LETTER

doi:10.1038/nature12171

Topographic diversity of fungal and bacterial communities in human skin

Keisha Findley[1], Julia Oh[1], Joy Yang[1], Sean Conlan[1], Clayton Deming[1], Jennifer A. Meyer[1], Deborah Schoenfeld[2], Effie Nomicos[2], Morgan Park[3], NIH Intramural Sequencing Center Comparative Sequencing Program†, Heidi H. Kong[2]* & Julia A. Segre[1]*

Traditional culture-based methods have incompletely defined the microbial landscape of common recalcitrant human fungal skin diseases, including athlete's foot and toenail infections. Skin protects humans from invasion by pathogenic microorganisms and provides a home for diverse commensal microbiota[1]. Bacterial genomic sequence data have generated novel hypotheses about species and community structures underlying human disorders[2–4]. However, microbial diversity is not limited to bacteria; microorganisms such as fungi also have major roles in microbial community stability, human health and disease[5]. Genomic methodologies to identify fungal species and communities have been limited compared with those that are available for bacteria[6]. Fungal evolution can be reconstructed with phylogenetic markers, including ribosomal RNA gene regions and other highly conserved genes[7]. Here we sequenced and analysed fungal communities of 14 skin sites in 10 healthy adults. Eleven core-body and arm sites were dominated by fungi of the genus *Malassezia*, with only species-level classifications revealing fungal-community composition differences between sites. By contrast, three foot sites—plantar heel, toenail and toe web—showed high fungal diversity. Concurrent analysis of bacterial and fungal communities demonstrated that physiologic attributes and topography of skin differentially shape these two microbial communities. These results provide a framework for future investigation of the contribution of interactions between pathogenic and commensal fungal and bacterial communities to the maintainenace of human health and to disease pathogenesis.

Since Hippocrates first described oral candidiasis in 400 BC, scientists have sought to explore the roles that commensal and pathogenic fungi and microbial communities have in human health and disease[8,9]. Culture-based studies have reported *Malassezia, Rhodotorula, Debaromyces, Cryptococcus* and, in some sites, *Candida* as fungal skin commensals[10]. Cutaneous fungal infections affect 29 million North Americans, but the role of dermatophytes in common toenail infections can be difficult to characterize using culture-based studies[11]. For other common skin disorders, such as seborrheic dermatitis (dandruff), fungal involvement remains incompletely understood[12,13]. Difficulty in establishing growth conditions[14,15] contribute to challenges to rapidly identify and direct treatment against pathogenic fungi.

To compare culture- and DNA-sequence-based identification of human skin-associated fungi, we obtained parallel samples from four skin sites of adult healthy volunteers (Supplementary Fig. 1 and Supplementary Table 1). We characterized isolates by morphological features and molecular markers. In total, we cultured more than 130 fungal isolates: 62 *Malassezia* (species *globosa, restricta* and *sympodialis*[13,16]), 25 *Penicillium* (species *chrysogenum* and *lanosum*) and 19 *Aspergillus* (species *candidus, terreus* and *versicolor*) (Supplementary Table 2). Five or fewer *Alternaria, Candida, Chaetomium, Chrysosporium, Cladosporium, Mucor, Rhodotorula* and *Trichophyton* isolates were cultured.

To explore fungal diversity with culture-independent methods, we prepared DNA from clinical swabs, and polymerase chain reaction (PCR)-amplified and sequenced two phylogenetic markers within the ribosomal RNA region: 18S rRNA and the intervening internal transcribed spacer 1 (ITS1) region[7,17,18]. We generated a custom ITS1 database based on sequences deposited in GenBank to classify sequences to genus level with greater than 97% accuracy (Supplementary Table 3). 18S rRNA sequences were classified using the SILVA database[19]. We determined the relative abundance of fungal genera of the occiput (back of head), nare (nostril), plantar heel and retroauricular crease (behind the ear). The genus *Malassezia* predominated in the retroauricular crease, nare and occiput; this was consistent across 18S rRNA and ITS1-characterized samples. Plantar heel was the most diverse site with representation of *Malassezia, Aspergillus, Cryptococcus, Rhodotorula, Epicoccum* and others (Supplementary Fig. 2). ITS1 sequencing enabled greater genus-level taxonomic resolution, reflecting the specificity of the genomic region and richness of the molecular database. Based on technical and analytic advantages, we selected the ITS1 region for subsequent sequencing and analyses of fungal diversity.

We generated more than 5 million ITS1 sequences from 10 healthy volunteers (from each of whom 14 skin-site samples were taken) (Supplementary Table 4). Both Ascomycetous and Basidiomycetous fungi were identified as normal skin flora. The genus *Malassezia* predominated at all 11 core-body and arm sites: antecubital fossa, back, external auditory canal, glabella, hypothenar palm, inguinal crease, manubrium, nare, occiput, retroauricular crease and volar forearm (Fig. 1). We explored *Malassezia* species-level resolution with a taxonomic data set that we developed based on reference ITS1 sequences and our human-associated *Malassezia* isolates. Pairwise comparisons of these *Malassezia* ITS1 sequences showed sequence identity within species to be greater than 91%, and identity between species to be 70 to 88%. These *Malassezia* sequences served as references within the phylogenetic pplacer[20] program to classify approximately 80 to 90% of *Malassezia* sequences per skin site to species level. Species-level identification revealed fungal specificity between body sites (Fig. 1). *M. restricta* predominated in external auditory canal, retroauricular crease and glabella, and *M. globosa* predominated on back, occiput and inguinal crease. Sites such as nares, antecubital fossa, volar forearm and hypothenar palm were characterized by multiple species (*M. restricta, M. globosa* and *M. sympodialis*). In total, we identified 11 of the 14 known *Malassezia* species among skin sites, suggesting that human skin is colonized with a wide range of *Malassezia*. Based on species-level resolution, we observed that fungal diversity is more dependent on body site than individual subject. ITS1 sequences also

[1]Genetics and Molecular Biology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [2]Dermatology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [3]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland 20852, USA.

*These authors contributed equally to this work.

†Lists of participants and affiliations appear at the end of the paper.

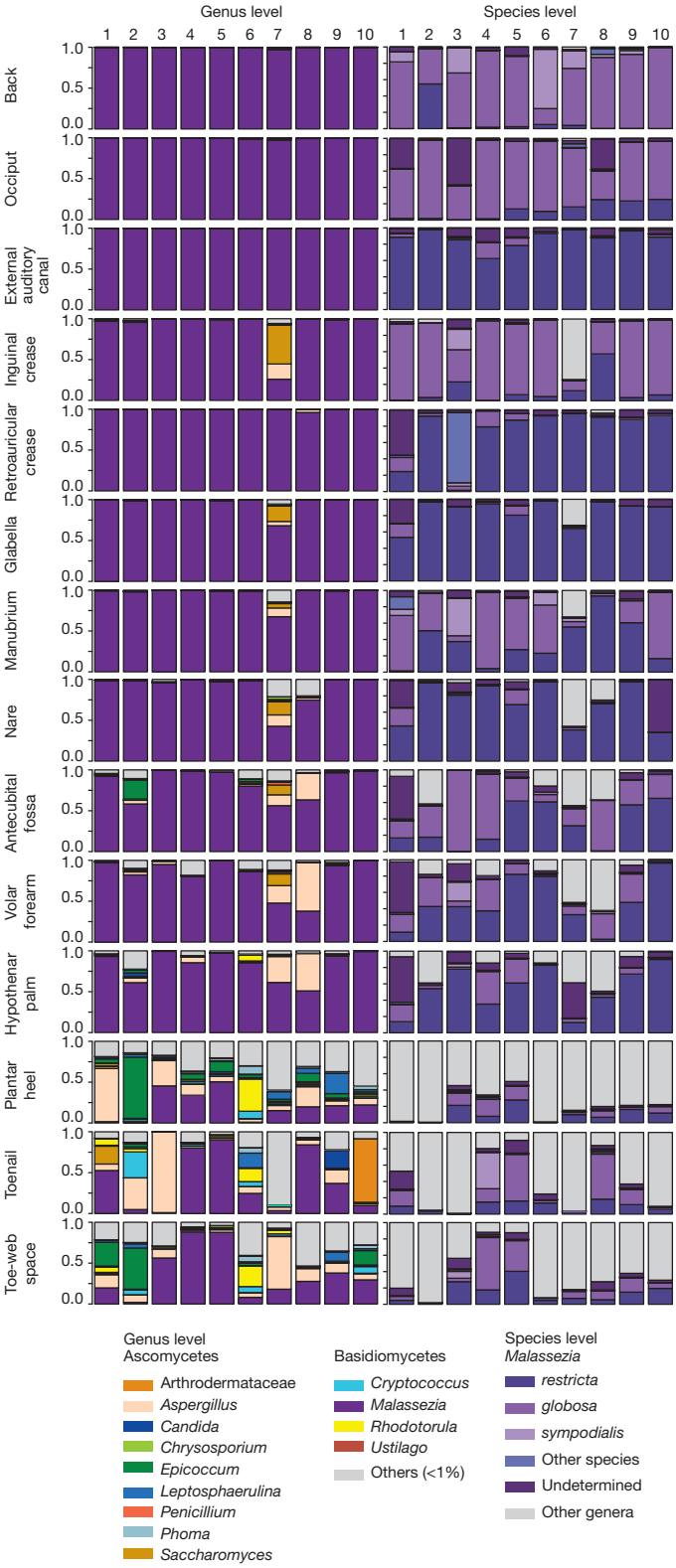

Figure 1 | Relative abundance of fungal genera and *Malassezia* species at different human skin sites. Fungal diversity of individual body sites of healthy volunteers (1–10) was taxonomically classified at the genus level, with further resolution of *Malassezia* species. For all body sites, the left side of the body was used, except for the right toenail of healthy volunteer 7.

matched *Candida* species *tropicalis, parapsilosis* and *orthopsilosis*, and *Cryptococcus* species *flavus, dimennae* and *diffluens*, which are considered to be part of the normal human flora and to be possible pathogens in wounds or immunocompromised patients[14].

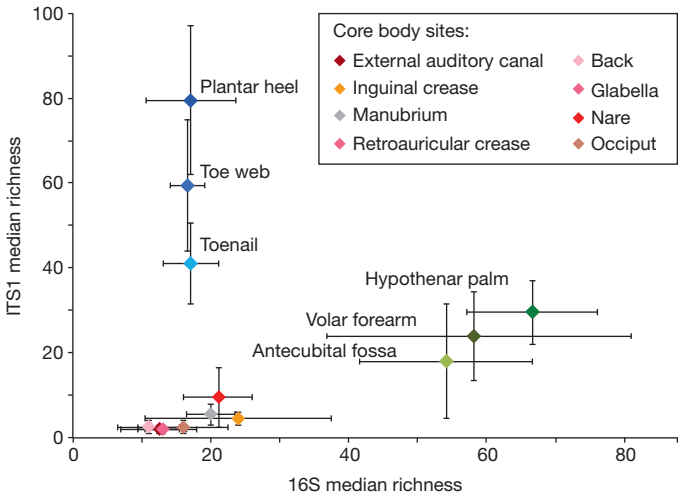

Figure 2 | Median richness of fungal and bacterial genera. Median taxonomic richness (or number of observed genera) of fungal and bacterial genera at 14 body sites. Error bars represent the median absolute deviation. The values for the core body sites retroauricular crease and glabella are identical and are therefore represented by a single data point on the graph and a shared colour in the key.

Substantially greater diversity was observed on three foot sites (plantar heel, toenail and toe web), in both the number of genera observed and the variation between individuals (Supplementary Fig. 3 and Supplementary Table 5). The fungal profile of one of the subjects (who we will refer to as healthy volunteer 7) was notably more diverse than other participants (Fig. 1). Healthy volunteer 7 completed a 2-month course of oral antifungal medication for a toenail infection 7 months before sampling. The remaining healthy volunteers reported no use of either oral or topical antifungal medication for at least 2 years before sampling. Healthy volunteer 7 is an outlier, but the additional genera that were identified (for example, *Aspergillus* and *Saccharomyces*) show that skin is capable of harbouring high fungal diversity. Although microbial sequencing is unable to determine causation, these data may suggest either that fungal community imbalance is associated with recurrent toenail infections or that alterations in fungal skin communities persist even 7 months after discontinuing antifungal medications. In comparison with culture-based analysis, ITS1 sequencing can provide a more complete view of the diversity of commensal microbiota, and also of potentially pathogenic microbiota.

To quantify and compare community similarity and taxonomic richness of skin sites, we assigned fungal sequences to taxonomic units based on genus-level phylogeny rather than percent sequence identity to obviate the high variation noted between species[21] with the latter metric. Plantar heel was the most complex fungal site (median richness of approximately 80 genera), and other foot sites had the next highest diversity (toe web and toenail, with approximately 60 and 40 genera, respectively; Fig. 2, Supplementary Table 6). Arm sites showed intermediate richness, ranging from 18 to 32 genera and core-body sites exhibited much lower richness, ranging from 2 to approximately 10 genera. Thus, regional location is a strong determinant of fungal richness. As observed in skin bacterial studies, left–right similarity within an individual was greater than between different individuals at the same body site (Supplementary Fig. 4 and Supplementary Table 7). To determine the temporal stability of the fungal microbiome, 6 healthy volunteers returned 1 to 3 months after initial sampling. Sites that showed *Malassezia* predominance at initial sampling displayed the same genus- and species-level predominance with strong community structure stability (Supplementary Figs 5 and 6, and Supplementary Table 8). Foot sites continued to show high diversity, perhaps reflecting frequent environmental exposure.

To explore the relationship between skin-associated fungi and bacteria, we sequenced the 16S rRNA gene from the same clinical samples. Consistent with previous studies[22,23], bacteria on healthy human skin

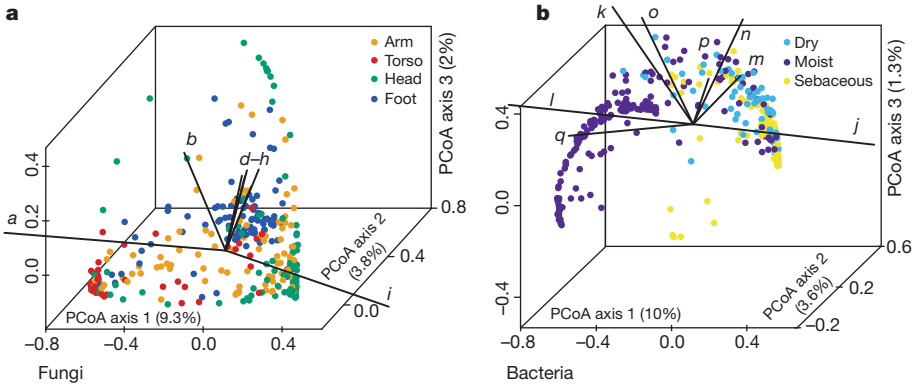

Figure 3 | Forces that shape fungal and bacterial communities. **a**, **b**, Principal coordinates analysis (PCoA) of the degree of fungal- and bacterial-community similarity at 14 body sites, based on predominant genera and species. Variation in fungal communities segregated strongly according to site location, with arm, foot, head and core-body sites forming discrete groups (**a**). Bacterial community structure was more dependent on site physiology (**b**). Axes that most significantly contribute to variation and their relative lengths are shown (these are defined in Supplementary Fig. 10 legend). For fungi, *M. restricta* (*i*; $\rho = 0.92$) and *M. globosa* (*a*; $\rho = -0.79$) are primary and opposing drivers of variation based on Spearman correlation with PCoA axis 1. For bacteria, *Propionibacterium* (*j*; $\rho = 0.95$) contributes to sebaceous site variation, whereas *Corynebacterium* (*l*; $\rho = -0.74$) and *Turicella* (*q*; $\rho = -0.56$) are the greatest contributors at moist sites based on Spearman correlation with PCoA axis 1.

were predominantly *Propionibacterium*, *Corynebacterium* and *Staphylococcus* (Supplementary Fig. 7). Similar to other moist skin sites, the toenail bacteria (not surveyed previously) were predominantly *Corynebacterium* and *Staphylococcus*. Interestingly, although healthy volunteer 7 was an outlier in terms of fungal diversity and membership, the bacterial profile was normal with respect to taxonomic and ecological measures of diversity (Supplementary Fig. 7). Directed studies may help elucidate how antibacterial and antifungal therapies perturb fungal and bacterial communities.

Bacterial and fungal richness were not linearly correlated, but were instead grouped into discrete clusters of sites from the same region; arm, foot and core-body (sites from the same regions had similar bacterial and fungal richness) (Fig. 2 and Supplementary Fig. 8). Arm sites displayed markedly greater bacterial diversity and lower fungal diversity than the foot and core-body sites. In contrast, foot sites displayed markedly greater fungal diversity with lower bacterial diversity than the arm and core-body sites. Core-body sites clustered together, and showed both lower bacterial and lower fungal diversity. These data reveal that the skin microbiome is complex and suggest that different characteristics shape skin bacterial and fungal communities.

Using principal coordinates analysis of community structure, we explored properties that may shape bacterial and fungal communities differentially. Consistent with previous studies[5], bacterial communities varied in the proportion of lipophilic bacteria (*Corynebacterium, Propionibacterium* and *Turicella*) and staphylococcal species, and grouped based on skin physiology into sebaceous, moist and dry sites (Fig. 3 and Supplementary Fig. 9). In contrast, fungal communities were segregated more clearly by site location than physiology, with foot, arm, head and torso sites forming discrete groups. Different *Malassezia* species drive variation in arms, torso and head, whereas a wide range of fungal genera drive variation in feet (Fig. 3 and Supplementary Fig. 9). Co-occurrence analysis of foot sites, based on Spearman correlation of fungal and bacterial taxonomic relative abundances (Supplementary Fig. 10), provided a preliminary evaluation of major fungal–bacterial associations. For example, a group of primarily Actinobacteria was anti-correlated with resident Ascomycota and Basidiomycota in contrast to a group of primarily Firmicutes and Proteobacteria that was positively correlated with these fungal taxa.

We observed that 20% (12 out of 60) of our study participants had clinical involvement (plantar-heel scaling, toe web scaling or toenail changes), consistent with possible fungal infections (Supplementary Table 1). Of the subjects with observed clinical involvement, positive mycological cultures were obtained from toenails (two samples) and plantar heel (one sample) (*Trichophyton, Penicillium* and *Aspergillus*). These observations are similar to the results of larger studies, which report signs of clinical involvement in up to 60% of feet of healthy individuals, and find 2 to 25% of cultures to be positive for fungi. The wide variation in prevalence of clinical involvement and positive fungal cultures is dependent on several factors, including population and climate[24,25]. As an initial inquiry into the aetiology of foot fungal disorders, we examined how observed clinical status (involved or uninvolved) at plantar heel, toe web or toenail affected fungal community structure. For uninvolved sites, interpersonal variation in community structure was highly consistent across all foot sites. In contrast, for sites with observed clinical involvement, similarity of community structure was much higher for plantar heel but much lower for toenail (Fig. 4). These data may suggest that there is a common fungal community shared among individuals with plantar-heel involvement, and high fungal diversity underlying toenail infections, but further studies are needed. These data sets can now be used to inform future clinical studies that examine microbial community shifts associated with fungal infections.

This systematic study clearly demonstrates that human skin surfaces are complex ecosystems, providing diverse environments for microorganisms that inhabit our bodies. Different factors determine bacterial and fungal communities, depending on the physiological properties of the skin. *Malassezia* species predominate on all core-body and arm sites. In contrast, foot sites display tremendous fungal diversity and markedly lower stability over time. Microbial community instability may provide an opportunity for potentially pathogenic microbes to establish disease. Plantar heels, toe webs and toenails are common sites of

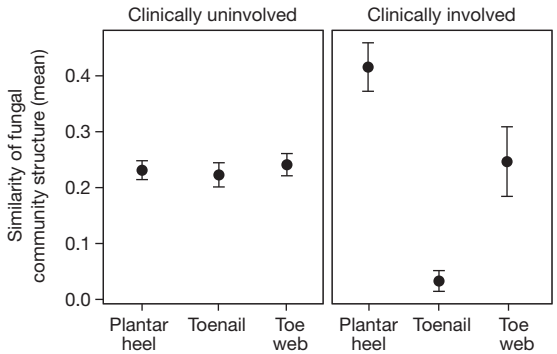

Figure 4 | Clinical involvement alters shared fungal community structure. Community structure measures the type and relative abundance of each genus. A value of 1 implies identical community structure and 0 implies dissimilar structures. Among uninvolved foot sites, community structure is fairly consistent at plantar heel, toenail and toe web sites. For involved sites, plantar heel has much greater shared community structure and toenails have much lower shared community structure. Error bars represent the s.e.m.

recurrent human fungal disease, which can be recalcitrant to treatment. This study also investigated fungal diversity at sites of predilection for other skin disorders, including seborrheic dermatitis, tinea cruris and subtypes of atopic dermatitis. With genomic advances, such as shotgun metagenomic sequencing, it is possible to begin to address interactions between microbes (bacterial–fungal, bacterial–bacterial, fungal–fungal) residing in these complex environments. The role of fungal commensals in educating the human immune system is gaining new appreciation in intestinal disease[26]. Further studies of healthy skin and dermatologic disorders are needed to explore these host–microbe interactions. In addition, antifungal medications, including azoles, echinocandins and amphotericin B, have potentially serious side effects such as liver or kidney damage[27]. Therefore, new treatment approaches are required to strategically target microbial dysbiosis and to combat the increasingly observed resistance against our current arsenal of antimicrobial therapies.

METHODS SUMMARY

Subject recruitment and sample preparation. This study 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 written informed consent before participation. For fungal culturing studies, skin was scraped with a surgical blade and placed directly in media. For sequencing studies, swabs from skin and environmental negative controls were placed in MasterPure Yeast DNA purification kit (Epicentre) lysis buffer augmented with lysozyme. Proteinase K (Invitrogen) was added to pre-digest toenail clippings and incubated overnight with shaking at 55 °C. Steel beads (5 mm in diameter) were added to mechanically disrupt fungal cell walls using Tissuelyser (Qiagen) for 2 min at 30 Hz and then using the PureLink Genomic DNA Kit (Invitrogen). For 18S rRNA sequencing, each DNA was amplified with SR6 (5′-TACCTGG TTGATTCTGC) and SR1R (5′-TGTTACGACTTTTACTT) primers. For ITS1 sequencing, each DNA was amplified with adaptor plus 18SF (5′-GTAA AAGTCGTAACAAGGTTTC) and 5.8S1R plus barcode primers (5′-GTTCA AAGAYTCGATGATTCAC). For 16S rRNA sequencing, each DNA was amplified with adaptor plus V1_27F (5′-AGAGTTTGATCCTGGCTCAG) and V3_534R plus barcode primers (5′-CAGCACGCATTACCGCGGCTGCTGG).

Sequence classification and analyses. Sequences were pre-processed to remove primers and barcodes. Possible chimaeras were identified with UCHIME in mothur[21,28]. ITS1 sequences were classified to genus level with BLAST (basic local-alignment search tool) and the *k*-nearest neighbour algorithm in mothur. 18S rRNA sequences were classified using the SILVA v108 database[19]. 16S rRNA sequences were classified to genus level using the RDP classifier and training set 6[29]. We curated and aligned a reference library of *Malassezia* ITS1 type-strain sequences retrieved from GenBank augmented with those from our human-associated fungal cultures. This curated library was used as a reference to phylogenetically place and classify ITS1 sequences to species level within pplacer[20]. Sequence placement on the reference tree was visualized in Archaeopteryx using the 'guppy' command for classifications with a likelihood score of greater than or equal to 0.65. Full methods are found in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

Received 5 December 2012; accepted 9 April 2013.
Published online 22 May 2013.

1. Marples, M. (ed.) *The Ecology of the Human Skin* (Bannerstone House, 1965).
2. Grice, E. A. & Segre, J. A. The human microbiome: our second genome. *Annu. Rev. Genomics Hum. Genet.* **13,** 151–170 (2012).
3. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
4. Pflughoeft, K. J. & Versalovic, J. Human microbiome in health and disease. *Annu. Rev. Pathol.* **7,** 99–122 (2012).
5. Peleg, A. Y., Hogan, D. A. & Mylonakis, E. Medically important bacterial–fungal interactions. *Nature Rev. Microbiol.* **8,** 340–349 (2010).
6. Dollive, S. *et al.* A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol.* **13,** R60 (2012).
7. James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443,** 818–822 (2006).
8. Ghannoum, M. A. *et al.* Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* **6,** e1000713 (2010).
9. Paulino, L. C., Tseng, C. H., Strober, B. E. & Blaser, M. J. Molecular analysis of fungal microbiota in samples from healthy human skin and psoriatic lesions. *J. Clin. Microbiol.* **44,** 2933–2941 (2006).
10. Roth, R. R. & James, W. D. Microbial ecology of the skin. *Annu. Rev. Microbiol.* **42,** 441–464 (1988).
11. Bickers, D. R. *et al.* The burden of skin diseases: 2004 a joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. *J. Am. Acad. Dermatol.* **55,** 490–500 (2006).
12. Gaitanis, G., Magiatis, P., Hantschke, M., Bassukas, I. D. & Velegraki, A. The *Malassezia* genus in skin and systemic diseases. *Clin. Microbiol. Rev.* **25,** 106–141 (2012).
13. Saunders, C. W., Scheynius, A. & Heitman, J. *Malassezia* fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. *PLoS Pathog.* **8,** e1002701 (2012).
14. Larone, D. H. *Medically Important Fungi: A guide to identification* (ASM Press, 2002).
15. St-Germain, G. & Summerbell, R. *Identifying Fungi: A Clinical Laboratory Handbook* (Star Publishing Company, 2011).
16. Gioti, A. *et al.* Genomic insights into the atopic eczema-associated skin commensal yeast *Malassezia sympodialis*. *mBio* **4,** e00572–12 (2013).
17. Bruns, T. D. *et al.* Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. *Mol. Phylogenet. Evol.* **1,** 231–241 (1992).
18. Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl Acad. Sci. USA* **109,** 6241–6246 (2012).
19. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–D596 (2013).
20. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11,** 538 (2010).
21. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
22. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–1697 (2009).
23. Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324,** 1190–1192 (2009).
24. Cohen, A. D., Wolak, A., Alkan, M., Shalev, R. & Vardy, D. A. Prevalence and risk factors for tinea pedis in Israeli soldiers. *Int. J. Dermatol.* **44,** 1002–1005 (2005).
25. Perea, S. *et al.* Prevalence and risk factors of tinea unguium and tinea pedis in the general population in Spain. *J. Clin. Microbiol.* **38,** 3226–3230 (2000).
26. Iliev, I. D. *et al.* Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science* **336,** 1314–1317 (2012).
27. Perfect, J. R., Lindsay, M. H. & Drew, R. H. Adverse drug reactions to systemic antifungals. Prevention and management. *Drug Saf.* **7,** 323–363 (1992).
28. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
29. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Heitman, A. Amend, Y. Shea, M. Turner, I. Brownell and M. Udey for helpful discussions. We thank J. Fekecs for assistance with the figures. This work was supported by the US National Institutes of Health (NIH) NHGRI and NCI Intramural Research Programs, and in part by NIH grant no. 1K99AR059222 (to H.H.K.). Sequencing was funded by grants from the NIH (1UH2AR057504-01 and 4UH3AR057504-02).

Author Contributions K.F., H.H.K. and J.A.S. designed the outline of the study. D.S. and E.N. recruited human subjects and assisted H.H.K. in sample collection for the experiment. K.F. and J.Y. assembled and curated the fungal database. J.A.M. and C.D. prepared the clinical samples for sequencing. The members of the NIH Intramural Sequencing Center Comparative Sequencing program carried out sequencing. K.F., J.O., S.C. and M.P. analysed sequence data. K.F., H.H.K. and J.A.S. drafted the manuscript with specific contributions from J.O., J.Y. and S.C. All authors read and approved the final version of the manuscript.

Author Information Sequence data from this study have been submitted to GenBank/EMBL/DDBJ under accession numbers KC669797–KC675175, and the Sequence Read Archive, and can be accessed through BioProject identification no. 46333. Patient and sample metadata have been deposited in the controlled-access Database of Genotypes and Phenotypes (dbGaP) under study accession phs000266.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.H.K. (konghe@mail.nih.gov) or J.A.S. (jsegre@nhgri.nih.gov).

NIH Intramural Sequencing Center Comparative Sequencing Program

Jesse Becker[1], Betty Benjamin[1], Robert Blakesley[1], Gerry Bouffard[1], Shelise Brooks[1], Holly Coleman[1], Mila Dekhtyar[1], Michael Gregory[1], Xiaobin Guan[1], Jyoti Gupta[1], Joel Han[1], April Hargrove[1], Shi-ling Ho[1], Taccara Johnson[1], Richelle Legaspi[1], Sean Lovett[1], Quino Maduro[1], Cathy Masiello[1], Baishali Maskeri[1], Jenny McDowell[1], Casandra Montemayor[1], James Mullikin[1], Morgan Park[1], Nancy Riebow[1], Karen Schandler[1], Brian Schmidt[1], Christina Sison[1], Mal Stantripop[1], James Thomas[1], Pam Thomas[1], Meg Vemulapalli[1] & Alice Young[1]

[1]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Maryland 20852, USA.

METHODS

Subject recruitment and sampling. Healthy adult male and female volunteers of 18 to 40 years of age were recruited from the Washington, DC metropolitan region, United States, from September 2009 to September 2011. This natural history study 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 written informed consent before participation. Subjects provided medical and medication history and underwent a physical examination. Exclusion criteria included history of chronic medical conditions, including chronic dermatologic diseases, and use of antimicrobial medication (antibiotic or antifungal treatments) 6 months before sampling (see Supplementary Table 1 for information on healthy volunteers). Bathing or showering with only non-antibacterial soap or cleansers was allowed during the 7 days before sample collection. No bathing, shampooing or moisturizing was permitted for 24 h before sample collection. Some healthy volunteers returned 1 to 3 months after their initial visit for follow-up sampling.

Fourteen skin sites representing a range of physiological characteristics and sites of predilection for fungus-associated dermatologic diseases were selected. Proximal and core-body sites were as follows: middle upper back, external auditory canal (inside the ear), retroauricular crease (behind the ear), occiput (back of scalp), glabella (central forehead, between eyebrows), inguinal crease (skin fold midway between hip and groin area), manubrium (upper central chest) and nare (inside the nostril). Distal body sites were as follows: antecubital fossa (inner elbow), volar forearm (mid-inner forearm), hypothenar palm (palm of hand, area closest to little finger), plantar heel (bottom of heel), toenail, and toe web (web-space between third and fourth toes) (Supplementary Fig. 1). All clinical findings observed at sampling sites were documented, including any scaling on the feet and toenail thickening, discoloration or subungal debris. Body sites with left–right symmetry (10 of the 14 body sites) were sampled bilaterally to calculate intrapersonal variation (see Supplementary Fig. 1 for sites sampled).

Fungal culturing and characterization. For fungal cultures, superficial skin scrapes were collected from a 4-cm^2 area with a sterile surgical blade and placed directly in media. Skin scrapings were spread on fungal culturing plates (under a laminar flow hood to minimize contamination) to isolate pathogenic and non-pathogenic fungi, including fastidious yeasts. Selective media containing antibiotic treatments to selectively suppress bacterial growth included: inhibitory mould agar with gentamicin (R01506); BHI agar with sheep blood, chloramphenicol and gentamicin (R01144); and Sabouraud dextrose agar, Emmons with chloramphenicol and cycloheximide (R01771) (Thermo Scientific) augmented with olive oil to promote *Malassezia* growth. Plates were incubated at 30 °C, checked daily for the first week and 2 to 3 days thereafter. Isolates that flourished in culture were re-streaked for single colonies, then subcultured to ensure purity and characterized by morphological features and molecular markers. DNA was extracted using the MasterPure Yeast DNA Purification Kit (Epicentre) according to the manufacturer's instructions with the addition of 5-mm steel beads to disrupt fungal cell wells mechanically (Qiagen). ITS1 and ITS2 regions were amplified from purified genomic fungal DNA using primers 18S-F (5′-GTAAAAGTCG TAACAAGGTTTC-3′) and 5.8S-1R (5′-GTTCAAAGAYTCGATGATTCAC-3′) for ITS1 and 5.8S-F (5′-GTGAATCATCGARTCTTTGAAC-3′) and 28S1-R (5′-TATGCTTAAGTTCAGCGGGTA-3′) for ITS2. PCR products were purified and sent to ACGT Inc. for sequencing and BLAST was carried out on the resulting amplicon sequence to identity each isolate[30].

Clinical sample collection, DNA extraction, PCR amplification and sequencing of 18S rRNA gene and ITS1. For DNA analyses, samples were collected, including negative controls, as described previously[31]. Catch-All Sample Collection Swabs (Epicentre) were used for skin-swab sample collection across all sites with the exception of the toenail (toenail clippings were collected)[32], and swabs were stored in lysis solution provided with the MasterPure Yeast DNA Purification Kit (Epicentre). To pre-digest the toenail clippings, Proteinase K (Invitrogen) was added to the sample and incubated overnight with shaking at 55 °C. Skin samples were incubated in yeast lysis buffer and lysozyme (20 mg ml^{-1}) for 1 h with shaking at 37 °C. Then, 5-mm steel beads were added to mechanically disrupt fungal cell walls using a Tissuelyser (Qiagen) for 2 min at 30 Hz. The Invitrogen PureLink Genomic DNA Kit (Invitrogen) was used for all subsequent steps.

For 18S rRNA amplicon sequencing, each DNA was amplified with SR6 (5′-TACCTGGTTGATTCTGC-3′) and SR1R (5′-TGTTACGACTTTTACTT-3′) primers. The following PCR conditions were used: 2.5 μl 10× AccuPrime Buffer II, 0.2 μl Accuprime Taq, 0.5 μl primer SR6 (20 μM), 0.5 μl primer SR1R (20 μM), and 4 μl of isolated microbial genomic DNA. PCR was carried out in duplicate when possible and a portion of the reaction was run on an agarose gel to verify the presence of the 18S PCR product. Cycle number was determined such that amplification was still in the linear range of the reaction and produced sufficient PCR product for cloning (maximum of 32 cycles). Negative controls on both the mock swab and water only (no template DNA) were performed with each set of amplifications to monitor procedures and reagents, respectively. The PCR product was ligated into the PCR4 TOPO vector (Invitrogen) according to the manufacturer's protocol. Of the resulting bacterial colonies, 384 per ligation were picked, plasmid DNA purified and inserts sequenced at NISC on an ABI 3730xl sequencer (Applied Biosystems) using M13 primers flanking the insert.

For ITS1 amplicon sequencing, each DNA was amplified with adaptor plus 18SF (5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGGTAAAAGTCGTA ACAAGGTTTC) and 5.8S-1R plus barcode (5′-GTTCAAAGAYTCGATGATT CAC) primers[33]. The following PCR conditions were used: 2.5 μl 10× AccuPrime Buffer II, 0.2 μl Accuprime Taq (Invitrogen), 0.1 μl primer B adaptor plus 18SF (100 μM), 2 μl primer 5.8S-1R plus barcode (5 μM), and 4 μl of isolated microbial genomic DNA. The PCR was carried out in duplicate for 32 cycles. Duplicate amplicons were combined, purified (Agencourt AMPure XP-PCR Purification Kit; Beckman Coulter), and quantified (QuantIT dsDNA High-Sensitivity Assay Kit; Invitrogen). An average of approximately 8 ng DNA of 94 amplicons were pooled together, purified (MinElute PCR Purification Ki; Qiagen) and sequenced on a Roche 454 GS20/FLX platform with Titanium chemistry (Roche). Flowgrams were processed with the 454 Basecalling pipeline (v.2.5.3).

For 16S rRNA amplicon sequencing, each DNA was amplified with adaptor plus V1_27F (5′-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGAGAGTTTG ATCCTGGCTCAG) and V3_534R plus barcode primers (5′-CAGCAC GCATTACCGCGGCTGCTGG)[33]. The following PCR conditions were used: 2 μl 10× AccuPrime Buffer II, 0.15 μl Accuprime Taq (Invitrogen), 0.04 μl adaptor plus V1_27F (100 μM), 2 μl primer V3_354R plus barcode (2 μM), and 2 μl of isolated microbial genomic DNA. PCR was carried out in duplicate for 30 cycles and then cleaning up of PCR, amplicon pooling of approximately 10 ng DNA, purification, and sequencing were performed as described above for ITS1.

Custom-generated fungal ITS1 reference database. ITS sequences were extracted from GenBank using the query: (ITS1[All Fields] OR ITS2[All Fields] OR 5.8S[All Fields]) AND Fungi[All Fields] NOT 'uncultured'[All Fields]. Taxonomy classifications associated with sequences were recorded as strings in the following order: kingdom, phylum, class, order, family and genus, and recorded as unclassified if the levels were not clearly defined. Sequence classification was curated manually and any discrepancies in taxonomy strings were resolved using the Taxonomy Database in Pubmed. When both anamorphic (asexual) and teleomorphic (sexual) names were represented for a species within GenBank, the strings were curated manually and the anamorphic taxonomic nomenclature was selected. The sequences were then clustered to 95% sequence identity using CD-HIT[34]. Representative sequences were chosen by CD-HIT and a consensus taxonomy string was generated for the sequence, starting from the highest level (kingdom) and moving to the lowest level (genus). If the most highly represented classification was twice as frequent as the next one, this classification was chosen for the level. If no classification satisfied this criterion, this and all lower levels were set as unclassified. Sequences that were clearly misclassified as fungi were removed from the curated database.

Sequence classification and analyses. Sequences were pre-processed to remove primers and barcodes. Possible chimaeras created during PCR amplification were identified with UCHIME in mothur[35,36]. Input category for 'reference' was set to self and included in the names file to check for chimaeras, thereby using more abundant sequences as references[36]. With the ITS database described above as the reference, these chimaera-checked sequences were classified to the genus level with the BLAST option and the *k*-nearest neighbour algorithm in mothur[36]. 16S rRNA sequences were classified to the genus or species level using the RDP classifier with training set (v.6) as described previously[37]. Sequences were assigned to taxonomic units based on their genus-level phylogenetic classification. R statistical software was implemented to generate plots representing the relative abundance of fungal genera.

Community richness (Chao1), diversity (Shannon Index), membership (Jaccard Index) and structure (Theta Index) were calculated within mothur as described previously after using a subsampling cut-off of 1,000 sequences per sample[31,38,39]. Diversity indices for left and right symmetric sites were averaged for body sites with bilateral symmetry. All statistical analyses are represented as the standard error of the mean unless otherwise indicated.

18S rRNA sequence classification. To pre-process the 18S rRNA sequences, traces were base-called using Phred (v.0.990722.g), trimmed with Crossmatch, and each clone assembled using Phrap (v.0.990329). The default parameters were used except that the force level was 9 and the mismatch penalty was −1 (refs 40, 41). For approximately 15% of read pairs, the overlap was not sufficient for *de novo* assembly and a scaffolded assembly was attempted. Scaffolded assembly was carried out using the AmosCmp16Spipeline (available from http://microbiomeutil.sourceforge.net) and non-redundant reference sequences from the SILVA small subunit rRNA database[42]. 18S rRNA sequences were classified using the SILVA

v.108 database. R was implemented to generate plots representing the relative abundance of fungal genera.

***Malassezia* sequence classification to the species level.** To classify *Malassezia* ITS1 sequences from the genus to species level, we used the get.lineage command in mothur to retrieve only *Malassezia* sequences. As an internal check, these skin-associated *Malassezia* ITS1 sequences were aligned with the *Malassezia* reference package in mothur, and all discrepancies were resolved. We next curated and aligned a reference library of *Malassezia* type-strain ITS1 sequences retrieved from GenBank and augmented by those from the fungal cultures described above. Two to ten representatives were included in the database for *Malassezia* species (*M. globosa, M. restricta, M. sympodialis, M. slooffiae, M. furfur, M. pachydermatis, M. dermatis, M. yamatoensis, M. obtusa, M. japonica* and *M. nana*) for a total of 52 ITS1 sequences. Sequences were aligned with MUSCLE to generate the reference alignment[43].

This curated library was used as the reference to place and classify novel skin-associated *Malassezia* ITS1 sequences phylogenetically to the species level with the software package pplacer[44]. Sequence placement on the reference tree along with confidence scores were visualized in Archaeopteryx using the 'guppy' command[44,45]. The 'guppy classify' output and a lightweight SQL database were used to make and store taxonomic classifications. A likelihood score of at least 0.65 was used for classifications produced by the guppy classify command. Finally, classifications were converted into mothur-compatible taxonomic strings to create the tax.summary file for community-based analyses as above. Similarly, species-level designations for bacterial *Staphylococcus* sequences were generated using pplacer with a 16S rRNA reference database built from rRNA records extracted from RefSeq genomes (as of April 2012) and RDP type species sequences (Release 10, Update 24)[37].

ITS1 and 16S rRNA comparisons. Taxonomic units were defined from genus- and, where available, species-level ITS1 and 16S rRNA phylotypes. Groups were each subsampled to 1,800 sequences and the Yue–Clayton theta index generated to compare the similarity between communities. Principal coordinate analysis of the theta index was performed and the Spearman correlation of the relative abundance of each taxonomic unit versus the top three axes was calculated to assess how each taxonomic unit contributed to variation along the axes.

Co-occurrence of bacteria and fungi was assessed by calculating the partial Spearman correlation of the relative abundances of the different taxa, adjusted for multiple within-patient measurements. Calculations were performed on Fisher-transformed *r* values. Comparisons were limited to those taxa that occurred in more than 25% of samples for either ITS1 or 16S rRNA, and for ITS1, if mean abundance across all samples exceeded 0.25%. Owing to the relatively high fungal diversity found at foot sites, only foot sites (plantar heel, toenail and toe web) were used.

30. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
31. Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324,** 1190–1192 (2009).
32. Grice, E. A. *et al.* A diversity profile of the human skin microbiota. *Genome Res.* **18,** 1043–1050 (2008).
33. Lennon, N. J. *et al.* A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol.* **11,** R15 (2010).
34. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
35. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
36. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
37. Conlan, S., Kong, H. H. & Segre, J. A. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. *PLoS ONE* **7,** e47075 (2012).
38. Kong, H. H. *et al.* Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome Res.* **22,** 850–859 (2012).
39. Yue, J. C. & Clayton, M. K. A similarity measure based on species proportions. *Comm. Statist. Theory Methods* **34,** 2123–2131 (2005).
40. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).
41. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8,** 175–185 (1998).
42. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
43. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
44. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11,** 538 (2010).
45. Han, M. V. & Zmasek, C. M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10,** 356 (2009).