

Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations

Morgane Petit,* Jean-Michel Astruc,[†] Julien Sarry,* Laurence Drouilhet,* Stéphane Fabre,*
Carole R. Moreno,* and Bertrand Servin*¹

*INRA, Génétique, Physiologie, Systèmes d'Élevage, and [†]Institut de l'Élevage, F-31326 Castanet-Tolosan, France

ORCID IDs: 0000-0002-2498-0926 (C.R.M.); 0000-0001-5141-0913 (B.S.)

ABSTRACT Recombination is a complex biological process that results from a cascade of multiple events during meiosis. Understanding the genetic determinism of recombination can help to understand if and how these events are interacting. To tackle this question, we studied the patterns of recombination in sheep, using multiple approaches and data sets. We constructed male recombination maps in a dairy breed from the south of France (the Lacaune breed) at a fine scale by combining meiotic recombination rates from a large pedigree genotyped with a 50K SNP array and historical recombination rates from a sample of unrelated individuals genotyped with a 600K SNP array. This analysis revealed recombination patterns in sheep similar to other mammals but also genome regions that have likely been affected by directional and diversifying selection. We estimated the average recombination rate of Lacaune sheep at 1.5 cM/Mb, identified ~50,000 crossover hotspots on the genome, and found a high correlation between historical and meiotic recombination rate estimates. A genome-wide association study revealed two major loci affecting interindividual variation in recombination rate in Lacaune, including the *RNF212* and *HEI10* genes and possibly two other loci of smaller effects including the *KCNJ15* and *FSHR* genes. The comparison of these new results to those obtained previously in a distantly related population of domestic sheep (the Soay) revealed that Soay and Lacaune males have a very similar distribution of recombination along the genome. The two data sets were thus combined to create more precise male meiotic recombination maps in Sheep. However, despite their similar recombination maps, Soay and Lacaune males were found to exhibit different heritabilities and QTL effects for interindividual variation in genome-wide recombination rates. This highlights the robustness of recombination patterns to underlying variation in their genetic determinism.

KEYWORDS recombination rate; genetic maps; QTLs; evolution; sheep

Meiotic recombination is a fundamental biological process that brings a major contribution to the genetic diversity and the evolution of eukaryotic genomes (Baudat *et al.* 2013). During meiosis, recombination enables chromosomal alignment resulting in proper disjunction and segregation of chromosomes, avoiding deleterious outcomes such as aneuploidy (Hassold *et al.* 2007). Over generations, recombination contributes to shaping genetic diversity in a population by creating new allelic combinations and

preventing the accumulation of deleterious mutations. Over large evolutionary timescales, divergence in recombination landscapes can lead to speciation; the action of a key factor in the recombination process in many mammals, the gene *PRDM9*, has been shown to have a major contribution to the infertility between two mouse species, making it the only known speciation gene in mammals today (Mihola *et al.* 2009).

Genetics studies on recombination were first used to infer the organization of genes along the genome (Sturtevant 1913). With advances in molecular techniques, more detailed physical maps and eventually whole-genome assemblies are now available in many species. The establishment of highly resolutive recombination maps remains of fundamental importance for the validation of the physical ordering of markers obtained from sequencing experiments (Groenen *et al.* 2012; Jiang *et al.* 2014). From an evolutionary

Copyright © 2017 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.117.300123>

Manuscript received June 30, 2017; accepted for publication July 31, 2017; published Early Online September 8, 2017.

Available freely online through the author-supported open access option.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300123/-/DC1.

¹Corresponding author: INRA, Génétique, Physiologie, Systèmes d'Élevage, 24 Chemin de Borde Rouge, Auzeville-Tolosane, CS 52627, F-31326 Castanet-Tolosan, France. E-mail: bertrand.servin@inra.fr

perspective, the relevant distance between loci is the genetic distance and recombination maps are essential tools for the genetic studies of a species, for estimation of past demography (Li and Durbin 2011; Boitard *et al.* 2016), detection of selection signatures (Sabeti *et al.* 2002; Voight *et al.* 2006), QTL mapping (Cox *et al.* 2009), and imputation of genotypes (Howie *et al.* 2009) for genome-wide association studies (GWAS) or genomic selection. Precise recombination maps can be estimated using different approaches. Meiotic recombination rates can be estimated from the observation of markers' segregation in families. Although this is a widespread approach, its resolution is limited by the number of meioses that can be collected within a population and the number of markers that can be genotyped. Consequently, highly resolutive meiotic maps have been produced in situations where large segregating families can be studied and genotyped densely (Shifman *et al.* 2006; Mancera *et al.* 2008; Groenen *et al.* 2009; Rockman and Kruglyak 2009; Kong *et al.* 2010) or by focusing on specific genomic regions (Cirulli *et al.* 2007; Stevison and Noor 2010; Kaur and Rockman 2014). In livestock species, the recent availability of dense genotyping assays has fostered the production of highly resolutive recombination maps (Tortereau *et al.* 2012; Johnston *et al.* 2016, 2017), particularly by exploiting reference population data from genomic selection programs (Sandor *et al.* 2012; Ma *et al.* 2015; Kadri *et al.* 2016).

Another approach to study the distribution of recombination on a genome is to exploit patterns of correlation between allele frequencies in a population (*i.e.*, linkage disequilibrium, LD) to infer past (historical) recombination rates (McVean *et al.* 2002; Li and Stephens 2003; Chan *et al.* 2012). Because the LD-based approach in essence exploits meioses accumulated over many generations, it can provide more precise estimates of local variation on recombination rate. For example, until recently (Pratto *et al.* 2014; Lange *et al.* 2016) this was the only known indirect approach allowing the detection of fine-scale patterns of recombination genome-wide in species with large genomes. Several highly recombining intervals (recombination hotspots) were detected from historical recombination rate maps and confirmed or completed those discovered by sperm-typing experiments (Crawford *et al.* 2004; Myers *et al.* 2005). One important caveat of LD-based approaches is that their recombination rate estimates are affected by other evolutionary processes, especially selection that affects LD patterns unevenly across the genome. Hence, differences in historical recombination between distant genomic regions have to be interpreted with caution. Despite this, historical and meiotic recombination rates usually exhibit substantial positive correlation (Rockman and Kruglyak 2009; Brunschwig *et al.* 2012; Chan *et al.* 2012; Wang *et al.* 2012).

The LD-based approach does not allow the study of individual phenotypes to directly identify loci influencing interindividual variation in recombination rates. In contrast, family-based studies in human (Kong *et al.* 2008; Chowdhury *et al.* 2009), *Drosophila* (Stevison and Noor 2010; Chan *et al.*

2012) mice (Shifman *et al.* 2006; Brunschwig *et al.* 2012), cattle (Sandor *et al.* 2012; Ma *et al.* 2015; Kadri *et al.* 2016), and sheep (Johnston *et al.* 2016) have demonstrated that recombination exhibits interindividual variation and that this variation is partly determined by genetic factors. Two recombination phenotypes have been described: the number of crossovers per meiosis (genome-wide recombination rate, GRR herein) and the fine-scale localization of crossovers (Individual Hotspot Usage, IHU). GRR has been shown to be influenced by several genes. For example, a recent GWAS found evidence for association with six genome regions in cattle (Kadri *et al.* 2016). Among them, one of the genomic regions consistently found associated with GRR in mammals is an interval containing the *RNF212* gene. In contrast to GRR, the IHU phenotype seems mostly governed by a single gene in most mammals: *PRDM9*. This zinc-finger protein has a key role in recruiting *SPO11*, thereby directing DNA double-strand breaks (DSBs) that initiate meiotic recombination. Because *PRDM9* recognizes a specific DNA motif, the crossover events happen in hotspots carrying this motif. However, this *PRDM9*-associated process is not universal as it is only active in some mammals; canids, for example, do not carry a functional copy of *PRDM9* and exhibit different patterns for the localization of recombination hotspots (Auton *et al.* 2013).

As mentioned above, recombination was studied recently in sheep (Johnston *et al.* 2016), which led to the production of precise genome-wide recombination maps, revealed a similar genetic architecture of recombination rates in sheep as in other mammals, and identified two major loci affecting individual variation. Quite interestingly, one of the QTL identified in this study, localized near the *RNF212* gene, was clearly demonstrated to have a sex-specific effect. This study was performed in a feral population of sheep that is quite distantly related to continental populations (Kijas *et al.* 2012) and has not been managed by humans for a long time. To understand how recombination patterns and genetic determinism can vary across populations, in this work we conducted a study in another sheep population, the Lacaune, from the south of France. The Lacaune breed is the main dairy sheep population in France, its milk being mainly used for the production of Roquefort cheese. Starting in 2011, a large genotyping effort started in the breed to implement a genomic selection program (Baloche *et al.* 2014), and young selection candidates are now routinely genotyped for a medium-density genotyping array (~50K SNP). This constitutes a large data set of genotyped families that can be used to study recombination, although limited to one sex as only males were used for genomic selection in this population. This data set offers an opportunity to study variation in recombination and its genetic determinism between very diverged populations of the same species. Hence, a first objective of this study was to elucidate whether these two sheep populations had similar distribution of recombination on the genome and whether they shared the same genetic architecture of the trait, and in particular the same QTL effects.

The second objective of this study was to compare different approaches to study recombination from independent data in the same population. To this end, in addition to the pedigree data, we exploit a sample of 51 unrelated individuals genotyped with a high-density genotyping array (~500K SNP). While, the family data were used to establish meiotic recombination maps, the sample of densely genotyped individuals was used to create historical recombination maps of higher resolution. This offered the opportunity to evaluate to what extent sheep ancestral recombination patterns match contemporary ones.

Materials and Methods

Study population and genotype data

In this work, we exploited two different data sets of sheep from the Lacaune breed: a pedigree data set of 8085 related animals genotyped with the medium-density Illumina Ovine Beadchip including 54,241 SNPs, and a diversity data set of 70 unrelated Lacaune individuals selected as to represent population genetic diversity, genotyped with the high-density Illumina Ovine Infinium HD SNP Beadchip (Moreno-Romieux *et al.* 2017; Rochus *et al.* 2017).

Standard data cleaning procedures were carried out on the pedigree data set using plink 1.9 (Chang *et al.* 2015) excluding animals with call rates below 95% and SNPs with call frequency < 98%. After quality controls, we exploited genotypes at 46,813 SNPs and 5940 meioses. For these animals, we only selected the sires that had their own sire known and at least two offspring and the sires that did not have their own sire known but at least four offspring. Eventually, 345 male parents, called focal individuals (FIDs) hereafter, met these criteria: 210 FIDs had their father genotype known while the remaining 135 did not (Figure 1).

Recombination maps

Meiotic recombination maps from pedigree data:

Detection of crossovers: Crossover locations were detected using LINKPHASE (Druet and Georges 2015). From the LINKPHASE outputs (recombination_hmm files), we extracted crossovers boundaries. We then identified crossovers occurring in the same meiosis < 3 Mb apart from each other (that we call double crossovers) and considered them as dubious. This number was chosen as it corresponded to clear outliers in the distribution of inter-crossover distances. They are also quite unlikely under crossover interference. We applied the following procedure: given a pair of double crossovers, we set the genotype of the corresponding offspring as missing in the region spanned by the most extreme boundaries and reran the LINKPHASE analysis. After this quality control step, we used the final set of crossovers identified by LINKPHASE to estimate recombination rates. This data set consisted of 213,615 crossovers in 5940 meioses.

Estimation of recombination rates: Based on the inferred crossover locations, meiotic recombination rates were estimated in windows of 1 Mb and between marker intervals of

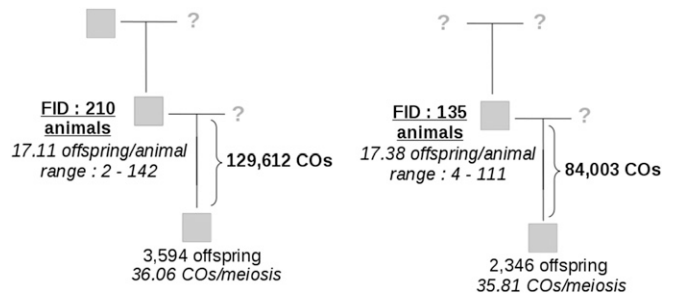


Figure 1 Families used to infer crossover (CO) events. COs were identified in meioses of 345 focal individuals (FIDs). Two-hundred and ten FIDs had their father known (left) while 135 FIDs did not (right).

the medium SNP array using the following statistical model, inspired by Cheung *et al.* (2007). For small genetic intervals such as considered here, the recombination rate (termed c in the following), is usually expressed in centimorgans per megabase and the probability that a crossover occurs in one meiosis in an interval j (measured in morgans) is $0.01 c_j l_j$ where l_j is the length of the interval expressed in megabases. When considering M meioses, the expected number of crossovers in the interval is $0.01 c_j l_j M$. When combining observations in multiple individuals, we want to account for the fact that they have different average numbers of crossovers per meiosis (termed R_s for individual s). To do so, we multiply the expected number of crossovers in the interval by an individual-specific factor equal to (R_s/R) , where R is the average number of crossovers per meiosis among all individuals. Finally, for individual s in interval j , the expected number of crossovers is $0.01 c_j l_j M_s R_s/R$. Given this expected number, a natural distribution to model the number of crossovers observed in an interval is the Poisson distribution so that the number y_{sj} of crossovers observed in the interval j for an individual s is modeled as:

$$y_{sj}|c_j \sim \text{Poisson}(0.01 c_j l_j M_s R_s/R). \quad (1)$$

To combine crossovers across individuals, the likelihood for c_j is the product of Poisson likelihoods from Equation 1.

We then specify a prior distribution for c_j :

$$c_j \sim \Gamma(\alpha, \beta). \quad (2)$$

To set α and β , first raw c_j estimates are computed using the method of Sandor *et al.* (2012) across the genome and then a γ distribution is fitted to the resulting genome-wide distribution (Supplemental Material, Figure S1 in File S11). Combining the prior (2) with the likelihoods in Equation (1), the posterior distribution for c_j is:

$$c_j|y_j \sim \Gamma\left(\alpha + \sum_s y_{sj}, \beta + 0.01 l_j \sum_s M_s R_s/R\right). \quad (3)$$

As the localization of crossovers was usually not good enough to assign them with certainty to a single genomic interval, final estimates of c_j are obtained as follows:

1. For each crossover overlapping interval j and localized within a window of size L , let x_c be an indicator variable that takes value 1 if the crossover occurred in interval j and 0 otherwise. Assuming that, locally, recombination rate is proportional to physical distance, set $P(x_c = 1) = \min(l_j/L, 1)$.
2. Using the probability in step 1, sample x_c for each crossover overlapping interval j and set $y_{sj} = \sum_c x_c$.
3. Given y_{sj} , sample c_j from Equation (3).

For each interval considered, perform step 2 and 3 above 1000 times to draw samples from the posterior distribution of c_j , thereby accounting for uncertainty in the localization of crossovers.

Historical recombination maps from the diversity data:

The diversity data contains 70 Lacaune individuals genotyped for a high-density (HD) SNP array comprising 527,823 autosomal markers (Rochus *et al.* 2017). Nineteen of these individuals are FIDs in the pedigree data. To perform the LD-based analysis on individuals unrelated to the pedigree study, these individuals were therefore removed from the data set and the subsequent analyses performed on the 51 remaining individuals. Population-scaled recombination rates were estimated using PHASE (Li and Stephens 2003). For computational reasons and to allow for varying effective population size along the genome, estimations were carried out in 2 Mb windows, with an additional 100 kb on each side overlapping with neighboring windows, to avoid a border effect in the PHASE inference. PHASE was run on each window with default options, except that the number of main iterations was increased to obtain larger posterior samples for recombination rate estimation (option -X10) as recommended in the documentation.

From the PHASE output, 1000 samples were obtained from the posterior distribution of:

The background recombination rate: $\rho_w = 4N_w c_w$, where N_w is the effective population size in the window, c_w is the recombination rate comparable to the family-based estimate.

An interval-specific recombination intensity λ_j , for each marker interval j of length l_j in the window, such that the population scaled genetic length of an interval is: $\delta_j = \rho_w \lambda_j l_j$.

The medians were used as point estimates of parameters λ_j and δ_j , computed over the posterior distributions $\{\lambda_j^{(k)}, \lambda_j^{(k)} \rho_w^{(k)} l_j ; k \in [1, 1000]\}$.

Intervals that showed an outlying λ_j value compared to the genome-wide distribution of λ_j were considered as harboring a crossover hotspot. Specifically, a mixture of Gaussian distribution was fitted to the genome-wide distribution of $\log_{10}(\lambda_j)$ using the mclust R package (Fraley and Raftery 2002; Fraley *et al.* 2012), considering that the major component of the mixture modeled the background distribution of λ_j in non-hotspot intervals. From this background distribution, a P -value was computed for each interval that corresponded

to the null hypothesis that it does not harbor a hotspot. Finally, hotspot-harboring intervals were defined as those for which false discovery rate (FDR) (λ_j) < 5%, estimating FDR with the Storey and Tibshirani (2003) method, implemented in the R qvalue package. This procedure is illustrated in Figure S2 in File S11.

Combination of meiotic and historical recombination rates and construction of a high-resolution recombination map:

To construct a meiotic recombination map of the HD SNP array requires that the historical recombination rate estimates be scaled by four times the effective population size. Due to evolutionary pressures, the effective population size varies along the genome, so it must be estimated locally. This can be done by exploiting the meiotic recombination rate inference obtained from the pedigree data analysis, as explained below.

Consider a window of 1 Mb on the genome; using the approach described above, we can sample values c_{jk} (window j , sample k) from the posterior distribution of the meiotic recombination rate c_j . Similarly, using output from PHASE, we can extract samples ρ_{jk} from the posterior distribution of the historical recombination rates ($\rho_j = \rho_w \lambda_j$). Now, considering that $\rho_j = 4N_e c_j$ where N_e is the local effective population size of window j , we get $\log(\rho_j) = \log(4N_e) + \log(c_j)$. This justifies using a model on both c_{jk} and ρ_{jk} values:

$$y_{ijk} = \mu + x_{ijk}\alpha + \beta_j + \nu_{ij} + e_{ijk} \quad (4)$$

where y_{ijk} is $\log(c_{jk})$ when $i = 1$ (meiotic-recombination rate sample) and y_{ijk} is $\log(\rho_{jk})$ when $i = 2$ (historical recombination rate sample). In this model, μ estimates the log of the genome-wide recombination rate, $x_{ijk} = 1$ if $i = 2$ and 0 otherwise, so that α estimates $\log(4N_e)$, where N_e is the average effective population size of the Lacaune population, $\mu + \beta_j$ estimates $\log(c_j)$ combining population and meiotic recombination rates, and $\alpha + (\nu_{2j} - \nu_{1j})$ estimates $\log(4N_e)$. μ and α_i were considered as fixed effects while β_j and ν_{ij} were considered as independent random effects. Using this approach allows us to combine, in a single model, LD- and pedigree-based inferences, while accounting for their respective uncertainties as we exploit posterior distribution samples.

Model 4 was fitted on 20 samples of the posterior distributions of c_j and ρ_j for all windows of 1 Mb covering the genome, with an additional fixed effect for each chromosome, using the lme4 R package (Bates *et al.* 2015). Windows lying < 4 Mb from each chromosome end were not used because inference on c_j was possibly biased in these regions (see *Results*). After estimating this model, historical recombination rate estimates of HD intervals were scaled within each window by dividing them by their estimated local effective population size (*i.e.*, $\exp(\hat{\alpha} + \widehat{\nu}_{2j} - \widehat{\nu}_{1j})$ for window j). For windows lying within 4 Mb of the chromosome ends, historical recombination rate estimates were scaled using the genome-wide average effective population size $\exp(\hat{\alpha})$. This

led eventually to estimates of the meiotic recombination rates, expressed in centimorgans per megabase, for all intervals of the HD SNP array, which we termed a high-resolution recombination map.

For each interval of the medium-density SNP array, we computed the number of significant hotspots detected as explained above and the hotspot density (number of hotspots per unit of physical distance). After having corrected for the chromosome effect, the GC content effect and for windows farther than 4 Mb of the chromosome end, we fitted a linear regression model to estimate the effect of hotspot density on the meiotic recombination rate.

Comparison with Soay sheep recombination maps and integration of the two data sets to produce new male recombination maps in sheep: To compare the recombination maps in Lacaune with the previously established maps in Soay sheep (Johnston *et al.* 2016), we downloaded the raw data from the dryad data repository (doi: 10.5061/dryad.pf4b7) and the additional information available on https://github.com/susjoh/GENETICS_2015_185553. As the approach used in Johnston *et al.* (2016) to establish recombination maps differs from the one used here, we chose to apply the method of this study to the Soay data to perform a comparison that would not be affected by differences in the methods. As the Lacaune data consist only of male meioses, we also only considered male meioses in the Soay data. The final Soay data set used consisted of 3445 individuals, among which were 299 male FIDs, defined as in the Lacaune analysis. After detecting crossovers with LINKPHASE, one FID exhibited a very high average number of crossover per meiosis (> 100) and was not considered in the analyses (Soay individual ID: RE4844), leaving 298 FIDs. The final data set consisted of 88,683 crossovers in 2609 male meioses and was used to estimate meiotic recombination maps using the exact same approach as described above, both on intervals of 1 Mb and on the same intervals as the ones considered in the Lacaune meiotic maps on the medium-density SNP array. Note that the Soay sheep are not necessarily polymorphic for the same markers as the Lacaune, but that our method is flexible and can nonetheless estimate recombination rates in intervals bordered by monomorphic markers: in such a case adjacent intervals will have the same estimated recombination rate. As the two populations were found to have very similar meiotic recombination maps (see *Results*), the two sets of crossovers were finally merged to create a combined data set of 302,298 crossovers in 8549 male meioses and to estimate new male sheep recombination maps, again on 1 Mb intervals and on intervals of the medium-density SNP array.

GWAS on recombination phenotypes: GRR

The set of crossovers detected was used to estimate the GRR of each FID in the family data set from their observed number of crossovers per meiosis, adjusting for covariates: year of birth of the parent, considered as a cofactor with 14 levels for years spanning from 1997 to 2010, and insemination month of

the offspring's ewe, treated as a cofactor with seven levels for months spanning from February to August. We used a mixed model for estimating the population average GRR μ , covariates fixed effects β , and individual breeding values u_s , while controlling for nongenetic individual-specific effects a_s :

$$y_{so} = \mu + x_{so}\beta + a_s + u_s + e_{so}$$

with $u_s \sim N(0, A\sigma_s^2)$, $a_s \sim N(0, I\sigma_a^2)$, and $e_{so} \sim N(0, I\sigma_e^2)$, where y_{so} is the number of crossovers in the meiosis between FID s and offspring o , A is the pedigree-based relationship matrix between FIDs, and x_{so} the line of the corresponding design matrix for observation y_{so} . We fitted this model using BLUPf90 (<http://nce.ads.uga.edu/software/>) and extracted: (i) estimates of variance components σ_e^2 , σ_a^2 , and σ_s^2 , which allows the estimation of the heritability of the trait (calculated as $\sigma_s^2/(\sigma_s^2 + \sigma_a^2 + \sigma_e^2)$) and (ii) prediction \tilde{u}_s of GRR deviation for each FID.

Genotype imputation

Nineteen of the 345 FIDs are present in the diversity data set of HD genotypes. For the 336 remaining FIDs, their HD genotypes at 507,784 SNPs were imputed with BimBam (Servin and Stephens 2007; Guan and Stephens 2008) using the 70 unrelated Lacaune individuals as a panel. To impute, BimBam uses the fastPHASE model (Scheet and Stephens 2006), which relies on methods using clusters of haplotypes to estimate missing genotypes and reconstruct haplotypes from unphased SNPs of unrelated animals. BimBam was run with 10 expectation-maximization (EM) starts, each EM was run for 20 steps on panel data alone, and an additional step was run on cohort data, with a number of clusters of 15. After imputation, BimBam estimates an average number of alleles for each SNP in each individual, termed mean genotype, computed from the posterior distribution of the three possible genotypes. This mean genotype has been shown to be efficient for performing association tests (Guan and Stephens 2008). In subsequent analyses, we used the mean genotypes provided by BimBam of the 345 FIDs at all markers of the HD SNP array. To assess the quality of genotype imputation at the most associated regions, 10 markers of the HD SNP array, one in the chromosome 6-associated region and nine in the chromosome 7-associated region (see *Results*) were genotyped for 266 FIDs for which DNA samples were still available. We evaluated the quality of imputation for the most significant SNPs by comparing each possible genotype's posterior probability estimated by BimBam to the error rate implied by calling it. We observed a very good agreement between the two measures (Figure S3 in File S11), which denoted good calibration of the imputed genotypes at top GWAS hits.

Single- and multi-QTL GWAS on GRR

We first tested association of individual estimated breeding values \tilde{u}_s (EBVs) with mean genotypes at 503,784 single SNPs imputed with BimBam. We tested these associations

using the univariate mixed model approach implemented in the Genome-wide Efficient Mixed Model Association (Gemma) software (Zhou and Stephens 2012). To account for polygenic effects on the trait, the centered genomic relationship matrix calculated from the mean genotypes was used. The *P*-values reported in the results correspond to the Wald test.

To go beyond single SNP association tests, we also estimated a Bayesian sparse linear mixed model (Zhou *et al.* 2013) as implemented in Gemma. This method allows the consideration of multiple QTL in the model, together with polygenic effects at all SNPs. The principle of the method is to have, for each SNP *l*, an indicator variable γ_l that takes value 1 if the SNP is a QTL and 0 otherwise. The strength of evidence that a SNP is a QTL is measured by the posterior probability $P(\gamma_l = 1)$, called posterior inclusion probability (PIP). Note that all SNPs are included in the model when doing so. Inference of the model parameters is performed using an iterative MCMC algorithm: the number of iterations was set to 10 million and inference was made on samples extracted every 100 iterations. When a genome region harbors a QTL, multiple SNPs in the region can have elevated PIPs. To summarize the strength of evidence for a region to carry a QTL, we calculated a rolling sum of PIPs over 50 consecutive SNPs using the rollsum function of the R zoo package (Zeileis and Grothendieck 2005). Given that the average physical distance between SNPs on the high-density SNP array is ~ 5 kb, this procedure interrogates the probability of the presence of a QTL in overlapping windows of ~ 250 kb.

For the univariate analysis, the FDR was estimated using the ash package (Stephens 2017), and SNPs corresponding to an FDR < 10% were deemed significant and annotated. For the multivariate analysis, regions where the rolling sum of PIPs were > 0.15 were further annotated. The annotation of the QTL regions consisted of extracting all genes from the Ensembl annotation v87 along with their Gene Ontology (GO) annotations and interrogating for their possible involvement in recombination.

Variant discovery and additional genotyping in RNF212: identification and assignation of the RNF212 sheep genome sequence

The *RNF212* gene was not annotated on the *Ovis aries* v3.1 reference genome. Nevertheless, a full sequence of *RNF212* was found in the scaffold01089 of *O. orientalis* [assembly Oori1, National Center for Biotechnology information (NCBI) accession NW_011943327]. By BLAST alignment of this scaffold, ovine *RNF212* could be located with confidence on chromosome 6 in the interval OAR6:116426000–116448000 of the Oari3.1 reference genome (Figure S4 in File S11). This location was confirmed by BLAST alignment with the bovine *RNF212* gene sequence. We also discovered that the Oari3.1 unplaced scaffold005259 (NCBI accession JH922970) contained the central part of *RNF212* (exons 4–9) and could be placed within a large assembly gap. Moreover, we also observed that automatically annotated noncoding RNA in the

RNF212 interval matched the exonic sequence of *RNF212* (Figure S4 in File S11).

Variant discovery in RNF212 in the Lacaune population

Based on the genomic sequence and structure of the *RNF212* gene annotated in *O. orientalis* (NCBI accession NW_011943327), a large set of primers were designed using PRIMER3 software (Table S1 in File S11) for amplification of each annotated exon and some intron parts corresponding to exonic regions annotated in *Capra hircus* (Chir_v1.0). PCR amplification (GoTaq; Promega, Madison, WI) with each primer pair was realized on 50 ng of genomic DNA from four selected homozygous Lacaune animals exhibiting the GG and AA (nonimputed) genotypes at the most significant SNP of the medium-density SNP array of the chromosome 6 QTL (rs418933055, *P*-value $2.56e-17$). Each PCR product was sequenced via the BigDye Terminator v3.1 Cycle Sequencing kit and analyzed on an ABI3730 sequencing machine (Applied Biosystems, Foster City, CA). Sequenced reads were aligned against the *O. orientalis* *RNF212* gene using CLC Main Workbench Version 7.6.4 (QIAGEN, Valencia, CA) to identify polymorphisms.

Genotyping of mutations in RNF212

The genotyping of 266 genomic DNAs from Lacaune animals for the four identified polymorphisms within the ovine *RNF212* gene was done by Restriction Fragment Length Polymorphism after PCR amplification using dedicated primers (Table S1 in File S11) (GoTaq; Promega), restriction enzyme digestion (*Bsr*BI for SNP_14431_AG; *Rsa*I for SNP_18411_GA; and *Bsu*36I for both SNP_22570_CG and SNP_22594_AG; New England Biolabs, Beverley, MA), and resolution on a 2% agarose gel.

Data availability

Genotype data and pedigree information on Lacaune individuals after quality controls are deposited on Zenodo (Astruc *et al.* 2017) as well as high-density genotypes of 70 unrelated Lacaune individuals (Moreno-Romieux *et al.* 2017). Computer code and scripts needed to reproduce all results are available on Github (<https://github.com/BertrandServin/sheep-recombination>) and described in supporting material File S10. Additional data, including output from PHASE, LINKPHASE, BimBam, and Gemma, are provided on the Zenodo repository (DOI: 10.5281/zenodo.821569) (Petit *et al.* 2017).

Results

High-resolution recombination maps

Meiotic recombination maps: genome-wide recombination patterns: We studied meiotic recombination using a pedigree of 6230 individuals, genotyped for a medium-density SNP array (50K) comprising $\sim 54,000$ markers. After quality controls, we exploited genotypes at 46,813 SNPs and identified 213,615 crossovers in 5940 meioses divided among 345 male

parents (FIDs) (see *Materials and Methods*). The pedigree information available varied among FIDs (Figure 1); 210 FIDs had their father genotype known while the remaining 135 did not. Having a missing parent genotype did not affect the detection of crossovers as the average number of crossovers per meiosis in the two groups was similar (36.1 with known father genotype and 35.8 otherwise) and the statistical effect of the number of offspring on the average number of crossovers per meiosis was not significant ($P > 0.23$). This can be explained by the fact that individuals that lacked father genotype information typically had a large number of offspring (17.4 on average, ranging from 4 to 111), allowing us to infer correctly their haplotype phase from their offspring genotypes only. Overall, given that the physical genome size covered by the medium-density SNP array is 2.45 Gb, we estimate that the mean recombination rate in our population is ~ 1.5 cM/Mb.

Based on the crossovers identified, we developed a statistical model to estimate meiotic recombination rates (see *Materials and Methods*) and constructed meiotic recombination maps at two different scales: for windows of 1 Mb and for each interval of the medium-density SNP array. As this statistical approach allowed us to evaluate the uncertainty in recombination rate estimates, we provide them in [File S1](#) and [File S2](#), along with the recombination rate estimates in each interval, their posterior variance, and 90% credible intervals. Graphical representation of the meiotic recombination maps of all autosomes are given in [File S3](#).

The recombination rate on a particular chromosome region was found to depend highly on its position relative to the telomere and to the centromere for metacentric chromosomes, *i.e.*, chromosomes 1, 2, and 3 in sheep (Figure S5 in [File S11](#)). Specifically, for acrocentric and metacentric chromosomes, recombination rate estimates were elevated near telomeres and centromeres, but very low within centromeres. In our analysis, recombination rate estimates were found to be low in intervals lying within 4 Mb of chromosome ends. While this could represent genuine reduction in recombination rates near chromosome ends, it is also likely due to crossovers being undetected in our analysis as only a few markers are informative to detect crossovers at chromosome ends. In the following analyses, we therefore did not consider regions lying within 4 Mb of the chromosome ends.

From local recombination rate estimates in 1-Mb windows or medium SNP array intervals, we estimated chromosome-specific recombination rates (Figure S6 in [File S11](#)). Difference in recombination rates between chromosomes was relatively well-explained by their physical size, with larger chromosomes exhibiting smaller recombination rates. Even after accounting for their sizes, some chromosomes showed particularly low (chromosomes 9, 10, and 20) or particularly high (chromosomes 11 and 14) recombination rates. In low recombining chromosomes, large regions had very low recombination, between 9 and 14 Mb on chromosome 9 and 36, 46 Mb on chromosome 10, and between 27 and 31 Mb on chromosome 20. In highly recombining chromosomes,

recombination rates were globally higher on chromosome 14, while chromosome 11 exhibited two very high-recombination windows between 7 and 8 Mb and between 53 and 54 Mb. In addition, we found, consistent with the literature, that GC content was quite significantly positively correlated with recombination rate both in medium SNP array intervals (P -value $< 10^{-16}$, $r = 0.20$) and in 1 Mb intervals (P -value $< 10^{-16}$, $r = 0.28$).

Estimation of historical recombination rates and identification of crossover hotspots: We used a different data set, with 51 unrelated individuals from the same Lacaune population genotyped for the Illumina HD SNP array (600K) comprising 527,823 autosomal SNPs after quality controls. Using a multipoint model for LD patterns (Li and Stephens 2003), we estimated, for each marker interval of the HD SNP array, historical recombination rates ρ (see *Materials and Methods*). Compared to meiotic maps, these estimates offer a greater precision as they in essence exploit meioses cumulated over many generations. However, the historical recombination rates obtained are scaled by the effective population size ($\rho = 4 N_e c$, where N_e is the effective population size and c the meiotic recombination rate), which is unknown and may vary along the genome due to evolutionary pressures, especially selection. Thanks to the higher precision in estimation of recombination rate, LD-based recombination maps offer the opportunity to detect genome intervals likely to harbor crossover hotspots. A statistical analysis of historical recombination rates (see *Materials and Methods*) identified $\sim 50,000$ intervals exhibiting elevated recombination intensities (Figure S2 in [File S11](#)) as recombination hotspots, corresponding to an FDR of 5%. From our historical recombination map, we could conclude that 80% crossover events occurred in 40% of the genome and that 60% of crossover events occurred in only 20% of the genome (Figure S7 in [File S11](#)).

High-resolution recombination maps combining family and population data: Having constructed recombination maps with two independent approaches and having data sets in the same population of Lacaune sheep, we first evaluated to what extent historical crossover hotspots explain meiotic recombination, and second estimated the impact of evolutionary pressures on the historical recombination landscape of the Lacaune population. We present our results on these questions in turn.

We studied whether variation in meiotic recombination can be attributed to the historical crossover hotspots detected from LD patterns only. For each interval between two adjacent SNPs of the density array, we (i) extracted the number of significant historical hotspots and (ii) calculated the historical hotspot density (in number of hotspots per unit of physical distance). We found both covariates to be highly associated with meiotic recombination rate estimated on family data [$r = 0.15$ with hotspot density ($P < 10^{-16}$) and $r = 0.19$ with the number of hotspots ($P < 10^{-16}$)]. These correlations

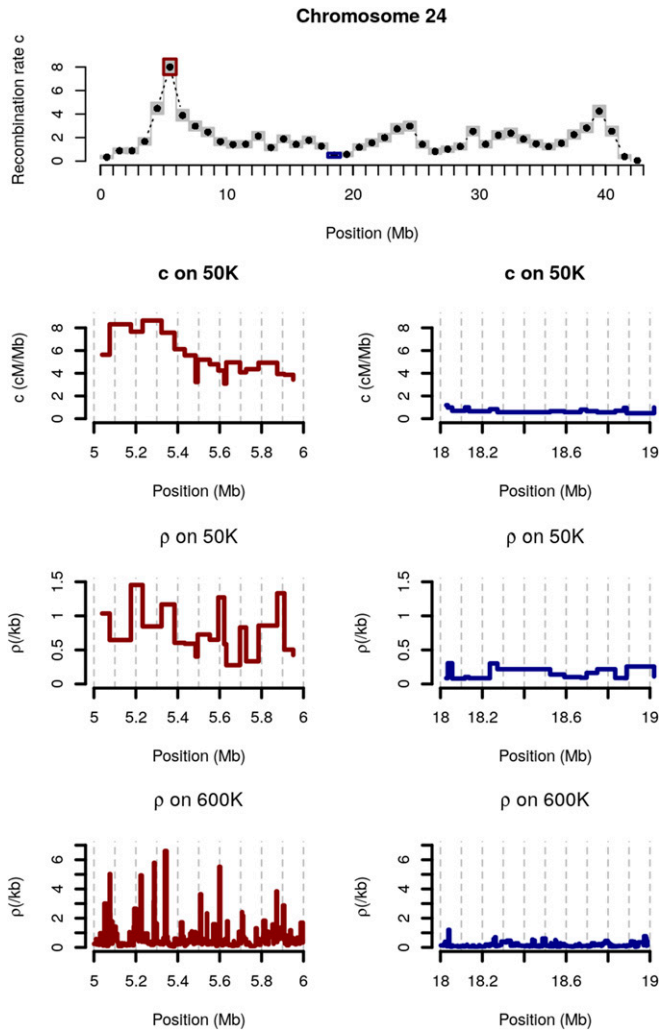


Figure 2 Comparison between population-based recombination rate and meiotic recombination rate for two 1-Mb windows on Sheep chromosome 24. Top: meiotic recombination rate along chromosome 24. Two windows with high (left, red) and low (right, blue) meiotic recombination rates estimates are zoomed in. Each panel represents, from top to bottom: meiotic recombination rate estimates (c) in SNP intervals of the 50K SNP array, population-based recombination rate estimates (ρ) in SNP intervals of the 50K SNP array, and population-based recombination rate estimates (ρ) in SNP intervals of the High-Density (600K) SNP array.

hold after correcting for chromosome and GC content effects [respectively, $r = 0.14$ ($P < 10^{-16}$) and $r = 0.18$ ($P < 10^{-16}$)]. Figure 2 illustrates this finding in two 1-Mb intervals from chromosome 24: one that exhibits a very high recombination rate (7.08 cM/Mb) and the second a low one (0.46 cM/Mb). In this comparison, the highly recombining window carries 36 recombination hotspots while the low recombinant one exhibits none. As the historical background recombination rates in the two windows are similar (0.7/kb for the one with a high recombination rate, and 0.2/kb for the other), the difference in recombination rate between these two regions is largely due to their contrasted number of historical crossover hotspots.

To study more precisely the relationship between historical and meiotic recombination rates, we fitted a linear mixed

model (see *Materials and Methods*) that allowed us to estimate the average effective population size of the population, the correlation between meiotic and historical recombination rates, and to identify genome regions where historical and meiotic recombination rates were significantly different. We found the effective population size of the Lacaune population to be ~ 7000 individuals and a correlation of 0.73 between meiotic and historical recombination rates (Figure 3). We discovered seven regions where historical recombination rates were much lower than meiotic ones and three regions where they were much higher (FDR < 0.02 , Figure S8 in File S11 and Table 1).

Seven of these 10 regions have extreme recombination rates compared to other genomic regions. To quantify to what extent a window is extreme, we indicate in Table 1, for each window, the proportion of the genome with a lower recombination rate (q_w). For six of these seven regions, the historical recombination rate is more extreme than the meiotic rate: four regions have very low meiotic recombination rate and even lower historical recombination rates (the two regions on chromosome 3 and two regions on chromosome 10, between 36 and 37 Mb and between 42 and 44 Mb), while two regions have very high meiotic recombination rates and even higher historical recombination rates (on chromosome 12 and on chromosome 23). For these six regions, the discrepancy between meiotic recombination and historical recombination estimates can be explained by the fact that we used a genome-wide prior in our model to estimate meiotic recombination rates that has the effect of shrinking our estimates toward the mean. Because historical estimates were not shrunk in the same way, for these six outlying regions the two estimates did not concur and it is possible that our meiotic recombination rate estimates were slightly overestimated (or, respectively, underestimated).

Out of the four remaining outlying windows, three had a low historical recombination rate but did not have particularly extreme meiotic recombination rates, so that the effect of shrinkage is not likely to explain the discrepancy between meiotic and historical recombination rates. Indeed, these three regions corresponded to previously identified selection signatures in sheep: a region on chromosome 6 spanning two intervals between 36 and 38 Mb contains the *ABCG2* gene, associated with milk production (Cohen-Zinder *et al.* 2005), and the *LCORL* gene associated to stature [recently reviewed in Takasuga (2016)]. This region has been shown to have been selected in the Lacaune breed (Fariello *et al.* 2014; Rochus *et al.* 2017). A region spanning one interval on chromosome 10, between 29 and 30 Mb, contains the *RXFP2* gene, which is associated with polledness and horn phenotypes (Johnston *et al.* 2013), is found to be under selection in many sheep breeds (Fariello *et al.* 2014), and a region on chromosome 13 between 63 and 64 Mb that contains the *ASIP* gene is responsible for coat color phenotypes in many breeds of sheep (Norris and Whan 2008), and again was previously demonstrated to have been under selection. For these three regions, we explain the low historical

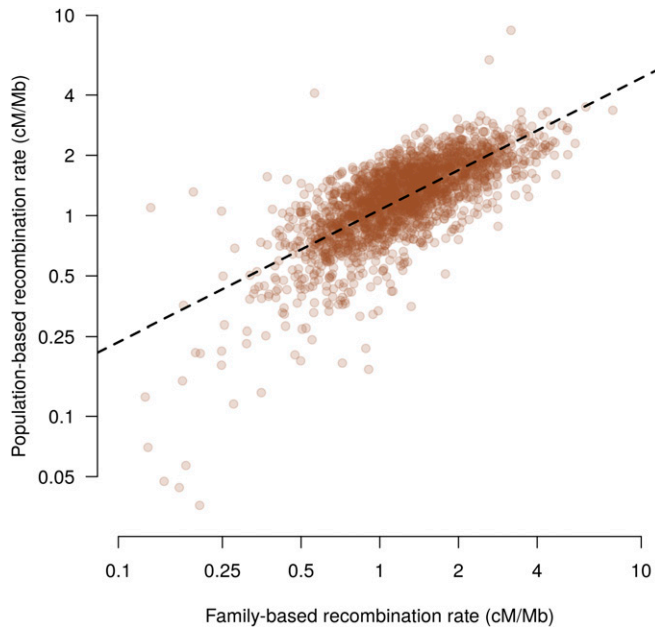


Figure 3 Population-based and meiotic recombination rates in windows of 1 Mb. The dashed line is the regression for population-based recombination rate on the family recombination rate. Values are shown on a logarithmic scale.

recombination estimates by a local reduction of the effective population size due to selection.

Finally, one of the three regions with a high historical recombination rate, on chromosome 20 between 28 and 29 Mb, had a low meiotic recombination rate, so that the effect of shrinkage cannot explain the discrepancy. This region harbors a cluster of olfactory receptor genes and its high historical recombination rate could be explained by selective pressure for increased genetic diversity in these genes (*i.e.*, diversifying selection), a phenomenon that has been shown in other species [*e.g.*, pig (Groenen *et al.* 2012), human (Ignatieva *et al.* 2014), and rodents (Stathopoulos *et al.* 2014)]. Finally, we used the meiotic recombination rates to scale the historical recombination rate estimates and produce high-resolution recombination maps on the HD SNP array (File S4).

Improved male recombination maps by combining Lacaune and Soay sheep data: Recently, recombination maps have been estimated in another sheep population, the Soay (Johnston *et al.* 2016). Soay is a feral population of ancestral domestic sheep living on an island located northwest of Scotland. The Lacaune and Soay populations are genetically very distant, their genome-wide F_{st} , calculated using the sheep HapMap data (Kijas *et al.* 2012), being ~ 0.4 . Combining our results with results from the Soay offered a rare opportunity to study the evolution of recombination over a relatively short time-scale as the two populations can be considered to have been separated at most dating back to domestication, $\sim 10,000$ years ago. The methods used in the Soay study are different from those used here, but the two data sets are similar, although the Soay data has fewer male meioses (2604 vs.

Table 1 Genome regions where meiotic and population-based recombination rates differ significantly

Chromosome	Window Span (Mb)	q_w	P -Value	ρ_s/c
3	103–104	0.06	$1.6 \cdot 10^{-5}$	0.28
3	109–110	0.04	$1.8 \cdot 10^{-5}$	0.28
6 ^a	36–38	0.14	$1.2 \cdot 10^{-7}$	0.22
10 ^a	29–30	0.77	$8.8 \cdot 10^{-5}$	0.31
10	36–37	0.01	$2.1 \cdot 10^{-5}$	0.29
10	42–44	< 0.01	$1.2 \cdot 10^{-14}$	0.11
13 ^a	63–64	0.33	$7.4 \cdot 10^{-6}$	0.31
12	4–5	0.92	$7.4 \cdot 10^{-6}$	3.7
20 ^a	28–29	0.01	$1.7 \cdot 10^{-5}$	3.6
23	10–11	0.97	$5.1 \cdot 10^{-6}$	3.8

ρ_s/c : ratio of population to meiotic recombination rate. q_w : proportion of genome regions with lower meiotic recombination rate. Details on the estimation of these parameters are given in the text. Regions with P -values $< 10^{-4}$ were considered outliers (FDR = 0.02).

^a Regions corresponding to potential selection signatures.

5940 in the present study). To perform a comparison that would not be affected by differences in estimation methods, we ran the method developed for the Lacaune data to estimate recombination maps on the Soay data. As the Soay study showed a clear effect of sex on recombination rates, we estimated recombination maps on male meioses only. Figure 4 presents the comparison of recombination rates between the two populations in marker intervals of the medium-density SNP array. The left panel shows that the two populations exhibit very similar recombination rates ($r = 0.82$, $P < 10^{-16}$), although Soay recombination rates appear higher for low recombining intervals ($c < 1.5$ cM/Mb in gray on the figure). We explain this by the shrinkage effect of the prior, which is more pronounced in the Soay as the data set is smaller: the right panel on Figure 4 shows that the posterior variance of the recombination rates are clearly higher in Soays for low recombining intervals while they are similar for more recombining intervals. Overall, our results on the comparison of the recombination maps in the two populations are consistent with the two populations having the same amplitude and distribution of recombination on the genome, at the scale of the medium-density SNP array. Therefore, we analyzed the two populations together to create new male recombination maps based on 302,298 crossovers detected in 8549 meioses (File S5). Combining the two data sets together led to a clear reduction in the posterior variance of the recombination rates, *i.e.*, an increase in their precision (Figure S9 in File S11).

Genetic determinism of genome-wide recombination rate in Lacaune sheep

Our data set provides information on the number of crossovers for a set of 5940 meioses among 345 male individuals. Therefore, it allows us to study the number of crossovers per meiosis (GRR) as a recombination phenotype.

Genetic and environmental effects on GRR: We used a linear mixed model to study the genetic determinism of GRR. The contribution of additive genetic effects was estimated by including a random FID effect with covariance structure

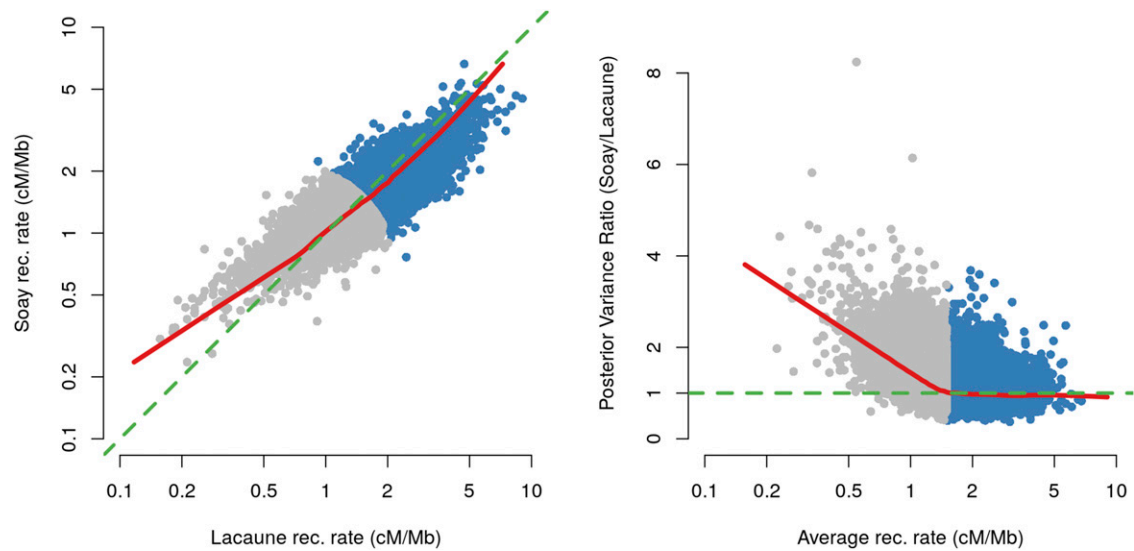


Figure 4 Comparison of recombination (rec.) rates in Soay and Lacaune populations. Left: scatterplot of posterior means of rec. rates (on a log scale). The green line is the line $y = x$ and the red line is a lowess smoothed line ($f = 0.05$). Right: Scatterplot of the ratio of posterior variance (Soay/Lacaune) as a function of the average of the posterior mean rec. rates in the two populations (on a log scale). The green line corresponds to equal variances and the red line is a lowess smoothed line ($f = 0.05$). Points in gray on both panels correspond to intervals with average rec. rate < 1.5 cM/Mb.

proportional to the matrix of kinship coefficients calculated from pedigree records (see *Materials and Methods*). We also included environmental fixed effects in the model: year of birth of the FID and insemination month of the ewe for each meiosis. We did not find significant differences between the FID year of birth; however, the insemination month of the ewe was significant ($P = 1.3 \cdot 10^{-3}$). There was a trend in increased recombination rates from February to May followed by a decrease until July and a regain in August, although the number of inseminations in August is quite low, leading to a high SE for this month (Figure S10 in File S11). Based on the estimated variance components (Table 2), we estimated the heritability of GRR in the Lacaune male population at 0.23.

GWAS identifies three major loci affecting GRR in Lacaune sheep: The additive genetic values of FIDs, predicted from the above model were used as phenotypes in a GWAS. Among the 345 FIDs with at least two offspring, the distribution of the phenotype was found to be approximately normally distributed (Figure S11 in File S11). To test for association of this phenotype with SNPs markers, we used a mixed model approach correcting for relatedness effects with a genomic relationship matrix (see *Materials and Methods*). Using our panel of 70 unrelated Lacaune, we imputed the 345 FIDs for markers of the HD SNP array. With these imputed genotypes, we performed two analyses. The first was an association test with univariate linear mixed models, which tested the effect of each SNP in turn on the phenotype (results in File S6); the second fitted a Bayesian sparse linear mixed model, allowing multiple QTL to be included in the model (results in File S7).

Figure 5 illustrates the GWAS results: the top plot shows the P -values of the single SNP analysis and the bottom plot

the posterior probability that a region harbors a QTL, calculated on overlapping windows of 20 SNPs. The single SNP analysis revealed six significant regions ($FDR < 10\%$): two on chromosome 1, one on chromosome 6, one on chromosome 7, one on chromosome 11, and one chromosome 19. Regions of chromosomes 6 and 7 exhibited very low P -values whereas the other three showed less-intense association signals. The multi-QTL Bayesian analysis was conclusive for two regions (regions on chromosome 6 and chromosome 7) while the rightmost region on chromosome 1 was suggestive (Table 3). Two additional suggestive regions were discovered on chromosome 3. Use of the multi-QTL approach of Zhou *et al.* (2013) led to estimate that, together, QTL explain $\sim 40\%$ of the additive genetic variance for GRR, with a 95% credible interval ranging from 28 to 53%.

The most significant region was located on the distal end of chromosome 6 and corresponded to a locus frequently associated with variation in recombination rate. In our study, the significant region contained 10 genes: *CTBP1*, *IDUA*, *DGKQ*, *GAK*, *CPLX1*, *UVSSA*, *MFSD7*, *PDE6B*, *PIGG*, and *RNF212*. For each of these genes except *RNF212*, which was not annotated on the genome (see below), we extracted their gene GO of the Ensembl v87 database, but none was clearly annotated as potentially involved in recombination. However, two genes were already reported as having a statistical association with recombination rate: *CPLX1* and *GAK* (Kong *et al.* 2014). *CPLX1* has no known function that can be linked to recombination (Kong *et al.* 2014) but *GAK* has been shown to form a complex with cyclin-G, which could impact recombination (Nagel *et al.* 2012). However, *RNF212* can be deemed a more likely candidate due to its function and given that this gene was associated with recombination rate variation in human (Kong *et al.* 2008; Chowdhury *et al.* 2009), cows (Sandor *et al.*

Table 2 Genetic parameters of the interindividual variation in GRR

Number of Sires	Additive Genetic Variance	Phenotypic Variance	Heritability
345	6.86 (0.75)	29.73 (0.84)	0.23 (0.02)

2012; Ma *et al.* 2015; Kadri *et al.* 2016), and mice (Reynolds *et al.* 2013). *RNF212* is not annotated in the sheep genome assembly oviAri3; however, this chromosome 6 region corresponds to the bovine region that contains *RNF212* (Figure S4 in File S11). We found an unassigned scaffold (scaffold01089, NCBI accession NW_011943327) of *O. orientalis musimon* (assembly Oori1) that contained the full *RNF212* sequence and that could be placed confidently in the QTL region. To confirm *RNF212* as a valid positional candidate, we further studied the association of its polymorphisms with GRR in results presented below.

The second most significant region was located between 22.5 and 23.1 Mb on chromosome 7. All significant SNPs in the region were imputed, *i.e.*, the association would not have been found based on association of the medium-density array alone. It matched an association signal on GRR in Soay sheep (Johnston *et al.* 2016). Consistent with our finding, in the Soay sheep study, this association was only found using regional heritability mapping and not using single SNP associations with the medium-density SNP array. This locus could match previous findings in cattle (association on chromosome 10 at ~20 Mb on assembly btau3.1); however, the candidate genes mentioned in this species (*REC8* and *RNF212B*) were located 2 and 1.5 Mb away from our strongest association signal, respectively. In addition, none of the SNPs located around these two candidate genes in cattle were significant in our analysis. Eleven genes were present in the region: *OR10G2*, *OR10G3*, *TRAV5*, *TRAV4*, *SALL2*, *METTL3*, *TOX4*, *RAB2B*, *CHD8*, *SUPT16H*, and *RPGRIP1*. The study of their GO, extracted from the Ensembl v87 database, revealed that none of them were associated with recombination, although *SUPT16H* could be involved in mitotic DSB repair (Kari *et al.* 2011). However, another functional candidate, *CCNB1IP1*, also named *HEI10*, was located between positions 23,946,971 and 23,951,850 bp, ~500 kb from our association peak. This gene is a good functional candidate as it has been shown to interact with *RNF212*: *HEI10* allows the elimination of the *RNF212* protein from early recombination sites and the recruitment of other recombination intermediates involved in crossover maturation (Qiao *et al.* 2014; Rao *et al.* 2016). Again, SNPs located at the immediate proximity of *HEI10* did not exhibit significant associations with GRR. Hence, our association signal did not allow us to pinpoint any clear positional candidate among these functional candidates (see Figure S12 in File S11). However, it was difficult to rule them out completely for three reasons. First, with only 345 individuals, our study may not be powerful enough to localize QTL with the required precision. Second, the presence of causal regulatory variants, even at distances of several 100 kb, is possible. Finally, the

associated region of *HEI10* exhibited apparent rearrangements with the human genome, possibly due to assembly problems in oviAri3. These assembly problems could be linked to the presence of genomic sequences coding for the T-cell receptor α chain. This genome region is in fact rich in repeated sequences, making its assembly challenging. Overall, identifying a single positional and functional candidate gene in this gene-rich misassembled genomic region was not possible based on our data alone.

Our third associated locus was located on chromosome 1 between 268,600 and 268,700 kb. In cattle, the homologous region, located at the distal end of cattle chromosome 1, has also been shown to be associated with GRR (Ma *et al.* 2015; Kadri *et al.* 2016). In these studies, the *PRDM9* gene has been proposed as a potential candidate gene, especially because it is a strong functional candidate given its proven effect on recombination phenotypes. In sheep, *PRDM9* is located at the extreme end of chromosome 1, ~275 Mb, 7 Mb away from our association signal (Ahlawat *et al.* 2016). Hence, *PRDM9* was not a good positional candidate for association with GRR in our sheep population. However, the associated region on chromosome 1 contains a single gene, *KCNJ15*, which has been associated with DNA DSB repair in human cells (Słabicki *et al.* 2010).

Finally, the two regions on chromosome 3 were analyzed. The first was located between 75,162 and 75,319 kb and contains only one annotated gene coding for the receptor for follicle-stimulating hormone (*FSHR*). Although it does not affect recombination directly, it is necessary for the initiation and maintenance of normal spermatogenesis in males (Tapanainen *et al.* 1997). The second region on the chromosome 3 was located between 201,198 and 201,341 kb but does contain any annotated gene.

Mutations in the *RNF212* gene are strongly associated to genome-wide recombination rate variation in Lacaune sheep: The QTL with the largest effect in our association study corresponded to a locus associated with GRR variation in other species and harboring the *RNF212* gene. As it was a clear positional and functional candidate gene, we carried out further experiments to interrogate specifically polymorphisms within this gene. As stated above, we used the sequence information available for the *RNF212* gene from *O. orientalis*, which revealed that *RNF212* spanned 23.7 kb on the genome and may be composed of 12 exons by homology with bovine *RNF212*. However, mRNA annotation indicated multiple alternative exons. Surprisingly, the genomic structure of ovine *RNF212* was not well-conserved with goat, human, and mouse syntenic *RNF212* genes (Figure S4 in File S11). As a first approach, we designed primers for PCR amplification

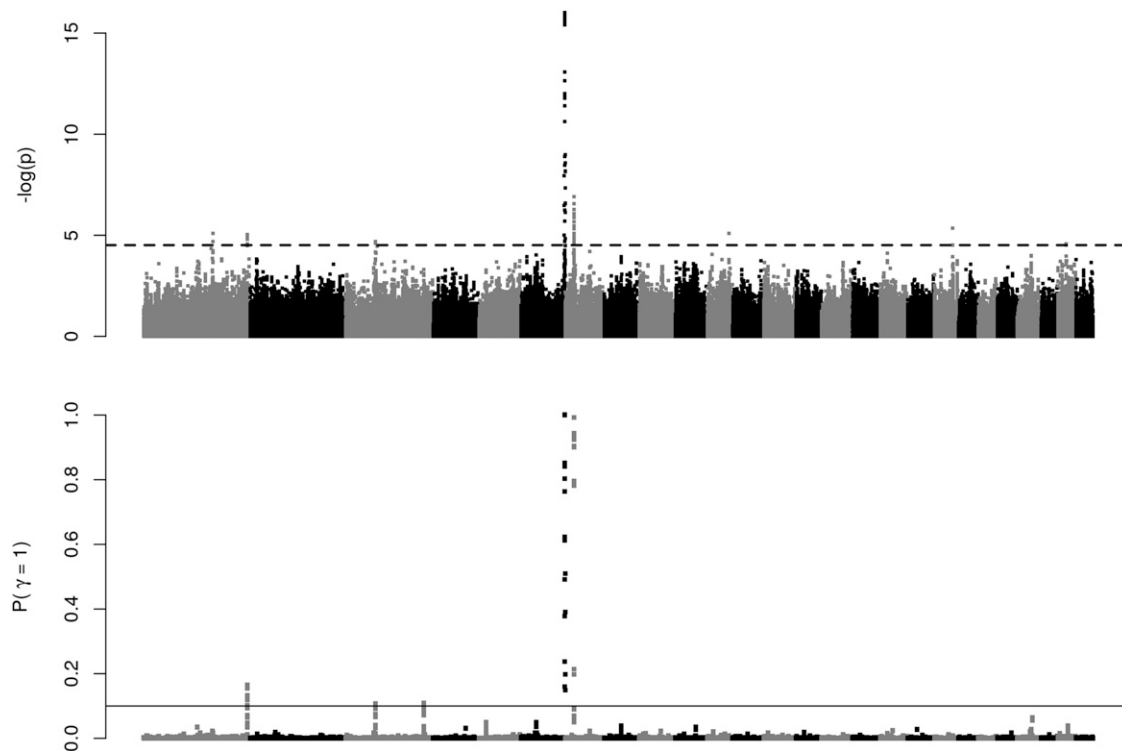


Figure 5 Genome-wide association study identifies three main QTL for GRR. Top: $-\log_{10}(P\text{-value})$ for single SNP tests for association. The genome-wide significance level (FDR = 5%) is represented by the horizontal dotted line. Bottom: posterior probability that a region of 20 SNPs harbors a QTL, using a Bayesian multi-QTL model. FDR, false discovery rate; GRR, genome-wide recombination rate.

(see *Materials and Methods*) and sequencing of all annotated exons and some intronic regions corresponding to exonic sequences of *C. hircus* *RNF212*. By sequencing *RNF212* from four carefully chosen Lacaune animals homozygous for GG or AA at the most significant SNP of the medium-density SNP array on chromosome 6 QTL (rs418933055, $P\text{-value}$ 2.56×10^{-17}), we evidenced four polymorphisms within the ovine *RNF212* gene (two SNPs in intron 9 and two SNPs in exon 10). The four mutations were genotyped in 266 individuals of our association study. We then tested their association with GRR using the same approach as explained above (results in [File S8](#)) and computed their LD (genotypic r^2) with the most associated SNPs of the high-density genotyping array (see [Figure S13](#) in [File S11](#)) ([Table 4](#)). Two of these mutations were found to be highly associated with GRR, their $P\text{-values}$ being of the same order of magnitude ($P < 10^{-16}$) as the most associated SNP (rs412583165), and one of them was even more significant than the most significant imputed SNP ($P = 6.25 \times 10^{-17}$ vs. $P = 9.8 \times 10^{-17}$). We found a clear agreement between the amount of LD between a mutation and the most associated SNPs and their association $P\text{-value}$ (see [Figure S13](#) in [File S11](#)). Overall, these results showed that polymorphisms within the *RNF212* gene were strongly associated with GRR, and likely tagged the same causal mutation as the most associated SNP. This confirmed that *RNF212*, a very strong functional candidate, was also a very strong positional candidate gene underlying our association signal.

The genetic determinism of recombination differs between Soay and Lacaune males

GWAS in the Soay identified two major QTL for GRR, with apparent sex-specific effects. These two QTL were located in the same genomic regions as our QTL on chromosome 6 and chromosome 7. The chromosome 6 QTL was only found to be significant in Soay females, while we detected a very strong signal in Lacaune males. Although the QTL was located in the same genomic region, the most significant SNPs were different in the two GWAS ([Figure 6](#)). Two possible explanations could be offered for these results: either the two populations have the same QTL segregating and the different GWAS hits correspond to different LD patterns between SNPs and QTL in the two populations, or the two populations have different causal mutations in the same region. Denser genotyping data, for example by genotyping the *RNF212* mutations identified in this work in the Soay population, would be needed to have a clear answer. For the chromosome 7 QTL, the signal was only found using regional heritability mapping ([Nagamine et al. 2012](#)) in the Soay, and after genotype imputation in our study, which makes it even more difficult to discriminate between a shared causal mutation or different causal mutations at the same location in the two populations.

Discussion

In this work, we studied the distribution of recombination along the sheep genome and its relationship to historical

Table 3 SNPs associated with GRR (*P*-values correspond to the single-SNP Wald test)

Rs Number	Chr	Position (bp)	Minor Allele	ρ	β	<i>P</i> -Value	pQTL
rs430436336	1	180044043	A	0.11	2.19	8.08×10^{-6}	0.006
rs400472211	1	268670581	A	0.33	0.86	9.41×10^{-6}	0.03
rs418551122	3	75216491	A	0.3	0.76	2.42×10^{-5}	0.04
rs407545143	3	201298545	G	0.24	1.13	9.36×10^{-4}	0.07
rs411987057	6	116517201	C	0.22	-2.3	1.31×10^{-16}	0.19
rs401206888	6	116440663	G	0.14	-1.95	2.04×10^{-16}	0.16
rs412583165	6	116525709	G	0.27	-2.38	9.80×10^{-17}	0.15
rs429477322	6	116509403	A	0.18	-2.17	3.94×10^{-16}	0.11
rs161854895	6	116491013	G	0.22	-2.17	2.53×10^{-16}	0.11
rs398811467	6	116472870	A	0.13	-1.94	2.51×10^{-16}	0.14
rs407110999	7	22859168	G	0.25	1.37	8.71×10^{-7}	0.1
rs413147562	7	22798236	A	0.23	1.61	1.20×10^{-7}	0.71

β corresponds to the effect of SNP (in number of crossover per meiosis) on GRR and pQTL is the probability of the SNP to be a QTL estimated using a Bayesian Sparse Linear Mixed Model (see *Materials and Methods*). Chr, chromosome; GRR, genome-wide recombination rate.

recombination rates. We showed that contemporary patterns of recombination are highly correlated to the presence of historical hotspots. We showed that the recombination patterns along the genome are conserved between distantly related sheep populations but that their genetic determinism of genome-wide recombination rates differ. In particular, we showed that polymorphisms within the *RNF212* gene are strongly associated to male recombination in Lacaune, whereas this genomic region shows no association in Soay males. Hence, combining three data sets, two pedigree data sets in distantly related domestic sheep populations and a densely genotyped sample of unrelated animals, revealed that recombination rate and its genetic determinism can evolve at short timescales, as we discuss below.

Fine-scale recombination maps

In this work, we were able to construct fine-scale genetic maps of the sheep autosomes by combining two independent inferences on recombination rate. Our study on meiotic recombination from a large pedigree data set revealed that sheep recombination exhibits general patterns similar to other mammals (Shifman *et al.* 2006; Chowdhury *et al.* 2009; Tortereau *et al.* 2012). First, sheep recombination rates were elevated at the chromosome ends, both on acrocentric and metacentric chromosomes. In the latter, our analysis revealed a clear reduction in recombination near centromeres. Second, recombination rate depended on the chromosome physical size, consistent with an obligate crossover per meiosis irrespective of the chromosome size. These patterns were consistent with those established in a very different sheep population, the Soay (Johnston *et al.* 2016), and indeed when reanalyzing the Soay data with the same approach as used in this study, the results showed a striking similarity between recombination rates in the two populations. Hence, our results show that recombination patterns were conserved over many generations, despite the very different evolutionary histories of the two populations and clear differences in the genetic determinism of GRR in males of the two populations. This similarity allowed us to combine the two data sets to create more

precise male sheep recombination maps than any of the two studies taken independently.

Our historical recombination maps revealed patterns of recombination at the kilobase scale, with small, highly recombining intervals interspaced by more wide, low recombining regions. This result was consistent with the presence of recombination hotspots in the highly recombinant intervals. A consequence was that, as observed in other species, the majority of recombination took place in a small portion of the genome: we estimated that 80% of recombination takes place in 40% of the genome. Kaur and Rockman (2014) suggested the use of a Gini coefficient as a measure of the heterogeneity in the distribution of recombination along the genome to facilitate interspecies comparisons. When calculated on the historical recombination data, the Lacaune sheep has a coefficient of 0.52, which is similar to what is observed in *Drosophila* but lower than that measured in humans or mice. However, the coefficient calculated here is likely an underestimate due to our limited resolution (a few kilobases on the HD SNP array) compared to the typical hotspot width (a few hundred base pairs). Overall, we identified 50,000 hotspot intervals, which was twice the estimated number of hotspots in humans (International HapMap Consortium *et al.* 2007). This difference can be explained by different nonmutually exclusive reasons. First, it is possible that what we detect as crossover hotspots are due to genome assembly errors, and we indeed found a significant albeit moderate effect (odds ratio = 1.4) of the presence of assembly gaps in an interval on its probability of being called a hotspot. Second, our method to call hotspots could be too liberal. Indeed, a more stringent threshold (FDR = 0.1%) would lead to ~25,000 hotspots, which would be similar to what is found in humans. Third, selection has been shown to impact hotspot discoveries, although not with the methods that we used here (Chan *et al.* 2012). Finally, there exists the possibility that sheep historically exhibit more recombination hotspots than humans. In any case, the strong association between meiotic recombination rate and density in historical hotspots showed that our historical recombination maps were generally accurate. We tried to find enrichment in sequence motifs

Table 4 Association of GRR with mutations in the RNF212 gene

Mutation Name	Base Change	Positions on OA Musimon Genome (Scaffold 01089)	Predicted Positions on v3.1 Sheep Genome (OAR6)	p	β	P-Value	pQTL
RNF212_14431_AG	A > G	132229	116438514	0.18	-3.98	6.25×10^{-17}	0.23
RNF212_18411_GA	G > A	136209	116442624	0.17	-5.58	4.93×10^{-15}	0.02
RNF212_22570_CG	C > G	140368	116446753	0.18	-3.94	4.61×10^{-16}	0.09
RNF212_22594_AG	A > G	140392	116446777	0.17	0.57	0.54	0.004

Association of mutations in the RNF212 gene with GRR in 345 individuals. Positions on different reference sequences as well as predicted positions on OAR v3.1 are indicated. p: allele frequency β : allele substitution effect. pQTL: probability that the SNP is a QTL after fitting a Bayesian sparse linear mixed model on the region (see details in the text). OA, *O. aries* reference genome; GRR, genome-wide recombination rate.

in the detected hotspots or specify their position relative to TSS (data not shown), but with no success mainly due to (i) the relative large hotspot intervals (~5 kb) compared to typical hotspot motifs and (ii) the quality of the sheep genome assembly, which still contains many small gaps that make such analyses difficult. Ultimately, these questions would need an improved genome assembly and better resolution of crossover hotspots, which should be addressed in the future from LD-based studies on resequencing data.

We combined, using a formal statistical approach, meiotic and LD-based recombination rate estimates. Using an approach conceptually similar to that of O'Reilly *et al.* (2008) led us to assess the impact of selection events on the sheep genome, in particular suggesting the possibility of an effect of diversifying selection at olfactory receptor genes. Based on this comparison, the correlation between historical and meiotic recombination rates was found to be high (r 0.7), but less than could be expected from previous results in humans, where the correlation was 97% on 5 Mb (Myers *et al.* 2006). However, it was closer to that of worms (Rockman and Kruglyak 2009), mice (Brunschwig *et al.* 2012), or *Drosophila* (Chan *et al.* 2012), 69, 47, and 50%, respectively. Again, more precise estimates of both meiotic and historical recombination rates could change this number but other causes can be put forward.

A first explanation could come from the fact that the model we used to estimate historical recombination rates is based on the assumption of a constant effective population size, both in the past and along the genome. To allow for varying population size along the genome, we estimated the model in 2-Mb intervals, but there is still a possibility that varying population size in the past affected our historical recombination rate estimates, as the method has been shown to be somewhat influenced by demography, although much less so for the identification of crossover hotspots (Li and Stephens 2003). Also, as already mentioned above, selection has been shown to have a substantial impact on the estimation of recombination rates with other approaches (Chan *et al.* 2012), although it has not been evaluated for the Li and Stephens (2003) model to our knowledge.

Second, our meiotic recombination maps are based on male meioses only, while historical recombination rates are averaged over both male and female meioses. The fact that male and female recombination differ substantially, particularly in

sheep (Johnston *et al.* 2016), could also explain this relatively lower correlation.

Third, it is also possible that selective pressure due to domestication and later artificial breeding had the impact of modifying extensively LD patterns on the sheep genome, degrading the correlation between the two approaches. Indeed, the historical recombination estimates summarize ancestral recombinations that took place in the past and it is possible that recombination hotspots that were present in an ancestral sheep population are no longer active in today's Lacaune individuals. This could arise, for example, if domestication led to a reduction in the diversity of hotspots defining genes, such as *PRDM9*, and hence a reduction in the number of motifs underlying hotspots, which would in turn change the distribution of recombination on the genome. For example, this has been shown in humans, where patterns of recombination differ between populations due to their different diversity at *PRDM9* (Baudat *et al.* 2010; Berg *et al.* 2010, 2011). Eventually, such a phenomenon would degrade the correlation between present day recombination (measured by the meiotic recombination rates) and past recombination (measured by historical recombination rates). Further studies on the determinism of hotspots in sheep, their related genetic factors, and their diversity would be needed to elucidate this question.

Despite these different effects, the substantial correlation between meiotic and historical recombination rates motivates the creation of scaled recombination maps that can be useful for interpreting statistical analysis of genomic data. As an illustration of the importance of fine-scale recombination maps for genetic studies, we found an interesting example in a recent study on the adaptation of sheep and goats (Kim *et al.* 2016). In this study, a common signal of selection was found using the integrated Haplotype Score (iHS) statistic (Voight *et al.* 2006) in these two species (Figure 5 in Kim *et al.* (2016)). This signature precisely matches the low recombining regions that we identified on chromosome 10. However, the iHS statistic has been shown to be strongly influenced by variation in recombination rates, and in particular to tend to detect low recombining regions as selection signatures (O'Reilly *et al.* 2008; Ferrer-Admetlla *et al.* 2014). Precise genetic maps such as the one we provide in this work could thus help in annotating and interpreting such selection signals.

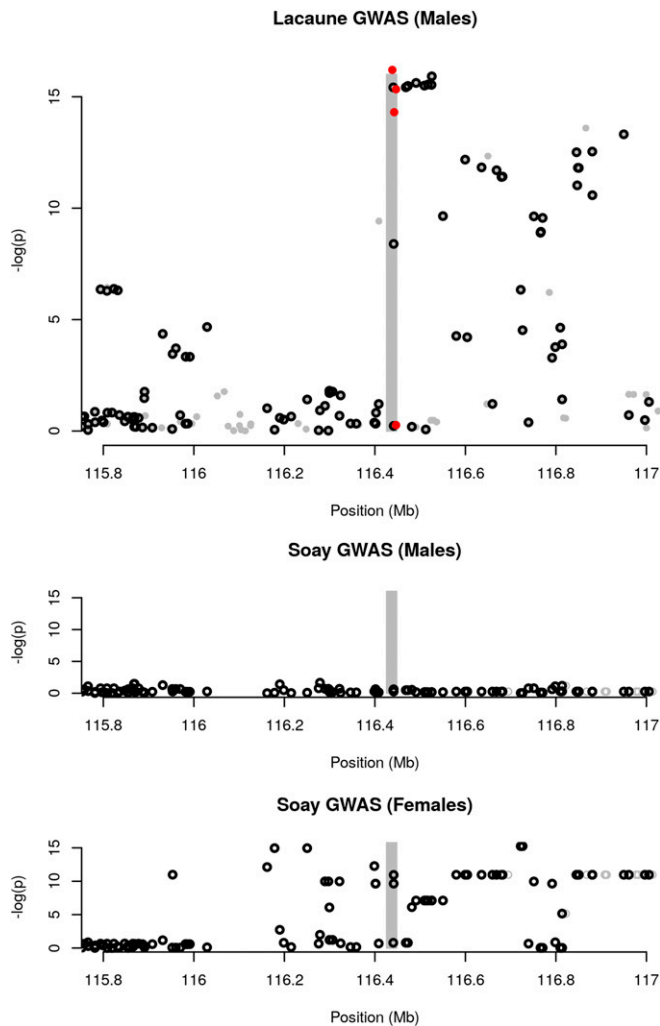


Figure 6 Comparison of GWAS results for the chromosome 6 QTL in Lacaune Males (top), Soay Males (middle), and Soay Females (bottom). The shaded area highlights the predicted position of the *RNF212* gene. Circle dots are markers tested in both populations. Red dots are the new mutations within the *RNF212* gene discovered in this study and genotyped in the Lacaune population. GWAS results in the Soay are from Johnston *et al.* (2016). GWAS, genome-wide association study.

Determinism of recombination rate in sheep populations

As mentioned in the introduction, two phenotypes have been studied with respect to the recombination process, but only one was studied here, GRR. We found that our data were not sufficient to study the Individual Hotspot Usage, which requires either a larger number of meioses per individual (Sandor *et al.* 2012; Ma *et al.* 2015; Kadri *et al.* 2016) or denser genotyping in families (Coop *et al.* 2008).

Our approach to study the genetic determinism of GRR in the Lacaune population was first to estimate its heritability, using a classical analysis in a large pedigree. This analysis also allowed us to extract additive genetic values (EBVs) for the trait in 345 male parents, which we used for a GWAS in a second step. The EBVs are, by definition, only determined by

genetic factors, as environmental effects on GRR are averaged out. Indeed, we found that the proportion of variance in EBVs explained by genetic factors in the GWAS was essentially one. A consequence was that, although this sample size could be deemed low by current standards, the power of our GWAS was greatly increased by the high precision on the phenotype. We estimated the heritability of GRR at 0.23, which was similar to estimates from studies on the same phenotype in ruminants [*e.g.*, 0.22 in cattle (Sandor *et al.* 2012) or 0.12 in male Soay sheep (Johnston *et al.* 2016), but see below for a discussion on the comparison with Soay sheep]. We had little information on the environmental factors that could influence recombination rate, but did find a suggestive effect of the month of insemination on GRR; in particular, we found increased GRR in the month of May. Confirmation and biological interpretation of this result would need dedicated studies, but it was consistent with the fact that fresh (*i.e.*, not frozen) semen is used for insemination in sheep and that the reproduction of this species is seasonal (Rosa and Bryant 2003).

The genetic determinism of GRR discovered in our study closely resembles what has been found in previous studies, especially in mammals. Two major loci and two suggestive ones affected the recombination rate in Lacaune sheep. The two main QTL are common to cattle and Soay sheep. The underlying genes and mutations for these two QTL are not yet resolved, but the fact that the two regions harbor interacting genes [*RNF212* and *HEI10* (Qiao *et al.* 2014; Rao *et al.* 2016)] involved in the maturation of crossovers makes these two genes likely functional candidates. Indeed, these two genes were identified as potential candidates underlying QTL for GRR in mice (Wang and Payseur 2017). The third gene identified here, *KCNJ15*, is a novel candidate, and its role and mechanism of action in the repair of DSBs needs to be confirmed and elucidated. Interestingly, these three genes are linked to the reparation of DSBs and crossover maturation processes. Finally, the fourth candidate *FSHR* has well-documented effects on gametogenesis but has not previously been linked to recombination.

In our study, 60% of the additive genetic variance in GRR remained unexplained by large-effect QTL and were due to polygenic effects. This could be interpreted in the light of recent evidence that has shown that other mechanisms, involved in chromosome conformation during meiosis, explain a substantial part of the variation in recombination rate between mouse strains (Baier *et al.* 2014) and bovids (Ruiz-Herrera *et al.* 2017). Furthermore, the variations at the major mammal recombination loci (*RNF212*, *CPLX1*, and *REC8* or the Human inversion *17q21.31*) explain only 3 to 11% (Ritz *et al.* 2017) of the phenotypic variance among individuals. Elucidating the genetic determinism of these different processes would thus require much larger sample sizes or different experimental approaches (Baier *et al.* 2014; Ruiz-Herrera *et al.* 2017).

The combination of data sets from the Lacaune population and one from the recent study of recombination in Soay sheep (Johnston *et al.* 2016) allowed us to study the evolution of

recombination at relatively short timescales. One of the most striking differences between our two studies is that the two QTL that were detected to be in common had no effect in Soay males, whereas they had strong effects in Lacaune males. However, the two populations had very similar polygenic heritability; accounting for the fact that the Lacaune QTL explain ~40% of the additive genetic variance, we could estimate the polygenic additive genetic variance in Lacaune males at 0.16, very similar to the 0.12 found in Soay males. Combined with our results that the two populations exhibit very similar male recombination maps, both in terms of intensity and genome distribution, the combination of the two studies shows that recombination patterns are conserved between populations under distinct genetic determinism, highlighting the robustness of mechanisms that drive them. We note that this robustness concerns recombination at a relatively broad scale (in the order of 10–100 kb), so it does not necessarily mean that the two breeds share recombination hotspots. Further work is needed to get a more detailed picture of the genetic control of recombination in sheep and will likely require the combination of multiple inferences from genetics, cytogenetics, molecular biology, and bioinformatics analyses.

Acknowledgments

Numerical analyses were performed on the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul). We are thankful to Tom Druet, Laurent Duret, Alain Pinton, Pierre Sourdille, Susan Johnston, and an anonymous reviewer for their helpful comments on an earlier version of the manuscript. Institut de l'Élevage (J.M.A.) and breed organizations (Ovitest and Confédération de Roquefort) provided the SNP genotypes and pedigree information. This work was partially funded by the BoDeliRe grant of the Institut National de la Recherche Agronomique Selgen Metaprogram and by Région Midi-Pyrénées.

Literature Cited

- Ahlawat, S., P. Sharma, R. Sharma, R. Arora, N. K. Verma *et al.*, 2016 Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep. *Anim. Genet.* 47: 740–751.
- Astruc, J.-M., G. Lagriffoul, C. Moreno-Romieux, M. Petit, and B. Servin, 2017 Raw genotyping data from: variation in recombination rate and its genetic determinism in sheep populations from combining multiple genomewide datasets. DOI: 10.5281/zenodo.804264. Available at: <https://doi.org/10.5281/zenodo.804264>. 10.5281/zenodo.804264
- Auton, A., Y. Rui Li, J. Kidd, K. Oliveira, J. Nadel *et al.*, 2013 Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genet.* 9: e1003984.
- Baier, B., P. Hunt, K. W. Broman, and T. Hassold, 2014 Variation in genome-wide levels of meiotic recombination is established at the onset of prophase in mammalian males. *PLoS Genet.* 10: e1004125.
- Baloche, G., A. Legarra, G. Sallé, H. Larroque, J.-M. Astruc *et al.*, 2014 Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J. Dairy Sci.* 97: 1107–1116.
- Bates, D., M. Mächler, B. Bolker, and S. Walker, 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67: 1–48.
- Baudat, F., J. Buard, C. Grey, A. Fledel-Alon, C. Ober *et al.*, 2010 PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327: 836–840.
- Baudat, F., Y. Imai, and B. de Massy, 2013 Meiotic recombination in mammals: localization and regulation. *Nat. Rev. Genet.* 14: 794–806.
- Berg, I. L., R. Neumann, K.-W. G. Lam, S. Sarbajna, L. Odenthal-Hesse *et al.*, 2010 PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.* 42: 859–863.
- Berg, I. L., R. Neumann, S. Sarbajna, L. Odenthal-Hesse, N. J. Butler *et al.*, 2011 Variants of the protein PRDM9 differentially regulate a set of human meiotic recombination hotspots highly active in African populations. *Proc. Natl. Acad. Sci. USA* 108: 12378–12383.
- Boitard, S., W. Rodríguez, F. Jay, S. Mona, and F. Austerlitz, 2016 Inferring population size history from large samples of genome-wide molecular data - an approximate Bayesian computation approach. *PLoS Genet.* 12: e1005877.
- Brunschwig, H., L. Levi, E. Ben-David, R. W. Williams, B. Yakir *et al.*, 2012 Fine-scale maps of recombination rates and hotspots in the mouse genome. *Genetics* 191: 757–764.
- Chan, A. H., P. A. Jenkins, and Y. S. Song, 2012 Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genet.* 8: e1003090.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.
- Cheung, V. G., J. T. Burdick, D. Hirschmann, and M. Morley, 2007 Polymorphic variation in human meiotic recombination. *Am. J. Hum. Genet.* 80: 526–530.
- Chowdhury, R., P. R. J. Bois, E. Feingold, S. L. Sherman, and V. G. Cheung, 2009 Genetic analysis of variation in human meiotic recombination. *PLoS Genet.* 5: e1000648.
- Cirulli, E. T., R. M. Kliman, and M. A. F. Noor, 2007 Fine-scale crossover rate heterogeneity in *Drosophila pseudoobscura*. *J. Mol. Evol.* 64: 129–135.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Looor, A. Everts-van der Wind *et al.*, 2005 Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15: 936–944.
- Coop, G., X. Wen, C. Ober, J. K. Pritchard, and M. Przeworski, 2008 High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* 319: 1395–1398.
- Cox, A., C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell *et al.*, 2009 A new standard genetic map for the laboratory mouse. *Genetics* 182: 1335–1344.
- Crawford, D. C., T. Bhangale, N. Li, G. Hellenthal, M. J. Rieder *et al.*, 2004 Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* 36: 700–706.
- Druet, T., and M. Georges, 2015 LINKPHASE3: an improved pedigree-based phasing algorithm robust to genotyping and map errors. *Bioinformatics* 31: 1677–1679.
- Fariello, M.-I., B. Servin, G. Tosser-Klopp, R. Rupp, C. Moreno International Sheep Genomics Consortium *et al.*, 2014 Selection signatures in worldwide sheep populations. *PLoS One* 9: e103813.
- Ferrer-Admetlla, A., M. Liang, T. Korneliussen, and R. Nielsen, 2014 On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* 31: 1275–1291.
- Fraley, C., and A. E. Raftery, 2002 Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* 97: 611–631.

- Fraley, C., A. E. Raftery, T. B. Murphy, and L. Scrucca, 2012 *N. d. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. University of Washington, Seattle, WA.
- Groenen, M. A. M., P. Wahlberg, M. Foglio, H. H. Cheng, H.-J. Megens *et al.*, 2009 A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res.* 19: 510–519.
- Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi *et al.*, 2012 Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393–398.
- Guan, Y., and M. Stephens, 2008 Practical issues in imputation-based association mapping. *PLoS Genet.* 4: e1000279.
- Hassold, T., H. Hall, and P. Hunt, 2007 The origin of human aneuploidy: where we have been, where we are going. *Hum. Mol. Genet.* 16: R203–R208.
- Howie, B. N., P. Donnelly, and J. Marchini, 2009 A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5: e1000529.
- Ignatieva, E. V., V. G. Levitsky, N. S. Yudin, M. P. Moshkin, and N. A. Kolchanov, 2014 Genetic basis of olfactory cognition: extremely high level of DNA sequence polymorphism in promoter regions of the human olfactory receptor genes revealed using the 1000 Genomes Project dataset. *Front. Psychol.* 5: 247.
- International HapMap Consortium, K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
- Jiang, Y., M. Xie, W. Chen, R. Talbot, J. F. Maddox *et al.*, 2014 The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344: 1168–1173.
- Johnston, S. E., J. Gratten, C. Berenos, J. G. Pilkington, T. H. Clutton-Brock *et al.*, 2013 Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* 502: 93–95.
- Johnston, S. E., C. Bérénos, J. Slate, and J. M. Pemberton, 2016 Conserved genetic architecture underlying individual recombination rate variation in a wild population of Soay sheep (*Ovis aries*). *Genetics* 203: 583–598.
- Johnston, S. E., J. Huisman, P. A. Ellis, and J. M. Pemberton, 2017 A high density linkage map reveals sexual dimorphism in recombination landscapes in red deer (*Cervus elaphus*). *G3 (Bethesda)* 7: 2859–2870.
- Kadri, N. K., C. Harland, P. Faux, N. Cambisano, L. Karim *et al.*, 2016 Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. *Genome Res.* 26: 1323–1332.
- Kari, V., A. Shcheteb, H. Neumann, and S. A. Johnsen, 2011 The H2B ubiquitin ligase RNF40 cooperates with SUPT16H to induce dynamic changes in chromatin structure during DNA double-strand break repair. *Cell Cycle* 10: 3495–3504.
- Kaur, T., and M. V. Rockman, 2014 Crossover heterogeneity in the absence of hotspots in *Caenorhabditis elegans*. *Genetics* 196: 137–148.
- Kijas, J., J. Lenstra, B. Hayes, S. Boitard, L. Porto Neto *et al.*, 2012 Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10: e1001258.
- Kim, E.-S., A. R. Elbeltagy, A. M. Aboul-Naga, B. Rischkowsky, B. Sayre *et al.*, 2016 Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116: 255–264.
- Kong, A., G. Thorleifsson, H. Stefansson, G. Masson, A. Helgason *et al.*, 2008 Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* 319: 1398–1401.
- Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson *et al.*, 2010 Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467: 1099–1103.
- Kong, A., G. Thorleifsson, M. L. Frigge, G. Masson, D. F. Gudbjartsson *et al.*, 2014 Common and low-frequency variants associated with genome-wide recombination rate. *Nat. Genet.* 46: 11–16.
- Lange, J., S. Yamada, S. E. Tischfield, J. Pan, S. Kim *et al.*, 2016 The landscape of mouse meiotic double-strand break formation, processing, and repair. *Cell* 167: 695–708.e16.
- Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome sequences. *Nature* 475: 493–496.
- Li, N., and M. Stephens, 2003 Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165: 2213–2233.
- Ma, L., J. R. O'Connell, P. M. VanRaden, B. Shen, A. Padhi *et al.*, 2015 Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet.* 11: e1005387.
- Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454: 479–485.
- McVean, G., P. Awadalla, and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231–1241.
- Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti, and J. Forejt, 2009 A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323: 373–375.
- Moreno-Romieux, C., F. Tortereau, J. Raoul, and B. Servin, 2017 High density genotypes of French sheep populations. DOI: 10.5281/zenodo.237116. Available at: <https://doi.org/10.5281/zenodo.237116>.
- Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310: 321–324.
- Myers, S., C. C. A. Spencer, A. Auton, L. Bottolo, C. Freeman *et al.*, 2006 The distribution and causes of meiotic recombination in the human genome. *Biochem. Soc. Trans.* 34: 526–530.
- Nagamine, Y., R. Pong-Wong, P. Navarro, V. Vitart, C. Hayward *et al.*, 2012 Localising loci underlying complex trait variation using regional genomic relationship mapping. *PLoS One* 7: e46501.
- Nagel, A. C., P. Fischer, J. Szawinski, M. K. La Rosa, and A. Preiss, 2012 Cyclin G is involved in meiotic recombination repair in *Drosophila melanogaster*. *J. Cell Sci.* 125: 5555–5563.
- Norris, B. J., and V. A. Whan, 2008 A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* 18: 1282–1293.
- O'Reilly, P. F., E. Birney, and D. J. Balding, 2008 Confounding between recombination and selection, and the Ped/Pop method for detecting selection. *Genome Res.* 18: 1304–1313.
- Petit, M., C. Moreno-Romieux, and B. Servin, 2017 Supplemental data to: variation in recombination rate and its genetic determinism in sheep (*Ovis aries*) populations from combining multiple genome-wide datasets. DOI: 10.5281/zenodo.821569. Available at: <https://doi.org/10.5281/zenodo.821569>.
- Pratto, F., K. Brick, P. Khil, F. Smagulova, G. V. Petukhova *et al.*, 2014 Recombination initiation maps of individual human genomes. *Science* 346: 1256442.
- Qiao, H., H. B. D. Prasada Rao, Y. Yang, J. H. Fong, J. M. Cloutier *et al.*, 2014 Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nat. Genet.* 46: 194–199.
- Rao, H. B. D. P., H. Qiao, S. K. Bhatt, L. R. J. Bailey, H. D. Tran *et al.*, 2017 A SUMO-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination. *Science* 355: 403–407.
- Reynolds, A., H. Qiao, Y. Yang, J. K. Chen, N. Jackson *et al.*, 2013 RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nat. Genet.* 45: 269–278.

- Ritz, K. R., M. A. F. Noor, and N. D. Singh, 2017 Variation in recombination rate: adaptive or not? *Trends Genet.* 33: 364–374.
- Rochus, C. M., F. Tortereau, F. Plisson-Petit, G. Restoux, C. Moreno-Romieux *et al.*, 2017 High density genome scan for selection signatures in French sheep reveals allelic heterogeneity and introgression at adaptive loci. *bioRxiv* DOI: <https://doi.org/10.1101/103010>.
- Rockman, M. V., and L. Kruglyak, 2009 Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* 5: e1000419.
- Rosa, H. J. D., and M. J. Bryant, 2003 Seasonality of reproduction in sheep. *Small Rumin. Res.* 48: 155–171.
- Ruiz-Herrera, A., M. Vozdova, J. Fernández, H. Sebestova, L. Capilla *et al.*, 2017 Recombination correlates with synaptonemal complex length and chromatin loop size in bovids—insights into mammalian meiotic chromosomal organization. *Chromosoma* DOI: 10.1007/s00412-016-0624-3.
- Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832–837.
- Sandor, C., W. Li, W. Coppieters, T. Druet, C. Charlier *et al.*, 2012 Genetic variants in REC8, RNF212, and PRDM9 influence male recombination in cattle. *PLoS Genet.* 8: e1002854.
- Scheet, P., and M. Stephens, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78: 629–644.
- Servin, B., and M. Stephens, 2007 Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* 3: e114.
- Shifman, S., J. T. Bell, R. R. Copley, M. S. Taylor, R. W. Williams *et al.*, 2006 A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol.* 4: e395.
- Ślabicki, M., M. Theis, D. B. Krastev, S. Samsonov, E. Mundwiller *et al.*, 2010 A genome-scale DNA repair RNAi screen identifies SPG48 as a novel gene associated with hereditary spastic paraplegia. *PLoS Biol.* 8: e1000408.
- Stathopoulos, S., J. M. Bishop, and C. O’Ryan, 2014 Genetic signatures for enhanced olfaction in the African mole-rats. *PLoS One* 9: e93336.
- Stephens, M., 2017 False discovery rates: a new deal. *Biostatistics* 18: 275–294.
- Stevison, L. S., and M. A. F. Noor, 2010 Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. *J. Mol. Evol.* 71: 332–345.
- Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100: 9440–9445.
- Sturtevant, A. H., 1913 The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* 14: 43–59.
- Takasuga, A., 2016 PLAG1 and NCAPG-LCORL in livestock. *Anim. Sci. J.* 87: 159–167.
- Tapanainen, J. S., K. Aittomäki, J. Min, T. Vaskivuo, and I. T. Huhtaniemi, 1997 Men homozygous for an inactivating mutation of the follicle-stimulating hormone (FSH) receptor gene present variable suppression of spermatogenesis and fertility. *Nat. Genet.* 15: 205–206.
- Tortereau, F., B. Servin, L. Frantz, H.-J. Megens, D. Milan *et al.*, 2012 A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* 13: 586.
- Voight, B. F., S. Kudravalli, X. Wen, and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. *PLoS Biol.* 4: e72.
- Wang, J., H. C. Fan, B. Behr, and S. R. Quake, 2012 Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150: 402–412.
- Wang, R. J., and B. A. Payseur, 2017 Genetics of genome-wide recombination rate evolution in mice from an isolated Island. *Genetics* 206: 1841–1852.
- Zeileis, A., and G. Grothendieck, 2005 Zoo: S3 infrastructure for regular and irregular time series. *J. Stat. Softw.* 14: 1–27.
- Zhou, X., and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44: 821–824.
- Zhou, X., P. Carbonetto, and M. Stephens, 2013 Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9: e1003264.

Communicating editor: B. Payseur