Not just contaminants: Uncovering unseen microbial biodiversity from plant DNA banks

Erin M Datlof [1] , **Anthony S Amend** [1] , **Kamala Earl** [2] , **Jeremy Hayward** [1] , **Clifford W Morden** [1] , **Rachael Wade** [1] , **Geoffrey Zahn** [1] , **Nicole A Hynson** Corresp. [1]

[1] Department of Botany, University of Hawai'i at Mānoa, Honolulu, HI, United States

[2] University of Florida, School of Forest Resources and Conservation, Gainesville, FL, United States

Corresponding Author: Nicole A Hynson
Email address: nhynson@hawaii.edu

DNA banks are used as storage repositories for genetic diversity of organisms ranging from plants to insects to mammals throughout the world. These banks preserve the genetic information for organisms of interest, however they 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 phyllosphere among species of the endemic Hawaiian plant genus, *Clermontia* (Campanulaceae). From just 20 DNA bank samples collected throughout the main Hawaiian Islands using next generation DNA amplicon sequencing, we uncovered approximately 1,780 fungal operational taxonomic units. Using these historic 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 also provides the first insights into elucidating patterns of microbial diversity through the use of DNA bank repository samples.

Not just contaminants: Uncovering unseen microbial biodiversity from plant DNA banks

Erin M. Datlof[1], Anthony S. Amend[1], Kamala Earl[2], Jeremy Hayward[1], Clifford W. Morden[1], Rachael Wade[1], Geoffrey Zahn[1], Nicole A. Hynson[1]

[1] Department of Botany, University of Hawai‘i at Mānoa, Honolulu, HI, United States

[2] School of Forest Resources and Conservation, University of Florida, Gainesville, FL, United States

Corresponding Author:

Nicole A. Hynson[1]

Department of Botany, University of Hawai‘i at Mānoa, Honolulu, HI, 96822, United States

Email address: nhynson@hawaii.edu

Abstract

DNA banks are used as storage repositories for genetic diversity of organisms ranging from plants to insects to mammals throughout the world. These banks preserve the genetic information for organisms of interest, however they 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 phyllosphere among species of the endemic Hawaiian plant genus, *Clermontia* (Campanulaceae). From just 20 DNA bank samples collected throughout the main Hawaiian Islands using next generation DNA amplicon sequencing, we uncovered approximately 1,780 fungal operational taxonomic units. Using these historic 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 also provides the first insights into elucidating patterns of microbial diversity through the use of DNA bank repository samples.

Introduction

Understanding biodiversity is an important goal of biology and ecology. 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 evolution is lost as well (Mattick et al., 1992). DNA banks were initially developed to collect genetic material in order to create a storage base for evolutionary

history, biological diversity, and genomic information (Mattick et al., 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 from specific locations. For example, plant bank samples not only preserve the targeted species' genomic information, but also preserve potentially important cryptic symbionts associated with their host, such as fungi known to inhabit the plant phyllosphere (Porras-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). These increases in estimates of fungal species richness are due in part to advances in direct environmental sequencing and extrapolations based on predictions of vascular plant to fungal ratios (O'Brien et al., 2011; Taylor et al., 2014). In order to obtain more accurate estimates of true fungal diversity, increased sampling using high throughput 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 90'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 (~86%; Hemmes and Desjardin 2002). Similar rates of endemism are found in the Hawaiian flora.

An estimated 89% of the Hawaiian vascular plant flora is endemic (Wagner et al., 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 & Wildlife, 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. This library preserves Hawaiian plant DNA and banks these samples for use in future studies of biodiversity (Morden, Caraway & Motley, 1996; Randell & Morden, 1999). All wild 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

93 species per tree and ranges of about 700-4,000 species of fungi per host (Jumpponen & Jones,
94 2009; Zimmerman & Vitousek, 2012). Thus, the HPDL has also likely and coincidentally
95 preserved a substantial portion of the diversity of Hawaiian fungi.
96 In this study, we utilize historic DNA bank samples to examine plant-associated fungal
97 diversity across space, and validate the use of plant DNA bank samples as a resource for
98 elucidating phyllosphere fungal biodiversity. As a model plant system, we selected a single
99 endemic Hawaiian plant genus, *Clermontia* (Campanulaceae), with species found across the
100 Hawaiian Islands (Givnish et al., 2009). Using DNA samples of eight species within this genus,
101 we sequenced the fungi found in these plants' phyllospheres. We took advantage of the
102 archipelago's geographic spatial gradient and the previously collected samples in the bank to test
103 for decreases in community similarity as the distance between communities increases, the
104 classical ecological pattern of distance-decay of community similarity (Nekola & White, 1999).
105
106 **Materials & Methods**
107 *Samples*
108 Twenty individual *Clermontia* foliar DNA extracts, representing eight species, were
109 selected from the Hawaiian Plant DNA Library (Morden, 2017). These specimens were collected
110 across the main Hawaiian Archipelago, from Hawai‘i Island to Kaua‘i (Table 1.; see
111 Supplementary Table S1. for more details). Samples obtained and stored in the DNA Library
112 were collected in the field, sealed in bags, and brought back to the lab. Leaves were not disturbed
113 by rinsing prior to DNA extraction. Approximately 1.0g of leaf tissue was extracted using a
114 modified CTAB method with cesium chloride banding and stored at -20°C (Doyle & Doyle,
115 1987; Morden, Caraway & Motley, 1996). For this study 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, 2015). 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.), followed by a second PCR attaching forward and reverse eight-base pair barcoded Illumina overhang adapters (i7 and i5; Illumina, 2015). See Supplementary 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 2x300 (Illumina, San Diego, CA, USA).

Bioinformatics

139 De-multiplexed fastq files were obtained from the sequencing facility from the ten
140 *Clermontia* plant bank samples. Raw sequencing data was deposited to the National Center for
141 Biotechnology Information Sequence Read Archive (NCBI SRA) under BioProject
142 PRJNA379349. These paired-end reads were merged with the Illumina Paired-End reAd mergeR
143 (PEAR), keeping reads with a minimum assembly length of 250-bp, average quality threshold of
144 15 and above, and discarding all reads with any uncalled bases (Zhang et al., 2014). Further
145 quality control was carried out using the FASTX-Toolkit
146 (http://hannonlab.cshl.edu/fastx_toolkit/), using the fastq_quality_filter command (Hannon Lab,
147 2016), where all reads with any base pairs containing a quality score below 15 were discarded
148 (Hannon Lab, 2016). Potential chimeras were removed in vsearch (Rognes et al., 2016) using the
149 uchime_ref command (Edgar et al., 2011), which referenced the User-friendly Nordic ITS
150 Ectomycorrhiza (UNITE) database, accessed on 11.03.2015 (Kõljalg et al., 2013). Operational
151 Taxonomic Units (OTUs) were clustered using the open-reference method (Navas-Molina et al.,
152 2013) in QIIME (Caporaso et al., 2010). Briefly, reads were matched to reference OTUs in the
153 UNITE dynamic database (ver7) (Kõljalg et al., 2010) with added *Clermontia* outgroups, then
154 remaining reads that failed to match were subsampled as seeds for three subsequent rounds of *de*
155 *novo OTU-picking.* The most abundant sequence for each OTU was chosen as a representative
156 sequence. Singleton reads were removed in QIIME prior to OTU table generation and taxonomy
157 was assigned against the UNITE database with the Basic Local Alignment (BLAST) algorithm.
158
159 *Statistics*
160 All statistical analyses were conducted in R version 3.3.0 (R Core Team, 2017). The
161 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 offer numerous advantages over other diversity indices, such as uniting species diversity and similarity, obeying the replication principle of species assemblages, and they are expressed in units of effective number of species (Chao et al., 2014). Two Hill number diversity values 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). To better visualize taxonomic abundances, a heat tree was created for each OTU at all taxonomic assignments in the *metacoder* package (Foster, 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*

185 package (Hijmans, 2016), time between sample collections in days, and Bray-Curtis community
186 dissimilarity using the *vegan* package (Oksanen et al., 2017). A partial mantel test for physical
187 distance and community dissimilarity, while controlling for temporal distances between each
188 sample, was run with 10,000 permutations (Oksanen et al., 2017).
189
190 **Results**
191 *Sequencing*
192 A total of 4,312,473 sequence reads were obtained from the plant DNA library samples.
193 Of these, 3,571,252 paired-end reads (82.8%) were successfully assembled and further quality
194 control removed low quality reads, keeping 2,680,945 reads (75.1%). After referencing UNITE,
195 121,618 (4.5%) chimeric sequences were removed, leaving a total of 2,559,327 high-quality
196 reads. Taxonomic assignment yielded a total of 1,648,971 fungal reads that were binned into
197 2,944 fungal OTUs for use in in downstream analyses.
198
199 *Fungal Diversity and Host Associations*
200 Each *Clermontia* DNA bank sample used in this study contained fungal DNA. In total,
201 we found 2,944 fungal phyllosphere OTUs associated with the ten *Clermontia* DNA bank
202 samples. After removing OTUs with less than ten reads and rarefying to the sample with the
203 minimum number of reads, we removed 1,164 OTUs (39.5%) and were left with a total of 1,780
204 fungal OTUs. The observed species accumulation curve for all ten samples did not reach an
205 asymptote, suggesting there is potentially far more fungal diversity to uncover with increased
206 sampling (Figure S1). Based on our ten samples the *iNEXT* extrapolation curves suggest fungal
207 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 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 (Figure 4). Observed richness per sample ranged from 108 to 682 fungal OTUs with an average of 295 OTUs per sample (± 54.69 standard error). Sequencing depth for each sample was sufficient to capture most fungal richness; all samples except for *C. fauriei* from Kauaʻi (K1) saturated their rarefaction curve (Figure S3).

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 phyllosphere samples, with phylum Ascomycota being most abundant, followed by Basidiomycota (Figures 1 & 3). 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; Figure 1). The top ten most abundant orders were Capnodiales, Chaetothyriales, Exobasidiales, Peltigerales, Pertusariales, Pleosporales, Tremellales, Ustilaginales, and two unknown orders (Figure 2). The abundances of each OTU and taxonomic assignments are shown as a heat tree in Figure 3.

Average OTU richness by island was 507.6 (± 128.458 standard error). Oʻahu had the highest richness, followed successively by Hawaiʻi, Maui, Molokaʻi, and Kauaʻi again had the lowest richness (Figure S3). This pattern was also apparent with the number of OTUs (Figure 5). Twenty OTUs were found on all of the five islands (Figure 5).

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 524.78km 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 phyllosphere communities exhibit a significant decrease in community similarity across increasing geographic distance (Figure 6, Partial Mantel test: $r = 0.423$, $p = 0.005$).

Discussion and Conclusion

In this study, we investigated the diversity of phyllosphere fungi associated with *Clermontia spp.* that were collected across the Hawaiian Islands and stored as DNA bank samples. 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 686 OTUs per plant sample. Despite high sequencing depth, the species accumulation curve for all samples did not saturate, indicating this sequencing effort likely underestimated true *Clermontia* phyllosphere fungal diversity. 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), as a single macroorganism associates with a multitude of microorganisms both within and on their surfaces (Turner, James & Poole, 2013).

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

253 fungal richness have been based off of plant to fungi ratios ranging from 1:6 (Hawksworth,
254 1991) to 1:17 (Taylor et al., 2014). Our data supplement these studies using environmental NGS
255 data. If we assume that the diversity of phyllosphere fungi associated with *Clermontia* species is
256 representative of the native Hawaiian flora, we would estimate based on Chao1 richness ($q = 0$)
257 extrapolations (determined by species; Figure S2) that the entire Hawaiian flora (c. 1,000
258 species) harbors about 4,000 fungi. This results in an approximate 1:4 plant to fungi species
259 ratio. However, just considering phyllosphere fungi associated with a single genus is likely an
260 underestimate of total fungal biodiversity due to potential host-fungi specificity (Hoffman &
261 Arnold, 2008). Supplementary to host specificity, only taking into consideration phyllosphere
262 fungi likely underestimates fungal richness due to niche partitioning among fungal species and
263 guilds (Hibbett, Gilbert & Donoghue, 2000).

264 In addition to the study of microbial diversity, questions regarding microbial
265 biogeography, host specificity, and the effects of global change on microbial communities could
266 be addressed with DNA banks. For example, we were able to confirm the distance decay of
267 microbial community similarity from DNA bank samples collected across the Hawaiian Islands.
268 This finding is similar to other microbial systems where significant distance decay patterns were
269 found in endophytic (Vaz et al., 2014) and ectomyorrhizal fungal communities (Bahram et al.,
270 2013), as well as bacteria and archaea (Barreto et al., 2014). Although our samples were not
271 collected in the same year or season, time was not a significant predictor of community
272 composition. However, in addition to geography, taking into account host genotype, specificity,
273 and differences in environmental factors may potentially explain more of the variation in the
274 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 of fungi in phylum Ascomycota followed by Basidiomycota (Rodriguez et al., 2008). It is not surprising that we found so many unknown fungal taxa (45.16% of total OTUs at the family level) including 28 OTUs unable to be placed 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, 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 in this 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 and focused on the host organism (in this case plants), microbial symbionts may interfere with downstream analyses and results. Microbial taxa associated with macroorganismns

298 should be taken into account when using NGS methods such as RAD seq, RNA seq, targeted
299 sequencing, among other techniques (da Fonseca et al., 2016).
300 Most DNA bank samples likely harbor unintended microbial communities associated
301 with each target individual from a specific location at distinct snapshots in time. While DNA
302 banks are a common genetic biodiversity repository (Seberg et al., 2016), to the best of our
303 knowledge this is the first study where they were used to investigate genetic material other than
304 that of the target organism. By using these archived samples we were able to rapidly recover
305 previously undocumented microbial diversity. The abundance of DNA bank samples stored
306 throughout the world represent a large proportion of the globes extant and extinct biological
307 diversity. This storage provides the opportunity for microbes associated with these organisms to
308 be easily investigated without the associated costs of sample collection. This may be important
309 for conservation efforts, giving insight into potentially important symbionts (van der Heijden,
310 Bardgett & Straalen, 2008; Busby et al., 2016). For those species that go extinct, their genomes
311 are preserved in DNA banks along with their corresponding microbial symbionts. These
312 associated microbes can be used to better understand the ecology of these organisms and
313 possibly identify coevolutionary patterns. Overall, this study highlights the potential use of DNA
314 bank samples for the study of global biodiversity. This study also demonstrated the benefits of
315 in-depth sample sequencing to uncover the majority of fungal diversity found in each plant bank
316 sample. With DNA bank samples stored throughout the world, already collected, processed, and
317 extracted, they harbor the potential for new and exciting investigations.
318
319 **Acknowledgements**

We thank Emily Johnston and Richard O'Rorke for experimental and lab help, further lab assistance by Leah Tooman. Also thanks to Sean Swift, Laura Tipton, and Sofia Gomes for R help. Thanks to Mitsuko Yorkston and Richard Pender for sample discussions.

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.

Arnold AE., Lutzoni F. 2007. Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* 88:541-549.

Bahram M., Kõljalg U., Courty P-E., Diédhiou AG., Kjøller R., Põlme 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.

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.

Blackwell M. 2011. The Fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany* 98:426-438.

Busby PE., Ridout M., Newcombe G. 2016. Fungal endophytes: modifiers of plant disease. *Plant Molecular Biology* 90:645-655.

342 Caporaso JG., Kuczynski J., Stombaugh J., Bittinger K., Bushman FD., Costello EK., Fierer N.,
343 Peña AG., Goodrich JK., Gordon JI., Huttley GA., Kelley ST., Knights D., Koenig JE.,
344 Ruth LE., Lozupone CA., McDonald D., Muegge BD., Pirrung M., *Reeder J*. 2010. QIIME
345 allows analysis of high-throughput community sequencing data. *Nature Methods* 7:335-336.
346 Chao A., Gotelli NJ., Hsieh TC., Sander EL., Ma KH., Colwell RK., Ellison AM. 2014.
347 Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and
348 Estimation in Species Diversity Studies. *Ecological Monographs* 84: 45-67.
349 Chen H. 2016. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package
350 version 1.6.17. *Available at https://CRAN.R-project.org/package=VennDiagram.*
351 da Fonseca RR., Albrechtsen A., Themudo GE., Ramos-Madrigal J., Sibbesen JA., Maretty L.,
352 Zepeda-Mendoza ML., Campos PF., Heller R., Pereira RJ. 2016. Next-generation biology:
353 Sequencing and data analysis approaches for non-model organisms. *Marine Genomics* 30:3-
354 13.
355 Doyle JJ., Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf
356 tissue. *Photochemical Bulletin* 19:11-15.
357 Edgar RC., Haas BJ., Clemente JC., Quince C., Knight R. 2011. UCHIME improves sensitivity
358 and speed of chimera detection**.** *Bioinformatics* 27(16):2194-2200.
359 Foster Z. 2016. metacoder: Tools for Parsing, Manipulating, and Graphing Hierarchical Data. R
360 package version 0.1.2. *Available at* https://CRAN.R-project.org/package=metacoder.
361 Gardes M., Bruns M. 1993. ITS primers with enhanced specificity for Basidiomycetes –
362 Application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2:113-118.
363 Givnish TJ., Millam KC., Mast AR., Paterson TB., Theim TJ., Hipp AL., Henss JM., Smith JF.,
364 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.

Hannon Lab. 2016. FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. *Available at http://hannonlab.cshl.edu/fastx_toolkit/.*

Hawksworth DL. 1991. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* 6:641-655.

Hemmes DE., Desjardin DE. 2002. Mushrooms of Hawai‘i. *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.

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.

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.

Illumina. 2013. 16S Metagenomic Sequencing Library Preparation. Illumina Inc. *Available at http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html*. (Accessed January 2015).

388 Jones MDM., Forn I., Gadelha C., Egan MJ., Bass D., Massana R., Richards TA. 2011.
389 Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* 474: 200-
390 203.
391 Jumpponen A., Jones KL. 2009. Massively parallel 454 sequencing indicates hyperdiverse
392 fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist*
393 184:438-448.
394 Kendrick B. 2001. *The Fifth Kingdom*. Mycologue Publications. Ontario: Mycologue
395 Publications.
396 Kõljalg U., Nilsson RH., Abarenkov K., Tedersoo L., Taylor AFS., Bahram M., Bates ST., Bruns
397 TD., Bengtsson-Palme J., Callaghan TM., *et al*. 2013. Towards a unified paradigm for
398 sequence-based identification of Fungi. *Molecular Ecology* 22:5271-5277.
399 Lips KR., Brem F., Brenes R., Reeve JD., Alford RA., Voyles J., Carey C., Livo L., Pessier AP.,
400 Collins JP. 2006. Emerging infectious disease and the loss of biodiversity in a Neotropical
401 amphibian community. *Proceedings of the national academy of sciences of the United*
402 *States of America* 103:3165-3170.
403 Mattick JS., Ablett EM., Edmonson DL. 1992. The Gene Library – Preservation and Analysis of
404 Genetic Diversity in Australasia. In: Adams RP., Adams JE., eds. *Conservation of Plant*
405 *Genes: DNA Banking and in vitro Biotechnology*. San Diego: Academic Press, 15-35.
406 McMurdie PJ., Paulson, JN. 2016. biomformat: An interface package for the BIOM file format.
407 *Available at https://github.com/joey711/biomformat/, http://biom-format.org.*
408 Morden CW., Caraway V., Motley TJ. 1996. Development of a DNA library for native Hawaiian
409 plants. *Pacific Science* 50:324-335.

410 Morden CW. 2017. Hawaiian Plant DNA Library. *Available at*
411 *http://www.botany.hawaii.edu/hawaiian-plant-dna-library/*. (Accessed 15 April 2017)
412 Myers N., Mittermeier RA., Mittermeier CG., Da Fonseca GA., Kent J. 2000. Biodiversity
413 hotspots for conservation priorities. *Nature* 403:853-858.
414 Navas-Molina JA., Peralta-Sánchez JM., González A., McMurdie PJ., Vázquez-Baeza Y., Xu Z.,
415 Ursell LK., Lauber C., Zhou H., Song SJ., Huntley J., Ackermann GL., Berg-Lyons D.,
416 Holmes S., Caporaso JG., Knight R. 2013. Advancing Our Understanding of the Human
417 Microbiome Using QIIME. *Methods in Enzymology*. Elsevier, 371-444.
418 Nekola JC., White PS. 1999. The distance decay of similarity in biogeography and ecology.
419 *Journal of Biogeography* 26:867-878.
420 Nilsson RH., Wurzbacher C., Bahram M., R. M. Coimbra V., Larsson E., Tedersoo L., Eriksson
421 J., Duarte C., Svantesson S., Sánchez-García M., Ryberg MK., Kristiansson E., Abarenkov
422 K. 2016. Top 50 most wanted fungi. *MycoKeys* 12:29-40.
423 O'Brien HE., Parrent JL., Jackson JA., Moncalvo J-M., Vilgalys R. 2005. Fungal Community
424 Analysis by Large-Scale Sequencing of Environmental Samples. *Applied and*
425 *Environmental Microbiology* 71:5544-5550.
426 Oksanen J., Blanchet FG., Friendly M., Roeland K., Legendre P., McGlinn D., Minchin PR.,
427 O'Hara RB., Simpson GL., Solymos P., Stevens MHH., Szoecs E., Wagner H. 2017. vegan:
428 Community Ecology Package. R package version. 2.4-2. *Available at https://CRAN.R-*
429 *project.org/package=vegan.*
430 Porras-Alfaro A., Bayman P. 2011. Hidden Fungi, Emergent Properties: Endophytes and
431 Microbiomes. *Annual Review of Phytopathology* 49:291-315.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *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.

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. Royal Botanic Gardens, Kew.

Sakai AK., Wagner WL., Mehrhoff LA. 2002. Patterns of endangerment in the Hawaiian flora. *Systematic biology* 51:276–302.

Santamaría J., Bayman P. 2005. Fungal Epiphytes and Endophytes of Coffee Leaves (*Coffea arabica*). *Microbial Ecology* 50:1-8.

Schoch CL., Seifert KA., Huhndorf S., Robert V., Spouge JL., Levesque CA., Chen W., Fungal Barcoding Consortium, *et al*. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* 109:6241-6246.

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.

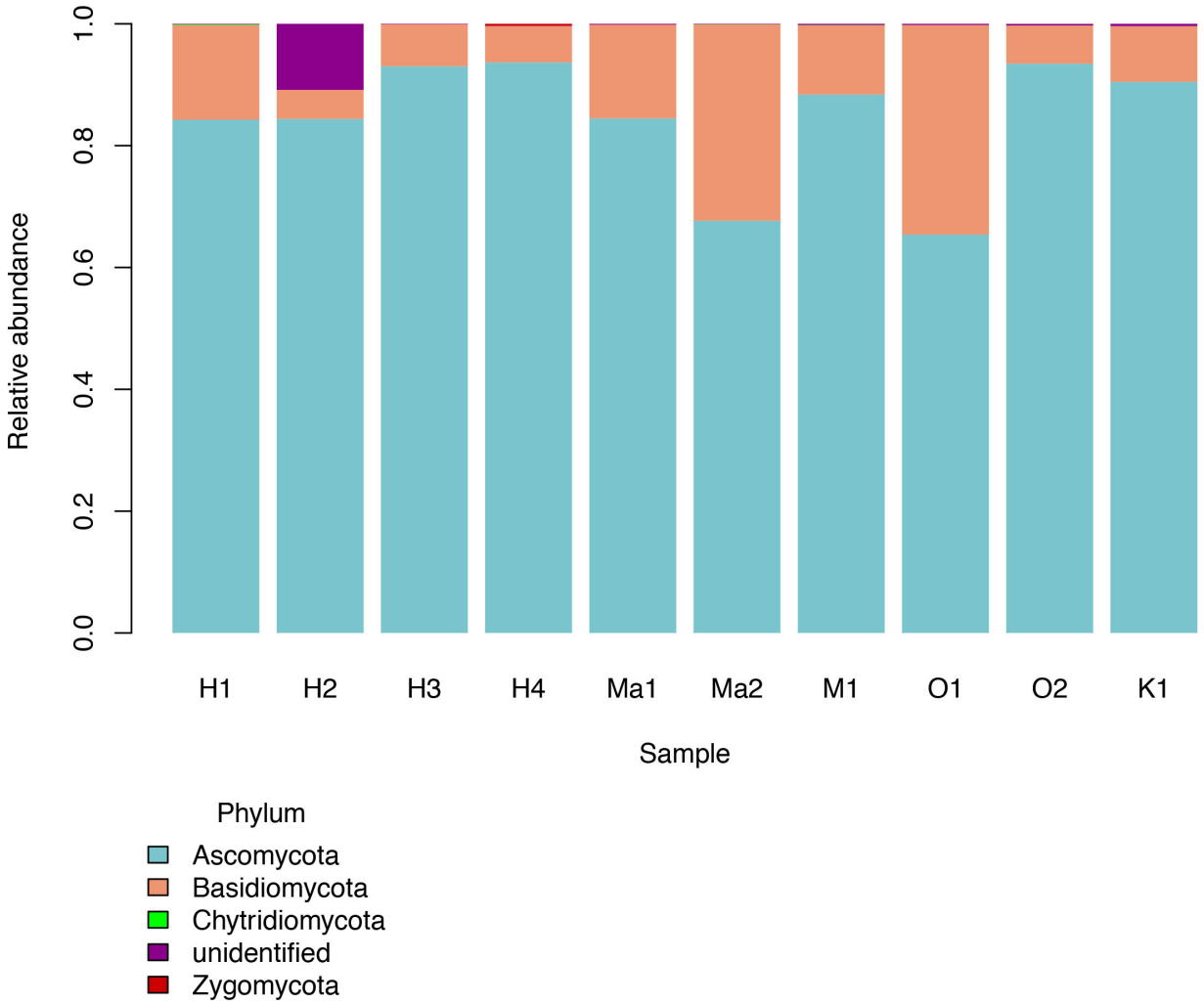
453 Smithsonian Institution. 2017. Flora of the Hawaiian Islands. *Available at*
454 *http://botany.si.edu/pacificislandbiodiversity/hawaiianflora/query2.cfm?radio* (Accessed 23
455 March 2017).
456 Spooner DM., Ruess, H. 2014. Curating DNA specimens. In: Applequist WL., Campbell LM.,
457 eds. *DNA Banking for the 21st Century*. St. Louis: William L. Brown Center, 71-80.
458 Taylor DL., Hollingsworth TN., McFarland JW., Lennon NJ., Nusbaum C., Ruess RW. 2014. A
459 first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche
460 partitioning. *Ecological Monographs* 84:3-20.
461 Turner TR., James EK., Poole PS. 2013. The plant microbiome. *Genome biology* 14:209.
462 US Fish and Wildlife Service. 2015. Listed species believed to or known to occur in Hawaii.
463 *Available at https://ecos.fws.gov/ecp0/reports/species-listed-by-state-*
464 *report?state=HI&status=listed* (Accessed 23 March 2017).
465 Vacher C., Hampe A., Porté AJ., Sauer U., Compant S., Morris CE. 2016. The Phyllosphere:
466 Microbial Jungle at the Plant–Climate Interface. *Annual Review of Ecology, Evolution, and*
467 *Systematics* 47:1-24.
468 van der Heijden MGA., Bardgett RD., van Straalen NM. 2008. The unseen majority: soil
469 microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology*
470 *Letters* 11:296-310.
471 Vaz ABM., Fontenla S., Rocha FS., Brandão LR., Vieira MLA., de Garcia V., Góes-Neto A.,
472 Rosa CA. 2014. Fungal endophyte β-diversity associated with Myrtaceae species in an
473 Andean Patagonian forest (Argentina) and an Atlantic forest (Brazil). *Fungal Ecology* 8:28-
474 36.

475 Vitousek PM., Mooney HA., Lubchenco J., Melillo JM. 1997. Human domination of Earth's
476 ecosystems. *Science* 277:494-499.
477 Vorholt JA. 2012. Microbial life in the phyllosphere. *Nature Reviews Microbiology* 10:828-840.
478 Wagner WL., Herbst DR., Sohmer SH. 1999. Manual of the flowering plants of Hawai'i. In:
479 Mill SW., ed. *Summary and brief analysis of the Hawaiian flora*. Honolulu: University of
480 Hawai'i Press and Bishop Museum Press, 3-6.
481 White TJ., Bruns T., Lee S., Taylor JW. 1990. Amplification and direct sequencing of fungal
482 ribosomal RNA genes for phylogenetics. In: Innis, MA., Gelfand DH., Sninsky JJ., White
483 TJ, eds. *PCR Protocols: A Guide to Methods and Applications*. New York: Academic Press,
484 Inc, 315-322.
485 Whitman WB., Coleman DC., Wiebe WJ. 1998. Prokaryotes: The unseen majority. *Proceedings*
486 *of the National Academy of Sciences* 95:6578-6583.
487 Zhang J., Kobert K., Flouri T., Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-
488 End reAd mergeR. *Bioinformatics* 30:614-6.
489 Zimmerman NB., Vitousek PM. 2012. Fungal endophyte communities reflect environmental
490 structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences*
491 109:13022-13027.
492
493
494
495
496
497

Tables and Figures

Table 1: Plant bank samples and accession numbers from the Hawaiian Plant DNA Library for each extract along with associated metadata. Two individual extracts were pooled for each location and given a sample code labeled by island.

Sample	Sample Code	HPDL Number	*Clermontia* species	Island	Date Extracted	Latitude	Longitude
1	M1	6843	*kakeana*	Moloka‘i	7/14/11	21.13	-156.92
1a		6844					
2	H1	6961	*calophylla*	Hawai‘i	11/18/11	20.09	-155.74
2a		6962					
3	H2	6888	*kohalae*	Hawai‘i	9/17/11	20.08	-155.74
3a		6889					
4	H3	6856	*clermoniotides*	Hawai‘i	8/3/11	19.21	-155.60
4a		6857					
5	H4	7339	*peleana ssp. singulariflora*	Hawai‘i	6/7/13	20.18	-155.80
5a		7940					
6	K1	5089	*fauriei*	Kaua‘i	9/6/05	22.09	-159.59
6a		5090					
7	O1	6809	*kakeana*	O‘ahu	6/17/11	21.34	-157.82
7a		6810					
8	O2	7008	*oblongifolia ssp. oblongifolia*	O‘ahu	3/15/12	21.41	-158.10
8a		7009					
9	Ma1	6875	*arborescens*	Maui	7/5/11	20.82	-156.28
9a		6876					
10	Ma2	6831	*kakeana*	Maui	7/14/11	20.80	-156.23
10a		6832					



503
504 **Figure 1: Relative abundances of fungal phyla for each *Clermontia spp.* DNA bank sample.**

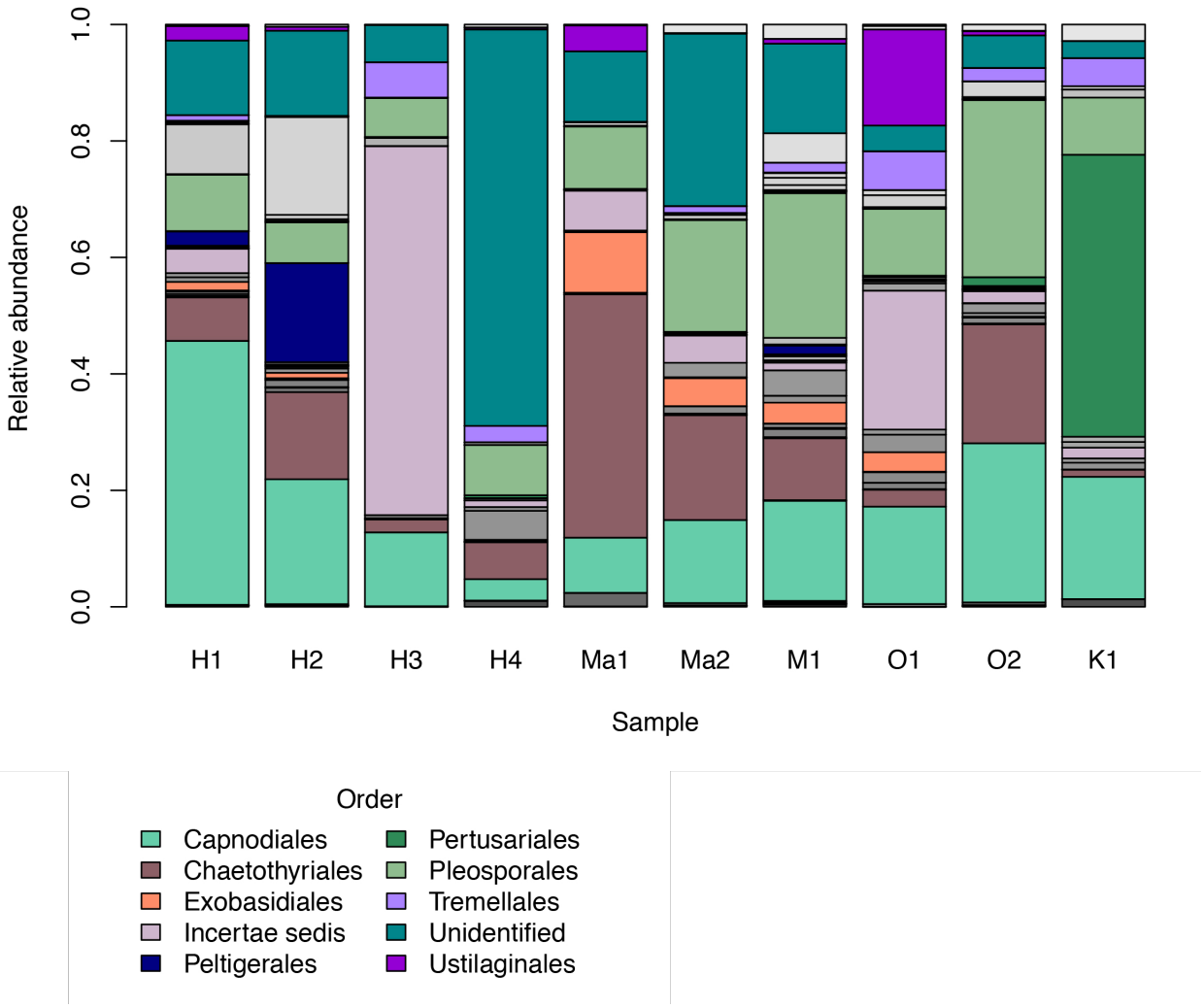


Figure 2: Relative abundances of the top 10 most abundant fungal orders for each *Clermontia spp.* DNA bank sample. All other orders are filled with grayscale.

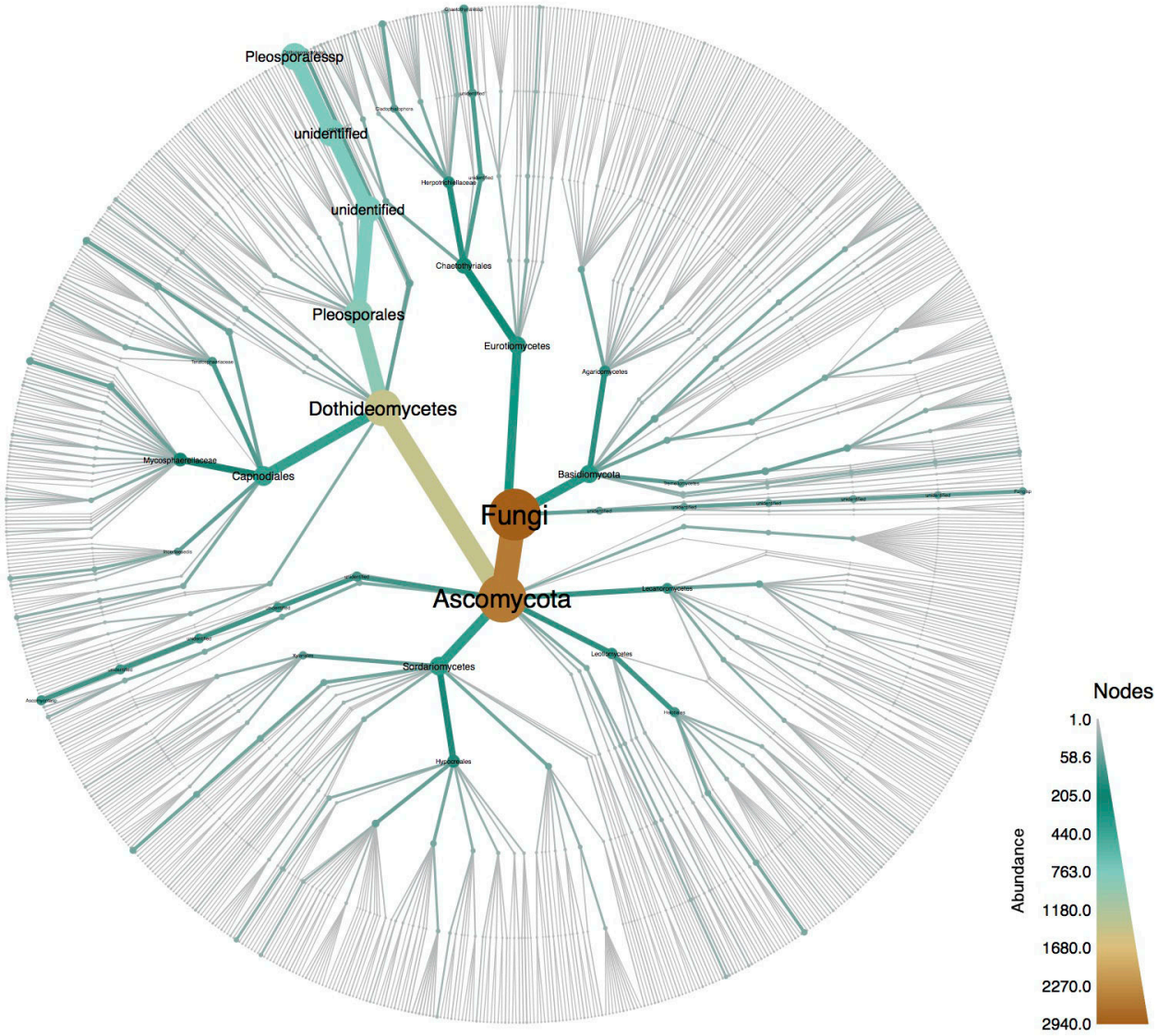


Figure 3: Heat tree for all fungal OTUs and higher taxonomy in the *Clermontia* phyllosphere. Size and color of nodes, from grey to orange, as well as edge widths are correlated with the abundance of each taxonomic assignment in the phyllosphere.

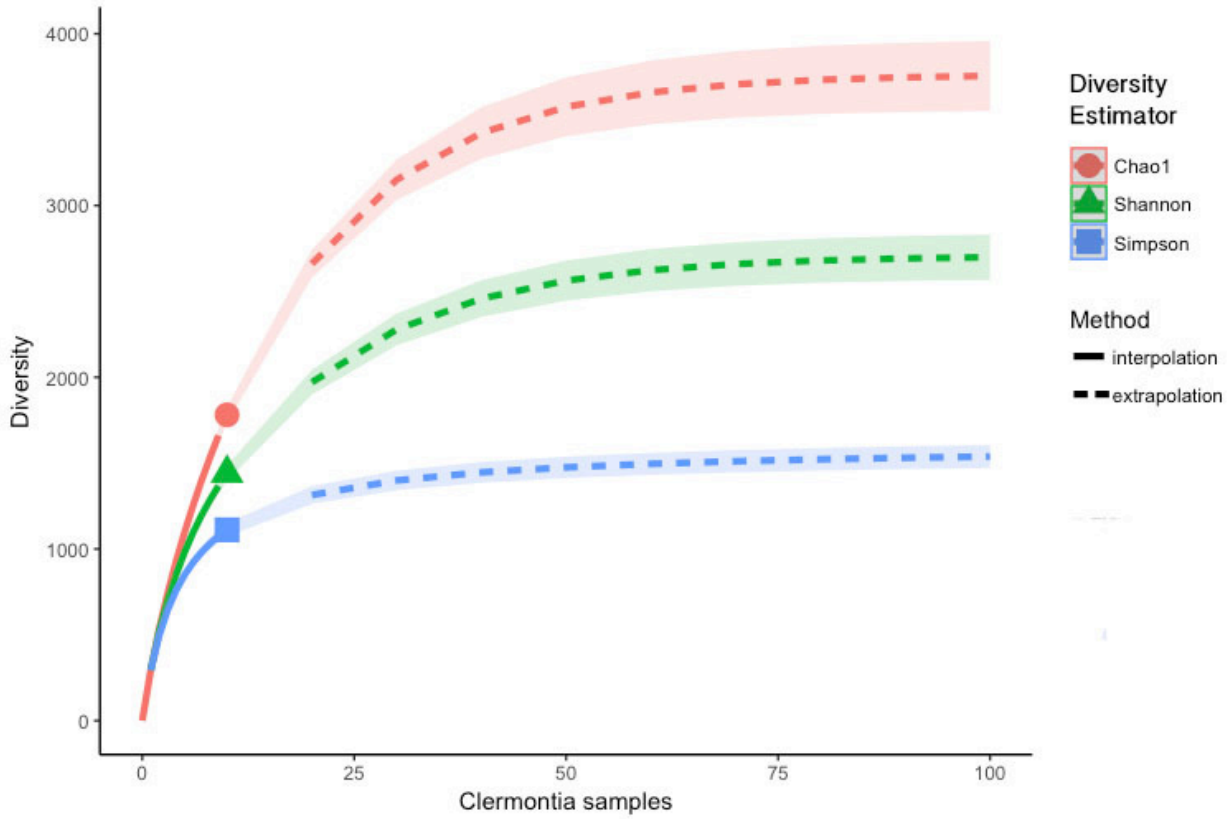
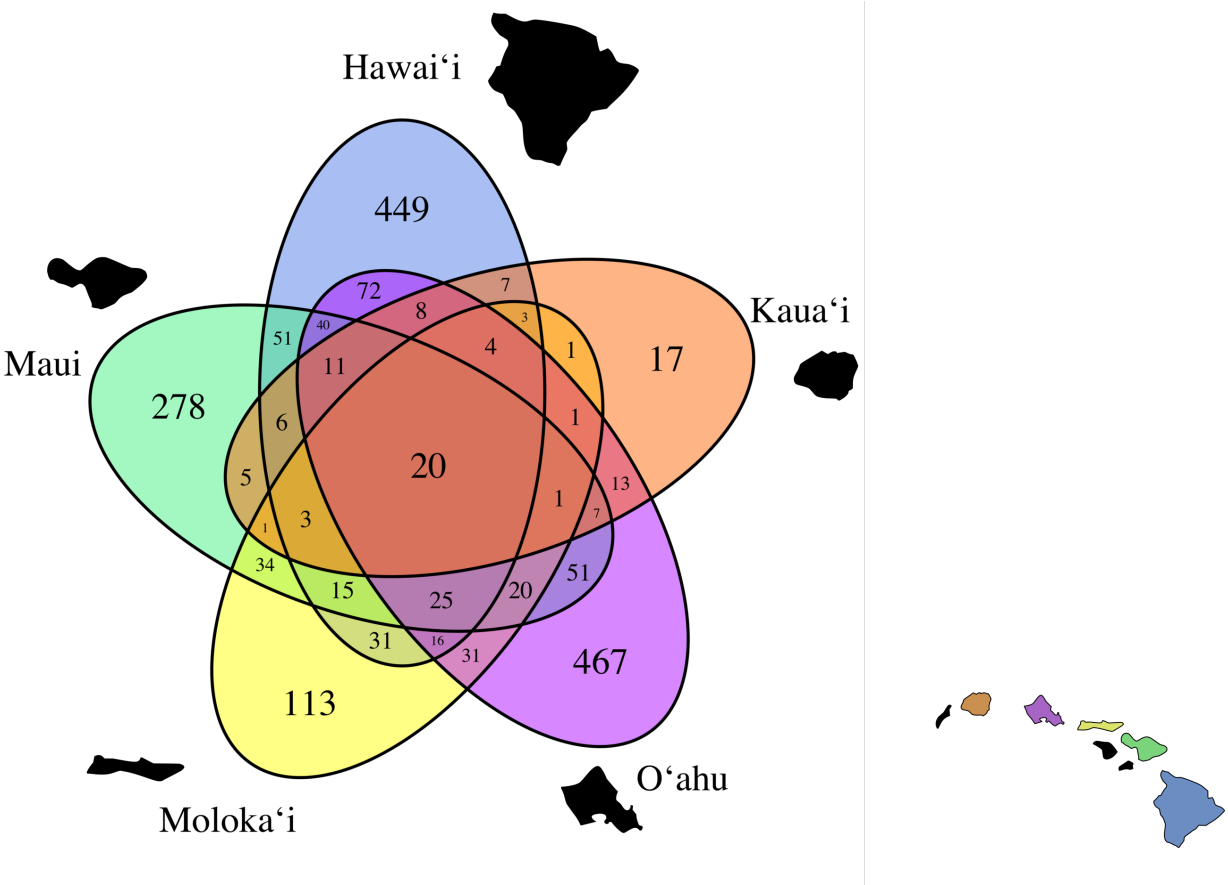


Figure 4: Sample interpolation (solid lines) and extrapolation (dashed lines) curves for all ten *Clermontia* plant bank samples using the rarefied OTU matrix. Three different diversity estimators were used (Chao1 richness, exponential of Shannon entropy, and inverse Simpson concentration indices) and are shown by the different colors with 95% confidence intervals shown by shading. Shapes represent observed plant bank sample diversity calculations.



519
520 **Figure 5: Venn Diagram displaying the number of overlapping fungal OTUs shared**
521 **between *Clermontia* samples from each of the five main Hawaiian Islands. The number of**
522 **OTUs unique to each island lie on the outermost portion of each ellipse. Color**
523 **corresponding islands are shown next to each ellipse and the Hawaiian island chain is**
524 **shown in the lower right hand corner.**
525
526
527

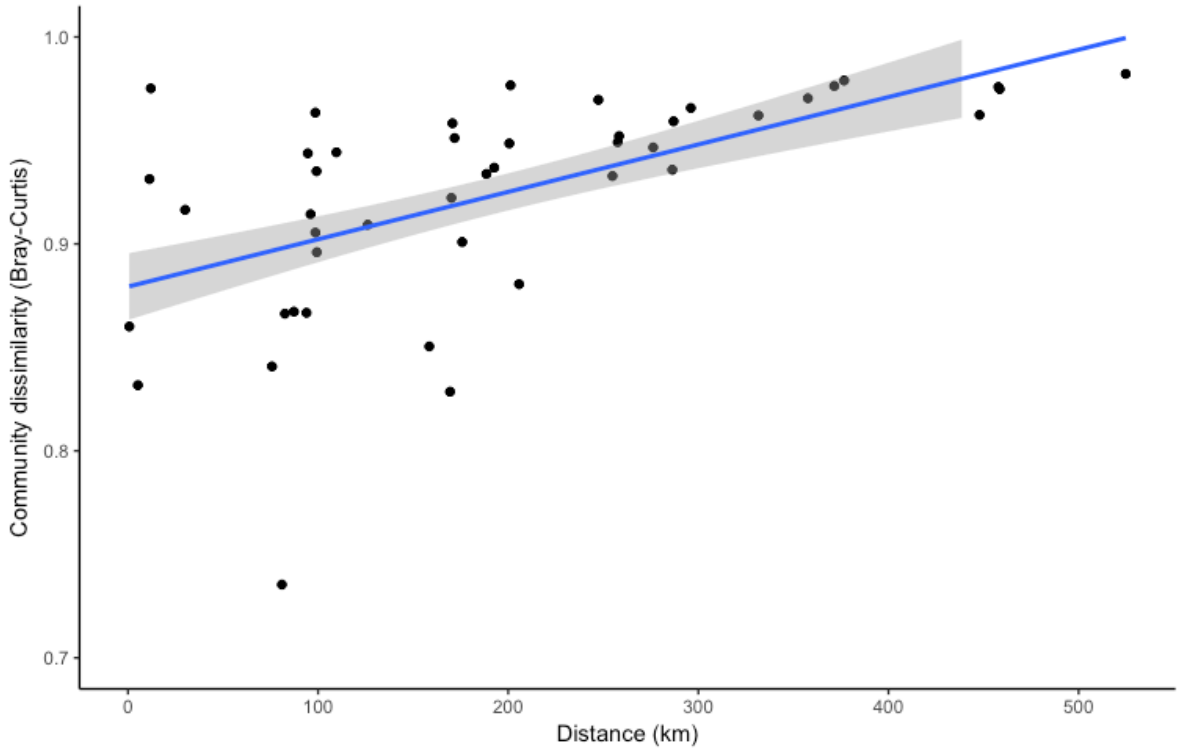


Figure 6: Pair-wise Bray-Curtis fungal community dissimilarity plotted against corresponding pair-wise physical distances for each *Clermontia* plant bank sample spanning the main Hawaiian Islands. A regression line was fit to the data, shown in blue, with 95% confidence intervals shown in grey. (Partial Mantel test: $r = 0.424$, $p = 0.005$, accounting for time in days).

554 ***Supplementary Information***
555
556 **Table S1: *Clermontia* DNA bank sample information and extended metadata.**

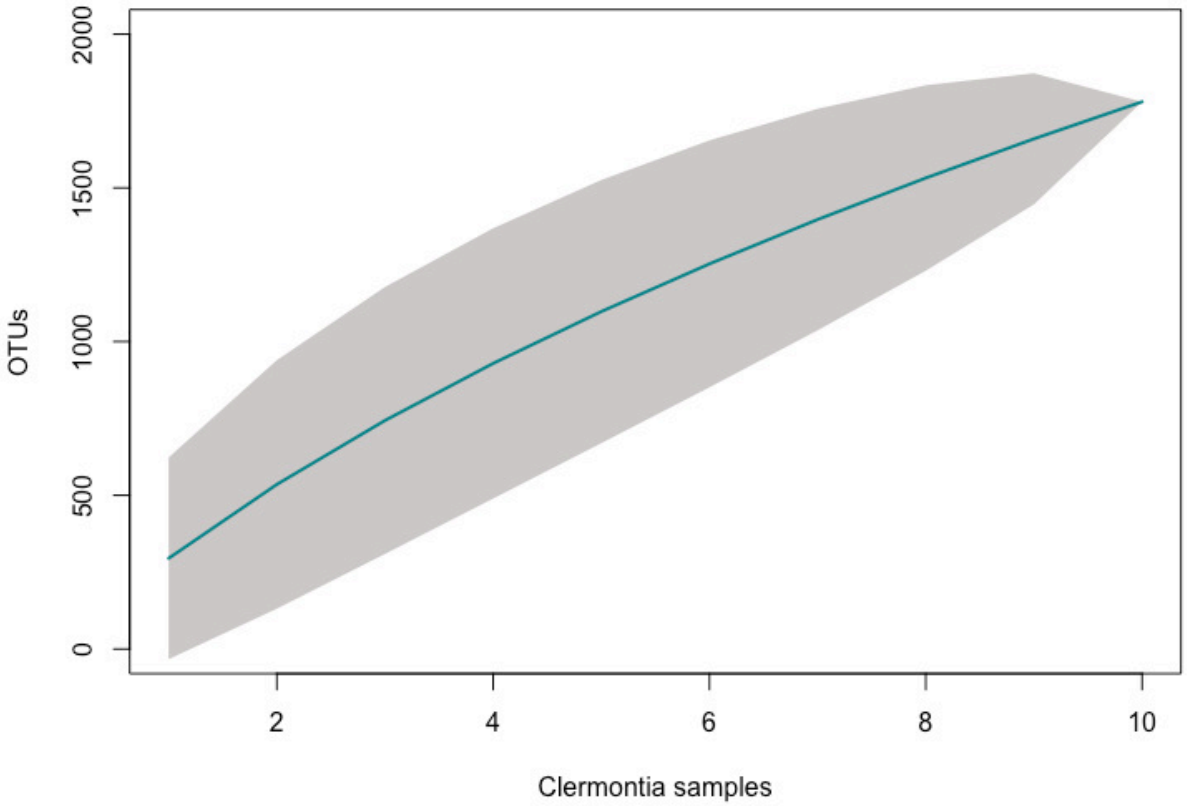
Sample Code	HPDL Number	*Clermontia* species	Island	Date Extracted	Latitude	Longitude	Location	Collector
M1	6843 6844	*kakeana*	Moloka'i	7/14/11	21.13	-156.92	Kamakou Preserve	Richard Pender
H1	6961 6962	*calophylla*	Hawai'i	11/18/11	20.09	-155.74	Pu'u O 'Umi Natural Area Reserve	Richard Pender
H2	6888 6889	*kohalae*	Hawai'i	9/17/11	20.08	-155.74	Kohala Mts.	Richard Pender
H3	6856 6857	*clermoniotides*	Hawai'i	8/3/11	19.21	-155.60	Ka'u Preserve Kaiholena	Richard Pender
H4	7339 7940	*peleana ssp. singulariflora*	Hawai'i	6/7/13	20.18	-155.80	Kohala Mts.	Richard Pender
K1	5089 5090	*fauriei*	Kaua'i	9/6/05	22.09	-159.59	Alaka'i Swamp	Clifford Morden
O1	6809 6810	*kakeana*	O'ahu	6/17/11	21.34	-157.82	Mt. Tantalus, Ko'olau Mts.	Richard Pender
O2	7008 7009	*oblongifolia ssp. oblongifolia*	O'ahu	3/15/12	21.41	-158.10	Palikea, Waianae Mts.	Richard Pender
Ma1	6875 6876	*arborescens*	Maui	7/5/11	20.82	-156.28	Waihiei Makawao Forest Reserve	Hank Oppenheimer
Ma2	6831 6832	*kakeana*	Maui	7/14/11	20.80	-156.23	Makawao Forest Reserve	Richard Pender

557
558
559
560
561
562
563
564
565
566
567
568
569
570
571

572 **Table S2: PCR volumes and thermal cycler settings for each amplicon library of the first**
573 **amplicon PCR and second index PCR.**
574

	Amplicon PCR		Index PCR
Reagent	Concentration	Volume/rxn	Volume/rxn
DNA	Full	1.0	1.0
H2O	-	10.3	23.0
Q5 Mastermix	2x	12.5	25.0
Forward Primer	10μM	0.6	0.5
Reverse Primer	10μM	0.6	0.5
	PCR Volume	25.0	50.0
	Temperature (°C)	Time (seconds)	Time (seconds)
Initial Denature	98	120	120
Denature	98	10	15
Annealing	51/54	10	15
Extension	72	15	25
Final Extension	72	120	120
	PCR Cycles	22	22

575



576
577 **Figure S1: Fungal OTU accumulation curve for all *Clermontia* plant bank phyllosphere**
578 **samples using the rarefied OTU matrix with 95% confidence intervals shown in grey.**

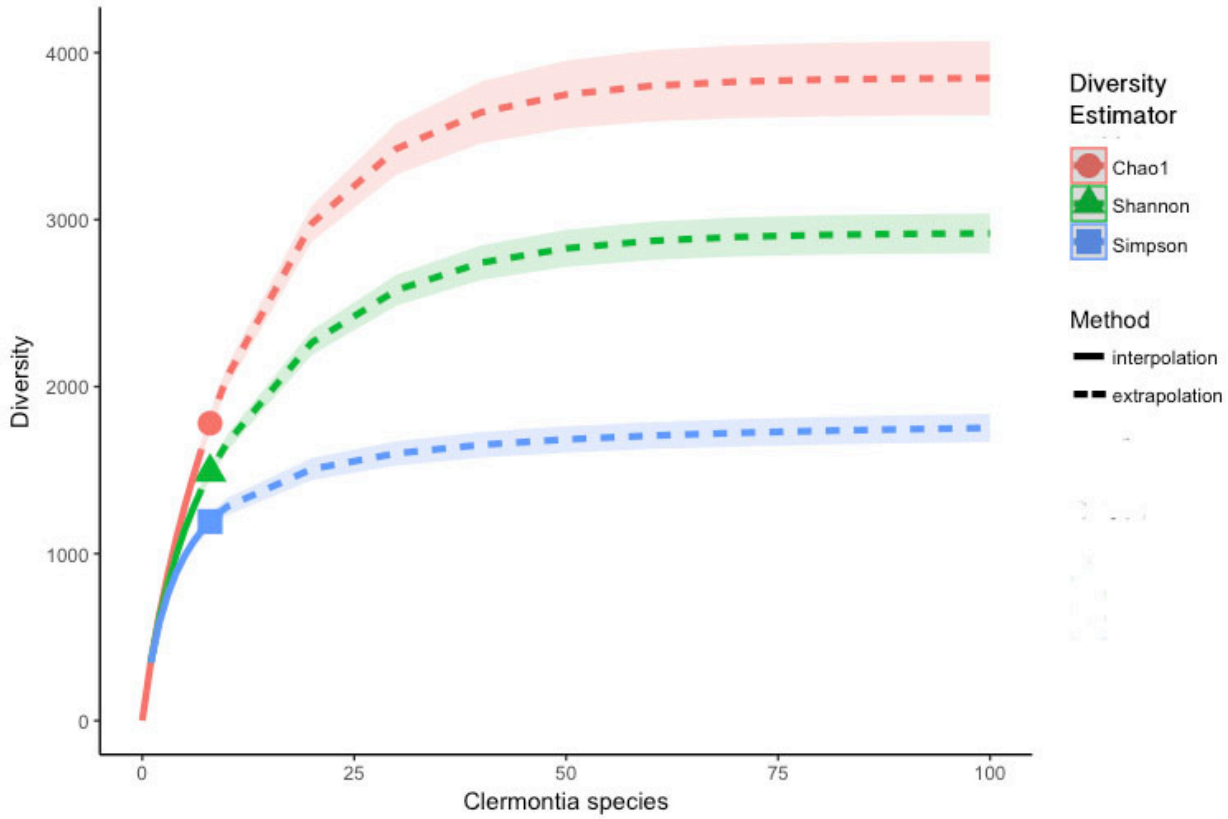
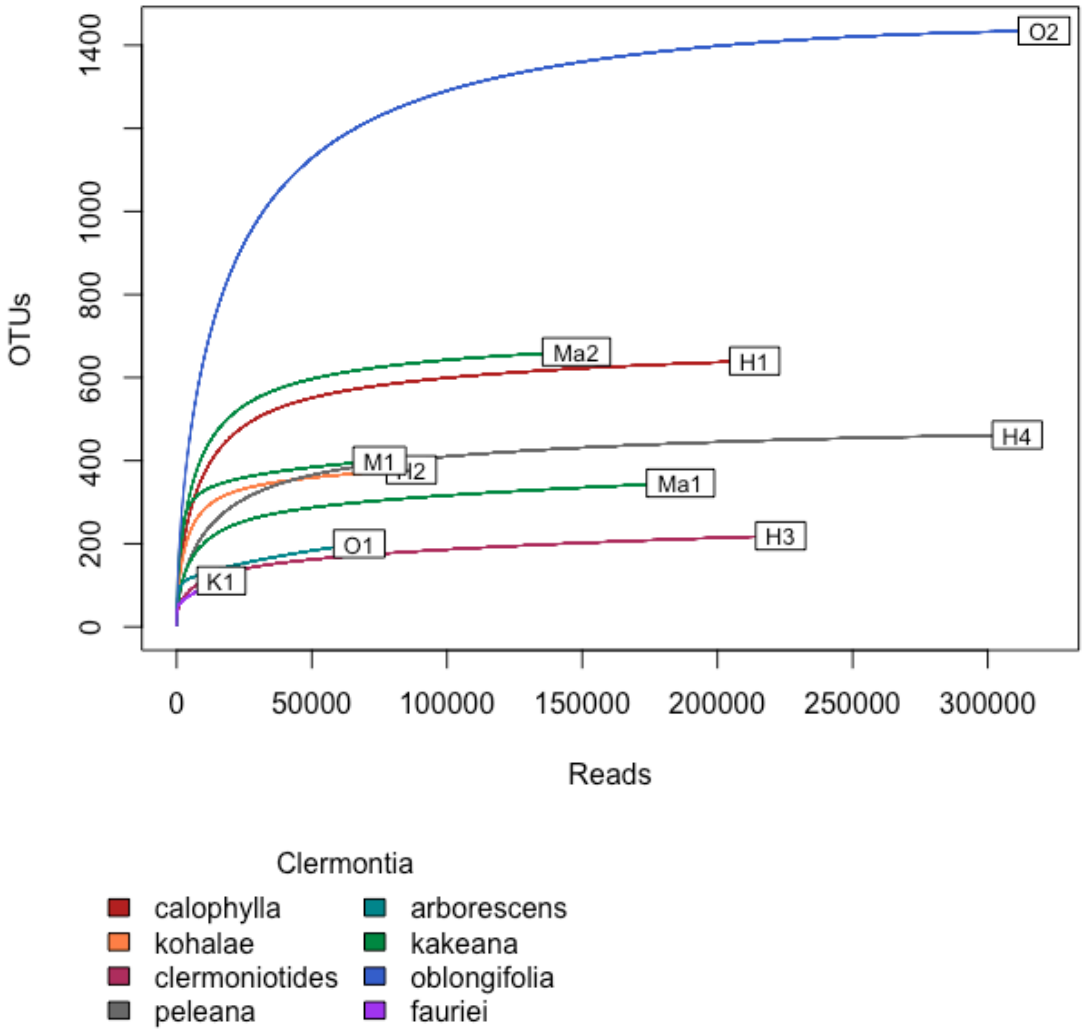


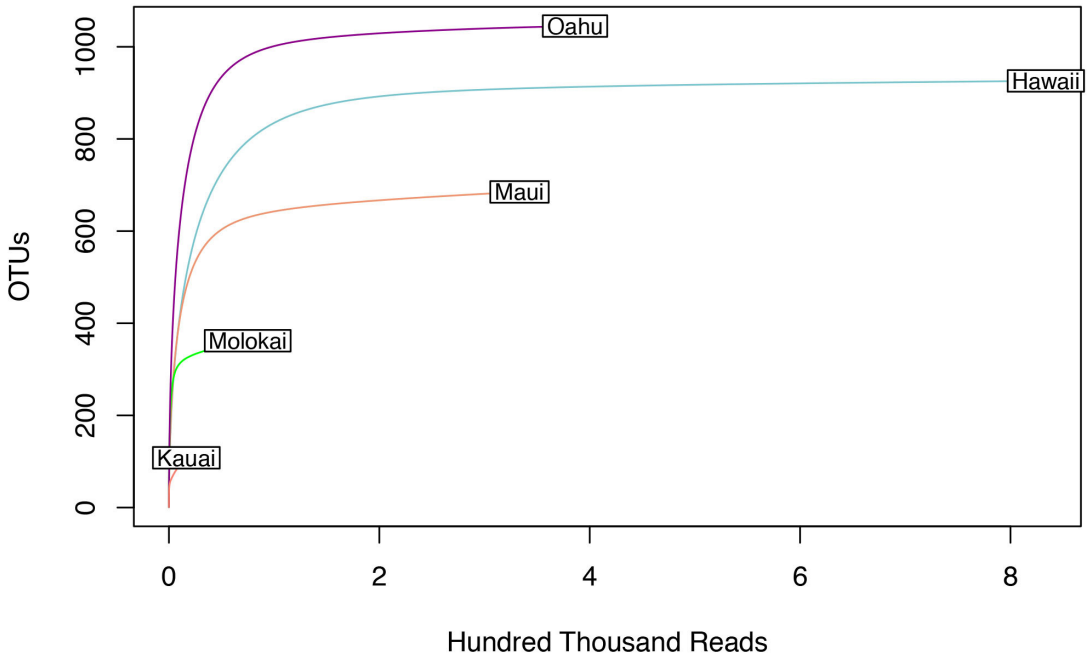
Figure S2: Species interpolation (solid lines) and extrapolation (dashed lines) curves for the eight *Clermontia* species sampled from the plant DNA bank using the rarefied OTU matrix. Three different diversity estimators were used (Chao1 richness, exponential of Shannon entropy, and inverse Simpson concentration indices) and are shown by the different colors with 95% confidence intervals shown by shading. Shapes represent observed plant bank species diversity calculations.



586

587

588 **Figure S3: Rarefaction curves for each *Clermontia* plant bank sample (non-rarefied),**
589 **fungal OTU accumulation over the corresponding number of DNA sequence reads, colors**
590 **display the different *Clermontia* species.**



591
592 **Figure S4: Rarefaction curves for *Clermontia* plant bank samples pooled by island (non-**
593 **rarefied), fungal OTU accumulation with the associated number of DNA sequences.**