



Introduction

Impatiens is a genus of flowering plant that has species hotspots in places like tropical Africa, the Himalayan region, Madagascar, South India and Sri Lanka, and Southeast Asia (Janssens et al. 2009). The plants prefer to be in locations that are these tropical regions for the excess moisture available; as a result, they tend to migrate up mountain slopes to avoid high fluctuation in heat or moisture, which often leads to isolation between closely related lineages (Janssens et al., 2009). The Eastern Arc Mountains (EAM) are a mountain range that is believed to be one of the oldest tropical rainforest for several reasons, but among them are geological age and long-lasting stability in terms of moisture and climate. This also makes them a major biodiversity hotspot especially for *Impatiens*, with many of the species' endemic to one or multiple mountain blocks (areas). Unfortunately, there isn't a lot of information available for many taxa in the EAM region because it has remained largely uninvestigated (Jump et al., 2014; Blackburn & Measey, 2009). Studying *Impatiens* in the Eastern Arcto understand its genetic variation, gene flow and structure, can help us understand how biodiversity evolved in this biologically important area. Gene flow and genetic variation can influence the amount of inbreeding, helping us to further understand the effects of environmental and anthropogenic changes to prevent major die-off or extinction. This will also fill in gaps in the general information we have on this genus, and by extension better understand plant endemism in the Eastern Arc. For this study, we generated and analyzed a SNP dataset for multiple populations of Eastern Arc endemic *Impatiens engleri* and *Impatiens keilii*.



Impatiens engleri



Impatiens keilii

Hypothesis

Because *I. engleri* is a primarily butterfly pollinated species, while *I. keilii* is primarily bird pollinated species, we hypothesized that *I. engleri* populations will have lower gene flow and thus more structure between mountain blocks likely leading to more inbreeding. On the other hand, *I. keilii* will have lower population structure and lower inbreeding. This is due to the bird having the ability to fly further distances than a butterfly, as such it is likely that it can land and pollinate multiple populations in different mountain blocks (Nazareno et al, 2021).

Methods

DNA Extraction + Library Prep:

Library prep was completed with a double digest restriction-site associated DNA (ddRAD) method. Total DNA was extracted using a modified CTAB DNA extraction protocol using silica-dried leaves (Doyle & Doyle 1987); the extracted DNA was then digested by EcoRI and MspI enzymes, barcodes were added, and samples pooled into 3 sub-libraries, then fragments were amplified. Fragments were size selected for 350 base pair (bp) inserts for sequencing on NovaSeq.

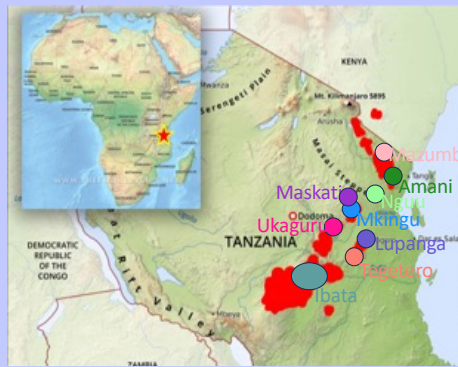


Figure 1. A map of the Eastern Arc Mountains. The red markers are the different mountain blocks. Colored circles correspond to populations included in sampling.

Data Analysis:

We used STACKS (Catchen et al. 2014) pipeline for Denovo assembly of raw reads: process_radtags followed by denovo_assembly.pl Samples with low reads were discarded. Parameter optimization was completed with various permutations of Minimum stack depth (-m), Distance allowed between stacks (-M), and Distance allowed between catalog loci (-n). The various parameter permutation assemblies were completed with both r80 and r60. m=3, M=3, and n=3 were selected as the best parameters for highest quality and quantity of loci recovered. The denovo_assembly.pl generated a .vcf of SNPs, which we filtered using vcftools (Danecek et al. 2011) for loci in Hardy Weinberg equilibrium: minimum minor allele frequency (maf) .05; minimum read depth 5, loci with at least 2 alleles, maximum missingness of 40%. Fst (Weir & Cockerham, 1984) was estimated using the hierfstats package. The expected and observed heterozygosity and inbreeding statistics were estimated using hierfstats (Goudet, 2005), adegenet (Jombart, 2008), and dartR (Gruber, 2018); PCs were estimated using plink (Weeks, 2010).

References

Blackburn, D.C. and Measey, G.J. (2009). Dispersal as or from an African biodiversity hotspot? *Molecular Ecology*, 18, 1904-1912.
 Catchen, J.M. et al. (2011) Stacks: Building and Genotyping Loci de Novo from Short-Read Sequences. *Genome Genomics Genetics*, Volume 1, Issue 3: 171-182.
 Danecek, P. et al. (2011) The variant call format and VCFtools. *Bioinformatics*, Volume 27, Issue 15: 2156-2158. <https://doi.org/10.1093/bioinformatics/btr330>
 Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin*, 19, 11-15.
 Janssens, S. B. et al. (2009) Rapid isolation of *Impatiens* (Balanophytes) during Pleistocene and Pliocene: Result of a global climate change. *Molecular Phylogenetics and Evolution*, Volume 52, Issue 3: 806-824.
 Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5: 184-186. <https://doi.org/10.1111/j.1471-8266.2004.00828.x>
 Gruber, B. et al. (2018), dart: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour*, 18: 691-699. <https://doi.org/10.1111/1755-0998.12345>
 Jump, A.S. et al. (2014). Genetic Divergence During Long-term Isolation in Highly Diverse Populations of Tropical Trees Across the Eastern Arc Mountains of Tanzania. *Biotropica*, 46, 565-574.
 Nazareno, A. G. et al. (2021) By Animal, Water, or Wind, Can Dispersal Mode Predict Genetic Connectivity in Riverine Plant Species? *Plant Plant Sci*, Volume 1.
 Jombart, S. (2008) adegenet: an R package for the multivariate analysis of genetic markers. *Bioinformatics*, Volume 24, Issue 7, 1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>
 Weeks, J. P. (2010). plink: An R Package for Linking Mixed-Format Data Using IRT-Based Methods. *Journal of Statistical Software*, 35(2), 1-33. <https://doi.org/10.18637/jss.v035.i02>
 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), 1358-1370.

Results

The stacks denovo assembly pipeline found variant 9204 loci for *I. engleri*. 260 loci were retained after filtering. 9171 loci were found for *I. keilii*; 604 loci were retained after filtering. *Im. keilii* pairwise Fst ranged from >0.001 – 0.52 with a median Fst of 0.29. *I. engleri* pairwise Fst ranged from 0.12 – 0.78 with a median Fst of 0.61.

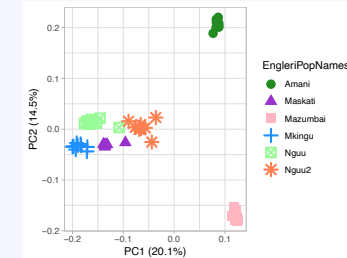


Figure 2. Principle Component Analysis of *Impatiens engleri* SNP dataset.

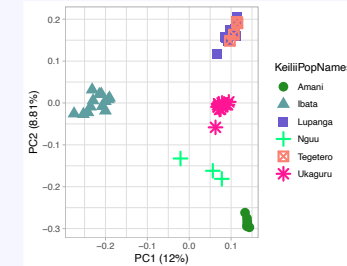


Figure 4. Principle Component Analysis of *Impatiens keilii* SNP dataset.

	Ho	He	Fis	genotype.counts
Mkingu	0.12513678	0.12256839	-0.02322687	6, 996960
Maskati	0.13435556	0.10725714	-0.21822093	5, 120635
Mazumbai	0.07200608	0.07055015	0.007806452	12, 860182
Amani	0.10161398	0.10587842	0.028244898	10, 805471
Nguu	0.08102564	0.07735452	-0.070574074	4, 288462
Nguu2	0.07430120	0.08888333	-0.151454545	2, 060241

Table 1. Homozygosity, heterozygosity and Fis of *Impatiens engleri* SNP dataset.

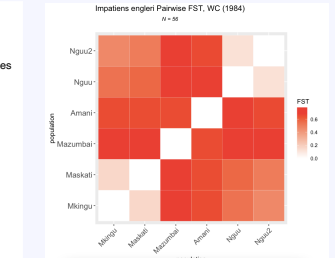


Figure 3. Pairwise Fst between *Impatiens engleri* populations.

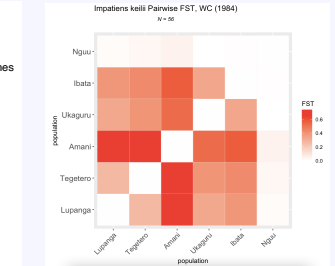


Figure 5. Pairwise Fst between *Impatiens keilii* populations.

	Ho	He	Fis	genotype.counts
Amani	0.05471028	0.04303037	-0.2880488	5, 502336
Ukuguru	0.14304861	0.13757407	-0.03304790	10, 023148
Tegetero	0.13234048	0.06722196	-0.91091379	3, 080952
Lupanga	0.08277844	0.07780723	-0.08753030	5, 309593
Nguu	0.00000000	0.00000000	NaN	2, 0000000
Ibatata	0.16307176	0.15512037	-0.03123445	11, 662037

Table 2. Homozygosity, heterozygosity and Fis of *Impatiens keilii* SNP dataset.

Discussion

- Pairwise Fst was overall higher in *I. engleri* compared to *I. keilii*, indicating that population structure is overall higher in this butterfly pollinated species. This supports the hypothesis that gene flow is higher in bird pollinated *Impatiens* such as *I. keilii*.
- Gene flow is occurring in populations that occur on the same mountain blocs, such as Tegetero & Lupanga, Nguu 1 & 2, and Maskati & Mkingu.
- Even for bird pollinated *I. keilii* populations have moderate to pronounced structure, indicating that Eastern Arc mountains do not have strong gene flow between them.

Acknowledgements

Thank you to Suzy Strickler for all the help with the coding, and to NSF-REU grant DBI-1757800.