Manuscript Draft : Aim3 Leaf Traits

Bolívar Aponte Rolón A. Elizabeth Arnold
Sunshine A. Van Bael

2023-11-29

1 Abstract

2 Introduction

3 Methods

3.1 Field

Growth and host plant inoculation seven tropical tree species was conducted at the greenhouses in the Gamboa Research Station, Smithsonian Tropical Research institute, Republic of Panama. The species, *Theobroma cacao*, *Dypterix* sp., *Lacmellea panamensis*, *Apeiba membranacea*, *Heisteria concinna*, *Chrysophyllum caimito*, and *Cordia alliodora* were chosen due to their variance in leaf traits (J.Wright unpublished data) and the availability of seeds in January- April 2019. Seeds of tree species were collected from the forest floor and grown in the greenhouse. Seedlings were kept in a chamber made out of PVC and clear plastic to prevent inoculation from spore fall inside the greenhouse. NEEDS INFORMATION ON THE SOIL MIXTURE AND AUTOCLAVING PROTOCOL. Seedlings reached a minimum of 4 true leaves before endophyte inoculation. Then 10 individual plants of each species were exposed to 10 nights of spore fall to achieve a high endophyte load (E+) and 10 homologous plants were kept inside the greenhouse plastic chamber to maintain a low endophyte load (E-) (Fig. ? MAKE A DIRAGRAM?). Plants exposed to spore fall were placed near (~10 m) the forest edge at dusk (~18:OO hours) and returned to the greenhouse at dawn (~07:00 hours) (Bittleston et al. 2011).

3.1.1 Leaf trait measurements

Three mature leaves were haphazardly collected from each of the individual plants in each treatment (E+, E-) within 7-10 days after inoculation treatment. Anthocyanin content and leaf thickness were measured while the leaf was still attached to the plant. We measured anthocyanin content with ACM-200plus (Opti-Sciences Inc. Hudson, New Hampshire, U.S.A.) on three haphazardly selected locations (working from the petiole out to the leaf tip) on the leaf surface of three haphazardly selected leaves for a total of nine measurements per plant (Tellez et al., 2022). The ACM-200 calculates an anthocyanin content index (ACI) value from the ratio of % transmittance at 931 nm/% transmittance at 525 nm (**opti-sciencesinc?**) . On compound leaves (i.e., *Dypterix* sp.) we measured at three different leaflets. Leaf thickness (m) was measured with a Mitutoyo 7327 Micrometer Gauge (Mitutoyo, Takatsu-ku, Kawasaki, Japan) in sthe same manner as the anthocyanin measurements, taking care to avoid major and secondary veins. After anthocyanin and leaf thickness measurements were completed, we removed the leaves from their stems, placed them inside a plastic bag (i.e. Ziploc), place in an ice chest and moved them to the lab for further measurements. Leaf punch strength was measured with an Imada DST-11a digital force gauge (Imada Inc., Northbrook, IL, United States) by conducting punch-and-die tests with a sharp-edged cylindrical steel punch (2.0 mm diameter) and a steel die with a sharp-edged aperture of small clearance (0.05 mm). The leaf punch measurements were taken by puncturing the leaf lamina at the base, mid-leaf and tip on both sides of the mid-vein, avoiding minor leaf veins when possible (Tellez et al., 2022). Once leaf toughness was measured, we used a 7 mm diameter punch hole to puncture disks for leaf mass per area (LMA) measurements. We collected one three disks per leaf (see Supplementary material for details). The disk punches dried at 60 °C for 48-72 hours. before being weighed.

3.1.2 Leaf tissue preparation for molecular work

The selected leaves were also used to profile endophyte community composition, abundance, and richness via amplicon sequencing (Illumina MiSeq). The leaf tissue remaining after the leaf trait measurements had the main vein and margins excised so that only the lamina remained. The lamina was haphazardly cut into 2 x 2 mm segments, enough to obtain a total of 16, and surface sterilized by sequential rinsing in 95% ethanol (10 s), 0.5 NaOCl (2 mins) and 70% ethanol (2 mins), as per (Arnold et al., 2003; Higgins et al., 2014; Tellez et al., 2022). After, leaves were air-dried briefly under sterile conditions. Sixteen leaf segments per leaf, a total of forty-eight leaf segments per plant, were plated in 2% malt extract agar (MEA), sealed with

Parafilm M (Bemis Company Inc., U.S.A.) and incubated at room temperature. The cultured leaf segments were used to estimate endophyte colonization of E+ and E- leaves. The presence or absence of endophytic fungi in the leaf cuttings was assessed 7 days after plating. The remaining sterilized leaf lamina was preserved in sterile 15 mL tubes with ~ 10 mL CTAB solution (1 M Tris–HCl pH 8, 5 M NaCl, 0.5 M EDTA, and 20 g CTAB). Leaf tissue in CTAB solution was used for amplicon sequencing (described in detail below). All leaf tissue handling was performed in a biosafety cabinet with all surfaces sterilized by exposure to UV light for 30 minutes and cleaned sequentially in between samples with 95% ethanol, 0.5% NaOCl and 70% ethanol to prevent cross contamination.

3.2 Amplicon sequencing

Leaf tissue in CTBA solution was stored for 2 months at room temperature prior to being placed at -80 C for 3 months before extracting DNA. In preparation for DNA extraction, we decontaminated all instruments, materials, and surfaces with DNAway (Molecular BioProducts Inc., San Diego, CA, United States), 95% Ethanol, 0.5 % NaOCl, and 70 % Ethanol, and subsequently treated with UV light for 30 minutes in biosafety cabinet. We then transferred 0.2 – 0.3 g of leaf tissue into duplicate sterile 2mL tubes, resulting in 2 subsamples. Total genomic DNA from subsamples was extracted as described in U'Ren & Arnold (2017). In brief, added two sterile 3.2 mm stainless steel beads to each tube and proceeded to lyophilize samples for 72 hours to fully remove CTAB content from tissue. After this period, we submerged the sample tubes in liquid nitrogen for 30s and proceeded to homogenize samples to a fine powder for 45 s in FastPrep-24 Tissue and Cell Homogenizer (MP Biomedicals, Solon, OH, USA). Afterwards, we repeated the decontamination procedure described before and used QIAGEN DNeasy 96 PowerPlant Pro-HTP Kit (U'Ren & Arnold, 2017) (QIAGEN, Valencia, CA, USA). After all genomic DNA was extracted, we pooled the subsamples for each individual sample before amplification. We used sterile equipment and pipettes with aerosol-resistant tips with filters in all steps before amplification. We followed a two-step amplification approach previously described by Sarmiento et al. (2017) and U´Ren & Arnold (2017). We used primers for the fungal ITSrDNA region, ITS1f (5'-CTTGGTCATTTAGAGGAAGTAA-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') with modified universal consensus sequences CS1 and CS2 and 0–5 bp for phase-shifting. Every sample was amplified in two parallel reactions containing 1-2 µL of DNA template (U'Ren & Arnold, 2017; see also Tellez et al., 2022). We visualized PCR (PCR1) reactions with SYBR Green 1 (Invitrogen, Carlsbad, CA, USA) on 2% agarose gel (Oita et al., 2021). Based on the electrophoresis band intensity, we combined parallel PCR1 reactions and diluted 5 µL of amplicon product with molecular grade water to standardize to a concentration of 1:15 (Sarmiento et al., 2017 for details; Tellez et al., 2022). We included DNA extraction blanks and PCR1 negatives in this step. We used a separate set of sterile pipettes, tips, and equipment to reduce contamination. We used a designated PCR area to restrict contact with pre-PCR materials (Oita et al., 2021).

We used 1 µL of PCR1 product from samples and negative control for a second PCR (PCR2) with barcode adapters (IBEST Genomics Resource Core, Moscow, ID, USA). Each PCR2 reaction (total 15 µL) contained 1X Phusion Flash High Fidelity PCR Master Mix, 0.075 µM of barcoded primers (forward and reverse pooled at a concentration of 2 µM) and 0.24mg/mL of BSA following (Sarmiento et al., 2017; U'Ren & Arnold, 2017). Before final pooling for sequencing, we purified the amplicons using Agencourt AMPure XP Beads (Beckman Coulter Inc, Brea, CA USA) to a ratio of 1:1 following the manufacturer's instructions. The products were evaluated with Bio Analyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) ((Tellez et al 2022) Tellez et al., 2022). We quantified the samples through University of Arizona Genetics Core, and subsequently diluted them to the same concentration to prevent over representation of samples with higher concentration, see (CITATION). Amplicons were normalized to 1 ng/µL, then pooled 2 µL of each for sequencing. No contamination was detected visually or by fluorometric analysis. To provide robust controls we combined 5 µL of each PCR1 negative and the DNA extraction blanks and sequenced them as samples. Ultimately, we combined samples into a single tube with 20 ng/µL of amplified DNA with barcoded adapters for sequencing on the Illumina MiSeq platform with Reagent Kit v3 (2×300 bp) following protocols from the IBEST Genomics Resource Core at the University of Idaho, USA. Again, we included the DNA extraction blanks and two PCR1 negatives and sequenced with samples. Sequencing yielded 3,778,081 total ITS1 reads.

3.2.1 Mock Communities

We processed and sequenced 12 mock communities following the methods described above. This allowed us to assess the quality of our NGS data set. We used two mock communities that consisted of PCR product from DNA extractions of 32 phylogenetically distinct fungi, representing lineages that are typically observed as endophytes: Ascomycota, Basidiomycota, Zygomycota and Chytridiomycota (Oita et al., 2021; see Daru et al., 2019 for details). In brief, we used six mock community with equimolar concentrations of DNA from all 32 fungal taxa and another six mock community with tiered concentrations of DNA from the same fungal taxa (Daru et al., 2019). Each mock community was sequenced five times (i.e., five replicates) (Oita et al., 2021). The read abundance from the equimolar and tiered communities was positively associated with the expected read number (with replicates as a random factor: R2Adj = 0.87, P = XXXX, see Supplementary material). Allowing us to evaluate the sequencing effectiveness in communities with known composition and structure (Bowman & Arnold, 2021). Henceforth, we used read abundance as a relevant proxy for biological OTU abundance (U'Ren et al., 2019).

3.2.2 Bioinformatic analyses

We used VSEARCH (v2.14.1) for *de novo* chimera detection, dereplication and sequence alignment. VSEARCH is an open-source alternative to USEARCH that uses an optimal global aligner (full dynamic programming Needleman-Wunsch), resulting in more accurate alignments and sensitivity (Rognes et al., 2016). For mock communities and experimental samples, we used forward reads (ITS1) for downstream bioinformatics analyses due to their high quality, rather than reverse reads (ITS4). Following Sarmiento et al. (2017), we concatenated all reads in a single file and used FastQC reports to assess Phred scores above 30 and determine the adequate length of truncation. We processed 892,713 of sequence reads from mock communities and 3,778,081 from experimental samples. We truncated mock community and experimental sample reads to a length of 250 bp with command `fast_truncIen` and filtered them at a maximum expected error of 1.0 with command `fast_maxee`. We then clustered unique sequence zero radius OTUs (that is, zOTUs; analogous to amplicon sequence variants (Callahan et al., 2016)), by using commands `derep_fulllength` and `minseqlength` set at 2. Sequentially we denoised and removed chimeras from read sequences with commands `cluster_unoise`, and `uchime3_denovo`, respectively (see Supplementary YYY for details). Finally, we clustered zOTUs at a 95% sequence similarity with command `usearch_global` and option `id` set at 0.95. After which, 3,035,960 sequence reads from experimental samples remained.

Taxonomy was assigned with the Tree-Based Alignment Selector Toolkit [v2.2; Carbone et al. (2019)] by placing unknowns within the Pezizomycotina v2 reference tree (Carbone et al., 2017). ITS sequences were blasted against the UNITE database by the ribosomal database project (RDP) classifier. A total 2147 OTUs hits were obtained and are composed of 68.6% Ascomycota, 26.8% Basidiomycota,<0.05% Chytridiomycota, <0.05% Glomeromycota, <0.05% Mortierellomycota, <0.05% Rozellomycota, 0.05% Kickxellomycota, and 4.2 % BLAST hit misses. Only OTUs representing Ascomycota were used for downstream statistical analyses since foliar endophyte communities in tropical trees are dominated by Ascomycota (Arnold & Lutzoni, 2007).

For each OTU identified, we removed laboratory contaminants from experimental samples by substracting the average read count found in control samples from the DNA extraction and PCR steps. Our analysis of mock communities allowed use to identify and remove false OTUs from experimental samples, those with fewer than 10 reads, and remove 0.1% of the read relative abundance across all samples (Oita et al., 2021). Removed reads represent the frequency of reads classified as contamination in the mock communities relative to the expected read count. Three experimental samples from *Theobroma cacao* (n=2) and *Apeiba membranacea* (n=1) were removed from all analyses due to incomplete entires. After pruning taxa with zero reads from experimental samples, we identified OTUs found exclusively (n= 260) in control ($E-$) plants (n=78) and deemed them as artifacts resulting from the greenhouse conditions. Consequently, these were consistently eliminated from treatment ($E+$) plants across all species. We converted reads for each fungal OTU to proportions of total sequence abundance per sample to reduce differences in sampling effort, following previous studies (Weiss et al. (2017); McMurdie & Holmes (2014)). We then removed singletons and obtained an average of 2,464,558 sequence reads in 529 Ascomycota OTUs accross 156 experimental samples of 7 tree species. All analyses post taxonomic assignment were performed in R [v. 4.3.2; R Core Team (2023)] using the `phyloseq` package (McMurdie & Holmes, 2013) and custom scripts (see Supplementary material).

3.2.3 Ant-endophyte interaction assays

A fresh fourth leaf was used in ant assays. To assess leaf-cutter ant damage, we introduced one detached leaf per plant per treatment to an actively foraging leaf-cutter ant colony for a two-hour assay. We presented leaf-cutter ant colonies with a choice of an E+ or an E- leaf on a disposable plastic plate next to an active nest trail. Carefully, we collected and placed debris from the trail leading up to the plate to lure ants into the plate. We initiated the ant assay as soon as an ant entered the plate and explored the leaf contents (for ~ 10-20 seconds). Every five minutes we took a digital photo of the choice arena until about 75% of the leaf content of one of the leaves was consumed. We used the digital photo at time zero and at the end of trial to quantify the leaf area removed using ImageJ [v1.52r; Schneider et al. (2012)]. Ant recruitment was estimated by counting individuals in the choice arena throughout trial event.

3.2.4 Pathogen assays

For the pathogen assays, we introduced an agar plug inoculated with hyphae of *Calonectria* sp. ($P+$ treatment), and an agar plug without the pathogen ($P-$ control) to similarly aged/sized leaves within 10-14 days after endophyte inoculations. Leaves with the $P+$ or $P-$ treatment were misted with sterile water two times a day (morning and afternoon) to maintain moisture. After four days, we removed the plugs and took digital photos to analyze leaf area damage using ImageJ [v1.52r; Schneider et al. (2012)].

3.2.5 Statistical Analyses

we used a multivariate analysis of variance (MANOVA) for the all the tree species using the "manova" function in the "stats" package in R (The R Core Team, 2013). A MANOVA allows for an analysis of variance with two or more covariates (i.e., endophyte load (E_load) and tree species)) and multiple dependent variables (i.e., ACI, LMA, Thickness, Toughness). With the "summary.aov" function and argument "split" set to "list ("E_load:Species")" I determined how independent variables, E_load and tree species, influenced variance of my covariates: Thickness, Toughness, LMA, and ACI. I computed Wilk's tests statistic for MANOVA where in the closer to zero the statistic is, the variable in question contributes more to the model, hence we reject the null hypothesis if zero. Additionally, to determine which interactions of E_load and tree species are significant in regard to abundance, I performed two way ANOVAs using "aov" function and post-hoc Tukey tests using "TukeyHSD" function in the "stats' package in R (The R Core Team, 2013). This allows for a more in-depth look at interactions between and among groups, not necessarily apparent from MANOVA tests.

To test for H2, I used a general linear mixed model (GLM). First, to determine which fixed effects to include in my model I created a correlation matrix with "cor()" function in R statistical software. Covariates (Thickness, Toughness, LMA, ACI, E_load, and Tree Species) with correlation coefficients greater than 0.25 were not included in linear model analyses. Secondly, Principal Component Analysis (PCA) was used to determine underlying interactions among covariates that could influence fungal endophyte abundance in seedlings (hereafter, abundance). The PCA was computed using the "prcomp" function in R statistical software (The R Core Team, 2013). The PCA was computed using the following covariates: Thickness, Toughness, LMA, and ACI, to test the effect of ACI, Species, Toughness, Thickness, LMA, endophyte load (E_load) as my independent variables (fixed effects) on the abundance of endophytes, dependent variable. The PCA was used to inform which covariates to include in linear model analyses. From PCA and correlation analysis, Thickness, Toughness and E_load were chosen to compute an initial linear model with interaction. Linear model was computed using "lm" function from the "stats" package in R (The R Core Team, 2013). I selected the best fit model using "stepAIC" function in the "MASS" package with "direction" argument set to "both" for backward and forward selection (The R Core Team, 2013). The best fit model was achieved when all covariates had a significant P-value. Results Results support my first hypothesis: H1) endophyte abundance will be significantly different across tree species and endophyte treatments. The results from the MANOVA for E_load (p-value < 0.0001), Species (p-value < 0.0001) and the interaction between E_load and Species (P-value < 0.0001) support this hypothesis, with a Wilk's test statistic of 0.44, 0.02, and 0.75 respectively, hence the null hypothesis is rejected (Fig. 2 and 3). Multiple interactions between endophyte abundance and endophyte treatment and tree species were significantly different. A two-way ANOVA and post-hoc Tukey test of these interactions show in detail which interaction between and among groups are significant (see Table 4 in supplementary materials).

Principal Component Analysis revealed how covariates (LMA, ACI, Thickness and Toughness) interact. I overlapped tree species groups on the PCA axes to show how the variance in the data is explained by PC1 (60%) and PC2 (27%) (Fig. 4). This is indicative of correlation among covariates. Thickness and toughness were orthogonal to each other in PCA, indicative of low correlation. Correlation analysis revealed that these covariates were negatively correlated to each other ($r = -0.12$). Additionally, both Toughness and Thickess had the greatest loadings on PC1 and PC2, respectively (Fig. 4).

The best fit model that resulted from using "stepAIC" function includes Toughness and E_load as fixed effects as well as their interaction effect (F-statistic=180.6, df = 463, p-value: $< 2.2e-16$). The AIC value for the best fit model is -108.1. I did not include in further analyses covariates that were above the 0.25 correlation cut-off established (see Table 1 in supplementary materials). Finally, two equations resulted from the multivariate regression analysis of the best fit model: for E- plants and for E+ plants (Fig. 5). Plants exposed E+ treatment have a negative correlation with leaf toughness. That is, as Toughness increased endophyte abundance decreased. Plants exposed to E- treatment had a negative correlation as well, albeit the slope of the line was not as steep (Fig. 5). No random effects were modeled.

4 References

Arnold, A. E., & Lutzoni, F. (2007). Diversity and host range of foliar fungal endophytes: Are tropical leaves biodiversity hotspots? *Ecology*, *88*(3), 541–549. https://doi.org/10.1890/05-1459

Arnold, A. E., Mejía, L. C., Kyllo, D., Rojas, E. I., Maynard, Z., Robbins, N., & Herre, E. A. (2003). Fungal endophytes limit pathogen damage in a tropical tree. *Proceedings of the National Academy of Sciences*, *100*(26), 15649–15654.

Bowman, E. A., & Arnold, A. E. (2021). Drivers and implications of distance decay differ for ectomycorrhizal and foliar endophytic fungi across an anciently fragmented landscape. *The ISME Journal*, *15*(12), 3437–3454. https://doi.org/10.1038/s41396-021-01006-9

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature*

Methods, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Kauff, F., U'Ren, J. M., May, G., & Lutzoni, F. (2017). T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: An example with the Pezizomycotina tree of life. *Bioinformatics*, *33*(8), 1160–1168. https://doi.org/10.1093/bioinformatics/btw808

Carbone, I., White, J. B., Miadlikowska, J., Arnold, A. E., Miller, M. A., Magain, N., U'Ren, J. M., & Lutzoni, F. (2019). T-BAS Version 2.1: Tree-Based Alignment Selector Toolkit for Evolutionary Placement of DNA Sequences and Viewing Alignments and Specimen Metadata on Curated and Custom Trees. *Microbiology Resource Announcements*, *8*(29), e00328–19. https://doi.org/10.1128/MRA.00328-19

Daru, B. H., Bowman, E. A., Pfister, D. H., & Arnold, A. E. (2019). A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1763), 20170395. https://doi.org/10.1098/rstb.2017.0395

Higgins, K. L., Arnold, A. E., Coley, P. D., & Kursar, T. A. (2014). Communities of fungal endophytes in tropical forest grasses: Highly diverse host- and habitat generalists characterized by strong spatial structure. *Fungal Ecology*, *8*(1), 1–11. https://doi.org/10.1016/j.funeco.2013.12.005

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Oita, S., Ibáñez, A., Lutzoni, F., Miadlikowska, J., Geml, J., Lewis, L. A., Hom, E. F. Y., Carbone, I., U'Ren, J. M., & Arnold, A. E. (2021). Climate and seasonality drive the richness and composition of tropical fungal endophytes at a landscape scale. *Communications Biology*, *4*(1), 313. https://doi.org/10.1038/s42003-021-01826-7

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Sarmiento, C., Zalamea, P. C., Dalling, J. W., Davis, A. S., Simon, S. M., U'Ren, J. M., & Arnold, A. E. (2017). Soilborne fungi have host affinity and host-specific effects on seed germination and survival in a lowland tropical forest. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(43), 11458–11463. https://doi.org/10.1073/pnas.1706324114

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Tellez, P. H., Arnold, A. E., Leo, A. B., Kitajima, K., & Van Bael, S. A. (2022). Traits along the leaf economics spectrum are associated with communities of foliar endophytic symbionts. *Frontiers in Microbiology*, *13*, 927780. https://doi.org/10.3389/fmicb.2022.927780

Tellez, P. H., Rojas, E., & Van Bael, S. (2016). Red coloration in young tropical leaves associated with reduced fungal pathogen damage. *Biotropica*, *48*(2), 150–153. https://doi.org/10.1111/btp.12303

U'Ren, J. M., & Arnold, A. E. (2017). 96 well DNA Extraction Protocol for Plant and Lichen Tissue Stored in CTAB. *Protocols.io*, 1–5.

U'Ren, J. M., Lutzoni, F., Miadlikowska, J., Zimmerman, N. B., Carbone, I., May, G., & Arnold, A. E. (2019). Host availability drives distributions of fungal endophytes in the imperilled boreal realm. *Nature Ecology & Evolution*, *3*(10), 1430–1437. https://doi.org/10.1038/s41559-019-0975-2

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 27. https://doi.org/10.1186/s40168-017-0237-y