Supplemental Information

Experimental Procedures

Subject Recruitment and Sampling
To expand our previous metagenomic survey, we re-sampled 12 healthy volunteers from our original study (Oh et al., 2014). Recruitment criteria, sampling procedure, and sample processing were as described previously. Briefly, 7 males and 5 females adults <45 years without chronic skin diseases were sampled three times between June 2011 and May 2014. Sample collection was approved by the Institutional Review Board of the National Human Genome Research Institute (http://www.clinicaltrials.gov/ct2/show/NCT00605878) and all subjects provided informed consent. Longitudinal samples were collected such that the span between time 1 and time 2 was 10-30 months, while 5-10 weeks separated time 2 and time 3 (Figure S1B). This study design allowed the comparison of stability over a long and short time span. Individuals with a history of chronic medical conditions, including chronic dermatologic diseases, were excluded. 3 patients did report use of oral antibiotics between timepoint 1 and timepoint 2. However, in this study, antibiotic usage did not appear to induce discernible shifts in the overall diversity or structures of skin communities. Separate studies are necessary to fully understand the effects of oral antibiotics on the skin. Full sample characteristics are available in Table S1.

17 sites were sampled to represent the diverse physiological characteristics of skin and the sites of predilection for certain dermatologic disease (Figure S1A): dry (hypothenar palm, volar forearm), moist (antecubital crease, inguinal crease, interdigital web space, popliteal crease), sebaceous (alar crease, back, cheek, external auditory canal, glabella, manubrium, occiput, retroauricular crease), and foot (plantar heel, toenail, toe web space). To obtain sufficient DNA for metagenomic sequencing, most sites were sampled using a swab-scrape-swab procedure, exceptions include the external auditory canal where only a swab was used and the toenail where a clipping was taken. All samples were stored in lysis buffer at -80C until DNA extraction.

Sample Sequencing
Procedures for library generation, sequencing, and processing of longitudinal samples were as previously described (Oh et al., 2014). Briefly, Nextera library kits were used to generate Illumina libraries per manufacturer's instructions with the exception of increasing from 6 to 10 PCR cycles. Libraries were sequenced on an Illumina HiSeq at the NIH Intramural Sequencing Center to a target of 15 to 50 million clusters of 2 x 100bp reads. In total, for 12 individuals, 3 timepoints, we obtained 594 samples or 8.4 trillon reads (722 Gbp) of non-human, quality-filtered paired-end and singleton reads (median 17.9 million reads (1.4 Gbp) per sample). After human removal based on mapping to the hg19 human reference genome, all samples were processed to trim bases with quality score below 20 and remove reads less than 50 bp. To reduce computational burden, post quality control, samples with >20 million reads were subsampled to 10 million paired end reads, and singletons were discarded.

Taxonomic classification of skin species and diversity estimates.
Taxonomic classifications were performed as previously described (Oh et al., 2014), except we updated the viral database, incorporating all Refseq viral genomes as of 06.2015. The microbial reference genome database in total included 2342 bacterial, 389 fungal, 6009 viral, and 67 archaeal. Reads not matching hg19 + hg19 rRNA were mapped to this genome collection using bowtie2's —very-sensitive parameter retrieving the top 10 hits (Langmead and Salzberg, 2012). Reads mapping to multiple genomes were then reassigned using Pathoscope v1.0 (Francis et al., 2013), which uses a Bayesian framework to examine each read's sequence and mapping quality within the context of a global reassignment. Read hit counts were then normalized by genome length and scaled to sum to one. Coverages were calculated using the genomeCoverageBed tool in the Bedtools suite (Quinlan and Hall, 2010). Because very low abundance organisms are represented by few reads, they are more susceptible to misclassification than more abundant genomes. To reduce the effects of low abundance misclassifications, we used genome coverage cutoffs for relative abundance and diversity calculations; genomes were binned with coverage cutoffs of $\geq$ 1, 0.1, 0.01 or 0.001. A coverage cutoff of $\geq$ 1 was used for major analyses, a conservative number that produced classifications that most closely corresponded with the results from other common metagenomic classifiers (e.g., Metaphlan (Truong et al., 2015) or analysis using other methodologies like 16S rRNA and ITS gene sequencing (Oh et al., 2014). This number typically accounts for >99.9% of the community abundance. We used the Shannon

diversity index as well as species observed for diversity comparisons for bacterial classifications. All taxonomies were reconstructed to the species level, combining hits to multiple strain subtypes to reduce the potential for erroneous strain-calling (Table S2).

Strain tracking of dominant skin species

Strain tracking of the dominant skin commensals *Propionibacterium acnes* (Table S3) and *Staphylococcus epidermidis* (Table S5) was accomplished as described previously (Oh et al., 2014). Briefly, reference databases for *P. acnes* and *S. epidermidis* were compiled from all complete and draft genomes available on NCBI, 78 and 61, respectively. Whole genome alignment, with nucmer, was then used to identity the "core" region shared between all sequenced strains for a species. SNVs identified in these core regions were subsequently used to generate dendograms with PhyML 3.0. We then grouped strains into subtypes based on phylogenic distance, 12 for *P. acnes* and 14 for *S. epidermidis* (Figures S4B and S6B). Metagenomic reads were mapped to each species database with bowtie2 (-score-min L,-0.6,0.006, -k number of genomes) (Langmead and Salzberg, 2012) with zero tolerance for mismatches. The resulting alignment file was then processed with Pathoscope (-theta_prior 10 x 10^88) (Francis et al., 2013) to deconvolute multiple mapping reads. Accuracy of this strain-tracking approach was previously validated with extensive simulations (Oh et al., 2014).

Identification of SNVs in the P. acnes core

For each sample, coverage of the *P. acnes* core was calculated with samtools (Li, 2011) and genomecoveragebed (Quinlan and Hall, 2010). High average coverage nicely related to percent coverage of the *P. acnes* core (Figure S4). Back and manubrium samples had the highest *P. acnes* sequencing depth, so were selected for more extensive SNV analysis (Figure S4). Because *P. acnes* strains are shared across sites of an individual, these results can be extrapolated to the rest of the body. For SNV analysis, metagenomic reads were first mapped against the *P. acnes* core genome using bowtie2 (--very-sensitive). The resulting alignment file was sorted by samtools and then processed with GATK's IndelRealigner (McKenna et al., 2010) to minimize mismatches resulting from insertions or deletions in the reads with respect to the reference genome. The corrected alignment file was then analyzed with samtools and bcftools to identify possible variants (samtools mpileup -uD -q30 -Q30, bcftools view -Abvcg, vcfutils.pl varFilter -D99percentileofcoverage -d4 -1 .00001 -4 .00001). Parameters were selected to filter false positive polymorphisms that were a result of sequencing error, recent sequence duplications not found in the draft genome, strand bias, or end distance bias. Possible variants were then filtered with custom scripts to meet criteria previously described (Lieberman et al., 2014). Briefly, an alternate allele was only considered if it was supported by >2 reads with a minimum mapping quality of 30, had an allele frequency >3%, and fewer than 20% of reads supporting the SNV also mapped to an indel. With rarefaction curves of SNVs discovered over increasing read depths (Figure 5C), we found that 1 million reads, 40X coverage of the *P. acnes* core, was sufficient for variant discovery. Thus, to reduce computational burden only subsamples of 1 million reads were used for further analysis.

Pangenome analyses of dominant skin species

To identify the functional capacity of dominant species in our metagenomic samples, we followed the procedure illustrated in Figures S5A. First, 196,083 *P. acnes* nucleotide-coding sequences were downloaded from NCBI and 147,257 *S. epidermidis* sequences were extracted from Manatee annotations of the genomes. The IGS Analysis Engine was used for structural and functional annotation of the sequences. (http://ae.igs.umaryland.edu/cgi/index.cgi, Galens et al., 2011). Manatee was used to view annotations (http://manatee.sourceforge.net/). Genes were then clustered into non-redundant orthologs with usearch (-cluster_fast -id 0.80 -centroids) (Edgar, 2010). To validate accuracy of the clustering, we verified the presence of 13 single copy marker genes (Greenblum et al., 2015). Singletons, clusters composed of a single sequence, were then filtered based on previously established criteria (Lefebure and Stanhope, 2007). Briefly, singletons were excluded if they 1) were shorter than 150 nucleotides, 2) were flagged as low complexity by Prinseq (Schmieder and Edwards, 2011), or 3) overlapped the beginning or end of a contig. 4) had a blast hit to a cluster at -e 1e-10. Based on this criteria 359 *P. acnes* and 874 *S. epidermidis* singletons were removed, leaving 3,774 and 5,627 gene clusters respectively (Table S4 and Table S5, respectively). Gene accumulation curves for these clusters mirrored previous pangenome studies for *P. acnes* (Tomida et al., 2013) and *S. epidermidis (*Conlan et al., 2012*)*. The curves showed that new genes discovered with additional genomes and the pangenome followed a power law curve, while core genome size fit an exponential decay curve (Figure S5B,C, Figure S6C,D). These gene clusters were then annotated by BLASTx against the KEGG database. To identify the functional capacity of a sample, reads were mapped to each of the gene cluster databases using bowtie2 (--very-sensitive). A gene was subsequently considered present only when 40% of its length was covered with reads. This criteria reduces gene calling due to spuriously mapped reads or

reads from orthologs of closely related species (Zhu et al., 2015). Average coverage of each gene was calculated with samtools (Li, 2011) and then normalized by the average coverage of 13 single copy marker genes (Greenblum et al., 2015) to yield a copy number estimate.

Statistics

All statistical analyses were performed in the R software. Data are represented as mean ± standard error of the mean unless otherwise indicated. Spearman correlations of non-zero values were used for all correlation coefficients. Site characteristics were treated as separate groups where indicated based on spatial physiological differences between these different body niches (Grice et al., 2009). Supervised random forest models to identify discriminatory taxa and modules were implemented with the randomForest package in R (Liaw and Wiener, 2002). For all boxplots, black center lines represent the median and box edges the first and third quartiles. The nonparametric Wilcoxon rank-sum test was used to determine statistically significant differences between microbial populations. Unless otherwise indicated, P-values were adjusted for multiple comparisons using the p.adjust function in R using method = "fdr". Statistical significance was ascribed to an alpha level of the adjusted P-values ≤ 0.05. Similarity between samples was assessed using the Yue–Clayton theta or Jaccard similarity index with relative abundances of species, sub-strains, or shared genomic variants. The theta coefficient assesses the similarity between two samples based on (1) number of features in common between two samples, and (2) their relative abundances with $\theta = 0$ indicating totally dissimilar communities and $\theta = 1$ identical communities (Yue and Clayton 2005). As θ takes into account species abundance, it is less susceptible to low-abundance species whose classifications are less robust. The Jaccard similarity index is a metric defined by the union of the species occurring between two samples. To avoid repeated measures, samples belonging to an individual were averaged before statistical comparisons between site characteristic when using summary metrics such as means, diversity, or theta indices.

Supplemental References

Conlan, S., Mijares, L.A., NISC Comp Seq Program , Becker, J., Blakesley, R.W., Bouffard, G.G., Brooks, S., Coleman, H., Gupta, J., Gurson, N.*, et al.* (2012). Staphylococcus epidermidis pan-genome sequence analysis reveals diversity of skin commensal and hospital infection-associated isolates. Genome biology *13*, R64.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Francis, O.E., Bendall, M., Manimaran, S., Hong, C., Clement, N.L., Castro-Nallar, E., Snell, Q., Schaalje, G.B., Clement, M.J., Crandall, K.A.*, et al.* (2013). Pathoscope: species identification and strain attribution with unassembled sequencing data. Genome research *23*, 1721-1729.

Galen, K., Orvis, J. Daugherty, S. Creasy, H.H., Angiuoli, S., White, O., Wortman, J., Mahurkar, A., Giglio, M.G., (2011). The IGS Standard Operating Procedure for Automated Prokaryotic Annotation. Stand Genomic Sci 4:244-251.

Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. Cell *160*, 583-594.

Grice, E.A., Kong, H.H., Conlan, S., Deming, C.B., Davis, J., Young, A.C., NISC Comp Seq Program, Bouffard, G.G., Blakesley, R.W., Murray, P.R.*, et al.* (2009). Topographical and temporal diversity of the human skin microbiome. Science *324*, 1190-1192.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987-2993.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2*, 18-22.

Lieberman, T.D., Flett, K.B., Yelin, I., Martin, T.R., McAdam, A.J., Priebe, G.P., and Kishony, R. (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. Nature genetics *46*, 82-87.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research *20*, 1297-1303.

Oh, J., Byrd, A.L., Deming, C., Conlan, S., Program, NISC Comp Seq Program, Kong, H.H., and Segre, J.A. (2014). Biogeography and individuality shape function in the human skin metagenome. Nature *514*, 59-64.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863-864.

Tomida, S., Nguyen, L., Chiu, B.H., Liu, J., Sodergren, E., Weinstock, G.M., and Li, H. (2013). Pan-genome and comparative genome analyses of propionibacterium acnes reveal its genomic diversity in the healthy and diseased human skin microbiome. mBio *4*, e00003-00013.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Sholz, M., Weingart, G., Passolli, E., Tett, A., Huttenhower, C., Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling.Nature methods 12:902-903.

Yue, J. C., Clayton, M. K. (2005) A similarity measure based on species proportions. Comm. Statist. Theory Methods 34, 2123–2131.

Zhu, A., Sunagawa, S., Mende, D.R., and Bork, P. (2015). Inter-individual differences in the gene content of human gut bacterial species. Genome biology *16*, 82.