

Universidade Estadual Paulista “Júlio de Mesquita Filho”

Instituto de Biociências de Botucatu

Departamento de Morfologia

Análises em larga escala de proteínas e construção de redes
biológicas com foco em estudos de cromossomos B

Rafael Takahiro Nakajima

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

Botucatu/SP

Maio de 2019

Instituto de Biociências - Seção Técnica de Pós-Graduação

Distrito de Rubião Júnior s/n CEP 18618-970 Cx Postal 510 Botucatu-SP Brasil

Tel (14) 3880-0780 posgraduacao@ibb.unesp.br

Universidade Estadual Paulista “Júlio de Mesquita Filho”

Instituto de Biociências de Botucatu

Departamento de Morfologia

Análises em larga escala de proteínas e construção de redes
biológicas com foco em estudos de cromossomos B

Rafael Takahiro Nakajima

Orientador: Prof. Dr. Cesar Martins

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

Botucatu/SP

Maio de 2019

Instituto de Biociências - Seção Técnica de Pós-Graduação

Distrito de Rubião Júnior s/n CEP 18618-970 Cx Postal 510 Botucatu-SP Brasil

Tel (14) 3880-0780 posgraduacao@ibb.unesp.br

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

Nakajima, Rafael Takahiro.

Análises em larga escala de proteínas e construção de redes biológicas com foco em estudos de cromossomos B / Rafael Takahiro Nakajima. - Botucatu, 2019

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

Orientador: Cesar Martins

Capes: 20204000

1. Proteômica. 2. Cromossomos. 3. RNA mensageiro. 4. Sequenciamento de nucleotídeos em larga escala. 5. Ciclídeos.

Palavras-chave: RNA codificante; RNA mensageiro; ciclídeo africano; cromossomo extra; proteômica.

Agradecimentos

Meus sinceros agradecimentos:

- Ao meu orientador Prof. Dr. Cesar Martins pela oportunidade, confiança, incentivo e aprendizagem ao longo desses quatro anos de trabalho.
- Aos membros do Laboratório de Genômica Integrativa.
- Aos membros do Laboratório SBGL e CEVAP.
- Ao José Eduardo por sempre nos ajudar e pelo amigo que se tornou.
- Aos meus pais por me apoiarem nas minhas decisões, sempre estando ao meu lado em momentos bons e ruins. São exemplos que desejo seguir.
- À Talita Aleixo, por me ajudar e me aguentar nos maus e bons momentos, sempre me apoiando e com um bom humor.
- À nova família que ganhei que sempre acreditaram em nosso potencial e capacidade.
- Ao meus amigos: Akira, Chokito, Érica, Gigio, Costinha, Jones, Patty e Rancheira pelo apoio e descontrações.

Apoio financeiro

- À Fundação de Amparo à Pesquisa do Estado de São Paulo – processo 2015/11485-0 (bolsa), 2013/04533-3, 2015/016661-1 e 2018/24610-6
- Ao CAPES (bolsa – 1 ano) e CNPq 474684/2013-0

Resumo

Os cromossomos B ocorrem em cerca de 2.828 espécies de diferentes táxons, sendo basicamente heterocromáticos e compostos de DNAs repetitivos. Recentemente, análises genômicas em larga escala estão sendo utilizadas para elucidar questões acerca dos cromossomos supranumerários. Os peixes ciclídeos recebem grande interesse científico, uma vez que muitas espécies passaram por um rápido e extenso processo de radiação adaptativa. Em algumas espécies do grupo, como *Astatotilapia latifasciata*, foi descrita a presença de cromossomos B. Neste trabalho foi caracterizado o perfil de expressão proteico em tecidos específicos na *A. latifasciata* e realizada análise funcional da presença do cromossomo B nesta espécie de teleósteo, elucidando a influência que este pode acarretar em vias metabólicas específicas. Além disso, esses dados foram integrados com os resultados de RNA-Seq dessa espécie, e construídas sub redes de co-expressão e interação proteína-proteína. Também foi calculada a entropia de *Shannon*, a qual não apresentou diversidade na expressão dos transcritos em cada biblioteca comparada. Além disso, foi analisada a expressão diferencial de RNAm em cada tecido em relação a presença do cromossomo B e ao sexo. Dentre os transcritos diferencialmente expressos, a análise de enriquecimento funcional apresentou processos relacionados ao ciclo celular, resposta imune e resposta ao estresse. Na maioria dos casos analisados entre a expressão de proteínas, os transcritos *up regulated* e as sub redes, apareceram os genes Aurora quinase, tubulina e proteína de cinetócoro que estão relacionados ao controle do ciclo celular. Os resultados obtidos, de forma geral, não apresentaram grande alteração no perfil proteico ou gênico relacionado a presença do cromossomo B, apesar de identificar proteínas e genes putativos para as vias do GnRH (Gonadotropin-Releasing Hormone), do ciclo celular e do estresse. Contudo, essa alteração no perfil de expressão sugere regulação mais sutil e pontual. Portanto, o cromossomo B pode atuar na regulação de processos vitais para sobrevivência da célula e facilitar sua manutenção e transmissão para as próximas gerações.

Palavras-chave: proteômica, cromossomo extra, RNA mensageiro, RNA codificante, ciclídeo africano

Abstract

The B chromosomes occur in about 2,828 species of different taxa, being heterochromatic and composed of repetitive DNA. Recently, large-scale genomic analyzes are being used to elucidate questions about supernumerary chromosomes. Cichlid fish are of great scientific interest, since many species have gone through a rapid and extensive process of adaptive radiation. In some species of the group, such as *Astatotilapia latifasciata*, the presence of B chromosomes was described. In this work, we characterize the profile of protein expression in specific tissues in *A. latifasciata* and performed a functional analysis of the presence of the B chromosome in this species of teleost, elucidating the influence that it can cause in specific metabolic pathways. In addition, we integrate these data with the RNA-Seq results of this species, and construct sub-networks of co-expression and protein-protein interaction. The Shannon entropy was also calculated, which did not show diversity in the expression of the transcripts in each library. In addition, differential expression analysis was performed on each tissue separately and the relationship between the presence of B chromosome and sex chromosome was analyzed. Among the differentially expressed transcripts, functional enrichment analysis presented processes related to the cell cycle, immune response and stress response. In relation to abundance of proteins, the up regulated transcripts and the sub-networks we identified genes like the Aurora kinase, tubulin and kinetochore protein genes that are related to cell cycle control. The results obtained, in general, did not show a great change in the protein or gene profile related to the presence of the B chromosome, although it identified putative proteins and genes for the GnRH (Gonadotropin-Releasing Hormone), cell cycle and stress pathways. However, this change in the expression profile suggests subtler and punctual regulation. Therefore, the B chromosome can act in the regulation of vital processes for the survival of the cell and facilitate its maintenance and transmission for the next generations.

Keywords: proteomic, extra chromosome, messenger RNA, coding RNA, African cichlids

Lista de Ilustrações

Figura 1. Hipótese da origem do cromossomo B.	17
Figura 2. Ciclídeo africano <i>Astatotilapia latifasciata</i>	18
Figura 3. Cariótipo do ciclídeo <i>Astatotilapia latifasciata</i>	19
Figura 4. Exemplo de grafos direcionados, não direcionados e com peso.....	24
Figura 5. Representação ilustrativa do aquário em que foram mantidos os peixes para a extração proteica.	27
Figura 6. Corte histológico das gônadas femininas e masculinas.	37
Figura 7. Análise quantitativa da extração proteica.....	38
Figura 8. Mapa do ciclo celular.	40
Figura 9. Mapa biológico da sinalização do GnHR.....	41
Figura 10. Mapa biológico da fosforilação oxidativa.	42
Figura 11. Mapa biológico da sinalização MAPK.....	44
Figura 12. Mapa biológico da sinalização do GnHR.....	45
Figura 13. Mapa biológico do transporte de RNA.	46
Figura 14. Análise de entropia dos transcritos.	47
Figura 15. Análise das bibliotecas de encéfalo de macho.....	49
Figura 16. Análise das bibliotecas de encéfalo de fêmea.....	51
Figura 17. Análise das bibliotecas de músculo de macho.....	53
Figura 18. Análise das bibliotecas de músculo de fêmea.....	55
Figura 19. Análise das bibliotecas de gônada de macho.....	56
Figura 20. Análise de cobertura genômica.	58
Figura 21. Análise das bibliotecas de gônada de fêmea.	59
Figura 22. Representação gráfica da análise das sub redes up regulated extraídas da rede de união Co-PPI-net que apresentam genes envolvidos no ciclo celular.....	71

Lista de Tabelas

Tabela 1. Genes envolvidos com ciclo celular e presentes no cromossomo B.	20
Tabela 2. Comparações realizadas para abundância das proteínas.....	30
Tabela 3. Mapas biológicos com a presença de proteínas com abundância alterada.....	39

Tabela 4. Número de acesso das enzimas encontradas para amostras de encéfalo de fêmeas B+.....	41
Tabela 5. Mapas biológicos com a presença de proteínas com abundância alterada.....	43
Tabela 6. Mapas biológicos com a presença de proteínas com abundância alterada.....	45
Tabela 7. Transcritos diferencialmente expresso em gônadas de fêmea.....	60
Tabela 8. Fisher's exact test dos genes em comum up regulated entre amostras de fêmea e macho de encéfalo.....	60
Tabela 9. Fisher's exact test dos genes em comum down regulated entre amostras de fêmea e macho de encéfalo.	61
Tabela 10. Fisher's exact test dos genes em comum up regulated entre amostras de fêmea e macho de músculo.....	62
Tabela 11. Termos ontológicos idênticos encontrados no transcriptoma e proteoma. ..	63
Tabela 12. Termos ontológicos idênticos encontrados no transcriptoma e proteoma. ..	64
Tabela 13. Termos ontológicos idênticos encontrados no transcriptoma e proteoma. ..	65
Tabela 14. Análise das sub redes.	69
Tabela 15. Genes relacionados ao ciclo celular entre os 10 maiores valores para degree e betweenness centrality.	70

Lista de abreviaturas e siglas

PCA	Análise da componente principal
B+	Indivíduos com cromossomo B
B-	Indivíduos sem cromossomo B
DNA	Ácido desoxirribonucleico
hnRNP Q-like	<i>Heterogeneous nuclear ribonucleoprotein Q-like</i>
TUBB1	<i>Tubulin beta 1</i>
KIF11	<i>Kinesin-like protein KIF11</i>
miRNAs	MicroRNA

RNA	Ácido ribonucleico
RNAm	RNA mensageiro
RNA-Seq	Sequenciamento de ácido ribonucleico
CDKs	<i>Cyclin-dependent kinases</i>
MPS	<i>Monopolar spindle 1 kinase</i>
SAC	<i>Spindle assembly checkpoint</i>
APC/C	<i>Anaphase-promoting complex</i>
NDC80	<i>Kinetochore protein NDC80 homolog</i>
Nuf2	<i>Kinetochore complex component</i>
Spc24 e Spc25	<i>NDC80 kinetochore complex component</i>
PCR	Reação em cadeia da polimerase
ACN	Acetonitrila
DDT	<i>Dithiothreitol</i>
EDTA	Ácido etilenodiamino tetra-acético
BSA	Proteica de albumina bovina
rDNA	DNA ribossomal
tRNA	RNA transportador
SAMP	<i>Single-Assembler Multiple-Parameter</i>
PPI	Interação proteína-proteína
PPI-net	Rede biológica de interação proteína-proteína
ORFs	<i>openreadingframes</i>

Sumário

1	Introdução	15
1.1	Cromossomos B	15
1.2	Origem do cromossomo B	16
1.3	<i>Astatotilapia latifasciata</i> (Perciformes, Cichlidae) como modelo para análises de cromossomos B	17
1.4	Análises genômicas em larga escala de cromossomos B em <i>Astatotilapia latifasciata</i>	19
1.5	Análises funcionais: Proteômica	21
1.6	Integração de dados	22
1.6.1	Redes biológicas	22
1.7	Ciclo celular	24
2	Hipótese	25
3	Proposta	26
4	Objetivos	26
4.1	Objetivo Geral	26
4.2	Objetivos Específicos	26
5	Material e Métodos	26
5.1	Animais, extração de DNA e genotipagem de animais B+ e B-	26
5.2	Estudo morfológico das gônadas	27
5.3	Extração de proteínas	28
5.4	Quantificação, digestão e espectrometria de massa	28
5.5	Análise dos dados - Proteômica	29
5.6	Reanálise do transcriptoma (RNAm)	30
5.6.1	Filtragem das <i>reads</i>	31
5.6.2	Montagem <i>de novo</i> e filtragem	31

5.6.3	Entropia de <i>Shannon</i>	33
5.6.4	Expressão diferencial de RNAm	34
5.6.5	Análise comparativa	34
5.7	Construção das redes.....	34
6	Resultados.....	36
6.1	Análise histológica.....	36
6.2	Proteômica	38
6.3	Abundancia proteica e mapas biológicos	38
6.3.1	Encéfalo	39
6.3.2	Músculo.....	43
6.3.3	Gônada	44
6.4	Transcriptoma	46
6.4.1	Entropia de <i>Shannon</i>	47
6.4.2	Expressão diferencial.....	48
6.5	Análise comparativa: Transcriptoma	60
6.5.1	Relação macho e fêmea	60
6.6	Análise comparativa: Transcriptoma e Proteoma	62
6.6.1	Encéfalo	62
6.6.2	Músculo.....	63
6.6.3	Gônada	64
6.7	Redes Biológicas.....	65
7	Discussão	71
7.1	Mapas biológicos: Proteoma.....	71
7.2	Transcriptoma	73
7.3	Análise comparativa	74
7.4	Redes biológicas	75
8	Conclusão	75

9	Materiais Suplementares.....	76
10	Referências	76

1 INTRODUÇÃO

1.1 Cromossomos B

Cromossomos B equivalem a uma classe de elementos supranumerários, descritos como os primeiros “elementos egoístas” (Wilson, 1907) e definidos por Jones e Rees (1982) como cromossomos não homólogos aos cromossomos normais do complemento (denominados de A), os quais não apresentam capacidade de recombinação com os mesmos. Inicialmente foram considerados elementos dispensáveis, sendo classificados como não funcionais por não conterem genes essenciais e, conseqüentemente, por apresentar seu próprio caminho evolutivo (Beukeboom, 1994).

Aproximadamente, 2.828 espécies de táxons diferentes apresentam cromossomos B (D’Ambrosi *et al.*, 2017), os quais estão presentes em apenas uma parcela da população, podendo variar em tamanho e número e alterar o cariótipo do hospedeiro (*Picea glauca* número diploide $2n = 24$ com indivíduos podendo apresentar de 0 a 6 cromossomos Bs ($2n = 24 \text{ As} + 0 - 6 \text{ Bs}$); *Vulpes vulpes* $2n = 43 \text{ As} + 0 - 8 \text{ Bs}$; *Rattus rattus* $2n = 42 \text{ As} + 0 - 5 \text{ Bs}$) (Miao *et al.*, 1991; Carlson, 2009; Jones *et al.*, 2008; Ploskaya-Chaibi *et al.*, 2015). Os cromossomos B, em relação à sua composição genômica, em geral são heterocromáticos, compostos basicamente por elementos repetitivos de DNA (Loreto *et al.*, 2008; Fantinatti *et al.*, 2011; López-León *et al.*, 2008; Teruel *et al.*, 2010; Nur *et al.*, 1988; Jamilena *et al.*, 1994; Gutknecht *et al.*, 1995; Puertas, 2002), seqüências de DNA de organela (Martis *et al.*, 2012), pseudogenes (Banaei-Moghaddam *et al.*, 2014) e genes de cópia única (Miao *et al.*, 1991; Graphodatsky *et al.*, 2005; Rajcic *et al.*, 2015; Martis *et al.*, 2012; Valente *et al.*, 2014; Navarro-Domínguez *et al.*, 2017a; Clark *et al.*, 2017).

Em geral, não seguem a segregação mendeliana, apresentando descendentes em proporções diferentes de 0,5 (presença/ausência de B, B+ / B-). Portanto, o sucesso de sua manutenção depende de uma transmissão preferencial por um mecanismo de *drive* durante a divisão celular. O *drive* pode ocorrer por meio de dois tipos de mecanismos: i) envolve a não-disjunção dos cromossomos homólogos ou cromátides irmãs (mitose e meiose) e pela segregação preferencial (mitose e meiose) (Clark *et al.*, 2017; Jones, 2018). Outro ponto interessante é que o cromossomo B pode apresentar importante papel biológico na origem de cromossomos sexuais, determinação sexual e possível envolvimento no controle do ciclo celular (revisado em Martins *et al.*, 2011; Yoshida *et*

al., 2011; Valente *et al.*, 2014; Helleu *et al.*, 2016; Clark *et al.*, 2017), reforçando a relevância de sua manutenção dentro da população.

Em peixes, a presença de cromossomos B tem sido relatada em distintas populações de diversas espécies, sendo variáveis em tamanho, frequência e composição. Entretanto, os estudos estão, basicamente, restritos a análises citogenéticas convencionais envolvendo descrição da presença e frequência deste polimorfismo e, em alguns casos, sua natureza eucromática/heterocromática. Em relação à análise de cromossomos B em representantes da família Cichlidae, os mesmos foram descritos em sete espécies pertencentes à subfamília sul-americana (Cichlinae) (Feldeberg & Bertollo, 1984; Martins-Santos *et al.*, 1995; Feldberg *et al.*, 2004) e por volta de 20 espécies da subfamília africana Pseudocrenilabrinae, incluindo a espécie *Astatotilapia latifasciata* (Poletto *et al.*, 2010b; Fantinatti *et al.*, 2011; Yoshida *et al.*, 2011). Análises mais detalhadas realizadas por Yoshida e colaboradores (2011) no ciclídeo *Lithochormis rubripinnis* oriundo do Lago Victoria, Leste da África, sugeriram que nesta espécie, a presença do cromossomo B pode ter um efeito na determinação do sexo, influenciando o desenvolvimento de fêmeas. Entretanto, em outras doze espécies desse mesmo lago, incluindo a espécie-modelo desse projeto, *A. latifasciata* (Poletto *et al.*, 2010a, 2010b; Yoshida *et al.*, 2011; Kuroiwa *et al.*, 2014), ambos os sexos podem apresentar o cromossomo B. Recentemente, foi identificado a presença do cromossomo B somente em fêmeas de sete espécies de ciclídeos do Lago Malawi, Leste da África (*M. lombardoi*, *Metriaclima zebra* “Boadzul,” *M. zebra* “Nkhata Bay”, *Metriaclima greshakei*, *Metriaclima mbenji*, *Labeotropheus trewavasae*, *Melanochromis auratus*) (Clark *et al.*, 2017).

1.2 Origem do cromossomo B

Em relação à sua origem, o cromossomo B não possui uma história comum entre espécies, apresentando variações em sua estrutura e propriedades da cromatina (Houben, 2017). Zhou e colaboradores (2012) sugerem que a origem do cromossomo B em *Drosophila albomicans* ocorreu pela fusão entre o cromossomo 3 e um cromossomo sexual ancestral. Martins e colaboradores (2012) propuseram que a origem do cromossomo B em *Secale cereale* advém da translocação e segregação desigual após segmentação ou duplicação gênica da espécie. Já no ciclídeo *A. latifasciata*, a duplicação de um segmento do cromossomo A (incluindo o centrômero) deu origem a um proto-

cromossomo (proto-B). Esse proto-B foi submetido a eventos de duplicação, formando um isocromossomo, e, em seguida, houve um enriquecimento de sequências provenientes do complemento A (Figura 1) (Valente *et al.*, 2014).

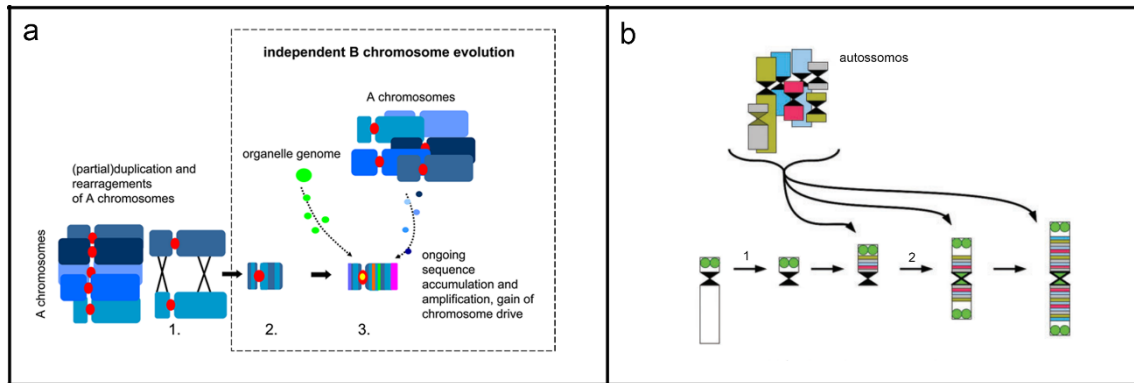


Figura 1. Hipótese da origem do cromossomo B. Em (a) o surgimento de um proto-B após a recombinação e um enriquecimento por sequências de DNA de organelas e DNAs repetitivos do genoma A (Houben et al., 2013). Em (b) um resquício de cromossomo (com centrômero) duplicado origina a um proto-B. Esse proto-B foi submetido a eventos de duplicação formando um isocromossomo que sofreu enriquecimento de sequências provenientes do complemento A, originando o cromossomo B (Valente *et al.*, 2014).

1.3 *Astatotilapia latifasciata* (Perciformes, Cichlidae) como modelo para análises de cromossomos B

A espécie *Astatotilapia latifasciata* (Figura 2) pertence à família Cichlidae (Perciformes), que, por sua vez, é uma das famílias de peixes com maior número de espécies (Nelson, 2006), com cerca de 3.000 representantes distribuídos pela América Central e do Sul, Madagascar, Sudeste da Índia e África (Kocher, 2004), sendo a maior diversidade ocorrente nos grandes lagos africanos (Trewavas, 1983). Este grupo de peixes apresenta uma rápida radiação adaptativa nos grandes lagos do leste da África, onde quase 2.000 espécies diversificaram-se nos últimos 10 milhões de anos (Kocher, 2004). Devido a essas características biológicas mencionadas, os ciclídeos são considerados excelentes modelos experimentais para estudos em diversas áreas da genética (Chang *et al.*, 2005; Ijiri *et al.*, 2007; Wang *et al.*, 2007; Cnaani *et al.*, 2008; Ferreira e Martins, 2008; Valente *et al.*, 2009; Poletto *et al.*, 2010a,b; Brawand *et al.*, 2014;).



Figura 2. Ciclídeo africano *Astatotilapia latifasciata*. Nessa espécie é possível observar o dimorfismo sexual, os machos apresentam uma coloração amarelada e vermelha e as fêmeas são acinzentadas. Parte inferior da foto um exemplar macho e na porção superior uma fêmea (foto: Adachi).

Estudos estruturais e funcionais neste grupo de peixes tem se mostrado promissores em análises de diversificação de espécies, diferenciação sexual, mecanismos evolutivos relacionados diretamente aos seus genomas e cariótipos e análises de processos metabólicos. O sequenciamento completo dos genomas e transcriptomas das espécies de ciclídeos (Brawand *et al.*, 2014) estão disponíveis no BouillaBase (bouillabase.org) e SaciBase (sacibase.ibb.unesp.br), o que possibilitará abrir novas fronteiras ao conhecimento estrutural, funcional e evolutivo neste grupo.

Além disso, *A. latifasciata* apresenta de 0 a 2 cromossomos B heterocromáticos, com tamanho relativamente grande, similar ao primeiro par cromossômico do complemento A (Figura 3). Estes cromossomos extras não seguem a segregação mendeliana e apresentam sequências de DNAs repetitivos, DNAs ribossomais e genes de cópia única (Poletto *et al.*, 2010b; Fantinatti *et al.*, 2011; Valente *et al.*, 2014). Recentemente, Carmello e colaboradores (2017) demonstraram a expressão diferencial de uma variante do gene hnRNP Q-like em fêmeas portadoras do cromossomo extra nesta espécie, evidenciando a relação entre cromossomo B e determinação sexual nos ciclídeos,

o que pode estar associado ao *drive* meiótico em fêmeas (Ramos *et al.*, 2016; Clark *et al.*, 2017). Adicionalmente, Venturelli e colaboradores (em preparação) analisaram cruzamentos direcionados de indivíduos com Bs (B+) e sem Bs (B-) de *A. latifasciata* e observaram uma frequência de aproximadamente 75% de proles B+ do cruzamento entre fêmeas B+ e machos B- e B+, evidenciando uma possível influência desse cromossomo em fêmeas (Yoshida *et al.*, 2011; Clark *et al.*, 2017).

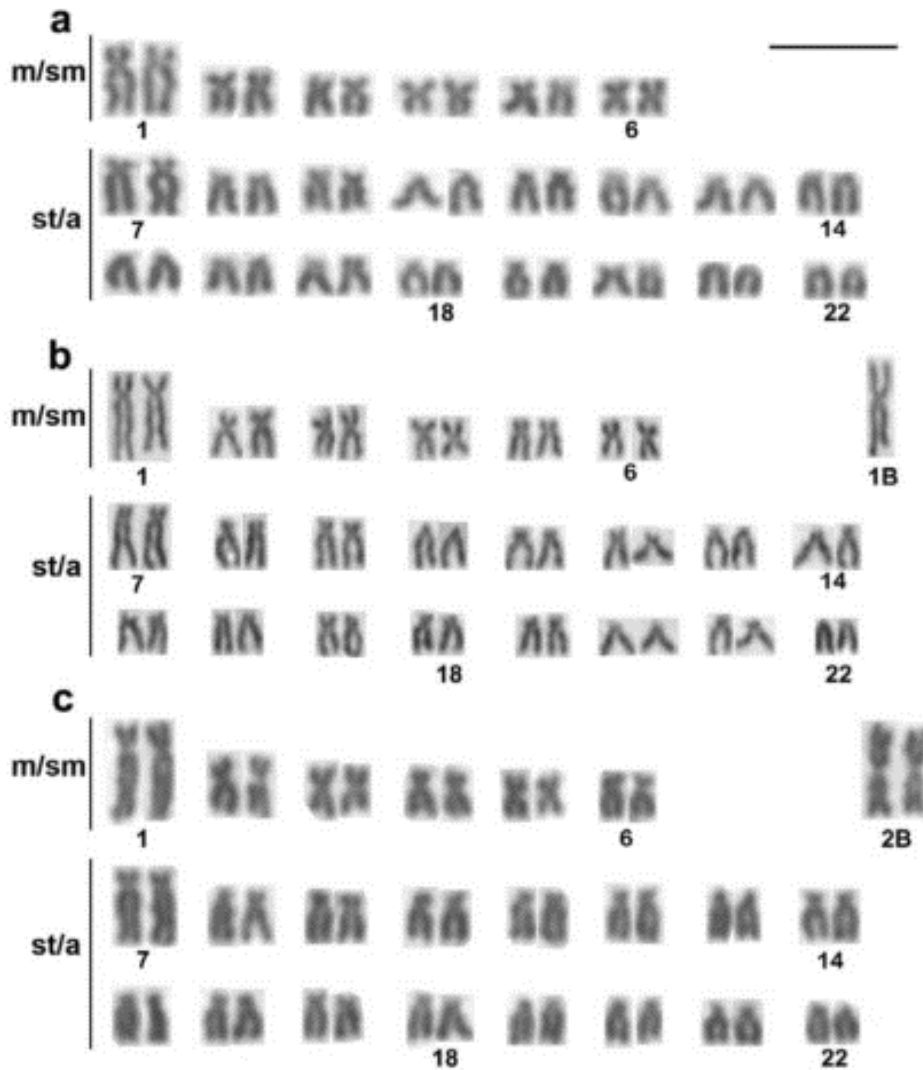


Figura 3. Cariótipo do ciclídeo *Astatotilapia latifasciata*. Cariótipo de indivíduos sem cromossomo B (a), com um (b) e dois (c) cromossomos supranumerários. Coloração em giemsa. (Poletto *et al.*, 2010).

1.4 Análises genômicas em larga escala de cromossomos B em *Astatotilapia latifasciata*

Junto a estudos citogenéticos, abordagens baseadas em análises genômicas em larga escala estão sendo utilizadas para elucidar questões acerca dos cromossomos

supranumerários. Valente e colaboradores (2014) analisaram os genomas de *A. latifasciata* com a presença e ausência do cromossomo B (B+ e B-), utilizando sequenciamento em larga escala. Neste estudo foram realizados procedimentos para identificação de regiões diferencialmente cobertas entre os dois genomas, permitindo a identificação de regiões “enriquecidas” nos genomas B+ e, possivelmente, presentes no cromossomo B.

Embora a maioria dos genes/sequências presentes no B sejam fragmentados, um pequeno número deles apresenta alta integridade. Entre os genes com alta integridade detectados, destaca-se um conjunto de genes envolvidos com controle da divisão celular. Assim, acredita-se que a presença de genes envolvidos com ciclo celular (Tabela 1) venha desempenhando papel importante na transmissão e manutenção do cromossomo B em *A. latifasciata* (Valente *et al.*, 2014).

Da mesma forma, o transcriptoma de *A. latifasciata* mostrou a presença de genes envolvidos com controle da divisão celular sendo transcritos pelo cromossomo B, o que também foi encontrado para o ciclídeo *Pundamilia nyererei* (Valente *et al.* 2014). As análises sugerem um possível papel funcional destes genes na manutenção dos cromossomos B nas células. Dessa forma, buscas realizadas em transcriptomas de *P. nyererei* (disponíveis em Bouillabase.org) revelaram que esta espécie contém transcritos B-específicos para os genes Separina, TUBB1 e KIF11 em diversos tecidos (Tabela 1) (Valente *et al.*, 2014). A partir destas informações, torna-se evidente que a presença de cromossomos B deve contribuir com novos transcritos ou mesmo alterar os padrões de expressão geral da célula.

Tabela 1. Genes envolvidos com ciclo celular e presentes no cromossomo B, de acordo com Valente *et al.* (2014). Os genes destacados em negrito apresentam evidências de atividade transcricional.

Nível de integridade (%)	No. de Acesso GenBank	Descrição gênica
70.93	XM_004570740.1	DNA repair protein XRCC2-like, mRNA
70.97	XM_004570733.1	Synaptonemal complex protein (SYCP) 2-like, mRNA
73.92	XM_004575607.1	Centromere-associated protein (CENP) E-like, mRNA
75.71	XM_004545355.1	Separin-like, mRNA

77.16	XM_004547754.1	Aurora kinase (AURK) A-B-like, mRNA
96.14	XM_004570623.1	Kinesin-like protein KIF11-like , mRNA
97.50	XM_004572890.1	Tubulin beta-5 (TUBB5) chain-like, mRNA
100	XM_004547671.1	Tubulin beta-1 (TUBB1) chain-like, mRNA
100	XM_004556260.1	Spindle and kinetochore-associated (SKA) protein 1-like, mRNA
100	XM_004576295.1	Regulator of telomere elongation helicase (RTEL) 1-like, mRNA

Legenda: Primeira coluna apresenta o nível de integridade dos genes encontrados. Segunda coluna, os números de acessos dos genes encontrados de acordo com o *GenBank*. Terceira coluna, a descrição dos respectivos genes. Os genes destacados em negrito apresentam evidências de atividade transcricional.

Para os miRNAs, com base nas mesmas amostras processadas para transcriptomas de *A. latifasciata*, foram construídas bibliotecas de microRNAs a fim de avaliar o perfil de expressão dos mesmos (Fantinatti, 2015). Em uma visão geral dos perfis de expressão de microRNAs, com base no banco de dados mirbase (<http://www.mirbase.org>), foi constatado que um total de 162 microRNAs apresentaram um padrão diferenciado entre as várias amostras considerando o sexo e a presença e ausência do cromossomo B.

1.5 Análises funcionais: Proteômica

Como mencionado acima, o avanço nas tecnologias de análise de DNA e RNA propiciou um melhor entendimento da composição genômica e expressão gênica nos organismos. Entretanto, estas informações ainda não são suficientes para elucidar precisamente os níveis de abundância proteica nas diferentes condições celulares, como observado por Vogel e Marcotte (2012), que demonstraram uma baixa correlação entre abundância de RNAm e proteínas em ratos ($R^2=0,4$). Além disso, nem sempre os bancos de dados fornecem informações quantitativas e qualitativas que possam ser relacionadas às proteínas. Tais questões podem ser mais bem estudadas utilizando-se a proteômica como abordagem. Sua proposta é a de analisar de forma global o conjunto de proteínas expressas em uma célula, tecido ou organismo (Wasinger *et al.*, 1995), permitindo avaliar quantitativa e qualitativamente as proteínas que atuam no metabolismo celular (Park, 2004; Chen e Harmon, 2006). Proteínas são determinantes na função celular e sua dinâmica de interações pode revelar muito mais sobre eventos célula/tecido específico que uma sequência genômica estática (Blakeley *et al.*, 2010).

Estudos proteômicos frequentemente utilizam técnicas de eletroforese bidimensional e de cromatografia líquida acopladas à espectrometria de massas,

permitindo a geração de dados em larga escala, como observado na comparação entre plasma seminal e sanguíneo de carpa (Dietrich *et al.*, 2014) ou no estudo do plasma humano (Jiang *et al.*, 2013). A proteômica permite estudar a estrutura, função e o controle dos sistemas biológicos pelas propriedades das proteínas, incluindo a sequência (identidade), abundância, atividade e estrutura das proteínas expressas por uma célula, assim como as modificações, interações e translocações sofridas por elas (Wilkins *et al.*, 1997). Logo o proteoma pode se modificar em resposta às condições e estímulos a que o organismo está exposto, além de características intrínsecas como presença ou ausência de componentes celulares.

Nesse projeto foi analisado pela primeira vez o perfil de abundância das proteínas dos indivíduos com presença e ausência do cromossomo B, visando identificar possíveis proteínas relacionadas ao cromossomo B em tecidos específicos de machos e fêmeas do ciclídeo *Astatotilapia latifasciata*. Em seguida, os dados foram integrados com os estudos de transcriptoma.

1.6 Integração de dados

O estudo genômico funcional se concentra nos aspectos dinâmicos, desde a transcrição de um gene, tradução e interações entre proteínas. Portanto, com dados de RNA-Seq (sequenciamento) e proteoma é possível integrar redes biológicas de interação entre proteínas, perfil de expressão de RNAs (incluindo miRNAs) e vias metabólicas em um contexto de biologia sistêmica (Rhodes *et al.*, 2005; Junker, 2008). Estas análises integrativas, a partir da construção de grafos de co-expressão dos dados do transcriptoma (dados brutos, Marques, 2016) e dos dados gerados desse trabalho, permitirão avanços na elucidação do papel biológico dos cromossomos B.

1.6.1 Redes biológicas

Conceitualmente, qualquer sistema complexo é constituído por um grande número de entidades (partículas ativas ou agentes) que podem interagir mutuamente e com o ambiente exterior (Bellomo e Carbonaro, 2011). Essa maquinaria complexa é realizada pelas interações entre as várias unidades que compõem o sistema vivo, tais como DNA, RNA, proteínas e pequenas moléculas, em conjugação com as informações vindas do meio externo. Essas interações podem ser representadas pela teoria de grafos ("redes").

São essas redes, portanto, que representam a maquinaria complexa que executa as instruções contidas nos genomas transformando-as em atividades biológicas pelas quais a vida se expressa. As redes são representadas por dois conjuntos: $G = \{V, A\}$, onde V é um conjunto de vértices e A um conjunto de pares formados por arestas que conectam cada dois vértices. Se os pares forem ordenados, a rede é classificada como direcionada, caso contrário é denominado de não-direcionada (Figura 4). As interações entre proteínas, por exemplo, por não importar a direção da interação entre elas são geralmente reportadas como grafos não-direcionados (Junker, 2008). As redes de regulação como o controle da expressão gênica, depende dos fatores que modulam a expressão do gene de interesse, no qual a ordem das interações é relevante para o resultado final. Nesses modelos, a rede é demonstrada por um grafo direcionado. As redes de transdução de sinais são representadas por grafos direcionados multi-arestas que representam as interações uni ou bi-direcionais entre proteínas, outras moléculas orgânicas ou inorgânicas (Junker, 2008; Pavlopoulos *et al.*, 2011).

Para obtermos respostas dessas redes complexas, como por exemplo, quais são os genes mais “importantes”, determinamos medidas que nos permitem caracterizar os vértices dessas redes, que são as topologias das redes complexas. As principais topologias de uma rede são: i) Grau (“*Degree*”) (o número de vizinhos do vértice v no caso de grafos direcionados, medindo o número de ligações que chegam e saem do vértice respectivamente), ii) Comprimento médio do caminho (“*Average path length*”)(número médio de conjuntos de nodos capazes de interligar dois nodos distintos v_i e v_j utilizando o menor número de nodos possível), iii) Coeficiente de agrupamento (“*Clustering coefficient*”) (mede o grau de conexões próxima a um determinado vértice), iv) Assortatividade (“*Assortativity*”) (representa a extensão dos nós de uma rede e como se associam a outros nós na rede, redes biológicas tendem a ser disassortativas, valores negativos para assortividade), v) Grau de intermediação (“*Betweenness centrality*”) (mede quanto um vértice está nos menores caminhos entre outros vértices na rede) (Albert e Barabási, 2002; Foster *et al.*, 2010).

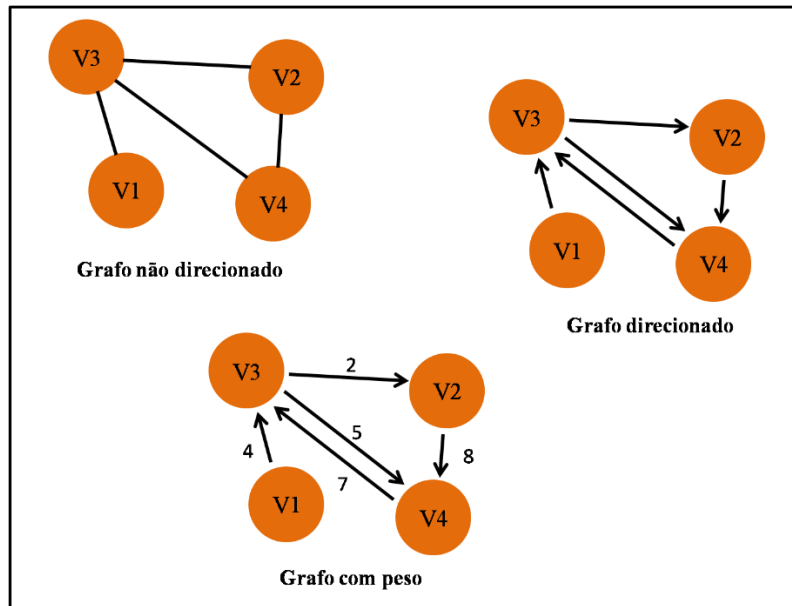


Figura 4. Exemplo de grafos direcionados, não direcionados e com peso. Figura baseada no trabalho de Pavlopoulos *et al.*, (2011).

1.7 Ciclo celular

O ciclo celular é um processo fundamental dos seres vivos multicelulares. Apresenta grande importância desde o desenvolvimento até a vida adulta, contribuem para a geração de células especializadas, que são necessárias para formar tecidos e desempenhar funções específicas. O controle do ciclo celular precisa ser intimamente preciso para não acarretar erros no desenvolvimento e surgimento de tumores. Nos eucariotos, esse processo está basicamente dividido em duas fases: Interfase e Mitose, com a presença de dois processos, entre elas G1 e G2. Nessas duas sub fases as células obtêm massa, integram sinais de crescimento, organizam a replicação gênica e preparam a segregação dos cromossomos (Barnun e O’Connell, 2014).

Durante o ciclo celular ocorrem vários processos regulatórios e de *checkpoint* para garantir o sucesso da divisão celular. Existem vários mecanismos de *checkpoint* já descritos como o crescimento celular adequado, resposta ao dano no DNA, monitoramento da replicação do DNA, dependência da fase S-M, e o *checkpoint* do fuso mitótico (SAC). Além disso, o ciclo celular é intimamente regulado, principalmente, por ciclinas dependentes de quinase (CDKs) e outras proteínas com grande importância como por exemplo: as proteínas MPS, a qual é uma das principais quinases envolvidas na localização do cinetocoro e no ponto de verificação do conjunto do fuso (SAC), que inibe

a transição da metáfase para anáfase até que as fibras do fuso estejam corretamente ligadas a todos os cromossomos, a aurora quinase A, a qual regula a ligação dos fusos mitóticos, mantendo a bipolaridade e também está relacionada com a maturação do centrômero; a proteína APC/C atua no *checkpoint* mitótico realizando a ubiquitinação da Securina liberando a Separase para separação das cromátides irmãs; o complexo NDC80 formado por um conjunto de quatro proteínas (Ndc80, Nuf2, Spc25 e Spc24), as quais são responsáveis pela ligação do cinetócoro ao microtubulos durante a divisão celular; e as proteínas da família 2.4.1.1, que atuam como estacionadores, ou seja, param o ciclo em diversos pontos da interfase e metáfase (Kaistha *et al.*, 2014; Ducat e Zheng, 2004; Hermeking e Benzinger, 2006; Ciferri *et al.*, 2007; Bolanos-Garcia e Blundell, 2011; Ruijtenberg e Heuvel, 2016).

Vários estudos referentes ao cromossomo B vem demonstrando a presença de genes envolvidos no ciclo celular (Valente *et al.* 2014; Helleu *et al.*, 2016; Clark *et al.*, 2017; Navarro-Domínguez *et al.* 2017a,b; Kichigin *et al.* 2019; Makunin *et al.* 2018; Jones, 2018). Makunin e colaboradores (2018) identificaram o gene *centromere protein N* na raposa *Vulpes vulpes L.*, Navarro-Domingues e colaboradores (2017b) encontraram genes ativos relacionados a divisão celular para os genes *CIP2A protein(cip2a)*, *cytoskeleton-associated protein 2(ckap2)*, *condensin I complex subunit G(cap-g)*, *kinesin-like protein kif20a (kif20a)* e *E3 ubiquitin-protein ligase mycbp2 (mycb2)* no gafanhoto *Eyprepocnemis plorans*, Kichigin e colaboradores (2019) encontraram genes relacionados a segregação cromossômica e divisão celular (*inner centromere protein (incep)* e *protein spire homolog 2 (spire2)*, respectivamente) no lagarto *Anolis carolinensis*. Portanto, há indícios que a manutenção do cromossomo B, em diferentes espécies, pode estar relacionadas com a presença de genes reguladores do ciclo celular.

2 HIPÓTESE

A presença do cromossomo B influencia a regulação gênica e interfere no perfil proteico no ciclídeo africano *Astatotilapia latifasciata*.

3 PROPOSTA

Analisar a possível interferência da presença do cromossomo B em *A. latifasciata*, comparando a expressão de RNAm e abundância proteica de tecidos específicos de indivíduos B+ e B-. Além disso, a partir dos dados do transcriptoma (RNAm) realizamos uma análise em um contexto de biologia de sistemas.

4 OBJETIVOS

4.1 Objetivo Geral

- Elucidar a influência da presença do cromossomo B em seus hospedeiros em uma abordagem de genômica funcional e processos moleculares complexos na espécie *A. latifasciata*.

4.2 Objetivos Específicos

- Analisar o perfil proteico de tecidos específicos de indivíduos com e sem cromossomo B e relacioná-los com sua função;
- Identificar as principais vias metabólicas em que ocorrem alterações no nível de expressão relacionado à presença do B.
- Integrar dados em um contexto de biologia de sistemas, relacionando a expressão diferencial de RNAm.

5 MATERIAL E MÉTODOS

5.1 Animais, extração de DNA e genotipagem de animais B+ e B-

Linhagens da espécie *A. latifasciata* provenientes de lojas de aquarofilia foram mantidas em aquários aclimatados no biotério do Laboratório Genômica Integrativa (LGI) (IBB-UNESP-Botucatu-SP). De acordo com análises prévias, esta população mantida no LGI possui indivíduos portadores de cromossomo B (comitê de ética – Protocolo n° 769-2015).

A extração de DNA foi feita a partir da nadadeira caudal de 12 indivíduos seguindo o protocolo de Sambrook e Russel (2001). Este grupo amostral foi formado por 6 indivíduos portadores de 1 cromossomos B (3 machos e 3 fêmeas denominados indivíduos B+) e 6 indivíduos sem cromossomos B (3 machos e 3 fêmeas denominados B-). A genotipagem foi realizada via reação em cadeia da polimerase (PCR) utilizando *primers* descrito por Fantinati e Martins (2016). Após a genotipagem, os animais foram aclimatados e mantidos em um único aquário, a fim de diminuir o viés experimental (Figura 5).

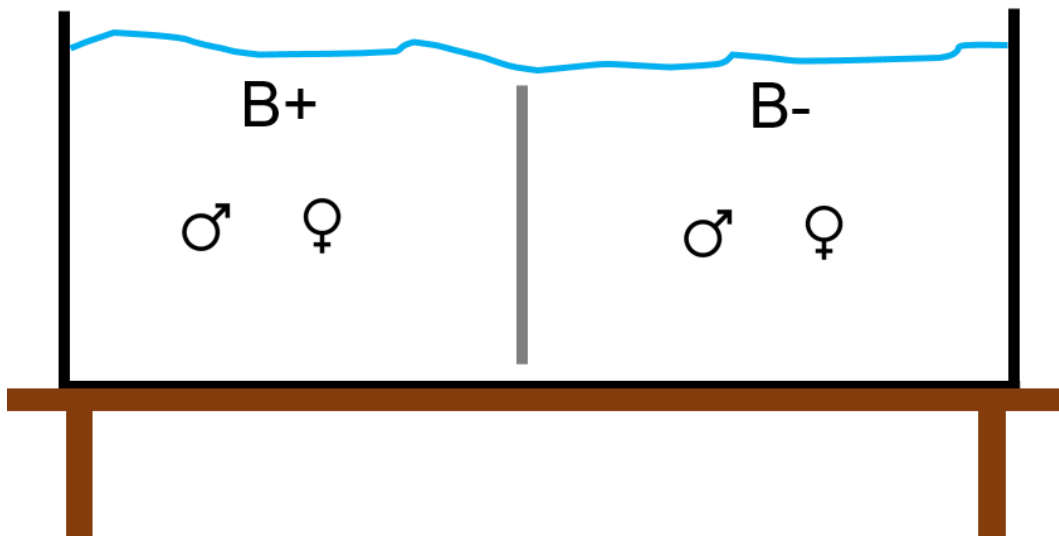


Figura 5. Representação ilustrativa do aquário em que foram mantidos os peixes para a extração proteica.

5.2 Estudo morfológico das gônadas

Para a confirmação do estágio de desenvolvimento das gônadas foi realizada a análise morfológica do tecido com a coloração de Azul de Toluidina (Pearse, 1972). Primeiramente, os tecidos extraídos em triplicata de indivíduos de ambos os sexos foram fixados em paraformaldeído 4% tamponado (pH 7,4) *overnight* (16 horas). Para a remoção do fixador dos materiais, estes foram desidratados em série alcoólica (70% e 95%) por 30 minutos cada a temperatura ambiente. A pré-infiltração foi feita em historesina + álcool 95% por pelo menos 12 horas. Em seguida, o material foi colocado em resina pura e deixado em *over night* para infiltração, sendo ambos os procedimentos mantidos em temperatura ambiente. A inclusão foi feita com resina básica pura + polimerizador em estufa à 37°C por 24 horas. Os blocos foram aderidos a um suporte de madeira e mantidos em estufa a 37°C. Os cortes histológicos foram corados com a

coloração Azul de Toluidina por 2 minutos e lavados em água corrente por 5 minutos e analisados em microscópio óptico convencional.

5.3 Extração de proteínas

Tecidos (músculo, encéfalo e gônada) de indivíduos adultos de genótipo B- e B+ foram extraídos em triplicatas em local esterilizado com água sanitária 10%, álcool 70% e rnaZeap. Imediatamente, após a coleta, os tecidos foram armazenados em nitrogênio líquido para conservação e posterior extração proteica. Na extração proteica foi utilizado um tampão de lise contendo urea a 7M, thiourea a 2M, Dithiothreitol (DTT) a 40mM e um coquetel de inibidor de proteases (Protease Inhibitor Cocktail Tablet, EDTA Free da Sigma-Aldrich). Para cada 100mg de tecido foram adicionados 500µl do tampão de lise seguido de homogenização. Após centrifugação à 14000 rpm a 4°C por 30 minutos, foi coletado o sobrenadante e armazenado em -80°C.

5.4 Quantificação, digestão e espectrometria de massa

A quantificação foi feita pelo método de Bradford (Bradford 1976) e a integridade foi analisada por meio da eletroforese em gel de poliacrilamida (SDSPAGE, Proto-Gel 30% da National Diagnostic) seguindo as recomendações do fabricante. Protocolo quantificação por Bradford: a) Diluir o reagente de Bradford (Sigma-Aldrich) em água ultra pura na proporção 4:1 e agitar levemente; b) Preparar a solução proteica de albumina bovina (BSA) 0,1% e a partir dela fazer quatro diluições: 1,2 µg/µl, 2,4 C, 4,8 µg/µl, e 6 µg/µl; c) Medir a absorbância (595 nm). A concentração das amostras foi baseada na equação linear gerada:

$$y=ax+b(1)$$

$$x= y-ba(2)$$

onde y é a absorbância e x a concentração em µg/µl.

A digestão enzimática das amostras (50 µg) foram diluídas em uma solução de bicarbonato de amônio 50mM e seguido de uma solução de redução e uma de alquilação com 10mM de DTT e 50mM de iodoacetamida, respectivamente. Em seguida, adicionou-se tripsina (1:50 com 50 mM de tampão de bicarbonato de amônio) (Promega) seguida

de 16 h de hidrólise finalizada com ácido fórmico a 1% (v / v). As amostras foram dessalinizadas com as colunas Sep-Pak C18 (Waters, Milford, MA, EUA), liofilizadas (SpeedVac; Thermo Scientific) e armazenadas a -80 ° C até a aplicação no espectrômetro de massas. Os péptidos extraídos foram redissolvidos em acetonitrilo a 70% (ACN) contendo solução de ácido fórmico a 0,1% antes da análise LC-MS / MS.

A análise dos peptídeos trópicos foi realizada no sistema nanoACQUITY UPLC (Waters, Milliford, USA) acoplado ao espectrômetro de massas Xevo Q-TOF G2 (Waters, Milliford, USA). Para tanto, o sistema UPLC nanoACQUITY foi equipado com uma coluna do tipo HSS T3 (Acquity UPLC HSS T3 column 75 mm x 150 mm; 1,8 µm, Waters), previamente equilibrada com 7% da fase móvel B (100% ACN + 0,1 % ácido fórmico). Os peptídeos foram separados através de um gradiente linear de 7-85 % de fase móvel B durante 70 min com fluxo de 0,35 µL/min e a temperatura da coluna mantida a 45°C. O MS foi operado em modo íon positivo, com o tempo de aquisição de dados de 75 min. Os dados obtidos foram processados através do software ProteinLynx GlobalServer (PLGS) versão 3.03 (Waters, Milliford, USA). A identificação das proteínas foi obtida através do algoritmo de contagem de íons incorporado ao software. Os dados obtidos foram buscados no banco de dados da espécie *Danio rerio* baixado do catálogo do UniProt (Universal Protein Resource) em setembro de 2017 e utilizamos os seguintes parâmetros: modificação fixa carbamidometilação; modificação variável oxidação da metionina; uma clivagem perdida pela enzima; massa molecular do tipo monoisotópica; erro de tolerância de peptídeos (MS) ± 0,2 Da e erro de tolerância (MS/MS) ± 0,2 Da; protonação + 2,+3+4, tipo de instrumento ESI-Tof.

5.5 Análise dos dados - Proteômica

Considerando ser a primeira análise em larga escala realizada utilizando a espectrometria de massa (*shotgun*) descrita para a espécie *A. latifasciata* e para estudos sobre cromossomo B, foi feita uma análise exploratória dos dados, onde as proteínas foram filtradas com corte mais baixo (*pair index* > 1 e *score* > 30). Foram comparadas amostras de acordo com o tecido (gônada, encéfalo e músculo) de macho e fêmea com a presença ou ausência do cromossomo B (gônada: macho B+versus macho B-; gônada fêmea B+versus fêmea B-; encéfalo: macho B+versus macho B-; encéfalo: fêmea B+versus fêmea B-; músculo: macho B+versus macho B-; músculo: fêmea B+versus fêmea B-, Tabela 2). Para as proteínas com abundâncias alteradas (expressão diferencial

– *up/down*), foram feitos s enriquecimento de termos ontológicos considerando um valor de corte com *p-value* < 0,05 através do G:profiler (Reimand, *et al.*, 2016) e para facilitar as análises dos termos enriquecidos foi feito agrupamento desses com o *software* REVIGO, considerando *p-value* (Supek, *et al.*, 2011). Além disso, as proteínas com as expressões alteradas foram analisadas com o pacote *Pathview* (Luo et al., 2013) do Bioconductor no ambiente R em busca das vias biológicas pertencentes. As vias foram obtidas do banco de dados KEGG (Ogata et al., 1999) e o organismo modelo selecionado foi o teleósteo *Danio rerio*. Foram plotadas todas as vias para as quais haviam proteínas com expressão alterada, as quais apresentavam um identificador no banco de dados de *D. rerio*. Assim, foram gerados mapas em conjunto para proteínas com *up/down* (caixas em vermelho/azul) e mapas em conjunto para proteínas exclusivas B+/B- (caixas em vermelho/azul)(Suplementar 1: Mapas Biológicos).

Tabela 2. Comparações realizadas para expressão diferencial das proteínas.

Tecido	Grupos
Gônada	Macho B+ versus Macho B- Fêmea B+ versus Fêmea B-
Encéfalo	Macho B+ versus Macho B- Fêmea B+ versus Fêmea B-
Músculo	Macho B+ versus Macho B- Fêmea B+ versus Fêmea B-

Legenda:B+ presença de cromossomo B, B- ausência do cromossomo B.

5.6 Reanálise do transcriptoma (RNAm)

Para as análises de anotação gênica, expressão de RNAm, predição de proteína e redes biológicas foi avaliado o sequenciamento do transcriptoma de RNAm realizado por Marques (2016). Os sub itens 5.5.1 e 5.5.2 até a última montagem gerada pelo CAP3 foram descritas por Marques (2016). A filtragem final, o mapeamento e as demais análises foram feitas nesse trabalho.

5.6.1 Filtragem das *reads*

As amostras de RNA foram sequenciadas utilizando a plataforma Illumina Hi-seq 2000, 100 x 2 *paired end reads* (4 amostras por *lane*). Este serviço foi realizado pela empresa LC Science (EUA). A montagem dos transcritos de cada biblioteca foi realizada a partir de 36 bibliotecas correspondentes a triplicatas de RNAm de gônadas, de encéfalo e de músculo de indivíduos machos e fêmeas com e sem Bs de *A. latifasciata* (Marques, 2016).

A análise de qualidade dos *reads* e identificação de sequências altamente representativas foram realizadas pelo *software* FASTQC. A eliminação de contaminantes do sequenciamento e seleção por qualidade do *read* foi baseada no valor de phred 30 de qualidade de *reads* em pelo menos 80%, utilizando o pacote Trimmomatic V0.30 (Bolger et al., 2014).

Os *reads* restantes foram alinhados contra um banco de dados de sequências de contaminantes, com 3,126,425 sequências de rDNA, tRNAs, genoma mitocondrial e bacteriano obtidos no banco de dados Refseq (NCBI). Os alinhamentos dos *reads* foram gerados utilizando o programa Bowtie2 com a opção “--very-fast-local”. Os *reads* não alinhados foram submetidos a verificação de adaptadores (Illumina) utilizando os *softwares* Trimmomatic V0.30 (Bolger et al., 2014) e blastn localmente com um corte de $e_value \leq 10e-5$ e 90% de identidade. Os *reads* não pareados foram separados pelo programa Pairfq. As bibliotecas relacionadas a um fragmento (*paired end*) foram agrupadas criando dois conjuntos de dados para montagem *de novo*.

5.6.2 Montagem *de novo* e filtragem

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

A montagem foi iniciada utilizando uma escala de k-mers de 21 a 81 nucleotídeos, alternando em 4 nucleotídeos de um para outro, com obtenção de 14 montagens independentes usando o Velvet (parâmetros “-ins_length 300 -min_trans_lgth 100”)

seguido do Oases (parâmetros “-ins_length2 300 -min_trans_lgth 100”). Os *reads* foram mapeados em cada montagem (opção *single read mode* and “--very-sensitive”) e as médias dos mapeamentos que apresentaram porcentagem de mapeamento $\geq 82.8\%$ foram selecionadas para o próximo passo (Suplementar 2: Tabela 1, linhas 2-22).

Para a montagem final, foi utilizado o programa CAP3 (Huang e Madan, 1999) em dois passos. As montagens consenso foram realizadas utilizando as montagens de k-mers 41 a 73 (9 montagens) e foram agrupadas em um único arquivo, os transcritos com porcentagem de “N” $\geq 10\%$ foram excluídos (Suplementar 2: Tabela 1, linhas 24-34). Devido ao grande número de sequências (1,435,763 transcritos), o arquivo da montagem foi dividido em dois grupos: um com k-mers de escala de 41 a 57 nucleotídeos e o segundo com escala de 61 a 73 nucleotídeos. Essa divisão em dois grupos foi estipulada medindo a quantidade de dados que o computador poderia suportar para uma única corrida CAP3. As primeiras montagens foram realizadas separadamente para cada grupo, utilizando a combinação de parâmetros (-f 5 ou 20 e -p com um range de 66-90, assim como “-h 20 -k1 -r 1”). Utilizando uma taxa (quantidade de *singlets* / quantidade de *contigs*) que poderia expressar a melhor montagem no sentido de ter tantos *contigs* quanto possível em comparação com o *singlets* (menor taxa avaliada para cada conjunto de dados), foi selecionado a melhor montagem (parâmetros do CAP3 “-f 5 -h 20 -k 1 -p 66 -r 1”) (Suplementar 2: Tabela 1, linhas 36-52).

Os melhores *contigs* e *singlets* gerados pelo CAP3 foram utilizados na montagem de ambos os grupos sendo submetidos a uma montagem final, utilizando os mesmos parâmetros (Suplementar 2: Tabela 1, linhas 56-58).

Os transcritos gerados foram mapeados no genoma de uma espécie próxima filogeneticamente, *Metriaclima zebra*, utilizando o *software* GMAP (gmap_build: parâmetro “-k 15” e para o gmap os parâmetros “-k 15 -B 5 --allow-close-indels 0 -p 3 -O -E "cdna" -f 2 -n 1 --nofails --min-trimmed-coverage 0.8 --min-identity 0.8 --no-chimeras”). Nesse passo de alinhamento não foi utilizado o genoma da espécie em estudo, *A. latifasciata*, por não estar finalizado. Para a anotação funcional da montagem do transcriptoma foi utilizado o programa Trinotate (parâmetro *default*) (Bryant et al., 2017). Foram selecionados os transcritos que obtiveram resultado para o blastp e blastx e considerando a presença de pelo menos 50% de cobertura de alinhamento da sequência do banco de dados sobre o transcrito.

5.6.3 Entropia de Shannon

A diversidade no nível de expressão dos transcritos foi analisada medindo sua entropia de *Shannon* (Eqs 3 e 4), a qual quantifica a complexidade do perfil de expressão dos transcritos. A entropia é dada

$$S = -\sum_{i=1}^n p_i \log_2 p_i (3)$$

onde

$$p_i = \frac{e_i}{\sum_{i=1}^n e_i} (4)$$

e_i é o valor de expressão do *transcrito*_{*i*} na amostra (DESeq count); corresponde a probabilidade de selecionar o *transcrito*_{*i*} dada a amostra; $-\log_2 p_i$ é a informação (mensurado em bits) obtido pelo evento observado; S é o valor esperado. No contexto atual, S é uma medida para a diversidade do perfil de expressão das amostras.

A entropia mínima é obtida se todos os transcritos de uma dada amostra for qualitativamente igual ($p = 1$ para um dado transcrito e $p = 0$ para todos os outros); reciprocamente, o valor máximo é obtido se todos os transcritos observados apresentarem a mesma frequência ($p_i = \frac{1}{n}$). Nesse caso, $S = \log_2 n$, or rather, o valor de entropia máxima depende do número n dos transcritos disponíveis. Conseqüentemente, existem duas fontes principais para o aumento da entropia: o número de transcritos e a dispersão da expressão de transcritos. Os valores baixos de entropia indicam um simples mecanismo regulatório, seja por expressão concentrada ou por ocorrência em poucos transcritos, enquanto valores de entropia maiores indicam um complexo processo regulatório causado pela expressão difusa de vários transcritos. Foi calculado o desvio padrão entre as triplicatas e por seguinte o intervalo de confiança de cada triplicata ($\alpha < 0,05$).

5.6.4 Expressão diferencial de RNAm

As *reads* de cada biblioteca (36 bibliotecas - 12 amostras) do transcriptoma de *A. latifasciata*, divididas em triplicatas de gônadas, encéfalo e músculo de indivíduos machos e fêmeas foram filtradas de acordo com a qualidade do sequenciamento e remoção de adaptadores feitas na montagem (descrito acima). Para análise de expressão diferencial das bibliotecas extraídas, as *reads* foram alinhadas contra a montagem final do transcriptoma de *A. latifasciata* utilizando o programa Bowtie2 com parâmetro “--b2-very-sensitive” (Langmead e Salzberg, 2012) e contagem dos transcritos pelo programa BEDtools multicov (Quinlan e Hall, 2010), além da análise das bibliotecas e do nível de expressão com a utilização do programa DESeq2 (Anders et al. 2014; Love et al., 2014). Foram considerados diferencialmente expressos transcritos com um *fold change* de 2 e com $p < 0,05$. Além disso, foi feito a busca por termos biológicos enriquecidos dos transcritos diferencialmente expressos com o software G:Profiler (Reimand J et al., 2016), baseado no banco de dados do ciclídeo *Oreochromis niloticus*.

5.6.5 Análise comparativa

Para verificarmos a relação entre duas variáveis, cromossomo B (presença/ausência) e sexo (macho/fêmea) nos transcritos diferencialmente expressos, foi aplicado o *Fisher's exact test (two tail)*, no qual foram utilizados valores normalizados pelo *software* DESeq2 (Anders et al., 2014; Love et al., 2014). Foram considerados valores significativos com $p < 0,05$, ou seja, a expressão diferencial depende da presença/ausência do cromossomo B e do sexo (macho/fêmea).

A prospecção de transcritos diferencialmente expressos e proteínas com abundância alteradas em comum, assim como os termos biológicos, foram agrupadas com base na teoria de conjuntos com a construção de diagramas de *Venn* (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

5.7 Construção das redes

A rede de co-expressão foi construída a partir das contagens dos *reads* das 36 bibliotecas normalizadas pelo nível de expressão em “rlog” a partir do programa DESeq2

(Anders *et al.*, 2014; Love *et al.*, 2014). Os dados obtidos foram organizados em uma tabela onde cada linha representa um gene e cada coluna um tecido/experimento. Esta tabela serviu de arquivo de entrada para Correlação de *Spearman* (*Rank Order Correlation*) e foram filtradas a 0,92 obedecendo assim a lei de potência, premissa para redes biológicas reais.

Em relação a rede de interação proteína-proteína (PPI-net), foi filtrada a montagem do transcriptoma que possuía porcentagem de repetição inferior a 40% e posterior seleção com base nos dados obtidos com o programa RepeatMasker (<http://repeatmasker.org>; Smit *et al.*, 2013) no Repbase (www.girinst.org/repbase; Jurka *et al.*, 2005).

Em seguida, foram selecionados os transcritos com as *openreadingframes* (ORFs) que iniciavam com uma metionina e com mais de 100 resíduos de aminoácidos com o programa getorf da suíte de aplicativos emboss (Rice *et al.*, 2000). As ORFs foram submetidas a análise de similaridade pelo *software* blastp (versão 2.3.0+), instalado localmente (Altschul *et al.*, 1990), onde ORFs com Identity percentage > 60,428% ou Hit Coverage > 48,899% ou Query Coverage > 64,901% ou ainda um e-value < 1,14e-08 foram selecionadas. Adicionalmente, as ORFs também foram submetidas a análise de domínios do banco de dados Pfam (www.pfam.xfam.org; Finn *et al.*, 2014) pelo o *software* HMMER (hidden Markov models, www.hmmer.janelia.org; Finn *et al.*, 2011) e aquelas com valores de e-value < 0,2 e i-value < 0,3 ou acc > 0,55 também foram selecionadas. As duas listas de ORFs foram mescladas e as redundâncias foram excluídas, formando o conjunto final de proteínas submetidas a predição de interação pelo *software* UNISPPI, onde somente interações com score superiores a 0,724 foram aceitas como interações na rede (Valente *et al.*, 2013). Como resultado desta etapa foi construído a rede PPI-net.

A partir dos transcritos diferencialmente expressos foram extraídos sub redes da co-expressão e da PPI-net para cada tecido. A união das redes de co-expressão e PPI-net(co-PPI-net) foram extraídas pelo programa Cytoscape (Shannon *et al.*, 2003) e as topologias foram analisadas por um script *in house* em linguagem *Python*. Desse modo, observamos a expressão em comum entre transcritos e a possível interação física que pode ocorrer entre eles. Além disso, para verificarmos de forma geral se as subredes (*down regulated* e *up regulated*) referentes a presença do cromossomo B apresentaram

diferenças, foi analisado a distribuição dos dados através do teste de *Anderson-Darling normality test* e a diferença entre as subredes pelo teste de *Wilcox test* com valor significativo para $p\text{-value} < 0,05$, ou seja, demonstrando ter diferença significativa entre as subredes. Os dois últimos testes foram realizados com pacotes “normtest” e a função “wilcox.test” em ambiente R, respectivamente.

6 RESULTADOS

6.1 Análise histológica

Os resultados morfológicos dos cortes histológicos das gônadas de macho demonstraram presença de todos os tipos celulares da maturação sexual das gônadas (espermatogônias A (indiferenciadas/diferenciadas), espermatogônias B, espermatócitos (em divisão), espermátides (I, II e III) de acordo com Schulz e colaboradores (2010)). Nas fêmeas, por apresentarem tipos celulares diferentes dependente do estágio de maturação (desenvolvimento contínuo assincrônico do oócito), foram selecionadas gônadas que não estavam completamente maduras para evitar extrair proteínas somente de oócitos em vitelogênese avançada, diminuindo o viés das análises seguintes e, assim, reduzir a quantidade de vitelo nas amostras complexas (proteínas) extraídas (Figura 6).

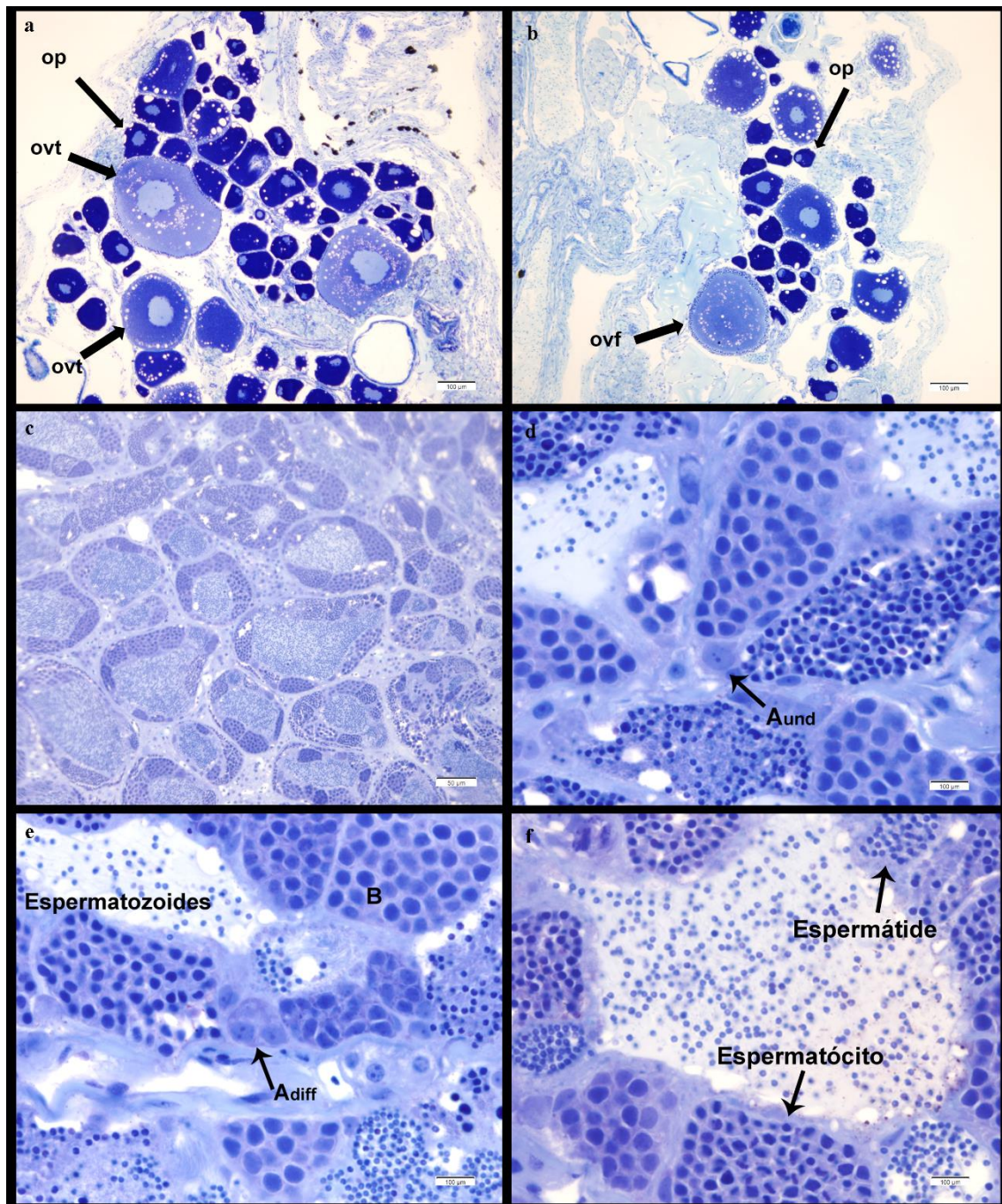


Figura 6. Corte histológico das gônadas femininas e masculinas. a e b, gônada de fêmea em estágio de desenvolvimento assincrônico, podendo observar oócitos pré-vitelogênicos, oócitos vitelogênicos intermediários e finais. c, d, e e f, gônada masculina apresentando os tipos celulares de maturação sexual (espermátogônias A (indiferenciadas/diferenciadas), espermátogônias B, espermátocitos (em divisão), espermátides (I, II e III)). **op** oócito pré-vitelogênico, **ovt** oócito vitelogênico intermediário, **ovf** oócito vitelogênico final, **Aund** espermátogônia A indiferenciada, **Adiff** espermátogônia A diferenciada e **B** espermátogônia tipo B.

6.2 Proteômica

A extração das proteínas se mostraram com boa qualidade e integridade, onde a curva (modelo linear $ax+b$) gerada pela quantificação de Bradford apresentou um R^2 de 0,98 (**Figura 7**). A partir dessa curva, as amostras foram quantificadas. Além disso, as amostras apresentaram boa integridade de acordo com a eletroforese em gel de poliacrilamida, SDS-PAGE.

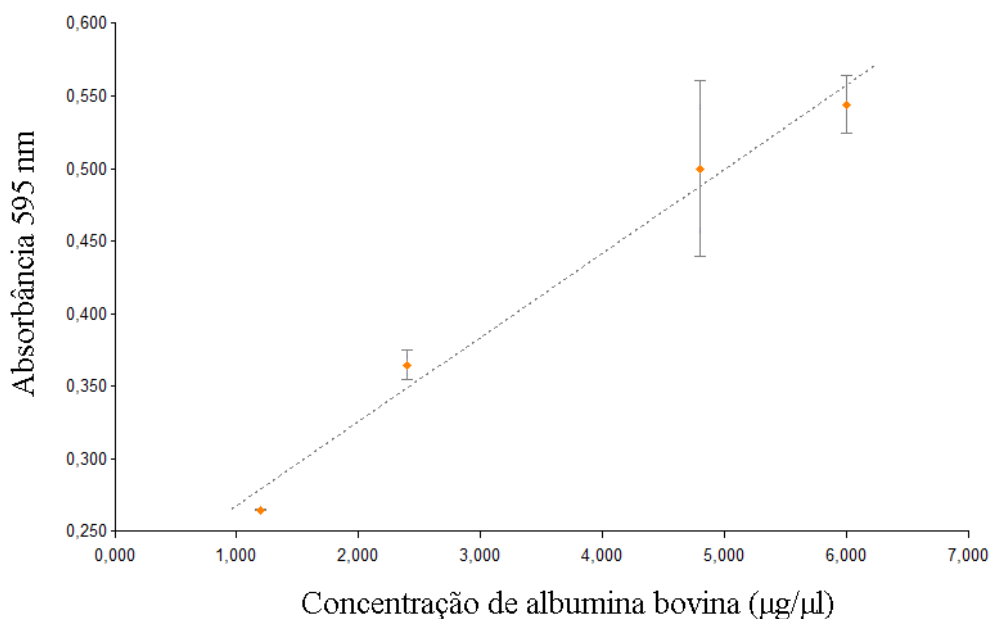


Figura 7. Análise quantitativa da extração proteica. Gráfico linear gerado para a quantificação por Bradford (Bradford 1976). $R^2 = 0.98$. Baseado na equação linear ($y=0,057x+0,21$) foram calculadas as concentrações das amostras substituindo a absorbância pela variável y .

6.3 Abundância proteica e mapas biológicos

Foram analisados os mapas biológicos das amostras resultantes da espectrometria de massas extraídas (em triplicata) de gônadas, encéfalo e músculo branco de indivíduos machos e fêmeas, comparando as amostras de fêmea B+ versus fêmea B- e machos B+ versus machos B- (Tabela 2 e Suplementar 1: Protein_Table_). Os mapas foram apresentados em dois formatos com maior/menor abundância e exclusividade B+ ou B- da proteína (Suplementar 1: Mapas Biológicos). Além disso, foram analisados os termos ontológicos enriquecidos relacionados (Suplementar 1: Protein_Table_).

6.3.1 Encéfalo

Em encéfalo foram extraídos 85 mapas biológicos (sem redundâncias) de acordo com a abundância de cada proteína (maior/menor) e presença em somente uma das condições (exclusivo B+ ou B-) (Tabela 3 e Suplementar 1: Mapas Biológicos). Nas fêmeas, os mapas biológicos apontaram 3 proteínas com menor abundância, 296 proteínas com maior abundância, 153 proteínas exclusivas de indivíduos B- e 208 proteínas exclusivas B+. Nos machos, os mapas biológicos foram obtidos a partir de 103 proteínas com menor abundância, 35 proteínas com maior abundância, 32 proteínas exclusivas de indivíduos B- e 297 proteínas exclusivas B+ (Suplementar 1: Protein_Table_encephalon_).

Tabela 3. Mapas biológicos com a presença de proteínas com abundância alterada. Foram comparadas amostras de fêmeas com (B+) e sem cromossomo B (B-). O mesmo foi realizado para machos.

Fêmea B+ versus Fêmea B-	n° de mapas
Menor e maior abundância em B+	40
Exclusivo B+ e B-	60
Macho B+ versus Macho B-	n° de mapas
Menor e maior abundância em B+	31
Exclusivo B+ e B-	45

Nesse conjunto de dados foi encontrada a via de sinalização da família *forkhead box O* (FOXO) em fêmeas, o FOXO é uma família de fatores de transcrição reguladores da expressão gênica em diferentes condições fisiológicas, os quais controlam o ciclo celular, o metabolismo de glicose, da apoptose e da resistência ao estresse oxidativo. Nessa via, foi encontrada a presença da proteína *serine/treonine kinase* (PLK) e a proteína *Sgk2b* (*serum/glucocorticoid regulated kinase 2b*) em amostras exclusivas B+ e B-, respectivamente. Além disso, foi identificada a proteína *Mps1* ou *TTK* (*dual-specificity protein kinase ttk*) presente nas amostras exclusivas B+, a qual é uma das principais quinases envolvidas na localização do cinetócoro e no *checkpoint* do conjunto do fuso (SAC) (Kaistha *et al.*, 2014) (Figura 8). Ainda nesta via, foi encontrada a proteína *Centrosomal protein 152* (Cep152), uma das principais proteínas relacionadas a duplicação do centríolo (Firat-Karalar e Stearns, 2014), as proteínas sinalizadoras da divisão celular, *Ywhae1* (*14-3-3 protein epsilon*) e *Ywhaz* (*tyrosine 3-*

monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide) em amostras exclusivas B+ e 11 proteínas putativas para 14-3-3 em amostras com abundância para B+. Além disso, foi identificado a proteína Mcm6 do complexo MCM exclusivas B-. Em relação as proteínas descritas a cima, foram obtidos os termos ontológicos enriquecidos para *Cell Cycle* (KEGG:04110), *Cell Cycle Checkpoints* (REAC:R-DRE-69620), *G2/M Checkpoints* (REAC:R-DRE-69481), *G2/M DNA damage checkpoint* (REAC:R-DRE-69473), *Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex* (REAC:R-DRE-75035).

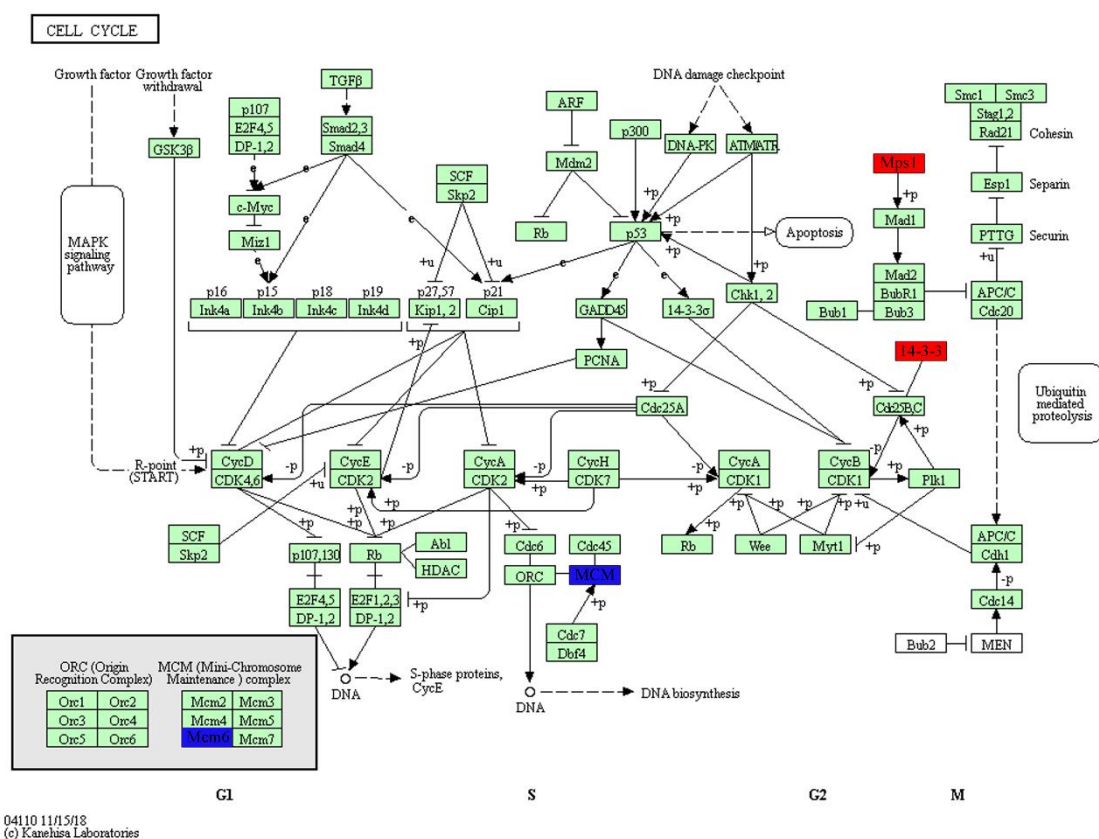


Figura 8. Mapa do ciclo celular. Proteína Mps1 (ou TTK) fosforila a proteína Mad1 para a formação inicial do complexo proteico de *check point* mitótico (MCC) da transição da fase G2/M. As proteínas sinalizadoras Ywhae1 e Ywhaz presentes na fase G2 do ciclo celular. Em vermelho a proteínas exclusivas B+, em azul proteínas exclusivas B-. Mapa extraído do KEGG, id: dre04110.

Outro mapa encontrado para proteínas exclusivas de amostras B+ em fêmeas foi a via de sinalização do GnRH, na qual também foi enriquecido. Nessa via foi encontrada a presença da proteína *calcium/calmodulin-dependent protein kinase* (CAMK) que atua

como fator de transcrição no promotor da subunidade LH β para regular a expressão do hormônio luteinizante (LH) (Pawson e McNeilly, 2005) (Figura 9).

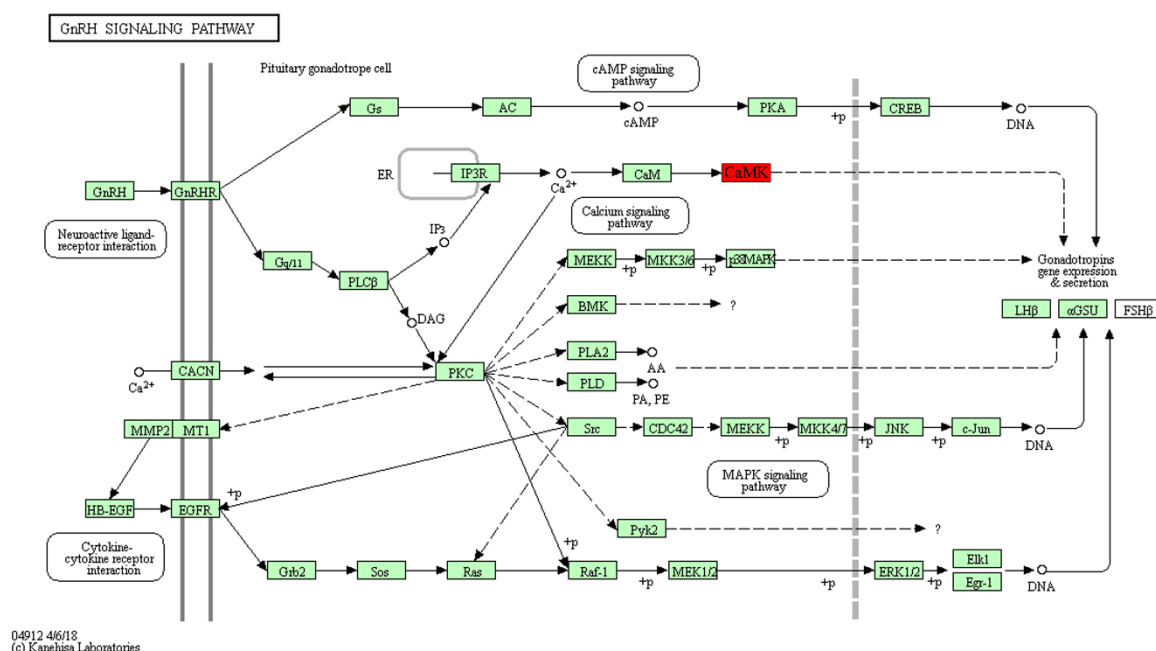


Figura 9. Mapa biológico da sinalização do GnHR. Em vermelho a proteína cálcio/calmodulin-dependente protein kinase (CaMK) presente nas amostras exclusiva B+. Mapa extraído do KEGG, id dre04920.

Em cérebro de fêmeas ainda foram encontradas, principalmente, proteínas com maior abundância relacionadas a mapas metabólicos básicos como glicólise/gliconeogênese, metabolismo do piruvato, arginina e prolina, pirimidina, cisteína e metionina, entre outros, no qual identificamos 12 enzimas com seus respectivos números de acesso (EC: *Enzyme Commission number*) (Tabela 4).

Tabela 4. Número de acesso das enzimas encontradas para amostras de cérebro de fêmeas B+.

Descrição	n° de EC
<i>Fructose-bisphosphate aldolase</i>	4.1.2.13
<i>Phosphopyruvate hydratase</i>	4.2.1.11
<i>Triose-phosphate isomerase</i>	5.3.1.1
<i>Glyceraldehyde-3-phosphate dehydrogenase</i>	1.2.1.12
<i>Malate dehydrogenase</i>	1.1.1.37
<i>Isocitrate dehydrogenase (NADP(+))</i>	1.1.1.42
<i>L-lactate dehydrogenase</i>	1.1.1.27

Anaphase-promoting complex subunit 1 (APC/C). Além disso, foram obtidos o enriquecimento funcional para o termo *chromatin* (GO:0000785, *p-value*: 0,043) com abundância nas amostras B+ para as proteínas Smarcc1a (*SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily c member 1a*), h3f3a e h3f3b.1 (Histonas H3), si:dkey-108k21.24 e a H4 (Histona H4).

6.3.2 Músculo

Em músculo foram extraídos 65 mapas biológicos (sem redundâncias) de acordo com a abundância de cada proteína (maior/menor) e presença em somente uma das condições (exclusivo B+/B-) (Tabela 5 e Suplementar 1: Mapas Biológicos). Nas fêmeas foram prospectados os mapas biológicos para: 176 proteínas com menor abundância, 23 proteínas com maior abundância, 214 proteínas exclusivas de indivíduos B- e 132 proteínas exclusivas B+. Nos machos foram prospectados os mapas biológicos para: 165 proteínas com menor abundância, 36 proteínas com maior abundância, 194 proteínas exclusivas de indivíduos B- e 109 proteínas exclusivas B+ (Suplementar 1: Protein_Table_muscle_).

Tabela 5. Mapas biológicos com a presença de proteínas com abundância alterada. Foram comparadas amostras de fêmeas com (B+) e sem cromossomo B (B-). O mesmo foi realizado para machos.

Fêmea B+ versus Fêmea B-	n° de mapas
Menor e maior abundância em B+	24
Exclusivo B+ e B-	33
Macho B+ versus Macho B-	n° de mapas
Menor e maior abundância em B+	22
Exclusivo B+ e B-	42

Amostras de músculo de macho B+ apresentaram mapas metabólicos básicos com ênfase na via de sinalização do MAPK, na qual foram encontradas a enzima *tyrosine-protein phosphatase non-receptor type 5* (PTP, EC:3.1.3.48), repressor das proteínas *mitogen-activated protein kinase 1/3* (erk), *mitogen-activated protein kinase 8/9/10* (*c-Jun N-terminal kinase*) e a proteína *p38 MAP kinase* (EC:2.7.11.24), as quais são quinases

importantes na regulação da proliferação, diferenciação, inflamação, apoptose, ciclo celular e via da sinalização do p53 (Figura 11).

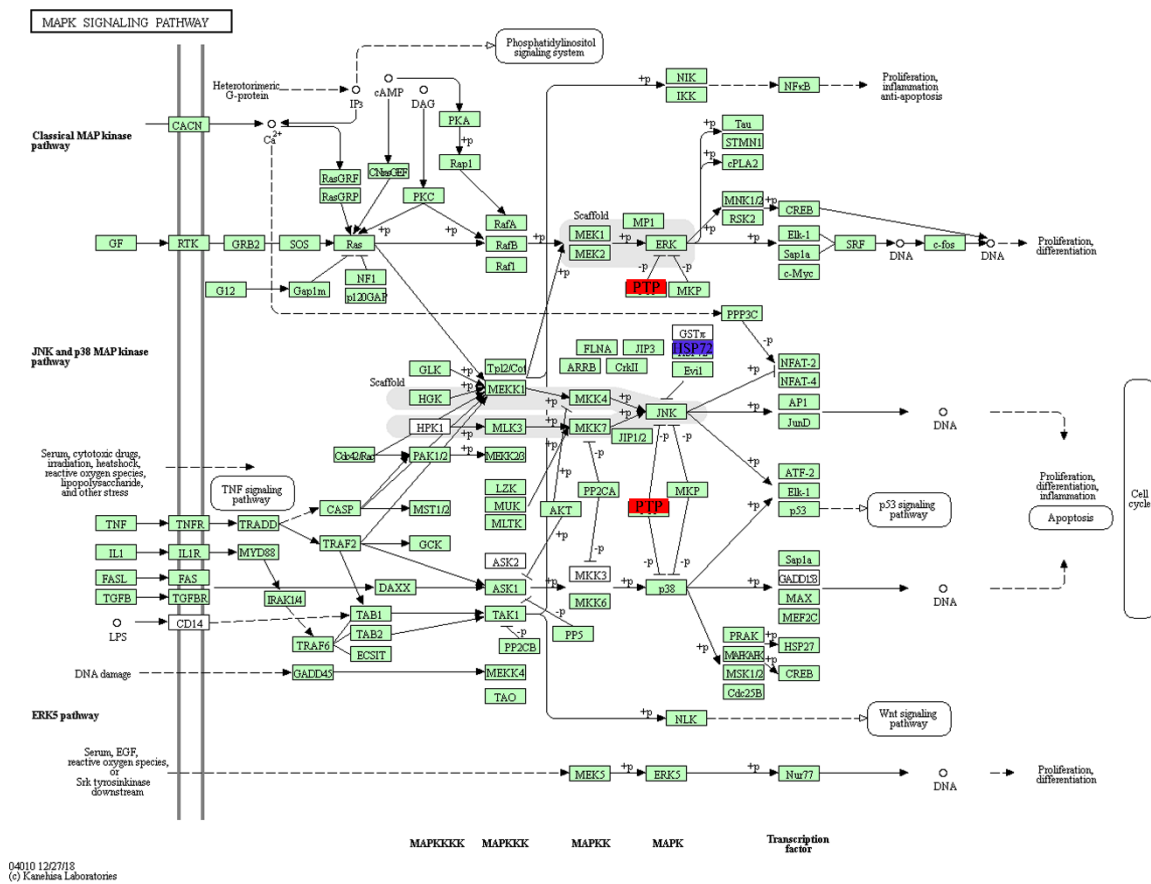


Figura 11. Mapa biológico da sinalização MAPK. Em vermelho a proteína com maior abundância em músculo de machos: enzima tyrosine-protein phosphatase non-receptor type 5 (PTP, EC:3.1.3.48). Em azul a proteína HSP72. Mapa extraído do KEGG, id dre04010.

6.3.3 Gônada

Nas gônadas foram extraídos no total 67 mapas biológicos (sem redundâncias) de acordo com a abundância de cada proteína (maior/menor) e presença em somente em uma das condições (exclusivo B+/B-) (Tabela 6 e Suplementar 1: Mapas Biológicos). Nas fêmeas foram prospectados os mapas biológicos de acordo com as 41 proteínas com menor abundância, 126 proteínas com maior abundância, 118 proteínas exclusivas de indivíduos B- e 106 proteínas exclusivas B+. Nos machos, os mapas encontrados foram de acordo com 131 proteínas com menor abundância, 127 proteínas com maior abundância, 185 proteínas exclusivas de indivíduos B- e 110 proteínas exclusivas B+ (Suplementar 1: Protein_Table_gonad_).

Tabela 6. Mapas biológicos com a presença de proteínas com abundância alterada. Foram comparadas amostras de fêmeas com (B+) e sem cromossomo B (B-). O mesmo foi realizado para machos.

Fêmea B+ versus Fêmea B-	n° de mapas
Menor e maior abundância em B+	25
Exclusivo B+ e B-	39
Macho B+ versus Macho B-	n° de mapas
Menor e maior abundância em B+	38
Exclusivo B+ e B-	44

Nas gônadas de fêmeas foi novamente identificado a via de sinalização da família *forkhead box O* (FOXO). Entretanto, com a presença das proteínas quinase AMP (AMPK) somente nas amostras B- e o receptor do fator de crescimento epidérmico (EGFR) nas amostras B+.

Outra via relacionada ao receptor EGFR é a via de sinalização do GnRH. O receptor do fator de crescimento epidérmico sofre transativação e leva a ativação de proteínas quinases ativadas por mitógeno (MAPKs), incluindo quinase regulada por sinal extracelular (ERK), quinase JUN N-terminal (JNK) e p38 MAPK (Figura 12).

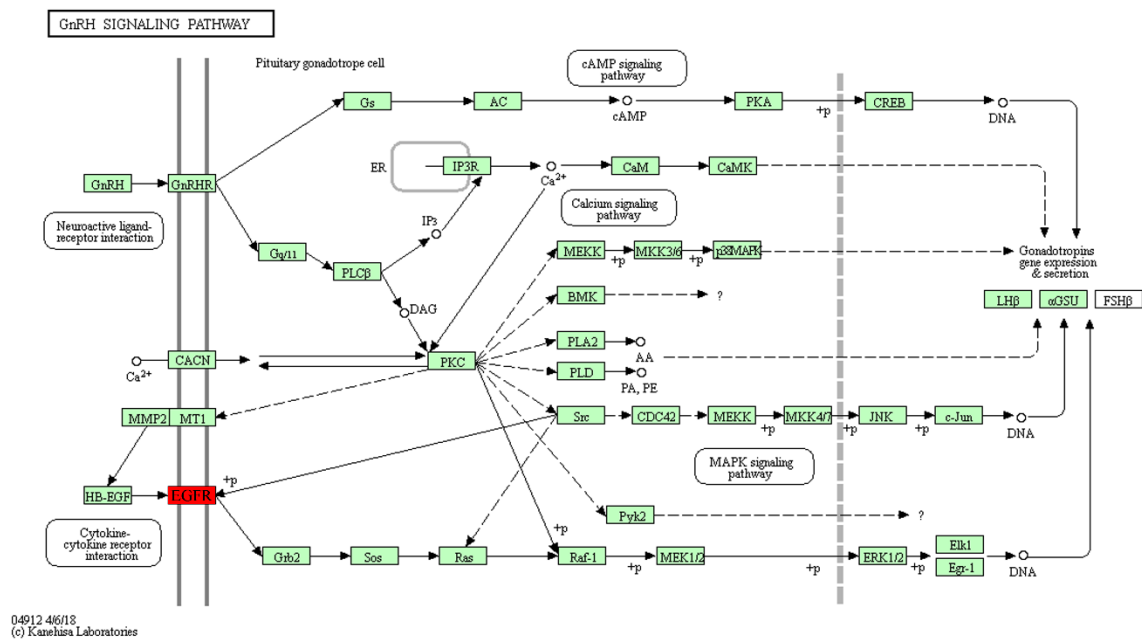


Figura 12. Mapa biológico da sinalização do GnRH. Em vermelho: receptor do fator de crescimento epidérmico (EGFR) presente nas amostras exclusiva B+. Mapa extraído do KEGG, id dre04920.

Nas gônadas de macho, foram encontradas a via de transporte de RNA com as proteínas *Elongation factor 1-alpha 1* (eEF1A), *Eukaryotic initiation fator 4A* e *4A3* (eIF4a e eIF4A3) e a proteína *Poly(A)-binding* (PABP) com baixa abundância ou somente nas amostras B- (proteínas exclusivas em amostras B-) (Figura 13).

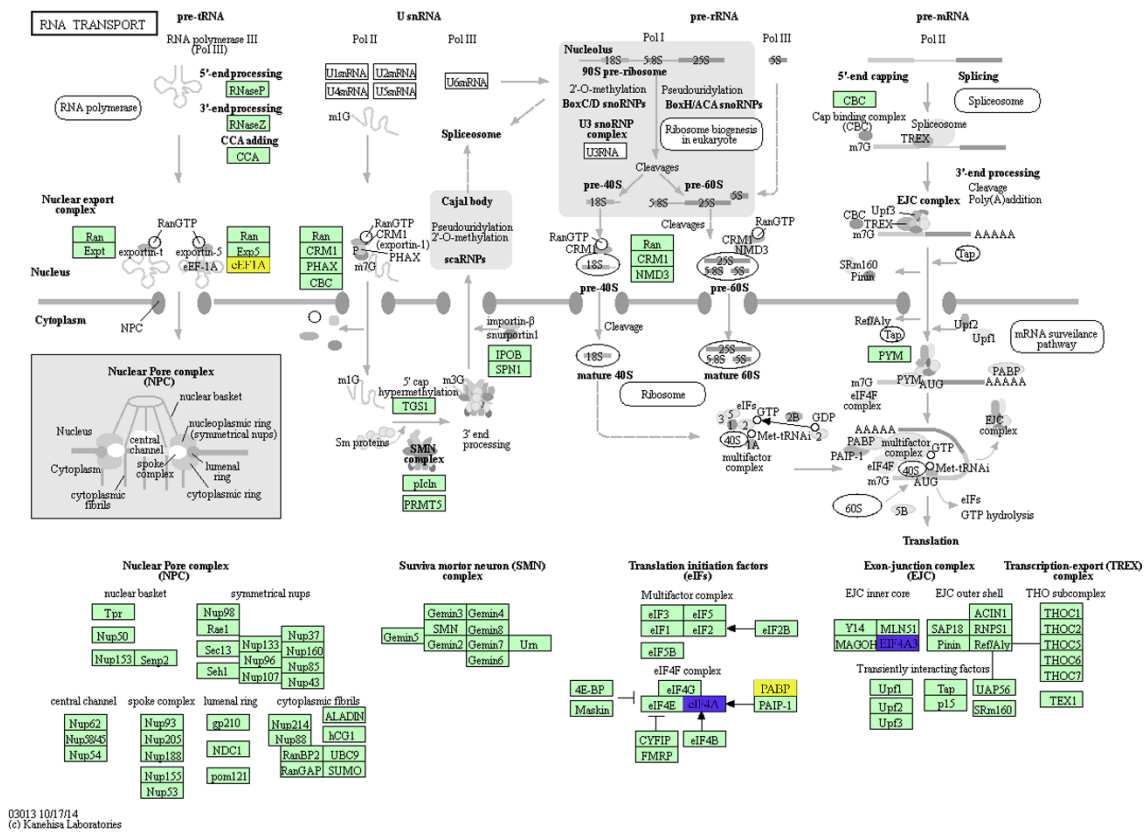


Figura 13. Mapa biológico do transporte de RNA. Em azul: *Eukaryotic initiation fator 4A* e *4A3* (eIF4a e eIF4A3), presente nas amostras exclusiva B- e em amarelo as proteínas *Elongation factor 1-alpha 1* (eEF1A) e proteína *Poly(A)-binding* (PABP) com menor abundância em amostras B+. Mapa extraído do KEGG, id dre03013.

6.4 Transcriptoma

As bibliotecas de gônada, cérebro e músculo branco geraram 841.712.040 *reads*, 999.793.794 *reads* e 843.420.662 *reads*, respectivamente. A montagem dos transcritos gerou 487.821 sequências com comprimento médio de 893 pb.

Após a montagem *de novo*, remoção de redundâncias e alinhamento no genoma da espécie *Metriaclima zebra*, a montagem final gerou 261.295 transcritos. Desses, a anotação funcional apresentou 27.526 genes (Suplementar 2: Anotação-Tr.gff3). Para as análises de entropia de *Shannon*, expressão diferencial e para as redes de interação, a fim

de englobar uma análise exploratória dos dados, foram utilizados os 261.295 transcritos como referência para o alinhamento das bibliotecas filtradas e para as contagens dos *reads* (transcritos: “transcr_”). Após as contagens, os transcritos foram normalizados segundo o programa DESeq2 (Anders *et al.* 2014; Love *et al.*, 2014). A entropia de *Shannon* seguiu a metodologia de Pinhal *et al.* (2018), a expressão diferencial seguiu um *pipeline* do próprio programa (DESeq2), e a análise das bibliotecas utilizando a análise de componente principal (PCA) e clusterização hierárquica. A rede de interação entre as redes de co-expressão e interação proteína-proteína apresentou uma distribuição de *degree* obedecendo a lei de potência (RHO=0.92 e RHO=0.94, respectivamente), representando aspectos reais de um sistema biológico.

6.4.1 Entropia de Shannon

A entropia de *Shannon*, calculada para as triplicatas, apresentou uma pequena diversidade de acordo com o tecido (encéfalo, músculo e gônada) analisado (Figura 14). Entretanto, não foi observado diferença entre as bibliotecas comparando a presença e ausência do cromossomo B. Nesse caso, a entropia demonstrou o perfil global dos valores de expressão dos transcritos identificados, a qual não apresentou flutuação discrepante e não foi possível observar especificidade referente ao cromossomo B (Figura 14).

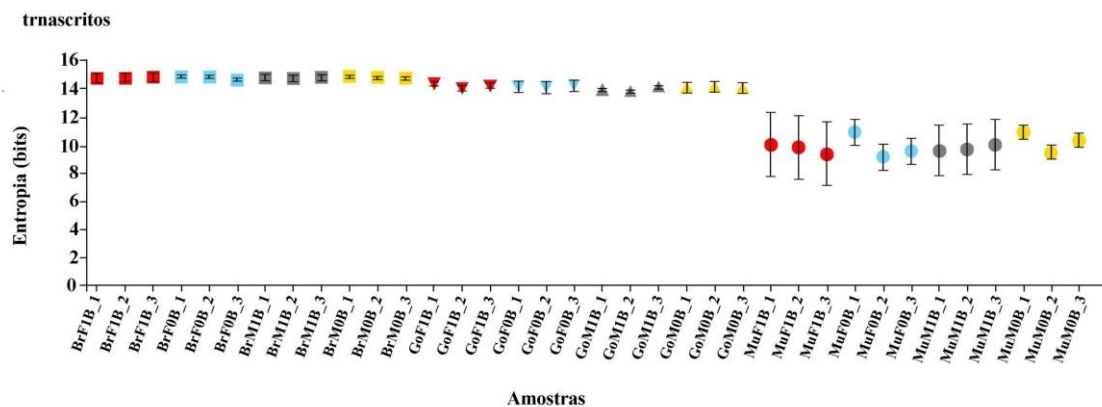


Figura 14. Análise de entropia dos transcritos. Entropia de *Shannon* foi utilizado para analisar a diversidade dos valores de expressão dos transcritos em cada biblioteca do transcriptoma. Em quadrado: amostras de encéfalo, triângulo: amostras de gônada; círculo: amostras de músculo; vermelho: fêmeas com cromossomo B; azul: fêmeas sem cromossomo B; cinza: macho com cromossomo B; amarelo: macho sem cromossomo B. A barra de erro representa o intervalo de confiança entre a triplicatas.

6.4.2 Expressão diferencial

A expressão diferencial (DE) foi realizada separadamente de acordo com o tecido e sexo analisado. Além disso, foram selecionados alguns genes candidatos com expressão diferencial e prospectados no genoma, verificando a cobertura de *reads* da montagem com referência no genoma de *M. zebra*, no qual é possível inferir se a região analisada pode estar presente no cromossomo B (Valente *et al.*, 2014).

6.4.2.1 Encéfalo - macho

Nos machos, as bibliotecas analisadas apresentaram um resultado positivo para análise do componente principal (PCA), na qual foi observado um padrão entre as bibliotecas analisadas. O mesmo foi observado na clusterização hierárquica da expressão (Figura 15). Foram encontrados no total 279 transcritos diferencialmente expressos, sendo anotados 17 transcritos *down regulated* e 15 transcritos *up regulated* (Figura 15, Suplementar 3: Table_S1, S3). A clusterização dos termos ontológicos enriquecidos para os transcritos *down regulated* foram: regulação circadiana da expressão gênica (GO:0032922), atividade NAD⁺ ADP ribosiltransferase (GO:0003950), espaço extracelular (GO:0005615) e complexo proteico-lipídico (GO:0032994) (Suplementar 3: Table_S2). Em relação aos transcritos *up regulated* não foram encontrados termos enriquecidos.

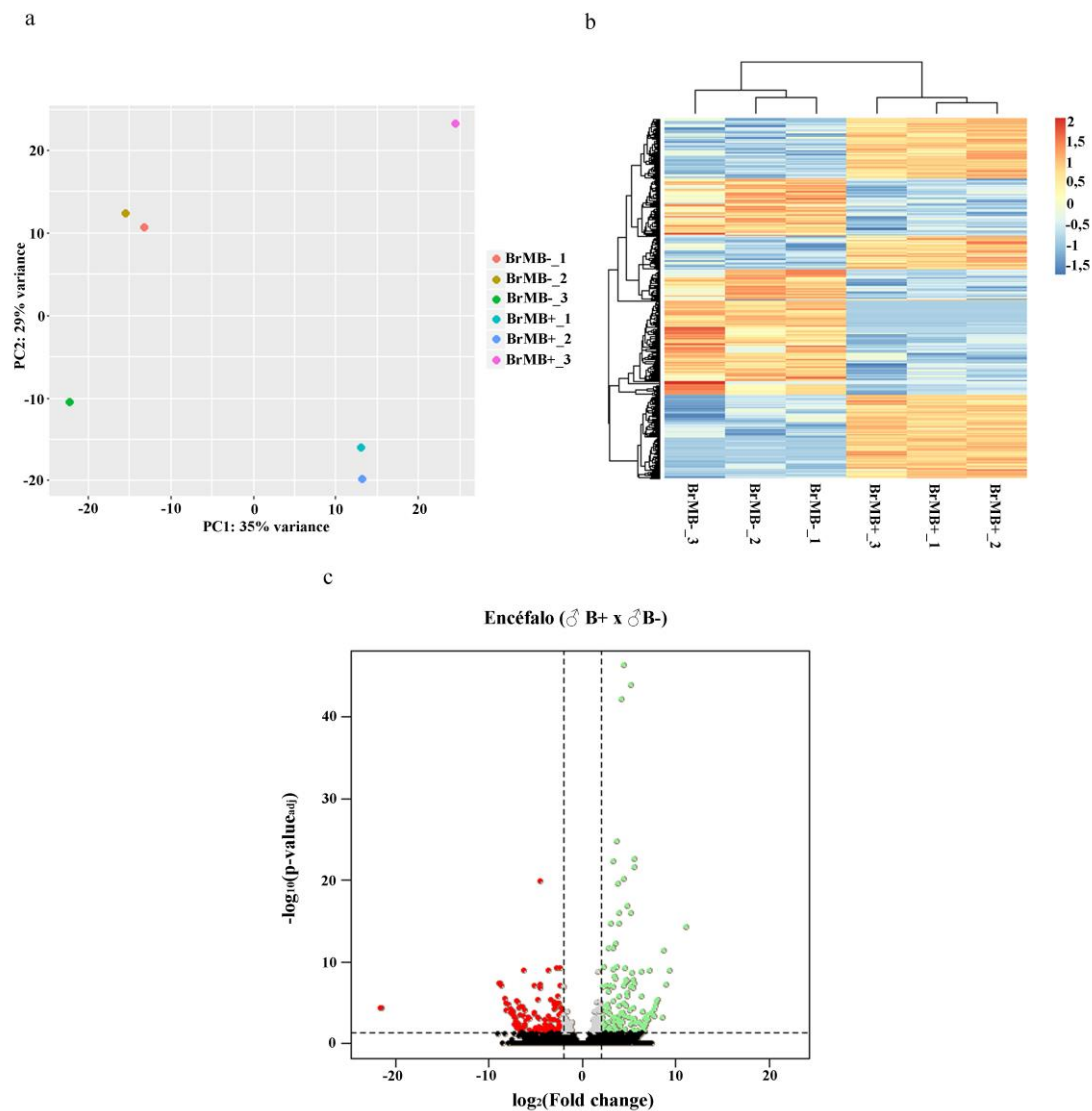


Figura 15. Análise das bibliotecas de encéfalo de macho. Em (a) análise da componente principal (PCA). De acordo com o dado obtido, é possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas: BrMB- (amostras sem cromossomo B) e BrMB+ (amostras com cromossomo B). Em (b) o *heat map* da clusterização hierárquica dos dados de expressão. A escala é a distância da centralização e a padronização de cada linha separadamente pelo Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior as bibliotecas são representadas com as siglas: BrMB- (amostras sem cromossomo B) e BrMB+ (amostras com cromossomo B). Em (c) *volcano plot* dos transcritos diferencialmente expressos em encéfalo de macho. Em vermelho transcrito com baixa expressão e em verde transcritos com alta expressão. Eixo das abscissas o valor de expressão em $\log_2(\text{fold change})$ e no eixo das ordenadas $-\log$ do $p\text{-value}_{adj}$.

Apesar dos poucos termos biológicos enriquecidos, dentre os genes *up regulated* são enfatizados os genes *secretagogin* (Scgn), *aurora kinase* (Aurka), *neural proliferation*

differentiation and control protein 1 (Npdc1). Dentre os genes *down regulated*, são destacados os genes *poly ADP-ribose polymerase 14* (Parp14) e o gene *nuclear receptor subfamily 1, group D, member 1* (Nr1d1; Rev-erba).

6.4.2.2 Encéfalo - fêmea

Nas fêmeas, a PCA e a clusterização hierárquica da expressão dos transcritos apresentaram uma biblioteca de indivíduo B- com alta discrepância, diferenciando tanto do padrão de amostra B- como B+, inviabilizando a análise (Figura 16: a e b). Devido a isso, a amostra foi removida e refeita a PCA e a clusterização hierárquica da expressão (Figura 16: c e d). Deste modo, a expressão diferencial foi realizada comparando as amostras B- (duplicata) contra a amostras B+ (triplicata). A expressão diferencial apresentou no total 488 transcritos, sendo anotados 105 transcritos *down regulated* e 25 transcritos *up regulated* (Suplementar 3: Table_S7 e S10). Em relação aos transcritos *down regulated* a clusterização dos termos ontológicos enriquecidos foram: *regulation and immune system process* (GO:0002682 e GO:0006955), *response to virus* (GO:0009615) e *transcription regulatory region DNA binding* (GO:0044212) (Suplementar 3: Table_S8). Para os transcritos *up regulated* foram obtidos enriquecimento para o termo *Ndc80 complex* (GO:0031262) (Suplementar 3: Table_S11).

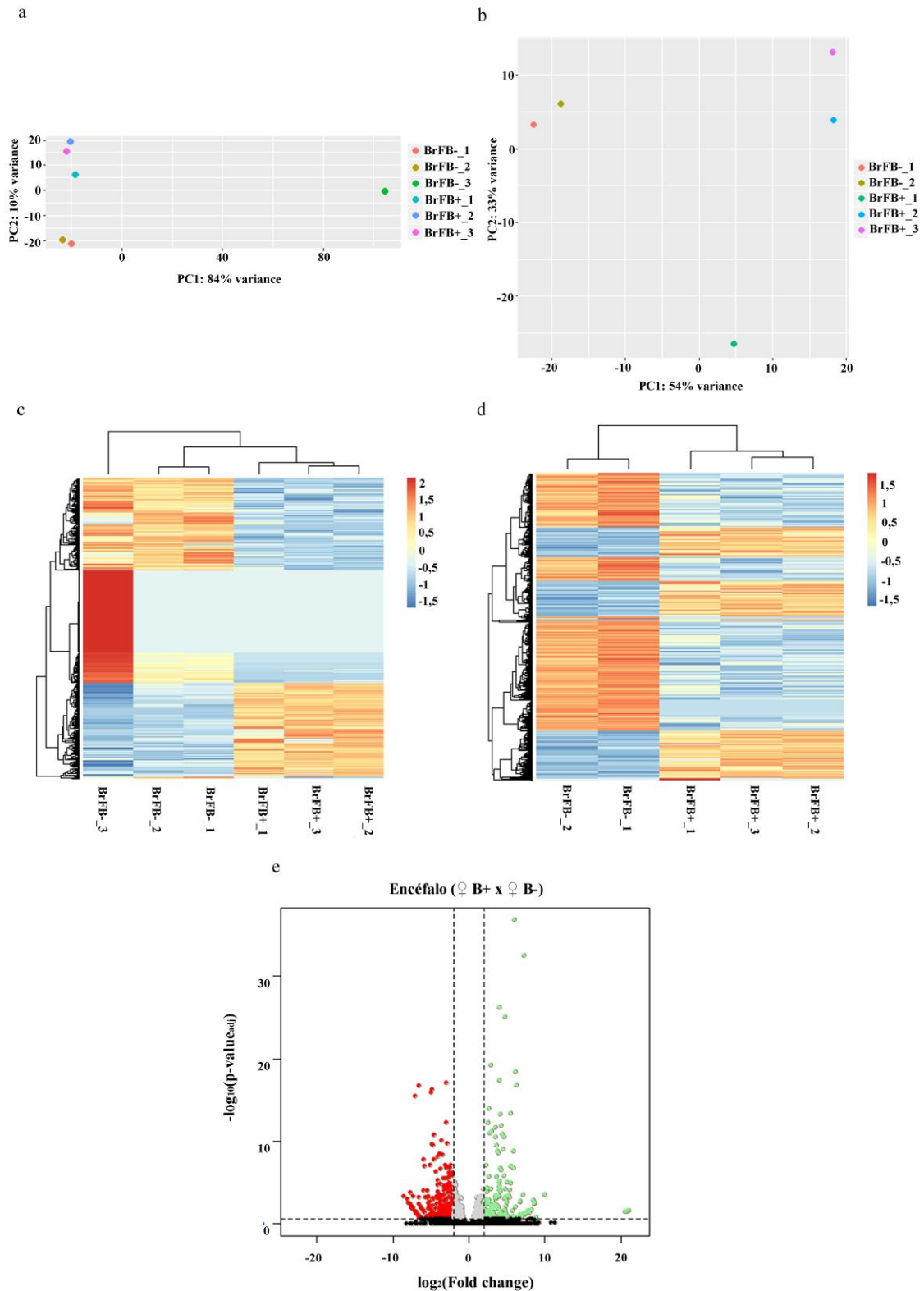


Figura 16. Análise das bibliotecas de encéfalo de fêmea. Em (a) e (b) análise da componente principal (PCA) e (c) e (d) o *heat map* da clusterização hierárquica dos dados de expressão. De acordo com o dado obtido, em (a) e (c) uma biblioteca de indivíduo B- apresenta uma diferença tanto do padrão das amostra B- como B+, inviabilizando a análise. Devido a isso, a amostra foi removida e refeita a PCA e o *heat map* da clusterização hierárquica da expressão (b) e (d). É possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas:

BrFB- (amostras sem cromossomo B) e BrFB+ (amostras com cromossomo B). Em (b) A escala é a distância da centralização e a padronização de cada linha separadamente pelo Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior, as bibliotecas são representadas com as siglas: BrFB- (amostras sem cromossomo B) e BrFB+ (amostras com cromossomo B). Em (c) *volcano plot* dos transcritos diferencialmente expressos em encéfalo de fêmea. Em vermelho transcrito com baixa expressão e em verde transcritos com alta expressão. Eixo das abscissas o valor de expressão em $\log_2(\text{fold change})$ e no eixo das ordenadas $-\log$ do $p\text{-value}_{adj}$.

Em relação aos transcritos *up regulated* foram identificados genes pertencentes a via do ciclo celular, como a: *tubulin beta 1* (*tubb1*), *aurora kinase* (*aurka*) e a proteína do *kinetochore* (*nuf2*). Também foram identificados a presença dos genes *B-cell lymphoma 6 protein homolog* (*bcl6*) *ezinc fingers* que atuam como fatores de transcrição regulando a expressão gênica e os genes relacionados ao termo ontológico de percepção a visão: *melanopsina* e *pinolsina*. Outro gene identificado foi o *microtubule-associated protein tau* (*mapta*). Os genes *down regulated* identificados foram *signal transducer and activator of transcription 1 and 3* (*stat1* e *stat3*), *poly ADP-ribose polymerase 14* (*parp14*), *polyubiquitini* e *RAS protein activator like-3* (*rasal3*).

6.4.2.3 Músculo – macho

Nos machos, as bibliotecas analisadas apresentaram um resultado positivo para a PCA e a clusterização hierárquica da expressão (Figura 17). Foram encontrados no total 38 transcritos diferencialmente expressos, sendo anotados 12 transcritos *down regulated* e 6 transcritos *up regulated* (Suplementar 3: *Table_S13* e *S16*). A clusterização dos termos ontológicos enriquecidos para os transcritos *down regulated* foram os processos biológicos: *proteolysis* (GO:0006508), *primary metabolism* (GO:0044238), *organonitrogen compound metabolism* (GO:1901564) e *nitrogen compound metabolism* (GO:0006807). As funções moleculares enriquecidas foram a *endopeptidase activity* (GO:0004175), *hydrolase activity* (GO:0016787) e *catalytic activity* (GO:0003824) (Suplementar 3: *Table_S14*). Os transcritos *up regulated* não apresentaram enriquecimento de termos.

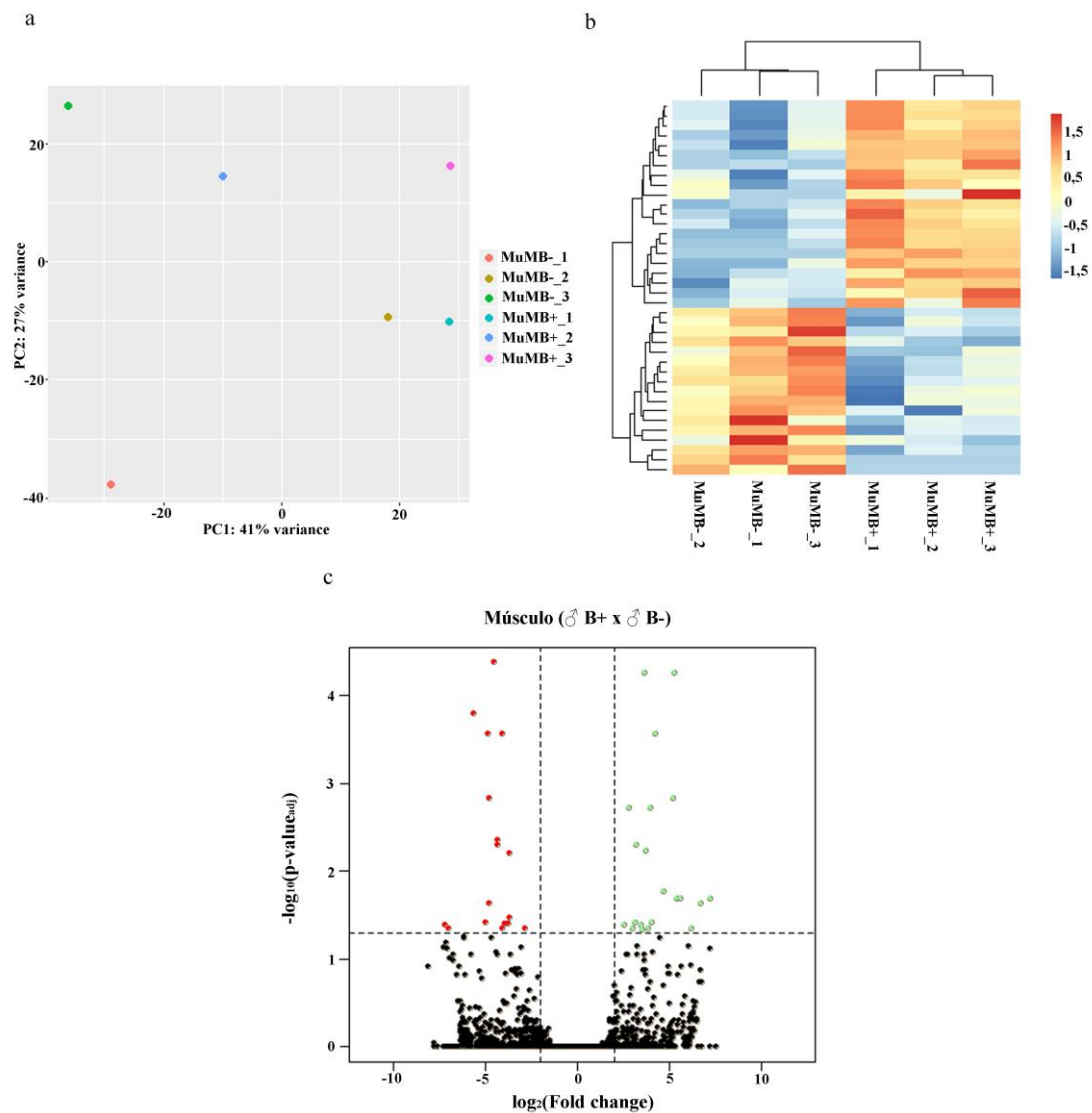


Figura 17. Análise das bibliotecas de músculo de macho. Em (a) análise da componente principal (PCA). De acordo com o dado obtido, é possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas: MuMB-_ (amostras sem cromossomo B) e MuMB+_ (amostras com cromossomo B). Em (b) o *heat map* da clusterização hierárquica dos dados de expressão. A escala é distância da centralização e padronização de cada linha separadamente pelo Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior, as bibliotecas são representadas com as siglas: MuMB-_ (amostras sem cromossomo B) e MuMB+_ (amostras com cromossomo B). Em (c) *volcano plot* dos transcritos diferencialmente expressos em músculo de macho. Em vermelho transcrito com baixa expressão e em verde transcritos com alta expressão. Eixo das abscissas o valor de expressão em $\log_2(\text{fold change})$ e no eixo das ordenadas $-\log$ do $p\text{-value}_{adj}$.

Em relação aos genes *down regulated*, é enfatizado a presença da enzima *serine/threonine-protein kinase pim-2* (pim2). Já para os genes *up regulated* são

destacados: *stanniocalcin-2* (stc2), *aurora kinase* (aurka) e *neural proliferation differentiation and control protein 1* (npdc1).

6.4.2.4 Músculo – fêmea

Nesse conjunto de dados a PCA apresentou uma biblioteca de indivíduo B- com diferença em relação às outras (Figura 18). Entretanto, para não perder força estatística (triplicatas) e pelo fato do resultado da clusterização hierárquica da expressão não apresentar “discrepância”, como encontrado nas bibliotecas de encéfalo de fêmeas, foram utilizadas todas as bibliotecas (triplicatas) para a análise de expressão diferencial. Foram obtidos no total 55 transcritos diferencialmente expressos, sendo 8 transcritos *down regulated* e 20 transcritos *up regulated* anotados (Suplementar 3: *Table_S19* e *Table_S22*). A clusterização dos termos ontológicos enriquecidos para os genes *down regulated* foram os processos biológicos: *mesenchyme migration* (GO:0090131); componente celular: *filopodium* (GO:0030175) e *cell body* (GO:0044297) (Suplementar 3: *Table_S20*). Os principais termos ontológicos para genes *up regulated* foram o processo biológico: *oxygen transport* (GO:0015671); função molecular: *oxygen transporter activity* (GO:0005344) e *oxygen binding* (GO:0019825); componente celular: *hemoglobin complex* (GO:0005833) (Suplementar 3: *Table_S23*).

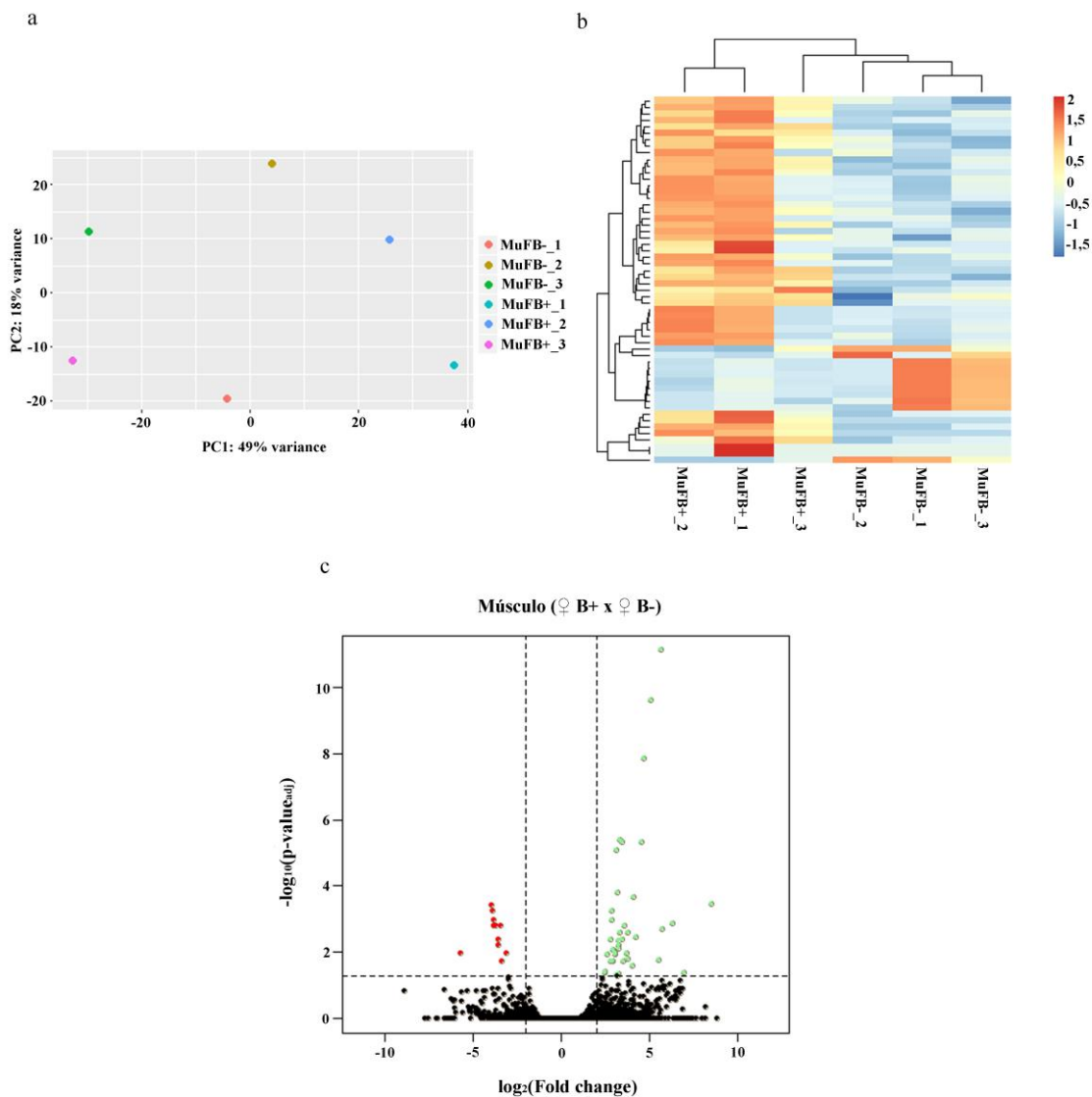


Figura 18. Análise das bibliotecas de músculo de fêmea. Em (a) análise da componente principal (PCA). De acordo com o dado obtido, é possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas: MuFB-_ (amostras sem cromossomo B) e MuFB+_ (amostras com cromossomo B). Em (b) o *heat map* da clusterização hierárquica dos dados de expressão. A escala é distância da centralização e padronização de cada linha separadamente pelo Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior, as bibliotecas são representadas com as siglas: MuFB-_ (amostras sem cromossomo B) e MuFB+_ (amostras com cromossomo B). Em (c) *volcano plot* dos transcritos diferencialmente expressos em músculo de fêmea. Em vermelho transcrito com baixa expressão e em verde transcritos com alta expressão. Eixo das abscissas o valor de expressão em $\log_2(\text{fold change})$ e no eixo das ordenadas $-\log$ do $p\text{-value}_{adj}$.

Nesse conjunto de dados foram identificados novamente a expressão dos genes putativos da aurka e tubb1, adicionalmente, o gene *G2/mitotic-specific cyclin-B1(ccnb1)*.

6.4.2.5 Gônada - macho

Nesse grupo foi obtido um resultado positivo para PCA e um padrão de expressão entre as bibliotecas comparadas visualizadas pela clusterização hierárquica (Figura 19). A DE identificou 110 transcritos, sendo 44 transcritos *down regulated* e 66 transcritos *up regulated*. Desses transcritos, foram anotados 8 genes *down regulated* e 11 genes *up regulated* (Suplementar 3: *Table_S25 e Table_S28*). Desses, considerando resultados já obtidos, foram encontrados transcritos *up regulated* para genes putativos para *kinetochore(nuf2)*, *tubulin beta 1 (tubb1)* e *BUB3-interacting e GLEBS motif-containing protein (znf207)*.

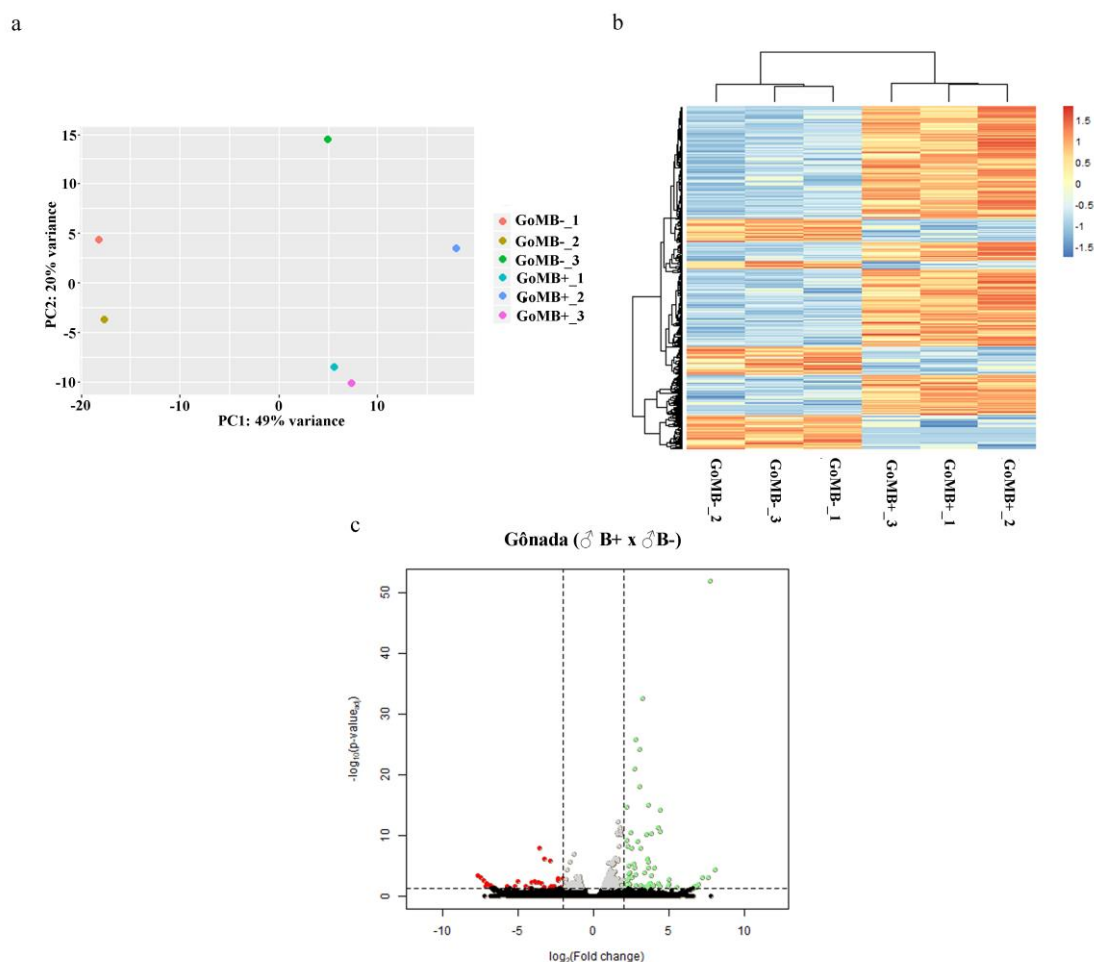


Figura 19. Análise das bibliotecas de gônada de macho. Em (a) análise da componente principal (PCA). De acordo com o dado obtido, é possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas: GoMB- (amostras sem cromossomo B) e GoMB+ (amostras com cromossomo B). Em (b) o *heat map* da clusterização hierárquica dos dados de expressão. A escala é distância da centralização e padronização de cada linha separadamente pelo

Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior, as bibliotecas são representadas com as siglas: GoMB- (amostras sem cromossomo B) e GoMB+ (amostras com cromossomo B). Em (c) *volcano plot* dos transcritos diferencialmente expressos em gônada de macho. Em vermelho transcrito com baixa expressão e em verde transcritos com alta expressão. Eixo das abscissas o valor de expressão em $\log_2(\text{fold change})$ e no eixo das ordenadas $-\log$ do $p\text{-value}_{adj}$.

A clusterização dos termos ontológicos enriquecidos dos genes *down regulated* foram: função molecular *stem cell factor receptor binding* (GO:0005173) e *potassium ion antiporter activity* (GO:0022821) (Suplementar 3: *Table_S26*). O termo ontológico enriquecido para os genes *up regulated* foi o componente celular *Ndc80 complex* (GO:0031262) (Suplementar 3: *Table_S29*).

A análise de cobertura genômica apresentou um cluster, incluindo 2 genes, *vip* e *fbox only protein 5 like* (*fbox5*) (Figura 20). Destes, o gene *vip* se encontra *up regulated*, embora esse gene seja geralmente expresso em cérebro de teleósteo atuando na liberação de hormônios, atuando nos sistemas nervoso e imune (Nam *et al.* 2009). Apesar de o gene *fbox5* não apresentar expressão diferencial, este possui um importante papel na regulação do ciclo celular e transdução de sinal, quando formado o complexo Skp1, Cdc53 e a proteína F-box (SCF) (Craig e Tyers. 1999).

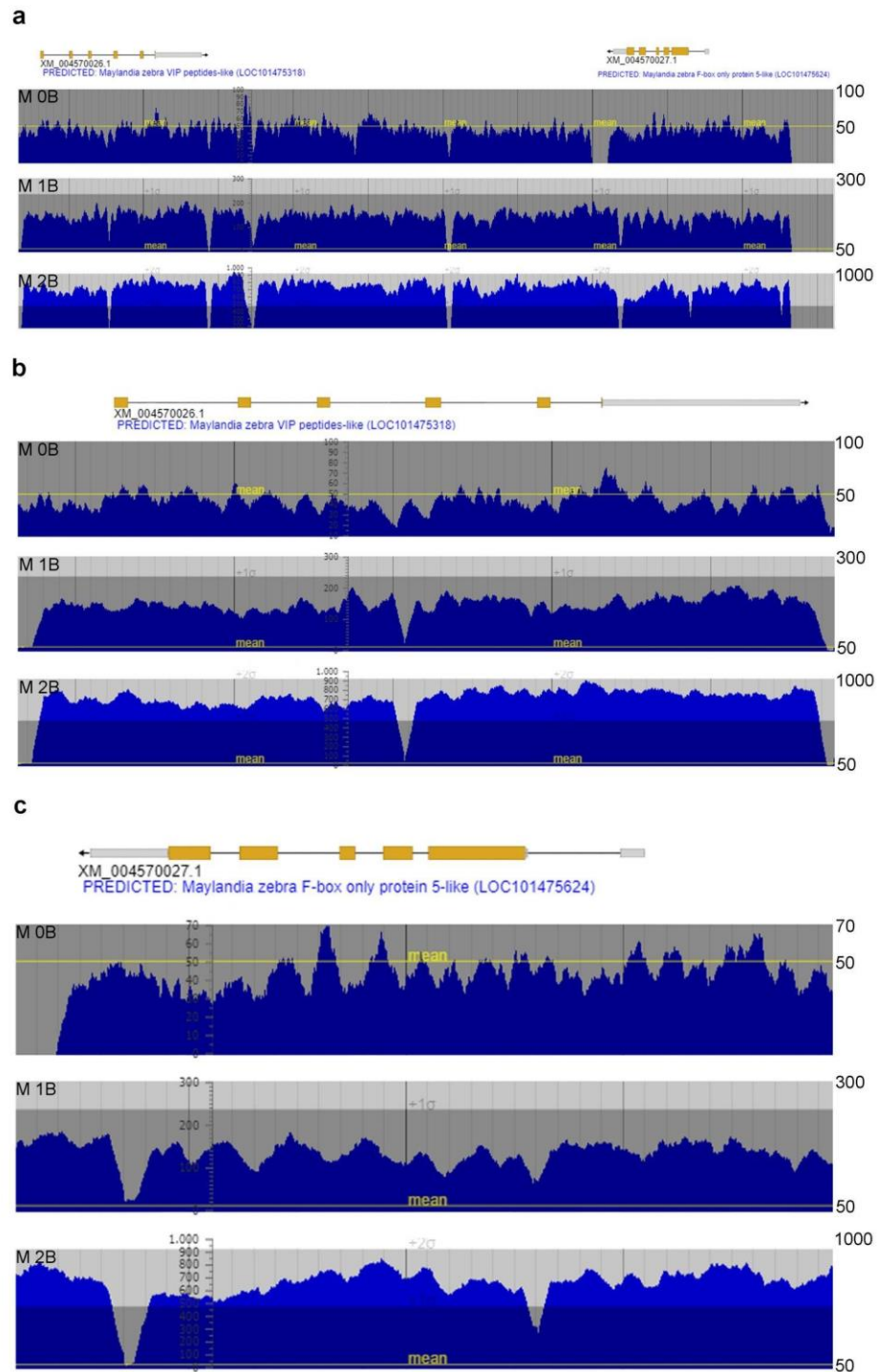


Figura 20. Análise de cobertura genômica. Região com alta cobertura gênica demonstrando possíveis regiões presentes no cromossomo B. Em (a) genoma 0B de *A. latifasciata*, em (b) genoma 1B de *A. latifasciata* e em (c) genoma 2B de *A. latifasciata*. Cobertura gênica calculada de acordo com Valente *et al.* (2014).

Nesse conjunto de dados também foi encontrado a expressão do gene *Mitochondrial cardiolipin hidrolase*. Além de ter um papel fundamental na conformação e fusão de mitocôndrias, este atua na espermatogênese e na biogênese de *PIWI-interacting RNA* (piRNA) (Pane *et al.* 2007; Huang *et al.* 2011).

6.4.2.6 Gônada - fêmea

Em gônadas de fêmeas a análise de componente principal (PCA) demonstrou que as bibliotecas não possuem um padrão entre os indivíduos do mesmo grupo (Figura 21a). Entretanto, ao realizar a DE foram obtidos 5 transcritos diferencialmente expressos. 3 *up regulated* e 2 *down regulated* (Tabela 7).

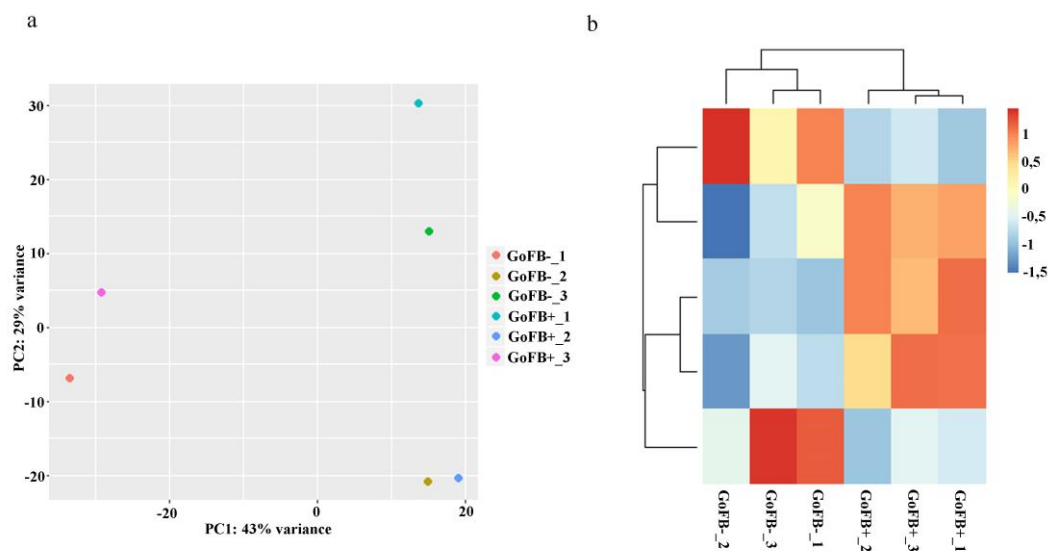


Figura 21. Análise das bibliotecas de gônada de fêmea. Em (a) análise da componente principal (PCA). De acordo com o dado obtido, é possível observar a separação em dois grupos. As bibliotecas são representadas com as siglas: GoFB-_ (amostras sem cromossomo B) e GoFB+_ (amostras com cromossomo B). Em (b) o *heat map* de clusterização hierárquica dos dados de expressão. A escala é distância da centralização e padronização de cada linha separadamente pelo Z-score ((valor real - média do grupo) / desvio padrão) variando de 2 a -1,5. Na parte inferior, as bibliotecas são representadas com as siglas: GoFB-_ (amostras sem cromossomo B) e GoFB+_ (amostras com cromossomo B).

Tabela 7. Transcritos diferencialmente expresso em gônadas de fêmea.

Transcrito id	Log2FoldChange	padj	Uniprot id	Name
transcr_111261	5.07	0.011	I3KPL9	<i>Gap junction protein</i>
transcr_74467	4.63	5.9E-06	I3JH95	<i>Protachykinin 1</i>
transcr_85175	2.8	0.0068	I3JJC2	<i>Iroquois homeobox 4a</i>
transcr_67759	-4.54	0.008	NA	NA
transcr_134124	-2.92	0.0001	I3KIC2	Uncharacterized

Transcrito id: identificação do transcrito; Log2FoldChange: *fold change*; padj: valor de *p-value* ajustado, significativo padj < 0,05; Uniprot id: identificador do banco de dados do uniprot; Nome: nome gênico.

6.5 Análise comparativa: Transcriptoma

6.5.1 Relação macho e fêmea

Com base na análise de expressão diferencial de cada grupo, os genes em comum que se apresentaram alterados entre tecidos e sexos analisados, foram submetidos ao *fisher's exact test (two tail)*, com a finalidade de verificar a dependência do sexo (macho-fêmea) na expressão do gene alterado (Suplementar 4: *comparative_tr*). Desse modo foi possível observar a relação entre duas variáveis analisadas, a presença do cromossomo B e o sexo do indivíduo. Nessa análise foi obtido resultado positivo somente em amostras de encéfalo e músculo.

6.5.1.1 Encéfalo

Nas amostras de encéfalo foram identificados 5 genes *up regulated* em comum entre machos e fêmeas (Suplementar 4: *comparative_tr: up*). Ao analisar a significância do teste estatístico, nenhum dos 5 genes apresentaram significância no teste ($p < 0,05$), ou seja, não apresentando relação entre o gênero (Tabela 8).

Tabela 8. *Fisher's exact test* dos genes em comum *up regulated* entre amostras de fêmea e macho de encéfalo.

Nome	Transcrito id	<i>Fisher's exact test</i>	Hipótese
aurab	transcr_147754	p=1	Não dependente do sexo
aurab	transcr_149468	p=0,67	Não dependente do sexo
aurab	transcr_153489	p=1	Não dependente do sexo
aurab	transcr_158766	p=0,75	Não dependente do sexo
tlr2	transcr_36668	p=1	Não dependente do sexo
pd111	transcr_123261	p=1	Não dependente do sexo

copia	transcr_63455	p=1	Não dependente do sexo
slmo2/prelid3b	transcr_168792	p=1	Não dependente do sexo

Nome: nome do gene; Transcrito id: identificador do transcrito; *Fisher's exact test*: considerado significativo valores $< 0,05$; hipótese: valores de $p < 0,05$ a expressão do gene varia de acordo com a presença do cromossomo B e gênero (dependente do sexo), valores de $p > 0,05$ a expressão do gene varia somente com a presença do cromossomo B (não dependente do sexo).

Para os genes *down regulated* foram identificados 3 genes em comum entre machos e fêmeas (Suplementar 4: *comparative_tr: up*). Ao analisar a significância do teste estatístico, nenhum dos 3 genes apresentaram significância ($p < 0,05$), ou seja, não apresentando relação entre o sexo (Tabela 9).

Tabela 9. *Fisher's exact test* dos genes em comum *down regulated* entre amostras de fêmea e macho de encéfalo.

Nome	Transcrito id	<i>Fisher's exact test</i>	Hipótese
Mog	transcr_130463	$p=0.75$	Não dependente do sexo
Mog	transcr_130517	$p=0.25$	Não dependente do sexo
par14	transcr_105889	$p=0.17$	Não dependente do sexo
Mucm	transcr_167920	$p=0.17$	Não dependente do sexo

Nome: nome do gene; Transcrito id: identificador do transcrito; *Fisher's exact test*: considerado significativo valores $< 0,05$; hipótese: valores de $p < 0,05$ a expressão do gene varia de acordo com a presença do cromossomo B e gênero (dependente do sexo), valores de $p > 0,05$ a expressão do gene varia somente com a presença do cromossomo B (não dependente do sexo).

6.5.1.2 Músculo

Nas amostras de músculo foram identificados 2 genes *up regulated* em comum entre machos e fêmeas (Suplementar 4: *comparative_tr: up*). Ao analisar a significância do teste estatístico, 2 transcritos anotados como aurora quinase e o transcrito anotado como npdc1 apresentaram significância no teste ($p < 0,05$), ou seja, apresentando relação entre o sexo e a presença do cromossomo B (Tabela 10).

Tabela 10. *Fisher's exact test* dos genes em comum *up regulated* entre amostras de fêmea e macho de músculo.

Nome	Transcrito id	<i>Fisher's exact test</i>	Hipótese
Aurab	transcr_147754	p=0.16	Não dependente do sexo
Aurab	transcr_149468	p=0.0009	Dependente do sexo
Aurab	transcr_153489	p=0.0001	Dependente do sexo
Aurab	transcr_158766	NA	NA
npdc1	transcr_99445	p=0.027	Dependente do sexo

Nome: nome do gene; Transcrito id: identificador do transcrito; *Fisher's exact test*: considerado significativo valores $< 0,05$; hipótese: valores de $p < 0,05$ a expressão do gene varia de acordo com a presença do cromossomo B e gênero (dependente do sexo), valores de $p > 0,05$ a expressão do gene varia somente com a presença do cromossomo B (não dependente do sexo).

6.6 Análise comparativa: Transcriptoma e Proteoma

Comparando os resultados obtidos da proteômica com os resultados do transcriptoma, foram identificadas somente algumas similaridades, nas quais tanto genes como proteínas se encontraram com alteração na expressão ou abundância. Além disso, foram verificados se os termos ontológicos em que as proteínas alteradas foram classificadas possuem os mesmos termos ontológicos para a anotação funcional dos genes encontrados pela análise do transcriptoma.

6.6.1 Encéfalo

Na análise do perfil protéico em encéfalo, foram identificadas no total 300 proteínas não redundantes em fêmeas e 243 proteínas não redundantes em macho, incluindo as com baixa e alta abundância, presença somente nas amostras B+ e B- (Suplementar 1: Protein_Table_encephalon_). O transcriptoma identificou 90 genes anotados em fêmeas e 28 genes anotados em machos, incluindo genes com baixa e alta expressão.

Ao analisar a presença de genes e proteínas em amostras de encéfalo de fêmeas foi identificado somente o gene e proteína Tubulina beta 1 com abundância elevada para proteínas e *up regulated* na DE (Suplementar 4: *comparative tr_pr: Table_S1*). Em

macho, foi identificado a *Poly [ADP-ribose] polymerase* (parp14) com baixa expressão gênica e exclusiva em amostra B- na proteoma (Suplementar 4: *comparative tr_pr: Table_S2*). Além disso, foram encontrados alguns termos ontológicos relacionados a processos básicos celulares entre os dados do transcriptoma e proteoma (Tabela 11).

Tabela 11. Termos ontológicos idênticos encontrados no transcriptoma e proteoma.

Gene Ontology ID	Tipo	Descrição	Grupos
GO:0005737	CC	<i>Cytoplasm</i>	
GO:0003924	MF	<i>GTPase activity</i>	Macho
GO:0016787	MF	<i>hydrolase activity</i>	↓transcriptoma e proteoma
GO:0005525	MF	<i>GTP binding</i>	
GO:0005524	MF	<i>ATP binding</i>	Macho
GO:0008017	MF	<i>microtubule binding</i>	↑transcriptoma e exclusivo B+ proteoma
GO:0005515	MF	<i>protein binding</i>	
GO:0030154	BP	<i>cell differentiation</i>	Fêmea
GO:0018105	MF	<i>peptidyl-serine phosphorylation</i>	↑ transcriptoma e exclusivo B+ proteoma
GO:0005515	MF	<i>protein binding</i>	
GO:0003924	MF	<i>GTPase activity</i>	Fêmea
GO:0005525	MF	<i>GTP binding</i>	↑ transcriptoma e proteoma
GO:0005200	MF	<i>structural constituent of cytoskeleton</i>	
GO:0005737	CC	<i>Cytoplasm</i>	Fêmea
GO:0005524	MF	<i>ATP binding</i>	↑ transcriptoma e proteoma exclusivo B+ proteoma

Gene Ontology ID: identificador do termo ontológico; tipo: BP: processo biológico, CC: componente celular, MF: função molecular. Descrição: nome referente ao termo ontológico; grupos: resultado referente ao transcriptoma ou proteoma, genes e proteínas com alteração na expressão ou abundância. ↑ e ↓: indicam alta e baixa expressão ou abundância, respectivamente.

6.6.2 Músculo

Na análise do perfil protéico em músculo foram identificadas no total 218 proteínas não redundantes em fêmeas e 237 proteínas não redundantes em macho, incluindo as com baixa e alta abundância, presença somente nas amostras B+ e B-

(Suplementar 1: Protein_Table_muscle_). O transcriptoma identificou 13 genes anotados em fêmeas e 15 genes anotados em machos, novamente incluindo genes com baixa e alta expressão.

Ao analisar a presença de genes e proteínas em amostras de músculo de fêmeas foi identificado somente o gene e a proteína *actin* (*actc*) *down regulated* na DE e com baixa abundância proteica (Suplementar 4: *comparative tr_pr: Table_S2*). Os termos ontológicos presentes em ambos os resultados da análise do transcriptoma e proteoma estão descritos na Tabela 12.

Tabela 12. Termos ontológicos idênticos encontrados no transcriptoma e proteoma.

Gene Ontology ID	Tipo	Descrição	Grupos
GO:0005975	BP	<i>carbohydrate metabolic process</i>	Macho ↓ transcriptoma e exclusivo B- proteoma
GO:0003824	CC	<i>catalytic activity</i>	
GO:0005524	MF	<i>ATP binding</i>	Macho ↓ transcriptoma e proteoma
GO:0005515	MF	<i>protein binding</i>	
GO:0005524	MF	<i>ATP binding</i>	Fêmea
GO:0005856	CC	<i>Cytoskeleton</i>	↓ transcriptoma e proteoma

Gene Ontology ID: identificador do termo ontológico; tipo: BP: processo biológico, CC: componente celular, MF: função molecular; Descrição: nome referente ao termo ontológico; grupos: resultado referente ao transcriptoma ou proteoma, genes e proteínas com alteração na expressão ou abundância. ↑ e ↓: indicam alta e baixa expressão ou abundância, respectivamente.

6.6.3 Gônada

Na análise do perfil protéico em gônada de macho foram identificadas no total 230 proteínas não redundantes incluindo as com baixa e alta abundância, presença somente nas amostras B+ e B- (Suplementar 1: Protein_Table_gonad_). O transcriptoma identificou 17 genes anotados em machos, novamente incluindo genes com baixa e alta expressão.

Ao analisar a presença de genes e proteínas em amostras de gônada de machos foi identificado somente o gene e a proteína *Tubulin beta 1* com abundância elevada para proteínas e *up regulated* na DE (Suplementar 4: *comparative tr_pr: Table_S1*). Os termos ontológicos presentes em ambos os resultados da análise do transcriptoma e proteoma estão descritos na Tabela 13.

Tabela 13. Termos ontológicos idênticos encontrados no transcriptoma e proteoma.

Gene Ontology ID	Tipo	Descrição	Grupos
GO:0005737	CC	<i>Cytoplasm</i>	Macho
GO:0005874	CC	<i>Microtubule</i>	↓ transcriptoma e proteoma
GO:0003924	MF	<i>GTPase activity</i>	
GO:0005200	MF	<i>structural constituent of cytoskeleton</i>	Macho
GO:0005525	MF	<i>GTP binding</i>	↑ transcriptoma e proteoma
GO:0005874	CC	<i>Microtubule</i>	
GO:0005856	CC	<i>Cytoskeleton</i>	Macho ↑ transcriptoma e proteoma exclusivo B+ proteoma

Gene Ontology ID: identificador do termo ontológico; tipo: BP: processo biológico, CC: componente celular, MF: função molecular; Descrição: nome referente ao termo ontológico; grupos: resultado referente ao transcriptoma ou proteoma, genes e proteínas com alteração na expressão ou abundância. ↑ e ↓: indicam alta e baixa expressão ou abundância, respectivamente.

6.7 Redes Biológicas

As redes de co-expressão e de interação proteína-proteína foram filtradas com o valor de Rho: 0,92 e 0,94 respectivamente, com a distribuição de *degree* obedecendo a lei de potência. As sub redes foram extraídas de acordo com os transcritos diferencialmente expressos e analisado as topologias. O número de interações, de transcritos e as topologias de cada sub rede está descrito na Tabela 14.

As topologias das sub redes de união demonstraram uma distribuição de *degree* obedecendo a lei de potência, disassortiva e apresentou alta conectividade, a qual representa uma rede real biológica (Albert e Barabási, 2002; Foster *et al.*, 2010). Foi verificado que os valores de *degree* não apresentaram distribuição normal, com valores abaixo de 0,05 para o p-value sob o teste de *Anderson-Darling normality test* e valores

significativos (p-value < 0,05) para o teste de *Wilcoxon*, demonstrando que há diferença entre as sub redes *up regulated* e *down regulated*, ou seja, a presença do cromossomo B influencia na expressão de RNAm (Suplementar 5: Subnetwork: Table S1).

A partir dessas sub redes, foram analisadas duas topologias: *Degree* e *Betweenness centrality*. O *degree* representa a possibilidade de que um nó na rede seja considerado um *hub*, ou seja, um gene com alta conectividade. O *betweenness centrality* informa a importância desse nó na rede. Para cada sub rede analisamos os 10 primeiros transcritos em cada topologia analisada (*degree e betweenness centrality*) (Suplementar 5: Subnetwork:Table S2).

As sub redes extraídas dos transcritos *up regulated* de encéfalo e gônada de macho, músculo de macho e fêmea apresentaram genes relacionados ao ciclo celular entre os 10 primeiros transcritos analisados, demonstrando a importância desses genes na presença do cromossomo B (Tabela 15 e Figura 22).

Tabela 14. Análise das sub redes.

Sub redes	Vértices	Arestas	Average Degree	Average betweenness	Average path length	Assortativity Degree
Encéfalo – Fêmea – <i>down regulated</i>	17061	92506	10,84	11518,08	2,52	-0,43
Encéfalo – Fêmea – <i>up regulated</i>	21911	36180	3,30	19166,63	2,76	-0,36
Encéfalo – Macho – <i>down regulated</i>	20273	33586	3,32	12559,87	2,40	-0,38
Encéfalo – Macho – <i>up regulated</i>	17616	23758	2,69	9976,37	2,14	-0,57
Músculo – Fêmea – <i>down regulated</i>	778	1049	2,69	468,70	2,23	-0,55
Músculo – Fêmea – <i>up regulated</i>	1004	4991	9,94	492,55	2,02	-0,86
Músculo – Macho – <i>down regulated</i>	13643	25527	3,74	6818,12	1,99	-0,60
Músculo – Macho – <i>up regulated</i>	673	1089	3,23	298,03	1,99	-0,88
Gônada – Macho – <i>down regulated</i>	1298	1720	2,65	386,05	2,01	-0,67
Gônada – Macho – <i>up regulated</i>	6824	11031	3,23	3959,33	2,16	-0,51

Tabela 15. Genes relacionados ao ciclo celular entre os 10 maiores valores para *degree* e *betweenness centrality*.

Sub rede	Gene	<i>Degree</i>	<i>Betweenness Centrality</i>
Gônada – macho <i>up</i>	Nuf2	629	—
	AuraA (transcr_153489)	631	0,66
Músculo – fêmea <i>up</i>	AuraA (transcr_158766)	630	—
	Ccnb1(transcr_126584)	—	1
Músculo – macho <i>up</i>	AuraA (transcr_153489)	631	0,77
	AuraA (transcr_158766)	3	—
Encéfalo – macho <i>up</i>	AuraA (transcr_153489)	631	—
	AuraA (transcr_158766)	630	—

Legenda: **Sub rede:** sub redes extraídas a partir de transcritos *up regulated*, **gene:** nome do gene e id gerado pela montagem do transcriptoma. Quanto maior o valor de *degree* e *betweenness centrality* maior importância do gene. *Betweenness centrality* varia de 0 a 1.

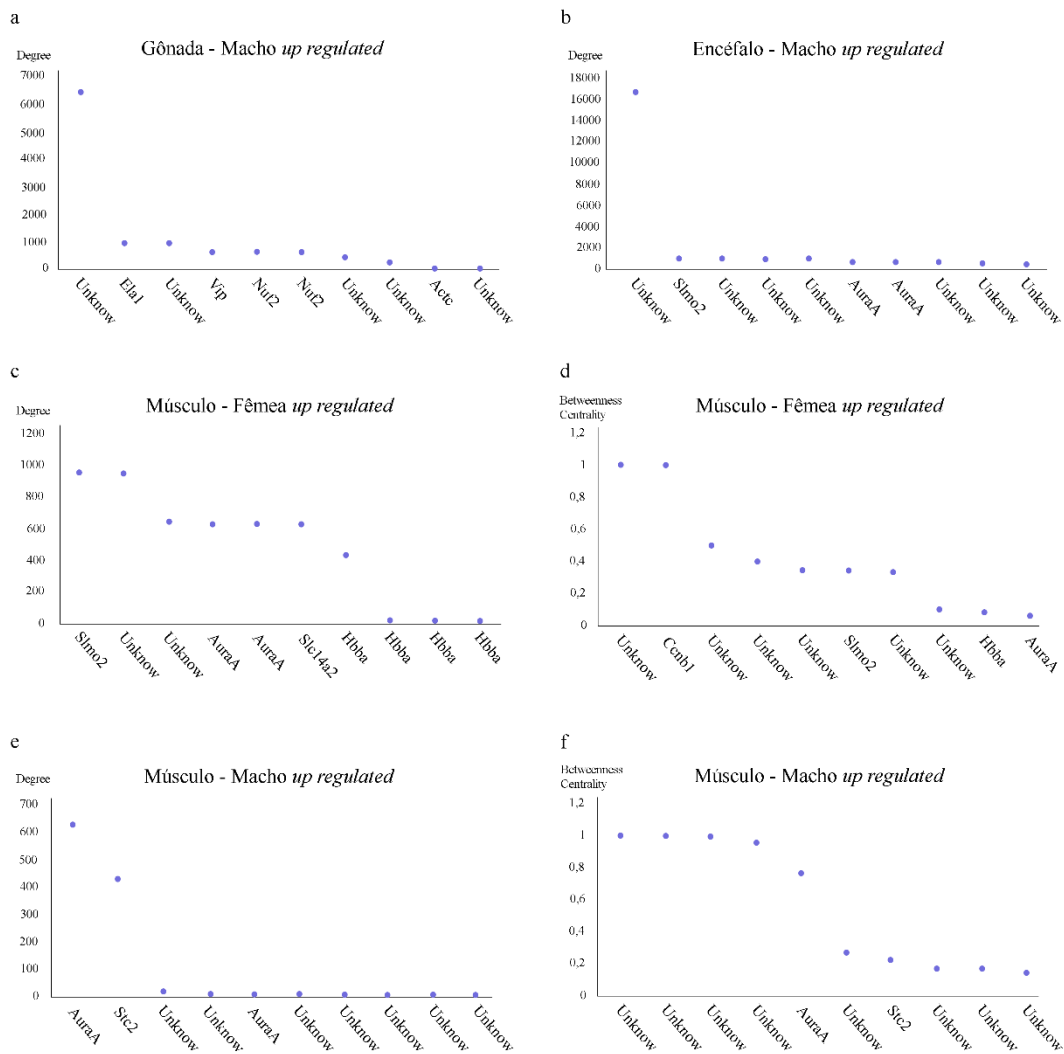


Figura 22. Representação gráfica da análise das sub redes *sup regulated* extraídas da rede de união Co-PPI-net que apresentam genes envolvidos no ciclo celular. Em **a**, **b**, **c** e **e** gráficos dos valores de *degree*, identificando potenciais hubs. No eixo das abscissas estão os transcritos e no eixo das ordenadas estão os valores de *degree*. Em **d** e **f** gráfico dos valores de *betweenness centrality*, identificando potenciais genes essenciais/importantes. No eixo das abscissas estão os transcritos e no eixo das ordenadas estão os valores de *betweenness centrality*. Genes envolvidos no ciclo celular: nuf2, AuraA e ccnb1.

7 DISCUSSÃO

7.1 Mapas biológicos: Proteoma

Nas amostras analisadas, a partir da abundância proteica, foram identificados processos básicos celulares nos quais essas proteínas participam. Entre os termos identificados, foi enfatizado a via de sinalização da família *forkhead box O* (FOXO) (encéfalo e gônada de fêmea), que é uma família de fatores de transcrição reguladores da expressão gênica em diferentes condições fisiológicas, os quais controlam o ciclo celular, o metabolismo de glicose, da apoptose e da resistência ao estresse oxidativo. Nessa via, em encéfalo de fêmeas B+ a ausência da proteína Sgkb2, inibidor do FOXO, e a presença da proteína PLK demonstra uma regulação positiva na via de sinalização do FOXO. Por outro lado, em gônadas foram identificadas as proteínas *kinase protein AMP* (AMPK) somente nas amostras B- e o *epidermal growth factor receptor* (EGFR) nas amostras B+. As proteínas AMPK e EGFR possuem uma função regulatória antagônica na sinalização do FOXO, demonstrando que na presença do cromossomo B o receptor EGF pode acarretar uma regulação negativa na sinalização do fator de transcrição e consequentemente pode alterar a expressão gênica. Por outro lado, o EGFR na via de sinalização do GnRH sofre transativação e leva a ativação de proteínas quinases ativadas por mitógeno (MAPKs), incluindo a *extracellular signal-regulated kinases* (ERK), *kinase JUN N-terminal* (JNK) e *p38 MAPK*. As MAPKs ativas translocam-se para o núcleo, resultando na ativação de fatores de transcrição e na indução rápida de genes (Farshori et al., 2003, Pawson e McNeilly, 2005). Adicionalmente, em encéfalo de fêmeas, ainda nessa via, foi encontrada a proteína *calcium/calmodulin-dependent protein kinase* (CAMK) que atua como fator de transcrição no promotor da subunidade LH β para regular a expressão do hormônio luteinizante (LH) (Pawson e McNeilly, 2005), na qual a presença do cromossomo B pode estimular a produção e liberação do hormônio luteinizante pela hipófise na maturação final e desova. Além disso, esta via pode estar relacionada ainda a controle da divisão celular, na qual pode atuar na manutenção e

perpetuação do cromossomo B (Valente *et al.* 2014; Clark *et al.* 2017; Navarro-Domínguez *et al.* 2017a,b; Kichigin *et al.* 2019; Makunin *et al.* 2018; presente trabalho).

Proteínas relacionadas ao ciclo celular foram encontradas em encéfalo de macho e de fêmea na presença do cromossomo B. A proteína Mps1 ou TTK (*dual-specificity protein kinase ttk*) presente nas amostras exclusivas B+ (encéfalo de fêmea) é uma das principais *kinases* envolvidas na localização do *kinetochoro* e no *checkpoint* do conjunto do fuso (SAC), que inibe a transição da metáfase para a anáfase até que a conexão adequada do fuso a todos os cromossomos seja alcançada (Kaistha *et al.*, 2014). Ainda nesta via foi encontrada a enzima 14-3-3 com alta abundância em fêmeas. As proteínas 14-3-3 se ligam a proteínas reguladoras e modulam sua função na transição das fases G1/S e G2/M durante o ciclo celular (Hermeking e Bnezinger, 2006; Cann e Hicks, 2007). Também foi identificado a proteína Mcm6 do complexo MCM com baixa abundância que está relacionado a formação da forquilha de replicação do DNA (Hotton *et al.*, 2018). Já nas amostras de encéfalo de macho encontramos a proteína *Anaphase-promoting complex subunit 1* (APC/C), que também atua na regulação da transição de metáfase para anáfase juntamente com o *mitotic checkpoint complex* (MCC ou SAC) (Acquaviva e Pines, 2006; Pines, 2011; Bolanos-Garcia e Blundel, 2011). Desse modo novamente foi observado que a presença do cromossomo B está alterando os níveis de proteínas que atuam na regulação do ciclo celular promovendo e facilitam sua manutenção, já que uma alteração drástica em vários processos biológicos poderia acarretar em sua eliminação do genoma.

Nas gônadas de macho encontramos a via de transporte de RNA com as proteínas *Elongation factor 1-alpha 1* (eEF1A), *Eukaryotic initiation fator 4A* e *4A3* (eIF4a e eIF4A3) e a proteína *Poly(A)-binding* (PABP) com baixa abundância ou somente nas amostras B- (proteínas exclusivas em amostras B-). Além disso, essas proteínas podem estar relacionadas a resposta ao estresse e a RNAs derivados de fragmentação de RNAs transportadores (tRFs), afetando a degradação e o transporte de moléculas entre núcleo e citoplasma (Keam e Hutvagner, 2015). Adicionalmente, foi encontrada a presença de tRFs com expressão diferencial em células B+ (Bovolenta *et al.* em preparação). O enriquecimento funcional das proteínas abundantes referentes à resposta imune e estresses na presença do cromossomo B pode estar conferindo resistência aos seus

portadores, principalmente em regiões ambientais adversas ocasionadas por poluições químicas (Pires *et al.* 2015; Clark *et al.* 2017).

As amostras de músculo de macho B+ apresentaram mapas metabólicos básicos com ênfase na via de sinalização do MAPK, na qual encontramos a enzima *tyrosine-protein phosphatase non-receptor type 5* (PTP, EC:3.1.3.48), repressor das proteínas *mitogen-activated protein kinase 1/3* (erk), *mitogen-activated protein kinase 8/9/10* (*c-Jun N-terminal kinase*) e a proteína *p38 MAP kinase* (EC:2.7.11.24), as quais são *kinases* importantes na regulação da proliferação, diferenciação, inflamação, apoptose, ciclo celular e via da sinalização do *p53*.

7.2 Transcriptoma

A entropia de *Shannon* demonstrou que a expressão global dos transcritos referentes a presença do cromossomo B não altera a ponto de modificar a expressão geral, ou seja, o cromossomo B altera a expressão de genes em pontos específicos como observado na análise de expressão diferencial. Desse modo, a diversidade da expressão dos transcritos mostra que a presença do B não acarreta mudanças drásticas e fenotípicas, o que provavelmente causaria sua eliminação. Além disso, análises prévias em dados de expressão de pequenos RNAs apresentou o mesmo comportamento (Bovolenta *et al.* em preparação). De acordo com esses resultados foram destacados processos biológicos pontuais evidenciados pela análise de expressão diferencial.

A expressão diferencial nas amostras do transcriptoma de RNAm de encéfalo, músculo e gônadas, na presença do cromossomo B, apresentou uma alteração na expressão de genes reguladores e genes de checagem durante uma fase específica do ciclo celular. Os genes da família *aurora kinase* (em encéfalo e músculo), proteína do *kinetochoro* (*nuf2*, encéfalo de fêmea e gônada de macho) e gene *G2/mitotic-specific cyclin-B1* (músculo de fêmea) e o enriquecimento para o *Ndc 80 complex* (em encéfalo de fêmea) participam do processo de formação do *spindle assembly* e *spindle assembly checkpoint* (SAC) durante a segregação cromossômica na divisão celular (Ducat e Zheng, 2004; Ciferri *et al.* 2007). O processo de *checkpoint* acontece durante a transição e confirmação do reconhecimento e da interação do *kinetochoro* aos microtúbulos para separação dos cromossomos homólogos (na meiose I) ou das cromátides irmãs (meiose

II ou mitose). Como descritos por Clark *et al.* (2017), a manutenção do cromossomo B e a frequência do número de cromossomos B em ciclídeos do lago Malawi depende do mecanismo de *drive* (mecanismos de transmissão vantajoso) do mesmo. Um desses mecanismos pode ser proveniente da não disjunção dos cromossomos homólogos ou das cromátides irmãs, ocasionando uma frequência de indivíduos com 0, 1 ou 2 cromossomos B (Clark *et al.* 2017). Além disso, a falha desse processo (SAC) é uma das principais causas de aneuploidias devido a segregação incorreta das cromátides irmãs (Bolanos-Garcia e Blundel, 2011). Genes relacionados ao ciclo celular foram encontrados com alta integridade no genoma da espécie em estudo por Valente *et al.* (2014) e em outras espécies como gafanhoto (*Eyprepocnemis plorans*), lagarto (*Anolis carolinensis*), raposa (*Vulpes vulpes L.*) e o cão chinês (*Nyctereutes procyonoides procyonoides Gray*) (Navarro-Domínguez *et al.* 2017a,b, Kichigin *et al.* 2019, Makunin *et al.* 2018).

7.3 Análise comparativa

A análise da relação do gênero (macho-fêmea) na expressão dos genes alterados pela presença do cromossomo B (valores de $p < 0,05$ para o *Fisher's exact test*) mostrou resultados positivos somente para dois genes, a *aurora kinase* e o gene *neural proliferation differentiation and control protein 1*, ambos nas amostras de músculo e *up regulated* em amostras B+. Novamente, trata-se de genes com envolvimento ao ciclo celular e sistema imune (Sansal *et al.*, 2000; Ducat e Zheng, 2004; Sun *et al.*, 2012).

Considerando a teoria de conjuntos e de que não há uma alta correlação entre transcriptoma e proteoma (Vogel e Marcotte, 2012), foram verificadas similaridades entre genes e proteínas alteradas e processos biológicos em que os diferentes genes e proteínas podem atuar. Em geral, os termos ontológicos encontrados foram processos celulares básicos. Novamente, foram observados genes relacionados a divisão celular e sistema imune, sendo frequente para quase todas as amostras o gene da tubulina beta 1 (*tubb1*). Considerando a presença da tubulina na maioria das amostras tanto no transcriptoma como na proteoma, esse gene possui uma grande chance de ser traduzido. Além disso, foi demonstrado que esse gene pode apresentar cópia específica do cromossomo B (Valente *et al.* 2014), o que pode evidenciar que genes presentes no cromossomo B podem apresentar propriedades funcionais, sendo transcritos e traduzidos.

7.4 Redes biológicas

As sub redes extraídas a partir da união das redes de co-expressão e interação proteína-proteína apresentaram genes envolvidos no ciclo celular para quase todas as amostras referentes a transcritos *up regulated*, demonstrando, segundo as topologias, serem genes com alto valor para *odegree* e essenciais. Além disso, a união entre as redes de co-expressão e interação proteína-proteína demonstram que não há somente uma correlação na expressão desses genes, mas também uma interação física entre elas. Portanto, a alteração de expressão desses genes relacionados ao ciclo celular na presença do cromossomo B confere forte evidencia da importância deste cromossomo no complexo processo biológico que envolve o ciclo celular.

8 CONCLUSÃO

As análises de proteômica recuperou, de uma forma geral, mapas biológicos de vias metabólicas fundamentais como, por exemplo, as vias de sinalização, regulação gênica, transporte de moléculas e ciclo celular, o que indica que a presença do cromossomo B não está atuando de modo global, mas sim em vias específicas e possivelmente regulatórias. Não diferente, a análise de bioinformática do transcriptoma de RNAm não apresentou diferença global no perfil de expressão causado pela presença do cromossomo B. A expressão diferencial, na presença do cromossomo B, apresentou diferenças pontuais, nas quais genes alterados participam principalmente na regulação do ciclo celular e resposta do sistema imune, além de processos celulares fundamentais. Ao ser analisados em conjunto os dados do transcriptoma de RNAm e proteoma, apesar de nem sempre demonstrarem a presença dos mesmos genes/proteínas em tecidos diferentes, os processos de regulação da divisão celular, como a SAC e a transição das fases G1/S e G2/M, estão presentes. Portanto, a regulação desses genes nesses processos pode atuar na manutenção do cromossomo B nas células.

Por outro lado, a presença de genes/proteínas relacionadas à resposta ao estresse e resposta do sistema imune não descarta a possível relação da frequência de cromossomo B em espécies onde o ambiente apresenta condições adversas como a poluição. Desse modo, o cromossomo B pode conferir resistência a certos poluentes ajudando na sobrevivência da espécie portadora e de sua própria manutenção.

Em relação às redes biológicas, as sub redes representaram ser redes biológicas reais de acordo com as topologias de distribuição de *degree*, disassortividade, e alta conectividade. As redes demonstraram novamente a importância de genes relacionados ao ciclo celular, evidenciando ainda mais a relação entre o controle do ciclo celular e a presença do cromossomo B.

De modo geral, os resultados obtidos a partir das análises de bioinformática mostraram que o cromossomo B além de apresentar elementos repetitivos, pseudogenes e genes em sua composição, podem ainda alterar a expressão gênica e abundância proteica, apresentando evidências funcionais relacionadas ao cromossomo B na espécie de ciclídeo *A. latifasciata*.

Análises futuras relacionadas aos genes encontrados nesse estudo como os genes *daaurora kinase*, *kinetochoro nuf2 protein*, *G2/mitotic-specific cyclin-B1*, *dual-specificity protein kinase ttk* e *anaphase-promoting complex subunit 1* que participam do processo de formação do *spindle assembly* e *spindle assembly checkpoint* (SAC) durante a segregação cromossômica, a enzima 14-3-3 que se ligam a proteínas reguladoras e modulam sua função na transição das fases G1/S e G2/M e a *tubulin beta 1* que é um dos maiores constituintes dos microtubulos que atuam na segregação cromossômica, abrem novos caminhos a serem estudadas sobre a funcionalidade do cromossomo B.

9 MATERIAIS SUPLEMENTARES

Para acessar os materiais suplementares segue o link [Materiais Suplementares](#).

10 REFERÊNCIAS

Acquaviva CJ, Pines J (2006) The anaphase-promoting complex/cyclosome: APC/C. *J Cell Sci* 2006 119: 2401-2404.

Albert R, Barabási AL (2012) Statistical mechanics of complex networks. *Rev Mod Phys* 74, 47.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 403-410.

Anders S, Paul Theodor Pyl, Wolfgang Huber (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2): 166-169.

Banaei-Moghaddam AM, Meier K, Karimi-Ashtiyani R, Houben A (2013) Formation and Expression of Pseudogenes on the B chromosome of Rye. *Plant Cell*, 25: 2536-2544.

Barnum KJ, O'Connell MJ (2014) Cell cycle regulation by checkpoints. *Methods Mol Biol*, 1170:29-40.

Bellomo N, Carbonaro B (2011) Toward a mathematical theory of living systems focusing on developmental biology and evolution: a review and perspectives. *Phys Life Rev*, 8(1):1-18.

Beukeboom LW (1994) Bewildering B: An impression of the 1st B—chromosome conference. *J Hered* 73:328–336.

Blakeley P, Siepen JA, Lawless C, Hubbard SJ (2010) Investigating protein isoforms via proteomics: a feasibility study. *Proteomics*, 10(6):1127–1140.

Bolanos-Garcia VM & Blundell TL (2011) BUB1 and BUBR1: multifaceted kinases of the cell cycle. *Trends in Biochemical Sci*, 36(3), 141–150.

Bolger AM, Lohse M, Usadel, B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, *et al.*, (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513: 375–381.

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, Lee TJ, Leigh ND, Kuo TH, Davis FG, Bateman J, Bryant S, Guzikowski AR, Tsai SL, Coyne S, Ye WW, Freeman RM Jr, Peshkin L, Tabin CJ, Regev A, Haas BJ, Whited JL (2017) A Tissue-30 Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep*; 18(3):762-776.

Burrows F, Zhang H, Kamal A (2004) Hsp90 Activation and Cell Cycle Regulation, *Cell Cycle*, 3:12, 1530-1536.

Cann KL, Hicks GG (2007) Regulation of the cellular DNA double-strand break response. *Biochem. Cell Biol.* 85: 663-674.

Carmello BO, Coan RLB, Cardoso AL, Ramos E, Fantinatti BEA, Marques DF, Oliveira RA, Valente GT, Martins C. (2017) The hnRNP Q-like gene is retroinserted into the B chromosomes of the cichlid fish *Astatotilapia latifasciata*. *Chromosome Res* 25(3-4):277-290.

Chang X, Kobayashi T, Senthilkumaran B, Kobayashi-Kajura H, Sudhakumari CC, Nagahama Y (2005) Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). *Gen. Comp Endocrinol* 141: 101–115.

Chen S, Harmon AC (2006) Advances in plant proteomics. *Proteomics*, 6(20):5504-16.

Ciferri C, Musacchio A, and Petrovic A. (2007) The Ndc80 complex: hub of kinetochore activity. *FEBS Lett*, 581, 2862–2869.

Clark FE, Conte MA, Ferreira-Bravo IA, Poletto AB, Martins C, Kocher TD. (2017) Dynamic Sequence Evolution of a Sex-Associated B Chromosome in Lake Malawi Cichlid Fish. *J Hered*, Volume 108, Issue 1, Pages 53–62.

Cnaani A, Lee B-Y, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller J-F, D'Cotta H, Penman DJ, Tomasino E, Coutanceau J-P, Pepey E, Shirak A, Kocher TD (2008) Genetics of Sex Determination in Tilapiine Species. *Sexual Dev* 2: 43-54.

Craig KL, Tyers M. (1999) The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog Biophys Mol Biol*, Volume 72, Issue 3.

D'Ambrosio U, Alonso-Lifante MP, Barros K, Kovarik A, Mas de Xaxars G, Garcia S. (2017) B-chrom: A database on B-chromosomes of plants, animals and fungi. *New Phytol*, 216, 635–642.

Dietrich MA, Adamek M, Hejmej A, Ciereszko A (2014) Characterization, expression and antibacterial properties of apolipoproteins A from carp (*Cyprinus carpio* L.) seminal plasma. *Fish Shellfish Immunol*, 41(2):389-401.

Ducat D, Zheng Y. (2004) Aurora kinases in spindle assembly and chromosome segregation. *Exp Cell Res*, Volume 301, Issue 1:60-67.

Fantinatti BE, 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 Genet*, 17(1):119.

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

Farshori PQ, Shah BH, Arora KK, Martinez-Fuentes A, Catt KJ (2003) Activation and nuclear translocation of PKCdelta, Pyk2 and ERK1/2 by gonadotropin releasing hormone in HEK293. cells. *J Steroid Biochem Mol Biol* 85:337-47.

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

Feldberg E, Bertollo LAC (1984). Discordance in chromosome number among somatic and gonadal tissue cells of *Gymnogeophagus balzanii* (Pisces, Cichlidae). *Rev. Brasil. Genet.* 7: 639–645.

Ferreira IA, Martins C (2008) Physical chromosome mapping of repetitive DNA sequences in Nile tilapia *Oreochromis niloticus*: evidences for a differential distribution of repetitive elements in the sex chromosomes. *Micron* 39: 411-418.

Finn RD, Clements J, Eddy SR (2011) HMMER Web Server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–W37.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) The Pfam protein families database. *Nucleic Acids Res* 42: D222–D230.

Firat-Karalar EN, Stearns T (2014) The centriole duplication cycle. *Philos Trans R Soc Lond B Biol Sci*, 369(1650):20130460.

Foster J, Foster DV, Grassberger P, Paczuski M (2010) "Edge direction and the structure of networks". *PNAS*, **107** (24): 2.

Graphodatsky AS, Kukekova AV, Yudkin DV et al (2005) The proto-oncogene CKIT maps to canid B-chromosomes. *Chromosome Res*, 13: 113-122.

Gutknecht J, Sperlich D, Bachmann L (1995) A species specific satellite DNA family of *Drosophila subsilvestris* appearing predominantly in B chromosomes. *Chromosoma* 103: 539-544.

He B, Zhao S, Chen Y, Cao Q, Wei C, Cheng X, Zhang Y (2015) Optimal assembly strategies of transcriptome related to ploidies of eukaryotic organisms. *BMC Genomics*, 16:65.

Hermeking H, Benzinger A (2006) 14-3-3 proteins in cell cycle regulation. *Semin Cancer Biol*, Volume 16, Issue 3:183-192, ISSN 1044-579X.

Hotton J, Agopiantz M, Leroux A, Charra-Brunaud C, Marie B, Busby-Venner H, Morel O, Guéant JL, Vignaud JM, Battaglia-Hsu SF, Gauchotte G (2018) Minichromosome maintenance complex component 6 (MCM6) expression correlates with histological grade and survival in endometrioid endometrial adenocarcinoma. *Virchows Arch*, 472:623-633.

Houben A (2017) B Chromosomes - A Matter of Chromosome Drive. *Front Plant Sci*, 8:210.

Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. *Genome Research*. 9:868–877.

Huang H, Gao Q, Peng X, et al. (2011) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell*, 20(3):376-87.

Ijiri S, Kaneko H, Kobayashi T, Wang D-D, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y (2007) Sexual Dimorphic Expression of Gene in Gonads during early differentiation of a Teleost Fish, the Nile Tilapia *Oreochromis niloticus*. *Bio Reprod*, 78: 333–341.

Jamilena M, Ruiz-Rejón C, Ruiz-Rejón M (1994) A molecular análisis of the origin of the *Crepis Capillaris* B chromosome. *J Cell Sci*, 107: 703-708.

Jiang L, et al. (2013). LC-MS/MS identification of doublecortin as abundant beta cell-selective protein discharged by damaged beta cells in vitro. *J Proteomics*, 80:268-80.

Jones RN, Rees H (1982) B chromosome. New York: Academic Press.

Jones RN, Viegas W, Houben A (2008) A century of B chromosomes in plants: so what?. *Ann Bot*, 101(6):767–775.

Jones RN (2018) Transmission and Drive Involving Parasitic B Chromosomes. *Genes (Basel)*. 9(8):388.

Junker B.H. (2008). Networks in Biology. In: Analysis of Biological Networks. Junker B.H., Schreiber F. (eds). John Wiley & Sons. p. 3-12.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*, 110:462–467.

Kaistha BP, Honstein T, Müller V, Bielik S, Sauer M, Kreider, R, ... Buchholz M (2014) Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. *Br J Cancer*, 111(9), 1780–1787.

Keam SP, Hutvagner G (2015) tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life*, 5(4), 1638–1651.

Kichigin IG, Lisachov AP, Giovannotti M. *et al.* (2019) First report on B chromosome content in a reptilian species: the case of *Anolis carolinensis*. *Mol Genet Genomics* 294: 13.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet*, 5: 288-98.

Kuroiwa A, Terai Y, Kobayashi N, Yoshida K, Suzuki M, Nakanishi A, Matsuda Y, Watanabe M, Okada N (2014) Construction of chromosome markers from the Lake Victoria cichlid *Paralabidochromis chilotis* and their application to comparative mapping. *Cytogenet Genome Res*, 142(2):112-20.

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9:357-359.

López-León MD, Cabrero J, Dzyubenko VV, Bugrov AG, Karamysheva TV, Rubtsov NB, Camacho JPM (2008) Differences in ribosomal DNA distribution on A and B chromosomes between eastern and western populations of the grasshopper *Eyprepocnemis plorans plorans*. *Cytogenet Genome Res*, 121: 260-265.

Loreto V, Cabrero J, López-León MD, Camacho JPM, Souza MJ (2008) Possible autosomal origin of macro B chromosomes in two grasshopper species. *Chromosome Res*, 16: 233-241.

Love, M. I., Huber, W., & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550.

Luo W, Brouwer C (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics (Oxford, England)*; 29:1830–1.

Makunin AI, Rajičić M, Karamysheva TV, *et al.* (2018) Low-pass single-chromosome sequencing of human small supernumerary marker chromosomes (sSMCs) and Apodemus B chromosomes. *Chromosoma* 127: 301.

Marques, D.F. Functional analysis of B chromosome presence using cichlid *Astatotilapia latifasciata* as model. Available online: <http://hdl.handle.net/11449/141953> (accessed on March 2016).

Martins C, et al. (2011). Animal genomes under the focus of cytogenetics. 1. ed. Hauppauge: Nova Science Publisher, 2011. v. 1. 160.

Martins-Santos IC, Portela-Castro ALB, Julio HF Jr (1995) Chromosome analysis of 5 species of the Cichlidae family (Pisces, Perciformes) from the Paraná River. *Cytologia*, 60:223–231.

Martis MM , Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, Scholz U, Gundlach H, Wicker T, Simková H, Novák P, Neumann P, Kubaláková M, Bauer E, Haseneyer G, Fuchs J, Dolezel J, Stein N, Mayer KFX, Houben A (2012) Selfish supernumerary chromosome reveals its origins a mosaic of host genome and organellar sequences. *PNAS*. 109:13343–13346.

Nam BH, Kim YO, Kong HJ, Kim WJ, Lee SJ, Choi TJ (2009) Identification and characterization of the prepro-vasoactive intestinal peptide gene from the teleost *Paralichthys olivaceus*, 127(3-4):249-58.

Navarro-Domínguez B, Ruiz-Ruano FJ, Camacho JPM, Cabrero J, López-León MD (2017a) Transcription of a B chromosome CAP-G pseudogene does not influence normal Condensin Complex genes in a grasshopper. *Sci Rep*, 7:17650.

Navarro-Domínguez B, Ruiz-Ruano FJ, Cabrero J, Corral JM, López-León MD, Sharbel TF, Camacho JPM (2017b) Protein-coding genes in B chromosomes of the grasshopper *Eyprepocnemis plorans*. *Sci. Rep*, 7, 45200.

Nelson JS: Fishes of the world, 4th ed (John Wiley and Sons, Inc. New York 2006).

Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH (1988) A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome. *Science* 240: 512-514

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*; 27:29–34.

Pane A, Wehr K, Schüpbach T (2007) zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev cell*, 12(6), 851-62.

Park OK (2004). *Proteomic Studies in Plants*. *J. Bioch. Mol. Biol*, v. 37, p. 133-138.

Pavlopoulos GA, *et al.* (2011) Using graph theory to analyze biological networks. *Biodata Mining*, 4: 10.

Pawson AJ, McNeilly AS (2005) The pituitary effects of GnRH. *Anim Reprod Sci* 88(1–2):75–94.

Pines, J. (2011) Cubism and the cell cycle: the many faces of the APC/C. *Nat. Rev. Mol. Cell Biol.* 12, 427-438.

Pinhal D, Bovolenta LA, Moxon S, *et al.* (2018) Genome-wide microRNA screening in Nile tilapia reveals pervasive isomiRs' transcription, sex-biased arm switching and increasing complexity of expression throughout development. *Sci Rep*, 8(1):8248.

Poletto AB, Ferreira IA, Martins C (2010a). The B chromosome of the cichlid fish *Haplochromis obliquidens* harbors 18S rRNA genes. *BMC Genetics* 11: 1.

Poletto AB, *et al.* (2010b). Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet* 11: 5-10.

Puertas MJ (2002). Nature and evolution of B chromosomes in plants: A noncoding but information-rich part of plant genomes. *Cytogenet Genome Res* 96: 198-205.

Quinlan, Aaron R., and Ira M. Hall (2010) “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26.6: 841–842. PMC. Web. 23 Oct. 2017.

Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J (2016) g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic acids research*, 44(W1), W83-9.

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. *Trends Genet* 16, 276–277.

Sansal I, Dupont E, Toru D, Rouget P (2000) NPDC-1, a regulator of neural cell proliferation and differentiation, interacts with E2F-1, reduces its binding to DNA and modulates its transcriptional activity. *Oncogene* 19, 5000-5009.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics Advance*.

Schulz RW, De Franca LR, Lareyre J-J, Legac F, Chiarini-Garcia H, Nobrega RH, Miura T (2010) Spermatogenesis in fish. *Gen Comp Endocrinol*, v. 165, n. 3, p. 390-411. ISSN 0016-6480.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498-504

Smit AFA, Hubley R, Green P (2013) *RepeatMasker Open-4.0*. 2013-2015 <http://www.repeatmasker.org>.

Sun Y, Senger K, Baginski TK, *et al.* (2012) Evolutionarily conserved paired immunoglobulin-like receptor α (PILR α) domain mediates its interaction with diverse sialylated ligands. *J Biol Chem*, 287(19):15837–15850.

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 2011;6(7):e21800.

Teruel M, *et al.* (2010) B chromosome ancestry revealed by histone genes in the migratory locust. *Chromosoma* 119: 217-225.

The International Cichlid Genome Consortium (2006). Genetic basis of vertebrate diversity: the cichlid fish model.

Trewavas E (1983). Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum, Natural History, London.

Valente GT, et al. (2009). Comparative cytogenetics of cichlid fishes through genomic in-situ hybridization (GISH) with emphasis in *Oreochromis niloticus*. *Chromosome Res*, 17(6):791-799.

Valente GT, Acencio ML, Martins C (2013) The Development of a Universal *In silico* Predictor of Protein-Protein Interactions. *Plos One*, 8: e65587.

Valente GT, Conte MA, Fantinatti BEA, Cabral-de-Mello DC, Carvalho RF, Vicari MR, Kocher TD, Martins C (2014) Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. *Mol Biol Evol*, 31:2061–2072.

Vogel C, Marcotte EM (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics* 13, 227-232.

Wang D-S, et al. (2007). Foxl2 up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. *Mol Endocrinol*, 21: 712–725.

Wasinger VC, et al. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16, 1090-1094.

Wilkins MR, et al. (1997). Detailed peptide characterization using PEPTIDEMASS--a World-Wide-Web accessible tool. *Electrophoresis*. 1997 Mar-Apr;18(3-4):403-8.

Wilson E (1907) The supernumerary chromosome of Hemiptera. *Science* 26:870-871.

Yoshida K, et al. (2011). B Chromosomes have a functional effect on female sex determination in Lake Victoria cichlid fishes. *Plos Genetics* 7(8): e1002203.

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

Zhou Q, et al. (2012). Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 2012, 13:109 doi:10.1186/1471-2164-13-109.