UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Análise genômica e transcriptômica das rotas *Argonauta* e *Piwi* em *Stenostomum leucops*

Deborah Cristina Schoemberger

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito para a obtenção do título de **Mestre em Genética e Biologia Molecular**

Orientador: Élgion Lúcio da Silva Loreto

PORTO ALEGRE

01/2024

Instituições e fontes Financiadoras

O desenvolvimento deste trabalho foi realizado no Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul (UFRGS) e Laboratório de Biologia Molecular e Sequênciamento (LabDros), situado na Universidade Federal de Santa Maria (UFSM). A pesquisa foi financiada pelo Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq).

Agradecimentos

Desejo expressar os meus agradecimentos a todos aqueles que, de alguma forma, permitiram que esta dissertação se concretizasse.

Primeiramente, agradeço ao Prof. Dr. Élgion da Silva Loreto, pelas valiosas contribuições dadas durante todo o processo, pela oportunidade e apoio na elaboração deste trabalho, por acreditar em mim e nas minhas capacidades.

Ao PPGBM e o corpo docente pelo ambiente criativo e amigável proporcionado.

A CNPq pela bolsa concedida, que representa não apenas um apoio financeiro, mas também um incentivo inestimável para meu crescimento acadêmico e pessoal.

Aos meus amigos e colegas do LabDros pela colaboração, apoio e troca de conhecimentos. Agradeço pelo companheirismo e as boas conversas.

Aos amigos que fiz durante o curso e àqueles, que apesar da distância, estiveram presentes. Agradeço por estarem ao meu lado, compartilhando risos, conselhos e tornando esta jornada acadêmica mais significativa e memorável.

Agradeço em especial minha amiga Anna Júlia. Suas palavras de incentivo, suporte e compreensão foram um alívio nos momentos mais desafiadores. Obrigada por estar sempre presente, pelas risadas, histórias compartilhadas e pelo ombro amigo.

A minha família pelo apoio e incentivo. Agradeço especialmente aos meus pais Edinéia e Paulo que sempre me apoiaram e me encorajaram na busca dos meus objetivos. Seu apoio, amor e compreensão foram fundamentais ao longo dessa jornada acadêmica.

A todos meus sinceros agradecimentos.

Sumário

1. Introdução 8

1.1 *Stenostemum leucops* 8

1.2 Regeneração 10

1.3 Os microRNAs..... 11

1.4 Os microRNAs e a regeneração em planárias 14

2. Objetivos 15

2.1 Gerais 15

2.2 Específicos 15

3. Materiais e Métodos 16

3.1 Obtenção dos dados genômicos e transcriptômicos..... 16

3.2 Predição gênica..... 18

3.3 Anotação 19

3.4 Filogenia das proteínas *Argonauta* e *Piwi*..... 19

3.5 Expressão diferencial..... 20

4. Resultados..... 20

4.1 Sequências de *Argonauta* e *Piwi*..... 20

4.2 Análise filogenética dos genes *Argonautas* e *Piwis*..... 24

4.3 Expressão diferencial dos genes..... 25

5. Discussão..... 30

6. Conclusões..... 31

7. Material Suplementar..... 33

8. Referências 37

LISTA DE ABREVIATURAS E SIGLAS

DNA – ácido desoxirribonucleico

gfp – green fluorescent protein

mRNA – RNA mensageiro

microRNA- Pequenos RNAs não codificatantes

RNA – ácido ribonucleico

siRNA – pequeno RNA interferente

piRNA – RNA de interação com *Piwi*

RISC – *RNA-induced silence complex*

RNAi – RNA de interferência

DGCR8 - *DiGeorge Syndrome Critical Region 8*

pb – pares de bases

Kb - kilobase

HMM - *Hidden Markov Model*

GFF – *General Feature Format*

CDD - *Conservated Domains Database*

NCBI - *National Center for Biotechnology Information*

Resumo

O gênero *Stenostomum* é composto por microturbelários de vida livre pertencentes a classe Catenulida, sendo este o grupo mais basal dos *Platyhelminthes*, são cosmopolitas, distribuídos globalmente. A maioria das espécies se reproduz assexuadamente por paratomia pela formação de zooides, novos vermes crescem a partir da extremidade posterior do animal, podendo formar uma cadeia de indivíduos. Esses animais podem ser apontados como potenciais organismos modelos devido a fácil manutenção em laboratório principalmente por apresentarem reprodução assexuada e grande capacidade regenerativa. Os microRNAs são pequenos RNAs não codificantes que têm importantes funções na regulação gênica. Podem ser divididos em dois grupos dependo com a proteína com que interagem: siRNAS que interagem com *Argonauta,* silenciam gene-alvos em níveis transcricionais e pós-transcricionais; e piRNAs que interagem com as proteínas da família *Piwi*, que protegem o genoma da atividade mutacional dos elementos transponíveis na linhagem germinativa em animais. Os microRNAs podem influenciar de diferentes formas durante o processo de regeneração em planárias e na manutenção da do estado de totipotência celular. Desta forma, uma análise detalhada da composição genômica, em *Stenostemum leucops,* dos genes das famílias *Argonauta* e *Piwi*, podem contribuir para o entendimento do papel dos miRNA no estado de totipotência celular dos *Platyhelminthes* e se esta característica já estava presente no grupo mais basal. Para o estudo foram obtidas sequências de proteínas *Argonauta* e *Piwi* de outros organismos do grupo de *Platyhelminthes* e grupos externos, para comparação com de *S. leucops,* essas foram obtidas através da revisão de literatura e bases de dados online (WormBase Parasite, NCBI e FlyBase). Os genomas e transcriptomas de *S.leucops* foram sequenciados pelo nosso grupo, o genoma através de *long-reads* (Nanopore) e short-reads (Ilumina). O transcriptoma para um e dois zooides (Iontorrent S5). Para predição dos genes foiutilizado *software* Augustus, com treinamento das respectivas sequências de proteínas para pos-terior localização no genoma de *S.leucops*. As sequências encontradas foram ainda anotadas manualmente utilizando o *software* Ugene e funcionalmente com eggNOG mapper. Para a iden-tificação de domínios conservados foi utilizado Batch-Web CDD-Search (NCBI). Para análises filogenéticas, as sequências foram alinhadas e refinadas, Mr. Bayes e IQTree foram utilizados para construção das árvores. Foram identificadas três sequências da família Argonauta, e duas da subfamília Piwi. Todas as sequências de Argonauta e Piwi revelaram a presença dos dominios proteicos característicos da família, nenhum splicing alternativo foi encontrado. No contexto filogenético, as sequências de Argonauta *de S. leucops* formaram um clado distintivo junto a representantes de planárias de vida livre, moluscos, trematódeos e cestodas, enquanto sequências de Piwi se agruparam de forma robusta com organismo que compartilham Piwi1. Não foi encontrada diferença na expressão dos genes *argonauta* e *piwi* em animais com I e II zooides em *Stenostemum*.

Palavras-chaves: Argonauta; Piwi; Platelmintos.

Abstract

The genus *Stenostomum* is composed of free-living microturbularians belonging to the Catenulida class, which is the most basal group of *Platyhelminthes*. They are cosmopolitan, distributed globally. Most species reproduce asexually by paratomy through the formation of zooids, new worms grow from the posterior end of the anima and can form chains of individuals. These animals can be identified as potential model organisms due to their easy maintenance in the laboratory, mainly because they have asexual reproduction and great regenerative capacity. MicroRNAs are small non coding RNAs that have important functions in gene regulation and can be divided into two groups depending on the protein they interact with: i) *Argonautes* that silence target genes at transcriptional and post transcriptional levels; ii) *Piwi* family proteins that protect the genome from the mutational activity of transposable elements in the germline in animals. MicroRNAs can influence the regeneration process in planarians in different ways. In this way, a detailed analysis of the genomic composition, in *S. leucops*, of the genes from the *Argonaute* and *Piwi* families, can contribute to the understanding of the role of miRNAs in the state of cellular totipotency of *Platyhelminthes* and whether this characteristic was already present in the most basal group. For the study, *Argonaute* and *Piwi* protein sequences were obtained from other organisms from the *Platyhelminthes* group and external groups, for comparison with those from *Stenostomum leucops*, these were obtained through literature and online databases (WormBase Parasite, NCBI and FlyBase). Our group, the genome using long-reads (Nanopore) and short-reads (Ilumina), sequenced the genomes and transcripts of *S.leucops*. The transcriptome for one and two zooids (Iontorrent S5). Augustus software was used to predict genes, with training of the respective protein sequences for subsequent location in the *S.leucops* genome. The sequences found were also annotated manually using the Ugene software and functionally with eggNOG mapper. To identify conserved domains, Batch-Web CDD-Search (NCBI) was used. For phylogenetic analyses, the sequences were aligned and refined, Mr. Bayes and IQTree were used to build the trees. Three sequences from the *Argonaute* family were identified, and two from the *Piwi* subfamily. All *Argonaute* and *Piwi* sequences revealed the presence of the protein domains characteristic of the family, no alternative splicing was found. In the phylogenetic context, *Argonaute* sequences from *Stenostemum leucops* formed a distinctive clade alongside representatives of free living planarians, molluscs, trematodes and cestodes, while *Piwi* sequences grouped robustly with organisms that share Piwi1. No difference was found in the expression of the *argonaut* and *piwi* genes in animals with I and II zooids in *Stenostemum*.

Keyworlds: Argonaute; Piwi; Platyhelminthes.

1. Introdução

1.1 - Stenostomum leucops

Platyhelmintes são animais triblásticos, caracterizados por serem achatados dorsoventralmente, acelomados e terem simetria bilateral. Atualmente compreendem 26.500 espécies, podendo ser encontrados principalmente em ambientes aquáticos e úmidos, com alguns representantes que conseguem habitar ambientes terrestres. Sem possuir uma proteção externa, podem assumir várias formas e tamanhos, desde microscópios até 30 cm de comprimento (Brusca et al, 2018).

O filo é monofilético e se divide em quatro classes: Turbellaria, Monogenea, Cestoda e Trematoda (Adell e Ruitort, 2021). Catenulida e Rhabditophora formam o grupo Turbellaria, sendo composto por organismos de vida livre, que habitam ambientes de água doce como pântanos, lagoas, riachos e habitats terrestre úmidos, distribuídos globalmente (Larsson e Jondelius, 2008).

A classe Catenulida é considerado o grupo mais basal dentro dos platelmintos. Atualmente cerca de 110 espécies são conhecidas. São organismos com poucos milímetros de tamanho, sendo conhecidos como microturbelários (Larsson e Willems, 2010). Apresentam uma cavidade bucal anterior, gânglio nervosos, podem apresentar estatocisto e um protonefrídio único biflagelado(Brusca et al. 2018).

Platelmintos do gênero *Stenostomum* (Figura 1) são microturbelários de vida livre pertencentes a classe Catenulida, existem 50 espécies conhecidas, das quais 25 podem ser encontradas na América do Sul. São animais cosmopolitas, identificados por ter um corpo acahtado de 0,5-2 mm de comprimento podendo formar cadeias de indivíduos (Noreña, 2005, Rosa et al. 2015). Possuem uma anatomia simples, com presença de gânglios cerebrais pré orais lobulados, excretóforos, fossas ciliadas e a sua reprodução é assexuada por paratomia (Noreña, 2005). Os poucos caracteres morfológicos podem se mostrar um desafio para determinação de espécie.

A maioria das espécies se reproduz assexuadamente por paratomia, pela formação de zooides, novos vermes crescem a partir da extremidade posterior do animal. Na paratomia o organismo se divide ao longo de um eixo perpendicular ao eixo ântero-posterior. A diferença entre a paratomia e a regeneração está na formação de uma "pré-geração" na qual há formação de um cérebro (Egger et al, 2007). Assim, cada zooide se desenvolve e se separa do corpo do animal primário dando origem a um novo verme, possuindo um desenvolvimento direto (Larsson e Willems, 2010). *Stenostemun leucops*, leva cerca de 42,5 horas, a uma temperatura de 28º

para realizar a paratomia completa, resultando em zooides contendo 2.400 células (Rosa e Loreto, 2021). Esses animais podem ser apontados como potenciais organismos modelos devido a fácil manutenção em laboratório (Rosa et al. 2015).

Stenostemum leucops também apresenta uma característica singular de poder expressar genes contidos em DNAs exógenos. Estudos de Rosa e Loreto (2019) demonstraram que esse organismo possui a capacidade de expressar genes contido em DNAs presentes no microbioma e, até mesmo quando moléculas de DNA são ingeridas diretamente do meio ambiente. Nesse estudo foi acompanhado a expressão de um gene repórter marcado por *gfp* introduzido por eletroporação nos animais, também foram testados plasmídeos transportados por bactérias e a ingestão de DNA nu. A avaliação ocorreu em relação a estabilidade de expressão do gene repórter e a manutenção do plasmídeo no verme (Rosa e Loreto, 2019). Os resultados indicaram que quando o plasmídeo era mantido no microbioma ele era expresso de maneira estável, mas quando ingerido o DNA nu ou de bactérias não mantidas no microbioma a expressão era transitória, demonstrando que o microbioma pode afetar a capacidade do verme em expressar determinados genes (Rosa e Loreto, 2019). Essa particularidade chama atenção para esse organismo, reforçando o potencial para utilização em pesquisas.

A interação entre organismo hospedeiro e seu microbioma pode contribuir para plasticidade fenotípica, impactando a capacidade de sobrevivência e evolutiva do organismo. Um holobionte é uma associação composta por um hospedeiro e seus microrganismos simbióticos associados, formando uma comunidade estável (Morris, 2018). Esses microrganismo que compõem o microbioma podem residir dentro ou fora do corpo e juntos formam uma unidade funcional. Eles contribuem para funções fisiológicas, proteção de patógenos e podem influenciar no comportamento e desenvolvimento do hospedeiro (Archie e Theis, 2011; Bordenstei e Theis,2015). O microbioma pode ser uma fonte variada de recursos, incluindo nutrientes, probióticos e metabólicos secundários (Promi et al, 2019). Isso por sua vez tem impacto na fisiologia e nos mecanismos de defesa do holobionte (Rosa e Loreto, 2019). A interação de *Stenostemum leucops* com seu microbioma e a plasticidade fenotípica foi descrito por Rosa e Loreto (2023).

Outros platelmintos que são comumente utilizados como organismos modelos, são as planárias de água doce. Principalmente as da ordem Tricladida como *Schmidtea mediterranea* e *Dugesia japonica,* que mostram a utilidade desses organismos tanto em estudos básicos de zoologia como nos estudos de genética e biologia molecu

lar. Especialmente em estudos que podem fornecer novas informações sobre processos de desenvolvimento não embrionários, estudos de regeneração, pluripotência celular e homeostase tecidual, que são de difícil acompanhamento em outros modelos de invertebrados (Eisenhoffer et al, 2008; Wagner et al, 2011; Guedelhoefer e Alvarado, 2012).

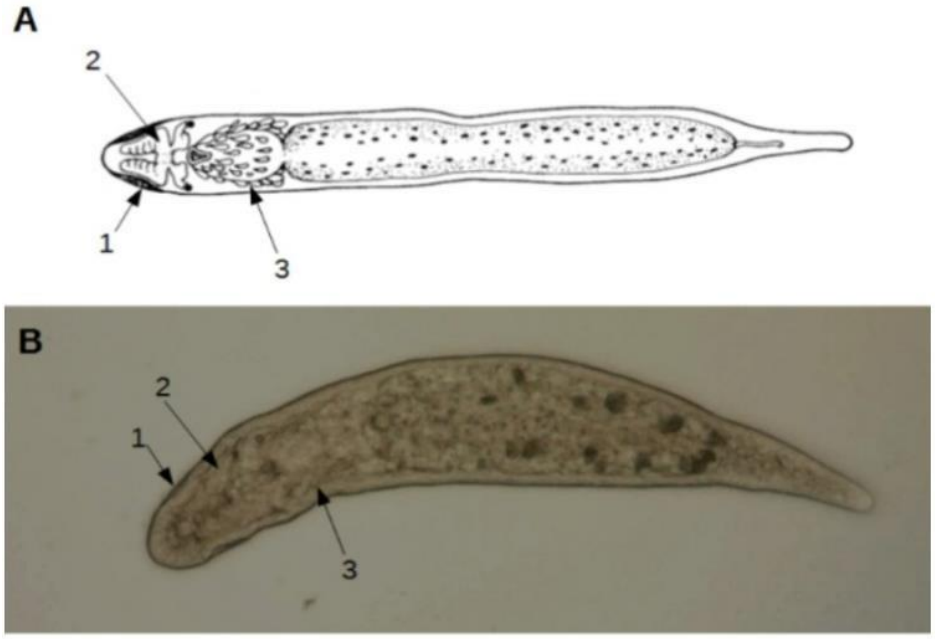


Figura 1. *Stenostemum leucops*. (A) representação anatômica correspondendo a 1:fossas ciliares, 2:glânglios cerebrais e 3: glândulas da faringe. (B) fotografia de *Stenostemum leucops*. Fonte*:* Noreña (2005) e Rosa et al. (2015)

1.2 – Regeneração

As planárias são popularmente conhecidas por sua alta capacidade de regeneração, desde reposição de partes faltantes do corpo até a regeneração completa do organismo. Essa capacidade, é dada pela presença de células-tronco, somáticas diferenciadas, chamadas de neoblastos, que estão presentes por todo o corpo, representado até 30% da composição corporal (Baguña et al, 1989). Possuem a característica de serem pluripotentes, sendo essenciais para renovação celular, manutenção da homeostase e regeneração (Ustyantsev et al, 2021).

Existem dois tipos de regeneração: A bidirecional, onde o animal é capaz de se regenerar em ambas as direções e monodirecional, onde uma estrutural distal, igual ou semelhante, pode ser regenerada. Elas ainda podem ser diferenciadas em epimórfica e morfaláxica. A regeneração epimórfica envolve a formação e crescimento de novas células para produzir a parte regenerada, enquanto a regeneração morfoláxica acontece através do rearranjo das células existentes (Ridden e Alvarado, 2004; Slack, 2010).

A regeneração bidirecional apresenta um controle de polaridade que é essencial para decisão da formação da cabeça ou da cauda (Ridden e Alvarado, 2004). Este tipo de regeneração também está envolvido na forma de reprodução assexuada, e renovação contínua de células em organismos como as planárias (Slack, 2010).

A capacidade de regeneração da família Stenostomidae tem sido debatida ao longo do tempo. Um estudo mais recente demonstra que o processo de regeneração de *Stenostemum leucops* é complexo e depende da parte do corpo envolvida. Diferentes rotas de regeneração podem ocorrer, incluindo absorção, migração ou até desenvolvimento de novas partes cefálicas (Rosa e Loreto, 2021).

Na região anterior do animal é possível que novas estruturas cefálicas sejam formadas, enquanto na região posterior a estrutura cefálica se desenvolve completamente, formando um verme com dois zooides. Nessa região também pode ocorrer o desenvolvimento de novas estruturas e absorção na região posterior. Na região posterior, pode se observar os desenvolvimentos de estruturas cefálicas ou a migração da região posterior para a anterior do zooide. Normalmente em vermes é possível observar até dois zooides, porém neste animal já foi observado a formação de cadeia com cinco zooides, dependendo das condições ambientais (Rosa e Loreto, 2021).

1.3– Os microRNAs

Os microRNAs fazem parte da classe de RNAs não codificantes de proteínas, e possuem a função principal de regulação da expressão gênica e de manter a estabilidade do genoma. Eles exercem sua função regulatória ao se ligarem à região 3’ não traduzida dos RNAs mensageiros (mRNA) alvo (Wu et al, 2020). Esse mecanismo resulta na diminuição dos níveis de proteína dos genes alvo (Filho e Kimura, 2006).

Os primeiros microRNAs foram descobertos através de estudos genéticos em *Caenorhabditis elegans*. Os pesquisadores estavam interessados em descobrir como o gene *lin-4* estava envolvido na regulação de desenvolvimento larval do nematoide.

O gene *lin- 4* é necessário para transição de desenvolvimento do estágio larval de L1 para L2 (Chalfie et al, 1981). Através do isolamento de uma mutação nula no gene *lin-4* os pesquisadores observaram que os animais apresentavam falha no desenvolvimento, repetiam processos larvais iniciais em estágios lavais tardios, não desenvolviam algumas estruturas adultas, e não conseguiam botar ovos (Chalfie et al, 1981; Almeida et al, 2011). Em 1987 Ferguson et al,

descobriram que uma mutação supressora no gene *lin-14* era capaz de reverter o fenótipo do efeito mutacional de *lin-4*. Dessa forma eles supuseram que *lin-4* poderiam estar regulando negativamente *lin-14* (Lee et al, 1993).

Em 1993, Lee et al descobriram que o gene *lin-4* não codificava uma proteína, mas sim pequenos RNAs de 22 a 61 nucleotídeos. Esses RNAs pequenos eram complementares a região repetida 3' UTR do mRNA do gene *lin-14*, e ligavam-se por uma interação RNA-RNA antis-sense. Com essa ligação havia a inibição da tradução do mRNA de *lin-14*, reduzindo a quantidade de proteína (Lee et al, 1993). Com base nisso, foram iniciados vários estudos para identificar microRNAs em diferentes grupos de animais e plantas.

A biogênese dos microRNAs envolve a ação de duas proteínas RNases III, nuclear e citoplasmática respectivamente, Drosha e Dicer. A maioria dos transcritos de microRNAs primários são gerados pela transcrição da polimerase II, não sendo descartado que outras polimerases podem estar envolvidas na transcrição de microRNAs. Esses microRNAs primários contém uma estrutura em haste-alça que é clivada pela proteína Drosha em conjunto com a proteína DGCR8, em invertebrados como *Caenorhabditis elegans* e *Drosophila melanogaster* é conhecida como Pasha. A junção dessas duas proteínas formam o complexo microprocessador, que gera os pré-RNAs. Posteriormente os pré-RNAs são exportados para o citoplasma pela exportina-5 (exportina-1em invertebrados) e processados pela Dicer que os cliva gerando duplexes. As fitas são separadas, sendo uma delas selecionada como microRNA maduro, enquanto a outra é degrada (Kim, 2005; Swarts et al, 2014; Fontenla et al, 2017).

Os microRNAs podem ser divididos em dois grupos, de acordo com as proteínas que interagem: O grupo de RNAs interferentes (siRNAs) que interagem com proteínas *Argonauta*s, agindo na regulação de biogênese de mRNAs e na defesa contra agente externos, e o grupo que interage com proteínas *Piwi* (piRNAs), responsáveis pela proteção dos efeitos mutacionais ocasionados pelos transposons, principalmente na linhagem germinativa dos animais, e também na defesa contra RNAs exógenos (Girard et al, 2006; Griffthis et al, 2013; Ugarte et al, 2023).

As proteínas *Argonauta*s são divididas em duas subfamílias, AGO E PIWI. Os siRNAs e microRNAs interagem com AGO formando o complexo RISC (RNA-*induced silencing complex*) que reprime seus alvos a partir da complementariedade de sequências, clivando o mRNA-

alvo ou direcionando a outras proteínas para a degradação (Figura 2B).

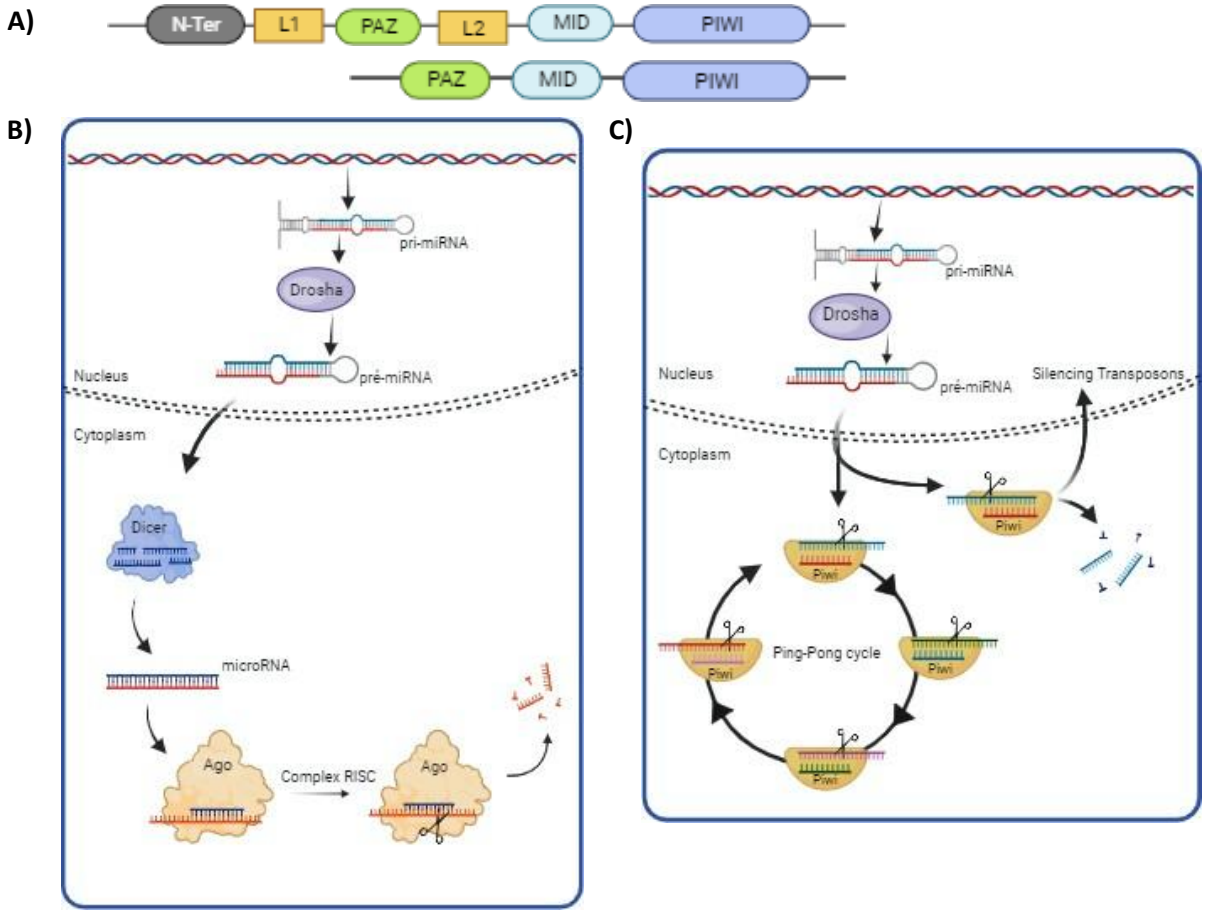


Figura 2. Vias de silenciamento de microRNA que interagem com AGO e PIWI. Em (A) A estrutura de domínios das proteínas *Argonauta*s e *Piwi* encontradas em eucariotos. Em (B) Os siRNAs ou miRNAs originados do processamento de Drosha são clivados pela Dicer se associam as proteínas da família *Argonauta* formando o complexo RISC. Quando encontram um mRNA complementar, ocorre a clivagem deste pelo complexo. Em (C) Os piRNAs que ao interagirem com a proteína *Piwi* podem passar por modificações no ciclo "ping-pong", gerando piRNAs secundários ou clivar um mRNA alvo diretamente. Ambos os mecanismos podem reconhecer uma sequência de mRNA complementar, resultando na degradação deste ou então recrutar metilases para desativação de transposons no genoma. Fonte: Autor

Diferentes domínios podem estar envolvidos na interação com RNAs-alvo, as *Argonauta*s são compostas por 4 domínios sendo eles N-terminal, MID, PAZ e PIWI, esses domínios são interligados por Linker 1 e 2 (Figura 2A), diferentes famílias dentro das *Argonauta*s podem ocorrer por variações entre os domínios, essas características estruturais são importantes na

função exercida pelas *Argonauta*s na regulação gênica e defesa contra patógenos (Jin et al, 2021).

As proteínas *Piwi* possuem dois mecanismos de reprimir a atividade de transposons e de mRNAs (Figura 2C). 1- Através do ciclo de "ping-pong": onde um piRNA-guia se liga uma proteína *Piwi*, formando um complexo, o qual encontra o mRNA-alvo e o cliva, este se torna um novo piRNA, chamado piRNA-alvo que pode formar novos complexos com proteínas *Piwi*. 2- Pela complementariedade, quando um piRNA está ligado a um transposon, ele sinaliza para proteínas *Piwi* o alvo que deve ser silenciado. A proteína *Piwi* então recruta outras proteínas com atividade de metilase impedindo que o transposon de mova ou se ative. O ciclo "ping-pong" gera uma amplificação exponencial de piRNAs, aumentando a eficiência contra elementos genéticos móveis (Loubalova et al, 2023).

1.4- Os microRNAs e a regeneração em planárias

Alguns estudos destacam relação entre a capacidade regenerativa dos platelmintos e a reprodução assexuada. Solana et al (2012) demonstrou que quando os neoblastos são eliminados por radiação ou por RNAi, não há regeneração e renovação celular levando o animal a morte.

Além disso também parece haver correlação entre a plasticidade celular e a expressão da proteína *Piwi* (Solana et al. 2012; Wolfswinkel van, 2014). Principalmente em planárias, onde essas proteínas são encontradas nos neoblastos que estão dispersos por todo o espaço mesodérmico. Estudo com *Schmidtea mediterrânea* demonstrou que diferentes genes *piwis* podem ser expressos em diferentes populações de neoblastos, e quando nocauteados podem gerar diferenças significativas na regeneração desses animais (Kashima et al, 2020).

Os microRNAs podem influenciar de diferentes formas durante o processo de regeneração em planárias, na estabilidade genômica prevenindo mutações indesejadas, influenciar na proliferação, diferenciação e apoptose celular, contribuindo para a reconstrução de tecidos, modular sinais de lesão gerando estímulos regenerativos, assim como impactar redes complexas em planárias influenciando múltiplas vias de sinalização (Tian et al, 2012; Pawlika et al, 2016; Cao et al, 2020).

Dessa maneira, examinar de forma detalhada a composição genômica dos genes das famílias *argonauta*s e *piwi* em *Stesnostemum leucops*, pode oferecer um entendimento mais profundo sobre o papel dos microRNAs na pluripotênca celular dos *Platyhelminthes* e se essa característica já estava presente no grupo ancestral mais antigo. Além disso, a complementação

com dados do transcriptoma focando na expressão das famílias de genes *argonauta* e *piwi*, em diferentes estados de desenvolvimentos de *Stenostemum leucops,* pode enriquecer nossa compreensão desse fenômeno.

2. Objetivos

2.1- Objetivos Gerais

Investigar com maior detalhe as sequências genômicas e transcriptômicas de membros das subfamílias *Piwi* e *Argonauta* em *Stenostomum leucops.*

2.2- Objetivos específicos

- Localizar sequências dos genes *piwi* e *argonauta* no genoma e transcriptôma de *Stenostemum leucops.*
- Analisar filogeneticamente as sequências de proteínas *Piwi* e *Argonauta* dentro do grupo *Platyhelminthes.*
- Analisar a expressão dos genes *piwi* e *argonauta* em *Stenostemum leucops* em diferentes estágios de um zooide e dois zooides.

3. Materiais e métodos

3.1 – Obtenção dos dados genômicos e transcriptômicos

O genoma e transcriptoma de *Stenostemum leucops* foram previamente sequenciados pelo nosso grupo de pesquisa (Rosa, 2022). Para a montagem do genoma foi utilizado um sequenciamento "*long reads*" com a tecnologia Oxford Nanopore, resultando em 433.104 *reads*, que apresentaram uma variação de tamanho entre 112 a 87.259 pares de bases (pb). A montagem revelou uma cobertura baixa, com média de 5,5 vezes, gerando 5.038 *contigs*. Dentre eles, 3.236 *contigs* têm mais de 10 Kb de extensão, com um N50 de 220.186 pb e o maior *contig* 1.164.702 pb. O genoma montado pelo software Fly (Lin et al, 2016) foi estimado em 296.492.128 pb.

Foram sequenciados transcriptomas de vermes com um e dois zooides pela plataforma S5 (IonTorrent). Foram gerados 80.831.095 *reads*, e a abundância desses *reads* foi estimada utilizando o software Kallisto (Bray et al, 2016). A montagem do trasncriptoma foi realizada pelo software Trinity (Grabherr et al, 2011) resultando em 279.881 *contigs*, onde o N50= 849. As regiões codificadoras foram anotadas com auxílio do software Augustus (Stanke e Waack, 2003) o qual foi treinado com genomas de *Macrostomum lignano*, *Caenorhabditis elegans* e *Schistosoma mansoni*.

As sequências de proteínas *Argonauta* e *Piwi* de outros organismos do grupo de *Platyhelmithes* e grupos externos (Tabela 1), para comparação com proteínas de *Stenostomum leucops*, foram obtidas através do artigo de Fontenla et al (2021) e bases de dados online como WormBase Parasite, WormBase, Uniprot e NCBI.

Tabela 1. Espécies utilizadas na Filogenia de proteínas *Argonauta* e *Piwi*.

Espécie	Acesso *Piwi*	Acesso *Argonauta*	Fonte
Stenostemum leucops	PRJNA1037460	PRJNA1037460	NCBI
Macrostomum lignano	PAA69384.1 PAA89917.1 PAA53062.1 ALD16252.1	ALD16264.1 ALD16264.1 PAA83744.1 PAA81224.1	NCBI
Prostheceraeus vittatus	Pvi28529 Pvi7924	Pvi28319	Fontenla et al (2021)

	Pvi28261 Pvi47833		
Geocentrophora aplanate	Gap14751 Gap14376	Gap26895 Gap	Fontenla et al (2021)
Rhyncomesostoma rostra-tum	Rro24399 Rro43774 Rro28014	Rro52833 Rro52813	Fontenla et al (2021)
Monocelis fusca	Mfu45028 Mfu83791 Mfu59068	Mfu46046 Mfu48140	Fontenla et al (2021)
Kronborgia amphipodicola	Kam1053 Kam3316 Kam3362	Kam15076	Fontenla et al (2021)
Bothrioplana semperi	Bse2112 Bse40309	Bse53849	Fontenla et al (2021)
Schmidtea mediterranea	Q2Q5Y9.1 Q2Q5Y8.1 ACC97187.1	ADY05336.1	NCBI
Fasciolopsis buski	-	KAA0184342	NCBI
Schistosoma mansoni	-	XP_018653609.1	NCBI
Hymenolepis microstoma	-	CDS26333.1 CDS30752	NCBI
Echinococcus multiocularis	-	CDS41461.1 CDS41463.2	NCBI

		KAH9283566.1 CDS36376.1	
Ceanorhabditis elegans	CE09083 NM_068593.6	CE31525 CE32063	WormBase/NCBI
Drosophila melanogaster	AF104354.1 NM_001043162.3 NM_057386.5	NM_079010.4 NM_140518.3	NCBI
Lottia gigantea	V3ZNA5 V4B7N4	V4B7N4 V3ZNA5 V3ZRX0	UniProt
Octopus bimaculoides	A0A0L8I6X6 A0A0L8HBN2	A0A0L8GBC7 A0A0L8GBV4 A0A0L8HBN2 A0A0L8I6X6	UniProt
Homo sapiens	XP054229831.1 NP001129139.1	NM_012199.5 XM_011516968	NCBI

Fonte: Autor

3.2 – Predição gênica

Para a predição gênica foi utilizado o software Augustus. Inicialmente, foi realizado um treinamento preliminar com genes das proteínas *Argonauta* e *Piwi* para identificá-las no genoma de *Drosophila melanogaster*. Após, o programa foi treinado novamente com o mesmo grupo de proteínas e retreinado, sendo aplicado na busca por genes no genoma de *Stenostemum leucops.*

O Augustus é um software para predição e anotação gênica *ab initio* em sequências eucarióticas, que utiliza modelo estatístico Oculto de Markov (HMM), e algoritmos de aprendizado de máquina para realizar as predições de gene num genoma de interesse. Após um treino inicial do modelo, com um conjunto de dados conhecidos, o programa analisa a sequência de DNA e

procura por regiões que correspondam a padrões típicos de genes. A saída gerada pelo Augustus é no formato GFF, indicando previsões de localização dos exóns, introns, regiões codificadoras e transcritas, além de fornecer informações sobre a confiabilidade (Stanke et al, 2004).

3.3 – Anotação

Após a predição gênica, realizamos uma filtragem para identificar os genes correspondentes e sua localização nos *contigs*. Esse processo envolveu o auxílio de Blastn para identificar os *contigs* e os respectivos genes. Esses genes foram separados para uma anotação manual utilizando o software Ugene (Okonechnikov et al, 2012), usando as saídas GFF geradas pelo Augustus para localizar e visualizar os genes, permitindo a análise da região, observação do tamanho e disposição das sequências gênicas.

Também foi realizada anotação funcional das sequências de proteínas encontradas no genoma de *Stenostemum leucops*, utilizando EggNOG-Mapper (Cantalapiedra et al, 2021). Além disso, domínios conservados foram identificados usando a ferramenta online CDD-Search (Shennan Lu et al, 2020) e as possíveis regiões promotoras foram localizadas pelo software MEME Suite (Bailey et al, 2015). Essas ferramentas se baseiam na similaridade entre sequências para oferecer uma compreensão detalhada sobre as funcionalidades e características biológicas das proteínas analisadas.

3.4 – Filogenia das proteínas *Argonauta* e *Piwi*

Para a análise filogenética, um conjunto de 25 sequências de proteínas *Argonauta*, representando 16 espécies e 35 sequências de proteínas *Piwi*, representando 12 espécies, foram alinhadas pelo programa MEGA X (Tamura et al, 2021) pelo alinhador ClustalW. Nessa etapa, a árvore filogenética foi construída considerando apenas táxons de platelmintos de vida livre. Posteriormente foi incorporado um segundo conjunto de sequências englobando táxons de outros grupos como, moluscos, cordados e platelmintos parasitas. Com essa inclusão, o conjunto expandido totalizou 41 sequências de proteínas *Argonauta*, abrangendo 19 espécies e 39 sequências de proteínas *Piwi* representando 15 espécies.

Em seguida, as sequências foram submetidas ao software Gblock (Tavalera e Castresana, 2007), afim de obter apenas blocos conservados. A determinação do modelo evolutivo foi realizada utilizando o software Modeltest-NG (Darriba et al, 2020) para todas as análises.

Para reconstrução da árvore filogenética, empregamos a análise Bayesiana (Material suplementar) e a Máxima Verossimilhança. Na análise por Máxima verossimilhança foi utilizado o programa IQtree (Minh et al, 2020) com 1.000 repetições de bootstrap, o software Mr. Bayes

(Huelsenbeck e Ronquist, 2001) foi utilizado para análise bayesiana, iniciando com 1.000.000 de gerações e adicionando gerações conforme necessário para aprimorar a convergência do resultado. As árvores filogenéticas foram visualizadas pelo software online ITOL (Ciccarelli FD, et al, 2006)

3.5 – Expressão diferencial

A análise comparativa de expressão entre animais com um e dois zooides, foi realizada a partir da expressão geral de todos os genes pelo programa Edge R (Robinson et al, 2010). Dentre a expressão total, foi verificada a expressão dos genes de proteínas *Argonauta* e *Piwi* nos diferentes estágios.

4. Resultados

4.1 Sequências de *Argonauta* e de *Piwi*

O genoma e transcriptoma de *S. leucops* foi montado conforme descrito em Rosa (2022). As estatísticas descritivas das montagens podem ser encontradas na Tabela 2 e 3, contendo os números de *reads* usados em cada montagem, o número de *contigs* obtidos, N50 e tamanho estimado do genoma. As análises de representatividade gênica realizada pelo software BUSCO mostrou uma boa completude gênica em que aproximadamente 70 e 80% dos genes "core" de eucariontes estão representados no genoma e transcriptoma, respectivamente.

Tabela 2. Estatísticas descritivas do genoma de *S. leucops* empregando a tecnologia Oxford Nanopore

Número de "reads"	433.104	**Análise pelo Busco**
Tamanho dos "reads"	112-87.259 pb	Completos 59.9 % copia única 57.4% completo e duplicado 2.5%
Número de contigs	5.038	Fragmentados 11.9 %
N50	220.186 pb	Ausentes 28.2%
Maior contig	1.164.702 pb	
Tamanho Estimado do genome	296.492.128 pb	Cobertura do genoma 5X

Tabela 3. Estatísticas descritivas do transcriptoma de *S. leucops*

Número de “reads"	80.831.095	**Análise pelo Busco**
Tamanho dos “reads”	40-350 pb	Completos 66.4% cópia única 30.6% completo e duplicaado 35.8%
Número de contigs	279.881	Fragmentados 14.0%
N50	849 pb	Ausentes 19.6%,

Fonte:Autor

Em nossas análises empregando o software Augustus no genoma de *S.leucops* foram encontradas três sequências da família *Argonauta,* correspondendo a duas sequências localizadas no *contig* 7 e uma sequência no *contig* 2375. As sequências apresentam aproximadamente 4Kb. Para a subfamília *Piwi* foram encontradas duas sequências, localizadas nos *contigs* 2714 e 5350. A sequência contida no *contig* 2714 tem aproximadamente 6.2 Kb e no *contig* 5350 a sequência tem aproximadamente 5.8 Kb (Tabela 4).

Tabela 4. Caracterização de sequências encontradas em *S.leucops*. São mostradas informações sobre as sequências encontradas como localização, o gene correspondente, o tamanho da sequência e orientação da sequência na fita.

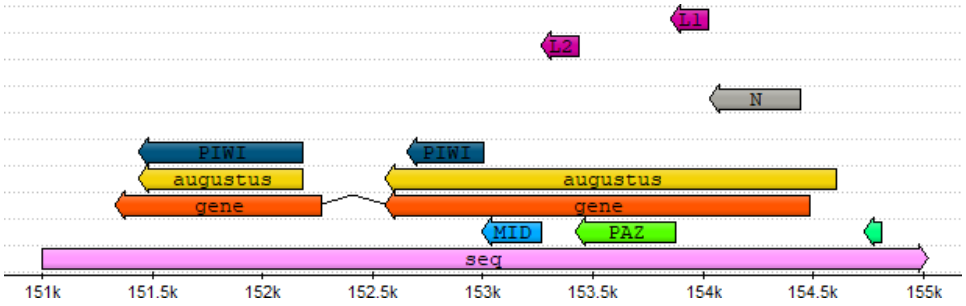
Localização	Gene	Tamanho	Orientação
contig 7a	*argonauta*	4Kb	(-)
contig 7b	*argonauta*	3Kb	(+)
contig 2375	*argonauta*	4Kb	(-)
contig 2714	*piwi*	6.2Kb	(+)
contig 5350	*piwi*	5.8Kb	(+)

Fonte: Autor

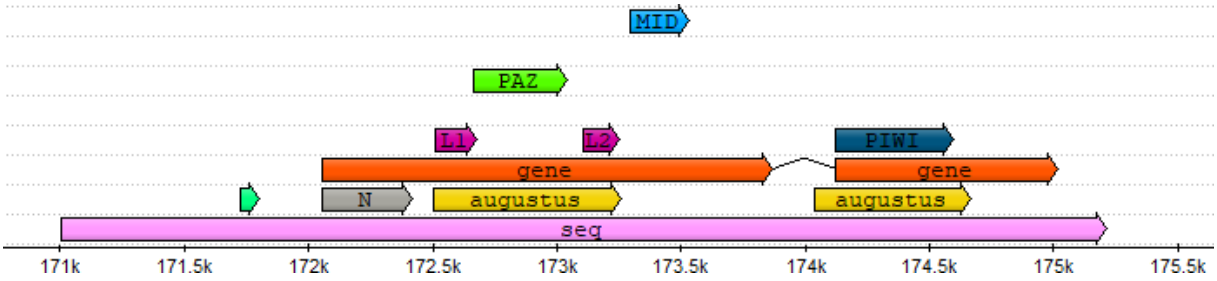
Comparamos as sequências genômicas com aquelas obtidas no transcriptoma de *S.leucops,* visando estimar com mais precisão as prováveis proteínas codificadas por estes genes. Nas Figuras 3 e 5 é possível visualizar detalhes das regiões codificadoras encontradas pelo software Augustus, as deduzidas a partir dos transcritos produzido pelos genes (transcriptomas), os domínios proteicos codificados pelos genes, identificados por bioinformática, assim como as regiões regulatórias dos genes como, por exemplo, possíveis promotores.

Os dois genes *Argonauta* encontrados no *contig* 7 apresentam dois éxons. A sequência do *contig* 2375 apresenta três éxons e também não foi encontrado *splicing* alternativo. Em todas as sequências de *Argonauta* encontradas foram identificadas todos os domínios proteicos característicos dessa família de proteínas, sendo eles N, L1, PAZ, MID, L2 e PIWI.

Sequência 1- *contig* 7a



Sequência 2- *contig* 7b



Sequência 3 – *contig* 2375

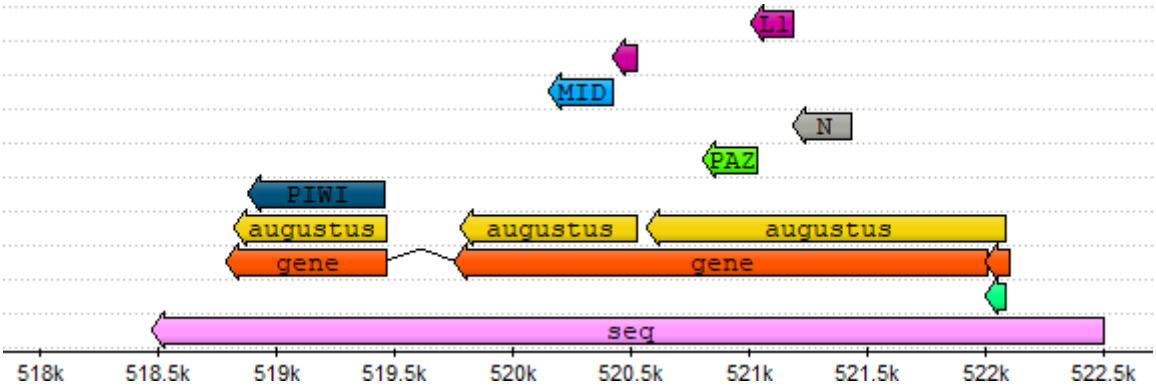


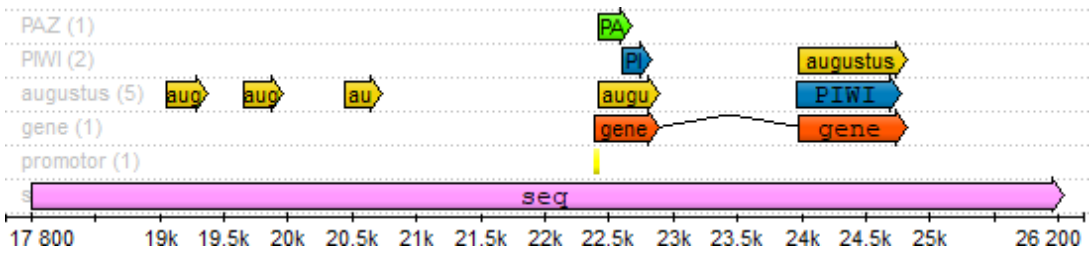
Figura 3. Sequências de *Argonauta*s. Na imagem, a região em rosa é representada como a região genômica abrangendo o gene em estudo. Em amarelo são as regiões codificantes que foram identificas pelo software Augustus e em laranja o transcrito gerado identificado no transcriptoma, em verde claro potenciais promotores. Os domínios referente a família *Argonauta* estão representados pelas cores: cinza: N, roxo: Linker 1 e 2, verde: PAZ, azul escuro: PIWI e azul claro: MID. Fonte: Autor.

Na figura 4 encontra-se os alinhamentos das proteínas *Argonauta*s de *S.leucops*. A divergência genética encontrada entre as proteínas do *contig* 7 é de 10,3 % e a divergência genética encontrada entre as proteínas do *conitg* 7 e a proteína do *contig* 2375 é de 36 %.

Figura 4. Alinhamento de proteínas *Argonautas* encontradas em *S. leucops*. Fonte:Autor

Foram encontradas duas sequências da subfamília *Piw*i, estando localizadas nos *contigs* 2714 e 5350 (Figura 5). A sequência correspondente ao *contig* 2714 apresenta 5 regiões codificantes identificadas pelo software Augustus, porém apenas duas dessas apresentam sequências correspondentes a proteínas *Piwi*. O gene encontrado no *contig* 5350 apresenta dois éxons, porém quando observado a região predita pelo software Augustus há identificação de domínio entre as regiões codificadoras para o domínio PAZ e identificada na região transcrita (identificadas no transcriptoma).

Sequência 1 – *contig* 2714



Sequência 2 – *contig* 5350

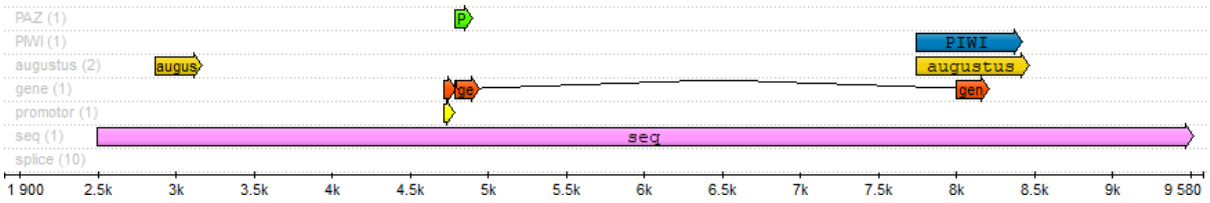


Figura 5.Sequências de *Piwis*. Em amarelo regiões codificadoras encontradas pelo software Augustus, em amarelo claro os potenciais promotores. Os domínios referentes a subfamília PIWI estão representados pelas cores: verde: PAZ, azul escuro: PIWI. Fonte:Autor

Na Figura 6 encontra-se os alinhamentos das proteínas *Piwi* de *S.leucops*. A divergência genética encontrada entre essas proteínas é de 55 %

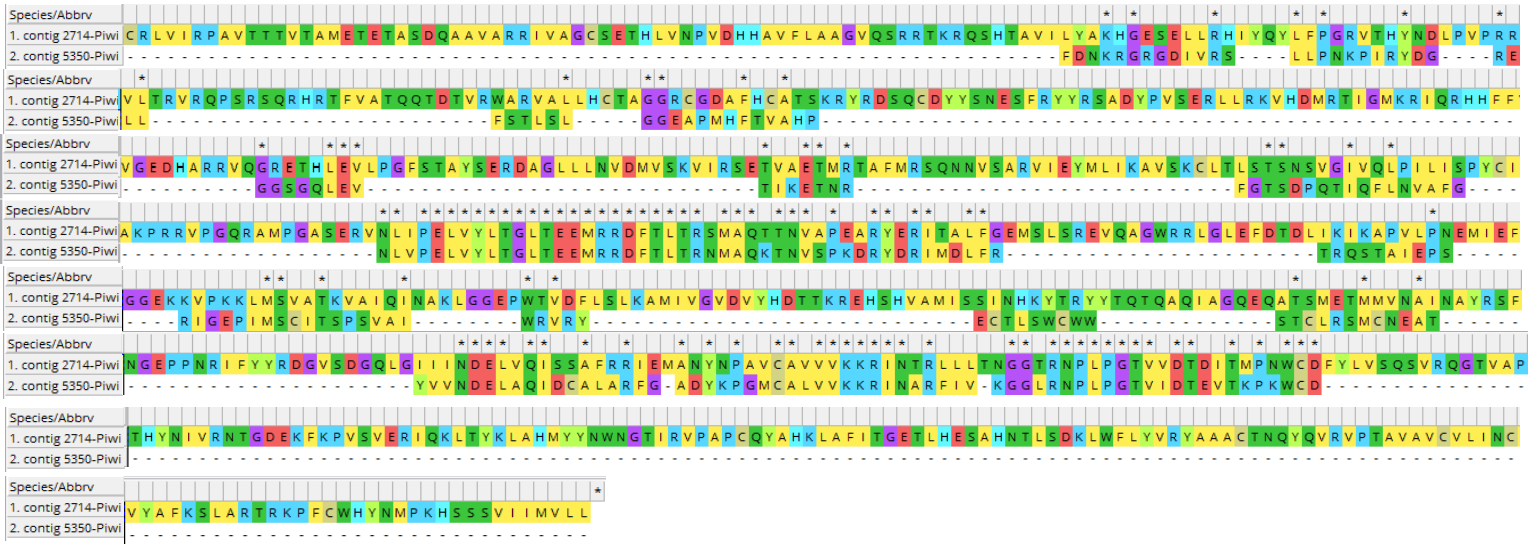


Figura 6. Alinhamentos de proteínas *Piwis* encontradas em *S. leucops*. Fonte:Autor

4.2 Análise filogenética dos genes *Argonautas* e *Piwi*

Para investigação das relações filogenéticas das proteínas *Argonauta* e *Piwi* foram realizadas comparações entre as sequências encontradas do nosso trabalho de outros platelmintos de vida livre e grupos externos sendo esses Artrópodes, Nematódeos, Cordados e Molusca.

Nossas sequências de *Argonauta* de *Stenostemum leucops* se classificaram em um clado único, com alto valor de bootstrap, dentro de um conjunto de representantes de planárias de vida livre incluindo táxons como molusca, trematódeos e cestoides. Quando considerado mais táxons, cordados, artrópodes e o nematódeo *Cenorhabditis elegans* encontram no mesmo conjunto, porém de maneira separada (Figura 7 e 8). Todos os táxons exibiram a presença de todos, ou quase todos, os domínios predominantes da família *Argonauta*.

Um clado com sequências semelhantes a *Piwi* pode ser diferenciado e um clado exclusivo composto por alguns representantes de planárias de vida livre, apresentando baixo suporte, e parasitas com maior suporte, mostraram sequências mais divergentes entre as demais espécies emenor presença de domínios.

As sequências de *Piwi* de *S. leucops* se agruparam em um clado único bem suportado, junto a organismos que compartilham *Piwi*1. Outro conjunto de organismos apresentam apenas *Piwi*2 e há um grupo formado por *Piwis* exclusivos de planárias de vida livres (Figura 9 e 10). Dentrodos diferentes conjuntos, clados únicos de espécies foram formados com bom suporte. Os grupos de *Piwi*1 e *Piwi*2 apresentam maior variedade de representantes, sugerindo sequências conservadas entre diversos organismos.

Stenostemum leucops também se apresenta em todas as árvores entre táxons das planárias de vida livres, principalmente *Macrostomum lignano*, um dos representantes Rhabditophora, grupo irmão dos Catenulidas e o nematódeo *C.elegans.*

4.3 Expressão diferencial dos genes

Com dados do transcriptoma fizemos a análise da expressão dos genes *argonautas* e *piwi* descritos nesta dissertação em organismos de *S.leucops* nas fases de 1 e 2 zoóides. A análise de expressão diferencial indicou que não há diferenças significativas na expressão desses genes entre em animais de I ou II zooides, sendo esses expressos constantemente nas duas condições.

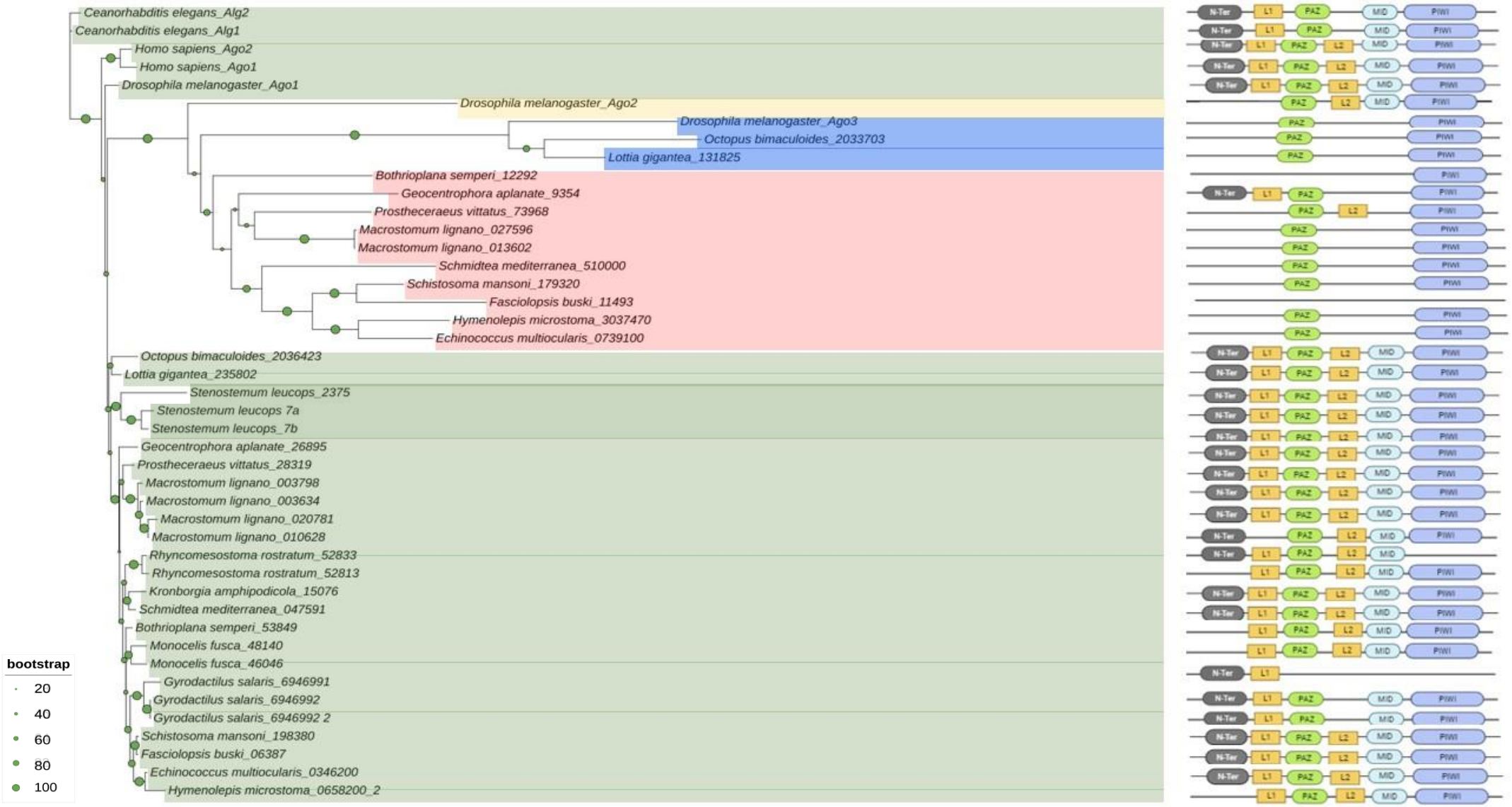


Figura 7: Árvore filogenética por Máxima Verossimilhança de proteínas *Argonauta* encontradas em diversos táxons como Turbelários, Nematoides, Trematódeos, Cestoides, Molusca, Artrópodes e cordados. Em verde *Argonautas* enontrados em apenas plánarias de vida livre, em vermelhos *Argonautas* encontrados em planarias e parasitas, amarelo siRNA e em azul *Piwi*. Ao lado se encontra os domínios específicos das proteínas em cada espécie.

Figura 8: Árvore filogenética por Máxima Verossimilhança de proteínas *Argonauta* encontradas em táxons de Turbelários, Nematoides, Trematódeos, Cestoides e Artrópodes. Em verde *Argonautas* enontrados em apenas plánarias de vida livres, em vermelhos *Argonauta*s encontrados em planarias e parasitas, amarelo siRNA e em azul Piwi. Ao lado se encontra os domínios específicos das proteínas em cada espécie

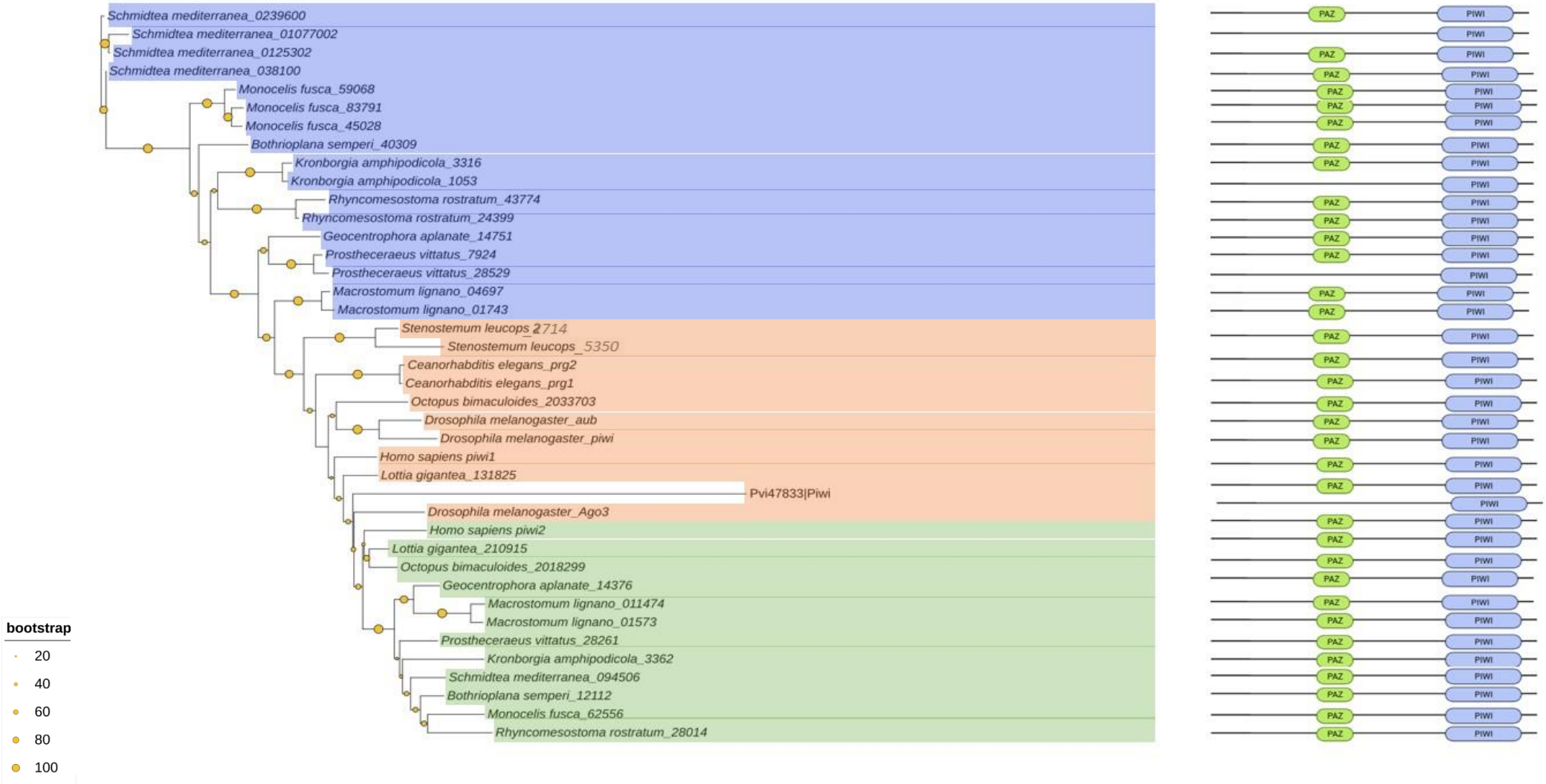


Figura 9: Árvore filogenética por Máxima Verossimilhança de proteínas *Piwi* encontradas diversos táxons como Turbelários, Nematoides, Molusca, Artrópodes e cordados. Em azul *Piwis* encontrados exclusivamente em Turbelários, vermelho Piw1 e verde *Piwi*2. Ao lado se encontra os domínios específicos das proteínas em cada espécie.

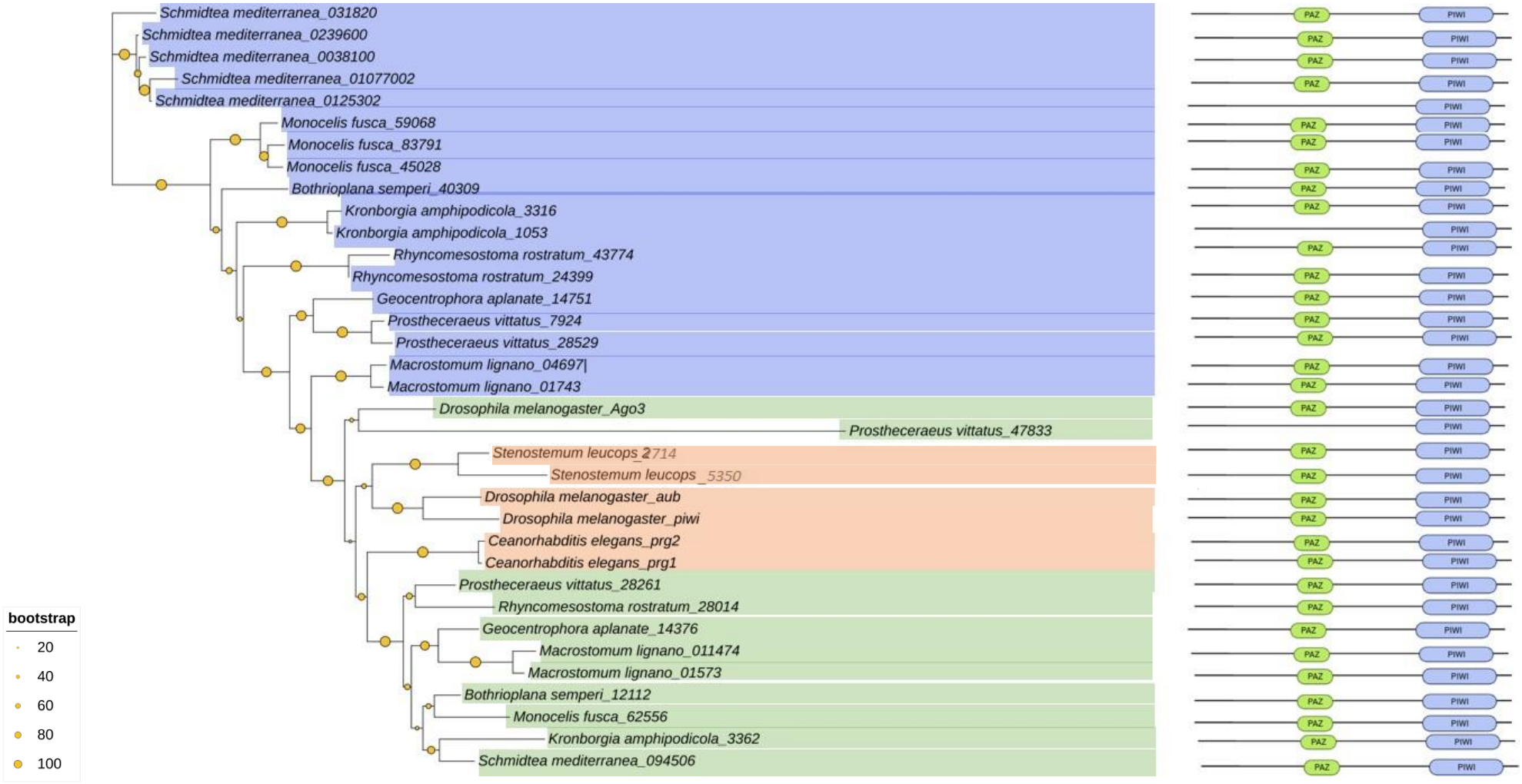


Figura 10: Árvore filogenética por Máxima Verossimilhança proteínas *Piwi* encontradas em táxons de Turbelários, Nematoides e Artrópodes. Em azul *Piwis* encontrados exclusivamente em Turbelários, em vermelho *Piwi*1 e verde *Piwi*2. Ao lado se encontra os domínios específicos das proteínas em cada espécie.

5. Discussão

Os genes *Argonautas* e *Piwi* tem funções importantes na defesa de patógenos exógenos e silenciamento de transposons, especialmente na linhagem germinativa em diversos grupos de animais. Nas planárias, também estão envolvidos na regulação da homeostase e regeneração, desempenhando um papel importante a sobrevivência desses animais (Kashima et al, 2020).

A presença dessas proteínas em *S. leucops* reforça a importância em platelmintos, especialmente de vida livre. Uma vez que em parasitas não é observado a presença da proteína *Piwi* (Protasio et al, 2020) levantando questões sobre a evolução dessas proteínas em organismos mais basais e a possível perda destes genes nos ramos derivados de platelmintos parasitos.

Estudos como de Chen et al (2010), ao empregar sequências de Argonautas em *Schistosoma japonicum,* identificaram que essas sequências exibiam entre 25% a 70% de similaridade, e possuíam os domínios proteicos PAZ e/ou PIWI da família. Zheng (2012) analisou o genoma de 6 platelmintos, abrangendo espécies de vida livre quanto parasitas, e demonstrou que *Argonauta*s dosplatelmintos foram agrupados em diferentes clados na árvore filogenética, indicando a diversidade e evolução dessas proteínas no filo. Zhou et al (2015) identificaram três proteínas *Piwi* em*Macrostemum lignano* que exibiram uma divergência de aproximadamente 40%, elas apresentaram os domínios PAZ e PIWI e as análises filogenéticas mostraram uma semelhança a *Piwis*da planária *Schimetea mediterranea*.

Com esses estudos há evidências de conservação de vias funcionais de microRNAs em platelmintos, ainda que pouco descritas para maioria das espécies (Kashima et al, 2020).

Com relação as relações filogenéticas entre os genes *Argonautas* e *Piwi*, nossos achados se assemelham a estudos de Fontenla et al. (2021), uma vez que *Stenostemum leucops* ocupa uma posição semelhante em nossas filogenias, apesar das sequencias utilizadas no nosso trabalho serem obtidas pelo nosso laboratório e apresentarem aproximadamente 74% de semelhança entre as utilizadas no estudo de Fontenla. Nossos resultados também mostram que a sequência de *Stenostemum leucops* em conjunto com a filogenia de *Argonauta*s, indica que uma das sequências presente no *contig* 7 pode representar duplicação recente, uma vez que apresenta alta similaridade e estão no mesmo *contig*. Além disso, observamos que essas sequências estão agrupadas juntas na filogenia, o que se repete em vários outros táxons de turbelários.

Nossos resultados não encontraram diferença na expressão dos genes *argonauta* e *piwi* em animais com I e II zooides em *Stenostemum*, mas estudos demonstram que há diferença na expressão especialmente em genes *piwi* de planárias, uma vez que são genes associados ao processo de regeneração. Estudos conduzidos com *S. mediterrânea*, observou-se uma redução nos níveis de piRNA quando os animais foram expostos a radiação (Li et al, 2021).

Diferentes proteínas *Piwis* estão associadas a regiões especificas do corpo como no estudo de Palakodeti et al (2008), eles observaram que certas proteínas *Piwis*, SMEDWI-1 e SMEDWI-2, são expressas tanto em neoblastos quanto nas linhagens germinativas e que SMEDWI-2 atua também na homeostase e regeneração. Tian et al (2011) também demonstraram que diferentes miRNAs são expressos diferencialmente no tecido regenerativo de planárias.

Eventualmente pesquisas especificas na expressão desses genes ou durante a regeneração deste organismo, possam oferecer uma compreensão mais aprofundada de como é regulada a expressão, e como outros genes podem desempenhar um papel na reprodução por paratomia.

Além dos sequenciamentos para ampliar o conhecimento de novas espécies, a anotação genômica é uma parte importante para identificar a localização de genes e atribuir funções a eles. (Yandell e Ence, 2012; Ekblom e Wolf, 2014). A anotação ainda é um desafio, já que ferramentas de predição são suscetíveis a erros, desta forma uma curadoria manual é importante para revisão e detalhamento da anotação, identificando e corrigindo anotações duvidosas, como presença ou ausência de domínios específicos ou genes não funcionais (Balbinot, 2020).

A combinação com outras ferramentas bioinformáticas se mostram necessárias para uma anotação mais precisa, uma vez que os preditores gênicos, como o *software* Augustus, são eficazes em identificar a região gênica de genes, mas podem deixar passar alguns detalhes quando utilizados em organismos não modelos, como no caso de *Stenostemum leucops.* A verificação de transcritos e outras ferramentas de alinhamento podem aperfeiçoar a busca e aproximar a representação da estrutura dos genes.

6. Conclusões

Identificamos e caracterizamos 3 genes *argonauta* e 2 genes *piwi* no genoma de *Stenostemum leucops*. A estrutura dos genes foi anotada e os elementos gênicos, como éxons, íntros, promotores e regiões UTRs foram identificadas. As prováveis proteínas foram deduzidas e os domínios, N, L1, PAZ, MID, L2 e PIWI foram identificados ao longo das sequências. A análise de

expressão transcipcional não mostrou expressão diferencial desses genes durante o processo de paratomia.

7. Material Suplementar

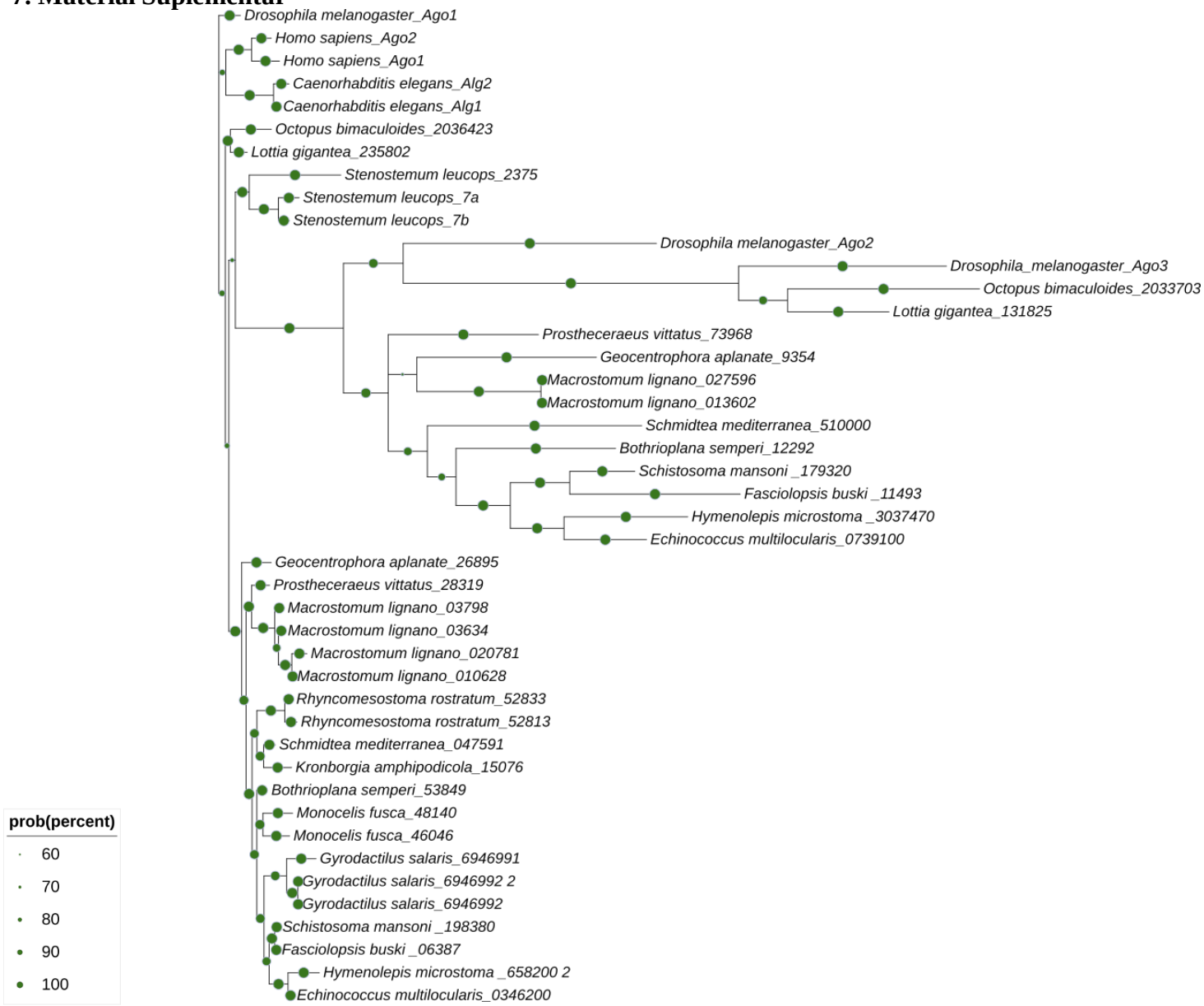


Figura 11. Árvore filogenética por Bayesiana de proteínas *Argonauta* encontradas em táxons como Turbelários, Nematoides, Trematódeos, Cestoides, Molusca, Artrópodes e cordados.

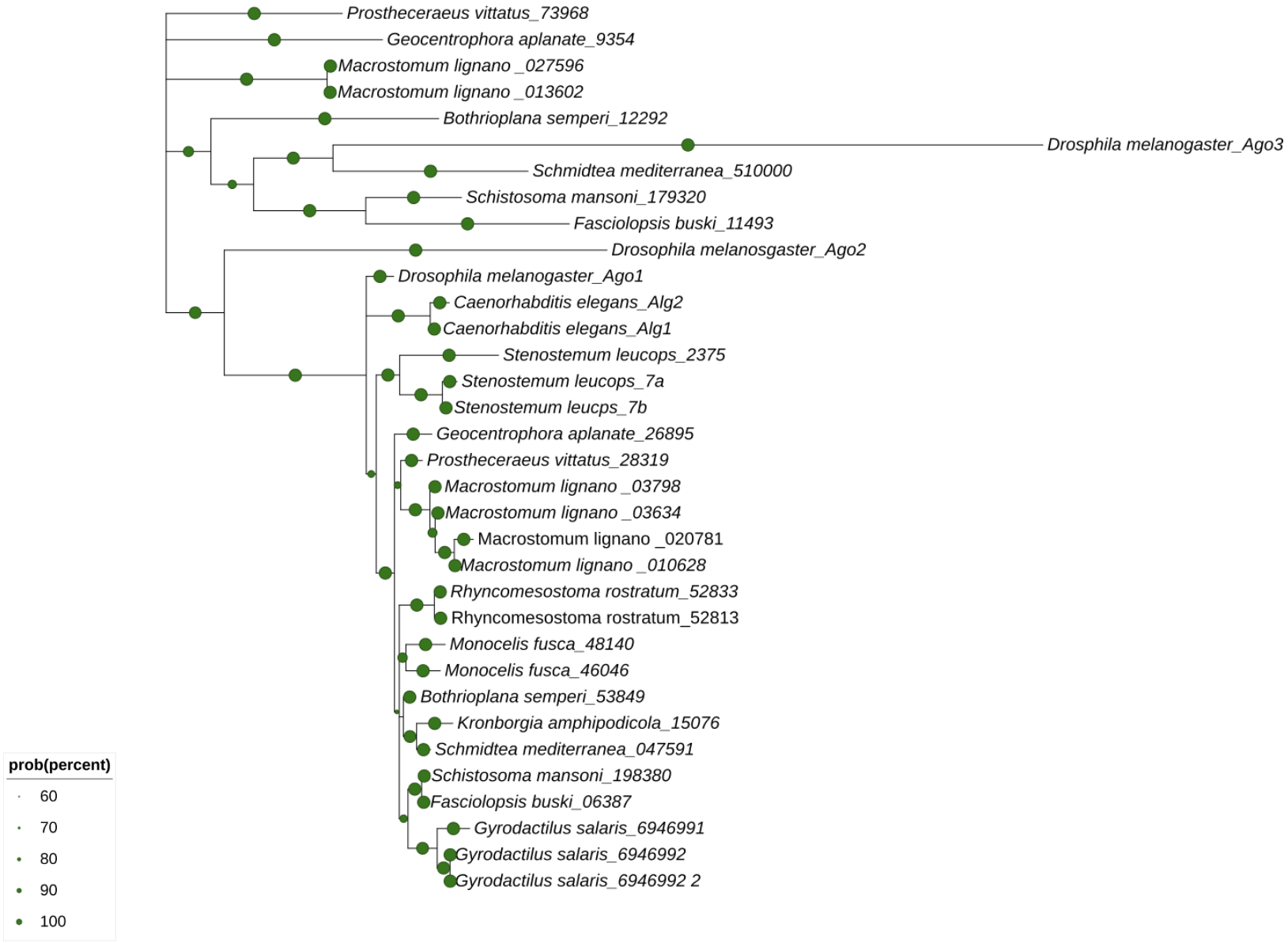


Figura 12. Árvore filogenética por Bayesiana de proteínas *Argonauta* encontradas em táxons de Turbelários, Nematoides, Trematódeos, Cestoides e Artrópode.

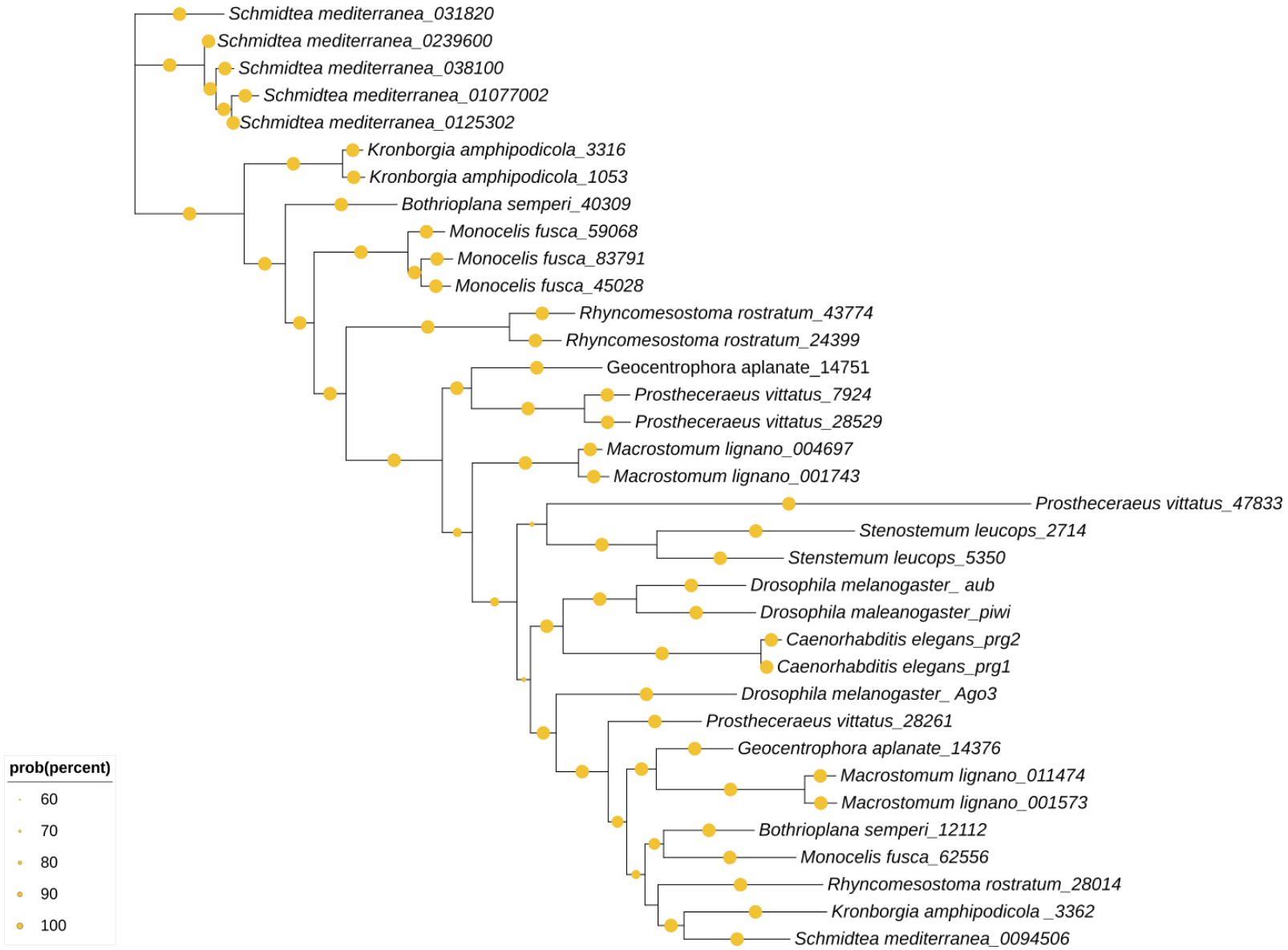


Figura 13. Árvore filogenética por Bayesiana de proteínas *Piwi* encontradas em táxons de Turbelários, Nematoides e Artrópodes.

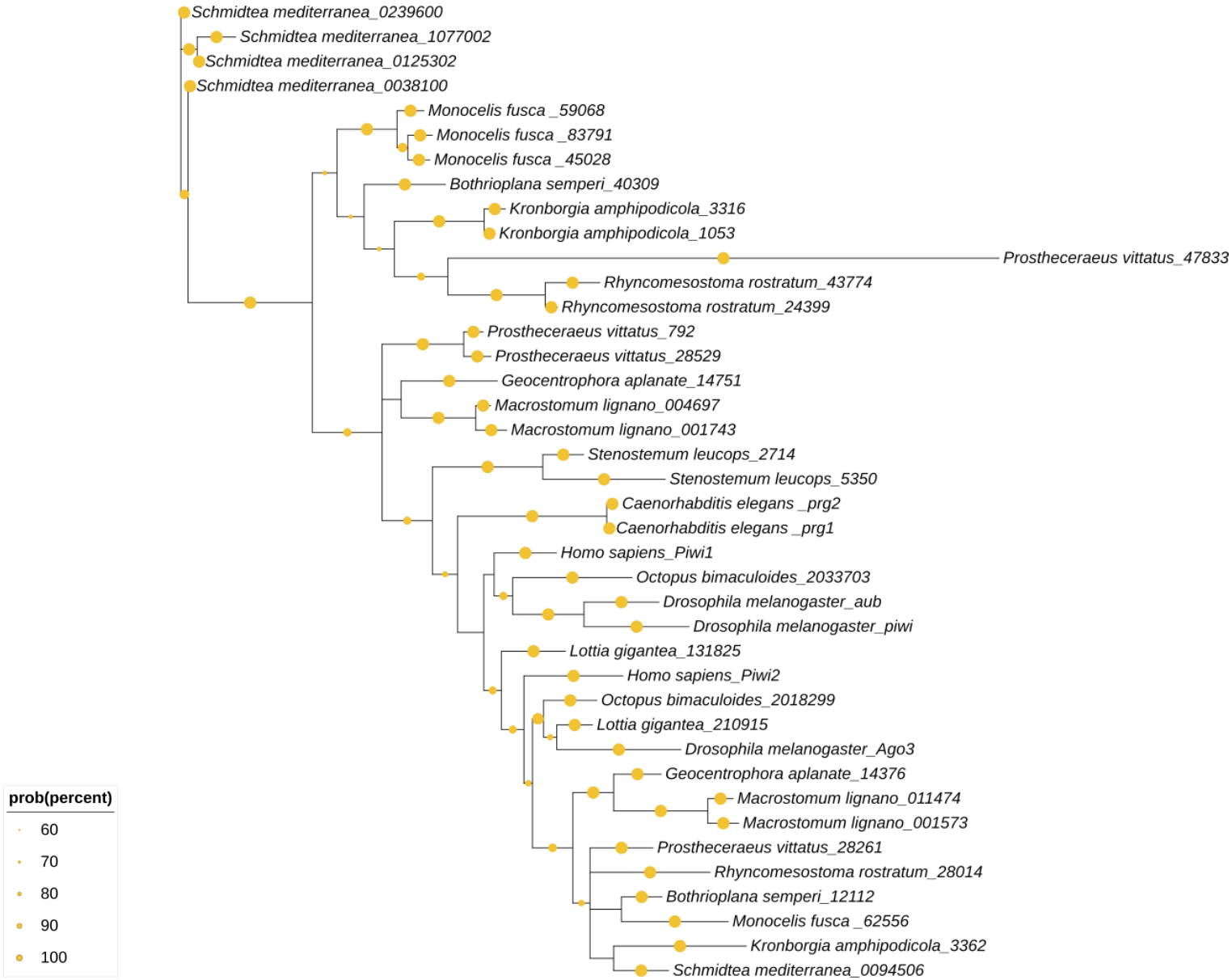


Figura 14. Árvore filogenética por Máxima Verossimilhança de proteínas *Piwi* encontradas diversos táxons como Turbelários, Nematoides, Molusca, Artrópodes e cordados.

8. Referências

Adell, T. e Riutort, M. Phyllum Platyheminthes. In Schierwater B. e DeSalle R. **Invertebrate Zoology: a tree of life approach**. New York: CRC Press, 2021 p. 219- 230.

Almeida, M. I., Reis, R. M., Calin, G. A. MicroRNA history: Discovery, recente applications, and next frontiers. **Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis**. v. 717, n. 1-2, p. 1-8. Dezembro, 2011. Disponível em: https://doi.org/10.1016/j.mrfmmm.2011.03.009.

Alvarado, S. A e Riddien, W. P. Fundamentals of Planarian Regeneration. **Annual Review of Cell and Developmental Biology**. v. 20, n. 1, p. 725-757. 2004. Disponível em: 10.1146/annurev.cellbio.20.010403.095114.

Archie, A. E. e Theis, R. K. Animal behavior meets microbial ecology. **Animal Behaviour**. v. 82, n. 3, p. 425-436. Setembro 2011. Disponível em: https://doi.org/10.1016/j.anbehav.2011.05.029.

Baguña, J. e Romero, R. Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. **Hidrobiologia**. v. 84, p.181-194. Outubro, 1981. Disponível em: https://doi.org/10.1007/BF00026179.

Bailey, T. L, Johnson J., Grant, C. E. et al. The MEME Suite. **Nucleic Acid Research**. v. 43, n. W1, p. W39- W49. 2015. Disponível em: https://doi.org/10.1093/nar/gkv416.

Balbinot, E. Workflow Cietífico e Anotação Genômica Funcional e Curadoria Manual de Genomas. (Dissertação) Mestrado em Biotecnologia. Universidade de Caxias do Sul. Caxias do Sul. p. 83. 2020.

Bordentein, R. S. e Theis, R. K. Host Biology in Light on the microbiome: Ten principles of Holobionts and Hologenomes. **Plos Biology**. v. 13, n. 8. Agosto, 2015. Disponível em: https://doi.org/10.1371/journal.pbio.1002226.

Bray, N., Pimentel, H., Melsted, P. et al. Near-optimal probabilistic RNA-seq quantification. **Nature Biotechnology**. v.34, n. 5, p. 525-527. 2016. Disponivel em: https://doi.org/10.1038/nbt.3519.

Brusca, C. R., Moore, W e Shuster, M. S. Filo Platyhelmithes. In:_____. (org). **Invertebrados**. Rio de Janeiro: Guanabara Koogan, 2018 p. 457- 503.

Cantalapiedra, C. P.; Hernández-Plaza, A.; Letunic, I. et al. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. **Molecular Biology and Evolution**. v. 38, n. 12, p. 5825-5829. 2021. Disponível em: https://doi.org/10.1093/molbev/msab293.

Cao, Z. et al. Different classes of small RNAs are essential for head regeneration in the planarian Dugesia japonica. **BMC Genomics**. v. 21. n. 1. 2020. Disponível em: https://doi.org/10.1186/s12864-020-07234-1.

Centro Nacional de Informações sobre Biotecnologia (NCBI) [Internet]. Bethesda (MD): Biblioteca Nacional de Medicina (EUA), Centro Nacional de Informações sobre Biotecnologia; [1988] – [consultado em 06 de abril de 2017]. Disponível em: https://www.ncbi.nlm.nih.gov/.

Chalfie, M., Horvitz, R. H., Sulston J. E. Mutation that to reiterations in the cell lineages of C. elegans. **Cell Press**. v. 24, n.1, p. 59-69. 1981. Disponível em: https://doi.org/10/1016/0092-8674(81)90501-8.

Chen, J., Yang, Y., Guo, S. et al. Molecular cloning and expression profiles of Argonaute proteins in *Schistosoma japonicum*. **Parasitol Research**. v.107, p.889–899 2010. Disponível em: https://doi.org/10.1007/s00436-010-1946-3.

Consortium UniProt: UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**. v. 49. n. D1, p. D480-D489. Janeiro, 2021. Disponivel em: https://doi.org/10.1093/nar/gkaa1100.

Darriba, D.; Posada, D.; Kozlov, A. M. et al. ModelTest-NG: A new and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. **Molecular Biology and Evolution**. v. 37, n. 1, p. 291-294. 2020. Disponivel em: https://doi.org/10.1093/molbev/msz189.

Egger, B. et al. Free-living flatworms under the knife: past and present. **Development genes and evolution.** v. 217, n**.** 2, p. 89-104. Fevereiro, 2007. Disponível em: https://doi.org/10.1007/s00427-006-0120-5.

Eisenhoffer, T. G., Kang, H., Alvarado, S. A. Molecular Analysis of Stem Cells and Their Descendants during Cell Turnover and Regeneration in the Planarian *Schmidtea mediterranea*. **Cell Stem Cell**. v. 3, n. 3, p. 327-339. Setembro, 2008. Disponível em: https://doi.org/10.1016/j.stem.2008.07.002.

Ekblom, R., Wolf, J. B. W. A field guide to whole-genome sequencing, assembly and annotation. **Evolutionary Applications**. v.7, n. 9, p. 1026-1042. 2014. Disponível em: https://doi.org/10.1111/eva.12178.

Filho, J.C. M. R. e Kimura, E. T. MicroRNAs: Nova Classe de Reguladores Gênicos Envolvidos na Função Endócrina e Câncer. **Arquivos Brasileiros de Endocrinologia e Metabologia**. v. 50, n. 6, p. 1102-1107. São Paulo,2006. Disponível em: https://www.scielo.br/j/abem/a/HMS663yYSyVvNmntfjMDQYQ/?format=pdf&lang=pt.

Fontenla, S., Rinaldi, G., Smircich, P. et al. Conservation and diversification of small RNA pathways within flatworms. **BMC Evolutionary Biology**. v. 17. 2017. Disponível em: https://doi.org/10.1186/s12862-017-1061-5.

Fontenla, S., Rinaldi, G., Tort, J. F. Lost and Found: Piwi and Argonaute Pathways in Flatworms. **Frontiers in Cellular an Infection Microbiology**. v. 11. Maio, 2017. Disponível em: https://doi.org/10.3389/fcimb.2021.653695'.

Girard, A. Sachidanandam, R., Hannon, G. J. e Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi porteins. **Nature**. v. 442, n. 7099, p. 199-202. Junho, 2006. Disponível em: https://doi.org/10.1038/nature04917.

Grabher, M. G., Brian, H. J., Yassour, M. et al. Trinity: reconstructing a full-length transcriptoma whitout a genome from RNA-Seq data. **Nature Biotechnology**. v. 29, n. 7, p. 644-652. Julho, 2011. Disponivel em: 10.1038/nbt.1883.

Gramates, L. S., Agapite, J., Attrill, H., et al. FlyBase: a guided tour of highlighted features. **Genetics**. v. 220, n. 4. Abril, 2022. Disponivel em: https://doi.org/10.1093/genetics/iyac035.

Griffiths, A. J. F. et al. RNA. In:_____.(org). **Introdução a Genética**. Rio de Janeiro: Guanabara Koogan, 2013 p. 245- 268.

Guedelhoefer, IV C. O., e Alvarado S. A. Amputation induces stem cell mobilization to sites of injury during planarian regeneration. **Development**. v. 139, n. 19, p. 3510-3520. Outubro, 2012. Disponível em: https://doi.org/10.1242/dev.082099.

Howe, K. L., Bolt, B. J., Shafie, M., et al. WormBase Parasite- a comprehensive resource for helminth genomics. **Molecular and Biochemical Parasitology**. v. 2, n. 10. Julho, 2017. Disponivel em: 10.1016/j.molbiopara.2016.11.005.

Huelsenbeck, J. P. e Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. **Bioinformatics**. v. 17, n. 8, p. 754-755. 2011. Disponível em: https://doi.org/10.1093/bioinformatics/17.8.754.

Letunic, I e Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. **Nucleic Acids Research**. v. 49, p. W293-W296. 2021. doi: 10.1093/nar/gkab301

Jin, S. Zhan, J. e Zhou. Y. Argonaute proteins: structures and thei endonuclease activity. **Molecular Biology Reports**. v. 48, p. 4837-4849. 2021. Disponível em: https://doi.org/10.1007/s11033-021-06476-w.

Kashima, M., Agata, K. e Shibata, N. What is the role of PIWI Family proteins in adult pluripotent stem cells? Insights from asexually reproducing animals, planarians. **Developmental, growth & differentiation**. v. 62, n. 6, p. 407-422. Agosto, 2020. Disponível em: 10.1111/dgd.12688.

Kim, N. V. MicroRNA biogenesis: coordinated cropping and dicing. **Nature Reviews Molecular Cell Biology**. v. 6, p. 376-385. Maio, 2005. Disponível em: https://doi.org/10.1038/nrm1644.

Larsson, K., Jondelius, U. Phylogeny of Catenulida and support for Platyhelmintes. **Organisms Diversity & Evolution**. v. 8, n. 5, p. 378-387. Dezembro, 2008. Disponível em: https://doi.org/10.1016/j.ode.2008.09.002.

Larsson, K., Willems, W. Report on freshwater Catenulida (Platyhelminthes) from Sweden with the description of four new species. **Zootaxa**, v. 2396, n. 1, p. 1-18. Março, 2010. Disponível em:10.5281/zenodo.193965.

Lee, R. C., Felnbaum, R. L., Ambros, V. The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*.**Cell Press**. v. 75, n. 5, p. 843-854. Dezembro, 1993. Disponível em: https://doi.org/10.1016/0092-8674(93)90529-Y.

Li, D.; Taylor, D. H., van Wolfswinkel, J. C. PIWI-mediated control of tissue-specific transposons is essential for somatic cell differentiation. **Cell Reports**. v.37, n.5. 2021. Disponível em: https://doi.org/10.1016/j.celrep.2021.109776.

Lin, Y. et al. Assembly of long error-prone reads using de Bruijn graphs. **Proceedings of the National Academy of Sciences**. v. 113, n. 52. Outubro, 2016. Disponível em: https://doi.org/10.1073/pnas.1604560113.

Loubalova, Z. Konstantidinidou, P. e Haase, A. D. Themes and variations on piRNA-guided transposon control. **Mobile DNA**. v. 14, n. 10.2023. Disponível em: https://doi.org/10.1186/s13100-023-00298-2.

Minh, B. Q., Schimdt, H. A., Chernomor, O. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. **Molecular Biology and Evolution**. v. 37, n. 5, p. 1530-1534. 2020. Disponível em: https://doi.org/10.1093/molbev/msaa015.

Morris, J. J. What is the hologenome concept of evolution?. **F1000 Research**. v. 7, n. 1, Outubro, 2022. Disponível em: 10.12688/f1000research.14385.1

Noreña, C., Damborenea, C., Brusa. A taxonomic revision of South American species of the genus Stenostomum O. Schmidt (Platyhelminthes: Catenulida) based on morphological characters. **Zoological Journal of the Linnean Society**. v.144, n. 1, p. 37–58. 2005. Disponível em: https://doi.org/10.1111/j.1096-642.2005.00157.x.

Okonechnikov, K.; Golosova, O.; Fursov, M. et al. Unipro UGENE: a unified bioinformatics toolkit. **Bionformatics**. v. 28, n. 8, p. 1166-1167. 2012. Disponível em: https://doi.org/10.1093/bioinformatics/bts091.

Palakodeti, D.; Smielewska, M.; Lu, Y. C. et al. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. **RNA** v. 14, n. 6, p. 1174-1186. 2008. Disponível em: https://doi.org/10.1261/rna.1085008.

Pawlicka, K. Perrigue, P. M., Barciszewski, J. Regulatory RNAs in Planarians. **Acta Biochimica Polonica**. v. 63, n. 4, p. 681-686. 2016. Disponível em*:* https://doi.org/10.18388/abp.2016_1355.

Promi, D., Babaei, P., Nielsen, J. Metagenomic mnalysis of microbe-mediated vitamin metabolism in the human gut microbiome. **BMC Genomics**. v. 20, n. 12. Março, 2019. Disponível em: https://doi.org/10.1186/s12864-019-5591-7.

Protasio, A. V.; Rawlinson, K. A.; Miska, E. A. et al. Evidence for transposable element control by Argonautes in a parasitic flatworm lacking the piRNA pathway. **bioRxiv**. 2020. Disponível em: doi: https://doi.org/10.1101/670372.

Reddien, P. W.; Bermange, A. L.; Murfitt, K. J.; et al. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. **Developmental cell**. v. 8, n. 5, p. 635-649. 2005. Disponível em: https://doi.org/10.1016/j.devcel.2005.02.014.

Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**. v. 26, n. 1, p. 139-140. Disponível em: https://doi.org/10.1093/bioinformatics/btp616.

Rosa, M. T. Descrição do Genoma, Transcriptoma e Microbioma de *Stenostemum lucops* (Platyhelminthes, Catenulida), com enfoque na filogenia, regeneração, plasticidade ecológica e elementos transponíveis. (Tese) Doutorado em Biodiversidade Animal. Universidade Federal de Santa Maria. Santa Maria. p. 107. 2022.

Rosa, M. T. et al. *Stenostomum leucops* Dugès, 1828 (Platyhelminthes, Catenulida): a putative species complex with phenotypic plasticity. **Papéis Avulsos de Zoologia**. v. 55, n. 27, p. 375-383. Dezembro 2015. Disponível em: https://doi.org/10.1590/0031-1049.2015.55.27.

Rosa, M.T., Loreto, E.L.S. *Stenostomum leucops* (Catenulida, Platyhelminthes) has a flexible microbiome in time and space. **Hydrobiologia**. v.850, p. 3675–3683. 2023. Disponivel em: *https*://doi.org/10.1007/s10750-022-04931-8.

Rosa, M.T., Loreto, E.L.S. Revisintind the regeneration of *Stenostemum leucops* (Catenulida, Platyhelminthes). **Invertebrate Reproduction & Development**. v. 66, n.1, p. 1-7. 2022. Disponível em: https://doi.org/10.1080/07924259.2021.1994475. Acesso em: Novembro de 2023.

Rosa, M.T., Loreto, E.L.S. The Catenulida flatworm can express genes from its microbiome or from the DNA it ingests. **Scientifics Reports**. v. 9, n. 19045. Dezembro, 2019. Disponível em: https://doi.org/10.1038/s41598-019-55659-w.

Shennan, L.; Wang, J.; Chitsaz, F., et al. CDD/SPARCLE: the conserved domain database in 2020. **Nucleic Acids Research.** v. 48, n. D1, p. D265-D268. 2020. Disponivel em: 10.1093/nar/gkz991.

Slack, M. W. J. Evolution of Regeneration. In: Harvey, P. Richard e Rosenthal, Nadia.(org). **Heart Development and Regeneration**. Academic Press. 2010, p. 827-837.

Solana, J. et al. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. **Genome Biology**. v. 13, R19. 2021. Disponível em: https://doi.org/10.1186/gb-2012-13-3-r19.

Stanke, M. Steinkamp, R., Waack, S. et al. AUGUSTUS: a web server for gene finding in eukaryotes. **Nucleic Acids Reserch**. v. 32, n. suppl_2, p. W309-W312. Julho, 2004. Disponivel em: https://doi.org/10.1093/nar/gkh379.

Stanke, M., e Waack, S. Gene prediction with a hidden Markov model and a new intron submodel.**Bioinformatics**. v. 19, n. suppl_2, p. ii215-ii225. Setembro, 2003. Disponivel em: https://doi.org/10.1093/bioinformatics/btg1080.

Swarts, D. C., Marakova, K., Wang, Y. et al. The evolutionary journey of Argonaute proteins. **Nature Structural & Molecular Biology**. v. 21, p. 743-753. 2014. Disponível **e**m: https://doi.org/10.1038/nsmb.2879.

Talavera, G., Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. **Systematic biology**. v. 56, n. 4, p. 564-577. 2007. Disponivel em: https://doi.org/10.1080/10635150701472164.

Taura, K., Stecher, G., Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. **Molecular Biology and Evolution**. v. 38, n. 7, p. 3022-3027. 2021. Disponivel em: https://doi.org/10.1093/molbev/msab120.

Tian, Q-N. et al. Differential expression of microRNA patterns in planarian normal and regenerative tissues. **Molecular Biology Reports**. v. 39, p. 2653-2658. 2012. Disponivel em: https://doi.org/10.1007/s11033-011-1018-5.

Tian, QN., Bao, ZX.; Lu, P. et al. Differential expression of microRNA patterns in planarian normal and regenerative tissues. **Mol Biol Rep 39**, 2653–2658. 2012. https://doi.org/10.1007/s11033-011-1018-5.

Ugarte, P. B., Barendse, P. Swarts, E. C. Argonaute proteins confer immunity in all domains of life. **Current Opinion in Microbiology**. v.74. 2023. Disponível em: https://doi.org/10.1016/j.mib.2023.102313.

Ustyantsev, K. V. et al. *Macrostomum lignano* as a model to study the genetics and genomics of parasitic flatworms. **Vavilovskii zhurnal genetiki i selektsii**. v. 25, n. 1, p. 108-116. Fevereiro, 2021. Disponível em: 10.18699/VJ21.013.

Wagner, E. D., Wang E. I., Reddien, W. P. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. **Science**. v. 332, n. 6031, p. 811-6. Maio, 2011.Disponivel em: 10.1126/science.1203983.

Wolfswinkel van, J. C., Wagner, D. E. e Reddien, P.W. Single-Cell Analysis Reveals Functionally Distinct Classes within the Planarian Stem Cell Compartment. **Cell Stem Cell**. v. 15, n. 3, p. 326-339. Setembro, 2014. Disponível em: https://doi.org/10.1016/j.stem.2014.06.007.

Wu, J. et al. Argonaute proteins: Structural features, functions and emerging roles. **Journal of Advanced Research**. v. 24, p. 317-324. 2020. Disponível em: https://doi.org/10.1016/j.jare.2020.04.017.

Yandell, M., Ence, D. A beginner's guide to eukaryotic genome annotation. **Nature Reviews Genetics**. v.13, p. 329-342. 2012. Disponivel em: https://doi.org/10.1038/nrg3174.

Zheng, Y. Phylogenetic analysis of the Argonaute protein family in platyhelminths. **Molecular Phylogenetics and Evolution**. v.66, n.3, p.1050 1054. 2010. Disponivel em: https://doi.org/10.1016/j.ympev.2012.11.014.

Zhou, X., Battistoni, G., El Demerdash, O. et al. Dual functions of Macpiwi1 in transposon silencing and stem cell maintenance in the flatworm *Macrostomum lignano*. **RNA.** v. 21, n. 11, p. 1885-1897. 2015. Disponível em: https://doi.org/10.1261/rna.052456.115.