

EADGENE European Animal Disease Genomics Network of Excellence for Animal Health and Food Safety

Genomics for Animal Health: Outlook for the Future
 13- 14th October 2009, Muséum National d'Histoire Naturelle, Paris, France



The scientific view of genomic selection
 Theo Meuwissen, Ås, Norway

This publication represents the views of the Authors, not the EC. The EC is not liable for any use that may be made of the information.

EADGENE

Introduction

- Genome sequence available in humans and many livestock species
- (By)product: massive numbers of SNPs
 - available through 2nd generation sequencing techn.
- Dramatic improvement of SNP genotyping
 - <<0.01\$ / SNP genotype
 - ~200\$ for 50k SNPs (for 500k SNPs)

Institute for Animal and Aquacultural Sciences

EADGENE

Aim

- How are we going to use this new technology in animal breeding?
 - For the prediction of genetic value
- Same aim:
 - How to use in plant breeding (outcross species)
 - Human genetic counselling for diseases
- Assuming complex trait:

$$P = \sum g_i + E$$

Institute for Animal and Aquacultural Sciences

EADGENE

How to use new genomics technology?

- Genomic selection:
 - Using dense SNP genotyping
 - ideally: some SNPs so close to genes, that they perfectly explain all genes
 - Perfect linkage disequilibrium (LD) with genes
 - Dense SNPs => all genes covered
 - Simultaneously estimate the effect of the SNPs
 - Select for SNP effects
 - no phenotypes required

Institute for Animal and Aquacultural Sciences

EADGENE

Genomic selection : the method in detail

(Meuwissen et al. 2001)

Genomics for Animal Health: Outlook for the Future
 13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France

EADGENE

Genome-wide EBV: $EBV = \sum CSE$

- CSE = Chromosome Segment Effect
 - CS identified by dense marker haplotypes
 - eg. 1 cM segments
 - Now: single SNPs
 - CSEs estimated in experimental data set
 - problem: ~100,000 CSEs and ~1,000 records ($k > n$)
 - **CSEs are not tested for significance**
 - Otherwise many real effects would be lost
- relies on LD to associate markers & QTL

Institute for Animal and Aquacultural Sciences

EADGENE

Estimation of CSE:

- QTL-MAS approach
 - Test CSE one-by-one for stat. signif.
 - Find biggest CSE
- GW-BLUP:
 - $V(\text{CSE}_i) = \text{constant}$
- BayesA:
 - $V(\text{CSE}_i) \sim \chi^2(\nu, S)$
- BayesB:
 - $V(\text{CSE}_i) = 0$ with prob. = p
 - $\sim \chi^2(\nu, S)$ with prob. = $(1-p)$

7

Institute for Animal and Aquacultural Sciences

EADGENE

Simulation Test

- 10 chroms of 100 cM
- 101 equally spaced markers per chromosome
- 100 potential QTL positions in middle between markers
- 1000 generations & $N_e = 100$:
 - mutation - drift balance
 - linkage disequilibria between markers & QTL

8

Institute for Animal and Aquacultural Sciences

EADGENE

Simulation Test (2)

- Gen. 1001 : 200 anim + Phenot + Genot
- Gen. 1002 : 2000 anim + Phenot + Genot
- Gen. 1003 : 2000 anim + Genot
- AIM : predict genotypic value of gen. 1003
 - makes selection without phenotypes possible
 - $EBV_i = \sum_j CSE_{ij}$

9

Institute for Animal and Aquacultural Sciences

EADGENE

Accuracy of Selection

	corr(TBV;EBV)
QTL-MAS approach	0.36
GS-BLUP	0.72
BayesA	0.80
BayesB	0.84

10

Institute for Animal and Aquacultural Sciences

EADGENE

Statistical Methods for GS

Genomics for Animal Health: Outlook for the Future
 13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France

EADGENE

BLUP of marker effects (GW-BLUP)

- Estimate effect of every SNP
 - Prior distribution: normal with constant variance
- Model:

$$y = X_m m + e$$

$$\text{Var}(m) = I \sigma_m^2$$

$$\text{Var}(y) = X X' \sigma_m^2 + I \sigma_e^2$$

$$(\text{Var}(y) = A_{ped} \sigma_a^2 + I \sigma_e^2) \leftarrow \text{Trad-BLUP}$$

12

(Habier et al. 2007; Goddard & Hayes 2007)

Institute for Animal and Aquacultural Sciences

EADGENE

Good idea to replace A_{ped} by XX' ?

- A_{ped} is only estimate of A_{true}
- XX' can be a better estimate of A_{true}
 - Given sufficient markers
- Alternative model to calculate GWBLUP:
 - traditional animal model with $A=XX'$
 - Much fewer Eqn
 - Easier to deal with missing genotypes

13

Institute for Animal and Aquacultural Sciences

EADGENE

BayesA

- Same model : $y = Xm + e$
- Prior distribution:
 - t-distribution (more fat-tailed than normal)
- Gibbs-sampling implementation:
 - Sample variance of SNP effect from inverse- χ^2
 - Sample SNP effect given this variance
 - So: relaxes the assumption of constant variance

14

Institute for Animal and Aquacultural Sciences

EADGENE

BayesB

- Prior distribution for SNP effect m_i :
 $m_i \sim$ t-distribution with prob. p
 $m_i = 0$ with prob. $(1-p)$
- Makes biologically sense:
 - Many SNPs are not close to QTL so no effect
 - Sequence data: only causative SNPs have effect

15

Institute for Animal and Aquacultural Sciences

EADGENE

BayesC

- Prior distribution for SNP effect m_i :
 $m_i \sim N(0, \sigma_{big}^2)$ with prob. p
 $m_i \sim N(0, \sigma_{small}^2)$ with prob. $(1-p)$
- Implementable by Gibbs
- small effects together act as a polygenic effect

16

Institute for Animal and Aquacultural Sciences

EADGENE

GWBLUP vs. BayesA/B/C

- Simulation studies:
 $GWBLUP < BayesA < BayesB/C$
- Real data studies:
 - Differences small
 - Often GWBLUP best, or as good as BayesA/B
- Theoretically:
 - If few genes & dense SNPs: BayesB best
 - If many genes / not enough density: GWBLUP best

17

Institute for Animal and Aquacultural Sciences

EADGENE

LASSO vs BayesA vs BayesB

18

Institute for Animal and Aquacultural Sciences

EADGENE

Including Interactions

- Reproducing Kernel Hilbert Space Regression (Gianola et al., 2006)
 - Takes automatically account of all interactions
- However:
 - Dominance interaction is not inherited
 - V_{additive} is often close 100% of V_{total} (Hill et al 2008)
 - Freq of mutant alleles often rare
 - Very difficult to estimate

19

Institute for Animal and Aquacultural Sciences

EADGENE

Applications

Genomics for Animal Health: Outlook for the Future
 13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France

EADGENE

Idea for difficult trait

- (eg. Tastiness of meat)
- Perform experiment
- Estimate SNP effects
- Select for SNP effects forever after
- Too good to be true?

21

Institute for Animal and Aquacultural Sciences

EADGENE

Re-estimation of SNP effects important

Sonesson et al. 2008

Institute for Animal and Aquacultural Sciences

EADGENE

Accuracy at genome sequence density

Generation	causative SNP	
	Excl	Incl
same	0.806	0.826
10 later	0.806	0.824

(density: 33,000 SNPs/M; 200 training recs)

23

Institute for Animal and Aquacultural Sciences

EADGENE

Effect of SNP density on accuracy

- accuracy is linear with R
- top accuracy (.9) requires
 - $10 * Ne * L$ markers
 - $1 * Ne * L$ records

Institute for Animal and Aquacultural Sciences

Slide 24

t1 move this slide
 theome, 22/02/2009

EADGENE

Dairy cattle breeding

- Selection of young bulls (no progeny test)
 - Saves costs of progeny test: ~40,000 \$/bull
 - Reduce generation interval by factor 2
 - Also use GS on bull dams
 - Schaeffer (2006):
 - double ΔG
 - Saves \$23 million / yr
- BUT: need scheme to update marker effects

25
 Institute for Animal and Aquacultural Sciences

EADGENE

Pig breeding schemes

- GS of test-boars for maternal traits
 - Estimate marker effects on sibs of test-boars
 - Genotype test-boars and select for marker effects
- Use crossbred-practical performance info
 - Record trait in practice on crossbreds
 - GS in the nucleus for crossbred performance

26
 Institute for Animal and Aquacultural Sciences

EADGENE

Fish breeding schemes

- Replace disease challenge tests on sibs by GS
 - Estimate marker effects from practical disease data
 - case-control design
 - GS in the nucleus
- No family housing needed
- Problem: ~30,000 selection candidates
 - Need pre-selection for other traits before GS

27
 Institute for Animal and Aquacultural Sciences

EADGENE

Humans : missing heritability

- Human height, Crohn's disease, schizophrenia
 - Heritability high: 60-80%
 - Dense SNP genotyping: ~1 milj SNPs
 - Large data sets: ~60,000
 - detected SNPs/genes GWAS: 30-50
 - Genetic variance explained: ~5%
- Why: infinitesimal model (+ few large genes)
 - Many mutations, small effects
 - Genomic selection: all $V_g \Rightarrow$ better than 30 SNPs

Genomics for Animal Health: Outlook for the Future
 13-14th October 2009, Muséum National d'Histoire Naturelle, Paris, France

EADGENE

Conclusions

- GS is MAS on a genome-wide scale
 - By using GW-dense markers : use all V_g
- predict EBV on non-phenotyped animals
 - Very different breeding structures
 - No recording of phen./pedigree on breeding animals
 - Shortening of generation intervals
 - Functional traits -> sustainable breeding
- Statistical methods:
 - GWBLUP \Rightarrow BayesB

29
 Institute for Animal and Aquacultural Sciences

EADGENE

Conclusions (2)

- No. of SNPs: $10 * N_e L \Rightarrow$ genome seq.
- No. of records: $1 * N_e L$
- Continuous updating of marker effects
 - with small SNP chip
 - Not if genome sequence data available
 - Then: 1 big reference population with lots of phenotypes and genotypes
- GS in (human) disease risk prediction:
 - Addresses all V_g instead of limited fraction

30
 Institute for Animal and Aquacultural Sciences