Genome landscape and genetic architecture of recombination in domestic goats (*Capra Hircus*).

Alice Etourneau[1*], Rachel Rupp[1], Bertrand Servin[1*]

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, 31320 Castanet-Tolosan, France

*Corresponding Authors

E-mail addresses:

AE: alice.etourneau@inrae.fr

RR: rachel.rupp@inrae.fr

BS: bertrand.servin@inrae.fr

Abstract

Background

Recombination is a fundamental biological process, both in participating to the creation of viable gametes and as a driver of genetic diversity. Characterising recombination is therefore of strong interest in breeding populations. In this study, we used ~50K genotyped data and pedigree from two French populations (Alpine and Saanen) of domestic goats (*Capra hircus*) to build sex-specific recombination maps, and to explore the genetic determinism of two recombination phenotypes: genome-wide recombination rate (GRR) and intra-chromosomic shuffling.

Results

Sex-specific recombination maps showed higher recombination in males than females for both breeds (Alpine autosomal map size = 35.1M in males and 30.5M in females; Saanen map size = 34.0M in males and 29.0M in females). Heterochiasmy is particularly notable on small chromosomes, and at both ends of the chromosomes. Yet, no difference in shuffling has been found between populations. Genetic parameters on recombination phenotypes could only be estimated in males, due to lack of data in females. Both phenotypes are significantly heritable (h^2=0.12 (0.03) for GRR and h^2=0.034 (0.015) for shuffling, when pooling breeds). GWAS on male GRR identified several significant loci, including RNF212, RNF212B and SSH1, the last one being a novel locus for this phenotype. Correlation of SNP effects between breeds is low for both recombination phenotypes (0.25 for GRR and 0.04 for shuffling), indicating different genetic determinism in the two breeds.

Conclusions

This study contributes to the understanding of recombination evolution in ruminants, both between breeds and species.

Background

Recombination is a fundamental biological process in eukaryotes. It is necessary for the proper segregation of chromosomes during meiosis (Koehler et al., 1996) and is also a great driver of genetic diversity by creating new combinations of alleles and avoiding the accumulation of deleterious mutations over time (Muller, 1964; Crow and Kimura, 1965). Studying recombination can therefore be important both to understand functional reproduction in a species, and to study evolutionary drivers of genetic diversity.

Of the different approaches to study recombination, pedigree-based studies, as opposed to population studies for instance, have several advantages: first they provide a direct measure of meiotic recombination, essentially unaffected by evolutionary pressures, through the transmission of gametes in the pedigree and second, they allow to assign crossovers to specific individuals and therefore estimate the genetic determinism of the recombination process. Pioneered in humans (Kong et al. (2002) , Coop et al. (2008)), the increased availability of cheap genotyping tools has led to a large number of recombination studies, especially in livestock. Indeed, the implementation of genotyping routines within genomic selection programs put at the disposal of researchers both the tools and the data to study recombination in pedigrees. As such, multiple studies were conducted in cattle (Sandor et al., 2012; Weng et al., 2014; Ma et al., 2015; Kadri et al., 2016; Shen et al., 2018; Brekke et al., 2023a), but also in sheep (Petit et al., 2017), chicken (Weng et al., 2019), pigs (Johnsson et al., 2021; Lozada-Soto et al., 2021; Brekke et al., 2022; Brekke et al., 2025) and Atlantic salmon (Brekke et al., 2023b). Apart from livestock, recombination in several natural populations with available pedigrees have also been studied: Soay sheep (Johnston et al., 2016), red deer (Johnston et al., 2017; Johnston et al., 2018) and wild house sparrow (McAuley et al., 2024).

This approach of characterising recombination through the use of genotypes and pedigree data allows to estimate recombination rates and build recombination maps of a species. Results have shown that heterochiasmy, or sex-differences in recombination, is widespread in many clades, but its sign varies: in some species, females recombine more than males while in others, males recombine more than females, as illustrated in fishes and birds (Cooney et al., 2021; McAuley et al., 2024).

This high diversity in recombination features shows the relevance of studying the recombination process itself. One way to study this process is through genetics, which allows

to estimate heritability and find associated QTLs for phenotypes such as recombination rate and, more recently, intra-chromosomic shuffling (Veller et al., 2019). The genetic control of recombination rate, despite showing common basis between species, was also shown to be very diverse, even between several populations or the two sexes of the same species. Focusing only on ruminants, heritability for phenotypes of recombination intensity varies between 0.04 in Norwegian Red cattle (Brekke et al., 2023a) and 0.22 in Dutch Holstein-Friesian cattle (Sandor et al., 2012). But variations in heritability can be observed between species, breeds within a species, and even sexes (Johnston et al., 2016; Kadri et al., 2016; Petit et al., 2017; Johnston et al., 2018). Identified QTL for these phenotypes can also vary between populations and species. Cattle (Sandor et al., 2012; Ma et al., 2015; Kadri et al., 2016; Shen et al., 2018; Brekke et al., 2023a), sheep (Johnston et al., 2016; Petit et al., 2017) and red deer (Johnston et al., 2018) all show a signal in the region of the genes RNF212B/HEI10/REC8/REC114, four genes in very close proximity to each other. Cattle and sheep also both have signal on RNF212. Yet despite these similarities, genes like MSH4 or PRDM9 have only been detected in cattle among ruminants, although this is possibly due to the increased statistical power in this species. When studying genetic control within sexes and breeds separately in a species, the differences are even starker. In sheep, Johnston et al. (2016) found a QTL at RNF212 only in Soay females, and not in males. Yet Petit et al. (2017) actually found a QTL for this same gene in males from another population, the Lacaune. In cattle, Ma et al. (2015) identified thirteen loci associated with recombination intensity, and only three of them were shared between sexes. In Kadri et al. (2016)'s bovine population, four of their ten detected QTLs are sex-specific. And once again, Brekke et al. (2023a) found sex differences in Norwegian red dairy cattle.

In this context of strong genetic diversity in very related groups, we present a new study on another ruminant species: the goat. In goats, the last genetic map was built with a linkage methodology in 1998 (Schibler et al., 1998). At the time, its primary goal was still to understand

the organisation of genes between one another, and the map was never updated since. Genome-wide recombination rate was estimated through the use of cytogenetics techniques by counting MLH1 foci (Fröhlich et al., 2015; Muñoz-Fuentes et al., 2015) but never confirmed using other methods.

In 2010, the first reference genome assembly was built for the domestic goat (Dong et al., 2013) and in 2013, an international goat genome consortium was created to acquire more genomic data and investigate genetic variations in the species. This lead to the development of a 50k SNP chip (Tosser-Klopp et al., 2014) and later on, the creation of a sequence-level SNP panel of more than a thousand goats from all over the world (Denoyelle et al., 2021).

In France, a goat breeding program for two breeds, Alpine and Saanen, has been in place since the 1960s, its main goal being to improve milk production for the purpose of cheese-making with the progressive inclusion of health and udder morphology characteristics. Since 2014 and the availability of the 50k SNP chip, selection candidates are now systematically genotyped. Furthermore, despite the fact that the breeding program is male-oriented, females have started to be genotyped as well in the last few years, both as part of the breeding program and within experimental programs. Therefore, we took advantage of genotyping and pedigree data available in the two Alpine and Saanen breeds to study recombination in this ruminant species using a pedigree-based approach, for the first time to our knowledge. In an opportunity to examine heterochiasmy and breed differences in this new species, we built updated, sex-specific recombination maps, developing novel methodologies to quantitatively measure sex differences in recombination patterns and take into account uncertainty about crossover detection. In a second time, we also studied the genetic determinism of two recombination phenotypes in both available breeds, by estimating genetic parameters and performing genome-wide association studies.

Methods

Study population and genotype data

In this work, we exploited pedigree and genotyping data of 7,588 goats born between 2008 and 2019. They amount to 3,169 males and 4,419 females from both the Alpine (5,051 samples) and Saanen (2,537 samples) breeds. The animals were genotyped with versions 1 and 2 of the Illumina Inc. GoatSNP50 chip. Only autosomal SNPs and markers common to both arrays were kept, resulting in 53,347 SNPs.

Quality control of physical marker ordering

Errors in the physical ordering of markers can lead to detect spurious recombination events. Recombination studies therefore require a robust marker ordering. Here, this ordering was built from two independent physical maps of the SNPs: the reference genome assembly ARS1 and the Radiation Hybrid (RH) map of Du et al., (2014). We decided to keep only SNP markers that have a consistent ordering on both physical maps. This was done by identifying, for each chromosome, the longest increasing subsequence (LIS, Schensted, 1961) in the mapping of the RH order to the assembly order. This requires knowing the order of the SNP markers on both maps. For the reference genome ordering, genomic positions of the 53,347 v1 autosomal markers of the GoatSNP50 Illumina chip were obtained from the ARS1 assembly as described on the International Goat Genome website (http://www.goatgenome.org/projects.html). For the RH map, additional analysis was required as it was built using SNP array of different species (ovine and bovine SNPs herein called external SNPs). We mapped the external SNPs to the ARS1 reference genome and then projected the position of each caprine SNP on the RH map from the positions of its two flanking external SNPs. For goat markers located at the extremities of the chromosome, which had only one flanking external SNPs and could not be projected on the RH map, we assumed the reference genome order was correct (but see below for further data quality control steps).

Figure S1 (**Additional File 1**) illustrates this quality control step on two chromosomes. After this step, 54,013 markers were kept.

Furthermore, an initial detection of crossovers on the genome (using methods detailed below) revealed 7 genomic regions (123 SNP markers in total) that showed extremely high recombination rates (see example on chromosome 1 on **Figure S2**, **Additional File 1**). Detailed analysis of the synteny conservation with the human genome in these regions showed that they corresponded systematically to conserved synteny breakpoints. As these events were most probably some remaining unidentified assembly errors, markers involved were discarded in downstream analysis. A list of these markers is given in **Additional File 2**.

Genotype quality control

Standard data cleaning procedures were applied using PLINK 1.9 (Chang et al., 2015). Animals and SNPs with call-rate below 99% were removed (**Figure S3**, **Additional File 1**). Monomorphic SNPs in the dataset were removed. No MAF filter was applied at this step as even a small MAF may be informative for crossover detection in families polymorphic for these markers. Markers with strong departures from Hardy-Weinberg genotype proportions ($p < 10^{-7}$) within each breed were removed. Note that families with a sire having more than 50 offspring were not used to compute the Hardy-Weinberg equilibrium (HWE) test as fathers carrying rare alleles in such designs can have a great impact on testing for HWE.

Mendelian errors were tested using the *mendel* command of the Yapp software version 0.3 (Servin, 2024). It evaluated the number of incoherent markers between parent and offspring and automatically corrected the kinship relationships in the pedigree which were beyond the threshold fixed by the software.

After these quality control steps, 46,081 markers and 7,457 animals, including 4,520 Alpine and 2,937 Saanen, were left.

Data phasing and crossover detection

Genotype phasing and crossover detection were done using Yapp version 0.4. First, the *phase* command was used in order to phase genotypes. In order to reconstruct the paternally and maternally inherited haplotypes of each individual (i.e. to infer the phase of heterozygous loci), Yapp uses a set of discrete steps. First, using all available genotypes, it identifies the unambiguous transmissions from parents to offspring, *e.g.* the alleles transmitted by a homozygous parent or received by a homozygous offspring are easily phased. This leads to a set of partially resolved phases. Second, for each parent in the pedigree, from oldest to newest, it improves the inference of parental phases by combining all partially resolved transmitted haplotypes using the algorithm of Favier et al. (2010). Third, after this first pass in the pedigree, the transmitted haplotypes from parent to offspring are improved using a Hidden Markov Model along the whole chromosome similar to the approach of Druet and Georges (2015) and segregation indicators are inferred. Then, the *recomb* command was used among a total of 6,434 meioses to detect crossovers. A crossover in the parent meiosis is defined as a change in phase observed in the offspring, on the inherited chromosome from the parent. Crossovers are detected between two heterozygous SNPs and can only be detected between the first and the last phased, heterozygous SNPs.

Meioses and crossovers were quality-controlled to ensure that only truly informative data was retained. All meioses with a parent or offspring phased at less than 90% of markers were removed. Furthermore, meioses with a parent that itself had none of its parents genotyped and only one offspring, were removed, as it is impossible to detect crossovers in such a setting.

Distance between two crossovers on the same chromosome was calculated for every meiosis. As crossovers are detected within an interval, distance between crossovers was calculated as the distance between the closest interval boundaries. All crossovers closer than 5 Mb from one another were considered double-crossovers, false-positives due to phasing errors

and removed (**Figure S4**, **Additional File 1**). After filtering, the dataset contained 6,204 meioses and 204,727 crossovers.

Recombination maps

Recombination maps were constructed by estimating recombination rates in non-overlapping intervals of one megabase (Mb) along the genome. The methodology for estimating recombination rates was adapted from Petit et al., (2017). For a given genomic interval i and a meiosis m, the count of the number of crossovers Y_{mi} is modelled using a Poisson distribution whose parameter (measured in Morgans) depends on the recombination rate c_i (in centiMorgans per Mb, cM/Mb) assumed constant in the interval and across meioses, and the length of the genomic interval l_{mi} (in Mb) that can depend on the meiosis (see below).

Then, the expected number of crossovers in an interval is $0.01 c_i l_{mi}$ and Y_{mi} is modelled as:

$$Y_{mi}|c_i \sim Poisson(0.01 c_i l_{mi}) \quad (1)$$

As in Petit et al. (2017), we define a prior on c_i as:

$$c_i \sim \Gamma(\alpha, \beta) \quad (2)$$

As in Petit et al. (2017), we used an empirical Bayes approach to set α and β from the genome-wide distribution of Maximum Likelihood Estimates of recombination rates ($\alpha = 3.02$ and $\beta = 2.32$, **Figure S5**, **Additional File 1**). Given the conjugacy of Poisson and Gamma distributions, the posterior distribution for c_i is then:

$$c_i|Y_{\cdot i} \sim \Gamma(\alpha + \sum_m Y_{mi}, \beta + 0.01 \sum_m l_{mi})$$

(3)

Stemming from the Bayesian model (3), the expected value for the Gamma law was used to estimate the chromosomic recombination rate:

$$\widehat{c_{chr}} = \frac{\alpha + \sum_m Y_{chr,m}}{\beta + \sum_m \mathrm{l}_{\mathrm{chr,m}}}$$

(4)

With α and β the prior parameters and m a meiosis. c_{chr} was estimated for both breeds and sexes, and a model explaining c_{chr} by the log of the chromosome size was fitted for all populations.

Meiosis Informativity

In the equations above, the length of the interval is allowed to depend on the interval and the meiosis considered. This is to account for the variation in the informativity of the data for the detection of crossovers. First, because samples are genotyped on a 50k SNP chip, no phasing information is available for the tips of chromosome before and after the first and the last SNP of the chromosome, which makes detecting crossovers impossible in these regions. Second, for a given meiosis, only SNPs that are heterozygous and phased in the parent are informative for crossover detection. Hence for a particular meiosis, this can further reduce the portion of the genome where crossovers can be detected (**Figure 1**).

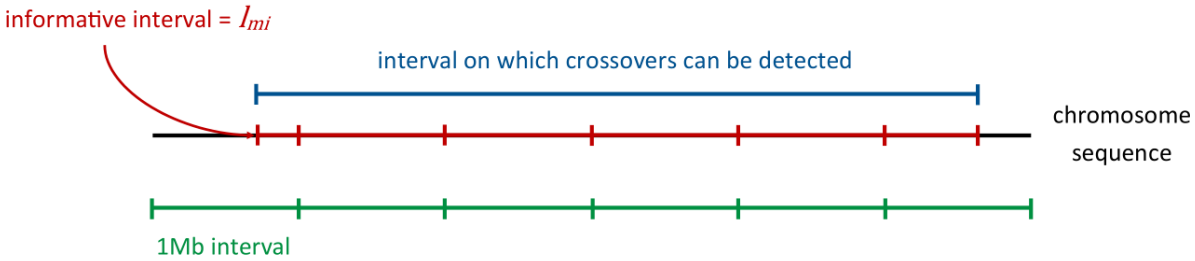

Figure 1: Illustration of the notion of interval informativity.

Accounting for uncertainty in crossover locations

Crossovers are only resolved within genome regions that vary in size depending on the parent and offspring genotypes. The average length of these intervals in our data is 1,4Mb with a 95% range of 0.1Mb to 10.2Mb (**Figure S6**, **Additional File 1**). To account for this uncertainty in the position of crossovers, we used a Monte Carlo approach as in Petit et al (2017): we first sampled crossover positions uniformly in the detection intervals and then the recombination rate from its posterior (Equation 4). This was done 100 times, providing for each interval 100 samples from the posterior distribution of recombination rate.

Let us name $E(c_i|Y_{.ik}) = b_{ik}$ the expected value of posterior k for genomic interval i. The expected value of all combined posteriors for interval i is therefore:

$$\widehat{c_i} = E(c_i|Y_{.i.}) = E_k\big(E(c_i|Y_{.ik})\big) = E_k(b_{ik})$$

(5)

And according to posterior variance law:

$$Var(\widehat{c_i}) = E\big(Var_k(c_i|Y_{.ik})\big) = Var(c_i) - Var_k(E(c_i|Y_{.ik}) = Var(c_i) - Var_k(b_{ik})$$

(6)

Thus, recombination rate c for genomic interval i was estimated as the mean of the 100 posteriors' expected values.

Assessing sex differences in local recombination rates

The Bayesian model described above is interesting in that it allows to account naturally for the uncertainties in the data and provides estimates of recombination rates even in regions with little information. Using this approach, we could in principle evaluate the sex differences in recombination rates from the posterior samples obtained by applying the model separately in both sexes. However, when interrogating the difference in local recombination rates between sexes, the model does not allow to separate the effect of differences in rates and those of differences in uncertainty. Therefore, the two sexes could be deemed to have different rates due to the amount of data in each sex being unbalanced (for example female estimates are typically more influenced by the prior than male estimates). To deal with this issue, we derived a permutation scheme that can separate both effects.

Our measure of heterochiasmy in an interval i is:

$$d_i = log2(c_{Mi}) - log2(c_{Fi})$$

(7)

That is, the log-ratio of male to female recombination (in base 2 for easier interpretation).

To account for the difference in data informativity between sexes, we permuted the sex of the parents 1,000 times to obtain a sample of d_i under the null of no difference between sexes. As permutations do not change the number of male and female meioses, they allow to account for difference in informativity between sexes. The resulting distribution of d_i in the permutation samples were very close to normally distributed (**Figure S7**, **Additional File 1**),

so in each interval the null distribution of d_i can be described by a normal distribution with parameters set at the sample mean and variance. This in turn allows to derive p-values for a difference between sexes in each interval. Finally, to account for multiple testing across intervals, the FDR approach of Storey and Tibshirani (2003) using the R "qvalue" package was applied on the p-values and the significance threshold set at 0.05.

Recombination phenotypes

The crossover dataset described above offers the opportunity to study the individual variation in some aspects of recombination (i.e. recombination phenotypes) and interrogate their genetic architectures. However, because we had very little data in females this was only possible to do in males. Two phenotypes were studied: the individual Genome-wide Recombination Rate (GRR) and the intra-chromosomic shuffling.

The intensity of recombination in a meiosis can be measured by summing the crossover count on all autosomes (Autosomal crossover count (ACC)). However, as explained above, the informativity of marker data varies between meioses and can influence the value of ACC. To take this into account, rather than the ACC as a measure of recombination intensity we use the genome-wide recombination rate (GRR), defined as the ACC divided by the informative portion of the genome (in Mb).

The other phenotype we studied is a measure of the efficacy of a meiosis to create new allelic combinations in the offspring: the rate of intra-chromosomic shuffling $\bar{r}_{intra}$ (Veller et al., 2019). It is defined as the probability that two randomly chosen loci uncouple their alleles during meiosis and is calculated as:

$$\bar{r}_{intra} = \sum_{k=1}^{n} 2p_k(1-p_k)L_k^2 \qquad (8)$$

Where for chromosome k, p_k is the proportion of SNPs inherited from one grandparent, L_k is the chromosome's proportion in the autosomal genome, and n is the number of autosomes.

Heritabilities

Heritabilities of the GRR and $\bar{r}_{intra}$ phenotypes were estimated using all males with a calculated phenotype, i.e. 520 parents accounting for 5,707 meioses (**Table 1** for detail).

To estimate heritability, variance components were estimated by restricted maximum likelihood (REML) using the Wombat software (Meyer, 2007). An animal model including a permanent environment effect was fitted:

$$Y \sim X\beta + Za + Zu + e \quad (9)$$

Y is the vector of observations (GRR or $\bar{r}_{intra}$); β is the vector of fixed effects, a is the vector of direct animal (additive) genetic effects; $\boldsymbol{u}$ is the vector of animal permanent effect; e is the vector of residual effects; $a \sim \mathcal{N}(0, A\sigma_A^2)$, $u \sim \mathcal{N}(0, I\sigma_B^2)$ and X, Z are the incidence matrixes. The pedigree matrix A was computed on five generations of the genealogy. The model was applied to all breeds separately or to the entire population at once. A fixed effect for breed was therefore added where necessary.

Once the variance components had been estimated, values were incorporated into the model as fixed values to obtain the Estimated Breeding Values (EBV) for each phenotype. Due to the absence of differences in estimated heritability between breeds (see Results), EBVs were estimated by pooling samples from both breeds. Note that although we did not fit a model for female recombination phenotypes, EBVs can still be predicted for females as they are connected to males in the pedigree.

Genome-wide association studies

Genotype Imputation

As an imputation reference panel, we used the whole-genome sequence VarGoats data presented in (Denoyelle et al., 2021) and retrieved from https://www.goatgenome.org/home.html. This data consisted of 1,372 goats out of which 252 European samples closely related to our study samples were used (from the Alpine, Fossés, Lorraine, Poitevine, Provençale, Pyrénées, Rove, Savoie and Saanen breeds).

SNP filtering was then applied on the resulting VCF based on several criteria, using bcftools (http://www.htslib.org/doc/1.1/bcftools.html). First, we removed SNPs having discordant reference alleles between the SNP array manifest and the resequencing data, and SNPs showing strictly less than two counts of the minor allele across all selected samples or with a proportion of missing genotypes greater than 10%. Then, SNPs were filtered based on the distribution of the following GATK 3.6 (Van der Auwera and O'Connor, 2020) variant-calling quality parameters: DP, FS, VQSLOD (**Figure S8**, **Additional File 1**). The chosen threshold for SNP removal were DP < 10,000; DP > 25,000; FS > 50; VQSLOD < 10 and VQSLOD > 25. The resulting number of SNPs remaining was 22,741,231 out of which 43,358 SNPs are present on the V1 GoatSNP50 Illumina chip. This corresponds to a drop rate of 5,6% for the chip SNPs.

The Beagle4 software (Browning and Browning, 2016) was used to impute the few missing genotypes of the sequence panel and correct the called genotypes by using the genotype likelihood (*gl*) method, a modelscale of 1.5 (Pook et al., 2020), an effective population size of 1000 and the recombination maps estimated from pedigree data *(see above)*. To assess the quality of the resulting panel genotypes, we leveraged information on 91 animals for which both sequencing and SNP genotyping data were available: we computed R^2 and genotype

concordance before and after imputation. Both metrics were highly improved by the imputation procedure: mean R2 increased from 0.928 to 0.990 and mean genotype concordance from 0.974 to 0.997. (**Figures S9** & **S10**, **Additional File 1**). Phasing of the reference panel was done using ShapeIt4 (Delaneau et al., 2019). As some animals from the reference panel were previously phased using pedigree information, their phased haplotypes for the SNP chip markers were used as scaffold reference for phasing with ShapeIt4. The resulting VCF file, phased and imputed, was used as reference panel for imputation.

Shapeit4 was also run on the phased VCF of 7,457 samples produced by Yapp from the SNP array data to solve the few unphased genotypes and impute the few remaining missing genotypes. The resulting fully phased 50k chip data was imputed to sequence using the Impute5 software (Rubinacci et al., 2020). Recombination maps were provided to increase power and N_e was set at 1000. Impute5 provides a measure of confidence for all imputed markers (imputation R2) which was used in later steps to filter SNPs *(see below).*

Statistical Analysis

GWAS were performed separately in each breed using the predicted EBV for male GRR and $\bar{r}_{intra}$ as phenotypes: 589 samples were used in Alpine and 254 in Saanen (**Table 1**). SNPs with MAF<5% and imputation R2<0.3 were removed from the analysis leaving N=15,988,092 and 15,734,503 SNPs in Alpine and Saanen respectively. A univariate linear mixed model was used for testing SNP association with the phenotype using the Gemma software (Zhou and Stephens, 2012), with the "leave one chromosome out" (LOCO) strategy (*i.e.* the genetic relationship matrix (GRM) used to analyse a chromosome is computed on all other chromosomes except this one).

Significance was assessed using the ash method (Stephens, 2016) implemented in the "ashr" R package. As this method takes into account the β estimates as well as their standard

errors, it can induce a reranking of SNPs compared to when only taking their p-values into account. All SNPs with s-value<0.01 were declared significant and SNPs with s-value<0.05 were declared suggestive.

Genetic correlation between breeds were measured by the correlation of SNP effect estimates ($\hat{\beta}$) for SNPs that were significant in at least one of the two breeds and using standard errors as weights.

Annotation of GWAS loci

A GWAS locus was defined from pairwise Linkage Disequilibrium (R2) between all significant SNPs on a given chromosome. R2 calculation was done using the PLINK 1.9 software. Significant SNPs were considered to belong to the same peak if their R2 was equal or superior to 0.2. For each lead SNP of each peak, a high LD area was defined as follows: an area whose boundaries correspond to the SNP furthest away in R2>0.6 from the lead variant, on both sides of the lead variant. We aimed at evaluating the evidence for association of all genes present in Ensembl (gff3 release 111). As this annotation did not include two major genes associated to the recombination process, RNF212 and PRDM9, we manually added them to the annotation file after blasting their bovine gene sequence on ARS1 (**Additional File 3**).

For all genes, we evaluated the following criteria: if the gene is within 2Mb of the lead variant; if it is the closest gene to the right, to the left or both; if it is the closest gene to another significant SNP of the peak and if it belongs within the “R2>0.6” area of the peak. Based on this we calculated a “candidate score”, ranging from 0 to 5, as the number of criteria satisfied by the gene.

Results

Recombination landscape

After quality control, genotype data on 7,457 animals led to the detection of 204,727 crossovers in 6,204 meioses from 843 parents. As a large portion of the dataset comes from a dairy selection program, most genotyped parents were males. As males also typically have more genotyped offspring than females, there are 10 times more male crossovers than female ones in the recombination data (**Table 1**). The software used to detect crossovers provides the information on the parts of the genome that are informative for the detection of crossovers (See Methods). Overall, all meioses had more than 94% of the genome informative for crossovers (**Figure S11**, **Additional File 1**). Females had slightly more informative meioses than males (mean=0.99% vs 0.98%) because the criteria used to filter meioses resulted in more stringency for keeping female meioses than male meioses, which are overall less connected in the pedigree.

Table 1: Summary table of data after all genotype and crossover quality control steps.

Breed	Sex	Count	Headcounts	Parents	Meiosis count	Crossover count
Alpine	M		1,739	327	3,562	120,057
	F		2,781	262	386	11,329
Saanen	M		1,339	193	2,145	70,307
	F		1,598	61	111	3,050
Total			7,457	843	6,204	204,727

Chromosome wide rates

A linear model was used to test differences in genetic map size between populations. Breed and sex parameters were found significant (sex effect p-value = 10^{-61}; breed effect p-value = 10^{-13}; interaction p-value = 0.52). The size of the genetic map was estimated at 35.1M for Alpine males, 34.0M for Saanen males, 30.5M for Alpine females and 29.0M for Saanen females (**Figure S12**, **Additional File 1**). Therefore, the genetic map is slightly shorter in Saanen than in Alpine and much shorter in females than males.

The mean recombination rate of each chromosome was estimated with the Bayesian model, and a linear model depending on the log size of the chromosomes was fit on the estimates (**Figure 2A**). Again, a significant difference between breeds was found (p-value = 0.01) but, as the effect is very small, the figure does not discriminate between breeds for clarity reasons. A significant difference between sexes is also observed (p-value $\approx 10^{-48}$): males overall have a higher chromosomic recombination rate than females, and the recombination is greater on small chromosomes than big chromosomes for both sexes. Overall, the chromosomic recombination rate has a log relation with the chromosome size. It is observed that the difference in recombination between the sexes is greater on small chromosomes than on big chromosomes, on which recombination rates are almost alike.

Chromosome landscape was calculated from 1Mb resolution Bayesian estimates of recombination rate: all positions in Mb were converted to positions relative to the centromere, and recombination rates of all chromosomes were studied depending on their position to the centromere. The results are presented in **Figure 2B**. Females have a much flatter recombination pattern along the chromosomes compared to males, and both sexes seem to recombine as much in the middle of chromosomes.

Local recombination rates

Recombination rates were estimated locally on 1Mb intervals on all autosomes (**Additional File 4**). Differences between sexes was estimated as the log ratio of male to female recombination rate and significance was assessed using its empirical null distribution calculated from permutations (see Methods). Recombination maps for each chromosome are available in **Additional File 5** and **Figure 2C** illustrates this inference on chromosome 25. It exhibits the general trend of the sex differences in recombination patterns where males recombine significantly more than females at the tips of the chromosomes. A total of 871 intervals of the genome are significant for sex differences (out of 2477 intervals). Of the significant intervals, 559 (64% of the significant intervals) are male biased and localized for the most part at the ends of chromosomes. The 312 remaining significant intervals (36% of the significant intervals) are female biased and are located in the middle of chromosomes. The list of all significant intervals can be found in **Additional File 6**.

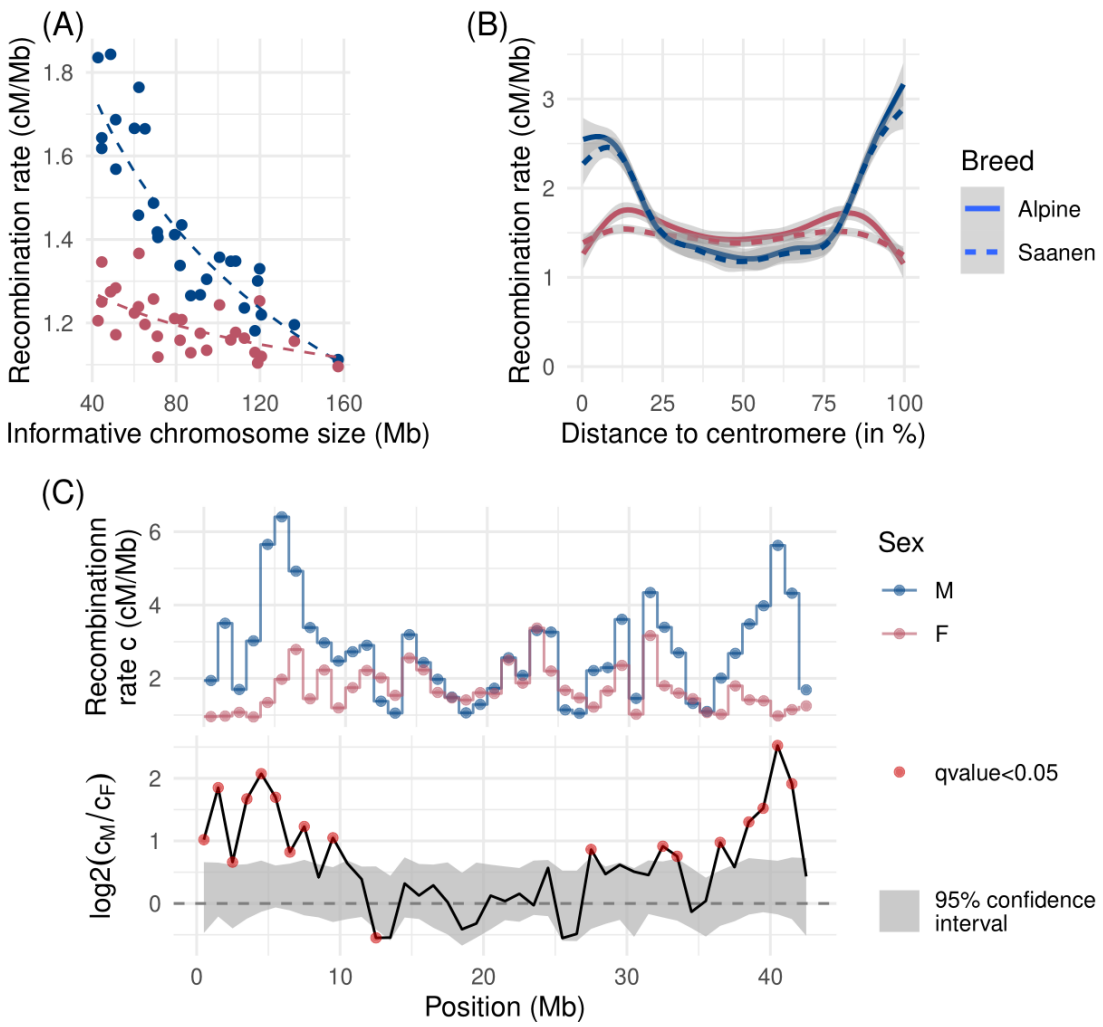

Figure 2: Estimated recombination rates for males and females at several scales. (A): Chromosomic recombination rate depending on chromosome size. (B): Bayesian estimates of the chromosome landscape of recombination rate depending on distance to centromere. Confidence intervals at 99%. (C): Recombination map on chromosome 25 at 1Mb resolution. Confidence interval under H_0 estimated using sex permutations.

Genetic Architecture of recombination phenotypes

Phenotype distributions

Two phenotypes were used to characterize the genetic determinism of recombination in the goat: the individual Genome-wide Recombination Rate (GRR), a measure of recombination

intensity and the intra-chromosomic shuffling, a measure of the *distribution* of crossovers on chromosomes. GRR is defined as the per meiosis crossover count divided by the length of informative genome. The average value of GRR over all meioses is 1.39 cM per Mb. The distributions of GRR between breeds and sexes are shown on **Figure S13** (**Additional File 1**).

Intra-chromosomic shuffling $\bar{r}_{intra}$ was defined as stated in Material & Methods. The average value of $\bar{r}_{intra}$ is 0.011. The distributions of $\bar{r}_{intra}$ between breeds and sexes are shown on **Figure S14** (**Additional File 1**). A linear model was fitted to estimate differences between populations. A representation of the estimates can be found on **Figure S15** (**Additional File 1**). No significant difference was found between sexes (p-value = 0.12) but a breed effect (Alpine $+1.01 \times 10^{-4}$; p-value = 0.03) and the interaction between breed and sex were found significant (female Alpine $+4.01 \times 10^{-4}$; p-value = 0.03).

Heritability estimates

Heritabilities of GRR and $\bar{r}_{intra}$ were estimated using an animal model taking into account repeated measures, only for the male phenotype. GRR heritability was estimated at 0.12±0.03 and was consistent in both breeds. $\bar{r}_{intra}$ heritability was estimated at 0.034±0.015 when both breeds were analysed together. No permanent environment was detected for any phenotype (**Table 2**).

Table 2: Variance components and heritability (SE) estimates for male genome-wide recombination rate and intra-chromosomic shuffling in a population of 843 Alpine and Saanen goats. Heritability was considered significant when (h^2±2SE) did not include 0.

Phenotype	Population	σ_A^2	σ_B^2	σ_E^2	τ^2	h^2

		All	0.50 (0.12)	0.05 (0.09)	3.52 (0.07)	0.01 (0.02)	0.12 (0.03) *
R	GRR	Alpine	0.47 (0.16)	0.11 (0.13)	3.57 (0.09)	0.03 (0.03)	0.11 (0.04) *
		Saanen	0.47 (0.17)	0.00 (0.13)	3.42 (0.11)	0.00 (0.03)	0.12 (0.04) *
		All	0.002 (0.001)	0.000 (0.001)	0.052 (0.001)	0.000 (0.01)	0.034 (0.015) *
$\bar{r}_{intra}$		Alpine	0.002 (0.001)	0.000 (0.001)	0.053 (0.001)	0.004 (0.019)	0.033 (0.019)
		Saanen	0.001 (0.001)	0.000 (0.001)	0.051 (0.002)	0.000 (0.021)	0.027 (0.020)

τ^2 = repeatability; * = significant heritability. With $\boldsymbol{\sigma_A^2}$ the genetic variance; $\boldsymbol{\sigma_B^2}$ the permanent environment variance and $\boldsymbol{\sigma_E^2}$ the residual variance.

Genome-Wide Associations

The EBVs predicted from the animal model were used as phenotypes for the GWAS on GRR and $\bar{r}_{intra}$. GWAS results using imputed genotypes to the sequence level are illustrated on **Figure 3** for Alpine and Saanen, on both phenotypes.

The analyses on GRR revealed 15 significant regions in Alpine and 4 suggestive regions in Saanen. For Alpine, there were one region on chromosomes 1, 6, 7, 13 and 29; three regions on chromosomes 5 and 10; as well as two regions on chromosomes 9 and 17. For Saanen, there

were one region on chromosomes 2, 4, 6 and 14. There is no overlap between the regions of interest of the two breeds. Correlation of $\hat{\beta}$ between breeds was calculated using standard errors as weights. It amounted to 0.25.

The analyses on $\bar{r}_{intra}$ revealed 7 significant regions in Alpine and 41 suggestive regions in Saanen. Correlation of $\hat{\beta}$ between the two breeds is 0.04. For Alpine, there were one region on chromosomes 1, 9 and 19; as well as two regions on chromosomes 8 and 10. The lead variant on chromosome 9 is 5kb away from the closest significant region for GRR GWAS, and the one on chromosome 10 is 268bp away from its respective significant region for GRR. Correlation of the $\hat{\beta}$ between the GWAS for GRR and $\bar{r}_{intra}$ is 0.89. For Saanen, there were five regions on chromosome 1; two regions on chromosomes 2, 4, 6, 8, 14, 17, 18, 19, 20, 21 and 24; one region on chromosomes 3, 5, 7, 9, 11, 13, 22 and 28; as well as three regions on chromosome 10 and 26. Despite some regions being on the same chromosomes as for GRR, no suggestive region is within 2Mb of suggestive regions for GRR. Correlation between the two GWAS in Saanen is 0.92.

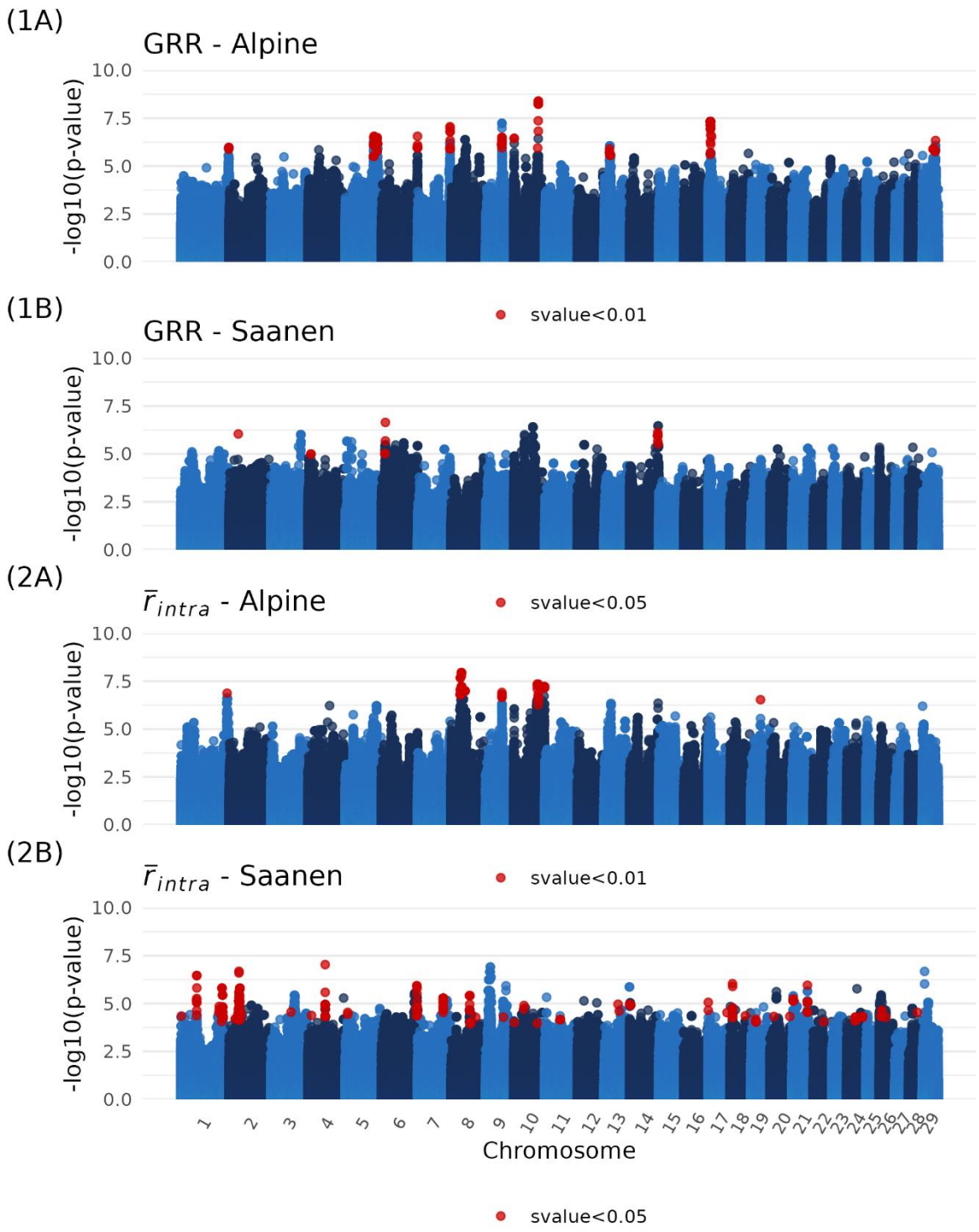

Figure 3: Manhattan plots of genome-wide association studies (GWAS) for male recombination phenotypes after imputation to sequence level. GRR and $\bar{r}_{intra}$ stand respectively for EBVs on male GRR and EBVs on male $\bar{r}_{intra}$.

485 Considering the GWAS on GRR, the most significant region of interest in Alpine is
486 located on chromosome 17 at around 7Mb (lead variant s-value=0.00236) and is illustrated on
487 **Figure 4A**. In this region, 58 candidate genes can be found, including SSH1 which scores 5/5.
488 The significant SNPs are located directly within the SSH1 gene, in an intronic region. The lead
489 variant has a MAF of 0.49 in the population. The region of interest illustrated on **Figure 4B**
490 contains significant SNPs in high LD in a region of 300kb and is rich in genes. It is the second
491 most significant region of interest (lead variant s-value=0.00238). Our standardised method for
492 ranking candidate genes identifies 86 genes with a score of at least 1/5. The highest scoring
493 (3/5) is CARMIL3 because of its physical proximity to the lead variant. Among the candidates
494 for this region are two genes previously associated with recombination: RNF212B (rank 17,
495 score 1/5) and REC8 (rank 51, score 1/5). The region of interest illustrated in **Figure 4C** is
496 located at the end of chromosome 6 and is the 6th most significant of the genome. It contains 36
497 candidate genes including SPON2 (rank 1, score 4/5) and RNF212 (rank 2, score 3/5).

498 In Saanen, the most suggestive region of interest can be found on chromosome 6 at
499 around 12.5Mb (lead variant s-value=0.0326) and is illustrated on **Figure 4D**. Several
500 significant SNPs of the area fall within the ANK2 gene which scores at 4/5 as a result. Eight
501 other candidate genes can be found in the region. Other notable genes in both breeds can be
502 found in **Table 3**. Exhaustive results for GWAS candidate genes in both breeds can be found
503 in **Additional File 7**, **tabs 1 & 2**.

504 **Table 3:** Genes scoring above 4/5 for the GWAS on male GRR in Alpine or Saanen.

Breed	Region of interest (chromosome: position of the lead variant)	Candidate gene (score/5)

	1:156523831	KCNH8 (5/5)
	5:94902996	GRIN2B (5/5)
	5:106271734	B4GALNT3 (5/5)
	6:117176129	SPON2 (4/5)
	7:105398227	SNAP47 (4/5)
Alpine	9:53226533	THEMIS (5/5)
	10:1686345	EFCAB11 (5/5)
	13:17395272	CCNY (4/5)
	17:6965601	SSH1 (5/5)
	17:7730110	GLTP (5/5)
	29:42944888	FERMT3 (5/5)
	2:29863833	TNS1 (5/5)
Saanen	6:12545797	ANK2 (4/5)
	14:91668757	B3GAT2 (5/5)

Considering the GWAS on $\bar{r}_{intra}$, the most significant region is the first signal of chromosome 8 in Alpine (lead variant s-value=0.0015). The gene PTPRD, which scores 4/5 for this region, has no known link with recombination. Another significant region can be found for this phenotype and breed at the end of chromosome 1 (**Figure 4E**). The highest scoring candidate genes are ERG (4/5), ETS2 (3/5) and KCNJ15 (2/5), but none of them seem to be involved in recombination either. The region is 6Mb away from PRDM9. The most significant region for $\bar{r}_{intra}$ in Saanen can be found on chromosome 4 (lead variant s-value=0.0026). The

gene AVL9 scores 5/5 for this region but does not seem to be involved in recombination. Exhaustive candidate genes results for GWAS on $\bar{r}_{intra}$ can be found in **Additional File 7, tabs 3 & 4**.

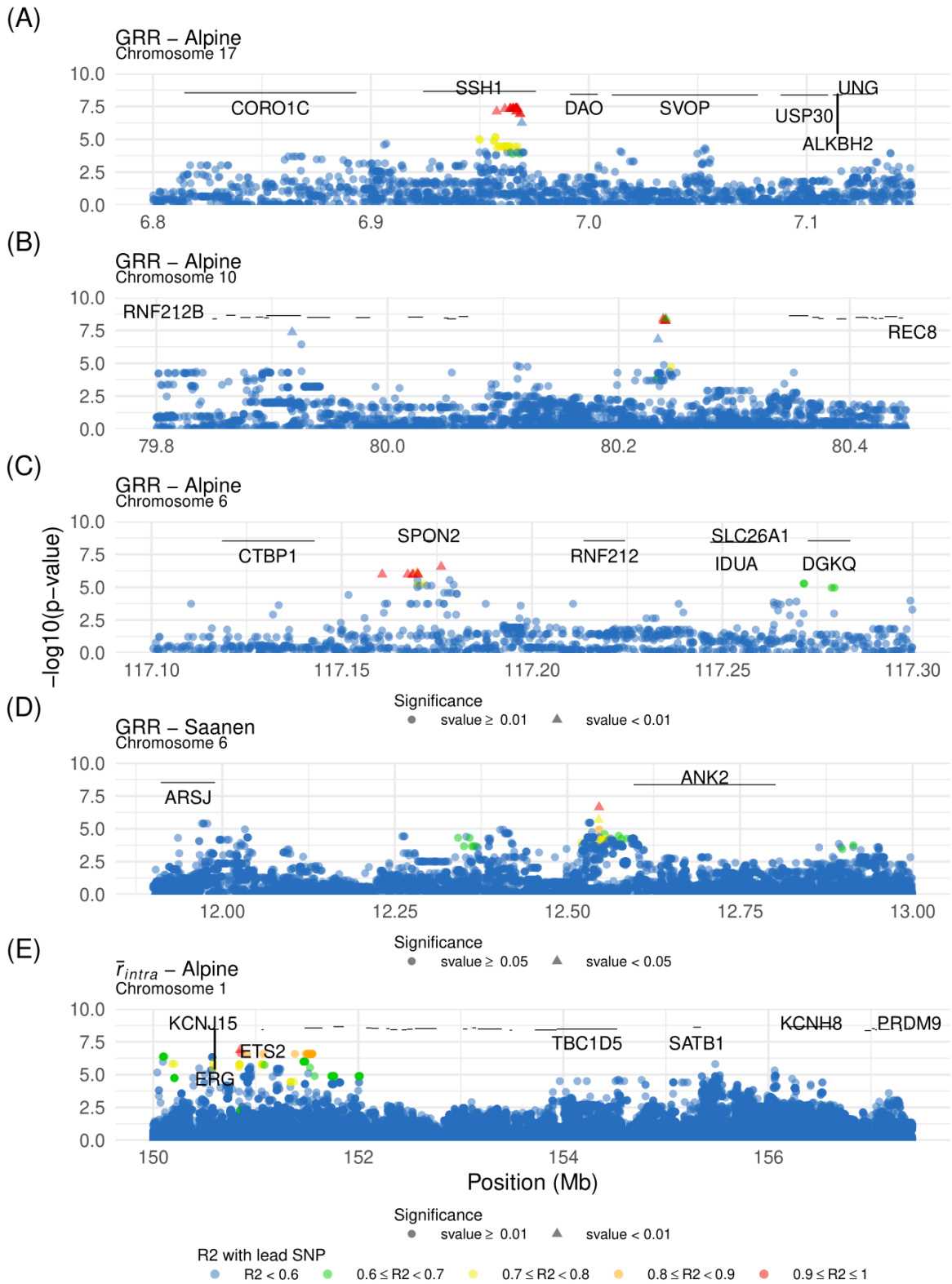

Figure 4: Zoom on GWAS Manhattan plots for regions of interest in both populations and phenotypes.

Discussion

This study is a detailed analysis of the recombination process and its genetic architecture in a new species: the goat (*Capra Hircus*). It provides precise estimates of recombination rates along the genome, using a new statistical inference approach that accounts for different levels of uncertainty in crossover data. This analysis revealed that the genetic length of the ARS1 goat genome assembly is 35.1M (Alpine males) corresponding to a genome-wide recombination rate of 1.42cM/Mb. However, recombination was found to be highly variable along the genome, with rates varying from 0.21 to 10.32cM/Mb. An important feature of the recombination landscape is the large heterochiasmy in some genomic regions, mostly located at chromosome ends. This is consistent with patterns of recombination in a closely related species, the sheep *Ovis Aries* (Johnston et al., 2016; Petit et al., 2017). In a second part of this study, we characterized the genetic architecture of recombination intensity and localisation. This analysis revealed low but significant genetic components to these traits and identified several QTLs for both traits, some of which match loci found in other species while others are new. Overall, our study adds to the growing knowledge on recombination and will contribute to understanding the evolution of recombination, a fundamental trait for individual fitness and adaptive potential of populations. We now discuss the importance of our findings and the future research questions it suggests.

Recombination landscape of the goat genome

Our study employs a similar strategy as previous studies in animal populations (Ma et al., 2015; Johnston et al., 2016; Johnston et al., 2017; Petit et al., 2017; Shen et al., 2018; Qanbari and Wittenburg, 2020; Brekke et al., 2022; Brekke et al., 2023a; Brekke et al., 2023b; McAuley et al., 2024), but we aimed at better accounting for the uncertainty in crossover data. The first source of uncertainty arises from the unprecise localisation of crossovers due to the

genotyping density, the second by the homozygosity of individuals in the pedigree, and the third concerns only female recombination rate estimation, as our available female data is much scarcer. To tackle the first problem, Petit et al. (2017) proposed a random sampling approach of crossover locations to average recombination rate estimation over crossover uncertainty. Here, we used the same strategy but extended it further to better identify genomic regions with significant differences in recombination rates between sexes. To tackle the second problem, we made use of an information provided by the software used to phase individuals (Yapp) that explicitly identifies regions of the genome where meioses are uninformative for crossover detection. In order to integrate both sources of uncertainty, we leveraged the Bayesian Gamma-Poisson model of Petit et al. (2017) that allows to incorporate this information naturally. One caveat of this model is that it is much harder to implement if one wants to add a hierarchical level to account for differences between groups (*e.g.* sex) but with non-null covariances. Recent developments around Poisson-Log-Normal models (Chiquet et al., 2021) could be used for this purpose (see e.g. Vassilief et al. (2022) in the context of recombination rate inference) but this will require further developments, in particular to better account for the longitudinal correlation of recombination rates along the genome. Finally, we proposed a permutation method to assess the significance of heterochiasmy along chromosomes. This method allows to draw conclusions from a lowly informative dataset in terms of sex differences, which could not have been confidently untangled otherwise. We believe it could be useful in future studies on heterochiasmy.

The average recombination rate of the goat genome (1.42cM/Mb varying from 1.12 to 1.71cM/Mb among chromosomes) is very similar to other ruminants (e.g. 1.5cM/Mb in sheep (Petit et al., 2017), 0.98cM/Mb in cattle (Qanbari and Wittenburg, 2020)). A striking feature of its landscape is a strong heterochiasmy that we could reveal even if the number of female meioses available in our datasets was relatively low. This heterochiasmy is mostly concentrated

in a few megabases at the chromosome ends, irrespective of the total chromosome size. As a consequence, the difference in average recombination rate between sexes is smaller for large chromosomes than small ones (**Figure 2A**). The biological causes underlying this heterochiasmy are still unknown. Further research combining genetics studies with cytogenetics and functional studies will most likely be required to unravel them.

Patterns of recombination and heterochiasmy in the goat match those found in sheep (Johnston et al., 2016; Petit et al., 2017; Vassilief et al., 2022) and cattle (Ma et al., 2015; Shen et al., 2018; Brekke et al., 2023a) although maybe to a lesser extent for this last species. Given that these three species have highly colinear genomes, it seems possible to make a quantitative comparison of recombination patterns along ruminant chromosomes to reveal how recombination evolved with speciation.

Genetic architecture of recombination phenotypes

Livestock breeding populations are good models to study the genetic architecture of the recombination process as they offer large datasets of genotyped individuals related through well characterised pedigrees. An important caveat of these datasets however lies in the unbalanced genotyping between sexes. In ruminant programs of genomic selection such as the one used here, male selection candidates are typically all genotyped while females are much less likely to be, although this tends to change with the decrease in genotyping costs. In our dataset, the number of females were too small to study the genetic architecture of recombination in this sex. We expect the number of genotyped females to grow in the future which will allow studying the genetic determinism of heterochiasmy in this species.

We studied two different phenotypes: the Genome-Wide Recombination Rate (GRR) which is a measure of recombination intensity, and intra-chromosomal shuffling ($\bar{r}_{intra}$) which depends on recombination intensity too but also measures differences in the localisation of

591 crossovers. As expected, we observed a high correlation between the two phenotypes (0.62).
592 However, despite strong correlation of the GWAS results on GRR and $\bar{r}_{intra}$, their genetic
593 architectures were found only partially overlapping, illustrating that they are probably capturing
594 different underlying biological processes (for example those involved in Double Strand Break
595 (DSB) localisation from those involved in DSB repair).

596 Heritability for male GRR was estimated at 0.12 with similar results for each breeds,
597 and similar to those found in other ruminant species for this phenotype (ACC or GRR): *e.g.*
598 0.22 for Dutch Holstein-Friesan males in cattle (Sandor et al., 2012), 0.13 in several cattle
599 breeds (Kadri et al., 2016), 0.12 for Soay sheep (Johnston et al., 2016) and 0.23 for Lacaune
600 sheep (Petit et al., 2017).

601 The heritability for male $\bar{r}_{intra}$ was estimated at 0.034 and consistent between the two
602 breeds, although only slightly significantly different from zero (SE=0.015). Heritability for this
603 phenotype seems to be smaller than in other species as Brekke et al. (2023) estimated it at 0.11
604 in salmon, a species using PRDM9 hotspots, and McAuley et al. (2024) estimated it at 0.14 in
605 sparrows, a species not using PRDM9 hotspots.

606 GWAS results for male GRR and $\bar{r}_{intra}$ showed low genetic correlation between Alpine
607 and Saanen. Despite identical heritability results, no QTL were found in common between the
608 two breeds. Besides, the correlation of the $\hat{\beta}$ for the SNPs that are significant in at least one
609 breed is only 0.25 for GRR and 0.04 for $\bar{r}_{intra}$. This suggests strong differences in the
610 underlying genetic determinism of recombination between these two breeds, although they are
611 both European breeds originating from the Alps and closely related in a worldwide diversity
612 panel (Colli et al. 2018). Despite this relatively close genetic proximity, those breeds have been
613 shown to have different genetic architecture for other traits. For example, Martin et al. (2018)
614 found genetic heterogeneity in the determinism of dairy-related traits between the two breeds,

615 despite similar breeding goals. Our results add to the literature that suggests that genome-wide
616 recombination landscapes tend to be mostly shared between populations while the underlying
617 genetic determinisms of the process can be quite different, *e.g.* Petit et al. (2017) in sheep,
618 Lozada-Soto et al., (2021); Brekke et al. (2022) in pigs, although results in cattle tend to show
619 they are more conserved (Shen et al., 2018).

620 Candidate genes associated with QTL loci in the GWAS were prioritised by
621 implementing a scoring methodology using criteria based on physical and genetic distance to
622 significant SNPs. This methodology was developed in order to provide an objective measure
623 for the investigation of candidate genes underlying GWAS peaks and quickly highlight strong
624 candidates in regions rich in annotations. For instance, 13Mb on chromosome 5 contain three
625 different peaks in rapid succession and 141 genes are annotated in the region. The scoring
626 allowed to quickly identify the most promising genes and order them all depending on their
627 physical and genetic proximity to each peak. Genes involved in meiosis in other species were
628 searched for in a second step to identify functional candidates at QTLs among the positional
629 candidates.

630 The GWAS results for GRR in Alpine goats show a signal towards the end of
631 chromosome 6, in very close proximity to the RNF212 gene. RNF212 has been found to be
632 associated with GRR in humans (Kong et al., 2014), cattle (Sandor et al., 2012; Kadri et al.,
633 2016) and sheep (Johnston et al., 2016; Petit et al., 2017). It has also been proved that RNF212
634 is involved in and is necessary for meiosis in mammals (Reynolds et al., 2013; Qiao et al.,
635 2014). Yet it only scores second candidate for the signal on chromosome 6 because it does not
636 directly contain the significant SNPs. Of course, this is still consistent with a regulatory variant
637 affecting RNF212 expression.

638 One signal with significant SNPs as far as 330kb apart can be found at the end of
639 chromosome 10 between 79.9 and 80.3Mb in the Alpine breed (**Figure 4B**). The loci RNF212B
640 and REC8 are in the region. They were previously suspected to be responsible for signals in
641 sheep (Johnston et al., 2016; Petit et al., 2017) and cattle (Sandor et al., 2012; Kadri et al., 2016)
642 but the studies had difficulty attributing the signal to one of the two genes due to a lack of
643 resolution. Among our candidates, RNF212B ranks 17th and REC8 ranks 51st, as the region is
644 very rich in annotated genes. Functionally, both these genes make very good candidates, with
645 RNF212B being closely related to RNF212 and REC8 being involved in meiotic recombination
646 (Parisi et al., 1999). But the gene richness of the region (62 genes in 690kb), their relative
647 physical distance to the lead variant and the fact that, surprisingly, the main part of the signal
648 is located in an area of 250kb completely devoid of annotated protein coding genes, makes
649 discriminating the two candidates, or even determining if they are good candidates, quite
650 difficult.

651 Among regions of interest, a noteworthy signal was found in Alpine on chromosome 17
652 around 7Mb (**Figure 4A**). The signal falls directly within the gene SSH1, which scored 5/5
653 according to our ranking method. Its paralog SSH2 is known to be involved in meiosis in mice
654 (Xu et al., 2023) and the signal suggests SSH1 might be involved too. This signal for GRR was
655 found in no other similar GWAS studies in mammalian species.

656 When finding regions of interest at the end of chromosome 1 in Alpine, both for GRR
657 and $\bar{r}_{intra}$, one could have assumed that these signals were associated with PRDM9, known
658 regulator of recombination hotspots (Baudat et al., 2010; Baker et al., 2017) and located at the
659 very end of chromosome 1 in goats (**Additional File 3**, **tab 1**). Therefore, it was surprising to
660 see that PRDM9 actually scores very low for the GRR signal (1/5 score) due to its physical
661 distance from the significant region and is even too far from the $\bar{r}_{intra}$ signal to gain any score

662 point. It could be explained that no association was found between PRDM9 and GRR because
663 GRR is a phenotype of intensity of recombination, and PRDM9 regulates the location of
664 recombination. But $\bar{r}_{intra}$ is a recombination phenotype linked with the position of crossovers
665 and the association was even more expected, especially considering that Brekke et al. (2023)
666 found a clear signal for PRDM9 in their own GWAS on salmon $\bar{r}_{intra}$ phenotype. Furthermore,
667 the third highest ranking candidate gene in goats for $\bar{r}_{intra}$ is KCNJ15 (2/5 score and **Figure**
668 **4E**). Although this gene seems to have no link with recombination, an association for this gene
669 with GRR has previously been found in sheep (Petit et al., 2017). Furthermore, Sandor et al.
670 (2012) found signal for the genes KCNJ2 and KCNJ16, paralogs of KCNJ15, in their GWAS
671 for GRR. Thus, further investigation on the involvement of KCNJ15 in recombination may be
672 relevant.

673 # Conclusions

674 In this study, we have found that recombination patterns are very alike between the
675 Alpine and Saanen breeds, but heterochiasmy is strongly present, manifested at the
676 chromosome ends. The recombination pattern we observed is well conserved with closely
677 related species like sheep and cattle. We also showed that male GRR and male $\bar{r}_{intra}$ are
678 heritable and that this value is the conserved in the two breeds. Heritability for GRR is
679 concordant with results found in sheep and cattle, but goat heritability for $\bar{r}_{intra}$ seems to be
680 lower than in the other species in which it was estimated. GWAS results showed low correlation
681 between the two breeds, for both phenotypes, and it seems that despite similar recombination
682 patterns and same heritability for recombination traits, the underlying genetic architecture of
683 recombination is different between Alpine and Saanen. Candidate genes found in GWAS are
684 partially in common with what was previously found in literature, with some novel genes. This

participates in showing that genetic control for recombination is well conserved among many lowly-related species, but also that the genes involved in this control can evolve quickly.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Initial genotypes and pedigree were provided by Capgenes and INRAE. Their access is restricted by the provider. Crossover data and informativity data, derived from these initial datasets, as well as genotypes of the phenotyped parents are available at: https://doi.org/10.57745/HPYLLT. Sequence data for the imputation reference panel can be found on the VarGoats project website: https://www.goatgenome.org/vargoats.html.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by INRAE.

Authors' contributions

All authors conceptualised this study, curated, analysed and investigated the data. BS acquired funding for this study through INRAE. All authors developed methodology for this study and administered the project. RR provided the initial dataset as resources for the project.

BS and AE did programming and software development to produce results. BS and RR supervised the project as PhD supervisors. All authors validated the results. AE and BS did visualisation work through diverse presentations and wrote the original draft. All authors reviewed and edited the current manuscript.

Acknowledgements

Many numerical analyses were performed on the genobioinfo bioinformatics platform Toulouse Midi-Pyrénées (Bioinfo Genotoul). We are thankful to Capgenes for giving us access to their genotypes used for genomic selection, and to VarGoats for giving us access to their sequences. We also thank the experimental unit P3R Bourges for granting us access to their genotypes as well. Full thanks to Susan Johnston, Tristan Mary-Huard and Leopoldo Sanchez Rodriguez for fruitful exchanges during the production of these results.

References

Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M (2017) Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. eLife **6**: e24133

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B (2010) PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. Science **327**: 836–840

Brekke C, Berg P, Gjuvsland AB, Johnston SE (2022) Recombination rates in pigs differ between breeds, sexes and individuals, and are associated with the RNF212, SYCP2, PRDM7, MEI1 and MSH4 loci. Genet Sel Evol **54**: 33

Brekke C, Gjuvsland AB, Berg P, Johnston SE (2025) Independent genetic basis of meiotic crossover positioning and interference in domestic pigs. Sci Rep **15**: 9260

Brekke C, Johnston SE, Gjuvsland AB, Berg P (2023a) Variation and genetic control of individual recombination rates in Norwegian Red dairy cattle. Journal of Dairy Science **106**: 1130–1141

Brekke C, Johnston SE, Knutsen TM, Berg P (2023b) Genetic architecture of individual meiotic crossover rate and distribution in Atlantic Salmon. Sci Rep **13**: 20481

Browning BL, Browning SR (2016) Genotype Imputation with Millions of Reference Samples. The American Journal of Human Genetics **98**: 116–126

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci **4**: 7

Chiquet J, Mariadassou M, Robin S (2021) The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances. Front Ecol Evol **9**: 588292

Cooney CR, Mank JE, Wright AE (2021) Constraint and divergence in the evolution of male and female recombination rates in fishes. Evolution **75**: 2857–2866

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M (2008) High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. Science **319**: 1395–1398

Crow JF, Kimura M (1965) Evolution in Sexual and Asexual Populations. The American Naturalist **99**: 439–450

Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET (2019) Accurate, scalable and integrative haplotype estimation. Nat Commun **10**: 5436

Denoyelle L, Talouarn E, Bardou P, Colli L, Alberti A, Danchin C, Del Corvo M, Engelen S, Orvain C, Palhière I, et al (2021) VarGoats project: a dataset of 1159 whole-genome sequences to dissect Capra hircus global diversity. Genet Sel Evol **53**: 86

Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Tosser-Klopp G, Wang J, Yang S, Liang J, et al (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotechnol **31**: 135–141

Druet T, Georges M (2015) LINKPHASE3: an improved pedigree-based phasing algorithm robust to genotyping and map errors. Bioinformatics **31**: 1677–1679

Du X, Servin B, Womack JE, Cao J, Yu M, Dong Y, Wang W, Zhao S (2014) An update of the goat genome assembly using dense radiation hybrid maps allows detailed analysis of evolutionary rearrangements in Bovidae. BMC Genomics **15**: 625

Favier A, Elsen J-M, de Givry S, Legarra A (2010) Optimal haplotype reconstruction in half-sib families. Edinburgh, UK, pp 27–15

Fröhlich J, Vozdova M, Kubickova S, Cernohorska H, Sebestova H, Rubes J (2015) Variation of Meiotic Recombination Rates and MLH1 Foci Distribution in Spermatocytes of Cattle, Sheep and Goats. Cytogenet Genome Res **146**: 211–221

Johnsson M, Whalen A, Ros-Freixedes R, Gorjanc G, Chen C-Y, Herring WO, de Koning D-J, Hickey JM (2021) Genetic variation in recombination rate in the pig. Genet Sel Evol **53**: 54

Johnston SE, Bérénos C, Slate J, Pemberton JM (2016) Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis aries*). Genetics **203**: 583–598

Johnston SE, Huisman J, Ellis PA, Pemberton JM (2017) A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus elaphus). G3 (Bethesda) **7**: 12

Johnston SE, Huisman J, Pemberton JM (2018) A Genomic Region Containing REC8 and RNF212B Is Associated with Individual Recombination Rate Variation in a Wild Population of Red Deer. G3 (Bethesda) **8**: 12

Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, Fritz S, Mullaart E, Baurain D, Boichard D, 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

Koehler KE, Hawley RS, Sherman S, Hassold T (1996) Recombination and nondisjunction in humans and flies. Human Molecular Genetics **5**: 1495–1504

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, et al (2002) A high-resolution recombination map of the human genome. Nat Genet **31**: 241–247

Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, Magnusdottir E, Olafsdottir SB, Thorsteinsdottir U, Stefansson K (2014) Common and low-frequency variants associated with genome-wide recombination rate. Nat Genet **46**: 11–16

Lozada-Soto EA, Maltecca C, Wackel H, Flowers W, Gray K, He Y, Huang Y, Jiang J, Tiezzi F (2021) Evidence for recombination variability in purebred swine populations. J Anim Breed Genet **138**: 259–273

Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, Bickhart DM, Cole JB, Null DJ, Liu GE, et al (2015) Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. PLoS Genet **11**: e1005387

Martin P, Palhière I, Maroteau C, Clément V, David I, Klopp GT, Rupp R (2018) Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. Journal of Dairy Science **101**: 5214–5226

McAuley JB, Servin B, Burnett HA, Brekke C, Peters L, Hagen IJ, Niskanen AK, Ringsby TH, Husby A, Jensen H, et al (2024) The Genetic Architecture of Recombination Rates is Polygenic and Differs Between the Sexes in Wild House Sparrows (Passer domesticus). Molecular Biology and Evolution **41**: msae179

Meyer K (2007) WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J Zhejiang Univ - Sci B **8**: 815–821

Muller HJ (1964) THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. Mutat Res **106**: 2–9

Muñoz-Fuentes V, Marcet-Ortega M, Alkorta-Aranburu G, Linde Forsberg C, Morrell JM, Manzano-Piedras E, Söderberg A, Daniel K, Villalba A, Toth A, et al (2015) Strong Artificial Selection in Domestic Mammals Did Not Result in an Increased Recombination Rate. Mol Biol Evol **32**: 510–523

Parisi S, McKay MJ, Molnar M, Thompson MA, van der Spek PJ, van Drunen-Schoenmaker E, Kanaar R, Lehmann E, Hoeijmakers JHJ, Kohli J (1999) Rec8p, a Meiotic Recombination and Sister Chromatid Cohesion Phosphoprotein of the Rad21p Family Conserved from Fission Yeast to Humans. Mol Cell Biol **19**: 3515–3528

Petit M, Astruc J-M, Sarry J, Drouilhet L, Fabre S, Moreno CR, Servin B (2017) Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. Genetics **207**: 767–784

Pook T, Mayer M, Geibel J, Weigend S, Cavero D, Schoen CC, Simianer H (2020) Improving Imputation Quality in BEAGLE for Crop and Livestock Data. G3 Genes|Genomes|Genetics **10**: 177–188

Qanbari S, Wittenburg D (2020) Male recombination map of the autosomal genome in German Holstein. Genet Sel Evol **52**: 73

Qiao H, Prasada Rao HBD, Yang Y, Fong JH, Cloutier JM, Deacon DC, Nagel KE, Swartz RK, Strong E, Holloway JK, et al (2014) Antagonistic roles of

ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. Nat Genet **46**: 194–199

Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N, Biswas K, Holloway JK, Baudat F, de Massy B, Wang J, et al (2013) RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat Genet **45**: 269–278

Rubinacci S, Delaneau O, Marchini J (2020) Genotype imputation using the Positional Burrows Wheeler Transform. PLoS Genet **16**: e1009049

Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M (2012) Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle. PLoS Genet **8**: e1002854

Schensted, C (1961) Longest increasing and decreasing subsequences. Can J Math **13**: 179–191

Schibler L, Vaiman D, Oustry A, Giraud-Delville C, Cribiu EP (1998) Comparative Gene Mapping: A Fine-Scale Survey of Chromosome Rearrangements between Ruminants and Humans. Genome Res **8**: 901–915

Servin B (2024) Yapp: a suite of tools to analyze genomic data in pedigrees.

Shen B, Jiang J, Seroussi E, Liu GE, Ma L (2018) Characterization of recombination features and the genetic basis in multiple cattle breeds. BMC Genomics **19**: 304

Stephens M (2016) False discovery rates: a new deal. Biostat kxw041

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA **100**: 9440–9445

Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, Donnadieu-Tonon C, Eggen A, Heuven HCM, Jamli S, et al (2014) Design and Characterization of a 52K SNP Chip for Goats. PLoS ONE **9**: e86227

Van der Auwera G, O'Connor BD (2020) Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st edition). O'Reilly Media

Vassilief H, Id Bella M, Hazard D, Tortereau F, Faraut T, Johnston SE, Servin B (2022) 202. Sex differences in recombination maps are associated with differential hotspot usage in sheep. Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers, Rotterdam, the Netherlands, pp 861–864

Veller C, Kleckner N, Nowak MA (2019) A rigorous measure of genome-wide genetic shuffling that takes into account crossover positions and Mendel's second law. Proceedings of the National Academy of Sciences **116**: 1659–1668

Weng Z, Wolc A, Su H, Fernando RL, Dekkers JCM, Arango J, Settar P, Fulton JE, O'Sullivan NP, Garrick DJ (2019) Identification of recombination hotspots and quantitative trait loci for recombination rate in layer chickens. J Animal Sci Biotechnol **10**: 1–13

Weng Z-Q, Saatchi M, Schnabel RD, Taylor JF, Garrick DJ (2014) Recombination locations and rates in beef cattle assessed from parent-offspring pairs. Genet Sel Evol **46**: 34

Xu K, Su X, Fang K, Lv Y, Huang T, Li M, Wang Z, Yin Y, Muhammad T, Liu S, et al (2023) The Slingshot phosphatase 2 is required for acrosome biogenesis during spermatogenesis in mice. eLife **12**: e83129

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet **44**: 821–824

Figures

Figure 1: Illustration of the notion of interval informativity.

Figure 2: Estimated recombination rates for males and females at several scales.

(A): Chromosomic recombination rate depending on chromosome size. (B): Bayesian estimates of the chromosome landscape of recombination rate depending on distance to centromere. Confidence intervals at 99%. (C): Recombination map on chromosome 25 at 1Mb resolution. Confidence interval under H_0 estimated using sex permutations.

Figure 3: Manhattan plots of genome-wide association studies (GWAS) for male recombination phenotypes after imputation to sequence level.

GRR and $\bar{r}_{intra}$ stand respectively for EBVs on male GRR and EBVs on male $\bar{r}_{intra}$.

Figure 4: Zoom on GWAS Manhattan plots for regions of interest in both populations and phenotypes.

Additional Files

Additional File 1

Supplementary figures: PDF file containing all supplementary figures to this paper.

Additional File 2

SNPs between synteny breakpoints: XLS file listing all 123 SNPs removed after identifying synteny breakpoints with human genome on ARS1 assembly.

Additional File 3

BLAST results: XLS file containing two tabs. The first tab contains BLAST results for the bovine gene sequence RNF212 on ARS1. The second tab contains BLAST results for the bovine gene sequence PRDM9 on ARS1.

Additional File 4

Recombination maps table: TXT file containing estimated recombination rates for each 1Mb interval of the autosomal goat genome, for each sex.

Additional File 5

Recombination maps figures: PDF file containing illustrations of recombination results (Additional File 4) for each chromosome, as well as an illustration of the log ratio of recombination rate between males and females, and when this log ratio indicates significant differences between sexes.

Additional File 6

Significant intervals for sex differences in recombination rate: TXT file listing all 1Mb intervals for which sex differences in recombination rate have been found significant.

Additional File 7

Candidate genes for GWAS: XLS file listing all genes that scored at least 1 as candidates in the GWAS on EBVs for male GRR and shuffling. First tab shows candidate genes for GRR in Alpine, second tab for GRR in Saanen, third tab for shuffling in Alpine, fourth tab for shuffling in Saanen. For each region of interest, genes have been sorted from higher scoring to lower scoring. Annotation has been added on genes previously found in GWAS for recombination phenotypes in ruminants, referencing the corresponding papers.