
# Genomic Selection
# From Livestock to Aquaculture
# Applications

John Woolliams

Alan Archibald, John Hickey & Ross Houston

The Roslin Institute, University of Edinburgh, U.K.

THE UNIVERSITY *of* EDINBURGH

# Outline

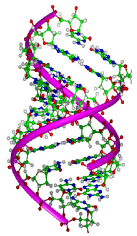
1. The State of Genomes
   - » livestock & aquaculture
2. Genomic Selection
   - » background & motivations
3. The State of the Art
   - » achievements & limitations
4. Genomic Selection 2.0
   - » sequencing populations and the final frontier?
5. Conclusions

1. The State of Genomes
   - Livestock
   - Aquaculture

# Genomic Selection & Sequence

- Genomic selection does not require a reference genome
  - » abbreviated GS
- BUT past & future tool development benefits from knowledge of sequence
  - » e.g. selection of SNP for spacing and coverage
- Stage of development of GS in species is associated with development of genome sequence
  - » amongst other things!

# Early Days: Towards a Sequence



**1953**
Watson and Crick

**1977**
DNA sequenced ΦX174 5,386 nt

**1990**
Human Genome Project launched

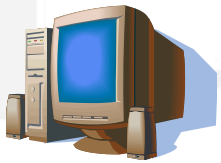
**1991**
PiGMaP project starts

'Halothane' gene test

**2001**
Draft human genome sequence

**1920s and 30s**
Fisher, Lush and others Population Genetics

**1970s +**
Advances in quantitative analysis



**1990s +**
Quantitative trait locus (QTL) mapping

**2001**
Genomic selection proposed

# Genomic Developments to 2011

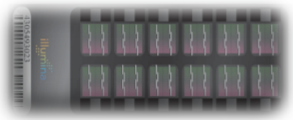
**2003**
Human genome sequence "finished"
$3 billion

**2004**
Chicken genome sequenced


**2005**
Dog genome sequenced

**2007**
Cat genome sequenced

**2008**
Bovine 50K SNP chip



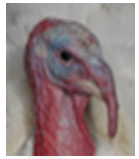
**2009**
Cattle genome sequenced

Horse genome Sequenced

**2009**
Pig 60K SNP chip

Sheep 60K SNP chip

**2010**
Turkey genome sequenced



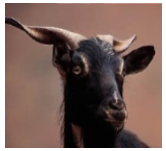
**2010**
750K bovine SNP chips

# Genomic Developments to 2015

**2012**
Chicken 600K SNP chip

**2012**
Pig genome sequenced


**2012**
Pig gene expression atlas



**2013**
Goat genome sequenced

**2013**
Duck genome sequenced

**2014**
Sheep genome sequenced

**2015**
Improved reference genomes – goat, pig, sheep, cattle, chicken,

**2015**
Pig 650K SNP chip

**2015**
Functional Annotation of Animal Genomes (FAANG) launched

THE UNIVERSITY OF EDINBURGH

BBSRC
20 Years of Pioneering
Great British Bioscience

# Tools for Genomic Selection

| | Dairy Cattle | Beef Cattle | Sheep | Goats | Pigs | Layer Chick's | Broiler Chick's | Turkey | Duck |
|---|---|---|---|---|---|---|---|---|---|
| Reference genome | +++ | +++ | ++ | ++ | ++ | +++ | +++ | + | + |
| High-density SNP 'chips' available | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | No |
| Genomic Selection? | +++ | + | ++ | + | +++ | ++(+) | +++ | No | No |

# Genomic Developments to 2011

ROSLIN

**2003**
Human genome sequence "finished"
$3 billion

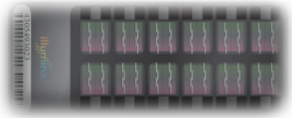
**2004**
Chicken genome sequenced


**2005**
Dog genome sequenced

**2006**
Salmon & trout cDNA micro-array

**2007**
Cat genome sequenced

**2008**
Bovine 50K SNP chip



**2008**
Sea bream cDNA micro-array

**2009**
Cattle genome sequenced

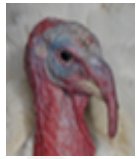
Horse genome Sequenced

**2009**
Pig 60K SNP chip

Sheep 60K SNP chip

**2009**
Salmon 6K SNP chip

**2010**
Turkey genome sequenced



**2010**
750K bovine SNP chips

**2010**
Sea bass cDNA micro-array

**2011**
Cod genome sequenced

**2011**
Salmon & trout RAD-sequencing

THE UNIVERSITY OF EDINBURGH

BBSRC
20 Years of Pioneering Great British Bioscience

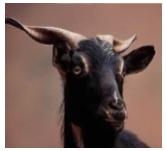
# Genomic Developments to 2015

ROSLIN

**2012**
Chicken 600K SNP chip

**2012**
Pig genome sequenced


**2012**
Pig gene expression atlas



**2013**
Goat genome sequenced

**2013**
Duck genome sequenced

**2014**
Sheep genome sequenced

**2015**
Improved reference genomes – goat, pig, sheep, cattle, chicken,

**2015**
Pig 650K SNP chip

**2015**
Functional Annotation of Animal Genomes (FAANG) launched

**2012**
Pacific oyster genome sequenced

**2013**
Salmon 130K SNP chip

**2014**
Trout genome sequenced

Salmon genome sequenced

Sea bass genome sequenced

Carp genome sequenced

Catfish HD SNP array

Carp HD SNP array

Rainbow trout HD SNP array

THE UNIVERSITY OF EDINBURGH

BBSRC
20 Years of Pioneering Great British Bioscience

# Tools for Genomic Selection

| | Atlantic salmon | Rainbow trout | European sea bass | European sea bream | Turbot | Common carp |
|---|---|---|---|---|---|---|
| Reference genome | + | + | + | (+) | (+) | + |
| High-density SNP 'chips' available | Yes | Yes | No | No | No | Yes |
| Genomic Selection? | ++ | (+) | No | No | No | No |

FISHBOOST
The next level of aquaculture breeding

SANTIAGO DE COMPOSTELA 2015

BBSRC
20 Years of Pioneering Great British Bioscience

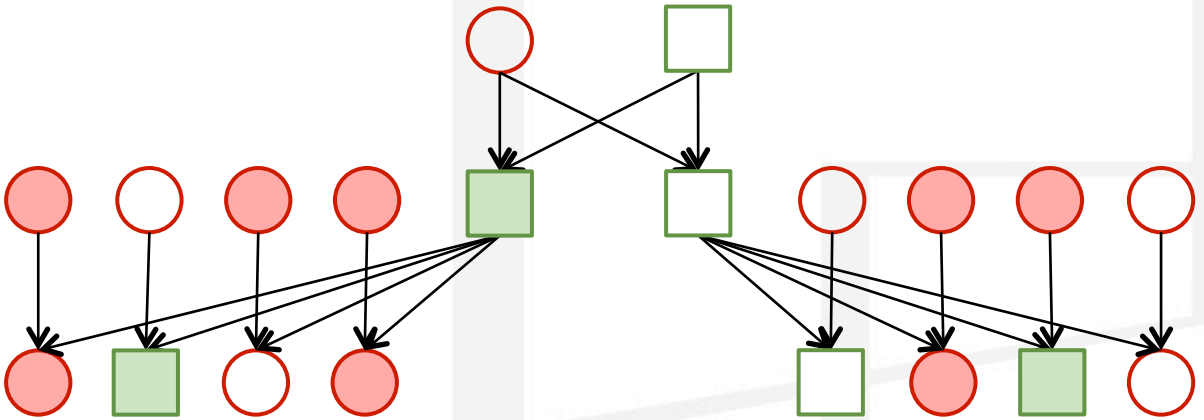
# Summary of Development

- Livestock species
  - Sequences published, refined
  - Expression atlases published and/ or major community efforts
  - Large number of SNP chips
    - multiple densities for most species to suit needs
    - imputation, GWAS etc
  - Genomic selection firmly established in major species
- Aquacultural species
  - Sequences at first draft
  - Early stage annotation and expression data
  - Some species with SNP chips
  - Genomic selection being introduced to some species

2. Genomic Selection
   - Background
   - Motivations

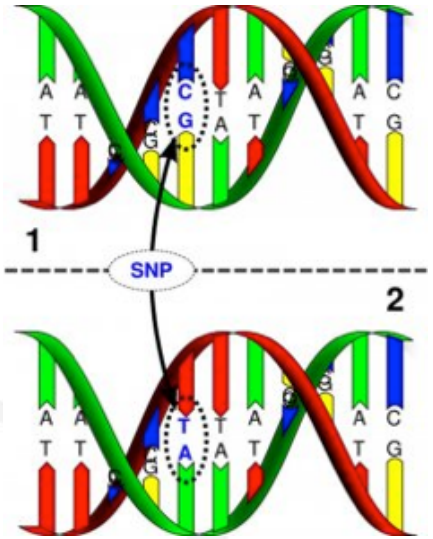
# Traditional Breeding

- Collect pedigree (○ □) phenotype (○ ●) information
  - » own performance (e.g. growth, …)
  - » sib-performance (e.g. carcass and meat quality, disease, …)
  - » progeny performance (e.g. milk yield, # eggs, …)
- Estimate genetic values by BLUP



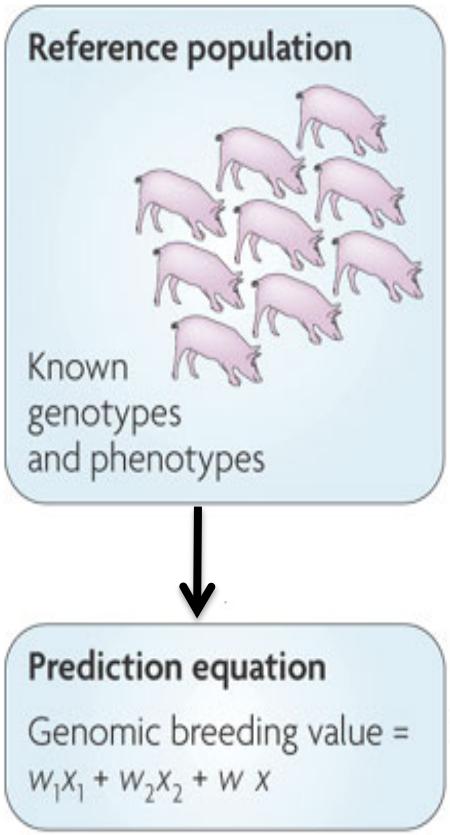
# Genomic Selection (1)

- Use abundant genome-wide markers
  - » e.g. SNP



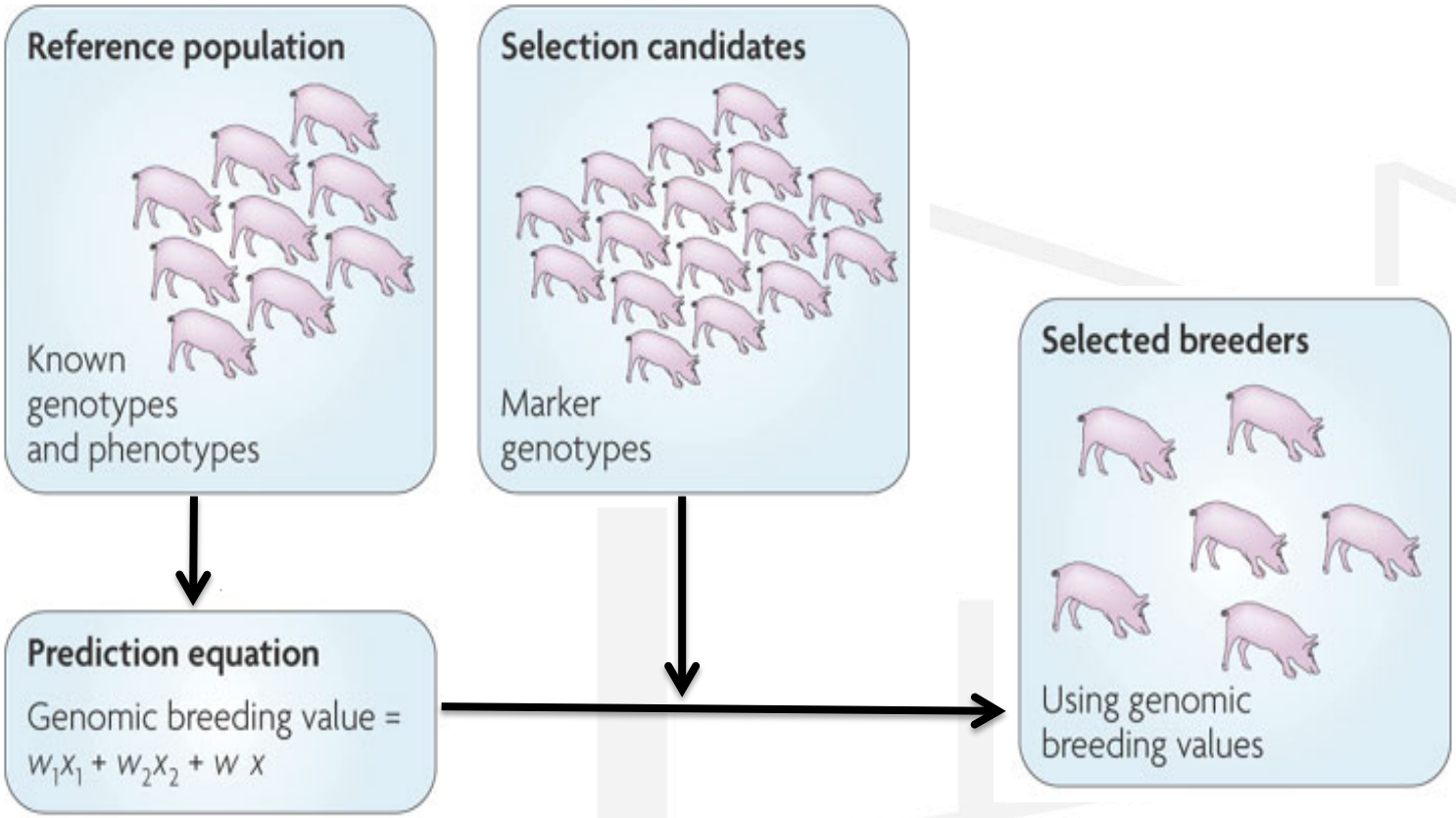
# Genomic Selection (2)

- Obtain predictions of Marker effects from Training Set



# Genomic Selection (3)

- Obtain predictions of candidates and select




Hayes & Goddard, 2009

# Some Important Properties

1. Accuracy of GS is dynamic
   - » bigger the training set, the more accurate the prediction
   - » BLUP accuracy is ~ static given structure of scheme
     - » f(data collected on close relatives, $h^2$)

# Some Important GS Properties

- Greater accuracy of Mendelian Sampling Term
  - » what makes us unique
  - » ancestors (e.g. sire) or collateral relatives (sibs) alone cannot estimate this
  - » source of continued gain beyond best parent
  - » selective advantage when controlling inbreeding
  - » explains ½ (or more) of genetic variance!
    - » i.e. within-family variance

# Some Useful GS Properties

- Prediction available early in life
  - » can consider selection in the absence of phenotypes
    - » e.g. disease phenotypes
  - » reduce generation interval

# GS & Gain

$$\text{Response} = \frac{\text{Selection Intensity x Accuracy x Diversity}}{\text{Generation Interval}}$$

- So GS adds gain through
  - » increases accuracy
  - » decreases generation interval
  - » can in some schemes add selection intensity
    - » e.g. avoiding random culling

3. The State of the Art
   - Livestock
   - Aquaculture

# GS Pioneers

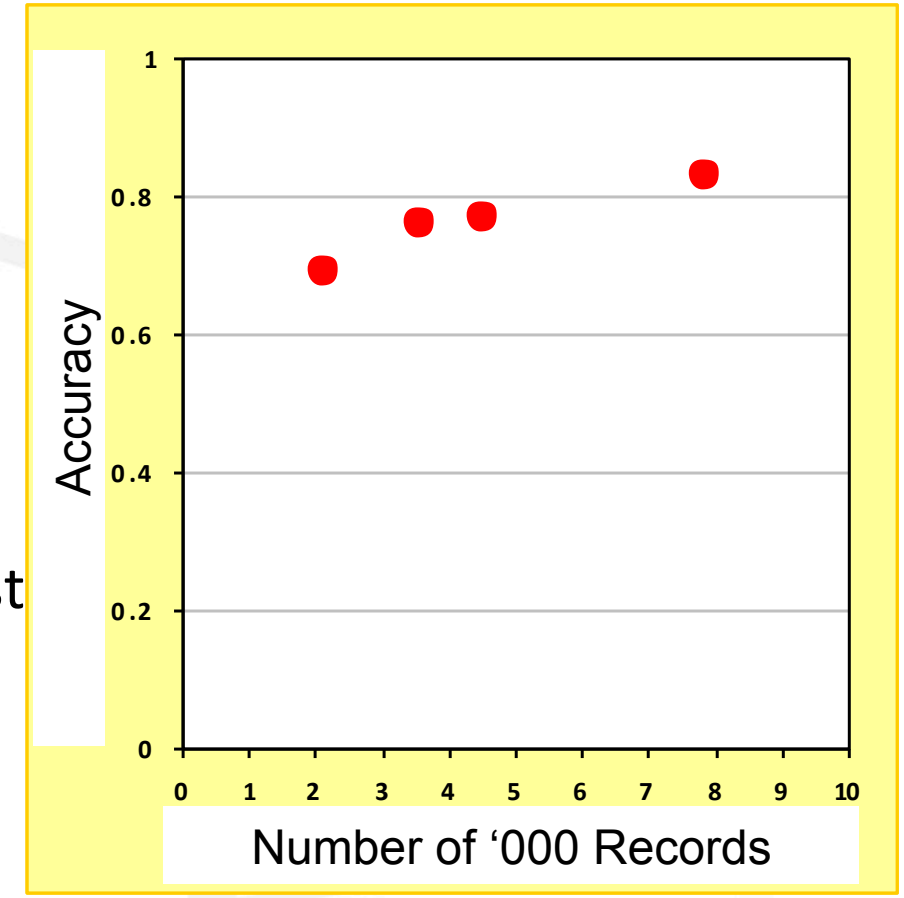
- Dairy breeding
  - » traditionally relied on progeny testing
  - » long generation interval
  - » high cost per bull tested

# GS Benefits in Dairy

- Benefits from prediction early in life
  - » screen more bulls more accurately before entering progeny test
    - » increase selection intensity
  - » only test the best bulls
  - » if GS predictions have high accuracy might avoid progeny test
    - » decrease generation interval

# GS Accuracy for Dairy

- Note
  - » high accuracy attained
    - » source USDA
  - » dynamic accuracy
- But
  - » require large number of bulls
  - » each bull has high accuracy test



# Dairy Advantages

- Costs of genotyping are covered by savings in progeny test
- Easy availability of large numbers of high-accuracy phenotypes on bulls
  - » i.e. the progeny test proofs
- Gains are potentially very large
- Safe implementation route as predictions on bulls can be compared emerging, accurate test results
- Animals are high value compared to cost of test

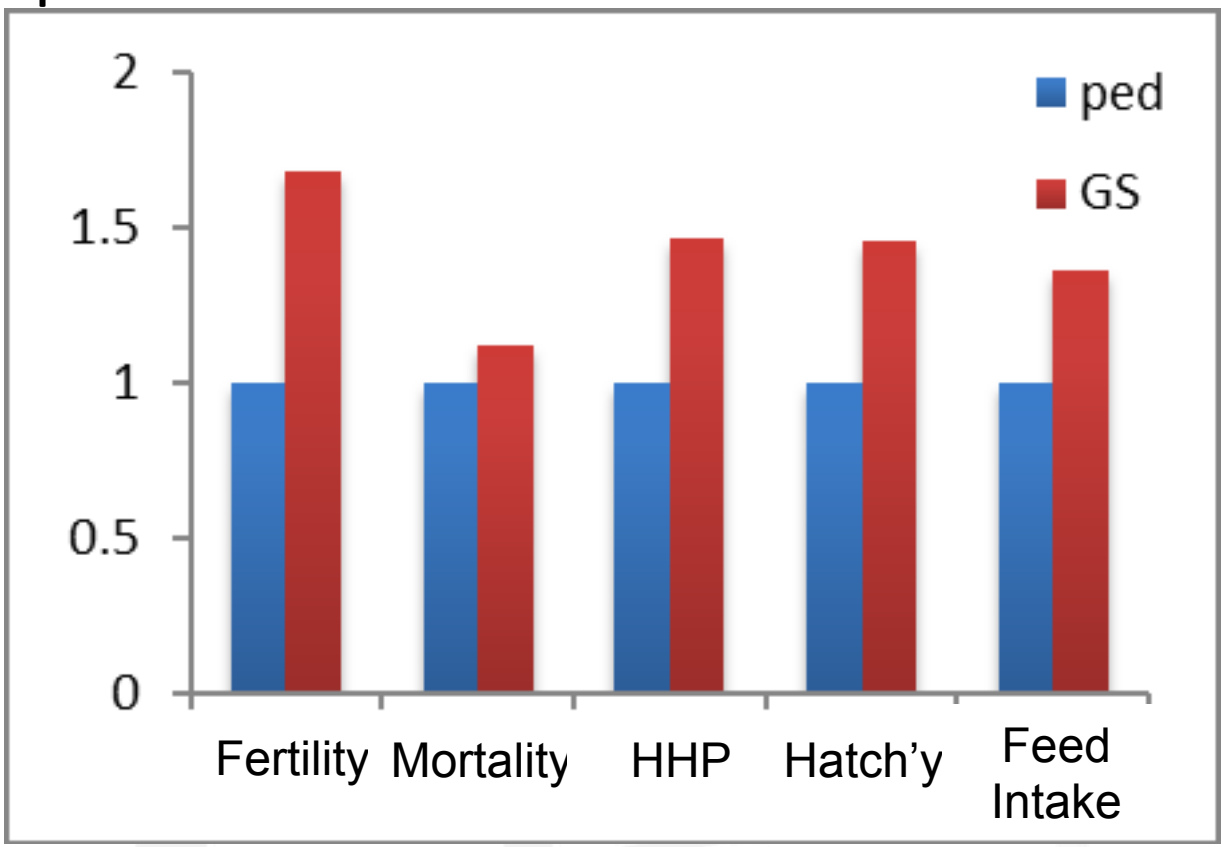
# Dairy Advantages

- Costs of genotyping are covered by savings in progeny test
- Easy availability of large numbers of high-
  phenotypes on bulls
  - » i.e. th
- G potentially very large
- Safe implementation route as predictions on bulls can be compared emerging, accurate test results
- Animals are high value compared to cost of test

What about sectors where animals have low individual value?

# GS Benefits in Broiler Chickens

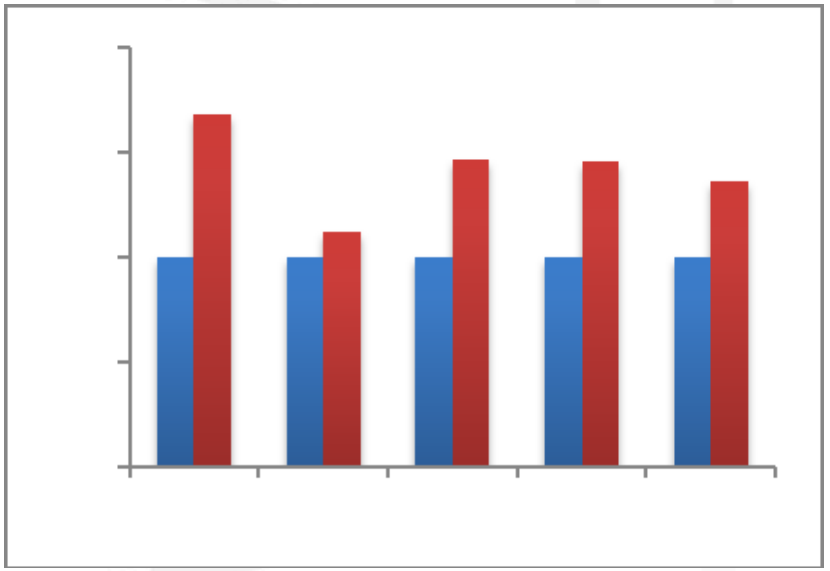
- Accuracy of GS compared to BLUP
  - » ~ 1.1 to 1.7
  - » will gain with time




From Wolc et al. 10th WCGALP, Vancouver, 2014

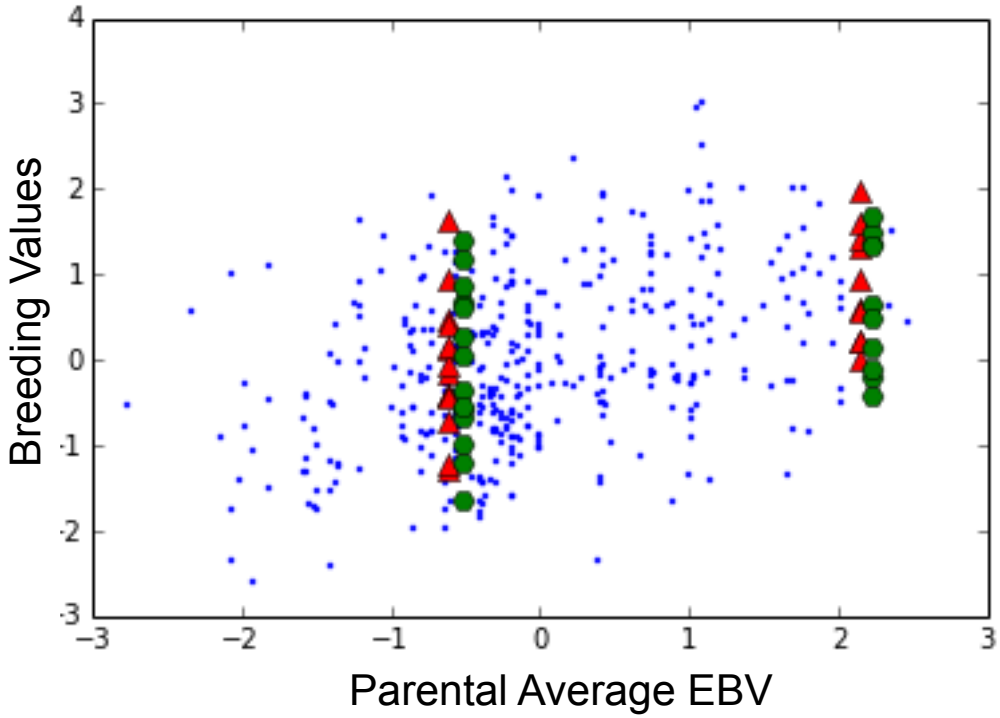
# GS Benefits in Broiler Chickens

- Valuable gains in sex-limited female traits with low $h^2$
  - » from estimate of Mendelian Sampling Term of males
  - » i.e. with no phenotype



# Value of Mendelian Sampling

- Red triangles: Mendelian genomic predictions
- Green circles: Validation from progeny means




From Wolc et al. 10th WCGALP, Vancouver, 2014

# Salmon Breeding

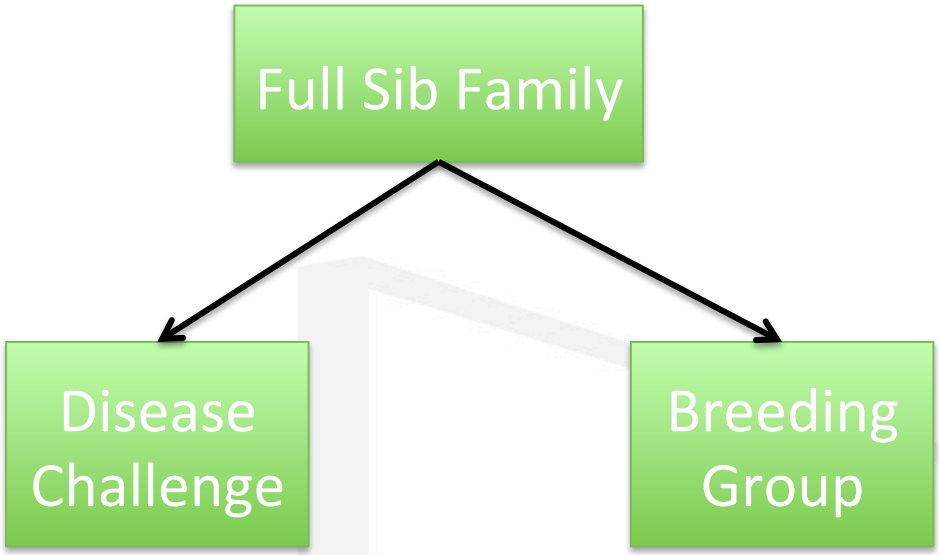
- Published results on GS are few
- Major study by Ødegård (2014)
  - » presenting this conference

# Salmon Breeding

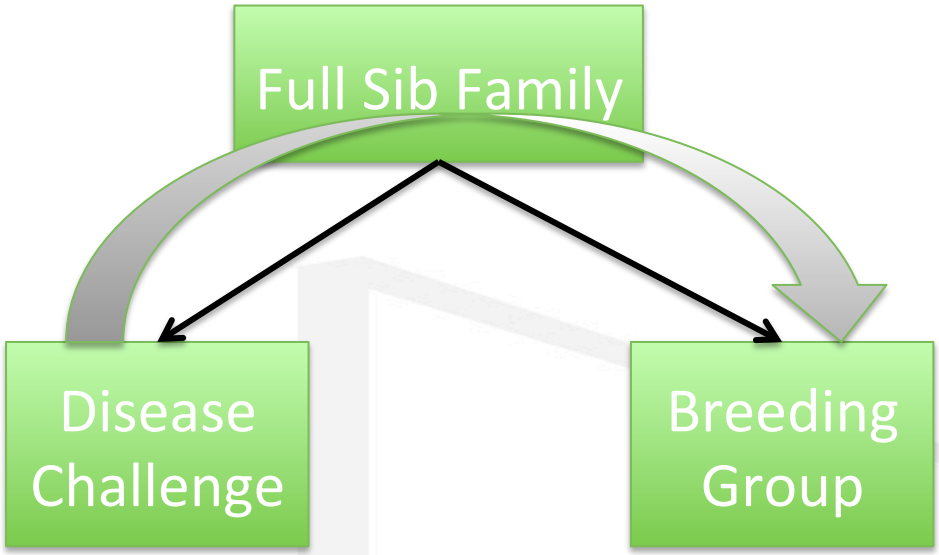
- Small to moderate training set size of 2000
- Accuracies from GS were better than BLUP
  - » 1.2-fold for Lice Resistance
  - » 1.1-fold for Fillet Colour
    - » BLUP for both traits was ~0.6

Ødegård et al. *Frontiers in Genetics* 5 (2014)

# Sib Testing

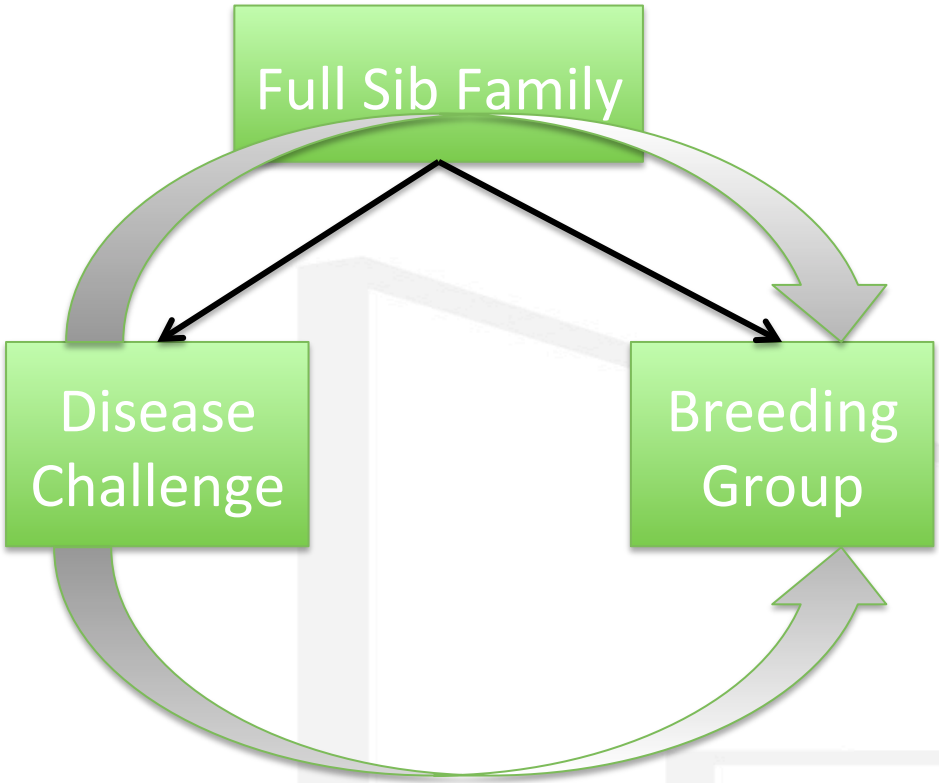


# Sib Testing with BLUP




- Information gained from test is only family mean

# Sib Testing with GS




- Information on family mean and desirable marker alleles within family both transferred to breeding group

# Sib Testing with GS

- Simulations show important benefits in accuracy
  - » estimate of Mendelian Sampling Term from markers
  - » require relatively few markers per chromosome (10-20)

Lillehammer et al (2013), Genet Sel Evol. 45

# Summary of Applications

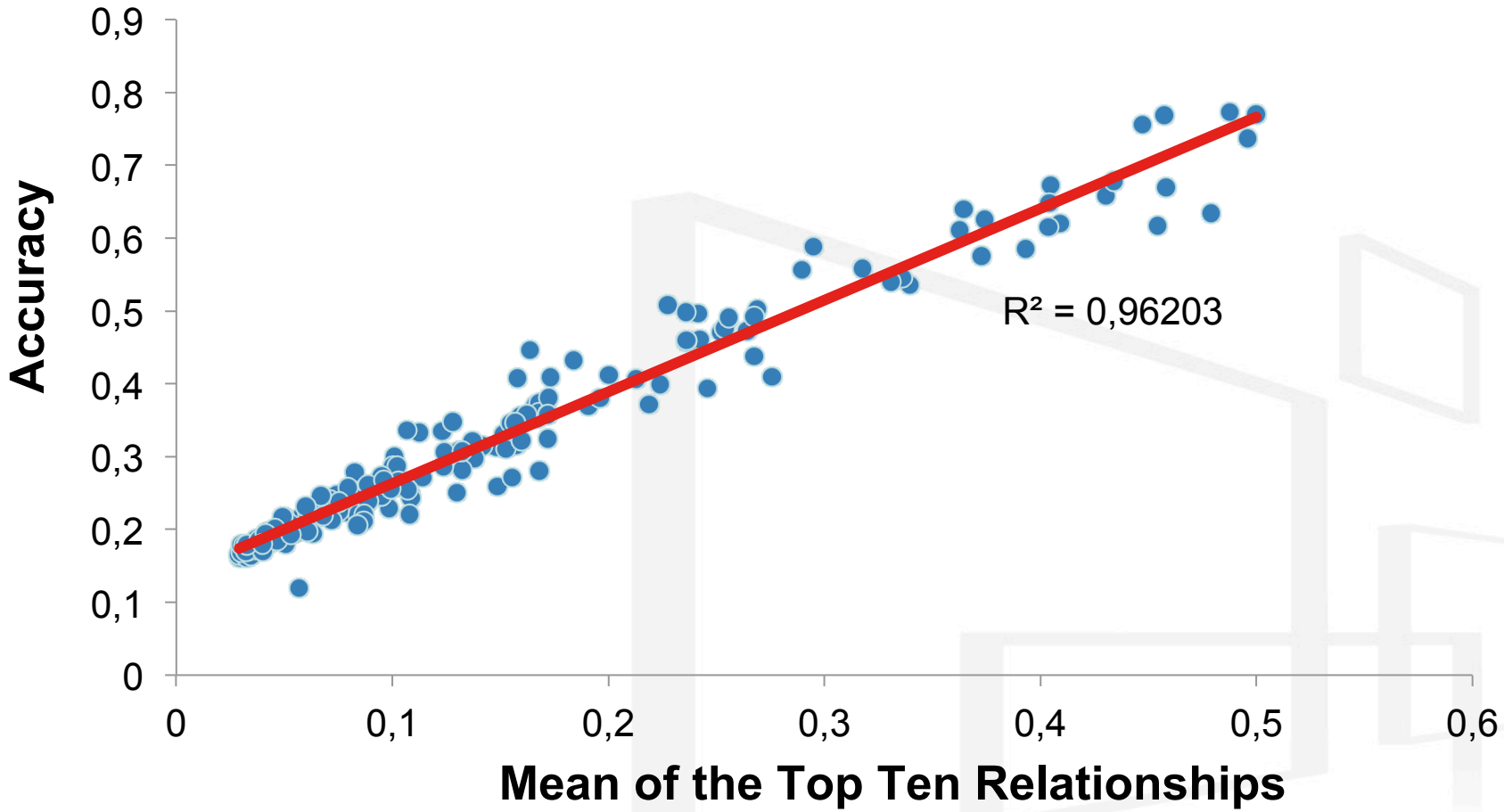
- GS has been implemented or is being implemented in all livestock species
  - » benefits over BLUP demonstrated
  - » benefits come from
    - » early prediction & reduced generation time
    - » better Mendelian sampling estimates, particularly for sex-limited traits
- Emerging evidence of benefits in salmon breeding
  - » exploits family structure in sib-testing
    - » c.f. livestock: more efficient training sets, lower SNP density needed
  - » GS models may benefit from separating within-family and between-family LD

4. Genomic Selection 2.0
   - Why isn't GS 1.0 Perfect?
   - Sequencing Populations
   - The Final Frontier?

# Why isn't GS 1.0 Perfect

- Markers effects are 'blind' predictions, not causal
  - » unable to predict unknown genetic correlations, e.g. GxE
  - » less than required for Precision Breeding
- Relies on linkage and LD, unlikely to capture full variance
  - » bovine 50k chip only captures 80% variance
- Can't predict well across breeds or lines
- Need large training sets for very good accuracies
  - » order 10,000s in livestock
  - » accuracy depends on how many close relatives
  - » but aquaculture populations often have many close relatives

# Accuracy and Relationships



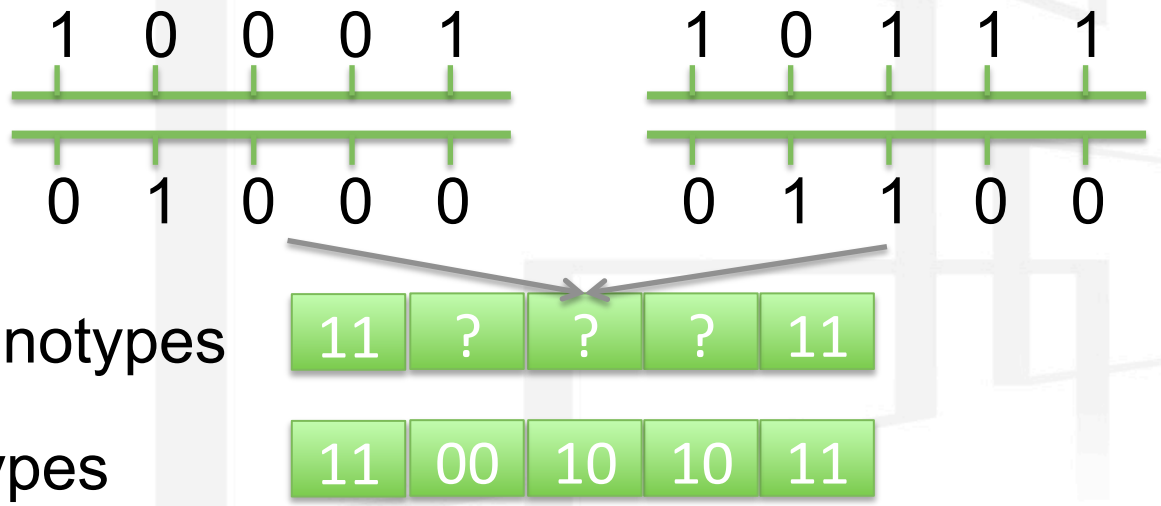

Clark et al. (2012)

# What can be done?

- Reduce cost of genotyping an individual
  - » obtain huge data

# Driving Down Unit Costs

- Imputing genotypes
  - » High density information for key individuals (e.g. sire)
  - » Low density information for candidates
  - » Exploit pedigree



# Power of Genotype Imputation

#Sires = 480
#Dams = 11,884
#Candidates = 100,000

60k chip = $120
6k chip = $48
3k chip = $35
384 chip = $20

| Scenario | MGD + PGD | Sire | Dam | Candidates | Individual cost | Accuracy of Imputation $R^2$ |
|---|---|---|---|---|---|---|
| SC1 | 0 | 60K | 0 | 384 | | 0.878 |
| SC2 | 384 | 60K | 384 | 384 | $20.58 | 0.929 |
| SC3 | 3K | 60K | 3K | 384 | $24.74 | 0.950 |
| … | | | | | | |
| SC16 | 60K | 60K | 60K | 60K | $120.00 | 1.000 |

Huang et al. (2012)

- AlphaImpute http://www.alphagenes.roslin.ed.ac.uk

# Driving Down Unit Costs

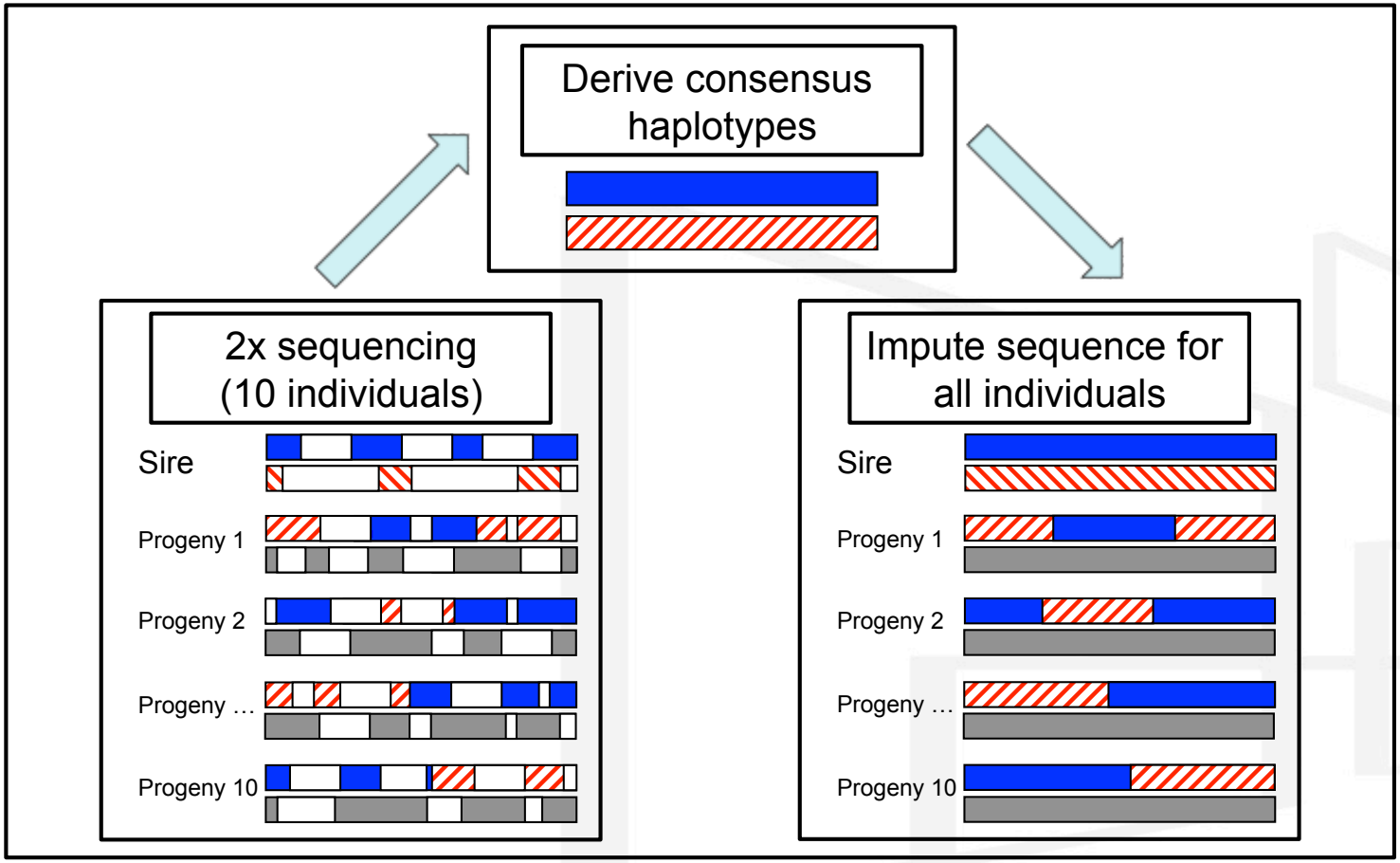
- Imputation
  - » High density information for key individuals (e.g. sire)
  - » Low density information for candidates
  - » Exploit pedigree

- Impute to what density?
  - » cost of sequencing data reducing!
  - » can we get it to $10 /individual with sequencing?

# Sequencing a Population

- ‘Traditional’ sequencing of individuals
  - » 10x or 30x coverage, still expensive
  - » therefore only sequence a few
- Sequencing a population
  - » how best to use total coverage available?
  - » genotyping by sequencing approach, 1x
  - » exploit pedigree and low Ne of livestock
    - » haplotype of key individuals are repeated many times in population
    - » sequence everyone at low coverage
    - » need new generation of imputation tools to cope with probabilistic nature of the sequencing at any locus

# Sequencing a Population



# Sequencing Populations

- Feasible to develop training sets in real populations of 300,000 animals all with sequence information
  - » before the next ISGA?

# The final frontier?

- Feasible to develop training sets in real populations of 300,000 animals all with sequence information
- Two orders of magnitude greater than '1000 Bulls' project in sequence data
- Order of magnitude greater than publically reported training sets for GS
- Will this be the final frontier?
  - » with our finite populations?

# What will final frontier look like?

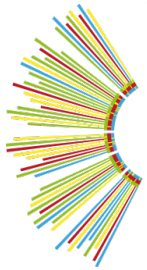

- Will accuracy be 1?
  - » it should be close
  - » if not, why not
    - » what is wrong with our models?
- How many QTN will be detected?
  - » 100's or 1000's?
  - » How much variation do detected QTN explain?


# Challenge of Precision Breeding

- Require to understand genetic correlation
  - » including GxE interactions
- Need to map QTN closely
  - » need sequence data, lots of it
  - » annotation of sequence
- To move from sequence to consequence need to understand expression and regulation
  - » to build the systems biology underling the consequences

# FAANG

- Functional Annotation of ANimal Genomes
  - » new international community consortium
- Target tissues
  - » musco-skeletal
  - » immune tissues
- Assays
  - » DNaseI, FAIREseq, ATAC-seq
  - » histone marks (promoters, enhancers)
  - » Methylation (BS-seq, RRBS)
  - » RNAseq (stranded), CAGE



FAANG
Functional Annotation of Animal Genomes


BBSRC
20 Years of Pioneering
Great British Bioscience

# Conclusions

- Major expansion of aquaculture genomes and tools emerging, while livestock are in a stage of refinement

# Conclusions

- Major expansion of aquaculture genomes and tools emerging, while livestock are in a stage of refinement
- Success of GS is determined by the quality and size of training set

# Conclusions

- Major expansion of aquaculture genomes and tools emerging, while livestock are in a stage of refinement
- Success of GS is determined by the quality and size of training set
- GS proven to be successful in key livestock species

# Conclusions

- Major expansion of aquaculture genomes and tools emerging, while livestock are in a stage of refinement
- Success of GS is determined by the quality and size of training set
- GS proven to be successful in key livestock species
- GS in aquaculture may be particularly beneficial due to sib testing

# Conclusions

- Major expansion of aquaculture genomes and tools emerging, while livestock are in a stage of refinement
- Success of GS is determined by the quality and size of training set
- GS proven to be successful in key livestock species
- GS in aquaculture may be particularly beneficial due to sib testing
- Genomic advances may allow us to boldly go to the final frontier

# Acknowledgements


- Sponsors of the presentation
  - » Red de Excelencia de Biotecnología en Acuicultura
  - » Aquagenomics-Net - Ministerio de Economía y Competitividad
- Sponsors of work presented
  - » EC FP7
  - » Biotechnology and Biological Sciences Research Council, UK
  - » Aviagen, Genus PIC

# Precision Breeding Goals

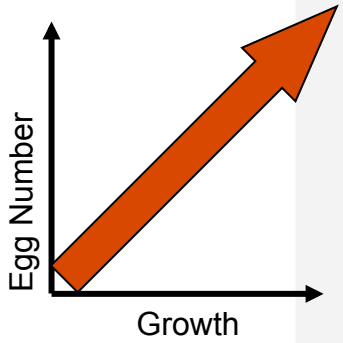
- To increase the scope and precision of predictions of the outcomes of breeding
- To avoid the introduction and advance of characteristics deleterious to animal well-being or, more generally, the well-being of the species
- To manage genetic resources and diversity between and within populations in accordance with the principles set out in the Convention on Biological Diversity

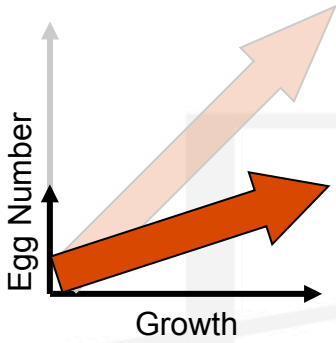
Flint & Woolliams, 2008, Proc.Roy.Soc.B

# Precise Breeding

- Differences in accuracy among traits arise from differences in amount of genetic information
  - » one trait sex-limited, one not
  - » different heritabilities
- Such differences create divergence from gains achieved using economic indices and desired gains

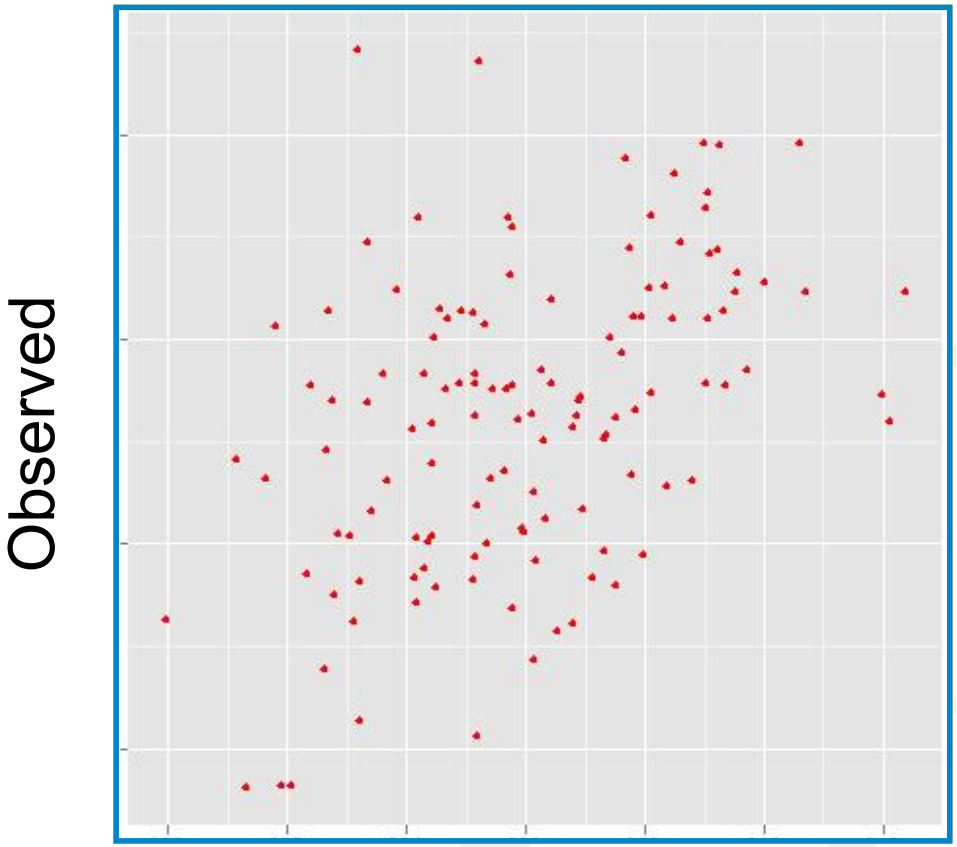
**Desired Gain**




**Achieved Gain**

# Cross Validation: Height




Set 1
r = 0.455

Set 2
r = 0.426

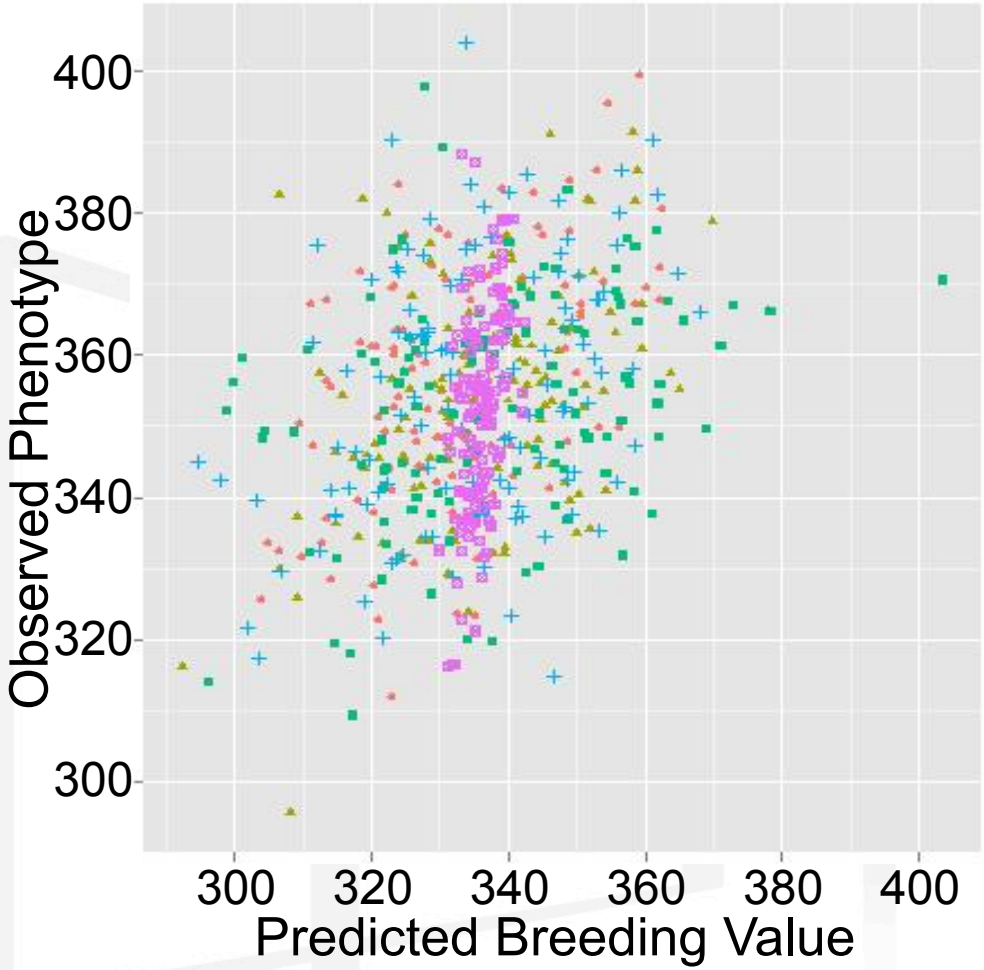
Set 3
r = 0.317

Set 4
r = 0.335

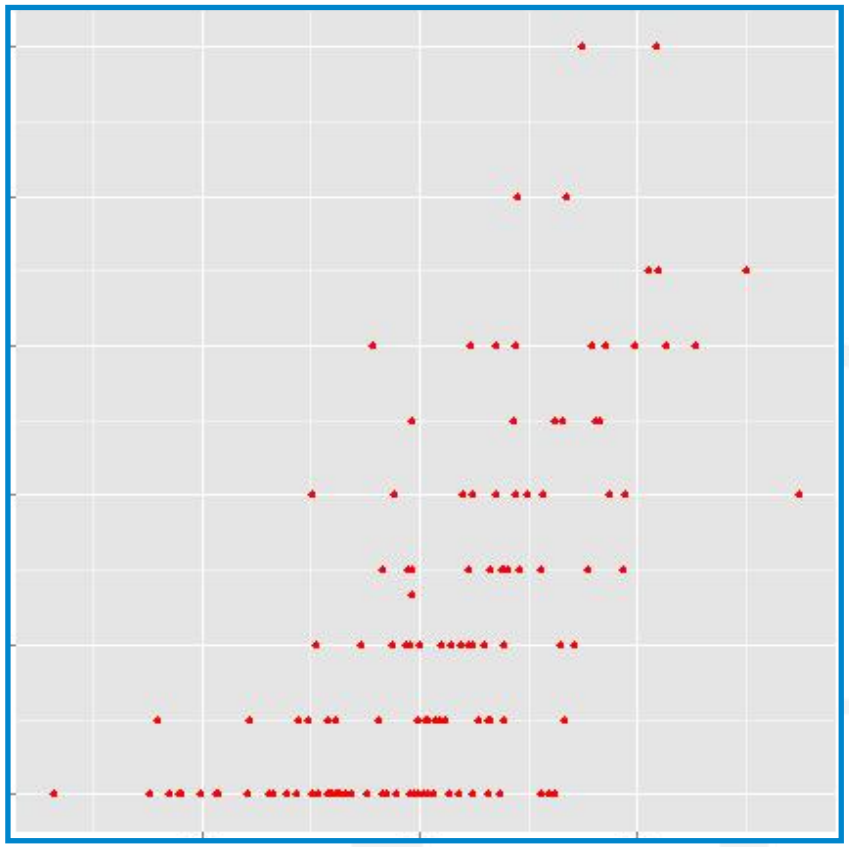
Set 5
r = 0.447

# Prediction of Height

- Accuracy = 0.400
- Max accuracy for predicting phenotype from BV is 0.548
  - » within families
  - » √0.3
- Corrected accuracy = 0.73
  - » from 497 phenotypes

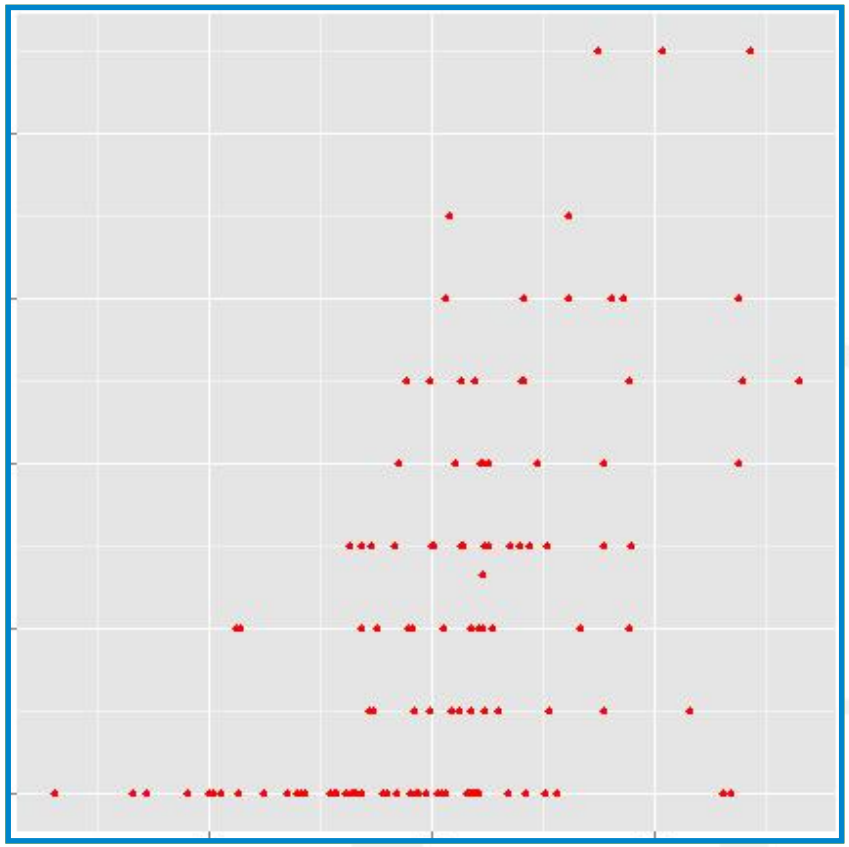


# Cross Validation: Bud Burst




Set 1
r = 0.646

# Cross Validation: Bud Burst




Set 1
r = 0.646

Set 2
r = 0.615

Set 3
r = 0.618

Set 4
r = 0.502

Set 5
r = 0.526

# Prediction of Bud Burst

- Accuracy = 0.581
- Max accuracy for predicting phenotype from BV is 0.632
  - within families
  - √0.4
- Corrected accuracy = 0.92
  - from 497 phenotypes