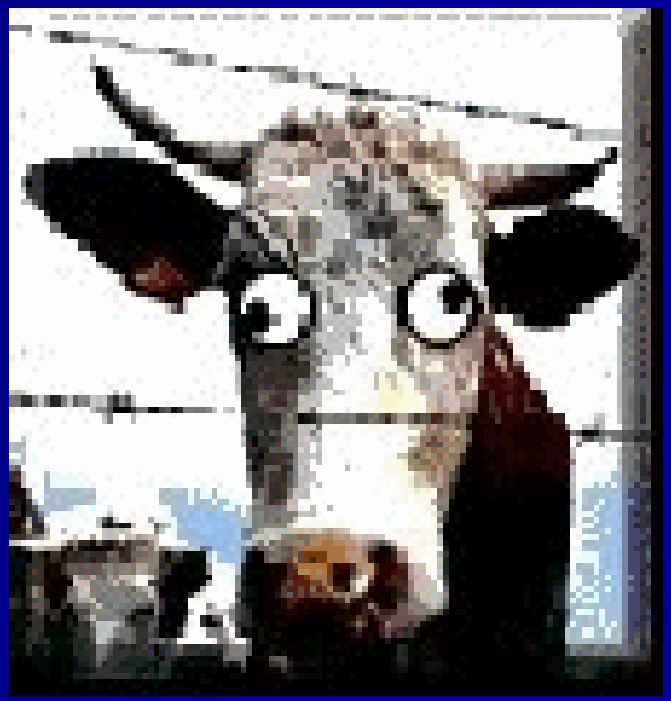
SNPs, CHIPs and WGS – Making Sense of Biotech Babble

Alison Van Eenennaam, Ph.D.

Cooperative Extension Specialist
Animal Biotechnology and Genomics
University of California, Davis

alvaneenennaam@ucdavis.edu



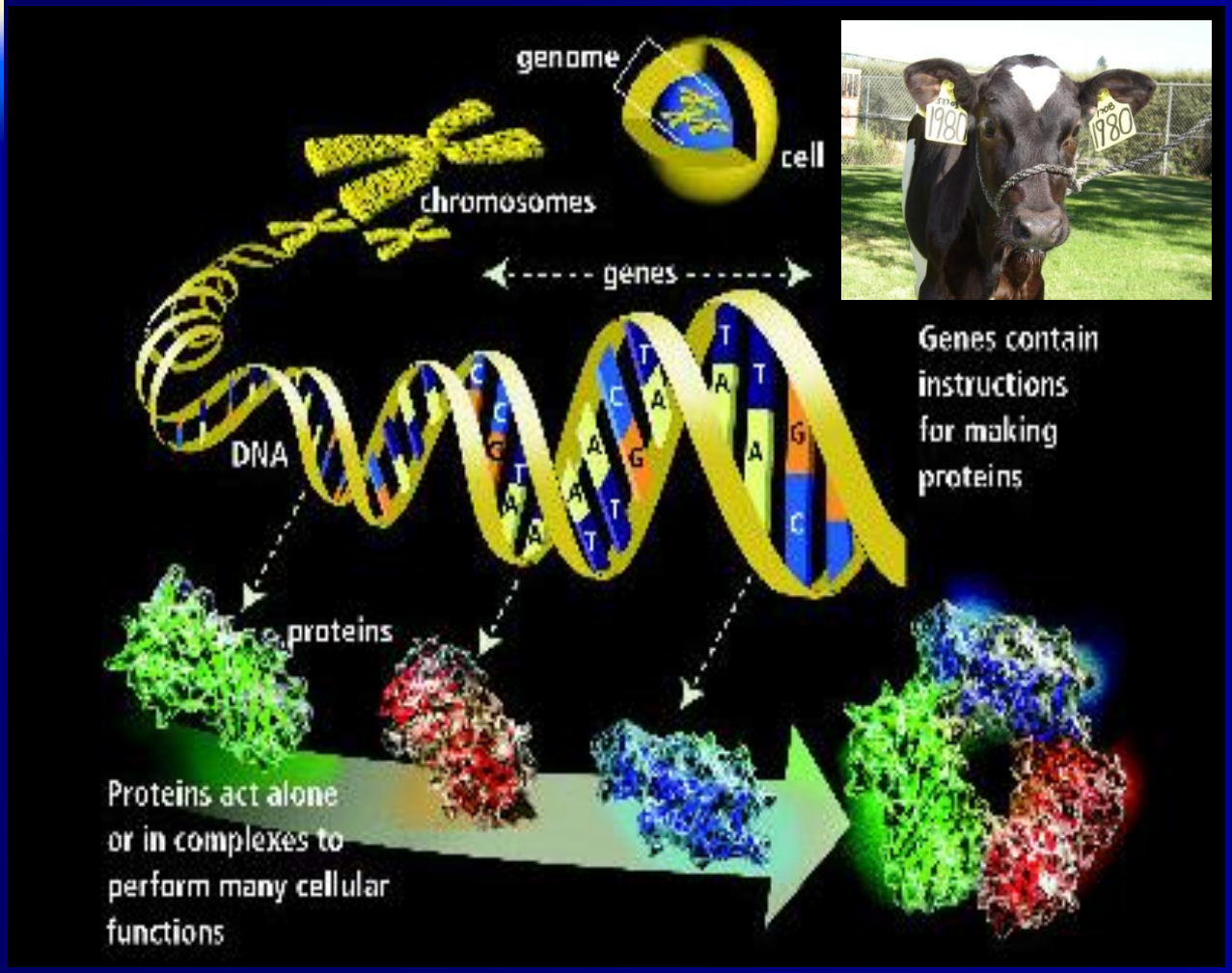
http://animalscience.ucdavis.edu/animalbiotech/

NBCEC Brownbagger 10/8/08

Overview

- Background
- STRs
- SNPs
- CHIPs
- MAS
- WGS
- Implications

The bovine genome is similar in size to the genomes of humans, with an estimated size of 3 billion base pairs.



Human & cattle genomes are 83% identical

Why is DNA sequence important to the cattle industry ?

- **Parentage**
- **DNA-Assisted Selection** genetically identification of superior animals through DNA genotyping
- **Traceability** – only DNA can link backwards and forwards through the production chain

Why is parentage important ?

- Identify bulls producing problem calves
- Identify extremes in phenotypes
- ID of cleanup bulls after AI
- Determine bull dominance – 50% of the bulls sire 80% of the calves
- Enable EPD calculations for commercial sires in a herd
- Genetic product/process validation

STRs and SNPs

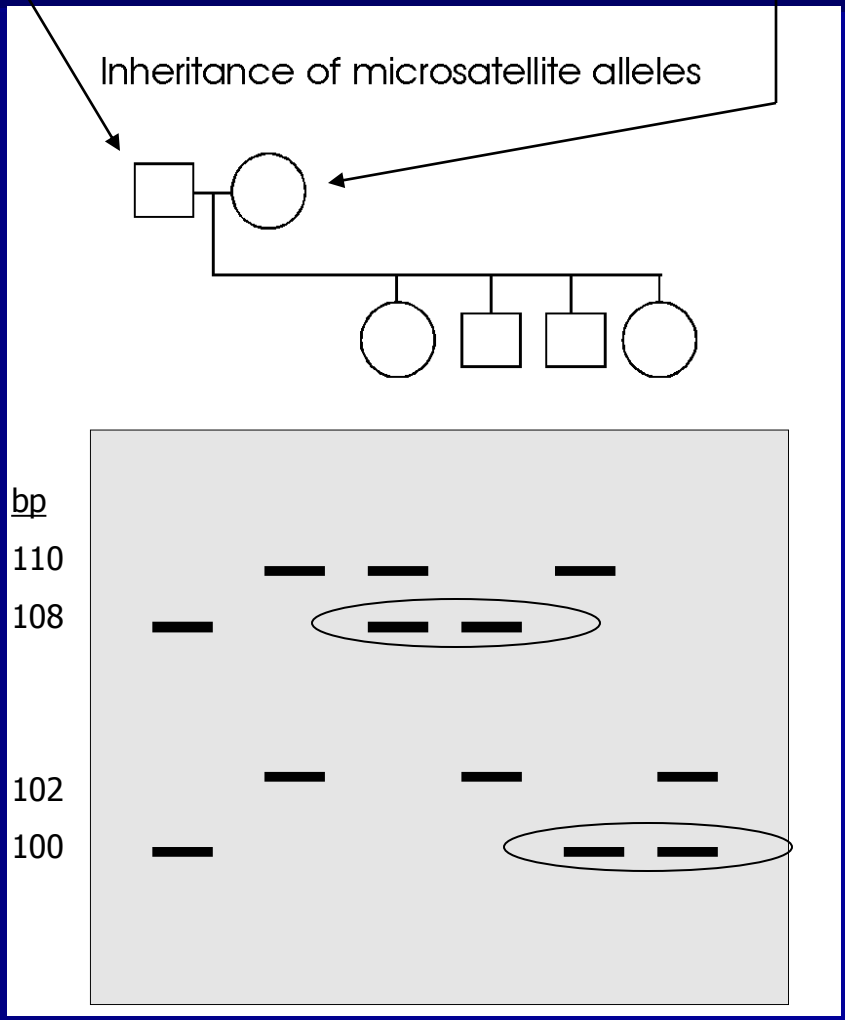
- There are two basic methods being used to determine the genetic identity and kinship (paternity) of an animal
- **Microsatellites** or short tandem repeat markers (**STRs**)
- **SNPs** = single nucleotide polymorphisms

How do microsatellites work ?

ctgaatatatatgcta
ctgagcta

ctgaatatatatatgcta
ctgaatgcta

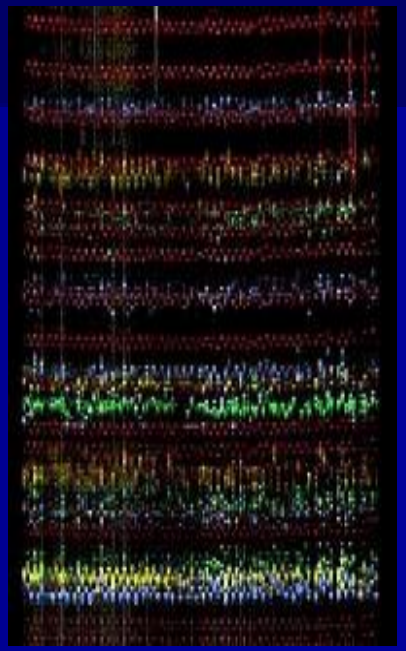
- Microsatellites or STRs are small tandem repeats (2,3 or 4 bp !!) that vary in number and size between individuals
- Inherit a copy from the dam and a copy from the sire
- Used for exclusion of parentage



Probability of exclusion (P_E)

- P_E = the probability that a random individual other than a true parent from a population in Hardy-Weinberg equilibrium is excluded as the parent of another randomly chosen individual.
- For unrelated sires, the probability of unambiguous parentage assignment is equal to P_E raised to the power of the number of non-parent candidate bulls

Microsatellites Pros and Cons



PROS

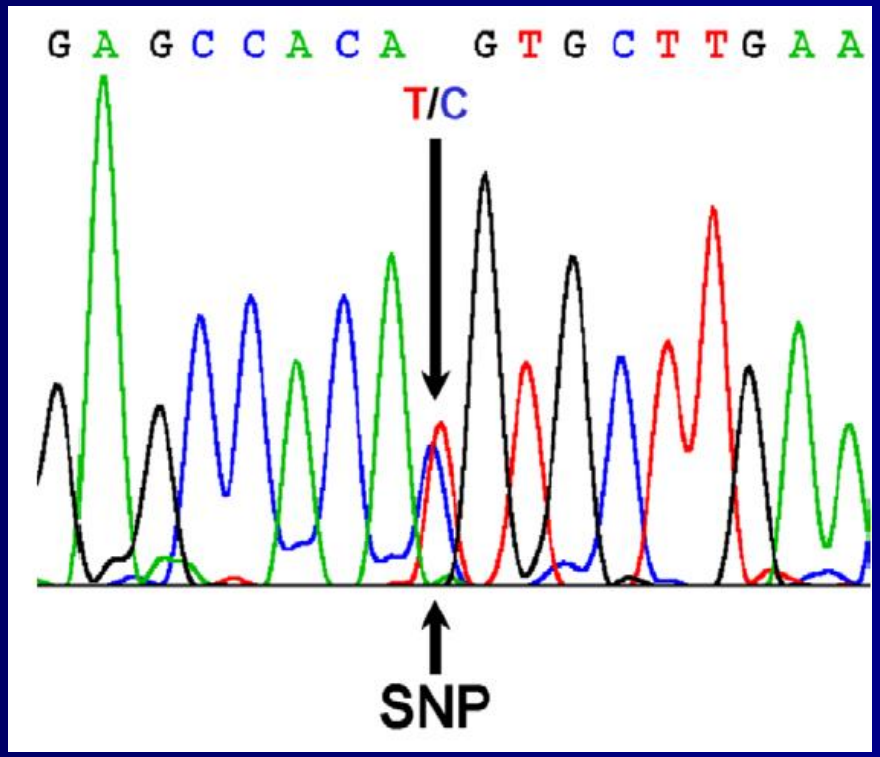
- Highly informative markers – many alleles
- Have been used by breed associations for years so historical database exists
- ISAG has a standardized marker set

CONS

- Hard to get consistent results across labs
- Not all microsatellites are equally informative across all breeds of cattle
- Can not be made to get much cheaper – currently running > $20/test
- Not much more research being done on finding new microsatellites

SNPs = Single nucleotide polymorphisms

SNPs are the most common and stable type of DNA marker in cattle and are ideally suited for automated, economical genetic testing



Ideal SNP for parentage

- **Allele in equal proportions (p = 0.5, q = 0.5)**
- **Evenly spaced throughout the genome**
- **Can be accurately scored**
- **Are commonly used across all labs**

SNPs Pros and Cons

PROS

- Abundance – 30 million in cattle!
- Potential for automation
- Low genotyping error rates
- Ease of standardization between labs
- Low mutation rates

CONS

-
-

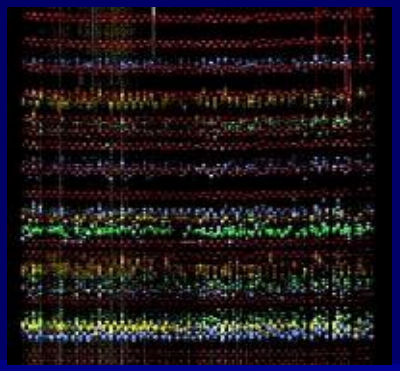
Calf	AA	AA	AA	AA	
					Probability
Bull 1	AA	AA	TT	AA	0
Bull 2	TA	TA	TA	TA	6
Bull 3	AA	AA	AA	AA	94

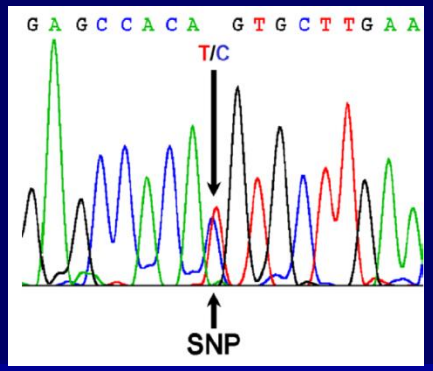
PATERNITY ANALYSIS IN LARGE COMMERCIAL CATTLE RANCH SETTING USING SNPs - UC DAVIS EXPERIENCE

- Blood collected on FTA cards from **27** herd sires and **624** calves derived from a multiple-sire pasture

Daniel J. Drake
M. Cecilia T. Penedo
University of California, Davis

Genotyping





- **Genotyping and paternity assignments based on microsatellites (STRs) were done by the UC Davis Veterinary Genetics Laboratory using a panel of 23 cattle markers (P_E=99.9%)**
- **Genotyping based on SNPs were done by a commercial genotyping company using a panel of 28 loci (P_E=95.5%)**

A. L. Van Eenennaam, R. L. Weaber, D. J. Drake, M. C. T. Penedo, R. L. Quaas , D. J. Garrick, E. J. Pollak. 2007. DNA-based paternity analysis and genetic evaluation in a large commercial cattle ranch setting. Journal of Animal Science. 85:3159–3169

Results of the paternity analysis

(PE=99.9%)

	23 Microsatellite (STR) panel	
One possible sire	533*	85.4%
More than one sire	4	0.6%
All excluded	76	12.2%
Resubmits	11	1.8%
TOTAL	624	

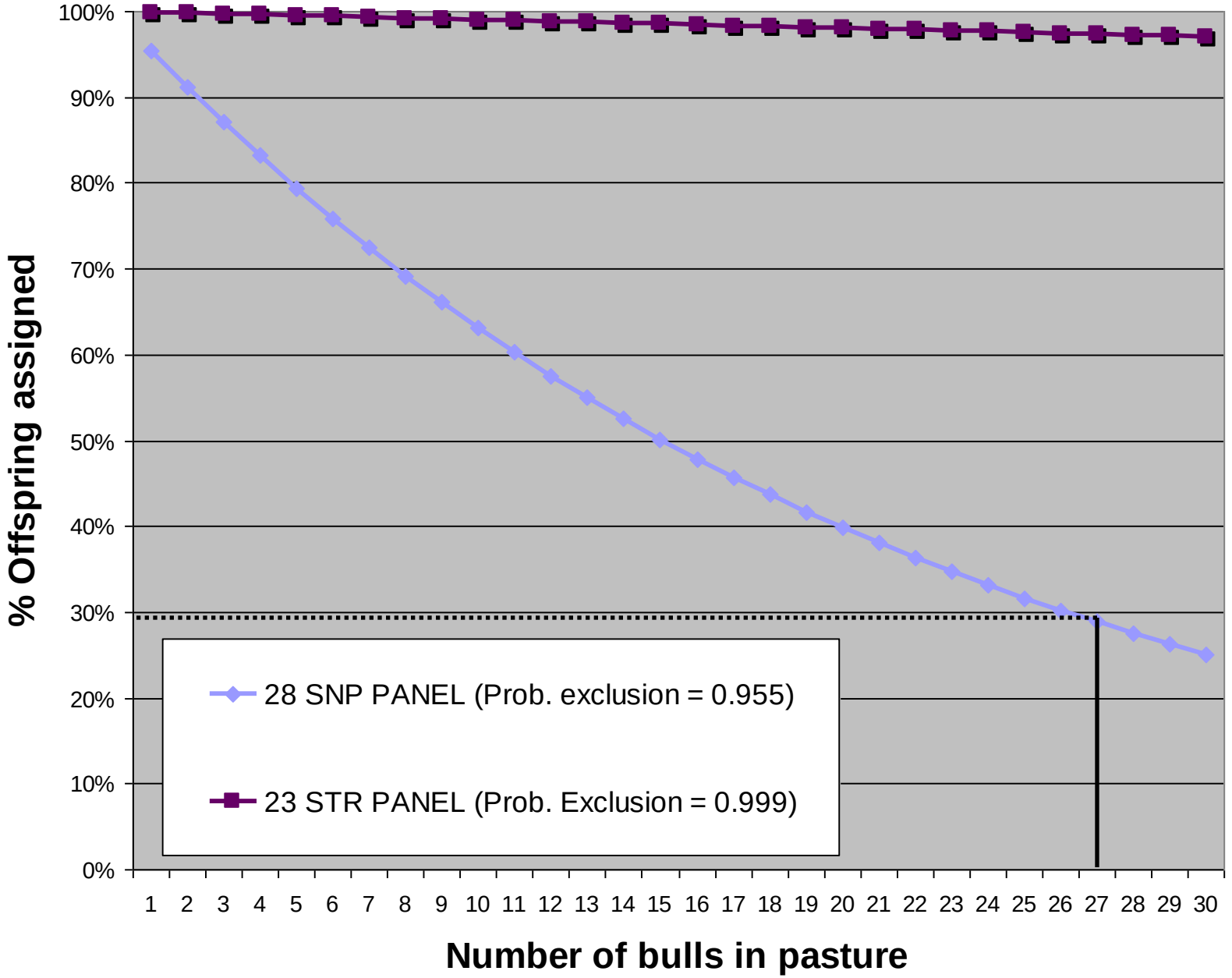
DNA from more than one animal

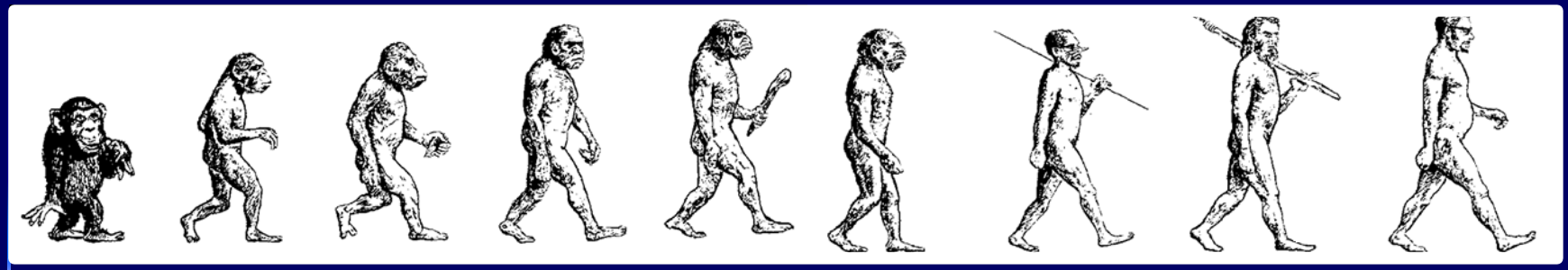
* 10 assignments allowed a one locus mismatch

	23 Microsatellite (STR) panel (PE=99.9%)		28 SNP panel (PE=95.5%)	
One possible sire	533*	85.4%	**175**	**23.3%**
More than one sire	4	0.6%	**420**	**67.3%**
All excluded	76	12.2%	**29**	**4.6%**
Resubmits	11	1.8%	**0**	**0%**
TOTAL	624		**624**	

* 10 assignments allowed a one locus mismatch

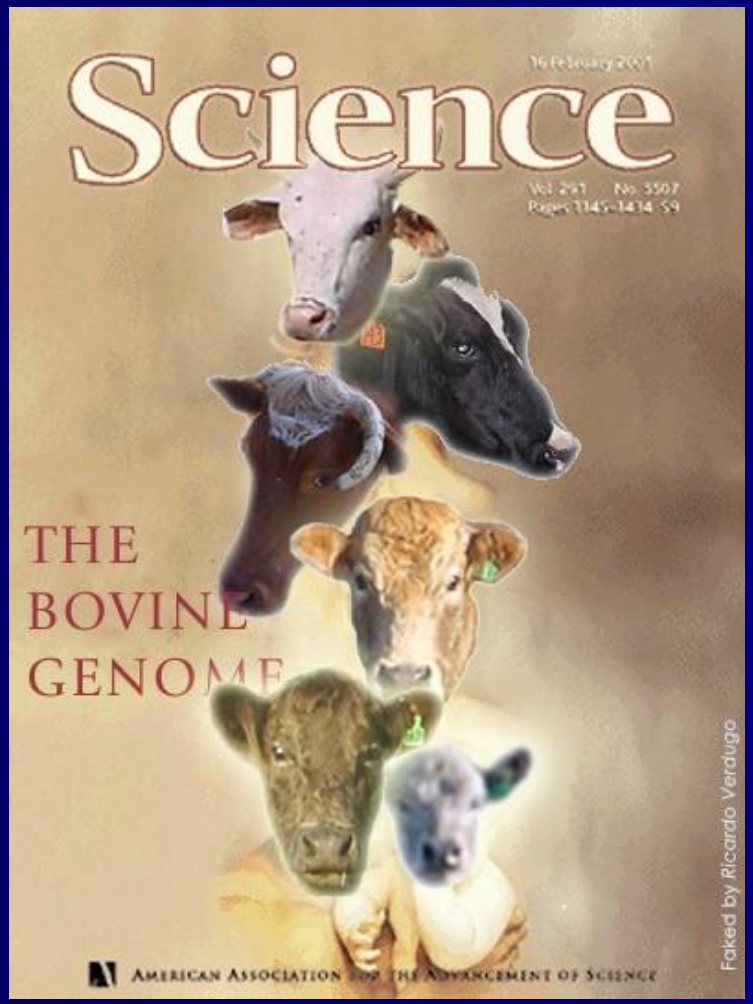
Unambiguous Assigment of Calves to a Single Sire Using a 28 SNP Panel versus a 23 STR Panel





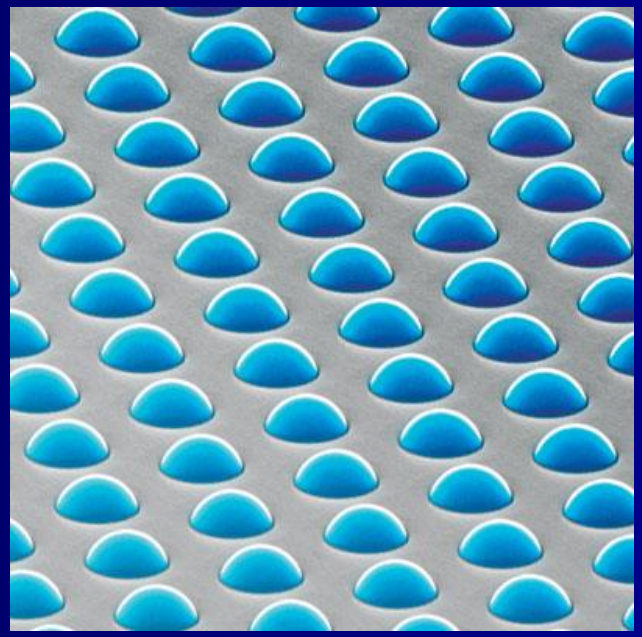
	28 SNP Panel – 27 sires 2005 (PE=95.5%)		62 SNP Panel – 23 sires 2006 (PE=99.975%)		99 SNP Panel – 28 sires 2007 (PE=99.999%)	
One sire assigned	175	23.3%	260	86.7%	294	97.0%
More than one sire	420	67.3%	16	5.3%	1	0.33%
All excluded	29	4.6%	24	8.0%	8	2.6%
TOTAL	624		300		303	

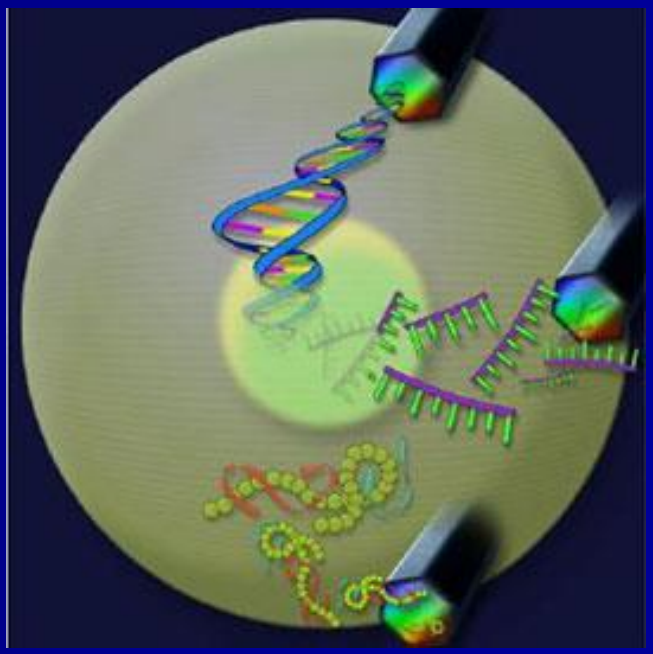
High-throughput SNP genotyping on 50,000 SNP CHIP (50K Chip)



The sequencing of the bovine genome allowed for a collaboration between MARC, BARC, UMC and UA to develop a set of 50,000 SNPs that are located throughout the entire genome

12 samples per BeadChip can be run on 50,000 SNPs at ~ $200/sample!





SNPs and parentage using the 50K chip

"The low rate of genotyping errors meant that less than five inconsistencies were usually found when parent-progeny assignment was correct. However, several thousand inconsistencies were usually found when the parent-progeny assignment was incorrect"

Wiggans et al. Genomic Evaluations in the United States and Canada: A collaboration. ICAR 2008

Implications

- Currently there are three competing SNP genotyping technologies – Affymetrix, Sequenom, and Illumina – prices are now less than 1 cent per SNP
- It is likely that SNP markers will replace alternatives (i.e. microsatellites) over the next 5 years

Commercial companies are offering DNA markers for use in **Marker-Assisted Selection** (MAS) for given traits

BOVIGEN

ANGUSURE

igenity

Marker-assisted selection is the process of using the results of DNA testing to assist in the selection of individuals to become parents in the next generation.

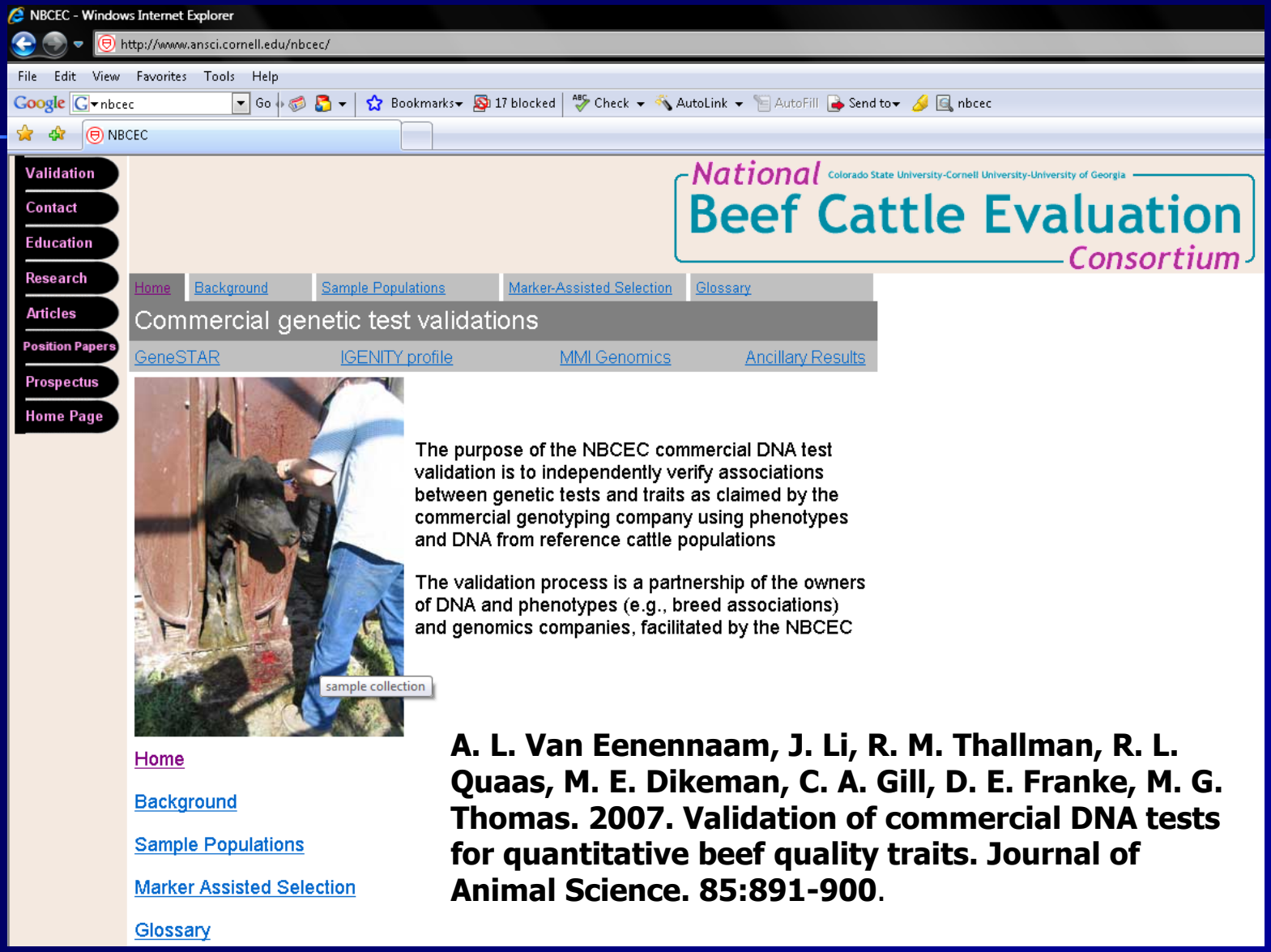
GENESEEK
Molecular Solutions for Breeding and Genetics

METAMORPHIX, INC.

Tests for quantitative traits – currently 10-100 SNPs

- Meat Tenderness
- Quality Grade (Marbling)
- Beef Cattle Feed Efficiency
- Meat Yield
- Disease Resistance
- Dairy Form
- Milk and Milk Component Yield

Independent validation of DNA tests http://www.nbcec.org/nbcec/



A. L. Van Eenennaam, J. Li, R. M. Thallman, R. L. Quaas, M. E. Dikeman, C. A. Gill, D. E. Franke, M. G. Thomas. 2007. Validation of commercial DNA tests for quantitative beef quality traits. Journal of Animal Science. 85:891-900.

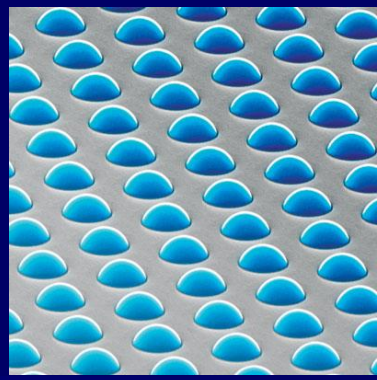
MAS (Marker-assisted selection)

- Currently available markers collectively account for 10% or less of the genetic variation
- A handful of markers is not enough for quantitative traits
- Hard to find all genes that affect a single trait
- Markers do not exist for many important traits
- Early adopters of genotyping for MAS in livestock have not experienced sufficient value capture i.e. they are too expensive

And DNA data is not being used in national cattle evaluation

- Only a small proportion of the population is being genotyped
- Individual producers may be reluctant to share results for animals that are shown to have inherited unfavorable marker alleles.
- There is no national structure, at the breed association or any other level, to routinely capture genotypic information in a consistent form for the purpose of national evaluation.

Whole genome-assisted selection (WGS)



- The use of dense SNP markers across the entire genome enables an estimation of the genetic merit of every chromosome fragment contributing variation in a population with phenotypic observations
- Can simultaneously test 50,000 markers
- Can be used to predict merit for all traits for which phenotyped populations exist

What is needed for whole genome-assisted selection?

THEORY

- Population
- Phenotypes
- Genotypes

Training = estimate the value of every chromosome fragment contributing variation in a population with phenotypic observations

Prediction = the results of training can then be used to predict the merit of new animals, not contained in the training data set

WGS effectively estimates an EPD for every chromosome fragment in the genome

Possible applications

- Product quality
- Feed efficiency
- Health
- Robustness
- Adaptability
- Stayability
- Reproductive traits
- Genetic disease resistance
  - Other difficult to phenotype traits

WGS compared to MAS

Genomic selection uses the estimated effect of many loci at once, not just the small number of statistically significant loci that are a feature of MAS (Dorian Garrick, Iowa State University)

As there are so many variants detected in WGS, the properties of them as a group becomes more important that their individual effects...It matters little if a specific variant fails under some circumstances as long at the majority of the variants are predictive. (John McEwan, NZ)

California to host BIF 2009! Mark your calendars!

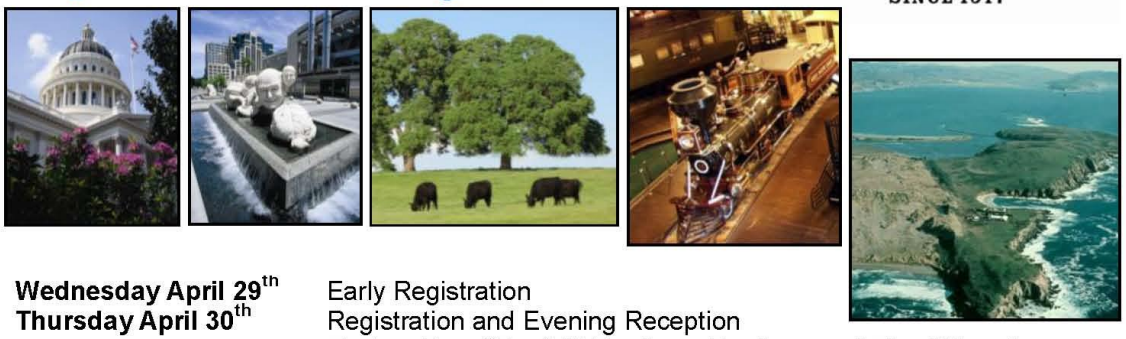
http://www.calcattlemen.org/bif2009.html

2009 Beef Improvement Federation Annual Research Symposium and Annual Meeting

Sacramento, California

April 30 – May 3, 2009

CALIFORNIA BEEF RUSH '09



Wednesday April 29th Early Registration
Thursday April 30th Registration and Evening Reception
Eastern Tour "**Foothill Bovines, Equines and Fine Wines**"
Friday May 1st Convention, Family/Spouse Tour, Evening Dinner
Saturday May 2nd Convention and Evening on your Own in Sacramento
Sunday May 3rd Western Tour "**Ocean Wines and Bovines**"

Questions ?