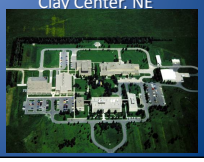
Low-cost, High-throughput Genotyping and Potential Applications to the Cattle Industry– Part 1 of 2

Heather Koshinsky
Eureka Genomics
Hercules, CA & Houston TX

R. Mark Thallman
USDA-MARC
Clay Center, NE



Who is Eureka Genomics?

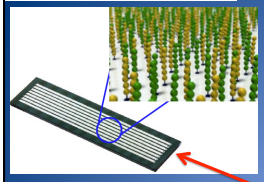
- Sequence data generation and analysis applied in three broad areas
- Low cost genotyping = 100s of SNPs starting at $5/animal
- Massive Genotyping by Sequencing Technology

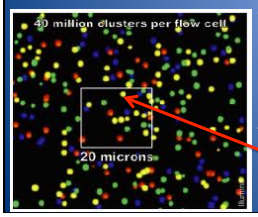
What is next generation sequencing?

- What is Next Generation Sequencing (NGS)?
  - High Throughput Sequencing (HTS)
- Extremely cost effective way to generate HUGE amounts of sequence data
  - $500 to $700 = 5 to 120 MILLION reads (sequence data)
- **$1 = 7,000 to 30,000 to 240,000 reads**

Next Generation Sequencing Platforms

Applied Biosystems
SOLiD Analyzer





8 lanes per run (takes 2 to 5 days)

30 million reads per lane

Read = a datapoint of sequence.
AGCTGCTGATCGTGATCGTAGCTGAT

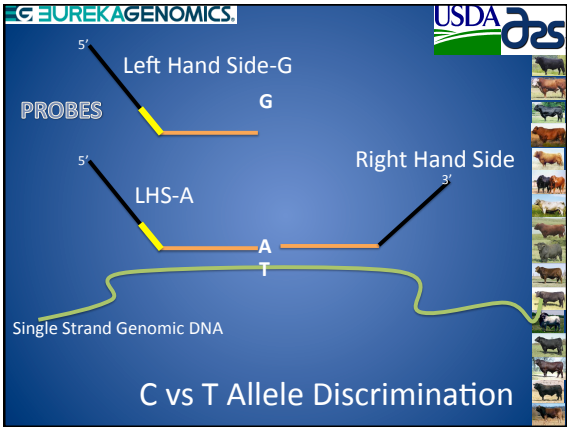
One read comes from one dot

How can next generation sequencing be applied to genotyping?

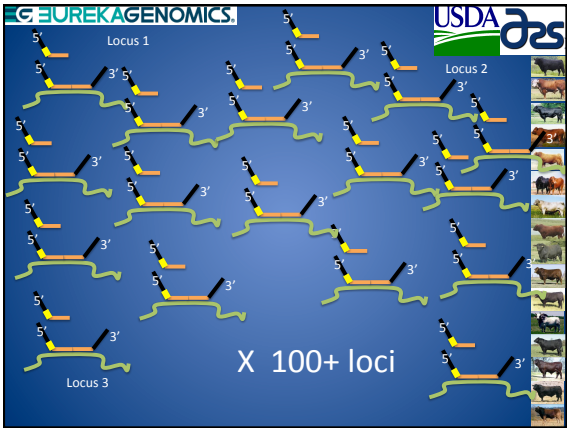
- Focus the sequence data on the informative loci
- Mix 1000+ animals in one lane in a manner that the sequence data can be taken apart

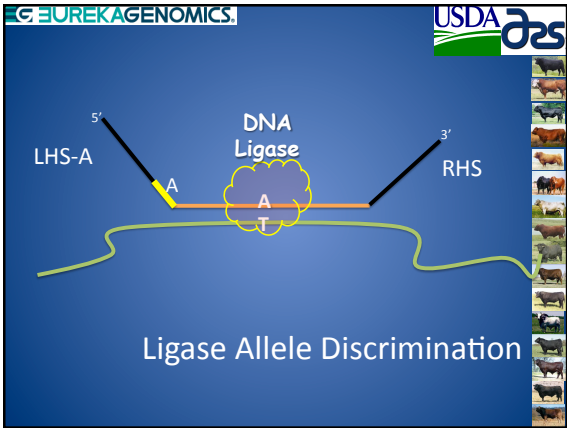
How to address the two tasks?

Ligase Dependent – Polymerase Chain Reaction (LD – PCR), within each well of these plates

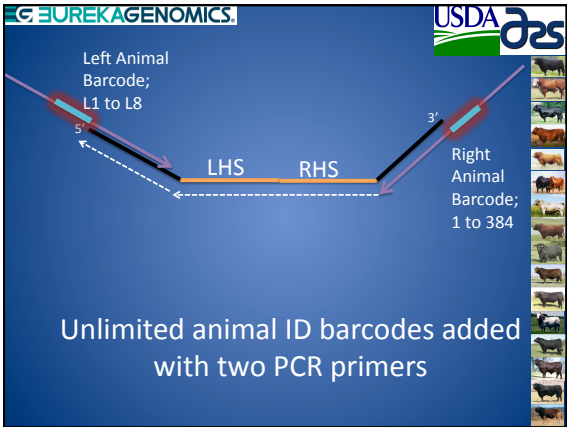


C vs T Allele Discrimination





Ligase Allele Discrimination



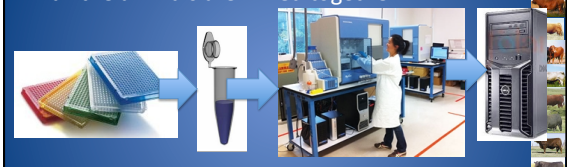
Unlimited animal ID barcodes added with two PCR primers

Now in each well (= 1 animal)

- Unique animal identification tags (barcodes)
- added to the alleles at all the loci
- with the bits that let them go on the sequencer

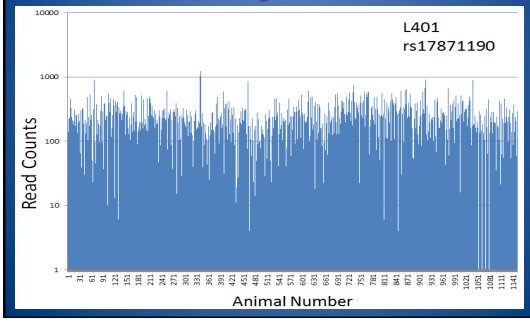
Generate the Sequence Data

- Pool the reaction wells
  = all the animals are mixed together

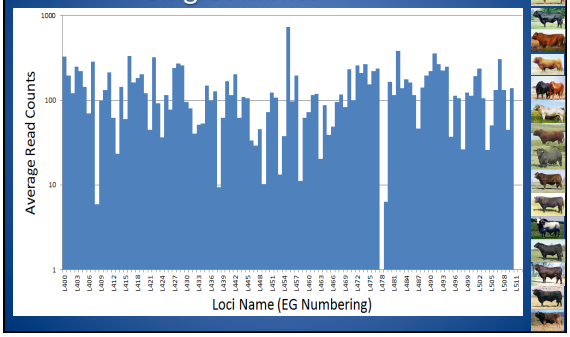


- Generate sequence data = individual reads
- Use the computers to pull it all back apart
  Internal controls, Animal identification, Locus & Allele(s)

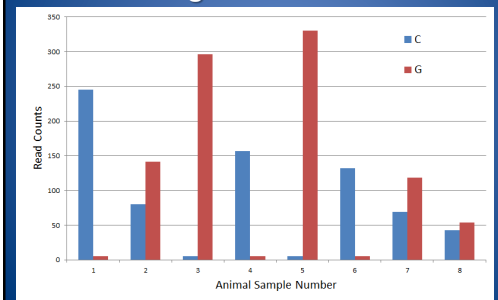
Reads Counts per Animal for a Single Locus



Read Counts per Locus for a Single Animal



Read Counts per Animal for a Single Locus.



Price and Turn Around Time

- Routine running (like the parentage panel)
  - $5/animal
  - DNA (1000 animals) → Genotype report
  - 2 weeks
- Optimized running

Price and Turn Around Time

- Routine running
- Optimized running
  - $2.50/animal
  - DNA (10,000+ animals on a routine basis) → Genotype report
  - 1 week

Not yet at "capacity"

- Low cost of the reads
- Improved instruments
- Overkill on reads per animal

- 1000s loci per animal (≠ 10X cost of 100)
- Increased animals per lane
- Lower cost per animal
- Shorter turn around time

Working with EG

Contact Information
Didier Perez
415-269-0666
didier@eurekagenomics.com

eurekagenomics.com

Low-cost, High-throughput Genotyping and Potential Applications to the Cattle Industry– Part 2 of 2

Heather Koshinsky
Eureka Genomics
Hercules, CA & Houston TX

R. Mark Thallman
USDA-MARC
Clay Center, NE

Overview

- Validation of Mass Genotyping by Sequencing Technology
- Applications of Mass Genotyping by Sequencing Technology
- Potential technology for sparse genome scan genotyping
- Applications of sparse genome scan

Validation Data

- 1080 animals genotyped with Illumina 50K chip
- 95 SNP out of Mike Heaton's 112 SNP parentage panel (all on 50K chip)
  - Used all genotypes at these SNP for which the 50K chip gave a call
- Genotype calling algorithm tuned for parentage application
  - Emphasizes accuracy over call rate

Results of Mass Genotyping by Sequencing Technology

Rank	Marker	Concordance	Call Rate
1	ARS-USMARC-Parent-AY842472-rs29001941	1	0.99
2	ARS-USMARC-Parent-DQ381153-rs29012842	1	0.98
3	ARS-USMARC-Parent-DQ786758-rs29024430	1	0.99
4	ARS-USMARC-Parent-DQ995976-no-rs	1	0.99
5	ARS-USMARC-Parent-DQ786766-rs29012070	1	0.99
6	ARS-USMARC-Parent-AY842474-rs29003226	1	0.99
90	ARS-USMARC-Parent-DQ647187-rs29010510	0.97	0.96
91	ARS-USMARC-Parent-AY851162-no-rs	0.97	0.91
92	ARS-USMARC-Parent-DQ916057-rs29009979	0.97	0.87
93	ARS-USMARC-Parent-DQ647188-rs29011099	0.96	0.82
94	ARS-USMARC-Parent-DQ984827-rs29012019	0.96	0.89
95	ARS-USMARC-Parent-DQ839235-rs29012691	0.95	0.88
Total		0.991	0.961

1080 Animals, 95 SNP Markers

Results of Mass Genotyping by Sequencing Technology

Animal Rank	Concordance	Call Rate
1 - 100	1.000	0.969
101 - 200	1.000	0.977
201 - 300	1.000	0.977
301 - 400	1.000	0.973
401 - 500	1.000	0.98
501 - 600	1.000	0.98
601 - 700	0.999	0.979
701 - 800	0.989	0.984
801 - 900	0.989	0.962
901 - 950	0.988	0.939
951 - 1000	0.978	0.966
1001 - 1050	0.970	0.928
1051 - 1060	0.951	0.915
1061 - 1070	0.652	0.694
1071 - 1080	0.181	0.304
1 - 1080	**0.991**	**0.961**

Cumulative Concordance = .994
Call Rate = .971

1080 Animals, 95 SNP Markers

Validation Results

- Error rate includes any errors in the 50K chip genotypes that we used as the standard

Applications of Mass Genotyping by Sequencing Technology

- Seedstock
  - Parentage Testing
  - Paternity Determination
  - Genetic Defect Testing
  - Testing for single gene traits
    - Color, polledness, F94L, etc.
  - Prediction of Economically Relevant Traits from Small Panels of SNP
  - New Recessive Defects Discovered by Individual Bull Sequencing

Applications of Mass Genotyping by Sequencing Technology

- Feedlot
  - Marker Assisted Management from Small Panels of SNP
- Commercial Ranch Project

Technology for Sparse Genome Scan Genotyping

- Roughly 1,000-5,000 markers spread evenly across the genome to allow tracking inheritance throughout the genome of a pedigreed population.
- Such a product is currently available for about $30/animal and used for parentage in Holsteins.
- In the coming years, such product can be a reality.
  - The objective is to make it sufficiently inexpensive for it to be feasible to genotype every registered animal with this technology.

Objective of Sparse Genome Scan

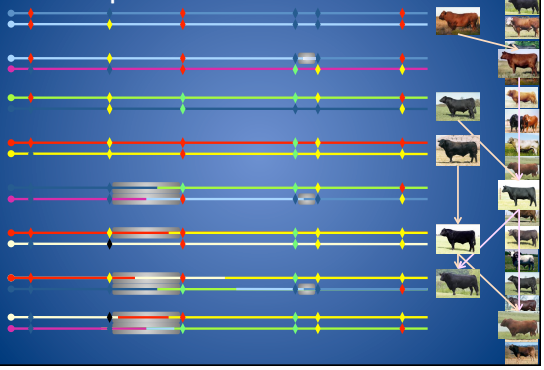
- To provide an inexpensive replacement to most other DNA testing that might be done on seedstock cattle.
  - But, to be effective, it needs to be applied to the whole herd, or better yet, the whole breed.

Sparse Genome Scan as Part of a Complete Pedigree Genotyping System

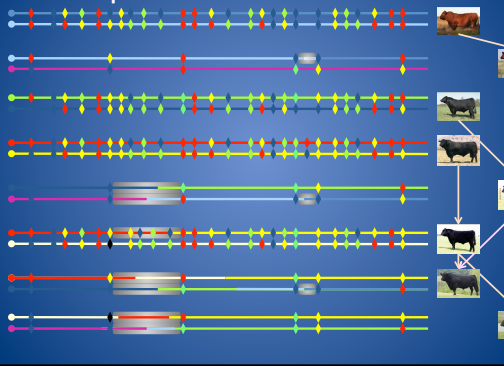
- Run sparse genome scan on most or all of the animals in the breed.
- Run the 50K chip on most or all AI sires in the breed.
- Individually sequence as many of the highly influential sires in the breed as is feasible.

Sparse Genome Scan

Sparse Genome Scan



Sparse Genome Scan



Applications of Sparse Genome Scan Genotyping

- Parentage verification
  - Could improve accuracy of conventional EPDs
- Paternity determination
  - Greater power than conventional tests
  - Extend genetic evaluation into commercial herds
- Genetic defect testing
- Prediction of quantitative traits
  - Through imputation of 50K or denser SNP
- Testing for single gene traits
  - Color, polledness, F94L, etc.

Applications of Sparse Genome Scan Genotyping

- For commonly recorded traits, it could greatly extend the feasible size of discovery and validation populations for genomics:
  - Discovery on sires with 50K genotypes
  - Validation on progeny with sparse genome scan

Longer Term Applications of Sparse Genome Scan Genotyping

- Detection and selection against lethal recessives that cause early embryonic mortality
- Use of better models (based on identity by descent) for genetic evaluation
  - Will require substantial statistical and software development

Thanks

- Amanda Lindholm-Perry
- Linda Flathman
- John Pollak

- John D Curry
- Paul Dier
- Maria Shin
- Jessica Nguyen
- Viacheslav Fofanov
- Nadeem Bulsara
- Jingtao Liu

Contact Information
Didier Perez
415-269-0666
didier@eurekagenomics.com