bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A high-density linkage map reveals sexually-dimorphic recombination landscapes in red deer (*Cervus elaphus*).

Susan E. Johnston, Jisca Huisman, Philip A. Ellis and Josephine M. Pemberton

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3FL, United Kingdom.

Corresponding author: Susan.Johnston@ed.ac.uk

Abstract

High density linkage maps are an important tool to gain insight into the genetic architecture of traits of evolutionary and economic interest. Here, we used information from the cattle genome to inform and refine a high density linkage map in a wild population of red deer (*Cervus elaphus*). We constructed a predicted linkage map of 38,038 SNPs and a skeleton map of 10,835 SNPs across 34 linkage groups. We identified several chromosomal rearrangements in the deer lineage, including six fissions, one fusion and two large inversion events. Our findings also showed strong concordance with map orders in the cattle genome. The sex-average linkage map length was 2739.7cM; the female autosomal map length was 1.21 longer than that of males (2767.4cM vs 2280.8cM, respectively). Differences in map lengths between the sexes was driven by markedly increased female recombination in centromeric regions and reduced male and female recombination at sub-telomeric regions. This observed pattern is unusual relative to other mammal species, where centromeric recombination is suppressed and telomeric recombination rates are higher in males. Overall, these maps provide an insight into recombination landscapes in mammals, and will provide a valuable resource for studies of evolution, genetic improvement and population management in red deer and related species.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

The advent of affordable next-generation sequencing and SNP-typing assays allows large numbers of polymorphic genetic markers to be characterised in almost any system. A common challenge is how to organise these genetic variants into a coherent order for downstream analyses, as many approaches rely on marker order information to gain insight into genetic architectures and evolutionary processes (Ellegren, 2014). Linkage maps are often an early step in this process, using information on recombination fractions between markers to group and order them on their respective chromosomes (Sturtevant, 1913; Lander & Schork, 1994). Ordered markers have numerous applications, including: trait mapping through quantitative trait locus (QTL) mapping, genome-wide association studies (GWAS) and regional heritability analysis (Bérénos *et al.*, 2015; Fountain *et al.*, 2016); genome-scans for signatures of selection and population divergence (Bradbury *et al.*, 2013; McKinney *et al.*, 2016); quantification of genomic inbreeding through runs of homozygosity (Kardos *et al.*, 2016); investigation of contemporary recombination landscapes (Kawakami *et al.*, 2014; Johnston *et al.*, 2016); and comparative genomics and genome evolution (Brieuc *et al.*, 2014; Leitwein *et al.*, 2016). Linkage maps also provide an important resource in *de novo* genome assembly, as they provide information for anchoring sequence scaffolds and allow prediction of gene locations relative to better annotated species (Fierst, 2015).

Nevertheless, creating linkage maps of many thousands of genome-wide markers *de novo* is a computationally intensive process requiring pedigree information, sufficient marker densities over all chromosomes and billions of locus comparisons. Furthermore, the ability to create a high resolution map is limited by the number of meioses in the dataset; as marker densities increase, more individuals are required to resolve genetic distances between closely linked loci (Kawakami *et al.*, 2014). Whilst *de novo* linkage map assembly with large numbers of SNPs is possible (Rastas *et al.*, 2016), one approach to ameliorate the computational cost and map resolution is to use genome sequence data from related species to inform initial marker orders. Larger and finer scale rearrangements can then be refined through further investigation of recombination fractions between markers.

In this study, we use this approach to construct a high density linkage map in a wild population of red deer (*Cervus elaphus*). The red deer is a large deer species widely distributed across

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the northern hemisphere, and is a model system for sexual selection (Kruuk *et al.*, 2002), behaviour (Clutton-Brock *et al.*, 1982), hybridisation (Senn & Pemberton, 2009), inbreeding (Huisman *et al.*, 2016) and population management (Frantz *et al.*, 2006). They are also an increasingly important economic species farmed for venison, antler velvet products and trophy hunting (Brauning *et al.*, 2015). A medium density map ($\sim$ 600 markers) is available for this species, constructed using microsatellite, RFLP and allozyme markers in a red deer $\times$ Père David's deer (*Elaphurus davidianus*) F_2 cross (Slate *et al.*, 2002). However, these markers have been largely superseded by the development of the Cervine Illumina BeadChip which characterises 50K SNPs throughout the genome (Brauning *et al.*, 2015). Estimated SNP positions are known relative to the cattle genome, but the precise order of SNPs in red deer remains unknown. Here, we integrate pedigree and SNP data from a long-term study of wild red deer on the island of Rum, Scotland to construct a predicted linkage map of $\sim$ 38,000 SNP markers and a skeleton linkage map of $\sim$11,000 SNP markers. As well as identifying strong concordance with the cattle genome and several chromosomal rearrangements, we also present evidence of strong sexual dimorphism in recombination rates (i.e. heterochiasmy) at centromeric regions of the genome. We discuss the implications of our findings for other linkage mapping studies and potential drivers of recombination rate variation and heterochiasmy within this system.

Materials and Methods

Study Population and SNP dataset.

The red deer population is located in the North Block of the Isle of Rum, Scotland (57°02'N, 6°20'W) and has been subject to an on-going individual-based study since 1971 (Clutton-Brock *et al.*, 1982). Research was conducted following approval of the University of Edinburgh's Animal Welfare and Ethical Review Body and under appropriate UK Home Office licenses. DNA was extracted from neonatal ear punches, post-mortem tissue, and cast antlers (see Huisman *et al.*, 2016 for full details). DNA samples from 2880 individuals were genotyped at 50,541 SNP loci on the Cervine Illumina BeadChip (Brauning *et al.*, 2015) using an Illumina genotyping platform (Illumina Inc., San Diego, CA, USA). SNP genotypes were scored using Illumina GenomeStudio software, and quality control was carried out using the *check.marker* function in GenABEL v1.8-0 (Aulchenko *et al.*, 2007) in R v3.3.2, with the following thresholds: SNP geno-

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

typing success >0.99, SNP minor allele frequency >0.01, and ID genotyping success >0.99. A total of 38,541 SNPs and 2,631 IDs were retained. The function identified 126 pseudoautosomal SNPs; heterozygous genotypes at non-pseudoautosomal X-linked SNPs within males were scored as missing. A pedigree was constructed in the software Sequoia (Huisman, 2017; see Huisman *et al.*, 2016 for information on deer pedigree construction).

Linkage map construction.

A standardised sub-pedigree approach was used for linkage map construction (Johnston *et al.*, 2016). The pedigree was split as follows: for each focal individual (FID) and its offspring, a sub-pedigree was constructed that included the FID, its parents, the offspring, and the other parent of the offspring (Figure 1), and were retained where all five individuals were SNP genotyped. This pedigree structure characterises crossovers occurring in the gamete transferred from the FID to the offspring. A total of 1355 sub-pedigrees were constructed, allowing characterisation of crossovers in gametes transmitted to 488 offspring from 83 unique males and 867 offspring from 259 unique females. Linkage mapping was conducted using an iterative approach using the software CRI-MAP v2.504a (Green *et al.*, 1990), with input and output processing carried out using the R package crimaptools v0.1 (S.E.J., available https://github.com/susjoh/crimaptools) implemented in R v3.3.2. In all cases, marker order was specified in CRI-MAP based on the criteria outlined in each section below. Code for the deer map construction is archived at https://github.com/susjoh/DeerMapv4.

Build 1: Order deer SNPs based on synteny with cattle genome. Mendelian incompatibilities were identified using the CRI-MAP *prepare* function, and incompatible genotypes were removed from both parents and offspring. SNPs with Mendelian error rates of >0.01 were discarded (N = 0 SNPs). Sub-pedigrees with more than 50 Mendelian errors between an FID and its offspring were also discarded (N = 4). All SNPs were named based on direct synteny with the cattle genome (BTA vUMD 3.0; N = 30), and so loci were ordered and assigned to linkage groups assuming the cattle order and a sex-averaged map of each chromosome was constructed using the CRI-MAP *chrompic* function (N = 38,261 SNPs, Figure S1).

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

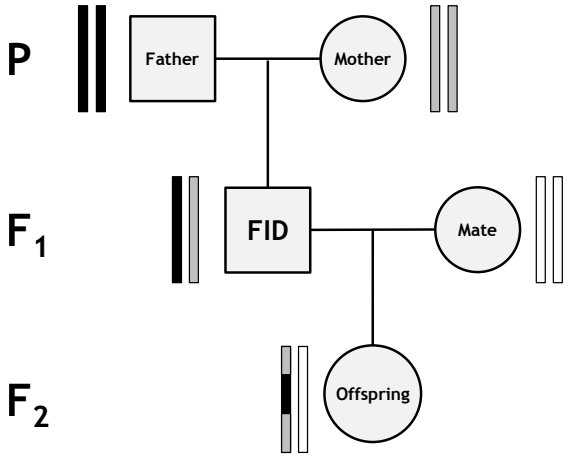

Figure 1: Sub-pedigree structure used to construct linkage maps. Rectangle pairs next to each individual represent chromatids, with black and grey shading indicating chromosome or chromosome sections of FID paternal and FID maternal origin, respectively. White shading indicates chromatids for which the origin of SNPs cannot be determined. Crossovers in the gamete transferred from the focal individual (FID) to its offspring (indicated by the grey arrow) can be distinguished at the points where origin of alleles origin flips from FID paternal to FID maternal and vice versa. From (Johnston *et al.*, 2016).

Build 2: Rerun cattle order with wrongly positioned chunks removed. All SNP loci from Build 1 were assigned to "chunks", defined as a run of SNPs flanked by map distances of ≥ 3 centiMorgans (cM). Several short chunks were flanked by large map distances, indicating that they were wrongly positioned in Build 1 (Figure S1); chunks containing <20 SNPs were removed from the analysis for Build 2 (N = 327 SNPs). A sex-averaged map of each chromosome was reconstructed using the *chrompic* function (N = 37,934 SNPs, Figure S2).

Build 3: Arrange chunks into deer linkage groups. SNPs from Build 2 were arranged into 34 deer linkage groups (hereafter prefixed with CEL) based on a previous characterisation of fissions and fusions from the red deer $\times$ Père David's deer linkage map (Slate *et al.*, 2002) and visual inspection of linkage disequilibrium (LD, R^2, calculated using the *r2fast* function in GenABEL; Figure S3). There was strong conformity with fissions and fusions identified in the previous deer map (Table 1); intra-marker distances of $\sim$100 cM between long chunks indicated that they segregated as independent chromosomes. In Build 2, chunks flanked by gaps of $\ll$ 100cM but >10cM were observed on the maps associated with BTA13 (CEL23) and BTA28 (CEL15; Figure S2). Visual inspection of LD indicated that these chunks were incorrectly orientated segments of $\sim$10.5 and $\sim$24.9 cM in length, respectively (Figure S3a and S3b; Table 1). Reversal of marker orders in these regions resulted in map length reductions of 19.4 cM and 20.9 cM, respectively. Visual inspection of LD also confirmed fission of CEL19 and CEL31 (syn-

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

tenic to BTA1), with a 45.4cM inversion on CEL19 (Figure S3c). The X chromosome (BTA30, CEL34) in Build 2 was more fragmented, comprising 9 chunks (Figure S4). Visual inspection of LD in females indicated that chunks 3 and 7 occurred at the end of the chromosome, and that chunks 4, 5 and 6 were wrongly-oriented (Figure S5). After rearrangement into new marker orders, a sex-averaged map of each deer linkage group was reconstructed using the *chrompic* function (N = 37,932 SNPs, Figure S6).

Build 4: Solve minor local re-arrangements. Runs of SNPs from Build 3 were re-assigned to new chunks flanked by recombination fractions of $\geq$ 0.01 (1 cM). Maps were reconstructed to test whether inverting chunks of <50 SNPs in length and/or the deletion of chunks of <10 SNPs in length led to decreases in map lengths by $\geq$ 1cM. One wrongly-orientated chunk of 25 SNPs was identified on CEL15 (homologous to part of the inversion site identified on BTA28 in Build 3), and the marker order was amended accordingly (reducing the map length from 101.4 cM to 98.1 cM). Three chunks on the X chromosome (CEL34) shortened the map by $\geq$ 1cM when inverted and were also amended accordingly, reducing the X-chromosome map by 10.8 cM. The deletion of 35 individual SNPs on 14 linkage groups shortened their respective linkage maps by between 1cM and 6.3cM. A sex-averaged map of each deer linkage group was reconstructed using the *chrompic* function (N = 37,897 SNPs, Figure S7).

Build 5: Determining the location of unmapped markers and resolving phasing errors. In Builds 1 to 4, 372 SNPs in 89 chunks were removed from the analysis. To determine their likely location relative to the Build 5 map, LD was calculated between each unmapped SNP and all other SNPs in the genome to identify its most likely linkage group. The CRI-MAP *chrompic* function provides information on SNP phase (i.e. where the grandparent of origin of the allele could be determined) on chromosomes transmitted from the FID to offspring. The correlation between allelic phase was calculated for each unmapped marker and all markers within a 120 SNP window around its most likely position. A total 186 SNPs in 18 chunks could be unambiguously mapped back to the genome; for all other markers, their most likely location was defined as the range in which the correlation of allelic phase with mapped markers was $\geq$ 0.9 (Adjusted R^2). A provisional sex-averaged map of each deer linkage group was reconstructed using the *chrompic* function (N = 38,083 SNPs). Marker orders were reversed on deer linkage groups 6, 8, 16, 22 and 31 to match the orientation of the cattle genome.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Errors in determining the phase of alleles can lead to incorrect calling of double crossovers (i.e. two or more crossovers occurring on the same chromosome) over short map distances, leading to errors in local marker order. To reduce the likelihood of calling false double crossover events, runs of grandparental origin consisting of a single SNP (resulting in a false double crossover across that SNP) were recoded as missing (Figure S8) and the *chrompic* function was rerun. Of the remaining double crossovers, those occurring over distances of $\leq$ 10cM (as measured by the distance between markers immediately flanking the double crossover) were also recoded as missing. Finally, sex-averaged and sex-specific maps of each deer linkage group were reconstructed using the *chrompic* and *map* functions (Figure 2, Figure S9).

Build 6: Building a skeleton map and testing fine-scale order variations. In Build 5, 71.6% of intra-marker distances were 0cM; therefore, a "skeleton map" was created to examine local changes in marker orders. All runs of SNPs were re-assigned to new chunks where all SNPs mapped to the same cM position; of each chunk, the most phase-informative SNP was identified from the *.loc* output from the CRI-MAP *prepare* function (N = 10,835 SNPs). The skeleton map was split into windows of 100 SNPs with an overlap of 50 SNPs, and the CRI-MAP *flips* function was used to test the likelihood of marker order changes of 2 to 5 adjacent SNPs (*flips2* to *flips5*). Rearrangements improving the map likelihood by >2 would have been investigated further; however, no marker rearrangement passed this threshold and so the Build 5 map was assumed to be the most likely map order (Map provided in Table S1).

Determining the lineage of origin of chromosome rearrangements.

Lineages of origin and/or verification of potential chromosomal rearrangements was attempted by aligning SNP flanking sequences (as obtained from Brauning *et al.*, 2015) to related genome sequences using BLAST v2.4.0+ (Camacho *et al.*, 2009). ; Cattle and sheep diverged from deer $\sim$ 27.31 Mya, and diverged from each other $\sim$ 24.6 Mya (Hedges *et al.*, 2015)); therefore, rearrangements were assumed to have occurred in the lineage that differed from the other two. Alignments were made to cattle genome versions vUMD3.0 and Btau_4.6.1 and to the sheep genome Oar_v3.1 using default parameters in *blastn*, and the top hit was retained where $\geq$ 85% of bases matched over the informative length of the SNP flanking sequence.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Variation in recombination rate and landscape.

Estimated genomic positions were calculated for each SNP based on the difference in cattle base pair position of sequential markers. At the boundaries of inversions and fusions, the base pair difference between markers was estimated assuming that map distances of 1cM were equivalent to 1 megabase (Mb). In cases of fissions, the first base pair position was estimated as the mean start position of the cattle chromosomes. Estimated genomic positions are given in Table S1. The relationship between linkage map and chromosome lengths for each sex were estimated using linear regression in R v3.3.2. To investigate intra-chromosomal variation in recombination rates, the probability of crossing over was determined within 1 Mb windows using the estimated genomic positions. This was calculated as the sum of recombination fractions *r* within the window; the *r* between the first and last SNPs and each window boundary was calculated as $r \times N_{boundary}/N_{adjSNP}$, where $N_{boundary}$ is the number of bases to the window boundary and N_{adjSNP} is the number of bases to the adjacent window SNP. All deer chromosomes are acrocentric, with the exception of the X chromosome (CEL34) and one unknown autosome (Gustavsson & Sundt, 1968), and are orientated in the same direction as the cattle genome, where centromeres are located at the beginning of the acrocentric chromosomes (Band *et al.*, 2000; Ma *et al.*, 2015). Here, we assumed the position of the centromere was at the start of the chromosome. Variation in crossover probability relative to telomeric and centromeric regions was modelled using a loess smoothing function.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1: Synteny between the cattle and deer genomes. Large-scale fissions and fusions are informed by Slate et al (Slate *et al.*, 2002) and confirmed in this study through sequence alignment. The estimated length is calculated based on homologous SNP positions on the cattle genome BTA vUMD 3.0.

Deer Linkage Group (CEL)	Cattle Chr (BTA)	Number of Loci	Estimated Length (Mb)	Sex-averaged map length (cM)	Male map length (cM)	Female map length (cM)	Notes
1	15	1158	82.7	88.7	75.2	96.7	
2	29	663	50.3	55.4	51.6	57.5	
3	5	885	57.7	63.8	56.5	67.8	Fission from CEL22 in deer lineage.
4	18	971	65.2	81.3	72.5	85.9	
5	17, 19	2039	137.9	126.8	119.7	130.8	Fusion of BTA17 & BTA19 in deer lineage.
6	6	723	52.6	59.6	52.8	63.5	Fission from CEL17 in deer lineage.
7	23	660	51.7	64.0	60.6	65.7	
8	2	860	58.0	62.1	54.4	66.7	Fission from CEL33 in deer lineage.
9	7	1690	111.8	109.4	96.7	116.7	
10	25	580	42.7	55.3	49.1	59.2	
11	11	1547	107.1	101.3	81.7	112.1	
12	10	1486	102.1	104.2	94.0	110.0	
13	21	986	69.8	76.3	61.9	84.3	
14	16	1113	82.2	85.0	79.4	88.2	
15	26, 28	1357	96.4	96.4	79.2	105.9	Fission BTA26 & BTA28 cattle lineage. ∼13Mb inversion in deer linage and ∼1.5Mb inversion in cattle lineage on segment syntenic with BTA28.
16	8	674	47.0	54.8	52.8	56.2	Fission from CEL29 in deer lineage.
17	6	1059	68.3	67.0	59.0	71.5	Fission from CEL6 in deer lineage.
18	4	1831	120.7	108.0	98.8	113.3	
19	1	1476	101.9	99.3	85.1	107.3	Fission from CEL31, then ∼36Mb inversion in deer lineage.
20	3	1810	118.6	112.9	95.6	122.9	
21	14	1236	84.1	85.5	69.7	94.6	
22	5	882	62.3	65.2	55.2	71.1	Fission from CEL3 in deer lineage.
23	13	1200	83.3	95.1	84.6	101.1	∼5.9Mb inversion in cattle lineage.
24	22	885	61.3	69.7	59.1	75.9	
25	20	1066	72.1	76.0	66.9	80.6	
26	9	633	41.7	51.7	50.9	52.2	Fission from CEL28 in deer lineage.
27	24	886	62.5	62.2	47.8	70.7	
28	9	938	65.5	64.7	60.3	67.2	Fission from CEL26 in deer lineage.
29	8	969	67.2	65.9	59.2	69.4	Fission from CEL16 in deer lineage.
30	12	1220	86.2	86.9	74.4	94.0	
31	1	892	57.7	59.1	53.3	62.3	Fission from CEL19 in deer lineage.
32	27	623	46.7	56.7	52.5	59.2	
33	2	1220	80.4	80.8	70.3	86.9	Fission from CEL8 in deer lineage.
34	X	1865	148.2	148.7	40.0	138.9	Three translocations (two in deer, one in cattle) and one ∼ 18Mb inversion in cattle lineage; see Figure S5.
All		38083	2644.1	2739.7	2320.8	2906.3	Male and female autosomal maps of 2280.8 cM and 2767.4 cM, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

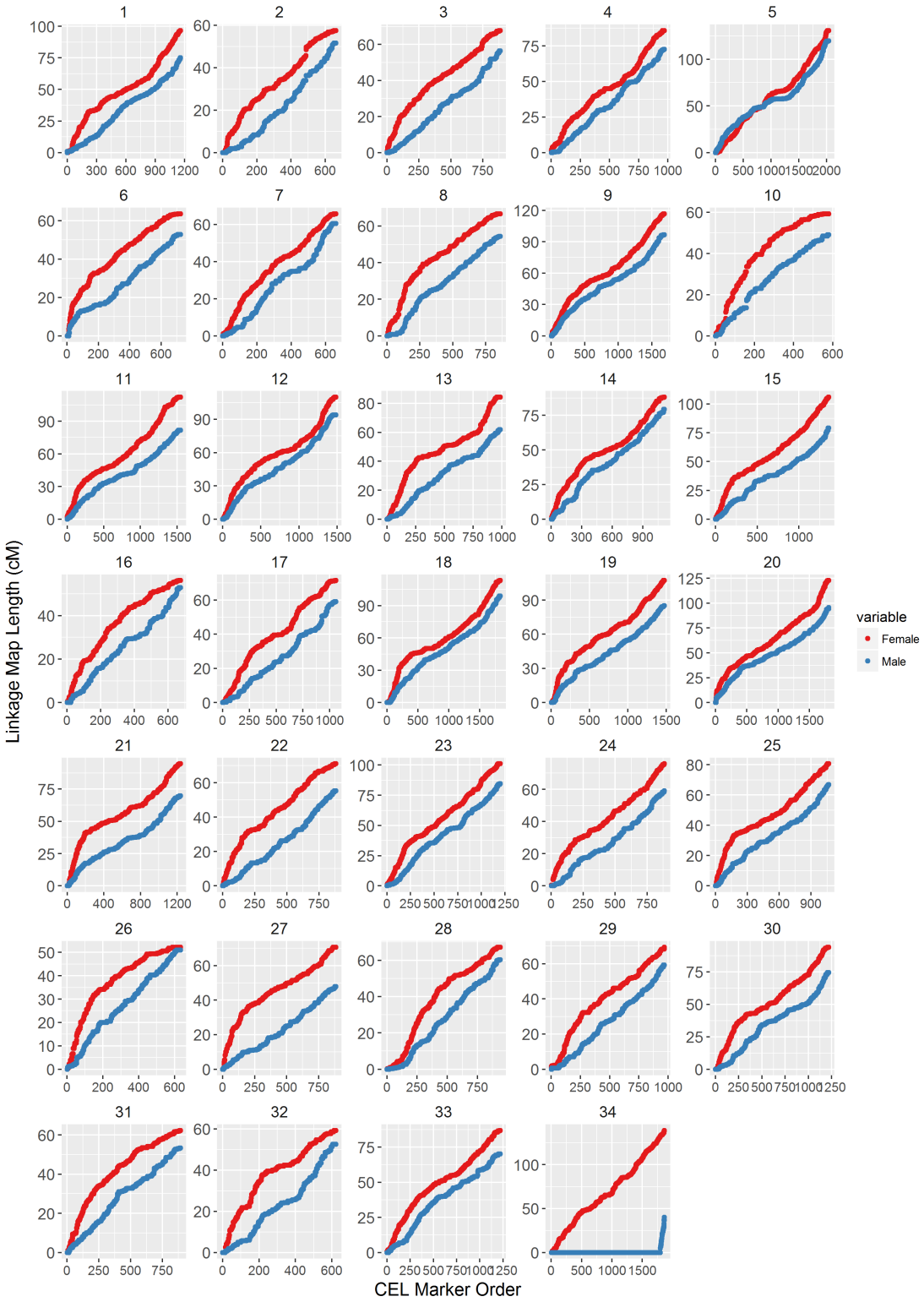

Figure 2: Sex-specific linkage maps for *Cervus elaphus* (CEL) linkage groups after Build 5. Map data is provided in Tables 1 and Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results

Linkage map.

The predicted sex-averaged red deer linkage map contained 38,083 SNP markers over 33 autosomes and the X chromosome (Full map provided in Table S1), and had a sex-averaged length of 2739.7cM (Table 1). A total of 71.6% of intra-marker recombination fractions were zero; a skeleton map of 10,835 SNPs separated by at least one meiotic crossover was also characterised (Table S1). The female autosomal map was 1.21 times longer than in males (2767.4 cM and 2280.8 cM, respectively, Table 1). In the autosomes, we observed six chromosomal fissions, one fusion and two large and formerly uncharacterised inversions occurring in the deer lineage (Table 1, Figure 3). Otherwise, the deer map order generally conformed to the cattle map order. The X chromosome had undergone the most differentiation from cattle, with evidence of three translocations, including two in the deer lineage and one in the cattle lineage, and one inversion in the cattle lineage (Figure S5, Table S2). The approximate positions of 90 unmapped markers are provided in Table S3.

Variation in recombination rate.

There was a linear relationship between estimated chromosome length and sex-averaged linkage map lengths (Adjusted R^2 = 0.961, Figure 4A). Smaller chromosomes had higher recombination rates (cM/Mb, Adjusted R^2 = 0.387, Figure 4B), which is likely to be a result of obligate crossing over. Female linkage maps were consistently longer than male linkage maps across all autosomes (Adjusted R^2 = 0.907, Figures S10) and correlations between estimated map lengths and linkage map lengths were similar in males and females (Adjusted R^2 = 0.910 and 0.954, respectively; Figure S11). In order to ensure that sex-differences in map lengths are not due to the over-representation of female meioses in the dataset, maps were reconstructed for ten subsets of 483 male and 483 female FIDs randomly sampled with replacement from the dataset; there was no significant difference between the true and sampled map lengths in males and females (Figure S13), suggesting accuracy in sex-specific map lengths. Fine-scale variation in recombination rate across chromosomes was calculated in 1Mb windows across the genome; recombination rate was considerably higher in females in the first $\sim$20% of the

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

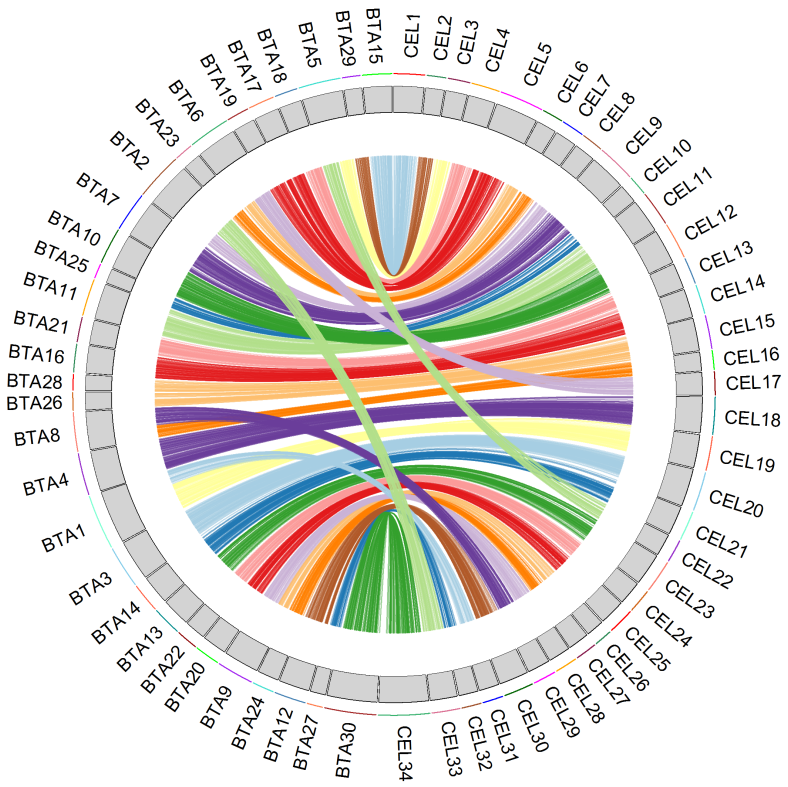

Figure 3: Comparison of marker positions on red deer linkage groups (right, ordered and prefixed CEL) and their predicted positions on cattle chromosomes (left, prefixed BTA). Data is shown for a sample of 2,000 SNPs. Large-scale rearrangement information and map data is provided in Tables 1 and Table S1, respectively. Plot was produced using the R package RCircos v1.1.3 (Zhang *et al.*, 2013) in R v3.3.2

chromosome, where the centromere is likely to be situated (Figure 5). This effect was consistent across nearly all autosomes, with the exception of CEL5 and CEL28 (Figure S12). Male and female recombination rates were not significantly different across the rest of the chromosome, although male recombination was marginally higher than females in sub-telomeric regions where the centromere was absent (Figure 5).

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

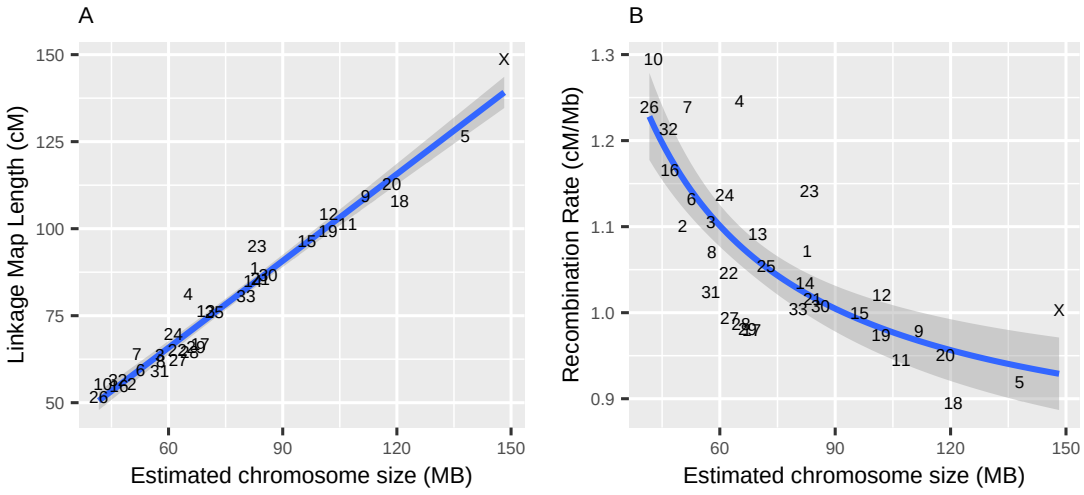

Figure 4: Broad-scale variation in recombination rate, showing correlations between (A) sex-averaged linkage map length (cM) and estimated chromosome length (Mb) and (B) estimated chromosome length (Mb) and chromosomal recombination rate (cM/Mb). Points are chromosome numbers, and lines and the grey-shaded areas indicate the regression slopes and standard errors, respectively.

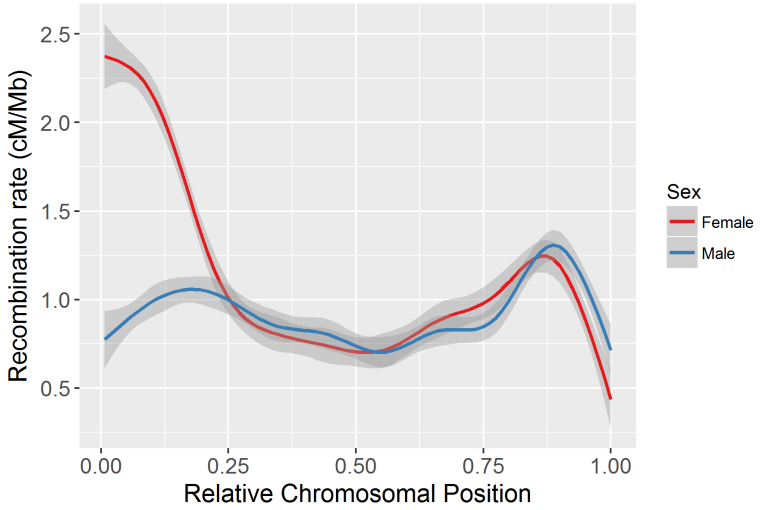

Figure 5: Loess smoothed splines of recombination rates across 33 autosomes for males and females. With the exception of one unknown linkage group, all chromosomes are acrocentric with the centromere at the beginning of the chromosome. Splines for individual chromosomes are shown in Figure S12

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Discussion

Utility of the red deer linkage map. In this study, we constructed a predicted linkage map of 38,083 SNPs and a skeleton map of 10,835 SNPs for a wild population of red deer. The predicted map included 98.8% of polymorphic SNPs within this population. Whilst several large-scale rearrangements were identified in the red deer lineage (Table 1, Figure 3), marker orders generally showed strong concordance to the cattle genome order. We are confident that the maps presented here are highly accurate for the purposes of genetic analyses outlined in the introduction; however, we also acknowledge that some errors are likely to be present. The limited number of meioses characterised means that we cannot guarantee a correct marker order on the predicted map at the same centiMorgan map positions, meaning that some small rearrangements may be undetected within the dataset. Furthermore, the use of the cattle genome to inform initial marker order may also introduce error in cases of genome misassembly. Considering these issues, we recommend that the deer marker order is used to verify, rather than inform any *de novo* sequence assembly in the red deer or related species.

Rearrangement of the X chromosome. The X chromosome (CEL34) showed the highest level of rearrangement, including two translocations in the deer lineage, one of which was a small region in the pseudoautosomal region (PAR) remapped to the distal end of the chromosome (Figure S5). The X chromosome also showed a similar pattern to the autosomes in the relationship between estimated chromosome length (Mb) and linkage map length (cM, Figure 4). This may seem counter-intuitive, as recombination rates in the X should be lower due to it spending one-third of its time in males, where meiotic crossovers only occurs on the PAR. However, female map lengths were generally longer, and 64% of the meioses used to inform sex-averaged maps occurred in females; furthermore, the female-specific map showed that the X conformed to the expected map length. Therefore, the linkage map length of the X is as expected, and not inflated due to mapping errors. Nevertheless, we acknowledge again that mapping errors may be present on the X, particularly given that fewer informative meioses can be detected on in non-PAR regions.

Sexual dimorphism in recombination landscape. Females had considerably higher recombination rates in pericentromeric regions, resulting in female-biased recombination rates overall;

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

recombination rates along the rest of the chromosome were similar in both sexes and lowest at the closest proximty to the telomere (Figure 5). Identifying female-biased heterochiasmy is not necessarily unusual, as recombination rates in placental mammals are generally higher in females (Lenormand & Dutheil, 2005), and are hypothesised to protect against aneuploidy (i.e. non-disjunction) after long periods of meiotic arrest (Morelli & Cohen, 2005; Coop & Przeworski, 2007; Nagaoka *et al.*, 2012). However, the pattern of heterochiasmy observed in this dataset is striking for several reasons. First, our findings are distinct from the other ruminants, namely cattle and sheep, which both exhibit male-biased heterochiasmy driven by elevated male recombination rates in sub-telomeric regions, with similar male and female recombination rates in pericentromeric regions (Ma *et al.*, 2015; Johnston *et al.*, 2016). Second, mammal studies to date show markedly increased male recombination rates at the sub-telomere, even if female recombination rates are higher overall (e.g. in humans and mice); this is not observed in the red deer map. Third, whilst female recombination rates tend to be relatively higher in pericentromeric regions in many species, they are generally suppressed in very close proximity the centromere (Brandvain & Coop, 2012); evidence from human studies suggests that increased centromeric recombination can result in aneuploid gametes (Lamb *et al.*, 2005).

At present, the mechanisms and biological significance of elevated female recombination rates around the centromere in female deer remains unclear, although ideas have been proposed to explain variation in sex differences more generally. One proposition is selection on gametes at the haploid stage in males may vary relative to the strength of sperm competition, favouring particular rates of recombination (Lenormand & Dutheil, 2005; Mank, 2009). In addition, increased telomeric recombination may allow more rapid sperm turnover during gametogenesis (Tankimanova *et al.*, 2004). Both of these ideas this may explain the difference in patterns observed between sheep and cattle (strong and moderate rates of sperm competition, respectively) with deer, which have relatively low sperm competition in the rut (Clutton-Brock *et al.*, 1982). However, support for this sperm competition hypothesis is only weakly supported (Trivers, 1988; Mank, 2009) and still fails to explain increased centromeric recombination in females. Other explanations may relate to reduced crossover interference at the centromere in females (Fledel-Alon *et al.*, 2009) and increased recombination rates between female meiotic drive elements and the centromere (Brandvain & Coop, 2012), although neither appear to the explain consistency of sex-differences and elevated female recombination across all acrocentric chromosomes (Figure S12). Regardless, more empirical investigation is required to elucidate the specific drivers of sex differences in recombination rate and elevated centromeric recom-

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

bination in this system. Future work will investigate the heritability and genetic architecture of recombination rate at an individual level and investigate the potential role of recombination modifiers in driving rate variation within this species.

Acknowledgements

We thank T. Clutton-Brock, F. Guinness, S. Albon, A. Morris, S. Morris, M. Baker and many others for collecting field data and DNA samples and their important contributions to the long-term Rum deer project. Discussions with Camillo Bérénos & Judith Risse aided the study in its early stages; we also appreciate feedback from Graham Coop, Loukas Theodosiou, Annaliese Mason and Yaniv Brandvain on interpreting some of the results. We thank Scottish Natural Heritage for permission to work on the Isle of Rum National Nature Reserve, and Wellcome Trust Clinical Research Facility Genetics Core in Edinburgh for performing the genotyping. This work has made extensive use of the resources provided by the University of Edinburgh Compute and Data Facility (http://www.ecdf.ed.ac.uk/). The long-term project Rum deer is funded by the UK Natural Environment Research Council, and SNP genotyping and current work were supported by a European Research Council Advanced Grant to J.M.P..

Author Contributions

J.M.P and J.H. organised the collection of samples. P.A.E. and J.H. conducted DNA sample extraction and genotyping. J.M.P. and S.E.J. designed the study. S.E.J. analysed the data and wrote the paper. All authors contributed to revisions.

References

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.

Band MR, Larson JH, Rebeiz M, *et al.* (2000) An ordered comparative map of the cattle and human genomes. *Genome Res.*, **10**, 1359–1368.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

322 Bérénos C, Ellis PA, Pilkington JG, Lee SH, Gratten J, Pemberton JM (2015) Heterogeneity of
323 genetic architecture of body size traits in a free-living population. *Mol. Ecol.*, **24**, 1810–1830.

324 Bradbury IR, Hubert S, Higgins B, *et al.* (2013) Genomic islands of divergence and their con-
325 sequences for the resolution of spatial structure in an exploited marine fish. *Evol. Appl.*, **6**,
326 450–461.

327 Brandvain Y, Coop G (2012) Scrambling eggs: meiotic drive and the evolution of female recom-
328 bination rates. *Genetics*, **190**, 709–23.

329 Brauning R, Fisher PJ, McCulloch AF, *et al.* (2015) Utilization of high throughput genome se-
330 quencing technology for large scale single nucleotide polymorphism discovery in red deer
331 and canadian elk. *bioRxiv*.

332 Brieuc MSO, Waters CD, Seeb JE, Naish KA (2014) A dense linkage map for chinook salmon
333 (oncorhynchus tshawytscha) reveals variable chromosomal divergence after an ancestral
334 whole genome duplication event. *G3: Genes|Genomes|Genetics*, **4**, 447–460.

335 Camacho C, Coulouris G, Avagyan V, *et al.* (2009) BLAST+: architecture and applications. *BMC*
336 *Bioinf.*, **10**, 1–9.

337 Clutton-Brock T, Guinness F, Albon S (1982) *Red Deer. Behaviour and Ecology of Two Sexes.*
338 University of Chicago Press.

339 Coop G, Przeworski M (2007) An evolutionary view of human recombination. *Nat Rev Genet*,
340 **8**, 23–34.

341 Ellegren H (2014) Genome sequencing and population genomics in non-model organisms.
342 *Trends Ecol. Evol.*, **29**, 51–63.

343 Fierst JL (2015) Using linkage maps to correct and scaffold de novo genome assemblies: meth-
344 ods, challenges, and computational tools. *Front. Genet.*, **6**, 220.

345 Fledel-Alon A, Wilson DJ, Broman K, *et al.* (2009) Broad-scale recombination patterns underly-
346 ing proper disjunction in humans. *PLos Genet.*, **5**, e1000658.

347 Fountain T, Ravinet M, Naylor R, Reinhardt K, Butlin RK (2016) A linkage map and QTL analysis
348 for pyrethroid resistance in the bed bug *Cimex lectularius*. *G3: Genes|Genomes|Genetics*, **6**,
349 4059–4066.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Frantz AC, Pourtois JT, Heuertz M, *et al.* (2006) Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population. *Mol. Ecol.*, **15**, 3191–3203.

Green P, Falls K, Crooks S (1990) *Documentation for CRIMAP, version 2.4*. Washington University School of Medicine.

Gustavsson I, Sundt CO (1968) Karyotypes in five species of deer (*Alces alces* l., *Capreolus capreolus* l., *Cervus elaphus* l., *Cervus nippon nippon* temm. and *Dama dama* l.). *Hereditas*, **60**, 233–248.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.*, **32**, 835–845.

Huisman J (2017) Pedigree reconstruction for SNP data: parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.*, **In revision**.

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM (2016) Inbreeding depression across the lifespan in a wild mammal population. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 3585–3590.

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

Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW (2016) Genomics advances the study of inbreeding depression in the wild. *Evol. Appl.*, **9**, 1205–1218.

Kawakami T, Smeds L, Backström N, *et al.* (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**, 4035–4058.

Kruuk LEB, Slate J, Pemberton JM, Brotherstone S, Guinness F, Clutton-Brock T (2002) Antler size in red deer: heritability and selection but no evolution. *Evolution*, **56**, 1683–1695.

Lamb NE, Sherman SL, Hassold TJ (2005) Effect of meiotic recombination on the production of aneuploid gametes in humans. *Cytogenet. Genome Res.*, **111**, 250–255.

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science*, **265**, 5181.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire PA (2016) A dense brown trout (*Salmo trutta*) linkage map reveals recent chromosomal rearrangements in the salmo genus and the impact of selection on linked neutral diversity. *bioRxiv*.

Lenormand T, Dutheil J (2005) Recombination difference between sexes: a role for haploid selection. *PLoS Biol.*, **3**, e63.

Ma L, O'Connell JR, VanRaden PM, *et al.* (2015) Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. *PLos Genet.*, **11**, e1005387.

Mank JE (2009) The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. *Genet. Res.*, **91**, 355–363.

McKinney GJ, Seeb LW, Larson WA, *et al.* (2016) An integrated linkage map reveals candidate genes underlying adaptive variation in chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol. Resour.*, **16**, 769–783.

Morelli MA, Cohen PE (2005) Not all germ cells are created equal: Aspects of sexual dimorphism in mammalian meiosis. *Reproduction*, **130**, 761–781.

Nagaoka SI, Hassold TJ, Hunt PA (2012) Human aneuploidy: mechanisms and new insights into an age-old problem. *Nat. Rev. Genet.*, **13**, 493–504.

Rastas P, Calboli FCF, Guo B, Shikano T, Merilä J (2016) Construction of ultradense linkage maps with Lep-MAP2: Stickleback F2 recombinant crosses as an example. *Genome Biol. Evol.*, **8**, 78–93.

Senn HV, Pemberton JM (2009) Variable extent of hybridization between invasive sika (*Cervus nippon*) and native red deer (*C. elaphus*) in a small geographical area. *Mol. Ecol.*, **18**, 862–876.

Slate J, Van Stijn TC, Anderson RM, *et al.* (2002) A deer (subfamily *Cervinae*) genetic linkage map and the evolution of ruminant genomes. *Genetics*, **160**, 1587–1597.

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.*, **14**, 43–59.

Tankimanova M, Hultén Ma, Tease C (2004) The initiation of homologous chromosome synapsis in mouse fetal oocytes is not directly driven by centromere and telomere clustering in the bouquet. *Cytogenet. Genome Res.*, **105**, 172–181.

bioRxiv preprint doi: https://doi.org/10.1101/100131; this version posted January 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

408 Trivers R (1988) Sex differences in rates of recombination and sexual selection. In *The evolution*
409 *of sex.* (edited by R Michod, B Levin), pp. 270–286. Sinauer Press., Sunderland, MA, USA.

410 Zhang H, Meltzer P, Davis S (2013) RCircos: an R package for Circos 2D track plots. *BMC*
411 *Bioinf.*, **14**, 244.