UNIVERSIDADE ESTATUAL PAULISTA “JÚLIO DE MESQUISTA FILHO”
INSTITUTO DE BIOCIÊNCIAS DE BOTUCATU
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA)

JORDANA INÁCIO NASCIMENTO-OLIVEIRA

CARACTERIZAÇÃO DE PEQUENOS RNAS NÃO-CODIFICANTES (microRNAs E PIWI-INTERACTING RNAs) NA PRESENÇA DO CROMOSSOMO B DO CICLÍDEO AFRICANO *ASTATOTILAPIA LATIFASCIATA*

Botucatu, SP
2022

Jordana Inácio Nascimento-Oliveira

CARACTERIZAÇÃO DE PEQUENOS RNAS NÃO-CODIFICANTES (microRNAs E PIWI-INTERACTING RNAs) NA PRESENÇA DO CROMOSSOMO B DO CICLÍDEO AFRICANO *ASTATOTILAPIA LATIFASCIATA*

Tese submetida ao Programa de Pós-graduação em Ciências Biológicas (Genética) do Instituto de Biociências de Botucatu da Universidade Estadual Paulista "Júlio de Mesquita Filho" para obtenção do título de Doutora em Ciências Biológicas (Genética).
Orientador: Prof. Dr. Cesar Martins

Botucatu, SP
2022

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Nascimento-Oliveira, Jordana Inácio.
Caracterização de pequenos RNAs não-codificantes (microRNAs E PIWI-interacting RNAs) na presença do cromossomo B do ciclídeo africano *Astatotilapia latifasciata* / Jordana Inácio Nascimento-Oliveira. - Botucatu, 2022

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Cesar Martins
Capes: 20204000

1. Publicações de divulgação científica. 2. Genoma. 3. Transcriptoma. 4. RNA-seq. 5. Elementos de DNA transponíveis.

Palavras-chave: Divulgação científica; Elementos transponíveis; Genoma; RNAseq; Transcriptoma.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE JORDANA INÁCIO NASCIMENTO OLIVEIRA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA), DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 18 dias do mês de fevereiro do ano de 2022, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de TESE DE DOUTORADO de JORDANA INÁCIO NASCIMENTO OLIVEIRA, intitulada **Caracterização de pequenos RNAs não-codificantes (microRNAs e PIWI-interacting RNAs) na presença do cromossomo B do ciclídeo africano *Astatotilapia latifasciata***. A Comissão Examinadora foi constituida pelos seguintes membros: Prof. Dr. CESAR MARTINS (Orientador(a) - Participação Virtual) do(a) Departamento de Biologia Estrutural e Funcional / Instituto de Biociencias de Botucatu , Prof. Dr. FREDERICO HENNING (Participação Virtual) do(a) Universidade Federal do Rio de Janeiro / UFRJ, Prof. Dr. ANDRE LUIS LAFORGA VANZELA (Participação Virtual) do(a) Universidade Estadual de Londrina - UEL, Prof. Dr. DIOGO CAVALCANTI CABRAL DE MELLO (Participação Virtual) do(a) Instituto de Biociências de Rio Claro - UNESP, Profa. Dra. CLAUDIA APARECIDA RAINHO (Participação Virtual) do(a) Departamento de Ciencias Quimicas e Biologicas / Instituto de Biociencias de Botucatu - UNESP. Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: APROVADA. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. CESAR MARTINS

Agradecimentos

Inspirada pelas aulas de Geografia e Ciências da quinta série e filmes de ficção científica, tenho sonho em ser cientista desde criança. O doutorado é o último passo para a formação antes de entrarmos no mercado de trabalho como pesquisadores. Nesses 10 anos, desde meu ingresso no Curso de Ciências Biológicas na UEPG, tive o privilégio da companhia de diversas pessoas que não poderia deixar de destacar meus mais profundos agradecimentos:

À minha família, por todo apoio, formação pessoal e amor. Mamis Darci, Pai Vilmar, Paidrasto JC, mano Rodolfo, Dinha Mundica, Vó Argentina e tia Dora, os dias longe de vocês não foram e não são fáceis. O bônus de estar trilhando meu caminho longe de casa nunca substituirá o ônus e vazio da distância. Mas carrego vocês no meu coração e sou extremamente grata por ter recebido tanto. Ao meu marido, Paulo, que me traz paz e ao mesmo tempo topa minhas loucuras, mas também não me deixa esquecer da realidade. “Sem amor, eu nada seria”, como dizia Renato Russo. Obrigada por sempre me apoiar e vamos juntos nessa embarcar em mais uma aventura.

Aos meus amigos, a vida é melhor com eles. Que preencheram o vazio da distância da família e tem um lugar especial no meu coração. Adauto, Poli, “Buna”, Camila, Ivan e Bia, obrigada pelas parcerias acadêmicas e alcoólicas. Em especial agradeço minha eterna “miga sua, louca”, Najila, por ser a irmã que eu não tive. A vida em Botucatu foi bem mais fácil em sua companhia.

Aos meus professores, Cesar, Marcelo, Adriane e Danillo. Gostaria de agradecer por todo apoio, por acreditarem em mim e por todas as oportunidades durante essa trajetória. De forma geral gostaria de externalizar minha mais profunda admiração a todos os professores. Foram professores que me inspiraram a chegar onde eu estou. São vocês que tornam o mundo um lugar mais esperançoso. Espero poder retornar a sociedade a dedicação que tiveram comigo.

À UNESP e ao Instituto de Biociências de Botucatu, em tom de saudosismo quero agradecer imensamente a instituição que foi minha casa durante os últimos 6 anos. Ter o privilégio de ser formada por uma instituição de excelência é o sonho de

qualquer criança que almeja ser cientista. Na UNESP foi possível colocar em prática toda a minha vontade por ensinar e fazer pesquisa.

À FAPESP, pelo investimento no meu doutorado (processo 2017/25193-7). Essa fundação exemplo para o Brasil torna possível o exercício da Ciência e colabora com o desenvolvimento científico do nosso país através do investimento em ideias que se executam na forma de projetos. Vida longa à FAPESP, por mais investimento em Ciência e pela autonomia universitária, muito obrigada.

A todos, que me permitiram voar alto.

"Contra o pessimismo da razão, o otimismo da prática."

"Deve-se convencer muita gente de que o estudo é também um trabalho, e muito cansativo, com um tirocínio particular próprio, não só intelectual, mas também muscular-nervoso: é um processo de adaptação, é um hábito adquirido com esforço, aborrecimento e até mesmo sofrimento."

— Antônio Gramsci

Resumo

Os supranumerários cromossomos B são elementos extras ao genoma encontrados em diversas espécies de eucariotos, desde fungos, plantas a animais. A origem independente nos mais diversos organismos também reflete a sua variedade quanto a número, morfologia e constituição. Desta forma, sabe-se que os cromossomos B são constituídos por DNA repetitivo, genes codificadores de proteína e genes que transcrevem RNAs não-codificantes (ncRNA). Os ncRNAs participam de uma infinidade de processos biológicos, desde os fundamentais como transcrição e tradução, até regulação durante o desenvolvimento embrionário e ciclo celular. Desta forma, conhecendo a baixa densidade de genes codificadores de proteínas nos cromossomos B, o presente estudo focou em caracterizar os ncRNAs, ainda pouco explorados e que trazem perspectivas interessantes no entendimento da origem e comportamento desse elemento extra. Foi utilizado como modelo de estudo o peixe *Astatotilapia latifasciata,* pertencente à família Cichlidae, modelo em estudos evolutivos. Dentre uma infinidade de classes de ncRNA optou-se por explorar a partir do RNAseq de *A. latifascita* os pequenos RNAs não-codificantes: microRNAs (miRNA) e PIWI-interacting RNAs (piRNA). Foram encontrados diversos miRNAs diferencialmente expressos em amostras com cromossomo B. Esses miRNAs tem como alvos diversos genes codificadores de proteínas e propõe-se que o cromossomo B, a partir da atuação dos miRNAs impacta em diversos processos biológicos. Adicionalmente, c*lusters* de piRNAs presentes no cromossomo B foram detalhadamente descritos indicando o envolvimento deste elemento extra em arquitetar o silenciamento de elementos transponíveis. Complementarmente, essa tese adicionou dois objetivos suplementares. Uma revisão sobre as pesquisas de RNAs e cromossomos B é apresentada e fornece uma linha do tempo e principais resultados que foram reunidos em uma árvore filogenética representativa dos organismos com cromossomo B e estudos relacionando RNAs. Outrossim, essa tese encerra com reflexões sobre divulgação científica, relatando alguns dos trabalhos publicados ao longo do período e propondo diretrizes para a expansão da popularização da ciência e educação científica no nosso país.

Palavras-chave: genoma, evolução, RNAseq, transcriptoma, elementos transponíveis, divulgação científica

Abstract

The supernumerary B chromosomes are extra elements to the genome found in several eukaryotes species, from fungi, plants to animals. The polyphyletic origin in the most diverse organisms also reflects their variety in terms of number, morphology and constitution. Thus, it is known that the B chromosomes are composed by repetitive DNA, protein coding genes and genes that transcribed non-coding RNAs (ncRNAs). ncRNAs participate in several biological processes, from fundamental ones such as transcription and translation, to regulation during embryonic development and cell cycle. Thus, taking into account the low density of protein-coding genes in the B chromosomes, the present study focused on characterizing the ncRNAs, that are poorly explored and can bring interesting perspectives to understand the origin and behavior of this extra element. The African cichlid *Astatotilapia latifasciata* fish, a family model in evolutionary studies, was used as a study model. Among a multitude of ncRNA classes, it was decided to explore from the RNAseq of *A. latifascita* the small non-coding RNAs: microRNAs (miRNA) and PIWI-interacting RNAs (piRNA). Several miRNAs were found differentially expressed in samples with B chromosome. These miRNAs target several protein-coding genes and it is proposed that the B chromosome, under the action of miRNAs, impacts on several biological processes. Additionally, piRNA clusters present on the B chromosome were described in detail, indicating the involvement of this extra element in architecting the silencing of transposable elements. Completely, this thesis added two supplementary objectives. A review of the research on RNAs and B chromosomes is presented and provides a timeline and main results that have been brought together in a representative phylogenetic tree of organisms with B chromosome and RNA studies. Furthermore, this thesis ends with reflections on science outreach, reporting some of the works published throughout the period and proposing guidelines for the expansion of the popularization of science and scientific education in our country.

Keywords: genome, Evolution, RNAseq, transcriptome, transposable elements, science outreach

Resumo para a sociedade

Você já parou para pensar que todos os seres vivos possuem DNA? Com exceção de alguns vírus, é fascinante imaginar que a mesma estrutura, o DNA, guarda as instruções que formarão um indivíduo. Nos seres que possuem núcleo delimitado, os eucariotos, o DNA se organiza empacotado com proteínas, formando os cromossomos. A variedade dos organismos também pode ser vista na variedade no número e morfologia (formato/tamanho) dos cromossomos que as espécies carregam. No meio dessa diversidade, temos alguns organismos que possuem um cromossomo extra, são os cromossomos B. Esses cromossomos extras geralmente se originam durante erros no pareamento dos cromossomos normais durante a divisão das células que formam os gametas (óvulos e espermatozoides, no caso dos animais). Esses pareamentos anormais podem provocar quebras nos cromossomos, dando origem a um novo cromossomo, que será o cromossomo B. Os cientistas já sabem que os cromossomos B carregam cópias de sequências de DNA dos cromossomos normais da espécie. Essas sequências podem ser genes, que são responsáveis por alguma função, como carregar o código para a formação de proteínas ou gerar um RNA que regula processos dentro da célula. Mas qual o impacto de um cromossomo a mais para as células? Como esse cromossomo se comporta? São perguntas ainda não respondidas e muito importantes para se entender o comportamento e função de cromossomos extras. Desta forma, esse trabalho de doutorado procurou investigar genes que pudessem contar mais sobre a função dos cromossomos B. Dentro de uma infinidade de classes de genes, foram escolhidos os “pequenos RNAs”, que são responsáveis por regular diversos processos nas células. O nosso modelo de estudo foi um peixe originário da África, chamado *Astatotilapia latifasciata*. A partir do sequenciamento de DNA e RNA desse peixe, de indivíduos com e sem cromossomo B, diversas análises foram realizadas no computador, numa área chamada bioinformática. Analisando o RNA das células de indivíduos com e sem cromossomo B nós identificamos dois resultados principais. (1) Descobrimos que o cromossomo B altera a quantidade (expressão) de um tipo de RNA, chamado microRNA, e isso pode impactar em diversos processos biológicos que esses microRNAs controlam. Outro resultado é que o cromossomo B tem sequências de outros pequenos RNAs, chamados piRNAs. Esses RNAs, por sua vez, protegem o conjunto de DNA dos organismos, chamado genoma e por isso são chamados de “guardiões do genoma”. Com isso, apresentamos também uma árvore da vida do cromossomo B e RNA: colocamos os principais grupos de organismos e indicamos quais deles já foram descobertos cromossomos B e quais possuem estudos de RNAs e cromossomos B. Foi mostrado que diversas classes de RNAs ainda não foram estudadas em relação ao

cromossomo B. Concluímos com esses estudos que o cromossomo B pode interferir em diversas vias que ainda não tinham sido descritas e podem direcionar novos estudos focando nessas vias regulatórias para melhor descrever as funções desses cromossomos extras.

Esse resumo foi revisado por familiares, não por pares. Agradeço os comentários e reflexões dos “revisores” que me fizeram analisar diversos pontos a melhorar a comunicação do meu trabalho ao grande público.

Lista de figuras

1 Introdução

3.1 Capítulo 1 - *Differential expression of miRNAs in the presence of B chromosome in the cichlid fish Astatotilapia latifasciata*

3.2 Capítulo 2 - *First characterization of PIWI-interacting RNA clusters and their role in genome protection against the accumulation of transposable elements in B chromosome*

3.3 Capítulo 3 - *B chromosome tree of life: an RNA world perspective*

3.4 Capítulo 4 - Relatos sobre divulgação científica: democratizar o conhecimento é construir um futuro melhor

Lista de tabelas

Sumário

1 Introdução

1.1 Cromossomo B

Os supranumerários cromossomos B ou simplesmente “Bs” são elementos extra e dispensáveis ao genoma, encontrados em aproximadamente 10 a 15% dos organismos já cariotipados, incluindo fungos, plantas e animais (Camacho et al., 2000; Houben et al., 2014; Jones et al., 2008; Jones, 1991; Miao et al., 1991; Peng & Cheng, 2011). Dentre esses grupos, plantas é o táxon com maior número de espécies descritas portadoras de cromossomos B (Peng & Cheng, 2011). Enquanto em animais, os Bs foram reportados em diversos invertebrados, em algumas espécies de mamíferos, mais de 100 espécies de peixes e apenas uma espécie de ave (Jones et al., 2008).

O número de espécies com registro de cromossomo B em cada táxon não deve necessariamente ser relacionado a biologia do grupo, pois, em determinados casos há uma escassez de informações devido à dificuldade de obtenção de amostras ou falta de estudos, como é o caso das aves (Camacho et al., 2000; D’Ambrosio et al., 2017). No filo Cnidária, por exemplo, ainda não foram encontradas espécies com cromossomo B, e entre as plantas, algas, samambaias e musgos possuem poucos registros representantes portadores de cromossomo supranumerário (D’Ambrosio et al., 2017). Mesmo sabendo que alguns grupos tendem a uma maior tolerância a eventos de ploidias, como é o caso das plantas, a prevalência de cromossomo B está associada a um evento polifilético e pode ser encontrada ao longo da árvore da vida (Valente et al., 2017).

O primeiro cromossomo supranumerário foi descrito em 1907 em algumas espécies do gênero *Metapodius* (Wilson, 1907). Esse cromossomo extra não pertencia nem ao conjunto de cromossomos sexuais (X, Y, Z ou W) tampouco ao conjunto dos autossomos, o complemento “A”. Desta forma, tornou-se convenção nomear esse elemento extra de cromossomo “B”. No geral, são fortemente heterocromáticos o que indica pouca atividade gênica (Camacho et al., 2000) e raramente a presença do B está associada a alguma alteração fenotípica (Valente et al., 2017).

A evolução e origem do cromossomo B em diversos *taxa* revela uma história independente, ou seja, varia em constituição, número e morfologia para cada espécie e até mesmo entre populações (Ploskaya-Chaibi et al., 2015) (**Figura 1**). Cromossomos B podem apresentar origem intraespecífica autossômica, quando a evolução é dada por eventos na própria espécie, ou interespecífica autossômica, quando cruzamentos interespecíficos originam esse elemento extra, mecanismo comumente encontrado em plantas (Camacho et al., 2000; Houben, 2017).

É consenso que duas forças contrárias atuem estabelecendo a frequência de cromossomos B em uma população: a variabilidade polimórfica que o cromossomo B pode trazer a uma população, aumentando a frequência; e os efeitos negativos ao *fitness* que o B pode causar, diminuindo a frequência (Camacho et al., 2000). O número de Bs em cada espécie também é variável (**Figura 1**). Na literatura é possível encontrar estudos com indivíduos que suportam diversos cromossomos B, como no caso da raposa vermelha (*Vulpes vulpes*, 2n=34 As + 0-8 Bs) e em milho (*Zea mays*, 2n=20 As + 0-34 Bs) (revisado por (Houben et al., 2014)) e ainda parece haver uma proporcionalidade do número de Bs suportado em cada espécie, um limite que se ultrapassado pode afetar negativamente o indivíduo, causando infertilidade, por exemplo (Houben, 2017).

Além disso, os Bs normalmente não se pareiam com nenhum outro cromossomo do complemento “A”. Sendo assim, é necessário que os cromossomos B ao longo da evolução adquiram um mecanismo de *drive* meiótico, ou seja, estratégias para que durante a não-disjunção meiótica não sejam eliminados (Houben, 2017). Consequência no *drive* meiótico é uma segregação não-mendeliana em que muitas vezes o cromossomo B é capaz de ser transmitido a frequências superiores a 50% (Benetta et al., 2019; Imarazene et al., 2021).

O comportamento do cromossomo B durante a meiose tem sido extensivamente investigado, e já se conhecem diversas formas de segregação não-mendeliana (Clark et al., 2017; Houben, 2017). Em plantas, por exemplo, foi observado uma não-disjunção direcionada, orquestrada por uma sequência não-codificante (Houben, 2017) e em peixes uma segregação preferencial ainda sem mecanismos moleculares descritos (Clark et al., 2017). Esses fenômenos foram reportados durante a meiose de fêmeas, em que o cromossomo B migra preferencialmente para o pólo germinativo, evitando o pólo vegetativo no caso das

plantas, ou para o oócito, evitando o corpúsculo polar no caso dos animais. Em espécies de ciclídeos africanos, por exemplo, o cromossomo B já foi reportado exclusivamente em fêmeas (Clark et al., 2017; Yoshida et al., 2011). Por sua vez, em *Astyanax mexicanus*, o cromossomo B tem uma prevalência maior em machos e possui uma cópia do gene *growth differentiation factor 6b* (*gdf6b*). O silenciamento dessa sequência através de CRISPR provocou a reversão sexual de machos para fêmeas, sugerindo uma possível função para gene determinante do sexo carregado pelo cromossomo B (Imarazene et al., 2021).

Essas evidências suportam uma relação estreita entre a presença do cromossomo B e sexo (Clark & Kocher, 2019). Além disso, os fenômenos de origem e evolução do cromossomo B estão muito próximos dos mecanismos de heteromorfia cromossômica sexual, como: origem independente, acúmulo de DNA repetitivo, fortemente heterocromático e eventualmente, genes específicos (Camacho et al., 2000) . Em algumas espécies de artrópodes, diversas evidências apontam a origem do cromossomo Y a partir de um cromossomo B (López-León et al., 1994; Nur et al., 1988). Adicionalmente, o cromossomo B de *Drosophila* contribuiu para a formação do cromossomo Y, que adquiriu mecanismos para se manter (Carvalho, 2002).

Sobre a composição, cromossomos B são ricos em DNA repetitivo, o que garante sua característica fortemente heterocromática. Embora a presença do B não seja relacionada com alterações fenotípicas, outras evidências mostram alterações fisiológicas e no ambiente celular (Camacho et al., 2000). Como fenômenos gerais, fora constatado um atraso na meiose das células que possuem cromossomo B, bem como um acréscimo no conteúdo de proteínas nucleares e RNAs.

Atualmente, sabe-se que o cromossomo B pode ter alguma atividade gênica, pois, há o registro de transcrição de genes em plantas e animais (Benetta et al., 2020; Leach et al., 2005; Mazzoni et al., 2021) sugerindo atividade do cromossomo B em orquestrar seu próprio mecanismo de transmissão e permanência nos genomas. Dentre os genes codificadores de proteínas, foram identificados no cromossomo B diversas sequências transcricionalmente ativas relacionadas a diversos processos (**Tabela 1**).

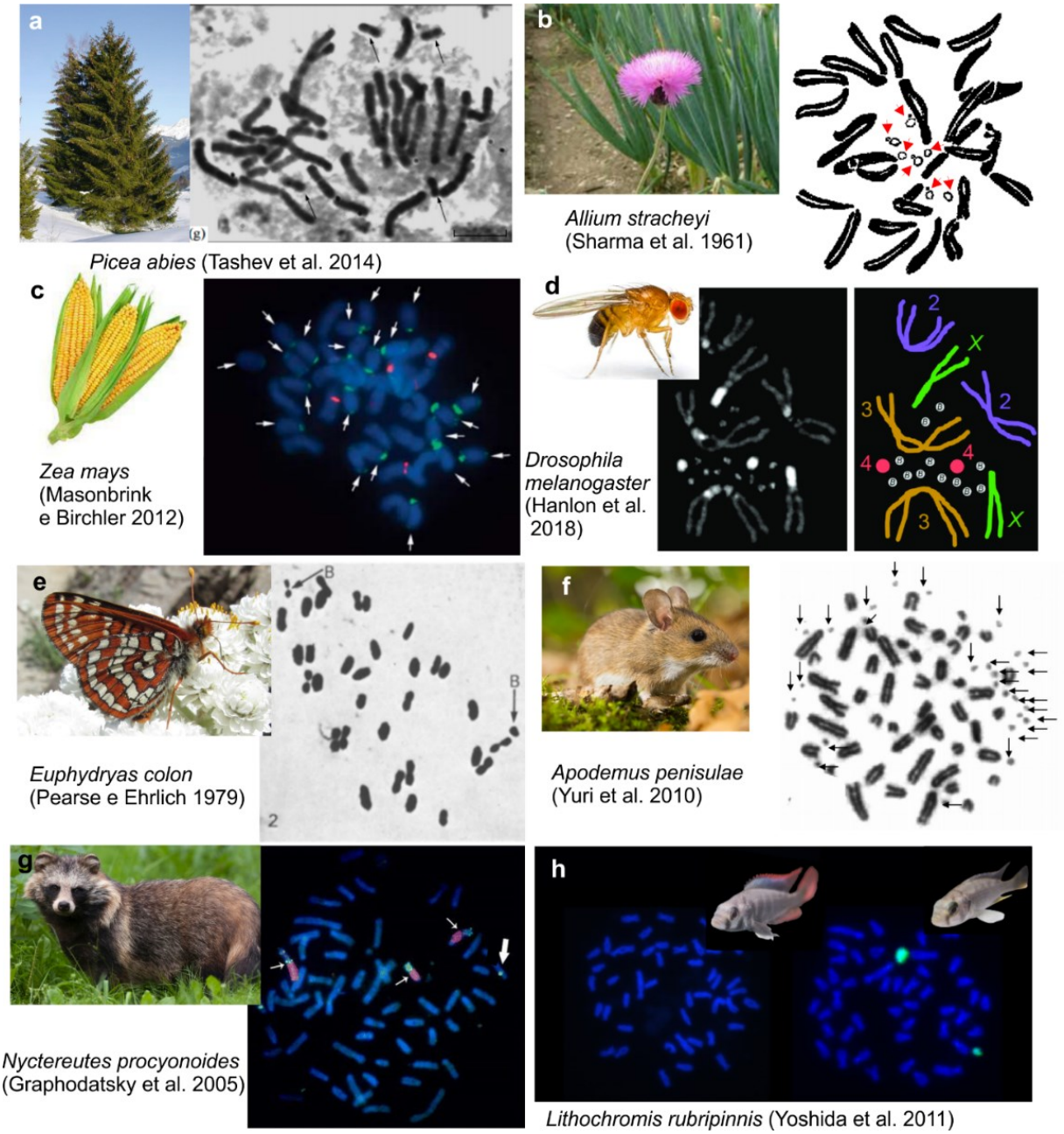

Figura 1 - Diversidade numérica e morfológica de cromossomos B. Os cromossomos B estão destacados com setas (a, b, c, e, f, g), ou representados graficamente (d). Em h, espécie onde cromossomo B é exclusivo nas fêmeas (metáfase da esquerda, cromossomos corados em verde).

Tabela 1 – Sequências descritas no cromossomo B

Espécie	Gene	Classe/ Função	Referência
Nectria haemotococca	*Cytochrome P-450 (Pda)*	Gene codificador de proteína/ Processos metabólicos	(Miao et al., 1991)
Astotilapia latifasciata	*Spindle and kinetocore-associated protein 1-like*	Gene codificador de proteína/ Segregação de cromossomos	(Valente et al., 2014)
Astotilapia latifasciata	BncDNA	RNA não codificante/ Função desconhecida	(Ramos et al., 2017)
Eyprepocnemis plorans	*Kinesin-like protein KIF20A*	Gene codificador de proteína/ Regulação da citocinese	(Navarro-Domínguez et al., 2017)
Secale cereale	*AGO4B*	Gene codificador de proteína/ Silenciamento induzido por RNA com função *slicer*	(Ma et al., 2017)
Zea mays	miRmB1	microRNA	(Huang et al., 2020)
Nasonia vitripennis	*Haploidizer*	Gene codificador de proteína/ função desconhecida	(Benetta et al., 2020)
Astyanax mexicanus	*Growth differentiation factor 6b*	Gene codificador de proteína/Candidato a gene determinante de sexo	(Imarazene et al., 2021)

Os avanços das tecnologias de sequenciamento, bioinformática e biologia molecular tem possibilitado o estudo em larga escala do conteúdo do B em diversas espécies (Ahmad & Martins, 2019; Ruban et al., 2017). Esses estudos trazem descobertas tanto para o conhecimento em Biologia Molecular como para o desenvolvimento de ferramentas a partir do entendimento do cromossomo B. Como é o caso do fungo patogênico *N. haemotococca* em que o cromossomo B carrega um gene que confere resistência contra um composto antimicrobiano produzido pela ervilha *Pisum sativum* (**Tabela 1**) (Miao et al., 1991).

No entanto, a maioria dos estudos que investigam a constituição dos cromossomos B são direcionados para a descrição de sequências repetitivas, como DNA satélites e elementos transponíveis, e genes codificadores de proteínas. Pouco tem se explorado a relação do cromossomo B com a porção não-codificante do genoma, com destaque para a classe de RNAs não-codificantes (ncRNAs) (Akbari et al., 2013; Huang et al., 2020; Li et al., 2017; Ramos et al., 2017). Os ncRNAs são

os principais moduladores da expressão gênica, pois, estão presentes em processos que afetam a transcrição e tradução (Adelman & Egan, 2017; Costa, 2007). Sabendo que, de forma geral, os cromossomos B possuem baixa densidade de genes codificadores de proteínas, caracterizar a porção não-codificante traz perspectivas para se explorar e descrever novas funções desse cromossomo parasita nos genomas hospedeiros.

1.2 RNAs não-codificantes (ncRNAs) no estudo de biologia cromossômica com foco no cromossomo B

Nos organismos mais derivados a porção não-codificante é predominante no genoma, compreendendo >90% (Mattick & Makunin, 2006). Com o avanço das tecnologias de sequenciamento e biologia molecular a lista de classes de ncRNAs tem aumentado, expandindo a visão sobre o papel do genoma não-codificante no ambiente celular (Mattick & Makunin, 2006; Moazed, 2009; Palazzo & Lee, 2015).

Dentre as principais classes de ncRNAs podemos citar três grandes grupos de acordo com sua função: os ncRNAs constitutivos, os intragênicos e os regulatórios (Moazed, 2009). Os ncRNAs constitutivos compreendem os RNAs transportadores, RNAs ribossomais e os pequenos RNAs nucleares que desempenham funções essenciais durante a transcrição e tradução. Já os ncRNAs intragênicos referem-se à porção não-codificantes dos RNAs mensageiros, como 3’ UTR (*untranslated region*), 5’UTR e íntrons, fundamentais para o processamento do RNA mensageiro. Para os ncRNA regulatórios há uma infinidade de subclassificações, destacando-se a classificação por tamanho: >200 nucleotídeos os RNAs longos não-codificantes (lncRNAs) e os <200 nucleotídeos os pequenos RNAs não-codificantes (sRNAs). A classificação das principais classes de ncRNAs está ilustrada na **Figura 2**.

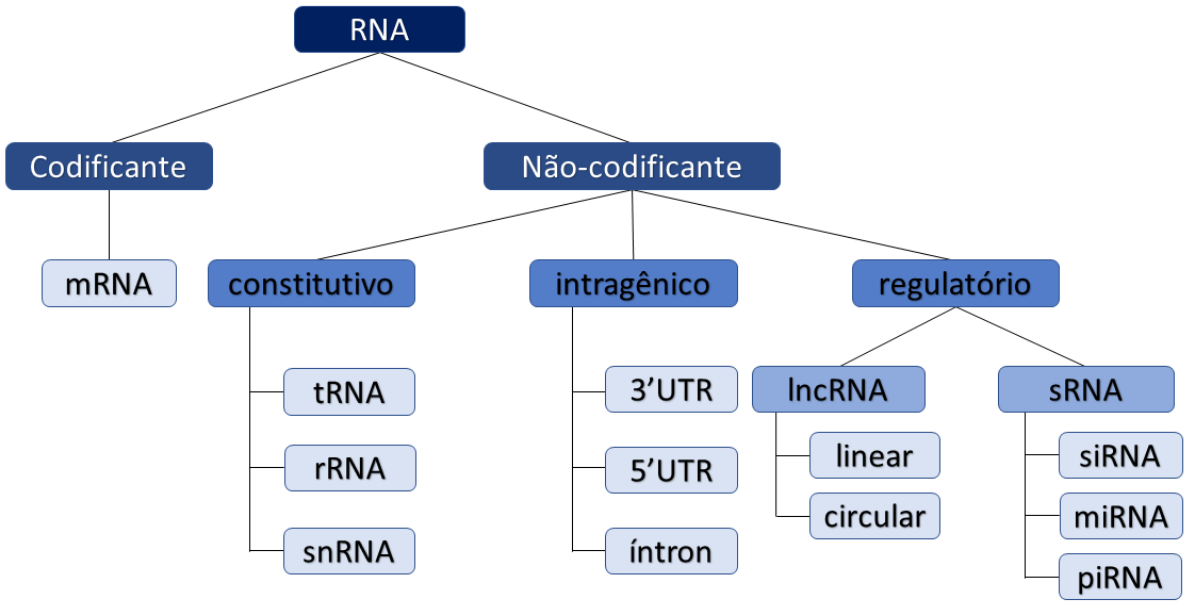

Figura 2 – Classificação dos RNAs não-codificantes (ncRNAs). Uma das classificações mais aceitas divide os tipos de ncRNAs em três grandes grupos: constitutivos (tRNA, RNA transportador; rRNA, RNA ribossomal; snRNA, pequeno RNA nuclear), intragênicos (3'UTR, untranslated region; 5' UTR; untraslated region; íntron) e regulatórios (lncRNAs, RNAs longos não-codificantes; sRNAs, pequenos RNAs). Esta última classe, como descrito no texto, é categorizada por tamanho do transcrito. Dentre os sRNAs detacam-se 3 principais classes: siRNAs (small interference RNA), miRNAs (microRNA) e piRNA (PIWI-Interacting RNAs).

Do ponto de vista de biologia cromossômica destacam-se os lncRNAs, como o *Xist* responsável pelo mecanismo de compensação de dose do cromossomo X em mamíferos (Hore et al., 2007). Adicionalmente, *flamenco*, um *cluster* de *PIWI-interacting RNA* (piRNA) foi descrito no cromossomo X de *Drosophila*, sendo o responsável pela produção da maior parte de piRNAs para o controle de elementos transponíveis (Goriaux et al., 2014). Algumas sequências de ncRNA já foram reportadas no cromossomo B (Akbari et al., 2013; Huang et al., 2020; Li et al., 2017; Ramos et al., 2017), dentre elas RNAs longos não-codificantes, lncRNAs (Ramos et al., 2017) e miRNAs (Huang et al., 2020). No entanto, ainda há poucos estudos com RNAs não-codificantes e cromossomos B, pois, nos últimos anos o foco esteve voltado aos genes codificadores de proteínas.

Devido a escassa informação sobre cromossomo B e sua relação com os pequenos RNAs não-codificantes, no presente trabalho foi investigado duas classes de sRNAs, os microRNAs (miRNAs) e os piRNAs na presença do cromossomo B.

Os sRNAs são conhecidos como os reguladores mais abundantes do genoma, por participarem de uma infinidade de vias que abrangem desde ciclo celular, desenvolvimento, até quadros patogênicos (Moazed, 2009). As proteínas da família Argonauta são essenciais para a atuação desses pequenos RNAs, destacando duas grandes subfamílias: Ago e Piwi. As proteínas Piwi (descobertas pioneiramente em *Drosophila* com a mutação *P-element induced wimpy testis*, Piwi) guiam pequenas moléculas de RNA de fita simples com biogênese independente de proteína Dicer, ou seja, os piRNAs. As proteínas Ago guiam siRNAs e miRNAs, moléculas derivadas de fita dupla de RNA e processamento dependente da proteína Dicer (Ponnusamy et al., 2017).

1.2.1 microRNAs e o controle da expressão gênica numa abordagem de genômica integrativa

Os miRNAs são sequências de aproximadamente 22 nucleotídeos (de 17 a 27 nucleotídeos) e têm sido reportados como os reguladores pós-transcricionais mais abundantes em animais (Lee et al., 1993; Ambros, 2004; Bartel, 2004; Wienholds et al., 2005; Fuller-carter et al., 2015) podendo somar de 1% a 5% do genoma (Olena & Patton, 2010).

O mecanismo de ação dos miRNAs ocorre pela complementaridade da sequência de miRNA e seus genes alvo (*targets*), que se constitui na interação da porção 5’ do miRNA com a região 3’UTR (*untranslated region*) do mRNA alvo. O segmento 5’UTR do miRNA que compreende do primeiro ao oitavo nucleotídeo é chamado de *seed*, sendo que do segundo ao sétimo nucleotídeo devem obrigatoriamente participar da interação miRNA:alvo para sítios canônicos de interação (Grimson, 2007; Bartel, 2009) (**Figura 3**).

Lee e colaboradores (1993) foram os pioneiros na descoberta dos miRNAs, ao detectarem em *Caenorhabditis elegans* um pequeno transcrito que estaria regulando um gene do desenvolvimento após sua transcrição, por ser constituído de uma sequência *antisense* ao gene.

A localização genômica dos genes de miRNAs ocorre de forma bastante heterogênea podendo encontrar-se em regiões intergênicas, regiões de íntrons e ainda regiões de éxons, esse último em menor frequência (Olena & Patton, 2010). Mas de forma geral, os genes de miRNAs se distribuem ao longo da eucromatina,

em regiões com alta densidade gênica (Bartel, 2018). É comum que estejam agrupados em *clusters* que designam as famílias de microRNAs, na sua maioria caracterizadas por conterem as mesmas sequências *seeds*. Para *Danio rerio*, por exemplo, há um número de 44 famílias agrupadas, dessas, 41 são compartilhadas com outros vertebrados (Heimberg et al., 2008; Thatcher et al., 2008). Desta forma é possível fazer uso de banco de dados de miRNAs, como o miRBase, para realizar predições *in silico* em espécies que ainda não tiveram microRNAs identificados.

As sequências de miRNA são geralmente conservadas ao longo do processo evolutivo, e o acúmulo de mutações pode acarretar perda de função (Bartel, 2004; Pang et al., 2006; Wahlestedt, 2006; Wang et al., 2007; Wang & Chang, 2011; Gebetsberger & Polacek, 2013; Fuller-Carter et al., 2015). Desta forma, pode-se dizer que os miRNAs estão sob fortes pressões seletivas, e isso provavelmente se deve ao fato de estarem localizado próximos a genes codificadores de proteínas.

Em animais a nomenclatura dos miRNAs é estabelecida pelas três primeiras letras referentes a espécie, seguidas do termo ‘miR’ e seu número de identificação. Ainda, pode carregar o termo 3p ou 5p que indicam à qual orientação a fita madura se refere (exemplo: microRNA 123 de *Homo sapiens* com fita ativa 3p: hsa-miR-123-3p). Ou seja, um mesmo microRNA pode estar presente em vários táxons seguindo o mesmo número de identificação e alterando apenas o nome da espécie como prefixo e/ou alguns nucleotídeos (Griffiths-Jones, 2004).

Leung & Sharp (2010) propuseram a síntese da molécula de miRNA em animais, na qual o transcrito primário de miRNA (pri-miRNA) possui uma estrutura secundária em forma de grampo de cabelo (*hairpin*). Em seguida, ocorre a ação da proteína Drosha, que cliva as extremidades do pri-miRNA, recebendo o nome de pre-miRNA, que, por sua vez, é carregado para o citoplasma pela Exportina-5. Em seguida, a proteína Dicer desconfigura a estrutura de *hairpin* formando um fragmento dupla fita com aproximadamente 22 nucleotídeos (Sohel, 2016). Uma das fitas (5p ou 3p), o miRNA maduro, quando associado ao complexo proteico Argonauta-RISC, é capaz de interagir com o sítio 3’UTR do mRNA alvo (Lee et al., 1993; Leung & Sharp, 2010) (**Figura 3**). Interação de miRNAs com outras regiões dos RNAs mensageiros, como 5’UTR e a porção codificante, também já foram descritas (Lytle et al, 2007; Liu et al., 2013).

Em vias gerais, a ligação do miRNA a seu alvo regula a tradução do mRNA através da sinalização para degradação ou impedindo o reconhecimento do tRNA para tradução (Lee et al.,1993; Pang et al., 2006; Wahlestedt, 2006; Wang & Chang, 2011). Um mesmo miRNA pode regular dezenas a centenas de genes (Grun et al., 2005), atuando em diferentes fases do ciclo celular e desenvolvimento, respostas a estímulos e doenças.

Apenas dois trabalhos caracterizam a presença de miRNAs em cromossomos B (Li et al., 2017; Huang, et al., 2019). Através de dados de RNAseq esses trabalhos mostraram que os miRNAs expressos pelo cromossomo B podem afetar a expressão dos miRNAs dos autossomos. Li et al., 2017, ainda reportaram algumas sequências de siRNAs e piRNAs expressos pelo cromossomo B da vespa *Nasonia vitripenis*.

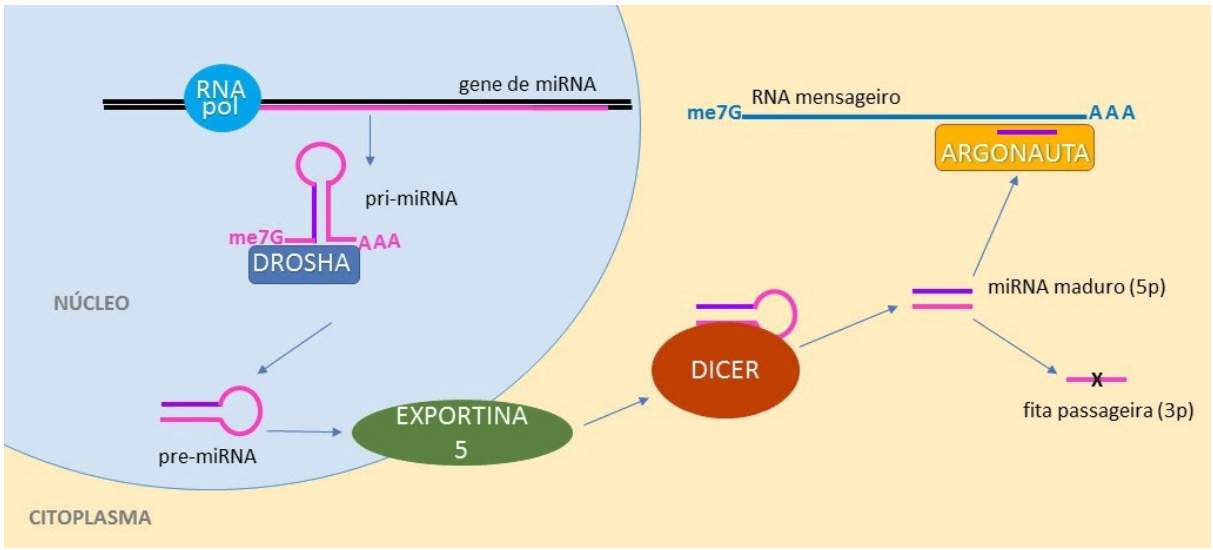

Figura 3 - Biogênese e atuação de miRNAs. Na imagem é exemplificado o processamento de um miRNA 5p, ou seja, após o processamento pela Dicer, ao separar as fitas, a sequência originária do braço 5p será o miRNA maduro, enquanto a 3p (fita passageira) será degradada. Um mesmo miRNA pode dar origem a miRNAs maduros 3p e 5p. Essa orientação é importante, pois, pode indicar atuação em alvos diferentes em momentos diferentes do ciclo celular (processo chamado de *arm-switching*).

1.2.2 PIWI-interacting RNAs (piRNAs) e manutenção da integridade do genoma

Dentre os pequenos RNAs não-codificantes os piRNAs foram descritos mais recentemente (Brennecke et al., 2007; Hartig et al., 2007). Junto com os miRNAs e siRNAs eles integram o grupo de sRNAs que se associam a proteínas Argonautas

como mecanismo de ação, formando os complexos RISC (RNA *induced silencing complex*, complexo de silenciamento induzido por RNA). No entanto, a principal diferença é que os piRNAs não formam estrutura em fita dupla e, portanto, não dependem da proteína Dicer durante o processamento (Moazed, 2009).

Até o presente momento, os piRNAs foram identificados apenas em animais, sendo descrito primeiramente em *Drosophila* e desempenham a função de silenciar elementos transponíveis principalmente nas células germinativas (Iwasaki et al., 2015). Em plantas esse processo é orquestrado pelos siRNAs (Grimson et al., 2008).

Algumas sequências de RNAs que pareciam controlar a atividade de elementos transponíveis foram descritas na década de 1990, os chamados rasi-RNAs (Pelisson et al., 1994). No entanto, apenas após a descrição dos domínios PAZ e PIWI das proteínas Argonautas, os rasi-RNAs foram reclassificados como *PIWI-interacting RNAs*: RNAs que interagem com proteínas PIWI para controle da mobilização de elementos transponíveis (Brennecke et al., 2007).

Os primeiros piRNAs foram observados em algumas linhagens de *Drosophila* que não possuem o elemento transponível *P-element*, chamadas de linhagens M, enquanto as linhagens que possuíam esse transponível são chamadas de linhagens P. Quando machos P são cruzados com fêmeas M a geração F1 sofre com a síndrome da disgenesia híbrida (*hybrid dysgenesis*), que dentre diversos problemas pode tornar a prole estéril. No entanto, cruzamentos entre, fêmeas P e machos M, fêmeas P e machos P, produzem proles normais. A conclusão é que a linhagem P desenvolveu piRNAs capazes de controlar as cópias do *P-element* ativo, evitando que esse elemento transponível cause problemas à prole. Sendo assim nos cruzamentos onde a fêmea pertence a linhagem P há a deposição dos piRNAs maternos que silenciam o *P-element.* Logo, a nomenclatura PIWI deriva desse fenômeno reportado em *Drosophila*, *P-element induced empty tests* (PIWI) (Luo & Lu, 2017).

Os piRNAs possuem de 24 a 35 nucleotídeos de tamanho e são encontrados no genoma agrupados em *clusters* maiores que mil pares de bases (1 kilobase) até centenas de kilobases. Os *clusters* de piRNAs são originados em regiões ricas em elementos transponíveis degenerados, os chamados “*junkyards”*, ou seja, regiões “ferro-velho” de elementos transponíveis (Han et al., 2015; Kofler, 2019; Luo et al.,

2020; Wierzbicki et al., 2021; Zeng & Tang, 2020). Portanto, a distribuição de piRNAs no genoma ocorre preferencialmente em regiões pericentroméricas e subteloméricas, ricas em elementos repetitivos, além de apresentarem marcas H3K9me3 e HP1, características de heterocromatina (Czech et al., 2018; Le Thomas et al., 2014). Esse processo de origem de *clusters* de piRNAs serve como uma "armadilha" para elementos transponíveis, pois, eles mesmos originam os *clusters* que irão atuar no seu silenciamento.

Os *clusters* de piRNA são classificados em *uni-strand* e *dual-strand,* segundo o enriquecimento de piRNAs maduros em uma fita do DNA ou nas duas fitas do DNA, respectivamente (Luo & Lu, 2017; Rosenkranz & Zischler, 2012). Esses *clusters* são transcritos em longos RNAs precursores que são transportados ao citoplasma próximos às mitocôndrias em uma região rica em proteínas, chamada de *nuage*, corpos Yb, corpos cromatóides ou *pi-bodies*. Nessa região os transcritos precursores podem ser processados por duas vias: a primária e a secundária (ou "ping-pong"). Na via primária, a atividade endonuclease da proteína PLD6 (homóloga a Zucchini de *Drosophila*) processa os transcritos precursores em pequenos transcritos primários, os pre-piRNAs (Han, 2015; Gainetdinov, 2018). A atividade de corte da proteína PLD6 ocorre preferencialmente em uma uracila 3', gerando o final do transcrito, e uma adenina 5', iniciando o próximo pre-piRNA. A distância 3' para 5' de um pre-piRNA para o outro é chamada de fase, por isso, os *clusters* dependentes de PLD6 são chamados de *phased* piRNAs (piRNAs faseados) (Han, 2015). Na sequência, os pre-piRNAs são processados pela proteína PIWI na sua extremidade 3' e direcionam-se ao ciclo "ping-pong". Acoplados às proteínas PIWI, forma-se o complexo de silenciamento (piRISC). Através da complementariedade de sequência, o complexo irá excisar o RNA mensageiro do elemento transponível, promovendo não apenas o silenciamento do transcrito, mas também criando um novo fragmento de sequência que se ligará a uma outra proteína PIWI tornando-se capaz de silenciar outros alvos (Aravin et al., 2007) (**Figura 4**). Sendo assim, conclui-se que a via primária aumenta a diversidade de piRNAs, correspondendo as mutações possíveis nos elementos transponíveis, enquanto o ciclo "ping-pong" aumenta a abundância de piRNAs (Han, 2015).

Além da atuação dos piRNAs silenciando elementos transponíveis no citoplasma há outra via atuação. O complexo piRISC pode ser transportado para o

núcleo e também por complementaridade de sequência, identificar sequências de elementos transponíveis sendo transcritas, atraindo a maquinaria de metilação de histonas para silenciar o gene de elemento transponível (Czech et al., 2018) (**Figura 4**).

Desta forma, os piRNAs são extremamente importantes para a manutenção da integridade do genoma, impedindo novas inserções de elementos transponíveis principalmente em células germinativas, sendo chamados de guardiões do genoma (Senti et al., 2010; Guo & Wu, 2013). Além disso, ao contrário dos miRNAs, os piRNAs e as proteínas envolvidas no processamento de piRNAs não são conservadas (Czech et al., 2018). E isso pode estar relacionado a rápida evolução de piRNAs a partir da degeneração de elementos transponíveis, fenômenos altamente variados entre as espécies (Kofler, 2019; Wierzbicki et al., 2021). Desta forma, piRNAs, em sua grande maioria, são espécie-específico, e correspondem a história evolutiva e variação de elementos transponíveis de determinado grupo.

Apenas um trabalho investiga piRNAs em cromossomos B (Li et al., 2017). A abordagem utilizada identificou no cromossomo B sequências de pequenos RNAs dentro do intervalo de tamanho para piRNAs. Pouco foi explorado a nível genômico a organização e a possível origem dessas sequências no cromossomo B.

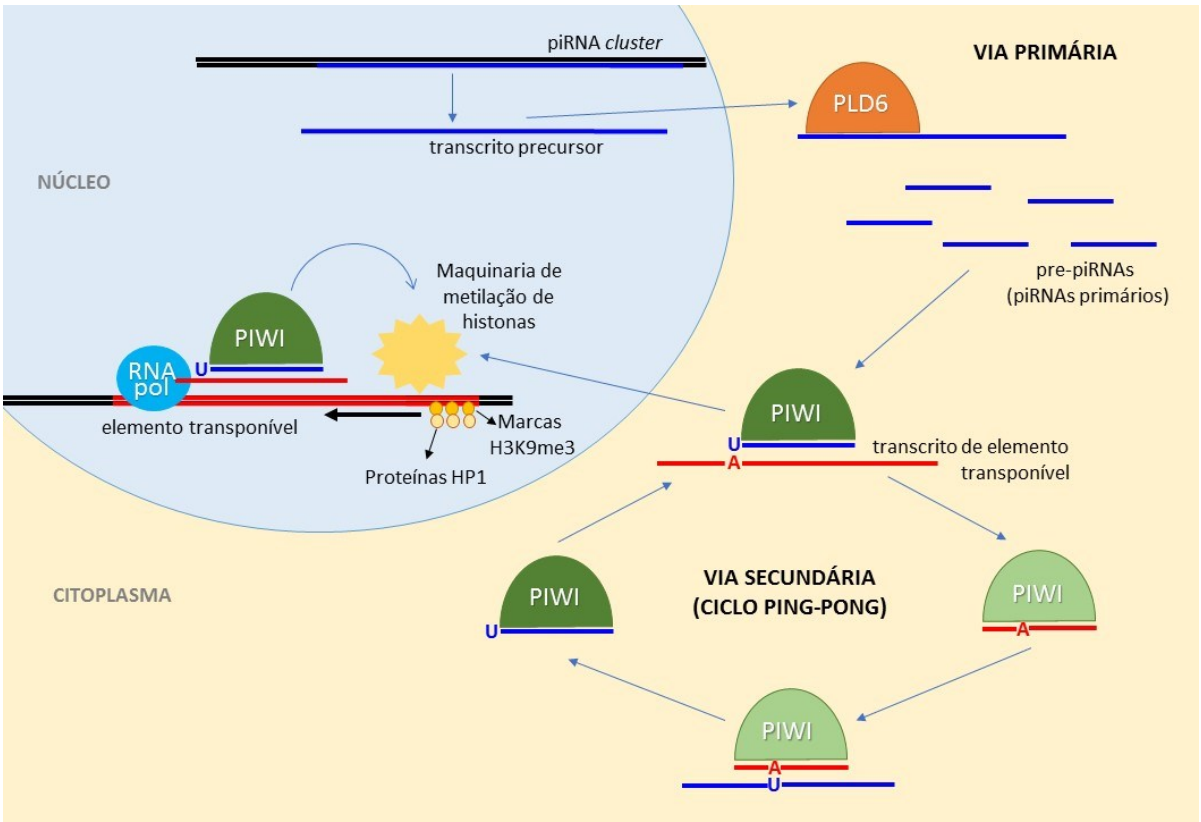

Figura 4 - Biogênese e atuação de piRNAs. Os transcritos de piRNAs são processados no citoplasma por duas vias: via primária, dependente da proteína PLD6 e via secundária ou ciclo ping-pong. As assinaturas das sequências de piRNAs estão destacadas, U (uracila) na primeira posição e a A (adenina) na décima posição do transcrito maduro. Ambas assinaturas são importantes para o reconhecimento e atividade de exonuclease pelas proteínas PIWI.

1.3 *Astatotilapia latifasciata* como modelo de estudo para cromossomo B em uma abordagem de genômica evolutiva e funcional

A família Cichlidae é uma família representativa dos Perciformes e está distribuída pela região tropical da África, Madagascar, América do Sul, e algumas regiões da Ásia (Kocher, 2004; Turner, 2007). Esta família tem oferecido grandes contribuições para o entendimento de processos evolutivos devido a rápida radiação adaptativa ocorrida pelos grupos que ocupam os grandes lagos africanos.

Além dos estudos de evolução, análises cromossômicas têm sido extensivamente realizadas no grupo, incluindo a caracterização de cromossomos B em várias espécies da família (Brawand et al., 2015a; D'Ambrosio et al., 2017; Kaufman et al., 1997; Mazzuchelli et al., 2012; Poletto et al., 2010). Dentre estas espécies, *Astatotilapia latifasciata* tem sido explorada sob o foco da citogenética clássica (Poletto et al., 2010), citogenética molecular (Poletto et al., 2012), e mais recentemente, genômica (Jehangir et al., 2019; Ramos et al., 2017; Valente et al., 2014).

Esta espécie é natural dos lagos Kyoga e Nawampasa (lagos satélites ao lago Vitória - África) é facilmente mantida em biotérios devido ao seu pequeno porte e fácil manipulação de condições como temperatura e qualidade da água (**Figura 5a** e **5b**).

O estudo de cromossomos B em peixes data da década de 1980, e a partir de então mais de 100 espécies com cromossomos B já foram descritas o que corresponde a 16,28% de todas as espécies cariotipadas (D'Ambrosio et al., 2017; Jones, 2017). Segundo o banco de dados de cromossomos B, Bchrom! v1.0 (disponível em http://www.bchrom.csic.es/, último acesso em 17 de novembro de 2021) a família Cichlidae possui 15 espécies com cromossomo B publicados nos últimos 30 anos. Em *A. latifasciata*, cromossomos B podem estar presentes em ambos os sexos e apresentam-se heterocromáticos (**Figura 5d** e **5e**). Além disso, a constituição do cromossomo B nesta espécie é um mosaico composto por

sequências repetitivas, pseudogenes, sequências degeneradas e sequências gênicas ligadas ao ciclo celular que parecem ter atividade transcricional (Valente et al., 2014). Além de genes codificadores de proteínas tem se explorado a relação da presença do cromossomo B com ncRNAs, envolvidos essencialmente em vias regulatórias. Ramos e colaboradores (2017) identificaram um lncRNA altamente disperso no cromossomo B que está diferencialmente expresso quando se compara indivíduos com e sem cromossomo B.

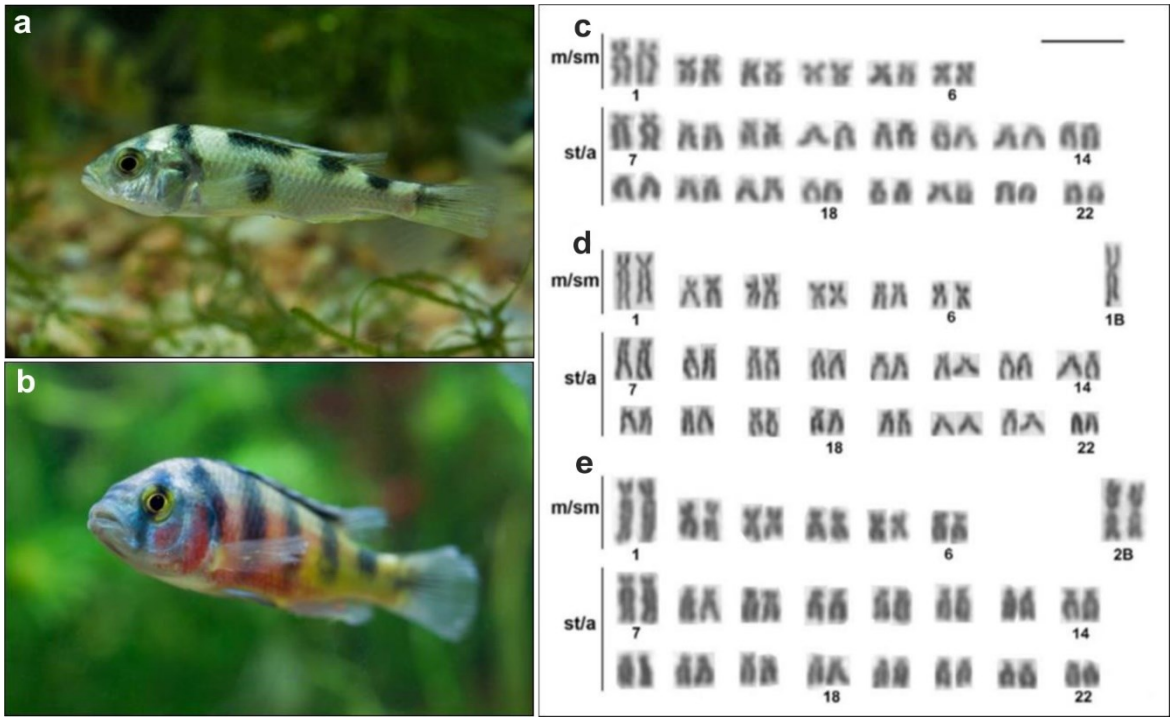

Figura 5 – Exemplares e cariótipo da espécie *A. latifasciata*. **a** Exemplar fêmea de coloração pálida. **b** Exemplar macho com coloração vermelha e amarela. Ambas imagens pertencem ao acervo de fotos do Laboratório Genômica Integrativa do Instituto de Biociências de Botucatu da UNESP (sem escala). **c-e** Cariótipo da espécie (Poletto et al., 2010). **c** Indivíduo sem cromossomo B (B-); **d-e** Indivíduos com um e dois cromossomos B (B+), respectivamente.

Valente e colaboradores (2014) reportaram uma vasta lista de genes com alta integridade, ou seja, região codificadora completa, presentes no cromossomo B de *A. latifasciata*. Entre esses genes foi encontrado *mitochondrial cardiolipin hydrolase-like* (LOC101465713) (*pld6*), com 100% de integridade. A família PLD tem envolvimento com a regulação da morfologia de mitocôndrias, e, mais especificamente, a proteína PLD6 está presente na formação do corpo denso na região perinuclear (*nuage*), localização onde concentram-se as proteínas da

biogênese e atuação dos PIWI-piRNAs e são responsáveis pelo processamento do transcrito precursor em piRNAs primários (Han et al., 2015).

Além de genes codificadores de proteínas, a análise da constituição do B mostrou a inserção de vários elementos transponíveis (Valente et al., 2014). Alguns desses elementos transponíveis estão diferencialmente expressos comparando com indivíduos não portadores de Bs (Coan & Martins, 2018), indicando uma possível diferença de atividade desses elementos relacionados a presença do cromossomo B. Estudos com cromossomos B de outros organismos também verificaram a diferença de expressão de elementos transponíveis presentes no B (Jie et al., 2019; Leach et al., 2005; Tanić et al., 2005).

Com evidências para o impacto do cromossomo B em vias regulatórias, mas pouca informação sobre a atuação de sequências não-codificantes nessas vias, o presente trabalho investigou a relação entre miRNAs e piRNAs à presença do cromossomo B em *A. latifasciata*.

Para cada classe, miRNAs e piRNAs, foram encontrados resultados que mostram o impacto do cromossomo B na regulação de diferentes vias. Utilizando diversas estratégias, nenhum gene de miRNA foi encontrado no cromossomo B. No entanto, foram encontrados diversos miRNA diferencialmente expressos na presença do cromossomo B. Ainda, predições *in silico* mostram a interação de miRNAs com genes duplicados no cromossomo B, indicando que sequências do complemento A pode estar interagindo com o conteúdo do B.

Por um outro lado, foram encontrados três *clusters* de piRNAs no cromossomo B, dois deles transcritos exclusivamente nas amostras B+. Esse resultado confrontou a proposta inicial desse projeto que esperava relacionar a atividade de piRNAs do genoma padrão de *A. latifasciata* controlando os elementos transponíveis expandidos no cromossomo B, assim, permitindo a permanência desse elemento extra no genoma.

Além disso, ao longo do período outros resultados foram obtidos e integrados nessa tese como detalhado na seção 3.

2 Objetivos

2.1 Geral:

Avaliar o impacto da presença do cromossomo B na atividade de pequenos RNAs não-codificantes de *Astatotilapia latifasciata.*

2.2 Específicos:

- Caracterizar microRNAs e PIWI-interacting RNAs nos genomas com e sem cromossomo B de *A. latifasciata*
- Comparar a expressão dos pequenos RNAs na presença do cromossomo B
- Investigar os alvos de interação dos pequenos RNAs relacionados à presença do cromossomo B

2.3 Suplementares:

- Fornecer um panorama histórico dos trabalhos envolvendo RNAs aplicados ao estudo de cromossomo B
- Organizar atividades de divulgação científica com base na literatura existente

3 Materiais, método, resultados e discussão: introdução aos capítulos

Materiais, método, resultados e discussão estão organizados em capítulos apresentados no formato de artigos científicos (capítulos 1, 2 e 3) e relatos sobre divulgação científica (capítulo 4).

O capítulo 1, intitulado "*Differential expression of miRNAs in the presence of B chromosome in the cichlid fish Astatotilapia latifasciata*" descreve os resultados da pesquisa sobre microRNAs na espécie modelo do estudo. Esse capítulo foi publicado na revista *BMC Genomics*.

O capítulo 2, "*First characterization of PIWI-interacting RNA clusters and their role in genome protection against the accumulation of transposable elements in parasitic B chromosomes*", reúne os resultados sobre piRNAs.

O capítulo 3 é um artigo de revisão sobre os trabalhos envolvendo RNA e cromossomos B. A revisão "*B chromosome tree of life: an RNA world perspective*" também apresenta desafios a serem superados e perspectivas para se avançar no conhecimento a respeito da evolução e função do cromossomo B através da investigação de RNAs.

Os materiais suplementares de cada capítulo estão disponíveis na pasta online com acesso pelo link: https://drive.google.com/drive/folders/1yaRS319ZGgUIW-MTVMbPz0d0W_H1OrL9?usp=sharing. Para acessá-los, basta copiar o endereço do link no navegador. Ao final de cada capítulo, os arquivos suplementares também estão com os respectivos links. Para acessá-los integralmente é necessário realizar o download dos arquivos. Eventuais problemas para acessar os arquivos favor contactar a autora pelo e-mail jordanainoliveira@gmail.com.

O último capítulo, "Relatos sobre divulgação científica: democratizar o conhecimento é construir um futuro melhor" descreve o trabalho de comunicação da ciência realizado ao longo do período. Esse capítulo apresenta algumas produções e materiais publicados com o objetivo de popularizar o conhecimento científico e discutir educação científica para o avanço e autonomia da sociedade.

As referências citadas ao longo dos capítulos bem como na seção de Introdução estão apresentadas na seção 5, ao final dessa tese.

3.1 Capítulo 1

"The 'Argonautes' looked at one another in amazement
and exclaimed with one voice: "Hercules!"
– Robert Graves

Differential expression of miRNAs in the presence of B chromosome in the cichlid fish *Astatotilapia latifasciata*

Jordana Inácio Nascimento-Oliveira[1], Bruno Evaristo Almeida Fantinatti[2], Ivan Rodrigo Wolf[1], Adauto Lima Cardoso[1], Erica Ramos[1], Nathalie Rieder[3], Rogerio de Oliveira[4] & Cesar Martins[1]

[1] Department of Structural and Functional Biology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, SP, Brazil

[2] Medical School, University of Ninth of July (UNINOVE), Bauru, SP, Brazil

[3] Faculty of Mathematics and Natural Sciences, University of Bonn, Bonn, Germany

[4] Department of Biostatistics, Plant Biology, Parasitology and Zoology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, SP, Brazil

Published *BMC Genomics*: doi.org/10.1186/s12864-021-07651-w

Abstract

B chromosomes (Bs) are extra elements observed in diverse eukaryotes, including animals, plants and fungi. Although Bs were first identified a century ago and have been studied in hundreds of species, their biology is still enigmatic. Recent advances in omics and big data technologies are revolutionizing the B biology field. These advances allow analyses of DNA, RNA, proteins and the construction of interactive networks for understanding the B composition and behavior in the cell. Several genes have been detected on the B chromosomes, although the interaction of B sequences and the normal genome remains poorly understood. We identified 727 miRNA precursors in the *A. latifasciata* genome, 66% which were novel predicted sequences that had not been identified before. We were able to report the *A. latifasciata*-specific miRNAs and common miRNAs identified in other fish species. For the samples carrying the B chromosome (B^+), we identified 104 differentially expressed (DE) miRNAs that are down or upregulated compared to samples without B chromosome (B^-) ($p < 0.05$). These miRNAs share common targets in the brain, muscle and gonads. These targets were used to construct a protein-protein-miRNA network showing the high interaction between the targets of differentially expressed miRNAs in the B^+ chromosome samples. Among the DE-miRNA targets there are protein-coding genes reported for the B chromosome that are present in the protein-

protein-miRNA network. Additionally, Gene Ontology (GO) terms related to nuclear matrix organization and response to stimulus are exclusive to DE miRNA targets of B^+ samples. This study is the first to report the connection of B chromosomes and miRNAs in a vertebrate species. We observed that the B chromosome impacts the miRNAs expression in several tissues and these miRNAs target several mRNAs involved with important biological processes.

Background

B chromosomes (Bs) are extra and nonessential elements found in approximately 10 to 15% of karyotyped organisms, ranging from fungi to plants and animals, and do not follow classical Mendelian inheritance patterns (Camacho et al., 2000; Houben et al., 2014; Jones et al., 2008; Jones, 1991; Miao et al., 1991). The origin, evolution, genome content and morphology of B chromosomes vary among organisms (Ploskaya-Chaibi et al., 2015).

A proto-B chromosome can emerge from chromosomal rearrangements, partial duplication of A chromosomes (normal chromosomes of the karyotype) or nonmeiotic disjunction (Houben et al., 2014). This new element increases its genomic content by insertion of A chromosome sequences copies, including various repetitive DNA classes (Bueno et al., 2013; Coan & Martins, 2018; Silva et al., 2017), protein-coding genes (Graphodatsky et al., 2005; Miao et al., 1991; Navarro-Domínguez et al., 2017; Rajičić et al., 2015; Valente et al., 2014), pseudogenes (Banaei-Moghaddam et al., 2013), retrogenes (Carmello et al., 2017), organellar DNA sequences (Martis et al., 2012) and noncoding sequences (Akbari et al., 2013; Huang et al., 2020; Li et al., 2017; Ramos et al., 2017). Duplicated sequences in B chromosomes have been proposed to facilitate its permanence in the host genome. These sequences may help the B chromosome drive during gametogenesis avoiding B elimination (Houben et al., 2014; Valente et al., 2014). In this way, the characterization of the B genomic content and its effects using genomics and bioinformatics tools is a promising approach to understand this extra element and its relation to the host genome (Ahmad & Martins, 2019; Valente et al., 2017).

Regarding to noncoding RNAs, some few sequences have already been reported either in the B chromosome or impacting in the expression of autosomal sequences in the B^+ samples. Noncoding RNAs exert strong effects on cell biological

processes and are potentially related to the presence of the B chromosome (Huang et al., 2020; Li et al., 2017; Ramos et al., 2017). Among noncoding RNAs, microRNAs (miRNAs) (~22 nucleotides long) act in the translation control by promoting the degradation or cleavage of mRNAs. That is why they are responsible for the control of important processes, such as development and differentiation, cell cycle regulation, stress and aging, and some diseases such as cancer. Notably, miRNAs are one of the most abundant regulators in the genome and several of them are highly conserved among organisms (Budak & Zhang, 2017; Grün et al., 2005; Moazed, 2009).

The biogenesis of miRNAs starts in the nucleus with the transcription of the primary miRNA (pri-miRNA) that has a hairpin structure and is processed to form a RNA duplex, named pre-miRNA. The pre-miRNA is transported to the cytoplasm and processed by the Dicer, that cleaves the RNA duplex into two single RNA molecules, 5p and 3p arms. Generally, only one arm will become the mature sequence while the another one will be degraded (Griffiths-Jones et al., 2011). When associated with the Argonaute protein, the mature miRNA interacts with its target based on antisense Watson-Crick pairing that occurs mainly in the 3' untranslated regions (3'UTR) of mRNAs (Leung & Sharp, 2010).

The connection between B chromosome and miRNAs has only been investigated in two organisms, an invertebrate and a plant species. The wasp *Nasonia vitripennis* carries a selfish supernumerary chromosome called PSR (Paternal Sex Ratio) that transcribes several small RNAs sequences, such as microRNAs, small interference RNAs and PIWI-Interacting RNAs (Li et al., 2017). In maize, B-derived miRNAs were found to affect A chromosome miRNA expression (Huang et al., 2020). However, to the best of our knowledge, these reports are the only two describing small noncoding RNA sequences in the context of B chromosomes. Thus, the impact of B chromosome duplications in the small noncoding RNAs expression is poorly understood.

Among vertebrates, B chromosomes have already been described in approximately 100 fish species (Jones et al., 2008), corresponding to 16.28% of karyotyped species (D'Ambrosio et al., 2017). Teleost fishes are important for evolutionary studies, especially in the Cichlidae family, due to their rapid adaptive radiation in East African great lakes (Brawand et al., 2015a; Kaufman et al., 1997;

Turner, 2007). B chromosomes were detected in several cichlid species (Kaufman et al., 1997; Martins & Jehangir, 2021; Poletto et al., 2010). Among them, the African cichlid *Astatotilapia latifasciata*, which carries one or two B chromosomes in both sexes, has been extensively investigated through classical cytogenetics (Fantinatti et al., 2011; Poletto et al., 2010), molecular biology (Cardoso et al., 2019; Coan & Martins, 2018) and, more recently, genomic approaches (Carmello et al., 2017; Jehangir et al., 2019; Ramos et al., 2017; Valente et al., 2014). Repetitive elements (Coan & Martins, 2018), coding genes (Valente et al., 2014) and a long noncoding RNA (Ramos et al., 2017) have already been identified in the B chromosome of *A. latifasciata.* Some of these sequences revealed a differential expression in the B^+ samples suggesting transcription activity and involvement of this extra element into several biological pathways (Huang et al., 2020; Navarro-Domínguez et al., 2019; Ramos et al., 2017).

The *A. latifasciata* B chromosome content has been investigated by comparing sequencing from B^- and B^+ DNA and RNA samples. The B chromosome gene content was first reported through genomic coverage rate analysis based on Illumina and 454 Life Sciences sequencing of microdissected B chromosome (Valente et al., 2014). In this work, the microdissected B chromosome sequences were compared to *Metriaclima zebra* reference genome, and the first B-genes of *A. latifasciata* were reported. Later, the *A. latifasciata* draft genome was constructed using Illumina high scale data and identified several duplicated contigs in the B chromosome (Jehangir et al., 2019). The coverage ratio compares coverage of sequenced among B^- and B^+ samples aligned against a reference assembled genome, which allows to identify higher coverage regions on the B^+ sequencing dataset, that represent duplicated regions in the B chromosome (Jehangir et al., 2019; Valente et al., 2014).

In this study, large-scale small RNA sequences (sRNAseq) were generated from the brain, muscle and gonads of B^- and B^+ individuals of both sexes of *A. latifasciata*. Using bioinformatics approaches, the *A. latifasciata* miRNA profile was described, and compared with other teleost miRNAs, mainly cichlid species. This allowed the identification of conserved and specific novel miRNAs. In this work, we introduce the application of several bioinformatics tools to investigate miRNA sequences in the context of B chromosomes based on coverage ratio analysis and

the generated “B-blocks” as previously reported (Valente et al., 2014). B-blocks are putative genomic regions observed on B chromosomes and detected via coverage ratio analysis as a result of NGS read coverage comparison between the two sequenced genomes (B^+ and B^-). We described 104 miRNAs that were differentially expressed (DE), either up or downregulated in the presence of the B chromosome compared to samples without B chromosome (the control). These miRNAs have common mRNA targets in the brain, muscle and gonads. Additionally, we found protein coding genes already described in the B chromosome (the B genes) as targets of DE miRNAs. Moreover, a network based on human protein-protein interactions of the DE miRNAs targets highlights the great potential of DE miRNAs in the influence of B chromosomes over several biological processes.

By combining the sRNAseq with the availability of *A. latifasciata* genomic and mRNA transcriptomic data we described the miRNome of this cichlid species. Further, this is the first study that relates the miRNA expression and the B chromosome presence in a vertebrate species. This is also the first report of coding and noncoding interactions related to B chromosome presence.

Materials and Methods

Samples

A. latifasciata fishes were obtained from the Integrative Genomics Laboratory fish room at São Paulo State University, Botucatu (SP), Brazil and were genotyped for presence/absence of the B chromosome using the previously developed marker for B chromosomes (Fantinatti & Martins, 2016). The animals with B (B^+) and without B (B^-) were kept in different aquariums until the tissue collection procedure. We used four animals to each group: females B^-, females B^+, males B^- and males B^+. Twelve animals (four of each group) were used to RNA-seq and twelve to DNA and RNA extraction to PCR, qPCR an RT-qPCR, totalizing 24 animals. The animals were submitted to euthanasia by immersion in eugenol 1% for three minutes, following the tissue extraction (brain, gonads and muscle) by liquid nitrogen flash frozen method. The tissues were stocked in -80 refrigerators until the RNA and DNA extraction, described below.

RNA sequencing of small RNAs in brain, gonads and muscle

The total RNA was extracted from the brain and gonads of six males and six females without B chromosomes (B-) and with B chromosomes (B^+) and the muscles of four males and four females B^- and B^+ using the TRIzol™ (Invitrogen) protocol. The samples were assessed for quality using electrophoresis on an agarose gel and RIN (RNA Integrity Number, >8) parameters and then sent for small RNA sequencing (sRNAseq). This service was executed by Sequencing Service at LC Sciences - Houston, TX, USA, using a single-end Illumina HiSeq 2000 platform after the library construction with TruSeq® Small RNA Sample Preparation (Illumina). The libraries are deposited in the NCBI database (access numbers SRR13040679-SRR13040710).

MicroRNome construction and expression analysis

The small RNA libraries were filtered for quality using Fastx-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) with the default parameters. Adaptors (TGGAA) and sequences shorter than 17 nucleotides and longer than 27 nucleotides were removed using Cutadapt (https://cutadapt.readthedocs.io/en/stable/) (**Additional File 11 - Table S1**).

The miRbase fish database (Kozomara et al., 2019) (Release 22.1, October 2018) (**Additional File 10 - Table S3**) was clustered using CD-hit software (default parameters) (Li & Godzik, 2006) to create a nonredundant reference list of fish miRNAs.

The filtered small RNA-Seq libraries were used as input for mapping based on default parameters (Friedländer et al., 2012) to predict and identify miRNAs (**Additional File 11 - Table S2**). Then, the nonredundant reference list of fish miRNAs was aligned to the *A. latifasciata* genome assembly (Jehangir et al., 2019) using Bowtie2 (v 2.3.3.1) (Langmead & Salzberg, 2012). The final prediction step consists of the submission of the mapper.pl output, the original nonredundant reference list and the aligned reference list to miRDeep2.0.1.2 analysis using default parameters (Friedländer et al., 2012). The miRDeep software allows identifying known and novel miRNAs based on the existence of the miRNAs in the miRBase. The novel miRNAs were analyzed with BLAST on the seed region (2-8 nucleotides) to characterize if the new miRNA belongs to either a known miRNA family or a known miRNA derived (Friedländer et al., 2012).

The differential expression analysis was performed using miaRma pipeline (Andrés-León et al., 2016) with the parameters desoft=EdgeR-Noiseq; filter=yes; cpmvalue=1; repthreshold=6; fc_threshold=0.5; edger_normethod=TMM; replicates=yes; bcvvalue=0.4; replicatevalue=biological; and noiseq_normethod=tmm. The comparisons were performed separately for each tissue (brain, gonads and muscle), sex (male and female) and the presence or absence of the B chromosome (B^+/B^-), always considering the B^- as control. The miRNAs with $p < 0.05$ were considered differentially expressed (DE).

Analysis of B chromosome copies

Three strategies were adopted to identify possible copies of miRNA genes related to the B chromosome. The first was based on "coverage ratio analysis" (Valente et al., 2014). The miRNA annotations of the *A. latifasciata* genome were compared to the B^- and B^+ read coverage ratios of the *A. latifasciata* genome. The regions with twice as high coverage read in the B^+ genome indicate a duplicated B chromosome sequence copy (Valente et al., 2014) (*i.e.*: 60x coverage B^- reads and 120x coverage B^+ reads). The second strategy was to use all miRNA reads (from B^- and B^+ samples) that failed to align the *A. latifasciata* genome during miRNome construction to identify B chromosome-specific miRNAs. The unaligned reads were used for miRNA identification by miRDeep using the B^+ genome assembly (Jehangir et al., 2019), followed by the prediction of miRNA:mRNA interactions by TargetScan (described further below). To analyze if these sequences are exclusive from B chromosome, we performed a BLAST search using *Metriaclima zebra* genome (v0, bouillabase.org, last accessed on October 16, 2020). The scaffold matched from *M. zebra* genome was visualized with *A. latifasciata* B^- and B^+ genomic read alignments on Saci Base Jbrowser (https://sacibase.ibb.unesp.br/index.html, last accessed on October 16, 2020). Finally, the third strategy was based on the duplicated B chromosome blocks (B-blocks) identified in a previous study (Jehangir et al., 2019). The previously predicted miRNAs located on B-blocks were used to predict the miRNA:mRNA interactions with TargetScan (described further below).

3'UTR prediction and miRNA:3'UTR interactions

The 3'UTR sequences from the *A. latifasciata* transcriptome (Nakajima, 2019) were predicted using Transdecoder (v5.5.0) (Haas et al., 2013). The annotation file for each transcript was used with Bedtools (v2.26.2) (Quinlan & Hall, 2010) to retrieve the predicted 3'UTR for further analyses of interactions. Additionally, the transcript tissue specificity was retrieved based on observed expression (Nakajima, 2019) to assess tissue-specific interactions.

The miRNAs (novel and known) identified on miRNome and the 3'UTR from the *A. latifasciata* transcriptome were submitted to TargetScan 6.0 (Ulitsky et al., 2012) to predict the miRNA:3'UTR interactions. Scores less than -0.2 were considered significant predictions in accordance with software parameters.

Genomic PCR, qPCR and RT-qPCR

Genomic DNA was extracted using (Sambrook & Russell, 2006) protocol. The samples were genotyped using the molecular marker for the presence of the B chromosomes (Fantinatti & Martins, 2016). Primers for mature miRNA sequences were constructed using Primer3Plus. Conventional genomic PCR was performed using *Taq* DNA Polymerase (Invitrogen-10342-053) submitted to cycling at 94°C for 5 minutes, followed by 35 cycles (94°C for 1 minute, 50°C to 60°C for 45 seconds, and 72°C for 10 minutes) and 72°C for 10 minutes. The result was verified by agarose gel electrophoresis (1%). The genomic DNA of eight B^- and eight B^+ samples, including samples from males and females, was used for GDR analysis with qPCR and SYBR Green detection (Ampliqon) to verify the putative miRNAs associated with the B chromosome copy. The gene expression analysis was based on the total RNA extracted from the brain, gonads and muscle of B^- and B^+ of both sexes, and performed in biological triplicates. The samples were converted to cDNA libraries (High-Capacity RNA-to-cDNA™ Kit, Applied Biosystems) and used for RT-qPCR with SYBR Green detection (Ampliqon). The $\Delta\Delta C_q$ method was used to analyze relative expression (de Santis et al., 2011). Q-Gene software (Simon, 2003) was used for normalization with the *ubiquitin-conjugating enzyme* (UBCE) as the reference gene. The primers are described in **Additional File 11 - Table S4**.

Network construction and functional analysis

The human PPI network was downloaded from Biogrid to analyze miRNA functions (Chatr-Aryamontri et al., 2015). The network was filtered to retain only nonredundant interactions between human proteins detected by affinity chromatography technology (MI:0004), X-ray crystallography (MI:0114), far western blotting (MI:0047), fluorescent resonance energy transfer (MI:0055), protein complementation assay (MI:0090), experimental interaction detection (MI:0045) and two hybrid (MI:0018) experiments. The annotated protein-coding transcripts (Nakajima, 2019) commonly regulated by B-DE-miRNAs in each tissue (brain, gonads and muscle), together with B genes reported in a previous study (Valente et al., 2014), were used to extract a PPI subnetwork from the filtered Biogrid. The miRNA:3'UTR interactions were added to the PPI subnetwork to create the final B-miR-net. The protein list was subjected to a Gene Ontology (GO) analysis using gProfiler to verify exclusive and specific GO terms for proteins derived from B-DE-miRNA targets and the remaining proteins from all *A. latifasciata* targets (Raudvere et al., 2019).

Results

The *A. latifasciata* miRNome

A nonredundant dataset was constructed based on miRBase fish miRNAs to create a miRNA reference list. The procedure resulted in 1,456 precursors and 1,234 mature sequences. This fish miRNA reference list was used to identify the miRNAs in the sRNA-seq data and annotate them in the *A. latifasciata* genome. We identified 727 miRNA precursors (pre-miRNAs) throughout the *A. latifasciata* genome (see **Additional File 1**). Among them, 246 (33.84%) pre-miRNAs have similarity with described miRNAs in miRBase, and they are called known miRNAs. On the other hand, sequences that were not similar to existing miRNAs are called novel. Additionally, novel miRNAs could only present seed similarity to existing miRNAs, indicating new miRNAs probably belonging to an existent miRNA family (Friedländer et al., 2008). The novel *A. latifasciata* pre-miRNAs represent 481 (66.16%) precursor sequences, and 29.31% of them exhibit miRBase seed similarity (nucleotides 2-8 from the 5’ end of the mature miRNA).

Clustered miRNAs can be arranged in a 5 kilobase genomic region long and generally are related to the same transcription factors (Xiong et al., 2019). Here, we

considered clustered miRNAs when the sequences were found in the same genomic contig and not exceeded 5 kilobases distance. The clustered miRNAs (pre-miRNAs on the same genome contig) accounted for 232 (31.91%) sequences; the longest cluster contained 9 pre-miRNAs. Also, 495 (68,08%) precursor sequences are single miRNAs in a genomic contig. Usually, one arm (5p or 3p) is more highly expressed in the cell, while the other arm can be degraded (Griffiths-Jones et al., 2011). Thus, comparing the expression of the arms in the sRNAseq, fifty-five percent of pre-miRNAs displayed higher expression in the 5p arm mature sequence. Transcription was identified on the minus strand for 368 (50.62%) pre-miRNAs and on the plus strand for 359 (49.38%) pre-miRNAs. All the results mentioned above are described in the **Figure 1a** and **b**.

The last miRBase release added seven new fish species and 2,050 new sequences, summing up 3,687 miRNAs sequences (**Additional File 11 – Table S3**). From these new species in the last release, five are cichlids and contribute to 1,300 miRNAs sequences. The miRNA seed sequences from *A. latifasciata* have similarity with 15 fish species (representatives of nine teleost families) in miRBase, corresponding to 246 pre-miRNAs (**Additional File 1**). The species were verified according to the three first letters of the miRNA ID, which corresponds to the species ID in the animal miRNA nomenclature pattern, as indicated next in the species name. Thirty-five percent of known miRNA seed sequences share similarity with cichlids (*Astatotilapia burtoni* – abu, *Metriaclima zebra* – mze, *Neolamprologus brichardi* - nbr, *Oreochromis niloticus* – oni, and *Pundamilia nyererei* – pny); 65% show similarity with other teleost families (one species of Adrianichthyidae, *Oryzias latipes* – ola; two species of the Cyprinidae family, *Cyprinus carpio* – ccr, and *Danio rerio*, – dre; one species of Gadidae, *Gadus morhua* – gmo; one species of Ictaluridae, *Ictalurus punctatus* – ipu; one species of Pleuronectidae, *Hippoglossus hippoglossus* – hhi; one species of Salmonidae, *Salmo salar* – ssa; and two species of Tetraodontidae, *Fugu rubripes* – fru, and *Tetraodon nigroviridis* – tni) (**Figure 1b and 1c**). Even not being the most representative reference species *Danio rerio* seeds are well represented among the known miRNAs, probably indicating the presence of highly conserved miRNAs in the *A. latifasciata* miRNome (**Additional File 9 – Table S3**). Approximately 70% of novel miRNAs are not similar to any seed from miRBase, indicating their potential as either new specific or nondescribed

miRNAs (**Figure 1b and 1c**). When considering only the novel miRNAs that present seed similarity from miRBase, 60% are similar to seed sequences from *A. burtoni, M. zebra* and *O. niloticus*, which probably represent exclusive conserved miRNAs families among cichlids (absent or not conserved in other groups).

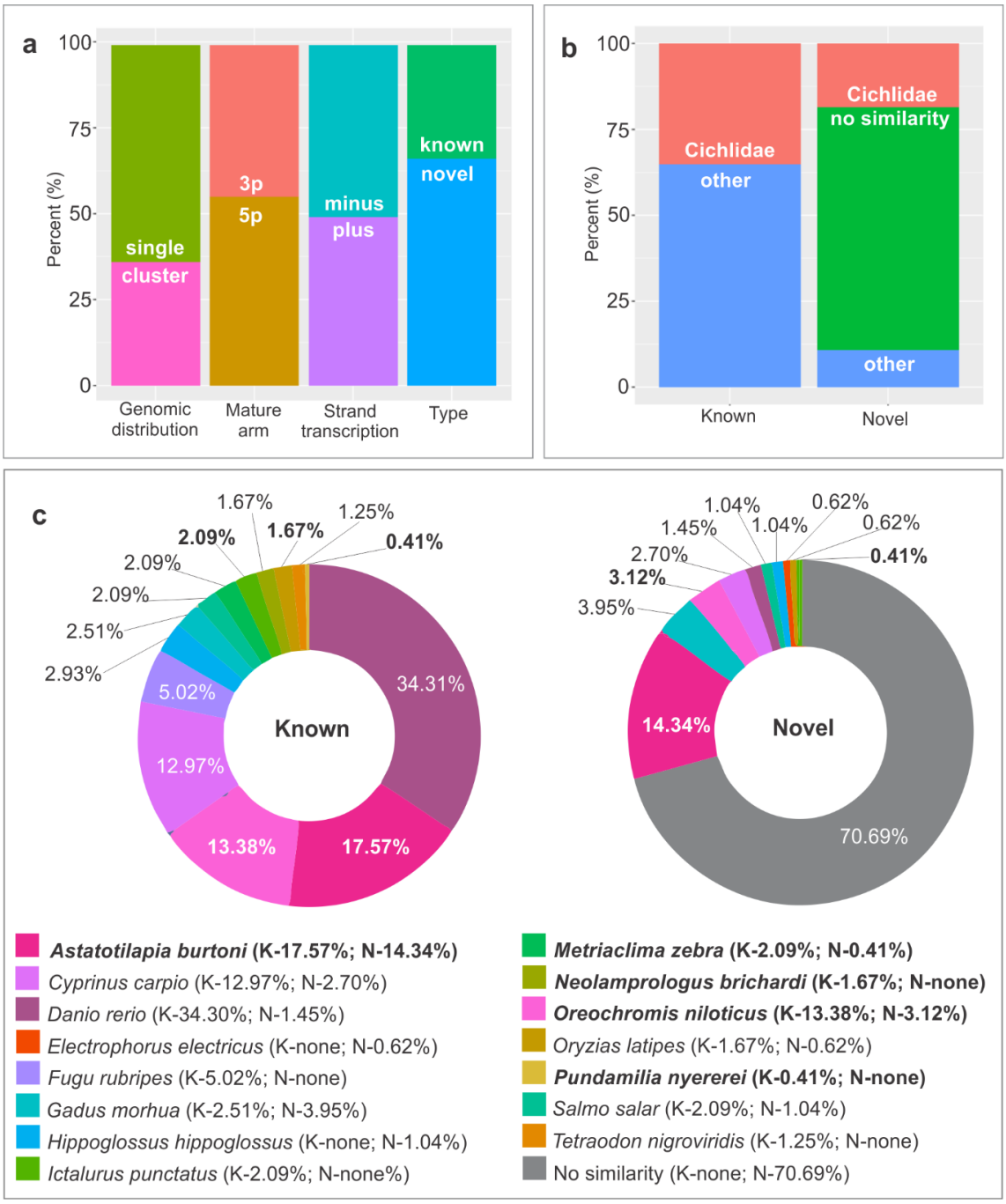

Figure 1 Description of the *A. latifasciata* miRNome. a Percentage of miRNA characteristics. Genomic distribution: the miRNA precursor arrangement on the genome; Mature arm: the mature sequence with more reads on RNA-Seq; Strand transcription: precursor transcription strand; Type: if the miRNA is similar to another known miRNA from miRbase (known) or if is probably a new miRNA sequence (novel). **b** Percentage of similarity with cichlids and other fishes. **c** Percentage of miRNA

similarity compared with fish sequences from miRBase. cichlid species are highlighted in bold. K, Known miRNAs; N, novel miRNAs.

Searching for miRNA genes on the B chromosome

We performed three different strategies in order to investigate the miRNA presence in the B chromosome. All strategies were based on DNA and sRNAseq comparison of B^- and B^+ samples (see Material and Methods).

The coverage ratio analysis (first strategy) consists in screening the coverage difference between the B^- and B^+ genomic reads aligned against the *A. latifasciata* assembled genome. This strategy allows us to find segments of A chromosomes that are duplicated on the B chromosome. We did not detect any miRNA gene inside a genomic region with coverage corresponding to the B chromosome (B^+ blocks). The second method was based on alignments using the sRNAseq reads from B^- and B^+ samples of all tissues that failed to align in the *A. latifasciata* reference genome. Then, we performed a second alignment using these unaligned sRNAseq reads to B^+ assembled genome (the *A. latifasciata* DNA with B chromosome). This method identified 21 novel miRNA genes, of which 6 were exclusively expressed in B^+ samples (**Additional File 2** and **Additional File 3**). Thus, ten B^+ assembly miRNA genes were selected for validation, as they had an adequate length for qPCR primer construction. Only one miRNA gene (called here novel_2026-B^+, **Figure 2a**, **Additional File 3**) located in contig NODE_313069 from the B^+ assembly was PCR-amplified (**Additional File 4**). However, amplification was observed in both B^- and B^+ samples (data not shown). qPCR experiments were performed to confirm that this genomic segment was in both groups of individuals (B^- and B^+). The gene dose ratio (GDR) compares the relative gene copies trough qPCR, which showed that the novel_2026-B^+ region has not GDR difference in the B^- and B^+ genomes, meaning the same number of copies in both genomes (**Figure 2b**). Additionally, novel_2026-B^+ is similar to scaffold_77 of *M. zebra*, where B^- and B^+ genomic reads are aligned (**Figure 2c**). The *A. latifasciata* B^- and B^+ genomes have several assembly gaps that might justify the absence of miRNA alignments. The evidence suggests that novel_2026-B^+ occurs in the *A. latifasciata* genome but is not present in the B chromosome. The gaps are probably resulting of an assembly bias in the *A. latifasciata* genome, where the short Illumina reads caused the region to be ignored during the assembly (**Additional File 4 – Figure S1**).

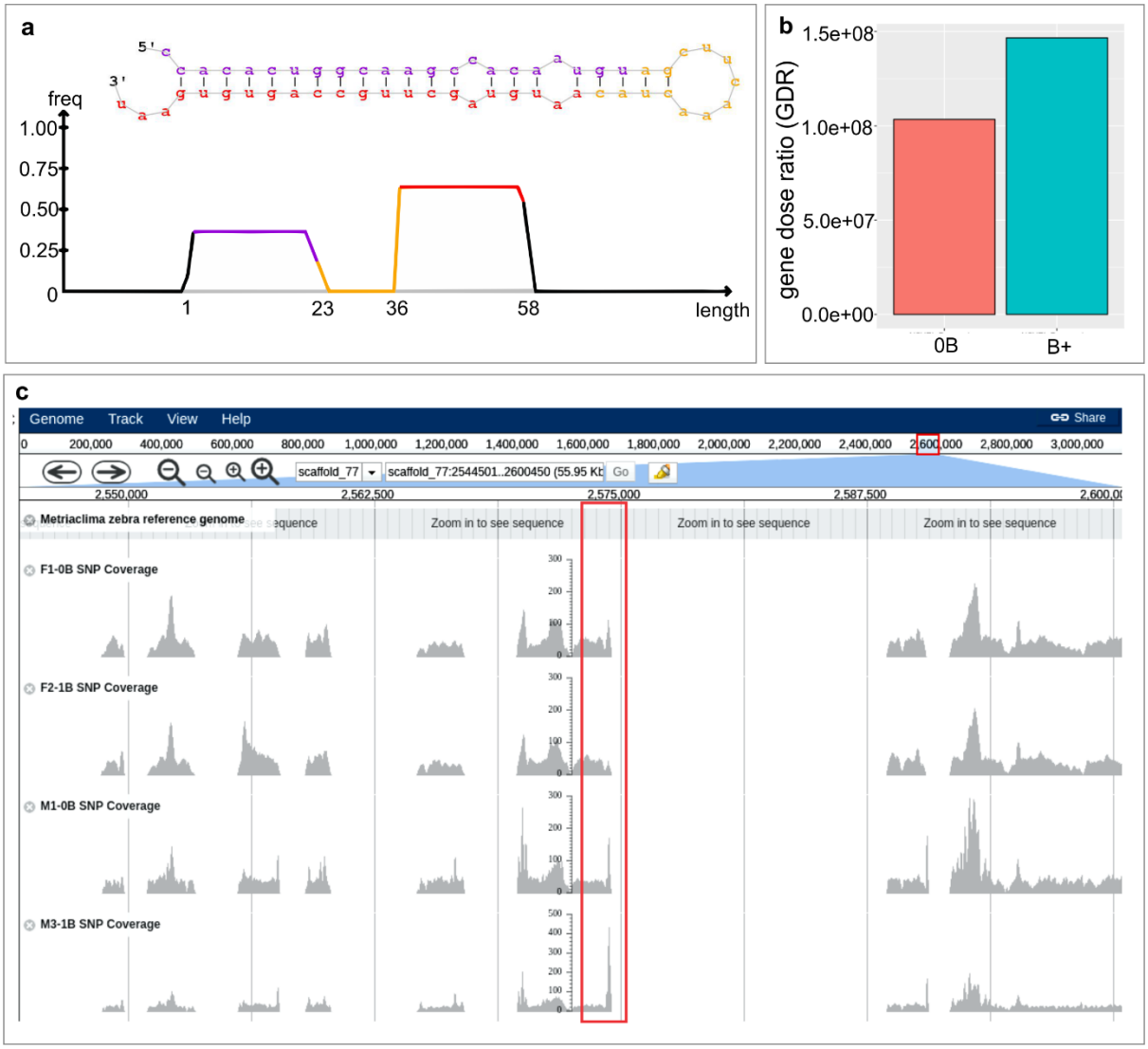

Figure 2 Functional miRNA absence in the B⁺ genome assembly. **a** The novel_2026-B⁺ predicted from the B⁺ assembly. This miRNA has a stem-loop secondary structure. **b** The NODE_313069 qPCR for B⁻ and B⁺ DNA samples. The difference between B⁻ and B⁺ amplification was not significant (p-value 13 >0.05). **c** Predicted novel_2026-B⁺ match with *M. zebra* scafold_77. The *A. latifasciada* genomic sequencing reads are shown below the *M. zebra* reference genome, F1-0B SNP coverage (female B⁻ sample), F-1B SNP coverage (female B⁺ sample), M1-0B SNP coverage (male B⁻ sample), M3-1B SNP coverage (male B⁺ sample). The gray area is the read coverage, the blank spaces show no aligned reads in this region, and the red rectangle highlights the novel_2026-B⁺ NODE_313069 region matching the *M. zebra* assembly.

The third approach was predicting miRNAs using the miRBase reference and the sRNAseq, as the previous strategies, but set the genomic “B-blocks” filtered by Jehangir et al. (Jehangir et al., 2019) as the reference background. With this strategy we would like to confirm if no duplicated miRNAs were missed in our manual coverage ratio strategy (first strategy). This prediction found 33 pre-miRNAs

on the B-blocks. Two miRNAs are similar to mir-2188, and the others are novel (no miRBase similarity) (**Additional File 6**). These 33 miRNAs did not show interaction with any mRNAs and were not considered for further analysis.

Therefore, based on the results obtained with these three strategies, we did not find strong evidences of miRNAs in the B chromosome. We discuss the limitations further.

Effects of B chromosomes on miRNA expression

The differential expression analysis was performed by comparing B^+ samples against the samples without B chromosome (as the control) in each tissue (brain, gonads and muscle) and each sex (male and female). Several miRNAs were differentially expressed (DE) in B^+ samples (either up or downregulated), which is why these sequences were called B DE miRNAs (B-DE-miRNAs). The heatmap in the **Figure 3a** shows the upregulated sequences (positive FoldChange in green gradient) and downregulated sequences (negative FoldChange in red gradient) in B^+ samples in each sex and tissue. The profile of DE miRNAs between tissues, sexes and the presence of the B chromosome detected 104 nonredundant miRNAs (**Figure 3a and b**). In brain, 12 novel and 2 known B-DE-miRNAs were found in females and 12 novel and 7 known B-DE-miRNAs were detected in males. In gonads, 29 novel and 18 known B-DE-miRNAs were found in females and 8 novel and 3 known B-DE-miRNAs were detected in males. In muscle, 6 novel and 4 known B-DE-miRNAs were found in females and 8 novel and 4 known B-DE-miRNAs were detected in males.

Five genomic regions carry B-DE-miRNAs, forming clustered miRNAs (**Figure 3c and Table 1**). The B-DE-miRNAs of the same cluster displayed the same DE pattern in a particular tissue. However, a cluster can have different expression profiles among tissues. Thus, a unique pattern for all tissues was not identified. Considering the genomic contigs, the contig NODE_615561 contained 4 novel mature B-DE-miRNAs that were upregulated in the brains of females. These miRNAs belong to 2 novel pre-miRNAs (novel-37864 and novel 37866) that are similar to mir-27c and mir-23c from *A. burtoni* seeds. In contig NODE_65503, 3 novel miRNA precursors (novel-5423, novel-5425 and novel 37866) with 5 downregulated B-DE-miRNAs were detected in the gonads of females, which are similar to the mir-

217 seeds from *C. carpio* and the mir-216a and mir-216b seeds from *A. burtoni.* However, the cluster on contig NODE_65503 also contained a mature B-DE-miRNA that was upregulated in the muscle of females; therefore, a cluster shows DE profiles according to the tissue (**Figure 3c and Table 1**). The distance between the pre-miRNAs genes is described in the start and end of pre-miRNAs sequences on **Table 1**. Although each tissue has a different miRNA expression profile, no difference in *drosha* and *dicer* gene expression was observed in B⁺ samples (**Additional File 5**).

Figure 3 Differential expression analysis. **a** Nonredundant differentially expressed miRNAs in B⁺ samples (green represents upregulated miRNAs and red represents downregulated, considering p <0.05 to >1.5 fold change). **b** Venn diagram of DE miRNAs among tissues. **c** Representation of two clustered DE-B-miRNAs structural organization in genomic contigs.

Table 1 – Clustered miRNAs composed by downregulated (↓) and upregulated (↑) B-DE-miRNAs in brain (BR), gonad (G), muscle (MU), female (F), male (M). FG, FoldChange.

Contig	Pre-miRNA name	DNA strand	Pre-miRNA start-end	Seed similarity	Mature B-DE-miRNA	Expression in B+ samples	FC in B+ sampes
NODE_173406	mir-99a	-	9625-9682	nbr-mir-99a	mir-99a-5p	↓ GO_F	-1.23
	novel_13044	+	9626-9684	gmo-mir-100b-5p	novel_13044-3p	↑ BR_M	+1.38
NODE_615561	novel_37864	-	499-563	abu-mir-27c	novel_37864-5p	↑ BR_F	+2.78
					novel_37864-3p	↑ BR_F	+2.49
	novel_37866	-	714-773	abu-mir-23c	novel_37864-5p	↑ BR_F	+2.68
					novel_37864-3p	↑ BR_F	+2.51
NODE_65503	novel_5423	-	4439-4498	ccr-mir-217	novel_5423-5p	↓ GO_F	-2.16
					novel_5423-3p	↓ GO_F	-1.76
	novel_5425	-	5012-5075	ccr-mir-216a	novel_5425-5p	↓ GO_F	-1.11
	novel_5427	-	5708-5768	ccr-mir-216b	novel_5427-3p	↓ GO_F	-2.19
					novel_5427-5p	↓ GO_F	-1.70
						↑ MU_F	+2.50
NODE_843581	mir-194	-	58381-58436	ccr-mir-194	mir-194-3p	↓ MU_M	-2.33
	mir-192	-	58185-58246	ccr-mir-192	mir-192-3p	↓ MU_M	-2.67
NODE_91705	mir-212-2	-	2623-2690	dre-mir-212-2	mir-212-2-5p	↓ GO_F	-1.89
	novel_7483	-	4415-4473	gmo-mir-2184	novel_7483-5p	↑ BR_F	+1.27

The miRNA 3'UTR interaction and protein-protein interaction (PPI) network

The 3'UTR miRNA binding site was chosen to predict the *A. latifasciata* miRNA targets to restrict and avoid false positives results (Oliveira et al., 2017). Other softwires were tested to miRNA:mRNA prediction, but they showed a huge number of interactions, being a problem to the filtering. The miRNA:mRNA interactions were predicted (**Additional File 7**) based on the miRNAs and 3'UTRs from the *A. latifasciata* transcriptome. Based on the best scores (< -0.2, following the software developer recommendations (Ulitsky et al., 2012), 2,080,942 interactions were identified in the brain, 2,061,604 in gonads and 2,016,807 in muscle.

The miRNA:3'UTR interactions detected for the B-DE-miRNAs formed a list composed of B-DE-miRNAs and targets (mRNAs)**.** The protein annotations of these transcripts (**Additional File 8**) from each tissue are presented in a Venn diagram, and 960 proteins were shared among all compared groups (defined as B-related

proteins) (**Figure 4a).** The protein-protein interaction (PPI) network from Biogrid was downloaded to evaluate whether B-related proteins had functional interactions (**Additional File 9**). The PPI indicates the physical and high specific contact of two or more proteins. Due the lack of fish protein-protein interaction data base, the interactions were filtered based on the human set. As reported in several studies, there are several miRNA targets conserved among organisms (Bartel, 2018; Rajewsky, 2006). In this way, our extrapolation report mostly conserved interactions that could be confirmed by presence in the data bases, as the online TargetScan (Ulitsky et al., 2012).

The original Biogrid PPI network contained 544,163 interactions (edges) between 24,913 proteins (nodes). After the experimental detection, redundant edges and species filtering procedures, 294,199 edges between 14,540 nodes were retained. Based on the 960 B-related proteins whose genes are B-DE-miRNA targets, and B genes identified in a previous study (Valente et al., 2014), a subnetwork extraction revealed 177 nodes and 335 edges in the Biogrid filtered network. The 177 nodes from the subnetwork are targets of 57 B-DE-miRNAs from the miRNA:3'UTR analysis. Finally, all miRNA targets and interactions were added to the subnetwork, resulting in 22,477 edges and 281 nodes (considering both proteins and miRNAs), which was called “B-miR-net” (**Figure 4b and Additional File 9**).

Despite the ability of B-DE-miRNAs to connect to several mRNAs in B-miR-net, proteins that are regulated in consensus between different tissues are interconnected with each other (**Figure 4b – red axis**). Interestingly, among the B-related proteins, connections exist other proteins that may by indirectly influenced by B-DE-miRNAs (**Figure 4b – edges between red and blue axes**). Additionally, 42 of the 102 genes reported on the B chromosome in a previous study (Valente et al., 2014) are in the B-miR-net, as they are B-DE-miRNA targets or interact with other proteins (**Table 2**). Nevertheless, 18 of these B-genes exhibit more than 80% of integrity in their coding region, meaning potentially translated proteins. The other 24 B-genes displayed less than 80% of integrity that may generate truncated transcripts what indicates protein translation problems (**Table 2** and **Additional File 7**).

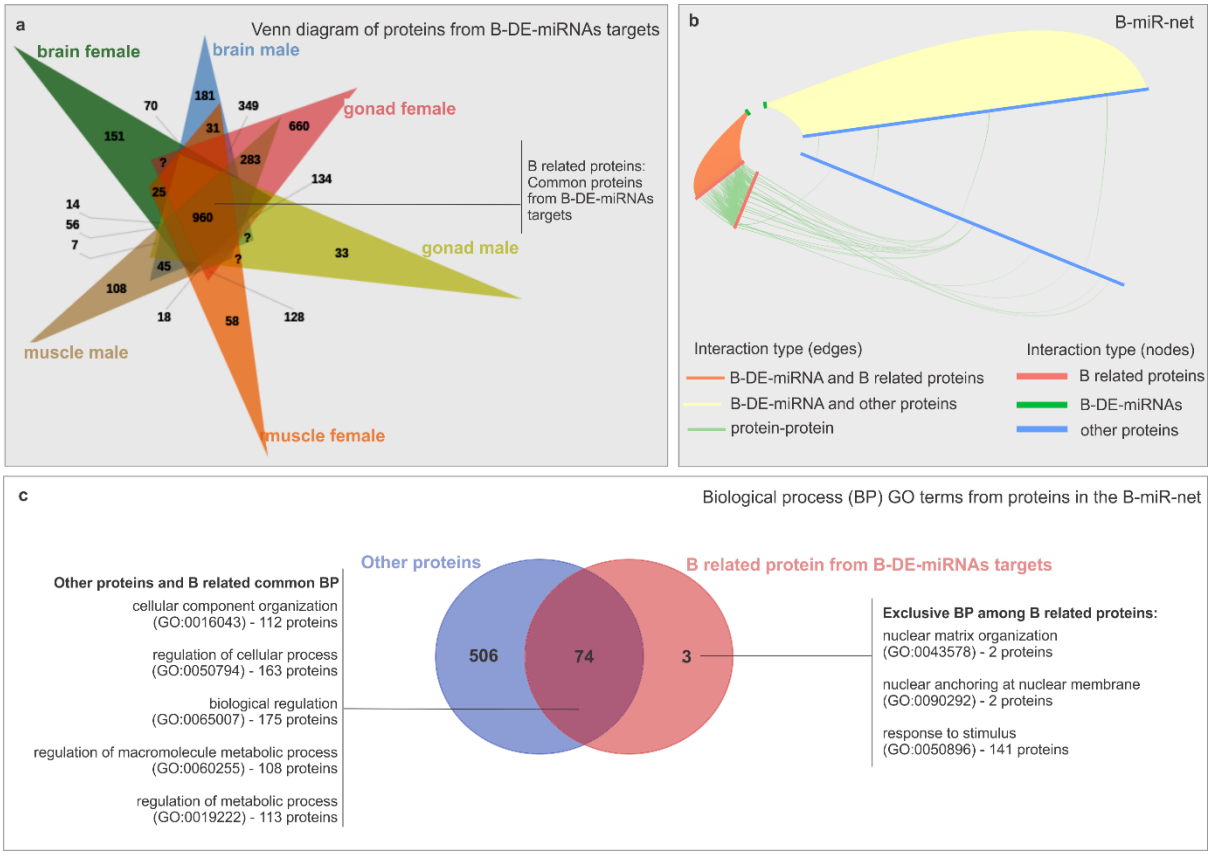

Figure 4 The B-mir-net. **a** Venn diagram of common protein annotation from B-DE-miRNAs targets. **b** The network axis represents the nodes ordered by degree (from low to high beginning from the center), and edges represent the connection between the axis nodes. Axes are duplicated to show the interconnections between its own subjects: red axis, B-related proteins (miRNA targets) and B genes reported in previous study (Valente et al., 2014); green axis, mature miRNAs; blue axis, tissue-specific miRNA targets; orange edges, miRNAs interacting with proteins commonly regulated by miRNAs in all analyzed tissues; yellow edges, miRNAs interacting with tissue-specific proteins; green edges, protein-protein interactions. **c** Venn diagram of GO terms on B-related proteins and all proteins from predicted *A. latifasciata* targets.

A Gene Ontology (GO) analysis was performed (**Additional Files 8**), and the proteins were analyzed in the groups of (I) B-related proteins (**Figure 4b – red axis**) and (II) all proteins whose genes are targets of *A. latifasciata* miRNAs (**Figure 4b – blue axis**). The results revealed 74 biological processes that were shared among the analyzed groups, meaning the common processes between B-related proteins and all *A. latifasciata* proteins (**Figure 4c**, **Additional File 10 – S2**). In other words, processes that are commonly present in the cell might be affected by B chromosome presence by miRNA targeting. However, three specific GO terms were found in the B-related protein group, which indicates exclusive processes in the PPI network for

genes controlled by the B-DE-miRNAs in all analyzed tissues. These process terms are nuclear matrix organization (GO:0043578), nuclear anchoring at nuclear membrane (GO:0090292), with members of these two terms being the SUN1 (*SUN domain-containing protein 1*), SUN2 (*SUN domain-containing protein 1*) and SYNE1 (*syne1*) proteins and response to stimulus (GO:0050896), with 141 proteins members (**Additional File 9 – S1**) (Figure 4c). Furthermore, a B-gene, ATP-binding cassette sub-Family A member 1 (*abca1*) with 80.25% of integrity, belongs to response of stimulus biological process. This gene is target of B^+ upregulated miRNAs in brain, gonad and muscle as describe in the **Figure 5**. As highlighted in the **Figure 5a**, these two biological processes exclusive of B-related proteins are connected to each other by miRNA targeting.

Furthermore, ABCA1 and SUN2 interact with the proteins CDC42 (cell division control protein 42 homolog) and SKP1 (S-phase kinase associated protein 1), respectively. These two proteins also have GO terms which are common to B-related proteins and all the other proteins. This indicates potential biological processes which occurs in the cells and might be affected by the B chromosome presence. The potential consequences of this interaction under B chromosome presence are developed further in the Discussion section.

Table 2 – B-genes present in the B-miR-net. B related proteins – proteins which either are targets of B-DE-miRNAs or interact with these targets. Other proteins: proteins which interact each other and are *A. latifasciata* miRNA targets.

	Gene name	Protein Symbol	Integrity (%)
B-related proteins	apoptosis regulator Bcl-2-like	BCL2	100
	coxsackievirus and adenovirus receptor-like	CXAR	100
	spindle and kinetochore-associated protein 1-like	SKA1	100
	kinesin-like protein KIF11-like	KIF11	96.13
	zinc finger protein 771-like	ZNF71	89.37
	serine/threonine-protein kinase RIO3-like	RIOK3	81.57
	ATP-binding cassette sub-family A member 1-like	ABCA1	80.25
	aurora kinase A-B-like	AURKA	77.16
	centromere-associated protein E-like	CENPE	73.92
	zinc finger protein 836-like	ZN836	64.52
	histone-lysine N-methyltransferase MLL3-like	KMT2C	56.16

Other proteins	butyrophilin subfamily 2 member A1-like	BT2A1	100
	butyrophilin-like protein 2-like	BTNL2	100
	myosin-10-like	MYH10	100
	polymeric immunoglobulin receptor-like	PIGR	100
	serine protease 23-like	PRS23	100
	peptide chain release factor 1-like. mitochondrial-like	RF1ML	100
	VIP peptides-like	VIP	100
	zinc finger protein 879-like	ZN879	100
	ATP-dependent RNA helicase DDX51-like	DDX51	92.22
	CD209 antigen-like	CD209	90.13
	protocadherin-10-like	PCD10	89.58
	heterogeneous nuclear ribonucleoprotein Q-like	HNRPQ	79.90
	leucine-rich repeat-containing protein 30-like	LRC30	71.93
	synaptonemal complex protein 2-like	SYCP2	70.97
	extracellular calcium-sensing receptor-like	CASR	69.46
	protein NLRC3-like	NLRC3	68.24
	guanine nucleotide-binding protein G	GNAI1	67.52
	vascular cell adhesion protein 1-like	VCAM1	65.79
	torsin-1A-interacting protein 2-like	ELOF1	64.25
	zinc finger protein 782-like	ZN782	63.42
	zinc finger protein 678-like	ZN678	63.01
	V-set domain-containing T-cell activation inhibitor 1-like	VTCN1	63.00
	poly [ADP-ribose] polymerase 14-like	PAR14	61.53
	DNA-directed RNA polymerase E subunit 1-like	RPA49	61.01
	endonuclease domain-containing 1 protein-like	ENDD1	58.21
	SAM domain and HD domain-containing protein 1-like	ESPL1	57.51
	targeting protein for Xklp2-A-like	TPX2	52.30
	interferon-induced very large GTPase 1-like	GVIN1	51.55
	sterile alpha motif domain-containing protein 12-like	SAM12	51.27
	xylulose kinase-like	XYLB	50.52
	glucose-6-phosphate 1-dehydrogenase-like	G6PD	50.43

Discussion

In the last few years, B chromosome science has undergone rapid advances due to the wide application of genomics and bioinformatics tools and functional approaches, including analyses of nonmodel species (Ruban et al., 2017). Here, we advanced the characterization of the miRNome in *A. latifasciata*, with a focus on the B chromosomes. Initial studies described approximately 200 miRNA genes in fish

species (Brawand et al., 2015a), and later, this number was increased to 400 genes with the advances in large-scale DNA and RNA sequencing (Xiong et al., 2019). Here, we identified more than 700 pre-miRNAs, 34% of which are known miRNAs identified in miRBase (**Figure 1**). A set of new pre-miRNAs in the *A. latifasciata* genome was described, confirming the rapid evolution of genomes in cichlid species (Brawand et al., 2015a; Franchini et al., 2019; Xiong et al., 2019). New miRNAs could arise and be lost quickly mainly between related groups (Xiong et al., 2019). In this manner, the miRNAs could be tightly involved to the diverse adaptation of Cichlidae family (Brawand et al., 2015b; Xiong et al., 2019). Some of novel pre-miRNA seeds are similar to cichlid seed sequences, and these new miRNAs are probably isomiRs originated from duplications or as products of RNA editing (Bartel, 2018). The new miRNAs are organized in clusters, which are usually controlled by the same factors to act in the same or related pathways (Xiong et al., 2019). The *A. latifasciata* species-specific isomiRs might perform new functions related to the B chromosome, as we can observe in the **Figure 5a**.

Concerning the investigation of small sequences in the B chromosome, it was not possible to describe a pre-miRNA in the B extra element based on the adopted approaches. We pointed two reasons: the miRNAs evolution and the limitations of Illumina sequencing to investigated small sequences with low coverage. One of the main reasons for the lack of miRNAs is probably related to the B chromosome structure, which in *A. latifasciata* originated from a mosaic of duplicated sequences that underwent mutations over time (Valente et al., 2014). Even though the difficulty of screening new miRNA sequences in a degenerated element (as the B chromosome), given the limitations of sRNAseq and genomic approaches, we cannot discard to find out small RNAs sequences in the B chromosome. Our results offered a miRNA genomic annotation and expression analysis of a new cichlid fish. This data can improve future studies by association with other genomic sequencing technologies that cover the B chromosome in a better way, such as Pacbio (Clark et al., 2018) and flow sorting (Karafiátová et al., 2020).

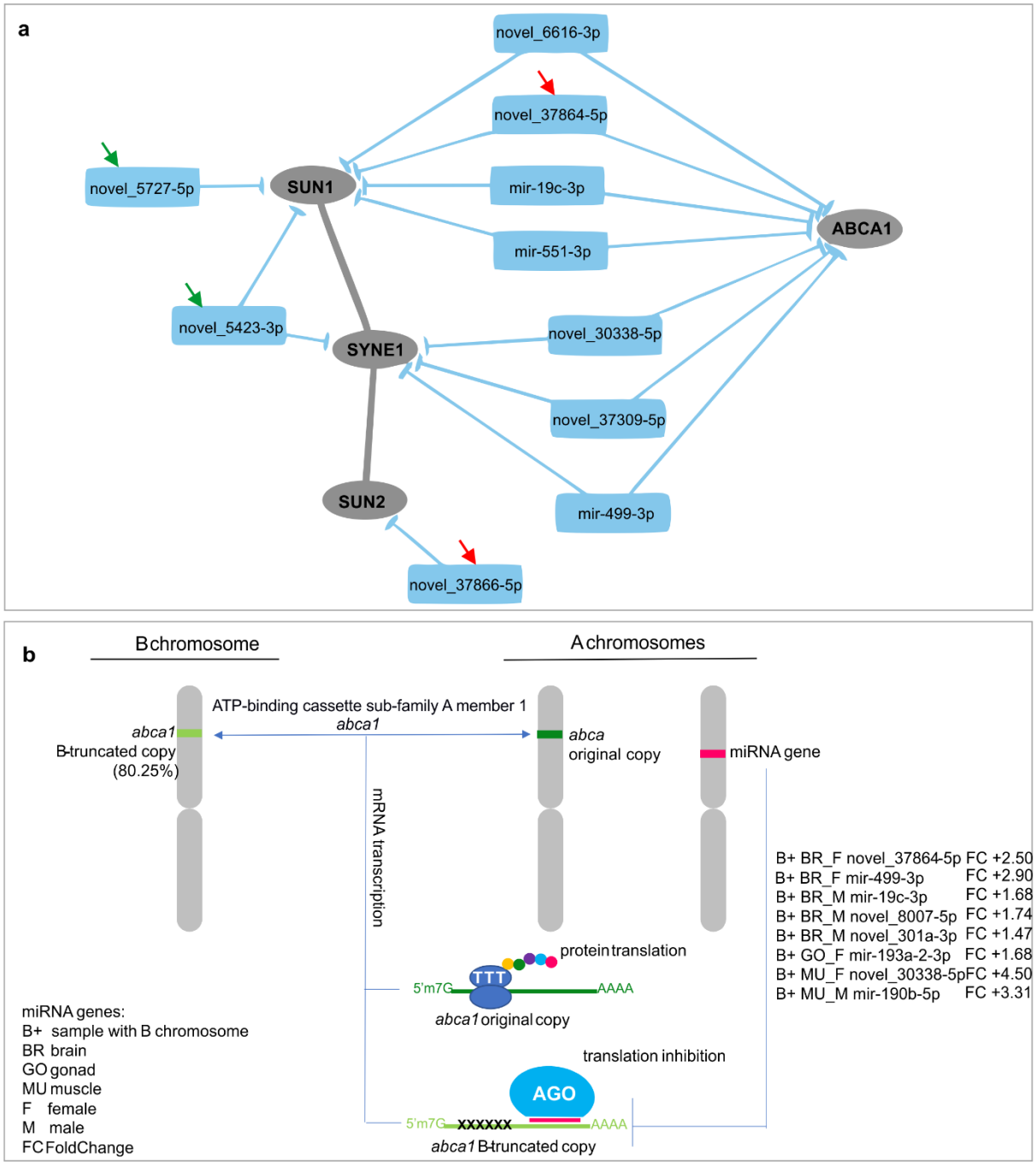

Figure 5 Potential activity of small RNA genes originated from A chromosomes and their action over the B chromosome. **a** B-DE-miRNAs targeting the genes of exclusive B-related protein GO terms. The arrows with the same color indicate clustered miRNAs. **b** The truncated B-genes of the B-miR-net compete to their original A copy and are controlled by B-DE-miRNAs in the way to benefit B chromosome maintenance.

We observed differences in the expression of miRNAs encoded by the regular chromosomes complement (B-DE-miRNAs) in B chromosome carriers. Similar results were also observed in maize (Huang et al., 2020). The mechanism by which

B chromosomes affect the expression of A complement miRNA genes is unclear. The presence of active genes related to miRNA biogenesis in the B chromosome might explain these variations. Argonaute (AGO) proteins are essential for processing small RNAs (Bartel, 2018; Moazed, 2009), and although AGO-like protein genes have already been observed on the B chromosomes of *Secale cereale* (Ma et al., 2017), copies of these genes and other genes related to miRNA biogenesis were not found in *A. latifasciata* B chromosomes (Jehangir et al., 2019; Valente et al., 2014). Moreover, the expression of A copies of *drosha* and *dicer* genes, the main regulators of miRNA biogenesis, were not affected by B chromosomes (**Additional File 5**). Thus, the miRNA pathway does not appear to have changes under the B chromosome presence.

The B-DE-RNAs are specific to each tissue (**Figure 3**), but their targets are usually the same (**Figure 4a**). In other words, we have different B-DE-miRNAs targeting common proteins among the tissues and sex. Further, some of these proteins interacts each other in biological processes with relation to B chromosome presence (**Figure 4c**). Among the target functions, the cell cycle pathway has attracted attention, since genes related to the cell cycle and chromosome segregation are registered as located on B chromosomes (for a review, see (Benetta et al., 2019)). What would explain the differential miRNA expression in the presence of the B chromosome? Gene fragments with different degrees of integrity are present in the *A. latifasciata* B chromosome (Valente et al., 2014). We suggest, if these truncated gene copies are transcribed, more binding sites will be available to miRNAs, generating regulatory competition with the A chromosome set (**Figure 5**) (Banaei-moghaddam et al., 2015). We reported a B-gene, *abca1,* which has 80% of integrity and is a B-DE-miRNA target (**Table 2** and **Figure 5a**). Thus, according to the miRNAs controlling these genes, even truncated copies might display increased activity in the cell, further investigations are needed to confirm such assumption. Finally, because the miRNA targets are similar in different tissues, the gene functions affected by their regulation may favor the B chromosome to some extent (Banaei-Moghaddam et al., 2013; Houben et al., 2014).

Conclusions

We presented the *A. latifasciata* miRNome and compared it with other available miRNA databases. Additionally, several miRNAs were DE in the brain, gonads and muscle in the B^+ samples and shared common genes as miRNA targets. The differentially expressed miRNAs detected in the presence of the B chromosome are not the same among the tissues, but the miRNA targets are involved in the same biological processes. Thus, we suggest that the B chromosome influences the cellular environment using miRNAs as a posttranscriptional control process that is probably for its own benefit of B drive and maintenance.

Availability of data and materials

The *A. latifasciata* B^+ assemblies and B-blocks are available in a previous study under Bioproject PRJN369442 access (Jehangir et al., 2019) and can be visualized on https://sacibase.ibb.unesp.br/. The *A. latifasciata* miRNAs sequences are shown in Additional Files. The targets sequences have been published (Nakajima, 2019). The small RNA-seq libraries are deposited in the NCBI database (accession numbers from SRR13040679 to SRR13040710b).

References

The references used in this chapter will be presented at the end of thesis.

Additional Files description and access

Additional File 1 miRDeep2 – *Astatotilapia latifasciata* miRNA prediction. **link**

Additional File 2 Html file of miRNA prediction. miRDeep2 – miRNA prediction in the B^+ assembly using B^- RNAseq which did not align in the *A. latifasciata* genome. **link**

Additional File 3 Html file of miRNA prediction. miRDeep2 – miRNA prediction in the B^+ assembly using B^+ RNAseq which did not align in the *A. latifasciata* genome. **link**

Additional File 4 Pre-miRNAs sequences on the B^+ assembled genome. **Figure S1** -Genome assembly bias. **link**

Additional File 5 Expression of *drosha* and *dicer* genes. **link**

Additional File 6 Html files containing miRNA prediction in B-blocks, separated by RNA-seq sample. **S1** miRDeep2 – miRNA prediction in B-blocks using brain female B^- RNA-seq. **S2** miRDeep2 – miRNA prediction in B-blocks using brain female B^+ RNA-seq. **S3** miRDeep2 – miRNA prediction in B-blocks using brain male B^- RNA-seq. **S4** miRDeep2 – miRNA prediction in B-blocks using brain male B^+ RNA-seq. **S5** miRDeep2 – miRNA prediction in B-blocks using gonads female B^- RNA-seq. **S6** miRDeep2 – miRNA prediction in B-blocks using gonads female B^+ RNA-seq. **S7** miRDeep2 – miRNA prediction in B-blocks using gonads male B^- RNA-seq. **S8** miRDeep2 – miRNA prediction in B-blocks using gonads male RNA-seq. **S9** miRDeep2 – miRNA prediction in B-blocks using muscle female B^- RNA-seq. **S10** miRDeep2 – miRNA prediction in B-blocks using muscle female B^+ RNA-seq. **S11** miRDeep2 – miRNA prediction in B-blocks using muscle male B^- RNA-seq. **S12** miRDeep2 – miRNA prediction in B-blocks using muscle male B^+ RNA-seq. **link**

Additional File 7 Multiple data sheets of samples separated by tissue and sex describing the B-DE-miRNA and its target. **link**

Additional File 8 Multiple data sheets of samples separated by tissue and sex describing the B-DE miRNA target gene and protein names – recovered from *A. latifasciata* transcriptome annotation (Nakajima, 2019). **link**

Additional File 9 Folder containing the .tsv files of networks. **S1.tsv** PPI network filtered from the Biogrid database. **S2.tsv** PPI of all proteins from *A. latifasciata* miRNA targets. **S3.tsv** B-related proteins and B genes subnetwork. **S4.tsv** B-miR-net. **link**

Additional File 10 GO datasets. **S1 Exclusive** GO terms for proteins derived from B-DE-miRNA targets and B genes. **S2** Common GO terms among B-related proteins and other proteins. **S3** GO terms for all *A. latifasciata* targets in the PPI network. **link**

Additional File 11 Samples and primers information. **Table S1** RNA-seq quality and filtering. **Table S2** Normalized reads mapped in the genome**. Table S3** miRBase fish database entries used to create the reference miRNA list. **Table S4** Primers used for genomic PCR, genomic qPCR and RT-qPCR. **link**

3.2 Capítulo 2

"Now, here, you see, it takes all the running you can do,
to keep in the same place. If you want to get somewhere else,
you must run at least twice as fast as that!"

— Red Queen quote in Through the looking-glass - Lewis Carroll

First characterization of PIWI-interacting RNA clusters and their role in genome protection against the accumulation of transposable elements in B chromosome

Jordana Inácio Nascimento-Oliveira[1], Adauto Lima Cardoso[1], Ivan Rodrigo Wolf[1], Rogério Antônio de Oliveira[2], Cesar Martins[1]

[1] Department of Structural and Functional Biology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, SP, Brazil

[2] Department of Biostatistics, Plant Biology, Parasitology and Zoology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, SP, Brazil

Abstract

B chromosomes are extra elements that are found in several eukaryote species. Usually, they do not express a phenotype in the host. Nonetheless, bioinformatics advances over the last decades have made it possible to describe several genes and molecular functions related to B chromosomes. These advances allow investigations of the relation between the B chromosome and the host and understanding of why and how this element has been sustained in genomes. However, there is a lack of knowledge concerning the dynamics of transposable element (TE) control in B-carrying cells, considering that TEs are highly abundant in extrachromosomal elements. Thus, the present study characterized PIWI-interacting RNA (piRNA) clusters and pathway responsible for silencing the mobilization of transposable elements in gonads of *Astatotilapia latifasciata* cichlid fish carrying the B chromosome. Through small RNA-seq and genome assembly, we predicted and annotated the piRNA clusters in the *A. latifasciata* genome. We observed that these clusters had a biased expression related to sex and B chromosome presence. Furthermore, three piRNA clusters, named *curupira*, were identified in the B chromosome. Two of them were exclusively expressed in gonads with the B chromosome. The composition of these *curupira* sequences was derived from LTR, LINE and DNA elements, representatives of old and young transposition events in the *A. latifasciata* genome and B chromosome. Testis with the B chromosome showed a decrease in reverse transcriptase and transposase expression, which could be related to the efficiency of piRNA activity in the presence of the B chromosome. The presence of the B chromosome also impacted expression of the

piRNA pathway genes. The mitochondrial cardiolipin hydrolase-like (*pld6*) gene is present in the B chromosome, as previously reported, and an increase in its expression was detected in gonads with the B chromosome. Due to the high composition of transposable elements in the B chromosome, it was possible to investigate the piRNA origin from these jumping genes. In this way, we suggest that the B chromosome is able to evolve its own genome guardians to prevent the decontrol of TEs mobilization. Furthermore, we also detected a bias of the B chromosome presence over *A. latifasciata* piRNA clusters and pathway genes.

Background

B chromosomes are extra elements found in several species, including fungi, plants and animals (D'Ambrosio et al., 2017). They were first described as a supernumerary chromosome in an insect species more than a century ago (Wilson, 1907). "B" is a reference to distinguish this extra element from regular chromosome complement (the "A" complement) (Randolph, 1941). Current estimates show that the B chromosome could be present in approximately 15% of karyotyped eukaryotic species (D'Ambrosio et al., 2017).

In general, B chromosomes arise from a proto-chromosome derived from an abnormal event in meiosis, such as chromosome breaks during nondisjunction (Houben, 2017). These proto-B chromosomes are later invaded by A complement and/or organellar (mitochondrion and chloroplast) sequences, which increase the diversity of their genomic content (Martis et al., 2012). For this reason, B chromosomes are often called "mosaic elements" and have accumulated several types of sequences, such as satellite DNA (Ruiz-Ruano et al., 2017), transposable elements (Lamb et al., 2007), pseudogenes (Banaei-Moghaddam et al., 2013), retrogenes (Carmello et al., 2017), protein coding genes (Ma et al., 2017; Navarro-Domínguez et al., 2017; Trifonov et al., 2013), long noncoding RNAs (Ramos et al., 2017) and miRNAs (Huang et al., 2020).

The B chromosome is generally recognized as an inert element without genetic activity based on its heterochromatic characteristics (Camacho et al., 2000). However, current molecular biology and bioinformatics advances have made it possible to identify the expression of B chromosome sequences (Ruban et al., 2017), also known as "B genes". The B chromosome of a fungal species, for

example, carries a gene that confers antibiotic resistance against a host compound, making the individuals more infectious (Miao et al., 1991). Additionally, in rye, RNA slicer activity for an Argonaute-like B chromosome copy has been observed *in vitro* (Ma et al., 2017), and some B chromosome peptides have been identified by mass spectrometry (Ma et al., 2021). Furthermore, several studies report that B chromosomes can affect the expression of A complement sequences (Huang et al., 2020; Nascimento-Oliveira et al., 2021) or impact various biological processes in the cell (Valente et al., 2017).

Repetitive DNAs are abundant in B chromosomes, with emphasis on the presence of large amounts of transposable elements (TEs) (Coan & Martins, 2018; Peng & Cheng, 2011). Regardless, the impact of transposition on the host genome due to the presence of the B chromosome is still not well understood. Although TEs are important for genome evolution, mobilization must be controlled to maintain genome integrity in order to avoid deleterious mutations (Kofler, 2019).

PIWI-interacting RNAs (piRNAs) are small noncoding sequences that are responsible for targeting TEs and promoting their silencing through the piRNA pathway in animals (Gainetdinov et al., 2018). The piRNAs originate from degenerated TE regions, the TE “junkyards” (Han et al., 2015), forming piRNA clusters that are first transcribed as long RNAs and later processed in the cytoplasm via a Dicer-independent pathway (Houwing et al., 2007; Moazed, 2009). In the cytoplasm, piRNA cluster transcripts are processed into small RNAs (piRNAs) in two ways: primary piRNAs and the ping-pong cycle. In primary or phased piRNA processing, the endonuclease mitochondrial cardiolipin hydrolase (*pld6*), a *Zucchini* homolog in *Drosophila*, cleaves piRNA cluster transcripts into mature piRNAs, allowing these sequences to bind PIWI proteins and form an RNA-inducing silencing complex, halting the mobilization of target TEs. Additionally, PIWI proteins are able to process piRNA cluster transcripts into piRNA sequences through the ping-pong cycle, which increases piRNA variability (Gainetdinov et al., 2018). Fish species carry two PIWI proteins encoded by *piwi-like* genes (*piwil1* and *piwil2*) (Song et al., 2019; Tao et al., 2016; Yi et al., 2014). In addition to the cytoplasm silencing pathway, the PIWI-piRNA complex can be directed to the nucleus to silence TE transcription, attracting methylation machinery to the TE chromosome region (Czech et al., 2018; Iwasaki et al., 2015). For this reason, piRNAs are commonly called

“genome guardians”, as they prevent the uncontrolled mobilization of TEs and thus help maintain genome integrity (Ishizu et al., 2012; Malone & Hannon, 2009).

How the B chromosome and its extra amount of TEs affect piRNAs or *vice versa* has not yet been investigated. Few studies have reported the relationship of the B chromosome and small RNAs, most of which are related to microRNAs (Huang et al., 2020; Li et al., 2017; Nascimento-Oliveira et al., 2021). Thus, this study was designed to investigate the piRNA and B chromosome relationship by examining the African Cichlid fish species *Astatotilapia latifasciata* as a model organism.

The B chromosome of *A. latifasciata* is a metacentric chromosome similar in size to the first pair of A complement (Poletto et al., 2010). Heteromorphic sex chromosomes were not identified in this species, and the B chromosome can be found in one-third of the population. Previous studies have already characterized the B chromosome and *A. latifasciata* genome through molecular and large-scale approaches, such as coding gene annotation (Valente et al., 2014), transposable element content (Coan & Martins, 2018), epigenetic DNA modifications (Cardoso et al., 2019), genome assembly (Jehangir et al., 2019), proteomics and transcriptomics (Nakajima, 2019), lncRNAs (Ramos et al., 2017) and miRNAs (Nascimento-Oliveira et al., 2021). Among the coding genes, *pld6*, which is important for primary piRNA generation, is present in the B chromosome of *A. latifasciata* (Valente et al., 2014).

In this manner, the present study reports how the accumulation of TEs through B chromosome evolutionary life could be followed by piRNA pathway expression. We suggest that the B chromosome accumulates TEs that are controlled by its own piRNAs. This coevolution generates a cycle that allows B chromosome maintenance, perpetuation and accumulation in cells and individuals.

Materials and Methods

Sample collection and small RNA sequencing

DNA and RNA samples were obtained from the *A. latifasciata* fish population maintained at the aquarium facility at the Integrative Genomics Laboratory of São Paulo State University, Botucatu (SP), Brazil (Protocol no. 769–2015). All the fishes were genotyped for B presence/absence using the previously developed marker for B chromosomes (Fantinatti & Martins, 2016) through extracted caudal fin DNA (Sambrook & Russell, 2006) and maintained in different aquariums until their use.

These materials were also used for quantitative PCR (qPCR), as described below. The fishes were euthanized by immersion in eugenol 1% for three minutes long. Total gonadal RNA of three females and three males with B chromosomes (B+) and without B chromosomes (B-) was collected, totaling 12 samples, and extracted following the TRIzol™ protocol recommendations. These samples were shipped for large-scale small RNA sequencing with next-generation Illumina-HiSeq sequencing (Sequencing Service at LC Sciences - Houston, TX, EUA).

Library filtering and piRNAs characterization

The raw data from sRNA-seq were filtered using the FastX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), trimming the sequences, removing the adapters and selecting reads ranging from 24 to 35 nucleotides in length to piRNA mature transcripts (Rosenkranz & Zischler, 2012). This small RNA dataset is available at the NCBI accession numbers SRR13040679 to SRR13040710.

The filtered and selected reads of each library (female without B, FB-; female with B, FB+; male without B, MB- and male with B, MB+) were collapsed using the script TBr2_collapse.pl (Rosenkranz & Zischler, 2012) and aligned against the *A. latifasciata* reference genome (Jehangir et al., 2019) with RNAmapper.pl (Rosenkranz et al., 2015). These steps are mandatory for piRNA prediction as follows.

A. latifasciata piRNA cluster prediction was performed using proTRAC (Rosenkranz & Zischler, 2012) based on the following aspects: 1) *A. latifasciata* genome and 2) genome annotation (Jehangir et al., 2019); 3) small RNA-seq mapping of the genome (from the RNAmapper.pl step); 4) *A. latifasciata* RepeatMasker annotation (Coan & Martins, 2018) and 5) *A. latifasciata* transcriptome coding gene annotation (Nakajima, 2019). The predicted piRNA clusters are shown in **Additional File 1**.

To analyze miRNA and piRNA distributions over the *A. latifasciata* genome, the genomic localization (*A. latifasciata* contigs) of piRNAs (this study) and miRNAs (Nascimento-Oliveira et al., 2021) was compared on a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Identification of piRNAs clusters on B chromosome

The genomic localization of predicted piRNA clusters was prospected to validate the presence of piRNAs in the B chromosome through the “coverage ratio” (Valente et al., 2014). This strategy consists of aligning the B- and B+ genomic reads against the reference genome (the genomic contigs with piRNA clusters from the *A. latifasciata* assembly). Due to the B chromosome duplicated sequence composition, higher coverage of B+ genomic reads (at least twice) than B- genomic reads alignment is expected. The reads alignment was performed using Bowtie (Langmead & Salzberg, 2012) default parameters, and the bam files are available at the SaciBase Database in Jbrowser viewer (https://sacibase.ibb.unesp.br/index.html). The contigs with a B+/B- coverage ratio >2 were selected as putative sequences belonging to the B chromosome. To visualize the depth coverage, the nonassembled reads (B- and B+) from the B selected contigs containing piRNA clusters that present higher coverage in the B+ alignments were extracted and submitted to bedcov analysis using Bedtools v2.29.2 (Quinlan & Hall, 2010). The depth coverage of the B selected contigs was visualized using the R package Sushi (Phanstiel et al., 2014).

***Pld6* B chromosome genomic copy presence validation**

Mitochondrial cardiolipin hydrolase-like (*pld6*) gene was detected in the *Metriaclima zebra* genome (v0, bouillabase.org, last accessed on October 16, 2020), and this genomic localization was used to check *pld6* in *A. latifasciata* based on an alignment of B- and B+ samples against the *M. zebra* genome available in SaciBase (https://sacibase.ibb.unesp.br/index.html). The presence of B+ was detected by “coverage ratio analysis” as previously described.

To validate the *pld6* duplication in the B chromosome, we performed the Gene Dose Ratio (GDR) through qPCR using B- and B+ DNA samples, as follow cycles: 95 °C for 20 seconds, 34 cycles (95 °C for 3 seconds and 60 °C for 30 seconds); melting curve of 95 °C for 15 seconds, 60 °C for one minute, 95 °C for 15 seconds. qPCR was executed using SYBR Green qPCR Master Mix (High ROX) Ampliqon (HY-K0521) detection measured by the ΔΔCq method (de Santis et al., 2011) based on the hypoxanthine phosphoribosyl transferase 1 (*hprt1*) single copy gene as a reference, calculated by Q-Gene software (Simon, 2003).

Conventional PCR was executed using primers for the *pld6* gene and *pld6* B-specific copy amplified from genomic DNA of B- and B+ genotyped individuals. Conventional genomic PCR was performed using recombinant T*aq* DNA Polymerase (Invitrogen-10342-053) submitted to cycles at 94 °C for 5 minutes, 35 cycles (94 °C for 1 minute, 50 °C to 60 °C for 45 seconds, 72 °C for 10 minutes) and 72 °C for 10 minutes. The result was verified by agarose gel electrophoresis (1%). The primers used are described in **Additional File 2**.

Molecular evolution analysis

To construct a phylogenetic tree for the molecular evolution analysis, the mRNA *pld6* sequences of 40 vertebrate species and a fly were downloaded from NCBI. The sequences cover the following groups: an insect sequence, 7 mammal species, 4 Chondrichthyes species, 27 teleost species (including 5 cichlids) and the two *pld6* sequences from the *A. latifasciata* transcriptome. The *pld6 A. latifasciata* sequences were obtained through a standalone Blast 2.2.31 (Madden, 2013) based on the *M. zebra* mRNA against the assembled *A. latifasciata* transcriptome (Nakajima, 2019). *Drosophila melanogaster* was included due to the good description of piRNA biogenesis, which is a good parameter to compare with the available data. We also included Chondrichthyes because there was no study about *pld6* in the group. We included all the available cichlid sequences due to our interest in *A. latifasciata pld6* characterization. The access ID of all sequences and the *A. latifasciata* sequence are described in **Additional File 3** and **Additional File 4**.

The coding sequence (CDS) region of each *pld6* gene was extracted using Geneious (https://www.geneious.com/). The CDSs were submitted to a codon alignment, also in Geneious, and used to construct a neighbor-joining tree based on the Tamura-Nei genetic distance model. The *pld6* codon alignment and the tree were used to perform several tests of molecular evolution analysis by hypothesis testing using the HyPhy (https://www.hyphy.org/) program. Based on the BUSTED method, it was possible to detect whether the *pld6* gene has experienced episodic selection events (Murrell et al., 2015). From these results, the aBSREL method was employed to determine which branch in the tree was under positive selection (Smith et al., 2015). In addition to the previously branch site models, we also employed FEL and

FUBAR to detect which sites in the CDS could be under pervasive selection, either positive or purifying selection (Pond & Frost, 2005; Murrell et al., 2013).

Reverse transcriptase (RVT) and transposase (DDE) consensus domain alignment

The identification of DDE (transposase) and RVT (reverse transcriptase) domains in the *A. latifasciata* transcriptome (Nakajima, 2019) was performed using the standalone version of the Hidden-Markov model HMMER 3.3 (hmmer.org) with the PFAM protein database (ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam34.0/). The sequences that matched the transposase and reverse transcriptase domains (E-value < 0.01 and acc > 0.8) were selected to perform nucleotide alignment on Geneious (https://www.geneious.com/). Based on this alignment, for each domain of transposase and reverse transcriptase, a consensus sequence was generated and used for primer design on Primer3Plus (Koressaar & Remm, 2007) to perform RT–PCR. The primers and the domain alignment are described in **Additional File 2** and **Additional File 6**, respectively.

The transposon elements landscape by families

The repetitive sequences of B- and B+ *A. latifasciata* assemblies were identified using RepeatMasker 4.1.0 (repeatmasker.org/RepeatMasker/). A custom library based on RepeatModeler (repeatmasker.org/RepeatModeler/) and merged with Repbase Update 20181026 (Schemberger et al., 2019) was used to run RepeatMasker with -s, -l, and -a parameters.

The repeat landscape and Kimura divergence values were generated using calcdivergencefromalign.pl and repeatlandscape.pl, both RepeatMasker pipelines that use repetitive annotation as input. The TEs that are present in the *curupira* clusters and validated in the B chromosome (Coan & Martins, 2018) were selected to generate a comparative Kimura chart of B- and B+ assemblies.

Functional validation (RT–PCR and RT–qPCR)

The total RNA of gonads, brain, gill and muscle from four B- and B+ genotyped males and females was extracted using TRIzol (see “Obtaining samples and small RNA sequencing”), resulting in twelve samples for each tissue, followed by

cDNA synthesis using an Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit (10400745). RT–qPCR was performed using SYBR Green qPCR Master Mix (High ROX) Ampliqon (HY-K0521) with the following cycles: 95 °C for 20 seconds, 34 cycles (95 °C for 3 seconds and 60° for 30 seconds); melting curve step: 95 °C for 15 seconds, 60 °C for one minute, 95 °C for 15 seconds. The expression based on the ΔΔCq method was analyzed using the ubiquitin-conjugating enzyme (*ubce*) as a reference through Q-Gene software (Simon, 2003). The primers are listed on **Additional File 2**.

Results

Characterization of piRNA clusters in *A. latifasciata*

The sRNA-seq libraries from FB-, FB+, MB- and MB+ were filtered by size to characterize piRNAs and compare them to *A. latifasciata* miRNAs (Nascimento-Oliveira et al., 2021), highlighting the two different length distributions in each type of small RNA dataset (**Figure 1a**). The miRNAs showed a greater abundance of 21-22 nucleotides, while the piRNAs showed more instances of 28-29 nucleotides, the most common length for these two classes (27). Additionally, miRNAs showed more expression than piRNAs in females, and the opposite was observed in males.

After identification, 372 piRNA clusters were predicted in the A. *latifasciata* genome. The shorter piRNA cluster has 1,006 base pairs (bp), while the longer piRNA cluster has 19,333 bp. The number of transcribed piRNA clusters in each group was 160 in FB-, 135 in FB+, 146 in MB- and 184 in MB+ (**Figure 1b** and **Additional File 1**). The clusters of each group were joined to form a single nomenclature for piRNA clusters of *A. latifasciata* (for more information about the clusters, such as name, genome localization, size and expression, see **Additional File 1**).

The differences in genome distribution between piRNAs and miRNAs are demonstrated in a Venn diagram showing contigs of these small RNA sequences (**Figure 1b**). Sixteen contigs of the *A. latifasciata* genome carried both miRNA and piRNA genes. Among them, thirteen contigs contained pre-miRNA sequences (miRNA genes that transcribe the hairpin structure) and were located inside piRNA clusters (**Figure 1b** and **Table 1**).

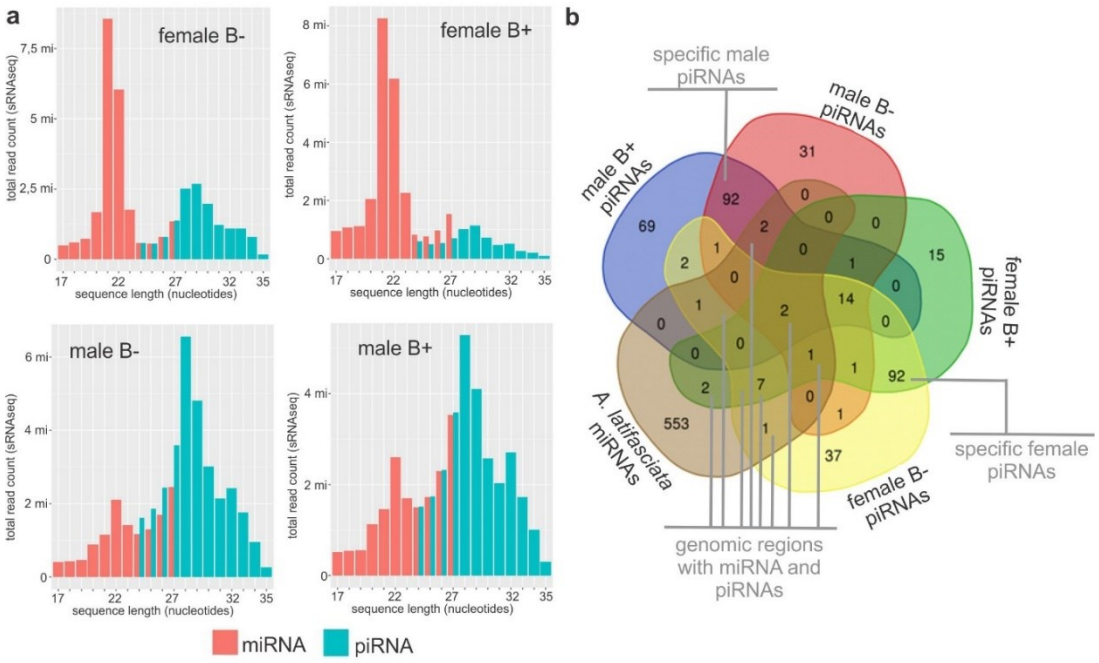

Figure 1 - Data on miRNAs and piRNAs in *A. latifasciata*. **a** Histogram of filtered datasets for each miRNA and piRNA analysis. **b** Genomic miRNA and piRNA Venn diagram depicting all the *A. latifasciata* small RNA genomic annotations. B-, samples without the presence of the B chromosome; B+, samples with the presence of the B chromosome.

Three contigs carried piRNA clusters and pre-miRNAs without overlaps. Some miRNAs were clustered into piRNA clusters, which is why there more than a pre-miRNA superposition on a piRNA cluster. These clustered miRNAs were miR-30a/b and miR-19a/b (**Table 1**). The contig NODE_540432 also carried miR-17b and novel sequences (miRNAs identified only in *A. latifasciata*), forming a clustered miRNA that was identified for the first time.

Table 1 - Colocalization of miRNA and piRNA sequences in the *A. latifasciata* genome. S, strand; T : S; transcription and strand.

Contig	**miRNA**			**piRNA**			
	miRNA start : end	**S**	**miRNA ID**	**piRNA cluster start : end**	**T : S**	**piRNA cluster**	**gonads expression**
miRNAs e piRNAs not superposed							
NODE_413592	23358 : 23437	+	novel_27198	5015 : 11188	mono:-	161	MB-, MB+
NODE_659880	5542 : 5604	+	novel_40237	1 : 5034	mono:-	297	FB+ FB-
NODE_530411	42958 : 43045	+	novel_33801	46001 : 52942	mono:+	230	FB+ FB-

miRNAs e piRNAs superposed							
NODE_540432	60460 : 60520	-	miR-17b-1	58002 : 63335	mono:+	235	FB-, FB+, MB-, FB+
	61062 : 61118	+	novel-34307				
	61235 : 61302	+	novel-34310				
	61365 : 61424	+	novel-34311				
	61471 : 61530	+	novel-34313				
NODE_374673	36480 : 36543	+	miR-146b	35407 : 36661	mono:+	146	FB-, FB+, MB-
NODE_347176	36121 : 36181	+	miR-21-1	35371 : 38453	mono:-	134	FB-, FB+
NODE_34239	37421 : 37486	+	miR-30a-1	36149 : 39395	mono:+	133	FB-, MB+
	37756 : 37816	+	miR-30b				
NODE_158712	303 : 365	+	miR-10c	304 : 3733	mono:+	33	FB+
NODE_566122	1109 : 1162	+	novel_35680	2 : 5452		248	FB+
NODE_749189	27328 : 27349	+	miR-19a-5p	25083 : 30798	mono:+	330	FB-, FB+, MB-, FB+
	27674 : 27696	+	miR-19b-5p				
NODE_319315	10039 : 10120	+	novel_21996	7 : 10614	mono:+	127	MB-, MB+
NODE_6163	99 : 153	+	novel_675	10 : 4241	bi:+/-	270	FB+ FB-
NODE_27951	2471 : 2536	+	novel_2628	6 : 3836	mono:+	107	FB+ FB-
NODE_960524	12443 : 12504	+	mir-21a	12344 : 14456	mono:-	368	FB+ FB-
NODE_581740	1963 : 2049	-	novel_36349	1 : 4010	mono:-	253	FB+ FB-
NODE_81238	848 : 916	-	novel_6616	9 : 2856	mono:-	344	FB-

We investigated which functions could be related to the colocalized miRNAs and piRNAs by examining the expression of their targets in gonads, which were already predicted (Nascimento-Oliveira et al., 2021). Three miRNAs did not have any target expression in gonads (miR-146b, miR-10c and miR-21a). Even though the targets of these three miRNAs could be expressed in other tissues, we restricted the analysis to gonads for comparison with piRNAs, which have higher expression in gonads. Conversely, 19 miRNAs colocalized to piRNAs targeting dozens of mRNAs expressed in gonads (**Additional File 5**). Additionally, seven domains related to mobilization are targets of the colocalized miRNAs: three associated with retrotransposition (LORF, *LINE-1 retrotransposable element ORF 1*; LIN1, *LINE-1 reverse transcriptase homolog*; RVT-like, *probable RNA directed DNA polymerase from transposon X element*) and four associated with transposition (PGBD, *PiggyBac transposable element derived protein*; POGK, *Pogo transposable element with*

KRAB domain; TC1, *Tc1 transposase*) (**Figure 2a**). The colocalized miRNAs annotated in the piRNA clusters formed pre-miRNA secondary structures (**Figure 2b**).

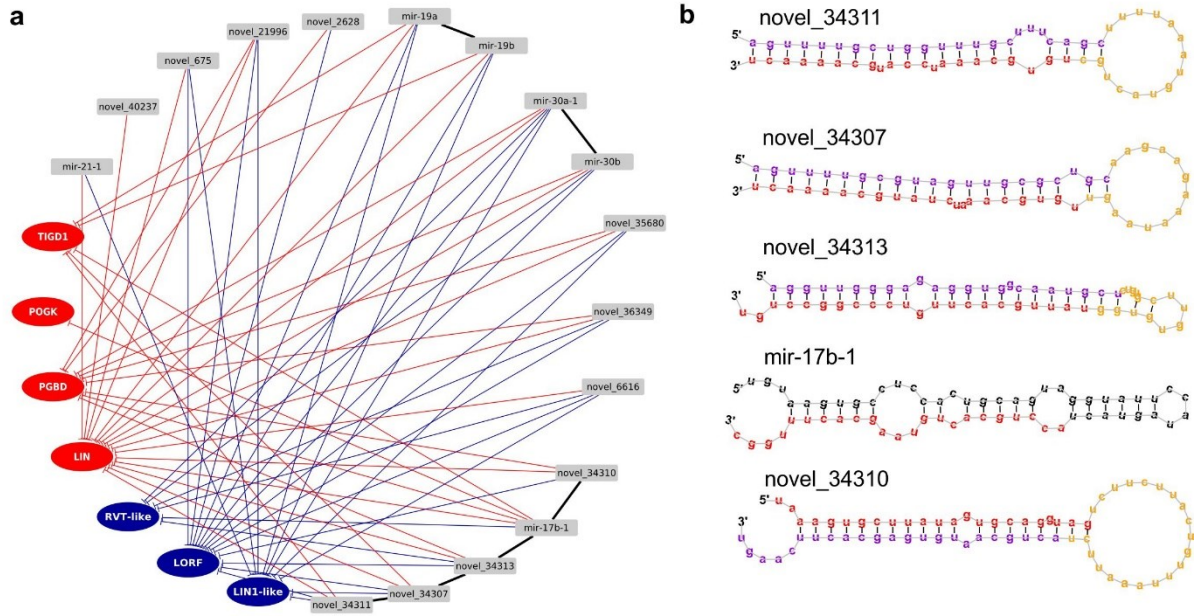

Figure 2 - MicroRNAs colocalized with piRNA clusters and their transposition gene targets. **a** miRNA:mRNA interaction network: the red circle nodes are targets associated with DNA transposons, and the blue nodes are targets derived from retrotransposition genes. The edges that connect the miRNAs (gray nodes) in red show connections of miRNA:transposons, and those in blue show connections of miRNA:retrotransposons. The bold black edges highlight the clustered miRNAs consequently derived from the same piRNA cluster in the same contig. **b** Clustered pre-miRNA secondary structures. The colors indicate each region identified in the small RNA-seq dataset: purple, passenger sequence; red, mature miRNA sequence; yellow, steam-loop sequence; black, not identified in the small RNA-seq dataset.

The *A. latifasciata* B chromosome carries piRNA clusters

Considering the similarity with the B chromosome and the entire standard genome, it is not possible to recover the B-assembled chromosome based only on Illumina data (Jehangir et al., 2019). An alternative strategy to find B chromosome copies is through the coverage ratio approach (Valente et al., 2014). This method consists of comparing the genomic read alignment of sequenced B- and B+ samples to the reference genome (*A. latifasciata* assembly). In this comparison, it is possible to detect genomic regions with higher coverage in B+ sequencing (B-blocks)

(Jehangir et al., 2019). In this way, we applied the coverage ratio method to find piRNAs in the B chromosome.

Three piRNA clusters were located in contigs with higher B+ genomic read coverage (**Figure 3a**, **b** and **c**), suggesting that these piRNA clusters were present in the B chromosome. Additionally, these clusters were enriched by TEs (**Figure 3d**, **e** and **f**), and two of them had significant differential expression in B+ samples (**Figure 3g**, **h** and **i**). These three piRNA clusters found in the B chromosome were named *curupira* (*curu*), followed by the contig number which the cluster derived: *curupira-59920* (*curu-59920*), *curupira-138667* (*curu-138667*) and *curupira-330285* (*curu-330285*). Curupira is a famous character in Brazilian folklore who protects the forests against hunters; he is the guardian of the forest. The *curupira* clusters are the guardians of the B chromosome.

The first cluster, *curupira-59920* (cluster 263 in **Additional File 1**), was 1,027 bp long and was predicted in contig NODE_59920. This contig had approximately 300x coverage considering the B- genomic reads, while the B+ genomic read count showed approximately 9,000x coverage (**Figure 3a**). The piRNA cluster corresponded to the entire contig of the reference genome and showed similarity to BEL/PAO retroelements (**Figure 3d**). Concerning piRNA transcription characteristics, this cluster was monodirectional, in which the processed piRNAs were derived from the plus strand. This cluster was expressed only in B+ samples, including both males and females (**Figure 3g**).

Curupira-138667 (cluster 19, **Additional File 1**) was 4,986 bp in size, which is almost the complete reference contig size. The B- genome reads made up approximately 60x coverage, and the B+ genome reads provided approximately 400x coverage (**Figure 3b**). Additionally, this contig showed similarity to a small portion of the retroelement DIRS (**Figure 3e**). The directionality of this cluster was monodirectional with expression on the minus strand. There was no significant difference in expression among the groups (**Figure 3h**).

The cluster *curupira-330285* (cluster 131, **Additional File 1**) was 8,024 bp long and corresponded to 80% of the contig size. Regarding coverage, the B- genome reads corresponded to 60x coverage, and the B+ genome reads provided 200x coverage (**Figure 3c**). Both contigs and clusters were enriched by several transposable elements and other simple repeats (**Figure 3f**). *hAT* and LINE

elements were representatives in this cluster, and the details are described in **Figure 4**. Furthermore, this cluster showed bidirectionality; in other words, transcription occurred in both strands. Expression was also observed in B+ samples, including both males and females (**Figure 3i**).

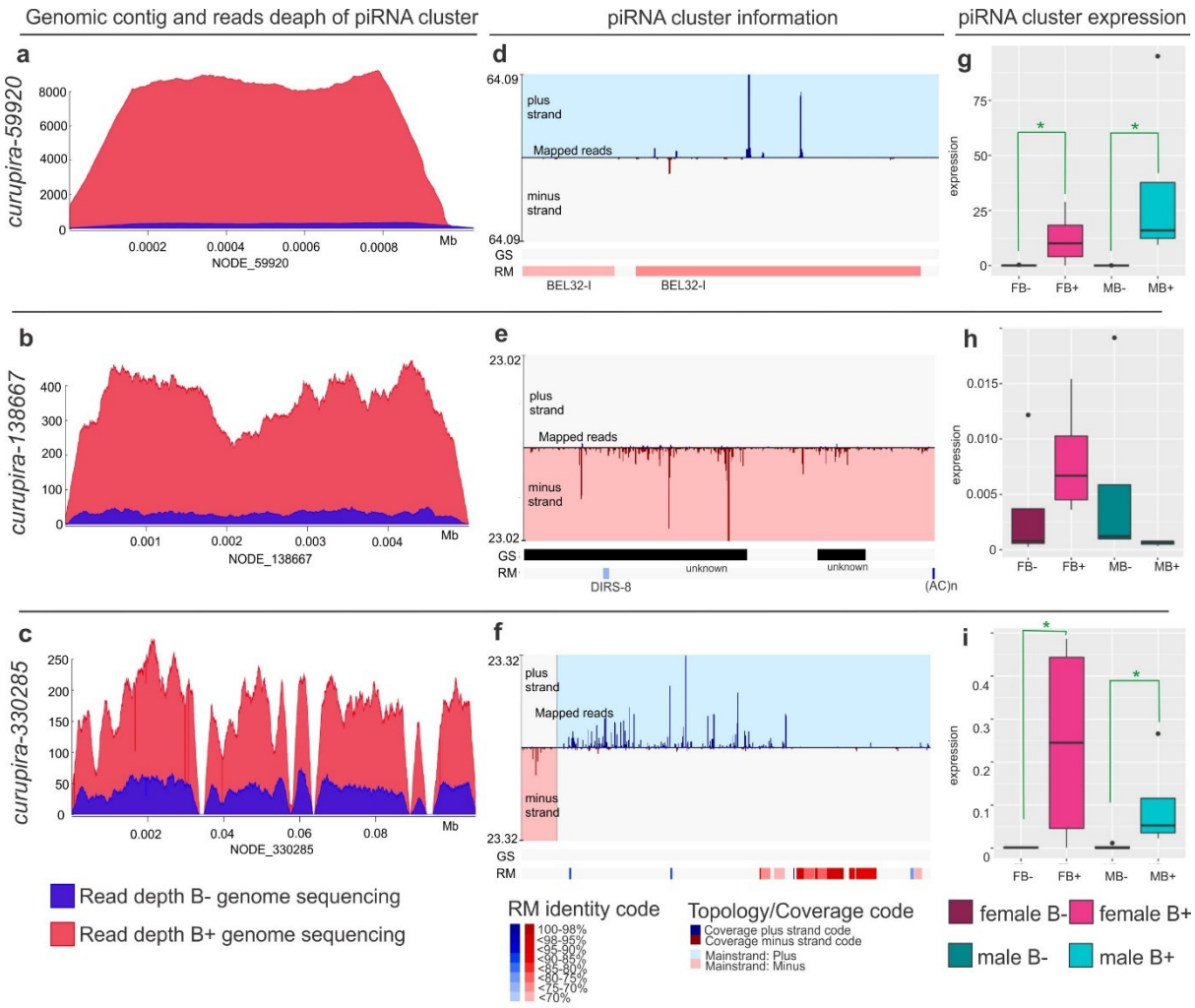

Figure 3 - Description of B chromosome piRNA clusters (named *curupira*). **a-c** Depth coverage graphs of piRNA clusters comparing B- (blue) and B+ (red) genomic reads. **d-f** piRNA cluster characterization from the prediction. The graphs show the piRNA sequences from the small RNA-seq dataset aligned to piRNA clusters (mapped reads); blue and red regions indicate plus and minus strands, respectively. GS, gene set from genome annotation; RM, RepeatMasker identity. **g-i** piRNA cluster expression in gonads. The Y axis shows the expression based on the ΔΔCq method, and the X axis shows the samples. The boxplot represents quartiles, the horizontal bar within the box is the median, the vertical lines outside of the boxes are the minimum or maximum value, and the points are outliers. Green asterisks represent significant differential expression (p <.0001).

	TE	Class/family	(%)
1	hAT-6	DNA/hAT-Ac	12.5
2	AlRepA-4	DNA/hAT-Charlie	27.7
3	AlRepA-4	DNA/hAT-Charlie	38.8
4	AlRepC-1433	DNA/hAT-Charlie	3.5
5	AlRepC-743	LINE/L2	20.3
6	AlRepC-1433	DNA/hAT-Charlie	7.3
7	AlRepC-743	LINE/L2	21.8
8	AlRepC-1433	DNA/hAT-Charlie	8.4
9	AlRepB-392	DNA/hAT-Charlie	1.7
10	AlRepB-392	DNA/hAT-Charlie	11.3
11	AlRepB-392	DNA/hAT-Charlie	14.6
12	Chap3a_Cis	DNA/hAT-Charlie	13.3

Figure 4 - Cluster 131 (c*urupira-330285*) RepeatMasker similarity. Details for the TEs and other repeats of the *curupira-330285* cluster are shown at the top of the figure. TE matches are highlighted in blue from the plus strand and red from the minus strand. The list describes each TE match, including the percentage of similarity between the piRNA cluster and the TE fragment match.

Table 2 shows a summary of the three piRNAs found in the B chromosome. Finally, of the three piRNA clusters, two were expressed only in the B chromosome (**Table 2**), suggesting that the paralog sequences in the *A. latifasciata* genome reference (NODE_59920 and NODE_330285) were either not piRNA clusters or were not active in autosomes.

Table 2 - Information on B chromosome piRNA clusters.

piRNA cluster name	piRNA cluster ID	Size (bp)	Reference genome localization	TE enrichment	T:S	RT–PCR gonads
curupira-59920	263	1027	NODE_59920	retroelement (LTR/PAO)	mono: +	B+ samples
curupira-138667	19	4986	NODE_138667	retroelement (LTR/DIRS)	mono: +	upregulated in B+ females
curupira-330285	131	8024	NODE_330285	transposon and retrotransposon (DNA/hAT, LINE/L2)	di:+/-	B+ samples

The B chromosome could affect TE mobilization

To verify whether there was a relationship between TE and piRNA evolution in the *A. latifasciata* genome and its B chromosome, we compared the masked TE in the previously *A. latifasciata* genome (Coan & Martins, 2018) and the TE content in the piRNA clusters predicted in the present study.

To obtain the TE constitution of piRNA clusters, we considered the RepeatMasker annotation (Coan & Martins, 2018) detected during piRNA prediction. The ratio of TE constitution in the piRNA clusters was reported separately to each group (FB-, FB+, MB- and MB+), describing the differences among the types of piRNA expression (**Figure 5a**, right side of bar plot). Starting with this information, we calculated the TE constitution in the piRNA clusters and *A. latifasciata* TE ratios to each TE subclass (**Figure 5b**).

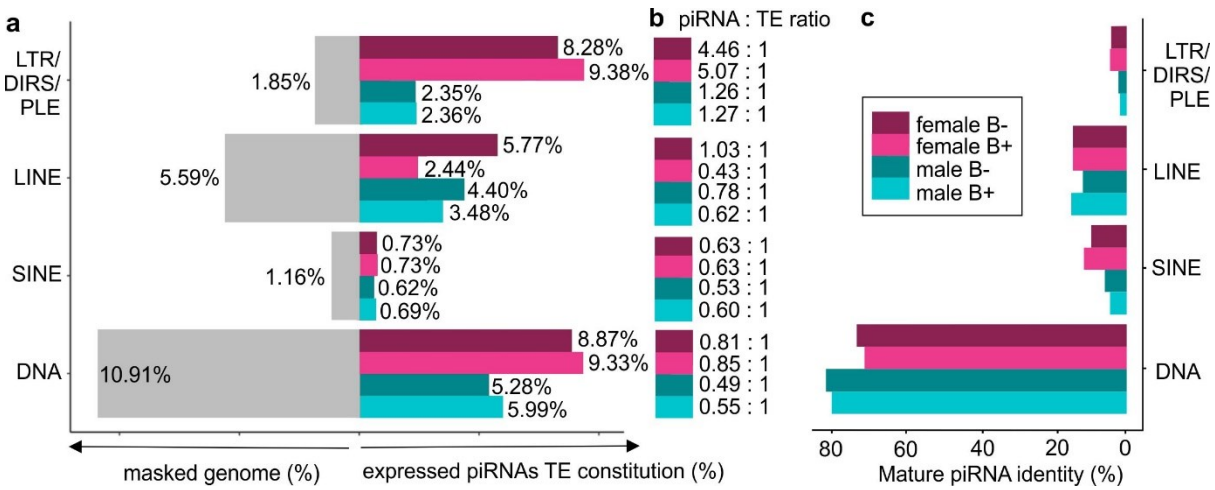

Figure 5 - TE and piRNA relations in *A. latifasciata*. **a** TE genome and piRNA cluster constitution. The left gray graphs show the *A. latifasciata* proportion of each TE class, and the colored right side represents the TE proportion of piRNA clusters expressed in gonads. **b** Expressed piRNA cluster TE constitution in relation to the masked genome region. **c** Mature piRNA similarity with TEs. The X axis represents the relative quantity of mature piRNA sequences with similarity to TEs.

Concerning the LTR, DIRS and PLE retroelements, which covered 1.85% of the *A. latifasciata* genome, the females (either B- or B+) expressed piRNA clusters covered by these subclasses, corresponding to 5x of its TE genome content. Thus, for each LTR/DIRS/PLE genome localization, 5 piRNA clusters derived from this subclass were expressed (LTR/DIRS/PLE). Conversely, in males, the relation was

1.2 piRNA-derived LTR clusters to each LTR in the genome. These elements have old degeneration events (**Figure 6a-c**).

Regarding LINE elements, which corresponded to 5.59% of the *A. latifasciata* genome, the females without B chromosome maintained a ratio of 1:1 for LINE-derived clusters, whereas for B+ females, this relation decreased to 0.43 piRNAs for each LINE in the genome. The same phenomenon was observed among males: the B- males had more piRNAs for each LINE, and a decrease in piRNA-derived LINE expression was observed in B+ males. This relation could be due to the constant and recent LINE insertions in the *A. latifasciata* genome, including the B chromosome (**Figure 6d**): more LINEs were found covering the genome than expressed piRNAs derived from this retroelement. While SINE elements covered 1.16% of the assembled genome, approximately 53% to 63% of the expressed clusters were derived from this retrotransposon and did not display great variation among the samples.

Conversely, DNA transposons, the most representative class in the *A. latifasciata* genome, were controlled by approximately a one-to-one ratio of transposon-derived female piRNAs (1.23 and 0.85 in FB- and FB+, respectively). This ratio was approximately halved in males (0.49 and 0.55 in MB- and MB+, respectively).

In general, there was no relationship between the TE distribution and the type of TE-piRNA-derived expression, and the presence of a sex and B chromosome bias that orchestrated piRNA cluster expression, independent of the genome proportion of *A. latifasciata* TEs, was reinforced. Another consistent finding concerned the mature piRNA identity. A blastn of the mature sequences over the library used in the RepeatMasker annotation (see Material and Methods) revealed that approximately 70% of the piRNAs were similar to transposon elements (**Figure 5c**), and the third wave of TE spread in the *A. latifasciata* genome was composed of transposon elements (**Figure 6e**), indicating that these elements could be expressed and that more piRNAs for this class could be requested. This point will be discussed further.

Due to the large presence of repetitive elements and duplicated sequences in the B chromosome, recovering B genes and B-specific sequences using Illumina reads is a major challenge (Jehangir et al., 2019; Nascimento-Oliveira et al., 2021). Thus, the *A. latifasciata* assembled genome was used as a reference to compare the

B- and B+ sequenced DNA and RNA data to find B chromosome-related sequences. However, based on the known TE families enriched in the B chromosome, we repeated the RepeatMasker and landscape analysis using the *A. latifasciata* reference genome and the B+ assembled genome (assembly performed with DNA samples from individuals with the B chromosome). Landscape analysis revealed differences between B- and B+ TE insertion events of duplicated sequences. Extracting the Kimura values from the landscape of specific TE families from B- and B+ allowed us to infer the time at which the TEs from B chromosomes could have emerged and differentiated from A TE sequences (**Figure 6a-e**). The BEL, L2 and *hAT* families are both present in the B chromosome and made up the piRNAs in the B chromosome. The Gypsy family is not present in the *curupira* clusters but is present in the B chromosome (Coan & Martins, 2018). These duplicated families in the B chromosome carried B-specific events in the Kimura landscape (**Figure 6a-e**). We suggest that these degeneration events could have contributed to the origin of the *curupira* cluster.

The representative major classes of TEs composing the three *curupira* sequences were LTR retrotransposons (PAO and DIRS), non-LTR retrotransposons (LINE/L2) and the *hAT* DNA transposon (**Table 2**).

Compared with all the selected TEs, LTR/PAO present in *curupira-59920* showed an older process of genome invasion. While LTR/DIRS had newer waves of invasion, one was older and the other was as recent as the new *hAT* element dispersion. The LINE/L2 element showed constant invasion, and genomic degeneration did not start (**Figure 6d**).

Comparisons of the B- and B+ TE contents revealed different copy numbers among the assemblies, these differences can suggest the moment of *curupira-59920* origin. This piRNA cluster was not expressed in B- samples, indicating that the original copy was just a TE region. However, in the B+ genome, this process enabled the piRNA cluster to originate in the B chromosome.

We evaluated the activity of reverse transcriptase (RVT) and transposase (DDE) domains through common consensus sequences identified in the *A. latifasciata* transcriptome (**Figure 6f-g**). The sequences annotated in the *A. latifasciata* transcriptome that corresponded to reverse transcriptase and transposase (regardless of the TE family) were aligned and used to obtain the

consensus sequences, which were used to design RT–qPCR primers for common regions of each gene (**Additional File 6**). The RVT domain was highly expressed in males compared with females, while the presence of the B chromosome seemed to decrease reverse transcriptase expression in testes (p value <.003) (**Figure 6f**). In turn, for transposase, no significant difference was observed between the sexes. Conversely, the presence of the B chromosome in males seemed to decrease expression in testes (p value <.04) (**Figure 6g**).

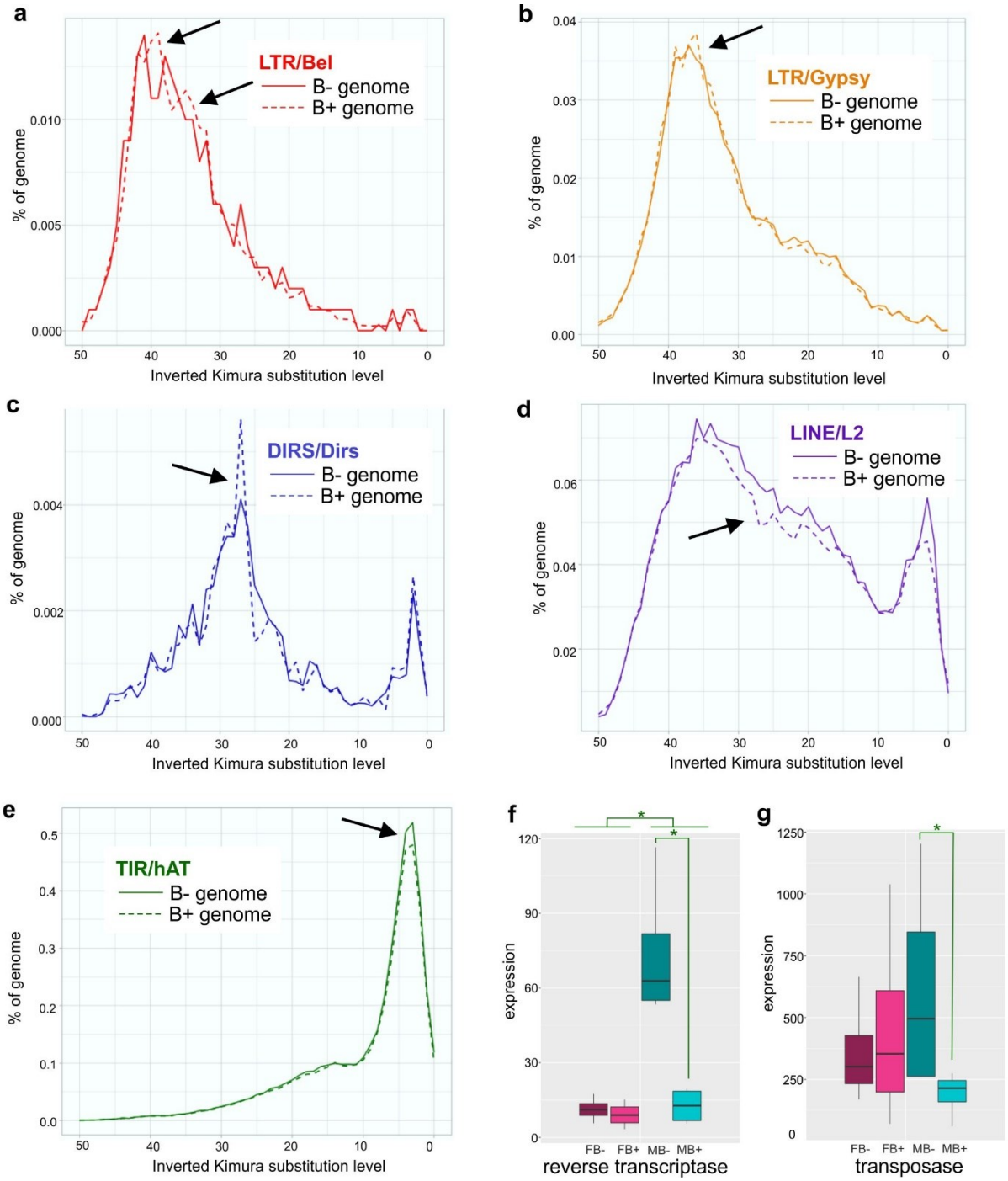

Figure 6 - Transposable element activity in *A. latifasciata*. **a-e** Kimura landscapes for nucleotide substitution rates. The black arrows highlight points of divergence between B- and B+ assemblies, indicating putative B+ chromosome TE diversification. **f-g** RVT and DDE expression in gonads. The Y axis shows expression based on the ΔΔCq method, and the X axis indicates the samples (FB-, FB+, MB-, MB+). The boxplot represents quartiles, the horizontal bar within the box is the median, and the vertical lines outside the boxes are the minimum or maximum values. The green asterisk represents significant differential expression (p <.003). FB-, female without B; FB+, female with B; MB-, male without B; MB+, male with B.

piRNA pathway and the B chromosome

We evaluated the expression of genes related to the piRNA pathway (*pld6, piwil1* and *piwi2*) in the testes and ovaries of B- and B+ samples (**Figure 7a**). Among these genes, the *pld6* gene has been previously reported in the B chromosome using Illumina sequencing (Valente et al., 2014).

The expression of *pld6* in the B samples, both male and female, was significantly higher (p .0017), which indicated that the *pld6* B-copy could be active and contribute to increased expression in B samples (**Figure 7a**). There was no significant difference in the expression of *piwil1* and *piwil2* in the presence of the B chromosome (p .09 and .08, respectively). Conversely, there was a difference in expression among males and females for the three analyzed genes (p <.0001). While *pld6* and *piwil1* were highly expressed in males, *piwil2* was more highly expressed in females.

We confirmed that these piRNA pathway genes were expressed in the gonad and found evidence for transcription of these genes in somatic samples, such as brain, muscle and gill (**Figure 7b**). The *pld6* gene was also expressed in the brain, including both males and females, with or without the B chromosome. In muscle and gill, it was expressed only in male B- and B+, respectively. The expression of *piwil1* in the brain and gill was detected in all samples, and in muscle, *piwil1* expression was not detected in female B-. It was possible to observe some variations in the detected amplicons in muscle and gill, suggesting the occurrence of different mRNA processing in these tissues. We also detected *piwil2* expression in the brain, which was clearly present in male B- and the background, suggesting truly low expression in other samples. Additionally, there was no evidence of *piwil2* expression in muscle and gill. These data suggested that sex and B presence could affect the expression of piRNA pathway genes in somatic tissues.

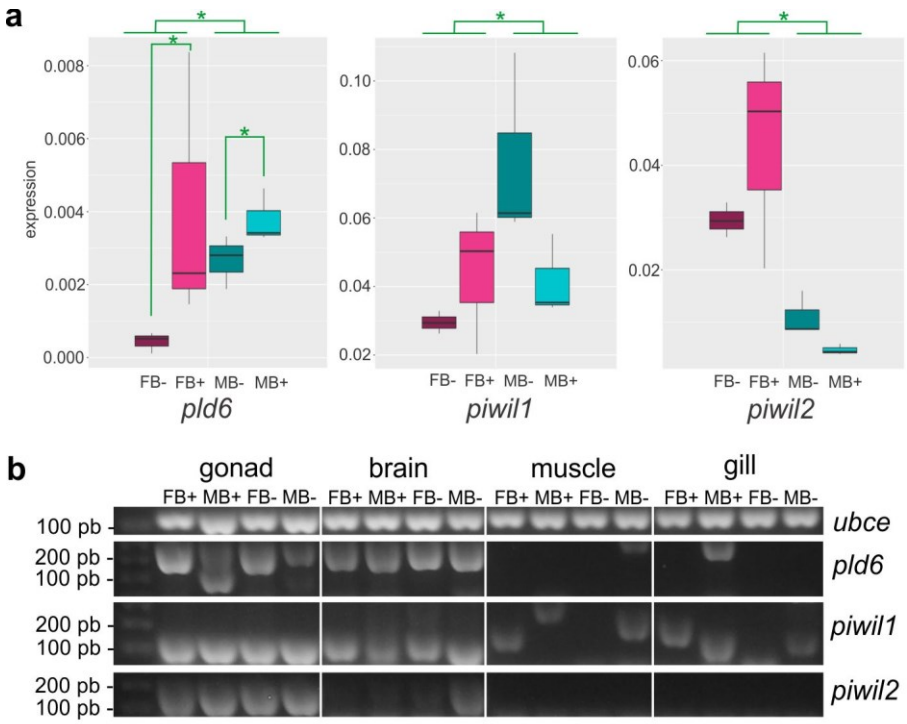

Figure 7 - piRNA pathway gene expression in *A. latifasciata*. **a** *pld6, piwil1* and *piwil2* gene expression in gonads. The Y axis shows expression based on the ΔΔCq method, and the X axis shows the samples. The boxplot represents quartiles, the horizontal bar within the box is the median, and the vertical lines outside the boxes are the minimum or maximum values. Green asterisk represents significant differential expression between the groups (p <.001). **b** PCR of piRNA pathway gene expressed in the gonad, brain, muscle and gill.

In addition to *pld6* expression, we also performed genome investigations to detect the B chromosome copy. Due to fragmentation of the draft genome, it was not possible to recover the complete *pld6* sequence from the *A. latifasciata* assembly, and thus, the *Metriaclima zebra* assembly was used as a reference genome. The *pld6* gene was present in scaffold_135 of the *M. zebra* genome, and the *A. latifasciata* B- and B+ sequencing read alignment on this region revealed the coverage difference in B+ compared with B- data. The B+ genome reads had approximately 200x coverage, while the B- genome had approximately 60x coverage (**Figure 8a**). GDR analyses indicated the occurrence of more copies in B+ samples, confirming the copies of this gene in the B chromosome (**Figure 8b**). In addition to the difference in coverage, it was possible to recover several B-specific mutations located in the *pld6* introns (**Additional File 7**). One of them was used to construct a primer to amplify the *pld6* B-specific copy (**Figure 8c and Additional File 7**).

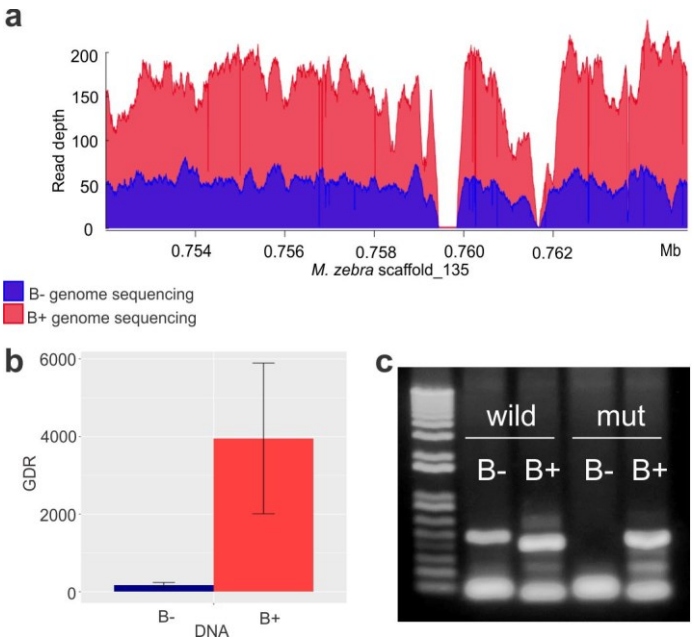

Figure 8 - *pld6* B-copy characterization. **a** Coverage chart of the B- (blue) and B+ (red) reads aligned against *M. zebra pld6*. **b** *pld6* gene dosage ratio (GDR) from B- (blue) and B+ (red) DNA samples ($p < .0001$). The Y axis shows the GDR based on the ΔΔCq method. **c** B-specific copy amplification. Wild is a nonmutated primer, and mut is a primer with a B-specific sequence that was found in the B+ genome sequenced reads. FB-, female without B; FB+, female with B; MB-, male without B; MB+, male with B.

We also recovered two transcripts of *pld6* from *A. latifasciata* mRNA-seq data using a standalone blastx based on the *M. zebra* sequence (XM_004567832). These sequences were used to construct a phylogenetic tree based on the CDS including more animal species, as described below.

Some studies have revealed the rapid evolution of piRNA pathway genes in teleosts (Song et al., 2019; Yi et al., 2014). Based on this evidence, we investigated *pld6* selective pressure in different ways. We manually identified the CDS region of *pld6* sequences from 41 animal species: one *D. melanogaster*, seven mammals, four Chondrichthyes, 27 teleosts and two *A. latifasciata* sequences recovered from the transcriptome. These sequences were used to construct a phylogenetic tree and perform evolutionary analysis. The *A. latifasciata pld6* sequences are well located in the cichlid clade (**Figure 9a**). The BUSTED test provided evidence (p value = .03 < .05) of gene-wide episodic diversifying selection in the selected test branches in the phylogeny, meaning that at least one tested branch had undergone diversifying selection. For this reason, we performed the aBSREL test, which identified those branches under diversifying selection. After correction, three nodes in the tree were

determined to be under selection, involving mammals, cypriformes and cichlids (**Figure 9a**).

To search for sites under selection, the FUBAR and FEL tests revealed 2 sites that were subjected to diversifying positive selection. Conversely, 121 sites were under negative selection ($p < 0.1$) (**Additional File 8**).

We also compared alignments of the *A. latifasciata pld6* sequence to *M. zebra* and *Astatotilapia caliptera*. It was possible to observe mismatches in the alignment that could represent 1) the specificity of *Astatotilapia* genus, where *Astatotilapia* species and *M. zebra* showed differences; and 2) *A. latifasciata*-specific nucleotides, where this species differs from *M. zebra* and *A. caliptera* (**Figure 9b**). This evidence could aid in understanding why *pld6* B-copy had no significant mutations over its CDS region. The selective pressure on this gene likely prevented the pseudogenization of the *pld6* B-copy.

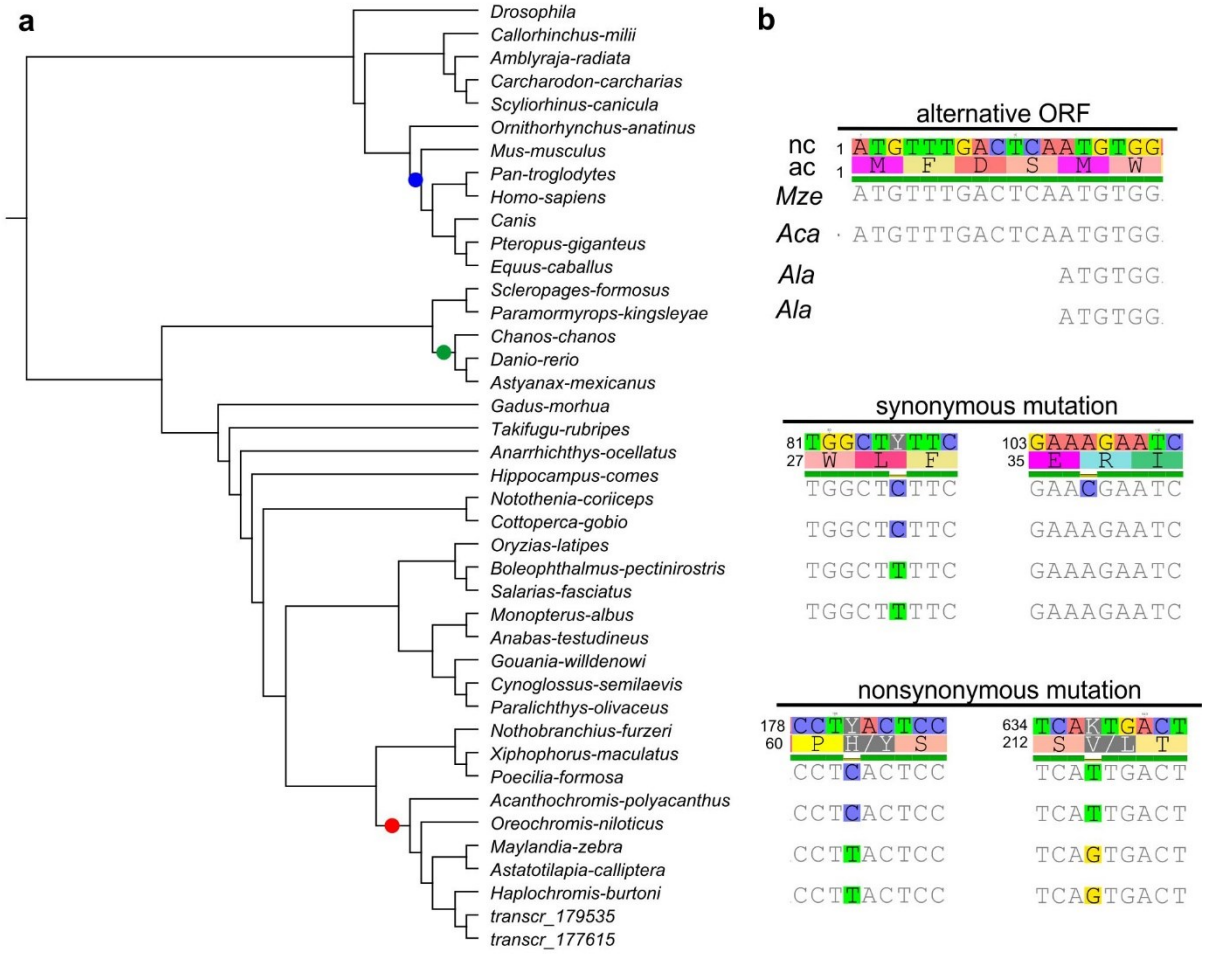

Figure 9 - *A. latifasciata pld6* transcript comparison. **a** *Pld6* phylogenetic tree; transcr_179535 and 177615 are *A. latifasciata* sequences from its transcriptomes. The colored dots represent the nodes

under diversifying selection. **b** Nucleotides (nc) and amino acids (ac) of *M. zebra* (*Mze*), *Astatotilapia caliptera* (*Aca*) and *A. latifasciata* (*Ala*) *pld6* CDS alignment. An initial alternative frame is detected in the *A. latifasciata* transcripts. Two detected synonymous mutations are shared among the *Astatotilapia* genus. Finally, two nonsynonymous mutations represent changes in the amino acid sequences in *A. latifasciata* PLD6.

Discussion

We identified a genomic locus of piRNA clusters in *A. latifasciata* cichlid fish. There are no available data on the piRNA sequences of any African cichlid, and the closest *A. latifasciata* species with a piRNA annotation deposited in the piRBase is *Danio rerio* (Wang et al., 2019; Zhang et al., 2014). We also used the *Metriaclima zebra* genome as a reference to predict the piRNA clusters through *A. latifasciata* sRNA-seq without success (data not shown). In contrast to miRNAs, piRNA sequences are poorly conserved and are considered species specific (Bartel, 2018; Moazed, 2009). Other differences among miRNAs and piRNAs are the length and the genomic distribution (Moazed, 2009). piRNAs are longer than miRNAs and are frequently found in heterochromatic regions (Siomi et al., 2011), while miRNAs are more common in intergenic regions (Bizuayehu & Babiak, 2014). Based on this knowledge, our characterization was efficient in differentiating the two small RNA classes from the same sRNA-seq data, and most miRNAs were not superposed or colocalized to piRNAs (**Figure 1b**).

In addition to piRNA cluster annotation, we identified miRNA genes that were colocalized with piRNA sequences. These miRNAs targeted several mRNA transcripts, including transposable element transcripts. We questioned whether the piRNA cluster evolved after the miRNA or whether the miRNA evolved from the piRNA cluster. As a third possibility, piRNA clusters carry small RNA sequences potentially with miRNA secondary structures, thus serving as miRNAs from piRNA clusters. The piRNA sequences were first described a few years ago (Hartig et al., 2007), and with the ongoing assessments of the piRNA pathway, new related proteins, new ways and new functions have been identified. Recently, new functions have been presented for piRNA targeting, and a wicky paring function similar to miRNA:mRNA interactions has been detected (Ramat & Simonelig, 2021). Thus, we suggest that the *A. latifasciata* piRNA clusters carry miRNA-like sequences, and these sequences are orchestrated to facilitate TE silencing (**Figures 1 and 2**).

Additionally, in comparisons of the miRNAs and piRNAs of *A. latifasciata*, miRNAs were more highly expressed than piRNAs in the female gonad, while piRNA expression was higher in the testis (**Figure 1a**). There are two explanations for these findings, which concern the different activities of the piRNA pathway and transposable elements in the ovary and testis (Houwing et al., 2007; Song et al., 2019), both of which will be further discussed. These phenomena are also reflected in the expression of the clusters shown in **Figure 1b**. Differences are observed among the samples, and in addition to sex bias (Houwing et al., 2007; Kawaoka et al., 2011; Kiuchi et al., 2014), we suggest a B chromosome bias for piRNA cluster activity.

Not only a B chromosome bias for piRNA activity in the gonads but also three piRNA clusters, called *curupira-59920*, *curupira-138667* and *curupira-330285,* were identified in this extra chromosome (**Figure 3**). The piRNA description in the B chromosome was determined only in the wasp B chromosome by identifying the putative piRNA range length using sRNA-seq (Li et al., 2017). Here, we performed the first detailed characterization of piRNA clusters encoded by B chromosomes.

Additionally, our findings revealed a connection between the *A. latifasciata* B chromosome TE story. The three *curupira* clusters are enriched by B chromosome TEs that were identified using *in situ* techniques (Coan & Martins, 2018). Furthermore, the BEL/PAO family is present in the B chromosome, but its expression was not identified (Coan & Martins, 2018). This family is representative of the *curupira-59920* cluster, which is exclusively expressed in B chromosome samples. In this manner, we suggest that *curupira-59920* is a B piRNA cluster exclusive and is able to control LTR elements, such as the BEL/PAO family (**Figure 3**).

Even not the most representative class in the *A. latifasciata* genome, the LTR elements are well covered in the piRNA clusters (**Figure 5**). The opposite was also observed; the DNA transposons were the most representative class, but the piRNA clusters did not follow this pattern (**Figure 5**). This misrelation was identified in the *Danio rerio* genome, in which 70% of the TEs are transposons but 10 to 20% are piRNAs composed of transposons; in contrast, only 8% of the piRNAs match LTR elements compared with 60% of the LTRs in *D. rerio* TEs (Houwing et al., 2007). This result could be explained based on the age and activity of the TEs. Consistent with this notion, the most mature piRNA with some TE similarity represents a DNA

transposon. The *hAT* and other transposons represent the youngest invasions in the *A. latifasciata* genome (**Figure 6c**, (Coan & Martins, 2018)), and this sort of TE-derived piRNA could be more active in silencing young transposons. Conversely, in **Figure 6a,** we can observe that LTR elements are the oldest TEs in the *A. latifasciata* genome (Coan & Martins, 2018). In agreement with this finding, LINE elements have demonstrated mobilization over time, starting with LTR and continuing to date (**Figure 6d**, (Coan & Martins, 2018)). In contrast, transposon elements that spread in the second burst followed by the third newest wave were composed of *hAT* elements (Coan & Martins, 2018). Thus, the piRNA and TEs constantly evolve together (Kelleher et al., 2018; Kofler, 2019; Kofler et al., 2018). We found piRNA clusters in the B chromosome composed of both old and new insertions (**Figures 5 and 6**).

These results reinforce our suggestions that old and new TE insertions could originate from piRNA clusters (Kofler et al., 2018) in the B chromosome. Thus, we suggest that *curupira-59920* could represent the first piRNA cluster to arise in the B chromosome due to LTR/PAO older invasion followed by sudden degeneration (**Figure 5c**). Even *curupira-330285* could represent the younger piRNA in the B chromosome due to evidence for newer *hAT* and constant LINE insertions. Additionally, we suggest that the original paralog copy in the *A. latifasciata* genome is merely a TE region, while in the B chromosome, these copied regions evolved in piRNA clusters. Therefore, we observed the expression of curupira-*59920* and *curupira-330285* only in B+ samples (**Figure 3**).

Concerning TE activity, we observed a decrease in reverse transcriptase and transposase expression in the male B+ (**Figure 6f-g**). We suggest that a correlation between time and TE activity arises for piRNA and efficiency supported by the B chromosome content. Transposon element *loci* are a “trap” to generate new piRNAs and promote their silencing (Kofler, 2019, 2020). Due to heterochromatic and TE enrichment, nonrecombined chromosomes evolve piRNA clusters that carry biased expression (Goriaux et al., 2014; Kawaoka et al., 2011; Kiuchi et al., 2014). We suggest that the B chromosome is a harbor to evolve new piRNAs and orchestrate its own TE silencing to guarantee the maintenance of Bs in the host genome (Czech et al., 2018; Kofler, 2019).

Another idea that could support the origin of new piRNAs in the B chromosome is consistent with the proposal of Kofler (Kofler, 2020). Using *Drosophila* as a model and following dozens of simulations, Kofler found a minimum piRNA size (proportion of piRNA clusters) in the genome to control TE mobilization and avoid extinction. In this way, the piRNA origin in the B chromosome could help to maintain host fitness and prevent elimination of the population carrying the B chromosome and its extra TEs.

We also investigated the influence of the presence of the B chromosome on piRNA pathway genes. We validated the presence of *pld6* in the B chromosome and verified that the gene was more highly expressed in gonads carrying the B chromosome (**Figure 7 and Figure 8**). The coding region of the *pld6* B chromosome copy is complete (Valente et al., 2014), and based on this evidence and our results, we conclude that the *pld6* B chromosome copy contributes to the piRNA pathway. The rapid adaptive evolution observed in teleost piRNA genes and the high TE variability found in these genomes (Song et al., 2019; Tao et al., 2016; Yi et al., 2014) could prevent degeneration of the *pld6* B copy (Banaei-Moghaddam et al., 2013).

Conversely, we did not observe any influence of the B chromosome presence on *piwil1* and *piwil2* gonad expression. TE insertion does not affect the expression of genes involved in piRNA biogenesis (Luo & Lu, 2017), which could explain why we were unable to identify a clear pattern of expression between the *piwil1* and *piwil2* genes when comparing samples with or without the B chromosome. In teleosts, *piwil1* and *piwil2* are usually more highly expressed in the testis (Houwing et al., 2007; Song et al., 2019; Tao et al., 2016); here, we observed this pattern only for the *piwil1* gene, whereas *piwil2* was upregulated in females compared with males (**Figure 7a**).

However, we detected the expression of these genes in other tissues (**Figure 7b**). In Nile tilapia, the expression of *piwil1* and *piwil2* genes has also been detected in muscle (Tao et al., 2016); here, only *piwil1* was expressed in the brain, muscle and gill. Conversely, this pattern is not observed in teleosts, since no expression was detected in soma tissue from *Cyprinus carpio* and *Scophthalmus maximus* (Wang et al., 2018; Zhou et al., 2012). The faster evolution of piRNA pathway genes in the cichlid seems to be orchestrated by TE diversity compared with the other groups,

contributing to adaptive processes in the TE silencing pathway (**Figure 9**, (Shao et al., 2019; Yi et al., 2014)). We also observed the expression of piRNA pathway genes in tissues other than gonads of cichlid species, suggesting that these genes could have functions other than TE silencing and necessitating further investigation (Tao et al., 2016; Yi et al., 2014).

Conclusion

Although the B chromosome is currently recognized as a selfish element enriched in TEs, the *A. latifasciata* B genome has accumulated new TE insertions involved in a piRNA arms-race pathway. The existence of piRNA clusters in the B chromosome could contribute to the ratio of genome size *versus* piRNA size to prevent elimination of the B chromosome in the species. Additionally, the *pld6* gene with 100% integrity copied in B provides additional evidence for the impact of the B chromosome on the piRNA pathway. In this way, we suggested several molecular evolutionary lines of evidence for novelties that ensure B chromosome survival in the host genome. The B chromosome carries its own guardians.

References

The references used in this chapter will be presented at the end of thesis.

Availability of data and materials

The *A. latifasciata* assembly is available under Bioproject PRJN369442 access and can be visualized at https://sacibase.ibb.unesp.br/. The *A. latifasciata* piRNA sequences are shown in Additional Files. The small RNA-seq libraries were deposited in the NCBI database (accession numbers SRR13040679 to SRR13040710b).

Additional files description and access

Additional File 1 – Zipped folder with fasta and interactive html piRNA cluster information for the *A. latifasciata* genome. The nomenclature is as follows: number-pirna-cluster_sex_B-presence (f, female; m, male; 0b, without B chromosome; 1b, with B chromosome). **link**

Additional File 2 – PDF file **Table S02** Primer sequences. **link**

Additional File 3 – PDF file **Table S03** *Pld6* NCBI ID access. **link**

Additional File 4 – Fasta file containing *A. latifasciata pld6* sequences from transcriptome. **link**

Additional File 5 – miRNA:mRNA multispreadsheet: **S05a** miRNA-mRNA interaction list of colocalized *A. latifasciata* miRNA-piRNA and transcript interactions. **S05b** Transcript annotation of the transcript blastx summary description. **S05c** Number of miRNA-mRNA interactions among transcripts (all) and annotated transcripts from blastx for each colocalized miRNA. **S05d** miRNA-TE interaction transposable element miRNA target list. **link**

Additional File 6 – PDF file of reverse transcriptase and transposase alignment information. **link**

Additional File 7 – PDF file of *pld6* B-mutation alignment. **link**

Additional File 8 – Spreadsheet of FEL and FUBAR selection site analysis. **link**

3.3 Capítulo 3

"Somewhere, something incredible is waiting to be known."

– Carl Sagan

B chromosome tree of life: an RNA world perspective

Jordana Inácio Nascimento-Oliveira[1] & Cesar Martins[1]

[1] Department of Structural and Functional Biology, Institute of Bioscience at Botucatu, São Paulo State University (UNESP), Botucatu, SP, Brazil

Abstract

B chromosomes are extra parasitic elements found in several eukaryotes, from fungi, plants and animals. Usually, the Bs are originated during some errors in meiosis and are composed by duplicated sequences from host genome. Even being mostly heterochromatic and low-dense in coding genes, the B chromosomes transcribe sequences and can impact the expression of host genomes. In this review we draw a timeline of studies investigating B chromosomes and RNAs, showing the advances and main findings during their history. Also, we highlighted which RNA classes have already been reported in the B chromosomes and which ones could be more explored in the focus of new perspectives to understand the B chromosome functions. In this way, we presented a B-tree of life indicating which branches has functional RNA studies for the B chromosome. We suggest the investigation of other unexplored RNA classes and functional analysis combined to cytogenetics studies to complete the B-tree of life in an RNA perspective.

Introduction

The supernumerary B chromosome first described in a *Hemiptera* species more than a century ago (Wilson, 1907), are spread over the eukaryotic tree of life, from fungi, plants and animals (D'Ambrosio et al., 2017; Jones, 2017). The occurrence of this selfish and parasitic element in a polyphyletic way reflects its variation as to origin, number, morphology and composition between the diverse taxa (D'Ambrosio et al., 2017; Valente et al., 2017).

The cytogenetic studies allowed us identify this accessory element during the mitosis and meiosis. These pioneer investigations showed a common heterochromatic state of the B chromosomes, and classify as an inert element, poorly in genes and expression activity (Camacho et al., 2000). Take into account, rarely the B chromosome is associated to a phenotype effect. An exception, for example, is the number of B chromosomes affecting negatively the size of plants (Jones et al., 2008).

Nevertheless, functional studies propose to solve how and why this extra element are maintained in the host genomes (Ahmad & Martins, 2019). Even heterochromatic and poor in genes the B chromosomes can impact the cell physiology, affect the host chromosomes expression and express its own sequences (Benetta et al., 2019).

When were the first questions about gene expression investigated on B chromosomes? What were the first techniques used to link RNAs and B chromosomes? Thus, in this review we address the B chromosome and RNA investigations history under a functional aspect. We also propose a B chromosome tree of life, based on the previous papers, which clades have B chromosome report and which ones have RNA-B effect investigation.

Talking about the future, we also present some challenging perspectives to solve unanswered questions about the RNA world in the B chromosomes, take into account not negligence the indispensable cytological techniques to complete the lack of B chromosome report in some branches into the tree of life.

The RNA and B chromosome history: **five big eras**

The first study to investigate the RNA and B chromosome relation date of 50's and reported the increase of RNA mass in the nucleus in the B chromosome presence in maize cells (Lin, 1955). In the next decades, studies also have focused into evaluate the RNA quantity in the B cells through labeling experiments (**Figure 1**). The mainly findings of these RNA and B reports, are: (1) The B does not change the RNA mass due to its heterochromatic characteristic (Himes, 1967), (2) the B increases the RNA content in the nucleus (Ayonoadu & Rees, 1970; Lin, 1955), (3) the B could affect the RNA decreasing in the cell (Kirk & Jones, 1970), and (4) the heterochromatin of B chromosome could be transcribed (Fox et al., 1974). Even not being convergent and having consensus results, these first findings discussed if the apparently inert extra element could be altering the RNA activity, content and expression in the cells. The methodology of these works was based on fluorescent and chemical techniques, such as to label the RNA molecules in the cells.

Advancing on the RNA and the B chromosome timeline, the next decades, starting from 80's were focus to find evidences for B chromosome transcription. In the absence of DNA and RNA sequencing techniques before 90's, some studies

investigated the B chromosome content through RNA hybridization (**Figure 1**). For example, a study used RNA transcribed in vitro (C-RNA, complementary RNA) as intermediated to B chromosome sequences hybridizations (Gibson & Godfrey, 1972). This result discusses that the B chromosome besides repetitive DNA could carry specific and duplicated sequences, which produce either coding or non-coding sequences. This study also suggested that in the future performing *in vitro* transcription can shed a light about the B chromosome functions. In fact, with the molecular biology and bioinformatics advances, nowadays we have thousands of B-sequences and B-transcripts reported, as we discuss further (Benetta et al., 2020; Ma et al., 2017; Miao et al., 1991; Ramos et al., 2017), including *in vitro* transcription and translation of B-copy genes (Ma et al., 2017).

Starting the new millennium, since the sequencing technology and oligonucleotide synthesis raised, the studies focus in sequences became possible. In the B chromosome area this was not different, the first B chromosome genes and expression sequences in the B presence were reported (Auger & Birchler, 2002; Leach et al., 2005; Miao et al., 1991; Tanić et al., 2005). Furthermore, in that time the first “large-scale” expression analysis in the mice *Apodemus flavicollis* was performed (Tanić et al., 2005). Using differential display reverse transcription-polymerase chain reaction (DD RT-PCR) technique, it was possible to detect specific and differentially expressed poly-A RNAs in the B samples. Similarly, using amplified fragment length polymorphism polymerase chain reaction (PCR-AFLP) with thousands of primers combinations, some of B-transcripts were reported in *Secale cereale* (Carchilan et al., 2009) and *Zea mays* (Lin et al., 2014). Through this PCR-based techniques, besides some B-specific transcripts it was possible to detect the B-dose impact in the expression over some genes (Lin et al., 2014).

Using primers of B chromosome sequences the transcription of some protein-coding genes in the cDNA of a roe deer were detected (Trifonov et al., 2013). Since BACs sequencing, several sequences were described, and it was possible to determinate some B-specific mutation in the B-cDNA. This is one of first works to investigated the B chromosome transcription in a vertebrate.

With all these molecular biology advances, the B chromosome runaway from inert characteristics to be described as a potential extra element with some transcription activity and gene expression impact in the host genome (Carchilan et

al., 2007). Nonetheless, the PCR techniques have two main limitations to recover B chromosome transcripts. One of them is the limit of primer combinations to perform PCR based on multiple primers techniques, such as AFLP (Carchilan et al., 2009). Another limitation is that B chromosome sequences usually are paralog copies from host genome, in this way, there is some difficult to distinct the A and B transcripts. These two reasons could lead to underestimated the B chromosome activity (Carchilan et al., 2009; Kato et al., 2005). In this way, the next step is identified the duplicated copies in the B chromosome based on single nucleotide polymorphism, as is possible in large-scale sequencing.

The B chromosome synonymous is “selfish element” generally because it has ways to escapes the meiosis elimination and segregates in a non-mendelian rate, increasing their number in the population (Houben, 2017). The extreme case of selfishness is the conflict of the haplodiploid wasp *Nasonia vitripenis* B chromosome. The B chromosome, also called Paternal Sex Ratio (PSR) in this species is responsible to eliminated the male genome in fertilized eggs, and instead of generate a female embryo, the +PSR fertilized eggs rise a male embryo (Aldrich et al., 2017). In this selfish scenario the first B transcriptome profile was generated, which reported nine PSR-specific transcripts, both coding and putative non-coding genes (Akbari et al., 2013). Since, we enter into RNAseq era. Additionally, both long and small RNAs in the *N. vitripenis* B chromosome-like were investigated (Li et al., 2017). Based on this first B-transcriptome, a putative new coding gene which could be responsible to cause the paternal genome elimination was described using RNAi technique (Benetta et al., 2020).

Gathering RNAseq and genome data, transcripts with B-specific mutations was identified, meaning that B genes are transcribed (Carmello et al., 2017; W. Huang et al., 2016; Ma et al., 2017). Besides, towards the “RNAseq era” is possible detect that B chromosome presence impacts in the A complement expression (Huang et al., 2016). That affirmation was propose in the past decades with the first B chromosome and RNA studies (Ayonoadu & Rees, 1970; Kirk & Jones, 1970; Lin, 1955). Exploring this purpose, several B-organisms had its RNA content sequenced: plants (Huang et al., 2016; Huang et al., 2020; Ma et al., 2017), insects (Akbari et al., 2013; Li et al., 2017; Navarro-Domínguez et al., 2017), fishes (Carmello et al., 2017;

Mazzoni et al., 2021; Nascimento-Oliveira et al., 2021) and mammals (Trifonov et al., 2013), are some examples.

In the grasshopper *Eyprepocnemis plorans*, through 0B and B+ RNAseq data was possible to identify at least 5 coding genes residing in the B chromosome which are actively transcribed (Navarro-Domínguez et al., 2017). In the RNAseq era besides to identify B-related transcripts in large scale data, is possible to focus at nucleotide level and find B-specific single nucleotide polymorphisms (Jie et al., 2019). And also investigate the functions that these transcripts are involved (Benetta et al., 2019; Jie et al., 2019).

Overcoming the “omics” age, more recently, the performing RNAi techniques shed light into B transcripts function on the cell. In fact, (Benetta et al., 2019) suggested that the better way to investigate B-copy and its functions is through gene silencing, such as CRISPR techniques. This prediction became reality and show putative molecular phenotypes involved to B chromosome genes (Benetta et al., 2020; Imarazene et al., 2021). Using RNAi technique, a gene in the PSR of *N. vitripennis, haploidizer*, was identified as responsible for the paternal genome elimination (Benetta et al., 2020). In turn, in *Astyanax mexicanus*, which the B chromosome either are absent or rare in females, through CRISPR a putative master sex gene carried by the B chromosome was detected. Promoting knock-out in the *growth differentiation factor 6b* gene, which is copied in the B chromosome, the B-male fish became females (Imarazene et al., 2021).

In this way, joining all the RNA and B chromosome timeline, we highlight five big eras: (1) physical and chemical experiments; (2) chromosome and RNA hybridization; (3) PCR era; (4) RNAseq and large-scale omics era; and (5) RNAi and CRISPR era (**Figure 1**).

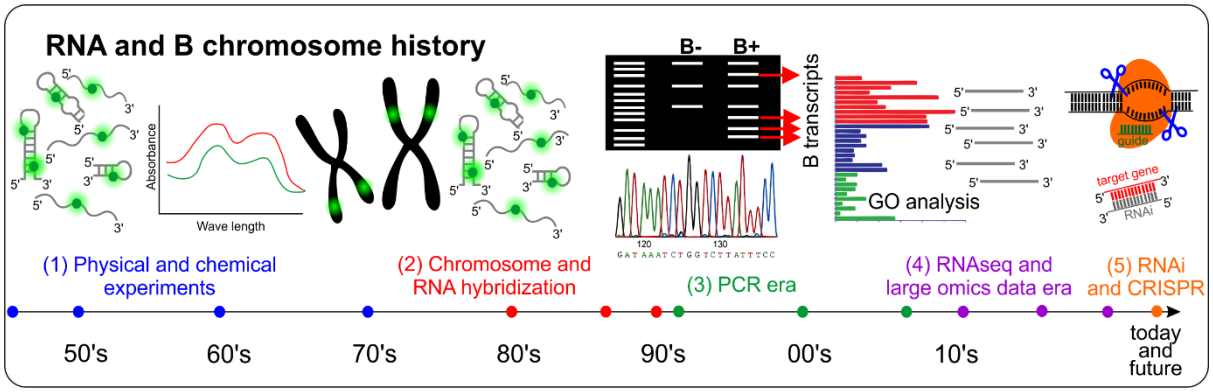

Figure 1 – The timeline of RNA studies and B chromosome.

B-transcribed sequences: functions and impacts

Repetitive DNAs are extensively studied in the B chromosomes and these extra elements are enriched by tandem and disperses repeats. In this way, reports about rRNA and TEs transcription by B chromosomes also have been explored. In the grasshopper *E. plorans*, B specific transcription of rRNA sequences was detected (Ruiz-Estévez et al., 2012, 2014) (**Figure 2A**). In the same way, LTR elements from B chromosome showed able to be transcribed in maize (Lamb et al., 2007) (**Figure 2B**). Several others studies has investigated the transcription activity of these B-repetitive regions, suggesting a role for epigenetic effects from these elements (Coan & Martins, 2018; Leach et al., 2005).

On the other hand, in a functional view, the RNA can be simply divided in two main groups: (1) coding sequences, the messenger RNAs (mRNAs); and (2) non-coding sequences, the non-coding RNAs (ncRNAs) (Palazzo & Lee, 2015). In this way, we report which coding and non-coding RNAs have already been described related to B chromosome, and which function its sequences can revel (**Figure 2**).

There are several reports for transcribed mRNAs from B chromosomes (Akbari et al., 2013; Benetta et al., 2019; Navarro-Domínguez et al., 2017; Valente et al., 2014). Looking for a messenger, the investigation about coding-genes allow to understand which proteins and consequently which processes can be affected by B chromosomes (Ahmad et al., 2020; Ahmad & Martins, 2019). Kinesin genes, for example, are found in B chromosome of a grasshopper (Navarro-Domínguez et al., 2017), a fish (Valente et al., 2014) and a plant (Banaei-Moghaddam et al., 2013). Three obviously distant organisms showing convergent evolution linked to putative functions in the B chromosome (Ahmad et al., 2020). An extensive B-mRNA gene

list, with known or none evidence of transcription is reviewed by (Benetta et al., 2019).

The functionality of B-proteins also was tested in rye. Carrying B-copy of a Argonaute gene, the transcription and translation were confirmed *in vittro*, indeed the translated B-copy also presented RNA slicer activity (Ma et al., 2017) (**Figure 2C**). Argonaute proteins are fundamental for small RNA biogenesis (Thomson et al., 2015), these findings could indicate the B-coding genes participation in such processes as discussed further. Besides known function from B-copy genes, new functions can be discovered in the B chromosome sequences. In the wasp *N. vitripennis* the PSR carry a new coding gene called *haploidizer* which has no similarity with a known protein and could indicate a new gene originated in the supernumerary chromosome (Benetta et al., 2020) (**Figure 2D**).

B chromosome can carrying truncated and pseudogenes sequences it is not a big novelty (Banaei-moghaddam et al., 2015; D'Ambrosio et al., 2017). Nonetheless, the pseudogenes transcription investigation allows to hypothesized several points to how the B chromosome could impact in the A chromosomes expression (Banaei-moghaddam et al., 2015; Banaei-Moghaddam et al., 2013). Since B chromosome carries truncated copies with dozens of mutations lead to concluded that these degenerated sequences are useless. However, even these odd sequences seems to have functions (Banaei-moghaddam et al., 2015; Banaei-Moghaddam et al., 2013). Besides to contribute for the landscape of expressed sequences in host genome, the pseudogenes from B chromosome could compete by regulation biding site with the original "A" copy. For example, the small non-coding genes prefer to bind the truncated transcript from the B, that could increase the A translation rate (Banaei-Moghaddam et al., 2013). That might be a manner of a selfish element controls the host transcription its own benefit. An example of this interaction was proposed in a fish species. The *abca1* B-copy gene (ATP-binding cassette sub-family A member 1-like gene) has 80% of coding sequence compared to original A copy. This gene is target of differentially expressed miRNAs in the B+ samples. In this way, is suggested that this interactions could increase the *abca1* translation through B-copy under miRNA regulation (Nascimento-Oliveira et al., 2021).

Regarding ncRNAs, these sequences can be subdivided into several groups (Mattick & Makunin, 2006).The classification may varies depend on context, since

the non-coding sequences can superposed its characteristics, such as length and function (Adelman & Egan, 2017; Costa, 2007; Palazzo & Lee, 2015). Here, we explore the RNA classification based on its function, in this way there are three main classes for non-coding sequences: (1) housekeeping, (2) intergenic and (3) regulatory.

The housekeeping ncRNAs are involved to transcription and translation processes, and are strongly conserved over the tree of life (Boivin et al., 2019). In fact, the ribosomal RNA, for example is focus of many studies in the B chromosome (Rajičić et al., 2015). The presence of active nuclear organization regions in the B chromosome can indicate the involvement of this extra element into fundamental processes of the cell (Valente et al., 2017). However, there is no information about other housekeeping ncRNAs and B chromosome. Investing studies in other housekeeping ncRNA, such as tRNAs and snoRNAs could bring information if the B chromosomes have impacts in the cell transcription and translation processes.

The non-coding part of mRNAs are important for stabilization and localization of coding transcripts in the cell, and 3' and 5' non translated regions are site binding for many regulatory molecules, such as small RNAs. Also, the intron sequences are related to protein translation variability through alternative splicing (Jeffares et al., 2006). We did not find any study involving intergenic ncRNA and B chromosome. Once these regions do not carry codon information, the selection pressures are more relaxed. Plus, due to the weekly natural selection presence in non-recombined chromosomes, as B chromosomes (Camacho et al., 2000; Mahajan & Bachtrog, 2017), how is the impact of mutations in these regions? Is there some impact to half-life of B-derived mRNAs? Do these regions keep the secondary structure similar to the original gene? On the other hand, 3' and 5' UTR carry sRNA binding sites, B-derived pseudogenes could compete to original copy and increase the A copy translation (Banaei-moghaddam et al., 2015). Additionally, introns are a large part in the coding genes, it can originate regulatory molecules during the mRNA processing, such as miRtrons and snoRNAs (Boivin et al., 2019). In this way, investing studies in these non-translated portions could shed light to B chromosome relations with mRNA processing.

The third class based on ncRNA function is regulatory sequences, which are divided into two main groups based on RNA length: the long-noncoding RNAs with

more than 200 nucleotides and the small RNAs, shorter than 200 nucleotides (Costa, 2007). Regulatory ncRNA are involved in dozens of processes, such as chromatin remodeling, gene expression and posttranscription regulation (Mattick & Makunin, 2006).

Concerning this class, there is some studies which explored regulatory ncRNAs and B chromosome impacts (Huang et al., 2020; Li et al., 2017; Ramos et al., 2017) (**Figure 3F-G**). A long non-coding RNA is largely spread over *A. latifasciata* B chromosome (Ramos et al. 2017). This sequence, the BncDNA, is expresses at least in eight tissues, and the levels could carry a sex bias (**Figure 2F**).

Expressed microRNAs sequences were found in maize B chromosome and also other miRNA genes seems to impact the A miRNA expression (Huang et al., 2020) (**Figure 2G**). In *N. vitripennis*, several sRNA sequences similar to miRNA, siRNA and piRNAs were found transcribed by PSR chromosome (Li et al., 2017). In *A. latifasciata* cichlid fish, three piRNA clusters were detailed describe in B chromosome. These sequences might be related to mobilization silencing control due to B-TE accumulation (Nascimento-Oliveira *in prep.*).

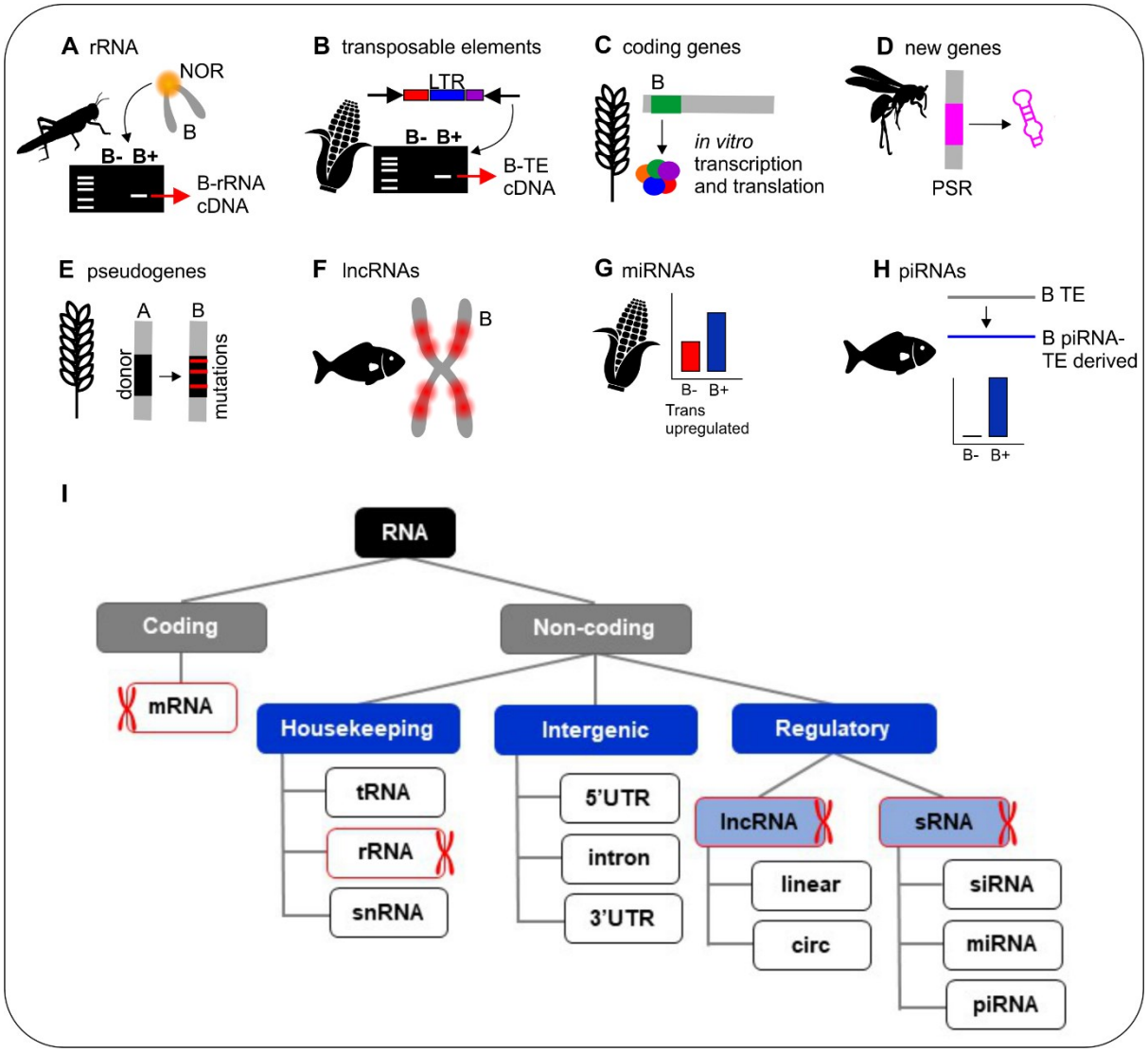

Figure 2 – B-transcribed sequences A-H examples of RNA sequences transcribed by B chromosome. **I** RNA overview classification and B chromosome studies. The classes which have reports in the B chromosome are highlighted with the red chromosome icon.

Joining the available information about which genes were described related to the B chromosome it was possible to draw which classes lacked description in the B-carriers (**Figure 2I**).

The small ncRNAs also have been poorly explored in the B chromosome studies. Through the small RNA sequencing, which get sequences over 10-60 nucleotides is possible characterized several types of non-coding genes (Boivin et al., 2019). Based on genome and RNAseq alignments we filtered the sRNAseq from *A. latifasciata* gonads with B chromosome and without B chromosome and show what kind of non-coding molecules could be described using small RNA sequencing (**Figure 3**). Into 15 to 35 nucleotides is possible to recover several small RNAs, as

well as, siRNAs, miRNAs, piRNAs and tRFs (transporter RNA fragments). Considering only the characterized miRNA (Nascimento-Oliveira et al., 2021) and piRNA (Nascimento-Oliveira *in preparation*) reported in previously studies, we count how much reads for these sRNA classes each sample (female or male, B- or B+) express. It is possible to observe that the miRNA and piRNA reads increase in the B+ (**Figure 3**).

Besides the known small RNA classes that is an unexplored universe to invest B chromosome studies, there is the possibility to describe potential new functional short sequences impacted by B chromosome. Several mid-size noncoding RNAs, range from 36 to 52 nucleotides (Boivin et al., 2019) are differentially present in B- and B+ samples of *A. latifasciata* gonads RNAseq (**Figure 3**).

In this way, besides the mRNA sequences, which is the small part of B chromosomes, there is a wide ncRNA field to deeply explore and suggest new functions related to the extra chromosomes.

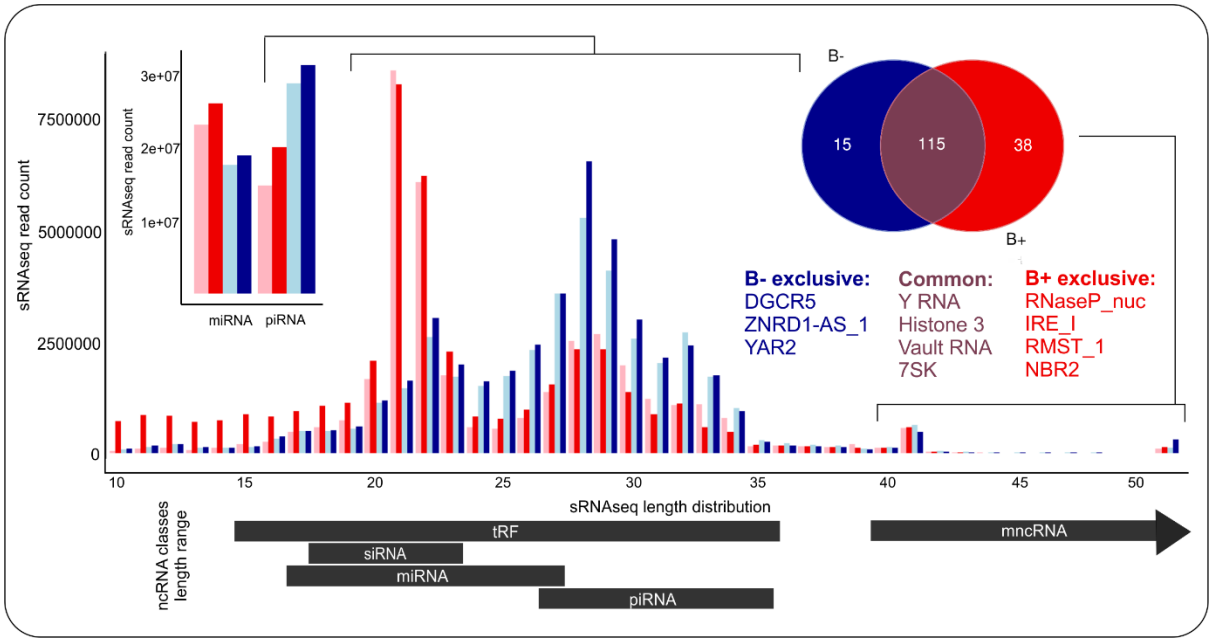

Figure 3 – Small RNAs from *A. latifasciata* gonads. The colors in the bar plot mean: light red, female B-; red, female B+; light blue, male B-; blue, male B+. Bellow to the bar plot with sRNAs histogram there is the specific sRNAs classes range length. The Venn diagram shows the number of mid-size RNAs in each group, bellow that some one of these families.

The B chromosome tree of life: an RNA perspective

It is a consensus in the B community that the extra chromosome are observed over the tree of life, and the increased number of species in specific clades is not only related to biology, but to higher number of cytogenetic studies in some groups (D'Ambrosio et al., 2017; Green, 1990; Jones, 2017).

In this way, we construct a "B chromosome tree of life" and highlight which clades have reported B chromosome presence and which ones have studies with B-RNAs. Firstly, we list the groups with B chromosome based on the two most recent reviews (D'Ambrosio et al., 2017; Jones, 2017). (D'Ambrosio et al., 2017) constructed a database of B chromosome studies from 1906 to July, 2016. The report about this database found 2828 species with B chromosome, and the more representative clades, as expected, are plants, insects, mammals and Actinopterygii fishes. Additionally, we also complement our list with (Jones, 2017) review. This report released the new species with B chromosome describe since 1980.

We also performed a manually search about the groups not represented in the two previous reviews to check if some new publications were recently available. In that way, some new groups were included in the B chromosome tree of life, are they: Platyhelminthes (Špakulová & Casanova, 2004), Annelida (Leitão et al., 2010), Mollusca (Leitão et al., 2008), Nematoda (Špakulová & Casanova, 2004) and avian (Hansson, 2019) (**Figure 4**).

Particularly, the B chromosome of avian are reported as a supernumerary germline-restricted chromosomes (GRCs) and are spread over the finches and song-birds (Hansson, 2019; Torgasheva et al., 2019). The avian species, also showed B chromosome with huge variation, from micro-chromosome to mega size chromosomes. That also indicates the polyphyletic event in the B chromosome origin and rapid phenomenon of gain and losses among the sister-organisms (Hansson, 2019; Torgasheva et al., 2019).

When the reviews mentioned "fishes" are meaning Actinopterygii species, the numerous, diverse, and most studied group (D'Ambrosio et al., 2017). In this way, both Chondrichthyes and lobe-ray fishes have not report about B chromosome presence (**Figure 4**). Algae, Porifera, Cnidaria and Equinodermata also have not report of B chromosome occurrence.

Overlapped to the B-tree of life, we also highlighted which clades have RNA investigation applied to B chromosome questions (**Figure 4**), and could confirm that the most studied species also are the species with B-RNA report.

Usually studied organisms through the “art-state” technologies applied to B questions are the model and most studies species, such as mammals, fishes, insects and plant with high economic importance. That is extremely related to cytogenetic studies: the numerous species reported to carry B chromosome are clades with some interest, like the mammals and fishes, and plants with some economic interest (D’Ambrosio et al., 2017).

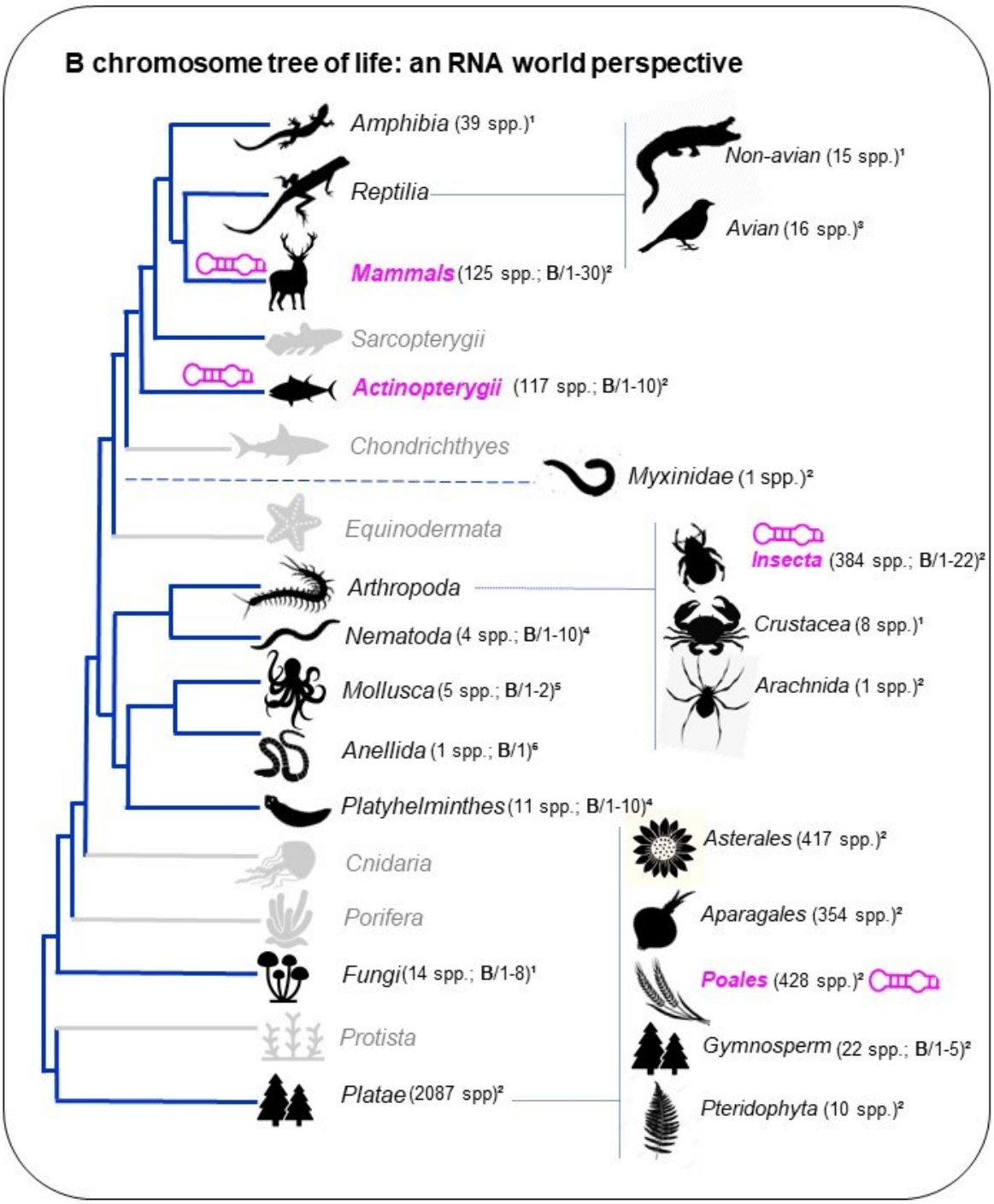

Figure 4- The B chromosome tree of life. The phylogenetic tree is descriptive representation (not scaled) of main eukaryotes groups. In blue branches are the organisms with B chromosome report, in gray branches the clades without description of B chromosome in the group. Some groups are detailed in the light line branches. Inside the brackets, some branches with either number of species or B chromosomes available in the literature. The groups with RNA investigation applied to B chromosomes are showed in bold pink letters with RNA icon. The references follows: (1) (Jones, 2017); (2) (D'Ambrosio et al., 2017); (3) (Hansson, 2019); (4) (Špakulová & Casanova, 2004); (5) (Leitão et al., 2008); (6) (Leitão et al., 2010).

Final remarks

Cytogenetics analysis are still need to complete the B-tree of life, as well as, the functional studies to better understand the convergent and divergent points of this polyphyletic phenomenon.

Material and Methods

The small RNAseq from *A. latifasciata* is available in NCBI database (SRR13040679 to SRR13040710b). The raw reads were filtered to quality based on (Nascimento-Oliveira et al., 2021) steps (with modifications – the range of filtered reads were performed along 15 to 60 nucleotides). To remove the redundancy in the libraries, the filtered reads were collapsed using TBr2_collapse.pl (Rosenkranz & Zischler, 2012) then aligned to *A. latifasciata* B- and B+ genome assemblies using sRNAmapper (Rosenkranz & Zischler, 2012). The aligned sequences were identified by Infernal v1.1 (Nawrocki & Eddy, 2013) based on last released Rfam database (Griffiths-Jones et al., 2003). The identified ncRNAs in B- and B+ libraries were compared in a Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/).

References

The references used in this chapter will be presented at the end of thesis.

3.4 Capítulo 4

"Divulga-se, enfim (e por que negar?), para satisfazer o desejo que alguns sentem, de partilhar com muitos outros o produto de sua experiência, adquirida seja diretamente no curso de seu próprio trabalho criador, seja mediante a absorção de informação colhida em fontes menos acessíveis ao grande público e o esforço de compreender essa informação, de situá-la dentro do quadro geral do conhecimento e de analisa-lhe as possíveis implicações. Esse desejo, essa alegria de transmitir fácil o que se obteve difícil."

— José Reis

Relatos sobre divulgação científica: democratizar o conhecimento é construir um futuro melhor

Percepção sobre a Ciência brasileira

O Ministério da Ciência, Tecnologia, Inovações e Comunicações e o Centro de Gestão e Estudos Estratégicos, através da pesquisa “Percepção Pública da Ciência e Tecnologia no Brasil”, realizada no ano de 2019, diagnosticaram que 73% dos brasileiros acreditam que a ciência e tecnologia trazem benefícios a sociedade. Esse resultado otimista é somado a outros 86% que confiam na ciência para o desenvolvimento da indústria e geração de mais oportunidades.

Através da mesma pesquisa, 62% dos entrevistados responderam ter algum interesse em ciência e tecnologia, e 41% considera os cientistas “pessoas inteligentes que fazem coisas úteis à humanidade”. No entanto, mesmo com esse alto interesse por ciência, a pesquisa também revelou que aproximadamente 90% dos entrevistados não souberam apontar o nome de um cientista brasileiro ou de uma instituição que realizasse pesquisa científica (**Figura 1**). O interesse elevado em ciência, mas a falta de afinidade com o assunto também se revela na percepção acerca de temas específicos: 78% dos entrevistados acreditam que antibióticos podem matar vírus.

As conclusões do assunto levam a acreditar que, no geral, a população entende a importância da ciência e sua relevância para a sociedade, mas sabe muito pouco sobre ela. Mesmo assim, 66,1% concordam que os governantes devam seguir as orientações dos cientistas, e 82,2% concorda que a maioria das pessoas é capaz de entender o conhecimento científico se ele for bem explicado.

A partir desses dados levanta-se o seguinte questionamento: quais fatores são responsáveis pelo desconhecimento sobre ciência? O jornalista científico Herton Escobar fala sobre um “abismo que sempre existiu” (Escobar, 2018). Esse abismo também pode ser atribuído ao acesso à Ciência, já que uma flutuação entre 61% a 93% dos entrevistados raramente ou nunca consomem conteúdos relativos à ciência nas mais diversas mídias (internet, TV, livros, jornais impressos, entre outros).

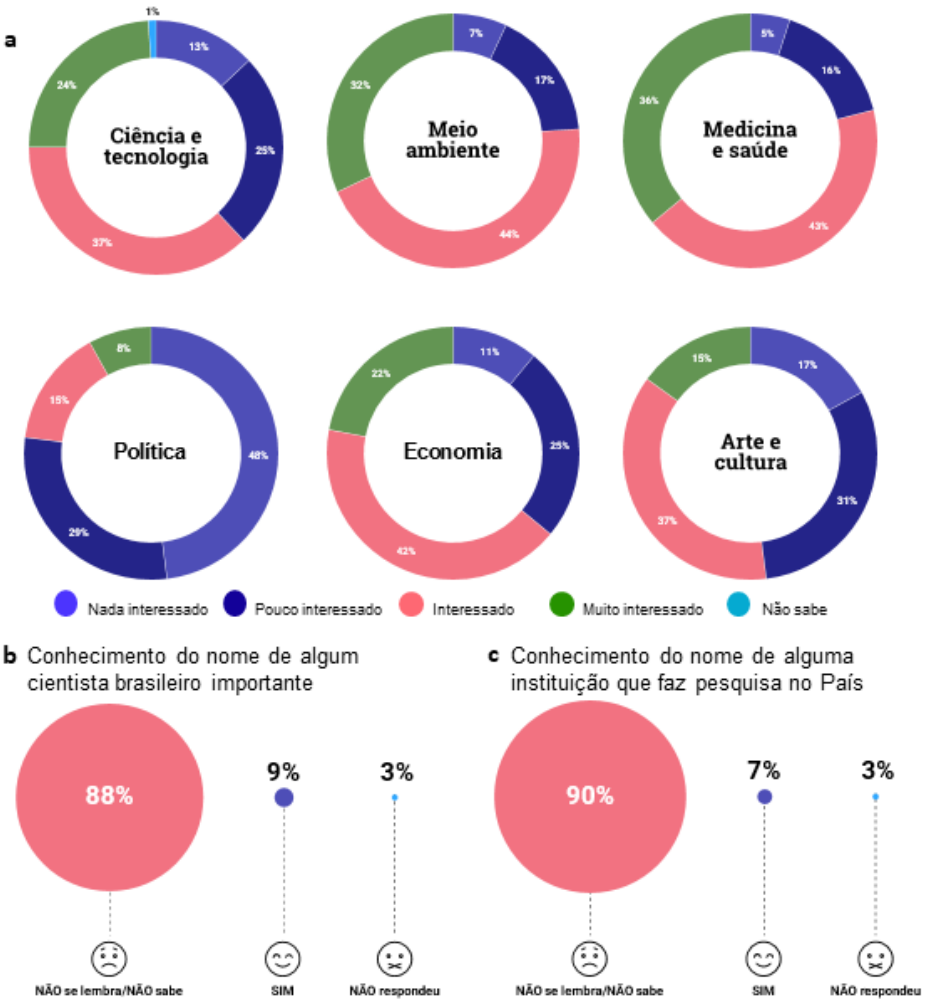

Figura 1 – Percepção da população brasileira sobre temas diversos. Fonte: Percepção Pública da C&T no Brasil 2019, CGEE, MCTIC.

É nesse cenário em que apresento as minhas contribuições para a divulgação da ciência brasileira baseando-me nos objetivos traçados por José Reis do porquê divulgar ciência (Massarani et al., 2018). Dois canais principais foram meu portal de comunicação da ciência com a sociedade. O Voando Alto Ciência reúne conteúdos que vão desde desmistificação de notícias falsas, relação sobre ciência e políticas públicas até temas relacionados a minha formação: Genética e Evolução. Esse perfil publica seus conteúdos no Instagram e YouTube.

Participo também como membro da equipe do “Ciência brasileira é de qualidade”, canal dedicado a divulgar os artigos científicos de pesquisadores brasileiros de forma mais acessível ao público leigo. O canal está ativo com perfis no Facebook, Instagram e Twitter.

- **“Divulga-se para atender à “fome individual de ciência”, [...] porque alguns dos feitos da ciência, ou que dela decorrem, se tornam ‘notícia’ de grande feito até sobre a massa dos que não tem ‘fome de ciência’.”** (Massarani et al., 2018)

No relatório da pesquisa “Percepção Pública da Ciência e Tecnologia no Brasil 2019”, 79% dos entrevistados disseram ter interesse em assuntos de medicina e saúde, enquanto 76% em meio ambiente (**Figura 1**). Certamente, precisaríamos retomar a pesquisa para investigar o impacto da pandemia de COVID-19 sobre esse aspecto.

Um dos trabalhos realizados envolveu a parceria com uma escola do interior do estado de São Paulo. Em contato com professores, foi encaminhado uma série de perguntas relacionadas à pandemia que seriam respondidas por um cientista (**Tabela 1**). A partir dessas perguntas dois vídeos foram elaborados (Parte 1: youtube.com/watch?v=OLhCJ2yVdfY&t=27s e Parte 2 youtube.com/watch?v=LTIGLuRoHj4&t=123s). Os alunos tiveram acesso as respostas e reproduziram o conteúdo aprendido na forma de desenho (**Figura 2**).

Tabela 1 – Perguntas formuladas pelos alunos que serviram como base para a construção dos vídeos sobre pandemia

Dúvidas em relação à pandemia
Por que o coronavírus tem contágio rápido?
Por que estamos observando ondas de contaminação?
O uso de máscara é realmente eficaz?
Como surgiu o coronavírus?
Por que algumas pessoas são assintomáticas?
Como o vírus entra no nosso corpo?
Há previsão de que a pandemia acabe?
Como acontecem as mutações do vírus?
As vacinas são eficazes? Qual a melhor?
Por que algumas vacinas causam reação?
Por que mesmo depois de tomar vacina a pessoa pode se contaminar?
A vacina pode causar alguma doença?

Figura 2 – Desenhos produzidos pelos alunos após assistirem aos vídeos sobre a pandemia. Os nomes dos alunos foram ocultados pela professora.

A professora relata que os alunos ficaram surpresos ao saberem que existem cientistas em uma cidade tão próxima a deles. O que vai ao encontro da pesquisa sobre a percepção da Ciência brasileira citada anteriormente, 90% dos entrevistados não conseguiram citar o nome de um cientista brasileiro, enquanto 88% não souberam dizer o nome de uma instituição brasileira de pesquisa (**Figura 1**).

Outros temas que repercutem fortemente na mídia estão relacionados a Genética e Evolução. O interesse em genética vem da curiosidade sobre doenças, testes de ancestralidade e até mesmo temas polêmicos como o racismo. Utilizei os conhecimentos de Genética para produzir dois vídeos sobre a variedade das

populações humanas e explicar o porquê não existem raças humanas na nossa espécie. Um deles trata do histórico do racismo, e porque movimentos como darwinismo social e frenologia (medição de crânios humanos) são pseudociências que devem ser fortemente repreendidas (link para acesso: youtube.com/watch?v=YTIMSVU1CB0&t=179s). No segundo material, a partir das novas pesquisas de sequenciamento de genomas e genética de populações expliquei a importância de sabermos genética no combate ao racismo (link para acesso: youtube.com/watch?v=3HwemiKjwDs). Esse último material foi utilizado em uma aula de ciência, onde a professora gentilmente me mandou o registro (**Figura 3**).

Figura 3 – Uso de vídeo produzido no canal Voando Alto Ciência em uma aula de Ciências do ensino básico.

Alguns temas relacionados a Genética também são discutidos informalmente entre amigos, e muitas vezes as pessoas não tem conhecimento da relação de um assunto do cotidiano e a Ciência por trás dele. Uma situação dessas motivou o artigo publicado na revista de educação “Genética na Escola”.

Em um momento descontraído com amigos, foi questionado a possibilidade dos cientistas “criarem uma pílula para cessar o encurtamento dos nossos telômeros e assim vivêssemos jovens para sempre”. Como geneticistas sabemos da importância dos telômeros para o controle do ciclo celular e vida útil saudável do

nosso organismo. Partindo desse assunto “telômeros” e “imortalidade” o artigo “Conectando os telômeros e o envelhecimento: é possível viver para sempre?” define a biologia das extremidades dos nossos cromossomos e responde o porquê a imortalidade não é tão viável assim (Nascimento-Oliveira, 2020) (**Figura 4a**). O artigo pode ser acessado através do link (precisa ser copiado e colado no navegador): https://www.geneticanaescola.com/_files/ugd/b703be_1160d23fd79047c094f16bbd837494e3.pdf.

E por que não falar de Genética de uma forma divertida? O Voando Alto Ciência promoveu uma transmissão ao vivo com integrantes do projeto temático “Cromossomos B, cromossomos sexuais e seus enigmas” (FAPESP 2015/16661-1) em um encontro para divulgar curiosidades sobre cromossomos. A transmissão “Arraiá dos cromossomos” contou com a presença de diversos profissionais da área, estudantes do ensino médio, estudantes de graduação e população em geral de diversas regiões do Brasil (**Figura 4b**). O evento está disponível em youtube.com/watch?v=cj_QAl231ZA&t=1547s.

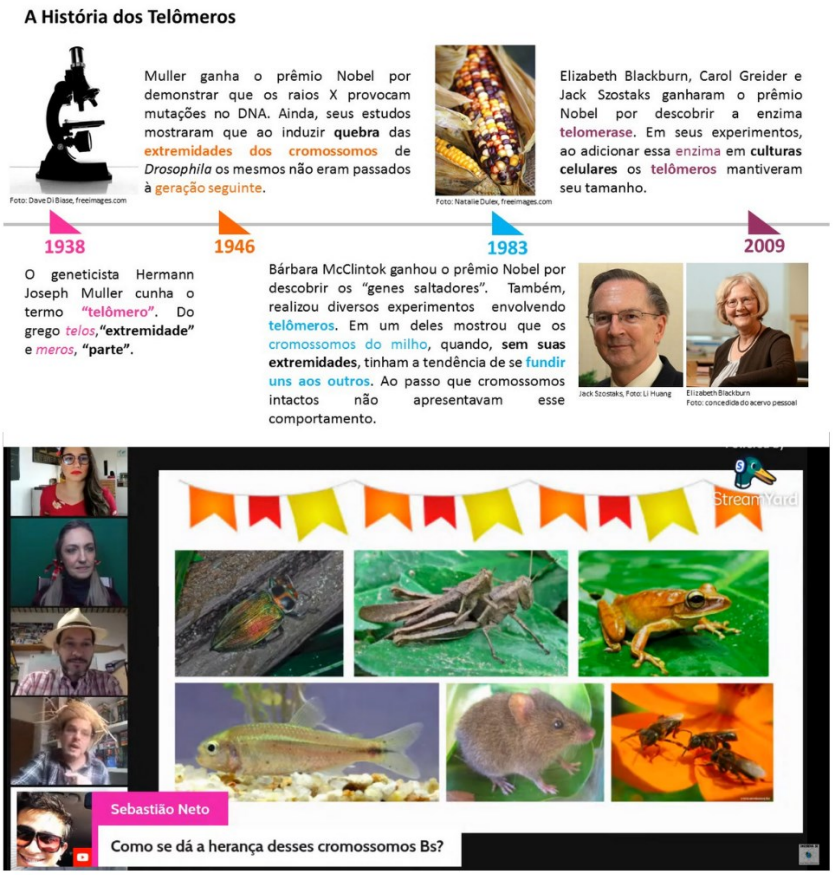

Figura 4 – A genética no cotidiano. Acima: Imagem publicada no artigo “Conectando os telômeros e o envelhecimento”. Abaixo: Evento “Arraiá dos cromossomos” transmitido no YouTube.

- **“Divulga-se no interesse da própria ciência, e por influência dos cientistas ou dos que compreendem o valor da ciência no mundo moderno, para conseguir apoio cada vez maior para as atividades científicas.”** (Massarani et al., 2018)

É preciso aproximar a população da realidade da Ciência no Brasil. Sozinhos, cientistas não são capazes de movimentar multidões em prol da Ciência. Essa reflexão é colocada por Herton Escobar no artigo “Divulgação científica: faça agora ou cale-se para sempre” (Escobar, 2018):

> *A comunidade científica não pode mais delegar à imprensa a responsabilidade de educar a sociedade sobre a importância da ciência — porque não cabe a ela essa responsabilidade, e mesmo que coubesse, ela não tem condições de fazer isso sozinha. O abismo é fundo demais para ser preenchido só com folhas de jornal e alguns minutos de televisão.*

Historicamente, no Brasil, os cientistas nunca precisaram diretamente do apoio da população para conseguir fomento para a Ciência. Desta forma sempre existiu um abismo entre os dois lados, cientistas e sociedade, que não gerava incômodo. No entanto, a atual situação das pesquisas brasileiras, sucateamento de universidades, cortes de bolsas e projetos, mostra que sozinhos os cientistas não conseguirão apoio que precisam (Escobar, 2018). Nesse sentido, o Voando Alto Ciência também dedicou diversas postagens para explicar a população qual o impacto das decisões políticas para a pesquisa e educação no nosso país (**Figura 5**).

O estrangulamento do investimento à Ciência brasileira compromete a soberania nacional (André et al., 2021). As consequências do sucateamento das verbas repassadas à ciência ativam um evento cascata sem fim. Vivemos a atual “fuga de cérebros” em que diversos pesquisadores encontram refúgio e amparo para continuar a se dedicar à ciência fora do Brasil. Outra fuga são de centenas de milhares de jovens doutores desempregados, que sem alternativas e perspectivas acabam migrando para outra área para garantir meios de sustento. Estima-se que a

taxa de desemprego entre os recém doutores no Brasil fica entre 35 e 25%, contra 2% no resto do mundo (Rosa, 2020).

Figura 5 – A Ciência e a política brasileira. Conteúdos produzidos pelo Voando Alto Ciência explicando as relações e o impacto das decisões políticas na ciência.

Esse desperdício de mão de obra altamente qualificada desencadeia a dependência do país para se importar soluções científicas e tecnológicas. Um exemplo é a corrida pelas vacinas durante a pandemia de COVID-19, se não fossem os laboratórios das instituições públicas, o Brasil não teria como produzir suas próprias vacinas. As consequências do constante abandono da Ciência apontam para um futuro incerto, dependente e nada soberano (André et al., 2021).

Desta forma, é urgente divulgar a situação em que a Ciência brasileira se encontra (Escobar, 2018). Um dado simples a ser apresentado é comparar o fomento à Ciência de 20 anos atrás. O valor de 4,8 bilhões de reais investidos em Ciência em 2000, se corrigidos pela inflação e seguissem o mesmo montante para os dias de hoje (ano 2021), deveríamos ter um investimento de aproximadamente 25 bilhões de reais. No entanto, em 2021 o investimento para Ciência foi de 2,8

bilhões de reais, metade do valor, sem correção, de 20 anos atrás. Considerando a correção, o valor investido corresponde a apenas aproximadamente 11% do investimento feito em 2000 (cálculo IGPM-FGV https://www3.bcb.gov.br/CALCIDADAO/publico/exibirFormCorrecaoValores.do?method=exibirFormCorrecaoValores).

Devemos lembrar que o contingente de alunos de pós-graduação, pós-doutorandos e projetos de pesquisa aumentaram significativamente nas últimas décadas. Em 1996 o número de pós-graduandos foi de 67.820 alunos, em 2021 esse número cresceu para 122.295 alunos (PORTAL MEC). Ou seja, apenas corrigir o valor da inflação não bastaria para garantir o progresso e o crescimento da Ciência em nosso país. Informações como essa deveria se tornar públicas, Ciência é cara, mas não é gasto, é investimento.

- **"Tentei ensinar aos outros o que de outrem não pude aprender." Dom Duarte Nunes de Leão** (Massarani et al., 2018)

Um dado otimista do relatório sobre a Percepção Pública da Ciência e Tecnologia no Brasil 2019 é que 82% dos entrevistados acredita que "a maioria das pessoas é capaz de entender o conhecimento científico se ele for bem explicado".

Pode-se concluir, que de forma geral, a população brasileira entende a importância da Ciência para o avanço da sociedade, mas sabe pouco sobre ela (Jornal da USP, 2019). Sobre esse diagnóstico podemos atribuir dois desafios: (1) acesso à informação de qualidade e (2) educação científica.

Segundo o relatório sobre a Percepção Pública da Ciência, a maioria das pessoas buscam assuntos relacionados a ciência em sites como YouTube e Facebook. No entanto, estudo aplicado na América Latina revela que mais de 60% dos brasileiros entrevistados não sabem reconhecer uma notícia falsa (Rodrigues, 2020). Padrão semelhante repete-se para outros países da América Latina e beira os 40% no Reino Unido, como revelou outra pesquisa (Breakstone et al., 2021). Desta forma, como a população sabe se a informação que ela consome na internet está cientificamente correta? A conclusão: não sabe (Breakstone et al., 2021).

Os obstáculos para se popularizar a ciência é muito mais que desmistificar notícias falsas. É preciso fortalecer a sociedade com o pensamento crítico, para que

ela por si só tenha autonomia de reconhecer uma pseudociência ou uma fonte duvidosa e assim não espalhar desinformação (Gomes et al., 2020; Santos, 2007).

Desta forma, torna-se inviável combater a desinformação em si, pois, toda vez que encontrarmos uma notícia falsa para desmistificá-la estaremos um passo atrás, uma vez que essa desinformação já foi passada adiante. Infelizmente os algoritmos promovem engajamento e nem sempre a ciência consegue competir com publicações sensacionalistas nas redes sociais (Gomes et al., 2020). O letramento científico, ou educação científica, auxiliaria na prática social da população leiga, para que ela por si só possa reconhecer as armadilhas das redes sociais (Santos, 2007).

E for fim, o acesso à Ciência é um direito descrito no artigo 27 da Declaração Universal dos Direitos Humanos: “Todo ser humano tem o direito de participar livremente da vida cultural da comunidade, de fruir as artes e de participar do progresso científico e de seus benefícios” (Piovesan, 2014).

É direito do ser humano ter acesso ao conhecimento científico e usufruir do seu progresso, e nesse quesito não limitando-se apenas ao conhecimento formal, mas também acesso à informação científica fora do ambiente escolar (Orsi, 2021). Desta forma, popularizar a ciência e promover educação científica é garantir um direito fundamental.

3.5 Considerações finais: discussão dos capítulos

"Na vida não há nada a temer, mas a entender"

– Marie Curie

O que os ncRNAs tem a revelar sobre o cromossomo B

Nos capítulos 1 e 2 foram apresentados os impactos do cromossomo B para a atividade de pequenos RNAs não-codificantes, mais especificamente miRNAs e piRNAs. Devido a fácil identificação de sequências de miRNAs pela alta conservação, disponibilidade de estudos e depósitos em bancos de dados, esperava-se inicialmente encontrar genes de miRNAs no cromossomo B. Para os piRNAs, por consequência ao acúmulo de elementos transponíveis no B, a hipótese inicial era que algum mecanismo do genoma do hospedeiro estaria atuando contra os TEs do cromossomo B. Ambas as hipóteses foram refutadas.

A partir das diversas estratégias utilizando sequenciamento de DNA e RNA, nenhum miRNA foi encontrado no cromossomo B. miRNAs são sequências altamente conservadas nos organismos sob forte pressão de seleção (Bartel, 2018). Desta forma, a origem de miRNAs pode ser um evento raro, por estar sob fortes pressões seletivas, novas sequências são rapidamente eliminadas (Bartel, 2018; Loh et al., 2011). Além disso, a detecção de novos miRNAs torna-se mais difícil devido a baixas taxas de expressão que geralmente moléculas de origem recente possuem (Bartel, 2018). Adicionalmente, miRNAs são comumente localizados em regiões intergênicas, eucromáticas com alta densidade de genes codificadores de proteínas (Ledda et al., 2020; Lin et al., 2006). Portanto, o ambiente fortemente heterocromático, baixa pressão de seleção em cromossomos não-recombinantes e baixa densidade gênica não parece ser propício para a origem de novos miRNAs.

A relação entre miRNAs e cromossomo B de *A. latifasciata* parece ter muito mais impacto fisiológico do que genômico. Diversos miRNAs foram encontrados diferencialmente expressos comparando amostras com e sem cromossomo B de machos e fêmeas, de cérebro, músculo e gônadas. Esses miRNAs mesmo sendo tecido específico possuem alvos em comum, mostrando que os mesmos processos controlados por miRNAs estão sofrendo influência do cromossomo B em diferentes tecidos. Além disso já foi proposto que pseudogenes do cromossomo B quando transcritos competem com suas cópias originais (Banaei-Moghaddam et al., 2013). Desta forma, sequências regulatórias, como miRNAs, preferencialmente interagem com esses transcritos truncados do cromossomo B elevando a expressão da cópia do complemento A. Foram encontrados diversos miRNAs diferencialmente

expressos que tem como alvos genes identificados no cromossomo B. Uma interação com um gene com cópia incompleta no cromossomo B foi detalhada, aplicando a afirmação anteriormente proposta por (Banaei-Moghaddam et al., 2013).

Contrariamente, esperava-se encontrar piRNAs do genoma de *A. latifasciata* atuando no silenciamento dos TEs do cromossomo B. Uma alternativa mais parcimoniosa foi revelada, mostrando que o próprio cromossomo B carrega sequências de piRNAs originadas a partir da degeneração de inserções de TEs.

Diferentemente dos miRNAs, piRNAs evoluem rapidamente (Biscotti et al., 2017; Song et al., 2019; Yi et al., 2014). A localização genômica de *clusters* de piRNAs em regiões altamente repetitivas e heterocromáticas indicam a baixa pressão de seleção dessas sequências e uma grande possibilidade para surgirem e se manterem nos genomas rapidamente (Kofler et al., 2018; Luo et al., 2020; Wierzbicki et al., 2021). Desta forma, o cromossomo B reúne características de um ambiente propício a evolução de novas sequências de piRNAs: fortemente repetitivo e heterocromático. Outra questão que reforça a possibilidade de origem de novos *clusters* de piRNAs no cromossomo B é o aumento da proporção de elementos transponíveis em relação ao genoma do hospedeiro. Parece haver um equilíbrio entre o tamanho do genoma, cobertura de elementos transponíveis e piRNAs atuando no silenciamento de TEs (Kofler, 2020). Durante a evolução do genoma a expansão de TEs se não seguida de surgimento de piRNAs pode drasticamente diminuir a aptidão da população e consequentemente levá-la a extinção (Kofler, 2020). Baseando-se nessa ideia, sugere-se que para não interferir na aptidão da população, o cromossomo B carrega seus próprios piRNAs e mantém a proporção TE e piRNAs compatíveis com as novas inserções. Investigações em cromossomos B de outros organismos podem auxiliar na confirmação dessa hipótese.

Além disso, as proteínas envolvidas no processamento de piRNAs e silenciamento de TEs, tais como PIWI, possuem taxas de evolução mais rápidas se comparadas com as proteínas envolvidas na via de miRNAs (Tao et al., 2016; Yi et al., 2014). Essa diferença parece estar relacionada com a pressão que a rápida evolução de elementos transponíveis exerce na via de silenciamento (Tao et al., 2016). Mais especificamente em peixes, a diversidade de TEs é mais alta em comparação a mamíferos. Consequentemente isso pode refletir na variedade de proteínas adaptadas à via de piRNAs, onde a diversidade de TEs exerce pressão no

sentido de uma via apta a controlar um maior número de famílias de TEs (Song et al., 2019; Yi et al., 2014; Zhou et al., 2014).

Como reportado no Capítulo 2, o cromossomo B de *A. latifasciata* contém cópias do gene *pld6* (Valente et al., 2014), responsável por processar os transcritos precursores de piRNAs em transcritos primários (Ishizu et al., 2012; Kabayama et al., 2017). Além da validação da presença dessas cópias no cromossomo B, encontramos evidências da transcrição desse gene nas gônadas. Outrossim, *pld6* carrega sítios que estão sob pressão de seleção. Por estar transcricionalmente ativo e com sinais de seleção, essas evidências podem explicar o porquê esse gene não sofreu processo de pseudogenização no cromossomo B e trazem luz a funções em que esse elemento extra participa (Banaei-moghaddam et al., 2015; Banaei-Moghaddam et al., 2013).

Não somente pequenos RNAs, mas também uma infinidade de classes das porções não traduzidas que compõem o cromossomo B, podem trazer informações importantes para o entendimento da origem, função e comportamento desse elemento extra no genoma. Como revisado no Capítulo 3, o estudo da relação de RNAs e cromossomos B iniciaram concomitantemente aos avanços de técnicas de biologia molecular. Embora esses avanços tenham sido explorados pela comunidade que investiga cromossomo B, a caracterização funcional de RNAs não-codificantes na presença desse elemento extra é pouco explorada. Além de comporem boa parte do genoma de eucariotos, a porção não-codificante também está presente em alta densidade no cromossomo B, uma vez que esse cromossomo tem baixa concentração de genes codificadores de proteínas. A partir dessa revisão, sugerimos que investigações funcionais principalmente concentradas na caracterização da porção não-codificante do cromossomo B são perspectivas interessantes para desvendar a origem e comportamento desse elemento extra nos genomas.

O que é cromossomo B?

Um dos objetivos suplementares dessa tese foi produzir conteúdo de divulgação científica ao longo dos quatro anos de doutorado. Como apresentado no Capítulo 4, esses conteúdos estiveram associados a alguns objetivos explanados por José Reis sobre “o porquê divulgar Ciência” (Massarani et al., 2018).

Os resultados obtidos a partir do engajamento com os conteúdos vão ao encontro do último relatório sobre “Percepção pública da Ciência e Tecnologia 2019”: a população em geral tem interesse sobre ciência, mas sabe pouco sobre ela. Além disso, destaca-se a importância da educação científica para a construção de uma sociedade autônoma (Gomes et al., 2020; Santos, 2007).

Desta forma além de tornar a ciência acessível ela precisa estar em uma linguagem entendível ao grande público. Exercendo esse compromisso, esta tese conta com um “Resumo para a sociedade” que foi apresentado a alguns familiares. A avaliação feita por eles, e não por pares, rendeu algumas perguntas: o que é eucarioto? O que é RNA? O cromossomo B é igual síndrome de Down? E até algumas reflexões para poder explicar ciência: como as mutações do cromossomo B foram parar lá? Nem os cientistas sabem responder essa última, e é por isso que continuamos estudando.

4 Conclusão

- A presença do cromossomo B provoca alteração na expressão de miRNAs em diversos tecidos analisados. Ainda, foi sugerido que o cromossomo B pode in *cis* impactar diversos processos celulares através das interações produzidas pelos miRNAs diferencialmente expressos.
- O acúmulo de elementos transponíveis no cromossomo B originou novos piRNAs. Desta forma, devido ao aumento de TEs e consequentemente aumento do tamanho do genoma do hospedeiro pela presença do cromossomo B, sugere-se que outros cromossomos B de outros organismos também tenham experenciado eventos de origem de piRNAs, equilibrando a mobilização de TEs e evitando a eliminação do B.
- As evidências funcionais do cromossomo B através do estudo de RNA foram descritas em organismos extensivamente estudados, como insetos, peixes, mamíferos e plantas. Esse o padrão é encontrado para as evidências citogenéticas quanto a presença do cromossomo B. A árvore da vida do cromossomo B em uma perspectiva de RNA demonstra a importância de avançarmos tanto nos estudos citogenéticos como funcionais para se completar questões não respondidas.
- A população sabe a importância da ciência, mas sabe pouco sobre como a Ciência funciona e sobre a própria Ciência. Três aspectos temporais podem ser levados em conta do porquê divulgar ciência. (1) A curto prazo: é urgente mobilizarmos a população e resgatarmos o investimento da ciência para a retomada do desenvolvimento do país. (2) a médio prazo: a educação científica para que a população por si só saiba reconhecer pseudociências e notícias falsas. (3) a longo prazo: entendendo ciência uma população autônoma pode pensar melhores estratégias de projeto de país. Outro aspecto transcendente é: aproximar a população da ciência é garantir um direito fundamental.

5 Referências

Adelman, K., & Egan, E. (2017). Non-coding RNA: More uses for genomic junk. *Nature*, *543*(7644), 183–185. https://doi.org/10.1038/543183a

Ahmad, S. F., Jehangir, M., Cardoso, A. L., Wolf, I. R., Margarido, V. P., Cabral-De-Mello, D. C., O'Neill, R., Valente, G. T., & Martins, C. (2020). B chromosomes of multiple species have intense evolutionary dynamics and accumulated genes related to important biological processes. *BMC Genomics*, *21*(1), 1–25. https://doi.org/10.1186/s12864-020-07072-1

Ahmad, S., & Martins, C. (2019). The Modern View of B Chromosomes Under the Impact of High Scale Omics Analyses. *Cells*, *8*(2), 156. https://doi.org/10.3390/cells8020156

Akbari, O. S., Antoshechkin, I., Hay, B. A., & Ferree, P. M. (2013). Transcriptome profiling of *Nasonia vitripennis* testis reveals novel transcripts expressed from the selfish B chromosome, paternal sex ratio. *G3: Genes, Genomes, Genetics*, *3*(9), 1597–1605. https://doi.org/10.1534/g3.113.007583

Aldrich, J. C., Leibholz, A., Cheema, M. S., Ausió, J., & Ferree, P. M. (2017). A "selfish" B chromosome induces genome elimination by disrupting the histone code in the jewel wasp *Nasonia vitripennis*. *Scientific Reports*, *7*(February), 1–14. https://doi.org/10.1038/srep42551

André, A., Helene, F., & Fernando, E. G. (2021). Estrangulamento da ciência compromete soberania. *Revista Questão de Ciência*. https://www.revistaquestaodeciencia.com.br/artigo/2021/10/25/estrangulamento-da-ciencia-compromete-soberania

Andrés-León, E., Núñez-Torres, R., & Rojas, A. M. (2016). miARma-Seq: A comprehensive tool for miRNA, mRNA and circRNA analysis. *Scientific Reports*, *6*, 1–8. https://doi.org/10.1038/srep25749

Auger, D. L., & Birchler, J. A. (2002). Maize tertiary trisomic stocks derived from B-A translocations. *J Hered*, *93*(1), 42–47. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12011174

Ayonoadu, U. W., & Rees, H. (1970). The effects of B chromosomes on the nuclear phenotype in root meristems of maize. *Heredity*, *i*, 365–383.

Banaei-moghaddam, A. M., Martis, M. M., Gundlach, H., Himmelbach, A.,

Altschmied, L., Mayer, K. F. X. X., Houben, A., Macas, J., Gundlach, H., Himmelbach, A., Altschmied, L., Mayer, K. F. X. X., & Houben, A. (2015). Genes on B chromosomes : Old questions revisited with new tools. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1849*(1), 64–70. https://doi.org/10.1016/ j.bbagrm.2014.11.007

Banaei-Moghaddam, A. M., Meier, K., Karimi-Ashtiyani, R., & Houben, A. (2013). Formation and Expression of Pseudogenes on the B Chromosome of Rye. *The Plant Cell*, *25*(7), 2536–2544. https://doi.org/10.1105/tpc.113.111856

Banaei-Moghaddam, Ali Mohammad, Meier, K., Karimi-Ashtiyani, R., & Houben, A. (2013). Formation and expression of pseudogenes on the B chromosome of rye. *Plant Cell*, *25*(7), 2536–2544. https://doi.org/10.1105/tpc.113.111856

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, *173*(1), 20–51. https://doi.org/ 10.1016/j.cell.2018.03.006

Benetta, E. D., Akbari, O. S., & Ferree, P. M. (2019). Sequence expression of supernumerary B chromosomes: Function or fluff? *Genes*, *10*(2). https://doi. org/10.3390/genes10020123

Biscotti, M. A., Canapa, A., Forconi, M., Gerdol, M., Pallavicini, A., Schartl, M., & Barucca, M. (2017). The small noncoding RNA processing machinery of two living fossil species, lungfish and coelacanth, gives new insights into the evolution of the argonaute protein family. *Genome Biology and Evolution*, *9*(3), 438–453. https://doi.org/10.1093/gbe/evx017

Bizuayehu, T. T., & Babiak, I. (2014). MicroRNA in teleost fish. *Genome Biology and Evolution*, *6*(8), 1911–1937. https://doi.org/10.1093/gbe/evu151

Boivin, V., Faucher-Giguère, L., Scott, M., & Abou-Elela, S. (2019). The cellular landscape of mid-size noncoding RNA. *Wiley Interdisciplinary Reviews: RNA*, *10*(4), 1–17. https://doi.org/10.1002/wrna.1530

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., … Di Palma, F. (2015a). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, *513*(7518), 375–381. https://doi.org/10.1038/nature13726

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' Civic Online Reasoning: A National Portrait.

Educational Researcher, *50*(8), 505–515. https://doi.org/10.3102/0013189X211017495

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating *loci* as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043

Budak, H., & Zhang, B. (2017). MicroRNAs in model and complex organisms. *Functional and Integrative Genomics*, *17*(2–3), 121–124. https://doi.org/10.1007/s10142-017-0544-1

Bueno, D., Palacios-Gimenez, O. M., & Cabral-de-Mello, D. C. (2013). Chromosomal mapping of repetitive DNAs in the grasshopper *Abracris flavolineata* reveal possible ancestry of the B chromosome and H3 histone spreading. *PLoS ONE*, *8*(6). https://doi.org/10.1371/journal.pone.0066532

Camacho, J. P. M., Sharbel, T. F., & Beukeboom, L. W. (2000). B-chromosome evolution. *Phil.Trans. R. Soc. Lond.*, *355*, 163–178.

Carchilan, M., Delgado, M., Ribeiro, T., Costa-Nunes, P., Caperta, A., Morais-Cecílio, L., Jones, R. N., Viegas, W., Houben, A., & Houbena, A. (2007). Transcriptionally active heterochromatin in rye B chromosomes. *The Plant Cell*, *19*(6), 1738–1749. https://doi.org/10.1105/tpc.106.046946

Carchilan, M., Kumke, K., Mikolajewski, S., & Houben, A. (2009). Rye B chromosomes are weakly transcribed and might alter the transcriptional activity of A chromosome sequences. *Chromosoma*, *118*(5), 607–616. https://doi.org/10. 1007/s00412-009-0222-8

Cardoso, A. L., Fantinatti, B. E. de A., Venturelli, N. B., Carmello, B. de O., Oliveira, R. A. de, & Martins, C. (2019). Epigenetic DNA modifications are correlated with B chromosomes and sex in the Cichlid *Astatotilapia latifasciata*. *Frontiers in Genetics*, *10*(April), 1–14. https://doi.org/10.3389/fgene.2019.00324

Carmello, B. O., Coan, R. L. B., Cardoso, A. L., Ramos, E., Fantinatti, B. E. A., Marques, D. F., Oliveira, R. A., Valente, G. T., & Martins, C. (2017). The hnRNP Q-like gene is retroinserted into the B chromosomes of the cichlid fish *Astatotilapia latifasciata*. *Chromosome Research*, *25*(3–4), 277–290. https://doi.org/10.1007/s10577-017-9561-0

Carvalho, A. B. (2002). Origin and evolution of the Drosophila Y chromosome.

Current Opinion in Genetics and Development, *12*(6), 664–668. https://doi.org/10.1016 /S0959-437X(02)00356-8

Chatr-Aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., … Tyers, M. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Research*, *43*(D1), D470–D478. https://doi.org/10.1093/nar/gku1204

Clark, F. E., Conte, M. A., Ferreira-Bravo, I. A., Poletto, A. B., Martins, C., & Kocher, T. D. (2017). Dynamic sequence evolution of a sex-Associated b chromosome in lake Malawi cichlid fish. *Journal of Heredity*, *108*(1), 53–62. https://doi.org/10. 1093/jhered/esw059

Clark, F. E., Conte, M. A., & Kocher, T. D. (2018). Genomic characterization of a B chromosome in lake Malawi cichlid fishes. *Genes*, *9*(12). https://doi.org/ 10.3390/genes9120610

Clark, F. E., & Kocher, T. D. (2019). Changing sex for selfish gain: B chromosomes of Lake Malawi cichlid fish. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038 /s41598-019-55774-8

Coan, R. L. B., & Martins, C. (2018). Landscape of transposable elements focusing on the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genes*, *9*(6). https://doi.org/10.3390/genes9060269

Costa, F. F. (2007). Non-coding RNAs: Lost in translation? *Gene*, *386*(1–2), 1–10. https://doi.org/10.1016/j.gene.2006.09.028

Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., & Hannon, G. J. (2018). PiRNA-guided genome defense: From biogenesis to silencing. *Annual Review of Genetics*, *52*, 131–157. https://doi.org/10.1146 /annurev-genet-120417-031441

D'Ambrosio, U., Alonso-Lifante, M. P., Barros, K., Kovařík, A., Mas de Xaxars, G., & Garcia, S. (2017). B-chrom: A database on B-chromosomes of plants, animals and fungi. *New Phytologist*, *2007*, 635–642. https://doi.org/10.1111/nph.14723

de Santis, C., Smith-Keune, C., & Jerry, D. R. (2011). Normalizing RT-qPCR data: are we getting the right answers? An appraisal of normalization approaches and internal reference genes from a case study in the finfish *Lates calcarifer. Marine*

Biotechnology, *13*(2), 170–180. https://doi.org/10.1007/s10126-010-9277-z

De Silva, D. M. Z. A., Utsunomia, R., Ruiz-Ruano, F. J., Daniel, S. N., Porto-Foresti, F., Hashimoto, D. T., Oliveira, C., Camacho, J. P. M., & Foresti, F. (2017). High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-12939-7

Escobar, H. (2018). Divulgação científica: faça agora ou cale-se para sempre. *ComCiência e Divulgação Científica*, *Dossiê 197*, 31. https://doi.org/ISSN 1519-7654

Fantinatti, B. E. A. A., & Martins, C. (2016). Development of chromosomal markers based on next-generation sequencing: The B chromosome of the cichlid fish *Astatotilapia latifasciata* as a model. *BMC Genetics*, *17*(1), 1–8. https://doi.org/10.1186/s12863-016-0427-9

Fantinatti, B. E. A. A., Mazzuchelli, J., Martins, C., Valente, G. T., Cabral-de-Mello, D. C., & Martins, C. (2011). Genomic content and new insights on the origin of the B chromosome of the cichlid fish *Astatotilapia latifasciata. Genetica*, *139*(10), 1273–1282. https://doi.org/10.1007/s10709-012-9629-x

Fox, D. P., Hewitt, G. M., & Hall, D. J. (1974). DNA replication and RNA transcription of euchromatic and heterochromatic chromosome regions during grasshopper meiosis. *Chromosoma*, *45*, 43–62.

Franchini, P., Xiong, P., Fruciano, C., Schneider, R. F., Woltering, J. M., Hulsey, C. D., & Meyer, A. (2019). MicroRNA gene regulation in extremely young and parallel adaptive radiations of crater lake cichlid fish. *Molecular Biology and Evolution*, *36*(11), 2498–2511. https://doi.org/10.1093/molbev/msz168

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., & Rajewsky, N. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology*, *26*(4), 407–415. https://doi.org/10.1038/nbt1394

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, *40*(1), 37–52. https://doi.org/10.1093/nar/gkr688

Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K., & Zamore, P. D. (2018). A single

mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. *Molecular Cell*, *71*(5), 775-790.e5. https://doi. org/10.1016/j.molcel.2018.08.007

Gibson, I., & Godfrey. (1972). Interpopulation variation in the satellite DNA from grasshoppers with B-Chromosomes. *Chromosoma*, *138*, 121–138.

Gomes, S. F., Penna, J. C. B. de O., & Arroio, A. (2020). Fake News Científicas: Percepção, Persuasão e Letramento. *Ciência & Educação (Bauru)*, *26*, 1–13. https://doi.org/10.1590/1516-731320200018

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., & Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Reports*, *15*(4), 411–418. https://doi.org/10.1002/embr.201337898

Graphodatsky, A. S., Kukekova, A. V., Yudkin, D. V., Trifonov, V. A., Vorobieva, N. V., Beklemisheva, V. R., Perelman, P. L., Graphodatskaya, D. A., Trut, L. N., Yang, F., Ferguson-Smith, M. A., Acland, G. M., & Aguirre, G. D. (2005). The proto-oncogene C-KIT maps to canid B-chromosomes. *Chromosome Research*, *13*(2), 113–122. https://doi.org/10.1007/s10577-005-7474-9

Green, D. M. (1990). Muller's ratchet and the evolution of supernumerary chromosomes. *Genome*, *33*(6), 818–824. https://doi.org/10.1139/g90-123

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., & Eddy, S. R. (2003). Rfam: An RNA family database. *Nucleic Acids Research*, *31*(1), 439–441. https://doi.org/10.1093/nar/gkg006

Griffiths-Jones, S., Hui, J. H. L., Marco, A., & Ronshaugen, M. (2011). MicroRNA evolution by arm switching. *EMBO Reports*, *12*(2), 172–177. https://doi.org/ 10.1038/embor.2010.191

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degnan, B. M., Rokhsar, D. S., & Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, *455*(7217), 1193–1197. https://doi.org/10.1038/nature07415

Grün, D., Wang, Y. L., Langenberger, D., Gunsalus, K. C., & Rajewsky, N. (2005). MicroRNA target predictions across seven drosophilo species and comparison to mammalian targets. *PLoS Computational Biology*, *1*(1), 0051–0066. https://doi.org/10.1371/journal.pcbi.0010013

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J.,

Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Han, B. W., Wang, W., Li, C., & Weng, Z. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, *348*(6236), 817–822.

Hansson, B. (2019). On the origin and evolution of germline chromosomes in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11570–11572. https://doi.org/10.1073/pnas.1906803116

Hartig, J. V., Tomari, Y., & Förstemann, K. (2007). piRNAs - The ancient hunters of genome invaders. *Genes and Development*, *21*(14), 1707–1713. https://doi.org/10.1101/gad.1567007

Himes, M. (1967). An analysis of heterochromatin in maize root tips. *The Journal of Cell Biology*, *35*(1), 175–181. https://doi.org/10.1083/jcb.35.1.175

Hore, T. A., Rapkins, R. W., & Graves, J. A. M. (2007). Construction and evolution of imprinted loci in mammals. *Trends in Genetics*, *23*(9), 440–448. https://doi.org/10.1016/j.tig.2007.07.003

Houben, A. (2017). B Chromosomes – A Matter of Chromosome Drive. *Frontiers in Plant Science*, *08*(February), 1–6. https://doi.org/10.3389/fpls.2017.00210

Houben, A., Banaei-Moghaddam, A. M., Klemme, S., & Timmis, J. N. (2014). Evolution and biology of supernumerary B chromosomes. *Cellular and Molecular Life Sciences*, *71*(3), 467–478. https://doi.org/10.1007/s00018-013-1437-7

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H. A., Hannon, G. J., Draper, B. W., & Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell*, *129*(1), 69–82. https://doi.org/10.1016/j.cell.2007.03.026

Huang, W., Du, Y., Zhao, X., & Jin, W. (2016). B chromosome contains active genes and impacts the transcription of A chromosomes in maize (*Zea mays* L.). *BMC*

Plant Biology, *16*(1), 1–14. https://doi.org/10.1186/s12870-016-0775-7

Huang, Y. H., Peng, S. F., Lin, Y. P., & Cheng, Y. M. (2020). The maize B chromosome is capable of expressing microRNAs and altering the expression of microRNAs derived from A chromosomes. *Chromosome Research*, *28*(2), 129–138. https://doi.org/10.1007/s10577-019-09620-2

Huang, Y., Peng, S., & Lin, Y. (2019). The maize B chromosome is capable of expressing microRNAs and altering the expression of microRNAs derived from A chromosomes. *Chromosome Res*. https://doi.org/10.1007/s10577-019-09620-2

Imarazene, B., Du, K., Beille, S., Jouanno, E., Feron, R., Pan, Q., Torres-Paz, J., Lopez-Roques, C., Castinel, A., Gil, L., Kuchly, C., Donnadieu, C., Parrinello, H., Journot, L., Cabau, C., Zahm, M., Klopp, C., Pavlica, T., Al-Rikabi, A., … Guiguen, Y. (2021). A supernumerary “B-sex” chromosome drives male sex determination in the Pachón cavefish, *Astyanax mexicanus*. *Current Biology*. https://doi.org/10.1016/j.cub.2021.08.030

Ishizu, H., Siomi, H., & Siomi, M. C. (2012). Biology of Piwi-interacting RNAs: New insights into biogenesis and function inside and outside of germlines. *Genes and Development*, *26*(21), 2361–2373. https://doi.org/10.1101/gad.203786.112

Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-interacting RNA: Its biogenesis and functions. *Annual Review of Biochemistry*, *84*(February), 405–433. https://doi.org/10.1146/annurev-biochem-060614-034258

Jeffares, D. C., Mourier, T., & Penny, D. (2006). The biology of intron gain and loss. *Trends in Genetics*, *22*(1), 16–22. https://doi.org/10.1016/j.tig.2005.10.006

Jehangir, M., Ahmad, S. F., Cardoso, A. L., Ramos, E., Valente, G. T., & Martins, C. (2019). De novo genome assembly of the cichlid fish *Astatotilapia latifasciata* reveals a higher level of genomic polymorphism and genes related to B chromosomes. *Chromosoma*, 81–96. https://doi.org/10.1007/s00412-019-00707-7

Jie, Z., Jun, H., Xiao, X., Fen, S., Yao, P., Lin, P., & Ming, Y. (2019). Novel B - chromosome - specific transcriptionally active sequences are present throughout the maize B chromosome. *Molecular Genetics and Genomics*, 1–13. https://doi.org/10.1007/s00438-019-01623-2

Jones, N. (2017). New species with B chromosomes discovered since 1980. *Nucleus*

(India), *60*(3), 263–281. https://doi.org/10.1007/s13237-017-0215-6

Jones, R. N., González-Sánchez, M., González-García, M., Vega, J. M., & Puertas, M. J. (2008). Chromosomes with a life of their own. *Cytogenetic and Genome Research*, *120*(3–4), 265–280. https://doi.org/10.1159/000121076

Jones, R. Neil, Viegas, W., & Houben, A. (2008). A century of B chromosomes in plants: So what? *Annals of Botany*, *101*(6), 767–775. https://doi.org/10.1093/aob/mcm167

Jones RN. (1991). B-Chromosome drive. *The American Naturalist*, *137*(3), 430–442.

Jornal da USP. (2019). Maioria dos brasileiros mantém visão otimista em relação à ciência e tecnologia, aponta pesquisa. *Jornal Da USP*. https://doi.org/ISSN - 2525-6009

Kabayama, Y., Toh, H., Katanaya, A., Sakurai, T., Chuma, S., Kuramochi-Miyagawa, S., Saga, Y., Nakano, T., & Sasaki, H. (2017). Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic Acids Research*, *45*(9), 5387–5398. https://doi.org/10.1093/nar/gkx027

Karafiátová, M., Bedná, M., Said, M., Jana, Č., Blavet, N., & Bartoš, J. (2020). The B chromosome of Sorghum purpureosericeum reveals the first pieces of its sequence. *Journal Of Experimental Botany*, *72*(5), 1606–1616. https://doi.org/10.1093/jxb/eraa548

Kato, A., Zheng, Y. Z., Auger, D. L., Phelps-Durr, T., Bauer, M. J., Lamb, J. C., & Birchler, J. A. (2005). Minichromosomes derived from the B chromosome of maize. *Cytogenetic and Genome Research*, *109*(1–3), 156–165. https://doi.org/10.1159/000082395

Kaufman, L. S., Chapman, L. J., & Chapman, C. A. (1997). Evolution in fast forward: *Haplochromine* fishes of the Lake Victoria region. *Endeavour*, *21*(1), 23–30. https://doi.org/10.1016/S0160-9327(96)10034-X

Kawaoka, S., Kadota, K., Arai, Y., Suzuki, Y., Fujii, T., Abe, H., Yasukochi, Y., Mita, K., Sugano, S., Shimizu, K., Tomari, Y., Shimada, T., & Katsuma, S. (2011). The silkworm W chromosome is a source of female-enriched piRNAs. *Rna*, *17*(12), 2144–2151. https://doi.org/10.1261/rna.027565.111

Kelleher, E. S., Azevedo, R. B. R., & Zheng, Y. (2018). The evolution of small-RNA-mediated silencing of an invading transposable element. *Genome Biology and Evolution*, *10*(11), 3038–3057. https://doi.org/10.1093/gbe/evy218

Kirk, D., & Jones, R. N. (1970). Nuclear genetic activity in B-chromosome rye, in terms of the quantitative interrelationships between nuclear protein, nuclear RNA and histone. *Chromosoma*, *31*(2), 241–254. https://doi.org/10.1007/BF00285151

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M. G., & Katsuma, S. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. In *Nature* (Vol. 509, Issue 7502, pp. 633–636). https://doi.org/10.1038/nature13315

Kocher, T. D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics*, *5*(4), 288–298. https://doi.org/10.1038/nrg1316

Kofler, R. (2019). Dynamics of transposable element invasions with piRNA clusters. *Molecular Biology and Evolution*, *36*(7), 1457–1472. https://doi.org/10.1093/molbev/msz079

Kofler, R. (2020). PiRNA clusters need a minimum size to control transposable element invasions. *Genome Biology and Evolution*, *12*(5), 736–749. https://doi.org/10.1093/gbe/evaa064

Kofler, R., Senti, K. A., Nolte, V., Tobler, R., & Schlötterer, C. (2018). Molecular dissection of a natural transposable element invasion. *Genome Research*, *28*(6), 824–835. https://doi.org/10.1101/gr.228627.117

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289–1291. https://doi.org/10.1093/bioinformatics/btm091

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). MiRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Lamb, J. C., Riddle, N. C., Cheng, Y. M., Theuri, J., & Birchler, J. A. (2007). Localization and transcription of a retrotransposon-derived element on the maize B chromosome. *Chromosome Research*, *15*(3), 383–398. https://doi.org/

10.1007/s10577-007-1135-0

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Le Thomas, A., Tóth, K. F., & Aravin, A. A. (2014). To be or not to be a piRNA: Genomic origin and processing of piRNAs. *Genome Biology*, *15*(1). https://doi.org/10.1186/gb4154

Leach, C. R., Houben, A., Field, B., Pistrick, K., Demidov, D., & Timmis, J. N. (2005). Molecular evidence for transcription of genes on a B chromosome in Crepis capillaris. *Genetics*, *171*(1), 269–278. https://doi.org/10.1534/genetics.105.043273

Ledda, B., Ottaggio, L., Izzotti, A., Sukkar, S. G., & Miele, M. (2020). Small RNAs in eucaryotes: new clues for amplifying microRNA benefits. *Cell and Bioscience*, *10*(1), 1–13. https://doi.org/10.1186/s13578-019-0370-3

Leitão, A., Carvalho, S., Ben-Hamadou, R., & Gaspar, M. B. (2010). Cytogenetics of *Hediste diversicolor* (Annelida: Polychaeta) and comparative karyological analysis within nereididae. *Aquatic Biology*, *10*(2), 193–200. https://doi.org/10.3354/ab00278

Leitão, A., Chaves, R., Joaquim, S., Matias, D., Ruano, F., & Guedes-Pinto, H. (2008). Supernumerary chromosomes on Southern European populations of the cockle *Cerastoderma edule*: Consequence of environmental pollution? *Estuarine, Coastal and Shelf Science*, *79*(1), 152–156. https://doi.org/10.1016/j.ecss.2008.03.015

Leung, A. K. L., & Sharp, P. A. (2010). MicroRNA Functions in Stress Responses. *Molecular Cell*, *40*(2), 205–215. https://doi.org/10.1016/j.molcel.2010.09.027

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y., Jing, X. A., Aldrich, J. C., Clifford, C., Chen, J., Akbari, O. S., & Ferree, P. M. (2017). Unique sequence organization and small RNA expression of a “selfish” B chromosome. *Chromosoma*, *126*(6), 753–768. https://doi.org/10.1007/s00412-017-0641-x

Lin, H. Z., Lin, W. De, Lin, C. Y., Peng, S. F., & Cheng, Y. M. (2014). Characterization of maize B-chromosome-related transcripts isolated via cDNA-

AFLP. *Chromosoma*, *123*(6), 597–607. https://doi.org/10.1007/s00412-014-0476-7

Lin, M. (1955). Chromosomal control of nuclear composition in maize. *Chromosoma*, *7*(1), 340–370. https://doi.org/10.1007/BF00329732

Lin, S. L., Miller, J. D., & Ying, S. Y. (2006). Intronic microRNA (miRNA). *Journal of Biomedicine and Biotechnology*, *2006*, 1–13. https://doi.org/10.1155/JBB/2006/26818

Loh, Y. H. E., Yi, S. V., & Streelman, J. T. (2011). Evolution of MicroRNAs and the diversification of species. *Genome Biology and Evolution*, *3*(1), 55–65. https://doi.org/10.1093/gbe/evq085

López-León, M. D., Neves, N., Schwarzacher, T., (Pat) Heslop-Harrison, J. S., Hewitt, G. M., & Camacho, J. P. M. (1994). Possible origin of a B chromosome deduced from its DNA composition using double FISH technique. *Chromosome Research*, *2*(2), 87–92. https://doi.org/10.1007/BF01553487

Luo, S., & Lu, J. (2017). silencing of transposable elements by piRNAs in *Drosophila*: an evolutionary perspective. *Genomics, Proteomics and Bioinformatics*, *15*(3), 164–176. https://doi.org/10.1016/j.gpb.2017.01.006

Luo, S., Zhang, H., Duan, Y., Yao, X., Clark, A. G., & Lu, J. (2020). The evolutionary arms race between transposable elements and piRNAs in Drosophila melanogaster. *BMC Evolutionary Biology*, *20*(1), 1–18. https://doi.org/10.1186/s12862-020-1580-3

Ma, W., Gabriel, T. S., Martis, M. M., Gursinsky, T., Schubert, V., Vrána, J., Doležel, J., Grundlach, H., Altschmied, L., Scholz, U., Himmelbach, A., Behrens, S. E., Banaei-Moghaddam, A. M., & Houben, A. (2017). Rye B chromosomes encode a functional Argonaute-like protein with in vitro slicer activities similar to its A chromosome paralog. *New Phytologist*, *213*(2), 916–928. https://doi.org/10.1111/nph.14110

Ma, W., Liu, Z. J., Beier, S., Houben, A., & Carpentier, S. (2021). Identification of rye B chromosome-associated peptides by mass spectrometry. *New Phytologist*, *230*(6), 2179–2185. https://doi.org/10.1111/nph.17238

Madden, T. (2013). The BLAST sequence analysis tool. *The BLAST Sequence Analysis Tool*, *Md*, 1–17. http://www.ncbi.nlm.nih.gov/books/NBK153387/

Mahajan, S., & Bachtrog, D. (2017). Convergent evolution of y chromosome gene

content in flies. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00653-x

Malone, C. D., & Hannon, G. J. (2009). Small RNAs as Guardians of the Genome. *Cell*, *136*(4), 656–668. https://doi.org/10.1016/j.cell.2009.01.045

Martins, C., & Jehangir, M. (2021). A genomic glimpse of B chromosomes in cichlids. *Genes and Genomics*, *43*(3), 199–208. https://doi.org/10.1007/s13258-021-01049-4

Martis, M. M., Klemme, S., Banaei-Moghaddam, A. M., Blattner, F. R., Macas, J., Schmutzer, T., Scholz, U., Gundlach, H., Wicker, T., Simkova, H., Novak, P., Neumann, P., Kubalakova, M., Bauer, E., Haseneyer, G., Fuchs, J., Dolezel, J., Stein, N., Mayer, K. F. X., & Houben, A. (2012). Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. *Proceedings of the National Academy of Sciences*, *109*(33), 13343–13346. https://doi.org/10.1073/pnas.1204237109

Massarani, L., Monteiro, E., & Dias, D. S. (2018). *José Reis: reflexões sobre a divulgação científica* (Fiocruz/COC (ed.)). http://labds.eci.ufmg.br:8080/bitstream/123456789/37/1/massarani_luisa%25dias_eliane_monteiro_santana%25reflexoes_divulgacao_cientifica_press%25livro%252018%2513082021.pdf

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15 Spec No*(1), 17–29. https://doi.org/10.1093/hmg/ddl046

Mazzoni, D., Andrade, Z. De, Ruiz-ruano, F. J., Utsunomia, R., Martín-peciña, M., Castro, J. P., Freire, P. P., Carvalho, R. F., Hashimoto, D. T., Suh, A., Oliveira, C., Porto-foresti, F., Artoni, R. F., Foresti, F., & Camacho, J. P. M. (2021). *Long-term persistence of supernumerary B chromosomes in multiple species of Astyanax fish*. 1–17.

Mazzuchelli, J., Kocher, T. D., Yang, F., & Martins, C. (2012). Integrating cytogenetics and genomics in comparative evolutionary studies of cichlid fish. *BMC Genomics*, *13*(1). https://doi.org/10.1186/1471-2164-13-463

Miao, V. P., Covert, S. F., & Vanetten, H. D. (1991). A fungal gene for antibiotic resistance on a dispensable (“B”) chromosome. *Science*, *254*(5039), 1773–1776. https://doi.org/10.1126/science.1763326

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome

defence. *Nature*, *457*(7228), 413–420. https://doi.org/10.1038/nature07756

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371. https://doi.org/10.1093/molbev/msv035

Nakajima, R. T. (2019). *Análises em larga escala de proteínas e construção de redes biológicas com foco em estudos de cromossomos B*. 1–20.

Nascimento-Oliveira, J. I. (2020). Conectando telômeros e o envelhecimento: é possível viver para sempre ? *Genética Na Escola*, *15*, 108–117. https://www.geneticanaescola.com/_files/ugd/b703be_1160d23fd79047c094f16bbd837494e3.pdf

Nascimento-Oliveira, J. I., Evaristo, B., Fantinatti, A., Wolf, I. R., Cardoso, A. L., Ramos, E., Rieder, N., Oliveira, R. De, & Martins, C. (2021). Differential expression of miRNAs in the presence of B chromosome in the cichlid fish *Astatotilapia latifasciata*. *BMC Genomics*, 1–16. https://doi.org/10.1186/s12864-021-07651-w

Navarro-Domínguez, B., Martín-Peciña, M., Ruiz-Ruano, F. J., Cabrero, J., Corral, J. M., López-León, M. D., Sharbel, T. F., & Camacho, J. P. M. (2019). Gene expression changes elicited by a parasitic B chromosome in the grasshopper *Eyprepocnemis plorans* are consistent with its phenotypic effects. *Chromosoma*, *128*(1), 53–67. https://doi.org/10.1007/s00412-018-00689-y

Navarro-Domínguez, B., Ruiz-Ruano, F. J., Cabrero, J., Corral, J. M., López-León, M. D., Sharbel, T. F., & Camacho, J. P. M. (2017). Protein-coding genes in B chromosomes of the grasshopper *Eyprepocnemis plorans. Scientific Reports*, *7*(September 2016), 1–12. https://doi.org/10.1038/srep45200

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nur, U., Werren, J. H., Eickbush, D. G., Burke, W. D., & Eickbush, T. H. (1988). A

"selfish" B chromosome that enhances its transmission by eliminating the paternal genome. *Science*, *240*(4851), 512–514. https://doi.org/10.1126/science.3358129

Oliveira, A. C., Bovolenta, L. A., Nachtigall, P. G., Herkenhoff, M. E., Lemke, N., & Pinhal, D. (2017). Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Frontiers in Genetics*, *8*(MAY), 1–10. https://doi.org/10.3389/fgene.2017.00059

Orsi, C. (2021). Ciência como direito humano. *Revista Questão de Ciência*.

Palazzo, A. F., & Lee, E. S. (2015). Non-coding RNA: What is functional and what is junk? *Frontiers in Genetics*, *5*(JAN), 1–11. https://doi.org/10.3389/fgene.2015.00002

Pelisson, A., Sun Song, U., Prud'homme, N., Smith, P. A., Bucheton, A., & Corces, V. G. (1994). *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO Journal*, *13*(18), 4401–4411. https://doi.org/10.1002/j.1460-2075.1994.tb06760.x

Peng, S., & Cheng, Y. M. (2011). Characterization of satellite CentC repeats from heterochromatic regions on the long arm of maize B-chromosome. *Chromosome Research*, *19*, 183–191. https://doi.org/10.1111/nph.14723

Phanstiel, D. H., Boyle, A. P., Araya, C. L., & Snyder, M. P. (2014). Sushi.R: Flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, *30*(19), 2808–2810. https://doi.org/10.1093/bioinformatics/btu379

Piovesan, F. (2014). Declaração Universal de Direitos Humanos: desafios e perspectivas. *Revista Brasileira de Estudos Jurídicos*, *9*(Iii), 31–56.

Ploskaya-Chaibi, M., Voitovich, A. M., Novitsky, R. V., & Bouhadad, R. (2015). B-chromosome and V-shaped spot asymmetry in the common frog (*Rana temporaria* L.) populations. *Comptes Rendus - Biologies*, *338*(3), 161–168. https://doi.org/10.1016/j.crvi.2014.12.005

Poletto, A. B., Ferreira, I. A., Cabral-de-Mello, D. C., Nakajima, R. T., Mazzuchelli, J., Ribeiro, H. B., Venere, P. C., Nirchio, M., Kocher, T. D., & Martins, C. (2012). Chromosome differentiation patterns during cichlid fish evolution. *BMC Genetics*, *13*. https://doi.org/10.1186/1471-2156-13-2

Poletto, A. B., Ferreira, I. A., & Martins, C. (2010). The B chromosomes of the African cichlid fish *Haplochromis obliquidens* harbour 18S rRNA gene copies. *BMC Genetics*, *11*(1), 1–8. https://doi.org/10.1186/1471-2156-11-1

Ponnusamy, M., Yan, K. W., Liu, C. Y., Li, P. F., & Wang, K. (2017). PIWI family emerging as a decisive factor of cell fate: An overview. *European Journal of Cell Biology*, *96*(8), 746–757. https://doi.org/10.1016/j.ejcb.2017.09.004

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093 /bioinformatics/btq033

Rajewsky, N. (2006). Microrna target predictions in animals. *Nature Genetics*, *38*(6S), S8–S13. https://doi.org/10.1038/ng1798

Rajičić, M., Adnadević, T., Stamenković, G., Blagojević, J., & Vujošević, M. (2015). Screening of B chromosomes for presence of two genes in yellow-necked mice, *Apodemus flavicollis* (Mammalia, Rodentia). *Genetika*, *47*(1), 311–321. https://doi.org/10.2298/GENSR1501311R

Ramat, A., & Simonelig, M. (2021). Functions of PIWI proteins in gene regulation: new arrows added to the piRNA quiver. *Trends in Genetics*, *37*(2), 188–200. https://doi.org/10.1016/j.tig.2020.08.011

Ramos, É., Cardoso, A. L., Brown, J., Marques, D. F., Fantinatti, B. E. A., Cabral-de-Mello, D. C., Oliveira, R. A., O'Neill, R. J., & Martins, C. (2017). The repetitive DNA element BncDNA, enriched in the B chromosome of the cichlid fish *Astatotilapia latifasciata*, transcribes a potentially noncoding RNA. *Chromosoma*, *126*(2), 313–323. https://doi.org/10.1007/s00412-016-0601-x

Randolph, L. F. (1941). Genetic Characteristics of the B Chromosomes in Maize. *Genetics*, *26*(6), 608–631. https://doi.org/10.1093/genetics/26.6.608

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Rodrigues, R. (2020). Pesquisa : os perigos ocultos da Internet. *Kaspersky Blog*. https://www.kaspersky.com.br/blog/digital-iceberg/13996/

Rosa, G. M. (2020). Brasil forma mestres e doutores para o desemprego. *Primeira Pauta*. https://primeirapauta.ielusc.br/2020/12/10/brasil-forma-mestres-e-

doutores-para-o-desemprego/

Rosenkranz, D., Han, C. T., Roovers, E. F., Zischler, H., & Ketting, R. F. (2015). Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence. *Genomics Data*, *5*, 309–313. https://doi.org/10.1016/j.gdata.2015.06.026

Rosenkranz, D., & Zischler, H. (2012). proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, *13*(1). https://doi.org/10.1186/1471-2105-13-5

Ruban, A., Schmutzer, T., Scholz, U., & Houben, A. (2017). How next-generation sequencing has aided our understanding of the sequence composition and origin of B chromosomes. *Genes*, *8*(11), 1–14. https://doi.org/10.3390/genes8110294

Ruiz-Estévez, M., Badisco, L., Broeck, J. Vanden, Perfectti, F., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2014). B chromosomes showing active ribosomal RNA genes contribute insignificant amounts of rRNA in the grasshopper *Eyprepocnemis plorans*. *Molecular Genetics and Genomics*, *289*(6), 1209–1216. https://doi.org/10.1007/s00438-014-0880-y

Ruiz-Estévez, M., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2012). B-chromosome ribosomal DNA is functional in the grasshopper *Eyprepocnemis plorans*. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0036600

Ruiz-Ruano, F. J., Cabrero, J., López-León, M. D., & Camacho, J. P. M. (2017). Satellite DNA content illuminates the ancestry of a supernumerary (B) chromosome. *Chromosoma*, *126*(4), 487–500. https://doi.org/10.1007/s00412-016-0611-8

Sambrook, J., & Russell, D. W. (2006). Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harbor Protocols*, *2006*(1), pdb.prot4455. https://doi.org/10.1101/pdb.prot4455

Santos, W. L. P. dos. (2007). Educação científica na perspectiva de letramento como prática social: funções, princípios e desafios. *Revista Brasileira de Educação*, *12*(36), 474–492. https://doi.org/10.1590/s1413-24782007000300007

Schemberger, M. O., Nascimento, V. D., Coan, R., Ramos, É., Nogaroto, V., Ziemniczak, K., Valente, G. T., Moreira-Filho, O., Martins, C., & Vicari, M. R.

(2019). DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma*, *128*(4), 547–560. https://doi.org/10.1007/s00412-019-00721-9

Shao, F., Han, M., & Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-51888-1

Simon, P. (2003). Q-Gene: Processing quantitative real-time RT-PCR data. *Bioinformatics*, *19*(11), 1439–1440. https://doi.org/10.1093/bioinformatics/btg157

Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: The vanguard of genome defence. *Nature Reviews Molecular Cell Biology*, *12*(4), 246–258. https://doi.org/10.1038/nrm3089

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Song, H., Xing, C., Lu, W., Liu, Z., Wang, X., Cheng, J., & Zhang, Q. (2019). Rapid evolution of piRNA pathway and its transposon targets in Japanese flounder (*Paralichthys olivaceus*). *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *31*(July), 100609. https://doi.org/10.1016/ j.cbd.2019.100609

Špakulová, M., & Casanova, J. C. (2004). Current knowledge on B chromosomes in natural populations of helminth parasites: A review. *Cytogenetic and Genome Research*, *106*(2–4), 222–229. https://doi.org/10.1159/000079291

Tanić, N., Vujošević, M., Dedović-Tanić, N., & Dimitrijević, B. (2005). Differential gene expression in yellow-necked mice *Apodemus flavicollis* (Rodentia, Mammalia) with and without B chromosomes. *Chromosoma*, *113*(8), 418–427. https://doi.org/ 10.1007/s00412-004-0327-z

Tao, W., Sun, L., Chen, J., Shi, H., & Wang, D. (2016). Genomic identification, rapid evolution, and expression of Argonaute genes in the tilapia, *Oreochromis niloticus*. *Development Genes and Evolution*, *226*(5), 339–348. https://doi.org/ 10.1007/s00427-016-0554-3

Thomson, D. W., Pillman, K. A., Anderson, M. L., Lawrence, D. M., Toubia, J., Goodall, G. J., & Bracken, C. P. (2015). Assessing the gene regulatory

properties of Argonaute-bound small RNAs of diverse genomic origin. *Nucleic Acids Research*, *43*(1), 470–481. https://doi.org/10.1093/nar/gku1242

Torgasheva, A. A., Malinovskaya, L. P., Zadesenets, K. S., Karamysheva, T. V., Kizilova, E. A., Akberdina, E. A., Pristyazhnyuk, I. E., Shnaider, E. P., Volodkina, V. A., Saifitdinova, A. F., Galkina, S. A., Larkin, D. M., Rubtsov, N. B., & Borodin, P. M. (2019). Germline-restricted chromosome (GRC) is widespread among songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11845–11850. https://doi.org/10.1073/pnas.1817373116

Trifonov, V. A., Dementyeva, P. V., Larkin, D. M., O'Brien, P. C. M., Perelman, P. L., Yang, F., Ferguson-Smith, M. A., & Graphodatsky, A. S. (2013). Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer (*Capreolus pygargus*). *BMC Biology*, *11*(1), 1. https://doi.org/10.1186/1741-7007-11-90

Turner, G. F. (2007). Adaptive radiation of cichlid fish. *Current Biology*, *17*(19), 827–831. https://doi.org/10.1016/j.cub.2007.07.026

Ulitsky, I., Shkumatava, A., Jan, C. H., Subtelny, A. O., Koppstein, D., Bell, G. W., Sive, H., & Bartel, D. P. (2012). Extensive alternative polyadenylation during zebrafish development. *Genome Research*, *22*(10), 2054–2066. https://doi.org/10.1101/gr.139733.112

Valente, G. T., Conte, M. A., Fantinatti, B. E. A., Cabral-De-Mello, D. C., Carvalho, R. F., Vicari, M. R., Kocher, T. D., & Martins, C. (2014). Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Molecular Biology and Evolution*, *31*(8), 2061–2072. https://doi.org/10.1093/molbev/msu148

Valente, G. T., Nakajima, R. T., Fantinatti, B. E. A., Marques, D. F., Almeida, R. O., Simões, R. P., & Martins, C. (2017). B chromosomes: from cytogenetics to systems biology. *Chromosoma*, *126*(1), 73–81. https://doi.org/10.1007/s00412-016-0613-6

Wang, H., Wang, B., Liu, J., Li, A., Zhu, H., Wang, X. B., & Zhang, Q. (2018). Piwil1 gene is regulated by hypothalamic-pituitary-gonadal axis in turbot (Scophthalmus maximus): A different effect in ovaries and testes. *Gene*. https://doi.org/ 10.1016/j.gene.2018.03.016

Wang, J., Zhang, P., Lu, Y., Li, Y., Zheng, Y., Kan, Y., Chen, R., & He, S. (2019).

PiRBase: A comprehensive database of piRNA sequences. *Nucleic Acids Research*, *47*(D1), D175–D180. https://doi.org/10.1093/nar/gky1043

Wierzbicki, F., Kofler, R., & Signor, S. (2021). Evolutionary dynamics of piRNA clusters in *Drosophila*. *BioRxiv*. https://doi.org/10.1101/2021.08.20.457083

Wilson, E. B. (1907). Note on the chromosome-groups of *Metapodius* and *Banasa*. *Biological Bulletin*, *XII*(5), 303–313.

Xiong, P., Schneider, R. F., Hulsey, C. D., Meyer, A., & Franchini, P. (2019). Conservation and novelty in the microRNA genomic landscape of hyperdiverse cichlid fishes. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-50124-0

Yi, M., Chen, F., Luo, M., Cheng, Y., Zhao, H., Cheng, H., & Zhou, R. (2014). Rapid evolution of piRNA pathway in the Teleost fish: Implication for an adaptation to transposon diversity. *Genome Biology and Evolution*, *6*(6), 1393–1407. https://doi.org/10.1093/gbe/evu105

Yoshida, K., Terai, Y., Mizoiri, S., Aibara, M., Nishihara, H., Watanabe, M., Kuroiwa, A., Hirai, H., Hirai, Y., Matsuda, Y., & Okada, N. (2011). B chromosomes have a functional effect on female sex determination in lake victoria cichlid fishes. *PLoS Genetics*, *7*(8). https://doi.org/10.1371/journal.pgen.1002203

Zeng, Q., & Tang, T. (2020). Role of PIWI-interacting RNAs on cell survival : Proliferation, apoptosis, and cycle. *IUBMB Life*, *April*, 1870–1878. https://doi.org/10.1002/iub.2332

Zhang, P., Si, X., Skogerbø, G., Wang, J., Cui, D., Li, Y., Sun, X., Liu, L., Sun, B., Chen, R., He, S., & Huang, D. W. (2014). PiRBase: A Web resource assisting piRNA functional study. *Database*, *2014*, 1–7. https://doi.org/10.1093/database/bau110

Zhou, Y., Wang, F., Liu, S., Zhong, H., Liu, Z., Tao, M., Zhang, C., & Liu, Y. (2012). Human chorionic gonadotropin suppresses expression of Piwis in common carp (*Cyprinus carpio*) ovaries. *General and Comparative Endocrinology*. https://doi.org/10.1016/j.ygcen.2011.11.044

Zhou, Y., Zhong, H., Liu, S., Yu, F., Hu, J., Zhang, C., Tao, M., & Liu, Y. (2014). Elevated expression of Piwi and piRNAs in ovaries of triploid *Crucian carp. Molecular and Cellular Endocrinology*. https://doi.org/10.1016/j.mce.2013.11.019