This is a post-peer-review, pre-copyedit version of an article published in Chromosoma. The final authenticated version is available online at: [http://dx.doi.org/10.1007/s00412-018-0662-0]

ORIGINAL ARTICLE

CrossMark

Low-pass single-chromosome sequencing of human small supernumerary marker chromosomes (sSMCs) and *Apodemus* B chromosomes

Alexey I. Makunin [1] · Marija Rajičić [2] · Tatyana V. Karamysheva [3] · Svetlana A. Romanenko [1,4] · Anna S. Druzhkova [1,4] · Jelena Blagojević [2] · Mladen Vujošević [2] · Nikolay B. Rubtsov [3] · Alexander S. Graphodatsky [1,4] · Vladimir A. Trifonov [1,4]

Abstract
Supernumerary chromosomes sporadically arise in many eukaryotic species as a result of genomic rearrangements. If present in a substantial part of species population, those are called B chromosomes, or Bs. This is the case for 70 mammalian species, most of which are rodents. In humans, the most common types of extra chromosomes, sSMCs (small supernumerary marker chromosomes), are diagnosed in approximately 1 of 2000 postnatal cases. Due to low frequency in population, human sSMCs are not considered B chromosomes. Genetic content of both B-chromosomes and sSMCs in most cases remains understudied. Here, we apply microdissection of single chromosomes with subsequent low-pass sequencing on Ion Torrent PGM and Illumina MiSeq to identify unique and repetitive DNA sequences present in a single human sSMC and several B chromosomes in mice *Apodemus flavicollis* and *Apodemus peninsulae*. The pipeline for sequencing data analysis was made available in Galaxy interface as an addition to previously published command-line version. Human sSMC was attributed to the proximal part of chromosome 15 long arm, and breakpoints leading to its formation were located into satellite DNA arrays. Genetic content of *Apodemus* B chromosomes was species-specific, and minor alterations were observed in both species. Common features of Bs in these *Apodemus* species were satellite DNA and ERV enrichment, as well as the presence of the vaccinia-related kinase gene *Vrk1*. Understanding of the non-essential genome elements content provides important insights into genome evolution in general.

Keywords Genome instability · *Apodemus peninsulae* · *Apodemus flavicollis* · Galaxy · Pipeline

Introduction

Supernumerary chromosomes are dispensable karyotype elements that can be treated as additional relative to the main

✉ Alexey I. Makunin
    alex.makunin@gmail.com

1   Institute of Molecular and Cellular Biology Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

2   Institute for Biological Research "Siniša Stanković", University of Belgrade, Belgrade, Serbia

3   Institute of Cytology and Genetics Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

4   Novosibirsk State University, Novosibirsk, Russia

chromosome set. In human, supernumerary chromosomes of various morphologies are sometimes formed de novo but frequencies of their occurrence in populations are quite low. For example, small supernumerary marker chromosomes (sSMCs) are found in 0.044% of newborn children, but less than 30% of those are inherited (Liehr and Weise 2007). In some other species, supernumeraries are significantly more frequent in populations due to effective non-Mendelian inheritance mode, known as drive (Jones 1991). These elements are called B chromosomes, or Bs—in contrast to A chromosomes of the main set. Bs usually do not provide any specific phenotypic effects, while human sSMCs are quite often (~ 30% carriers) associated with various disorders, most frequently mental retardation and male infertility (Liehr et al. 2004; Liehr and Weise 2007). Most supernumeraries originate from the host chromosome fragments, although several examples of Bs resulting from interspecific hybridization exist (Kour et al. 2009; Lamatsch et al. 2011). The regions of origin are not

pre-determined but tend to occur in specific regions: human sSMC are most frequently homologous to proximal regions of chromosome HSA15 (Liehr et al. 2004); protooncogene *KIT* is located in the studied Bs of foxes (Graphodatsky et al. 2005), raccoon dogs (Yudkin et al. 2007), and South American gray brocket deer (Makunin et al. 2016). Evolution of B chromosomes usually includes recombinational detachment from the main genome, accumulation of mutations as well as repetitive elements, and eventual degradation (Houben et al. 2014).

Advancements in high-throughput sequencing and related technologies made it possible to create genomic resources for a multitude of species, including non-model ones (Koepfli et al. 2015). Thus, it became possible to utilize conserved synteny of DNA sequences or genes in order to reconstruct ancestral karyotypes (Kim et al. 2017), characterize the breakpoint regions (Larkin et al. 2009), investigate whole-genome duplications (Lien et al. 2016; Session et al. 2016), etc. However, chromosome-level assemblies are still scarce, as they require costly and labor-intensive mapping procedures. Besides, not all species are subject to whole-genome sequencing and assembly. Thus, lineage-specific rearrangements and non-standard genome instability cases, such as supernumerary chromosomes, remain understudied.

Previously, we utilized Illumina MiSeq data generated from flow-sorted chromosomes amplified with degenerated oligonucleotide-primed-polymerase chain reaction (DOP-PCR) (Telenius et al. 1992) to investigate the origin of B chromosomes in two deer species (Makunin et al. 2016), as well as to connect anole lizard (*Anolis carolinensis*) scaffolds to previously uncharacterized microchromosomes and to analyze rearrangements between two *Anolis* species (Kichigin et al. 2016). Here, we integrate our pipeline into Galaxy interface and successfully apply it to single chromosome copies of microdissected human sSMC and B chromosomes of two mice species: *Apodemus flavicollis* and *Apodemus peninsulae*. We demonstrate that the studied human sSMC was derived from the pericentric region of chromosome 15, while B chromosomes originate from multiple genomic regions in a species-specific manner, with one protein-coding gene, *Vrk1* kinase, shared between species and only subtle genetic content differences within species.

Results

Galaxy tools and workflow for isolated chromosome sequencing

The majority of steps needed for the inference of chromosome-specific regions can be performed with the tools already available in Galaxy ToolShed: cutadapt (Martin 2011) for trimming of primers and adapters from reads, bowtie2 (Langmead and Salzberg 2012) for read alignment to reference genome, and bedtools (Quinlan and Hall 2010) for Binary Alignment Map

(BAM) to Browser Extensible Data (BED) format conversion and BED merging. Two additional tools were implemented based on scripts from command-line version of DOPSeq_analyzer (Makunin et al. 2016):

contam_filter takes two BAM alignments of the same reads to two genomes: target and contaminant (in case of *Apodemus* those were mouse and human, respectively). It outputs alignments for which $MAPQ_{target} \geq MAPQ_{contaminant}$ (MAPQ - MAPping Quality). This procedure removes reads resulting from contamination with human DNA introduced during sample preparation. Additionally, user can specify minimal MAPQ for alignment retention (default value 20). This is useful as repetitive regions of the reference genome can produce false-positive signals.

region_dnacopy is a core tool of the pipeline providing prediction of chromosome-specific regions. Its first input is a BED file of positions covered by reads with additional column containing the number of reads within each position. It can be generated with bedtools BAM to BED and subsequent bedtools MergeBED (with "Count" operation on any column). Second input is a tab-separated file with chromosomes and their sizes. The tool produces two files. First one is a plot (Fig. 1) of distances between positions covered by reads in logscale, in which lower values correspond to chromosome-specific regions (i.e., reads are more densely located). This plot is generated only if no more than 42 chromosomes are analyzed. Second output is a tabular file with statistics on every inferred region. Regions with lower mean_pd (pairwise distances between positions covered by reads) are most likely present in the sampled chromosome.

Workflows in Galaxy are series of tool actions that can be run in a batch. We generated workflows for all types of data used in our study. "DOPSeq SE notrim" was used for human sSMC analysis and includes single-end read alignment to human genome, MAPQ filtering with contam_filter, BAM to BED conversion and prediction of target regions by region_dnacopy. "DOPSeq PE WGA" and "DOPSeq PE DOP" were used for *Apodemus* chromosomes amplified with Whole-Genome Amplification (WGA) and DOP-PCR, respectively. They start by trimming primers and Illumina adapter paired-end reads. Trimmed read pairs are aligned to both mouse and human genomes, and then human contaminant reads are removed. Next steps (coversion, region prediction) are performed similarly to single-end pipeline.

Human sSMC

Fluorescence in situ hybridization (FISH) of the microdissected chromosome against patient metaphase plate (Fig. S1a)
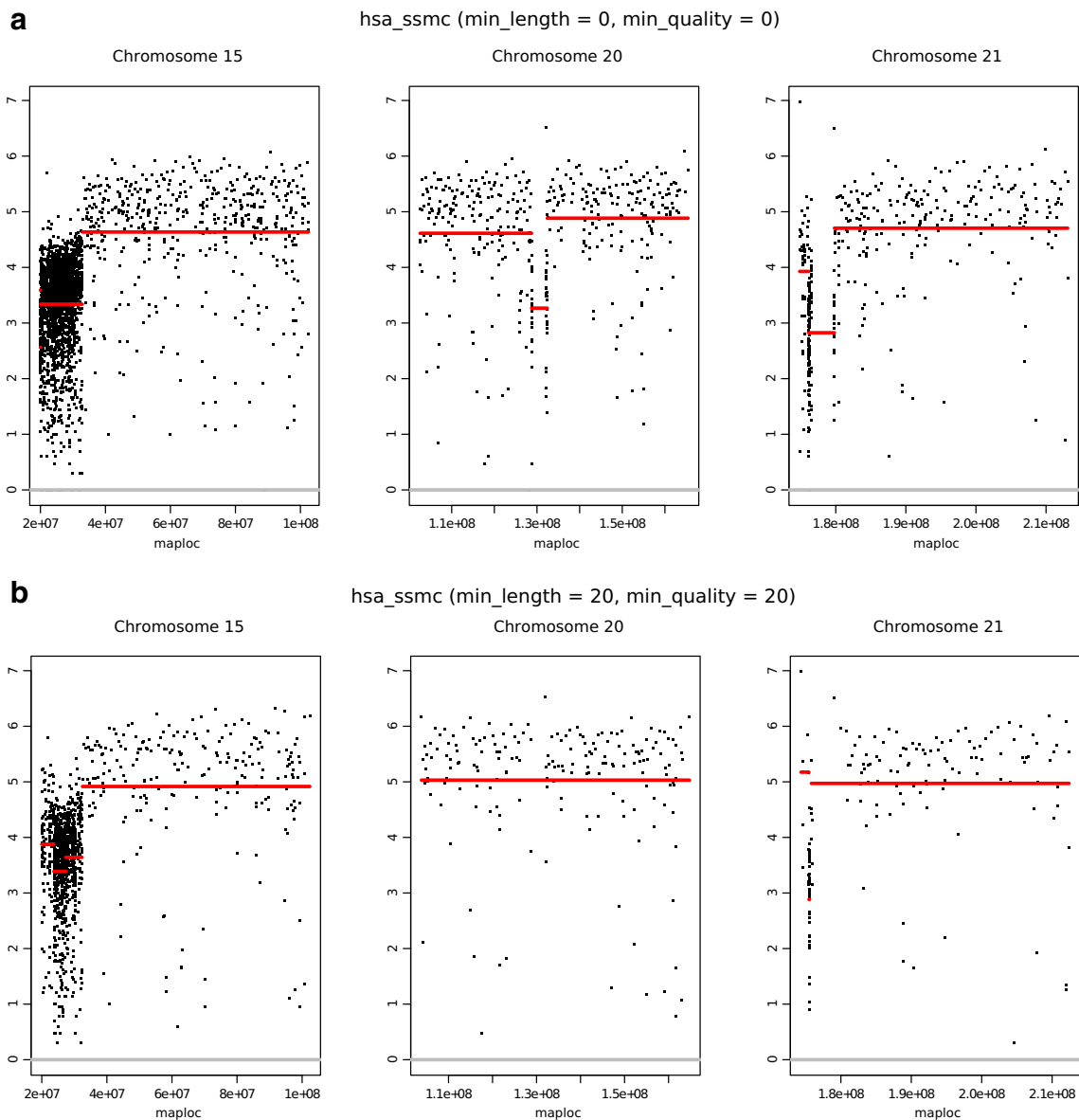
a hsa_ssmc (min_length = 0, min_quality = 0)

b hsa_ssmc (min_length = 20, min_quality = 20)

Fig. 1 **Region_dnacopy plot for human sSMC. a** Minimum MAPQ = 20. **b** Minimum MAPQ = 0. Only chromosomes 15, 20, and 21 are presented. Black dots correspond to individual distances between positions covered by reads (in logscale), while red lines are mean values inferred by the tool

indicated that the sSMC originated as an inverted duplication of the short arm and the pericentric region of the long arm of chromosome 15 (inv dup(15)(q13)), one of the frequent human sSMC types (Liehr et al. 2004). A total of 496,953 single-end reads were obtained with Ion Torrent PGM and subsequently aligned to human genome. After filtering (MAPQ ≥ 20), 91,918 high-quality alignments remained. These were merged into 12,912 read positions. Two clear chromosome-specific regions were detected: their coordinates after manual correction were HSA15:20,004,411–32,494,635 and HSA21:9,593,662–11,150,959 (Fig. 1a). First region corresponds to the long arm region identified with

FISH, while the second one seems to be an artifact resulting from centromeric satellite DNA mapping. It is interesting that without MAPQ filtering, the region of HSA15 extends to 32,886,960. This extension represents a field of tandem repeats similar to centromeric satellites. Besides, additional signals in pericentromeric regions of most chromosomes were detected resulting from additional satellite DNA mappings (Fig. 1b). Thus, the breakpoint leading to the formation of this sSMC is located within the field of tandem repeats at HSA15:32,495–32,887 kbp. Largest clusters of repetitive DNA for this sample include Alu SINE (Fig. S1b), centromeric satellites, and L1 LINE. Only a few reads of

rRNA distributed over the short arm of the chromosome 15 were found.

Apodemus flavicollis autosome and B chromosomes

A control autosome and five B chromosomes were isolated from metaphase preparations obtained from bone marrow and fibroblast cell cultures of three specimens: two from Serbia and one from Russia (Table S1). Chromosome-specific libraries were constructed using WGA kit (Sigma). Specificity of libraries was confirmed by FISH (Rajičić et al. 2017; Fig. S3c). A total of 1.6–5.6 million paired reads per sample were obtained on Illumina MiSeq. After trimming and removal of human DNA contamination, 3000–31,000 high-quality alignments per library remained. Those were merged into 1000–6000 positions covering 83–585 kbp throughout the mouse genome for each sample (Table S2).

In agreement with comparative painting data (Matsubara et al. 2004), AFL23 corresponded to mouse chromosome regions MMU10: 34–78 Mbp and MMU17: 27–38 Mbp (Table S3). The information on evolutionary breakpoint coordinates between *A. flavicollis* and *Mus musculus* chromosomes was obtained here for the first time. Besides, several small regions represented putative micro-rearrangements from 35 to 649 kbp in size. The latter cases require additional verification, e.g., with FISH localization of BAC clones.

We identified target regions independently for each B-chromosomal library. The libraries AflB1, AflB7, and Afl_B produced high-quality results, while for AflB5 and AflB6, mapping results included fewer positions and thus some of the smaller regions were not revealed. In the case of AflB5, it was caused by a lower content of target reads (2 vs 4% in higher quality libraries) coupled with lower read counts. AflB6 library presumably included less unique amplicons, as suggested by the increased coverage of reads within positions. The regions detected were in a good agreement with biological replicates (AflB5-7) and with populations (AflB1 and AflB5-7 from Serbia, Afl_B from Russia). The only population-specific region at MMU9: 72 Mbp was identified

in a sample from Russia (Fig. 2). A total of 22 common regions were identified spanning 3.5 Mbp in the mouse genome (Table S4). With most significant deletions (manifesting as gaps in coverage) removed, the total size of non-repetitive regions reduced to 2.2 Mbp. Unexpectedly, a high overlap was found between mapped sequences of WGA libraries from different biological samples of Afl Bs (AflB1, AflB7, and Afl_B) (Table S5), which is more likely due to non-random amplification, rather than because of read mappability issues.

Functional clustering could be performed for 29 of 38 genes located within the regions present on B chromosomes. This method indicated an enrichment of microtubule and cell cycle proteins (enrichment score 1): *Cenpe* (centromere protein E), *Dync1i2* (dynein cytoplasmic 1 intermediate chain 2), *Mns1* (meiosis-specific nuclear structural protein 1), and *Mapre1* (microtubule-associated protein, RP/EB family, member 1). Other less significant categories included nucleotide-binding (0.41), membrane (0.13), and metal-binding (0.12) proteins.

B chromosomes accumulated more substitutions and indels than autosomes (1 derived position per 28–31 bp vs 1 per 44 bp) relative to mouse genome, which is in agreement with the proposed degradation of additional chromosomes. For the same reason, an elevated proportion of non-synonymous mutations was observed in protein-coding genes. High-confidence protein-disrupting variants accompanied by multiple missense substitutions were observed in genes *Cenpe*, *Rraga*, and *Kdm6*.

The largest clusters of repeats in AFL23 were mostly composed of B2 SINE and L1 LINE families. In B chromosomes, satellite repeats (including MurSatRep1) together with ERVL (MaLR) and ERVK LTRs were also observed. Transposable elements were tandemly arranged in some cases.

Apodemus peninsulae B chromosomes

B chromosomes of two morphologies (macro-Bs—Ape_mB—and dot-like Bs—Ape_dB) were obtained from the cell cultures of two individuals (Table S1), and chromosome-specific
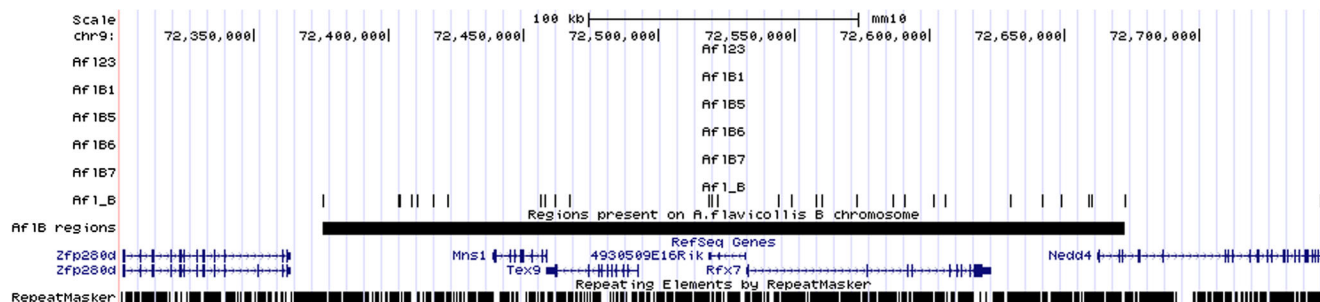

Fig. 2 B chromosome region specific to Russian population of *Apodemus flavicollis*. Reads are lacking from control autosome Afl23 and B chromosomes from Serbian population individuals (AflB1, AflB5, AflB6, AflB7), but reads from Russian population B chromosomes (Afl_B) do map on the region. Reference genome—mouse (mm10)

libraries were generated with DOP-PCR. Specificity of libraries was confirmed by FISH (Fig. S3a, b) (Karamysheva et al. 2002; Rubtsov et al. 2004). Sequencing results seemingly covered most available amplicons despite the high level of human contamination in Ape_dB and significant trimmed read shrinking in Ape_mB (Table S2).

Target regions were predicted for both types of B chromosomes. The total size of regions after removal of putative deletions was 9.4 Mbp for Ape_mB and 5.5 Mbp for Ape_dB, which is significantly higher than for Afl Bs (2.2 Mbp). Regions of Ape_dB represent a subset of Ape_mB regions (Table S6), but their size is often larger.

Some regions included representatives of gene families with high levels of sequence homology between members in the mouse genome and thus were detected more than once:

1) Three regions contained NLR family related to inflammation (*Nlrp9a*, *Nlrp9b*, *Nlrp9c*) from MMU7. Given the 80–83% homology between these genes in mouse genome at the positions covered by reads, we proposed that only one of these genes might be represented on Bs. The best candidate is the *Nlrp9c*, the only gene with upstream region covered.

2) Three regions contained genes from neurexin family. Their homologs showing 86% sequence identity are found on MMU1 (*Cntnap5a*, *Cntnap5b*) and MMU17 (*Cntnap5c*). Here, the homolog of the gene from MMU17 is more likely to be present on Bs, as it demonstrates higher sequence similarity and its 3′ end is covered by reads (no coverage at 3′ end in other two genes).

3) Two tandem arrays of *Vmn2r* (*Vmn2r8-Vmn2r17* at MMU5: 109 Mbp, *Vmn2r84-Vmn2r87* at MMU10: 130 Mbp—G protein-coupled receptors active in vomeronasal sensory neurons) were covered uniformly (90–99% homology). Some portion of array from MMU10 is likely to be present, as neighboring unique gene *Tespa1* is covered.

Regions with representatives of *Cntnap5* and *Vmn2r* families are present on both dot-like and macro-Bs, while *Nlrp9*—only on macro-Bs. Exclusion repetitively discovered regions as putative false positives resulting from gene family homology within mouse genome reduce B-chromosomal region sizes to 7296 kbp for Ape_mB and to 3935 kbp for Ape_dB.

For functional analysis, we used the combined gene list for both dot-like B and macro-B regions. Thirty out of 32, these genes could be used by DAVID GO search engine. Enrichment scores were high for nucleotide-binding (2.84), laminin and EGF-like domain-containing (1.93), cytoskeleton (0.7), and ion-binding (0.09) proteins. With Panther classifications enabled, a cluster of receptors arose with a high score (7.89).

Sequence variant density was higher in dot-like B chromosomes (one variant per 27 bp) than in macro B chromosomes (one variant per 35 bp). We did not observe higher variant density in the regions shared between B chromosome types, which would be expected if all common regions were older than macro-B-specific. In Ape_mB, two regions had no differences from mouse genome throughout 0.7 kbp (MMU12: 17 Mbp) and 1.9 kbp (MMU13: 74 Mbp) of the obtained sequence, although sequence conservation level among placental mammals (Miller et al. 2007) was not high in these regions. A low ratio of missense to synonymous substitutions was observed in dot-like Bs (0.95) compared to macro-Bs (2.50) and *A. flavicollis* Bs (1.68–1.71). Frameshift variants in *Kif23* gene were observed in both types of Bs.

Both morphotypes of B-chromosomes were enriched with LINE L1 elements. A mixture of *A. peninsulae* major centromeric repeats (clones APE_HindIII_7, APE_BstXI_S11, APE_BstXI_S23) (Matsubara et al. 2008) was found to be abundant in dot-like Bs and quite depleted in macro-Bs. The minor centromeric repeat APE_EcoRI_1 was detected in a very low copy number only in dot-like Bs. Satellite repeats, including MurSatRep1; ERVK, and ERVL (MaLR) LTRs; and gene families (*NLR*, *Vmn2r*) were spotted among smaller clusters.

Discussion

Comparison of B chromosomes in two *Apodemus* species

Our results clearly indicate that B chromosomes studied here are composed of amplified DNA from single or several genomic regions. The process of B chromosomes origin was independent in each species, although several common regions, including *Vrk1* (vaccinia related kinase 1) gene, were identified. Interestingly, copies of this gene were recently found on Bs of *A. flavicollis* in a simple sequence repeat profiling experiment (Bugarski-Stanojević et al. 2016). Margins of *Vrk1* region were not identical in two *Apodemus* species (Fig. 3), thus suggesting independent recruitment of the same genomic region to Bs. Enrichment with cell cycle-associated genes is characteristic for B chromosomes not only in *A. flavicollis*, but also in the gray brocket deer (Makunin et al. 2016), fox and raccoon dog (Graphodatsky et al. 2005; Yudkin et al. 2007; Duke Becker et al. 2011), and even grasshopper *Eyprepocnemis plorans* (Navarro-Domínguez et al. 2017). Receptor genes abundant in Bs of *A. peninsulae* were previously associated with evolutionary breakpoint regions of pig genome (Groenen et al. 2012). X-chromosomal region of *A. flavicollis* encompassing *Kdm6* gene is located close to the eutherian ancestral pseudoautosomal region (Raudsepp and Chowdhary 2015), indicating a potential functional connection between supernumerary and sex chromosomes. Taken together, these facts give evidence in favor of common mechanisms underlying various kinds of genomic instability.
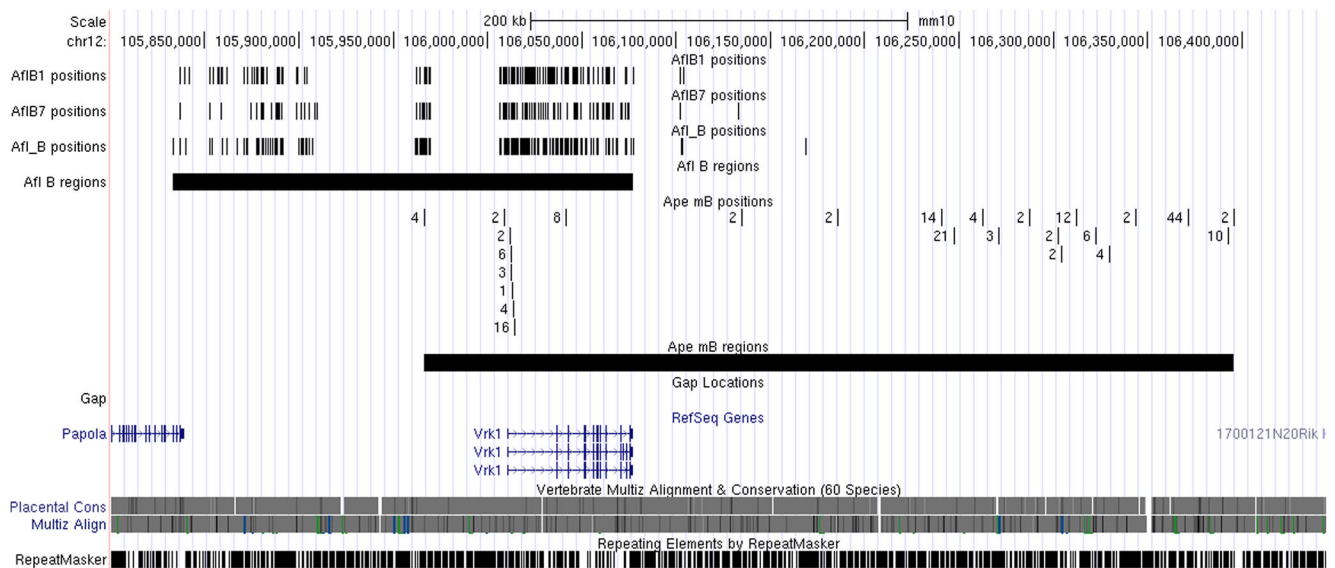
Fig. 3 ***Vrk1* gene region on B chromosomes of two *Apodemus* species.** AflB1, AflB7, Afl_B—B chromosomes from different *A. flavicollis* individuals, Ape_mB—macro-Bs of *A. peninsulae*. Reference genome—mouse (mm10 at UCSC Genome Browser)

Evolutionary path of B chromosomes in *Apodemus* species is shaped by the recombinational detachment from the main genome and lack of selective pressures. As a result, protein-coding genes accumulate missense and nonsense mutations and thus become pseudogenes. Accumulation of repetitive DNA, mostly satellites and endogenous retroelements, occurs in both species. Differential accumulation of centromeric repeats between Bs with different morphologies in *A. peninsulae* is in agreement with cytogenetics data (Rubtsov et al. 2004; Matsubara et al. 2008).

Here, we observed the intraspecific differentiation of B chromosomes in both *Apodemus* species. Only one cytological type of B chromosome has been detected in *A. flavicollis*, but on molecular level, we did observe some small structural differences between Bs from different populations. In *A. peninsulae*, two morphological types of Bs were slightly different, but studies involving other B chromosome morphologies and broader population sampling are yet to be done. Both loss and gain of the genetic material are possible during the evolution of B chromosomes (Houben et al. 2014), although some authors state that sequence degradation and loss are more common (Utsunomia et al. 2016). Unfortunately, the amount and the quality of variant calling data did not allow us to directly estimate a relative degeneration degree of the diverged regions against the conserved ones. However, in *A. peninsulae*, smaller dot-like Bs do not have any regions additional to those present on larger macro-Bs, which makes the degradation scenario more plausible.

Chromosome sequencing for clinical rearrangements

The marker chromosome analyzed in this study represents the most common type of human sSMC, an inverted duplication derived from the pericentromeric region of chromosome 15 long arm. In case of maternal inheritance, clinical phenotype of the inv dup(15)(q11-q13) is similar to the corresponding region duplication syndrome (OMIM: 608636). It is characterized by autism, mental retardation, ataxia, seizures, developmental delay, and behavior problems (Bundey et al. 1994; Burnside et al. 2011). As both breakpoints were located within alphoid DNA arrays, we suggest the origin of this sSMC results from the illegitimate recombination between the pericentromeric satellite array and the block at HSA15: 32.5–32.9 Mbp. The region involved in sSMC formation (15q11-q13) was previously identified with fluorescent in situ hybridization and array comparative genomic hybridization (Liehr et al. 2004). Our method gives additional detail on breakpoint coordinates, gene content, and, potentially, sequence variation, thus providing a framework to connect genetic content of the supernumeraries with the phenotype.

The similarity between small supernumerary chromosomes in human and B chromosomes found in some mammalian species was previously recognized (Liehr et al. 2008). Both types of chromosomes contain duplications of the host genome regions usually not exceeding several megabase pairs in size. Similar duplications occurring within the standard chromosome set are known as segmental duplications (Bailey et al. 2002). These duplications are sometimes additionally amplified, resulting in inverted duplication of the sSMC under study, and in multiple gene copies found in B chromosomes of diverse species (Duke Becker et al. 2011; Yoshida et al. 2011; Trifonov et al. 2013). Thus, sSMCs might correspond to an early stage of B chromosome evolution, and acquisition of drive mechanism for more efficient transmission could in principle transform those elements into true B chromosomes. Supernumeraries without disease phenotype and active centromeres are the best

candidates, as these are not eliminated by selection and are not so frequently lost during cell division (Liehr et al. 2008). However, sSMCs are usually derived from a single genomic region, while B chromosomes in many species, including *Apodemus*, are composed of multiple regions (Duke Becker et al. 2011; Yoshida et al. 2011; Valente et al. 2014; Makunin et al. 2016), which could have been acquired sequentially (Houben et al. 2014) and/or in a single catastrophic rearrangement event, such as chromothripsis (Stephens et al. 2011).

Here, we demonstrated that application of high-throughput sequencing allows for rapid characterization of chromosome genetic content. Previously, efforts for high-throughput sequencing of microdissected chromosomes were made in human (Weise et al. 2010), as well as in other species (Seifertova et al. 2013; Traut et al. 2013; Zlotina et al. 2016), but the analysis included only visual inspection or read counting. A more sophisticated sliding window coverage analysis was utilized to locate breakpoint margins in balanced translocations based on flow sorted human chromosome sequencing (Chen et al. 2008, 2010). Compared to this approach, microdissection-based method proposed here significantly reduces costs and experimental complexity. Besides, utilization of individual read positions rather than sliding windows at breakpoint identification step in principle provides higher resolution at lower sequencing coverage.

General concerns on single chromosome sequencing

Our results indicate that a single copy of a microdissected chromosome amplified with DOP-PCR or WGA kit produces sequencing results informative enough to characterize its genetic content in terms of both unique and repetitive sequences and, to some extent, to interpret specific sequence variants. With 0.5–6% reads having unique mapping positions in the reference genome, several hundred thousand reads are sufficient for the downstream analysis. Increasing the number of reads has diminishing returns, as sequencing exhausts the available DNA fragments (Fig. 4).

Not surprisingly, WGA kits perform better than DOP-PCR. However, over a half of the mapped DNA fragments were reused between the three biological samples of *A. flavicollis* B chromosomes (Table S5). At the same time, these fragments covered less than 10% of the underlying regions in the mouse genome. This can be explained by either a non-random DNA amplification at the initial stages of sample preparation (and thus intrinsic to the WGA kit), or a true lack of sequences that can be mapped to the gaps (e.g., because of the low homology or deletions in B chromosomes). The practical outcome is that the accuracy of the breakpoint mapping is bound to the distribution of recoverable positions and the minimum size of the detected regions is limited to at least 20 kbp (five read positions with one position per 5 kbp).

Non-random amplification, sequence divergence and repeat-rich regions result in false-positive deletions
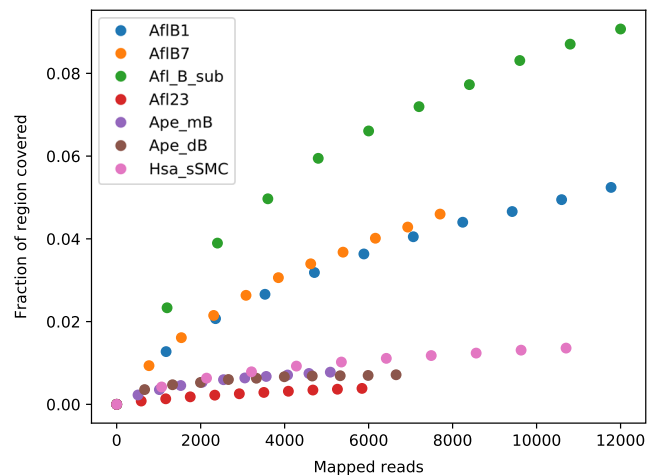

Fig. 4 **Sequence recovery from chromosome specific libraries against the number of mapped reads.** For Afl_B only a subset of reads was used: a total of 31,195 mapped reads allowed to recover 0.102 of the reference region sequence. Note that recovery depends on amplification protocol (WGA is better than DOP-PCR) and chromosomal region size (e.g., 3.4 Mbp for AFLBs vs 56.7 Mbp for AFL23)

manifesting as gaps in mapped read coverage (Makunin et al. 2016). In the worst case, those can be present at the margins of the target regions, as it was demonstrated for the human sSMC, where one of the breakpoints was located within the block of alpha-satellite repeats. In this case, inclusion of low-quality read mappings demonstrated that at least a part of this block can be present in the studied chromosome. Accounting for read mappability or utilization of higher sensitivity read aligners can be used to improve chromosomal region predictions.

Implementation of the tools and the pipeline for processing of single chromosome sequencing data in Galaxy not only makes the analysis more user-friendly and reliable. It also adds flexibility for the choice of the read processing and alignment software as well as easy re-run of the analysis with different sets of options and reference genomes. This allows Galaxy users to fine-tune the analysis to obtain better predictions of the chromosome genetic content.

Methods

Sampling

A total of five *Apodemus* animals were analyzed in this study (Table S1). Animals from Serbia (*A. flavicollis* 3727 and 3980) were collected using Longworth traps and treated according to the legal and ethical guidelines as indicated in Directive 2010/63/EU of the European Parliament and the Council of September 22, 2010 on the protection of animals used for scientific purposes. All animal procedures were approved by the Ethical Committee for the Use of Laboratory Animals of the Institute for Biological Research "Siniša Stanković,"

University of Belgrade. Capture, handling, and euthanasia of the mice from Russia (*A. flavicollis* 24985, *A. peninsulae* Sib04 and 01-09) followed protocols approved by the Animal Care and Use Committee of the Institute of Cytology and Genetics. No additional permits are required for research on non-listed species in Russia. Experiments described in this manuscript were carried out in accordance with the approved national guidelines for the care and use of animals.

Cell cultures and metaphase chromosome preparations

Fibroblast cell culture was obtained using collagenase/hyaluronidase treatment of tissues (Romanenko et al. 2015). All cell lines were deposited in the IMCB SB RAS cell bank ("The general collection of cell cultures," No. 0310-2016-0002). Metaphase chromosome preparations were performed as described previously for both cell cultures (Rubtsov et al. 2000; Rajičić et al. 2017) and bone marrow preparations (Baker et al. 1982).

Chromosome microdissection

Microdissection of single copies B-like chromosomes was performed according to (Yang et al. 2017). Amplification was performed either with DOP-PCR with MW-6 primer (CCGACTCGAGNNNNNNATGTGG) (Telenius et al. 1992; Rubtsov et al. 2000) or with a whole genome amplification kit (WGA Sigma-Aldrich, USA) according to the manufacturer's protocol (Table S1).

DNA sequencing

Sequencing of human sSMC was performed on Ion Torrent PGM. Sequencing of *Apodemus* B chromosomes was performed on Illumina MiSeq. Libraries were constructed with Illumina TruSeq v.2 kit (250 bp paired-end reads) for Afl_B, Ape_dB, and Ape_mB and Illumina TruSeq v.3 kit (300 bp paired-end reads) for AflB1, AflB5, AflB6, AflB7, and Afl23.

Target region identification

Regions of the reference genome present on the sampled chromosomes were identified with DOPseq_analyzer pipeline implemented as a set of Galaxy tools. Briefly, DOP or WGA primers and Illumina adapters were trimmed off with cutadapt, alignment to target (mouse mm10), and contamination (human hg19) genomes was performed with bowtie2 ("very-sensitive-local" option). Contaminant and low-quality (MAPQ < 20) alignments were discarded with contam_filter tool. Target regions present on chromosomes were identified based on changes in mean density of merged read positions with

region_dnacopy tool. Only regions including over five positions were subsequently analyzed.

Sequence variation analysis

Variant calling within target regions was performed with GATK HaplotypeCaller (McKenna et al. 2010); variants were annotated based on mm10 genome with snpEff (Cingolani et al. 2012). Variant calling and variant density calculations were automated by vca_reg pipeline of DOPseq_analyzer.

Repetitive DNA analysis

Trimmed reads (> 19 bp) were submitted to RepeatExplorer clustering and annotation pipeline (Novák et al. 2010, 2013) hosted on Galaxy-based IMCB server. Read similarities over 55% of the read length were interpreted as edges connecting the similar reads (nodes). Clusters of connected reads (each including > 0.1% of initial reads) were annotated with mammalian RepeatMasker database. Assembled contigs were aligned against NCBI nt database (August 8, 2016); clusters with best matches in Muridae or without any hits were treated as belonging to the sampled chromosome.

Compliance with ethical standards

References

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science 297(5583):1003–1007. https://doi.org/10.1126/science.1072047

Baker RJ, Haiduk MW, Robbins LW et al (1982) Chromosomal studies of South American bats and their systematic implications. Mamm Biol S Am 4:303–327

Bugarski-Stanojević V, Stamenković G, Blagojević J, Liehr T, Kosyakova N, Rajičić M, Vujošević M (2016) Exploring supernumeraries—a new marker for screening of B-chromosomes presence in the yellow necked mouse Apodemus flavicollis. PLoS One 11(8): e0160946. https://doi.org/10.1371/journal.pone.0160946

Bundey S, Hardy C, Vickers S, Kilpatrick MW, Corbett JA (1994) Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. Dev Med Child Neurol 36(8):736–742

Burnside RD, Pasion R, Mikhail FM, Carroll AJ, Robin NH, Youngs e, Gadi IK, Keitges E, Jaswaney VL, Papenhausen PR, Potluri VR, Risheg H, Rush B, Smith JL, Schwartz S, Tepperberg JH, Butler MG (2011) Microdeletion/microduplication of proximal 15q11. 2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. Hum Genet 130(4):517–528. https://doi.org/10.1007/s00439-011-0970-4

Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares IL, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser B, Vingron M, Ropers HH (2008) Mapping translocation breakpoints by next-generation sequencing. Genome Res 18(7):1143–1149. https://doi.org/10.1101/gr.076166.108

Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, Markus S, Wiedersberg E, Kistner G, van Ravenswaaij-Arts CMA, Kleefstra T, Kalscheuer VM, Ropers HH (2010) Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. Eur J Hum Genet EJHG 18(5): 539–543. https://doi.org/10.1038/ejhg.2009.211

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80–92. https://doi.org/10.4161/fly.19695

Duke Becker SE, Thomas R, Trifonov VA, Wayne RK, Graphodatsky AS, Breen M (2011) Anchoring the dog to its relatives reveals new evolutionary breakpoints across 11 species of the Canidae and provides new clues for the role of B chromosomes. Chromosom Res 19(6):685–708. https://doi.org/10.1007/s10577-011-9233-4

Graphodatsky AS, Kukekova AV, Yudkin DV, Trifonov VA, Vorobieva NV, Beklemisheva VR, Perelman PL, Graphodatskaya d, Trut LN, Yang F, Ferguson-Smith MA, Acland GM, Aguirre GD (2005) The proto-oncogene C-KIT maps to canid B-chromosomes. Chromosom Res 13(2):113–122. https://doi.org/10.1007/s10577-005-7474-9

Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens HJ, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, Bystrom M, Capitanu B, Carvalho-Silva D, Chardon P, Chen C, Cheng R, Choi SH, Chow W, Clark RC, Clee C, Crooijmans RPMA, Dawson HD, Dehais P, de Sapio F, Dibbits B, Drou N, du ZQ, Eversole K, Fadista J, Fairley S, Faraut T, Faulkner GJ, Fowler KE, Fredholm M, Fritz E, Gilbert JGR, Giuffra E, Gorodkin J, Griffin DK, Harrow JL, Hayward A, Howe K, Hu ZL, Humphray SJ, Hunt T, Hornshøj H, Jeon JT, Jern P, Jones M, Jurka J, Kanamori H, Kapetanovic R, Kim J, Kim JH, Kim KW, Kim TH, Larson G, Lee K, Lee KT, Leggett R, Lewin HA, Li Y, Liu W, Loveland JE, Lu Y, Lunney JK, Ma J, Madsen O, Mann K, Matthews L, McLaren S, Morozumi T, Murtaugh MP, Narayan J, Truong Nguyen D, Ni P, Oh SJ, Onteru S, Panitz F, Park EW, Park HS, Pascal G, Paudel Y, Perez-Enciso M, Ramirez-Gonzalez R, Reecy JM, Rodriguez-Zas S, Rohrer GA, Rund L, Sang Y, Schachtschneider K, Schraiber JG, Schwartz J, Scobie L, Scott C, Searle S, Servin B, Southey BR, Sperber G, Stadler P, Sweedler JV, Tafer H, Thomsen B, Wali R, Wang J, Wang J, White S, Xu X, Yerle M, Zhang G, Zhang J, Zhang J, Zhao S, Rogers J, Churcher C, Schook LB (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491(7424):393–398. https://doi.org/10.1038/nature11622

Houben A, Banaei-Moghaddam AM, Klemme S, Timmis JN (2014) Evolution and biology of supernumerary B chromosomes. Cell Mol Life Sci 71(3):467–478. https://doi.org/10.1007/s00018-013-1437-7

Jones RN (1991) B-chromosome drive. Am Nat 137(3):430–442. https://doi.org/10.1086/285175

Karamysheva TV, Andreenkova OV, Bochkaerev MN, Borissov YM, Bogdanchikova N, Borodin PM, Rubtsov NB (2002) B chromosomes of Korean field mouse Apodemus peninsulae (Rodentia, Murinae) analysed by microdissection and FISH. Cytogenet Genome Res 96(1-4):154–160. https://doi.org/10.1159/000063027

Kichigin IG, Giovannotti M, Makunin AI, Ng BL, Kabilov MR, Tupikin AE, Barucchi VC, Splendiani A, Ruggeri P, Rens W, O'Brien PCM, Ferguson-Smith MA, Graphodatsky AS, Trifonov VA (2016) Evolutionary dynamics of Anolis sex chromosomes revealed by sequencing of flow sorting-derived microchromosome-specific DNA. Mol Gen Genomics 291(5):1955–1966. https://doi.org/10.1007/s00438-016-1230-z

Kim J, Farré M, Auvil L, Capitanu B, Larkin DM, Ma J, Lewin HA (2017) Reconstruction and evolutionary history of eutherian chromosomes. Proc Natl Acad Sci 114(27):E5379–E5388. https://doi.org/10.1073/pnas.1702012114

Koepfli K-P, Paten B, O'Brien SJ (2015) The genome 10K project: a way forward. Annu Rev Anim Biosci 3(1):57–111. https://doi.org/10.1146/annurev-animal-090414-014900

Kour G, Kour B, Kaul S, Dhar MK (2009) Genetic and epigenetic instability of amplification-prone sequences of a novel B chromosome induced by tissue culture in Plantago lagopus L. Plant Cell Rep 28(12):1857–1867. https://doi.org/10.1007/s00299-009-0789-9

Lamatsch DK, Trifonov V, Schories S, Epplen JT, Schmid M, Schartl M (2011) Isolation of a cancer-associated microchromosome in the sperm-dependent parthenogen Poecilia formosa. Cytogenet Genome Res 135(2):135–142. https://doi.org/10.1159/000331271

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9(4):357–359. https://doi.org/10.1038/nmeth.1923

Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA (2009) Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res 19(5): 770–777. https://doi.org/10.1101/gr.086546.108

Liehr T, Claussen U, Starke H (2004) Small supernumerary marker chromosomes (sSMC) in humans. Cytogenet Genome Res 107(1-2):55–67. https://doi.org/10.1159/000079572

Liehr T, Mrasek K, Kosyakova N, Ogilvie C, Vermeesch J, Trifonov V, Rubtsov N (2008) Small supernumerary marker chromosomes

(sSMC) in humans; are there B chromosomes hidden among them. Mol Cytogenet 1(1):12. https://doi.org/10.1186/1755-8166-1-12

Liehr T, Weise A (2007) Frequency of small supernumerary marker chromosomes in prenatal, newborn, developmentally retarded and infertility diagnostics. Int J Mol Med 19(5):719–731

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, Grammes F, Grove H, Gjuvsland A, Walenz B, Hermansen RA, von Schalburg K, Rondeau EB, di Genova A, Samy JKA, Olav Vik J, Vigeland MD, Caler L, Grimholt U, Jentoft S, Inge Våge D, de Jong P, Moen T, Baranski M, Palti Y, Smith DR, Yorke JA, Nederbragt AJ, Tooming-Klunderud A, Jakobsen KS, Jiang X, Fan D, Hu Y, Liberles d, Vidal R, Iturra P, Jones SJM, Jonassen I, Maass A, Omholt SW, Davidson WS (2016) The Atlantic salmon genome provides insights into rediploidization. Nature 533(7602):200–205. https://doi.org/10.1038/nature17164

Makunin AI, Kichigin IG, Larkin DM, O'Brien PCM, Ferguson-Smith MA, Yang F, Proskuryakova AA, Vorobieva NV, Chernyaeva EN, O'Brien SJ, Graphodatsky AS, Trifonov VA (2016) Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. BMC Genomics 17(1):618. https://doi.org/10.1186/s12864-016-2933-6

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17(1):10–12. https://doi.org/10.14806/ej.17.1.200

Matsubara K, Nishida-Umehara C, Tsuchiya K, Nukaya D, Matsuda Y (2004) Karyotypic evolution of Apodemus (Muridae, Rodentia) inferred from comparative FISH analyses. Chromosom Res 12(4):383–395. https://doi.org/10.1023/B:CHRO.0000034103.05528.83

Matsubara K, Yamada K, Umemoto S, Tsuchiya K, Ikeda N, Nishida C, Chijiwa T, Moriwaki K, Matsuda Y (2008) Molecular cloning and characterization of the repetitive DNA sequences that comprise the constitutive heterochromatin of the A and B chromosomes of the Korean field mouse (Apodemus peninsulae, Muridae, Rodentia). Chromosom Res 16(7):1013–1026. https://doi.org/10.1007/s10577-008-1259-x

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303. https://doi.org/10.1101/gr.107524.110

Miller W, Rosenbloom K, Hardison RC, Hou M, Taylor J, Raney B, Burhans R, King DC, Baertsch R, Blankenberg D, Kosakovsky Pond SL, Nekrutenko A, Giardine B, Harris RS, Tyekucheva S, Diekhans M, Pringle TH, Murphy WJ, Lesk A, Weinstock GM, Lindblad-Toh K, Gibbs RA, Lander ES, Siepel A, Haussler D, Kent WJ (2007) 28-Way vertebrate alignment and conservation track in the UCSC Genome Browser. Genome Res 17(12):1797–1808. https://doi.org/10.1101/gr.6761107

Navarro-Domínguez B, Ruiz-Ruano FJ, Cabrero J, Corral JM, López-León MD, Sharbel TF, Camacho JPM (2017) Protein-coding genes in B chromosomes of the grasshopper Eyprepocnemis plorans. Sci Rep 7:45200. https://doi.org/10.1038/srep45200

Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11(1):378. https://doi.org/10.1186/1471-2105-11-378

Novák P, Neumann P, Pech J et al (2013) RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29(6):792–793. https://doi.org/10.1093/bioinformatics/btt054

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Oxf Engl 26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033

Rajičić M, Romanenko SA, Karamysheva TV, Blagojević J, Adnađević T, Budinski I, Bogdanov AS, Trifonov VA, Rubtsov NB, Vujošević

M (2017) The origin of B chromosomes in yellow-necked mice (Apodemus flavicollis)—break rules but keep playing the game. PLoS One 12(3):e0172704. https://doi.org/10.1371/journal.pone.0172704

Raudsepp T, Chowdhary BP (2015) The eutherian pseudoautosomal region. Cytogenet Genome Res 147(2-3):81–94. https://doi.org/10.1159/000443157

Romanenko SA, Biltueva LS, Serdyukova NA et al (2015) Segmental paleotetraploidy revealed in sterlet (Acipenser ruthenus) genome by chromosome painting. Mol Cytogenet 8:1

Rubtsov NB, Karamisheva TV, Astakhova NM, Liehr T, Claussen U, Zhdanova NS (2000) Zoo-FISH with region-specific paints for mink chromosome 5q: delineation of inter-and intrachromosomal rearrangements in human, pig, and fox. Cytogenet Genome Res 90(3-4):268–270. https://doi.org/10.1159/000056786

Rubtsov NB, Karamysheva TV, Andreenkova OV, Bochkaerev MN, Kartavtseva IV, Roslik GV, Borissov YM (2004) Comparative analysis of micro and macro B chromosomes in the Korean field mouse Apodemus peninsulae (Rodentia, Murinae) performed by chromosome microdissection and FISH. Cytogenet Genome Res 106(2-4):289–294. https://doi.org/10.1159/000079301

Seifertova E, Zimmerman LB, Gilchrist MJ, Macha J, Kubickova S, Cernohorska H, Zarsky V, Owens NDL, Sesay AK, Tlapakova T, Krylov V (2013) Efficient high-throughput sequencing of a laser microdissected chromosome arm. BMC Genomics 14(1):357. https://doi.org/10.1186/1471-2164-14-357

Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, van Heeringen SJ, Quigley I, Heinz S, Ogino H, Ochi H, Hellsten U, Lyons JB, Simakov O, Putnam N, Stites J, Kuroki Y, Tanaka T, Michiue T, Watanabe M, Bogdanovic O, Lister R, Georgiou G, Paranjpe SS, van Kruijsbergen I, Shu S, Carlson J, Kinoshita T, Ohta Y, Mawaribuchi S, Jenkins J, Grimwood J, Schmutz J, Mitros T, Mozaffari SV, Suzuki Y, Haramoto Y, Yamamoto TS, Takagi C, Heald R, Miller K, Haudenschild C, Kitzman J, Nakayama T, Izutsu Y, Robert J, Fortriede J, Burns K, Lotay V, Karimi K, Yasuoka Y, Dichmann DS, Flajnik MF, Houston DW, Shendure J, DuPasquier L, Vize PD, Zorn AM, Ito M, Marcotte EM, Wallingford JB, Ito Y, Asashima M, Ueno N, Matsuda Y, Veenstra GJC, Fujiyama A, Harland RM, Taira M, Rokhsar DS (2016) Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538(7625):336–343. https://doi.org/10.1038/nature19840

Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings l, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger l, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, Futreal PA, Campbell PJ (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144(1):27–40. https://doi.org/10.1016/j.cell.2010.11.055

Telenius H, Carter NP, Bebb CE, Nordenskjo¨ld M, Ponder BAJ, Tunnacliffe A (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13(3):718–725. https://doi.org/10.1016/0888-7543(92)90147-K

Traut W, Vogel H, Glöckner G, Hartmann E, Heckel DG (2013) High-throughput sequencing of a single chromosome: a moth W chromosome. Chromosom Res 21(5):491–505. https://doi.org/10.1007/s10577-013-9376-6

Trifonov VA, Dementyeva PV, Larkin DM et al (2013) Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (Capreolus pygargus). BMC Biol 11:1–11

Utsunomia R, de Andrade Silva DMZ, Ruiz-Ruano FJ et al (2016) Uncovering the ancestry of B chromosomes in Moenkhausia

sanctaefilomenae (Teleostei, Characidae). PLoS One 11(3): e0150573. https://doi.org/10.1371/journal.pone.0150573

Valente GT, Conte MA, Fantinatti BEA, Cabral-de-Mello DC, Carvalho RF, Vicari MR, Kocher TD, Martins C (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. Mol Biol Evol 31(8):2061–2072. https://doi.org/10.1093/molbev/msu148

Weise A, Timmermann B, Grabherr M, Werber M, Heyn P, Kosyakova N, Liehr T, Neitzel H, Konrat K, Bommer C, Dietrich C, Rajab A, Reinhardt R, Mundlos S, Lindner TH, Hoffmann K (2010) High-throughput sequencing of microdissected chromosomal regions. Eur J Hum Genet EJHG 18(4):457–462. https://doi.org/10.1038/ejhg.2009.196

Yang F, Trifonov V, Ng BL, et al (2017) Generation of paint probes from flow-sorted and microdissected chromosomes. In: Liehr T (ed) Fluorescence in situ Hybridization (FISH) Application Guide. Springer, Berlin, Heidelberg, pp 63–79

Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, Hirai H, Hirai Y, Matsuda Y, Okada N (2011) B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. PLoS Genet 7(8):e1002203. https://doi.org/10.1371/journal.pgen.1002203

Yudkin DV, Trifonov VA, Kukekova AV, Vorobieva NV, Rubtsova NV, Yang F, Acland GM, Ferguson-Smith MA, Graphodatsky AS (2007) Mapping of *KIT* adjacent sequences on canid autosomes and B chromosomes. Cytogenet Genome Res 116(1-2):100–103. https://doi.org/10.1159/000097424

Zlotina A, Kulikova T, Kosyakova N, Liehr T, Krasikova A (2016) Microdissection of lampbrush chromosomes as an approach for generation of locus-specific FISH-probes and samples for high-throughput sequencing. BMC Genomics 17(1):126. https://doi.org/10.1186/s12864-016-2437-4