The Buffalo Expression Atlas

Rachel Young
PAG Conference
Jan 2017

Buffalo species

- Family Bovidae
- Sub-Family Bovinae
- Tribe Bovini

- Genus Bubalus
  - Bubalus arnee, Wild water buffalo
  - Bubalus bubalis, Domesticated water buffalo

- Genus Bos
  - Bos taurus, Domesticated cattle
  - Bos mutus, Wild yak
  - Bos indicus, Zebu

- Genus Pseudoryx
  - Pseudoryx nghetinhensis, Saola

- Genus Syncerus
  - Syncerus caffer, African buffalo

- Genus Bison
  - Bison bison, American bison
  - Bison bonasus, Wisent

Distribution of water buffalo population

Number of heads by country Average 2006 - 2014

http://faostat3.fao.org/browse/Q/QA/E

Increasing global buffalo population

Top 5 countries (number of heads) Average 2006 - 2014

http://faostat3.fao.org/browse/Q/QA/E

Significant Indian buffalo population

India is home to over half the world’s water buffalo population (54%, FAO 2014)
Buffalo production is a major source of employment in India
Mainstay of the dairy industry in India
Contributes 55% of milk production, from 35% of the total bovine population
Number of major buffalo breeds generated (12 registered breeds)

The Buffalo Atlas

A fine scale gene expression atlas generated by RNA-Seq from various tissues collected from three riverine buffalo breeds
Main aims of the atlas:
- define a comprehensive buffalo transcriptome based upon quantitative RNA profiling of tissues and cell types
- prevalence and significance of expressed genetic variation and allelic imbalance within and between buffalo breeds
- potential variation between breeds will form the basis for the development of predictive marker-assisted selection and breed improvement in the buffalo industry.
- provide an online resource for the ruminant genomics research community where free access to all data will be available together with the results of our analyses
Atlas outline

Eighty tissue and cell types collected per animal from eight adult buffalo (4 male, 4 female)

- Mediterranean, Padharpuri, Bhadawari breeds
- Animals aged between 6 months (Mediterranean) and 4-5 years (Indian)
- Main organ systems, immune tissues and GI tract

RNA-Seq based atlas of gene expression

- Core atlas
  - RNA-Seq of 164 tissues from 4 Mediterranean buffalo
- Supplemented with Indian buffalo samples
  - RNA-Seq of 56 tissues from 4 Indian buffalo

Sequencing and Analysis Approach

Stranded Illumina TruSeq libraries were prepared and sequenced by Edinburgh Genomics

125bp PE sequencing
- 68 samples sequenced at high depth (100M reads) Total RNA-Seq
- 152 samples sequenced at medium depth (25M reads) mRNA-Seq

Sequencing process complete

Analysis ongoing

Bioinformatic pipelines developed by Steve Bush
Kallisto, rapid alignment-free quantification of gene expression
HISAT2-Stringtie-Ballgown alignment-based pipeline (novel transcripts)
Differential expression visualised as a network graph in Miru

https://genomics.ed.ac.uk/  http://kaajek.com/miru

Visualisation of Atlas data

MIRU

inspecting the clusters

EMR1 cluster – macrophage-specific expression

Annotating the genome

Efforts to update the existing reference genome by J. Williams, Dovetail and colleagues

Plus the generation of a PacBio long-read assembly for the Mediterranean buffalo

Annotation of the reference genome with transcriptome data...

Lead to submission of the water buffalo genome to be hosted on Ensembl

Cap Analysis Gene Expression (CAGE)

CAGE analysis characterises transcriptional start sites within promoters to single-nucleotide resolution

allows the study of regulatory inputs driving gene expression
enables high-throughput identification of promoters and the construction of transcriptional networks
instrumental in globally mapping specific TSSs in eukaryotes and emphasised the existence of alternatively regulated TSSs, novel regulatory elements and allowed predictions of TF binding sites
CAGE analysis of buffalo macrophages

Various macrophage populations were isolated or generated from primary cells and CAGE libraries prepared (alveolar macrophages, bone marrow derived macrophages + LPS, monocyte derived macrophages)

Libraries were sequenced by Edinburgh Genomics (Illumina, 50M reads/sample)

Analysis of CAGE data ongoing

Allele Specific Expression

Mutations affecting expression of an allele on the same chromosome are cis expression quantitative trait loci (cis eQTL)

Individuals heterozygous for these mutations are expected to show allele specific expression or allelic imbalance between two alleles

This ASE can be detected by RNAseq if there is a heterozygous site in the coding sequence of the gene

ASE can also be caused by imprinting, when an epigenetic mark distinguishes the paternal or maternal chromosomes and causes them to be expressed at different levels

Allele Specific Expression in buffalo

Collaboration with INRA-Jouy as part of FAANG-Europe COST Action

Allele specific expression (ASE) analysis of water buffalo transcriptome data generated by Roslin Institute and analysed using INRA-Jouy ASE-pipeline

Analysis of ASE in datasets obtained from other ruminant species (sheep and goat) at Roslin Institute as part of FAANG initiative

Making the data available

Sample metadata

Sequence data

Gene expression data

...along with a link to the Miru network graph

Samples can be made available through

Resequencing of diverse Indian buffalo breeds

Collaboration with S Kumar, CCMB and S Gokhale, BAIF

Resequencing multiple representative individuals from diverse Indian buffalo breeds (SciGenom, India and Edinburgh Genomics, UK)

Already resequenced 36 buffalo from 6 Indian breeds, 10X coverage (Illumina)

A further 60 buffalo from the same 6 breeds will be resequenced by Edinburgh Genomics at 30X coverage (HiSeq X, Illumina)

Sequences currently undergoing assembly at Roslin by Steve Bush

Improve the assembly of the buffalo genome and to document the variome

Identify SNPs in Indian breeds to develop a new SNPchip encompassing data from Mediterranean and various Indian riverine buffalo breeds

Summary and conclusions

Generated an atlas of gene expression covering three riverine breeds of the domestic water buffalo

Describes the transcriptional landscape of the river buffalo

Will be used to annotate the existing reference genome, update gene models and provide an online resource for the ruminant genomics research community

Resequencing of diverse Indian buffalo breeds

Improve the assembly of the buffalo genome and to document the variome

Identify potential variation between breeds and build tools for predictive marker-assisted selection and breed improvement in the buffalo industry

Sample metadata

Sequence data

Gene expression data

...along with a link to the Miru network graph
Acknowledgements

Maclab
David Hume
Kim Summers
Lucas Lefèvre
Steve Bush
Praful Dutta
Zofia Lisowski
Lindsey Waddell
Kristin Sauter
Clare Pridans
Gemma Davis
Emily Clark
Anna Raper
Alita Carlisle
Tim Regan
Mark Barnett
Sara Clohisey
Alan Archibald
Steve Bishop
Dryden Farm
Tim King
Peter Tennant
Adrian Ritchie
Alison Mackenzie
Edinburgh Genomics
Mick Watson
Karim Gharbi
Richard Tabot
Helen Gunter
INRA Jouy
Dominique Rocha
Gabriel Guillocheau
CCMB
Satish Kumar
Akhay Joshi
Herojeet Singh
PTP Lodi
John Williams
Daniela Iamartino
Nadia Fiandese
Sara Botti
Funding from:
BAIF - Uruli Kanchan
Suresh Gokhale
RJ Santosh
Velu Dhanikachalam
Edinburgh Genomics
Mick Watson
Karim Gharbi
Richard Tabot
Helen Gunter
INRA Jouy
Dominique Rocha
Gabriel Guillocheau

Thank you for your attention

Unannotated genes

Gene names assigned to unknown genes (128 so far) by implementing an annotation framework

Names assigned after blastp of their longest peptide against NCBI nr

A – match to existing peptide

the blastp alignment must:

- score at p <= 1e-25
- % identity within the aligned region of >=90%
- alignment length >= 90% of the length of query protein
- alignment length >= 50 amino acids
- align with no gaps
- not be labelled low quality