Analysis of pedigree data and whole-genome sequences in 12 cattle breeds reveals extremely low within-breed Y-chromosome diversity

C. Escouflaire*,[†] [iD] and A. Capitan*,[†] [iD]

*ALLICE, Paris 75012, France. [†]Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas 78350, France.

Summary

In this article, we analyzed pedigree information on males from 12 bovine breeds born in France between 2015 and 2019. We report an overall small number of paternal lineages with, for example, a minimal number of ancestors accounting for 95% of the Y-chromosome pool of their breed ranging from only 2 to 15 individuals. Then, we mined whole-genome sequence data from 811 sires ($2 \leq n \leq 510$ per breed) and built a median-joining network using 1411 SNPs. Most branches were breed-specific and in agreement with the geographic and genetic relatedness of these populations. The within-breed haplotype diversity was lower than expected based on genealogical information, which supports the existence of major male founder effects predating pedigree recording. In addition, we observed *de novo* mutation events among the descendants of the same ancestors, which are of interest to define paternal sub-lineages. Our results pave the way to future studies on the estimation of the effects of Y-chromosome haplotypes on male reproductive performances and on the conservation of Y-chromosome diversity.

Keywords cattle, haplogroup, haplotype, paternal lineage, whole-genome sequence, Y1, Y2, Y-chromosome

In mammals, the Y chromosome is essential for male sex determination and fertility. In spite of its potential interest for breeding more fertile males, only a few research studies have focused on the genetic diversity and gene composition of the Y chromosome in cattle (Yue *et al.* 2015; Hughes *et al.* 2020). This lack of interest for the Y chromosome is due to the fact that breeding schemes have long focused on production traits (milk or meat) only, and massively used a few bulls with high genetic breeding value through artificial insemination (AI), without other consideration. In addition, the presence of a large number of highly repetitive and palindromic sequences, have posed a technological challenge in terms of sequencing and assembly of this chromosome's sequence. Nevertheless, the absence of chromosomal recombination outside the pseudo-autosomal region preserves the original haplotypes, and the transmission of the Y chromosome can therefore be followed by studying paternal lines in pedigrees. In this study, we investigated Y-chromosome diversity in 12 beef and dairy breeds bred in France (Table 1) using both pedigree and molecular marker information.

As a preliminary analysis, we mined the genealogies of males born between 2015 and 2019 using the Lineage program, one of the latest updates of the PEDIG package (Boichard 2002). We considered only the individuals with both parents known and, for each breed, we calculated the proportion of males born from AI versus natural mating, as well as the contributions to the present Y-chromosome pool of the most ancient paternal ancestors recorded within each pedigree.

Regardless of the breed, the number of male lineages was very small and the main ancestor (defined as the individual with the highest contribution to the Y pool of its breed among the most ancient paternal ancestors) accounted for 23.2–58% of these (MainA, Table 1). However, we observed significant differences between dairy versus dual-purpose and beef breeds for all the metrics calculated. Dairy breeds showed a higher use of AI and, as a consequence, a smaller number of ancestors contributing to more than 1% of the paternal lineages (NbA > 1), smaller minimal numbers of ancestors accounting for 95% of the paternal lineages (MinNbA95), and higher contributions from the main ancestors (MainA; Table 1, Table S1). This phenomenon is particularly significant in our dataset for Holstein cattle in which 98.3% of the Y-chromosomes of the investigated population derive from only two ancestors: Wis Burke Ideal (HOLUSAM000001013415 born in 1947 with a contribution of 47.5%) and Rosafe Pearl Hannibal,

Address for correspondence

A. Capitan, INRAE – Centre IdF Jouy-Antony, UMR1313 GABI – Equipe G2B, bât. 440, 4 rue Jean Jaurès, Jouy-en-Josas 78 350, France.
E-mail: aurelien.capitan@inrae.fr

Table 1 Indicators of paternal lineages diversity for males born between 2015 and 2019 using pedigree information.

Purpose	Breed	Number of males	AI	NbA > 1	MinNbA95	MainA
Dairy	Holstein	8 259 344	93%	3	2	50.8%
	Montbéliarde	1 866 789	91%	6	6	36.2%
	Normande	986 395	97%	4	4	52.3%
	Abondance	111 424	80%	5	5	44.3%
	Brown Swiss	75 638	95%	5	5	34.2%
	Tarentaise	39 927	83%	5	4	58.0%
Dual	Simmental	80 741	94%	12	11	38.5%
Beef	Charolaise	1 893 381	32%	9	11	23.2%
	Limousine	1 359 913	16%	12	15	39.5%
	Blonde d'Aq.	805 828	30%	11	15	36.1%
	Salers	199 433	13%	9	9	46.1%
	Rouge des Prés[1]	98 254	25%	10	9	28.6%
P-value dairy vs beef[2]			7.4×10^{-8}	1.3×10^{-5}	9.5×10^{-4}	3.8×10^{-2}

AI: proportion of males born from artificial insemination or multiple ovulation embryo transfer; NbA > 1: number of male ancestors contributing for more than 1% of the paternal lineages in the population analyzed; MinNbA95: minimal number of ancestors accounting for 95% of the paternal lineages; MainA: contribution of the main ancestor to the paternal lineages; Aq.: Aquitaine; Nb: number.
[1]The Rouge Des Prés breed was previously known as Maine Anjou.
[2]Two sample t-test with the alternative hypothesis that the means are not equal.

(HOLUSAM000001322381, 1955, 50.8%), which are the paternal grand sires of the influential AI bulls Elevation (HOLUSAM000001491007) and Chief (HOLUSAM000001427381), respectively. Yue *et al.* (2015) reported similar results in a study on North American Holsteins, and also traced these paternal lineages back to two bulls born in 1880.

To evaluate Y-chromosome diversity at the molecular level, we mined the vcf files from run 8 of the 1000 Bull Genomes Project, which were obtained after aligning whole-genome sequences of 4104 *Bos taurus* against the ARS-UCD1 genome assembly and a 43-Mb long draft sequence of chromosome Y (accession no. CM001061.2 in GenBank; http://www.ncbi.nlm.nih.gov/genbank/), as described in Bouwman *et al.* (2018). Due to the highly repetitive nature of the Y chromosome, we applied a series of filters to reduce the risk of considering false variants. We retained only SNPs: (i) that were present in two or more animals; (ii) that had a root mean square mapping quality higher than 40 and a FisherStrand value lower than 60 (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants); (iii) that were carried by less than 5% of the 479 purebred *Bos taurus* females available in run 8 of the 1000 Bull Genomes Project; and (iv) for which less than 5% of the heterozygous genotypes among males carried the alternative allele. In addition, we considered only the males from the breeds investigated and for which genotypes were available for at least 90% of the variants (with heterozygous genotypes considered as missing). We used Fimpute3 without pedigree information to impute missing genotypes (Sargolzaei *et al.* 2014), and POPART v1.7 (Leigh & Bryant 2015) to create a preliminary median-joining network. Then, we verified all individual genotypes that generated

web-like structures in the network and manually edited putative errors of variant calling or imputation according to the principle of parsimony (number of corrected genotypes n = 56/1 144 321). The final median-joining network was based on information from 811 sires (2 ≤ n ≤ 453 per breed; Tables S2 & S3) genotyped for 1411 SNPs, which to our knowledge represents the largest dataset analyzed to date in cattle in terms of individuals combined with number of markers. In total, 1266 SNPs (89.7%) supported the existence of the two main North-East (Y1) and South-West (Y2) European haplogroups that were initially described by Edwards *et al.* (2011) based on sequence information from three Y-chromosome genes and two microsatellite markers (Fig. 1). Among these 1266 SNPs, 66 were included in the panel of 68 markers used by Chen *et al.* (2018) to differentiate the Y1 and Y2 haplogroups. Within these haplogroups, most of the branches were breed-specific and reflected the geographic and genetic relatedness of these populations as previously evaluated using autosomal markers (Gautier *et al.* 2010; Park *et al.* 2015). For example, some Blonde d'Aquitaine bulls shared identical Y-haplotypes with Limousin and Salers bulls, and the same occurred between Montbéliarde and Simmental or between Abondance and Tarentaise. The within-breed haplotype diversity was lower than expected based on pedigree information with independent influential ancestors carrying the same Y-chromosome haplotype. For example, in Normande, we did not observe any variation between the Y-chromosomes of 30 bulls descending from five ancestors. These results support the existence of major male founder effects dating back to the creation of the breeds, and thus several decades before the advent of modern reproduction and selection techniques. The availability of 453 male whole genomes for the sole Holstein breed enabled us to
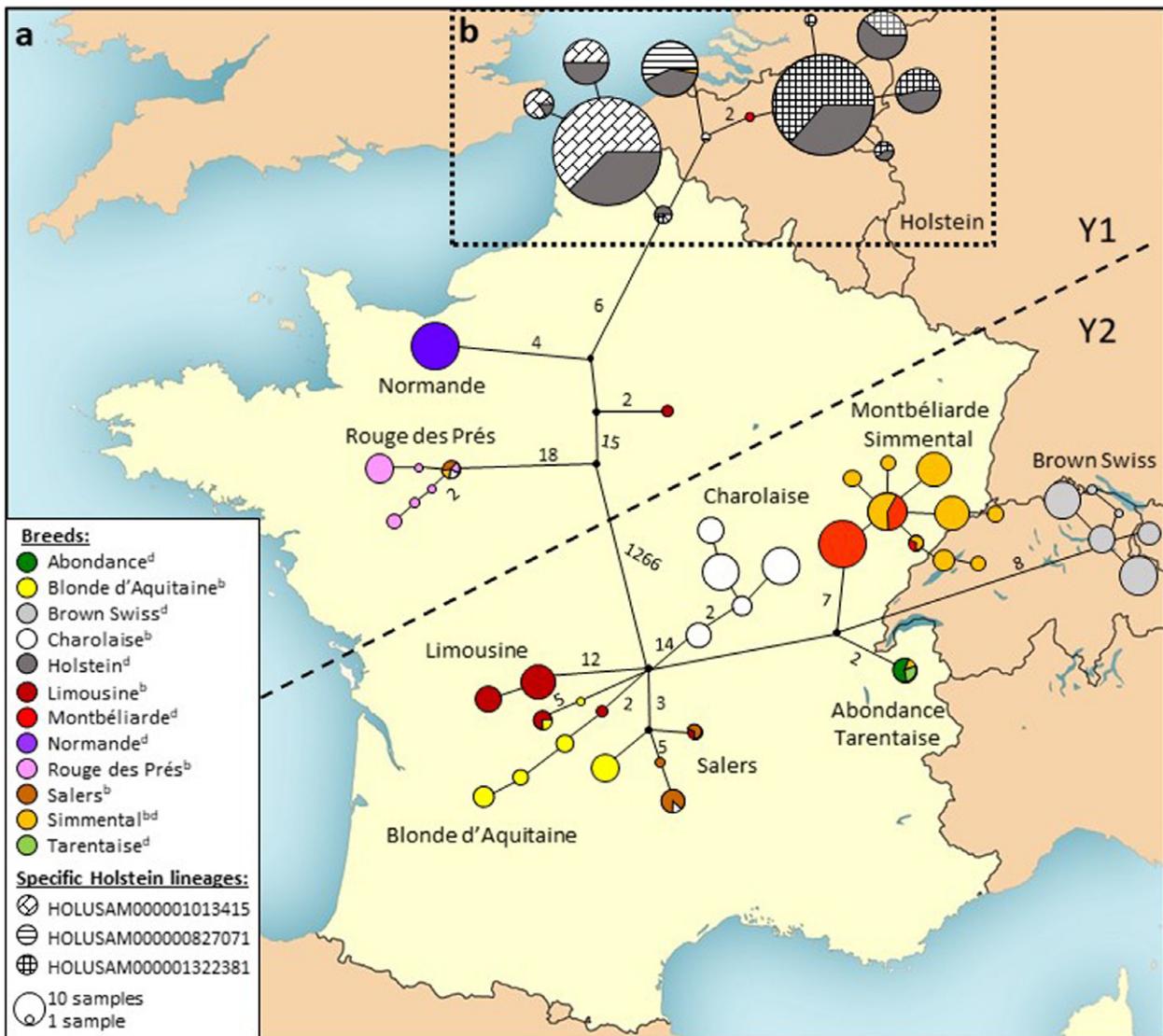
Figure 1 Median joining network based on 1411 Y-chromosome SNPs and 811 individuals. (a) Results for the whole data set. Each circle represents a haplotype and the size of the circle is proportional to the number of bulls bearing this haplotype. Branches connecting circles represent nucleotide changes between haplotypes. The number of variants is indicated on the branches when they are supported by more than one SNP. Colors refer to breeds of origin. The most represented haplotypes of each breed have been placed on a map on the geographic cradle of that breed. (b) Focus on Holstein derived Y1-haplotypes. Patterns refer to the most ancient male ancestor of carriers of each haplotype. Absence of pattern means that no pedigree information was available for the carrier of this specific haplotype in the French pedigree database. Annotations [b], [d], and [bd] refer respectively to beef, dairy, and dual-purpose breeds.

identify 41 carriers of a third Y-chromosome haplotype in Holstein, which originates from the bull HOLU-SAM000000827071 (born before 1950) that is the paternal great-grand sire of the influential AI bull Ivanhoe Star (HOLUSAM000001441440; Fig. 1). Nowadays, this rare haplotype account for only 1.2% of the Y paternal lines (See Suppl. Table S1). Moreover, our results show that only five of the 1411 SNPs differentiated the three basal Y-chromosome haplotypes found in the Holstein breed and we report six additional mutation events that occurred among the descendants of the two main paternal lines from

Rosafe Pearl Hannibal, (HOLUSAM000001322381, $n = 4$) and Wis Burke Ideal (HOLUSAM000001013415, $n = 2$; Fig. 1). Adding these variants to custom SNP chips (Boichard *et al.* 2018) would be very useful to distinguish sub-lineages and estimate the effects on male-related traits of other types of mutations that may have accumulated over time, and require long-read sequencing technologies or quantitative PCR to be detected. Previously published analyses demonstrated that: (i) the ampliconic region of the Y chromosome can experience substantial loss or gain of gene copies within a few generations (Yue *et al.* 2015); (ii)

some of these copy number variants are associated with scrotal circumference and male fertility in Holsteins (Yue *et al.* 2013, 2014); and (iii) consequently that the estimation of the impact of the Y-chromosome on traits based on paternal lineage is inaccurate (Yue *et al.* 2015).

In addition, we observed very rare carriers of the Holstein Y1 haplotype within the Simmental (1/70) and Montbéliarde (1/45) breeds, which result from planned introgressions during the 1970s and 1980s to improve milk production. Based on pedigree information, there is a single Holstein paternal lineage still segregating at a low rate (0.3%, seventh most important lineage) in the population of Montbéliarde males born between 2015 and 2019. The founder of this lineage is the Red Holstein bull Janes Royal (HOLUSAM000001599157, born in 1970), a remote descendant of the ancestor bull HOLUSAM000000798226 (born around 1950), the lineage of which has disappeared from the present Y-chromosome pool of French Holstein.

Given the large genetic distance between the Y1 and Y2 haplogroups, it is likely that they have accumulated many mutations in addition to the SNPs we are studying, with some of them potentially affecting male reproduction. The presence of Y1 and Y2 haplotypes on a similar genomic background in Montbéliarde represent a unique opportunity to study the impact of the Y variations on fertility traits in the future, when the number of Y1-haplotype carriers among Montbéliarde AI bulls will be sufficient. Estimating the phenotypic consequences of the variants associated with the different Y haplogroups and haplotypes found in West-European cattle breeds is a prerequisite before starting conservation introgression programs to improve Y-chromosome variability and/or male reproductive performances. Finally, we also identified six additional examples of Y1-haplotype carriers within breeds that are normally fixed for the Y2 haplotype. After verification, three of them were attributed to errors in pedigree data, breed affiliation, or in the correspondence between DNA samples and animal ID (Data S1). These results call for a systematic verification of the identity (and breed affiliation) of the individuals whole genome sequenced for example by studying the concordance between genotypes derived from SNP array and from WGS, prior to any analysis. Such errors may have important consequences on imputation and subsequent GWAS or on filtering of breed-specific variants when searching for mutations responsible for recessive genetic defects (Michot *et al.* 2016).

In conclusion, using pedigree information and whole-genome sequence data, we report an extremely low within-breed Y-chromosome diversity in 12 dairy and beef cattle breeds. This situation due to ancient bottlenecks was particularly exacerbated in dairy breeds due to the increased use of modern reproduction techniques. The large set of informative SNPs identified in this study are of particular interest for future studies dedicated to the estimation of the effects of the various Y-chromosome haplotypes on male reproductive performances and to the conservation of Y-chromosome diversity.

Conflict of interest

The authors report no conflict of interest.

Data availability statement

The whole-genome sequence data of 4110 taurine cattle individuals analysed in this study are part of the 1000 Bull Genomes Run 8.0. Part of these are already publicly available in the BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) at accession numbers PRJNA238491, PRJEB9343, PRJNA176557, PRJEB18113, PRNJA343262, PRJNA324822, PRJNA324270, PRJNA277147, PRJEB 5462 (Table S4). The complete dataset with all the Y-chromosome variants of the 811 males used in this study is available from the corresponding author on reasonable request.

References

Boichard D. (2002) PEDIG: a fortran package for pedigree analysis suited for large populations. In: *7th World Congress on Genetics Applied to Livestock Production*, pp 28–9. 23 août 2002, Montpellier, France.

Boichard D., Boussaha M., Capitan A. *et al.* (2018) Experience from large scale use of the EuroGenomics custom SNP chip in cattle. In: *11th World Congress on Genetics Applied to Livestock Production*, pp. 1–6.

Bouwman A.C., Daetwyler H.D., Chamberlain A.J. *et al.* (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics* **50**, 362–7.

Chen N., Cai Y., Chen Q. *et al.* (2018) Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature Communications* **9**, 1–13.

Edwards C.J., Ginja C., Kantanen J. *et al.* (2011) Dual origins of dairy cattle farming – evidence from a comprehensive survey of european Y-chromosomal variation. *PLoS One* **6**, e15922.

Gautier M., Laloë D. & Moazami-Goudarzi K. (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* **5**, e13038.

Hughes J.F., Skaletsky H., Pyntikova T. *et al.* (2020) Sequence analysis in *Bos taurus* reveals pervasiveness of X-Y arms races in mammalian lineages. *Genome Research* **31**, 1716–26.

Leigh J.W. & Bryant D. (2015) POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6**, 1110–6.

Michot P., Chahory S., Marete A. *et al.* (2016) A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. *Genetics Selection Evolution* **48**, 1–15.

Park S.D.E., Magee D.A., McGettigan P.A. *et al.* (2015) Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biology* **16**, 1–15.

Sargolzaei M., Chesnais J.P. & Schenkel F.S. (2014) A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 478.

Yue X.P., Chang T.C., DeJarnette J.M., Marshall C.E., Lei C.Z. & Liu W.S. (2013) Copy number variation of PRAMEY across breeds and its association with male fertility in Holstein sires. *Journal of Dairy Science* **96**, 8024–34.

Yue X.P., Dechow C., Chang T.C., DeJarnette J.M., Marshall C.E., Lei C.Z. & Liu W.S. (2014) Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. *BMC Genomics* **15**, 1–12.

Yue X.P., Dechow C. & Liu W.S. (2015) A limited number of Y chromosome lineages is present in North American Holsteins. *Journal of Dairy Science* **98**, 2738–45.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. List of ancestors accounting for at least 1% of the paternal lineage of their breed.

Table S2. Details on Y-chromosome SNPs considered in this study.

Table S3. Genotype information for the distinct haplotypes identified in this study.

Table S4. Haplogroups, sub-haplogroups and accession numbers of the 811 bulls.

Data S1. Investigations on Y1-haplotype carriers within breeds that are normally fixed for the Y2-haplotype.