Evolution of the facial skin microbiome during puberty in normal and acne skin.

A.M. Schneider,[1] Z.T. Nolan,[1] K. Banerjee,[2] A.R. Paine,[1] Z. Cong,[1] S.L. Gettle,[1] A.L. Longenecker,[1] X. Zhan,[3] G.W. Agak,[4] A.M. Nelson[1]

Supplementary Information

Materials and Methods

Human Subjects

Forty-eight volunteers were recruited under an approved Institutional Review Board protocol (#8042) from the Pennsylvania State University College of Medicine. All participants provided written informed assent with parental consent prior to the study. Participants included males and females, ages 7-17, with normal or acne skin. All children were healthy with no underlying skin/medical conditions other than acne, if present. The menstrual cycle was not recorded for females. Information regarding use of oral contraceptives (OCPs), including brands and frequency, was not obtained and could not be assessed in this study. No subjects had a history of systemic retinoid use. Only subjects who verbally confirmed no oral antibiotic use within the past 4 weeks were enrolled. A chart review of the 12 months prior to study enrollment did not identify systemic or topical antibiotic exposure for any subject. Exclusion criteria included use of topical antibiotics, benzoyl peroxide, or salicylic washes within 2 weeks prior to the start of the study; and use of topical retinoids, dermabrasion or facial laser therapy within 4 weeks prior to the start of the study. No instructions on modifying or limiting each subject's regular hygiene practices (bathing, lotion use) prior to sampling were given and bathing frequency/timing prior to sampling were not documented. For each subject, we collected the following demographic and clinical data: 1) age, 2) sex, 3) race/ethnicity, 4) Tanner stage of pubertal development staged by a board-certified dermatologist,[1,2] 5) presence or absence of acne; if present, acne was staged by dermatologist on the indicated IGA scale,[3] 6) sebumeter readings and 7) presence or absence of lotions and makeup on the skin at sampling time. Demographics are provided in Supplemental Table S1.

Cohort Characteristics (a) by Tanner Stage, (b) by early and late Tanner groups

(a)

Tanner Stage	1	2	3	4	5	Total
Normal	6	7	2	1	5	21
Acne	6	2	6	7	6	27
Total	12	9	8	8	11	48
Age Median	7.9	10.6	12	13.5	15.6	11.8
(Range)	(7-9)	(7-14)	(10-13)	(12-16)	(12-17)	(7-17)

(b)

Tanner Stage	Early (T1-T2)	Late (T3-T5)	Total
Normal	13	8	21
Acne	8	19	27
Total	21	27	48
Age Median	9.0	13.9	11.8
(Range)	(7-14)	(10-17)	(7-17)

IGA Scale Adapted from the FDA center for drug evaluation and research on acne vulgaris[3]

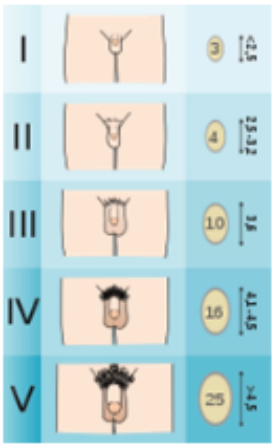
Score	Grade	Description
0	Clear	Normal, clear skin with no evidence of acne vulgaris
1	Almost Clear	Rare non-inflammatory lesions may be present, with rare non-inflamed papules (papules must be resolving and may be hyperpigmented, though not pink-red)
2	Mild	Some non-inflammatory lesions are present, with few inflammatory lesions (papules/pustules only; no nodulocystic lesions)
3	Moderate	Many non-inflammatory lesions. Multiple inflammatory lesions evident with several to many papules/pustules, and there may be one small nodulocystic lesion
4	Severe	Inflammatory lesions are more apparent, many comedones and papules/pustules, there may be a few nodulocystic lesions
5	Very Severe	Highly inflammatory lesions predominate, variable number of comedones, many papules/pustules and many nodulocystic lesions

Tanner Stage Evaluation Form

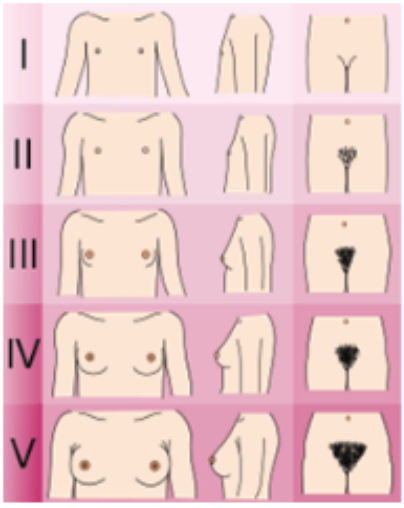
Tanner staging was done according to Marshall and Tanner, using the criteria below.[1,2] The format and wording of this evaluation form was adopted by our group using descriptions of pubertal stages obtained from the original publications, The American Academy of Pediatrics, and WIKIPEDIA. Images were obtained from WIKIPEDIA.

TANNER STAGE

Genitals (male)		□N/A
□	Tanner I	Testicular volume less than 1.5 ml; small penis of 3 cm or less (pre pubertal) (typically age nine and younger)
□	Tanner II	Testicular volume between q.6 and 6 ml; skin on scrotum thins, reddens and enlarges; penis length unchanged (9-11)
□	Tanner III	Testicular volume between 6 and 12 ml; scrotum enlarges further; penis begins to lengthen to about 6 cm (11-12.5)
□	Tanner IV	Testicular volume between 12 and 20 ml; scrotum enlarges further and darkens; penis increases in length to 10 cm (12.5-14)
□	Tanner V	Testicular volume greater than 20 ml; adult scrotum and penis of 15 cm in length (14+)



Breasts (female)		□ N/A
□	Tanner I	No glandural tissue: areola follows the skin contours of the chest (pre pubertal) (typically age 10 and younger)
□	Tanner II	Breast bud forms, with small area of surrounding glandular tissue; areola begins to widen (10-11.5)
□	Tanner III	Breast begins to become more elevated, and extends beyond the borders of the areola, which continues to widen but remains in contour with surround breast (11.5-13)
□	Tanner IV	Increased breast size and elevation; areola and papilla form a secondary mound projecting the contour of the surrounding breast (13-15)
□	Tanner V	Breast reaches final adult size; areola return to contour of the surrounding breast, with a projecting central papilla. (15+)



Pubic hair (both male and female)		
□	Tanner I	No pubic hair at all (pre pubertal)(typically age 10 and younger)
□	Tanner II	Small amount of long, downy hair with slight pigmentation at the base of the penis and scrotum (males) or on the labia majora (females) (10-11.5)
□	Tanner III	Hair becomes more coarse and curly, and begins to extend laterally (11.5-13)
□	Tanner IV	Adult-like hair quality, extending across pubis but sparing medial thighs (13-15)
□	Tanner V	Hair extends to medial surface of the thighs (15+)

Sebum Measurements

The sebum output on the forehead of each subject was measured with the Sebumeter 810™ Instrument according to manufacturer directions. Sebumeter readings were captured prior to skin microbiome sampling. Briefly, single sebumeter readings were captured from the left, center and right regions of the forehead. These three readings for each individual were averaged together to calculate the overall sebum measurement for each subject.

Microbiome Sampling

Sampling occurred within the Dermatology Clinical Research Unit at Penn State Health. Sterile synthetic fiber applicators (with plastic handle) were pre-moistened with yeast lysis buffer (Epicentre, Illumina). A 7 cm^2 area (using a plastic template) in the center of the forehead was scrubbed with moderate pressure for thirty seconds and placed in a sterile 1.5 mL tube containing 600 µL of yeast lysis buffer (Epicentre, Illumina). Both the sebum measurement and skin microbiome collection was done on the forehead and acne was present on the forehead in acne subjects. No washes or solutions were applied to the forehead before sampling. Sebum measurements were collected first, using the Sebumeter 810 device. The probe surface area that touches the skin's surface is 64 mm^2. We sampled the sebum on the left, middle and right areas of the forehead (little circles in Diagram 1). We then sampled the microbiome from a 7 cm^2 (700 mm^2) area of the middle forehead, which did overlap with one of the sebum measurement areas (larger blue circle in Diagram 1). In total, ~10% of the skin microbiome sampling area overlapped with the sebum measurement area; the majority of the sampling area was undisturbed prior to sampling. The swabs were immediately stored at -80 °C. For negative control sample, a sterile swab was removed from its packaging, held in the air for ~30 seconds to mimic the time that a swab is in contact with human skin, and placed into a sterile 1.5 mL tube containing 600 µL of yeast lysis buffer and stored at -80 ºC with all samples. Mock community controls (MCC); generalized community control (ATCC MSA-1005) and skin-specific community control (ATCC MSA-1002), and negative

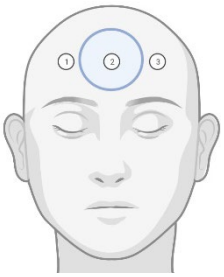


Diagram 1

sampling controls (n=5) (e.g., sampling swabs without skin contact) were processed alongside samples from human subjects.

Metagenomic sequencing, processing, and taxonomic identification

Frozen swab samples were shipped to Microbiome Insights (British Columbia, Canada) for DNA extraction, library preparation and sequencing. DNA was extracted using the Qiagen MagAttract PowerSoil DNA KF kit (Formerly MO Bio PowerSoil DNA Kit) using a KingFisher robot. DNA quality was evaluated by gel electrophoresis and quantified with Qubit 3.0 fluorometer (Thermo-Fischer, Waltham, MA, USA). Libraries were made following manufacturer protocols and Illumina Nextera library preparation kit (Illumina, San Diego, CA, USA). Sequencing was done using a high output Illumina Nestseq run (2x150 paired-end reads) which yielded 197.46 Gbases (median = 3.9 Gbases). Negative control 6 was added as a sequencing control.

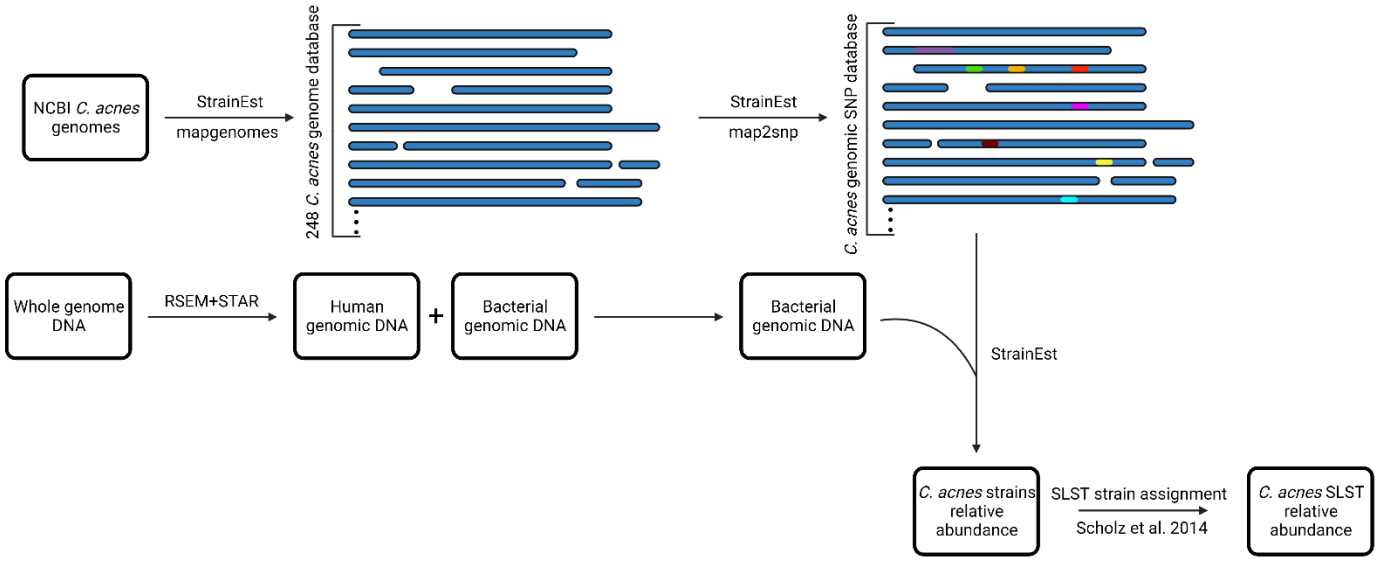
Initial quality evaluation was done using FastQC (v0.11.5).[4] Raw sequences were processed for: adapter removal, read trimming, low-complexity-reads removal, and host-sequence removals. Adapter removal was done using cutadapt (v2.6).[5] Trimming was done with Trimmomatic (v0.36)[6] using custom parameters (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). Low-complexity sequences were detected with Komplexity v0.3.6.[7] High-quality reads were mapped to the human genome (Genome Reference Consortium Human Reference 37) and those that mapped were removed from the analysis. After quality control, the remaining high-quality sequences were used for taxonomic profiles using MetaPhlAn2,[8] consistent with prior studies.[9] The median number of high-quality filtered reads per sample was 2,840,816 (Table S3). Consistent with previous whole genome sequencing studies, samples with > 50,000 non-human quality controlled reads were included in our study.[9,10] Positive mock community controls (n=2), negative sampling controls (n=5), and a negative sequencing control (n=1) were analyzed in parallel with patient samples (Fig. S1).

Diversity Analyses and Statistics

Relative abundance data was available for 332 species (308 bacteria, 22 viruses, 1 Eukaryota, and 1 Unclassified-group). Prior to statistical analysis, we eliminated 10 (3%) species which were only present in the positive controls and 101 (30.4%) species which were detected in a single sample with a relative abundance of less than 1% (Table S4). Furthermore, two Escherichia species were also dropped as they were detected in the negative controls at high levels and are likely contaminants. No species were detected in the negative sequencing control sample (NSC, Table S3). All further analysis was done with a microbiome community of the remaining 219 species (201 bacteria, 16 viruses, 1 Eukaryota, and 1 Unclassified-group). The Wilcoxon rank sum test was used to compare relative abundances and Shannon indices. The Bray-Curtis dissimilarity index is used to calculate beta diversities, which were compared using PERMANOVA (Permutational Multivariate Analysis of Variance).[11] Unless otherwise specified, the reported p-values are not adjusted for multiple testing. If adjusted for multiple testing, adjusted p-values are obtained using the Bonferroni correction. Viral sequences dominated 8 of 48 samples, but no significant association between viral relative abundance and demographic (age, sex, ethnicity) or clinical data (sebum levels, Tanner stage, IGA score) were found.

***C. acnes* Strain Level Analysis**

To identify and analyze the *C. acnes* strain composition in our dataset, we followed the pipeline illustrated and described here.



All publicly available *C. acnes* genomes (complete and draft) available at NCBI (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/1140/) as of October 7th 2020 were downloaded to be utilized in the identification of *C. acnes* strains within our dataset. StrainEst (v1.2.4) was used to build an index of single nucleotide variants (SNV) present in the *C. acnes* strains (mapgenomes, map2snp).[12] A custom reference index of C. acnes SNV was built using Bowtie2 (v2.3.2.4) (bowtie-build). Our data was aligned against this custom reference index to identify each individual *C. acnes* strain with Bowtie2.[13] Samtools (v1.10.0)[14] was used to reformat the .sam output of Bowtie2 to a sorted and indexed .bam format compatible for use in StrainEst. Strains were identified using StrainEst (with filtered reads from sequence processing as previously described) default parameters, except 3X coverage for the *C. acnes* genome. A second custom *C. acnes* SNV reference index containing strains only identified in our samples was built. This was used with StrainEst to determine the relative abundance of each *C. acnes* strain per sample.

The single locus sequence typing (SLST) classification scheme was used to classify *C. acnes* strains.[15] Based on degree of homology, the SLST scheme uses a single locus to assign strains to a specific cluster (A-H,

K, L) and further subclassification to a nested strain group (e.g., A1, A2, B1, B2, etc.). To determine strain SLST classifications in this study, we performed a BLAST search of SLST sequence as determined by Scholz and colleages[15] against the identified *C. acnes* strains in our dataset. We identified a total of 167 unique C. acnes strains in our dataset. 153 of 167 *C. acnes* strains were assigned to SLST clusters and SLST groups based on 100% homology to SLST sequences defined by the Scholz 2014 publication.[15] Eleven strains were only assigned to a specific cluster (A-H, K, L) based on 97-99% sequence homology to other strains within that cluster. By definition of the SLST scheme, even 1 SNP would constitute a different group, therefore, an identified strain must have a 100% homology for strain group assignment (for example, A1). Three unique strains could not be assigned to a specific strain cluster because they were less than 97% homologous; subsequently, we grouped these strains together in a single cluster, denoted X (unclassified). Corresponding C. *acnes* phylotypes are denoted next to SLST classification in brackets.[16–18]

We analyzed the relative abundance data according to SLST cluster for diversity analysis at both the cluster and group level. α–diversity was calculated using the Shannon Index; comparisons between α–diversity values were made using the Wilcoxon rank sum test. The Bray-Curtis dissimilarity index was used to calculate β-diversity values, which were compared using PERMANOVA.

Functional Profile Analysis

For functional profile analyses filtered data as described above was used. Filtered data was further processed to remove rRNA reads from the dataset using SortMeRNA (v4.1.0)[19] utilizing all of its available databases before functional analysis. Paired-end and remaining singleton reads for each sample were then concatenated into a single input per sample for the HUMAnN2 (v2.8.1) analysis package.[20] Data was then analyzed in HUMAnN2 using its default analysis pipeline. As *E. coli* had been identified as a potential contaminant in our negative control samples, all reads associated with *E. coli* were removed from the data. All reads within a given functional classification was then grouped together regardless of the species associated with the read and converted from UniRef to KEGG Orthology (KO) terms using a function in

HUMAnN2(humann2_regroup_table -g uniref90_ko). 7,134 KO terms were then used as input into MicrobiomeAnalyst on October 16th 2020.[21] Using its default filtering parameters for low counts and low variance 2,562 KOs were removed. 4,572 KEGG orthology enzymes underwent univariate analysis and pathway enrichment analysis. FDR was set at q=0.05 to identify enzymes and pathways of biological interest.

Supplemental Tables

Table S1: Patient demographics and metadata

ID	Age (yrs)	Sex	Ethnicity	Hispanic/ Latino (Y/N)	Tanner Stage	Acne or Normal	IGA Score	Sebumeter (mean)	Lotion (Y/N)	Make-up (Y/N)
S01	12	F	White	Y	4	Normal	n/a	112.67	N	N
S02	12	F	Hispanic	Y	4	Acne	3	57.67	N	N
S03	14	F	Black	N	4	Acne	2	107.00	N	N
S04	10	M	White	Y	2	Normal	n/a	9.67	N	N
S05	17	F	White	N	5	Acne	3	129.33	N	N
S06	7	F	White	N	1	Normal	n/a	3.33	N	N
S07	10	F	White	N	3	Acne	1	30.00	N	N
S08	7	M	White	N	1	Normal	n/a	1.33	N	N
S09	17	F	Asian	N	5	Acne	1	55.33	N	N
S10	8	M	Egyptian	Y	1	Normal	n/a	8.67	N	N
S11	12	F	Egyptian	N	3	Acne	3	77.33	N	N
S12	15	M	White	Y	5	Acne	3	54.67	N	N
S13	7	F	Black	N	2	Normal	n/a	28.33	Y	N
S14	17	F	Asian	N	5	Acne	1	29.67	Y	N
S15	13	M	White	N	4	Acne	2	82.33	N	N
S16	12	F	White	N	3	Acne	2	88.67	N	N
S17	13	M	Hispanic	Y	4	Acne	2	31.33	N	N
S18	14	F	Asian	N	2	Normal	n/a	34.33	Y	N
S19	8	M	White	Y	1	Normal	n/a	6.33	N	N
S20	14	M	Dominican	Y	4	Acne	3	50.67	N	N
S21	12	F	Black	N	3	Normal	n/a	40.33	N	N
S22	12	F	White	N	5	Acne	3	90.00	N	N
S23	8	M	White	N	1	Normal	n/a	0.00	N	N
S24	14	M	White	N	4	Acne	3	99.00	N	N
S25	8	M	White	N	1	Acne	1	23.00	N	N
S26	12	M	Black	N	3	Acne	2	45.00	N	N
S27	16	M	White	Y	5	Normal	n/a	53.67	N	N
S28	11	M	White	Y	2	Acne	1	4.33	N	N
S29	9	M	Black	N	1	Acne	1	131.00	N	N
S30	13	F	White	N	3	Acne	2	115.67	Y	N
S31	9	F	Black	N	1	Acne	1	45.67	N	N
S32	8	M	Black	N	1	Acne	1	13.00	N	N
S33	10	F	White	N	2	Acne	2	28.00	N	N
S34	8	F	White	N	1	Acne	1	32.33	N	N
S35	13	F	White	N	3	Acne	3	58.00	Y	N
S36	11	M	White	N	2	Normal	n/a	1.67	N	N
S37	16	F	White	N	4	Acne	2	40.00	N	N
S38	16	F	White	N	5	Acne	2	36.67	N	N
S39	10	M	Asian	N	2	Normal	n/a	42.00	N	N
S40	7	M	White	N	1	Acne	1	3.67	N	N
S41	15	F	Asian	N	5	Normal	n/a	69.00	Y	Y
S42	12	M	White	N	2	Normal	n/a	4.67	N	N
S43	8	F	White	N	1	Normal	n/a	0.00	N	N
S44	10	F	White	N	2	Normal	n/a	12.33	N	N
S45	14	M	White	N	5	Normal	n/a	53.33	N	N
S46	12	F	White	N	3	Normal	n/a	51	N	N
S47	16	F	White	N	5	Normal	n/a	180.67	Y	N
S48	17	F	White	N	5	Normal	n/a	149	Y	N

Table S2: Summary statistics for covariates

	N=48
Age, Mean (SD)**	11.8 (3.11)
Sex, N (%):**	
F	27 (56.2%)
M	21 (43.8%)
Ethnicity, N (%):	
Hispanic or Latino	10 (20.8%)
Not Hispanic or Latino	38 (79.2%)
Race, N (%):	
Asian	5 (10.4%)
Black/African American	7 (14.6%)
Dominican*	1 (2.08%)
Egyptian*	2 (4.17%)
Hispanic*	2 (4.17%)
White	31 (64.6%)
Tanner Stage (Tanner), N (%):	
1	12 (25.0%)
2	9 (18.8%)
3	8 (16.7%)
4	8 (16.7%)
5	11 (22.9%)
Acne or Normal (Status), N (%):	
Acne	27 (56.2%)
Normal	21 (43.8%)
IGA, N (%):	
1	10 (20.8%)
2	9 (18.8%)
3	8 (16.7%)
n/a	21 (43.8%)
Sebum, Mean (SD)	50.5 (43.5)
Lotion, N (%):	
N	40 (83.3%)
Y	8 (16.7%)
Makeup, N (%):	
N	47 (97.9%)
Y	1 (2.08%)

*Due to low frequencies, Dominican, Hispanic, and Egyptian have been combined into a single category in association analysis. Reported in Figure 2C of the main text.

**There is no statistical difference in age ($p = 0.212$) and sex ($p = 0.771$) between normal and acne subjects.

Table S3: Total number of quality reads.

Subject	Total reads	Non-human quality controlled reads	Average PHRED score	Subject	Total reads	Non-human quality controlled reads	Average PHRED score
S01	43,351,062	515,468	33.94	S29	16,015,446	1,985,180	32.79
S02	23,586,832	73,364	33.86	S30	33,989,198	1,857,776	34.02
S03	43,820,318	2,534,836	33.92	S31	22,595,618	124,416	33.95
S04	33,209,234	792,602	34.05	S32	31,948,270	2,833,916	34.03
S05	22,164,074	58,202	33.83	S33	35,945,874	9,778,772	34.09
S06	35,276,514	1,036,098	33.87	S34	31,778,862	4,642,038	34.09
S07	35,797,380	6,794,592	34.06	S35	30,829,244	7,584,492	33.93
S08	45,548,600	915,714	33.99	S36	30,870,460	5,980,952	34.12
S09	22,819,428	418,528	34.07	S37	32,888,394	4,739,262	34.07
S10	22,115,734	3,183,802	34.11	S38	32,075,012	9,830,498	34.04
S11	25,127,610	6,010,934	33.96	S39	29,965,254	2,700,704	34.04
S12	23,877,084	2,930,414	33.98	S40	36,105,820	432,632	33.99
S13	22,796,170	3,574,582	34.04	S41	32,371,298	392,540	34.02
S14	25,257,700	8,643,988	33.95	S42	32,622,664	3,808,152	34.03
S15	22,658,912	1,553,638	34.03	S43	41,733,306	998,522	33.98
S16	31,396,046	4,817,976	33.94	S44	38,258,514	1,179,840	33.99
S17	35,949,892	787,406	34	S45	29,999,928	1,717,440	33.95
S18	36,121,656	2,933,596	34.06	S46	34,914,058	1,444,546	33.97
S19	39,820,732	2,709,100	34.02	S47	38,215,518	777,366	33.96
S20	38,657,384	5,979,350	34.01	S48	41,378,916	381,920	33.9
S21	34,844,518	1,903,112	34.03	MCC MSA1002	24,934,686	22,405,522	34.03
S22	32,524,150	2,796,734	33.97	MCC MSA1005	18,843,196	16,675,032	33.95
S23	33,762,522	1,550,054	34.03	NC 1	1,499,874	1,393,780	34.14
S24	31,679,492	4,144,590	33.95	NC 2	4,393,326	4,078,798	34.16
S25	32,955,376	3,628,742	34.05	NC 3	4,265,224	4,023,998	34.1
S26	27,908,660	798,702	34.08	NC 4	4,461,482	4,190,042	34.14
S27	29,172,440	385,468	33.89	NC 5	5,426,748	5,051,296	34.14
S28	23,865,108	1,696,590	34.08	NSC	38,046	7,890	33.93

*PHRED score represents the quality of the sequencing run. A score of 30 represents a base accuracy of 99.9% or the probability of an incorrect base is 1 in 1,000. A score of 40 is 99.99% accurate (1 in 10,000 chance).[22,23]

MCC = Mock Community Control, NC = Negative Control, NSC = Negative Sequencing Control

Table S4: 101 species removed before analyses. Prior to analyses, we removed species that were present in positive controls (10 species) or those only detected within a single patient sample (101 species; <1% relative abundance). Negative sampling controls contained >90% Escherichia; thus, sequences corresponding to Escherichia were removed from all samples before analysis (Fig. S1c, Supplemental Methods). All further analyses were done with the remaining 219 species: 201 bacteria, 16 viruses, 1 Eukaryota, and 1 unclassified group. Thus, all species that were removed following the criteria above are listed in the table below. Proprioni. - Propionibacterium, Staph. - Staphylococcus, Strep. - Streptococcus. un. – unclassified.

Kingdom	Species		
Viruses	*Feline leukemia virus*	*Propioni. phage P14 4*	*Staph. phage 80alpha*
	Mupapillomavirus 1	*RD114 retrovirus*	*Strep. pyogenes phage 315 6*
Bacteria	*Acinetobacter bereziniae*	*Delftia un.*	*Neisseria subflava*
	Acinetobacter guillouiae	*Dermacoccus sp Ellin185*	*Oligella urethralis*
	Acinetobacter gyllenbergii	*Dialister micraerophilus*	*Oribacterium sinus*
	Actinobaculum urinale	*Dorea longicatena*	*Paracoccus denitrificans*
	Aggregatibacter aphrophilus	*Eikenella corrodens*	*Parvimonas micra*
	Agrobacterium unclassified	*Empedobacter brevis*	*Parvimonas un.*
	Agromyces unclassified	*Enterobacter aerogenes*	*Peptoniphilus rhinitidis*
	Akkermansia muciniphila	*Eubacterium biforme*	*Peptostreptococcus stomatis*
	Alistipes unclassified	*Eubacterium brachy*	*Porphyromonas gingivalis*
	Bacteroides uniformis	*Eubacterium siraeum*	*Porphyromonas gulae*
	Bacteroides vulgatus	*Halomonas stevensii*	*Porphyromonas uenonis*
	Bergeyella zoohelcum	*Halomonas un.*	*Prevotella bivia*
	Bifidobacterium angulatum	*Hymenobacter un.*	*Prevotella copri*
	Bifidobacterium animalis	*Jonquetella anthropi*	*Prevotella maculosa*
	Bifidobacterium pseudocatenulatum	*Kingella oralis*	*Prevotella saccharolytica*
	Brachybacterium paraconglomeratum	*Lachnospiraceae bacterium ICM7*	*Propionibacterium avidum*
	Brevibacterium mcbrellneri	*Lachnospiraceae oral taxon 107*	*Pseudomonas tolaasii*
	Campylobacter concisus	*Lactobacillus acidophilus*	*Rickettsia felis*
	Candidatus Prevotella conceptionensis	*Lactobacillus casei paracasei*	*Rickettsia parkeri*
	Capnocytophaga granulosa	*Lactobacillus gasseri*	*Roseburia un.*
	Caulobacter vibrioides	*Lactobacillus rhamnosus*	*Ruminococcus obeum*
	Clostridium nexile	*Lactobacillus vaginalis*	*Scardovia wiggsiae*
	Collinsella aerofaciens	*Leptotrichia goodfellowii*	*Selenomonas flueggei*
	Coriobacteriaceae bacterium BV3Ac1	*Leptotrichia hofstadii*	*Sphingobacterium un.*
	Corynebacterium amycolatum	*Leuconostoc carnosum*	*Sphingobium un.*
	Corynebacterium aurimucosum	*Leuconostoc lactis*	*Staph. equorum*
	Corynebacterium bovis	*Massilia un.*	*Strep. equi*
	Corynebacterium jeikeium	*Mitsuokella un.*	*Strep. oligofermentans*
	Corynebacterium massiliense	*Mobiluncus un.*	*Strep. pyogenes*
	Corynebacterium urealyticum	*Moraxella catarrhalis*	*Variovorax un.*
	Deinococcus radiodurans	*Neisseria gonorrhoeae*	*Xanthomonas citri*
	Delftia acidovorans	*Neisseria polysaccharea*	

Table S5: Top 20 most abundant species.

Kingdom	Species	Abundance Percentage
Bacteria	*Cutibacterium acnes*	40.65
Eukaryota	*Malassezia globosa*	4.35
Viruses	*Betapapillomavirus 3*	4.18
Bacteria	*Staphylococcus epidermidis*	3.92
Unclassified	Unclassified	3.51
Bacteria	*Streptococcus mitis oralis pneumoniae*	3.42
Viruses	*Human papillomavirus 161 like viruses*	3.15
Viruses	*Betapapillomavirus 1*	3.06
Viruses	*Betapapillomavirus 5*	2.26
Viruses	*Propionibacterium phage P101A*	2.23
Viruses	*Polyomavirus HPyV6*	1.71
Bacteria	*Enhydrobacter aerosaccus*	1.54
Bacteria	*Streptococcus thermophilus*	1.19
Bacteria	*Neisseria unclassified*	1.06
Bacteria	*Haemophilus parainfluenzae*	0.99
Bacteria	*Streptococcus sanguinis*	0.97
Bacteria	*Staphylococcus caprae capitis*	0.97
Bacteria	*Haemophilus influenzae*	0.92
Bacteria	*Gemella haemolysans*	0.79
Bacteria	*Rothia mucilaginosa*	0.77

Table S6: Top 20 most abundant species within Tanner stages 1-2.

Kingdom	Species	Abundance Percentage
Bacteria	*Cutibacterium acnes**	21.95
Eukaryota	*Malassezia globosa**	8.85
Viruses	*Betapapillomavirus 5**	5.18
Viruses	*Betapapillomavirus 3**	4.86
Bacteria	*Streptococcus mitis oralis pneumoniae**	4.39
Unclassified	Unclassified*	2.99
Viruses	*Human papillomavirus 161 like viruses**	2.61
Bacteria	*Staphylococcus epidermidis**	2.58
Viruses	*Betapapillomavirus 1**	2.16
Bacteria	*Haemophilus influenza**	2.07
Bacteria	*Neisseria unclassified**	1.92
Bacteria	*Streptococcus sanguinis**	1.74
Bacteria	*Haemophilus parainfluenzae**	1.69
Bacteria	*Rothia mucilaginosa**	1.53
Bacteria	*Gemella haemolysans**	1.36
Viruses	*Streptococcus phage EJ 1*	1.11
Bacteria	*Granulicatella elegans*	1.06
Bacteria	*Corynebacterium kroppenstedtii*	1.06
Bacteria	*Staphylococcus hominis*	1.06
Viruses	*Alphapapillomavirus 4*	1.04

* Present in Supplemental Table S5

Table S7: Top 20 most abundant species within Tanner stages 3-5.

Kingdom	Species	Abundance Percentage
Bacteria	*Cutibacterium acnes**	55.19
Bacteria	*Staphylococcus epidermidis**	4.95
Unclassified	Unclassified*	3.91
Viruses	*Propionibacterium phage P101A**	3.88
Viruses	*Betapapillomavirus 1**	3.77
Viruses	*Betapapillomavirus 3**	3.65
Viruses	*Human papillomavirus 161 like viruses**	3.57
Viruses	*Polyomavirus HPyV6**	3.05
Bacteria	*Streptococcus mitis oralis pneumoniae**	2.66
Bacteria	*Enhydrobacter aerosaccus**	2.19
Bacteria	*Streptococcus thermophiles**	1.43
Bacteria	*Staphylococcus caprae capitis**	1.17
Viruses	*Propionibacterium phage P100D*	0.86
Eukaryota	*Malassezia globosa**	0.85
Viruses	*Alphapapillomavirus 2*	0.81
Bacteria	*Haemophilus parainfluenzae**	0.45
Viruses	*Avian endogenous retrovirus EAV HP*	0.43
Bacteria	*Neisseria unclassified*	0.4
Bacteria	*Streptococcus sanguinis**	0.37
Bacteria	*Lactobacillus iners*	0.35

* Present in Supplemental Table S5

Table S8: 45 species (37 bacteria, 8 viruses) present only on normal skin.

Kingdom	Species	Abundance (%)	Prevalence (%)
Bacteria	*Neisseria flavescens*	0.3686	12.5
Bacteria	*Porphyromonas sp oral taxon 279*	0.0895	2.08
Bacteria	*Capnocytophaga sp oral taxon 329*	0.0613	2.08
Bacteria	*Kingella denitrificans*	0.0506	4.17
Bacteria	*Campylobacter showae*	0.0189	4.17
Bacteria	*Haemophilus paraphrohaemolyticus*	0.0232	4.17
Bacteria	*Cardiobacterium hominis*	0.0056	4.17
Bacteria	*Streptococcus anginosus*	0.0151	6.25
Bacteria	*Actinomyces graevenitzii*	0.0028	4.17
Bacteria	*Megasphaera micronuciformis*	0.0038	4.17
Bacteria	*Atopobium parvulum*	0.0069	6.25
Bacteria	*Prevotella denticola*	0.0023	4.17
Bacteria	*Capnocytophaga ochracea*	0.0027	4.17
Bacteria	*Kytococcus sedentarius*	0.0338	2.08
Bacteria	*Brevundimonas diminuta*	0.0346	4.17
Bacteria	*Peptoniphilus lacrimalis*	0.0445	6.25
Bacteria	*Pantoea ananatis*	0.0087	4.17
Bacteria	*Enterobacter cloacae*	0.0054	8.33
Viruses	*Streptococcus phage EJ 1*	0.4860	4.17
Bacteria	*Peptoniphilus harei*	0.1258	10.42
Bacteria	*Anaerococcus hydrogenalis*	0.0344	6.25
Bacteria	*Porphyromonas asaccharolytica*	0.0123	4.17
Bacteria	*Facklamia hominis*	0.0285	4.17
Bacteria	*Prevotella buccalis*	0.0091	4.17
Bacteria	*Actinomyces turicensis*	0.0077	4.17
Bacteria	*Anaerococcus lactolyticus*	0.0109	4.17
Bacteria	*Campylobacter ureolyticus*	0.0030	4.17
Bacteria	*Anaerococcus obesiensis*	0.0085	6.25
Bacteria	*Campylobacter hominis*	0.0030	4.17
Bacteria	*Prevotella pallens*	0.0078	4.17
Bacteria	*Atopobium rimae*	0.0074	4.17
Bacteria	*Peptoniphilus duerdenii*	0.0097	4.17
Bacteria	*Peptostreptococcus unclassified*	0.0012	4.17
Viruses	*Enterobacteria phage HK633*	0.0516	2.08
Bacteria	*Veillonella sp oral taxon 780*	0.0310	2.08
Bacteria	*Dialister invisus*	0.0024	4.17
Viruses	*Alphapapillomavirus 4*	0.4547	2.08
Viruses	*Betapapillomavirus 3*	4.1791	8.33
Bacteria	*Ruminococcus torques*	0.0193	4.17
Bacteria	*Acinetobacter oleivorans*	0.0739	2.08
Bacteria	*Pedobacter unclassified*	0.0024	4.17
Viruses	*Betapapillomavirus 5*	2.2642	4.17
Viruses	*Alphapapillomavirus 2*	0.4567	2.08
Viruses	*Staphylococcus phage phiETA3*	0.0967	2.08
Viruses	*Staphylococcus phage phi2958PVL*	0.0791	2.08

Table S9: 16 species (14 bacteria, 2 viruses) present only on acne skin.

Kingdom	Species	Abundance (%)	Prevalence (%)
Bacteria	*Staphylococcus haemolyticus*	0.0035	6.25
Bacteria	*Propionibacterium sp KPL1844*	0.0609	2.08
Bacteria	*Acinetobacter baumannii*	0.0023	4.17
Bacteria	*Pseudonocardia unclassified*	0.0050	4.17
Viruses	*Polyomavirus HPyV6*	1.7150	2.08
Bacteria	*Roseomonas unclassified*	0.0117	4.17
Bacteria	*Atopobium vaginae*	0.1417	6.25
Bacteria	*Prevotella amnii*	0.0346	4.17
Bacteria	*Gordonia terrae*	0.0034	8.33
Bacteria	*Actinomyces neuii*	0.0054	4.17
Bacteria	*Faecalibacterium prausnitzii*	0.0918	2.08
Bacteria	*Eubacterium rectale*	0.0876	4.17
Bacteria	*Staphylococcus lugdunensis*	0.1010	4.17
Bacteria	*Streptococcus vestibularis*	0.0095	4.17
Bacteria	*Megasphaera genomosp type 1*	0.0066	4.17
Viruses	*Propionibacterium phage P100D*	0.4811	6.25

Table S10: Frequency distribution of strains within the SLST clusters.

SLST cluster	Phylotype*	No. of strains identified in SLST cluster
A	IA_1	58
B	IA_1	2
C	IA_1	18
D	IA_1	8
E	IA_1	6
F	IA_2	25
G	IC	6
H	IB	20
K	II	15
L	III	6
X (unclassified)		3
Sum		167

*Corresponding C. *acnes* phylotypes as identified in the literature.[16–18]

Table S11: *C. acnes* strains present in this study. SLST and Phylotype classifications for each strain, based on homology to sequenced strains.

Strain	SLST Assigned	SLST Group	Phylotype Assigned	Strain	SLST Assigned	SLST Group	Phylotype Assigned
SK137	C1	C	IA	523_PAVI	G1	G	IC
HL002PA1	F1	F	IA	KCOM 1861 (= ChDC B594)	K2	K	II
HL067PA1	F2	F	IA	PMH5	L1	L	III
HL005PA4	F1	F	IA	PMH7	L1	L	III
HL002PA2	A2	A	IA	PA_30_2_L1	98.35% F1	F	IA
HL025PA2	F1	F	IA	PA_15_1_R1	C1	C	IA
HL053PA1	C2	C	IA	PA_21_1_L1	H1	H	IB
HL110PA1	E1	E	IA	PA_15_2_L1	A1	A	IA
HL083PA1	C1	C	IA	PA_12_1_R1	99.17% F1	F	IA
HL072PA1	A6	A	IA	PA_12_1_L1	A1	A	IA
HL050PA1	99.79% F2, F7	F	IA	50.1.L1	A1	A	IA
HL046PA1	F1	F	IA	51.1.L1	A3	A	IA
HL005PA1	C2	C	IA	52.1.L4	A1	A	IA
HL050PA3	F1	F	IA	48.1.L1	A1	A	IA
HL005PA3	A1	A	IA	49.1.L1	A1	A	IA
HL045PA1	C2	C	IA	32.1.L2	A1	A	IA
HL005PA2	A1	A	IA	46.1.L1	A1	A	IA
HL086PA1	E4	E	IA	43.1.L1	A1	A	IA
HL038PA1	C1	C	IA	37.1.L1	A1	A	IA
HL050PA2	K4	K	II	44.1.L1	A1	A	IA
HL087PA3	F1	F	IA	NTS_2003_1719	D1	D	IA
HL020PA1	A1	A	IA	LRY_BL	H1	H	IB
HL013PA2	A1	A	IA	NTS_2004_10708	A26	A	IA
HL063PA1	A1	A	IA	NTS_31306190	A1	A	IA
HL036PA1	A2	A	IA	UBA1564	99.59% D1	D	IA
HL036PA2	A2	A	IA	UBA3960	A5	A	IA
HL027PA2	A1	A	IA	Dec-89	A1	A	IA
HL063PA2	F4	F	IA	Nov-78	A1	A	IA
HL043PA2	C1	C	IA	Nov-88	H1	H	IB
HL002PA3	A2	A	IA	Oct-43	K2	K	II
HL025PA1	D1	D	IA	09-263	A1	A	IA
HL110PA2	E3	E	IA	10-113	C5	C	IA
HL074PA1	C2	C	IA	Nov-49	K1	K	II
HL059PA1	F1	F	IA	9-Sep	K1	K	II
HL046PA2	A2	A	IA	CA17	K1	K	II
HL110PA4	K2	K	II	CA39	A2	A	IA
HL056PA1	C2	C	IA	CA51	A2	A	IA
HL030PA1	H1	H	IB	09-193	D1	D	IA
HL087PA1	F1	F	IA	09-322	D1	D	IA
HL083PA2	F3	F	IA	29-Sep	A1	A	IA

HL007PA1	C1	C	IA	UMB0211	H1	H	IB
HL043PA1	C1	C	IA	JCM 18919	L5	L	III
HL072PA2	A6	A	IA	Asn12	L2	L	III
HL082PA1	E5	E	IA	S2_005_002_R2_31	A1	A	IA
HL092PA1	E1	E	IA	P15-207	G1	G	IC
HL027PA1	F1	F	IA	P15-206	G1	G	IC
HL059PA2	99.79% F2, F3, F7, F11	F	IA	M13605	D1	D	IA
HL036PA3	A2	A	IA	P15-186	A1	A	IA
HL030PA2	F4	F	IA	T20574	99.79% K1	K	II
HL078PA1	B1	B	IA	T29350	K1	K	II
HL037PA1	F6	F	IA	T32516	H1	H	IB
HL013PA1	F1	F	IA	T35709	H1	H	IB
HL087PA2	A1	A	IA	T35743	H1	H	IB
J139	K8	K	II	T35877	X	X	X
J165	A1	A	IA	KCOM 1315	H1	H	IB
SK187	E2	E	IA	P15-014	H1	H	IB
HL096PA3	A1	A	IA	P15-021	A1	A	IA
HL096PA2	C1	C	IA	P15-159	D1	D	IA
HL103PA1	K1	K	II	UBA11121	A1	A	IA
HL097PA1	G1	G	IC	UBA9075	X	X	X
HL099PA1	C1	C	IA	T14076	H1	H	IB
266	A1	A	IA	C-45	99.59% A1	A	IA
6609	H1	H	IB	FDAARGOS_577	C1	C	IA
SK182	C1	C	IA	FDAARGOS_503	A1	A	IA
P.acn33	F1	F	IA	S2_006_000_R1_58	X	X	X
P.acn17	F5	F	IA	T28794	G1	G	IC
P.acn31	F4	F	IA	P15-181	D1	D	IA
PRP-38	G1	G	IC	P15-088	A1	A	IA
C1	A5	A	IA	NBRC 107605	A1	A	IA
FZ1/2/0	B1	B	IA	ATCC 11827	A1	A	IA
HL096PA1	C1	C	IA	TP-CU389	F1	F	IA
DSM 1897	A1	A	IA	MIT 1869-c3-2	H1	H	IB
HL042PA3	K2	K	II	MIT 1857-A1	K2	K	II
PA2	H1	H	IB	MIT 1879-a3-2	H1	H	IB
P6	H1	H	IB	MIT 1851-a4-2	H1	H	IB
JCM 18909	99.38% L5	L	III	DA10166-1	99.79% A1&A6	A	IA
JCM 18916	A2	A	IA	NLAE-zl-G260	H1	H	IB
JCM 18918	99.79% H1	H	IB	KPA171202	H2	H	IB
JCM 18920	K1	K	II	ATCC 6919	A1	A	IA
hdn-1	A1	A	IA	P15-231	A1	A	IA
HL411PA1	A1	A	IA	P15-077	99.79% A1&A6	A	IA
HL202PA1	K13	K	II	P15-089	F4	F	IA
HL201PA1	99.38% L7&L8	L	III	T36318	F4	F	IA
119_PAVI	A1	A	IA				

Table S12: p-values comparing α-diversities (Shannon indices) of the *C. acnes* SLST clusters. Significant results are highlighted in yellow. Early (T1-T2) and Late (T3-T5).

SLST cluster	Early Normal vs. Early Acne	Early Normal vs. Late Normal	Early Acne vs. Late Acne	Late Normal vs. Late Acne
A	0.213	0.007	0.919	0.099
B	1	1	1	0.782
C	0.266	0.41	0.4	0.83
D	0.864	0.027	0.57	0.047
E	0.481	0.841	0.702	0.791
F	0.279	0.044	0.543	0.418
G	0.242	0.406	0.697	0.096
H	1	0.187	0.791	0.04
K	NA	NA	0.2	NA

Table S13: p-values comparing β–diversities of the *C. acnes* SLST clusters.
Significant results are highlighted in yellow. Early (T1-T2) and Late (T3-T5).

SLST cluster	Early Normal vs. Early Acne	Early Normal vs. Late Normal	Early Acne vs. Late Acne	Late Normal vs. Late Acne
A	0.562	0.011	0.982	0.158
B	1	0.1	1	0.272
C	0.606	0.284	0.822	0.28
D	0.254	0.022	0.808	0.087
E	0.696	0.336	0.341	0.793
F	0.452	0.009	0.718	0.12
G	0.695	0.411	0.702	0.627
H	0.046	0.076	0.408	0.215
K	NA	NA	0.4	NA

Table S14: α-diversity comparison of SLST groups.
p-values comparing Shannon indices. Significant results are highlighted in yellow.

SLST group	Early Normal vs. Early Acne	Early Normal vs. Late Normal	Early Acne vs. Late Acne	Late Normal vs. Late Acne
99.79% A1&A6	1	1	1	1
A1	0.053	0.004	1	0.099
A2	0.786	0.498	0.509	0.47
A5	0.914	NA	0.083	NA
A6	NA	1	NA	1
B1	1	1	1	0.782
C1	0.168	0.664	0.505	0.83
C2	1	1	0.543	0.8
D1	0.65	0.103	0.545	0.178
E1	NA	1	NA	1
F1	0.766	0.736	0.101	0.534
F4	0.456	0.53	0.745	1
G1	0.242	0.406	0.697	0.096
H1	0.96	0.223	0.685	0.136
K1	NA	NA	0.1	NA
K2	NA	NA	1	NA

Table S15: β-diversity comparison of SLST groups.

p-values comparing β-diversities. Significant results are highlighted in yellow.

SLST group	Early Normal vs. Early Acne	Early Normal vs. Late Normal	Early Acne vs. Late Acne	Late Normal vs. Late Acne
99.79% A1&A6	0.457	0.4	0.46	0.111
A1	0.759	0.026	0.987	0.368
A2	0.11	0.488	0.146	0.344
A5	0.943	NA	0.346	NA
A6	NA	1	NA	0.18
B1	1	0.1	1	0.272
C1	0.483	0.435	0.716	0.217
C2	0.2	0.1	0.343	0.733
D1	0.284	0.413	0.503	0.469
E1	NA	0.333	NA	NA
F1	0.465	0.084	0.854	0.294
F4	0.641	0.009	0.683	0.014
G1	0.695	0.411	0.702	0.627
H1	0.043	0.075	0.375	0.248
K1	NA	NA	1	NA
K2	NA	NA	0.667	NA

Table S16: Network mapping of significant enzymes.

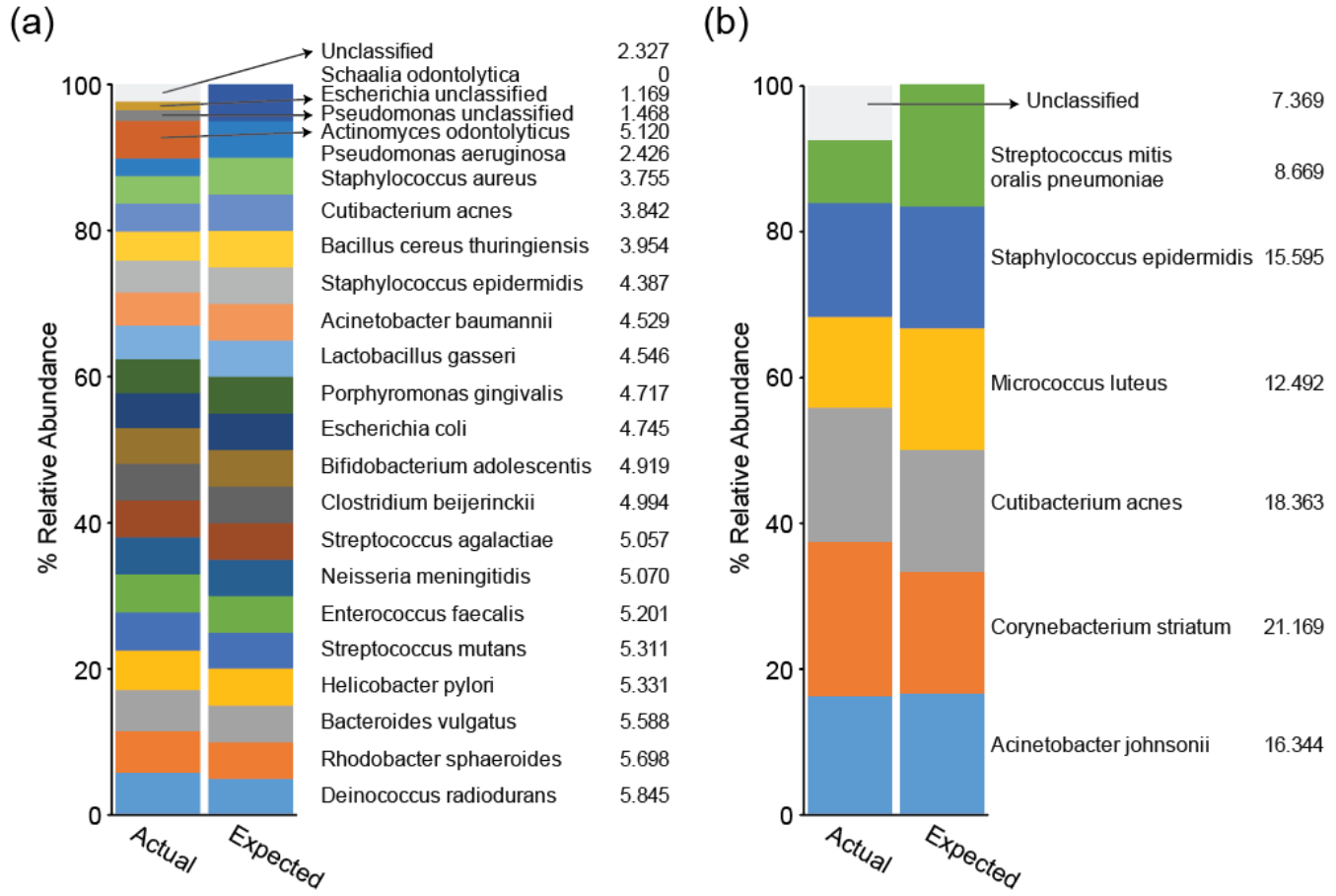
Pathway	Significant/ Total (%)	p-value	FDR	Enzymes
Porphyrin and chlorophyll metabolism	16/69 (23.19)	0.00000411	**0.000608**	K03405, K13542, K02303, K02233, K00595, K02228, K02232, K03404, K02227, K05936, K02189, K00231, K00798, K01749, K03399, K01845
Histidine metabolism	10/35 (28.57)	0.0000431	**0.00319**	K01468, K02501, K00765, K01814, K00274, K00817, K01523, K02500, K01745, K00013
Thiamine metabolism	6/23 (26.09)	0.00268	0.0991	K03153, K00946, K00941, K03149, K00788, K01662
Terpenoid backbone biosynthesis	6/23 (26.09)	0.00268	0.0991	K13787, K01823, K01770, K03527, K01662, K00919
Biosynthesis of amino acids	23/222 (10.36)	0.0133	0.394	K01817, K00930, K00611, K02501, K00765, K00620, K00215, K01814, K00817, K04092, K01609, K01523, K04517, K08094, K00058, K00145, K02500, K03786, K01778, K01647, K01735, K00812, K00013
Novobiocin biosynthesis	3/12 (25.00)	0.038	0.852	K00817, K04517, K00812
Riboflavin metabolism	4/22 (18.18)	0.0495	0.852	K14652, K11753, K00794, K01497
Phenylalanine, tyrosine and tryptophan biosynthesis	8/64 (12.50)	0.0514	0.852	K01817, K00817, K04092, K01609, K04517, K03786, K01735, K00812
Citrate cycle (TCA cycle)	7/53 (13.21)	0.0518	0.852	K18118, K00161, K00240, K01903, K01647, K00162, K00658
Propanoate metabolism	7/55 (12.73)	0.0613	0.87	K00140, K01847, K00016, K11263, K01692, K01903, K05606

Table S17: Unique and significant enzymes identified by network mapping.

Enzyme	p-value	FDR	Trend	Enzyme	p-value	FDR	Trend	Enzyme	p-value	FDR	Trend
K00013	0.00314	0.040671	↑ LA	K00941	0.00071	0.018978	↑ LA	K02232	0.000393	0.016152	↑ LA
K00016	0.00014	0.01165	↑ LA	K00946	0.000341	0.015448	↑ LA	K02233	0.000169	0.012454	↑ LA
K00058	0.001199	0.02415	↑ LA	K01468	0.000121	0.010868	↑ LA	K02303	8.51E-05	0.010868	↑ LA
K00140	1.59E-05	0.009657	↑ LA	K01497	0.002203	0.033221	↑ EA	K02500	0.001553	0.027847	↑ LA
K00145	0.001331	0.025793	↑ LA	K01523	0.000764	0.01951	↑ LA	K02501	0.000221	0.01385	↑ LA
K00161	8.88E-05	0.010868	↑ LA	K01609	0.000677	0.018897	↑ LA	K03149	0.001088	0.022725	↑ LA
K00162	0.003558	0.043971	↑ LA	K01647	0.002371	0.034523	↑ LA	K03153	8.54E-05	0.010868	↑ LA
K00215	0.000376	0.016103	↑ LA	K01662	0.002687	0.03691	↑ LA	K03399	0.00199	0.031051	↑ EN
K00231	0.001049	0.022198	↑ LA	K01692	0.00063	0.018647	↑ LA	K03404	0.000462	0.016268	↑ LA
K00240	0.000435	0.016152	↑ LA	K01735	0.002913	0.038714	↑ LA	K03405	2.29E-05	0.009657	↑ LA
K00274	0.000432	0.016152	↑ LA	K01745	0.001595	0.028139	↑ LA	K03527	0.000878	0.020529	↑ LA
K00595	0.000184	0.012778	↑ LA	K01749	0.001541	0.027731	↑ LA	K03786	0.001676	0.028549	↑ LA
K00611	0.000165	0.012454	↑ LA	K01770	0.000752	0.019327	↑ LA	K04092	0.000576	0.018162	↑ LA
K00620	0.00034	0.015448	↑ LA	K01778	0.0018	0.029713	↑ LA	K04517	0.000851	0.020474	↑ LA
K00658	0.003793	0.045342	↑ LA	K01814	0.000384	0.016103	↑ LA	K05606	0.002206	0.033221	↑ LA
K00765	0.000252	0.014778	↑ LA	K01817	0.000116	0.010868	↑ LA	K05936	0.000687	0.018897	↑ LA
K00788	0.001089	0.022725	↑ LA	K01823	0.000551	0.017626	↑ LA	K08094	0.001037	0.022172	↑ LA
K00794	0.001043	0.022178	↑ LA	K01845	0.003164	0.040862	↑ LA	K11263	0.000187	0.012778	↑ LA
K00798	0.001475	0.027245	↑ LA	K01847	2.64E-05	0.009657	↑ LA	K11753	0.000683	0.018897	↑ LA
K00812	0.002944	0.038907	↑ EN	K01903	0.000859	0.020529	↑ LA	K13542	8.46E-05	0.010868	↑ LA
K00817	0.000542	0.01758	↑ LA	K02189	0.000774	0.019556	↑ EN	K13787	0.000497	0.017079	↑ LA
K00919	0.003394	0.04265	↑ LA	K02227	0.000528	0.017502	↑ LA	K14652	0.000295	0.014931	↑ LA
K00930	0.000164	0.012454	↑ LA	K02228	0.000272	0.014931	↑ LA	K18118	2.35E-05	0.009657	↑ LA

Supplemental Figures

Supplemental Figure S1

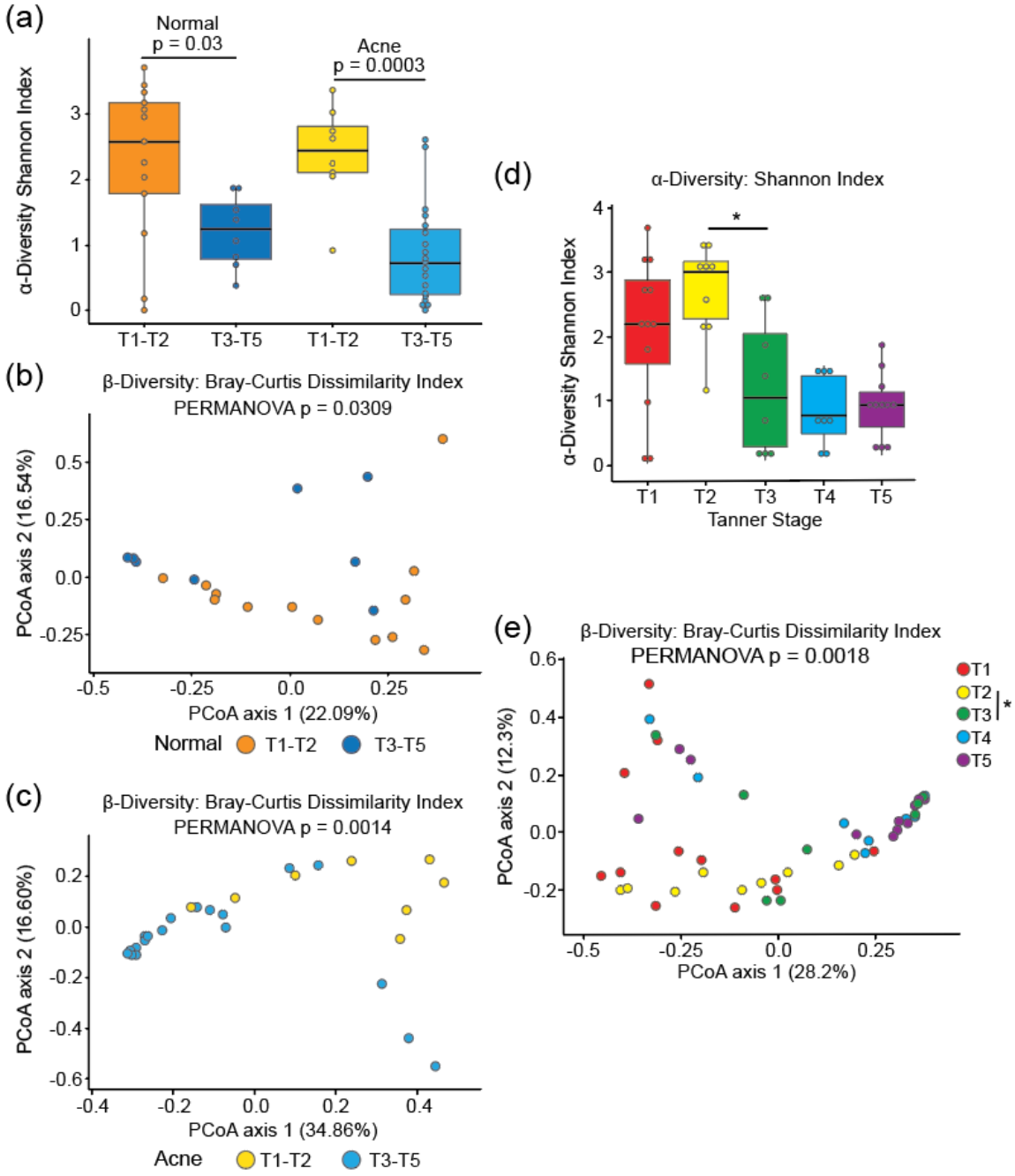


(c)

	NC 1 (%)	NC 2 (%)	NC 3 (%)	NC 4 (%)	NC 5 (%)	Average
Atopobium vaginae	0	0	0	0	0.026	0.005
Cutibacterium acnes	0	0	0	0.021	0	0.004
Escherichia coli	90.182	94.829	96.912	95.563	94.621	94.421
Escherichia unclassified	9.818	4.914	2.985	4.356	5.261	5.467
Pantoea unclassified	0	0.174	0.103	0	0.079	0.071
Prevotella bivia	0	0	0	0	0.013	0.003

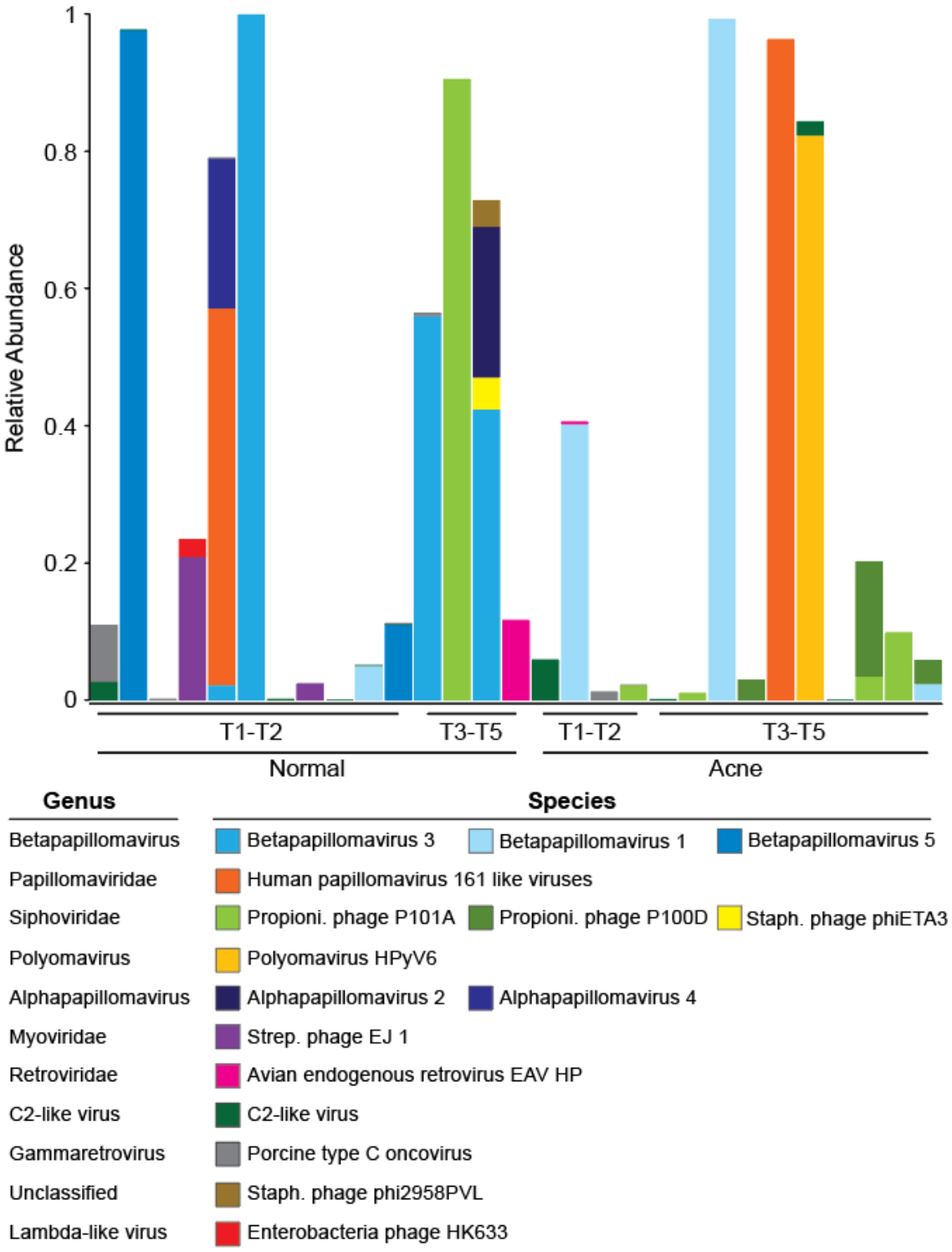
Supplemental Figure S1 Positive and negative control data. (a) 20 Strain Even Mix Genomic Material MSA-1002 (purchased from ATCC). Species level detection of MSA-1002 showing actual sequenced values and expected composition. MCC is composed of 20 expected species at 5% relative abundance, we detected 22 species and 1 unclassified group with percent abundance denoted next to name. *Schaalia odontolytica* present at 5% abundance in the control was not detected in our sequencing run. Three species not included in the mock community control were also detected at variable levels of abundance. (b) Skin Microbiome Genomic Mix MSA-1005 (purchased from ATCC). Species level detection of actual sequenced values compared to expected composition of six species at ~16.7% composition. All six species were detected in our sequencing run at variable levels, approximately 7% of the sample was unclassified at the species level. (c) Negative controls 1-5 were collected during sample collection. *Escherichia coli* was the most dominant species detected in the controls and is not a common skin commensal. Therefore, *Escherichia coli* and Escherichia unclassified were removed from patient sample data and further analysis. The negative control for DNA extraction (not shown above) did not yield any classifiable sequence and was removed from analysis completely (Table S3).

Supplemental Figure S2



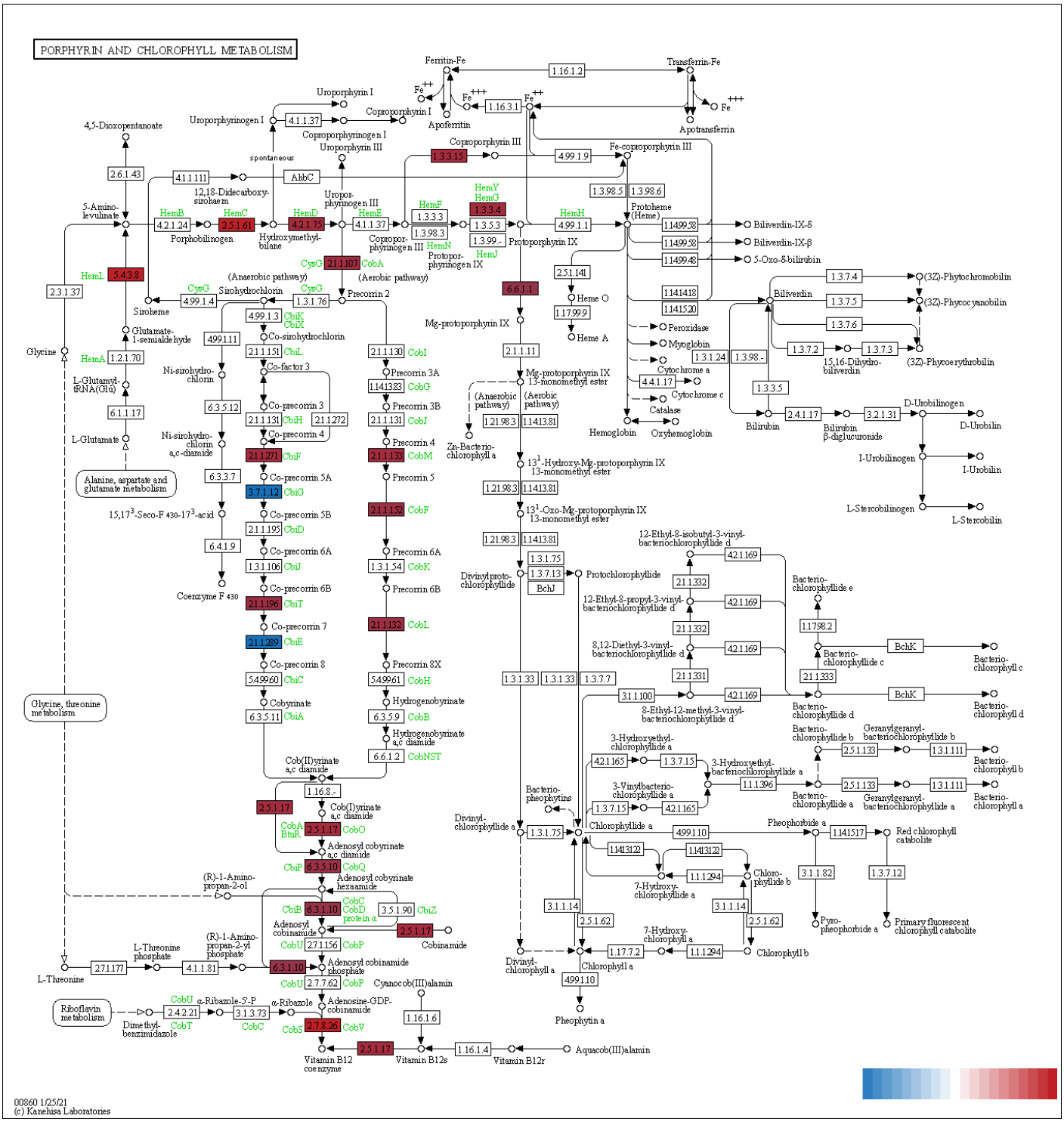
Supplemental Figure S2 α- and β-diversity shift significantly over puberty regardless of disease state. (a) Comparison of α-diversity by disease state between early (T1-T2) and late (T3-T5) Tanner stages, Wilcoxon rank-sum test, normal p = 0.03, acne p = 0.0003. (b) Comparison of β-diversity between early and late Tanner stages in normal skin, PERMANOVA, p = 0.0309. (c) Comparison of β-diversity between early and late Tanner stages in acne skin, PERMANOVA, p = 0.0014. All Shannon indices are positive, the minimum is 0.0074. (d) α-diversity of all samples stratified by Tanner stage; Wilcoxon rank sum test between consecutive Tanner stages; T2 vs T3 p=0.0079. (e) Principal Coordinates Analysis (PCoA) plot visualizing Bray Curtis Dissimilarity Index (β-diversity); PERMANOVA between all Tanner stages, p = 0.0018; PERMANOVA between consecutive Tanner stages, T2 vs T3 p=0.0336.

Supplemental Figure S3



Supplemental Figure S3 Viral composition by species and individual patient. Viral composition of each individual patient (29/48) by species, patients with no viral species were excluded from this graph. Each species in the legend is clustered by Genus. Proprioni. - Propionibacterium, Staph. - Staphylococcus, Strep. - Streptococcus.

Supplemental Figure S4



Supplemental Figure S4 KEGG Pathway map of *Porphyrin and Chlorophyll Metabolism* identifying location of the 16 significantly enriched KO terms. From Supplemental Table S14, significantly enriched KO terms associated with KEGG Pathway *Porphyrin and Chlorophyll Metabolism* (n = 16; colored) all map to the porphyrin synthesis/vitamin B usage arm of the pathway, with limited involvement of the chlorophyll arm of the pathway. Pathway obtained from https://www.genome.jp/kegg/ko.html after input of 16 KO terms. Scale: least to greatest expression- blue to red; white = no change in expression levels)

Supplemental Results

Non-human Sequences Summary

C. acnes dominated the bacterial landscape, but other skin commensals, *Staphylococcal* and *Streptococcal* spp. were detected in young and older pediatric skin, regardless of the presence of acne. We detected minimal *Corynebacterium* in our samples. This finding may be attributed to the fact that *Corynebacterium* prefers moist body sites over sebaceous body sites and that the relative abundance of *Corynebacterium* increases with increasing adult age.[24–26] Aside from bacteria, we also identified viruses that are uniquely present in either acne or normal skin in our pediatric cohort (Tables S8-S9, Fig. S3). While a thorough assessment of the viral composition in pediatric skin or in acne skin has not been done, viruses and bacteriophages present on healthy skin in adults including Papilloma and Pox viruses, and *Propionibacterium* and *Staphylococcus* phages.[27] Similar to bacteria, viruses display body-site specificity, although viral composition is much less stable overtime. Given that children have a more diverse and variable skin microbiome than adults and the transient nature of the virome, a larger, more comprehensive study in a pediatric population is needed to identify resident viruses and determine their contribution to skin health or diseases, such as acne.

Supplemental References:

1 Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child* 1970; **45**:13–23.

2 Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Arch Dis Child* 1969; **44**:291–303.

3 (CDER) C for DE and R. Acne Vulgaris: Developing Drugs for Treatment. 2005.

4 Andrews S (2010). FastQC: A quality control tool for high throughput sequence data [WWW Document]. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

5 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 2011; **17**:10.

6 Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**:2114–20.

7 Clarke EL, Taylor LJ, Zhao C, *et al.* Sunbeam: An extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome* 2019; **7**:46.

8 Truong DT, Franzosa EA, Tickle TL, *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015; **12**:902–3.

9 Oh J, Byrd AL, Deming C, *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature* 2014; **514**:59–64.

10 Meisel JS, Hannigan GD, Tyldsley AS, *et al.* Skin Microbiome Surveys Are Strongly Influenced by Experimental Design. *J Invest Dermatol* 2016; **136**:947–56.

11 Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001; **26**:32–46.

12 Albanese D, Donati C. Strain profiling and epidemiology of bacterial species from metagenomic sequencing. *Nat Commun* 2017; **8**:2260.

13 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; **9**:357–9.

14 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**:2078–9.

15 Scholz CFP, Jensen A, Lomholt HB, *et al.* A novel high-resolution single locus sequence typing scheme for mixed populations of Propionibacterium acnes in vivo. *PLoS One* 2014; **9**:e104199.

16 McDowell A. Over a Decade of recA and tly Gene Sequence Typing of the Skin Bacterium Propionibacterium acnes: What Have We Learnt? *Microorganisms* 2017; **6**:1.

17 Scholz CFP, Brüggemann H, Lomholt HB, *et al.* Genome stability of Propionibacterium acnes: A comprehensive study of indels and homopolymeric tracts. *Sci Rep* 2016; **6**:1–12.

18 Brüggemann H, Salar-Vidal L, Gollnick HPM, Lood R. A Janus-Faced Bacterium: Host-Beneficial and -Detrimental Roles of Cutibacterium acnes. *Front Microbiol* 2021; **12**:673845.

19 Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 2012; **28**:3211–7.

20 Franzosa EA, McIver LJ, Rahnavard G, *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods* 2018; **15**:962–8.

21 Dhariwal A, Chong J, Habib S, *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* 2017; **45**:W180–8.

22 Ewing B, Hillier LD, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175–85.

23 Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:186–94.

24 Grice EA, Kong HH, Conlan S, *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* 2009; **324**:1190–2.

25 Oh J, Conlan S, Polley EC, *et al.* Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med* 2012; **4**:77.

26 Dimitriu PA, Iker B, Malik K, *et al.* New insights into the intrinsic and extrinsic factors that shape the human skin microbiome. *MBio* 2019; **10**.

27 Hannigan GD, Meisel JS, Tyldsley AS, *et al.* The human skin double-stranded DNA virome: topographical and temporal diversity, genetic enrichment, and dynamic associations with the host microbiome. *MBio* 2015; **6**:e01578-15.