

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

High throughput SNP identification and genotyping
Martien Groenen

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.

ANIMAL SCIENCES GROUP WAGENINGEN UR

EADGENE Introduction

Overview

- High throughput SNP identification using 2nd generation sequencing
 - (1) Species whose genome has been sequenced (reference genome)
 - (2) Species without a genome sequence
- Applications of high throughput SNP typing
- Future applications

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

ANIMAL SCIENCES GROUP WAGENINGEN UR

EADGENE Introduction

Solexa sequencing (Illumina GA)

- 1 flow cell, 8 lanes
- Read length 36 bp
- 3-8 million reads per lane
- 1 sequence run = 2 days
- Error rate of ~ 1%

T G C T A C G A T ...

TTTTTTTGT ...

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

ANIMAL SCIENCES GROUP WAGENINGEN UR

EADGENE Introduction

Solexa sequencing (Illumina GA)

- (1) Species whose genome has been sequenced (reference genome)
 - Example: Pig (Swine SNP chip consortium)
- (2) Species without a genome sequence
 - Example: Turkey

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

ANIMAL SCIENCES GROUP WAGENINGEN UR

EADGENE SNP discovery with reference genome Example: pig

Overview of the pig SNP project

~ 25 animals/breed

150-200 bp

6x 6x 6x 6x 6x → Estimate of MAF within breeds

30x → Highly reliable SNP identification

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

ANIMAL SCIENCES GROUP WAGENINGEN UR

EADGENE SNP discovery with reference genome Example: pig

Preliminary data of HaeIII library

- DNA pools digested with 3 restriction enzymes: *HaeIII*, *MspI* and *AclI* (2x)
- Separate DNA fragments on polyacrylamide gels
- Sequence fragments in the 120-240 bp range, depending on the enzyme used

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

ANIMAL SCIENCES GROUP WAGENINGEN UR

High throughput SNP identification and genotyping

Martien Groenen, Wageningen University, The Netherlands

EADGENE SNP discovery without reference genome
Example: Turkey

30 % of the sequence in contigs > 50 bp

Contig size	Observed frequency	Uniquely mapping to chicken number	fraction (80 % sequence identity for > 80 % of the contig)
50-74	126953	62970	0.50
75-99	36810	21058	0.57
100-149	25163	19372	0.77
150-199	8734	6572	0.75
200-299	6379	4434	0.70
300-399	2349	1359	0.58
400-499	1189	551	0.46
500-599	677	259	0.38
>=600	1369	282	0.21

Overall 58 % of the sequences align uniquely to the chicken genome

Can be improved by combining with 454 or by using paired end

~ 20 Mb
~ 30 % of the sequence data

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

ANIMAL SCIENCES GROUP
WAGENINGEN

EADGENE SNP discovery without reference genome
Example: Turkey

Preliminary analysis using MAQ

- Criteria used for SNP identification
 - Minor allele frequency > 0.1
 - SNP identification tested with high quality sequences as well as with all sequences (depth of 58x)

SNP selection criteria	SNPs detected*
occurrence minor allele >= 8	16144
occurrence minor allele in high quality* >= 4	7656

* on assembled contigs representing 30 % of the genome


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

ANIMAL SCIENCES GROUP
WAGENINGEN

EADGENE High density SNP typing

Applications of high throughput SNP typing

- A few examples in chicken and pigs based on Illumina Golden Gate assays (1536 SNPs) and a chicken 18K iSelect chip




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

ANIMAL SCIENCES GROUP
WAGENINGEN

EADGENE High density SNP typing


LD, Genomic selection and gene identification

Extent of linkage disequilibrium (LD)

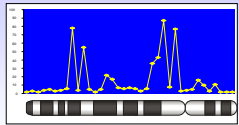


Amaral et al (2008) Genetics

Fine mapping of QTL



Genomic selection



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

ANIMAL SCIENCES GROUP
WAGENINGEN

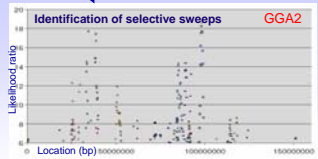
EADGENE High density SNP typing

Analysis of selective sweeps

Low frequency of ancestral allele

Decreased Heterozygosity

Identification of selective sweeps **GGA2**



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

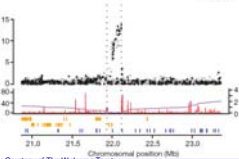
ANIMAL SCIENCES GROUP
WAGENINGEN

EADGENE High density SNP typing
Future


Whole genome association and LD studies

- Needed: High density SNP arrays (~100,000 - 500,000 SNPs)

Whole genome association studies



Extent of LD for the complete genome



Courtesy of The Wellcome Trust case control consortium, Nature (2007)

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

ANIMAL SCIENCES GROUP
WAGENINGEN

High throughput SNP identification and genotyping

Martien Groenen, Wageningen University, The Netherlands

High density SNP typing
Future

Analysis of Copy Number Variation (CNV)

- Needed: High density SNP arrays (~500,000 SNPs)

Comparative intensity analysis
Affymetrix 500K early access SNP chip

Combine samples

Redon et al (2006) Nature 444: 444

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

Future applications

Towards the \$1000 genome?

- Start thinking about (re)sequencing your favorite species – breed - individuals

SNP typing

Sequencing

↓

Data Handling

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

Summary

Ongoing SNP discovery projects

Your Favorite Species/breed?

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

Acknowledgements

Thank You

- Wageningen University
 - Richard Croijmans
 - Andreia Amaral
 - Marcos Ramos
 - Hinri Kerstens
 - Niki Winn van Bers
 - Pieter van As
 - Hendrik-Jan Megens
 - Martin Elferink
 - Bert Dibbits
 - Tineke Veenendaal
- Roslin Institute
 - Alan Archibald
 - Andy Law
- Sanger Institute
 - Carol Churcher
 - Carol Scott
- University of Aarhus
 - Christian Bendixen
 - Jakob Hedegaard
- ServiceXS, Leiden
 - Thomas Chi-A-Woeng
- Leiden Genome Techn Center
 - Johan den Dunnen
- NIOO
 - Kees van Oers
 - Marcel Visser
- Hendrix genetics
 - Gerard Albers
 - Addie Vereijken
 - Annemieke Rattink
- USDA – ARS, Beltsville
 - Curt van Tassel
- USDA – ARS, Clay Center
 - Gary Rohrer
 - Tim Smith
 - Dan Nonneman
- Iowa State University
 - Max Rothschild
 - James Reecey
- INRA
 - Denis Milan
- University of Missouri
 - Jerry Taylor
- University of Illinois
 - Larry Schook
 - John Beever

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