7th International Safflower Conference
WAGGA WAGGA AUSTRALIA

Diversity Arrays Technology (DArT): A generic high-density genotyping platform

Peter Wenzl*, Eric Huttner, Jason Carling, Ling Xia, Hélène Blois, Vanessa Caig, Kasia Heller-Uszynska, Damian Jaccoud, Colleen Hopper, Gosia Aschenbrenner-Kilian, Margaret Evers, Puthik Hok, Michael Duncan, Karl Miller, Grzegorz Uszynski, Andrzej Kilian
Diversity Arrays P/L, PO Box 7141 Yarralumla, Canberra, ACT 2600, Australia;
*peter@DiversityArrays.com

Abstract
Diversity Arrays Technology (DArT) is a sequence-independent genotyping method that generates genome-wide genetic fingerprints. It was conceived to enable high-throughput genotyping of plant species irrespective of the level of previous investments in genomics resources for a species. Here we provide an overview of the technology and some of its most frequent applications.

Keywords: molecular markers – genotyping – high throughput – microarrays

Introduction

Marker technologies are undergoing a transition from predominantly serial assays that measure the size of DNA fragments to hybridization-based assays with high multiplexing levels [8,20]. Two hybridization-based technologies have emerged during the last years: SNP (Single Nucleotide Polymorphisms) [5] and DArT (Diversity Arrays Technology) [12,33]. In contrast to current SNP technologies, DArT performs well in polyploid species such as wheat [1], banana [15] or sugarcane [10], does not require any existing DNA-sequence information [12] and can be deployed with a fraction of the resources required for SNP platforms (at least an order of magnitude less). Here we present an overview of the technology and some of its applications.

Technology Features

The principle
DArT reveals DNA polymorphism by querying representations of genomic DNA samples for the presence/absence of individual fragments (Figure 1).

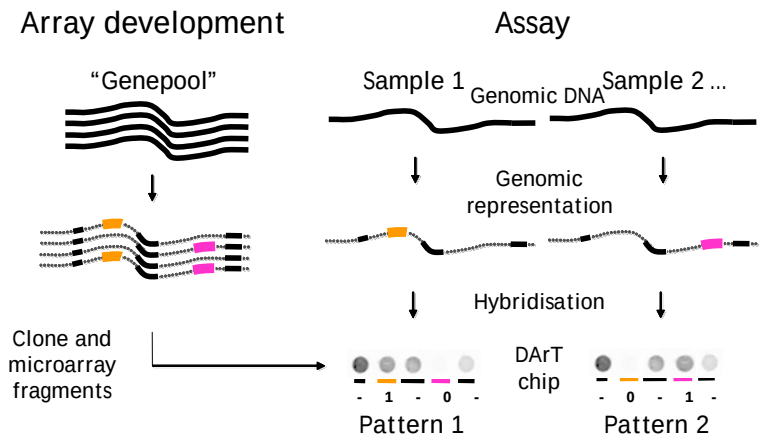

Figure 1: Principles of DArT array development and routine genotyping.

A genomic representation is the product of any method that reproducibly captures a defined set of DNA fragments from genomic DNA. For example, genomic DNA is digested with *Pst*I and a second restriction enzyme such as *Taq*I, adapters are ligated to *Pst*I ends and short adapter-ligated fragments are amplified [1,16,33]. SNP and insertion/deletion (InDel) polymorphisms at (or between) restriction-enzyme sites determine whether or not individual fragments are captured in the representation of a particular genotype [37].

The composition of the representation of an unknown sample is decoded through hybridisation to a microarray. This array is built once for each species and contains representation fragments (probes) produced from a set of genotypes that cover the gene pool of the species. When these probes are amplified from cloned representation fragments, part of the polylinker region of the cloning vector is co-amplified. A DNA fragment complementary to this polylinker region ('reference DNA') is co-hybridised to the array to quantify, for each probe, the amount of DNA spotted on the array during routine assays (see 'DArT assay" section below). A marker (= representation fragment) is polymorphic if the relative hybridisation intensity across genotypes (arrays) falls into distinct clusters. Depending on the cluster into which it falls a marker is called 'present' or 'absent' for a particular genotype [12,16,33].

Genotyping assay
The assay (Figure 2) is usually performed in a dominant fashion, although a subset of markers could be scored hemi-dominantly [16]. The accuracy of the assay is approximately 99.8%, that is, approximately 1 error per 500 genotype calls [1,33]. The algorithm used to score DArT markers calculates an estimate of marker quality, which can be used to select marker subsets of varying quality stringencies for different applications [1,16,33].

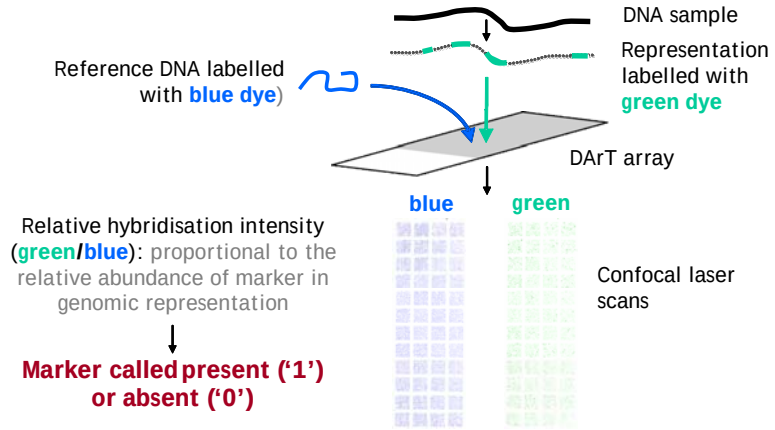

Figure 2: The current format of DArT assay.

Polyploidy
In contrast to other SNP-typing techniques, DArT performs well in species of virtually any ploidy level, including sugarcane (x = 5-14) [1,4,10,15,16,32]. The good performance is at least partly due to the fact that SNP detection in DArT is mediated by the high fidelity of restriction enzymes rather than primer annealing [1].

Marker distribution across the genome
The SNP and InDels surveyed by DArT are 'randomly' selected across the whole genome. However, the 'methylation filtration' effect arising from using *Pst*I (a methylation-sensitive restriction enzyme) partly enriches genomic representations for hypo-methylated 'gene space' regions (Figure 3).

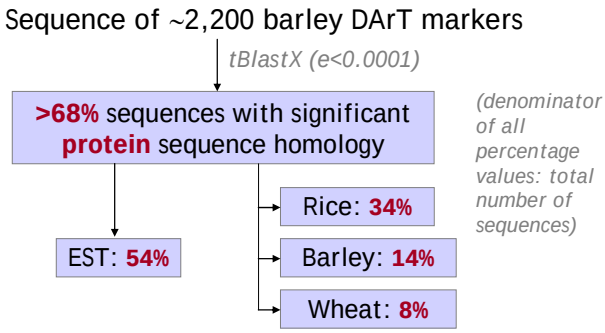

Figure 3: Analysis of barley DArT marker sequences.

The actively expressed, low-copy regions of a species' gene space tend to be located in distal chromosome areas, a pattern that is clearly reflected in the DArT-marker density along chromosome arms in barley (Figure 4) [34] Because the size of the 'gene space' varies much less than total genome size, DArT is fairly insensitive with respect to variations in genome size [16].

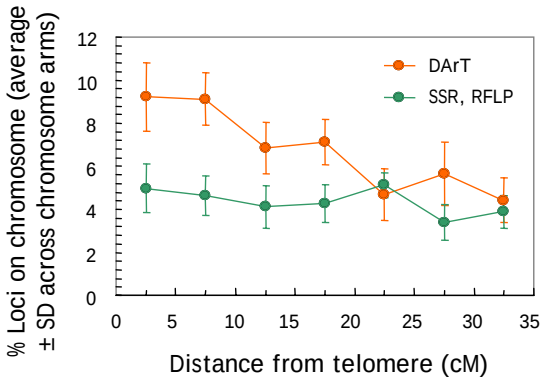

Figure 4: Average distribution of DArT marker along chromosome arms in barley [34].

.

Applications

Diversity studies

The large number of markers that are simultaneously assayed by DArT provide a high level of resolution in genetic-diversity studies (Figure 5) [1,2,4,10,13,15-17,22,32,33,36,38,39].

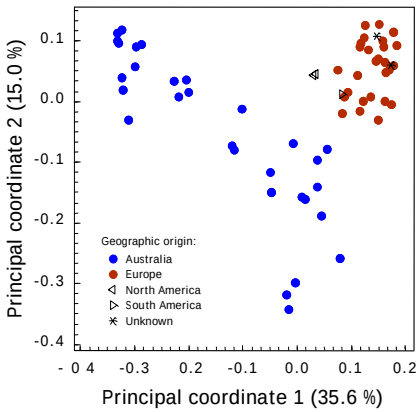

Figure 5: Diversity analysis of a set of bread wheat cultivars. The more diverse set of Australian cultivars are well separated from European and American cultivars [1].

Genetic-distance estimates derived with DArT are more likely to be accurate because the 'random' nature of DArT markers should reduce the ascertainment bias when compared to technologies relying on targeted marker development [20].

Genetic mapping

DArT markers are readily used to build genetic maps [1,9,16,21-23,26,30,34,37], which tend to be highly collinear across different populations (Figure 6). At the time of writing, approximately 250 mapping populations have been genotyped with DArT for wheat and barley alone.

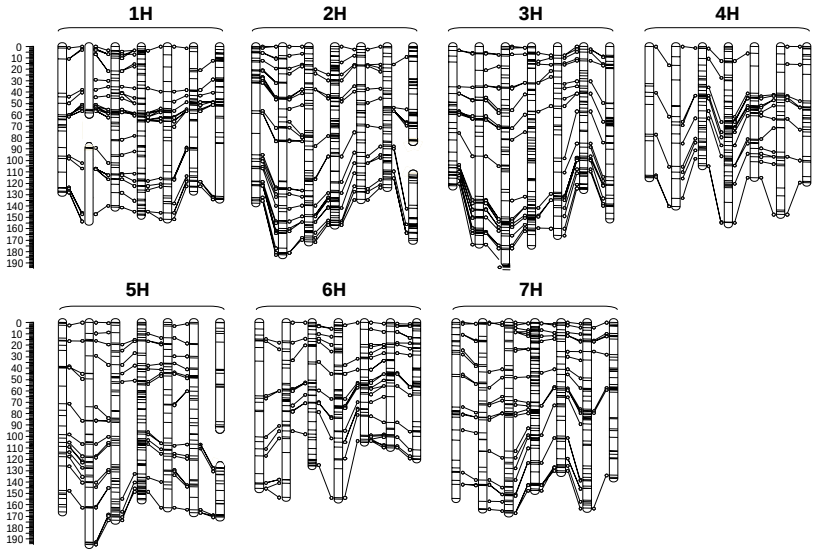

Figure 6: Colinearity of barley linkage maps independently built for seven different populations of doubled-haploids or recombinant-inbred lines [34].

Marker-trait associations

Given the speed with which genome-wide fingerprints can be generated, DArT markers are being extensively used for QTL and association-mapping studies [3,6,7,11,18,19,24,25,27, 28,29,40]. In recognition of the fact that many less-researched species lack genetic linkage maps, colleagues at National ICT Australia (NICTA) have developed a method for establishing marker-trait relationships that does not require a linkage map [3]. This method is based on statistical machine-learning (SML) algorithms that are used in areas outside of biology or agriculture. Figure 7 displays the results of an SML analysis of plant height in barley, in which the known map positions of individual markers were only used to display the SML results.

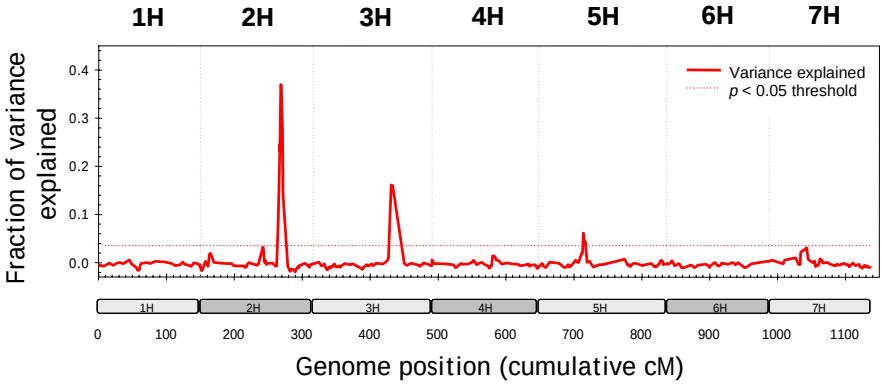

Figure 7: QTL analysis of plant height in barley using a novel approach based on statistical machine learning [3]. Map positions of markers were only used to display the results of the analysis.

Bulked segregant analysis

Although DArT markers are typically scored dominantly, the hybridisation intensities recorded in DArT assays are quantitative in nature because they reflect the abundance of individual DNA fragments in genomic representations. Given sufficient replication of the assay, the hybridisation intensities can thus be used to measure allele frequencies in DNA pools. One straightforward application of this property is quantitative Bulked Segregant Analysis (qBSA) [35].

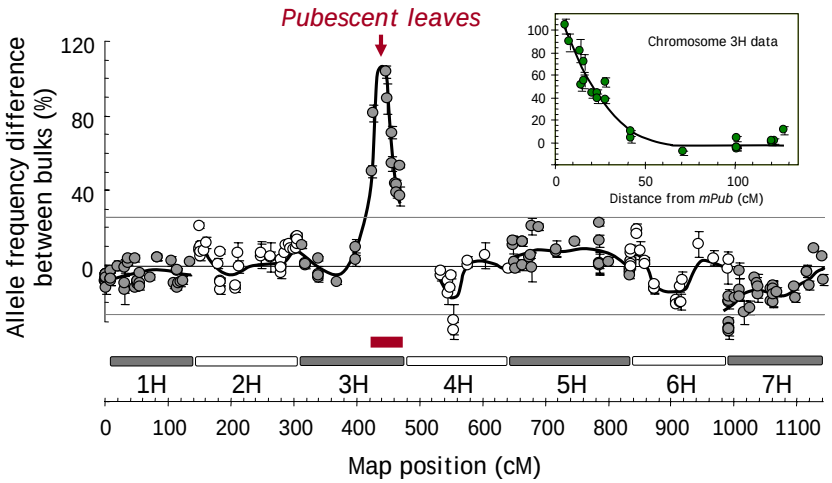

Figure 8: Quantitative Bulked Segregant Analysis identifies the pubescent leaves (*mPub*) locus in a population of 80 doubled haploids derived from a cross between two phenotypically contrasting barley cultivars [35].

In qBSA, the DNA pools from two phenotypically contrasting groups of segregants are hybridised to a single array. The hybridisation-signal difference (and hence allele-frequency difference) is 100% for markers in complete linkage disequilibrium with the target locus and decays in its vicinity in a monotonous fashion. The latter allows the estimation of the cM distance between individual markers and the targeted locus [35].

Marker-assisted selection

Once marker-trait relationships have been established, groups of DArT markers linked to loci of interest can be selected to establish a lower-plex assay on a different platform. This platform allows the cost-effective typing of several hundred markers for a large number of samples. Ideally, a set of markers spread throughout the genome is included. This simultaneously enables foreground selection for a defined group of loci and background selection for quality control purposes or other applications such as the rapid introgression of novel alleles into an adapted background.

Delivery of the Technology

The uptake of the technology is steadily increasing; the number of DNA samples genotyped on DArT arrays has approximately been doubling each year for the last few years. A DArT array for a new plant species can be developed in a very limited period of time. For example, more than 1,000 polymorphic markers were identified for the moss *Garovaglia* during a three-month project [13]. To this date DArT has not only been established for approximately 40 plant species, but also for two fungal pathogens, a bacterium and two animal species [14,31, unpublished data]. DArT was conceived as an open technology platform. We are working towards creating a network of like-minded people and organisations interested in deploying DArT in their research and breeding programs.

Acknowledgements
We thank our colleagues at the Value Added Wheat Cooperative Research Centre (VAWCRC) for their commitment to our past Triticarte P/L partnership and the Australian Grains Research and Development Corporation (GRDC) for ongoing financial support. We also thank our collaborators, former employees and students and a large number of visitors for their contributions.

References

Akbari et al. (2006) Diversity Arrays Technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113: 1409

Badea A, Eudes F, Graf RJ, Laroche A, Gaudet DA, Sadasivaiah RS (2008) Phenotypic and marker-assisted evaluation of spring and winter wheat germplasm for resistance to fusarium head blight. Euphytica, in press

Bedo J, Wenzl P, Kowalczyk A, Kilian A. Precision-mapping and statistical validation of quantitative trait loci by machine learning. BMC Genetics 9: 35

Bradeen JM, Iorizzo1 M, Gao L, Aversano R, Quirin EA, Carputo D (2008) Comparative structural genomics of the potato tertiary genepool: improving access to disease resistance genes. Plant & Animal Genome XVI Conference. San Diego, USA, 12-16 January

Chee et al. (1996) Accessing genetic information with high-density DNA arrays. Science 274: 610

Crossa J, Burgueño J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177: 1889-1913

Grewal TS, Rossnagel BG, Pozniak CJ, Scoles GJ (2008) Mapping quantitative trait loci associated with barley net blotch resistance. Theor Appl Genet 116: 529-539

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101: 5-18

Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglinton JK, Chalmers KJ (2007) A genetic map of 1,000 SSR and DArT markers in a wide barley cross. Theor Appl Genet 115: 383-39

Heller-Uszynska et al. (2006) Diversity Arrays Technology (DArT) for high throughput, whole-genome molecular analysis in sugarcane. Tropical Crops Biotechnology Conference. Cairns, Australia, 16-19 August

Huynh BL, Wallwork H, Stangoulis JCR, Graham RD, Willsmore KL, Olson St, Mather DE (2008) Quantitative trait loci for grain fructan concentration in wheat (*Triticum aestivum* L.). Theor Appl Genet117: 701-709

Jaccoud et al. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29: e25

James KE, Schneider H, Ansell SW, Evers M, Robba L, Uszynski G, Pedersen N, Newton AE, Russel SJ, Vogel JC, Kilian A (2008) Diversity Arrays Technology (DArT) for pan-genomic evolutionary studies of non-model organisms. PLoS ONE 3: e1682

Kema et al. (2007) Sequencing the major *Mycosphaerella* pathogens of wheat and banana. Plant & Animal Genome XV Conference. San Diego, USA, 13-17 January

Kilian (2007) Towards effective deployment of Diversity Arrays Technology (DArT) in banana genomics and sequencing. Plant & Animal Genome XV Conference. San Diego, USA, 13-17 January

Kilian et al. (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. *In* Tuberosa et al.: Proceedings of the International Congress In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution. Avenue Media, Bologna; pp 443-461

Lezar et al. (2004) Development and assessment of microarray-based DNA fingerprinting in *Eucalyptus grandis*. Theoret Appl Genet 109: 1329

Li H, Vaillancourt R, Mendham N, Zhou M (2008) Comparative mapping of quantitative trait loci associated with waterlogging tolerance in barley (*Hordeum vulgare* L.). BMC Genomics 9: 401

Lillemo M, Asalf B, Singh RP, Huerta-Espino J, Chen XM, He ZH, Bjørnstad A (2008) The adult plant rust resistance loci Lr34/Yr18 and Lr46/Yr29 are important determinants of partial resistance to powdery mildew in bread wheat line Saar. Theor Appl Genet, in press

Luikart et al. (2003) Molecular markers for population genomics and genome typing. Nature Rev Genet 4: 981

Mace ES, Rami JF, Bouchet S, Klein PE, Klein RR, Kilian A, Wenzl P, Xia L, Halloran K, Jordan DR. A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. Under review

Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9: 26

Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Huttner E, DeAmbrogio E, Kilian A. An integrated DArT-SSR linkage map of durum wheat. Molecular Breeding, in press

Ortiz RO (2008) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Plant & Animal Genome XVI Conference. San Diego, USA, 12-16 January

Parh DK, Jordan DR, Aitken EAB, Mace ES, Jun-ai P, McIntyre CL, Godwin ID (2008) QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117: 369-382

Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder MS, Kilian A, Korol AB, Fahima T (2008) High-density genetic map of durum wheat x wild emmer wheat based on SSR and DArT markers. Theor Appl Genet 117: 103-115

Pozniak CJ, Knox RE, Clarke FR, Clarke JM (2007) Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. Theor Appl Genet 114: 525-537

Rheault et al. (2007) Using DArT and SSR markers for QTL mapping of *Fusarium* head blight resistance in six-row barley. Plant & Animal Genome XV Conference. San Diego, USA, 13-17 January

Roy et al. (2006) Association mapping of disease resistance genes in wild barley. Plant & Animal Genome XIV Conference. San Diego, USA, 14-18 January

Semagn et al. (2006) Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49: 545

Sessitsch et al (2006) Diagnostic microbial microarrays in soil ecology. *New Phytologist* 171: 719

Tinker NA, Kilian A, Jackson E, The OAT DArT Consortium (2008) DArT marker development and applications in oat. Plant & Animal Genome XVI Conference. San Diego, USA, 12-16 January

Wenzl et al. (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA 101: 9915

Wenzl et al. (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7: 206

Wenzl et al. (2007) A DArT platform for quantitative bulked segregant analysis. BMC Genomics 8: 196

White J, Law JR, MacKay I, Chalmers KJ, Smith JS, Kilian A, Powell W (2008) The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor Appl Genet 116: 439-453

Wittenberg et al. (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Genet Genomics 274: 30

Xia et al. (2005) DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110: 1092

Yang et al. (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology. Theor Appl Genet 113: 585

Zhang L, Marchand S, Belzile F (2008) Genome-wide linkage disequilibrium analysis and association study for agronomic traits in canadian barley. Plant & Animal Genome XVI Conference. San Diego, USA, 12-16 January