Microbial Ecology
https://doi.org/10.1007/s00248-021-01730-x

FUNGAL MICROBIOLOGY

Hawaiian Fungal Amplicon Sequence Variants Reveal Otherwise Hidden Biogeography

Laura Tipton[1] · Geoffrey L. Zahn[2] · John L. Darcy[3] · Anthony S. Amend[4] · Nicole A. Hynson[5]

Received: 9 October 2020 / Accepted: 7 February 2021

Abstract

To study biogeography and other ecological patterns of microorganisms, including fungi, scientists have been using operational taxonomic units (OTUs) as representations of species or species hypotheses. However, when defined by 97% sequence similarity cutoff at an accepted barcode locus such as 16S in bacteria or ITS in fungi, these OTUs can obscure biogeographic patterns, mask taxonomic diversity, and hinder meta-analyses. Amplicon sequence variants (ASVs) have been proposed to alleviate all of these issues and have been shown to do so in bacteria. Analyzing ASVs is just emerging as a common practice among fungal studies, and it is unclear whether the benefits found in bacterial studies of using such an approach carryover to fungi. Here, we conducted a meta-analysis of Hawaiian fungi by analyzing ITS1 amplicon sequencing data as ASVs and exploring ecological patterns. These surveys spanned three island groups and five ecosystems combined into the first comprehensive Hawaiian Mycobiome ASV Database. Our results show that ASVs can be used to combine fungal ITS surveys, increase reproducibility, and maintain the broad ecological patterns observed with OTUs, including diversity orderings. Additionally, the ASVs that comprise some of the most common OTUs in our database reveals some island specialists, indicating that traditional OTU clustering can obscure important biogeographic patterns. We recommend that future fungal studies, especially those aimed at assessing biogeography, analyze ASVs rather than OTUs. We conclude that similar to bacterial studies, ASVs improve reproducibility and data sharing for fungal studies.

Keywords ITS · ASV · OTU · Mycobiome · Hawaiian Islands · Island biogeography

Introduction

For cryptic microorganisms such as fungi and bacteria, operational taxonomic units (OTUs) have been the currency used as species equivalents in targeted amplicon sequencing for as long as DNA sequencing has been available [1]. OTUs are often used to assess community diversity and membership in environmental and often hyper-diverse study systems such as soil, water, and host-associated microbiomes [2–4]. Most commonly defined as 97% sequence similarity of the target barcode region (16S for bacteria, ITS for fungi; [5, 6]) using one of many types of agglomerative clustering algorithms, OTUs have been likened to the equivalent of microbial species [7], or species hypotheses, despite our limited ability to confidently assign taxonomies at the species level for many groups of microbes [8, 9]. For the purposes of this analysis, our OTUs are defined as clusters delimited by the 97% sequence similarity level but can be generalized to any level of sequence similarity.

While OTUs based on short sequence reads of hypervariable genome regions may be limited in their ability to resolve microbial taxonomy, they have been used in fungi to evaluate many ecological patterns including latitudinal gradients [10], endemism [11], neutral community assembly [12], and relationships between diversity and soil biochemistry [13]. However, comparisons of OTUs across studies are

✉ Nicole A. Hynson
nhynson@hawaii.edu

[1] School of Natural Science and Mathematics, Chaminade University of Honolulu, Honolulu, HI, USA

[2] Biology Department, Utah Valley University, Orem, UT, USA

[3] Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[4] School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, USA

[5] Pacific Biosciences Research Center, University of Hawai'i at Mānoa, Honolulu, HI, USA

Published online: 20 March 2021

challenging given the inherent pseudo-randomness involved in the clustering algorithms, which can arise from either the data itself or from heuristic shortcuts designed to speed up the process. Thus, synthesizing across studies has been a major obstacle and limits our ability to divine large-scale patterns of fungal ecology and biogeography.

Recently, oligotyping and amplicon sequence variants (ASVs) were devised to examine the diversity of bacteria among habitats at a finer scale than OTUs allowed. This is because both oligotyping and ASVs were designed to target variation at the finest level possible within a given gene region, such as bacterial strains [14, 15]. Whereas OTUs are generated by clustering together reads at a given similarity threshold, oligotypes and ASVs are formed by combining identical reads and then using "denoising" procedures to determine and remove or consolidate those variants that are more likely to be caused by PCR error (that would have been absorbed by the similarity threshold in OTUs) than natural variation. While there are differences in the methods used to generate oligotypes and ASVs, we refer to them both as ASVs for the purposes of this analysis. In bacteria, the strain-level resolution provided by ASVs is desirable given the amount of functional variability within a genus and even within a species [16, 17]. However, the use of ASVs to resolve finer scale biotic and abiotic interactions for other microorganisms such as fungi [18] and diatoms [19], and in environmental DNA surveys [20], is only just beginning.

An additional potential benefit of ASVs versus OTUs is that while OTUs, particularly those chosen de novo, are study-specific and represent clouds of similar sequences, ASVs are exact matches, meaning they are clearly delineated real sequences and are comparable. This facilitates meta-analyses or combining of data across studies. Generating ASVs across experiments also allows for the development of databases that are location or system specific and that can be appended with any new ASVs discovered by the latest experiment. Such databases have the potential to reduce inter-experiment variability, a known problem in internal transcribed spacer (ITS) amplicons for fungi [21], by modeling and excluding experiment-specific errors while preserving true biological diversity. Location-specific ASV databases have the potential to reveal micro-niches undetected in single studies, while system-specific ASV databases can be used to reveal patterns on geographic or time scales too large for a single study to capture.

We set out to create a fungal location-specific ASV database for the Hawaiian Islands and perform a meta-analysis of fungal community surveys examining different ecosystems (Table 1). In order to reproducibly combine these 10 surveys, each dataset was reanalyzed using ASVs rather than the OTUs used in each original analysis (Table 2). The resulting ASVs were combined into one large Hawaiian fungal ASV database, which covers three islands or island clusters: Hawaii Island, Maui Nui (consisting of Maui, Molokai, Lanai, and Kahoolawe), and Oahu and five habitats: plant foliar epiphytes (outside the plant), plant foliar endophytes (inside the plant), marine, soil, and air (Table 2, Supplemental File 1). Once compiled, we were able to use this Hawaiian fungal ASV database to test whether ASVs produced finer granularity or recapitulate both large- and small-scale ecological patterns, specifically measures of diversity, island specialization, and species abundance distributions relative to traditional OTUs.

Methods

Datasets

We compiled ITS sequences from 10 fungal surveys of the Hawaiian Islands (Table 1). These surveys spanned the major Hawaiian Islands; however, we combined samples from Molokai, Lanai, and Maui into "Maui Nui" (a geologic term that encompasses Maui, Molokai, Lanai, and Kahoolawe based on their shared origin as a single island 1.2 million years ago; [28]) to make the sampling more even across islands. Each survey had sequenced the ITS1 region on the Illumina MiSeq platform. In order to process all datasets the same way, we obtained the .fastq files for each sample in each dataset.

Sequence Processing and ASV Formation

Each sample was run through ITSx version 1.1b [29] using custom R scripts to facilitate .fastq processing, resulting in .fastq files containing only ITS1 region for each read. These custom scripts are available at https://github.com/darcyj/fastq-from-ITSx. The extracted ITS1 reads were then used to create ASVs using dada2 version 1.6.0 [30]. We used dada2 because it has been updated to accommodate ASVs of different lengths, a common attribute of fungal ITS amplicons that is relatively rare in 16S bacterial amplicons. An error model was calculated for each survey and used to create 10 independent ASV tables. These tables were merged, bimeras (two-parent chimeras) were removed, and identical ASVs were collapsed across surveys. The resulting table contained 2118 samples and 44,383 ASVs. We removed ASVs present in only one sample (26,338 or 60%) and samples that then contained only one ASV (224 or 11%), resulting in an ASV table with 1894 samples and 18,045 ASVs. This table was used for all subsequent analyses.

OTU Formation

In order to compare the ASVs to OTUs while preserving the data cleanup performed during ASV creation, we formed OTUs by combining the ASVs at 97% similarity using the

Table 1 Fungal surveys

Survey name	Habitat(s)	Island(s)	Number of samples	Publication	SRA accession
BIG	Endophytes	Hawaii	276	[22]	PRJNA474551
FEF2	Endophytes	Oahu, Hawaii, Maui Nui	190	[23]	PRJNA470970
FEF3	Endophytes	Maui Nui, Hawaii	252	[23]	PRJNA470970
Mesophotic algae	Marine	Maui Nui	23	[24]	PRJNA355018
MLO	Aerobiota	Hawaii, Oahu	382	[25]	PRJNA386517
Nguyen	Soil	Oahu	5	Unpublished	
Peay, Vitousek	Soil	Hawaii	156	[26]	PRJNA379981
Swabs	Epiphytes	Oahu	195	[24]	PRJNA355011
Lau	Epiphytes, Marine	Oahu	335	Unpublished	PRJNA594061
Kipauka	Soil	Hawaii	80	[27]	PRJNA316729

uclust method [31] in QIIME version 1.9 [32] without any further filtering. This resulted in 10,713 OTUs, of which 7265 (68%) were composed of a single ASV.

Taxonomic Assignment

For the largest OTUs (see results for definitions of largest), we assigned taxonomies to all ASVs by top megaBLAST hit to the NCBI nucleotide database and by top massBLASTer hit to the UNITE fungal database using the criteria: percent identity > 98%, query coverage = 100%, bit score > 240, *E*-value < 1E-61 (SI Tables 1-3) [33]. In the event of multiple, equally scored top hits, we recorded the first fungal hit. Taxonomic assignments were conducted after ASV formation with no collapsing of ASVs based on taxonomy.

Diversity Metrics

Initial diversity metrics, including both alpha and beta diversity, were calculated in the *vegan* R package version 2.5-1 [34]. Hill numbers, which also measure alpha diversity, were calculated in the *hillR* R package version 0.4.0 [35]. The first Hill number ($q = 0$) is an estimate of richness without regard for relative abundance. The second Hill number ($q = 1$) is the exponential of Shannon entropy, related to the effective number of common OTUs, while the third Hill number ($q = 1$) is the inverse Simpson concentration, related to the effective number of dominant OTUs [35]. Finally, for the procrustes analysis of beta diversity, samples with fewer than 1000 reads ($N = 600$) were removed and binary Jaccard distance was used to avoid conflating read and species abundance.

Table 2 Final sample counts by island and ecosystem

Ecosystem	Hawaii	Maui Nui	Oahu
Soil	236	0	5
Epiphytes	0	0	287
Aerobiota	374	0	6
Endophytes	365	121	157
Marine	0	23	175

Octaves

To compare cluster abundances distributions, abundances were binned by log base 2 abundance cutoffs or octaves [36], using the *sads* R package version 0.4.2 [37]. The distributions were compared using a non-parametric Kolmogorov-Smirnov test [38, 39].

Specialization

As a measure of fungal specialization, bipartite networks were formed between islands and OTUs or ASVs. An edge was drawn if the OTU or ASV was found in a sample from the island. Specialization within the resulting network was measured by the H2' index [40]. A larger H2' indicates more OTUs/ASVs that are more selective or specialized to their island or ecosystem and more islands that are OTU/ASV specific. This was calculated using the H2fun() in the R package *bipartite* version 2.08 [38, 39].

Results

We examined the biogeographic structure by island and ecosystem of the ASVs that make up the largest (defined either as containing the most ASVs, being the most abundant, or being the most ubiquitous) OTUs in our database. OTU

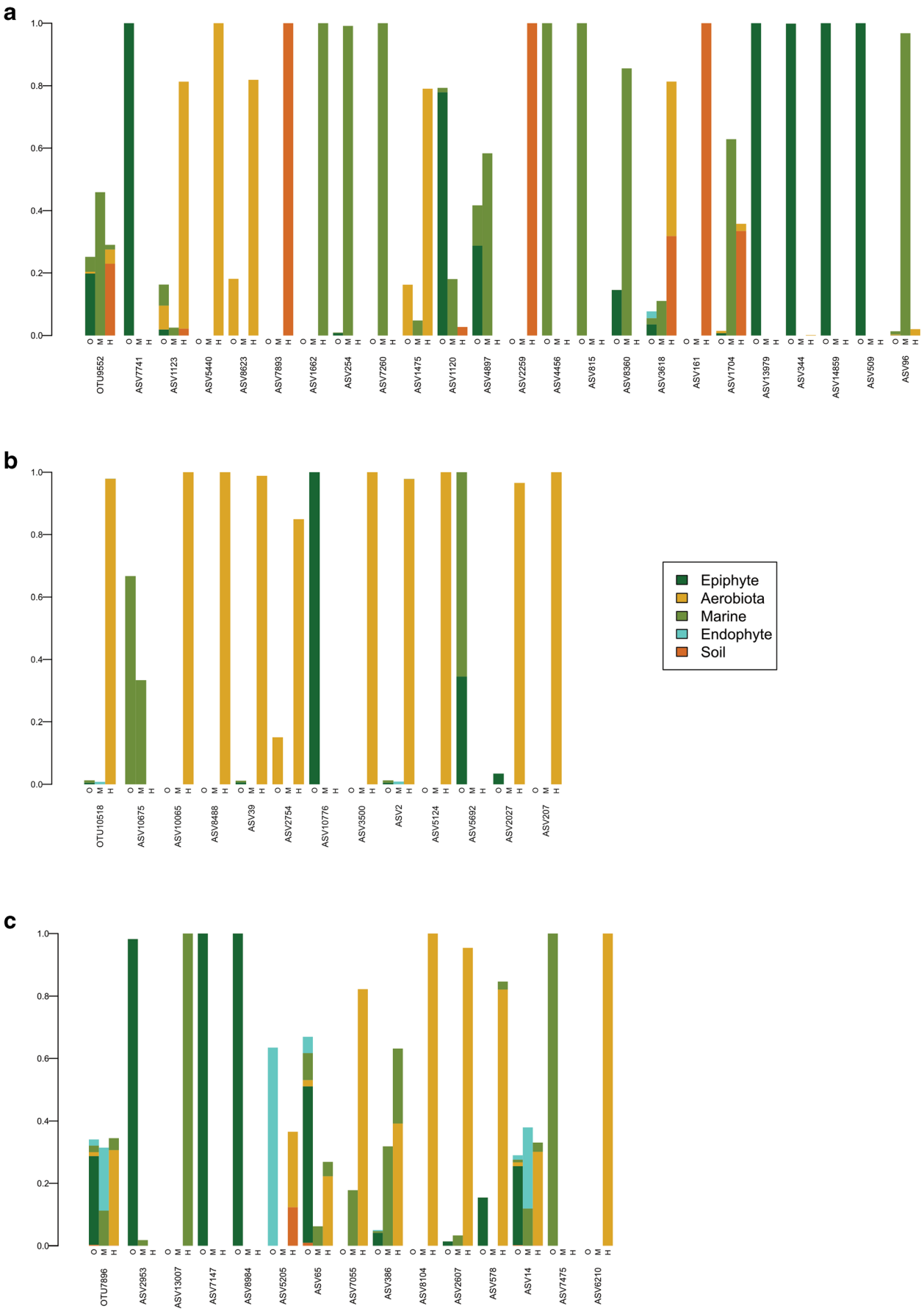
"denovo9552" (*Didymella* sp., a possible plant pathogen) was made up of the largest number of ASVs and was found occurring across three islands and island groups: Oahu (25.2% of reads), Maui Nui (45.8% of reads), and Hawaii (29.0% of reads), yet twelve (52%) of the ASVs contained in this OTU were found on single islands only (Fig. 1a), including ASV161 which had the highest ubiquity of any ASVs in the OTU (present in 57 out of 241 samples), which was found exclusively in soil samples from Hawaii Island. Conversely, the OTU with the highest average relative abundance across all samples ("denovo10518," *Plectosphaerella* sp., a plant pathogen) was seen predominantly on Hawaii Island (97.9% of reads), yet two of its 12 ASVs were seen exclusively on Oahu (Fig. 1b). The most ubiquitous OTU ("denovo7896," *Cladosporium* sp., a common mold) was seen in 869 out of 2118 (41%) samples, evenly distributed across the three islands: 34.1% of reads in samples from Oahu, 31.4% of reads in samples from Maui Nui, and 34.5% of reads in samples from Hawaii. Four of the 14 ASVs that make up this OTU were seen on all three islands, but six ASVs were seen on single islands only—three on Oahu only and three on Hawaii only (Fig. 1c). The island endemism described here is not attributed to sample type, as only one ASVs in each of the three OTUs described were seen in single samples; the median number of samples each ASV was seen in was 8 samples (IQR 2–16) for denovo9552, 3.5 samples (IQR 2–48.8) for denovo10518, and 4 samples (IQR 2–40.5) for denovo7896 (Supplemental Figure 1).

Despite most OTUs containing single ASVs (68%), the remaining OTUs were found to mask taxonomic variation relative to ASVs. For example, the representative sequence for OTU denovo9552 was taxonomically assigned to a single *Didymella* sp. in the NCBI database, and while most (96%) of the other ASVs within this OTU were assigned to various members of the Didymellaceae, only one other ASV was assigned to the same GenBank accession number as the representative OTU sequence (Supplemental Table 1). Additionally, one of the 23 ASVs was assigned to a fungus outside the Didymellaceae, *Leptosphaerulina trifolii*, which is a member of the Pleosporaceae in the same order as the Didymellaceae. For the same OTU, using the UNITE curated taxonomic reference database resulted in twenty ASVs (87%) assigned to the same species hypothesis (SH), *Didymella exigua*. One of the ASVs assigned to a different SH, ASV161, was assigned only to the phylum Ascomycota, but NCBI assigned this ASV to the species *Epicoccum hordei*, based on similarity to a voucher sequence that was deposited in 2018 [41]. Similarly, the most abundant OTU, denovo10518, had all but one ASV, out of 12, assigned to members of the *Plectosphaerella* genus in the NCBI database (Supplemental Table 2). Assignments based on the UNITE database were more consistent with 11 ASVs being assigned to the same SH, SH1644387.08FU *Plectosphaerella oratosquillae*. However, the single ASV assigned to a different SH was the first or representative ASV that would be used for OTU taxonomic assignment; this ASV was assigned to SH1644410.08FU *Plectosphaerella cucumerina*.

The largest discrepancy between NCBI and UNITE databases came in the most ubiquitous OTU, denovo7896 (Supplemental Table 3). Using the NCBI database, 11 ASVs (79%) were assigned to members of the genus *Cladosporium*, but using the UNITE, 4 ASVs (29%), including the OTU reference sequence, were assigned to SH1572792.08FU *Mycosphaerella tassiana* and 8 ASVs (57%) were assigned to SH1572816.08FU which could only be named to Dothideomycetes, to which *Mycosphaerella tassiana* belongs. The remaining ASVs in the OTU were assigned to uncultured fungal clones in the NCBI database on simply fungi in the UNITE database. We are unable to identify the cause of such taxonomic discrepancies, which may be owed to misnamed or undersampled taxa in the databases or to limited taxonomically informative data from short sequence reads such as those analyzed here.

Diversity metrics changed minimally between OTUs and ASV datasets. Using ASVs, there was an inevitable increase in richness, but this increase was uniform across all ecosystems (Fig. 2). For richness, or Hill number 0, the average increase from OTUs to ASVs was 6.65 "species" (SD 9.40) per sample. The other Hill numbers also increased as expected; entropy, or Hill number 1 exponential Shannon index, increased by 3.06 (SD 6.33) and diversity, or Hill number 2 inverse Simpson index, increased by 1.13 (SD 2.79). These small upticks in diversity metrics did not change the ranking of samples from most to least diverse. Similarly, Jaccard distance beta diversity distances rankings were maintained and clusters of the samples were preserved when switching from OTUs to ASVs (Fig. 3). A symmetric procrustes analysis to directly compare the nonmetric multidimensional scaling of the Jaccard distance beta diversities reveals a sum of squares (m12 squared) of 0.000025 between the OTU and ASV Jaccard-based matrices, where a perfect match would have a sum of squares of 0. Due possibly to the relatively limited geographic distance (less than 600 km) sampled confounded by the breadth of the habitats examined, we saw no discernable pattern between physical distance and Jaccard distance beta diversity (Supplemental Figure 2). This lack of pattern was observed in both OTU and ASV-based distances.

Other broad-scale ecological patterns were also similar whether looking at OTUs or ASVs. The species abundance distributions, based on read abundances, of the OTUs and ASVs, when binned by octaves or a doubling of abundance, are not significantly different from each other (K-S test p-value = 0.2591). Both distributions peak in the 6th octave, which spans 32 to 64 reads (Fig. 4). We then looked at island specificity using a whole-network measure of specialization, H2', where a value of 0 would indicate a perfectly generalized


a
b
c
Epiphyte
Aerobiota
Marine
Endophyte
Soil
OTU9552
ASV7741
ASV1123
ASV5440
ASV8623
ASV7893
ASV1662
ASV254
ASV7260
ASV1475
ASV1120
ASV4897
ASV2259
ASV4456
ASV815
ASV8360
ASV3618
ASV161
ASV1704
ASV13979
ASV344
ASV14859
ASV509
ASV96
OTU10518
ASV10675
ASV10065
ASV8488
ASV39
ASV2754
ASV10776
ASV3500
ASV2
ASV5124
ASV5692
ASV2027
ASV207
OTU7896
ASV2953
ASV13007
ASV7147
ASV8984
ASV5205
ASV65
ASV7055
ASV386
ASV8104
ASV2607
ASV578
ASV14
ASV7475
ASV6210

◀ **Fig. 1** Islands and ecosystems where reads making up OTUs were detected. Each OTU/ASV along the X-axis is broken down into the percentage of reads originating from each island (represented by a trio of abutting bars) and ecosystem (represented by color). The left-most bar trio represents the entire OTU while the remaining bars each represent an ASV that is contained in the OTU. **a** OTU denovo9552 contains 23 ASVs, the most of all OTUs. **b** OTU denovo10518, the most abundant OTU in the dataset, was found predominantly on the island of Hawaii, but the 12 ASVs it contains were spread out across the three islands. **c** OTU denovo7896, the OTU seen in the most samples, was evenly distributed across the three islands, but the 14 ASVs it contains are not

network with every taxa occurring on every island and values approaching 1 indicating a more specialized network with each taxa occurring only on a single island. Using OTUs, our fungal communities had an H2' of 0.756, while using ASVs the communities had an H2' of 0.827. Both are indicative of highly specialized or specific relationships between taxa and islands within the networks.

Discussion

Identifying large-scale ecological and biogeographic patterns among fungal communities has been a technical and logistical challenge. Technically, the OTU-based approaches that have dominated the field of fungal molecular ecology for the last 25 years limit comparisons across studies and do not lend themselves to meta-analysis due to differences among clustering algorithms and random seeding. Logistically, it is generally unrealistic for single lab groups to perform global surveys of fungi (but see [42, 43] for groups that are able to mount global surveys). By adopting an ASV-based approach, we are now able to overcome these challenges. As proof of concept, we performed a meta-analysis of fungi sampled from five habitats and across three of the major Hawaiian Islands. After combining these 10 surveys into a Hawaiian fungal database of ASVs, patterns of fungal biogeography, short-comings of taxonomic databases, and stability between ASVs and OTUs for broad ecological patterns were revealed.

By taking the approach of examining biogeographic patterns of fungi via ASVs, new information has emerged that is masked by the traditional OTU methods. For example, the three largest OTUs all contained ASVs that were found only on single islands (Fig. 1). This island specificity is masked using OTU-based clustering where all three OTUs appear on all islands. Further evidence of this concealed specificity was shown by the H2' specificity index between taxa and islands, which increased when measuring taxa by ASVs instead of OTUs. This becomes particularly salient in light of concerns surrounding local species extinctions or extirpations and microbial conservation [44], as well as attempts to resolve microbial geographic distributions. However, that broad patterns of distance-based changes in community composition and species abundance distributions (SADs) were basically recapitulated by ASVs indicates that overall patterns of community structure are robust independent of clustering algorithms. What ASVs contribute that traditional OTUs may obscure are hypotheses for which taxa may be of particular concern or interest for conservation purposes or future studies of host/habitat specificity.

Given that early ASV studies were used to show bacterial community differences within the human mouth that were not seen by the initial OTU-based studies [45], it is not surprising that we find fungal community differences among habitats that were obscured by OTUs. These findings have important implications for understanding the distribution of fungi, potentially at the population-scale. Because the ITS spacer region is not conserved and should not be under strong

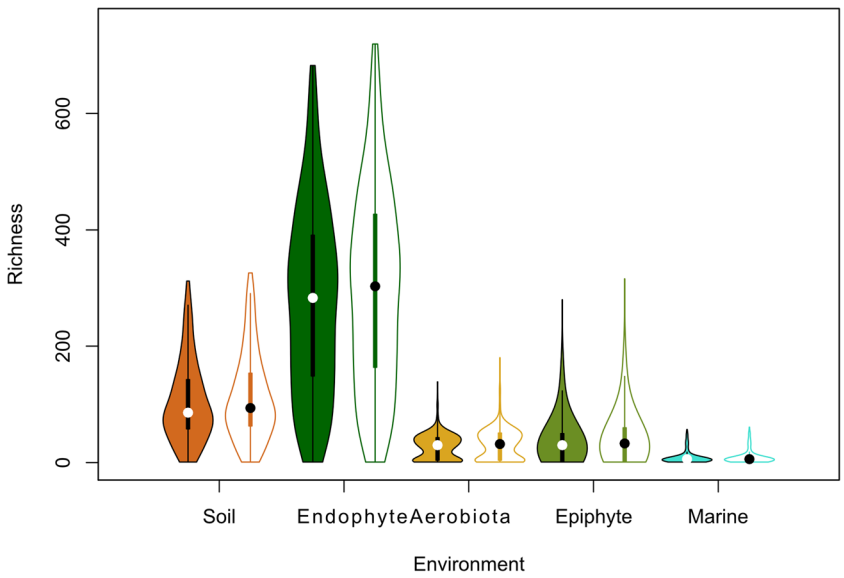


Fig. 2 Violin plots of richness by the ecosystem. The distribution of fungal community richness (Hill number 0), as measured by a number of OTUs or ASVs observed, for each sample is plotted by ecosystem, regardless of source study. Richness for OTU analyses is shown in solid violins; richness for ASV analyses is shown in empty violins. The increase in observed ASVs compared to OTUs is expected, but, importantly, the ordering of ecosystems by richness is preserved

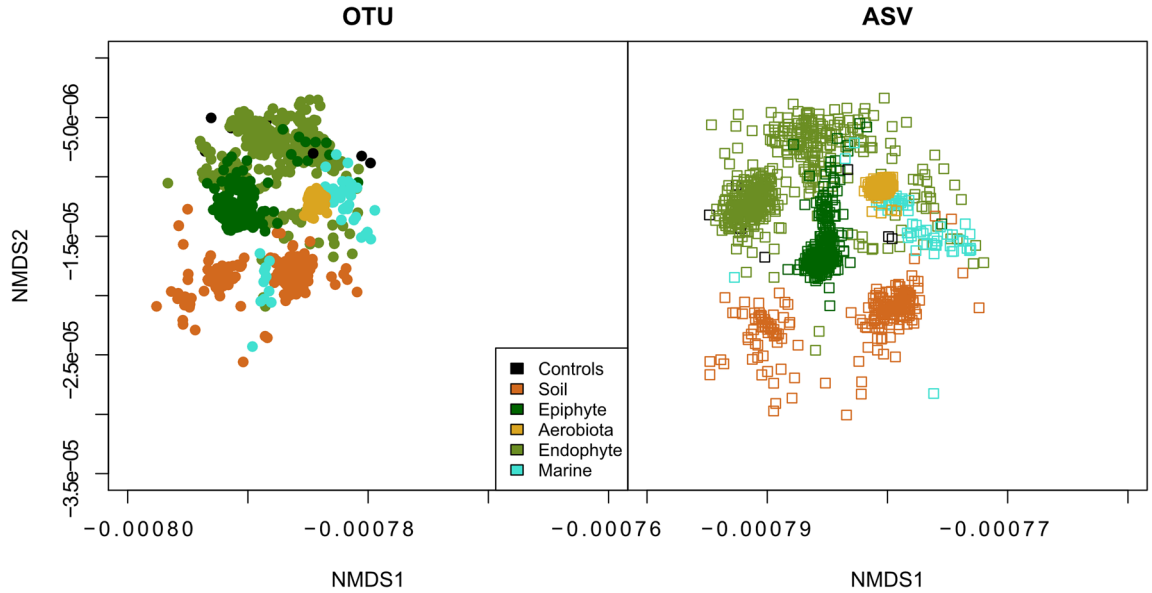


Fig. 3 NMDS plot. Ordination of all fungal surveys based on Jaccard distance measured between OTUs on the right and ASVs on the left. Each point represents a sample; each color represents an ecosystem. Both OTUs and ASVs show distinct grouping by ecosystem

selection, variation among fungal ASVs is likely due to founder effects or random mutations. Follow-up studies comparing island strains across more conserved regions of their genomes are needed to assess the relative importance of the ITS region as an indicator of natural selection.

Despite clustering at ≥ 97% similarity, we found that the taxonomy of ASVs from many of our OTUs was incongruent with their representative OTU sequences. Three possible causes for these discrepancies are (1) incorrect or inconsistent identifications within and between databases such as NCBI's GenBank [46] and UNITE [33] which could lead to the same fungus being assigned different taxonomy based on its sequence similarity to differently named accessions, (2) insufficient variation of the taxon in the database including absent conspecifics, particularly in smaller databases such as UNITE, or (3) similar to bacterial 16S ASVs, fungal ITS ASVs provide better resolution than OTU-based clustering and variation in taxonomic assignments reflects true biological variation. Without curated databases that include voucher specimens from the geographic area of interest, it remains challenging to determine which of these three factors dominates. However, if it is the latter, the additional taxonomic resolution revealed by ASVs is particularly important when it comes to downstream taxonomy-based analyses such as FUNGuild and

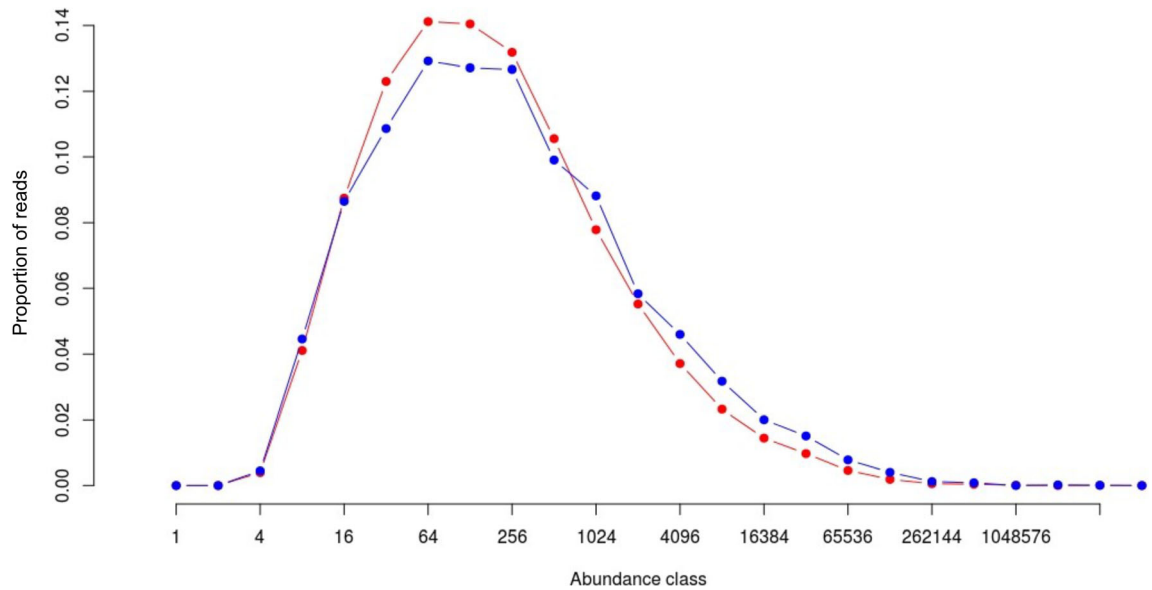


Fig. 4 Octave plot. Cluster abundance distributions for OTUs (in blue) and ASVs (in red). Cluster abundances are binned by octave or log base 2 increments meaning each octave contains OTUs or ASVs with double the abundance, measured by the number of reads, of the previous octave

habitat origins [47, 48] as these traits may not be conserved at the genus or family level represented by the OTU.

Like others before us, we found that while slightly inflating richness, ASVs recapitulated the broad ecological patterns seen by OTUs [18]. The use of ASVs to study ecological patterns, such as species abundance distributions, is dependent upon the same assumptions as using OTUs, specifically that relative read abundance is proportional to relative taxon abundance and that the community was sufficiently and unbiasedly sampled. These assumptions have long been part of the discussion around DNA-based fungal surveys [49, 50] and should be considered at the onset of any new study. Nevertheless, depending upon the questions being asked and the goals of the study, OTU-based approaches may remain valid and comparable to ASVs.

Conclusions

By performing a meta-analysis of fungal communities across the Hawaiian Islands and comparing ASV-based and OTU-based results, we have shown that ASVs can be used to combine fungal surveys reproducibly, in much the same way they can be used for bacterial surveys. While additional biogeographic and potentially taxonomic information can be gained by using ASVs, the results of OTU-based studies are unlikely to be overturned by re-analyzing the data. However, analyzing the sequences resulting from older studies as ASVs eases meta-analyses such as the ones performed here. The ASV-based approach also results in otherwise obscured patterns that can lead to additional and fruitful lines of inquiry, especially in the context of meta-community studies, and mapping the fine and large-scale spatial distributions of fungi. Going forward, we recommend fungal researchers consider the use ASVs rather than OTUs in order to improve reproducibility, enable meta-analyses or comparisons across studies, and, where applicable, gain the most information possible about biogeographic patterns.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00248-021-01730-x.

Acknowledgments The authors would like to thank Dr. Nhu Nguyen for contributing previously unpublished data. We would also like to thank Katie Lund for assisting with sample processing. This research was made possible in part by an award to A.S.A. and N.A.H. from the W.M. Keck Foundation and The National Science Foundation, award numbers 1556856 & 1255972. L.T. was funded by the Alfred P. Sloan Foundation Microbiome of the Built Environment Postdoctoral Fellowship and would like to thank the Hawaii Data Science Institute and the UH Cyber Infrastructure team for their HPC support.

Data Accessibility and Code Availability DNA sequences are in the NCBI SRA under the accession numbers shown in Table 1. Final ASV table, sample metadata, and code can be found at https://github.com/ltipton/HIMycobiome. Complete ASV database will be uploaded to the Hawaii Data Science Institute repository at https://himycobiome.its.hawaii.edu.

Author Contributions N.A.H., A.S.A., and L.T. designed the research; G.L.Z. and A.S.A. performed the research; L.T., J.L.D., and G.Z. analyzed the data; all authors contributed to manuscript preparation and editing.

Funding This research was made possible in part by an award to A.S.A. and N.A.H. from the W.M. Keck Foundation and The National Science Foundation, award numbers 1556856 & 1255972. L.T. was funded by the Alfred P. Sloan Foundation Microbiome of the Built Environment Postdoctoral Fellowship.

Declarations

Conflict of Interest The authors have no conflict of interest.

References

1. Benlloch S, Martínez-Murcia AJ, Rodríguez-Valera F (1995) Sequencing of bacterial and archaeal 16S rRNA genes directly amplified from a hypersaline environment. Syst Appl Microbiol 18:574–581. https://doi.org/10.1016/S0723-2020(11)80418-2
2. Group JCHMPDGW (2012) Evaluation of 16S rDNA-based community profiling for human microbiome research. PLoS One 7: e39315. https://doi.org/10.1371/journal.pone.0039315
3. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Boss E, Bowler C, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sieracki M, Velayoudon D (2015) Structure and function of the global ocean microbiome. Science 348:1261359. https://doi.org/10.1126/science.1261359
4. Thompson LR, Sanders JG, McDonald D et al (2017) A communal catalogue reveals Earth's multiscale microbial diversity. Nature. https://doi.org/10.1038/nature24621
5. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An KD, Bai FY, Barreto RW, Begerow D, Bergeron MJ, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS, del-Prado R, Dentinger B, Dieguez-Uribeondo J, Divakar PK, Douglas B, Duenas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, Garcia MA, Ge ZW, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo LD, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera C, Hirayama K, Hirooka Y, Ho HM, Hoffmann K, Hofstetter V, Hognabba F, Hollingsworth PM, Hong SB, Hosaka K, Houbraken J, Hughes K, Huhtinen S,

Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EBG, Kelly LJ, Kirk PM, Knapp DG, Koljalg U, Kovacs GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SSN, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo JM, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul TL, Ruibal C, Sarmiento-Ramirez JM, Schmitt I, Schussler A, Shearer C, Sotome K, Stefani FOP, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh SO, Sung GH, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vagvolgyi C, Vialle A, Vu TD, Walther G, Wang QM, Wang Y, Weir BS, Weiss M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora JC, Zhang N, Zhuang WY, Schindel D (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. PNAS 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

6. Weisburg WG, Barns SM, Pelletier DA, Lane DJ (1991) 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173:697–703. https://doi.org/10.1128/jb.173.2.697-703.1991
7. Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 44:846–849. https://doi.org/10.1099/00207713-44-4-846
8. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9:e93827. https://doi.org/10.1371/journal.pone.0093827
9. Will KW, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55. https://doi.org/10.1111/j.1096-0031.2003.00008.x
10. Khomich M, Davey ML, Kauserud H, Rasconi S, Andersen T (2017) Fungal communities in Scandinavian lakes along a longitudinal gradient. Fungal Ecol 27:36–46. https://doi.org/10.1016/j.funeco.2017.01.008
11. Talbot JM, Bruns TD, Taylor JW, Smith DP, Branco S, Glassman SI, Erlandson S, Vilgalys R, Liao HL, Smith ME, Peay KG (2014) Endemism and functional convergence across the North American soil mycobiome. Proc Natl Acad Sci 111:6341–6346. https://doi.org/10.1073/pnas.1402584111
12. Cui L, Lucht L, Tipton L, Rogers MB, Fitch A, Kessinger C, Camp D, Kingsley L, Leo N, Greenblatt RM, Fong S, Stone S, Dermand JC, Kleerup EC, Huang L, Morris A, Ghedin E (2015) Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med 191:932–942. https://doi.org/10.1164/rccm.201409-1583OC
13. Zhang Y, Dong S, Gao Q, Liu S, Ganjurjav H, Wang X, Su X, Wu X (2017) Soil bacterial and fungal diversity differently correlated with soil biochemistry in alpine grassland ecosystems in response to environmental changes. Sci Rep 7. https://doi.org/10.1038/srep43077
14. Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11:2639–2643. https://doi.org/10.1038/ismej.2017.119
15. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML (2013) Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 4:1111–1119. https://doi.org/10.1111/2041-210X.12114
16. Martiny AC, Treseder K, Pusch G (2013) Phylogenetic conservatism of functional traits in microorganisms. ISME J 7:830–838. https://doi.org/10.1038/ismej.2012.160
17. Schmidt VT, Reveillaud J, Zettler E, Mincer TJ, Murphy L, Amaral-Zettler LA (2014) Oligotyping reveals community level habitat selection within the genus Vibrio. Front Microbiol 5. https://doi.org/10.3389/fmicb.2014.00563
18. Glassman SI, Martiny JBH (2018) Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. mSphere 3:e00148–e00118. https://doi.org/10.1128/mSphere.00148-18
19. Tapolczai K, Selmeczy GB, Szabó B, B-Béres V, Keck F, Bouchez A, Rimet F, Padisák J (2021) The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. Ecol Indic 122:107322. https://doi.org/10.1016/j.ecolind.2020.107322
20. Porter TM, Hajibabaei M (2020) Putting COI metabarcoding in context: the utility of exact sequence variants (ESVs) in biodiversity analysis. Front Ecol Evol 8. https://doi.org/10.3389/fevo.2020.00248
21. Song Z, Schlatter D, Gohl DM, Kinkel LL (2018) Run-to-run sequencing variation can introduce taxon-specific bias in the evaluation of fungal microbiomes. Phytobiomes J 2:165–170. https://doi.org/10.1094/PBIOMES-09-17-0041-R
22. Cobian GM, Egan CP, Amend AS (2019) Plant–microbe specificity varies as a function of elevation. ISME J 13:2778–2788. https://doi.org/10.1038/s41396-019-0470-4
23. Darcy JL, Swift SOI, Cobian GM, Zahn GL, Perry BA, Amend AS (2020) Fungal communities living within leaves of native Hawaiian dicots are structured by landscape-scale variables as well as by host plants. Mol Ecol 29:3102–3115. https://doi.org/10.1111/mec.15544
24. Wainwright BJ, Zahn GL, Spalding HL, Sherwood AR, Smith CM, Amend AS (2017) Fungi associated with mesophotic macroalgae from the 'Au'au Channel, west Maui are differentiated by host and overlap terrestrial communities. PeerJ 5:e3532. https://doi.org/10.7717/peerj.3532
25. Tipton L, Zahn G, Datlof E, Kivlin SN, Sheridan P, Amend AS, Hynson NA (2019) Fungal aerobiota are not affected by time nor environment over a 13-y time series at the Mauna Loa Observatory. PNAS 116:25728–25733. https://doi.org/10.1073/pnas.1907414116
26. Peay KG, von Sperber C, Cardarelli E, Toju H, Francis CA, Chadwick OA, Vitousek PM (2017) Convergence and contrast in the community structure of bacteria, fungi and archaea along a tropical elevation–climate gradient. FEMS Microbiol Ecol 93. https://doi.org/10.1093/femsec/fix045
27. Vannette RL, Leopold DR, Fukami T (2016) Forest area and connectivity influence root-associated fungal communities in a fragmented landscape. Ecology 97:2374–2383. https://doi.org/10.1002/ecy.1472
28. US Geological Survey Hawaiian Volcano Observatory scientists and colleagues (2003) Once a big island, Maui Count now four small islands. Volcano Watch
29. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, de Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend AS, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V, Nilsson RH (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol n/a-n/a. https://doi.org/10.1111/2041-210X.12073
30. Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

31. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461
32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303
33. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res 47:D259–D264. https://doi.org/10.1093/nar/gky1022
34. Oksanen J, Blanchet FG, Friendly M, et al (2016) vegan: community ecology package
35. Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, Ellison AM (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol Monogr 84:45–67. https://doi.org/10.1890/13-0133.1
36. Preston FW (1948) The commonness, and rarity, of species. Ecology 29:254–283. https://doi.org/10.2307/1930989
37. Prado PI, Miranda MD, Chalom A (2016) sads: maximum likelihood models for species abundance distributions
38. Dormann CF (2011) How to be a specialist? Quantifying specialisation in pollination networks. Network Biology
39. Dormann CF, Gruber B, Fründ J (2008) Introducing the bipartite Package: analysing Ecological Networks. 8:4
40. Blüthgen N, Menzel F, Blüthgen N (2006) Measuring specialization in species interaction networks. BMC Ecol 6:9. https://doi.org/10.1186/1472-6785-6-9
41. (2018) Epicoccum hordei CGMCC 3.18360 ITS region; from TYPE material
42. Davison J, Moora M, Öpik M et al (2015) FUNGAL SYMBIONTS. Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. Science 349:970–973. https://doi.org/10.1126/science.aab1161
43. Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson KH, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo LD, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, de Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K (2014) Global diversity and geography of soil fungi. Science 346:1256688. https://doi.org/10.1126/science.1256688
44. Carthey AJR, Blumstein DT, Gallagher RV, Tetu SG, Gillings MR (2020) Conserving the holobiont. Funct Ecol 34:764–776. https://doi.org/10.1111/1365-2435.13504
45. Eren AM, Borisy GG, Huse SM, Welch JLM (2014) Oligotyping analysis of the human oral microbiome. PNAS 111:E2875–E2884. https://doi.org/10.1073/pnas.1409644111
46. Bidartondo MI (2008) Preserving accuracy in GenBank. Science 319:1616. https://doi.org/10.1126/science.319.5870.1616a
47. Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG (2016) FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecol 20:241–248. https://doi.org/10.1016/j.funeco.2015.06.006
48. Sinclair L, Ijaz UZ, Jensen LJ, Coolen MJL, Gubry-Rangin C, Chroňáková A, Oulas A, Pavloudi C, Schnetzer J, Weimann A, Ijaz A, Eiler A, Quince C, Pafilis E (2016) Seqenv: linking sequences to environments through text mining. PeerJ 4:e2690. https://doi.org/10.7717/peerj.2690
49. Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? Mol Ecol 19:5555–5565. https://doi.org/10.1111/j.1365-294X.2010.04898.x
50. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol 10:189. https://doi.org/10.1186/1471-2180-10-189