



OPEN

Biodiversity of mycobial communities in health and onychomycosis

Michael Olbrich^{1,2,5}, Anna Lara Ernst^{1,5}, Foteini Beltsiou¹, Katja Bieber¹, Sascha Ständer³, Melanie Harder⁴, Waltraud Anemüller³, Birgit Köhler³, Detlef Zillikens³, Hauke Busch^{1,2,5}✉, Axel Künstner^{1,2,5} & Ralf J. Ludwig^{1,3,5}✉

Onychomycosis (OM) is a common fungal nail infection. Based on the rich mycobial diversity in healthy toenails, we speculated that this is lost in OM due to the predominance of a single pathogen. We used next generation sequencing to obtain insights into the biodiversity of fungal communities in both healthy individuals and OM patients. By sequencing, a total of 338 operational-taxonomic units were found in OM patients and healthy controls. Interestingly, a classifier distinguished three distinct subsets: healthy controls and two groups within OM patients with either a low or high abundance of *Trichophyton*. Diversity per sample was decreased in controls compared to cases with low *Trichophyton* abundance (LTA), while cases with a high *Trichophyton* abundance (HTA) showed a lower diversity. Variation of mycobial communities between the samples showed shifts in the community structure between cases and controls—mainly driven by HTA cases. Indeed, LTA cases had a fungal β -diversity undistinguishable from that of healthy controls. Collectively, our data provides an in-depth characterization of fungal diversity in health and OM. Our findings also suggest that onychomycosis develops either through pathogen-driven mechanisms, i.e., in HTA cases, or through host and/or environmental factors, i.e., in cases with a low *Trichophyton* abundance.

Onychomycosis (OM) is a fungal infection of the nail that accounts for around 50% of all worldwide nail disorders^{1,2}. The infection occurs in about 10% of the adult population with a higher prevalence in males and elderly people^{2–4}. Patients exhibit symptoms such as white or yellow nail discoloration, thickening, roughness, and separation of the nail from the nail bed. The current understanding of OM, as a (mostly) single pathogen-driven infection is based on its current clinical diagnostics, specifically microscopy and culture of clinical samples.

By contrast, multiplex PCR, a recently established diagnostic tool for fungal skin infections, showed a presence of more than one fungal pathogens in up to 32% of the investigated nail samples in OM patients^{5,6}. However, multiplex systems are a targeted approach that is designed to detect fungal DNA using species-specific primers⁷. Hence, they are limited to the detection of species that are targeted in the respective assay. By contrast, next-generation sequencing (NGS) yields large numbers of reads and can provide additional, profound insights into the fungal diversity of OM, as it facilitates detection of dominant communities as well as low-abundant and rare taxa^{8–11}. In an effort to map the fungal diversity of 14 anatomical sites in clinically healthy individuals, it was found that the toenail fungal community was largely dominated by *Malassezia* at the genus level¹².

Thus, we hypothesized that *Malassezia* might have a protective effect against other fungal skin pathogens by occupation of the niche. In addition, predicted on the data from multiplex PCR that detects more than one pathogen in OM, we assumed that the composition of pathogens in OM is even more diverse. To challenge these assumptions, we applied internal transcribed spacer 2 (ITS2) sequencing to toenail samples of confirmed OM cases to quantify the fungal microbiome (mycobiome) in human OM. We further included nail specimen from clinically healthy nails, in which neither microscopy, culture nor multiplex PCR detected any fungal pathogen. This identified 338 fungal operational-taxonomic units (OTUs) in all samples. Overall, the results provide

¹Lübeck Institute for Experimental Dermatology, Institute for Cardiogenetics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany. ²Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany. ³Department of Dermatology, Allergy and Venerology, University of Lübeck, Lübeck, Germany. ⁴EUROIMMUN AG, Lübeck, Germany. ⁵These authors contributed equally: Michael Olbrich, Anna Lara Ernst, Hauke Busch, Axel Künstner and Ralf J. Ludwig. ✉email: hauke.busch@uni-luebeck.de; ralf.ludwig@uksh.de

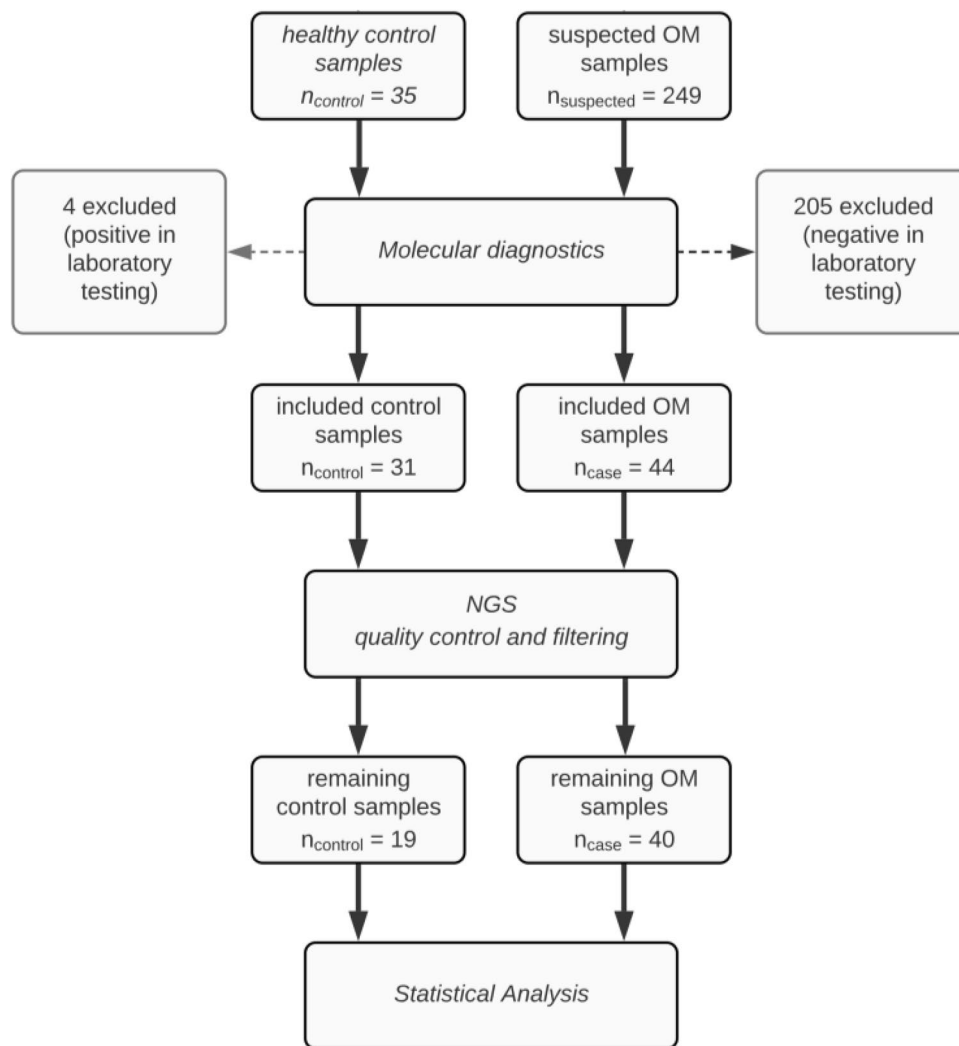


Figure 1. Flowchart of cohort collocation.

detailed insights into the composition of fungal communities in toenails, suggesting two distinct pathogenic pathways in OM pathogenesis.

Materials and methods

Patient and sample collection. Samples were collected from patients treated in the in- and out-patient clinic of the Department of Dermatology at the UKSH in Lübeck in the form of toenail scrapings¹³. Patients with diseases that are associated with nail changes, such as psoriasis, alopecia areata or others were excluded. Sample material was obtained from a total of 249 patients with suspected OM. In addition, 35 healthy controls with no clinical evidence of OM were recruited. All patients and controls underwent routine and molecular OM laboratory testing. Of the 249 patients with suspected OM, the diagnosis was confirmed in 44 cases. Thus 205/249 patients with inconclusive results were excluded. Among the 35 healthy controls, routine and/or molecular OM diagnosis showed positive results. Thus, 4/35 controls were excluded and 31 healthy controls were included in the analysis. Quality control procedure post sequencing resulted in a total of 59 remaining samples, of which 40 were confirmed OM cases (25 males, 15 females, average age 61.3 ± 12.5) and 19 healthy controls (6 males, 13 females, average age 42.7 ± 16.3) (Fig. 1, Table 1). All experiments with human samples were performed after written informed consent had been obtained and were approved by the ethical committee of the Medical Faculty of the University of Lübeck (#17-066) in accordance with the Declaration of Helsinki.

Routine OM laboratory testing. For routine OM laboratory testing, samples were prepared for microscopic/histologic examination and culture. Material from nail scrapings was taken from petri dishes and wet mounts were created by combining the material with one drop of Blankophor dye (Indulor, Ankum, Germany) on a glass slide for fluorescence microscopic examination. The remaining material was then implanted in Dermatophyte-Agar with or without the mold inhibitor Cycloheximid (Merck, Darmstadt, Germany). The plates

Group	Sex (f m)	Age (mean sd)	Pets ¹ (no yes)	Therapy ² (no yes)
Control	13 6	42.74 16.26	15 4	19 0
Case	15 25	61.33 12.47	29 11	33 7
Case LTA	12 12	60.17 12.71	18 6	21 3
Case HTA	3 13	63.06 12.29	11 5	12 4

Table 1. Covariate data. List of covariates grouped by case and control samples as well as the identified classification into low *Trichophyton* abundance (LTA) and high *Trichophyton* abundance (HTA) subgroups ¹indicates if pets are in the household and/or the individual has contact to animals on a regular basis ²indicates if the person had applied any treatment (topically and/or systemically) to the suspected onychomycosis.

were then incubated for three to four weeks at room temperature with weekly assessment by an experienced dermatologist.

Molecular OM laboratory testing. DNA was extracted using the QIAamp DNA mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions for tissue samples. 40 ml of Proteinase K (Qiagen, Hilden, Germany) were added prior to the incubation at 56 °C and 1000 rpm overnight. Half of the sample volume was further used for sequencing preparation. From the remaining quantity, multiplex PCR followed by incubation on a chip for detection of specific fungal sequences were performed using the EUROArray Dermatomyces kit (EUROIMMUN, Lübeck, Germany) according to the manufacturer's instructions. The PCR settings prior to hybridization onto the microarray chip were as follows: initial denaturation for 3 min at 94 °C, 36 cycles of each 15 s at 94 °C, 15 s at 56 °C and 40 s at 72 °C, followed by 1 min final elongation at 72 °C.

Library preparation and NGS. From the isolated DNA samples, the nuclear ribosomal ITS2 region was amplified by using a dual indexing approach and the ITS specific sequencing primers ITS4 and fITS7 that were linked to a unique eight-base multiplex identifier, as previously described⁸. PCR reactions were carried out in a 25 µl reaction volume using the Phusion[®] Polymerase (ThermoFisher, Waltham, MA, USA) with the following PCR settings: initial denaturation for 30 s at 98 °C; 35 cycles of 9 s at 98 °C, 30 s at 50 °C, and 30 s at 72 °C, final extension for 10 min at 72 °C. For all forward and reverse primer combinations, template-free reactions were performed. PCR products were quantified on a 1.5% agarose gel (Biozym, Hessisch-Oldendorf, Germany) after 5 min at 120 V followed by 1 h at 110 V. Gel pictures were taken on a Vilber E-Box CX5.TS (Vilber Lourmat SAS, Collegien, France) and image analysis was done using the image analysis software Vision quant version 16.16.0.0 (Vilber Lourmat SAS, Collegien, France). Fungal PCR products were equimolarly subpooled, run on an UltraPure™ Agarose gel (ThermoFisher, Waltham, MA, USA) and the 250 bp sized bands were extracted from the gel using the MinElute Gel Extraction Kit (Qiagen, Venlo, Netherlands). Subpool concentrations were determined using the NEBNext Library quantification Kit for Illumina (New England Biolabs, Frankfurt am Main, Germany) on the Eppendorf Mastercycler ep Realplex (Eppendorf, Hamburg, Germany) following the manufacturer's instructions. The subpools were combined to an equimolar library which was further purified using the AMPure[®] Beads XP Kit (Beckman&Coulter, Brea, CA, USA) and quantified with the NEBNext Library quantification Kit. The average amplicon size of the library was determined using the Agilent High Sensitivity DNA Kit (Agilent, Santa Clara, CA, USA) and the library (17.5 pM) was sequenced on a MiSeq (Illumina, San Diego, CA, USA) using the MiSeq v3 600 cycles sequencing chemistry (Illumina, San Diego, CA, USA) together with 20% PhiX Control v3 (Illumina, San Diego, CA, USA). Sequencing data used for this study were submitted to the European Nucleotide Archive (ENA) and is available under accession number PRJEB37496.

Bioinformatics analysis. Raw FASTQ files that were obtained from the MiSeq sequencing platform included about 18.8 million reads. Raw data was processed using PIPITS⁸ (v2.3) with default parameters and the RDP Classifier¹⁴ (v2.1211) against UNITE database¹⁵ (04.02.2020) to retrieve taxonomic annotation for each sequence. Statistical analysis was performed using R¹⁶ (v4.0.2). The R packages *phyloseq*¹⁷ (v1.28), *vegan*¹⁸ (v2.5.6), *PhILR*¹⁹ (v1.14.0), *DivNet*²⁰ (v0.3.6), *caret*²¹ (v6.0-86) and *indicspecies*²² (v1.7.6) were used to facilitate the subsequent analyses. Visualization was done with *ggplot2*²³ (v3.2.1) and *ggpubr*²⁴ (v0.2.3). Post-processing was facilitated as follows: Filtering of the data set comprised removal of unassigned phyla, removal of OTUs prevalent in less than five percent of samples and removing samples with less than 1000 counts in total. The filtered data set was comprised of 59 samples (40 cases and 19 controls) containing 338 OTUs. The lack of exhaustively curated databases for fungal genomes resulted in a growing number of unassigned entities corresponding to the taxonomic rank, e.g., on family level about one in four are unassigned while on species level more than half of the entities remain unspecified (Suppl. Fig. 1). Nevertheless, NGS followed by taxonomic annotation allows to identify fungal species. An overview of per sample communal composition, over all taxonomic ranks, is provided in Supplementary Fig. 2. Cumulative Sum Scaling (CSS) was employed as normalization method to mitigate the bias introduced by strong variability in total counts per sample (mean: 41,265; standard deviation: ± 62,937). The subsequent transformation to relative abundances enabled comparison of samples.

For linear regression analysis, *lm* (R *stats* package v4.0.2) was used and the Bayesian model was build using *rstanarm*²⁵ (v2.21.1). Results from both models were plotted using the R package *sjPlot*²⁶ (v2.8.5). The R package *randomForest*^{27,28} (4.6–14) was used to build the Random forest classifier. Genus count abundances were transformed to relative abundances and 70% of the data were used as training data set. The remaining 30% were

used as validation set (held-out data). For classification, 10,000 trees were generated, and 40 variables were tested (mtry parameter). Additionally, the proximity between rows was calculated.

Sample-wise abundance plots were generated from phylum to genus level by selecting the 20 taxa with highest mean abundance and agglomerating the remaining entities into the single category called ‘others’. In the analysis we quantified the fungal diversity on the family level, as it constituted the best trade-off between degree of detail and mapping quality. Group-wise tests for difference in taxonomic abundance were performed using Mann–Whitney U test and p values were corrected for multiple testing using Benjamini–Hochberg correction. Alpha diversity (Shannon diversity) on non-normalized counts was estimated using *DivNet* (sample-wise and community-wise) according to the assumption that taxa in the community cooccur in an ecological network²⁹. Differences in alpha diversity were assessed using the beta function (*breakaway* v4.7.2) for sample-wise estimates with group as a fixed factor and testHypothesis (*DivNet*) for community-wise estimates, respectively. Beta diversity was estimated by first performing a phylogenetic isometric log-ratio transformation on the filtered phyloseq object. To that end, the filtered (non-normalized or scaled) compositional data set was transformed via the *PhILR* package, which facilitates the estimation of diversity in terms of Euclidean distance measure. Subsequently, a two-way permutational multivariate analysis of variance (PERMANOVA) testing (99,999 permutations) was conducted on the resulting distance matrix in order to evaluate the contribution of the covariates “Therapy”, “Pets”, “Sex” and “Age” as well as the difference between controls and cases. Whereby “Therapy” indicates if any treatment was applied for suspected OM, and “Pets” indicates if the person has regular contact with pets or domestic animals. The significance of testing was controlled by analyzing the respective dispersion of beta diversity, i.e., evaluating if compared groups are homogeneously distributed. The association between species and membership to a site-group was calculated using the *signassoc* function as implemented in the *indicspecies* package.

Ethics approval. Ethics Committee of the University of Lübeck, AZ: 17-066.

Results

A total of 284 participants were enrolled in this study, the majority ($n = 249$) presented to our in- and out-patient clinic with suspected OM. Clinical presentation, “conventional” and molecular diagnostics were used to classify the participants as either healthy controls or confirmed OM cases. Samples were categorized as healthy, if clinical examination led to no suspicions of OM and none of the diagnostic tests was positive for OM, and as cases if OM was clinically suspected and at least one positive finding from any of the diagnostic tests applied. Overall, 44 case samples and 31 controls were included in the ITS sequencing procedure. From the total of 75 sequenced samples, 59 samples (40 cases, 19 controls) passed the quality control of the PIPITS analysis pipeline. They contained more than 1000 contigs respectively and comprised a total of 338 OTUs.

Composition of toenail fungal communities identifies distinct subgroups in the case samples. In healthy controls, we observed a large variation in the abundance of different fungal communities. For example, on the family level, in 5/19 samples the family *Bulleribasidiaceae* was most abundant, accounting for 35% or more of all reads (Suppl. Fig. 1). In addition, 4/19 samples were dominated by *Mycosphaerellaceae*, whilst *Cladosporiaceae* was most abundant in 3 samples. The remaining 7 samples did not have a common dominating family (Suppl. Fig. 1). This variation in healthy controls was even more apparent on the genus level, where most compositions of the toenail mycobiome are unique to the individual (Fig. 2A). This was confirmed by the large variance in alpha diversity (Fig. 2B).

In patient samples, we noted a remarkable separation of the mycobiome family and genus level into two distinct groups (Suppl. Fig. 1, Fig. 2A). In 16/40 OM cases, the family level was dominated by *Arthrodermataceae*, which were almost exclusively of the *Trichophyton* genus, accounting for at least 80% of the reads in various instances. In the remaining 24 OM patient samples, fungal diversity was very similar to that of healthy controls. Thus, the OM patient group can be stratified according to the presence of the genus *Trichophyton*: Those with a low *Trichophyton* abundance (below 15%), and those with high *Trichophyton* abundance (above 19%, Fig. 2A, Suppl. Fig. 3A).

OM samples can be classified into two distinct subgroups. According to this finding, samples were stratified into healthy controls, low *Trichophyton* abundance (LTA) cases or high *Trichophyton* abundance (HTA) cases, for subsequent analyses (Fig. 2B–D). The robustness of this grouping was assessed using linear and a Bayesian regression model alongside the assigned group as explanatory variable and relative abundance of *Trichophyton* as response. Both models confirmed the marked increase in *Trichophyton* abundance in the high group (linear model: $R^2_{\text{adj}} = 0.753$, $p < 0.001$; Bayesian model $R^2 = 0.748$, 95% CI 0.55–0.80; Suppl. Fig. 3C–D). We further classified the samples by a Random forest approach, which makes use of the relative abundance of all identified genera. The trained model had no out-of-bag error and was able to stratify the held-out data correctly into groups with low and high *Trichophyton* abundance. The genus with the highest overall impact on the stratification was *Trichophyton* (mean decrease in Gini index of 5.2; Sup. Fig. 3E).

We subsequently determined the fungal species richness and their heterogeneity using the alpha diversity of the fungal communities. Differences in sample-wise estimates of diversity were modelled to test for differences of total diversity (observed plus unobserved) across multiple samples³⁰. A significant decrease in diversity was detected between healthy controls and cases (beta test, $p < 0.001$, decrease 0.5204, s.e. 0.1350). However, no differences in alpha diversity were observed between healthy controls and LTA samples (beta test, $p = 0.105$). Furthermore, the HTA group showed a decreased Shannon diversity compared to healthy controls ($p < 0.001$, decrease 0.8934, s.e. 0.1886) as well as compared to the LTA group ($p = 0.002$, decrease 0.6217, s.e. 0.1677, Fig. 2B). Besides estimating alpha diversity for each sample, Shannon index was estimated for each group

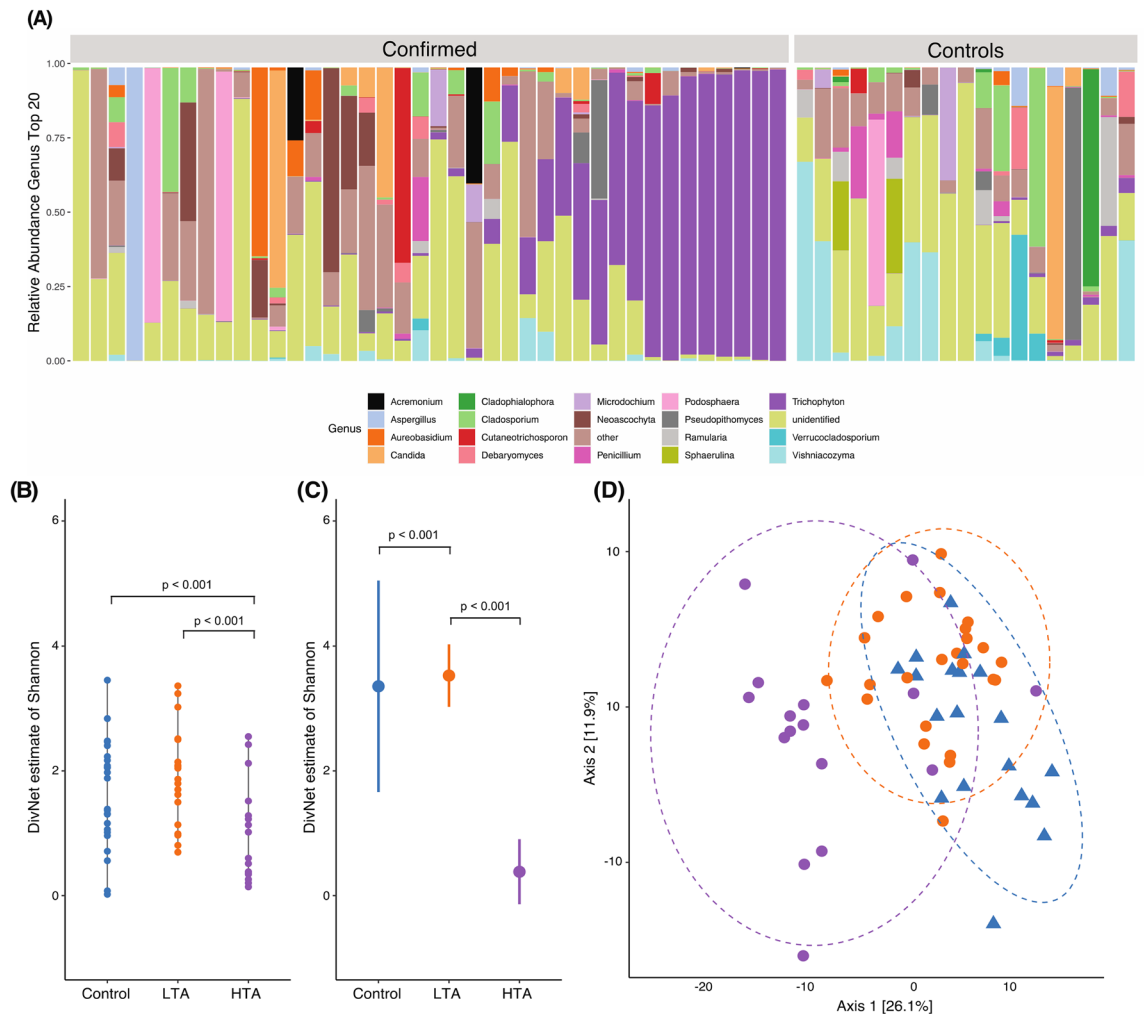


Figure 2. Composition of toenail fungal communities. **(A)** The 20 most abundant fungal genera (average across samples) were selected for visualization. The category others comprises all remaining genera. Samples are depicted in cohort grouping of confirmed case- and healthy control samples. Within those groups, the samples were arranged based on the abundance of genus *Trichophyton* in order to visualize the found subdivision in the case group. **(B)** Alpha diversity in sample-wise estimates showed no difference between controls and LTA group. Compared to the control group, a significant decrease in Shannon diversity was observed ($p < 0.001$, decrease 0.8934, s.e. 0.1886) in HTA samples. **(C)** Estimated alpha diversity in community-wise comparison using Shannon. The low *Trichophyton* abundance (LTA) group showed a significant increase in diversity compared to the healthy controls ($p = 0.001$, increase 0.1719, s.e. 0.0511). The high *Trichophyton* abundance (HTA) group had a significantly lower diversity compared to the control group ($p < 0.001$, increase 2.9728, s.e. 0.0652). **(D)** Principle coordinate analysis (PCoA) of case/control samples based on the Euclidean distance on PHILR transformed counts with blue triangles referring to controls, orange circles to LTA cases and violet circles to HTA cases. Axis 1 captures 26.1% of the variation and separates the cases into two distinct groups.

(community-wise) as well (Fig. 2C) and further investigated applying a model that accounts for both strong heterogeneity between samples and rare taxa³⁰. As in the sample-wise data, Shannon diversity was significantly decreased in cases compared to controls ($p < 0.001$, decrease 0.7137, s.e. 0.0882). The LTA group showed a significant increase in the community richness compared to the healthy controls ($p = 0.001$, increase 0.1719, s.e. 0.0511). Furthermore, the HTA group had a significantly lower diversity compared to the healthy controls ($p < 0.001$, increase 2.9728, s.e. 0.0652, Fig. 2B–C).

Next, beta diversity, i.e., the variation in mycobial composition between groups, was investigated and showed a significant difference ($p = 4 \times 10^{-5}$, $R^2 = 0.08442$) between cases and controls, accounting for approximately 8% of variation in the data set. Examination of group dispersions showed significant differences in the dispersion of case and control samples. This is congruent with the separation of cases into LTA and HTA samples, as it indicates that the difference between case and control group is influenced to a large degree by the difference of composition within groups. After accounting for the subdivision into LTA and HTA groups, we found that the differences are significant ($p = 1 \times 10^{-5}$, $R^2 = 0.18055$) and account for 18% of variation in the data set. The covariates age, sex, presence of pets and previous antifungal therapy were not found to be significant and account in total for 14.5% of variation in the data set. Validation by testing for differences in beta dispersion showed no

significant difference in composition within the three groups, i.e., the results of the multivariate analysis are not an artifact of heterogeneous dispersions (Fig. 2D).

Fungal sequencing does not significantly outperform existing diagnostics. Lastly, we addressed the question whether sequencing-based diagnosis outperforms current diagnostic approaches, including PCR-based methods. To that end, a supervised Random forest approach estimated the efficacy of separating cases from controls based on patient data, i.e., covariates, PCR results and OTUs. We considered the division into control, LTA and HTA groups as ground truth and estimated the classifier performance in a bootstrapped training and test approach. A detailed description of the procedure as well as receiver operator curve (ROC) statistics are shown in Supplementary Fig. 4. In fact, the classification based on OTUs marginally, but not significantly, outperformed the other features with an average probability of 91.27% of correct classification. However, as expected based on the strong similarity in composition between control and LTA groups, the later could not be identified reliably.

Discussion

We used current diagnostic tools of OM to curate a cohort of healthy and OM samples. The utilization of NGS enabled us to investigate the complex composition of the mycobiome in healthy and infected toenails. Based on previous reports, we had expected a predominance of genus *Malassezia* in healthy samples. We further hypothesized that by occupation of the toenail niche with *Malassezia* species protects from onychomycosis—albeit some *Malassezia* species may cause OM³¹. The obtained results were, however, quite unexpected in several ways:

In our study, only those nail samples were included, in which clinical examination, microscopy, histology and multiplex analysis excluded OM. In these samples, a much greater diversity was observed and none of the samples was dominated by *Malassezia*. In the study by Findley and colleagues the predominant species were different *Malassezia* species¹². By contrast, we herein report a much higher diversity on both genus and species level, in toenails from healthy controls. At the moment, we can only speculate on the reason for this difference. We found the ongoing effort to curate more fungal genomes which is leading to a lack of comprehensive databases to map against, to be the likeliest explanation. Currently, this leads to decreasing accuracy in assignment on genus and species level, up to an amount of 53% unassigned reads on species level. Other, mutually non-exclusive explanations, are: (i) a high plasticity of the mycobiome, as described for the intestinal mycobiome (ii) relative low sample sizes of both studies, (iii) different selection criteria, (iv) regional, and/or (v) seasonal differences.

At an about equal distribution, the case group was stratified by the abundance of the genus *Trichophyton*. Within this cohort, we found an abundance of 15% *Trichophyton* to be the boundary that separates the high *Trichophyton* abundance (HTA) samples from the low *Trichophyton* abundance (LTA) samples. This finding suggests that OM can be either pathogen- or host-/environment-driven. In accordance with this, we showed that on average, the sample-wise calculated fungal alpha diversity was significantly lower ($p < 0.001$) in HTA samples compared to healthy controls, which is an expected confirmation of our alternative stratification in the case group. No significant differences were found between controls and LTA group. However, considering each group as a single environment, i.e., compute group-wise alpha diversity, results in significantly more diversity in the LTA group as compared to healthy controls. This means that dysregulated communities, associated with disorders, do not necessarily have to be less diverse than healthy communities. Furthermore, the state of dysregulation as a whole is more diverse than the stable healthy state. One conceivable explanation is that the LTA samples are a snapshot of the onset of OM while the HTA group represents late-stage disease, i.e., during the state of dysregulation a single pathogen found favorable conditions to become the dominant entity. This is congruent with the growing severity of symptoms during disease progression.

Due to the similarity in fungal beta diversity between a subclass of OM patients with healthy controls, ITS2 sequencing, followed by taxonomic annotation, might currently not be suited for the diagnosis of OM. Thus, diagnosis of OM still seems to be a domain of conventional microscopy and culture, which may be partially replaced by sequencing-based multiplex assays^{6,32}. However, ITS sequencing, in addition to conventional/molecular OM diagnosis, may provide significant clues regarding the underlying mechanics of infection and disease progression, i.e., whether or not additional host- and/or environmental factors need to be considered for treatment and prevention. Current treatment of OM predominantly targets the causative fungal pathogen, while considering host- and environmental factors, e.g. age, presence of diabetes or occlusive footwear, as risk factors for an unsatisfactory treatment response and recurrence³³. A more comprehensive understanding of disease progression and interaction with host- and environmental factors could potentially offer new targeted approaches for individual treatment. In this context the LTA group is of particular interest as the relationship between mycobial composition disease status is more ambiguous than in the HTA group.

The distribution of fungi in individual patients points towards potentially clinically relevant and interesting findings. More specifically, we observed a relative high abundance of *Neoscochyta graminicola* in 5/24 OM patients of the LTA group. By contrast, this species was not detected among the 20 most abundant species in the HTA or the healthy control group (Supplementary Fig. 1). *Neoscochyta* are commonly found on plants³⁴ and have not been associated with human disease so far. In addition, two OM patients in the LTA group show a predominance of *Podosphaera* (Fig. 2). This genus of fungi causes powdery mildew in plants³⁵. Up to date, there is no report on *Podosphaera* as the causative agent of OM or tinea. Hence, the presence of *Podosphaera* is most likely due to colonization rather than disease causing-agent. Furthermore, one OM patient in the LTA group had a high presence of the species *Schizopora paradoxa* (*Xylodon paradoxus*), which is a rot wood degrader commonly found in Korea³⁶. Again, this species has so far not been associated with human pathology. Thus, the detection of *S. paradoxa* most likely is most likely due to colonization. However, larger cohorts of OM patients are needed to confirm or reject these assumptions.

Our study is not without limitations: First, the age distribution among the groups is different. Healthy individuals are on average 20 years younger than the diseased patients. This may have an impact on the mycobiome as skin physiology changes with age³⁷. But this is, in accordance with the relevant literature, as it points out that OM predominantly affects the elderly^{2–4}. However, the setup of the experiments aimed at identification of OM patients confirmed by at least one diagnostic method and excluding clinically normally appearing controls where fungi were detected by any of the current diagnostic methods. We believe the later to be of more importance for this exploratory study. The impact of age on the fungal diversity would be an interesting follow-up study.

Second, the reported higher prevalence of OM in men^{2,3} is interesting with respect to the HTA and LTA group separation. We found a significant difference in distribution of the sexes between controls and HTA group. While the distribution of sexes in the LTA group is equal, the HTA is predominantly composed of male subjects. The existing bias towards males in the case group could be a result of the enhanced health-care behavior of women compared to that of men. It has been documented that men tend to delay seeking medical advice until a disorder becomes a severe inconvenience^{38,39}. Thus, we hypothesize that the higher abundance of *Trichophyton rubrum* is a phenomenon observed in infections at a later stage.

Collectively, we here provide new insights into the human nail mycobiome in health and OM. In addition to the detailed insights on fungal diversity, our results also indicate that OM develops either due to the high pathogenicity of certain fungal pathogens, or, alternatively, by yet-to-be-defined host or environmental factors.

Data availability

Sequencing data used for this study were submitted to the European Nucleotide Archive (ENA) and is available under accession number PRJEB37496.

Received: 9 July 2021; Accepted: 3 May 2022

Published online: 25 May 2022

References

- Gupta, A. K. *et al.* Systematic review of nondermatophyte mold onychomycosis: Diagnosis, clinical types, epidemiology, and treatment. *J. Am. Acad. Dermatol.* **66**(3), 494–502. <https://doi.org/10.1016/j.jaad.2011.02.038> (2012).
- Heikkilä, H. & Stubb, S. The prevalence of onychomycosis in Finland. *Br. J. Dermatol.* **133**(5), 699–703. <https://doi.org/10.1111/j.1365-2133.1995.tb02741.x> (2006).
- Piraccini, B. & Alessandrini, A. Onychomycosis: A review. *JoF* **1**(1), 30–43. <https://doi.org/10.3390/jof1010030> (2015).
- Roseeuw, D. Achilles foot screening project: Preliminary results of patients screened by dermatologists. *J. Eur. Acad. Dermatol. Venereol.* **12**(Suppl 1), S6–9 (1999) (**discussion S17**).
- Dhib, I. *et al.* Multiplex PCR assay for the detection of common dermatophyte nail infections. *Mycoses* **57**(1), 19–26. <https://doi.org/10.1111/myc.12096> (2014).
- Hayette, M.-P. *et al.* Clinical evaluation of the DermaGenius[®] Nail real-time PCR assay for the detection of dermatophytes and *Candida albicans* in nails. *Med. Mycol.* **57**(3), 277–283. <https://doi.org/10.1093/mmy/myy020> (2019).
- Ständer, S. *et al.* DNA-chip-based molecular testing as a clue for the diagnosis of tinea: A case series. *J. Eur. Acad. Dermatol. Venereol.* <https://doi.org/10.1111/jdv.17296> (2021).
- Gweon, H. S. *et al.* PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods Ecol. Evol.* **6**(8), 973–980. <https://doi.org/10.1111/2041-210X.12399> (2015).
- Chen, H. & Jiang, W. Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front. Microbiol.* **5**, 508. <https://doi.org/10.3389/fmicb.2014.00508> (2014).
- Schoch, C. L. *et al.* Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **109**(16), 6241–6246. <https://doi.org/10.1073/pnas.1117018109> (2012).
- Joyce, A., Gupta, A. K., Koenig, L., Wolcott, R. & Carviel, J. Fungal diversity and onychomycosis: An analysis of 8816 toenail samples using quantitative PCR and next-generation sequencing. *J. Am. Podiatr. Med. Assoc.* **109**(1), 57–63. <https://doi.org/10.7547/17-070> (2019).
- Findley, K. *et al.* Topographic diversity of fungal and bacterial communities in human skin. *Nature* **498**(7454), 367–370. <https://doi.org/10.1038/nature12171> (2013).
- Ständer, S. *et al.* The impact of lesional inflammatory cellular infiltrate on the phenotype of bullous pemphigoid. *J. Eur. Acad. Dermatol. Venereol.* <https://doi.org/10.1111/jdv.17303> (2021).
- 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**(16), 5261–5267. <https://doi.org/10.1128/AEM.00062-07> (2007).
- Nilsson, R. H. *et al.* The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* **47**(D1), D259–D264. <https://doi.org/10.1093/nar/gky1022> (2019).
- R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2008). Available: <http://www.R-project.org>.
- McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**(4), e61217. <https://doi.org/10.1371/journal.pone.0061217> (2013).
- Oksanen, J. *et al.* *vegan: Community Ecology Package*. Available: <https://CRAN.R-project.org/package=vegan> (2019).
- Silverman, J. D., Washburne, A. D., Mukherjee, S. & David, L. A. A phylogenetic transform enhances analysis of compositional microbiota data. *eLife* **6**, e21887. <https://doi.org/10.7554/eLife.21887> (2017).
- Willis, A. D. & Martin, B. D. DivNet: Estimating diversity in networked communities. *Bioinformatics*, Preprint. <https://doi.org/10.1101/305045> (2018).
- Kuhn, M. *caret: Classification and Regression Training*. Available: <https://CRAN.R-project.org/package=caret> (2020).
- Cáceres, M. D. & Legendre, P. Associations between species and groups of sites: Indices and statistical inference. *Ecology* **90**(12), 3566–3574. <https://doi.org/10.1890/08-1823.1> (2009).
- Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016). Available: <https://ggplot2.tidyverse.org>.
- Kassambara, A. *ggpubr: “ggplot2” Based Publication Ready Plots* (2019).
- Goodrich, B., Gabry, J., Ali, I. & Brilleman, S. *rstanarm: Bayesian applied regression modeling via Stan*. Available: <https://mc-stan.org/rstanarm> (2020).
- Lüdtke, D. *sjPlot: Data Visualization for Statistics in Social Science*. Available: <https://CRAN.R-project.org/package=sjPlot> (2020).
- Breiman, L. Random forests. *Mach. Learn.* **45**(1), 5–32. <https://doi.org/10.1023/A:1010933404324> (2001).
- Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**(3), 18–22 (2002).
- Willis, A. D. Rarefaction, alpha diversity, and statistics. *Front. Microbiol.* **10**, 2407. <https://doi.org/10.3389/fmicb.2019.02407> (2019).

30. Willis, A., Bunge, J. & Whitman, T. Improved detection of changes in species richness in high diversity microbial communities. *J. R. Stat. Soc. C* **66**(5), 963–977. <https://doi.org/10.1111/rssc.12206> (2017).
31. Khosravi, A. R. *et al.* Yeasts as important agents of onychomycosis: In vitro activity of propolis against yeasts isolated from patients with nail infection. *J. Altern. Complement. Med.* **19**(1), 57–62. <https://doi.org/10.1089/acm.2011.0722> (2013).
32. Koo, S. H. *et al.* Development and validation of a real-time multiplex PCR assay for the detection of dermatophytes and *Fusarium* spp. *J. Med. Microbiol.* **68**(11), 1641–1648. <https://doi.org/10.1099/jmm.0.001082> (2019).
33. Lipner, S. R. & Scher, R. K. Onychomycosis: Treatment and prevention of recurrence. *J. Am. Acad. Dermatol.* **80**(4), 853–867. <https://doi.org/10.1016/j.jaad.2018.05.1260> (2019).
34. Crous, P. W. *et al.* Fungal Planet description sheets: 469–557. *Persoonia* **37**, 218–403. <https://doi.org/10.3767/003158516X694499> (2016).
35. Kim, S. *et al.* Genome resource of *Podosphaera xanthii*, the host-specific fungal pathogen that causes cucurbit powdery mildew. *Mol. Plant Microbe Interact.* **34**(4), 457–459. <https://doi.org/10.1094/MPMI-11-20-0307-A> (2021).
36. Min, B. *et al.* Genome sequence of a white rot fungus *Schizophora paradoxa* KUC8140 for wood decay and mycoremediation. *J. Biotechnol.* **211**, 42–43. <https://doi.org/10.1016/j.jbiotec.2015.06.426> (2015).
37. Tončić, R. J., Kezić, S., Hadžavdić, S. L. & Marinović, B. Skin barrier and dry skin in the mature patient. *Clin. Dermatol.* **36**(2), 109–115. <https://doi.org/10.1016/j.clindermatol.2017.10.002> (2018).
38. Thompson, A. E. *et al.* The influence of gender and other patient characteristics on health care-seeking behaviour: A QUALICOPC study. *BMC Fam. Pract.* **17**(1), 38. <https://doi.org/10.1186/s12875-016-0440-0> (2016).
39. Redondo-Sendino, A., Guallar-Castillón, P., Banegas, J. R. & Rodríguez-Artalejo, F. Gender differences in the utilization of health-care services among the older adult population of Spain. *BMC Public Health* **6**, 155. <https://doi.org/10.1186/1471-2458-6-155> (2006).

Acknowledgements

We are thankful for the computational support from the OMICS compute cluster at the University of Lübeck.

Author contributions

Conception of the work: H.B., A.K. and R.J.L. Acquisition, analysis, or interpretation of data: all authors. Drafting the work or critical revision for important intellectual content: all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL. Research Training Group “Genes, Environment and Inflammation” (GRK1743), the Cluster of Excellence “Precision Medicine in Chronic Inflammation” (EXC2167) from the Deutsche Forschungsgemeinschaft, and the Schleswig–Holstein Excellence-Chair Program from the State of Schleswig Holstein.

Competing interests

MH is an employee of Euroimmun AG that has commercialized a DNA-chip based diagnostic tool for the diagnosis of fungal skin infections, DZ and RJL have received funding for research from Euroimmun AG. All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13074-8>.

Correspondence and requests for materials should be addressed to H.B. or R.J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022