

DNA – A competitor or a Partner?

John McEwan

World Hereford Conference July 2008

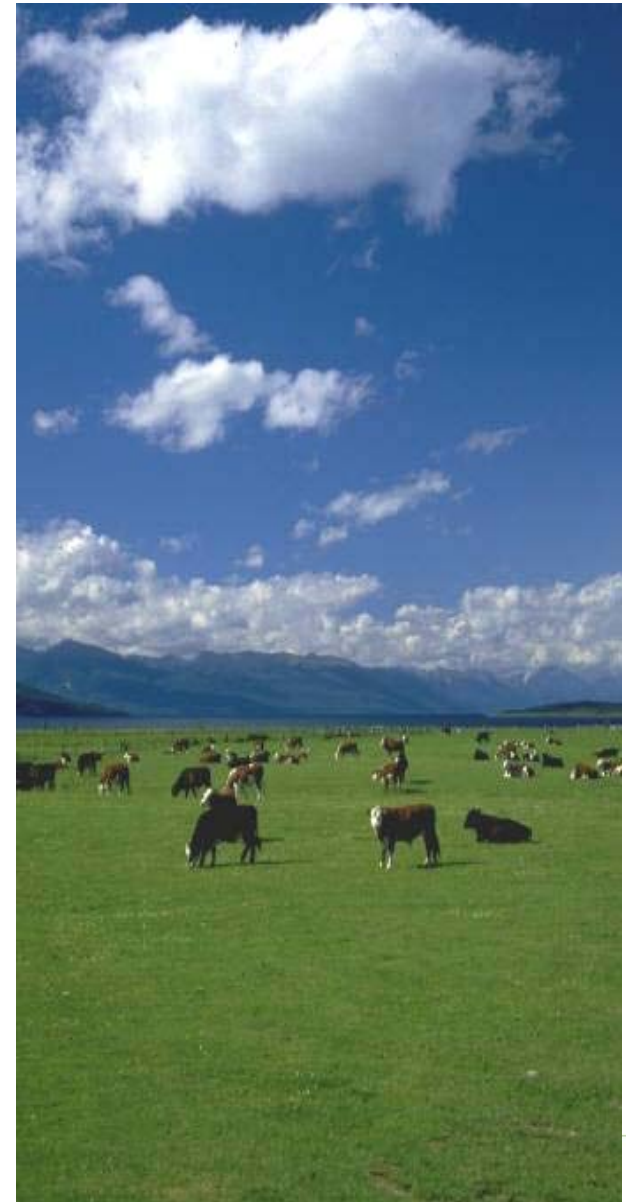


Farming, Food and Health. **First**

Te Ahuwhenua, Te Kai me te Whai Ora. Tuatahi

Overview

- **Genome structure**
- **What can DNA information provide?**
 - *Identity*
 - *Parentage (breed)*
 - *Genetic worth*
 - *Production worth*
- **New technology**
 - **SNP chips**
 - **Whole Genome Selection**
 - **Cost changes**
- **Creating the resources?**
- **Putting it all together**



Genome Structure

- Genetic selection retains favourable DNA variants
- We now have the ability to identify and select the variants
- The current opportunity is to select on variants using **LD**

Many types of variants

- Single Nucleotide Polymorphisms (**SNPs**)

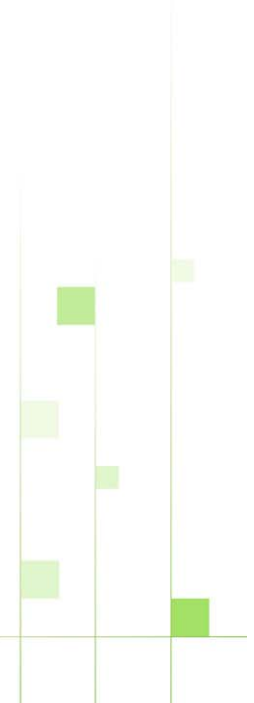
AGCTACATC**C****GCTAGACGT** fathers chromosome
TCGATGTAG**G****CGATCTGCA**

AGCTACATC**T****GCTAGACGT** mothers chromosome
TCGATGTAG**A****CGATCTGCA**

- Microsatellites

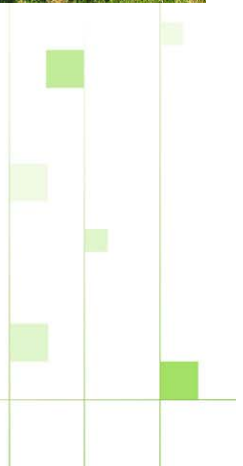
AGCTACATC**CGCGCGCGCGCGCG**.....**GCTAGACGT**

- Insertion/deletion or indels
- Copy number variants or CNVs



SNPs key facts

- In cattle 1 SNP/500-1000 bp
- 3 Billion bp genome = ~3-6 Million differences!
- About 20-50 new ones created each generation
 - More created males perhaps 80%
 - Plenty lost each generation
 - Most? ~94% have no effect
 - 1% larger effects
 - 50-100K plus 1 extra /generation
 - 5-100 variants explain 50% of a traits genetic variation
- Freq of SNPs & distribution in genome species
 - Population history
 - Size
 - Selection



What is a SNP chip?

SNP = single nucleotide polymorphism

SNP chip = test 60,000 to 1,000,000 SNPs

WGS works by being able to:

- predict status of other SNP variants nearby
 - Uses a concept called linkage disequilibrium
- includes variants that affect production traits

MELD	atcgcgtgtagctagtgctagctgctagctagctgatgca
ROM1_read12667t.....
AWA1_read00345
SBF1_read06734
TEX1_read00234t.....
ROM1_read10385t.....
TEX1_read39890



Linkage disequilibrium

Track DNA segment

- Recombination scrambles
- Markers need to be close

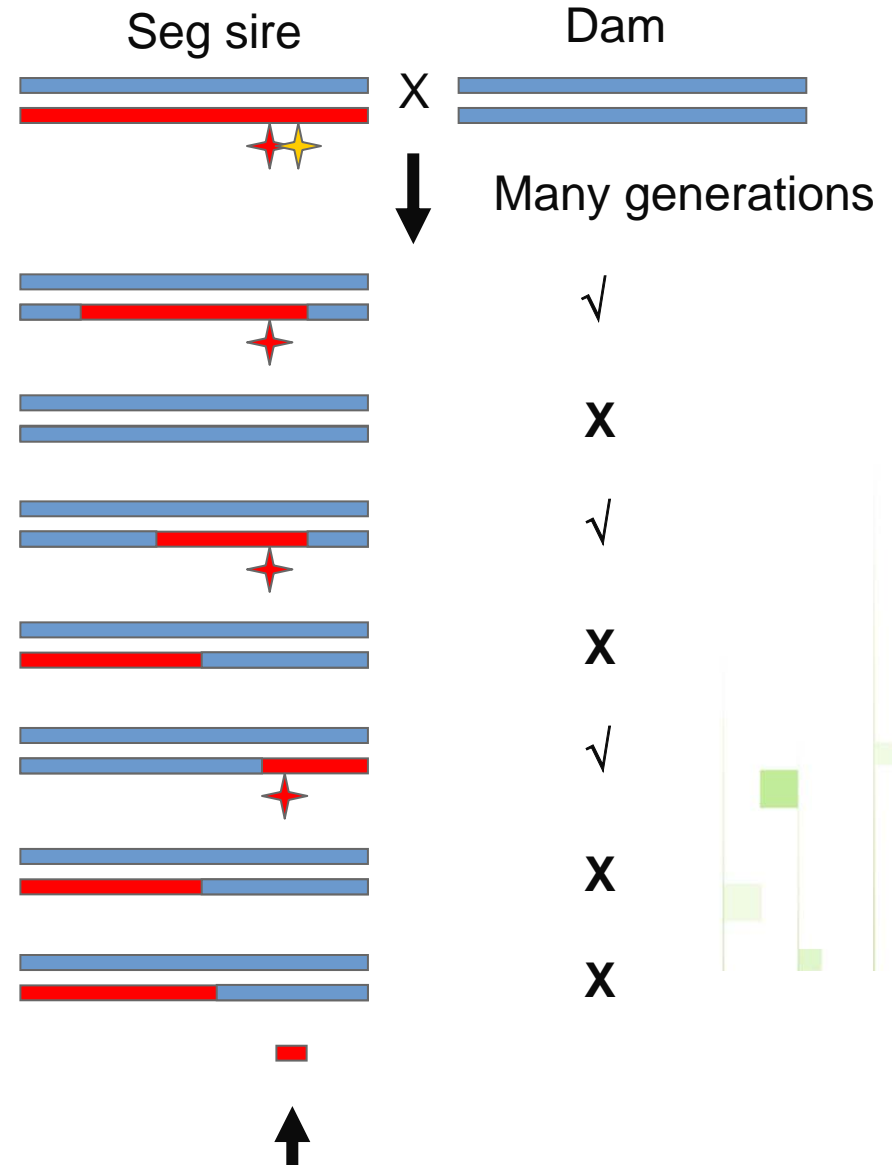
No known pedigree

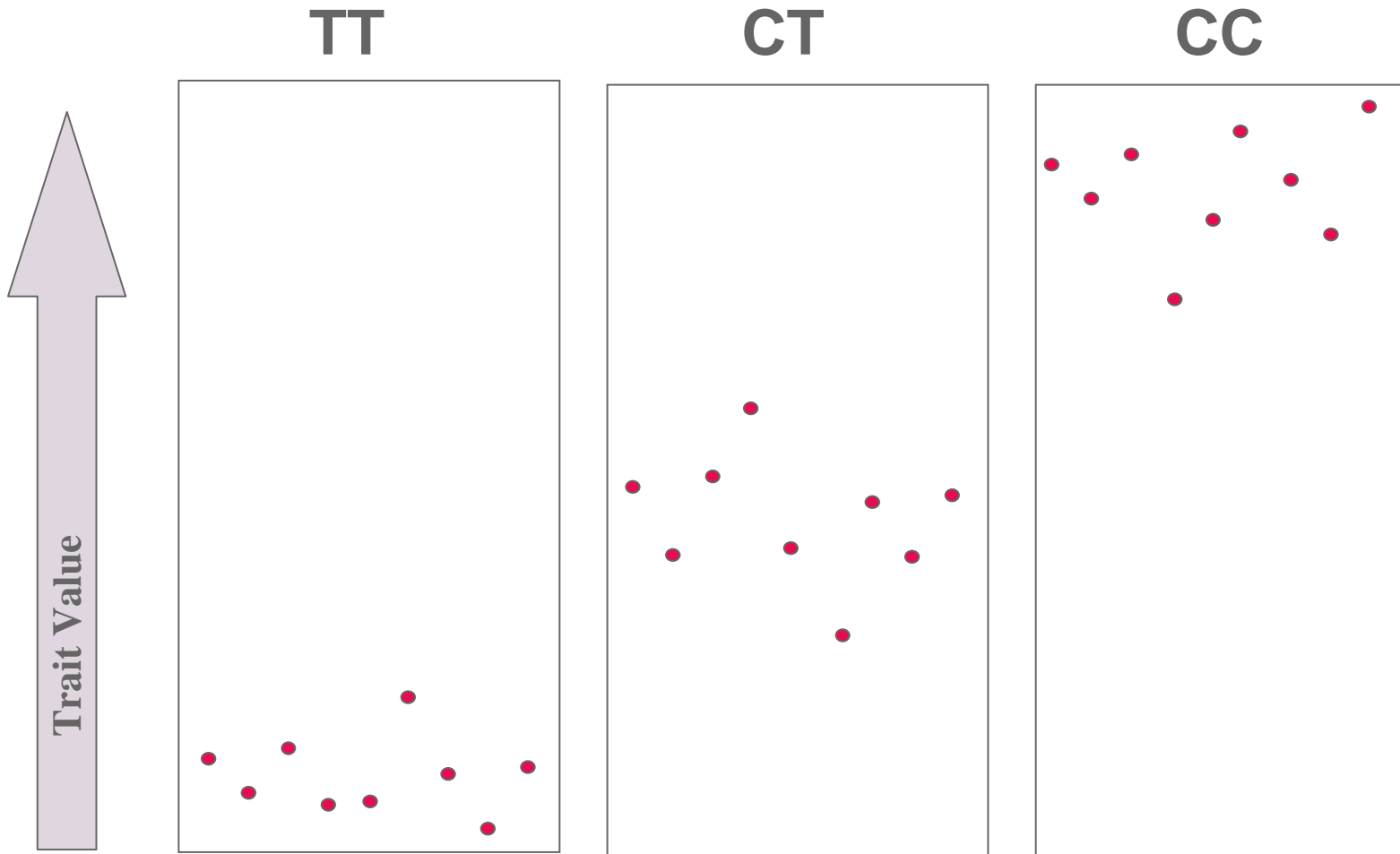
Find associated SNP

Association effect depends

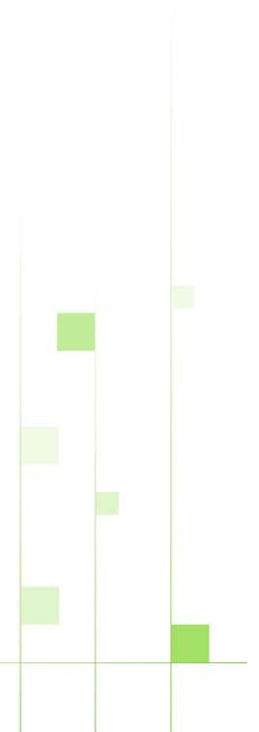
- Size of effect
- How close
- How well linked

		C	G
A		1000	40
T		55	1200

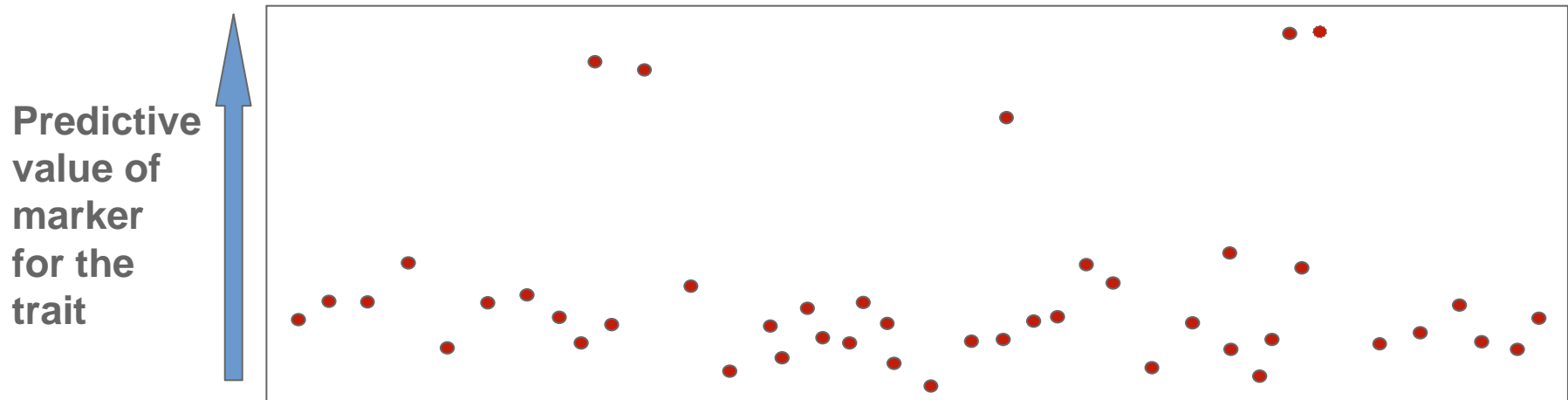




Each Marker is given a predictive value for the trait
(In this case a strong association)



Then genotype each of the animals using the SNP chip



Each of 60,000 SNP markers



What can DNA information provide?

Identity

- Within small flocks need ~10-20 SNPs
- Across industry require ~ 50 SNPs

Parentage

- Within small flocks need ~50 SNPs
- Across industry need ~100-400 SNPs
- Breed proportions ~50-100 SNPs
- Reconstructed pedigrees 1000-2000 SNPs



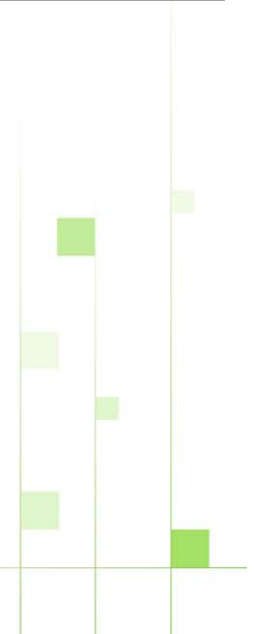
Breeding value

- ~300SNPs/trait
- Perhaps 1000-5000 selected SNPs explain most variation (WGS)

Production value

- Additive plus combination of alleles
- Same numbers

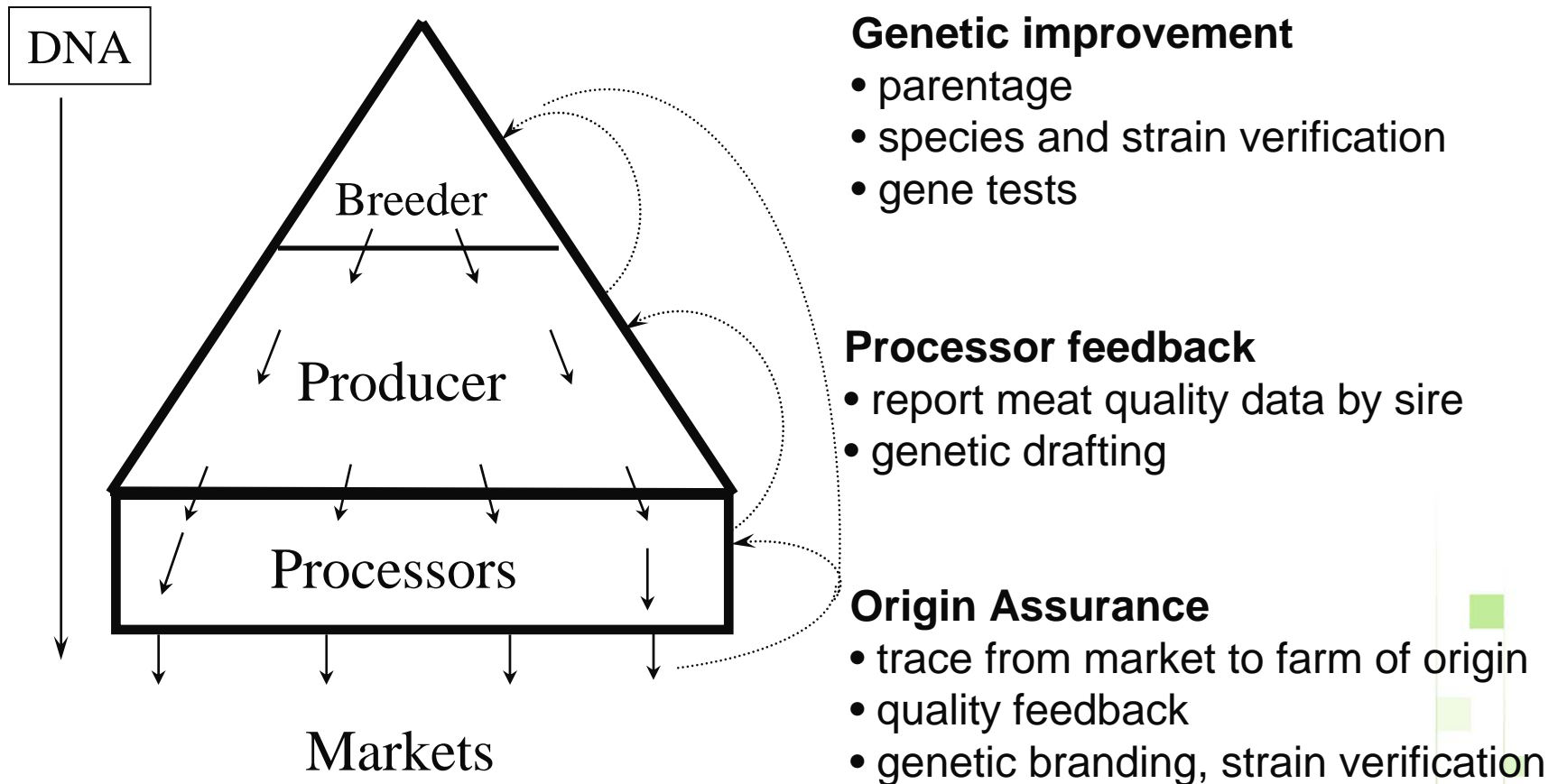
Future is 1 sample and test for all four



DNA information systems



“Closing the loop”



Putting it together (DNA is a partner)

- Not only cheaper & faster: need better tests
- Need a system
 - Cheap robust sampling and storage
 - Industry genetic database
 - Automation of as many steps as possible
 - Internet reporting
- NZ have internet systems and databases
 - Need national IDs

$$\mathbf{gEBV = a*dnaEBV + b*EBV}$$

- “Blending” most likely in interim
- Produces a result identical to existing system
- Easy to understand
- Benefits: sex limited, difficult to measure traits



Summary

- DNA variants: identity, parentage, breeding & production
- Offer potential to accelerate genetic gain (for difficult traits)
- Costs declining rapidly
- Need to develop good systems to use it well
 - Industry phenotypes
 - Benefits in extensive industries
- Dairy Cattle NZ WGS implemented
- Beef cattle
 - Identity now
 - Parentage now
 - Individual marker tests now
 - Whole genome selection?
 - Genetic drafting?





New technology

Genotyping and sequencing costs declining
10 fold every 7 years

- 60K SNPs \$250
- 1.5K SNPs \$30-50
- 100 SNPs \$10

Plan for the future: not the present or past

- 1 year can do 600 SNPs in all stud tier
- 7 years can do ~6000 SNPs in all stud tier
- 14 years can do 60K SNPs in all stud tier
- 20 years all key stud sires are sequenced?!



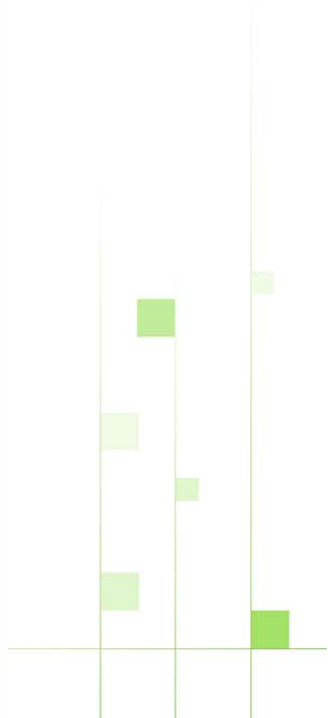
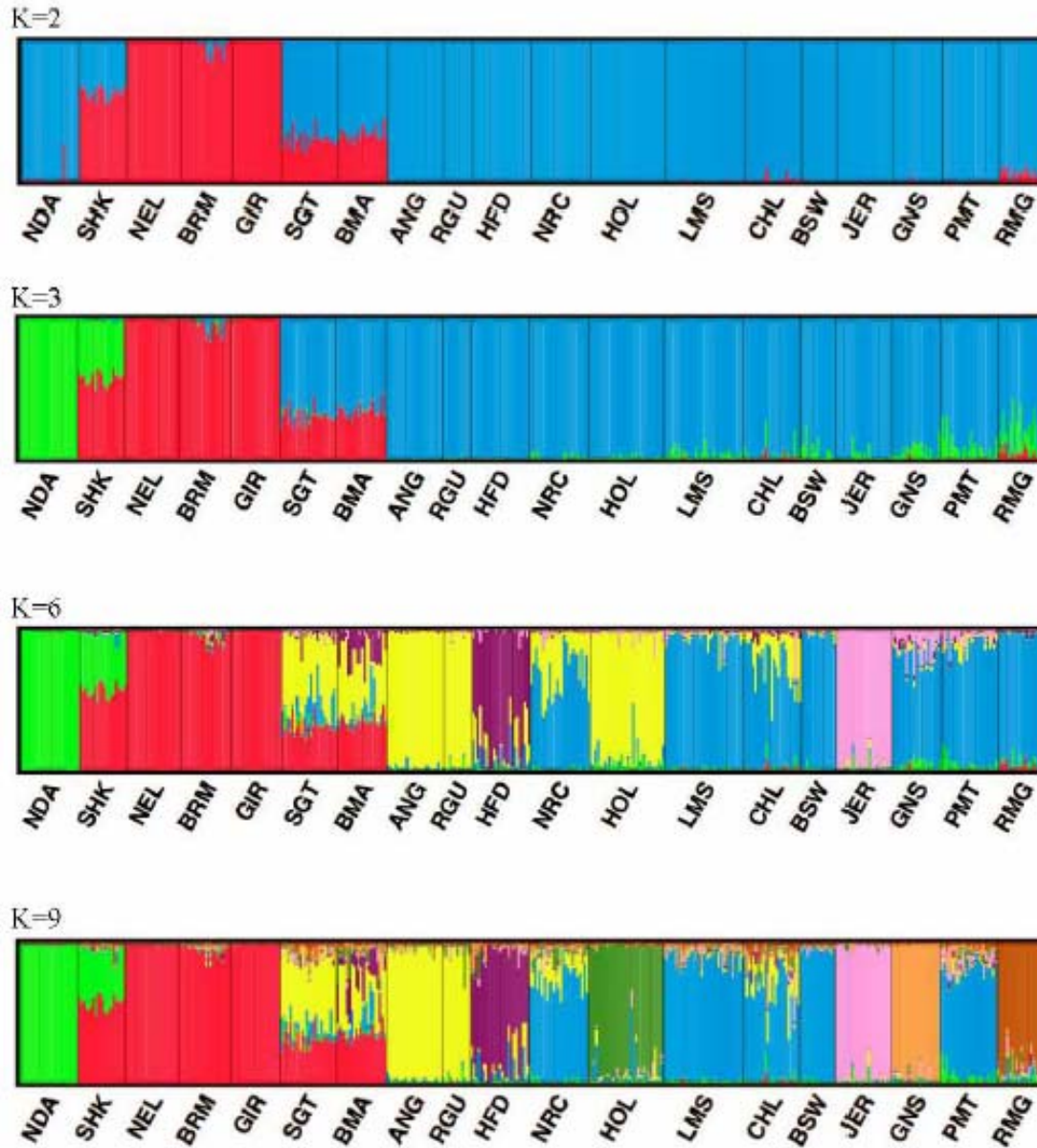
Genome Sequencing why and how?



- We need the genome sequence for many reasons:
 - Current context is to **find** SNPs for whole genome selection
 - Have sufficient sequence around SNPs to **detect** them
 - **order and spacing**: SNP selection and analysis of results
- **Human 2000** cost ~**\$1000M**
 - 300K-1M SNP chips
- **Cattle 2002-2006**
 - **\$10M** physical map
 - **\$52M** sequence (7X coverage ~1M SNPs)
 - 10-30-60K SNP chips
 - WGS implemented 2008
- **Sheep 2007**
 - **\$2.5M** (3X coverage ~200K SNPs)
 - 60K SNP chip
 - 1 year delay cost sheep industry \$10M!
- **Deer 2009?**
 - <\$1M?



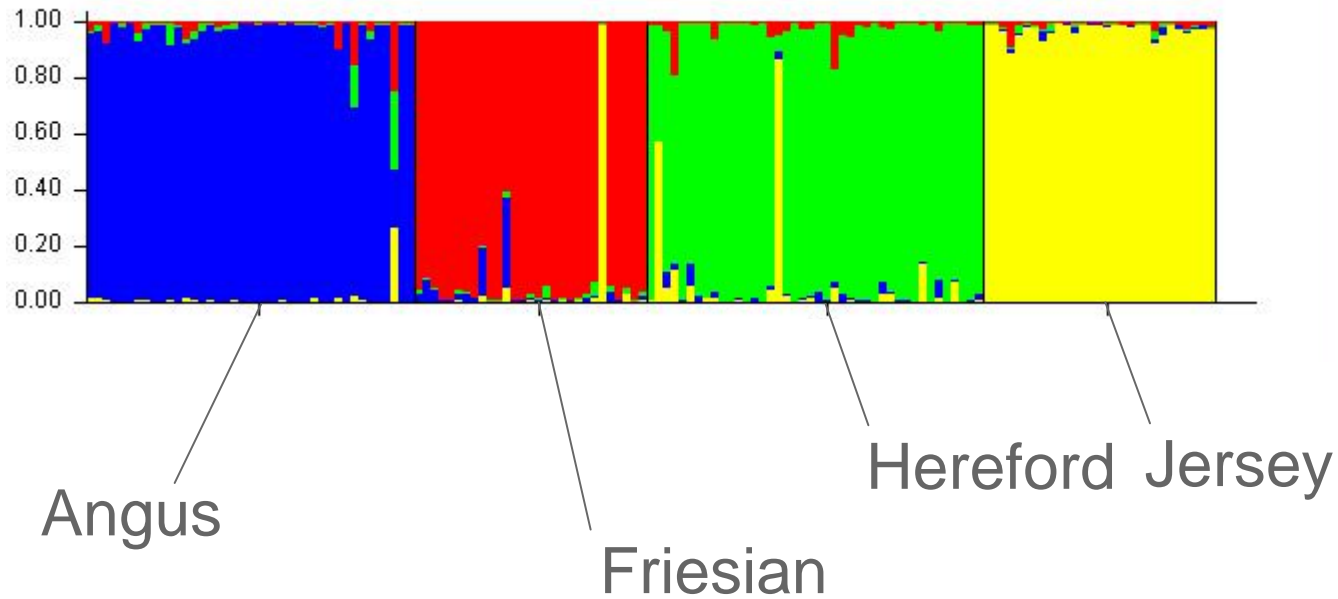
Breed differentiation



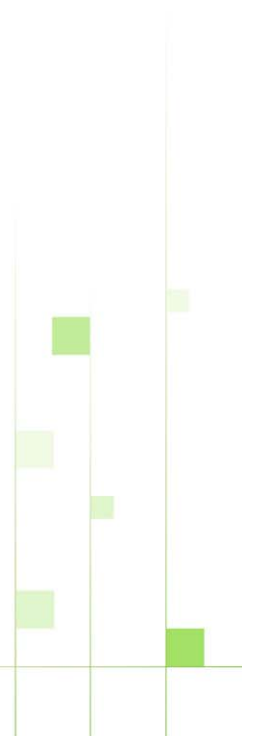
Homogeneity in New Zealand cattle breeds



K = 4



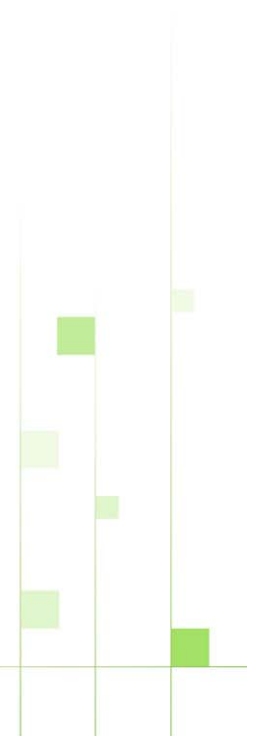
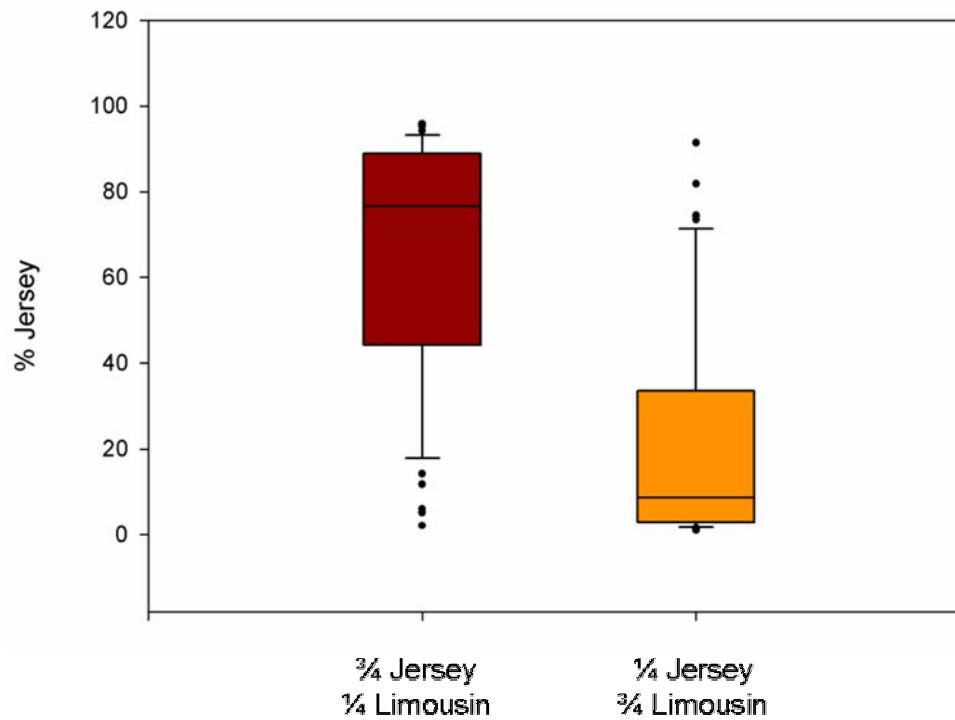
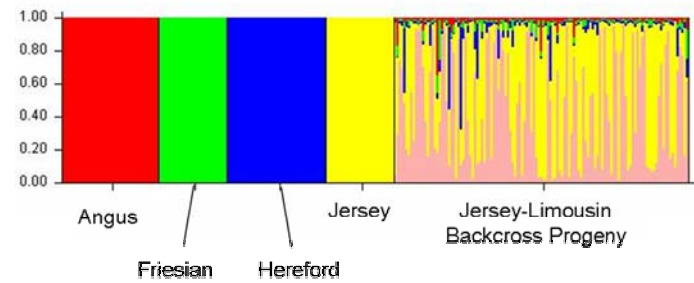
Uses DNA parentage markers!



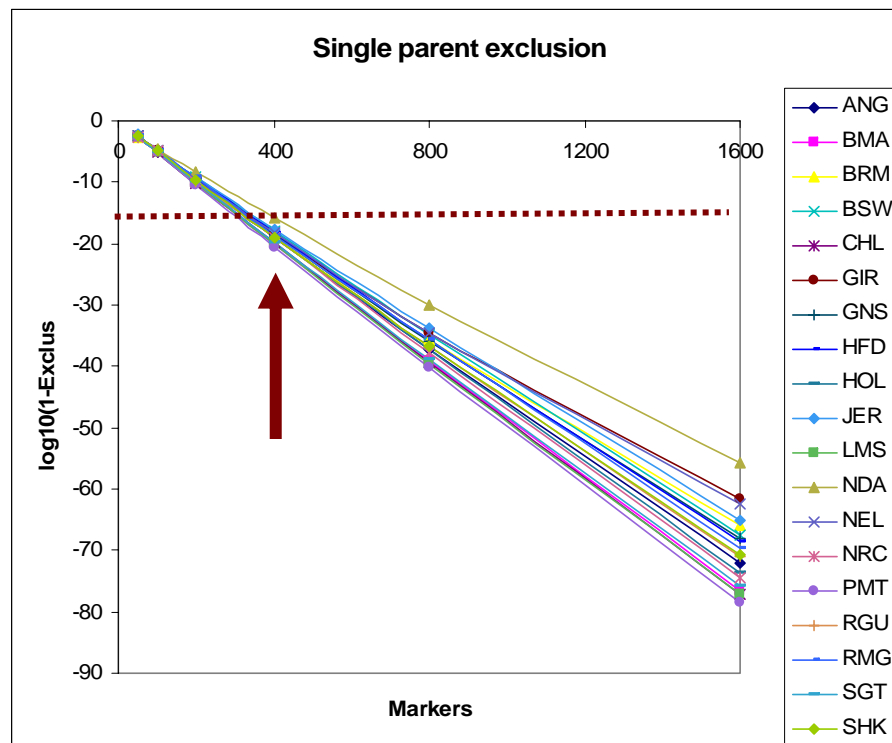
Assignment of breed composition in hybrid animals



- Known hybrid ancestry
- ◆ Jersey x Limousin
 - ◆ Reciprocal Backcross

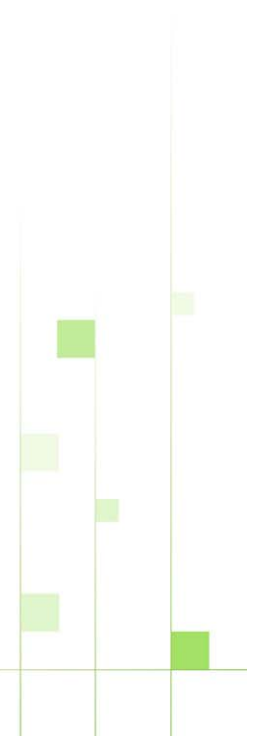


Traceability via paternity



Plot of $\log_{10}(1-\text{exclusion probability})$ versus SNP marker number by breed for paternity dam unknown

- Based on cattle HapMap data
- ~1 billion cattle
- “world” threshold $\sim 1\text{E-}15$
- require ~400 “good” SNPs



Linkage disequilibrium

Track DNA segment

- Recombination scrambles
- Markers need to be close

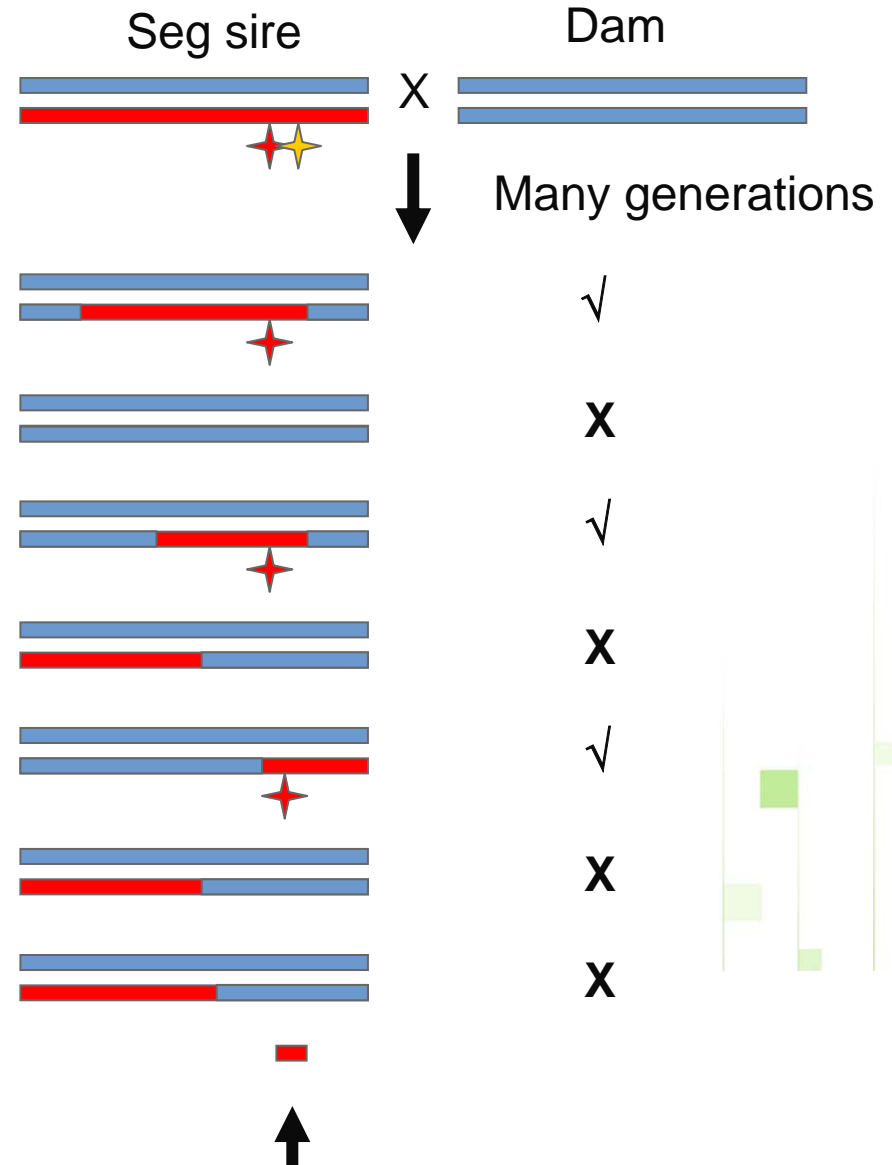
No known pedigree

Find associated SNP

Association effect depends

- Size of effect
- How close
- How well linked

		C	G
A	1000	40	
T	55	1200	



Possible beef cattle system



	Stud sires	Stud bulls	Commercial cows
	2500	250000	20,000,000
	1000	100,000	?
SNPs	60K	1600	50
Cost	\$500	\$30	\$5
\$/cc	\$0.03	\$0.15	?

Diagram description: A table comparing three stages of a beef cattle system. The first stage (Stud sires) has 2500 animals, 1000 SNPs, a cost of \$500, and \$0.03/cc. The second stage (Stud bulls) has 250,000 animals, 100,000 SNPs, a cost of \$30, and \$0.15/cc. The third stage (Commercial cows) has 20,000,000 animals, 50 SNPs, a cost of \$5, and an unknown value. A red arrow labeled 'SNP Key' points from the Stud sires stage to the Stud bulls stage. A black arrow points from the Stud sires stage to the Commercial cows stage.

The first 2 steps process can potentially double the genetic gain