Deterministic processes have limited impacts on foliar fungal endophyte communities along a savanna-forest successional gradient

Mathew Harris¹, Martin Kemler², Bernard Slippers¹, Samantha-Leigh Jamison-Daniels¹, Frederick Witfeld², Monique Botha¹, Dominik Begerow², Andreas Brachmann³, and Michelle Greve¹

¹University of Pretoria
²Ruhr-Universität Bochum
³Ludwig-Maximilians-Universität München

July 6, 2022

Abstract

Patterns and drivers of succession provide insight into the mechanisms that govern community assembly and are indicators of community resilience and stability but are still poorly understood in microbial communities. We assessed whether the successional trends of woody vegetation are mirrored by foliar fungal endophyte communities of three tree species that are abundant across the woody successional gradient using a total amplicon sequencing approach. Additionally, we test the relative contribution of host identity, abiotic predictors, biotic factors, and spatial distance between sites in predicting community composition and species richness of endophyte communities. Unlike the woody community, endophyte communities showed no consistent evidence of deterministic successional trends. Host identity was the most important factor structuring fungal endophyte community composition. Spatial distance played some role in explaining differences in community composition, but the effects of this and other environmental variables were small and not consistent between different host species. Much of the variation in endophyte composition remained unexplained. Host identity was most important in predicting endophyte richness. Although endophyte communities showed no deterministic succession, community assembly was most strongly influenced by host identity and spatial distance.

Title: Deterministic processes have limited impacts on foliar fungal endophyte communities along a savanna-forest successional gradient

Mathew Andrew Harris¹,²; Martin Kemler³; Bernard Slippers²,⁴; Samantha-Leigh Jamison-Daniels¹; Frederick Witfeld³; Monique Botha¹; Dominik Begerow³; Andreas Brachmann⁵ & Michelle Greve¹,²

1 – Department of Plant and Soil Science, University of Pretoria, Private Bag X20, Pretoria, 0002, South Africa
2 – Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0002, South Africa
3 – Evolution of Plants and Fungi, Ruhr-Universität Bochum, Universitätsstr. 150, 44801 Bochum, Germany
4 – Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X20, Pretoria, 0002, South Africa
Patterns and drivers of succession provide insight into the mechanisms that govern community assembly and are indicators of community resilience and stability, but are still poorly understood in microbial communities.

We assessed whether the successional trends of woody vegetation are mirrored by foliar fungal endophyte communities of three tree species that are abundant across the woody successional gradient using a total amplicon sequencing approach. Additionally, we test the relative contribution of host identity, abiotic predictors, biotic factors, and spatial distance between sites in predicting community composition and species richness of endophyte communities. Unlike the woody community, endophyte communities showed no consistent evidence of deterministic successional trends. Host identity was the most important factors structuring fungal endophyte community composition. Spatial distance played some role in explaining differences in community composition, but effects of this and other environmental variables were small and not consistent between different host species. Much of the variation in endophyte composition remained unexplained. Host identity was most important in predicting endophyte richness. Although endophyte communities showed no deterministic succession, community assembly was most strongly influenced by host identity and spatial distance.

Key words
Community assembly, Ecology, Fungi, Microbiome, Succession, Symbiosis, Trees

Introduction
Understanding the relative contribution of deterministic and stochastic processes during community assembly represents a major unresolved issue in microbial ecology (Dini-Andreote et al., 2015; Antwis et al., 2017; Zhou and Ning, 2017; Tripathi et al., 2018). Unravelling the mechanisms that drive shifts in community composition is crucial for understanding the functions and ultimately the ecosystem processes that microbial communities are able to deliver (Salles et al., 2009; Crowther et al., 2014; Laforest-Lapointe et al., 2017). Succession, the study of how biological communities reorganise through time after disturbances (Johnson, 1979; Chang and Turner, 2019), provides valuable insight into community assembly (Chang and HilleRisLambers, 2016). Understanding succession in microbial communities can, therefore, help to disentangle the relative importance of deterministic and stochastic processes on community assembly (Tripathi et al., 2018).

The composition of communities during succession can be driven by deterministic mechanisms (environmental conditions and biotic filters) and stochastic processes (dispersal, drift, speciation and priority effects) (Hubbell, 2001; Vellend, 2010; Nemergut et al., 2013; Chang and HilleRisLambers, 2016). For microbial communities it has been hypothesised that during the early stages of succession, communities are largely governed by stochastic events, and only as succession continues does the relative strength of deterministic processes begin to cause directional changes in community composition (Dini-Andreote et al., 2015,
Studies on fungal succession have mostly focused on belowground soil and root-associated communities (Blaalid et al., 2012; Brown and Jumpponen, 2014; Davey et al., 2015; Dini-Andreote et al., 2016; Dong et al., 2016; Turner et al., 2019). These studies have repeatedly shown that fungal community assembly during later successional stages is deterministic (Brown and Jumpponen, 2014; Dini-Andreote et al., 2016; Gao et al., 2019; Turner et al., 2019). Both abiotic (e.g., carbon, nitrogen, pH and phosphorus) and biotic (e.g., plant species richness, composition, root exudates and microbial competition) factors are drivers of belowground microbial community composition patterns (Dini-Andreote et al., 2016; Dong et al., 2016; Turner et al., 2019). While some evidence exists for both linked (Davey et al., 2015) and decoupled (Turner et al., 2019) trends of succession between plant communities and belowground fungal communities, generalisable patterns across different fungal guilds are still unclear.

Disentangling the relative influence that host-associated and environmental factors play in determining host-endophyte composition represents a major theme in endophyte ecology (Antwis et al., 2017; Harrison and Griffin, 2020), as these fungi are an important aspect in plant performance and health (Compant et al., 2019). Host identity appears to be one of the most important drivers of foliar fungal endophyte community composition (Terhonen et al., 2019). Host-specific defence mechanisms, the host-specific production of various enzymes, secondary metabolites and concentration differences of nutrients and molecules within their leaves directly influence the composition of foliar endophyte communities (Kembel and Mueller, 2014; Cordovez et al., 2019; Darlison et al., 2019; Tellezet et al., 2020). Geographic distance is another important factor influencing community composition, with community dissimilarity generally increasing with geographic distance due to dispersal limitation (Soininen et al., 2007). Evidence for distance decay in foliar fungal endophytes is mixed. Some evidence points to the absence of such a relationship at both small (Cordier et al., 2012; Oono et al., 2017) and large scales (Vincent et al., 2016; Barge et al., 2019; U'Ren et al., 2019), while other studies show evidence of distance decay, especially for rare foliar fungal endophyte taxa (Vaz et al., 2014; David et al., 2016; Koide et al., 2017; Oono et al., 2017). Climate, a major factor driving the community composition of plants, seems to have less influence on determining the composition of fungal endophytes, particularly at fine scales (Compant et al., 2010; Santoyo et al., 2017). Water availability in particular affects some fungal endophytes’ ability to germinate and persist (Arnold, 2007; Peay et al., 2016). The effect of climate on endophyte composition may also be indirect: through its effect on host composition and physiology fungal endophyte community composition may be affected (Compant et al., 2010; Terhonen et al., 2019).

Plant community composition structures soil and root-associated fungal communities due to the strong biotic filter imposed by plant hosts (Carney and Matson, 2006; Hausmann and Hawkes, 2009; Hoch et al., 2019; Hu et al., 2019). However, how plant community composition structures fungal endophyte composition is yet to be directly assessed (Griffin and Carson, 2018; Griffin et al., 2019). Therefore, directly linking plant community composition to foliar fungal endophyte composition may help to disentangle the factors responsible for structuring these fungal communities and ultimately help to understand how this scales-up to mediate plant microbial ecosystem functioning relationships (Laforest-Lapointe et al., 2017; Griffin et al., 2019; Harrison and Griffin, 2020).

In contrast to community composition, species richness considers only the number of species. A number of factors affect foliar fungal endophyte species richness (Arnold and Lutzoni, 2007; Unterscher et al., 2007; Lau et al., 2013; Griffin et al., 2019; Harrison and Griffin, 2020). Host identity appears to be one of the most important drivers (Lau et al., 2013; Peay et al., 2016; U’Reen et al., 2019; Yao et al., 2019). Other factors, including host age or height, microhabitat (e.g., moisture, light intensity and temperature) and plant richness, can also shape foliar fungal endophyte richness (Bernstein and Carroll, 1977; Unterscher et al., 2007; Zimmerman and Vitousek, 2012; Scholtysek et al., 2013; Oono et al., 2015; Griffin et al., 2019). Yet, few studies have simultaneously assessed how multiple factors drive patterns of foliar fungal endophyte richness.
The aim of this study is to assess patterns and drivers of foliar fungal endophyte community assembly in a system undergoing deterministic vegetation succession. Our objectives were to a) test whether foliar fungal endophyte communities follow a deterministic successional trend as displayed by the woody tree communities in the same study system; and b) determine the relative contribution of host species identity, surrounding tree composition, geographic distance and abiotic variables to endophyte composition and richness.

**Experimental procedures**

**Study area**

The study was conducted at Buffelskloof Private Nature Reserve (25°19′22.21″ S, 30°29′15.41″ E), Mpumalanga, South Africa. Buffelskloof supports vegetation types from three biomes: Afromontane forest, mid-altitude savanna and montane grassland (Mucina and Rutherford, 2011). The reserve is approximately 1500 ha in size with an altitudinal variation of 1000 to 1800 m (Buffelskloof Nature Reserve, 2019). Samples were collected between 1314 and 1477 m elevation.

Within the grassland and savanna vegetation on Buffelskloof, bush clumps (BCs) have been establishing (Jamison-Daniels et al., 2021). A BC is defined as an association of one or more trees (> 1.2 m) with tree saplings (< 1.2 m) growing beneath the canopy, that are separated from each other by grassy vegetation (Jamison-Daniels et al., 2021). A space-for-time substitution study, in which the woody vegetation of 40 BCs ranging in size from 10.05 m² to 1342.99 m² (mean = 266.41 m², SE = 347.75 m²) was characterised, where it was found that the formation of these BCs follows a deterministic trend of succession in woody species, with a turnover from open-habitat trees to shade-tolerant trees as the BCs increase in size (Jamison-Daniels et al., 2021). These directional changes in tree community composition are driven by directional changes in microclimatic conditions (temperature, soil moisture and humidity) and light availability as BCs increase in size (Jamison-Daniels et al., 2021).

**Sampling**

Endophyte communities were sampled from the same BCs that were surveyed by Jamison-Daniels et al., (2021). Three tree species that are widely distributed and common across BCs of different sizes (and thus different successional stages) (Jamison-Daniels et al., 2021) were selected: *Euclea crispa* subsp. *ovata* (Burch.) F.White, *Searsia chirindensis* (Baker f.) Moffett and *Canthium inerme* Kuntze. Trees of these species were present in 38 of the 40 BCs.

Field sampling took place between 20 and 23 November 2018, to characterise the summer leaf endophyte communities. In every BC containing the host species, up to four trees (> 1.2 m) per host species were selected. The coordinates of each tree that was sampled were recorded using a Garmin Etrex GPS (Garmin Ltd., Olathe, KS, USA) and the height of each tree was estimated using a 1.2m dowel stick. All hosts did not occur in all BCs, and BCs did not always contain four individuals from each host. For BCs that had more than four individuals of a host, the first four individuals that were encountered were sampled. *E. crispa*, *C. inerme* and *S. chirindensis* trees were present in 27, 21 and 28 of the BCs, and 84, 50 and 55 individual trees were sampled from each host species, respectively. From each tree, five leaves from each of the four cardinal directions were collected half-way between the highest and lowest leaves. Only fully unrolled leaves that had reached maturity and had no visible signs of infection or insect damage were selected. Leaves were stored in envelopes within a cooler box, and processed within 8 hours of collection.

Microclimatic conditions within each BC were characterised to determine whether differences in microclimate influenced endophyte community composition and richness (see Jamison-Daniels et al., 2021 for full details). Light intensity, temperature (average, minimum, maximum and standard deviation measured over 130 days), relative humidity and soil moisture were all previously measured and subsequently extracted (from Jamison-Daniels et al., 2021).

The species richness (count of tree species per BC), Shannon-Wiener diversity, species composition and the tree basal area of all trees >1.2 meters per BC were obtained from Jamison-Daniels et al., (2021). Only trees >1.2 meters were considered, as these represent established vegetation within the BCs. Additionally,
proximity of each sampled tree to the large indigenous forest which acted as a source population for most BC trees (Jamison-Daniels et al., 2021) was calculated to represent a proxy for potential endophyte inoculum pressure. Within Google Earth Pro (v7.3.2.5776) a polygon was drawn around the indigenous forest at Buffelskloof Nature Reserve and subsequently extracted as a shape file. The distance from each tree sampled to the edge of the indigenous forest within the reserve was obtained using the points to polygon function in Esri, ArcMap (Esri, Redlands, CA, USA).

Leaves underwent surface washing (Arnold and Herre, 2003) to reduce and possibly eliminate the epiphytic burden from each sample. All leaves sampled from one tree were washed together successively in 70% EtOH (30 seconds), 2% NaOCl (60 seconds), 70% EtOH (60 seconds) and autoclaved dH₂O (60 seconds). The leaves were then placed on paper towel and left to dry. Leaf disks from the dried leaves were cut using a 6 mm cork-borer (sterilised between each sample) and subsequently stored on silica gel in falcon tubes until DNA extraction. For E. crispa and C. inerme 12-18 leaf disks were cut per leaf, while 22-25 disks were cut from larger S. chirindensis leaves.

**Homogenisation, DNA extraction and sequencing**

Prior to homogenisation of each sample, a metal cylinder and ball bearing were sterilised in a 6% sodium hypochlorite (NaOCl) solution for three minutes, followed by immersion in 70% ethanol (EtOH) for one minute. All leaf disks from one sample were homogenised together in the metal cylinder fastened to a Retsch® MM2000 laboratory mixer mill (Retsch® GmbH, Haan, Germany) for one minute at 70% of the maximum oscillation frequency and subsequently stored at -200C until DNA extraction. DNA of 60 mg dried homogenised leaf material per sample was extracted using the my-Budget plant DNA extraction kit (Bio-Budget Technologies GmbH, Krefeld, Germany) following manufacturer’s instructions. The ITS region of all extracted samples was amplified using the ITS1-F (Gardes and Bruns, 1993) and ITS4 (White et al., 1990) primer combination and was subsequently visualised on a 0.8% agarose gel with a 100bp ladder, together with positive and negative controls to ensure that we had indeed extracted fungal DNA without contamination.

Illumina amplicon library preparation utilised a nested PCR approach (Unterseher et al., 2016) (Supporting Information Figure 1, Methods S1). The ITS region was amplified using the ITS1-F (Gardes and Bruns, 1993) and ITS4 (White et al., 1990) primer combination. The final amplicons were subsequently visualised on a 0.8% agarose gel with added ethidium bromide (Supporting Information Methods S1). Concentration of DNA for pooling was quantified from the band intensity of the gel images using ImageJ (Schneider et al., 2012) (Supporting Information Methods S1). The amplicon pools were cleaned using the CleanPCR magnetic bead kit (CleanNA, Waddinxveen, Netherlands) (Supporting Information Methods S2). In total three samples were lost during homogenisation or due to low DNA concentration resulting in 186 samples that were sent for sequencing.

The final pooled amplicon library was sequenced at the Genomics Service Unit of the Ludwig Maximilians University (LMU) Biocenter on an Illumina MiSeq® sequencer (Illumina Inc., San Diego, CA, USA) using the MiSeq® Reagent Kit v3 Chemistry, for 2 × 250 bp paired-end sequencing. All raw sequences were deposited on the NCBI portal under the following accession codes: BioProject – PRJNA674320; BioSample (SAMN16634781 – SAMN16634966).

**Bioinformatics**

Bioinformatic processing was performed in QIIME1 (Caporaso et al., 2010) and QIIME 2 (Bolyen et al., 2019). Samples were demultiplexed according to their unique tag/index combinations (Supporting Information Table S1), which were removed during the process, together with the primer sequences (Supporting Information Table S2). For subsequent analyses only forward reads were used, as reverse reads often suffer from a lower PHRED quality and due to length differences within the ITS gene region, which often prevents merging of both reads. Three S. chirindensis samples which contained less than 10 000 sequences were removed prior to downstream bioinformatic analyses. The raw sequences from the remaining 183 samples were subsequently passed through deblur (Amir et al., 2017) implemented in the QIIME 2 pipeline, which assigns
raw sequence reads to Amplicon Sequence Variants (ASVs). Reads were trimmed at 180 bp. The UNITE database was used as reference sequences (version 8; 020219) (https://unite.ut.ee/). Only ASVs which were classified as belonging to the Kingdom Fungi were retained. Fungal ASVs were written into a feature table, which was used for subsequent downstream analyses. The full ASV feature table and all metadata relating to the manuscript can be obtained from figshare: 10.6084/m9.figshare.14518200 – ASV feature table and 10.6084/m9.figshare.14518218 – metadata.

Analyses

All analyses were performed on the full, unrarified, ASV table. This was done for two reasons. First, rarefying the ASV table to the smallest sample size to account for differences in library sizes between samples made no difference to the interpretation of the results (results not shown). Second, from a statistical point of view, rarefaction is inept for the comparison of relative abundances (McMurdie and Holmes, 2014; Willis, 2019).

As predictor variables that are highly correlated can lead to spurious effects on analyses, all continuous predictor variables were tested for multi-collinearity prior to analyses (Supporting Information Table S3). When two variables were highly correlated, i.e. \( r > 0.75 \), one of these variables was removed (Supporting Information Table S3). Bush clump area, bush clump tree basal area and bush clump tree species richness were highly correlated. Therefore, only bush clump tree basal area was retained for analyses on endophyte composition and richness, as bush clump tree basal area gives a good representation of available woody tree host density within individual BCs. Bush clump area was only retained in the analyses on successional trends, as BC area was a good proxy for BC maturity and woody vegetation successional stage (Jamison-Daniels et al., 2021).

Successional trend

A two-step approach was used to determine the processes of endophyte community assembly and to evaluate whether these fungal communities followed the deterministic successional trend observed for their woody tree hosts (Jamison-Daniels et al., 2021).

First, we assessed whether the amount of variation in community composition of our actual endophyte assemblages, explained by a number of predictor variables, was greater than (implying deterministic processes), less than (implying deterministic processes) or not different (implying stochastic processes) to what could be expected in randomly generated communities (Dini-Andreote et al., 2015). Initially, a canonical correspondence analysis (CCA) was conducted to test which factors, namely host identity, spatial distance between sampled trees, abiotic conditions (maximum and minimum temperature, temperature standard deviation, light intensity, soil moisture measured per BC) and biotic factors (tree basal area per BC, tree height and distance to forest edge), best explained fungal community composition. Weighted linear regression was performed on the constraining variables (Ter Braak, 1986). To incorporate the effect of spatial distance on community composition, a weighted principal coordinates of neighbourhood matrix (PCNM) analysis (created using the geographic coordinates of every sampled tree) was performed. This analysis transforms coordinates to a rectangular distance matrix that is acceptable for constrained ordination techniques like CCA (Borcard and Legendre, 2002; Legendre and Borcard, 2008). Backward and forward stepwise permutation tests, for 1000 permutations, were used to determine the best fitting model, based on the model with the lowest AIC score (Venables and Ripley, 2002). Rare species contribute heavily to the chi-squared distance used to plot site and species scores in CCA analysis (Legendre and Legendre, 2012). Therefore, to reduce the spurious effects of rare species within the CCA, an eigenvalue decomposition approach was used to determine in how many samples a particular ASV must have occurred for it to be retained when performing the final ordination (Legendre and Legendre, 2012). When a considerable drop in inertia for one of the first five eigenvalues is detected, it indicates in how many samples an ASV must have occurred to be retained for the final CCA analysis (Supporting Information Figure S2) (Legendre and Legendre, 2012). The eigenvalue decomposition detected a drop in the inertia for the third eigenvalue after dropping ASVs that occurred in less than 20 samples (Supporting Information Figure S2). Therefore, all ASVs \( (n= 4905) \) that occurred in 20 samples or less were removed before the final CCA analysis (N= 411) and the randomisations.
The amount of variation explained by the best model with the retained predictor variables (host identity, light intensity and four spatial eigenvectors (spatial eigenvector 1, spatial eigenvector 2, spatial eigenvector 13 and spatial eigenvector 17)) was calculated. Then, 10,000 randomly assembled abundance-weighted community data matrices were constructed based on the true community identity, richness and abundance (Greve et al., 2008). The randomly assembled matrices conserved species richness per sample as observed in the true community, and set the probability of species being selected proportional to ASV read abundance (Gotelli and Graves, 1996; Gotelli, 2000). For each of the 10,000 randomly assembled communities, a new CCA was performed with the predictors from the best CCA model (see above). The 2.5% and 97.5% quantiles of the percentage variation explained by the CCAs were calculated for the 10,000 random communities, and it was assessed whether the percentage variation explained by the CCA of the true community was larger than (> 97.5% quantile), smaller than (< 2.5% quantile) or not significantly different (between 2.5% and 97.5% quantiles) to the percentage variation explained by the randomly generated communities using a z-test (Greve et al., 2008).

Because CCA randomisation analyses provided evidence for deterministic community assembly (see Results), a second analysis was conducted to test whether changes in community composition could be explained by BC size, as BC size increases with tree succession (Jamison-Danielset al., 2021). A significant directional change in endophyte species composition with BC size would indicate deterministic succession, while no predictable change in composition with BC size would suggest stochastic succession (Dini-Andreote et al., 2015). We tested this for the fungal communities extracted from each of the three host species using two different pair-wise similarity indices (Morisita and Raup-Crick indices). The Morisita index is weighted towards assessing similarity in common taxa, while the Raup-Crick index gives more weight to co-occurring rare taxa (Morisita, 1962; Raup and Crick, 1979). Pair-wise similarity in fungal community composition was calculated between all possible combinations of each of the smallest \( \frac{1}{4} \) of the BCs and each of the largest \( \frac{3}{4} \) of the BCs for each host species (following Jamison-Danielset al., 2021), to establish if there was directional change in community composition as BCs increase in size. For each of the largest \( \frac{3}{4} \) BCs, similarity values with the smallest \( \frac{1}{4} \) of the BCs were averaged. Generalised linear mixed effects models (GLMMs) were used to model the effect of BC area (of the largest \( \frac{3}{4} \) BCs) on foliar fungal endophyte community similarity for both Raup-Crick and Morista similarity indices. BC identity of the largest \( \frac{3}{4} \) BCs was included as a random effect in the model (McCulloch, 1997). Since the similarity values for both the indices are scaled between 0-1, models were fitted using a binomial distribution and logit link function (Zuur et al., 2009). If the foliar fungal community similarity between the largest and the smallest BCs decreased or increased significantly with BC area of the largest BCs, it was interpreted as an indication of deterministic succession, while no relationship was taken as an indication of communities being ecologically neutral and primarily governed by stochastic processes that structure community composition (Hubbell, 2001; Dini-Andreote et al., 2015; Jamison-Danielset al., 2021).

Drivers of assemblage composition

We tested which factors, namely host identity, spatial distance between sampled trees, abiotic conditions (maximum and minimum temperature, temperature standard deviation, light intensity and soil moisture measured per BC) and biotic factors (tree basal area per BC, tree height, distance to forest edge and the tree composition per BC), affected fungal endophyte community composition.

To do this we used a generalized dissimilarity modelling (GDM) approach. This flexible approach enabled us to simultaneously incorporate categorical (i.e. host), linear (e.g. climate), compositional (i.e. tree compositional dissimilarity), and spatial data (i.e. geographic distance) predictors of endophyte composition into one analysis (Ferrier et al., 2007). GDM is a nonlinear extension of matrix regression, which has specifically been designed to deal with two types of nonlinearity commonly encountered in biological data: 1) the curvilinear relationship between ecological or spatial separation and the observed compositional dissimilarity, and 2) non-stationarity, i.e. differences in the rate of compositional turnover along environmental or spatial gradients (Ferrier et al., 2007; Fitzpatrick et al., 2013).

The default of three I-spline basis functions (knots) per predictor variable was used in all GDM analyses
Drivers of endophyte richness

To test the effects of host identity (i.e. host species), abiotic variables (maximum and minimum temperatures, temperature standard deviation, light intensity, relative humidity and soil moisture) and biotic variables (tree basal area per BC, tree height and distance to the forest edge) on ASV richness, GLMMs, with a Poisson distribution and a log-link function were used (Zuur et al., 2009). Interaction terms between host identity and all abiotic and biotic variables were fit within the model and BC identity was included in the model as a random variable (McCulloch, 1997). Best subset modelling, based on the lowest AIC-value, was used to assess which predictor variables from the global model should be retained (Burnham and Anderson, 2002). The model was overdispersed, therefore overdispersion was corrected by employing the observation-level random effects approach (Lawson et al., 1999; Elston et al., 2001). Marginal and conditional R²-values were calculated (Nakagawa et al., 2017).

Statistical software

All analyses were conducted in R, v3.6.0 (R Core Team, 2019) using the packages vegan v2.5-6 (Oksanen et al., 2019), lme4 v1.1-23 (Bates et al., 2015), MuMIn v1.43.17 (Barton, 2019), car v3.0-9, gdm v1.4.2.2. (Fitzpatrick et al., 2021), and effects v4.2-0 (Fox, 2003; Fox and Weisberg, 2019).

Results

The 183 leaf samples from three native host species yielded 7 355 098 demultiplexed sequences. On average there were 40 192 sequences per sample, the highest number of sequences from one sample was 103 727 and the lowest 12 533 (Supporting Information Table S5). In total 5 326 unique ASVs were recovered with an average of 193.80 ± 101.82 ASVs per sample. The number of ASVs recovered from C. inerme, E. crispa and S. chirindensis was 1 473, 2 221 and 1 847, respectively.

Successional trend

Variables retained in the CCA (host identity, spatial distance between samples and light intensity) explained 11.76% of the variation of the actual endophyte community. In contrast, only between 4.27% - 4.75% of the variation in ASV composition of randomly generated communities could be explained by these same predictor variables. Therefore, the predictor variables predicted significantly (2.5 times) more variation in community composition of the true than the randomly generated communities (z-value = 56.812, p-value = <16x10^-16). This indicates that communities are non-randomly assembled and that deterministic selective forces structure fungal endophyte communities.

The similarity of foliar fungal communities showed no consistent trends across the gradient of BC size (Table 1). The average Raup-Crick similarity decreased significantly with increasing BC area for endophyte communities of E. crispa (Figure 1 & Table 1), as expected under deterministic succession. However, no significant trends in community similarity with BC area could be observed for rare communities of the other two host species (Table 1). Similarly, the Morisita similarity index showed no host-specific successional trends across BC area (Table 1). Therefore, five of the six analyses showed no evidence of deterministic succession.

Assemblage composition

The final GDM for the full dataset retained two host species (E.crispa and S. chirindensis), tree basal area, distance to the indigenous forest edge, geographic distance between samples, light intensity and tree compositional dissimilarity as predictor variables (Figure 2). The final model for the full dataset was significant (Null Deviance = 691.996, GDM Deviance = 491.635, p -value = 0.000001), as were all the predictor
variables, except light intensity (Figure 2; Table 2). The final model explained 28.95% of the deviance in the turnover of foliar fungal endophyte community composition. With host identity alone explaining ~23% of this variance. Of the remaining predictors distance to the forest and the difference in tree basal area were the most important, both having variable rates of turnover along the gradients and displaying the quickest turnover at short distances from the forest and small differences in tree basal area (Figure 2c & d). The rate of turnover in fungal endophyte composition along the tree compositional dissimilarity gradient was the steepest of any of the gradients; with the maximum magnitude of turnover reached when the difference in surrounding tree composition was only slightly dissimilar, i.e. samples not from the same BC (Figure 2f).

When GDMs were run per host species, different predictors were retained as important in explaining the composition of endophytes within different hosts. Only geographic distance between samples was retained and significant in all models (Figure 2 & Figure 3). The final GDM model for *E. crispa* was significant (Null Deviance = 147.69, GDM Deviance = 122.65, p-value = 0.000001) and explained 16.96% of the deviance in turnover of foliar fungal endophyte community composition. It retained five variables (Supporting Information Table S5), of which distance to the forest edge, geographic distance between samples and the difference in tree basal area were significant (Figure 3a-c). Distance to the forest was the most important predictor, with the rate of turnover being approximately linear (Figure 3a). The rate of turnover in endophyte composition along the geographic distance and tree basal area gradients were non-linear with the highest rates of turnover occurring at larger geographic separation between samples and small differences in tree basal area (Figure 3b & 3c).

For *C. inerme*, the final GDM model retained five predictor variables (Supporting Information Table S6); however, only minimum temperature and geographic distance between samples were significant (Figure 3d & 3e). Minimum temperature was the most important predictor with both larger differences in minimum temperature and geographic distance between samples resulting in higher rates of turnover. The magnitude and rates of turnover in endophyte composition was almost non-existent at low differences in minimum temperatures and short geographic distances between samples but increased dramatically at the higher end of these gradients (Figure 3d & 3e). The final model was significant (Null Deviance = 52.85, GDM Deviance = 45.92, p-value = 0.00503), and explained 13.12% of the deviance in turnover of foliar fungal endophyte community composition.

The final GDM model for *S. chirindensis* fungal communities retained five predictor variables (Supporting Information Table S7); but the only retained variable which was significant was geographic distance between samples (Figure 3e). The rate of turnover in endophyte composition increased with distance between samples; this increase was most rapid at short geographic distances (Figure 3e). The final *S. chirindensis* GDM model was significant (Null Deviance = 42.78, GDM Deviance = 38.622, p-value = 0.000001), and explained 9.52% of the deviance in turnover of foliar fungal endophyte community composition.

### ASV richness

Host was the most important predictor of ASV fungal richness (Table 3), with one host – *S. chirindensis* – consistently supporting fewer ASVs (mean = 83.94 ± 52.79) than *C. inerme* (mean = 218.5 ± 95.8) and *E. crispa* (mean = 245.85 ± 70.82) (Figure 4). The best subset GLMM model testing the effect of abiotic, biotic and spatial factors on ASV fungal endophyte richness retained light intensity; the interaction between host species and tree height and the interaction between host species and maximum temperature. Both the interactions between host and maximum temperature, and between host and tree height were significant (Figure 4, Table 3). Fungal ASV richness increased as light intensity decreased (Figure 4a). ASV endophyte richness decreased in *S. chirindensis* individuals that experienced higher maximum temperatures (Figure 4b) but was unaffected by temperature differences experienced by individuals of *C. inerme* and *E. crispa* (Figure 4b). Additionally, ASV endophyte richness decreased with tree height for *S. chirindensis* but did not change with tree height in *C. inerme* and *E. crispa* (Figure 4c). Fixed effects explained 63% of variation in ASV richness (Table 3).

### Discussion
Foliar fungal endophyte communities did not display consistent patterns of deterministic succession within individual host species, despite the surrounding plant community undergoing deterministic successional changes. Instead, host species identity was the most important factor shaping community composition and richness patterns for endophytes. The strong effect imposed by the host potentially limited the ability to observe consistent deterministic succession patterns across the successional gradient, i.e. BC size and its associated changes in abiotic conditions.

**Successional patterns**

Across the common and rare endophyte assemblages of three host tree species, evidence for deterministic succession was only found for the rare endophyte communities of one of the three woody host species (*E. crispa*) and in none of the three assemblages of common endophyte species. Therefore, we find no consistent evidence of deterministic succession of endophyte communities within individual hosts across a woody-plant successional gradient. Under a scenario of deterministic succession, similar directional shifts in fungal endophyte community composition would have been observed since abiotic environmental conditions changed fairly consistently with BC size (Jamison-Daniels *et al.*, 2021). The lack of consistent directional changes in endophyte composition with BC size, but significant effects of the host identity on species composition, suggest that the effect of the plant host on endophyte composition trumps the impact that the consistent changes of environmental variables has on fungal endophyte communities. A large percentage of variation in the community assembly models remain unexplained, suggesting that unmeasured or random processes may contribute to the assembly of endophyte communities (Hubbell, 2001).

The lack of strong evidence for deterministic succession within endophyte communities does not preclude succession of endophyte communities of all trees per bush clump. Our study specifically assessed changes in endophyte composition within single host species across a tree successional gradient. Because tree composition changed in a directional manner with BC size (Jamison-Daniels *et al.*, 2021), and host identity is an important determinant of endophyte composition and richness (Liu *et al.*, 2019; U’Ren *et al.*, 2019; Harrison and Griffin, 2020) (Figure 2; Figure 4; Table 1), endophyte communities of entire BCs (i.e. across different tree species or within tree genera or families) could show stronger evidence of deterministic changes in endophyte composition. As a certain suite of tree species are associated with a specific successional stage in this system (Jamison-Daniels *et al.*, 2021), their associated endophyte communities may also be associated with the associated tree successional stage. Under this scenario, the succession of endophyte communities would be driven by changes in host species composition with successional stage, and not changes in environmental factors with successional stage.

**Drivers of community composition**

Although no consistent evidence of deterministic succession in endophyte communities was observed; there was some evidence of deterministic processes shaping the community assembly of fungal endophyte communities. Host identity had the largest impact on endophyte community composition (Figure xx, Table xx). Host identity alone explained ~23% of the total ~29% variation in turnover of endophyte composition explained by the final model. Host identity is known to be a major determinant of the composition of foliar fungal endophyte communities (Christian *et al.*, 2016; Davidet *et al.*, 2016; Vincent *et al.*, 2016; Liu *et al.*, 2019; Yao *et al.*, 2019). The specificity of endophytes to a particular host or group of taxonomically related hosts ranges from extremely narrow to relatively wide (Arnold and Lutzoni, 2007; Pölme *et al.*, 2018). This may explain why, even though host identity was the important factor structuring foliar fungal endophyte communities, the identity of only two host species was retained in the final model, with *C. inerme* consistently being removed as one of the variables which explained the least turnover in endophyte community composition.

Geographic distance between samples was the only significant variable consistently retained in every GDM (i.e. the full dataset and each of the three hosts individually) (Figure 2 & Figure 3)). This indicates that the similarity in composition of foliar fungal endophytes decays with increasing distance, a pattern commonly observed in ecological communities (Nekola and White, 1999; Soininen *et al.*, 2007). Indeed, others have
shown that foliar fungal endophytes are dispersal limited over distances from 10km to >100km (David et al., 2016; Koide et al., 2017; Oono et al., 2017). However, here we show that even over a shorter spatial distance, i.e. <1500m, there is consistent evidence for distance decay driving turnover in endophyte community composition.

The GDMs performed for each host individually, showed that the factors responsible for the turnover in foliar fungal endophyte community composition within a host is variable as is the rate and magnitude of this turnover (Figure 3, Supporting information Table S5-S7). For example, distance to the indigenous forest, the difference in minimum temperature and geographic separation were the most important variables explaining the turnover in foliar fungal endophyte communities for E. crispa, C. inerme and S. chirindensis respectively, holding all other variables constant in their individual models. Others have shown how the distance to a potential inoculum source, in our case a large indigenous forest, can be important for causing compositional differences in fungal communities (Glassman et al., 2017). It has also been shown how abiotic conditions, like temperature differences, can shape endophyte community composition due to differences in the physiological performance of endophyte species in dealing with the prevailing conditions (Arnold and Herre, 2003; Peay et al., 2016; Unterseher et al., 2016). Such findings highlight how both deterministic (i.e. temperature differences) and stochastic (i.e. distance to an inoculum source or geographic separation) forces can be important factors causing turnover in foliar fungal endophyte communities in the same system but for different hosts.

Endophyte richness

Endophyte richness was most strongly influenced by host identity, with endophyte richness of one host (S. chirindensis) being consistently lower than that of the other hosts. In contrast, several abiotic and biotic variables showed inconsistent effects on endophyte richness across the three host species. Reduced endophyte richness at higher light intensity agrees with other studies: high levels of UVB negatively impact endophyte persistence, with high UVB increasing leaf desiccation and/or the activation of the plant’s defence responses, ultimately leading to lower endophyte richness (Unterseher et al., 2007, 2012).

Other abiotic effects on endophyte richness differed between host species which may shed some light on how such factors drive endophyte richness in these host species. Lower endophyte richness in S. chirindensis occurred in warmer BCs that this species occupied. Temperature differences within BCs may drive differences in chemistry and secondary metabolite production produced by the leaves of S. chirindensis that experience higher temperatures (Reich et al., 1999; Veteli et al., 2002), potentially leading to lower ASV richness (Arnold and Herre, 2003; Unterseher et al., 2012, 2013). High temperatures can activate the plant defence response, thereby increasing the production of secondary metabolites with potential anti-microbial properties (Unterseher et al., 2016) which decreases the observed endophyte richness by reducing the number of successful endophyte colonisations. Taller trees are usually older, and older hosts have been shown to support decreased endophyte richness compared to their younger conspecifics (Oono et al., 2015). It has been postulated that older trees invest more in defence mechanisms which resist endophyte colonisation; alternatively established groups of endophytes in older trees outcompete newly arriving endophytes, ultimately leading to lower endophyte richness in older trees (Unterseher et al., 2007; Oono et al., 2015).

Conclusion

Although foliar fungal endophyte community assembly within this natural system was structured by deterministic selective forces (particularly host identity), there was no consistent evidence of deterministic successional trends in endophyte communities. A fair amount of variation in endophyte community composition remained unexplained. This could be due to several reasons. Variables which were not measured here, e.g. competitive interactions between endophyte species (Crowther et al., 2014), could be more important for structuring endophyte communities. Alternatively, stochastic processes of community assembly may have been more important than deterministic processes in structuring foliar fungal endophyte composition (Hubbell, 2001). These stochastic processes include drift, dispersal limitation, speciation and extinction events acting together to shape the observed community composition (Vellend, 2010). Additionally, the scale at
which endophyte communities were quantified may not accurately reflect the scale at which processes of endophyte community assembly are operating and thus ultimately obscure our ability to detect strong patterns (Harrison and Griffin, 2020). Thus, continued work to gain a deeper understanding of how fungal endophyte communities organise themselves through space and across time within natural systems will be essential to appreciate the functions and ultimately the ecosystem services these microbes are able to deliver.

Acknowledgements

The authors thank Buffelskloof Private Nature Reserve, more specifically Mr John Burrows, for providing them with access to the research site, and to Mpumalanga Tourism and Parks Agency for permits to collect plant material. A special thanks to Mandy Messal and Laura Milhe for their help during the collection and processing of samples collected. The authors would also like to thank the Tree Protection Co-operative Program (TPCP), the centre of excellence in plant health and biotechnology (CPHB) and the South African National Research Foundation (Grant numbers 98889 and 116333) for funding this research.

Author Contributions

MAH, MK, BS & MG contributed to the design of the research, performance of the research, interpretation of the data and writing of the manuscript. MAH, S-L J-D & MB contributed to the data collection and data analyses of the research. MAH, FW, DB & AB contributed to the performance of the laboratory work, sequencing and bioinformatics of the fungal communities.

Data Availability

All raw Illumina sequence data are available on the NCBI portal under the following accession codes: Bio-Project – PRJNA674320; BioSample (SAMN16634781 – SAMN16634966). The full ASV feature table can be obtained from figshare: 10.6084/m9.figshare.14518200. All metadata relating to the manuscript can be obtained from figshare: 10.6084/m9.figshare.14518218. All R scripts relating to the project can are available upon request to the corresponding author.

Competing Interests

We know of no conflicts of interest that are associated with this potential publication. Additionally, we declare that there has been no significant financial support for conducting this research that could have influenced its overall outcome. As the corresponding author, I confirm that the entire manuscript with all supplementary materials has been read and approved for submission by all authors named above.

<table>
<thead>
<tr>
<th>Host species</th>
<th>Similarity index</th>
<th>Fixed effects</th>
<th>Estimate</th>
<th>SE</th>
<th>z-value</th>
<th>p-value</th>
<th>Conditional R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclea crispa</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>-1.3301</td>
<td>0.4968</td>
<td>-2.677</td>
<td>7.43X10^-3</td>
<td>0.2185</td>
</tr>
<tr>
<td>Euclea crispa</td>
<td>Morisita</td>
<td>Area</td>
<td>-0.16368</td>
<td>0.1528</td>
<td>-1.071</td>
<td>2.84X10^-1</td>
<td>0.00133</td>
</tr>
<tr>
<td>Canthium inerme</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>-0.09949</td>
<td>0.594</td>
<td>-0.167</td>
<td>8.67X10^-1</td>
<td>0.00013</td>
</tr>
<tr>
<td>Canthium inerme</td>
<td>Morisita</td>
<td>Area</td>
<td>-0.6755</td>
<td>0.3599</td>
<td>-1.877</td>
<td>6.05X10^-2</td>
<td>0.00032</td>
</tr>
<tr>
<td>Searsia chirindentis</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>0.0064</td>
<td>0.0051</td>
<td>1.248</td>
<td>2.12X10^-1</td>
<td>0.00013</td>
</tr>
<tr>
<td>Searsia chirindentis</td>
<td>Morisita</td>
<td>Area</td>
<td>0.0001</td>
<td>0.00031</td>
<td>0.323</td>
<td>7.47X10^-1</td>
<td>5.13X10^-4</td>
</tr>
</tbody>
</table>

Table 1. Results from the generalised linear mixed effects models assessing the effect of bush clump area on the similarity of fungal communities. Pair-wise similarity values are calculated for all samples taken from each of the three different host species. Raup-Crick index is the probabilistic similarity of co-occurrences between rare and common ASVs, while the Morisita similarity gives weighting to more abundant ASVs.

<table>
<thead>
<tr>
<th>Host species</th>
<th>Similarity index</th>
<th>Fixed effects</th>
<th>Estimate</th>
<th>SE</th>
<th>z-value</th>
<th>p-value</th>
<th>Conditional R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euclea crispa</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>-1.3301</td>
<td>0.4968</td>
<td>-2.677</td>
<td>7.43X10^-3</td>
<td>0.2185</td>
</tr>
<tr>
<td>Euclea crispa</td>
<td>Morisita</td>
<td>Area</td>
<td>-0.16368</td>
<td>0.1528</td>
<td>-1.071</td>
<td>2.84X10^-1</td>
<td>0.00133</td>
</tr>
<tr>
<td>Canthium inerme</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>-0.09949</td>
<td>0.594</td>
<td>-0.167</td>
<td>8.67X10^-1</td>
<td>0.00013</td>
</tr>
<tr>
<td>Canthium inerme</td>
<td>Morisita</td>
<td>Area</td>
<td>-0.6755</td>
<td>0.3599</td>
<td>-1.877</td>
<td>6.05X10^-2</td>
<td>0.00032</td>
</tr>
<tr>
<td>Searsia chirindentis</td>
<td>Raup-Crick</td>
<td>Area</td>
<td>0.0064</td>
<td>0.0051</td>
<td>1.248</td>
<td>2.12X10^-1</td>
<td>0.00013</td>
</tr>
<tr>
<td>Searsia chirindentis</td>
<td>Morisita</td>
<td>Area</td>
<td>0.0001</td>
<td>0.00031</td>
<td>0.323</td>
<td>7.47X10^-1</td>
<td>5.13X10^-4</td>
</tr>
</tbody>
</table>

Table 2. Relative importance of the retained predictor variables for fungal endophyte compositional turnover for the full dataset. Estimates are determine by summing the coefficients of the I-splines from the GDM. Significant p-values of retained predictor variables are indicated in bold.
<table>
<thead>
<tr>
<th>Gradient</th>
<th>estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searsia chirindensis</td>
<td>0.364518</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Euclea crispa</td>
<td>0.325869</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Distance to forest edge</td>
<td>0.251447</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Δ Tree basal area</td>
<td>0.234689</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Geographic distance</td>
<td>0.174214</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Δ Tree compositional dissimilarity</td>
<td>0.154513</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inverse light intensity</td>
<td>0.154031</td>
<td>0.075</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Df</th>
<th>Chi-square</th>
<th>p-value</th>
<th>Marginal R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host</td>
<td>2</td>
<td>190.7435</td>
<td>&lt; 2.2X10⁻¹⁶</td>
<td>0.63055</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>1</td>
<td>1.5634</td>
<td>2.11X10⁻¹</td>
<td></td>
</tr>
<tr>
<td>Tree height</td>
<td>1</td>
<td>3.6442</td>
<td>5.63X10⁻²</td>
<td></td>
</tr>
<tr>
<td>Tree species richness</td>
<td>1</td>
<td>2.3464</td>
<td>1.26X10⁻¹</td>
<td></td>
</tr>
<tr>
<td>Inverse light intensity</td>
<td>1</td>
<td>4.0973</td>
<td>4.30X10⁻²</td>
<td></td>
</tr>
<tr>
<td>Host X Maximum temperature</td>
<td>2</td>
<td>12.6229</td>
<td>1.82X10⁻³</td>
<td></td>
</tr>
<tr>
<td>Host X Tree height</td>
<td>2</td>
<td>6.1574</td>
<td>4.60X10⁻²</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance</th>
<th>Standard deviation</th>
<th>AIC</th>
<th>Conditional R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1</td>
<td>Bush clump ID)</td>
<td>0.03451</td>
<td>0.1867</td>
<td>2078.7</td>
</tr>
<tr>
<td>(1</td>
<td>Sample ID)</td>
<td>0.13616</td>
<td>0.369</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Type II Wald chi-square test of the best subset GLMM assessing which factors affect ASV foliar fungal endophyte richness across a tree successional gradient. The retained variables from the best subset model, and their interactions between each other are fixed effects. Bold p-values represent significant effects at p > 0.05. Bush clump identity and Sample identity were included as random variables.

References


R Core Team (2019) R: A language and environment for statistical computing.


**Figure legends**

**Figure 1.** Effect plot showing the relationship between Raup-Crick similarity values of foliar fungal endophyte communities on Euclea crispa and BC area as a proxy for age of the BCs. Light blue shading represents the 95% confidence interval for the fitted model. Only significant trends are shown.
**Figure 2.** a-f) Significant retained predictor variables for the GDM performed on the full dataset. Each panel (a-f) represents the fitted I-splines (partial regression fits) of each retained predictor variable associated with the compositional turnover in foliar fungal endophyte communities. The maximum height reached by each curve represents the total amount of compositional turnover explained by that variable, holding all other variables constant.

**Figure 3.** a-c) Significant retained predictor variables for the GDM performed on the endophyte communities within the host Euclea crispa. d & e) Significant retained predictor variables for the GDM performed on the endophyte communities within the host Canthium inerme. f) Significant retained predictor variable for the GDM performed on the endophyte communities with the host Searsia chirindensis. Each panel (a-f) represents the fitted I-splines (partial regression fits) of each retained predictor variable associated with the compositional turnover in foliar fungal endophyte communities. The maximum height reached by each curve represents the total amount of compositional turnover explained by that variable, holding all other variables constant.

**Figure 4.** Predictor effect plots for the best subset GLMM model for ASV richness of foliar fungi. a) ASV richness decreased with light intensity - the x-axis represents inverse light intensity measured as photosynthetic active radiation (PAR) outside a BC minus PAR inside the BC. b) Maximum temperatures experienced within the BCs had no effect on ASV richness in C. inerme and E. crispa, but decreased with maximum temperature in S. chirindensis. c) Height of the trees from which samples were taken had no effect on ASV richness in C. inerme and E. crispa, whereas ASV richness decreased with maximum temperature in S. chirindensis. Light blue shading represents the 95% confidence interval for the fitted model.