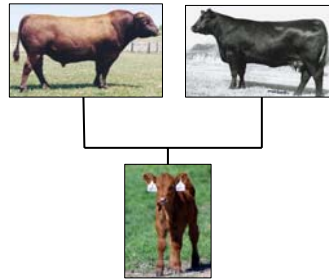


Parentage-based DNA traceback in beef and dairy cattle



Mike Heaton, Ph.D.

USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska

Updated 7-15-2008

Why is animal ID needed?

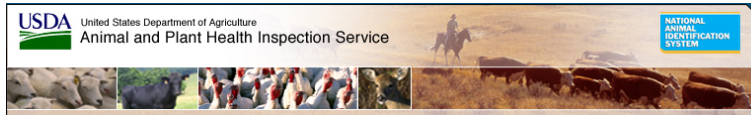
From a production perspective:

- Track carcasses at slaughter for beef improvement programs
- Verify source and age for branded beef programs
- Identify superior sires in multi-sire mating systems
- Identify mis-mothering cases
- Verify artificial insemination records
- Solve rustling cases
- Verify claims of tissue-residue violations at slaughter



Why is animal ID needed?

From an animal health perspective:



“...to help protect American animal agriculture from foreign or domestic disease threats.” APHIS

Goal: to traceback within 48 hours of discovery.

Desired outcome: reduced financial and social impact of disease outbreaks

Examples:

- brucellosis and tuberculosis eradication programs
- foreign animal disease - BSE traceback



Our goal:

to support DNA-based disease traceback in cattle.

Ways to physically label individual cattle

Branding



Tattooing



Tags and passports



The dispute

“There's some confusion about the paperwork....”

Which of the 9 downer cattle slaughtered that day had the BSE-infected brain?

“DNA testing by the best experts available could compare samples from the mad cow and its offspring or parents.”



Dr. Brian Evans
Chief Veterinary Officer
Canadian Food Inspection Agency

The resolution



USDA's Chief Veterinarian
Dr. Ron DeHaven of APHIS

December 31, 2003

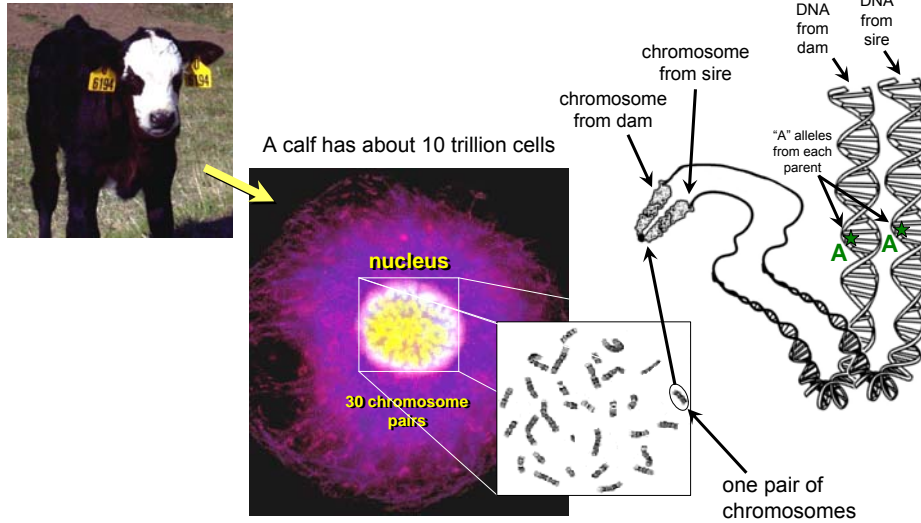
“...we are sending multiple samples to two laboratories -- one in Canada and one in the United States.”

“... the U.S. laboratory is in Nebraska, [and] It's a USDA laboratory that has that expertise.”

January 6, 2004

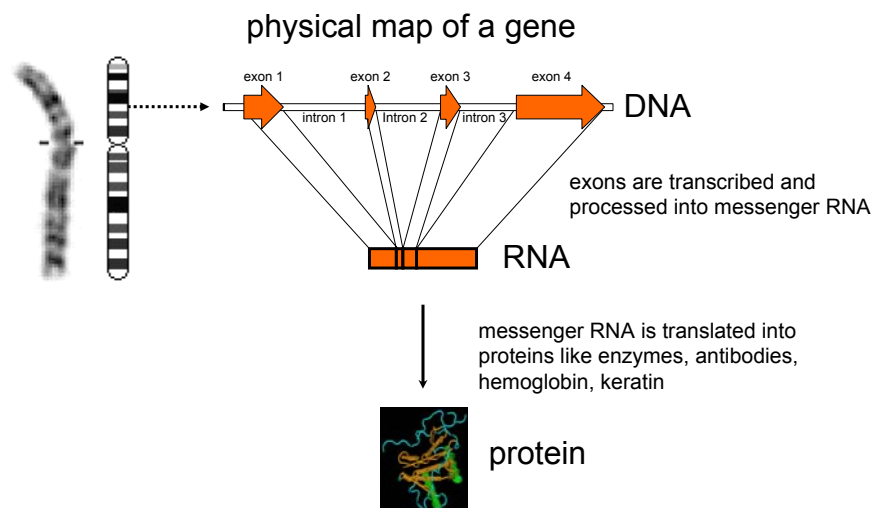
"We now have DNA evidence that allows us to verify with a high degree of certainty, the [Canadian] birthplace of the BSE-infected cow."

DNA is the label never removed from beef

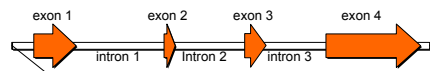
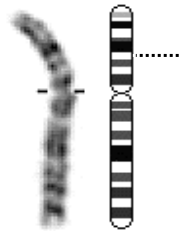


A chromosome has genes and other DNA

- A chromosome may contain more than 1000 genes.
- Genes comprise about 1 to 3 % of the genome.



Genes are encoded by sequences of DNA



nucleotides: A, C, G, T

```

1  tcgataaac  cctccaatgy  aaaagacatt  gggaggttcc  ggtaaaagcc  tactgcattt
61  gcaagtaact  gggagaggct  gtttactgtg  cagctccat  tcacacggca  ggaatacctg
121  aatagctgga  ccacaggtga  gcaaactcac  cttctccact  tgcctccctg  gaaagaagaa
181  atcaaaagga  aatctctctc  tccagtggtc  aaaatcgacc  gccaatgget  atctgtagca
241  tttctctggt  ctagaacaca  actgagaagg  aaaccttctg  cagggttggg  attgagctta
301  ggcgtatcta  caggagccac  acgtgcttga  acaattcct  gggcgactt  gacaggaatc
361  tcaacagctt  ggcaagcaag  gtaagctgtg  ctatggcaac  agcctgcctt  ccaatggatg
421  gcagtactac  ttcttaaaat  attctcctt  atgactagca  gtacactttt  aatcagcttg
481  ttcaatcatt  tgacaaatc  tatttggtga  gactctgtgc  gtgttgctta  taggtactgt
541  tctaggcact  ggggatacat  ttctgtctc  cacagaatgt  tgtaatgggt  cagttagaaa
601  agcaacaag  catacagcat  gatgtctagt  agtgataagt  gctatgaaga  aaaataaagg
661  ggtaaagagc  ataggggtca  agagggagag  ggtgatata  gtagctggcg  acatcatcta
721  agcagctatt  tgaattaagc  aagagagttg  gccacatcac  taacttgaa  ggaacattcc
781  aggcagaggy  aacagctgta  caaaggccct  aagtagaaa  taacttcatt  tgttcaaga
841  tcaacaggag  gccagttgtg  gctggatctg  agtaagttag  gggcttttag  gaggagacca
901  cacagggcct  atggattctg  ccccacgat  ggtgggaac  ttccacaggy  ttctgagcag
961  ggcagagact  accacctgat  ttatgtgatt  taaatgatc  attctggctt  ctgtgtggtt
1021  aacagactga  gtaggggtaa  gcacagaagt  aagatggctt  gttgggaagc  taccacagta
1081  ctccagctgc  agatgatggt  ggggtgttga  gagtaagagt  ggcagacatt  gttggatttt
1141  ggctatgatc  aaaaatcga
    
```

At what level does *most* of the genetic diversity occur in cattle?

Most of the genetic diversity in cattle occurs at the nucleotide level

DNA sequence of **maternal** chromosome: 5' -A-T-C-G-A-T-T-3'

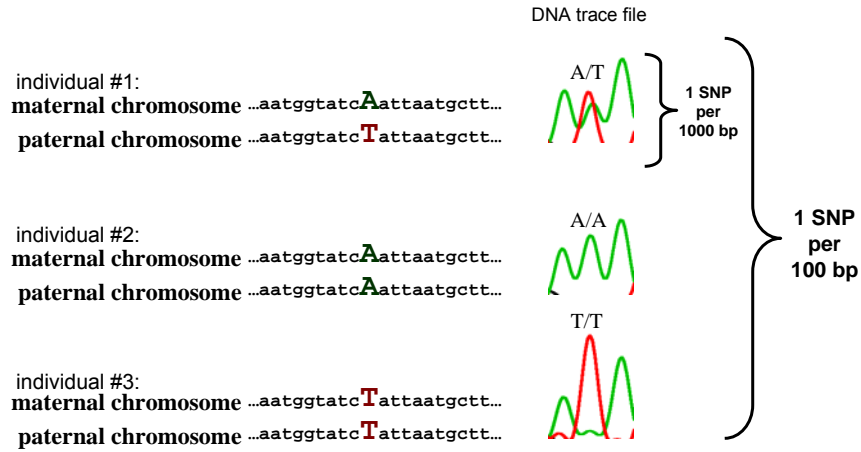
DNA sequence of **paternal** chromosome: 5' -A-T-C-T-A-T-T-3'

↑
SNP

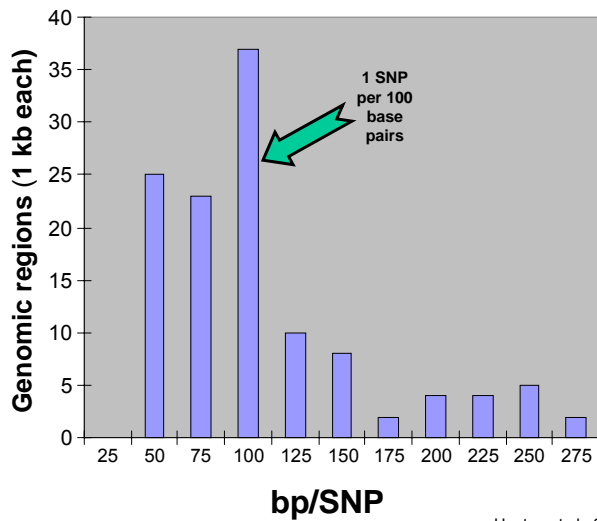
- Single nucleotide polymorphisms (SNPs) occur, on average, once in every 100 nucleotides in a diverse group of 96 U.S. cattle
 - Heaton et al., Mamm. Genome, 2002.

What are SNPs?

SNPs are sites in the genome where two different nucleotides may be observed



SNP density distribution per 1 kb genomic region (122 parentage markers tested, 216 beef and dairy cattle sequenced)



Heaton et al., 2006, unpublished results

Properties of SNPs

- Abundant (approximately 30 million sites in cattle)
- Stable (low mutation reversal rate)
- Amenable to high-throughput automatic scoring
- Low cost per SNP genotype
- Many genotyping platforms available
- SNP scoring is comparable between labs & platforms

Bottom line: SNP markers are the new “gold standard” for genetic analysis

Ways to use DNA for traceback

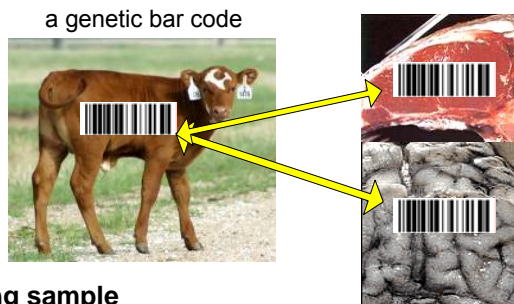
- **DNA fingerprinting (sample matching)**
 - comparing genotypes between samples
 - resolves disputes if samples were collected at the point of origin *before* a disease outbreak occurred.

- **Advantages:**

- high degree of power
- all genotypes useful

- **Disadvantages:**

- requires a preexisting sample



- **Example:** 20 SNPs were sufficient for verifying sample tracking in dairy beef processing (Heaton et al. J Am Vet Med Assoc. 2005).

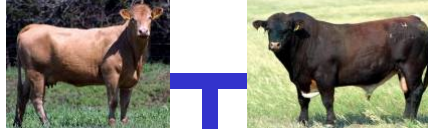
Ways to use DNA for traceback

- **Parentage analysis**

- determining whether alleles are shared between parents and offspring
- may confirm the origin of a diseased animal if tissues from a parent are available.

- **Advantages:**

- preexisting sample of “case” not needed



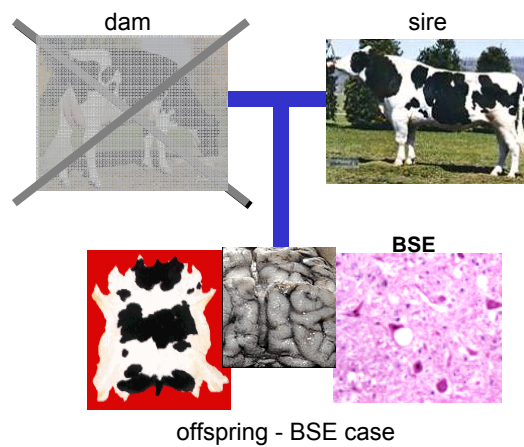
- **Disadvantages:**

- not all genotypes useful
- requires more markers
- requires more samples



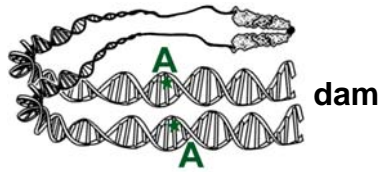
Sometimes parentage testing is the last resort for DNA-based traceback

- **Worst case scenario: only one parent available**
- 2003 Washington State BSE case

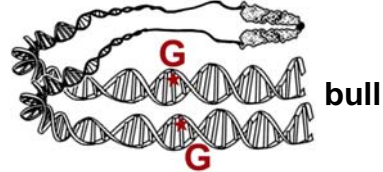


Using SNP markers for parentage

The offspring must share an allele with *each* parent



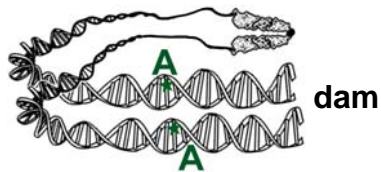
calf **A,A**



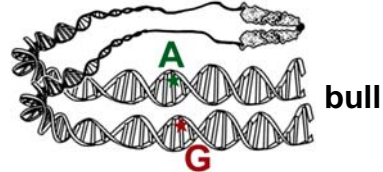
excluded sire **G,G**

Using SNP markers for parentage

The offspring must share an allele with *each* parent



calf **A,A**

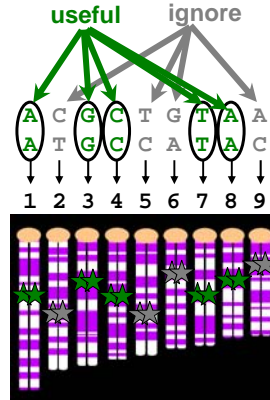


possible sire **A,G**

Accurate sire determination requires many DNA markers (example: 9 markers on 9 chromosome pairs)

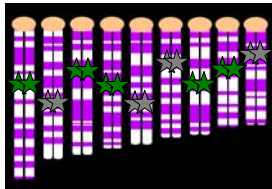
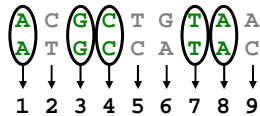
When comparing only sires and offspring (i.e. dams not used).....

- the calf's homozygous DNA markers are the most useful
- the calf's heterozygous markers are ignored because every sire will share an allele with the calf for those markers



Comparing the calf to the sire: elimination

Calf 1



Bull 1 $\begin{pmatrix} G \\ G \end{pmatrix}$ * * * * * * * *
* * * * * * * *
1
(eliminated)



Bull 2 $\begin{matrix} G & T & G & \begin{pmatrix} T \\ T \end{pmatrix} \\ A & T & G & \end{matrix}$ * * * * * *
* * * * * * * *
1 2 3 4
(eliminated)



Bull 3 $\begin{matrix} G & T & G & C & C & A & \begin{pmatrix} C \\ C \end{pmatrix} \\ A & T & G & T & C & A & \end{matrix}$ * *
* * * * * * * *
1 2 3 4 5 6 7
(eliminated)



Bull 4 $\begin{matrix} G & T & G & C & C & A & C & A & C \\ A & T & G & T & C & A & T & A & C \end{matrix}$
1 2 3 4 5 6 7 8 9
(not eliminated)



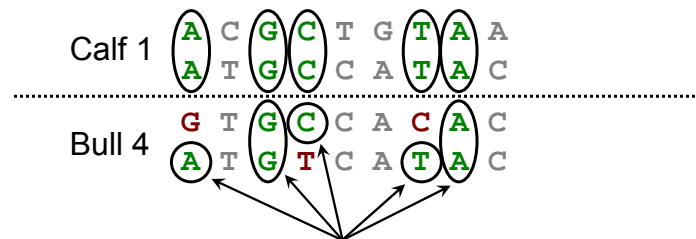
Comparing the calf to a possible sire



Calf 1



Bull 4



The ideal parentage SNP markers

- **have both alleles in equal proportions in all breeds**
 - balanced alleles provide more power
- **are evenly spaced throughout the genome**
 - more independent inheritance provides more power
- **can be accurately scored**
 - one wrong genotype can confound a traceback dispute
- **are publicly available**
 - encourages standardization and fair competition

The ideal marker is frequent in all breeds

A collaborative effort was undertaken to assemble many beef and dairy breeds for testing (screening) allele frequency

96 diverse sires from 19 beef breeds (Drs. Heaton and Laegreid; ARS, MARC)

464 cross-bred Canadian beef cattle containing germplasm primarily from Angus, Charolais, Hereford, Simmental, Galloway, and other breeds (Dr. Moore, University of Alberta)



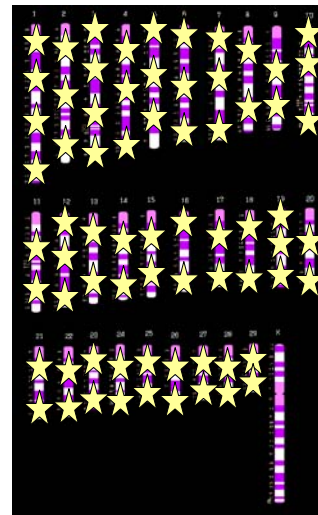
120 prominent sires from 4 dairy breeds (Drs. Van Tassell and Sonstegard; ARS, BARC)



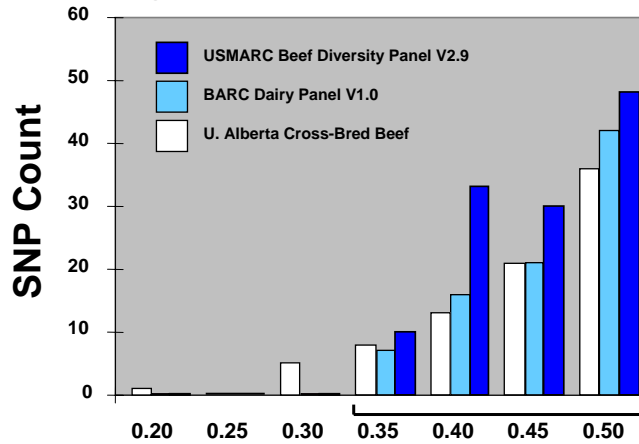
More than 4000 candidate SNPs, mostly from the Bovine Genome Project, were genotyped to select those with best minor allele frequencies (Drs. Heaton, McKay, Moore, and Murdock; MARC and U. Alberta)

The ideal markers are evenly distributed across the genome

- We used the MARC composite map to select regions for markers
- Developed by Dr. Warren Snelling at MARC, this map contains positions of thousands of DNA markers.
- We chose SNP candidates in chromosomal regions that were approximately 20 to 50 cM apart
- X and Y chromosomes not used



Distribution of minor allele frequencies for 121 parentage SNPs in US and Canadian cattle



Minor allele frequency

Collaborators:
Drs. Bennett and Keele, MARC
Dr. Moore, U. Alberta.

The ideal parentage marker scores accurately

- One wrong genotype may exclude the *true* parent.
- In disease traceback this causes delays, confusion, and loss of credibility.

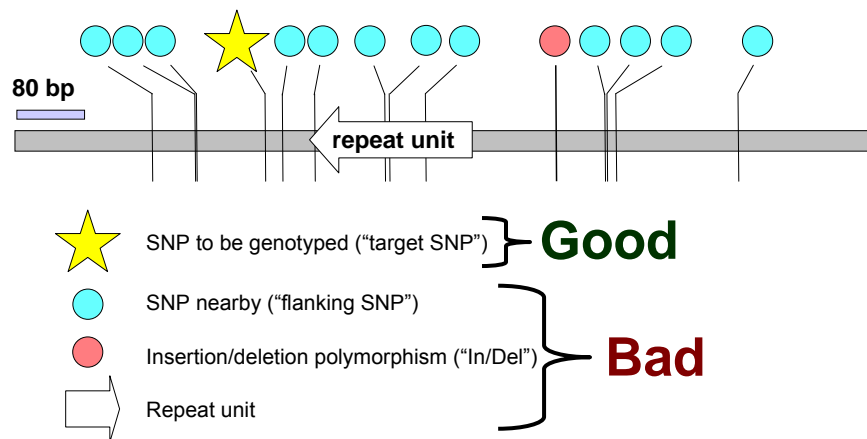


Accurate scoring requires that SNPs:

- are in unique genomic regions (non-repetitive)
- are in regions with well characterized flanking SNPs
- have no major insertion/deletions nearby

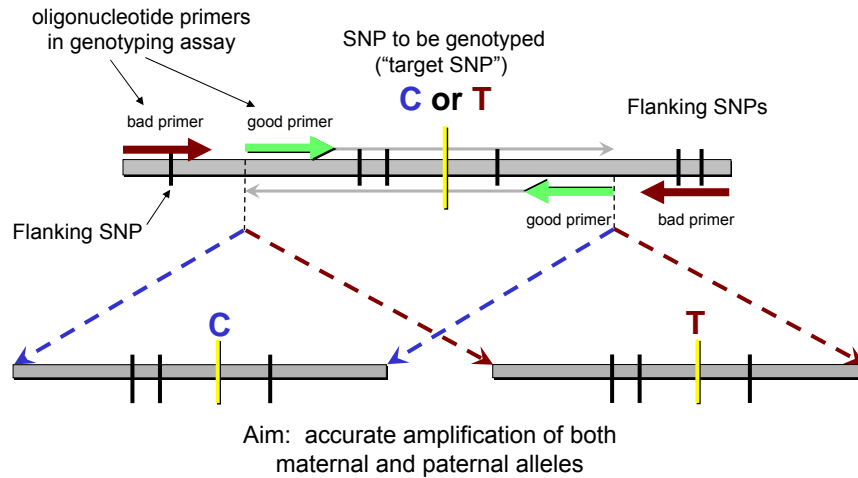
The nucleotide diversity in a typical region of the bovine genome

1 SNP occurs every 100 bp when comparing the 96 diverse beef sires from 19 breeds



The consequence of 1 SNP every 80 bp

- **Wrong genotype assigned to some animals**
 - Because some oligonucleotide primers do not bind correctly

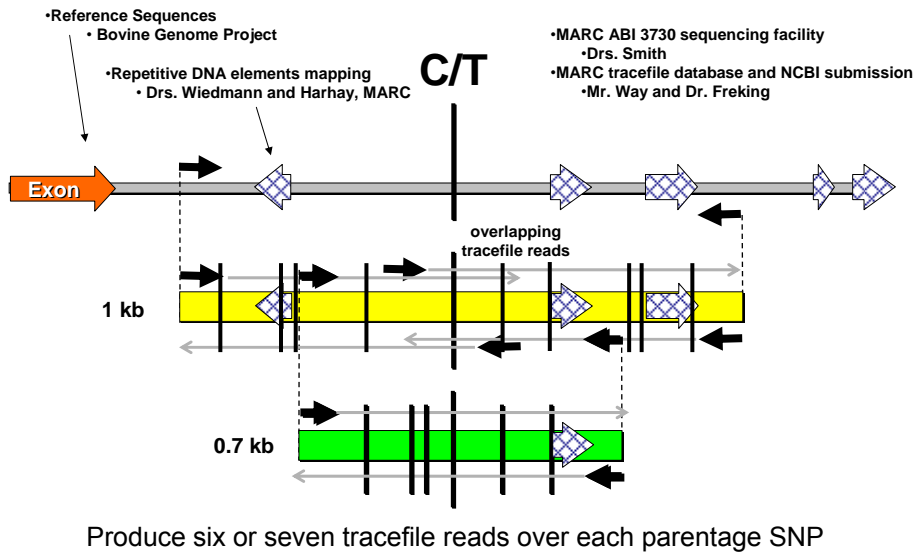


Bottom line: hidden SNPs may cause the wrong genotypes to be scored

- Increases costs
- Decreases throughput
- Frustrates customers *and* genotyping companies
- Decreases platform flexibility
- Decreases competition in the genotyping markets

Solution: determine the flanking DNA sequence of parentage SNPs in populations to be genotyped

DNA sequencing strategy: nested PCR



Parentage SNP population sequencing results

- 121 different regions with parentage SNPs analyzed
 - [USDA Bovine SNP Set for Parentage-Based Traceback \(UPT_BTA_2007.121.V1\)](#)
- 42 breeds with 24 diverse animals from each breed (1008 total cattle)

| | | | |
|--------------------|---------------|-----------------|--|
| Angus | Gelbvieh | Piedmontese | Collaborators: Dr. Neibergs, WSU Dr. Chase, ARS, STARS Dr. Bob Bohlender Mr. Goode, Goode Cattle |
| Ankole-Watusi | Guernsey | Pinzgauer | |
| Ayrshire | Hereford | Red Angus | |
| Beefmaster | Highland | Red Poll | Recent additions: Indu-Brazil Dexter |
| Belgian Blue | Holstein | Romagnola | |
| Blonde D'Aquitaine | Jersey | Salers | |
| Brahman | Limousin | Santa Gertrudis | |
| Brahmousin | Maine-Anjou | Senepol | |
| Brangus | Marchigiana | Shorthorn | |
| Braunvieh | Mini-Hereford | Simmental | |
| Brown Swiss | Mini-Zebu | Tarentaise | |
| Charolais | Montbeliard | Texas Longhorn | |
| Chianina | Murray Grey | Tuli | |
| Corriente | Nelore | Wagyu | |

Physical map of one region with a parentage SNP

BTA 27, 45 cM, GenBank no. EF034084

5273 bp

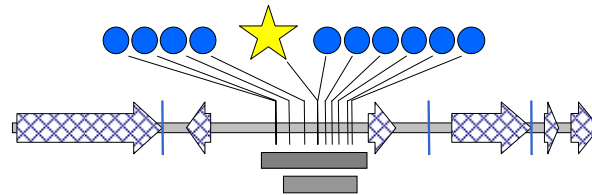
BTA27

★ Parentage SNP

● Flanking SNP

▢ Bovine repetitive elements

■ STS (amplicon)



For 121 parentage SNPs (first 216 of 1008 cattle)

0.425 average minor allele frequency beef
0.436 average minor allele frequency dairy
1646 adjacent SNPs
259 overlapping amplicons sequenced
137k tracefiles deposited at GenBank
89 Mb of sequence
2004 bovine repetitive elements
258 exons
183 CDSs

•Average of 9.8 kb annotated per file (1.2 Mb total)

•Collaborators in depositing GenBank annotation

•Dr. Anjanette Johnston
GenBank Submissions Staff

•Dr. Ilene Karsch Mizrahi
GenBank Coordinator

Immediate public access requested

- During marker development a number of companies wanted immediate access to the results.
 - Some companies needed to design tests as fast as we could generate the marker information
- How to provide fair public access?
- GenBank provides access but *not* easily

Public access to detailed SNP information

http://cgemm.louisville.edu/usmarc/MARC_web_page/traceback.html



University of Louisville

A SNP Marker Set for DNA-based Traceback in North American Beef and Dairy Cattle.

SNP Summary Table: SNPs that have been thoroughly screened to accomplish both DNA fingerprinting and parentage testing in U.S. beef and dairy populations, an ability that only a small fraction of known DNA markers have. These are markers are specially selected for optimum power, genome-wide distribution, accuracy in genotyping, and high-throughput "multiplexability". The target SNPs and surrounding DNA was sequenced in a group of 216 diverse sires from 19 beef breeds and 4 dairy breeds representing the vast majority of U.S. cattle. The average minor allele frequencies in beef and dairy are greater than 0.410. The DNA diversity in the adjacent regions has been documented for use in designing DNA tests that will be accurate in more than 99.9% of the North American population. Available information includes, flanking context, frequency summaries, cross references to other data.

Available Genotype Data: A table of consensus genotypes scored in the USMARC Beef Cattle Diversity Panels V1.0 and Dairy Cattle Diversity Panel V1.0.

Cattle Population Summaries: A table of animal population information corresponding breed information.

Research results and information provided by Dr. Michael Ramos, Animal Research Center (USMARC), Clay Center, NE.

USDA, ARS Project Number: 5438-32000-023-02 Specific Cooperative Agreement accession Number: 409004. For more information about the project please see "Project Info".

This project is in collaboration with scientists at the USDA ARS, Bovine Functional Genomics Laboratory in Beltsville, MD, the University of Alberta's Beef Genomics Laboratory in Edmonton Canada, the Center for Genetics and Molecular Medicine at the University of Louisville, KY, and Cogenics in Morrisville, NC.

Center for Genetics and Molecular Medicine (CGeMM)
 Director: Kenneth S. Ramos Ph.D.
 Director of Bioinformatics Operations: Ted Kalbfleisch Ph.D.
 Please direct questions or comments to [Ted Kalbfleisch](mailto:ted.kalbfleisch@louisville.edu)

Collaborator:
Dr. Ted Kalbfleisch,
U. of Louisville

•We first started with a "static" page for downloading key summary information
 •UPT_BTA_2007.121.V1

•Next, a set of interactive pages were developed that access relational database information.




Public internet access to flanking SNPs and allele frequencies by breed

Polymorphism Confirmations:

• gctgggcca[A-G]ccataggag(Fwd) AY335912.928

Minor Allele Frequencies:

| Breed | A | C | G | T |
|--------------------|-------|-------|-------|-------|
| Total | 0.410 | 0.000 | 0.590 | 0.000 |
| Wagyu | 0.410 | 0.000 | 0.590 | 0.000 |
| Hereford | 0.410 | 0.000 | 0.590 | 0.000 |
| Simmental | 0.410 | 0.000 | 0.590 | 0.000 |
| Red Angus | 0.410 | 0.000 | 0.590 | 0.000 |
| Charolais | 0.410 | 0.000 | 0.590 | 0.000 |
| Belgian Blue | 0.410 | 0.000 | 0.590 | 0.000 |
| Corriente | 0.410 | 0.000 | 0.590 | 0.000 |
| Limousin | 0.410 | 0.000 | 0.590 | 0.000 |
| Brahman | 0.410 | 0.000 | 0.590 | 0.000 |
| Lenghorn | 0.410 | 0.000 | 0.590 | 0.000 |
| Murray Grey | 0.410 | 0.000 | 0.590 | 0.000 |
| Maine Anjou | 0.410 | 0.000 | 0.590 | 0.000 |
| Watusi | 0.410 | 0.000 | 0.590 | 0.000 |
| Jersey | 0.410 | 0.000 | 0.590 | 0.000 |
| Brown Swiss | 0.410 | 0.000 | 0.590 | 0.000 |
| Chianina | 0.410 | 0.000 | 0.590 | 0.000 |
| Marchigiana | 0.410 | 0.000 | 0.590 | 0.000 |
| Blonde D'Aquitaine | 0.410 | 0.000 | 0.590 | 0.000 |
| Pinzgauer | 0.410 | 0.000 | 0.590 | 0.000 |
| Piedmontese | 0.410 | 0.000 | 0.590 | 0.000 |
| Tuli | 0.410 | 0.000 | 0.590 | 0.000 |
| Saeta Gertrudis | 0.410 | 0.000 | 0.590 | 0.000 |
| Montbeliard | 0.410 | 0.000 | 0.590 | 0.000 |
| Romagnola | 0.410 | 0.000 | 0.590 | 0.000 |
| Ayrshire | 0.410 | 0.000 | 0.590 | 0.000 |
| Gelbvieh | 0.410 | 0.000 | 0.590 | 0.000 |
| Chianus | 0.410 | 0.000 | 0.590 | 0.000 |
| Holstein | 0.410 | 0.000 | 0.590 | 0.000 |

allele count
39:0:15:0

SEQUENCE

target SNP: gctgggcca[A-G]ccataggag(Fwd)

flanking SNPs: [A-G]ccataggag(Fwd)

flanking SNP frequencies: [A-G]ccataggag(Fwd)

| FEATURE TYPE | POSITION | ALLELES | GENOTYPE COUNT | FEATURE ID | DB3 |
|--------------|-----------------------|---------|----------------|------------|-------|
| UTR | 46594033 (r 46594102) | C/T* | 589 | 8129424 | 81294 |
| polymorphNam | 46594036 | C/T* | 589 | 10712543 | 10712 |
| polymorphNam | 46594024 | C/T* | 588 | 10712543 | 10712 |
| polymorphNam | 46594018 | A/G* | 355 | 10712543 | 10712 |
| polymorphNam | 46593948 | A/T* | 229 | 10712543 | 10712 |
| polymorphNam | 46593967 | A/G* | 212 | 10712543 | 10712 |
| polymorphNam | 46593956 | C/T* | 223 | 10712543 | 10712 |
| polymorphNam | 46593976 | C/T* | 1008 | 10712543 | 10712 |
| polymorphNam | 46493766 | A/G* | 1009 | 10712543 | 10712 |

Individual animal genotypes linked to tracefiles

The screenshot displays the CCGeMM web interface. The main panel shows a table titled "Genotypes for Polymorphism '10369106 A/G'". The table has columns for ID, EXPERIMENTAL SAMPLE NAME, ETHNICITY, ALLELES, and TRACE. Several rows are visible, with the first few having circled values in the ID, ETHNICITY, and TRACE columns. A green arrow points from the TRACE column of the first row to a detailed view of a tracefile for SNP '10371598/A/G'. This detailed view shows chromatograms for three different samples, each with a genotype confirmation ID and a reference to the experimental sample.

| ID | EXPERIMENTAL SAMPLE NAME | ETHNICITY | ALLELES | TRACE |
|-----------|--------------------------|-----------|---------|------------------|
| 10371399 | 19999824_Dad_BullSample | Angus | A/G | Genotype Details |
| 11181366 | 19300403_Dad_BullSample | Balera | G/G | Genotype Details |
| 11181397 | 19300503_Dad_BullSample | Balera | G/G | Genotype Details |
| 11181635 | 19300549_Dad_BullSample | Balera | | |
| 11181644 | 19300552_Dad_BullSample | Balera | | |
| 11181830 | 19381043_Dad_BullSample | Romagnolo | | |
| 111819420 | 19981657_Dad_BullSample | Wagyu | | |
| 111819426 | 19981659_Dad_BullSample | Wagyu | | |
| 111819462 | 19981673_Dad_BullSample | Balera | | |
| 111819503 | 19981682_Dad_BullSample | Balera | | |
| 111819533 | 19981694_Dad_BullSample | Shorthorn | | |
| 111819508 | 20071927_Dad_BullSample | Shorthorn | | |
| 10371933 | 19931619_Dad_BullSample | Simmental | | |
| 10371961 | 19981623_Dad_BullSample | Charolais | | |
| 10371966 | 19981625_Dad_BullSample | Red Angus | | |
| 10371990 | 19981637_Dad_BullSample | Hereford | | |
| 10372223 | 19981639_Dad_BullSample | Limousin | | |
| 10372720 | 19981646_Dad_BullSample | Simmental | | |
| 10372739 | 19981648_Dad_BullSample | Simmental | | |
| 10372754 | 19981650_Dad_BullSample | Simmental | | |
| 10371623 | 19231212_Dad_BullSample | Hereford | | |
| 10371676 | 19711114_Had_Kuduswala | Hereford | | |

Where are we now?

- **Many of the 121 USDA Parentage-Based Traceback SNPs**
 - are on the USDA-Illumina Bov50k SNP chip
 - have been adapted by companies into their parentage tests
 - have been used in additional peer-review research publications
- Database development and uploading of detailed SNP information is ongoing at U. of Louisville and MARC
- The 121 USDA Parentage SNPs are continually being sequenced and tested in new breeds to evaluate their utility in the event that an emergency DNA-based traceback is needed.
 - The results are being made available as they are produced
- This set of SNPs currently represents the best group of well-characterized parentage SNPs available.

Mike Heaton, Ph.D.
 USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska

Updated 7-15-2008