Genetic diversity of three plant species used in restoration using high throughput sequencing techniques for the Coorong, Lower Lakes and Murray Mouth Recovery Project

Ed Biffin[1,3], Kor-jent van Dijk[1], Michelle Waycott[1,2,3,4], Gareth Belton[1]

[1] Australian Centre for Evolutionary Biology and Biodiversity (ACEBB), School of Biological Sciences, The University of Adelaide

[2] Environment Institute, The University of Adelaide

[3] State Herbarium of South Australia, Department of Environment, Water and Natural Resources (DEWNR)

[4] Contact: michelle.waycott@sa.gov.au, PO Box 1047, State Herbarium of South Australia, Department of Environment, Water and Natural Resources (DEWNR), Adelaide South Australia 5001.

Report to Department of Environment, Water and Natural Resources, Major Projects, Coorong, Lower Lakes and Murray Mouth Recovery Project, Major Projects Branch and The State Herbarium of South Australia, Science, Monitoring and Knowledge Branch.

Please cite this report as:

Biffin, E, Van Dijk, KJ, Waycott, M, Belton, G. 2015. Genetic diversity of three plant species used in restoration using high throughput sequencing techniques for the Coorong, Lower Lakes and Murray Mouth Recovery Project. Final report to Department of Environment, Water and Natural Resources, Major Projects, Coorong, Lower Lakes and Murray Mouth Recovery Project. University of Adelaide, Adelaide, South Australia, 2015.

Executive Summary

- Populations of three important species (*Allocasuarina verticillata, Kunzea pomifera, Schoenoplectus validus*) used in the Coorong, Lower Lakes and Murray Mouth Revegetation Project were assessed for genetic diversity and population genetic structure using novel *High Throughput Sequencing* techniques. A total of 128 samples drawn for 4-6 populations per species were screened for single nucleotide polymorphisms (SNPs), which were used to estimate population genetic parameters.
- For *Allocasuarina* and *Kunzea,* the sequencing approach yielded c. 50 and 80 SNPs, respectively. Analyses of these data suggest relatively high levels of genetic diversity within populations, and for the sample overall. With respect to population genetic structure, the *Allocasuarina* populations showed weak differentiation across the study region. For *Kunzea,* the southernmost population (Beachport) formed a cluster that was distinct from the remaining populations. These contrasting patterns may reflect higher levels of gene flow among wind (*Allocasuarina*) versus animal pollinated (*Kunzea)* species.
- The genetic data for *Schoenoplectus* suggest very low levels of diversity within populations, and across the majority of loci examined, all individuals from the same population shared an identical genotype. Three of the four populations also displayed an extremely low level of differentiation, although a single population, TST1166, was strongly divergent from the rest. These findings are suggestive of clonality, although it is unclear whether TST1166 represents a divergent clonal genotype or a distinct taxonomic entity meaning that it is possible that material used for site TST1166 may be a different species.
- The approach adopted here has shown good potential to generate informative genetic data for previously little known species that can directly inform management. For *Allocasuarina*, high genetic diversity and low differentiation suggest that seed sourcing based upon proximity to a revegetation site should be secondary to the size of the source population, as on average, larger populations should harbour more genetic diversity. For *Kunzea*, there is some evidence of geographic structure which should be taken into account for seed sourcing. Further research is needed for *Schoenoplectus*, in particular to examine the breeding system for evidence of clonality and to establish the taxonomic status of populations.

Background

Genetic data can assist restoration studies on several levels. For example, among poorly known and poorly characterised taxa, genetic approaches have proven valuable in the identification of cryptic lineages (e.g. Ndlovu et al. 2012). Molecular data have been widely applied to understanding the role of historical processes on shaping standing genetic diversity within species (e.g. McCallum et al. 2014). Interpreting the role of historical and contemporary gene flow on population genetic structure can help to inform seed collection protocols for restoration projects and for instance, strong geographic structure might argue for localised seed sourcing to maintain locally adapted genotypes. Among taxa with a high dispersal capacity we might expect little opportunity for local adaptation and a seed sourcing strategy targeting the most genetically diverse populations, irrespective of their location, might be appropriate (e.g. Broadhurst, 2011). For most species, however, little is known of the levels of genetic diversity that exist across the geographic range. Currently all plant species grown for the CLLMM (Coorong, Lower Lakes and Murray Mouth) Revegetation Project are grown from locally sourced propagules (http://www.gwlap.org.au/news.php?news=62), which might maximise local adaptation but could also compromise the immediate and long term success of revegetation projects if source populations have low levels of genetic diversity. Understanding the trade-offs between seed sourcing to preserve local adaptation and fitness versus reductions in fitness owing to a narrow genetic base are essential to developing effective revegetation strategies (Broadhurst et al., 2006).

The present study focuses on evaluating the genetic diversity within three key species used in the CLLMM Revegetation Project: *Allocasuarina verticillata* (Casuarinaceae), *Kunzea pomifera* (Myrtaceae) and *Schoenoplectus validus* (Cyperaceae). These species differ in life history attributes that are considered reasonable predictors of important population genetic parameters such as within population diversity and its distribution among populations. In particular, wind pollinated and wind/gravity seed dispersed species such as *A. verticillata* and *S. validus* might be expected to display lower genetic differentiation among populations relative to *K. pomifera* (animal pollination and seed dispersal). Woody plants (*A. verticillata* and *K. pomifera*) tend to display lower among

population differentiation and higher genetic diversity relative to non-woody species (*S. validus*) with similar life history traits (Hamrick and Godt, 1996).

In recent years, new DNA sequencing technologies have provided capacity to generate large scale, informative data sets for non-model taxa (i.e. taxa with little or no existing genomic resources) (McCormack et al. 2013) at relatively low cost. In the present study, we leverage these approaches, referred to as Next Generation (NGS) or High Throughput Sequencing (HTS) to genetically screen species populations collected from the CLLMM Management Area to explore the within and among population diversity of the three target species.

Our overall aims are:

1. To develop baseline genetic data for three key plant species used for the CLLMM Revegetation Project using HTS approaches.
2. To provide an initial estimate of within and among population genetic diversity for the three species.
3. To assess the feasibility of this approach in informing seed collection and revegetation strategies in the CLLMM region.

Methods

Sample collection and DNA extraction

Leaf tissue samples were collected from individuals of *A. verticillata, K. pomifera* and *S. vallidus* in the CLLMM region. For each species, approximately 50 individuals were sampled at each of between 6 and 9 locations (cf. population) (Appendix 1). Tissue samples were placed in silica gel for subsequent DNA extraction and long-term storage.

For baseline genetic analyses, DNA was extracted from the leaf tissues of 128 samples at the Australian Genomics Research Facility (AGRF) in Adelaide using standard protocols (http://www.agrf.org.au/services/dna-extraction). Sampling included 6 populations, each of 8 individuals, for *A. verticillata* (Figure 1) and *K. pomifera* (Figure

2). Four populations, each of 8 individuals, were included for *S. validus* the populations (Figure 3).

Library preparation and DNA Sequencing

The DNA samples were genotyped using the *CRoPS* method developed for the Ion Torrent sequencing platform by the Plant Evolutionary and Population Genetics Group at the University of Adelaide. The initial library development and screening for SNPs (single nucleotide polymorphisms) is as outlined in van Dijk and Waycott (2014) with the exception that the Ion Proton platform was used for sequencing. Post sequencing data processing as is outlined by van Dijk et al. (2014).

Following assembly of the sequence data, we extracted contigs for each individual with the consensus options: minimum coverage=50; call ambiguity if the least frequent base was present in at least 25% of the reads. The contigs (consensus sequences) for all individuals were then assembled to the reference to create alignments for each locus. Loci with less than 60% of samples represented were excluded from further analysis.

Every locus was manually screened for SNPs in Geneious v.6 (Kearse et al, 2012). Loci that had several SNPs in one sequence or showed fixed heterozygosity at a position where discarded from the analysis as these suggest the presence of paralogues (i.e. gene copies related by genome duplications). SNPs with a frequency of more than 5% where selected for subsequent analysis, and for loci with more than one such SNP, we selected the first in the alignment, i.e. one SNP was extracted per locus to ensure each data point was independent of the others. Each SNP was subsequently phased into a binary genotype format (calling the ambiguous positions heterozygotes) for downstream analyses using GenALX (Peakal and Smouse, 2006; genetic diversity and spatial structure) and STRUCTURE v2.3.4 (Pritchard et al, 2000; Population subdivision). For the STRUCTURE analyses, 20 short runs were performed (10,000 burnin and 25,000 iterations) to establish the most likely number of clusters using the *ad hoc* delta K method (Evanno et al. 2005). Subsequently, a long run was performed forcing the value of K selected by the delta K statistic. Default settings were used using an admixture model and burnin of 50,000 and 500,0000 iterations.

For the *Schoenoplectus* data, full loci were retained for phylogenetic analyses in PhyML (Guindon and Gascuel, 2003). The concatenated data were analysed using a GTR model of sequence evolution and branch support was assessed by 100 bootstrap replicates.

Results and Discussion

Data

For each species, we generated approximately 20 million sequences in total, and approximately 500,000 sequences for each individual. This equated to c. 180 high coverage (> 60% individuals represented) loci for *Allocasuarina,* yielding 51 SNPs. For the *Kunzea* data set, we recovered c. 150 high coverage loci and 81 SNPs.

After manually screening the *Schoenoplectus* data, the evident pattern was of fixed differences between populations, and virtually no variation within. The majority of variable sites within populations showed fixed heterozygosity, which can be indicative of clonality (Balloux et al., 2003). These data were therefore analysed using phylogenetic approaches only. For these analyses we extracted approximately 30 contigs that had data present for all individuals, comprising c. 3500bp when aligned and concatenated.

Genetic diversity

Genetic diversity statistics for *A. verticillata* and *K. pomifera* are presented in Table 1. The mean number of alleles provides (N_a) provides a reasonable indicator of genetic variation, as does expected heterozygosity (H_e). While these data are difficult to compare among studies, the values here are consistent with relatively high levels of genetic diversity within populations, and for the samples overall. A large difference in values of observed (H_o) and expected heterozygosity can suggest processes such as inbreeding and population stratification, but here, show fairly close correspondence consistent with a random mating hypothesis (Table 1).

Table 1: Population genetic parameters for *Allocasuarina verticillata* and *Kunzea pomifera.* For population localities refer Figures 1 and 2.

	Population	*Na*	*Ho*	*He*
Allocasuarina verticillata	TST1148	1.765	0.237	0.274
	TST1154	1.627	0.231	0.201
	TST1165	1.824	0.248	0.274
	TST1167	1.765	0.294	0.298
	TST1175	1.784	0.325	0.292
	TST1176	1.608	0.275	0.236
	Overall Mean	**1.729**	**0.268**	**0.262**
	TST1156	1.580	0.268	0.199
Kunzea pomifera	TST1160	1.543	0.240	0.202
	TST1164	1.556	0.266	0.181
	TST1170	1.593	0.303	0.235
	TST1174	1.691	0.259	0.242
	TST1180	1.704	0.300	0.251
	Overall Mean	**1.611**	**0.273**	**0.218**

Na average number of alleles
Ho observed heterozygosity
He heterozygosity expected under Hardy-Weinberg equilibrium

Population subdivision

For both *Allocasuarina* and *Kunzea,* the most like number of clusters or panmictic units identified in STRUCTURE was K=2. For *Allocasuarina,* populations TST1159, TST1165 and TST1181 are predominantly assigned to a single cluster, while the remaining populations are admixed (i.e. inferred ancestry belonging to both clusters). There is little evidence for spatially driven genetic structure across the study region (Figure 1).

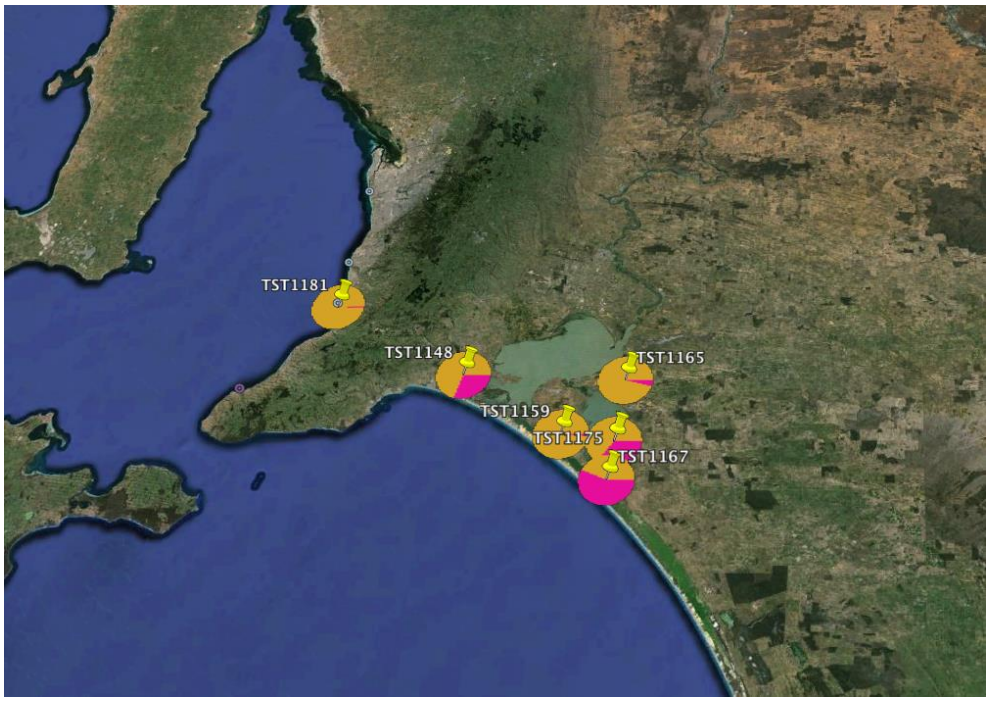

Figure 1 : Population localities for *Allocasuarina* with pie charts indicating the probability of assignment to each of 2 genetic clusters inferred using STRUCTURE

For the analyses of the *Kunzea* data, we found the 5 populations from the northern part of the study region form a cluster that is distinct from the southernmost population (Figure 2). Spatial analyses in GenALeX revealed significant support ($p \leq 0.01$) for an isolation by distance (IBD) model. That is, the observed genetic structure could reasonably be explained by the accrual of genetic differences under limited dispersal.

Figure 2: Population localities for *Kunzea* with pie charts indicating the probability of assignment to each of 2 genetic clusters inferred using STRUCTURE

The phylogenetic analysis of the *Schoenoplectus* data (Fig. 3) strongly indicates the presence of two entities. All individuals from populations TST1147, TST1150 and TST1173 form a well-supported group (BS 100%) that is distinct from TST1166. Branch lengths are proportional to the amount of evolutionary change along each branch, indicating the vast majority of differences occur between these groups (Figure 3).

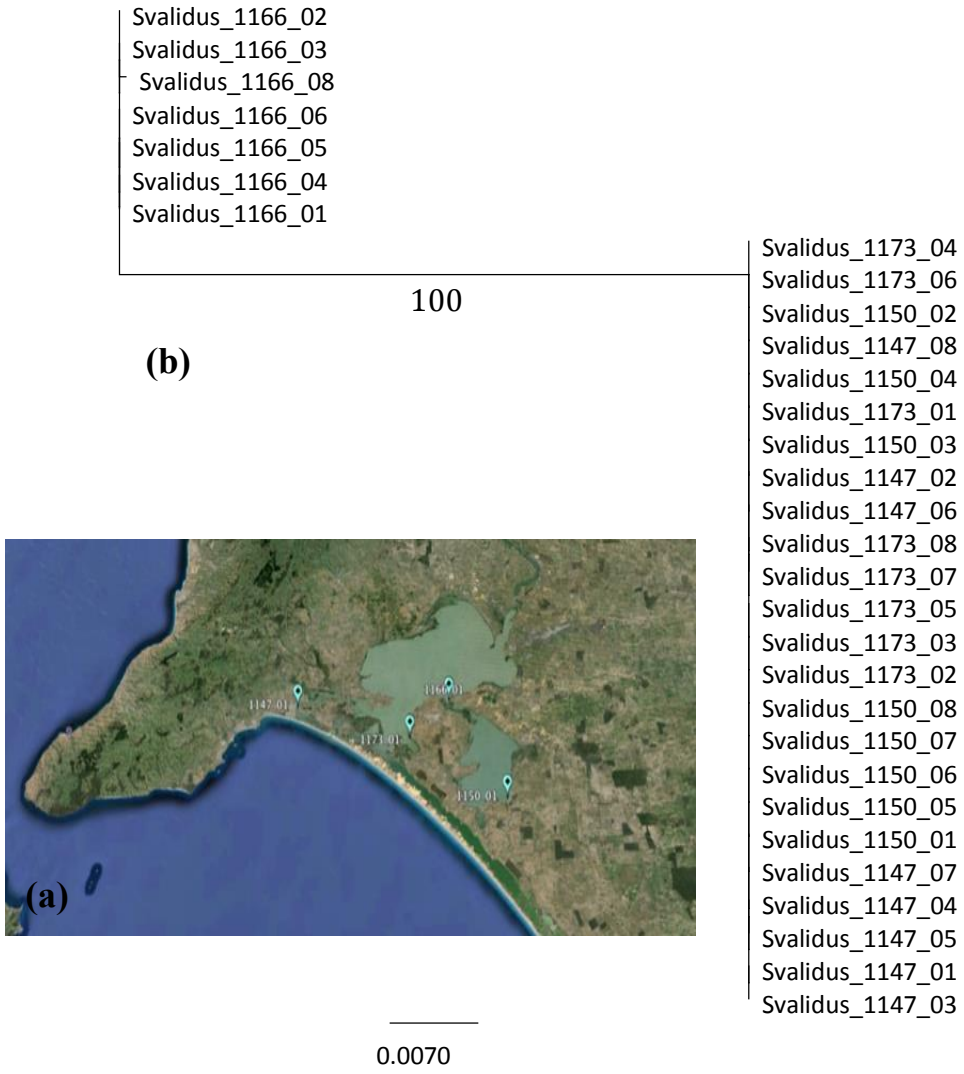

Figure 3: (a)Population localities and (b) phylogenetic relationships among *S. validus* samples. Branch lengths are proportional to the number of changes along that branch. The branch separating TST1166 from the remaining populations has 100% bootstrap support.

Implications

Analysis of both the *Allocasuarina* and *Kunzea* data sets suggest relatively high genetic diversity within populations and fairly weak evidence for population structure, particularly in *Allocasuarina*. A previous study of *A. verticillata* (Broadhurst, 2011) found patterns of high genetic diversity and low spatial structure similar to those observed here. The absence of a clear spatial genetic structure among several *A. verticillata* populations in south-western Victoria was attributed to high levels of inter-population gene flow, which could effectively prevent the development genetic

structure at the landscape scale (Broadhurst, 2011). Similar findings have been reported for other wind-pollinated species and serve to emphasise the importance of life history characteristics as predictors of population genetic structure (Broadhurst et al., 2006).

In contrast, *Kunzea* showed some differentiation with the southern-most population (TST1170; Beachport) forming a distinct genetic cluster. A disjunction across the Coorong region might reflect historical (climatic) processes that have isolated the Southeast from the Greater Mount Lofty Ranges (e.g. Guerin and Lowe, 2012) and have prevented gene flow. On the other hand, the genetic analyses reveal evidence for significant isolation by distance – that is, inter-population gene flow decreases in proportion to inter-population distance, which may reflect the behaviour of animal dispersal vectors in this species. In either case, a finding of some geographic structure suggests that seed sourcing strategies for *Kunzea* should take this into account.

For *Schoenoplectus,* the data strongly suggest there are 2 entities included in the sample as the individuals included in TST1166 (Meningie) were found to be strongly differentiated from the remaining population samples. On the other hand, there was no apparent variation *within* each of these clusters, which is suggestive of a predominantly asexual reproductive mode. Modelling approaches have found that at high levels of clonality, genetic diversity declines sharply (Balloux et al. 2003), which is the pattern observed here. On the other hand, several studies have reported strong local structure among clonal populations, and high levels of overall diversity as a consequence (e.g. Ivey and Richards, 2001). In the context of the above, the taxonomic status of the various *Schoenoplectus* populations included in this study requires clarification to determine whether the divergent clusters are different species or highly divergent clones of the same species prior to defining a seed/propagule sourcing strategy for revegetation projects. Similarly, further investigation into the reproductive biology of this species is warranted.

In general, the approach employed here shows good potential for informing revegetation strategies. The molecular methods used here are low cost, require no prior knowledge of the species genetics, and can generate high quality genomic data

for virtually any taxon. These baseline data can be sufficient to directly inform management (as for *Allocasuarina* and *Kunzea*), or can suggest future directions for study (as for *Schoenoplectus*).

The key recommendations from this study are:

- Given the relatively high level of genetic diversity and absence of strong geographic structure in *Allocasuarina verticillata,* germplasm from this species could be sourced quite broadly. According to Broadhurst (2011), genetic diversity was decreased among small, isolated populations (<30 individuals) of *A. verticillata* representing a reasonable benchmark for revegetation practitioners.
- The level of differentiation among *K. pomifera* populations suggests that are more localised sampling strategy would be prudent. While the possibility of phylogeographic (i.e. historical) structuring is not fully supported by an isolation by distance hypothesis, there remains a possibility of some local adaptation in geographically distant populations.
- Further work is required to adequately assess diversity among *Schoenoplectus,* and in particular, the taxonomic status of the populations should be established. While the genetic data are consistent with clonality, more work is required to determine reproductive modes in this species.

References

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

Broadhurst LM 2011. Genetic diversity and population genetic structure in fragmented Allocasuarina verticillata (Allocasuarinaceae) – implications for restoration *Australian Journal of Botany*, **59**, 770–780.

Broadhurst LM, Young AG, Thrall PH, Murray BG 2006. Sourcing seed for *Acacia acinacea*, a key revegetation species in south-eastern Australia. *Conservation Genetics* **7**, 49–63.

Evanno G, Regnaut S, Goudet J 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611–2620.

Guindon S and Gascuel O 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood *Systematic Biology* 52: 696-704.

Guerin, GR and Lowe, AJ 2012. Multi-species distribution modelling highlights the Adelaide Geosyncline, South Australia, as an important continental-scale arid-zone refugium. *Austral Ecology* 38: 427-435.

Hamrick, JL and Godt, MJW 1996. Effects of Life History Traits on Genetic Diversity in Plant Species. *Philosophical Transactions: Biological Sciences* 51: 1291-1298.

Ivey CT and Richards JH 2001. Genotypic Diversity and Clonal Structure of Everglades Sawgrass, *Cladium jamaicense* (Cyperaceae). *International Journal of Plant Sciences* 162: 1327-1335.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P and Drummond A 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

McCallum KP, Guerin GR, Breed MF and Lowe AJ 2104. Combining population genetics, species distribution modelling and field assessments to understand a species vulnerability to climate change. *Austral Ecology* 39: 17-28.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC and Brumfield RT 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66:526-38

Ndlovu J, Richardson DM, Wilson JRU, O'Leary M and Le Roux JJ 2013. Elucidating the native sources of an invasive tree species, *Acacia pycnantha*, reveals unexpected native range diversity and structure. *Annals of Botany* 111: 895-904.

Peakall R, Smouse PE 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Van Dijk KJ, Waycott M 2014. Genetic analysis on *Acacia pinguifolia* seedlings using deep sequencing techniques. Final report to Department of Environment, Water and Natural Resources (DEWNR). University of Adelaide, Adelaide, South Australia, 2014.

Van Dijk KJ, Waycott M, Biffin E, Cross H 2014. Population genetic analysis of *Eucalyptus paludicola*. Final report to Department of Environment, Water and Natural Resources (DEWNR). University of Adelaide, Adelaide, South Australia, 2014.