Disentangling demography and selection effects of cattle domestication - new insights from the 1000 bull genomes project

Simon S. Boitard, Marlies Dolezal, Bertrand Servin, Daniel Fischer, Jared Decker, Iona Mac Leod, Qianqian Zhang, Bernt Guldbrandtsen, Mogens Lund, Alessandro Bagnato, et al.

To cite this version:

Simon S. Boitard, Marlies Dolezal, Bertrand Servin, Daniel Fischer, Jared Decker, et al.. Disentangling demography and selection effects of cattle domestication - new insights from the 1000 bull genomes project. Annual Meeting of the Society for Molecular Biology and Evolution (SMBE), Jul 2015, Vienne, Austria. hal-02796438

HAL Id: hal-02796438
https://hal.inrae.fr/hal-02796438
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Disentangling demography and selection effects of cattle domestication - new insights from the 1000 bull genomes project

Simon Boitard, Marlies Dolezal, Bertrand Servin, Daniel Fischer, Jared Decker, Iona Macleod, Qianqian Zhang, Bernt Guldbrandtsen, Mogens Lund, Alessandro Bagnato, Johanna Vilkki and the 1000 bull genomes project

Institut National de la Recherche Agronomique, GABI, Jouy-en-Josas, France
Ecole Pratique des Hautes Etudes, ISYEB, Paris, France
Taurine cattle history

- **Domestication** in Western Asia, \( \approx 10,000 \) years ago.
- Introduction in Europe through **two migration routes**.
- Creation of modern breeds, \( \approx 200 \) years ago.
- **Selection** for milk and meat production, fertility, stature, coat color, mild temper . . .

The 1000 bull genomes project

- 1,147 bulls
- 41 breeds, mainly industrial and from North-west Europe.
- Sequencing at 10.6 fold coverage on average.
- 33.6 million variants.
- Inference of **demography and selection**:
  - 390 animals from 16 breeds
  - Close relatives and recently admixed individuals filtered out.
Outline

1. Evolution of effective population size through time?

2. Genomic regions under selection?

3. Conclusions and perspectives
1 Evolution of effective population size through time?

2 Genomic regions under selection?

3 Conclusions and perspectives
Simulation results

- Expected demographic scenario in cattle (MacLeod et al., 2013).
- Applied PSMC (Li and Durbin, 2011) and MSMC (Schiffels and Durbin, 2014) to genomic samples simulated under this scenario.
Simulation results

- Expected demographic scenario in cattle (MacLeod et al, 2013).
- Applied PSMC (Li and Durbin, 2011) and MSMC (Schiffels and Durbin, 2014) to genomic samples simulated under this scenario.
Simulation results

- Expected demographic scenario in cattle (MacLeod et al, 2013).
- Applied PSMC (Li and Durbin, 2011) and MSMC (Schiffels and Durbin, 2014) to genomic samples simulated under this scenario.
Simulation results

- Expected demographic scenario in cattle (MacLeod et al, 2013).
- Applied PSMC (Li and Durbin, 2011) and MSMC (Schiffels and Durbin, 2014) to genomic samples simulated under this scenario.
An approximate Bayesian computation (ABC) approach

- Observed sample of $n$ genomes summarized using:
  - the **allele frequency spectrum (AFS)**
  - the average **linkage disequilibrium (LD)** for different bins of physical distance between SNPs.

- Genomic samples of $n$ genomes simulated under random population size histories, with a prior distribution.

- Histories leading to **similar statistics as the observed sample accepted** and used to build the posterior distribution.

Manuscript under revision with co-authors Willy Rodriguez-Valcarce, Flora Jay, Stefano Mona and Frédéric Austerlitz.
Simulation results - ABC

- Population size history reconstructed up to recent past (large sample size).
- Unaffected by phasing errors (unphased genotype data).
- Robust to sequencing errors if rare alleles removed.
Histories inferred in all breeds **diverge from domestication**.

Continuous population size decline **starting before domestication**.

Ranking of recent population sizes consistent with expectations.
Differences between breeds: 3 geographical groups

- BrownSwiss
- Fleckvieh
- Holstein
- SwedishRed
- Charolais
- Limousin
- Hereford
- Jersey

- "Danubian" origin.
- "Mediterranean" origin.
- More complex origin.
Differences between breeds: 3 geographical groups

- "Danubian" origin.
- "Mediterranean" origin.
- More complex origin.
Differences between breeds: 3 geographical groups

- "Danubian" origin.
- "Mediterranean" origin.
- More complex origin.
Differences between breeds: 3 geographical groups

- Brown Swiss
- Fleckvieh
- Holstein
- Swedish Red
- Charolais
- Limousin
- Hereford
- Jersey

- "Danubian" origin.
- "Mediterranean" origin.
- More complex origin.
Differences between breeds: 3 geographical groups

- “Danubian” origin.
- “Mediterranean” origin.
- More complex origin.
1 Evolution of effective population size through time?

2 Genomic regions under selection?

3 Conclusions and perspectives
Within each breed:

- Look for **hard sweep signatures** (regions with excess of rare alleles) using the CLR statistic (Nielsen et al, 2005).

- Estimate the **neutral distribution of CLR** by simulating genomic samples from the **posterior distribution of population size history** provided by ABC.
Overview of selection signatures

- **Significant hard sweeps** (genome-wide type I error 5%) for 9 breeds with more than 20 sampled animals:
  - 7.5% of the genome in total, 32% of annotated genes.
  - 46% of theses regions detected in several breeds.
  - 2 sweeps detected in 8 out of 9 breeds, around PLAG1 (stature) and MEF2A (muscle growth factor).

- Within most significant regions, many genes related to:
  - **coat color**: PMEL, KIT, KITLG, TYRP1, MC1R, MITF
  - **growth / stature**: MSTN, TRAM2, NCAPG-LCORL, GRID2, ASAP1
MEF2A region

- **Holstein**

  - Heterozygosity

- **Limousin**

- **Hereford**
Outline

1. Evolution of effective population size through time?

2. Genomic regions under selection?

3. Conclusions and perspectives
Conclusions

- 1000 bull genomes project: first dataset including large samples of entire genomes.
- Promise for re-discovery of domestic cattle history.
- Continuous decline of effective population size since domestication.
- Different dynamics between geographical groups.
- Significant hard sweep signatures, often related to breed standards.
Account for population structure, migration and . . .

. . . violations of the Kingman coalescent?

- Large skew of reproductive success induced by modern breeding
  → excess of genetic diversity (Der and Plotkin, 2014).
- Recent population sizes estimated by ABC larger than expected from pedigree data.

Sweep signatures related to production traits: focus on standing variation or polygenic selection.
Co-authors:

- **Marlies Dolezal**, Vetmed Uni, Vienna, Austria
- **Bertrand Servin**, INRA, GenPhySE, Toulouse, France
- **Daniel Fischer & Johanna Vilkki**, MTT AgriFood Research, Jokioinen, Finland
- **Jared Decker**, University of Missouri, Columbia, USA
- **Iona Macleod**, University of Melbourne, Melbourne, Australia
- **Qianqian Zhang & Bernt Guldbrandtsen & Mogens Lund**, Aarhus University, Aarhus, Denmark
- **Alessandro Bagnato**, UNIM, Medicina Veterinaria, Milano, Italy

1000 bull genomes collaborators:

- **Ben Hayes**, DPI, Biosciences Research Division, Bundoora, Australia
- **Aurélien Capitan & Didier Boichard**, INRA, GABI, Jouy-en-Josas, France
- www.1000bullgenomes.org

ABC collaborators:

- **Willy Rodriguez**, Université de Toulouse, IMT, Toulouse, France
- **Stefano Mona**, EPHE, ISYEB, Paris, France

Computing Facilities:

- Genotoul bioinformatics platform Toulouse Midi-Pyrénées.