

How I use ICIS and how I think it should be developed

Howard Eagles

Objective of our research

- Estimate genotypic values of known genes from plant breeding datasets
 - Collaborative with Karen Cane and primarily applied wheat breeders
 - Largely with diagnostic markers
 - Very large, unbalanced datasets
 - Biased estimates a substantial problem

Operations

- Identify genes of importance
- Determine if alleles can be identified

If answers are 'yes'

- Identify variation for genes in target germplasm
- We use ICIS GMS to do this

An example from current research

- Diagnostic markers became available for Vrn-A1, one of 4 genes which determine winter-spring growth habit in wheat
- 3 alleles likely to be important in Australian wheat were known: a, b, and v

Identification of likely variation

- Select key parents from known germplasm structure: worked out iteratively using ICIS GMS
 - Old (early 1890s): Federation, Florence
 - Intermediate (about 1920s): Dundee
 - Intermediate (about 1940s): Gabo
 - Intermediate (about 1950s): Olympic
 - Early CIMMYT (1960s – 1970s): WW15, Pitic 62, Pavon

Emphasis

- Parents that left substantial progeny
 - ICIS is very useful for this
- No necessarily crop production
- However, often related
- Usually less than 20 at this initial stage

With Vrn-A1 example

- Found the b allele, which is infrequent elsewhere, was in Federation and Florence
- Was maintained down the pedigrees to modern cultivars like Yenda
- The a allele came in early, and was in Dundee; maintained down the lineage through Dagger and Spear. Also in WW15
- The v (winter allele) was in very old cultivars, and also in Pavon

Operations of our research

- Assemble large data-sets, for Vrn-A1 we currently have 128 experiments between 1986 and 2009 with heading dates
- Type close to 1000 lines for alleles of the target gene.
- Do this efficiently using pedigrees and known genetic information (Glu, Pin, Rht, Lr and other genes can identify incorrect crosses)

Data storage

- Pedigree data is all stored in ICIS
- Phenotypic and genotypic data is all stored in simple Access databases. Note the plural: databases.
- I always create a new database to assemble and export data for statistical analyses in Genstat

Tables in Access

- Phenotypic data (fields are environments, lines, observations)
- Lines and their genes (linked to the phenotypic data through “lines”)
- Environment specifications (linked to the phenotypic data through “environments”)
- Specialised tables added as required

Advantages and disadvantages

- Advantage: Maximum flexibility
- Disadvantage: Have to know how to write queries
- Can use as calculating machines; eg in the Vrn-A1 example I have queries calculating vernal days and daylengths

Molecular typing complete

- Use SETGEN to make a list of entries that are going to be analysed
- Each entry must occur only once in the list
- Use BROWSE to calculate the coefficient of parentage matrix
- With numbers > 1000 , this is usually an iterative process with substantial solving of errors in the pedigree structures
- Pedigrees always best if entered first as the cross (dummy name if unknown) and selections from the cross

New Access database for export to Genstat

- Has the following groups of tables:
 - COPs (GID1, GID2, COP)
 - Entries (always 3 or more, now including field(s) for GIDs); one has the genes
 - Environments (usually only 1 in this database)
 - Data
- And sophisticated queries that assemble 2 output text files for Genstat
 - Data
 - Square COP matrix

Genstat analyses

- Use REML
- Genes and environmental covariates always fixed
- Entries, as relationship or COP matrix always random

My simple view

- Two parts of a statistical model
 1. Genetic
 2. Environmental
 3. Spatial analyses long used to minimise biases on the environmental side of the model
 4. Relationship matrices do a similar thing on the genetic side

Scope of research

- This is NOT limited to specific crops or specific traits
- Can be used with any crop
- Need to identify allelic variation for genes
- Now mostly with diagnostic markers
- Need to minimise bias

Principal

- 4 genes -> 1 genotypic value for heading date
- Quality: 8 genes -> 1 genotypic value for dough extensibility, etc
- No reason this can't be extended to 8, 9, 10, or more genes!
- Just need large datasets to predict them with reasonable accuracy

Cross predictor

- Owned by MPBCRC
- Using Glu, Pin Srp allele data (10368 combinations) predicts grain quality outcomes for specific crosses
- Breeder can compare frequencies to meet quality requirements for specific crosses
- Easy to use
- Must know alleles in parents
- Written in C#, no recent development
- No search capability

ParentInfor

	CID	Name	Genotype	Rmax	Ext	DDT	WA
▶	0	PASTOR	a i d c b b b a	430.3	19.09	5.49	64.77
	1	BAVIACORA	b c d c h b b a	367.0	19.20	5.35	65.59
	2	YITPI	a b d c h c a b	370.4	20.06	5.37	63.95
	3	ANNUELLO	a b a b b b a b	326.2	20.96	4.92	63.47

AllResults

	CID	CName	RIL	Freq(%)	PS95	PS99	Rmax	Ext	DDT	WA
▶	1	PASTO	a b d c b	3.750	79	121	415.7	19.91	5.78	63.45
	1	PASTO	a b d c b	3.750	79	121	427.0	19.26	5.49	64.93
	1	PASTO	a b d c b	3.750	79	121	406.4	20.04	5.54	63.58
	1	PASTO	a b d c b	3.750	79	121	417.7	19.38	5.25	65.06
	1	PASTO	a b d c h	8.750	33	51	379.7	19.94	5.61	63.82
	1	PASTO	a b d c h	8.750	33	51	390.9	19.28	5.32	65.30

SelectionCriterion

	RMax	Ext	DDT	WA	Rust
Min	<input type="text" value="320"/>	<input type="text" value="20.5"/>	<input type="text" value="3.0"/>	<input type="text" value="20.0"/>	<input type="text" value="0.0"/>
Max	<input type="text" value="450"/>	<input type="text" value="30"/>	<input type="text" value="5.2"/>	<input type="text" value="100.0"/>	<input type="text" value="100.0"/>

Select

SaveAll

Selected

	Cross ID	CrossNa	RIL	Freq(%)	PS95	PS99	Rmax	Ext	DDT	WA
	2	PASTOR	##	12.5	23	35	327.900	20.875	4.920	63.390
	4	BAVIAC	a b a b b	2.625	113	174	326.2	20.96	4.92	63.47
	4	BAVIAC	b b a b b	1.750	170	261	330.4	20.99	4.91	63.49
	4	BAVIAC	##	4.4	67	103	328.300	20.975	4.915	63.480
	1	PASTOR	##	3.8	79	121	373.800	20.650	5.160	63.800
	2	PASTOR	##	12.5	23	35	327.900	20.875	4.920	63.390
	4	BAVIAC	##	4.4	67	103	328.300	20.975	4.915	63.480

SaveSummary

<<Back

Exit

A spin-off of our project

- Karen Cane uses diagnostic markers for a wide range of genes
- Adding new genes as the project progresses
- We have almost all Australian cultivars characterised for these genes
- Distribute an Access database to Australian breeders of cultivars. Breeding lines on a company-by-company basis
- They like these!

Where to now?

- There are people that know a lot more about this than me, but:
- Could cross prediction capability be bolted onto ICIS?
- Could allelic information flow be linked to GMS? For example, track Vrn-A1 alleles through pedigrees, as I have done using ICIS GMS + paper?
- This is not molecular identification methods
- I am aware that other programs can do this

Facilitate usage: example

- Generate COP matrices with simple screen selections instead of Browse
- Select export formats: eg, I want 3 columns, GID1, GID2, COP. I link GID to names in Access

From Peter Martin

- Uses Kathmandu for data storage
- Import and export in spreadsheets
- Export for sophisticated data manipulation, including statistical analyses
- Re-import results, rather than try to do calculations within databases