

Article

Microeukaryotic Communities on the Fruit of *Gardenia thunbergia* Thunb. with a Focus on Pathogenic Fungi

Bastian Steudel ¹, Himansu Baijnath ², Thorben Schwedt ^{3,4} and Armin Otto Schmitt ^{5,6,*} 

¹ Health and Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China; bastiansteudel@aol.com

² Ward Herbarium, School of Life Sciences, University of KwaZulu-Natal, Durban 4000, Kwa-Zulu Natal, South Africa; baijnathh@ukzn.ac.za

³ DLR German Aerospace Center, Bunsenstrasse 10, 37073 Göttingen, Germany; thorben.schwedt@gmail.com

⁴ Institute of Maritime Energy Systems, DLR German Aerospace Center, Max-Planck-Straße 2, 21502 Geesthacht, Germany

⁵ Breeding Informatics, Department of Animal Sciences, Georg-August University of Göttingen, Margarethe von Wrangell-Weg 7, 37075 Göttingen, Germany

⁶ Center of Integrative Breeding Research, Albrecht Thaer-Weg 3, 37075 Göttingen, Germany

* Correspondence: armin.schmitt@uni-goettingen.de

Abstract: Woody fruit which stay on ornamental plants for a long time may present a risk of infection to other organisms due to the presence of pathogens on their surface. We compared the microbe communities on the fruit surfaces of garden ornamental *Gardenia thunbergia* Thunb. with those on other surfaces in the study region. As *Gardenia* fruit contain antifungal substances, the focus of this study was on the fungal communities that exist thereon. We used Illumina sequencing to identify Amplicon Sequence Variants (ASV) of the internal transcribed spacer 2 (ITS2) of the ribosomal RNA. The microbial communities of the *Gardenia* fruit are distinct from the communities from the surrounding environments, indicating a specialized microhabitat. We employed clustering methods to position unidentified ASVs relative to known ASVs. We identified a total of 56 ASVs representing high risk fungal species as putative plant pathogens exclusively found on the fruit of *Gardenia*. Additionally, we found several ASVs representing putative animal or human pathogens. Those pathogens were distributed over distinct fungi clades. The infection risk of the high diversity of putative pathogens represented on the *Gardenia* fruit needs to be elucidated in further investigations.

Keywords: environmental DNA; ITS2; global trading; ornamental plants



Citation: Steudel, B.; Baijnath, H.; Schwedt, T.; Schmitt, A.O. Microeukaryotic Communities on the Fruit of *Gardenia thunbergia* Thunb. with a Focus on Pathogenic Fungi. *Pathogens* **2021**, *10*, 555. <https://doi.org/10.3390/pathogens10050555>

Academic Editor: Lawrence S. Young

Received: 13 April 2021

Accepted: 30 April 2021

Published: 4 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to human activities, organisms have been moved around the globe over the last few centuries. Such translocations have led to invasions and naturalizations of alien species on all continents [1,2]. Some of these transports have been unintended, e.g., pathogens on crops, microbes on boots, crabs in water reservoirs of ships, etc. However, the introduction of hazardous organisms to other environments is a major threat to ecosystems and the agricultural and forest economy [3–5]. Recent studies revealed that microbial organisms are not “everywhere,” as previously stated as a paradigm [6,7]. Hence, studies on the microbial communities brought into new environments can be a possible measure of plant and animal/human disease prevention. In this study, methods of detection and strategies for preventing infections of plant pathogens are discussed intensively [8–10]. Although this study was initially focused on fungal plant pathogens as a source of threats to ecosystems as well as crop and ornamental plants in new environments, we identified several animal and human pathogens as well. Prominent taxa hazardous to both plants and animals are fungi of the Ascomycetes and Basidiomycetes. This study revealed different DNA sequences from those taxa to be present only on the fruit of *Gardenia thunbergia* Thunb. (Rubiaceae). To estimate the hazardous potential of transplanting *G. thunbergia* to other

environments, molecular surveys of the ITS2 region were used. The ITS2 region is a very variable region between the 5.8S and the 28S ribosomal genes of eukaryotes. ITS2 is appropriate to determine taxa, especially fungi on the species level, if reference sequences are available [11,12].

Several fungi are well known as plant pathogens. However, different genera include pathogenic and nonpathogenic species [13–15], which can be difficult to distinguish. Further, host specificity varies substantially between pathogenic fungi [16]. On the other hand, the identification of host plants of pathogens is sometimes difficult. For example, it was shown that ornamental plants like orchids [15,17–20], agricultural plants [21], and woody plants [22,23] can be infected by *Fusarium* spp., causing molds and leaf spots. Interestingly, *Fusarium* spp. and *Acremonium* spp. are among the taxa causing serious infections in humans [24,25]. Often, human or animal pathogens are of other taxa than plant pathogens, even if they may be sister clades to each other. Prominent animal pathogens are members of the Hypocreomycetidae like *Beauveria*, while other genera of this subclass like *Verticillium* are plant pathogens. *Fusarium* and *Acremonium* also belong to this subclass of Sordariomycetes in the Ascomycota. However, it is known that newly introduced plant pathogens, especially fungi, can be traced back to agricultural and ornamental plant imports and nurseries [26–28].

G. thunbergia, commonly called the white or forest gardenia, is an evergreen shrub or small tree usually growing up to five meters in height with strongly scented white flowers [29]. Its hard woody indehiscent fruit is consumed by large mammals, thus distributing the seeds. The species occurs naturally in South Africa (Eastern Cape and KwaZulu-Natal Provinces), Southern Mozambique, and eastern eSwatini (Swaziland). It was observed that the fruit of *G.thunbergia* persists for several years on the tree and may therefore host pathogens. Hence, the fruit of this species was studied. It was reported that different pathogens were introduced to Egypt through *Gardenia jasminoides* J.Ellis [30]. We expected adapted fungi communities in the samples of *G.thunbergia*, as it was previously shown that the fruit of other *Gardenia* species bear antifungal ingredients [31,32], although contaminations with fungi on the dried fruit of *G. jasminoides*, which is used as medical drug, are known [33].

2. Results

2.1. General Overview on the Microeukaryote Communities of All Samples

G. thunbergia fruit samples (Figure 1) were found to be clustered closely together (Figure 2).

The samples from the whole study area are more widespread than the *Gardenia* fruit samples, while sample Afr_4 [*Syagrus romanzoffiana* (Cham.) Glassman] is placed in the lower-right part of the graphs, next to the *G. thunbergia* samples.

Including the nonidentified ASVs under study, 58 Alveolata, 3243 fungi, 3 Heterobolosa, 32 Metazoa, 44 Protista, and 539 microalgae were found, making a total of 3919 ASVs. Most of the Alveolata, Heterobolosa, and Metazoa were identified at the genus level, whereas the Protista were not been identified at a lower taxon level. However, this study focused on the ASV of fungi for further analyses to identify putative pathogens.

A comparison between the sequences found on the fruit of *G. thunbergia* (samples Afr_10 to Afr_16) and the other samples (Afr_1 to Afr_9) showed that 1029 sequences of the fungi (33%) were found in both groups, while 261 (8%) were only found on the *Gardenia* fruit. In contrast, 172 sequences of microalgae (32%) were present in both groups, but only 1 (0.2%) was found exclusively on the fruit. Thus, the fungi found exclusively on the fruit of *Gardenia* do not stem from lichen of the surrounding environment. Hence, they may represent lichen fungi which are infecting the same microalgae as other fungi in the study region, or do not stem from lichen growing on the fruit. If those fungi are not endosymbionts of lichen, they may represent pathogens.

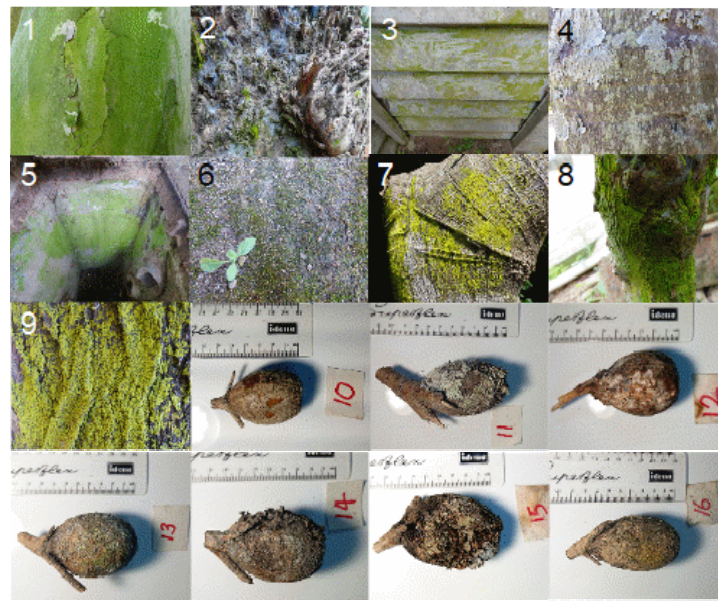


Figure 1. Samples from South Africa used in this study. Samples Afr_1toAfr_9 (pictures 1 to 9) are from different surfaces of plants and artificial substrates and are used to characterize the microbial communities of the whole study site, while samples Afr_10 to Afr_16 are *Gardenia thunbergia* fruit (see Table S1).

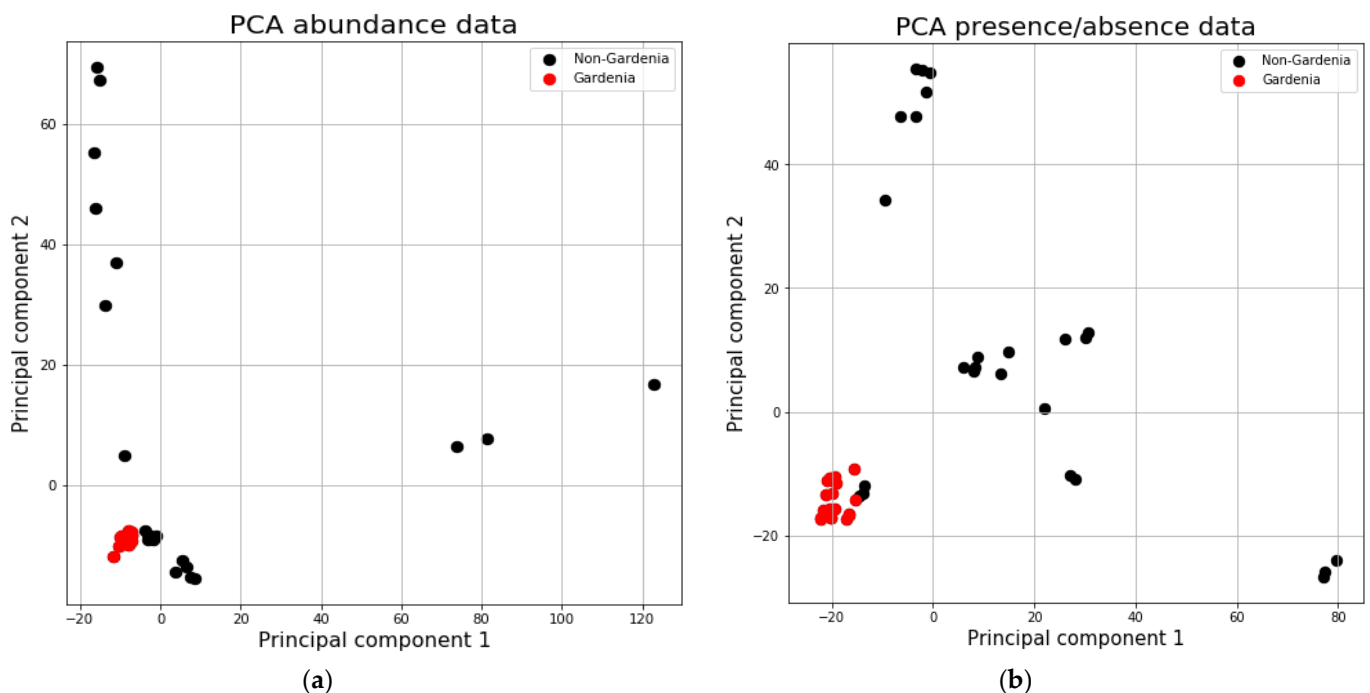


Figure 2. Principal Component Analyses (PCA) of the microbial communities of all samples with abundance data (a) and presence/absence data (b). Note that *Gardenia* fruit samples are clustered on the down-left side (red dots) while the other samples (black dots) are widely distributed for both analyses. Principal Component 1 has an explanatory power of 15.8% and 16.5%, and Principal Component 2 of 10.7% and 13.2% for the abundance and the presence/absence data, respectively.

2.2. Abundances of the Microbial Communities

In total, 2.57 million reads were recorded, of which 1.06 million were from *Gardenia thunbergia* Thunb. fruit. The most common organisms were fungi and lichen algae. Out of the records with more than 1% of total reads, 12 fungi and 2 microalgae were found;

the latter had the lowest read numbers. Focusing on the *Gardenia* fruit, we identified 62,541 reads from organisms that were present only on the fruit and not in the other local communities. After removal of ASVs from other organisms, 261 fungus ASVs were recorded by 62,337 reads. The most abundant sequence was ASV_000076 (*Candida homilientona*, 11,153 reads), followed by ASV_000487 (*Penicillium* sp., 3204 reads), and ASV_003838 (*Cosmospora* sp., 2445 reads). 113 (43%) ASVs which were found only on the *Gardenia* fruit accounted for more than 0.01% of the reads from the *Gardenia* samples, while 27 (48%) of the 56 high risk ASVs accounted for more than 0.01% of the reads from the *Gardenia* fruit.

2.3. Diversity and Composition of Fungi on Fruit of *Gardenia Thunbergia*

The dendrogram including all fungus sequences exclusively from the *Gardenia* fruit (Figure S1) showed a cluster of Basidiomycota (cluster A), four clusters of Ascomycota (clusters B to E), and a nonphylogenetic group of Ascomycota (F), which was merged into one clustering as otherwise three very small groups would have been necessary (Figure S1). A color code was used to identify high risk clades for plant and animal pathogens (red for plant, and brown for human/animal pathogens) in the six single group dendrograms. Clades that most probably bear only nonpathogenic species are marked in dark green, while those that are unlikely to bear pathogens are marked in bright green. Yellow marked clades cannot be estimated regarding their pathogen potential. A combination of the respective dendrogram letter and a consecutive number to identify clades or sequences that were not identified to genus level was used.

The dendrogram of cluster A (Basidiomycota) revealed five taxa (Figure S2). Most of the sequences belonged to the Tremellomycetes, forming three (or four if A1 is in the Tremellomycetes) taxa (*Kockovaella*/*Fellomyces*, A2, and *Tremella*), of which one comprised the genus *Septobasidium* (Pucciniomycetes). The A1 clade may be any class of Basidiomycota, as it stands between *Septobasidium* and the Tremellomycetes clade.

The dendrogram of cluster B (Figure S3) shows four groups and two single sequences that cannot be determined at the genus level or are unknown. Clade B1 comprises two sequences of Lecanoromycetes. Clade B2 contains Didymosphaeriaceae, clade B3 is not further classified, and the sequence B4 is sister to *Funbolia*. Clade B5 is sister to sequence B6 and the *Xanthomendoza weberi* reference sequence. Further, sequence ASV_004060 was identified as *Diplodia* and was placed next to the pathogen *Macrophomina phaseolina*. Hence, this sequence may represent a possible plant pathogen fungus. The pathogens *Thielaviopsis* and *Ceratocystidiaceae* sp. "MF952419" are clustered with the lichen genus *Xylaria*, all other reference pathogen sequences included in cluster B are the sister group of these.

Cluster C shows pathogen taxa from the Hypocreales (Figure 3).

One main clade is built by the animal and human pathogen genera *Cyphellophora*, *Exophilia*, and *Cladophialophora*, including the C1 sequence ASV_003637. The clade of the genus *Cyphellophora* is monophyletic and includes the previously unidentified sequence ASV_000635. The clade of the genus *Exophilia* is monophyletic and includes three sequences (ASV_001177, ASV_000886, and ASV_004028) which could not be identified.

The other main clade is formed by the plant pathogens, including all plant pathogen references of this clade, the genera *Eutype* and *Sarcocladium*, and the genera *Diaporthe* and *Phomopsis* from the Diaporthales. Clades C2 (Pleosporaceae) and C3 (Nectriaceae) are nested in this clade. However, these clades include single sequences determined to genus level. The sequences ASV_003826 and ASV_001699 of C2 belong to the Pleosporaceae, while the sequence ASV_003068 was identified as *Neoocultibambus*, and the sequence ASV_002435 as *Crassiparies*, both from the Pleosporaceae. The remaining three sequences of this clade were not identified further. ASV_004044 belongs to *Pectenaria* (Peltigerales), and is the basis to the lichen reference sequences of *Cladonia* and sequence C4, while *Sarcocladium* and the Nectriaceae are nested into this clade.

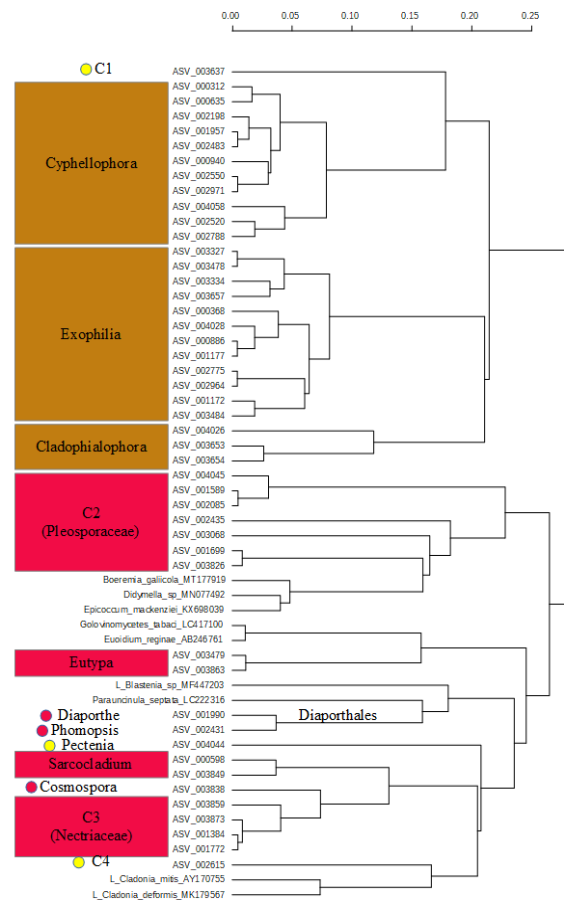


Figure 3. Dendrogram of cluster C. Almost all sequences belong to pathogens. Brown clusters represent pathogens to animals and humans with only few plant pathogens, red clusters are likely plant pathogens. Dots are used for singleton sequences. Due to space limitation “Diaporthales” was written on the branch of the cluster of *Diaporthe* and *Phomopsis*. The level of difference is given in the scale.

Cluster D (Figure S4) includes mainly *Penicillium* and *Talaromyces* sequences as a sister group to lichen fungi (Pezizomycotina). *Talaromyces* A clade is the basis of D1, *Penicillium* A, and the *Talaromyces* B clade. *Talaromyces* C sequence ASV_001465 (*Talaromyces verrucosus*) is the sister to the two Hypocreales sequences D4 (ASV_002546 and ASV_003642). The basis of the *Penicillium* cluster is built by the D5 (Eurotiomycetes) clade. D6 is included in the *Penicillium* B and *Penicillium* C cluster and most likely represents a *Penicillium* species. D7 is the sister group to *Penicillium* D, including *Xylographa trunciseda*, while *X. trunciseda* seems to be misplaced.

In cluster E (Figure 4), only Sordariomycetes, with only four sequences from *Phaeocremonium* (Togniales) split into two branches (ASV_003665, ASV_000840, and ASV_000997) and ASV_004056, and all other classified sequences from the Hypocreales were identified.

However, some of the sequences were not identified deeper than to the level of Sordariomycetes, and may hence represent other orders. Overall, the three genera *Acremonium* (three clades), *Phaeocremonium* (two clades), and *Niesslia* (two clades) are split, while the reference sequence pairs of *Verticillium* and *Fusarium* are distinct. The genera *Beauveria*, *Sarcocladium*, *Trichothecium*, *Clonostachys*, *Pseudocosmophora*, and *Rosasphaeria* each form monophyletic clades. The E5 (Nectriaceae A) clade is the basis of E6 (Nectriaceae), the genus *Pseudocosmophora* (Nectriaceae), E7, E8, and *Verticillium leptobracteatum*. Clade E10 bears the sequence ASV_004048 (*Rosasphaeria*), E11 cannot be placed, although it is sister to *Microcera*.

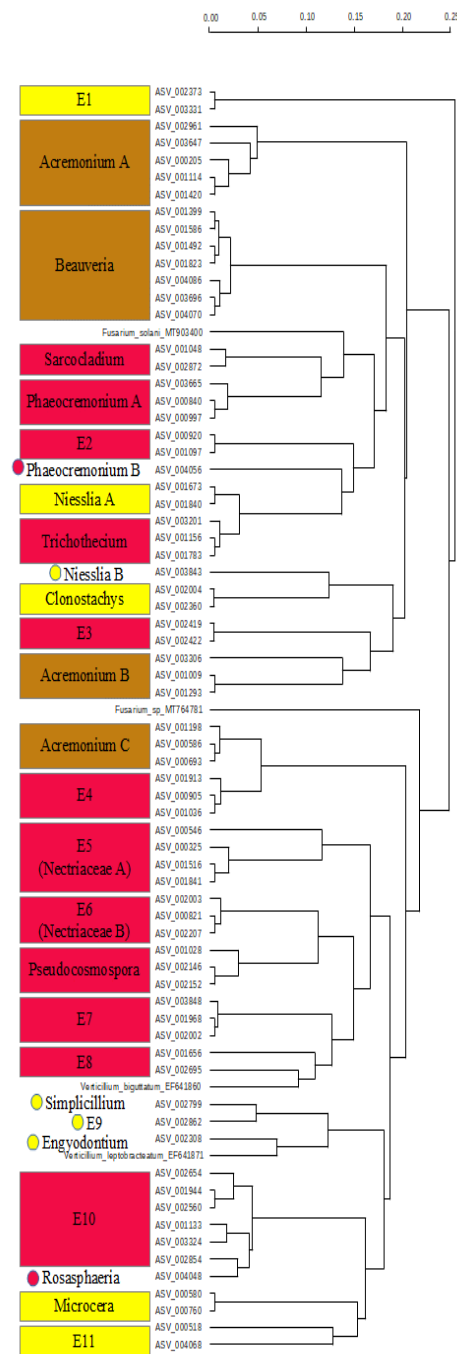


Figure 4. Dendrogram of cluster E. Yellow clades cannot be classified as plant pathogens, brown clades are animal pathogens, and red clades represent most likely plant pathogens. Dots are used for singleton sequences. The level of difference is given in the scale.

Cluster F (Figure S5) was built by the sequences not included in the clusters A–E, as it would be inappropriate to build three very small clusters of those remaining according to their position in the overall phylogeny. F1 is the sister to F2 (Eurotiales), the genus *Arachnomycetes* is the sister clade to F3, *Candida*, and the reference sequences of *Taphrina*.

Sequences representing organisms with a high risk of being plant pathogens with sequence abundances and fruit age are shown in Table 1.

Table 1. Overview of the sequences rated as representing high risk plant pathogens. Sequence ID stands for sequence identity, no. seq stands for total number of sequences, no. samples stands for number of samples in which the sequence was found, the age was classified as “young” (y), “medium” (m), and “old” (o). If the respective sequence was found in all age classes, this was noted as “all”. If the family or the genus of the organism represented by the ASV is not determined this is indicated by a “?”.

Sequence ID	Family	Genus	Cluster	No. seq.	No. samp.	Age
ASV_004045	Pleosporaceae	?	C	21	1	y
ASV_001589	Pleosporaceae	?	C	94	2	y
ASV_002085	Pleosporaceae	?	C	55	1	y
ASV_002435	Pleosporaceae	?	C	133	5	all
ASV_003068	Pleosporaceae	?	C	33	1	o
ASV_001699	Pleosporaceae	?	C	57	1	y
ASV_003826	Pleosporaceae	?	C	10	1	y
ASV_003479	Diatrypaceae	Eutypa	C	31	1	o
ASV_003863	Diatrypaceae	Eutypa	C	18	1	o
ASV_001990	?	?	C	365	1	m
ASV_002431	?	?	C	309	1	m
ASV_000598	incertae sedis	Sarcocladium	C	737	3	y/o
ASV_003849	incertae sedis	Sarcocladium	C	26	1	y
ASV_003838	Nectriaceae	Cosmospora	C	2445	7	all
ASV_003859	Nectriaceae	?	C	160	1	o
ASV_003873	Nectriaceae	?	C	23	1	o
ASV_001384	Nectriaceae	?	C	176	1	o
ASV_001772	Nectriaceae	?	C	144	1	o
ASV_001048	incertae sedis	Sarcocladium	E	435	1	o
ASV_002872	incertae sedis	Sarcocladium	E	51	1	o
ASV_003665	Togniniaceae	Phaeocremonium	E	66	1	o
ASV_000840	Togniniaceae	Phaeocremonium	E	240	1	o
ASV_000997	Togniniaceae	Phaeocremonium	E	168	1	o
ASV_000920	?	?	E	184	1	y
ASV_001097	?	?	E	148	1	y
ASV_004056	Togniniaceae	Phaeocremonium	E	48	1	m
ASV_003201	incertae sedis	Trichothecium	E	28	2	y/m
ASV_001156	incertae sedis	Trichothecium	E	146	2	y/m
ASV_001783	incertae sedis	Trichothecium	E	87	2	y/m
ASV_002419	?	?	E	39	3	y/o
ASV_002422	?	?	E	34	3	y/o
ASV_001913	?	?	E	147	2	y/o
ASV_000905	?	?	E	164	3	all
ASV_001036	?	?	E	190	4	y/o
ASV_000546	Nectriaceae	?	E	662	1	m
ASV_000325	Nectriaceae	?	E	769	1	o
ASV_001516	Nectriaceae	?	E	111	2	m/o
ASV_001841	Nectriaceae	?	E	83	1	o
ASV_002003	Nectriaceae	?	E	56	2	y/m
ASV_000821	Nectriaceae	?	E	236	1	m
ASV_002207	Nectriaceae	?	E	37	1	m
ASV_001028	Nectriaceae	Pseudocosmophora	E	512	1	m
ASV_002146	Nectriaceae	Pseudocosmophora	E	461	1	m
ASV_002152	Nectriaceae	Pseudocosmophora	E	189	1	m
ASV_003848	?	?	E	17	2	y/o
ASV_001968	?	?	E	60	1	o
ASV_002002	?	?	E	55	1	o
ASV_001656	?	?	E	88	2	o
ASV_002695	?	?	E	56	2	y/o
ASV_002654	?	?	E	178	4	all
ASV_001944	?	?	E	414	4	m/o
ASV_002560	?	?	E	128	4	all
ASV_001133	?	?	E	386	3	all
ASV_003324	?	?	E	469	5	all
ASV_002854	?	?	E	273	2	m/o
ASV_004048	Niessliaceae	Rosasphaeria	E	37	5	all

The abundances of sequence numbers of the 56 high risk taxa in our sampling are very different, ranging from 10 to 2445.

We recorded an estimated age gradient of the fruit of *G. thunbergia* under study from sample Afr_10 (youngest), Afr_16, Afr_13, Afr_12, Afr_14, Afr_15, to Afr_11 (oldest) based on the density and size of the on-growing lichens, but the last three samples (Afr_14,

Afr_15, and Afr_11) were very similar regarding their lichen growth, and thus, may be similar in age. We classified the samples Afr_10 and Afr_16 as young, the samples Afr_13 and Afr_12 as medium age, and the samples Afr_14, Afr_15, and Afr_11 as old based on the coverage and size of the on-grown lichen. We found 27 of the high risk sequences on the young, 24 on the medium aged, and 35 on the old fruit.

3. Discussion

3.1. Overall Microeukaryote Communities

Microeukaryote communities from the fruit of *G. thunbergia* (samples Afr_10-Afr_16) are different from those of other host plants and artificial surfaces (samples Afr_1-Afr_9) from our study region. For these special habitats, a more detailed analysis is required of the putative pathogens of these communities. These distinct microbial communities of *G. thunbergia* fruit were expected, as the genus is known to bear antifungal substances. However, it is remarkable that we found only one microalga on the fruit, but 261 fungi ASVs. For the other microalgae and fungi, more than 30% of the found ASVs were recorded on the *Gardenia* fruit and the other samples, while the rest was only found in the non-*Gardenia* samples. While we expected to observe only a part of the microbial community of a region on one specialized habitat (*Gardenia* fruit), this clearly shows that the lichen on *G. thunbergia* fruit bear photobionts that are not specialized to this habitat. In fact, we found only very few microfungi considered as lichen forming on the *Gardenia* fruit.

3.2. Pathogen Fungi on *Gardenia thunbergia* Fruit

A great variety of fungus sequences representing taxa of possible plant and animal pathogens was only present on *G. thunbergia* fruit. These are from the Basidiomycetes and the Ascomycetes. Generally, estimates of the hazardous potential of different species of a genus can be difficult, and morphological determination can mismatch the underlying phylogeny [13,34]. Therefore, we identified clades that may be hazardous if the genus is known to bear mainly pathogenic species or is nested into hazardous species/genera in cases of species that are unknown to genus level.

In the Basidiomycetes group of our sequences, we found only non- or likely nonhazardous clades. *Septobasidium* is known to form symbiotic associations with insects, but is not associated with plant parasitism. The remaining clades are unlikely to be hazardous to plants due to their position in the dendrogram.

In the Ascomycetes groups of the fungi occurring only on the *Gardenia* fruit, clusters were identified with high risk of plant (and animal) pathogens and others with low risk of containing plant pathogens. While clusters C and E include the most plant pathogens, cluster F represents numerous animal pathogens. In cluster C, the genera *Xyphellophora*, *Exophilia*, and *Cladophialophora* are known as animal and human pathogens or infectious. The other half of the dendrogram bears the included pathogen reference sequences and needs to be considered as plant pathogen taxa, except the sequence ASV_004044 (*Pectenia*) and ASV_002615 (unidentified). Note that the reference sequences of *Cladonia* are included there. Cluster E includes known animal pathogens (*Acromonium* on mammals and *Beauveria* on Arthropods) and known plant pathogens, while smaller groups cannot be identified as plant pathogens but may nonetheless be pathogenic.

Clusters B and D may include taxa representing plant pathogens, but this is difficult to estimate as the clade B1 is nested in plant pathogen reference sequences, and *Xylaria*, which is not known as plant pathogenic but rather a saprophytic genus, is the crown group of the pathogen reference sequences. On the other hand, the B3 and B5 clades are related to lichen reference sequences. In cluster D the closely related genera *Penicillium* and *Talaromyces* form the major group in the dendrogram. Both genera inhabit pathogens and nonpathogenic species. Hence, it cannot be judged whether they represent a possible threat of infection in foreign habitats.

Focusing on the number of likely pathogenic fungi taxa, we identified 18 sequences in the clade C, and 38 sequences in the clade E as „high risk“ taxa for plant pathogens (total

56). Of these, 13 (clade C) and 25 (clade E), respectively, were not determined to genus level (Table 1). Hence, these taxa may represent thus far unknown pathogens. The human and animal pathogen sequences could be determined on genus level, however.

The abundance of the ASVs representing high risk organisms is very different. It may be speculated that the infection risk depends on the number of organisms in the different samples. However, as we used Illumina sequencing with a prior PCR, the number of sequences detected may not represent the corresponding abundance or relations in the living biofilm. Irrespective of that, about half of the high risk ASVs were found with a read abundance of more than 106 counts (0.01% of the reads from *Gardenia* fruit), indicating that these are present at a certain quantity in the biofilms studied. As the number of ASVs is highest (35 out of 56) on the oldest fruit, the infectious potential of old *G. thunbergia* fruit may be considered higher than that of young or medium-aged fruit.

3.3. Lichen on *Gardenia thunbergia* Fruit

As the old fruit of *G. thunbergia* that were studied were covered with thick lichen biofilms or crusts, we expected some extraordinary lichens as well. However, exclusively on the fruit of *G. thunbergia*, only three sequences (ASV_000986, ASV_001217, and ASV_004044) were found that were identified as lichen fungi (Pezizomycetes), and two additional sequences (ASV_004052 and ASV_002615) that may represent a lichen fungus. This, however, corresponds to the fact that only a single microalga was found exclusively on the fruit. Hence, lichen growing on the fruit of *G. thunbergia* are most likely not restricted to this habitat. This is somewhat surprising, as it we expected to observe lichen and fungi on the fruit surfaces, which are adapted to the fruit ingredients of *G. thunbergia*, since the fruit of *G. jasminoides* contains cerbinal, an antimicrobial ingredient [31]. Further, it was reported that also *G. brighamii* H.Mann leaf extracts contain antifungal substances [32]. On the other hand, many fungi taxa found exclusively on the fruit of *G. thunbergia*, which may be a threat to other environments, crops and garden ornamentals, were identified.

4. Materials and Methods

4.1. Methods Overview

We analyzed 16 samples of microbial mats or biofilms from different plant and artificial surfaces in the Durban region in South Africa (Figure 1). We used nine samples from different habitats to determine the general microbial communities at the study site (samples Afr_1 to Afr_9) and compared them with the samples of *Gardenia* fruit (samples Afr_10 to Afr_16). Three replicates were taken from each sample and processed individually, resulting in a total of 48 DNA extractions. For identification of the microbial communities, we used Illumina sequencing to produce libraries of ITS2 sequences of all eukaryotes in these samples. After identification of sequences that were unique to the fruit surfaces of *G. thunbergia* possible pathogenic taxa were identified.

4.2. Study Region and Sample Collection

Samples were collected in “Strawberry Fields” in KwaZulu-Natal, Durban, Sherwood, RSA, and nearby. Sherwood is known as the “garden suburb” of Durban, South Africa. Durban is characterized by a warm temperate climate with an annual mean temperature of almost 21 °C and a mean annual precipitation of about 1000 mm. From May to October, temperatures are lower than the year mean temperature, precipitation is lower from April to September than in the other months.

The seven samples of *G. thunbergia* fruit (samples Afr_10 to Afr_16) were obtained exclusively from Strawberry Fields, while the others were collected at Strawberry Fields (samples Afr_1, Afr_4, Afr_5, Afr_6, and Afr_9) and other sites (samples Afr_2, Afr_3, Afr_7, and Afr_8). Non-*Gardenia* samples were taken from different surfaces like concrete, tree barks, and a leaf scale. The *Gardenia* fruit was collected whole, while the hard surface samples were scratched with a scalpel and the plant surface samples were cut from the plant. An overview on the samples is given in Table S1. All samples were dried indoor at

room temperature on separate paper sheets to avoid contamination in Durban, put into plastic jars for shipment, and stored after shipping in a freezer at $-20\text{ }^{\circ}\text{C}$ until sample processing.

4.3. DNA Preparation and PCR

Three replications of DNA extraction per sample were obtained using surface material from different spots of the sample surface by scalpel scratching. This material included the surface material of the sample, i.e., epidermis and some of the tissue below the epidermis of the plant samples. This was done to include organisms growing on the surface and within the plant material. After removal, the material of each replicate was ground separately in a mortar without liquid nitrogen to obtain an even sample material; later, it was further ground during the DNA extraction procedure. The DNeasy PowerSoil Pro Kit (Qiagen, Hilden, Germany) was used as specified in the manual for DNA extraction. After DNA extraction, a PCR with the primers ITS3_KYO2 (GATGAAGAACGYAGYRAA) ([35]) and ITS4 (TCCTCCGCTTATTGATATGC) ([36]) was done, including the default MiSeq adapters. The volume of the PCR reactions was 50 μL , containing 1 μL of each primer solution (10 μM), 10 μL 5 \times GC Buffer (Thermo Scientific, Waltham, MA, USA), 0.2 μL MgCl_2 (25 mM, Thermo Scientific, Waltham, MA, USA), 2.5 μL DMSO (5%, Thermo Scientific, Waltham, MA, USA), 1 μL dNTPs (10 mM, Thermo Scientific, Waltham, MA, USA), 0.5 μL Phusion High-Fidelity DNA polymerase (10 mM, Thermo Scientific, Waltham, MA, USA), 1 μL template (containing 25 ng DNA), and 32.8 H_2O . Denaturation was done at $98\text{ }^{\circ}\text{C}$ for 1 min, 25 cycles were performed at $98\text{ }^{\circ}\text{C}$ for 45 s, $60\text{ }^{\circ}\text{C}$ for 45 s, and $72\text{ }^{\circ}\text{C}$ for 30 s, and a final extension was done at $72\text{ }^{\circ}\text{C}$ for 5 min. Negative controls were prepared without template but with 1 μL H_2O , and positive controls were prepared with a template known to bear DNA from a previous study. PCR products were concentrated and purified with MagSi-NGS^{Prep} magnetic beads, as recommended by the manufacturer (Steinbrenner, Wiesenbach, Germany). DNA was eluted in 30 μL elution buffer EB (Qiagen, Hilden, Germany). Purified PCR products were quantified and sequenced as described by Schneider and colleagues [37] using a MiSeq instrument and v3 chemistry (Illumina, San Diego, CA, USA).

4.4. Data Processing and Selection of Sequences

The replicates Afr_3a, Afr_3b, and Afr_15c could not be amplified successfully, and were therefore excluded from further analyses. From the other runs, we obtained on average 57,277 sequence reads per sample replicate, with a minimum of 11,382 and a maximum of 128,359.

Paired-end sequencing data from the Illumina MiSeq were processed as follows. Sequences were quality-filtered with fastp (version 0.19.6) ([38]) using a per base phred score of 20, base pair correction by overlap (option -c), as well as 5'- and 3'-end read trimming with a sliding window of 4, a mean quality of 20 and minimum sequence size of 50 bp. After quality control, the paired-end reads were merged using PEAR (version 0.9.11) ([39]), primers were clipped using cutadapt (version 1.18) ([40]) with default settings and further processed using VSEARCH (v 2.9.1) ([41]). This included size sorting and filtering of the paired reads to a minimum of 140 bp (`-sortbylength -minseqlength 140`) and dereplication (`-derep_fulllength`). Dereplicated ASVs were denoised with UNOISE3 using default settings (`-cluster_unoise -minsize 8`) and chimeras were removed (`-uchime3_denovo`). Additional reference-based chimera removal was performed (`-uchime_ref`) against the SILVA SSU NR database (<https://www.arb-silva.de/>, accessed on 6 January 2020). A total of 4290 ASVs were identified in the samples. Taxonomic assignment was performed against the UNITE database v8.2 (<https://plutof.ut.ee/#/doi/10.15156/BIO/786372>, accessed on 6 January 2020) with an identity of at least 90% to the query sequence resulting in a total of 3437 ASVs, of which 3146 have been identified as “fungi” and 291 had no blast hits. As most microalgae were excluded by the processing, probably because they may not be included in the data base, we reconsidered those excluded ASVs that had a minimum of 80 sequence reads in at least two different sample replicates. The resulting “no blast hit”

sequences were manually grouped into fungi or Viridiplantae using the NCBI database (<https://www.ncbi.nlm.nih.gov/>, accessed on 7–12 September 2020) for inclusion into further analyses. This additional threshold to omit artifacts was used as it is very unlikely to record 80 sequence reads resulting from chimera or other artifacts [42]. Only for a single sequence there was no similarity found in the NCBI data base with standard settings. This resulted in a total of 4067 different ASVs. We removed 148 sequences from Pteridophyta, Bryophyta, and Anthophyta (e.g., *Euphorbia*, *Fallopia*, *Gardenia*, *Libidibia*, and *Yucca*) to keep only sequences of the microalgae as photosymbiotic organisms, resulting in a total of 3919 sequences which we included in the analyses.

ASVs instead of operational taxonomic units (OTU) were used for our analyses as ASV can resolve the biological diversity more exactly [42,43].

For the description of the sequence abundances, the mean values of the sequence numbers of the three replicates for all samples was used. This was chosen as generally very strong correlation between the replicates of each sample (Table S2), with a mean correlation coefficient of almost 0.9 was found.

4.5. Data Records

The data ITS2 amplicon raw sequences were deposited at the National Center for Biotechnology Information under the BioProject PRJNA672041.

4.6. Statistics

Principal Component Analyses (PCA) were performed using the free machine learning library “scikit-learn” for the Python programming language. Data were provided in CSV format and read using Pandas (pandas.read_csv—pandas 1.2.3 documentation (pydata.org)) with 45 rows representing the different replicates and 4290 columns, representing the ASVs. The dataset was normalized using the scikit-learn StandardScaler (sklearn.preprocessing.StandardScaler—scikit-learn 0.24.1 documentation (scikit-learn.org)) to a mean of 0 and standard deviation 1. The first 10 principal components were calculated using sklearn (sklearn.decomposition.PCA—scikit-learn 0.24.1 documentation (scikit-learn.org) ([44]) and “fit_transform(x)”, where x is the normalized dataset and the explained variance was available with the attribute (“explained_variance_ratio”). Results were plotted using matplotlib (Matplotlib: Python plotting—Matplotlib 3.3.4 documentation). The same procedure was performed for the presence/absence data set in which the read abundances were omitted.

4.7. Construction of Dendrograms

Focusing on the 261 sequences of fungi exclusively found on the *Gardenia* fruit, we constructed a dendrogram including known pathogens and lichen fungi as references. This was necessary for sequence assembly for the ASV that were not identified during the previous data processing (e.g., “uncultured fungi”, “no blast hit”). For this, distance matrices produced by using the package “kmer” ([45]) in R 4.0.2 ([46]) were built. The six groups identified in this dendrogram were used to construct individual dendrograms accordingly. Eight reference sequences (two *Phytophthora*, two *Pythium*, two *Lichenomphalia*, and two *Colletotrichum*) were omitted in the separate analyses of the six groups as no related taxa were found (Oomycota, Agaricomycetes, and Glomerellales, respectively) in the ASVs exclusively from the samples from the *Gardenia* fruit.

5. Conclusions

The fruit of *G. thunbergia* bears a great variety of pathogens to plants as well as to humans and animals. The pathogens found on the fruit of *G. thunbergia* may pose a threat to plants and animals, especially if they are brought into environments to which they have not yet adapted. How the infection risk of these microbial communities needs to be considered regarding phytopathology measures requires further study.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/pathogens10050555/s1>, Figure S1: Overview on the clusters based on the dendrogram including all sequences, Figure S2: Dendrogram of the Basidiomycota clade A. Dark green boxes indicate clusters that are very unlikely to contain plant pathogens, light green clusters may contain plant pathogens, Figure S3: Dendrogram of cluster B. Light green clusters that are unlikely to contain plant pathogens, yellow clusters can not be classified regarding their pathogenic potential, Figure S4: Dendrogram of cluster D, consisting mainly of *Taloromyces* and *Penicillium*, Figure S5: Dendrogram of cluster F. The yellow clade can not be classified regarding the plant pathogen potential, the brown clades represent likely animal pathogens, Table S1: Samples used in this study, Table S2: Similarity values between the replicates of one sample.

Author Contributions: The study was initiated by B.S. and H.B. Samples were collected by H.B., and field observations have been done by H.B. over the last years. DNA extraction was performed by T.S. and B.S. Analyses of the data was done by A.O.S., T.S., and B.S. The manuscript was drafted by B.S. with contributions from all other authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data ITS2 amplicon raw sequences were deposited at the National Center for Biotechnology Information under the BioProject PRJNA672041.

Acknowledgments: We thank Birgit Sohnrey, Rosemarie Clemens, Judith Witte, Anja Poehlein, Dominik Schneider, and Rolf Daniel for laboratory support and Dominik Schneider for comments on a previous version of the manuscript. We acknowledge support by the Open Access Publication Funds of the Göttingen University.

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

References

1. Richardson, D.M.; Pysek, P.; Rejmanek, M.; Barbour, M.G.; Panetta, F.D.; West, C.J. Naturalization and invasion of alien plants: Concepts and definitions. *Divers. Distrib.* **2000**, *6*, 93–107. [[CrossRef](#)]
2. Frenot, Y.; Chown, S.L.; Whinam, J.; Selkirk, P.M.; Convey, P.; Skotnicki, M.; Bergstrom, D.M. Biological invasions in the Antarctic: Extent, impacts and implications. *Biol. Rev.* **2005**, *80*, 45–72. [[CrossRef](#)]
3. Bandyopadhyay, R.; Frederiksen, R.A. Contemporary global movement of emerging plant diseases. *Ann. N. Y. Acad. Sci.* **1999**, *894*, 28–36. [[CrossRef](#)] [[PubMed](#)]
4. MacLeod, A.; Pautasso, M.; Jeger, M.J.; Haines-Young, R. Evolution of the international regulation of plant pests and challenges for future plant health. *Food Secur.* **2010**, *2*, 49–70. [[CrossRef](#)]
5. Santini, A.; Ghelardini, L.; De Pace, C.; Desprez-Loustau, M.L.; Capretti, P.; Chandelier, A.; Cech, T.; Chira, D.; Diamandis, S.; Gaitniekis, T.; et al. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytol.* **2013**, *197*, 238–250. [[CrossRef](#)]
6. O'Malley, M.A. 'Everything is everywhere: But the environment selects': Ubiquitous distribution and ecological determinism in microbial biogeography. *Stud. Hist. Philos. Sci. Part. C Stud. Hist. Philos. Biol. Biomed. Sci.* **2008**, *39*, 314–325. [[CrossRef](#)]
7. Xu, X.; Wang, N.; Lipson, D.; Sinsabaugh, R.; Schimel, J.; He, L.; Soudzilovskaia, N.A.; Tedersoo, L. Microbial macroecology: In search of mechanisms governing microbial biogeographic patterns. *Glob. Ecol. Biogeogr.* **2020**, *29*, 1870–1886. [[CrossRef](#)]
8. Miller, S.A.; Beed, F.D.; Harmon, C.L. Plant Disease Diagnostic Capabilities and Networks. *Annu. Rev. Phytopathol.* **2009**, *47*, 15–38. [[CrossRef](#)]
9. Parnell, S.; Bosch, F.V.D.; Gottwald, T.; Gilligan, C.A. Surveillance to Inform Control of Emerging Plant Diseases: An Epidemiological Perspective. *Annu. Rev. Phytopathol.* **2017**, *55*, 591–610. [[CrossRef](#)]
10. Rimbaud, L.; Dallot, S.; Bruchou, C.; Thoyer, S.; Jacquot, E.; Soubeyrand, S.; Thébaud, G. Improving Management Strategies of Plant Diseases Using Sequential Sensitivity Analyses. *Phytopathology* **2019**, *109*, 1184–1197. [[CrossRef](#)] [[PubMed](#)]
11. Yang, R.-H.; Su, J.-H.; Shang, J.-J.; Wu, Y.-Y.; Li, Y.; Bao, D.-P.; Yao, Y.-J. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS ONE* **2018**, *13*, e0206428. [[CrossRef](#)]
12. Fan, S.; Miao, L.; Li, H.; Lin, A.; Song, F.; Zhang, P. Illumina-based analysis yields new insights into the diversity and composition of endophytic fungi in cultivated *Huperzia serrata*. *PLoS ONE* **2020**, *15*, e0242258. [[CrossRef](#)]
13. Berbee, M.L. The phylogeny of plant and animal pathogens in the Ascomycota. *Physiol. Mol. Plant. Pathol.* **2001**, *59*, 165–187. [[CrossRef](#)]

14. Van Der Does, H.C.; Rep, M. Virulence Genes and the Evolution of Host Specificity in Plant-Pathogenic Fungi. *Mol. Plant.-Microbe Interact.* **2007**, *20*, 1175–1182. [[CrossRef](#)] [[PubMed](#)]
15. Srivastava, S.; Kadooka, C.; Uchida, J.Y. Fusarium species as pathogen on orchids. *Microbiol. Res.* **2018**, *207*, 188–195. [[CrossRef](#)] [[PubMed](#)]
16. Sexton, A.C.; Howlett, B.J. Parallels in Fungal Pathogenesis on Plant and Animal Hosts. *Eukaryot. Cell* **2006**, *5*, 1941–1949. [[CrossRef](#)]
17. Kim, W.G.; Lee, B.D.; Kim, W.S.; Cho, W.D. Root rot of moth orchid caused by *Fusarium* spp. *Plant. Pathol. J.* **2002**, *18*, 225–227. [[CrossRef](#)]
18. Almanza-Álvarez, J.; Garibay-Orijel, R.; Salgado-Garciglia, R.; Fernández-Pavía, S.P.; Lappe-Oliveras, P.; Arellano-Torres, E.; Ávila-Díaz, I. Identification and control of pathogenic fungi in neotropical valued orchids (*Laelia* spp.). *Trop. Plant. Pathol.* **2017**, *42*, 339–351. [[CrossRef](#)]
19. Sarsaiya, S.; Jia, Q.; Fan, X.; Jain, A.; Shu, F.; Chen, J.; Lu, Y.; Shi, J. First Report of Leaf Black Circular Spots on *Dendrobium nobile* Caused by *Trichoderma longibrachiatum* in Guizhou Province, China. *Plant. Dis.* **2019**, *103*, 3275. [[CrossRef](#)]
20. Fernández-Herrera, E.; Rentería-Martínez, M.E.; Ramírez-Bustos, I.I.; Moreno-Salazar, S.F.; Ochoa-Meza, A.; Guillén-Sánchez, D. *Colletotrichum karstii*: Causal agent of anthracnose of *Dendrobium nobile* in Mexico. *Can. J. Plant. Pathol.* **2020**, *42*, 514–519. [[CrossRef](#)]
21. Gebru, S.T.; Mammel, M.K.; Gangiredla, J.; Tournas, V.H.; Lampel, K.A.; Tartera, C. Draft Genome Sequences of 12 Isolates from 3 *Fusarium* Species Recovered from Moldy Peanuts. *Microbiol. Resour. Announc.* **2019**, *8*, e01642-18. [[CrossRef](#)]
22. Liew, E.C.Y.; Laurence, M.H.; Pearce, C.A.; Shivas, R.G.; Johnson, G.L.; Tan, Y.; Edwards, J.; Perry, S.; Cooke, A.; Summerell, B.A. Review of *Fusarium* species isolated in association with mango malformation in Australia. *Australas. Plant. Pathol.* **2016**, *45*, 547–559. [[CrossRef](#)]
23. Nhung, N.P.; Thu, P.Q.; Dell, B.; Chi, N.M. First report of canker disease in *Dalbergia tonkinensis* caused by *Fusarium lateritium* and *Fusarium decemcellulare*. *Australas. Plant. Pathol.* **2018**, *47*, 317–323. [[CrossRef](#)]
24. Walsh, T.J.; Groll, A.; Hiemenz, J.; Fleming, R.; Roilides, E.; Anaissie, E. Infections due to emerging and uncommon medically important fungal pathogens. *Clin. Microbiol. Inf.* **2004**, *10*, 48–66. [[CrossRef](#)] [[PubMed](#)]
25. Guitard, J.; Degulys, A.; Buot, G.; Aline-Fardin, A.; Dannaoui, E.; Rio, B.; Marie, J.P.; Lapusan, S.; Hennequin, C. *Acremonium sclerotigenum-Acremonium egyptiacum*: A multi-resistant fungal pathogen complicating the course of aplastic anaemia. *Clin. Microbiol. Inf.* **2014**, *20*, O30–O32. [[CrossRef](#)]
26. Jones, D.R.; Baker, R.H.A. Introductions of non-native plant pathogens into Great Britain, 1970–2004. *Plant. Pathol.* **2007**, *56*, 891–910. [[CrossRef](#)]
27. Moralejo, E.; Pérez-Sierra, A.M.; Álvarez, L.A.; Belbahri, L.; Lefort, F.; Descals, E. Multiple alien *Phytophthora* taxa discovered on diseased ornamental plants in Spain. *Plant. Pathol.* **2009**, *58*, 100–110. [[CrossRef](#)]
28. Jung, T.; Orlikowski, L.; Henricot, B.; Abad-Campos, P.; Aday, A.G.; Aguín Casal, O.; Bakonyi, J.; Cacciola, S.O.; Cech, T.; Chavarriaga, D.; et al. Widespread *Phytophthora* infestations in European nurseries put forest, semi-natural and horticultural ecosystems at high risk of *Phytophthora* diseases. *For. Pathol.* **2016**, *46*, 134–163. [[CrossRef](#)]
29. Palgrave, K.C. *Trees of Southern Africa*, 3rd ed.; Struik Nature: Cape Town, South Africa, 2002.
30. Hilal, A.A. New diseases of ornamentals in Egypt: V: Cut flower plant: *Gardenia*. *Egypt. J. Phytopathol.* **2004**, *32*, 145–146.
31. Ohashi, H.; Tsurushima, T.; Ueno, T.; Fukami, H. Cerbinal, a pseudoazulene iridoid, as a potent antifungal compound isolated from *Gardenia jasminoides* Ellis. *Agric. Biol. Chem.* **1986**, *50*, 2655–2657. [[CrossRef](#)]
32. Kafua, L.; Kritzinger, Q.; Hussein, A. Antifungal activity of *Gardenia brighamii* leaf extracts. *S. Afr. J. Bot.* **2010**, *76*, 411. [[CrossRef](#)]
33. Hitokoto, H.; Morozumi, S.; Wauke, T.; Sakai, S.; Kurata, H. Fungal contamination and mycotoxin detection of powdered herbal drugs. *Appl. Environ. Microbiol.* **1978**, *36*, 252–256. [[CrossRef](#)] [[PubMed](#)]
34. Feau, N.; Vialle, A.; Allaire, M.; Tanguay, P.; Joly, D.L.; Frey, P.; Callan, B.E.; Hamelin, R.C. Fungal pathogen (mis-) identifications: A case study with DNA barcodes on *Melampsora* rusts of aspen and white poplar. *Mycol. Res.* **2009**, *113*, 713–724. [[CrossRef](#)] [[PubMed](#)]
35. Toju, H.; Tanabe, A.S.; Yamamoto, S.; Sato, H. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* **2012**, *7*, e40863. [[CrossRef](#)] [[PubMed](#)]
36. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*; Innis, M., Gelfand, D., Shinsky, J., White, T., Eds.; Academic Press: New York, NY, USA, 1990; pp. 315–322.
37. Schneider, D.; Thürmer, A.; Gollnow, K.; Lugert, R.; Gunka, K.; Groß, U.; Daniel, R. Gut bacterial communities of diarrheic patients with indications of *Clostridioides difficile* infection. *Sci. Data* **2017**, *4*, 170152. [[CrossRef](#)]
38. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
39. Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **2014**, *30*, 614–620. [[CrossRef](#)]
40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
41. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ.* **2016**, *4*, e2584. [[CrossRef](#)]

42. Tsuji, S.; Miya, M.; Ushio, M.; Sato, H.; Minamoto, T.; Yamanaka, H. Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. *Environ. DNA* **2019**, *2*, 42–52. [[CrossRef](#)]
43. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**, *11*, 2639–2643. [[CrossRef](#)] [[PubMed](#)]
44. Tipping, M.E.; Bishop, C.M. Probabilistic Principal Component Analysis. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **1999**, *61*, 611–622. [[CrossRef](#)]
45. Wilkinson, S.P. kmer: An R Package for Fast Alignment-Free Clustering of Biological Sequences. Version 1.1.2.. 2019. Available online: <https://shaunpwilkinson.github.io/post/kmer-vignette/> (accessed on 6 January 2020).
46. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2020.