

Genomic Selection: Promise And Concerns

William Muir Purdue University



Efficiency of MAS Relative to Individual Selection



Lande and Thompson (1990)

Problem: Cannot identify QTL that account for even moderate proportion of genetic variation with lowly heritable traits

Detection of QTL

- Also requires phenotypic data
- Low Heritability Traits
 - Cannot Establish Marker-QTL associations
 - MAS is Most Needed for Such Traits

PARADOX

Potential Solution

- Genome Wide MAS (GMAS)
 - Meuwissen et al 2001
 - Uses all Markers
 - Dense
 - Every 1cM



Implementation (Meuwissen et al, 2001)

- Combine All Data In Mixed Model
 - QTL effect Assumed
 Sampled From
 Distribution

Random Effect

Estimation

$Y = XB + ZG + \varepsilon$

where G_i represents the genetic effect of the ith haplotype, Z_i is an incidence matrix and has a 0, 1, 2 for the number of haplotypes of type G_i present in the jth animal

Prediction

$$\widehat{Y}_j = GEBV = \sum_i Z_i \widehat{G}_i$$

Estimation

- Requires Multiple Generations of Data
 - All Individuals
 - Genotyped at All Loci (Z matrix)
 - Phenotypes Measured for All Traits (Y Matrix)
 As Many Traits as Desired

 $Y = XB + ZG + \varepsilon$

- Fixed Effects (Age, Sex, Block, etc)
 - Recorded
 - (X Matrix)

Maximum Likelihood Estimation of B and G



Assumes that genetic variance associated with each marker is equal to the total genetic variance divided by the total number of makers (Bayesian Ridge Regression)

How Well Does it Work?

Compare Accuracy of Genomic EBV (GEBV) With BLUP EBV (BEBV) Using Gene Level Simulations

Assumptions are Critical

- All MAS requires Linkage Disequilibrium (LD)
 - What is LD
 - How is it generated

Linkage Disequilibrium



LD=p(AB/ab)-p(Ab/aB)

Any factor changing the relative frequency of coupling *vs.* repulsion phase impacts LD

LD Generation

- All forces that change allele frequencies
 - Mutation
 - Migration
 - Selection
 - Genetic Drift

Starting Conditions

- Hardy-Weinberg Equilibrium
 - Within and between loci
 - Generate LD by Random Drift
- Mutation-Drift Equilibrium
 - Pre-existing LD
 - LD has not decayed from original mutation event
 - Dependent on mutation rates (types of makers) and population size

Mutation Drift Equilibrium (MDE) $N_e=100$ Generations=1000



Generations of Training Duration and Accuracy of Prediction



Accuracy=Correlation Between Predicted and True Breeding Value

Results

GEBV Accuracy with h²=.5

Starting in HWE

Starting in MDE



N=128, N_e=16

GEBV Accuracy with h²=.1

Starting in HWE

Starting in MDE



N=128, N_e =16

Effect of Random vs. Directional Selection on Accuracy

Starting in HWE





h²=.1 N=256, Ne=32, 100/100 Marker/QTL loci distributed on 100cM. (average over 60 replicates, SEM=.02).

Alternative Approach Genomic Relationship Matrix

- Assumes
 - Dense markers evenly spaced across the genome
 - Assumes markers are in LD with QTL affecting trait(s) of interest
 - Alike in State (AIS) alleles were at one time a result of a single mutation, thus IBD when traced back in evolutionary time
 - Each marker account for an equal proportion of genetic variance (infinitesimal model)
 - Genetic Effects are Normally Distributed

Compute (AIS) relationship matrix (G)

$$TA_{k} = 2 \frac{\sum_{i=1}^{2} \sum_{j=1}^{2} I_{ij}}{4}$$

$$\mathbf{G} = \boldsymbol{\sigma}_{A^*}^2 \mathbf{G}^*$$

 $\sigma^2_{{}_{A*}}$

 TA_k =total allelic relationship at kth locus TA_k =2x coefficient of relationship (Malecot. 1948)



Is the additive genetic variance associated with the markers for the trait



Note: with low marker density the markers may not capture any genetic variance

			LOCUS											
	A		E	3	(C	[)	E		Peo		gree	
Individual	1	2	1	2	1	2	1	2	1	2				
1	2	2	1	1	1	2	1	1	2	2		1	12	
2	1	2	1	2	2	2	1	2	1	1			$\langle \rangle$	
3	1	2	1	1	1	2	1	2	1	2			\rightarrow	4
4	2	2	1	1	2	2	1	1	2	1	3	4	5	6
5	2	1	1	2	2	2	1	1	2	1				
6	2	2	1	1	2	2	1	1	2	1				
dividuals (X,	,Y)	<u> </u>			1				1 - 2		Iotal	rxy		
x=1	$ ^2$	$\langle 2 \rangle$	$ $ $ $ 1 \rangle			2	$ $ $ $ 1 \rangle	$\begin{bmatrix} 1 \\ . \end{bmatrix}$	2	2				
y=1		▶ 2↓		▶ 1↓	↓1 <u>~</u>	▶ 2↓	ļ1 🖌	▶ 1↓		▶ 2				
sum		4		4		2		4		4				
hared allele	es	2		2		1		2		2	9	1.8		
x=1	2	/ 2	1	1	1	2	1	1	2	/ 2				
y=2	1	2		▶ 2	2	► 2↓	↓1 ▲	2	1	▶ 1				
sum		2		2		2		2		0				
hared allele	es.	1		1		1		1		0	4	0.8		
						•				•		0.0		
		AIS	G=GRM							IBD	PEDIGREE	А		
	1	2	3	4	5	6								
1	1.8 👞	0.8	1.2 👞	1.6	1.2 🔨	1.6 👞		x 1	_ 0	0.5	0.5	0,5	Q.5	
2	0.8	1.4	1	1.2	4.2	1.2		0 /	1	0.5	0 † 5	Ø.5	0.5	
3	1.2		1.2	1.2	1.8	1.2		0.5	0.5	1	0.5	0.5	0.5	
4	1.6	1.2	1.2	1.8	1.4	1.8	\searrow	0.5	0.5	0.5	h	/ 0.5 /	0.5	
5	1.2	1.2		1.4	1.4	1.4	\bowtie	8.5	0.5	0.5	Ø.5	1/	0.5	
6	1.6	1.2	1.2	1.8	1.4	1,8	$ \searrow $	Q.5	0.5	0.5	þ.5 /	Ø.5	1	
				$\overline{\}$			/	\sum	$\overline{}$			/		
									\sim		$\backslash / / /$			
					<				\sim	\mathbf{N}	$\backslash / / /$			
										$ \longrightarrow $				
											\mathbb{N}			
	P	arents assu	umed not re	elated (Fals	e) F	Parents ass	umed non i	nbred (false	e) Ful	l sibs ass	umed = relati	onship (fal	se)	
														1
													2	. I

G* Computed Directly from M

	LOCUS												
		A		В		С	0)		E		22=1	
Individual	1	2	1	2	1	2	1	2	1	2		12=0	
1	2	2	1	1	1	2	1	1	2	2		11=-1	
2	1	2	1	2	2	2	1	2	1	1			
3	1	2	1	1	1	2	1	2	1	2			
4	2	2	1	1	2	2	1	1	2	1			
5	2	1	1	2	2	2	1	1	2	1			
6	2	2	1	1	2	2	1	1	2	1			
			М	N individua	als x p marl	kers			M'	p markers	x N individ	uals	
1	1	-1	0	-1	1		1	0	0	1	0	1	
2	0	0	1	0	-1		-1	0	-1	-1	0	-1	
3	0	-1	0	0	0		0	1	0	1	1	1	
4	1	-1	1	-1	0		-1	0	0	-1	-1	-1	
5	0	0	1	-1	0		1	-1	0	0	0	0	
6	1	-1	1	-1	0								
	0.8	-0.2	0.2	0.6	0.2	0.6		1.8	0.8	1.2	1.6	1.2	1.6
	-0.2	0.4	0	0.2	0.2	0.2		0.8	1.4	1	1.2	1.2	1.2
	0.2	0	0.2	0.2	0	0.2		1.2	1	1.2	1.2	1	1.2
	0.6	0.2	0.2	0.8	0.4	0.8		1.6	1.2	1.2	1.8	1.4	1.8
	0.2	0.2	0	0.4	0.4	0.4		1.2	1.2	1	1.4	1.4	1.4
	0.6	0.2	0.2	0.8	0.4	0.8		1.6	1.2	1.2	1.8	1.4	1.8
			MM'/5							G*			
			dimensior	n nxn		Note that	1+ MM' /5= G	*					
							The are the	e same					
													22

Mixed Model Equations



Note one can multiply both sides of the second equation by G to avoid an inverse



Note, only ¹/₂ the additive genetic variance was captured by the markers ²⁴

Y				Z								
7		1	0	0	0	0	0	V(A)=	5			
9		0	1	0	0	0	0	V(E)=	20			
10		0	0	1	0	0	0					
6		0	0	0	1	0	0		XX			
9		0	0	0	0	1	0		6			
11		0	0	0	0	0	1					
Х				MM'/5					ΧY			
1		0.8	-0.2	0.2	0.6	0.2	0.6		52			
1		-0.2	0.4	0	0.2	0.2	0.2					
1		0.2	0	0.2	0.2	0	0.2					
1		0.6	0.2	0.2	0.8	0.4	0.8					
1		0.2	0.2	0	0.4	0.4	0.4					
1		0.6	0.2	0.2	0.8	0.4	0.8					
· · ·						- · ·						
√(A*)G*7'X				V(A*)G7'7+	V(F)I				V(A*)GZY			
11		24	-1	1	3	1			89			
4			22		1	1	1		37			
4		1	0	21	1	0	1		34			
15		3	1	1	24	2	4		126			
8		1	1		2	22	2		68			
15		3	1	1	<u> </u>	22	24		126			
10				· · ·		2	<u> </u>		120			
6	1	1	1	1	1	1	h		52			
11	24	-1	1	י א	1	י ז	U1		80			
		22	0	1	1	1			37			
4	1	0	21	1	0	1	113		3/			
15	2	1	1	2/	2	1	114		126			
2	1	1	۱ ۵	24	2	4	u 4 u5		68			
15	2	1	1	Z	22	21 21	116		126			
10	3	1	1	4 1 HS	2	24	<u>u0</u>		RHC			
									NH0			
h		0 220260	_0 00709	_0.01022	0 01027	_0_00620	0 00000	-0 00620	50		9 760004	
		0.230200	-0.00190	-0.01022	0.01037	-0.00029	-0.00000	-0.00029	90 90		_0.702304	
		-0.07000	0.040009	0.000799	0.001607	-0.00240	-0.001000	-0.00240	27		-0.20929	
		0.00000	0.000001	0.047204	0.001027	0.0007		-0.00007	24	_1	0.034400	
u3	1	-0.03092	-0.00058	0.001379	0.04919	-0.00075	0.001005	-0.00075	34	=	-0.02194	
U4	=	-0.11254	-0.00078	0.003063	0.003334	0.040055	0.001074	-0.00335	60		-0.1048	 07
US		-0.06107	0.000807	0.000747	0.003012	-0.0015	0.048432	-0.0015	68		-0.05796	 25
u6		-0.11254	-0.00078	0.003063	0.003334	-0.00335	0.001074	0.046655	126		-0.1648	

Equivalent Model Estimation of Marker effects

$$\begin{bmatrix} \mathbf{X'}_{1,N} \, \mathbf{X}_{N,1} & \mathbf{X}_{1,N}' \, \mathbf{M}_{N,p} \\ \mathbf{M'}_{p,N} \, \mathbf{X}_{N,1} & \mathbf{M'}_{p,N} \, \mathbf{M}_{N,p} + \frac{\sigma_e^2}{\left(\frac{\sigma_{A^*}^2}{L}\right)} \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_{1,1} \\ \mathbf{g}_{p,1} \end{bmatrix} = \begin{bmatrix} \mathbf{X'}_{1,N} \, \mathbf{Y}_{N,1} \\ \mathbf{M'}_{p,N} \, \mathbf{Y}_{N,1} \end{bmatrix}$$

Dimension pxp
Each Marker effects is solved for

Model of Meuwissen et al (2001)

$$\hat{u}_i = GEBV_i = \mathbf{Mg} = \sum_j M_{ij} \hat{g}_j$$

		M'						Μ					7
1	0	0	1	0	1		1	-1	0	-1	1		-
-1	0	-1	-1	0	-1		0	0	1	0	-1		
0	1	0	1	1	1		0	-1	0	0	0		
-1	0	0	-1	-1	-1		1	-1	1	-1	0		
1	-1	0	0	0	0		0	0	1	-1	0		
		MM'					1	-1	1	-1	0		
3	-3	2	-3	1			M'X		M'Y		Y	Х	
-3	4	-2	3	-1			3		24		7	1	
2	-2	4	-3	-1			-4		-34		9	1	
-3	3	-3	4	-1			4		35		10	1	
1	-1	-1	-1	2			-4		-33		6	1	
		1					0		-2		9	1	
1	0	0	0	0							11	1	
0	1	0	0	0									
0	0	1	0	0			V(A)=	5			Χ'Χ		
0	0	0	1	0			V(E)=	20			6		
0	0	0	0	1									_
		LHS						RHS			ΧΎ		_
6	3	-4	4	-4	0	В		52			52		_
3	23	-3	2	-3	1	g1		24					_
-4	-3	24	-2	3	-1	g2		-34					_
4	2	-2	24	-3	-1	g3	='	35					_
-4	-3	3	-3	24	-1	g4		-33					_
0	1	-1	-1	-1	22	g5		-2					_
				inverse(LH	S)			RHS					_
В		0.238268	-0.02055	0.030921	-0.03165	0.029414	0.002238	52		8.762384			_
g1		-0.02055	0.046795	0.002074	-0.00012	0.002068	-0.00194	24		-0.08491			_
g2	='	0.030921	0.002074	0.047116	-0.00139	-0.00057	0.001958	-34	='	0.021935			_
g3		-0.03165	-0.00012	-0.00139	0.047031	0.00085	0.002119	35		0.012177			_
g4		0.029414	0.002068	-0.00057	0.00085	0.047091	0.002059	-33		0.070134			_
g5		0.002238	-0.00194	0.001958	0.002119	0.002059	0.045822	-2		-0.08231			_
													_
				M			g		0.05000				_
<u>u1</u>		1	-1	0	-1	1	-0.08491		-0.25929				_
u2		0	0	1	0	-1	0.021935		0.094488		,		_
<u>u3</u>	='	0	-1	0	0	0	0.012177	='	-0.02194	same as b	etore		_
u4		1	-1	1	-1	0	0.070134		-0.1648				
u5		0	0	1	-1	0	-0.08231		-0.05796				_ 2
u6		1	-1	1	-1	0			-0.1648				

Genomic Selection in Poultry, Results with Broilers and Comparison with Traditional BLUP

Acknowledgements

- This project was supported by National Research Initiative Competitive Grant no.2008-35205-04723 and Agriculture and Food Research Initiative Competitive Grant no. 2009-65205-05665 from the USDA National Institute of Food and Agriculture
- Lead Project Director:
- Hans H. Cheng, USDA, ARS, East Lansing, MI
- •

•

- Co-Investigators:
- William Muir Purdue U., West Lafayette, IN
- Ron Okimoto, Terry Wing, Sean Maceachern, Rachel Hawken Cobb Vantress, Siloam Springs, AR
- Albert Paszek now with Norsvin USA
- Gerard Albers and Addie Vereijken Hendrix Genetics, The Netherlands
- Theo Meuwissen Norwegian U. of Life Sciences, Norway
- Guilherme Rosa U. of Wisconsin, Madison, WI
- Ignacy Misztal , King, Joy, Chen, C. Y., U. of Georgia, Athens, GA
- Martien Groenen Wageningen U., The Netherlands
- •
- Collaborators:
- Charles Pick and Tun-Ping Yu DNA Landmarks, Montreal, Canada

Potential Benefits

- Reduced Generation Interval
 - Select breeders as chicks
 - Housing and feed savings
- Increased accuracy
 - Low heritability traits
- Alternative for difficult or expensive traits to measure
 - Feed efficiency
 - Carcass composition
 - Disease resistance
 - Can address animal wellbeing concerns

Design of experiment

- Lines
 - Male line
 - Female line
- Traits of selection
 - Breast Meat (high h²)
 - Weight (medium h^2)

Methods for estimation of breeding values (EBVs and GEBVs)

- EBV's
 - BLUP (multi-trait)
- GEBVs
 - GBLUP (multi-trait)
 - Single Step (ssGBLUP)
 - Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.* 2010 Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal Of Dairy Science **93**: 743-752.
 - Legarra, A., I. Aguilar, and I. Misztal, 2009 A relationship matrix including full pedigree and genomic information. Journal Of Dairy Science 92: 4656-4663.
 - Advantages
 - Corrects for multi-trait selection bias
 - Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra, 2011 Bias in genomic predictions for populations under selection. Genetics Research 93: 357-366.
 - Utilizes records from ungenotyped animals
 - Concurrent
 - Ancestral

EBVs and GEBVs combined using an index

- Equal weight to all traits
- Standardized relative to additive variance (standard deviation)

GBLUP and **BLUP** Training

- Two training generations
 - no selection
 - used historical phenotypes
 - Banked DNA samples
- Numbers genotyped (GBLUP)
 - 2,500 each line
- Phenotypes
 - BLUP and GBLUP
 - 280,000 each line

Selection Program-GEBV

- Tiered
 - Tier 1
 - 800 Genotyped and Phenotyped
 - Tier 2
 - 200 Phenotyped
- Plays into strength of ssGBLUP
 - Uses all records
- Only birds from Tier 1 selected based on index
- Number selected
 - 20 Males
 - 200 Females

Selection Program BLUP

- Tiered
 - Tier 1 (800) phenotyped
 - Tier 2 (200) phenotyped
- Only birds from Tier 1 selected based on index
- Number selected
 - -20 Males
 - 200 Females

Duration

- 3 generations of selection
- Generation 4
 - Expanded to 4,000
 - Phenotyped only
 - Progeny test of Generation 3
 - Bases for comparison between methods

Accuracy

- Correlation between the true and predicted breeding values
 - Don't know true EBV
- Equivalent formula
 - Legarra, A., C. Robert- Granie´, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180: 611-618
 - Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir, 2011 Effect of different genomic relationship matrices on accuracy and scale. Journal Of Animal Science 89: 2673-2679
 - Does not require true value of EBV to be known
 - Requires heritability to be known
 - Calculation of EBV does not include phenotypes of the generations they were estimated in

 $r(EBV, EB\hat{V})$

 $r(EB\hat{V}, Y)$

h

Results

Generations 3-4

Accuracy



Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir, 2011 Effect of different genomic relationship matrices on accuracy and scale. Journal Of Animal Science **89**: 2673-2679

Genetic Trend Female Line Index



Genetic Trend Male Line Index



Challenges

GBLUP lines

- Some mismatch in early generations between DNA samples, pedigree, phenotypes
- Impact $\Delta G = \text{accuracy *intensity} * \sigma_a$

Reduced selection intensity in GBLUP lines

- Mismatches were random
- Proportion of those selected was therefore at random
- Reduced accuracy in GBLUP program
 - Training and SNP effects were updated each generation including current generation
 - Inaccurate SNP effects (phenotype- genotype relationships were incorrect)



- Blood sampling on a large scale by farm workers in a chicken house.
 - Sample identification and tracking
 - Bar-coding, portable scanners and printers suitable for a chicken house
- Timely pedigree checking
 - Parent and offspring pedigree checking where possible
 - Full sib pedigree checking when parent information is not available
- Sample processing and genotyping to allow timely error checking
 - Timing is critical in chicken breeding (weeks not months)
 - Ability to keep selection candidates for pedigree verification if necessary
 - Space for storing selection candidates
 - Very different to cattle breeding..less space, less time

Conclusions

- GBLUP more accurate than BLUP
- Genetic trends reflect increased accuracy of GBLUP
- GBLUP at a minimum was able to keep up with BLUP[©]
- Quality control essential for translation of technology to applications
- For some traits, especially those measured late in life, being as good as BLUP is good enough

Conclusions

- Results are very encouraging
- Company will continue GBLUP program, particularly for
 - low heritability traits
 - Traits difficult or expensive to measure
 - Traits measured late in life cycle