

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



Caracterização e evolução de parálogos de CenH3 em espécies de *Drosophila* do grupo *repleta*

Orientado: Erick Weberth de Lima Junqueira

Orientador: Dr. Gustavo Campos e Silva Kuhn

Belo Horizonte

2019

Erick Weberth de Lima Junqueira

Caracterização e evolução de parálogos de CenH3 em espécies de *Drosophila* do grupo *repleta*

Dissertação apresentada ao programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais como pré-requisito obrigatório para obtenção do título de Mestre em Genética, área de concentração Genômica e Bioinformática.

Orientador: Dr. Gustavo Campos e Silva Kuhn

Belo Horizonte

2019

043 Junqueira, Erick Weberth de Lima .
 Caracterização e evolução de parálogos de CenH3 em espécies de drosophila
do grupo repleta [manuscrito] / Erick Weberth de Lima Junqueira . - 2019.
 50 f. : il. ; 29,5 cm.

 Orientador: Dr. Gustavo Campos e Silva Kuhn.

 Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de
Ciências Biológicas. Programa de Pós-Graduação em Genética.

 1. Genética. 2. Duplicação gênica. 3. Centrômero. 4. Drosophila. I. Kuhn,
Gustavo Campos e Silva. II. Universidade Federal de Minas Gerais. Instituto
de Ciências Biológicas. III. Título

CDU: 575



Pós-Graduação em Genética
Departamento de Genética, Ecologia e Evolução, ICB
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - C.P. 486 - Pampulha - 31270-901 - Belo Horizonte - MG
e-mail: pg-gen@icb.ufmg.br FAX: (+31) - 3409-2570



**"Caracterização e evolução de parálogos de CenH3 em espécies de
Drosophila do grupo repleta "**

Erick Weberth de Lima Junqueira

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Marta Svartman
Marta Svartman
UFMG

Fernanda Antunes Carvalho
Fernanda Antunes Carvalho
UFMG

Laila Alves Nahum
Laila Alves Nahum
FIOCRUZ

Belo Horizonte, 28 de novembro de 2019.

Dedico este trabalho aos meus sobrinhos, Sabryna, Alerrandro, Maria Eduarda, Valery, Cecília e José Carlos, para que saibam que podemos ser pioneiros em difíceis jornadas, e ao meu anjinho Gabriel, que jamais será esquecido.

Agradecimentos

É possível ser sucinto quando se tem tantas pessoas para agradecer e sem as quais nada disso seria possível?

Em primeiro lugar, gostaria de agradecer aquele que encontro todos os dias diante do espelho. Em você reconheço a existência do divino e a persistência do profano. Continuaremos juntos na batalha pelo equilíbrio da vida, da prevalência do divino, nas aventuras do caminho escolhido e na expectativa do amanhã. Logo começaremos um novo capítulo.

Minha gratidão aos **professores do LCEv**. Ao meu orientador **Dr. Gustavo Campos e Silva Kuhn**, por me receber como aluno, me ensinar uma área da genética que antes era totalmente obscura para mim e me ajudar na jornada de ser um cientista, fica aqui registrado o meu mais sincero obrigado. E a professora, **Dr^a. Marta Svartman**, que em um ano de convivência me ensinou dez. Obrigado por toda experiência compartilhada. Além de participarem do meu crescimento profissional tenho a certeza que ajudaram a lapidar um ser humano melhor.

Aos membros das **bancas examinadoras**, pela disponibilidade em analisar o trabalho apresentado em Seminários B e esta dissertação, obrigado por refrescarem meus olhos quando a vista cansada já não enxergava direito mais.

A **Universidade Federal de Minas Gerais**, ao **Departamento de Genética, Ecologia e Evolução**, ao **Programa de Pós-graduação em Genética** da UFMG e todos os **professores e funcionários**, em especial a **Raissa, Daniela e Vitoria**, que com muita paciência, alegria e eficiência tornaram os processos burocráticos toleráveis.

Aos funcionários da UFMG e do ICB **Daniel dos Santos Filho, Marlene, Maria da Paixão, Dona Maria, Jeferson, Elaine e todos os outros** que trabalham nos bastidores da ciência brasileira, o trabalho de formiguinha de vocês é que mantém esse formigueiro funcionando. Gratidão!

À equipe dos laboratórios LBEM (**Fabricio dos Santos**), LBMM (**Evanguedes Kalapothakis**), LGM (**Monica Bucciarelli**), GenePop (**Bernadete Lovato**), LINC (**Adriana Abalen**), LGH (**Raquel Carvalho**), TecnoGen (**Anderson Miyoshi**), LAB (**Francisco Lobo**) e Farmacogenômica e Farmacogenética (**Marcelo Luizon**) que sempre estiveram de portas abertas para que eu pudesse realizar meus experimentos e tirar dúvidas.

Aos amigos do LCEv, **Barbie, Babau, Pepita, Rafa, Mi, Rada, Licinha, Naiara, Gustavinho, Lukete, Lu, Zé e Gui**. O que posso dizer é que sem vocês não existiria o James, sem o James não existiria esse mestrado. **Rafa**, nosso casamento na ciência foi o melhor casamento que esse ICB já viu. Obrigado por estar sempre ao meu lado.

As amigas **Nikole, Hortênsia, Aline e Laurinha**, vocês não têm noção do quanto eu as admiro. Obrigado por serem o meu muro das lamentações, minhas incentivadoras e por me ensinarem tanto. Aos **amigos da PPGG** que direta ou indiretamente contribuíram para a minha caminhada, em especial a **Tatiane, Flavia, Marina, Thomaz, Thais, Amanda e Nazaré**.

A **minha família**, sendo os mais importantes a mamãe (**Cirene**), o papai (**José**), os manos (**Roberth e Jean**), os avôs (**Jerônimo e Antônio**) e as avós (**Orozina e Carmelita**). A fé que vocês depositam em mim só não é maior do que o amor que sinto por vocês. É por saber que sempre posso contar com vocês que tenho tanta coragem em desbravar o desconhecido. **Tias, tios, primas e primos**, meu agradecimento se estende a todos.

Aos **novos amigos** que fiz em BH e tornaram esse período mais agradável, em especial ao **César, Ricardo, Marliete, Ana, Flaviane, Viviane, Fernanda, Thales, Alessandra, Rafa e Felipe**. Aos **velhos amigos** (Caldas Novas, Goiânia, Brasília, Diretoria e CsF) que durante minha ausência continuaram a se preocupar comigo e torceram pelo meu sucesso. Um agradecimento especial para **Raquel, Gui e Alba**, que me ajudam a sonhar e ao mesmo tempo manter os pés no chão.

To my Aussie friends, **Max, James and Joel**, who have followed my journey as a scientist and have cheered for my success.

As famílias belorizontinas que me receberam de braços abertos, obrigado **Tia Selma, Tia Adélia e Regina**.

A todos aqueles que encontraram as minhas chaves nas seis vezes que as perdi. Não sei quem são vocês, mas obrigado. Até minha mãe já desistiu das minhas chaves.

Ao meus **housemates** nesses dois anos de BH, em especial ao **Iuri**, que na fase mais complicada do mestrado me ofereceu um lar calmo e tranquilo. Desejo a você o mesmo quando chegar a sua hora de defender o mestrado.

Pelo apoio financeiro, gostaria de agradecer ao **CNPq** e a **CAPES** pela bolsa concedida e apoio a pesquisa. E claro, aos **meus pais**, que possibilitaram a minha permanência em BH já que a bolsa de mestrado é, claramente, insuficiente para que se tenha um mestrado tranquilo do ponto de vista financeiro.

Resumo

Centrômeros são regiões importantes para a segregação dos cromossomos, sendo os DNAs satélites e elementos transponíveis os seus componentes mais abundantes. Durante a divisão celular, a proteína centromérica CenH3, que é uma variante da histona H3, interage com o DNA centromérico formando nucleossomos exclusivos dos centrômeros. Tanto o DNA centromérico quanto as proteínas centroméricas possuem uma rápida taxa de evolução, o que é um fato inesperado, considerando a importante função do centrômero. De acordo com a hipótese do impulso centromérico, a expansão do DNA centromérico em um determinado cromossomo implica no maior recrutamento de proteínas centroméricas que, por fim, favorece a transmissão preferencial do cromossomo para o óvulo e conseqüentemente para a próxima geração. Entre as desvantagens desse fenômeno estão o aumento da infertilidade em machos e o possível aumento da frequência de cromossomos que podem carregar mutações deletérias. Em *Drosophila*, a proteína centromérica CenH3 é conhecida como Cid. Estudos anteriores demonstraram a existência de parálogos desse gene em espécies do subgênero *Drosophila*. O presente projeto teve como objetivo ampliar os estudos dos parálogos de *Cid* em espécies do grupo *repleta*, determinar a origem filogenética da duplicação gênica *Cid1-Cid6*, caracterizar todos os parálogos encontrados e testar a possível existência de seleção positiva sobre eles. Para isso, técnicas de biologia molecular associadas com análises de bioinformática foram utilizadas. Os resultados apontam que o parálogo *Cid6* já estava presente no ancestral comum do complexo *buzzatii*. Somente os parálogos *Cid1* e *Cid5* foram encontrados no *cluster mojavensis* e *Cid6* e *Cid5* no complexo *buzzatii*. Além disso, nossa análise indica que *Cid5*, parálogo encontrado no macho de *Drosophila*, está sob seleção positiva, o que sugere que este parálogo poderia atuar como supressor do impulso centromérico em espécies de *Drosophila*.

Palavras-chave: cenH3; duplicação gênica; centrômero; *drosophila*

Abstract

Centromeres are important regions for chromosome segregation, with satellite DNAs and transposable elements being their most abundant components. During the cell division, the centromeric protein CenH3, which is a variant of histone H3, interacts with the centromeric DNA forming a specific nucleosome in the centromere. Both, centromeric DNA and centromeric proteins, have a rapid evolution rate, which is an unexpected fact considering the important role of the centromere. According to the centromere drive hypothesis, the expansion of centromeric DNA in a specific chromosome implies in increase recruitment of centromeric proteins, which ultimately favours preferential transmission of the chromosome to the egg and hence to the next generation. Disadvantages of this phenomenon include increased infertility in male and the possible increase in the frequency of chromosomes that may carry deleterious mutations. In *Drosophila*, the centromeric protein CenH3 is known as Cid, which previous studies have shown the existence of paralogs of this gene in species of the *Drosophila* subgenus. The present project aimed to broaden the studies of *Cid* paralogs in species from the *repleta* group, to determine the phylogenetic origin of *Cid1-Cid6* gene duplication, to characterize all found paralogs and to test the possible existence of positive selection on them. For this, molecular biology techniques associated with bioinformatics analysis were used. The results indicate that the *Cid6* paralog was already present in the common ancestor of the *buzzatii* complex. Only *Cid1* and *Cid5* paralogs were found in *mojavensis* cluster, whereas *Cid6* and *Cid5* were found in the *buzzatii* complex. Furthermore, our analysis indicates that *Cid5*, a paralog found in the *Drosophila* male, is under positive selection, suggesting that this paralog could act as a centromere drive suppressor in *Drosophila* species.

Keywords: cenH3; gene duplication; centromere; *drosophila*

Sumário

1. Introdução.....	10
1.1 Centrômero e DNA centromérico.....	10
1.2 Proteína centromérica CenH3.....	12
1.3 A hipótese do impulso centromérico.....	14
1.4 <i>Drosophila</i> como organismo modelo.....	16
1.4.1 O grupo <i>repleta</i> do subgênero <i>Drosophila</i>	17
2. Objetivo.....	21
2.1 Objetivo geral.....	21
2.2 Objetivos específicos.....	21
3. Metodologia.....	22
3.1 Espécies de <i>Drosophila</i> e linhagens.....	22
3.2 Amplificação de <i>Cid</i> e purificação do produto de PCR.....	22
3.3 Ligação do inserto no vetor plasmidial e transformação bacteriana.....	24
3.4 Sequenciamento de DNA e montagem das sequências	25
3.5 Reconstrução de árvores filogenéticas.....	25
3.6 Análise de seleção positiva.....	26
4. Resultados e discussão.....	27
4.1 Origem da duplicação <i>Cid1-Cid6</i>	28
4.2 Isolamento de parálogos de <i>Cid</i> em espécies do grupo <i>repleta</i>	30
4.3 Alinhamento e caracterização de parálogos de <i>Cid</i> em espécies do grupo <i>repleta</i>	33
4.4 Inferência filogenética.....	40
4.5 Teste de seleção positiva nos parálogos de <i>Cid5</i> e <i>Cid6</i>	43
5. Conclusões.....	46
6. Referência Bibliográfica.....	48

1. Introdução

1.1. Centrômero e DNA centromérico

Durante a divisão celular em eucariotos a região do centrômero desempenha uma importante função na interação das proteínas cinetocóricas com os microtúbulos, permitindo a adequada segregação dos cromossomos durante a formação das células-filhas (Cheerambathur e Desai, 2014). Letalidade, câncer e infertilidade estão, em alguns casos, relacionados com aberrações cromossômicas decorrentes de erros de segregação, o que demonstra a importância dos centrômeros para a correta divisão celular meiótica e mitótica (Thompson et al., 2010; Santaguida e Amon, 2015).

Em cromossomos monocêntricos, o centrômero é visualizado como uma constrição visível durante a metáfase, sendo uma região especializada onde ocorre a formação do cinetócoro, um complexo multiproteico, que intermedia a conexão do DNA centromérico com proteínas cinetocóricas e microtúbulos responsáveis pela migração dos cromossomos para os polos celulares (Alberts et al., 2017). Entretanto, alguns organismos, como o *Caenorhabditis elegans* (nemátoda) e *Lanuzia nivea* (planta), possuem cromossomos holocêntricos, onde os microtúbulos se ligam a regiões centroméricas dispersas por toda a extensão do cromossomo (Dernburg, 2001; Nagaki et al., 2005) (**Figura 1**).

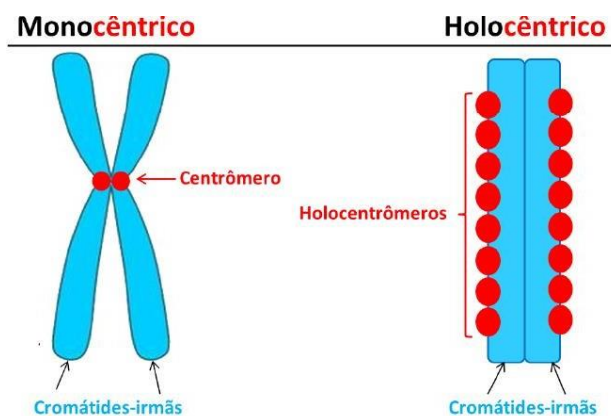


Figura 1. Modelo esquemático de um cromossomo monocêntrico e um cromossomo holocêntrico (retirado de Silva, 2016).

O DNA centromérico ainda é tópico de grande interesse para os pesquisadores, uma vez que sua elucidação perpassa pelas dificuldades do sequenciamento e montagem genômica. Localizado em uma região de heterocromatina, entre seus componentes mais abundantes

estão os DNAs satélites (satDNA) (Heslop-Harrison e Schwarzacher 2013; Melters et al., 2013) e os elementos transponíveis (TEs) (Charlesworth et al., 1994).

Os satDNAs são formados por sequências de DNA que se repetem em *tandem* formando arranjos que podem chegar a milhões de pares de bases (pb) (Charlesworth et al., 1994). Eles evoluem de acordo com os princípios da evolução combinada, onde sequências repetitivas são homogeneizadas e fixadas na população (Dover, 1982) através de mecanismos de recombinação não-recíproca, como *crossing-over* desigual e conversão gênica (Dover, 1986; Dover 2002). Apesar da região centromérica já ter sido considerada como livre de recombinação, *crossing-over* desigual e conversão gênica já foram identificados no dinamismo de satDNAs (Mahtani e Willard 1998; Talbert e Henikoff, 2010).

Os elementos transponíveis (TEs) são sequências de DNA que podem se deslocar para diferentes regiões do genoma e são capazes de formar repetições quando replicados durante o processo de transposição (Tollis e Boissinot, 2012). Podem ser categorizados como autônomos, quando codificam as enzimas mediadoras da transposição, e não-autônomos, quando utilizam as enzimas de outros TEs (Griffiths, et al., 2016). Quanto ao mecanismo de transposição, podem ser divididos em um grupo mediado por RNA, chamado de retroelementos (LTR e não-LTR) e outro grupo mediado por DNA, chamado de transposons de DNA (Plohl et al., 2014).

Apesar do acima exposto, nem a composição rica em DNA repetitivo ou a presença de TEs parecem ser determinantes ou suficientes para identificar o que seria o centrômero (Karpen e Allshire, 1997), uma vez que, em neocentrômeros, proteínas centroméricas são encontradas mas não necessariamente há satDNAs ou TEs (Marshall et al., 2008). Contudo, um centrômero funcional pode ser definido pela interação entre o DNA centromérico e a proteína centromérica CenH3 (*Centromeric Histone 3*), que é uma variante da histona H3, encontrada apenas na cromatina do centrômero (Palmer et al., 1991; Dawe e Henikoff 2006; Blower et al., 2002) (**Figura 2**).



Figura 2. Modelo de montagem do centrômero mostrando a interação entre CenH3 com o DNA centromérico (Adaptado de Carroll et al., 2010).

1.2. Proteína centromérica CenH3

Algumas proteínas associadas ao centrômero foram descritas na década de 1980 (Earnshaw et al., 1984; Hadlaczky et al., 1989). Entre elas a CenH3, denominada *CENTromeric Protein A* (CENP-A), que inicialmente foi identificada em humanos como um antígeno reconhecido pelo antissoro de pacientes com a síndrome CREST (**C**alcinose, fenômeno de **R**aynaud, dismotilidade **E**sofágica, e **S**clerodermia e **T**elangiectasia) (Earnshaw e Rothfield, 1985). Em 1991, experimentos de Palmer et al., demonstraram que a CENP-A não era uma histona H3 com modificações pós-traducionais, mas sim uma variante da histona H3 e que, portanto, advinha de um produto gênico distinto do gene da histona H3 canônica (**Figura 3**).

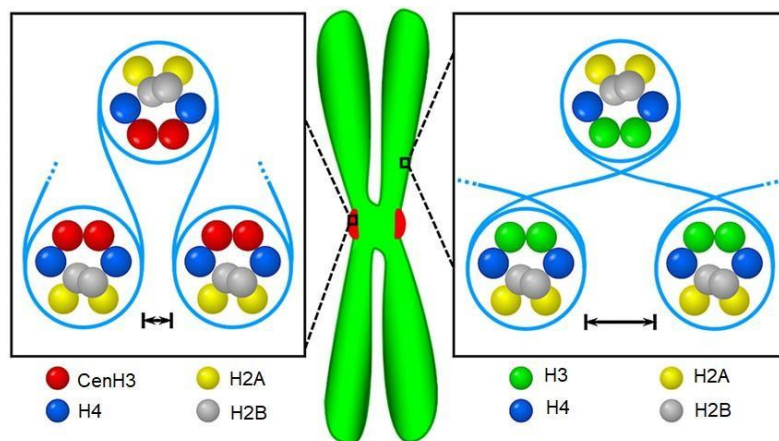


Figura 3. Disposição da histona H3 e da variante centromérica CenH3 no cromossomo. (Retirado de Panchenko et al., 2011).

Homólogos da CenH3 logo começaram a ser descritos em diferentes organismos modelos, como o Cse4 em *Saccharomyces cerevisiae* (Meluh et al., 1998), HCP-3 em *Caenorhabditis elegans* (Buchwitz et al., 1999), Cnp1 em *Saccharomyces pombe* (Takahashi et al., 2000), Cid em *Drosophila melanogaster* (Henikoff et al., 2000), HTR12 *Arabidopsis thaliana* (Talbert et al., 2002) e em outras espécies de eucariotos. Sua importância para a formação do cinetócoro tornou-se evidente após uma série de experimentos demonstrarem o efeito da depleção da CenH3, com proteínas do cinetócoro localizando-se em lugares inesperados, e da superexpressão da CenH3, levando a deposição de CenH3 em regiões não centroméricas e formação de cinetócoros ectópicos (Heun et al., 2006; Blower e Karpen, 2001; Blower et al., 2002).

Basicamente, as histonas e a CenH3 podem ser divididas em duas regiões, o domínio da cauda N-terminal (NTT) e o domínio C-terminal. A NTT varia tanto em tamanho quanto em

composição dos aminoácidos quando comparada entre espécies ou com a histona canônica H3 (Sullivan et al., 1994; Yoda et al., 2000; Henikoff e Dalal, 2005; Henikoff e Smith 2015). Em *Drosophila melanogaster*, a NTT possui aproximadamente 130 aminoácidos enquanto humanos e fungos possuem 45 e 20 aminoácidos, respectivamente, (Henikoff et al., 2001; Henikoff e Dalal, 2005).

Já a região C-terminal da CenH3, também conhecida como *Histone Fold Domain* (HFD), é mais conservada entre as espécies e mostra maior identidade com a mesma região da histona canônica H3 (Sullivan et al., 1994, Malik e Henikoff, 2003), possuindo a αN e 3 α -hélices separadas por dois *loops* (Black et al., 2004). Black et al., (2007) mostraram que a região que compreende o *loop1* e α -hélice2 seria essencial no direcionamento da CenH3 para o nucleossomo centromérico, sendo denominada inicialmente de CATD (*CENP-A target domain*). É na HFD da CenH3 que o domínio NTT da chaperona Cal1 interage e forma uma estrutura receptível para a interação de outra proteína centromérica, conhecida como CENP-C (Schittenhelm et al., 2010), sendo essa relação interdependente entre as proteínas centroméricas necessária para sua correta função e localização no centrômero (Chen et al., 2014; Roure et al., 2019) (**Figura 4**).

Na histona canônica H3, o *loop1* faz contato com o DNA no nucleossomo canônico e, sendo assim, esperava-se que em CenH3 a função fosse a mesma. Porém, estudos revelaram que a região do *loop1* na CenH3 permanece exposta na superfície do nucleossomo centromérico (Black et al., 2004). Rosin e Mellone (2016), sugeriram que o *loop1* estaria em contato com outras proteínas centroméricas, sendo, portanto, responsável pelo reconhecimento e recrutamento das mesmas (Roure et al., 2019).

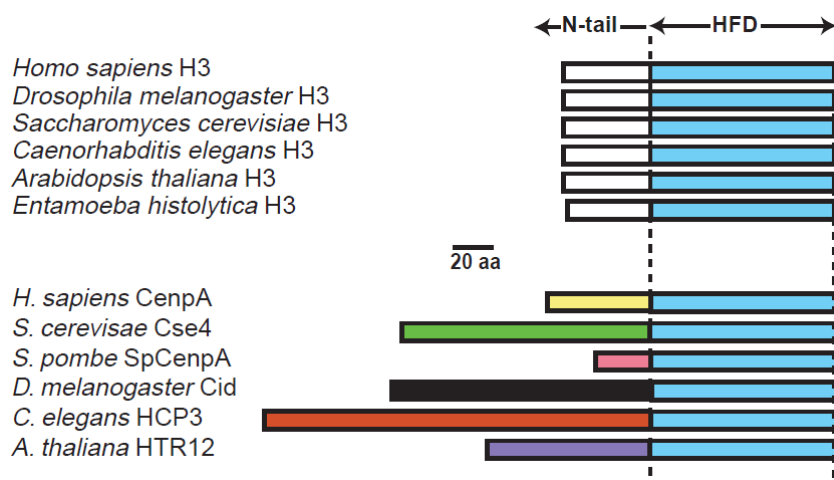


Figura 4. Esquema da comparação da histona H3 com a CenH3 de diferentes linhagens. O HFD é conservado entre as duas proteínas. Entretanto, enquanto a NTT da histona H3 é quase invariável, a NTT da CenH3 varia em sequência e tamanho entre as linhagens (retirado de Malik e Henikoff, 2003).

Devido à importância funcional do centrômero, esperaria-se uma conservação evolutiva do DNA centromérico e das proteínas centroméricas. Entretanto, evidencia-se uma variada e acelerada taxa de evolução em ambos, tanto em animais como em plantas (Lee e et al., 2005; Kuhn et al., 2008; Ugarkovic, 2009; Finseth et al., 2015).

1.3 A hipótese do impulso centromérico

Em 2001, Henikoff et al. propuseram uma hipótese que poderia explicar o paradoxo de proteínas e DNA centromérico evoluírem rapidamente apesar de comporem uma região cromossômica com papel tão essencial para a manutenção da vida em eucariotos. De acordo a hipótese do Impulso Centromérico (Henikoff et al., 2001; Malik 2009, Kursel e Malik, 2017), em consequência da expansão do DNA centromérico em um cromossomo, mais proteínas centroméricas seriam recrutadas, aumentando as chances desse cromossomo ser transmitido para a próxima geração durante a meiose feminina, uma vez que o ócito contém apenas um dos quatro produtos meióticos (Caryl et al., 2003; Schu e Ellenberg, 2008; Zou et al., 2008). O efeito dessa expansão e sua influência na meiose das fêmeas podem aumentar a frequência desse cromossomo na população e, concomitantemente, resultar em efeitos negativos, uma vez que esses cromossomos expandidos podem reter mutações deletérias (Mckee et al., 1998; Eaker et al., 2001) e em machos podem estar relacionados com o aumento da infertilidade (Fishman e Saunders, 2008; Mckee et al., 1998; Eaker et al., 2001) (**Figura 5**). Dessa forma, qualquer alelo supressor do impulso centromérico, que tenha o efeito de restaurar a paridade meiótica dos centrômeros com diferentes forças meióticas, poderá ser vantajoso e, conseqüentemente, aumentar a sua frequência na população (Kursel e Malik, 2017; Teixeira et al., 2018).

O resultado desse processo gera um ciclo onde os satDNAs centroméricos expandem, recrutam mais CenH3 e causam o impulso centromérico, que é então suprimido pelas mudanças em CenH3 que potencialmente restauram a paridade meiótica entre os cromossomos com diferentes quantidades de satDNAs (Henikoff et al., 2001; Kursel e Malik, 2017).

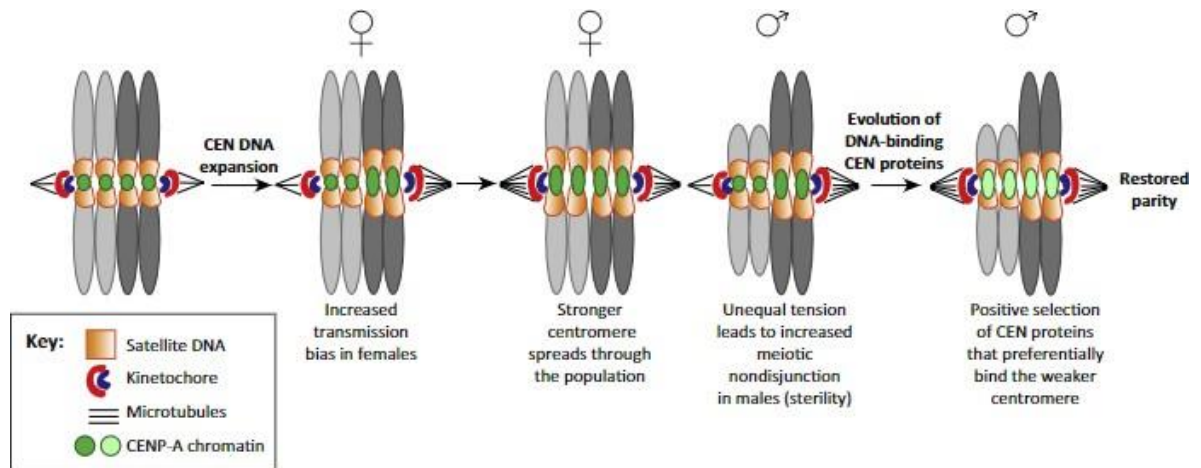


Figura 5. Modelo do impulso centromérico. Em uma primeira fase, a expansão do DNA centromérico aumenta a interação com proteínas centroméricas e fibras do fuso, o que pode levar a uma vantagem na meiose feminina ou a uma série de efeitos deletérios em meiose masculina. Em uma outra situação, a proteína centromérica sofre uma alteração que acaba suprimindo o impulso centromérico por reduzir a afinidade com o DNA centromérico, restaurando, assim, a paridade durante a divisão celular. (Extraído de Rosin e Mellone, 2017)

Experimentos na planta *Mimulus* demonstraram que cromossomos que continham o *locus D*, onde houve expansão na região do centrômero, eram transmitidos com maior frequência e que a duplicação de CenH3 em *Mimulus* poderia ser resultado de um processo de supressão do impulso centromérico (Finseth et al., 2015). Já em estudos usando camundongos, foi observado que a segregação preferencial dos cromossomos está associada com o maior recrutamento de microtúbulos e proteínas centroméricas para a formação do cinetócoro, como a CENP-A, além de correlacionar essa segregação preferencial com a presença de repetições em *tandem* no centrômero (Chmatal et al., 2014; Iwata-Otsubo et al., 2017),

O gene que codifica CenH3 está presente em cópia única na grande maioria das espécies de eucariotos estudada (Malik, 2009). Recentemente, Teixeira (2016) descreveu que em espécies do subgênero *Drosophila* há duas cópias funcionais de *Cid* (*Centromere identifier*), homólogo de CenH3 em *Drosophila*, inicialmente chamadas de *Cid-A* (ancestral) e *Cid-B* (nova cópia). Nas espécies do subgênero *Sophophora* (ao qual *D. melanogaster* pertence), existe apenas uma cópia de *Cid*. Teixeira (2016) ainda relatou que a cópia *Cid-A* possui expressão em todos os tecidos, mas com expressão reduzida nos machos, enquanto *Cid-B* possui expressão exclusiva nos machos. Este resultado levantou a hipótese de que após a duplicação gênica de *Cid*, *Cid-A* manteve a função centromérica canônica enquanto *Cid-B* se diferenciou e, possivelmente, desempenharia papel supressor do impulso

centromérico na meiose masculina. Em apoio a esta hipótese, foi evidenciado que *Cid-A* está evoluindo sob seleção purificadora, enquanto *Cid-B* evolui por seleção positiva (Teixeira, 2016). No entanto, apenas cinco espécies do grupo *repleta* (subgênero *Drosophila*) foram utilizadas nesta análise, sendo necessária uma maior amostragem de espécies para que resultados mais robustos sejam obtidos.

Esta duplicação acima descrita de *Cid* foi independentemente descoberta e investigada por Kursel e Malik (2017), que chamaram estes parálogos de *Cid1* (*Cid-A*) e *Cid5* (*Cid-B*). Por este motivo, no presente trabalho, estes parálogos também foram denominados de *Cid1* e *Cid5*. Kursel e Malik (2017) ainda descreveram três outros parálogos funcionais de *Cid*, todos eles em diferentes espécies do subgênero *Sophophora*: *Cid2* em *Drosophila eugracilis* e *Cid3* e *Cid4* em espécies do subgrupo *montium*.

O trabalho realizado por Teixeira et al., (2018) revelou ainda um novo parálogo, chamado de *Cid6*, presente em duas espécies investigadas do cluster *buzzatii* do grupo *repleta* (subgênero *Drosophila*), *Drosophila buzzatii* e *Drosophila seriema*. Nestas espécies, o gene *Cid1* sofreu uma duplicação inter-cromossômica (primeiro relato em eucariotos para CenH3), gerando a cópia *Cid6*. A cópia *Cid1*, por sua vez, sofreu degeneração por várias inserções de elementos transponíveis. Portanto, a hipótese é a de que *Cid6* desempenha o papel canônico de *Cid1* nestas espécies. Estas espécies também possuem *Cid5*, que como mencionado anteriormente, acredita-se atuar em função supressora de impulso centromérico na meiose de machos (**Figura 6**).

1.4 *Drosophila* como organismo modelo

Grande parte dos estudos moleculares, evolutivos e de função centromérica da CenH3 foram realizados utilizando algum organismo modelo, sendo *Drosophila* um dos mais utilizados (Teixeira, 2016).

As moscas do gênero *Drosophila* são facilmente capturadas no meio natural ou cultivadas no laboratório, têm proles numerosas, curtos períodos de geração, ocupam pouco espaço físico para manutenção de estoques, são facilmente manipuladas geneticamente e possuem, além disso, um genoma compacto em poucos cromossomos onde há genes ortólogos associados com doenças humanas (Bier, 2005; Prüßing et al., 2013).

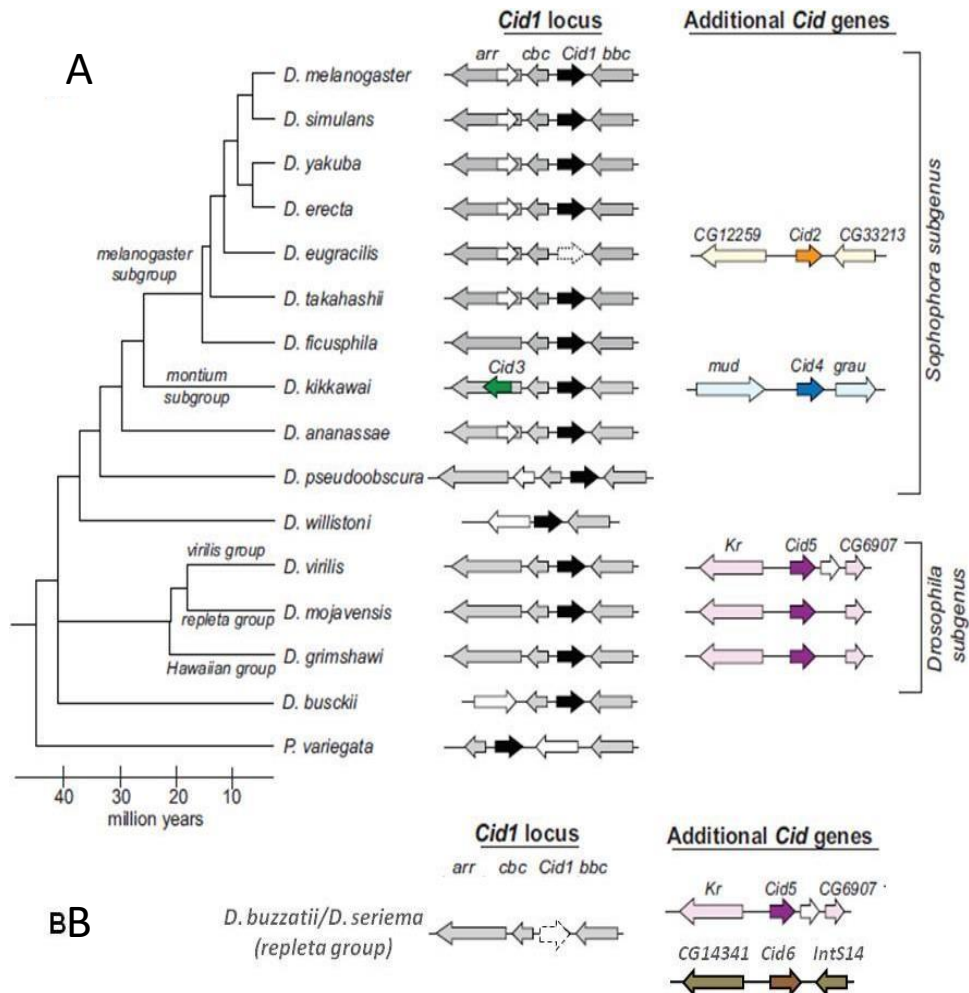


Figura 6. Parálogos de *Cid* identificados em *Drosophila*. **(A)** Parálogos *Cid1*, *Cid2*, *Cid3*, *Cid4* e *Cid5*, descritos por Kursel e Malik (2017) e **(B)** parálogo *Cid6*, descrito por Teixeira e et al. (2018). (Figura adaptada de Kursel e Malik, 2017)

A primeira espécie de *Drosophila* a ter o genoma sequenciado foi *Drosophila melanogaster* (Adams et al., 2000). Posteriormente, o genoma de mais 11 espécies de *Drosophila* foi publicado no projeto *Drosophila 12 Genomes Consortium* (Clark et al., 2007). Atualmente, mais de 20 espécies de *Drosophila* possuem o genoma sequenciado, sendo alguns desses genomas resultado de projetos individuais de alguns laboratórios (Garrigan et al., 2012; Zhou e Bachtrog, 2012; Nolte et al., 2013; Guillén et al., 2014; Teixeira et al., 2018). Esses dados são importantes para que análises evolutivas possam ser estudadas com maior profundidade.

1.4.1 O grupo *repleta* do subgênero *Drosophila*

Existem mais de 2.000 espécies descritas para o gênero *Drosophila*, popularmente conhecidas como “moscas das frutas” (Powel, 1997). Devido à grande quantidade de espécies, drosofilistas criaram um sistema de classificação que reconhece, após gênero,

grupos, complexos e clusters de espécies. De acordo com dados biogeográficos e fossilíferos, a separação entre os dois subgêneros, *Sophophora* e *Drosophila*, ocorreu entre 40-60 milhões de anos atrás (Powel, 1997; Sturtevant, 2001; O'Grady e DeSalle, 2018; Yassin, 2013). A **Figura 7** mostra as relações filogenéticas de espécies de *Drosophila* representativas dos principais grupos de *Drosophila*.

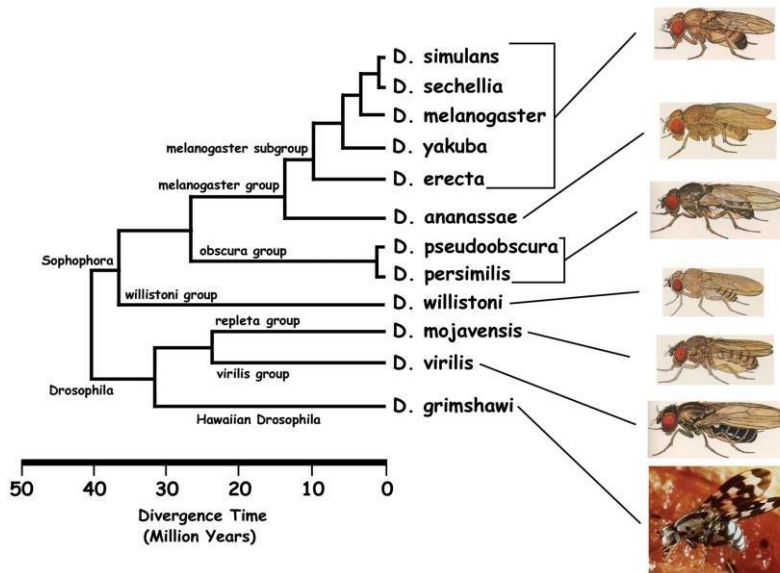


Figura 7. Relação filogenética de espécies de *Drosophila* representativas dos principais grupos de *Drosophila*. (Retirado em reinnovsac.com/)

Dentro do subgênero *Drosophila* temos o grupo *repleta* (**Figura 7**), representado por mais de 100 espécies organizadas em cinco subgrupos, sendo eles: *repleta*, *mercatorum*, *hydei*, *mulleri* e *fascicola* (O'Grady e DeSalle, 2018; Durando et al., 2000). Mesmo algumas espécies sendo generalistas, o grupo *repleta* é conhecido por possuir espécies cactofílicas que são capazes de degradar uma série de toxinas produzidas pelos cactos durante a decomposição dos mesmos e conseguem explorar esse recurso natural (Markow e O'Grady, 2006). A cactofilia pode ser considerada uma apomorfia dessas espécies, uma vez que não é observada em outras espécies fora do grupo *repleta* (Kircher, 1982; Starmer et al., 1986).

O cluster *mojavensis* pertence ao complexo *mulleri*, subgrupo *mulleri*, grupo *repleta*. Esse cluster é composto por três espécies, sendo elas: *Drosophila mojavensis* (Patterson 1940), *D. arizonae* (Ruiz, Heed e Wasserman, 1990) e *D. navojoa* (Ruiz, Heed e Wasserman, 1990), sendo as duas primeiras espécies irmãs diferenciadas por inversões cromossômicas e a última considerada a espécie basal do cluster, com características morfológicas distintas em relação as outras duas (Ruiz, Heed e Wasserman, 1990). Essas espécies são encontradas nas regiões áridas da América do Norte sendo *D. mojavensis* a espécie de maior distribuição geográfica (Markow e O'Grady, 2006).

Já o complexo *buzzatii*, subgrupo *mulleri*, grupo *repleta*, compreende 13 espécies que são identificadas morfológicamente pelo aparelho reprodutor masculino, edeago, e citologicamente por inversões cromossômicas observadas em lâminas de cromossomos politênicos (Ruiz e Wasserman, 1993). Dentro do complexo *D. buzzatii* ainda é possível separar as espécies em três clusters, sendo eles o cluster *martensis*: *D. martensis* (Wasserman e Wilson, 1957); *D. starmeri* (Wasserman et al., 1973); *D. uniseta* (Wasserman et al., 1973); *D. venezolana*, (Wasserman, Fontdevila e Ruiz, 1983), o cluster *D. stalker*: *D. richardsoni* (Vilela, 1983) e *D. stalker* (Wheeler, 1954), e por fim o cluster *buzzatii*: *D. buzzatii*, (Patterson e Wheller, 1942); *D. seriema* (Tidon-Sklorz e Sene, 1995); *D. koepferae* (Fontdevila e Wasserman, 1988); *D. antonieta* (Tidon-Sklorz e Sene, 2001), *D. gouveai* (Tidon-Sklorz e Sene, 2001); *D. borborema* (Vilela e Sene, 1977) e *D. serido* (Vilela e Sene, 1977) (**Figura 8**). Todas as espécies são endêmicas da América do Sul, com exceção da espécie semi-cosmopolita *D. buzzatii*, encontrada em outros continentes como consequência da dispersão feita pelo homem e de seus cactos hospedeiros do gênero *Opuntia* (Caryophyllales, Cactaceae) (Manfrin e Sene, 2006; Hasson et al., 2019, Barrios-Lealet al., 2019).

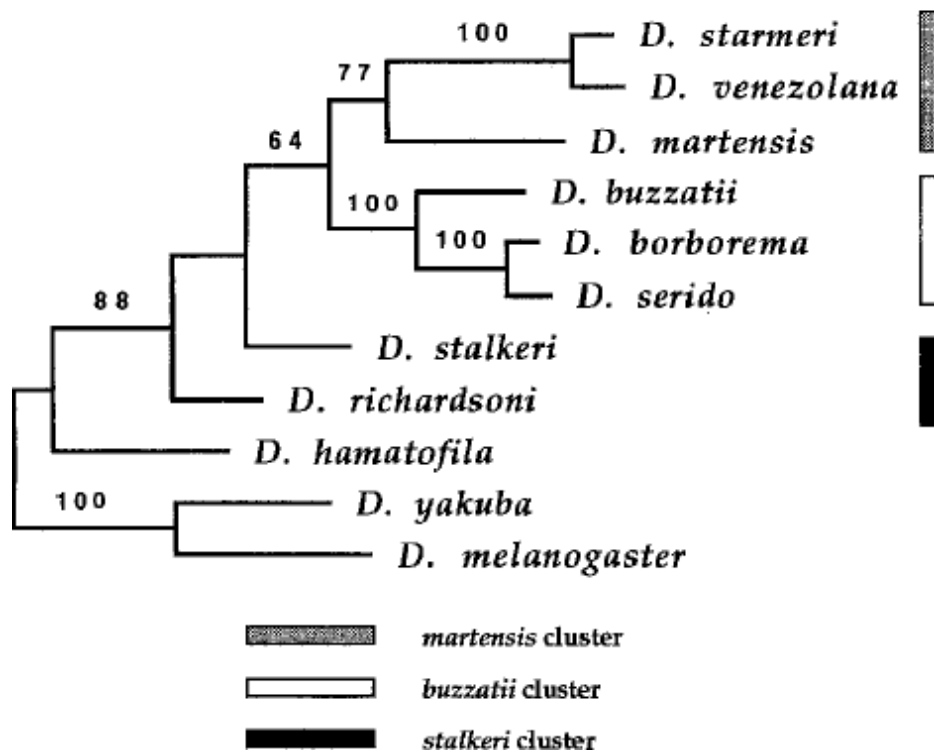


Figura 8. Filogenia do complexo *Drosophila buzzatii* usando o gene mitocondrial citocromo oxidase. (Retirado de Spicer, 1995).

Estudos com abordagem genética e evolutiva utilizando diferentes espécies de *Drosophila*, como as pertencentes do cluster *mojavensis* e do complexo *buzzatii*, são úteis para compreender processos evolutivos e moleculares. Dessa forma, levando em consideração as recentes descobertas sobre estrutura e função da CenH3, assim como a duplicação do gene que codifica a proteína centromérica em *Drosophila*, este trabalho procurou contribuir para o entendimento da CenH3 no grupo *repleta*.

2. Objetivo

2.1. Objetivo geral

O objetivo principal dessa pesquisa foi caracterizar e estudar a evolução dos parálogos *Cid1*, *Cid6* e *Cid5* em espécies do grupo *repleta*, com foco no cluster *Drosophila mojavensis* e no complexo *Drosophila buzzatii*.

2.2. Objetivos específicos

- Determinar a origem filogenética da duplicação *Cid1-Cid6*, através da investigação destes parálogos em *D. stalker* (cluster *stalker*) e *D. starmeri* (cluster *martensis*), espécies do complexo *buzzatii* que constituem grupos externos ao cluster *buzzatii*;
- Investigar se existem espécies no complexo *buzzatii* contendo ambos parálogos *Cid1* e *Cid6* funcionais (ao contrário da situação presente em *D. buzzatii* e *D. seriema*, onde *Cid1* se degenerou) e em caso afirmativo, caracterizá-los;
- Caracterizar os parálogos de *Cid* em espécies do grupo *repleta*;
- Testar se existe seleção positiva atuando nos parálogos de *Cid1*, *Cid5* e *Cid 6* nas espécies do grupo *repleta*.

3. Metodologia.

3.1 Espécies de *Drosophila* e linhagens.

Para a coleta de espécies e populações de *Drosophila* foram utilizadas armadilhas semi-fechadas contendo meio composto por banana madura macerada e fermento biológico seco. As armadilhas permaneceram no local de coleta por aproximadamente 48 horas até a captura dos indivíduos com o auxílio de uma rede entomológica. Os pontos de coleta foram definidos com base em estudos prévios de coleta de *Drosophila* no Estado de Minas Gerais (Franco e Manfrin, 2012).

As espécies do cluster *buzzatii* são morfologicamente indistinguíveis, com exceção da morfologia do edeago, aparelho reprodutor do macho (Tidon-Sklorz e Sene, 1995). Sendo assim, machos coletados foram identificados no nível de espécie e fêmeas foram selecionadas com o objetivo de criar isolinhagens para posterior identificação taxonômica a partir de machos nascidos no laboratório.

D. stalker (cluster *stalker*) e *D. starmer* (cluster *martensis*) foram obtidas diretamente da Cornell University (*Drosophila* Stock Centre) e *D. koepferae* e *D. buzzatii* (cluster *buzzatii*) foram gentilmente cedidas pelo professor Alfredo Ruiz da Universitat Autònoma de Barcelona.

3.2 Amplificação de Cid e purificação do produto de PCR

O DNA genômico foi extraído utilizando o kit comercial (Promega Kit Wizard® Genomic DNA Purification) seguindo as instruções do fabricante e quantificado por espectrofotometria (NanoDrop™ 2000/2000c Spectrophotometers).

Cada reação de PCR foi realizada de acordo com as seguintes condições (volume final de 25µL): 0,8mM de dNTP, 1.5mM de MgCl₂, 5pmol de cada *primer* (*Forward* e *Reverse*) específico para cada amplificação (**Tabela 1**), 5µL de tampão (Green GoTaq® Reaction Buffer), 1u de Taq polimerase (GoTaq® DNA Polymerase - Promega) e 25ng de DNA genômico da espécie estudada. O programa de amplificação das etapas de PCR consistiu em uma desnaturação inicial de 95°C por 5 minutos, seguido de 30 ciclos a 95°C por 1 minuto, de 40°C a 55°C (variando entre as espécies) por 1 minuto e 72°C por 1 minuto, uma extensão final a 10°C por 5 minutos e encerrando a reação com redução na temperatura a 4°C.

Tabela 1. *Primers* utilizados para amplificar *Cid1*, *Cid6* e *Ci5* por PCR convencional,

Nome do <i>primer</i>	Região de anelamento	Sequência 5'-3'	Orientação
mojC5CidF	<i>UpstreamCid1</i>	CACATCGTTCTAAAAGTTGC	<i>Forward</i>
mojC5CidR	<i>DownstreamCid1</i>	TGCATCCCTTACATACACAG	<i>Reverse</i>
GGD1601F	<i>UpstreamCid6</i>	TCAGTTTGTTTTATTGCCGC	<i>Forward</i>
GGD1602R	<i>DownstreamCid6</i>	GCACGGTTCAGTTTCGATT	<i>Reverse</i>
GGD1501F	<i>UpstreamCid5</i>	TAGCCAACGTGTAAGCTGTG	<i>Forward</i>
GGD1502R	<i>DownstreamCid5</i>	CACGAGCGTATTTAGAATCG	<i>Reverse</i>
GGD4101F	<i>bbc</i>	CGTTACTGTGCGCAGGTT	<i>Forward</i>
GGD4102R	<i>cbc</i>	GAGTCTGCAAACCAGA	<i>Reverse</i>
GGD4104R	<i>Cid1/Cid6</i>	CAACGCTGCGCAAGAACTC	<i>Reverse</i>
GGD4105F	<i>Cid1</i>	CGTGAAAACGCTAATCGG	<i>Forward</i>
GGD4601F	<i>CG14341</i>	TCCTCTCCGGGATCTGTGG	<i>Forward</i>
GGD4602R	<i>IntS14</i>	TGCAGAAGGCCAACGGCTGC	<i>Reverse</i>
GGD4604R	<i>Cid6</i>	CGTGAAAATGGCAATCGTGG	<i>Reverse</i>
GGD5601F	<i>Cid6</i>	CCACTAGCCGTTTAGCAC	<i>Forward</i>
GGD5602R	<i>Cid6</i>	CGCGACATGGCAATGGTC	<i>Reverse</i>
GGD5501F	<i>Cid5</i>	TGCGTTCGATCTGCGTTAC	<i>Forward</i>
GGD5502R	<i>Cid5</i>	TGGAGCTGCGTGACATGG	<i>Reverse</i>

Os *primers* mojC5CidF e mojC5CidR foram utilizados para amplificar o parálogo de *Cid1*, tendo como referência as regiões flanqueadoras de *Cid1* de *D. mojavensis*. Já os *primers* GGD1601F e GGD1602R, para o gene *Cid6*, e os *primers* GGD1501F e GGD1502R, para o gene *Cid5*, foram desenhados a partir da região flanqueadora de *Cid6* e *Cid5* de *D. buzzatii*, respectivamente.

Para investigação mais detalhada da presença dos parálogos *Cid1* e *Cid6* em espécies do complexo *buzzatii*, também foram utilizados os *primers* GGD4601F e GGD4602R que anelam, respectivamente, nos genes *CG14341* e *IntS14*, e os *primers* GGD4101F e GGD4102R, que anelam nos genes *bbc* e *cbc*. Quando necessário, os *primers* GGD4104R e GGD4105F (*Cid1*) e GGD4104R e GGD4604R (*Cid6*) em combinação com os *primers* acima descritos foram utilizados (**Tabela1**). Os genes *Kr* e *CG6907* possuem uma distância acima de 2kb em relação a *Cid5* de *D. buzzatii*, o que inviabilizou a utilização de *primers* que auxiliariam na investigação mais detalhada de *Cid5* em outras espécies do complexo *D. buzzatii*.

Os produtos da PCR foram visualizados em gel de agarose 1,5%, com auxílio de um transiluminador de LED azul (Blue Transilluminator 470nm) da Uniscience. A banda de DNA de interesse foi excisada para purificação utilizando o kit "illustra™ GFX™ PCR DNA and Gel Band Purification", da GE Healthcare, seguindo as especificações do fabricante. A

concentração e a pureza do produto purificado foram estimadas através de quantificação utilizando espectrofotômetro (NanoDrop™ 2000/2000c Spectrophotometers).

3.3 Ligação do inserto no vetor plasmidial e transformação bacteriana

Para a ligação dos produtos da PCR no vetor utilizou-se 15ng do DNA amplificado purificado, 50ng do plasmídeo vetor pGEM®-T Vector Systems, 5µl de tampão de ligação 10x, 1µl de T4 DNA ligase e água estéril suficiente para 10µl de reação e incubação por 16 horas a 4 °C. Os plasmídeos foram incubados com 45µl de bactérias eletrocompetentes de *E. coli*, linhagem XL1 Blue (Phonectria) por 1 minuto no gelo e então transferidos para uma cubeta e submetidos a um choque de 2500mV no eletroporador de pulso (Gene PulserXcell™ - Bio-Rad). Após o choque, foi adicionado 1ml de solução SOC (10 mM de MgCl₂, 20 mM de glicose), homogeneizado, transferido para tubos Eppendorf de 1,5ml e incubado a 37 °C por 1 hora em banho-maria.

Em placas contendo LB-ágar (10g de NaCl, 10g de triptona, 5g de extrato de levedura, 15g de ágar, água destilada q.s.p. 1 L, pH 7,5) e ampicilina (100µg/ml), foram acrescentados 30µl de X-gal (50mg/ml) e 4µl de IPTG (0,6mM). Nestas placas, foram distribuídos 150µl da suspensão de bactérias e as placas foram incubadas por aproximadamente 16 horas a 37°C. Para selecionar as colônias que possuíam o plasmídeo com o fragmento de DNA exógeno foi utilizado o método de seleção por antibiótico e o sistema IPTG/X-gal. A inserção do DNA de interesse ao plasmídeo acontece dentro do gene lacZ, que produz a enzima β-galactosidase que ao clivar o substrato X-gal libera um produto insolúvel de cor azul. Dessa forma, colônias que não possuem o inserto mantêm a integridade do lacZ e, portanto, terão a coloração azul. Já as colônias que possuem o DNA exógeno não terão a produção de β-galactosidase devido a inserção do DNA no gene lacZ, apresentando, assim, coloração branca (Lehninger et al., 2000).

Cada colônia branca foi isolada e transferida pra tubo com 5ml de meio LB-líquido (10g de NaCl, 10g de triptona, 5g de extrato de levedura, água destilada q.s.p. 1L, pH 7,5) com ampicilina (100µg/ml) contido em um tubo falcon de 15mL que foi em seguida colocado em um agitador a 220 rpm à temperatura de 37°C durante 16 horas. Tubos com meio turvo foram selecionados para teste de amplificação do vetor contendo o inserto. O teste de amplificação do vetor contendo o inserto consiste em uma PCR convencional, seguindo os mesmos parâmetros da PCR descrita no tópico 3.2, com a utilização dos *primers* que anelam ao vetor (M13F: 5' GTA AAA CGA CGG CCA GT 3'; M13R: 5' CAG GAA ACA GCT ATG ACC 3'). Espera-se que o produto amplificado possua, em Kb, a somatória do DNA de interesse mais aproximadamente 260pb que corresponde ao plasmídeo sem o inserto, sendo este visível em

eletroforese. Dessa forma, as amostras selecionadas foram submetidas a extração de DNA plasmidial utilizando o kit PureLink® QuickPlasmidMiniprep (Invitrogen).

3.4 Sequenciamento de DNA e montagem das sequências

O DNA plasmidial foi quantificado utilizando um espectrofotômetro (NanoDrop™ 2000/2000c Spectrophotometers). Para o sequenciamento, em um tubo contendo 100ng do DNA plasmidial, foi adicionado 1µL do *primer* M13 *forward* ou *reverse* e água estéril suficiente para o volume final de 7,5µL. As amostras foram sequenciadas utilizando o método de Sanger dideoxy na plataforma ABI 3130 (AppliedBiosystems). Para aumentar a qualidade do DNA sequenciado, o mesmo DNA plasmidial foi sequenciado usando os *primers* M13 *Forward* e o M13 *Reverse* em reações independentes.

As sequências de *Cid* (nucleotídeo e aminoácidos) foram alinhadas usando o programa *ClustalW* (Larkin et al., 2007), implementado no Geneious Primer® 2019.0.4 (Kearse et al. 2012). Quando necessário, o alinhamento foi refinado manualmente para corrigir *gaps* e regiões mal alinhadas.

3.5 Reconstrução de árvores filogenéticas

Árvores filogenéticas para os parálogos de *Cid* foram construídas utilizando os métodos de Máxima Verossimilhança (MV) e inferência Bayesiana (IB). Para a árvore de MV foi utilizado o modelo de substituição nucleotídica HKY (Hasegawa, Kishino & Yano, 1985). As configurações específicas do programa permaneceram no modo padrão sendo o suporte estatístico das árvores filogenéticas calculado utilizando *bootstrap* (Felsenstein, 1985) com 1.000 réplicas. Já para árvores IB o *software* utilizado foi MrBayes 3.2.6 (Ronquist et al., 2012) utilizando-se parâmetros no modo padrão, onde a análise passou por 10 milhões de gerações a partir de uma árvore inicial com cadeias MCMC (Monte Carlo via Cadeias de Markov). Usando o comando *sumt* a árvore consenso foi calculada e, posteriormente, visualizada e editada no programa Figtree v.1.4.2 (Rambaut, 2014).

3.6 Análise de seleção positiva

A análise de seleção positiva nos parálogos de *Cid* foi realizada por dois métodos:

i) Para detecção de seleção positiva do gene *Cid6* e *Cid5*, utilizamos o teste MK (McDonald e Kreitman, 1991) disponível no programa DnaSP v.6.12.03 (Rozas et al., 2017) usando quatro populações de *D. buzzatii* e duas populações de *D. seriema*. O teste MK detecta seleção positiva distribuída em um gene ou em partes específicas do gene, onde assume que se a proteína não sofreu grandes alterações ao longo do processo evolutivo, a razão entre mutações não-sinônimas e mutações sinônimas fixas entre espécies (dN/dS) seria aproximadamente igual a razão entre polimorfismo não-sinônimo e polimorfismo sinônimo dentro das espécies (pN/pS). O índice de seleção positiva após a divergência das espécies pode ser inferido quando o número maior que o esperado de mutações fixas não-sinônimas é observado (Yang, 2006).

ii) Utilizamos o algoritmo CODEML do pacote PAML (Yang, 1997) implementado no programa EasyCodeML v1.21 (Gao et al., 2019) para detectar seleção positiva nos parálogos de *Cid* em espécies do grupo *repleta*. Foram comparados três grupos de modelos: M2a versus M1a e M8 (β e ω) versus M7, sendo M1a e M7 seleção neutra e M2a e M8 para seleção positiva. Se o resultado do teste de verossimilhança (Likelihood ratio test – LRT) rejeitasse o modelo de seleção neutra em favor do modelo de seleção positiva, sítios que hipoteticamente estariam sobre essa seleção poderiam ser identificados utilizando os valores de probabilidade posterior > 0.95 e $\omega > 1$ a partir da abordagem bayesiana empírica (Bayes Empirical Bayes - BEB).

4. Resultados e discussão.

Como resultado de nossas coletas, quatro espécies foram identificadas como pertencentes ao cluster *D. buzzatii* (*D. serido*, *D. gouveai*, *D. seriema* e *D. buzzatii*) que, juntamente com as espécies adquiridas do Cornell Stock Centre (*D. stalker*i e *D. starmer*i) e as espécies e populações adquiridas de outros laboratórios (*D. koepferae* e *D. buzzatii* – Bu28), estão sumarizadas na **Tabela 2**.

Tabela 2. Espécies e populações de *Drosophila* do grupo *repleta* utilizadas nesse trabalho

Espécie	População	Local de Coleta ou origem do estoque.	Data Base IDs	
			Cid1/Cid6	Cid5
<i>D. buzzatii</i>	ST-01	LCEv Lab- MG	http://dbuz.uab.cat/blast.php (<i>D. buzzatii</i> Freeze 1 Scaffolds)	
	J-19	LCEv Lab - MG	-	-
	Bu28	Bolívia / GGBE Lab	-	-
	SC	Serra do Cipó- MG	-	-
<i>D. seriema</i>	D73C3	Bahia /LCEv Lab	ERX2037878	ERX2037878
	SC	Serra do Cipó - MG	-	-
<i>D. koepferae</i> *	KO-2	Argentina / GGBE Lab	-	-
<i>D. serido</i> *	SC	Serra do Cipó - MG	-	-
<i>D. gouveai</i> *	F	Furnas - MG	-	-
<i>D. starmer</i> i*	N/C	Stock Centre - EUA	-	-
<i>D. stalker</i> i*	N/C	Stock Centre - EUA	-	-
<i>D. mojavensis</i>	N/C	LCEv Lab - MG	XM_002006887.2	XM_018104694.1
<i>D. arizonae</i>	N/C	Banco de dados	XM_018010248.1	XM_018009273.1
<i>D. navojoa</i>	N/C	Banco de dados	XM_018104694.1	XM_018103086.1

* Espécies em que Cid foram estudadas pela primeira vez no presente trabalho

N/C: Não consta.

4.1 Origem da duplicação *Cid1-Cid6*

O estudo realizado por Teixeira et al. (2018) mostrou que o parólogo *Cid6*, que teve origem a partir da duplicação gênica, está presente em duas espécies do cluster *D. buzzatii*, o que mostra que esta duplicação já existia na espécie ancestral do cluster *buzzatii*. Ainda, o mesmo estudo mostrou que *Cid6* está ausente no cluster *mojavensis*, que se divergiu do cluster *buzzatii* há pelo menos 11 milhões de anos (Oliveira et al., 2012).

Para acessarmos a origem filogenética da duplicação que deu origem ao parólogo *Cid6*, investigamos a presença de *Cid1/Cid6* em espécies do clusters *martensis* e *stalker*, considerados basais no complexo *D. buzzatii* (**Figura 8**). Inicialmente, foram realizados experimentos de PCR utilizando os *primers* *mojC5CidF/mojC5CidR* e *GGD1601F/GGD1602R* (**Tabela 1**), que anelam, respectivamente, em regiões flangeadoras dos genes *Cid1* e *Cid6*, (**Figura9**).

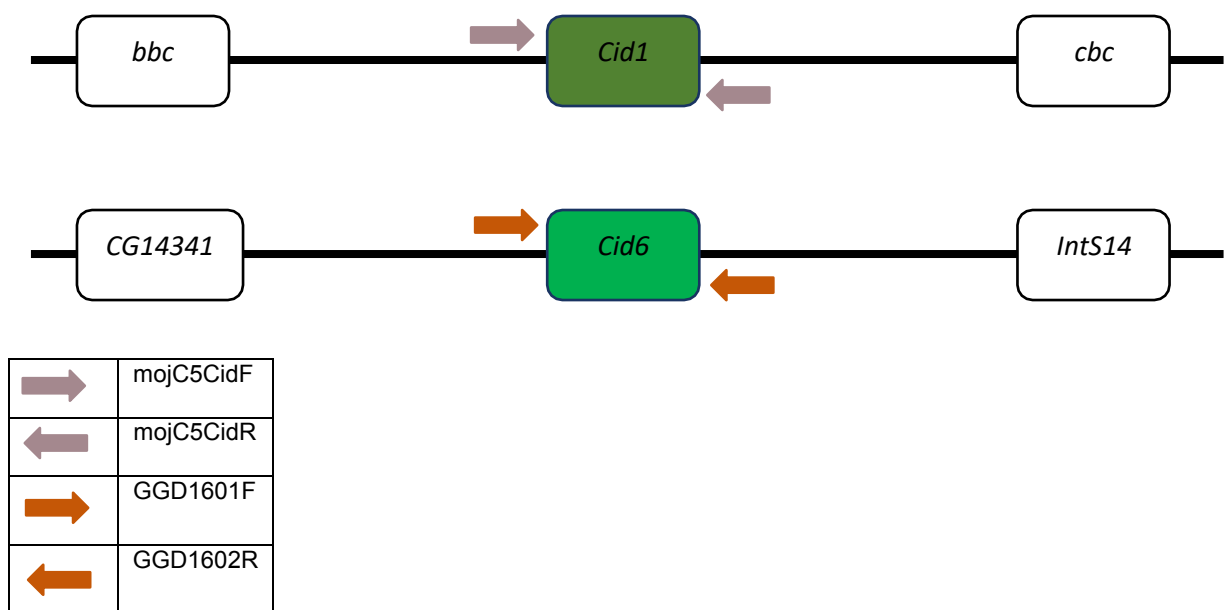


Figura 9. Esquema ilustrativo da região de anelamento dos primers flangeadores de *Cid1* e *Cid6*.

Nenhum produto de PCR foi observado para o gene *Cid1* e *Cid6* nas espécies *D. stalker* e *D. starmeri*. Este resultado ainda poderia ser consequência da falta de anelamento dos *primers* em uma região nucleotídica variável entre as espécies. Por este motivo, um teste com um novo conjunto de *primers* foi realizado. Dessa vez, foram utilizados os *primers* *GGD4601F* e *GGD4604R* (para amplificar do gene *CG14341* até o gene *Cid6*); *GGD4104R* e *GGD4602R* (para amplificar do gene *Cid6* até o gene *IntS14*) e *GGD4601F* e *GGD4602R* (para amplificar do gene *CG14341* até o gene *IntS14*) (**Tabela 1**) (**Figura 10**).

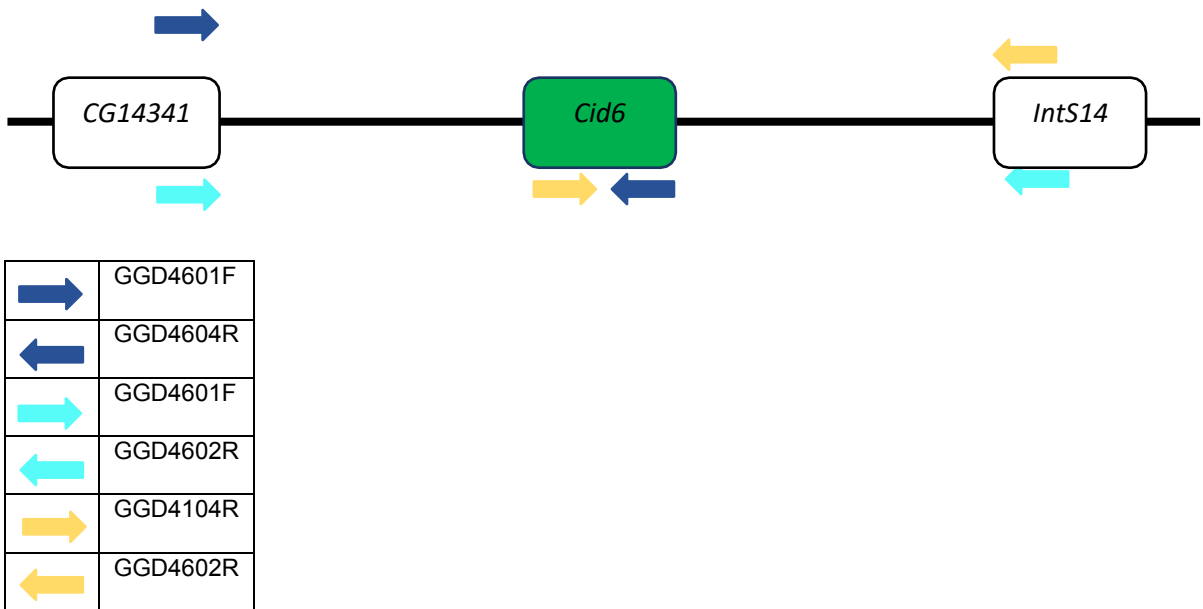


Figura 10. Esquema ilustrativo da região de anelamento dos *primers* em *CG14341*, *Cid6* e *IntS14*.

Após o resultado positivo da amplificação em *D. stalker* (**Figura 11**), o produto de PCR foi clonado e sequenciado, confirmando a presença de *Cid6* nesta espécie.

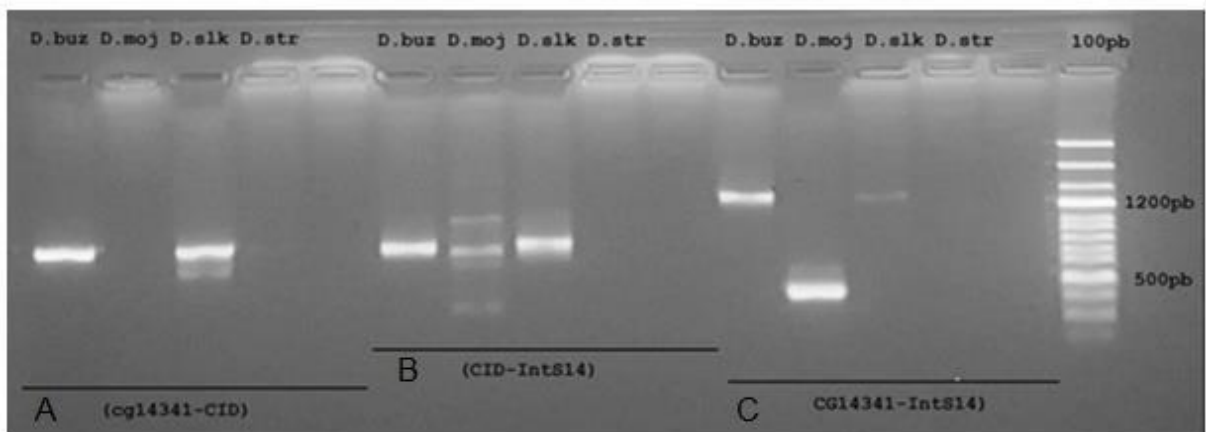


Figura 11. Resultado da amplificação entre os genes flanqueadores de *Cid6* em *D. stalker* e *D. starmeri*, usando como controles positivos *D. buzzatii* e *D. mojavensis*. **A.** Do gene *CG14341* até *Cid*. **B.** Do gene *Cid* até o gene *IntS14*. **C.** Do gene *CG14341* até o gene *IntS14*.

Não houve amplificação de *Cid6* em *D. starmeri* (**Figura 11**). Regiões intergênicas podem sofrer inserções e expansão de material genético, o que, nesse caso, poderia ter interferido na eficiência da amplificação por PCR. Portanto, foram utilizados os *primers* GGD5601F e GGD5602R (**Tabela 1**), que anelam internamente no gene *Cid6*, para verificar a presença ou ausência de *Cid6* em *D. starmeri* (**Figura 12**).

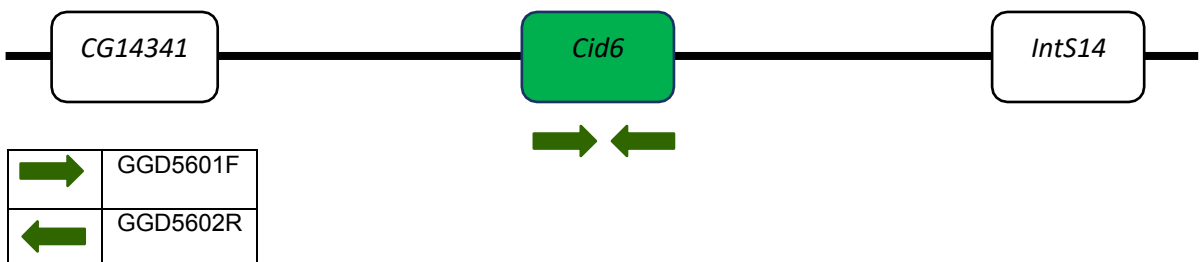


Figura 12. Esquema ilustrativo da região de anelamento dos *primers* internos de *Cid6*.

o resultado da PCR (**Figura 13**) indica a presença do parálogo Cid6 em *D. starkeri*, representante utilizada no presente trabalho do cluster *martensis*.

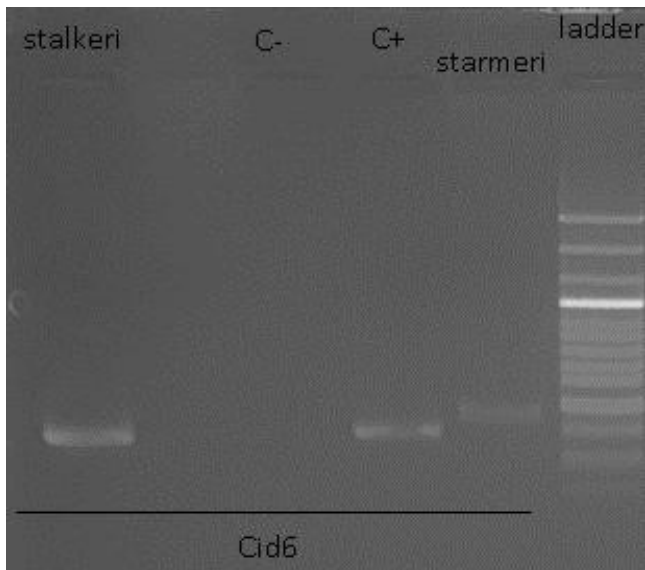


Figura 13. Resultado da amplificação do gene *Cid6* em *D. starkeri*, usando como controle positivo *Cid6* de *D. buzzatii*.

Em resumo, os resultados obtidos mostram a presença de *Cid6* em *D. starkeri* e muito provavelmente em *D. starkeri*, sugerindo assim, que a duplicação *Cid1-Cid6* já existia no ancestral comum do complexo *buzzatii*.

4.2 Isolamento de parálogos de *Cid* em espécies do grupo *repleta*.

Com a descoberta de que a duplicação *Cid1-Cid6* já estava presente no ancestral do complexo *buzzatii*, o próximo passo foi isolar e caracterizar os parálogos presentes nas espécies dos cluster *mojavensis* e do complexo *buzzatii*.

Para isolamento do parálogo *Cid1* foram utilizados os *primers* *mojC5CidF* e *mojC5CidR* (**Tabela 1**) (**Figura 9**). Como esperado, o gene *Cid1* foi amplificado apenas em *D. mojavensis*. Não foram observados produtos de PCR para *Cid1* em nenhuma das espécies do complexo *buzzatii* analisadas (*D. starkeri*, *D. starkeri*, *D. buzzatii*, *D. koepferae*, *D. gouveai*, *D. seriema* e *D. serido*), indicando que *Cid1* já deve ter sofrido degeneração na espécie ancestral do complexo *buzzatii* e sua função transferida para a nova cópia *Cid6*.

Já para o isolamento do gene *Cid6* em *D. mojavensis*, *D. buzzatii*, *D. koepferae*, *D. gouveai*, *D. seriema* e *D. serido* foram utilizados os *primers* *GGD1601F* e *GGD1602R* (**Tabela 1**) (**Figura 9**). Como esperado, *D. mojavensis* foi a única espécie em que não houve

amplificação de *Cid6*. Para *D. stalker*i e *D. starmer*i o processo de isolamento dos parálogos *Cid1-Cid6* foi descrito no tópico 4.1.

Para a amplificação do parálogo *Cid5* foram utilizados os *primers* GGD1501F e GGD1502R (**Tabela 1**) (**Figura 14**). Com exceção de *D. stalker*i, todas as outras espécies (*D. mojaven*is, *D. starmer*i, *D. buzzat*ii, *D. koepfer*ae, *D. gouve*ai, *D. seriema* e *D. serido*) tiveram o gene *Cid5* amplificado por PCR. Com o objetivo de verificar a presença de *Cid5* em *D. stalker*i, realizamos uma nova PCR com *primers* internos (GGD5501F e GGD5502R) do gene *Cid5* (**Tabela 1**, **Figura 14**), que resultou em amplificação de produto com tamanho esperado, confirmando a presença do parálogo *Cid5* nesta espécie (**Figura 15**).

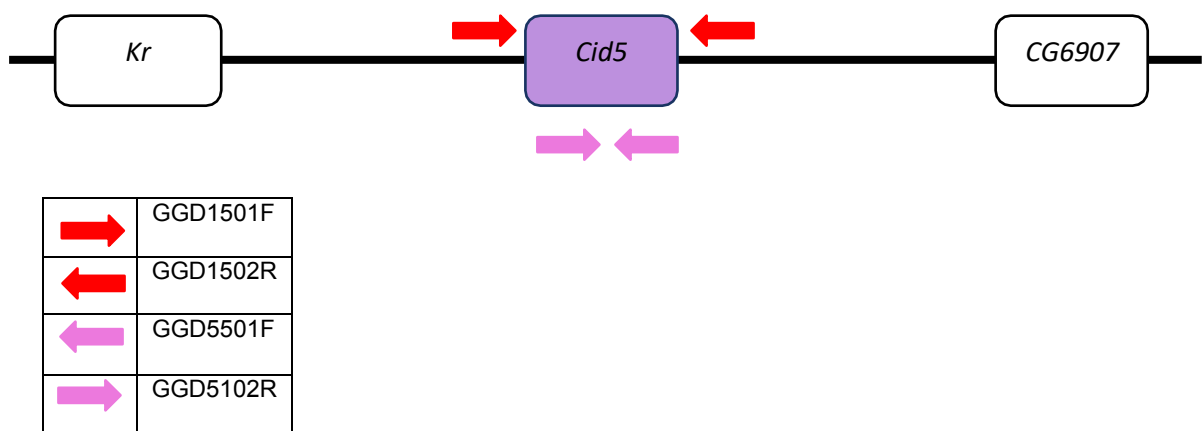


Figura 14. Esquema ilustrativo da região de anelamento dos primers flanqueadores e internos de *Cid5*.

Os resultados das PCRs e os *primers* utilizados para isolamento dos parálogos *Cid1*, *Cid6* e *Cid5* em *D. mojaven*sis, *D. stalker*i, *D. starmer*i, *D. buzzat*ii, *D. koepfer*ae, *D. gouve*ai, *D. seriema* e *D. serido* estão sumarizados na **Tabela 3** e a **Figura 16** mostra a presença dos parálogos *Cid1*, *Cid6* e *Cid5* no contexto da filogenia das espécies estudadas.

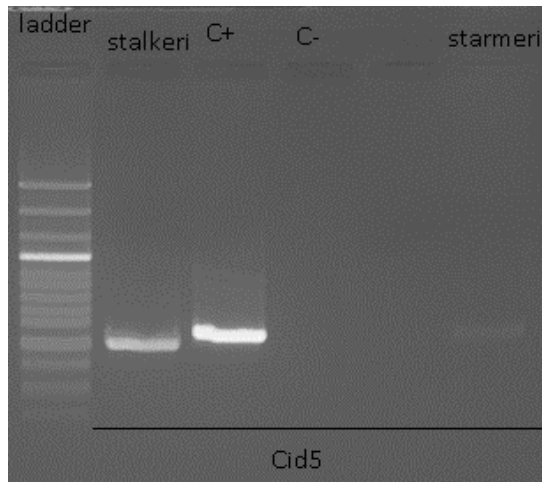


Figura 15. Resultado da amplificação dos gene *Cid5* em *D. stalkeri*, usando como controle positivo *Cid5* de *D. buzzatii*.

Tabela 3. Espécies do grupo *repleta*, primers usados para isolar *Cid1*, *Cid6* e *Cid5* e presença ou ausência do parálogo

Cluster	Espécie	Gene		
		<i>Cid1</i>	<i>Cid6</i>	<i>Cid5</i>
<i>mojavensis</i>	<i>D. mojavensis</i>	P	A	P
	Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R
	<i>D. arizonae</i>	P	A	P
	Primers	*	*	*
	<i>D. navojoa</i>	P	A	P
	Primers	*	*	*
<i>martensis</i>	<i>D. starmeri</i>	A	P	
	Primers	mojC5CidF e mojC5CidR	GGD5601FF e GGD5602R	GGD1501F e GGD1502R
<i>stalkeri</i>	<i>D. stalkeri</i>	A	P	P
	Primers	mojC5CidF e mojC5CidR	GGD4104R, GGD4601F, GGD4602R e GGD4604R	GGD5501F e GGD5502R
<i>buzzatii</i>	<i>D. buzzatii</i>	A	P	P
	Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R
	<i>D. seriema</i>	A	P	P
	Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R
	<i>D. koepferae</i>	A	P	P
	Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R
	<i>D. gouveai</i>	A	P	P
	Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R
<i>D. serido</i>	A	P	P	
Primers	mojC5CidF e mojC5CidR	GGD1601F e GGD1602R	GGD1501F e GGD1502R	

A: Ausência de amplificação

P: Presença de amplificação

* : Publicado em Teixeira et al., 2018

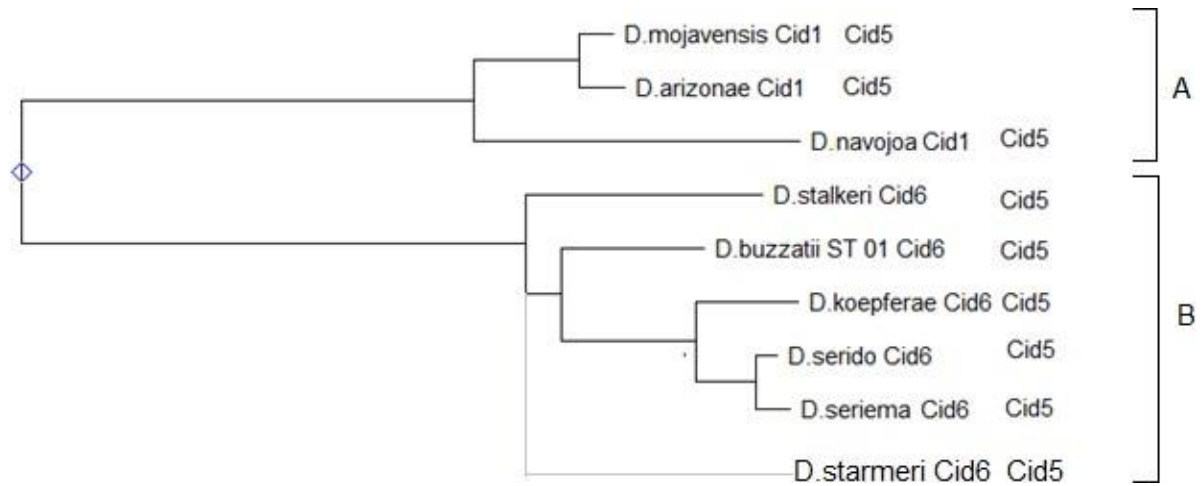


Figura 16. Representação dos parálogos de Cid no grupo repleta. **A.** cluster *mojavensis*. **B.** complexo *buzzatii*.

4.3 Alinhamento e caracterização de parálogos de Cid em espécies do grupo repleta.

De posse das novas sequências de *Cid1*, *Cid6* e *Cid5*, realizamos um alinhamento destas sequências com as previamente descritas por Teixeira et al.,(2018). Desta forma, foi possível verificar a presença de todos os principais domínios de Cid (**Figura 17**), destacando os dois domínios, a calda N-terminal (NTT) e a *Histone Fold Domain* (HFD) com suas principais regiões, αN , α -hélice 1, α -hélice 2, α -hélice 3, *loop1* e *loop2*. A **Tabela 4** sumariza as principais características dos parálogos Cid1, Cid6 e Cid5 em relação ao tamanho em pb e os aminoácidos mais abundantes nas espécies estudadas.

Tabela 4. Principais características dos parálogos *Cid1-Cid6* e *Cid5* em relação ao tamanho em pb, composição nucleotídica e de aminoácidos mais abundantes.

Espécie	Tamanho em pb							
	<i>Cid1/Cid6</i>				<i>Cid5</i>			
	Gene	NTT	HFD	Aminoácidos mais abundantes	Gene	NTT	HFD	Aminoácidos mais abundantes
<i>D. buzzatii</i> **	687	387	300	Thr, Arg	660	354	306	Pro, Arg,
<i>D. koepferae</i>	690	390	300	Thr, Arg	639	333	306	Arg, Pro
<i>D. seriema</i> **	693	393	300	Thr, Arg	627	321	306	Arg, Pro
<i>D. serido</i>	693	393	300	Thr, Arg	639	333	306	Arg, Pro
<i>D. gouveai</i>	*	*	*	*	627	321	306	Arg, Pro
<i>D. starmeri</i>	*	*	*	*	621	312	309	Arg, Pro
<i>D. stalker</i>	690	390	300	Thr, Arg	*	*	*	*
<i>D. mojavensis</i> **	624*	324	300	Leu, Thr	645	339	306	Arg, Pro
<i>D. arizonae</i> **	624*	324	300	Leu, Arg	645	339	306	Arg, Leu
<i>D. navjoa</i> **	618*	318	300	Ser, Arg	657	351	306	Glu, Arg
Media	665	365	300	Thr, Ser	640	334	306	Arg, Pro

* Produto de PCR não foi sequenciado ou caracterizado, sugerindo apenas a presença do gene.

** Dados previamente publicados.

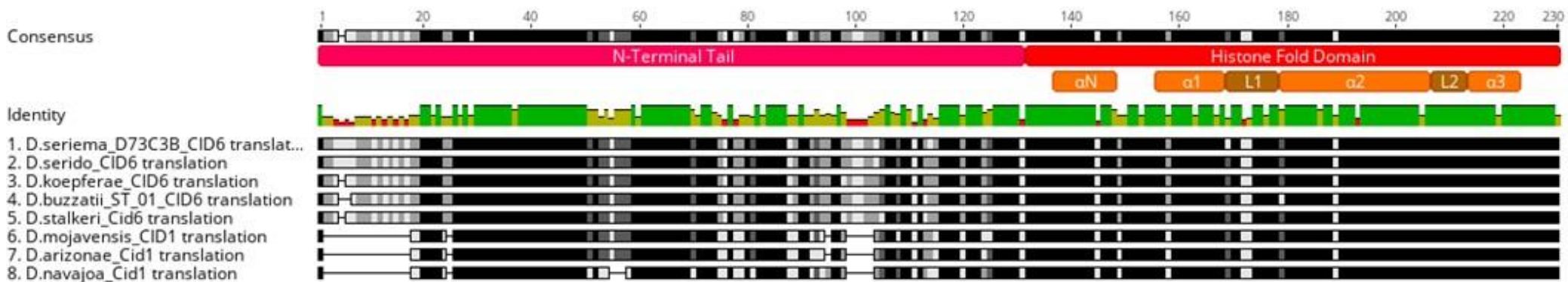
Modificações que acontecem em resíduos de aminoácidos influenciam nas propriedades químicas das proteínas e suas funções biológicas (Clarke e Tamanoi, 2006; Bedford e Clark, 2009). Arginina, por exemplo, em complexos envolvendo a interação proteína-DNA é um sítio doador de elétron para o grupo fosfato, onde uma metilação nesse resíduo não apenas modifica a forma da proteína como essa capacidade doadora, inibindo, por exemplo, interações com outras proteínas (Luscombe, et al., 2001; Bedford et al., 2000). Em proteínas *Tau*, associadas com doenças neurodegenerativas, as regiões ricas em prolina vem sendo estudadas como um potencial sítio de ligação de tubulinas e microtúbulos (McKebben e Rhoades, 2019). Kursel e Malik (2017) descreveram na *Cid5* do grupo *repleta* uma região rica em prolina, sendo assim, futuros estudos relacionando as regiões da proteína ricas em determinados aminoácidos podem ajudar a elucidar a função e interação com outras proteínas, como a CenH3 na interação com o DNA centromérico e na interação com outras proteínas na formação do cinetócoro.

Na região NTT de *Cid1* e *Cid6* (**Figura 18**) é possível observar um *indel* de aproximadamente 42 nucleotídeos do cluster *D. mojavensis* em relação a *D. buzzatii*. Esta região, que compreende aproximadamente 16 aminoácidos, está presente em *Drosophila virilis* (grupo *virilis*) (Kursel e Malik, 2017) (**Figura 18C**), apontando que houve um evento de deleção dessa região no cluster *D. mojavensis*. Além disso, outros trabalhos já apontam que a diferença observada na NTT da CenH3 entre as espécies pode estar relacionada com a função biológica das proteínas, como por exemplo, reconhecer e recrutar outras proteínas para a formação do cinetócoro (Malik e Henikoff, 2001 ; Kursel e Malik, 2017).

Já a região HFD das espécies do cluster *D. mojavensis* e do complexo *D. buzzatii* apresenta uma maior uniformidade quando alinhada (**Figura 19**). Estudos iniciais com proteínas centroméricas já demonstravam a conservação do domínio HFD entre a histona centromérica entre as espécies e da histona centromérica com a histona canônica H3 de diferentes espécies (Malik e Henikoff, 2000, Kursel e Malik, 2017; Teixeira et al., 2018). A região HFD parece manter sua conservação devido sua importância na interação com outras proteínas centroméricas assim como na relação com o DNA centromérico. Um estudo recente demonstra que a região CATD (Cenp-A *Target Domain*), que ocupa maior parte da HFD (da α N até α -hélice 2 em humanos e da α N até a α -hélice 3 em *Drosophila*), seria essencial na interação da CenH3 com Cenp-C e Cal1 para a localização centromérica e formação do cinetócoro (Roure et al., 2019).

Usamos o método de SlidingWindow, implementado no DnaSP v.6.12.03 (Rozas et al., 2017), para analisar a variabilidade nucleotídica entre os dois parálogos Cid1/Cid6 e Cid5. Nos parálogos *Cid1*, *Cid6* e *Cid5* o domínio NTT é altamente variável, enquanto no domínio HFD os picos de maior variação nucleotídica correspondem com a região *loop1* (**Figura 20**). Esses dados estão de acordo com outros trabalhos que já relataram a alta variação da NTT em relação a HFD entre os parálogos de CenH3 e na região do Loop 1 (Malik e Henikoff, 2001; Kursel e Malik, 2017; Teixeira et al., 2018).

A



B

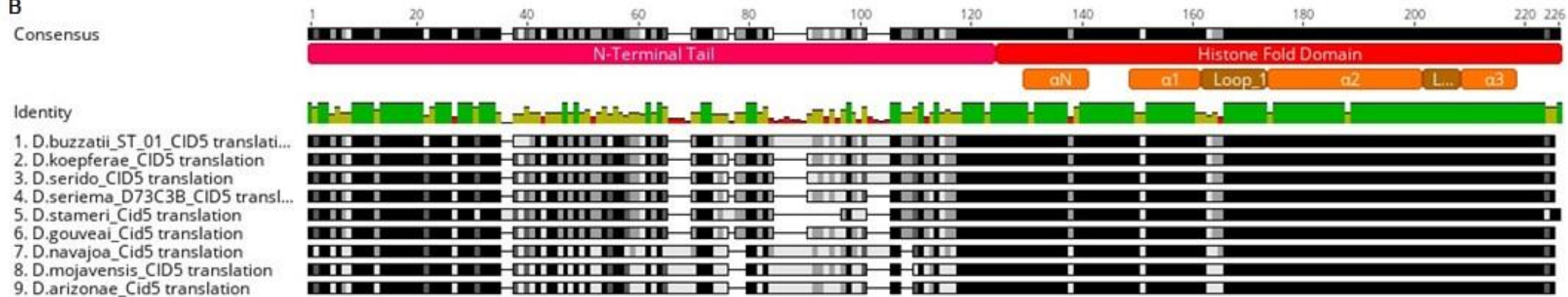


Figura 17. Alinhamento e caracterização dos parálogos *Cid1-Cid6* e *Cid5*. (A) *Cid1* e *Cid6* (B) *Cid5*.

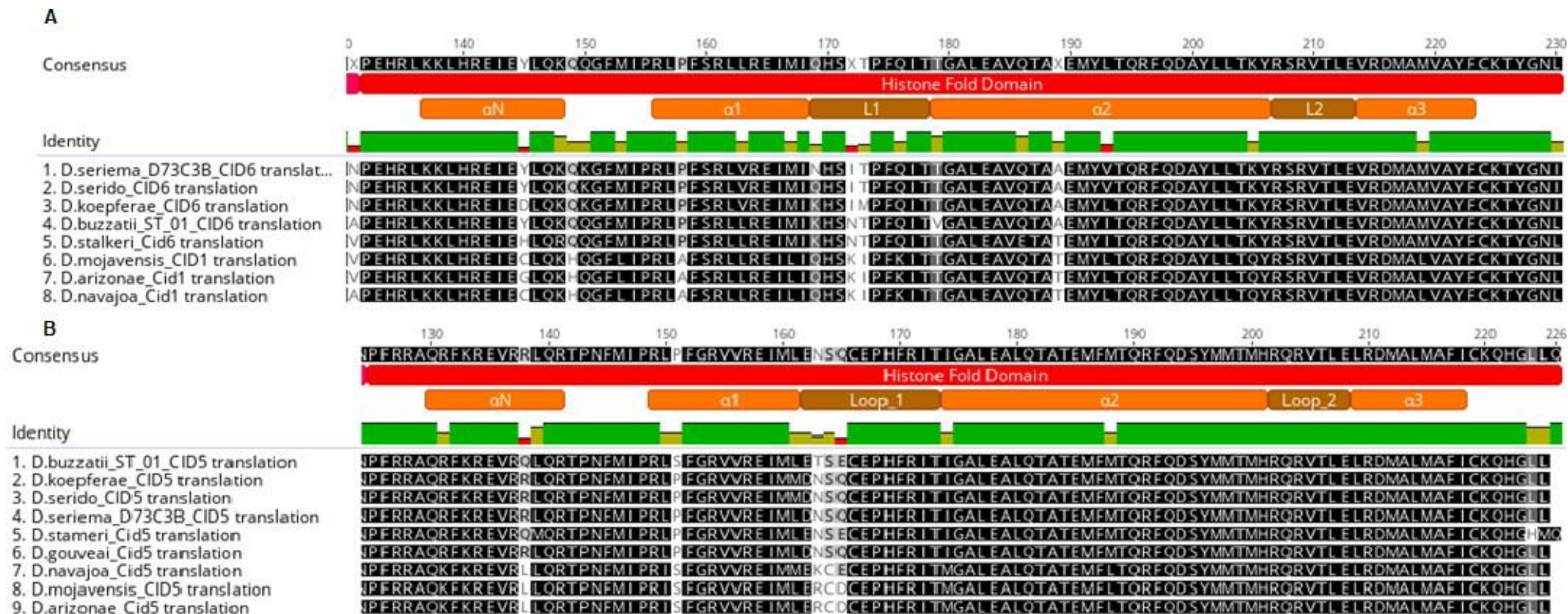


Figura 19. Alinhamento e caracterização do domínio HFD dos parálogos Cid1-Cid6 e Cid5. (A) Cid1 e Cid6 (B) Cid5.

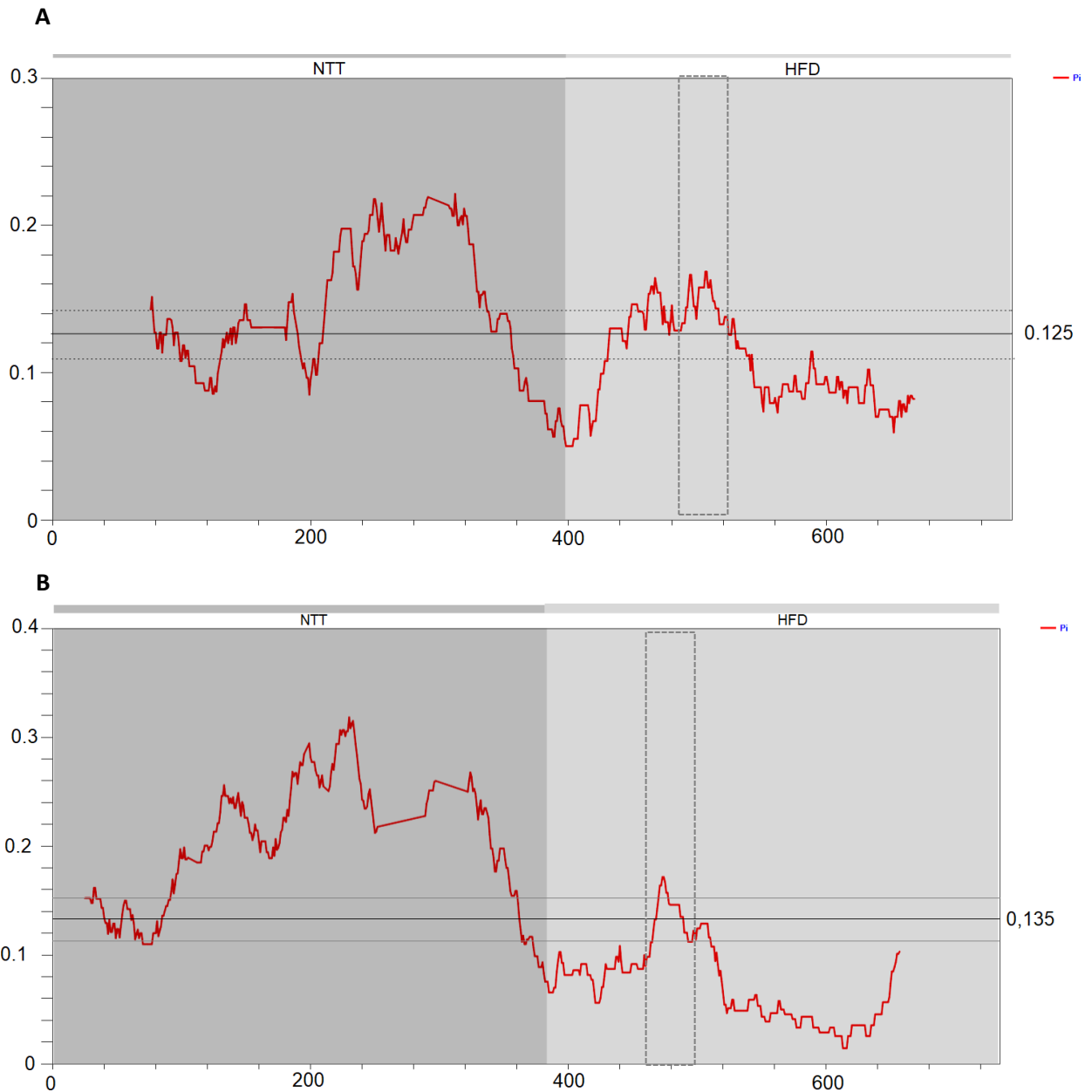


Figura 20. Variação nucleotídica. Sliding Window [Window length = 50nt – Step size = 1nt]. Área cinza escuro corresponde ao domínio NTT e área cinza claro corresponde ao domínio HFD. A área tracejada corresponde a região do *Loop1* Média da diversidade nucleotídica está representada pela linha vermelha enquanto a diversidade média \pm SD está representado pelas linhas tracejadas. **A.** *Cid1-Cid6* **B.** *Cid5*.

4.4 Inferência filogenética

Devido à alta divergência encontrada na NTT de *Cid*, árvores filogenéticas foram construídas com a região gênica correspondente a HFD usando os métodos de Máxima Verossimilhança (MV) e Inferência Bayesiana (IB). As árvores filogenéticas de MV apresentaram topologia semelhante à encontrada em trabalhos realizados com outros marcadores nas espécies desse estudo (**Figura 21**), onde as espécies do cluster *mojavensis* aparecem em um clado a parte do complexo *buzzatii* e mantendo *D. stalker* e *D. starmeri* como espécies próximas do cluster *D. buzzatii* (Ruiz e Wasserman, 1993; Franco et al., 2010; Kuhn e Sene, 2005; Ferreira, 2011). Já nas árvores geradas por IB (**Figura 22**) o resultado foi semelhante para as árvores MV, com exceção do posicionamento de *D. stalker* em relação as outras espécies do complexo *buzzatii*, colocando-a próximo de *D. koepferae*. Isso mostra a importância de se ter mais marcadores moleculares para definir relações filogenéticas, principalmente tratando-se de espécies próximas. Sendo assim, *Cid6* e *Cid5* mostraram-se como bons marcadores taxonômicos e filogenéticos para o grupo *repleta*, indicando que o gene *CenH3* também pode ser útil como marcador filogenético em outras espécies de eucariotos.

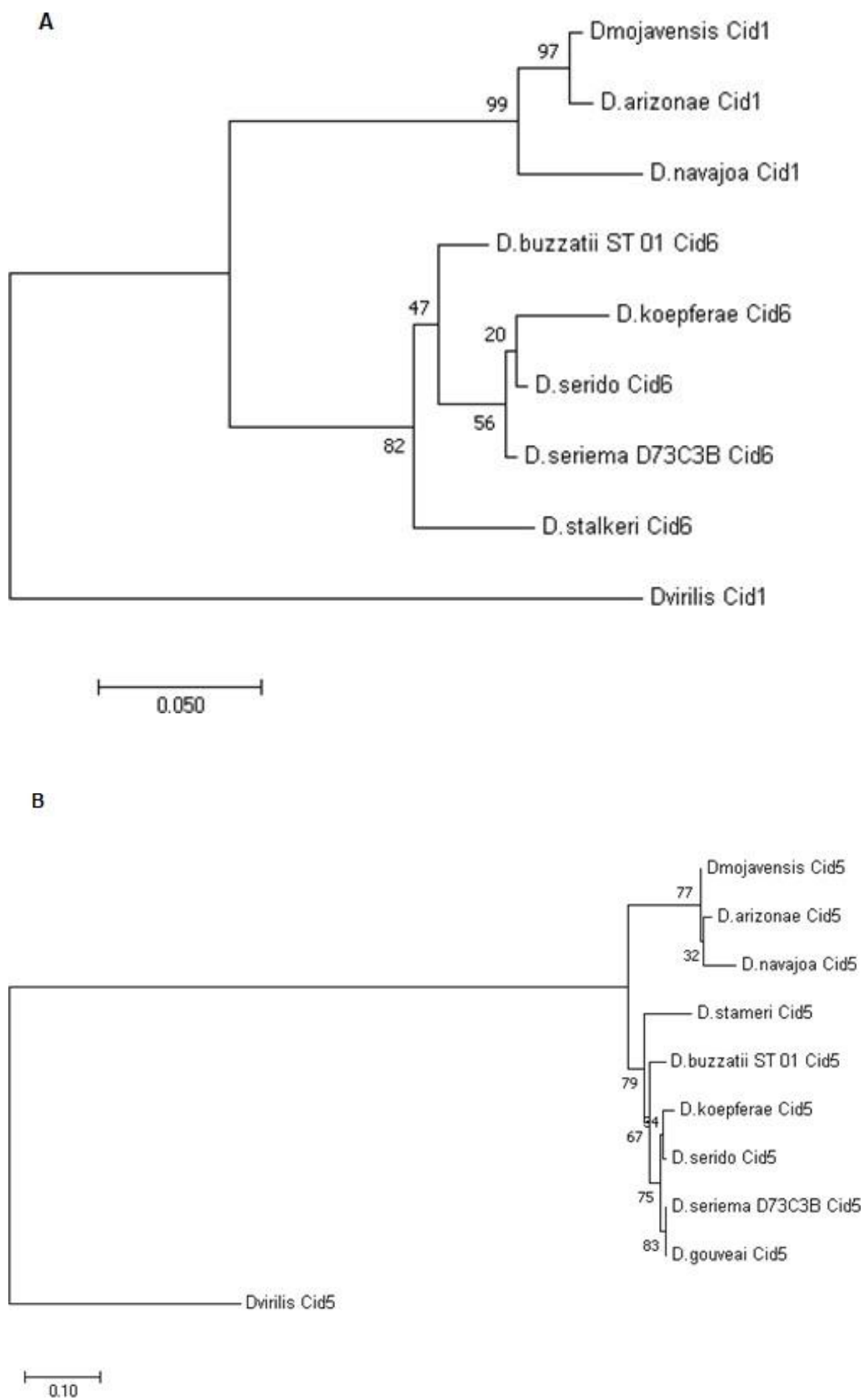


Figura 21. Árvore filogenética de Máxima Verossimilhança para os parálogos de *Cid* no grupo *repleta*. O suporte de *bootstrap* é mostrado em cada nó. A escala representa o número de substituições de nucleotídeo por sítio. **(A)** *Cid1-Cid6* **(B)** *Cid5*.

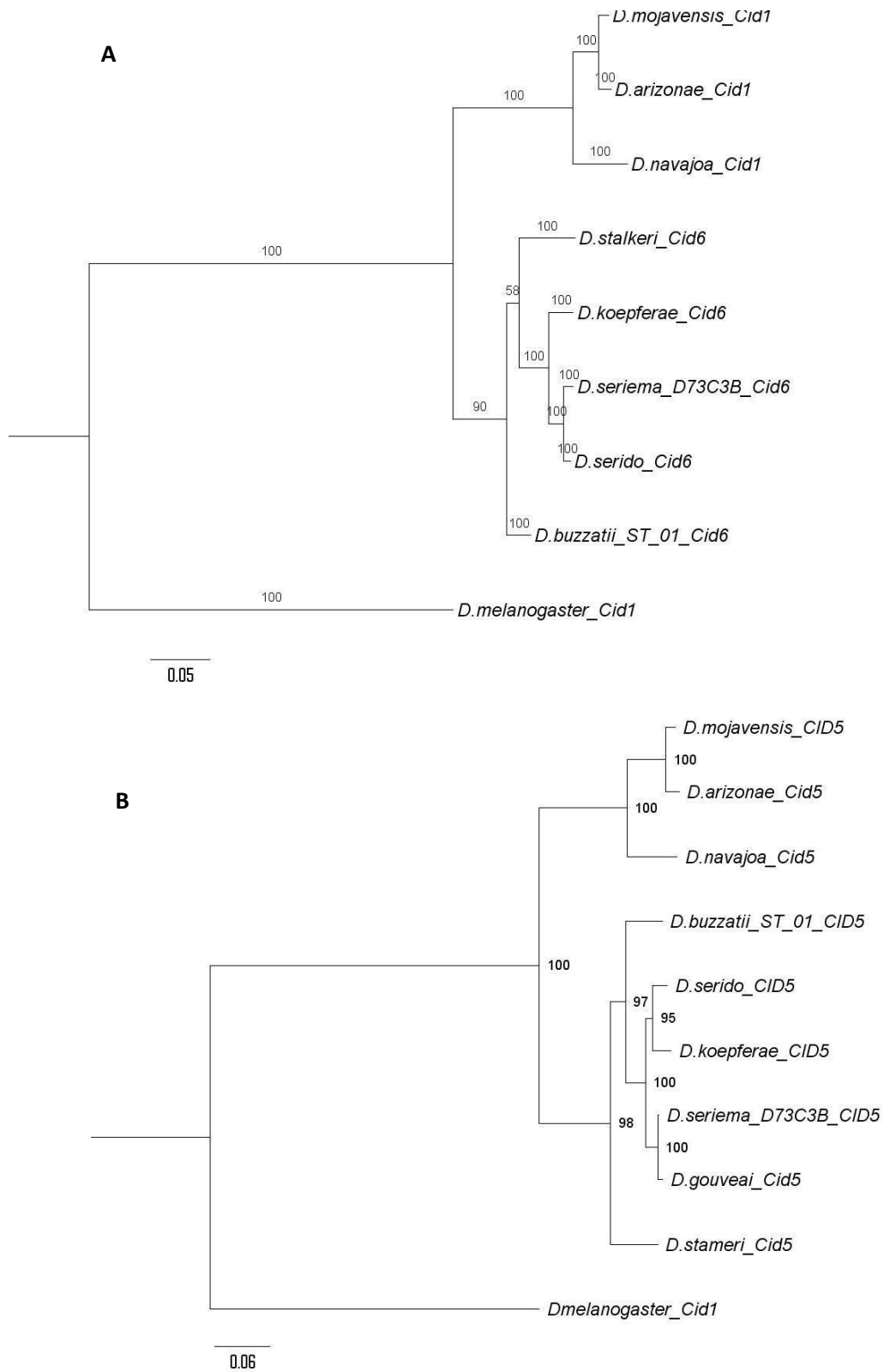


Figura 22. Árvore filogenética de com Inferência Bayesiana para os parálogos de *Cid* no grupo *repleta*. (A) *Cid1-Cid6* (B) *Cid5*.

4.5 Teste de seleção positiva nos parálogos de *Cid5* e *Cid6*

Para o teste MK (McDonald e Kreitman, 1991) foram utilizadas quatro populações de *D. buzzatii* (St-01, J-19, Bu28 e SC) e duas populações de *D. seriema* (D73C3 e SC). Árvores filogenéticas de MV contendo sequencias de *Cid6* e *Cid5* foram construídas para validação molecular do status taxonômico das populações amostradas (**Figura 23**).

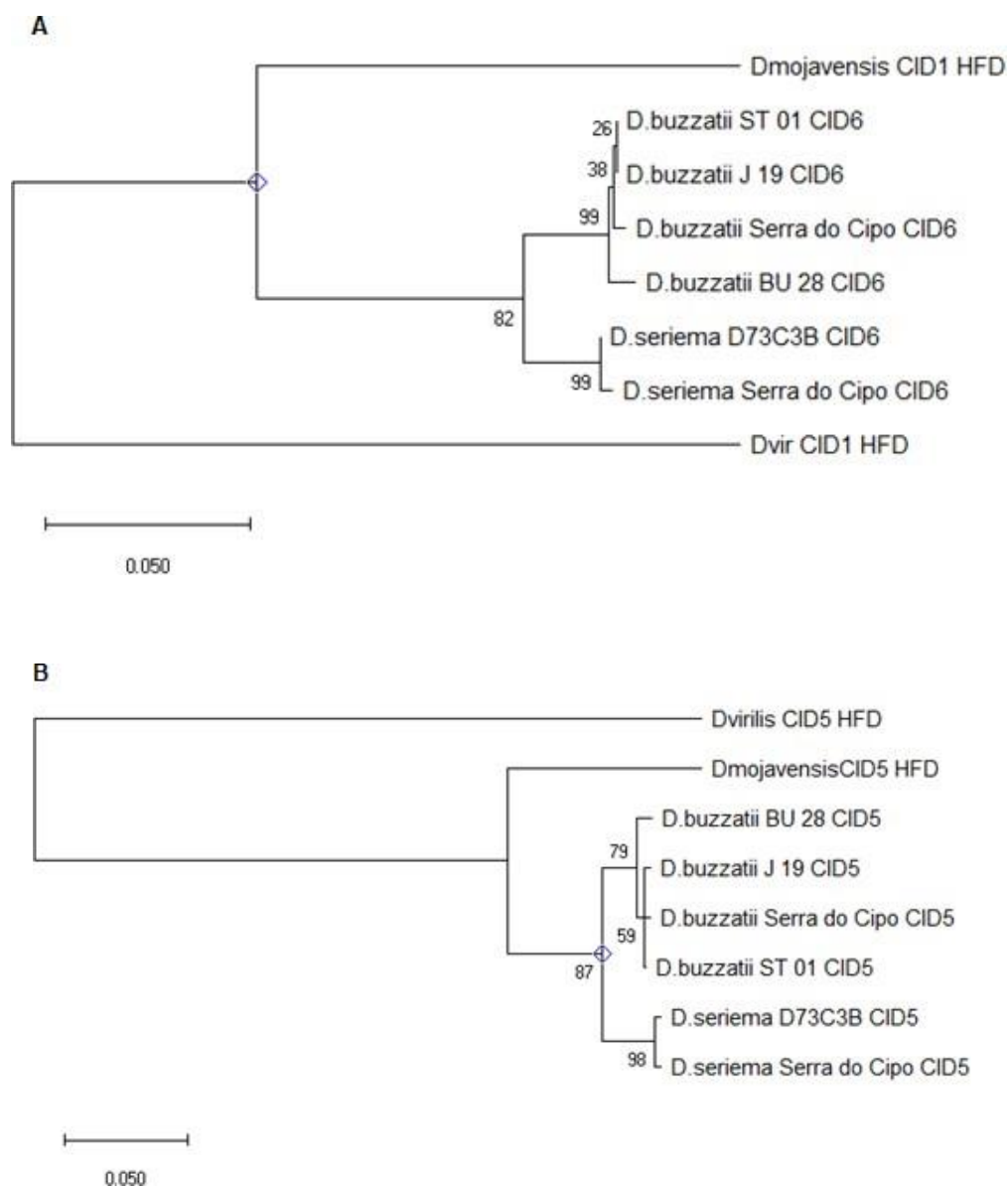


Figura 23. Árvores filogenética de Máxima Verossimilhança para os parálogos de *Cid* das populações de *D. buzzatii* e *D. seriema*. O suporte de *bootstrap* é mostrado em cada nó. A escala representa o número de substituições de nucleotídeo por sítio. **(A)** *Cid1-Cid6* **(B)** *Cid5*.

O resultado do teste MK não aponta seleção positiva para nenhuma das regiões analisadas (*full length*, NTT, HFD e Loop1), uma vez que a razão dN/dS é estatisticamente menor que a razão pN/pS e $p >$ que 0.05 (**Tabela 5**). Porém, o número amostral utilizado das populações de *D. buzzatii* e *D. seriema* pode ter mascarado um possível sítio sob seleção positiva, uma vez que uma substituição fixa pode acabar sendo considerada um polimorfismo se o número amostral aumentar significativamente (Graur e Li, 1999).

Tabela 5. Resultado do teste MK* com Neutrality index e p-value*.

Gene	Cluster	Região	#Codons	Nao Sin. fixa.	Sin. fixa	Nao Sin. Pol	Sin. Pol	Razão	Neutrality Index	p-value
<i>Cid6</i>	<i>buzzatii</i>	Full length	228	23	15	4	9	23:15:4:9	0.290	0.106
		NTT	129	15	12	4	6	15:12:4:6	0.533	0.475
		HFD	99	8	3	0	3	8:3:0:3	0.000	0.054
		Loop1	10	2	1	0	0	2:2:0:0	-	-
<i>Cid5</i>	<i>buzzatii</i>	Full length	208	22	15	7	12	22:15:7:12	0.398	0.158
		NTT	107	18	10	6	6	18:10:6:6	0.556	0.489
		HFD	101	4	5	1	6	4:5:1:6	0.208	0.307
		Loop1	12	0	3	0	2	0:3:0:2	0.000	0.100

*Estatisticamente significante quando $p < 0.05$

Para teste de seleção positiva envolvendo todas as espécies do grupo *repleta* presentes nesse estudo, foi utilizado o algoritmo CODEML implementado no EasyCodeML. O resultado aponta seleção positiva na região NTT de *Cid5*, no resíduo glutamina na posição 78 (78Q) (**Figura 24**), com probabilidade posterior $> 95\%$ (**Tabela 6**). Seleção positiva para a região NTT em *Cid5* foi descrita por Teixeira (2016) e em *Cid3* por Kursel e Malik (2017).

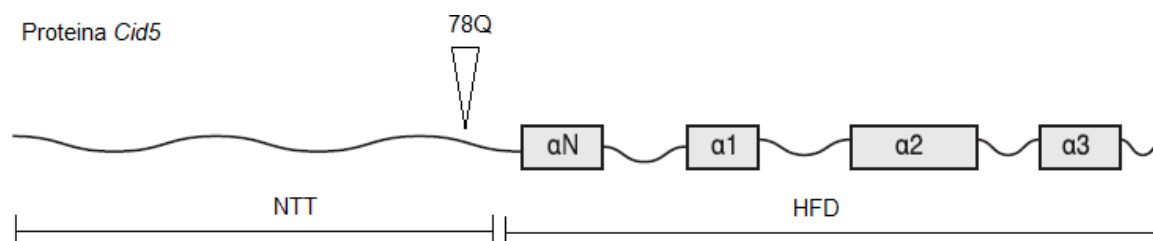


Figura 24. Esquema representando a proteína *Cid5* mostrando o sítio que está evoluindo sob seleção positiva identificado pela análise BEB – probabilidade posterior > 0.95 . (Adaptado de Kursel e Malik, 2017)

Sendo o gene *Cid5* expresso nos testículos de *Drosophila* (Kursel e Malik, 2017; Teixeira et al., 2018), este resultado de seleção positiva reforça a ideia de que este parálogo seria um candidato a gene supressor do impulso centromérico em células germinativas dos machos. De fato, experimentos recentes conduzidos por Kursel e Malik (2018), além de demonstrarem que a depleção de *Cid5* resulta em esterilidade no macho ou efeito letal de origem paterna, ainda sugerem que as diferenças presentes na NTT dos parálogos *Cid1* e *Cid5* podem estar relacionadas com interações proteicas distintas e especificidades celulares na formação do cinetócoro, sustentando a hipótese de função específica de *Cid5* como supressor do impulso centromérico nos machos.

Tabela 6. Parâmetros estimados para teste de seleção positiva usando o algoritmo CODEML

		EasyCodeML								Sitio positivo*
		- 2ΔI M1a vs M2a LRT p- value	- 2ΔI M7 vs M8** LRT p- value	- 2ΔI M8a vs M8 LRT p- value	Parâmetros estimados para M8					
	Gene				p0	p	q	p1	ω	
Grupo <i>repleta</i>	<i>Cid6</i>	1.00	0.6899	0.5228	0.89054	0.18863	0.72129	0.10946	1.48205	-
	<i>Cid5</i>	0.1517	0.1065	0.043	0.98769	0.19510	0.41259	0.01231	5.47813	78 Q (0.964)

* Probabilidade Posterior > 0.95 e $\omega > 1$.

** Graus de Liberdade =2 para comparações M7/8.

Kursel e Malik (2017) também detectaram seleção positiva na região *loop1* do parálogo *Cid4*, que Rosin e Mellone (2016, 2017) demonstraram estar envolvida na interação da CenH3 com a chaperona Cal1. Porém, nenhum resultado significativo de seleção positiva foi encontrado no presente trabalho para a região *loop1* de *Cid6* ou *Cid5*. Análises envolvendo mais espécies e populações podem ajudar a elucidar as pressões seletivas que ocorrem nesta região que apresentam maior diversidade nucleotídica.

Conclusões

Em relação aos nossos objetivos:

- **Determinar a origem filogenética da duplicação *Cid1-Cid6*, através da investigação destes parálogos em *D. stalker* (cluster *stalker*) e *D. starmeri* (cluster *martensis*), espécies do complexo *buzzatii* que constituem grupos externos ao cluster *buzzatii*.**

O presente estudo mostrou que o parálogo *Cid6* já se encontrava presente no ancestral comum do complexo *buzzatii*. Desta forma, não conseguimos determinar a origem filogenética precisa da duplicação *Cid1-Cid6*. Para este fim, novas espécies deverão ser estudadas, pertencentes à clados que se situam entre o cluster *mojavensis* e o complexo *buzzatii*.

- **Investigar se existem espécies no complexo *buzzatii* contendo ambos parálogos *Cid1* e *Cid6* funcionais (ao contrário da situação presente em *D. buzzatii* e *D. seriema*, onde *Cid1* se degenerou) e em caso afirmativo, caracterizá-los.**

Não conseguimos isolar *Cid1* de nenhuma espécie do complexo *buzzatii* estudada. Em contraste, encontramos *Cid6* em todas as espécies deste complexo. Este resultado significa que na espécie ancestral do complexo *buzzatii*, o gene *Cid1* já sofrera degeneração, tendo o parálogo *Cid6* retido a função original de *Cid1*.

- **Isolamento e caracterização de parálogos de *Cid* em espécies do grupo *repleta*.**

Caracterizamos os parálogos *Cid1*, *Cid6* e *Cid5* nas espécies do grupo *repleta*. Na análise das sequências de nucleotídeos, não encontramos nenhum indício de degeneração/pseudogenização, com exceção de *Cid1* no complexo *buzzatii*. Caracterizamos os domínios NTT e HFD (com suas regiões α N e as tres α -hélices separadas por dois *loops*) destes parálogos e encontramos uma maior variação nucleotídica em NTT em relação a HFD (com exceção da região Loop1), resultado que está de acordo com outros trabalhos realizados anteriormente.

Nossos resultados, portanto, indicam que nas espécies do grupo *repleta* existem dois parálogos de *Cid* atuando em função centromérica, *Cid1* com *Cid5* ou *Cid6* com *Cid5*. Resta, portanto, investigar em detalhes como estes parálogos atuam em função centromérica de machos e fêmeas (haja vista que *Cid5* possui expressão exclusiva em machos) e qual é o papel das regiões NTT e HFD neste processo.

- **Testar se existe seleção positiva atuando nos parálogos de *Cid1*, *Cid5* e *Cid 6* nas espécies do grupo *repleta*.**

Os resultados com CODEML indicaram que o parálogo *Cid5* no grupo *repleta* evoluiu sob seleção positiva, reforçando a ideia de que *Cid5* atua como supressor do impulso centromérico nos machos, uma hipótese previamente sugerida por Kursel e Malik (2017) e Teixeira e cols. (2018). Esta seleção positiva foi detectada especificamente em um resíduo de aminoácido da região NTT. Sabe-se que a região NTT de CenH3 está em contato intenso com o DNA da cromatina centromérica e seleção positiva nesta região também foi detectada em vários organismos (Finseth et al. 2015). Portanto, os dados sugerem que a região NTT pode desempenhar um papel importante na co-evolução de CenH3 com o DNA centromérico.

Adicionalmente, o presente trabalho mostrou que os parálogos de *Cid* corretamente inferiram as relações filogenéticas entre as espécies do grupo *repleta*, estabelecidas com o uso de outros marcadores genéticos. É possível que o gene CenH3 possa também ser um ótimo marcador filogenético em outras espécies de eucariotos.

Referência Bibliográfica

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PE, Hoskins RA, Galle RF, et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science*; 287: 2185–2195
- Alberts B, et al. *Biologia molecular da célula*. 6. ed. Porto Alegre: Artmed, 2017.
- Barrios-Leal DY, Neves-Da-Rocha J, Manfrin MH. (2019). Genetics and Distribution Modeling: The Demographic History of the Cactophilic *Drosophila buzzatii* Species Cluster in Open Areas of South America. *J Hered*;110(1), 22-33.
- Belford MT, Frankel A, Yaffe MB, Clarke S, Leder P, and Richard S. (2000). Arginine methylation inhibits the binding of proline-rich ligands to *Srchomology* 3, but not WW, domains. *J. Biol. Chem.* 275, 16030–16036.
- Belford MT e Clark SG. (2009). Protein Arginine Methylation in Mammals: Who, What, and Why. *Molecular Cell* 33, January 16, Elsevier Inc
- Bier E. (2005). *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet.* Jan;6(1):9-23.
- Black BE, Foltz DR, Chakravarthy S, Luger K, Woods Jr VL, Cleveland DW. (2004). Structural determinants for generating centromeric chromatin. *Nature* Vol 430 29 July.
- Black BE, Jansen LE, Maddox PS, Foltz DR, Desai AB, Shah JV, Cleveland DW (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell* 25:309–322
- Blower MD and Karpen GH. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* 3, 730-739.
- Blower MD, Sullivan BA, Karpen GH. (2002). Conserved organization of centromeric chromatin in flies and humans. *Dev Cell* 2:319 –330.
- Böhne A, et al., (2008). A Bayesian Model Comparison Approach to Inferring Positive Selection Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research*, v. 16, n. 1, p. 203-215.
- Buchwitz BJ, Ahmad K, Moore LL, Roth MB and Henikoff S. (1999). A histone-H3-like protein in *C. elegans*. *Nature* 401, 547-548.
- Carroll CW, Milks KJ, Straight AF. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. *The Journal of Cell Biology*, 189 (7) 1143-1155
- Caryl AP, Jones GH, Franklin CH. (2003). Dissecting plant meiosis using *Arabidopsis thaliana* mutants. *J. Exp. Bot.* 54:25-38.
- Charlesworth B, Sniegowski PE, Stephan W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*37, 215-220.
- Chmatal L, Gabriel SI, Mitsainas GP, Martínez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr Biol*, 24:2295-2300.

- Cheerambathur, D. K.; Desai, A. (2014). Linked in: formation and regulation of microtubule attachments during chromosome segregation. *Current Opinion in Cell Biology*, v. 26, p.113-22.
- Chen CC, Dechassa ML, Bettini E, Ledoux MB, Belisario C, Heun P, Luger K, and Mellone, BG. (2014). CAL1 is the *Drosophila* CENP-A assembly factor. *J. Cell Biol.* 204, 313–329.
- Clark A, Eisen M, Smith D. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218
- Clarke SG, and Tamanoi F, eds. (2006). Protein Methyltransferases. *The Enzymes*, Third Edition, Volume XXIV (San Diego, CA: Academic Press).
- Clément Y, Tavares R, Marais GA. (2006). Does lack of recombination enhance asymmetric evolution among duplicate genes? Insights from the *Drosophila melanogaster* genome. *Gene*. 385:89-95.
- Comeron JM, Ratnappan R, Bailin S. (2012). The Many Landscapes of Recombination in *Drosophila melanogaster* *PLoS Genet.* 8: e1002905.
- Cooper JL, Henikoff S. (2004). Adaptive Evolution of the Histone Fold Domain in Centromeric Histones. *Mol. Biol. Evol.* 21:1712-1718.
- Dawe RK, Henikoff S. (2006). Centromeres put epigenetics in the driver's seat. *Trends Biochem. Sci.* 31:662-669
- Dernburg AF. (2001). Here, there, and everywhere: kinetochore function on holocentric chromosomes. *J Cell Biol* 153: F33-F38.
- Dover GA. (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299:111-117.
- Dover GA. (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilate D. *Trends Genet.* 2:159-165.
- Dover GA. (2002). Molecular drive. *Trends Genet.* 18:587-589.
- Durando CM, Baker RH, Etges WJ, Heed WB, Wasserman M, DeSalle R, (2000). Phylogenetic Analysis of the repleta Species Group of the Genus *Drosophila* Using Multiple Sources of Characters. *Molecular Phylogenetics and Evolution* Vol. 16, No. 2, August, pp. 296–307
- Eaker S, Pyle A, Cobb J, Handel MA. (2001). Evidence for meiotic spindle checkpoint from analysis of spermatocytes from Robertsonian-chromosome heterozygous mice. *J. Cell. Sci.* 114:2953-2965.
- Earnshaw WC, Halligan N, Cooke C, e Rothfield N. (1984) The kinetochore is part of the metaphase chromosome scaffold. *J. Cell Biol.* 98, 352-357
- Earnshaw WC e Rothfield N. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91, 313-321
- Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

- Finseth FR, Dong Y, Saunders A, Fishman L. (2015). Duplication and Adaptive Evolution of a Key Centromeric Protein in *Mimulus*, a Genus with Female Meiotic Drive. *Mol. Biol. Evol.* 32:2694-2706.
- Fishman L, Saunders A. (2008). Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* 322:1559–1562.
- Fontdevila A. et al. 1988. *Drosophila koepferae*: A new member of the *Drosophila serido* (Diptera:Drosophilidae) superspecies taxon. *Ann. ent. Soc. Am.*, Maryland, 81:380-385
- Franco FF, Silva-Bernardi ECC, Sene FM, Hasson ER, Manfrin MH. (2010). Intra- and interspecific divergence in the nuclear sequences of the clock gene *period* of the *Drosophila buzzatii* cluster. *Journal of Zoological Systematics and Evolutionary Research.* 48:322-331.
- Franco FF, Manfrin MH. (2012). Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. *J. Biogeogr*
- Ferreira RF. (2011). Filogenia do complexo *Drosophila buzzatii* (grupo repleta): inferências de análises multilocus mitocondriais e nucleares
- Gallach M, Chandrasekaran C and Betran E. (2010). "Analyses of nuclearly encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic conflict in *Drosophila*. *Genome Biol Evol* 2: 835-850.
- Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9, 3891–3898.
- Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR, Presgraves DC. (2012). Genome sequencing reveal complex speciation in the *Drosophilasimulans* clade. *Genome Res.* 22:1499-1511.
- Graur D e Li WH. (1991). Neutral mutation hypothesis test. *Nature* 354:114-115.
- Guillén Y, Rius N, Delprat A, Williford A, Muyas F, *et al.* (2014). Genomics of ecological adaptation in cactophilic *Drosophila*. *Genome Biol. Evol.* 7:349-366.
- Griffiths AJ, Wessler SR, Lewontin RC, Carrol SB. (2016). *Introdução à Genética.* 11ª eD. Rio de Janeiro: Guanabara Koogan.
- Hadlaczky G, Praznovsky T, Rasko I e Kereso J. (1989). Centromere proteins. I. Mitosis specific centromere antigen recognized by anti-centromere auto antibodies. *Chromosoma* 97, 282-28
- Hasegawa M, Kishino H, e Yano T. (1985). Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution.* 22 (2): 160-174.
- Hasson E, De Panis D, Hurtado J, Mensch J. (2019). Host plant adaptation in cactophilic species of the *Drosophila buzzatii* cluster: fitness and transcriptomics. *J Hered.* 110(1), 46- 57.
- Heslop-Harrison JS, Schwarzacher T. (2013). Nucleosomes and centromeric DNA packaging. *Proc Natl Acad Sci U S A.* 2013 Dec 10;110(50):19974-5.
- Henikoff S, Ahmad K, Malik H. (2001). The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* 293:1098-1102

Henikoff S, Ahmad K, Platero JS, van Steensel B. (2000). Heterochromatic deposition of centromeric histone H3-like proteins. *Proc Natl Acad Sci U S A.* 97:716–721.

Henikoff S, Ahmad K, Platero JS e van Steensel B. (2000) Heterochromatic 1 deposition of centromeric histone H3-like proteins. **Proc. Natl. AcaD. Sci.** USA 97: 716–721

Henikoff S, Dalal Y. (2005). Centromeric chromatin: what makes it unique? *Curr Opin Genet Dev.* Apr;15(2):177-84.

Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH. (2006). Mislocalization of the *Drosophila* Centromere-Specific Histone CID Promotes Formation of Functional Ectopic Kinetochores. *Dev. Cell* 10:303-315

Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chma' tal L, Yang K, Sullivan BA, Schultz RM, Lampson MA, Black BE. (2017). Expanded satellite repeats amplify a discrete CENPA nucleosome assembly site on chromosomes that drive in female meiosis. *Curr Biol* 27:2365-2373 e2368.

Karpen GH, Allshire RC. (1997). The case for epigenetic effects on centromere identity and function. *Trends Genet* 13:489–496.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al., (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649

Kircher HW, (1982). Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. In: Barker, J.S.F., Starmer, W.T. (Eds.), *Ecological Genetics and Evolution. The Cactus–Yeast–Drosophila Model System*. Academic Press, Sydney, pp. 143–158.

Kuhn GCS, Sene FM. (2005). Evolutionary turnover of two pBuM satellite DNA subfamilies in the *Drosophila buzzatii* cluster (*repleta* group): from *alpha* to *alpha/beta* arrays. *Gene* 349: 77–85.

Kuhn GCS, Sene FM, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS. (2008). Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosome Res* 16:307-324.

Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. (2018). *Molecular Biology and Evolution* 35:1547-1549

Kursel LE, Malik HS. (2017). Recurrent gene duplication leads to diverse repertoires of centromeric histones in *Drosophila* species. *Mol Biol Evol*, 34:1445-1462.

Kursel LE, Malik HS. (2018). The cellular mechanisms and consequences of centromere drive. *Current Opinion in Cell Biology*, 52:58–65.

Larkin MA, Black G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, e et al., (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Natl Acad Sci USA.* 102:11793–11798.

- Lehninger A.L., Nelson D L, e Cox MM. (2000). *Lehninger principles of biochemistry*. New York: Worth Publishers
- Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* 29, 2860–2874
- Mahtani MM, Willard HF. (1998) Physical and genetic mapping of the human X chromosome centromere: repression of recombination. *Genome Res* 8:100–110
- Malik HS, Henikoff S. (2001). Adaptive Evolution of *Cid*, a Centromere-Specific Histone in *Drosophila*. *Genetics* 157:1293-1298.
- Malik H.S. e Henikoff S. (2002). Conflict begets complexity: the evolution of centromeres. *Curr. Opin. Genet. Dev.* 12, 711–718.
- Malik HS, Henikoff S. (2003). Phylogenomics of the nucleosome. *Nature Structural Biology* 10:882-891.
- Malik HS, Henikoff S. (2009). Major evolutionary transitions in centromere complexity. *Cell* 138:1067-1082.
- Manfrin MH e Sene FM. (2006). Cactophilic *Drosophila* in South America: a model for evolutionary studies. *Genetica*, 126, 57–75.
- Markow TA e O'Grady P. (2006) *Drosophila: a Guide to Species Identification and Use*. Academic Press (Elsevier), London.
- Marshall OJ, Chueh AC, Wong LH, and Choo KH. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am. J. Hum. Genet.* 82, 261–282.
- McDonald JH, Kreitman M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- Melters DP, Bradnam KR, Young HA, Telis N, May MR, et al., (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* 14: R10.
- Meluh PB, Yang P, Glowczewski L, Koshland D, Smith MM. (1998). Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell*;94:607–613.
- McKee BD, Wilhelm K, Merrill C, Ren X. (1998). Male sterility and meiotic drive associated with sex chromosome rearrangements in *Drosophila*. Role of X-Y pairing. *Genetics* 149:143- 155.
- McKebben K e Rhoades E. (2019). Tau's proline rich region dominates tubulin binding. *Biophysical Journal* 116, 3, 1, 157A-158A.
- NAGAKI K, KASHIHARA K, MURATA M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzulanivea*. *Plant Cell*, v. 17, n. 7, p. 1886-1893.
- Nambiar M, Smith GR. (2016). Repression of harmful meiotic recombination in centromeric regions. *Semin. Cell Dev. Biol.* 54:188-197.

- Nolte V, Pandey RV, Kofler R, Schlötterer C. (2013). Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. *Genome Res.* 23:99-110.
- O'Grady P.M, and DeSalle R. (2018). Phylogeny of the Genus *Drosophila*. *Genetics* 209: 1-25
- Oliveira DCSG, Almeida FC, O'Grady PM, Armella MA, DeSalle R. e Etges WJ. (2012). Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophilarepleta* species group. *Mol. Phylogenet. Evol.* 64: 533–544.
- Panchenko T, Sorensen TC, Woodcock, CL, Kan Z, Wood S, Resch MG, Luger K, Black BE, (2011). Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *PNAS* October 4, 108 (40) 16588-16593
- Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl AcaD. Sci. USA* 88:3734-3738.
- Patterson JT, and Wheeler MR. (1942). Description of new species of the subgenera *Hirtodrosophila* and *Drosophila* Austin: University of Texas Publication No. 4213. 1942.
- Patterson JT, and Crow JF. (1940). Hybridization in the *mulleri* group of *Drosophila* Austin: University of Texas Publication No. 4032.
- Powel J. (1997). Progress and prospects in Evolutionary Biology: The *Drosophila* model. New York: Oxford University Press. 562 p.
- Plohl M, Meštrović N, Mravinac B. (2014). Centromere identity from the DNA point of view. *Chromosoma* 123:313-325.
- Prüßing. K, Voigt. A, Schulz. JB. (2013). *Drosophila melanogaster* as a model organism for Alzheimer's disease. *Molecular Neurodegeneration* 8:35
- Rambaut, A. (2014). FigTree. Version 1.4.2. Inst. Evol. Biol.; Univ. Edinburgh.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, and Huelsenbeck JP. (2012). MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model selection across a large model space. *Syst. Biol.* 61:539-542. MEGA X. Molecular Evolutionary Genetics Analysis across computing platforms
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* 34: 3299-3302.
- Rosin L, Mellone BG. (2016). Co-evolving CENP-A and CAL1 Domains Mediate Centromeric CENP-A Deposition across *Drosophila* Species. *Dev. Cell* 37:136-147.
- Rosin LF, Mellone BG. (2017). Centromeres drive a hard bargain. *Trends Genet.* 33: 101-117.
- Roure V, Medina-Pritchard B, Lazou V, Rago L, Anselm E, Venegas D, Jeyaprakash AA, Heun P. (2019). Reconstituting *Drosophila* Centromere Identity in Human Cells. *Cell Reports* Volume 29, Issue 2, 8 October, Pages 464-479.e5
- Ruiz A, Wasserman M, and Heed WB. (1990). Evolution of the mojavensis cluster of cactophilic *Drosophila* with descriptions of two new species. – *J. Heredity* 81: 30–42.

- Ruiz A and Wasserman M. (1993). Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* 70: 582-596.
- Santaguida S, Amon A. (2015). Short- and long-term effects of chromosomemis-segregation and aneuploidy. *Nat Rev Cell Biol* 16:473– 485.
- Schittenhelm RB, Althoff F, Heidmann S, and Lehner CF. (2010). Detrimental incorporation of excess Cenp-A/Cid and Cenp-C into *Drosophila* centromeres is prevented by limiting amounts of the bridging factor Cal1. *J. Cell Sci.* 123, 3768–3779.
- Schu M, Ellenberg J. (2008). A new model for asymmetric spindle positioning in mouse oocytes. *Curr. Biol.* 18:1986-1992.
- Silva ASM. (2016). Estrutura centromérica e adaptações meióticas em espécies holocêntricas do gênero *Rhynchospora* (Cyperaceae). Tese (Doutorado em Biologia Vegetal – Sistemática e Evolução). Universidade Federal de Pernambuco. Pernambuco, p.144.
- Spicer GS. (1995). Phylogenetic utility of the mitochondrial cytochrome oxidase gene: molecular evolution of the *Drosophila buzzatii* species complex. *J. Mol. Evol.* 41(6): 749–759.
- Sullivan KF, Hechenberger M, Masri K.(1994) Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* 1994;127:581–592.
- Starmer WT, Barker JS, Phaff HJ, Fogleman JC. (1986). Adaptations of *Drosophila* and yeasts: their interactions with the volatile 2-propanol in the cactus-microorganism-*Drosophila* model system. *Aust. J. Biol. Sci.* 39:69-77.
- Sturtevant AH (2001) A History of Genetics. Cold Spring Harbor Laboratory Press Electronic Scholarly Publishing Project
- Takahashi Y, Mizoi J, Toh-E A, Kikuchi Y.(2000). Yeast Ulp1, an Smt3-specific protease, associates with nucleoporins. *J Biochem* 128(5):723-5
- Talbert PB, Henikoff S. (2010). Centromeres convert but don't cross. *PLoS Biol* 8:e1000326.
- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S. (2002). Centromeric Localization and Adaptive Evolution of an Arabidopsis Histone H3 Variant. *The Plant Cell* 14:1053-1066.
- Teixeira JR. (2016).Evolução do gene CID em espécies do subgênero *Drosophila* (Diptera: Drosophilidae). Trabalho de Conclusão de Curso (Bacharel em Ciências Biológicas). Universidade Federal de Minas Gerais. Minas Gerais. p.41.
- Teixeira JR, Dias GB, Svartman M, Ruiz A e Kuhn GCS. (2018). Concurrent Duplication of *Drosophila* Cid and Cenp-C Genes Resulted in Accelerated Evolution and Male Germline-Biased Expression of the New Copies. *Journal of Molecular Evolution.* Jul;86(6):353-364
- Thompson SL, Bakhoun SF, Compton DA. (2010). Mechanisms of chromosomal instability. *Curr Biol* 20:R285-95.
- Tidon-sklorz R e Sene FM. (1995). *Drosophila seriema* n. sp.: New Member of the *Drosophila serido* (Diptera: *Drosophilidae*) Superspecies Taxon. *Annals of the Entomological Society of America.* 88. 2. 139-142.

Tidon-sklorz R e Sene FM. (2001). *Drosophila antonietae* sp. n. and *Drosophila gouveai* sp. n.: Two new species of the *Drosophila serido* superspecies taxon (Diptera, Drosophilidae). Iheringia, serie Zoologia , 90, 141-146.

Tollis M, Boissinot S. (2012). The evolutionary dynamics of transposable elements in eukaryote genomes. In: Garrido-Ramos, M.A., editors. Repetitive DNA, Genome dynamics, 7th edn. Karger, Basel. pp 68-91.

Ugarković D (2009) Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol 48:53–76

Vilela CR. (1983). A revision of the *Drosophila repleta* species group (Diptera, Drosophilidae). Revista Brasileira Entomologica 27:1-114.

Vilela CR. e Sene FM. (1977). Two new neotropical species of the *repleta* group of the Genus *Drosophila* (Diptera, Drosophilidae). Revta bras. Ent., São Paulo, 30:295-299.

Wasserman M, Fontdevila A, and Ruiz A. (1983). Potential gene exchange between South American *Drosophila* species, with a description of a new species in the *D. repleta* group. Ann Entomol Soc Am, 76, 675–677.

Wasserman M, Koeper HR e Ward BL. (1973). Two New *repleta* Group Species of the Genus *Drosophila* (Diptera: Drosophilidae) from Venezuela. Annals of the Entomological Society of America 66(6):1239-1242

Wasserman M, Wilson FD. (1957). Further studies on the *repleta* group. University of Texas Publications 5721: 132-156.

Wheeler MR. (1954). A new genus and two new species of neotropical flies (Diptera; Drosophilidae). Entomol. News 21:207-213

Yang Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556

YANG Z. (2006). Computational molecular evolution. Oxford, Oxford University Press.

Yassin A. (2013). Phylogenetic classification of the *Drosophilidae* Rondani. Diptera.: the role of morphology in the postgenomic era. Syst. Entomol. 38:349-364.

Yoda K, Ando S, Morishita S, Houmura K, Hashimoto K, Takeyasu K, Okazaki T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc Natl Acad Sci U S A 97:7266–7271

Zou J, Hallen MA, Yankel CD, Endow SA. (2008). A microtubule-destabilizing kinesin motor regulates spindle length and anchoring in oocytes. J. Cell. Biol. 180:459-466.

Zhou Q, Bachtrog D. (2012). Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337:341-345.