Potential of DIVERSITY ARRAYS TECHNOLOGY (DArT) for rice authentication

Andrzej Kilian
a.kilian@DiversityArrays.com

Diversity Arrays Technology (DArT) Pty Ltd

Plant Authentication Workshop

Belgium

21-22 September 2009

Diversity Arrays Technology Pty Ltd

- Company established in 2001in Canberra, Australia
- Mission: improving the efficiency of using natural resources through modern genetic and information technologies
- Main business: genome profiling services and IT support for agriculture + consultancy to private and public organisations
- Global service, performed >1.4 G marker assays
- Mostly genetics and breeding, but increasingly involved in authentication for various industries Strong interest in international agriculture: training for >30 visitors and technology co-developers
- “Open access” and partnership-based business model (“franchise” platforms established in India and France and a new one in Brazil under negotiations)

Thanks to: **Eric Huttner, Peter Wenzl**, **Grzegorz Uszynski**, Kasia Heller-Uszynska,Jason Carling, Vanessa Caig, Margaret Evers, Damian Jaccoud, Ling Xia, Shiying Yang,Sujin Patarapuwadol, Pavan Lakkineni, Gosia Aschenbrenner-Kilian, Aurelie Bonin, Brent Thompson, Cyril Cayla PLUS many visitors and previous students

DArT: whole genome profiling

Features

- Detects all types of polymorphism (mostly SNP but also InDels, CNVs and methylation variation)
- Unbiased for the location of polymorphism ("candidate genes"), but highly enriched for genetically active regions of the genome
  - initial SNP approaches affected by a small number of genotypes used for discovery (ascertainment bias)
- Profiles the whole genome (thousands of loci) in a single assay on automated platforms but can be done as monoplex (low plex not effective for breeding or authentication)
- Fast and inexpensive technology development
  - no need for assay development or sequence information

Overcoming limitations of earlier technologies

- Throughput and costs (a few cents per datapoint)
- Reducing initial investment 10-100 fold
- Data integrity (automation of data acquisition and storage)

Utility for breeders/agriculture limited by lack of integration with other data types and efficient software to exploit the data AND slow breeders' operations change

DArT has been established for...

Plants

Arrays developed (2007)

- apple
- arabidopsis[1]
- bambara groundnut
- banana & plantain
- barley & wild relatives
- cassava
- chickpea
- coconut
- fern species
- festuca
- hops
- lupin
- moss species
- oats
- pigeonpea
- potato
- quinoa
- rice & wild relatives
- rye
- ryegrass
- sorghum
- sugarcane
- tomato
- wheat: bread, durum & wild relatives

Recently developed

- brassicas (several)
- capsicum
- cotton
- cucumber
- eucalypt
- Grapevine
- pine
- pumpkin
- Triticale
- Peanut
- Sweetpotato
- sugarbeet

In the 'pipeline'

- 12 new species

Microorganisms

- Mycosphaerella graminicola
- Mycosphaerella fijensis
- Salmonella[3]
- Ustilago scitaminea

Animals

- mosquito
- mouse

[1] Plant Research International
[2] Proof-of-concept by University of Pretoria; full development at DArT P/L
[3] ARC Seibersdorf

(DArT arrays for new species are usually developed in a collaborative spirit through consortia). From January 2010 ALL organisms are enabled

DArT - "Parallel reverse RFLP"

Array development

"Genepool"

Clone and microarray fragments

Marker discovery & Routine assays

Sample 1 Genomic DNA Sample 2 ...

Genomic representation

Hybridisation

DArT chip

- 1 - 0 -

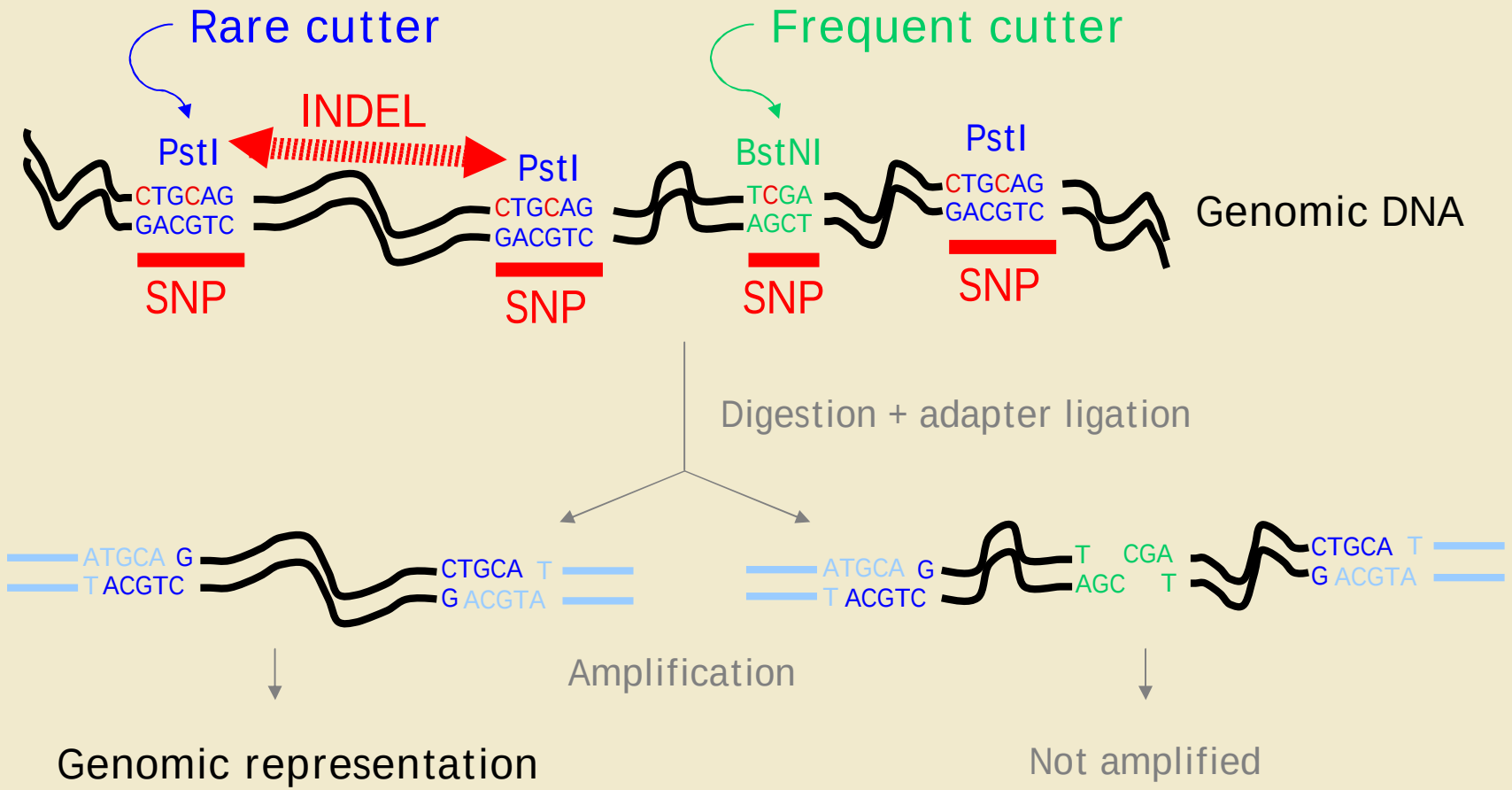
Pattern 1

- 0 - 1 -

Pattern 2

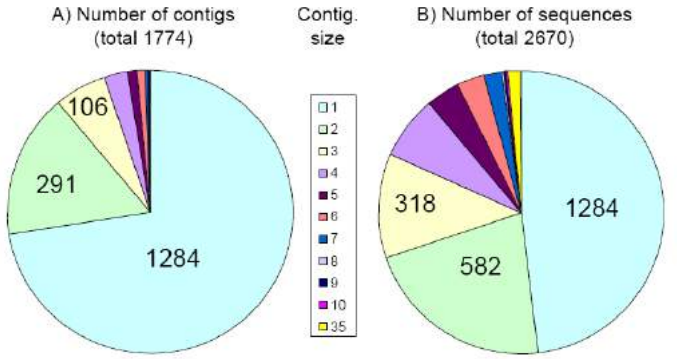
Technical details @ www.DiversityArrays.com

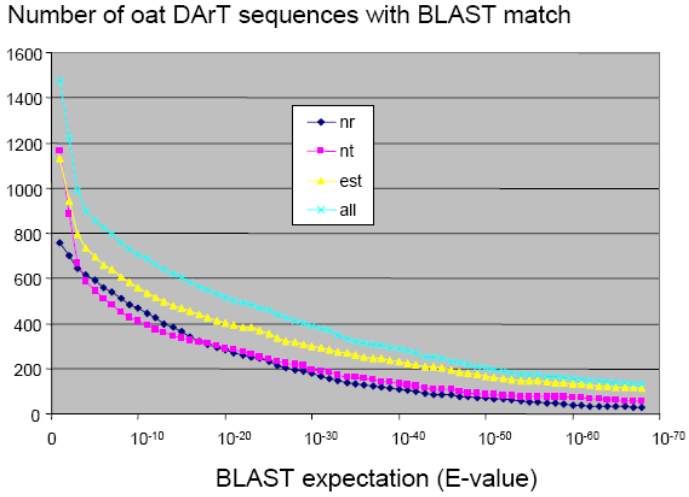
Example of complexity reduction method



DArT markers are in of “gene space”

- PstI enzyme: ‘methylation filtration’ effect
  - enriches for hypo-methylated, expressed genome regions and low copy sequences
  - makes DArT robust with respect to genome size
  - enables detection of epigenetic variants/”sports”
- High frequency (around 50%) of markers with high homology to “genes” in already processed species (e.g. barley, wheat, sugarcane, oat, sorghum, potato, eucalyptus, etc.)
- DArT deployed in seven genome sequencing projects mostly for linking HD genetic maps with sequence assemblies

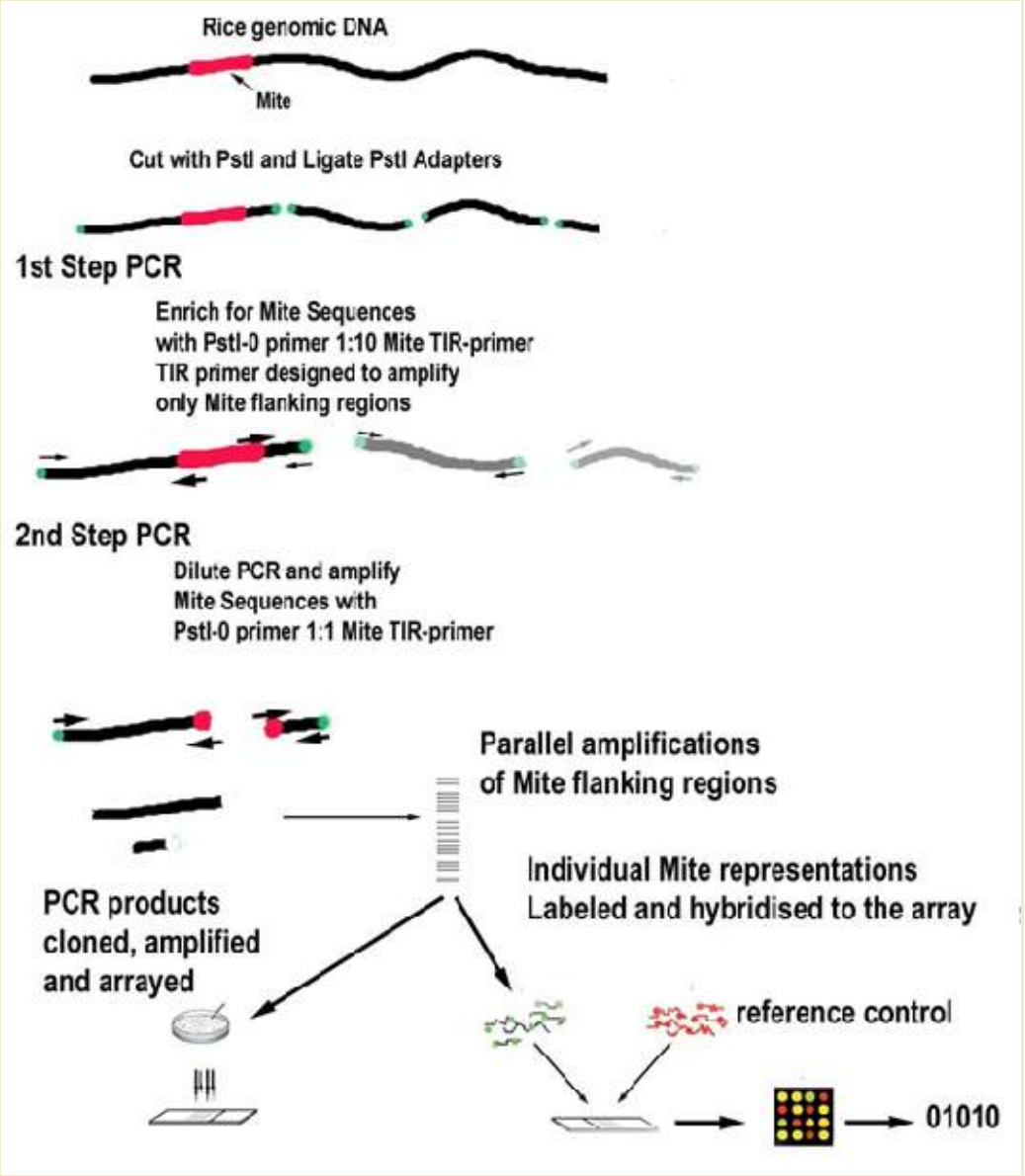




Tinker et al, 2009

Rice MITE-DArT method

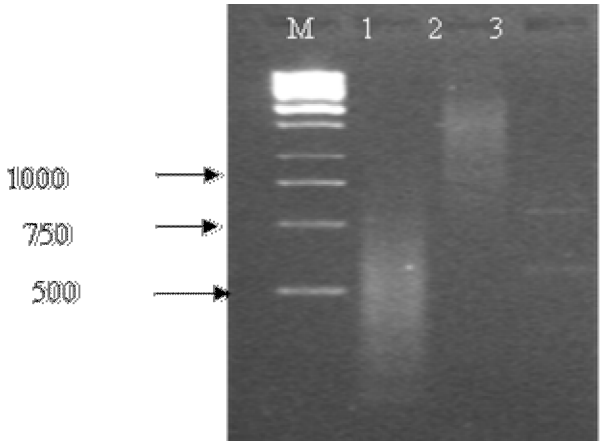
- Developed initially using rice genome (Peng et al) using Stowaway family and PstI RE.
- Many alternative MITE families tested and refined method established for rice – increased complexity (Patarapuwadol 2007)
- Adopted to many other organisms like sorghum, sugarcane, mosquito (Bonin et al, 2008) by simple change of MITE primer sequence and adjustment of complexity through selective base on Bsp1286I site
- Lack of methylation sensitivity on RE (Bsp1286I) – "strictly genetic" markers
- Requires some DNA sequence information



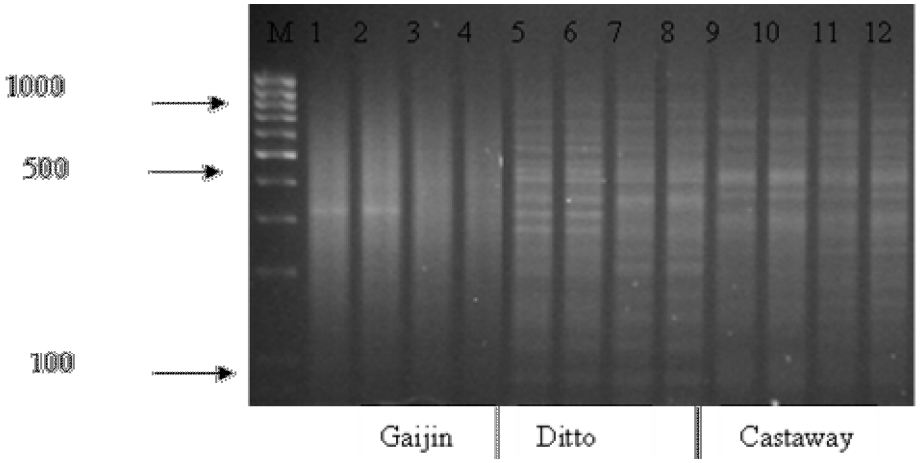
Optimisation of MITE-DArT for rice

Patarapuwadol 2007

- Replacement of PstI enzyme with Bsp1286I to increase representation size >10,000 fragments
- Selection of MITE-TIR with the largest representation and most heterodispersed amplicons
- Gaijin's good apparent performance verified by sequencing and mapping
- Smaller representations of Ditto and Castaway showing some banding patters

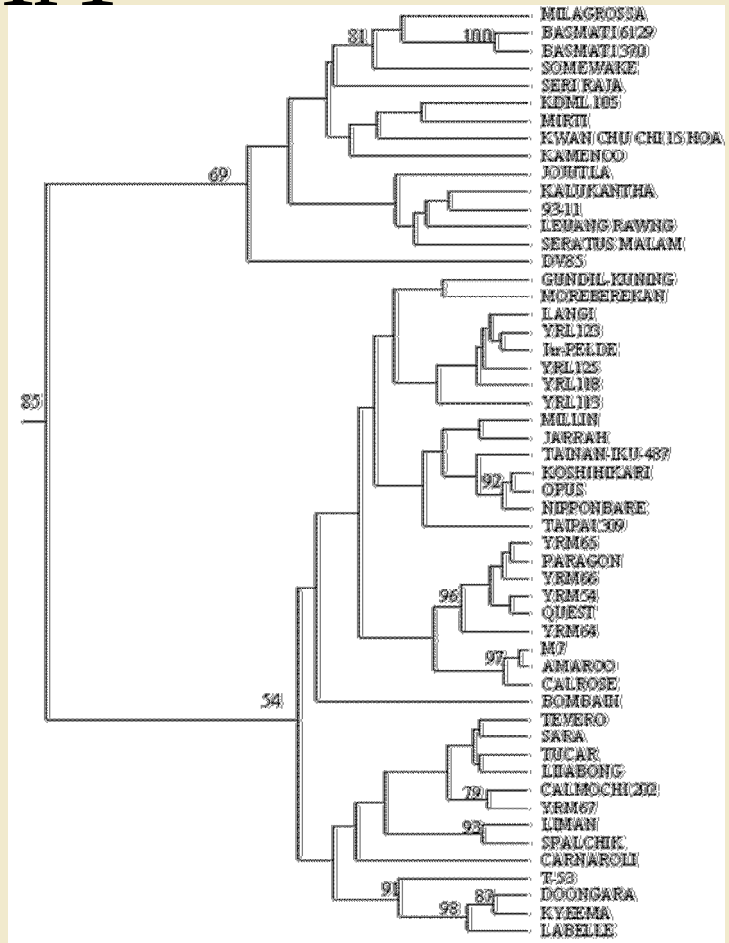


M - 1kb DNA Ladder marker PCR products of 93-11obtained with primer combination of Bsp1286I-Stowaway-TIR primer lane 1; BspI1286 primers - lane 2 and Stowaway-TIR primer- lane 3

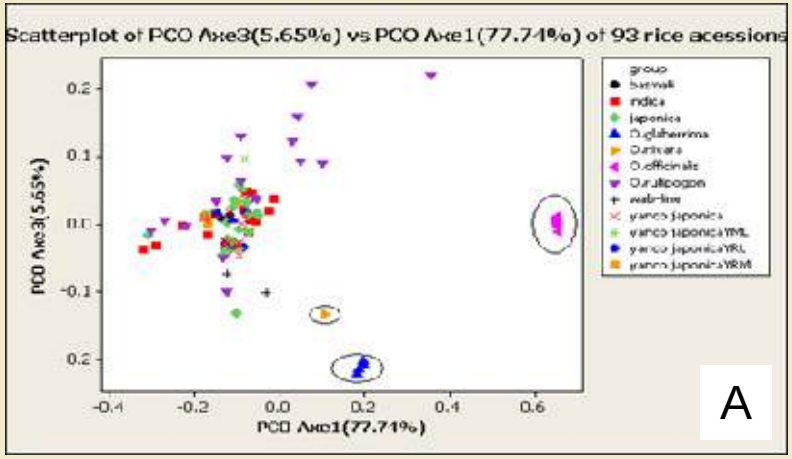


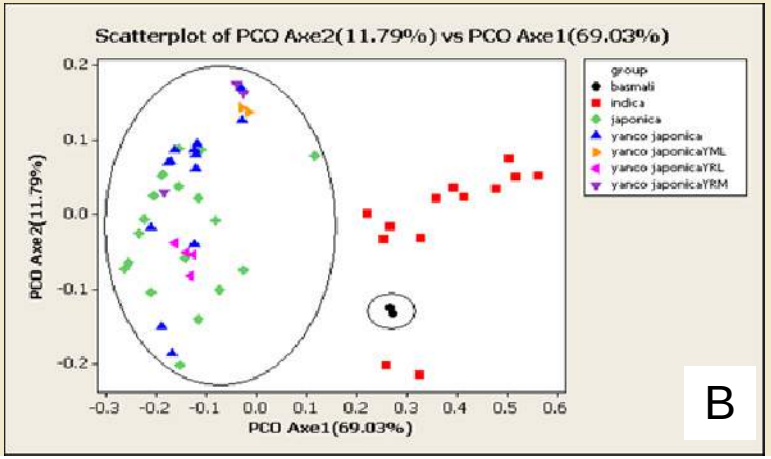
M - 100bp DNA Ladder. Two independent amplicons form 93-11 and Nipponbare respectively. Primer combinations: Bsp1286I and MITE-TIRs from Gaijin (1-4), Ditto (5-8) and Castaway (9-12).

Rice genetic diversity revealed by MITE-DArT



UPGMA dendogram constructed form 252 MITE-DArT markers of a group of cultivated accessions. Some of them are a landrace accessions. For example Jojutla, DV85 and Seri Raja originate from Mexico, Bangladesh and Malaysia respectively. (Patarapuwadol 2007)

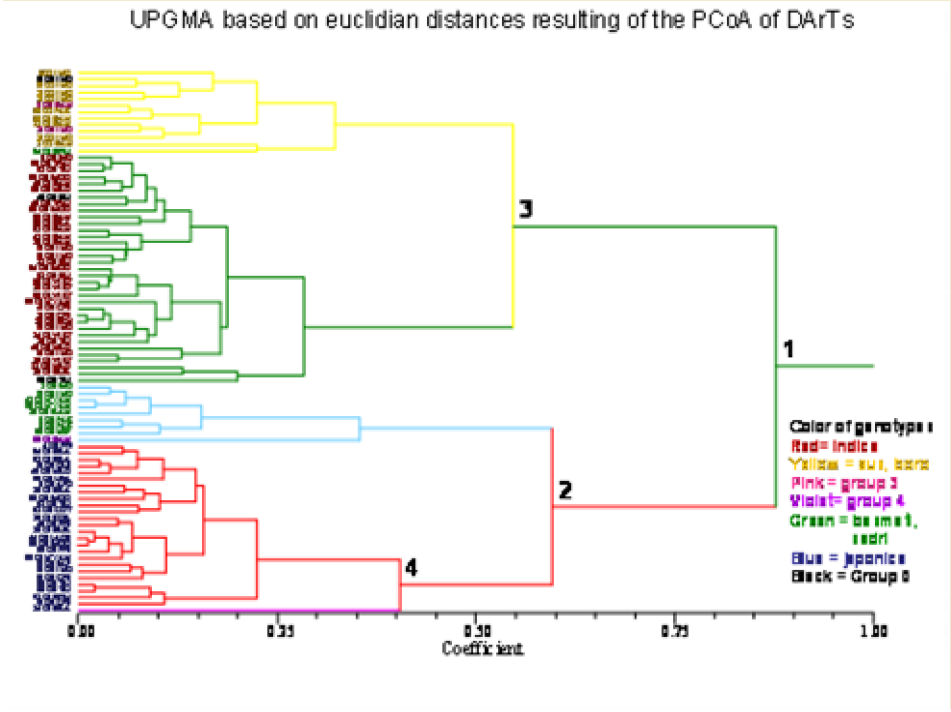


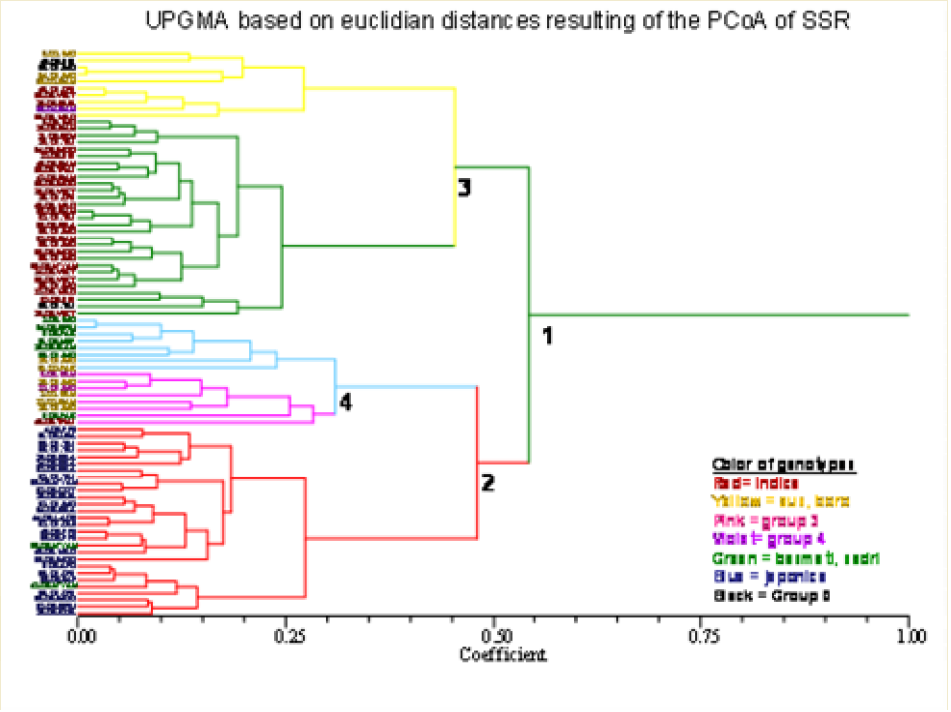


Principal coordinates analysis (PCO) using Nei& Li distance matrix computed from MITE-DArT markers form 93 rice accession (A), Cultivated rice accessions of O.sativa (B) (Patarapuwadol, 2007)

Comparison of DArT and SSR Hurtado et al, in prep.

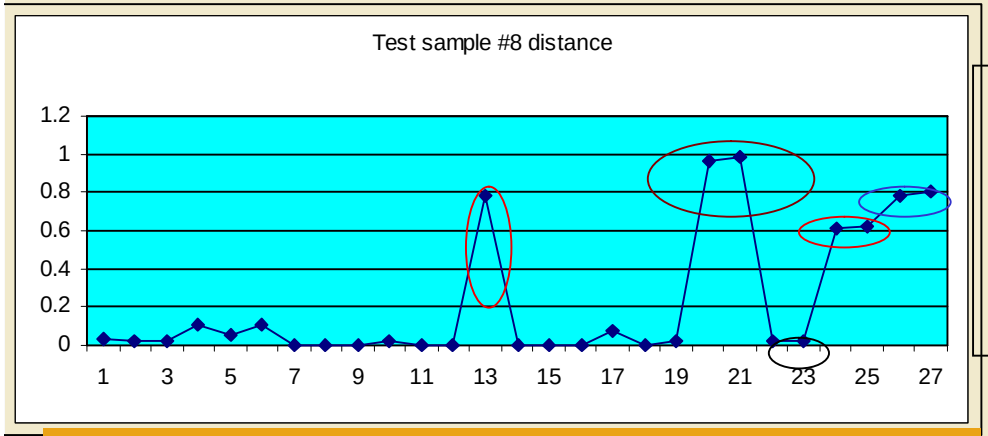
- DArT tree reflected very clearly the organization of the accession set into isozymic groups and fits “exceptionally well” other than SSR markers
- Country of origin separation in most groups revealed by DArT markers
- SSRs did not fit well with what was observed with other marker types, neither with other studies conducted with SSR
- Rapid evolution and high likelihood of scoring ad/or tracking errors?





Authentication tests in rice

- Tracking down planting mistakes in foundation seed production
  - High level of phenotypic similarity increasingly strong impediment in “classical” detection
- Detecting adulterations in the product supply chain
  - High level of precision of DArT analysis verified by “reconstruction” experiment
  - Detection limit not tested, but 5% admixture easy on current platform
  - Cultivar heterogeneity likely upper limit for detecting “admixture”
  - Sourcing DNA template from polished rice grain presents a manageable challenge

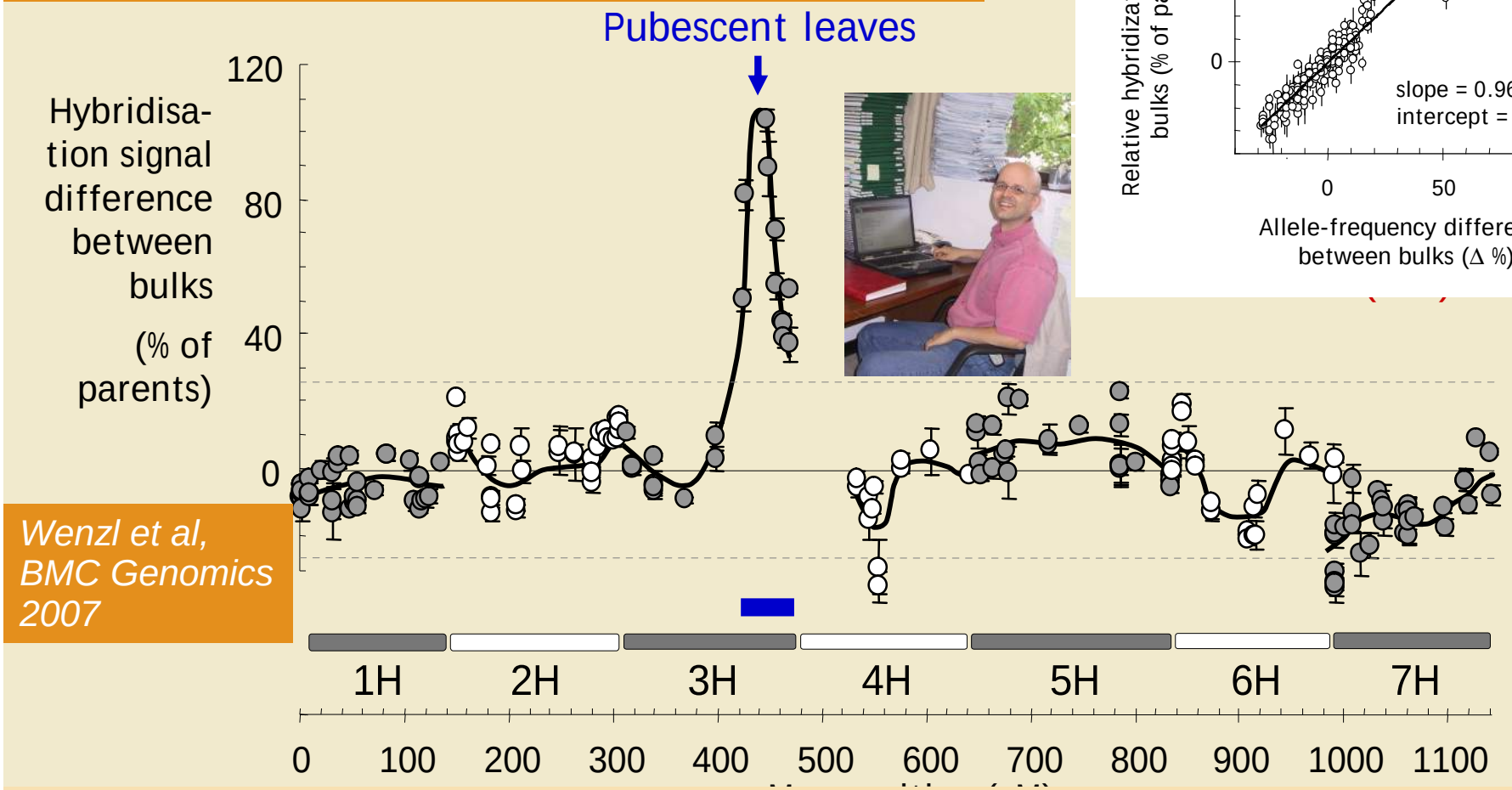


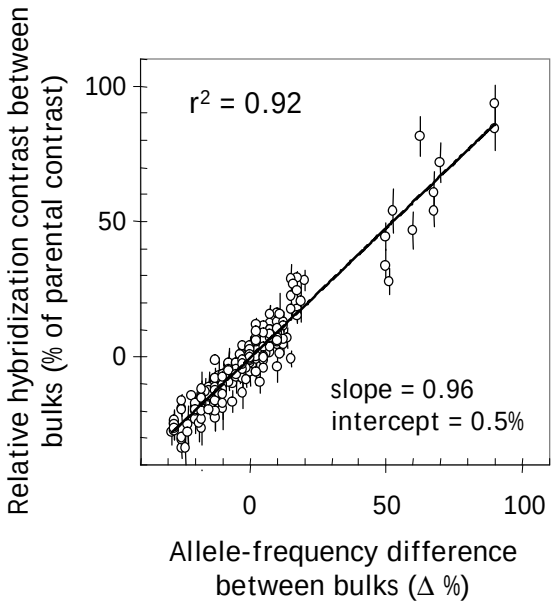
Humming distance of a test sample from other test samples and controls

1-19	test samples
20-21	suspected contaminant control
22-23	“pure” control
24-25	20% contamination reconstruction
26-27	50% contamination reconstruction

Bulked segregant analysis

Rapid and cost effective marker-trait association tool





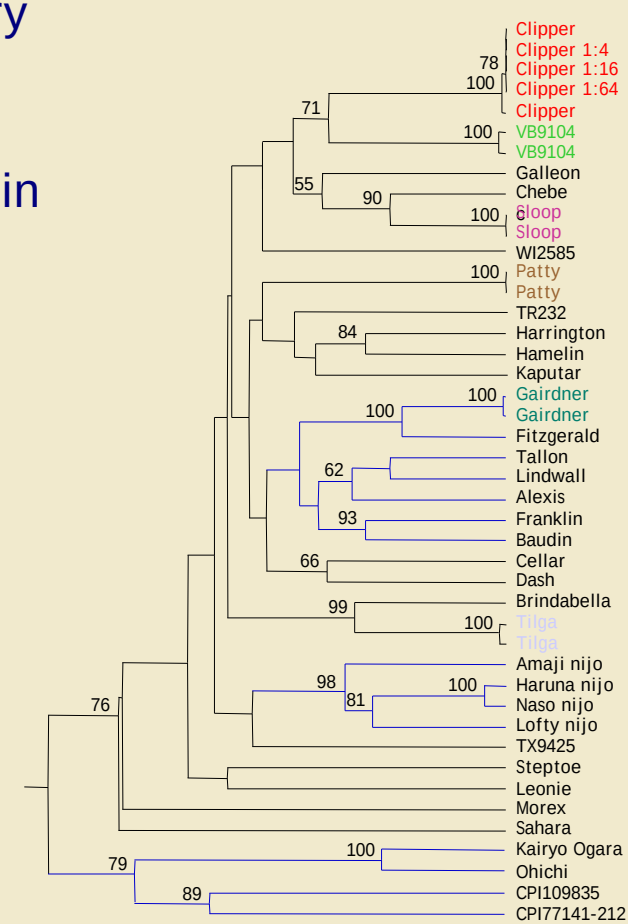
Wenzl et al, BMC Genomics 2007

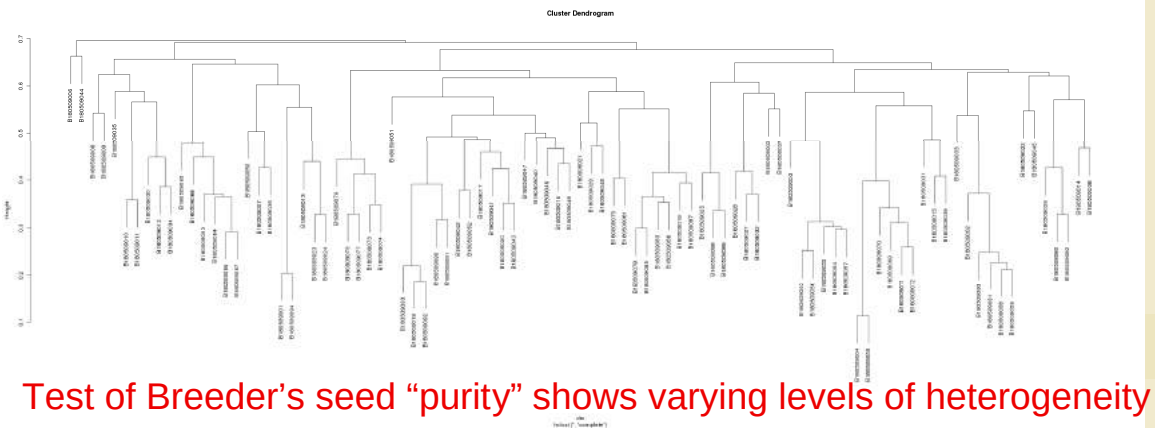
DArT assays is fully quantitative enabling estimation of the distance of linked markers from the targeted locus and precise allele frequency determination including admixture detection, estimation and likely “deconvolution”

Barley: first large genome with DArT expanding to purity testing applications

- Optimisation of complexity reduction method for plants with complex genomes
- Heterogeneity detected in many cultivars even with very limited genotype sampling in first experiments
- Heterogeneity present at all level of seed certification process (breeder, foundation, certified) and often high in production system
- Initiating industry commissioned project on PBRed varieties heterogeneity evaluation, seed purity testing (<0.5% contamination limit!)
- Export requirements and trading between farmers ("declared seed"?) - end point royalty system

Wenzl et al, PNAS, 2004





Test of Breeder's seed "purity" shows varying levels of heterogeneity

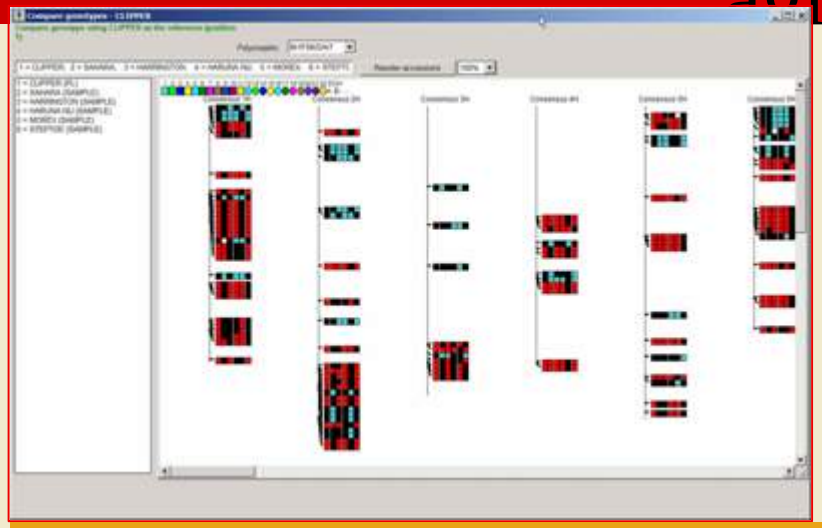
DArT in wheat: largest service volume in large and complex genome crop

- >50,000 wheat samples processed (>95% as service at ~1 cent per marker assay)
- over 250 mapping populations of wheat on the array containing common marker set – map integration effort
- High quality of individual and integrated maps
- Array expanded now to 7,700 markers
- Average call rate: 95% (DNA quality and "purity" dependent)
- Allele-calling consistency:
  - 99.7 % for wheat and other organisms
- Unambiguous discrimination of all currently grown varieties (tested for Australia and inferred for the global production)

Wheat 3B integrated map: 17 mapping populations

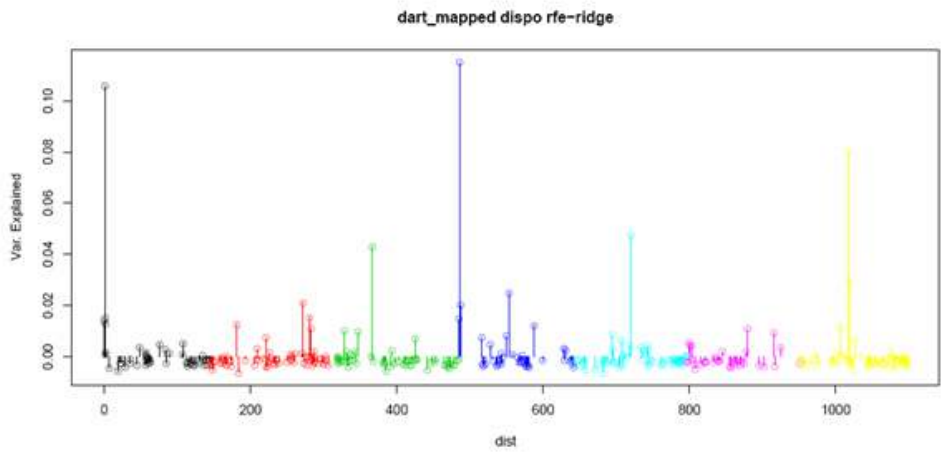
Total number of markers: 982
820 DArTs
162 others

From genotyping service to player in "digital agriculture"



Geneflow screenshot of germplasm module (Geneflow Inc)

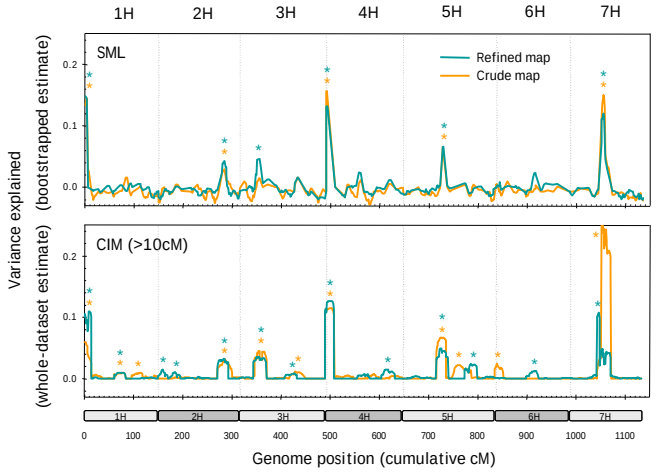
- Integration of molecular data, breeding records and field/environment data essential
- IT infrastructure and in-house software (DArTdb, Client interaction, DArTsoft, DArTools) built on Open Source software (LAMP)
- Critical collborations in IT:
  - NICTA (Australia): data mining/machine learning technologies
  - Statisticians network of GRDC (Australia) Data storage and analysis (Katmandoo)
  - Geneflow Inc. (USA): molecular breeding software

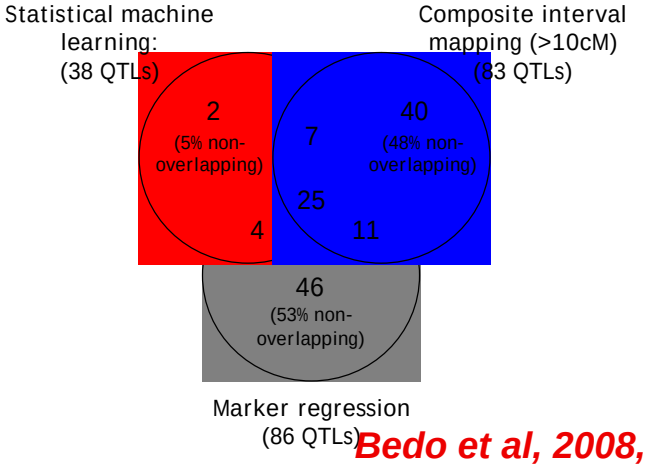


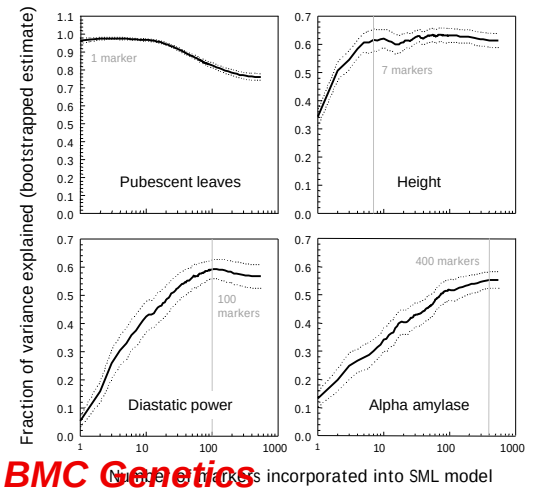
Machine learning for marker-trait associations (collaboration with NICTA)

Statistical Machine Learning for genetic data mining

- "Maps" QTL without a map – robust to genotyping errors
- Modest overlap in QTL discovered by SML and several softwares
- Measures trait "complexity"
- From QTL (detection of major effects) to genomic selection (accounting for contribution to trait at all positions)
- New application in purity testing allowing assay cost reduction
  - How pure?
  - How many contaminants?
  - What are the main contaminants?



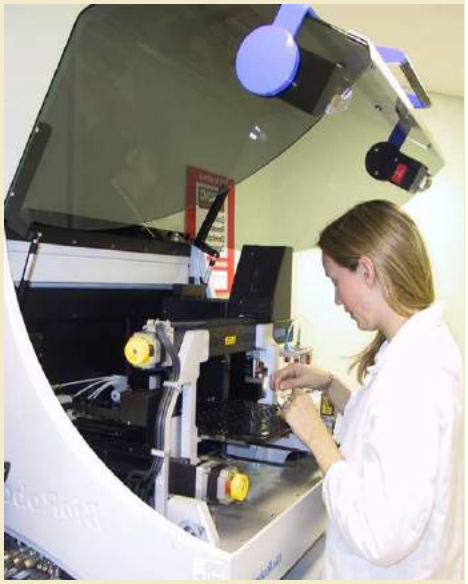




Bedo et al, 2008, BMC Genetics

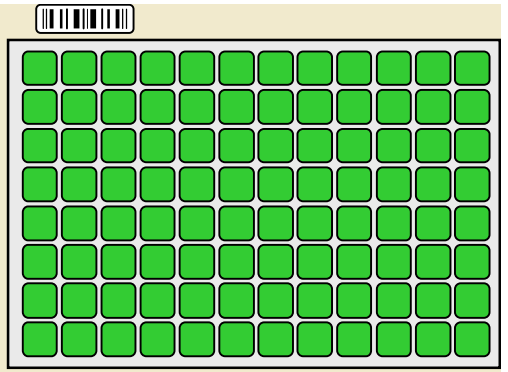
Applications of DArT in (epi-) genetics and breeding

- Comprehensive characterisation of germplasm/diversity studies: guides selection of parents
- Genetic ID/PBR (most interesting case of epigenetic discrimination of "sports")
- Seed purity/product integrity testing
- Genetic, physical (assisting genome sequencing) and QTL mapping
- Epigenetic studies (e.g. tissue culture effects)
- Association mapping and pedigree mapping
- Accelerated introgression from wild germplasm– especially in diversity-poor crops (e.g. legumes)
- Genomic selection (genome profile-based breeding): thousands to tens of thousands samples per year for larger customers

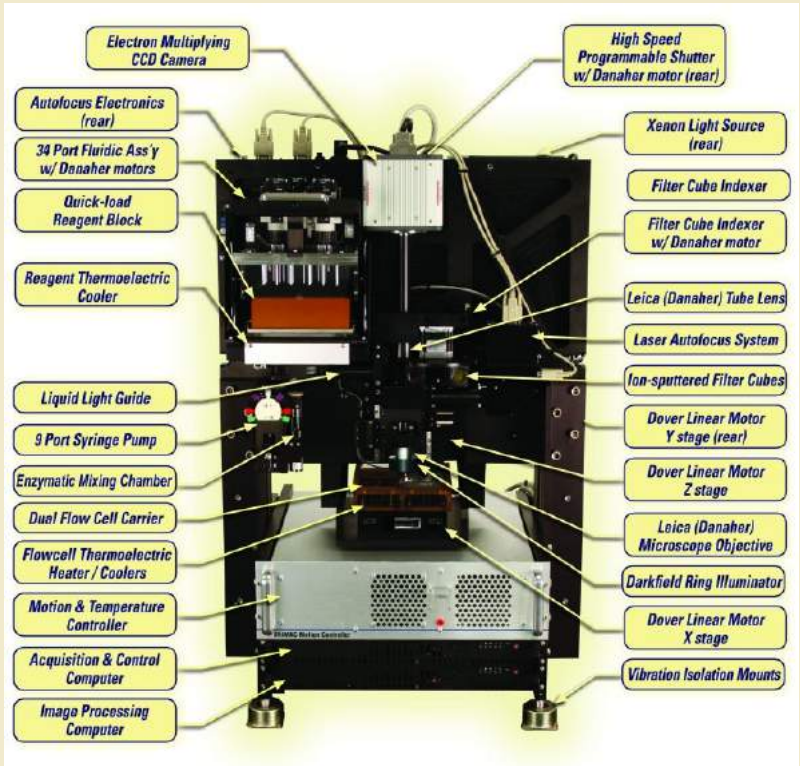


Delivering the service: platforms for various levels of multiplexing

- Standard DArT arrays
  - 7,700 markers (up to 15,000)
- "DArT Plate" format
  - 100 -1000 markers per sample
  - Genetic ID and genome profile based selection
- Optically encoded beads
- Polonator – NGS platform adopted to DArT applications
  - Open access technology (design, OS software and consumables)- cheap and "adaptable"
  - Collaboration with system developers to enable throughput increase & price reduction
  - Developing our own "chemistry"



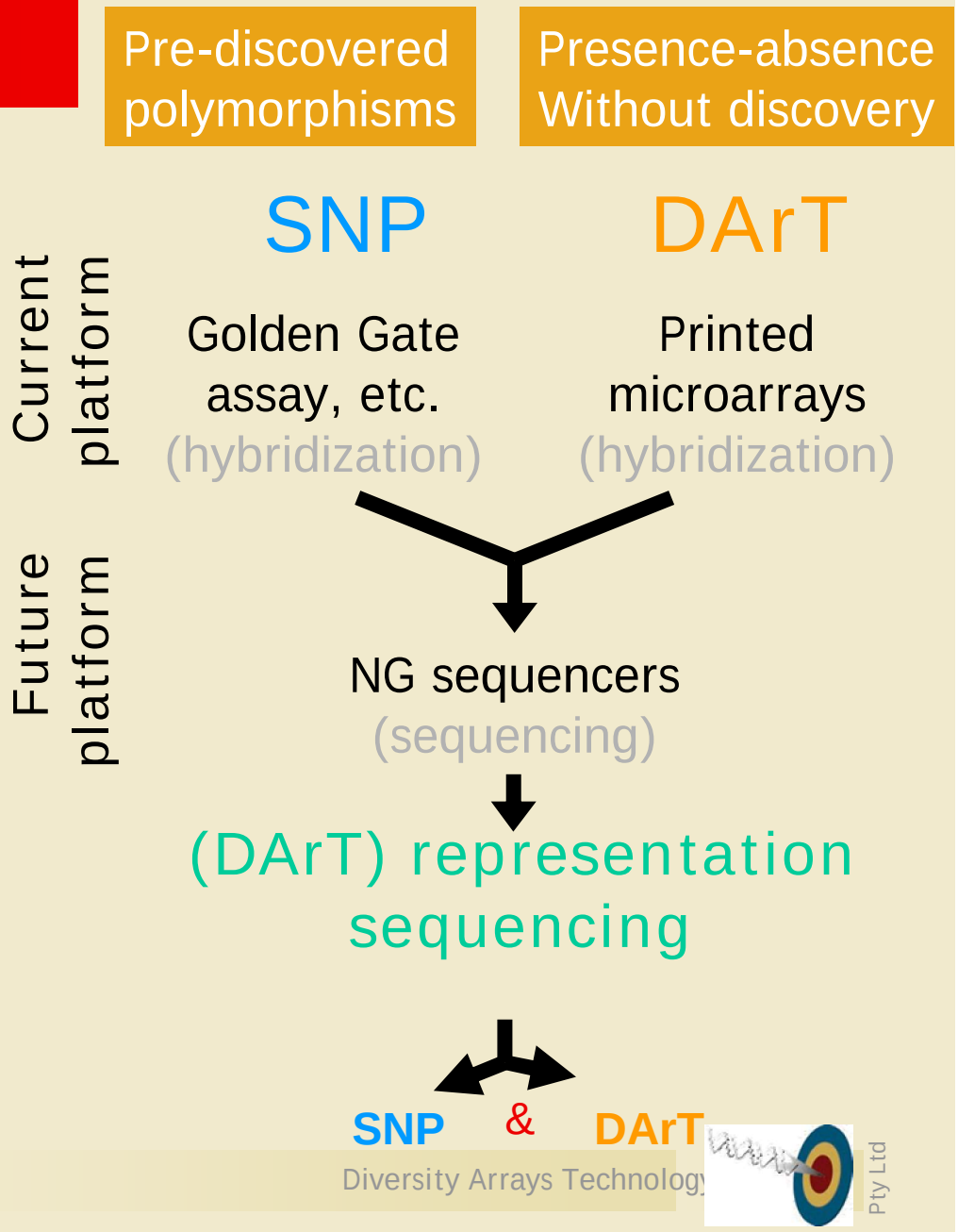
DArT Plate for parallel analysis of 96 samples at medium multiplex



***Poonator* open access, high quality NGS platform**

Platform convergence and pricing issues

- Sequencing of representations anticipated in DArT patent
- Until now price prohibitive both for hardware and consumables
- First "research" attempts on other NGS platforms show promise, but are expensive
- Still substantial effort required to make the system competitive against "old" DArT in terms of assay cost (10- 30 Euro per sample in high volume of service and 96 plate unit)
- Price of authentication test will be determined by:
  - Staff cost (including training)
  - Hardware investment
  - Volume of service
  - Precision of answer required
  - consumables



Conclusions

- DArT developed for >55 organisms representing wide range of breeding systems, genome size and ploidy levels
- Technology of choice for crops with limited resources or complex. polyploid genomes but working well in “simple” ones like rice.
- Exploiting high precision of restriction enzymes compared to oligonucleotide hybridisation kinetics; ability to discriminate “sports”
- Addressing the need for technologies and services providing genetic fingerprints at the density appropriate for application
- Increasing throughput and resolution by evolving hardware platforms combined with development of dedicated software solutions
- Heterogeneity of cultivars poorly measured/understood and most likely critical parameter of any authentication strategy (“reference “)
- Limitations of Authentication substantially non-technological especially in light of rapid convergence of modern DNA platforms
- Need for clarity of definitions and requirements and need to be guided to strike a fair balance of benefits between consumer, producer and environment