Molecular Genetics and Genomics (2018) 293:665–684
https://doi.org/10.1007/s00438-017-1412-3

ORIGINAL ARTICLE

The red deer *Cervus elaphus* genome CerEla1.0: sequencing, annotating, genes, and chromosomes

Nóra Á. Bana[1,2] · Anna Nyiri[1] · János Nagy[2] · Krisztián Frank[1,2] · Tibor Nagy[1] · Viktor Stéger[1] · Mátyás Schiller[1] · Péter Lakatos[4] · László Sugár[2] · Péter Horn[2] · Endre Barta[1,5] · László Orosz[1,3]

Received: 1 June 2017 / Accepted: 19 December 2017 / Published online: 2 January 2018

Abstract

We present here the de novo genome assembly CerEla1.0 for the red deer, *Cervus elaphus*, an emblematic member of the natural megafauna of the Northern Hemisphere. Humans spread the species in the South. Today, the red deer is also a farm-bred animal and is becoming a model animal in biomedical and population studies. Stag DNA was sequenced at 74× coverage by Illumina technology. The ALLPATHS-LG assembly of the reads resulted in 34.7×10^3 scaffolds, 26.1×10^3 of which were utilized in Cer.Ela1.0. The assembly spans 3.4 Gbp. For building the red deer pseudochromosomes, a pre-established genetic map was used for main anchor points. A nearly complete co-linearity was found between the mapmarker sequences of the deer genetic map and the order and orientation of the orthologous sequences in the syntenic bovine regions. Syntenies were also conserved at the in-scaffold level. The cM distances corresponded to 1.34 Mbp uniformly along the deer genome. Chromosomal rearrangements between deer and cattle were demonstrated. 2.8×10^6 SNPs, 365×10^3 indels and 19368 protein-coding genes were identified in CerEla1.0, along with positions for centromerons. CerEla1.0 demonstrates the utilization of dual references, i.e., when a target genome (here *C. elaphus*) already has a pre-established genetic map, and is combined with the well-established whole genome sequence of a closely related species (here *Bos taurus*). Genome-wide association studies (GWAS) that CerEla1.0 (NCBI, MKHE00000000) could serve for are discussed.

Keywords *Cervus elaphus* genome · Deer genome · *Bos taurus* genome · Cattle genome · Next-generation sequencing · De novo assembly

Introduction

Genome-wide association studies (GWAS) have gained importance in the arsenal of today's genetic analyses. Whole genome sequence assemblies provide the most important support for GWAS. In this paper, we present the genome assembly CerEla1.0 for the red deer, and *Cervus elaphus*. CerEla1.0 could serve manifold Genome-wide association studies (GWAS), including, for instance, organ regeneration, comparative analyses of orthologous mammalian genes, their promoters, regulatory networks, SNP sets, development of chromosome-specific microsatellites for individual identifications, and DNA markers for refined population and evolutionary studies. Red deer is an emblematic member of the natural megafauna of the Northern Hemisphere, and is displayed in Neolithic cave paintings. Humans introduced and spread the species in the Southern Hemisphere. Today, the red deer is both a wild animal and also a half-domesticated farm-bred livestock for venison and velvet antler products, and is getting recognized as a model animal for bone, osteoporosis and regeneration research as well as for population and evolutionary studies. Red deer, the Royal Game of the Middle Ages in Europe, is surrounded by respect and amazement in many cultures. For instance, in historical legends, a mythological Wonder Deer led the Hungarians to their present land. Antler and skeleton remains excavated from

Communicated by S. Hohmann.

Nóra Á. Bana and Anna Nyiri equal contributor.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-017-1412-3) contains supplementary material, which is available to authorized users.

✉ László Orosz
orosz.laszlo@abc.naik.hu

Extended author information available on the last page of the article

time to time document that red deer have lived in the Carpathian Basin for at least 7000 years. Due to the very high average quality and often record size of our trophies as well as for spiritual and emotional reasons, red deer is a "charismatic" member of the Hungarian megafauna and stands in the center of attention (Bán 1998). The estimated wild population size was around 100,000 in 2010 (Szabolcsi et al. 2014).

The closest relative of *C. elaphus* is the sika deer *Cervus nippon*. The two species can interbreed and produce fertile offspring in co-occurring natural populations (McDevitt et al. 2009; Biedrzycka et al. 2012). In the Far-East, the sika deer (*C. nippon*) has also been in the focus of interest for a long time, including farming, meat production, and alternative medicine. In China, antler extracts have been used for tonic since ancient times.

In modern biomedical research, the antler mainly serves as a model for studies on bone development or regeneration of a complete organ (Li and Suttie 2001, 2012; Li et al. 2002, 2007; Price and Allen 2004; Park et al. 2004; Price et al. 2005; Gu et al. 2008; Kierdorf et al. 2009). The advent of next-generation sequencing (NGS) technologies added further impetus to antler research with transcriptome analyses (Yao et al. 2012; Ba et al. 2016), which were extended toward more complex developmental studies of the whole body of the deer (Jia et al. 2016).

In the past decade, we introduced deer research in our laboratory on a molecular biological basis; we cloned and functionally identified a set of genes which are differentially expressed during the rapid but controlled proliferation and robust mineralization of the developing antler, or involved in skeletal osteoporosis coupled with the annual antler cycle of the deer (Molnár et al. 2007; Gyurján et al. 2007; Stéger et al. 2010). Physiological deer osteoporosis and pathological human osteoporosis were compared, and new genes of significant roles were identified (Borsy et al. 2009; Balla et al. 2008). In another line of our deer project, we developed a highly sensitive tetranucleotide STR multiplex PCR for identification of individuals and for population and forensic studies ("DeerPlex", Szabolcsi et al. 2014). The mtDNA sequence of the Central European red deer was established by the NGS approach (Frank et al. 2016); the mitochondrial DNA lineages of the Carpathian Basin were described (Frank et al. 2017). Our ambition of whole genome sequencing and to assemble the first draft reference genome for *C. elaphus* stems from these efforts, as well as from the strong tradition in bioinformatics and functional genomics at the NARIC Agricultural Biotechnology Institute, Gödöllő, Hungary (Barta et al. 2016).

Compilation of DNA markers and the application of interspecific hybridization and back-crosses provided a way to develop the basis for the classical genetic analysis for deer, a so far non-model animal. The excellence and the power of the concept were demonstrated by the construction of the *C. elaphus* genetic map (Tate et al. 1995; Slate et al. 2002a).

This genetic map of 621 sites (2532 cm in length with 5.7 cm average spacings) integrated modern day technologies and methods, including among others: comparative genomics and orthologous DNA marker alleles drawn from ruminants and other mammals, making interspecific F1 hybrids between milu (Pere David's deer, *Elaphurus davidianus*) and red deer *C. elaphus*, building an F2 mapping herd of a reasonable size by artificial inseminations from the cross of F1 stags × red deer hinds, computing recombination frequencies by LOD support, map making using the MapMaker V3.0 computer program and high resolution statistics. The deer genetic map served as the origin of further studies such as among others, the evolution of ruminant genomes (Slate et al. 2002a), QTL scan (Slate et al. 2002b), genome-wide SNP search (Fisher et al. 2015; Brauning et al. 2015), and whole genome annotation and assembling pseudochromosomes (this work).

In this study, using next-generation sequencing technologies, we assembled the pseudochromosomes and annotated the resulting reference genome CerEla1.0 of the Central-European red deer, *Cervus elaphus hippelaphus*. Two references, the (recombination) genetic map of red deer by Slate et al. (2002a), the genome sequence of *Bos taurus* (an evolutionally close species, NCBI Btau_5.0.1) and the rule of co-linearity between the genetic map and the chromosome were utilized, as was an independent work published in bioRxiv by Fisher et al. (2015) and Brauning et al. (2015). The concept may be useful for genome annotation in the case of other non-model animals. Shortly before the completion of the CerEla1.0 manuscript, Johnston et al. (2017) published the very high-density genetic map of *C. elaphus* (based on 38,000 SNPs) in bioRxiv. We believe that this genetic map will provide a rich source for further improvement of CerEla1.0.

The reference genome sequence CerEla1.0 and the pseudochromosome complement of Central-European red deer (*C. elaphus hippelaphus*) may provide a basis for broader interests including, among others, conservation genetics, refined evolution, and population studies within the family Cervidae and Pecora, for identification of descents for autosomal, maternal and paternal lineages, for forensic identification, for defining allelic compositions behind phenotypes important, for example, in game management, or in biomedical research and applications, like in bone research, osteoporosis, organ regeneration and tumour biology.

Materials and methods

Collection of samples

Blood (3 × 10 ml) was taken from a living animal of 7 years of age (Fig. 1), a capital stag with shed antlers weighing 12 kg at the age of 11 years, at the Deer Farm of the Game Management and Landscape Center of Kaposvár University (Bőszénfa, Hungary) (the definition of a capital individual is: a stag of exceptionally strong antlers, whose trophy would score at least 170 CIC points, i.e., that would be at least a bronze medalist.). In this case, the stag would have been golden medalist (at least 240 CIC points) by far, significantly above 210 CIC points, the minimum for gold medal. Sample collection was performed by a trained veterinarian according to standard veterinary medical practice with a permission from the Hungarian Veterinary Chamber [Hungarian Animal Rights Law (243/1998, XII.31)]. Samples were preserved in EDTA buffer and stored at − 20 °C.

DNA preparation and sequencing

Total genomic DNA was extracted from the blood samples using the Duplicα Prep Automated DNA/RNA Extraction System (EuroClone S.p.A., Italy). Isolated DNA samples were sent to Aros Applied Biotechnology (Aarhus, Denmark). As a custom service, one paired (~ 500 bp) and one mate pair (1.2 kb) library were prepared and sequenced in two and one illumina HiSeq 2000 lane (2 × 101 bp), respectively.

Fig. 1 Photograph of the *C. elaphus* stag (crot. No. 3016). This stag donated the DNA for the CerEla1.0 genome program. The stag was born in 2003. This snapshot was taken in 2010 at the Deer Farm of the Game Management and Landscape Center of Kaposvár University (Bőszénfa, Hungary). The blood samples for DNA preparation were taken in 2011. At the age of 11, his shed antlers weighted 12.0 kg (would be above 240 CIC points highly gold medalist, i.e., above 210 CIC points)

Sequence assembly and sequence analysis

The reads were assembled into contigs and scaffolds by the Broad Institute's ALLPATHS-LG (Gnerre et al. 2011) program using mostly the default parameters. The program ran on an SGI UV 1000 machine (1152 intel CPU core and 6 TB memory) provided by the Hungarian National Infrastructure Development Program. All the further sequence analyses were done using standard bioinformatics/genomics programs EMBOSS (Rice et al. 2000), BWA (Li and Durbin 2009), lastz (Harris 2007), BLAST (Altschul et al. 1990; Korf et al. 2003), etc., and in-house custom-made Bash, Python, and Perl scripts.

Pseudochromosome assembly

Genetic marker sequences were identified in UCSC, UNIPROT, NCBI, and ENA databases and retrieved for further analysis. Microsatellite marker sequences were extended by the surrounding regions using the genome sequence of the species from where they originated. They were mapped to both the cattle reference genomes (Baylor Btau_4.6.1, Bos_taurus_UMD3.1.1, and Btau_5.0.1) and the scaffolds using blastn (with the following options: -1e-10-outfmt 6-best_hit_score_edge 0.05-best_hit_overhang 0.25-max_target_seqs 3). In the following iterative steps, we used the MUMmer3.0 program (Kurtz et al. 2004) with nucmer option and bwa-mem together with our custom Bash scripts to map selected scaffolds and contigs into the masked bovine genomes to determine the order and orientation of scaffolds and contigs in the red deer pseudochromosomes. The final generation of pseudochromosome sequences was carried out by a custom-made Python script.

Genome annotation

Pseudochromosome sequences were used in the MAKER version 2.31.8 (Cantarel et al. 2008; Cambell et al. 2014) genome annotation pipeline. MAKER identifies and masks repetitive elements in the genome with RepeatMasker-open-4.0.5 and RepeatRunner programs. Therefore, first, an organism-specific repeat library suitable for repeat masking was created. This was achieved by the RepeatModeler-1.0.4 software using RECON version 1.08 and Repeat Scout version 1.0.5. While RepeatMasker identifies known repeats, RepeatModeler (Smit et al. 2013–2015) predicts novel repeat sequences in the genome. In the next step, MAKER generated the initial (evidence-based) gene models based on expressed sequence tag (EST), mRNA, and protein evidences. As EST evidence, cDNA from cattle (*Bos taurus*)

was used (ftp://ftp.ensembl.org/pub/release-88/fasta/bos_taurus/cdna/). For RNA-seq evidence, we downloaded the raw RNA-seq reads from a sika deer (*Cervus nippon*) experiment (Yao et al. 2012) and the transcriptome using the TrinityRNAseq pipeline (http://trinityrnaseq.sf.net) (Grabherr et al. 2011). We downloaded the cattle, the human (*Homo sapiens*) and the sheep (*Ovis aries*) complete proteomes from the ENSEMBL database (ftp://ftp.ensembl.org/pub/release-88/fasta/homo_sapiens/pep/; ftp://ftp.ensembl.org/pub/releas-88/fasta/ovis_aries/pep/), and used them as protein homology evidence. For repeat masking, artiodactyl was applied as model organism for RepBase masking in RepeatMasker, together with the previously generated organism-specific library and MAKER's internal database containing transposable elements. BLAST was used for aligning EST, mRNA, and protein sequences to the genome and repeat identification. Soft-masking rather than hard-masking was applied in BLAST. MAKER used Exonerate software to refine BLAST alignments. Next, SNAP (Korf 2004) and AUGUSTUS (Stanke et al. 2006) gene prediction programs were trained and employed to create ab-inito (non evidence-based) gene predictions. The evidence-based gene models were used to train AUGUSTUS and SNAP (Cambell et al. 2014). After training SNAP, the predicted set of genes is used to retrain (bootstrap) SNAP. To obtain the final annotation set, the evidence-based and the ab-initio gene predictions were integrated, filtered, and optimized by MAKER. The MAKER's derived gene model is in gff3 format, which can be easily loaded into genome browsers. Finally, to identify protein functions, the InterProScan (Jones et al. 2014) software was used to identify and to further annotate protein-coding genes. Different types of RNA-coding genes were annotated using MAKER. We identified ribosomal RNAs and microRNAs with the blastn program using the small and large subunit (SSU, LSU) sequences of mammalian ribosomal RNA of the SILVA123 (Quast et al. 2013) database and 21 mammalian sequences of the miRBase (Kozomara et al. 2014) database. We searched 5S units of ribosomal RNA with barrnap-0.6 (http://www.vicbioinformatics.com/software.barrnap.shtml). We found the transfer RNAs with tRNAscan–SE-1.3.1 (Schattner et al. 2005) software by eliminating the short interspersed and the pseudo tRNAs. In the command, we used -H -f -m -o options.

Genetic variant detection

The raw reads of both the paired-end and mate pair libraries were mapped to the reference genome using bwa-mem (Version: 0.7.10-r789). The unmapped read pairs were withdrawn for further analysis. We used SAMtools software (Li et al. 2009) (Version: 0.1.19–44428cd) to detect genetic variants, with "mpileup -D -S -E -uf" command line options. The vcf files were generated with BCFtools "view -bvcg" parameters and we applied the vcfutils.pl script's "varFilter -D 188" option to filter those variants, where the coverage was maximum 188 (3× the average genome coverage). The variants whose Phred quality score was higher than 30 were filtered with custom Perl script. The annotation of the variant's function was performed with the ANNOVAR software (Wang et al. 2010), where we created our own gene definition database using a MAKER-derived gff3 file. To annotate the genetic variants' functions, we used the table_annovar script.

Data availability

The raw reads have been deposited into the SRA database (SRR4013902). The reference genome sequence has been submitted to the NCBI database and can be accessed at the following accession number (MKHE00000000). The gene annotation and the variation tracks are available for browsing and downloading from the JBrowse web page http://emboss.abc.hu/wonderdeer/JBrowse.

Results

Genome sequence and assembly: reads, contigs, and scaffolds

A stag, a 7-year-old Central European red deer (*Cervus elaphus hippelaphus*) with capital antlers (shed antlers of 12 kg) was the source of the DNA, prepared from blood samples. The stag was bred in the Bőszénfa Deer Farm of Kaposvár University, Kaposvár, Hungary (Fig. 1).

The red deer stag has 33 pairs of autosomes, plus the X and Y chromosomes. For DNA sequencing, a paired-end and a mate pair library were constructed in Aros Applied Biotechnology (Aarhus, Denmark) as a custom service, and sequenced by Illumina technology (Illumina HiSeq2000). The combined length of all reads (2.2×10^9) added up to 222.7 Gbp, which corresponded to about 74× coverage of the sequence libraries and 62× of the haploid genome of *C. elaphus*, assuming that the genome size was similar to that of related ruminants like cattle (*Bos taurus*) or sheep (*Ovis aries*), i.e., around 3 Gbp. Both the initial de novo assembly and the scaffolding were carried out with the ALLPATHS-LG program. According to the ALLPATHS-LG assembly, the reads added up to 437,412 contigs with a total contig length of 1.95 Gbp. The contigs assembled to 34,724 scaffolds with a combined length of 3.4 Gbp (i.e., "total scaffold length, with gaps"). Read statistics is shown in Table 1, ALLPATH-LG report on statistics in Table S1.

Table 1 Read statistics: number of reads and bases of paired-end and mate pair libraries

Read sequence libraries	Number of reads	Number of bases
Paired-end		
Ce_PE_s1_1		
L003	202,009,674	20,402,977,074
Ce_PE_s1_2		
L004	204,044,499	20,608,494,399
Ce_PE_s2_1		
L005	184,790,536	18,663,844,136
Ce_PE_s2_2		
L006	188,274,662	19,015,740,862
Mate pair		
Ce_MP_1		
L007	163,020,536	16,465,074,136
Ce_MP_2		
L008	160,624,627	16,223,087,327
Total		
1×	1,102,764,534	111,379,217,934
2×	2,205,529,068	222,758,435,868

Assembly, guided by the reference genetic map of *C. elaphus*, mapmarker scaffolds (MMSc-s)

A genetic map of *C. elaphus* with 5.7 cM average spacings, derived from data of the interspecific back-cross F2 population between *C. elaphus* × Elaphurus *davidianus* (David peer's deer) by Slate et al. (2002a), was utilized as the first step for reference-guided assembly. In accordance with the haploid chromosome number, the genetic map split into 34 linkage groups (X and Y drawn together) with 2532 cM in combined lengths. From the 621 genetic markers mapped by Slate et al., we could map 365 (Table S2). These were defined by spacious DNA marker alleles (derived from ESTs, RFLVs, STSs, and protein sequences, which were adapted mainly from *Artiodactyla*, less from hominids and rodents). The rest were AFLPs, with no sequence data available. We unambiguously identified the DNA sequence of these 365 map markers in the deer scaffolds and called them "mapmarker" scaffolds (abbreviated as MMSc). By this step, the MMSc-s gained a defined position in the map array: in exchange, the site of the genetic map became part of an extended and specific DNA sequence of the deer genome (i.e., as MMSc, Fig. 2 provides a typical example, *C. elaphus* linkage group 1).

Co-linearity between the genetic map of *C. elaphus* and the sequence of the *B. taurus* genome

The 365 *C. elaphus* "mapmarker" scaffolds were probed against the two versions of the DNA sequence of the *B. taurus* genome (NCBI Btau_4.6.1, Btau_5.0.1). We assumed that the bovine genome as template would provide high-fidelity alignments to the *C. elaphus* scaffolds. Our previously developed "zoo clonings" (Gyurján et al. 2007; Stéger et al. 2010) and sequencings over a number of deer genes strengthened this belief, since 98–100% similarity was registered in exons and more than 90% in the 5′ regulatory regions/promoters. The 365 deer MMSc-s uniquely identified the corresponding orthologous positions in the bovine genome. The array of the MMSc-s on the deer genetic map and the array of the bovine orthologues in the bovine genome were co-linear along all the chromosomes of the two species. Moreover, the syntenies for the in-MMSc genes were identical (see below in the section "Intra-scaffolds and intra-contigs syntenies").

The majority of deer chromosomes (19 autosomes plus X, Y) could be directly paired with single *B. taurus* homologous chromosomes (for a typical example, see Fig. 2). Interestingly, eight acrocentric deer chromosomes appear to be tandemly fused into four acrocentric bovine ones (for a typical example, see Fig. S3), and conversely, one acrocentric deer chromosome appears to be split to two acrocentric bovine ones (Fig. S1). One metacentric deer chromosome appears to be centrically split to two acrocentric ones in the *B. taurus* complement (Fig. 3). The structural relation between deer chromosomes 19 and 31 vs. bovine 1 (i.e., Ce19 and 31 vs. Bt1) as well as for Ce26 and 28 vs. Bt9 seemed more complex. In the case of "Ce19 and 31 vs. Bt1", a tandem fusion of the two acrocentric *C. elaphus* chromosomes and a translocation are combined (Fig. 4), whereas the tandem fusion of acrocentric chromosomes Ce26 and Ce28 in acrocentric Bt9 displayed a paracentric inversion in relation of Ce28 and Bt9 (Fig. S2). All structural relations of Ce and Bt chromosomes are summarized in Table 2. These observations are in complete accordance with the comparative cytogenetic analyses by Bonnet et al. (2001) between cattle and sika deer (*C. nippon*).

The co-linearity test between the deer map and the sheep (*O. aries*) genome (ISGC Oar_v3.1) was repeated with a very similar outcome (data not shown).

Filling up the segments of the *C. elaphus* genetic map with scaffolds, guided by reference genes of *B. taurus* (RGSc-s)

The high similarity of the DNA sequences of individual genes of cattle and deer and the co-linearity of the map markers (MMSc-s) made it possible that aligning the deer scaffolds along the bovine genome sequence as template also led to a reliable genomic order for the deer scaffolds, i.e., the co-linearity and syntenies would manifest themselves at a lower, sub-chromosomal scale.

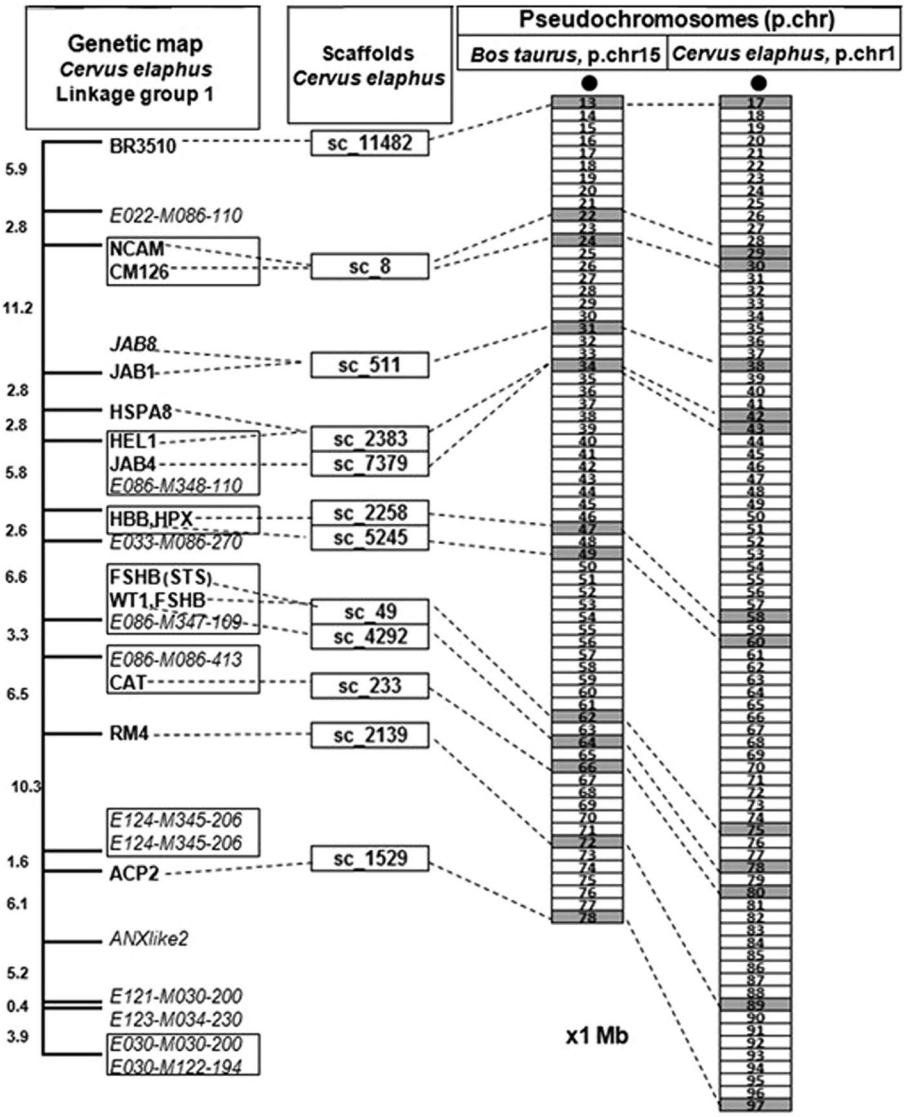


Fig. 2 Co-linearity of genetic map markers and map marker scaffolds (MMSc) of deer *Cervus elaphus* pseudochromosome and linkage group 1 (in genome program CerEla1.0) and *Bos taurus* pseudochromosome 15 (in genome program Btau_5.0.1). Numbers on the deer genetic map segments correspond to cM distances taken from Slate et al. (2002a). Numbering along the pseudochromosomes correspond to distances in Mbp. Black dots correspond to centromerons. Similar relations were found for pseudochromosomes Ce2/Bt29, Ce4/Bt18, Ce7/Bt23, Ce9/Bt7, Ce10/Bt25, Ce11/Bt11, Ce12/Bt10, Ce13/Bt21, Ce14/Bt16, Ce18/Bt4, Ce20/Bt3, Ce21/Bt14, Ce23/Bt13, Ce24/Bt22, Ce25/Bt20, Ce27/Bt24, Ce30/Bt12, Ce32/Bt27, CeX/BtX, and CeY/BtY. Note: Slate et al. 2002a applied the Kosambi mapping function to calculate cM map distances (Kosambi 1943) instead of the standard Haldane function (Haldane 1919)

First, bovine reference genes corresponding to the cognate genetic map segments of deer were selected (UCSC Btau_4.6.1 and Bos_taurus_UMD.3.1.1) one by one along the bovine genomic DNA segments. Next, using the bovine reference gene sequences as "bait" and the deer scaffolds as "prey", deer scaffolds that carried the orthologous deer gene of the bovine reference (we named these RGSc, reference gene-containing scaffolds) were angled. Consequently, after probing the deer RGScaffold sequences against the bovine genome sequence, the deer (genetic) map segments (defined by the deer/bovine DNA markers/MMSc-s at the flanks) were "filled up" with the deer RGScaffolds. In this way, beyond the 365 MMSc-s, a further 6013 RGScaffolds were arrayed—following the bovine gene order—in the corresponding deer genetic map segments. The final outcome, with the scaffolds arranged along the *C. elaphus* pseudochromosomes, is shown at http://emboss.abc.hu/wonderdeer/JBrowse, and a detailed example for one chromosome segment is shown in Fig. 5.

Filling up inter-scaffold gaps

The bait/prey relation was reversed in this search, i.e., DNA sequences of *C. elaphus* scaffolds served as bait and the bovine genomic segments as prey. All the map marker- and the reference gene-containing scaffolds (MMSc-s and RGSc-s) were omitted here from use as baits; hence, only those carrying so far unidentified genetic elements like protein-coding genes, rRNA, tRNA, and miRNA genes remained on the screen. We called these IRGSc-s, inter-reference genes' scaffolds. Furthermore, the IRGSc-s should be located in the bovine genome in the gaps between an MMSc and an RGSc or between two RGSc-s. Altogether, 13,748 scaffolds contained genetic elements. In-scaffold syntenies of genes

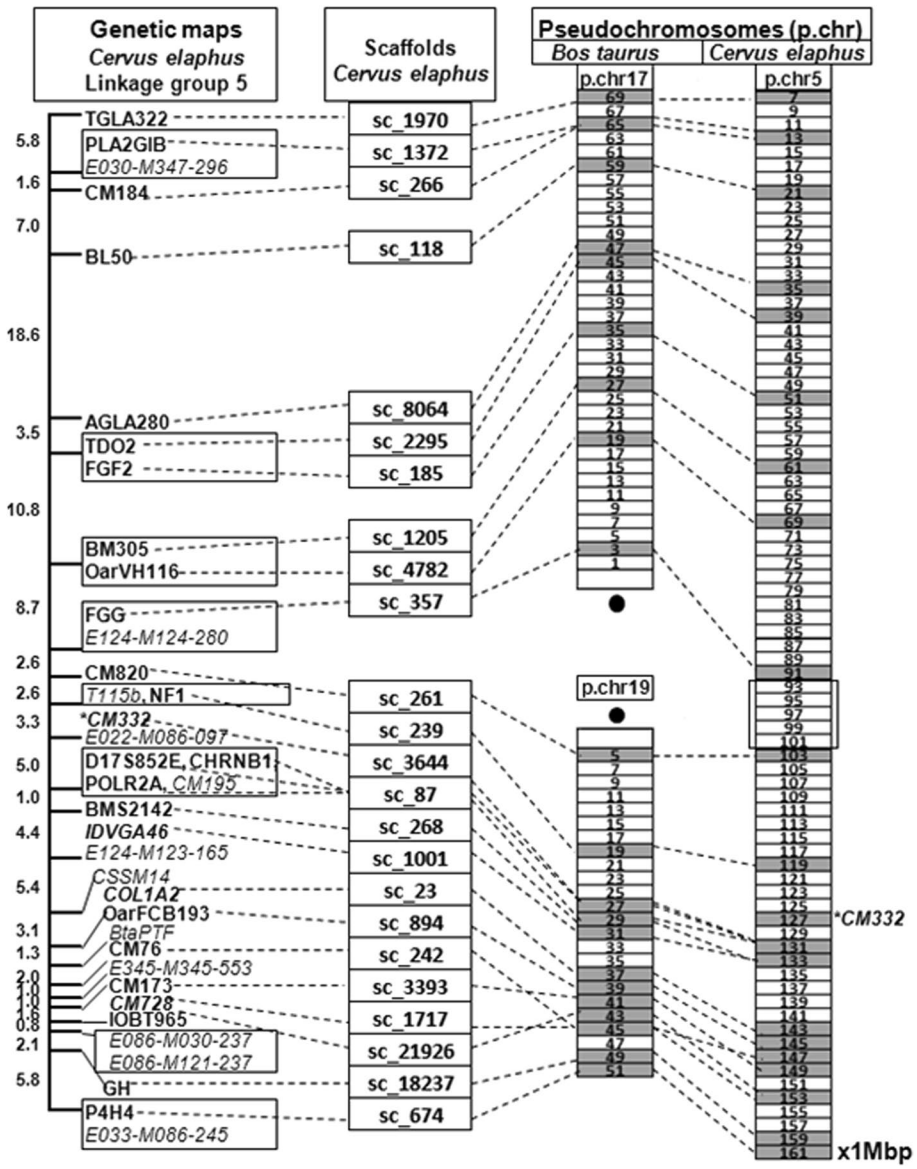


Fig. 3 Example for the fission of deer metacentric chromosome Ce5 in two acrocentric bovine chromosomes Bt17 and Bt19. Numbering for cM and Mbp distances is similar to that in Fig. 2. Black dots correspond to centromeres

were then tested for deer–bovine relation (see the section below and Fig. 5).

After aligning the MMSc-s, RGSc-s, and IRGSc-s (13,748 combined), 15,205 scaffolds longer than 2Kb still remained non-localized. To find their positions, regions of the bovine genome sequence overlapping the previously mapped 13,748 deer scaffolds were masked out, and the remaining regions served as template for probing the 15,205 still non-localized scaffolds. In this way, 9845 new scaffolds (called GFSc-s, gap filling scaffolds) gained chromosomal positions (although in bovine order) resulting in 23,593 scaffolds having homologous regions in the bovine reference genome.

Although most of these homologous scaffolds (99.6%, 23,491/23,593) could be unambiguously aligned with the bovine genome, the localization for 102 scaffolds still remained ambiguous. These scaffolds were next broken up and rearranged in 35 new scaffolds and 2582 individual contigs, and then re-aligned scaffold by scaffold and contig by contig. In this way, we could solve all the intra-scaffold contradictions and find their final positions in the red deer pseudochromosomes.

Finally, the pseudochromosome complement of the haploid red deer (*C. elaphus hippelaphus*) reference genome (CerEla1.0) was assembled from 23,491 plus 35 scaffolds (MMSc, RGSc, IRGSc, GFSc-s plus the 35 new ones) and 2582 individual contigs, adding up to 26,108 sequence elements and spanning 3.4 Gbp.

We were unable to locate 11,444 scaffolds, adding up to 52,989,442 bp sequence. We classified these into the unplaced category.

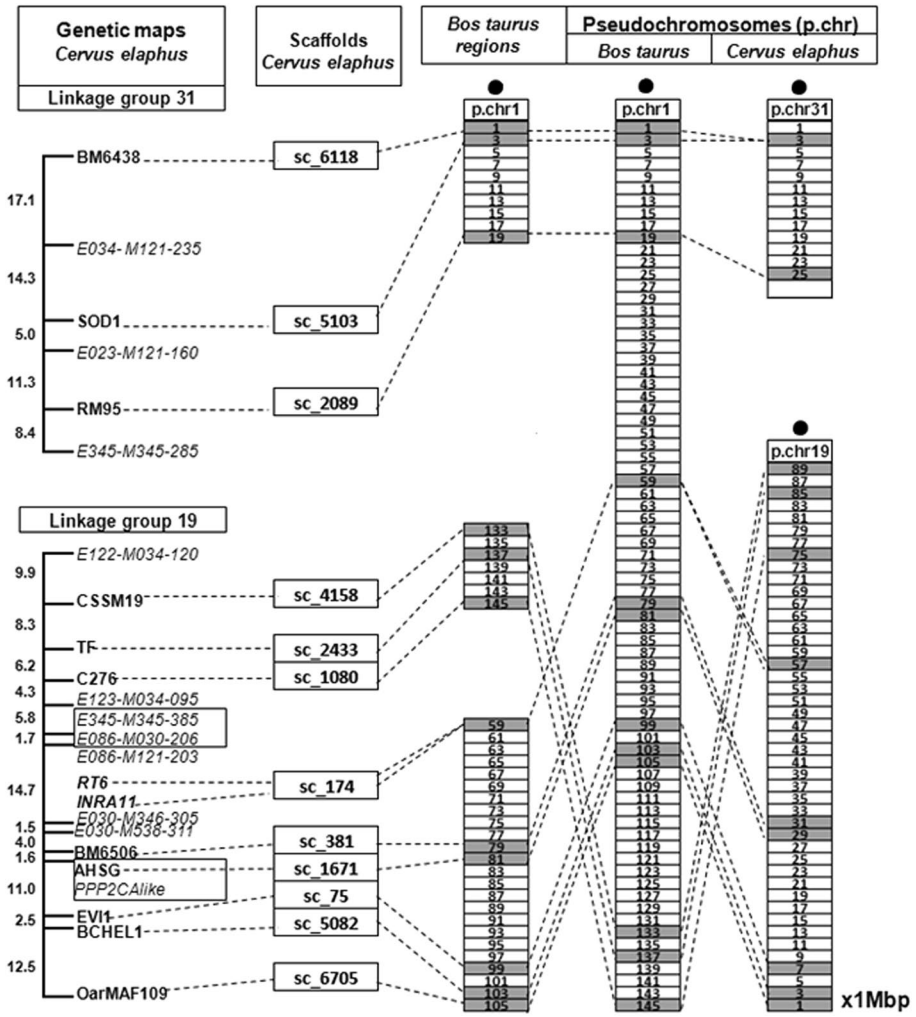


Fig. 4 Fusion of two deer chromosomes Ce19 and Ce31 in one Bt chromosome Bt1 coupled with a translocation. Note, a putative break in the C276–RT6/INRA11 region could convert the Ce19 and Bt1 sequences into each other. Numbers on the deer genetic map segments correspond to cM distances. Numbering along the pseudochromosomes correspond to distances in Mbp. Black dots correspond to centromerons

Intra-scaffold and intra-contig syntenies

We selected scaffolds containing more than one genetic element (i.e., syntenic genes), and tested these intra-scaffold syntenies against the bovine genome (Btau_5.0.1). In cases when uncertainty arose, we also considered the sheep genome. In all cases investigated (3422 scaffolds), we registered identical local, in-scaffold syntenies for the deer and for the cognate, orthologous overlapping bovine chromosomal segments. Hence, the syntenies appeared not only at the chromosomal level for the map marker sites, but also at the sub-chromosomal (scaffold) level (examples are shown in Fig. 5 for scaffolds sc-8, sc-9, and sc-511).

Annotation of the *C. elaphus* genome

To annotate the genes on the red deer pseudochromosomes, we applied the MAKER annotation pipelines. As evidences, we used the complete bovine transcriptome, the human and ovine proteome, and the de novo partial transcriptome of sika deer. Using this pipeline, we could identify 19,368 genes. As mentioned above, the syntenies and order of the intra-scaffold deer genes were identical with those of the cognate bovine genes. The number of annotated genes is shown in Table 3.

We could also annotate 589 rRNA genes (LSU, SSU), adding up to 98.3 Kbp sequence (0.0029% of the *C. elaphus* genome CerEla1.0), 1029 5 s rRNA genes (96 Kb and 0.0028% of CerEla1.0), 2096 tRNA genes (128 Kbp and 0.0038% of CerEla1.0), and 264 microRNA genes (27.7 Kbp, 0.0008% of CerEla1.0). These figures were in very good agreement with those found in other mammalian genomes.

Repetitive sequences were annotated both as part of the MAKER pipeline and also separately with dedicated programs. We applied the RepeatMaker program, which marked 22.73% of the genome as repetitive. In aggregate, it represents a 769,492,957-bp-long DNA sequence (Table 4).

It is worth mentioning that our previously developed tetranucleotide STRs for 10 sites (DeerPlex for multiplex PCR analyses, Szabolcsi et al. 2014) were all identified in the scaffolds. Eight of these were in different chromosomes,

Table 2 Pseudochromosome complement of *Cervus elaphus*: some basic characteristics

C.e. chr	Mapmarker position (Mbp) C.e.1.0[a]		Corresponding for B.t. chr		
	Centromeron				
	Proximal	Distal	B.t. chr	C*	Structure/alignment
1 A	BR3510	ACP2	15 A	15	*---Ce1
	17,149,645	97,596,143			*---Bt15
2 A	TGLA86	KRN1	29 A	29	*---Ce2
	60,822,183	463,507			*---Bt29
3 A	ILSTS42	OarMAF23	5 A	5	*---Ce3 (*---Ce22)
	4,768,942	66,970,424			*--------Bt5----------
4 A	INRA38	JP23	18 A	18	*---Ce4
	11,079,846	76,570,752			*---Bt18
q	FGG	TGLA322	17 A	f *	Ce5q----*----Ce5p
5 M	92,136,286	7,052,020		17/19	
	CM820	P4H4	19 A		Bt17—* *---Bt19
p	102,829,424	161,668,917			
6 A	PDGFRA	PDE6B	6 A	n.c.*	(*---Ce17) *---Ce6
	62,582,512	2,029,781			*----------Bt6-------
7 A	Bta_BoLaDIB	Delta14	23 A	23	*---Ce7
	9,457,097	49,425,282			*---Bt23
8 A	TGLA226	ALPL	2 A	n.c.*	(*---Ce33) *---Ce8
	52,007,152	5,210,075			*----------Bt2-------
9 A	RM12	ILSTS6	7 A	7	*---Ce9
	6,623,210	1,213,442,281			*---Bt7
10 A	HBA	CM21	25 A	25	*---Ce10
	229,713	50,800,499			*---Bt25
11 A	SLC8A1	ASS	11 A	11	*---Ce11
	109,429,674	9,405,010			*---Bt11
12 A	BM3033	BMS2614	10 A	10	*---Ce12
	124,818,172	7,220,242			*---Bt10
13 A	IGF1R	T193	21 A	21	*---Ce13
	8,595,779	79,944,215			*---Bt21
14 A	CR2	BM1706	16 A	16	*---Ce14
	6,509,868	87,172,047			*---Bt16
15 A	RT5	CYP17	28 A	28	*--------Ce15------
	6,402,305	89,468,482	26 A		*---Bt28 *---Bt26
16 A	LPL	ORM1	8 A	n.c.*	(*---Ce29)*---Ce16
	58,006,661	10,424,686			*----------Bt8-------
17 A	ILSTS93	SPP1	6 A	6	*---Ce17 (*---Ce6)
	886,652	47,385,900			*----------Bt6------
18 A	RM188	OarHH064	4 A	4	*---Ce18
	22,742,514	146,986,314			*---Bt4
19 A	CSSM19	OarMAF109	1 A	n.c.*	(*---Ce31) *---Ce19
	90,353,433	157,115			*----------Bt1--------
20 A	INRA6	TGLA127	3 A	3	*---Ce20
	11,820,085	131,656,106			*---Bt3
21 A	CSSM66	BM2934	14 A	14	*---Ce21
	6,097,957	87,929,022			*---Bt14
22 A	TEXAN15	ACO2	5 A	n.c.*	(*---Ce3) *---Ce22
	59,941,417	10,015,108			*----------Bt5--------
23 A	PLC154	OarMAF18	13 A	13	*---Ce23
	2,389,036	102,873,535			*---Bt13

Table 2 (continued)

C.e. chr	Mapmarker position (Mbp) C.e.1.0[a]		Corresponding for B.t. chr		
	Centromeron				
	Proximal	Distal	B.t. chr	C*	Structure/alignment
24 A	HUJ175	HIS-H1	22 A	22	*---Ce24
	27,402,800	70,501,351			*---Bt22
25 A	BM1225	BM4107	20 A	20	*---Ce25
	7,980,198	54,541,102			*---Bt20
26 A	CM102	SOD2	9 A	n.c.*	(Ce28–i--*) *---Ce26
	9,035,997	44,533,880			*----------Bt9---------
27 A	JP38	PAI2	24 A	24	*---Ce27
	34,310,557	84,583,784			*---Bt24
28 A	OarCP021	ETH225	9 A	9(i)	Ce28–i--* (*---Ce26)
	81,944,836	13,060,454			*----------Bt9----------
29 A	UWCA47	CM100	8 A	8	*---Ce29(*---Ce16)
	1,242,245	76,016,626			*----------Bt8-------
30 A	RM178	ILSTS33	12 A	12	*---Ce30
	21,709,783	108,535,018			*---Bt12
31 A	BM6438	RM95	1 A	1	*---Ce31 (*---Ce19)
	2,858,627	25,506,844			*----------Bt1--------
32 A	BM6526	BM203	27 A	27	*---Ce32
	15,764,812	55,918,160			*---Bt27
33 A	INRA40	INHBB	2 A	2	*---Ce33 (*---Ce8)
	2,090,945	93,576,890			*----------Bt2--------
X A	B9	T27b	X M	n.c.*	*-------CeX
	11,882,131	51,457,707			--BtX–*----

A/M, acrocentric/metacentric chromosome; p/q, short/long chromosomal arms; C.e./B.t., C. elaphus/B. taurus; i, inversion; *, centromeron; c*, centromeron of C.e. chromosome corresponds to centromeron of B.t. chromosome; f*, fused centromerons; n.c.*, no corresponding centromeron for B.t. chromosomes, *---- *----- > *---------- tandem fusion of two acrocentric chromosomes, --------* *-------- > --------*-------- centric fusion (Robertsonian translocation) of two acrocentric chromosomes

and the pair (C01–T26) which matched the same linkage group in the bovine genome (Bt16) was also linked to the homologous deer chromosome 14 in the deer genome.

SNP pattern/heterozygosity along CerEla1.0 assembly

Since we sequenced the DNA of a stag, all the autosomes provided two copies of DNA. It allowed us to find and annotate the heterozygous autosomal variations in this individual (Table 5 and http://emboss.abc.hu/wonderdeer/JBrowse). We aligned the original reads to the pseudochromosomal sequences and determined the SNVs and the small indels. We could identify 2,807,458 SNVs. This number is only slightly lower than the 3.2–3.7 million that we found when we sequenced three Hungarian Mangalica pigs (Molnár et al. 2014). We also found 364,689 indels. Using the result of the MAKER annotation pipeline, we further annotated the heterozygous SNVs. In this way, we found 17,700 nonsynonymous and 14,252 synonymous variations.

Centromeron positions in the CerEla1.0 assembly

The karyotype for *C. elaphus* (as well as for some other *Cervidae*) and for *B. taurus* have been well established by classical chromosome cytogenetics (Gustavsson and Sundt 1968; Fontana and Rubini 1990; Bonnet et al. 2001). The structures of the chromosome complements are regarded "primitive", since nearly, all the chromosomes are acrocentric (A chromosomes) and only rare metacentric ones (M chromosomes) can be recorded as a result of centric fusions of acrocentrics (i.e., Robertsonian translocations). Interestingly, however, tandem fusions and fissions (i.e., two acrocentrics to one acrocentric and vice versa) are often selected karyologically during the phyletic evolution of *Pecoran* descendant lineages (e.g., bovide, sheep, and deer). It is worth to note here that in this work, the comparisons for

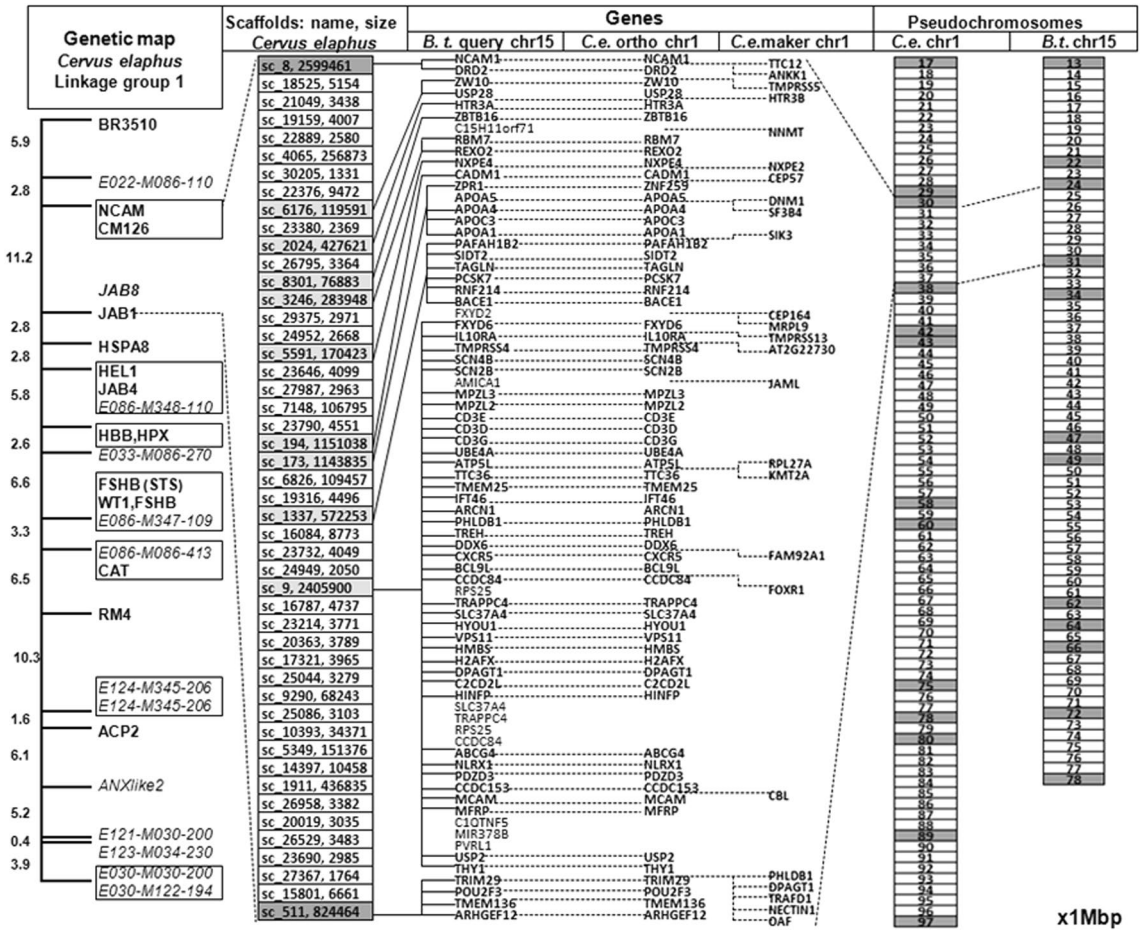


Fig. 5 Examples for in-scaffold and in-contig syntenies in the CerEla1.0 genomic segment defined by MMSc-s sc_8 and sc_511. Note, syntenies parallel in deer genomic scaffolds and in the corresponding bovine genomic segments. Numbers at left: cM distances in deer genetic map. Middle: note the perfect syntenies for in-scaffold genes (e.g., along the 2.4-Mbp-long RGSc sc_9). Similar syntenies were recorded along the entire CerEla1.0 genome (see Results: "Intra-scaffolds and intra-contigs syntenies")

the *C. elaphus* and *B. taurus* pseudogenomes displayed the same chromosomal associations and homeologies (one-to-one correspondence, fusion, scission, and translocation) as those reported in the detailed comparative studies by Bonnet et al. (2001) on the chromosome RBG banding patterns of the sika deer (*C. nippon*) and *B. taurus*, which allowed us to combine their cytogenetic results with the genomics of this study (see below, Table 2). In contrast to the cytogenetic studies, the genetic map of deer did not define the positions of the centromerons (Table S3). Unlike in deer, in the case of cattle genetic markers close to the centromeron had been defined for all 29 autosomes and the X chromosome (Ma et al. 1996; Everts-van der Wind et al. 2004, 2005). We identified these markers in the successive *B. taurus* genome projects, although no markings or notes were found for the centromerons in the sequences of the pseudochromosomes (NCBI Bos_taurus_UMD3.1.1, Btau_5.0.1). However, comparisons of the orthologous genetic arrays of the deer and bovine pseudochromosomes combined with the cytogenetic comparisons of deer and cattle chromosomes (i.e., alignment of banding patterns of metaphase chromosomes) made possible to assign centromerons to all deer linkage groups (i.e., 34) and to the cognate pseudochromosomes (Table 2, slots 2 and 3). Furthermore, all the bovine homologies for deer centromerons were established (Table 2, slots 4 and 5). In certain cases, comparisons of the orthologous genetic arrays of the pseudochromosomes combined with the cytogenetic comparisons of deer and cattle chromosomes (i.e., alignment of RBG banding patterns of metaphase chromosomes, Bonnet et al. 2001) made possible, alone and independently, to assign centromerons to 6 deer linkage groups and to the cognate pseudochromosomes: Ce5, 15, 19, 26, 28, and 31. Briefly, Ce5 M is proven independently by cytology and genomics (Fig. 3 of this work and Fig. 2 in Bonnet et al. 2001); Ce15 A, the centromeron corresponds to that of Bt28, proven by the correspondence of its centromeron–proximal part to Bt28, and of the distal part to Bt26 (i.e., "tandem fusion", Fig. S1 in this work and Fig. 2 in Bonnet et al.

Table 3 Summary of data of denovo assembly of CerEla1.0

Cervus elaphus chr	Total length (bp)	Ungapped length (bp)	Chr[a] length (cM)	Sequence elements (sc, ctg)	Number of genes	No. of deer markers used	Relation to chromosomes			
							Sika deer[b]	Cattle[a]	Sheep[a]	Human[a]
1	104,504,130	60,726,700	78.1	963	622	16	11	15	15	11
2	63,256,308	35,767,465	74.9	572	541	11	29	29	21	11
3	88,455,658	48,939,352	57	627	530	12	27	5	3	12
4	81,198,604	48,908,856	98.4	622	1035	11	1p	18	14	19
5	178,027,732	108,827,520	99	1046	1698	26	2	17, 19	17, 11	4, 12, 17
6	73,112,236	38,596,877	79	705	333	13	22	6	6	4
7	66,841,537	38,713,878	48.2	498	492	10	23	23	20	6
8	55,920,303	34,458,394	54.8	287	455	11	18	2	2	1, 2
9	141,945,777	83,052,372	84.7	1016	1025	15	5	7	5	5, 19
10	55,943,575	33,198,136	52.2	361	702	17	30	25	24	7, 16
11	140,388,259	82,430,176	64.3	1017	910	16	8	11	3	2, 9
12	127,780,087	75,235,002	116.4	901	794	20	16	10	7	14, 15
13	89,791,019	52,029,727	112.5	646	472	8	21	21	18	14, 15
14	103,589,943	59,604,123	80.2	680	587	15	14	16	12	1
15	125,274,996	74,133,175	53.8	782	712	10	9	26, 28	22, 25	1, 10
16	62,950,657	35,715,860	55.6	577	327	6	32	8	2	8, 9
17	79,719,380	43,062,969	58.7	619	246	11	16	6	6	4
18	152,661,129	87,311,698	97.2	1178	626	11	4	4	4	7
19	127,240,616	73,001,574	83.9	928	619	10	7	1	1	3, 21
20	149,338,601	92,562,240	102.6	1129	1132	17	3	3	1	1
21	107,358,006	61,629,440	76.8	1036	409	8	13	14	9	8
22	63,924,421	36,199,049	80.2	474	520	7	19	5	3	22
23	109,465,363	65,617,900	74.3	690	687	14	1q	13	13	10, 20
24	78,163,535	47,342,068	76.3	450	519	11	26	22	19	3
25	96,537,449	52,762,107	58.5	720	289	7	20	20	16	5
26	55,103,472	31,310,669	66.3	360	240	9	31	9	8	6
27	84,643,473	47,073,830	63.1	606	321	6	24	24	23	18
28	82,073,021	44,568,841	85.1	653	245	8	17	9	8	6, 9
29	80,165,896	45,644,972	59.9	599	307	7	15	8	2	9
30	117,797,301	63,651,907	100.2	1066	382	8	10	12	10	13
31	75,455,120	38,504,182	56.1	686	193	3	25	1	1	21
32	60,010,529	32,242,645	46.3	433	196	6	28	27	26	4, 8
33	121,428,352	67,204,072	119	855	463	6	12	2, 22	2	2, 3
X	181,543,317	86,341,570	18.4	2265	716	4	X	X	X	X
Y	4,026,935	1,810,667	–	61	23	1	Y	Y	Y	Y
Unplaced	52,989,442	–	–	11,444	–	–		–	–	–

Table 3 (continued)

Cervus elaphus chr	Total length (bp)	Ungapped length (bp)	Chr[a] length (cM)	Sequence elements (sc, ctg)	Number of genes	No. of deer markers used	Relation to chromosomes			
							Sika deer[b]	Cattle[a]	Sheep[a]	Human[a]
Total	3,438,626,179	1,928,180,013	2532	37,552	19,368	371		–	–	–

[a]Data transferred from Slate et al. (2002a)

[b]Deduced from this work and from Bonnet et al. (2001)

2001); and Ce19 A corresponds to the centromeron distal half of Bt1. The centromeron position is proven by the observation that a well-defined centromeron–proximal segment of Ce19 "translocated during the evolution" to the end of Bt1. It is worth noting that the perfect congruence of the cytology and genomics is demonstrated by Fig. 4 in this work and by Figs. 2 and 5 in the paper of Bonnet et al. (2001); Ce26 A, the position of the centromeron is proven by the correspondence of Ce26 to the distal half of Bt9 (what is revealed as if a "tandem fusion of Ce26 and 28" happened, Fig. S2 of this work and Fig. 2 in Bonnet et al. 2001); and Ce28 A, the centromeron corresponds to that of Bt9. However, putting the centromeron in place, we considered that the banding pattern of Ce28 and that of the proximal part of Bt9 can be aligned, if a large paracentric inversion occurred in evolution (as demonstrated by Bonnet et al. 2001), and if, as a further precondition, this inversion overlapped the entire region defined by the orthologous map marker sites in both species (Fig. S2); and Ce31 A, the centromeron corresponds to that of Bt1. Its position is proven by the alignment of Ce31 with the centromeron–proximal half of Bt1, furthermore by Bt1's appearing as a "tandem fusion of Ce31 and 19" (see Fig. 4 and also at Ce19 A, above).

On physical lengths and cM distances

The genetic map of *C. elaphus* we used for reference is 2532 cM long (Slate et al. 2002a); the assembled CerEla1.0 reference genome, as shown in this work, spans 3.4 Gbp, i.e., 1 cM corresponds to 1.34 Mbp in CerEla1.0. This figure is significantly higher than the 1Mbp/1 cM "thumb rule" or the 0.8 Mb/1 cM value established for the bovine genome (Arias et al. 2009). It is worth mentioning, however, that the deer and bovine mapping systems were significantly different, since the deer mapping system was an interspecific back-cross population, whereas in the case of cattle, it was based on an F2 (i.e., intra-specific) outbred population. It is also noteworthy that in the recently announced high-density genetic map of *C. elaphus*, the cM/Mbp ratio was proven to be 1.04 in a pedigree-based mapping system (Johnston et al. 2017).

The length of the deer pseudogenome CerEla1.0, we present here is 25% longer than that published for cattle in Btau_5.0.1 (NCBI, Science), namely, 3.4 and 2.7 Gbp, respectively. It is worth noting that the 2.7 Gbp Btau_5.0.1 genome that we used for reference and for template is built up only from sequenced DNA sections (i.e., no NNNs inserted). To test the 0.7 Gbp "surplus" length of CerELA1.0 in smaller genomic distances, the homologous genomic segments in the CerEla1.0 and Btau_5.0.1 pseudogenomes were compared. These segments were the ones defined by the orthologous map markers syntenic in both species. As shown in Fig. 6, with few exceptions, all along the deer and

Table 4 Repetitive sequences in the CerEla1.0 genome

File name: C.e.1.0.fasta				
Sequences: 35				
Total length: 3,385,636,737 bp (1,928,192,114 bp excl N/X runs)				
GC level: 41.51%				
Bases masked: 769,492,957 bp (22.73%)				
Number of length percentage				
Elements[a] occupied of sequence				
SINEs		1,443,280	209,069,252 bp	6.18%
	Alu/B1	0	0 bp	0%
	MIRs	348,401	49,923,175 bp	1.47%
LINEs		1,039,158	393,951,179 bp	11.64%
	LINE1	520,359	231,587,901 bp	6.84%
	LINE2	227,931	56,065,241 bp	1.66%
	L3/CR1	30,317	6,214,363 bp	0.18%
	RTE	259,649	99,947,899 bp	2.95%
LTR elements		300,340	96,940,119 bp	2.86%
	ERVL	66,416	24,866,835 bp	0.73%
	ERVL–MaLRs	107,869	34,041,529 bp	1.01%
	ERV_classI	77,337	31,321,116 bp	0.93%
	ERV_classII	33,875	3,264,732 bp	0.1%
DNA	Elements	250,418	49,072,131 bp	1.45%
	hAT-Charlie	141,928	26,225,569 bp	0.77%
	TcMar–Tigger	38,629	10,057,487 bp	0.3%
Unclassified		4585	769,585 bp	0.02%
Total interspersed repeats			749,802,266 bp	22.15%
Small RNA		216,401	35,304,753 bp	1.04%
Satellites		2624	699,977 bp	0.02%
Simple repeats		377,096	15,587,604 bp	0.46%
Low complexity		62,511	2,988,665 bp	0.09%

[a]See more in text: Results, "Annotation of the *C. elaphus* genome"

Table 5 Indels and single nucleotide variations (SNV) in CerEla1.0

Variants	Total variant	INDEL	SNV	
			Nonsynonymous	Synonymous
∑	2,807,458	364,689	17,770	14,252
Downstream	28,973	3264	–	–
Exonic	33,240	758	17,770	14,252
Splicing	–	–	1	2
Intergenic	2,271,127	297,246	–	–
Intronic	438,856	59,273	–	–
Splicing	228	18	–	–
Upstream	26,659	3122	–	–
UTR3	7992	975	–	–
UTR5	1430	151	–	–

bovine pseudogenomes, the deer segments are uniformly 1.25-fold longer than the bovine segments, which also means that the lengths of the deer scaffolds in CerEla1.0 are uniformly 1.25-fold longer than the corresponding bovine genomic segments.

The 25% (0.7 Gbp) virtual excess length of the deer pseudogenome CerEla1.0 (3.4 Gbp) vs. the bovine Btau_5.0.1 (2.7 Gbp) can be accounted for by the too many "NNNs" inserted between the in-scaffold contigs by the ALLPATH-LG program. Provided that the lengths of the *B. taurus* and of the *C. elaphus* genomes are essentially concordant; furthermore, taking into account that in CerEla1.0, the combined DNA sequences of the contigs add up to 1.9 Gbp (Table 3, slot 3), 0.8 Gbp for the total length for the "in-scaffold NNNs" would be more feasible than 1.5 Gbp. However, Fig. 6 also shows that the ALLPATH-LG program distributed this 1.5 Gbp of "NNNs" proportionally with the corresponding physical distances. We believe that in CerEla1.0, the contigs cover 70% of the bovine genome Btau_5.0.1. (i.e., 1.9 Gbp per 2.7 Gbp).

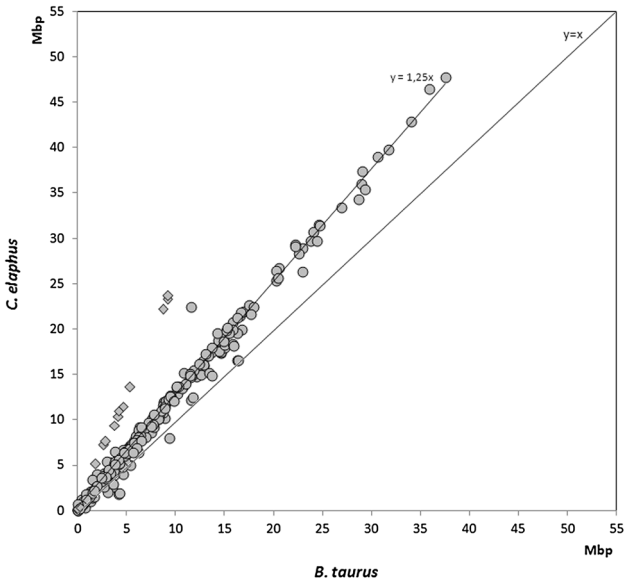


Fig. 6 Genomic segments defined by the *C. elaphus* map markers in CerEla1.0 are compared with the *B. taurus* orthologous segments in Btau_5.0.1. Note, (i) circles: the great majority of pseudogenomic distances are 1.25-fold longer in CerEla1.0 than in Btau_5.0.1 and (ii) squares: distances along the corresponding pseudochromosomes deer 11 and cattle 11 are 2.2-fold longer in deer 11

Genomic segments created by rearrangements

Inversions, translocations, chromosomal fusions, and splits translocate or separate the mapmarkers and make new neighbourhoods around the break points of these rearrangements. In the deer–bovine relation, these "evolution-created" rearrangements resulted in joint sequences with new deer–bovine neighbourhoods. In this CerEla1.0 project, we registered 26 "deer–bovine" rearrangements, including 18 inversions, 2 translocations, and 6 chromosomal fusions/splits that led to new linkages of the orthologous mapmarker sites (Table 6). In the six cases, when one acrocentric bovine chromosome corresponded to two acrocentric deer chromosomes, the last segment of the "front" Ce chromosome and the first region of the "rear" Ce chromosome could not be defined unambiguously from our sequencing data (of scaffolds and contigs) combined with the alignments to the bovine genome. Scaffolds and contigs aligned to the corresponding bovine regions were shared between the deer chromosomes (in CerEla1.0) by considering the recombination distances in the deer genetic map and arranged in "bovine" order. The combined DNA sequences of this type added up to 0.166 Gbp, 5% of the CerEla1.0 genome (Table 6 part A).

Inversions led to 54 deer–bovine sequence transition points at the flanking regions totalling 462.9 Mbp (13.5% of the CerEla1.0) as defined by the neighbouring MMSc-s (Table 6 part B). So far, we have not completed defining the points of the deer–bovine sequence switches. At the present stage in CerEla1.0, these flanking segments were filled up with the scaffolds and contigs exclusively following the bovine order. We believe, however, that future comparisons with other mammalian genomes and with a higher density genetic map of deer will allow more precise identifications.

All in all, in 81.5% of CerEla1.0, the genes were ordered according to the bovine order (between MMSc-s) or deer order (arrays of MMSc-s and of in-scaffold syntenic genes), whereas in 18.5% (at chromosomal splits/fusions, at flanks of inversions), syntenic blocks of deer and bovine genes were combined.

Discussion

In conclusion, in this study, we sequenced the genome of the red deer *Cervus elaphus* with the Illumina New Generation Sequencing technology. Guided by two references, i.e., by the co-linearity of the recombination map of the red deer *Cervus elaphus* and by the bovine reference genome sequence, we successfully assembled the whole genome as CerEla1.0. The sequences in CerEla1.0 were assorted in the pseudochromosome complement of the red deer, one of the most valuable members of the European megafauna, especially in the Carpathian Basin. CerEla1.0 made available deer, a so far non-model animal, for Genome-Wide Association Studies. This study demonstrates the power, both in the annotation of a new mammalian genome and in the pseudochromosome set assembly, of the combination of the genetic map, based on molecular markers, of the target species (the deer) with the existing genome reference of a taxonomically related species (*Bos taurus*)—in the present case, both belonging to ruminants *Artiodactyla* and *Pecora*.

This deer genome (CerEla1.0) was assembled from 26,108 sequence elements (scaffolds and contigs), assorted in 33 auto plus X and Y pseudochromosomes with a total length of 3.4 Gbp (1.9 Gbp without gaps). This means that the scaffolds of the draft deer genome covered virtually the whole *B. taurus* genome, the contig sequences covered 70% of the same, and 90% of the bovine orthologous genes were identified along the deer pseudochromosomes.

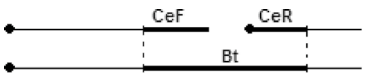
The relationships between the deer and bovine (and also of a number of ovine) chromosomes, i.e., congruencies, fusions, fissions, and inversions, were identified in a "semi-fine" scale (Tables 2, 3; Figs. 2, 3, 4, 5, S1, S2, S3).

In its present state, the gene array in the deer pseudochromosomes is a mosaic of deer and bovine segments. The order of the "mapmarker" scaffolds (MMSc-s) and the intra-scaffold syntenies represent the valid deer arrays.

The length of the CerEla1.0 genome is 3.4 Gbp: due to the somewhat long inserted intercontig NNNs, it is 1.25-fold longer than the bovine genome, 2.7 Gbp in Btau_5.0.1. All

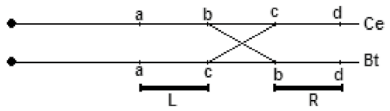
Table 6 Genomic rearrangements in CerEla1.0 in relation with the bovine genome

A Topology:chr fusion/fission



			∑ Mbp shared	Mbps shares in Ce chrs. Mbp	
Bt1	Ce31	Ce19	49.8	Ce31: 49,8	Ce19: <1
Bt2	Ce33	Ce8	31.6	Ce33: 27,8	Ce8 : 3,7
Bt5	Ce3	Ce22	24.8	Ce3 : 21,4	Ce22: 3,4
Bt6	Ce17	Ce6	42	Ce17: 32,0	Ce6 : 10,0
Bt8	Ce29	Ce16	8.9	Ce29: 4,0	Ce16: 4,9
Bt9	Ce28	Ce26	8.9	Ce28: 0,002	Ce26: 8,9
∑			166		

B Flanking segments of inversions



	Flanking segment Mbp	
	L	R
	0.8	4.2
Ce4	1.4	4.6
Ce5	2	5.5
	1.9	14.7
Ce6	4.8	16.3
	7.7	9.2
Ce7	0.9	16.7
Ce8	17.2	1.2
Ce11	22.1	13.6
	14	0.1
Ce12	0.1	7.4
	1.7	11.6
Ce14	7.6	14.4
	6.5	10.1
Ce18	4.7	5.5
Ce20	4.8	1.4
	18.3	9.3
	0.3	6.3
Ce23	1.9	6
Ce24	2.3	5.1
Ce26	13.7	10.2
Ce27	34.2	14.9
	1.9	16.3
Ce28	28	0.002
Ce29	3.1	34.6
Ce32	13.7	4.8
Ce33	2.1	1.2
∑	218	245.2

Part A: tandem fissions/fusions of chromosomes (see Table for details), Part B: inversions. Thickened parts in the simplified chromosome illustrations: the genomic regions, where the sites for the deer–bovine sequence switches occur. CeF and CeR (see pic.A): deer chromosomes tandemly fused in the bovine genome. L and R (see pic.B): the flanking regions of the inversion

along the two genomes, at sub-genomic, sub-chromosomal, and scaffoldic levels, the physical distances are directly correlated with each other in the same ratio of 1.25 (as shown in Fig. 6). The approximately 0.7 Gb (virtual) surplus length for the deer genome can be accounted for by the ALLPATH-LG program characteristics (i.e., insertion of NNN tracts in the scaffolds proportionally between the contigs).

The validity of this deer genome sequence and pseudochromosome complement (CerEla1.0) is supported by:

1. Slate's 365 genetic mapmarkers and the 10 Deerplex STRs were also identified in the bovine genome, and the deer and bovine arrays were co-linear along entire chromosomes as well as in chromosome rearrangements (fusions, fissions and translocations) relative to each other, i.e., between deer and cattle. It is worth noting that no reciprocal translocations were found in deer vs. bovine relation.
2. In all those cases (3422), when two or more genes were carried in a deer scaffold, the syntenies were identical in the deer scaffolds and in the corresponding bovine genome segments.
3. All genes cloned previously (related to antler development and cyclic physiological osteoporosis of deer stag, Molnár et al. 2007; Gyurján et al. 2007; Borsy et al. 2009; Stéger et al. 2010) as well as the STR loci previously developed for multiplex PCR analyses (DeerPlex, Szabolcsi et al. 2014) were recognized in the scaffold/contig sequences and localized in the pseudochromosomes. Both of the two STRs, C01 and T26, which indicated linkage by linkage disequilibrium (LD) test, were localized on *C. elaphus* pseudochromosome 14.
4. Although the *C. elaphus* genome CerEla1.0 was arranged in a mosaic deer–bovine order, the fair correlations for the orthologous deer and bovine genomic regions indicated that the deer scaffolds and contigs covered nearly the whole deer genome. This was supported by the fact that some 90% of the deer orthologs of the bovine protein-coding genes were identified in deer (19,368 for CerEla1.0 vs. 21,427 Btau_5.0.1).

Possible further studies: The reference genome CerEla1.0 of the red deer (*Cervus elaphus hippelaphus*) and its annotation, in accordance with fresh data from other programs, is under continuous monitoring and updating. If the sequence data of the SNP-based map markers will be available (Johnston et al. 2017), the updating of CerEla1.0 will become possible using the approach described in this work. The sequence and the pseudochromosome complement of CerEla1.0 may provide a basis and a rich source for broader interests, including, among others: conservation genetics, refined evolution, and population studies within the family *Cervidae* [e.g., fallow deer (*Dama dama*) or roe deer (*Capreolus capreolus*)] as well as in a wider range of ruminants and *Pecora*. CerEla1.0 also provides a source for developing chromosome-specific microsatellite sets.

A large number of SNP/heterozygotic sites were identified (2.8×10^6 SNVs, 3.6×10^5 indels) and aligned to the deer pseudochromosomes. CerELA1.0 is a leading basis for future genome-wide SNP and microsatellite studies, which may shed light on inbreeding/outbreeding, may help in the identification of gene introgressions and of descents for autosomal, maternal, and paternal lineages. Forensic identification, or definition of allelic compositions underlying phenotypes important, for example, in game, management could also be possible areas of utilization. The exploration of the genetic secret of record antlers becomes possible by genome-wide association studies. Applications and utilizations in several fields of medical research (e.g., bone and osteoporosis research, organ development and regeneration, and robust tissue proliferation/tumour biology) are also feasible.

Acknowledgements This work was supported by the Doctoral School of Animal Science of the Kaposvár University; by the Ministry of Agriculture, Grant No. NAIK–MBK/M71411; by the Ministry of Health, Social and Family Affairs, Grant ETT–ESKI 006/2009; by the Hungarian Academy of Sciences, Grant No. E-127/9/1/2012; by NKFP, Grant 1A/007/2004. We acknowledge NIIF for awarding us access to resources based in Hungary at Pécs, Szeged and Debrecen.

Author contributions NÁB (bioinformatics), KF (molecular genetics), JN (deer breeding) PhD students of Kaposvár University, AN (bioinformatics), TN (bioinformatics), MS (bioinformatics) research assistants of NARIC, VS (molecular genetics and genomics) supervised the laboratory work, PL (medical research, non-model animals), LS (veterinarian zoology) organized the collection of samples, EB supervised bioinformatics, PH animal geneticist, led the Doctoral School of Animal Science of the Kaposvár University, LO designed the study, analysed the data and wrote the paper. All authors contributed to revisions. Results are available online at emboss.abc.hu/wonderdeer/Jbrowse.

Funding This work was supported by the Ministry of Agriculture, Grant No. NAIK–MBK/M71411; by the Ministry of Health, Social and Family Affairs, Grant ETT–ESKI 006/2009; by the Hungarian Academy of Sciences, Grant No. E-127/9/1/2012; and by NKFP, Grant 1A/007/2004.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. Blood was collected from a living animal performed by a trained veterinarian according to the standard veterinary medical practice with a permission from the Hungarian Veterinary Chamber. Sample collection was performed by a trained veterinarian according to standard veterinary medical practice with a permission from the Hungarian Veterinary Chamber [Hungarian Animal Rights Law (243/1998, XII.31)].

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arias J, Keehan M, Fisher P, Coppieters W, Spelman R (2009) A high density linkage map of the bovine genome. BMC Genet 10(1):18. https://doi.org/10.1186/1471-2156-10-18

Ba H, Wang D, Li C (2016) MicroRNA profiling of antler stem cells in potentiated and dormant states and their potential roles in antler regeneration. Mol Genet Genom 291:943–955. https://doi.org/10.1007/s00438-015-1158-8

Balla B, Kósa J, Kiss J, Borsy A, Podani J, Takács I, Lazáry Á, Nagy Z, Bácsi K, Speer G, Orosz L, Lakatos P (2008) Different gene expression patterns in the bone tissue of aging postmenopausal osteoporotic women. Calcif Tissue Int 82(1):12–26. https://doi.org/10.1007/s00223-007-9092-3

Bán I (1998) The Hungarian wonder deer. EP Systema, Debrecen **(ISBN:963-03-4955-8)**

Barta E, Bánfalvi Z, Havelda Z, Hiripi L, Jeney Z, Kiss J, Kolics B, Marincs F, Silhavy D, Stéger V, Várallyay É (2016) Agricultural genomics: an overview of the next generation sequencing projects at the NARIC-Agricultural Biotechnology Institute in Gödöllő. Hung Agric Res 25:10–21

Biedrzycka A, Solarz W, Okarma H (2012) Hybridization between native and introduced species of deer in Eastern Europe. J Mammal 93(5):1331–1341. https://doi.org/10.1644/11-MAMM-A-022.1

Bonnet A, Thévenon S, Claro F, Gautier M, Hayes H (2001) Cytogenetic comparison between Vietnamese sika deer and cattle: R-banded karyotypes and FISH mapping. Chromosome Res 9:673–687. https://doi.org/10.1023/A:1012908508488

Borsy A, Podani J, Stéger V, Balla B, Horváth A, Kósa J, Gyurján I, Molnár A, Szabolcsi Z, Szabó L, Jakó E, Zomborszky Z, Nagy J, Semsey S, Vellai T, Lakatos P, Orosz L (2009) Identifying novel genes involved in both deer physiological and human pathological osteoporosis. Mol Genet Genom 281:301–313. https://doi.org/10.1007/s00438-008-0413-7

Brauning R, Fisher PJ, McCulloch AF, Smithies RJ, Ward JF, Bixley MJ, Lawley CT, Rowe SJ, McEwan JC (2015) Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk. bioRxiv. https://doi.org/10.1101/027318

Campbell MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinform 48:4.11.1–4.11.39. https://doi.org/10.1002/0471250953.bi0411s48

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188–196. https://doi.org/10.1101/gr.6743907

Everts-van der Wind A, Kata SR, Band MR, Rebeiz M, Larkin DM, Everts RE, Green CA, Liu L, Natarjan S, Goldammer T, Lee JH, McKay S, Womack JE, Lewin HA (2004) A 1463 gene cattle–human comparative map with anchor points defined by human genome sequence coordinates. Genome Res 14(7):1424–1437. https://doi.org/10.1101/gr.2554404

Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, Chiu R, Schein JE, Marra MA, Womack JE, Lewin HA (2005) A high-resolution whole-genome cattle–human comparative map reveals details of mammalian chromosome evolution. Proc Natl Acad Sci USA 102(51):18526–18531. https://doi.org/10.1073/pnas.0509285102

Fisher PJ, Brauning R, McCulloch AF, Smithies RJ, Ward JF, Bixley MJ, Lawley CT, Everett-Hincks JM, McEwan JC (2015) Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk. bioRxiv. https://doi.org/10.1101/027318

Fontana F, Rubini M (1990) Chromosomal evolution in *Cervidae*. Biosystems 24(2):157–174. https://doi.org/10.1016/0303-2647(90)90008-O

Frank K, Barta E, Bana N, Nagy J, Horn P, Orosz L, Stéger V (2016) Complete mitochondrial genome sequence of a Hungarian red deer (*Cervus elaphus hippelaphus*) from high-throughput sequencing data and its phylogenetic position within the family *Cervidae*. Acta Biol Hung 67:133–147. https://doi.org/10.1556/018.67.2016.2.2

Frank K, Bleier N, Tóth B, Sugár L, Horn P, Barta E, Orosz L, Stéger V (2017) The presence of Balkan and Iberian red deer (*Cervus elaphus*) mitochondrial DNA lineages in the Carpathian Basin. Mammal Biol 86:48–55. https://doi.org/10.1016/j.mambio.2017.04.005

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci USA 108(4):1513–1518. https://doi.org/10.1073/pnas.1017351108

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

Gu L, Mo E, Zhu X, Jia X, Fang Z, Sun B et al (2008) Analysis of gene expression in four parts of the red deer antler using DNA chip microarray technology. Anim Biol 58:67–90. https://doi.org/10.1163/157075608X303654

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. https://doi.org/10.1111/j.1601-5223.1968.tb02204.x

Gyurján I, Molnár A, Borsy A, Stéger V, Hackler L, Zomborszky Z, Papp P, Duda E, Deák F, Lakatos P, Puskás LG, Orosz L (2007) Gene expression dynamics in deer antler: mesenchymal differentiation toward chondrogenesis. Mol Genet Genom 277:221–235. https://doi.org/10.1007/s00438-006-0190-0

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between linked factors. J Genet 8:299–309

Harris RS (2007) Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University

Jia BY, Ba HX, Wang GW, Yang Y, Cui XZ, Peng YH, Zheng JJ, Xing XM, Yang FH (2016) Transcriptome analysis of sika deer in China. Mol Genet Genom 291:1941–1953. https://doi.org/10.1007/s00438-016-1231-y

Johnston SE, Huisman J, Ellis PA, Pemberton JM (2017) Sexual dimorphism in recombination landscape in red deer (*Cervus elaphus*): a role for meiotic drive? bioRxiv. https://doi.org/10.1101/100131

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kierdorf U, Li C, Price JS (2009) Improbable appendages: Deer antler renewal as a unique case of mammalian regeneration. Sem Cell Dev Biol 20:535–542. https://doi.org/10.1016/j.semcdb.2008.11.011

Korf I (2004) Gene finding in novel genomes. BMC Bioinform 5:59. https://doi.org/10.1186/1471-2105-5-59

Korf I, Yandell M, Bedell J (2003) An essential guide to the basic local alignment search tool: BLAST. O'Reilly & Associates, Inc., Sebastopol

Kosambi DD (1943) The estimation of map distance from recombination values. Ann Eugen 12:172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73. https://doi.org/10.1093/nar/gkt1181

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biol 5(2):R12. https://doi.org/10.1186/gb-2004-5-2-r12

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li C, Suttie JM (2001) Deer antlerogenic periosteum: a piece of postnatally retained embryonic tissue? Anat Embryol 204:375–388. https://doi.org/10.1007/s004290100204

Li C, Suttie JM (2012) Morphogenetic aspects of deer antler development. Front Biosci 4:1836–1842. https://doi.org/10.2741/e505

Li C, Clark DE, Lord EA, Stanton JA, Suttie JM (2002) Sampling technique to discriminate the different tissue layers of growing antler tips for gene discovery. Anat Rec 268:125–130. https://doi.org/10.1002/ar.10120

Li C, Stanton JL, Robertson TM, Suttie JM, Sheard PW, Harris JA, Clark DE (2007) Nerve growth factor mRNA expression in the regenerating antler tip of Red Deer (*Cervus elaphus*). PLoS One 10(2):e148. https://doi.org/10.1371/journal.pone.0000148

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Ma RZ, Beever JE, Da Y, Green CA, Russ I, Park C, Heyen DW, Everts RE, Fisher SR, Overton KM, Teale AJ, Kemp SJ, Hines HC, Guérin G, Lewin HA (1996) A male linkage map of the cattle (*Bos taurus*) genome. J Hered 4:261–271. https://doi.org/10.1093/oxfordjournals.jhered.a022999

McDevitt AD, Edwards CJ, O'Toole P, O'sullivan P, O'Reilly C, Carden RF (2009) Genetic structure of, and hybridisation between, red (*Cervus elaphus*) and sika (*Cervus nippon*) deer in Ireland. Mammal Biol 74:263–273. https://doi.org/10.1016/j.mambio.2009.03.015

Molnár A, Gyurján I Jr, Korpos E, Borsy A, Stéger V, Buzás Z, Kiss I, Zomborszky Z, Papp P, Deák F, Orosz L (2007) Identification of differentially expressed genes in the developing antler of red deer *Cervus elaphus*. Mol Genet Genom 277:237–248. https://doi.org/10.1007/s00438-006-0193-x

Molnár J, Nagy T, Stéger V, Tóth G, Marincs F, Barta E (2014) Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. BMC Genom 15:761. https://doi.org/10.1186/1471-2164-15-761

Park HJ, Lee DH, Park SG, Lee SC, Cho S, Kim HK et al (2004) Proteome analysis of red deer antlers. Proteomics 11:3642–3653. https://doi.org/10.1002/pmic.200401027

Price J, Allen S (2004) Exploring the mechanisms regulating regeneration of deer antlers. Philos Trans R Soc Lond Biol 359:809–822. https://doi.org/10.1098/rstb.2004.1471

Price J, Allen S, Faucheux C, Althanaian T, Mount J (2005) Deer antlers: a zoological curiosity or the key to understanding organ regeneration in mammals? J Anat 207:603–618. https://doi.org/10.1111/j.1469-7580.2005.00478.x

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16(6):276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–W689. https://doi.org/10.1093/nar/gki366

Slate J, Van Stijn TC, Anderson RM, McEwan KM, Maqbool NJ, Mathias HC, Bixley MJ, Stevens DR, Molenaar AJ, Beever JE, Galloway SM, Tate ML (2002a) A deer (subfamily *Cervinae*) genetic linkage map and the evolution of ruminant genomes. Genetics 160:1587–1597

Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML, Pemberton JM (2002b) A genome scan for quantitative trait loci in a wild population of red deer (*Cervus elaphus*). Genetics 162:1863–1873

Smit AFA, Hubley R, Green P (2013–2015) RepeatMasker Open-4.0. http://www.repeatmasker.org/. Accessed Jan 2016

Stanke M, Tzvetkova A, Morgenstern B (2006) AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. Genome Biol 7:S11–S18. https://doi.org/10.1186/gb-2006-7-s1-s11

Stéger V, Molnár A, Borsy A, Gyurján I, Szabolcsi Z, Dancs G, Molnár J, Papp P, Nagy J, Puskás L, Barta E, Zomborszky Z, Horn P, Podani J, Semsey S, Lakatos P, Orosz L (2010) Antler development and coupled osteoporosis in the skeleton of red deer *Cervus elaphus*: expression dynamics for regulatory and effector genes. Mol Genet Genom 284:273–287. https://doi.org/10.1007/s00438-010-0565-0

Szabolcsi Z, Egyed B, Zenke P, Padar Z, Borsy A, Steger V, Pasztor E, Csanyi S, Buzas Z, Orosz L (2014) Constructing STR multiplexes for individual identification of Hungarian red deer. J Forensic Sci 59:1090–1099. https://doi.org/10.1111/1556-4029.12403

Tate ML, Mathias HC, Fennessy PF, Dodds KG, Penty JM, Hill DF (1995) A new gene mapping resource: interspecies hybrids between Père David's deer (*Elaphurus davidianus*) and red deer (*Cervus elaphus*). Genetics 139:1383–1391

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Res 38:e164. https://doi.org/10.1093/nar/gkq603

Yao B, Zhao Y, Zhang H, Zhang M, Liu M, Liu H, Li J (2012) Sequencing and de novo analysis of the Chinese Sika deer antler-tip transcriptome during the ossification stage using Illumina RNA-Seq technology. Biotechnol Lett 34:813–822. https://doi.org/10.1007/s10529-011-0841-z

Affiliations

Nóra Á. Bana[1,2] · Anna Nyiri[1] · János Nagy[2] · Krisztián Frank[1,2] · Tibor Nagy[1] · Viktor Stéger[1] · Mátyás Schiller[1] · Péter Lakatos[4] · László Sugár[2] · Péter Horn[2] · Endre Barta[1,5] · László Orosz[1,3]

Nóra Á. Bana
bana.agnes.nora@abc.naik.hu

Anna Nyiri
nyiri.anna.emese@abc.naik.hu

János Nagy
nagy.janos@ke.hu

Krisztián Frank
frank.krisztian@abc.naik.hu

Tibor Nagy
black007@abc.hu

Viktor Stéger
steger.viktor@abc.naik.hu

Mátyás Schiller
schiller.matyas@abc.naik.hu

Péter Lakatos
lakatos.peter@med.semmelweis-univ.hu

László Sugár
sugar.laszlo@ke.hu

Péter Horn
horn.peter@ke.hu

Endre Barta
barta.endre@abc.naik.hu
http://abc.naik.hu/en/organizational-units/abc-department-of-genomics-agricultural-genomics-and-bioinformatics#link_3
http://emboss.abc.hu/wonderdeerhomepages

[1] Agricultural Biotechnology Institute, National Agricultural Research and Innovation Center, Szent-Györgyi Albert str. 4, Gödöllő 2100, Hungary

[2] Department of Animal Breeding Technology and Management Faculty of Agricultural and Environmental Sciences, Kaposvár University, Guba Sándor str. 40, Kaposvár 7400, Hungary

[3] Department of Genetics, Faculty of Sciences, Eötvös Loránd University, Pázmány Péter ave. 1/c, Budapest 1117, Hungary

[4] 1st Department of Internal Medicine, Semmelweis University, Korányi Sándor str. 2/a, Budapest 1083, Hungary

[5] Department of Biochemistry and Molecular Biology, University of Debrecen, Nagyerdei ave 98, Debrecen 4032, Hungary