Short communication:

Identification of large selective sweeps associated with major genes in cattle

T. Druet[*], L. Pérez-Pardal[*], C. Charlier[*] and M. Gautier[§]

[*]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, B-4000 Liège, Belgium

[§]Institut National de la Recherche Agronomique, UMR CBGP (INRA – IRD – CIRAD – Montpellier SupAgro), F-34988 Montferrier sur Lez, France

Corresponding author:

Tom Druet

Tel: 003243669172/ Fax:

tom.druet@ulg.ac.be

Summary:

Selection for new favorable variants can lead to selective sweeps. However, such sweeps might be rare in the evolution of different species where polygenic adaptation or selection on standing variation might be more common. Still, strong selective sweeps have been described in domestic species such as chicken lines or dog breeds. The goal of our study was to use a panel of individuals from 12 different cattle breeds genotyped at high density (800K SNPs) to perform a whole genome scan for selective sweeps defined as unexpectedly long stretches of reduced heterozygosity. To that end we developed a hidden Markov model where one of the hidden states corresponds to regions of reduced heterozygosity. Some unexpectedly long regions were identified. Among those, six contained genes known to affect traits with simple genetic architecture such as coat color or horn development. However, there was little evidence for sweeps associated to genes underlying production traits.

Key words: Selective sweep, hidden Markov model, selection signatures, cattle

Selection acting on a newly arisen or rare favorable variant tends to sweep out genetic diversity in its surrounding. This phenomenon has first been termed a selective sweep by Maynard Smith & Haigh (1974) and later a hard sweep (e.g., Pritchard *et al*. 2010) to distinguish it from alternative kinds of footprints of selection. Indeed, selection might also be acting on pre-existing or standing variation (e.g. after an environmental shift) leading to a so-called « soft sweep » (Hermisson & Pennings 2005). Similarly, in the case of a polygenic architecture, the optimum phenotype might be defined by various and subtle combinations of favorable variants (Chevin & Hospital 2008). Importantly, polygenic adaptation or « soft-sweeps » are expected to have a much less pronounced impact on local genetic variability than hard sweeps and hence being far less easy to identify by classical approaches aiming at detecting footprints of selection (e.g., Pritchard *et al*. 2010; Peter *et al*. 2012). Recent studies pointed out that hard sweeps might be rare in human evolution (Hernandez *et al*. 2011) and that most of adaptation might be due to polygenic adaptation as reviewed by Pritchard *et al*. (2010).

Populations from domesticated species have recently been subjected to intense selection sometimes on traits with simple genetic architecture (e.g. coat color). Hence hard sweeps might be more common and more intense in such species. As a recent example, by comparing domesticated chicken breeds with their ancestor using whole genome sequence data, Rubin *et al*. (2010) identified a 40 kb region containing the *thyroid stimulating hormone receptor* (*TSHR*) and presenting a strong reduction in heterozygosity.

The objective of the present study was to determine whether long stretches of markers with reduced heterozygosity are common in cattle, possibly indicating selective sweeps and giving the opportunity to discover novel genes subjected to selection during domestication or

breed creation. In cattle, most genome scans for footprints of selection were based on the comparison of allelic frequencies across various breeds (Barendse *et al*. 2009; Flori *et al*. 2009; Gautier *et al*. 2009; Stella *et al*. 2010). In these studies, a few tens of thousands SNPs (e.g. the Illumina BovineSNP50 genotyping assay) were used leading to a typical density of 10 to 20 markers per Mb. Such a density remained too sparse to allow the detection of sweeps based on heterozygosity profiles without ambiguity, particularly for short sweeps. The recently developed Illumina BovineHD genotyping assay proposes a 10-fold increased marker density and thus represents a valuable resource to address this issue.

For the purpose of our study, we used samples from twelve breeds (see Table 1 for details) genotyped on the Illumina BovineHD genotyping assay. We used 725293 SNPs mapping on the 29 autosomes on the UMD 3.1 assembly. Within each breed, only markers with a genotyping call rate above 0.95 were conserved.

We developed a hidden Markov model (**HMM**) to identify within each population regions of reduced heterozygosity in contrast to background regions. Briefly, our HMM relies on three hidden states ('neutral', 'intermediate' and 'sweep') and was inspired from the one proposed by Boitard *et al*. (2009) to detect signatures of selection based on the characterization of the site frequency spectrum. To account for possible residual observed variability (e.g. due to genotyping error, recently arisen SNP, incomplete sweep), heterozygosity level h_{sweep} of markers within sweep regions (i.e. emission probability of the sweep state in our HMM model) were assumed Beta distributed following:

$$h_{sweep} \sim Beta(0.5,49.5)$$

The emission probabilities were categorized into bins of size 0.01 (e.g., from 0.00 to 0.01,

from 0.01 to 0.02, etc). For a marker with heterozygosity x, the emission probability was equal to the density of the Beta distribution in the corresponding bin. With such a parameterization, the distribution assumed for h_{sweep} is concentrated on low values - e.g. $P(h_{sweep}<0.01)=0.68$ and $P(h_{sweep}<0.05)=0.975$ (note also that $P(h_{sweep}>0.5)$ was set equal to 0). In practice, emission probabilities below 0.001 were replaced by a constant value which might be interpreted as genotyping error rate across SNPs (the constant was such that the distribution sums to 1). For the neutral and intermediate regions distribution, the emission probabilities ($h_{intermediate}$ and $h_{neutral}$) were unknown. Finally, transition probabilities (probabilities to move from a hidden state to another between two consecutive markers) were such that going from a neutral to a sweep region (or vice versa) required going through the intermediate region. The transition matrix was identical to the one in Boitard *et al.* (2009) with p = 0.001:

$$T=\begin{pmatrix} 1-p & p & 0 \\ p/2 & 1-p & p/2 \\ 0 & p & 1-p \end{pmatrix}$$

Initial state probabilities were equal to 0.999 for neutral regions and 0.0005 for other regions. $h_{intermediate}$ and $h_{neutral}$ distribution parameters and probabilities of each SNP to be in different hidden states were estimated using an the forward-backward and EM algortihms (e.g., Rabiner 1989). The method is implemented in a software available at http://www.giga.ulg.ac.be/jcms/prod_381171/fr/software.

As a matter of expedience we further defined hard sweeps as regions i) with SNP 'sweep' state probability strictly above 0.999 (i.e. of highly reduced heterozygosity) ii) with a marker

density above 100 SNPs per Mb (to discard uninformative regions), iii) with no gaps between successive SNPs exceeding 40 kb (to avoid selecting regions which are long due to markers spacing) and iv) an average heterozygosity below 0.05.

We were able to identify hard sweeps in all the breeds considered according to the criteria presented above and as summarized in Table 1 (a list of regions identified by breeds is available in Supplementary Material). The longest sweep was 1.08 Mb long in Brown Swiss and the breeds harbored from 0 to 3 sweeps longer than 0.5 Mb and 2 to 9 sweeps longer than 0.25 Mb (see Table 1), which represent unexpectedly long regions of reduced heterozygosity. The median length of sweeps was much lower, ranging from 57 to 104 kb. In total, sweeps covered from 0.12 % (in Piedmontese) to 0.51% (in Guernsey) of the genome (31 to 126 regions according to the breed). Finally, no sweep was found in common to all the twelve breeds as might have been expected in case of a systematic bias introduced by the SNP assay (e.g. genomic regions represented by an unexpectedly high number of lowly polymorphic SNPs across populations).

Interestingly, most of the strongest sweeps identified included genes underlying traits under selection in some breeds (these associated genes are highlighted in the Supplementary Material). For instance, the 504 kb sweep observed on chromosome (BTA) 2 in Belgian Blue beef (**BBB**) cattle (Figure 1) which displays an average heterozygosity of 0.002 contains the *myostatin* gene (*MSTN*). An eleven bp deletion within *MSTN* causing double-muscling (Grobet *et al.* 1997) has been selected for and fixed in BBB (the Piedmontese also displays a sweep which is due to selection for another *MSTN* variant causing double muscling (Kambadur *et al.* 1997)). Other examples correspond to genes affecting coat color or pattern such as the *melanocortin receptor 1* (*MC1R*) also known as the *Extension* locus (Klungland *et al*. 1995), possibly the *KIT* gene associated to the *spotting* locus (e.g., Fontanesi *et al.* 2009),

the *premelanosome protein* (*PMEL*) formerly known as the *SILV* gene responsible for white color in Charolais (Gutierrez-Gil *et al*. 2007), the *POLL* locus causing absence of horns (Seichter *et al*. 2012) and fixed in Angus and Hereford samples or related to stature such as *PLAG1* (Karim *et al*. 2011). For the *MC1R* and *PLAG1* (Figure 2) sweeps, respectively 9 and 7 breeds displayed a sweep with varying sizes. For most of these known genes, surrounding regions displayed a reduced heterozygosity (most often below 0.02) for segments above 0.25 Mb, suggesting recent selection.

These different examples act as positive controls and inform us about the patterns of variation expected around selected genes in cattle. Indeed, for most sweeps identified, the underlying polymorphisms under selection responsible for the observed signals remains to be identified. Yet, some of these sweeps overlap with regions identified by other approaches (e.g. F_{ST}-based genome scans) such as the region around chr2:62,000,000 on UMD3.1 (Barendse *et al*. 2009) or the QTL regions on BTA6 encompassing *NCAPG-LCORL* (Flori et al., 2009; Lindholm-Perry *et al*. 2011; Setoguchi *et al*. 2009) and presumably involved in stature (Lango Allen *et al*. 2010; Pryce *et al*. 2011; Signer-Hasler *et al*. 2012).

Although our HMM model allows to efficiently identify sweeps, it might be sensitive in its current form to both parameterization and to demographic characteristics of the populations under study. Indeed, the model was found sensitive to the shape of the *Beta* distribution assumed for h_{sweep}. For instance and as expected, keeping the same mean but decreasing its variance (i.e. putting more emphasis on low heterozygosity level) resulted in the loss of the sweeps displaying the weakest signals. More generally, because overall genetic variability is tightly related to the intensity of random genetic drift, we expect our model to be sensitive to the effective population size. As expected, we observed that both the estimated proportions of the genome covered by sweeps (as defined above) and the number of sweeps

among the different breeds were highly negatively correlated with the effective population size estimated from the samples (ρ=-0.87 and ρ=-0.81, respectively). Hence, caution needs to be taken in interpreting the weakest sweeps by adapting parameters to specificities of the populations under study and the characteristics of the markers considered (ascertainment scheme and density). Extensive simulation studies might further be required to determine to which extent demographic history of the population under study might generate spurious sweeps under this model.

For the empirical purpose of this study however, positive controls help to better define the expected size of actual sweeps. Among the uncharacterized sweeps identified, those that are as long as or even larger than known sweeps represent the best candidates for further investigation. Similarly, sweeps confirmed in several breeds (Rubin *et al*. 2010), mapping within known QTL (in cross-breeding experiments) or within region where variants were identified in other species (e.g. *NCAPG-LCORL*) might safely be considered as true positive. The advent of high-throughput sequencing technologies makes it possible to provide a detailed picture of sequence variability within such sweep and alternative approaches such as those based on the allele frequency spectrum (Boitard *et al*. 2009; Boitard *et al*. 2012) will help in confirming, refining and possibly identifying the causal variant underlying these observed sweeps. To that end, collaborative efforts such as the cattle 1000 genomes projects are encouraging.

In conclusion, the HMM proposed here allowed us to identify major sweeps around genes with major effects in cattle. However, these sweeps were not so common and the characterized ones (50% (28%) of the detected ones larger than 0.50 Mb (0.25 Mb)) were mostly related to traits with simple architecture (e.g. coat color, horn development). These traits were selected at breed creation, some of them defining the breed. Conversely, there

was little evidence for sweeps associated to genes underlying production traits which have been the target of recent selection and might be expected to present longer sweeps. This suggests a rather polygenic architecture for these traits.

Acknowledgements

We thank André Eggen, Illumina and the BovineHD genotyping array consortium for sharing data from BovineHD genotyping, BBG, Belgimex, Fabroca and BBCI for providing Belgian Blue sires semen samples and Arnaud Sartelet for help in collecting samples. We are grateful to the GIGA-Geno-Transcriptomics Technology Platform and particularly Naima Ahariz, Nadine Cambisano and Sarah Géron for sample genotyping. The authors thank also Michel Georges for his comments and suggestions on this work. Tom Druet and Carole Charlier are respectively Research Associate and Senior Research Associate of the Fonds de la Recherche Scientifique – FNRS. This work was supported by a grant from the Walloon Ministry of Agriculture (D31-1271).

References

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. & Turner L.B. (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* **10**, 178.

Boitard S., Schlotterer C. & Futschik A. (2009) Detecting selective sweeps: a new approach based on hidden markov models. *Genetics* **181**, 1567-78.

Boitard S., Schlotterer C., Nolte V., Pandey R.V. & Futschik A. (2012) Detecting selective sweeps from pooled next-generation sequencing samples. *Molecular Biology and Evolution* **29**, 2177-86.

Chevin L.M. & Hospital F. (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**, 1645-60.

Fontanesi L., Tazzoli M., Russo V. & Beever J. (2009) Genetic heterogeneity at the bovine *KIT* gene in cattle breeds carrying different putative alleles at the *spotting* locus. *PAnimal Genetics* **41**, 295-303.

Flori L., Fritz S., Jaffrezic F., Boussaha M., Gut I., Heath S., Foulley J.L. & Gautier M. (2009) The genome response to artificial selection: a case study in dairy cattle. *PLoS One* **4**, e6595.

Gautier M., Faraut T., Moazami-Goudarzi K., Navratil 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**, 1059-70.

Gautier M., Flori L., Riebler A., Jaffrezic F., Laloe D., Gut I., Moazami-Goudarzi K. & Foulley J.L. (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* **10**, 550.

Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Menissier F., Massabanda J., Fries R., Hanset R. & Georges M. (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 71-4.

Gutierrez-Gil B., Wiener P. & Williams J.L. (2007) Genetic effects on coat colour in cattle: dilution of eumelanin and phaeomelanin pigments in an F2-Backcross Charolais x Holstein population. *BMC Genetics* **8**, 56.

Hermisson J. & Pennings P.S. (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335-52.

Hernandez R.D., Kelley J.L., Elyashiv E., Melton S.C., Auton A., McVean G., Sella G. &

Przeworski M. (2011) Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920-4.

Kambadur R., Sharma M., Smith T.P. & Bass J.J. (1997) Mutations in myostatin (GDF8) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Research* **7**, 910-6.

Karim L., Takeda H., Lin L., Druet T., Arias J.A., Baurain D., Cambisano N., Davis S.R., Farnir F., Grisart B., Harris B.L., Keehan M.D., Littlejohn M.D., Spelman R.J., Georges M. & Coppieters W. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics* **43**, 405-13.

Klungland H., Vage D.I., Gomez-Raya L., Adalsteinsson S. & Lien S. (1995) The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* **6**, 636-9.

Lango Allen H., Estrada K., Lettre G., Berndt S.I., Weedon M.N., Rivadeneira F., Willer C.J., Jackson A.U., Vedantam S., Raychaudhuri S., Ferreira T., Wood A.R., Weyant R.J., Segre A.V., Speliotes E.K., Wheeler E., Soranzo N., Park J.H., Yang J., Gudbjartsson D., Heard-Costa N.L., Randall J.C., Qi L., Vernon Smith A., Magi R., Pastinen T., Liang L., Heid I.M., Luan J., Thorleifsson G., Winkler T.W., Goddard M.E., Sin Lo K., Palmer C., Workalemahu T., Aulchenko Y.S., Johansson A., Zillikens M.C., Feitosa M.F., Esko T., Johnson T., Ketkar S., Kraft P., Mangino M., Prokopenko I., Absher D., Albrecht E., Ernst F., Glazer N.L., Hayward C., Hottenga J.J., Jacobs K.B., Knowles J.W., Kutalik Z., Monda K.L., Polasek O., Preuss M., Rayner N.W., Robertson N.R., Steinthorsdottir V., Tyrer J.P., Voight B.F., Wiklund F., Xu J., Zhao J.H., Nyholt D.R., Pellikka N., Perola M., Perry J.R., Surakka I., Tammesoo M.L., Altmaier E.L., Amin N., Aspelund T., Bhangale T., Boucher G., Chasman D.I., Chen C., Coin L., Cooper M.N., Dixon A.L., Gibson Q., Grundberg E., Hao K., Juhani Junttila M., Kaplan L.M., Kettunen J., Konig I.R., Kwan T., Lawrence R.W.,

Levinson D.F., Lorentzon M., McKnight B., Morris A.P., Muller M., Suh Ngwa J., Purcell S., Rafelt S., Salem R.M., Salvi E., Sanna S., Shi J., Sovio U., Thompson J.R., Turchin M.C., Vandenput L., Verlaan D.J., Vitart V., White C.C., Ziegler A., Almgren P., Balmforth A.J., Campbell H., Citterio L., De Grandi A., Dominiczak A., Duan J., Elliott P., Elosua R., Eriksson J.G., Freimer N.B., Geus E.J., Glorioso N., Haiqing S., Hartikainen A.L., Havulinna A.S., Hicks A.A., Hui J., Igl W., Illig T., Jula A., Kajantie E., Kilpelainen T.O., Koiranen M., Kolcic I., Koskinen S., Kovacs P., Laitinen J., Liu J., Lokki M.L., Marusic A., Maschio A., Meitinger T., Mulas A., Pare G., Parker A.N., Peden J.F., Petersmann A., Pichler I., Pietilainen K.H., Pouta A., Ridderstrale M., Rotter J.I., Sambrook J.G., Sanders A.R., Schmidt C.O., Sinisalo J., Smit J.H., Stringham H.M., Bragi Walters G., Widen E., Wild S.H., Willemsen G., Zagato L., Zgaga L., Zitting P., Alavere H., Farrall M., McArdle W.L., Nelis M., Peters M.J., Ripatti S., van Meurs J.B., Aben K.K., Ardlie K.G., Beckmann J.S., Beilby J.P., Bergman R.N., Bergmann S., Collins F.S., Cusi D., den Heijer M., Eiriksdottir G., Gejman P.V., Hall A.S., Hamsten A., Huikuri H.V., Iribarren C., Kahonen M., Kaprio J., Kathiresan S., Kiemeney L., Kocher T., Launer L.J., Lehtimaki T., Melander O., Mosley T.H., Jr., Musk A.W., Nieminen M.S., O'Donnell C.J., Ohlsson C., Oostra B., Palmer L.J., Raitakari O., Ridker P.M., Rioux J.D., Rissanen A., Rivolta C., Schunkert H., Shuldiner A.R., Siscovick D.S., Stumvoll M., Tonjes A., Tuomilehto J., van Ommen G.J., Viikari J., Heath A.C., Martin N.G., Montgomery G.W., Province M.A., Kayser M., Arnold A.M., Atwood L.D., Boerwinkle E., Chanock S.J., Deloukas P., Gieger C., Gronberg H., Hall P., Hattersley A.T., Hengstenberg C., Hoffman W., Lathrop G.M., Salomaa V., Schreiber S., Uda M., Waterworth D., Wright A.F., Assimes T.L., Barroso I., Hofman A., Mohlke K.L., Boomsma D.I., Caulfield M.J., Cupples L.A., Erdmann J., Fox C.S., Gudnason V., Gyllensten U., Harris T.B., Hayes R.B.,

Jarvelin M.R., Mooser V., Munroe P.B., Ouwehand W.H., Penninx B.W., Pramstaller P.P., Quertermous T., Rudan I., Samani N.J., Spector T.D., Volzke H., Watkins H., Wilson J.F., Groop L.C., Haritunians T., Hu F.B., Kaplan R.C., Metspalu A., North K.E., Schlessinger D., Wareham N.J., Hunter D.J., O'Connell J.R., Strachan D.P., Wichmann H.E., Borecki I.B., van Duijn C.M., Schadt E.E., Thorsteinsdottir U., Peltonen L., Uitterlinden A.G., Visscher P.M., Chatterjee N., Loos R.J., Boehnke M., McCarthy M.I., Ingelsson E., Lindgren C.M., Abecasis G.R., Stefansson K., Frayling T.M. & Hirschhorn J.N. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8.

Lindholm-Perry A.K., Sexten A.K., Kuehn L.A., Smith T.P., King D.A., Shackelford S.D., Wheeler T.L., Ferrell C.L., Jenkins T.G., Snelling W.M. & Freetly H.C. (2011) Association, effects and validation of polymorphisms within the NCAPG - LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC Genetics* **12**, 103.

Maynard Smith J. & Haigh J. (1974) The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23-35.

Peter B.M., Huerta-Sanchez E. & Nielsen R. (2012) Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genetics* **8**, e1003011.

Pritchard J.K., Pickrell J.K. & Coop G. (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology* **20**, R208-15.

Pryce J.E., Hayes B.J., Bolormaa S. & Goddard M.E. (2011) Polymorphic regions affecting human height also control stature in cattle. *Genetics* **187**, 981-4.

Rabiner L.R. (1989) A tutorial on Hidden Markov Chaines and selected applications in speech recognition. *Proc. IEEE* **77**, 257-86.

Rubin C.J., Zody M.C., Eriksson J., Meadows J.R., Sherwood E., Webster M.T., Jiang L., Ingman

M., Sharpe T., Ka S., Hallbook F., Besnier F., Carlborg O., Bed'hom B., Tixier-Boichard M., Jensen P., Siegel P., Lindblad-Toh K. & Andersson L. (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587-91.

Seichter D., Russ I., Rothammer S., Eder J., Forster M. & Medugorac I. (2012) SNP-based association mapping of the polled gene in divergent cattle breeds. *Animal Genetics* **43**, 595-598.

Setoguchi K., Furuta M., Hirano T., Nagao T., Watanabe T., Sugimoto Y. & Takasuga A. (2009) Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC Genetics* **10**, 43.

Signer-Hasler H., Flury C., Haase B., Burger D., Simianer H., Leeb T. & Rieder S. (2012) A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One* **7**, e37282.

Stella A., Ajmone-Marsan P., Lazzari B. & Boettcher P. (2010) Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* **185**, 1451-61.

Figure Legends

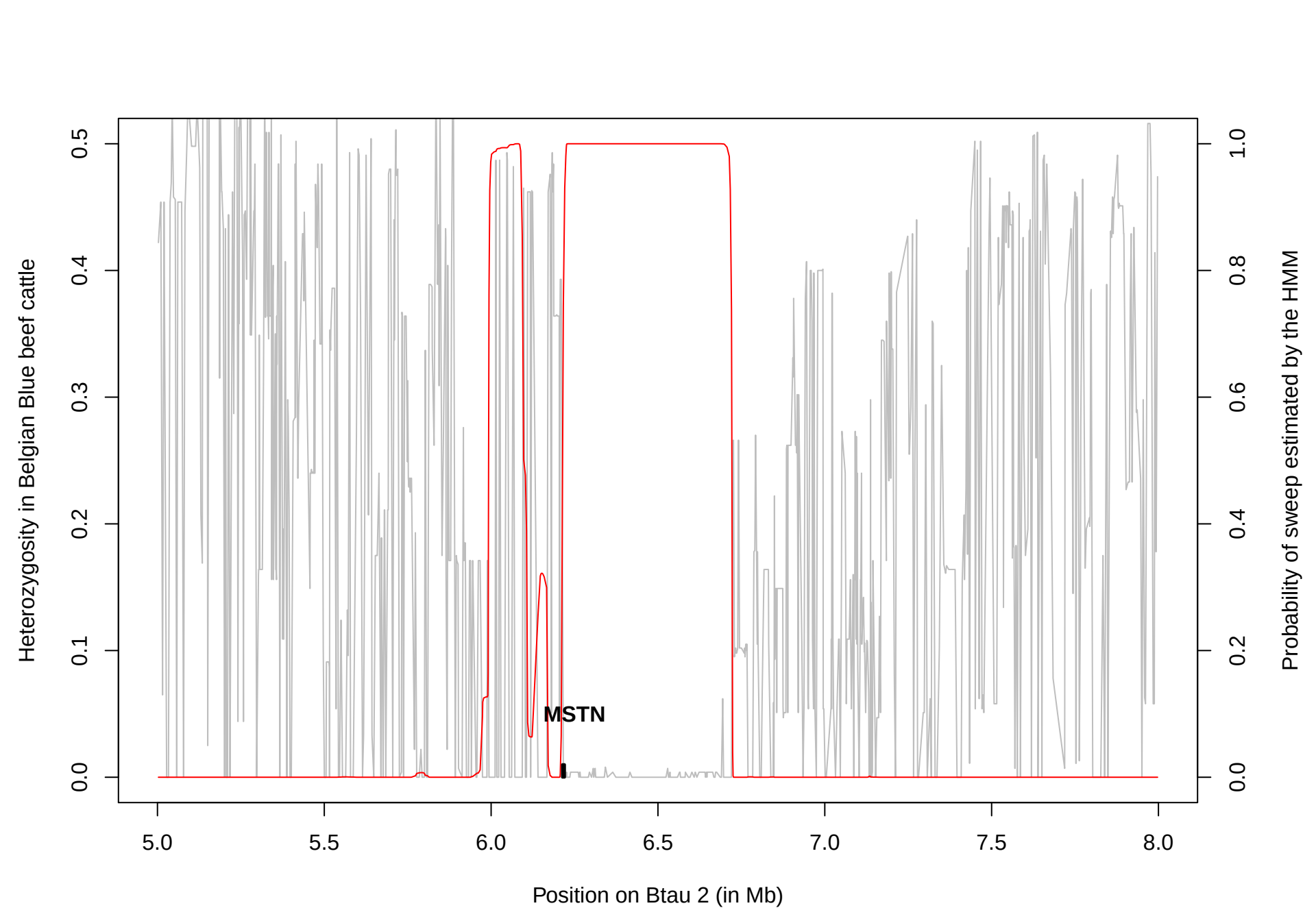
Figure 1: Marker heterozygosity (in grey), in the Belgian Blue beef (BBB) cattle breed in the *MSTN* region mapping to BTA2. The dark line represents the sweep state probability estimated by a hidden Markov model (**HMM**) which contains three hidden states: 'sweep', 'intermediate' and 'neutral' region. Emission probabilities are defined as heterozygosity of the genotyped SNPs. The HMM identifies a clear sweep 0.5 Mb long encompassing *MSTN* in which the heterozygosity level is almost null, contrasting with the surrounding region.

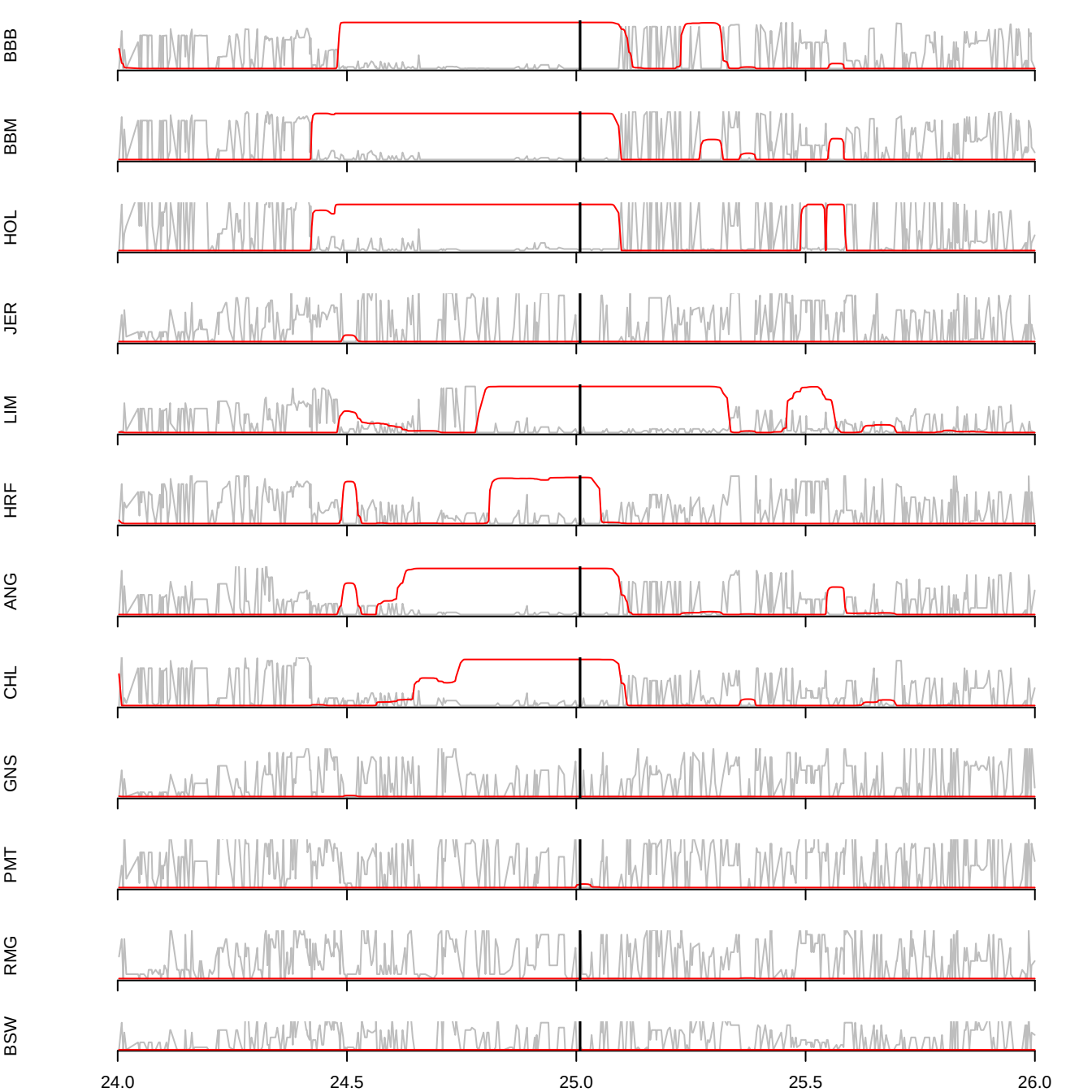
Figure 2: Marker heterozygosity (in grey) in different breeds in the *PLAG1* region on BTA14. The dark line represents the sweep state probability estimated by a hidden Markov model which identifies a clear sweep in seven breeds out of the twelve investigated. The sweep is greater than 0.6 Mb in some breeds and the swept region common to the seven breeds encompasses *PLAG1*. (BBB Belgian Blue beef cattle, BBM dual purpose Belgian Blue, HOL Holstein, JER jersey, LIM Limousine, HRF Hereford, ANG angus, CHL Charolais, GNS Guernsey, PMT Piedmontese, RMG Romagnola, BSW Brown Swiss).

Table 1. Number of identified segments with the HMM by breed.

Breed	Number of individuals	Number of segments	Median segment length	Number of Segments > 0,50 Mb	Number of segments >0,25 Mb	Maximum segment size	%genome	Effective population size[1]
Belgian Blue Beef breed (BBB)	275	126	65 kb	3	4	0,657 Mb	0,46	109
Dual purpose Belgian Blue (BBM)	52	80	81 kb	1	4	0,659 Mb	0,33	127
Holstein (HOL)	60	125	70 kb	1	5	0,607 Mb	0,46	92
Jersey (JER)	38	80	101 kb	3	7	0,598 Mb	0,45	74
Limousin (LIM)	50	51	91 kb	0	3	0,452 Mb	0,22	267
Hereford (HRF)	35	58	104 kb	1	5	0,512 Mb	0,31	124
Angus (ANG)	42	104	82 kb	0	7	0,466 Mb	0,42	101
Charolais (CHL)	37	40	58 kb	1	3	0,994 Mb	0,17	285
Guernsey (GNS)	21	112	84 kb	1	9	0,535 Mb	0,51	100
Piedmontese (PMT)	21	31	61 kb	0	2	0,371 Mb	0,12	372
Romagnola (RMG)	21	70	57 kb	0	3	0,328 Mb	0,21	139
Brown Swiss (BSW)	22	106	74 kb	1	6	1,080 Mb	0,43	65

[1]Effective population size was estimated from the linkage-disequilibrium between markers separated by 10 Mb with the following approximation: $r^2 = 1/(1 + 4N_e c)$ where c is the distance in cM between markers (assuming 1cM≈1Mb (Gautier *et al.* 2007) in cattle populations).


Heterozygosity in Belgian Blue beef cattle
0.0
0.1
0.2
0.3
0.4
0.5
Probability of sweep estimated by the HMM
0.0
0.2
0.4
0.6
0.8
1.0
MSTN
5.0
5.5
6.0
6.5
7.0
7.5
8.0
Position on Btau 2 (in Mb)


BBB
BBM
HOL
JER
LIM
HRF
ANG
CHL
GNS
PMT
RMG
BSW
24.0
24.5
25.0
25.5
26.0