Development of genomic resources and whole-genome prediction in the Pacific Whiteleg shrimp *Litopenaeus vannamei*.

Dean Jerry[1], Herman Raadsma[1,3],Mehar Khatkar[3], Hein van der Steen[2], Jeff Prochaska[2], David Jones[1] & Kyall Zenger[1]

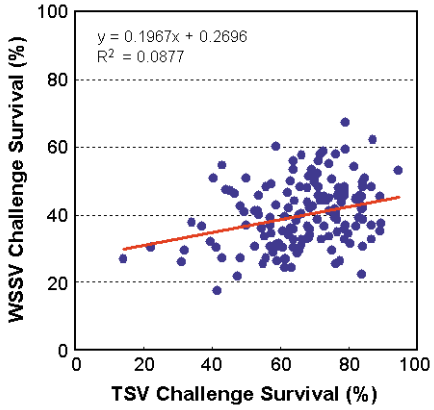
1. Centre for Sustainable Tropical Fisheries & Aquaculture, and College of Marine and Environmental Sciences, James Cook University, Townsville Australia.
2. Global Gen, Indonesia.
3. ReproGen, Faculty of Veterinary Science, University of Sydney, Australia.

Traditional breeding programs

JAMES COOK UNIVERSITY AUSTRALIA

Some limitations with traditional shrimp breeding programs

- Hard to measure traits or can't measure early
  - Disease resistance – labour intensive / accuracy
  - Carcass quality – mature animals



- Some traits low or variable heritability
  - Disease resistance – WSSV (h^2= 0.01-0.05)
  - Fertility - no of eggs (h^2= 0.09)

- Negative or low genetic correlations.
  - Disease & growth - WSSV (up to $_gc$ = -0.40)
  - Low between diseases - TSV and WSSV survival ($_pc$ = 0.09)
  - Increased difficulty in performing multi-trait selection

Traditional breeding programs

JAMES COOK UNIVERSITY AUSTRALIA

- **Limitations continued..**
  - Selection candidates not always directly evaluated
    - sacrificed in testing – eg., disease & carcass composition
  - Genetic change = (**accuracy of selection***selection intensity*genetic standard deviation)/generation interval

Solution

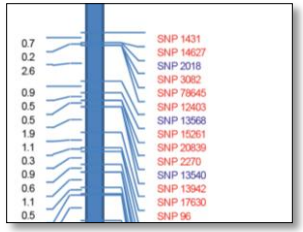
Directly evaluate genes or parts of the genome responsible for favourable traits

1. QTL mapping / GWAS and MAS
2. Genomic Selection

Program Objectives

Years

0	1)	Generate a large genomic sequence and SNP resource for *L. vannamei*
1	2)	Develop trait recording program and pedigree / genetic parameter evaluation
2	3)	Create dense genetic maps for genome structure and trait association studies
3	4)	Perform GWAS / QTL investigations
4+	5)	Evaluate genomic selection options based on dataset and farm resources



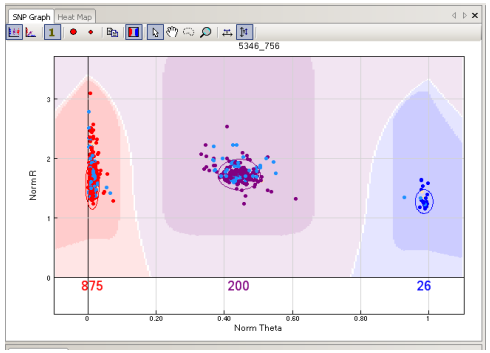
1. Illumina Infinium SNP array

JAMES COOK UNIVERSITY AUSTRALIA

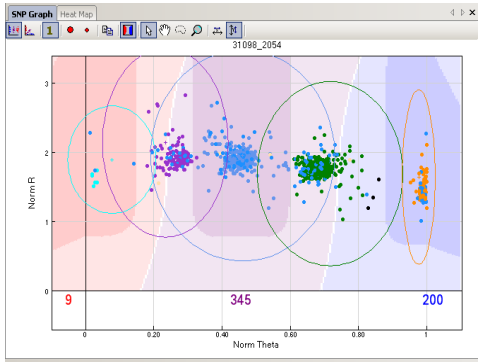
❑ SNP validation of **8,967** SNPs across 1,327 *L. vannamei* (GenomeStudio)

SNP Exclusion Reason	# SNPs (Percentage)
Probe didn't bind	262 (2.91)
Ambiguous Clusters (poor probe)	1,391 (15.51)
Het Excess	163 (1.82)
MAF < 0.01	276 (3.10)
Mendelian Inheritance Errors	425 (4.77)
MSV (genome duplication)	43 (0.5)
Total Useable	***6,407 (71.55)***
Total SNPs	8,967

Good SNP Cluster



MSV



❑ Average MAF = 31.21%

❑ Average call rate = 98.16 %

2. *L. vannamei* animal resources

- Indonesian based shrimp breeding company
  - Nucleus breeding centre
  - Broodstock multiplication
  - Hatchery PL production
  - SPF (9 pathogens)

- Diverse foundation stock (5 sources)
- Family based selection began 2008 (288+ FS families / year)
  - Growth
  - Reproduction
  - Survivability
  - Low salinity tolerance
  - WSSV, IMNV, TSV resistance

~300-1000 family average records

3. Genome resource development

Genetic Linkage Maps

- 13 mapping reference families

  → 688 informative meiosis

- 4,390 SNPs successfully mapped

  → LOD 2 & 3 framework = 2,898 SNPs

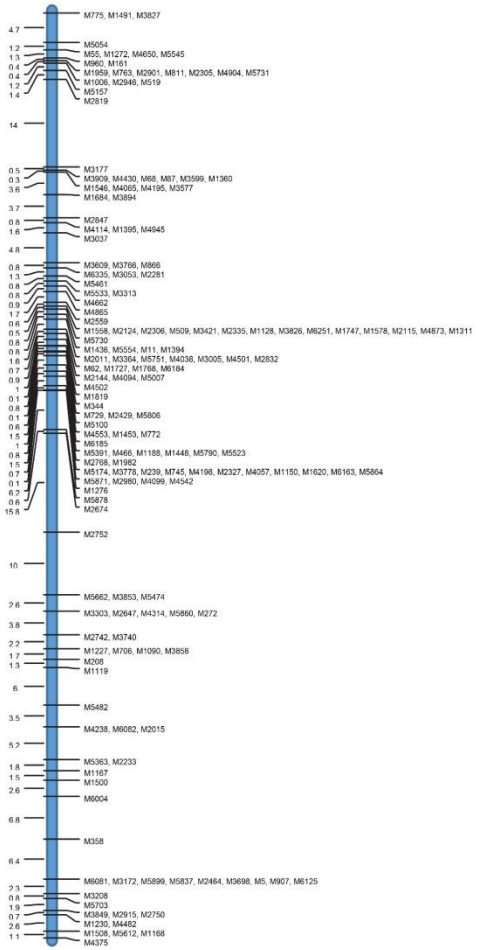
- 45 linkage groups → 97.89% coverage
- Sex average total length → 4559 cM
- Average inter-locus distance (no 0's)

  → 2.67cM

QTL analysis – sex

- 3 major QTL - 63%, 23%, 39% effect

Linkage group 6



4. GWAS analysis

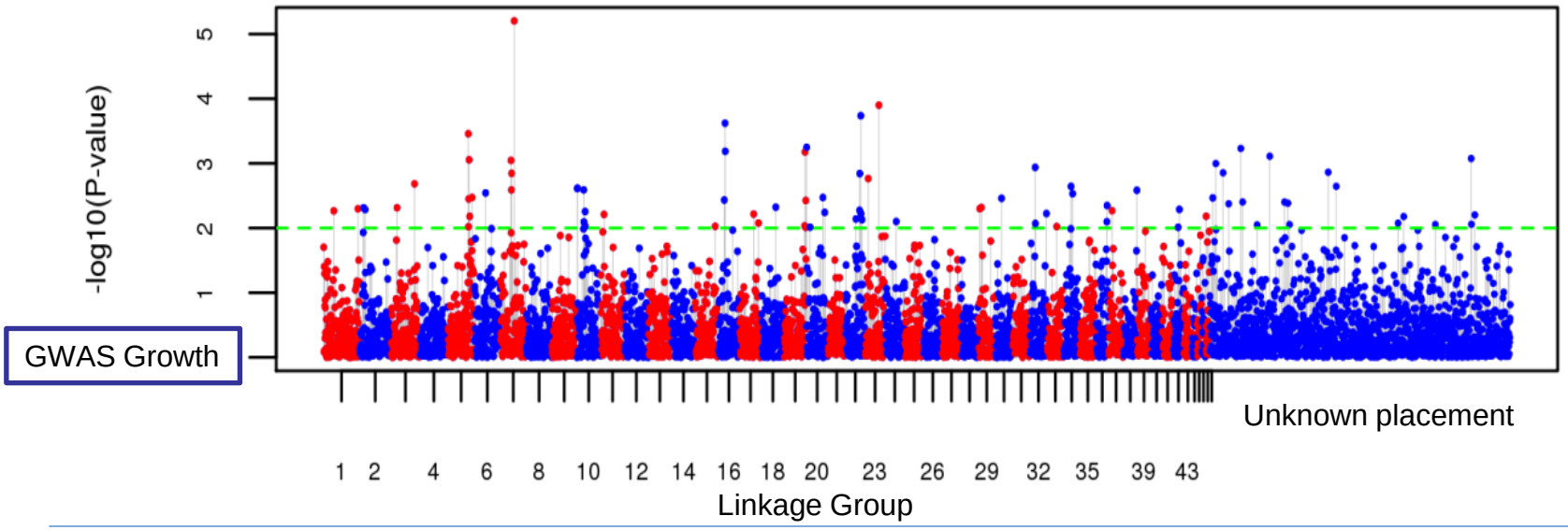
JAMES COOK UNIVERSITY AUSTRALIA

Genes of major effect - power analysis

- >90% power for moderate-high heritability (>0.15, ie., growth)
- 30-50 % power for low heritability (0.02-0.08, ie., disease resistance)

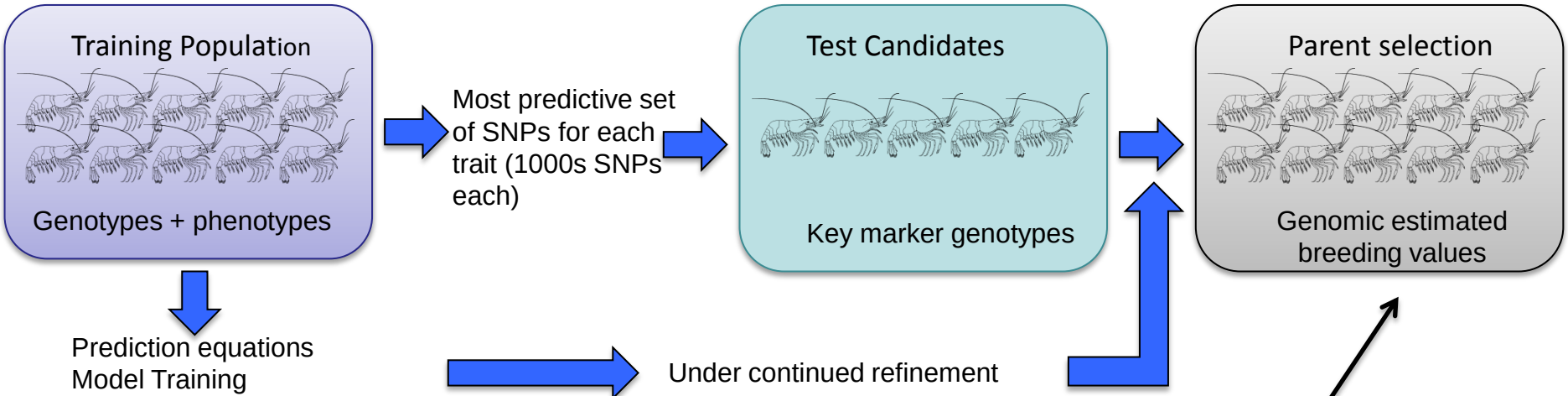
GWAS results

- No significant SNPs following FDR correction across all traits
- No genes of major effect for moderate-high heritable traits
- Insufficient power to detect genes of small effects



5. Genomic selection

- ***Most traits are complex involving many genes of small effect***
  - *Need to simultaneously search for all genes of small effects*



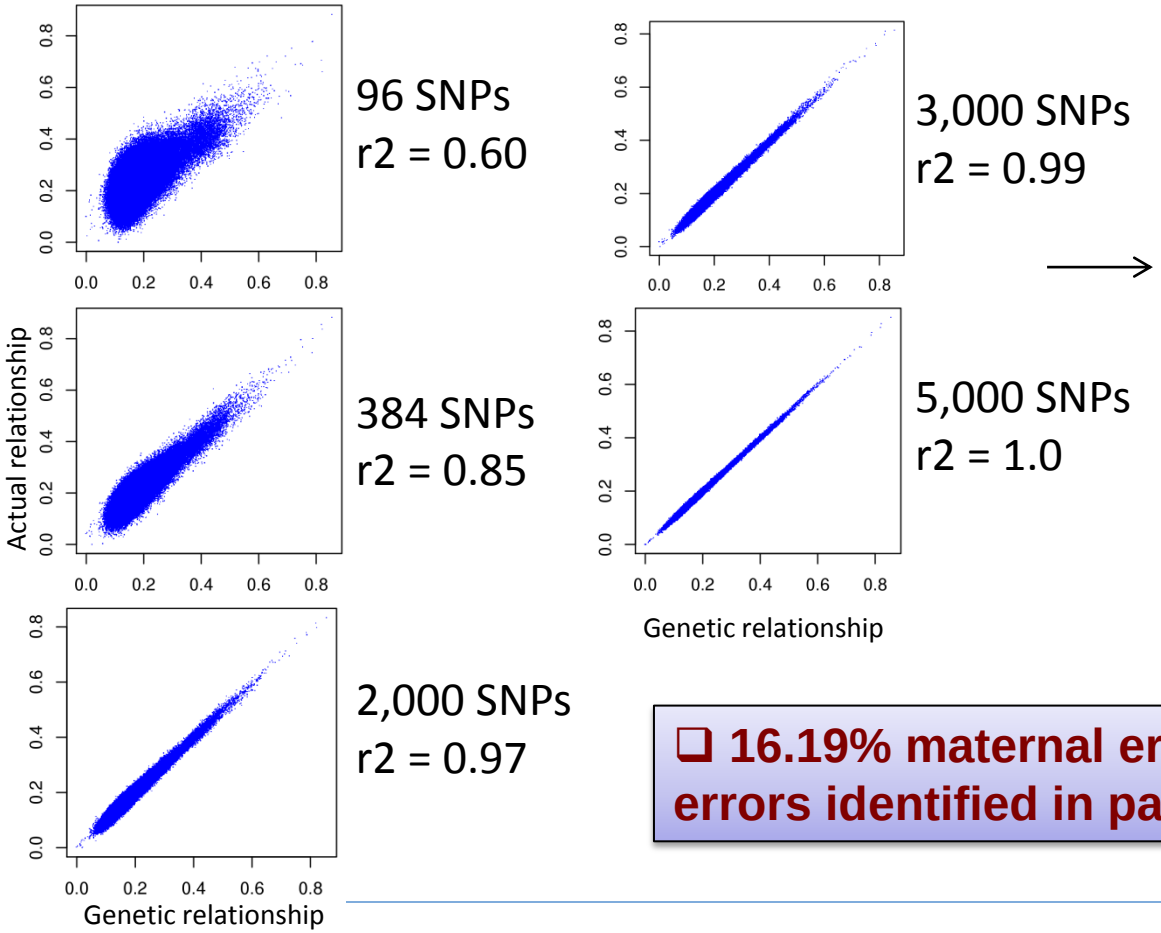
ID #	SNP1	SNP2	SNP3	SNP4	SNPn	GEBV
1M	8	3	4	8	6	29
2M	6	1	3	0	1	11
3M	0	1	0	1	-3	-1
4F	6	3	2	5	1	17
5F	1	2	--1	2	-4	2

Meuwissen et al, (2001)

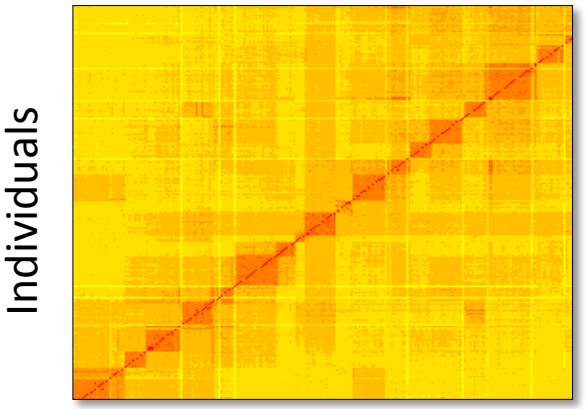
5. Genomic Selection application

What's required?

Accurate pedigrees – Genomic relationship matrix



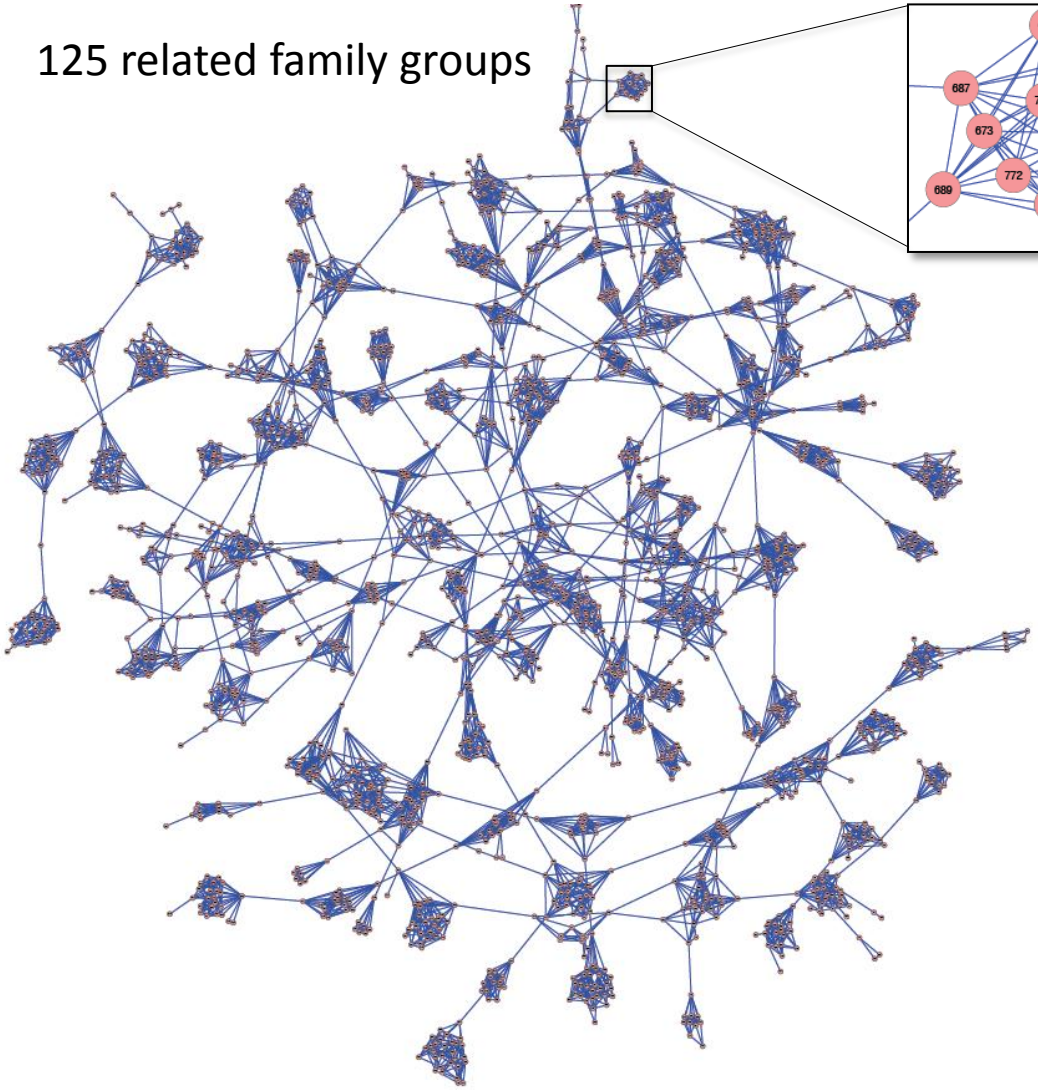
Pair wise IBS Heat map



Red blocks = family groups

❑ **16.19% maternal errors and 21.17% paternal errors identified in paper pedigree**

5. Genomic Selection application



NETVIEW analysis

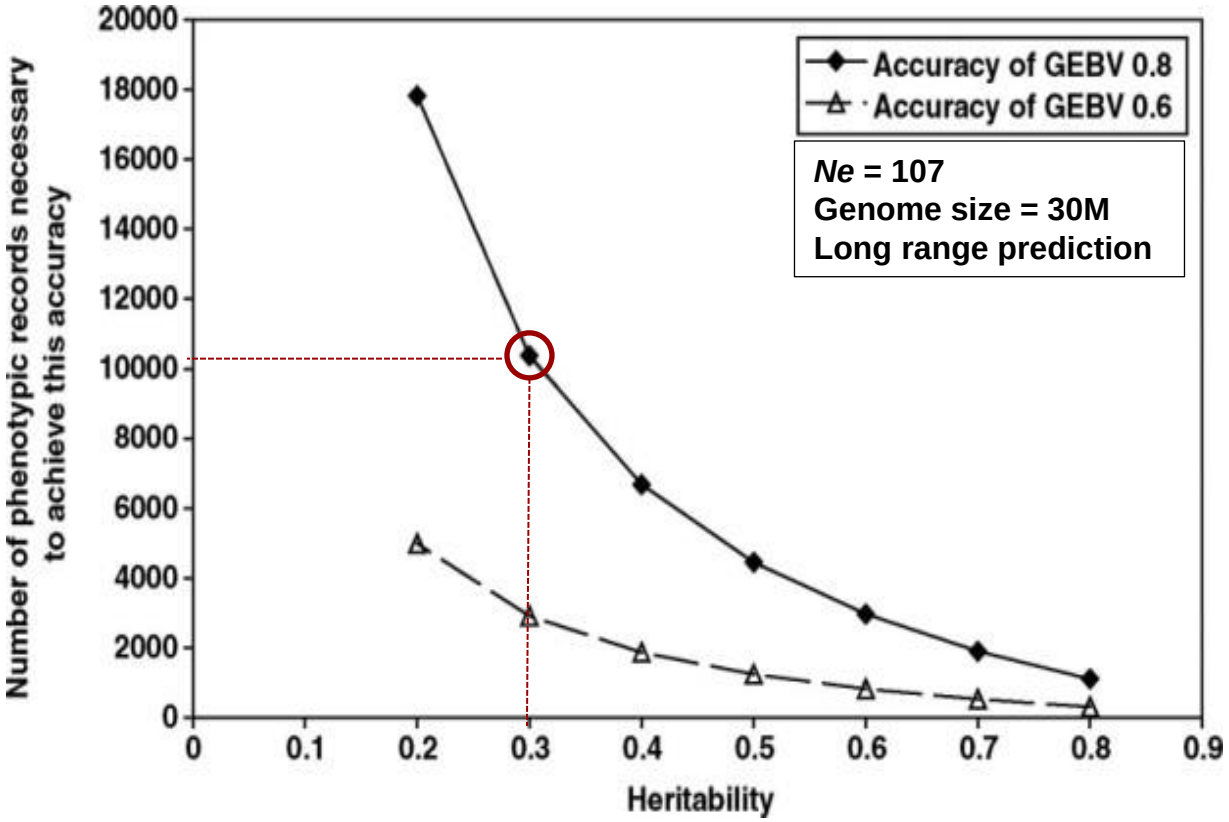
- Uses genome-wide SNPs
- Groups individuals into clusters without prior ancestry info
- Reveals fine-scale individual relatedness
- Reveals common ancestors and family connections
- Useful in correcting / inferring pedigree information

Neuditschko et al, *PLOS one* (2012), Steinig et al, *Mol Ecol Res* (2015)

5. Genomic Selection application

What's required?

Unrelated Animal Numbers[1]



Random SNP Numbers[1,2]

Unrelated individuals

10*Ne*L = 32,500

8*Ne*L = 26,000

6*Ne*L = 19,500

2*Ne*L = 6,500

1. Hayes & Goddard *Genome* (2010); 2. Sonesson et al, *GSE* (2010)

5. Genomic Selection application

JAMES COOK UNIVERSITY AUSTRALIA

What's available ?

Large numbers of accurate phenotypic records required

- ❑ Commercial farm data only
  - Limited number of family average phenotypic records (~300-900)
  - Both parents not always genotyped
  - Needed to infer family parental allele frequencies
    - Matching family average phenotype & genotype
    - Used pooled genotyping approach on nauplii (300+) to reconstruct family parent allele frequencies (ie, 0, 0.25, 0.5, 0.75, 1)
    - Pooled genotyping approach had 99.49% accuracy at reconstructing parental allele frequencies / genotypes

Although limited data – what's the GS results for growth and TSV resistance!

5. Genomic Selection application

JAMES COOK UNIVERSITY AUSTRALIA

Accuracy of genomic prediction (MIRROR) for Growth – moderate-high heritability

% in training	Number of families		gBLUP results for Test set			rrBLUP results for Test set		
	Training	Test	Accuracy	Bias	MSE	Accuracy	Bias	MSE
75	312	104	**0.667**	1.251	0.072	**0.693**	1.256	0.053
67	277	139	**0.597**	0.965	0.075	**0.632**	1.039	0.056
50	208	208	**0.618**	1.428	0.077	**0.647**	1.478	0.051

Accuracy of genomic prediction (MIRROR) for TSV resistance – low heritability

% in training	Number of families		gBLUP results for Test set			rrBLUP results for Test set		
	Training	Test	Accuracy	Bias	MSE	Accuracy	Bias	MSE
75	237	79	**-0.069**	-8.759	0.943	**-0.026**	-0.492	0.947
67	211	105	**0.135**	16964	0.897	**0.045**	16.352	0.896
50	158	158	**-0.038**	-24678	0.863	**0.152**	90.081	0.862

Final Comments

- Comprehensive SNP and map resources developed
- Family average phenotype and parent average allele frequency used (single record and genotype)
- Genomic relationships used to replace pedigree matrix @ $r^2 = 0.97$
  - ~3000 SNPs required
- Current SNP density appears to be sufficient for GS
  - more required for GWAS analysis (small effect genes)
- No indication of genes of major effect
- Genomic Selection a real option in this commercial farm
  - More phenotypic/genotypic data required for low heritable traits


Questions
ISGA 2018 in JCU Cairns!