universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Exploring Genomic Relationships in *Tillandsia* subgenus *Tillandsia*"

verfasst von / submitted by

Norma Lisette Rivera BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna 2019 Masterstudium Botanik UG2002

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 832
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Christian Lexer
Mitbetreut von / Co-Supervisor:	

Contents

Acknowledgements

Thanks to Erik Szamosvari and Lukas Grossfurthner fort their support, and to Gil Yardeni, Michael Barfuss, and Christian Lexer for their endless patience and guidance.

Abstract

The genus *Tillandsia* of the Bromeliaceae family is an ideal study system for addressing the processes facilitating or accompanying adaptive radiation. Our objectives for this study were to test phylogenetic hypotheses from previous studies based on fewer markers, provide more resolution for phylogenomic research on *Tillandsia* subgenus *Tillandsia*, and investigate the differences between the subgenus' nuclear and plastid phylogenetic trees. Plant materials sampled from North, Central, and South America were subjected to DNA extraction, library preparation, whole-genome re-sequencing, and bioinformatic analyses. The composition of clades recovered by our phylogenomic study was very similar to that observed in previous studies with much fewer markers, both for nuclear and plastid genomes. Complex patterns of genome-wide variation in tree topologies seen could stem from incomplete lineage sorting, and the differences between the nuclear and plastid genomes could be due to the plastid's uniparental inheritance and its considerably smaller size and greater degree of conservation. The known timing of diversification and observed patterns of transitions between C3 and CAM photosynthesis among the studied clades are suggestive of Pleistocene glaciations and refugia being among the drivers for the adaptive radiation of *Tillandsia* subgenus *Tillandsia*.

Keywords: *Tillandsia*, adaptive radiation, CAM photosynthesis, whole genome sequencing, plastid genome, incomplete lineage sorting

Die Gattung *Tillandsia* aus der Familie der Bromeliaceae ist ein ideales Studiensystem, um zu untersuchen, welche Prozesse adaptive Radiation erleichtern oder begleiten. Unser Ziel war es, phylogenetische Hypothesen aus früheren Studien, die auf weniger molekularen Markern basieren, zu testen, eine höhere Auflösung für die Phylogenie an *Tillandsia* Untergattung *Tillandsia* bereitzustellen und die Unterschiede zwischen kern- und plastidenbasierten Phylogenien des Subgenus zu untersuchen. Zu diesem Zweck wurde DNA aus Pflanzenmaterial aus Nord-, Mittel- und Südamerika extrahiert, NGS-libraries vorbereitet, und bioinformatisch, basierend auf einer Gesamtgenom-sequenzierung, analysiert. Die Zusammensetzung der durch unsere phylogenomische Studie gewonnenen Kladen war derjenigen sehr ähnlich, die in früheren Studien mit viel weniger Markern sowohl für das Kern- als auch für das Plastidengenom beobachtet wurde. Die resultierenden Muster in genomweiter Variation könnten auf unvollständige Sortierung der Abstammungslinien zurückzuführen sein, und die Unterschiede zwischen dem Kern- und dem Plastidengenom könnten auf die uniparentale Vererbung des Plastidengenoms sowie auf dessen erheblich geringere Größe und größere Konservierungsrate zurückzuführen sein. Das bekannte Timing der Diversifikation und die beobachteten Muster der Übergänge zwischen C3- und CAM-Photosynthese in die studierten Kladen lassen darauf schließen, dass Pleistozäne Vergletscherungen und Refugien zu den Treibern für die adaptive Strahlung der *Tillandsia*-Untergattung *Tillandsia* gehören.

Schlagwörter: *Tillandsia*, adaptive Radiation, CAM Photosynthese, Gesamtgenom-Sequenzierung, Plastidengenom, Incomplete Lineage Sorting

Introduction

Bromeliaceae is a diverse plant family with its main distribution in the American tropics (Figure 1). The genus *Tillandsia* is comprised of ca. 650 species, which are mostly small, perennial, evergreen epiphytes (Benzing, 2000; Mendoza *et al.*, 2017). According to Benzing *et al*., (2000) *Tillandsia* possesses characters that facilitated success in often demanding habitats – success that led to rapid diversification. A synapomorphy for Bromeliaceae are its absorptive trichomes that serve a vital role in both water conservation and acquisition – a functionally important adaptive trait, sometimes referred to as a key innovation (Tomlinson, 1969; Benzing, 2000; Crayn *et al*., 2015).

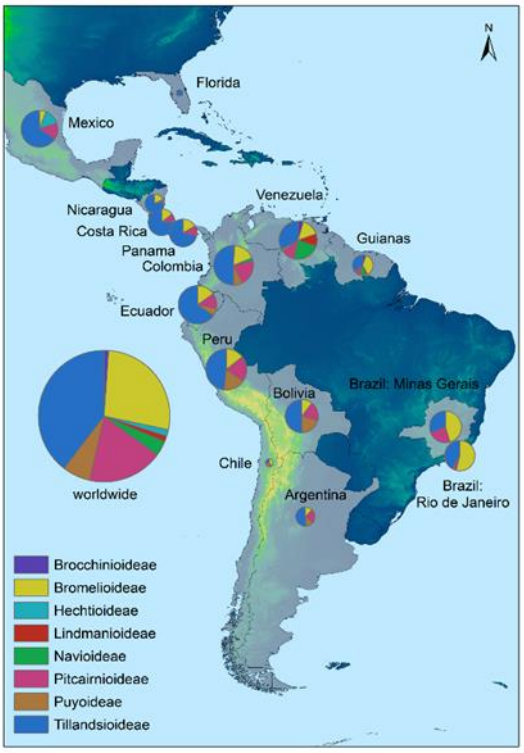


Figure 1 . Diversity of Bromeliad subfamilies (Cáceres González et al., 2011b).

Scarcity and the necessity of economical water usage led to another key innovation: crassulacean acid metabolism (CAM) as a photosynthetic pathway, independently evolved thrice within Bromeliaceae (Crayn, Winter and Smith, 2004). It enables elevated water use efficiency by restricting gas exchange during the daytime. The plants' stomata open nocturnally, when less water is lost via transpiration, and CO_2 taken up is processed by PEP carboxylase and temporarily stored as malic acid in the vacuole. Diurnally, the stomata close again, and the malic acid is then decarboxylated and enters the plastid and the Calvin Cycle (Winter & Smith, 1996). CAM photosynthesis is thought to have multiple origins within the Bromeliaceae family; within Tillandsioidae, 28% are CAM plants – all of them belonging to *Tillandsia* (Crayn, Winter and Smith, 2004; Crayn *et al.*, 2015), though not all members of *Tillandsia* are CAM plants. According to Givnish *et al.* (2014), the second distinct origin of CAM photosynthesis occurred in the Middle Miocene sub-epoch (between 15.97-11.6 Ma) in subfamily Tillandsioideae, but fuzzy resolution in phylogenetic trees call into question the estimated date of *Tillandsia*'s evolution. Barfuss *et al.,* (2016) and Mendoza *et al.,* (2017) shed light on the genus *Tillandsia*'s phylogenetic relationships through the use of plastid and nuclear markers, but the species' recent adaptive radiation calls for a more comprehensive study. What has also been found within the *Tillandsia* subgenus is that not only are there both C3 and CAM plants, but facultative plants as well (Mooney *et al.*, 1989; Crayn *et al.*, 2015) that have the capability to switch between the C3 and CAM photosynthetic pathways, depending on environmental cues (Winter and Holtum, 2014). Such recently-evolved innovations make this genus an interesting study system for adaptive radiation, defined by Seehausen (2004) as "the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage."

In addition, the plastid genome is an ideal study system for adaptive radiation because it is smaller, inherited maternally, and is more conserved than the nuclear genome. Given its uniparental inheritance and non-recombinant nature, plastid genomes offer an insightful view into adaptive radiation and species hybridization (Liu *et al.*, 2018),when compared to the nuclear genome, as these two genomic compartments are expected to be differentially affected by lineage sorting processes and gene flow.

The introduction of new innovative technology has made phylogenomics a rapidly-growing new field in science. With the aid of Next Generation Sequencing (NGS) and bioinformatic software that

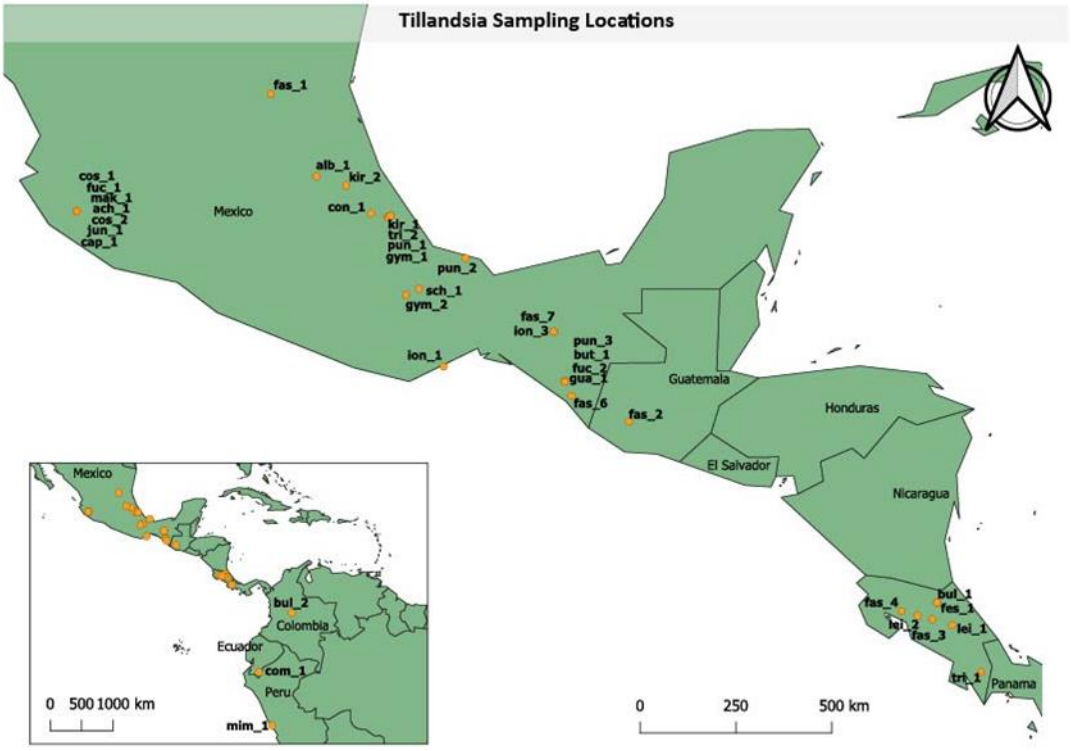
makes processing a plethora of calculations possible, we can uncover relationships between organisms that were before only linked by their morphological traits- species that may have diverged so recently that they are currently clumped into the same species morphologically. Utilizing NGS, we focus upon 19 species of *Tillandsia* subgenus *Tillandsia*, and seek to i) investigate the differences between phylogenies inferred using few markers as Barfuss *et al*. (2016), and Mendoza *et al*. (2017) and whole genome phylogenetic patterns, ii) explore how the phylogenetic trees differ when using plastid and nuclear markers, and lastly, iii) learn more about this group's phylogeny and the processes facilitating or accompanying its evolution.

Methods

Sampling

Plant material was taken from 23 *Tillandsia* species in Mexico, Guatemala, Costa Rica, Colombia, Guatemala, Peru, and Honduras (Figure 1 and Table 1). Plant material was cut lengthwise and dried in silica powder (Chase & Hills, 1991).

Figure 2. Map of Sampling locations.



It should be noted, that some samples stem from living specimens, and the original collection coordinates could not be retrieved, but only the general locality. In Figure 1, only the sites with coordinates are shown.

Code	Species	Collection ID	Elevation (m)	Latitude	Longitude
pun_1	*T. punctulata*	GY089	1417	19° 32’ 57.7” N	096° 55' 05.7” W
pun_2	*T. punctulata*	MH016_h	1105	18° 32’ 52.908” N	95° 9’ 3.528” W
but_1	*T. butzii*	LSM_132C	1800	15° 38' 07.3” N	92° 48' 14.3” W
fuc_1	*T. fuchsii var. stephanii*	LSM_368F	1400	19° 39' 20.17” N	104° 19' 45.86” W
pun_3	*T. cf juncea*	LSM_147C	1800	15° 38' 08.1” N	92° 48' 14” W
jun_1	*T. juncea*	LSM_366D	1500	19° 39' 05.2” N	104° 20' 06” W
kir_1	*T. kirchhoffiana*	GY157	1611	19° 31' 00.3” N	097° 00' 07.9” W
tri_1	*T. tricolor*	OL053_A	1134	8° 47’ 36.4992” N	82° 58’ 16.9968” W
ach_1	*T. achyrostachys*	LSM_370A	1400	19° 39' 20.17” N	104° 19' 45.86” W
sch_1	*T. schiedeana*	MH076_B1	300	17° 49' 27.5988” N	96° 14’ 37.5” W
con_1	*T. concolor*	GY133	2425	19° 35' 52.8” N	097° 22' 40.4” W
fes_1	*T. festucoides*	OL009_c	77	10° 26' 01.5” N	84° 00' 44.8” W
bul_1	*T. bulbosa*	OL010_f	77	10° 26' 1.5” N	84° 0’ 44.7998” W
gym_1	*T. gymnobotrya*	MH027_C	2040	19° 29’ 55.2984” N	96° 58’ 17.2992” W
gym_2	*T. gymnobotrya*	MH124_c	2110	17° 40’ 49.872” N	96° 33’ 12.204” W
gua_1	*T. guatemalensis*	LSM_15D	1200	15° 39' 32.4” N	92° 48' 28.5” W
mak_1	*T. makoyana*	LSM_374B	1500	19° 39' 05.2” N	104° 20' 06” W
fuc_2	*T. fuchsii*	LSM_170A	1600	15° 37' 49.7” N	92° 48' 12.1” W
fas_1	*T. fasciculata*	GY004	1250	22° 25' 12.3” N	099° 44' 58.5” W
tri_2	c.f. *T. tricilor*	GY084	1417	19° 32' 44.3” N	096° 55' 15.0” W
kir_2	*T. kirchhoffiana*	B1308	1163	20° 15' 31.3” N	97° 58' 06.9” W
bul_2	*T. bulbosa*	HBVSN	2640	4° 40' 06.6” N	74° 05' 58.7” W
fas_2	*T. fasciculata*	HBV 0026950	1562	14° 41' 43.1” N	91° 17' 25.3” W
lei_1	*T. leiboldiana*	HBV 0000663	600	9° 54’ 0” N	83° 39’ 30” W
cap_1	*T. caput-medusae*	LSM_369C	1400	19° 39' 20.17” N	104° 19' 45.86” W
cos_1	*T. cossonii*	LSM_381C	2000	19° 35' 11.9” N	104° 15' 57.9” W
cos_2	*T. cossonii*	LSM_365C	1500	19° 39' 05.2” N	104° 20' 06” W
fas_3	*T.fasciculata*	HBV 0024655	1200	10° 01' 59.6” N	84° 07' 25.0” W
fas_4	*T.fasciculata*	HBV 0024657	750	10° 14' 09.1” N	84° 50' 51.9” W
ion_1	*T.ionantha.*var.*maxima*	HBV 0024697	100	15° 59' 49.0” N	95° 39' 55.2” W
lei_2	*T.leiboldiana*	HBV 0024715	1030	10° 07' 27.5” N	84° 28' 25.3” W
fas_5	*T.fasciculata*	HBV 0025334			
fas_6	*T.fasciculata*	HBV 0025390	140	15° 18' 02.0” N	92° 38' 48.0” W
alb_1	*T.albida*	HBV 0000269	1342	20° 28' 21.1” N	98° 40' 07.3” W
ion_2	*T. ionantha*	HBV B219/91			
ion_3	*T.ionantha.*var.*van*	HBV B437/82	1200	16° 48' 59.4” N	93° 03' 59.1” W
com_1	*T. complanata*	IO041A	1200	4° 06’ 26.4” S	78° 58’ 05.5” W
fas_7	*T.fasciculata*	HBV 0024656	1200	16° 48' 59.4” N	93° 03' 59.1” W
adp_1	*T. adpressiflora*	BR3.2015 (8951)			
mim_1	*T. mima* var. *chiletensis*	HBV 0020417	1000	12° 02' 58.0” S	77° 01' 02.5” W

Table 1. Sample information. Whenever possible, coordinates were provided.

DNA Extraction

10-20 mg of silica-dried plant material were added to a 2ml tube along with 5 glass beads. The tubes were subsequently frozen in liquid nitrogen and ground in the TissueLyser II (Qiagen, cat No./ID: 85300) for 8-12 minutes in cycles of 4 minutes, until the plant material was a homogenous powder. DNA extractions were carried out with a modified CTAB protocol (Supplementary 1), based on Doyle & Doyle’s (1987) protocol, and with the DNeasy® Plant Mini Kit (Qiagen) as per the manufacturer’s instructions. Samples were subsequently cleaned with the NucleoSpin® gDNA Clean-up kit (Macherey-Nagel). If a sample was found to have less than 4 ng/µl of DNA when checked with the Qubit 3.0 Flurometer (Invitrogen), then the process was repeated with more plant material and respective extracts were pooled.

Library Preparation, PCR, and Size Selection

The extracted DNA was fragmented using the Bioruptor® Pico, using the settings "15 sec on," and "90 sec off." One cycle was run at a time, and the fragment lengths subsequently verified with gel electrophoresis. These steps were repeated until the desired length of 200 bp was reached. Dual-indexed libraries following Meyer & Kircher (2010) and Kircher *et al.* (2012) were then prepared according to the KAPA protocol (KAPA LTP Library Preparation Kit), using Agencourt AMPure magnetic beads. Illumina TruSeq adaptors were added, and PCR settings were: Denaturation 98° C 30 sec; 8 cycles of Denaturation 98° C 10 sec, Annealing 60° C 20 sec, Elongation 72° C 20 sec; final extension step of Elongation 72° C 10 min. A 1:1 ratio of beads: sample was added and incubated for 10 minutes before being washed twice with 80% ethanol on the magnetic stands, and then eluted in ddH_2O. Size selection was then performed according to the SPRIselect protocol (Beckman Coulter Life Sciences). Fragment size distribution was then viewed with the Agilent 2100 Bioanalyzer. Completed libraries were sent to the Vienna Biocenter Core Facility (VBCF) for paired-end (PE) Illumina sequencing.

Data Pre-Processing

The raw reads were demultiplexed with deML (Renaud *et al*., 2014), and split with SAMtools (Li *et al.,* 2009). The adapters were trimmed using Trimmomatic (Bolger *et al.,* 2014) with the settings "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:15". Once it was certain that the adapter was removed, Bowtie 2 (Langmead *et al*., 2012) was utilized to build a pseudo reference for the nuclear genome. An *Ananas comosus* nuclear genome (Ming *et al.*, 2015) and a *T. usneoides* plastid genome served as references for read mapping. Bowtie2 also aligned the trimmed fastq files under the setting "- -very-sensitive." SAMtools was used to sort the files by coordinate, and duplicates were then subsequently marked with Picard Tools (available from: https://broadinstitute.github.io/picard). Variants were called with Freebayes (Garrison & Marth, 2012).

For the nuclear genome- in VCFtools (Danecek *et al*., 2011), the following filters were utilized: a filter for genotypes called for at least 36 out of 38 individuals across all individuals, SNPs with a minor allele count of less than 3 were removed, and a minimum depth of 3 reads for a genotype call. With the program Vcflib (Garrison, 2016), a filter for keeping allele balance between25% and 75% was applied, as well as mapping quality above 90%, paired status of 50% for reads supporting reference or alternate alleles, and removing loci with quality scores in the lowest and highest quartiles of the sequencing depth. After indels were removed, we were left with 1,740,402 sites for analysis in the nuclear genome dataset.

For the plastid genome, the following filters were utilized in VCFtools: the maximum missing count was set to 45 (out of 47), with a minimum depth of 4, and then the indels were removed. This left us with 1816 sites in the plastid genome dataset.

The vcf files were subsequently converted to phylip and nexus text files using PGDSpider (Lischer & Excoffier, 2012).

Phylogenetic Analyses

The plastid genome file did not contain enough SNPs to perform this step, therefore the following refers only to the nuclear genome. The filtered VCF file was cut into overlapping windows of 100 kb, sliding by 20kb. The windows were then filtered so that each had at least 200 SNPs. This was run through ParGenes (Morel, Kozlov and Stamatakis, 2018) , which takes multiple sequence alignments and conducts multiple individual tree inferences in parallel, resulting in a better-fitting model. The

output was subsequently run through RAxML-NG (Kozlov *et al*., 2018) and was used to produce a maximum likelihood tree (Figure 5) with the model GTR-GAMMA with random assignment, the option to correct for ascertainment bias, and 1.6 million SNPS. ASTRAL (Mirarab *et al.*, 2014) , using a coalescent-based formula, then produced many trees and output the best supported phylogeny - the formula was shown to produce a smaller missing branch rate in comparison to RAxML-NG. In contrast to RAxML-NG, the node labels on the ASTRAL trees display consensus percentages instead of bootstrap support.

A comparison tree was produced without genomic windows in RAxML-NG with the same settings as above: "GTR-GAMMA" model, ascertainment bias correction, and random assignment of 1.6 million SNPs to create a maximum likelihood tree. Angsd (Korneliussen, Albrechtsen and Nielsen, 2014) used called genotypes to create a principal component analysis of our individuals, assuming minimum allele frequency of 0.07, filtering out alleles that were in less than 50% of individuals, and otherwise using default settings. Graphics were created using R (packages gplots and ggplot2). Variance was calculated in R by dividing the eigen values by the factors.

FineRADstructure (Malinsky *et al.*, 2018) was used with default settings to create a simple co-ancestry heatmap of the nuclear genome in R (packages ape and XML), allowing distinct groupings to be discerned. Phylogenetic Networks were created in Splitstree4 (Huson & Bryant, 2006) using NeighborNet distance transformation to view reticulations between individuals. Individual nexus files were created for each genome. A Densitree (Bouckaert, 2010) graphic was produced using 500 100kb windows as alignments to run through RAxML-NG with the same aforementioned settings. 1000 trees were produced and then superimposed onto each other with the Densitree software. The "root canal" option was used to create a backbone. Allele frequencies were calculated using Simon Martin's "freq.py" script (available from https://github.com/simonhmartin/genomics_general), Patterson's D-statistic was calculated in R using a script by Michael Matschiner called calculate_abba_baba.r (available from https://github.com/mmatschiner/tutorials), based on commands from Simon Martin's tutorial (available from http://evomics.org). *Tillandsia complanata* was used most often as the outgroup, although *T. fuchsii* was occasionally used as well.

Results

After filtering, 1,740,402 sites were left for analysis in the nuclear region dataset, and 1816 sites in the plastid region dataset. As an exploratory visualization, a principal component analysis (Figure 3) was carried out. 19.3% of the variance is explained by PC1, and 6.6% is explained by PC2. A pattern already emerges, as we can see species grouping together. . A similar grouping pattern can be seen in the fineRADstructure-generated coancestry heatmap (Figure 4) for the nuclear genome, in which four distinct clusters appear, which are numbered according to their order of appearance in the RAxML-NG tree (Figure 5). They were named according to their positions on the trees generated in previous studies. *Tillandsia* subg. *Tillandsia* was named 'Clade K' in Barfuss *et al.* (2005), and then separated into subclades K.1 and K.2 by Mendoza Granados *et al.,* (2017). For convenience, we will continue expanding upon this naming scheme, using Mendoza Granados' tree as a reference.

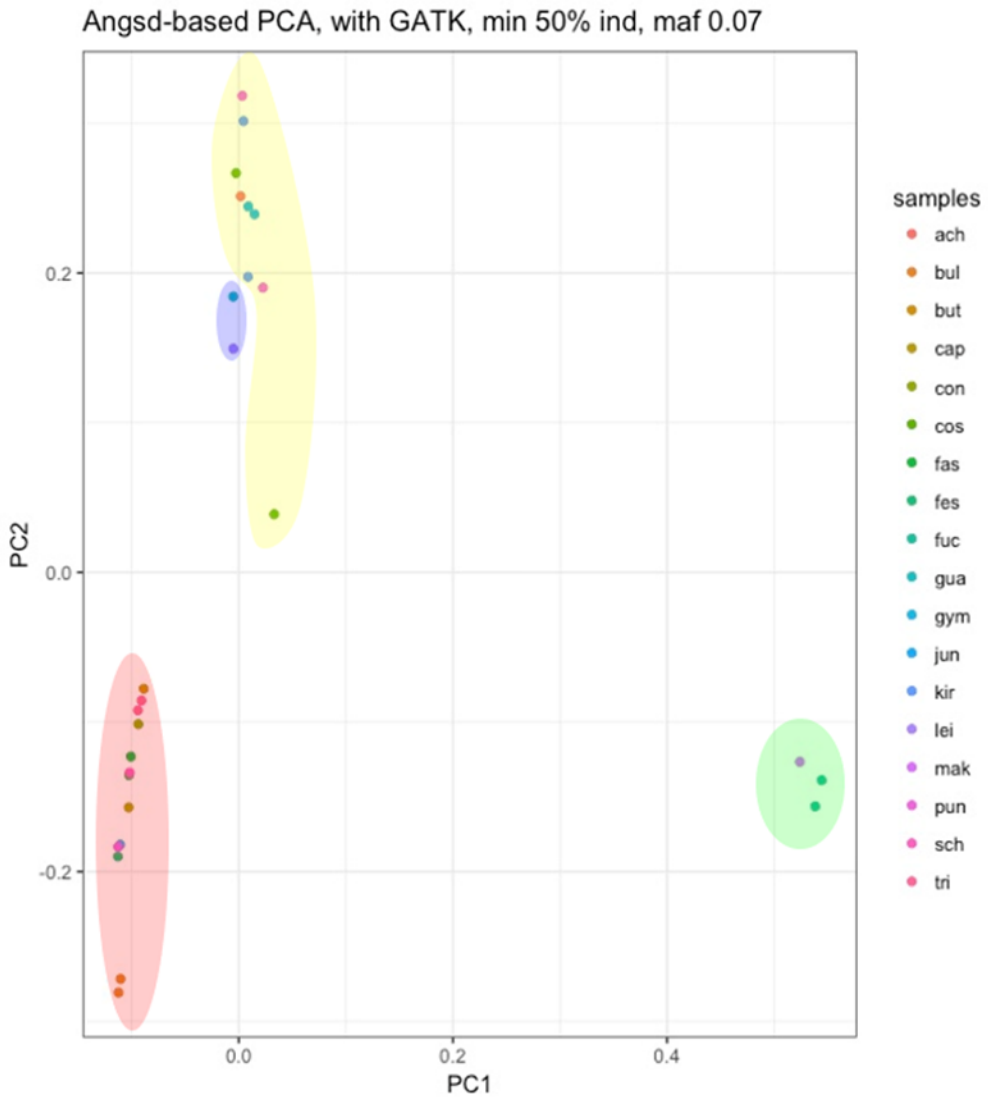


Figure 3. Principal component analysis, nuclear genome. Samples are named by the first three letters of their specific epithet. The pink group corresponds to K.2.3 , the green group corresponds to K.2.1 , yellow corresponds to K.1 , and blue corresponds to K.2.2, as per Figure 4.

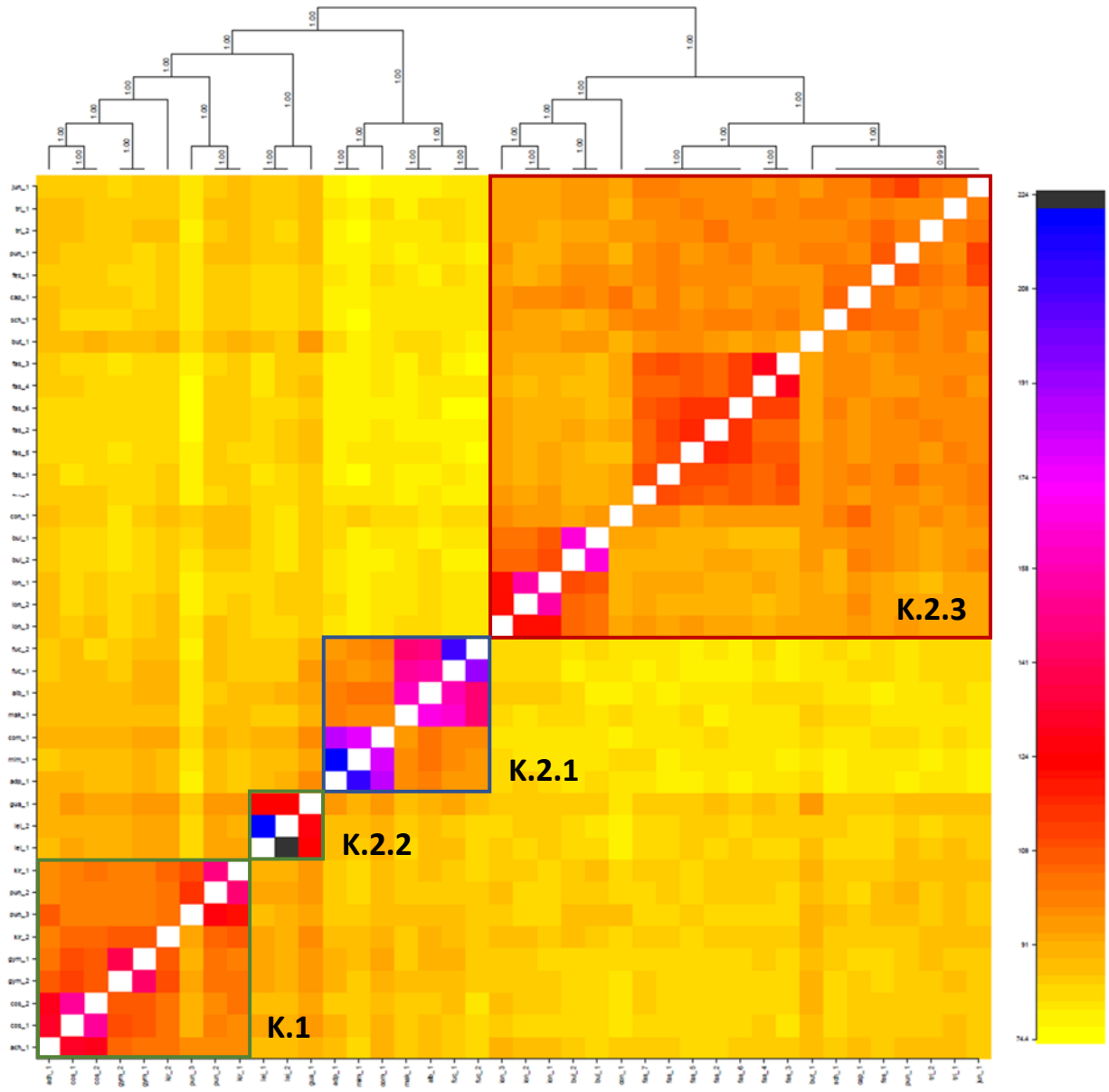


Figure 4. Simple Co-ancestry Heatmap, Nuclear Genome. Each of the axes depicts all of our samples, which are then mapped against each other for shared coancesty. A coalescent tree was formed based upon this information (top of figure)

Our clade K.1 generated with whole-genome data corresponds to clade K.1 in Mendoza *et al.,* (2017). Our remaining clusters correspond to sub-clades within Clade K.2. Our clade K.2.1 corresponds to Clade K.2, *Tillandsia* sect. *Tillandsia*. Our clade K.2.2 corresponds with Clade K.2, *Tillandsia* sect. *Allardtia*. Our clade K.2.3 corresponds to *Tillandsia* sect. *Eriophyllum*. The same clustering pattern is supported throughout our analyses. In the heatmap (Figure 4), the groupings are easily distinguishable, grouped by their shared coancestry. A subgroup can be seen nested within clade K.2.3, which shares more coancestry within the clustered samples than with the rest of clade K.2.3, but due to low bootstrap support, it was not considered its own distinct clade.

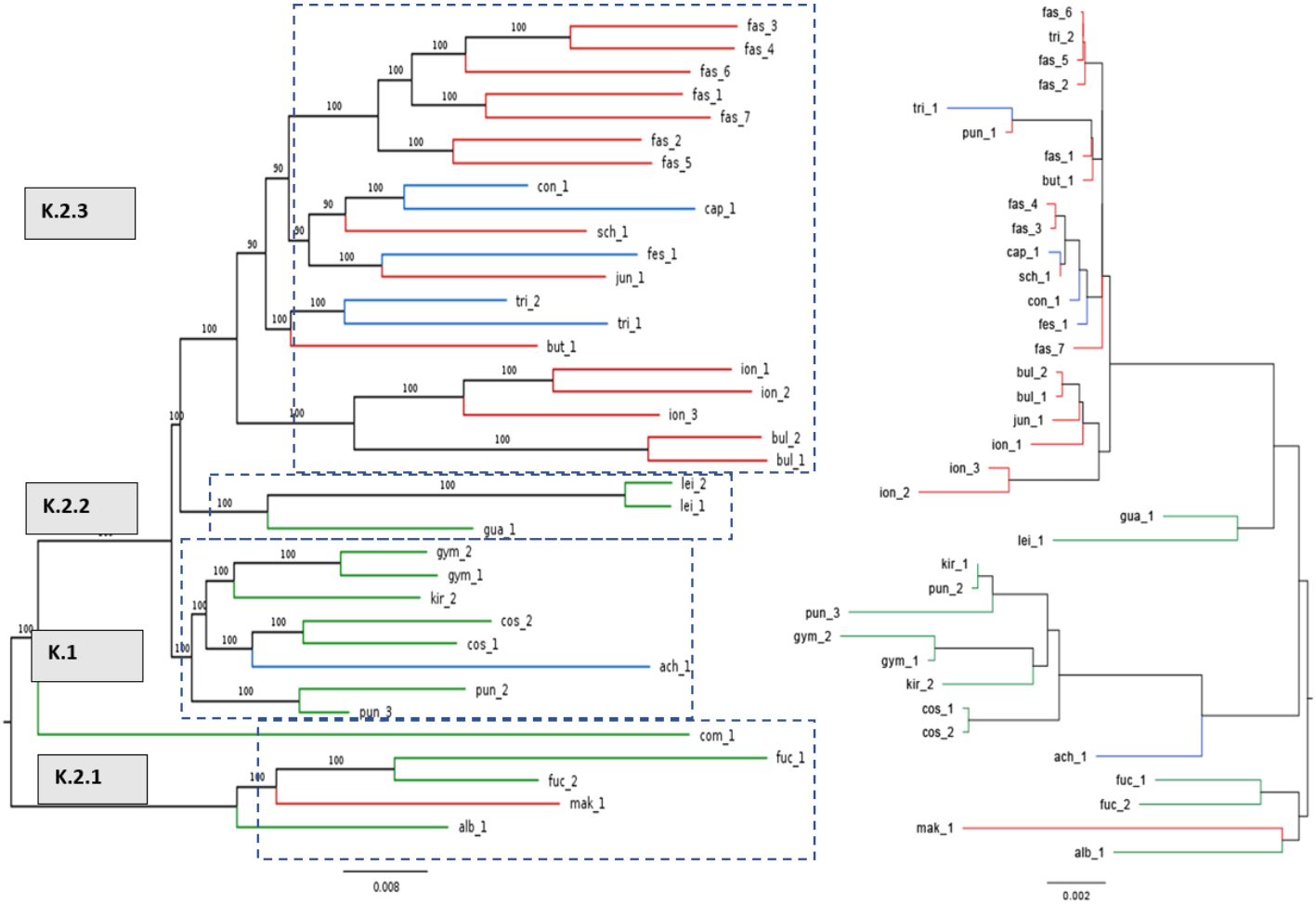


Figure 5. Maximum likelihood trees of the nuclear (left) and plastid (right) genomes produced in RAxML-NG. The groupings are based on the FineRADstructure heatmap (Figure 4). The branch colors represent photosynthetic pathways: green is C3, red is CAM, and blue is facultative.

Similar groupings are reflected on the RAxML-NG tree generated with the plastid data (Figure 5, right), although differences in branch lengths can be seen. Not all individuals are shown consistently throughout analyses because they were filtered as needed. The colors of the RAxML-NG branches assist one to visualize patterns created by photosynthetic pathways. Red represents CAM photosynthesis, green represents C3, and blue represents facultative C3/CAM, based upon stable isotope data from Mooney *et al*. (1989) and Crayn *et al*. (2014). The maximum likelihood tree has good bootstrap support – interestingly, the clades that are slightly less certain are the ones with the most individuals with facultative capabilities.

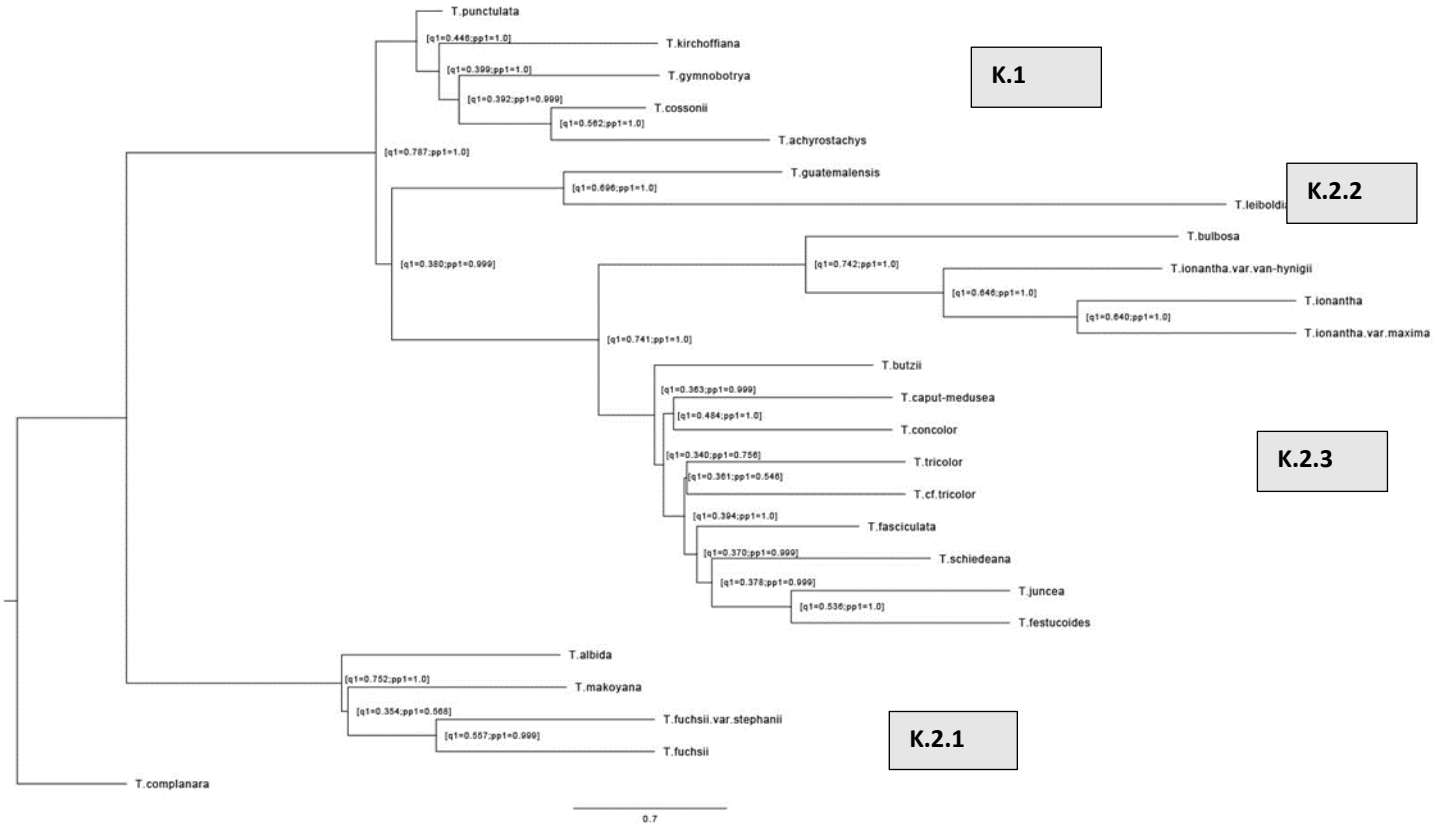


Figure 6. ASTRAL tree created with the nuclear genome, depicting the same supported clades as in the previous figures. The nodes depict branch lengths in coalescent units (q1) and branch support in the form of local posterior probability (pp1).

The coalescent tree created in ASTRAL (Figure 6) supports the RAxML-NG maximum likelihood tree. The coalescent values (q1 value on the tree) indicate the percentage of trees in which this clade was present. Although it is rooted differently, it shows clades K.1, K.2.1, K.2.2, and K.2.3 with the same relationships. This supports the hypothesis that our whole-genome data produced comparable trees to previous studies that used much fewer markers. Although the order of individuals within respective clades differs slightly, the groupings of individuals is consistent between the whole genome trees and the few marker trees.

The ABBA-BABA D-statistics positive values averaged at 0.055, with the highest value at 0.12. The negative values averaged at -0.047. D- statistics with a P value greater than 0.05 were discounted. In addition, most of the ABBA-BABA signal was due to chromosome 19, which deserves further attention in future research.

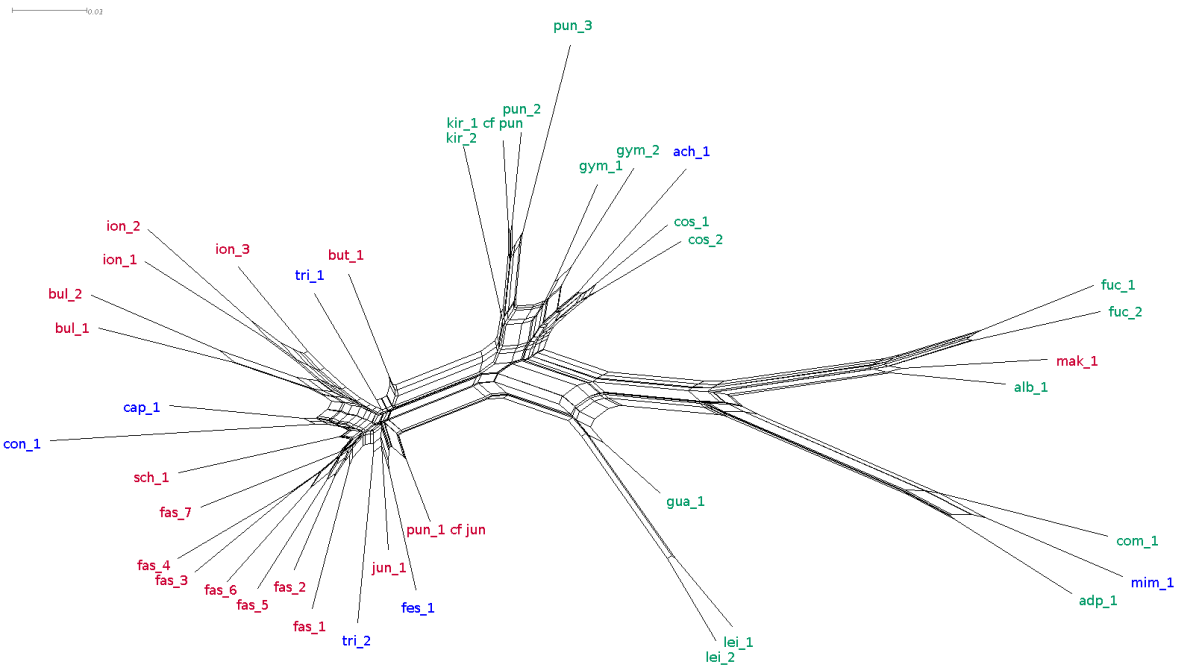


Figure 7 Splitstree phylogenetic networks. Nuclear genome. Individuals were colored based on their photosynthetic system: Red represents obligate CAM, green represents obligate C3, and blue represents facultative.

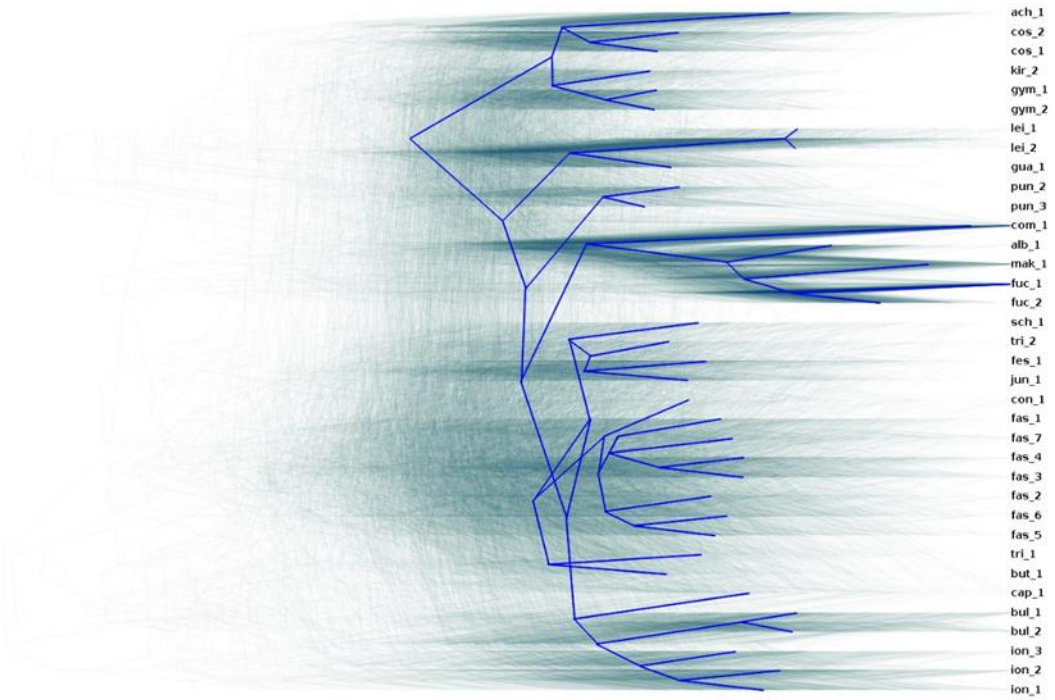


Figure 8. Densitree showing the complexity of the relationships within our dataset with 1000 trees created from 100kb windows, with the root canal option utilized.

The underlying complexity of relationships along the genomes of the studied species is supported by the noisy resolution of the overlaid trees (Figure 8) and of the reticulations formed in the Splitstree network (Figure 7). The aforementioned clusters are more easily viewed in the Splitstree than in the Densitree graph but are present in both. In the Densitree representation, the overlaid trees extend far beyond the root canal tree, suggesting that the complexity arose early in the clades.

Discussion

Throughout all analyses, four groupings have consistently been covered. Both the maximum likelihood and coalescent trees based on our whole-genome data (Figures 5 & 6) corroborate the gene trees presented in previous studies (Barfuss *et al.,* 2016; Mendoza *et al.,* 2017). Differences may be accounted for by the larger number of markers throughout the whole genome. Small discrepancies occur not only between our analyses and previous studies, but also between analyses within our study. Single individuals' relationships differ between plastid and nuclear genome analyses due to the plastid dataset's fewer sites, along with its uniparentally inherited and more conserved nature. Furthermore, branch lengths in the plastid analyses are shorter than the nuclear counterparts. That is likely due to the more conserved nature of plastid genomes, and thus, their lower levels of variation (Liu *et al.*, 2018), connected with the general absence of recombination in plastid genomes. Additionally, mutations among functional genes could be detrimental. According to Maier and Schmitz-Linneweber (2004), the majority of the genes found in the plastid genome regulate photosynthetic processes.

The noisy resolution that can be viewed in the Densitree plot (Figure .) and by the ASTRAL coalescent values (Figure 6.) support the previous hypothesis about *Tillandsia* spp. having recently diverged (Givnish *et al.*, 2011, 2014). The complexity of their relationships can be explained by incomplete lineage sorting (ILS) and hybridization (Abbott *et al.*, 2013; Choleva *et al.*, 2014; Goetze *et al.*, 2017) , also suggested by Patterson's D-statistics supported by significant P values (0.05 significance). As can be seen in Dasmahapatra *et al*. (2012) and Kronforst *et al.* (2013), it is not uncommon for Patterson's D-statistics results to be averaged around 0.05, as ours were. Given that most of the values were positive, this is implicative of ABBA (as opposed to BABA) conformation. In this project, we did not statistically test whether hybridization or ILS was the more probable cause behind the variation in subgenus *Tillandsia,* but it is an interesting matter that deserves attention in future work.

Tillandsia spp. are thought to have formed in the Miocene (28.4 – 7.2 Mya), which was characterized by a generally warmer and more humid climate worldwide. This was followed by climate change oscillations (Sass and Specht, 2010), the most significant of which were those that occurred in the Pleistocene. Known for its glacial/climate fluctuations, its ice shields waxed and waned, and the drought and cooler temperatures are hypothesized to have been the drivers for many species' adaptive radiations. The running hypothesis already raised by several researchers (Hewitt, 2000; Davis and Shaw, 2001; Knowles, 2003; Palma-Silva *et al.*, 2009; Paggi *et al.*, 2015; Mendoza Granados *et al.*, 2017) is that the Pleistocene glaciations may have been a driver for species divergence. Since CAM is an adaptation for aridity, which was an effect of the Pleistocene in many regions, and certain (younger) clades in our study are mostly CAM, a reasonable hypothesis connecting the origin of CAM with Pleistocene climatic changes can be made, to be further investigated in the future.

Perhaps the obligate CAM species settled in a more arid area while the obligate C3 were in refugia (Gilmartin, 1983). In our case, our subset radiated quickly, but seemed to hold on to the ability to revert to other photosynthetic processes when necessary, which can be seen in Figure 5 with the different photosynthetic processes and Figures 7 & 8 regarding ILS.

Clades K.2.1, K.2.2, and K.1 are predominantly C3, whereas clade K.2.3 is predominantly CAM. Clade K.2.3 is made up of mostly CAM plants (Mooney *et al.*, 1989; Crayn *et al.*, 2015; de La Harpe *et al.*, 2018), while the smaller three are a mixture, although mostly C3. Crayn *et al.*, (2015) determined whether a plant utilized C3 or CAM photosynthesis with carbon isotope ratios (13C). If the sample was determined to have a greater negative value than -20%, the plant was considered as C3; if it was less negative, it was named CAM (Silvera *et al*., 2010). There exist plants that are capable of both aforementioned types of photosynthesis, reacting to the environment around them. Mooney *et al*. (1989) used slightly different criteria: species within the range of -20% to -14% were considered

facultative CAM, whereas obligate CAM was anything less negative than -14% (Mooney *et al*., 1989). Based upon these criteria, the branch labels were colored in Figures 5 & 7 to help visualize this.

In the nuclear genome, the first individual of interest is mak_1 (*T. makoyana*), within clade K.2.1, which is predominantly C3, except for mak_1. Its δ13C value of -13.9% makes it an obligate CAM plant, within a group of obligate C3's. As we do not have individuals of the entire subg. *Tillandsia* within our dataset, we do not yet know the entire story, but it would appear that *Tillandsia makoyana* evolved CAM photosynthesis independently from the rest of the individuals in our dataset. This suggests that *T. makoyana* adapted to a niche with more arid conditions than those of *T. complanata*, *fuchsii*, and *T. albida*. Group 2 has a similar occurrence, with the exception that ach_1 (*T. achyrostachys*) became facultative CAM, as opposed to obligate CAM. This suggests an environment in flux, forcing the *T. achyrostachys* to be flexible. As with Group 1, the species that came after ach_1 in Group 2 kept, or possible reverted to, the C3 pathway (Crayn, Winter and Smith, 2004). Group 3 is made up of obligate C3 individuals, which share a large amount of coancestry (Figure 4.).

Clade K.2.3 is of interest because it is predominantly CAM, but has facultative CAM species sprinkled in. Two possible scenarios can be inferred from this: i) K.2.3 evolved as obligate CAM species, and then a few select species reverted to C3, or ii) obligate CAM species evolved from facultative CAM species.

Interestingly, our clades are also supported morphologically. According to Gardner's (1986) *Tillandsia* Classification key, our clade K.2.1, the outgroup, is placed into Gardner's Group II, based on its round filament cross-sections and open corolla throats, while the rest of our samples belong to Group I, based on flat filament cross-sections and closed corolla throats. Although these species were not separated based on photosynthetic pathways, this observation further supports our phylogenetic hypotheses based on the fact that they are morphologically different. Nevertheless, our knowledge grows with every discovery, and this is a topic that deserves further investigation. *Tillandsia*'s adaptive radiations will be better understood with more complete sampling of taxa and genomes in the future.

References

Abbott, R. et al. (2013) 'Hybridization and speciation', Journal of Evolutionary Biology, 26(2), pp. 229–246. doi: 10.1111/j.1420-9101.2012.02599.x.

Barfuss, M., Till, W., Leme, E., Pinzón, J., Manzanares, J., Halbritter, H., Samuel, R., & Brown, G. (2016). Taxonomic revision of Bromeliaceae subfam. Tillandsioideae based on a multi-locus DNA sequence phylogeny and morphology. Phytotaxa, 279(1), 1–97 doi:http://dx.doi.org/10.11646/phytotaxa.279.1.1

Benzing , David. (2000). Bromeliaceae: Profile of An Adaptive Radiation. 10.1017/CBO9780511565175.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. Bioinformatics, 26(10), 1372–1373. https://doi.org/10.1093/bioinformatics/btq110

Cáceres González, D. A., Schulte, K., Schmidt, M., & Zizka, G. (2011). A synopsis of the Bromeliaceae of Panama, including new records for the country. Willdenowia, 41(2), 357–369. https://doi.org/10.3372/wi.41.41216

Chase MW, Hills HG. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40:215-220

Choleva, L. et al. (2014) 'Distinguishing between incomplete lineage sorting and genomic introgressions: Complete fixation of allospecific mitochondrial DNA in a sexually reproducing fish (Cobitis; Teleostei), despite clonal reproduction of hybrids', PLoS ONE, 9(6). doi: 10.1371/journal.pone.0080641.

Crayn, D. M., Winter, K., & Smith, J. A. C. (2004). Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. Proceedings of the National Academy of Sciences, 101(10), 3703–3708. https://doi.org/10.1073/pnas.0400366101

Crayn, D. M., Winter, K., Schulte, K., & Smith, J. A. C. (2015). Photosynthetic pathways in Bromeliaceae: Phylogenetic and ecological significance of CAM and C3based on carbon isotope ratios for 1893 species. Botanical Journal of the Linnean Society, 178(2), 169–221. https://doi.org/10.1111/boj.12275

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert Handsaker, Gerton Lunter, Gabor Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin and 1000 Genomes Project Analysis Group, Bioinformatics, 2011 The Variant Call Format and VCFtools

de La Harpe, M. et al. (2018) 'Genomic footprints of repeated evolution of CAM photosynthesis in tillandsioid bromeliads', bioRxiv, p. 495812. doi: 10.1101/495812.

Dasmahapatra, K. K. et al. (2012) 'Butterfly genome reveals promiscuous exchange of mimicry adaptations among species', Nature, 487(7405), pp. 94–98. doi: 10.1038/nature11041.

Doyle & Doyle. 1987. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochemical Bulletin* 19(1): 11-15.

Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN]* 2012. v1.0.0

Garrison E. Vcflib, a simple C++ library for parsing and manipulating VCF files. 2016. https://github.com/vcflib/vcflib.

Goetze, M. et al. (2017) Incomplete lineage sorting and hybridization in the evolutionary history of closely related, endemic yellow-flowered Aechmea species of subgenus Ortgiesia (Bromeliaceae), American Journal of Botany, 104(7), pp. 1073–1087. doi: 10.3732/ajb.1700103.

Huson, D. H. and Bryant, D., Application of Phylogenetic Networks in Evolutionary Studies, Mol. Biol. Evol., 23(2):254-267, 2006.

Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012 Jan;40(1):e3. doi: 10.1093/nar/gkr771. Epub 2011 Oct 21. PubMed PMID: 22021376; PubMed Central PMCID: PMC3245947.

Korneliussen, T., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15:356. https://doi.org/10.1007/978-1-4614-1174-1_8

Kozlov, A. M., Darriba, D., Morel, B., & Stamatakis, A. (2018). RAxML-NG : A fast , scalable , and user-friendly tool for maximum likelihood phylogenetic inference, 1–5.

Kronforst, M. R. et al. (2013) Hybridization Reveals the Evolving Genomic Architecture of Speciation, Cell Reports. The Authors, 5(3), pp. 666–677. doi: 10.1016/j.celrep.2013.09.042.

Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012, 9:357-359.

Li H.*, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9. [PMID: 19505943]

Lischer HEL and Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28: 298-299.

Liu, L. et al. (2018) Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (Saxifragaceae), using genome skimming data, BMC Genomics. BMC Genomics, 19(1), pp. 1–17. doi: 10.1186/s12864-018-4633-x.

Maier R.M., Schmitz-Linneweber C. (2004) Plastid Genomes. In: Daniell H., Chase C. (eds) Molecular Biology and Biotechnology of Plant Organelles. Springer, Dordrecht

Malinsky, M., Trucchi, E., Lawson, D., & Falush, D. (2018). RADpainter and fineRADstructure: population inference from RADseq data. BioRxiv, 57711. https://doi.org/10.1101/057711

Mendoza Granados, C., Granados-Aguular, X., Donadio, S., Salazar, G. A., Flores-Cruz, M., Hagsater, E., … Magallon, S. (2017). Geographic structure in two highly diverse lineages of *Tillandsia* (Bromeliaceae). NRC Research Press, 95, 641–651.

Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010 Jun;2010(6):pdb.prot5448. doi: 10.1101/pdb.prot5448. PubMed PMID: 20516186.

Mirarab,S., R. Reaz, Md. S. Bayzid, T. Zimmermann, M. S. Swenson, T. Warnow; ASTRAL: genome-scale coalescent-based species tree estimation, Bioinformatics, Volume 30, Issue 17, 1 September 2014, Pages i541–i548, https://doi.org/10.1093/bioinformatics/btu462

Morel, B., Kozlov, A. M., & Stamatakis, A. (2018). ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. BioRxiv, 373449. https://doi.org/10.1101/373449

Mooney, A. H. A., Bullock, S. H., Ehleringer, J. R., & Ecology, S. F. (1989). Carbon Isotope Ratios of Plants of a Tropical Dry Forest in Mexico Published by : British Ecological Society

Paggi, G. M. et al. (2015) Limited pollen flow and high selfing rates toward geographic range limit in an Atlantic forest bromeliad, Flora: Morphology, Distribution, Functional Ecology of Plants. Elsevier GmbH., 211, pp. 1–10. doi: 10.1016/j.flora.2015.01.001.

Palma-Silva, C. et al. (2009) Range-wide patterns of nuclear and chloroplast DNA diversity in Vriesea gigantea (Bromeliaceae), a neotropical forest species, Heredity, 103(6), pp. 503–512. doi: 10.1038/hdy.2009.116.

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. (2014). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics (Oxford, England), 31(5), 770-2.

Silvera, K. et al. Evolution along the crassulacean acid metabolism continuum. Funct. 563 Plant Biol 37, 995–1010 (2010).

Tomlinson, P. B., & Metcalfe, C. R. (1969). Anatomy of the monocotyledons: III.

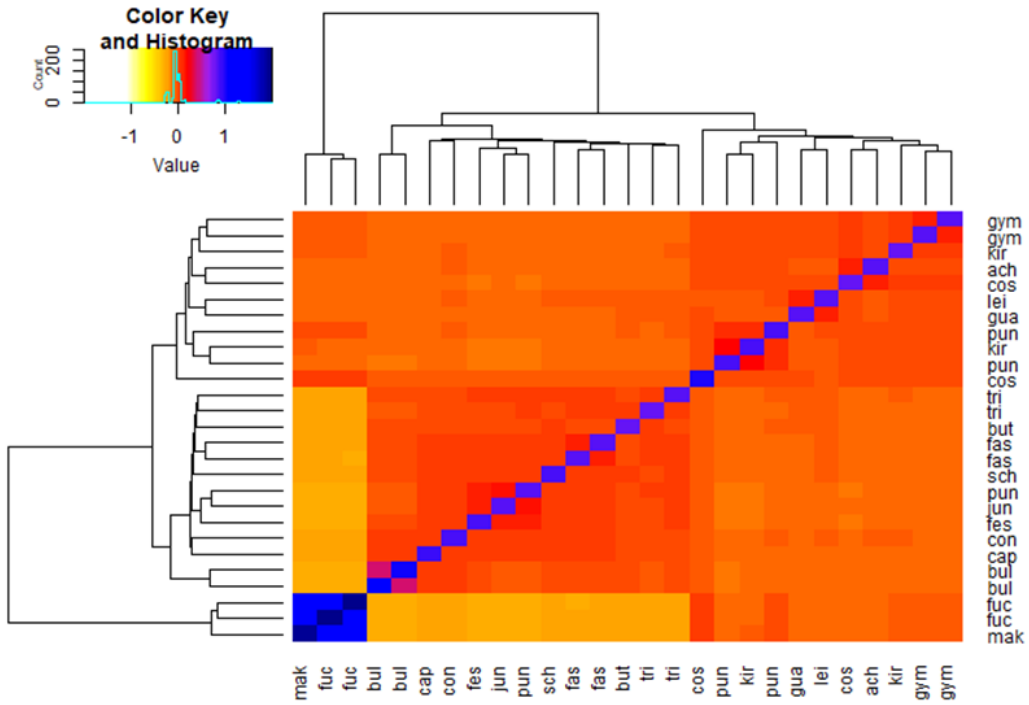
Winter, K. and Holtum, J. A. M. (2014) 'Facultative crassulacean acid metabolism (CAM) plants: Powerful tools for unravelling the functional elements of CAM photosynthesis', Journal of Experimental Botany, 65(13), pp. 3425–3441. doi: 10.1093/jxb/eru063.

Winter, K., & Smith, J. A. C. (1996). An Introduction to Crassulacean Acid Metabolism. Biochemical Principles and Ecological Diversity, (January). https://doi.org/10.1007/978-3-642-79060-7

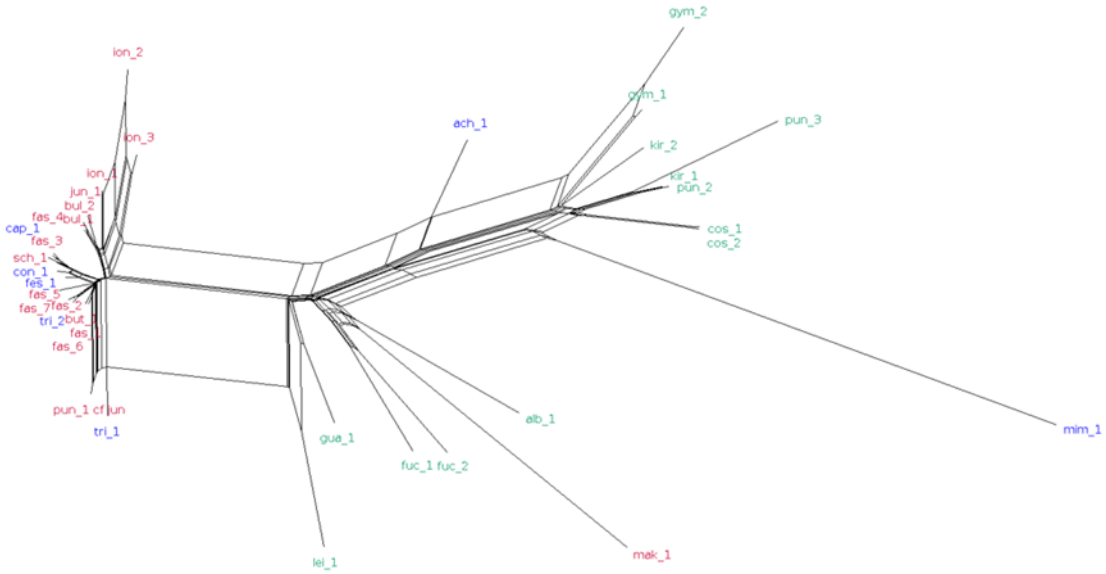
Supplement

Quick-start protocol CTAB (modified)

1. Centrifuge ground plant material 10 sec.
2. Add 400 µl high-salt 2x CTAB isolation buffer*, 40 µl PVP-40 (20%), 20 µl sarkosyl (20%). Mix by vortexing. ≥ 15 min at 60° C . Invert 2-3 times.
   *Optional: 1% volume 2-mercaptoethanol.
3. Centrifuge 10 sec.
4. Add 400 µl CIA. Mix by inverting. Incubate 5 min at room temperature.
5. Centrifuge 5 min at ≥18,000 x g.
6. Supernatant to new 1.5 ml tube.
7. Add 25 µl RNase A (10 mg/ml). Vortex. Incubate ≥ 10 min at 37° C.
8. Add 1/10 volume sodium acetate, and 400 µl CIA. Mix by inverting. Incubate ≥ 5 min at 4° C.
9. Centrifuge 5 min at ≥ 18,000 x g.
10. Supernatant to new 2 ml tube.
11. Add 2.5x volume 100% ethanol. Mix by inverting. ≥5 min at 4° C.
12. (Cooling) centrifuge ≥10 min at ≥ 18,000 x g, 4° C. Discard supernatant.
13. Add 500 µl 70% ethanol. Incubate 5 min at room temperature.
14. Centrifuge ≥ 5 min at ≥ 18,000 x g. Discard supernatant.
15. Repeat steps 13 and 14.
16. Add 500 µl 100% ethanol. Incubate 5 min at room temperature.
17. Centrifuge ≥ 5 min at ≥ 18,000 x g. Discard supernatant.
18. Dry the pellet.
19. Dissolve the pellet in ddH_2O.



Supplementary Figure 1. Angsd heatmap using the nuclear data supports Figure 4, although shown in a different arrangement.



Supplementary Figure 2. Splitstree network of plastid genomic data. Reticulations represent homoplasy, not hybridization.