Peer-reviewed version available at Genes 2018, 9, 269; doi:10.3390/genes9060269

Article

Structural and functional analysis of transposable elements focusing on B chromosomes of the cichlid fish *Astatotilapia latifasciata*

Rafael Coan [1] and Cesar Martins [1,*]

[1] Sao Paulo State University (UNESP), Institute of Biosciences, Department of Morphology, Botucatu, SP, Brazil.; rafaelbcoan@gmail.com (R.C.); cmartins@ibb.unesp.br (C.M.)

* Correspondence: cmartins@ibb.unesp.br; Tel.: +55-14-3880-0462 (C.M.)

Abstract: B chromosomes (B) are supernumerary elements found in many taxonomic groups. Most B chromosomes are rich in heterochromatin and composed of abundant repetitive sequences, especially transposable elements (TEs). Bs origin is generally linked to the A chromosome complement (A). The first report of a B chromosome in African cichlids was on *Astatotilapia latifasciata*, which can harbor 0, 1 or 2 B chromosomes. Classical cytogenetics studies found high TE content on the species B chromosome. In this study, we aim to understand TE composition and expression on *A. latifasciata* genome and its relation to the B chromosome. We use bioinformatics analysis to explore TEs genome organization and also their composition on the B chromosome. Bioinformatics findings were validated by fluorescent *in situ* hybridization (FISH) and real-time PCR (qPCR). *A. latifasciata* has a TE content similar to other cichlid fishes and several expanded elements on its B chromosome. With RNA sequencing data (RNA-seq) we showed that all major TE classes are transcribed in brain, muscle and male/female gonads. The evaluation of TE expression between B- and B+ individuals showed that few elements have differential expression among groups and expanded B elements were not highly transcribed. Putative silencing mechanisms may the acting on the B chromosome of *A. latifasciata* to prevent adverse consequences of repeat transcription and mobilization in the genome.

Keywords: Repetitive elements, RNA-Seq, genomics, evolution, cytogenetics, supernumerary elements, extra chromosomes

1. Introduction

B chromosomes are supernumerary elements additional to the autosome chromosomes (A) and have been observed in several species of animals, plants and fungi. Bs have a distinct evolutionary pathway compared to the A chromosomes, with a non-Mendelian form of inheritance [1]. Bs can have neutral, deleterious or beneficial effects to the hosts: presence of additional chromosome correlates to sex determination on the cichlid fish *Lithocromis rubripinnis*, V-shaped phenotype on the common frog *Rana temporaria* and antibiotic resistance on *Nectria haematococca* fungus [1–4].

B chromosomes are usually heterochromatic due to abundance of repetitive elements in their composition [5]. Regardless of the repetitive content predominance on Bs, they can carry genes with different levels of integrity, including fully transcribed copies. Gene copies found on B chromosome can modulate gene expression of A complement genes and even influence metabolic pathways [6–8]. Bs are usually a mosaic of sequences from the A complement and B origin is generally linked to genomic instability of A chromosomes, with the formation of a proto-B and later expansion of repetitive elements [9–11]. Repetitive elements including transposable elements (DNA transposons and retrotransposons) [12–15] represent a huge portion of most eukaryotic genomes and are also a major component of B chromosome constitution [3,13]. Furthermore, gene clusters, such as U2 snRNA [16], 18S rRNA and H1 [11], H3 and H4 [17] histones were also found on Bs from many

species. Classical cytogenetics analysis tries to delimit B chromosome origin on the different A complement pairs with the use of probes coming from multigenic families and TEs [11,18–21]. The TE content on B chromosomes is related to the TE content of the A complement. Expansion of repeats on B is a common characteristic of the supernumerary caused by lack of recombination and low selective pressure [3]. In rye's B chromosome, there is accumulation of satellite DNA, Ty1/Copia, a LTR retrotransposon, and a few unclassified sequences [22]. On B of *Alburnus alburnus* fish, expanded Ty3/Gypsy sequence had similarity with the reverse transcriptase codifying gene [23].

In addition to the accumulation of repetitive elements on Bs, there are also evidences that those elements can be transcribed. The *StarkB* element specific to maize B chromosome has variable expression among individuals and loci [24]. This element, in a different study, was found expressed together with Gypsy and Copia TEs. In fact, *StarkB* expression showed evidence of dose-dependence; expression increases as the number of Bs rises. This demonstrates the influence of the number of B chromosomes in their sequences transcription [25]. Another recent study found differentially expressed repetitive sequences between 0B and 4B individuals in rye. It was also noted difference of transcription among different tissues [26].

Among African cichlids, B chromosomes were observed on twenty species [2,19,27–29], being first described in *Astatotilapia latifasciata*, with the presence of 0, 1 or 2 B chromosomes [27]. Bs of *A. latifasciata* are among the largest B chromosomes already investigated. They are heterochromatic and abundant on repetitive elements, as evidenced by classical cytogenetics mappings based on 18S ribosomal DNA (rDNA), Rex1 and Rex3 elements [19,27]. Recent next generation sequencing (NGS) analysis also revealed accumulation of diverse classes of repetitive sequences on the B chromosome of this species [10]. In this way, our study aims to understand TE content, distribution and transcription on *A. latifasciata* genome and its impact on the B chromosome. We use a combination of classical molecular cytogenetics and NGS to evaluate the TE landscape of the species and find representative sequences on the B chromosome. We analyzed the transcription levels of repeats and their possible relation to B enriched sequences. In summary, the understanding of repeat content on the A complement, and they relation to the B chromosome, will help elucidate B's constitution and perpetuation, as well as their influence on the cell biology.

2. Materials and Methods

Samples

All animal samples were available at the fish facility from the Integrative Genomics Laboratory, Sao Paulo State University, Botucatu, Brazil. We respected the ethical principles adopted by the Brazilian College of Animal Experimentation, with approval from the Institute of Biosciences/UNESP – Sao Paulo State University ethics committee (protocol no 486-2013). We used samples from males and females with B chromosomes (B+) or without Bs (B-). Samples were genotyped into B- and B+ by PCR, with specific primers for B chromosome presence or absence [30].

All datasets for bioinformatics analysis were previously sequenced. DNA sequencing consisted in four male (M1-0B, M2-1B, M3-1 and M4-2B) and two female samples (F1-0B and F2-1B) from two previous studies [10,31]. RNA sequencing was performed in three tissues: brain, muscle and gonads. Within each tissue, six samples were B- and six B+. Each B condition (B- or B+) has three males and three females, with 36 total RNA-Seq samples. All data are available in Sacibase (sacibase.ibb.unesp.br/). For FISH, we used a B+ sample, confirmed though molecular PCR and cytogenetic analysis. For qPCR, we used B- (sample ID 04) and B+ (samples IDs 907, 908, 910, 913, 918 and 919) samples confirmed by PCR.

Repeat identification and landscape pipelines

In order to estimate the repetitive content on the genome of *A. latifasciata*, we used an assembled 0B genome, containing male and female reads as reference [31]. We first created a custom repeat library with RepeatModeler 1.0.8 [32], according to their instructions and default parameters. Any ID issue in the created *fasta* file was resolved manually. We merged the custom repeat library with

RepBase Update 20150807 [33] in order to obtain a comprehensive repeat library to input in RepeatMasker. Despite the manual curation, some sequences maintained the "*Unknown*" status. On the second step in the repeat identification, the merged library was used as input for RepeatMasker 4.0.5 [34] to search for repeat copy number and organization on the assembled genome. RepeatMasker was run with "*slow (-s)*", "*align (-a)*" and "*library (-lib)*" parameters. To summarize RepeatMasker results we used *buildSummary.pl*, a perl script from their package. Output files from RepeatMasker were also used as input for *createrepeatlandscape.pl* and *calcdivergencefromalign.pl* scripts, in order to calculate Kimura divergence values and plot the repeat landscape. Both are helper perl scripts from RepeatMasker package.

Comparative analysis pipeline

For the first search of expanded elements on the B chromosome of *A. latifasciata* we used RepeatExplorer [35] which performs a graph-based clustering of raw Illumina reads. RepeatExplorer uses a small subset of sequenced reads (0.1 – 0.5x coverage) [36] as input, providing a fast and accurate way to compare two or more datasets.

Raw Illumina datasets from genomes without B chromosome ("0B") and with 2B chromosomes ("2B") available at Sacibase (http://sacibase.ibb.unesp.br/) [10] were used as input for RepeatExplorer pipeline. RepeatExplorer provides a set of helper scripts to prepare the data for clustering. Reads were quality filtered based on default parameters of *paired_fastq_filtering.R* and a random sample of 5 million reads for each genome comprising of approximately 0.5x genome coverage was selected. Finally, graph-based clustering was applied for *de novo* repeat identification and comparative analysis [35,36] using RepeatExplorer pipeline with default parameters and developer's recommendations (https://repeatexplorer-elixir.cerit-sc.cz/; oral communication from authors). Results from clustering were visually inspected with respect to their graphic composition, which indicates the type of repeat, and proportion of reads from each genome. We manually chose clusters with the highest content of 2B reads in comparison with 0B reads for further steps. Longest contigs from the selected cluster were annotated with RepeatMasker and BLAST, using the command line and default parameters. To solve eventual ambiguities in the annotation, priority was given to RepeatMasker, with a joint search for conversed domains via BLAST-CDD; the identification of TE related proteins was key to classify the assembled contigs. We used the annotated contigs found by RepeatExplorer to design primers for probe construction for fluorescence in situ hybridization (FISH) and qPCR validation (see next sections).

To further characterize the repetitive content of *A. latifasciata* B chromosome we performed a coverage ratios analysis on alignments of six sequenced samples (Supplementary Table S1). With this methodology, we could analyze TE expansion on the B chromosome for individual loci, fine tuning the results. A similar approach was previously used in order to find B chromosome blocks [8,10]. The coverage ratios analysis was done as follows. Raw illumina reads were filtered to eliminate adapters and bacterial contaminants. We created a library of adapters and bacterial genomes and aligned the reads against this library. We used bowtie2 2.1 [37] with "*--very-fast-local*" parameter, thus excluding reads similar to the contaminants/adapters library. This created a contaminant free dataset. We checked the reads quality distribution with fastqc 0.10.1 [38] software and performed filtering with fastx-toolkit 0.0.13 [39]. We used as quality cut-off a Phred of 28 over 80% of the read. We then applied pairfq 0.11 [40] to restore pair-end reads. Filtered reads were aligned with bowtie2 against *A. latifasciata* reference genome (see Repeat identification and landscape pipeline topic) using the "*--very-sensitive*" parameter. Alignment statistics from BAM files were extracted with qualimap 2.2 [41].

BAM files were used to extract coverage values. We used previously generated RepeatMasker results (see Repeat identification and landscape pipeline topic) in order to extract coverage only for TE regions. Since our analysis was focused on TEs, repetitive elements with "*Simple_repeat*" and "*Low_complexity*" were excluded from the annotation files. To extract coverage information from BAM files we used bedtools 2.25 [42]. The single copy gene hypoxanthine phosphoribosyltransferase (HPRT) was a cut-off for low coverage regions (see details in Supplementary Methods). The average

coverage for each TE interval was calculated using custom bash and python scripts. We used the average coverage of each interval to calculate the coverage ratio between the different genomes, always using the M1-0B sample as reference (Supplementary Table S2; Supplementary Methods). The ratios between HPRT coverages (Supplementary Table S3) were used to normalize the different read coverages among samples.

Transcriptome analysis

We had access to 36 libraries of messenger RNA (mRNA) previously sequenced (available at Sacibase), from muscle, gonads and brain from 0B (B-) and 1B (B+) males and females [43]. They were sequenced in an Illumina HiSeq2000 and include approximately 30 million reads each. The data is freely available at Sacibase. Reads were filtered to remove adapters and contaminants, similarly to the genomic pre-processing. They were quality filtered to maintain reads with at least a phred of 28 over 80% of the read. Filtered reads were submitted to RepEnrich 1.2 [44] which performs an enrichment of transcripts on the assembled genome using repeat coordinates as reference. RepEnrich returns a count table from each element and takes account the repetitive nature of these elements; results can be used for expression estimation within a tissue, or for differential expression. We followed the developer's tutorial to perform these analyses (https://github.com/nerettilab/RepEnrich). To find TE expression values, we used the Bioconductor package edgeR 3.4.2 and the counts table produced by RepEnrich. We calculated TE expression within a specific B- tissue, in order to evaluate if there is TE expression in all of them. To find a particular tissue expression, we used edgeR to calculate the RPKM (reads per kilobase per million mapped reads) values of each B- tissue. We separated male and female gonads in the analysis as they are morphologic and functionally different tissues. Later, the differential expression levels between B- and B+ samples across tissues were evaluated based on the RepEnrich and edgeR developer's recommendations, using a GLM (generalized linear model) statistical model. We calculated log2 fold-change and false discovery rate (FDR) for each tissue and compared B- and B+ values within tissues. As of previous steps, male and female gonads were analyzed separately. To consider an element differentially expressed, we used a log2 fold-change cut-off of 1.2 and FDR cut-off of 0.05.

Validation steps

Total DNA was extracted with Qiagen's DNeasy® Blood & Tissue Kit, using the fish caudal fin. Primers were designed on annotated contigs from selected RepeatExplorer clusters (Supplementary Table S4; Supplementary Table S5) using OligoExplorer® 1.2 [45], Primer-BLAST [46] and PCR Primer Stats [47]. Amplicons produced by conventional PCR were sequenced using the Sanger method in order to confirm the sequences found by RepeatExplorer. After confirmation, sets of FISH probes were PCR labeled with biotin-dUTP.

Mitotic chromosome preparations were performed according to [48] with modifications. Kidney tissue was dilacerated with forceps and syringe, and placed on a KCL 0,075M solution and incubated for 30 minutes. The cell suspension was fixed with methanol 3:1 acetic acid mixture, following three washes and 900rpm centrifugations (10 minutes each) were conducted, discarding the supernatant at the end of each centrifugation. Preparation was stained with Giemsa and metaphases were visualized under an optic microscope.

FISH was performed on chromosome squash preparations according to protocol by [49] with modifications. Slides were dehydrated with 70% ethanol, pretreated with RNAse A (37ºC for 20 minutes) and pepsin/HCl (3 minutes), washed (3x 2xSSC for 2 minutes each), dehydrated (2x 70% ethanol for 2 minutes; 2x 90% ethanol for 2 minutes; 1x 100% ethanol for 4 minutes) and dried (1 hour at 60ºC). For denaturation, formamide 70% was applied to slides during 37 seconds at 65ºC and readily dehydrated (2x 70% ethanol for 2 minutes; 2x 90% ethanol for 2 minutes; 1x 100% ethanol for 4 minutes). Overnight hybridization was performed with previous PCR labeled probes. Slides were washed (42ºC – formamide 50% for 5 minutes; 2 times 2x SSC for 5 minutes; 2 times 4xSSC/Twenn for 5 minutes) and then detected with fluorescein isothiocyanate (FITC) (0,2µl FITC: 100µl 4xSSC tween – 37ºC for 30 minutes). A final wash was made (3 times 4xSSC tween for 3 minutes) and the

slides stained with 4',6-diamidino-2-phenylindole (DAPI). Slides were visualized on a BX61 Olympus microscope and hybridized metaphases registered under an Olympus DP71 digital camera coupled to the microscope. The metaphases were trimmed using Adobe Photoshop CS2.

Quantitative PCR (qPCR) of genomic DNA was performed on a StepOnePlus™ Real-Time PCR equipment, from Applied Biosystems, using GoTaq® qPCR Master Mix kit from Promega. Cycling conditions were 95ºC for 10 minutes; 40 cycles of denaturation at 95ºC – 15 seconds and annealing/extension at 60ºC - 60 seconds and 1 final cycle of dissociation at 95ºC. Results were used to calculate the gene dosage ratio (GDR) using the $2^{-\Delta CT}$ [50], by relative quantification to the autosomal single copy gene *hypoxanthine phosphoribosyltransferase* (HPRT), with an Excel spreadsheet.

3. Results

DNA transposable elements are the most representative class in A. latifasciata genome

Estimation of transposable elements copy number and their percentage on the 0B genome, here considered the canonical genome, was obtained with RepeatMasker (Figure 1a, Supplementary Table S6 and Supplementary file 1). *A. latifasciata* genome is comprised of 28.41% of repetitive DNA, and the majority are transposable elements. DNA transposons are the most representative category, followed by LINE, LTR and SINE retrotransposons. In total, retrotransposons (LTR, LINE and SINE) have 354,288 copies versus 464,043 of DNA transposons. Among retroelements, LINEs have both higher copy number and bases masked on the genome. SINEs are in higher number than LTRs, but LTRs have more bases masked on the genome, probably due to the larger size of each copy. We found little over 7% of unclassified elements on the genome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2018 doi:10.20944/preprints201803.0145.v1

Peer-reviewed version available at Genes 2018, 9, 269; doi:10.3390/genes9060269

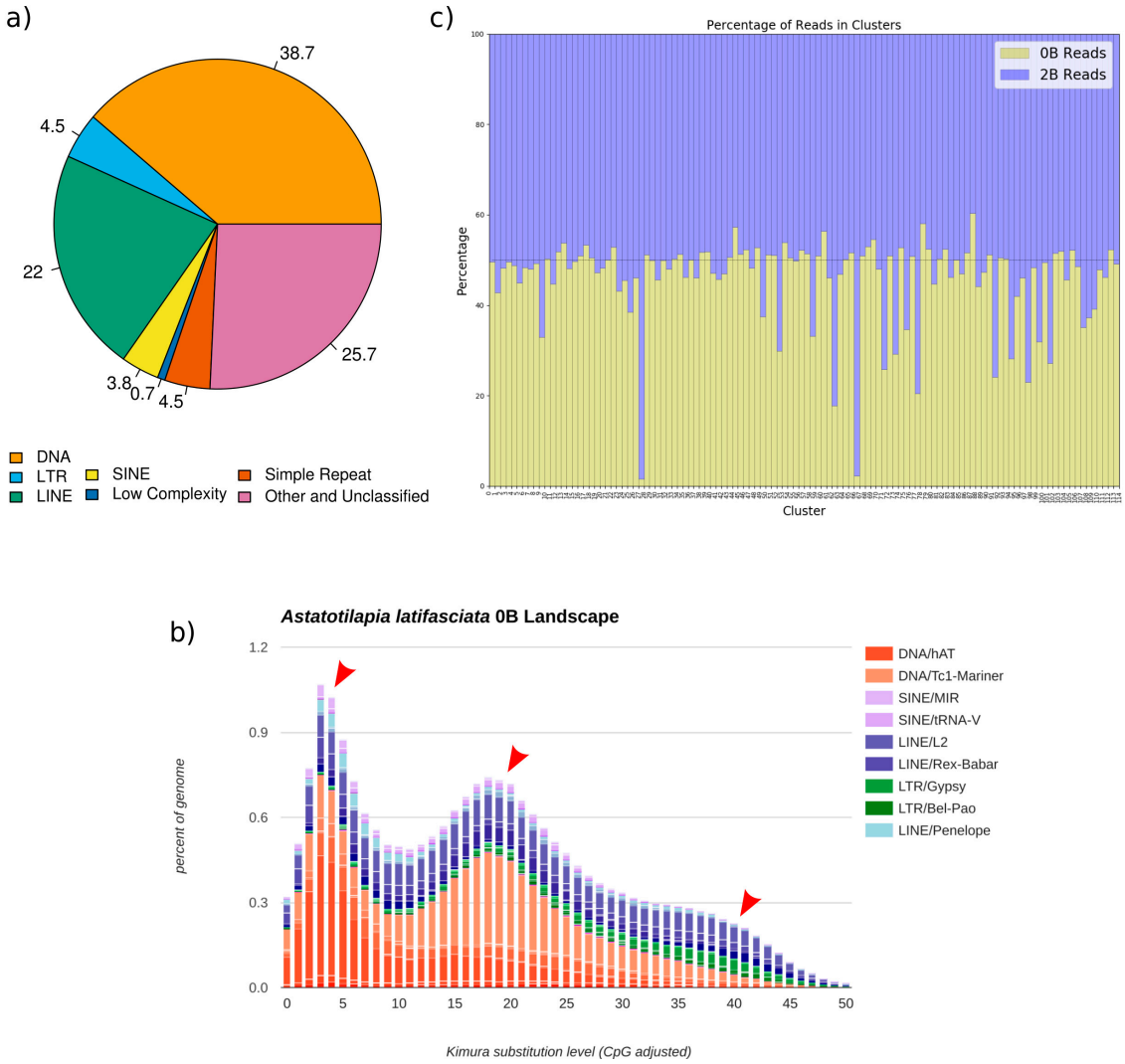

Figure 1. Genomic characteristics of repeated DNAs in *A. latifasciata* genome. **a)** Percentage of each type of repetitive sequences, based on the whole repetitive portion detected by RepeatMasker on the 0B genome. DNA transposons and LINE dominate the repetitive content. Unclassified sequences comprise mostly of multigenenic families and gene clusters. **b)** Repeat landscape of *A. latifasciata* genome. The graph shows, for each element, the sequence divergence from their consensus (x axis), in relation to their number of copies on the genome (y axis). Peaks represent insertion waves (red arrows) of elements on the genome. Elements with older insertion waves are present on the right side of the graph, while newer insertions are depicted on the left side. Different colors show distinct element types described on the right side. For a detailed, interactive version of the graph, please refer to Supplemental file 2. **c)** Comparative analysis of repeats on 0B and 2B genomes, via RepeatExplorer. Each column represents 100% of reads of a cluster; reads proportion from 0B and 2B genomes are in yellow and blue, respectively. Clusters with higher proportion of 2B reads (in blue) are expanded through the lower part of the graph (yellow).

Tc1-Mariner (4.37% of the genome) and hAT-Ac (1.77% of the genome) are the most representative DNA elements. Together they represent over 50% of DNA transposons on *A. latifasciata* genome. There is a high diversity of other DNA elements, but each in a small percentage. The predominant LINE is L2 (2.44%) but the previously found element Rex (1.06% of the genome) is also representative. All major classes of LTR are found on the genome, with Gypsy (0.61%) and Pao (Bel/Pao – 0.21%) classes being the most abundant. Among SINEs, there is high variation of element copy number, with families derived from transporter RNA (tRNA) the most abundant.

Three main bursts of DNA elements during evolution explain their genome abundance

Using the assembled 0B genome and RepeatMasker results, a repeat landscape was constructed (Figure 1b, and Supplementary file 2 for an interactive version). It indicates the relative waives of insertion of specific families in the genome [51]. There are three main burst events on *A. latifasciata* evolution. In an older burst, there is a slightly accumulation of retrotransposons, both LTR (Bel/Pao, Gypsy and ERV) and LINE (Rex-Babar and L2). The second major burst of TEs has DNA elements as major forces. Tc1-Mariner has the most increment in copy number during this period. The third most recent burst of TE insertions is due to the accumulation of hAT, also a DNA element. Although hAT copy number begins to increase in the second insertion wave, the third insertion wave has predominantly hAT elements.

As a trend, we see DNA elements dominating the landscape of TE insertions during *A. latifasciata* genome evolution. Retroelements shows signs of constant mobilization, *e. g.* L2 and Rex, which seems to have both young and older copies, represented by a constant appearance on the landscape.

Accumulation of TEs on the B chromosome of A. latifasciata

To establish whether there is accumulation of transposable elements on the B chromosome of *A. latifasciata*, we performed a comparative analysis using RepeatExplorer and coverage ratios. The RepeatExplorer pipeline resulted in 114 clusters (Supplementary file 3), with approximately 75 percent of reads in clusters and 25 percent in singlets. Clusters having higher proportion of 2B reads were selected for manual inspection (Figure 1c, Table 1, Supplementary file 3). From a total of 20 analyzed clusters, four were chosen for FISH probe construction; clusters with better contig annotations and genome relevance were selected. The clusters with higher proportions of 2B reads were clusters 28 and 67, annotated as Bel/Pao and Gypsy, respectively. This indicates an expansion of those TEs on the extra chromosome. Other clusters selected for FISH were cluster 63 (hAT element) and cluster 74 (L2 element).

Table 1. Selected clusters chosen for FISH probe construction and their annotations. Clusters from RepeatExplorer were selected for having mainly 2B reads. Retro, retrotransposon; DNA, DNA transposon; LTR, long terminal repeat retrotransposon; LINE, long interspersed nuclear elements.

Cluster	0B Reads	2B Reads	Annotation	Repeat Class
28	359	23,389	Bel/Pao	Retro/LTR
63	2,939	13,563	DNA/hAT	DNA
67	351	15,603	Gypsy	Retro/LTR
74	4,575	11,117	L2	Retro/LINE

Although RepeatExplorer returned different expanded elements on the B chromosome, we chose to further categorize these elements and have a broad picture of B's TE composition. In this way we could computationally validate RepeatExplorer's data and build a database of putative expanded elements with their respective loci. We performed a coverage ratios analysis on the various NGS datasets available (0B, 1B and 2B, males and females).

Coverage ratios found several elements expanded on the B chromosome, and they match RepeatExplorer data (Table 2; Supplementary file 4). In general, the repeat class with the most accumulation on B's is DNA transposons, followed by LINEs and LTRs. A high number of unclassified sequences were also found on B, showing a diverse mobility of sequences to the supernumerary.

Table 2. Coverage ratios analysis between the six sequenced individuals. Values are normalized in reference to the M1-0B genome (0B male). Values show the number of extra copies in a sample in relation to the reference. Numbers are the sum of all loci expanded on the B chromosome. DNA, DNA transposon; LTR, long terminal repeat retrotransposon; LINE, long interspersed nuclear elements.

Element	Class	Family	F1-0B/M1	F2-1B/M1	M2-1B/M1	M3-1B/M1	M4-2B/M1
Gypsy-188_DR-I	LTR	Gypsy	0.00	357.90	478.75	505.95	977.98
AlRepB-26	LINE	L2	2.07	454.47	438.12	454.37	976.85
BEL32-I_DR	LTR	Pao	11.24	276.16	324.04	343.45	731.48
Gypsy-23_GA-I	LTR	Gypsy	0.00	311.83	336.77	360.23	709.62
Mariner-N13_DR	DNA	TcMar-Tc1	17.81	276.37	294.65	312.57	648.31
AlRepB-738	DNA	hAT-Ac	23.67	320.79	356.62	382.23	631.88
BEL32-LTR_DR	LTR	Pao	5.46	215.99	278.12	309.23	593.38
Maui	LINE	L2	777.82	916.34	626.59	665.77	312.38
AlRepC-927	DNA	hAT-Ac	253.98	250.69	96.81	92.91	0.00
AlRepB-157	Satellite	Satellite	545.89	537.41	39.69	23.75	44.40
TZSAT	Satellite	Satellite	181.04	177.25	4.73	4.52	6.86

Examples of major TEs expanded on the B chromosome found by coverage ratios analysis include Gypsy-188_DR-I (Gypsy), AlRepB-26 (L2), Bel32-I_DR (Bel/Pao), Mariner-N13_DR (Tc1-Mariner) and AlRepB-738 (hAT-Ac) (Tables 1 and 2). RepeatExplorer also found elements from these families and some were validated by qPCR, sequencing and FISH (see next section). Copy number for those elements varies with the number of Bs on the genome, that is, the 2B individual have predominantly accumulated more copies than the 1B individual. Several other TEs follow this pattern (Table 2).

The data also shows different trends, with accumulation of elements in one of the sexes, *e. g.* male or female, or accumulation related to B presence and sex. For example, the Maui element (member of L2 family) is expanded on females and B chromosome carriers. Also, the element AlRepC-927 (hAT family) is absent on the 2B individual, which may indicate a recent mobilization or a populational bias. Although the individual analysis of each element is out of the scope of this work, such examples shows the plasticity and mobility that TEs can achieve, especially considering the presence of a B chromosome.

B chromosome TE accumulation is validated by cytogenetics and molecular techniques

Sequences found by RepeatExplorer were used to design a custom set of primers and to validate elements expanded on the B chromosome of *A. latifasciata* (Supplementary Table S4 and S5). Custom probes were obtained by PCR and hybridized on mitotic chromosome preparations (Supplementary Table S4). FISH analysis revealed intense hybridization signal mostly on the metacentric B chromosome (Figure 2a). As predicted by the bioinformatics pipeline, signals from custom probes concentrated on the B indicate accumulation of the element on the extra chromosome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2018 doi:10.20944/preprints201803.0145.v1

Peer-reviewed version available at Genes 2018, 9, 269; doi:10.3390/genes9060269

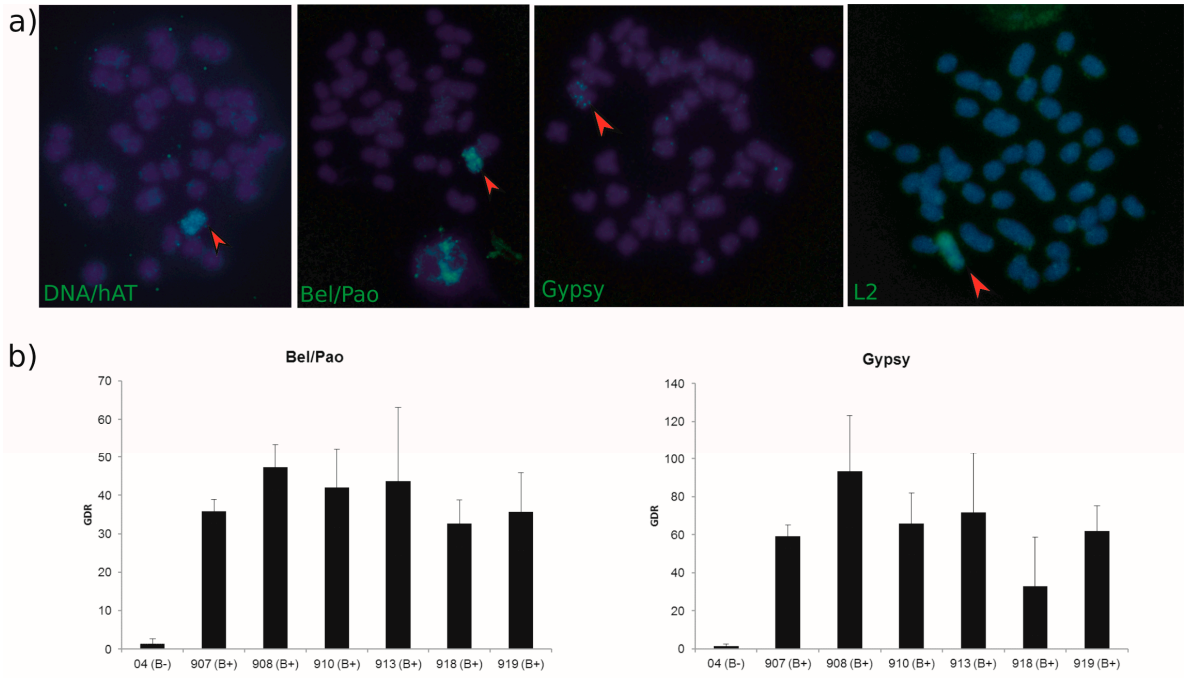

Figure 2. Distribution and copy number variation of repetitive DNAs in the *A. latifasciata* genome. **a)** FISH mapping of four selected elements (DNA/hAT, Bel/Pao, Gypsy and L2) on *A. latifasciata* metaphasic chromosomes. Probes were PCR labeled by biotin-dUTP and the signal was detected by FITC (green). Red arrowheads indicate the B chromosomes. **b)** Quantitative PCR of Bel/Pao and Gypsy elements detected by RepeatExplorer pipeline. Sample number 04 has no B chromosome (B-), while the others have at least one B (B+). Data shows B+ individuals with higher copy number than B- individuals. GDR: gene dosage ratio.

qPCR was used to quantify the sequences in relation to the normalizer, single copy gene HPRT. Bel/Pao and Gypsy elements were chosen due to their abundance on the B chromosome (Supplementary Table S5). Relative quantification showed higher copy number of the elements on B carrying individuals (Figure 2b). All B+ samples, from Bel/Pao and Gypsy elements, show higher GDR than the individual 04 (B-). Results shows B carrying individuals with higher element copy number than B- individuals. Data from qPCR agrees with results from RepeatExplorer and FISH.

TEs transcription across tissues and higher expression on the B chromosome samples

RPKM values shows that many TEs are transcribed on various *A. latifasciata* tissues (Supplementary Figure 1; Supplementary file 5). All major families are expressed, with high variability among individual elements. Although not suitable for comparison between tissues, RPKM values reveal that TE transcription occurs on *A. latifasciata* genome. Elements with recent insertion waves, thus putative activity, are expressed on different tissues. Tc1-Mariner and hAT have the highest expression levels among DNA transposons. L2, RTE-BovB, Rex, L1, Penelope and MIR have high expression within retrotransposons. As a trend, families' expression is constant in the tissues.

We sought to find if elements on the B chromosome have different expressions levels than the rest of the genome. This would show putative activity of elements present on the B, if compared to the A complement. All tissues showed differential expression of repetitive elements between B+ and B- genomes (Figure 3; Table 3; Supplementary file 6). There is a great variability in families expressed in the different tissues, but DNA transposons have differential expression in all of them. On the brain, there are 15 upregulated elements and 5 downregulated. Retroelements dominates the differential expression; 14 out of 20 differentially expressed are retroelements. On the muscle, only 3 elements have differential expression, all upregulated on the B+ tissue. Excluding an unclassified element, the other two are DNA transposons. On the male gonads, we detected 14 differentially

Peer-reviewed version available at Genes 2018, 9, 269; doi:10.3390/genes9060269

expressed elements, with only 3 upregulated on B+ genome. On the female gonads, only 3 elements have differential expression, all upregulated on the B+ genome. Our results indicate that, with exception of male gonads, there is a trend for upregulation of elements on B+ tissues.

Table 3. Repetitive elements with differential expression between B- and B+ individuals. B expansion represents how many more copies an individual has, related to the reference M1-0B. The number is a sum of all loci of analyzed individuals (F1-0B/M1-0B, F2-1B/M1-0B, M2-1B/M1-0B, M3-1B/M1-0B, M4-2B/M1-0B).

Element	Superfamily	Class	Fold-Change	B Expansion
Brain				
AlRepD-1119	L2	LINE	1.4829	22/88/115/126/187
CR1-28_HM	CR1	LINE	3.2887	No
REX1-3_XT	Rex-Babar	LINE	1.7297	No
GYPSY2-I_CB	Gypsy	LTR	5.2102	No
AlRepC-299	Unknown	Unknown	1.5748	9/16/4/6/0
I-6_AAe	I	LINE	3.6624	No
L1-6_DR	L1	LINE	-3.8609	2/0/0/0/2
BovBa-1_EF	RTE-BovB	LINE	4.1560	No
AlRepD-1964	Unknown	Unknown	1.1420	24/117/121/120/160
Gypsy-34-I_DR	Gypsy	LTR	4.5448	No
AlRepD-520	Unknown	Unknown	1.8113	47/30/6/2/0
Mariner-1_SP	TcMar-Fot1	DNA	4.0150	0/0/2/5/12
ERV1-3N-EC_I-int	ERV1	LTR	2.4732	No
Gypsy-17-I_DR	Gypsy	LTR	-1.7877	2/2/2/4/16
Gypsy52-I_DR	Gypsy	LTR	5.4846	No
Tc1-2Eso	TcMar-Tc1	DNA	2.4478	No
Rex1-52_DR	Rex-Babar	LINE	-3.1220	No
Helitron-2_DR	Helitron	RC	-1.2385	No
RMER17C-int	ERVK	LTR	2.6684	No
Gypsy-20-I_DR	Gypsy	LTR	-2.1144	0/0/2/0/0
Muscle				
AlRepC-299	Unknown	Unknown	7.5228	9/16/4/6/0
AlRepB-358	hAT-Ac	DNA	3.1883	207/209/146/139/11
Mariner-1_SP	TcMar-Fot1	DNA	11.0098	0/0/2/5/12
Gonad Male				
AlRepD-4130	hAT-Ac	DNA	2.7431	37/87/51/56/31
HEROTn	R2-Hero	LINE	-6.4246	No
7SLRNA	srpRNA	srpRNA	3.4512	No
P-27_HM	P	DNA	-2.5097	No
AgaP15	P	DNA	-5.8314	No
Charlie16	hAT-Charlie	DNA	-5.7227	No
AlRepE-134	DNA	DNA	-2.9391	6/3/3/5/3
ERV-4_CPB-I	ERV1	LTR	-2.9007	No
Gypsy-10_GA-LTR	Gypsy	LTR	-2.6009	No
AlRepE-2243	Unknown	Unknown	3.6990	5/2/2/0/2
CR1-20_CQ	CR1	LINE	-2.9297	No
Dada-1_ON	Dada	DNA	-3.1755	No
Copia3-I_XT	Copia	LTR	-5.6443	No
AlRepD-3555	Unknown	Unknown	-4.5855	No

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2018 doi:10.20944/preprints201803.0145.v1

Peer-reviewed version available at Genes 2018, 9, 269; doi:10.3390/genes9060269

Gonad Female				
AlRepD-1636	Unknown	Unknown	5.5199	23/94/96/104/111
hAT-27_LCh	DNA	DNA	3.0980	No
AlRepD-4141	Unknown	Unknown	6.7706	2/0/0/0/2

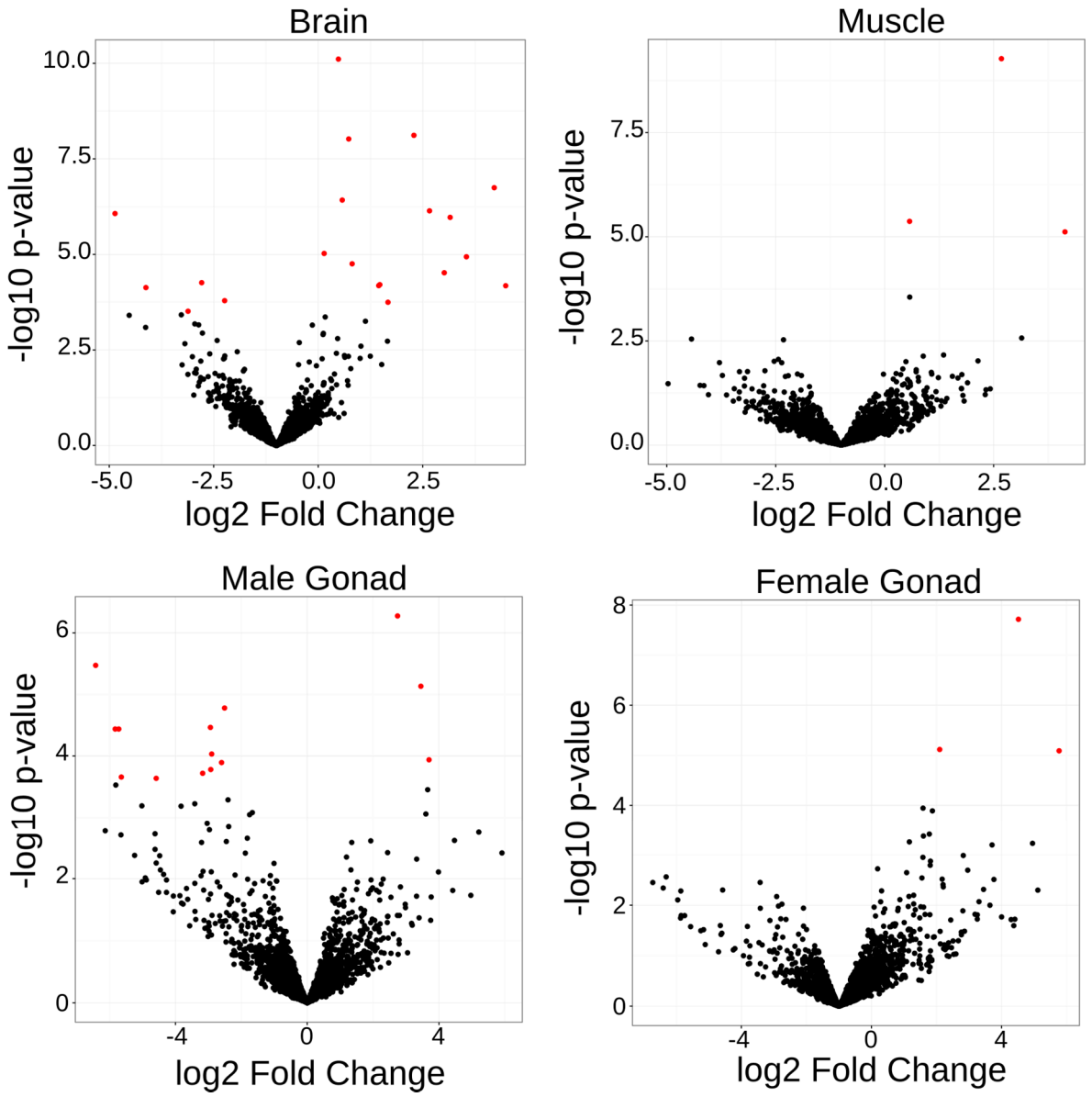

Figure 3. Volcano plots showing differential expression among B- and B+ individuals for analyzed tissues (brain, muscle and male and female gonads). Red points indicate elements with differential expression above threshold (log2 fold-change 1.2 and FDR < 5%). Upregulated repeats on B+ genomes are represented in the right, while downregulated repeats on B+ genomes to the left of each figure.

Although most elements found differentially expressed between B- and B+ genomes are not related to an accumulation on the B chromosome, there are exceptions. For example, element AlRepB-358 (hAT family) is differently expressed on the B+ muscle and has a high copy number on 1B individuals according to the coverage ratio analysis.

The only two elements that have differential expression in two tissues (brain and muscle) are Mariner-1_SP and AlRepC-299. A BLAST on AlRepC-299 reveled 98% similarity with sequences from V2R gene cluster from *Haplochromis chilotes* (accession number AB780556.1) and 97% identity

with sequences from *Oreochromis niloticus*, in a region with no annotation but close to pseudogenes (accession number AB270897.1).

4. Discussion

Genomic organization of TEs in A. latifasciata

We characterized the repetitive content of the reference 0B genome of *A. latifasciata* as a first step to further analysis. DNA transposons were the major TE class found, with retroelements in a smaller fraction. This composition is similar to other teleost fishes, as well as other published African cichlids. In general, teleost fish shows one of the greatest diversity of TEs among vertebrates, with a high number of families populating the genome [52,53]. Among DNA transposons, Tc1-Mariner and hAT (class II TEs) are the most representative, with higher percentage on the genome of *A. latifasciata*. This is similar to other African cichlids, where class II elements are responsible for about 10% of repetitive content, a number observed on *A. latifasciata*. L2 (a LINE retroelement) also predominates on the genome and together with Tc1-Mariner, have the highest genome percentage, both in *A. latifasciata* and other cichlid fish [52,54].

The repeat landscape shows several waves of TE insertion on the genome. We found that LTR, such as Gypsy and Bel/Pao, have entered the genome on early stages. DNA transposons, *e. g.* Tc1-Mariner, have latter insertions, and are accompanied by a large increase in copy number. In general, families with high copy number on the genome of *A. latifasciata* shows prominent insertion waves and have expansion on the B chromosome. Elements with higher number of copies seem to be dominating B genome content, probably due to B's low selective pressure.

Recent insertions of Penelope, Rex-Babar, L2, Tc1-Mariner and hAT elements may indicate recent activity. In fact, these families have several elements with high expression on analyzed tissues. Some copies of these elements may still be active and populating new loci. Furthermore, some families are constant on the landscape, which indicates active mobilization during the evolution of the genome. Our data shows a mark of TEs: their activity is variable trough time, with bursts following purification cycles [55].

Among most transcribed elements, several have high number of copies on the genome. Although transcription is not directly correlated to mobilization events, some expressed elements are related to recent insertion waves; some sequences even shows evidence of open reading frames (ORFs) capable of expressing protein coding sequences (data not shown). Considering that transposition events are a major cause of genomic instability and rearrangement [56], we should consider that, although transcribed, these copies might have lost the capacity to mobilize.

Environmental conditions can alter TEs activity, change their copy number and cause large genome alterations, especially trough modifications in epigenetic silencing. Post-transcriptional alterations, such as siRNA, can silence TEs and limit their copy number increase, making TE transcription not correlated with a transposition event [55,57,58]. Even with several transcribed elements on the genome, post-transcriptional silencing can limit their transposition and copy number alteration.

TEs on the B chromosome

For the first comparison between B- and B+ genome datasets we used RepeatExplorer which is a *de novo* method for element identification and assembly [35,53]. It can be used for comparative quantification between two or more datasets [59]. RepeatExplorer has the advantage of using raw Illumina reads with very low coverage; we chose this method for FISH probe construction and to rapidly test our hypothesis. Coverage information from clusters obtained by RepeatExplorer is indicative of expanded elements on a particular dataset [36]. According to [35], the number of reads in clusters is proportional to the quantity of that element on the genome. Here we used a combination of methods, RepeatExplorer and coverage ratios, to give a broad picture of *A. latifasciata* B chromosome TE composition.

With RepeatExplorer we could identify and validate a few expanded elements on B chromosome. All FISH-mapped elements showed a similar pattern of hybridization, demonstrating a high number of copies on the B and thus corroborating clustering data. We chose to perform a second validation, with qPCR, with two elements present mainly on the B. Gypsy and Bel/Pao were selected for their abundance on the B chromosome, if compared to the A complement. Both analyses showed B expansion of these specific transposable elements.

After our findings of expanded elements on the B chromosome of *A. latifasciata*, our aim was to specify repeat families and their loci; for that, we used the coverage ratios approach. Data from coverage ratios showed the diversity of TE composition on the B of the species, and also corroborated previous findings from RepeatExplorer. Elements detected by RepeatExplorer have more coverage on B+ genomes. Results from coverage ratios are related to the sum of all copies that are at least duplicated on the B chromosome. We choose this conservative threshold due to the inherent variation of repeats copy number. Fluctuations of B blocks copy number are present even among siblings, as found by [29] in *Metriaclima lombardoi*. Our sequencing data comes from different populations and is also prone to variation among individuals, which was evidenced by coverage analysis and qPCR.

Another characteristic detected by RepeatExplorer and by the coverage ratio analysis is the number of other types of sequences on the B chromosome. The V2R gene sequence was found in some contigs assembled by RepeatExplorer, together with HOX and SOX genes fragments. Considering the quantity of retroelements expanded on the B, such elements may have carried a number of sequences during their activity cycles, as they are facilitators of sequence mobilization [60]. Retroelements, such as LINEs, have weak poly-A signals and are good candidates to facilitate mobilization of adjacent sequences. Furthermore, enzymes can also incorporate non-TE mRNA [61,62], leading to the incursion of different types of sequences on the B chromosome. In general, transposable elements take advantage of B chromosome low selective pressure for their insertion and perpetuation [3]. The combination of low selective pressure and drive (a B chromosome characteristic) allows repeats to expand and increase copy number on the B chromosome [9,10]. Such expansion or contraction of sequences on Bs is a landmark of supernumeraries of diverse species [59].

Our data shows evidence of transcription of B chromosome sequences, although only a few TEs have differential expression between B- and B+ genomes. Differentially expressed elements expanded on the B chromosome, usually have a lower coverage difference than highly expanded ones, such as Gypsy-188 and Bel32. Therefore, the majority of expanded elements on B+ genomes do not show differential expression between B- and B+ individuals. Expanded elements may have been present on early stages of B chromosome formation, and we hypothesize that silencing mechanisms targets them. If not silenced, elements with high copy number on the B chromosome could generate high genomic instability [55]. In general, increases in repeat copy number causes their transcription reduction thought silencing mechanisms [63,64]. Elements with high copy number can have high expression, as the hAT family in maize (Vicient 2010). Expression levels also suffer influence from the quantity of B chromosomes in the cell. Higher number of Bs is followed by higher differential expression [25]. Our data is based on 1B samples, therefore small changes in differential expression from B sequences may be difficult to be detected.

Expression of repetitive elements on B chromosomes is highly dependent of the species and repeat localization on supernumerary. Expression is also related to the TE family. In rye, elements expressed on the B chromosome are correlated to their high copy number on the genome, and even modulating transcription of A complement copies [6]. Differential expression among 0B and 4B rye individuals reveled high quantity of expressed B repeats, together with several gene fragments [26]. Gypsy and Mariner elements are expanded and expressed on the B chromosome of the grasshopper *Eyprepocnemis plorans*, with predisposition to euchromatic regions [66]. Considering that methylation patterns vary according to each B chromosome, their transcriptional levels may be altered. DNA methylation inactivates B chromosomes, reducing their effects on the cell [67]. It is plausible that the high copy elements on the B+ genome of *A. latifasciata* have their transcription regulated to avoid

disruptive interference in the cell physiology. Bs from *A. latifasciata* have potential to generate non coding RNA regulatory sequences [68], therefore the regulation of the extra chromosome could help to stabilize the genome.

5. Conclusions

The research of transposable elements, especially on B chromosomes, brings technical and analytical challenges. The understand of their organization on the genome can shed lights on the intrinsic mechanisms of gene and genome regulation. In this regard, this study is one of the steps given to clarify B chromosome organization and structure. We found several expanded elements on the B chromosome of *A. latifasciata*, confirmed through high throughput sequencing, bioinformatics, FISH mapping and qPCR quantification. We have also described a general landscape of repeat copy number, their relative insertion times on the genome and transcriptional levels.

The influence of TEs on shaping the genome is clear and their role on formation and perpetuation of B chromosome requires different approaches, such as the investigation of transcriptional levels of messenger RNAs and noncoding RNAs. We had evidences that the expansion of TEs on the B chromosome can be followed by gain of regulatory mechanisms that control de TE activity. Deeper understanding of epigenetic mechanisms acting on the accessory chromosome is also necessary to determine the level of influence of discovered elements on B.

Supplementary Materials: Figure S1: Repetitive elements transcription levels quantified by RPKM, Table S1: Samples and statistics for the six read datasets alignments against *A. latifasciata* reference genome, Table S2: Metrics of the HPRT gene coverage for the six alignments, Table S3: Ratios between HPRT coverages in reference to M1-0B dataset, Table S4: Primers used for construction of custom FISH probes, Table S5: Primers used for qPCR, Table S6: Copy number estimation for several elements found on *A. latifasciata* genome, File S1: Repeat copy number on the *A. latifasciata* 0B genome, File S2: *A. latifasciata* 0B Landscape, File S3: Clusters from RepeatExplorer, File S4: Coverage ratios normalized to the M1-0B genome as reference, File S5: RPKM values of repetitive elements from brain, muscle and male/female gonads, File S6: Differential expression values between B- and B+ genomes on brain, muscle and male/female gonads, Supplementary Methods: Determination of cut-off for coverage regions.

Acknowledgments: This work was financially supported through grants from the São Paulo Research Foundation (FAPESP) (2013/04533-3; 2014/16763-6; 2015/16661-1) and the National Counsel of Technological and Scientific Development (CNPq) (134446/2014-3; 305321/2015-3).

Author Contributions: R.C. and C.M. conceived and designed the experiments; R.C. performed the experiments; R.C. and C.M. analyzed the data; R.C. and C.M. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Camacho, J. P.; Sharbel, T. F.; Beukeboom, L. W. B-chromosome evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2000**, *355*, 163–78, doi:10.1098/rstb.2000.0556.
2. Yoshida, K.; Terai, Y.; Mizoiri, S.; Aibara, M.; Nishihara, H.; Watanabe, M.; Kuroiwa, A.; Hirai, H.; Hirai, Y.; Matsuda, Y.; Okada, N. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *PLoS Genet.* **2011**, *7*, e1002203, doi:10.1371/journal.pgen.1002203.
3. Houben, A.; Banaei-Moghaddam, A. Evolution and biology of supernumerary B chromosomes. *Cell. Mol. …* **2014**, *71*, 467–78, doi:10.1007/s00018-013-1437-7.
4. Ploskaya-Chaibi, M.; M. Voitovich, A.; V. Novitsky, R.; Bouhadad, R. B-chromosome and V-shaped spot asymmetry in the common frog (Rana temporaria L.) populations. *C. R. Biol.* **2015**, *338*, 161–168, doi:10.1016/j.crvi.2014.12.005.
5. Beukeboom, L. E. O. W. Review article Bewildering Bs◎: an impression of the 1st B-Chromosome Conference. *Heredity (Edinb).* **1994**, *73*, 328–336.
6. Carchilan, M.; Kumke, K.; Mikolajewski, S.; Houben, A. Rye B chromosomes are weakly transcribed and might alter the transcriptional activity of A chromosome sequences. *Chromosoma* **2009**, *118*, 607–16, doi:10.1007/s00412-009-0222-8.
7. Adnađević, T.; Jovanović, V. M.; Blagojević, J.; Budinski, I.; Cabrilo, B.; Bijelić-Čabrilo, O.; Vujošević, M. Possible Influence of B Chromosomes on Genes Included in Immune Response and Parasite Burden in Apodemus flavicollis. *PLoS One* **2014**, *9*, e112260, doi:10.1371/journal.pone.0112260.
8. Valente, G. T.; Nakajima, R. T.; Fantinatti, B. E. A.; Marques, D. F.; Almeida, R. O.; Simões, R. P.; Martins, C. B chromosomes: from cytogenetics to systems biology. *Chromosoma* **2017**, *126*, 73–81, doi:10.1007/s00412-016-0613-6.
9. Banaei-Moghaddam, A. M.; Martis, M. M.; Macas, J.; Gundlach, H.; Himmelbach, A.; Altschmied, L.; Mayer, K. F. X.; Houben, A. Genes on B chromosomes: Old questions revisited with new tools. *Biochim. Biophys. Acta* **2014**, doi:10.1016/j.bbagrm.2014.11.007.
10. Valente, G. T.; Conte, M. a; Fantinatti, B. E. a; Cabral-de-Mello, D. C.; Carvalho, R. F.; Vicari, M. R.; Kocher, T. D.; Martins, C. Origin and Evolution of B Chromosomes in the Cichlid Fish Astatotilapia latifasciata Based on Integrated Genomic Analyses. *Mol. Biol. Evol.* **2014**, doi:10.1093/molbev/msu148.
11. Silva, D. M. Z. D. a; Pansonato-Alves, J. C.; Utsunomia, R.; Araya-Jaime, C.; Ruiz-Ruano, F. J.; Daniel, S. N.; Hashimoto, D. T.; Oliveira, C.; Camacho, J. P. M.; Porto-Foresti, F.; Foresti, F. Delimiting the origin of a B chromosome by FISH mapping, chromosome painting and DNA sequence analysis in Astyanax paranae (Teleostei, Characiformes). *PLoS One* **2014**, *9*, e94896, doi:10.1371/journal.pone.0094896.
12. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–20, doi:10.1038/371215a0.
13. Shapiro, J. a; von Sternberg, R. Why repetitive DNA is essential to genome function. *Biol. Rev. Camb. Philos. Soc.* **2005**, *80*, 227–50.
14. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J. L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; Paux, E.; SanMiguel, P.; Schulman, A. H. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–82, doi:10.1038/nrg2165.
15. Rebollo, R.; Romanish, M. T.; Mager, D. L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **2012**, *46*, 21–42, doi:10.1146/annurev-genet-110711-155621.
16. Bueno, D.; Palacios-Gimenez, O. M.; Cabral-de-Mello, D. C. Chromosomal Mapping of Repetitive DNAs

in the Grasshopper Abracris flavolineata Reveal Possible Ancestry of the B Chromosome and H3 Histone Spreading. *PLoS One* **2013**, *8*, doi:10.1371/journal.pone.0066532.

17. Teruel, M.; Cabrero, J.; Perfectti, F.; Camacho, J. P. M. B chromosome ancestry revealed by histone genes in the migratory locust. *Chromosoma* **2010**, *119*, 217–225, doi:10.1007/s00412-009-0251-3.
18. Cabral-de-Mello, D. C.; Moura, R. C.; Martins, C. Chromosomal mapping of repetitive DNAs in the beetle Dichotomius geminatus provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity (Edinb).* **2010**, *104*, 393–400, doi:10.1038/hdy.2009.126.
19. Fantinatti, B. E. a; Mazzuchelli, J.; Valente, G. T.; Cabral-de-Mello, D. C.; Martins, C. Genomic content and new insights on the origin of the B chromosome of the cichlid fish Astatotilapia latifasciata. *Genetica* **2011**, *139*, 1273–82, doi:10.1007/s10709-012-9629-x.
20. Utsunomia, R.; Silva, D. M. Z. de A.; Ruiz-Ruano, F. J.; Araya-Jaime, C.; Pansonato-Alves, J. C.; Scacchetti, P. C.; Hashimoto, D. T.; Oliveira, C.; Trifonov, V. A.; Porto-Foresti, F.; Camacho, J. P. M.; Foresti, F. Uncovering the Ancestry of B Chromosomes in Moenkhausia sanctaefilomenae (Teleostei, Characidae). *PLoS One* **2016**, *11*, e0150573, doi:10.1371/journal.pone.0150573.
21. Silva, D. M. Z.; Daniel, S. N.; Camacho, J. P. M.; Utsunomia, R.; Ruiz-Ruano, F. J.; Penitente, M.; Pansonato-Alves, J. C.; Hashimoto, D. T.; Oliveira, C.; Porto-Foresti, F.; Foresti, F. Origin of B chromosomes in the genus Astyanax (Characiformes, Characidae) and the limits of chromosome painting. *Mol. Genet. Genomics* **2016**, *291*, 1407–18, doi:10.1007/s00438-016-1195-y.
22. Martis, M. M.; Klemme, S.; Banaei-Moghaddam, A. M.; Blattner, F. R.; Macas, J.; Schmutzer, T.; Scholz, U.; Gundlach, H.; Wicker, T.; Šimková, H.; Novák, P.; Neumann, P.; Kubaláková, M.; Bauer, E.; Haseneyer, G.; Fuchs, J.; Doležel, J.; Stein, N.; Mayer, K. F. X.; Houben, A. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13343–6, doi:10.1073/pnas.1204237109.
23. Ziegler, C. G.; Lamatsch, D. K.; Steinlein, C.; Engel, W.; Schartl, M.; Schmid, M. The giant B chromosome of the cyprinid fish Alburnus alburnus harbours a retrotransposon-derived repetitive DNA sequence. *Chromosom. Res.* **2003**, *11*, 23–35, doi:10.1023/A:1022053931308.
24. Lamb, J. C.; Riddle, N. C.; Cheng, Y. M.; Theuri, J.; Birchler, J. A. Localization and transcription of a retrotransposon-derived element on the maize B chromosome. *Chromosom. Res.* **2007**, *15*, 383–398, doi:10.1007/s10577-007-1135-0.
25. Huang, W.; Du, Y.; Zhao, X.; Jin, W. B chromosome contains active genes and impacts the transcription of A chromosomes in maize (Zea mays L.). *BMC Plant Biol.* **2016**, *16*, 88, doi:10.1186/s12870-016-0775-7.
26. Ma, W.; Gabriel, T. S.; Martis, M. M.; Gursinsky, T.; Schubert, V.; Vrána, J.; Doležel, J.; Grundlach, H.; Altschmied, L.; Scholz, U.; Himmelbach, A.; Behrens, S.; Banaei-Moghaddam, A. M.; Houben, A. Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its A chromosome paralog. *New Phytol.* **2016**, doi:10.1111/nph.14110.
27. Poletto, A. B.; Ferreira, I. A.; Martins, C. The B chromosomes of the African cichlid fish Haplochromis obliquidens harbour 18S rRNA gene copies. *BMC Genet.* **2010**, *11*, 1, doi:10.1186/1471-2156-11-1.
28. Kuroiwa, a; Terai, Y.; Kobayashi, N.; Yoshida, K.; Suzuki, M.; Nakanishi, a; Matsuda, Y.; Watanabe, M.; Okada, N. Construction of chromosome markers from the Lake Victoria cichlid Paralabidochromis chilotes and their application to comparative mapping. *Cytogenet. Genome Res.* **2014**, *142*, 112–20, doi:10.1159/000356128.
29. Clark, F. E.; Conte, M. a.; Ferreira-Bravo, I. a; Poletto, A. B.; Martins, C.; Kocher, T. D. Dynamic Sequence

Evolution of a Sex-Associated B Chromosome in Lake Malawi Cichlid Fish. *J. Hered.* **2017**, *108*, 53–62, doi:10.1093/jhered/esw059.

30. Fantinatti, B. E. A.; Martins, C. Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish Astatotilapia latifasciata as a model. *BMC Genet.* **2016**, *17*, 119, doi:10.1186/s12863-016-0427-9.
31. Jehangir, M. Genome assembly of the cichlid fish Astatotilapia latifasciata with focus in population genomics of B chromosome polymorphism Available online: http://hdl.handle.net/11449/151740.
32. Smit, A.; Hubley, R. RepeatModeler Open-1.0 2015.
33. Bao, W.; Kojima, K. K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11, doi:10.1186/s13100-015-0041-9.
34. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0 2013.
35. Novak, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; Macas, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **2013**, *29*, 792–793, doi:10.1093/bioinformatics/btt054.
36. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **2010**, *11*, 378, doi:10.1186/1471-2105-11-378.
37. Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **2012**, *9*, 357–359, doi:10.1038/nmeth.1923.
38. Andrews, S. Fastqc 2016.
39. Hannon FASTX-Toolkit 2010.
40. Staton, E. Pairfq 2016.
41. Garcia-Alcalde, F.; Okonechnikov, K.; Carbonell, J.; Cruz, L. M.; Gotz, S.; Tarazona, S.; Dopazo, J.; Meyer, T. F.; Conesa, A. Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **2012**, *28*, 2678–2679, doi:10.1093/bioinformatics/bts503.
42. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. In *Current Protocols in Bioinformatics*; Wiley Online Library, 2014; Vol. 0 ISBN 0471250937.
43. Marques, D. F. Functional analysis of B chromosome presence using cichlid Astatotilapia latifasciata as model Available online: http://hdl.handle.net/11449/141953.
44. Criscione, S. W.; Zhang, Y.; Thompson, W.; Sedivy, J. M.; Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **2014**, *15*, 583, doi:10.1186/1471-2164-15-583.
45. Gene Link Oligo Explorer 1.2 2005.
46. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **2012**, *13*, 134, doi:10.1186/1471-2105-13-134.
47. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **2000**, *28*, 1102,1104.
48. Bertollo, L.; Takahashi, C.; Moreira-Filho, O. Citotaxonomic consideration on Hoplias lacerdae (Pisces, Erythrinidae). *Braz J Genet* **1978**, 103–120.
49. Pinkel, D.; Straume, T.; Gray, J. W. Cytogenetic analysis using quantitative , high-sensitivity , fluorescence hybridization. *PNAS* **1986**, *83*, 2934–2938.
50. Bel, Y.; Ferré, J.; Escriche, B.; Redon, R.; Ishikawa, S.; Fitch, K.; Feuk, L.; Perry, G.; Andrews, T.; Fiegler, H.; Shapero, M.; Carson, A.; Chen, W.; Cho, E.; Dallaire, S.; Freeman, J.; Gonzalez, J.; Gratacos, M.; Huang, J.;

Kalaitzopoulos, D.; Komura, D.; MacDonald, J.; Marshall, C.; Mei, R.; Montgomery, L.; Nishimura, K.; Okamura, K.; Shen, F.; Somerville, M.; Tchinda, J.; Valsesia, A.; Woodwark, C.; Yang, F.; Zhang, J.; Zerjal, T.; Zhang, J.; Armengol, L.; Conrad, D.; Estivill, X.; Tyler-Smith, C.; Carter, N.; Aburatani, H.; Lee, C.; Jones, K.; Scherer, S.; Hurles, M.; Knuutila, S.; Bjorkqvist, A.; Autio, K.; Tarkkanen, M.; Wolf, M.; Monni, O.; Szymanska, J.; Larramendy, M.; Tapper, J.; Pere, H.; El-Rifai, W.; Hemmer, S.; Wasenius, V.; Vidgren, V.; Zhu, Y.; Knuutila, S.; Aalto, Y.; Autio, K.; Bjorkqvist, A.; El-Rifai, W.; Hemmer, S.; Huhta, T.; Kettunen, E.; Kiuru-Kuhlefelt, S.; Larramendy, M.; Lushnikova, T.; Monni, O.; Pere, H.; Tapper, J.; Tarkkanen, M.; Varis, A.; Wasenius, V.; Wolf, M.; Zhu, Y.; Vontas, J.; Small, G.; Hemingway, J.; Field, L.; Blackman, R.; Tyler-Smith, C.; Devonshire, A.; Mouches, C.; Pasteur, N.; Berge, J.; Hyrien, O.; Raymond, M.; de, S.; de, S.; Georghiou, G.; Chiang, P.; Song, W.; Wu, K.; Korenberg, J.; Fogel, E.; Keuren, M. Van; Lashkari, D.; Kurnit, D.; Chiang, P.; Wei, W.; Gibson, K.; Bodmer, R.; Kurnit, D.; Bustin, S.; Benes, V.; Garson, J.; Hellemans, J.; Huggett, J.; Kubista, M.; Mueller, R.; Nolan, T.; Pfaffl, M.; Shipley, G.; Vandesompele, J.; Wittwer, C.; Chiang, P.; Kurnit, D.; Preter, K. De; Speleman, F.; Combaret, V.; Lunec, J.; Laureys, G.; Eussen, B.; Francotte, N.; Board, J.; Pearson, A.; De, P.; Van, R.; Vandesompele, J.; Papin, J.; Vahrson, W.; Dittmer, D.; Paton, M.; Karunaratne, S.; Giakoumaki, E.; Roberts, N.; Hemingway, J.; Gahan, L.; Gould, F.; Heckel, D.; Xu, X.; Yu, L.; Wu, Y.; Ferré, J.; Rie, J. Van; MacIntosh, S.; Bel, Y.; Siqueira, H.; Siegfried, B.; Ferre, J.; Escriche, B.; Gujar, G.; Khawale, R.; Kalia, V.; Zhao, J.; Jin, L.; Yang, Y.; Wu, Y.; Baxter, S.; Zhao, J.; Gahan, L.; Shelton, A.; Tabashnik, B.; Heckel, D.; Lindsley, D.; Zimm, G.; Tay, W.; Behere, G.; Heckel, D.; Lee, S.; Batterham, P.; Coates, B.; Sumerford, D.; Hellmich, R.; Lewis, L.; Bel, Y.; Escriche, B.; Murthy, S.; Magliocco, A.; Demetrick, D.; Poitout, S.; Bues, R.; Dorak, M. Quantitative real-time PCR with SYBR Green detection to assess gene duplication in insects: study of gene dosage in Drosophila melanogaster (Diptera) and in Ostrinia nubilalis (Lepidoptera). *BMC Res. Notes* **2011**, *4*, 84, doi:10.1186/1756-0500-4-84.

51. Chalopin, D.; Fan, S.; Simakov, O.; Meyer, A.; Schartl, M.; Volff, J. N. Evolutionary active transposable elements in the genome of the coelacanth. *J. Exp. Zool. Part B Mol. Dev. Evol.* **2014**, *322*, 322–333, doi:10.1002/jez.b.22521.
52. Chalopin, D.; Naville, M.; Plard, F.; Galiana, D.; Volff, J. N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **2015**, *7*, 567–580, doi:10.1093/gbe/evv005.
53. Sotero-Caio, C. G.; Platt, R. N.; Suh, A.; Ray, D. A. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* **2017**, *9*, 161–177, doi:10.1093/gbe/evw264.
54. Brawand, D.; Wagner, C. E.; Li, Y. I.; Malinsky, M.; Keller, I.; Fan, S.; Simakov, O.; Ng, A. Y.; Lim, Z. W.; Bezault, E.; Turner-Maier, J.; Johnson, J.; Alcazar, R.; Noh, H. J.; Russell, P.; Aken, B.; Alföldi, J.; Amemiya, C.; Azzouzi, N.; Baroiller, J.-F.; Barloy-Hubler, F.; Berlin, A.; Bloomquist, R.; Carleton, K. L.; Conte, M. a.; D'Cotta, H.; Eshel, O.; Gaffney, L.; Galibert, F.; Gante, H. F.; Gnerre, S.; Greuter, L.; Guyon, R.; Haddad, N. S.; Haerty, W.; Harris, R. M.; Hofmann, H. a.; Hourlier, T.; Hulata, G.; Jaffe, D. B.; Lara, M.; Lee, A. P.; MacCallum, I.; Mwaiko, S.; Nikaido, M.; Nishihara, H.; Ozouf-Costaz, C.; Penman, D. J.; Przybylski, D.; Rakotomanga, M.; Renn, S. C. P.; Ribeiro, F. J.; Ron, M.; Salzburger, W.; Sanchez-Pulido, L.; Santos, M. E.; Searle, S.; Sharpe, T.; Swofford, R.; Tan, F. J.; Williams, L.; Young, S.; Yin, S.; Okada, N.; Kocher, T. D.; Miska, E. a.; Lander, E. S.; Venkatesh, B.; Fernald, R. D.; Meyer, A.; Ponting, C. P.; Streelman, J. T.; Lindblad-Toh, K.; Seehausen, O.; Di Palma, F. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **2014**, *513*, 375–81, doi:10.1038/nature13726.
55. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **2014**, *27*, 2573–84, doi:10.1111/jeb.12513.

663 56. Hedges, D. J.; Deininger, P. L. Inviting instability: Transposable elements, double-strand breaks, and the
664 maintenance of genome integrity. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **2007**, *616*, 46–59,
665 doi:10.1016/j.mrfmmm.2006.11.021.
666 57. Feschotte, C.; Jiang, N.; Wessler, S. R. Plant transposable elements: where genetics meets genomics. *Nat.*
667 *Rev. Genet.* **2002**, *3*, 329–41, doi:10.1038/nrg793.
668 58. Slotkin, R. K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat.*
669 *Rev. Genet.* **2007**, *8*, 272–85, doi:10.1038/nrg2072.
670 59. Klemme, S.; Banaei-Moghaddam, A. M.; Macas, J.; Wicker, T.; Novák, P.; Houben, A.; Banaei-
671 Moghaddam, A. High-copy sequences reveal distinct evolution of the rye B chromosome. *New …* **2013**,
672 *199*, 550–8, doi:10.1111/nph.12289.
673 60. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **2010**, *20*, 1313–1326,
674 doi:10.1101/gr.101386.109.
675 61. Maestre, J.; Tchénio, T.; Dhellin, O.; Heidmann, T. mRNA retroposition in human cells: processed
676 pseudogene formation. *EMBO J.* **1995**, *14*, 6333–8.
677 62. Moran, J. V.; DeBerardinis, R. J.; Kazazian, H. H. Exon shuffling by L1 retrotransposition. *Science* **1999**, *283*,
678 1530–4, doi:10.1126/science.283.5407.1530.
679 63. Meyers, B. C.; Tingey, S. V; Morgante, M.; Meyers, B. C.; Tingey, S. V; Morgante, M. Abundance ,
680 Distribution , and Transcriptional Activity of Repetitive Elements in the Maize Genome Abundance ,
681 Distribution , and Transcriptional Activity of Repetitive Elements in the Maize Genome. **2001**, 1660–1676,
682 doi:10.1101/gr.188201.
683 64. González, L. G.; Deyholos, M. K. Identification, characterization and distribution of transposable elements
684 in the flax (Linum usitatissimum L.) genome. *BMC Genomics* **2012**, *13*, 644, doi:10.1186/1471-2164-13-644.
685 65. Vicient, C. M. Transcriptional activity of transposable elements in maize. *BMC Genomics* **2010**, *11*, 601,
686 doi:10.1186/1471-2164-11-601.
687 66. Montiel, E. E.; Cabrero, J.; Camacho, J. P. M.; López-León, M. D. Gypsy, RTE and Mariner transposable
688 elements populate Eyprepocnemis plorans genome. *Genetica* **2012**, *140*, 365–374,
689 doi:10.1007/s10709-012-9686-1.
690 67. Bugno-Poniewierska, M.; Solek, P.; Wronski, M.; Potocki, L.; Jezewska-Witkowska, G.; Wnuk, M. Genome
691 organization and DNA methylation patterns of B chromosomes in the red fox and Chinese raccoon dogs.
692 *Hereditas* **2014**, *8*, 1–8, doi:10.1111/hrd2.00049.
693 68. Ramos, É.; Cardoso, A. L.; Brown, J.; Marques, D. F.; Fantinatti, B. E. A.; Cabral-de-Mello, D. C.; Oliveira,
694 R. A.; O'Neill, R. J.; Martins, C. The repetitive DNA element BncDNA, enriched in the B chromosome of
695 the cichlid fish Astatotilapia latifasciata, transcribes a potentially noncoding RNA. *Chromosoma* **2016**,
696 doi:10.1007/s00412-016-0601-x.
697