Population structure in wild emmer wheat (*Triticum dicoccoides*) based on EST-SSR markers

Emebiri L[1], Ogbonnaya F[2] and Moody D[1]
[1] ***Biosciences Research Division, Department of Primary Industries, Grains Innovation Park, Natimuk, Horsham Vic. 3401, Australia***
[2] ***International Center for Agricultural Research in the Dry Areas (ICARDA), PO Box 5466, Aleppo, Syrian Arab Republic.***

INTRODUCTION

A high strategic priority for practical cereal improvement worldwide is to enrich the cultivated gene pools by incorporating favourable alleles, genes or gene complexes from wild relatives (Feuillet *et al*. 2007). Wild emmer wheat (*Triticum dicoccoides*; 2n=4x=28; genome AABB), the tetraploid progenitor of cultivated durum and bread wheat, shows considerable variability across its range in Israel (Nevo *et al*. 1982). Many alleles at its loci have been shown to be associated with resistance to environmental stress (Peleg *et al*. 2008) and a long list of agronomic traits (see review by Xie and Nevo 2008). Hence, it represents a valuable source of allelic variation for improving the yield stability, quality and important agronomic traits in wheat, either by producing synthetics (Mujeeb-Kazi *et al*. 2008) or by the technique of advanced backcross QTL (Tanksley *et al*. 1996).

In the present study, we analysed the population structure of 149 accessions of wild emmer wheat, based on EST-SSR markers, as a first step towards choosing parents for production of synthetics. Due to the limited degree of out-crossing in wild emmer wheat, we hypothesised a high level of linkage disequilibrium, low polymorphism (due to limited recombination rates), and high population subdivision (structure). Here, we tested these predictions by quantifying levels of molecular diversity and population structure across EST-SSR loci in the sample collection of wild emmer wheat. For comparison, we also analysed a population of nine cultivated durum varieties and two *Aegilop tauschii* accessions.

MATERIALS AND METHODS

Plant materials used for the study comprised 149 accessions of wild emmer wheat, obtained through the CIMMYT-Australian Germplasm Evaluation (CAGE) suite of projects (http://mendel.lafs.uq.edu.au:8080/ICIS5/GWIS_SYNT.htm). In addition, 9 durum (*Triticum turgidum durum;* AABB genome) cultivars and 2 accessions of *Aegilops tauschii* (DD genome) were also included in the study. Genomic DNA was extracted from single plants of each material, using the standard phenol/chloroform method as described by Martin *et al*. (2004).

EST-SSR Analysis: Seven EST-SSR primers, derived from durum and wild emmer wheat EST sequences (Eujayl *et al*. 2002; Peng *et al*. 2005), were used for genotyping. The microsatellite regions were amplified by PCR with flurorescent-labeled primers as described by Imtiaz *et al*. (2008), and size-separated on ABI 3730 automatic DNA sequencer (Applied Biosystems). The DNA fragments were sized automatically and assigned to specific alleles based on binning a range of sizes (± 0.5 base pair), as determined by GeneMarker (SoftGenetics LLC) software using the local Southern algorithm (Elder and Southern 1987).

Diversity Analysis: Alleles were recorded as co-dominant to avoid potential loss of information and allow accurate assessment of true genetic relationships. Basic statistics such as observed heterozygosity, gene diversity (or expected heterozygosity), allele richness, alleles per locus and polymorphism information content (PIC) were calculated per-locus using PowerMarker *V3.25* (Liu and Muse 2005).

Population structure analysis: We evaluated population structure by inspection of a neighbour joining dendrogram built from a genetic distance, which was calculated by the method introduced by Peakall *et al*. (1995) for codominant markers. We used MEGA 4 (2007) to construct the phylogenetic tree.

Conserved linkage between functional alleles: Linkage between pairs of polymorphic EST loci mapped on different chromosomes was evaluated using the software package TASSEL (http://www2.maizegenetics.net/). LD was estimated by squared allele-frequency correlations (r^2).

RESULTS AND DISCUSSIONS

Diversity statistics: A total of 188 genotypes were found in the population of wild emmer wheat accessions, based on different combinations of the 129 alleles generated by 7 primer pairs (Table 1). Of the two sources of EST primers used, the DuPw developed by Eujayl *et al*. (2002) were the most polymorphic (PIC = 0.78), and generated cleaner PCR products that were easily resolved as single bands. Across EST-SSR loci, global gene diversity in the wild emmer accessions

Table 1. Summary statistics of genetic diversity in wild emmer wheat, based on EST-SSR markers.

Locus	No. of Genotypes	No. of Alleles	Nei's unbiased Gene Diversity	PIC
cwem12C	23	22	0.70	0.68
cwem34g1	16	11	0.54	0.52
cwem34g2	18	7	0.75	0.71
cwem14B	16	10	0.60	0.57
cwem38D1	7	6	0.52	0.47
cwem38D2	7	5	0.66	0.61
DuPw004	47	29	0.82	0.81
DuPw038	30	20	0.87	0.86
DuPw023	24	19	0.71	0.68
Average	*21*	*14*	*0.69*	*0.66*

Table 2. Summary statistics of population genetic parameters for each group of germplasm.

Wheat Germplasm Group	No. of accessions	Nei's unbiased Gene Diversity	Alleles per Locus	PIC
Wild emmer	144	0.69	14.56	0.66
Durum	9	0.62	4.56	0.55
Ae. tauschii	2	0.74	2.56	0.46

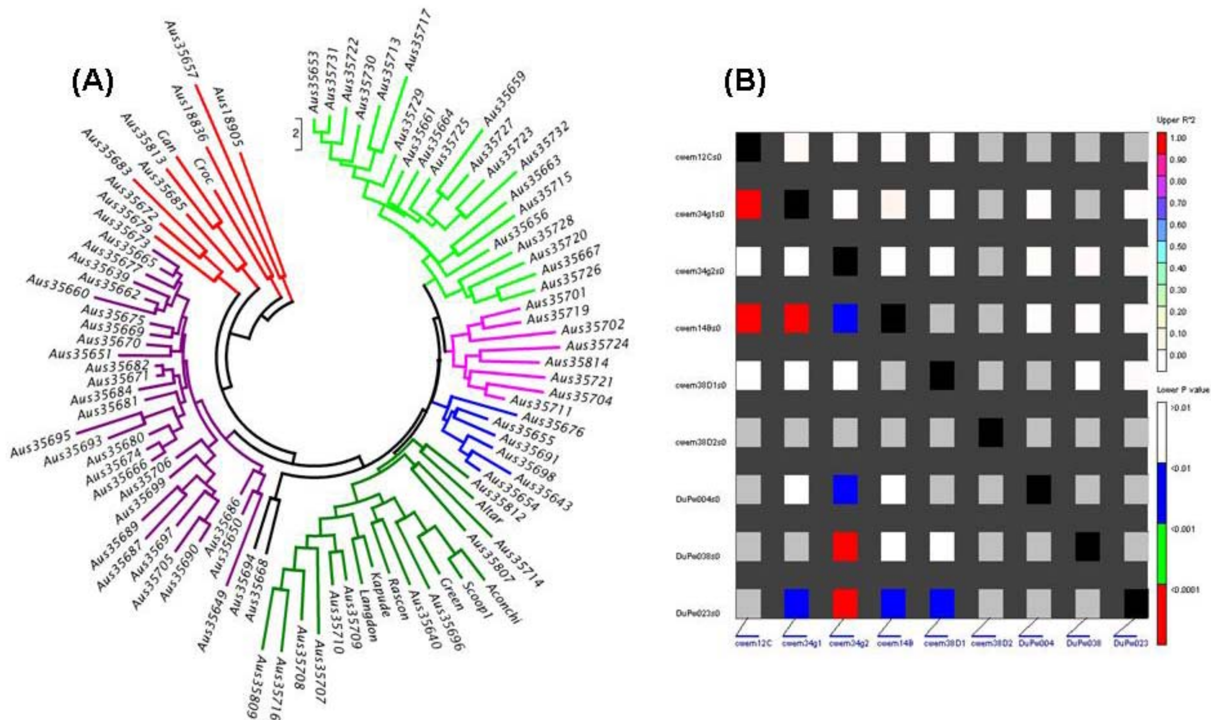

Figure 1. (A) Neighbour-Joining tree of individual plants in 160 accessions of wheat diverse germplasm, and (b) Pattern of inter-chromosomal pair-wise associations, based on 9 EST-SSR markers.

averaged 0.69 (SD = 0.04), while mean number of alleles per locus was 14.6 (SD = 8.7). These values are much higher than those reported by Nevo (2006) and Luo *et al*. (2007), based on allozyme and RFLP analyses respectively, but showed good agreement to those reported by Fahima *et al*. (2002) and Peleg *et al*. (2008).

Comparison of allelic diversity: A comparison of gene diversity and degree of polymorphism in the different wheat accessions is shown in Table 2. Expected heterozygosity was much higher than the observed in the *Ae*. *tauschii* accessions, probably due to the small number of accessions examined. Gene diversity of the wild emmer wheat accessions was comparable to that of the cultivated durum, but allelic diversity was higher in the wild emmer wheat accessions (Table 2). On average, the number of alleles per locus in the wild emmer wheat accessions was 14.6, compared to 4.6 in the cultivated durum cultivars.

We used a phylogenetic tree, constructed for all 160 wheat germplasm accessions (Figure 1a), to identify clusters of genetically similar lines. The tree showed good agreement with results (not shown) from a model-based approach suggested by Pritchard *et al*. (2000), and supported the hypothesis of high-level population subdivision. The accessions were grouped into 13 main clusters, ranging in size from 2 to 35 lines per group (Figure 1a). Conserved linkage between functional alleles was assessed at the inter-chromosomal level, and of the 36 pairs of loci evaluated, 28% were in tight linkage ($r^2 > 0.7$) with $P < 0.01$ in the wild emmer accessions (Figure 1b). This is higher than the level observed linkage disequilibrium in maize (10%) and sorghum (8.7%), and is consistent with the report of non-random distribution of molecular markers in *T. dicoccoides* (Peng *et al*. 2000).

CONCLUSIONS

The collection of wild emmer wheat examined in this study demonstrated a high level of molecular divergence and significant inter-chromosomal linkage disequilibrium. These findings suggest conserved gene clusters, and are supportive of a targeted approach towards exploitation of this resource for useful genes. One avenue of our continuing research is to identify a core set of accessions suitable for producing synthetics with maximal diversity and subsequent analysis of the effect of genetic background.

REFERENCES

Elder JK & Southern EM (1987) Computer-aided analysis of one-dimensional restriction fragments gels. p. 165-172

Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2002) Theor Appl Genet, 104: 399-407.

Feuillet C, Langridge P & Waugh R (2007) Trends in Genetic, 24: 24-32.

Imtiaz M, Ogbonnaya FC, Oman J & van Ginkel M (2008). Genetics, 178: 1725-1736.

Liu K & Muse SV (2005) Bioinformatics, 21: 2128-2129.

Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG & Dvorak J (2007) Theor Appl Genet, 114: 947-959.

Malysheva-Otto LV, Ganal MW & Röder MS (2006) BMC Genetics, 2006, 7.

Martin EM, Eastwood RF & Ogbonnaya FC (2004) Aust J Agric Res, 55: 1205-1211.

Mujeeb-Kazi A, Alvina Gul A, Farooq M, *et al*. (2008) Aust J Agric Res, 59: 391-398.

Nevo E, Golenberg EM, Beiles A, Brown AHD & Zohary D (1982) Theor Appl Genet, 62: 241-254.

Peakall R, Smouse PE & Huff DR (1995) Mol Ecol, 4: 135-147.

Peleg Z, Fahima T, Abbo S, Krugman T Saranga Y (2008) Genome, 5: 187-195

Peleg Z, Saranga Y, Krugman T, Abbo S, Nevo E & Fahima T (2008) Plant Cell & Environ, 31: 39-49.

Peng JH, Korol AB, Fahima T, Roder MS, Ronin Y.I, Li Y C & Nevo E (2000) Genome Res. 10: 1509-1531.

Peng JH, Nora L, Lapitan V (2005) Funct Integr Genomics, 5: 80–96.

Pritchard JK, Stephens M & Donnelly P (2000) Genetics, 155: 945-959.

Tamura K, Dudley J, Nei M & Kumar S (2007) Mol Biol & Evol, 24: 1596-1599.

Tanksley SD, Grandillo S, Fulton TM, Zamir D, *et al*. (1996) Theor Appl Genet, 92: 213-24.

Xie W & Nevo E (2008) Euphytica, DOI 10.1007/s10681-008-9703-8.