

Chillingham: inbreeding and heterozygosity patterns

Inbreeding and Purging at the Genomic Level: the Chillingham Cattle Reveal Extensive, Non-Random SNP Heterozygosity

John L. Williams^{1,4}, Stephen J.G. Hall^{2,5}, Marcello Del Corvo¹, Keith T. Ballingall³, L. Colli⁶, P. Ajmone Marsan⁶ and Filippo Biscarini^{1*}

¹ Parco Tecnologico Padano (PTP), Via Einstein - Loc. Cascina Codazza, 26900, Lodi, Italy

² School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK

³ Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, Midlothian, EH26 0PZ, UK

⁴ Present address: School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA 5371, Australia

⁵ Present address: Livestock Diversity Ltd., 3 Cross O'Cliff Hill, Lincoln LN5 8PN, UK.

⁶ Università Cattolica del Sacro Cuore, via Emilia Parmense 84, 29122, Piacenza, Italy

* Corresponding author: filippo.biscarini@tecnoparco.org (+39 340 7499754)

Keywords: Cattle, inbreeding, linkage disequilibrium, lethal recessive haplotypes, livestock conservation, HD SNP-chip, animal genetic resources

Inbreeding and Purging at the Genomic Level: the Chillingham Cattle Reveal Extensive, Non-Random SNP Heterozygosity

Summary

Local breeds of livestock are of conservation significance as components of global biodiversity and as reservoirs of genetic variation relevant to the future sustainability of agriculture. One such rare historic breed, the Chillingham cattle of northern England, have a 350 year history of isolation and inbreeding, yet show no diminution of viability or fertility. The Chillingham cattle have not been subjected to selective breeding. It has been suggested previously that the herd has minimal genetic variation. In the present study, high density SNP genotyping with the 777K SNP chip showed that 9.1% loci on the chip are polymorphic in the herd, compared with 62-90% seen in commercial cattle breeds. Instead of being homogeneously distributed along the genome, these loci are clustered at specific chromosomal locations. A high proportion of the Chillingham individuals examined were heterozygous at many of these polymorphic loci, suggesting that some loci are under balancing selection. Some of these frequently heterozygous loci have been implicated as sites of recessive lethal mutations in cattle. Linkage disequilibrium equal or close to 100% was found to span up to 1350 kbps, and LD was above $r^2=0.25$ up to more than 5000 kbps. This strong LD is consistent with the lack of polymorphic loci in the herd. The heterozygous regions in the Chillingham cattle may be the locations of genes relevant to fitness or survival which may help elucidate the biology of local adaptation in traditional breeds, and facilitate selection for such traits in commercial cattle.

Introduction

Management of inbreeding is fundamental to the conservation of genetic resources. DNA techniques have enabled the degree of inbreeding to be inferred, even where pedigree information is not available, in a wide range of animal species including endangered species, so-called “model species”, and commercial livestock. Examples include Attwater’s prairie chicken (Hammerly et al. 2013), the zebra finch (Santure et al. 2010), and commercial Kenyan and Norwegian dairy cattle (Gorbach et al. 2010; Hillestad et al. 2015). The genetic architecture of inbreeding depression and of the purging of disadvantageous alleles has also been investigated using genetic markers, principally in livestock species, and especially in modern commercial cattle breeds (e.g. Purfield et al. 2013). These breeds were initially created about 200 years and a relatively small number of generations ago, from closed populations (Feliuss et al. 2014). More recently they have been intensively selected for production characteristics, further restricting the effective population size (Taberlet et al. 2008).

New genomic techniques that are being used to investigate and manage the more inbred commercial populations may contribute to the conservation genetics of endangered wild species (Kristensen & Sørensen 2005). The use of these techniques will also provide a better understanding of the genetic architecture of locally adapted livestock breeds which have been less intensely selected, but which have suffered inbreeding. These breeds are genetic resources in their own right, fundamental to human wellbeing. They are acknowledged under the 1992 Convention on Biodiversity as components of global biodiversity (Hoffmann 2011), which signatory countries have undertaken to conserve. Studies of such breeds may also elucidate the functional architecture of the genome and this may prove to be of particular relevance to conservation biology in general, as these breeds have been thought for many years to exhibit signatures of natural selection.

In some local populations there has been no management of inbreeding, and here we study a particularly rare cattle breed where this has been the case: the Chillingham white cattle of northern

England. These cattle have not been subjected to selection for commercial traits, and have remained a closed herd for at least 350 years. Herd records indicate no diminution of fertility or viability since records began to be kept in the mid-19th century (Hall & Hall 1988; Hall et al. 2005). The herd currently numbers approximately 50 males and 50 females. Inbreeding for 67 generations and an effective population size (N_e) of 8.0 was inferred by Visscher et al. (2001).

Under favourable conditions, females begin to show ovarian activity at 10 months of age and bulls have a mature weight of up to 400 kg (Hall, unpublished). The only recorded congenital abnormalities are in the molar arcade (Ingham 2002), and there is a relatively high incidence of testicular hypoplasia (Hall et al. 2005); the occurrence of the latter has been correlated with increased inbreeding in commercial cattle (Čítek et al. 2009). In the light of an earlier demonstration of complete homozygosity at 24 out of 25 microsatellite loci distributed across the genome (Visscher et al. 2001), it was proposed that the Chillingham cattle have been purged of lethal recessive mutations. However, it is now evident (Curik et al. 2014) that 25 microsatellite loci are too few to permit inference of overall lack of genomic variation. The aim of the present study was, therefore, to confirm, by high density SNP genotyping, the apparent lack of variation, to gain further insights into properties of the genome of Chillingham cattle, and to extend knowledge of the genomic architecture of inbreeding as it may be applied to conservation management.

Materials and methods

Animals and genotypes

Biological samples, blood or tissue, were collected from 20 Chillingham animals (seven males, 13 females). As no pedigree records exist for the Chillingham herd and organized sample collection is not possible, samples were obtained from animals that died between 2005 and 2011. Two tissue samples were rejected as being of too poor quality, and DNA was extracted from the remaining 18 samples using the Qiagen DNeasy kit according to the manufacturer's instructions. These samples were genotyped with the BovineHD SNP Chip (777,962 SNPs; Illumina, San Diego,

http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf).

Data were edited for call-rate only. Two samples and 2172 SNPs were discarded because of low call-rate (< 95%). Unmapped SNPs (3302) were not used. In total, 16 animals and 772,488 SNPs were retained for the analysis.

Estimation of genetic parameters

The following genetic parameters were estimated in the Chillingham population: i) the fraction of polymorphic (and, conversely, monomorphic) loci over the total number of loci (polymorphic loci are defined as loci showing multiple alleles segregating in the population); ii) observed homozygosity (HOM_{OBS}); iii) observed heterozygosity (HET_{OBS}); iv) the inbreeding coefficient

$F_{IS} = 1 - \left(\frac{HET_{OBS}}{HET_{EXP}} \right)$ (Wright, 1949), where HET_{EXP} is the expected heterozygosity based on allele frequencies ($2pq$); and v) linkage disequilibrium (LD). Observed homozygosity and heterozygosity were calculated as fractions of, respectively, homozygous and heterozygous animals at each locus over the sample size. LD between each pair of SNPs was measured as r^2 :

$$r^2(p_A, p_B, p_{AB}) = \frac{D^2(p_{AB} - p_A p_B)^2}{p_A(1 - p_A)p_B(1 - p_B)} \quad (1)$$

where p_A , p_B and p_{AB} are the frequencies of the AB haplotype and of alleles A and B in the first and second locus respectively, and $D^2 = (p_{AB} - p_A p_B)^2$ is the squared difference between observed and expected haplotype frequency (Van Liere & Rosenberg 2008). The r^2 values were grouped into bins based on the base-pair distance between SNPs from the physical map. The average per-bin LD as a function of the base-pair distance was then used to estimate LD decay in the Chillingham cattle.

Runs of homozygosity and heterozygosity

Runs of homozygosity (ROHom) were defined as consecutive stretches of homozygous SNP genotypes. A maximum of 2 missing genotypes and 1 heterozygous SNP were allowed in a ROHom. Similarly, runs of heterozygosity (ROHet) were defined as consecutive stretches of heterozygous SNP genotypes. A maximum of 2 missing genotypes and 4 homozygous genotypes were allowed in a ROHet. These criteria were based on previous works on ROHom in cattle: Purfield et al. (2012) allowed 2 missing and 1 heterozygous SNP in a ROHom using the 54K SNP chip; with the HD SNP chip, Ferenčaković et al. (2013) were more lenient, and allowed up to 64 missing and 16 heterozygous SNP in a ROHom.

Autocorrelations

To test whether the distribution of observed heterozygosity in the Chillingham genome is significantly non-random (i.e. clustered into "runs of heterozygosity"), the autocorrelations between adjacent SNP loci were compared against autocorrelations from a random data series ("white noise series"; Hyndman & Athanasopoulos, 2014). Autocorrelations are widely used in time series (longitudinal) data to measure the strength of the relationship between consecutive values. The same line of reasoning was applied to the series of SNP loci physically aligned along the Chillingham genome. Autocorrelations were computed for increasing distances between SNP loci.

For a white noise series, 95% of autocorrelations are expected to fall within $\pm \frac{2}{\sqrt{T}}$ where T is the length of the series. Therefore, if > 5% of autocorrelations are outside these boundaries, the distribution can be said to be significantly non-random. Chromosome- and genome-wide p-values

were obtained from z-deviates $\left(z = \frac{(\text{freq} > 0.95) - 0.05}{\sqrt{\frac{0.05 * 0.95}{T}}} \right)$ tested against a standard normal distribution under $H_0: P = 0.05$ (with $H_1: P > 0.05$); freq is the per-chromosome or genome-wide proportion of autocorrelations $> / < \pm \frac{2}{\sqrt{T}}$.

Software

Data handling and editing, estimation of HOM_{OBS} , HET_{OBS} and F_{IS} , identification of ROH_{om} and ROH_{het} , calculation of autocorrelations and plots were produced with the open source programming environment for statistical analysis R (R Development Core Team, 2008). LD was calculated using PLINK v1.07 (Purcell et al., 2007).

Results

After data editing, 772,488 SNP genotypes for 16 animals were available, with an average MAF of 0.023 (± 0.079). Of these 772,488 SNPs, only 70,150 (9.1%) were polymorphic in the Chillingham samples, which implies 90.9% of the SNP were monomorphic: the greatest number of monomorphic loci was on chromosome 28 (98.32%) and the lowest on chromosome 12 (81.95%) (Table 1). The overall observed heterozygosity (number of heterozygous genotypes over $16 \times 772,488 = 12,359,808$ total genotypes) was 0.026 (2.6%), which translates to 0.974 (97.4%) homozygosity. The inbreeding coefficient was estimated as $F_{IS} = 0.92$ (± 0.28). Nevertheless, there is still a significant amount of residual polymorphic loci which are frequently seen in the heterozygous state in the Chillingham cattle population. These loci appear to be distributed non-randomly and are clustered in blocks, or “runs of heterozygosity” (BTA1 and BTA7 in Figure 1, as illustration; all chromosomes in Supplementary Figure 1). The autocorrelation test confirmed that this SNP

heterozygosity is not distributed randomly (both chromosome- and genome-wide p-values \ll 0.00001) but is clustered in specific regions. Heterogeneity in recombination rates along the genome was not considered; however, it is not likely to affect results too severely, given the clear observed clusters of heterozygosity and the very low p-value obtained from the test. On average there were 9.5 runs of heterozygosity longer than 150Kb per chromosome. Chromosome 7 had 25 runs of heterozygosity longer than 150Mb and 21% of the chromosome in runs of heterozygosity, while chromosome 28, which had the lowest heterozygosity, had 2 runs of heterozygosity between 5-10Kb. The longest uninterrupted run of heterozygous SNP loci was on chromosome 7, covering 371 Mb (Supplementary Table 1 and Supplementary Figure 2). Complete LD was estimated to span up to 350 Kbp in the Chillingham genome, and LD close to 100% spanning up to 1350 kbps was detected. LD above $r^2=0.25$ spans > 5000 kbps (Figure 2).

Discussion

Among the 16 Chillingham cattle 9.1% of the SNP loci on the HD SNP panel exhibit allelic polymorphism, compared with the estimated 62-90% polymorphic loci found in fourteen taurine and indicine breeds (Pertoldi et al. 2014; Porto-Neto et al. 2014; Ramey et al. 2013); these studies had sample sizes ranging from 39 to 262 individuals, and applied no filter on low MAF SNP (except Pertoldi et al. who removed SNP that were monomorphic across breeds); the average MAF was between 0.194 and 0.246 (compared to 0.023 in the Chillingham), and expected heterozygosity was estimated in the range 0.25 – 0.34 (0.026 in the Chillingham). The true level of polymorphism in the Chillingham cattle could indeed be somewhat higher, as there may be ascertainment bias of SNP included in the panel. It should also be noted that the SNP on the panel were selected from data on a limited number of breeds (Matukumalli et al, 2009), therefore Chillingham cattle may be polymorphic at loci not included in the SNP panel. In addition, rarer alleles segregating in the Chillingham herd may not have been detected in the small sample size (16 individuals). Nevertheless, these data indicate a considerably lower level of variation in the Chillingham than

other cattle breeds. Calculations based on the assumptions of Visscher et al. (2001), which were an ancestral heterozygosity of 0.70 and 67 generations of inbreeding with an effective population size of 8.0, predicted that 19,000 (2.4%) loci would be polymorphic in Chillingham cattle, which given the uncertainty in the estimates is not much dissimilar from the observed 9.1%. The residual polymorphic loci are not homogeneously distributed across the Chillingham genome, but are clustered into runs of heterozygosity, suggesting that some loci are under balancing selection. Interestingly, these patterns of clustered heterozygosity are not observed in other cattle breeds: Figure 3 shows the distribution of polymorphic loci from the HD Bovine SNP chip on BTA1 for the Chillingham and three other cattle breeds (one dairy, two beef white breeds): Holstein Friesian, Chianina, and Romagnola.

SNP genotypes were also used to assess the relatedness among individual animals, estimating the genomic relationships as in Van Raden (2008). Apart from animals CHIL_11 and CHIL_15 there were no particularly related animals (see the heatmap in Figure 4) and no strong family structure emerged in the sampled Chillingham population. Based on 54K SNP genotype data, a recent paper on the diversity of 56 cattle breeds (Orozco-ter Wengel et al., 2015) showed that the Chillingham was distant from other breeds, as may be expected from its unusual genome structure. This was confirmed by our results with the HD SNP panel: Figure 5 shows the phylogenetic tree for Chillingham, Holstein-Friesian, Romagnola and Chianina cattle based on the maximum likelihood graph model implemented in the software *TreeMix* (Pickrell & Pritchard, 2012).

Overall, the SNP data confirmed the general lack of variation implied by the microsatellite study of Visscher et al. (2001), with 90.9% monomorphic loci, 97.4% homozygosity, runs of homozygosity (ROHom) covering 95% of the genome and 14 ROHom longer than 30MB. This contrasts with the proportion of the genome found in monomorphic runs in a study of 42 cattle breeds (Purfield et al. 2012) where on average 198Mb (7.5%) of the genome was found in ROHom, with the greatest proportions found on chromosomes 14 and 16 (13.71% and 14.16% respectively).

The chromosomal regions where polymorphic SNP were frequently found in the heterozygous state

in the Chillingham genome (runs of heterozygosity, ROHet) may harbour loci associated with fitness and where diversity is an advantage (Curik et al. 2014). In dairy cattle, 25 haplotypes which were not observed in the homozygous state, were identified at specific chromosomal positions in particular breeds (Van Raden et al. 2011; Fritz et al. 2013; Sahana et al. 2014; Cole et al., 2015). It is argued that these positions harbour genes with recessive lethal mutations. Of these 25 haplotypes, 18 fall within ROHet detected in the Chillingham genome. These regions are listed in Table 2, with the average heterozygosity and SNP polymorphism for each region as it occurs in the Chillingham genome. For 17 and 18 out of the 25 regions, the percentage of polymorphic loci and the average heterozygosity, respectively, were greater than the overall polymorphism and heterozygosity of the respective Chillingham chromosome. These chromosomal regions are of functional significance in dairy cattle, and are likely to contain genes involved in embryonic survival. A functional mutation in the CWC15 gene (Sonstegard et al. 2013) which is involved in placental attachment has been identified at one of these regions on BTA15 in the Jersey breed. These apparently lethal mutations are breed specific and presumably arose after the formation of the breeds, and therefore are not likely to be present in Chillingham cattle. However, maintenance of haplotypic diversity at such loci may be an advantage for fitness and could be subject to balancing or countervailing selection.

The biallelic nature of SNPs facilitates the investigation of linkage disequilibrium (Slatkin 2008). LD is conserved over an extremely long range in the Chillingham cattle which contrasts with other cattle breeds, where high values of LD are typically detected only over much shorter distances (< 10 kbps) and LD drops below $r^2=0.20$ from ~ 200 kbps (Gautier et al., 2007; Porto-Neto et al. 2014). However, the strong LD observed in the Chillingham population is consistent with the high level of monomorphic loci, which means recombination events may go undetected. In this respect, the general observation is that domesticated animals have more chiasmata than wild animals of the same life span (Burt & Bell 1987). High recombination rates may be selected under regimes of strong directional selection for performance (Notter 1999) or by genetic drift in small populations (Otto & Barton 2001). The Chillingham herd is not under selection for productivity (Hall & Hall

1988; Hall et al. 2005) which makes it similar to a wild population. However, given the high homozygosity, it is likely that recombination events are undetected rather than less frequent.

Classical genetics predicts (Falconer & Mackay 1996) that after 50 generations of sib-mating, under neutrality, only about 5% of the genome would be expected to remain polymorphic. That 9.1% of the SNP on the HD chip are polymorphic in the Chillingham indicates that loss of variation may have been retarded by balancing or countervailing selection. The persistence of blocks of loci where animals show a large number of heterozygous genotypes in the highly inbred Chillingham cattle suggests that there are regions of the bovine genome that contain genes with a major effect on fitness, which are resilient to genetic drift in an unselected population.

Conservation of genetic variation within commercial breeds is of key importance for the sustainability of livestock farming, and the runs of heterozygosity identified may contain loci that contribute to fitness or survival traits (McParland et al. 2009). Databases exist for production traits in commercial livestock, while data on fitness and survival traits are sparse. It has been argued that the goal of selective breeding should be to fix the most beneficial alleles for the traits under selection (Cole and VanRaden, 2011). However, as the effects of the alleles, or genes linked to the optimal alleles for the traits under selection may be detrimental to fitness in general, it would be prudent to maintain diversity at all or most loci to avoid affecting fitness. Thus knowledge of genes contributing to overall fitness is of importance to agriculture and conservation biology. Further study of these regions where there is high heterozygosity in relatively unselected, locally adapted breeds, may facilitate the identification of the genes that are under natural selection to maintain fitness.

Acknowledgments

We thank the Chillingham Wild Cattle Association for the provision of tissue samples. ANAFI (Associazione Nazionale Allevatori Frisone Italiana) contributed HD SNP data for the Holstein

Frisian breed, within the research project INNOVAGEN funded by the Italian Ministry of Agriculture (MiPAAF). ANABIC (Associazione Nazionale Allevatori Bovini Italiani da Carne) contributed HD SNP data for the Chianina and Romagnola breeds, within the research project PARAMOGGQUAL, funded by CCBI (Consorzio Produttori Carne Bovina Pregiata delle Razze Italiane). This research was supported by the “GenHome” project from the Italian Ministry of Science (MIUR). FB acknowledges support from the Marie-Curie European Reintegration Grant NEUTRADAPT (PERG08-GA-2010-276699) of the European Union.

Conflict of interests

The authors declare that they have no conflict of interest.

Supporting material

Supplementary Table 1: summary of the distribution and total chromosomal coverage for runs of homozygosity and runs of heterozygosity in the Chillingham cattle genome.

Supplementary Figure 1: Graphical representation of the frequency of runs of homozygosity and runs of heterozygosity by size class.

Supplementary Figure 2: Distribution of polymorphic SNP loci across all chromosomes for the 16 Chillingham individuals passing quality controls.

References

- Burt A. & Bell G. (1987) Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326**, 803-805.
- Čítek J., Rehout V., & Hájková J. (2009) Congenital disorders in the cattle population of the Czech Republic. *Czech Journal of Animal Science* **54**, 55-64.
- Cole, J. B., & VanRaden, P. M. (2011). Use of haplotypes to estimate Mendelian sampling effects and selection limits. *Journal of Animal Breeding and Genetics*, **128**(6), 446-455.
- Cole J.B., van Raden P.M., Null D.J., Hutchison J.L., Cooper T.A. & Hubbard S.M. (2015) Haplotype tests for recessive disorders that affect fertility and other traits http://www.aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html. USDA Animal Improvement Program Research Report Genomics 09-13.
- Curik I., Ferencakovic M. & Sölkner J. (2014) Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science* **166**, 26-34.
- Falconer D. S. & Mackay T.F.C. (1996) Introduction to quantitative genetics. Fourth edition. Longman, Harlow, UK.
- Felius M., Beerling M.L., Buchanan D.S., Theunissen B., Koolmees P.A. & Lenstra J.A. (2014) On the history of cattle genetic resources. *Diversity* **6**, 705-750.
- Ferenčaković M., Sölkner J. & Curik I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics, selection, evolution* **45**(1), 42
- Fritz S., Capitan A., Djari A., Rodriguez S.C., Barbat A., Baur A., Grohs C., Weiss B., Boussaha M., Esquerré D., Klopp C., Rocha D. & Boichard D. (2013) Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS One* **8**, e65550.
- Gautier M., Faraut T., Moazami-Goudarzi K., Navrati, V., Foglio M., Grohs C., Boland A., Garnier J.G., Boichard D., Lathrop G.M., Gut I.G. & Eggen A. (2007) Genetic and haplotypic

- structure in 14 European and African cattle breeds. *Genetics* **177**(2), 1059-1070.
- Gorbach D.M., Makgahlela M.L., Reecy J.M. Kemp S.J., Baltenweck I., Ouma R., Mwai O., Marshall K. Murdoch B., Moore S. & Rothschild M.F. (2010) Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *Journal of Animal Breeding and Genetics* **127**, 348-351.
- Hall S.J.G. & Hall J.G. (1988) Inbreeding and population dynamics of the Chillingham cattle (*Bos taurus*). *Journal of Zoology*, London **216**, 479-493.
- Hall S.J.G., Fletcher T.J., Gidlow J.R., Ingham B., Shepherd A., Smith A. & Widdows A. (2005) Management of the Chillingham wild cattle. *Government Veterinary Journal* **15**, 4-11.
- Hammerly S.C., Morrow M.E. & Johnson J.A. (2013) A comparison of pedigree- and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Molecular Ecology* **22**, 5313-5328.
- Hillestad B., Woolliams J.A., Meuwissen T., Våge D.I. & Klemetsdal G. (2014) Estimating rate of inbreeding and effective population size using genomic data in Norwegian Red cattle. *Proceedings 10th World Congress of Genetics Applied to Livestock Production*. Vancouver BC, Canada. https://www.asas.org/docs/default-source/wcgalp-posters/470_paper_8757_manuscript_659_0.pdf?sfvrsn=2
- Hoffmann I. (2011) Livestock biodiversity and sustainability. *Livestock Science* **139**, 69-79.
- Hyndman R. & Athanasopoulos G. (2014) Forecasting: principles and practice. *OTexts* (<https://www.otexts.org/fpp>).
- Ingham B. (2002) Dental anomalies in the Chillingham Wild White Cattle. *Transactions of the Natural History Society of Northumbria* **62**, 169-175.
- Kristensen T.N. & Sørensen A.C. (2005) Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science* **80**, 121-133.
- McParland S., Kearney F. & Berry D.P. (2009) Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genetics Selection Evolution* **41**, 16.

- Notter D.R. (1999) The importance of genetic diversity in livestock populations of the future. *Journal of Animal Science* **77**, 61-69.
- Orozco-ter Wengel, P., Barbato M., Nicolazzi E.L., Biscarini F., Milanesi M., Davies W., Williams D., Stella A., Ajmone-Marsan P. & Bruford M.W. (2015) Revisiting demographic processes in cattle with genome-wide population genetics analysis. *Frontiers in Genetics* **6**, 191.
- Otto S.P. & Barton N.H. (2001) Selection for recombination in small populations. *Evolution* **55**, 1921-1931.
- Pertoldi C., Purfield D.C., Berg P., Jensen T.H., Bach O.S., Vingborg R. & Kristensen T.N. (2014) Genetic characterization of a herd of the endangered Danish Jutland cattle. *Journal of Animal Science* **92**, 2372-2376.
- Pickrell, J. K., & Pritchard, J. K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**(11): e1002967
- Porto-Neto L.R., Kijas J.W. & Reverter A. (2014) The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution* **46**, 22.
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R, Bender D., Maller J., Sklar P., de Bakker P.I.W, Daly M.J. & Sham P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**.
- Purfield, D. C., D. P. Berry D. P., McParland S. & D. G. Bradley D. G. 2012. Runs of homozygosity and population history in cattle. *BMC Genetics* **13**:70.
- R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>. ISBN 3-900051-07-0
- Ramey H. R., J. E. Decker J. E., S. D. McKay S. D., M. M. Rolf M. M., R. D. Schnabel R. D. & J. F. Taylor J. F. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* **14**, 382.

- Sahana G., Nielsen U.S., Aamand G.P., Lund M.S. & Guldbbrandtsen B. (2014) Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLoS One* **8**, e82909.
- Santure A.W., Stapley J., Ball A.D., Birkhead T.R., Burke T. & Slate J. (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* **19**, 1439-1451.
- Slatkin M. (2008) Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477-485.
- Sonstegard T.S., Cole J.B., van Raden P.M., van Tassell C.P., Null D.J., Schroeder S.G., Bickhart D.M. & McClure M.C. (2013) Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLoS One* **8**, e54872.
- Taberlet P., Valentini A., Rezaei H.R., Naderi S., Pompanon F., Negrini R. & Ajmone-Marsan P. (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology* **17**, 275-284.
- VanLiere J.M. & Rosenberg N.A. (2008) Mathematical properties of the r^2 measure of linkage disequilibrium. *Theoretical Population Biology* **74**, 130-137.
- Van Raden P.M. (2008) Efficient methods to compute genomic predictions. *J. of Dairy Science* **91**(11), 4414–4423.
- van Raden P.M., Olson K.M., Null D.J. & Hutchison J.L. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science* **94**, 6153-6161.
- Visscher P.M., Smith D., Hall S.J.G. & Williams J.L. (2001) A viable herd of genetically uniform cattle. *Nature* **409**, 303.
- Wright S. (1949). The genetical structure of populations. *Annals of eugenics*, **15**(1), 323-354.

Tables

Table 1: number of monomorphic and polymorphic loci, corresponding proportions, and observed heterozygosity (H_{OBS}) in the Chillingham population. Overall and per-chromosome values.

BTA	N. monomorphic	N. polymorphic	% monomorphic	% polymorphic	HET_{OBS}
1	40854	5502	88.13%	11.87%	0.034
2	36685	3238	91.89%	8.11%	0.033
3	33077	2404	93.22%	6.78%	0.029
4	29871	5000	85.66%	14.34%	0.045
5	30990	3738	89.24%	10.76%	0.044
6	32063	3363	90.51%	9.49%	0.031
7	27426	5643	82.94%	17.06%	0.050
8	31518	1934	94.22%	5.78%	0.017
9	27623	3346	89.20%	10.80%	0.033
10	27987	2378	92.17%	7.83%	0.029
11	30488	1459	95.43%	4.57%	0.018
12	21320	4697	81.95%	18.05%	0.080
13	21117	2425	89.70%	10.30%	0.028
14	23680	1049	95.76%	4.24%	0.016
15	23275	1424	94.23%	5.77%	0.015
16	21222	2903	87.97%	12.03%	0.039
17	20654	1537	93.07%	6.93%	0.027
18	17294	2023	89.53%	10.47%	0.047
19	17508	1362	92.78%	7.22%	0.021
20	20253	1176	94.51%	5.49%	0.020
21	18516	2598	87.70%	12.30%	0.051
22	15851	2127	88.17%	11.83%	0.044
23	14000	1156	92.37%	7.63%	0.032
24	16316	2258	87.84%	12.16%	0.043
25	11503	1386	89.25%	10.75%	0.050
26	14581	619	95.93%	4.07%	0.017
27	12157	950	92.75%	7.25%	0.018
28	12785	218	98.32%	1.68%	0.006
29	13569	1106	92.46%	7.54%	0.028
X*	38155	1131	97.12%	2.88%	0.015
Total	702338	70150	90.92%	9.08%	0.026

*Based on females only

Table 2 Chromosomal regions implicated as harbouring detrimental haplotypes in dairy cattle and heterozygosity and polymorphism of each as they occur in the Chillingham genome

Haplotype *	Breed	BTA	Start Mb	End Mb	Mean het. of region in Chillingham	% polymorphic
HH4	Holstein	1	1.28		0.300	33.3%
	Brown					
BHP/JHP	Swiss	1	1.71	1.99	0.172	19.4%
HH4	Holstein	1	69.76		0.056	7.0%
HH2	Holstein	1	94.86	96.55	0.063	22.0%
HHB	Holstein	1	145.12		0.029	4.7%
62.7	Jersey	2	116.04	121.05	0.117	12.7%
HDR	Holstein	3	9.48		0.004	5.2%
HHC	Holstein	3	43.41		0.014	20.0%
	Brown					
BHW	Swiss	4	49.62	49.74	0.008	11.7%
HH1	Holstein	5	63.15		0.017	9.5%
175.5	Holstein	7	3.12	9.57	0.129	18.6%
	Brown					
BH1	Swiss	7	42.81	47.00	0.080	29.8%
186.139	Jersey	7	58.26	62.98	0.454	51.5%
HH3	Holstein	8	95.41		0.089	33.7%
HH5	Holstein	9	92.35	93.91	0.065	9.8%
	Brown					
BHD	Swiss	11	14.74		0.056	21.8%
JH1	Jersey	15	15.71		0.078	12.3%
369.1	Holstein	15	71.98	76.13	0.114	18.2%
HHM	Holstein	15	77.66	77.70	0.004	7.7%
AH1	Ayrshire	17	65.92		0.030	4.1%
HBR/HHR	Holstein	18	14.75	14.76	0.074	25.0%
	Brown					
BH2	Swiss	19	10.62	11.73	0.039	14.0%
HH0	Holstein	21	21.18	21.19	0.118	14.3%
	Brown					
BHM	Swiss	24	62.12	62.16	0.122	16.6%
JH2	Jersey	26	8.81	9.41	0.005	4.2%

*From Cole et al., 2015 (http://www.aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html)

Figures

Figure 1: distribution of heterozygous SNP on BTA1 and BTA7 of the Chillingham genome. The locations of the harmful haplotypes HH2 and BH1 are indicated.

Figure 2: distribution of pairwise LD (r^2) between SNP markers as a function of their distance. For each distance bin, the interquartile range of LD values is reported.

Figure 3: distribution of heterozygous SNP from the HD Bovine SNP chip on BTA1 and BTA7 in the Chillingham, Holstein Frisian, Chianina and Romagnola cattle breeds. Unlike in the Chillingham, polymorphic loci in the other breeds are clearly not clustered and rather distributed evenly along the chromosomes.

Figure 4: heatmap of genomic relationships between the 16 Chillingham animals used in the study. Genomic relationships were estimated following Van Raden (2008) and based on HD SNP chip data.

Figure 5: phylogenetic tree between four cattle breeds genotyped with the Bovine HD SNP chip: Chillingham, Holstein Frisian, Chianina and Romagnola.