

# Uncovering unseen fungal diversity from plant DNA banks

Erin M. Datlof<sup>f</sup>, Anthony S. Amend<sup>1</sup>, Kamala Earl<sup>2</sup>, Jeremy Hayward<sup>1,†</sup>, Clifford W. Morden<sup>1</sup>, Rachael Wade<sup>1</sup>, Geoffrey Zahn<sup>1</sup> and Nicole A. Hynson<sup>1</sup>

<sup>1</sup> Department of Botany, University of Hawai'i at Mānoa, Honolulu, HI, United States of America

<sup>2</sup> School of Forest Resources and Conservation, University of Florida, Gainesville, FL, United States of America

<sup>†</sup> Deceased.

## ABSTRACT

Throughout the world DNA banks are used as storage repositories for genetic diversity of organisms ranging from plants to insects to mammals. Designed to preserve the genetic information for organisms of interest, these banks also indirectly preserve organisms' associated microbiomes, including fungi associated with plant tissues. Studies of fungal biodiversity lag far behind those of macroorganisms, such as plants, and estimates of global fungal richness are still widely debated. Utilizing previously collected specimens to study patterns of fungal diversity could significantly increase our understanding of overall patterns of biodiversity from snapshots in time. Here, we investigated the fungi inhabiting the phylloplane among species of the endemic Hawaiian plant genus, *Clermontia* (Campanulaceae). Utilizing next generation DNA amplicon sequencing, we uncovered approximately 1,780 fungal operational taxonomic units from just 20 DNA bank samples collected throughout the main Hawaiian Islands. Using these historical samples, we tested the macroecological pattern of decreasing community similarity with decreasing geographic proximity. We found a significant distance decay pattern among *Clermontia* associated fungal communities. This study provides the first insights into elucidating patterns of microbial diversity through the use of DNA bank repository samples.

Submitted 6 May 2017  
Accepted 3 August 2017  
Published 28 August 2017

Corresponding author  
Nicole A. Hynson,  
nhynson@hawaii.edu

Academic editor  
Antonio Amorim

Additional Information and  
Declarations can be found on  
page 13

DOI 10.7717/peerj.3730

© Copyright  
2017 Datlof et al.

Distributed under  
Creative Commons CC-BY 4.0

## OPEN ACCESS

**Subjects** Biodiversity, Biogeography, Microbiology, Mycology

**Keywords** Microbial diversity, DNA banks, Fungi, Biodiversity, Biobanks

## INTRODUCTION

Understanding biodiversity is an important goal of biology. This is particularly critical in a changing world with habitat degradation and fragmentation, population declines, and species extinctions (*Vitousek et al., 1997*). Once a species becomes extinct, the genetic history resulting from its evolution is lost as well (*Mattick, Ablett & Edmonson, 1992*). DNA banks were initially developed to collect genetic material to create a storage base for evolutionary history, biological diversity, and genomic information (*Mattick, Ablett & Edmonson, 1992*). Throughout the world, samples are collected and stored in these banks to document and preserve genetic diversity (*Spooner & Ruess, 2014*). For extinct species, DNA bank samples act as storage deposits for their genomes (*Adams, 1994; Spooner & Ruess, 2014*).

In addition to the importance of DNA bank repositories for archiving target organisms' genetic information, these samples also harbor the microbial diversity associated with each accession. These samples represent well-preserved DNA at snapshots in time and space. For example, plant DNA bank samples not only preserve the genomic information of the target species, but also preserve potentially important cryptic microbial symbionts associated with the host, such as fungi known to inhabit the plant phyllosphere ([Porrás-Alfaro & Bayman, 2011](#); [Vorholt, 2012](#)).

Despite much work on patterns of plant diversity, comparatively little is known about the diversity of fungi. Fungi play crucial functions in ecosystems by acting as decomposers and nutrient cyclers, important mutualists such as mycorrhizae, and pathogens influencing host species populations ([Kendrick, 2001](#); [Lips et al., 2006](#)). Globally, <100,000 species of fungi have been described ([Blackwell, 2011](#)), which is far less than total estimated fungal diversity, and also less than vascular plants, with <400,000 species currently described ([Royal Botanic Gardens Kew, 2016](#)). Estimates of global fungal species richness have increased almost 3–5 fold in the past 20 years, from 1.5 million ([Hawksworth, 1991](#)) to 3.5–6 million species ([O'Brien et al., 2005](#); [Taylor et al., 2014](#)). This increase is due in part to advances in direct environmental DNA sequencing and extrapolations based on predictions of vascular plant to fungal ratios ([O'Brien et al., 2005](#); [Taylor et al., 2014](#)). In order to obtain more accurate estimates of true fungal diversity, increased sampling using high throughput DNA sequencing of many different types of environments is needed, and DNA banks may significantly contribute to filling this knowledge gap.

Hawai'i is a biodiversity hotspot, making it an exceptional location to study patterns of species diversity ([Myers et al., 2000](#)). However, we know very little about Hawaiian fungi, their potential rates of endemism, and patterns of biodiversity. A survey of mushrooms throughout the Hawaiian Islands conducted in the 1990's found 310 species. The majority of these taxa were introduced, however 52 were putatively native and 46 of these taxa were considered potentially endemic (~88%; [Hemmes & Desjardin, 2002](#)). Similar rates of endemism are found in the Hawaiian flora.

An estimated 89% of the Hawaiian vascular plant flora is endemic ([Wagner, Herbst & Sohmer, 1999](#)). The unique Hawaiian flora is threatened by habitat degradation and loss, coupled with species invasions, which have led to native species becoming endangered or extinct ([Morden, Caraway & Motley, 1996](#)). There are currently 1,175 recognized native (endemic plus indigenous) Angiosperm species in Hawai'i ([Smithsonian Institution, 2017](#)) and 422 of these plants are currently endangered (35.9%; [US Fish and Wildlife Service, 2015](#)) with 104 taxa extinct or possibly extinct (8.8%; [Sakai, Wagner & Mehrhoff, 2002](#)). As a result of these extinctions and a strong potential for additional future losses, the Hawaiian Plant DNA Library (HPDL) was created to preserve the genetic diversity of the Hawaiian flora ([Morden, Caraway & Motley, 1996](#)).

The HPDL was started in 1992 and currently has over 10,000 accessions representing over 86% of all of the Hawaiian plant genera and over 50% of all species ([Morden, 2017](#)). Similar to other DNA banks across the world, collections for common species contain multiple individuals with their own accession number. However, the HPDL is unique due to the relatively disproportionate number of threatened and endangered native Hawaiian

plants compared to other localities, and thus serves as a genetic repository for many species that are endangered or extinct in the wild (Morden, Caraway & Motley, 1996). The main goal of the HPDL is to preserve Hawaiian plant DNA and bank samples for use in future studies of biodiversity (Morden, Caraway & Motley, 1996; Randell & Morden, 1999).

All naturally occurring plant tissues harbor fungi as both endophytes, living in between plant cells (Rodriguez *et al.*, 2009) and epiphytes living on plant surfaces (Santamaría & Bayman, 2005), collectively known as phyllosphere fungi (Vacher *et al.*, 2016). These communities form diverse assemblages with some studies showing an average of about 100 species per individual tree and ranges of about 700–4,000 species of fungi per host species (Jumpponen & Jones, 2009; Zimmerman & Vitousek, 2012). Thus, along with the HPDL's banked plant samples it also has likely and coincidentally preserved a substantial portion of the diversity of Hawaiian fungi, acting as a repository for not only plant genetics, but their microbiomes as well.

Utilizing several plant bank samples from the HPDL we investigated the diversity of unintentionally co-sampled fungi found within banked plant samples. Phyllosphere fungal communities can be cryptic and hyperdiverse (Jumpponen & Jones, 2009; Zimmerman & Vitousek, 2012) and this poses a challenge to studies with the goal of assessing microbial diversity. Even with the adoption of next generation DNA sequencing technologies, such as high throughput amplicon sequencing from environmental samples, observed diversity is often an under representation of true diversity (Chiu & Chao, 2016). Thus, rather than superficially sequencing the fungi from all available replicate DNA extracts from a single plant species, we chose to deeply sequence ten samples of a common endemic Hawaiian plant genus, *Clermontia* (Campanulaceae), with species found across the Hawaiian Islands (Givnish *et al.*, 2009). This deep sequencing was done in hopes that the vast majority of fungi from our samples would be recovered from each sample.

In this study, we utilized historical DNA bank samples to validate the use of plant bank samples as a resource for elucidating phyllosphere fungal biodiversity, while subsequently examining plant-associated fungal diversity across space. Our two main questions were do DNA bank samples store microbial diversity? And can these previously collected samples be used to uncover ecological patterns, such as changes in microbial community similarity over space? Using DNA samples of eight species of *Clermontia* stored in the HPDL, we sequenced the fungi found in these plants' phylloplanes (the portion of the phyllosphere made up by leaves). We took advantage of the archipelago's geographic spatial gradient and the previously collected samples in the HPDL to test for decreases in community similarity as the distance between communities increases, the classical ecological pattern of distance-decay of community similarity (Nekola & White, 1999).

## MATERIALS & METHODS

### Samples

We examined foliar DNA extracts stored in the HPDL from the genus *Clermontia*, representing eight species collected from Hawai'i Island, O'ahu, Maui, Moloka'i and

**Table 1** DNA bank samples and accession numbers from the HPDL for each extract along with associated metadata. In this study, two individual extracts were pooled for each location and given a sample code labeled by island.

Sample	Sample code	HPDL number	<i>Clermontia</i> species	Island	Date extracted	Latitude	Longitude
1	M1	6,843	<i>kakeana</i>	Moloka'i	7/14/11	21.13	−156.92
1a		6,844					
2	H1	6,961	<i>calophylla</i>	Hawai'i	11/18/11	20.09	−155.74
2a		6,962					
3	H2	6,888	<i>kohalae</i>	Hawai'i	9/17/11	20.08	−155.74
3a		6,889					
4	H3	6,856	<i>clermoniotides</i>	Hawai'i	8/3/11	19.21	−155.60
4a		6,857					
5	H4	7,339	<i>peleana</i> ssp. <i>singulariflora</i>	Hawai'i	6/7/13	20.18	−155.80
5a		7,940					
6	K1	5,089	<i>fauriei</i>	Kaua'i	9/6/05	22.09	−159.59
6a		5,090					
7	O1	6,809	<i>kakeana</i>	O'ahu	6/17/11	21.34	−157.82
7a		6,810					
8	O2	7,008	<i>oblongifolia</i> ssp. <i>oblongifolia</i>	O'ahu	3/15/12	21.41	−158.10
8a		7,009					
9	Ma1	6,875	<i>arborescens</i>	Maui	7/5/11	20.82	−156.28
9a		6,876					
10	Ma2	6,831	<i>kakeana</i>	Maui	7/14/11	20.80	−156.23
10a		6,832					

Kaua'i (Table 1; see Table S1 for more details). When initial collections were made, young leaf samples were collected in the field, sealed in bags, and stored at 4 °C until DNA extraction (less than a week). Leaves were not disturbed by rinsing prior to DNA extraction. Approximately 1.0 g of leaf tissue was extracted using a modified CTAB method with cesium chloride banding optimized to recover high quality DNA intended for long-term storage, and stored at −20 °C (Doyle & Doyle, 1987; Morden, Caraway & Motley, 1996). Samples used in this study were in a −20 °C freezer from 2–10 years. Two individual plant DNA extracts of each species per location were equally pooled, yielding a total of ten samples from 20 *Clermontia* individuals ( $n = 10$ ), with *C. kakeana* replicates on three different islands: O'ahu, Moloka'i, and Maui.

### PCR and sequencing

These pooled extracts were individually prepared for fungal DNA sequencing with slight modifications to the Illumina 16S Metagenomic Sequencing Library Preparation protocol using a two-step PCR and index attachment (Illumina, 2013). Fungal DNA amplicons of the ~250–400-bp targeted nuclear ribosomal Internal Transcribed Spacer 1 (ITS1) locus were amplified using ITS1F primers with Illumina adapter overhangs (5' Adapter-CTTGGTCATTTAGAGGAAGTAA-3'; Gardes & Bruns, 1993) and modified ITS2 primers (5' Adapter-GCTGCGTTCTTCATCGATGC-3'; White et al., 1990). The ITS locus is the official fungal DNA barcode (Schoch et al., 2012). Amplicons were purified and size-selected

using SPRIselect beads (Beckman Coulter, Inc., Brea, CA, USA), followed by a second PCR attaching forward and reverse eight-base pair barcoded Illumina overhang adapters (i7 and i5; *Illumina, 2013*). See [Table S2](#) for PCR recipes and thermalcycler parameters. These indexed libraries were bead purified and quantified using the Qubit dsDNA HS kit (Life Technologies Inc. Gaithersburg, MD, USA). Libraries were then pooled at equimolar concentrations and sent to the Hawai'i Institute for Marine Biology Genetics Core Facility (HIMB) for quality control on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and sequencing on the Illumina MiSeq platform v.3 paired-end  $2 \times 300$  (Illumina, San Diego, CA, USA).

## Bioinformatics

De-multiplexed fastq files were obtained from the sequencing facility from the ten *Clermontia* plant bank samples. Raw sequencing data was deposited to the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under BioProject [PRJNA379349](#). These paired-end reads were merged with the Illumina Paired-End reAd mergeR (PEAR), keeping reads with a minimum assembly length of 250-bp, average quality threshold of 15 and above, and discarding all reads with any uncalled bases (*Zhang et al., 2014*). Further quality control was carried out using the FASTX-Toolkit, using the `fastq_quality_filter` command (*Hannon Lab, 2016*), where all reads with any base pairs containing a quality score below 15 were discarded (*Hannon Lab, 2016*). Potential chimeras were removed in `vsearch` (*Rognes et al., 2016*) using the `uchime_ref` command (*Edgar et al., 2011*), which referenced the User-friendly Nordic ITS Ectomycorrhiza (UNITE) database, accessed on 11.03.2015 (*Köljalg et al., 2013*). Operational Taxonomic Units (OTUs) were clustered within QIIME (*Caporaso et al., 2010*) using the open-reference method (*Navas-Molina et al., 2013*) following the Usearch algorithm (*Edgar, 2010*). Briefly, reads were matched to reference OTUs in the UNITE dynamic database (ver7) (*Köljalg et al., 2013*) with added *Clermontia* outgroups, then remaining reads that failed to match were subsampled as seeds for three subsequent rounds of *de novo* OTU-picking. The most abundant sequence for each OTU was chosen as a representative sequence. Singleton reads were removed in QIIME prior to OTU table generation and taxonomy was assigned against the UNITE database with the Basic Local Alignment (BLAST) algorithm.

## Statistics

All statistical analyses were conducted in R version 3.3.0 (*R Core Team, 2017*). The OTU table from QIIME was imported into R with the package *biomformat* (*McMurdie & Paulson, 2016*). OTUs that mapped to plant taxonomies or those that had no BLAST hit were removed from the OTU matrix and all OTUs with greater than ten reads were kept for analyses. Samples were rarefied to 16,546 reads, the minimum sample depth. Rarefaction, species accumulation curves were generated using the *vegan* package for: all samples, individual samples, and samples pooled by island (*Oksanen et al., 2017*). Because observed species richness often under estimates true species richness (*Hughes et al., 2001*), asymptotic extrapolations of species richness and diversity for all samples and species were estimated based on the first three Hill numbers using the *iNEXT* package for raw incidence

data (Hsieh, Ma & Chao, 2016). These are namely species richness, the exponential of Shannon entropy, and the inverse Simpson concentration, represented by  $q = 0, 1, 2$ , respectively (Chao et al., 2014). Hill numbers and extrapolations were generated based on individual samples and individual species. A Venn diagram was generated to visualize overlapping taxa between islands using the *VennDiagram* package (Chen, 2016).

### Distance matrices

To investigate ecological patterns, we accounted for variables that may be influencing the fungal communities found in these banked samples. These factors were temporal and physical distances between sample collections, as well as fungal community dissimilarity. Pairwise distance matrices were calculated for physical distance in kilometers using the *geosphere* package (Hijmans, 2016), time between sample collections in days, and Bray-Curtis community dissimilarity using the *vegan* package (Oksanen et al., 2017). Separate Mantel tests for each combination of the following pairwise distance matrices: time between sample collections (days), as well as Euclidean physical distance between samples (km), and community dissimilarity (Bray-Curtis), were run for 10,000 permutations. To investigate the effects of these variables a final partial Mantel test for physical distance and community dissimilarity, while controlling for time, was run for 10,000 permutations (Oksanen et al., 2017).

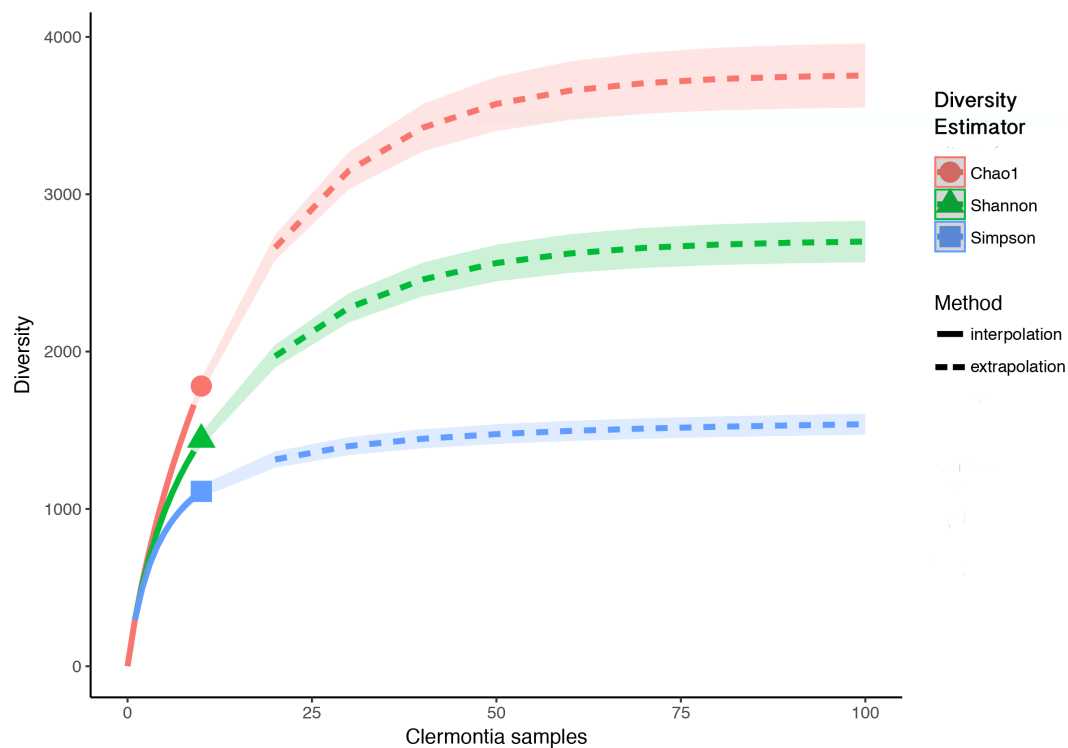
## RESULTS

### Sequencing

A total of 4,312,473 sequence reads were obtained from the plant DNA library samples. Of these, 3,571,252 paired-end reads (82.8%) were successfully assembled and further quality control removed low quality reads, keeping 2,680,945 reads (75.1%). After referencing UNITE, 121,618 (4.5%) chimeric sequences were removed, leaving a total of 2,559,327 high-quality reads. Taxonomic assignment yielded a total of 1,648,971 fungal reads that were binned into 2,944 fungal OTUs for use in downstream analyses.

### Fungal diversity and host associations

Each *Clermontia* DNA bank sample used in this study contained fungal DNA. In total, we found 2,944 fungal OTUs associated with the ten foliar DNA bank samples. After removing OTUs with less than ten reads and rarefying to the sample with the minimum number of reads, we removed 1,164 OTUs (39.5%) and were left with a total of 1,780 fungal OTUs. While the observed OTU accumulation curve for all ten samples combined did not reach an asymptote (Fig. S1), OTU accumulation curves by sample and by island (except in the case of *C. fauriei* from Kaua'i and *C. kakeana* Moloka'i) generally reached their asymptotes, indicating that overall, we successfully recovered the majority of fungi present in our samples (Figs. S2 and S3). Observed richness per sample after rarefying, ranged from 108 to 682 fungal OTUs with an average of 295 OTUs per sample ( $\pm 54.69$  standard error). From our ten samples combined the *iNEXT* extrapolation curves suggest fungal richness based on the Hill number  $q = 0$  (Chao1 richness) will saturate around 3,947 OTUs which would require at least 50 samples. Similarly fungal diversity based on  $q = 1$  (exponential

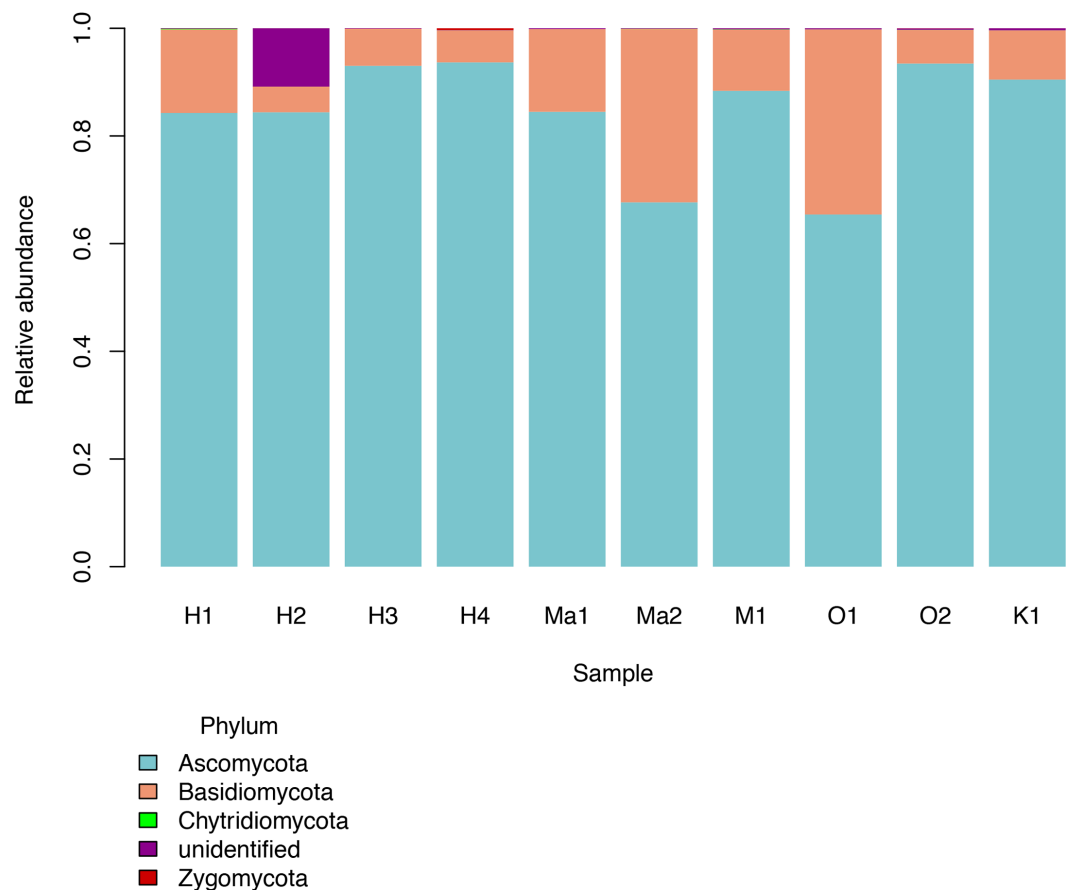


**Figure 1** Sample interpolation (solid lines) and extrapolation (dashed lines) curves for phylloplane fungi from all ten *Clermontia* DNA bank samples. Based Hill's numbers three different diversity estimators were used (Chao1 richness, exponential of Shannon entropy, and inverse Simpson concentration indices) and are shown by the different colors with 95% confidence intervals shown by shading. Shapes represent observed phylloplane fungal OTU diversity for the ten samples of *Clermontia* spp.

Shannon entropy) was estimated to saturate at around 2,750, and diversity based on  $q = 2$  (inverse Simpson concentration) was estimated to saturate at about 1,591 (Fig. 1).

We investigated patterns of fungal diversity at the phyla and ordinal levels. Overall, the majority of fungi in the subkingdom Dikarya dominated all of the phylloplane samples, with phylum Ascomycota being most abundant followed by Basidiomycota (Fig. 2). Fungi belonging to the phylum Chytridiomycota and Zygomycota were also present in lower abundances. Additional OTUs mapped to kingdom Fungi but could not be identified further (Unidentified; Fig. 2). The top ten most abundant orders were Capnodiales, Chaetothyriales, Exobasidiales, Peltigerales, Pertusariales, Pleosporales, Tremellales, Ustilaginales, and two unknown orders (Fig. 3).

Based on our observed data, total average fungal OTU richness by island was 507.6 ( $\pm 128.46$  standard error). O'ahu had the highest total observed richness with 1,045 OTUs, followed successively by Hawai'i (926 OTUs), Maui (685 OTUs), Moloka'i (362 OTUs), and Kaua'i again had the lowest richness (108 OTUs), though unlike the other islands we did not saturate our OTU richness by sequencing effort curve for *C. fauriei* from Kaua'i and *C. kakeana* from Moloka'i indicating that these are an underrepresentation of fungal OTU richness (Fig. S3). Overall about twenty OTUs were found on all of the five islands (Fig. 4).



**Figure 2** Relative abundances of fungal phyla for each *Clermontia* DNA bank sample.

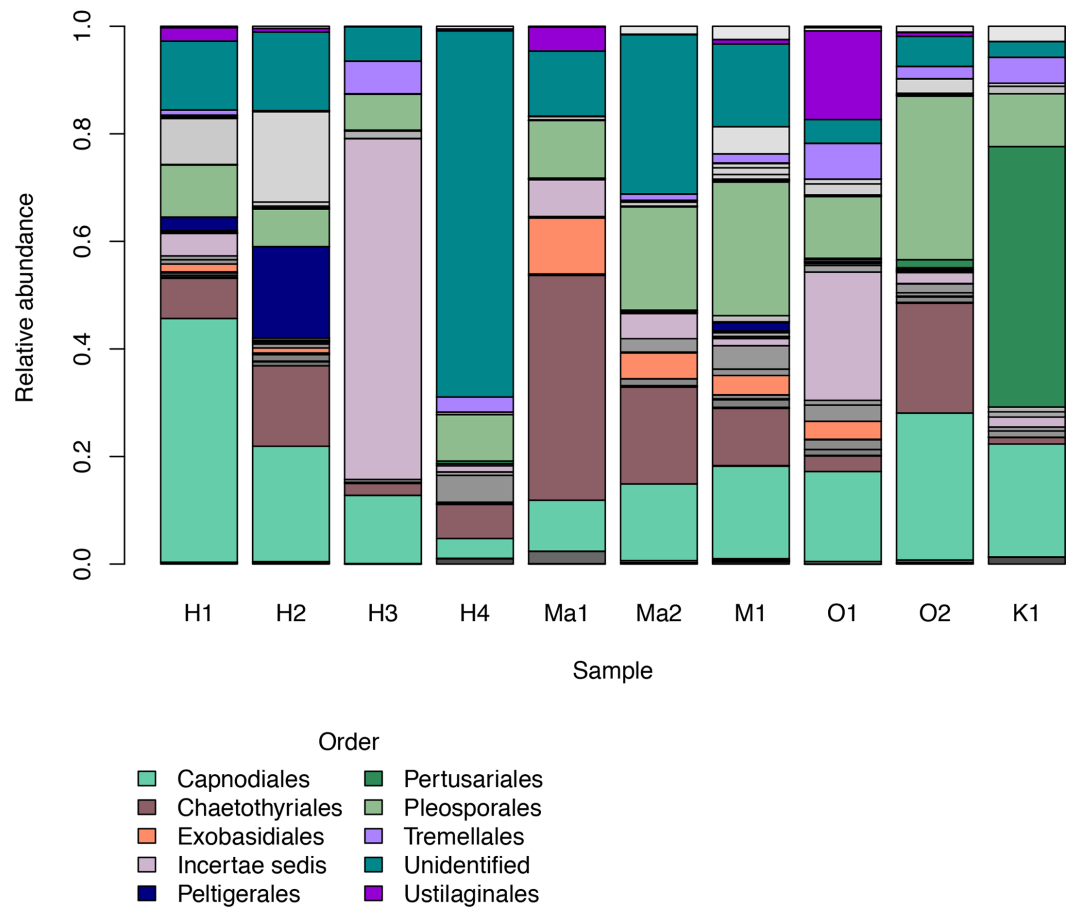
### Physical distance decay

*Clermontia* DNA bank extracts used in this study spanned across the main Hawaiian Islands. The nearest samples were collected less than one kilometer apart from a single site in Kohala, Hawai'i Island, and the furthest distance was 524.78 km from Kohala, Hawai'i Island to the Alaka'i Swamp, Kaua'i. Over this spatial range, while taking into account time (number of days) between sample collections, the fungal phylloplane communities exhibit a significant decrease in community similarity across increasing geographic distance (Fig. 5, Partial Mantel test:  $r = 0.423$ ,  $p = 0.005$ ). Time was also significantly correlated with physical distance between sample sites (Fig. S4, Mantel test:  $r = 0.455$ ,  $p = 0.048$ ). Time between sampling and community dissimilarity was marginally significantly correlated (Fig. S5, Mantel test:  $r = 0.619$ ,  $p = 0.051$ ). Despite the significant relationships with sampling time the Partial Mantel between distance and community dissimilarity while accounting for time, was significant (Fig. 5).

## DISCUSSION AND CONCLUSION

In this study we investigated the diversity of phylloplane fungi associated with *Clermontia* spp. that were collected across the Hawaiian Islands and stored as DNA bank samples.

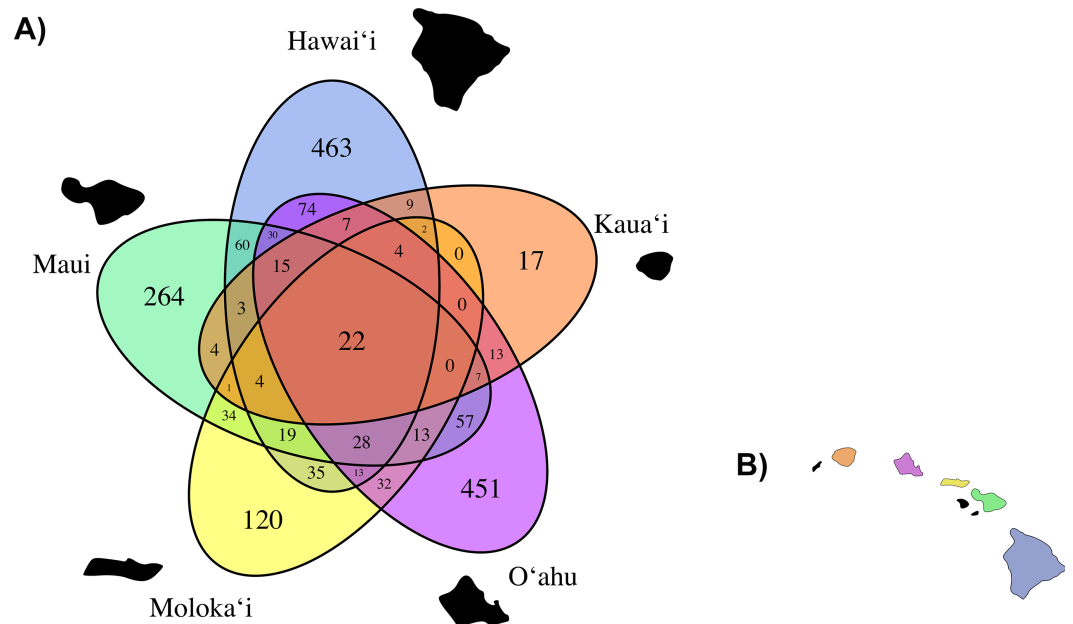




**Figure 3** Relative abundances of the top ten most abundant fungal orders for each *Clermontia* DNA bank sample represented by color bars. The less abundant orders are represented by grayscale bars.

While sample collection methods were not developed with the intention of preserving or analyzing phyllosphere fungi, they coincide with common practices for these purposes (Jumpponen & Jones, 2009; Zimmerman & Vitousek, 2012).

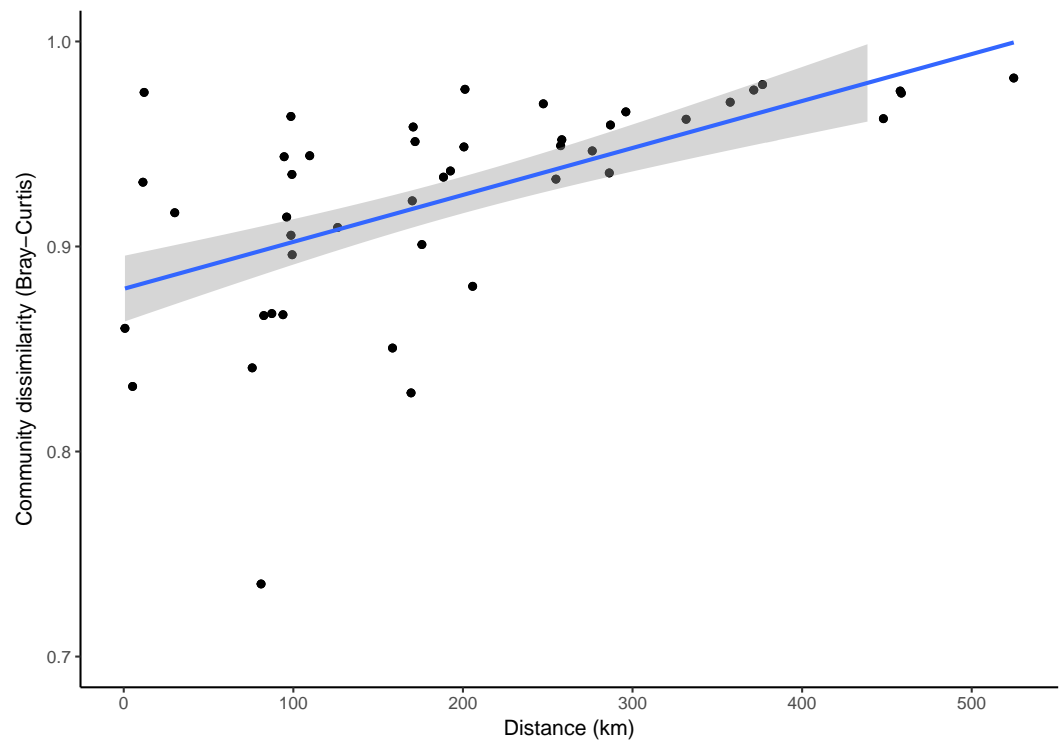
We found that these specimens harbored a considerable diversity of fungi. After quality control, we found 1,780 fungal OTUs from just ten samples, representing 20 *Clermontia* individuals and eight species. Fungal richness ranged from 108 to 682 OTUs per plant sample. Incredibly, this diversity was recovered from a total of just 20 grams of leaf tissue from which DNA was extracted and preserved. Despite our high sequencing depth, the observed species accumulation curves for all samples and islands did not saturate, indicating our sequencing efforts likely underestimated true *Clermontia* phylloplane fungal diversity. However, this novel use of DNA bank samples revealed substantial undiscovered fungal biodiversity stored in plant samples. These results provide further evidence of microbes making up the “unseen majority” of biodiversity (Whitman, Coleman & Wiebe, 1998), where a single macroorganism associates with a multitude of microorganisms both within and on their surfaces (Turner, James & Poole, 2013).



**Figure 4** (A) A Venn Diagram displaying the number of overlapping fungal OTUs shared between *Clermontia* samples from each of the five main Hawaiian Islands, represented by a different color. The number of OTUs unique to each island lie on the outermost portion of each ellipse. (B) The geographic layout of the main Hawaiian Islands.

This study highlights a new and underutilized function of biological collections, as well as gives insights into regional fungal diversity patterns. Previous estimates of total regional fungal richness have been based off of plant to fungi ratios ranging from 1:6 (*Hawksworth, 1991*) to 1:17 (*Taylor et al., 2014*). Our data supplement these studies using environmental NGS data. If we assume that the diversity of phylloplane fungi associated with *Clermontia* species is representative of the native Hawaiian flora, we would estimate based on Chao1 richness ( $q = 0$ ) extrapolations (determined by species; *Fig. S6*) that the entire Hawaiian flora (c. 1,000 species) harbors about 4,000 fungi. This results in an approximate 1:4 plant to fungi species ratio. However, just considering phylloplane fungi associated with a single genus is likely an underestimate of total fungal biodiversity due to potential host-fungi specificity (*Hoffman & Arnold, 2008*). Supplementary to host specificity, only taking into consideration phylloplane fungi likely underestimates total regional fungal richness due to habitat niche partitioning among fungal species and guilds (*Hibbett, Gilbert & Donoghue, 2000*).

In addition to the study of microbial diversity, questions regarding microbial biogeography, host specificity, and the effects of global change on microbial communities could be addressed with DNA banks. For example, we were able to confirm the distance decay of microbial community similarity from DNA bank samples collected across the Hawaiian Islands. This finding is similar to other microbial systems where significant distance decay patterns were found in foliar endophytic (*Vaz et al., 2014*) and ectomyorrhizal fungal communities (*Bahram et al., 2013*), as well as bacteria and archaea (*Barreto et al., 2014*). We were also able to identify a temporal partitioning of phylloplane



**Figure 5** Pair-wise Bray-Curtis fungal community dissimilarity plotted against corresponding pair-wise physical distances for each *Clermontia* plant bank sample spanning the main Hawaiian Islands. A regression line was fit to the data, shown in blue, with 95% confidence intervals shown in grey. (Partial Mantel test:  $r = 0.424$ ,  $p = 0.005$ , accounting for time between sample collection in days).

fungal communities (Fig. S5), indicating that DNA bank collections may be used to study changes in microbial communities over time. However, in addition to geography and temporal factors, taking into account host genotype, age, specificity, and differences in environment, such as light exposure and humidity, may potentially explain additional variation in fungal communities (Hoffman & Arnold, 2008).

In agreement with other phyllosphere studies, the majority of fungal taxa were identified as belonging to the subkingdom Dikarya, with the majority in phylum Ascomycota followed by Basidiomycota (Rodriguez et al., 2009). It is not surprising that we found so many unknown fungal taxa (45.16% of total OTUs at the family level) including 28 OTUs we were unable to place at the phylum level. The plant samples from this study represent an endemic Hawaiian genus whose microbial associates are previously unstudied, and possibly associate with undescribed fungi endemic to Hawai'i. However, this degree of unassigned fungal OTUs is not unique to our system and highlights our limited current knowledge of fungal diversity (Nilsson et al., 2016). For example, recent discoveries using environmental DNA sequencing have reshaped the fungal tree of life by uncovering a new fungal Phylum, the Cryptomycota (Jones et al., 2011). This stresses the need for further investigations of fungal biodiversity, their cryptic nature and diverse functions make for intriguing new discoveries that have the potential to change evolutionary and ecological theories based primarily on macroorganisms.

With the recent advent of next generation sequencing (NGS) techniques genomic investigations of non-model organisms have become readily accessible ([Da Fonseca et al., 2016](#)). However, there are important caveats to consider when using these methods and analyses. For example, working with environmental samples poses the challenging prospect of encountering hyperdiverse microbial communities such as the fungi found here, and other studies of plant phyllosphere fungi ([Arnold, 2007](#); [Arnold & Lutzoni, 2007](#)). As seen in this NGS study, thousands of fungi can be associated with a small number of plant leaf samples. While uncovering this diversity is a goal of some microbial ecologists, for researchers using NGS techniques focused on the host organism (in this case plants), microbial symbionts may interfere with downstream analyses and results. Microbial taxa associated with macroorganisms should be taken into account when using NGS methods such as RAD seq, RNA seq, targeted sequencing, among other techniques ([Da Fonseca et al., 2016](#)). Additionally, future research into the microbiomes of hosts preserved in DNA banks should take into consideration how sample processing and storage may affect microbes. When initially collecting samples for these purposes, care should be taken to avoid microbial contamination from non-target hosts or environments.

Most DNA bank samples likely harbor unintended microbial communities associated with each target individual from a specific location at distinct snapshots in time. While DNA banks are a common genetic biodiversity repository ([Seberg et al., 2016](#)), to the best of our knowledge this is the first study where they were used to investigate genetic material other than that of the target organism. By using these archived samples we were able to rapidly recover previously undocumented microbial diversity. The abundance of DNA bank samples stored throughout the world represent a large proportion of the globe's extant and extinct biological diversity. This storage provides the opportunity for microbes associated with these organisms to be easily investigated without the associated costs of sample collection. This may be important for conservation efforts, giving insight into potentially important symbionts ([Van der Heijden, Bardgett & Straalen, 2008](#); [Busby, Ridout & Newcombe, 2016](#)). For those species that go extinct, their genomes are preserved in DNA banks along with their corresponding microbial symbionts. For extant organisms, DNA bank samples could be used to better understand the ecology of symbiosis and possibly identify coevolutionary patterns. Overall, this study highlights the potential use of DNA bank samples for the study of global biodiversity. This study also demonstrates the benefits of in-depth sample sequencing to uncover the majority of fungal diversity found in each plant bank sample. With DNA bank samples stored throughout the world, already collected, processed, and extracted, they harbor the potential for new and exciting investigations.

## ACKNOWLEDGEMENTS

We thank Emily Johnston and Richard O'Rorke for experimental and lab help, and Leah Tooman for further lab assistance. Also thanks to Sean Swift, Laura Tipton, and Sofia Gomes for R help. Thanks to Mitsuko Yorkston and Richard Pender for sample discussions. We would also like to thank Ricardo Araujo and two additional anonymous reviewers for their helpful suggestions on a previous version of this manuscript.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by a donation from the Illumina Corporation and National Science Foundation (No. 1255972) to Anthony S. Amend. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
Illumina Corporation and National Science Foundation: 1255972.

### Competing Interests

Anthony S. Amend is an Academic Editor for PeerJ.

### Author Contributions

- Erin M. Datlof conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Anthony S. Amend conceived and designed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Kamala Earl and Rachael Wade conceived and designed the experiments, performed the experiments, reviewed drafts of the paper.
- Jeremy Hayward conceived and designed the experiments, analyzed the data, prepared figures and/or tables.
- Clifford W. Morden reviewed drafts of the paper and provided DNA bank samples.
- Geoffrey Zahn analyzed the data, prepared figures and/or tables, reviewed drafts of the paper.
- Nicole A. Hynson assisted with data analysis, wrote the paper, reviewed drafts of the paper.

### Data Availability

The following information was supplied regarding data availability:

The raw sequence reads are deposited in NCBI Sequence Read Archive as BioProject [PRJNA379349](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA379349).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.3730#supplemental-information>.

## REFERENCES

- Adams RP. 1994.** DNA bank-net—an overview. In: Adams RP, Miller JS, Golderberg EM, Adams JE, eds. *Conservation of plant genes II: utilization of ancient and modern DNA*. St. Louis: Monographs in Systematic Botany from the Missouri Botanical Garden, 1–13.

- Arnold AE. 2007.** Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal Biology Reviews* **21**:51–66 DOI [10.1016/j.fbr.2007.05.003](https://doi.org/10.1016/j.fbr.2007.05.003).
- Arnold AE, Lutzoni F. 2007.** Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* **88**:541–549 DOI [10.1890/05-1459](https://doi.org/10.1890/05-1459).
- Bahram M, Kõljalg U, Courty P-E, Diédhiou AG, Kjølner R, Pölme S, Ryberg M, Veldre V, Tedersoo L. 2013.** The distance decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. *Journal of Ecology* **101**:1335–1344 DOI [10.1111/1365-2745.12120](https://doi.org/10.1111/1365-2745.12120).
- Barreto DP, Conrad R, Klose M, Claus P, Enrich-Prast A. 2014.** Distance-decay and taxa-area relationships for bacteria, archaea and methanogenic archaea in a tropical lake sediment. *PLOS ONE* **9**:e110128 DOI [10.1371/journal.pone.0110128](https://doi.org/10.1371/journal.pone.0110128).
- Blackwell M. 2011.** The fungi: 1, 2, 3...5.1 million species? *American Journal of Botany* **98**:426–438 DOI [10.3732/ajb.1000298](https://doi.org/10.3732/ajb.1000298).
- Busby PE, Ridout M, Newcombe G. 2016.** Fungal endophytes: modifiers of plant disease. *Plant Molecular Biology* **90**:645–655 DOI [10.1007/s11103-015-0412-0](https://doi.org/10.1007/s11103-015-0412-0).
- 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, Ruth LE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R. 2010.** QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**:335–336 DOI [10.1038/nmeth.f.303](https://doi.org/10.1038/nmeth.f.303).
- 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. *Ecological Monographs* **84**:45–67 DOI [10.1890/13-0133.1](https://doi.org/10.1890/13-0133.1).
- Chen H. 2016.** VennDiagram: generate high-resolution Venn and Euler Plots. R package version 1.6.17. Available at <https://CRAN.R-project.org/package=VennDiagram>.
- Chiu C-H, Chao A. 2016.** Estimating and comparing microbial diversity in the presence of sequencing errors. *PeerJ* **4**:e1634 DOI [10.7717/peerj.1634](https://doi.org/10.7717/peerj.1634).
- Da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, Zepeda-Mendoza ML, Campos PF, Heller R, Pereira RJ. 2016.** Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics* **30**:3–13 DOI [10.1016/j.margen.2016.04.012](https://doi.org/10.1016/j.margen.2016.04.012).
- Doyle JJ, Doyle JL. 1987.** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Photochemical Bulletin* **19**:11–15.
- Edgar RC. 2010.** Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–2461 DOI [10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461).
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011.** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**(16):2194–2200 DOI [10.1093/bioinformatics/btr381](https://doi.org/10.1093/bioinformatics/btr381).
- Gardes M, Bruns M. 1993.** ITS primers with enhanced specificity for Basidiomycetes—application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**:113–118 DOI [10.1111/j.1365-294X.1993.tb00005.x](https://doi.org/10.1111/j.1365-294X.1993.tb00005.x).

- Givnish TJ, Millam KC, Mast AR, Paterson TB, Theim TJ, Hipp AL, Henss JM, Smith JF, Wood KR, Sytsma KJ. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences* 276:407–416 DOI 10.1098/rspb.2008.1204.
- Hannon Lab 2016.FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. Available at [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/).
- Hawksworth DL. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* 6:641–655 DOI 10.1016/S0953-7562(09)80810-1.
- Hemmes DE, Desjardin DE. 2002. Mushrooms of Hawai'i. In: *Introduction*. Berkeley: Ten Speed Press, 8–9.
- Hibbett DS, Gilbert L-B, Donoghue MJ. 2000. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature* 407:506–508 DOI 10.1038/35035065.
- Hijmans RJ. 2016. geosphere: spherical trigonometry. R package version 1.5-5. Available at <https://CRAN.R-project.org/package=geosphere>.
- Hoffman MT, Arnold AE. 2008. Geographic locality and host identity shape fungal endophyte communities in cupressaceous trees. *Mycological Research* 112:331–344 DOI 10.1016/j.mycres.2007.10.014.
- Hsieh TC, Ma KH, Chao A. 2016. iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.12. Available at <http://chao.stat.nthu.edu.tw/blog/software-download/>.
- Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology* 67:4399–4406 DOI 10.1128/AEM.67.10.4399-4406.2001.
- Illumina 2013.16S metagenomic sequencing library preparation. Available at [http://support.illumina.com/downloads/16s\\_metagenomic\\_sequencing\\_library\\_preparation.html](http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html) (accessed on 00 January 2015).
- Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA. 2011. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474:200–203 DOI 10.1038/nature09984.
- Jumpponen A, Jones KL. 2009. Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist* 184:438–448 DOI 10.1111/j.1469-8137.2009.02990.x.
- Kendrick B. 2001. *The fifth kingdom*. Ontario: Mycologue Publications.
- Köljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Pöldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of Fungi. *Molecular Ecology* 22:5271–5277 DOI 10.1111/mec.12481.

- Lips KR, Brem F, Brenes R, Reeve JD, Alford RA, Voyles J, Carey C, Livo L, Pessier AP, Collins JP. 2006. Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. *Proceedings of the National Academy of Sciences of the United States of America* **103**:3165–3170 DOI [10.1073/pnas.0506889103](https://doi.org/10.1073/pnas.0506889103).
- Mattick JS, Ablett EM, Edmonson DL. 1992. The gene library—preservation and analysis of genetic diversity in Australasia. In: Adams RP, Adams JE, eds. *Conservation of plant genes: DNA banking and in vitro biotechnology*. San Diego: Academic Press, 15–35.
- McMurdie PJ, Paulson JN. 2016. biomformat: an interface package for the BIOM file format. Available at <https://github.com/joey711/biomformat/>, <http://biom-format.org>.
- Morden CW. 2017. Hawaiian plant DNA library. Available at <http://www.botany.hawaii.edu/hawaiian-plant-dna-library/> (accessed on 15 April 2017).
- Morden CW, Caraway V, Motley TJ. 1996. Development of a DNA library for native Hawaiian plants. *Pacific Science* **50**:324–335.
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403**:853–858 DOI [10.1038/35002501](https://doi.org/10.1038/35002501).
- Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knight R. 2013. Advancing our understanding of the human microbiome using QIIME. In: *Methods in enzymology*. Amsterdam: Elsevier, 371–444.
- Nekola JC, White PS. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* **26**:867–878 DOI [10.1046/j.1365-2699.1999.00305.x](https://doi.org/10.1046/j.1365-2699.1999.00305.x).
- Nilsson RH, Wurzbacher C, Bahram M, Coimbra VRM, Larsson E, Tedersoo L, Eriksson J, Duarte C, Svantesson S, Sánchez-García M, Ryberg MK, Kristianson E, Abarenkov K. 2016. Top 50 most wanted fungi. *MycoKeys* **12**:29–40 DOI [10.3897/mycokeys.12.7553](https://doi.org/10.3897/mycokeys.12.7553).
- O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R. 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* **71**:5544–5550 DOI [10.1128/AEM.71.9.5544-5550.2005](https://doi.org/10.1128/AEM.71.9.5544-5550.2005).
- Oksanen J, Blanchet FG, Friendly M, Roeland K, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2017. vegan: community ecology package. R package version 2.4-2. Available at <https://CRAN.R-project.org/package=vegan>.
- Porrás-Alfaro A, Bayman P. 2011. Hidden fungi, emergent properties: endophytes and microbiomes. *Annual Review of Phytopathology* **49**:291–315 DOI [10.1146/annurev-phyto-080508-081831](https://doi.org/10.1146/annurev-phyto-080508-081831).
- R Core Team. 2017. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>.
- Randell RA, Morden CW. 1999. Hawaiian plant DNA library II: endemic, indigenous, and introduced species. *Pacific Science* **53**:401–417.



- Rodriguez RJ, White Jr JF, Arnold AE, Redman RS. 2009. Fungal endophytes: diversity and functional roles. *New Phytologist* 182:314–330  
DOI 10.1111/j.1469-8137.2009.02773.x.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ Preprints* 4:e2409v1.
- Royal Botanic Gardens Kew 2016. *State of the world's plants report— 2016*. Kew: Royal Botanic Gardens.
- Sakai AK, Wagner WL, Mehrhoff LA. 2002. Patterns of endangerment in the Hawaiian flora. *Systematic Biology* 51:276–302 DOI 10.1080/10635150252899770.
- Santamaría J, Bayman P. 2005. Fungal epiphytes and endophytes of coffee leaves (*Coffea arabica*). *Microbial Ecology* 50:1–8 DOI 10.1007/s00248-004-0002-1.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences of the United States of America* 109:6241–6246 DOI 10.1073/pnas.1117018109.
- Seberg O, Droege G, Barker K, Coddington JA, Funk V, Gostel M, Petersen G, Smith PP. 2016. Global genome biodiversity network: saving a blueprint of the tree of life—a botanical perspective. *Annals of Botany* 118:393–399 DOI 10.1093/aob/mcw121.
- Smithsonian Institution. 2017. Flora of the Hawaiian Islands. Available at <http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/query2.cfm?radio> (accessed on 23 March 2017).
- Spooner DM, Ruess H. 2014. Curating DNA specimens. In: Applequist WL, Campbell LM, eds. *DNA banking for the 21st century*. St. Louis: William L. Brown Center, 71–80.
- Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nusbaum C, Ruess RW. 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* 84:3–20 DOI 10.1890/12-1693.1.
- Turner TR, James EK, Poole PS. 2013. The plant microbiome. *Genome Biology* 14:209 DOI 10.1186/gb-2013-14-6-209.
- US Fish and Wildlife Service. 2015. Listed species believed to or known to occur in Hawaii. Available at <https://ecos.fws.gov/ecp0/reports/species-listed-by-state-report?state=HI&status=listed> (accessed on 23 March 2017).
- Vacher C, Hampe A, Porté AJ, Sauer U, Compant S, Morris CE. 2016. The phyllosphere: microbial jungle at the plant–climate interface. *Annual Review of Ecology, Evolution, and Systematics* 47:1–24 DOI 10.1146/annurev-ecolsys-121415-032238.
- Van der Heijden MGA, Bardgett RD, Van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* 11:296–310 DOI 10.1111/j.1461-0248.2007.01139.x.
- Vaz ABM, Fontenla S, Rocha FS, Brandão LR, Vieira MLA, De Garcia V, Góes-Neto A, Rosa CA. 2014. Fungal endophyte  $\beta$ -diversity associated with Myrtaceae species in an Andean Patagonian forest (Argentina) and an Atlantic forest (Brazil). *Fungal Ecology* 8:28–36 DOI 10.1016/j.funeco.2013.12.008.

- Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. 1997.** Human domination of earth's ecosystems. *Science* **277**:494–499 DOI [10.1126/science.277.5325.494](https://doi.org/10.1126/science.277.5325.494).
- Vorholt JA. 2012.** Microbial life in the phyllosphere. *Nature Reviews Microbiology* **10**:828–840 DOI [10.1038/nrmicro2910](https://doi.org/10.1038/nrmicro2910).
- Wagner WL, Herbst DR, Sohmer SH. 1999.** Manual of the flowering plants of Hawai'i. In: Mill SW, ed. *Summary and brief analysis of the Hawaiian flora*. Honolulu: University of Hawai'i Press and Bishop Museum Press, 3–6.
- White TJ, Bruns T, Lee S, Taylor JW. 1990.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR protocols: a guide to methods and applications*. New York: Academic Press, Inc, 315–322.
- Whitman WB, Coleman DC, Wiebe WJ. 1998.** Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America* **95**:6578–6583 DOI [10.1073/pnas.95.12.6578](https://doi.org/10.1073/pnas.95.12.6578).
- Zhang J, Kobert K, Flouri T, Stamatakis A. 2014.** PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**:614–616 DOI [10.1093/bioinformatics/btt593](https://doi.org/10.1093/bioinformatics/btt593).
- Zimmerman NB, Vitousek PM. 2012.** Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences of the United States of America* **109**:13022–13027 DOI [10.1073/pnas.1209872109](https://doi.org/10.1073/pnas.1209872109).