data

MDPI

Data Descriptor

Illumina 16S rRNA Gene Sequencing Dataset of Bacterial Communities of Soil Associated with Ironwood Trees (*Casuarina equisetifolia*) in Guam

Tao Jin [1], Robert L. Schlub [2] and Claudia Husseneder [1,*]

[1] Department of Entomology, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA; jtao@agcenter.lsu.edu
[2] College of Natural & Applied Sciences, University of Guam, Cooperative Extension Service, Mangilao 96923, Guam; rlschlub@triton.uog.edu
* Correspondence: chusseneder@agcenter.lsu.edu

Abstract: Ironwood trees, which are of great importance for the economy and environment of tropical areas, were first discovered to suffer from a slow progressive dieback in Guam in 2002, later referred to as ironwood tree decline (IWTD). A variety of biotic factors have been shown to be associated with IWTD, including putative bacterial pathogens *Ralstonia solanacearum* and *Klebsiella* species (*K. variicola* and *K. oxytoca*), the fungus *Ganoderma australe*, and termites. Due to the soilborne nature of these pathogens, soil microbiomes have been suggested to be a significant factor influencing tree health. In this project, we sequenced the microbiome in the soil collected from the root region of healthy ironwood trees and those showing signs of IWTD to evaluate the association between the bacterial community in soil and IWTD. This dataset contains 4,782,728 raw sequencing reads present in soil samples collected from thirty-nine ironwood trees with varying scales of decline severity in Guam obtained via sequencing the V1–V3 region of the 16S rRNA gene on the Illumina NovaSeq (2 × 250 bp) platform. Sequences were taxonomically assigned in QIIME2 using the SILVA 132 database. Firmicutes and Actinobacteria were the most dominant phyla in soil. Differences in soil microbiomes were detected between limestone and sand soil parent materials. No putative plant pathogens of the genera *Ralstonia* or *Klebsiella* were found in the samples. Bacterial diversity was not linked to parameters of IWTD. The dataset has been made publicly available through NCBI GenBank under BioProject ID PRJNA883256. This dataset can be used to compare the bacterial taxa present in soil associated with ironwood trees in Guam to bacteria communities of other geographical locations to identify microbial signatures of IWTD. In addition, this dataset can also be used to investigate the relationship between soil microbiomes and the microbiomes of ironwood trees as well as those of the termites which attack ironwood trees.

Dataset: Repository name: National Center for Biotechnology Information. Data identification number: BioProject ID PRJNA883256 [Accession Numbers: SRX20942017-SRX20942059]. Direct URL to data: https://www.ncbi.nlm.nih.gov/sra?LinkName=bioproject_sra_all&from_uid=883256 (accessed on 1 September 2023).

Keywords: soil; bacteria; diversity; taxonomic index; metataxanomics; amplicon sequencing; 16S

Citation: Jin, T.; Schlub, R.L.; Husseneder, C. Illumina 16S rRNA Gene Sequencing Dataset of Bacterial Communities of Soil Associated with Ironwood Trees (*Casuarina equisetifolia*) in Guam. *Data* **2024**, *9*, 54. https://doi.org/10.3390/data9040054

Academic Editor: Sujeevan Ratnasingham

Received: 29 January 2024
Revised: 20 March 2024
Accepted: 1 April 2024
Published: 7 April 2024

1. Summary

The decline of ironwood trees (*Casuarina equisetifolia*) in Guam has been investigated extensively since it was first detected in 2002 [1]. A complex of factors was found to be associated with IWTD including the presence of the wilt-causing bacteria *Ralstonia solanacearum*, wetwood bacteria of the genus *Klebsiella*, the fungus *Ganoderma australe*, as well as attacks by numerous termite species [2,3]. Soilborne pathogens and soil microbiome composition have been previously reported to be linked to the decline of trees other

than ironwood [4,5]. This project aims at the assessment of bacterial compositions in soil collected from the root region of healthy and declining ironwood trees in Guam using metataxanomics. Bacterial communities were described from soil samples that were collected from thirty-nine ironwood trees at seventeen different locations on the island of Guam. The climate of Guam is tropical marine. The collection dates were in February and March, 2022, which is the dry season of the island with one-third of annual rainfall of 84–116 inches [6]. The temperature range during collection was 76–87 °F. The collection sites were spread over the entire island, which encompasses five distinct soil orders. Entisols are prevalent in the soil type called Guam series on Northern Guam, which is shallow soil covering coral limestone, while a combination of Oxisols, Mollisols, and Alfisols characterizes the volcanic terrain in Southern Guam, composing Pulantat, Atate and Akina series. Inceptisols are typically formed in the bottom lands of southern Guam [7]. The soil moistures and parent materials (PM) are listed in Table 1. Briefly, the soil of collection locations on limestone consists of clay, sand, and silt, with a cation exchange capacity (CEC) of 20–45 meq/100 g, pH values of 6.6–7.8, and organic matters (OM) of 8–15 Pct; the soil belonging to the sand PM category is dominant in sand and alkaline, with the lowest CEC < 10 meq/100 g and the lowest OM < 5 Pct; and the soil belonging to the tuff PM category is rich in clay and acidic, with a CEC of 10–50 meq/100 g, and a OM of 5–10 Pct [8]. Bacterial community composition and diversity were determined via amplicon sequencing using the V1–V3 region of the 16S rRNA gene on the Illumina NovaSeq (2 × 250 bp) platform following the Illumina Nextera Dilute library protocol and evaluated using QIIME 2 version 2022.2 with SILVA 132 as the taxonomic reference database and iNEXT. The dataset in the present study was generated to test whether the soil collected from under ironwood trees in Guam contained putative pathogenic bacteria, such as the genera *Ralstonia* and *Klebsiella*, and whether factors, such as *Ralstonia* presence and decline severity of ironwood trees had an effect on bacterial diversity.

Table 1. Metadata describing the locations where 39 soil samples used for bacterial diversity analysis were collected.

Sample ID	Location	Tree GPS	Altitude Classification (m)	Parent Material	Site Management	Percentage of Soil Moisture (%)	Depth of Soil Cores Taken (cm)
22-122	UOG, Mangilao	13.43109, 144.80084	low (67)	limestone	none	25.57	4, 7, 4, 10, 10, 4
22-126	Mangilao Golf Course	13.47111, 144.84528	high (129)	limestone	high	16.26	10, 10, 10, 10, 10
22-127	UOG Yigo Station	13.53309, 144.87161	high (178)	limestone	moderate	34.33	10, 10, 10, 10, 10
22-128	UOG Yigo Station	13.53327, 144.87158	High (144)	limestone	high	40	10, 10, 10, 10, 10
22-129	UOG Yigo Station	13.53297, 144.87364	high (142)	limestone	moderate	29.87	10, 10, 10, 10, 10
22-131	UOG Yigo Station	13.53231, 144.87267	high (173)	limestone	moderate	38.99	10, 10, 10, 10, 10
22-133	Mangilao Golf Course	13.47082, 144.84503	high (127)	limestone	high	39.19	10, 10, 10, 10, 10
22-137	Watson's Farm, Yigo	13.56456, 144.87746	high (168)	limestone	moderate	34.46	10, 6, 8, 10, 4
22-139	Watson's Farm, Yigo	13.56660, 144.87416	high (161)	limestone	moderate	40.84	8, 10, 10, 6, 10
22-140	Watson's Farm, Yigo	13.56598, 144.87462	high (164)	limestone	moderate	31.2	10, 10, 10, 10, 10
22-141	Watson's Farm, Yigo	13.56583, 144.87688	high (170)	limestone	moderate	33.83	8, 10, 10, 10, 10
22-143	Watson's Farm, Yigo	13.56707, 144.87654	high (171)	limestone	none	49.4	8, 8, 10, 8, 4
22-145	Watson's Farm, Yigo	13.56692, 144.87740	high (163)	limestone	moderate	38.84	10, 10, 10, 8, 10
22-149	UOG Ija Station	13.26595, 144.71707	high (110)	tuff	moderate	50.52	10, 10, 10, 10, 10
22-155	AAFB	13.56314, 144.93079	high (160)	limestone	moderate	28.31	4, 8, 10, 8, 10
22-156	AAFB	13.56830, 144.93179	high (164)	limestone	high	56.23	10, 10, 10, 10, 9.5
22-157	AAFB	13.56132, 144.93056	high (155)	limestone	moderate	54.76	8, 10, 8, 10, 10
22-158	AAFB	13.55782, 144.93009	high (155)	limestone	high	38.51	10, 10, 10, 10, 6
22-159	Paseo Park, Hagatna	13.47956, 144.75429	low (6)	limestone	high	7.51	4, 4, 4, 8, 4
22-160	Governor's Complex, Aniquia	13.47870, 144.73200	low (6)	tuff	high	17.76	6, 10, 6, 10, 10
22-161	Apaca Point, Agat	13.40239, 144.66307	low (6)	tuff	moderate	52.65	10, 10, 10, 10, 10
22-164	Windward Hills Golf Course	13.37788, 144.74162	high (126)	tuff	high	35.97	6, 10, 8, 8, 10
22-165	Windward Hills Golf Course	13.37650, 144.73789	high (115)	tuff	high	36.79	8, 8, 10, 6, 4
22-166	Country Club of the Pacific	13.37163, 144.76817	low (26)	tuff	high	17.59	4, 8, 6, 8, 8, 4
22-168	Duenas Beach	13.25978, 144.73688	low (11)	sand	moderate	7.24	10, 10, 10, 10, 10
22-169	Duenas Beach	13.25963, 144.73735	low (12)	sand	moderate	15.08	10, 10, 10, 10, 10
22-170	Ysrael Beach	13.24797, 144.72690	low (5)	sand	none	11.23	4, 10, 10, 10, 10
22-171	Ysrael Beach	13.24741, 144.72708	low (7)	sand	none	22.46	8, 10, 8, 8, 4
22-173	Tarague Beach	13.62518, 144.89525	low (9)	sand	high	7.87	10, 10, 10, 10, 10
22-174	Tarague Beach	13.62340, 144.89664	low (21)	sand	high	6.36	10, 10, 10, 10, 8
22-187	Ritidian	13.64889, 144.85289	low (12)	sand	moderate	17.69	10, 10, 10, 10, 10
22-188	Watson's Farm, Yigo	13.56741, 144.87413	high (163)	limestone	none	53.9	4, 6, 6, 8, 10
22-189	Watson's Farm, Yigo	13.56864, 144.87701	high (173)	limestone	none	43.11	4, 4, 6, 6, 6, 10
22-190	Watson's Farm, Yigo	13.56594, 144.87816	high (168)	limestone	none	39.51	2, 4, 4, 4, 4, 8
22-191	Watson's Farm, Yigo	13.56553, 144.87749	high (159)	limestone	moderate	20.74	10, 10, 10, 10, 6
22-192	Watson's Farm, Yigo	13.56528, 144.87704	high (159)	limestone	moderate	31.58	6, 8, 10, 10, 10
22-208	Cocos Island	13.23476, 144.64574	low (20)	sand	none	24.59	10, 4, 8, 10, 10
22-209	Cocos Island	13.23701, 144.65102	low (12)	sand	none	24.98	2, 4, 10, 4, 10, 6
22-223	Sagan Kotturan Chamoru	13.50311, 144.78416	low (45)	limestone	none	8.8	10, 10, 10, 10, 8

2. Data Description

The V1–V3 variable region of the 16S rRNA gene was amplified using the Illumina NovaSeq platform to identify the bacterial taxa and diversity present in soil around ironwood trees. The links and accession numbers to the fastq files in this dataset are provided in Table 2. The bacterial sequences present in the 39 samples were taxonomically assigned using the Quantitative Insights into Microbial Ecology (QIIME2 version 2022.2) pipeline [9].

Table 2. Accession numbers and links for raw fastq sequences of soil samples collected from thirty-nine ironwood trees in Guam.

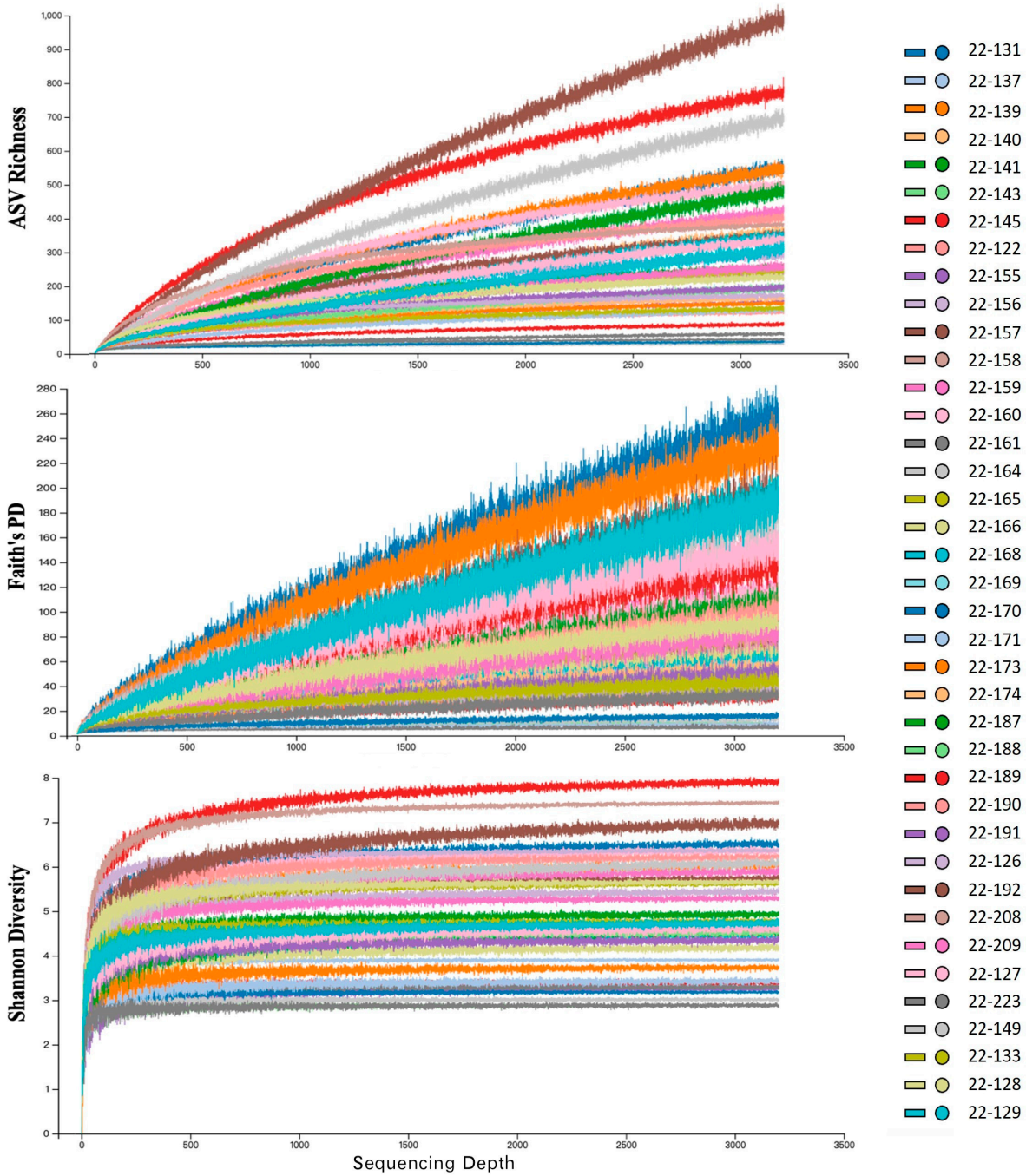
Sample Name	SRA Number	Accession Link
22-122	SRR25194844	https://www.ncbi.nlm.nih.gov/sra/SRX20942017[accn] (accessed on 1 September 2023)
22-126	SRR25194843	https://www.ncbi.nlm.nih.gov/sra/SRX20942018[accn] (accessed on 1 September 2023)
22-127	SRR25194832	https://www.ncbi.nlm.nih.gov/sra/SRX20942029[accn] (accessed on 1 September 2023)
22-128	SRR25194821	https://www.ncbi.nlm.nih.gov/sra/SRX20942040[accn] (accessed on 1 September 2023)
22-129	SRR25194810	https://www.ncbi.nlm.nih.gov/sra/SRX20942051[accn] (accessed on 1 September 2023)
22-131	SRR25194806	https://www.ncbi.nlm.nih.gov/sra/SRX20942055[accn] (accessed on 1 September 2023)
22-137	SRR25194805	https://www.ncbi.nlm.nih.gov/sra/SRX20942056[accn] (accessed on 1 September 2023)
22-139	SRR25194803	https://www.ncbi.nlm.nih.gov/sra/SRX20942058[accn] (accessed on 1 September 2023)
22-140	SRR25194802	https://www.ncbi.nlm.nih.gov/sra/SRX20942059[accn] (accessed on 1 September 2023)
22-141	SRR25194842	https://www.ncbi.nlm.nih.gov/sra/SRX20942019[accn] (accessed on 1 September 2023)
22-143	SRR25194841	https://www.ncbi.nlm.nih.gov/sra/SRX20942020[accn] (accessed on 1 September 2023)
22-145	SRR25194840	https://www.ncbi.nlm.nih.gov/sra/SRX20942021[accn] (accessed on 1 September 2023)
22-155	SRR25194838	https://www.ncbi.nlm.nih.gov/sra/SRX20942023[accn] (accessed on 1 September 2023)
22-156	SRR25194837	https://www.ncbi.nlm.nih.gov/sra/SRX20942024[accn] (accessed on 1 September 2023)
22-157	SRR25194836	https://www.ncbi.nlm.nih.gov/sra/SRX20942025[accn] (accessed on 1 September 2023)
22-158	SRR25194835	https://www.ncbi.nlm.nih.gov/sra/SRX20942026[accn] (accessed on 1 September 2023)
22-159	SRR25194834	https://www.ncbi.nlm.nih.gov/sra/SRX20942027[accn] (accessed on 1 September 2023)
22-160	SRR25194833	https://www.ncbi.nlm.nih.gov/sra/SRX20942028[accn] (accessed on 1 September 2023)
22-161	SRR25194831	https://www.ncbi.nlm.nih.gov/sra/SRX20942030[accn] (accessed on 1 September 2023)
22-164	SRR25194830	https://www.ncbi.nlm.nih.gov/sra/SRX20942031[accn] (accessed on 1 September 2023)
22-165	SRR25194829	https://www.ncbi.nlm.nih.gov/sra/SRX20942032[accn] (accessed on 1 September 2023)
22-166	SRR25194828	https://www.ncbi.nlm.nih.gov/sra/SRX20942033[accn] (accessed on 1 September 2023)
22-168	SRR25194827	https://www.ncbi.nlm.nih.gov/sra/SRX20942034[accn] (accessed on 1 September 2023)
22-169	SRR25194826	https://www.ncbi.nlm.nih.gov/sra/SRX20942035[accn] (accessed on 1 September 2023)

Table 2. *Cont.*

Sample Name	SRA Number	Accession Link
22-170	SRR25194825	https://www.ncbi.nlm.nih.gov/sra/SRX20942036[accn] (accessed on 1 September 2023)
22-171	SRR25194824	https://www.ncbi.nlm.nih.gov/sra/SRX20942037[accn] (accessed on 1 September 2023)
22-173	SRR25194823	https://www.ncbi.nlm.nih.gov/sra/SRX20942038[accn] (accessed on 1 September 2023)
22-174	SRR25194822	https://www.ncbi.nlm.nih.gov/sra/SRX20942039[accn] (accessed on 1 September 2023)
22-187	SRR25194819	https://www.ncbi.nlm.nih.gov/sra/SRX20942042[accn] (accessed on 1 September 2023)
22-188	SRR25194818	https://www.ncbi.nlm.nih.gov/sra/SRX20942043[accn] (accessed on 1 September 2023)
22-189	SRR25194817	https://www.ncbi.nlm.nih.gov/sra/SRX20942044[accn] (accessed on 1 September 2023)
22-190	SRR25194816	https://www.ncbi.nlm.nih.gov/sra/SRX20942045[accn] (accessed on 1 September 2023)
22-191	SRR25194815	https://www.ncbi.nlm.nih.gov/sra/SRX20942046[accn] (accessed on 1 September 2023)
22-192	SRR25194814	https://www.ncbi.nlm.nih.gov/sra/SRX20942047[accn] (accessed on 1 September 2023)
22-208	SRR25194813	https://www.ncbi.nlm.nih.gov/sra/SRX20942048[accn] (accessed on 1 September 2023)
22-209	SRR25194812	https://www.ncbi.nlm.nih.gov/sra/SRX20942049[accn] (accessed on 1 September 2023)
22-223	SRR25194811	https://www.ncbi.nlm.nih.gov/sra/SRX20942050[accn] (accessed on 1 September 2023)
22-149	SRR25194808	https://www.ncbi.nlm.nih.gov/sra/SRX20942053[accn] (accessed on 1 September 2023)
22-133	SRR25194807	https://www.ncbi.nlm.nih.gov/sra/SRX20942054[accn] (accessed on 1 September 2023)

A total of 4,782,728 raw sequencing reads were obtained with 27F as the forward primer and 519Rmod and 519Rmodbio as reverse primers [10] to capture a broad array of bacterial diversity across soil samples collected from the 39 ironwood trees in Guam. A total of 3,337,420 sequence reads represented by 28,563 unique Amplicon Sequence Variants (ASVs) remained after quality-filtering using DADA2 and merging of amplicons obtained with both reverse primer sets for each sample. The ASVs with no taxonomical assignment, i.e., sequences with less than 99% identity to references in the SILVA 132 database, were removed, resulting in 200,499 sequence reads and 2303 ASVs with taxonomic assignment.

Alpha-rarefaction in QIIME2 and iNEXT was used to assess whether sequencing depth, sample numbers, and coverage were sufficient to capture the majority of bacterial diversity present in the soil samples. The sequence-depth-based rarefaction curves for ASV richness and Faith's phylogenetic distance (PD) (Figure 1a) showed that diversity in many samples started to level off but did not reach an asymptote in some of the samples, indicating that at least in highly diverse samples, the soil microbiome richness and phylogenetic diversity were not completely captured. In contrast, Shannon diversity plateaued at a sequencing depth of around 1000 reads indicating that overall bacteria diversity was sufficiently captured, and the continued increase in richness and phylogenetic distance was likely based on rare species as it would be expected in microbiome-rich environments like soil.



(**a**)

Figure 1. *Cont.*

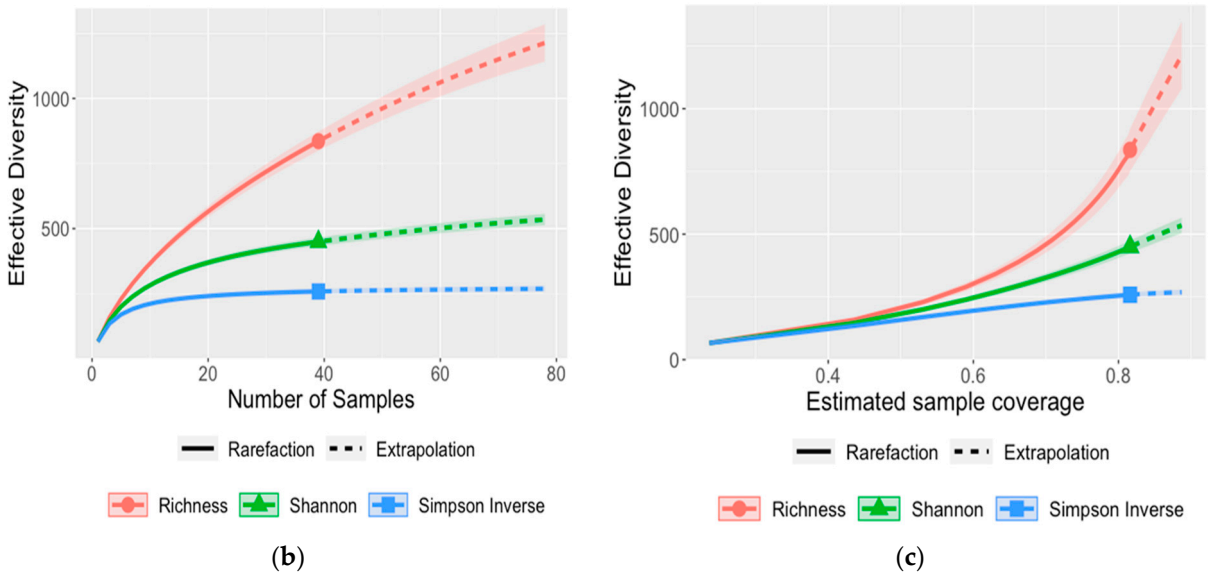


Figure 1. The rarefaction curves based on (**a**) sequencing depth of each sample depict the ASV richness, Faith's phylogenetic distance, and Shannon diversity indices of bacteria diversity in 39 soil samples; (**b**) the number of samples depict effective bacterial diversity for richness, Shannon and Simpson Inverse metrics; and (**c**) estimated sample coverage depict effective diversity for the same three metrics. Solid lines were drawn up to the actual sample size; dashed lines represent extrapolation to twice the sample size. Rarefaction was conducted across the entire bacteria diversity (including ASVs without taxonomical assignment). Sequencing depth was assessed by the number of reads.

In addition, rarefaction curves were created using sample-based and coverage-based approaches (Figure 1b,c) to depict the relationship between effective diversity and the number of samples or the estimated sample coverage. Effective diversity considers both the relative abundance and richness of the data and is quantified using Hill numbers (q) with q = 0, 1, and 2 representing ASV richness, Shannon diversity, and Simpson Inverse, respectively. The sample- and coverage-based rarefaction curves were extrapolated to twice the sample size to allow for the computation of effective diversity values beyond the original sample size.

The sample-based rarefaction curves (Figure 1b) for Shannon diversity and Simpson Inverse mostly had levelled out at the actual sample size reaching an effective diversity of approximately 500 and 250, respectively. Extrapolation of these two curves did not result in significant increases in these values. On the other hand, the rarefaction curve for ASV richness continued to rise even after extrapolation. However, this increase in ASV richness seemed to be driven by rare ASVs since there was no corresponding increase in Shannon and Simpson Inverse. The 39 samples in this dataset provided over 80% coverage and doubling the sample size would increase the coverage to 95% (Figure 1c).

The taxonomic assignment of ASVs revealed seventeen different phyla in soil collected around ironwood trees in Guam. The dominant phyla and their relative abundances were as follows: Firmicutes (50.67%), Actinobacteria (31.52%), Proteobacteria (10.57%), Acidobacteria (3.96%), and others (3.28%) (Figure 2). The twenty most abundant ASVs with the highest number of reads were assigned to phyla Firmicutes, Actinobacteria, Proteobacteria, and Acidobacteria (Table 3). The ASV with the highest number of reads belonged to the genus Bacillus, which is typically predominant in soil (Table 3). None of the ASVs from the 39 samples were assigned to putative pathogens associated with IWTD in Guam such as *Ralstonia* spp. and *Klebsiella* spp. However, the positive controls that were spiked with *R. solanacearum* did show *Ralstonia* presence, confirming that DNA extraction methods and primers were appropriate for detection.

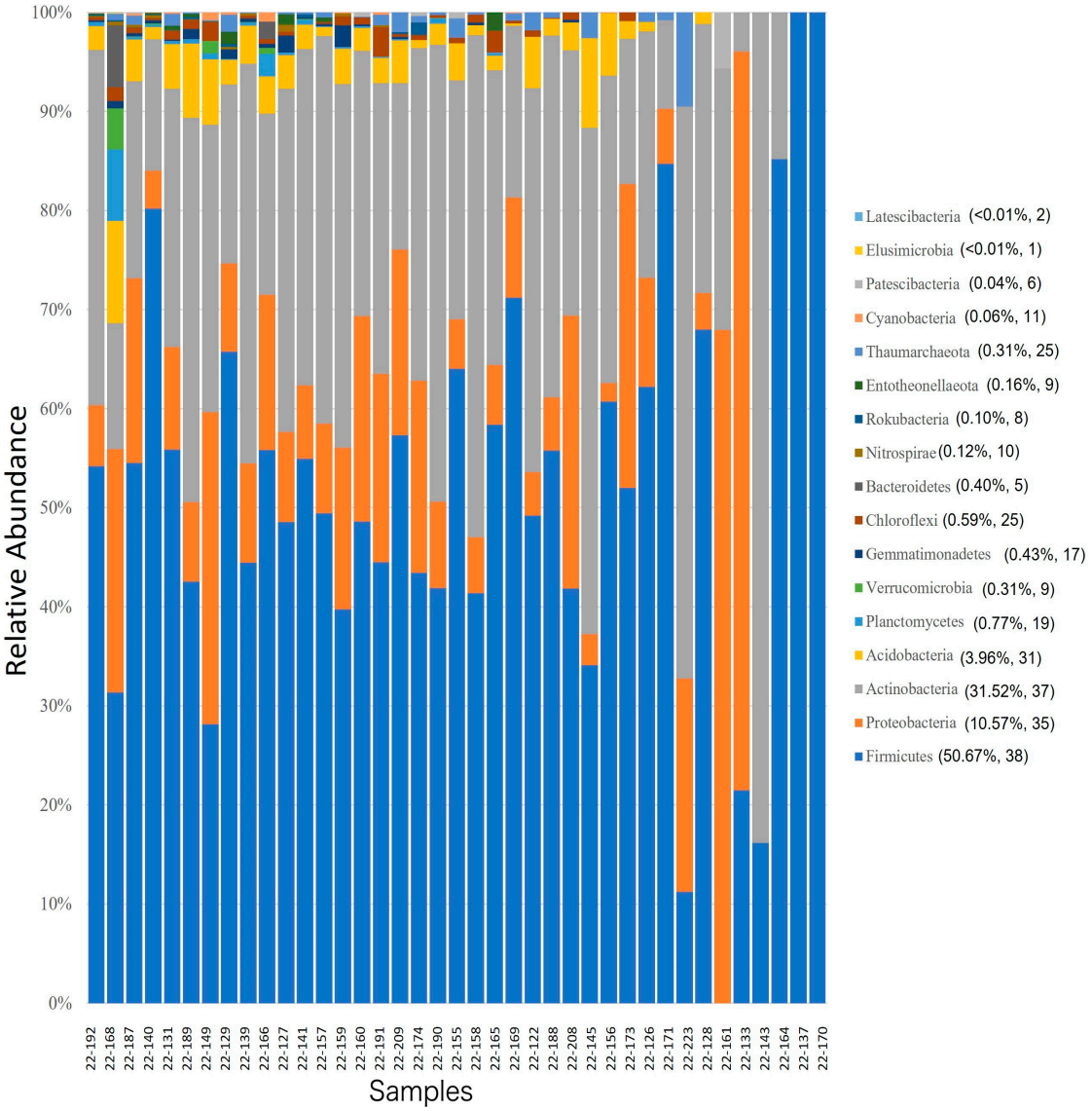


Figure 2. Relative abundance of bacterial phyla associated with 39 samples of soil collected from ironwood trees in Guam. The legend contains each phylum with its percent abundance and number of samples in which that phylum was found.

Four indices, i.e., ASV richness, Pielou's evenness, Shannon diversity, and Faith's PD, were used to perform alpha-diversity analysis on the dataset after rarefaction to a common sequencing depth of 3200 per sample. Kruskal–Wallis ANOVA with Benjamini–Hochberg correction was used to determine the group significance for factors with categorical data: location, presence of *Ralstonia* in trees as measured with the Agdia Strip test, tree decline severity as measured by visual inspection of fullness and dieback of branches (symptomless, slightly damaged, distinctly damaged, heavily damaged, and nearly dead), altitude classification (high $\geq$ 100 m and low < 100 m above sea level), parent material (limestone, sand, and tuff), and site management (no maintenance, moderately managed, and highly managed) [10,11]. Spearman rank tests were used to assess the correlations for factors with numerical data including altitude, percentage of dead trees in plot, and percentage of trees with termites in plot (for a detailed description of these factors, see [10,11]). However, no significant effects on alpha-diversity were detected for any of these factors.

Table 3. The 20 most abundant ASVs in soil samples according to total numbers of reads with their assignment in SILVA and the number of samples in which they were detected.

Phylum	Order	Lowest SILVA Assignment	Number of Reads	Number of Samples	Average Number of Reads per Sample	Standard Deviation
Firmicutes	Bacillales	*Bacillus*	25,369	34	650	1420.2
Firmicutes	Bacillales	*Bacillus*	24,899	33	638	1473.3
Actinobacteria	Gaiellales	Gaiellales	15,500	33	397	1110.6
Actinobacteria	Solirubrobacterales	Solirubrobacterales	9381	29	241	703.1
Firmicutes	Bacillales	*Bacillus*	8560	31	219	434.7
Firmicutes	Bacillales	*Bacillus abyssalis*	6286	17	161	557.7
Firmicutes	Bacillales	*Lysinibacillus*	6281	15	161	457.8
Proteobacteria	Rhizobiales	Xanthobacteraceae	3607	28	92	170.2
Firmicutes	Bacillales	*Bacillus* sp. A-10	3312	29	85	105.9
Actinobacteria	Propionibacteriales	*Microlunatus*	3197	21	82	221.2
Actinobacteria	Unknown	Thermoleophilia	3195	19	82	228.3
Firmicutes	Bacillales	*Cohnella*	2959	19	76	187.6
Actinobacteria	Streptomycetales	*Streptomyces*	2841	26	73	120.3
Actinobacteria	Gaiellales	*Gaiella*	2719	23	70	183.4
Proteobacteria	Rhizobiales	*Pedomicrobium*	2503	23	64	118.3
Actinobacteria	Micromonosporales	*Micromonospora*	2496	20	64	144.2
Firmicutes	Bacillales	*Paenibacillus*	2429	20	62	156.4
Actinobacteria	Micromonosporales	Micromonosporaceae	2401	15	62	193.1
Acidobacteria	Unknown	Subgroup 6	2214	20	57	161.5
Actinobacteria	Microtrichales	Microtrichales	2022	17	52	160.6

For beta-diversity, single-factorial Permutational Multivariate Analysis of Variance (PERMANOVA) at 1000 permutations based on both of the weighted and unweighted Unifrac distance metrics [12,13] was used to evaluate the differentiation of the microbial composition among soil samples grouped by the categorical factors listed above. The weighted Unifrac distance metric takes both abundance and phylogenetic distance into account; however, no significant influences on bacterial community differentiation were found for any of these factors. For unweighted Unifrac, which only considers presence/absence and phylogenetic distance of ASVs, altitude classification (p = 0.00999, pseudo-F = 1.26, n = 39) and parent material (p = 0.04995) showed significant effects on the beta-diversity of the bacterial communities. The effects of the parent material on the soil microbiome were driven by significant differences between limestone and sand (p = 0.047952, pseudo-F = 1.16, n = 30); however, no significant differences in the microbiota were observed in other factors.

3. Methods

3.1. Sample Collection and Processing

Soil samples were collected by the team from the University of Guam from February to March in 2022 from under 39 different ironwood trees in Guam using heat-sterilized equipment (Tables 1 and 4). The trees with different scales of IWTD severity (symptomless, slightly damaged, distinctly damaged, heavily damaged, and nearly dead) were distributed across seventeen different locations. Approximately 100 mL of soil was collected from 5 to 6 soil cores around each tree. Samples were collected between 1 and 1.5 m from the base of the tree at a depth of 2–10 cm. The cores collected from the same tree were mixed and an aliquot of the soil was transferred to a 15 mL plastic ultra-high-performance centrifuge tube containing 8 mL of 95% ETOH until a volume of 11 mL was reached. In addition, 10 g of six soil samples were spiked with ooze from slices of tissue from trees marked positive for *Ralstonia solanacearum* species complex by Agdia Strip tests and also preserved in 95% ethanol for use as positive control.

Table 4. Metadata indicating the health of the trees at the sample location. Note that some metadata were not available for all samples (NA = not applicable).

Sample ID	Presence of *Ralstonia* [1]	Tree Health Ranking	Tree Decline Severity	Percentage of Dead Trees in Plot (%)	Percentage of Trees with Termites in Plot (%)	Plot Average Health
22-122	+	sick	nearly dead	13.33	26.67	sick
22-126	−	healthy	symptomless	2.94	55.88	sick
22-127	−	healthy	symptomless	16.67	56.67	sick
22-128	+	sick	nearly dead	NA	NA	NA
22-129	−	healthy	slightly damaged	0	56.56	sick
22-131	+	sick	nearly dead	9.09	36.36	sick
22-133	+	sick	nearly dead	0	40	sick
22-137	−	healthy	slightly damaged	0	61.76	sick
22-139	−	healthy	symptomless	2.86	65.71	sick
22-140	−	sick	nearly dead	3.03	39.39	healthy
22-141	+	sick	heavily damaged	53.33	33.33	sick
22-143	−	healthy	symptomless	8.33	8.33	sick
22-145	+	sick	heavily damaged	38.1	23.81	sick
22-149	+	sick	nearly dead	58.33	20.83	sick
22-155	−	healthy	slightly damaged	14.29	71.43	sick
22-156	−	sick	heavily damaged	0	60	sick
22-157	−	healthy	slightly damaged	NA	NA	NA
22-158	−	sick	nearly dead	NA	NA	NA
22-159	−	healthy	symptomless	NA	NA	NA
22-160	−	healthy	symptomless	NA	NA	NA
22-161	−	healthy	symptomless	NA	NA	NA
22-164	+	healthy	slightly damaged	0	100	sick
22-165	+	sick	nearly dead	NA	NA	NA
22-166	−	sick	heavily damaged	NA	NA	NA
22-168	−	healthy	symptomless	10	60	sick
22-169	−	healthy	slightly damaged	12.94	38.82	sick
22-170	−	sick	nearly dead	7.59	16.46	healthy
22-171	−	healthy	slightly damaged	9.09	16.97	sick
22-173	+	sick	nearly dead	NA	NA	NA
22-174	−	healthy	symptomless	8.11	21.62	sick
22-187	−	healthy	slightly damaged	0	100	sick
22-188	−	healthy	slightly damaged	2.78	36.11	sick
22-189	−	healthy	slightly damaged	0	33.33	sick
22-190	−	healthy	slightly damaged	25.71	28.57	sick
22-191	+	sick	nearly dead	NA	NA	NA
22-192	+	sick	heavily damaged	NA	NA	NA
22-208	−	sick	distinctly damaged	7.61	21.74	sick
22-209	−	healthy	slightly damaged	7.56	25.58	sick
22-223	−	healthy	slightly damaged	8.57	51.43	sick

[1] Positive (+) means presence of *Ralstonia* detected and negative (−) means absence of *Ralstonia*.

3.2. DNA Extraction and Sequencing

Samples were sent to Louisiana State University Agricultural Center where DNA was extracted from 250 mg of dried soil of each sample using the DNeasy PowerSoil Pro kit (Qiagen, Germantown, MA, USA). All the procedures were conducted using sterile techniques and supplies to eliminate contamination. The DNA concentrations were quantified on a Qubit 4 Fluorometer (Thermo Fisher Scientific, Wilmington, DE, USA) with the Qubit dsDNA BR Assay Kit (Invitrogen, Thermo Fisher Scientific). The samples were then sent to the University of New Hampshire Hubbard Center for Genome Studies for subsequent sequencing. The V1–V3 region of the 16S rRNA gene of the bacterial DNA was amplified using the 27F as the forward primer and 519Rmod and 519Rmodbio as the reverse primers [14,15] and sequenced on the Illumina NovaSeq (2 × 250 bp) platform using Illumina Nextera Dilute library protocol with a spike-in of 1% Phi X (Illumina, San Diego, CA, USA).

3.3. Data Analysis

Bioinformatic analysis was performed using the QIIME2 pipeline version 2022.2 [9]. Demultiplexed fastq sequences were obtained after sequencing from the University of New Hampshire Hubbard Center for Genome Studies. The primers and chimera sequences were removed and denoising was performed using DADA2 [16]. Sequence reads of 251-nucleotide lengths were obtained. The amplicons obtained by the two sets of reverse primers were combined for each sample. Due to a lack of overlap between some of the forward and reverse reads, only forward sequences were used for the analysis. To assess diversity, sequence-, sample size- and coverage-based rarefaction curves were generated with the QIIME2 alpha-rarefaction plugin and the R package iNEXT version 3.0.0 (iNterpolation/EXTrapolation) [17]. Taxonomic assignments were performed by comparing the sequence reads or ASVs obtained through DADA2 quality filtering against the SILVA 132 reference database [18] using the consensus method in BLAST, with a 99% pairwise identity cutoff. ASVs that could not be assigned to any taxonomic group were excluded from the dataset prior to generating taxa bar plots, which illustrated the relative abundance of ASVs at the phylum level. Alpha-diversity analyses based on the four indices, ASV richness, Pielou's evenness, Shannon diversity, and Faith's PD, were performed using alpha-correlation and alpha-group-significance plugins in QIIME2. The beta-diversity was calculated using beta-group-significance method of qiime diversity plugin in QIIME2.

4. User Notes

- This dataset contributes to the investigation of the association between bacterial communities in soil and IWTD [4,19];
- Bacteria *Ralstonia solanacearum* and *Klebsiella* species (*K. variicola* and *K. oxytoca*) were previously isolated from diseased ironwood trees [2]. This dataset could be used to assess whether soilborne microbiomes are the source of these or other putative pathogens;
- This dataset describes bacterial communities present in soil collected from under ironwood trees with varying scales of decline severity in Guam. Microbiologists can use this dataset to compare the bacteria in soil with those found in other geographic regions that are affected by IWTD;
- These data can be utilized by microbiologists and plant pathologists to understand if there is an association between the microbiota of soil and the microbiota of the ironwood trees and the termites infesting those trees [10,11].

Author Contributions: C.H.: funding acquisition, conceptualization, supervision, writing—review and editing; R.L.S.: project administration, funding acquisition, sample collection, writing—review and editing; T.J.: formal analysis, writing—original draft, data analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2018-38640-28418 through the Western Sustainable Agriculture Research and Education program under project number SW19-906.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in National Center for Biotechnology Information (NCBI) at https://www.ncbi.nlm.nih.gov/sra?LinkName=bioproject_sra_all&from_uid=883256 (accessed on 1 September 2023).

Acknowledgments: We express our thanks to the lab members at the University of Guam for their assistance in gathering the soil samples, as well as Garima Setia and Junyan Chen for their help in the bioinformatics analysis. We also express thanks to Kelley Thomas and his team from the University of New Hampshire Hubbard Center for Genomic Studies for their support in DNA sequencing and for providing the server for data analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. University of Guam Cooperative Extension Service, Gago. Guam Ironwood Tree, Casuarina equisetifolia Past, Present, Future. Available online: https://projects.sare.org/wp-content/uploads/Guam-Ironwood-Manual_8-26-22.pdf (accessed on 26 August 2022).
2. Ayin, C.M.; Schlub, R.L.; Yasuhara-Bell, J.; Alvarez, A.M. Identification and characterization of bacteria associated with decline of ironwood (*Casuarina equisetifolia*) in Guam. *Australas. Plant Pathol.* **2015**, *44*, 225–234. [CrossRef]
3. Park, J.; Husseneder, C.; Schlub, R.L. Morphological and molecular species identification of termites attacking Ironwood trees, *Casuarina equisetifolia* (Fagales: Casuarinaceae), in Guam. *J. Econ. Entomol.* **2019**, *112*, 1902–1911. [CrossRef] [PubMed]
4. Pinho, D.; Barroso, C.; Froufe, H.; Brown, N.; Vanguelova, E.; Egas, C.; Denman, S. Linking tree health, rhizosphere physicochemical properties, and microbiome in acute oak decline. *Forests* **2020**, *11*, 1153. [CrossRef]
5. Ruiz-Gómez, F.J.; Pérez-de-Luque, A.; Navarro-Cerrillo, R.M. The involvement of Phytophthora root rot and drought stress in holm oak decline: From ecophysiology to microbiome influence. *Curr. For. Rep.* **2019**, *5*, 251–266. [CrossRef]
6. Gingerich, S.B.; Keener, V.; Finucane, M.L. *Climate Trends and Projections for Guam*; East West Center: Honolulu, HI, USA, 2015. Available online: https://pubs.usgs.gov/publication/70145880 (accessed on 13 March 2024).
7. Deenik, J. Soils of Guam. In *Guam Grazing and Livestock Management Workshop*; University of Hawaii: Honolulu, HI, USA, 2010. Available online: https://manoa.hawaii.edu/ctahr/tpalm/pdfs-marianas/pdfs/vol_one/2_soils_management/soils_guam_2010_notes.pdf (accessed on 13 March 2024).
8. Web Soil Survey. Available online: https://websoilsurvey.nrcs.usda.gov/app/ (accessed on 13 March 2024).
9. Estaki, M.; Jiang, L.; Bokulich, N.A.; McDonald, D.; Gonzalez, A.; Kosciolek, T.; Martino, C.; Zhu, Q.; Birmingham, A.; Vazquez-Baeza, Y. QIIME 2 enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. *Curr. Protoc. Bioinform.* **2020**, *70*, e100. [CrossRef] [PubMed]
10. Setia, G.; Schlub, R.L.; Husseneder, C. Next-generation sequencing dataset of bacterial communities of *Microcerotermes crassus* workers associated with Ironwood trees (*Casuarina equisetifolia*) in Guam. *Data Brief* **2023**, *48*, 109286. [CrossRef] [PubMed]
11. Setia, G.; Chen, J.; Schlub, R.L.; Husseneder, C. Taxonomic profiling of *Nasutitermes takasagoensis* microbiota to investigate the role of termites as vectors of bacteria linked to ironwood tree decline in Guam. *PLoS ONE* **2023**, *18*, e0296081. [CrossRef] [PubMed]
12. Anderson, M.J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **2001**, *26*, 32–46. [CrossRef]
13. Lozupone, C.; Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **2005**, *71*, 8228–8235. [CrossRef] [PubMed]
14. Lane, D.J.; Pace, B.; Olsen, G.J.; Stahl, D.A.; Sogin, M.L.; Pace, N.R. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6955–6959. [CrossRef] [PubMed]
15. Lane, D.J. 16S/23S rRNA Sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*; Stackebrandt, E., Goodfellow, M., Eds.; John Wiley and Sons: New York, NY, USA, 1991; pp. 115–175.
16. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [CrossRef] [PubMed]
17. Hsieh, T.C.; Ma, K.H.; Chao, A. iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **2016**, *7*, 1451–1456. [CrossRef]

18. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596. [CrossRef] [PubMed]
19. Gomez-Aparicio, L.; Dominguez-Begines, J.; Villa-Sanabria, E.; García, L.V.; Munoz-Pajares, A.J. Tree decline and mortality following pathogen invasion alters the diversity, composition and network structure of the soil microbiome. *Soil. Biol. Biochem.* **2022**, *166*, 108560. [CrossRef]