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 basis of two recombination phenotypes: genome-wide recombination rate (GRR) and intra-chromosomal 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 determinants 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 [1] and is also a great driver of genetic diversity by creating new combinations of alleles and avoiding the accumulation of deleterious mutations over time [2,3]. 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 basis of the recombination process. Pioneered in humans [4,5], 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 [6–11], but also in sheep [12], chicken [13], pigs [14–17] and Atlantic salmon [18]. Apart from livestock, recombination in several natural populations with available pedigrees have also been studied: Soay sheep [19], red deer [20,21] and wild house sparrow [22].

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 most species like pigs [14–16], humans [4], deer [20], chicken [13] or Atlantic salmon [18], females recombine more than males; while in other species like cattle [8–

11] or sheep [19], males show higher recombination than females. [22] and [23] illustrate this inconsistent heterochiasmy in birds and fishes, respectively.

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-chromosomal shuffling [24]. 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 [11] and 0.22 in Dutch Holstein-Friesian cattle [6]. But variations in heritability can be observed between species and, most notably, sexes within a species [9,12,19,21]. Identified QTL for these phenotypes can also vary between populations and species. Cattle [6,8–11], sheep [12,19] and red deer [21] all show a signal in the region of the genes RNF212B/HEI10/REC8/REC114, four genes in very close proximity to each other in these three species. In cattle [6,9] and sheep [12,19], associations with the *RNF212* gene have also been found. 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, [19] found a QTL at RNF212 only in Soay females, and not in males. Yet [12] actually found a QTL for this same gene in males from another population, the Lacaune. In cattle, [8] identified thirteen loci associated with recombination intensity, and only three of them were shared between sexes. In [9]'s bovine population, four of their ten detected QTLs are sex-specific. And once again, [11] found sex differences in Norwegian red dairy cattle.

85 In this context of strong genetic diversity in very related groups, we present a new study on
86 another ruminant species: the goat. In goats, the last genetic map was built with a linkage
87 methodology in 1998 [25]. At the time, its primary goal was still to understand the organisation
88 of genes between one another, and the map was never updated since. Genome-wide
89 recombination rate was estimated through the use of cytogenetics techniques by counting
90 MLH1 foci [26,27] but never confirmed using other methods.

91 In 2010, the first reference genome assembly was built for the domestic goat [28] and in 2013,
92 an international goat genome consortium was created to acquire more genomic data and
93 investigate genetic variations in the species. This lead to the development of a 50k SNP chip
94 [29] and later on, the creation of a sequence-level SNP panel of more than a thousand goats
95 from all over the world [30].

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

108 addition, we also studied the genetic basis of two recombination phenotypes in both available
109 breeds, by estimating genetic parameters and performing genome-wide association studies.

110 Methods

111 Study population and genotype data

112 In this work, we exploited pedigree and genotyping data of 7,588 goats born between 2008 and
113 2019. They amount to 3,169 males and 4,419 females from both the Alpine (5,051 samples)
114 and Saanen (2,537 samples) breeds. The animals were genotyped with versions 1 and 2 of the
115 Illumina Inc. GoatSNP50 chip. Only autosomal SNPs and markers common to both arrays were
116 kept, resulting in 53,347 SNPs.

117 Quality control of physical marker ordering

118 Errors in the physical ordering of markers can lead to detect spurious recombination events.
119 Recombination studies therefore require a robust marker ordering. Here, this ordering was built
120 from two independent physical maps of the SNPs: the reference genome assembly ARS1 and
121 the Radiation Hybrid (RH) map of [31]. We decided to keep only SNP markers that have a
122 consistent ordering on both physical maps. This was done by identifying, for each chromosome,
123 the longest increasing subsequence (LIS)[32] in the mapping of the RH order to the assembly
124 order. This requires knowing the order of the SNP markers on both maps. For the reference
125 genome ordering, genomic positions of the 53,347 v1 autosomal markers of the GoatSNP50
126 Illumina chip were obtained from the ARS1 assembly as described on the International Goat
127 Genome website (http://www.goatgenome.org/projects.html). For the RH map, additional
128 analysis was required as it was built using SNP array of different species (ovine and bovine
129 SNPs herein called external SNPs). We mapped the external SNPs to the ARS1 reference
130 genome and then projected the position of each caprine SNP on the RH map from the positions
131 of its two flanking external SNPs. For goat markers located at the extremities of the

132 chromosome, which had only one flanking external SNP and could not be projected on the RH
133 map, we assumed the reference genome order was correct (but see below for further data quality
134 control steps). **Figure S1** (**See Additional File 1**) illustrates this quality control step on two
135 chromosomes. After this step, 54,013 markers were kept.

136 Furthermore, an initial detection of crossovers on the genome (using methods detailed below)
137 revealed 7 genomic regions (123 SNP markers in total) that showed extremely high
138 recombination rates (See example on chromosome 1 in **Additional File 1, Figure S2**). Detailed
139 analysis of the synteny conservation with the human genome in these regions showed that they
140 corresponded systematically to conserved synteny breakpoints. As these events were most
141 probably some remaining unidentified assembly errors, markers involved were discarded in
142 downstream analysis. For a list of these markers, see **Additional File 2, Table S1**.

143 Genotype quality control

144 Standard data cleaning procedures were applied using PLINK 1.9 [33]. Animals and SNPs with
145 call-rate below 99% were removed (**See Additional File 1, Figure S3**). Monomorphic SNPs in
146 the dataset were removed. No MAF filter was applied at this step as even a small MAF may be
147 informative for crossover detection in families polymorphic for these markers. Markers with
148 strong departures from Hardy-Weinberg genotype proportions ($p < 10^{-7}$) within each breed
149 were removed. Note that families with a sire having more than 50 offspring were not used to
150 compute the Hardy-Weinberg equilibrium (HWE) test as fathers carrying rare alleles in such
151 designs can have a great impact on testing for HWE.

152 Mendelian errors were tested using the *mendel* command of the Yapp software version 0.3 [34].
153 It evaluated the number of incoherent markers between parent and offspring and automatically
154 corrected the kinship relationships in the pedigree which were beyond the threshold fixed by
155 the software.

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

158 Data phasing and crossover detection

159 Genotype phasing and crossover detection were done using Yapp version 0.4. First, the *phase*
160 command was used in order to phase genotypes. In order to reconstruct the paternally and
161 maternally inherited haplotypes of each individual (i.e. to infer the phase of heterozygous loci),
162 Yapp uses a set of discrete steps. First, using all available genotypes, it identifies the
163 unambiguous transmissions from parents to offspring, *e.g.* the alleles transmitted by a
164 homozygous parent or received by a homozygous offspring are easily phased. This leads to a
165 set of partially resolved phases. Second, for each parent in the pedigree, from oldest to newest,
166 it improves the inference of parental phases by combining all partially resolved transmitted
167 haplotypes using the algorithm of [35]. Third, after this first pass in the pedigree, the transmitted
168 haplotypes from parent to offspring are improved using a Hidden Markov Model along the
169 whole chromosome similar to the approach of [36] and segregation indicators are inferred.
170 Then, the *recomb* command was used among a total of 6,434 meioses to detect crossovers. A
171 crossover in the parent meiosis is defined as a change in phase observed in the offspring, on the
172 inherited chromosome from the parent. Crossovers are detected between two heterozygous
173 SNPs and can only be detected between the first and the last phased, heterozygous SNPs.

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

178 Distance between two crossovers on the same chromosome was calculated for every meiosis.
179 As crossovers are detected within an interval, distance between crossovers was calculated as

180 the distance between the closest interval boundaries. All double-crossovers closer than 5 Mb
181 from one another were considered false-positives due to phasing errors and both were removed
182 (**See Additional File 1, Figure S4**). After filtering, the dataset contained 6,204 meioses and
183 204,727 crossovers.

184 Recombination maps

185 Recombination maps were constructed by estimating recombination rates in non-overlapping
186 intervals of one megabase (Mb) along the genome. The methodology for estimating
187 recombination rates was adapted from [12]. For a given genomic interval i and a meiosis m,
188 the count of the number of crossovers Y_{mi} is modelled using a Poisson distribution whose
189 parameter (measured in Morgans) depends on the recombination rate c_i (in centiMorgans per
190 Mb, cM/Mb) assumed constant in the interval and across meioses, and the length of the genomic
191 interval l_{mi} (in Mb) that can depend on the meiosis (see below).

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

193 $$Y_{mi} \mid c_i \sim Poisson(0.01 c_i l_{mi}) \qquad (1).$$

194 As in [12], we define a prior on c_i as:

195 $$c_i \sim \Gamma(\alpha, \beta) \quad (2).$$

196 As in [12], we used an empirical Bayes approach to set α and β from the genome-wide
197 distribution of Maximum Likelihood Estimates of recombination rates ($\alpha = 3.02$ and $\beta =$
198 2.32, see **Additional File 1, Figure S5**). Given the conjugacy of Poisson and Gamma
199 distributions, the posterior distribution for c_i is then:

200 $$c_i \mid Y_{\cdot i} \sim \Gamma(\alpha + \textstyle\sum_m Y_{mi}, \beta + 0.01 \sum_m l_{mi}) \qquad (3).$$

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

203 $$\widehat{c_{\mathrm{chr}}} = \frac{\alpha+\sum_m Y_{chr,m}}{\beta+\sum_m l_{chr,m}} \quad (4),$$

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

207 ***Meiosis Informativity***

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

216 **[Figure 1 here]**

217 ***Accounting for uncertainty in crossover locations***

218 Crossovers are not determined at a specific position but resolved within genome regions that
219 vary in size depending on the parent and offspring genotypes. The average length of these
220 intervals in our data is 1,4Mb with a 95% range of 0.1Mb to 10.2Mb (**See Additional File 1,**
221 **Figure S6**). To account for this uncertainty in the position of crossovers, we used a Monte Carlo
222 approach as in [12]: we first sampled crossover positions uniformly in the detection intervals
223 and then the recombination rate from its posterior (Equation 4). This was done 100 times,
224 providing for each interval 100 samples from the posterior distribution of recombination rate.

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

227 $$\widehat{c_i} = E(c_i|Y_{\cdot i\cdot}) = E_k\big(E(c_i|Y_{\cdot ik})\big) = E_k(b_{ik}) \qquad (5),$$

228 And according to posterior variance law:

229 $$Var(\widehat{c_i}) = E\big(Var_k(c_i|Y_{\cdot ik})\big) = Var(c_i) - Var_k(E(c_i|Y_{\cdot ik}) = Var(c_i) - Var_k(b_{ik}) \qquad (6).$$

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

232 ***Assessing sex differences in local recombination rates***

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

243 Our measure of heterochiasmy in an interval i is:

244 $$\widehat{\mathrm{d}_i} = log2(\widehat{\mathrm{c}_{\mathrm{M}i}}) - log2(\widehat{\mathrm{c}_{\mathrm{F}i}}) \qquad (7).$$

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

246 To account for the difference in data informativity between sexes, we permuted the sex of the
247 parents 1,000 times to obtain a sample of $\mathbf{d_i}$ under the null of no difference between sexes. As
248 permutations do not change the number of male and female meioses, they allow to account for
249 difference in informativity between sexes. The resulting distribution of $\mathbf{d_i}$ in the permutation
250 samples were very close to normally distributed (**See Additional File 1, Figure S7**), so in each
251 interval the null distribution of $\mathbf{d_i}$ can be described by a normal distribution with parameters set
252 at the sample mean and variance. This in turn allows to derive p-values for a difference between
253 sexes in each interval. Finally, to account for multiple testing across intervals, the False
254 Discovery Rate (FDR) approach of [37] using the R "qvalue" package was applied on the p-
255 values and the significance threshold set at 0.05.

256 Recombination phenotypes

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

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

268 The other phenotype we studied is a measure of the efficacy of a meiosis to create new allelic
269 combinations in the offspring: the rate of intra-chromosomal shuffling $\overline{r}_{intra}$ [24]. It is defined

270 as the probability that two randomly chosen loci uncouple their alleles during meiosis and is
271 calculated as:

272 $$\overline{r}_{intra} = \sum_{k=1}^{n} 2p_k(1-p_k)L_k^2 \qquad (8),$$

273 Where for chromosome k, p_k is the proportion of SNPs inherited from one parent of the focal
274 animal (or grandparent of the offspring), L_k is the chromosome's proportion in the autosomal
275 genome, and n is the number of autosomes.

276 Heritabilities

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

279 To estimate heritability, variance components were estimated by restricted maximum likelihood
280 (REML) using the Wombat software [38]. An animal model including a permanent
281 environment effect was fitted:

282 $$\mathbf{y} \sim \mathbf{X\beta} + \mathbf{Za} + \mathbf{Zu} + \mathbf{e} \qquad (9).$$

283 $\mathbf{y}$ is the vector of observations (GRR or $\overline{r}_{intra}$); $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{a}$ is the vector
284 of direct animal (additive) genetic effects; $\mathbf{u}$ is the vector of animal permanent environment
285 effect; $\mathbf{e}$ is the vector of residual effects; $\mathbf{a} \sim N(0, \mathbf{A}\sigma_A^2)$, $u \sim N(0, \mathbf{I}\sigma_B^2)$ and $\mathbf{X}, \mathbf{Z}$ are the
286 incidence matrixes. The pedigree matrix $\mathbf{A}$ was computed on five generations of the genealogy.
287 The model was applied to all breeds separately or to the entire population at once. A fixed effect
288 for breed was therefore added where necessary.

289 Once the variance components had been estimated, values were incorporated into the model as
290 fixed values to obtain the Estimated Breeding Values (EBV) for each phenotype. Due to the
291 absence of differences in estimated heritability between breeds (see Results), EBVs were
292 estimated by pooling samples from both breeds. Note that although we did not fit a model for

293 female recombination phenotypes, EBVs can still be predicted for females as they are connected
294 to males in the pedigree.

295 Genome-wide association studies

296 *Genotype Imputation*

297 As an imputation reference panel, we used the whole-genome sequence VarGoats data
298 presented in [30] and retrieved from https://www.goatgenome.org/home.html. This data
299 consisted of 1,372 goats out of which 252 European samples closely related to our study
300 samples were used (from the Alpine, Fossés, Lorraine, Poitevine, Provençale, Pyrénées, Rove,
301 Savoie and Saanen breeds).

302 SNP filtering was then applied on the resulting VCF based on several criteria, using bcftools
303 (http://www.htslib.org/doc/1.1/bcftools.html). First, we removed SNPs having discordant
304 reference alleles between the SNP array manifest and the resequencing data, and SNPs showing
305 strictly less than two counts of the minor allele across all selected samples or with a proportion
306 of missing genotypes greater than 10%. Then, SNPs were filtered based on the distribution of
307 the following GATK 3.6 [39] variant-calling quality parameters: DP, FS, VQSLOD (**See**
308 **Additional File 1, Figure S8**). The chosen threshold for SNP removal were DP < 10,000; DP
309 > 25,000; FS > 50; VQSLOD < 10 and VQSLOD > 25. The resulting number of SNPs
310 remaining was 22,741,231 out of which 43,358 SNPs are present on the V1 GoatSNP50
311 Illumina chip. This corresponds to removing 5.6% of the chip SNPs.

312 The Beagle4 software [40] was used to impute the few missing genotypes of the sequence panel
313 and correct the called genotypes by using the genotype likelihood (*gl*) method, a modelscale of
314 1.5 [41], an effective population size of 1000 and the recombination maps estimated using a
315 pedigree-based approach *(see above)*. To assess the quality of the resulting panel genotypes, we
316 leveraged information on 91 animals for which both sequencing and SNP genotyping data were

317 available: we computed R^2 and genotype concordance before and after imputation. Both metrics
318 were highly improved by the imputation procedure: mean R2 increased from 0.928 to 0.990
319 and mean genotype concordance from 0.974 to 0.997. (**See Additional File 1, Figures S9 &**
320 **S10**). Phasing of the reference panel was done using ShapeIt4 [42]. As some animals from the
321 reference panel were previously phased using pedigree information, their phased haplotypes for
322 the SNP chip markers were used as scaffold reference for phasing with ShapeIt4. The resulting
323 VCF file, phased and imputed, was used as reference panel for imputation.

324 Shapeit4 was also run on the phased VCF of 7,457 samples produced by Yapp from the SNP
325 array data to solve the few unphased genotypes and impute the few remaining missing
326 genotypes. The resulting fully phased 50k chip data was imputed to sequence using the Impute5
327 software [43]. Recombination maps were provided to increase power and N_e was set at 1000.
328 Impute5 provides a measure of confidence for all imputed markers (imputation R2) which was
329 used in later steps to filter SNPs *(see below).*

330 ***Statistical Analysis***

331 GWAS were performed separately in each breed using the predicted EBV for male GRR and
332 $\overline{r}_{intra}$ as phenotypes: 589 samples were used in Alpine and 254 in Saanen (Table 1). SNPs with
333 MAF<5% and imputation R2<0.3 were removed from the analysis leaving N=15,988,092 and
334 15,734,503 SNPs in Alpine and Saanen respectively. A univariate linear mixed model was used
335 for testing SNP association with the phenotype using the Gemma software [44], with the "leave
336 one chromosome out" (LOCO) strategy (*i.e.* the genetic relationship matrix (GRM) used to
337 analyse a chromosome is computed on all other chromosomes except this one).

338 Significance was assessed using the ash method [45] implemented in the "ashr" R package, by
339 calculating s-values for each SNP. In the same way as for q-values [37], thresholding s-values
340 at level α gives a set of significant features so that a proportion of α is expected to be false-

341 positives. While [37]'s FDR gives indication on the confidence that effects are different from
342 zero, [45]'s False Sign Rate (FSR) gives indication on the confidence of the *sign* of effects. As
343 this method takes into account the β estimates as well as their standard errors, the resulting s-
344 values can induce a reranking of SNPs compared to methods only taking p-values into account.
345 All SNPs with s-value<0.01 were declared significant and SNPs with s-value<0.05 were
346 declared suggestive.

347 Correlations of SNP effects between breeds were measured by the correlation of β estimates
348 for SNPs that were significant in at least one of the two breeds, and using standard errors as
349 weights.

350 ***Annotation of GWAS loci***

351 A GWAS locus was defined from pairwise Linkage Disequilibrium (R2) between all significant
352 SNPs on a given chromosome. R2 calculation was done using the PLINK 1.9 software.
353 Significant SNPs were considered to belong to the same peak if their R2 was equal or superior
354 to 0.2. For the lead SNP of each peak, a high LD area was defined as follows: an area whose
355 boundaries correspond to the SNP furthest away in R2>0.6 from the lead variant, on both sides
356 of the lead variant. We aimed at evaluating the evidence for association of all genes present in
357 Ensembl (gff3 release 111). As this annotation did not include two major genes associated with
358 the recombination process, *RNF212* and *PRDM9*, we manually added them to the annotation
359 file after blasting their bovine gene sequence on ARS1 (**See Additional File 3, Tables S2 &**
360 **S3**).

361 For all genes, we evaluated the following criteria: if the gene is within 2Mb of the lead variant;
362 if it is the closest gene to the right, to the left or both; if it is the closest gene to another
363 significant SNP of the peak and if it belongs within the "R2>0.6" area of the peak. Based on

364 this we calculated a "candidate score", ranging from 0 to 5, as the number of criteria satisfied
365 by the gene.

366 Results

367 Recombination landscape

368 After quality control, genotype data on 7,457 animals led to the detection of 204,727 crossovers
369 in 6,204 meioses from 843 parents. As a large portion of the dataset comes from a dairy
370 selection program, most genotyped parents were males. As males also typically have more
371 genotyped offspring than females, there are 10 times more male crossovers than female ones in
372 the recombination data (Table 1). The software used to detect crossovers provides the
373 information on the parts of the genome that are informative for the detection of crossovers (See
374 Methods). Overall, all meioses had more than 94% of the genome informative for crossovers
375 (**See Additional File 1, Figure S11**). Females had slightly more informative meioses than
376 males (mean=0.99% vs 0.98%). Indeed, although females are usually less well connected in the
377 pedigree than males, the ones that passed our quality control criteria for meioses (**see Methods**)
378 tended to be better connected than the average male.

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

Breed	Sex	Headcount	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

380 Chromosome-wide rates

381 A linear model was used to test differences in genetic map size between populations. Breed and
382 sex parameters were found significant (sex effect p-value = 10^{-61}; breed effect p-value =
383 10^{-13}; interaction p-value = 0.52). The size of the genetic map was estimated at 35.1M for

384 Alpine males, 34.0M for Saanen males, 30.5M for Alpine females and 29.0M for Saanen
385 females (**See Additional File 1, Figure S12**). Therefore, the genetic map is slightly shorter in
386 Saanen than in Alpine and much shorter in females than males.

387 The mean recombination rate of each chromosome was estimated with the Bayesian model, and
388 a linear model depending on the log size of the chromosomes was fit on the estimates (**Fig. 2a**).
389 Again, a significant difference between breeds was found (p-value = 0.01) but, as the effect is
390 very small, the figure shows breed-averaged estimates to avoid over-plotting. A significant
391 difference between sexes is also observed (p-value $\approx 10^{-48}$): males overall have a higher
392 chromosomic recombination rate than females, and the recombination rate is greater on small
393 chromosomes than big chromosomes for both sexes. Overall, the chromosomic recombination
394 rate has a log relation with the chromosome size. It is observed that the difference in
395 recombination between the sexes is greater on small chromosomes than on big chromosomes,
396 on which recombination rates are almost alike.

397 Chromosome landscape was calculated from 1Mb resolution Bayesian estimates of
398 recombination rate: all positions in Mb were converted to positions relative to the centromere,
399 and recombination rates of all chromosomes were studied depending on their position to the
400 centromere. Considering that all goat autosomes are acrocentric, the beginning of all
401 chromosomes (or centromeric tip) is located at 0%, and the end (or telomeric tip) is located at
402 100%. The results are presented in **Fig. 2b**. Females have a much flatter recombination pattern
403 along the chromosomes compared to males, and both sexes seem to recombine as much in the
404 middle of chromosomes.

405 Local recombination rates

406 Recombination rates were estimated locally on 1Mb intervals on all autosomes (**See Additional**
407 **File 4, Table S4**). Differences between sexes were estimated as the log ratio of male to female

408 recombination rate and significance was assessed using its empirical null distribution calculated
409 from permutations (see Methods). Recombination maps for each chromosome are available in
410 **Additional File 5** (**Figures S18 to S46**), and **Fig. 2c** illustrates this inference on chromosome
411 25. It exhibits the general trend of the sex differences in recombination patterns where males
412 recombine significantly more than females at the tips of the chromosomes. As all goat
413 autosomes are acrocentric, this means that male recombination rate is higher not only at the
414 telomeric end, but also at the centromeric end. A total of 871 intervals of the genome are
415 significant for sex differences (out of 2477 intervals). Of the significant intervals, 559 (64% of
416 the significant intervals) show higher recombination rate in males and are localized for the most
417 part at both ends of chromosomes. The 312 remaining significant intervals (36% of the
418 significant intervals) show higher recombination rate in females and are located in the middle
419 of chromosomes. For a list of all significant intervals, see **Additional File 6, Table S5**.

420 **[Figure 2 here]**

421 Genetic parameters and architecture of recombination phenotypes

422 *Phenotype distributions*

423 Two phenotypes were used to characterize the genetic basis of recombination in the goat: the
424 individual Genome-wide Recombination Rate (GRR), a measure of recombination *intensity* and
425 intra-chromosomal shuffling ($\overline{r}_{intra}$) a measure of the *distribution* of crossovers on
426 chromosomes. This is because crossovers located in the middle of a chromosome lead to create
427 more novel allelic combinations and therefore increase intra-chromosomal shuffling while
428 crossovers located at the tips of a chromosome decrease intra-chromosomal shuffling.

429 GRR is defined as the per meiosis crossover count divided by the length of informative genome.
430 The average value of GRR over all meioses is 1.39 cM per Mb. The distributions of GRR for
431 different breeds and sexes are shown in **Figure S13** (**See Additional File 1**).

Intra-chromosomal shuffling $\overline{r}_{intra}$ was defined as stated in Material & Methods. The average value of $\overline{r}_{intra}$ is 0.011. The distributions of $\overline{r}_{intra}$ between breeds and sexes are shown in **Figure S14** (**See Additional File 1**). A linear model was fitted to estimate differences between populations. A representation of the estimates can be found in **Figure S15** (**See Additional File 1**). No significant difference was found between sexes (p-value = 0.12) but Alpine goats had significantly higher $\overline{r}_{intra}$ (p-value = 0.03). Furthermore, the interaction between breed and sex was found significant (female Alpine $+4.01 \times 10^{-4}$; p-value = 0.03), meaning that in average, Alpine females shuffled more than Alpine males, but Saanen females shuffled less than Saanen males.

Heritability estimates

Heritabilities of GRR and $\overline{r}_{intra}$ 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. $\overline{r}_{intra}$ heritability was estimated at 0.034±0.015 when both breeds were analysed together. No significant permanent environment effect was detected for any phenotype (**Table 2**).

Table 2: Variance components and heritability (SE) estimates for male genome-wide recombination rate and intra-chromosomal shuffling in a population of 843 Alpine and Saanen goats.

Phenotype	Population	σ_A^2	σ_B^2	σ_E^2	τ^2	h^2
GRR	All	0.50 (0.12)	0.05 (0.09)	3.52 (0.07)	0.01 (0.02)	0.12 (0.03) [a]
	Alpine	0.47 (0.16)	0.11 (0.13)	3.57 (0.09)	0.03 (0.03)	0.11 (0.04) [a]
	Saanen	0.47 (0.17)	0.00 (0.13)	3.42 (0.11)	0.00 (0.03)	0.12 (0.04) [a]
$\overline{r}_{intra}$	All	0.002 (0.001)	0.000 (0.001)	0.052 (0.001)	0.000 (0.01)	0.034 (0.015) [a]
	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)

Heritability was considered significant when ($h^2 \pm 2SE$) did not include 0. τ^2 = permanent environment effect; [a] = significant heritability. With σ_A^2 the genetic variance; σ_B^2 the permanent environment variance and σ_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 $\overline{r}_{intra}$. GWAS results using imputed genotypes to the sequence level are illustrated in **Fig. 3** for Alpine and Saanen, on both phenotypes.

The analyses on GRR revealed 15 significant QTLs in Alpine and 4 suggestive QTLs in Saanen. For Alpine, there were one QTL on chromosomes 1, 6, 7, 13 and 29; three QTLs on chromosomes 5 and 10; as well as two QTLs on chromosomes 9 and 17. For Saanen, there were one suggestive QTL on chromosomes 2, 4, 6 and 14. There is no overlap between the significant QTLs in Alpine and the suggestive QTLs in Saanen. Correlation of SNP effects between the two breeds for this phenotype was calculated using standard errors as weights. It amounted to 0.25.

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

473 within 2Mb of suggestive QTLs for GRR. Correlation between the two GWAS in Saanen is
474 0.92.

475 **[Figure 3 here]**

476 Considering the GWAS on GRR, the most significant QTL in Alpine is located on chromosome
477 17 at around 7Mb (lead variant s-value=0.00236) and is illustrated in **Fig. 4a**. In this region, 58
478 candidate genes can be found, including *SSH1* which scores 5/5. The significant SNPs are
479 located directly within the *SSH1* gene, in an intronic region. The lead variant has a MAF of 0.49
480 in the population. The QTL illustrated in **Fig. 4b** contains significant SNPs in high LD in a
481 region of 300kb and is rich in genes. It is the second most significant QTL (lead variant s-
482 value=0.00238). Our standardised method for ranking candidate genes identifies 86 genes with
483 a score of at least 1/5. The highest scoring (3/5) is *CARMIL3* because of its physical proximity
484 to the lead variant. Among the candidates for this region are two genes previously associated
485 with recombination: *RNF212B* (rank 17, score 1/5) and *REC8* (rank 51, score 1/5). The QTL
486 illustrated in **Fig. 4c** is located at the end of chromosome 6 and is the 6th most significant of the
487 genome. It contains 36 candidate genes including *SPON2* (rank 1, score 4/5) and *RNF212* (rank
488 2, score 3/5).

489 In Saanen, the most suggestive QTL can be found on chromosome 6 at around 12.5Mb (lead
490 variant s-value=0.0326) and is illustrated in **Fig. 4d**. Several significant SNPs of the area fall
491 within the *ANK2* gene which scores at 4/5 as a result. Eight other candidate genes can be found
492 in the region. Other notable genes in both breeds can be found in Table 3. For exhaustive results
493 for GWAS candidate genes in both breeds, see **Additional File 7**, **Tables 6 & 7**.

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

Breed	QTL of interest (chromosome: position of the lead variant)	Candidate gene (score/5)
Alpine	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)
	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)
Saanen	2:29863833	TNS1 (5/5)
	6:12545797	ANK2 (4/5)
	14:91668757	B3GAT2 (5/5)

Considering the GWAS on $\overline{r}_{intra}$, the most significant QTL 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 the recombination process. Another significant QTL can be found for this phenotype and breed at the end of chromosome 1 (**Fig. 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 QTL for $\overline{r}_{intra}$ 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 the recombination process. For exhaustive candidate genes results for GWAS on $\overline{r}_{intra}$, see **Additional File 7, Tables 8 & 9**.

[Figure 4 here]

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,

512 with rates varying from 0.21 to 10.32cM/Mb. An important feature of the recombination
513 landscape is the large heterochiasmy in some genomic regions, mostly located at chromosome
514 ends. This is consistent with patterns of recombination in a closely related species, the sheep
515 *Ovis aries* [12,19].

516 In a second part of this study, we characterized the genetic architecture of recombination
517 intensity and localisation through the study of two recombination phenotypes, GRR and $\overline{r}_{intra}$.
518 This analysis revealed low but significant genetic components to these traits and identified
519 several QTLs for both traits, some of which match loci found in other species while others are
520 new. Overall, our study adds to the growing knowledge on recombination and will contribute
521 to understanding the evolution of recombination, a fundamental trait for individual fitness and
522 adaptive potential of populations. We now discuss the importance of our findings and the future
523 research questions it suggests.

524 Studying recombination from livestock breeding programmes data

525 Livestock breeding populations are good models to study the genetic architecture of the
526 recombination process as they offer large datasets of genotyped individuals related through well
527 characterised pedigrees. An important caveat of these datasets however lies in the unbalanced
528 genotyping between sexes. In ruminant programs of genomic selection such as the one used
529 here, male selection candidates are typically all genotyped while females are much less likely
530 to be, although this tends to change with the decrease in genotyping costs. In our dataset, the
531 number of dams was very small. We still managed to correctly characterise heterochiasmy
532 along the genome, but the number of dams was too small to study the genetic architecture of
533 recombination in this sex. We expect the number of genotyped females to grow in the future
534 which will allow studying genetic determinants of heterochiasmy in this species. However, it

535 should be noted that offspring size will still typically be smaller for females, so that taking into
536 account uncertainty as we did here will remain relevant.

537 Recombination landscape of the goat genome

538 Our study employs a similar strategy as previous studies in animal populations [8,10–12,16,18–
539 20,22,46], but we aimed at better accounting for the uncertainty in crossover data. The first
540 source of uncertainty arises from the unprecise localisation of crossovers due to the genotyping
541 density, the second by the homozygosity of individuals in the pedigree, and the third concerns
542 only female recombination rate estimation, as our available female data is much scarcer. To
543 tackle the first problem, [12] proposed a random sampling approach of crossover locations to
544 average recombination rate estimation over crossover uncertainty. Here, we used the same
545 strategy but extended it further to better identify genomic regions with significant differences
546 in recombination rates between sexes. To tackle the second problem, we made use of an
547 information provided by the software used to phase individuals (Yapp) that explicitly identifies
548 regions of the genome where meioses are uninformative for crossover detection. In order to
549 integrate both sources of uncertainty, we leveraged the Bayesian Gamma-Poisson model of [12]
550 that allows to incorporate this information naturally. One caveat of this model is that it is much
551 harder to implement if one wants to add a hierarchical level to account for differences between
552 groups (*e.g.* sex) but with non-null covariances. Recent developments around Poisson-Log-
553 Normal models [47] could be used for this purpose (see e.g. [48] in the context of recombination
554 rate inference) but this will require further developments, in particular to better account for the
555 longitudinal correlation of recombination rates along the genome. Finally, we proposed a
556 permutation method to assess the significance of heterochiasmy along chromosomes. This
557 method allows to draw conclusions from a lowly informative dataset in terms of sex differences,
558 which could not have been confidently untangled otherwise. We believe it could be useful in
559 future studies on heterochiasmy.

560 The average recombination rate of the goat genome (1.42cM/Mb varying from 1.12 to
561 1.71cM/Mb among chromosomes) is very similar to other ruminants (e.g. 1.5cM/Mb in sheep
562 [12], 0.98cM/Mb in cattle [46]). A striking feature of its landscape is a strong heterochiasmy
563 that we could reveal even if the number of female meioses available in our datasets was
564 relatively low. This heterochiasmy is mostly concentrated in a few megabases at the
565 chromosome ends, irrespective of the total chromosome size. As a consequence, the difference
566 in average recombination rate between sexes is smaller for large chromosomes than small ones
567 (**Fig. 2a**). The biological causes underlying this heterochiasmy are still unknown. Further
568 research combining genetics studies with cytogenetics and functional studies will most likely
569 be required to unravel them.

570 Patterns of recombination and heterochiasmy in the goat match those found in sheep [12,19,48]
571 and cattle [8,10,11] although maybe to a lesser extent for this last species. Given that these three
572 species have highly colinear genomes, it seems possible to make a quantitative comparison of
573 recombination patterns along ruminant chromosomes to reveal how recombination evolved
574 with speciation.

575 Genetic parameters and architecture of recombination phenotypes

576 We studied two different phenotypes: the Genome-Wide Recombination Rate (GRR) which is
577 a measure of recombination intensity, and intra-chromosomal shuffling ($\overline{r}_{intra}$) which depends
578 on recombination intensity too but also measures differences in the localisation of crossovers.
579 As expected, we observed a high correlation between the two phenotypes (0.62). However,
580 despite strong correlation of the GWAS results on GRR and $\overline{r}_{intra}$, their genetic architectures
581 were found only partially overlapping, illustrating that they are probably capturing different
582 underlying biological processes (for example those involved in Double Strand Break (DSB)
583 localisation from those involved in DSB repair).

584 Heritability for male GRR was estimated at 0.12 with similar results for each breed, and similar
585 to those found in other ruminant species for this phenotype (ACC or GRR): *e.g.* 0.22 for Dutch
586 Holstein-Friesan males in cattle [6], 0.13 in several cattle breeds [9], 0.12 for Soay sheep [19]
587 and 0.23 for Lacaune sheep [12].

588 The heritability for male $\overline{r}_{intra}$ was estimated at 0.034 and consistent between the two breeds,
589 although only slightly significantly different from zero (SE=0.015). Heritability for this
590 phenotype seems to be smaller than in other species as [18] estimated it at 0.11 in salmon, a
591 species with *PRDM9* hotspots, and [22] estimated it at 0.14 in sparrows, a species without
592 *PRDM9* hotspots.

593 The animal models used for estimating genetic parameters also allow to predict the additive
594 genetic values (EBVs) of every animal in the pedigree. These EBVs on male GRR and $\overline{r}_{intra}$
595 were used as phenotypes for our GWAS. This permits to have only one measure of phenotype
596 for each individual, and therefore to run analyses using the fast computing procedures required
597 for whole genome sequence imputed datasets. Using EBVs as phenotypes for GWAS is
598 sometimes discouraged for different reasons. First, to avoid too much shrinkage in the
599 phenotypes: ideally all animals should have either own and/or direct offspring phenotypes.
600 Here, out of the 843 individuals in the GWAS, only 323 (38%) did not have any own
601 phenotypes. These were all females, as they obviously cannot be directly measured for male
602 recombination phenotypes. Among these females, only 43 (5% of all animals) were not directly
603 related to a male with own information for these phenotypes, *i.e.* had a coefficient of
604 relationship smaller than ½ with measured samples. So overall, our dataset meets this first
605 requirement. Another caveat of using EBVs as phenotype is that their precisions can sometimes
606 vary widely between individuals, especially when there are large differences in their number of
607 repeated measures. In our case, the posterior variances of EBVs ranged from 2.6 to 7.1, *i.e.* a

608 factor smaller than 3 between the most and least precise predictions. Also, although precisions
609 of females EBVs tended to be smaller than males their distribution was unimodal (**See**
610 **Additional File 1, Figures S16 & S17**). Given this homogeneity in the precision of EBVs, we
611 did not deem deregression, a procedure that comes with its own caveats, necessary. Overall, for
612 this dataset we believe that phenotypic records are sufficiently informative and homogeneously
613 distributed among animals to use EBVs as phenotypes. Finally, we note that we used the
614 pedigree relationship matrix to derive EBVs while the GRM was used in the GWAS, in addition
615 to using the LOCO methodology, avoiding potential convergence issues (see **Methods**).

616 GWAS results for male GRR and $\overline{r}_{intra}$ showed low genetic correlation between Alpine and
617 Saanen. Despite identical heritability results, no QTL were found in common between the two
618 breeds. Besides, the correlation of the $\hat{\beta}$ for the SNPs that are significant in at least one breed
619 is only 0.25 for GRR and 0.04 for $\overline{r}_{intra}$. This suggests strong differences in the underlying
620 genetic determinants of recombination between these two breeds, although they are both
621 European breeds originating from the Alps and closely related in a worldwide diversity panel
622 [49]. Despite this relatively close genetic proximity, those breeds have been shown to have
623 different genetic architecture for other traits. For example, [50] found genetic heterogeneity in
624 the determinism of dairy-related traits between the two breeds, despite similar breeding goals.
625 Our results add to the literature that suggests that genome-wide recombination landscapes tend
626 to be mostly shared between populations while the underlying genetic determinants of the
627 process can be quite different, *e.g.* [12] in sheep, [15,16] in pigs, although results in cattle tend
628 to show they are more conserved [10].

629 Candidate genes associated with QTL loci in the GWAS were prioritised by implementing a
630 scoring methodology using criteria based on physical and genetic distance to significant SNPs.
631 This methodology was developed in order to provide an objective measure for the investigation

632 of candidate genes underlying GWAS peaks and quickly highlight strong candidates in regions
633 rich in annotations. For instance, 13Mb on chromosome 5 contain three different peaks in rapid
634 succession and 141 genes are annotated in the region. The scoring allowed to quickly identify
635 the most promising genes and order them all depending on their physical and genetic proximity
636 to each peak. Genes involved in meiosis in other species were searched for in a second step to
637 identify functional candidates at QTLs among the positional candidates.

638 The GWAS results for GRR in Alpine goats show a signal towards the end of chromosome 6,
639 in very close proximity to the *RNF212* gene. *RNF212* has been found to be associated with
640 GRR in humans [51], cattle [6,9] and sheep [12,19]. It has also been proved that *RNF212* is
641 involved in and is necessary for meiosis in mammals by playing a central role to the designation
642 of crossover sites and the formation of crossover-specific recombination complexes [52,53].
643 Yet it only scores second candidate for the signal on chromosome 6 because it does not directly
644 contain the significant SNPs. Of course, this is still consistent with a regulatory variant affecting
645 *RNF212* expression.

646 One signal with significant SNPs as far as 330kb apart can be found at the end of chromosome
647 10 between 79.9 and 80.3Mb in the Alpine breed (**Fig. 4b**). The genes *RNF212B* and *REC8* are
648 in the region. They were previously suspected to be responsible for signals in sheep [12,19] and
649 cattle [6,9] but the studies had difficulty attributing the signal to one of the two genes due to a
650 lack of resolution. Among our candidates, *RNF212B* ranks 17th and *REC8* ranks 51st, as the
651 region is very rich in annotated genes. Functionally, both these genes make very good
652 candidates. On the one hand, *RNF212B* is a homolog of *RNF212*, this gene's involvement in
653 recombination being described just above, and the two proteins *rnf212* and *rnf212b* have been
654 showed to interact with each other and to be necessary for their stability during the
655 recombination process [54]. On the other hand, REC8 is also known to be involved in meiotic

656 recombination [55] through DSB formation [56] as well as in implementing homologue bias
657 during DSB repairs [57–59]. But the gene richness of the region (62 genes in 690kb), their
658 relative physical distance to the lead variant and the fact that, surprisingly, the main part of the
659 signal is located in an area of 250kb completely devoid of annotated protein coding genes,
660 makes discriminating the two candidates, or even determining if they are good candidates, quite
661 difficult.

662 Among significant QTLs, the most significant signal was found in Alpine on chromosome 17
663 around 7Mb (**Fig. 4a**). The signal falls directly within the gene *SSH1*, which scored 5/5
664 according to our ranking method. This gene belongs to the slingshot phosphatase family,
665 constituted of three genes [60]. Investigation of the GTEx database [61,62] showed that it is
666 expressed in testis and automatic annotations [63,64] indicate it may be involved in chromatin
667 remodelling mechanisms. Furthermore, its paralog *SSH2* showed protein-enrichment in a study
668 aiming to predict mouse meiosis-essential genes [65] and is known to be involved in later
669 spermatogenesis [66]. But current knowledge of these genes' functions as well as the strong
670 signal found for a QTL within *SSH1* suggest this family may be involved in recombination.

671 When finding QTLs of interest at the end of chromosome 1 in Alpine, both for GRR and $\overline{r}_{intra}$,
672 one could have assumed that these signals were associated with *PRDM9*, known regulator of
673 recombination hotspots [67,68] and located at the very end of chromosome 1 in goats (**See**
674 **Additional File 3**, **Table 1**). Therefore, it was surprising to see that *PRDM9* actually scores
675 very low for the GRR signal (1/5 score) due to its physical distance from the significant QTL
676 and is even too far from the $\overline{r}_{intra}$ signal to gain any score point. It could be explained that no
677 association was found between *PRDM9* and GRR because GRR is a phenotype of intensity of
678 recombination, and *PRDM9* regulates the location of recombination. But $\overline{r}_{intra}$ is a
679 recombination phenotype linked with the position of crossovers and the association was even

more expected, especially considering that [17] found a clear signal for *PRDM9* in their own GWAS on pig $\overline{r}_{intra}$ phenotype. Furthermore, the third highest ranking candidate gene in goats for $\overline{r}_{intra}$ is *KCNJ15* (2/5 score and **Fig. 4e**). Although this gene seems to have no link with recombination, an association for this gene with GRR has previously been found in sheep [12]. Furthermore, [6] found signal for the genes *KCNJ2* and *KCNJ16*, paralogs of *KCNJ15*, in their GWAS for GRR. Thus, further investigation on the involvement of *KCNJ15* in recombination may be relevant.

Conclusions

In this study, we have found that recombination patterns are very alike between the Alpine and Saanen breeds, but heterochiasmy is strongly present, manifested at the chromosome ends. The recombination pattern we observed is well conserved with closely related species like sheep and cattle. We also showed that male GRR and male $\overline{r}_{intra}$ are heritable and that this value is conserved in the two breeds. Heritability for GRR is concordant with results found in sheep and cattle, but goat heritability for $\overline{r}_{intra}$ seems to be lower than in the other species in which it was estimated. GWAS results showed low correlation between the two breeds, for both phenotypes, and it seems that despite similar recombination patterns and same heritability for recombination traits, the underlying genetic architecture of recombination is different between Alpine and Saanen. Candidate genes found in GWAS are partially in common with what was previously found in literature, with some novel genes. This supports the hypothesis 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 and scripts for analysis are available at https://doi.org/10.57745/HPYLLT. Sequence data for the imputation reference panel can be obtained from the VarGoats consortium: https://www.goatgenome.org/vargoats.html.

Competing interests

The authors declare that they have no competing interests.

Funding

Genotype data from the genomic selection programme have received funding from: the French Genomcap and Phenofinlait programmes (ANR, Apis-Gène, CASDAR, FranceAgriMer, France Génétique Elevage, the French Ministry of Agriculture Agrifood, and Forestry), CASDAR (Genovicap et Maxi'male). Genotype data from the INRAE research database have received funding from: UE 7th framework programme ("3SR" grant #245140), Region Centre-Val de Loire (CAPRIMAM grant) and APISGENE (ACTIVE GOAT grant). AE was funded by the INRAE Animal Genetics and Ecology and Biodiversity divisions.

Authors' contributions

All authors participated in conceptualizing this study and to the data curation and analyses. 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. All authors validated the

results. AE and BS did visualisation work through diverse presentations and wrote the original draft. All authors reviewed, edited and approved the final manuscript.

Acknowledgements

Many numerical analyses were performed on the genobioinfo bioinformatics platform Toulouse Midi-Pyrénées (Bioinfo Genotoul). We are thankful to Capgenes and the INRAE P3R unit (https://doi.org/10.15454/1.5483259352597417E12) for providing access to SNP array genotypes, and to the VarGoats consortium for early access to whole genome sequencing data. We would like to thank Susan Johnston, Tristan Mary-Huard and Leopoldo Sanchez Rodriguez for fruitful discussions during the production of these results.

References

1. Koehler KE, Hawley RS, Sherman S, Hassold T. Recombination and nondisjunction in humans and flies. Human Molecular Genetics. 1996;5:1495–504.

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

3. Crow JF, Kimura M. Evolution in Sexual and Asexual Populations. The American Naturalist. 1965;99:439–50.

4. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, et al. A high-resolution recombination map of the human genome. Nat Genet. 2002;31:241–7.

5. Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. Science. 2008;319:1395–8.

6. Sandor C, Li W, Coppieters W, Druet T, Charlier C, Georges M. Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle. Paigen K, editor. PLoS Genet. 2012;8:e1002854.

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

8. Ma L, O'Connell JR, VanRaden PM, Shen B, Padhi A, Sun C, et al. Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. Auton AJ, editor. PLoS Genet. 2015;11:e1005387.

9. Kadri NK, Harland C, Faux P, Cambisano N, Karim L, Coppieters W, et al. Coding and noncoding variants in *HFM1* , *MLH3* , *MSH4* , *MSH5* , *RNF212* , and *RNF212B* affect recombination rate in cattle. Genome Res. 2016;26:1323–32.

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

11. Brekke C, Johnston SE, Gjuvsland AB, Berg P. Variation and genetic control of individual recombination rates in Norwegian Red dairy cattle. Journal of Dairy Science. 2023;106:1130–41.

12. Petit M, Astruc J-M, Sarry J, Drouilhet L, Fabre S, Moreno CR, et al. Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. Genetics. 2017;207:767–84.

13. Weng Z, Wolc A, Su H, Fernando RL, Dekkers JCM, Arango J, et al. Identification of recombination hotspots and quantitative trait loci for recombination rate in layer chickens. J Animal Sci Biotechnol. 2019;10:1–13.

14. Johnsson M, Whalen A, Ros-Freixedes R, Gorjanc G, Chen C-Y, Herring WO, et al. Genetic variation in recombination rate in the pig. Genet Sel Evol. 2021;53:54.

15. Lozada-Soto EA, Maltecca C, Wackel H, Flowers W, Gray K, He Y, et al. Evidence for recombination variability in purebred swine populations. J Anim Breed Genet. 2021;138:259–73.

16. Brekke C, Berg P, Gjuvsland AB, Johnston SE. 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. 2022;54:33.

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

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

19. Johnston SE, Bérénos C, Slate J, Pemberton JM. Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis aries*). Genetics. 2016;203:583–98.

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

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

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

23. Cooney CR, Mank JE, Wright AE. Constraint and divergence in the evolution of male and female recombination rates in fishes. Evolution. 2021;75:2857–66.

24. Veller C, Kleckner N, Nowak MA. 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. 2019;116:1659–68.

25. Schibler L, Vaiman D, Oustry A, Giraud-Delville C, Cribiu EP. Comparative Gene Mapping: A Fine-Scale Survey of Chromosome Rearrangements between Ruminants and Humans. Genome Res. 1998;8:901–15.

26. Fröhlich J, Vozdova M, Kubickova S, Cernohorska H, Sebestova H, Rubes J. Variation of Meiotic Recombination Rates and MLH1 Foci Distribution in Spermatocytes of Cattle, Sheep and Goats. Cytogenet Genome Res. 2015;146:211–21.

27. Muñoz-Fuentes V, Marcet-Ortega M, Alkorta-Aranburu G, Linde Forsberg C, Morrell JM, Manzano-Piedras E, et al. Strong Artificial Selection in Domestic Mammals Did Not Result in an Increased Recombination Rate. Mol Biol Evol. 2015;32:510–23.

28. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotechnol. 2013;31:135–41.

29. Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and Characterization of a 52K SNP Chip for Goats. Liu Z, editor. PLoS ONE. 2014;9:e86227.

30. Denoyelle L, Talouarn E, Bardou P, Colli L, Alberti A, Danchin C, et al. VarGoats project: a dataset of 1159 whole-genome sequences to dissect Capra hircus global diversity. Genet Sel Evol. 2021;53:86.

31. Du X, Servin B, Womack JE, Cao J, Yu M, Dong Y, et al. An update of the goat genome assembly using dense radiation hybrid maps allows detailed analysis of evolutionary rearrangements in Bovidae. BMC Genomics. 2014;15:625.

32. Schensted, C. Longest increasing and decreasing subsequences. Can J Math. 1961;13:179–91.

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

34. Servin B. Yapp: a suite of tools to analyze genomic data in pedigrees [Internet]. INRAE; 2024. Available from: https://yapp.readthedocs.io

35. Favier A, Elsen J-M, de Givry S, Legarra A. Optimal haplotype reconstruction in half-sib families. In: Dovier A, Dal Palù A, Will S, editors. Edinburgh, UK: EasyChair; 2010 [cited 2022 May 5]. p. 27–37. Available from: https://easychair.org/publications/paper/SpJ

36. Druet T, Georges M. LINKPHASE3: an improved pedigree-based phasing algorithm robust to genotyping and map errors. Bioinformatics. 2015;31:1677–9.

830 37. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad
831 Sci USA. 2003;100:9440–5.

832 38. Meyer K. WOMBAT—A tool for mixed model analyses in quantitative genetics by
833 restricted maximum likelihood (REML). J Zhejiang Univ - Sci B. 2007;8:815–21.

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

836 40. Browning BL, Browning SR. Genotype Imputation with Millions of Reference Samples.
837 The American Journal of Human Genetics. 2016;98:116–26.

838 41. Pook T, Mayer M, Geibel J, Weigend S, Cavero D, Schoen CC, et al. Improving Imputation
839 Quality in BEAGLE for Crop and Livestock Data. G3 Genes|Genomes|Genetics. 2020;10:177–
840 88.

841 42. Delaneau O, Zagury J-F, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable
842 and integrative haplotype estimation. Nat Commun. 2019;10:5436.

843 43. Rubinacci S, Delaneau O, Marchini J. Genotype imputation using the Positional Burrows
844 Wheeler Transform. Scheet P, editor. PLoS Genet. 2020;16:e1009049.

845 44. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies.
846 Nat Genet. 2012;44:821–4.

847 45. Stephens M. False discovery rates: a new deal. Biostat. 2016;kxw041.

848 46. Qanbari S, Wittenburg D. Male recombination map of the autosomal genome in German
849 Holstein. Genet Sel Evol. 2020;52:73.

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

852 48. Vassilief H, Id Bella M, Hazard D, Tortereau F, Faraut T, Johnston SE, et al. 202. Sex
853 differences in recombination maps are associated with differential hotspot usage in sheep.
854 Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)
855 [Internet]. Rotterdam, the Netherlands: Wageningen Academic Publishers; 2022 [cited 2025
856 Feb 12]. p. 861–4. Available from: https://www.wageningenacademic.com/doi/10.3920/978-
857 90-8686-940-4_202

858 49. Colli L, Milanesi M, Talenti A, Bertolini F, Chen M, Crisà A, et al. Genome-wide SNP
859 profiling of worldwide goat populations reveals strong partitioning of diversity and highlights
860 post-domestication migration routes. Genet Sel Evol. 2018;50:58.

861 50. Martin P, Palhière I, Maroteau C, Clément V, David I, Klopp GT, et al. Genome-wide
862 association mapping for type and mammary health traits in French dairy goats identifies a
863 pleiotropic region on chromosome 19 in the Saanen breed. Journal of Dairy Science.
864 2018;101:5214–26.

51. Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, et al. Common and low-frequency variants associated with genome-wide recombination rate. Nat Genet. 2014;46:11–6.

52. Reynolds A, Qiao H, Yang Y, Chen JK, Jackson N, Biswas K, et al. RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. Nat Genet. 2013;45:269–78.

53. Qiao H, Prasada Rao HBD, Yang Y, Fong JH, Cloutier JM, Deacon DC, et al. Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. Nat Genet. 2014;46:194–9.

54. Ito M, Yun Y, Kulkarni DS, Lee S, Sandhu S, Nuñez B, et al. Distinct and interdependent functions of three RING proteins regulate recombination during mammalian meiosis. Proceedings of the National Academy of Sciences. 2025;122:e2412961121.

55. Parisi S, McKay MJ, Molnar M, Thompson MA, van der Spek PJ, van Drunen-Schoenmaker E, et al. Rec8p, a Meiotic Recombination and Sister Chromatid Cohesion Phosphoprotein of the Rad21p Family Conserved from Fission Yeast to Humans. Mol Cell Biol. 1999;19:3515–28.

56. Kugou K, Fukuda T, Yamada S, Ito M, Sasanuma H, Mori S, et al. Rec8 guides canonical Spo11 distribution along yeast meiotic chromosomes. Mol Biol Cell. 2009;20:3064–76.

57. Klein F, Mahr P, Galova M, Buonomo SB, Michaelis C, Nairz K, et al. A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell. 1999;98:91–103.

58. Brar GA, Hochwagen A, Ee LS, Amon A. The Multiple Roles of Cohesin in Meiotic Chromosome Morphogenesis and Pairing. MBoC. 2009;20:1030–47.

59. Kim KP, Weiner BM, Zhang L, Jordan A, Dekker J, Kleckner N. Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. Cell. 2010;143:924–37.

60. Patterson KI, Brummer T, O'brien PM, Daly RJ. Dual-specificity phosphatases: critical regulators with diverse cellular targets. Biochemical Journal. 2009;418:475–89.

61. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.

62. GTEx Portal - SSH1 gene page [Internet]. GTEx Portal. [cited 2025 Aug 14]. Available from: https://www.gtexportal.org/home/gene/SSH1

63. SSH1 Gene - GeneCards | SSH1 Protein | SSH1 Antibody [Internet]. GeneCards. [cited 2025 Aug 14]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=SSH1&keywords=SSH1

64. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. 2023;51:D1003–9.

65. Fang K, Li Q, Wei Y, Zhou C, Guo W, Shen J, et al. Prediction and Validation of Mouse Meiosis-Essential Genes Based on Spermatogenesis Proteome Dynamics. Molecular & Cellular Proteomics. 2021;20:100014.

66. Xu K, Su X, Fang K, Lv Y, Huang T, Li M, et al. The Slingshot phosphatase 2 is required for acrosome biogenesis during spermatogenesis in mice. Yan W, Harper DM, Hammoud S, editors. eLife. 2023;12:e83129.

67. Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. Science. 2010;327:836–40.

68. Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. de Massy B, editor. eLife. 2017;6:e24133.

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. All 29 chromosomes are acrocentric. 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 $\overline{r}_{intra}$ stand respectively for EBVs on male GRR and EBVs on male $\overline{r}_{intra}$.

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

Additional Files

Additional File 1 Figure S1

File format: PDF

Title: Position of (b)ovine (top) and goat (bottom) chip SNPs on both the RH map and the ARS1 assembly, on chromosomes 2 and 18.

Description: Goat SNPs have been filtered using the longest increasing subsequence algorithm, and only remaining SNPs have been represented.

Additional File 1 Figure S2

File format: PDF

Title: Crossover distribution on the 35-45 Mb area of chromosome 1.

Description: Red lines = position of conserved synteny breaks.

Additional File 1 Figure S3

File format: PDF

Title: Additive SNP and sample call-rate cumulative distributions.

Description: Left: SNP call-rate; right: sample call-rate. Black line: chosen threshold=0.01.

Additional File 1 Figure S4

File format: PDF

Title: Distribution of the distance between two crossovers per chromosome per meiosis.
Description: Red line: chosen threshold=5Mb

Additional File 1 Figure S5

File format: PDF

Title: Density distribution of the Maximum Likelihood estimates of recombination rates on 1Mb intervals of the genome.

Description: Black: gamma law; extracted parameters α=3.08 & β=2.20.

Additional File 1 Figure S6

File format: PDF

Title: Distribution of crossover detection interval sizes.

Additional File 1 Figure S7

File format: PDF

Title: Distribution of all permuted recombination rate log ratio for all genomic intervals.

Additional File 1 Figure S8

File format: PDF

Title: SNP distribution of sequencing and variant calling quality parameters.

Description: A: DP parameter; B: FS parameter; C: VQSLOD parameter.

Additional File 1 Figure S9

File format: PDF

Title: R^2 evolution according to allele frequency.

Description: (A) Before genotype imputation. Mean R^2=0.928. (B) After genotype imputation. Mean R^2=0.990.

Additional File 1 Figure S10

File format: PDF

Title: Genotype concordance according to allele frequency.
Description: (A) Before genotype imputation. Mean genotype concordance=0.974. (B) After genotype imputation. Mean genotype concordance=0.997.

Additional File 1 Figure S11

File format: PDF

Title: Distribution the proportion of the autosomal genome informative for crossovers in parents (M=males vs F=females).

Additional File 1 Figure S12

File format: PDF

Title: Comparison of recombination map size estimates between populations.

Description: Confidence interval = $\mu \pm 2 \times standard\ error$.

Additional File 1 Figure S13

File format: PDF

Title: Distribution of genome-wide recombination rates (GRR) in meioses.

Additional File 1 Figure S14

File format: PDF

Title: Distribution of intra-chromosomal shuffling in meioses.

Additional File 1 Figure S15

File format: PDF

Title: Comparison of intra-chromosomal shuffling estimates between populations.

Description: Confidence interval = $\mu \pm 2 \times standard\ error$.

Additional File 1 Figure S16

Additional File 1 Figure S17

Additional File 2 Table S1

File format: XLS

992 Title: SNPs within synteny breakpoints

993 Description: File listing all 123 SNPs removed from dataset after identifying synteny
994 breakpoints with human genome on ARS1 assembly.

995 **Additional File 3 Table S2**

996 File format: XLS

997 Title: BLAST results for the bovine gene sequence RNF212 on ARS1 assembly.

998 **Additional File 3 Table S3**

999 File format: XLS

1000 Title: BLAST results for the bovine gene sequence PRDM9 on ARS1 assembly.

1001 **Additional File 4 Table S4**

1002 File format: XLS

1003 Title: Sex-specific estimated recombination rates for each 1Mb interval of the autosomal goat
1004 genome.

1005 **Additional File 5 Figure S18**

1006 File format: PDF

1007 Title: Estimated recombination rates for males and females on chromosome 1, at 1Mb
1008 resolution.

1009 Description: Log ratio between male and female recombination rates were plotted in the bottom
1010 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1011 significant for sex differences are marked with a red dot on the bottom panel.

1012 **Additional File 5 Figure S19**

1013 File format: PDF

1014 Title: Estimated recombination rates for males and females on chromosome 2, at 1Mb
1015 resolution.

1016 Description: Log ratio between male and female recombination rates were plotted in the bottom
1017 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1018 significant for sex differences are marked with a red dot on the bottom panel.

1019 **Additional File 5 Figure S20**

1020 File format: PDF

1021 Title: Estimated recombination rates for males and females on chromosome 3, at 1Mb
1022 resolution.

1023 Description: Log ratio between male and female recombination rates were plotted in the bottom
1024 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1025 significant for sex differences are marked with a red dot on the bottom panel.

1026 **Additional File 5 Figure S21**

1027 File format: PDF

Title: Estimated recombination rates for males and females on chromosome 4, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S22

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 5, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S23

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 6, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S24

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 7, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S25

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 8, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S26

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 9, at 1Mb resolution.

1065 Description: Log ratio between male and female recombination rates were plotted in the bottom
1066 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1067 significant for sex differences are marked with a red dot on the bottom panel.

1068 **Additional File 5 Figure S27**

1069 File format: PDF

1070 Title: Estimated recombination rates for males and females on chromosome 10, at 1Mb
1071 resolution.

1072 Description: Log ratio between male and female recombination rates were plotted in the bottom
1073 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1074 significant for sex differences are marked with a red dot on the bottom panel.

1075 **Additional File 5 Figure S28**

1076 File format: PDF

1077 Title: Estimated recombination rates for males and females on chromosome 11, at 1Mb
1078 resolution.

1079 Description: Log ratio between male and female recombination rates were plotted in the bottom
1080 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1081 significant for sex differences are marked with a red dot on the bottom panel.

1082 **Additional File 5 Figure S29**

1083 File format: PDF

1084 Title: Estimated recombination rates for males and females on chromosome 12, at 1Mb
1085 resolution.

1086 Description: Log ratio between male and female recombination rates were plotted in the bottom
1087 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1088 significant for sex differences are marked with a red dot on the bottom panel.

1089 **Additional File 5 Figure S30**

1090 File format: PDF

1091 Title: Estimated recombination rates for males and females on chromosome 13, at 1Mb
1092 resolution.

1093 Description: Log ratio between male and female recombination rates were plotted in the bottom
1094 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1095 significant for sex differences are marked with a red dot on the bottom panel.

1096 **Additional File 5 Figure S31**

1097 File format: PDF

1098 Title: Estimated recombination rates for males and females on chromosome 14, at 1Mb
1099 resolution.

1100 Description: Log ratio between male and female recombination rates were plotted in the bottom
1101 panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found
1102 significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S32

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 15, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S33

Title: Estimated recombination rates for males and females on chromosome 16, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S34

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 17, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S35

Title: Estimated recombination rates for males and females on chromosome 18, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S36

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 19, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S37

Title: Estimated recombination rates for males and females on chromosome 20, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S38

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 21, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S39

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 22, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S40

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 23, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S41

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 24, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S42

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 25, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S43

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 26, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S44

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 27, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S45

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 28, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 5 Figure S46

File format: PDF

Title: Estimated recombination rates for males and females on chromosome 29, at 1Mb resolution.

Description: Log ratio between male and female recombination rates were plotted in the bottom panel. Confidence interval of log ratio under null hypothesis appears in grey. Intervals found significant for sex differences are marked with a red dot on the bottom panel.

Additional File 6 Table S5

File format: XLS

Title: Significant 1Mb intervals for sex differences in recombination rate.

Additional File 7 Table S6

File format: PDF

Title: Candidate genes for GWAS on EBVs for male genome-wide recombination rate in Alpine goats.

Description: All genes scoring at least 1 as candidates are showed. For each QTL of interest, genes have been sorted from higher scoring to lower scoring. Manual annotation has been added

on genes previously found in GWAS for recombination phenotypes in ruminants, referencing the corresponding papers.

Additional File 7 Table S7

File format: PDF

Title: Candidate genes for GWAS on EBVs for male genome-wide recombination rate in Saanen goats.

Description: All genes scoring at least 1 as candidates are showed. For each QTL of interest, genes have been sorted from higher scoring to lower scoring. Manual annotation has been added on genes previously found in GWAS for recombination phenotypes in ruminants, referencing the corresponding papers.

Additional File 7 Table S8

File format: PDF

Title: Candidate genes for GWAS on EBVs for male intra-chromosomal shuffling in Alpine goats.

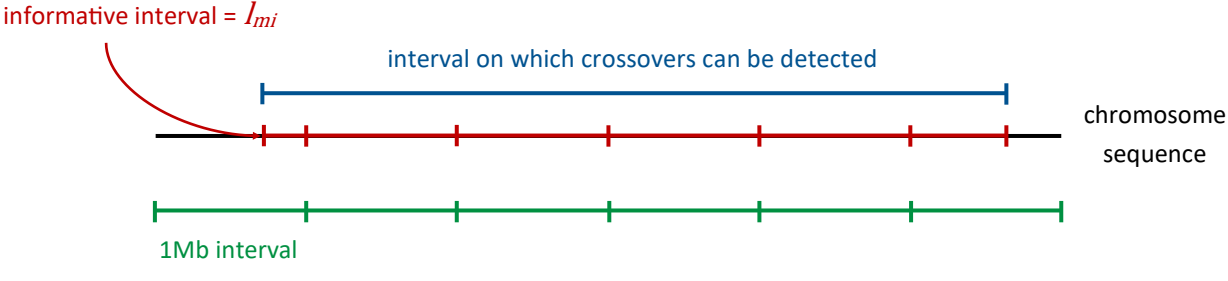
Description: All genes scoring at least 1 as candidates are showed. For each QTL of interest, genes have been sorted from higher scoring to lower scoring. Manual annotation has been added on genes previously found in GWAS for recombination phenotypes in ruminants, referencing the corresponding papers.

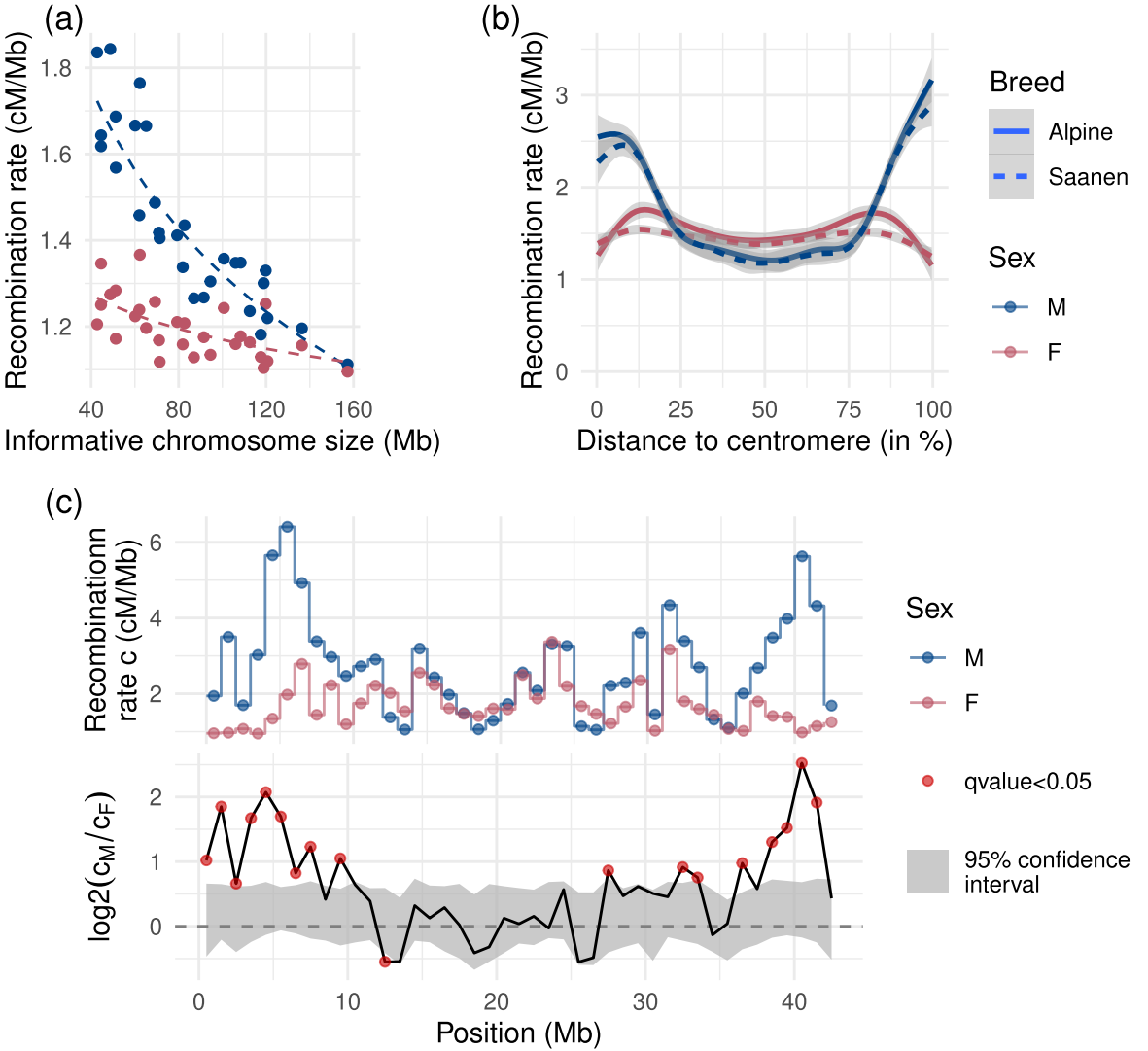
Additional File 7 Table S9

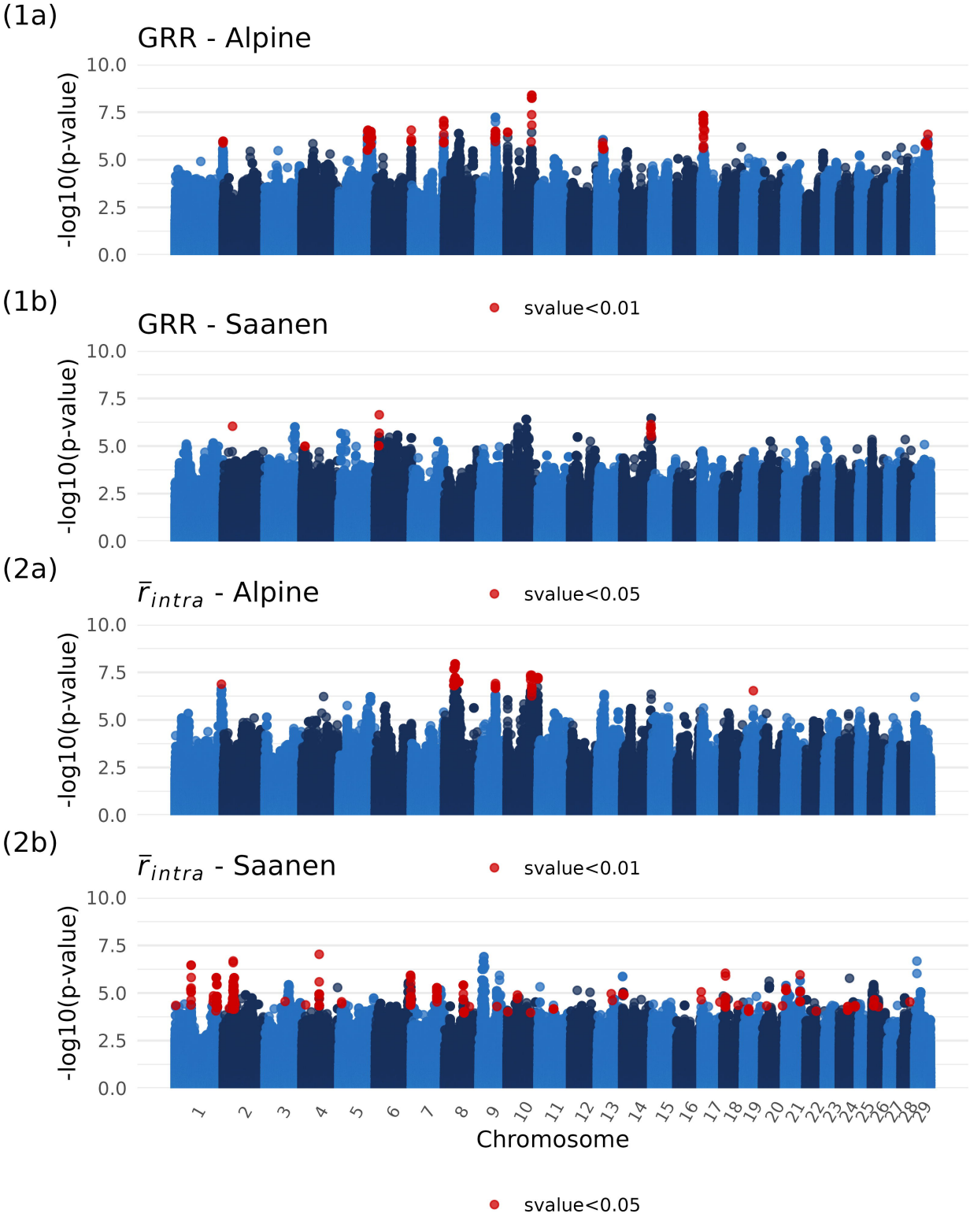
File format: PDF

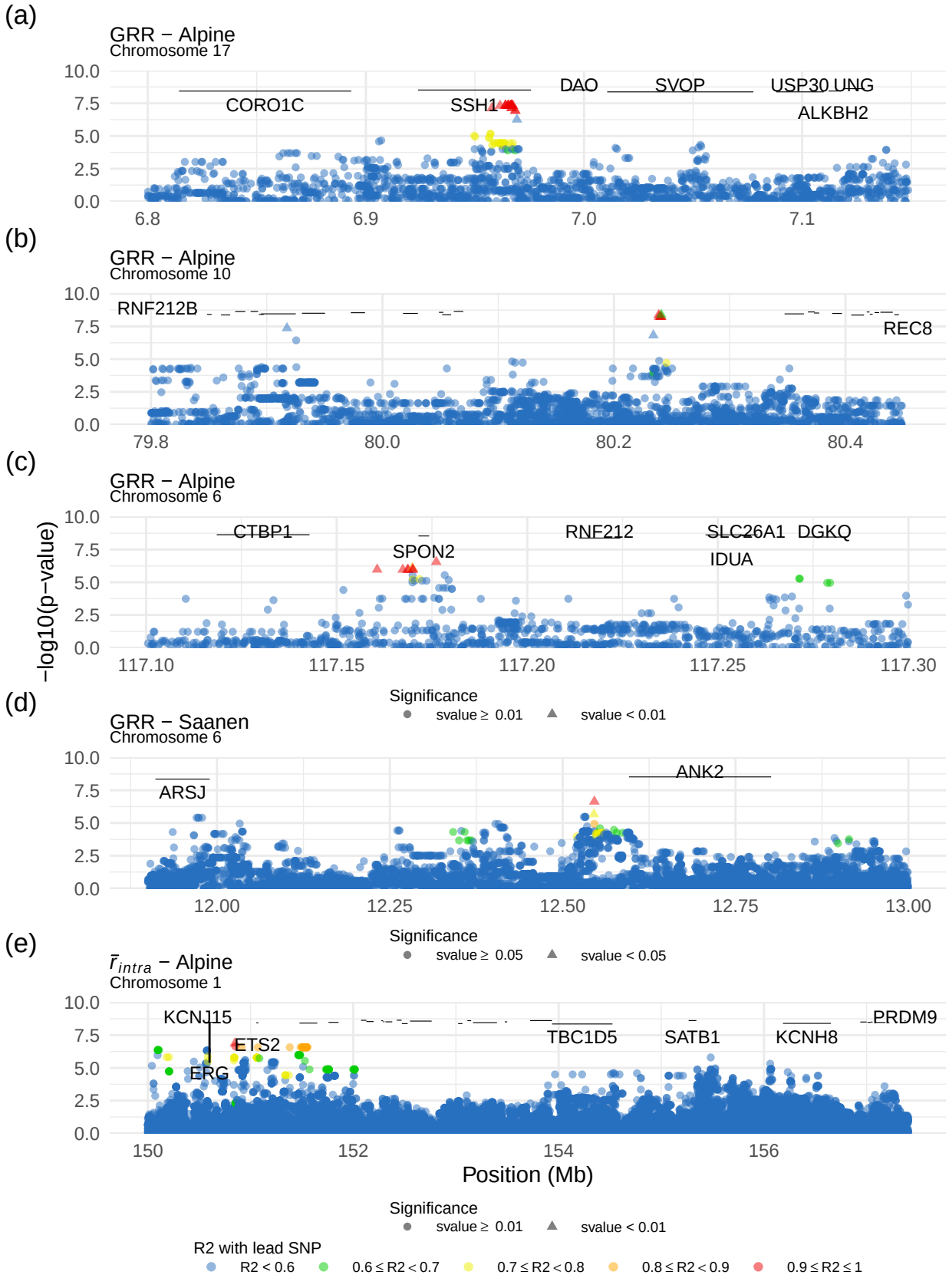
Title: Candidate genes for GWAS on EBVs for male intra-chromosomal shuffling in Saanen goats.

Description: All genes scoring at least 1 as candidates are showed. For each QTL of interest, genes have been sorted from higher scoring to lower scoring. Manual annotation has been added on genes previously found in GWAS for recombination phenotypes in ruminants, referencing the corresponding papers.


informative interval = l_{mi}
interval on which crossovers can be detected
chromosome sequence
1Mb interval


(a)
Recombination rate (cM/Mb)
1.8
1.6
1.4
1.2
40
80
120
160
Informative chromosome size (Mb)
(b)
Recombination rate (cM/Mb)
3
2
1
0
0
25
50
75
100
Distance to centromere (in %)
Breed
Alpine
Saanen
Sex
M
F
(c)
Recombinationn rate c (cM/Mb)
6
4
2
log2(cM/cF)
2
1
0
0
10
20
30
40
Position (Mb)
Sex
M
F
qvalue<0.05
95% confidence interval


(1a)
GRR - Alpine
(1b)
GRR - Saanen
svalue<0.01
(2a)
$\bar{r}_{intra}$ - Alpine
svalue<0.05
(2b)
$\bar{r}_{intra}$ - Saanen
svalue<0.01
-log10(p-value)
10.0
7.5
5.0
2.5
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
Chromosome
svalue<0.05


(a)
GRR − Alpine
Chromosome 17
CORO1C
SSH1
DAO
SVOP
USP30
UNG
ALKBH2
(b)
GRR − Alpine
Chromosome 10
RNF212B
REC8
(c)
GRR − Alpine
Chromosome 6
CTBP1
SPON2
RNF212
SLC26A1
DGKQ
IDUA
−log10(p−value)
Significance
svalue ≥ 0.01
svalue < 0.01
(d)
GRR − Saanen
Chromosome 6
ARSJ
ANK2
Significance
svalue ≥ 0.05
svalue < 0.05
(e)
$\bar{r}_{intra}$ − Alpine
Chromosome 1
KCNJ15
ETS2
ERG
TBC1D5
SATB1
KCNH8
PRDM9
Position (Mb)
Significance
svalue ≥ 0.01
svalue < 0.01
R2 with lead SNP
R2 < 0.6
0.6 ≤ R2 < 0.7
0.7 ≤ R2 < 0.8
0.8 ≤ R2 < 0.9
0.9 ≤ R2 ≤ 1