


Whole genome sequencing in livestock species : The end of the beginning ?
 André Eggen, INRA, France

EADGENE European Animal Disease Genomics Network of Excellence for Animal Health and Food Safety

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

Whole genome sequencing in livestock species : The end of the beginning ?

André Eggen
 Laboratoire de Génétique biochimique et de Cytogénétique (LGBc)
 INRA Département de Génétique Animale
 JOUY EN JOSAS, France





This publication represents the views of the Authors, not the EC. The EC is not liable for any use that may be made of the information.


EADGENE

Human Genome Project (HUGO)

- A catalyser
- contributed to the improvement of technologies and strategies
- Awoke dreams for the study of livestock genomes

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

Why are we doing it ?

- Basic knowledge
 - What is a genome ?
 - How it functions ? (genetic mechanisms)
 - To understand the history of life.
- Applied knowledge
 - Genetic disorders
 - « Pharmaco-genomics »
- Livestock species
 - Genetic control for complex Phenotypes
 - Biomedical models (Growth, Development, Health, Reproduction, ...)

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

Limited

- For livestock, one thought that only few « representative genomes » will be sequenced
- White papers to make the case
 - Improving human health
 - Informing human biology
 - Expand knowledge of basic biological process relevant to human health
 - Provide surrogate systems for human experimentation
 - **Facilitating the ability to do experiments (positional mapping)**

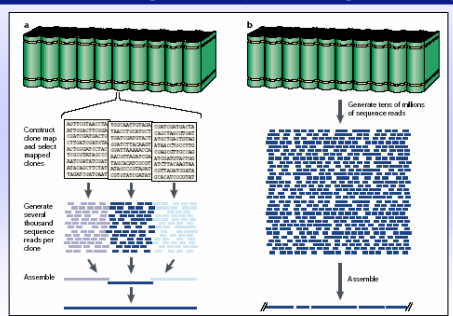
Example : Cattle

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

Two main shotgun-sequencing strategies

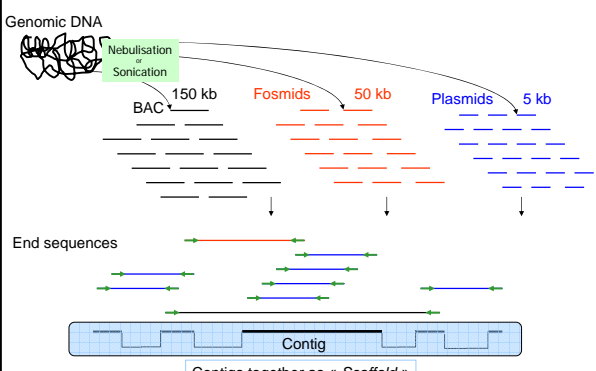


Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK



EADGENE

Whole genome sequencing (shotgun)




Contigs together as « Scaffold »

Whole genome sequencing in livestock species : The end of the beginning ?

André Eggen, INRA, France

EADGENE

The Horse genome



- Broad Institute
- 7 x coverage, female thoroughbred "Twilight"
- Sequences : **Whole Genome Shotgun (700 pb)**

Clone Size	Sequences	Coverage
4 kb	20 . 10 ⁶	4,97 x
10 kb	6 . 10 ⁶	1,42 x
40 kb	1,7 . 10 ⁶	0,34 x
BES 200 kb	0,2 . 10 ⁶	
Total		6,77 x

10th - 11th June 2008, Edinburgh, UK

EADGENE

Livestock Genomes sequenced (2008)

Species	Genus	Class	Assembly Method	Project	Status
Cow	<i>Bos taurus</i>	Mammal	Draft Assembly (7x)	BCM-HGSC	Completed
Cat	<i>Felis catus</i>	Mammal	Draft Assembly	WUGSC	Completed
Chicken	<i>Gallus gallus</i>	Non-Mammalian Vertebrate	Draft Assembly (6,6x)	WUGSC	Completed
Dog	<i>Canis familiaris</i>	Mammal	Draft Assembly	BI/MIT	Completed
Horse	<i>Equus caballus</i>	Mammal	Draft Assembly (7x)	BI/MIT	Completed
Pig	<i>Sus scrofa</i>	Mammal	Draft Assembly (BAC to BAC)	Sanger	In Process
Rabbit	<i>Oryctolagus cuniculus</i>	Mammal	Low Coverage WGS (~2X)	BI/MIT	Completed
Sheep	<i>Ovis aries</i>	Mammal	Draft Assembly	BCM-HGSC	In progress

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

EADGENE

A take home message

- Although draft sequences of several genomes are available, still much of work remains to fully benefit from the tremendous amount of data generated by the different sequencing initiatives in livestock species


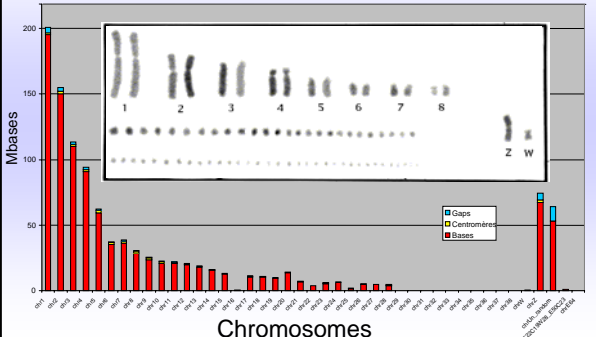
- Chicken
- Pig
- Cattle

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

EADGENE

The chicken genome

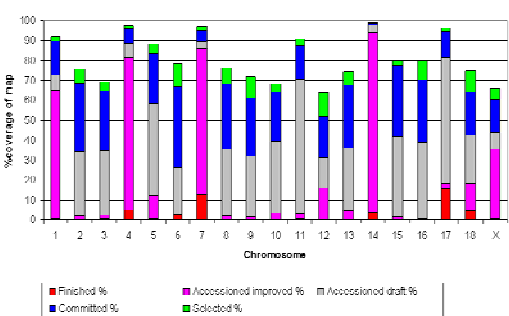
- Washington University
- Genome Size : 1,2 . 10⁹ bp
- 6 x WGS, Physical map

Chromosomes

EADGENE

Sequence clone progress 20/03/08




13009 clones selected and sent for sequencing covering 81.1% of the map.
Total sequence = 1406 Mb from 8500 sequenced clones
4521 Improved/finished, 58.8 Mb finished.

EADGENE

The Bovine Genome

(Baylor College of Medicine, Houston, USA)



- Coverage : 7,1 with WGS, BES, BAC sequences from the international physical map
- Reference animal : « Dominette », Hereford

	Btau 1.0	Btau 2.0	Btau 3.1	Btau 4.0
Availability	September 2004	March 2005	August 2006	October 2007
Strategie	WGS	WGS	WGS + BAC + BES	Improved assembly
Coverage	3 x	6.2 x	7.1 x	
Sequences	9 Gb	17,7 Gb	19 Gb	

Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK


Whole genome sequencing in livestock species : The end of the beginning ?
 André Eggen, INRA, France

EADGENE

SNPs as added value from Genome sequencing

- Reference sequence (6-12x Coverage)
- Several 100k sequences from several animals belonging to different breeds
- Align Sequences to identify several 100k SNPs
- SNP validation with a restricted number of animals from different breeds
- Build a whole genome SNP array (International Consortium)

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

High Throughput Sequencing

	Illumina Genome Analyzer	Roche NimbleGen 454 FLX system	Applied Biosystems SOLiD Analyzer
Amount of DNA	100 ng-1µg	?	5-10 µg
Volume of sequence	1,5 Gb (3 Gb)	0,1 Gb	1-1,5 Gb (1,5-2 Gb)
Size	36 pb (50 -70 bp)	200-300 pb (400 pb)	35 pb (2 x 25 pb)
<i>de novo Sequencing</i>	NO	YES	NO
Resequencing	YES	YES	YES
Expression studies	YES	YES	YES

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

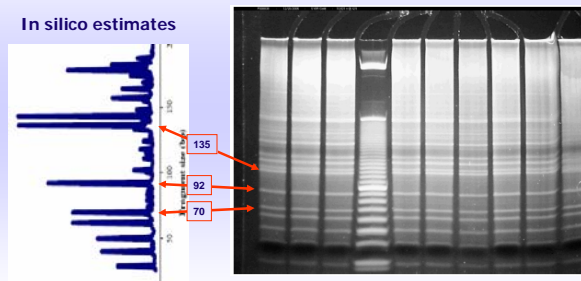


EADGENE


Van Tassel *et al.*, 2008
 Nature Methods 5, 247-252

In silico digestion matches gel

In silico estimates



Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

Van Tassel *et al.*, 2008
 Nature Methods 5, 247-252

High Throughput Sequencing

- 71 millions reads produced
 - 25 bases-read, total of 1.7 Gb
- 3.7 millions were adaptor-sequences
- 7.6 millions not starting with CC (*Hae*III GGCC)
- 15 millions reads eliminated through QC
- 49 millions accurate and "useful" reads (1.3 Gb)

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK



EADGENE

Van Tassel *et al.*, 2008
 Nature Methods 5, 247-252

SNP Identification (Reduced Representation Library)

GGCCAGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGGC
HaeIII Reference sequence HaeIII

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG } Holstein

CCACGTAATCGGCACCTTAGGATCTGC

CCACGTAATCGGCACCTTAGGATCTGC

CCACGTAATCGGCACCTTAGGATCTGC

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG } Angus

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG


CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG } BEEF

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG

CCACGTAATCGGCACCTTAGGATCTGCaacct....cactggATGAAGGCTCAATTGGCCATCATCGG

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




EADGENE

Van Tassel *et al.*, 2008
 Nature Methods 5, 247-252

Identification of Bovine SNP

- Analysis of 49 millions reads
 - 5 022 143 unique tags
- 62 042 putative SNPs identified @ high stringency
- 1 SNP / 1780 pb
- 25 125 SNPs validated and integrated to the «BovineSNP50 genotyping BeadChip» (ILLUMINA) with over 54 000 SNPs

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

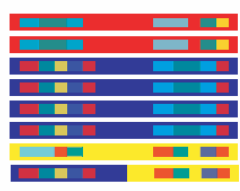


Whole genome sequencing in livestock species : The end of the beginning ?
 André Eggen, INRA, France

EADGENE

Bypassing the mapping step ...!

- Tens to hundreds of thousands of SNPs genotyped for the same price as a few hundred microsatellites
- Direct interrogation of Linkage Disequilibrium (LD) (relaxes the requirements for specific pedigree relationships)
- In Dog
 - causal mutations embedded in shared short-range (<10-Kb) haplotypes predating breed creation
 - breed-specific long-range (>100-Kb) haplotypes created at the time of breed formation
- Drastically reduction of the localisation interval of the QTLs (< 500Kb)

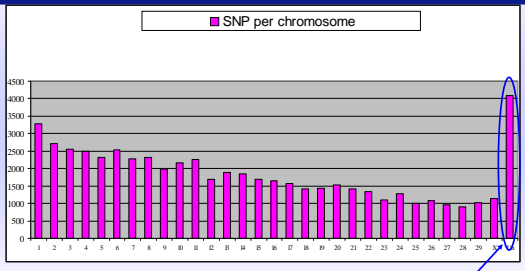


Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

INRA

EADGENE

The bovine SNP50 Genotyping BeadChip



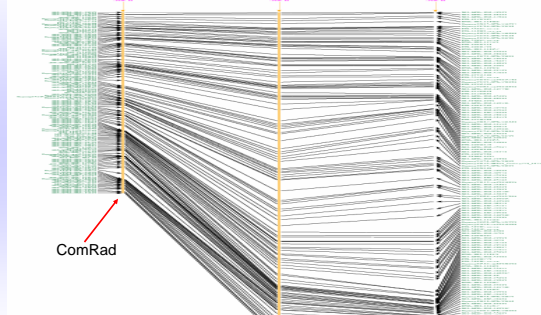
Total Assays : 58 331
 Mean Interval : 45,878 Kb

Unassigned SNPs

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

INRA

RH – Bovine Assembly Comparative Map



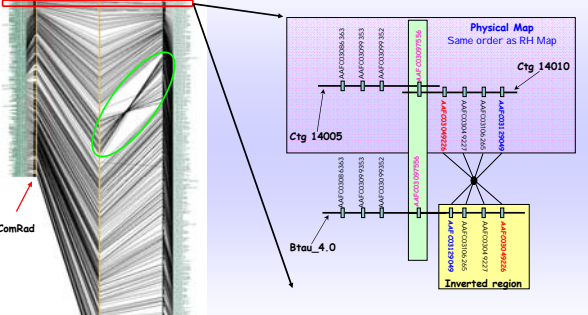
ComRad

Texas A&M

Btau 4.0

BTA22

RH – Bovine Assembly Comparative Map



ComRad

Texas A&M

Btau 4.0

Physical Map
 Same order as RH Map

Ctg 14005

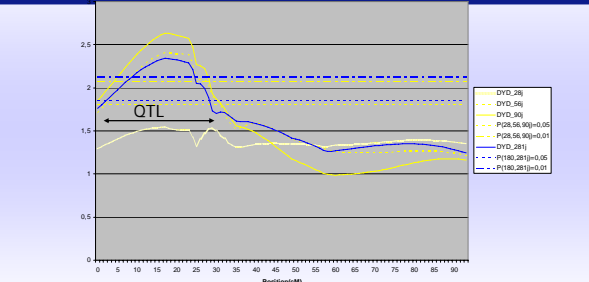
Ctg 14010

Inverted region

BTA14

EADGENE

A female fertility QTL on BTA03



QTL

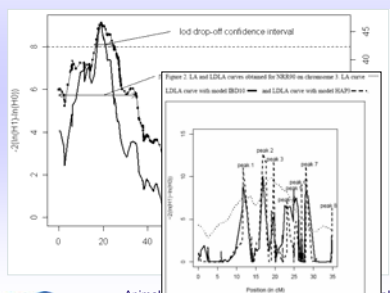
Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

INRA

EADGENE

Fine mapping in Holstein Population

(Druet *et al.*, 2008)



lod drop-off confidence interval

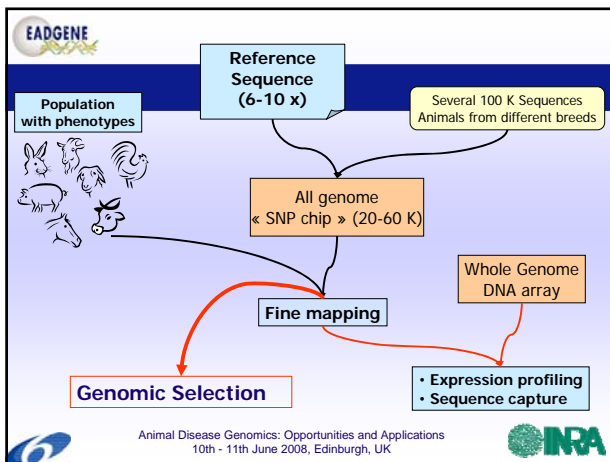
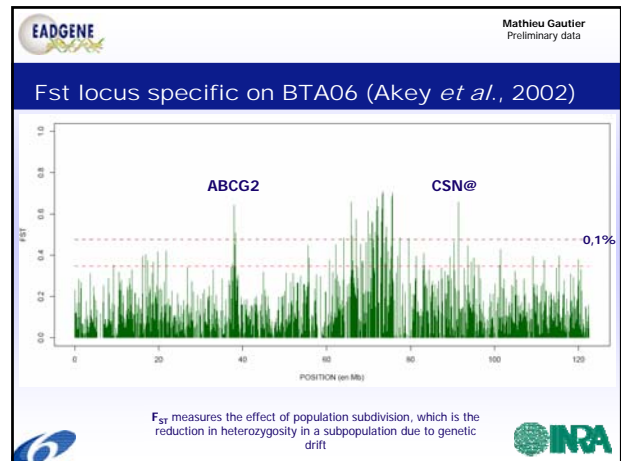
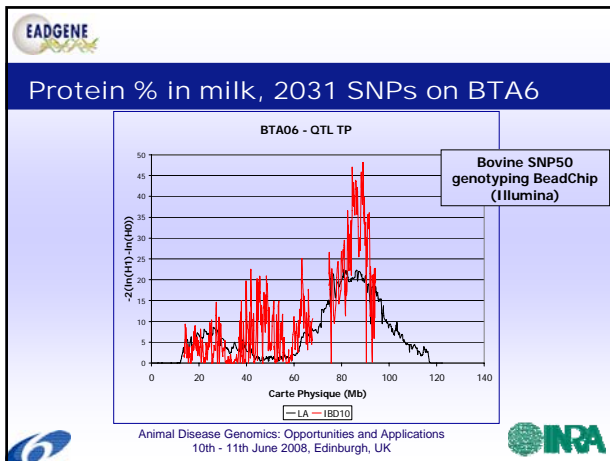
Female Fertility

- Dense map for BTA03 (460 SNP)
- Validation and fine mapping of the QTL in a single step

Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

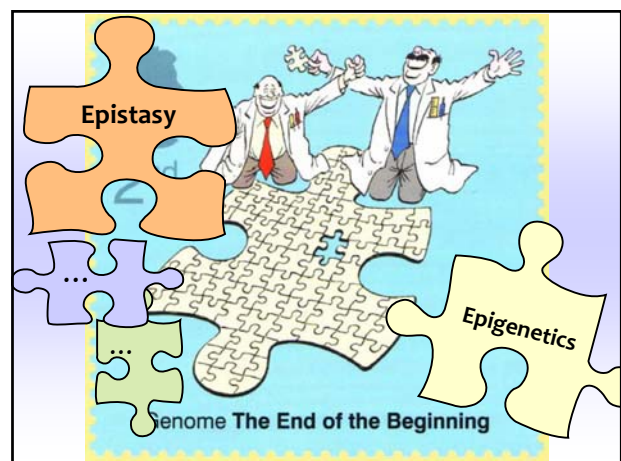
INRA

Whole genome sequencing in livestock species : The end of the beginning ?
 André Eggen, INRA, France




- Genomic Selection**
- Furthest advanced : Cattle
 - Australia, US, NRF : cor (GWEBV,proof) ~ 0.6
 - 25k SNP chip
 - Australia used QTL-MAS approach (Ben Hayes, pers Com)
 - NRF -> BayesB & BLUP of marker effects
 - General rumor :
 - 'It is not as easy as we thought'
 - More research is needed to get Genomic Selection to work
 - What statistical models and breeding structures?
- Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK

- Domestic Animal Resources**
- Unique opportunities to decipher the molecular architecture of complex traits
 - The technology seems no more to be a limitation
 - High Throughput Sequencing
 - Expression profiling
 - Methylation studies
 - New dreams for the near future (whole genome for €500.-)
 - The quality of the data is an issue
 - Assembly, Annotation but also phenotypes !
- Animal Disease Genomics: Opportunities and Applications
 10th - 11th June 2008, Edinburgh, UK




Whole genome sequencing in livestock species : The end of the beginning ?
André Eggen, INRA, France



Acknowledgments

- **USDA, USA**
 - Curt van Tassel, Tim Smith, Tad Sonstegard
- **University of Missouri, USA**
 - Robert Schnabel, Jerry Taylor
- **Members of the iBMC consortium**
- **CEA-CNG, Evry, France**
 - Ivo Gut and his team
- **INRA**
 - Alain Vignal, Denis Milan, Gérard Guérin, Simon de Givry, Thomas Scheix, Thomas Faraut, Bertrand Servin
- **UNCEIA**
 - Alain Malafosse, Sébastien Fritz
- **LABOGENA**
 - Yves Amigues, Marie-Yvonne Boscher and others
- **INRA – Jouy en Josas**
 - Didier Boichard
- **Bovine Genomics' Team**
 - Mathieu Gautier, Mekki Boussaha
 - and others ...

Financial support :
ANR GENANIMAL, EADGENE, INRA, UNCEIA



Animal Disease Genomics: Opportunities and Applications
10th - 11th June 2008, Edinburgh, UK

