UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

(GENÉTICA)

Elucidação do efeito de cromossomos supernumerários na espécie de ciclídeo africano *Astatotilapia latifasciata*, com base na análise de expressão de micro-RNAs

BRUNO EVARISTO DE ALMEIDA FANTINATTI

Botucatu/SP

2015

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

(GENÉTICA)

Elucidação do efeito de cromossomos supernumerários na espécie de ciclídeo africano *Astatotilapia latifasciata*, com base na análise de expressão de micro-RNAs

BRUNO EVARISTO DE ALMEIDA FANTINATTI

Tese apresentada ao Instituto de Biociências de Botucatu, Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, como requisito para a obtenção do título de Doutor pelo Programa de Pós-Graduação em Ciências Biológicas (Genética).

Botucatu/SP

2015

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Marisa e Erivelto. As atividades desenvolvidas aqui em Botucatu não seriam possíveis sem o apoio deles.

Agradeço também ao meu irmão, Thiago, que sempre me estimulou a pensar de forma crítica, contribuindo indiretamente com minha formação científica.

Agradeço à Ana Silvia, minha noiva, que sempre esteve ao meu lado, lidando com meus momentos bons e ruins e me ajudando em tudo o que fosse possível.

Agradeço ao meu orientador Dr. Cesar Martins, por ter me concedido a oportunidade de cursar mestrado e doutorado.

Agradeço imensamente aos amigos do LGI, e de laboratórios vizinhos. Muito obrigado pelas discussões, churrascos, cervejas e risadas.

Agradeço ao Dr. Igor Jurisica e ao Dr. Mike Tsay, que me ajudaram durante os trabalhos desenvolvidos em Toronto.

Agradeço também às agências de fomento que estiveram envolvidas neste trabalho, fornecendo bolsas de estudo e suportes financeiros: CNPq (Processo: 151697/2011-6), CAPES-PDSE (Processo: 3127-14-1) e FAPESP (Processo 2013/04533-3).

A Rafael Henrique Correa Batista

RESUMO

Cromossomos B (ou simplesmente Bs) são elementos cromossômicos adicionais presentes em uma grande variedade de espécies de eucariotos. Apesar de muitos trabalhos reportarem a distribuição de cromossomos B em várias espécies, uma teoria definitiva sobre a origem e manutenção deste elemento ainda não foi elaborada, mostrando que mais estudos são necessários. Tendo em vista que as técnicas de citogenética clássica e molecular não são capazes de responder muitas questões em relação aos cromossomos B, é importante procurar respostas a nível molecular. Micro-RNAs são pequenas moléculas de RNA com tamanho variando de 17 a 25 nucleotídeos, e que agem como controladores pós-transcricionais se ligando a RNAs mensageiros, e silenciando sua atividade, moldando assim o panorama funcional da célula. Entre espécies de peixes ciclídeos, *Astatotilapia latifasciata* possui de um a dois grandes cromossomos B heterocromáticos. Considerando micro-RNAs como reguladores de expressão gênica, efeitos trazidos pela presença de cromossomos B foram analisados baseando-se na espressão diferencial de micro-RNAs utilizando técnicas de sequenciamento de próxima geração, de modo a obter mais informações em relação aos efeitos que cromossomos B causam no genoma hospedeiro. Os dados obtidos mostram que a presença do cromossomo B na espécie *Astatotilapia latifasciata* apresenta um certo impacto no perfil de expressão de miRNAs, resultando no controle de um grande número de genes, com destaque para genes envolvidos com processos de divisão celular e desenvolvimento. Também foi observado que tais efeitos ocorrem de uma forma dependente do tecido analisado. Deste modo, nossos dados trazem forte evidência de que os cromossomos B não são elementos inativos como imaginado antes, mas de fato podem interferir in processos biológicos muito importantes.

ABSTRACT

B chromosomes (Bs) are additional chromosomal elements found in a high diversity of eukaryote groups including yeast, plants and animals. Despite thousands of reports describing the distribution of this chromosome type, a comprehensive theory for its origin, maintenance and evolution has not emerged, and new studies are necessary. Since most cytogenetic techniques does not answer major questions surrounding Bs, it is important to look at the molecular level in order to go further. MicroRNAs are small single nucleotide RNA molecules, ranging from 17 to 25 nucleotides that act as a post-transcriptional controller, binding and silencing mRNAs, consequently shaping the genomic profile of expressed genes. Among cichlid fish species, *Astatotilapia latifasciata* might carry 1-2 large heterochromatic B chromosomes. Considering miRNAs as major players in gene regulation we analyzed the effects of Bs on host individuals through microRNA analysis based on next-generation sequencing data in order to better understand B chromosome biology. The obtained data shows that the presence of the B chromosome in the species under study presents a huge impact on the miRNA expression profile, resulting in a influence over a high number of genes, where many of them are related to cell cycle control. It was also observed that B chromosome effects take place in a tissue-dependent fashion. In this way, our data give strong evidences that B chromosomes are no longer inactive elements as expected before, but indeed can interfere in very important biological process of cells and organisms.

RESUMO

ABSTRACT

1. REVISÃO DE LITERATURA 8

1.1. Cichlidae – Modelo de estudo 8

1.2. Cromossomos supernumerários 9

1.3. RNAs não codificantes com ênfase em Micro-RNAs 15

2. OBJETIVOS 22

2.1. Objetivo geral e justificativa 22

2.2. Objetivos específicos 22

3. MATERIAL E MÉTODOS 24

3.1. Amostragem 24

3.2. Extração de DNA genômico 24

3.3. Genotipagem 24

3.3.1. Genotipagem qualitativa 25

3.3.2. Genotipagem quantitativa 27

3.4. Coleta dos tecidos para análises de expressão 29

3.5. Extração de RNA total e seqüenciamento 30

3.6. Processamento e análise dos dados 31

3.7. Redes de interação 34

4. RESULTADOS E DISCUSSÃO 36

4.1. Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish *Astatotilapia latifasciata* as model 37

4.2. B chromosomes effects based on large scale bioinformatics analysis of micro-RNA in the African cichlid fish *Astatotilapia latifasciata* 56

5. CONCLUSÕES 85

6. REFERÊNCIAS BIBLIOGRÁFICAS 86

1. REVISÃO DE LITERATURA

1.1. Cichlidae – Modelo de estudo

Inserida na ordem Perciformes - considerada a maior ordem existente de vertebrados - encontram-se os peixes da família Cichlidae, a qual possui mais de 3.000 espécies (Kocher, 2004). Tal família é dividida em quatro sub-famílias, sendo elas Etroplinae, Ptychochrominae, Pseudocrenilabrinae e Cichlinae. A espécie *Astatotilapia latifasciata*, encontra-se inserida na sub-família Pseudocrenilabrinae (Smith et al. 2008) originária do continente africano. Outras sub-famílias (Cichlinae, Etroplinae e Ptychochrominae) são características da região Neotropical, Índia e Madagascar, e Madagascar, respectivamente.

A grande diversidade de ciclídeos presente nos grandes lagos africanos tem despertado interesse em pesquisadores da áreade biologia evolutiva, pois aproximadamente 2.000 espécies sofreram um rápido processo dediversificação nos últimos 10 milhões de anos (Kocher, 2004), o que representa um período evolutivo relativamente curto, evidenciando a alta capacidade de diferenciação deste grupo.

A capacidade de suportar grandes variações ambientais é também observada nos ciclídeos. Por exemplo, a espécie *Orechromis alcalicus grahami* suporta viver em águas extremamente salinas (com pH de até 10,5) e/ou extremamente quentes como em lagoas próximas ao lago Magadi (Kênia) (Maina, 2000). A alta plasticidade fenotípica presente neste grupo (Pfennig and Ehrenreich, 2014; Schneider et al. 2014) pode ser considerada responsável pela considerável radiação adaptativa de membros desta família nos variados ambientes dos lagos africanos.

Os peixes desta família também são bastante apreciados para alimentação, pesca esportiva e aquariofilia. Em tal contexto, a espécie *Oreochromis niloticus* (comumente denominada “Tilápia do Nilo”) é vista como a mais apreciada para consumo, por conta de seu rápido desenvolvimento, facilidade de cultivo, elevada capacidade de adaptação e carne de sabor agradável. As espécies do gênero *Cichla* (comumente denominadas “Tucunarés”) são apreciadas para consumo e pesca esportiva. Adicionalmente, outras espécies, por apresentarem padrões de coloração bastante variados e marcantes, são apreciadas para aquariofilia.

Considerando estudos cromossômicos, grande parte das informações da família Cichlidae concentra-se basicamente na espécie *Oreochromis niloticus.* Para as demais espécies, tais informações são em sua grande maioria restritas a análises de citogenética clássica. Apesar disso, alguns estudos mais atuais têm trazido informações acerca da composição e organização genômica de espécies desta família (Mazzuchelli e Martins2009; Teixeira et al.2009; Valente et al. 2011; Mazzuchelli et al. 2012).

Atualmente também tem sido utilizadas tecnologias de sequenciamento de próxima geração para obtenção dados em larga escala, com finalidade de buscar informações mais detalhadas em relação a problemas como (i) rápida radiação adaptativa dos ciclídeos Africanos (Brawand et al. 2014), (ii) análises mais aprofundadas em conteúdo e evolução de cromossomos B (Valente et al. 2014) e (iii) plasticidade fenotípica (Schneider et al. 2014).

1.2. Cromossomos supernumerários

Cromossomos supernumerários, também conhecidos como cromossomos B ou denominados simplesmente “Bs”, são cromossomos extras aos cromossomos autossômicos (denominados complemento A), que se encontram presentes em 10-15% das espécies eucariotas (Camacho et al. 2000; Camacho, 2005). Atualmente este número pode ser muito maior, uma vez que espécies animais e vegetais são constantemente descobertas como portadoras de Bs, como observado em Yoshida et al. (2011), onde mais de 10 novas espécies de ciclídeos africanos foram detectadas como possuidoras de Bs.

Cromossomos B têm sido encontrados em diversos grupos, e representam um problema bastante interessante do ponto de vista evolutivo e, ultimamente, também do ponto de vista funcional. Estes cromossomos possuem um enorme potencial de nos ensinar sobre evolução de genomas nos mais variados grupos em que são encontrados. Algumas espécies de ciclídeos foram detectadas como portadoras de chromossomos B. Tais espécies estão distribuídas tanto pela américa do sul, quanto pelo continente africano (Feldberg et al. 2003, Feldberg et al. 2004, Poletto et al. 2010a, Fantinatti et al. 2011, Yoshida et al. 2011). Dentre os ciclídeos com cromossomos B, *Astatotilapia latifasciata* (Figura 1), pode apresentar um ou dois grandes cromossomos B em seu cariótipo (Figura 2).

Figura 1: Macho adulto da espécie *Astatotilapia latifasciata.*

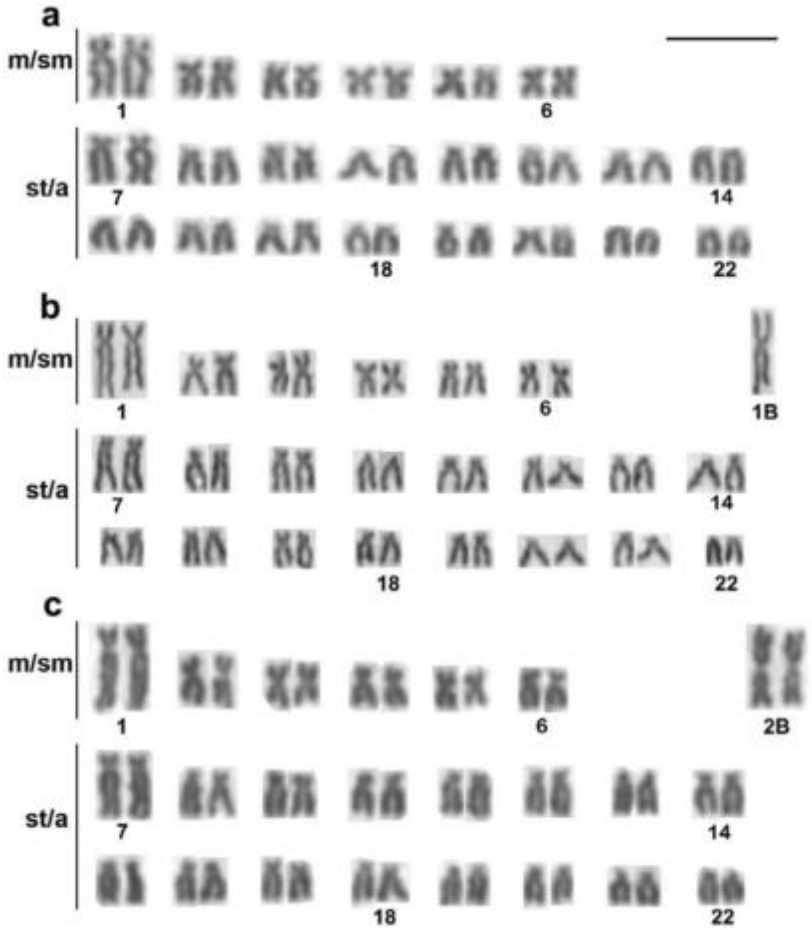


Figura 2: Cariótipos da espécie *Astatotilapia latifasciata* evidenciando (a) condição normal, (b) presença de 1 ou (c) 2 cromossomos B. Barra indica 5 µm. (Poletto et al. 2010a, com autorização dos autores).

Cromossomos B podem apresentar “drive” e, por conta disto, são denominados “elementos parasitas”, pois tal peculiaridade faz com que permaneçam no genoma (Camacho et al. 2000). Sendo assim, normalmente o indivíduo que porta cromossomos B tem seu genoma denominado “hospedeiro”.

Cromossomos B podem apresentar muitas variações. Tais variações podem ocorrer em se tratando de tamanho, número ou morfologia, que podem ser encontradas dentro de uma mesma espécie, ou até mesmo dentro de um mesmo indivíduo, como relatado por diversos estudos (Kasahara, 1982; Trifonov et al. 2002; Szczerbal e Switonski, 2003; Moreira-Filho et al. 2004). Neste sentido, cromossomos B podem ser tão pequenos quanto um simples ponto, sendo também chamados de ”*dot-like*” (Trifonov et al. 2002), ou podem ser encontrados em formas tão grandes quanto o maior par do complemento A (Maistro et al. 1992; Poletto et al. 2010). Também podem chegar a ser considerados gigantes genômicos, atingindo assim tamanhos superiores aos maiores pares de cromossomos de todo o cariótipo como visto na espécie de peixe *Alburnus alburnus*, detendo até o momento a posição de maior cromossomo B encontrado para vertebrados (Ziegler et al. 2003; Schmid et al. 2006).

Cromossomos B apresentam geralmente uma morfologia distinta dos outros cromossomos do complemento A. Tais variações também estão associadas a conteúdo genômico, sendo que a grande maioria dos cromossomos B descritos mostram-se heterocromáticos e repletos de elementos repetitivos, salvo algumas exceções onde Bs eucromáticos foram detectados, como visto na planta *Scilla vvedenskyi* (Greilhuber et al. 1976), no

peixe *Characidium cf. zebra* (Venere et al. 1999) e no roedor *Apodemus flavicollis* (Tanic et al. 2005).

Cromossomos B são conhecidos por não apresentarem recombinações entre si quando presentes em grande número, nem com outros cromossomos do complemento A. Apesar disso, existem evidências de que recombinações ocorram em cromossomos B da espécie de mamífero *Vulpes vulpes* (Basheva et al. 2010).

A presença deste elemento no genoma hospedeiro pode trazer inúmeros resultados ao *fitness* do hospedeiro. Existem casos onde a presença deste cromossomo pode afetar beneficamente o hospedeiro (Zima et al. 2003) por meio de efeitos heteróticos. Em outros casos, tais efeitos podem manifestar-se de forma deletéria (Jones et al. 2008) fazendo com que o hospedeiro apresente problemas quando em presença de Bs. Teruel et al. (2011) relatam um caso onde diferentes linhagens de cromossomos B causam um decréscimo na expressão da proteína *Hsp70*no gafanhoto *Eyprepocnemis plorans*. Por se tratar de uma proteína de resposta a estresse, reflete a direta relação que cromossomos B possuem com o *fitness* do hospedeiro. Esta espécie de gafanhoto também apresenta cromossomos B com diferentes níveis de parasitismo, resultando em variados níveis de quiasmas observados (Camacho et al. 2002; Camacho, 2005) em resposta ao efeito de drive apresentado pelo cromossomo B.

Cromossomos B possuem uma ampla variedade de formas de surgimento. A análise de sequências compartilhadas entre o complemento A e B é uma forma de obter informações sobre sua origem (Beukeboom, 1994). Se sequências são compartilhadas entre um dado cromossomo B e seu

complemento A, pode tratar-se de uma forma intra-específica de origem, representada aqui por aquela quando o cromossomo B é originado por segmentos do próprio genoma em que se encontra (Sapre e Deshpande, 1987; Schartl et al. 1995; McAllister e Werren, 1997). Origens inter-específicas também são possíveis, e são geralmente notadas quando são encontradas no cromossomo B sequências que estão ausentes no complemento A da espécie hospedeira, porém presentes no genoma de uma espécie próxima (McAllister e Werren, 1997). Ultimamente, com o uso de tecnologias de seqüenciamento de próxima geração, tem sido detectado em cromossomos B até mesmo sequências oriundas de organelas citoplasmáticas (Martis et al. 2012).

O sequenciamento completo dos genomas de várias espécies de ciclídeos (Brawand et al. 2014), associado a novas abordagens em bioinformática, compreendem um conjunto de ferramentas de extrema importância para a contribuição da elucidação dos mecanismos que norteiam a origem e evolução de cromossomos supernumerários. Isto não se aplica somente ao grupo de peixes, mas em um panorama mais amplo, contribuindo também para um melhor entendimento dos processos evolutivos de uma forma geral.

Embora sequenciamentos genômicos vem contribuindo massivamente para um conhecimento mais aprofundado dos aspectos evolutivos, é necessário que haja uma maior atenção em relação à análise prévia da espécie estudada. Isto deve ser realizado com base em cariotipagem prévia da espécie modelo do estudo, antes que seja executado o sequenciamento completo de seu genoma. A presença de cromossomos B em um genoma sequenciado, pode resultar em maiores dificuldades durante o processo de montagem do

genoma, ou até mesmo ocasionar erros na montagem do genoma. Isto ocorre em decorrência da grande quantidade de genes degradados e/ou duplicados que possam estar presentes nos cromossomos B (Valente et al. 2014), e até mesmo por outros elementos que são comumente encontrados em grande número nestes cromossomos.

Estudos mais atuais trouxeram informações mais detalhadas em relação à composição de tais cromossomos. Foi observado em *Mycosphaerella graminicola* que cromossomos extras possuem uma alta quantidade de genes, sendo vários relacionados a processos de divisão celular (Goodwin et al. 2011). A presença de genes em cromossomos B também foi observada para *Astatotilapia latifasciata* onde, igualmente para *M. graminicola*, grande quantidade destes genes é associada a processos de divisão celular (Valente et al. 2014). Em relação a conteúdo, foi observado em *Secale cereale* que além de segmentos provenientes de cromossomos do complemento A, também encontram-se insertos de DNA organelar (Martis et al. 2012). Funcionalmente, foi observado também que em *Secale cereale*, fragmentos de pseudo-genes são ativos, e apresentam comportamento variável dependente do tecido analisado (Banaei-Moghaddam et al. 2013). Análises em larga escala tendo como base miRNAs mostram-se eficazes para obtenção de informações relacionadas a cromossomos B, uma vez que a presença de tais cromossomos resultam em efeitos nos organismos hospedeiros.

1.3. RNAs não codificantes com ênfase em Micro-RNAs

Non-coding RNAs (ncRNAs) representam uma classe de RNAs que não resultarão em proteínas. Inúmeras classes de ncRNAs são conhecidas, dentre

elas os RNAs transportadores (tRNAs) responsáveis pelo transporte de aminoácidos durante a síntese de proteínas, RNAs ribossomais (rRNAs) responsáveis pela formação dos ribossomos, pequenos RNAs nucleares (snRNAs) e pequenos RNAs nucleolares (snoRNAs). ncRNAs regulatórios são divididos em classes de acordo com seu comprimento. ncRNAs longos são os que apresentam tamanhos maiores que 200 pb. Pequenos ncRNAs apresentam tamanhos menores que 200 pb. Dentro da classe de pequenos ncRNAs ainda existem os micro-RNAs (miRNAS) e PIWI-interacting RNAs (piRNAs). MiRNAs são pequenas moléculas de RNA contendo geralmente de 18 a 22 nucleotídeos (Bartel, 2004; Chen et al, 2014), que agem como reguladores pós-transcricionais, envolvendo-se na degradação e/ou silenciamento de RNAs mensageiros (mRNAs) de acordo com um certo grau de complementariedade. Tal controle pós-transcricional se dá pela incorporação do miRNA a um complexo protéico denominado RISC (*RNA Induced Silencing Complex*) onde o miRNA serve de guia para que regiões complementares sejam encontradas e resulta na degradação de um mRNA ou nainibição de sua tradução, o que representa um processo bastante complexo, envolvendo participação de inúmeras proteínas (Jonas and Izaurralde 2015). Os primeiros miRNAs descobertos foram descritos em *Caenorhabditis elegans* (Lewis et al. 2003), tendo sido denominados *lin-4* e *let-7*. Posteriormente miRNAs foram sendo descobertos também em outras espécies. Em peixes, miRNAs foram primeiramente descritos e caracterizados para *Danio rerio* (espécie comumente denominada zebrafish – tida como um importante modelo biológico), onde verificou-se a importância destes elementos para várias funções (Giraldez et al. 2005).

Os níveis de expressão de determinados miRNAs variam de acordo com o tempo e tecido. Por exemplo, durante a formação dos olhos em zebrafish, miR-181a e b são expressos expecificamente no tecido da retina (Kapsimali et al., 2007). MiRNAs também são bastante importantes durante os processos de diferenciação sexual. Em *Hippoglossus hippoglossus*, por volta de 17 miRNAs foram detectados como altamente expressos em ovários. Em testículos, 14 miRNAs foram detectados como diferencialmente expressos, ambos os grupos apresentando um *fold-change* de ao menos 2 (Bizuayehu et al. 2012).

MiRNAs são transcritos a partir de diversas regiões do genoma. Estas regiões genômicas transcricionais dos miRNAs podem estar organizadas de forma individual, ou ainda em clusters, podendo localizar-se em regiões de éxons ou íntrons. Inicialmente a trasncrição é realizada pela RNA polimerase II ou III resultando nos pri-miRNAs. A enzima *Drosha* então processa estes elementos em fragmentos de aproximadamente 70 nucleotídeos de comprimento, ainda em formatos de *hairpins*, que são então levados ao citoplasma por intermédio da exportina-5. A enzima *Dicer* então realiza o último processamento, resultando em sequências de 18-22 nucleotídeos, caracterizando seu estado maduro (Figura 3). Após atingir seu estado estrutural maduro, tais miRNAs adquirem funcionalidades que englobam desde o controle do desenvolvimento dos inúmeros tecidos durante os processos de diferenciação celular, até a manutenção de funções básicas, como por exemplo, regeneração e manutenção de tecidos e respostas a estresse.

A funcionalidade dos micro-RNAs se dá por conta da inserção destes elementos em grandes complexos protéicos denominados RISC (*RNA Induced Silencing Complex*) (Bartel et al. 2004). Após sua inserção em tais complexos,

os miRNAs servem de guia para que todo o complexo de ligue em determinadas regiões dos RNAs mensageiros, causando assim, a completa degradação ou quebra dos segmentos de mRNA, resultando por final na diminuição ou até mesmo na inibição da expressão do gene. A região dos mRNAs que normalmente é alvo de interações entre miRNAs são as 3'-UTRs, embora existam casos em que outras regiões são sujeitas a interações com miRNAs. Isto é observado no vírus HIV-1, onde miR-146a tem como alvo uma região da proteína do capsídeo do vírus (Chen et al 2014), o que resulta em uma diminuição da produção viral.

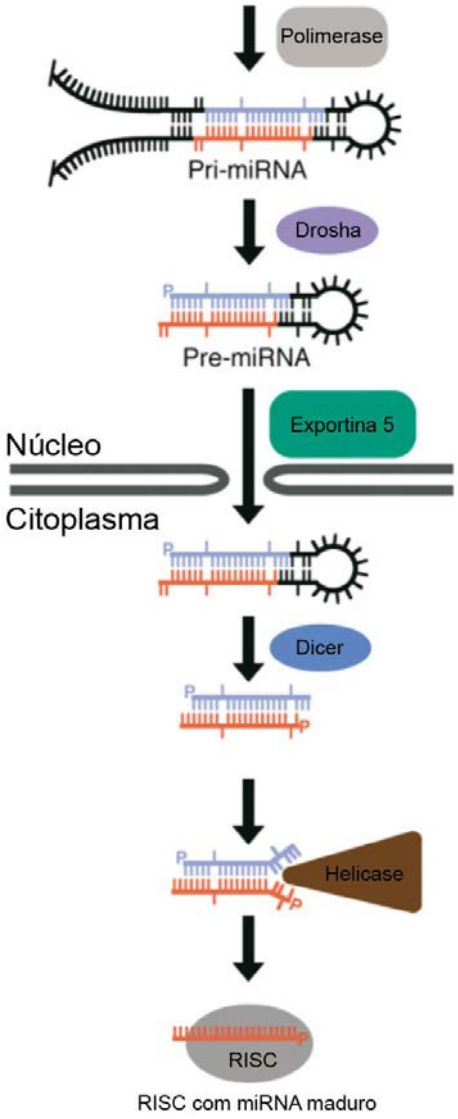


Figura 3: Esquema da biogênese dos microRNAs em metazoários. Adaptado de Bartel 2004.

Pareamentos de miRNAs com seus alvos em espécies animais, não obedecem simplesmente à complementaridade entre ambas as partes (pareamento denominado "*watson-and-crick*") que é comumente observada em espécies vegetais. Tais pareamentos podem ocorrer de várias formas diferentes, o que dificulta a criação dos métodos de predição de alvos. Tais formas de pareamento se resumem basicamente a três tipos: canônico,

suplementar e compensatório. O tipo de pareamento canônico, ainda se divide em outras formas diferentes, dependendo da presença de uma adenina na região *seed* e também do tamanho da referida região (Figura 4a). A região *seed* é uma região localizada na extremidade 5' do miRNA, e é caracterizada por uma região onde um pareamento perfeito ocorre na grande maioria das vezes (Witkos et al. 2011). Para o pareamento suplementar, a região *seed* aparece como um pareamento perfeito, adicionado de três ou quatro outros pareamentos perfeitos localizados na outra extremidade do miRNA, geralmente nas posições 13-16 (Figura 4b). Para o pareamento compensatório, um *mismatch* aparece na região *seed*, seguido de pareamentos perfeitos nas posições 13-16 do miRNA (Figura 4c). Tais variações nas formas como um determinado miRNA se liga a uma região alvo, faz com que tal miRNA possua uma ampla quantidade de possibilidades, resultando em um elevado número de genes em que possa atuar.

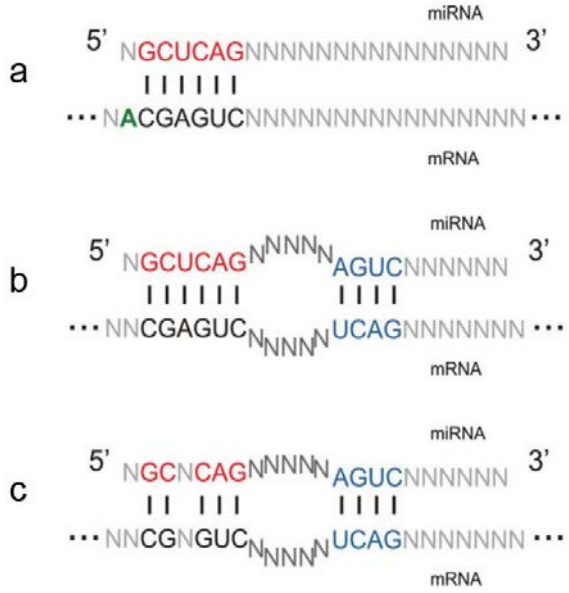


Figura 4: Exemplos de diferentes formas de interação miRNA:mRNA. Região *seed* é apresentada em vermelho. Pareamentos suplementares e compensatórios são apresentados em azul. Adaptado de Witkos et al. 2011.

Várias ferramentas computacionais foram produzidas para realizar predição de interações entre miRNAs e mRNAs. Porém, pelo fato de que não há um conhecimento muito bem estabelecido em relação às leis que governam os pareamentos miRNA-mRNA em grupos animais, desenvolvedores de *softwares* acabam por basear seus algoritmos em diferentes métodos para a análise de possíveis alvos, o que normalmente gera um certo número de falsos positivos, levando a conclusões errôneas, caso a atenção devida não seja aplicada. A alta discrepância entre a forma como miRNAs pareiam com seus alvos é tão grande, que *softwares* específicos para espécies vegetais e animais foram produzidos (Witkos et al. 2011). Tal fato torna necessário que dados obtidos por tais preditores passem por um processo de validação experimental, o que contribui diretamente com a melhoria de tais algoritmos, diminuindo por consequência o número de falsos positivos preditos.

Considerando de que miRNAs são elementos que desempenham importantes funções referentes a controle de processos biológicos, análises de tais controladores podem representar ferramentas eficazes para obtenção de informações relacionada a atividade de cromossomos B.

2. OBJETIVOS

2.1. Objetivo geral e justificativa

Tendo em vista que os efeitos da presença de cromossomos B no genoma hospedeiro encontra-se ainda para ser elucidado, o presente trabalho teve por objetivo análises dos efeitos causados pela presença destes cromossomos tendo como modelo a espécie de ciclídeo africano *Astatotilapia latifasciata*, portadora de um ou dois grandes cromossomos B.

Tais estudos foram executados com ênfase no perfil molecular, baseando-se em análise *in silico* em larga escala de expressão de miRNAs em caráter comparativo entre amostras de tecidos de animais 0B e 1B. Estas análises foramr ealizadas em duplicatas/triplicatas biológicas, considerando-se os dois sexos e quatro diferentes tecidos (cérebro, músculo, gônadas e brânquias), totalizando 40 bibliotecas sequenciadas e analisadas.

Tal estudo mostra-se promissor para o conhecimento evolutivo e funcional de genomas de ciclídeos, com ênfase aos efeitos causados pela presença de cromossomos B.

2.2. Objetivos específicos

Desenvolvimento de marcadores moleculares, baseados em PCR e qPCR, para genotipagem de indivíduos quanto à presença de cromossomos B em caráter qualitativo e quantitativo;

Análises *in silico* de perfis de expressão de miRNAs conhecidos;

Predição e análises *in silico* de perfis de expressão de novos miRNAs;

Geração de redes de interação miRNA-mRNA e identificação de genes alvos.

Identificação de processos biológicos relacionados à presença/ausência de Bs.

3. MATERIAL E MÉTODOS

3.1. Amostragem

Foi utilizada a espécie de ciclídeo africano *Astatotilapia latifasciata* amplamente difundida na aquariofilia, como modelo de estudo para conduzir a presente proposta de pesquisa. Esta espécie é natural dos lagos Kyoga e Nawampasa (lagos satélites do lago Vitória, na África). Tais animais foram obtidos a partir de lojas de aquariofilia da cidade de Botucatu/SP que oferecem ciclídeos de pequeno porte para a comunidade aquariofilista da região.

Os animais foram transportados ao biotério do Laboratório Genômica Integrativa, localizado no Instituto de Biociências da UNESP, Campus de Botucatu, e mantidos em ambiente controlado em se tratando de temperatura, alimentação e luminosidade. Os animais foram mantidos em tais condições durante todo o processo de genotipagem, até o momento de coleta de tecidos para posteriores análises.

3.2. Extração de DNA genômico

As extrações de DNA genômico utilizando tecidos das nadadeiras foram realizadas por meio do *kit DNeasy* (Qiagen), seguindo as especificações do fabricante. Tal extração objetivou a obtenção de material para realização das genotipagens quanto à presença/ausência de cromossomos B, bem como seu número, tendo em vista que até dois cromossomos B foram encontrados nesta espécie (Poletto et al. 2010; Fantinatti et al. 2011).

3.3. Genotipagem

3.3.1. Genotipagem qualitativa

Para a obtenção de preparações citogenéticas, utiliza-se normalmente um processo de infecção por fungos através de injeção intraperitonial, causando assim um aumento na taxa de divisão celular, seguido da coleta do tecido dos rins e processos específicos para obtenção de preparações cromossômicas (Bertollo et al. 1978). Tal procedimento se torna inviável, de acordo com os passos posteriores, por se tratar de extração e análise de expressão de micro-RNAs. Frente a isto, tornou-se necessária a criação de um método capaz de detectar a presença de cromossomos supernumerários sem que houvesse a necessidade de sacrifício do animal estudado, envolvendo tratamento prévio que tivesse efeito na fisiologia do organismo levando a possíveis alterações nos perfis de expressão gênica.

Para o marcador qualitativo, um cromossomo B foi microdissectado utilizando um microscópio invertido (Olympus IX51) com ajuda de um micromanipulador (Narishige, Tokyo, Japan). O material microdissectado foi então transferido para tubos onde uma amplificação utilizando o kit Genome Plex® Single Cell - Whole Genome Amplification - WGA4 (Sigma-Aldrich, St. Louis, MO, USA) foi realizada, seguindo as instruções do fabricante. O produto da amplificação foi purificado usando o kit GenEluteTM PCR Clean-Up (Sigma-Aldrich). Os produtos purificados foram utilizados para sequenciamento de próxima geranção ou reamplificados para obtenção de sondas a serem usadas em experimentos de FISH para validação dos fragmentos (Figura 3).

Para a reamplificação, uma reação utilizando o *kit* WGA3 (*Genome Plex® Whole Genome Amplification* – WGA3, Sigma-Aldrich) o

nucleotideodigoxigenina 11-dUTP foi incorpodado aos fragmentos de modo a marcar a sonda.

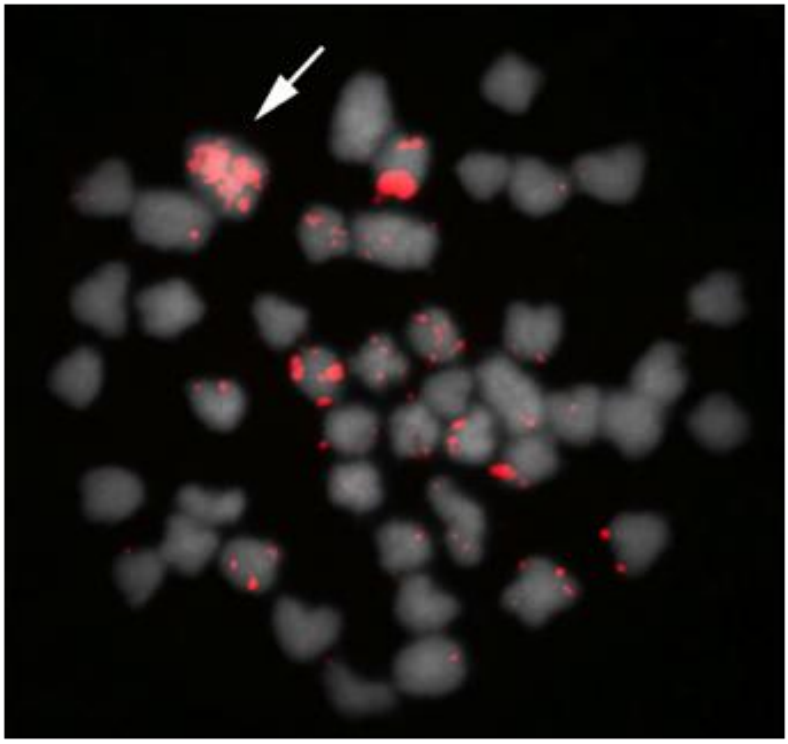


Figura 3: FISH evidenciando sonda do cromossomo B microdissectado hibridada ao cromossomo B.

DNA gerado por meio da amplificação total do cromossomo B microdissectado foi sequenciado utilizando a plataforma 454 GS Jr. (Roche, Basel, Switzerland) e os fragmentos do sequenciamento foram montados utilizando o software *Newbler*. Os contigs montados foram analisados por meio do algorítmo BLAST, utilizando a coleção de nucleotídeos do NCBI. Contigs apresentando *hits* com species de peixes foram recuperados e então uma busca por regiões específicas do cromossomo B foi realizada. Primers para PCR convencional foram desenhados sobre regiões apresentando mutações específicas do cromossomo B juntamente com primers reversos, e controles positivos que foram desenhados em regiões conservadas. Tais primers foram

desenhados para serem aplicados em uma reação de PCR multiplex, de modo que somente um fragmento (fragmento controle) fosse amplificado em DNA sem cromossomos B, e dois fragmentos (fragmento controle + fragmento específico) fossem amplificados em amostras contendo cromossomos B.

A melhor temperatura de anelamento foi testada por meio de uma termociclagem com um gradiente de temperatura de anelamento variando de 45 a 65°C. As reações foram preparadas da seguinte forma: 1 U de *Platinum Taq DNA Polymerase*, 1.5 mM de $MgCl_2$, tampão RXN a 1 X, 0.32 mM de dNTP, 0.2 µM de cada primer (totalizando 0.6µM de primers) e água ultrapura para completar 25 µl. Ciclagens de temperatura foram realizadas da seguinte forma: 5min a 95°C, 34 x (1min a 95ºC, 30seg a 50ºC, 45seg a 72ºC) e 5min a 72°C. O material foi então analisado em gel de agarose 1%.

3.3.2. Genotipagem quantitativa

As amostras previamente detectadas como portadoras de cromossomos B (B+) foram avaliadas quanto à quantidade de tais cromossomos. Tais análises se deram através de um marcador molecular com base em PCR quantitativo (qPCR) capaz de detectar diferentes números de cópias de determinado segmento entre várias amostras.

DNA genômico de amostras 0B e 2B, genotipadas citogeneticamente, foram submetidas a sequenciamento por meio da plataforma Illumina. Tal material foi fragmentado em porções medindo por volta de 500 pb usando um Ultrasonicador S220 (Covaris Inc., Woburn, MA, USA). As bibliotecas separadas foram preparadas utilizando o kit TruSeq DNA sample preparation kit ver.2 rev.C (Illumina Inc., San Diego, California, USA). Sequenciamentos em

modo "*paired-end*" foram executados em *lanes* separados para cada amostra por meio da plataforma Illumina HiSeq 1000.

Os *reads* resultantes do sequenciamento de ambas as amostras 0B e 2B foram separadamente submetidos a alinhamento contra o genoma de referência do ciclídeo *Metriaclima zebra,* disponível no banco de dados Bouillabase (http://cichlid.umd.edu/cichlidlabs/kocherlab/bouillabase.html), utilizando o software *Bowtie2* (Langmead e Salzberg, 2012). Os resultados dos alinhamentos foram submetidos a uma análise de taxas de cobertura entre os dois grupos de dados, de acordo com Valente et al. 2014, e as regiões com altas taxas de cobertura foram selecionadas para realização de qPCR.

As reações de qPCR foram realizadas utilizando DNAs conhecidos de amostras previamente genotipadas citogeneticamente. Foram utilizadas três amostras diferentes para cada genótipo (3x0B, 3x1B e 3x2B). Estas mesmas amostras foram utilizadas como controle nas reações de qPCR posteriores.

Os experimentos de qPCR foram realizados da seguinte forma: Primeiramente as amostras foram diluídas para uma concentração padrão de 30ng/µl, e tal concentração foi confirmada por meio de espectrofotometria por meio do equipamento Nanovue (GE). O mix de reação foi preparado usando *GoTaq* qPCR *Master Mix* (Promega) 1X, adicionado de 0,16 µM de cada primer, 600ng de DNA e água ultrapura para completar 75 µl. O gene phosphoribosyltransferase (HPRT) amplamente utilizado em experimentos de qPCR, foi utilizado como gene controle para as análises dos dados que foram obtidos por meio de "*gene dose ratio*" (GDR) (Bel et al. 2011).

Após os primeiros testes utilizando amostras conhecidas, novas amostras foram analisadas utilizando somente uma amostra controle para cada

genótipo. Os animais foram então separados em três grupos distintos, de acordo com os resultados das genotipagens (0B, 1B e 2B). Tais análises também foram realizadas em diferentes tecidos de um mesmo indivíduo, com a finalidade de observar uma possível presença de mosaicismo entre os animais selecionados, uma vez que cromossomos B podem apresentar tal característica (Perfectti e Werren, 2001).

3.4. Coleta dos tecidos para análises de expressão

Animais 0B e 1B previamente genotipados foram selecionados para proceder com a coleta de tecidos voltada à extração de RNA total. Quatro diferentes tecidos foram selecionados, sendo eles: cérebro, brânquias, músculo e gônadas.

Inicialmente, os animais foram anestesiados por meio de gelo. Posteriormente, posicionados sobre uma placa acrílica previamente resfriada em nitrogênio líquido, para que os tecidos se mantivessem resfriados durante todo o procedimento de coleta. Tal placa foi embalada em papel alumínio previamente autoclavado e tratado em sequência, com etanol 70%, água sanitária 10% e RNase ZAP. O papel alumínio foi substituído a cada espécime processado. Cérebro, músculo, gônadas e brânquias foram coletados com auxílio de bisturis, tesouras e pinças individuais previamente autoclavadas e esterilizadas em sequência, com etanol 70%, água sanitária 10% e RNase ZAP. Bisturis, tesouras e pinças também foram substituídas a cada espécime processado.

Logo após a coleta dos diferentes tecidos, os mesmos foram imediatamente acondicionados em criotubos, congelados em nitrogênio líquido e mantidos estocados a -80°C para posterior extração de RNA total.

3.5. Extração de RNA total e seqüenciamento

Os três diferentes tecidos previamente coletados, congelados em nitrogênio líquido e estocados a -80° foram submetidos à extração de RNA total utilizando TRIZOL (Invitrogen), seguindo as especificações do fabricante.

Todas as amostras extraídas foram submetidas à análise de qualidade por meio do equipamento *Nanovue* (GE), e de integridade por meio do equipamento *Bioanalyzer* (Agilent Technologies). Amostras apresentando um valor de *RIN* (*RNA Integrity Number*) igual ou superior a 8 foram selecionadas para realização dos sequenciamentos.

Todo o processo de preparação de bibliotecas e sequenciamento foi realizado via empresa *LC Sciences* (Houston, TX, EUA) com base na utilização do kit TruSeq Small RNA Sample Prep. Os procedimentos consistiram em ligação de adaptadores nas extremidades 3' e 5' dos fragmentos de RNA total, seguido de transcrição reversa. Uma purificação através de gel de poliacrilamida foi realizada, onde fragmentos de tamanhos condizentes com miRNAs foram coletados. O material purificado foi submetido a sequenciamento com base na plataforma *next-generation* Illumina - HiSeq. Os dados resultantes foram disponibilizados pela empresa através de um *link* FTP, e transferidos aos computadores do Laboratório Genômica Integrativa, de modo a prosseguir com os processamentos.

3.6. Processamento e análise dos dados

Todos os dados foram processados utilizando uma estrutura computacional de alto desempenho, localizada no Laboratório Genômica Intergrativa, baseada em sistemas operacionais *GNU/Linux* em arquitetura de *64bits* com 64, 128 ou 1024 *gigabytes* de memória e 24, 32 ou 128 núcleos de processamento, respectivamente. Os ambientes de processamento de dados se deram através das distribuições *Linux Mint 14* e *17* (http://www.linuxmint.com), e *SuSE Linux EnterpriseServer 11*, também conhecida como "SLES11" (https://www.suse.com).

Os dados recebidos foram primeiramente analisados quanto à qualidade de leitura das bases. Para isso, foi empregado o *software FastQC* (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) versão 0.10.1. Após a análise dos dados de qualidade do sequenciamento, o material passou por um processo de remoção dos adaptadores utilizados nos procedimentos de sequenciamento, por meio do *software Fastx-toolkit* (http://hannonlab.cshl.edu/fastx_toolkit/index.html). *Reads* que não continham adaptadores foram mantidos nas bibliotecas. *Reads* que, após a remoção dos adaptadores, encontraram-se com menos de 17 nucleotídeos de comprimento, foram excluídos das bibliotecas.

Após a remoção dos adaptadores, as bibliotecas foram submetidas a uma filtragem com base em parâmetros de qualidade de leitura, utilizando o *software Fastx-toolkit*, de modo que 90% do conteúdo dos *reads* fossem representados por um *phred score* de ao menos 30 (parâmetros "90x30").

Reads não satisfazendo tais pré-requisitos de qualidade foram também excluídos.

Posteriormente aos processos de filtragem e limpeza, o primeiro conjunto de dados (*reads* do grupo denominado "*set_0"*) foi alinhado contra o genoma de referência da espécie de ciclídeo africano *Metriaclima zebra*, utilizando o *software Bowtie2* (Langmead e Salzberg, 2012) com a finalidade de remover possíveis *reads* contaminantes. Os *reads* não alinhados contra o genoma de referência foram considerados contaminantes e, portanto, removidos permanentemente de todas as bibliotecas.

Os *reads* alinhados contra o genoma de referência (*reads* do grupo denominado "*set_1*") foram utilizados para a obtenção de um perfil de expressão de micro-RNAs conhecidos, tendo como base listas de miRNAs conhecidos de espécies de peixe, utilizando dados de precursores de micro-RNAs obtidos por meio do banco de dados *MirBase,* versão 21 (http://www.mirbase.org).

Tal processo se deu por um alinhamento utilizando o *software Bowtie* (Langmead et al. 2009) onde os dados de alinhamentos resultantes foram analisados. Os *reads* que falharam ao se alinhar contra o banco de dados do *miRBase* para peixes (*reads* do grupo denominado "*set_2*") foram separados e utilizados em um processo de predição de novos micro-RNAs. Para o processo de predição, os *reads* (*set_2*) foram analisados por meio do *software MiRCat*, componente do pacote "*The UEA sRNA Workbench*" (Stocks et al. 2012), que resultaram em sequências precursoras de micro-RNAs juntamente com sequências de micro-RNAs maduros, e suas respectivas localizações em relação ao genoma de referência utilizado (*M. zebra*).

Todos os diferentes micro-RNAs preditos em todas as bibliotecas foram copiados em um novo arquivo e unificados de modo a criar um banco de dados geral que contivesse todos os micro-RNAs preditos em todas as bibliotecas. Todos os elementos deste banco de dados foram renomeados utilizando identificações numéricas (Exemplo: *Novel_1*, *Novel_2*, e assim sucessivamente), mantendo-se uma versão original com os dados de cada biblioteca de forma separada. Um *bash script* personalizado foi escrito para renomear isoladamente todos os micro-RNAs separados de todas as bibliotecas, com base no banco de dados geral previamente criado. Deste modo, os diferentes nomes foram direcionados aos mesmos elementos presentes entre todas as bibliotecas. Tal procedimento possibilitou que as bibliotecas se tornassem passíveis de comparação em termos de identificação de seus elementos, não havendo desta forma, cruzamento de dados entre micro-RNAs diferentes.

Após estes últimos passos de processamento, uma nova rodada de análises foi realizada utilizando o *software Bowtie* (Langmead et al. 2009) seguido do pacte DESeq (Anders and Huber, 2010), com a finalidade da obtenção de um perfil de expressão de todos os micro-RNAs (um fluxograma é apresentado na figura 4).

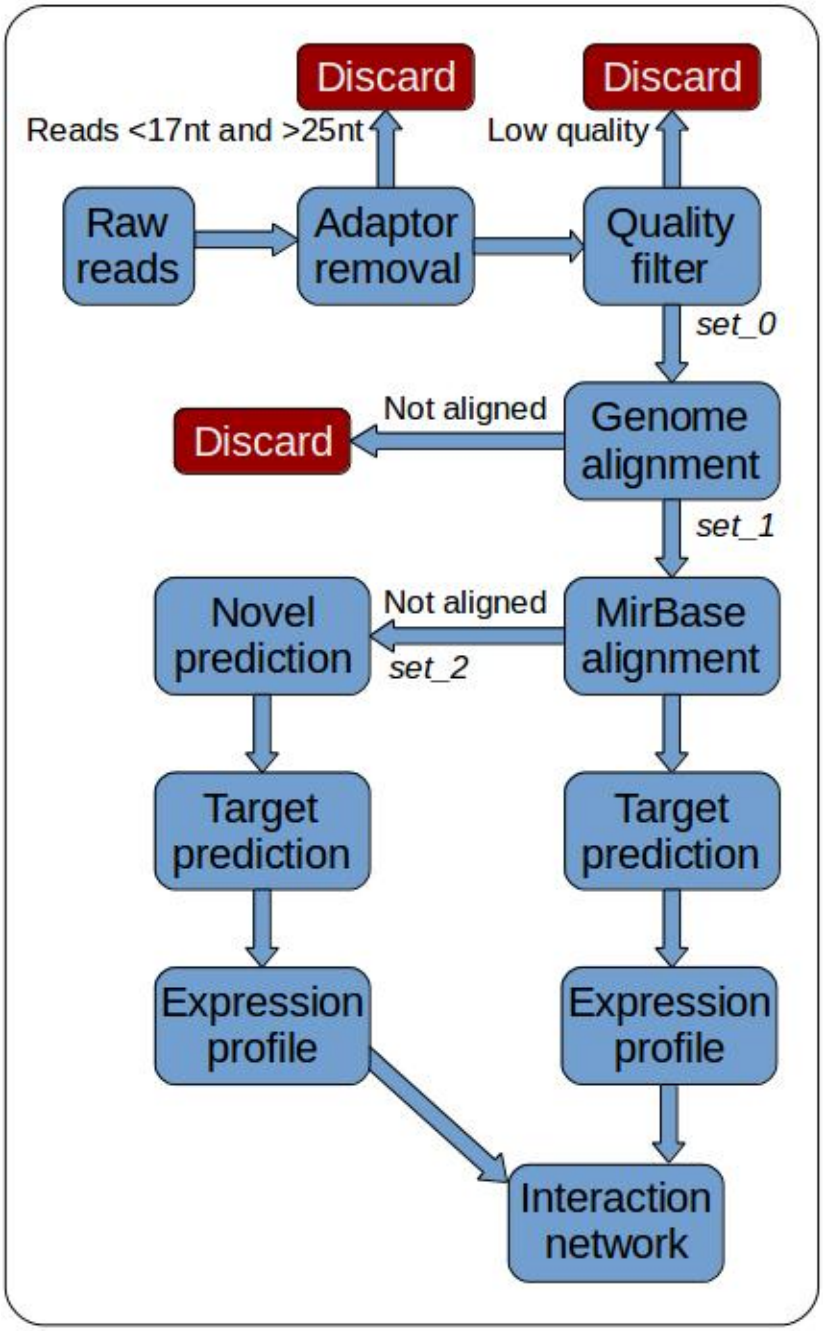


Figura 4: Fluxograma apresentando os passos executados (azul). Blocos em vermelho representam pontos em que *reads* foram excluídos por não perfazerem os requisitos mínimos.

3.7. Redes de interação

Para as redes de interação, todos os miRNAs foram utilizados em uma análise de *target prediction* por meio dos *softwares* miRanda (Enright et al.

2004), RNAhybrid (Krüger et al. 2006) e PITA (Kertesz et al. 2007), utilizando como sequências alvo, as regiões não traduzidas (3'-UTRs) da espécie de peixe *Danio rerio*, coletadas diretamente do banco de dados *Ensenbl-Biomart* (http://www.ensembl.org/). Tal banco de dados de UTRs foi organizado de forma a conter nos cabeçalhos de cada sequência os códigos de acesso relativos ao banco de dados, bem como o nome do gene ao qual tal UTR é diretamente relacionada.

As predições foram executadas utilizando parâmetros padrões dos algorítmos, e os resultados foram submetidos a uma filtragem baseada nos valores de energia livre dos pareamentos entre os micro-RNAs e as UTRs, medidos em kcal/mol. Tal limite foi estabelecido inicialmente em -18 kcal/mol. Este número representa a força de interação entre os miRNAs e as UTRs, onde os menores valores representam interações mais fortes. Neste caso, a aplicação de um limite de valor energético bastante baixo ajuda substancialmente a diminuir a quantidade de falsas predições que são possíveis de serem encontradas em resultados deanálises de predição de alvos utilizando ferramentas computacionais. Um segundo passo de filtragem foi realizado com o uso de um script escrito em *Python* de modo a coletar somente predições existentes nos três algoritmos.

A criação das redes de interação foi realizada por meio do *software* NAViGaTOR (http://ophid.utoronto.ca/navigator/). Tal *software* possibilita uma análise gráfica bastante detalhada das interações entre os elementos em questão.

4. RESULTADOS E DISCUSSÃO

O presente tópico será apresentado em formato de manuscritos, a serem submetidos para publicação em revistas científicas da área de genética, genômica e evolução. Os manuscritos encontram-se divididos em dois capítulos do presente volume.

4.1. Development of chromosomal markers based on next-generation sequencing: the B chromosome of the cichlid fish *Astatotilapia latifasciata* as model

Bruno EA Fantinatti and Cesar Martins

Universidade Estadual Paulista - UNESP, Instituto de Biociências, Departamento de Morfologia, Botucatu, SP, Brazil.

Corresponding Author: C Martins, Departamento de Morfologia, Instituto de Biociências, UNESP – Universidade Estadual Paulista, CEP 18618-970, Botucatu, SP, Brazil. Telephone: +55 (14) 3880-0462; E-mail: cmartins@ibb.unesp.br

Abstract

Based on 454 sequencing data of a microdissected B chromosome and Illumina whole genome sequencing data generated for 0B and 2B animals, we develop PCR- and qPCR-based markers for the B chromosomes of the cichlid fish *Astatotilapia latifasciata* (that possess 0, 1 or 2 B chromosomes). Specific PCR primers were designed in the way to have two amplified fragments for B positive samples, and only the control fragment for the B negative samples. In this way, the PCR maker only detected the B presence/absence, not providing information about the number of Bs. On the other hand, the qPCR markers clearly discriminated 1B and 2B samples. The copy number variation identified in the B chromosomes was confirmed by FISH-mapping. The analysis of polymorphic chromosome segments (as here investigated) based in the next-generation approach represents a powerful strategy to obtain markers to detect the presence/absence of extra chromosomes or gain or loss of genomic blocks. Furthermore, the qPCR also gives information in the copy number of a specific DNA fragment. Such approach finds application to investigate diverse chromosome polymorphism including B and sex chromosomes, and chromosomal duplications and deletions.

Key Words: Evolution, supernumerary chromosome, molecular markers, odd chromosome, selfish element.

Introduction

Chromosome studies have advanced after the introduction of *in situ* hybridization of nucleic acids (Pardue and Gall 1969; Gall and Pardue 1969) to chromosomes and improvement of molecular biology in the second half the 20th century. As a consequence, we had the development of DNA markers as chromosome probes and the emerging of molecular cytogenetics that find a wide application in the chromosome biology science. The development of fluorescent *in situ* hybridization (Pinkel et al. 1986), chromosome painting (Pinkel et al. 1988) and bacterial artificial chromosomes (BAC) probes (Korenberg et al. 1999), allowed an important advance in the chromosome knowledge. The more recent advances on the sequencing of several genomes brings the possibility to explore karyotypes and chromosome problems based in the nucleotide sequence of whole chromosomes and genomes allowing the karyotype reconstruction of species based in *in silico* data of complete nucleotide sequences of genomes. This new approach opens the possibility of a powerful integration of genomics and cytogenetics directed to investigation of chromosome problems. The ancestral syntenies of nucleotide sequences across different groups can be established based on sequence orthologies among species. Such an approach has allowed the application of electronic chromosome painting (E-painting) and the foundation of "*in silico* cytogenetics" as a new perspective for analyzing chromosomes and karyotypes (Kohn et al. 2006) and even the identification of conserved linkage groups between very distant related animal groups as human and sea anemone (Putnam et al. 2007).

The development of next-generation sequencing (NGS) in first decade of XX century changed the universe of genome sequencing, providing low-cost

and high-throughput sequencing available to ordinary laboratories. The applications of NGS seem almost endless, allowing for rapid advances in many fields related to the biological sciences, such resequencing genomes, comparative biology studies, public health, epidemiology, physiology, gene expression, among others. The recent technical advances in genomics contrasts and its contrast to classical cytogenetics lead several biologists to emphasize that cytogenetics was supposed to disappears in the modern genomic era. Here we explore NGS data in the obtention of chromosome markers to investigate classical problems in cytogenetics. Conventional and quantitative polymerase chain reaction (PCR and qPCR, respectively) procedures were developed based on NGS data as genetic markers to genotype the presence/absence and number of B chromosomes, using the cichlid fish *Astatotilapia latifasciata* as model. Our analysis shows that NGS data can also find application in the development of DNA markers to investigate chromosomal polymorphism.

Material and methods

DNA samples and karyotyping

Specimens of the cichlid fish *Astatotilapia latifasciata* (native to lakes Kyoga and Nawampasa in Uganda, satellite lakes of Lake Victoria) were obtained from a stock established from the trade and maintained in the fish facility of the Integrative Genomics Laboratory at Sao Paulo State University (Botucatu, Brazil). The animals were karyotyped by classical cytogenetic procedures using a Giemsa stain to identify 0B, 1B and 2B karyotypes as previously described (Poletto et al. 2010; Fantinatti et al. 2011). DNA samples

of 0B, 1B and 2B animals were extracted from fin clips and stored in -80°C for the next steps of analysis.

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).

B chromosome microdissection

One B chromosome was microdissected using an inverted microscope (Olympus IX51) equipped with a mechanical micromanipulator (Narishige, Tokyo, Japan), transferred to microtubes and posteriorly amplified by the kit Genome Plex® Single Cell - Whole Genome Amplification - WGA4 (Sigma-Aldrich, St. Louis, MO, USA), 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 in fluorescence *in situ* hybridization (FISH) to validate that the DNA obtained come from the B chromosome. In the reamplification reaction with WGA3 kit (Genome Plex® Whole Genome Amplification – WGA3, Sigma-Aldrich) the nucleotide digoxigenin 11-dUTP was incorporated to label the B probe.

Genome sequencing

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 newbler software package. The assembled contigs were blasted against the National Center for Biotechnology Information (NCBI) nucleotide collection. The contigs with hits against fish DNA sequences were retrieved for posterior analysis.

Genomic DNA of a 0B and 2B male samples were submitted to Illumina sequencing. Genomic DNA from each individual was sheared to an average size of 500bp using an S220 focused ultrasonicator (Covaris Inc., Woburn, MA, USA) and separate libraries were constructed for each individual using the TruSeq DNA sample preparation kit ver.2 rev.C (Illumina Inc., San Diego, California, USA). Paired end (2x100bp) sequencing of each library was performed in separate lanes on an Illumina HiSeq 1000 sequencer. The Illumina reads were aligned to the *Metriaclima zebra* cichlid reference genome using Bowtie2 software (Langmead and Salzberg, 2012).

All sequences and alignments are available in the genome browser of Sacibase database (www.sacibase.ibb.unesp.br).

B chromosome markers development: B presence/absence

For the development of PCR marker for the B chromosome presence/absence, the contigs generated from the 454 sequencing were submitted to alignment against cichlid genome data available at NCBI. A total of 32 different oligos were designed over nine different contigs that produced alignment against cichlid genomes. Conserved regions between the B genomic

data and the cichlid genomes were retrieved in order to have a control for the reaction. Primers were designed over the conserved regions in order to amplify a control fragment in the 0B, 1B and 2B samples. Subsequently, a third oligo was designed with its 3' end exactly over the nucleotide variation that is characteristic of the B genome in order to amplify a DNA fragment only in the presence of B chromosome DNA. The oligos were designed to work in a multiplex reaction in the way to amplify only one fragment in 0B animals (regarding the control fragment) and two fragments for the 1B and 2B samples (regarding the B specific fragment plus the control fragment).

A total of 31 oligos were designed, including B specific and control primers, covering nine different genomic regions from the 454 nucleotide database. The optimal annealing temperature was tested through a gradient termocycler ranging from 45 to 65°C. The reactions were carried out as follows: 1 U Platinum Taq DNA Polymerase, 1.5 mM $MgCl_2$, 1 X RXN Buffer, 0.32 mM dNTP, 0.2 µM of each primer (totaling 0.6µM) and ultrapure water up to 25 µl. Cycling was carried out as follows: 5min at 95°C, 34 x (1min at 95ºC, 30sec at 50ºC, 45sec at 72ºC), 5min at 72°C. The material was then analyzed in a 1% agarose gel.

B chromosome markers development: number of Bs

Illumina reads from both 0B and 2B datasets, were separately submitted to alignment against the cichlid *Metriaclima zebra* reference genome (available at Bouilabase) using the Bowtie2 software (Langmead and Salzberg, 2012). The result of the alignment was submitted to the analysis of coverage ratios between the two datasets, according to Valente et al. (2014). Since B

chromosomes present a high amount of repetitive DNAs, the genomic regions of the B chromosome would present a higher coverage of reads compared to the regular A chromosomes. In this way, a total of six genomic regions with high coverage in the 2B genome compared to the 0B genome were selected for qPCR analysis. Experiments of qPCR were carried out using known DNA samples previously genotyped by cytogenetic approach, i.e. three of each 0B, 1B and 2B samples. Such samples were used as controls in all qPCR genotyping procedures.

The qPCR experiments were prepared as follows: Firstly all the samples were diluted to 30ng/µl and the expected concentration was confirmed using Nanovue Spectrophotometer (GE). The reaction mix was prepared using 1X GoTaq qPCR Master Mix (Promega) added by 0.16 µM of each primer, 20 µl of DNA (30ng/µl) and ultrapure water up to 75 µl. The reactions were run in triplicates, (three of each 0B, 1B and 2B known samples). The hypoxanthine phosphoribosyltransferase gene (HPRT), widely used on qPCR experiments, was used as a control gene for the analysis of the obtained data based on gene dose ratio analysis (GDR) consisting in $2^{-\Delta Ct}$ (Bel et al. 2011).

After the first round of the PCR and qPCR tests using known samples, unknown samples were analyzed using only one sample for each control, including the qualitative PCR and the quantitative qPCR genotyping. In the last round of tests, DNA samples of different tissues (liver, brain, muscle, eye and heart) were collected and analyzed in order to test the presence of B mosaics among the tissue sampled.

Fluorescence in situ hybridization

The microdissected B chromosome and DNA fragments of genomic regions of the B markers were mapped onto the chromosomal complement containing B chromosomes of *A. latifasciata* by FISH. FISH was performed using the protocol described by Pinkel et al. (1986) with modifications (Cabral-de-Mello et al. 2012). After hybridization the metaphases of *A. latifasciata* were analyzed in an epifluorescence Olympus BX61 microscope (Olympus, Tokyo, Japan) and the images were captured using an Olympus DP73 system.

Results

The sequencing of the microdissected B chromosome using the 454 platform resulted in a total of 125,601 reads comprising 48,637,895 base pairs (bp). The assembly resulted in 3,836 contigs with the average size of 372 base pairs. For all the sets of PCR primers designed over the 454 data (Table 1) (Fig. 1a), the best amplifications were obtained using primers designed in the contig-_182 that corresponds to the scaffold_26 of the *M. zebra* reference genome. The use of such primers resulted in fragments comprising the expected size for both control and specific fragment, i.e. ~163bp for the control fragment, which appears in all samples, and ~260bp for the B-specific fragment, which appears only in B positive (B+) samples (Fig. 1b). Considering the length divergence between the control and the B-specific fragment, it was very clear to distinguish B+ and B negative (B-) animals. The FISH experiment using as probe the B-specific fragment of contig_182 (related to scaffold_26 on reference genome) shows a very high accumulation of such DNA throughout the entire length of the B chromosome and no signal was observed in the A complement (Fig. 1c).

The 0B and 2B whole sequenced genomes using Illumina platform comprised 401,017,570 reads and 306,823,512 reads respectively. All B+ animals, previously genotyped using the qualitative marker based on 454 genome data was then used to proceed to the next step of genotyping, which is related to the quantitative qPCR marker. Among the six genomic regions analyzed (see Table 2) the region corresponding to scaffold_13 presented the best amplification rates, with low standard deviation values between the samples. All the others regions (scaffold_3, 19, 26 and 31) presented amplification, but the standard deviation for these regions presented very high values. We can not discard the fact that standard deviations can be affected by pippeting techniques. Despite this, the scaffold_13 was then selected for analysis of samples. The results of the qPCR reactions were very clear and efficient in the differentiation of the 1B samples from the 2B samples based on 9 known control samples (three of each 0B, 1B and 2B samples). As expected, the copy number of the region selected for the tests are approximately twice in 2B samples in comparison to the 1B samples (Fig. 1e). Such result proves that the marker is able to clearly differentiate between the 1B and 2B genomes. The FISH experiments using as probe the quantitative qPCR marker fragment showed a very high accumulation of such element over the B chromosome, and no signals were observed in any of the A complement chromosomes (Fig. 1d).

Discussion

B chromosomes are one of the most astonishing issues on cytogenetics. Despite all the efforts directed to cytogenetics and B chromosomes, there is still no theory applicable to all B chromosomes concerning its origins and evolution.

One important problem is that B chromosomes probably arisen in different groups at different times, making the B chromosomes of different species completely unrelated to each other. Molecular cytogenetics has helped to answer many questions surrounding this issue by looking sequence sharing between A and B complement. Using this approach, most B-harboring species is thought to have a B chromosome originated from its own A chromosome set.

As seen in few species, B chromosomes can bring effects on host fertility and development (see Jones et al. 2008; Zima et al. 2003). It is also observed that different types of B chromosome presence causes an increase in the chiasma frequency related to the parasitism level of such B chromosome, leading to think that these effects take place through inducible recombination (Camacho et al. 2002).

Since cytogenetic techniques have a low resolution limit, high amount of information can be lost during the procedures and data analysis. Adding this to the fact that cytogenetics present data only regarding composition, the use of molecular tools represent a new approach to go further in B chromosome analysis, and help to uncover a little bit more the B chromosome conundrum.

The obtention of fish chromosome preparations involves in most cases cell division stimulus based on bread yeast injection and the sacrifice of the animal followed by kidney tissue extraction for chromosome preparation. Some problems can appear at this point: (i) The yeast injection causes contamination leading to immunological response, and disturbing any further functional analysis; (ii) As only the kidney tissue is used, the B chromosome presence in other tissues is not checked. (iii) Although chromosome preparations can also be achieved through cell culture, this is avery slow and expensive procedure.

One of the more desired questions surrounding B chromosome biology is in fact its function in the cell. New technologies make possible to analyze B chromosome by another approach, i.e. functional characteristics which represents a new way to study B chromosomes. NGS analysis based on transcriptomes and microRNAomes are effective approaches to find answers about B chromosome functions, because such techniques can point directly in alterations not necessarily possible to be observed through morphological data. Following such facts, it is necessary to develop a technique capable of detecting B chromosomes on DNA samples.

The new sequencing technologies in constant development and sequencing costs constantly decreasing, the obtention and analysis of NGS data is becoming much more accessible. Important information can be achieved with this approach, so the steps used to create the markers we presented here can be applied in other chromosome polymorphism problems in a faster and large scale approach.

FISH mapping

Both genomic regions explored for PCR and qPCR genotyping were detected in B chromosomes and not detected in the A genome. The absence of FISH signals in the A complement is due to the very low copy number of such sequences in the A genome, as seen by the qPCR data. The analyzed genomic regions are spread over the B chromosome accumulation of DNA fragments, which is characteristic of B chromosomes, due to the low selective pressure applied over segments from B chromosomes. This huge divergence in copy number can be observed in the qPCR data.

Final remarks

Such qualitative and quantitative genotyping can contribute to make possible the separation of animals in groups, according to their genotypes, i.e. number of Bs, for further analysis concerning B chromosome segregation, including a more accurate content analysis over B chromosome sequences through massive sequencing, providing data also over possible effects caused by B chromosomes in host fitness, based on transcriptome and microRNAome analysis. It is also possible to use this approach for analyzing different Bs that, cytogenetically, seems to have exactly the same content in terms of repetitivity of some specific segments. The number of investigations involving genome sequencing is increasing very fast due to the advances in the sequencing technologies. We know that B chromosomes are present in eukaryote groups in about 15% (Camacho, 2005) which represents a high rate. To sequence genomes without knowledge about B chromosomes on such species represents a very important problem. For solving that problem, it is important to develop new technologies for detecting such B chromosomes before sequencing procedures.

Acknowledgement

This work was supported by grants from São Paulo Research Foundation (2011/03807-7; 2013/04533-3).

References

As referências deste manuscrito encontram-se reunidas ao final deste volume no item “6. Referências Bibliográficas”.

Tables

Table 1: List of designed primers for qualitative PCR markers. The numbers 49, 80, 182, 207, 323, 1100 or 1987 are references to the contigs obtained with the 454 data, and the letter B or C, an identification regarding the specificity of the primer as and B specific primer or positive control (B or C+), followed by the reverse/forward annotation (F or R).

Primer name	Primer sequence
49B1-F	5' GAGCTTCACACTTGCAGAGGTAAGTCATTTTTGCAGAGAC 3'
49B2-F	5' GCTTCACACTTGCAGAGGTAAGTCATTTTT 3'
49C+-F	5' GTTTACAGTCTGATGATGGGACATCATGCTCTGC 3'
49-R	5' TGTCCAGAGTATAATCGCAGCCTTTGCGGT 3'
80B-F	5' GAGGCATTACATCGGTCTTTCCATCA 3'
80C+-F	5' GGTGAGCAGCAGGATTTTGAATTGAATGCG 3'
80-R	5' CCTGATTGAGTGCTTCTCACAC 3'
182B1-F	5' GGGTGTGTTTGGTTGTGGTTTGACAAGGAGTG 3'
182B2-F	5' GGAGTGAATTGTGATGGT 3'
182B3-F	5' GGTTTGACAAGGAGTGAATTGTGATGGTTAGATC 3'
182B4-F	5' GAGTGAATTGTGATGGTTAGATCACTAGGTAT 3'
182C+-F	5' AGAATGGTCCAAGGAAGG 3'
182-R	5' CCATCAGAACCAGCATTAA 3'
207B-F	5' GAGACACTTCTTGGAGAAAATGAAATGCCCAC 3'
207C+-F	5' ACCAGGCCAGGAGACGACTGAAGAACT 3'
207-R	5' GACCTGCAGAAATGTGAACATGGTTGCAGTTTACAA 3'
323B-F	5' GGGGGTGTTTTGCTTTTGGTTTTCCTACATTAGTTA 3'
323C+-F	5' GTATAAGCCATCTCTGTCATCTAAGGTACA 3'
323-R	5' GACACAGTACAGCTGACACAGACGAAGCAACAG 3'
764B-F	5' CCTGAGATGGTCCGATTGGGCTGGTAA 3'
764C+-F	5' GGTGAAGCATCAAAGAGCTCTCTGAGTCT 3'
764-R	5' GGAGACAAGGAGATGCGTGTTGGTGAAGTCCTAA 3'
1100B-F	5' GGGTGTGTGGAGATGTACATCAGCACACATGTT 3'
1100C+-F	5' CACTGAGACGGCATTGGCATGAGAAA 3'
1100-R	5' AGCATGGTGGCAGAGGTCTTTA 3'
1987B-F	5' CCCTCCTGTTATTCATTCCCTA 3'
1987C+-F	5' TACTTTGCTGTGTGTTTTGCCTGTC 3'
1987-R	5' AAGTGTGGCTGTGTGCAGGCAGGAAT 3'

2519B-F	5' GCAGGATTCAGGAGTGAAGCATCTGTGTGA 3'
2519C+-F	5' CACTAAACTGCAGACATCAGGCTG 3'
2519-R	5' CATTGTTCTGCTGCAGTCAATGGAC 3'

Table 2: Selected genomic regions for qPCR primer design. The scaffolds can be accessed in Bouillabase database (www.bouillabase.org).

Scaffold	Start position	End position	Size (bp)
Scaffold_3	9112721	9126380	13659
Scaffold_13	5682225	5684956	2731
Scaffold_19	960417	963148	2731
Scaffold_26	1707513	1710244	2731
Scaffold_31	5673821	5680650	6829
Scaffold_324	77097	77779	682

Figures

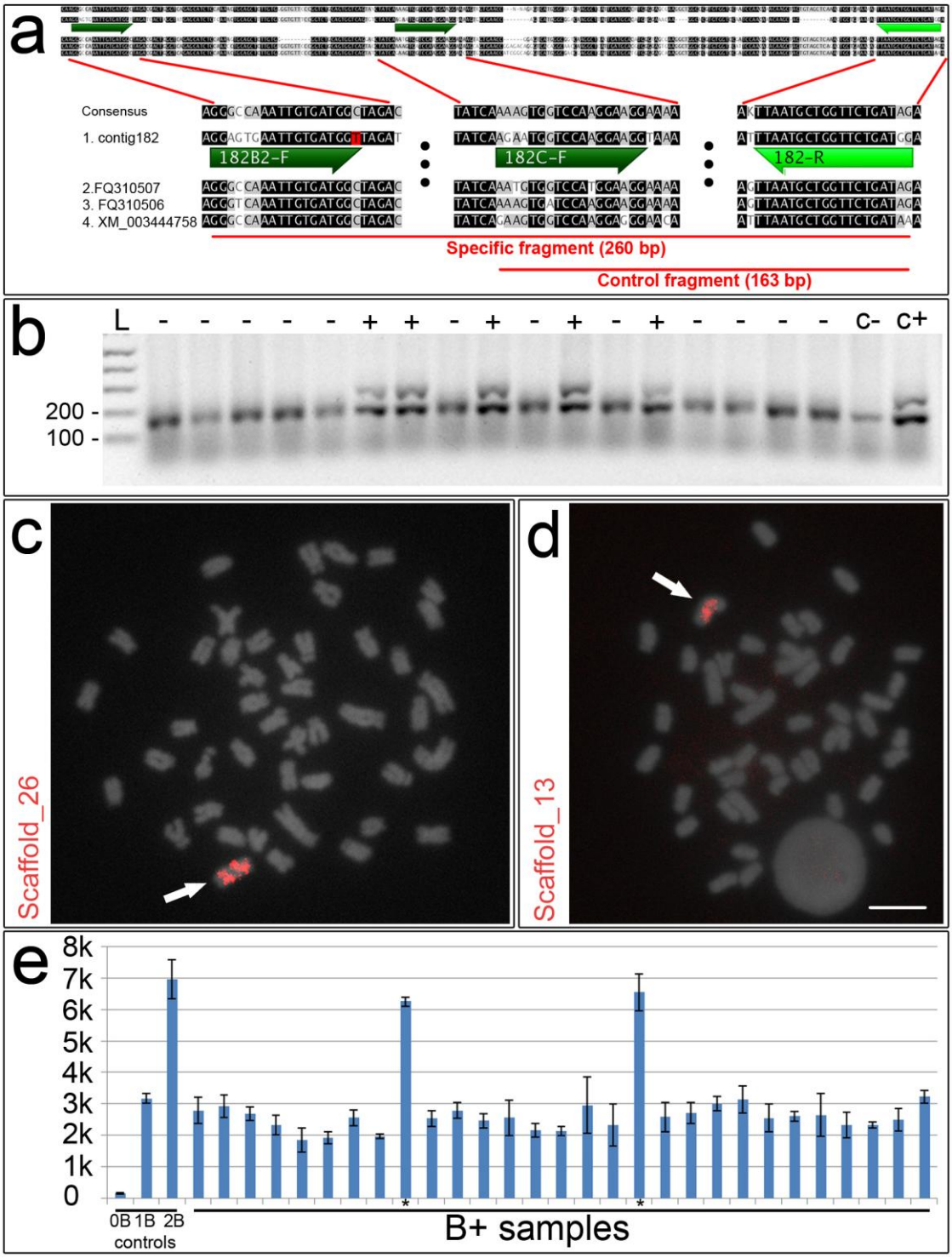


Figure 1: (a-c) Design and results for the qualitative PCR marker. (a) Scheme for the primer design with emphasis on the B-specific and control fragments of the scaffold_26. Three genomic sequences of different fishes: *Dicentrarchus labrax*, *Dicentrarchus labrax* and *Oreochromis niloticus* respectively (FQ310507, FQ310506, XM_003444758) were used to establish a consensus to be compared to the 454 sequence data of the microdissected B (contig_182). (b) 1% agarose gel showing PCR

products from B+ and B- DNA samples. Note that the B- samples presents only one DNA fragment (control fragment) and the B+ samples present two fragments, one related to the control fragment and a second one related to the B-specific fragment. (c) FISH using the PCR marker region as probe. Arrows indicates the B chromosome. (d-e) Development and results observed for the qPCR marker analysis. (e) Graphic plot showing the relative number of copies detected among B+ samples and control samples. Note that 2B samples (asterisks) present twice copies compared to the 1B samples. The standard deviation is showed for each sample. (d) FISH using the qPCR marker region as probe. Arrows indicates the B chromosome. Bar indicates 5 μm.

4.2. B chromosomes effects based on large scale bioinformatics analysis of micro-RNA in the African cichlid fish *Astatotilapia latifasciata*

Bruno EA Fantinatti[1], Mike Tsay[2], Igor Jurisica[2] and Cesar Martins[1]

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

[2]Princess Margaret Cancer Centre, Toronto Medical Discovery Tower, Toronto, ON, Canada

Corresponding Author: C Martins, São Paulo State University - UNESP, Institute of Biosciences, Department of Morphology, 18618-970, Botucatu, SP, Brazil. Telephone: +55 (14) 3880-0462; E-mail: cmartins@ibb.unesp.br

Abstract

B chromosomes are additional chromosomes present in about 15% of eukaryote groups and its function and evolutionary history is still not clearly elucidated. For that reason, B chromosomes are an interesting investigation issue in chromosome biology. We comparatively analyzed 0B and 1B samples from four different tissues of the cichlid fish *Astatotilapia latifasciata*. Known and Novel micro-RNAs expression profile were obtained through Illumina small RNA deep sequencing. Our data suggests the possibility that B chromosomes does not present effects itself. Instead, B chromosome presence act over the A complement genome causing functional modificationsin the cell biology. Such alterations seems to involve genes related to cell cycle control and development. In this way, our data give strong evidences that B chromosomes are no longer inactive elements as expected before, but indeed can interfere in very important biological process of cells and organisms.

Key Words: Non-coding, Evolution, Supernumerary chromosome, Effect, Target gene, micro-RNAs

Introduction

B chromosomes, also known as supernumerary or even “Bs”, were firstly reported by Wilson (1907), and became one of the most intriguing cytogenetic characteristic in the species they are found, which corresponds to ~15% of the eukaryote groups (Camacho 2005). Bs are not considered essential to host life. This high transmission rate can be a result of drive mechanisms, which increases the probability of Bs being kept on host genome (Camacho 2005). Consequently, the presence of B chromosomes in the genome might cause some effects, which are observed in some species. Such effects can be presented as positive or deleterious characteristics depending on the number of Bs in the genome and in the parasitism level of the B chromosome (Camacho et al. 2002; Zima et al. 2003; Jones et al. 2008).

Such extra elements do not follow the Mendelian laws of inheritance (Camacho et al. 2000; Camacho 2005) and present a huge variation regarding morphology, size and content (Ziegler et al. 2003; Schmid et al. 2006; Martis et al. 2012; Valente et al. 2014; Trifonov et al. 2002). They appear aside the normal set of chromosomes (called A chromosomes), and seem not to perform recombination with any of them. In fact, only one case of Bs x As recombination has been observed in a mammal species (Basheva et al. 2010). In general this lack of recombination makes B chromosomes prone to a low selective pressure, resulting in a high accumulation of repetitive elements (Fantinatti et al. 2011; Martis et al. 2012).

Cichlid fishes are a group that possesses a considerably number of species harboring B chromosomes. Bs have been identified in seven South American and fourteen African species of cichlid fish (Poletto et al. 2010a,

2010b; 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 Lake Kyoga (Lake Victoria system) (Poletto et al. 2010a), and can possess one or two large B chromosomes. Additionally, supernumeraries were detected in all the Victoria cichlids karyotyped, which includes in total thirteen species (Poletto et al. 2010a; Yoshida et al. 2011; Kuroiwa et al. 2014). On the other hand, among fourteen karyotyped species of Lake Malawi, B chromosomes were detected in only one species, *Metriaclima lombardoi* (Poletto et al. 2010b). Cytogenetic analyses found no B chromosomes in another eight African and Asian cichlids (Poletto et al. 2010b).

Classical and molecular cytogenetic approaches have been usually applied over B chromosomes in an attempt of answering questions regarding its origin and evolution. Such techniques present a very low-resolution level, resulting in superficial view of the genome. In this way, new technologies are needed to advance in the cytogenetics of Bs. Recent works have directed high throuput next-generation sequencing for uncovering B chromosomes biology in plant, yeast and animal species (Goodwin et al. 2011; Zhou et al. 2012; Martis et al. 2012; Valente et al. 2014). Large scale analyses of genome and transcriptome content is a very powerfull approach for obtaining information on biological issues and have been used for various purposes like development (McDaneld et al. 2009), diseases (Vaz et al. 2010) and evolution (Nachtigall et al. 2014).

Under the focus of massive data analysis, the non-coding RNAs have been highly explored in recent years considering their strong impact on

controlling several biological process (Rinn and Guttman 2014; Morris and Mattick, 2014). Among the non-coding, micro-RNAs (miRNAs), a well reported class of non coding RNAs ranging from 18 to 22 nucleotides, are known to perform a fine regulation of gene activity in a post-transcriptional mode after associated with the RNA induced silencing complex (RISC) and binding to the 3'UTR regions of mRNA (Bartel et al. 2004; Chen et al. 2014). Such functional regulations acts via degradation or cleavage of mRNAs during the entire lifetime of the organism and are responsible for the control of important processes such as development control of tissues during differentiation, basic functions maintenance such as tissue regeneration and stress response. The way in which B chromosomes affect genome functions is still not well understood and the expression of such small controllers as miRNAs can be affected by the presence of B chromosomes in the genome. Considering miRNAs are important players in the cell physiology and B chromosomesare still enigmatic considering their possible biological role, here we look into miRNA profiles as a powerful approach to better understand B chromosome biology.

Material and methods

Animal Samples

Specimens of the cichlid fish species *Astatotilapia latifasciata* (native to lakes Kyoga and Nawampasa in Uganda, satellite lakes of Lake Victoria) were obtained from a stock established from the trade and maintained in the fish facility of the Integrative Genomics Laboratory at Sao Paulo State University (Botucatu, SP, Brazil). 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 Institute of Biosciences/UNESP - Sao Paulo State University ethic committee on use of animals (Protocol no. 486-2013).

The animals were genotyped regarding B chromosome presence by using the molecular markers developed for this purpose (Chapter one of this volume). For the RNA extraction, a total of 40 samples were collected. The samples included four different tissues (brain, muscle, gills and gonads) from both male and female specimens, both with and without B chromosomes and organized in duplicates (muscle and gills) and triplicates (gonads and brain).

Total RNA extraction, micro-RNA sequencing and data filtering

Total RNA was extracted from 0B and 1B specimens, using TRIzol reagent (Life Technologies) following the manufacturer's specifications. The integrity of the samples was evaluated using Bioanalyzer (Agilent) for sequencing. The sequencing procedures were performed by LC Sciences (www.lcsciences.com) using an Illumina HiSeq platform. The sequenced miRNA data were downloaded through a FTP link provided by LC Sciences. Adaptor removal was carried out using Fastx-toolkit software (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The quality of the data was analyzed by FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and a quality filtering process was carried out using FastX-toolkit. For this filtering, all the reads presenting at least 90% of its length with a phred score of at least 30 were kept for further analysis. Reads with quality scores under this threshold were permanently discarded. Filtered reads (quality filtering) were submitted to a

genome alignment against *Metriaclima zebra* genome as reference using Bowtie2 software (Langmead e Salzberg, 2012) in order to remove possible cross-species contamination. Only reads aligned to *M. zebra* genome were used on downstream analysis.

Known micro-RNAs profiles

For the identification of known miRNA profiles, the filtered reads (quality and contamination filtering) were aligned against fish mature miRNA dataset created using fish microRNA data collected from MirBase version 21 (mirbase.org). Such database was created using all the miRNAs described for fish species and a custom Bash script was developed in order to filter the reference data regarding repetition of elements throughout the used fish species. Alignments against such fish reference using Bowtie software (Langmead et al. 2009) were run. No mismatches were accepted, and raw count data was obtained from BAM files using Samtools software (Li et al, 2009). Count data normalization and differential expression (DE) analysis were carried out using R/Bioconductor DESeq package (Anders and Huber, 2010) generating heatmaps regarding expression values. A list of DE miRNAs was extracted from this data considering a fold-change ≥ 2 and p-value ≤ 0.05. Expression values over the fold-change threshold but with p-value > 0.05 were permanently discarded.

Novel micro-RNAs prediction

Filtered reads with no alignment against fish miRNA database were recovered and used for novel miRNAs prediction procedures. A custom Perl script was used to filter those reads to a size between 17 and 23 nucleotides. Reads comprising sizes < 17 and > 23 were removed. Novel prediction was carried out using the remaining reads and miRCat software, a module from the UEA sRNA Workbench software (Stockset al. 2012) using standard parameter settings for animal species. The *Metriaclima zebra* genome was used as reference for novel prediction procedures (available at www.bouillabase.org). After the novel prediction, novel miRNAs sequences were added to the fish miRNA database, and the alignments were run again for the obtention of a global miRNA expression data containing known and novel information for all the samples.

Target prediction

Target prediction procedures were carried out using known and novel miRNA lists and three different target prediction algorithms: miRanda (Enright et al. 2004), RNAhybrid (Krüger et al. 2006) and PITA (Kertesz et al. 2007). As there is no UTR data described for *A. latifasciata*, *D. rerio* 3'-UTRs obtained from Ensembl-Biomart database were used as target input. After the target predictions were run, a custom Bash script was designed to filter such data to a free energy threshold below -18 kcal/mol. Predictions presenting values higher than -18 kcal/mol were permanently discarded from the analysis in order to decrease false positive rate on miRNA:target interaction predictions. A custom Python script was designed to collect only the overlapping predictions over all the three algorithms. As a result, only interactions predicted by the three

algorithms with free energy values below -18 kcal/mol were kept to proceed to the interaction network assembly.

Are DE miRNAs transcribed from B chromosome genomic content?

The possibility of DE miRNAs being originated directly from B chromosome content was considered, repsesenting thus an effect caused directly by B chromosome activity. High coverage peaks present in the 2B *A. latifasciata* genomic reads in comparission with 0B genomic reads is supposed to represent highly represented regions in the B chromosome content (Valente et al. 2014). Considering such information, a blast search was performed using all the miRNAs detected and predictedusing *M. zebra* genome as database, in order to check if miRNAs sites match such high coverage regions of B chromosome genome.

miRNA:mRNA Interaction Networks

Interaction networks were assembled using NAViGaTOR software (http://ophid.utoronto.ca/navigator/) based on overlapping prediction results. After creating a main network containing all the nodes for all the samples, such nodes were filtered out in order to create four different specific networks representing only DE miRNAs for each tissue analyzed.

GO analysis

For analyzing biological significande of genes targeted by DE miRNAs, GO analysis were carried out using BinGO, a Cytoscape extension (Maere et al. 2005). A full search was carried out using *Danio rerio* annotations regarding molecular function, biological process and cellular component.

Results

Total RNA extraction, micro-RNA sequencing and data filtering

The read numbers in all samples varied from 7,710,362 to 19,914,188. After adaptor removal procedures, quality filtering and genome alignment, an average of 67% of all reads was kept on the libraries for being used on downstream analysis (Supplementary Material 1).

Known micro-RNAs expression profiles

For the known miRNAs profiles, the filtered reads were subbmited to Bowtie alignment using fish known mature miRNAs database as reference for expression profile obtention. The results shows that 632 out of 1,029 micro-RNAs described for fish species were expressed in *A. latifasciata* considering all the libraries, with specific variations between different tissues. Such expressed miRNAs corresponds to 61.4%. Among the 632 expressed known miRNAs in *A. latifasciata*, 53 presented differential expression values (fold-change ≥ 2 and p-value ≤ 0.05) between B- and B+ samples. Such DE miRNAs presented variations between the tissues, regarding B chromosome presence. A heatmap presenting expression values for known miRNAs can be observed in Supplementary Material 2.

Novel micro-RNAs prediction and expression profiles

The reads that failed to align against known miRNAs database were applied on prediction procedures. Such predictions were carried out using miRcat software and resulted in a total of 57 new miRNAs over all the samples. Both precursor and mature sequences were predicted, and the precursor hairpin can be onserved for all the novel miRNAs (Figure 1). As for the known miRNAs, novel elements presenting fold-change ≥ 2 were considered differentially expressed. Eleven out of 57 presented differential expression (novel_12, _15, _21, _22, _23, _26, _28, _32, _47, _52 and _55) (Figure 1). As for the known miRNAs, novel DE elements also presented variation between different tissues and had a prevalence of 19.2% of all novel miRNAs. The expression values for novel miRNAs are presented in a heatmap (Suppementary Material 3).

Despite heatmap data identified clusters among the samples (Figure 2; Supplementary Material 2 and 3), such analysis did not clearly differentiate tissues, probably due to the variations observed in biological samples related to B chromosome presence. Brain is the only tissue clearly clustered apart despite such differentiation can only be observed when known data is considered in the analisys. For all the three heatmap datasets (known-only, novel-only and both novel and known) it was not possible to detect clusters linked to sex and/or B chromosome. Some global inversed patterns can be observed, characterized by groups of miRNAs with low expression in a given tissue and high expression in another tissue. Such patterns help separating the different tissues more clearly (Figure 2, Supplementary Material 2 and 3).

Target prediction

The target prediction procedures yielded a total of 42,728,923 interactions between novel/known miRNAs and *D. rerio* 3'-UTRs. After the filtering process based in an interaction threshold of -18 kcal/mol, a total of 16,943,862 interactions remained for both known and novel miRNAs. Overlapping prediction resulted from RNAhybrid, miRanda and PITA analysis based on a Python script decreased the number of interactions to a total of 12,360. Such remaining interactions were then used to create a network in order to obtain a visualization of all the interaction data. The target prediction threshold causes the network to have less DE miRNAs than volcano plots due to the fact that not all DE miRNAs presented positive target prediction considering the three prediction algorithms.

Interaction networks

Only DE elements presenting predicted targets were considered for assembling the interaction network, resulting in a decreased number of elements composing the interactions. In such way, only DE miRNAs presenting target prediction are presented in the networks. Four networks were created regarding each tissue: brain, gills, gonads and muscle (Figure 3). Nodes representing up-regulated and down-regulated miRNAs appears in red and blue respectively. Edges linking genes targeted only by novel miRNAs, only by known miRNAs and by both novel and known miRNAs appears in light red, light green and light blue, respectively. Male DEs are represented by square nodes,

and female DEs are represented in elliptic nodes. Target genes are represented by small black nodes. Each node has an identification name.

The four tissue networks presented variation in graph structure (Figure 3). Brain and gill tissues network has only two graphs followed by gonads network presenting six graphs and muscle which presented a total of eight graphs. Brain tissue network graph has a total of eight nodes comprising one novel miRNA, two known miRNAs and five target genes, totaling six interactions. It comprises the most simple interaction, divided in two graphs where all of them are presented as over-expressed. Two miRNAs are over-expressed in males (ccr-miR-205 and ipu-miR-205) and only one is over-expressed in females (Novel_32). A total of five genes are targeted by these miRNAs. Novel_32 targets *appb*, *rnf128a* and *cltcb* genes. ccr-miR-205 targets only *slc22a7b* gene and ipu-miR-205 targets both *slc22a7b* and *arid4a* genes. For muscle tissue, a total of 67 nodes remained in the network. 63 interactions was observed comprising two novel miRNAs and eleven known miRNAs. 54 target genes were detected. Considering only the thirteen DE miRNAs on muscle tissue, almost all of them (11) are up-regulated and only hhi-miR-449 is up-regulated in males. Only two miRNAs are down-regulated, each one in a different gender (dre-miR-499-3p in males and novel_32 in females). Gonads network presented 58 interactions. Its 44 nodes were divided between nine miRNAs (two novels) and 35 target genes. For gonads, the most interesting aspect is the graph composed by novel_26. Such graph has inversal expression level for novel_26 where it appears to be up-regulated in males, and down-regulated in females. Novel_26 and ssa-miR-203b-3p jointly targets *dhrs7cb* gene, but such inversion

do not occurs in ssa-miR-203b-3p. Gills tissue network is divided in two graphs, and comprised 11 nodes, six target genes and five miRNAs (one novel miRNA).

All the networks presented shared graphs. There were detected four shared graphs in total: (i) gills and muscle, (ii) muscle and brain, (iii) brain and gonads and (iv) gonads and muscle. Muscle present sharing of a specific graph containing three known miRNAs (ipu-miR-124a, ola-miR-124-3p and dre-miR-124-3p) and one novel miRNA (novel_15). Both graphs have five target genes, *cdc42bpb*, *plcd1a*, *neurog1*, *gria4a* and *prkci*, and *cdc42bpb* and *plcd1a* are targeted only by dre-miR-124-3p. Such graphs appear exactly with the same structure in both gill and muscle tissues, and these miRNAs are detected as over-expressed exclusively in females. The second sharing group appears between muscle and brain. Novel_32 appears in both tissues presenting three genes as targets (*appb*, *cltcb* and *rnf128a*). The expression level of novel_32 in both graphs is inverse, where in muscle such miRNA appears to be down-regulated only in females, and in brain such miRNA appears to be up-regulated only in females. A third sharing graph occurs between brain and gonads. It presents two known miRNAs (ccr-miR-205 and ipu-miR-205) up-regulated only in males, and two genes appears to be targeted by these miRNAs (*arid4a* is targeted only for ipu-miR-205 and *slc22a7b* is targeted by both ccr-miR-205 and ipu-miR-205). A fourth is related to novel_12. It appears to be differentially expressed in gonads and muscle but it is down-regulated in male gonads, and up-regulated in female muscle. Such novel has as target the *odf3b* gene. An isolated graph appears to be interesting. In gonads, novel_26 is down-regulated in females and up-regulated in males and targets the gene *dhrs7cb*.

Discussion

Differential Expressed miRNAs

Our study shows a total of 64 DE miRNAs (fold-change ≥ 2 and pvalue ≤ 0.05) (53 for known miRNAs plus 11 for novel miRNAs) with interactions with a huge number of genes, suggesting extensive effect over the transcription profile of cells. Such increased number of targets can be responsible for important effect on the cell structure and function, taking into account that one specific miRNA can act over a huge number of target genes.

Based on our interaction network data, it was observed that most DE variations (19 out of 30 DE miRNAs) occurs in females. From those, only three miRNAs were detected as down-regulated and 16 female-specific DE miRNAs were up-regulated throughout the four tissues analysed. Differences between tissues were also observed, and brain tissue was found to be the less affected one with a very small fraction of DE miRNAs (three miRNAs only). Gills came in second place, presenting five DE miRNAs, gonads had nine DE miRNAs followed by muscle, which had 13 DE miRNAs. An overall analysis of miRNA profiles evidences clear DE among tissues, but absence of clear DE signal among B- and B+ samples. In this way, B chromosome activity appears to depend directly on the transcriptional activity of A chromosomes regarding both tissue and gender, resulting in such differential expression levels. Differential expression of miRNAs occur during development of a given specimen considering different tissues to be formed. It is observed that even inside a specific tissue, miRNA expression might vary depending on the region of such tissue. Such variation occurs, for instance, in the expression of miR-92b, miR-124, miR-9 and miR-135c in zebrafish (Bizuayehu et al. 2014). Regarding eye

formation, there is also variation in miRNA expression. A huge number of miRNAs are related to eye formation in zebrafish, and such miRNAs are also related to the same function in *Lates calcarifer* - the asian seabass. But miR-181a and b were specifically expressed in retina cells of zebrafish (Kapsimali et al., 2007). Micro-RNAs variation also occurs considering sex determination and development. As observed for *Hippoglossus hippoglossus*, a given set of miRNAs is higher expressed in testis than in ovary tissue (Bizuayehu et al. 2012) evidencing the discrepance over miRNA expression. Such alteration is responsible for leading cell differentiation process and conduct such differentiationin order to result in male or female gonads.

In fact, BLAST search did not detect any presence of DE miRNAs mapped to B chromosome genome segments. Such data shows that the alterations detected over the four different tissues are not direcly linked to miRNAs activity coming from B chromosome. Instead, B chromosome appears to interfere on miRNAs profile of A chromosomes complement.

Genes regulated by DE miRNAs

The biological processes under the control of DE miRNAsis very diversificated. For gills and muscle (considering here the shared graph), *neurog1* gene appears to be targeted by novel_15, wich appears to be up-regulated only in females. According to ontology data *neurog1* is presented as an important gene related to brain development. *Neurog1* is related to regulation of cell fate during neuronal development (Onoguchi et al. 2012) and is necessary for the formation of zebrafish ganglia (Andermann et al. 2002). *Neurog1* appears to control the activity of *neurod* gene during the development

process. The knockdown of *neurog1* causes the absence of expression of *neurod*. Such knockdown causes the decrease of production of neurons (Andermann et al. 2002) representing an important effect. As *neurog1* is targeted by an over-expressed miRNA (novel_15) in two distinct tissues, it does not seems to occurs in brain tissue, resulting in the fact that such up-regulated miRNA does not represents alterations over brain tissue. *Neurog1* morpholino injection resulted in almost no production of neurons in zebrafish. Such result would be expected for *A. latifasciata* if novel_15 were up-regulated in brain tissue, representing thus an negative effect of B chromosome as observed when high number of Bs are present in the host species (Jones et al. 2008).

Appb gene seems to suffer opposite influences in brain and muscle tissue. Considering brain tissue, *appb* is targeted by the up-regulated novel_32. Such novel_32 is down-regulated in muscle tissue. In both tissues, this alteration is caused exclusively in female gender. The *appb* gene, also known as amyloid precursor protein-b, is necessary for synaptic formation in zebrafish and has been related to alzeimer disease. Morpholino assays have shown that *appb* expression is necessary in the inicial stages but not during subsequent processes (Abramsson et al. 2013). As *appb* is targeted by an up-regulated miRNA in brain tissue, it is expected to be causing a possible delay in brain morphogenesis. In fact, such alteration seems not to be completely disrupting any important function since B-positive speciments of *A. latifasciata* can normaly reach adult stage.

The *prkci* gene, also present in the shared graph between muscle and gill, and presents a very complex biological process activity. According to ontology analisys, it is related to neuronal differentiation and neural tube formation. *Prkci*

is important to keep precursors cells during cell division (Roberts and Appel 2009) in order to only one daughter get differentiated in neurons while the other one keeps itself as a precursor cell. It is also related to embryonic morphogenesis and heart formation (Uhalte et al. 2012), where specimens lacking *prkci* expression shows abnormal formation.

Considering shared graphs from gonads and brain tissue, *arid4a* gene appears targeted by ipu-miR-205. *Arid4a* gene, also known as AT rich interactive domain 4A, is a cell cycle gene which down regulates cell proliferation (Kumar et al. 2012) and is also considered a leukemia suppressor gene (Wu et al. 2008). It was observed that cells over-expressing *arid4a* gene appears to develop less colonies than control cells (Kumar et al. 2012). MicroRNA ipu-miR-205 is detected as up-regulated in both brain and gonads tissues. Such alteration is present only in male specimens. Valente and co-authors (2014) detected a group of genes related to cell cycle in the B chromosome of *A. latifasciata.* The fact that these genes were found to present a high conservation rate in the Bs uggest they might be active, or at least they have been recently transferred to the B chromosome. Action of genes related to cell cycle is thought to be helpful for the B chromosome to be kept on the genome. *Arid4a* is considered a cell cycle down-regulator gene and it is targeted by an up-regulated miRNA in gonad tissue (ipu-miR-205). In brain, such alteration would not bring any important effect. But the fact that cell cycle controller gene is affected by an up-regulated miRNA on gonads tissue is important. Such effect or alteration, when appearing in gonad tissue, is prone to help the B chromosome to be transmitted to gametes and then helping Bs to be kept in the species by increasing cell ploriferation. That would lead us to

hypothesize that the B transmission rate through male specimens would be more frequent than in females. In fact, Fantinatti and co-workers (2011) showed that when only one B chromosome is present in the genome, it is more frequent on females. On the other hand, when two B chromosomes are present, it seems to be more frequent on males. Despite this data only reflects differences between sex and 1B and 2B samples (Fantinatti et al. 2011) and not B transmission rate between genders, B transmission rate of *A. latifasciata* could be altered due to an accumulation of miR-205 over-expression when 2B is present in the cell. Such accumulation of up-regulated miR-205 could increase the targeting rate over *arid4a* and thus decreasing *arid4a* activity. Such down-regulation of *arid4a* would increase cell ploriferation and thus might contribute to B chromosome drive. Unfortunately, there is still no data regarding B transmission rate between male and female, which could be obtained through directed crossing analysis.

B chromosomes and miRNA effects on A. latifasciata *cells*

Our data shows miRNAs with differentially expression levels related to B chromosome presence/absence in the cichlid fish *A. latifasciata*. This might contribute to B chromosome segregation by causing an unbalance on genes activity profile. Such genetic unbalance might be helpful for the B chromosome to increase the strength of drive mechanisms. In fact, B chromosomes do not present a Mendelian mode of inheritance. Despite the molecular mechanisms of meiotic drive is still not well understood (Houben et al. 2014), drive mechanisms are considered responsible for B chromosome advantages during meiosis and such mechanisms might vary between different species. *Crepis capillaries*

(smooth hawksbeard), for example, presents B chromosomes that tend to accumulate in the cells that will become floral organs, increasing then the probability of being transmitted to next generations (Jones 1995). In maize, there are three different processes helping to increase B chromosome segregation that includes (i) non-disjunction at pollen grain mitosis, (ii) preferential fertilization and (iii) suppression of meiotic loss (Jones et al. 2008).

Variation in chiasmata rate on B-carrier cells was demonstrated for the grasshopper *Eyprepocnemis plorans* (Camacho et al. 2002) and those variations can be a result of such unbalance over cell division related genes, consequently improving B chromosome drive. Such alterations might represent a very thin limit between a positive effect that will be responsible for the B chromosome to be kept on the genome, and deleterious effect that will be responsible for reproductive/surviving issues, consequently leading the species to lose the B chromosome. It was observed that high number of B chromosomes in rye causes decrease in development and in *Aegilops speltoides* decreases the pollen grains viability (Jones et al. 2008).

Recent data have shown B chromosomes are no longer considered gene poor, but can harbours a considerable number of genes (Goodwin et al. 2011, Martis et al. 2012, Valente et al. 2014). Some of those genes are found with a high level of conservation, which diverges from the though that B genes must be inactivated. In fact, some of those genes are found to be transcriptionally active (Banaei-Moghaddam et al. 2013; Valente et al. 2014) and could act modulating the transcriptional profile of host genome.

The obtained data shows that DE miRNAs levels do not follow the B chromosome presence in all the different tissues analyzed. This information

leads to think that the nuclear environment surrounding B chromosome is important for its activity, which will result in variable effects on host genome depending on tissue type. This observation is in agreement with data observed in rye (*Secale cereale*) where B-specific pseudogenes-like fragments were found to be differentially expressed in a tissue-specific manner (Banaei-Moghaddam et al. 2013). Also it was observed that the majority of DE miRNAs targets genes related to development and cell cycle control. Such information is in agreement with data observed for B chromosome content in *A. latifasciata* (Valente et al. 2014) and *Mycosphaerella graminicola* (Goodwin et al. 2011) where genes related to cell cycle were found to be present in B chromosome content.

Although we did not detect any effect of miRNAs over B chromosome genes previously identified (Valente et al. 2014), we can not rule out the possibility of any effect of miRNAs in the regulation of such genes. It is believed that B chromosome transcriptionally active genes could probably help B chromosome maintenance in the genome. B chromosome sequences have originated through duplication events from original copies in the A genome (Valente et al. 2014). After duplication the B copies are believed to 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. It was observed that alterations on UTR sequences might create different miRNA target sites (Amato et al. 2013). Such alterations can change the gene regulatory structure by modifying the miRNA target region. B chromosome located genes are prone to pass through such alterations due to low selective pressure that rules their

evolutionary history, acquiring new regulatory regions and assuming new functions.

Conclusion

Our results shows that B chromosome biology is more complex than thought before. It seems clear miRNA transcriptomes reflect an balance of the B chromosome genome presence and its interactions affecting the regulation of the A chromosomes genome. MicroRNAs seems to be acting over development and cell cycle related genes, and thus, shaping the environment favorable to B chromosome transmission. Other non-coding RNAs and also other molecules such proteins can also be affected by B chromosome presence in the nuclei environment, generating important modifications in the cell physiology. Such alterations can be the key for a heterotic effect or deleterious effect, depending on the intensity in wich such alterations occurs.

Acknowledgement

This work was supported by grants from FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo (2013/04533-3), CAPES - Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (3127-14-1) and CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (474684/2013-0 and 151697/2011-6) and ProPe-UNESP (0045/021/13-PROPE/CDC).

References

As referências deste manuscrito encontram-se reunidas ao final deste volume no item “6. Referências Bibliográficas”.

Figures

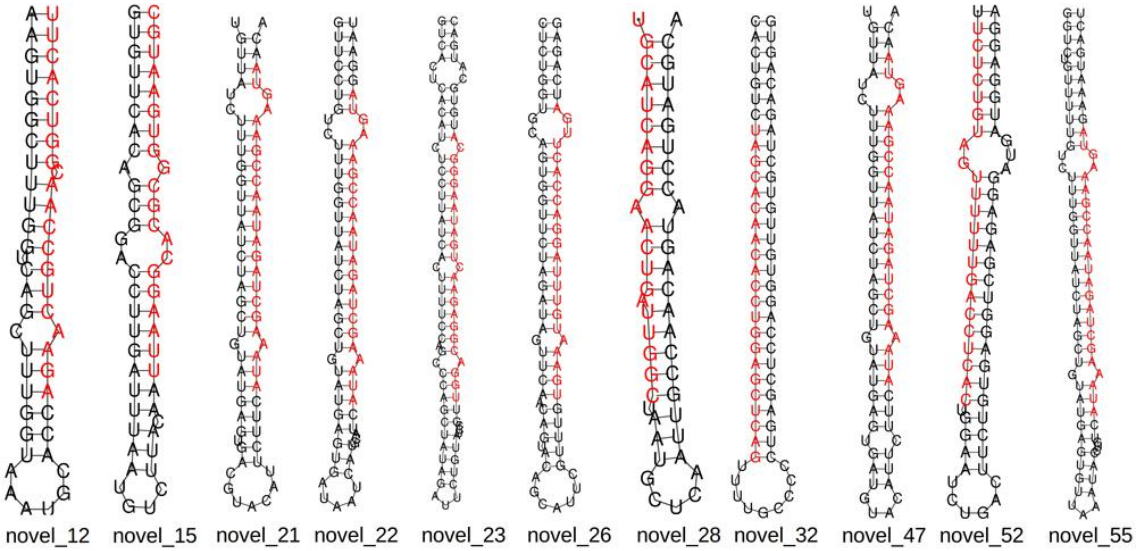


Figure 1: Precursor hairpin structures of the eleven differential expressed novel miRNAs predicted with miRCat software. Mature sequences appear in red.

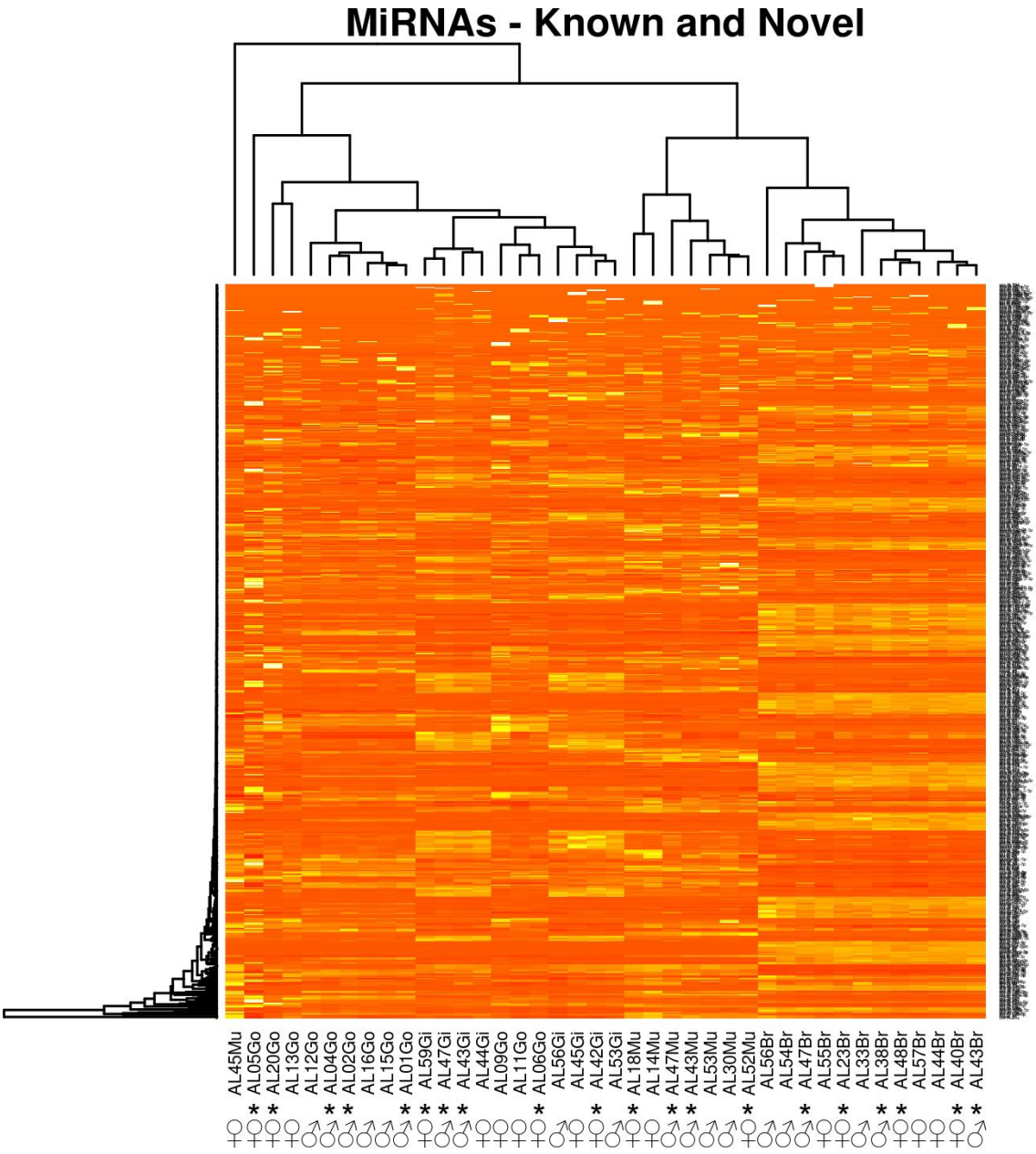


Figure 2: Clustered heatmap showing known and novel normalized expression values for micro-RNAs. MiRNAs identifications is on right. Samples identifications on bottom. Asterisks represents B positive samples. High and low expression levels are represented in dark and light colors respectively.

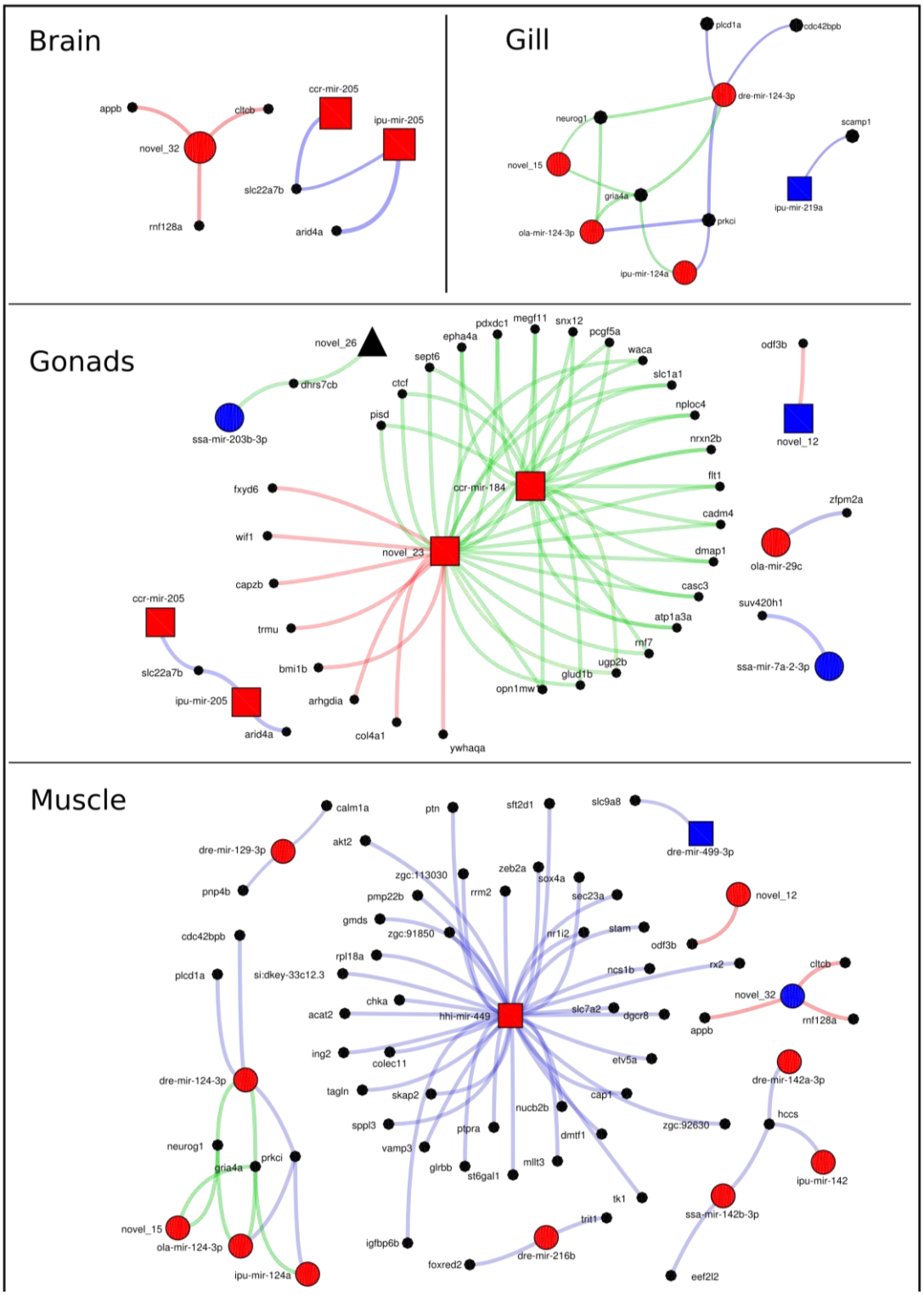
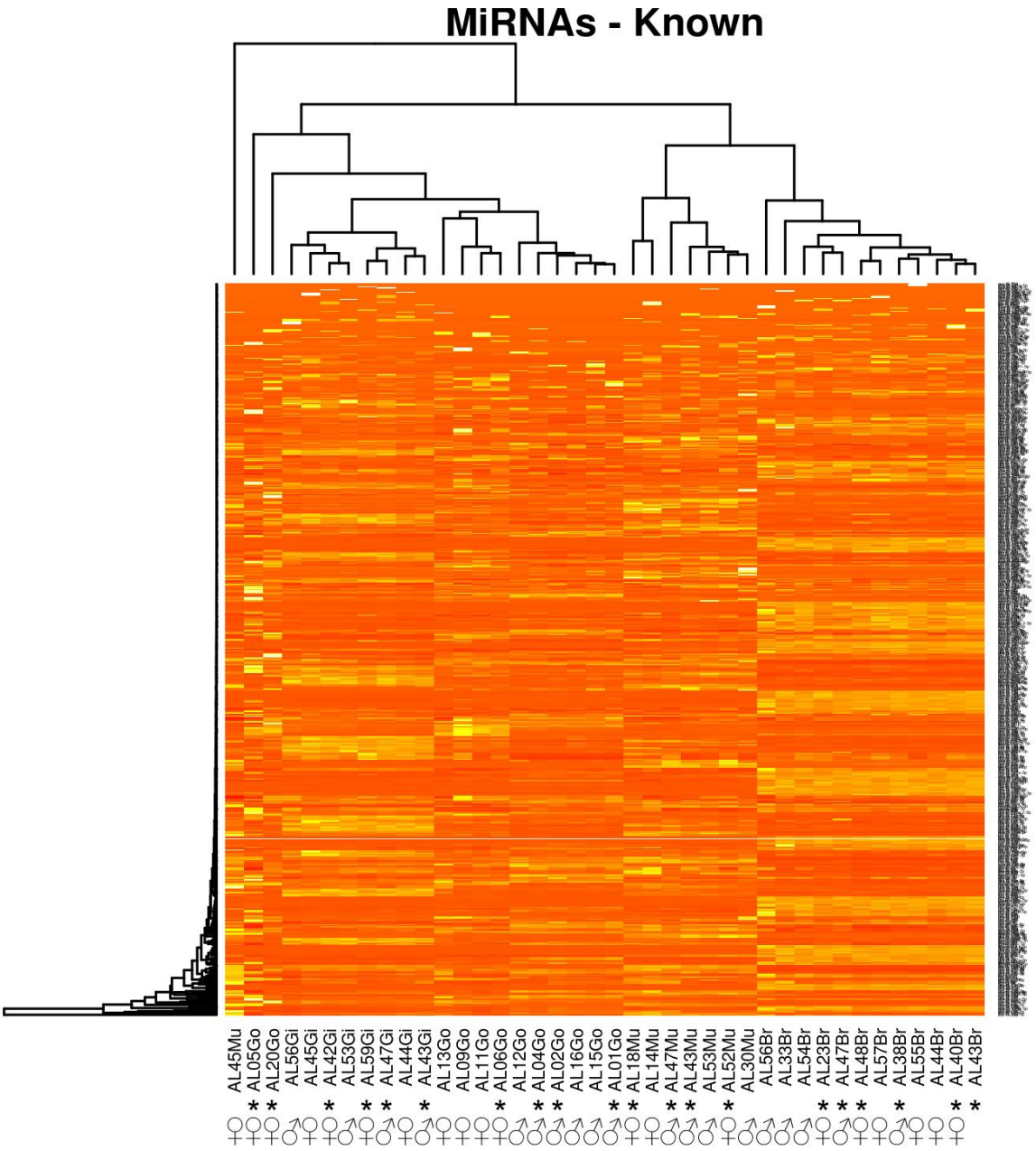


Figure 3: Interaction miRNA-gene networks for brain, gills, gonads and muscle tissues. Nodes representing up-regulated and down-regulated miRNAs appears in red and blue respectively. Edges linking genes targeted only by novel miRNAs, only by known miRNAs and by both novel and known miRNAs appears in light red, light green and light blue respectively. Male DE miRNAs are represented by square nodes, and female DE miRNAs are represented in elliptic nodes. Target genes are represented by small black nodes.

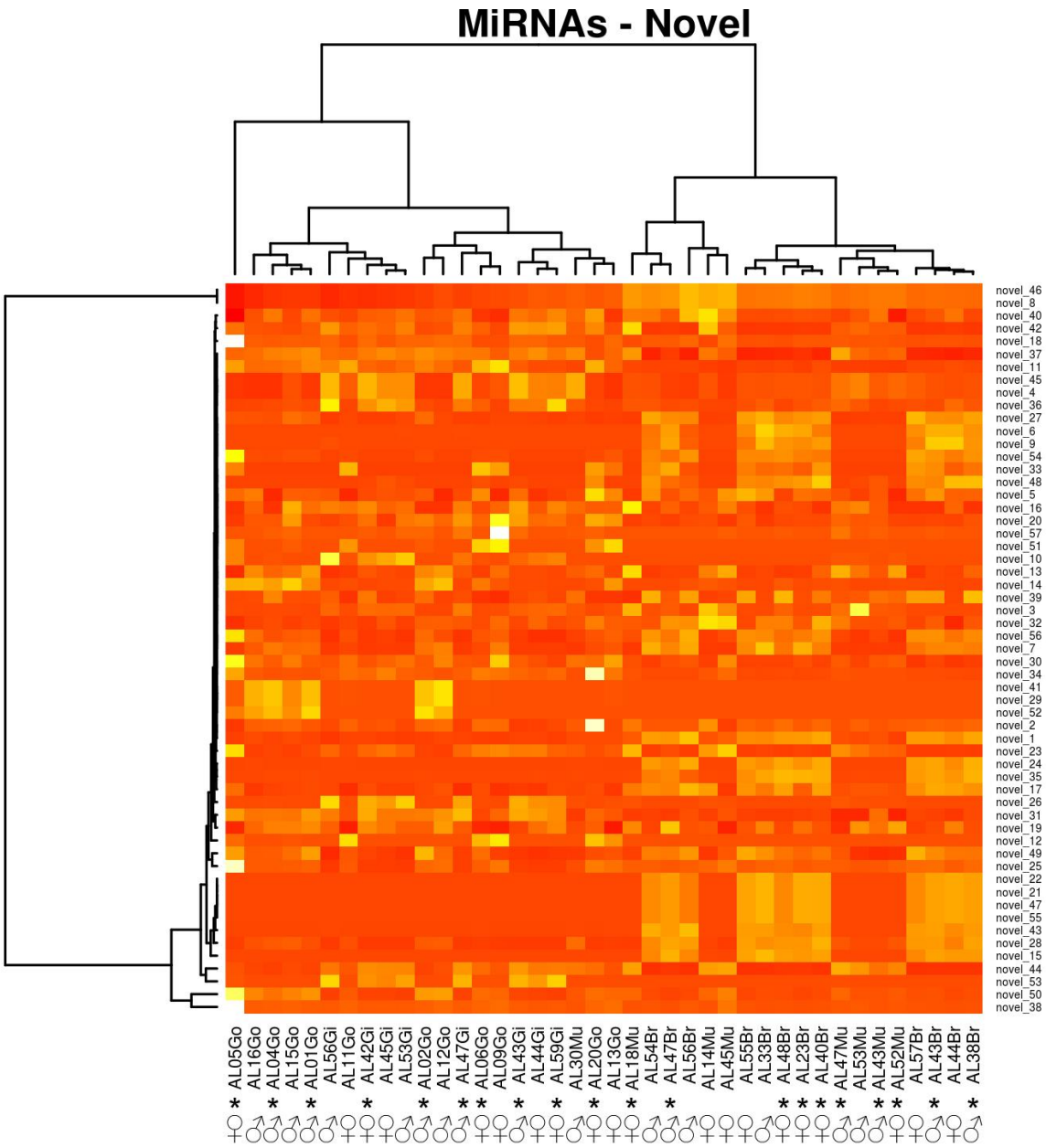
Supplementary Material

Suppementary Material 1: Number of obtained reads and read number after each processing step. B positive samples appear in bold.

Samples	Obtained reads (raw data)	Reads after adaptor removal	Reads after quality filtering	Reads after genome alignment	Remaining reads (%)
AL01Go	**12914266**	**12429551**	**11501605**	**9008708**	**69,8**
AL02Go	**13043032**	**12375133**	**11437819**	**9102762**	**69,8**
AL04Go	**19914188**	**17549736**	**16154832**	**12532511**	**62,9**
AL05Go	**16160535**	**9814219**	**9062238**	**4647236**	**28,8**
AL06Go	**19471118**	**16754087**	**15557321**	**9604944**	**49,3**
AL09Go	11075285	10651226	9884087	5686282	51,3
AL11Go	11342803	10373024	9658038	5455490	48,1
AL12Go	12462763	11647825	10799333	8339383	66,9
AL13Go	15910422	15288600	14231212	9177845	57,7
AL14Mu	18414513	14834624	12319480	11378385	61,8
AL15Go	15851784	15144108	14033920	10976130	69,2
AL16Go	12930943	12248309	11311252	8857725	68,5
AL18Mu	**17036677**	**13099407**	**11058891**	**10230574**	**60,1**
AL20Go	**11240729**	**10322303**	**9586565**	**5745210**	**51,1**
AL23Br	**12140843**	**11989245**	**11039358**	**10211500**	**84,1**
AL30Mu	11397967	9752202	8461409	6038528	53,0
AL33Br	10706434	10471940	9550192	8829328	82,5
AL38Br	**10846143**	**10346946**	**9096348**	**8038807**	**74,1**
AL40Br	**9735863**	**9549015**	**8410003**	**7768558**	**79,8**
AL42Gi	**16577859**	**15129149**	**14362123**	**12360518**	**74,6**
AL43Br	**10484927**	**10141617**	**8937496**	**8303556**	**79,2**
AL43Gi	**8706688**	**8546338**	**8101383**	**7202778**	**82,7**
AL43Mu	**14223489**	**12804701**	**11158251**	**9180433**	**64,5**
AL44Br	13104704	12372308	11173181	10458789	79,8
AL44Gi	8202282	8020653	7610759	6733199	82,1
AL45Gi	8507920	7646325	7242196	6305509	74,1
AL45Mu	12156497	10232019	9736377	9034366	74,3
AL47Br	**11533859**	**11173256**	**9784337**	**9219903**	**79,9**
AL47Gi	**19775982**	**18289510**	**17351433**	**15328378**	**77,5**
AL47Mu	**9143547**	**7718242**	**7304712**	**5966853**	**65,3**
AL48Br	**12729573**	**12307692**	**10854990**	**9855624**	**77,4**
AL52Mu	**13100116**	**9262141**	**8734038**	**5528066**	**42,2**
AL53Gi	10078089	8199145	7742808	5168626	51,3
AL53Mu	17339044	14671972	13865334	10597184	61,1
AL54Br	9290791	9094862	8003065	7502534	80,8
AL55Br	11303486	11111003	9789607	9021258	79,8
AL56Br	9459105	9296615	8191815	7716602	81,6
AL56Gi	7710362	7090918	6647036	3487571	45,2
AL57Br	10056094	9624693	8457112	7496459	74,5
AL59Gi	**9829273**	**9513420**	**9028079**	**7508024**	**76,4**



Supplementary Material 2: Clustered heatmap showing known micro-RNAs normalized expression values. MiRNAs identifications is on right. Samples identifications on bottom. Asterisks represents B positive samples. High and low expression levels are represented in dark and light colors respectively.



Supplementary Material 3: Clustered heatmap showing novel micro-RNAs normalized expression values. MiRNAs identifications is on right. Samples identifications on bottom. Asterisks represents B positive samples. High and low expression levels are represented in dark and light colors respectively.

5. CONCLUSÕES

O presente conjunto de dados mostra que tecnologias de sequenciamento de próxima geração possuem um grande potencial de contribuição para com o estudo dos cromossomos B, que representam um problema biológico extremamente importante, e em grande parte aguardando elucidação.

As atividades dos micro-RNAs diferencialmente expressos em relação à presença de cromossomos B em amostras de quatro tecidos diferentes, mostram que há uma alteração em relação a genes envolvidos com os processos de divisão celular, o que pode representar uma estratégia evolutiva pela qual o cromossomo B aumente seu grau de transmissão. Os dados também mostram que os efeitos trazidos pela presença dos cromossomos B são de certa forma dependentes das atividades do complemento A, assim como dos diferentes tecidos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Abramsson A, Kettunen P, Banote RK, Lott E, Li M, Arner A and Zetterberg H. The zebrafish amyloid precursor protein-b is required for motor neuron guidance and synapse formation. Developmental Biology, 381(2), 377–388. 2013.

Amato F, Seia M, Giordano S, Elce A, Zarrilli F, Castaldo G and Tomaiuolo R. Gene Mutation in MicroRNA Target Sites of CFTR Gene: A Novel Pathogenetic Mechanism in Cystic Fibrosis? PLoS ONE, 8(3), 4–9. 2013.

Andermann P, Ungos J and Raible DW. Neurogenin1 defines zebrafish cranial sensory ganglia precursors. Developmental Biology, 251(1), 45–58. 2002.

Anders S andHuber W.Differential expression analysis for sequence count data.Genome Biology, 11, pp. R106. 2010.

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R and Houben A. Formation and expression of pseudogenes on the B chromosome of rye. The Plant Cell.25(7), 2536-44. 2013.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 116(2), 281–97. 2004.

Basheva EA, Torgasheva AA, Sakaeva GR, Bidau C, Borodin PM. A- and B-chromosome pairing and recombination in male meiosis of the silver fox (*Vulpes vulpes* L., 1758, Carnivora, Canidae). Chromosome Research.18(6), 689-96. 2010.

Bel Y, Ferré J and Escriche B. Quantitative real-time PCR with SYBR Green detection to assess gene duplication in insects: study of gene dosage in *Drosophila melanogaster* (Diptera) and in *Ostrinia nubilalis* (Lepidoptera). BMC Research Notes, 4(1), 84. 2011.

Bertollo LAC, Takahashi CS, Moreira-Filho O. Citotaxonomic consideration on *Hoplias lacerdae* (Pisces, Erythrinidae). Braz J Genet 1:103-120. 1978.

Beukeboom LW: Bewildering Bs: An impression of the 1st B- Chromosome Conference. Heredity 73:328-336. 1994.

Bizuayehu TT, Babiak J, Norberg B, Fernandes JMO, Johansen SD and Babiak I. Sex-Biased miRNA Expression in Atlantic Halibut (*Hippoglossus hippoglossus*) Brain and Gonads. Sexual Development 6, 257-266. 2012.

Bizuayehu TT and Babiak I. MicroRNA in Teleost Fish. Genome Biology and Evolution. 6(8), 1911–1937. 2014.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, … Di Palma F. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513, 375–381. 2014.

Cabral-de-Mello DC, Valente GT, Nakajima RT and Martins C. Genomic organization and comparative chromosome mapping of the U1 snRNA gene in cichlid fish, with an emphasis in *Oreochromis niloticus*. Chromosome Research, 20(2), 279–92. 2012.

Camacho J, Sharbel T, Beukeboom l: B-chromosome evolution. Evolution. 163-178. 2000.

Camacho JPM, Bakkali M, Corral JM, Cabrero J, López-León MD, Aranda I, Martín-Alganza A, Perfectti F. Host recombination is dependent on the degree of parasitism. Proc R Soc Lond B 269: 2173–2177. 2002.

Camacho JPM. B chromosomes. In: Gregory TR, ed. The Evolution of the Genome, San Diego: Elsevier, pp. 223-286. 2005.

Chen AK, Sengupta P, Waki K, Van Engelenburg SB, Ochiya T, Ablan SD, et al. MicroRNA binding to the HIV-1 Gag protein inhibits Gag assembly and virus production. Proceedings of the National Academy of Sciences of the United States of America, 111(26), E2676–83. 2014.

Enright AJ, John B, Gaul U, Tuschl T, Sander C and Marks DS.MicroRNA targets in *Drosophila*.Genome Biology 2003, 5:R1. 2004.

Fantinatti BEA, Mazzuchelli J, Valente GT, Cabral-de-Mello DC and Martins C. Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata*. Genetica, 139(10), 1273–82. 2011.

Feldberg E, Porto JIR, Bertollo LAC. Chromosomal changes and adaptation of cichlid fishes during evolution. In Fish adaptations Edited by: Val AL, Kapoor BG. New Dehli and New York: Science Publishers 285-308. 2003.

Feldberg E, Porto JIR, Alves-Brinn MN, Mendonça MNC, Benzaquem DC. B chromosomes in Amazonian cichlid species. Cytogenet Genome Res 106:195-198. 2004.

Gall J and Pardue M. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proceedings of the National Academy of Sciences, (1), 378–383. 1969.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, et al. MicroRNAs regulate brain morphogenesis in zebrafish. Science 308(5723), 833–8. 2005.

Goodwin SB, M'barek SB, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, … Kema GHJ. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. PLoS Genetics, 7(6), e1002070. 2011.

Greilhuber J and Speta F. C-banded karyotypes in the Scilla hohenackeri group, S. persica and Puschkinia (Liliaceae), Plant Syst. Evol. 126 149–188. 1976.

Houben A, Banaei-Moghaddam AM, Klemme S and Timmis JN. Evolution and biology of supernumerary B chromosomes. Cellular and Molecular Life Sciences : CMLS, 71(3), 467–78. 2014.

Jonas S and Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing.Nature Reviews Genetics, 16(7), 421-433. 2015.

Jones R. Tansley review no. 85.B chromosomes in plants. New Phytologist, 131(4), 411–563. 1995.

Jones RN, González-Sánchez M, González-García M, Vega JM, Puertas MJ. Chromosomes with a life of their own. Cytogenet Genome Res. 120:265-280. 2008.

Kapsimali M, Kloosterman WP, de Bruijn E, Rosa F, Plasterk RHA and Wilson SW.. MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. Genome Biol. 8:R173. 2007.

Kasahara S, Yonenaga-Yassuda Y. Chromosomal Variability in *Akodon* sp. (Rodentia, Cricetidae). Cytologia 47:317-324. 1982.

Kertesz M, Iovino N, Unnerstall U, Gaul U and Segal E. The role of site accessibility in microRNA target recognition.Nature Genetics, 39(10), 1278–84. 2007.

Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. Nature v5, p288-298. 2004.

Kohn M, Högel J, Vogel W, Minich P, Kehrer-Sawatzki H, Graves JM and Hameister H. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. Trends in Genetics: TIG, 22(4), 203–10. 2006.

Korenberg JR. Human Genome Anatomy: BACs Integrating the Genetic and Cytogenetic Maps for Bridging Genome and Biomedicine. Genome Research, 9(10), 994–1001. 1999.

Kumar MG, Patel NM, Nicholson AM, Kalen AL, Sarsour EH and Goswami PC. Reactive oxygen species mediate microRNA-302 regulation of AT-rich interacting domain 4a and C-C motif ligand 5 expression during transitions between quiescence and proliferation. Free Radical Biology and Medicine, 53(4), 974–982. 2012.

Krüger J and Rehmsmeier M. RNAhybrid: microRNA target prediction easy,fast and flexible. Nucleic Acids Research 34, 451-454. 2006.

Kuroiwa A, Terai Y, Kobayashi N, Yoshida K, Suzuki M, Nakanishi A, Matsuda Y, Watanabe M, Okada N. Construction of chromosome markers from the Lake Victoria cichlid *Paralabidochromis chilotes* and their application to comparative mapping. Cytogenet Genome Res. 142:112–120. 2014.

Langmead B, Trapnell C, Pop M and Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009.

Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2.Nature Methods 9, 357–359. 2012.

Lewis BP, Shih I, Jones-Rhoades NW, Bartel DP and Burge CB. Prediction of mammalian microRNA targets.Cell, 115(7), 787–98. 2003.

Maere S, Heymans K and Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics, 21(16), 3448–3449. 2005.

Maina JN. The highly specialized secretory epithelium in the buccal cavity of the alkalinity adapted Lake Magadi cichlid, *Oreochromis alcalicus grahami* (Teleostei: Cichlidae): a scanning and transmission electron microscope study. J Zool Lond 251:427-438. 2000.

Maistro EL, Foresti F, Oliveira C, Almeida-Toledo LF. Occurence of macro B chromosomes in *Astyanax scabripinnis* paranae (Pisces, Characiformes, Characidae).Genetica 87:101-106. 1992.

Martis MM, Klemme S, Banaei-Moghaddam AM, et al. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. Proceedings of the National Academy of Sciences of the United States of America, 109(33), 13343–6. 2012.

Mazzuchelli J, Martins C. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. Genetica 136:461-469. 2009.

Mcallister BF, Werren JH. Hybrid origin of a B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis*. Chromosoma 243-253. 1997.

McDaneld TG, Smith TPL, Doumit ME, Miles JR, Coutinho LL, Sonstegard TS, … Wiedmann RT. MicroRNA transcriptome profiles during swine skeletal muscle development. BMC Genomics, 10, 77. 2009.

Moreira-Filho O, Galetti Jr. PM, Bertollo LAC. B chromosomes in the fish *Astyanax scabripinnis* (Characidae, Tetragonopterinae): An overview in natural populations. Cytogenet Gen Res 106:230-234. 2004.

Morris KV and Mattick JS.The rise of regulatory RNA. Nature Reviews. Genetics, 15(6), 423–37. 2014.

Nachtigall P, Dias M and Pinhal D. Evolution and genomic organization of muscle microRNAs in fish genomes.BMC Evolutionary Biology, 14(1), 196. 2014.

Onoguchi M, Hirabayashi Y, Koseki H and Gotoh Y. A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development. Proceedings of the National Academy of Sciences, 109(42), 16939–16944. 2012.

Pardue ML and Gall JG. Molecular Hybridization of Radioactive DNA To the DNA of Cytological Preparations.Proceedings of the National Academy of Sciences.64(2), 600-604. 1969.

Perfectti F and Werren J. The interspecific origin of B chromosomes: experimental evidence. Evolution, 55(5), 1069–1073. 2001.

Pfennig DW and Ehrenreich IM.Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation.Molecular Ecology, 23(18), 4438–40. 2014.

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J and Gray J. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proceedings of the National Academy of Sciences of the United States of America, 85(23), 9138–42. 1988.

Pinkel D, Straume T and Gray J. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proceedings of the National Academy of Sciences of the United States of America, 83(May), 2934–2938. 1986.

Poletto AB, Ferreira IA, Martins C. The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. BMC genetics, 11, 1.2010.

Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, … Rokhsar DS. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317(5834), 86–94. 2007.

Rinn J and Guttman M. RNA and dynamic nuclear organization. Science 345, 1240. 2014.

Roberts RK and Appel B. Apical polarity protein PrkCi is necessary for maintenance of spinal cord precursors in zebrafish. Developmental Dynamics, 238(7), 1638–1648. 2009.

Sapre AB, Deshpande DS.Origin of B chromosomes in *Coix* L. through spontaneous interspecific hybridization. J. Hered. 78, 191-196. 1987.

Schartl M, Nanda I, Schlupp I, Wilde B, Epplen JT, Schmidt M, Parzefall J. Incorporation of subgenomic amounts of DNA as compensation for mutational load in a gynogenetic fish. Nature 373, 68-7 1. 1995.

Schmid M, Ziegler CG, Steinlein C, Nanda I, Schartl M. Cytogenetics of the bleak (*Alburnus alburnus*), with special emphasis on the B chromosomes. Chromosome Research 14:231–242. 2006.

Schneider RF, Li Y, Meyer A and Gunter HM. Regulatory gene networks that shape the development of adaptive phenotypic plasticity in a cichlid fish. Molecular Ecology.23, 4511–4526. 2014.

Smith WL, Chakrabarty P, Sparks JS.Phylogeny, taxonomy, and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae).Cladistics 24:625-641. 2008.

Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L, Schwach F, Tamas D and Moulton V. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics (Oxford, England), 28(15), 2059–61.2012.

Szczerbal I, Switonski M. B chromosomes of the Chinese raccoon dog (*Nyctereutes procyonoides* Gray) contain inactive NOR-like sequences. Caryologia 56 2:213-216. 2003.

Tanić N, Vujošević M, Dedović-Tanić N, Dimitrijević B.Differential gene expression in yellow-necked mice *Apodemus flavicollis* (Rodentia, Mammalia) with and without B chromosomes.Chromosoma, 133(8), 418-427. 2005.

Teixeira WG, Ferreira IA, Cabral-de-Mello DC, Mazzuchelli J, Valente GT, Poletto AB, Venere PC, Martins C. Organization of repeated DNA elements

in the genome of the cichlid fish *Cichla kelberi* and its contributions to the knowledge of fish genomes. Cytogenet Genome Res. 125: 224-234. 2009.

Teruel M, Sørensen JG, Loeschcke V, Cabrero J, Perfectti F and Camacho JPM. Level of heat shock proteins decreases in individuals carrying B-chromosomes in the grasshopper *Eyprepocnemis plorans*. Cytogenetic and Genome Research, 132(1-2), 94–9. 2011.

Trifonov VA, Perelman PL, Kawada S-I, Iwasa MA, Oda S-I, Graphodatsky. Complex structure of B-chromosomes in two mammalian species: *Apodemus penisulae* (Rodentia) and *Nyctereutes procyonoides* (Carnivora). Chromosome Research 10:109-116. 2002.

Uhalte E, Kirchner M, Hellwig N, Allen JJ, Donat S, Shokat KM, … Abdelilah-Seyfried S. In vivo conditions to identify Prkci phosphorylation targets using the analog-sensitive kinase method in zebrafish. PLoS ONE, 7(6), 2–7. 2012.

Valente GT, Mazzuchelli J, Ferreira IA, Poletto AB, Fantinatti BEA, Martins C. Cytogenetic Mapping of the Retroelements *Rex1*, *Rex3* and *Rex6* among Cichlid Fish: New Insights on the Chromosomal Distribution of Transposable Elements. Cytogenetic and Genome Research. 133(1), 34-42.2011.

Valente GT, Conte M, Fantinatti BEA, Cabral-de-Mello DC, Carvalho RF, Vicari MR, Kocher T, Martins C. Origin and Evolution of B Chromosomes in the Cichlid Fish *Astatotilapia latifasciata* Based on Integrated Genomic Analyses. Molecular Biology and Evolution. 31(8), 2061-72. 2014.

Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R and Bhattacharya A. Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. BMC Genomics, 11, 288. 2010.

Venere PC, Miyazawa C, Galetti PM. New cases of supernumerary chromosomes in characiform fishes. Genet Mol Biol 22:345-349. 1999.

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

Witkos TM, Koscianska E and Krzyzosiak WJ. Practical Aspects of microRNA Target Prediction.Current Molecular Medicine, 11(2), 93–109. 2011.

Wu MY, Eldin KW and Beaudet AL. Identification of chromatin remodeling genes Arid4a and Arid4b as leukemia suppressor genes. Journal of the National Cancer Institute, 100(17), 1247–1259. 2008.

Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe, et al. B chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. PLoS Genetics, 7(8), e1002203. 2011.

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

Zima J. Piálek J. and Macholán M. Possible heterotic effects of B chromosomes on body mass in a population of *Apodemus flavicollis*.Canadian Journal of Zoology, 81(8), 1312-1317.2003.

Zhou Q, Zhu H, Huang Q, Zhao L, Zhang G, Roy SW, … Bachtrog D. Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. BMC Genomics, 13(1), 109. 2012.