Identification of clonemates and genetic lineages using next-generation sequencing (ddRADseq) guides conservation of a rare species, Bossiaea vombata (Fabaceae)

Identification of clonemates and genetic lineages using next-generation sequencing (ddRADseq) guides conservation of a rare species, *Bossiaea vombata* (Fabaceae).

Authors: Michael D. Amor[1], Joshua C. Johnson[2], Elizabeth A. James[1]

[1] Royal Botanic Gardens Victoria, Birdwood Ave, South Yarra, Victoria 3141, Australia

[2] Institute for Sustainable Industries and Liveable Cities, College of Engineering and Science, Victoria University, Melbourne, Australia

Highlights

- *Bossiaea vombata* was found to consist of five clones using ddRADseq. Three sites were monoclonal, and two clones were identified at the fourth and largest site.
- The application of genetic thresholds allowed non-identical but similar multilocus genotypes to be assigned as clonemates reducing the risk of over-estimating clonal diversity.
- ddRADseq was an efficient, cost-effective method for clonal identification in a non-model plant and applicable to phylogenetic studies of a group of leafless bossiaeas from eastern Australia using a single library.
- ddRAD loci from the same library may have limited utility for chloroplast haplotyping due to attributes inherent in library preparation and/or the absence of a reference genome from *B. vombata* or a closely related species.

ABSTRACT

Plant species capable of clonal reproduction range from rare, sterile species that have a high extinction risk to invasive plants that influence the structure of ecosystems. There is increasing evidence that clonality in combination with reduced fecundity and limited dispersal capacity

increases extinction risk. As many conservation targets are not well-characterised genetically, our objective was to determine the utility of sequencing a reduced representation of the genome to inform the conservation of a non-model plant species. We sequenced a single DNA library from a recently described, rare, clonal species; *Bossiaea vombata*. Multiple assemblies of that library were used to evaluate our ability to assess genetic variation and clonal assignment, to identify chloroplast haplotypes and to obtain phylogenetic information. Next generation sequencing (ddRADseq) provided a cost and time effective method for identifying clones and assigning clonemates despite increased levels of missing data in comparison to more traditional methods (e.g. microsatellites). We applied a threshold of genetic difference to determine whether individual samples belonged to the same or different clones. *Bossiaea vombata,* was found to comprise only five clones with all but one site being monoclonal - indicating that the genotypic diversity of the species cannot be determined from a census of stems. The ddRADseq method showed utility for phylogenetic analysis but we identified possible shortcomings in chloroplast haplotyping using loci from the same library.

Keywords: chloroplast haplotype; genet; multilocus lineage; low fecundity; next generation sequencing; population structure

1. Introduction

The field of conservation biology has focused predominantly on sexually reproducing species although conservation requirements can differ between clonal and sexual species. On paper, clonal species with limited fecundity are poor candidates for conservation but we know that clones can survive many times the generation span of their sexual conspecifics and related taxa (e.g., Lynch and Balmer, 2004; Rossetto et al., 1999; Mock et al., 2012). Population fitness and adaptability are dependent on genetic diversity which is a proxy for species health (Mable, 2019). The extended longevity of a genet increases the chance of occasional sexual recruitment and can moderate genetic decline because the loss of individual ramets does not mean the loss of a genet (Jongejans et al., 2008; de Witte and Stöcklin, 2010). However, if asexual reproduction reduces opportunities for genetic novelty via meiotic recombination, subsequent adaptative capacity may also be reduced (Hoffmann and Sgrò, 2011).

A changing environment is likely to disproportionately affect species that are constrained either genetically or geographically (Fordham et al., 2013). The resulting abundance of many species will rest on their capacity to respond either by adapting to new conditions in their current environments or avoiding them by dispersing to new areas (Hoffmann and Sgrò, 2011). Cunze et al., (2013) predict that most plant species will decline when faced with contemporary climate change as dispersal limitations restrict their capacity to shift geographic ranges. If true, species reliant on clonal growth, including *Bossiaea vombata* Ross, 2018, the focal species of this study, form a group with an elevated extinction risk because their rate of population expansion and/or ability to shift range due to limited dispersal is generally more restricted compared to their sexually reproducing counterparts. However, the success and persistence of clonal plants over historic periods of climate change suggest alternative adaptive strategies are also in play. One such strategy suggested by Dodd and Douhovnikoff (2016), is the accumulation of epigenetic changes, as an alternative to DNA mutation and recombination that may imbue clonal plants with a rapid response to environmental variation by enabling phenotypic plasticity and selection in situ.

As a pre-requisite to the management of clonal plants for conservation, their capacity to disperse, adapt and balance sexual and asexual reproduction must be addressed. Ideally, these conservation management strategies would incorporate evolutionary principles (Sgrò et al., 2011; Christmas et al., 2015), highlighting the need to expand genetic studies of clonal species. The loss of reproductive capacity is considered a threatening process for species survival and the extent of clonality within a species can be a factor in its fecundity (Eckert, 2002; Silvertown, 2008). Knowledge of the underlying patterns of genetic variation and the genetic lineages generated by historic population processes is of particular relevance when planning conservation strategies (Young et al., 1996; Hoffmann and Sgrò, 2011), particularly for species where clonality can have a negative impact on the production of new genotypes (Charpentier, 2002).

Within 50 years most Australian ecosystems are expected to experience climatic changes resulting in environmental stress for organisms that are adapted to current conditions (Dunlop et al., 2012). The extent and distribution of clonality among most of Australia's 22,000+ vascular plants is unknown; however, data is accumulating on its association with rare, relict species with reduced fecundity or dysfunctional breeding systems (for example, Coates, 1988; Lynch et al., 1998, Sydes and Peakall, 1998; Bartolome et al., 2002; Kimpton et al., 2002; Peakall et al., 2003; Gross and Caddy, 2006; Gross et al., 2011). The assumption that clonal growth enables local persistence when sexual reproduction is suppressed has been borne out by empirical studies but data are scarce for Australian species (e.g. Rossetto et al., 1997; Coates et al., 2002; James and McDougall, 2007; Millar et al., 2010).

Genetic studies of species capable of clonal reproduction become important if, for example, a demographic census of populations is likely to misrepresent the effective population size (Arnaud-Haond et al., 2007). The expanding use of genomic techniques has shifted the focus of population studies from a small number of neutral genetic markers or candidate genes to genome-wide screening of genetic variation of evolutionary, adaptive and conservation importance (Ouborg et al., 2010; McMahon et al., 2014). The use of single nucleotide polymorphisms (SNPs) is highlighting the evolutionary continuum that links lineages to populations and species (Coates et al., 2018) but clonal plants have not been a target of these

studies to date. Genomic data can improve our phylogenetic understanding of species emergence and provide an evolutionary context for conservation efforts (Byrne, 2007). McMahon et al., (2014) propose the development of 'conservation priors' based on genomic data to guide management of populations, and decision frameworks for biodiversity conservation are also being developed that utilise genomic information in an evolutionary context (Allendorf et al., 2010; Hoffmann et al., 2015).

Deciding on an appropriate method for the generation of genome-wide data for clonal species identified as conservation candidates can be influenced by several factors. The availability of existing data (e.g. reference genomes or transcriptomes or microsatellite loci) reduces the time taken to identify informative genetic regions but is limited or unavailable for most species. Microsatellites have traditionally been the marker of choice for genetic studies of clonal plants but their utility for taxa where genetic diversity is low, is determined by whether an adequate number of polymorphic loci can be identified, and cross-species amplification may not be adequate for comparative studies. When few or no data exist for a species, the most cost effective and productive approach may be to sequence a reduced representation of the genome to identify SNPs (Elshire et al., 2011; Peterson et al., 2012). Reduced representation sequencing has become increasingly accessible for those non-model species that have become conservation targets (Andrews et al., 2016). Although Pantoja et al., (2017) used SNPs to characterise genotypic variation and population structure in a facultative clonal herb, there have been few studies reporting the use of SNPs in the assessment of clonality.

Here, we evaluate the utility of genome-wide SNP data obtained via double digest restriction site-associated DNA sequencing (ddRADseq, Peterson et al., 2012) to address questions relevant to the conservation of a non-model, clonal plant. *Bossiaea vombata* was selected for analysis due to its endangered conservation status. The species has been confirmed to occur within an area of less than 5 km^2 in the Wombat State Forest, Victoria, Australia. Our primary aim was to investigate the number of clonal individuals that the species contains - thus, the extent of clonality among all known discrete patches and, therefore, the vulnerability of *B. vombata* to extinction. We

126 also aimed to assess the utility of the same dataset to address phylogenetic- and cpDNA haplotye-
127 based questions relevant to conservation.

128

129 **2. Materials and Methods**

130 *Bossiaea vombata* (Fabaceae) is one of ~78 species in the endemic Australian genus, *Bossiaea*. Most
131 species are small, woody shrubs to about 1 – 2 m. In one group, the "leafless" bossiaeas, leaves
132 are reduced to scales. Generally, this life form is considered to be an adaptation to harsh dry
133 conditions in arid areas in Australia but most leafless bossiaeas are found in higher rainfall areas
134 with annual precipitation >500 mm. Thompson (2012) recognised 12 eastern Australian leafless
135 *Bossiaea* species and assigned them to one of four subgroups. Of the 12 species, six were described
136 between 1808 and 1879 and six species were described relatively recently - between 2008 and
137 2012. Five of the post-revision species were described from *B. bracteosa* (Ross, 2008; McDougall,
138 2009 and one from *B. ensata* (Thompson, 2012, and key therein). Clonal growth has been noted in
139 ten species and can be extensive (Thompson, 2012). The recently recognised species have very
140 restricted geographic ranges which makes the whole group a candidate for conservation.

141 *Bossiaea vombata* is known from four sites comprising a total combined patch area of <500 m^2
142 spread over a maximum distance of 3 km. The species reproduces rhizomatously and flowers in
143 spring (Fig. S2, S3). The longevity of individuals is unknown but at least one site presumed to
144 contain ramets of a single plant has been known prior to 2008 (Ross, 2008). Only site 4 contained
145 more than one patch and stems were less dense in patches 4 and 5 compared to other patches
146 (Fig. S4). The breeding system is unknown. Pollen viability ranged from 15 – 90% in spring 2018
147 (E. James, unpublished data) suggesting that for at least some sites, pollen viability should be
148 sufficient for sexual reproduction in the absence of other impediments such as self-
149 incompatibility. Anecdotal reports of the presence of pods with undeveloped seed were
150 confirmed during this study (personal observations, pers. comm Donna Thomas, Ballarat Botanic
151 Gardens, Victoria) but seedlings have never been observed so the species is considered
152 functionally sterile. The listing of *B. vombata* as Endangered under Victorian State Government
153 Legislation provides some protection and is appropriate given the extremely restricted confirmed

distribution and very small combined patch size. The species is highly vulnerable to threats such as fire and damage from vehicles and road maintenance activities due to the small number (and size) of patches, their occurrence on road verges and the apparent lack of seedling recruitment.

2.1 Sampling

As *B. vombata* is leafless, 40 samples of young stem tissue were collected for genetic analysis across all known locations in the Wombat State Forest, Victoria, Australia (under DELWP permit 10007717; Table 1, Fig. 1). Five samples were collected from each of eight discrete patches located across four sites. Because of the clonal nature of the species, we refer to four "sites" to indicate general location and describe discrete groups of stems with a minimum distance between them of 7 m as "patches". We estimated the size of each patch with geolocations recorded at their centres. Samples were taken from four equidistant points around the perimeter and one central point of each patch, wrapped in labelled aluminium foil and kept on ice until they were snap frozen in liquid N_2 within 5 h and stored at -20 ˚C. All sites consisted of a single patch except Site 4 where five patches occurred within an area of ca.1000 m^2. In addition, three samples of a single *B. riparia* individual and one each of *B. bracteosa, B. grayi* and *B. walkeri* were included for comparison (cultivated at RGB Victoria; Table 1).

[Fig. 1]

[Table 1]

2.2 DNA isolation

Frozen stem material (~100 mg) was ground to a fine powder using a mortar and pestle with liquid N_2 and the addition of a small amount of sand. Genomic DNA was isolated using the manufacturer's CTAB protocol for Bioline ISOLATE II Plant DNA Kit except that DNA was **eluted in a total volume of 40 µL**. DNA quality was confirmed by using 1.5 % agarose gel electrophoresis in 1xTBE buffer for 45 min at 100 V and stained with SybrSafe (Invitrogen, Thermo Fisher Scientific Australia). DNA isolations were quantified using a Qubit v3.0 fluorometer (Invitrogen Life Technologies) and stored at -20 °C.

2.3 Library preparation and sequencing

A modified version of the Peterson et al., (2012) ddRADseq protocol was used to prepare DNA libraries. A detailed protocol for preparation and amplification is available at michaelamor.com/protocols. Our final DNA library contained 53 samples including nine technical replicate pairs. In summary, genomic DNA was digested for 18 hours with EcoRI-HF and AseI restriction enzymes (New England Biolabs), barcoded adapters were ligated to digested DNA fragments. Non-ligated adapters were removed before libraries were size-selected by magnetic bead purification using Jetseq Clean (Bioline)/PEG 8000 buffer solution at 0.5x then 0.9x the DNA solution volume.

PCR-based indexing of the individual libraries was conducted using real-time PCR (rtPCR) in a Bio-Rad CFX96 thermocycler, with amplification stopped after 10 cycles in total. Amplification adequate for sequencing, whilst minimising PCR bias, was assessed by visualising fluorescence curves and ensuring that they had not plateaued. Amplified libraries were pooled in equal concentrations based on Relative Fluorescence Unit outputs from rtPCR and concentrated/purified using Jetseq Clean beads/PEG 8000 buffer solution (1.8x DNA solution volume). The pooled library was size-selected at 350-550 bp in a Pippin Prep (Sage Science) using a 2% agarose (100-600 bp) cassette and quantified via qPCR using a Jetseq Library Quantification Hi-ROX kit (Bioline) on a Bio-Rad CFX96 thermocycler. The library was diluted to 10 nM, denatured and diluted to 20 pM for sequencing using a 600 cycle (paired-end) v3 MiSeq Reagent Kit on an Illumina MiSeq with 10% PhiX spiked into the run.

2.4 Quality filtering and bioinformatics pipeline

Raw paired-end reads were trimmed if the quality dropped below a score of phred20, based on a sliding-window of four bases using Trimmomatic v0.38 (Bolger et al., 2014). Reads below our minimum length requirement of 250 bases were discarded. Finally, reads were trimmed if Illumina adapters were present, filtered for microbial and fungal contaminants using Kraken v2.0.6 (Wood and Salzberg, 2014) and merged using PEAR v0.9.8 (Zhang et al., 2014). Merged and unmerged reads (read one only) were demultiplexed into individual sample read-sets using the 'process_radtags' feature of STACKS v1.4.6 (Catchen et al., 2013).

2.5 De novo assembly of RAD loci

We performed *de novo* assembly of unclassified RAD loci using ipyrad v0.7.29 (Eaton, 2014). Further sequence quality filtering was performed to convert base calls with a score of <30 into Ns, whilst excluding reads with ≥15 Ns. A minimum depth of six reads per individual was required for clustering of putative loci. Loci containing more than two alleles were excluded as potential paralogs.

Three datasets were assembled to address independent questions. 1) clonal identification assembly (n=37): All *B. vombata* individuals with suitable read numbers, ≥25/37 individuals present at each locus; 2) chloroplast mapping assembly (n=13): 2-3 *B. vombata* individuals per site, ≥5/13 individuals present to retain each locus; 3) phylogenetic assembly (n=12): one *B. vombata* individual per patch (n=8) plus one individual each from *B. bracteosa, B. grayi, B. riparia* and *B. walkeri*, retaining only loci with ≥6/12 samples present.

For assembly 1, reads were clustered at 99% similarity within and among each individual (2.5 [250 bp locus] to 5.6 [560 bp locus] SNPs per locus). **This similarity was selected for several reasons; (i) reducing clustering similarity to 95% (allowing for up to 28 SNPs per locus) did not substantially alter the number of loci assembled, suggesting that all assembled loci had relatively few SNPs, (ii) we only required a single SNP per locus for analyses, (iii) this level of similarity encompassed variation between technical replicate pairs and, finally, although population/species-level studies typically use comparatively more relaxed similarity clustering, (iv) we considered 2-5 SNPs per locus as a suitable number as we aimed to investigate finer-scale diversity among clones/individuals. For assemblies 2 and 3, we clustered reads at a similarity level of 95%** to allow for diversity among species.

To remove potentially linked SNPs in all assemblies, we used a '--thin 1000' filter in VCFtools v0.1.15 (Danecek et al., 2011). This allowed us to retain only the first SNP when multiple SNPs occurred within 1,000 bases of each other, which was greater than our largest locus length. The resulting VCF files were converted to 'genind' objects using the 'vcfR2genind' command. Finally, we assessed the occurrence of outlier loci using Bayescan v2.1 (Foll and Gaggiotti, 2008) and a PCA approach implemented via the 'pcadapt' package (Luu et al., 2019) in R.

2.6 Assigning clones

Ramets from a single genet may differ in their genotypes due to genotyping errors and somatic mutation and closely related but different clones may exhibit very similar genotypes. The application of thresholds is typically applied to clonal data to determine how much variation is acceptable between clonemates (Douhovnikoff and Dodd, 2003; Obrien et al., 2014).

We used multiple analyses to investigate potential thresholds among multi-locus lineages and identify *B. vombata* clones. First, we calculated a pairwise genetic distance matrix among *B. vombata* individuals using the 'bitwise.dist' function in the 'poppr' package (Kamvar et al., 2014) in R. A histogram and scatter plot of the genetic distance matrix was used to identify potential threshold values for *B. vombata* to distinguish clonemates from distinct multilocus lineages (Meirmans and Van Tienderen, 2004). We constructed an ultrametric tree ('upgma' function in 'phangorn' package; Schliep, 2011) based on the same distance matrix and overlaid a discrimination threshold based on the previous graphs (Kamvar et al., 2015). We used this combined approach to determine the number of clades supported at the point of our determined threshold. Finally, we investigated the number of genetic clusters, via Discriminant Analysis of Principal Components (DAPC; Jombart et al., 2010) using the 'adegent' package (Jombart, 2008), without a priori site input. The number of genetic clusters supported via DAPC was determined via the Bayesian Information Criterion (BIC; Schwarz, 1978).

2.7 Genetic diversity and demographics

To investigate clonal diversity we calculated F_{IS} and pairwise F_{ST} via the 'boot.ppfis' and 'pairwise.fst' functions, respectively, in the 'hierfstat' R package (Goudet 2005). Pairwise G_{ST} was calculated via the 'pairwise_Gst_Hedrick' function in the R package 'mmod' (Winter 2012). Global observed heterozygosity (Ho) and gene diversity (expected heterozygosity, He) were calculated using the 'basic.stats' function in 'heirfstat'.

2.8 Chloroplast haplotypes from ddRADseq data

We tested the viability of identifying chloroplast haplotypes using ddRADseq data in place of amplicon sequencing using assembly 2. We mapped reads from 2-3 individuals per identified clone (those with the highest read count) with ipyrad to a consensus sequence generated from the alignment of two chloroplast genomes from the family Fabaceae: *Glycine max* (GenBank

accession: CM010429.1) and *Indigofera tinctoria* (GenBank accession: KJ468098.1). Geneious v9.1.8 (https://www.geneious.com) was used to perform the alignments and to visualise the resulting phylip sequence files.

2.9 Suitability of ddRADseq for phylogenetic study of eastern Australian leafless bossiaeas

We used ipyrad to construct the phylogenetic assembly *de novo.* Similarity thresholds for within and among sample assembly was set to 95%. All other parameters were defined as above. The resulting phylip file was used to produce a Maximum Likelihood phylogeny via RAxML v8.2.9 (Stamatakis, 2014) based on the GTR+G model of evolution with 1,000 bootstrap replicates.

3. Results

3.1 Sequencing run statistics

Sequencing of our ddRADseq library resulted in 8.7 million paired-end reads that contained a maximum of a single error per 1,000 bases (Phred quality score of 30). Of the quality paired-end reads, approximately 7.8 million (90%) were retained after trimming for quality and Illumina adapters. Less than 1% of these reads were filtered out by attempting to classify them to microbial and fungal databases. On average, 30% of the retained reads overlapped sufficiently and were combined into 'single end data'. We retained only read one of the non-merged data (Unmerged read two sequences were discarded as they were considered linked to their associated read one sequence and were lesser quality). Merged and non-merged (read one only) reads (5.1 million total) were demultiplexed according to their barcodes, which resulted in 4.96 million reads being retained (97%). The mean number of reads per sample was 93,505 (SD=24,249) after discarding three samples with insufficient read numbers (<1%, relative to read numbers of other samples) (see table 1) that were attributed to error during library preparation process.

3.2 Assembly statistics

The minimum read depth (number of identical reads) per individual of each dataset was 6. Across all three assemblies, an average of 2% of our total putative loci were filtered out as they did not meet our set quality criteria. An average of 91% of loci were filtered out as they did not occur

within our threshold for minimum sample number. Approximately 1% of our final loci were filtered from all three assemblies due to treating loci with >2 alleles as potential paralogs.

Across all *B. vombata* individuals used for clonal identification (assembly 1), the average read depth per site was 8.3. This assembly produced 1,368 loci with 755 informative sites. Further filtering gave us 720 unlinked SNPs with 720 informative sites. The average read depth for our chloroplast dataset (assembly 2) was 9.2. The assembly produced 8 loci with 3 informative sites. In our phylogenetic dataset (assembly 3), the average read depth per site was 9.6. This assembly resulted in 2,893 loci with 2,866 informative sites (minor alleles occurred in >1 sample). *Bossiaea vombata, B. bracteosa, B. grayi, B. riparia* and *B. walkeri* individuals were represented in 63-93%, 27%, 11%, 29% and 13% of assembled loci, respectively. BayeScan identified zero outliers and 28 outliers were detected by pcadapt. As no outliers were common across both approaches, we considered those identified by pcadapt to be false positives and all following analyses were conducted using our three complete datasets.

3.3 Clonal assignment

We used a genetic distance-based approach to visualise a percentage threshold between diversity levels that reflect i) library preparation/sequencing error and somatic mutations within a clone and ii) inter-clonal variation representing multi-locus lineages. A clear valley was present at 0.4% genetic divergence (Fig. 2a). This threshold value was overlaid on our ultrametric topology and allowed us to identify five lineages which we treated as putative clones (Fig. 2b).

[Fig. 2]

The distribution of clones of *B. vombata* revealed that sites 1, 2 and 3 were monoclonal and each consisted of a single patch identified as a unique genet ranging in area from ca. 8 – 100 m^2. **The largest distance between clones was approximately 3 km (between sites 1 and 4)**. The fourth site, covering the greatest area, contained two clones. **One large fragmented clone dominated the site with ramets identified at least 40 m apart**. Two samples situated along a trackside edge of patch four and adjacent farmland belonged to clone five **with ramets a minimum distance of 2 m from the dominant clone**. There was no correlation between the geographic distance between sites and the genetic similarity between genets including the two clones from site 4.

DAPC analysis also supported the presence of five *B. vombata* clones in the Wombat State Forest (Fig. 3a). The probabilities of assignment to a particular group (clone) for each individual were all 100% (Fig. 3b). These results were congruent with those obtained using genetic thresholds.

[Fig. 3]

3.4 Population structure and diversity

The population structure observed via DAPC did not reflect the geographical distribution of our clones within the Wombat State Forest. **This may reflect a lack of modern admixture, historical loss of genetically intermediate individuals or may be an artefact of few viable populations/sites**. Pairwise G_{ST} values supported this finding as differences reported among clones did not correspond to proximity. Population differentiation values ranged from 0.2233 to 0.5583 and supported distinction among five clones (Table 2).

[Table 2]

Each clone showed an excess of heterozygotes, as all F_{IS} values were negative (clone 1: -4.12315, clone 2: -0.61055, clone 3: -0.8436, clone 4: -0.7785 and clone 5: -1.0749). We note that clones one and five had the most extreme negative F_{IS} values and had the fewest samples (n=2 each). Genetic variability of *B. vombata*, inferred by Ho and He, was relatively low; 0.220 and 0.130, respectively.

3.5 Chloroplast haplotypes

Mapping our chloroplast dataset (assembly 2) to our reference resulted in eight loci (2,738 nucleotides combined, <2% of the reference genome) all of which were conserved across *B. vombata* samples and references. None of the loci corresponded to sequences for cpDNA regions trnL-F and ndhF that exhibit variation among other eastern Australian leafless *Bossiaea* species (available on GenBank).

3.6 Phylogenetic relationships

Phylogenetic relationships among species were largely well supported. In our unrooted tree, *B. bracteosa* was supported as the closest relative of *B. vombata* (Fig. 4b). *Bossiaea vombata* was supported as a distinct taxon with no phylogenetic structure supported within the clade.

[Fig. 4]

4. Discussion

We found the next generation sequencing method, ddRADseq (Peterson et al., 2012), to be an efficient, cost-effective method for quantifying genetic variation and identifying clones in *B. vombata* - a highly restricted plant species lacking available genetic information. We prepared and sequenced a single library and performed three independent assemblies to determine that *B. vombata* comprises five clonally reproducing genetic individuals (genets) that range in size from ca. 8 – 102 m^2. We also obtained **eight** completely conserved cpDNA loci from *B. vombata* and distantly related members of the family Fabaceae. Finally, using data from *B. vombata* and all available close relatives, we were able to construct a phylogenetic tree with high bootstrap support - showing promise for the ability to resolve phylogenies and investigate clonality from a single library.

The identification of five distinct genets of *B. vombata* from the four known sites (three monoclonal) indicates that the genotypic diversity of the species is extremely low and cannot be determined from an in-situ census of stems. With so few genetic individuals, no site is of greater conservation value than another and all genets should be considered a single ESU. While ddRADseq is known to produce datasets with random missing data, the inclusion of technical replicates and the application of a genetic threshold allowed us to minimise the risk of overestimating clonal diversity.

Our results support clonal growth as the primary mode of reproduction at all four *B. vombata* sites, a finding that is consistent with a lack of direct evidence of seedling recruitment despite apparently viable pollen and casual observations of pod formation. In the absence of sexual reproduction, differences in clonal fitness are likely to result in loss of genets over time (Honnay and Bossuyt, 2005). The population structure of mostly monoclonal sites points to *B. vombata* being a relict species with extant populations maintained via vegetative rather than sexual reproduction (Kruckeberg and Rabinowitz, 1985). Negative F_{IS} is one signature of asexual reproduction (Halkett et al. 2005). Stoeckel et al. (2006) showed that significant negative F_{IS} in

374 *Prunus avium* could be explained by asexual reproduction and Balloux et al. (2003) predict highly
375 negative F_{IS} values only for levels of clonality approaching 1, as found for three of the four *B.*
376 *vombata* sites. As F_{IS} estimates for all *B. vombata* clones were highly negative, our genetic results
377 support clonality as the primary mode of reproduction. The identification of a second, smaller
378 clone at the perimeter of one patch at the largest site, site 4, may be an indication of rare recent
379 recruitment, however, its position at the edge of a road and nearby cleared farmland may also
380 reflect the decline of most of this clone due to habitat loss.

381 Based on limited cpDNA data available (GenBank), variation is present among (and within some)
382 leafless *Bossiaea* species. Mapping ddRADseq reads to a chloroplast genome gave us the potential
383 to investigate cpDNA haplotypes without additional amplicon sequencing. As no chloroplast
384 genomes were available for closer relatives, we were unable to determine whether the large
385 phylogenetic distance between *B. vombata* and the reference sequence limited locus identification
386 to only highly conserved regions across the family and/or whether the ddRAD process also
387 contributed due to read length or choice of restriction enzymes. However, we consider that the
388 phylogenetic distance between our reference and the leafless bossiaeas combined with the
389 technical limitations of ddRADseq are likely to have contributed to our recovery of only
390 conserved regions.

391 We consider that mapping *Bossiaea* ddRADseq loci to chloroplast reference genomes from within
392 the leafless bossiaeas will be necessary before we can assess whether variable cpDNA loci are
393 recoverable and informative using the shorter reads from ddRADseq compared to amplicon
394 sequencing. A combination of careful enzyme selection, availability of an appropriate reference
395 sequence and longer read lengths produced from alternative sequencing platforms are
396 recommended for future ddRADseq experiments. The ddRAD process results in restriction
397 fragments that mostly originate from the nuclear genome due to the relative abundance of nuclear
398 compared to plastid restriction sites. However, the number of overall fragments obtained,
399 including from the chloroplast, is highly tunable. Therefore, the use of a single library for several
400 purposes, as in our study or specifically for haplotyping, may be limited for now, however, as

technology improves and cost-per-sequencing-read continues to decrease, the viability of this approach may improve in the future.

Broadhurst et al., (2016) found that range size, life form and breeding system are some of the important predictors of genetic diversity in Australian plant species and changing environmental conditions are known to affect plant mating systems (Eckert et al., 2010). Reproductive strategies influence the distribution of genetic diversity which has a strong effect on plant population fitness and future viability (Christmas et al., 2015). Sexual failure in most exclusively clonal populations is attributed to environmental factors (Vallejo-Marín et al., 2010), and sexual extinction can eventuate where sexual reproduction is suppressed for a prolonged period and growth is almost exclusively clonal. (Honnay and Bossuyt, 2005). If reproductive constraints lead to permanent sterility within a species, its extinction is inevitable in the absence of asexual reproduction (Silvertown, 2008). Loss of genotypes can lead to monoclonal populations of a species with significant implications for population viability (Eckert, 2002; Vallejo-Marín and O'Brien, 2007). This would appear to be applicable to *B. vombata*, adding to Australian examples of woody species with functionally sterile monoclonal populations that include *Halagorodendron lucasii* (Sydes and Peakall, 1998), *Acacia anomola* (Coates, 1988) *Santalum lanceolatum* (Warburton et al., 2000), *Lomatia tasmanica* (Lynch et al., 1998) and *Grevillea renwickiana* (James and McDougall, 2014) all of which are actively managed for conservation. However, analysis of the breeding system is required to confirm sexual reproductive failure in *B. vombata* and identify its underlying basis. Investigating barriers to sexual recruitment in-situ or using an ex-situ collection, using controlled pollinations within and among clones, is feasible now that the location of clones is known.

The role of species with the capacity to buffer environmental fluctuations may be important in stabilising systems as species suites change even if those individual species are uncommon and rarely recruit from seed (Isbell et al., 2011). However, the current rarity of *B. vombata* limits its possible ecological influence unless altered habitat conditions promote rapid clonal expansion or sexual reproduction. The incremental loss of rare species will ultimately decrease unique biodiversity in ecological systems, and in the absence of detailed ecological information, there is potential for subsequent loss of species reliant on interactions with these vulnerable species for

429 survival. We have no direct observations relevant to the impact of climate change on *B. vombata.*
430 However, indirectly, the distribution of leafless bossiaeas in south-eastern Australia is consistent
431 with fragmentation influenced by historical climate processes including increasing aridity
432 (Quilty, 1994) and is likely to be impacted by ongoing changing climatic conditions (Dunlop et
433 al., 2012). Post-colonisation land use changes in the Wombat State Forest (gold mining, timber
434 harvesting, agriculture & grazing) and surrounding areas combined with changing temperature
435 and rainfall patterns in recent decades are likely to have had a significant impact on
436 environmental conditions in the areas where *B. vombata* is found now. In association with the
437 small number of clones and apparent functional sterility of *B. vombata*, those factors also have the
438 potential to affect its future persistence, without intervention, particularly if the dispersal capacity
439 of the species is limited to the rate of clonal growth. However, the small number of clones makes
440 it feasible to propagate each clone and develop an ex-situ population. This would enable study
441 of reproductive constraints to determine whether self-incompatibility or inbreeding is affecting
442 the production and viability seed given that three of the four sites are monoclonal, and the
443 remaining site has only two clones. If successful, artificial pollinations between clones may be an
444 effective way to increase genetic novelty and genotypic diversity via seed. Vegetatively
445 propagating multiples of each clone would allow the augmentation of extant populations with
446 clone/s already present at a site and raises the possibility of introducing a mix of all clones at new
447 locations to re-establish genetic connectivity.

448 The efficacy of ddRADseq data for phylogenetic studies is already well documented (Lemmon
449 and Lemmon, 2013; Ree and Hipp, 2015) and our results support its utility for phylogenetic and
450 clonal diversity studies within southeastern Australian leafless bossiaeas (and other clonally
451 reproducing species of broader interest; such as other plants, algae and corals). Developers of
452 biodiversity conservation policy acknowledge the need to retain dynamic evolutionary processes
453 (McGuigan and Sgrò, 2009; Sgrò et al., 2011) and biologically relevant conservation units can be
454 identified by characterising the evolutionary history of species and populations and clarifying
455 the genetic relationships between them. *Bossiaea vombata* is clearly a single ESU that should be
456 managed as such, but the relevant information is lacking for the remaining leafless *Bossiaea*. The
457 use of ddRADseq to generate genomic data from populations of all eastern leafless bossiaeas

would facilitate the identification of genetic lineages and the extent of clonality and clarify phylogenetic relationships between currently recognised taxa. Such a genetic overview is one step towards developing conservation guidelines for the southeastern Australian leafless bossiaea group by taking into consideration the evolutionary processes that have shaped the group to this point. Uncovering the relationships between lineages across all twelve leafless bossiaeas in eastern Australia would provide a basis for identifying both ESUs (Moritz, 1994) and MUs (Palsbøll et al., 2007), irrespective of current taxonomy, to design a conservation strategy that recognises the evolutionary history of the whole group. Practical guidelines could include hand-pollinations between natural populations or ex-situ plants to increase genetic connectivity and as a seed source to establish new populations or to augment existing populations. Ultimately, the decisions used to prioritise targets for conservation must account for their potential evolutionary and ecological contribution, in addition to factors such as extinction probability, the feasibility of removing threats and the efforts required to ensure their future.

Declarations of competing interest: none

Acknowledgements

We would like to thank Neville Walsh (RBGV) for providing information on the discovery and taxonomy of *B.vombata*, Sumali Jayaweera for volunteering her technical assistance in the lab and during sample collection, Donna Thomas (Ballarat Botanic Garden Victoria) for help with fieldwork, site information and field observations, Donna McMaster (Department of Environment Land Water and Planning) for assistance with funding as a component of project BOA2017RLTA Grampians SoE, and Charmaine DiQuattro (Victoria University) for facilitating access to laboratories.

Supplementary information

The following is supplementary information for this article:

Supplementary information Amor et al.pdf

CRediT author statement

Michael Amor: Methodology, Software, Investigation, Validation, Formal analysis, Data curation, Vizualisation, Writing- Reviewing and Editing. **Joshua Johnson:** Resources, Writing-Reviewing and Editing. **Elizabeth James:** Conceptualization, Funding Acquisition, Project Administration, Investigation, Vizualisation, Writing- Original draft preparation, Reviewing and Editing, Supervision.

Data accessibility

Raw data and all working phylip/vcf files (post-filtering and assembly) are available from GenBank (Accession: XXXXXXXXXX).

References

Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697-709.

Andrews, K.R., Good, J.M., Miller, R.M., Luikart, G., Hohenlohe, P.A., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. **17**, 81–92.

Arnaud-Haond, S., Duarte, C.M., Alberto, F., Serrao, E.A., 2007. Standardizing methods to address clonality in population studies. Mol. Ecol. 16, 5115-5139.

Balloux, F., Lehmann, L., de Meeus, T. 2003. The population genetics of clonal and partially clonal diploids. Genetics, 164, 1635–1644.

Bartolome, M., Walsh, N.G., James, E.A., Ladiges, P.Y., 2002. A new, rare species of *Acacia* from north-east Victoria. Aust. Syst. Bot. 15, 465-475.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114-2120.

Byrne, M., 2007. Phylogeography provides an evolutionary context for the conservation of a diverse and ancient flora. Aust. J. Bot. 55, 316-325.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., Cresko, W., 2013 Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124-3140.

Charpentier A., 2002. Consequences of clonal growth for plant mating. Evol. Ecol. 15, 521-530.

Christmas, M.J., Breed, M.F., Lowe, A.J., 2015. Constraints to and conservation implications for climate change adaptation in plants. Conserv. Genet. 17, 305-320.

Coates, D.J., 1988. Genetic diversity and population genetic structure in the rare chittering grass wattle, *Acacia anomala* Court. Aust. J. Bot. 36, 22-27.

Coates, D., Byrne, M., Moritz, C., 2018. Genetic diversity and Conservation units: Dealing with the species-population continuum in the age of genomics. Front. Ecol. Evol. 6. Retrieved from doi:10.3389/fevo.2018.00165

Coates, F., Walsh, N. G., James, E. A., 2002. Threats to the survival of the Grampians pincushion lily (*Borya mirabilis*, Liliaceae)—a short-range endemic from western Victoria. Aust. Syst. Bot. 15, 477-485.

Cunze, S., Heydel, F., Tackenberg, O., 2013. Are plant species able to keep pace with the rapidly changing climate? PLoS One 8, e67909.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools, Bioinformatics 27, 2156–2158.

de Witte, L. C., Stöcklin, J., 2010. Longevity of clonal plants: why it matters and how to measure it. Ann. Bot. 106, 859-870.

Dunlop, M., Hilbert, D.W., Ferrier S., House, A., Liedloff, A., Prober, S.M., Smyth, A., Martin, T.G., Harwood, T., Williams, K.J., Fletcher, C., and Murphy, H., 2012. The Implications of Climate Change for Biodiversity Conservation and the National Reserve System: Final Synthesis. A report prepared for the Department of Sustainability, Environment, Water, Population and Communities, and the Department of Climate Change and Energy Efficiency. CSIRO Climate Adaptation Flagship, Canberra.

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844-1849.

Eckert, C.G., 2002. The loss of sex in clonal plants. Evol. Ecol. 15, 501-520.

Eckert, C.G., Kalisz, S., Geber, M.A., Sargent, R., Elle, E., Cheptou, P.O., Winn, A.A., 2010. Plant mating systems in a changing world. Trends Ecol. Evol. 25, 35-43.

541 Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Buckler, E.S., Mitchell, S.E., 2011. A robust,
542 simple genotype-by-sequencing approach for high diversity species. PLoS One 6, e19379.
543 Foll, M., Gaggiotti, O.E., 2008. A genome scan method to identify selected loci appropriate for
544 both dominant and codominant markers: A Bayesian perspective. Genetics 180, 977-993.
545 Fordham, D.A., Akcakaya, H.R., Arau´jo, M.B., Keith, D.A., Brook, B.W., 2013. Tools for
546 integrating range change, extinction risk and climate change information into
547 conservation management. Ecography 36, 1–9.
548 Goudet, J., 2005. Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol Ecol
549 Notes. 5, 184-186.
550 Gross, C.L., Caddy, H.A.R., 2006. Are differences in breeding mechanisms and fertility among
551 populations contributing to rarity in *Grevillea rhizomatosa* (Proteaceae)? Am. J. Bot. 93,
552 1791-1799.
553 Gross, C.L., Nelson, P.A., Haddadchi, A., Fatemi, M., 2011. Somatic mutations contribute to
554 genotypic diversity in sterile and fertile populations of the threatened shrub, *Grevillea*
555 *rhizomatoza* (Proteaceae). Ann. Bot. 109, 331-342.
556 Halkett, F., Simon, J.C., Balloux, F. 2005. Tackling the population genetics of clonal and partially
557 clonal organisms. Trends Ecol. Evol. 20, 194–201.
558 Hoffmann, A.A., Sgrò, C.M., 2011. Climate change and evolutionary adaptation. Nature 470,
559 479-485.
560 Hoffmann, A.A., Willi, Y., 2008. Detecting genetic responses to environmental change. Nat. Rev.
561 Genet. 9, 421-432.
562 Hoffmann, A.A., Griffin, P., Dillon, S., Catullo, R., Byrne, M., Jordan, R., Oakeshott, J., Joseph,
563 L., Lockhart, P. Borevitz, J., Sgrò, C., 2015. A framework for incorporating evolutionary
564 genomics into biodiversity conservation and management. Climate Change Responses
565 2:1.
566 Honnay, O., Bossuyt, B., 2005. Prolonged clonal growth: escape route or route to extinction?
567 Oikos 108: 427–432.
568 Honnay, O., Jacquemyn, H., 2007. Susceptibility of common and rare plant species to the genetic
569 consequences of habitat fragmentation. Cons. Biol. 21, 823-831.

570 Honnay, O., Jacquemyn, H., 2010. Clonal plants: beyond the patterns—ecological and
571 evolutionary dynamics of asexual reproduction. Evol. Ecol. 24 1393–1397.
572 James, E.A., McDougall, K.L., 2007. Extent of clonality, genetic diversity and decline in the
573 endangered mallee *Eucalyptus imlayensis*. Aust. J. Bot. 55, 548-553.
574 James, E.A., McDougall, K.L., 2014. Spatial genetic structure reflects extensive clonality, low
575 genotypic diversity and habitat fragmentation in *Grevillea renwickiana* (Proteaceae), a
576 rare, sterile shrub from south-eastern Australia. Ann. Bot. 114, 413-423.
577 Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers.
578 Bioinformatics 24, 1403-1405.
579 Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new
580 method for the analysis of spatially structured populations. BMC Genetics 11, 94.
581 Jongejans, E., de Vere, N., de Kroon, H., 2008. Demographic vulnerability of the clonal and
582 endangered meadow thistle. Plant Ecol. 198, 225-240.
583 Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of
584 populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281.
585 Kamvar, Z.N., Brooks, J.C., Grünwald, N.J. 2015. Novel R tools for analysis of genome-wide
586 population genetic data with emphasis on clonality. Front. Genet. 6.
587 Kimpton, S.K., James, E.A., Drinnan, A.N., 2002. Reproductive biology and genetic marker
588 diversity in *Grevillea infecunda* (Proteaceae), a rare plant with no known seed production.
589 Aust. Syst. Bot. 15, 485-492.
590 Kruckeberg, A.R., Rabinowitz, D., 1985. Biological aspects of endemism in higher plants. Ann
591 Rev. Ecol. System. 16, 447-479.
592 Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and
593 phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99-121.
594 Luu, K., Blum, M., Privé, F., 2019. pcadapt: Fast Principal Component Analysis for Outlier
595 Detection. R package version 4.1.0. https://CRAN.R-project.org/package=pcadapt
596 Lynch, A.J.J., Barnes, R.W., Cambecedes, J., Vaillancourt, R.E., 1998. Genetic evidence that *Lomatia*
597 *tasmanica* (Proteaceae) is an ancient clone. Aust. J. Bot. 46, 25-33.

598 Lynch, A.J.J., Balmer, J., 2004. The ecology, phytosociology and stand structure of an ancient
599 endemic plant, *Lomatia tasmanica* (Proteaceae) approaching extinction. Aust. J. Bot. 52,
600 619-627.
601 Mable, B.K. 2019. Conservation of adaptive potential and functional diversity: integrating old
602 and new approaches. Cons. Genet. 20, 89-100.
603 McDougall, K.L., 2009. Four new species related to *Bossiaea bracteosa* F. Muell. Ex Benth. In
604 southeastern Australia. Telopea 12, 347-360.
605 McGuigan, K., Sgrò, C., 2009. Evolutionary consequences of cryptic genetic variation. Trends
606 Ecol. Evol. 24, 305-311.
607 McMahon, B.J., Teeling, E.C., Höglund, J., 2014. How and why should we implement genomics
608 into conservation? Evol. Appl. 7, 999-1007.
609 Meirmans, P.G., Van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: two programs for the
610 analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4, 792–794.
611 Millar, M.A., Byrne, M., Coates, D.J., 2010. The maintenance of disparate levels of clonality,
612 genetic diversity and genetic differentiation in disjunct subspecies of the rare *Banksia*
613 *ionthocarpa*. Mol. Ecol. 19, 4217-4227.
614 Mock, K.E., Callahan, C.M., Islam-Faridi, M.N., Shaw, J.D., Rai, H.S., Sanderson, S.C., Wolf, P.G.,
615 2012. Widespread triploidy in western North American aspen (*Populus tremuloides*). PLoS
616 One 7, e48406.
617 Moritz, C., 1994. Defining 'Evolutionarily Significant Units' for conservation. Trends Ecol. Evol.
618 9, 373-375.
619 Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R.K., Hedrick, P.W., 2010. Conservation
620 genetics in transition to conservation genomics. Trends Genet. 26, 177-187.
621 Palsbøll, P.J., Berube, M., Allendorf, F.W., 2007. Identification of management units using
622 population genetic data. Trends Ecol. Evol. 22, 11-16.
623 Peakall, R., Ebert, D., Scott, L.J., Meagher, P.F., Offord, C.A., 2003. Comparative genetic study
624 confirms exceptionally low genetic variation in the ancient and endangered relictual
625 conifer, *Wollemia nobilis*. Mol. Ecol. 12, 2331-2343.

626 Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double Digest
627 RADseq: An inexpensive method for de novo SNP discovery and genotyping in model
628 and non-model species. PLoS One 7, 1-11.
629 Quilty, P.G., 1994. The background: 144 million years of Australian palaeoclimate and
630 palaeogeography. In History of the Australian Vegetation: Cretaceous to Recent (Ed R.S.
631 Hill), pp. 14-43. Adelaide University Press.
632 Ree, R.H. & Hipp, A.L., 2015. Inferring phylogenetic history from restriction site-associated
633 DNA (RADseq). In Next -Generation Sequencing in Plant Systematics. (Eds E Hörandl,
634 MS Appelhans), pp. 181-204. International Association for Plant Taxonomy, IAPT:
635 Königstein, Germany.
636 Ross, J. H., 2008. A new species of *Bossiaea* (Fabaceae: Bossiaeeae) from Victoria. Muelleria 26,
637 54-56.
638 Rossetto, M., Jezierski, G., Hopper, S.D., Dixon, K.W., 1999. Conservation genetics and clonality
639 in two critically endangered eucalypts from the highly endemic south-western
640 Australian flora. Biol. Cons. 88, 321-331.
641 Rossetto, M., Lucarotti, F., Hopper, S.D., Dixon, K.W., 1997. DNA fingerprinting of *Eucalyptus*
642 *graniticola*: a critically endangered relict species or a rare hybrid? Heredity 79, 310-318.
643 Silvertown, J., 2008. The evolutionary maintenance of sexual reproduction: Evidence from the
644 ecological distribution of asexual reproduction in clonal plants. Int. J. Plant Sci. 169, 157-
645 168.
646 Schwarz, G., 1978. Estimating the Dimension of a Model. Ann. Stat. 6, 461-464.
647 Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592-593.
648 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
649 phylogenies. Bioinformatics 30, 1312-1313.
650 Sydes, M., Peakall, R., 1998. Extensive clonality in the endangered shrub *Haloragodendron lucasii*
651 (Haloragaceae) revealed by allozymes and RAPDs. Mol. Ecol. 7, 87-93.
652 Thompson, I. R., 2012. A revision of eastern Australian *Bossiaea* (Fabaceae: Bossiaeeae).
653 Muelleria 30, 106-174.

Vallejo-Marín, M., Dorken, M.E., Barrett, S.C.H., 2010. The ecological and evolutionary consequences of clonality for plant mating. Annu. Rev. Ecol. Evol. Syst. 41, 193-213.

Winter, D.J., 2012. MMOD: an R library for the calculation of population differentiation statistics. Mol. Ecol. Resour. 12, 1158-1160.

Wood, D.E., Salzburg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 1415, R46.

Young, A.G., Boyle, T., Brown, T., 1996. The population genetic consequences of habitat fragmentation for plants. Trends Ecol. Evol. 11, 413-418.

Zhang, J, Kobert, K, Flouri, T, Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614-20.

Tables

Table 1. Accession, locality and assembly information for samples included in this study.

Species	Accession	Locality (patch size)	Assembly
B. bracteosa	NGW8873	Dargo Rd, Dargo VIC	3
B. grayi	CBG7800854	Cultivated: RBGV (CANB741502) Origin: Cotter pumping station, Murrumbidgee River ACT	3
B. riparia	RBGV-070563	Cultivated: RBGV-030506. Origin: Big River 7.5 km SW of Jamieson	3
B. vombata	BV31	Wombat State Forest, VIC, site 1 (9x11 m)	NA, low reads
B. vombata	BV32	Wombat State Forest, VIC, site 1 (9x11 m)	1,2
B. vombata	BV33	Wombat State Forest, VIC, site 1 (9x11 m)	NA, low reads
B. vombata	BV34	Wombat State Forest, VIC, site 1 (9x11 m)	NA, low reads
B. vombata	BV35*	Wombat State Forest, VIC, site 1 (9x11 m)	1,2,3
B. vombata	BV36*	Wombat State Forest, VIC, site 2 (2x4 m)	1
B. vombata	BV37	Wombat State Forest, VIC, site 2 (2x4 m)	1,2,3
B. vombata	BV38	Wombat State Forest, VIC, site 2 (2x4 m)	1
B. vombata	BV39	Wombat State Forest, VIC, site 2 (2x4 m)	1,2
B. vombata	BV40	Wombat State Forest, VIC, site 2 (2x4 m)	1,2
B. vombata	BV41	Wombat State Forest, VIC, site 3 (5x11 m)	1,2
B. vombata	BV42	Wombat State Forest, VIC, site 3 (5x11 m)	1,2
B. vombata	BV43	Wombat State Forest, VIC, site 3 (5x11 m)	1
B. vombata	BV44	Wombat State Forest, VIC, site 3 (5x11 m)	1,2,3
B. vombata	BV45	Wombat State Forest, VIC, site 3 (5x11 m)	1
B. vombata	BV51	Wombat State Forest, VIC, site 4 - patch 1 (9x8 m)	1
B. vombata	BV52*	Wombat State Forest, VIC, site 4 - patch 1 (9x8 m)	1
B. vombata	BV53	Wombat State Forest, VIC, site 4 - patch 1 (9x8 m)	1
B. vombata	BV54	Wombat State Forest, VIC, site 4 - patch 1 (9x8 m)	1,2,3
B. vombata	BV55	Wombat State Forest, VIC, site 4 - patch 1 (9x8 m)	1
B. vombata	BV46	Wombat State Forest, VIC, site 4 - patch 2 (3x4 m)	1
B. vombata	BV47	Wombat State Forest, VIC, site 4 - patch 2 (3x4 m)	1
B. vombata	BV48	Wombat State Forest, VIC, site 4 - patch 2 (3x4 m)	1
B. vombata	BV49	Wombat State Forest, VIC, site 4 - patch 2 (3x4 m)	1
B. vombata	BV50	Wombat State Forest, VIC, site 4 - patch 2 (3x4 m)	1,3
B. vombata	BV56	Wombat State Forest, VIC, site 4 - patch 3 (5x13 m)	1
B. vombata	BV57	Wombat State Forest, VIC, site 4 - patch 3 (5x13 m)	1
B. vombata	BV58	Wombat State Forest, VIC, site 4 - patch 3 (5x13 m)	1,2,3
B. vombata	BV59*	Wombat State Forest, VIC, site 4 - patch 3 (5x13 m)	1
B. vombata	BV60	Wombat State Forest, VIC, site 4 - patch 3 (5x13 m)	1
B. vombata	BV61*	Wombat State Forest, VIC, site 4 - patch 4 (17x6 m)	1
B. vombata	BV62*	Wombat State Forest, VIC, site 4 - patch 4 (17x6 m)	1,2
B. vombata	BV63	Wombat State Forest, VIC, site 4 - patch 4 (17x6 m)	1,2
B. vombata	BV64	Wombat State Forest, VIC, site 4 - patch 4 (17x6 m)	1
B. vombata	BV65	Wombat State Forest, VIC, site 4 - patch 4 (17x6 m)	1,2,3
B. vombata	BV66*	Wombat State Forest, VIC, site 4 - patch 5 (14x6 m)	1,3
B. vombata	BV67	Wombat State Forest, VIC, site 4 - patch 5 (14x6 m)	1
B. vombata	BV68*	Wombat State Forest, VIC, site 4 - patch 5 (14x6 m)	1
B. vombata	BV69	Wombat State Forest, VIC, site 4 - patch 5 (14x6 m)	1
B. vombata	BV70	Wombat State Forest, VIC, site 4 - patch 5 (14x6 m)	1
B. walkeri	RBGV-070595*	Cultivated: RBGV. Origin: unknown. Natural distribution NW Victoria, western New South Wales, SW South Australia, SW Western Australia	3

Assembly numbers represent 1) clonal identification, 2) chloroplast haplotype & 3) phylogenetic assembly.
Asterisk reflects accession with technical replicate pair.

670 Table 2: Pairwise population differentiation (G_{ST}: lower left, Jost's D: upper right) among five
671 identified clones of *Bossiaea vombata*. Calculations are based on 720 ddRADseq SNPs.

	Clone 1	Clone 2	Clone 3	Clone 4	Clone 5
Clone 1	-	0.091	0.113	0.113	0.057
Clone 2	0.340	-	0.139	0.124	0.091
Clone 3	0.423	0.508	-	0.140	0.112
Clone 4	0.453	0.499	0.558	-	0.050
Clone 5	0.223	0.350	0.423	0.253	-

672

673

Figure Captions

Fig 1. Distribution map showing (a) the location of the Wombat State Forest, Victoria, Australia and (b,c) indicative locations of *Bossiaea vombata* individuals collected during this study. Maps were generated using Google's 'terrain' map data in R via the (a) 'maps' and (b) 'ggmap' packages.

Fig 2. Genetic divergence threshold identification (dashed line) among five *Bossiaea vombata* clones in the Wombat State Forest, Victoria, Australia. Analyses are based on 720 SNPs. The x-axis represents genetic distance. The y-axis represents; (a) pairwise comparisons of genetic distance for each SNP and (b) a summary of pairwise comparison occurrences (density) for a given genetic distance. We find (a, b) a distinct difference between genetic diversity representing library preparation, sequencing error and somatic mutations (left of dashed line) and that representing inter-clonal variation (right of dashed line) and (c) the presence of five multi-locus lineages when applying this threshold to an ultrametric topology based on genetic distances. Sites 1-3 were monoclonal (clone 1-3, respectively). Site 4 comprised clones 4 and 5.

Fig 3. DAPC plot of *Bossiaea vombata* individuals based on 720 SNPs obtained via ddRADseq with no input regarding a priori site information. The x-axis represents discriminant function one, which accounts for 52% of the overall variability in our dataset. The y-axis represents discriminant function two and accounts for 22% of the overall genetic variability. (a) Here, we see five distinct clusters representing independent clonal lineages. (b) Graphical representation showing membership assignment probabilities (y-axis) for 37 individuals (x-axis) as calculated via DAPC analysis. Sites 1-3 were monoclonal (clone 1-3, respectively). Site 4 comprised clones 4 and 5.

Fig 4. Unrooted phylogram **of five leafless *Bossiaea* species. Maximum Likelihood analysis was** based on 2,893 SNPS obtained via ddRADseq. **Bootstrap support values above 85 are shown at each node.** Distributions based on Thompson (2012) and voucher information from The Australasian Virtual Herbarium (https://avh.ala.org.au).

701