Impact of climate on the genetic diversity of native species using *Microlaena stipoides* as a model

Final Report March 2011

S. McDonald[1], F. M. Shapter[1], I.Chivers[3] , D.L.E. Waters[1] and R.J. Henry[1,2]

[1] Centre for Plant Conservation Genetics, Southern Cross University.

[2]Queensland Alliance for Agriculture and Food Innovation, University of Queensland

[3] Native Seeds Pty Ltd, Melbourne.

Summary

Genetic diversity in wild populations of a distant rice relative, (*Microlaena stipoides* (Labill.) R. Br.), was examined along a longitudinal transect. The *Isa* defence locus from 95 individuals included 7 haplotypes and was more diverse in populations from hotter and dryer environments at lower altitudes that had been disturbed for agriculture when compared to less disturbed populations in the mountains. This confirmed the earlier finding that this gene was more diverse in wild barley populations from dryer environments. Whole genome shotgun sequencing of bulked individuals from populations from contrasting environments was used to explore diversity at the genome level. Analysis of approximately 80% of the chloroplast genome sequences indicated a significantly lower level of diversity in undisturbed mountain environment. The nuclear ribosomal RNA gene sequences of the populations showed similar diversity, indicating that the pattern of variation in diversity in different environments was not genome wide. SNP were detected at between 6 and 7 SNP per 1000bp in the chloroplast and around 40 SNP per 1000bp in nuclear ribosomal genes in bulks of 11 individuals from two contrasting environments. Increased genetic diversity specifically at defence loci may be important in adapting plant populations to the greater environmental variation in warmer and dryer climatic conditions. The data support the view that these wild grass populations are retaining diversity even in highly disturbed areas and that these areas include germplasm that will provide diversity that supports adaptation to environmental change. These results emphasise the importance of conserving populations in more marginal environments as these may contain genetic diversity important for species survival in response to environmental change.

Project Scope

The aim was to characterize the genetic variation which occurs in wild populations of *Microlaena stipoides* across a range of environmental and climatic conditions. The change in average rainfall, temperature and altitude from coastal to alpine areas of Victoria can be used as a reference for how this species varies genetically and phenotypically with changes in the environment. Analysis of these adaptations in conjunction with climatic data will enable hypotheses about the genetic potential of native grasses to tolerate climate change.

Methods summary

Plant collection

Live plant specimens were collected at 36 sites along a transect from Melbourne to Paynesville. These plants have been established in two common gardens, one at the Native seeds Research site in Dumbalk Victoria and the other at Southern Cross University in Northern NSW.

Leaf specimens were harvested for 10 individuals at six sites along the transect to use for analysis of the genomic variation within a site compared with that between each of the sites. Leaf samples were also harvested for the 36 whole plants and DNA was extracted for all leaf specimens. Seeds were harvested from 30 of the plants in the common garden and stored for future breeding work. The 36 plants in the common garden experiment were sampled when the new growth was optimal, for DNA extraction and were lodged in the Australian Plant DNA Bank.

Phenotypic evaluation of the common garden experiment identified significant variation to plant size, seed quantity, colour and size, tiller number, plant architecture, leaf shape and colour, and, dry matter and grain yield.

Environmental data

Average rainfall and temperature, along with soil type and altitude was tabulated for all 37 sites (the site at the highest altitude was above the snowline and had no *M. stipoides* present) for comparison with the genetic analysis.

Genomic characterisation

The *isa* gene has been used in previous studies looking at the impact of climatic change on genetic variability in wild barleys. The homologue of this gene has been characterised in *M. stipoides* and primers for PCR amplification have been optimised. The 95 DNA samples were analysed for sequence variability.

Whole genome shot gun sequencing was conducted on bulk samples of DNA from contrasting populations at high and low altitude.

Detailed methods

Field Sampling of M. stipoides

An East-West transect was established from Studley Park Boathouse, inner Melbourne, to Paynesville on the Gippsland coast of Victoria (Australia) over a total distance of 238.6 kilometres (Fig.1).

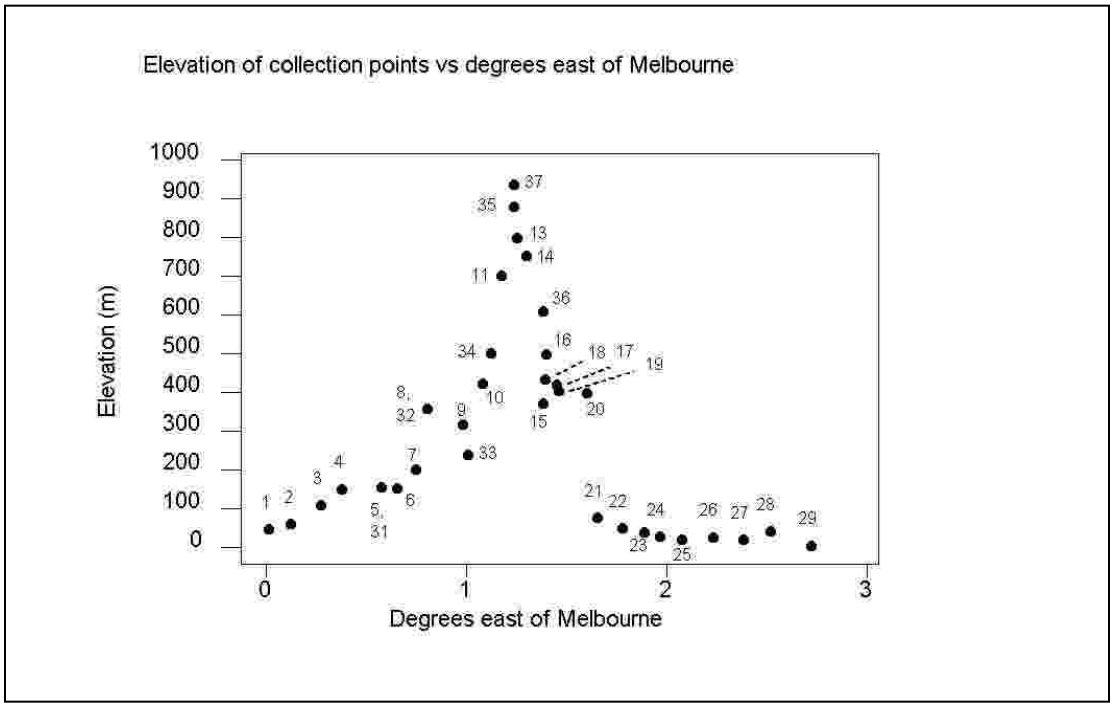

Figure 1: Comparison of the altitudes of the 37 *M. stipoides* sample sites

The transect lay within a narrow latitude range between 37.77°S and 38.02°S, a latitudinal variation of only 0.24°. Samples of *M .stipoides* were collected (Fig. 2.) from a range of altitudes ranging from just above sea level to the permanent snowline of Mount Baw Baw (939m). The populations collected were not continuous. All available mature seed, two representative live plants and leaf samples were collected at each location. Live plants were collected following the methods described in (Whalleyand Brown, 1973). GPS location and elevation data were recorded with a description of the site and soil characteristics. Soil was analysed for pH using a Manutec soil pH test kit and the values were recorded. A total of 37 sample sites were selected. These were a minimum of 10 to 15 km apart, in different valleys and where possible not on roadsides where mowing machinery can move seed from one location to the next. Sites chosen were alongside paths, in parks, along creeks or along minor dirt roads. Sites were assessed for phenotypic variation to confirm that the entire population on that site was the same phenotype. Each population was sampled and treated separately. Six of the original sampling sites (at altitudes of 3, 58, 317, 497, 882 and 939 m) were then more intensively sampled. Leaf samples were collected from ten plants within the *M. stipoides* population at each site and stored as dry specimens in silica gel. Field samples were taken from plants spaced at least 1 m apart.

Two morphologically similar sample plants were taken from each site and potted into 150 mm nursery pots. Native Seeds P/L at Sandringham, Melbourne, retained one pot from each site and the other was evaluated at Southern Cross University, Lismore in Northern New South Wales. All 35 pots at each storage location were subsequently housed in a common environment to eliminate the effect of their original environmental conditions over the following growing season.

Phenotype Evaluation

The potted samples were grown on for three months. They then were photographed and measurements were taken of the leaf blade width, leaf blade length, stolon length and the stolon width.

Climatic data

Climatic data for the *M. stipoides* population in Australia was collated from the Bureau of Meteorology web site (http://www.bom.gov.au/climate/cdo/about/sitedata.shtml). Current and past metrological stations were chosen as close as possible to the sample sites from this database the following information was obtained. The mean annual maximum and minimum temperature, the highest and lowest mean monthly temperatures, the mean of the highest and lowest recorded temperatures and the mean annual rainfall.

Maximum Temperatures

The Mean Annual Maximum temperatures decline from 20.7°C to 15.4°C with increasing altitude, this represents a decrease of 0.66°C /100m, also consistent with atmospheric adiabatic decrease in temperature

Minimum Temperatures

Mean Annual Minimum temperatures ranged from 10.2°C at 31 m to 5.3°C at 838 m, this represents a decline of 0.59°C /100m. This is consistent with atmospheric adiabatic temperature rate of change. Mean Monthly Minimum temperatures declined from 3.8°C to -0.1°C and the Mean of Lowest Temperatures declined from 0.1°C to -6.95°C

The regions with the highest recorded temperatures were from central Melbourne to Mt Evelyn and from Cowwarr region to Paynesville while the lower temperatures were in the central highlands.

Precipitation

This transect can be separated into two rainfall categories. One where the precipitation is less than 1000 mm annually and the other precipitation is greater than 1000 mm annually. This coincides with the same geographical regions as the temperatures.

The highlands tend to be the wetter regions while the lower coastal plains are drier.

DNA Preparation

DNA was isolated from all 95 samples in total. Samples were prepared by cutting leaves into 1-3 mm in length pieces and placed into microtubes. There were 60 samples from the dried specimens consisting of 10 samples each from Sites 02, 09, 18, 29, 35 and 37. All microtubes contained approximately 30-50 µg of tissue capped and stored at –20° C. DNA was extracted using Qiagen plate plant extraction kit in accordance with the manufacturer's guidelines. DNA was eluted twice eluded with 100 µL of elution buffer to give maximum yield at reduced concentration.

PCR and sequencing

Primers based upon the *Isa* gene sequence from rice (*Oryza sativa*) were not successful in gene amplification. An alignment between the *Isa* gene of *Oryza sativa* and *Isa* gene of *Hordeum vulgare* was subsequently used to identify several highly conserved areas for universal primer design. Phire® Hot Start II DNA Polymerase was used for PCR. Amplification conditions were 98° C for the melt temperature and 72° C for annealing and elongation. The PCR product was gel extracted using the Qiagen gel extraction kit as per the manufacturers guidelines and sequenced by the Sanger method.

Whole genome sequencing of population bulks

DNA samples from the two most environmentally differentiated sites, site29 and site 37, were selected for massively parallel sequencing. Eleven individuals were sampled per site. DNA from the 11 leaf specimens from each site were quantified by nanodrop spectroscopy and gel visualisation and were pooled in equimolar amounts into two pools respectively. Each sample contributed 270 ng of DNA, giving a total of 2.7 µg per pool. Each of these pooled samples were submitted for sequencing on the Illumina GAII Platform (Illumina, San Diego, CA, USA) and processed as per the methodology described by Nock *et al*., 2010.

Additional *M. stipoides* sequence data from the Nock *et al*., 2010 study was also analysed to estimate a baseline level of diversity which could be expected in this species. Data was trimmed according to the methods set out in that study however reference assembly and SNP detection parameters followed the protocol outlined below with the exceptions that reference assembly parameters were adjusted for the shorter read length (limit =5) and the minimum and maximum distance were set as 80 and 200 respectively.

Sequence data were trimmed on a CLC genomics workbench (www.clcbio.com) where reads with a quality score of less than 0.01 were discarded, paired end reads were trimmed to a minimum of 30bp and single end reads trimmed to a minimum of 20bp in length. Reads from the two sites were then assembled to a *M .stipoides* chloroplast sequence (Genbank accession GU592211) and a putative *M. stipoides* nuclear ribosomal gene sequence was determined from a *de novo* assembly of M. stipoides Illumina GA II sequence reads to a consensus of rice (M11585,

M168450, AY097328, AP005091, AF069218), corn (BT088101), Wheat (AY049041) and M .stipoides (DQ888637) nuclear ribosomal genbank accessions.

Reference assembly was completed with the following parameters; mismatch cost of 2, insertion and deletion costs of 3, length fraction of 0.8 and similarity of 0.8, minimum distance for paired end reads was set at 180 with a maximum distance of 340, criteria was set to ignore non-specific matches and vote for ambiguous calls. SNP detection parameters were assigned as follows; window length of 21, maximum number of gaps or mismatches of 2, minimum quality score of 30 for SNP site and surrounding bases, minimum coverage was initially set as 1X and a minimum variant frequency of 1%.

Secondary analysis of the CLC output was conducted using Microsoft Excel 2007 where SNP's at any reference position were eliminated from the analysis if the number of variants =1 (ie there was no polymorphism within the pool, just between the pool and the reference used). Similarly the minimum coverage was limited to 88X or greater at any given locus for the SNP to be reported. Minimum acceptable SNP frequency at any given locus was determined to be 5% for the Chloroplast analysis and 1% for the nuclear ribosomal genes. SNP were reported as total number SNP/gene and as unique (defined as the SNP only occurring at one site for any given reference position) or as a common SNP (defined as the SNP being identified in data pools from both Site 37 and Site 29). SNP frequency/1000bp was also calculated as; total number of SNP/number of bases in the alignment consensus to the reference gene sequence *1000.

Results

Phenotypic diversity in M. stipoides

The collected plants from each location were replanted in potting mix and grown for a period of five months under uniform growing conditions. This time period was sufficient to overcome the effects of the original location (eg. shading) and for the underlying phenotype to express itself. After that period four measurements were taken from each plant – the blade length and width, the internode length and the width of the stem. There was considerable variation in all of these characteristics. For example, blade width varied from 1.5 mm to 11.0 mm and blade length varied from 45 to 200 mm. Statistical analysis of these measures against both high/low rainfall and disturbed/natural environments was undertaken using two-way Analysis of Variance (General Lineal Model). No significant results were obtained indicating that there was no relationship between rainfall and site disruption and the phenotypic characteristics displayed.

Diversity at Isa locus in M. stipoides

Sequencing of the Isa locus of 95 samples of *M. stipoides* from 35 different sites samples revealed SNP at base 63,132,159,162 and 232 in the sequence. In total 7 haplotypes were detected. Three of these SNP would result in changes in the amino acid encoded suggesting that this variation is functional in plant defence.

Variation in Isa diversity at different sites

Dryer sites with a rainfall below 1000mm pa had a greater diversity at the Isa locus. At 90% of the high rainfall sites the common (A) haplotype was found. The higher diversity at the dryer sites was highly significant (P=0.001 Fishers exact test). Haplotypes F and H were only found at

the lower rainfall sites. The 11 individuals from the high rainfall site bulked for whole genome sequence analysis had only a single haplotye of *Isa* while the low rainfall site had 4 haplotyes in the 11 individuals bulked.

Intra-individual diversity in chloroplast and nuclear genomes
Data for a single plant was analysed to determine the apparent level of diversity due to variation within an individual and sequencing error. The SNP density determined for an individual *M. stipoides* plant was 6.52 SNP/1000bp for the nuclear ribosomal gene and 0.64 SNP/1000bp for the Chloroplast genome. This was determined from a reference assembly which aligned 205588 and 486494 reads respectively, resulting in average coverage of 1129 X and 121X coverage.

Chloroplast diversity
It is accepted that the chloroplast genome (cpDNA) within an individual angiosperm leaf sample predominantly exists as maternally inherited circular molecules, in multiple copies (Small et al 2004). Therefore in a pool of 11 plant samples of the same species there would be copies of the chloroplast genome from each of the11 genotypes, potentially with some polymorphism between the individuals, and the expected SNP between the individuals for a variant that represented a single individual should occur at a frequency of 9.09% in the bulk. Due to the potential bias introduced in the preparation of samples for the Illumina platform and any minor inaccuracy in the pooling of the genomic DNA quality it was assumed that there would be a normal distribution of SNP frequencies around this 9.09% level for an individual variant.

Although there should be limited intra-individual variation present in the chloroplast data, it is reported that massively parallel sequencing of genomic DNA which is then assembled against a chloroplast reference will align portions of homologous sequence from the nuclear genome in the Chloroplast assembly. These homologous genomic fragments would have an increased level of polymorphism relative to chloroplast genomic fragments, however not these differences would not necessarily be enough to for these sequences to be excluded from the alignment and hence SNP analysis of the assembly would identify these homologous polymorphisms as SNP with very low frequencies (Nock et al 2010). In order to capture the maximum number of the true chloroplast SNP while excluding the homologous nuclear genomic SNP from the analysis, 5% was determined to be the lowest acceptable frequency for a SNP in the chloroplast analysis.

Individual plant analysis determined a SNP density of 0.64 SNP/1000bp for the chloroplast. This analysis supports the application of this 5% limit and indentifies that the increase in SNP density established for the pooled samples 9.59 SNP/1000bp for Site 29 and 7.46 SNP/1000bp for Site 37, was indeed due to variation between the individuals in the pool rather than within the individuals or from contamination from homologous nuclear genomic reads. This indicates a statistically significant increase in the SNP diversity of the chloroplast was observed at the hot, dry coastal site, 29. It is noteworthy that although the two collections sites were geographically isolated and had very different environmental conditions, the greater proportion of the SNP (566) were indentified at both sites, ie were common SNP, with only 231 and 54 unique SNP at sites 29 and 37 respectively.

Diversity in ribosomal RNA genes

M. stipoides is tetraploid with a base chromosome number of 12, and estimated genome size of 880 Mbp. Using a whole genome shotgun approach for a pool of 11 individuals with such a large genome size, at the time of this experiment, would result in a theoretical coverage of 0.33X for each individual or ~3X coverage of the nuclear single copy genes. After alignment back to the nuclear genes sequencing actually resulted in incomplete coverage of all nuclear genes with the exception of the nuclear ribosomal genes as they exist in the genome in 1000's of tandem arrays and hence are represented in very high copy number (Small et al 2004).

A single copy of a nuclear allele would be expected at minimum of around 2.3% frequency in a bulk of 11 potentially heterozygous tetraploid individuals An awareness of the limitations in the quantitative assembly and analysis of bulk samples, lead to the application of a qualitythreshold where only SNP with a frequency of 1% or greater where analysed from the nuclear genome data. SNP with these frequencies are likely to be well above the frequency of sequencing errors and below the level that would exclude a single allele in one individual in the bulk.

The two sites had 40.3 and 40.6 SNP per 1000bp, respectively in the ribosomal RNA gene sequence analysed, and therefore no significant difference was identified between sites. This suggests diversity between individuals (6.5SNP per 1000bp for a single plant due to variation between the large number of copies of these genes) but indicates no difference in diversity in the nuclear genome between locations.

Discussion

Diversity at the Isa locus, especially that which induces a SNP resulting in a change to the sequence of the protein, is likely to be significant in plant defence. The protein encoded by this gene has been shown to be expressed in the pericarp of the seed (Furtado and Henry 2003) and may protect the seed from bacteria and fungi.

Wild barley populations had greater diversity at this locus in populations found in drier locations (Cronin et al 2007) and this may reflect the greater diversity of the drier environments especially in the diversity of biotic stress in these environments. These observations imply that adaptation to dry environments may require greater diversity at loci protecting the plant from the greater biological diversity of diseases in these environments. Wetter environments may support higher pathogen populations but they may be less diverse in the more uniform and favourable environments where water is not so limiting.

This study has shown that in *M. stipoides* (a distant relative of rice from Australia) is also observed to exhibit greater diversity at this locus in dryer environments. This suggests that the observation of sequence diversity variation in wild barley populations may reflect those likely to be found throughout the grasses.

The use of a wild relative of rice, a species with a reference genome sequence, has facilitated the analysis of variation at the wider genome level. The lack of differences in diversity in the in the ribosomal genes as representatives of the nuclear genome suggest that this adaptive variation

may be restricted to specific genes and does not necessarily reflect a generally greater diversity across the genome in drier environments.

Student training

The project successfully recruited a Masters student in April 2009, It was determined that the breadth of this project would support the development of a Masters thesis, rather than Honours as originally planned, without significant alteration of the project scope. The student is expected to submit a thesis in early 2011.

Publications

A major journal publication reporting the findings of this study in greater depth is being prepared for publication.

References

Cronin, JK, Bundock, PC, Henry, RJ and Nevo, E (2007). Adaptive climatic molecular evolution in wild barley at the Isa defense locus. *Proceedings of the National Academy of Science U S A* 104: 2773-2778.

Furtado A, Henry RJ and Meech SB (2003). The promoter of the asi gene directs expression in the maternal tissues of the seed in transgenic barley. *Plant Molecular Biology* 52: 787-799.

Nock C, Waters DLE, Edwards MA, Bowen S, Rice N, Cordeiro GM and Henry RJ (2010). Chloroplast genome sequence from total DNA for plant identification. *Plant Biotechnology Journal.* Published on line August, *doi: 10.1111/j.1467-7652.2010.00558.x*

Small, R. L., Cronn, R. C. and Wendel, F. (2004). L.A.S. Johnson Review No. 2: Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**: 145-170.

Whalley, R.B.D. and Brown, R.W. (1973). A method for the collection and transport of native grasses from the field to the glasshouse. *Journal of Rangeland Management* 26(5), 376-77.