

ARTICLE OPEN



Germline-restricted chromosome of songbirds has different centromere compared to regular chromosomes

Jakub Rídl^{1,2}✉, Dmitrij Dedukh³, Zuzana Halenková^{1,2}, Stephen A. Schlebusch^{1,2}, Vladimír Beneš⁴, Mireia Osuna Lopez⁴, Tomasz S. Osiejuk⁵, Francisco J. Ruiz-Ruano^{6,7}, Alexander Suh^{6,7,8}, Tomáš Albrecht^{1,9}, Jiří Reif^{10,11} and Radka Reifová¹✉

© The Author(s) 2025

Centromeres are an important part of chromosomes which direct chromosome segregation during cell division. Their modifications can therefore explain the unusual mitotic and meiotic behaviour of certain chromosomes, such as the germline-restricted chromosome (GRC) of songbirds. This chromosome is eliminated from somatic cells during early embryogenesis and later also from male germ cells during spermatogenesis. Although the mechanism of elimination is not yet known, it is possible that it involves a modification of the centromeric sequence on the GRC, resulting in problems with the attachment of this chromosome to the mitotic or meiotic spindle and its lagging during anaphase, which eventually leads to its elimination from the nucleus. However, the repetitive nature and rapid evolution of centromeres make their identification and comparative analysis across species and chromosomes challenging. Here, we used a combination of cytogenetic and genomic approaches to identify the centromeric sequences of two closely related songbird species, the common nightingale (*Luscinia megarhynchos*) and the thrush nightingale (*L. luscinia*). We found a 436-bp satellite repeat present in the centromeric regions of all regular chromosomes (i.e., autosomes and sex chromosomes), making it a strong candidate for the centromeric repeat. This centromeric repeat was highly similar between the two nightingale species. Interestingly, hybridization of the probe to this satellite repeat on meiotic spreads suggested that this repeat is missing on the GRC. Our results indicate that the change of the centromeric sequence may underlie the unusual inheritance and programmed DNA elimination of the GRC in songbirds.

Heredity; <https://doi.org/10.1038/s41437-025-00779-5>

INTRODUCTION

Centromeres are an essential component of chromosomes. By recruiting the kinetochore—a protein structure that mediates the attachment of the chromosomes to the microtubule spindle—they control chromosome segregation during cell division (McKinley and Cheeseman 2016). Defects in centromere function can lead to the loss or amplification of individual chromosomes in the dividing cells, with potentially deleterious consequences for the organisms (Holland and Cleveland 2009). However, alterations in the centromere have also been implicated in the unusual non-Mendelian inheritance (Clark and Aker 2021; Chmátal et al. 2017; Henikoff et al. 2001; Rosin and Mellone 2017) or programmed DNA elimination (Dedukh and Krasikova 2022) of certain chromosomes, including the parasitic B chromosomes (Johnson Pokorná and Reifová 2021).

Although the centromeric function is essential and well conserved, the centromeric DNA sequences, which are typically composed of satellite tandem repeats, represent one of the most rapidly evolving sequences in eukaryotic genomes (Henikoff et al.

2001; Poignet et al. 2021; Voleníková et al. 2023). This “centromeric paradox” can be explained by the frequent female meiotic drive occurring at centromeric loci, where centromeres with the ability to preferentially move the chromosome to the oocyte or ovum rather than polar bodies can spread rapidly in the population and cause rapid turnover of centromeric sequences between species (Kursel and Malik 2018). In addition, centromeric sequences can evolve without much constraint, as epigenetic mechanisms leading to the loading of the centromere-specific histone H3 variant CENP-A, rather than the sequence itself, typically determine whether DNA acquires centromeric function (McKinley and Cheeseman 2016).

The rapid evolution of centromeric sequences, combined with their repetitive nature, makes them notoriously difficult to study. The position of the centromere within chromosomes can be inferred from cytogenetic data by observing primary constriction on metaphase chromosomes or by using various centromeric markers (Poignet et al. 2021; Talbert and Henikoff 2020). Furthermore, as genomic regions surrounding the centromere

¹Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic. ²Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics, Czech Academy of Sciences, Prague, Czech Republic. ³Laboratory of Non-Mendelian Evolution, Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Liběchov, Czech Republic. ⁴Genomics Core Facility, European Molecular Biology Laboratory, Heidelberg, Germany. ⁵Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology Adam Mickiewicz University, Poznań, Poland. ⁶Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany. ⁷Bonn Institute for Organismal Biology—Animal Biodiversity, University of Bonn, Bonn, Germany. ⁸Department of Organismal Biology—Systematic Biology (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ⁹Institute of Vertebrate Biology, Czech Academy of Sciences, Studenec, Czech Republic. ¹⁰Institute for Environmental Studies, Faculty of Science, Charles University, Prague, Czech Republic. ¹¹Department of Zoology, Faculty of Science, Palacký University, Olomouc, Czech Republic. Associate editor: Stacey Hanlon. ✉email: jakub.ridl@img.cas.cz; radka.reifova@natur.cuni.cz

Received: 6 November 2024 Revised: 26 June 2025 Accepted: 26 June 2025

Published online: 30 July 2025

show reduced nucleotide diversity and increased linkage disequilibrium due to suppressed recombination, the centromeric position can be estimated from population genomic data (Begun and Aquadro 1992; Begun et al. 2007; Weissensteiner et al. 2017; Weighill et al. 2019). However, centromeric repeats themselves are often missing from genome assemblies and their identification requires a more detailed analysis of the repetitive genome content. As a result, centromeres have only been characterized in a limited number of species (e.g., Yamada et al. 2002; Yi et al. 2015; Khost et al. 2017; Weissensteiner et al. 2017; Uno et al. 2019; Blommaert et al. 2024; Chang et al. 2024). Only recently has the more detailed analysis of centromere repeat array structure been comprehensively possible with (nearly) gap-free genome assemblies for example in humans (Altemose et al. 2022; Gao et al. 2024; Logsdon et al. 2024) or chicken (Huang et al. 2023).

In songbirds, the most species-rich group of birds comprising nearly half of all bird species, the candidate centromeric repeats have only been described in few species, the chaffinch (*Fringilla coelebs*) (Saifuddinova et al. 2001), the hooded crow (*Corvus corone cornix*) (Weissensteiner et al. 2017), the zebra finch (*Taeniopygia guttata*) (Takki et al. 2022) and birds-of-paradise (Paradisaeidae species) (Peona et al. 2023). Songbirds have relatively conserved karyotypes consisting of a similar number of chromosomes across species (Kapusta and Suh 2017). In addition to the standard set of chromosomes, songbirds contain the so-called germline-restricted chromosome (GRC) (Kinsella et al. 2019; Pigozzi and Solari 1998; Torgasheva et al. 2019). The GRC shows unusual mitotic as well as meiotic inheritance (Borodin et al. 2022). It is eliminated from somatic cells during early embryogenesis and from male germ cells during spermatogenesis (Pigozzi and Solari 2005, but see Pei et al. 2022). Although it is normally present as a single copy in males and two copies in females, polymorphism in copy number has been observed between individuals (Borodin et al. 2022; Malinovskaya et al. 2020; Torgasheva et al. 2021) and sometimes even between cells within an individual (Sotelo-Muñoz et al. 2022). Although the mechanisms behind this unusual inheritance are still unknown, it is possible that modification of the GRC centromeric sequence may be involved.

Here we used a multi-evidence approach to investigate the position and repetitive content of the centromeres in two closely related songbirds, the common nightingale (*Luscinia megarhynchos*) and the thrush nightingale (*L. luscinia*). The two species diverged approximately 1.8 Mya (Storchová et al. 2010) and still occasionally hybridize in the sympatric zone (Sottas et al. 2023). Recently, a high-quality chromosome-level genome assembly has been published for both species (Schlebusch et al. 2023). In addition, almost the whole GRC has been assembled for both nightingale species, although the quality of the assembly is lower compared to regular chromosomes (Schlebusch et al. 2023). First, we estimated the centromere position on individual chromosomes based on patterns of nucleotide diversity and linkage disequilibrium along the genome as well as previously published cytogenetic data (Poignet et al. 2021). Second, we searched for repetitive sequences corresponding in position to the estimated centromere positions from genomic and cytogenetic data. Finally, we generated probes for fluorescence in situ hybridization (FISH) against the candidate repeats to cytogenetically verify that they are localized within the centromeric region.

We found a 436-bp satellite repeat present in the centromeric region of all regular chromosomes (i.e., autosomes and sex chromosomes) making it a good candidate for the centromeric sequence in nightingales. The cytogenetic and genomic data, however, suggest that the repeat is either missing or fragmented and degenerated on the GRC. Our results indicate that the GRC may have a different centromere than the regular chromosomes, which could explain its unusual mitotic and meiotic behavior.

MATERIALS AND METHODS

Animals and sample collection

Four unrelated individuals (two females and two males) from each nightingale species were captured using mist nets and/or collapsible traps in allopatric populations to avoid sampling of interspecific hybrids. The common nightingales (*L. megarhynchos*) were captured in South-Western Poland by the Odra River, between the towns Wrocław (N 51.203°, E 16.965°) and Lubiąż (N 51.265°, E 16.470°). The thrush nightingales (*L. luscinia*) were captured in the vicinity of Łomża city (between N 52.970°, E 21.899° and N 53.134°, E 22.416°) in North-Eastern Poland. The birds were sampled in 2018 at the beginning of the breeding season between May 6 and May 15. Birds were euthanized by cervical dislocation. From female individuals, heart samples, representing the somatic tissue without the GRC, were dissected, flash-frozen in liquid nitrogen and stored at −80 °C for further DNA isolation and whole genome sequencing. These data complemented previously obtained genomic data from another three individuals of each species (see below). From male individuals, the testis (containing meiotically dividing cells with the GRC) and the tibia (containing bone marrow with a high proportion of mitotically dividing cells) were dissected for preparation of meiotic and mitotic spreads, respectively. The work was carried out in accordance with ethical animal research requirements of Poland according to Polish law (the Act On the Protection of Animals Used for Scientific or Educational Purposes, 15.01.2015, item 266, implementing Directive 2010/63/EU of the European Parliament and of the European Council of 22.09.2010). Experiments on birds were approved by the General Directorate for Environmental Protection (permission no. DZP-WG.6401.03.123.2017.dl.3).

Whole genome sequencing

DNA from heart tissue from each sampled female individual was isolated using MagAttract HMW DNA Kit (Qiagen) according to the manufacturer's instructions. The DNA samples were sent to GeneCore (EMBL, Heidelberg) for standard Illumina paired-end sequencing using the HiSeq 2500.

In addition, we utilized previously published (Schlebusch et al. 2023) standard Illumina sequencing data from kidneys from two male individuals of each species (BioProject accession code PRJNA808609) and chromosome-level somatic genome assemblies together with respective Illumina data from one female individual of each species (BioProject accession codes PRJNA810511 and PRJNA810515).

Estimates of nucleotide diversity and linkage disequilibrium along chromosomes

All the sequencing reads were quality trimmed using Trimmomatic v0.39 (Bolger et al. 2014). The trimmed Illumina reads from two female (this study) and two male (Schlebusch et al. 2023) individuals of each species were aligned to their respective reference genome (accessions GCA_034336665.1 and GCA_034336685.1) (Schlebusch et al. 2023) using bwa v0.7.17 (Li 2013). Samtools v1.14 (Danecek et al. 2021) was used to remove duplicate alignments with the fixmate and markdup functions. BCFtools's mpileup and call commands (Danecek et al. 2021) were used for single nucleotide polymorphism (SNP) variant calling. SNPs with a quality score greater than 20 were selected using the varFilter command from VCFtools (Danecek et al. 2011). Nucleotide diversity (π) and linkage disequilibrium (LD) were calculated using VCFtools and visualized in 50 kb long windows along individual chromosomes using the R (R Core Team 2023) package ggplot2 (Wickham 2016). Specifically, we used VCFtools option --windowed-pi to calculate π , and --geno-r2 option (with filtering options --thin 500 --ld-window-bp 10,000) to calculate LD.

Identification of candidate centromeric repeats

The Illumina reads from one female individual from each species (Schlebusch et al. 2023) were used for the identification of repetitive elements using RepeatExplorer2 (Novák et al. 2010, 2013), including the TAREAN pipeline (Novák et al. 2017) in the Galaxy environment, both according to the author's guidelines (Novák et al. 2020). From each species, 200,000 paired-end reads were randomly selected using seqtk v1.3-r106 (<https://github.com/lh3/seqtk>) and used separately for each species with RepeatExplorer2 with the cluster merging option. Consensus sequences of each resulting putative tandem repeat were duplicated to create dimers before being used with RepeatMasker v4.1.1 (Smit et al. 2015) to identify their location and abundance in each reference genome. The cumulative length of each satellite within 50 kb windows was calculated in R (R Core Team 2023) and plotted using ggplot2 (Wickham 2016).

BLASTN (Camacho et al. 2009) was used to calculate the sequence similarity between the consensus sequences of the candidate centromeric satellite found in both species (see “Results”). To search for a possible occurrence of the candidate centromeric satellite in the published assembly of the GRC in both nightingale species (Schlebusch et al. 2023), we used BLASTN search against the respective GRC sequences (NCBI accessions GCA_030412455.1 and GCA_030412475.1). To identify homologs in other species, we used BLASTN search of the common nightingale candidate centromeric satellite against NCBI database and also against previously identified repetitive elements in the collared flycatcher (*Ficedula albicollis*) (Suh et al. 2018).

The identified candidate centromeric satellites were inspected for the presence of the 17-bp CENP-B box motif by aligning (A/C) TTCGTTGGAACGGGA, the proposed mammalian CENP-B box consensus sequence (Masumoto et al. 1989), using MAFFT v7.511 (Katoh et al. 2019).

Preparation of mitotic and meiotic chromosome spreads

Mitotic metaphase chromosomes were prepared from bone marrow as has been described in Poignet et al. (2021). Briefly, after washing out bone marrow from the tibias of each bird, cells were cultivated in 5 ml of D-MEM medium (Sigma Aldrich), with the addition of 75 µl of colcemid (Roche) solution for 40 min at 37 °C. Afterward, D-MEM medium (Sigma Aldrich) was replaced by pre-warmed hypotonic 0.075 M KCl solution for 25 min at 37 °C and then cells were fixed in 3:1 (methanol: acetic acid) fixative solution. After washing three times in a fixative solution (methanol:acetic acid, 3:1), tissues were stored at −20 °C until their use. To prepare chromosomal spreads, the cell suspension was dropped on a slide and the fixative solution was allowed to evaporate. The remaining chromosomal spreads were stained with Giemsa (5% Giemsa in 0.07 M phosphate buffer, pH 7.4).

Meiotic spreads of pachytene chromosomes were prepared from the testes of reproductively active males following Peters et al. (1997). Testes were placed in hypotonic solution (30 mM Tris, 50 mM sucrose, 17 mM trisodium citrate dihydrate, and 5 mM EDTA; pH 8.2) and cut to release cells. After 50 min of incubation in a hypotonic solution, testis tissue fragments were transferred to 200 µl of 100 mM sucrose and disaggregated. The resulting cell suspension (40 µl) was dropped onto a slide previously treated with 1% PFA and 0.15% of Triton X100 (Sigma Aldrich), and spread by slide inclination. Afterward, slides were placed in a humid chamber for 90 min, washed for 2 min in 1× PBS, and used for immunostaining. Pachytene chromosomes were immunostained with rabbit polyclonal anti-SYCP3 antibody (ab15093, Abcam), labeling the lateral elements of the synaptonemal complex (dilution 1:200) combined with human anticentromere serum (CREST, 15-234, Antibodies Incorporated) binding to centromere-specific proteins CENP-A, CENP-B, and CENP-C (dilution 1:50) (Mchugh 2007) according to the protocol described in Moens (2006). We used the following secondary antibodies: Alexa-594-conjugated goat anti-Rabbit IgG (H+L) (A32740, Invitrogen; dilution 1:200) and Alexa-488-conjugated goat anti-Human IgG (H+L) (A-11013, Invitrogen; dilution 1:200). Primary and secondary antibodies were diluted in PBT (3% BSA and 0.05% Tween 20 in 1× PBS) and incubated in a humid chamber for 90 min each. Slides were washed 3 times in 1× PBS and dehydrated through an ethanol series (50%, 70%, and 96%, 3 min each). Finally, all slides were dried and stained with DAPI in mounting medium Vectashield (Vector Laboratories, Burlingame, CA, USA, cat. number H-1200-10).

Fluorescence in situ hybridization (FISH) of candidate centromeric repeats

We used FISH to cytogenetically verify that the candidate centromeric repeat (see “Results”) is located in the centromeric regions of chromosomes. We used two kinds of probes against its sequence. First, we used probes synthesized and labeled by PCR using the following primers (annealing temperature 61 °C):

L. luscinia:

Forward 5'-GCTGTGCACACTTTCGCTTT-3',

Reverse 5'-CCGTCTACCCCTCTCAGCA-3';

L. megarhynchos:

Forward 5'-GAAGTGCCGTCTACCCCTCT-3',

Reverse 5'-GTGAGATCTGTGGTGTGCTGT-3'.

PCR probes were labeled with biotin-16-dUTP (Roche, Mannheim, Germany) or digoxigenin-11-dUTP (Roche, Mannheim, Germany). The primers were designed inside of the satellite monomer with the estimated PCR product length of 221 bp in *L. megarhynchos* and 199 bp in *L. luscinia*. However, since the satellite is organized in tandem repeats, the outcome

of the PCR reaction is a mixture of the probes of different lengths (657 bp and 635 bp in *L. megarhynchos* and *L. luscinia*, respectively, if a satellite dimer is amplified, 1093 bp and 1071 bp if a trimer is amplified and so on). The obtained mixture of probes should thus cover the whole length of the satellite.

Second, we used the following commercially synthesized oligoprobes. The 20-bp probe for *L. luscinia* (5'-GCTGTGCACACTTTCGCTTT-3) was labeled with biotin-16-dUTP, and the 22-bp probe for *L. megarhynchos* (5'-GTTGTAGCAAAGCTGTTCTGG-3') was labeled with digoxigenin-11-dUTP (see Supplementary Fig. 2 for the positions of PCR primers and oligoprobes in the satellite motif).

For FISH, chromosome slides were pretreated with 0.01% pepsin in 0.01 M HCl for 10 min, fixed with 2% paraformaldehyde for 10 min, washed in 1× PBS, dehydrated in an ethanol series (50%, 70% and 96%, 3 min each) and air-dried. In the case of pachytene chromosomal spreads, the pretreatment steps were omitted. The hybridization mixture for PCR probes contained 50% formamide, 10% dextran sulfate, 2× SSC (saline-sodium citrate buffer), 200 ng (per slide) of either *L. luscinia* or *L. megarhynchos* probe, and 10-fold excess of salmon sperm DNA (Sigma-Aldrich) to the amount of the probe. For two-colored FISH, we added 100 ng of *L. luscinia* probe labeled with biotin and 100 ng of *L. megarhynchos* probe labeled with digoxigenin to the hybridization mixture which was otherwise the same as previously described. In the case of oligoprobes, the hybridization mixture included 40% formamide, 10% dextran sulfate, 2× SSC, 200 ng of oligoprobe, and a 10-fold excess of salmon sperm DNA (Sigma-Aldrich) to the amount of the probe. Both PCR and oligoprobes were denatured at 86 °C for 6 min and placed on ice for 10 min. In the case of two-colored FISH, the denatured probes were also incubated at 37 °C for 30 min. Simultaneously, chromosomal DNA on slides with mitotic chromosomes and pachytene chromosomes was denatured at 76 °C for 3 min, dehydrated in a series of ice-cold ethanol washes (50%, 70%, and 96%, 3 min each), and air dried. Afterward, a hybridization mixture with denatured probe was dropped on slides, covered by cover slides, mounted with rubber cement, and incubated overnight at either 37 °C in the case of PCR probes or at room temperature in the case of oligoprobes. After hybridization overnight, slides with PCR probes were washed thrice in 0.2× SSC for 5 min at 50 °C; slides with oligoprobes were washed in 2× SSC for 5 min at 46 °C. Detection of probes was performed with streptavidin conjugated with Alexa 488 (S11223, Invitrogen, San Diego, CA, United States); digoxigenin was detected by anti-digoxigenin antibodies conjugated with rhodamine (11207750910, Merck, Darmstadt, Germany). After incubation for 3 h, slides were washed in 4× SSC for 5 min at 45 °C, dehydrated in an ethanol series (50%, 70%, and 96%, 3 min each), and mounted with Vectashield/DAPI (1.5 mg/mL) antifade medium (Vector Laboratories, Burlingame, CA, USA, cat. number H-1200-10).

Since we did not observe any difference between the mapping of PCR probes and oligoprobes, only results from PCR probes are shown in the “Results” section.

Image processing

Chromosomal slides were examined using an Olympus Provis AX 70 epifluorescence microscope and a ZEISS Axio Imager.Z2 epifluorescence microscope. Images of metaphase chromosomes were captured with an Olympus DP30BW CCD camera and a CoolCube 1 camera (MetaSystems, Altlusheim, Germany). The images were adjusted in Adobe Photoshop software, version CS5, and analyzed using the IKAROS and ISIS imaging software (Metasystems, Altlusheim, Germany).

RESULTS

Estimation of centromere position within individual chromosomes

To estimate the position of the centromere on individual chromosomes, we first utilized whole genome sequence data from four individuals per species and calculated nucleotide diversity (π) and linkage disequilibrium (LD) along individual chromosomes in 50 kb windows. Figure 1 shows the distribution of π and LD along the first four largest chromosomes (data for other chromosomes are shown in Supplementary Fig. 1). In most chromosomes we found a single prominent region with reduced π values and at the same time increased LD compared to the rest of the chromosome, representing a candidate centromeric region (Fig. 1 and Supplementary Fig. 1).

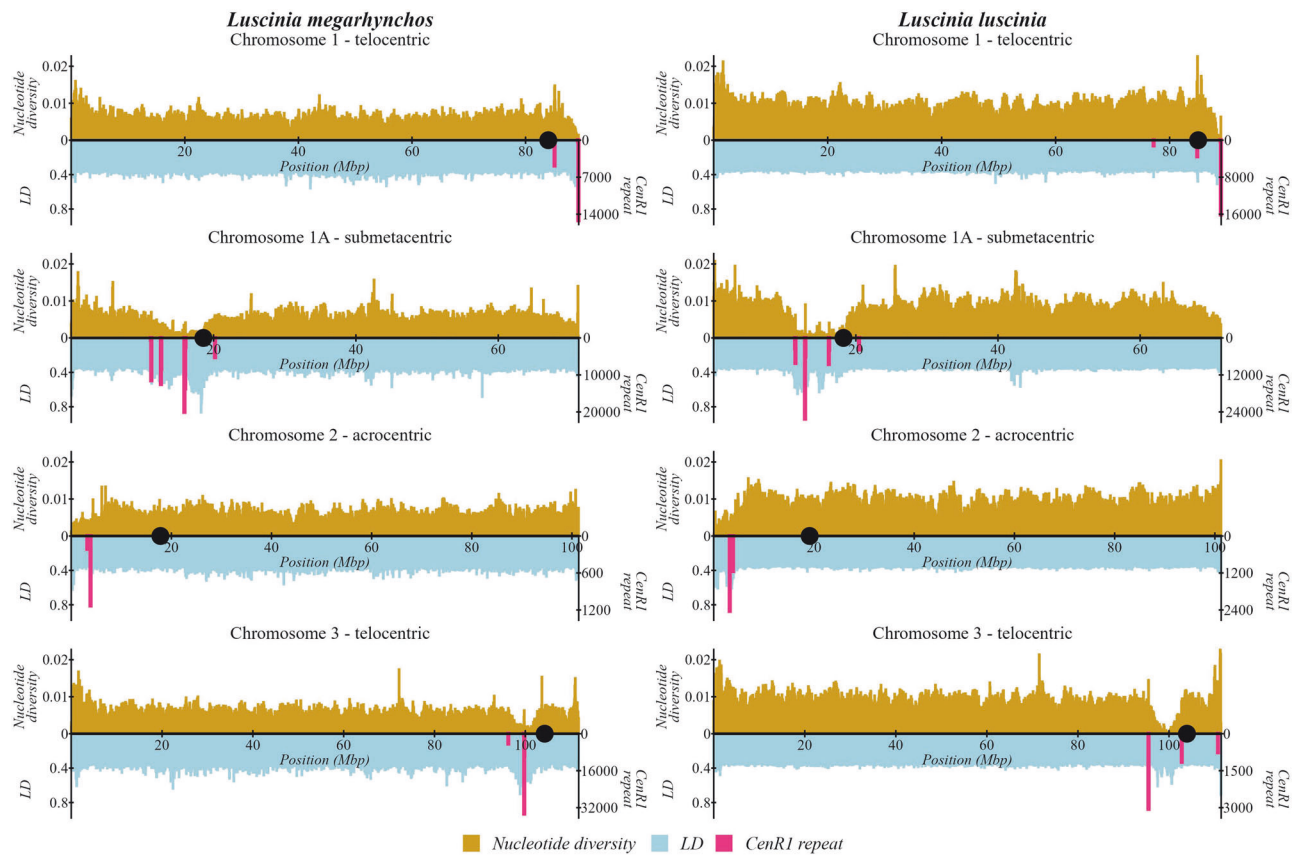


Fig. 1 Nucleotide diversity (yellow), linkage disequilibrium (LD) (blue) and the cumulative length of centromeric satellites *LmegCenR1* and *LlusCenR1* (magenta) in 50 kb sliding windows along four largest nightingale chromosomes. Chromosome morphology classifications adapted from Poignet et al. (2021). Black dots mark the positions of centromeres estimated from the cytogenetic data. The estimated positions of centromeres roughly colocalize with the genomic regions of decreased nucleotide diversity and elevated linkage disequilibrium and with the regions containing *CenR1* repeats.

We then used previously published cytogenetic data (Poignet et al. 2021), where centromeres on pachytene chromosomes were stained with the CREST antibody. The centromere position on each chromosome was estimated by comparing the relative lengths of the short and long chromosome arms in 15 cells per individual and calculating the median arm ratio. The chromosomes were classified as telocentric, acrocentric, submetacentric or metacentric based on the estimated centromere position. The positions of centromeres from cytogenetic data corresponded well with the predicted centromere position based on the levels of π and LD for all chromosomes, where we were able to homologize the assembled chromosomes with chromosomes from cytogenetic experiments based on the relative chromosome lengths (Fig. 1 and Supplementary Fig. 1). A few observed discrepancies, where the position of the centromere from the cytogenetic data is slightly shifted compared to the estimated centromere position from the genomic data (Supplementary Fig. 1), could be explained by missing small parts of our assemblies.

The final estimated position of the centromeres on individual chromosomes was the same for both nightingale species, suggesting that no structural changes affecting the centromere position occurred after the two species diverged.

Identification of candidate centromeric repeat sequences

RepeatExplorer2 was used to identify repetitive sequences in both nightingale genomes, while RepeatMasker was used to identify the locations and abundance of these repeat sequences in the genomes. We found 5 tandem repeats in the common nightingale and 8 tandem repeats in the thrush nightingale (Supplementary

Tables 1 and 2). The most frequent repeat in both species was a 436-bp satellite, which was present in 6.4% of the analyzed reads in *L. megarhynchos* and in 4.0% of the analyzed reads in *L. luscinia*. The genome localization of the identified repeats was further compared to the estimated position of the centromere based on nucleotide diversity and cytogenetic data. The most abundant tandem repeat was the only repeat whose position corresponded with the predicted centromeric regions (Fig. 1 and Supplementary Fig. 1). It was found almost exclusively in centromeric regions and was missing in other parts of the genome. None of the other putative repeats showed such a clear colocalization across chromosomes (data not shown). This makes the most common tandem repeat the best candidate for the centromeric repeat in nightingales, hence we refer to it as *CenR1* generally in both species and more specifically *LmegCenR1* in the common nightingale and *LlusCenR1* in the thrush nightingale.

The overall sequence identity between the *CenR1* consensus sequences of the two nightingale species was 98.4% (Supplementary Fig. 2), showing the high similarity of this candidate centromeric repeat in the two nightingale species. The *LmegCenR1* sequence of *L. megarhynchos* also showed 81% identity to the previously identified 436-bp satellite repeat (*fAlbSat4*) in the genome of collared flycatcher (*Ficedula albicollis*) (Suh et al. 2018). In addition, nucleotide blast search against NCBI nr database showed hits to chromosomes of European robin (*Erithacus rubecula*, GenBank assembly accession GCA_903797595.2) with an average identity 78%.

In humans, a 17-bp long motif called CENP-B box was identified in centromeric satellites as a binding site for centromere protein B

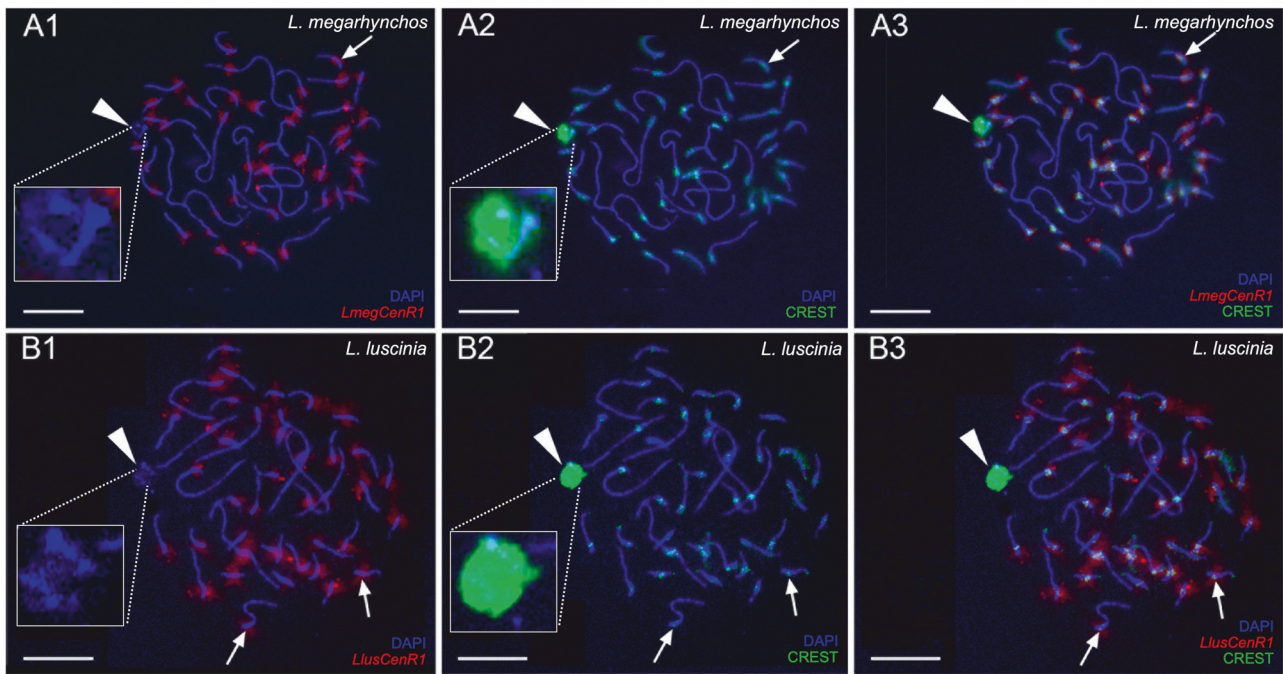


Fig. 2 Fluorescence in situ hybridization (FISH) of the *CenR1* repeat on pachytene chromosomes of two nightingale species. Pachytene chromosomes are stained by an antibody against the lateral components of the synaptonemal complex, anti-SYCP3 antibody (blue). FISH-mapping of *LmegCenR1* repeat on *L. megarhynchos* pachytene spreads (**A1**) and *LlusCenR1* repeat on *L. luscinia* pachytene spreads (**B1**) (red). The centromeres on regular chromosomes (indicated by arrows) and the whole GRC (indicated by arrowheads and magnified in insets) are labeled by the CREST antibody (green) (**A2**, **B2**). *CenR1* repeat signal colocalizes with the CREST signal in all chromosomes except the GRC, which show no *CenR1* repeat signal (**A3**, **B3**). Scale bar = 10 μ m.

which plays a significant role in centromere assembly (Masumoto et al. 1989). A putative CENP-B box with sequence similarity to human consensus sequence was later identified in other species (Gamba and Fachinetti 2020) including several bird taxa, e.g. pigeon, chaffinch and chicken (Solovei et al. 1996; Krasikova et al. 2012). We identified a putative CENP-B box motif in the *CenR1* satellite in both nightingale species, sharing the same sequence with an overall 65% similarity to the mammalian CENP-B box consensus (Supplementary Fig. 3).

Candidate centromeric repeat co-localizes with centromeric markers

To further verify that the *CenR1* represents the centromeric repeat, we prepared probes against these satellites in both nightingale species and hybridized them on mitotic and meiotic chromosome spreads of the corresponding species. In mitotic spreads, the signals were observed in centromeric regions, visible as chromosomal primary constrictions, in all macrochromosomes as well as microchromosomes. The intensity of the signals, nevertheless, varied in size among the chromosomes in both species (Supplementary Fig. 4A, B), suggesting that individual chromosomes may differ in the copy number of satellite repeats. A similar pattern was observed on pachytene spreads, where synaptonemal complexes were labeled with antibodies against their lateral components (SYCP3) and the centromeres were stained using the CREST antibody. We observed a complete overlap of signals between the CREST antibodies and *CenR1* repeat in all regular chromosomes of both species (Fig. 2), strongly suggesting that this satellite represents the centromeric repeat on all regular chromosomes.

We further performed a cross-species experiment and hybridized the *L. megarhynchos* *CenR1* PCR probe to *L. luscinia* chromosomes and vice versa. Hybridization was effective in both directions (Supplementary Fig. 4C, D), which is consistent with the finding that the *CenR1* satellite shows high similarity between the

nightingale species. The two-colored FISH experiment also showed perfect colocalization of the PCR probes of the two nightingale species in the centromeric regions on spreads of both species (Supplementary Fig. 5).

GRCs have different centromeric sequences than other chromosomes

We identified the GRC in pachytene spreads from testis in both nightingale species. The GRC exists as a small univalent and is labeled diffusely along its entire length by the CREST antibody (Fig. 2A2, B2), while the CREST staining is restricted to centromeric regions in other chromosomes (Poignet et al. 2021; Sotelo-Muñoz et al. 2022; Torgasheva et al. 2019). Interestingly, in contrast to other chromosomes, the *CenR1* probe showed no signal on the GRC in either of the two nightingale species (Fig. 2A1–B3), suggesting that the *CenR1* repeat is missing, has very low abundance or is modified on this chromosome.

To verify the absence or modification of the centromeric repeat on the GRC, we performed the BLASTN search of the *CenR1* repeat against published GRC sequences (Schlebusch et al. 2023). The BLASTN resulted in a single hit in the thrush nightingale GRC assembly localized to one scaffold (accession JAOYSP010000059.1). Although it showed high similarity (99% to 100%), the hit was split into three alignment blocks localized at both ends of the scaffold and covered only 78% of the *LlusCenR1* motif (see Supplementary Figs. 6A and 7 for dot plots and alignments). In the common nightingale, there were three GRC scaffolds showing hits to the *LmegCenR1* motif (accessions JAOYSO010001234.1, JAOYSO010000748.1, and JAOYSO010000749.1). The scaffold JAOYSO010001234.1 contained just a small piece of the motif (28%) (Supplementary Figs. 6B and 8). The scaffolds JAOYSO010000748.1 and JAOYSO010000749.1 showed alignments covering 80% and 100% of the *LmegCenR1* motif, respectively, however, they were not continuous but interrupted by gaps in the scaffolds (stretches of Ns in the contig sequences of undefined length). Alignments to the scaffolds JAOYSO010000748.1 and

JA0YSO010000749.1 also showed lower sequence similarity ranging from 89% to 97% (see Supplementary Figs. 6C, D, 9 and 10).

DISCUSSION

Centromeres play a crucial role in chromosome segregation during cell division through their ability to attach chromosomes to the mitotic or meiotic spindle. However, although high-quality reference genome assemblies now exist for many species, centromeric sequences are often missing or unidentified in genome assemblies due to their repetitive content. Here we have identified a candidate centromeric sequence in two closely related nightingale species. It consists of a 436-bp satellite repeat and is present in centromeric regions of all chromosomes except the GRC. Below we discuss our findings in the context of centromere evolution in birds and consider the possible implications of the different centromeric sequences for the inheritance of the GRC.

In birds, centromere composition has so far only been well described in the chicken. In this species, most acrocentric chromosomes have centromeres that are composed of a tandem array of the same 41-bp repeat called CNM (Huang et al. 2023). Although the CNM repeat appears to be an important part of the centromeres, the kinetochore itself does not usually bind to its sequence, but to its boundaries or to a smaller variable repeat motif embedded within the CNM (Huang et al. 2023). On the other hand, centromeres of metacentric or submetacentric chromosomes consist of chromosome-specific tandem repeats (Piégu et al. 2018).

Among songbirds, centromeric repeats have been identified in the zebra finch (Takki et al. 2022), chaffinch (Saiftdinova et al. 2001), and Corvoidea (crows, birds-of-paradise and relatives; Weissensteiner et al. 2017; Peona et al. 2023). In the zebra finch, analysis of the repetitive content of the genome, followed by FISH with the candidate repeats on the metaphase chromosomes, revealed that the centromeres of all chromosomes are composed of two distinct tandem repeats, *Tgut191A* and *Tgut716A*, with monomer lengths of 191 bp and 716 bp, respectively (Takki et al. 2022). In the chaffinch, genome restriction digestion followed by cloning of the repeat sequence and verification of the centromeric location by FISH led to the identification of a 505–506-bp tandem repeat present in the centromeric regions of all chromosomes (Saiftdinova et al. 2001). In Corvoidea, the combination of tandem repeat identification across species with population genomic data of hooded crows identified candidate centromeric repeats, pericentromeric repeats, and centromere-containing assembly gaps across autosomes (Weissensteiner et al. 2017; Peona et al. 2023). Here we identified the centromeric tandem repeat, called *CenR1*, with a monomer length of 436 bp in two other songbird species, the common nightingale and the thrush nightingale. While these two closely related species have the same tandem repeat in their centromeres, the repeat shows no similarity to the centromeric repeats identified in zebra finch and chaffinch. However, we have identified the homologous tandem repeat to *CenR1* in the genomes of the collared flycatcher, previously described as *fAlbSat4* (Suh et al. 2018), and the European robin. These two species, together with nightingales, belong to the family Muscicapidae (Zhao et al. 2023). Therefore, it is possible that this centromeric sequence represents an ancestral centromeric sequence in this taxon, although this needs to be verified.

The centromeric function of the *CenR1* satellite repeat in nightingales is supported not only by its presence in the centromeric regions of all regular chromosomes, as has been verified by the FISH analysis, but also by the presence of the CENP-B box in its sequence. The CENP-B box is a 17-bp sequence motif that is recognized by the kinetochore-associated protein CENP-B and is common in centromeric satellite sequences of many species (Gamba and Fachinetti 2020; Saffery et al. 1999). Moreover, a probe against the *CenR1* tandem repeat colocalizes with the

CREST antacentromere serum, which is known to bind the kinetochore proteins CENP-A, CENP-B, and CENP-C (Mchugh 2007). However, to prove that the kinetochore binds directly to the *CenR1* satellite repeat and not to a small embedded or nearby sequence, an antibody to the CENP-A kinetochore protein would need to be developed and used to immunoprecipitate the DNA directly bound to it, as has been done in chicken (Shang et al. 2010, 2013). Even if the *CenR1* satellite does not form a core of the centromere and does not bind directly to the kinetochore, it can form a pericentric heterochromatin, which is important for sister chromatin cohesion and accurate chromosome segregation.

Interestingly, the GRC was the only chromosome that showed no signal in the FISH experiment with the *CenR1* probe, suggesting that this repeat is either missing on the GRC, or has a lower number of copies than can be detected by FISH, or has diverged sufficiently so that probe does not bind to it. In accordance with this, analysis of the currently available GRC assembly (Schlebusch et al. 2023) revealed only a few sequences similar to the centromeric repeat, but these were either fragmented or relatively highly divergent. It is therefore possible that centromeric repeats present on regular chromosomes are degenerated on the GRC. However, this result needs to be interpreted with caution, given the highly fragmented and still incomplete nature of the GRC assembly. Notably, the GRC also shows different staining with the CREST serum compared to the regular chromosomes. While CREST only stains the centromeric regions on regular chromosomes, it labels the entire length of the GRC in male pachytene cells. Although it is not clear what the CREST serum binds to given its autoimmune character, this may indicate that the kinetochore binds to the GRC in a different way compared to regular chromosomes, which might be related to its unusual behavior in male meiosis.

In several systems, it has been shown that the inability of chromosomes to attach to the mitotic/meiotic spindle or the failure of chromatid separation and their subsequent lagging in the anaphase can lead to their elimination from the nucleus (Dedukh and Krasikova 2022; Ishii et al. 2016). We, therefore, propose that a modification of the centromeric sequence on the GRC could be the underlying mechanism for the programmed elimination of this chromosome from somatic cells and male germ cells. The problems with chromatid separation and chromosome lagging in anaphase can also lead to the duplication of the chromosome in one daughter cell while it is eliminated from the other cell if cell division is asymmetric (Banaei-Moghaddam et al. 2012; Johnson Pokorná and Reifová 2021; Ruban et al. 2020; Wu et al. 2019). Modification of the GRC centromere could thus also explain how the chromosome is duplicated in the female germline at some point during germline development and how the GRC copy number variation occasionally arises. Further cytogenetic analyses of early embryos and male and female gonads are required to test this hypothesis.

DATA AVAILABILITY

The Illumina whole genome sequencing data generated in this study have been deposited in the NCBI's SRA database under the BioProject accession code (PRJNA1178309). The consensus sequences of *LmegCenR1* and *LlusCenR1* have been deposited in the NCBI's GenBank database (accession numbers PQ530385 and PQ530386, respectively).

REFERENCES

- Altomose N, Logsdon GA, Bzikadze AV, Sidhwani P, Langley SA, Caldas GV et al. (2022) Complete genomic and epigenetic maps of human centromeres. *Science* 376:eabl4178
- Banaei-Moghaddam AM, Schubert V, Kumke K, Weiß O, Klemme S, Nagaki K et al. (2012) Nondisjunction in favor of a chromosome: the mechanism of Rye B chromosome drive during pollen mitosis. *Plant Cell* 24:4124

- Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356:519–520
- Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW et al. (2007) Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5:e310
- Blommaert J, Sandoval-Castillo J, Beheregaray LB, Wellenreuther M (2024) Peering into the gaps: long-read sequencing illuminates structural variants and genomic evolution in the Australasian snapper. *Genomics* 116:110929
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Borodin P, Chen A, Forstmeier W, Fouché S, Malinovskaya L, Pei Y et al. (2022) Mendelian nightmares: the germline-restricted chromosome of songbirds. *Chromosome Res* 30:255–272
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- Chang X, He X, Li J, Liu Z, Pi R, Luo X et al. (2024) High-quality *Gossypium hirsutum* and *Gossypium barbadense* genome assemblies reveal the landscape and evolution of centromeres. *Plant Commun* 5:100722
- Chmátal L, Schultz RM, Black BE, Lampson MA (2017) Cell biology of cheating: transmission of centromeres and other selfish elements through asymmetric meiosis. *Prog Mol Subcell Biol* 56:377–396
- Clark FE, Aker A (2021) Unravelling the mystery of female meiotic drive: where we are. *Open Biol* 11:210074
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. (2021) Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008
- Dedukh D, Krasikova A (2022) Delete and survive: strategies of programmed genetic material elimination in eukaryotes. *Biol Rev Camb Philos Soc* 97:195–216
- Gamba R, Fachinetti D (2020) From evolution to function: two sides of the same CENP-B coin? *Exp Cell Res* 390:111959
- Gao S, Zhang Y, Bush SJ, Wang B, Yang X, Ye K (2024) Centromere landscapes resolved from hundreds of human genomes. *Genom Proteom Bioinform* 22:qzae071
- Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293:1098–1102
- Holland AJ, Cleveland DW (2009) Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* 10:478–487
- Huang Z, Xu Z, Bai H, Huang Y, Kang N, Ding X et al. (2023) Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA* 120:e2216641120
- Ishii T, Karimi-Ashtiyani R, Houben A (2016) Haploidization via chromosome elimination: means and mechanisms. *Annu Rev Plant Biol* 67:421–438
- Johnson Pokorná M, Reifová R (2021) Evolution of B chromosomes: from dispensable parasitic chromosomes to essential genomic players. *Front Genet* 12:727570
- Kapusta A, Suh A (2017) Evolution of bird genomes—a transposon's-eye view. *Ann NY Acad Sci* 1389:164–185
- Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166
- Khost DE, Eickbush DG, Larracuente AM (2017) Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. *Genome Res* 27:709
- Kinsella CM, Ruiz-Ruano FJ, Dion-Côté A-M, Charles AJ, Gossmann TI, Cabrero J et al. (2019) Programmed DNA elimination of germline development genes in songbirds. *Nat Commun* 10:5468
- Krasikova A, Fukagawa T, Zlotina A (2012) High-resolution mapping and transcriptional activity analysis of chicken centromere sequences on giant lampbrush chromosomes. *Chromosome Res* 20:995–1008
- Kursel LE, Malik HS (2018) The cellular mechanisms and consequences of centromere drive. *Curr Opin Cell Biol* 52:58–65
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* <https://doi.org/10.48550/arXiv.1303.3997>
- Logsdon GA, Rozanski AN, Ryabov F, Potapova T, Shepelev VA, Catacchio CR et al. (2024) The variation and evolution of complete human centromeres. *Nature* 629:136–145
- Malinovskaya LP, Zadesenets KS, Karamysheva TV, Akberdina EA, Kizilova EA, Romanenko MV et al. (2020) Germline-restricted chromosome (GRC) in the sand martin and the pale martin (Hirundinidae, Aves): synopsis, recombination and copy number variation. *Sci Rep* 10:1058
- Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* 109:1963–1973
- Mchugh NJ (2007) 20—Centromere autoantibodies. In: Shoenfeld Y, Gershwin ME, Meroni PL (eds) *Autoantibodies*, 2nd edn. Elsevier, Burlington, pp 151–157
- McKinley KL, Cheeseman IM (2016) The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* 17:16–29
- Moens PB (2006) Zebrafish: chiasmata and interference. *Genome* 49:205–208
- Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11:378
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29:792–793
- Novák P, Ávila Robledillo L, Kobližková A, Vrbová I, Neumann P, Macas J (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res* 45:e111
- Novák P, Neumann P, Macas J (2020) Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc* 15:3745–3776
- Pei Y, Forstmeier W, Ruiz-Ruano FJ, Mueller JC, Cabrero J, Camacho JPM et al. (2022) Occasional paternal inheritance of the germline-restricted chromosome in songbirds. *Proc Natl Acad Sci USA* 119:e2103960119
- Peona V, Kutschera VE, Blom MPK, Irestedt M, Suh A (2023) Satellite DNA evolution in Corvoidea inferred from short and long reads. *Mol Ecol* 32:1288–1305
- Peters AH, Plug AW, van Vugt MJ, de Boer P (1997) A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. *Chromosome Res* 5:66–68
- Piégu B, Arensburg P, Guillou F, Bigot Y (2018) But where did the centromeres go in the chicken genome models? *Chromosome Res* 26:297–306
- Pigozzi MI, Solari AJ (1998) Germ cell restriction and regular transmission of an accessory chromosome that mimics a sex body in the zebra finch, *Taeniopygia guttata*. *Chromosome Res* 6:105–113
- Pigozzi MI, Solari AJ (2005) The germ-line-restricted chromosome in the zebra finch: recombination in females and elimination in males. *Chromosoma* 114:403–409
- Poignet M, Johnson Pokorná M, Altmanová M, Majtánová Z, Dedukh D, Albrecht T et al. (2021) Comparison of karyotypes in two hybridizing passerine species: conserved chromosomal structure but divergence in centromeric repeats. *Front Genet* 12:768987
- R Core Team (2023) R: a language and environment for statistical computing. <https://www.R-project.org/>
- Rosin LF, Mellone BG (2017) Centromeres drive a hard bargain. *Trends Genet* 33:101–117
- Ruban A, Schmutzer T, Wu DD, Fuchs J, Boudichevskaya A, Rubtsova M et al. (2020) Supernumerary B chromosomes of *Aegilops speltoides* undergo precise elimination in roots early in embryo development. *Nat Commun* 11:2764
- Saffery R, Earle E, Irvine DV, Kalitsis P, Choo KH (1999) Conservation of centromere protein in vertebrates. *Chromosome Res* 7:261–265
- Saifitdinova AF, Derjushcheva SE, Malykh AG, Zhurov VG, Andreeva TF, Gaginikaya ER (2001) Centromeric tandem repeat from the chaffinch genome: isolation and molecular characterization. *Genome* 44:96–103
- Schlebusch SA, Rídl J, Poignet M, Ruiz-Ruano FJ, Reif J, Pajer P et al. (2023) Rapid gene content turnover on the germline-restricted chromosome in songbirds. *Nat Commun* 14: 4579
- Shang W-H, Hori T, Toyoda A, Kato J, Pependorf K, Sakakibara Y et al. (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res* 20:1219–1228
- Shang W-H, Hori T, Martins NMC, Toyoda A, Misu S, Monma N et al. (2013) Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev Cell* 24:635–648
- Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0. 2013–2015. <http://www.repeatmasker.org>
- Solovei IV, Joffe BI, Gaginikaya ER, Macgregor HC (1996) Transcription on lampbrush chromosomes of a centromerically localized highly repeated DNA in pigeon (*Columba*) relates to sequence arrangement. *Chromosome Res* 4:588–603
- Sotelo-Muñoz M, Poignet M, Albrecht T, Kautz A, Dedukh D, Schlebusch SA et al. (2022) Germline-restricted chromosome shows remarkable variation in size among closely related passerine species. *Chromosoma* 131:77–86
- Sottas C, Reif J, Piálak L, Poignet M, Kverek P, Dolata PT et al. (2023) Patterns of hybridization in a secondary contact zone between two passerine species, the common nightingale *Luscinia megarhynchos* and the thrush nightingale *Luscinia luscinia*. *J Avian Biol* 2023:e03061
- Storchová R, Reif J, Nachman MW (2010) Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution* 64:456–471
- Suh A, Smeds L, Ellegren H (2018) Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Mol Ecol* 27:99–111
- Takki O, Komissarov A, Kulak M, Galkina S (2022) Identification of centromere-specific repeats in the zebra finch genome. *Cytogenet Genome Res* 162:55–63
- Talbert PB, Henikoff S (2020) What makes a centromere? *Exp Cell Res* 389:111895

- Torgasheva AA, Malinovskaya LP, Zadesenets KS, Karamysheva TV, Kizilova EA, Akberdina EA et al. (2019) Germline-restricted chromosome (GRC) is widespread among songbirds. *Proc Natl Acad Sci USA* 116:11845–11850
- Torgasheva A, Malinovskaya L, Zadesenets K, Shnaider E, Rubtsov N, Borodin P (2021) Germline-restricted chromosome (GRC) in female and male meiosis of the great tit (*Parus major*, Linnaeus, 1758). *Front Genet* 12:768056
- Uno Y, Nishida C, Hata A, Ishishita S, Matsuda Y (2019) Molecular cytogenetic characterization of repetitive sequences comprising centromeric heterochromatin in three Anseriformes species. *PLoS ONE* 14:e0214028
- Voleníková A, Lukšířová K, Mora P, Pavlica T, Altmanová M, Štundlová J et al. (2023) Fast satellite DNA evolution in *Nothobranchius* annual killifishes. *Chromosome Res* 31:33
- Weighill D, Macaya-Sanz D, DiFazio SP, Joubert W, Shah M, Schmutz J et al. (2019) Wavelet-based genomic signal processing for centromere identification and hypothesis generation. *Front Genet* 10:487
- Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Petterson O, Suh A et al. (2017) Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Res* 27:1–12
- Wickham H (2016) ggplot2: elegant graphics for data analysis. <https://ggplot2.tidyverse.org>
- Wu D, Ruban A, Fuchs J, Macas J, Novák P, Vaio M et al. (2019) Nondisjunction and unequal spindle organization accompany the drive of *Aegilops speltoides* B chromosomes. *New Phytol* 223:1340–1352
- Yamada K, Shibusawa M, Tsudzuki M, Matsuda Y (2002) Molecular cloning and characterization of novel centromeric repetitive DNA sequences in the blue-breasted quail (*Coturnix chinensis*, Galliformes). *Cytogenet Genome Res* 98:255–261
- Yi C, Wang M, Jiang W, Wang D, Zhou Y, Gong Z et al. (2015) Isolation and identification of a functional centromere element in the wild rice species *Oryza granulata* with the GG genome. *J Genet Genomics* 42:699–702
- Zhao M, Gordon Burleigh J, Olsson U, Alström P, Kimball RT (2023) A near-complete and time-calibrated phylogeny of the Old World flycatchers, robins and chats (Aves, Muscicapidae). *Mol Phylogenet Evol* 178:107646

ACKNOWLEDGEMENTS

This research was funded by the Czech Science Foundation (grant 23-07287S to RR, TA and DD), the Grant Agency of Charles University (grant 314222 to ZH) and the Institutional Research Support grant of the Charles University (SVV 260684/2024).

AUTHOR CONTRIBUTIONS

The project was conceptualized by RR and JRi; Samples were collected by JRi, TA and TSO; Cytogenetic analysis was performed by DD; DNA extraction was done by JRi and Illumina sequencing was done by MOL and VB; Bioinformatic analyses were done by JRi, ZH, and SAS with consultation from AS and FJR-R; Manuscript was written by JRi, RR, DD, ZH and SAS and all authors contributed to text editing.

FUNDING

Open access publishing supported by the institutions participating in the CzechELib Transformative Agreement.

COMPETING INTERESTS

The authors declare no competing interests.

RESEARCH ETHICS STATEMENT

The work was carried out in accordance with ethical animal research requirements of Poland according to Polish law (the Act On the Protection of Animals Used for Scientific or Educational Purposes, 15.01.2015, item 266, implementing Directive 2010/63/EU of the European Parliament and of the European Council of 22.09.2010). Experiments on birds were approved by the General Directorate for Environmental Protection (permission no. DZP-WG.6401.03.123.2017.dl.3).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-025-00779-5>.

Correspondence and requests for materials should be addressed to Jakub Řídl or Radka Reifová.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025