Microsatellite Development via Next-Generation Sequencing in *Acacia stenophylla* (Fabaceae) and *Duma florulenta* (Polygonaceae): Two Ecologically Important Plant Species of Australian Dryland Floodplains

Bruce F. Murray[1], Michael A. Reid[1] and Shu-Biao Wu[2]

[1]Geography and Planning, School of Humanities Arts and Social Sciences, University of New England, Armidale, NSW, 2351, Australia.

[2]School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351, Australia.

Corresponding Author:

Bruce Murray

Email address: bmurra20@myune.edu.au

Postal address: Earth Studies Building C02, Geography and Planning, University of New England, Armidale, NSW, 2351, Australia.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Microsatellite Development via Next-Generation Sequencing in *Acacia stenophylla* (Fabaceae) and *Duma florulenta* (Polygonaceae): Two Ecologically Important Plant Species of Australian Dryland Floodplains

Abstract

Duma florulenta and *Acacia stenophylla* are two ecologically important but understudied species that naturally occur on the floodplains and riverbanks of Australia's arid and semi-arid river systems. This paper describes the discovery and characterization of 12 and 13 polymorphic microsatellite markers for *D. florulenta* and *A. stenophylla* respectively. The number of alleles per locus for *D. florulenta* ranged from 2-12 with an average of 6.1. Across all samples, observed and expected heterozygosities ranged from 0.026 to 0.784 and 0.026 to 0.824, respectively and mean polymorphic information content was equal to 0.453. For *A. stenophylla*, the number of alleles per locus ranged between 2 and 8 with an overall mean of 4.8. Across all samples, observed and expected heterozygosities ranged from 0.029 to 0.650 and 0.029 to 0.761, respectively and mean polymorphic information content was 0.388. The developed suites of 12 and 13 microsatellite markers for *D. florulenta* and *A. stenophylla*, respectively provide opportunity for novel research into mechanisms of gene flow, dispersal and breeding system and how they operate under the extreme variability these species are exposed to in the environments in which they live.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Introduction

Microsatellites continue to be one of the most useful genetic markers for studies in molecular ecology (Guichoux et al., 2011; Vieira et al., 2016). The use of microsatellites is particularly common in the field of plant sciences with over 87 microsatellite development articles having been published in the American Journal of Botany's 'Primer Notes and Protocols in Plant Sciences' in the period spanning 2013-2105 (Vieira et al., 2016). Their continued widespread use is the result of a number of desirable characteristics, such as co-dominance, ease of use, abundant distribution throughout the genome as well as high polymorphism and reproducibility (Jarne and Lagoda, 1996; Mittal and Dubey, 2009; Santana et al., 2009; Singham et al., 2014; Sunnucks, 2000). These characteristics make them useful not only for genome mapping projects, but also in biological research, answering questions ranging in level from species (phylogenetics), through population (genetic structure) and family (parentage relatedness) to the individual (identity, sex) (Buschiazzo and Gemmell, 2006). Unfortunately, this versatility comes at a cost as microsatellite discovery and validation used to be both expensive and labor intensive. In addition, while there can be some transferability between closely related species, markers can also be species specific, meaning that for non-model organisms de-novo development of microsatellite loci is often necessary (Lepais and Bacles, 2011a). As a result, there has been a constant search for ever more cost effective and time efficient methods for the de-novo isolation of microsatellite markers since the detection of microsatellites in eukaryote genomes approximately 30 years ago (Jarne and Lagoda, 1996).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Traditionally the vast majority of DNA sequence production has relied on some form of the Sanger method that was first developed in 1977 (Sanger et al., 1977a; Sanger et al., 1977b) and sequencing in a genome scale using such a method is costly. However, the advent of next-generation sequencing (NGS) has decreased the cost of sequencing by several orders of magnitude. Following the publication of the first studies to utilize NGS for the sequencing of species with no prior genome information in 2007/8, there has been a dramatic swing towards NGS in research where large amounts of sequence data are required (Ekblom and Galindo, 2011). This is because NGS is capable of producing vast amounts of data in a relatively cost-effective manner. Applications of NGS range from full genome resequencing and more targeted discovery of mutations or polymorphisms to genome wide mapping of DNA protein interactions (Shendure and Ji, 2008). Specifically, Takayama et al. (2011) identifies the rapid and cost-effective development of microsatellite loci in non-model plant species as a particularly useful application of NGS. NGS services are offered through a number of different commercial products which include 454 sequencing (used in the 454 Genome Sequencers, Roche Applied Science; Basel), Solexa Technology (used in the Illumina (San Diego) genome analyser), the SOLiD platform (Applied Biosystems; Foster City, CA, USA), the Polonator (Dover/Harvard) and the HeliScope Single Molecule Sequencer technology (Helicos; Cambridge, MA, USA) (Shendure and Ji, 2008). The most commonly used NGS platforms for the isolation of microsatellite loci in plants are the Illumina and 454 sequencing platforms (Zalapa et al., 2012). The 454 NGS platform was the first of the NGS platforms to become commercially viable (Margulies et al., 2005; Nyrén, 2007) and as a result of longer reads (approximately 350-600 bp per read) and its ability to uncover hundreds if not thousands of microsatellite loci even at low genome coverage, it continues

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

to be the most widely used NGS platform for microsatellite loci development in plants (Zalapa et al., 2012).

The rivers and wetlands of semi-arid and arid areas of Australia are some of the most hydrologically variable and unpredictable systems in the world (Puckridge et al., 1998). Vegetation of these systems consists largely of annual herbaceous grass and forb species that rely on dormant seed banks to persist through long periods of unfavourable conditions. Large woody perennial species are much less diverse and do not produce dormant seedbanks. The dynamics of dryland floodplain seedbanks have been well documented in the literature (Capon, 2007; Capon and Brock, 2006; Capon and Reid, 2016; James et al., 2007; Reid and Capon, 2011; Webb et al., 2006). However, the mechanisms that allow these larger, structurally dominant tree and shrub species to disperse and persist in these extremely variable environments is not well known.

Acacia stenophylla A. Cunn. Ex Benth. and *Duma florulenta* (Meisn) T.M. Schust. are two of these larger woody perennial species that are common throughout the Murray Darling Basin, NSW, Australia. *A. stenophylla* is a large shrub/small tree from the Fabaceae family that has a lifespan of up to 50 years (Thomson, 1987). The species is widely distributed along watercourses and on floodplains and low lying areas of arid and semi-arid inland Australia. *Acacia stenophylla*'s weeping habit and propensity to occur in the vicinity of streams and waterbodies gives rise to two of its common names, river cooba and native willow (Doran and Turnbull, 1997). *Duma florulenta* is a woody perennial shrub from the Polygonaceae family. The species also commonly occurs in wetlands and along water courses of semi-arid and arid Australia. Thin interwoven

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

branches provide *D. florulenta* with its common name, tangled lignum. Both species are capable of reproducing sexually through seed and asexually through vegetative means, *A. stenophylla* through root suckering and *D. florulenta* through rhizomes, branch layering and stem fragments that break off the parent plant and take root. *Acacia stenophylla* and *D. florulenta* are essential to the success of bird breeding events in wetlands of dryland Australia and along with river red gum (*Eucalyptus camaldulensis*) were found to provide the main nesting substrate for more than 30 colonial and migratory bird species in the Narran Lakes wetland system (Birdlife International, 2009). *D. florulenta* in particular is the preferred nesting material of many of these bird species and was identified as a feature of critical importance in the conservation of waterbirds in a study of the Paroo wetlands (Maher and Braithwaite, 1992). Although the two species co-exist they appear to have contrasting responses to the highly variable and unpredictable environment in which they live.

A number of studies have used microsatellites to explore breeding system, genetic structure, gene flow and dispersal in plant species of riverine environments (e.g. Fér and Hroudová, 2009; Smulders et al., 2008; Wei et al., 2015; Werth and Scheidegger, 2014;). These studies have identified dispersal corridors and barriers (Wei and Jiang, 2013; Werth et al., 2014), the presence of long distance vegetative dispersal (Fér and Hroudová, 2008; Mosner et al., 2012), prevalence of hydrochory (Love et al., 2013; Pollux et al., 2007) and levels of clonality to name a few. However, the majority of these studies have taken place in European or Northern hemisphere countries where the hydrology and biology of riverine systems is often more predictable than their Australian counterparts (particularly large dryland rivers) with distinct environmental gradients and seasonal flow patterns.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Chong et al. (2013) identified different patterns of sprouting and recruitment in the tropical riverine paperbark tree *Melaleuca leucadendra* that resulted in higher levels of clonality at frequently flooded sites in comparison to sites that were subject to less frequent flooding. Hurry et al. (2013) and Robinson et al. (2012) used microsatellites to explore the genetic structure of the common reed *Phragmites australis* and the woody wetland plant *Melaleuca ericifolia* respectively in the Ramsar protected coastal wetlands of the Gippsland Lakes ecosystem, Victoria, Australia. Hurry et al. (2013) found that no clear associations between salinity level and genetic structure could be drawn with geographic distance having a greater influence on the genetic structure of *P. australis*. Robinson et al. (2012) found significant clonal structure with single stands of *M. ericifolia* corresponding to single genets with no intermingling between adjacent stands/genets identified. The sole study that used microsatellites to investigate a larger woody plant species within the dryland river ecosystems of Australia is that of Butcher et al. (2009) in their study of *Eucalyptus camaldulensis* (or river red gum), a tree species, whose natural geographic range spans virtually the entire Australian mainland. In their study spanning this entire range they found that downstream seed dispersal had less influence than geographic distance on dispersal pattern with 40 % of the genetic variation explained by latitude and moisture index. This study indicated that *E. camaldulensis* should be treated as a number of different sub-species rather than a single variable taxon.

Like many river systems worldwide, Australian dryland river systems are subject to altered hydrological conditions as a result of water resource development (Nilsson et al., 2005). In order to fully understand the implications of these changes and predict consequences of further changes, it is important to understand how dispersal mechanisms,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

gene flow and genetic variation operate under different hydrological conditions. As implied earlier, recent developments in spatial and statistical analyses coupled with the emergence of cost-effective and highly-resolving genetic markers, microsatellites for instance, have meant that landscape connectivity and population processes such as dispersal and breeding systems can be more easily and readily evaluated (Sunnucks and Taylor, 2008). Although a limited number of genetic markers have been developed in other species of Acacia (e.g. *A. harpohylla*, (Lepais and Bacles, 2011a)), currently no species specific genetic markers exist for either species *D. florulenta* or the genus Duma. This paper describes the use of 454 pyrosequencing for the discovery and validation of microsatellite loci in *D. florulenta* and *A. stenophylla*, two tree/shrub species adapted to the highly variable and unpredictable environments that constitute Australia's dryland river ecosystems.

Methods

Sample Collection and DNA Extraction

A total of 40 *A. stenophylla* and 39 *D. florulenta* samples were collected from 3 sites located on 3 different rivers (The Darling, The Warrego and The Balonne) of the Northern Murray Darling Basin (Figure 1). Phyllode samples (*A. stenophylla*) and leaf and stem samples (*D. florulenta*) were dried in the field with silica gel. A lack of moisture at the Warrego River site meant that leaves were absent from all but one of the *D. florulenta* individuals, in these cases leaf samples were replaced by samples of green stem. Approximately 20 mg of each of the silica dried *A. stenophylla* and *D. florulenta*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

samples was ground mechanically using a Mixer Mill MM 301 (Retsch GmbH & Co., Haan, Germany) at 30 Hertz for 2 minutes (4 minutes for stem samples). Genomic DNA was isolated using a Bioline ISOLATE II plant DNA kit (Bioline, Sydney, Australia) according to the manufacturer's protocol. Following isolation, purity and concentration of DNA samples was determined using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA).

Sequencing, Microsatellite Discovery and Primer Design

A high-quality DNA sample from each species was then sent to the Australian Genome Research Facility for shotgun library preparation and a ¼ picotiter-plate run (⅛ per species) of next generation sequencing with the 454 Roche GS FLX sequencing platform (Roche / 454 Life Sciences, Branford, Ct, USA). Sequences were subject to the standard quality filtering and trimming performed by GS-FLX software.

Sequence data in FASTA format was then run through the QDD v 1.3 pipeline (Meglécz et al., 2010) in order to identify microsatellite repeat regions and design locus specific primers for PCR amplification. Primer design was carried out with Primer3 (Rozen and Skaletsky, 2000) which is imbedded in the QDD pipeline. Microsatellite regions were identified using a minimum search criterion of 5 di-nucleotide repeats. Primer design was carried out based on the following criteria; a final product size of between 150 and 500 bp, optimal GC content of 50% with a range between 20% and 70%, an optimal melting temperature of 60°c with a range between 55°c and 63°c and a primer length ranging between 18 and 27 base pairs. From the microsatellite regions for which

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

primers were designed only those with dinucleotide repeats were selected as they were the most common and are less frequent in gene regions than larger trinucleotide sequence repeats (Morgante et al., 2002). A total of 48 primer pairs were chosen for further testing based on their product size (less than 350 bp and occupying all of 4 size range groups; 150-199, 200-249, 250-299 and 300-350), self-primer and primer dimerization scores (≤ 4 for *A. stenophylla* and ≤ 6 for *D. florulenta*) and melting temperatures ranging from 58.5-60.5°c. Overall this included: 15 primer pairs in the 150-199 bp category, 14 primer pairs in the 200-249 bp category, 14 primer pairs in the 250-299 bp category and 5 in the 300-350 bp category for *D. florulenta* and 15 primer pairs in the 150-199 bp category, 14 primer pairs in the 200-249 bp category, 12 primer pairs in the 250-299 bp category and seven primer pairs in the 300-350 bp category for *A. stenophylla*.

PCR Amplification and Microsatellite Validation

Initial PCR reactions comprised of 1x Bioline MyTaq reaction buffer (includes 5 mM dNTPs and 15 mM $MgCl_2$), 320 nM of forward and reverse primers, 1 U of Bioline MyTaq DNA polymerase, approximately 40 ng DNA template and sterilized water up to a final reaction volume of 20 µL. A touchdown PCR program was used, which consisted of an initial denaturation step of 5 mins at 94°C followed by three cycles of denaturation for 30s at 94°C, annealing for 45s at 60°C and elongation for 45s at 72°C. This step was repeated for three cycles with annealing temperatures of 57°C and 54°C and for 30 cycles at an annealing temperature of 52°C. The last step was a final elongation at 72°C for 10 mins. PCR products were stained with GelStar (Lonza Rockland, ME, USA) and screened for amplification using 1.5% agarose gel. Microsatellites that were successfully amplified

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

were further screened for polymorphism with ten individuals using a polyacrylamide gel stained with GelStar. A total of 15 primer pairs, that resulted in PCR products with gel patterns showing polymorphism, were selected for fragment analysis by capillary electrophoresis. To reduce the costs by multiplexing, PCR products were labelled using M13 universal primers as outlined in Sheulke (2000). Total PCR reaction volumes were 15 µL consisting of 160 nM reverse primer and fluorescently tagged M13 universal primer sequence (TGT AAA ACG ACG GCC AGT), 40 nM of the forward primer with M13 tail, 1x Bioline MyTaq reaction buffer, 0.75 U Bioline MyTaq DNA polymerase and 30 ng DNA template. PCR was carried out with a total of 40 individuals for *A. stenophylla* (15 Darling, 15 Balonne and 10 Warrego) and 39 individuals for *D. florulenta* (14 Darling, 15 Balonne and 10 Warrego). The PCR program remained the same as previously described. Multiplex microsatellite analysis was performed using a multiplex genotyping method where PCR products were amplified in simplex and then mixed before loading into the same electrophoresis gel channel, i.e., sequencer capillary (Vieira et al., 2016). Microsatellite groupings for multiplex genotyping were determined using Multiplex Manager 1.0 (Holleley and Geerts, 2009) resulting in 3 groups of five loci for each species. PCR products were analysed with applied Biosystems Genescan LIZ-500 on a 3730 genetic analyser (California, USA). Alleles were scored using GeneMapper v 4.0 and 18 bp were subtracted from total fragment sizes in order to account for the effect of adding the M13 primer tail to locus-specific forward primers.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Statistical Analysis

GenAlEx 6.503 (Peakall and Smouse, 2006; Peakall and Smouse, 2012) was used to estimate the number of alleles, Observed and expected heterozygosities, probability of identity (PI) and inbreeding co-efficients for each of the 12 and 13 markers for 39 and 40 samples of *D. florulenta* and *A. stenophylla*, respectively. Calculations for probability of identity included both a regular PI equation that do not take into account the possibility of related individuals being sampled and a more stringent equation that accounts for the sampling of relatives, PI_{sibs}. INEst (Chybicki and Burczyk, 2009) was used to estimate null allele frequencies as it provides methods that simultaneously estimate null alleles and inbreeding coefficients producing null allele frequency estimates that account for the effect of inbreeding. Polymorphic Information Content (PIC) was estimated using CERVUS (Kalinowski et al., 2007) and the presence of linkage disequilibrium and deviations from Hardy-Weinberg Equilibrium (HWE) were estimated using GenePop 4.4.3 (Rousset, 2008). Markers with a PIC > 0.5 are considered to be highly informative, markers with PIC > 0.25 are considered to be moderately informative while markers with PIC < 0.25 are considered to have low information content (Langen et al., 2011).

Results

Sequencing and Microsatellite Identification

A ¼ plate of 454 next-generation sequencing revealed a total of 301,006 de-multiplexed reads. Of these reads slightly more were obtained for *A. stenophylla* than

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

D. florulenta. The total number of reads for *A. stenophylla* was 158,392 with an average sequence length of 439 bp, while the total number of reads for *D. florulenta* was 142,614 with an average sequence length of 438 bp. Following analysis with QDD and Primer3, primers were designed for 893 perfect and 247 compound microsatellites in *A. stenophylla* and 354 perfect and 59 compound microsatellites in *D. florulenta* from 1004 and 372 sequences, respectively. The vast majority of microsatellites consisted of dinucleotide repeats with 731 (82%) and 241 (68%) for *A. stenophylla* and *D. florulenta* respectively. Trinucleotides were the next most frequent with 147 (16%) and 104 (29%) while tetra, penta and hexanucleotides collectively made up less than 3% in both species (Figures 1 and 2). The AT/TA repeat type was the most common for both species with 348 for *A. stenophylla* and 128 for *D. florulenta* or 48 and 53 per cent of the total number of dinucleotide repeat types respectively (Figures 1 and 2).

The majority of loci with dinucleotide repeat sequences in both species consisted of only five repeats, 59% for *D. florulenta* and 46% for *A. stenophylla* (Figures 3 and 4). Only 5.4% of *D. florulenta* and 13.7% of *A. stenophylla* loci consisted of 10 or more repeats with the AT repeat type making up 46.2% and 82% respectively (Figures 3 and 4).

PCR Amplification and Microsatellite Selection

Of the 48 primer pairs selected for each species, amplification of a PCR product within the expected size range was successful for all but 3 (94%) of the *A. stenophylla* primer pairs and all but 8 (83%) of the *D. florulenta* primer pairs. Polyacrylamide gel

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

patterns were consistent with polymorphism for 17 *D. florulenta* primer pairs and 19 *A. stenophylla* primer pairs in ten respective individuals selected from different populations. On inspection of the electrophoretograms of the 15 primer pairs that were selected for fragment analysis for each species, it was apparent that one *A. stenophylla* marker failed to produce a clear product with the fluorescently labelled universal primers, two of the *D. florulenta* primer pairs were monomorphic and one primer pair of each species produced multiple uninterpretable peaks. These markers were discarded leaving 13 *A. stenophylla* and 12 *D. florulenta* markers for for further validation (Table 1.).

Validation

Duma florulenta

For the 12 microsatellite loci selected for further validation, the number of alleles across all samples ranged from 2-12 with an overall mean of 6.1. Observed and expected heterozygosities ranged from 0.026 to 0.784 and 0.026 to 0.824 respectively (Table 2.). PIC ranged from 0.025 to 0.789 with an overall mean across all samples and loci of 0.453 (Table 2.). Within the three populations the number of alleles ranged between 1 and 9, observed and expected heterozygosities ranged from 0.0 to 0.867 and 0.0 to 0.880 respectively and PIC ranged between 0.0 and 0.834 (Table 2). Mean F values were high for Darling (0.216) and Warrego (0.288) River populations while F was lower (0.070) but still positive at the Balonne river site (Table 2.). This suggests a high level of inbreeding which is not surprising given the tendency of the species to reproduce vegetatively. Null allele frequency estimates ranged from 0.014 to 0.240 with frequencies above 0.100

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

occurring in 5 of the 12 alleles, namely Df 20, Df 40, Df 78, Df 87 and Df 88 (Table 2.). PI and PI_{sibs} values ranged from 0.039 to 1.0 and 0.334 to 1.0 respectively. Cumulative PI for all populations and overall were well below 0.001 while cumulative PI_{sibs} were all below 0.002 (Table 2.). A total of four markers showed significant deviation from HWE (initial $\alpha = 0.05$ following sequential Bonferroni correction) as a result of heterozygote deficit; Df 87 and 88 at Darling, Warrego and Global population levels; Df 40 at Darling and global population levels; and Df 3 solely at the global population level. Following sequential Bonferroni correction none of the loci showed significant signs of linkage disequilibrium at an initial α level of 0.05.

Acacia stenophylla

Across all loci and samples, the number of alleles ranged between 2 and 8 with an overall mean of 4.8. Observed and expected heterozygosities ranged from 0.029 to 0.650 and 0.029 to 0.761 respectively. PIC ranged between 0.028 and 0.715 with a mean across all samples and loci of 0.388. The number of alleles across the three populations of *A. stenophylla* ranged between 1 and 8, observed and expected heterozygosities ranged from 0.0 to 0.800 and 0.0 to 0.811 respectively and PIC ranged from 0.0 to 0.754. Mean F values were positive for all populations with the highest value of 0.167 recorded at the Darling River site (Table 3.). The overall mean F value was 0.108 suggesting some level of inbreeding in *A. stenophylla* populations. Null allele frequencies of 0.1 or above were recorded for 3 markers As 90 at all populations (0.164, 0.211, and 0.245), As 65 at Darling and Warrego River sites (0.100, 0.173) and As 39 at the Balonne River site (0.229) (Table 3.). PI and PI_{sibs} values ranged from 0.087 to 1.0 and 0.390 to 1.0. Cumulative PI values were well below 0.001 and cumulative PI_{sibs} values were below 0.003 for all populations

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

(Table 3.). Only one loci, As 90, showed heterozygote deficit significantly different from HWE at all populations and global level. All other loci did not differ significantly from HWE at an initial α of 0.05 following sequential Bonferroni corrections nor did any loci show signs of linkage disequilibrium.

Discussion

A Total of 142,614 *D. florulenta* and 158,392 *A. stenophylla* sequences were obtained through a ¼ plate run of 454 shotgun next generation sequencing. This is representative of much higher numbers of sequences that can be obtained using NGS in comparison to older more traditional sequencing methods and is in congruence with numbers reported for other studies using 454 sequencing for microsatellite development in plants (e.g. Csensics et al., 2010; Fatemi et al., 2013). Despite both species having similar sequencing statistics the number of perfect microsatellites detected that were suitable for primer design was considerably higher in *A. stenophylla* (893) than *D. florulenta* (354). As reads obtained using NGS with shotgun library should be randomly distributed over the genome, this is a good indication of the relative abundance microsatellite repeat sequences present in these species. While the variation in frequency of microsatellites has been found to be relatively stable in angiosperm genomes a significant negative relationship between microsatellite frequency and genome size has been recorded (Shi et al., 2013). This suggests that a larger genome size may be responsible for the smaller number of microsatellite sequences identified in *D. florulenta*. Dinucleotide repeats were by far the most dominant repeat size among the microsatellites discovered constituting 82 and 68 per cent of the total microsatellites discovered for *A.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

stenophylla and D. florulenta respectively. AT/TA repeats were the most common, comprising approximately half of the dinucleotide repeat types in both species. The AG repeat type was the next most common for both *A. stenophylla* and *D. florulenta* representing 30 and 32 per cent of the total number of dinucleotide repeats respectively. This is in agreement with previous studies of the abundance of various microsatellite motifs in plants that found that the AC repeat type, that is common in animals and mammals, is not so common in plants. AT has been found to be the overwhelmingly dominant type in plants with the AG type also common (Lagercantz et al., 1993; Morgante and Olivieri, 1993).

From the 15 primer pairs selected for genotyping, 13 *A. stenophylla* and 12 *D. florulenta* microsatellites successfully amplified and produced interpretable polymorphic peaks. Mean observed and expected heterozygosities across all populations were equal to 0.383 and 0.506 for *D. florulenta* and 0.354 and 0.432 for *A. stenophylla*. Average PIC across all populations for *D. florulenta* and *A. stenophylla* was equal to 0.453 and 0.388 respectively. This suggests that these sets of markers are moderately to highly informative and will be useful for population genetic studies. These markers were selected based largely on their primer characteristics and the presence of bands indicative of polymorphism on inspection of polyacrylamide gels. This method resulted in a high proportion of suitable markers being present in the 15 markers selected for each species. Given the known positive relationship between repeat length and polymorphism, it may have been more efficient to select microsatellites with the highest number of repeats to achieve highly informative markers. However, the proportion of dinucleotide microsatellites with 10 or more repeats was very low in both species (5.4% for *D.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

florulenta and 13.7% for *A. stenophylla*) meaning that this would have severely limited the number of high repeat marker candidates for selection. Average inbreeding co-efficients were moderate to high for both species, exceeding 0.2 at all populations for *D. florulenta* and ranging between 0.028 and 0.167 for *A. stenophylla*. This is not surprising given that both species have the capability to regenerate and spread through vegetative means (Roberts and Marston, 2011). Cumulative PI_{sibs} values did not exceed 0.003 for either species at any of the populations, given their propensity for vegetative reproduction this means that these markers can be confidently used for clonal identification. Deviations from HWE were only observed for one marker in *A. stenophylla* while four markers recorded deviations from HWE in *D. florulenta*. However, none of the four *D. florulenta* markers recorded deviations from HWE across all populations. Null allele frequencies greater than 10% were estimated for 5 and 3 *D. florulenta* and *A. stenophylla* loci respectively, however, only one loci (As90) had estimated null allele frequencies of above 10% at all populations. This suggests the strong presence of null alleles at some loci, while the presence of null alleles is not necessarily detrimental to estimation of population genetic parameters (Lepais and Bacles, 2011b), precautions may need to be taken in order to account for null allele frequencies and avoid bias when dealing with these loci (Chapius and Estoup, 2007).

As previously mentioned *D. florulenta* and *A. stenophylla* are two understudied yet ecologically important species inhabiting the extremely variable and unpredictable environments that constitute Australia's dryland river systems. Despite their structural dominance, woody perennial tree and shrub species such as these are far less diverse than their herbaceous counterparts in these systems. While herbaceous species in these systems

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

are known to survive unfavorable conditions through the maintenance of soil seed banks, the mechanisms that allow these larger woody perennial species to persist is less well known. This study provides a suite of 12 and 13 microsatellite markers, for *D. florulenta* and *A. stenophylla* respectively, that will facilitate the exploration of genetic structure, gene flow, breeding system and dispersal of these species in a highly variable and unpredictable environment.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

REFERENCES

Buschiazzo E., Gemmell N. J. (2006) The Rise Fall and Renaissance of Microsatellites in Eukaryotic Genomes. *BioEssays* **28**: 1040-1050.

Butcher P. A., McDonald M. W., and Bell J. C. (2009) Congruence Between Environmental Paramaeters, Morphology and Genetic Structure in Australia's Most Widely Distributed Eucalypt, *Eucalyptus camaldulensis*. *Tree Genetics and Genomes* **5**: 189-210.

Capon S. J. (2007) Effects of Flooding on Seedling Emergence From the Soil Seed Bank of a Large Desert Floodplain. *Wetlands* **27**(4): 904-914.

Capon S. J. and Brock M. A. (2006) Flooding, Soil Seed Bank Dynamics and Vegetation Resilience of a Hydrologically Variable Desert Floodplain. *Freshwater Biology* **51**(2): 206-223.

Capon S. J. and Reid M. A. (2016) Vegetation Resilience to Mega-Drought Along a Typical Floodplain Gradient of the Southern Murray-Darling Basin. *Australia. Journal of Vegetation Science*: 1-12: DOI: 10.1111/jvs.12426.

Chapius M-P. and Estoup A. (2007) Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution* **24**(3): 621-631.

Chong C., Edwards W., Pearson R. and Waycott M. (2013) Sprouting and Genetic Structure vary with Flood Disturbance in the Tropical Riverine Paperbark *Melaleuca leucendra* (Myrtaceae). *American Journal of Botany* **100**(11):2250-2260.

Chybicki I. J. and Burczyk J. (2009). Simaltaneous Estimation of Null Allels and Inbreeding Coefficients. *Journal of Heredity* **100**(1): 106-113.

1 Csensics D., Brodbeck S. and Holdregger R. (2010) Cost-Effective, Species-Specific
2 Microsatellited Development for the Endangered Dwarf Bulrush (*Typha*
3 *minima*) Using Next-Generation Technology. *Journal of Heredity* **101**: 789-793.
4 Doran J. C. and Turnbull J. (1997) *Australian Trees and Shrubs: Speies for Land*
5 *Rehabilitation and Farm Planting in the Tropics*. Canberra, Australia:
6 Australian Centre for International Agricultural Research. 384 p.
7 Ekblom R. and Galindo J. (2011) Applications of Next Generation Sequencing in
8 Molecular Ecology of Non-Model Organisms. *Heredity* **107**: 1-15.
9 Fatemi M., Houliston G. J., Haddadchi A. and Gross C. L. (2013) Cost Effective
10 Microsatellite Markers for *Banksia Integrifilia* (Proteaceae). *Applications in*
11 *Plant Sciences* **1**(2):1-3.
12 Fér T. and Hroudová Z. (2008) Detecting Dispersal of *Nuphar lutea* in River Corridors
13 using Microsatellite Markers. *Freshwater Biology* **53**: 1409-1422.
14 Fér T. and Hroudová Z. (2009) Genetic Diversity and Dispersal of *Phragmites australis*
15 in a Small River System. *Aquatic Botany* **90**: 165-171.
16 Guichoux E., Lagache L., Wagner S., Chaumeil P., Leger P., Lepais O., Lepoittevin C.,
17 Malausa T., Revardel E., Salin F. and Petit R. J. (2011) Current Trends in
18 Microsatellite Genotyping. *Molecular Ecology Resources* **11**: 591-611.
19 Holleley C. E. and Geerts P. G. (2009) Multiplex Manager 1.0: a Cross-Platform
20 Computer Program that Plans and Optimizes Multiplex PCR. *BioTechniques* **46**:
21 511-517.
22 Hurry C. R., James E. A. and Thompson R. M. (2013) Connectivity, Genetic Structure
23 and Stress Response of *Phragmites australis*: Issues for Restoration in a
24 Salinising Wetland System. *Aquatic Botany* **104**: 138-146.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Birdlife International (2009) Important Bird Area. Fact Sheet: Narran Wetlands: Birdlife International (BI), Australia www.birdlife.org accessed on 3/9/2010.

James C. S., Capon S. J., White M. G., Rayburg S. C. and Thoms M. C. (2007) Spatial Variability of the Soil Seed Bank in a Heterogeneous Ephemeral Wetland System in Semi-Arid Australia. *Plant Ecology* 190(2): 205-217.

Jarne P. and Lagoda P. J. L. 1996. Microatellites, From Molecules to Populations and Back. *Trends in Ecology & Evolution* **11**(10): 424-429.

Kalinowski S. T., Taper M. L. and Marshall C. (2007) Revising How the Computer Program CERVUS Accomodates Genotyping Error Increases Success in Paternity Assignment. *Molecular Ecology* **16**: 1099-1106.

Lagercantz U., Ellgren H. and Adndersson L. (1993) The Abundance of Various Polymorhic Microsatellite Motifs Differs Between Plants and Vertebrates. *Nucleic Acids Research* **21**(5): 1111-1115.

Langen K., Schwarzer, J., Kullman H., Bakker T. C. M. and Thünken T. (2011) Microsatellite Support for Active Inbreeding in Cichlid Fish. *Plos One* **6**(9): 1-9.

Lepais O. and Bacles C. F. E. (2011a) Comparison of Random and SSR-Enriched Shotgun Pyrosequencing for Microsatellite Discovery and Single Multiplex PCR Optimisation in Acacia Harpophylla F. Muell. Ex Benth. *Molecular Ecology Resources* **11**: 711-724.

Lepais O. and Bacles C. F. E. (2011b) De Novo Discovery and Multiplexed Amplification of Microsatellite Markers in Black Alder (Alnus glutinosa) and Related Species Using SSR-Enriched Shotgun Pyrosequencing. *Journal of Heredity* **102**(5): 627-632.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

1 Love H. M., Maggs C. A., Murray T. E. and Provana J. (2013) Genetic Evidence for
2 Predominantly Hydrochoric Gene Flow in the Invasive Riparian Plant *Impatiens*
3 *glandulifera* (Himalayan balsam). *Annals of Botany* **112**(9): 1743-1750.
4 Maher M. T. and Braithwaite L. W. (1992) Patterns of Waterbird use in Wetlands of the
5 Paroo, a River System of Inland Australia. *Rangeland Journal* **14**(2): 128-142.
6 Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka
7 J., Braverman M. S., Chen Y-J., Chen Z., Dewell S. B., Du L., Fierro J. M.,
8 Gomes X. B., Godwin B. C., He W., Helgesen S., Ho C. H., Irzyk G. P., Jando,
9 S. C., Alenquer M. L. I., Jarvie T. P., Jirage K. B., Kim J-B. Knight J. R., Lanza
10 J. R., Leamon J. H., Lefkowitz S. M., Lei M., Jing L., Lohman K. L., Lu H.,
11 Makhijani V. B. McDade, K. E., McKenna M. P., Myers E. W., Nickerson E.,
12 Nobile J. R., Plant R., Puc B. P., Ronan M. T., Roth G. T., Sarkis G. J., Simons
13 J. F., Simpsons J. W., Srinivasan M., Tartaro K. R., Tomasz A. Vogt K. A.,
14 Volkmer G. A., Wang S. H., Wang Y., Weiner M. P., Yu P. Begley R. F. and
15 Rothberg J. M. (2005) Genome Sequencing in Microfabricated High-Density
16 Picolitre Reactors. *Nature* **437**: 376-380.
17 Meglécz E., Costedoat C., Dubut V., André G., Malausa T., Pech N.and Martin J-F.
18 (2010) QDD: a User-Friendly Program to Select Microsatellite Markers and
19 Design Primers from Large Sequencing Projects. *Bionformatics* **26**(3): 403-404.
20 Mittal N. and Dubey K. (2009) Microsatellite Markers - a New Practiec of DNA Based
21 Markers in Molecular Genetics. *Pharmacognosy Review* **3**(6):235-246.
22 Morgante M., Hanafey M. and Powell W. (2002) Microsatellites are Preferentially
23 Associated with Nonrepetitive DNA in Plant Genomes. *Nature Genetics* **30**:
24 194-200.

Morgante M. and Olivieri A. M. (1993) PCR-Amplified Microsatellites as Markers in Plant Genetics. *The Plant Journal* **3**(1): 175-182.

Mosner E., Liepelt S., Ziegenhagen B. and Leyer I. (2012) Floodplain Willows in Fragmented River Landscapes: Understanding Spatio-Temporal Patterns as a Basis for Restoration Plantings. *Biological Conservation* **153**: 211-218.

Nilsson C., Reidy C. A., Dynesius M. and Revenga C. (2005) Fragmentation and Flow Regulation of the World's Large River Systems. *Science* **308**(5720): 405-408.

Nyrén P. (2007) The History of Pyrosequencing. *Methods in Molecular Biology* **373**: 1-13.

Peakall R. and Smouse P. E. (2006) GenAlEx 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Molecular Ecology Notes* **6**: 288-295.

Peakall R. and Smouse P. E. 2012. GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-an Update. *Genetics and Population Analysis* **28**(19) :2537-2539.

Pollux B. J. A, Jong M. D. E, Steegh A., Verbruggen E., Van Groendael J. M, Ouborg N. J. (2007) Reproductive Strategy, Clonal Structure and Genetic Diversity in Populations of the Aquatic Macrophyte *Sparganium emersum* in River Systems. *Molecular Ecology* **16**: 313-325.

Puckridge J. T., Sheldon F., Walker K. F. and Boulton A. J. (1998) Flow Variability and the Ecology of Large Rivers. *Marine & Freshwater Research* **49**(1): 55-72.

Reid M. A. and Capon S. (2011) Role of Soil Seed Bank in Vegetation Responses to Environmental Flows on a Drought -Affected Floodplain. *River Systems* (June): 249-259.

Roberts J. and Marston F. (2011) *Water Regime for Wetland and Floodplain Plants*. Canberra: A source book for the Murray-Darling Basin, National Water Commission.

Robinson R. W., James E. A. and Boon P. I. (2012) Population Structure in the Clonal, Woody Wetland Plant *Melaleuca ericifolia* (Myrtaceae): an Analysis Using Historical Aerial Photographs and Molecular Techniques. *Australian Journal of Botany* **60**: 9-19.

Rousset F. (2008) GENEPOP'007: a Complete Re-implementation of the GENEPOP Software for Windows and Linux. *Molecular Ecology Resources* **8**: 103-106.

Rozen S. and Skaletsky H. 2000. Primer3 on the WWW for General Users and for Biologist Programmers. *Methods in Molecular Biology* **132**: 365-386.

Sanger F., Air G. M., Barrell B. G., Brown N. L., Coulson A. R., Fiddes J. C., Hutchison C. A., Slocombe P. M. and Smith M. (1977a) Nucleotide Sequence of the Bacteriophage ΦX174 DNA. *Nature* **265**: 687-695.

Sanger F., Nicklen S. and Coulson R. (1977b) DNA Sequencing with Chai-Terminating Inhibitors. *Proceedings of the National Academy Of Sciences of the United States of America* **74**(12): 5463-5467.

Santana Q. C., Coetzee M. P. A., Steenkamp E. T., Mlonyeni O. X., Hammond G. N. A., Wingfield M. J, Wingfield B. D. (2009) Microsatellite Discovery by Deep Sequencing of Enriched Genomic Libraries. *Biotechniques* **46**(3): 217.223.

Shendure J. and Ji H. (2008) Next Generation DNA Sequencing. Nature Biotechnology **26**(10):1135-1145.

Sheulke M. (2000) An Economic Method for the Fluorescent Labeling of PCR Fragments. *Nature Biotechnology* **18**: 233-234.

Shi J., Huang S., Fu D., Yu J., Wang X., Hua W., Liu S., Liu G. and Wang H. (2013) Evolutionary Dynamics of Microsatellite Distribution in Plants: Insights from the Comparison of Sequenced Brassica, Arabidopsis and Other Angiosperm Species. *PlosOne* **8**(3): 1-16.

Singham G. V., Othman A. S., Vargo E., Booth W. and Lee C-Y. (2014) The Advantages of Next Generation Sequencing Against the Enrichment Technique in Isolating Biomarkers from the Termite Genome. Paper Presented at the 10th Rim Termite Research Group Conference, Kuala Lumpur, Malaysia.

Smulders M. J. M., Cotrell J. E., Lefèvre F., van der Schoot J., Arens P., Vosman B., Tabbener H. E., Grassi F., Fossati T., Castiglione S. Krystufek V., Fluch S., Burg K., Vornam B., Pohl A., Gebhardt K., Alba, N., Agúndez D., Maestro C., Notivol E., Volosyanchuk R., Pospíškova M., Bordács S., Bovenschen J., van Dam B. C. Koelewijn H. P., Halfmaerten D., Ivens B., van Slycken J., Vanden Broeck A., Storme V. and Boerjan W. (2008) Structure of the Genetic Diversity in Black Poplar (*Populus nigra* L.) Populations across European River Systems: Consequences for Conservation and Restoration. *Forest Ecology and Management* **255**: 1388-1399.

Sunnucks P. (2000) Efficient Genetic Markers for Population Biology. *Trends in Ecology & Evolution* **15**(5): 199-203.

Sunnucks P. and Taylor A. C. (2008) The Application of Genetic Markers to Landscape Management. In: Pettit C, Cartwright W, Bishop I, Lowell K, Pullar D, Duncan D, (eds). *Landscape Analysis and Visualisation: Spatial Models for Natural Resource Management and Planning*. Berlin: Springer: 211-234.

Takayama K., López P. S., König C., Kohl G., Novak J. and Stuessy T. F. (2011) A Simple and Cost-Effective Approach for Microsatellite Isolation in Non-Model Plant Species Using Small-Scale 454 Pyrosequencing. *Taxon* **60**(5): 1442-1449.

Thomson L. A. J. (1987) Australian Acacias for Saline, Alkaline Soils in the Hot, Dry Subtropics and Tropics. In: Turnbull J.W. (ed). *Australian Acacias in Developing Countries*: Proceedings of and International Workshop Held at the Forestry Training Centre. Gympie, Qld, Australia: Australian Centre for Agricultural Reseach.

Vieira M. L. C., Santini L. Diniz A. L. and Munhoz C. F. (2016) Microsatellite Markers: What they Mean and Why they are so Useful? *Genetics and Molecular Biology* **39**(3): 312-328.

Webb M., Reid M., Capon S., Thoms M., Rayburg S. and James C. (2006) Are Floodplain-Wetland Plant Communities Determined by Seed Bank Composition or Inundation Periods? *IAHS* **306**: 241-248.

Wei X., Meng H., Bao D. and Jiang M. (2015) Gene Flow and Genetic structure of a mountain Riparian Tree Species, *Euptelea pleiospermum* (Eupteleaceae): How Important is the Stream Dendritic Network? *Tree Genetics and Genomes* **11**(64): 1-11.

Wei X., Meng H. and Jiang M. (2013) Landscape Genetic Structure of a Streamside Tree Species *Euptelea pleiospemum* (Eupteleaceae): Contrasting Roles of River Valley and Mountain Ridge. *Plos One* **8**(6): 1-8.

Werth S. and Scheidegger C. (2014) Gene Flow within and between Catchments in the Threatened Riparian Plant *Myricaria germanica*. *Plos One* **9**(6): 1-13.

Werth S., Schödl M. and Scheidegger C. (2014) Dams and Canyons Disrupt Gene Flow Among Populations of a Treatened Riparian Plant. *Freshwater Biology* **59**: 2502-2515.

Zalapa J. E., Cuevas H., Zhu H., Steffan S., Senalik D., Zeldin E., McCown B., Harbut R. and Simon P. (2012). Using Next-Generation Sequencing Approaches to Isolate Simple Sequence Repeat (SSR) Loci in the Plant Sciences. *American Journal of Botany* **99**(2): 193-208.

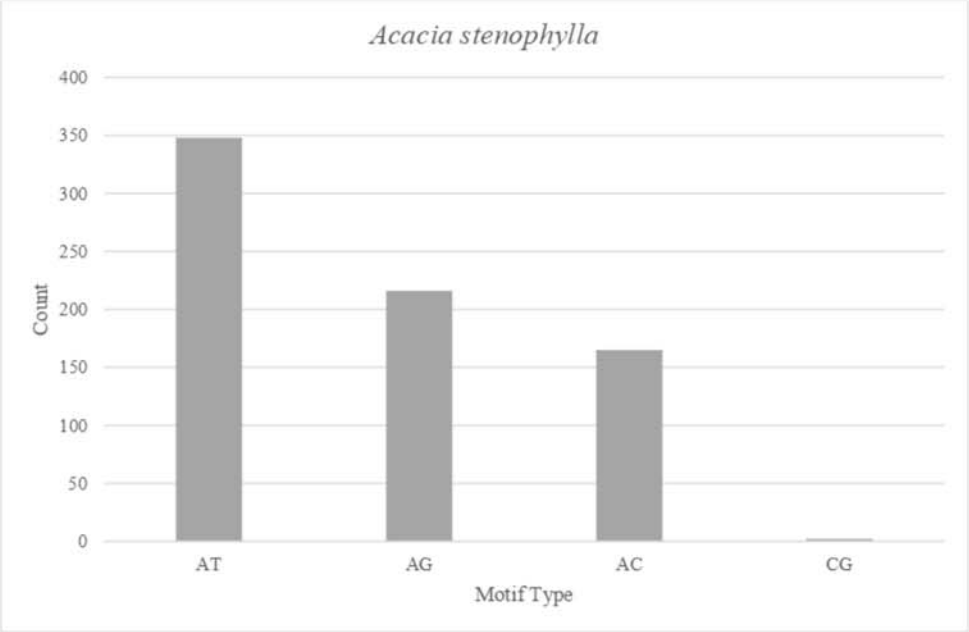

Figure 1. Number of *A. stenophylla* microsatellites with different motif types for which primer pairs were successfully designed. Note: AT motif type includes at and ta motifs; AG motif types include ag, ga, ct and tc motifs; AC motif type incudes ac, ca, tg and gt; and the CG motif type includes cg and gc motif types.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

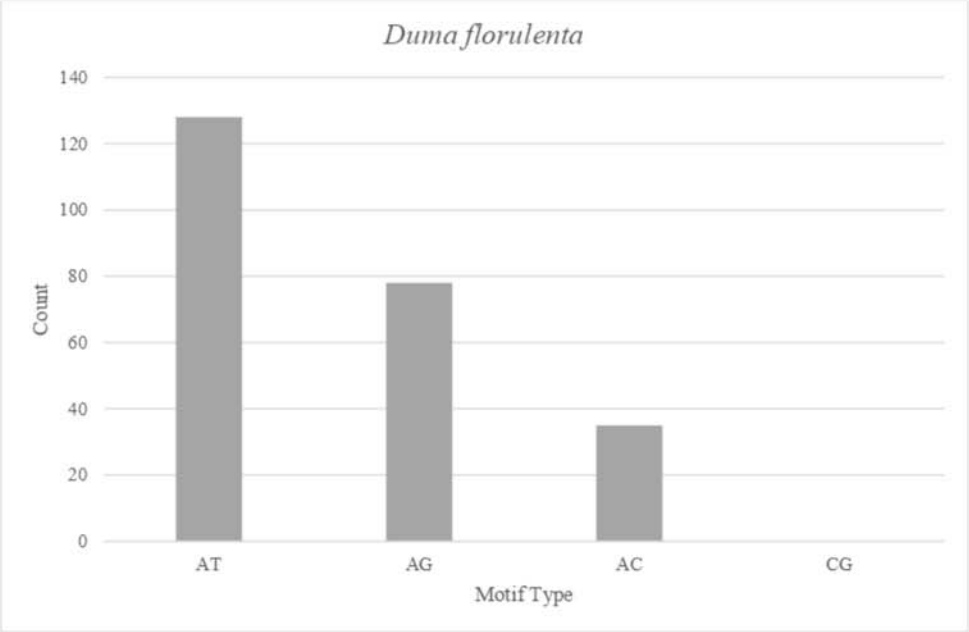

Figure 2. Number of *D. florulenta* microsatellites with different motif types for which primer pairs were successfully designed. Note: AT motif type includes at and ta motifs; AG motif types include ag, ga, ct and tc motifs; AC motif type incudes ac, ca, tg and gt; and the CG motif type includes cg and gc motif types.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

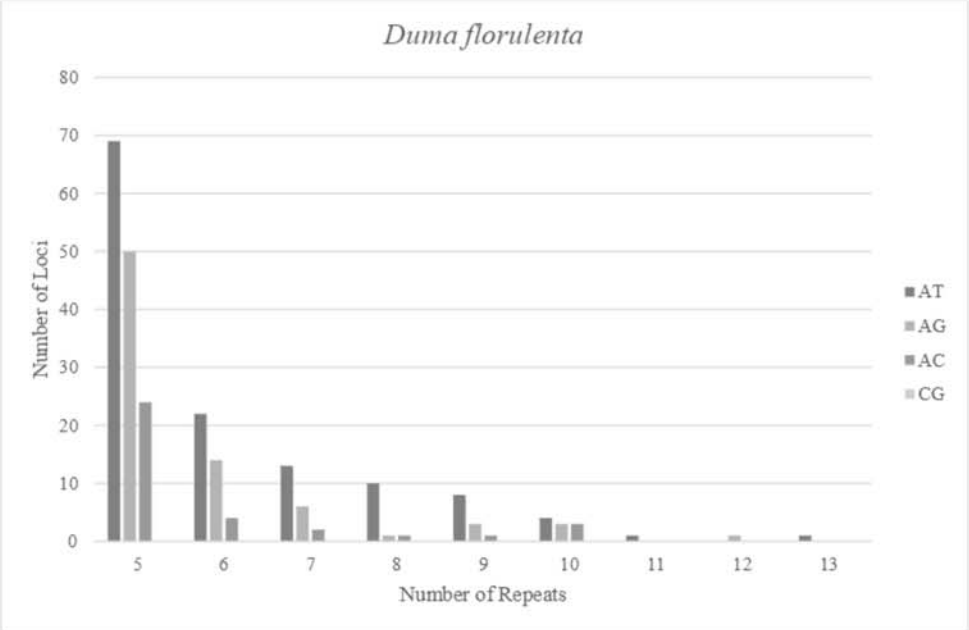

Figure 3. Distribution of dinucleotide *D. florulenta* microsatellite loci with differing numbers of repeats across motif types. Note: AT motif type includes at and ta motifs; AG motif types include ag, ga, ct and tc motifs; AC motif type incudes ac, ca, tg and gt; and the CG motif type includes cg and gc motif types.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

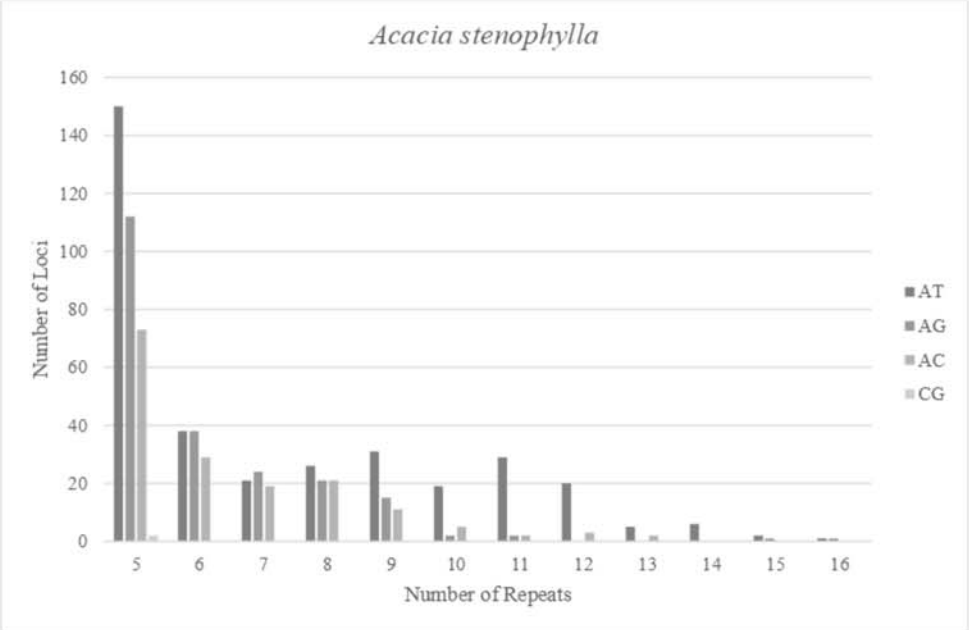

Figure 4. . Distribution of dinucleotide *A. stenophylla* microsatellite loci with differing numbers of repeats across motif types. Note: AT motif type includes at and ta motifs; AG motif types include ag, ga, ct and tc motifs; AC motif type incudes ac, ca, tg and gt; and the CG motif type includes cg and gc motif types.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Table 1. Characteristics of 12 *Duma florulenta* and 13 *A. stenophylla* microsatellite loci.

Species	Locus	Primer Sequence	Repeat Motif	Size (bp)	Fluorescent Label	GenBank Accession No.
D. florulenta	Df 3	F: TGAACCTCAACACAACTCCTCT R: AGATGTTTCCGCACGATAGC	$(TA)_7$	150	FAM	KX762273
D. florulenta	Df 5	F: AACACTCGCCATTGATGACA R: ACCCATTTTGTCTTCCTCCTG	$(GA)_6$	150	FAM	KX762281
D. florulenta	Df 20	F: CACCTGGGTTTCTATTGGAGA R: GCCACTCCTTTTCCTTTCCT	$(TA)_5$	159	FAM	KX762277
D. florulenta	Df 40	F: GAAATTACGGAAACAAGGGGA R: GGAGTTGCGATAAGGGAAGA	$(AT)_7$	182	PET	KX762280
D. florulenta	Df 45	F: CAAGTAAAGTGCGGAGGGAA R: GACATTTCTTATATCTTGGAGTTTGC	$(GA)_9$	185	VIC	KX762272
D. florulenta	Df 62	F: CTGATCTGCCTTGTTCTTGC R: TGGACACGTTCATTCTTGGA	$(CT)_6$	203	NED	KX762279
D. florulenta	Df 78	F: GAAGAACAAGGAAAACCCCA R: CCCAACATGCCCTGTATTCT	$(TA)_6$	242	NED	KX762276
D. florulenta	Df 80	F: TTTCAAAGGATTTCAACGCC R: TCACAGCACAAAACAAACCC	$(TA)_8$	244	VIC	KX762274
D. florulenta	Df 84	F: ACGCAGTTAGGCTCCTTCAA R: AGTTCCATTTGGGCCTCTCT	$(TA)_5$	257	FAM	KX762278
D. florulenta	Df 87	F: GTGGTGGAGGCCAAATTCTA R: TGCCAACTTCTTTTCTGTTGC	$(GA)_{12}$	264	PET	KX762283
D. florulenta	Df 88	F: AAGGTCAATGGGATGGAACA R: ACCTTCCCCTTTCATCGACT	$(AG)_6$	266	FAM	KX762282
D. florulenta	Df 100	F: TTGATAGGTTATTATCTTCCTGACACA R: TTGGGATGGGAATCCTAACA	$(AC)_6$	330	FAM	KX762275
A. stenophylla	As 1	F: TCCATCCTCTTCCTCTGTCC R: CGTAATGTTGTGTTCAAGGTGG	$(TC)_7$	150	FAM	KX762263
A. stenophylla	As 19	F: AATCCAACCGTGCCTACATC R: AATCAAGTGAGGAGGAGGGG	$(AT)_{11}$	160	FAM	KX762266
A. stenophylla	As 31	F: CCATTGATGTTGATCTCCTACG R: CTTTCAAGTGTCATTCCCCAA	$(AG)_7$	184	FAM	KX762268
A. stenophylla	As 39	F: CATCGTCAAATCCACGGTTA R: CCTCTCGATTGTTTTCCCCT	$(GA)_7$	197	PET	KX762265
A. stenophylla	As 51	F: TCAGGGACATCTTGGACCTC R: CTCTGACACTTCGTTCGCTG	$(GT)_8$	206	NED	KX762271
A. stenophylla	As 56	F: CTGCGTCAGAACTTGATGGA R: CCTCTCATTCCGAAAACCAG	$(TA)_{10}$	213	VIC	KX762264
A. stenophylla	As 65	F: AAAGCATTATAGCCCCAGCA R: CGACGAGGAGAATAGGCAAG	$(AT)_5$	237	VIC	KX762259
A. stenophylla	As 68	F: GCTGCCATCATCTTCAACAG R: TAAAAGGAATGGCTCGGATG	$(GT)_8$	243	PET	KX762269
A. stenophylla	As 72	F: TTCGTTTTCCCTTCATAGCC R: CTGAACCGTCGAGGTAGGAG	$(CT)_9$	252	NED	KX762260
A. stenophylla	As 73	F: GTCAAACCCAGAATCGCAGT R: CCCAGAAGCTCTGCTACCTG	$(GA)_9$	252	VIC	KX762270
A. stenophylla	As 89	F: TATCAGGTAGGGTATGCCGC R: TGATGATTCCACATTTTGGG	$(AC)_5$	285	FAM	KX762261
A. stenophylla	As 90	F: TTGACACATGGCGTCGTTAT R: GTTTGTCATGTTGGGGTTCC	$(CA)_{11}$	288	FAM	KX762267
A. stenophylla	As 96	F: AAGCTTGTTCCAATCTCCGA R: TGGCGATCTCTTCTGAATCC	$(GT)_6$	315	FAM	KX762262

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27027v1 | CC BY 4.0 Open Access | rec: 10 Jul 2018, publ: 10 Jul 2018

Table 2. Characteristics of 12 microsatellite loci in *Duma florulenta* tested in 39 individuals from 3 populations.

Duma florulenta																										
	Darling (n = 14) Geographic coordinates: 29°57'55.0"S, 146°08'48.8"E								Balonne (n = 15) Geographic coordinates: 28°23'46.7"S, 148°18'31.3"E								Warrego (n = 10) Geographic coordinates: 29°19'01.1"S, 145°50'27.4"E									
Locus	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	
Df 3	8	0.643	0.759	0.698	0.122	0.049	0.105	0.410	6	0.500	0.638	0.575	0.187	0.095	0.188	0.490	4	0.300	0.553	0.480	0.429	0.049	0.271	0.555	11	**0.**
Df 5	3	0.214	0.315	0.274	0.294	0.086	0.514	0.727	3	0.200	0.191	0.175	-0.084	0.033	0.674	0.826	2	0.200	0.189	0.164	-0.111	0.045	0.689	0.832	4	0.
Df 20	2	0.250	0.431	0.328	0.395	0.240	0.430	0.651	2	0.286	0.254	0.215	-0.167	0.095	0.600	0.778	2	0.375	0.458	0.337	0.127	0.189	0.418	0.640	2	0.
Df 40	5	**0.308**	0.637	0.574	0.498	0.186	0.188	0.491	6	0.333	0.586	0.540	0.412	0.061	0.214	0.520	6	0.500	0.726	0.658	0.275	0.033	0.129	0.437	8	**0.**
Df 45	8	0.700	0.647	0.597	-0.138	0.053	0.167	0.484	7	0.429	0.537	0.498	0.172	0.083	0.253	0.554	6	0.600	0.721	0.650	0.124	0.027	0.134	0.441	12	0.
Df 62	3	0.357	0.500	0.395	0.259	0.075	0.355	0.598	3	0.600	0.522	0.428	-0.189	0.018	0.322	0.578	2	0.300	0.268	0.222	-0.176	0.039	0.588	0.769	3	0.
Df 78	4	0.308	0.345	0.310	0.071	0.115	0.468	0.701	2	0.400	0.460	0.346	0.100	0.035	0.407	0.630	3	0.400	0.563	0.436	0.252	0.042	0.315	0.561	5	0.
Df 80	7	0.692	0.818	0.755	0.120	0.077	0.077	0.376	9	0.867	0.880	0.834	-0.018	0.014	0.039	0.334	5	0.778	0.712	0.617	-0.156	0.054	0.163	0.454	12	0.
Df 84	2	0.071	0.071	0.067	-0.037	0.079	0.869	0.933	1	0.000	0.000	0.000	n/a	n/a	1.000	1.000	1	0.000	0.000	0.000	n/a	n/a	1.000	1.000	2	0.
Df 87	8	**0.333**	0.808	0.742	0.570	0.226	0.837	0.383	7	0.733	0.828	0.774	0.083	0.020	0.066	0.366	4	**0.100**	0.711	0.619	0.852	0.131	0.162	0.453	8	**0.**
Df88	3	**0.143**	0.474	0.380	0.687	0.145	0.372	0.615	2	0.467	0.517	0.375	0.067	0.033	0.375	0.594	2	**0.000**	0.505	0.365	1.000	0.129	0.386	0.606	3	**0.**
Df100	3	0.571	0.474	0.380	-0.251	0.038	0.372	0.615	2	0.333	0.434	0.332	0.206	0.043	0.425	0.646	2	0.222	0.523	0.372	0.550	0.189	0.378	0.598	3	0.
Mean/ Cumulative	4.7	0.383	0.523	0.458	0.216	-	$9.5E^{-06}$	$9.8E^{-04}$	4.2	0.429	0.487	0.424	0.070	-	$2.2E^{-01}$	$1.6E^{-03}$	3.3	0.324	0.498	0.410	0.288	-	$9.5E^{-01}$	$1.8E^{-03}$	6.1	0.

Note: Bolded observed heterozygosities indicate significant deviation form Hardy-Weinberg Equilibrium as a result of heterozygote deficit.

Table 3. Characteristics of 13 microsatellite loci in *Acacia stenophylla* tested on 40 individuals from 3 populations.

Acacia stenophylla																										
	Darling (n = 15) Geographic coordinates: 29°57'55.0"S, 146°08'48.8"E								Balonne (n = 15) Geographic coordinates: 28°23'46.7"S, 148°18'31.3"E								Warrego (n = 10) Geographic coordinates: 29°19'01.1"S, 145°50'27.4"E									
Locus	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	**H_O**	**H_E**	**PIC**	**F**	**$null_{ITM}$**	**PI**	**PI_{sib}**	**A**	**H**
As 1	2	0.333	0.287	0.239	-0.200	0.013	0.560	0.751	3	0.733	0.536	0.414	-0.416	0.012	0.336	0.575	2	0.400	0.337	0.269	-0.250	0.034	0.514	0.718	3	0.5
As 19	5	0.600	0.674	0.612	0.078	0.014	0.161	0.464	8	0.667	0.791	0.732	0.128	0.021	0.087	0.390	6	0.700	0.784	0.710	0.060	0.026	0.100	0.403	8	0.6
As 31	3	0.467	0.515	0.445	0.063	0.014	0.305	0.577	6	0.600	0.632	0.566	0.018	0.014	0.196	0.494	3	0.400	0.637	0.527	0.339	0.062	0.234	0.506	6	0.5
As 39	1	0.000	0.000	0.000	n/a	n/a	1.000	1.000	2	0.077	0.077	0.071	-0.040	0.229	0.860	0.928	1	0.000	0.000	0.000	n/a	n/a	1.000	1.000	2	0.0
As 51	3	0.267	0.343	0.294	0.195	0.018	0.484	0.706	2	0.200	0.186	0.164	-0.111	0.028	0.689	0.832	3	0.300	0.279	0.247	-0.132	0.038	0.558	0.757	4	0.2
As 56	8	0.600	0.754	0.704	0.177	0.011	0.099	0.410	5	0.667	0.674	0.612	-0.024	0.013	0.161	0.465	5	0.400	0.747	0.659	0.437	0.081	0.135	0.429	8	0.5
As 65	2	0.400	0.460	0.346	0.100	0.016	0.407	0.630	3	0.333	0.522	0.428	0.339	0.053	0.322	0.578	2	0.111	0.111	0.099	-0.059	0.173	0.807	0.899	3	0.3
As 68	3	0.133	0.246	0.221	0.439	0.023	0.597	0.781	4	0.267	0.251	0.232	-0.101	0.023	0.584	0.775	4	0.800	0.679	0.587	-0.240	0.019	0.184	0.474	5	0.3
As 72	3	0.400	0.605	0.495	0.316	0.021	0.263	0.523	3	0.600	0.543	0.440	-0.144	0.017	0.310	0.565	5	0.667	0.712	0.617	0.009	0.077	0.163	0.454	5	0.5
As 73	1	0.000	0.000	0.000	n/a	n/a	1.000	1.000	3	0.133	0.131	0.123	-0.053	0.031	0.767	0.878	2	0.100	0.100	0.090	-0.053	0.053	0.824	0.908	4	0.0
As 89	3	0.267	0.421	0.347	0.344	0.023	0.412	0.650	3	0.400	0.503	0.396	0.178	0.032	0.354	0.595	2	0.600	0.526	0.375	-0.200	0.032	0.375	0.594	4	0.4
As 90	4	**0.385**	0.662	0.575	0.395	0.164	0.193	0.480	6	**0.267**	0.811	0.754	0.660	0.211	0.769	0.377	5	**0.222**	0.660	0.580	0.644	0.245	0.186	0.484	8	**0.2**
As 96	2	0.133	0.129	0.117	-0.071	0.018	0.774	0.881	2	0.133	0.129	0.117	-0.071	0.032	0.774	0.881	2	0.100	0.100	0.090	-0.053	0.054	0.823	0.908	2	0.1
Mean/ Cumulative	3.1	0.307	0.392	0.338	0.167	-	$5.2E^{-05}$	$1.7E^{-03}$	3.8	0.391	0.445	0.388	0.028	-	$5.2E^{-01}$	$1.7E^{-03}$	3.2	0.369	0.436	0.373	0.042	-	$1.0E^{-08}$	$2.2E^{-03}$	4.8	0.3

Note: Bolded observed heterozygosities indicate significant deviation from Hardy-Weinberg Equilibrium as a result of heterozygote deficit