Genome of *Malassezia arunalokei* and Its Distribution on Facial Skin

Yong-Joon Cho,a Taeyune Kim,b,* Daniel Croll,c Minji Park,b§ Donghyeun Kim,b Hye Lim Keum,b Woo Jun Sul,b Won Hee Jungb

aSchool of Biological Sciences and Research Institute of Basic Sciences, Seoul National University, Seoul, South Korea
bDepartment of Systems Biotechnology, Chung-Ang University, Anseong, South Korea
cLaboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

ABSTRACT *Malassezia* is a fungal genus found on the skin of humans and warm-blooded animals, with 18 species reported to date. In this study, we sequenced and annotated the genome of *Malassezia arunalokei*, which is the most recently identified *Malassezia* species, and compared it with *Malassezia restricta*, the predominant isolate from human skin. Additionally, we reanalyzed previously reported mycobiome data sets with a species-level resolution to investigate *M. arunalokei* distribution within the mycobiota of human facial skin. We discovered that the *M. arunalokei* genome is 7.24 Mbp in size and encodes 4,117 protein-coding genes, all of which were clustered with *M. restricta*. We also found that the average nucleotide identity value of the *M. arunalokei* genome was 93.5, compared with the genomes of three *M. restricta* strains, including *M. restricta* KCTC 27527. Our findings demonstrate that they indeed belong to different species and that *M. arunalokei* may have experienced specific gene loss events during speciation. Furthermore, our study showed that *M. arunalokei* was diverged from *M. restricta* approximately 7.1 million years ago and indicated that *M. arunalokei* is the most recently diverged species in the *Malassezia* lineage to date. Finally, our analysis of the facial mycobiome of previously recruited cohorts revealed that *M. arunalokei* abundance is not associated with seborrheic dermatitis/dandruff or acne, but was revealed to be more abundant on the forehead and cheek than on the scalp.

IMPORTANCE *Malassezia* is the fungus predominantly residing on the human skin and causes various skin diseases, including seborrheic dermatitis and dandruff. To date, 18 species have been reported, and among them, *M. restricta* is the most predominant on human skin, especially on the scalp. In this study, we sequenced and analyzed the genome of *M. arunalokei*, which is the most recently identified *Malassezia* species, and compared it with *M. restricta*. Moreover, we analyzed the fungal microbiome to investigate the *M. arunalokei* distribution on human facial skin. We found that *M. arunalokei* has higher relative abundance on the forehead and cheek than on the scalp.

KEYWORDS genome, *Malassezia arunalokei*, *Malassezia restricta*, mycobiome, skin fungal community, skin mycobiome

*Malassezia* is a fungal genus predominantly residing on the skin of humans and warm-blooded animals. A major physiological characteristic of *Malassezia* is its lipid dependency, which reflects the sebum-rich environment it inhabits. The lipophilic nature of *Malassezia* is due to the absence of a gene encoding fatty acid synthase in the genome of most *Malassezia* species (1). To date, 18 species have been reported, of which 10 (*M. restricta*, *M. globosa*, *M. arunalokei*, *M. sympodialis*, *M. dermatis*, *M. slooffiae*, *M. furfur*, *M. obtusa*,
M. japonica, and M. yamatoensis) are found on human skin. Among the Malassezia species, Malassezia restricta is the most predominant on human skin, especially on the scalp. It is associated with skin diseases such as seborrheic dermatitis. In addition to M. restricta, Malassezia arunalokei is of particular interest because of its high internal transcribed spacer (ITS) sequence similarity to Malassezia restricta (2). M. arunalokei was first identified and reported in 2016 by Honnavar et al., who originally isolated it from the scalp and nasolabial fold of patients with seborrheic dermatitis/dandruff (SD/D) and healthy individuals in northwestern India (2). Honnavar et al. showed that the appearance of M. arunalokei colonies was slightly different from those of M. restricta and that the average single-cell size of M. arunalokei was also larger than that of M. restricta, with a mean area of 5.5 versus 3.2 \( \mu \text{m}^2 \), respectively. Additionally, they showed that M. arunalokei is catalase negative. This phenotype was previously found in M. restricta, whose genome lacks a catalase gene, but was not observed in other Malassezia species (2). A comparison of ITS sequences between M. arunalokei and M. restricta showed 6.4% variation, and a fluorescent amplified fragment length polymorphism (FAFLP) analysis showed only 25% similarity to M. restricta, underpinning that M. arunalokei is a novel species and differentiated from M. restricta (2).

In this study, we sequenced and annotated the genome of the M. arunalokei strain NCCPF 127130 (CBS 13387) and compared it with M. restricta KCTC 27527 (3, 4). Moreover, we reanalyzed previously reported amplicon sequencing data sets at a species-level resolution to investigate the distribution of M. arunalokei within facial skin mycobiota of East Asian individuals (5, 6). We also estimated the time at which M. arunalokei diverged within the Malassezia genus.

**RESULTS AND DISCUSSION**

**Analysis of the M. arunalokei genome.** The genomic DNA of M. arunalokei NCCPF 127130 (equivalent to CBS 13387 and MTCC 12054) was extracted and sequenced using an Illumina NextSeq. We obtained 19 contigs, resulting in a genome size of 7.24 Mbp. The results of BUSCO analysis are presented in Table 1; the genome assembly is high quality in terms of completeness and contamination compared to the completely assembled genome of

| Parameter                              | M. arunalokei | M. restricta |
|----------------------------------------|--------------|--------------|
| Status                                 | Draft        | Complete     |
| Total genome size (bp)                 | 7,247,604    | 7,330,907    |
| No. of contigs                         | 19           | 9 (1)*       |
| \( N_\text{50} \) (bp)                 | 771,082      | 1,222,814    |
| GC ratio (%)                           | 55.6         | 55.8 (31.4)* |
| Total no. of tRNAs                    | 82           | 74 (24)*     |
| Total no. of CDSs \( ^b \)            | 4,117        | 4,390 (16)*  |
| CDS length (avg/median)               | 1,525.4/1,287.0 | 1,473.8/1,236.0 |
| Exon length (avg/median)              | 1,147.9/942.0 | 990.0/711.0  |
| Intron length (avg/median)            | 68.1/40.0    | 37.2/30.0    |
| No. of introns                         | 1,354        | 2,150        |
| No. of exons                           | 5,471        | 6,556        |
| No. (%) of spliced genes               | 858 (20.8)   | 1,194 (27.2) |
| Gene density (genes/Mb)               | 568.1        | 601.1        |
| No. (%) of BUSCOs:                    |              |              |
| Complete BUSCOs                        | 712 (93.9)   | 715 (94.3)   |
| Complete and single-copy BUSCOs        | 711 (93.8)   | 714 (94.2)   |
| Complete and duplicated BUSCOs         | 1 (0.1)      | 1 (0.1)      |
| Fragmented BUSCOs                      | 6 (0.8)      | 8 (1.1)      |
| Missing BUSCOs                         | 40 (5.3)     | 35 (4.6)     |

*The value in parentheses shows the information for the mitochondrial genome.

\( ^b \)CDSs, coding DNA sequences.

M. japonica, and M. yamatoensis) are found on human skin. Among the Malassezia species, Malassezia restricta is the most predominant on human skin, especially on the scalp. It is associated with skin diseases such as seborrhoeic dermatitis. In addition to M. restricta, Malassezia arunalokei is of particular interest because of its high internal transcribed spacer (ITS) sequence similarity to M. restricta (2).

M. arunalokei was first identified and reported in 2016 by Honnavar et al., who originally isolated it from the scalp and nasolabial fold of patients with seborrhoeic dermatitis/dandruff (SD/D) and healthy individuals in northwestern India (2). Honnavar et al. showed that the appearance of M. arunalokei colonies was slightly different from those of M. restricta and that the average single-cell size of M. arunalokei was also larger than that of M. restricta, with a mean area of 5.5 versus 3.2 \( \mu \text{m}^2 \), respectively. Additionally, they showed that M. arunalokei is catalase negative. This phenotype was previously found in M. restricta, whose genome lacks a catalase gene, but was not observed in other Malassezia species (2). A comparison of ITS sequences between M. arunalokei and M. restricta showed 6.4% variation, and a fluorescent amplified fragment length polymorphism (FAFLP) analysis showed only 25% similarity to M. restricta, underpinning that M. arunalokei is a novel species and differentiated from M. restricta (2).

In this study, we sequenced and annotated the genome of the M. arunalokei strain NCCPF 127130 (CBS 13387) and compared it with M. restricta KCTC 27527 (3, 4). Moreover, we reanalyzed previously reported amplicon sequencing data sets at a species-level resolution to investigate the distribution of M. arunalokei within facial skin mycobiota of East Asian individuals (5, 6). We also estimated the time at which M. arunalokei diverged within the Malassezia genus.
M. restricta KCTC 27527

M. arunalokei NCCPF 127130

FIG 1 MAUVE genome alignment between M. restricta KCTC 27527 and M. arunalokei NCCPF 127130. Red lines denote chromosomal or contig partition.
phylogenetic analysis indicated that *M. arunalokei* is the most recently diverged species in the *Malassezia* lineage currently known.

Previous phylogenetic analysis conducted by Honnavar et al. as well as other previous analyses using D1/D2 domains of large subunit (LSU) ribosomal DNA (rDNA) sequences showed that *M. globosa*, *M. restricta*, and *M. arunalokei* are closely clustered within the same clade in the phylograms (1, 2, 9), and our result was in accordance with these findings. However, we also observed two distinct subspecies groups in the *M. restricta* species (*M. restricta* CBS 8741 and KCTC 27527 versus *M. restricta* CBS 7877) in the phylogram we generated. The existence of subspecies groups (amplicon strain variants) of *M. restricta* on the human skin was also described elsewhere (5). To confirm the presence of subspecies groups in genetic levels in detail, more genomes of *M. restricta* strains should be fully sequenced.

Orthologous gene comparisons between *M. arunalokei* and *M. restricta*. As indicated by the whole-genome alignment presented in Fig. 1, most *M. arunalokei* genes are expected to belong to *M. restricta*. To confirm this, we compared orthologous genes between *M. arunalokei* NCCPF 127130 and *M. restricta* KCTC 27527. Type strains of other *Malassezia* species, such as *M. globosa* strain CBS 7966, *M. sympodiales* ATCC 42132, and *M. pachydermatis* CBS 1879, the genomes of which were sequenced and annotated, were also included for more comprehensive analysis. As shown in Table 1, the genome of *M. arunalokei* NCCPF 127130 contains 4,117 protein-coding genes. Overall, most of orthologs in *M. arunalokei* were shared with *M. restricta* (Fig. 4A). Notably, 444 (11.4%) orthologs of *M. arunalokei* were shared with *M. restricta* only, a significantly higher proportion than the number of orthologs shared with other strains (Fig. 4B). In comparison with *M. arunalokei* and *M. restricta* alone, 214 *M. restricta*-specific orthologs were identified against 36 *M. arunalokei*-specific ones, implying that exclusive gene loss occurred in *M. arunalokei* during speciation (Fig. 4A; Table S4). (3, 4). We then closely examined the absent genes in the *M. arunalokei* genome compared with *M. restricta*. Gene Ontology (GO) analysis found no significant enrichment of any GO term in the categories of biological process and molecular function. However, regarding GO terms related to the cellular component category, we discovered that mitochondrial localization was the most enriched term (67 out of 214) among genes absent in the *M. arunalokei* genome compared with *M. restricta*. Although most of the deleted genes encode mitochondrial proteins, we found no defects in the growth of *M. arunalokei* in aerobic environments and under oxidative stress conditions (data not shown).

Among the genes that are absent in *M. arunalokei*, we paid particular attention to the ortholog of *M. restricta* MRET_0913, which encodes imidazole glycerol-phosphate
dehydratase (IGPD) required for histidine biosynthesis. IGPD is encoded by \textit{HIS3} in the model yeast \textit{Saccharomyces cerevisiae}, and deletion of the gene caused histidine auxotrophy of the fungus (10). On human skin, the proteogenic amino acid histidine plays an important role. Histidine is a natural moisturizing factor and is produced from the proteolysis of the histidine-rich protein filaggrin, which is mainly found in granular layer keratinocytes in the upper epidermis (11). Furthermore, studies have shown that histidine supplementation increases the barrier function of the skin and reduces atopic dermatitis symptoms (12, 13). Taken together, we hypothesize that the histidine-rich environment in the epidermal layer may have triggered the loss of the \textit{HIS3} homolog in the \textit{M. arunalokei}\linebreak lineage.

Such gene loss is consistent with regressive evolution frequently observed in parasitic organisms, which are themselves associated with hosts that show metabolic redundancy when parasitized (14). Gene loss upon environmental change or condition has also been observed, most commonly in unicellular organisms, including pathogenic bacteria and fungi. Furthermore, in some cases, gene loss influenced the pathogenicity of the organisms (14). Examples include loss of the functional BNA genes in a pathogenic yeast, \textit{Candida glabrata}. The BNA genes are required for nicotinic acid (NA). Therefore, loss of the gene caused NA auxotrophy, which in turn increased expression of the EPA genes encoding a family of adhesin proteins and the pathogenicity of \textit{C. glabrata} (15). However, we do not have additional evidence to determine whether gene loss in \textit{M. arunalokei}, such as the \textit{HIS3} homolog,
is associated with the pathogenicity of the fungus. This is mainly due to the lack of a tool for genetic manipulation.

*M. arunalokei* has high relative abundance on the forehead and cheek compared to the scalp. In our previous studies, we analyzed the mycobiomes on the scalp, forehead, and cheeks of healthy individuals and patients with SD/D or acne residing in Seoul, South Korea, using culture-independent amplicon sequencing (5, 6). These studies analyzed differences in the diversity of fungal communities between the patient groups and healthy groups, but the distribution of *M. arunalokei* among the fungal communities on the scalp, forehead, and cheeks could not be determined. Therefore, we reanalyzed the previously obtained mycobiome data sets at the species level in our current study in order to gain more precise information on the distribution of *M. arunalokei* on the scalp, forehead, and cheeks in our patient group and healthy group. Moreover, in addition to our own data sets, the amplicon sequencing data set generated by Tong et al. (16), who collected samples from multiple skin sites, including the forehead, on a healthy cohort residing Hong Kong, China, were included in the current study to expand the analysis (Table 2). A total of 356 species were detected in our mycobiome analysis. Among them, we assigned 346 species, while the remainder were unassigned (Table S5). The results of our mycobiome analysis showed that *M. arunalokei* was present in 80.7% of samples (209 out of 259), indicating that the fungus is universally present on the scalp, forehead, and cheeks, although its relative abundance is minimal compared with those of other *Malassezia* species (Fig. 5). We next performed a Wilcoxon rank sum test to determine the significance of the *M. arunalokei* abundance with respect to either disease status (SD/D or acne) or specific facial skin site (scalp, forehead, or cheeks). No significant differences in the abundance of *M. arunalokei* were observed between the patient group and the healthy group (Fig. 6A). However,
we did find differences between \textit{M. arunalokei} abundance and different skin sites (scalp versus forehead and cheeks). The abundance of \textit{M. arunalokei} on the forehead and cheeks was significantly higher than that on the scalp in both the patient and healthy groups (Fig. 6B and C). These findings indicate that \textit{M. arunalokei} abundance is not related to D/SD or acne but rather to the skin site. We also compared the number of nonchimeric reads with detection frequency of \textit{M. arunalokei} and found no correlation between them (Fig. S1). This result further indicated that skin site has a greater influence on mycobiome composition than sequencing depth.

Overall, our data showed that \textit{M. arunalokei} is a fungal species possessing a genome highly homologous to that of \textit{M. restricta} and indicated that it is the most recent species in the lineage of \textit{Malassezia} that has adapted to the skin environment. Furthermore, our analysis of the facial mycobiome in humans showed that \textit{M. arunalokei} had distinct niche preferences.

\textbf{MATERIALS AND METHODS}

\textbf{Genomic DNA library construction and sequencing.} \textit{M. arunalokei} NCCPF 127130 was obtained from the Westerdijk Fungal Biodiversity Institute (https://www.knaw.nl/en/institutes/westerdijkinstitute). The strain was cultured, and its genomic DNA was extracted as described previously (4). Genomic DNA was sheared using a g-TUBE device (Covaris, Woburn, MA, USA) to fragments 20 kb in size, and DNA library construction was then performed using the PacBio platform (PacBio Biosciences, Menlo Park, CA, USA) according to the manufacturer’s instructions. Sequencing was performed using a SMRT Cell and PacBio Sequel system (PacBio Biosciences). Additionally, an Illumina sequencing library was constructed using a NEBNext Ultra II FS DNA library prep kit for

\begin{table}
\centering
\caption{Occurrence of \textit{M. arunalokei} in swab samples}
\begin{tabular}{|l|c|c|c|c|c|c|c|}
\hline
\textbf{Parameter} & \multicolumn{2}{c|}{\textbf{Seoul}} & \multicolumn{2}{c|}{\textbf{Hong Kong,}} & \multicolumn{2}{c|}{\textbf{Hong Kong,}} \\
 & \textbf{Scalp} & \textbf{Forehead} & \textbf{Cheek} & \textbf{forehead, normal} & \textbf{forehead, normal} & \\
\hline
\% of occurrences & 66.7 & 54.4 & 100 & 100 & 100 & 90.6 \\
No. of samples & 45 & 57 & 18 & 10 & 22 & 11 \\
No. of occurrences & 30 & 31 & 18 & 10 & 22 & 11 \\
\hline
\end{tabular}
\end{table}

\textsuperscript{a}Samples from Seoul represent facial skin mycobiome data sets generated in our previous study (5, 6). Samples from Hong Kong represent skin mycobiome data sets generated by Tong et al. (16). SD/D, seborrheic dermatitis/dandruff.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Relative abundance of various fungal species (y axis) in samples from different skin sites of the face (labeled on the x axis). Each column represents either healthy or diseased individuals. Fungal species, skin site, and affected status are indicated in the legend on the right.}
\end{figure}
Illumina (E7805S; New England Biolabs, Ipswich, MA, USA) according to the manufacturer’s instructions, followed by sequencing using a NextSeq 550 (Illumina, San Diego, CA, USA) to generate 2/36-bp paired-end reads.

**Genome assembly and analysis.** Generated raw sequences were assembled using the hybrid assembler Unicycler v.0.4.8 (17), and the general statistics are shown as Table S1 in the supplemental material. Gene prediction and annotation were performed using FunGAP pipeline v.1.1.0 using previous RNA sequencing data of *M. restricta* KCTC 27527 (3, 18). Protein sequences were clustered for an ortholog search using DIAMOND v.2.0.6, with the parameter “-id 50” (sequence identity) and mcl 14–137 with the parameter “I 2.0” (inflation value, where higher values indicate more fine-grained clustering) (19). To evaluate the quality of genome assembly, we conducted an analysis using BUSCO v.5.3.0 (20) based on fungi_odb10 (13 December 2019). Genome sequences were aligned using MAUVE v.2.4.0, and their average nucleotide identity (ANI) values were calculated by ANIb method based on BLAST+ from JSpecies-WS (21–23).

**Divergence time estimation.** We performed divergence time estimation based on Bayesian inference between species of the Malassezia genus. At first, we collected the genome sequences from the 25 strains belonging to 14 different species (Table S2). In addition, we used *S. cerevisiae* (BioProject accession no. PRJEB7245; assembly accession no. GCA_000205763.1) as an outgroup. Amino acid sequences of seven highly conserved fungal barcoding proteins were selected, namely, RPB2 (DNA-directed, RNA polymerase II subunit), phospholipase, lipase, actin, MCM7 (DNA replication licensing factor MCM7), TEF1 (elongation factor

**FIG 6** Relative abundance of *M. arunalokei* at different facial skin sites in study cohorts. (A) Relative abundance of *M. arunalokei* between patient (seborrheic or acne) and healthy (normal) groups. (B and C) Relative abundance of *M. arunalokei* at different skin sites (scalp versus forehead and cheeks) in the (B) healthy and (C) patient groups. Values above the bar graphs are *P* values. The numbers of the samples used in the analysis are listed in Table 2.
Supp. Material: 5.8 MB.

**References**

1. Wu G, Zhao H, Li C, Rajapakse MP, Wong WC, Xu J, Saunders CW, Reeder NL, Reilmann RA, Scheynius A, Sun S, Billmyre BR, Li W, Averette AF, Mieczkowski P, Heitman J, Theelen B, Schroder MS, De Sessions PF, Butler G, Maurer-Stroh S, Boekhout T, Nagarajan N, Dawson TL, Jr. 2015. Genus-wide comparative genomics of Malassezia delineates its phylogeny, physiology, and niche adaptation on human skin. PLoS Genet 11:e1005614. https://doi.org/10.1371/journal.pgen.1005614.

2. Honnavar P, Prasad GS, Ghosh A, Dogra S, Handa S, Rudramurthy SM. 2016. Malassezia arunalokei sp. nov., a novel yeast species isolated from seborrhoeic dermatitis patients and healthy individuals from India. J Clin Microbiol 54:1826–1834. https://doi.org/10.1128/JCM.00683-16.

3. Cho YJ, Park M, Jung WH. 2019. Resequencing the genome of Malassezia restricta strain KCTC 27527. Microbiol Resour Announc 8:e00213-19. https://doi.org/10.1128/MRA.00213-19.

4. Park M, Cho YJ, Lee YW, Jung WH. 2017. Whole genome sequencing analysis of the cutaneous pathogenic yeast Malassezia restricta and identification of the major lipase expressed on the scalp of patients with dandruff. Mycoses 60:188–197. https://doi.org/10.1111/myc.12586.

5. Kim J, Park T, Kim HJ, An S, Sul WJ. 2021. Inferences in microbial structural signatures of acne microbiome and mycobio. J Microbiol 59:369–375. https://doi.org/10.1007/s12275-021-0647-1.

6. Park T, Kim HJ, Myeong NR, Lee HG, Kwack I, Lee J, Kim BJ, Sul WJ, An S. 2017. Collapse of human scalp microbiome network in dandruff and seborrhoic dermatitis. Exp Dermatol 26:835–838. https://doi.org/10.1111/exd.13293.

7. Jain C, Rodríguez RL, Phillippy AM, Konstantindis K, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:15114. https://doi.org/10.1038/s41467-018-07641-9.

8. Zhao R-L, Li G-J, Sánchez-Ramírez S, stata M, Yang Z-L, Wu G, Dai Y-C, He S-H, Cui B-K, Zhou J-L, Wu F, He M-Q, Moncalvo J-M, Hyde KD. 2017. A six-geeen phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomic perspective. Fungal Diversity 84:43–74. https://doi.org/10.1007/s13225-017-0381-5.

9. Ianiñi G, Heitman J. 2020. Approaches for genetic discoveries in the skin commensal and pathogenic Malassezia yeasts. Front Cell Infect Microbiol 10:393. https://doi.org/10.3389/fcimb.2020.00393.
22. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.

23. Richter M, Rossello-Mora R, Glockner FO, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931. https://doi.org/10.1093/bioinformatics/btw681.

24. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst101.

25. Steenwyk JL, Buida TJ, Ill, Li Y, Shen XX, Rokas A. 2020. ClipKIT: a multiple sequence alignment trimming software for accurate phylogenetic inference. PLoS Biol 18:e3001007. https://doi.org/10.1371/journal.pbio.3001007.

26. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuhnert D, De Maio N, Matschiner M, Mendes FK, Muller NF, Ogilvie HA, Du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

27. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67: 901–904. https://doi.org/10.1093/sysbio/syy032.

28. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA, Schoenfeld D, Nomicos E, Park M, Program N, Kong HH, Segre JA, NIH Intramural Sequencing Center Comparative Sequencing Program. 2013. Topographic diversity of fungal and bacterial communities in human skin. Nature 498:367–370. https://doi.org/10.1038/nature12171.

29. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet 17:1–10. https://doi.org/10.14806/ej.17.1.200.

30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon sequencing reads. EMBnet 17:10. https://doi.org/10.1038/nmeth.3869.

31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.

32. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33:D501–D504. https://doi.org/10.1093/nar/gkl025.

33. Domingos P, Pazzani M. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learn 29:103–130. https://doi.org/10.1023/A:1007431511361.