol **9:**666. http://dx.doi.org/10.1038/msb.2013.22.
27. **Huson DH, Mitra S, Ruscheweyh H-, Weber N, Schuster SC.** 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res **21:**1552–1560. http://dx.doi.org/10.1101/gr.120618.111.
28. **Human Microbiome Project Consortium.** 2012. Structure, function and diversity of the healthy human microbiome. Nature **486:**207–214. http://dx.doi.org/10.1038/nature11234.
29. **Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Becker J, Benjamin B, Blakesley R, Bouffard G, Brooks S, Coleman H, Dekhtyar M, Gregory M, Guan X, Gupta J, Han J, Hargrove A, Ho SL, Johnson T, Legaspi R, Lovett S, Maduro Q, Masiello C, Maskeri B, McDowell J, Montemayor C, Mullikin J, Park M, Riebow N, Schandler K, Schmidt B, Sison C, Stantripop M, Thomas J, Thomas P, Vemulapalli M, Young A, Kong HH, Segre JA.** 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature **498:**367–370 http://dx.doi.org/10.1038/nature12171.
30. **Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, NISC Comparative Sequencing Program, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA.** 2009. Topographical and temporal diversity of the human skin microbiome. Science **324:**1190–1192. http://dx.doi.org/10.1126/science.1171700.
31. **Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F.** 2006. The marine viromes of four oceanic regions. PLoS Biol **4:**e368. http://dx.doi.org/10.1371/journal.pbio.0040368.
32. **Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R.** 2009. Bacterial community variation in human body habitats across space and time. Science **326:**1694–1697. http://dx.doi.org/10.1126/science.1177486.
33. **Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, Bouffard GG, Blakesley RW, Murray PR, Green ED, Turner ML, Segre JA.** 2009. Topographical and temporal diversity of the human skin microbiome. Science **324:**1190–1192. http://dx.doi.org/10.1126/science.1171700.
34. **Leplae R, Hebrant A, Wodak SJ, Toussaint A.** 2004. ACLAME: a CLAssification of mobile genetic elements. Nucleic Acids Res **32:**D45–D49. http://dx.doi.org/10.1093/nar/gkh084.
35. **Kanehisa M, Goto S.** 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res **28:**27–30. http://dx.doi.org/10.1093/nar/28.1.27.
36. **Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, Rodriguez-Mueller B, Zucker J, Thiagarajan M, Henrissat B, White O, Kelley ST, Methé B, Schloss PD, Gevers D, Mitreva M, Huttenhower C.** 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol **8:**e1002358. http://dx.doi.org/10.1371/journal.pcbi.1002358.
37. **Schloss PD, Handelsman J.** 2008. A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. BMC Bioinformatics **9:**34. http://dx.doi.org/10.1186/1471-2105-9-34.
38. **Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA III, Loomer P, Armitage GC, Relman DA.** 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J **6:**915–926. http://dx.doi.org/10.1038/ismej.2011.169.
39. **McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJV, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD.** 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother **57:**3348–3357. http://dx.doi.org/10.1128/AAC.00419-13.
40. **Brötz E, Kulik A, Vikineswary S, Lim CT, Tan GY, Zinecker H, Imhoff JF, Paululat T, Fiedler HP.** 2011. Phenelfamycins G and H, new elfamycin-type antibiotics produced by Streptomyces albospinus Acta 3619. J Antibiot **64:**257–266. http://dx.doi.org/10.1038/ja.2010.170.
41. **Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q.** 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res **33:**D325–D328. http://dx.doi.org/10.1093/nar/gki008.

42. **Soffer N, Zaneveld J, Vega Thurber R.** 2015. Phage-bacteria network analysis and its implication for the understanding of coral disease. Environ Microbiol **17:**1203–1218. http://dx.doi.org/10.1111/1462-2920.12553.
43. **Barabasi A, Albert R.** 1999. Emergence of scaling in random networks. Science **286:**509–512. http://dx.doi.org/10.1126/science.286.5439.509.
44. **Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X.** 2010. Functional molecular ecological networks. mBio **1:**e00169. http://dx.doi.org/10.1128/mBio.00169-10.
45. **Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow CT, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, Fuhrman JA.** 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J **5:**1414–1425. http://dx.doi.org/10.1038/ismej.2011.24.
46. **Marinelli LJ, Fitz-Gibbon S, Hayes C, Bowman C, Inkeles M, Loncaric A, Russell DA, Jacobs-Sera D, Cokus S, Pellegrini M, Kim J, Miller JF, Hatfull GF, Modlin RL.** 2012. *Propionibacterium acnes* bacteriophages display limited genetic diversity and broad killing activity against bacterial skin isolates. mBio **3:**e00279–12. http://dx.doi.org/10.1128/mBio.00279-12.
47. **Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW.** 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **12:**87. http://dx.doi.org/10.1186/s12915-014-0087-z.
48. **Hannigan GD, Hodkinson BP, McGinnis K, Tyldsley AS, Anari JB, Horan AD, Grice EA, Mehta S.** 2014. Culture-independent pilot study of microbiota colonizing open fractures and association with severity, mechanism, location, and complication from presentation to early outpatient follow-up. J Orthop Res **32:**597–605. http://dx.doi.org/10.1002/jor.22578.
49. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421. http://dx.doi.org/10.1186/1471-2105-10-421.
50. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.
51. **Schmieder R, Edwards R.** 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One **6:**e17288. http://dx.doi.org/10.1371/journal.pone.0017288.
52. **Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J.** 2012. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol **13:**R122. http://dx.doi.org/10.1186/gb-2012-13-12-r122.
53. **UniProt C.** 2014. Activities at the universal protein resource (UniProt). Nucleic Acids Res **42:**D191–D198. http://dx.doi.org/10.1093/nar/gkt1140.
54. **Angly F, Rodriguez-Brito B, Bangor D, McNairnie P, Breitbart M, Salamon P, Felts B, Nulton J, Mahaffy J, Rohwer F.** 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics **6:**41. http://dx.doi.org/10.1186/1471-2105-6-41.
55. **Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, Antonopoulos DA, Barott K, Cottrell MT, Desnues C, Dinsdale EA, Furlan M, Haynes M, Henn MR, Hu Y, Kirchman DL, McDole T, McPherson JD, Meyer F, Miller RM, Mundt E, Naviaux RK, Rodriguez-Mueller B, Stevens R, Wegley L, Zhang L, Zhu B, Rohwer F.** 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLOS Comput Biol **5:**e1000593. http://dx.doi.org/10.1371/journal.pcbi.1000593.
56. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H.** 2014. Vegan: community ecology package. R package version 2.2-0. http://CRAN.R-project.org/package=vegan.
57. **Wickham H.** 2009. ggplot2: elegant graphics for data analysis. Springer, New York, NY.
58. **McGill R, Tukey JW, Larsen WA.** 1978. Variations of box plots. Am Stat **32:**12–16. http://dx.doi.org/10.1080/00031305.1978.10479236.
59. **Zheng Q, Wang X-.** 2008. GOEAST: a web-based software toolkit for gene ontology enrichment analysis. Nucleic Acids Res **36:**W358–W363. http://dx.doi.org/10.1093/nar/gkn276.
60. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics **23:**673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.
61. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461.
62. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7:**335–336. http://dx.doi.org/10.1038/nmeth.f.303.
63. **Chen L, Xiong Z, Sun L, Yang J, Jin Q.** 2012. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res **40:**D641–D645. http://dx.doi.org/10.1093/nar/gkr989.
64. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
65. **Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C.** 2012. Microbial co-occurrence relationships in the human microbiome. PLoS Comput Biol **8:**e1002606. http://dx.doi.org/10.1371/journal.pcbi.1002606.
66. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.** 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:**2498–2504. http://dx.doi.org/10.1101/gr.1239303.
67. **Sarkar SK, Chang C.** 1997. The Simes method for multiple hypothesis testing with positively dependent test statistics. J Am Stat Assoc **92:**1601–1608. http://dx.doi.org/10.1080/01621459.1997.10473682.
68. **Benjamini Y, Hochberg Y.** 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodological **57:**289–300.
69. **Assenov Y, Ramirez F, Schelhorn S-, Lengauer T, Albrecht M.** 2008. Computing topological parameters of biological networks. Bioinformatics **24:**282–284. http://dx.doi.org/10.1093/bioinformatics/btm554.
70. **Edgar RC.** 2007. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics **8:**18. http://dx.doi.org/10.1186/1471-2105-8-18.
71. **Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.** 2009. Circos: an information aesthetic for comparative genomics. Genome Res **19:**1639–1645. http://dx.doi.org/10.1101/gr.092759.109.