Origin and Evolution of B Chromosomes in the Cichlid Fish *Astatotilapia latifasciata* Based on Integrated Genomic Analyses

Guilherme T. Valente,[1] Matthew A. Conte,[2] Bruno E.A. Fantinatti,[1] Diogo C. Cabral-de-Mello,[3] Robson F. Carvalho,[1] Marcelo R. Vicari,[4] Thomas D. Kocher,[2] and Cesar Martins*[,1]

[1]Department of Morphology, Biosciences Institute, São Paulo State University, São Paulo, Brazil

[2]Department of Biology, University of Maryland, College Park

[3]Department of Biology, Biosciences Institute, São Paulo State University, São Paulo, Brazil

[4]Department of Structural and Molecular Biology and Genetics, State University of Ponta Grossa, Paraná, Brazil

***Corresponding author:** E-mail: cmartins@ibb.unesp.br, cmartinsunesp@gmail.com.

Associate editor: Miriam Barlow

Abstract

Approximately 15% of eukaryotes contain supernumerary B chromosomes. When present, B chromosomes frequently represent as much as 5% of the genome. Despite thousands of reports describing the distribution of supernumeraries in various taxa, a comprehensive theory for the origin, maintenance, and evolution of B chromosomes has not emerged. Here, we sequence the complete genomes of individual cichlid fish (*Astatotilapia latifasciata*) with and without B chromosomes, as well as microdissected B chromosomes, to identify DNA sequences on the B. B sequences were further analyzed through quantitative polymerase chain reaction and in situ hybridization. We find that the B chromosome contains thousands of sequences duplicated from essentially every chromosome in the ancestral karyotype. Although most genes on the B chromosome are fragmented, a few are largely intact, and we detect evidence that at least three of them are transcriptionally active. We propose a model in which the B chromosome originated early in the evolutionary history of Lake Victoria cichlids from a small fragment of one autosome. DNA sequences originating from several autosomes, including protein-coding genes and transposable elements, subsequently inserted into this proto-B. We propose that intact B chromosome genes involved with microtubule organization, kinetochore structure, recombination and progression through the cell cycle may play a role in driving the transmission of the B chromosome. Furthermore, our work suggests that karyotyping is an essential step prior to genome sequencing to avoid problems in genome assembly and analytical biases created by the presence of high copy number sequences on the B chromosome.

Key words: **chromosome evolution, genome evolution, next generation sequencing, supernumerary chromosome.**

Introduction

Eukaryote genomes are composed of the normal (A) set of chromosomes, and extra, accessory, or supernumerary (B) chromosomes that do not follow Mendelian laws of inheritance. B chromosomes were first described more than a century ago (Wilson 1907) and have since been observed in approximately 15% of eukaryotic species that have been karyotyped (Camacho 2005). They are not essential for life and therefore can be considered "dispensable" additional chromosomes (White 1973; Jones 1991; Camacho et al. 2000; Palestis et al. 2004; Jones et al. 2008). Although B chromosomes have been observed in more than 500 animal species (Camacho 2005), major questions concerning their origin, evolution and function remain unanswered.

The extensive distribution of B chromosomes among diverse taxa might suggest beneficial effects on their hosts, but there is no strong evidence to support the idea that B chromosomes contribute to host fitness. Current theory suggests that B chromosomes arise from the A chromosomes and are maintained as parasitic elements or "selfish" DNA (Burt and Trivers 2006). Extra chromosomes are typically deleterious because they create unbalanced gene dosage. Furthermore, unpaired chromosomes are typically lost during cell division. The survival of B chromosomes therefore depends on having a transmission advantage during meiosis (meiotic drive) (Camacho 2005). Because they do not undergo recombination with other chromosomes, B chromosomes are inherited clonally. The evolutionary dynamics of mutation and weak selection lead to the accumulation of large amounts of repetitive DNA sequence on the B. This high content of repetitive elements makes it difficult to trace the origin of B chromosomes from specific A chromosomes.

Previous studies of B chromosomes have mostly used classical and molecular cytogenetic techniques with relatively limited resolution. An exception is a recent study that sequenced B chromosomes of rye (*Secale cereale*) that had been isolated from the A genome by flow cytometry (Martis et al. 2012). Analysis of these sequences showed that rye B chromosomes were mostly derived from two autosomes, with subsequent accumulation of sequences from the other autosomes and organelle genomes. Related work suggests the mechanism of B chromosome drive in this species involves a combination of nondisjunction and unequal spindle formation (Banaei-Moghaddam et al. 2012). Furthermore, recent studies also demonstrate that B

Article

chromosomes harbor transcriptionally active DNA sequences (Banaei-Moghaddam et al. 2013; Trifonov et al. 2013) that could benefit their hosts.

Cichlid fishes are best known for their astonishing adaptive radiations in East Africa, but they have also become an interesting model to study B chromosome biology. B chromosomes have been identified in 7 South American and 14 African species of cichlid fish (Poletto, Ferreira, Cabral-de-Mello, et al. 2010; Poletto, Ferreira, Martins 2010; Fantinatti et al. 2011; Yoshida et al. 2011; Kuroiwa et al. 2014). Among the African species, B chromosomes were first described in *Astatotilapia latifasciata* from Lake Nawampasa, a satellite lake of the Lake Kyoga system (Poletto, Ferreira, Martins 2010). Supernumerary chromosomes have been detected in all the Victoria cichlids karyotyped, which includes in total thirteen species (Poletto, Ferreira, Martins 2010; Yoshida et al. 2011; Kuroiwa et al. 2014). On the other hand, among 14 karyotyped species of Lake Malawi, B chromosomes were detected in only one species, *Metriaclima lombardoi* (Poletto, Ferreira, Cabral-de-Mello, et al. 2010). Cytogenetic analyses found no B chromosomes in another eight African and Asian cichlids (Poletto, Ferreira, Cabral-de-Mello, et al. 2010).

B chromosomes in *A. latifasciata* have been studied with a focus on cytogenetics (Poletto, Ferreira, Martins 2010; Fantinatti et al. 2011). Both sexes of *A. latifasciata* can have either one or two cytogenetically similar B chromosomes. In situ hybridization shows that these chromosomes have many repetitive DNA sequences and show no sex-specific differences (Poletto, Ferreira, Martins 2010; Fantinatti et al. 2011). Advances in genomic sequencing have created new opportunities for studying classical problems in cytogenetics. Here, we investigate the B chromosomes of *A. latifasciata* using high-throughput sequencing to better understand their gene content, origin, and pattern of evolution. Our data show that the B chromosome is a genomic mosaic composed of degenerated sequences derived from most of the A chromosome set. We also find evidence of transcriptionally active genes from the B chromosome. The presence of apparently intact genes involved with the cell cycle suggests mechanisms by which the B chromosome escaped elimination during meiosis and became established in *A. latifasciata* populations. Moreover, we suggest the B chromosomes of *A. latifasciata* and Lake Victoria cichlids may have a common origin.

Results

Next-Generation Sequencing and Quantitative Polymerase Chain Reaction

Illumina HiSeq sequencing of whole genomes generated a total of 306,823,512 reads from the B+ individual and 401,017,570 reads from the B− individual. A total of 263,502,665 (85.88%) B+ reads and 350,283,423 (87.35%) B− reads were aligned to the *M. zebra* reference assembly using Bowtie2 (Langmead and Salzberg 2012). This represents 22.92× and 30.99× coverage (mean coverage ratio of 0.74) of the respective genomes over the entire length (848,776,495 bp) of the reference genome assembly. Regions along the genome that had significantly higher B+ coverage than B− coverage were identified (see Materials and Methods) and are referred to as B chromosome "blocks." The exact number of B chromosome blocks, as well as their exact boundaries in the reference sequence, varies depending on the parameters adopted for the analysis (table 1; supplementary data set S1, Supplementary Material online; see Materials and Methods).

Table 1. Gene Lists Obtained After the Establishment of Four Sets of Blocks under Different Conditions of Coverage Ratios (≥1.48, 0 stdv; ≥1.88, 0 stdv + 2 stdv) and Block Size Tolerances (≤100 bp; ≤1 kb).

Gene Lists	Number of Blocks	Total Length of Blocks (bp)	Number of Genes
0 stdv/≤100 bp	9,347	2,943,058	5,312
0 stdv/≤1 kb	28,872	18,664,260	16,540
+ 2 stdv/≤100 bp	2,478	1,323,254	1,061
+ 2 stdv/≤1 kb	3,491	3,350,071	2,476

A total of 125,601 reads totaling 48,637,895 bp were obtained from the microdissected B chromosome using the 454 Life Sciences platform. This was expected to correspond to approximately 0.8X coverage of *A. latifasciata* B chromosome. However, a large fraction (96.4%) of the reads were most similar to human, which may indicate contamination during the enzymatic amplification of the single microdissected chromosome. The contaminated reads were filtered, and the remaining reads (0.57%) were most similar to the sequence of cichlids or closely related fish species. These sequences correspond to 209,181 bp in the *M. zebra* reference genome assembly. The 454 data are in general agreement with our Illumina data and confirm many of the B sequence blocks (fig. 1*A*).

We used quantitative polymerase chain reaction (qPCR) for relative copy number quantification of selected B sequence blocks in fish with 0, 1, or 2 B genomes. The qPCR estimates of copy number are tightly related with the B chromosome number determined from chromosome spreads (fig. 1*A*, supplementary fig. S1 and data set S2, Supplementary Material online). Moreover, the qPCR results were highly correlated with the mean coverage ratio (2B/0B) for these regions ($R^2 = 0.99$). Because we sampled many individuals, these analyses also demonstrate that the B blocks are not related to copy number variation on the autosomes.

Comparative Genomic Analysis

Mapping of the B blocks on the genetically anchored map of the *M. zebra* reference genome (O'Quin et al. 2013) demonstrates that the B chromosome contains DNA sequences derived from almost all linkage groups. The largest number of blocks map to linkage groups 1 (165,284 bp, ~11.5% of B), 3 (177,225 bp, ~12.3% of B), and 9 (146,110 bp, ~10.1% of the B). Because the anchored map is still incomplete, the greatest fraction of B sequences map to scaffolds not yet placed in any linkage group (347,198 bp, ~24.1% of B) (fig. 1*B*).

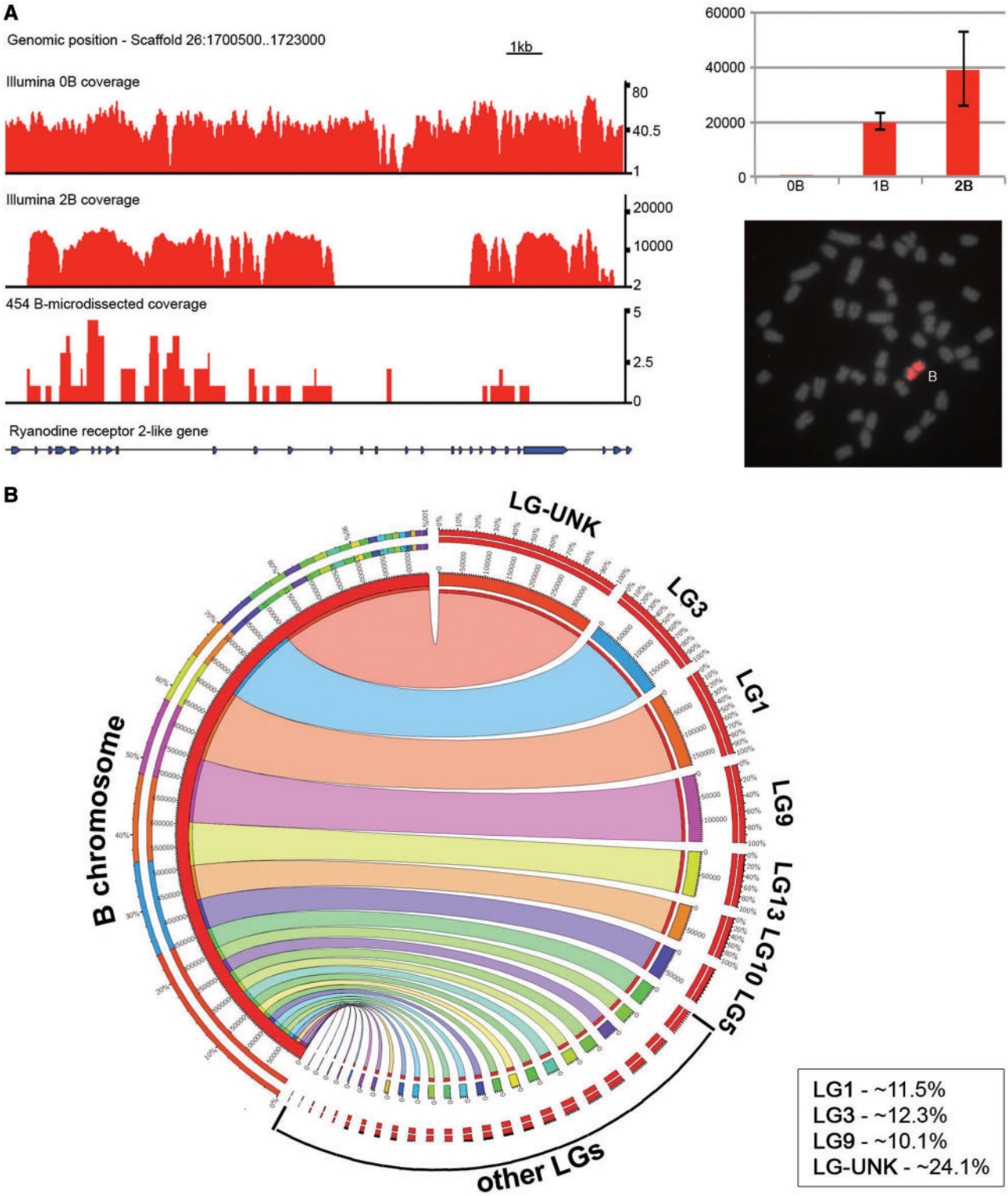

Fig. 1. Comparative genomic analyses of B sequences. (*A*) Genomic analysis of scaffold 26. Read coverage of representative regions on scaffold 26 of 0B and 2B genomes and the microdissected B (note the coverage as indicated on the *Y* axis). The gene annotation is also indicated. The bar graph and the metaphasic plate (right) represent copy number by qPCR and FISH mapping, respectively. (*B*) Relationship of the B blocks to the LGs of *Metriaclima zebra*. B blocks are shown on the left and corresponding LGs on the right. The genomic percentage of the major LGs (1, 3, and 9) represented on the B is highlighted in the box.

Yoshida et al. (2011) used DNA sequences derived from the B chromosome of the Lake Victoria African cichlid *Lithochromis rubripinnis* to identify and sequence a B-derived clone from a *Haplochromis chilotes* bacterial artificial chromosome (BAC) library. Alignment of the 18 sequence contigs from this BAC to the *M. zebra* genome revealed a close correspondence to B chromosome blocks of *A. latifasciata*. Moreover, the alignment of B+ and B− Illumina reads

from *A. latifasciata* shows higher coverage from the B+ than the B− sample over the regions corresponding to the 18 *H. chilotes* BAC contigs (supplementary fig. S2, Supplementary Material online).

An examination of the read coverage from the *Pundamilia nyererei* (also from Lake Victoria) genome project (Broad Institute, http://www.ncbi.nlm.nih.gov/bioproject/60367, last accessed May 7, 2014) reveals that this species also contains the same high coverage blocks from the B chromosome (supplementary fig. S2, Supplementary Material online).

Repetitive Elements and Gene Search

A search for repetitive DNAs in the *A. latifasciata* B chromosome blocks found 1,216 repeats distributed across 623 blocks with at least one repetitive element. Approximately 19.7% and 18.7% of those sequences are DNA transposons and retrotransposons, respectively, whereas approximately 61.5% are unknown, simple, and low complexity repeats (fig. 2*A*). Among the transposable elements (TEs), DNA transposons and LINEs are the most abundant.

We used the read coverage data to define blocks of B chromosome sequences using two different coverage criteria (coverage ratios $\geq$1.48 and $\geq$1.88) and two block size tolerances ($\leq$100 bp and $\leq$1 kb). The gene content of each block was then determined from the annotation of the *M. zebra* reference genome. Table 1 and figure 2*B* show the number of genes identified in each of the four resulting gene lists. The combined lists identified a total of 5,858 genes, which constituted the combined list of putative B-genes (B combined list [BCL]) (supplementary data set S3, Supplementary Material online). In most cases, the B chromosome block is continuous across the gene or gene fragment, indicating that a duplicated DNA fragment of the A genome was incorporated into the B. In a few cases, the regions of high coverage corresponded to just the exons of a gene, suggesting a reverse transcription event followed by insertion in the B chromosome. For example, the *hnRNP Q*-like (heterogeneous nuclear ribonucleoprotein Q-like) gene shows the characteristics of a retrogene (supplementary fig. S3, Supplementary Material online).

Physical Mapping of B Sequences

Previous fluorescence in situ hybridization (FISH) mapping found that the B chromosome is enriched with 18S rDNA, Rex1, and Rex3 TEs, among other repeat elements (Poletto, Ferreira, Martins 2010; Fantinatti et al. 2011). Diverse classical and molecular cytogenetic analyses show that this B chromosome is the only large metacentric chromosome and corresponds to an isochromosome in which the two chromosome arms are highly similar (figs. 1*A* and 2*C*; supplementary fig. S1, Supplementary Material online). Moreover, sequences from the B chromosome blocks were physically mapped over B chromosome spreads. qPCR results show that the B chromosome is enriched for those sequences (fig. 1*A*; supplementary fig. S1, Supplementary Material online).

Function and Evolution of High-Integrity Genes

To identify the intact genes on the B chromosomes, we calculated an integrity score for each gene sequence in the combined list. The majority of BCL genes (96.8%) have integrity scores less than 50% (fig. 2*D*). The remaining 3.2% of genes with integrity scores more than 50% are listed in supplementary data set S4, Supplementary Material online.

Approximately 59% of the genes in BCL and approximately 66% in *M. zebra* gene sets were annotated with Gene Ontologies (GO) (supplementary fig. S4, Supplementary Material online). A detailed functional analysis (GO, Uniprot [www.uniprot.org, last accessed May 7, 2014] and National Center for Biotechnology Information [NCBI, www.ncbi.nlm.nih.gov, last accessed May 7, 2014] descriptions) of the BCL genes with an integrity $\geq$70% was performed and allows us to label these genes according their major functions (supplementary data set S4, Supplementary Material online). The GO enrichment analysis comparing BCL and *M. zebra* GOs identified an enrichment for terms related to cell adhesion, signaling, actin polymerization, neurological aspects, ion transport, and mainly several basic cell process in the relatively intact genes (integrity $\geq$ 70%) (supplementary fig. S5, Supplementary Material online). Among the high-integrity B chromosome genes, there are several genes involved with cell cycle control (table 2). Comparative analysis of these particular genes revealed the presence of an A copy of the gene shared among cichlid species and a B-specific copy only observed in *A. latifasciata* and *P. nyererei* (supplementary fig. S6, Supplementary Material online). Molecular phylogenetic analysis of these high-integrity genes showed that most genes detected in the B chromosome of *A. latifasciata* group with *P. nyererei* putative B−sequences (supplementary fig. S7, Supplementary Material online).

An analysis of transcriptome sequences from *P. nyererei* (available at Bouillabase.org) revealed that this species contains B-encoded transcript variants of the *Separin*, *TUBB1*, and *KIF11* genes in several tissues (supplementary fig. S8, Supplementary Material online). The selective pressure analysis showed that all three sequences tested are under neutral selection (supplementary table S1, Supplementary Material online).

Discussion

Our analyses show that the B chromosomes of *A. latifasciata* are gene rich and composed of DNA sequences accreted from almost all of the autosomes. They have a TE composition similar to that found in the rest of the genome. We now propose a model for the origin and evolution of this chromosome type and argue that some intact genes on the B play a role in maintaining the B chromosome. We further discuss the relationship between B chromosomes and sex determination.

Chromosomal Origin of the B

Our data clearly show that almost all the autosomes have contributed sequences to the B chromosome through a process of gene duplication that might be mediated by mobile

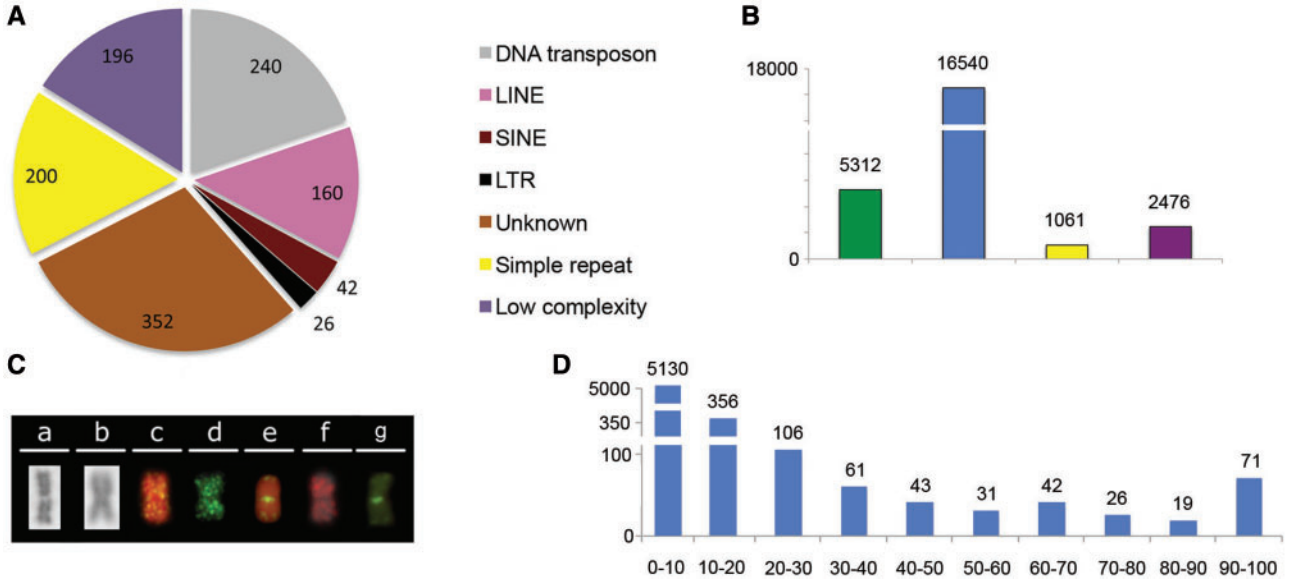

Fig. 2. B genomic content. (*A*) Number and types of TEs in the B chromosome. (*B*) Analysis of the B gene lists. *Y* axis, number of genes: Green, 0 stdv/100 bp; blue, 0 sdtv/1 kb; yellow, +2 stdv/100 bp; and purple, +2 stdv/1 kb. (*C*) B chromosomes cytogenetic analysis: (a) Giemsa staining, (b) C-banding, FISH mapping of (c) Rex1, (d) Rex3, (e) 45 S rDNA and (f) BAC b03TI066P02 probes, and (g) CMA_3 staining. (*D*) Results of gene integrity analysis showing the number of genes (*Y* axis) in each integrity percentage bins (*X* axis).

Table 2. High-Integrity Genes Involved with Cell Division Detected in the B Chromosome Blocks.

Integrity (%)	GenBank	Gene Description
70.93	XM_004570740.1	DNA repair protein *XRCC2*-like (LOC101476596), mRNA
70.97	XM_004570733.1	SC protein (*SYCP*) 2-like (LOC101474742), mRNA
73.92	XM_004575607.1	Centromere-associated protein (*CENP*) E-like (LOC101466893), mRNA
75.71	XM_004545355.1	*Separin*-like (LOC101471736), mRNA
77.16	XM_004547754.1	Aurora kinase (*AURK*) A-B-like (LOC101467173), mRNA
96.14	XM_004570623.1	Kinesin-like protein *KIF11*-like (LOC101466420), mRNA
97.50	XM_004572890.1	Tubulin beta-5 (*TUBB5*) chain-like (LOC101478022), mRNA
100	XM_004547671.1	Tubulin beta-1 (*TUBB1*) chain-like (LOC101468168), mRNA
100	XM_004556260.1	Spindle and kinetochore-associated (*SKA*) protein 1-like (LOC101476636), mRNA
100	XM_004576295.1	Regulator of telomere elongation helicase (*RTEL*) 1-like (LOC101468303), mRNA

elements or retrogene insertions. Linkage groups 1, 3, and 9 are the main sequence donors, which might indicate that these linkage groups were the source of the proto-B or which might simply reflect the relative size of these donor chromosomes. This is in contrast to the results for rye, where the B is mostly derived from just two autosomes (Martis et al. 2012). Previous papers have suggested that B chromosomes incorporate sequences from multiple A chromosomes (Alfenito and Birchler 1993; Houben et al. 2001; Peng et al. 2005). However, these analyses used physical mapping of repetitive sequences and were not able to assess the origin of all the sequences on the B chromosome. Interestingly, FISH mapping showed that the B chromosome of *Paralabidochromis chilotes*, another African cichlid, shares sequences with LG3 (Kuroiwa et al. 2014), supporting the idea that this linkage group was important in the origin of these B chromosomes.

Yoshida et al. (2011), using microdissected Bs from *L. rubripinis* as chromosome probes, detected homology of B chromosomes among 11 Lake Victoria cichlids. Our results comparing the *A. latifasciata* B sequences to the *H. chilotes* BACs and B chromosome of *L. rubripinis* (Yoshida et al. 2011) suggest that the B chromosomes of Lake Victoria cichlids, including those from small satellite lakes, have a common origin. We propose that the *A. latifasciata* proto-B formed before the split of the major lineages of Lake Victoria Haplochromini cichlids and have been maintained despite the rapid speciation in this group. The maintenance of the B in diverse species over evolutionary time also suggests that the B has acquired special mechanisms for segregation and maintenance. Because B chromosomes also have been described in the Lake Malawi cichlid *M. lombardoi* (Poletto, Ferreira, Cabral-de-Mello, et al. 2010), it is possible that the East African cichlid Bs have an even more ancient origin.

It is commonly assumed that B chromosomes are filled with repetitive sequences and contain relatively few genes. Physical mapping (Fantinatti et al. 2011) showed that the B chromosome contains many repetitive DNA sequences at

higher density than in the A chromosomes. We identified the main TEs in this B chromosome, and our annotations show that the relative abundance of TEs on the B chromosomes parallels that of the A genome of other cichlids (Shirak et al. 2010). Our results also show that the B chromosomes include sequences homologous to known genes. Although some of these genes might still be functional in some way, most are simply fragments in various states of decay. Only two classes of genes are expected to be intact: Those recently inserted into the B and those that are preserved because of their importance in B chromosome maintenance and transmission. Although many of the genes with high-integrity scores may be recent arrivals on the B chromosome, others may be functional. For example, the B-copy of *Separin*, *TUBB1*, and *KIF11* are transcriptionally active in several tissues of *P. nyererei*. Our study is the first to identify an abundance of gene sequences on the B chromosome and represents a first step in identifying functional B chromosome genes.

TEs are one of the most common migrants into the clonally inherited B chromosome and could be responsible for the insertion of sequences into the B. Gene sequences could have been inserted as fragments, or they could have been broken after insertion on the B by subsequent TE insertions. Another source of sequences on the B chromosome is retroposed copies (retrogenes). In fact, the *hnRNP Q*-like gene we found in the B chromosome of *A. latifasciata* shows characteristics of a retrogene. Although we do not have any evidence of transcription of *hnRNP Q*-like B-copy, there is increasing evidence that retroposition, which generates duplicated gene copies in new genomic positions, is important for the origin of new gene functions (Brosius 1999; Betrán et al. 2002).

Genes Involved in Chromosome Segregation

Our results offer the first insight into the molecular mechanisms by which B chromosomes might alter chromosome segregation to their advantage. To persist, unpaired B chromosomes must have a mechanism to bias their segregation to more than 50% of the daughter cells and gametes. Among the high-integrity genes detected, there are several that play important roles in cell division, including proteins involved with microtubule organization (*TUBB1* and *TUBB5*), kinetochore structure (*SKA1*, *KIF11*, and *CENP-E*), recombination (*XRCC2*, *SYCP2*, and *RTEL1*), and progression through the cell cycle (*Separase*, *AURK*).

The association of microtubules with the chromosome is mediated by a complex that includes SKA1 and which integrates the kinetochore-microtubule structure, regulating microtubule depolymerization and playing a critical role in proper chromosome segregation. All these processes are required to start anaphase. Depletion of this element results in misaligned chromosomes and loss of kinetochore-microtubule attachments (Welburn et al. 2009; Schmidt et al. 2012). Beta tubulins (TUBB) are assembled in association with alpha tubulins genes to construct microtubule filaments, including spindle fibers. The kinetochore-microtubule structure is also stabilized by CENP-E, which is an essential mitotic kinesin (Weaver et al. 2003).

AURK proteins play a vital role in cell cycle control (Macarulla et al. 2008). Defective copies of these genes are associated with severe mitotic abnormalities and cancers. Downregulation of AURK genes can result in mitotic abnormality, leading to problems in centrosome organization, spindle assembly, chromosome alignment, and cytokinesis (Fu et al. 2007). Overexpression of AURK-A has been shown to be associated with aneuploidy, chromosome instability, tumorigenic transformation, and progression in mammalian cells (Bischoff et al. 1998; revised in Gómez-López et al. 2014). AURK-B has been shown to regulate SKA1 (Schmidt et al. 2012). KIF11 is a protein required from prophase until metaphase, which participates in spindle assembly, centrosome separation, postmitotic centrosome movement, and bipolar spindle formation. The inhibition or depletion of KIF11 contributes to the cessation of the cell division process by disrupting centrosomes and spindle poles, by building monopolar microtubules, and by releasing the condensed chromosome surrounding the unseparated centrosomes (Rapley et al. 2008; Ferenz et al. 2010).

Another important set of genes detected on the B chromosome are involved with recombination and repair. Separin (also known as separase) is necessary for the metaphase–anaphase transition and for proper DNA repair (Hauf et al. 2001). XRCC2 is involved with the maintenance of chromosome stability acting in the homologous recombination repair (Nagaraju et al. 2009). SYCP2 is a major component of the lateral elements of synaptonemal complexes (SCs), which regulate the arrangement of homologous chromosomes during meiosis (Winkel et al. 2009). RTEL1 is an essential helicase that is crucial for DNA break repair via homologous recombination and telomere maintenance (Walne et al. 2013)

Furthermore, we detect B-specific transcript variants of the *Separin*, *TUBB1*, and *KIF11* genes in several tissues of *P. nyererei*, suggesting the possibility that these B chromosome copies retain some function. The selective pressure analysis of these genes suggests that they have been largely released from selective constraints and are evolving neutrally. Purifying selection against duplicated gene copies is typically relaxed. Under some conditions, the duplicate copies may acquire new functions. We suggest that the duplicates of these cell cycle genes inserted into the proto-B/B chromosome have become B-variant copies that may alter the spindle structure, spindle dynamics, and/or the SC to promote B chromosome transmission.

Other Functions Encoded by B Genes

There are several examples of B chromosomes involved in sex determination and the evolution of sex chromosomes (reviewed in Burt and Trivers 2006). Manipulation of the gender of the host may improve rates of transmission of the B chromosome. Yoshida et al. (2011) found that B chromosomes biased development toward females in *L. rubripinnis* but not in other species of Lake Victoria cichlid. A significant fraction of the *A. latifasciata* B chromosome is

derived from LGs 1 and 3, which are sex chromosomes in tilapiine cichlids (Cnaani and Kocher 2008; Cnaani et al. 2008; Ser et al. 2009). Kuroiwa et al. (2014) also found a relationship between LG3 and B chromosomes in *Pa. chilotes*. We also found B genes with GO annotations related to "sex determination," a homolog of the Wilms tumor gene (essential for maintaining the integrity of testes). Several genes had GO annotations of "sex differentiation" including three transcript variants of pre-B-cell leukemia transcription factor 1 (involved in steroidogenesis and sexual differentiation in humans), FKBP4 (role in intracellular trafficking of steroid hormone receptors between cytoplasm and nucleus), and FNDC3A (involved in spermatid-Sertoli adhesion during spermatogenesis). However, all these genes have integrity scores less than 15.6%. Furthermore, B chromosomes are detected at similar frequencies in both sexes of *A. latifasciata*, so it does not appear that they play a role in sex determination in this species. On the other hand, we cannot rule out that the idea that these genes were intact and active in sex differentiation early in the evolutionary history of the B, and later lost function and became relics. At this point, we cannot determine whether the effects of B chromosomes on sexual development were a characteristic of the ancestral B chromosome or were recently derived in *L. rubripinnis*.

Another interesting set of genes are the olfactory receptors *5F1* (or *OR11-10*), *6C4* (or *OR12-10*), *6N1* (or *OR6N1*), and *51E1* (or *OR51E1*) (integrity scores of 100%), and the acidic leucine-rich nuclear phosphoprotein 32 family member E (or *Cpd1*) (integrity score ~81.3%), which is differentially expressed during cerebellar development in mouse (Radrizzani et al. 2001). There is a tight correlation between olfactory receptors and sex behavior among cichlids (Maruska and Fernald 2010). These data suggest that genes involved with development, nervous system, and sex behavior might also have a functional role in maintaining the B chromosome.

A Model for the Origin of B Chromosomes

Taking together our results and the findings of previous studies (Poletto, Ferreira, Martins 2010; Fantinatti et al. 2011; Yoshida et al. 2011; Kuroiwa et al. 2014), we constructed a model for the origin of B chromosomes in *A. latifasciata*, which opens new possibilities for the understanding the origin and evolution of B chromosomes in other animal species.

We suggest that the *A. latifasciata* B chromosome arose as a segmental duplication of an autosome. The chromosome fragment must have included a centromere and might have arisen by nondisjunction. A small fragment containing relatively few genes would avoid the negative selection associated with unbalanced gene dosage that dooms most aneuploidies. Subsequently this proto-B has been continuously bombarded by DNA fragments from the A chromosomes. It also underwent isochromosome formation and amplification of some DNA sequences to produce the structure of the current B chromosome (fig. 3). Isochromosomes are chromosomes with two equal arms originating from a biarmed chromosome by the loss of one of its arms and its replacement with an exact copy of the other arm. B chromosomes are frequently associated with isochromosome formation in plants (Jones and Puertas 1993; Jones and Houben 2003) and animals (Mestriner et al. 2000; Bueno et al. 2013), and this phenomenon probably contributes to late stages of B chromosome differentiation.

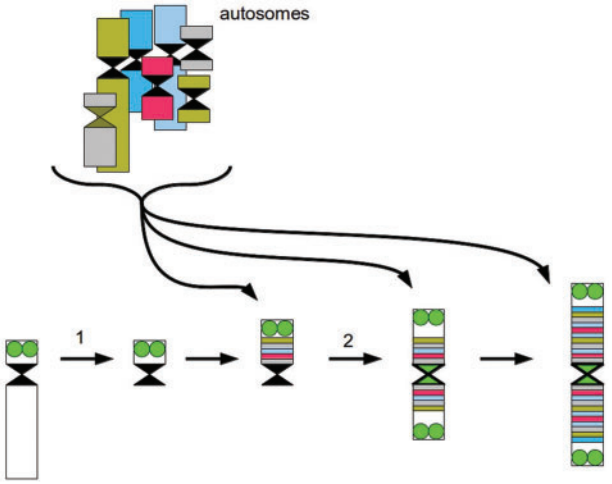

Fig. 3. A model for the B chromosome evolution in *Astatotilapia latifasciata*. 1, segmental duplication from an autosome forms the proto-B; 2, isochromosome formation. The different colors represent DNA segments originated from diverse autosomes and light green at the centromeric and telomeric regions represents rDNA sequences.

Burt and Trivers (2006) and Houben et al. (2013) discuss how a nonessential chromosome fragment containing pericentromeric and centromeric sequences might have the ability to bind to the spindle, giving rise to a proto-B chromosome. However, some kind of drive mechanism is necessary for a proto-B chromosome to rise in frequency. The meiotic machinery promotes competition for spindle binding, and so selection favors the increase of centromeric sequences to promote segregation (Werren 2011). The *A. latifasciata* proto-B has been continuously invaded by DNA fragments from all A chromosomes, mainly by TEs mobilization (which can also insert genes), retrogene insertion, and ectopic recombination. These processes have inserted many sequences during the evolution of B chromosomes. The insertion of genes related to spindle and kinetochore formation might create new proteins that could favor the maintenance of B chromosomes through the cell cycle. Later in the evolutionary process of B evolution, the *A. latifasciata* B chromosome underwent isochromosome formation and amplification of DNA sequences to produce the structure of the current B chromosome, with similar DNA content and sequence distribution in both arms. The formation of B chromosome rings in metaphase I (Poletto, Ferreira, Martins 2010) corroborates this hypothesis.

Conclusions

Our data show that DNA sequences from many different A chromosomes contributed to the B chromosome in *A. latifasciata* and that many individual transposition events

(including retrogenes) were important to the insertion of those sequences. Although most B genes have low levels of integrity, we detected a few high-integrity genes and found evidence of transcription of some B-specific gene sequences. Moreover, we also proposed a model for B chromosome evolution based on the gene content found here and showed that the B chromosome in Lake Victoria cichlids might predate the radiation of that group. The investigation of B chromosomes in species from lakes Malawi and Tanganyika will certainly contribute to understanding the astonishing evolutionary history of cichlids in East Africa.

Furthermore, our data show that karyotyping the species is essential prior to genomic sequencing. The presence of a B chromosome can create problems for genome assembly and introduce bias into the final data generated. The genomic approaches employed here broaden the study of karyotypes and chromosomes in ways not possible using classical or even molecular cytogenetics alone. Next-generation sequencing (NGS) creates new opportunities for understanding the function and evolution of sequences on B chromosomes. In the same way, the application of large-scale analyses of coding and noncoding RNAs of B+ and B− tissue and animals will open new approaches to address questions about B chromosomes.

Materials and Methods

Ethics Statement

The experimental research on animals here employed agree with ethical principles in animal research adopted by the Brazilian College of Animal Experimentation and was approved by the Biosciences Institute/UNESP, Sao Paulo State University ethic committee on use of animals (Protocol no. 486-2013).

DNA Samples and Cytogenetic Analysis

Astatotilapia latifasciata is native to lakes Kyoga and Nawampasa in Uganda, satellite lakes of Lake Victoria. Specimens were obtained from a stock established from the trade and maintained in the fish facility at Sao Paulo State University (Botucatu, Brazil). They were karyotyped by classical cytogenetic procedures using a Giemsa stain. We identified males without any B chromosome (B−), or with two B chromosomes (B+), for further analysis. C-banded chromosome preparation was performed according to Sumner (1972) and Chromomycin A_3 (CMA_3) staining according to Schweizer (1976). FISH was conducted with the following probes (selected from Fantinatti et al. 2011): Rex1 and Rex3 TEs, 45 S rDNA, and BAC b03TI066P02. The data were analyzed according to the centromeric index (ci) that represents the ratio of the length of the short arm (p) of the chromosome to that of the total chromosome ($p + q$, where q is the length of the long arm). The chromosomes were classified as metacentric (ci of 50.00-40.10), submetacentric (ci of 40.00-25.10), telocentric (ci of 25.00-0.01), or acrocentric (ci of 0.00) according to Guerra (1986).

B Chromosome Microdissection and Painting

B chromosome microdissection was performed on an inverted microscope (Olympus IX51) equipped with a mechanical micromanipulator (Narishige, Tokyo, Japan). The glass capillaries used for microdissection were produced by a micropipette puller (Narishige) and had a 0.7 µm diameter. One B chromosome was microdissected, transferred to microtubes and posteriorly amplified by the kit Genome Plex Single Cell, Whole Genome Amplification, WGA4 (Sigma-Aldrich, St. Louis, MO), according to the instructions of the manufacturer. The amplification product of the B chromosome was purified using the kit GenEluteTM PCR Clean-Up (Sigma-Aldrich). The purified products were used for 454 NGS or reamplified to be used as chromosome probes using the kit Genome Plex Whole Genome Amplification, WGA3 (Sigma-Aldrich). In the amplification reaction with WGA3, the nucleotide digoxigenin 11-dUTP was incorporated for the B probe with the ratio (7 dTTP: 3 dUTP 11-digoxigenin). The whole chromosome painting hybridization process followed the protocol described by Pinkel et al. (1986) under high stringency conditions—77% (200–400 ng probe/slide, 50% formamide, 10% dextran sulfate, 2xSSC, pH 7.0–7.2, at 37 °C overnight). After hybridization, the metaphases of *A. latifasciata* were analyzed in an epifluorescence Olympus BX51 microscope (Olympus, Tokyo, Japan), and the images were captured using an Olympus DP72 system.

Genome Sequencing

For the Illumina libraries, genomic DNA (gDNA) from each individual was sheared to an average size of 500 bp using an S220 focused ultrasonicator (Covaris Inc., Woburn, MA). Separate libraries were constructed for each individual using the TruSeq DNA sample preparation kit ver.2 rev.C (Illumina Inc., San Diego, CA). Paired end (2 × 100 bp) sequencing of each library was performed in separate lanes on an Illumina HiSeq 1000 sequencer. The Illumina reads were aligned to the *M. zebra* reference genome using paired-end mode of Bowtie2 with the –very sensitive option.

DNA generated from whole-DNA amplification of a single microdissected B chromosome was sequenced using the 454 GS Jr platform (Roche, Basel, Switzerland) and assembled using the Newbler software package. The assembled contigs were blasted against the NCBI nucleotide collection. The 454 reads were also aligned to the *M. zebra* reference using BWA (with the BWA-SW algorithm). Coverage, read tracks, and alignment for Illumina and 454 data are available at the Bouillabase database (www.bouillabase.org, last accessed May 7, 2014).

Identification of B Chromosome "Blocks" via Coverage Ratio Analysis

Sites with at least 15X Illumina coverage in both B+ and B− samples were selected to calculate the mean B+/B− coverage ratio for each position on scaffold_0, which does not appear to contain any B chromosome sequence. The mean ratio was 0.74 (22.92/30.99) with a standard deviation (stdv) of 0.20. This presumably estimates the expected ratio for

sequences that have the same copy number in both genomes (i.e., the autosomal copies in both genomes). Thus, sequences that have one copy on an autosome and at least one copy on the B are expected to have a coverage ratio about 1.48.

We extracted all sites with a coverage ratio $\geq$1.48 and $\geq$1.88 (1.48 + 2stdv), separately, for further analysis. A custom script was used to identify blocks of contiguous sequence from those selected sites. We used two levels of "tolerance" to build sequence blocks, either 100 bp or 1 kb. For instance, the tolerance of 100 bp means that if two selected sites are $\leq$100 bp apart, they are considered part of the same sequence block. Thus, four set of blocks were obtained to further analysis: 0 stdv and plus 2 stdv, both with tolerance of 100 bp and 1 kb.

Several blocks of the B chromosome, and also the 454 reads, were examined in a genome browser of the Bouillabase.org database to confirm the extent of the block, the coverage ratio, and to detect any possible artifacts arising from the reference genome assembly.

Annotation Procedures, Gene Integrity Analysis, and GO Search

The blocks were compared with the reference genome annotation to estimate the presence of TEs and genes on the B chromosomes. The reference consisted of the *M. zebra* gene set (v0 – Broad Institute) genome assembly. For the TEs search, only the block set + 2 stdv/$\leq$100 bp was examined because it represents the most stringent condition used to build the block sets. If a gene is present in more than one list, we select it to make a combined list of putative B chromosome genes (BCL). The position of each BCL gene was retrieved from the four sets of blocks.

Because the read mapping approach does not allow us to retrieve consensus sequences to predict the encoded protein sequences, we choose to calculate the B genes length and integrity using the DNA sequences. Usually, the length of each B-gene identified is slightly different in each of the four analyses, so we overlapped the DNA pieces related of the same gene (each "piece" is a different gene length in each list of blocks) and recovered the length of each B-gene as the sum of all overlapped pieces. The integrity percentage of each B-gene was calculated comparing its length to the corresponding gene length in the annotation of the *M. zebra* genome.

The BLAST2GO pipeline (Conesa et al. 2005) was applied over both the *M. zebra* gene set (40,008) (Bouillabase.org) and the BCL. Fisher's exact test (using corrected *P* values to control the false discovery rate) was used to compare the set of GO terms from both BCL and *M. zebra* genes for a GO enrichment analysis. GO terms related to the keywords "sex determination" and "sex differentiation" were retrieved from Gene Ontology database (www.geneontology.org/, last accessed May 7, 2014) and were searched against the GO of BCL.

GO-enriched terms and the complete GO annotation from the genes with an integrity $\geq$50% and $\geq$70%, respectively, were analyzed independently. In the last case, Uniprot and NCBI function descriptions, as well as reports available in the literature, were also used to understand the function of each gene.

Comparative and Evolutionary Analysis

The B chromosome BAC contigs of the African cichlid from Lake Victoria *H. chilotes* (Yoshida et al. 2011) were aligned against the *M. zebra* reference using the BWA-SW algorithm. In addition, all B+ and B− Illumina reads were aligned over these "B-BAC" contigs using Bowtie2.

In attempt to identify the A chromosomes that the B chromosome were derived from, we mapped the B blocks to the *M. zebra* anchored genome assembly (O'Quin et al. 2013) and used Circos (Krzywinski et al. 2009) to generate a graphic view. The + 2 stdv/100 bp tolerance set of predicted B blocks was used for this analysis because it represents a conservative set of predicted blocks.

Comparisons of the read coverage from the *P. nyererei* genome and transcriptomes (Broad Institute, http://www.ncbi.nlm.nih.gov/bioproject/60367, last accessed May 7, 2014) and B chromosome blocks were also performed in the same manner as described below. The reads of high-integrity genes from *P. nyererei*, *M. zebra*, and *A. latifasciata* (0B and 2B genomes) were retrieved and consensus sequences for each gene of all three species were built using SAMtools (Li et al. 2009). The 2B gene sequences were recovered comparing the ambiguous base positions with 0B genome (e.g., supplementary fig. S6, Supplementary Material online). The *M. zebra* sequences were used as queries to look for those genes in other cichlid (*A. burtoni*, *Neolamprologus brichardi*, and *Oreochromis niloticus*) and noncichlid (*Oryzias latipes*, *Xiphophorus maculatus*, *Lepisosteus oculatus*, and *Takifugu rubripes*) genomes by BLAST searches at Bouillabase and NCBI. The mRNA and genomes sequences were aligned using MAUVE (Darling et al. 2004), and the alignments were manually edited. The parameters (-t ML -s 11 -i -g 4 -S BEST -v -p -AICc -f -c 100) of JmodelTest (Posada 2008) were applied over all alignments to estimate the best evolutionary models. Bayesian inference using random local clock model were conducted using Beast v1.7.0 (Drummond and Rambaut 2007) using the CIPRES Science Gateway (Miller et al. 2010). The maximum likelihood (using aLRT reliability test) analyses were performed using Phyml v3.0.1 (Guindon and Gascuel 2003). Neighbor-joining (1,000 replicates for bootstrap) were applied using Seaview software (Gouy et al. 2010). The burn-in adopted in Bayesian inference was determined in Tracer (Rambaut 2009), and the summarization and exclusion of trees out of convergence area were performed using TreeAnnotator (Drummond and Rambaut 2007).

The consensus sequences of each *Separin*, *TUBB1*, and *KIF11* genes (genes with mutational points in agreement with *P. nyererei* transcriptomes; supplementary fig. S8, Supplementary Material online) of 0B and 2B genomes were aligned against genomic and mRNA sequences of *M. zebra* (mRNA sequences XM_004570623.1, XM_004545355.1, and XM_004547671.1) using MAUVE aligner (Darling et al. 2004). Based on mRNAs, the introns were manually excluded and the amino acid (aa) residues were predicted

for each sequence. The proteins were aligned against *M. zebra* aa residues (XP_004570680.1, XP_004545412.1, and XP_004547728.1) using Muscle (Edgar 2004) to certify the predictions and if all introns, and 3′- and 5′-UTRs were excluded. The edited nucleotide sequences were further aligned based on their protein alignment using Seaview software (Gouy et al. 2010). The dN/dS (nonsynonymous/synonymous substitutions) ratio calculation to estimate the selective pressure was applied using PAL2NAL software (Goldman and Yang 1994) comparing 0B versus 2B, 0B versus *M. zebra*, and 2B versus *M. zebra* sequences.

Validation of NGS and Integration of Data via qPCR and FISH

Genomic block characteristics of the B genome were selected for the construction of qPCR primers (supplementary table S2, Supplementary Material online) to screen *A. latifasciata* genomes for the presence of B chromosome and to quantify their abundance in three animals of each 0B, 1B, and 2B genomes (each previously checked for B cytogenetically). The selected regions comprise B blocks with the highest, medium, and lowest coverage ratios. qPCR of gDNA was used to calculate the gene dose by a ΔCt method of relative quantification (Nguyen et al. 2013). Gene dosage ratios (GDR) of the target genes were compared with a single-copy autosomal gene, hypoxanthine phosphoribosyltransferase (*Hprt*). If the target gene is autosomal, its copy number ratio compared with Hprt is expected to be 1:1. RT-qPCR was carried out in the StepOne Real-Time PCR Systems (Life Technologies, Carlsbad, CA). The target and reference genes were analyzed simultaneously in triplicates of three independent samples. The cycling conditions were 95 °C for 10 min; 45 cycles of 95 °C for 15 s, and 60 °C for 1 min. Specificity of the PCR products was confirmed by analysis of the dissociation curve. A correlation analysis between GDR and mean coverage ratios (2B/0B) was performed for each B region of scaffolds 3, 13, 19, 26, and 31 selected for qPCR.

The same DNA segments of the 2B blocks used for qPCR were used as probes for FISH to determine whether they were located in the B chromosomes. FISH was performed using the protocol described by Pinkel et al. (1986) with modifications (Cabral-de-Mello et al. 2012). The slides were denatured in 70% formamide/2xSSC, pH 7, for 36 s, and dehydrated in an ice-cold ethanol series (70%, 85%, and 100%). The probes consisted of markers on scaffolds 13, 19, 26, 31, and 324 (selected from this work).

To search for correspondence of the B+ illumina blocks and the sequences generated from the microdissected sequenced B chromosome, reads generated from the whole sequenced microdissected B chromosome were aligned to the *M. zebra* reference genome (Bouillabase database, www.bouillabase.org, last accessed May 7, 2014) and compared with the coverages of the *A. latifasciata* B− and B+ genomes.

Supplementary Material

Supplementary figures S1–S8, tables S1 and S2, and data sets S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by grants from São Paulo Research Foundation (2011/03807-7) and the National Science Foundation (DEB-1143920).

References

Alfenito MR, Birchler JA. 1993. Molecular characterization of a maize B chromosome centric sequence. *Genetics* 135:589–597.

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A. 2013. Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell* 25:2536–2544.

Banaei-Moghaddam AM, Schubert V, Kumke K, Weiβ O, Klemme S, Nagaki K, Macas J, González-Sánchez M, Heredia V, Gómez-Revilla D, et al. 2012. Nondisjunction in favor of a chromosome: the mechanism of rye B chromosome drive during pollen mitosis. *Plant Cell* 24:4124–4134.

Betrán E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* 12:1854–1859.

Bischoff JR, Anderson L, Zhu Y, Mossie K, Ng L, Souza B, Schryver B, Flanagan P, Clairvoyant F, Ginthe C, et al. 1998. A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* 17:3052–3065.

Brosius J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238: 115–134.

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. Cambridge: Belknap Press of Harvard University.

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, editor. The evolution of the genome. San Diego (CA): Elsevier. p. 223–286.

Camacho JP, Sharbel TF, Beukeboom LW. 2000. B-chromosome evolution. *Philos Trans R Soc Lond B.* 355:163–178.

Cnaani A, Kocher TD. 2008. Sex-linked markers and microsatellite locus duplication in the cichlid species *Oreochromis tanganicae*. *Biol Lett.* 4: 700–703.

Cnaani A, Lee BY, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller JF, D'Cotta H, Penman DJ, et al. 2008. Genetics of sex determination in tilapiine species. *Sex Dev.* 1:43–54.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.". *Bioinformatics* 21: 3674–3676.

Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–403.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

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.

Ferenz NP, Gable A, Wadsworth P. 2010. Mitotic functions of kinesin-5. *Semin Cell Dev Biol.* 21:255–259.
Fu J, Bian M, Jiang Q, Zhang C. 2007. Roles of aurora kinases in mitosis and tumorigenesis. *Mol Cancer Res.* 5:1–10.
Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11: 725–736.
Gómez-López S, Lerner RG, Petritsch C. 2014. Asymmetric cell division of stem and progenitor cells during homeostasis and cancer. *Cell Mol Life Sci.* 71:575–597.
Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol.* 27:221–224.
Guerra MS. 1986. Reviewing the chromosome nomenclature of Leval et al. *Braz J Genet.* 9:741–743.
Guindon S, Gascuel O. 2003. PhyML: a simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.
Hauf S, Waizenegger IC, Peters JM. 2001. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science* 293: 1320—1323.
Houben A, Banaei-Moghaddam AM, Klemme S. 2013. Biology and evolution of B chromosomes. In: Leitch IJ, editor. Plant genome diversity, Vol. 2. Berlin (Germany): Springer. p. 149–165.
Houben A, Verlin D, Leach CR, Timmis JN. 2001. The genomic complexity of micro B chromosomes of *Brachycome dichromosomatica*. *Chromosoma* 110:451–459.
Jones RN. 1991. B-chromosome drive. *Am Nat.* 137:430–442.
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.
Jones RN, Houben A. 2003. B chromosomes in plants: escapes from the A chromosome genome? *Trends Plant Sci.* 8:417–423.
Jones RN, Puertas MJ. 1993. The B chromosomes of rye (*Secale cereale* L.). In: Dhir KK, Sareen TS, editors. Frontiers in plant science research. Delhi (India): Bhagwati Enterprises. p. 81–112.
Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19:1639–1645.
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.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
Macarulla T, Ramos FJ, Tabernero J. 2008. Aurora kinase family: a new target for anticancer drug. *Recent Pat Anticancer Drug Discov.* 3: 114–122.
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.
Maruska KP, Fernald RD. 2010. Reproductive status regulates expression of sex steroid and GnRH receptors in the olfactory bulb. *Behav Brain Res.* 213:208–217.
Mestriner CA, Galetti PM Jr, Valentini SR, Ruiz IR, Abel LD, Moreira-Filho O, Camacho JP. 2000. Structural and functional evidence that a B chromosome in the characid fish *Astyanax scabripinnis* is an isochromosome. *Heredity* 85:1–9.
Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 2010 Nov 14; New Orleans, LA. p 1–8.
Nagaraju G, Hartlerode A, Kwok A, Chandramouly G, Scully R. 2009. XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversions between sister chromatids. *Mol Cell Biol.* 29: 4283–4294.
Nguyen P, Sýkorová M, Šíchová J, Kůta V, Dalíková M, Čapková Frydrychová R, Neven LG, Sahara K, Marec F. 2013. Neo-sex chromosomes and adaptive potential in tortricid pests. *Proc Natl Acad Sci U S A.* 110:6931–6936.
O'Quin CT, Drilea AC, Conte MA, Kocher TD. 2013. Mapping of pigmentation QTL on an anchored genome assembly of the cichlid fish, *Metriaclima zebra*. *BMC Genomics* 14:287.
Palestis BG, Trivers R, Burt A, Jones RN. 2004. The distribution of B chromosomes across species. *Cytogenet Genome Res.* 106: 151–158.
Peng SF, Lin YP, Lin BY. 2005. Characterization of AFLP sequences from regions of maize B chromosome defined by 12 B-10L translocations. *Genetics* 169:375–388.
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, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C. 2010. Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet.* 11:50.
Poletto AB, Ferreira IA, Martins C. 2010. The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. *BMC Genet.* 11:1.
Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.
Radrizzani M, Vila-Ortiz G, Cafferata EG, Di Tella MC, Gonzalez-Guerrico A, Perandones C, Pivetta OH, Carminatti H, Idoyaga VP, Santa-Coloma TA. 2001. Differential expression of CPD1 during postnatal development in the mouse cerebellum. *Brain Res.* 907: 162–174.
Rambaut A 2009. Molecular evolution, phylogenetics and epidemiology: tracer [Online]. [updated 2009 Nov 30]. Available from: http://tree.bio.ed.ac.uk/software/tracer/.
Rapley J, Nicolàs M, Groen A, Regué L, Bertran MT, Caelles C, Avruch J, Roig J. 2008. The NIMA-family kinase Nek6 phosphorylates the kinesin Eg5 at a novel site necessary for mitotic spindle formation. *J Cell Sci.* 121:3912–3921.
Schmidt JC, Arthanari H, Boeszoermenyi A, Dashkevich NM, Wilson-Kubalek EM, Monnier N, Markus M, Oberer M, Milligan RA, Bathe M, et al. 2012. The kinetochore-bound Ska1 complex tracks depolymerizing microtubules and binds to curved protofilaments. *Dev Cell.* 23:968–980.
Schweizer D. 1976. Reverse fluorescent chromosome banding with chromomycin and DAPI. *Chromosoma* 58:307–324.
Ser JR, Roberts RB, Kocher TD. 2009. Multiple interacting loci control sex determination in Lake Malawi cichlid fish. *Evolution* 64: 486–501.
Shirak A, Grabherr M, Di Palma F, Lindblad-Toh K, Hulata G, Ron M, Kocher TD, Seroussi E. 2010. Identification of repetitive elements in the genome of *Oreochromis niloticus*: tilapia repeat masker. *Mar Biotechnol.* 12:121–125.
Sumner A. 1972. A simple technique for demonstrating centromeric heterochromatin. *Exp Cell Res.* 75:304–306.
Trifonov VA, Dementyeva 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.
Walne AJ, Vulliamy T, Kirwan M, Plagnol V, Dokal I. 2013. Constitutional mutations in RTEL1 cause severe dyskeratosis congenita. *Am J Hum Genet.* 92:448–453.
Weaver BA, Bonday ZQ, Putkey FR, Kops GJ, Silk AD, Cleveland DW. 2003. Centromere-associated protein-E is essential for the

mammalian mitotic checkpoint to prevent aneuploidy due to single chromosome loss. *J Cell Biol.* 162:551–563.

Welburn JPI, Grishchuk EL, Backer CB, Wilson-Kubalek EM, Yates JR 3rd, Cheeseman IM. 2009. The human kinetochore Ska1 complex facilitates microtubule depolymerization-coupled motility. *Dev Cell.* 16: 374–385.

Werren JH. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci U S A.* 108:10863–10870.

White MJD. 1973. Animal cytology and evolution. London: Cambridge University Press.

Wilson B. 1907. The supernumerary chromosomes of Hemiptera. *Science* 26:870–71.

Winkel K, Alsheimer M, Ollinger R, Benavente R. 2009. Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma* 118:259–267.

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.