Conservation genomics of *Caesia parviflora var. minor* in support of management and translocation activities.

August 2022

FINAL REPORT

Research Centre for Ecosystem Resilience

Australian Institute for Botanical Science

This report was prepared for the environmental consulting group, RPS Australia Asia Pacific.

Acknowledgements

We acknowledge the Traditional Custodians of the land on which the plant species in this study are found on and pay respects to Elders past and present. We acknowledge the field collectors who helped with the sampling.

Citation

Yap J-YS, Rossetto M (2022) Conservation genomics of *Caesia parviflora var. minor* in support of management. Research Centre for Ecosystem Resilience, Royal Botanic Garden Sydney.

Contents

EXECUTIVE SUMMARY

Taxonomic status and genetic health of the endangered *Caesia parviflora* var. *minor* was assessed using high quality genome scans. We provide empirical evidence confirming that the variety and *C. parviflora* var. *parviflora* are genetically similar and thus should be reassessed as belonging to one taxon. As our study only included a subset of both, broader sampling across the geographic distribution of both is recommended to determine the true geographic extent of this taxon.

As for *C. parviflora* var. *minor* at Newnes, low genetic diversity is present, indicating that genetic rescue involving the mixing with nearby populations will be beneficial.

1. INTRODUCTION

1.1 Background

Caesia R.Br. comprises approximately 12 species of tufted, grass-like plants with delicate, short-lived, white, blue or purple flowers in which the perianth twists spirally when they fade (Henderson, 1987). The main diversity of the genus is in Australia, where eight species are found and seven are endemic.

One of the species, *C. parviflora* R.Br. also known as pale grass-lily, is a species complex and occurs widely in Australia. The species has lily-like flowers with three grey or purple stripes on each petal. There are three varieties of *C. parviflora*, with the larger varieties taller than 20cm and have broader leaves, such as var. *parviflora* which has white flowers and var. *vittata* which has blue flowers, whereas the smaller variety, var. *minor* is less than 20 cm tall, has narrower leaves and inconsistent flower colour (blue or white). Identification of *C.* var. *minor* can be challenging as it depends on flowering time, and because a *C.* var. *minor* can be confused with a stunted var. *parviflora* or *C. alpina*, both of which are more widespread. It is important to rigorously test the relationships among the varieties and species, as *C. parviflora* var. *minor* is currently listed as endangered on New South Wales' Biodiversity Conservation Act 2016, and verifying its taxonomic status is crucial to better manage its conservation.

Although *Caesia parviflora* var. *minor* has a wide distribution, in New South Wales, Tasmania, southern Victoria and South Australia, it is uncommon and only found in damp places in open forests on sandstone. One of its populations is situated near the Angus Place Colliery, approximately 15km northwest of Lithgow in the Blue Mountains region of New South Wales, where Centennial Coal is under obligation by the Environment Protection and Biodiversity Conservation Act 1999 (EPBC) to implement the Western Region Biodiversity Offset Strategy. As part of the offset obligations, Centennial identifies the potential impacts to threatened species and undertake the relevant conservation actions. Three individuals of *C. parviflora* var. *minor* were initially identified at Angus place in the Offset report. However, RPS an ecological consulting company suggests there could be up to 200 individuals based on unpublished findings and Bionet database (RPS 2016). Given the cryptic nature of *C. parviflora* and difficulty to identify the species, there is a need to verify the number of individuals (or diversity) present at Angus Place and undertake a phylogenetic reconstruction at species/subspecies level in relation to *C. parviflora* var. *minor*, in support of the assessment of the impact of vegetation clearing on the local population at Angus Place.

The Royal Botanic Gardens & Domain Trust (RBG&DT) was contracted by RPS to conduct a conservation genomics study on *C. parviflora* var. *minor* to provide foundational knowledge essential to the development of effective conservation strategies and help guide management strategies. The benefits of a single genetic study in the framework of conservation work are manifold as outputs can be used to rigorously test species status, quantitatively assess genetic diversity, determine genetic health, identify hybrids and provide practical solutions to long-term management strategies (Rossetto et al. 2021).

1.2 Aims of the conservation genomics study of *Caesia parviflora* var. *minor*

To support long-term management and conservation of *Caesia parviflora* var. *minor*, a conservation genomics study using whole-genome scanning (DArTseq and Targeted sequencing of 353 nuclear genes) was conducted with the following aims:

1. Ascertain whether *Caesia parviflora var. minor* merits taxonomic revision by examining it within its current phylogenetic framework.
2. Test for evidence of gene flow between varieties/species.
3. Assess the presence and extent of clonality, level of kinship and the genetic diversity in the Newnes population of *C. parviflora* var. *minor*.

2. METHODS:

2.1 Sampling

This study used two whole genome scanning approaches, DArTseq and targeted sequencing to gain an understanding of the genetic diversity in *C. parviflora* var. *minor* in Newnes and examine its taxonomic standing. Different sampling strategies were applied to the two approaches particularly because targeted sequencing is rigorous at clarifying taxonomic status of various species/subspecies/varieties whereas DArTseq is highly sensitive at distinguishing genetically differentiated individuals from clones within a population, measure overall diversity and heterozygosity, and differentiate between species and hybrids.

The sampling of *C. parviflora var. minor* was undertaken by the RPS officers. Sampling for the variety consisted of 29 samples from Angus Place Colliery (hereon referred to as the Newnes population), approximately 15km northwest of Lithgow in the Blue Mountains region of New South Wales. Additional two herbarium curated specimens of the variety were obtained as part of this sampling to verify the identity of the individuals at Angus Place Colliery. The specimens are NSW799729 from Arrawarra, NSW858060 from the study site in 2015. All 31 samples were included for DArTseq, whereas, the two mentioned herbarium samples and one sample from Newnes were included for targeted sequencing.

Other varieties/species were sampled for this study. For DArTseq, 25 individuals of *C. parviflora* var. *parviflora,* five individuals of *C. parviflora* var. *vittata* from Woodford, New South Wales and eight *C. alpina* from a range of locations within Alpine National Park, Victoria were sampled. For targeted sequencing of 353 nuclear genes, three *C. parviflora* var. *parviflora*, two *C. parviflora* var. *vittata*, two *C. alpina*, one *C. occidentalis* and an outgroup *Arthropodium minus* were included (Figure 1, Table 1).

2.2 DNA extraction and targeted sequencing analysis

Twelve samples were sent to AGRF for DNA extraction and targeted sequencing.

The Sequence data was generated using a targeted sequencing approach involving the Angiosperms353 universal probe set.

Once the data is returned, the software, HybPiper was used to generate target locus sequences, including assembly of gene regions, extraction of exon and intron sequences and identification of paralogous gene copies, for 11 samples of *Caesia* and the outgroup, *Arthropodium minus* (Table 1). The program was run using HybPiper-RBGV (https://github.com/chrisjackson-pellicle/HybPiper-RBGV), a Nextflow pipeline containerised using Singularity with Nextflow v21.04.3 and Singularity v3.8.3. The protein sequences for each recovered gene were aligned using Muscle v5.1, and gene trees were generated using iqtree2 v2.2.0. Then, the gene trees were combined into a single phylogenetic tree using ASTRAL v5.7.7.

2.3 DNA extraction and DArTseq analysis

Forty-four samples were sent to Diversity Arrays Technology (DArT) Pty Ltd in Canberra for DNA extraction and genotype-by-sequencing analysis (referred to as DArTseq analysis).

2.3.1 Quality screening and control of Single Nucleotide Polymorphism data

Single nucleotide polymorphisms (SNPs) data was quality checked using the filtering scripts implemented by an in-house designed package called RRtools package v1.0 (as described in Rossetto *et al.* 2019) in the open source program, R (version 3.3.0, R Core Development Team 2013). Loci that did not pass standard quality thresholds were removed from the data and were not used in downstream analysis. To ensure that only the higher quality DArTseq markers were used for analyses, all SNPs with a reproducibility (proportion of replicate assay pairs for which the marker score is consistent) of less than 96% and which had more than 30% missing data were excluded from the dataset.

2.3.2 Principal coordinate analyses

Adegenet 2.1.1 package in R (version 3.3.0, R Core Development Team) was used to perform a Principal Component Analysis (PCA) to better understand relationships between individuals and populations. This method of PCA derives an ordination based on Euclidean transformed dissimilarity matrix of the data.

2.3.3 Kinship

Genetic similarity between individuals located at the same site and corresponding cultivated material was estimated using the unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering method as implemented in the phanghorn package v2.4.0 in R. Kinship (relatedness) measurements were used in assessing the degree of kinship across *C. parviflora* var. *minor* in Newnes. Pairwise kinship coefficient was estimated from the genotype data using an Identity-by-descent analysis in SNPrelate package v1.17.1 in R.

3. RESULTS AND INTERPRETATION

3.1 Summary

We report results based on whole genome scanning of *Caesia parviflora* var. *minor* and related varieties and species, to assess species concepts and genetic history of the endangered *C. parviflora* var. *minor* and provide an understanding of remaining genetic diversity. Targeted 353 sequencing enabled differentiation between subspecies/species. DArTseq provided support for species differentiation, a quantification of genetic diversity within populations and an assessment of kinship and admixture.

The significant findings are:

- *Caesia parviflora* var. *minor* is not a distinct taxon as it is genetically akin to *C. parviflora* var. *parviflora;*
- Hybridisation was not detected between *C. parviflora* var. *minor* and other varieties/species;
- DArTseq data identified low genetic variability within the Angus Place Colliery population (Fig. 1);
- The low genetic variability is caused by the high relatedness among the individuals at Angus Place Colliery (Fig. 2), with 10 individuals belonging to Family 1 and sibling/half-sibling relationships are observed among 8 individuals, and 20 individuals belonging to Family #2 and sibling/half-sibling relationships are observed among 15 individuals;
- Individuals at Angus Place Colliery are genetically different from the individual from Arrawarra (Fig. 1).

3.2 *Caesia parviflora* var. *minor* and var. *parviflora* are one taxon.

Our genetic analyses using both approaches showed that specimens representing *C. parviflora* var. *minor* is not genetically distinct from *C. parviflora* var. *parviflora* despite both being distributed geographically distant from each other (Fig. 1, 2). The phylogenetic tree based on targeted sequencing data showed that the clade consisting of *C. parviflora* var. *minor* and var. *parviflora* is paraphyletic as there was a *C. parviflora* var. *parviflora* individual nested among the *C. parviflora* var. *minor* individuals. The phylogenetic tree however

resolved that *C. parviflora* var. *vittata* is distinct from this clade, and the outgroups *C. alpina* and *C. occidentalis* are distinct from all varieties of *C. parviflora*.

The Splitstree network based on DArTseq data of 7,945 genome-wide markers (SNPs) support that *C. parviflora* var. *minor* is not genetically distinct from *C. parviflora* var. *parviflora*. Both varieties were in a cluster on the main branch and are genetically distinct from *C. parviflora* var. *vittata* which comes out as a tip of a fork on the other end of the main branch. The other tip of the fork is further away from the main branch showing that *Caesia alpina* is the most genetically distant from the cluster of *C. parviflora* var. *minor* and var. *parviflora*.

Our study overall showed that in order to resolve the relationships among the *Caesia parviflora* complex (species/subspecies/varieties), a wider geographic sampling and taxonomic representation will be needed.

3.2 Genetic health of *Caesia parviflora* var. *minor* at Newnes

We report the genetic health of *C. parviflora* var. *minor* at Newnes given the interest by RPC. The PCA of the DArTseq data identified low genetic variability is present in the Newnes population (Fig. 3). The low genetic variability is contributed by the high relatedness among the individuals at Angus Place Colliery (Fig. 4), with 10 individuals belonging to Family 1 and sibling/half-sibling relationships are observed among 8 individuals, and 20 individuals belonging to Family 2 and sibling/half-sibling relationships are observed among 15 individuals.

4. CONCLUSION

This project highlights the following results:

- *Caesia parviflora* var. *minor* is genetically akin to *C. parviflora* var. *parviflora* such that both should be lumped as one taxonomic entity and the endangered status of C. parviflora var. minor should be reassessed;
- More intensive sampling of *C. parviflora* var. *minor* and var. *parviflora* is recommended to ascertain the geographic extent of the taxon that encompasses both *C. parviflora* var. *minor* and var. *parviflora;*
- Low genetic variability is present at the Newnes population indicating that genetic rescue through mixing with nearby *C. parviflora* var. *minor* or var. *parviflora* populations is likely to be beneficial.

5. FIGURES AND TABLES

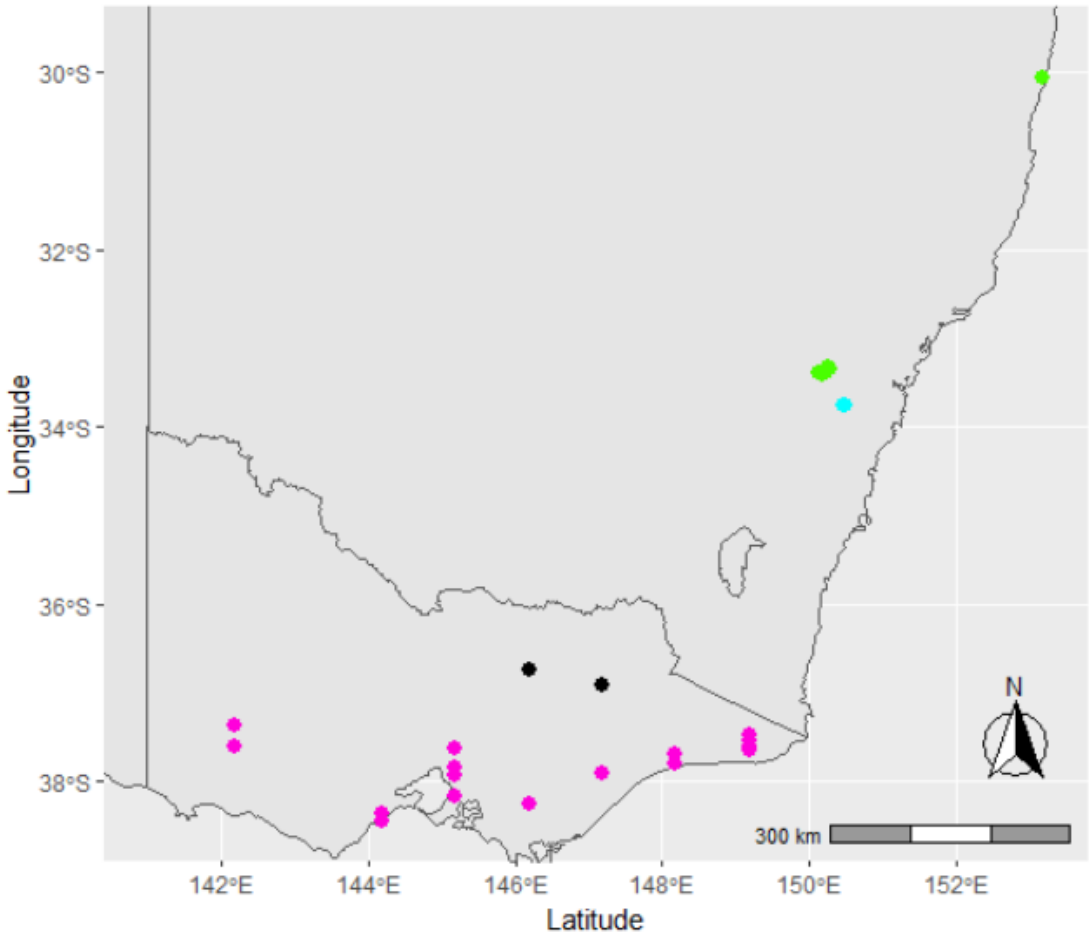


Figure 1: Distribution of sampling for various Caesias.

Each dot indicates a sampling site: green dots represent *C. parviflora* var. *minor*, blue dots represent *C. parviflora* var. *vittata*, maroon dots represent *C. parviflora* var. *parviflora* (approximate location provided) and black dots represent *C. alpina*. *Caesia occidentalis* from Western Australia was not included in the map.

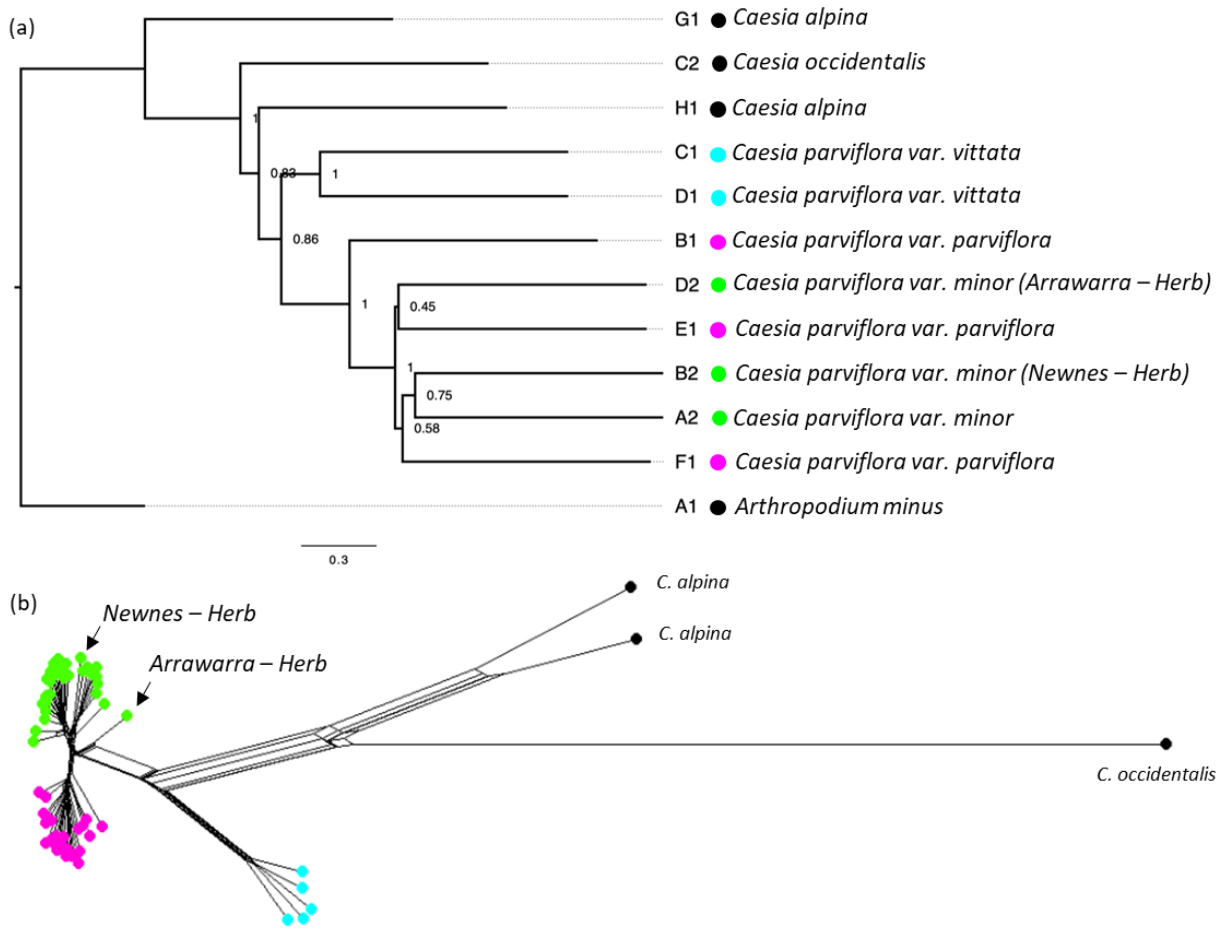


Figure 2: (a) Astral phylogenetic tree of *Caesia* samples (N=12) constructed from targeted sequencing data and (b) Splitstree network analysis of the DArTseq SNP data across *C. parviflora* var. *minor* (N=31), *C. parviflora* var. parviflora (N=25), *C. parviflora* var. *vittata* (N=5) and *C. alpina* (N=8).

In (a), the phylogenetic tree was constructed from 339 gene trees that are of good quality (Appendix 1). Branch lengths in coalescent units and branch support values measure the support for a quadripartition. Refer to Table 1 for details on the 12 samples.

Table 1: Sample metadata on *Caesia* used in targeted sequencing.

Sample	NSW number	Species	Herbarium specimen	Location
A1	NSW1093884	*Arthropodium minus*		NA (outgroup)
B1	NSW1078306	*C. parviflora* var. *parviflora*	Y	Cumberland Plain TERN Supersite, NSW
C1	NSW1078303	*C. parviflora* var. *vittata*	Y	North east of upper Sheba Dam, Hanging Rock NSW
D1	NSW1078304	*C. parviflora* var. *vittata*	Y	Mt Annan Botanic Garden, Woodland Conservation Area, NSW
E1	NSW1093883	*C. parviflora* var. *parviflora*	Y	Thirlmere Lakes National Park, NSW
F1	NSW1093714	*C. parviflora* var. *parviflora*		Bemm Forest, Little River Track, VIC
G1	NSW1093750	*C. alpina*		Bogong High Plains, Alpine National Park, VIC
H1	NSW1093648	*C. alpina*		Mt Buffalo NP, ca. 100m N of the base of The Cathedral, VIC
A2	NSW1092379	*C. parviflora* var. *minor*		Beecroft Firetrail near Kangaroo Creek, Marrangaroo NSW
B2	NSW1078302	*C. parviflora* var. *minor*	Y	Newnes Plateau NSW
C2	NSW1093855	*C. occidentalis*	Y	Beekeepers Rd, Eneabba WA
D2	NSW1093868	*C. parviflora* var. *minor*	Y	Garby Nature Reserve, Arrawarra NSW

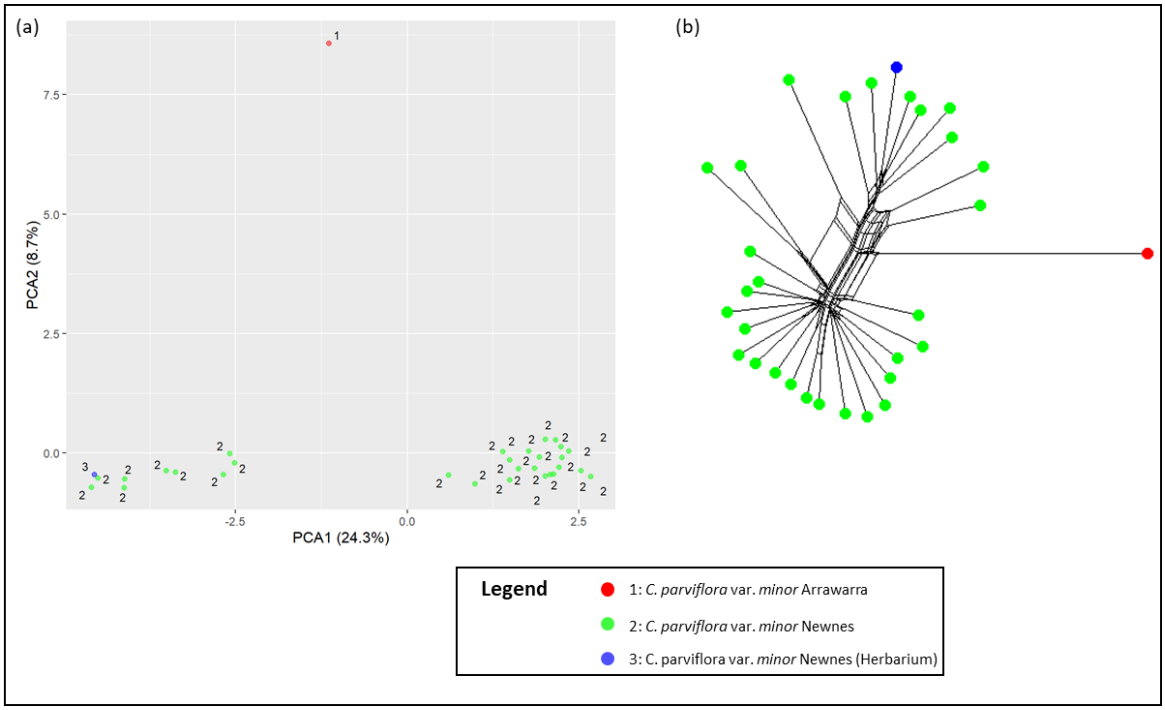


Figure 3. **Principle Component analysis (a) and Splitstree network analysis (b) of the SNP data from 31 specimens of *Caesia parviflora* var. *minor.***

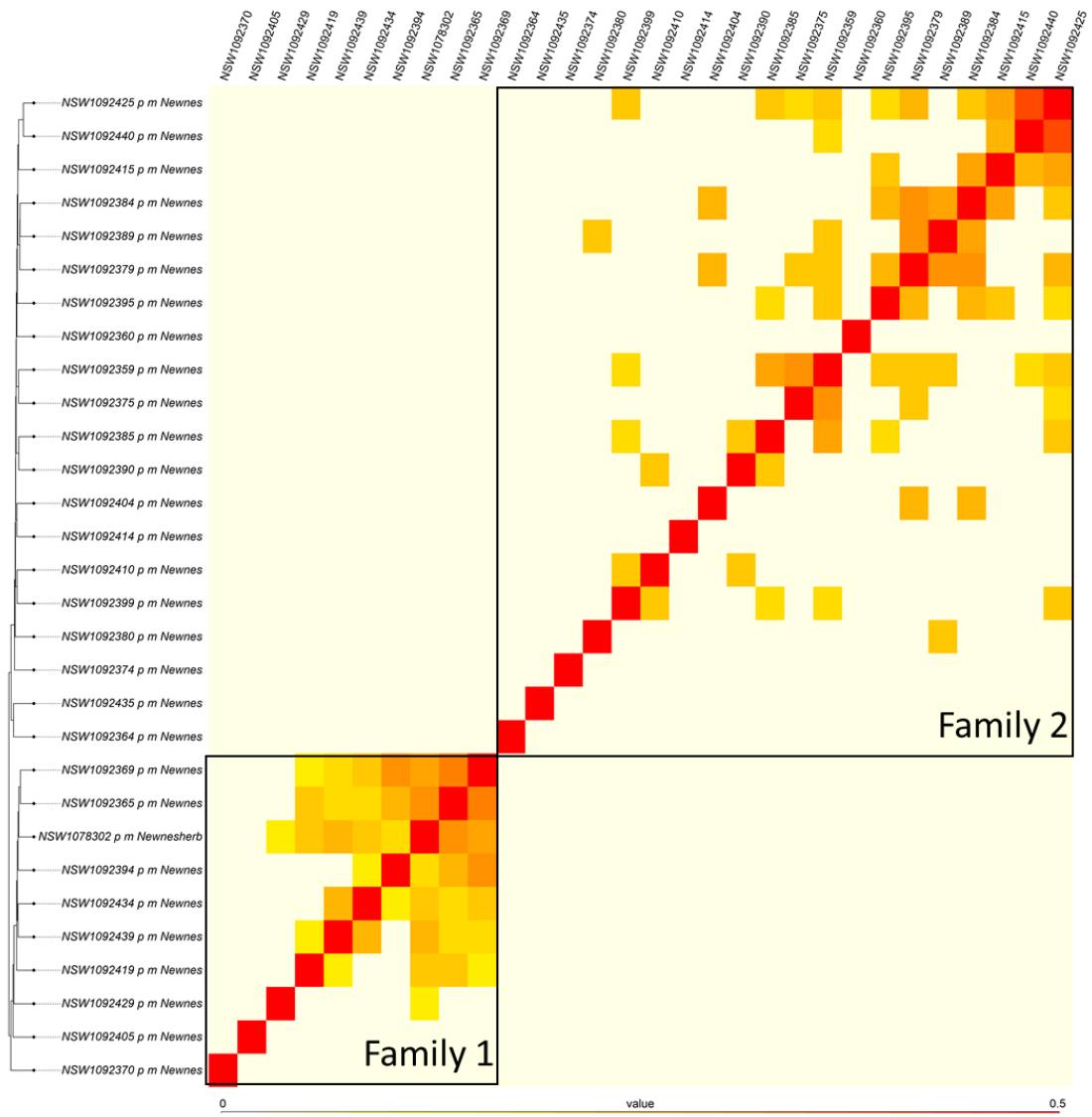
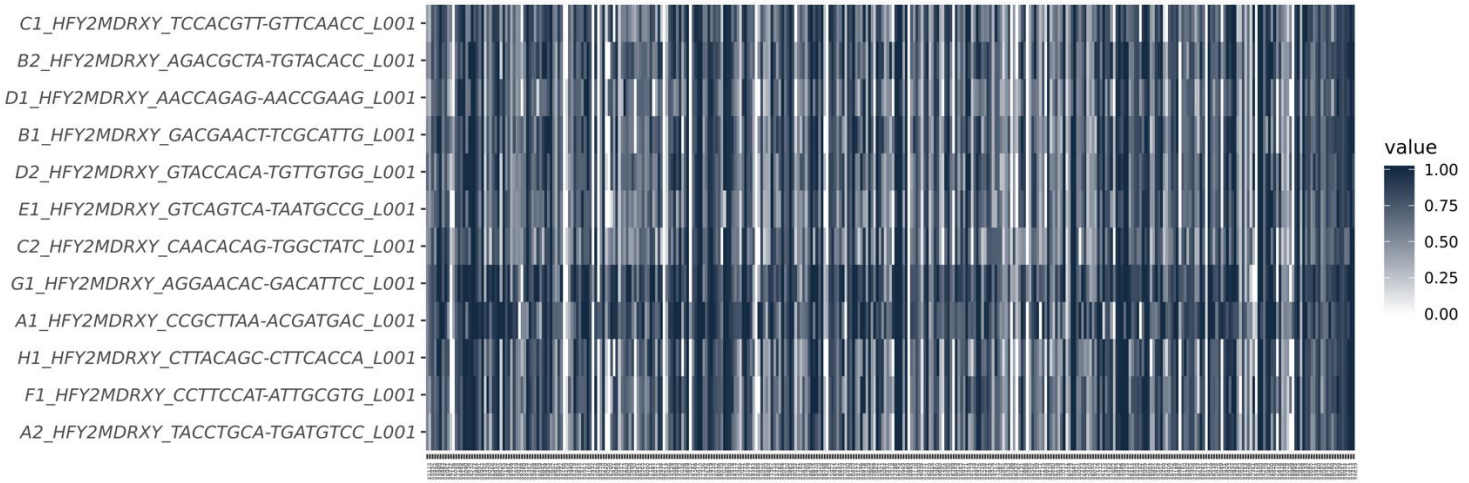


Figure 4. **Composite UPGMA tree/Kinship heatmap analysed from the SNP data for 31 specimens of *Caesia parviflora* var. *minor*.**

All 31 specimens were studied in a pairwise kinship analysis, and this resulted in the heatmap above consists of pairwise kinship coefficients displayed as colours: Red (high Kinship; 0.4 or greater = clone), Orange-Yellow (siblings) -White (no relationship). The descending red diagonal on the graph is the result of an individual matched with itself.

To the left of the heatmap is a UPGMA tree generated from a distance-based matrix of the SNP data. The tree tip labels include each *C. parviflora var. minor* individual's unique ID (i.e. NSW number) and site information (see Table 2 for details on genets of multiple ramets).

Appendix



Appendix Figure 1: Heatmap of gene recovery and length per sample from *Caesia* specimens prepared using the Angiosperms353 universal probes.

Appendix Table 1. Statistics on the HybPiper-RBGV run. Columns with percentage represent number of genes with sequences > X% of the target length.

Name	Reads	Mapped	On target (%)	Mapped	With contigs	With seqs	25%	50%	75%	150%	Paralog warnings	Without supercontigs	Supercontigs	Trimmed supercontigs	Supercontig skipped	Potential paralogs
A2	21320578	1140958	5.4	349	346	329	315	251	167	0	8	111	218	175	0	50
F1	29506873	1486057	5	351	348	329	316	270	178	0	5	108	221	183	0	65
H1	19671297	974149	5	348	343	324	307	253	173	0	23	106	218	203	0	40
A1	10564863	1862875	17.6	346	345	340	334	299	210	0	4	147	193	143	0	78
G1	22811110	1319421	5.8	352	347	340	334	297	226	0	42	95	246	238	0	35
C2	19711671	2277063	11.6	352	345	332	305	243	135	0	1	125	207	151	0	70
E1	19015028	2322875	12.2	352	348	328	314	249	143	0	2	98	230	178	0	70
D2	30473596	3157609	10.4	351	350	335	322	266	158	0	2	105	230	186	0	80
B1	27180597	1979370	7.3	351	351	332	318	264	166	1	14	94	238	211	0	48
D1	36062475	4266348	11.8	352	349	324	310	243	135	0	14	88	236	227	0	43
B2	30783564	2756205	9	352	349	332	323	272	169	0	6	104	228	186	0	84
C1	32298600	2643413	8.2	352	349	328	314	259	156	0	17	97	231	219	0	47

6. REFERENCES

Henderson RJF (1987) Caesia. In 'Flora of Australia' Vol. 45. (Ed. AS George) pp. 281–288. (Australian Government Publishing Service: Canberra)

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rossetto, M., Bragg, J., Kilian, A., McPherson, H., van der Merwe, M., & Wilson, P. D. (2019). Restore and Renew: a genomics-era framework for species provenance delimitation. *Restoration Ecology*, *27*, 538–548.

Rossetto, M., Yap, J. Y. S., Lemmon, J., Bain, D., Bragg, J., Hogbin, P., ... & Wilson, T. C. (2021). A conservation genomics workflow to guide practical management actions. Global Ecology and Conservation, 26, e01492.

RPS (2016) Biodiversity Assessment Report Springvale Water Treatment Project.