



**ANÁLISE FUNCIONAL DA PRESENÇA DE CROMOSSOMO B, UTILIZANDO O
PEIXE CICLÍDEO *Astatotilapia latifasciata* COMO MODELO**

Diego Ferreira Marques

Botucatu, Julho de 2016

INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA)

**ANÁLISE FUNCIONAL DA PRESENÇA DE CROMOSSOMO B, UTILIZANDO O
PEIXE CICLÍDEO *Astatotilapia latifasciata* COMO MODELO**

Diego Ferreira Marques

Orientador: Cesar Martins

Tese apresentada ao Instituto de Biociências, Câmpus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Genética).

Botucatu, Julho de 2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÊC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Marques, Diego Ferreira.

Análise funcional da presença de cromossomo B, utilizando o peixe ciclídeo *Astatotilapia latifasciata* como modelo / Diego Ferreira Marques. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Cesar Martins

Capes: 20204000

1. Peixe - Genética. 2. Ciclídeos. 3. Expressão gênica.
4. Transcriptoma. 5. Genômica. 6. Cromossomos.

Palavras-chave: Elemento extra numerário; expressão gênica; genômica funcional; transcriptoma.

**O fracasso jamais te surpreenderá se
sua decisão de vencer for
suficientemente forte.**

Og Mandino

AGRADECIMENTOS

À DEUS.

Palavras não serão suficientes pra demonstrar o porquê que devo agradecer a minha **FAMÍLIA**. Obrigado pelo exemplo de honestidade, sinceridade, caráter e afetividade.

À meus **PAIS** (Ana e Bosco) por terem me ajudado e me apoiado em todas as dificuldades que fui levado a lidar. Obrigado por ser flexíveis e rido quando necessário, e muito obrigado por ainda estarem presentes de forma muito sólida em minha vida apesar da distância.

À meus **IRMÃOS** (Dannylo, Dannyel, Dulcyana) pelo apoio, e pela companhia diária, amo vocês. Abraço também ao cunhado Moisés, cunhadas Jana e Nyra e ainda mais aos meus dois sobrinhos Arthur e João Vitor que me alegram grandemente com seus recadinhos e fotos diárias.

Ao meu **ORIENTADOR** que apostou em mim desde o começo de dando suporte financeiro desde o início quando isso tudo era apenas um sonho.

Aos meus **AMIGOS DO LABORATÓRIO**. Durante esse período convivi mais tempo com vocês do que com minha própria família, muitos de nós seguiremos caminhos opostos mas vou sempre levar os momentos que passamos juntos, os ensinamentos.

Aos meus **AMIGOS QUE A VIDA UNIU** em especial Leonardo Mendes, Carina Leonelli e Lívia Silveira (UNESP), Adriane, Luiza, Rodrigo, Vânia e minhas parceiras da UNZER Carol e Tati!! além de muitos outros fora da universidade que deixaram os momentos mais leves e divertidos. Um muito obrigado a minhas amigas e parceiras Vanda e Dona Terezinha!!!!

Gostaria de agradecer também aos órgãos que forneceram apoio financeiro para realização desse trabalho: FAPESP (processo 2012/16288-0), CAPES e CNPq.

MUITO OBRIGADO!!!!

SUMÁRIO

1. Introdução.....	9
1.1 Cromossomos supernumerários.....	9
1.2 Peixes ciclídeos como modelo para estudos de biologia evolutiva.....	11
1.3 O ciclídeo <i>Astatotilapia latifasciata</i> como modelo para estudos de cromossomos B	12
1.4 Transcriptomas na elucidação de aspectos funcionais celulares	14
2. Objetivos.....	16
2.1 Objetivo Geral e Justificativa	16
2.2 Objetivos específicos	17
3. Hipótese.....	17
4. Materiais e Métodos	18
4.1 Obtenção das amostras de RNA	18
4.2 Genotipagem de animais B+ e B-.....	18
4.3 Extração do RNA total	19
4.4 Sequenciamento e processamento dos <i>reads</i>	19
4.5 Montagem <i>de novo</i>	20
4.6 Diferencial de expressão, identificação de transcritos redundantes entre as amostras e anotação funcional.....	21
5. Resultados e discussão	22
6. Conclusões.....	37
7. Referências Bibliográficas.....	38

Resumo

Os cromossomos supernumerários (cromossomos B) são elementos extras, não homólogos ao complemento A e já foram descritos em uma gama de espécies representando os principais grupos de eucariotos. Estudos com esses elementos têm sido mais direcionados as investigações em nível estrutural, sendo que aspectos funcionais têm sido pouco explorados. Dentre as espécies de peixes, os ciclídeos tem recebido grande interesse científico uma vez que sofreram um rápido e extenso processo de radiação adaptativa e até o momento, cromossomos B já foram descritos em 21 espécies de ciclídeos, variando em quantidade de um a três. Considerando os aspectos citados, atrelados à possibilidade do uso de ferramentas que integram dados funcionais, genômicos e citogenéticos, o presente projeto visou analisar o perfil transcricional de tecidos-alvo de *A. latifasciata* afim de investigar possíveis interferências motivadas pelo cromossomo B relacionadas ao nível de transcrição em indivíduos com presença ou ausência desse elemento extra. Uma gama de genes diferencialmente expressos evidenciada nos tecidos analisados, mostra que B cromossomo pode interferir em diferentes vias fisiológicas da célula. Os resultados mostraram grande contribuição para a discussão sobre a importância funcional do cromossomo B, identificando vários genes relacionados ao ciclo celular, manutenção da estrutura celular e resposta imune. Foram identificados genes altamente expressos localizados no B, o que mostra sua interferência direta. Entretanto a maioria dos genes diferencialmente expressos estão localizados nos cromossomos do complemento normal, cuja função está sendo afetada pela presença deste elemento. Outro ponto importante foi a observação de uma possível ação de repressão do cromossomo B em testículo e aumento dos níveis de expressão em ovário, sugerindo uma possível ação no mecanismo de "drive" desse cromossomo nos gametas femininos.

Palavras-chave: Elemento extra numerário, expressão gênica, transcriptoma, genômica funcional.

Abstract

Supernumerary chromosomes (B chromosomes) are extra elements, not homologous to the A complement and have been described in a range of species in eukaryotes. Studies with these elements have been focused in structural level, and functional aspects have been poorly explored. Among fish species, cichlids have gaining scientific interest because the rapid and extensive adaptive radiation process occurred in the group and B chromosomes have been described in 21 cichlid species, ranging in number from one to three. Considering these aspects, linked to the possibility of using tools that integrate functional and genomic data and cytogenetics, we aimed to analyze transcriptional profiles of target tissues of *A. latifasciata* in order to investigate possible interference motivated by B chromosome related to the transcript expression levels in individuals with the presence or absence of this extra element. A range of genes differentially expressed evidenced in the analyzed tissues shows that B chromosome may be interfering in different pathways. The results contributed to the discussion of functional aspects of B chromosome, identifying several genes related to cell cycle, maintenance of cell structure and immune response. Highly expressed genes located on B were identified. However, the majority of highly expressed genes is located on chromosomes of the normal complement and seems to be affected by the presence of B. Furthermore, our data suggest a possible repressive action of B chromosome in tests and increasing expression genes in ovarian confirming the action of drive mechanism of this chromosome in female gametic cells.

Keywords: Extra Element, gene expression, transcriptome, functional genomics.

1. Introdução

1.1 Cromossomos supernumerários

Os cromossomos B, também conhecidos como elementos supernumerários, não são homólogos aos cromossomos do complemento normal (complemento A) (Jones e Rees 1982). São descritos como “elementos egoístas” (Wilson, 1907) e podem ser encontrados em alguns indivíduos de uma mesma população e não em outros, o que os caracterizam como elementos não essenciais e dispensáveis (revisado em Burt and Trivers, 2006). Esses cromossomos são descritos como elementos parasitas e o genoma de indivíduos portadores do B é classificado como “hospedeiro” (Camacho *et al.* 2000). Seu surgimento pode estar diretamente relacionado a rearranjos cromossômicos do tipo fusão e fissão ou à partir da geração de indivíduos híbridos através do cruzamento entre espécies próximas (Camacho, 2000).

Quanto a composição genômica, em geral são heterocromáticos e compostos basicamente de elementos repetitivos, esse enriquecimento de sequências mostra-se frequente; pode causar a inativação de genes, modificar a ordem e a estrutura gênica e assim, ocasionar a gradual degeneração deste cromossomo por exemplo (Camacho *et al.* 2000; Valente *et al.* 2014).

Análises estruturais descrevem a presença de regiões eucromáticas com descrição de genes nucleares e ribossomais (Rajicic *et al.* 2015), como famílias multigênicas de RNAr 45S (Loreto *et al.* 2008; Fantinatti *et al.* 2011), sequências de microssatélites (Piscor and Parise-Maltempo 2016) e histonas (López-León *et al.* 2008; Teruel *et al.* 2010; Bueno *et al.* 2013; Silva *et al.* 2014; Utsunomia *et al.* 2016), entre outros, que podem interferir no genoma de seus hospedeiros (Graphodatsky *et al.* 2005; Coleman *et al.* 2009; Acosta and Moscone 2011; Valente *et al.* 2014).

Os cromossomos B tendem a se manter na população e dentre os mecanismos que facilitam essa permanência ressalta-se o “drive”. Nesse processo o B dirige-se durante a meiose na fêmea movendo em direção ao ovócito, escapando assim de uma possível eliminação; em outros casos o B mantêm-se destruindo cromossomos do complemento A, como no caso de *Nasonia vitripennis*, por processos ainda não claramente elucidados (revisado em Burt and Trivers, 2006; Pereira *et al.* 2009; Akbari *et al.* 2013). Em sua maioria, os cromossomos B se beneficiam da não disjunção como parte de seu mecanismo de auto-acumulação (revisado em Burt and Trivers, 2006; Lin *et al.* 2014).

No grupo de peixes, estudos citogenéticos os têm revelado em distintas populações de diversas espécies; estes variam em tamanho, frequência e composição. Os estudos estão, em sua maioria, restritos a análises convencionais envolvendo descrição da presença e frequência deste polimorfismo além da sua natureza eucromática/heterocromática (Mestriner *et al.* 2000; Ziegler *et al.* 2003; Poletto *et al.* 2010a; Voltolin *et al.* 2010; Fantinatti *et al.* 2011, Silva *et al.* 2014; Valente *et al.* 2014; Rajicic *et al.* 2015).

Apesar da idéia de ser um elemento inerte, atualmente há muitos relatos da influência do cromossomo B na modulação gênica do hospedeiro (Martis *et al.* 2012; Huang *et al.* 2016). Estes cromossomos podem estar envolvidos em recombinação ectópica entre diferentes cromossomos e estarem envolvidos no movimento de DNAs de organelas para os cromossomos B (Martis *et al.* 2012; Valente *et al.* 2014). Estudos recentes relatam que esses cromossomos podem ter originado cromossomos sexuais e também possuem um papel vital na determinação sexual em algumas espécies de peixes e insetos (Yoshida *et al.* 2011; Martis *et al.* 2012; Houben *et al.* 2013; Akbari *et al.* 2013; Valente *et al.* 2014).

Não obstante o maior conhecimento do cromossomo B ser relacionado apenas a sua questão estrutural, alguns estudos mostram interferência transcricional direta desse elemento. No estudo em *Drosophila albomicans* foi evidenciado transcritos do cromossomo B (Zhou *et al.* 2012), além dos trabalhos com o roedor (*Capreolus pygargus*) (Trifonov *et al.* 2013), com centeio (*Secale cereale*) (Carchilan *et al.* 2007; Carchilan *et al.* 2009; Banai-Moghaddan *et al.* 2013; 2015), em milho (*Zea mays L.*) (Jin *et al.* 2016) e no gafanhoto (*Eyprepocnemis plorans*) (Teruel *et al.* 2010; Ruiz-Estévez *et al.* 2012) ambos evidenciando transcritos específicos provenientes do cromossomo B, atestando sua interferência direta no genoma “hospedeiro”.

Apesar das evidências da funcionalidade do cromossomo B em espécies de animais, plantas e insetos tanto em nível direto quanto na modulação de genes do complemento A (Carchilan *et al.* 2007; Carchilan *et al.* 2009; Yoshida *et al.* 2011; Zhou *et al.* 2012; Akbari *et al.* 2013; Trifonov *et al.* 2013; Jin *et al.* 2016) o que se observa é ainda um nível de investigação pontual no qual as análises são restritas a poucas regiões e/ou determinados genes, fornecendo assim informações interessantes, entretanto, de certa forma limitadas.

Nesse panorama, o grupo de peixes que vem ganhando mais destaque no estudo e análise a cerca do cromossomo B, refere-se a família cichlidae. Na qual, dentre as características mais marcantes desse grupo citam-se sua rápida evolução adaptativa além da descrição de inúmeras espécies portadoras desse elemento extra numerário em seu genoma (revisado por Kocher, 2004; Yoshida *et al.* 2011; Brawand *et al.* 2014). Outro fator importante

que auxilia a análise comparativa refere-se a disponibilidade de genomas de referência de vários indivíduos dessa família permitindo análises em larga escala.

1.2 Peixes ciclídeos como modelo para estudos de biologia evolutiva

Os peixes caracterizam-se por sua diversidade e são comumente utilizados em estudos de diversidade e evolução entre os vertebrados (Nelson, 2006). Neste cenário, as espécies da família Cichlidae representam excelentes modelos experimentais para estudos em diversas áreas da genética (Chang *et al.* 2005; The International Cichlid Genome Consortium, 2006; Ijiri *et al.* 2007; Wang *et al.* 2007a; Cnaani *et al.* 2008; Ferreira e Martins, 2008; Valente *et al.* 2009, Poletto *et al.* 2010a,b). Este grupo de peixes mostrou uma rápida radiação adaptativa nos grandes lagos do leste da África, onde quase 2.000 espécies diversificaram-se nos últimos 10 milhões de anos, sendo potenciais modelos para estudos em genética, evolução e especiação (revisado por Kocher, 2004).

Os ciclídeos pertencem à ordem Perciformes e estão incluídos entre as famílias de peixes com maior número de espécies (Nelson, 2006), com cerca de 3.000 representantes distribuídos pela América Central e do Sul, Madagascar, Sudeste da Índia, África e Arábia (Skelton, 2001; revisado por Kocher, 2004). A maior diversidade de espécies ocorre principalmente nos grandes lagos africanos (Trewavas, 1983; Skelton, 2001). Os representantes desta família são subdivididos em quatro subfamílias: Etoplinae (ciclídeos de Madagascar e sul da Ásia – Índia e Sri Lanka), Ptychochrominae (endêmicos da Ilha de Madagascar), Cichlinae (espécies da região neotropical) e Pseudocrenilabrinae (ciclídeos africanos) (Sparks and Smith, 2004; Smith *et al.* 2008).

Estudos cromossômicos e genômicos neste grupo tem se mostrado promissores em análises de diversificação de espécies, diferenciação sexual e mecanismos evolutivos (Yoshida *et al.* 2011; Schneider *et al.* 2013). Atualmente, o sequenciamento completo dos genomas de diversas espécies de ciclídeos abriu novas fronteiras ao conhecimento estrutural, funcional e evolutivo neste grupo. Dentre os genomas sequenciados até o momento, destacam-se as espécies *Oreochromis niloticus*, *Astatotilapia burtoni*, *Neolamprologus brichardi*, *Pundamilia nyerereie* e *Metriaclicma zebra* (disponíveis em <http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html>).

A investigação e análise genômica desses cinco novos genomas de ciclídeos mostram uma acelerada evolução gênica, com acúmulo de duplicações gênicas e interferência de expressão mediada por elementos transponíveis e sequências não codificantes (microRNAs).

O estudo revelou também evidências de sua rápida diversificação, corroborando a teoria de radiação adaptativa (Brawand *et al.* 2014).

Dentro da família Cichlidae, a presença de cromossomos B foi descrita em 7 espécies da região neotropical e 14 espécies africanas, variando em quantidade de um a três cromossomos extras, com diferenças cromossômicas morfológicas entre as espécies. Esses estudos forneceram evidências de que os cromossomos B podem ter um efeito funcional na determinação do sexo feminino como foi observado na espécie *Metriaclima zebra*, além de evidenciar uma gama de genes presentes nesse cromossomo com potencial de serem transcritos (Poletto *et al.* 2010a; Poletto *et al.* 2010b; Yoshida *et al.* 2011; Kuroiwa *et al.* 2013; Valente *et al.* 2014).

Os avanços no sequenciamento em larga escala e análises comparativas utilizando genomas de representantes deste grupo permitiram um melhor entendimento dos mecanismos que nortearam a rápida diversificação e evolução deste grupo de peixes. A associação de dados cromossômicos e genômicos permitem o entendimento mais amplo acerca da origem e evolução do cromossomo B. Essas iniciativas abrem novas perspectivas para estudos quanto a funcionalidade desse elemento, uma vez que essa questão ainda permanece obscura. Nesse cenário, um bom modelo de estudo para as análises funcionais é a espécie *Astatotilapia latifasciata*, a qual já possui inúmeros trabalhos tanto a nível citogenético quanto genômico permitindo assim uma análise mais robusta incorporando diferentes metodologias.

1.3 O ciclídeo *Astatotilapia latifasciata* como modelo para estudos de cromossomos B

A espécie de ciclídeo africano *Astatotilapia latifasciata*, amplamente difundida no mercado de aquarofilia é natural dos lagos Kyoga e Nawampasa (lagos satélites ao lago Vitória, do Leste Africano), está considerada extinta no primeiro e ameaçada de extinção no segundo lago (Steeves *et al.* 2010). Estudos prévios mostraram a ocorrência de um a dois cromossomos B em indivíduos desta espécie (Poletto *et al.* 2010a, 2010b; Fantinatti *et al.* 2011).

Análises estruturais de citogenética, utilizando o mapeamento de sequências de DNAs ribossomais, do elemento repetitivo SATA, BACs (Bacterial Artificial Chromosomes), além de fração genômica contendo DNA altamente repetitivo (*Cot-1* DNA) e hibridização genômica comparativa (CGH), indicaram a presença de uma grande quantidade de DNAs repetitivos nos cromossomos B desta espécie. Os dados evidenciam a hipótese de uma

possível origem desse elemento proveniente do primeiro par cromossômico do complemento A, o que é suposto ser o seu cromossomo sexual (Poletto *et al.* 2010a; Fantinatti *et al.* 2011).

Paralelamente às investigações em nível citogenético, outras abordagens baseadas em análises genômicas em larga escala serviram para elucidar questões acerca dos cromossomos B. O primeiro estudo em larga escala do genoma do cromossomo B microdissectado de *Astatotilapia latifasciata* forneceu subsídios ainda básicos a respeito de seu conteúdo, entretanto, se mostrou essencial para construção de um marcador por PCR capaz de identificar indivíduos portadores de cromossomo B sem que para isso o espécime fosse eutanasiado (Fantinatti *et al.* submetido), permitindo a continuação de pesquisas a respeito de análises relacionadas a função biológica desse cromossomo quando presente no organismo (Valente *et al.* 2014).

O sequenciamento do genoma de peixes 0B e 2B da espécie forneceu mais dados a respeito da organização genômica e conteúdo gênico nesse cromossomo. Foi observado que o cromossomo B de *A. latifasciata* sofreu invasão de sequências de todos os autossomos sendo mais evidentes as sequências dos grupos de ligação 1, 3 e 9 (Valente *et al.* 2014). Em suma, sugere-se que o cromossomo supernumerário nesta espécie possa ter evoluído a partir de um pequeno fragmento cromossômico seguido pelo processo de duplicação gênica, invasão por várias famílias de DNA repetidos e de outros fragmentos de cromossomos do complemento A (Poletto *et al.* 2010a; Fantinatti *et al.* 2011, Valente *et al.* 2014), como descrito previamente nos trabalhos de Jones e Rees (1982) e Camacho *et al.* (2000) realizados em outras espécies.

Com a análise genômica foi possível verificar o enriquecimento de sequências de diversos cromossomos do complemento A, genes com alta integridade e presença de retrogene hnRNP Q-Like (heterogeneous nuclear ribonucleoprotein Q-Like) descrito pela primeira vez em cromossomos B, indicando uma forte relação da presença de elementos transponíveis na modulação de seu conteúdo gênico (Valente *et al.* 2014).

Apesar da maioria dos genes encontrados no cromossomo B estarem fragmentados, um número considerável encontra-se intacto. Dentre os genes conservados, muitos foram identificados atuando no processo de divisão celular (organização de microtúbulos, estrutura do cinetocoro e recombinação) e receptores olfativos (Valente *et al.* 2014). A descoberta destes genes levanta a questão de saber se esses estão ativos e qual o papel que desempenham em relação ao cromossomo B, para isso, uma metodologia que vem ganhando mais destaque atualmente é a análise de transcriptomas.

1.4 Transcriptomas na elucidação de aspectos funcionais celulares

Os avanços recentes nas tecnologias de sequenciamento em larga escala e dos novos algoritmos de análise e montagem de sequências, reduziram drasticamente o tempo, custo e dificuldade na decifração de novos genomas (Mardis 2011), permitindo o desenvolvimento de projetos ousados como o “Genome 10K” (genome10k.soe.ucsc.edu/) que objetivou sequenciar 10 mil genomas de vertebrados obtendo um rank atual de 102 genomas de diferentes espécies (acesso em 28/06/2016), o “i5K” (arthropodgenomes.org/wiki/i5K), que pretende sequenciar genomas de 5 mil espécies de insetos contendo atualmente 58 genomas completos e 120 genomas em desenvolvimento (acesso em 28/06/2016) e o “Fish-T1K” (fisht1k.org/) que visa gerar o transcriptoma de 1000 peixes da classe Actinopterygii e que até o momento forneceu dados de 180 espécimes, representando 142 espécies, cobrindo 51 ordens e abrangendo 109 famílias (acesso em 20/06/2016).

Embora estas abordagens tenham grande potencial nos estudos comparativos e mesmo de anotação gênica, o conhecimento de muitos aspectos relacionados à função biológica permanecem ainda limitados (Hass *et al.* 2008). Neste cenário, o advento das técnicas de sequenciamento de RNA (RNA-seq) em larga escala vem permitindo investigações mais profundas a cerca da expressão global de genes nos mais diversos grupos de organismos, incluindo plantas (Sánchez *et al.* 2012; Castilho e Buell 2012; revisado em Staiger *et al.* 2015; Jin *et al.* 2016), insetos (Vogel *et al.* 2011; Ferreira *et al.* 2013; Yu *et al.* 2016), peixes (Salem *et al.* 2007; Serrana *et al.* 2012; Schunter *et al.* 2014; Sun *et al.* 2016), entre outros, possibilitado a recuperação de transcriptomas completos.

Essa abordagem propiciou um salto de uma forma pontual de análise, para uma escala que permite análise completa dos genes transcritos em determinado organismo, tecido, célula ou condição biológica (Logan and Buckley, 2015; Yu *et al.* 2016). Esse novo enfoque de análise de expressão gênica, tem sido amplamente difundido por se mostrar vantajoso em estudos baseados em espécies não-modelo, uma vez que não há a limitação de análises pela falta de um genoma de referência, além de permitir a quantificação dos níveis de expressão e identificação de polimorfismos de único nucleotídeo (SNP) (Sánchez *et al.* 2012; Logan and Buckley 2015).

Um estudo interessante que evidencia as vantagens na utilização de transcriptomas foi o trabalho com o peixe *Tripterygion delaisi* no qual foram utilizadas análises de cérebro de machos e fêmeas no período reprodutivo e os resultados evidenciaram que durante o período

de reprodução a plasticidade fenotípica é um fator mais importante na expressão diferencial do que genes relacionados ao dimorfismo sexual nessa espécie (Schunter *et al.* 2014).

A metodologia também fornece a possibilidade de incorporação de outros dados de sequenciamentos de larga escala como no caso do estudo desenvolvido com piraquê *Electrophorus electricus* o qual analisou o transcriptoma e o mirnoma com o intuito de avaliar o perfil transcricional de genes e de elementos regulatórios em diferentes tecidos da espécie (Traeger *et al.* 2015). Abordagem semelhante foi empregada em um estudo envolvendo a síndrome de klinefelter associada a assinaturas epigenéticas afim de evidenciar as possíveis alterações a nível transcricional e epigenômico no cérebro de portadores dessa síndrome (Viana *et al.* 2014).

No que se refere a anotação e análise genômica, a utilização de transcriptomas mostrou-se promissora na resolução de questões como no caso da espécie de ave *Parus major*, em que foram utilizados as sequências transcritas de RNA, afim de construir um primeiro transcriptoma de referência para esse grupo (Santure *et al.* 2011) e permitiu também a investigação de genes parálogos, splice variantes e distribuição de microssatélites de músculo esquelético em dourada (*Spaurus aurata*) (Serrana *et al.* 2012).

O desenvolvimento dessa metodologia, propiciou também uma análise real dos mecanismos envolvidos na biologia do organismo, como por exemplo, a análise do perfil transcricional frente a privação de alimento em *Oncorhynchus mykiss* com o intuito de se entender de que forma se dá a resposta proteolítica para obtenção de energia (Salem *et al.* 2007), além da identificação e monitoramento de mecanismos biológicos envolvidos a respostas de estresse como as variações de temperatura e salinidade em truta (Sanchez *et al.* 2011). Ligando as respostas de variações transcriptômicas às mudanças em níveis mais elevados de organização aumentando muito o poder da genômica funcional para explorar as interações organismo-ambiente em escalas amplas (Kassahn *et al.* 2009; Logan and Buckley, 2015).

Inúmeros trabalhos relatam estudos sobre a análise do perfil transcricional em resposta às diferentes variações ambientais e condições biológicas, porém, até o presente momento, pouco se sabe em relação a análise de transcriptomas de espécies que apresentam um elemento extra numerário (cromossomo B). Devido aos inúmeros exemplos, a presente metodologia se mostra bastante eficiente e pode auxiliar a descrição do comportamento transcricional de genomas com o cromossomo B, buscando entender não somente a função biológica, como possivelmente, o mecanismo que propicia a manutenção do B no genoma desses organismos.

2. Objetivos

2.1 Objetivo Geral e Justificativa

A geração do sequenciamento do cromossomo B microdissectado, bem como a análise mais aprofundada do genoma 0B e 2B evidenciaram um maior entendimento a respeito da origem e evolução do cromossomo B na espécie *Astatotilapia latifasciata*. Foram identificados genes com alta integridade e com potencial a serem transcricionalmente ativos, dando novas perspectivas para explorar aspectos funcionais dos genes presentes neste cromossomo.

Dessa forma, a partir de dados de sequenciamento genômico do transcriptoma obtidos por *next generation sequencing*, o objetivo do presente trabalho foi analisar o perfil transcricional de tecidos-alvo de *A. latifasciata* afim de investigar possíveis interferências acarretadas pelo cromossomo B sobre o nível de expressão de transcritos em indivíduos com presença (B+) e ausência (B-) desse cromossomo extra.

Foram obtidas amostras em triplicatas de cérebro, músculo e gônadas que, com exceção de cérebro, foram provenientes de mesma geração parental F2. Foram considerados a presença e ausência do cromossomo B e o sexo dos indivíduos. Os resultados visam contribuir ainda mais para as discussões acerca do modo como o cromossomo B escapa dos processos de eliminação celular e se mantém presente nas populações além de fornecer dados relevantes acerca da modulação da expressão gênica em indivíduos hospedeiros do cromossomo B.

2.2 Objetivos específicos

- Construir o transcriptoma de referência de *A. latifasciata* através da montagem *de novo* utilizando os três tecidos sequenciados;
- Diagnosticar os níveis globais de expressão gênica nos diferentes tecidos;
- Identificar quais genes estão diferencialmente expressos nas amostras B+ e B-.

3. Hipótese

O cromossomo B de *A. latifasciata* possui genes ativos e é capaz de interferir na modulação do perfil de expressão de genes do complemento cromossômico A permitindo sua manutenção no genoma.

4. Materiais e Métodos

4.1 Obtenção das amostras de RNA

Foram analisados espécimes do ciclídeo africano *Astatotilapia latifasciata*, amplamente difundida no mercado de aquarioria. Esta espécie é natural dos lagos Kyoga e Nawampasa (lagos satélites ao lago Vitória) e está considerada extinta no primeiro e ameaçada de extinção no segundo lago (Steeves *et al.* 2010). Atualmente esta espécie é mantida no biotério de peixes do Laboratório Genômica Integrativa, UNESP-Botucatu.

Todos os procedimentos de manipulação dos animais estão de acordo com os princípios éticos em pesquisa adotado pelo Colégio Brasileiro de Experimentação Animal aprovado pelo Instituto de Biociências / UNESP - comitê de ética no uso de animais (Protocolo n.º. 769).

4.2 Genotipagem de animais B+ e B-

A genotipagem de *A. latifasciata* quanto a presença/ausência/quantidade de cromossomos B foi realizada com base em dois marcadores previamente desenvolvidos. Um primeiro marcador está baseado em reação em cadeia da polimerase (PCR) e foi desenvolvido com base nas sequências geradas (plataforma 454 Jr. Roche) para um cromossomo B microdissectado da espécie (Fantinatti *et al.* em preparação). A partir de análises comparativas entre o draft gerado e genomas de outros teleósteos depositados no *National Center for Biotechnology Information* (NCBI), foram construídos conjuntos de três primers com anelamento em diversas regiões do genoma. Os *primers* foram projetados de forma que reações utilizando o DNA de animais B+ resultassem duas bandas ao passo que as reações utilizando DNA de animais B- apresentassem somente em uma banda (banda controle) (Fantinatti *et al.* submetido).

Embora os primers projetados anteriormente mostraram-se muito eficientes na genotipagem de animais B+ e B-, a identificação da presença de 1 ou 2 Bs permanecia ainda obscura. Dessa forma, um segundo marcador foi desenvolvido com base nos dados de sequenciamento do genoma total de animais 0B e 2B (Valente *et al.* 2014). Este novo marcador foi baseado em PCR em tempo real (qPCR) e permite identificar se o espécime possui um ou dois cromossomos B (Fantinatti *et al.* em preparação).

4.3 Extração do RNA total

O RNA total foi extraído usando Trizol (*life technology, EUA*) de acordo com as especificações do fabricante seguida pela análise por espectrofotometria para fins de quantificação do material extraído utilizando o equipamento *Nano Vue (General Electrics, EUA)* com a integridade do RNA analisada pela determinação do “RIN” (*RNA Integrity Number*) utilizando o equipamento *2100 Bioanalyzer (Agilent, EUA)*. Sendo selecionadas para o sequenciamento amostras com RIN superior a 8.

4.4 Sequenciamento e processamento dos reads

Foram geradas 36 bibliotecas (12 amostras) de transcriptomas de *A. latifasciata* divididas em triplicatas de gônadas, cérebro e músculo de indivíduos machos e fêmeas com presença e ausência de cromossomo B objetivando produzir bibliotecas de transcritos com pelo menos 30 milhões de reads (Tabela 1).

Tabela 1: Relação do número de bibliotecas sequenciadas para cada tecido. Amostras relacionadas a presença de cromossomo B e sexo.

	B +		B-		Total
	Macho	Fêmea	Macho	Fêmea	
Cérebro	3	3	3	3	12
Músculo	3	3	3	3	12
Gônadas	3	3	3	3	12
Total	9	9	9	9	36

As amostras de RNA foram sequenciadas utilizando a plataforma Illumina Hi-seq 2000, 100 x 2 *paired end reads* (4 amostras por lane). Este serviço foi realizado pela empresa LC Science (EUA).

Cada biblioteca foi submetida ao *software* FASTQC para análise de qualidade dos reads além da identificação de sequências altamente representadas. A eliminação de contaminantes do sequenciamento e seleção por qualidade do read, identificados pelo *software* foi realizada utilizando o pacote Trimmomatic V0.30 (Bolger *et al.* 2014) com valor de phred 30 de qualidade de reads em pelo menos 80% deste (Material suplementar S1).

Os *reads* restantes foram alinhados contra um banco de dados de sequências de contaminantes, com 3,126,425 sequências de rDNA, tRNAs, genoma mitocondrial e bacteriano obtidos no banco de dados Refseq (NCBI); os alinhamentos dos *reads* contra o

Refseq database (NCBI) foram gerados utilizando o programa Bowtie2 com a opção “--very-fast-local”. Os *reads* não alinhados foram submetidos a verificação de adaptadores Illumina usando a pesquisa blastn, as sequências que mostraram similaridade (não importando a posição) foram excluídos ($e_valor \leq 10e^{-5}$ e 90% de identidade foram os parâmetros de corte). Os reads não pareados foram separados pelo programa Pairfq. As bibliotecas relacionadas a um fragmento (*paired end*) foram agrupados criando dois conjuntos de dados para montagem de novo.

4.5 Montagem *de novo*

Para a montagem *de novo* foi utilizada a estratégia SAMP (Single-Assembler Multiple-Parameter) usando os programas Velvet versão 1.2.08 (Zerbino e Birney 2008) e Oases versão 0.2.08 (Schulz *et al.* 2012), como recomendado para espécies diplóides (He *et al.* 2015).

A montagem foi iniciada utilizando uma escala de *k-mers* de 21 a 81 nucleotídeos alternando em 4 nucleotídeos de um para outro obtendo-se 14 montagens independentes usando o Velvet (parâmetros “-ins_length 300 -min_trans_lgth 100”) seguido do Oases (parâmetros “-ins_length2 300 -min_trans_lgth 100”). Os *reads* foram mapeados em cada montagem (opção single read mode and “--very-sensitive”) e a média do mapeamento foi calculada, as montagens que apresentaram porcentagem de mapeamento $\geq 82.8\%$ foram selecionada para o próximo passo (material complementar S2, linhas 2-22).

Para a montagem final, foi utilizado o programa CAP3 (Huang e Madan 1999) em dois passos. A montagem consenso utilizando as montagens de *k-mers* 41 a 73 (9 montagens) foram agrupadas em um único arquivo, transcritos porcentagem de “N” $\geq 10\%$ foram excluídos (material complementar S2, linhas 24-34). Devido ao grande número de sequências (1,435,763 transcritos) o arquivo da montagem foi dividido em dois grupos, um com *k-mers* de escala 41 a 57 e o segundo com escala de 61 a 73 nucleotídeos. Foi estipulado medindo a quantidade de dados que o computador poderia suportar para uma única corrida CAP3. As primeiras montagens do CAP3 foram realizadas para cada grupo, separadamente, utilizando a combinação de parâmetros (-f 5 ou 20 e -p escala de 66-90; outros parâmetros foram “-h 20 -k 1 -r 1”) em uma escala (quantidade de *singlets* / quantidade de contigs), que poderia expressar a melhor montagem no sentido de ter o máximo de contigs quanto possível em comparação com o *singlets* (menor taxa avaliada para cada conjunto de dados), foram selecionadas as melhores montagens (parâmetros do CAP3 “-f 5 -h 20 -k 1 -p 66 -r 1”)

(material complementar S2, linhas 36-52).

Os melhores *contigs* e *singlets* gerados pelo CAP 3 foram utilizados na montagem de ambos os grupos sendo submetidos a uma montagem final, utilizando os mesmos parâmetros (material complementar S2, linhas 54, 56-58). O mesmo mapeamento e filtragem foi feita para os transcritos ≤ 300 pb (não utilizada nesse trabalho). Por fim, o transcriptoma final foi constituído por 203,968 sequências com tamanho médio de 1,879 pb (material complementar S2, linha 61).

4.6 Diferencial de expressão, identificação de transcritos redundantes entre as amostras e anotação funcional

Para a identificação dos níveis de expressão foi utilizado o pacote DEseq do R-Bioconductor (Gentleman *et al.* 2004; Anders e Huber 2010) e foram considerados apenas transcritos com valor de $p_{adj} \leq 0,05$.

Para a construção dos diagramas de venn foi utilizado o pacote Venny 2.0 (Oliveros 2015). A anotação dos genes diferencialmente expressos em comum das amostras com presença de cromossomo B foi realizada utilizando o *software* Blast2GO (Götz *et al.* 2008), as sequências foram investigadas utilizando o banco de dados de sequências não redundantes (nr) usando o BLASTx com *e-value* de 0.001 seguindo os parâmetros padrão do programa.

5. Resultados e discussão

O presente tópico será apresentado em formato de manuscrito, a ser submetido para publicação em revistas científicas da área de genética, genômica ou evolução.

B chromosomes effects on gene transcription patterns in the cichlid fish

Astatotilapia latifasciata

Marques DF¹, Conte MA², Valente GT³, Nakajima RT¹, Coan RLB¹, Fantinatti BEA¹, Kocher TD², Martins C¹

1. Department of Morphology, Institute of Biosciences, UNESP - Sao Paulo State University, Botucatu, SP, Brazil.

2. Department of Biology, University of Maryland, College Park, MD, USA.

3. Bioprocess and Biotechnology Department, Agronomical Science Faculty, UNESP – Sao Paulo State University, Botucatu, SP, Brazil.

Abstract

Supernumerary chromosomes (B chromosomes) are extra elements not homologous to the regular complement and have been described in a range of species representing most eukaryote lineages. Recent studies with these elements have been focused on structural level, and functional aspects have been poorly investigated. B chromosomes have been described in 21 cichlid fish species, including *Astatotilapia latifasciata* from Lake Victoria system that can harbors 1 or 2 large heterochromatic B chromosomes. Cytogenomics analysis described the presence of thousands of highly fragmented genes in the B chromosome of this species. Besides the advances on the genomic content of this element, its functional relevancy is still to be elucidated. Considering the relevance of B chromosomes, mRNA transcriptional profiles of brain, muscle and gonads of *A. latifasciata* were obtained in order to investigate possible interference modulated by B chromosome presence or absence. A range of transcripts evidenced in the analyzed tissues shows that B chromosome may be interfering in different pathways. Our data contributed to the discussion of functional aspects of B chromosome, identifying several genes related to a cell cycle, cell structure maintenance and immunity response. We identified few up regulated genes located on the B chromosome. On the other hand, most genes differential expressed are located in the regular chromosome complement and are affected by the presence of this extra element. Furthermore, our data suggest a possible repressive gene expression action of B chromosome in tests and increasing expression level of genes in ovarian confirming the occurrence of drive mechanism of this chromosome in female gametic cells.

Keywords: Supernumerary chromosome, gene expression, transcriptome, functional genomics.

Introduction

B chromosomes (Bs) are extra chromosomal elements present in a huge range of eukaryotic species, reported over 1,300 species of plants and 500 species of animals (Camacho 2005). The heterochromatic content of Bs has highlighted them as non-functional, parasitic and dispensable chromosomes for their hosts (Camacho 2000; Jones *et al.* 2008). Although many studies report deleterious effects when the number of Bs per cell increases, their presence in lower number in many taxa suggests beneficial effects (Jones & Rees 1982; revised in Burt and Trivers 2006; Valente *et al.* 2014). Cytogenetic and genomic evidence shows the B chromosomes originated from the standard chromosomes (A complement) and were subsequently invaded by repetitive elements (Utsunomia *et al.* 2016; Piscor and Parise-Maltempi 2016), duplicated/degenerated genes, retrogene insertions (Valente *et al.* 2014) and organelle genome sequences (Martis *et al.* 2012).

B chromosomes showed an unstable behavior during cell division enabling non-Mendelian inheritance, transmission rates exceeds those normal chromosomes and their irregular mitotic and meiotic behavior allows accumulation selfishly in the germline (Camacho, 2000). Meiotic accumulation of Bs has been described from female meiosis in various species of plants and animals as a consequence of asymmetric inheritance of this chromosome during chromosome segregation in meiosis; the B chromosome migrates preferentially to the secondary oocyte instead of to the first polar body (revised in Burt and Trivers 2006). A failure to pair with standard chromosomes prompted a model of clonal inheritance (Camacho 2005). Maintenance of these chromosomes in so many species suggests a segregation advantage during meiosis and instability during mitosis (Banaei-Moghaddam *et al.* 2013).

In animals, B chromosomes have been extensively described among fishes, including cichlids, being described in seven South American (Feldberg and Bertollo, 1984; Martins-Santos *et al.* 1995; Feldberg *et al.* 2004) and fourteen African species (Poletto *et al.* 2010a, 2010b; Fantinatti *et al.* 2011; Yoshida *et al.* 2011; Kuroiwa *et al.* 2014). Supernumerary chromosomes have been detected in all 13 species of cichlids karyotyped from Lake Victoria (Poletto *et al.* 2010a; Yoshida *et al.* 2011; Kuroiwa *et al.* 2014). On the other hand, among 14 karyotyped species of Lake Malawi, B chromosomes were detected only in *Metriaclima lombardoi* (Poletto *et al.* 2010b).

Astatotilapia latifasciata can carry zero, one or two B chromosomes (Poletto *et al.* 2010; Fantinatti *et al.* 2011). Molecular cytogenetic data reveals B chromosome enriched of

transposable elements, as well as 18S rDNA sequences and satellite DNAs (Poletto *et al.* 2010; Fantinatti *et al.* 2011). Comparative molecular cytogenetic mapping showed B chromosome sharing genome segments with the first pair of the standard complement (Poletto *et al.* 2010a; Fantinatti *et al.* 2011; Valente *et al.* 2014) suggesting this chromosome contributed to the origin of the B. Additionally, the distribution of DNA sequences in the B implying the formation of an isochromosome (Fantinatti *et al.* 2011) followed by invasion of genes and transposable elements (Valente *et al.* 2014). Large scale genomic analysis in this species identified thousands of duplicated sequences from essentially all chromosomes of the ancestral karyotype. Although most genes on the B chromosome are fragmented, few are largely intact. These mostly intact sequences include genes involved in the cell cycle (microtubule organization, kinetochore structure, recombination and progression through the cell cycle) and olfactory receptors. The discovery of these genes raises the question of whether these genes are active and what role they play in the hosts (Valente *et al.* 2014).

Several functional analyses have evidenced transcripts derived from B chromosomes. It was observed in *Drosophila albomicans* transcribed sequences with similarity to the subcentromeric regions of both the ancient X and the neo-X chromosome (Zhou *et al.* 2012); In the rodent *Capreolus pygargus* expressed partial sequence from gene FPGT-TNNI3K (Trifonov *et al.* 2013) was also detected; rye *Secale cereal* expresses pseudogene-like fragments (Carchilan *et al.* 2007; Carchilan *et al.* 2009; Banai-Moghaddan *et al.* 2013; 2015); and the grasshopper *Eyprepocnemis plorans* expresses rRNA (Teruel *et al.* 2010; Ruiz-Estévez *et al.* 2012). Those findings attest the potential of B chromosomes to harbor transcriptionally active genes.

The development of RNAseq methods has been a large step for functional studies based in the analysis of whole transcriptomes of cells, tissues or entire organisms (Salem *et al.* 2007; Santure *et al.* 2012; Serrana *et al.* 2012, Logan and Buckley 2015). RNAseq is particularly advantageous for studies of non model organisms even in the absence of a reference genome sequence (Sanchez *et al.* 2012, Logan and Buckley 2015). NGS has contributed to important discoveries about the origin and evolution of supernumerary chromosomes and their gene content (Goodwin *et al.* 2011; Martis *et al.* 2012; Valente *et al.* 2014). Although these recent reports have shown B chromosomes having more genes than previously expected, information about their function remains scarce with just punctual investigations. Here we investigate the functional effect of B chromosomes and their possible contribution in the transcriptional regulation in *Astatotilapia latifasciata*. Our data show that the B chromosome alters the expression of many genes, and provide new insights about the mechanisms by how B

chromosomes could have been maintained at high frequency in cells and organisms.

Methods

Animal samples

The cichlid fish *A. latifasciata* (Pseudocrenilabrinae) is native to lakes Kyoga and Nawampasa (satellite lakes of Lake Victoria in Africa) (Steeves *et al.* 2010). Previous classical and molecular cytogenetic studies detected the presence of one or two B chromosomes in individuals from populations farmed in Brazil for the aquarium hobbyist market (Poletto *et al.* 2010; Fantinatti *et al.* 2011). Currently, a stock of this species containing animals with 0B, 1B and 2B chromosomes is maintained in the fish facility of the Integrative Genomics Laboratory at São Paulo State University.

Biological samples were handled according to ethical principals adopted by Brazilian College for animal experiments accepted by the Biosciences Institute, UNESP-Sao Paulo State University ethic committee on use of animals (Protocol no. 769-2015).

Genotyping samples for 0B, 1B and 2B chromosomes

The presence of B chromosomes was identified by polymerase chain reaction (PCR) and quantitative PCR (qPCR) based on specific B chromosome polymorphisms (Fantinatti and Martins, not published). Primers were designed to amplify a fragment of 186bp for the 0B, 1B and 2B genomes and a 283 bp fragment exclusively for the 1B and 2B samples. The qPCR was designed to quantify the copy number of specific B segments allowing the identification of 1B or 2B samples based on the copy number variation among the samples. For the present work we analyzed 0B and 1B chromosome samples.

Sample collection, RNA preparation and sequencing

Muscle, gonads and brain tissues were collected from three adult males and females with either 0B or 1B chromosomes. The samples were frozen on liquid nitrogen and storage in -80°C until use. RNA extraction was performed using TRI reagent (Molecular Research Center). Tissues were homogenized using an Ultra Turrax homogenizer (MM200) at 30Hz for two intervals of 30 seconds. Phase separation was done with 1:1 volume of chloroform

(Invitrogen) and centrifuged for 15 minutes at 12.000g at 4°C. RNA was precipitated with 1:2 volume of isopropanol (Invitrogen) and centrifuged at 12.000g for 10 minutes at 4°C. The total RNA pellet was washed twice with 75% ethanol by vortexing and centrifuging for 5 minutes at 7.500g and finally eluted in 30µl of ultra pure water. The RNA was quantified by spectrophotometry using a NanoVue (GE Healthcare Life Sciences, USA) and the quality of RNA was obtained as the RNA integrity number by bioanalyzer (RNA Integrity number, RIN ≥ 8). Total RNA samples were used for library construction and sequencing using the Illumina Hiseq-2000platform.

Reads filtering

Were sequenced 36 libraries (18 samples) being 12 samples to each tissue (brain, muscle and gonads), low quality reads were subjected to the filtering process by Phred quality of 30 using the software Trimmomatic V0.30 (Bolger *et al.* 2014) reads less than 80bp were removed. The selected reads were aligned against a “contaminant database” which has 3,126,425 sequences performing rDNAs, tRNAs, mitochondrial and bacterial genome from Refseq database (NCBI); those alignments were doing using Bowtie2 with the “--very-fast-local” option. The unmapped reads were subjected to Illumina adaptors checking using blastn search, which reads with some hit (does not care the position) were excluded (e-value $\leq 10e^{-5}$ and 90% of identity are the cutoff parameters). The not paired reads were excluded by Pairfq software. Libraries related to one fragment paired-end were merged creating two datasets for de novo assembling (see Supplementary S1).

The de novo assembling and filtering process

For de novo assembling it was used the SAMP strategy (Single-Assembler-Multiple-Parameter) using Velvet and Oases, as recommended for diploid species (He *et al.* 2015).

We started assembling using a ranging of k-mers from 21 to 81 with 4-mers jump, giving us fourteen independent assembling using Velvet (“-ins_length 300 -read_trkg yes -min_contig_lgth 100 -conserveLong yes” parameters) followed by Oases (“-ins_length2 300 -min_trans_lgth 100” parameters). Reads were mapped over each assembling (single read mode and “--very-sensitive“ option) and the average of mapping was calculated, in which if some assembling has a mapping percentage $\geq 82.8\%$ were selected for the next step (see Supplementary S2, lines 2-22).

For the final assembling it was used CAP3 in two steps. The assembling using k-mers 41-73 (9 assembling) were merged in a single file and the transcripts with “N” $\geq 10\%$ were excluded (see Supplementary S2, lines 24-34). Due to the high amount of sequences (1,435,763 sequences) the assembling files were split in two groups - one with k-mers ranging from 41 to 57 and another one ranging from 61 to 63. It was stipulated by measuring how much data the computer could support for a single CAP3 running. The first CAP3 assembling were performed for each group, separately, using many combination of parameters (-f 5 or 20 and -p ranging from 66-90; other parameters were “-h 20 -k 1 -r 1”). Using a kind of rate (amount of singlets/amount of contigs) that could express the best assembling in the sense to have as much contigs as possible compared to the singlets (lower rate was better for each dataset), we selected the best first assembling (“-f 5 -h 20 -k 1 -p 66 -r 1” CAP3 parameters) (see Supplementary S2, lines 36-52). The contigs and singles generated by these first best CAP3 assembling of both groups were subjected to a final assembling using the same parameters and program as well (see Supplementary S2, lines 54, 56-58). The transcripts lower than 300 bp were excluded from both contigs and singlets file and (see Supplementary S2, lines 60-61) resulting an assembly with 203,968 transcripts.

Differential expression, identification of redundant transcripts between samples and functional annotation

DEseq package from R-Bioconductor (Gentleman *et al.* 2004; Anders and Huber 2010) was used to identify the expression levels. Only transcribed with *adj* ≤ 0.05 were considered.

The Venny 2.0 package (Oliveros 2015) was used for the construction of Venn diagrams. The annotation of differentially expressed transcripts in common of the samples with the presence of chromosome B was performed using Blast2GO software (Götz *et al.* 2008), sequences were investigated using the database of non-redundant sequences (nr) using BLASTx with an e-value of 0.001 following the default program parameters.

Results

Transcriptome assembly

Thirty-six libraries from brain, muscle and gonads of males and females were sequenced generating a total of 1,812,771,500 reads and after trimming 503,825,612 reads remain. Approximately 30% of reads were acceptable for the analysis composed only by paired-end reads (supplementary data S1).

In the first step for the de novo assembly, kmer size of 41 to 69 nucleotides recovered better assemblies individually with good metrics (avg length 927 to 1,047 bp). The second step of merging the assemblies reconstructed the reference transcriptome contained 203,968 contigs and average length of 1,879 nucleotides (supplementary data S2).

Differential expressed transcripts

Differential expression (DE) analysis was performed in four different comparisons related to presence/absence of B chromosome and sex. It was observed in brain eight DE transcripts between sex, including three up regulated in females and five in males (Supplementary data S3). We identified 219 DE transcripts between 0B and 1B individuals including 64 are up regulated in 1B and 155 in 0B samples (Supplementary data S4). Within each sex, it was seen 118 DE transcripts in females with 97 up regulated in females 1B and 21 in females 0B (Supplementary data S5). It was identified 149 DE transcripts in males including 96 up regulated in males 1B and 53 in males 0B (Supplementary data S6) (Figure 1).

The DE results from muscle showed a different pattern compared to brain. It was identified three DE transcripts between sex: one up regulated in males and two in females (Supplementary data S7). Comparative DE analysis among 0B and 1B samples detected 25 DE transcripts in 0B samples (Supplementary data S8). It was observed eight transcripts up regulated in only 1B female samples (Supplementary data S9). In males it was observed nine DE transcripts with eight up regulated in males 1B and one in males 0B (Supplementary data S10) (Figure 2).

We identified 85,663 gonad transcripts differentially expressed between sex and 45,989 and 39,674 are up regulated in males and females, respectively (Supplementary data S11). Comparisons among 0B and 1B samples did not detect any differential expressed gene

(Supplementary data S12). Sex comparisons identified 88 DE transcripts in females including 81 up regulated in females 1B and seven in females 0B (Supplementary data S13). We observed 438 DE transcripts in males with 46 up regulated in male 1B and 392 in 0B samples (Supplementary data S14) (Figure 3).

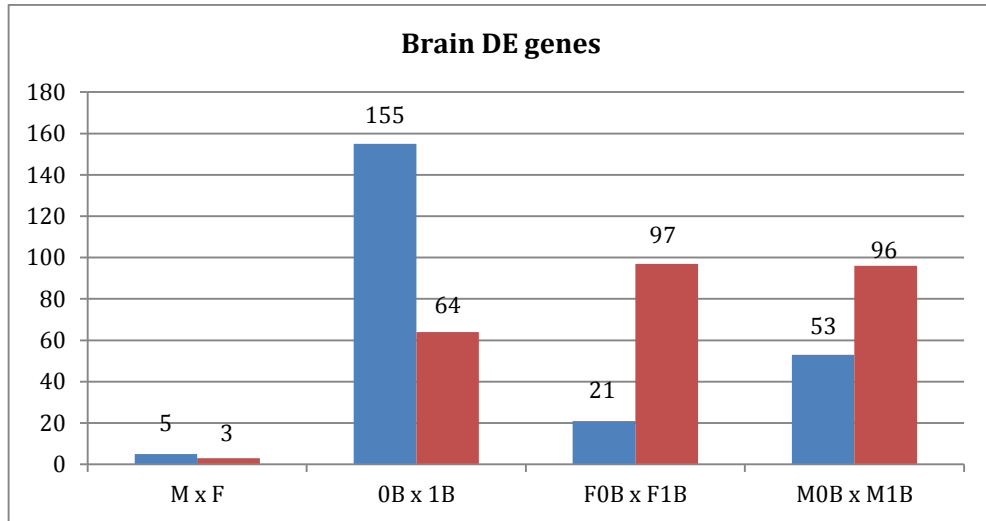


Figure1: Differentially expressed transcripts in brain. Number of DE transcripts (Y axis) in comparisons with samples (X axis). Blue bars indicate male, 0B, female 0B and male 0B; Red bars indicate female, 1B, female 1B and male 1B.

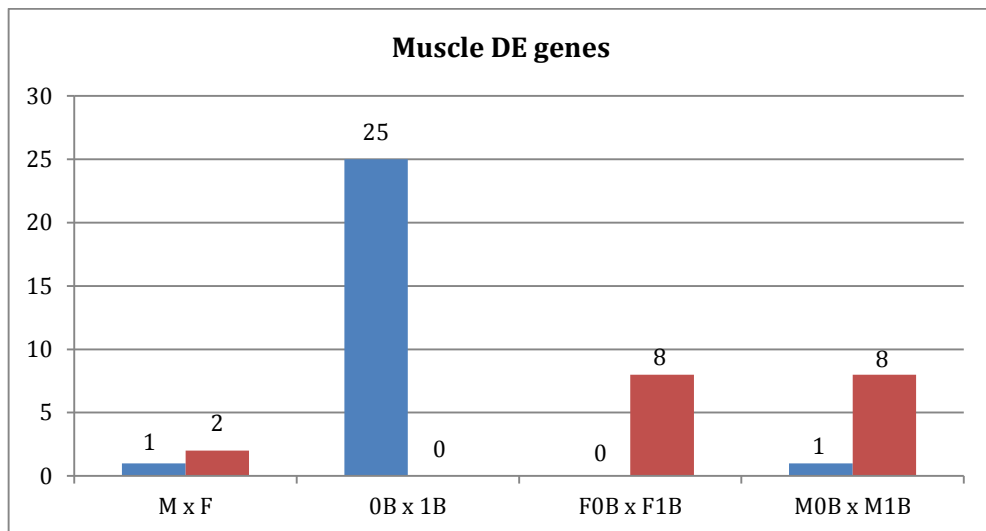


Figure2: Differentially expressed transcripts in muscle. Number of DE transcripts (Y axis) in comparisons with samples (X axis). Blue bars indicate male, 0B, female 0B and male 0B; Red bars indicate female, 1B, female 1B and male 1B.

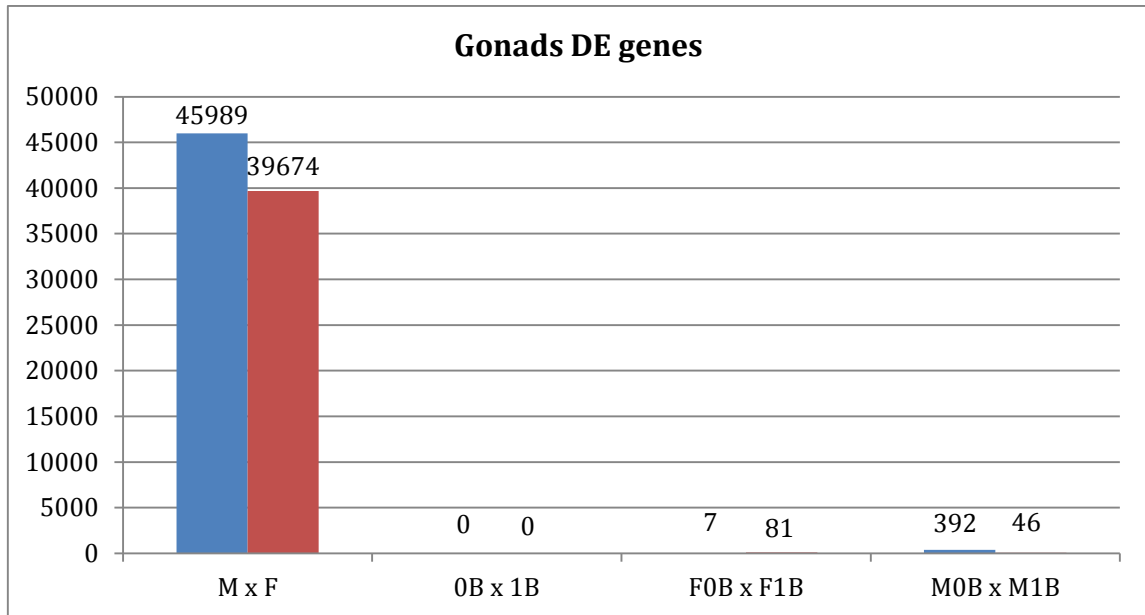


Figure3: Differentially expressed transcripts in gonads. Number of DE transcripts (Y axis) in comparisons with samples (X axis). Blue bars indicate male, 0B, female 0B and male 0B; Red bars indicate female, 1B, female 1B and male 1B.

Transcripts differentially expressed in each condition indicate groups of genes involved with different biological process. Out of all transcripts differentially expressed in the tissue samples investigated, 57 transcripts of brain, seven of muscle, and three of gonads, are common in comparisons of males and females 1B (Fig. 4).

Among transcripts up regulated in 1B samples none of them showed similarity to all tissues. Only was shown similarity between two samples separately (Fig. 5). The basic ontology analysis retrieves 74 transcripts in brain, ten in muscle and 147 in gonads. Blast database search of these transcripts did not recover identity to any known gene/sequence for most of them, but were identified few genes. (Supplementary data S15).

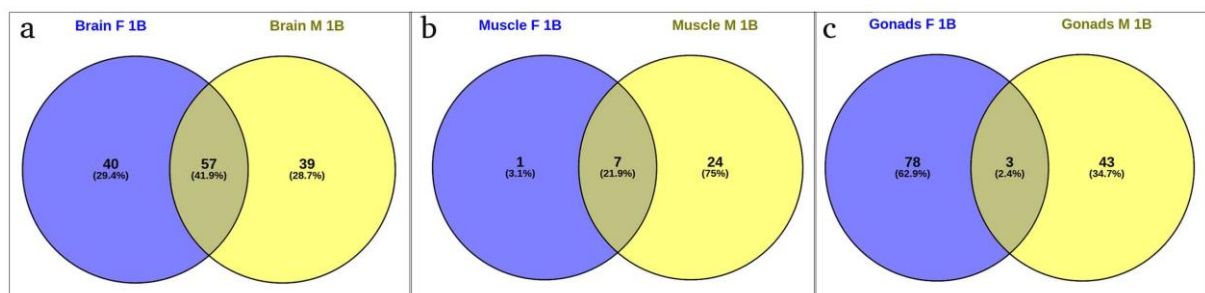


Figure 4: Venn diagrams of DE transcripts of 1B samples in different sex. a) DE transcripts of brain, b) muscle and c) Gonads, the blue circles indicate DE transcripts in female 1B, yellow circles indicate DE transcripts in males 1B.

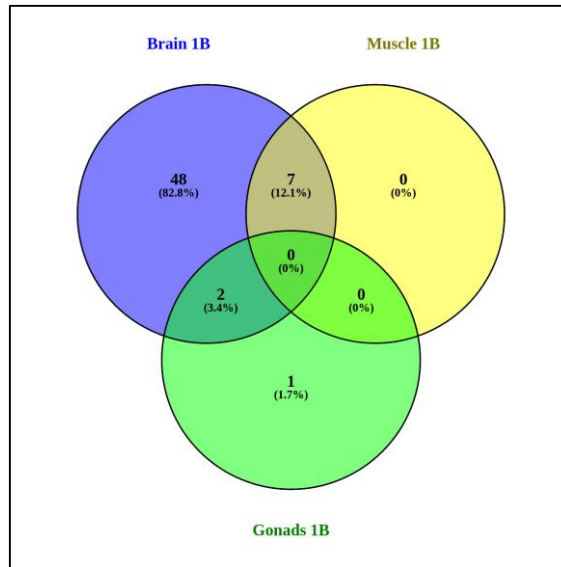


Figure 5: Venn diagram of brain, muscle and gonads up regulated transcripts shared between male and female.

Discussion

Recovered transcriptomes for *A. latifasciata*

The most recent study using NGS technology to investigate B chromosomes revealed several intact genes on B chromosome of the cichlid fish *A. latifasciata* (Valente *et al.* 2014). With exception of few studies involving large scale analysis and B chromosome (Akbari *et al.* 2013; Huang *et al.* 2016), several studies have reported transcripts originated from B chromosome DNA sequences in lower scale and make reference to one or few sequences (Zhou *et al.* 2012; Trifonov *et al.* 2013; Carchilan *et al.* 2007; 2009; Teruel *et al.* 2010; Ruiz-Estévez *et al.* 2012; Banai-Moghaddan *et al.* 2013; 2015).

In the present work, we obtained transcripts of brain, muscle and gonads of B- and B+ samples of *A. latifasciata* based in a large dataset of transcriptomes generated through Illumina sequencing. Our recovered transcriptomes had better quality parameters compared to datasets obtained for other fish species as gilthead sea bream (*Sparus aurata*) (454-936bp and N50 of 1,269bp) (Serrana *et al.* 2012), black-faced blenny (*Tripterygion delaisi*) (N50 of 298bp) (Schunter *et al.* 2014) and pacu fish (*Piaractus mesopotamicus*) (1,334bp of average length and N50 of 2,772 bp) (Mareco *et al.* 2015). We can see higher differences in our metrics based on multiple kmers approach to reconstruct a reference transcriptome for *A. latifasciata*. These results offer good perspectives for transcriptomes reconstruction and

assembly, giving a strong dataset for further analysis to identify gene expression differences and advance to locate genes and other functional sequences.

Differential expression and functional analysis

The higher number of up regulated DE transcripts in brain of 0B samples evidences that the B chromosome is interfering and down-regulating gene expressions in this tissue. Different from brain, B chromosome presence has increased gene expression levels in muscle and gonads. The decreasing of DE transcripts in 1B male testis could indicate that 1B males could have genes suppressed by the B chromosome.

Several up regulated genes in 1B samples were identified by GO analysis acting over (I) cell cycle, (II) maintenance of cell structure and (III) immunity response (Table 1). Among cell cycle genes, we identified the transcriptional regulator *myc* that activate or repress RNA transcription. The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation (Dominguez-Sola *et al.* 2007). Another gene related to cell cycle identified was A-kinase anchor protein13 (AKAP-13), that encodes a member of the AKAP family, a group of structurally diverse proteins, that have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell (Schwartz, 2001).

Another interesting detected transcript up regulated genes in 1B samples codify the maintenance of chromosomes protein 5-like (SMC5) that represents a large family of ATPases and participates in many aspects of higher-order chromosome organization and dynamics (Losada and Hitaro *et al.* 2005; Nasmyth and Haering, 2005). We also observed high expression of the gene aurora kinase (AURK) that is over expressed in brain and muscle 1B. Copies of AURK were reported before located in the B chromosome of *A. latifasciata* (Valente *et al.* 2014). Aurora kinase is serine/threonine-protein kinase that is essential for cell cycle control (Maraculla *et al.* 2008). The enzyme helps the dividing cell dispense its genetic materials to its daughter cells, play a crucial role in cellular division by controlling chromatid segregation. Defects in this segregation can cause genetic instability, a condition that is highly associated with tumorigenesis (Bolanos-Garcia 2005; revised in Gómez-López *et al.* 2014).

In relation to the process of maintenance of cell structure we detected over expression of the genes responsible for the map7 domain-containing protein, Spectrin beta non-erythrocytic 5 (Stabach and Morrow, 2000), transcription elongation factor 1 (Becker *et al.*

2013; Morrissey *et al.* 2015) and SUN domain-containing protein 1 (Tzur *et al.* 2006; Uzer *et al.* 2015). All of them are involved with organization of microtubules or cytoskeleton possibly favoring interkinetic nuclear migration.

Another biological process that was observed in the DE expressed genes refer to a modulation of genes related to the immune response. Here among up regulated genes we found two genes located in the B chromosome evidenced before by Valente *et al.* (2014) that act in the immune response mechanism as the toll-like receptor-2 in brain that plays a fundamental role in pathogen recognition and activation of innate immunity (Schröder and Schumann 2005; Nachtigall *et al.* 2014) and major histocompatibility complex class i-related gene that is antigen-presenting molecule specialized in presenting microbial vitamin B metabolites, involved in the development and expansion of a small population of T-cells expressing an invariant T-cell receptor alpha chain called mucosal-associated invariant T-cells (MAIT) (reviewed in Hewitt 2003).

Many studies have reported the relation of RNA and heterochromatin acting as a defense against TEs as a mechanism to genome maintenance (Grewal & Elgin 2007; Slotkin & Martienssen 2007; revised by Obbard *et al.* 2008). B chromosomes are widely described enriched of heterochromatin and TEs so we believe that the higher levels of expressed genes related to immunity response could be a reaction by the possible activity of these elements that leads to activating immunity in the B+ cells.

Table 1: Up regulated genes in 1B samples with respective functions. Genes located on the B chromosome are highlighted in bold.

CELL CYCLE			
Genes	Brain	Muscle	Gonads
A-kinase anchor protein13 (AKAP-13)	x		
Aurora kinase (AURK)	x	x	
Regulator myc		x	
Structural maintenance of chromosomes protein 5-like (SMC5)	x	x	
MAINTENANCE OF CELL STRUCTURE			
Genes	Brain	Muscle	Gonads
Cadherin	x		
Cytoskeleton-associated protein 5			x
Transcription elongation factor 2			x
Talin-2-like		x	
IMUNITY RESPONSE			
Genes	Brain	Muscle	Gonads
Toll-like receptor-2	x		

The higher number of DE genes in 1B samples comparing female gonads 0B *versus* female gonads 1B, suggests the extra element of B+ cells interfere in the gene expression as a

alternative to maintain themselves in the germ line cells and benefits the B chromosome to segregate to the gametic female cells. The interference of B chromosome in gonads give support to the drive theory that has been proposed as a mechanism that favors B chromosome migration to the gametic cell in females in some very unusual ways (Kayano 1957; revised in Burt and Trivers, 2006). Many studies have described drive as the mechanism of maintenance of B chromosome in several species such as plants and animals (Jones 1991; revised in Burt and Trivers, 2006). The drive mechanism contributes to a non-Mendelian segregation of the B chromosome increasing the frequency of Bs among female gametes. Despite of emerging theories about benefits of B itself to be maintained in the cell, the molecular machinery over this mechanism remains unclear. Here we propose that the B chromosome of *A. latifasciata* could influence the female meiosis, causing the preferential segregation of the B to the gametic cells.

Conclusion

B chromosomes had already been described involved in sex determination and the evolution of sex chromosomes (revised 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). The identification of differentially expressed genes in B+ chromosome cells in *A. latifasciata* reveals an interference of B chromosomes in a high scale level of gene regulation.

Our results showed that B chromosome act in many ways; contributing with transcripts by his own genes (located in the B) or affecting genes of the A complement. It should be highlighted that B chromosome affects mainly genes related to a cell cycle, cell structure maintenance and immunity response. Such gene modulation control could favor the B survive in the cells.

Our work also evidences that B chromosome presence affects gene expression, increasing DE genes in female gonads. These results adds contribution to understanding the drive mechanism in female gonads and open the perspective for more refined analysis on the reproductive studies in order to have a definitive answer about the genetic basis of B chromosome segregation during cell cycle.

References

As referências deste manuscrito encontram-se reunidas ao final deste volume no item “Referências Bibliográficas”.

6. Conclusões

Atualmente, cromossomos B vem ganhando maior foco nas investigações cromossômicas, afim de identificar os mecanismos pelos quais se mantém no genoma dos hospedeiros. Até o momento, pouco foi esclarecido quanto ao seu comportamento na célula e as maiores descobertas estão relacionadas ao seu conteúdo gênico ou em identificações funcionais de forma pontual, restrita a sequências específicas.

No presente estudo foi evidenciado através da análise em larga escala de RNA uma possível interferência do cromossomo B no genoma do seu hospedeiro, não só contribuindo através de uma ação direta, com a expressão de genes e sequências presentes no cromossomo B, como também modulando a transcrição de genes presentes no complemento A. Outro ponto importante foi a utilização de uma abordagem técnica de montagem, a qual pode ser reproduzida e utilizada como modelo em montagens de transcriptomas de amostras que possuem elementos extra numéricos em seu genoma.

Além dos resultados encontrados, o desenvolvimento do estudo propiciou um ganho extra para aquisição de dados de RNA que já estão sendo utilizados para abordagens funcionais em outros trabalhos que utilizam *A. latifasciata* como modelo, enriquecendo e contribuindo ainda mais para o conhecimento mais preciso do comportamento do cromossomo B na célula bem como seus mecanismos de manutenção e estabilidade durante os ciclos de divisão celular meiótico e mitótico.

7. Referências Bibliográficas

- Adnadevic T, Jovanovic VM, Blagojevic J, Budinski I, Cabrilo B, Bijelic-Cabrilo, Vujosevic M (2014). Possible Influence of B Chromosomes on Genes Included in Immune Response and Parasite Burden in *Apodemus flavicollis*. Plos one. 9:11.
- Akbari OS, Antoshechkin I, Hay BA, Ferree PM (2013). Transcriptome Profiling of *Nasonia vitripennis* Testis Reveals Novel Transcripts Expressed from the Selfish B Chromosome, Paternal Sex Ratio. *Genes Genomes Genetics*. 3: 1597-1605.
- Anders S, Huber W (2010). Differential expression analysis for sequence count data. *Genome Biol*. 11:R106.
- Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A (2013). Formation and Expression of Pseudogenes on the B Chromosome of Rye. *The Plant Cell*. 25: 2536–2544.
- Banaei-Moghaddam AM, Martis MM, Macas JC, Gundlach HB, Himmelbach AA, Altschmied LA, Mayer KFXB, Houben A (2015). Genes on B chromosomes: Old questions revisited with new tools. *Biochimica et Biophysica*. 64–70
- Basheva EA, Torgasheva AA, Sakaeva GR, Bidau C, Borodin PM (2010). A- and B-chromosome pairing and recombination in male meiosis of the silver fox (*Vulpes vulpes* L., 1758, Carnivora, Canidae). *Chromosome Research*. 18:689–696.
- Becker M, Kuhse J, Kirsch J (2013). Effects of two elongation factor 1A isoforms on the formation of gephyrin clusters at inhibitory synapses in hippocampal neurons. *Histochemistry and Cell Biology* 140.6: 603–9.
- Beukeboom LW (1994). Bewildering Bs: An impression of the 1st B-chromosome conference. *Heredity*. 73: 928-996.
- Brawand D, Wagner CE, Li YI, Malinsky M, Keller I et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*. 513: 375–381.
- 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, Trivers (2006). *Genes in Conflict: The Biology of Selfish Genetic Elements*. Copyright. 613: 325-380.
- Cabral-de-Mello DC, Valente GT, Takajima 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 Research*. 20:279-292.
- Camacho JPM, Sharbel TF, Beukeboom LW (2000). B chromosome evolution. *Phil Trans R Soc Lond B*. 355: 163-178.
- Camacho, J.P.M. (2005). B chromosomes. In T.R. Gregory. *The Evolution of the Genome*. San Diego: Elsevier. 223–86.
- Carchilan M, Delgado M, Ribeiro T, Costa-Nunes P, Carpetta A, Morais-Cecílio L, Jones RN, Viegas W, Houben A (2007). Transcriptionally active heterochromatin in rye B chromosomes. *Plant Cell*. 19:1738–1749
- Carchilan M, Kumke K, Mikolajewski S, Houben A (2009). Rye B chromosomes are weakly transcribed and might alter the transcriptional activity of A chromosome sequences. *Chromosoma*. 118:607–616
- Castillo G, Buell R (2013). *Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence*. RSC Publishing.
- Chang X, Kobayashi T, Senthilkumaran B, Kobayashi-Kajura H, Sudhakumari CC, Nagahama Y (2005). Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). *Gen Comp Endocrinol*. 141: 101–115.
- Cnaani A, Lee BY, Zilberman C, Ozouf-Costaz C, Hulata G, Ron M, D'Hount a, Baroiller JF, Cotta HD, Peuman DJ, Tomasino E, Coutanceau JP, Pepey E, Shirak A, Kocher TD (2008). Genetics of sex determination in Tilapiine species. *Sexual Development*. Karger AG. 2:43-54.
- Dominguez-Sola D, Ying CY, Grandori C, Ruggiero L, Chen B, Li M, Galloway DA, Gu W, Gautier J, Dalla-Favera R (2007). Non-transcriptional control of DNA replication by c-Myc. *Nature*. 448: 445-453.
- Fantinatti BEA, Mazzuchelli J, Valente GT, Cabral-de-Mello DC, Martins C (2011). Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genetica*. 139: 1273-1282.
- Feldberg E, Bertollo LAC (1984). Discordance in chromosome number among somatic and gonadal tissue cells of *Gymnogeonhagus balzani*. *Brazilian Journal of Genetics*, Ribeirão Preto, SP. VIII:639-645.
- Feldberg E, Porto JIR, Bertollo LAC (2003). Chromosomal changes and adaptation of cichlid fishes during evolution. In *Fish adaptations* Edited by: Val AL, Kapoor BG. New Delhi and New York: Science Publishers. 285-308.

- Feldberg E, Porto JIR, Brinn MNA, Mendonça MNC, Benzaquem DC (2004). B-Chromosomes in Amazonian Cichlid species. *Cytogenetic and Genome Research*. 106:195-198.
- Ferreira IA and Martins C (2008). Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. *Micron*. 39: 411-418.
- Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldón T, Guigó R, Sumner S (2013). Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biology*. 14:R20.
- Franks TK, Houben A, Leach CR, Timmis JN (1996). The molecular organization of a B chromosome tandem repeat sequence from *Brachycome dichromosomata*. *Chromosoma*. 105: 223-230.
- Gentleman RC, Gentleman RC, Carey VJ, Carey VJ, Bates DM, Bates DM, *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 5:R80.
- 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.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, *et al.* (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 36:3420–35.
- Graphodatsky AS, Kukekova AV, Yudkin DV, Trifonov VA, Vorobieva NV, Beklemisheva VR, Perelman PL, Graphodatskaya DA, Trut LN, Yang F, Ferguson-Smith MA, Acland GM and Aguirre GD (2005). The proto-oncogene C-KIT maps to canid B-chromosomes. *Chromosome Research*. 13:113-122.
- Grewal, S. I. S. & Elgin, S. C. R. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* 447, 399–406.
- Gutknecht J, Sperlich D, Bachmann L (1995). A species specific satellite DNA family of *Drosophila subsilvestris* appearing predominantly in B chromosomes. *Chromosoma*. 103:539-544.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008). Automated eukaryotic gene structure annotation using Evidence Modeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 9:R7
- He B, Zhao S, Chen Y, Cao Q, Wei C, Cheng X, Zhang Y (2015). Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genomics*. 16:65
- Hewitt EW (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology*. 110, 163-169.
- Hoer PACT, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen GJB, den Dunnen JT (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res*. 36:21.
- Huang X, Madan A (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*. 9:868–877.
- Huang W, Du Y, Zhao X, Jin W (2016). B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC Plant Biol* 16:88
- Ijiri S, Kaneko H, Kobayashi T, Wang D-D, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y (2007). Sexual Dimorphic Expression of Gene in Gonads during early differentiation of a Teleost Fish, the Nile Tilapia *Oreochromis niloticus*. *Bio Reprod*. 78:333–341.
- Jamilena M, Ruiz-Rejón C, Ruiz-Rejón M (1994). A molecular análisis of the origin of the *Crepis capillaris* B chromosome. *J Cell Sci*. 107:703-708.
- Jones RN, Rees H (1982). B chromosome. New York: Academic Press.
- Jones RN (1991). B-chromosome drive. *American Naturalist*. 3:430-442.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*. 110:462–467
- Kassahn KS, Crozier RH, Pörtner HO and Caley MJ (2009). Animal performance and stress: responses and tolerance limits at different levels of biological organization. *Biology Reviews*. 84: 277-292.
- Kayano H (1957). Cytogenetic studies in *Lilium callosum*. III. Preferential segregation of a supernumerary chromosome in EMC's. *Proc Japan Acad*. 33: 553-558.
- Kobayashi T, Kajiura-Kobayashi H, Nagahama Y (2002). Two isoforms of vasa homologs in a teleost fish: their differential expression during germ cell differentiation. *Elsevier Science Ireland Ltd*. 111:167-171.
- Kocher TD (2004). Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet*. 5:288-98.
- Kohany O, Gentles AJ, Hankus L, Jurka J (2006). Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics*. 7:474.
- Krovel AV and Olsen LC (2002). Expression of a *vas:EGFP* transgene in primordial germ cells of the zebrafish. *Elsevier Science Ireland Ltd*. 116:141-150.
- Kumar S and Blaxter ML (2010). Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* 11:571.

- Kuroiwa A, Terai Y, Kobayashi N, Yoshida K, Suzuki M, Nakanishi A, Matsuda Y, Watanabe M, Okada N (2013). Construction of chromosome markers from the Lake Victoria cichlid *Paralabidochromis chilotes* and their application to comparative mapping. *Cytogenetic Genome Research*. 142:112–120.
- Logan C and Buckley BA (2015). Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. *The Company of Biologists*. 218:1915-1924.
- López-León MD, Cabrero J, Dzyubenko VV, Bugrov AG, Karamysheva TV, Rubtsov NB, Camacho JPM (2008). Differences in ribosomal DNA distribution on A and B chromosomes between eastern and western populations of the grasshopper *Eyprepocnemis plorans*. *Cytogenet Genome Res*. 121: 260-265.
- Loreto V, Cabrero J, López-León MD, Camacho JPM, Souza MJ (2008). Possible autosomal origin of macro B chromosomes in two grasshopper species. *Chromosome Research*. 16:233-241.
- Losada A, Hirano T (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes*. 11:1269–1287
- Lin HZ, Lin WD, Lin CY, Peng SF, Cheng YM (2014). Characterization of maize B-chromosome-related transcripts isolated via cDNA-AFLP. *Chromosoma* 123: 597-607.
- Maina JN (2000). The highly specialized secretory epithelium in the buccal cavity of the alkalinity adapted Lake Magadi cichlid, *Oreochromis alcalicus grahamsi* (Teleostei: Cichlidae): a scanning and transmission electron microscope study. *J Zool Lond*. 251:427-438.
- Macarulla T, Ramos FJ, Tabernero J (2008). Aurora kinase family: a new target for anticancer drug. *Recent Pat Anticancer Drug Discov*. 3:114–122.
- Maki T, Grimaldi AD, Fuchigami S, Kaverina I, Hayashi I (2015). CLASP2 Has Two Distinct TOG Domains That Contribute Differently to Microtubule Dynamics. *J Mol Biol*. 427:2379–2395.
- Mardis ER (2011). A decade's perspective on DNA sequencing technology. *Nature*. 470:198-203.
- Mareco EA, Serrana DGL, Johnston IA, Dal-Pai-Silva M (2015). Characterization of the transcriptome of fast and slow muscle myotomal fibres in the pacu (*Piaractus mesopotamicus*). *BMC Genomics*. 16:182.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. 18(9):1509-1517.
- Martins C, Cabral-de-Mello DC, Valente GT, Mazzuchelli J, Oliveira SG, Pinhal D (2011). Animal genomes under the focus of cytogenetics. 1. ed. Hauppauge: Nova Science Publisher. 1:160.
- Martins C (2012). Projeto de pesquisa: Integrando citogenética e genômica na compreensão da estrutura e evolução de cromossomos B em eucariotos utilizando o peixe ciclídeo *Astatotilapia latifasciata* como modelo. Relatório Científico Anual. FAPESP.
- Martins-Santos IC, Portela-Castro ALB, Julio HF Jr (1995) Chromosome analysis of 5 species of the Cichlidae family (Pisces, Perciformes) from the Paraná River. *Cytologia*. 60:223–231.
- Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, Scholz U, Gundlach H, Wicker T, Simková H, Novák P, Neumann P, Kubaláková M, Bauer E, Haseneyer G, Fuchs J, Dolezel J, Stein N, Mayer KFX, Houben A (2012). Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *PNAS*. 109:13343–13346.
- Nasmyth K, Haering CH (2005). The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem*. 74: 595–648.
- Mazzuchelli J, Yang F, Kocher T, Martins C (2011). Comparative cytogenetic mapping of Sox2 and Sox14 in cichlid fishes and inferences on the genomic organization of both genes in vertebrates. *Chromosome Research*. 19:657-667.
- McLoughlin DM, Miller CCJ (1996). The intracellular cytoplasmic domain of the Alzheimer's disease amyloid precursor protein interacts with phosphotyrosine-binding domain proteins in the yeast two-hybrid system. *Federation of European Biochemical Societies*. 397:197-200.
- Mestriner CA, Galetti Jr. PM, Valentini SR, Ruiz IRG, Abel LDS, 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.
- Montgomery EA, Huang SM, Langley CH, Judd BH (1991). Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics*. 129:1085-1098.
- Moreton J, Dunham SP, Emes RD (2014). A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: assembly of the duck (*Anas platyrhynchos*) transcriptome. *Frontiers in Genetics*. 5:190.
- Morrissey C, Schwefel D, Ennis-Adeniran V, Taylor IA, Crow YJ, Webb M (2015). The eukaryotic elongation factor eEF1A1 interacts with SAMHD1. *The Biochemical Journal*. 466. 1: 69–76.
- Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 5(7):621-628.
- Nachtigall I, Tamarkin A, Tafelski S, Weimann A, Rothbart A, Heim S, Wernecke KD, Spies C (2014). Polymorphisms of the toll-like receptor 2 and 4 genes are associated with faster progression and a more severe course of sepsis in critically ill patients. *J Int Med Res* 42.1: 93–110.

- Nelson J.S. Fishes of the world, 4thed (John Wiley and Sons, Inc. New York 2006).
- Nokkala S, Kuznetsova V, Maryanska-Nadachowska A (2000). Achiasmata segregation of a B chromosome from the X chromosome in two species of psyllids (Psylloidea: Homoptera). *Genetica*. 108:181-189.
- Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH (1988). A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome. *Science*. 240:512-514.
- Obbard DJ, Gordon KHJ, Buck AH and Jiggins FM (2008). The evolution of RNAi as a defence against viruses and transposable elements. *Phil. Trans. R. Soc. B*. 364, 99–115.
- Ohashi H, Umeda N, Hirazawa N, Ozaki Y, Miura C, Miura T (2007). Expression of vasa (vas)-related genes in germ cells and specific interference with gene functions by double-stranded RNA in the monogean, *Neobenedeniagirellae*. Elsevier Science Ireland Ltd. 37:515-523.
- Oliveros JC (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams. <http://bioinfo.gp.cnb.csic.es/tools/venny/index.html>
- Pearson G, Robinson F, Gibson TB, Xu BE, Karandikar M, Berman K, Cobb MH (2001). Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocrine Reviews*. 22. (2): 153–83.
- Pereira HS, Barão A, Caperta A, Rocha J, Viegas W, Delgado M (2009). Rye Bs disclose ancestral sequences in cereal genomes with a potential role in gametophyte chromatid segregation. *Molecular and Biology Evolution*. 26:1683–1697.
- Piscor D & Parise-Maltempi PP (2016). Microsatellite Organization in the B Chromosome and A Chromosome Complement in *Astyanax* (Characiformes, Characidae) Species. *Cytogenet Genome Res*.
- Poletto AB, Ferreira IA, Martins C (2010a). The B chromosome of the cichlid fish *Haplochromis obliquoidens* harbors 18S rRNA genes. *BMC Genetics*. 11:1.
- Poletto AB, Ferreira IA, Cabral-de-Mello DC, Nakajima RT, Mazzuchelli J, Ribeiro HB, Venere PC, Nirchio M, Kocher TD, Martins C (2010b). Chromosome differentiation patterns during cichlid fish evolution. *BMC Genetics*. 11:5-10.
- Puertas MJ (2002). Nature and evolution of B chromosomes in plants: A noncoding but information-rich part of plant genomes. *Cytogenet Gen Res*. 96:198-205.
- Pyburn TM, Bensing BA, Xiong YQ, Melancon BJ, Tomasiak TM, Ward NJ, *et al.* (2011). A Structural Model for Binding of the Serine-Rich Repeat Adhesin GspB to Host Carbohydrate Receptors. *PLoS Pathog*. 7:7.
- Rajicic M, Adnadevic T, Stamenkovic G, Blagojevic J, Vujosevic M (2015). Screening of B chromosomes for presence of two genes in yellow-necked mice, *Apodemus flavicollis* (Mammalia, Rodentia). *Genetika*. 47-1: 311-321.
- Ruiz-Estévez M, López-León MD, Cabrero J, Camacho JPM (2012). B-Chromosome Ribosomal DNA Is Functional in the Grasshopper *Eyprepocnemis plorans*. *PLOS ONE* 7(5).
- Salem M, Silverstein J, Rexroad III C and Yao J (2007). Effect of starvation on global gene expression and proteolysis in rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics*. 8:328.
- Sánchez C, Weber G, Gao G, Cleveland B, Yao J, Rexroad III C (2011). Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. *BMC Genomics*. 12:626.
- Santure A, Gratten J, Mossman J, Sheldon B, Slate J (2011). Characterization of the transcriptome of a wild great tit *Parus major* population by next generation sequencing. *BMC Genomics*. 12:283.
- Schulz MH, Zerbino DR, Vingron M, Birney E (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics Advance*.
- Schunter C, Vollmer SV, Macpherson E and Pascual M (2014). Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics*. 15:167.
- Serrana D, Estévez A, Andree K, Johnston I (2012). Fast skeletal muscle transcriptome of the Gilthead sea bream (*Sparus aurata*) determined by next generation sequencing. 13:181.
- Schwartz JH (2001). The many dimensions of cAMP signaling. *Proc. Natl. Acad. Sci. U.S.A.* 98. 24: 13482–4.
- Schröder NW, Schumann RR (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis* 5. 3: 156–64.
- Silva D, Pansonato-Alves JC, Utsunomia R, Araya-Jaime C, Ruiz-Ruano FJ, Daniel SN, Hashimoto DT, Oliveira C, Camacho JPM, Porto-Foresti F, Foresti F (2014). Delimiting the Origin of a B Chromosome by FISH Mapping, Chromosome Painting and DNA Sequence Analysis in *Astyanax paranae* (Teleostei, Characiformes). *Plos one*. 9:1-10.
- Skelton, P (2001). A complete guide to the freshwater fishes of Southern Africa. 2th. New Holland publishing. South Africa.
- Slotkin, R. K. & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet*. 8, 272.
- Smith LW, Chakrabarty P, Sparks JS (2008). Phylogeny, taxonomy, and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae). *Cladistics* 24:625–641.

- Sparks JS, Smith WL (2004). Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics*. 20:501-517.
- Staiger D (2015). Shaping the Arabidopsis Transcriptome through Alternative Splicing. *Advances in Botany*. 13.
- Steeves G, Hansen D, Lamboj A (2010). Cichlids of Africa. Vol. I: Haplochromines. Grandã Publishing.
- Stabach PR, Morrow JS (2000). Identification and Characterization of β V Spectrin, a Mammalian Ortholog of Drosophila β H Spectrin. *The Journal of Biological Chemistry*. 275:21385–21395.
- Teixeira WG, Ferreira IA, Cabral-de-Mello DC, Mazzuchelli J, Valente GT, Poletto AB, Venere PC, Martins C (2009). Organization of repeated DNA elements in the genome of the cichlid fish *Cichlakelberi* and its contributions to the knowledge of fish genomes. *Cytogenet Genome Research*. 125:224-234.
- Teruel M, Cabrero J, Perfectti F, Camacho JP (2010). B chromosome ancestry revealed by histone genes in the migratory locust. *Chromosoma*. 119:217-225.
- The International Cichlid Genome Consortium (2006). Genetic basis of vertebrate diversity: the cichlid fish model.
- Traeger LL, Volkening JD, Moffett H, Gallant JR, Chen P, Novina CD, Junior GNP, Anand R, Wells GB, Pinch M, Güth R, Unguez GA, Albert JS, Zakon H, Sussman MR, Samanta MP (2015). Unique patterns of transcript and miRNA expression in the South American strong voltage electric eel (*Electrophorus electricus*). *BMC Genomics*. 16:243.
- Trewavas E (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum, Natural History, London.
- Tzur YB, Wilson KL, Gruenbaum Y (2006). SUN-domain proteins: 'Velcro' that links the nucleoskeleton to the cytoskeleton. *Nat Rev Mol Cell Biol*. 7.10: 782–8.
- Utsunomia R, Silva DMZdA, Ruiz-Ruano FJ, Araya-Jaime C, Pansonato-Alves JC, Scacchetti PC, et al. (2016) Uncovering the ancestry of B chromosomes in *Moenkhausia sanctaefilomenae* (Teleostei, Characidae). *PLoS One* 11: e0150573
- Uzer et al. (2015). Cell mechanosensitivity to extremely low magnitude signals is enabled by a LINCed nucleus. *Stem Cells*. 33. 6:2063–2076.
- Valente GT, Schneider CH, Gross MC, Feldberg E, Martins C (2009). Comparative cytogenetics of cichlid fishes through genomic in-situ hybridization (GISH) with emphasis on *Oreochromis niloticus*. *Chromosome Research*. 17:791–799.
- Valente GT, Conte MA, Fantinatti BEA, Cabral-de-melo DC, Carvalho R, Vicari MR, Kocher TD, Martins C (2014). Origin and Evolution of B Chromosomes in the Cichlid Fish *Astatotilapia latifasciata* Based on Integrated Genomic Analyses. *Molecular Biology Evolution*.
- Viana J, Pidsley R, Troakes C, Spiers H, Wong CCY, Al-Sarraj S, Craig I, Schalkwyk, Mill J (2014). Epigenomic and transcriptomic signatures of a Klinefelter syndrome (47,XXY) karyotype in the brain. *Epigenetics*. 4:587–599.
- Viñas J, Piferrer F (2008). Stage-Specific gene expression during fish spermatogenesis as determined by laser-capture microdissection and quantitative-PCR in sea bass (*Dicentrarchus labrax*) gonads. *Biology of Reproduction*. 79:738-747.
- Vogel H, Altincicek B, Glöckner G, Vilcinskis A (2011). A comprehensive transcriptome and immunegene repertoire of the lepidopteran model host *Galleria mellonella*. *BMC Genomics*. 12:308.
- Voltolin TA, Laudicina A, Senhorini JA, Bortolozzi J, Oliveira C, Foresti F, Porto-Foresti F (2010). Origin and molecular organization of supernumerary chromosomes of *Prochilodus lineatus* (Characiformes, Prochilodontidae) obtained by DNA probes. *Genetica*. 138:11-12.
- Wang D-S, Kobayashi T, Zhou Y, Paul-Prasanth B, Ijiri S, Sakai F, Okubo K, Morohashi K-I, Nagahama Y (2007). Foxl2 up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. *Mol Endocrinol*. 21:712–725.
- Wilson EB (1907). The supernumerary chromosomes of Hemiptera. *Science*. 26:870-871.
- Wu TD, Watanabe CK (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 21:1859–1875.
- Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, Hirai H, Hirai Y, Matsuda Y, Okada N (2011). B Chromosomes Have a Functional Effect on Female Sex Determination in Lake Victoria Cichlid Fishes. *Plos Genetics*. 7:8.
- Yu K, Yu Y, Tang X, Chen H, Xiao J, Su X (2016). Transcriptome analyses of insect cells to facilitate baculovirus-insect expression. *Protein Cell*. 7(5):373–382.
- Zhou Q, Zhu H, Huang Q, Xuan Z, Zhang G, Zhao L, Ding Y, Roy S, Vicoso B, Ruan J, Zhang Y, Zhao R, Mu B, Min J, Zhang Q, Li J, Luo Y, Liang Z, Ye C, Li R, Zhang X, Wang J, Wang W, Bachtrog D (2012). Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics*. 13:109.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 18:821–829.

Ziegler CG, Lamatsch DK, Steinlein C, Engel W, Scharl M, Schmid M (2003). The giant B chromosome of the cyprinid fish *Alburnus alburnus* harbours a retrotransposon-derived repetitive DNA sequence. *ChromosResearch* 11:23-35.