GOVERNMENT OF WESTERN AUSTRALIA

Department of **Biodiversity, Conservation and Attractions**

**Biodiversity** and **Conservation Science**

# Component 2 Report

# Assessment of genetic diversity in sub-populations of *Marianthus aquilonaris*

**Dr Tara Hopley and Dr Margaret Byrne**

Biodiversity and Conservation Science, Department of Biodiversity, Conservation and Attractions

December 2019

Report submitted to Audalia Resources

## Executive Summary

This research is in response to a request from Botanica Consulting for a research project that provides information on the population genetic diversity, structure and connectivity of *Marianthus aquilonaris* to inform management of the population in relation to proposed mining activity. Audalia Resources is seeking more information about the connectivity of plants of *Marianthus aquilonaris* to inform environmental impact assessment and conservation management.

*Marianthus aquilonaris* is declared as Rare Flora under the Biodiversity Conservation Act 2016 and is recorded from six sub-populations in three population clusters at one location in the Bremer Range. Little is known of the genetic diversity and structure of the species, or the connectivity of subpopulations through gene flow. Proposed mining activity in the area may have impacts on two of the six sub-populations. This research aims to determine the genetic diversity and structure of the six subpopulations, the contribution of each population to the total diversity present in the species and the level of genetic connectivity among populations. This can inform management to maximise retention of genetic diversity.

The project addressed the requirement through research into the assessment of the genetic diversity present in each of the five sub-populations currently present (no individuals were found at the sixth sub-population), the spatial genetic structure present among the sub-populations, and assessment of connectivity and gene flow of the five sub-populations. Genetic diversity and structure research was accomplished by sampling 30 individuals from each of the five sub-populations and undertaking genetic assessment using a reduced representation genomic sequencing approach. Several population diversity parameters were measured for each sub-population as well as overall genetic structure and differentiation. The contribution of each sub-population to the total maximal gene diversity was also evaluated. Connectivity assessment was accomplished by undertaking paternity analysis of seed collected from ten mother plants in sub-population 1B to determine the source of the pollen contribution to the seed by identifying whether the pollen is local, from within the sub-population, or from another sub-population.

The main findings include:

- All sub-populations of *Marianthus aquilonaris* were found to have moderate levels of genetic diversity.
- The level of differentiation among the sub-populations is high given the small geographical distance between them, suggesting that there is limited genetic connectivity.
- Population differentiation analysis showed sub-population 1A to have the greatest differentiation from all other sub-populations, consistent with the greater isolation of this subpopulation.
- Analysis of contribution of each sub-population to the total gene diversity found subpopulation 1D, as well as sub-populations 1C and 1E, represent the largest proportion of the gene diversity present across the species.
- Sub-populations 1A and 1B are less representative of the gene diversity present than other sub-populations; however, they do contain more than half of the private alleles present.

- The majority of seedlings from sub-population 1B tested for paternity (96%) were fathered by plants within sub-population 1B.
- There is a high rate (49%) of self-pollination, where mothers are also the fathers of the seedling.
- Every progeny cohort is receiving pollen from multiple fathers, and paternal source plants are often spread throughout the sub-population showing pollen movement is occurring across the sub-population.
- 16% of plants were involved in fathering the outcrossed seedlings that were sampled, suggesting good contribution of plants to reproduction.
- A small number of seedlings are receiving a pollen contribution from other sub-populations, with evidence of contributions from sub-populations, 1A, 1C and 1D.

All sub-populations of *Marianthus aquilonaris* were found to represent unique genetic clusters, indicating that there has been limited historical connectivity and gene flow amongst all subpopulations. All sub-populations were found to harbour private alleles, representing unique diversity present within each sub-population. While sub-populations 1A and 1B represented the highest numbers of private alleles, gene diversity present in sub-populations 1C, 1D and 1E were the most representative of total gene diversity present in the species. Results suggest that while the majority of pollination is by fathers within sub-population 1B, there is a small amount of pollen coming from other sub-populations. A large number of plants within the sub-population are contributing to the reproductive process, of which we only assessed a snapshot. Every progeny cohort assessed had pollen contribution from multiple fathers, indicating mixing of genetic material throughout the sub-population during seed production. There is limited movement of pollen between sub-populations, which is consistent with the high differentiation seen between sub-populations.

## Project Objective and Outcome

The research project shall provide information about the genetic diversity, structure, connectivity and gene flow amongst the sub-populations of *Marianthus aquilonaris*.

## Background Proposed Research

This research is in response to a request from Botanica Consulting for a research project that provides information on the population genetic diversity, structure and connectivity of *Marianthus aquilonaris* to inform management of the population in relation to proposed mining activity. *Marianthus aquilonaris* is a rare species that is currently found in six sub-populations in three populations clusters at one location in the Bremer Range. Little is known of the genetic diversity and structure of the species, or the connectivity of sub-populations through gene flow. Proposed mining activity in the area may have impacts on one of the clusters of sub-populations. Audalia Resources is seeking more information about the population genetic diversity and structure of plants of *Marianthus aquilonaris* to inform environmental impact assessment and conservation management.

## Research Plan

This research will assess genetic structure and estimate connectivity among the sub-populations of *Marianthus aquilonaris*. The genetic analysis will be undertaken with next generation genomic sequencing that provides the greatest power to identify localised genetic structure and evidence of connectivity.

Samples from all sub populations of *Marianthus aquilonaris* will be collected and analysed using DArTseq to determine genetic diversity and genetic structure among the sub populations. DNA samples from up to 30 individuals per subpopulation (all individuals from sub-populations 1e and 1f that have less than 30 individuals) will be sequenced by Diversity Arrays Technology (150 samples in total). Population genetic parameters will be obtained for the species using a range of appropriate population genetics software, using the Pawsey supercomputing facilities where required.

Connectivity will be assessed using paternity analysis of seed collected from sub-population 1B to determine the location of the pollen contribution to the seed and whether the pollen is local or from another sub-population. DNA analysis using DArTseq of seed from collections from 10 plants in populations 1B will be undertaken, as well as all plants from sub-population 1B. This sub-population is smaller than the others and thus all plants in the sub-populations can be genotyped giving power to identify those seed sired by plants from within the sub-population, and those sired from plants in other sub-populations.

# Research Methodology

**Sample Collection**

Leaf samples from 30 individuals at each of the five sub-populations (1A, 1B, 1C, 1D, 1E) were received. No plants were found at sub-populations 1F. The spatial relationships among populations is represented in Figure 1. Seed and leaf were also received for 10 plants (mother plants) and 350 leaf samples representing all individuals present in sub-population 1B. Spatial relationships among individuals in sub-population 1B are represented in Figure 2.

*Figure 1 Spatial relationship among sub-populations of* Marianthus aquilonaris *in the Bremer Range.*

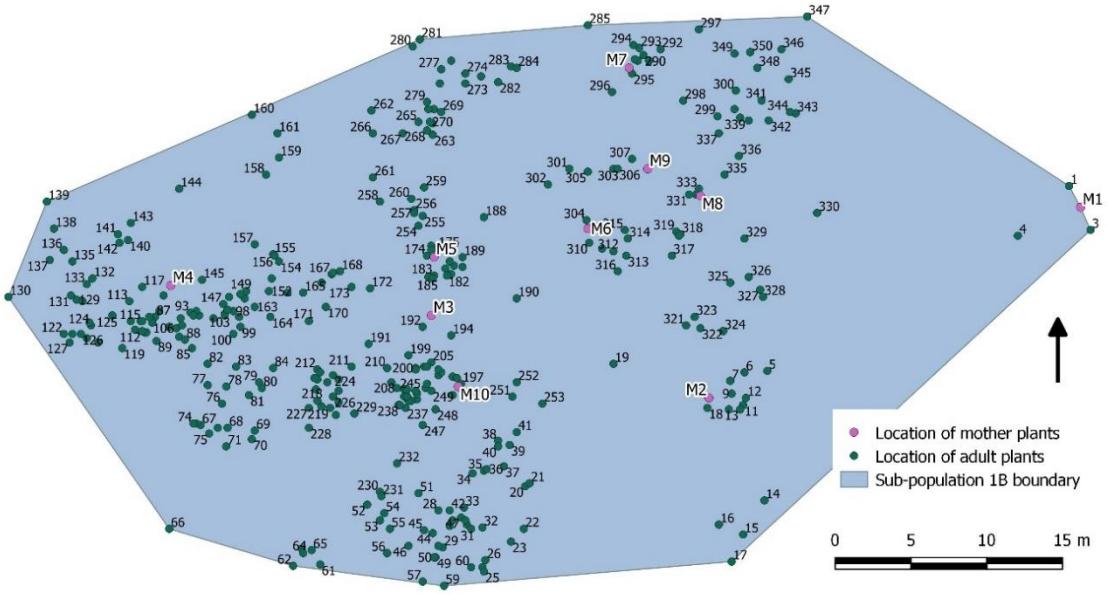


*Figure 2 Spatial relationship among individuals of sub-population 1B of* Marianthus aquilonaris *in the Bremer Range.*

**Laboratory Analysis**

The seed collections of *Marianthus aquilonaris* were cleaned then counted, listed in Table 1 below is the number of seeds received for each mother plant. For germination, 45 seed from each mother had the seed coat nicked with a scalpel blade. Seeds were then soaked in a 10% solution of PPM (Plant Preservative Material supplier, (Plant Cell Technology)) for 15 min before being placed onto agar containing 100 mg/L Gibberellic Acid (GA3). Gibberellic Acid (filter sterilised) was added to autoclaved water agar that had cooled to a temperature of 60°C. Plates were incubated at 15°C with light/dark cycles of 12 hours.

*Table 1: The number of seed received for each mother plant.*

| **Mother** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of seeds received** | 45 | 110 | 100 | 64 | 30 | 90 | 48 | 90 | 105 | 102 |
| **Number of seedlings successfully grown** | 16 | 12 | 17 | 19 | 7 | 23 | 12 | 24 | 21 | 29 |

Leaf material from adult plants and from the seedlings was freeze-dried before genomic DNA was extracted using a modified CTAB method (Doyle & Doyle 1987), with the addition of 1% w/v PVP (polyvinylpyrrolodine) to the extraction buffer. DNA of samples was checked for quality and the amount of DNA quantified before DNA concentrations were standardised. DNA samples were then sent to Diversity Arrays Technology (DArT) (Canberra) for DArTSeq analysis.

**Data Analysis**

The results received from the DArTSeq analysis were filtered using the R packages dartR (Gruber & Georges 2019), poppr (Kamvar et al. 2014, 2015) and SNPRelate (Zheng et al. 2012) in R (R Core Team 2016). The data were filtered to a loci call rate of 95%, an individual call rate of 95%, a reproducibility score of 1, a hardy-weinberg equilibrium with a 5% level of significance, a minor allele frequency greater than 2%, a linkage disequilibrium threshold of 20%, removal of monomorphic loci, and finally filtered on hamming distance to remove potential paralogues

As the collection of all adult leaf samples at sub-population 1B was undertaken at a separate time to the collection of seed, it included recollection of leaf material form the 10 mother plants. Therefore, the first step was to identify which of the leaf collections was a repeat collection of the mother leaf samples. This was undertaken by creating a distance matrix from the final snp dataset using the dist function in the stats package in R (R Core Team 2016). The repeat samples were removed from the analysis.

*Outlier removal*

Outliers were removed from the filtered SNP dataset as most genetic structure programs used assume neutrality within data. BayPass (Gautier 2015) was used to identify outliers within the SNP datasets. This was done using the XtX differentiation measure, which is analogous to the SNP FST corrected for covariance of population allele frequencies. Initially the core model was run four times with default settings, with a nval of 100,000, burnin of 50,000, npilot of 30, and pilotlength of 5000, results were averaged over runs. Calibration of the XtX statistic was undertaken using the function simulate.baypass() to create a pseudo-observed dataset, and subsequently run using the same settings on the core model to calculate 1% and 99% thresholds to discriminate between neutral and outlier loci. Those SNPs having XtX statistics above the 99% and below the 1% threshold, representing directional and balancing selection respectively, were removed to create a neutral dataset.

*Neutral population structure*

To identify clusters of individuals and visualise the major axes of variation between clusters, principle coordinates analysis (PCO) was undertaken, implemented in the adegenet package (Jombart & Ahmed 2011) in R (R Core Team 2016). Expected and observed heterozygosity, private alleles, inbreeding coefficients and pairwise population differentiation (FST) were assessed using the adegenet (Jombart & Ahmed 2011), hierfstat (Goudet & Jombart 2015) and Poppr (Kamvar et al. 2014) packages in R.
Population genetic structure was explored using Structure 2.3.4 (Pritchard et al. 2000) using the neutral data set obtained after filtering and outlier removal. Analysis using K-values from 2 to 7 were undertaken, with ten independent runs for each K-value with a burnin of 50,000 and 250,000 MCMC iterations. The R package pophelper (Francis 2017) was used to visual results and select the most probable K based on the ΔK metric (Evanno et al. 2005).

Analyses were performed in Metapop 1.0.3 (Pérez-Figueroa et al. 2009) to determine the relative contributions of populations toward overall genetic diversity and to allow an assessment of the impact of their removal.

Adult leaf material representing all individuals in sub-population 1B was used to confirm the relationship between the full samples collection of all individuals in the sub-population with that of analysis using a subsample of 30 plants. Principal coordinates analysis (PCO), expected and observed heterozygosity, private alleles, inbreeding coefficients and pairwise population differentiation (FST) were assessed as described above.

*Paternity analysis*

To ensure a dataset that was informative for paternity analysis, a stringent filtering to loci with a minor allele frequency above 0.4 was undertaken and loci with mismatches between mother and progeny were removed. Paternity analyses were conducted with the SNP genotype data using CERVUS version 3.0, which uses a maximum-likelihood assignment based approach to infer parentage (Marshall *et al.* 1998; Kalinowski *et al.* 2007). CERVUS calculates the natural logarithm of the likelihood ratio (LOD score), which provides the likelihood of paternity of each candidate male relative to a random male in the population for each offspring. CERVUS uses simulations of the allele frequencies of adults in the population to calculate critical differences in LOD scores between the most likely father and all other candidate fathers to assign paternity at either 80% or 95% confidence. Paternity was simulated for 100,000 offspring to determine the critical LOD scores for the assignment of paternity. CERVUS assignments of the most likely fathers were made using Delta scores, Delta is defined as the difference in LOD scores between the most likely candidate parent and the second most likely candidate parent. The advantage of using Delta over LOD is that it guards against potentially incorrect assignment of parentage when two or more candidate parents have similar large positive LOD scores.

For any seedling that was not assigned paternity with at least 95% confidence in CERVUS, a population assignment method was used to predict the most likely sub-population to have produced that seedling. This was implemented in the R package assignPOP v1.1.7 (Chen et al. 2018). The assignPOP process performs population assignment using a machine-learning framework; it employs supervised machine-learning methods to evaluate the discriminatory power the data. It then uses a cross-validation procedure followed by PCA to evaluate assignment accuracy and membership probabilities. First, the data set is partitioned into training (baseline) and test (holdout) data sets using a resampling cross-validation procedure, with the user specifying the number or proportion of individuals from each source to be used in the training data set. Next, the features of the training data sets are reduced in dimensionality using PCA, the output of which is used to build predictive models from user-chosen classification machine-learning functions. Finally, these models are used to estimate membership probabilities of test individuals and assign them to a source population, while also evaluating the baseline data and conducting assignment tests on individuals for which the origin is unknown.

**Genetic diversity and differentiation - 30 individuals per sub-population**

The results from the samples that included 30 individuals from each of the five sub-populations, had a single sample fail sequencing and as such the DArTSeq results contained 149 samples and 9503 loci. After filtering, as outlined above, the filtered SNP dataset set contained 4065 loci and 146 individuals. Outlier detection analysis found 24 loci under directional selection and 24 loci under balancing selection. These outliers were removed from further data analysis of population differentiation and structure as most programs used assume neutrality within data. This resulted in a final dataset of 4017 loci.

Analysis of the samples of *Marianthus aquilonaris* found moderate levels of nuclear genetic diversity across all sub-populations (Table 2, Figure 3). The observed heterozygosity values ranged from 0.239 to 0.321, with sub-population 1D having the highest value and sub-population 1A having the lowest. Sub-population 1D also had the highest mean allelic richness at 1.92, followed by sub-population 1C at 1.87 while the lowest was found at 1B with 1.72. However, sub-population 1A was found to have the highest number of private alleles with 37 alleles unique to the sub-population, followed by sub-population 1C which has 23 private alleles. All sub-populations were found to have negative inbreeding coefficients suggesting that mating is not occurring between related or genetically similar individuals.

*Table 2: Genetic diversity characteristics of the five sub-populations of* Marianthus aquilonaris.

| Sub-population | Number of individuals | Mean allelic richness | Private alleles | Expected heterozygosity | Observed heterozygosity | Inbreeding coefficient | Population size estimate* |
|---|---|---|---|---|---|---|---|
| **1A** | 30 | 1.74 | 37 | 0.227 (0.003) | 0.239 (0.003) | -0.051 | 260/2259 |
| **1B** | 28 | 1.72 | 17 | 0.220 (0.003) | 0.246 (0.004) | -0.121 | 138/247 |
| **1C** | 29 | 1.87 | 23 | 0.279 (0.003) | 0.297 (0.003) | -0.065 | 1142/3205 |
| **1D** | 30 | 1.92 | 12 | 0.300 (0.003) | 0.321 (0.003) | -0.071 | 2090/NA |
| **1E** | 29 | 1.85 | 11 | 0.273 (0.003) | 0.281 (0.003) | -0.030 | 1029/NA |

*Population size estimates taken from counts by Botanica Consulting in 2013-2014 and DBCA in 2015.

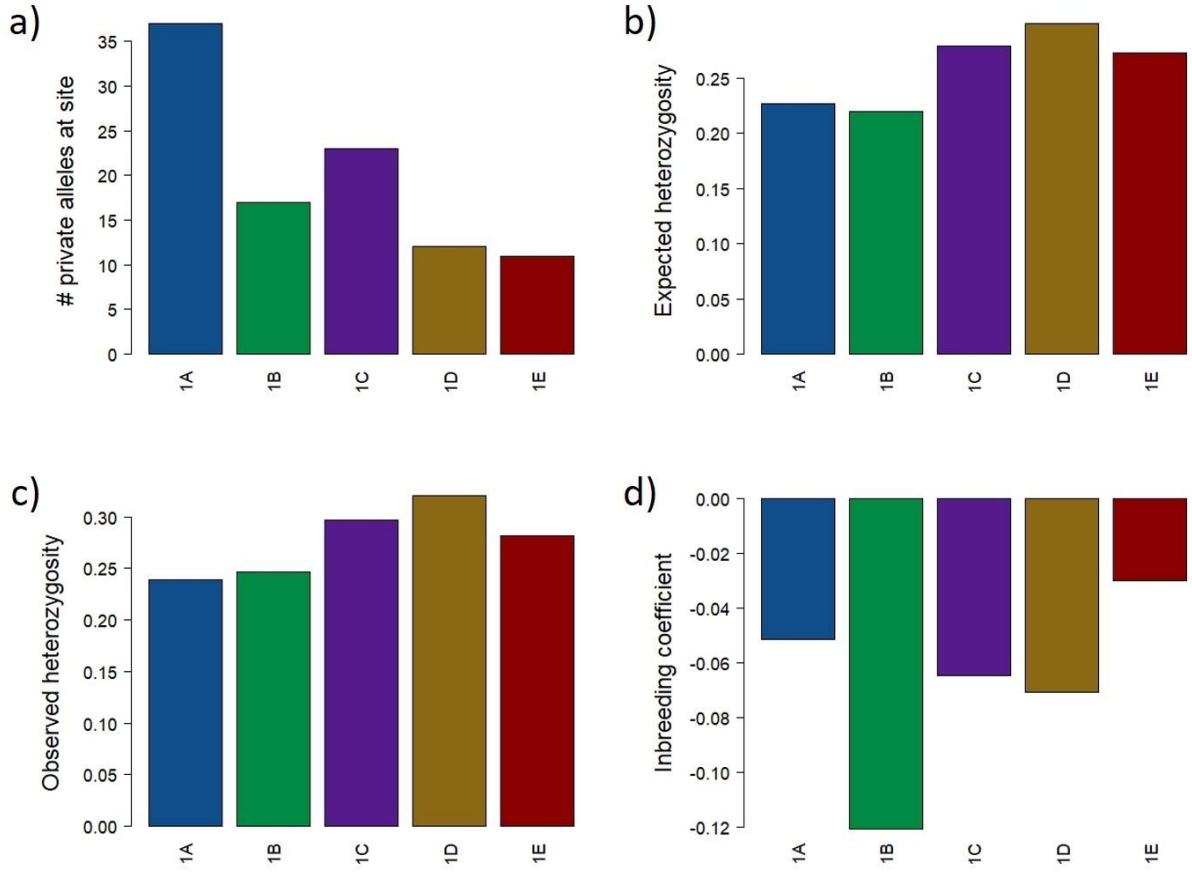


*Figure 3 Visual representation for comparison of genetic diversity characteristics, a) number of private alleles, b) expected heterozygosity, c) observed heterozygosity and d) inbreeding coefficient for each of the five sub-populations.*

Measures of genetic differentiation ($F_{ST}$) found a range of values from a low of 0.042 to a high of 0.235 (Table 3). The highest differentiation was between sub-populations 1A and 1B and the lowest differentiation was found between sub-populations 1D and 1E. Sub-populations 1A and 1B showed moderate differentiation from all sub-populations. Supporting the results of the principal coordinate analysis sub-populations 1C, 1D and 1E were found to have low levels of genetic differentiation. Principal components analysis also highlighted the highest differentiation amongst sub-populations to be between sub-populations 1A and 1B, separated along the first axis and the closer grouping of sub-populations 1C, 1D and 1E (Figure 4). With the second axis differentiating sub-population 1A from the grouping of sub-populations 1C, 1D and 1E. The principal components analysis also shows overlap between individuals in sub-population 1D and 1E.

*Table 3 Pairwise FST comparison amongst sub-populations of* Marianthus aquilonaris.

| Sub-population | 1A | 1B | 1C | 1D | 1E |
|---|---|---|---|---|---|
| **1A** | - | | | | |
| **1B** | 0.235 | - | | | |
| **1C** | 0.164 | 0.156 | - | | |
| **1D** | 0.141 | 0.159 | 0.067 | - | |
| **1E** | 0.181 | 0.197 | 0.106 | 0.042 | - |

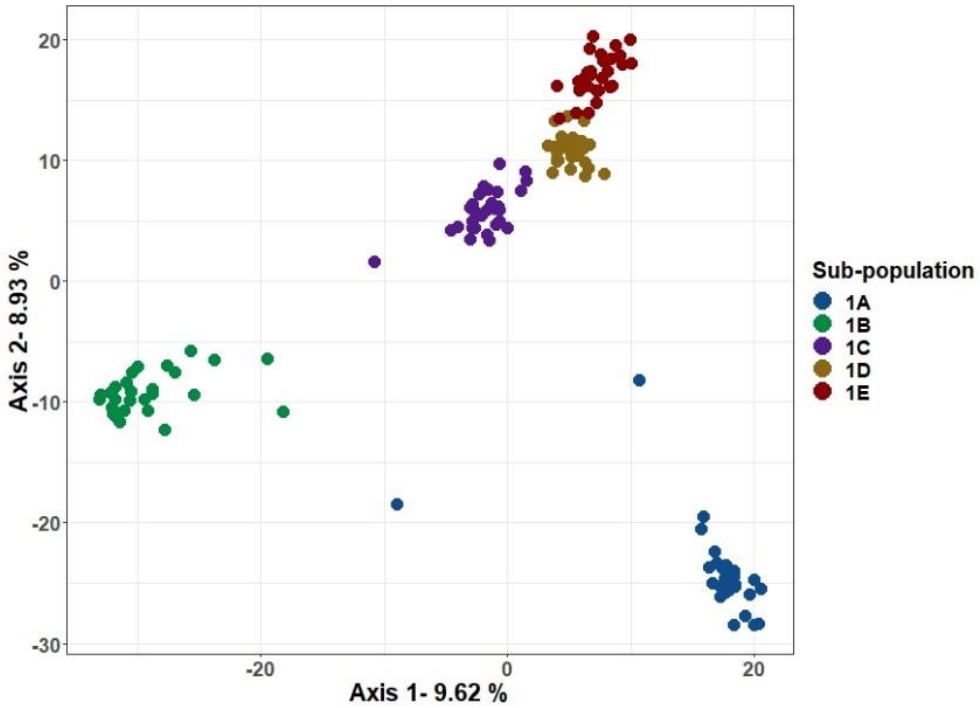


*Figure 4 Principal coordinates analysis of genetic differentiation based on 4017 single nucleotide polymorphisms. The first two axis shown represent 18.55% of the total genetic variation.*

Population genetic structure, assessed in Structure, identified five genetic clusters present, generally representing each of the five sub-populations (Figure 5). Sub-population 1D was found to be represented by a mixture of two genetic clusters, its own unique cluster and that genetic cluster represented by sub-population 1E. This shows that these two sub-populations have some connectivity, supporting the results found in the measures of genetic differentiation discussed above. The Structure results also highlight that several individuals in all populations are represented to some degree by their home sub-population genetic cluster and that of a different sub-population, generally neighbouring sub-populations. This indicates that some mating is occurring between sub-populations facilitating gene flow amongst sub-populations.

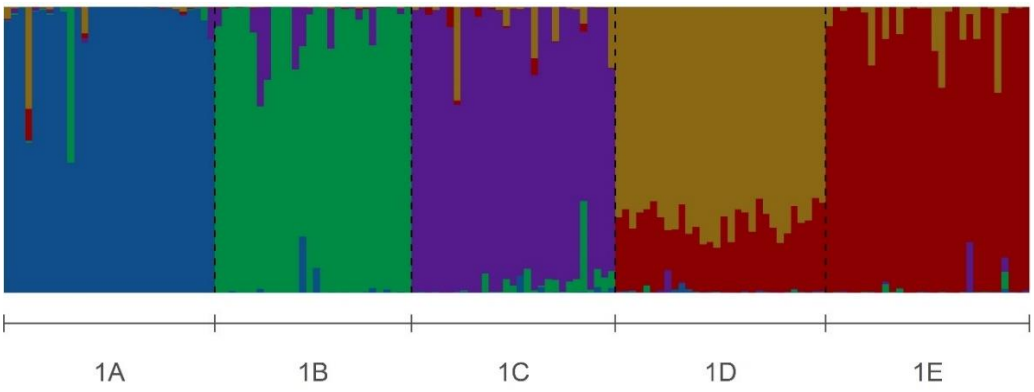


*Figure 5 Structure results showing the five identified genetic clusters present, each individual is represented by a vertical bar which is apportioned into its kinship to each of the identified genetic clusters.*

*Sub-population contributions*

Analysis of the contribution of each sub-population to the total maximal gene diversity found sub-population 1D to harbor a large proportion of the total gene diversity present across all the sub-populations, followed by sub-population 1C (Figure 6a). The impacts on total genetic diversity caused by removing each sub-population showed variable but small outcomes (Figure 6b). The gene diversity is slightly increased if sub-populations 1A and 1B are removed, this is likely a reflection of the lower heterozygosity found at these sites. Gene diversity is decreased the most when sub-population 1D is removed, with similar impacts when removing sub-populations 1C and 1E.

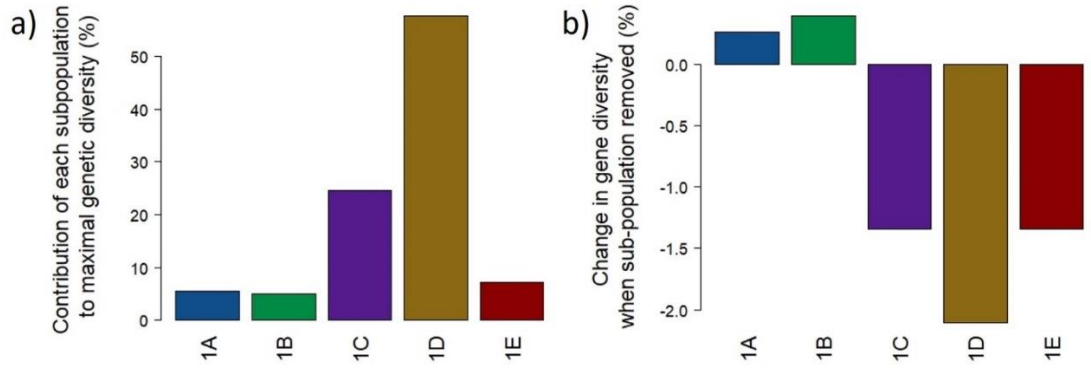


*Figure 6 Influence of sub-populations to a) proportional contribution of each subpopulation to a pool with maximal genetic diversity (%) and b) impacts of removing each sub-population on total gene diversity.*

**Genetic diversity and differentiation including all adult samples from sub-population 1B**

The genetic diversity and differentiation estimates from sub-population 1B with all individuals sampled were consistent with previous results, confirming that the sub-sampling provided a reliable sample of the genetic diversity in the sub-population. However, there was a slight reduction in diversity estimates for the second dataset between the smaller original sample of 30 individuals and that estimated from the whole sub-population sampling (Table 4). This may be due to more samples likely being related as all individuals were sampled whereas original sampling of a smaller number of plants would have been spread out across the sub-population in order to avoid sampling of related individuals. This is also likely reflected in the inbreeding coefficient, which shows a positive value for the whole population sampling, while all values are negative for those with only 30 representative samples from a sub-population.

*Table 4: Genetic diversity characteristics of the five sub-populations with 30 samples (1A-1E) and sub-population labelled M1BA which has all individuals of sub-population 1B represented.*

| Sub-population | Number of individuals | Mean allelic richness | Private alleles | Expected heterozygosity | Observed heterozygosity | Inbreeding coefficient |
|---|---|---|---|---|---|---|
| 1A | 30 | 1.768 | 0 | 0.261 (0.003) | 0.276 (0.004) | -0.056 |
| 1B | 30 | 1.799 | 0 | 0.251 (0.003) | 0.282 (0.004) | -0.122 |
| 1C | 30 | 1.888 | 0 | 0.314 (0.003) | 0.341 (0.003) | -0.088 |
| 1D | 30 | 1.904 | 0 | 0.326 (0.003) | 0.350 (0.003) | -0.074 |
| 1E | 30 | 1.849 | 0 | 0.301 (0.003) | 0.309 (0.003) | -0.026 |
| M1BA | 344 | 1.786 | 1 | 0.244 (0.003) | 0.223 (0.003) | 0.086 |

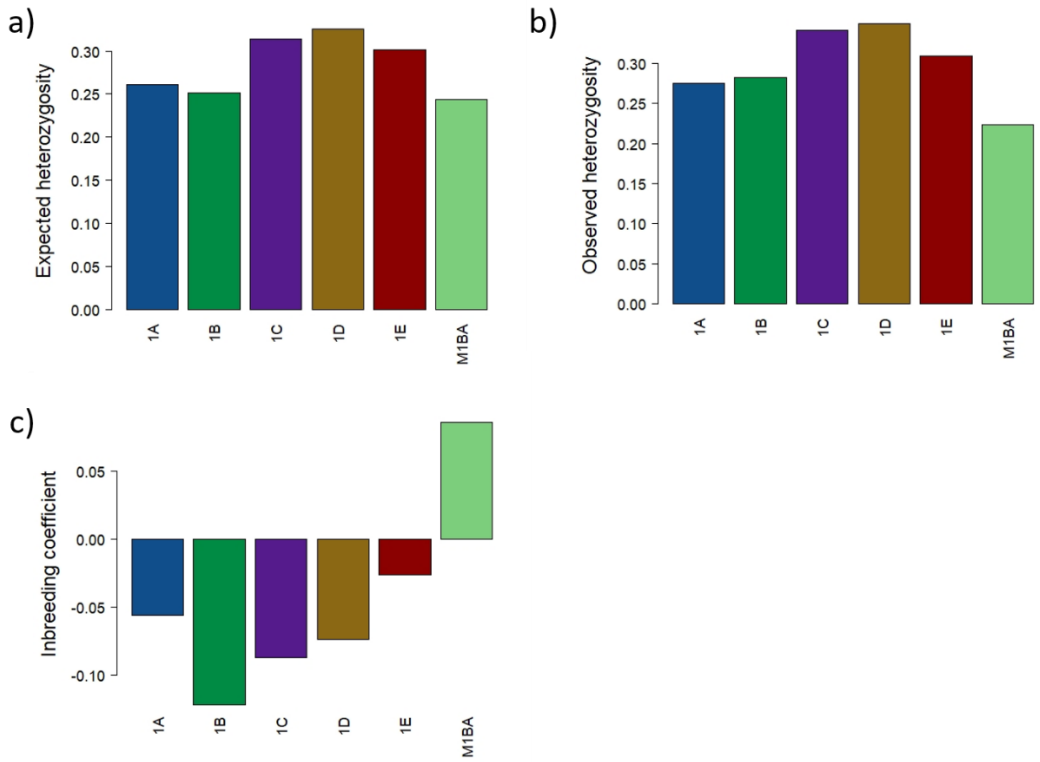


*Figure 7 Visual representation for comparison of genetic diversity characteristics, a) expected heterozygosity, b) observed heterozygosity and c) inbreeding coefficient for each of the five sub-populations and the whole populations sampling of sub-population 1B (M1BA).*

Measures of genetic differentiation among sub-populations ($F_{ST}$) reflected earlier work with estimates of differentiation from sub-population 1B with all individuals sampled slightly higher with the sub-sample (Table 5). The two sub-population 1B samples showed little differentiation (-0.002) as would be expected if the sub-sample was an accurate representation of the whole population. Principal components analysis showed the reduced sample of 30 individuals were clustered with the 350 individual samples from across the whole population (Figure 8).

*Table 5: Pairwise FST comparison amongst sub-populations of* Marianthus aquilonaris.

| **Sub-population** | **1A** | 1B | 1C | 1D | 1E | M1BA |
|---|---|---|---|---|---|---|
| 1A | - | | | | | |
| 1B | 0.234 | - | | | | |
| 1C | 0.157 | 0.156 | - | | | |
| 1D | 0.143 | 0.161 | 0.066 | - | | |
| 1E | 0.18 | 0.194 | 0.103 | 0.041 | - | |
| M1BA | 0.246 | -0.002 | 0.184 | 0.192 | 0.223 | - |

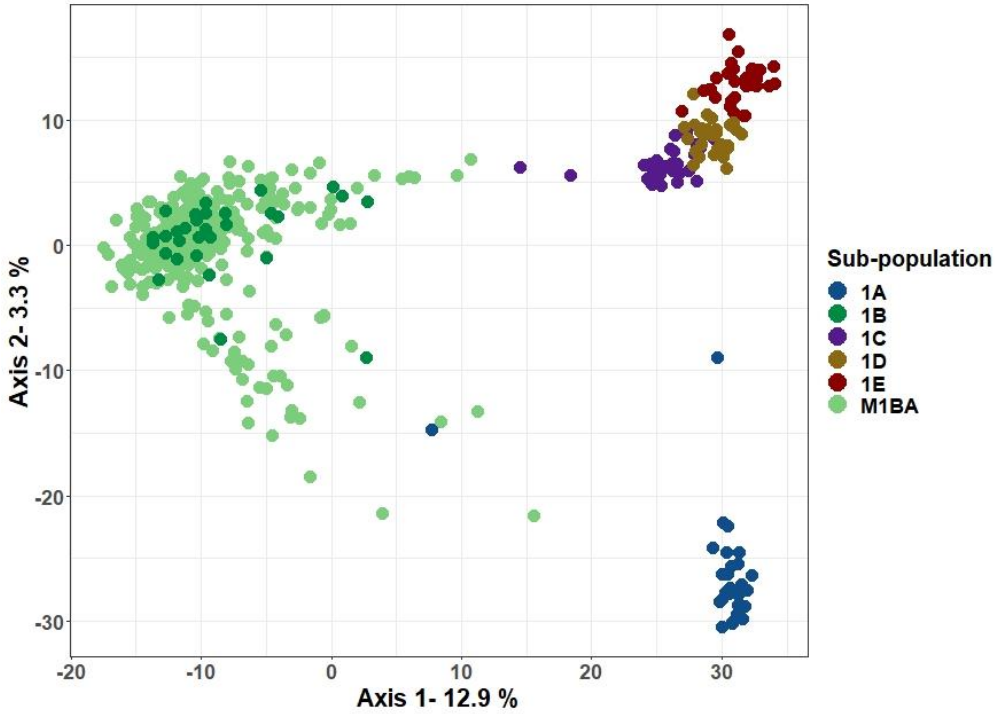


*Figure 8 Principal coordinates analysis of genetic differentiation based on 3499 single nucleotide polymorphisms. The first two axis shown represent 16.26% of the total genetic variation.*

**Paternity analysis**

Seed gemination was high and while variable among mothers, generally approached 100%. Seed germinates were planted into a pre-mix soil and seedling survival was low. Those mothers with low initial numbers of seed generally had lower germination and seedling survival. Seedlings were harvested once they reached about 5cm tall and DNA extraction was undertaken. A total of 180 seedlings were harvested and DNA extracted (The numbers of seedlings per mother are listed in Table 6).

The 350 adult DNA samples including the 10 mother DNA samples and the 180 seedling DNA samples were sequenced. Of the samples, 3 adult samples and 8 seedling samples failed sequencing, consequently DArTSeq results contained 347 adults and 172 seedlings sequenced at 9967 loci. After filtering, as outlined above, the filtered SNP dataset set contained 3548 loci across 344 adults and 165 seedlings. Outlier identification analysis found 28 loci under directional selection and 21 loci under balancing selection. These outliers were removed from further data analysis of population statistics and differentiation as programs used assume neutrality within data. This resulted in a final dataset of 3499 loci. This dataset was used to for confirming genetic diversity and differentiation and for population assignment.

The stringent filtering to create the most informative set of loci for paternity analysis resulted in 116 loci. These were used to assign paternity in CERVUS. While all seedlings were assigned paternity within sub-population 1B, not all of these were with a high confidence. Those seedlings with a trio Delta score correlating to a 95% confidence were considered as known paternity. Of the 165 seedlings genotyped, 148 were assigned paternity with 95% confidence, the numbers of seedlings for each mother with assigned paternity is shown in Table 4. The remaining 17 individuals were not able to be assigned to a specific father and may represent pollen from outside the sub-population or may represent paternity from a father with very close relatives that can't be differentiated. The proportion of seedlings from each mother assigned paternity varied from 75% in Mother 4 to 100% in three mothers. Of the 148 seedlings with known paternity, 75 (46%) were assigned as selfed seed, where the mother plant is also the father. The selfing rate (the proportion of seedlings that were selfed) varied between mothers, with no seedlings produced by selfing in Mother 6 to 87% (13) in Mother 1 (Table 4). This may be due to the Mother 1 plant being more isolated with a lower density of plants surrounding in the vicinity and on the edge of the sub-population boundary (Figure 1). However, Mother 10 had an 82% selfing rate and is in the middle of the sub-population and surrounded by other plants. There may be a combined effect of phenology (flowering timing) and density that influences the rate of outcrossing.

Of the 343 adult plants tested for paternity, 11% (39) contributed to the outcrossed (not selfed) seedlings assigned paternity. Each mother plant had multiple fathers contributing pollen to seedling cohort (Table 6). As the number of outcrossed seedlings per mother varied greatly so did the numbers of fathers contributing per seedlings cohort. Only 8 (21%) of the fathers contributed to three or more seedlings, the majority (79%) fathering only one or two seedlings. Of the 39 fathers, 8 (21%) contributed to more than one mother's seedling cohort. Figure 9-18 show the number of fathers, the location of fathers and the number of seedlings they fathered for each mother. While some mothers (M3, M5 and M7) show more localised pollen contributions, most mothers have pollen contributions from a broader area. There are a small number of mother-father pairs that produce a large number of

offspring, mother 8 and father A333 with 11 progeny (Figure 16) and mother 6 and father A295 with seven progeny (Figure 14), although this will be biased by the number of seedling in each cohort. However, most mothers had seed with multiple fathers. There didn't seem to be an impact of location of mother plants on the proportion of progeny assigned paternity, even though it might be expected that mothers on the edge of the sub-population would be more likely to receive pollen from outside the population.

*Table 6: The number of seedlings successfully grown for each mother plant, the number of seedlings per mother after sequencing and filtering of data, the number of seedlings assigned paternity with 95% confidence in CERVUS and the numbers of fathers contributing to the seedling cohort for each mother plant.*

| **Mother** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of seedlings in final filtered sequencing dataset** | 15 | 12 | 16 | 16 | 4 | 23 | 9 | 21 | 21 | 28 |
| **Number of seedlings assigned paternity with 95% confidence** | 15 | 11 | 14 | 12 | 4 | 19 | 8 | 20 | 17 | 28 |
| **Number of seedlings that were selfed** | 13 | 4 | 11 | 10 | 2 | 0 | 3 | 3 | 6 | 23 |
| **Numbers of fathers contributing to non-selfed seedlings** | 2 | 6 | 3 | 2 | 2 | 11 | 5 | 6 | 10 | 4 |

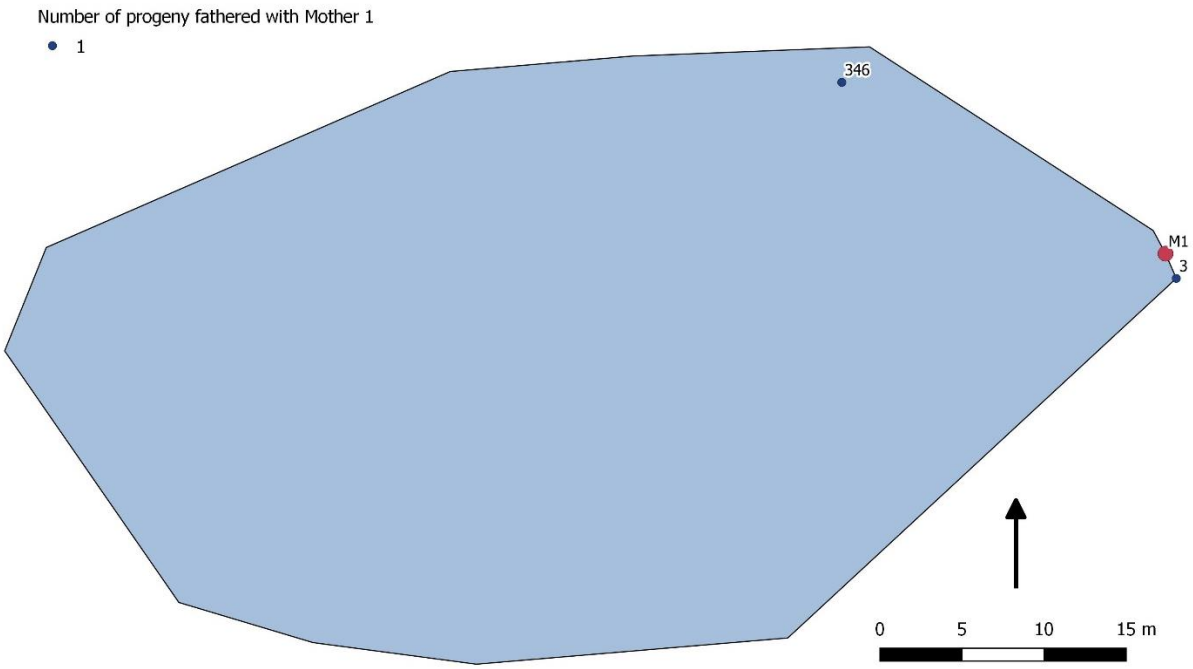


*Figure 9 Spatial relationship among Mother plant 1 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

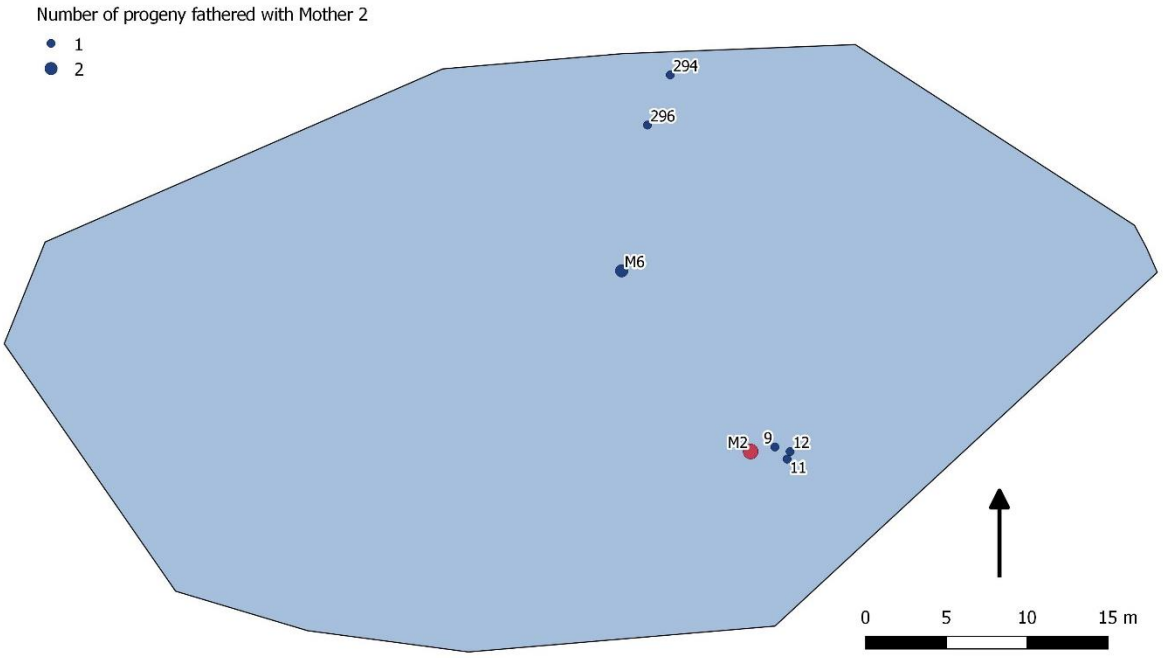


*Figure 10 Spatial relationship among Mother plant 2 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

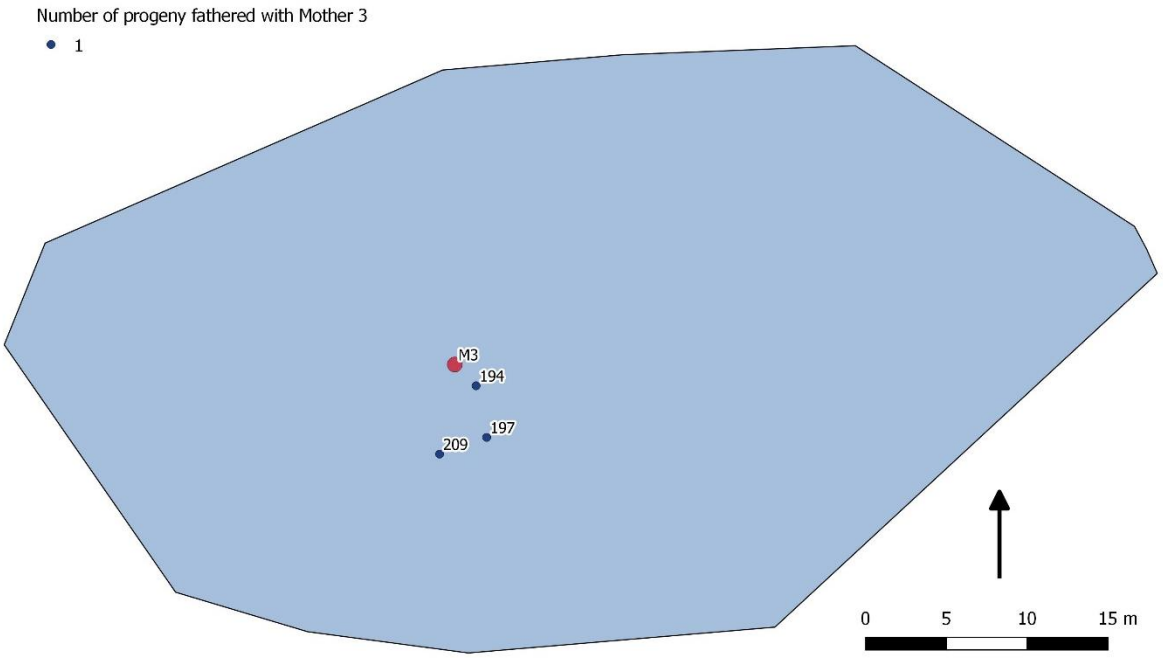


*Figure 11 Spatial relationship among Mother plant 3 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

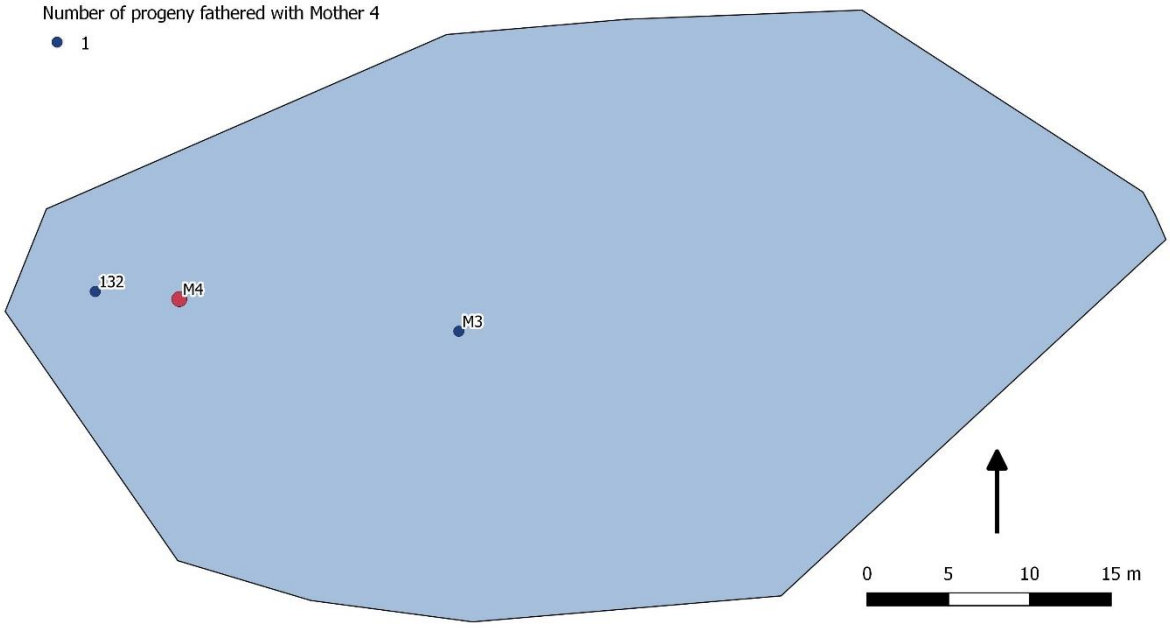


*Figure 12 Spatial relationship among Mother plant 4 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

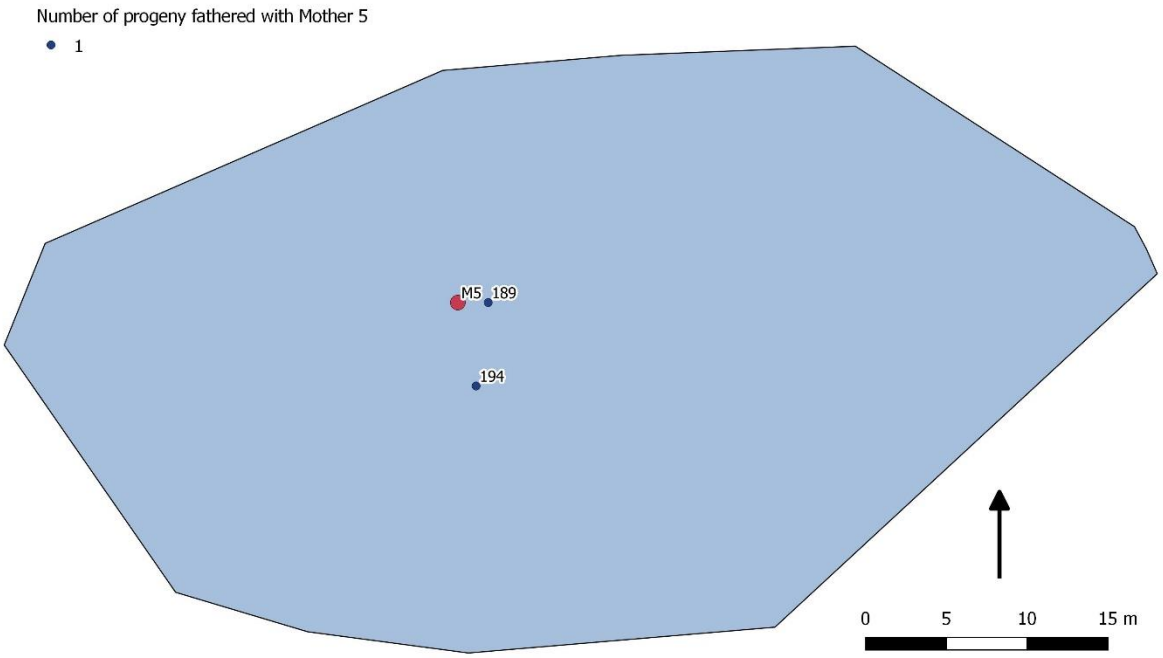


*Figure 13 Spatial relationship among Mother plant 5 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

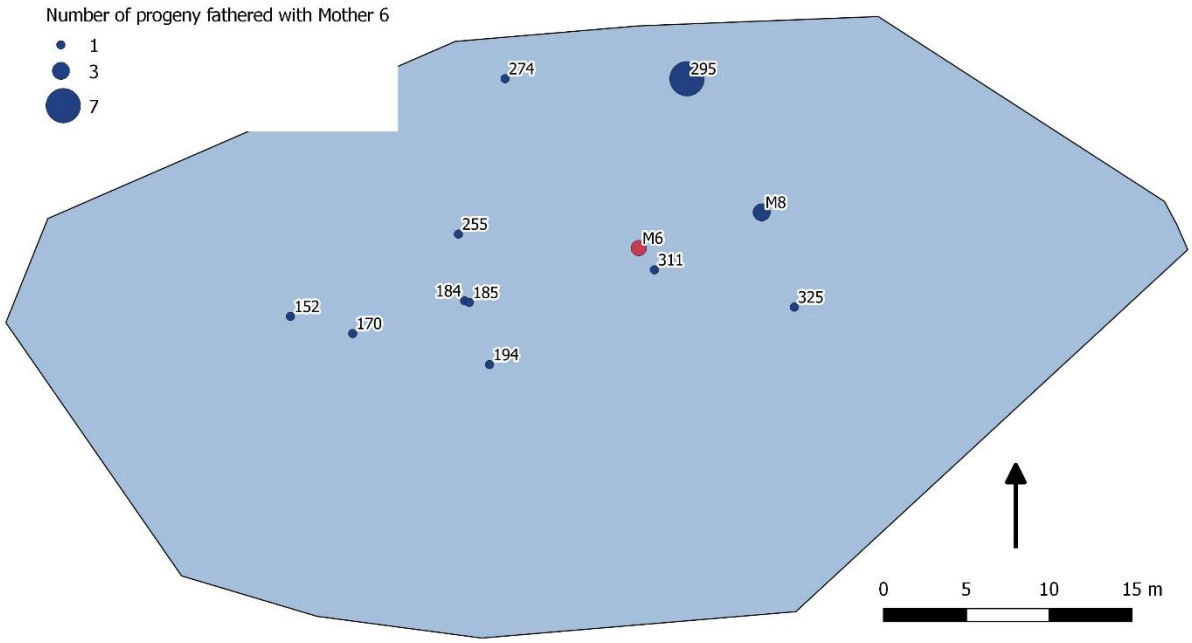


*Figure 14 Spatial relationship among Mother plant 6 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

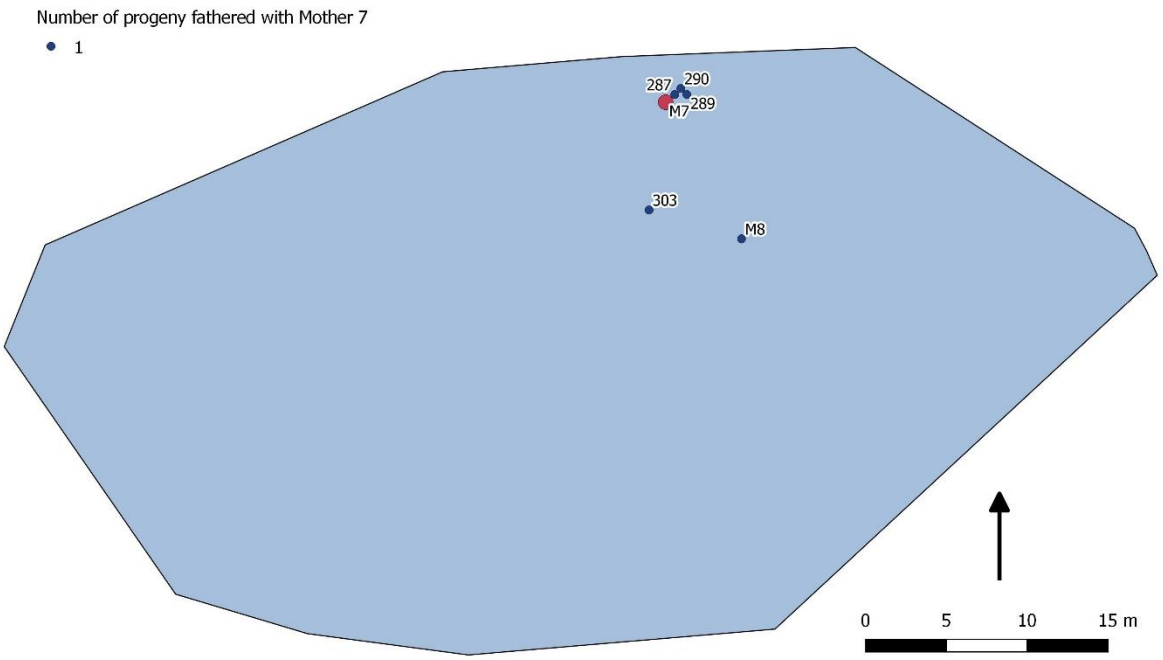


*Figure 15 Spatial relationship among Mother plant 7 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

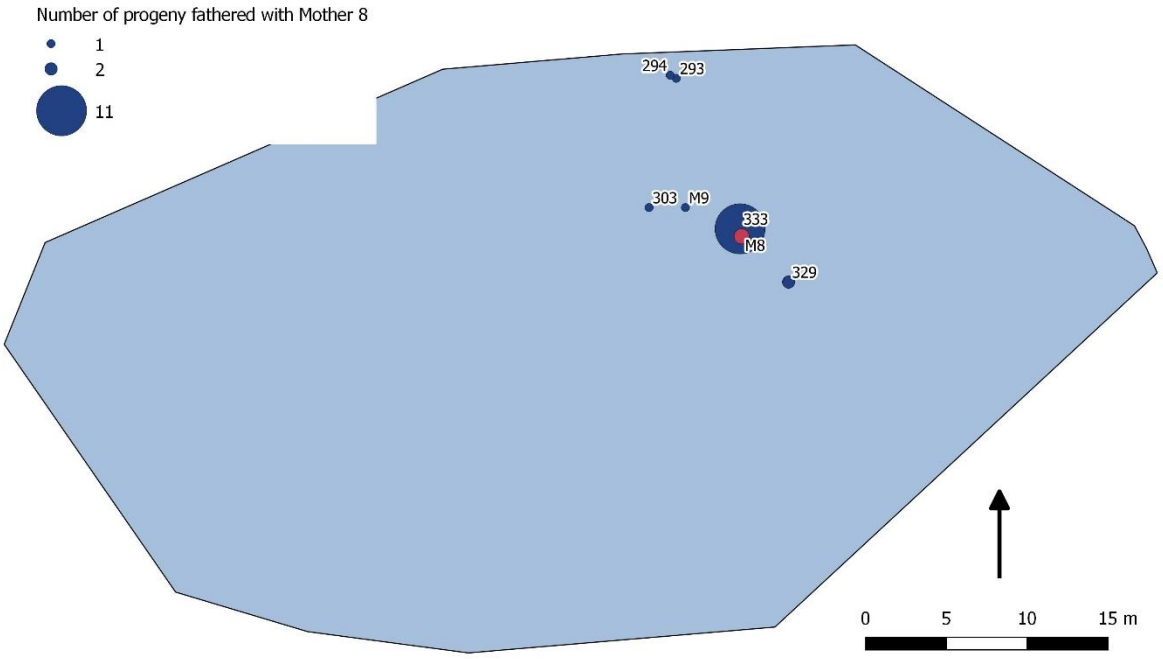


*Figure 16 Spatial relationship among Mother plant 8 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

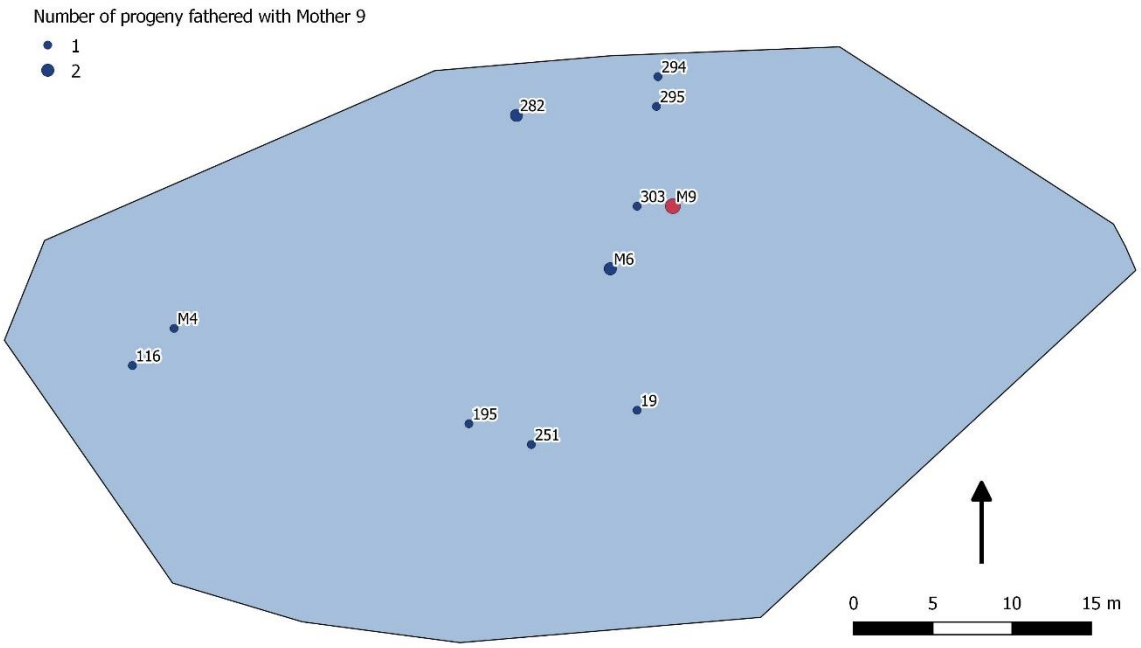


*Figure 17 Spatial relationship among Mother plant 9 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

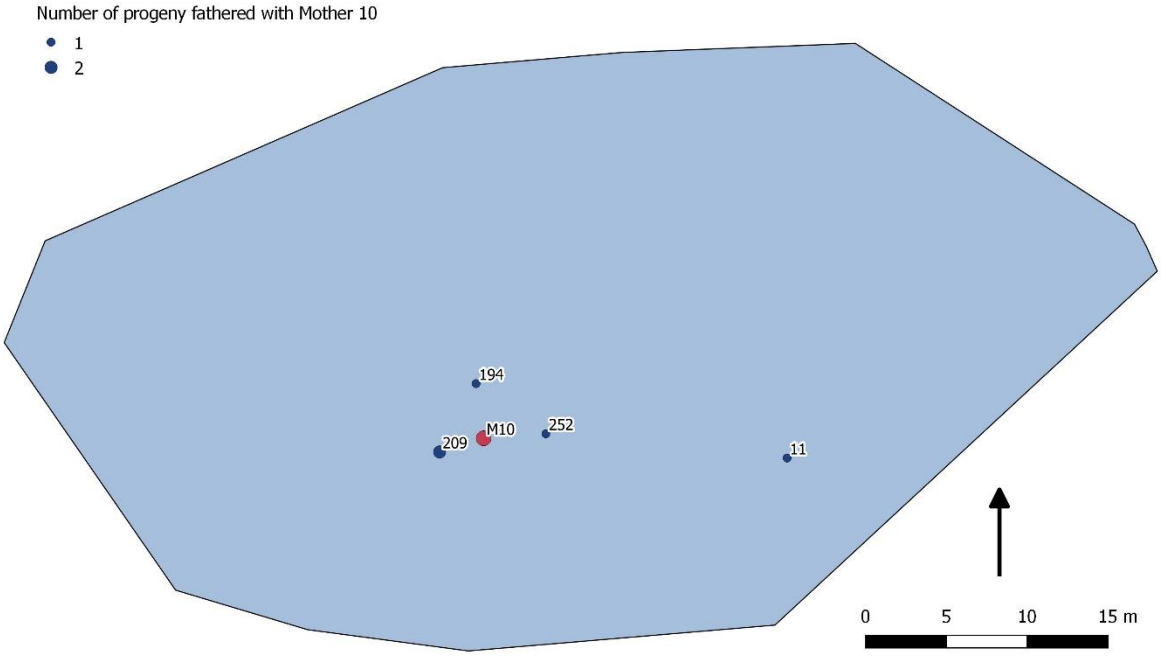


*Figure 18 Spatial relationship among Mother plant 10 and the fathers of seedlings assigned with 95% confidence in CERVUS, shaded area shows boundary of sub-population 1B of* Marianthus aquilonaris.

**Population assignment**

For those 17 seedlings that were not assigned paternity with a 95% confidence a population assignment approach was used to determine whether they are more likely to have been produced by pollen from another sub-population. The original 30 samples from each sub-population were used for the population assignment as this will not bias the results due to variation in sampling size.

To check the reliability of the assignment method we also examined the population assignment of the 148 seedlings that were assigned paternity with 95% confidence. For each of the seedlings being assessed, a probability of membership to each reference population was generated. Figure 19 shows these membership probabilities for each seedling, with a separate colour representing the proportional contribution to each of the sub-populations. All seedlings showed the majority assignment to sub-population 1B confirming the paternity assignment.

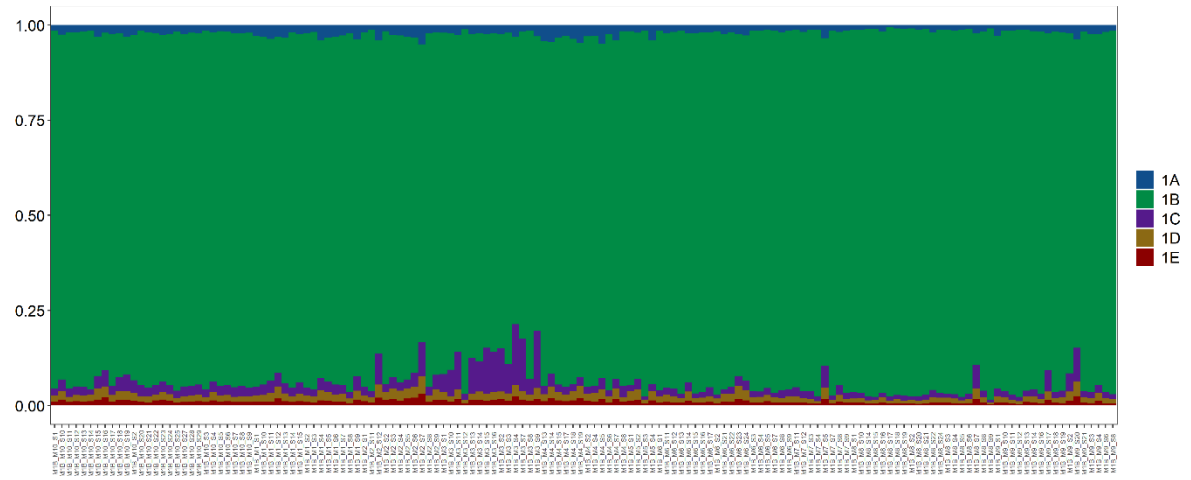


*Figure 19 Probability of membership for the 148 seedlings assigned paternity with 95% confidence, to the five sub-populations of* Marianthus aquilonaris*. Results estimated based on 3548 loci.*

For each of the 17 seedlings being assessed, a probability of membership to each reference population was generated. Figure 20 shows these membership probabilities for each seedling, with a separate colour representing the proportional contribution to each of the sub-populations. Of the 17 seedlings, 10 showed the majority assignment to sub-population 1B suggesting they have been fathered by plants within this sub-population. The other seven seedlings showed assignment to multiple sub-populations. Of these, four had majority assignment to sub-population 1C and 1B as expected from a seedling from sub-population 1B with pollen from sub-population 1C. A single seedling showed roughly 50% assignment to sub-population 1A along with some assignment to sub-populations 1C, 1D and 1B, and likely originated from pollen from sub-population 1A. The other two seedlings showed majority assignment to sub-population 1D with some assignment to 1B but also 1C, and it is likely that these three seedlings were fathered with pollen from sub population 1D. The two seedlings are from Mother 3 and seedlings from this mother also showed greater level of membership to sub-population 1C (Figure 19). This suggests that Mother 3 is a progeny of an earlier pollen migration event from sub-population 1C and this explains the mixed membership of the two seedlings fathered by pollen from sub-population 1D. Thus, the results suggest seven seedlings likely received pollen from outside sub-population 1B, from sub-populations 1C, 1D and 1A.

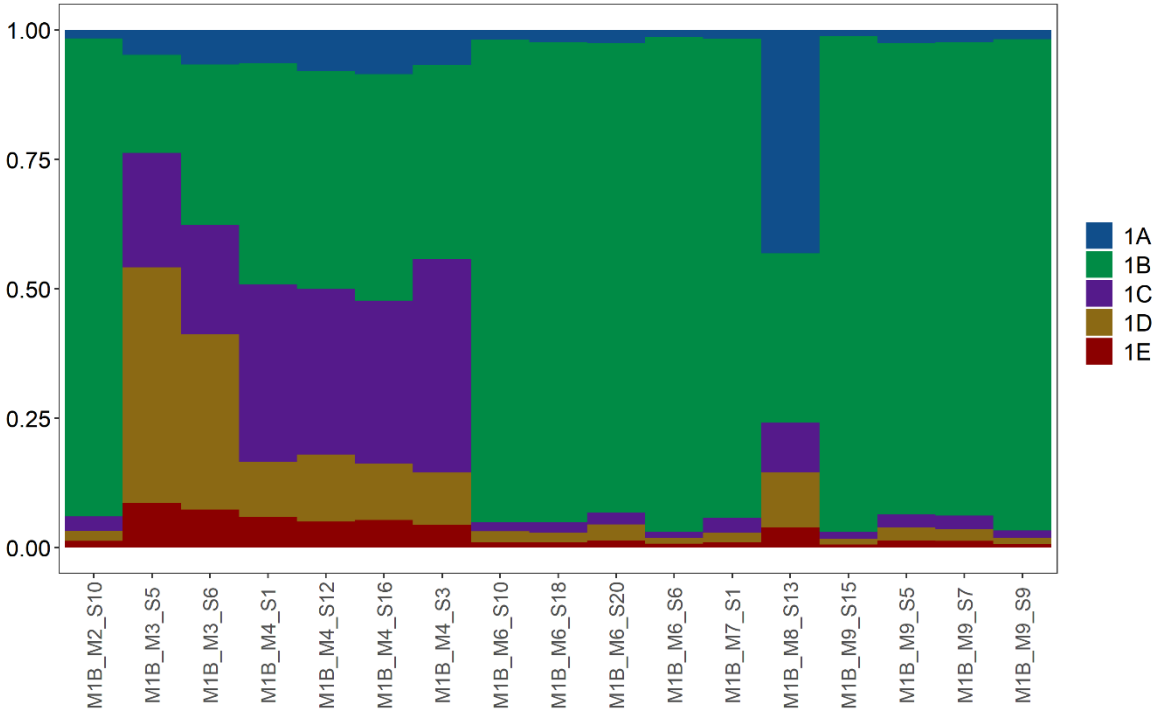


*Figure 20 Probability of membership for the 17 seedlings not assigned paternity with 95% confidence, to the five sub-populations of* Marianthus aquilonaris*. Results estimated based on 3548 loci.*

## Summary

Genetic analysis of all sub-populations of *Marianthus aquilonaris* showed moderate levels of genetic diversity. Sub-population 1D was found to have the highest heterozygosity and allelic richness levels, with sub-population 1C had the second highest heterozygosity, mean allelic richness and number of private alleles, while sub-population 1A had the lowest levels of heterozygosity. However, sub-population 1A was found to have the highest number of private alleles suggesting that this sub-population harbours the highest levels of genetic diversity that is unique from the other sub-populations. This is consistent with the greater isolation of this sub-population.

Population differentiation analysis showed sub-population 1A to have the greatest differentiation from all other sub-populations, consistent with the greater isolation of this sub-population, approximately 600 m from the nearest sub-population 1B. Sub-population 1B also showed high levels of differentiation from other sub-populations even though it is separated from sub-population 1C by only approximately 250m. Low levels of differentiation were found amongst sub-populations 1C, 1D and 1E. Sub-populations 1D and 1E appear to be genetically connected with a lower differentiation and some admixture between genetic clusters, as expected due to their closer geographic relationship. The level of differentiation among the sub-populations is high given the small geographical distance between them. This suggests that there is limited genetic connectivity among the sub-populations, except for 1D and 1E.

Analysis of contribution of each sub-population to the total gene diversity found sub-population 1D, as well as sub-populations 1C and 1E, contain the largest proportion of the gene diversity present across the species. Sub-populations 1A and 1B have less genetic diversity present, although these two sub-populations contain more than half of the private alleles present and removing these would likely result in a loss of allelic diversity.

The genetic diversity and differentiation estimates from sub-population 1B with genotypes of all individuals were consistent with the results from 30 samples from each sub-population, confirming that sub-sampling for genetic analysis was a reliable estimate of genetic relationships among sub-populations. The full sampling of sub-population 1B showed a slightly higher inbreeding coefficient and slightly lower heterozygosity estimates. This is likely due to more related individuals being included in the whole population sampling whereas the original sampling of a smaller number of plants would have been carried out across the sub-population to avoid sampling of related individuals.

Successful germination and growth of seedlings was variable between the seed cohorts, ranging from 5-29. Of the 180 seedlings sampled for analysis, 165 were successfully sequenced and passed quality and filtering checks. Of the 165 seedlings, 148 were assigned paternity to a sampled plant in sub-population 1B with 95% confidence. Of these 148 seedlings assigned paternity, 75 were assessed as arising from self-pollination where the mother is also the father. Of the outcrossed progeny, the numbers of seedlings assigned paternity per mother was variable, with every cohort receiving pollen from multiple fathers. The plants contributing pollen were spread throughout the sub-population showing pollen dispersal is occurring across the sub-population. Overall 11% of plants were involved in fathering the portion of seedlings that we sampled, suggesting good representation of plants involved with reproduction. Phenology has a strong influence on the plants involved in producing seed at any point in time. Generally, not all plants in a population will be flowering at the same time, as

such only those flowering synchronously will be captured in a seedling cohort. It is therefore likely that plants in the population not represented in the current paternity analysis are also involved in reproduction across the sub-population.

Population assignment showed that seven seedlings, or 4% of all seedlings assessed, likely received pollen from outside sub-population 1B. While this amount is small, it is consistent with the high differentiation seen among sub-populations.

Analysis of seed has shown that pollen dispersal is occurring across sub-populations 1B over distances of approximately 42m. Pollen dispersal between sub-population 1B and other populations is low with only 4% of seedlings fathered from sub-populations 1C, 1D and 1A, that range from 150-465m away from sub-population 1B. The assignment of these seedlings confirms the power of this approach to detect pollen immigration. It may be that pollen immigration between closer population is greater.

Overall, the results demonstrate high levels of self-pollination, effective pollen dispersal among plants across the sub-population, and limited pollen immigration into the sub-population from other sub-populations.

## Acknowledgements

We thank Audalia Resources for funding and support of the research project. We thank Jim Williams and staff at Botanica Consulting for information on *Marianthus aquilonaris* and for leaf and seed collections.

# References


Chen KY, Marschall EA, Sovic MG *et al.* (2018) assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. *Methods in Ecology and Evolution*, **9**, 439–446.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, **14**, 2611–20.

Francis RM (2017) pophelper: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*, **17**, 27–32.

Goudet J, Jombart T (2015) hierfstat: estimation and tests of hierarchical F-statistics. , R package version 0.04-22.

Gruber B, Georges A (2019) dartR: Importing and analysing SNP and silicodart data generated by genome-wide restriction fragment analysis.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology*, **16**, 1099–106.

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, **6**, 208.

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Pérez-Figueroa A, Saura M, Fernández J, Toro MA, Caballero A (2009) METAPOP-A software for the management and analysis of subdivided populations in conservation programs. *Conservation Genetics*, **10**, 1097–1099.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–59.

R Core Team (2016) R: A language and environment for statistical computing.

Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.