**Recombination, selection and population history effects on the distribution of runs of homozygosity (ROH) in red deer** **(*Cervus* *elaphus*)**

**Hewett, A.M. , Pemberton, J.M. , Johnston, S. E. , and Stoffel, M A..**Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, EH9 3FL, United Kingdom

**Abstract**

## **Introduction**

The inheritance of two genomic segments that are identical by descent resultant from mating of related parents can generate runs of homozygosity (ROH) in their offspring. With increased availability of high density genomic data due to SNP-chips and whole genome sequencing , in recent years the study of ROH in research has increased exponentially (Curik, Ferenčaković, and Sölkner 2014; Pryce et al. 2014). Previous studies have shown that the length, abundance and genomic location of ROHs varies considerably in a population and can be affected by a number of related processes, including selection, recombination, and population history (Kardos, Qvarnstrom, and Ellegren 2017). Firstly, an allele under strong positive selection is expected to increase in incidence in the population which may result in increased homozygosity at the site and eventually lead to area region where ROH are common in the population. This concept has been widely used to identify signatures of selection in artificially selected organisms (refs). By searching for areas in the genome where ROH are unusually common, often called ROH hotspots or ROH islands, it is possible to identify genomic regions of interest and even identify genes under selection. For example, genes associated with lactation and milk yield have been found to be present in dairy cattle, in regions where more than 50% of the population harboured a ROH (Peripolli et al. 2018). Once identified these regions can then be used to aid genetic improvement of livestock.

ROH size and position have also been shown to correlate with recombination rate. Recombination rate varies within and between chromosomes, and previous studies have shown that ROH hotspots tend to be sites of lower recombination rate (Pemberton et al. 2012; Johnston et al. 2017; Kawakami et al. 2014). In low recombination regions with high linkage disequilibrium long haplotypes are rarely broken up through meiosis, meaning they are more likely to come together in a ROH and form a ROH hotspot. Conversly, in high recombination regions, long haplotypes are broken down, resulting in fewer and shorter ROH. However, selection and recombination rate should not be thought of as mutually exclusive. Selection may interact with recombination rate to influence ROH distribution (Kardos, Qvarnstrom, and Ellegren 2017). Selection can reduce genetic variation and Ne at closely linked loci (called genetic hitchhiking or a selective sweep), with the extent dependent on the selection coefficient and local recombination rate (Charlesworth 2009). As such, in an area of low recombination, a site under positive selection is more likely to have an extended region of linked loci, which may result in a ROH. Therefore, both selection and recombination rate may be influencing the formation of a ROH hotspot.

An additional factor that can influence ROH distribution is population history. Intuitively, population history can directly influence ROH abundance, as the more inbred a population is, the more ROH will be present in the population. For example, populations of both wild and domesticated pigs having undergone a known population bottleneck have a higher incidence of ROHs than populations not having a bottleneck (Bosse et al. 2012). It was further shown by Bosse et al. (2012) that animals from the same populations tended to have similar ROH distribution patterns reflecting a shared demography. This implies that population history can have dramatic effects on ROH location, and shared population history events can result in a common ROH pattern in all individuals which experience the same history. Factors such as drift and selection will also play an important role in shaping the ROH landscape following such an event. For example, following a population bottleneck, genetic drift may result in certain alleles becoming fixed by chance, increasing the likelihood of generating ROHs at said site following generations of inbreeding. In addition, sites of past selection can result in common ROH hotspots between populations (Pemberton et al. 2012). For example Grilz-Seger et al. (2019) showed that a gene associated with coat colour suspected to be under positive artificial selection in horses is in a common ROH hotspot between breeds which shared the same founders and ancestors. On the other hand, population-specific ROH hotspots can indicate unique sites of selection (Grilz-Seger et al. 2019; Pemberton et al. 2012). It is also important to note the potential for interactions between recombination rate and population history. Species or populations which are closely related will have similar recombination rate patterns across the genome, which will again result in more similar distributions of ROHs (Nothnagel et al. 2010).

Together the factors discussed above, recombination rate, selection and population history all play interacting roles in shaping the genomic locations and incidence of ROH in a population. Here we aim to investigate the influence of each of these factors on ROH distribution in a wild population of individually-monitored red deer inhabiting the island of Rum, Scotland. We use >35,000 genome-wide autosomal SNPs to search for ROH in the population sampled over a 40-year-period. We also use a population from mainland Scotland to compare inbreeding level and ROH distribution between different populations. In addition, a linkage map for red deer enabled us to quantify the recombination rate around the gneome, allowing for the assessment of recombination rate effects on ROH. We also ran forward simulations based on known history of the population of red deer on Rum to investigate the effects of selection, recombination and population history on the distribution of ROHs.

## **Methods**

***Study populations***

The red deer population inhabiting the north block of the Isle of Rum, Scotland (57°0’N, 6°20’W) has been studied at an individual level since 1971 (Clutton-Brock, Guinness, and Albon 1982), and is the main study population for this project. DNA was extracted from ear punches from calves captured soon after birth or darted adults, post-mortem tissue or cast antlers (See Huisman et al. (2016) for full details). DNA samples were genotyped at 50,541 attempted SNP loci on the Cervine Illumina 50K BeadChip (Brauning 2015). SNP genotypes were scored using GenomeStudio (Illumina). SNP quality control was carried out in GenABEL with the following parameters: SNP minor allele frequency > 0.01, ID genotyping success > 0.99, SNP genotyping success > 0.99. Following these steps 35,132 autosomal SNPs and 3046 individuals were retained for analysis. See Johnston et al. (2017) for full details on SNP genotyping and quality control.

A mainland population of red deer from Kintyre, Scotland was compared with the Rum deer. Details of sample collection and DNA extraction can be found in McFarlane et al. (2020). Samples were genotyped and quality controlled as above, leading to 34,673 autosomal SNPs and 157 individuals being retained for analysis. These individuals were originally genotyped as part of a study of red x sika (*Cervus nippon*) hybridisation and were pure red deer according to that study (McFarlane et al. 2020).

***Calling runs of homozygosity***

We searched for runs of homozygosity in each genotyped individual using the --*homzyg* function in Plink v1.90 (Purcell et al. 2007). We used two different estimates of marker positions to call ROH: recombination distance in centimorgans (cM) and physical distancs in basepairs (bp). Recombination distance was based on direct estimates of recombination in the Rum red deer pedigree (Johnston et al. 2017). Physical markers were based on homologous SNP positions on the cattle genome BTA vUMD 3.0 and the sheep genome Oar\_v3.1 (Johnston et al. 2017). For consistency between searches we assumed 1Mb 1cM (Actual value: 1Mb = 1.04cM, (Johnston et al. 2017). The following parameters were used to identify ROH in both cases: --*autosome-num* 33, *--homozyg-snp* 40 (minimum number of SNPs in a ROH), *--homozyg-kb* 2500 (minimum length of a ROH in kb), *--homozyg-denisty* 70, *--homozyg-window-snp* 35, *--homozyg-window-missing* 4, *homozyg-het* 0, --*maf* 0.01.

***ROH Hotspots and coldspots***

To determine ROH hotspots and coldspots across the genome we estimated the proportion of individuals with a ROH at a SNP, or *PropROH*. PropROH was calculated as a SNP-by-SNP measure as follows:

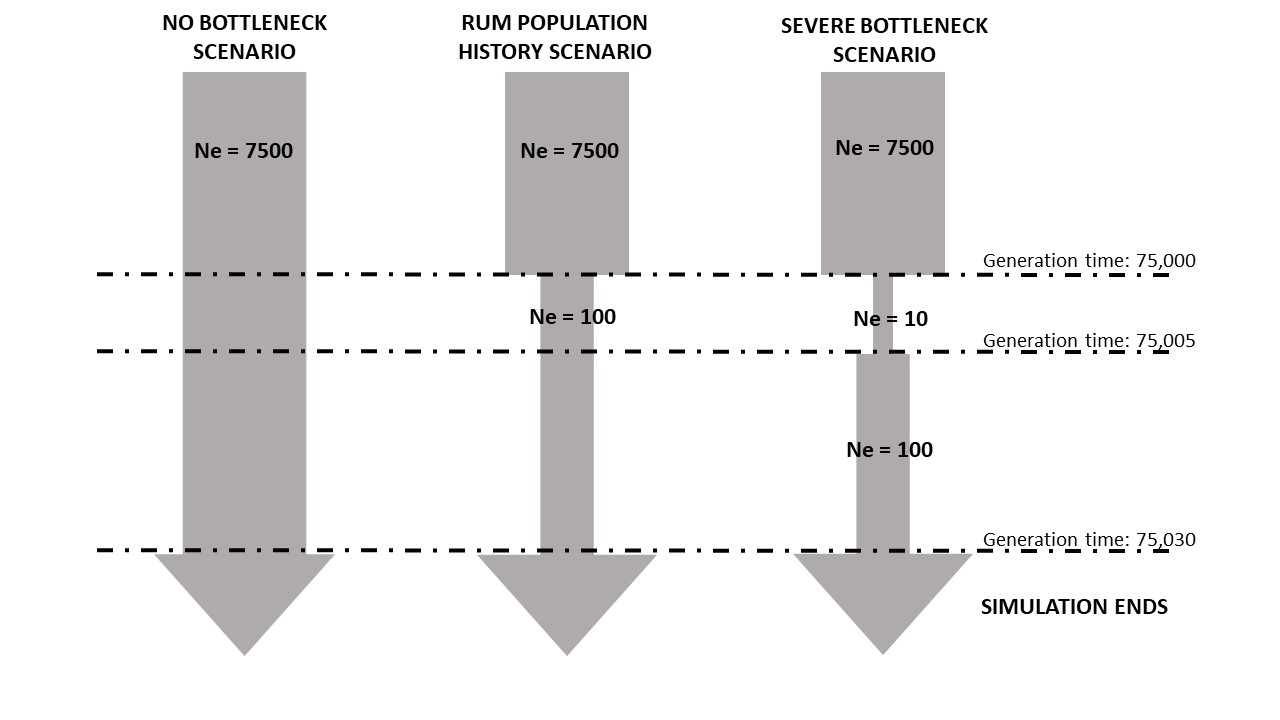
We classed the 99th percentile value for genomewide *P* as the hotspot threshold, such that any SNP with a value equal to or above this threshold was classed as a hotspot SNP. The 1st percentile value for genomewide P was classed as the coldspot threshold, and any SNP equal to or below this value was classed as a coldspot SNP.

In the course of this activity we found that *PropROH* was positively correlated with the density of SNPs such that fewer ROH were identified in regions where fewer SNPs were genotyped. For example for the Rum population, using genetic map positions and a 1500kb/cM window (which is smaller than the minimum length set for a ROH when searching for ROH, but still allows many SNPs to pass quality control) the Pearson’s correlation was 0.58). This artefact of the density of SNPs in a region and the Plink algorithm for calling ROH has previously been identified as an issue for this type of analysis (Nandolo et al. 2018). A density of 23 SNPs per 1500kb/cM window with a 100kb/cM sliding window was chosen as the minimum acceptable SNP density because this was the point at which the relationship saturated and after this threshold the correlation was greatly reduced (Pearsons correlation: 0.2, Supplementary information). All SNPs within windows that fell below this density were discarded from the analysis of hotspots and coldspots. Additional QC for this analysis removed the first and last 40 SNPs from each linkage group to account for the fact that fewer ROH will be called in these regions as a ROH cannot span past the extremities of a linkage group. These QC steps resulted in 25,798 SNPs remaining using the genetic map and 26,318 SNPs using the physical map for the Rum population. In the Kintyre population 25,194 SNPs were retained using the genetic map and 25,638 SNPs using physical map.

***Simulations***

Simulations were carried out in SLiM 3 (Haller and Messer 2019) to test the effect of selection, recombination rate and population history on the distribution of ROH across a simulated 100Mb chromosome. Each simulation was run 25 times. We used three different population history scenarios: one reflecting the Rum population, one reflecting the mainland population and one with a more severe population bottleneck than the Rum population. All scenarios started with an effective population size of 7500 set to run for 75,000 generations as a burn-in (10\*Ne) (Haller and Messer 2019). This start effective population size was based on Ne = ~1/10 population size (Frankham 2007). An estimation of the red deer population in Scotland at the beginning of the 20th century was ~150,000 individuals (The Deer Working Group 2019). The population size has since doubled to ~300,000 at the start of the 21st century (The Deer Working Group 2019). If we assume the same rate of population increase, the red deer population of Scotland (and the rest of the UK) was ~75,000 at the beginning of the 19th century.

The population history of Rum is somewhat unknown, an unrecorded number of individuals were introduced from various red deer herds in England beginning in 1845 (Marshall, 1998). These herds would have been separated from the main population of red deer prior to the introduction to Rum. Therefore, to simulate this bottleneck event in the Rum population history we dropped to Ne = 100 at generation 75,000. The simulation ended at generation 20,030 (simulated present day) and outputted the current Ne = 100 individuals.   
The more severe bottleneck scenario dropped to Ne = 10 at generation 75,000, and then increased to Ne = 100 at generation 75,005 to reflect the increase of the Rum population. As above the simulation ended at generation 75,030 and outputted Ne = 100 individuals. The mainland population scenario maintained Ne = 7500 until generation 20,030 when all Ne = 7500 individuals were output. See Fig.



Three different models were tested for each population history scenario: neutral, varied recombination rate plus selection, and varied recombination rate plus stronger selection. All models had a mutation rate of 1e-8. The neutral model was set to have a constant recombination rate across the chromosome at 1.038 cM/Mb based on estimates from the linkage map (Johnston et al. 2017) and every mutation was set as neutral i.e s=0.   
The second model incorporated neutral mutations (fixed s=0), slightly beneficial mutations (mean s=0.0001, exponential distribution), beneficial mutations (mean s=0.001, exponential distribution), slightly deleterious mutations (mean s=-0.001, gamma distribution shape parameter 0.2) and deleterious mutations (mean s=-0.01, exponential distribution) occurring in the ratio 4:1:1:1:1 respectively. This same model also had a varied recombination rate across the chromosome (rather than a constant rate as in the neutral model). Varying recombination rate was based on estimates across 32 acrocentric autosomes from Johnston et al. (2017). We split the simulated 100Mb chromosome into 10 Mb regions and assigned a recombination rate to each region based on averaged recombination rate of autosomes using their relative chromosomal position. For regions 1-10 the recombination rates were as follows: 1.75 cM/Mb, 1.23 cM/Mb, 0.89 cM/Mb, 0.81 cM/Mb, 0.74 cM/Mb, 0.80 cM/Mb, 0.87 cM/Mb, 1.05 cM/Mb, 1.20 cM/Mb, 0.67 cM/Mb.

The model with recombination and higher selection coefficients changed the selection coefficients by a factor of five, i.e. neutral mutations (fixed s=0), slightly beneficial mutations (mean s=0.0005, exponential distribution), beneficial mutations (mean s=0.005, exponential distribution), slightly deleterious mutations (mean s=-0.005, gamma distribution shape parameter 0.2) and deleterious mutations (mean s=-0.05, exponential distribution).

## **Results**

The Rum population had overall more ROH and higher inbreeding coefficients than the Kintyre population. Across all 3046 Rum deer we found 45,814 ROH using genetic map (cM positions) and 55,414 ROH using the physical map (bp positions). In 157 Kintyre deer we found 1227 ROH using the genetic map and 1448 ROH using the physical map. The mean and maximum population inbreeding coefficients were higher using the physical map positions for both populations, Table 2.

**Table 2** – Comparison of mean, minimum and maximum number of ROH per individual, mean inbreeding coefficient and maximum individual inbreeding coefficient between two populations and using two map positions to search for ROH.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean number ROH per id | Min ROH | Max ROH | Mean FROH | Max FROH |
| Rum population  (Genetic map) | 15.04 | 0 | 34 | 0.037 | 0.193 |
| Rum population  (Physical map) | 18.2 | 0 | 36 | 0.0488 | 0.261 |
| Kintyre pop  (Genetic map) | 7.82 | 0 | 22 | 0.019 | 0.107 |
| Kintyre pop  (Physical map) | 9.22 | 0 | 22 | 0.024 | 0.118 |

***ROH hotspots***

As demonstrated above we found that the marker positions used (cM or bp positions) affected the number of ROH found in the population. We next explored this effect on the genomic distribution of ROH. In the Rum population using the genetic map positions there were five regions classed as ROH hotspots across five linkage groups (Fig. 2, Table X). In contrast, using the physical map positions seven regions were shown to be ROH hotspot on five linkage groups. 142 SNPs in four regions were classed as hotspots in both methods. Of these shared hotspots those on linkage group 28 and 15 had a higher propROH  using the genetic positions than the physical positions, Fig. X.

Using the genetic map the Kintyre population had four regions classed as a ROH hotspot across four linkage groups. Using the physical map for the Kintyre population 5 regions were shown as ROH hotspots on 5 linkage groups. Only one hotspot was consistent between the two methods, on linkage group 20, with 38 SNPs shared.

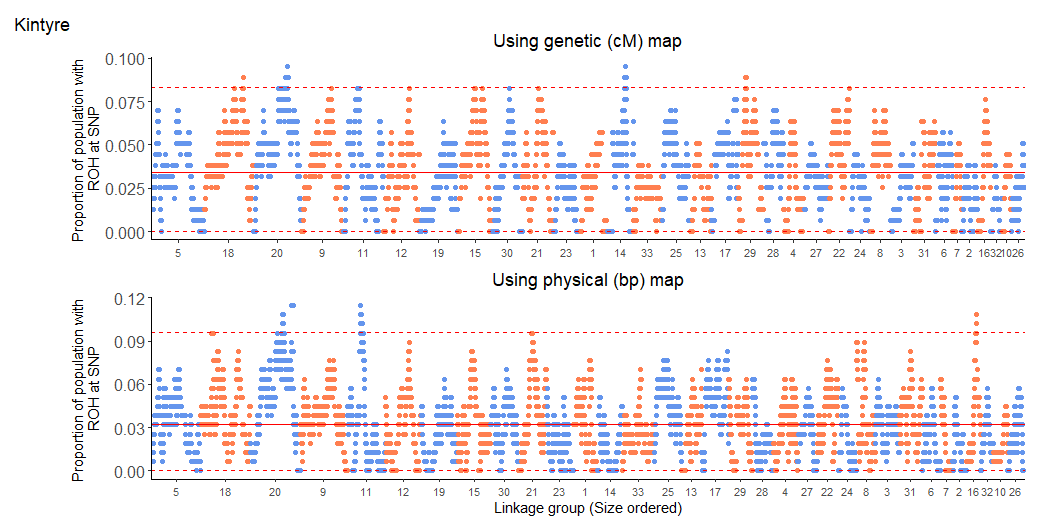
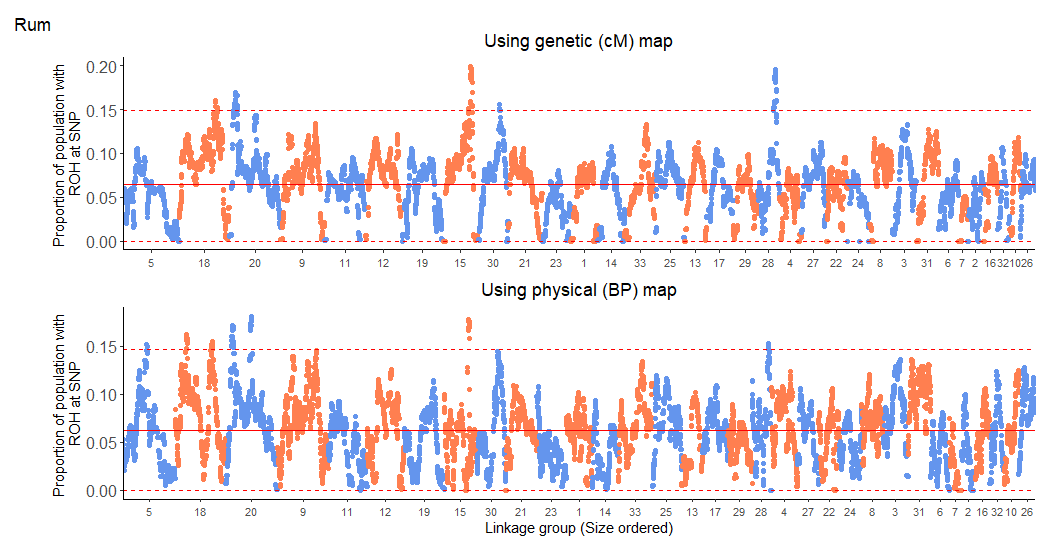
When using the genetic map for both populations no hotspot SNPs are shared between the populations. However, when using the physical map for both populations 45 hotspot SNPs are shared, in one region on linkage group 18, Fig. This shared hotspot was not classed as a hotspot in either population using the genetic map, Fig.

**Table 3** – Comparison of ROH hotspots across two populations and using two map positions to search for ROH. Shown is the mean proportion of individuals with a ROH at a SNP (PropROH),with standard error. The ROH hotspot threshold (99th percentile of PropROH), any SNP with a value of PropROH above this threshold value is classed as a ROH hotspot. The number of SNPs classed as ROH hotspots and the linkage groups containing ROH hotspots. Linkage groups containing multiple hotspots are labelled as hotspot a) or b).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Population | Map Position used | Mean *P*ropROH | ROH Hotspot threshold  (99th percentile *PropROH*) | # of SNPs over ROH hotspot threshold | Linkage groups containing ROH hotspots |
| Rum dataset | Genetic | 0.064 ± 0.0002 (SE) | 14.9% | 257 | 15,18,20,28,30 |
| Physical | 0.062 ± 0.0002 (SE) | 14.7% | 142 | 5,15,18(a),18(b),20(a),20(b),28 |
| Kintyre dataset | Genetic | 0.034 ±0.0002 (SE) | 8.3% | 92 | 14,18,20,29 |
| Physical | 0.032 ± 0.0001 (SE) | 9.6% | 324 | 11,16,18,20,21 |

We found that the pattern of ROH distribution were moderately positively correlated between the two populations. Using the genetic map the correlation of *PropROH*  between the Rum and Kintyre populations was r=0.567 (Pearson’s correlation, p-value: 2.2e-16, df=24200, t=107.01). The physical map showed a correlation of *P* of r=0.529 (Pearson’s correlation, p-value:2.2e-16, df=24046, t=96.841).

**Figure 2** – Proportion of **Rum** (top) and **Kintyre** (Bottom) individuals with a ROH at each SNP (P) across all linkage groups. Linkage groups are ordered by size from largest on the left to smallest on the right. The top panel shows the distribution of P using the genetic map positions (cM) and the bottom panel shows the same when using physical map positions (bp). Upper and lower dashed lines indicate the 99th and 1st percentile values representing hotspot and coldspot thresholds respectively. The solid red line indicates the mean P.



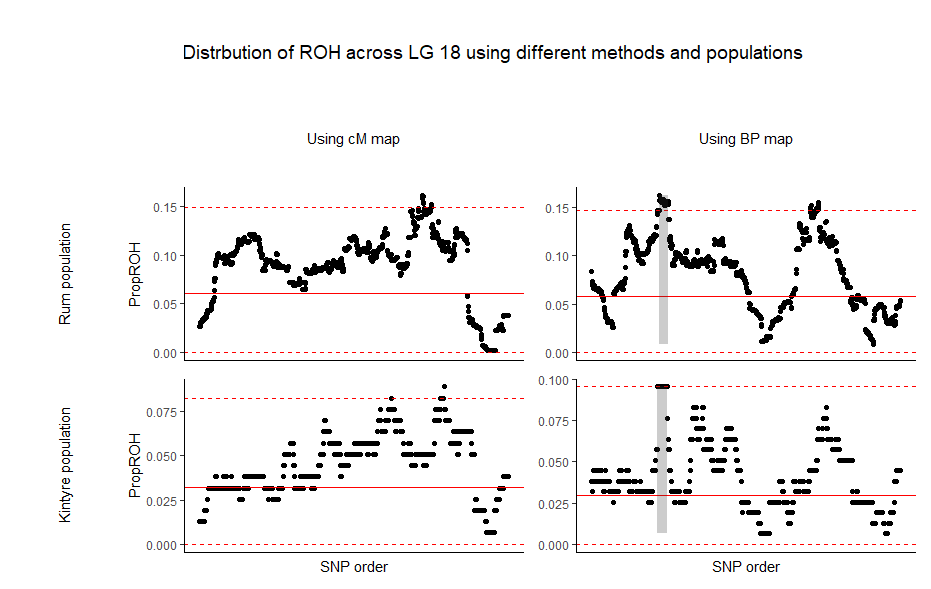


Fig – Comparison of the proportion of individuals with a ROH at a SNP (PropROH) in linkage group 18 for two populations of red deer and using two map positions to search for ROH – cM and bp positions. The region shaded grey shows the hotspot regions shared between two populations when using the physical map positions. Upper and lower red dashed line show the 99th and 1st percentile PropROH respectively. The solid red line shows the mean PropROH. NB: the hotspots using the cM map is not shared between the populations, as no SNPs are common between both.

***Simulations***

In our simulations we ran 25 iterations of three model designs within three different population histories. In Fig. 6 we show the mean (Fig. 6A) and 99th percentile (Fig 6B) PropROH for each of the 25 iterations, grouped by model design. For each population history scenario we compared various models to a neutral model with a constant recombination rate throughout the chromosome and no selection. Comparing to this neutral model we investigated the effects of varying recombination rates over the chromosome and selection, and varied recombination rates with stronger selection.

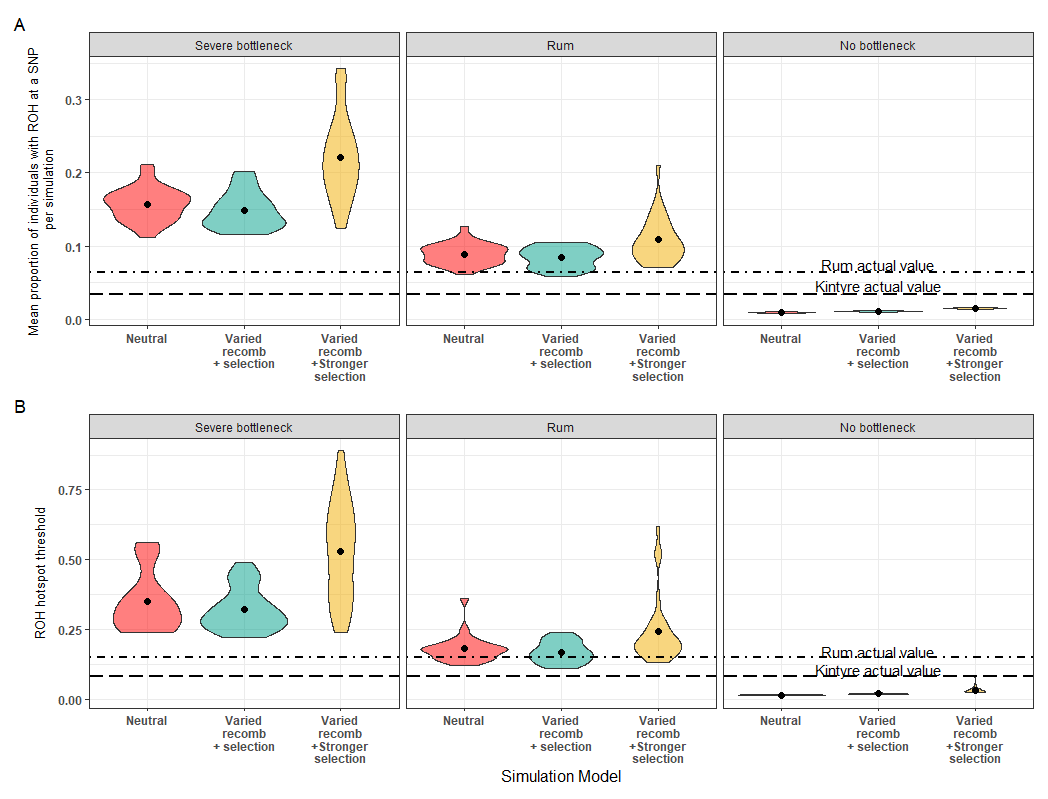
We found that of all parameter tested, population history had the greatest effect on ROH number and localisation. When we compare between population histories (compare across panels in Fig.6A) there is a noticeable variation in mean propROH. A severe bottleneck gives a higher mean PropROH, and more variation between iterations of the same model (Fig. 6A left). These higher mean PropROH values reflect a higher number of ROH in this scenario as a result of more inbreeding events. Whereas, the simulated Rum population history model has an intermediate mean *PropROH* between the other two scenarios (Fig. 6A middle). Finally, the population model with no bottleneck has a much lower mean *PropROH* reflecting a lower incidence of ROH resulting from less inbreeding (Fig 5A right).

Comparing within each population history, no model differs greatly from the neutral model for the mean *PropROH*. As stated above there is the most variation between iterations of the same model in the severe bottleneck, and the stronger selection model shows the greatest variation between iterations. This variation between iterations is likely due to the effects that genetic drift can have on populations following a severe bottleneck. The no bottleneck population shows the biggest difference in mean propROH between models, with stronger selection showing higher propROH that other models. (selection is more efficient in larger pops).

While mean propROH can give a good indication of the ROH number in a population, the 99th percentile propROH (i.e. the ROH hotspot threshold) can show the localisation of ROH, and whether there are any regions which have excessive numbers of ROH. As above, the most variation in 99th percentile propROH is between population history, reflecting the level of inbreeding occurring in the population. In all population histories the model including stronger selection and varied recombination shows outliers with higher values indicating that in some iterations this hotspot threshold is driven up,. As this is only present in the stronger selection model it implies that selection is creating a ROH hotspot. However, as can be seen in the width of the violin plot this does not occur in all iterations.

Comparing the simulation results to the empirical data from Rum and Kintyre – where the no bottleneck simulations are representing the Kintyre population and the Rum simulations represent the population on Rum – the mean and 99th percentile values from simulations reflect well those found in the empirical data, Fig 6. This similarity between results allows for a direct comparison between simulations and empirical data. From the simulation results using the Rum population history we cannot conclude that the ROH hotspots found in the Rum dataset are sites of positive selection. We show that no model differs greatly from the neutral model. Indicating that the results from the empirical data were equally likely to occur from a neutral model than that of a model which included selection. In addition, the Kintyre values for PropROH are higher in the empirical dataset than the simulated results, suggesting that the Kintyre population history is more complex than this simulation.

**Figure 5** – Violin plots showing A) mean B) 99th percentile proportion of individuals with a ROH at a SNP values for every iteration. Black dots indicate the mean value for all 25 iterations of a simulation, the width of the violin indicates the number of simulations within said area. Each box contains a different simulated population history from left to right: severe historical population bottleneck, Rum population history, no historical population bottleneck. Each population history had 5 models tested: neutral model with a constant recombination rate and no selection, model including selection, model including varied recombination rate across the simulated chromosome, model including both varied recombination rate and selection and finally a model with selection coefficients 5-fold higher than the previous model.



## Discussion

**Population history and inbreeding level**

We found that the Rum population was more inbred that the Kintyre population (Rum mean FROH from genetic map: 0.037, Kintyre FROH from genetic map: 0.019), and this difference consistent when using the physical map. It reflects different population histories. The Rum population became isolated from the mainland ~150 years ago, has had few introductions since and currently numbers around 1000 individuals, while the Kintyre population is continuous with the rest of the Scottish mainland population which numbers hundreds of thousands.

ROH distribution can be used to compare populations of the same species and/ or different related species, inferring ROH hotspots common between populations or species are shared sites of selection (Grilz-Seger et al. 2019; Pemberton et al. 2012). Here we find ROH location is positively correlated between two red deer populations using both maps. This result is akin to previous studies showing ROH distribution patterns were similar between individuals within the same population due to a shared population history (Bosse et al. 2012). The Rum and Kintyre population will have shared a large part of their population history – only diverging ~150 years ago. Therefore, the similar patterns in ROH distribution may be due to this shared ancestry. However, a the possibility of a technical artefact in these results should not be ignored. The correlation in the ROH distributions between two populations became stronger when using the unfiltered SNPs. As the SNPs removed in the filtered dataset were in low density regions this may mean that the number of ROH found and in turn the ROH distribution was dependent on SNP density. We have already shown this artefact to be somewhat true, hence the filtering (see methods). Others have also highlighted this issue of misidentifying ROH islands due to SNP coverage in the bovine genome (Nandolo et al. 2018). Therefore, it is possible that the correlation between ROH distribution in the two populations may also be driven by the SNP density, as this will be similar between the populations. On the other hand, the filtering steps performed prior to this analyses to account for this artefact should allow some assurance across our results, nonetheless it is an important issue to highlight.

**Recombination rate**

We used two methods when searching for ROH in *plink,* the genetic map reflecting recombination, and the physical map that does not, centimorgan (cM) position, and base pair (bp) position respectively. Kardos, Qvarnstrom, and Ellegren (2017) first suggested using the genetic map in this circumstance under the reasoning that if one accounts for recombination, remaining hotspots are more likely to be sites of interest (e.g. areas of positive selection). Kardos, Qvarnstrom, and Ellegren (2017) further noted that, when compared to the physical map, the genetic map showed ROH hotspots in the same positions but to a ‘lesser extent’. As in Kardos, Qvarnstrom, and Ellegren (2017) we find similar hotspots that have lower PropROH when using the genetic map rather than the physical map in both populations of red deer. Four ROH hotspots in the Rum population using the physical map are no longer classed as hotspots when using the genetic map. In addition, we found two hotspots reach further over the hotspot threshold in the genetic map than in the physical map and one new hotspot that reaches over the threshold. Our results support those from Kardos, Qvarnstrom, and Ellegren (2017) showing the effect that using the genetic map can have on results analysing ROH hotspots. The differing results are due to the effect that recombination rate has on the distribution of ROH across the genome, which is known from previous studies. Pemberton et al. (2012) clearly show a correlation between ROH distribution and recombination rate, with a number of ROH hotspots occurring in regions of low recombination rate. In using the genetic map positions we can hence account for recombination rate variability across the genome allowing for a better determination of ROH hotspots of interest. Therefore, in agreement with Kardos, Qvarnstrom, and Ellegren (2017) we suggest that, where possible, studies should preferentially use genetic map positions to search for ROH hotspots of interest, particularly to infer selection.

When comparing the ROH distributions between populations a noteworthy result is that a shared hotspot between the two populations is only present when using the physical map. As the physical map does not take into account recombination, this suggests that this hotspot is a consequence of the recombination rate landscape shared between populations. Nothnagel et al. (2010) also concluded that in humans, locations of ROH hotspots were alike in different subpopulations due to similar genome-wide patterns of recombination. Our results here support these conclusions, also suggesting recombination rate is driving similarities in ROH patterns between the two populations. We suggest caution over concluding that ROH hotspots in common between populations are sites of positive selection without taking into account recombination. As demonstrated here, common ROH hotspots may be regions of lower recombination rate, which are common between populations as a result of a shared recombination landscape. Hence, a further benefit of using a genetic map, is in comparing populations will hold fewer biases, giving more accurate representations of ROH distributions occurring from unique population demography or selection.

**Selection**

Once we have accounted for recombination using the genetic map, the remaining hotspots may be areas of interest - a number of studies suggest ROH hotspots are sites of positive selection (Peripolli et al. 2018). When using the genetic map positions the red deer population on Rum showed five genome regions with a ROH in more than 15% of the population i.e. a ROH hotspot. In the mainland population from Kintyre four regions were present in more than 8% of the population. One ROH hotspot was in common between the two populations. It is possible these regions may be sites of positive selection, but more analysis is needed to validate this. In ROH hotspots in the Rum population, haplotype diversity was lower in comparison to ROH coldspots and the remainder of the genome. However, the number of different haplotypes did not differ between these groups. Additionally, within specific linkage groups, haplotype diversity was reduced at the site of a hotspot. Together these results suggest certain haplotypes are at a higher frequency than others. These haplotypes may then be beneficial, representing sites of positive selection. However, inferring sites of selection should be taken in caution. The percentage of individuals with a ROH at a hotspot was much lower here than in other studies suggesting ROH as sites of selection, as high as >80% in livestock and ~20% in humans (Pemberton et al. 2012; Purfield et al. 2012). Therefore, as there is a fairly low number of individuals in the whole population, these may be regions of genetic drift occurring by chance.

**Simulations**

We used simulations as a tool to establish whether the ROH hotspots found in the Rum population may be a result of selection or other mechanisms. In the Rum population history simulations, the mean P and top 1% P are similar to the empirical results from Rum, giving confidence that the simulation parameters chosen can reflect the Rum population with some accuracy. From the simulations we show no model in the Rum population history scenario differs greatly from the neutral model, indicating that the results from the empirical data were equally likely to occur as a result of neutral processes as under selection. However, we cannot conclude these regions are not sites of selection. One noteworthy result common across all population history scenarios is that the model including higher selection coefficients (5x higher than selection coefficients used in other models) resulted in some outliers with a higher top 1% threshold. These simulations with a higher threshold indicate regions of stronger selection driving up this threshold. These outliers were more similar to results from artificially selected organisms, which generally have higher selection coefficients than naturally selected organisms (ref). Therefore, ROH hotspots may only contain sites of selection if the selection coefficients for the population are high enough and only in a certain number of outlier cases. In the case of the Rum population, the simulations suggest that selection coefficients may not be high enough to influence ROH distribution. On the other hand, in populations with higher selection coefficients (e.g. artificially selected organisms) the selection can drive ROH distribution.

From our simulations, we show that of all the variables tested, population history actually had the greatest influence on ROH distribution. Between population history scenarios there is a noticeable difference in both mean P and top 1% P. The simulations with a severe bottleneck had the highest number of ROH and each model had considerable variability per simulation shown in the spread of the violin plot. These results reflect those observed in real bottlenecked populations (REF). A smaller effective population size means inbreeding will be more common than in populations with no bottleneck, resulting in more ROH in the population. On the other hand, simulations with no bottleneck (i.e. reflecting the mainland population) show a very low incidence of ROH when on the same scale as the other population scenarios. Reflecting the differences found in the empirical data between the mainland and Rum population. The mainland population, with a higher effective population size, will have a lower level of inbreeding resulting in less ROH. We also see a significantly higher mean P and top 1% P in the stronger selection coefficient model relative to the other models in this population history scenario, a pattern which is not present in the other two population history scenarios. This again reflects what is found in literature, that in larger populations selection is more efficient. As such, mutations with higher selection coefficients are likely occurring at a higher frequency and therefore are able to be selected for or against, resulting in a higher incidence of ROH compared to the other models. In a smaller population these mutations may not reach fixation as a result of genetic drift (REF!!).

Our simulations have offered a valuable way to assess how different factors may be affecting our main study population on Rum, overall showing that population history has the greatest effect on ROH distribution and density in a simulated wild population. Interestingly selection and recombination do not appear to influence the distribution of ROH in our simulations. Selection only appears to have an effect when the selection coefficients are increased to possibly unrealistic levels for a wild population. In a slightly contrary outcome to our results above and past literature showing the effects of recombination rate on ROH distribution, a model including recombination rate does not differ greatly to a null model. It may be that recombination rate does effect ROH distribution but not to a such dramatic state that would result in a ROH hotspot.

## **References**

Bosse, M., H. J. Megens, O. Madsen, Y. Paudel, L. A. Frantz, L. B. Schook, R. P. Crooijmans, and M. A. Groenen. 2012. 'Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape', *PLoS Genet*, 8: e1003100.

Brauning, R. ;Fisher, P.J,; A.F. McCulloch; R. J. Smithies; J. F. Ward; M. J. Bixley; C.T. Lawley; S.J. Rowe; J.C. McEwan. 2015. 'Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk', *bioRxiv*.

Charlesworth, B. 2009. 'Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation', *Nat Rev Genet*, 10: 195-205.

Clutton-Brock, Tim H, Fiona E Guinness, and Steve D Albon. 1982. *Red deer: behavior and ecology of two sexes* (University of Chicago press).

Curik, Ino, Maja Ferenčaković, and Johann Sölkner. 2014. 'Inbreeding and runs of homozygosity: a possible solution to an old problem', *Livestock Science*, 166: 26-34.

Frankham, R. 2007. 'Effective population size/adult population size ratios in wildlife: a review', *Genet Res*, 89: 491-503.

Grilz-Seger, G., T. Druml, M. Neuditschko, M. Mesaric, M. Cotman, and G. Brem. 2019. 'Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size', *Anim Genet*, 50: 334-46.

Haller, B. C., and P. W. Messer. 2019. 'SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model', *Mol Biol Evol*, 36: 632-37.

Huisman, J., L. E. Kruuk, P. A. Ellis, T. Clutton-Brock, and J. M. Pemberton. 2016. 'Inbreeding depression across the lifespan in a wild mammal population', *Proc Natl Acad Sci U S A*, 113: 3585-90.

Johnston, S. E., J. Huisman, P. A. Ellis, and J. M. Pemberton. 2017. 'A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus elaphus)', *G3 (Bethesda)*, 7: 2859-70.

Kardos, M., A. Qvarnstrom, and H. Ellegren. 2017. 'Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences', *Genetics*, 205: 1319-34.

Kawakami, T., L. Smeds, N. Backstrom, A. Husby, A. Qvarnstrom, C. F. Mugal, P. Olason, and H. Ellegren. 2014. 'A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution', *Mol Ecol*, 23: 4035-58.

McFarlane, S. E., D. C. Hunter, H. V. Senn, S. L. Smith, R. Holland, J. Huisman, and J. M. Pemberton. 2020. 'Increased genetic marker density reveals high levels of admixture between red deer and introduced Japanese sika in Kintyre, Scotland', *Evol Appl*, 13: 432-41.

Nandolo, W., Y. T. Utsunomiya, G. Meszaros, M. Wurzinger, N. Khayadzadeh, R. B. P. Torrecilha, H. A. Mulindwa, T. N. Gondwe, P. Waldmann, M. Ferencakovic, J. F. Garcia, B. D. Rosen, D. Bickhart, C. P. van Tassell, I. Curik, and J. Solkner. 2018. 'Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances', *Genet Sel Evol*, 50: 43.

Nothnagel, M., T. T. Lu, M. Kayser, and M. Krawczak. 2010. 'Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans', *Hum Mol Genet*, 19: 2927-35.

Pemberton, T. J., D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg, and J. Z. Li. 2012. 'Genomic patterns of homozygosity in worldwide human populations', *Am J Hum Genet*, 91: 275-92.

Peripolli, E., N. B. Stafuzza, D. P. Munari, A. L. F. Lima, R. Irgang, M. A. Machado, Jcdc Panetto, R. V. Ventura, F. Baldi, and Mvgb da Silva. 2018. 'Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle', *BMC Genomics*, 19: 34.

Pryce, J. E., M. Haile-Mariam, M. E. Goddard, and B. J. Hayes. 2014. 'Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle', *Genet Sel Evol*, 46: 71.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am J Hum Genet*, 81: 559-75.

Purfield, D. C., D. P. Berry, S. McParland, and D. G. Bradley. 2012. 'Runs of homozygosity and population history in cattle', *BMC Genet*, 13: 70.