

The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish *Astatotilapia latifasciata*, transcribes a potentially noncoding RNA

Érica Ramos¹ · Aduino L. Cardoso¹ · Judith Brown² · Diego F. Marques¹ · Bruno E. A. Fantinatti¹ · Diogo C. Cabral-de-Mello³ · Rogério A. Oliveira⁴ · Rachel J. O'Neill⁵ · Cesar Martins¹

Received: 8 February 2016 / Revised: 3 April 2016 / Accepted: 3 May 2016 / Published online: 12 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Supernumerary chromosomes have been studied in many species of eukaryotes, including the cichlid fish, *Astatotilapia latifasciata*. However, there are many unanswered questions about the maintenance, inheritance, and functional aspects of supernumerary chromosomes. The cichlid family has been highlighted as a model for evolutionary studies, including those that focus on mechanisms of chromosome evolution. Individuals of *A. latifasciata* are known to carry up to two B heterochromatic isochromosomes that are enriched in repetitive DNA and contain few intact gene sequences. We isolated and characterized a transcriptionally active repeated DNA, called B chromosome noncoding DNA (BncDNA), highly represented across all B chromosomes of *A. latifasciata*. BncDNA transcripts are differentially processed among six different tissues, including the production of smaller transcripts, indicating transcriptional variation may be linked to B chromosome presence and sexual

phenotype. The transcript lengths and lack of similarity with known protein/gene sequences indicate BncRNA might represent a novel long noncoding RNA family (lncRNA). The potential for interaction between BncRNA and known miRNAs were computationally predicted, resulting in the identification of possible binding of this sequence in upregulated miRNAs related to the presence of B chromosomes. In conclusion, Bnc is a transcriptionally active repetitive DNA enriched in B chromosomes with potential action over B chromosome maintenance in somatic cells and meiotic drive in gametic cells.

Keywords Cytogenomics · Repetitive element · Evolution · Supernumerary chromosome

Introduction

B chromosomes (Bs) are extra, or supernumerary, chromosomes that have been described in 10–15 % of karyotyped species of all major eukaryote groups (Camacho 2005), including cichlid fishes. Cichlids are an established model for evolutionary studies due to their fast adaptive radiation in East Africa (Kocher 2004) and have also become a model for the study of B chromosomes. Notably, Bs have been identified in 7 South American (Feldberg and Bertollo 1984; Martins-Santos et al. 1995; Feldberg et al. 2004) and 14 African (Poletto et al. 2010a, b; Yoshida et al. 2011; Kuroiwa et al. 2014) species of cichlids. *Astatotilapia latifasciata* is an African cichlid fish found in Lake Nawampasa (satellite lake of Lake Kyoga, part of Lake Victoria system) and is a widely distributed species in the aquarium hobbyist trade (Skelton 2001). Previous studies identified the presence of either 1 or 2B chromosomes in both

Electronic supplementary material The online version of this article (doi:10.1007/s00412-016-0601-x) contains supplementary material, which is available to authorized users.

✉ Cesar Martins
cmartins@ibb.unesp.br

¹ Department of Morphology, Institute of Biosciences, Sao Paulo State University, 18618-689 Botucatu, SP, Brazil

² Allied Health Sciences Department and Institute for Systems Genomics, University of Connecticut, 06269 Storrs, CT, USA

³ Department of Biology, Institute of Biosciences, Sao Paulo State University, 13506-900 Rio Claro, SP, Brazil

⁴ Department of Biostatistics, Institute of Biosciences, Sao Paulo State University, 18618-689 Botucatu, SP, Brazil

⁵ Department of Molecular and Cell Biology and Institute for Systems Genomics, University of Connecticut, 06269 Storrs, CT, USA

sexes of *A. latifasciata* aquarium populations and revealed that these chromosomes are heterochromatic and contain highly repeated sequences. Sequences such as ribosomal DNA and diverse repetitive elements were cytogenetically characterized on the B chromosomes (Poletto et al. 2010a, b; Fantinatti et al. 2011). A high-throughput sequencing approach was applied to investigate the B chromosomes of *A. latifasciata*, showing that B is a genomic mosaic composed of thousands of degenerate sequences, including genes and transposable elements, derived from most of the A (autosomal) chromosomes. Among the genes and pseudogenes detected, few of them are intact (Valente et al. 2014).

B chromosomes do not follow the Mendelian laws of inheritance and are considered “dispensable” chromosomes because they are not responsible for essential life processes (White 1973; Jones 1991; Camacho et al. 2000; Palestis et al. 2004; Jones et al. 2008). The “selfish DNA theory” has been proposed as one possible explanation for the maintenance of B chromosomes in host genomes (Burt and Trivers 2006). Bs can be found in different numbers among and within individuals of a population. In high number, extra chromosomes lead to deleterious effects due to unbalanced gene dosage (Jones and Rees 1982; Bougourd and Jones 1997; Carlson 2009) and can contribute to phenotypic variation (for review, Houben et al. 2013). It has been proposed that proto-B chromosomes are derived from events of duplication and translocation from the subregions of the A chromosomes and subsequently become further enriched in A chromosome and organelle DNA sequences through transposition events (Houben et al. 2013; Valente et al. 2014). As a consequence, they have accumulated large blocks of repetitive DNA sequences (Bueno et al. 2013), organelle sequences (Martis et al. 2012), a few genes (Miao et al. 1991; Graphodatsky et al. 2005; Valente et al. 2014; Rajcic et al. 2015), and pseudogenes (Banaei-Moghaddam et al. 2013). Additionally, functional studies have demonstrated gene and pseudogene expression from B chromosomes (Leach et al. 2005; Ruiz-Estévez et al. 2012; Banaei-Moghaddam et al. 2013; Trifonov et al. 2013; Lin et al. 2014), some of which carry direct effects over transcriptional activity of autosomal genes (Carchilan et al. 2009).

In recent years, functional studies have focused on noncoding sequences, or “noncoding” DNA (ncDNA), which comprise the largest proportion of eukaryotic genomes (Ahnert et al. 2008). Noncoding RNA sequences are now known to play a role in gene regulation and genomic organization dynamics (Sandelin et al. 2004; Tian et al. 2015). Included in this class are the ribosomal RNAs (rRNAs), transporter RNAs (tRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), small interfering RNAs (siRNAs), microRNAs (miRNAs), and long noncoding RNAs (lncRNAs) (Eddy 2001; Mattick and Makunin 2006). One

of the most intriguing classes of ncRNA is the lncRNAs, which has been conventionally described as a group of sequences that transcribes ncRNAs longer than 200 bp in length (Kapranov et al. 2007; Wilusz et al. 2009; Wang and Chang 2011). Indeed, new subclassifications have been proposed for lncRNAs considering RNA function and association with subcellular structures and other features (Laurent et al. 2015). Many studies have demonstrated the involvement of lncRNAs in processes such as cancer metastasis (Gupta et al. 2010), apoptosis in breast cancer cells (Pickard and Williams 2014), X chromosome silencing (Brockdorff et al. 1992; Brown et al. 1992; McHugh et al. 2015), adaptive immunity in mammals (Pang et al. 2009), development (Standaert et al. 2014), and heart hypertrophy protection (Han et al. 2014). In fact, lncRNAs are traditionally considered poorly conserved (Pang et al. 2006), but recent studies have shown the presence of mammalian lncRNA homologs in fish species, indicating a possible conservation of these elements in different groups (Basu et al. 2013). Furthermore, a lack of identifiable sequence conservation as a consequence of rapid sequence evolution does not necessarily indicate a lack of functional conservation in vertebrates (Ulitsky et al. 2011).

We have identified a transcriptionally active repetitive element called B chromosome noncoding DNA element (BncDNA) enriched over the entirety of the B chromosome in the cichlid *A. latifasciata*. This element is expressed in at least six tissues (heart, brain, muscle, liver and eye) of *A. latifasciata*, producing different transcript sizes (varying from ~200 to ~500 bp). BncRNA transcriptional levels vary among tissues and are affected by B chromosome presence and sexual phenotype. Thus, Bnc represents a candidate lncRNA with potential contribution to B chromosome maintenance and drive mechanisms.

Materials and methods

Animal samples and genotyping

Specimens (total of 95) of *A. latifasciata* were obtained from a stock establishment from the trade and maintained in the fish facility of the Integrative Genomics Laboratory at Sao Paulo State University (Botucatu, Brazil). Tissue samples and chromosome preparations were obtained for animals without (0B), carrying one (1B), and carrying two (2B) B chromosomes, including individuals of each sex. DNA and RNA were obtained from these specimens according to the protocol of Sambrook and Russel (2001). All procedures are in accordance with ethical principles of the Brazilian College of Animal Experimentation and were approved by Institute of Biosciences/São Paulo State University (UNESP) ethics committee on use of animals (Protocol no. 486–2013). Genotyping for the B chromosome presence/absence was

conducted using a polymerase chain reaction (PCR) marker previously developed (Fantinatti et al., in preparation) that amplifies DNA fragments carrying B-specific mutations. The number of B chromosomes (one or two B chromosomes) in each animal was determined previously using classical cytogenetic metaphase chromosome procedures and a Giemsa stain. In the case of RNA samples used for reverse transcription real-time polymerase chain reaction (RT-qPCR) and Northern blot assays, we were not able to perform cytogenetic metaphase analyses without impacting RNA expression profiles. In these cases, we referred to the absence of B chromosome as “B−” and the presence of B as “B+” as determined below without delineating the B copy number.

Genomic approach

The presence of ribosomal DNA (rDNA) sequences in the B chromosome of *A. latifasciata* (Poletto et al. 2010a; Fantinatti et al. 2011) motivated the design of several sets of primers to further define the organization of rRNA genes in the 0B, 1B, and 2B genomic content. PCR fragments were cloned into pGEM-T plasmid vectors (Promega) and the recombinants obtained were subjected to Sanger sequencing. The sequences obtained (accession numbers KX226399–KX226404) were subject to basic searches for any similarity using NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, Altschul et al. 1990) and Bouilabase (<http://www.bouilabase.org>) BLAST tools, as well as repeat masker (<http://repeatmasker.org>; Smit et al. 2013) in Repbase (www.girinst.org/repbase; Jurka et al. 2005). Among all characterized rRNA gene sequences, one set of primers, RAND_1 (Supplementary Table 1), amplified a non-rDNA sequence (probably by unspecific annealing of primers during the PCR reaction), which was further characterized (BncDNA).

The BncDNA sequence was subject to analysis using BLAST against cichlid fish genomes (Bouilabase) and the *A. latifasciata* first draft reference genome (Jehangir et al., in preparation). The BncDNA regions identified in the *A. latifasciata* genome were isolated and aligned against the short-read datasets of 0B and 2B individuals with sequenced genomes from the species available at Sacibase (sacibase.ibb.unesp.br) using Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>; Langmead and Salzberg 2012) implementing fast parameter options; from these alignments, B-specific nucleotide polymorphisms were assessed. The minimum depth of coverage to consider point mutations in the polymorphism analysis was 30, based on the mean coverage of the genome. Furthermore, Southern blot analyses were performed using the full BncDNA sequence probed against digested genomic DNA using the restriction enzymes *EcoRI*, *BclI*, *HphI*, and *HaeIII* to compare the banding pattern obtained and infer differences in genomic organization of BncDNA among B+ and B− samples.

Digestions were conducted with *EcoRI* alone and a mix of *BclI*, *HphI*, and *HaeIII* in the same reaction.

In an attempt to compare the relative copy number of BncDNA sequence among tissues and B groups (B− and B+ individuals), we performed quantitative PCR (qPCR) using the StepOnePlus™ Real-Time PCR System (Applied Biosystems, USA) and SYBR Green PCR master mix kit (Applied Biosystems, USA) in biological triplicates of each group (B−/+, male/female) and technical duplicates. Several sets of primers were designed in four regions that span the entire BncDNA sequence: Bnc_1, Bnc_2, Bnc_3, and Bnc_4 (Supplementary Table 1 and Supplementary Fig. 1). The data was normalized to a reference gene, *hypoxanthine-guanine phosphoribosyltransferase* (*HPRT*), using Q-Gene software (Muller et al. 2002; Simon 2003). All qPCR results of this work were analyzed using the statistical approach of a generalized linear model (GLM) with the averages adjusted by asymmetric distribution Gamma to all variables implemented in SAS version 9.3, procedure GENMOD. The analyzed variables were sex, B chromosome presence, or both. The significant results were selected based on maximum *p* value of 0.05.

BncDNA was also characterized by fluorescence in situ hybridization (FISH) mapping to chromosome spreads in all groups according to Pinkel et al. (1986) with modifications (Cabral-de-Mello et al. 2012). We colocalized BncDNA, labeled with digoxigenin-11-dUTP and detected using an anti-digoxigenin-rhodamine antibody (Roche), with 18S rDNA, labeled with biotin-14-dATP and detected by an avidin–fluorescein isothiocyanate (FITC) conjugate (Sigma-Aldrich), as the latter is a marker sequence for B chromosome identification (Poletto et al. 2010a).

Functional approach

Blastn searches against cichlid fish transcriptomes (<http://www.bouilabase.org>), microRNAs (www.mirbase.org; Griffiths-Jones 2004), RNA families (<http://rfam.xfam.org>; Nawrocki et al. 2014), and long ncRNA (<http://www.lncrnadb.org>; Quek et al. 2014) databases were performed for the BncDNA sequence. BncDNA was further analyzed for open reading frames (ORFs) using Geneious Pro 4.8.5 software (<http://www.geneious.com>; Kearse et al. 2012) and submitted to HMMER (hidden Markov models, www.hmmerr.janelia.org; Finn et al. 2011), pfam (www.pfam.xfam.org; Finn et al. 2014) and UniProt (<http://www.uniprot.org/blast>; The UniProt Consortium 2015); in addition, gene predictions were performed using FGENESH at softberry (<http://www.softberry.com>; Solovyev et al. 2006).

Based on the identification of transcripts of BncDNA in cichlid transcriptomes, tissues (gonads, brain, eye, muscle, heart, and liver) from B− and B+ male/female of *A. latifasciata* were selected to assess BncDNA transcript

lengths and transcription levels using Northern blot and RT-qPCR, respectively. Total RNA extraction of tissue samples was performed using Trizol Reagent (Invitrogen Ambion, USA), and RNA concentration and quality was determined with NanoVue Plus Spectrophotometer (GE Healthcare, USA) and 2100 Bioanalyzer (Agilent, EUA). RNA extractions were performed from animals collected at the same period of the day from individuals of the same parental cross to limit RNA expression variation due to environmental factors or population variations. All RNA samples were treated using DNase (Thermo Fisher Scientific, USA) to exclude DNA contamination. To verify BncDNA transcript presence, we designed two sets of primers called BncRT1 and BncRT2 (Supplementary Table 1) based in the sequence regions with more hits against transcripts in blastn analysis in the Bouillabase transcriptomes of cichlids. To trace the tissues that express BncDNA transcripts and to define the complete size of transcripts produced, we developed Northern blot assays for large RNAs (>200 bp) and small RNAs (<200 bp) separately as per (Carone et al. 2009). The probe and the hybridization were the same described for the Southern blot analysis.

Moreover, we performed blastn searches of BncDNA in *A. latifasciata* transcriptomes (Marques et al., in preparation) in order to identify BncDNA transcripts and isoforms. Additionally, the transcriptome reads were aligned against the *A. latifasciata* assembled genome implementing very fast parameter options and a search for B-specific nucleotide polymorphisms was conducted based on the nucleotide polymorphisms genome dataset reference. Relative BncDNA transcriptional levels were determined via reverse transcription quantitative PCR (RT-qPCR) in biological triplicates of each group and technical duplicates, as above, using the same sets of primers that span the BncDNA sequence (Bnc_1, Bnc_2, Bnc_3 and Bnc_4) (Supplementary Table 1 and Supplementary Fig. 1) and normalizing expression values to the housekeeping gene *ubiquitin-conjugating enzyme (UBCE)* using Q-Gene software. *UBCE* was selected as a reference from comparison of expression levels among various sets of genes using RefFinder software (<http://www.leonxie.com/referencegene.php>).

To determine if BncDNA transcripts (assembled transcripts of *A. latifasciata* transcriptomes) were capable of forming secondary structures, we applied single-stranded nucleic acid secondary structure prediction using RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>; Lorenz et al. 2011) and Genebee (http://www.geneBee.msu.su/services/rna2_reduced.html; Brodsky et al. 1995). Additionally, we predicted interactions of BncRNA with differentially expressed microRNAs of B+ samples of *A. latifasciata* microRNomes obtained from brain, muscle, and gonads (Fantinatti et al., in preparation) using PITA (http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html; Kertesz et al.

2007) and miRanda (<http://www.microna.org/microna/getGeneForm.do>; Enright et al. 2003) software. The predicted interaction results were filtered using a maximum of −20 kcal free energy (for miRanda) and −10 kcal free energy (for PITA).

Results

Genomic features

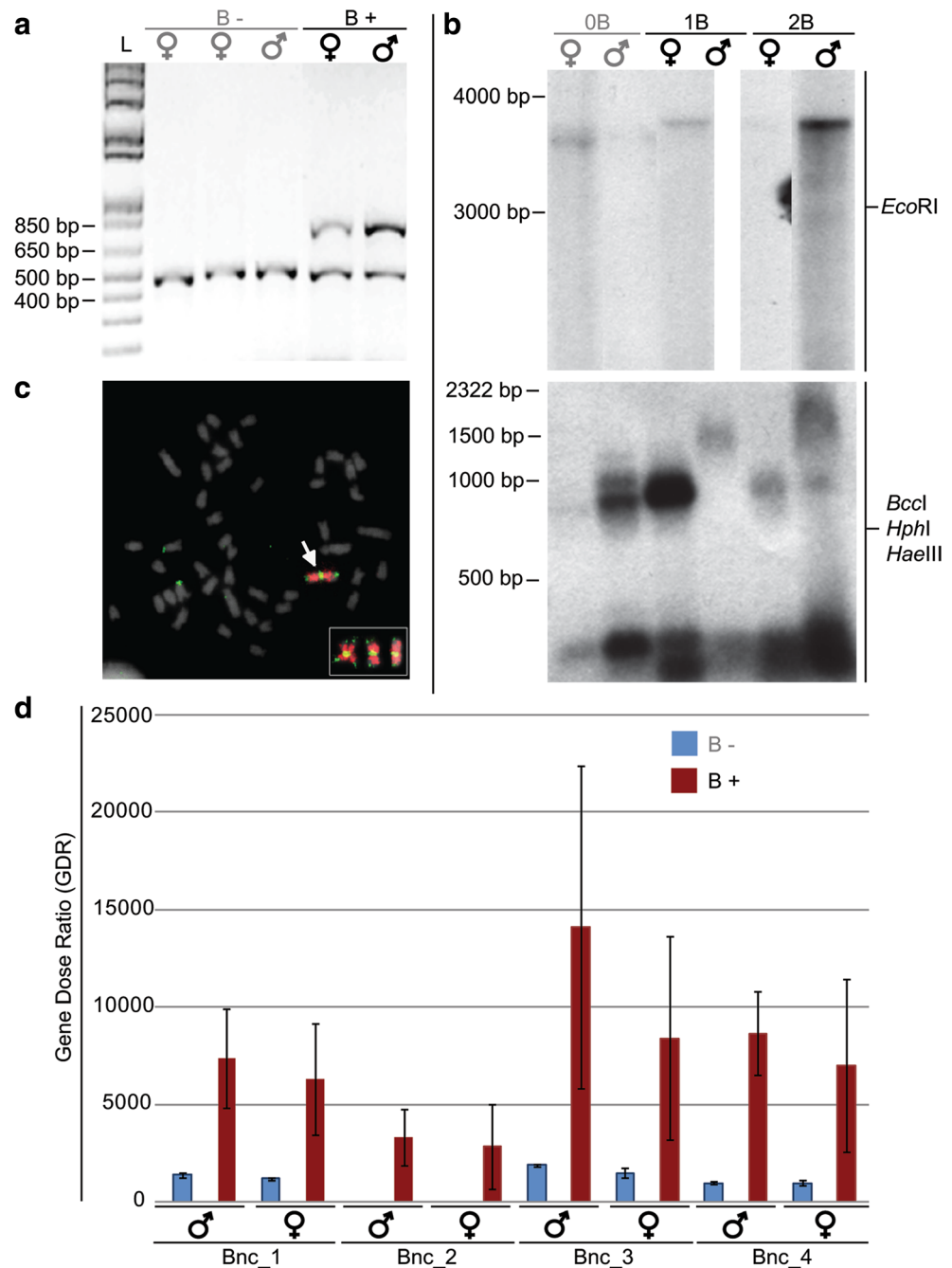
In the present study, we obtained PCR products from RAND_1 primers that showed a different banding pattern between B− and B+ samples, with a specific B+ fragment of approximately 700 bp (Fig. 1a) named BncDNA. Considering our interest in B chromosomes, this B+–specific band was cloned, sequenced and its nucleotide sequence (accession number KX196446) submitted to searches against the NCBI nonredundant nucleotide database, Repeat Masker, pfam, and UniProt databases. However, these searches failed to find significant sequence similarity with any previously characterized sequence.

Southern blot analysis confirmed our initial PCR results, showing discordant banding patterns between B+ and B− individuals (Fig. 1b and Supplementary Fig. 2a) and validating the presence of this novel sequence in this cichlid species. *EcoRI* digestions of genomic DNA generated bands between 3000 and 4000 bp for B− samples and slightly larger bands for B+ samples (both 1B and 2B). Using a mix of the enzymes *BccI*, *HphI*, and *HaeIII* in the same digestion that cut inside BncDNA sequence, we observe two bands under 500 bp for B+ individuals and only one band for B− individuals, of approximately the same size (Fig. 1b, bottom panel and Supplementary Fig. 2a). Moreover, a band ~1000 bp is observed in all groups, except in the 1B male. Additional hybridizing bands above 1000 bp and under 1500 bp were observed in males, but are absent in females. These additional bands are smaller (1000 bp) in 0B individuals than in 1 or 2B individuals (1500 bp).

FISH mapping confirmed the presence of abundant BncDNA copies in the B chromosome of males and females (Fig. 1c), with the element occupying almost the entire extent of this chromosome, with the exception of centromeres and telomeres. In contrast, the A chromosome set did not present any visible FISH signal. The results of qPCR of the four BncDNA regions (Fig. 1d) showed that the B+ group carries more copies of the entire BncDNA sequence when compared to the B− group, corroborating the Southern blot and FISH results.

In a BLAST search against other cichlid genomes, we identified genomic regions that shared partial identity with BncDNA among *Metriaclima zebra*, *Astatotilapia burtoni*, and *Pundamilia nyererei* genomes (Supplementary Fig. 1),

Fig. 1 Genomic features of BncDNA. **a** B+ specimens are distinguished from B- specimens by the presence of an additional amplicon at ~700 bp in PCR analysis. **b** Southern blot of DNA probed with Bnc sequence (bands pasted due to sample organization and missing data for some groups, see original image in Supplementary Fig. 1). **c** Fluorescence in situ hybridization on metaphase chromosomes of a 1B individual (gray DAPI) shows that BncDNA (red) is juxtaposed to the 18S rDNA (green) on the B chromosome, white arrow indicates B chromosome. **d** Genomic quantification by qPCR analysis showing the relative abundance of BncDNA among B+ samples, *y*-axis shows relative gene dose ratio (GDR), *x*-axis indicates the different Bnc primer sets. *L*, molecular weight marker; *B*-, specimens without B chromosome; *B*+, specimens with B chromosome; *0B*, specimens without B chromosome; *1B*, specimens carrying one B chromosome; *2B*, specimens carrying two B chromosomes



indicating the presence of this element in all cichlid fish species with available genome sequences. Furthermore, we have identified BncDNA in different regions of the *A. latifasciata* genome assembly, indicating Bnc is spread across the genome but remains fragmented (Supplementary Fig. 1). The search for genomic point nucleotide variation (Supplementary Table 2) in BncDNA regions of the *A. latifasciata* assembled genome revealed two categories of variation: nucleotide polymorphisms found exclusively in B+ samples and polymorphisms enriched in B+ samples (found with at least double the frequency over B- samples). The genomic regions

containing BncDNA in *M. zebra*, *A. burtoni*, *P. nyererei*, and *A. latifasciata* were aligned (Supplementary Fig. 3), indicating variable organization of BncDNA regions among these species.

BncRNA transcript features

Database searches failed to find similarity of BncDNA to any well-characterized DNA sequence, gene, protein, or RNA sequence. WU-Blasts against transcriptomes of *P. nyererei*, *A. burtoni*, *M. zebra*, and *A. latifasciata* (Supplementary

Fig. 1) showed transcripts from different tissues (embryo, gonads, heart, brain, among others) with high identity to different regions of BncDNA transcripts. From this analysis, two longer and four smaller regions are predominantly represented in these transcriptomes. To validate these transcripts, we performed RT-PCR using primers designed for BncDNA (Supplementary Table 1). RT-PCR showed variable patterns of bands across different tissues of *A. latifasciata* (Supplementary Fig. 4). Sanger sequencing did not show any nucleotide differences among the fragments obtained for B+ and B- samples (accession numbers KX226408-KX226412).

The results from *A. latifasciata* reference transcriptome blastn allowed the identification of three isoforms, 322, 414, and 680 bp in length (Supplementary Fig. 5, accession numbers KX226405-KX226407). When submitted to BLASTN against known sequence databases, we could find identity to transcribed and non-characterized sequences in other cichlid fishes. The B-type genomic nucleotide polymorphisms (Supplementary Table 2) were not detected within the transcript reads, showing that those regions carrying mutations, enriched or exclusive of the B chromosome, are likely not transcribed.

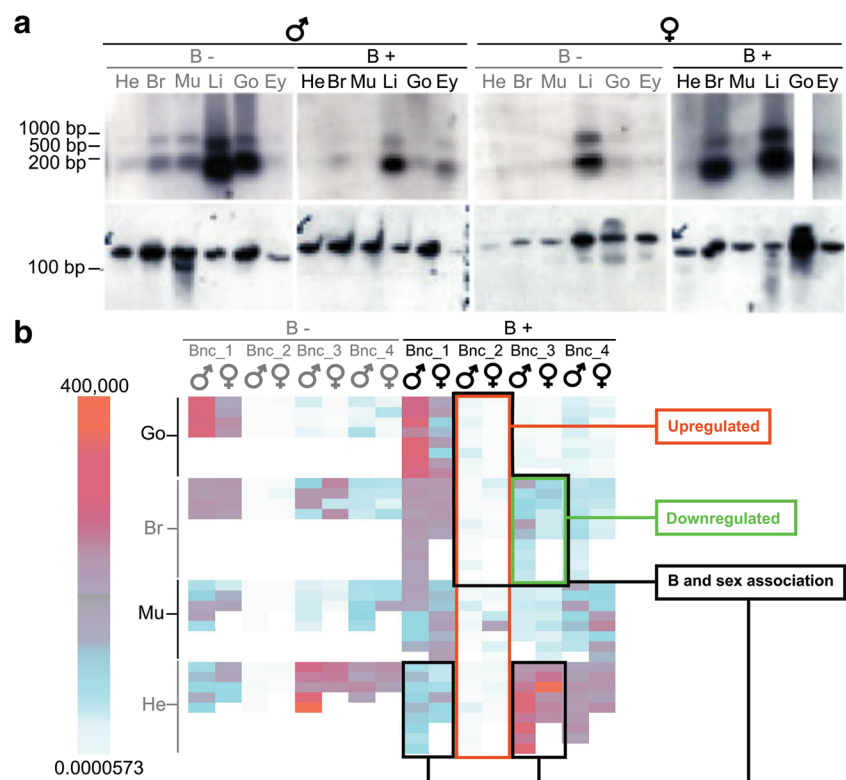
Six tissues (brain, gonads, heart, muscle, liver, and eye) were subjected to Northern blot analysis to determine the size of Bnc transcripts (Fig. 2a and Supplementary Fig. 2b). The results confirmed BncDNA is transcribed in all analyzed tissues and revealed the approximate transcript sizes common to all groups (B+ and B- male and female samples) of 200 and 500 bp. For large RNAs, some samples failed to

show hybridizing bands, likely due to technical limitations in the quantity or quality of RNA used. Furthermore, small transcripts were detected via small RNA Northern blot with a discrete fragment of <100 bp in B- female (liver, gonads, and eye), B- male (muscle), and B+ female (liver and gonads).

The transcriptional levels of BncDNA were investigated in different tissues using comparative RT-qPCR across four BncDNA subregions (Fig. 2b and Supplementary Table 3). We verified statistically significant differences in the analyzed tissues and regions and correlated these data to B chromosome presence and phenotypic sex. Significant differences were found in all tissues, wherein BncDNA region 2 was upregulated in all B+ individuals. Moreover, brain tissue showed downregulation of region 3 in B+ samples. When analyzing both characteristics (sex and B presence) together, we have identified heart (regions 1 and 3), brain (regions 2 and 3), and gonad (region 2) as showing differences in expression levels somewhat related to these two features. Muscle (region 3) and gonad (regions 1 and 3) samples from males were upregulated without any correlation to B chromosome presence. Region 4 did not present any differential expression, showing that it may be a canonical transcript without impact by B chromosome presence; in other words, region 4 may represent an A chromosome transcript.

Although BLASTX, BLASTN, and FGENSESH searches did not reveal any similarity to known or predicted proteins, the BncDNA sequence ORF prediction showed six ORFs of

Fig. 2 BncDNA functional features. **a** Large and small RNA Northern blots in different tissues and specimens indicating the transcripts length of BncRNA (bands pasted due to sample organization and missing data to some groups, see original image in Supplementary Fig. 1). **b** Heat map of relative RNA expression levels from RT-qPCR analysis using different set of primers (Bnc_1, Bnc_2, Bnc_3 and Bnc_4) showing the up and downregulation of BncRNA regions in different tissues and regions correlated to B chromosome presence and sexual phenotype; B-, specimens without B chromosome; B+, specimens with B chromosome; Go, gonads; Br, brain; He, heart; Mu, muscle; Li, liver; Ey, eye; white squares in heat map means missing data



167, 63, 42, 41, 34, and 34 amino acids (aa). These amino acid sequences were submitted to BLASTX and BLASTN at pfam and rfam databases, respectively, and did not show any identity to known protein family domains nor RNA families. Secondary structure of the full BncRNA sequence was predicted and it generated five different kinds of secondary structures (Supplementary Fig. 6), pointing to a possible fold of the longest transcript and further interaction with other molecules. From this point, we predicted possible binding sites in BncDNA transcripts using differentially expressed (up and downregulated) microRNAs in B+ *A. latifasciata* microRNomes. We identified six binding sites in BncRNA isoforms to three different miRNAs: miR-129-3p, miR-9-5p, and miR-153a-5p (Supplementary Table 4). All miRNAs were upregulated in female B+ muscle tissue.

Discussion

Genomic organization of BncDNA element

*Eco*R1 digested genomic DNA assessed by Southern blot indicates a unique banding pattern likely represent the entire BncDNA cluster of copies. These bands, however, differ in size between B+ and B− samples, suggesting that the BncDNA array has a large number of copies on B chromosomes. The absence of multimeric units in Southern blot analyses suggests BncDNA is not tandemly arrayed in the genome (neither B− nor B+). Southern band patterns obtained under *Bcl*I, *Hph*I, and *Hae*III combinatorial digestions corroborate that BncDNA is present on both A and B chromosomes and indicates that BncDNA copies are differently organized in the B+ genomes. Although there are clear differences in the band intensity among B+ and B−-delineated samples and individual sex, this Southern hybridization analysis was not quantitative and thus could not be used to infer copy number of BncDNA. Rather, the quantification of BncDNA copies was obtained by qPCR, a more precise and reliable method of quantification.

FISH results confirmed the enrichment of BncDNA in B chromosomes and showed no localization of BncDNA in B− individuals. This lack of signal on A chromosomes in these individuals is likely due to the FISH technique resolution employed herein, since BncDNA is present in a lower number of clustered copies in 0B individuals as determined by other methods. Notably, the relative copy number quantification using qPCR corroborated the BncDNA enrichment in the B+ samples. The wide standard deviation values obtained in the qPCR analyses for the B+ samples are likely indicative of the presence of both 1B and 2B animals in the same group (defined simply as B− or B+) and/or an extensive variation in copy number of BncDNA among individuals with the same number of B chromosomes.

Searches in fish genomes, including *A. latifasciata*, showed BncDNA is present in different genomic regions of a variety of cichlids in a fragmented arrangement. This fragmentation is probably a consequence of the assembly bias expected for repetitive DNA regions of eukaryotic genomes (Treangen and Salzberg 2011). In addition, our polymorphism analysis identified both exclusive and enriched B+ nucleotide variation, likely a consequence of the low selective pressure on B chromosomes and the repetitive nature of BncDNA. We hypothesize that BncDNA was among the founder sequences that defined the proto-B chromosome, or migrated later to the B, and underwent successive events of duplication and amplification. The B chromosome of *A. latifasciata* contains thousands of sequences duplicated from essentially every chromosome within the karyotype, including transposable elements and genes (Valente et al. 2014). BncDNA represents a newly discovered repetitive element over represented in the B chromosome genomic content.

Characterization of BncDNA transcripts

We detected BncRNA in all analyzed tissues, yet the observed differing transcript sizes could be due to differential processing or non-canonical transcription of BncDNA. Furthermore, analyses via Northern blots detected final transcript sizes of 200–500 bp, without apparent differences between B−/B+ individuals. However, different patterns of hybridizing bands were found for BncDNA transcripts in the small RNA pools, but the resolution of these bands was not of sufficient quality to delineate differences between B+ and B− individuals, nor male vs female groups. Alternatively, RT-qPCR results indicate BncDNA is transcribed differentially across tissues and among individual groups. Considering that we found different transcription levels between the BncRNA regions (1–4) in the same tissue sample, as well as across different tissues, we infer that its transcripts are highly processed. B chromosome presence affects BncDNA transcription in all analyzed tissues, often correlated with phenotypic sex (muscle was an exception), although there is no evidence for transcription of BncDNA that is exclusive to B chromosomes. One possible explanation for the observed variation in expression levels correlated with B presence could be that transcription of BncDNA canonical copies occurs from both A and B chromosomes. Alternatively, transcription of B chromosome copies indirectly influences the expression of A copies, as demonstrated in rye (Carchilan et al. 2009).

BncRNA differential expression correlated with phenotypic sex, coupled with its association to B chromosomes, reinforces previous evidence supporting a role for B chromosomes in sex chromosome evolution in several organisms. B chromosomes have contributed to the origin of the *Drosophila* Y chromosome (Carvalho 2002; Nokkala et al. 2003) and appear to be under some form of selection that has led to

maintenance of B chromosomes within *Drosophila albomicans* populations (He et al. 2000). Based on whole-genome sequencing, it was observed that B chromosomes originated from the standard chromosomes as a product of the sex chromosome–autosome fusion in *D. albomicans* (Zhou et al. 2012). Among fishes, the characid, *Astyanax scabripinnis*, has B chromosomes found more frequently in females (Vicente et al. 1996; Neo et al. 2000). In cichlids of Lake Victoria (East Africa), the B chromosomes are female-specific in *Lithochromis rubripinnis* and cross-breeding using females with and without the B chromosomes demonstrated that the presence of the B leads to a female-biased sex ratio (Yoshida et al. 2011). In all of these examples, it is believed that the difference in B-carrier sex ratio is a result of meiotic drive of the B. Although B and sex chromosome association has been described for various species, their relationship to one another is still poorly understood and needs further investigation. The effect of B chromosomes over the cell cycle seems the most influential mechanism that could favor the mitotic stability and maintenance of B chromosomes in somatic tissues.

Our findings show that BncDNA produces transcripts above 200 bp, a range classically defined as lncRNA (Kapranov et al. 2007). Furthermore, lncRNAs can be originated from transposable element (TEs) exonic domains (Kelley and Rinn 2012; Johnson and Guigó 2014). Somehow, BncDNA is impacted by the presence of a B chromosome. Based on the functional features of BncRNA identified herein, including *in silico* predictions, coupled with the lack of similarity to known, protein-coding sequences, we presume this sequence is an lncRNA. Previous studies have described B-specific sequences that are transcribed (Carchilan et al. 2009; Banaei-Moghaddam et al. 2013; Trifonov et al. 2013; Lin et al. 2014; Valente et al. 2014); however lncRNAs in these systems were not identified from, nor affected by, B chromosomes.

Long ncRNAs regulate cellular processes by different mechanisms, including via a miRNA interaction that impacts mRNA regulation (increase or decrease of expression levels) (Yoon et al. 2014). For example, lncRNAs can act through secondary structure interactions, as observed in the half-STAU1-binding site RNAs (1/2-sbsRNAs) that interact and regulate Staufin 1 (STAU1) mRNA in humans (Gong and Maquat 2011). Our *in silico* predictions indicate the possible formation of secondary structures and miRNA binding sites in BncRNAs. We hypothesize that predicted miRNA binding sites in the stem-loop secondary structure of BncDNA transcripts may regulate mature miRNAs, capturing them and acting as a “sponge” mechanism (Ebert et al. 2007). Alternatively, BncRNA transcription itself could also be regulated by miRNAs through transcript degradation (Leucci et al. 2013). Either mechanism could affect miRNA binding activity over other nascent mRNA target sequences. The

activity of miRNAs upregulated in female B+ muscle cells has been related to cilia assembly regulation and actin dynamics (mir-129-3p) (Cao et al. 2012) and regulation of neurogenesis and microglial inflammatory response mediation (mir-9-5p) (Coolen et al. 2013; Yao et al. 2014) in vertebrates.

Assuming BncDNA is a long noncoding RNA, this molecule could cause changes in the cell environment favoring B chromosome maintenance. Since BncDNA showed expression levels in gonad tissue correlated with B chromosome, we suggest it is involved in the meiotic drive of B chromosomes in female oocytes, preferentially transmitting B containing gametes to the next generation. The maintenance of B chromosomes is possible due to their transmission frequencies at a rate higher than Mendelian frequencies through the action of a variety of possible mechanisms, including segregation failure, by which B chromosomes undergo transmission distortion, or meiotic drive. While meiotic drive is the key to understanding B chromosome function and evolution, the molecular mechanisms that lead to B chromosome meiotic drive is far from clear (Houben et al. 2013). Meiotic accumulation has been described in various species of plants and animals as a consequence of the preferential migration of the B to the secondary oocyte instead of the polar body (Camacho et al. 2000). The noncoding BncRNA could represent a new player in the control mechanisms over meiotic drive and subsequent maintenance of B chromosomes.

Conclusion

The repeated BncDNA is a transcriptionally active element in the *A. latifasciata* genome, enriched on B chromosomes, and is found in other cichlid genomes and transcriptomes. The B enrichment of BncDNA leads to changes in BncRNA transcription levels. The observed length of BncDNA transcripts, coupled with an absence of identifiable coding sequences, indicate that BncDNA is a candidate lncRNA. The BncDNA transcripts are processed among different tissues with remarkable variation among sex and B– and B+ samples. Thus, the results obtained indicate BncDNA transcribes an lncRNA that might have an important effect on cell function, favoring B chromosome maintenance and segregation during cell division. Finally, the link between Bnc transcription and phenotypic sex warrants a reevaluation of the relationship between B chromosomes and sex chromosome origin and evolution.

Acknowledgments This work was supported by grants from São Paulo Research Foundation (13/04533-3, 10/17212-2, 12/20959-8, 13/01688-6), National Council for Scientific and Technological Development (474684/2013-0, 301728/2011-9), and São Paulo State University (0045/021/13-PROPE/CDC).

Compliance with ethical standards All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

This article does not contain any studies with human participants performed by any of the authors.

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

References

- Ahnert SE, Fink TMA, Zinovyev A (2008) How much non-coding DNA do eukaryotes require? *J Theor Biol* 252:587–592
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A (2013) Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell* 25:2536–44
- Basu S, Müller F, Sanges R (2013) Examples of sequence conservation analyses capture a subset of mouse long non-coding RNAs sharing homology with fish conserved genomic elements. *BMC Bioinformatics* 14:S14
- Bougourd SM, Jones RN (1997) B chromosomes: a physiological enigma. *New Phytol* 137:43–54
- Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71:515–526
- Brodsky LI, Ivanov VV, YaL K d, Leontovich AM, Nikolaev VK, Feranchuk SI, Drachev VA (1995) GeneBee-NET: internet-based server for analyzing biopolymers structure. *Biokhimiia* 60:923–928
- Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527–542
- Bueno D, Palacios-Gimenez OM, Cabral-de-Mello DC (2013) Chromosomal mapping of repetitive DNAs in the grasshopper *Abracris flavolineata* reveal possible ancestry of the B chromosome and H3 histone spreading. *PLoS One* 8, e66532
- Burt A, Trivers R (2006) *Genes in conflict: the biology of selfish genetic elements*. Belknap Press of Harvard University, Cambridge
- Cabral-de-Mello DC, Valente GT, Nakajima RT, Martins C (2012) Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. *Chromosome Res* 20:279–292
- Camacho JP (2005) B chromosomes. In: Gregory TR (ed) *The evolution of the genome*. Elsevier, San Diego, pp 223–286
- Camacho JP, Sharbel TF, Beukeboom LW (2000) B-chromosome evolution. *Philos T Roy Soc B* 355:163–178
- Cao J, Shen Y, Zhu L, Xu Y, Zhou Y, Wu Z, Li Y, Yan X, Zhu X (2012) miR-129-3p controls cilia assembly by regulating CP110 and actin dynamics. *Nat Cell Biol* 14:697–706
- Carchilan M, Kumke K, Mikolajewski S, Houben A (2009) Rye B chromosomes are weakly transcribed and might alter the transcriptional activity of A chromosome sequences. *Chromosoma* 118:607–616
- Carlson W (2009) The B chromosome of maize, in: Bennetzen JL, Hake S (Eds.), *Maize handbook—Volume II: genetics and genomics*, Springer, pp. 459–480
- Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ (2009) A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. *Chromosoma* 118:113–125
- Carvalho AB (2002) Origin and evolution of the *Drosophila* Y chromosome. *Curr Opin Genet Dev* 12:664–668
- Coolen M, Katz S, Bally-Cuif L (2013) mir-9: a versatile regulator of neurogenesis. *Front Cell Neurosci* 7:220
- Ebert MS, Neilson JR, Sharp PA (2007) MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* 4:721–726
- Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2:919–929
- Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS (2003) MicroRNA targets in *Drosophila*. *Genome Biol* 5:R1
- Fantinatti BEA, Mazzuchelli J, Valente GT, Cabral-de-Mello DC, Martins C (2011) Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genetica* 139:1273–1282
- Feldberg E, Bertollo LAC (1984) Discordance in chromosome number among somatic and gonadal tissue cells of *Gymnogeonhagus balzani*. *Braz J Genet* 8:639–645
- Feldberg E, Porto JIR, Brinn MNA, Mendonça MNC, Benzaquem DC (2004) B-chromosomes in Amazonian cichlid species. *Cytogenet Genome Res* 106:195–198
- Finn RD, Clements J, Eddy SR (2011) HMMER Web Server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) The Pfam protein families database. *Nucleic Acids Res* 42:D222–D230
- Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTR via Alu elements. *Nature* 470:284–288
- Graphodatsky AS, Kukekova AV, Yudkin DV, Trifonov VA, Vorobieva NV, Beklemisheva VR, Perelman PL, Graphodatskaya DA, Trut LN, Yang F, Ferguson-Smith MA, Acland GM, Aguirre GD (2005) The proto-oncogene C-KIT maps to canid B-chromosomes. *Chromosome Res* 13:113–122
- Griffiths-Jones S (2004) The microRNA registry. *Nucleic Acids Res* 32: D109–D111
- Gupta RA, Shah N, Wang KC, Jeewon K, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071–1076
- Han P, Li W, Lin C-H, Yang J, Shang C, Nurnberg ST, Jin KK, Xu W, Lin C-Y, Lin C-J, Xiong Y, Chien H-C, Zhou B, Ashley E, Bernstein D, Chen P-S, Chen H-SV, Quertermous T, Chang C-P (2014) A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* 514:102–106
- He LP, Ling FY, Zheng XZ, Wang W, Kuang RP (2000) The effect of B chromosome on the reproduction of *Drosophila albomicans*. *Yi Chuan Xue Bao* 27:114–120
- Houben A, Banaei-Moghaddam AM, Klemme S (2013) Biology and evolution of B chromosomes. In: Leitch IJ (ed) *Plant genome diversity*, vol 2. Berlin, Springer, pp 149–165
- Johnson R, Guigó R (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 20:959–976
- Jones RN (1991) B-chromosome drive. *Am Nat* 137:430–442
- Jones R, Rees H (1982) B chromosomes. Academic Press, London
- Jones RN, Gonzalez-Sanchez M, Gonzalez-Garcia M, Vega JM, Puertas MJ (2008) Chromosomes with a life of their own. *Cytogenet Genome Res* 120:265–280
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467
- Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR

- (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316:1484–1488
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649
- Kelley D, Rinn J (2012) Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Gen Biol* 13:R107
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of accessibility in microRNA target recognition. *Nature Genet* 39:1278–1284
- Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet* 5:288–98
- Kuroiwa A, Terai Y, Kobayashi N, Yoshida K, Suzuki M, Nakanishi A, Matsuda Y, Watanabe M, Okada N (2014) Construction of chromosome markers from the Lake Victoria cichlid *Paralabidochromis chilotes* and their application to comparative mapping. *Cytogenet Genome Res* 142:112–120
- Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
- Laurent GS, Wahlestedt C, Kapranov P (2015) The landscape of long noncoding RNA classification. *Trends in Genet* 31:239–251
- Leach CR, Houben A, Field B, Pistrick K, Demidov D, Timmis JN (2005) Molecular evidence for transcription of genes on a B chromosome in *Crepis capillaries*. *Genetics* 171:269–278
- Leucci E, Patella F, Waage J, Holmström K, Lindow M, Porse B (2013) microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. *Sci Rep* 3:2535
- Lin HZ, Lin WD, Lin CY, Peng SF, Cheng YM (2014) Characterization of maize B-chromosome-related transcripts isolated via cDNA-AFLP. *Chromosoma* 123:597–607
- Lorenz R, Bernhart SH, Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA package 2.0. *Algorithm Mol Biol* 6:26
- Martins-Santos IC, Portela-Castro ALB, Julio HF Jr (1995) Chromosome analysis of 5 species of the Cichlidae family (Pisces, Perciformes) from the Paraná River. *Cytologia* 60:223–231
- Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, Scholz U, Gundlach H, Wicker T, Šimková H et al (2012) Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc Natl Acad Sci U S A* 109:13343–13346
- Mattick JS, Makunin V (2006) Non-coding RNA. *Hum Mol Genet* 15:R17–R29
- McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A, Sweredoski MJ, Shishkin AA, Su J, Lander ES, Hess S, Plath K, Guttman M (2015) The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521:232–236
- Miao VP, Covert SF, VanEtten HD (1991) A fungal gene for antibiotic resistance on a dispensable (“B”) Chromosome. *Science* 254:1773–1776
- Muller PY, Janovjak H, Miserez AR, Dobbie Z (2002) Processing of gene expression data generated by quantitative real-time RT-PCR. *BioTechniques* 32:1372–1379
- Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD (2014) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* 43:D130–D137
- Neo DM, Moreira-Filho O, Camacho JMP (2000) Altitudinal variation for B chromosome frequency in the characid fish *Astyanax scabripinnis*. *Heredity* 85:136–141
- Nokkala S, Grozeva S, Kuznetsova V, Maryańska-Nadachowska A (2003) The origin of the achiasmatic XY sex chromosome system in *Cacopsylla peregrina* (Frst.) (Psylloidea, Homoptera). *Genetica* 119:327–32
- Palestis BG, Trivers R, Burt A, Jones RN (2004) The distribution of B chromosomes across species. *Cytogenet Genome Res* 106:151–158
- Pang KC, Frith MC, Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genet* 22:1–5
- Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, Mattick JS (2009) Genome-wide identification of long noncoding RNAs in CD8+ T cells. *J Immunol* 182:7738–7748
- Pickard MR, Williams GT (2014) Regulation of apoptosis by long non-coding RNA GASS in breast cancer cells: implications for chemotherapy. *Breast Cancer Res Tr* 145:359–370
- Pinkel D, Straume T, Gray JW (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci U S A* 83:2934–2938
- Poletto AB, Ferreira IA, Martins C (2010a) The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. *BMC Genet* 11:1
- Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C (2010b) Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet* 11:50
- Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME (2014) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res* 43:D168–D173
- Rajicic M, Adnadevic T, Stamenkovic G, Blagojevic J, Vujosevic M (2015) Screening of B chromosomes for presence of two genes in yellow-necked mice, *Apodemus flavicollis* (Mammalia, Rodentia). *Genetika* 47:311–321
- Ruiz-Estévez M, López-León MD, Cabrero J, Camacho JPM (2012) B chromosome ribosomal DNA is functional in the Grasshopper *Eyprepocnemis plorans*. *PLoS One* 7, e36600
- Sambrook J, Russell DW (2001) *Molecular cloning. A laboratory manual* (3rd ed) Cold Spring, New York
- Sandelin A, Bailey P, Bruce S, Engström PG, Klos JM, Wasserman WW, Ericson J, Lenhard B (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* 5:99
- Simon P (2003) Q-gene: processing quantitative real-time RT-PCR data. *Bioinformatics* 19:1439–1440
- Skelton P (2001) *A complete guide to the freshwater fishes of Southern Africa* (2th) New Holland publishing, South Africa
- Smit AFA, Hubley R, Green P (2013–2015) RepeatMasker Open-4.0. <<http://www.repeatmasker.org>>
- Solovyyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7:S10
- Standaert L, Adriaens C, Radaelli E, Keymeulen AV, Blanpain C, Hirose T, Nakagawa S, Marine J-C (2014) The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* 20:1844–1849
- The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43:D204–D212
- Tian E, Borsset M, Sawyer JR, Brede G, Vatsveen TK, Hov H, Waage A, Barlogie B, Shaughnessy JD Jr, Epstein J, Sundan A (2015) Allelic mutations in noncoding genomic sequences construct novel transcription factor binding sites that promote gene overexpression. *Genes Chromosomes Cancer* 54:692–701
- Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 13:36–46
- Trifonov VA, Demytyeva PV, Larkin DM, O’Brien PC, Perelman PL, Yang F, Ferguson-Smith MA, Graphodatsky AS (2013)

- Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (*Capreolus pygargus*). *BMC Biol* 6:90
- Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* 147:1537–1550
- 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* 31:2061–2072
- Vicente VE, Moreira-Filho O, Camacho JPM (1996) Sex-ratio distortion associated with the presence of a B chromosome in *Astyanax scabripinnis* (Teleostei Characidae). *Cytogenet Cell Genet* 74:70–75
- Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. *Mol Cell* 43:904–914
- White MJD (1973) *Animal cytology and evolution*. Cambridge University Press, London
- Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23:1494–1504
- Yao H, Ma R, Yang L, Hu G, Chen X, Duan M, Kook Y, Niu F, Liao K, Fu M, Hu G, Kolattukudy P, Buch S (2014) MIR-9 promotes microglial activation by targeting MCP1. *Nat Commun* 5:4386
- Yoon J-H, Abdelmohsen K, Gorospe M (2014) Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol* 34:9–14
- Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, Hirai H, Hirai Y, Matsuda Y et al (2011) B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genet* 7, e1002203
- Zhou Q, Zhu H, Huang Q, Li Z, Zhang G, Roy SW, Vicoso B, Xuan Z, Ruan J, Zhang Y, Zhao R, Ye C, Zhang X, Wang J, Wang W, Bachtrög D (2012) Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 13:109