# Genetic analysis of *Banksia marginata* from the central region of Victoria


**Prepared by:**
Dr Adam Miller
Centre for Integrative Ecology
School of Life and Environmental Sciences
Deakin University
PO BOX 423
Warrnambool, 3280, VIC

**Partners:** Deakin University, Connecting Country, Harcourt Valley Landcare, Kara Kara Conservation Management Network, La Trobe University, CSIRO, Australian Network for Plant Conservation

## Summary of project findings

A genetic study of *Banksia marginata* from the central region of Victoria was performed to determine contemporary patterns of population connectivity and genetic diversity. A primary driver of this research was to identify genetically diverse seed sources for assisting regional restoration efforts (including the establishment of seed production areas (SPAs)), and inbred remnants in need of genetic rescue. Genetic analyses were performed on 16 remnant stands spanning the central region of Victoria, with results indicating significant genetic differentiation among sample locations and evidence of local populations being dependent on self-recruitment for several generations. Estimate of relatedness estimates indicate that remnant stands consist of highly related individuals, with an average of 16% of sampled individuals being direct relatives. These findings indicate limited genetic connectivity among remnant *B. marginata* stands from the central region of Victorian, suggesting minimal cross pollination between remnants, and recruitment for the past few generations being primarily from locally produced seed.

The observed patterns of local genetic structuring probably reflect contemporary demographic processes associated with population fragmentation following European settlement. Allelic diversity appears to be significantly lower in some remnants, while estimates of heterozygosity are comparable across remnants, a pattern typical of recent local population decline. This is further supported by additional analyses indicating that most remnants across the central region of Victoria have suffered from recent genetic bottlenecks. Therefore, it is likely that the observed patterns of genetic differentiation among remnants is due to recent reductions in local population sizes, and connectivity among remnants. These findings raise questions about the potential factors influencing spatial patterns of pollination / gene flow in *B. marginata*, highlighting the need for further studies into the synchrony of plant phenology among remnants, pollinator behaviour, and landscape factors influencing pollinator movement.

Findings from this study suggest that many small and isolated remnants in the region are highly vulnerable to inbreeding in the absence of managed intervention. It is uncertain if local populations are already suffering from inbreeding effects, or if the species can tolerate some level of inbreeding, as this was not explicitly assessed in this study. Research evaluating the the relative fitness of small isolated vs large contiguous populations will help provide some clarification in this regard (i.e. seed viability, flowering intensity, recruitment). In theory 20% gene flow / migration is needed to reduce the negative effects of inbreeding, however this study demonstrates that gene flow is currently inadequate (zero migration in most cases). Restoration efforts should focus on bolstering local population sizes and the connectivity among remnants to reduce inbreeding threats.

Selecting appropriate seed sources for restoration activities will be equally as important for reducing inbreeding threats and enhancing the environmental resilience of *B. marginata* remnants, through the selection of seed mixes with a broad genetic base. Based on the findings of this study and a recent study on *B. marginata* from the Victorian Volcanic Plains (which produced similar results), the following recommendations are provided to help guide future conservation and restoration planning:

1) Patterns of genetic diversity are comparable across remnants included in this study, suggesting these remnants are suitable for future seed collection activities. Other

remnants in the region are also likely to be suitable, but manager should target the larger remnants to capture the most genetic diversity and avoid collecting from small and potentially inbred remnants.

2) Smaller remnants (10s of individuals) are more at risk of inbreeding and extirpation than large remnants (100s of individuals). This should be used as a framework for prioritising restoration efforts

3) Seed production Seed production areas (SPAs) are a useful mechanism for meeting seed shortfalls and providing regionally generic seed mixes for establishing new populations or augmenting existing populations. SPAs will also reduce the pressure from struggling remnants currently targeted for seed. SPAs should include source material from multiple provenances spanning the region, to ensure seed mixes have a broad genetic base. A blend of composite (see point 4) and climate-adjusted provenancing approaches (see point 5) is recommended. Linda Broadhurst (CSIRO) and I can provide specific advice for the development of Banksia SPAs when needed.

4) When augmenting an existing remnant seed mixes should aim to include a composite of local provenance (~60%), and material from non-local sources (~40%; SPA regionally generic seed mix). This calculated approach will broaden the genetic basis and adaptability of the population, without compromising locally adapted genes/traits (if present). When establishing new populations (no local provenance available), diverse multi-provenance seed mixes should be adopted (and sourced from the SPA once established)

5) Given the imminent threat of climate change, and lack of gene flow in this species, managers should consider supplementing local seed mixes with 'climate ready' seed. This can be achieved through the careful selection of seed from climates matching expected future climates at a given location. There are several programs underway in this space and both Linda Broadhurst (CSIRO) and I can provide advice on this when required.

The recommendations provided above are based on the outcomes of the genetic study alone, but of course other factors require consideration when investing in restoration projects. However, the results presented here do provide an important resource for guiding future management decisions associated with the conservation of *B. marginata* in the central region of Victoria. This study provides new insights into the species biology and ecology and highlights the vulnerability of many *B. marginata* remnants to future inbreeding threats and environmental change, and the urgent need for intervention. This study also indicates the need for further research to characterise factors limiting gene flow on fine spatial scales in this species. For example, research on characterising the key pollinators, their behaviour, and landscape influences on pollinator movement would be extremely beneficial. Also, comparisons of plant phenology across remnants, and pollen viability would assist in identifying potential barriers to gene flow among remnants. A detailed outline of the methods used to undertake this study, and the results of the associated genetic analyses are provided below.

## Methods

### *Field collections*

A total of 16 remnant *B. marginata* stands from the central region of Victoria were selected for genetic analysis (Table 1). Leaf tissue was collected over the winter / spring period of 2017, with a maximum of 30 ~ one gram samples of fresh growth collected from individual trees from each site. At each site, individuals were sampled at random avoiding sampling adjacent individuals to reduce possible sampling of close relatives. Individual samples with unique identifiers were preserved in paper coffee filters and rapidly desiccated with silica gel ensuring minimal degradation of genetic material. Geospatial data for each of the collection sites was not provided, meaning that spatially explicit genetic analyses could not be performed.

**Table 1.** Site information, corresponding codes and sample sizes for each of the 16 collection sites. *note: geospatial information for each of the collection sites has not been provided by the project partner.

| Collection site | Code | n |
|---|---|---|
| Balborra | BAL | 4 |
| Billy-Ho Nature Walk | BHO | 6 |
| Burke and Wills track north | BWN | 7 |
| Burke and Wills track south | BWS | 12 |
| Havistock Hill | HAV | 20 |
| Koala Park | KOA | 14 |
| Long Point | LOP | 30 |
| Mt Beckworth | MBK | 17 |
| NYC | NYC | 12 |
| Orr Street | ORR | 4 |
| Pastoria East | PSE | 25 |
| Petticoat Slatery Creek lower | PSL | 17 |
| Petticoat Slatery Creek lower | PSU | 5 |
| Smeaton Cemetary | SMC | 5 |
| Tooboorac | TOO | 25 |
| Wilsons Rd, Kooroocheang | WIL | 5 |
| Whyte Rd, Kara Kara | WTE | 30 |

### *DNA extraction and genotyping*

Genomic DNA was extracted from 30 mg sample of tissue for individual specimens using the NucleoSpin® 96 Plant II protocol (Machery-Nagel Inc., Düren, KO, GER) and DNA quantitation was performed as per the QuantiFluor® dsDNA System (Promega Inc, Madison, NY, USA). *Banksia marginata* DNA samples were genotyped at 10 microsatellite loci using a composite of genetic markers developed by He *et al.* 2013 and Fatemi *et al.* 2014, and additional markers developed in the present study (Table 2). In order to distinguish PCR products upon capillary separation, primers for the ten microsatellite markers were tagged with a unique fluorescent label during PCR using the method outlined in Blacket et al.

(2012). Reactions matrices for PCR amplification consisted of 5 μl Qiagen multiplex mix (Qiagen, Chadstone, Victoria, Australia), 4 μl of primer mix (0.2 μM of each primer) and 2 μl of template DNA. PCR conditions consisted of an initial 15 min denaturing step at 94 °C, followed by 40 cycles of 94 °C for 30 s, 59 °C for 1:30 min, and 72 °C for 1:00 min, with a final extension step of 60 °C for 30 min. Genotyping was subsequently performed using an Applied Biosystems 3730 capillary analyser and product lengths were determined relative to a GS500LIZ_3730 size standard. Fragment analyses were conducted using an ABI3730 XL DNA analyser. Microsatellite profiles were examined and scored manually and assessed for polymorphisms using GeneMapper version 4.0 (Applied Biosystems).

**Table 2.** List of microsatellite loci used in the current study

| Locus | Allelic range | # of alleles | Reference |
|---|---|---|---|
| BM11 | 192 - 202 | 3 | This study |
| BM12 | 213 - 258 | 13 | " |
| BM2 | 167 - 190 | 5 | " |
| BM22 | 220 - 254 | 11 | " |
| BM27 | 123 - 156 | 13 | " |
| BM37 | 260 - 284 | 9 | " |
| BM13 | 170 - 182 | 20 | " |
| BM4 | 241 - 248 | 11 | " |
| BR3 | 155 - 175 | 21 | He *et al.* 2008 |
| Bint05VIC | 228 - 265 | 9 | Fatemi *et al*. 2013 |
| Bint07NED | 174 - 183 | 5 | Fatemi *et al*. 2014 |

### *Population genetic analyses*

Descriptive statistics were calculated for the microsatellite data using FSTAT ver. 2.9.3 (Goudet 1995) including: (1) allelic richness per population averaged over loci, (2) Weir and Cockerham's inbreeding coefficient ($F_{IS}$), a global estimate of population differentiation ($F_{ST}$) with 95% confidence limits (Weir and Cockerham 1984), (3) population pairwise measures of $F_{ST}$ with significance determined using permutation (10,000), and (4) tests for linkage disequilibrium between loci using a loglikelihood ratio test. The software MICRO-CHECKER ver. 2.2 (Van Oosterhout et al. 2004) was used to assess microsatellite loci for null alleles and scoring errors using formula 1 outlined by Brookfield (1996), as evidence of null homozygotes was not apparent. The sequential Bonferroni correction (Rice 1989) was used when performing multiple simultaneous comparisons.

Estimates of observed ($H_O$) and expected ($H_E$) heterozygosity were determined using the Excel Microsatellite Toolkit (Park 2001) and deviations from HWE were determined using GENEPOP ver. 3.4 (Raymond and Rousset 1995). Mean allelic richness and observed heterozygosity were compared among sample sites using a two-sided permutation test (10,000 permutations) also implemented in FSTAT. An analysis of molecular variance (AMOVA) was performed in GenAlEx using pairwise $F_{ST}$ as the distance measure, with 10,000 permutations and missing data for loci set at 10%. Identical multi-locus genotypes were identified using the multi-locus matching tool also implemented in GenAlEx.

A Discriminant Analysis of Principle Components (DAPC), implemented in the adegenet package for *R* (Jombart 2008; Jombart & Ahmed 2011), was used to summarize patterns of genetic differentiation between sample sites. Relationships between individuals at each site were estimated with the program ML-RELATE (Kalinowski et al. 2006). ML-RELATE calculates coefficients of relatedness (r) and putative relationships among individuals (e.g., unrelated, siblings, parent/offspring) using a maximum likelihood approach.

The software package BOTTLENECK was also run to test for evidence of recent reductions in the effective population size based on a comparison of allele numbers and gene diversity at polymorphic loci (Cornuet and Luikart 1996). BOTTLENECK tests were performed on all remnants using the infinite allele model (IAM), stepwise mutation model (SMM) and the two-phased model of mutation (TPM), with the intermediate TPM considered most suitable for microsatellite loci (Cornuet and Luikart 1996). The variance for TPM was set to 30 % and the proportion of SMM in TPM set to 70 %. Due to the relatively small number of loci, the Wilcoxon's signed-rank test was applied to determine significance (Cornuet and Luikart 1996) based on 1,000 iterations. A qualitative descriptor of the allele frequency distribution ('mode-shift' indicator), which discriminates bottlenecked populations from stable populations, was also calculated in BOTTLENECK (Luikart et al. 1998). To account for multiple comparisons, we applied the FDR procedure (Benjamini & Hochberg 1995).

Rates of recent migration were estimated among each of the sampling locations using a Bayesian algorithm implemented in BAYESASS v.3.0.3 (Wilson and Rannala 2003). The program estimates migration among populations within the last three generations. To identify movements among populations, 5 independent runs of $10^{-7}$ MCMC iterations were used following a burn-in period of $10^{-7}$ and a sampling interval of 500 steps. Chains were compared to a stationary posterior distribution for convergence by performing multiple runs with dispersed starting values. Proportion of individuals that were assigned as migrants (migration rates) and associated 95% credible intervals (CIs) were estimated among each of the sampling locations.

## Results

A total of 238 individual *B. marginata* samples from the 16 collection sites were successfully genotyped at 11 microsatellite loci (Tables 2 and 3). Marker independence was confirmed across all sample sites with linkage disequilibrium analyses indicating no significant linkage between loci, while MICRO-CHECKER analyses found no evidence of scoring errors or null alleles at any locus. A total of 120 alleles were detected, with a mean of 3.38 alleles per locus over all sites (Table 3), and observed heterozygosity was moderate to high, ranging from 0.48 to 0.67 (mean $H_E$ = 0.56). Estimates for BAL, BHO, BWN, ORR, PSU, SMC, WIL should be treated with caution due to a small sample size.

**Table 3.** Statistics for 16 *Banksia marginata* collection sites screened with 11 microsatellite loci. ***Relatedness*** = values representing the percentage of unrelated, half-sibling, full-sibling, and parent-offspring relationships (respectively) per site, with values in parentheses representing the total percentage of related individuals. ***Migration*** = the fraction of individuals in the population that are migrants derived from another population per generation (source population code provided). Mean values averaged across loci for number of alleles (***a***), expected ($\boldsymbol{H_E}$) and observed ($\boldsymbol{H_O}$) heterozygosity, Hardy–Weinberg equilibrium P-values (**HWE**), and inbreeding coefficients($F_{IS}$). Statistical significance ($\alpha = 0.05$) after correction for multiple comparisons is indicated by bold text.

| Collection site | Code | n | Loci typed | Relatedness | Migration | *a* | HE | HO | HWE | FIS |
|---|---|---|---|---|---|---|---|---|---|---|
| Balborra | BAL | 4 | 11 | 0 | 0% | 2.73 | 0.5463 | 0.5227 | 0.99 | 0.048 |
| Billy-Ho Nature Walk | BHO | 6 | 11 | 0 | 0% | 3.91 | 0.5895 | 0.6061 | 0.99 | -0.031 |
| Burke and Wills track north | BWN | 7 | 11 | 86, 5, 10, 0 (14) | 10%(BWS) | 2.91 | 0.4535 | 0.5714 | 0.95 | -0.288 |
| Burke and Wills track south | BWS | 12 | 11 | 92, 3, 3, 3 (8) | 0% | 3.55 | 0.4516 | 0.5592 | 0.90 | -0.252 |
| Havistock Hill | HAV | 20 | 11 | 74, 7, 12, 7 (26) | 0% | 3.64 | 0.4568 | 0.5591 | 0.65 | -0.231 |
| Koala Park | KOA | 14 | 11 | 67, 4, 26, 2 (33) | 0% | 2.73 | 0.4291 | 0.5639 | **P<0.01** | -0.331 |
| Long Point | LOP | 30 | 11 | 75, 11, 6, 8 (25) | 0% | 4.00 | 0.5246 | 0.5758 | 0.02 | -0.099 |
| Mt Beckworth | MBK | 17 | 11 | 82, 7, 9, 2 (18) | 0% | 4.27 | 0.6106 | 0.6738 | 0.01 | -0.107 |
| NYC | NYC | 12 | 11 | 73, 14, 6, 8 (27) | 0% | 3.45 | 0.5171 | 0.5682 | 0.72 | -0.104 |
| Orr Street | ORR | 4 | 11 | 0 | 0% | 3.18 | 0.5682 | 0.6364 | 0.99 | -0.143 |
| Pastoria East | PSE | 25 | 11 | 67, 13, 7, 13 (33) | 0% | 4.60 | 0.5500 | 0.4800 | P<0.01 | 0.14 |
| Petticoat Slatery Creek lower | PSL | 17 | 11 | 80, 14, 4, 2 (20) | 0% | 3.91 | 0.5496 | 0.5384 | 0.10 | 0.021 |
| Petticoat Slatery Creek lower | PSU | 5 | 11 | 0 | 9%(PSL) | 2.73 | 0.5253 | 0.5636 | 1.00 | -0.083 |
| Smeaton Cemetary | SMC | 5 | 11 | 90, 0, 0, 10 (10) | 0% | 2.18 | 0.3715 | 0.4273 | 1.00 | -0.173 |
| Tooboorac | TOO | 25 | 11 | 75, 13, 5, 7 (25) | 0% | 4.09 | 0.5749 | 0.5877 | 0.35 | -0.023 |
| Wilsons Rd, Kooroocheang | WIL | 5 | 11 | 80, 0, 0, 20 (20) | 5%(TOO) | 2.36 | 0.4727 | 0.5636 | 0.99 | -0.222 |
| Whyte Rd, Kara Kara | WTE | 30 | 11 | 73, 9, 6, 12 (27) | 0% | 3.27 | 0.5018 | 0.5758 | 0.72 | -0.15 |

All sites were found to conform to HWE suggesting random mating (Table 2), except for sites KOA which showed significant deviations ($P < 0.01$). This estimate was influenced by a single locus only. Significant inbreeding was not observed at any sample location, however, relatedness analyses revealed that a significant number of individuals at each collection sites are direct relatives, with an average of 16% of samples being half- or full-siblings, or parent/offspring. The lowest number of related individuals was recorded at BAL, BHO, ORR, and PSU (0%) and the highest at NYC and WTE (27%). Again, estimates for BAL, BHO, BWN, ORR, PSU, SMC, WIL should be treated with caution due to a small sample size.

A global estimate of $F_{ST}$ across all loci was significantly different from zero ($F_{ST} = 0.169$; 95% confidence interval (CI) = 0.149 - 0.186) indicating limited gene flow and genetic structuring between sampling sites. Pairwise population comparisons of $F_{ST}$ indicated weak-moderate and significant genetic differentiated between all site pairs (36 pairwise comparisons) (Table 4). Only sites with sample sizes >10 individuals were included in $F_{ST}$ calculations to avoid biases of genetic structure. AMOVA analyses also indicated a high level of genetic variation between sites (16%, $P < 0.01$). Results from the Bayesian migration analyses were highly congruent, in indicating limited migration among sites within the last three generations (Table 3). Estimates of the strength and directionality of migration indicate that each site has been largely dependent on recruitment from local sources, with minimal migration from non-local sources. Evidence of weak recent migration was recorded between 3 site pairs, in each case being unidirectional and not exceeding 10% per generation (average 1.5% per generation). Collectively, these findings indicate a significant genetic structuring and limited gene flow among sites sampled across the central region of Victorian.

BOTTLENECK analyses found evidence of recent reductions in effective population size (bottleneck events) for all *B. marginata* remnants. The Wilcoxon's sign-rank test for heterozygote excess was significant after correction for multiple comparisons under the IAM and TPM models for most remnants, while significant excess was observed under all three models for remnants KAL and TRW only. In most cases these results were further supported by evidence of mode shifts, suggesting that many *B. marginata* remnants are not in mutation-drift equilibrium and have undergone significant reductions in effective population size in the recent past.

**Table 4.** Pairwise estimates of $F_{ST}$ between 9 *Banksia marginata* collection sites (sites with >10 individuals available for genetic analysis. Values shown in bold are non-significant (P > 0.001) and remained significant after corrections for multiple comparisons.

| | BWS | HAV | KOA | LOP | MBK | NYC | PSL | TOO |
|---|---|---|---|---|---|---|---|---|
| BWS | | | | | | | | |
| HAV | 0.25 | | | | | | | |
| KOA | 0.23 | 0.24 | | | | | | |
| LOP | 0.17 | 0.22 | 0.21 | | | | | |
| MBK | 0.16 | 0.16 | 0.17 | 0.16 | | | | |
| NYC | 0.11 | 0.23 | 0.23 | 0.15 | 0.11 | | | |
| PSL | 0.14 | 0.17 | 0.09 | 0.13 | 0.09 | 0.10 | | |
| TOO | 0.12 | 0.13 | 0.17 | 0.15 | 0.10 | 0.12 | 0.08 | |
| WTE | 0.24 | 0.21 | 0.24 | 0.18 | 0.14 | 0.21 | 0.13 | 0.13 |

DAPC analyses retained 40 principal components, and the first two discriminant functions capturing 95% of the total variance within the microsatellite dataset. When plotted across the x- and y-axes individuals from the WTE, MBK, LOP, HAV, NYC and KOA sites cluster distantly from all other sites which have overlapping centroids, (Figure 4 top). These findings indicate significant genetic differentiation between a number of sites included in this analysis, however the lack of discrimination between several sites (which have been found to be genetically unique based on other independent analyses discussed above), may again be due to recent population isolation and incipient early stages of genetic divergence. DAPC analyses were repeated to include sites from the Victorian Volcanic Plains (VVP) to provide insights into regional patterns in genetic differentiation. Again, DAPC analyses retained 40 principal components, with first two discriminant functions capturing approximately 95% of the total variance. When plotted the sites from the central region of Victoria and the VVP separate across the x- and y-axes suggesting regional genetic differences driven by potentially limited historical and/or contemporary gene flow between regions.

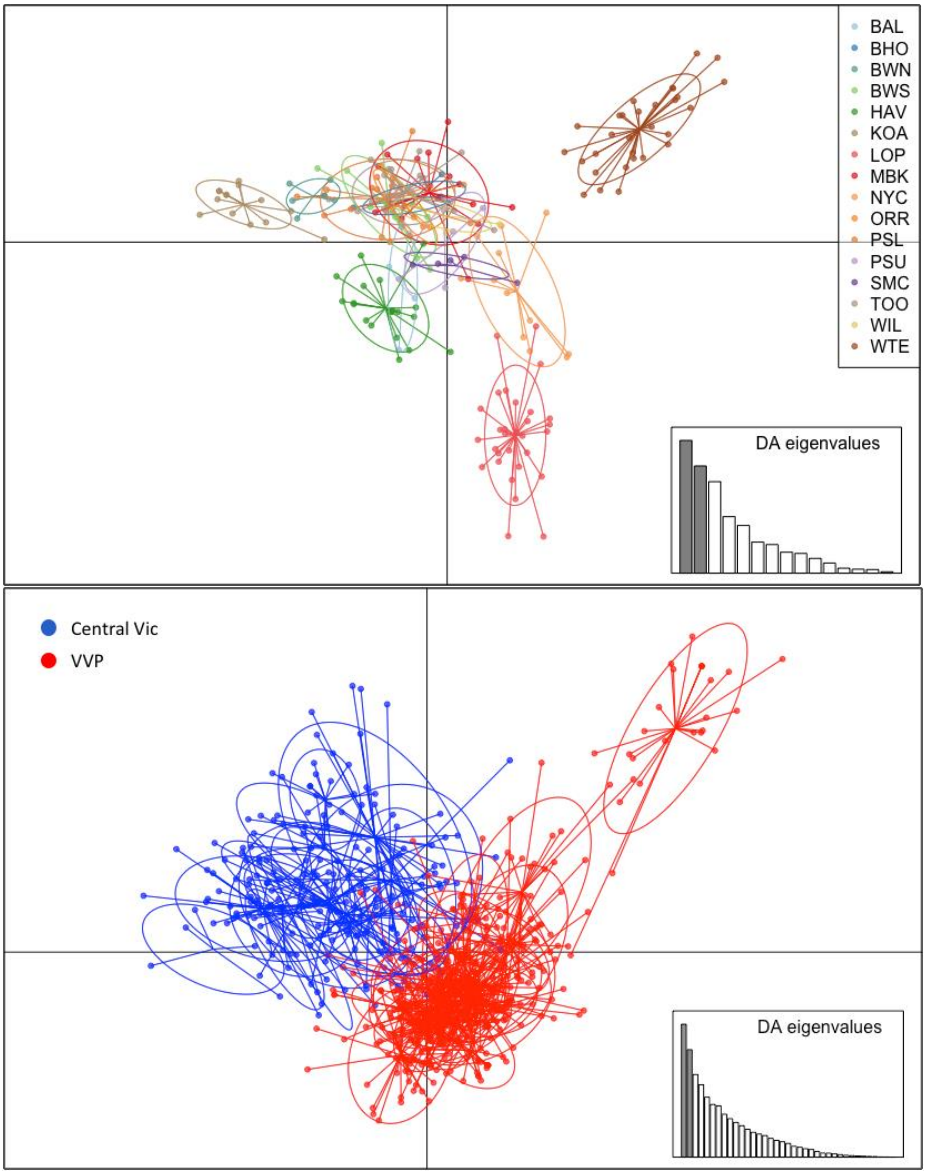


**Figure 4.** Discriminant Analysis of Principle Components; (TOP) including individuals from all 16 *B. marginata* remnants from the central region of Victoria, (BELOW) individuals from all central region (red) and Victorian Volcanic Plains(blue) remnants.

**Acknowledgements:** Many people have made significant contributions to this study. Firstly, Bonnie Humphreys (Connecting Country) initiated the project and coordinated the fund raising for the genetic study through funding largely provided by the National Landcare Network. Bonnie and Debra Saxon-Campbell (Kara Kara Conservation Management Network) coordinated the field collections along with colleagues from local landcare, community and government groups. The Harcourt Valley Landcare group hosted a public forum in Harcourt in March 2018 where results from this study were presented to the general public (funded by the Harcourt Valley Landcare group and the Norman Wettenhall Foundation). Finally, many thanks to Linda Broadhurst (CSIRO) and Susan Hobee (LaTrobe University) for their support with genetic analyses and drafting of the report content.

## References

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate − a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57: 289−300

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5: 453−455

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144: 2001−2014

Goudet J (1995) Fstat (version 1.2): A computer program to calculate F-statistics. J Hered 86: 485−486

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24: 1403–1405

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, Wagner AP, Taper ML (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Mol Ecol Notes. 6: 576-579

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89: 238−247

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. University of Dublin

Peakall R, Smouse PE (2006) Genalex 6: Genetic analysis in excel. Population genetics software for teaching and research. Mol Ecol Notes 6: 288−295

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49: 1280−1283

Rice WR (1989) Analyzing tables of statistical tests. Evolution. 43, 223–225

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358−1370

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 163: 1177-1191