Modern plant breeding supported by high density genetic information

Eric Huttner

Outline

- Intro: plant breeding and the “Information Revolution”.
- Manipulating gene one by one: one example
- Cost-effective high density genome profiles
- Their applications

Crop improvement matters !

Improved plant varieties

- Improved variety: combinations of genes delivering performance
  - For the farmer:
    yield, economic and environmental sustainability
  - For the consumer of the products
    quality, price, environmental sustainability
- Combining genes
  - “Classical” breeding
  - Genetic modification

This is “Science for Society”
Have you thanked a plant breeder this morning ?

Bt Cowpea for Africa (TJ Higgins et al.)

Cowpeas are an important protein food for 200 million people in Sub-Saharan Africa.

- Maruca Podborer is the Major Target
- Use Gene Technology to Introduce Insect Resistance Genes
- Complement Traditional Cowpea Breeding Programmes in Africa

Nigerian cowpea breeder with Bt cowpea

Breeding

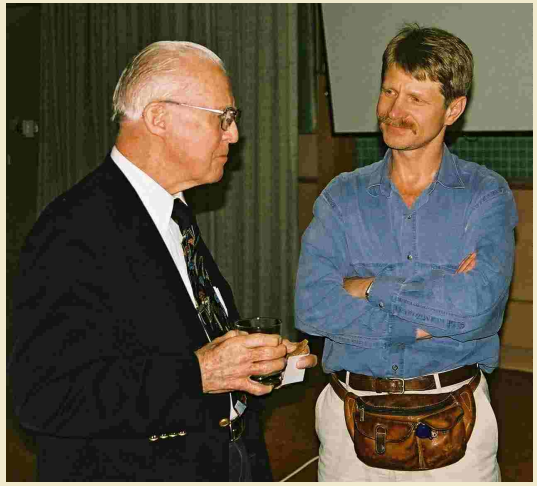
- Breeding is Integration: combining large number of genes
- Routine use of molecular genetic (DNA) markers
  - Germplasm diversity
  - Identify and track useful genomic regions
  - Small number of markers used in routine breeding
- Past limitations
- Can we do better ?
  - For complex traits
  - For complex germplasm
  - For orphan or “small” crops

Breeding with molecular markers

- Simple traits
  - For example disease resistance
    one gene – large effect
  - Marker = alternative to phenotyping
  - Medium density molecular profiles:
    accelerate introgressions of small genomic regions
- Complex traits
  - Need to combine many genes
  - Require high density molecular profiles
  - Only recently practical and cost-effective
  - Applications of these data require new IT tools

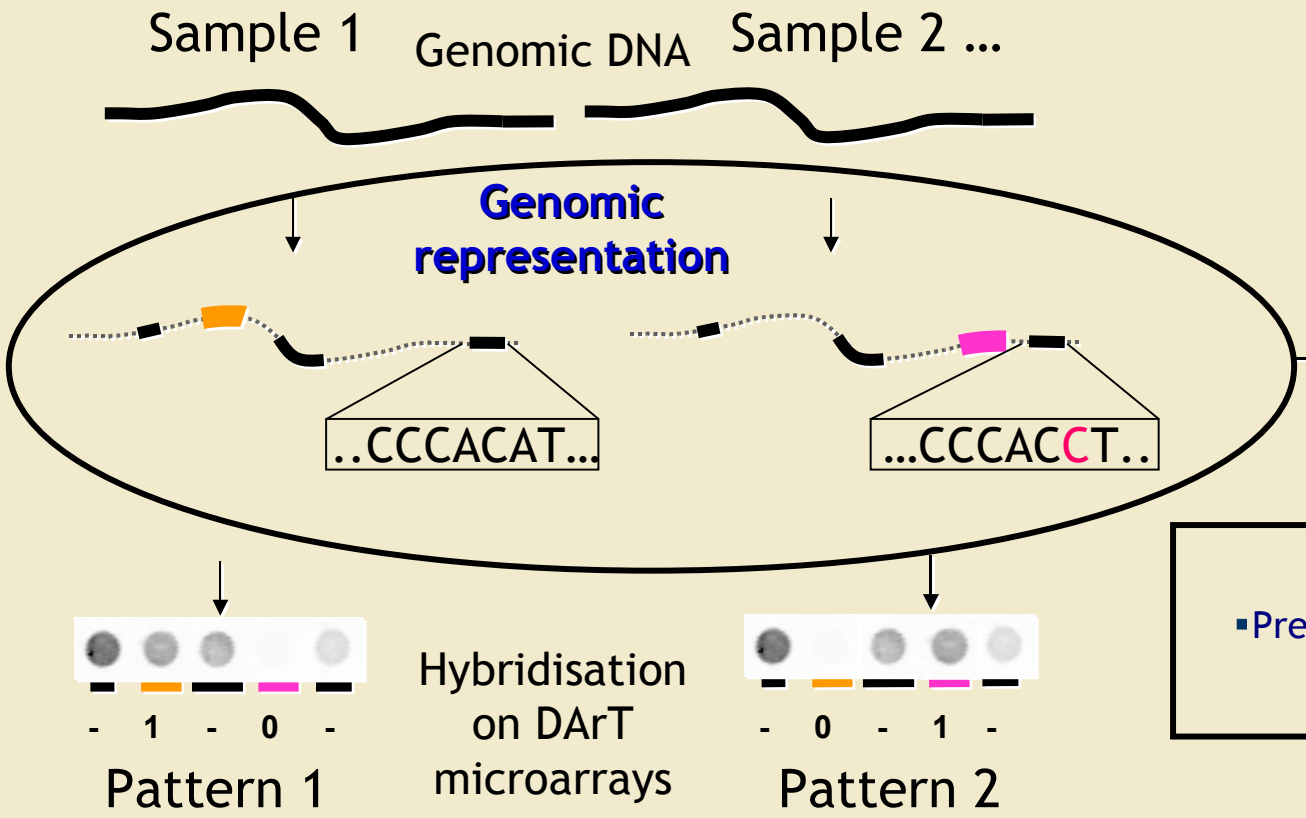
Diversity Arrays for high density molecular profiles

- Original method invented by Andrzej Kilian
- Complexity of the genome to be analysed is reduced to about 1%: Genomic Representation consisting mostly of low copy DNA
- Representation analysed:
  - Microarray platform
  - DNA sequencing platform



How it works

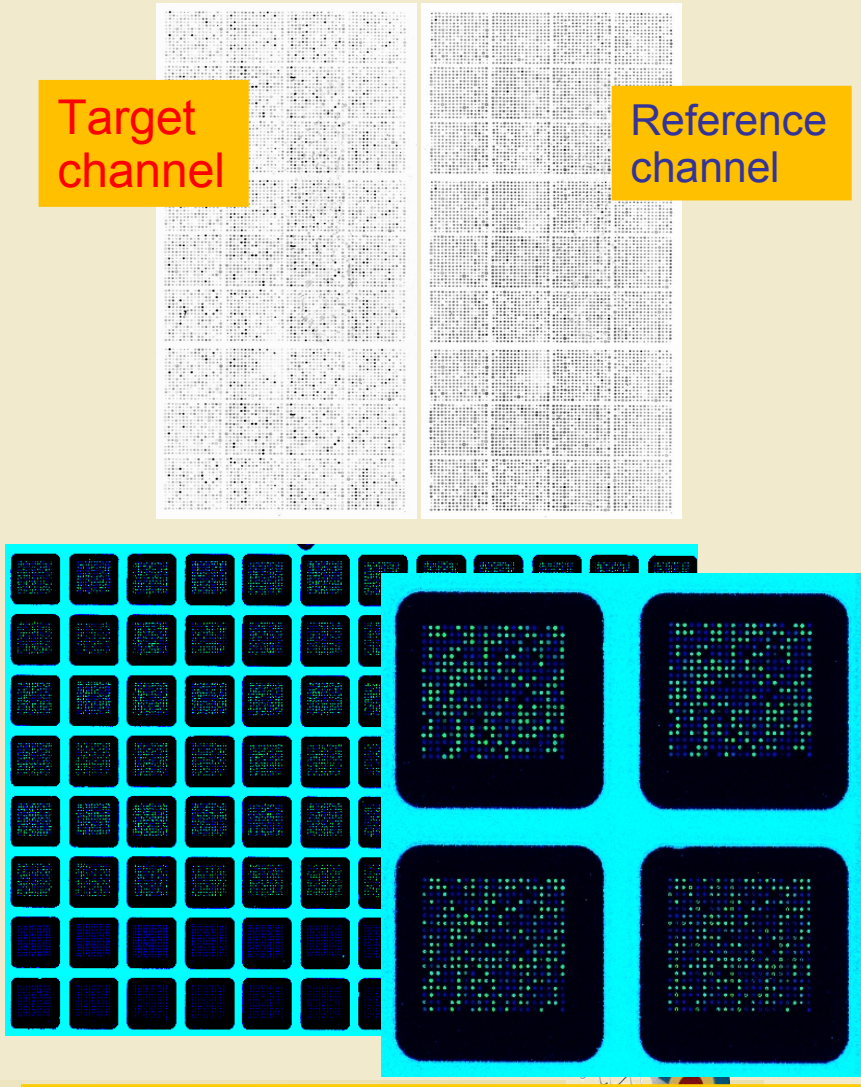
DArT Marker: genomic DNA fragment, **the presence (quantity) or sequence of which** is polymorphic in a Genomic Representation



- Count each fragment
- Presence / Absence polymorphism
- Identify SNPs

DNA-to-data service: Microarray

- Report the score: 0 or 1, for each marker and each sample typed.
- High density arrays
  - 7K to 15K markers assayed
  - Genetic mapping, QTL mapping, association mapping
  - About 0.01$ per marker assayed
- Low density arrays
  - 200-500 markers
  - Genetic ID, seed purity, Genomic Selection



DArT Plate for parallel analysis of 96 samples at medium multiplex (200-500)

Example of result: scoring table

				Extract Plate Barcode		١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠	١٩-٤-٠
						A	B	C	D	E	F	G	H	A	B
						١	١	١	١	١	١	١	١	٢	٢
MarkerName	CloneID	Chromosome	P	Call Rate	PIC	LW١٢	LW٦١	LW١١٩	LW١٥٦	LW٢١٣	LW٢٥٩	LW٣٢٠	LW٣٤٨	LW١٦	LW٦٣
wPt-٦٣٢٩	٦٦٤٦٩	١A	٧٤.٣٣	٨٠.٨٥	٠.٥٠	١	١	١	٠	١	X	١	X	٠	١
wPt-٦٦٥٤	١١٦٨٦٠	١A	٨٧.٠١	٩٥.٧٤	٠.٥٠	٠	٠	١	٠	١	١	١	X	X	X
wPt-٢٨٧٢	١١٥٣١٥	١A	٨٥.٣٠	٩١.٤٩	٠.٥٠	X	٠	٠	X	X	١	٠	٠	١	٠
wPt-٩٣١٧	١١٥٩١٨	١A	٨٣.٩٢	٨٨.٣٠	٠.٥٠	X	X	١	X	X	١	١	X	١	١
wPt-٠٤٣٢	١١٦٠٥٤	١A	٧٨.٨٨	٨٨.٣٠	٠.٥٠	١	٠	٠	١	٠	١	١	١	X	X
wPt-٦٧٠٩	١١٦٨٦٣	١A	٧٨.٥٠	٨٠.٨٥	٠.٤٩	٠	٠	X	X	X	١	٠	١	٠	٠
wPt-٠١٢٨	١١٦٥٨٧	١A	٧٩.٢٧	٨٩.٣٦	٠.٤٩	١	X	٠	١	X	١	٠	٠	٠	X
wPt-٨٣٤٧	١١٦٠٦٤	١A	٨٨.٦٣	٩٦.٨١	٠.٤٨	٠	١	١	١	١	X	٠	١	١	١
wPt-٩٥٩٢	١٢٠٦٨٨	١A	٨٣.٠٩	٨٩.٣٦	٠.٤٨	٠	٠	١	٠	٠	١	X	١	X	X
wPt-١١٦٧	١١٩٨٥٥	١A	٨٦.٣١	٩٢.٥٥	٠.٤٨	١	٠	١	١	X	X	١	١	١	١
wPt-٨٠١٦	١١٥٤٢٦	١A	٨٨.٨١	٩٤.٦٨	٠.٤٦	١	١	١	١	X	١	١	٠	X	١
wPt-٦٠٧٤	١١٥٤٩٧	١A	٨٧.٤٠	٩٦.٨١	٠.٤٥	١	١	١	٠	١	١	٠	١	١	١
wPt-٢٥٢٧	١١٦٩١٣	١A	٨٧.٨٢	٩٥.٧٤	٠.٤٤	١	١	٠	٠	١	X	٠	٠	٠	X
wPt-٦٣٥٨	١١٦٤٢٤	١A	٧٥.١٢	٨٩.٣٦	٠.٤٤	٠	X	١	١	١	١	٠	١	١	١

DNA to data service: Genotyping By Sequencing

- Genomic Representation is sequenced
  - Millions of tags generated per sample
  - 100-1000 tags per sequenced fragment
- Presence – Absence markers: report the score: 0 or 1, for each marker and each sample typed.
- Single Nucleotide Polymorphisms, identified by comparing to a reference, reported separately

Markername	Sequence	SNP	Discordan	CallRate	PIC	One	Tags pmr	Het	SNP	Ref	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6	Ex 7	Ex 8	Ex 9	Ex 10	Ex 11	Ex 12	Ex 13	Ex 14
100000264| F| 0	TGCAGAACAA…	Ref	0.016	98.93	0.426	0.692	1740	0.014	0.335	0.674	0	1	1	0	0	0	-	0	1	0	0	0	1	0
100000264| F| 0	TGCAGAACAA…	45:T>C	0	99.47	0.437	0.323	884	0.014	0.335	0.674	1	0	0	1	1	1	1	1	1	1	1	1	0	1
100000309| F| 0	TGCAGTAATA…	Ref	0	100.00	0.165	0.909	1589	0.000	0.081	0.919	1	1	1	1	1	1	1	1	1	1	1	1	1	1
100000309| F| 0	TGCAGTAATA…	48:G>A	0	100.00	0.165	0.091	120	0.000	0.081	0.919	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100000374| F| 0	TGCAGATTAT…	Ref	0	100.00	0.129	0.93	1945	0.005	0.077	0.928	1	1	1	1	0	1	1	0	1	1	1	1	1	1
100000374| F| 0	TGCAGATTAT…	30:G>A	0	100.00	0.139	0.075	157	0.005	0.077	0.928	0	0	0	0	1	0	0	1	0	0	0	1	0	0

Services available for 72 crop plants and relatives

Aegilops	Coffee	Millet	Sorghum
Agropyrum	Common bean	Miscanthus	Soybean
Alfalfa	Cotton	Mung bean	Spinach
Apple	Cucumber	Oat	Spruce
Bambara groundnut	Dactylis	Oil palm	Strawberry
Banana	Date Palm	Olive	Sugar beet
Barley	Eucalyptus	Pear	Sugarcane
Brassica	Fir	Pigeonpea	Sweet potato
Cacao	Grape	Pine	Switchgrass
Canarygrass	Groundnut	Pineapple	Taro
Capsicum	Hemp	Plum	Tea
Carrot	Hop	Poppy	Tobacco
Cassava	Jatropha	Potato	Tomato
Castor bean	Lolium	Pumpkin	Triticale
Chickpea	Lupin	Quinoa	Tritordeum
Citrus	Macadamia	Rice	Wheat
Clover	Maize	Rubber tree	Willow
Coconut	Medicago	Rye	Yam

We can develop the technology for any other crops in a few weeks

Applications of DArT (1)

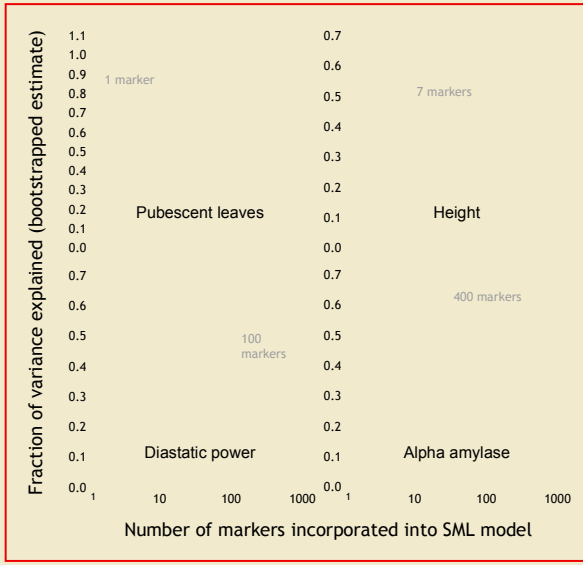
- Variety fingerprinting, diversity analysis
- Rapid creation of linkage maps for QTL analysis, etc.
  - Map as you go
- Accelerated introgression from exotic germplasm
  - Example of a GM trait to transfer to local varieties
- Single-step quantitative bulked segregant analysis.
- Any DArT marker can be converted to a single marker and applied on its own but:
  - Assay development required
  - Cost effectiveness ?

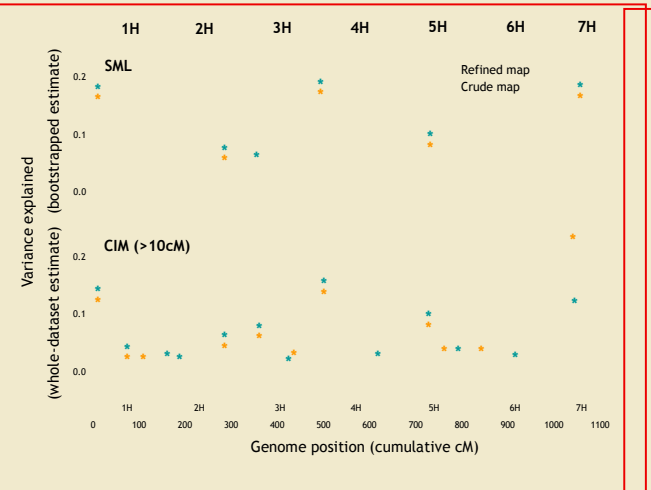
Applications of DArT (2)

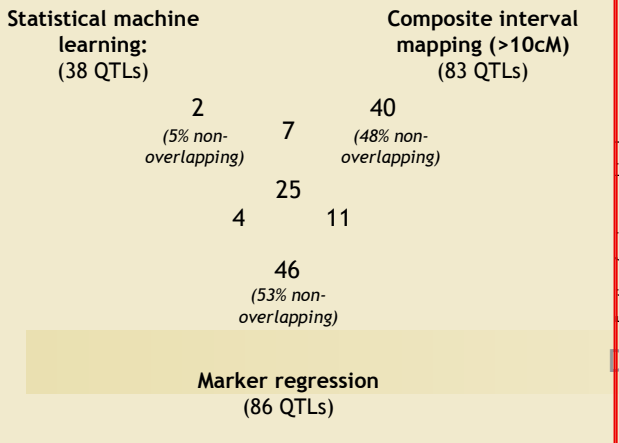
- "Whole-genome profile" applications
  - Comprehensive single-step characterisation of large collections germplasm.
  - Genetic background screening.
  - Identification of multiple regions responsible for complex traits ("adaptation complexes").
  - Mapping of genome rearrangements.
  - Marker-assisted selection for many traits simultaneously.
- Will this enable new breeding methodologies ?
  - "Bolder" use of genetic diversity, increased use of exotic germplasm.
  - Data mining approaches to marker-trait association.
  - Genomic Selection

Statistical Machine Learning for data mining

- Method described by Bedo et al, BMC Genetics (2008) using biparental barley population for proof of concept
- Several algorithms, with and without structure component accounting
- Measures trait "complexity" in number of markers to create best model
- Applicable to any population type including breeding programs
- Used either for QTL/association mapping or for genomic selection (to account for contribution to trait at all positions)







The limitations of Marker Assisted Selection

- Genes associated with traits can be found
  - In biparental or multiparental populations
  - In association populations
- But: for many traits a large number of genes each contribute very little
- Yang et al. Nature Genetics 2010
  - human height (heritability 80%)
  - 50 best markers only explain 5% of additive variance
  - All markers explain 46% of additive variance

Genomic Selection (GS)

- First developed for animal breeding
- Phenotype and genotype a Training Population
- Use all marker data: estimate the effect of all markers
  - Many possible statistics can be applied
- Calculate a Breeding Value: the GEBV
  - Depends on the trait(s) selected
  - Yield, product quality, disease resistance, …
- Validate the model on a Validation Population
- Apply by genotyping breeding populations
  - Use the GEBV (instead of the phenotype) to select the best individuals

Features of GS

- Unbiased use of all marker data
- No need to discover QTLs
- Breeders can use markers without understanding the underlying biology
- GS is practical because whole genome molecular profiles are now available at an acceptable cost
- Models show increased breeding efficiency compared to phenotypic and Marker Assisted Selection
- Practical considerations
  - Linkage Disequilibrium required
  - Breeding systems, population structure
  - Population size
  - Marker density
- Does it work for plant breeding ?

GS example: Eucalyptus
Resende et al. 2012, New Phytologist, 194, 116-128

- 2 breeding populations: 11 and 51 parents
- 4 traits, 700-900 individual genotyped
- 3000-3500 polymorphic DArT markers used
- Correlation of predicted and observed Breeding Values about 0.7
- Use of GEBV
  - Early selection before phenotypes can be measured
  - Reduce breeding cycle

Contact author: Dario Grattapaglia
dario@cenargen.embrapa.br

GS example: Wheat historical lines at CIMMYT
Crossa et al, 2010, Genetics, 186, 713-724

- 599 lines, 4 environments, 1279 DArT markers
- One trait: Grain yield
- Tried several methods to estimate marker effects
- Correlation of predicted and observed yield: between 0.4 and 0.6 depending on environments and methods

Contact author: Jose Crossa j.crossa@cgiar.org

Wheat bi-parental populations (Cornell University)
Heffner et al. 2011, Crop Science,51, 2597-2606.

- Two biparental populations: high LD
- Nine grain quality traits, high heritability but difficult phenotyping
- Marker density of 300-500
- Population size: 24 to 256
- Multiple environments
- Accuracy of 0.5 achieved for population size of 96

Contact author: Mark Sorrells mes12@cornell.edu

GS Review and Conclusion

- Recent review (Cornell University): Lorenz et al. (2011) Advances in Agronomy 110:77-122
  - Contact author: JL. Jannink jeanluc.jannink@ars.usda.gov
- Data now showing that GS can be used in plant breeding in a range of contexts.

Wheat stem rust (image USDA)

Conclusion

- Application of GM technology can serve society
  - Safety issues addressed
  - How Europe is slowing Africa's progress
- New breeding methods are becoming possible
  - In the lab: new technologies, instruments
  - In silico: computer power, new algorithms
- Information becomes knowledge
- Knowledge becomes new products

Remember to hug a plant breeder tomorrow !

Current DArT team

Andrzej Kilian

IT:
- Grzegorz Uszynski
- Frank Detering
- Puthick Hok
- Matthew Sellers
- Michael Kirk
- Zahidul Islam

Laboratory:
- Vanessa Caig
- Jason Carling
- Colleen Hopper
- Gosia Kilian
- Damian Jaccoud
- Adriane Machado
- Kasia Heller-Uszynska
- Cina Vipin
- Hui Ming Yang
- Ling Xia

Acknowledgements

Current DArT team

Former students & employees

Mona Akbari
Ilaria Catizone
Cyril Cayla
Mark Jones
Pavan Krishna
Sujin Patarapuwadol
Kaiman Peng
Brent Thomson
Peter Wenzl
Shi Ying Yang

NICTA Collaborators

Justin Bedo
Adam Kowalczyk

Visitors to DArT:

Hania Bolibok
Aurélie Bonin
Sophie Bouchet
Arianna Bruzzone
Yong-gu Cho
Andrew Craig
Isabel De La Casa
Evelyn Hackl
Eshwar Kankanti
Marianne Konrad-Koeszler
David Kudrna
Li Rongbai
Ali Mehrabi
Chandrika Perera
Ange Marie Risterucci
Senthilvel Senapathy
Mike Stevens
Antoinette van Schalkwyk
Alexander Wittenberg
Prapit Wongtiem
Zong Xuxiao
Dione Mendes
Samanta Marengo
Darius Grzebelus
Boris Brinez
Vu Thi Thuy Hang
Cesar Petroli
Cina Zachariah
Feng Yan Li
Carolina Sansaloni
Anna Seroczynska
Gui Yi Jie

Discussion and questions welcome