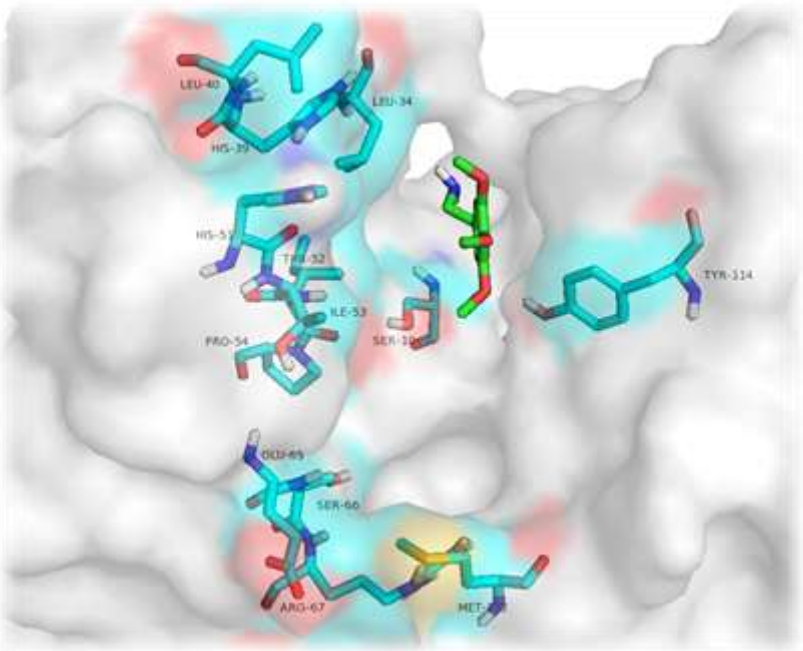
UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

Estrutura e função: evolução da proteína Glutationa S - Transferase D1 nas espécies do cluster *Drosophila buzzatii* (grupo *Drosophila repleta*)



Giuliatti, S. (2018)

ADRIANO SILVA DOS SANTOS

RIBEIRÃO PRETO
2018

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

Estrutura e função: evolução da proteína Glutationa S - Transferase D1 nas espécies do cluster *Drosophila buzzatii* (grupo *Drosophila repleta*)

ADRIANO SILVA DOS SANTOS

Dissertação apresentada ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto-USP, como parte do requisito para a obtenção do título de Mestre em Ciências.

Área de Concentração: Genética.

Orientador: Profª. Drª. Maura Helena Manfrin

RIBEIRÃO PRETO
2018

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Santos, Adriano Silva dos

Estrutura e função: evolução da proteína Glutationa S - Transferase D1 nas espécies do cluster *Drosophila buzzatii* (grupo *Drosophila repleta*). Ribeirão Preto, 2018.

78p. : 22 il. ; 30 cm

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Genética.

Orientador: Maura Helena Manfrin.

1. *Drosophila*. 2. Cactaceae. 3. proteína GSTD1. 4. sítio GSH. 5. desintoxicação. 6. adaptação.

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

Este trabalho foi realizado com o auxílio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

FOLHA DE APROVAÇÃO

Adriano Silva dos Santos

Título da Dissertação: Estrutura e função: evolução da proteína Glutationa S - Transferase D1 nas espécies do cluster *Drosophila buzzatii* (grupo *Drosophila repleta*).

Dissertação apresentada ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto-USP, como parte do requisito para a obtenção do título de Mestre em Ciências.

Área de Concentração: Genética.
Orientador: Prof[a]. Dr[a]. Maura Helena Manfrin.

Aprovado em:

Banca Examinadora:

Prof. Dr._____

Instituição: _____Assinatura_____

Prof. Dr._____

Instituição: _____Assinatura_____

Prof. Dr._____

Instituição: _____Assinatura_____

Prof. Dr._____

Instituição: _____Assinatura_____

Prof. Dr._____

Instituição: _____Assinatura_____

Dedico

Ao meu avô Sr. Joaquim (*in memoriam*) e a minha avó dona Maria.

À minha mãe Deusirene e ao meu pai Jorge.

Ao meu irmão André e ao meu primo Jhonatham.

AGRADECIMENTOS

À Deus por ter me fornecido capacidade, força, determinação e oportunidade para a realização do meu mestrado.

À Profª. Drª. Maura Helena Manfrin (FFCLRP-USP) pela amizade, confiança, ajuda em diversos momentos, pelas discussões sobre ciências, pelos ensinamentos, momentos descontraídos no laboratório, pelo apoio e orientação na vida acadêmica.

À Profª. Drª. Silvana Giuliatti (FMRP-USP) pela parceria na realização desse trabalho, pelos conhecimentos, ensinamentos, por me apresentar o "mundo das proteínas" e por tornar possível o estudo da proteína GSTD1, eu agradeço.

À Profª. Drª. Therese Markow pelos ensinamentos, aprendizados e financiamento de uma experiência profissional internacional e cultural, oferecida no congresso Evolution and Ecology: The cactophilic *Drosophila* model system, na cidade de Alamos, Sonora, México. Eu sou grato!

A Profª. Drª. Tania Hernandez (LANGEBIO) e ao Prof. Dr. Alfredo Ruiz (Barcelona-Spain) pelo compartilhamento de experiências profissionais, ensinamentos, discussões sobre ciência e motivação na carreira científica.

Ao Prof. Dr. Fábio de Melo Sene (FMRP-USP) pelos bons momentos de descontração, discussões e pelo incentivo na carreira acadêmica.

À Profª. Drª. Nilce Maria Martinez Rossi (FMRP-USP) e a Profª. Drª. Zilá Luz Paulino Simões (FFCLRP) por gentilmente disponibilizarem o uso de equipamentos em seus laboratórios para a realização de várias etapas práticas desse trabalho.

Ao Prof. Dr. Tiago Campos Pereira (FFCLRP-USP) pelas boas discussões e ideias sobre ciência.

Ao meu avô Sr. Joaquim (*in memoriam*) pelo apoio, confiança, carinho, amizade, pelo cuidado paternal, por acreditar no meu potencial, por me ajudar a realizar os meus sonhos, pelas histórias e valores, que levarei para toda a minha vida.

Á minha avó dona Maria por ter me ajudado a realizar esse sonho, pela confiança, amizade, pelo carinho, atenção e pelos ótimos cafezinhos durante os meus estudos na seleção do mestrado em genética.

À minha mãe Deusirene pelo incentivo, por acreditar nos meus sonhos, pelo apoio constante que foram de fundamental importância para o desenvolvimento desse trabalho.

Ao meu pai Jorge por depositar confiança na minha realização profissional e pelo carinho constante.

Ao meu irmão André e ao meu primo Jhonatham pelo apoio, ajuda e amizade.

Á minha amiga Dora Yovana Barrios Leal pela convivência agradável no laboratório, amizade, carinho, companheirismos, pela ajuda na realização das técnicas de Biologia Molecular, em diversas etapas desse trabalho, e na vida acadêmica.

Aos meus amigos do Laboratório de Genética Evolutiva (FFCLRP-USP): Ana B. Zichinelli, Taís C. Lavagnini-Pizzo, Eliete C. Delduca e Natácia E. de Lima, pela boa convivência, pelas discussões, companheirismos e ajuda fornecida no decorrer da realização desse trabalho.

Aos meus colegas: Pablo M. G. Rosa (LANGEBIO-CINVESTAV); Eunice P. Zaragoza (LANGEBIO-CINVESTAV); Alejandra Casteneda (LANGEBIO-CINVESTAV); Emily Durkin (University of Alberta - Canada); Nicolas Moreyra (University of Nuenos Aires - Argentina); Javier C. Ponce (University of Maryland-USA) pelas discussões sobre ciência, pelos bons momentos em Alamos (México) e vivência cultural.

Ao técnico Paulo Ricardo Epifânio (FMRP-USP), do laboratório de Genética Evolutiva, pela amizade, convivência e pelo bom trabalho realizado na manutenção dos estoques de *Drosophila*.

Aos meus colegas do Laboratório de Genética e Biologia Molecular de Fungos (FMRP-USP): Jéssica F. A. Silva e Tamires A. Bitencourt pela ajuda e discussão de métodos de Biologia Molecular.

Á minha amiga Graziella R. de Souza pela amizade e apoio.

À CAPES, pela concessão da bolsa de mestrado.

Ao Departamento de Genética (FMRP-USP) pela disponibilização de recursos e condições para a minha formação profissional.

Á secretária Susie A. P. Nalon do Departamento de Genética (FMRP-USP), pela amizade, carinho, atenção e pela ajuda em diversos momentos.

Às instituições da CAPES, FAPESP, CNPq e USP, pela disponibilização de verbas ao Laboratório de Genética Evolutiva.

Tudo é considerado impossível até acontecer.
Nelson Mandela

RESUMO

SANTOS, A.S. **Estrutura e função: evolução da proteína Glutationa S - Transferase D1 nas espécies do cluster *Drosophila buzzatii* (grupo *Drosophila repleta*).** 2018. 78f. Dissertação (Mestrado em Ciências). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

O cluster *Drosophila buzzatii* (grupo *D. repleta*) é constituído por sete espécies crípticas, com desenvolvimento larval ocorrendo exclusivamente em tecidos de cactos em decomposição. Os cactos são constituídos por diversos compostos químicos, que incluem substâncias tóxicas (alcaloides, esteroides e fenóis) para insetos. Nas espécies cactofílicas *Drosophila mojavensis* e *D. buzzatii*, o gene *GSTD1,* da família Glutationa S- transferases, é superexpresso em larvas submetidas a substratos tóxicos. *GSTD1* é indicado com função de desintoxicação, sendo importante alvo de seleção e necessário no processo de adaptação em espécies de *Drosophila*. No presente trabalho, foram isoladas regiões codificadoras completas do gene *GSTD1* das sete espécies do cluster *D. buzzatii* com o objetivo de entender a evolução nas sequências de DNA e nos fenótipos (proteínas), com análise da estrutura e função da proteína GSTD1, no processo de desintoxicação, entre as espécies do cluster. Foram sequenciados 630 pb do gene *GSTD1* para cada espécie e a estrutura primária da proteína, composta por 209 resíduos de aminoácidos, foi inferida "*In silico*". Nas inferências filogenéticas Bayesiana com o gene *GSTD1*, é indicado monofilietismo do cluster, com formação do primeiro aglomerado para *D. buzzatii* e *D. koepferae*, o segundo composto por *D. antonietae* e *D. serido* e, por último, um agrupamento entre *D. gouveai*, *D. borborema* e *D. seriema*. Sinais de seleção positiva nas sequências de DNA resultam em alterações nos resíduos de aminoácidos nas posições 39, 133, 136 e 209, com mudanças na estrutura química da proteína GSTD1, entre as espécies do cluster. Foram obtidos sete modelos da proteína GSTD1, realizada com modelagem por homologia, uma para cada espécie do cluster *D. buzzatii*. No *docking* molecular, entre as proteínas GSTD1 com o ligante mescalina, foi observado que as proteínas de *D. seriema*, *D. gouveai* e *D. koepferae* apresentaram melhores interações químicas com o ligante e menor gasto energético. O ligante mescalina foi posicionado entre duas regiões GSH (glutationa redutase) em *D. seriema*, *D. gouveai* e *D. koepferae*. A evolução do gene *GSTD1* entre as espécies do cluster *D. buzzatii* envolve sinais de seleção positiva em sequências de DNA com mudanças químicas em resíduos de aminoácidos no sítios-G e na estrutura interna da proteína GSTD1, bem como a capacidade do sítio de ligação GSH reconhecer o composto alcaloide.

Palavras-chave: *Drosophila*. Cactaceae. proteína GSTD1. sítio GSH. desintoxicação. adaptação.

ABSTRACT

SANTOS, A.S. **Structure and function: evolution of the protein Glutathione S - Transferase D1 in the species of the cluster *Drosophila buzzatii* (group *Drosophila repleta*).** 2018. 78f. Dissertação (Mestrado em Ciência). Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

The cluster *Drosophila buzzatii* (group *D. repleta*) is constituted by seven cryptic species, with larval development occurring exclusively in tissues of decomposing cacti. Cacti are composed of several chemical compounds, which include toxic substances (alkaloids, steroids and phenols) for insects. In the cactophilic species *Drosophila mojavensis* and *D. buzzatii*, the *GSTD1* gene, from the Glutathione S-transferases family, is overexpressed in larvae submitted to toxic substrates. *GSTD1* is indicated with detoxification function, being important target of selection and necessary in the adaptation process in *Drosophila* species. In the present work, complete coding regions of the *GSTD1* gene were isolated from the seven species of the *D. buzzatii* cluster with the objective of understanding the evolution in DNA sequences and phenotypes (proteins), with analysis of the structure and function of the GSTD1 protein in the process detoxification, among the species of the cluster. 630 bp of the *GSTD1* gene was sequenced for each species and the primary structure of the protein, composed of 209 amino acids, was inferred "In silico". In the Bayesian phylogenetic inferences with the *GSTD1* gene, monophyly of the cluster is indicated, with the formation of the first grouping between *D. buzzatii* and *D. koepferae*, the second composed by *D. antonietae* and *D. serido* and finally, a grouping between *D. gouveai*, *D. borborema* and *D. seriema*. Signs of positive selection on DNA sequences result in changes in amino acid residues at positions 39, 133, 136 and 209, with changes in the chemical structure of the GSTD1 protein among the species in the cluster. We obtained seven models of the *GSTD1* protein, performed with homology modeling, one for each species of the *D. buzzatii* cluster. In the molecular docking between the GSTD1 proteins with the mescaline ligand, it was observed that the proteins of *D. seriema*, *D. gouveai* and *D. koepferae* presented better chemical interactions with the ligand and lower energy expenditure. Interestingly, the mescaline linker was positioned between two GSH (glutathione reductase) regions, in *D. seriema*, *D. gouveai* and *D. koepferae*. The evolution of the *GSTD1* gene among *D. buzzatii* cluster species involves positive selection signals in DNA sequences with chemical changes in amino acid residues at the G-sites and in the internal structure of the GSTD1 protein, as well as the ability of the GSH binding site to recognize the alkaloid compound.

Key-words: *Drosophila*. Cactaceae. GSTD1 protein. GSH site. detoxification. adaptation.

SUMÁRIO

1. INTRODUÇÃO 13
1.1 Evolução em proteínas 13
1.2 Glutationa S-Transferases 15
1.3 Cluster *Drosophila buzzatii* 17
2. OBJETIVOS 22
3. MATERIAL E MÉTODOS 23
3.1 Amostras Biológicas 23
3.2 Protocolo Molecular e Análise das Sequências 24
3.3 Análise da região codificadora (CD) e resíduos de aminoácidos de GSTD1 25
3.4 Inferências filogenéticas para as espécies do cluster *Drosophila buzzatii* 26
3.5 Teste de Seleção 27
3.6 Modelagem e análise *in silico* da proteína GSTD1 29
4. RESULTADOS 31
4.1 Análise das sequências de DNA do gene *GSTD1* 31
4.2 Composição da região codificadora (CD) e resíduos de aminoácidos de GSTD1 31
4.3 Inferências filogenéticas para as espécies do cluster *Drosophila buzzatii* 37
4.4 Testes de Seleção 40
4.5 Modelagem e análise *in silico* da proteína GSTD1 44
4.6 *Docking* molecular da proteína GSTD1 com o ligante mescalina 46
4.7 Hidrofobicidade 51
4.8 Seleção Positiva e Alterações Estruturais Proteicas 53
5. DISCUSSÃO 55
5.1 Variações em sequências codificadoras e em resíduos de aminoácidos de GSTD1 56
5.2 Inferências filogenéticas no cluster *D. buzzatii* 59
5.3 Teste de seleção em sequências do gene *GSTD1* 62
5.4 Modelagem *in silico* e análise da proteína GSTD1 65
6. CONCLUSÕES 69
7. REFERÊNCIAS BIBLIOGRÁFICAS 70

1. INTRODUÇÃO

1.1 Evolução em proteínas

Em regiões gênicas codificadoras, as sequências de DNA contêm informações que quando consideradas em trincas, após processamento adequado, são traduzidas em resíduos de aminoácidos. As sequências de resíduos de aminoácidos são denominadas estrutura primária da proteína e possuem as informações necessárias para as ligações químicas entre os resíduos de aminoácidos e formação de dobras, estruturas de α-hélice e β-folha, durante a conformação da estrutura secundária. As estruturas de α-hélice e β-folha interagem por ligações entre os grupos de cadeias laterais dos resíduos de aminoácidos, com formação de dobras e torções até atingir o máximo de estabilidade com menor gasto energético, que resultam em uma proteína tridimensional, denominada proteína terciária. A proteína nessa conformação apresenta atividade biológica, com capacidade de interação com outras proteínas e com atividade de ligação a substratos químicos específicos (Hartl & Hayer-Hartl, 2009; Gupta et al. 2014; Ghosh et al. 2017).

Para a formação das estruturas terciárias é necessário a ação de proteínas denominadas chaperonas, que atuam estabilizando ao processo de tradução, com inibição do dobramento prematuro da proteína e garantindo a conformação terciária nativa (Hartl & Hayer-Hartl, 2009).

A sequência gênica, estrutura e função de proteínas são processos centrais para a compreensão de processos biológicos e celulares (Betts & Russell, 2003). A diversificação em proteínas ocorre por meio de modificações ou otimização em estruturas e/ou funções proteicas (Sali, 1995). Grande parte das proteínas apresentam diferenças superficiais no dobramento, mas possuem similaridade na composição de elementos estruturais que indicam eventos de duplicações gênica, com regiões homologas, no processo de evolução de proteínas e permitem agrupar as proteínas em famílias (Sali, 1995; Pál et al. 2006; Saisawang et al. 2012). Por outro lado, proteínas são alteradas por mutações de inserção, deleção e por substituições sinônimas e não sinônimas. As mudanças em sequências gênicas podem estar sobre processo de seleção de forma a modelar os melhores fenótipos (estruturas proteicas), nos processos de adaptação dos organismos, em caracteres de aptidão/mortalidade frente a um ambiente (Pál et al. 2006).

A evolução em proteínas envolve diversos fatores, tais como mutações na estabilidade proteica, cinética de dobramento, interações químicas, dinâmica funcional e modificações conformacionais. O mapeamento de mudanças no espaço sequência-estrutura de proteínas

possibilita avaliar a ação de mutações sítio-especificas, conflito adaptativo, epistasia e redes de interações intracelulares (Sikosek & Chan, 2014).

A taxa e o mecanismo de evolução em sequências proteicas tem recebido grande importância nos estudos de biologia evolutiva desde 1960. O estudo das taxas de evolução em proteínas resulta de conhecimentos obtidos pela evolução molecular e da genômica comparativa. A taxa de evolução em sequências proteicas (*k*) é estimada pelo número de substituições não sinônimas por sinônimas em sequências de DNA por sítio, entre par de proteínas ortólogas (*d*); dividido pelo dobro do tempo desde a divergência entre as espécies em análise (*t*). O valor de *t* é inferido por meio de registros fósseis ou estimados a partir de datações moleculares (Nei & Kumar, 2000; Zhang & Yang, 2015).

Outro método de avaliar a evolução proteica é estimada por meio da razão entre quantidade de substituições não sinônimas pelo número de substituições sinônimas em sequências de DNA (dN/dS), que indica alterações na composição de resíduos de aminoácidos da proteína. As substituições em resíduos de aminoácidos podem ocorrer por eventos recombinação gênica e mutações pontuais. O resultado obtido nesse método indica processos de seleção positiva, seleção purificadora ou seleção neutra, que podem estar atuando em uma sequência gênica e modelando o processo de evolução na proteína (Pál et al. 2006).

A quantidade de substituições em bases nucleotídicas em regiões gênicas pode indicar processos: (1) seleção positiva, quando ocorre maior número de substituições não-sinônimas do que sinônimas, sugerindo um reajuste em resíduos de aminoácidos de forma a melhorar a atividade da proteína e possivelmente favorecer a adaptação; (2) seleção purificadora, quando se observa maior número de substituições sinônimas do que não-sinônimas, indicando um reajuste com eliminação de substituições que resultam em alterações menos favoráveis para as atividades da proteína e para o processo de adaptação (Pál et al. 2006); (3) seleção neutra ocorre com número semelhante de substituições sinônimas e não sinônimas (McDonald & Kreitman, 1991; Pál et al. 2006), que sugerem eventos de deriva genética.

As modificações em resíduos de aminoácidos podem influenciar no processo de interação entre proteínas, função enzimática e no processo de adaptação de um organismo (Pál et al. 2006).

As modificações na composição gênica e no fenótipo (proteína) são provenientes de alterações em sequências de DNA que resultam em substituições de resíduos de aminoácidos da estrutura primária da proteína, que podem acarretar em alterações na geometria das estruturas terciárias, estabilidade termodinâmica, hidrofobicidade, peso molecular e função

enzimática (Lee et al. 2004; Pál et al. 2006; Zeldovich & Shakhnovich, 2008; Hartl & Hayer-Hartl, 2009; Ghosh et al. 2017).

Em geral, as variações em proteínas ocorrem de forma lenta em comparação com as alterações em sequência de DNA (Kumar & Singh, 2016). Isso se explica pela leitura do código genético, com múltiplas formas de trincas para um mesmo resíduo de aminoácido. Por outro lado, alterações em resíduos de aminoácidos em regiões do sítio ativo proteico podem resultar no aumento de sua eficiência funcional, com possível reflexo na aptidão dos indivíduos em uma população (Pál et al. 2006; Hartl & Hayer-Hartl, 2009).

1.2 Glutationa S-Transferases

Em *Drosophila melanogaster*, Glutationa S-Transferases ou GSTs é uma superfamília gênica, com 36 genes conhecidos e 41 proteínas na família (Saisawang et al. 2012). Nos genes *Gstd11*, *Gsto2*, *Gstt3* e *Gstz2* ocorrem *splicing* alternativo que aumenta a diversidade proteica na família (Saisawang et al. 2012). Por outro lado, o número de genes e de proteínas na família GSTs é diferente entre as espécies de *Drosophila*, como por exemplo: *D. erecta* apresenta 37 genes e 42 proteínas, *D. yakuba* apresenta 43 genes e 47 proteínas, *D. ananassae* apresenta 43 genes e 47 proteínas, *D. pseudoobscura* apresenta 31 genes e 36 proteínas, *D. willistoni* apresenta 40 genes e 43 proteínas e *D. mojavensis* apresenta 28 genes e 32 proteínas (Saisawang & Ketterman, 2014).

Os genes GSTs são divididos em seis classes: Delta (D), Epsilon (E), Omega (O), Theta (T), Zeta (Z) e Sigma (S). A classificação dos genes GSTs é determinada por análise da semelhança em sequências de resíduos de aminoácidos, posição genômica e por inferências filogenéticas. As sequências de aminoácidos da mesma classe apresentam grande estado de conservação na composição de resíduos de aminoácidos (Saisawang et al. 2012, 2014).

Os 36 genes GSTs em *D. melanogaster* apresentam localização citológica nos cromossomos 2R, 3 braços R e L e no cromossomo X. Os genes estão distribuídos em tandem ao longo dos cromossomos. Isso sugere que os genes de GSTs passaram por uma série de eventos de duplicação em tandem, que deram origem a formas parálogas, com expressão diferencial, diversidade de membros e diferenças funcionais (Saisawang et al. 2012).

Os genes da família GST são multifuncionais, com funções reconhecidas, como por exemplo, em desintoxicação, resposta a estresse oxidativo e em cascata de transdução de sinal (Low et al. 2007; revisão Huber et al. 2008; Saisawang et al. 2012).

Genes ortólogos de GSTs apresentam funções importantes no processo de desintoxicação e sobrevivência celular, conservadas em diversos organismos, como por exemplo, em bactérias (Allocati et al. 2009; Arivalagan et al. 2017), plantas (Dixon et al. 2002), peixes (Donnarumma et al. 1988), moscas (Morton, 1993, Matzkin, 2008; Low et al. 2007, 2010; Saisawang et al. 2012, 2014; Panis et al. 2016), mosquitos (Rason et al. 1997; King et al. 2017) e seres humanos (Vorojeikina et al. 2017).

Em meio celular, as enzimas da família GSTs são encontradas na forma de homo e heterodímero, com a presença de dois sítios ativos por dímero, que apresentam atividades enzimáticas independentes. O sítio ativo apresenta, no mínimo, duas regiões de ligações químicas, uma para ligação com glutationa (GSH), composto tripeptídico contendo resíduos de cisteína, glutamato e glicina, e outro sítio de ligação química, com menor especificidade para interação com eletrófilos (revisão de Huber et al. 2008).

De forma geral, os seres vivos estão continuamente em exposição a substâncias químicas, com capacidade de toxicidade variada. Com isso, o processo de desintoxicação é um processo necessário para a sobrevivência e metabolismo celular dos organismos. O processo de desintoxicação a produto xenobiótico é realizado em três fases. Na fase I e III ocorre a transformação dos produtos xenobiótico em um composto químico com maior solubilidade em água e menor toxicidade. A fase II possui extrema importância para o metabolismo e eliminação de produtos xenobióticos do meio celular, por ocorrer nessa etapa à conjugação de glutationa/xenobióticos. Na fase III ocorre o transporte dos metabólitos para regiões externas do ambiente celular (revisão de Huber et al. 2008).

O mecanismo de desintoxicação por glutationa tem sido sugerido como fundamental para adaptação biológica e sobrevivência dos organismos (revisão de Huber et. al. 2008).

Entre os genes da família GST, o gene *Glutationa S - Transferase D1* tem sido indicado como importante no processo de desintoxicação ao pesticida ambiental DDT (dicloro-difenil-tricloroetano) (Low et al. 2007). Além disso, tem sido muito discutido no processo de adaptação de espécies de *Drosophila* em diferentes ambientes nutricionais (Matzkin, 2008; Low et al. 2010; Panis et al. 2016; López-Olmos et al. 2017). Em experimentos com larvas de *Drosophila buzzatii* cultivadas em meio de cultura do cacto colunar *Trichocereus terscheckii*, considerado hospedeiro secundário para essa espécie (Soto et al. 2014; Panis et al. 2016), foram observadas diferenças no perfil de expressão gênica com evidências para genes de desintoxicação, como por exemplo, *alcohol dehydrogenase*, *cytochrome p450* e *Glutationa S - Transferase D1* (Panis et al. 2016). Essa diferença na

expressão gênica indicam mudanças genéticas que podem auxiliar na colonização de novos nichos ecológicos e favorecer a adaptação aos estresses impostos pelo uso de tecidos de cactos em decomposição, que contém compostos tóxicos (Panis et al. 2016).

1.3 Cluster *Drosophila buzzatii*

O cluster *Drosophila buzzatii* (grupo *Drosophila repleta*) é um grupo monofilético de espécies cactófilas, pois suas larvas na natureza se desenvolvem exclusivamente em tecidos de cactos em decomposição (Pereira et al. 1983; Ruiz & Wasserman, 1993; Manfrin et al. 2001). As sete espécies nominais que compõem o cluster *Drosophila buzzatii* são: *D. buzzatii*, *D. koepferae*, *D. antonietae*, *D. serido*, *D. gouveai*, *D. borborema* e *D. seriema*, identificadas taxonomicamente por características da genitália externa dos machos (Vilela, 1983) e naturalmente endêmicas da América do Sul (Manfrin & Sene, 2006). Dentre estas, apenas *D. buzzatii* é subcosmopolita, sendo encontrada em várias regiões do mundo, como Europa, África e Oceania (Fontdevila et al. 1981; Barker et al. 1985), sendo que essa colonização ocorreu nos últimos 200 anos, devido ao transporte de seu cacto hospedeiro da América do Sul para as demais regiões por ação antrópica (Barker et al. 1985). A distribuição geográfica atual das espécies do cluster *D. buzzatii* na América do Sul está mostrada na **Fig. 1**.

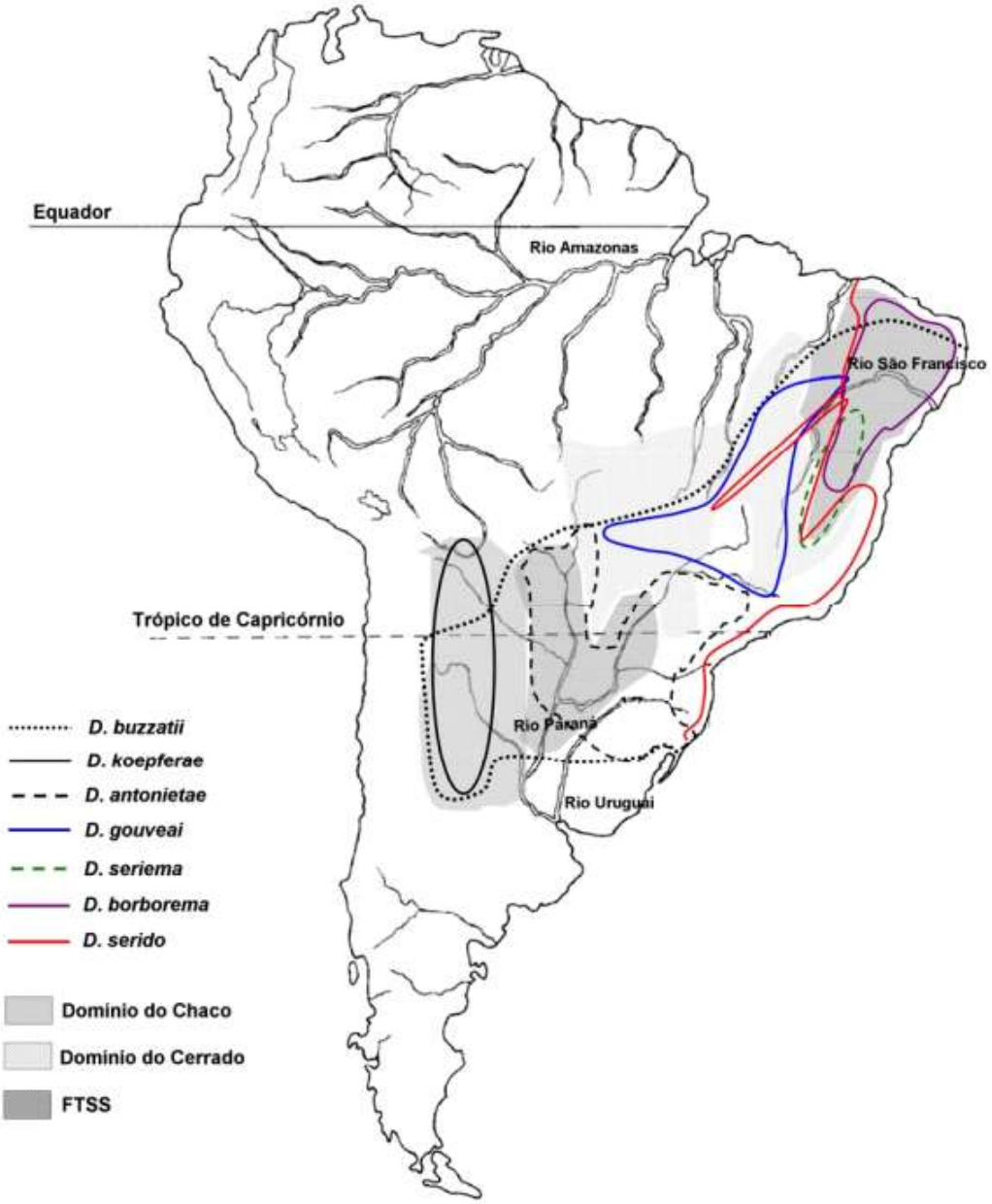


Fig. 1: Distribuição geográfica conhecida das espécies do cluster *Drosophila buzzatii* na América do Sul. As principais bacias hidrográficas, bem como a distribuição aproximada das principais áreas de florestas secas da América do Sul, estão apresentadas no mapa. FTSS: Florestas Tropicais Sazonalmente Secas. Mapa extraído de Franco (2009).

A morfologia do edeago é a principal característica diagnóstica para a identificação das espécies do cluster (Silva & Sene, 1991). Porém, muitas vezes as diferenças são de caráter quantitativo e detectado por meio de análises morfométricas do edeago (Tidon & Sene, 2001; Franco et al. 2006).

Hipóteses de relações filogenéticas entre as espécies do cluster *D. buzzatii* foram estabelecidas a partir de dados de inversões cromossômicas fixas (Ruiz & Wasserman, 1993), do gene mitocondrial c*itrocromo oxidase I* (Manfrin et al. 2001) e dos genes nucleares *α* -

esterase5 (Santos et al. 2009) e *period* (Franco et al. 2010). Embora as filogenias obtidas não sejam concordantes quanto às suas topologias, em todas foi inferida o monofiletismo do cluster *Drosophila buzzatii*. As relações internas, em geral, dividem o grupo em três linhagens evolutivas: a mais próxima do ancestral é representada pelas espécies *D. buzzatii* e *D. koepferae*, a segunda é representada pela espécie *D. antonietae*, e a terceira agrupa *D. serido* e uma politomia formada por *D. gouveai*, *D. borborema* e *D. seriema.*

Os padrões contrastantes, encontrados quando diferentes marcadores são utilizados para estudar as relações filogenéticas entre as espécies do cluster *D. buzzatii*, parecem ser uma consequência das dificuldades intrínsecas à reconstrução filogenética de espécies filogeneticamente próximas (Machado & Hey, 2003).

Em análises populacionais, foram sugeridos eventos de introgressão dos genes mitocondriais e nucleares entre as espécies (Manfrin et al. 2001; Franco et al. 2010; Kokudai et al. 2011). O tempo do ancestral comum do cluster foi inferido na transição Plio-Pleistoceno e pode estar relacionado com os ciclos glaciais e a dinâmica dos cactos (Oliveira et al. 2012).

A diversificação entre as espécies de *Drosophila* pode ser considerada uma radiação adaptativa (Manfrin et al. 2001; Low et al. 2007) e pode ter uma relação ao processo de adaptação ao uso de cactos palmares, com menor complexidade química e ao uso de cactos colunares, com maior complexidade química, que podem ser mais tóxicos (Oliveira et al. 2012).

Sabe-se que indivíduos adultos representantes das espécies do cluster *D. buzzatii* emergem de diferentes espécies de cactos, como: *D. koepferae* de cactos *Opuntia*, *Neocardenasia*, *Cereus* e *Trichocereus*; *D. buzzatii* em cladódios em decomposição de *Opuntia*, *Cereus*, *Pilosocereus* e *Trichocereus* (Fontdevila et al. 1981; revisão Manfrin & Sene, 2006); *D. gouveai* em cactos *Pilosocereus machrisii*, *Cereus* sp. e *Opuntia fícus-indica* (Moraes & Sene, 2003; Soto et al. 2010); *D. antonietae* em cladódios de *Cereus hildmaniannus* e *Opuntia monacantha* (Tidon-Sklorz & Sene, 2001); *D. serido* em cactos *Cereus hildmannianus*, *Cereus fernambucensis*, *Cephalocereus piauhyensis*, *O. monacantha*, *O. fícus-indica* e *Pilosocereus* sp. (Pereira et al. 1983; Tidon-Sklorz & Sene, 2001; De-Brito et al. 2002; revisão Manfrin & Sene, 2006); *D. seriema* em cladódios em decomposição de *Pilosocereus* sp.; e *D. borborema* em tecidos de *Pilosocereus piauhyensis* (revisão Manfrin & Sene, 2006). Isso demonstra a capacidade das espécies do cluster *D. buzzatii* em se adaptar ao uso de diferentes nichos ecológicos.

Por outro lado, os cactos são constituídos por diversos compostos químicos, que incluem substâncias tóxicas, como por exemplo, alcaloides, esteroides e fenóis (Davet, 2005; Brito-Filho, 2015), com graus diferentes de toxicidades para as espécies de *Drosophila*, conforme a capacidade de adaptação metabólica dos organismos (Kircher et al. 1967). Tem sido sugerido que a composição de alcaloides possui graus diferentes de toxicidade. Em experimentos com cultivo em laboratório de larvas de *Drosophila*, em meio contendo o alcaloide lofocereína do cacto senita, *Lophocereus schottii*, foram observados os fenótipos de morte e infertilidade de adultos (F1) em *Drosophila aldrichi*, *D. longicornis* e *D. nigrospiracula* (Kircher et al. 1967). Por outro lado, *D. mojavensis* e *D. pesudoobscura* apresentaram desenvolvimento da prole F1, mas não resulta em prole F2, indicando um processo de infertilidade ou inviabilidade no desenvolvimento larval (Kircher et al. 1967).

Variações em fenótipos de tempo de desenvolvimento, sobrevivência, fertilidade e fecundidade em *D. buzzatii* e *D. koepferae* têm sido relacionadas ao uso de tecidos em putrefação de espécies de cactos (Soto et al. 2014). *Drosophila antonietae* e *D. gouveai* apresentaram maior tamanho da asa ao utilizar extratos do cacto *Cereus hildmannianus*. Para *D. antonietae* e *D. gouveai*, as alterações no uso de cacto hospedeiro resultam em diferenças na viabilidade larval, na forma da asa e no tamanho do indivíduo adulto (Soto et al. 2010). *Drosophila gouveai* apresenta variações na estrutura do órgão reprodutor masculino (edeago), em resposta ao cultivo de indivíduos em meio com *P. machrisii* e *C. hildmannianus* (Soto et al. 2010). Em vista disso, o uso de diferentes extratos de cactos em decomposição é um fator que influência em caracteres fenotípicos importantes de viabilidade, desenvolvimento e reprodução em espécies de *Drosophila*.

Drosophila buzzatii e *D. koepferae* se desenvolvem em tecidos em decomposição dos cactos *Opuntia sulphurea* e *Trichocereus terschekii*, porém *D. koepferae* tem uma queda de viabilidade ao utilizar os recursos de *O. sulphurea* (Soto et al. 2014). Isso demonstra que extratos de cacto pode fornecer um ambiente estressante e diminuir a viabilidade em espécies cactofílicas de *Drosophila* (Padró et al. 2018).

Experimentos com cultivo em laboratório de larvas de *D. buzzatii*, em meio contendo substratos de *O. sulphurea* (hospedeiro primário) e em outro meio constituído por substratos de *T. terscheckii* (hospedeiro secundário), sugerem que o cacto *T. terscheckii* apresenta maior toxicidade para o desenvolvimento das larvas dessa espécie (Panis et al. 2016).

Por outro lado, as larvas de *D. buzzatii* cultivadas em *T. terscheckii* apresentaram mudanças no perfil de expressão gênica em comparação com o meio de *O. sulphurea*, com

maiores expressões gênicas evidenciadas para os genes *ADHs*, *GSTs*, *P450s* e *UGTs*, com funções reconhecidas em processos de oxidação-redução, desintoxicação, desenvolvimento e neurobiológicos (Panis et al. 2016). Com isso, nas larvas de *D. buzzatii* ocorreram reajustes das expressões gênicas em resposta a composição química do cacto *T. terscheckii*, que sugerem minimizar os efeitos nocivos dos constituintes tóxicos e possibilitar o uso de recursos oferecidos pelo cacto (Panis et al. 2016).

Nesse sentido, os dados apresentados por diversos autores indicam que as espécies de *Drosophila* cactofílicas respondem de forma diferente em termos de caracteres fenotípicos, desenvolvimento biológico e expressão gênica, com diferentes respostas adaptativas aos compostos tóxicos presentes em espécies de cactos (Kircher et al. 1967; Soto et al. 2010, 2014; Panis et al. 2016).

Em *D. mojavensis* (Matzkin, 2014) e em genes diferencialmente expressos de *D. buzzatii* durante o cultivo em meio com *T. terschekii* (Panis et al. 2016), tem sido mencionado o gene *Glutationa S - Transferase D1* como um dos importantes "genes alvos" necessários para a adaptação das espécies de *Drosophila* em diferentes condições alimentares de cactos com substâncias tóxicas. Além disso, o gene é mencionado no processo de desintoxicação ao pesticida DDT (dicloro-difenil-tricloroetano) (Low et al. 2007). Desta forma o gene *GSTD1* possui grande importância em termos biológicos, porém não se tem informações disponíveis sobre as sequências de DNA, os alvos de seleção positiva em sequências gênicas de *GSTD1*, composição de resíduos de aminoácidos, alterações bioquímicas e comportamento sobre a dinâmica da proteína com alcaloide de cactos, entre as espécies do cluster *D. buzzatii*, que permitam elucidar os processos evolutivos do grupo ao uso de diferentes hospedeiros/cactos.

Nesse sentido, as espécies do cluster *D. buzzatii* demonstram serem importantes modelos para avaliar os processos de adaptações metabólicas e evolutivas, na interação inseto-hospedeiro. Com isso, surgiram algumas questões relevantes sobre o processo de adaptação entre as espécies de *Drosophila* cactofílicas, como: as espécies do cluster *D. buzzatii* apresentam diferenças na composição nucleotídica ou em resíduo de aminoácido no gene *GSTD1*? Há sinais de seleção positiva na sequência do gene *GSTD1*? A proteína GSTD1 apresenta características estruturais e bioquímicas específicas para cada espécie? Hipótese avaliada: sinais de seleção, em nível filogenético de espécie, nas sequências gênicas de *GSTD1* resultam em modificações na estrutura ou na função proteica de GSTD1, com o aumento da eficiência no processo de desintoxicação aos substratos de cactos, entre as espécies do cluster *D. buzzatii*.

2. OBJETIVOS

O presente trabalho teve como objetivo geral avaliar a estrutura e atividade da proteína GSTD1, nas espécies cactofílicas do cluster *Drosophila buzzatii*.

Os objetivos específicos foram:

- Descrever a composição nucleotídica da região codificadora (CDs) do gene *GSTD1* nas espécies do cluster *D. buzzatii*;
- Descrever a composição em resíduos de aminoácidos da estrutura primária da proteína GSTD1;
- Inferir as relações filogenéticas entre as espécies do cluster *D. buzzatii* e outros representantes do grupo *D. repleta*, por meio de análises de sequências do gene nuclear *GSTD1*;
- Identificar os sinais de seleção e mapear a distribuição dos sítios nas sequências gênicas;
- Compreender as relações estruturais proteicas "*In silico*" de GSTD1 entre as espécies do cluster;
- Avaliar as potenciais atividades da proteína GSTD1 na interação com alcaloide de cacto.

3. MATERIAL E MÉTODOS

3.1 Amostras Biológicas

Para a realização deste trabalho, foi analisado um representante de cada espécie do cluster *Drosophila buzzatii* (Tabela 1). Os indivíduos foram coletados entre 2005-2016, pelo grupo de pesquisa, armazenados e preservados em freezer -80°C no Laboratório de Genética Evolutiva, da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, da Universidade de São Paulo. As identificações taxonômicas foram realizadas por meio de dados morfológicos da genitália masculina (edeago) (Vilela, 1983).

Tabela 1: Relação das espécies do cluster *D. buzzatii* e localidade dos indivíduos, utilizados para a obtenção das sequências do gene *Glutationa S- transferase D1*.

ID	Espécie	Localidade	CG
R41	*D. borborema*	Casa Nova – BA	09°14’15”S, 01°09’19.2”W
N58	*D. seriema*	Santana do Riacho - MG	19°17’25”S, 43°35’27.5”W
R71	*D. antonietae*	Sertãozinho – SP	21°09’S, 47°58’00”W
B26	*D. koepferae*	Famatina - La Rioja, Argentina	29°11’37”S, 67°23’10”W
R38	*D. serido*	Bom Jesus – RN	06°01’0.5”S, 35°38’52.7”W
N36	*D. gouveai*	Petrolina – PE	09°06’30.7”S, 40°38’49.3”W

ID = código de localidade no livro de campo, CG = coordenadas geográficas.

Foram extraídas as sequências nucleotídicas do gene *GSTD1* no GenBank para as espécies do grupo externo, que abrangeram os seguintes representantes: *Drosophila aldrichi* (acesso GenBank: MF372215.1), *Drosophila mojavensis* (acesso GenBank: EU079442.1), *Drosophila arizonae* (acesso GenBank: EU079471.1), *Drosophila navojoa* (acesso GenBank: EU079446.1), *Drosophila hexastigma* (acesso GenBank: MF372275.1), *Drosophila eremophila* (acesso GenBank: MF372253.1), *Drosophila anceps* (acesso GenBank: MF372377.1) e *Drosophila virilis* (acesso GenBank: XM_002054301.2). Dentre as espécies do cluster *D. buzzatii*, a sequência de *GSTD1* de *D. buzzatii* foi extraída do seu genoma sequenciado (Disponível em: <http://dbuz.uab.cat/welcome.php> código: lcl|scaffold305:1573-2228. Acesso: 20 ago 2017).

3.2 Protocolo Molecular e Análise das Sequências

O DNA genômico dos indivíduos adultos foi extraído com o uso de reagentes do *PureLink® Genomic DNA Purification Kit* (*Qiagen*), seguindo as instruções do fabricante. Para as amplificações do gene nuclear *GSTD1* (*Glutationa S- transferase D1*), foram realizadas reações em cadeia de polimerase (PCR), que continham entre 50 a 70 ng de DNA, 200 uM de dNTP, 0,4 uM de cada um dos *primers* (*forward* e *reverse*), 1X de tampão flexi (200 mM Tris-HCL pH 8,4; 500 mM KCl), 0,3 U de *Taq polimerase* (Promega ®) e água mili-Q para completar o volume final de 25 µl. A PCR foi realizada no *termociclador AB Applied Biosystems* (Veriti 96 well Thermal Cycler) e as condições de amplificação foram: 94° C por 1'30", 35 ciclos de 30" a 94°C, 30" a 46°C e 5' por 72°C. Para amplificar o gene *GSTD1*, foram desenhados os seguintes *primers* degenerados no programa IDT (disponíveis em: <https://www.idtdna.com>): 5'-MATGGYTGACTTCTAYTAYC-3' (*forward*), 5'-TTCTCGTACCACTTGTTCACGTTNKCR-3' (*reverse*), 5'-CCCGAGTTCGTGAAGATCAAYCCNCA-3' (*forward*) e 5'-GGCTYAWTCGAAGTAYTTCTTG-3' (*reverse*) (Fig. 2). Para o desenho dos *primers*, foram utilizadas como referência as sequências codificadoras completas do gene *GSTD1* de *D. mojavensis* (acesso GenBank: EU079442.1) e *D. arizonae* (acesso GenBank: EU079471.1).

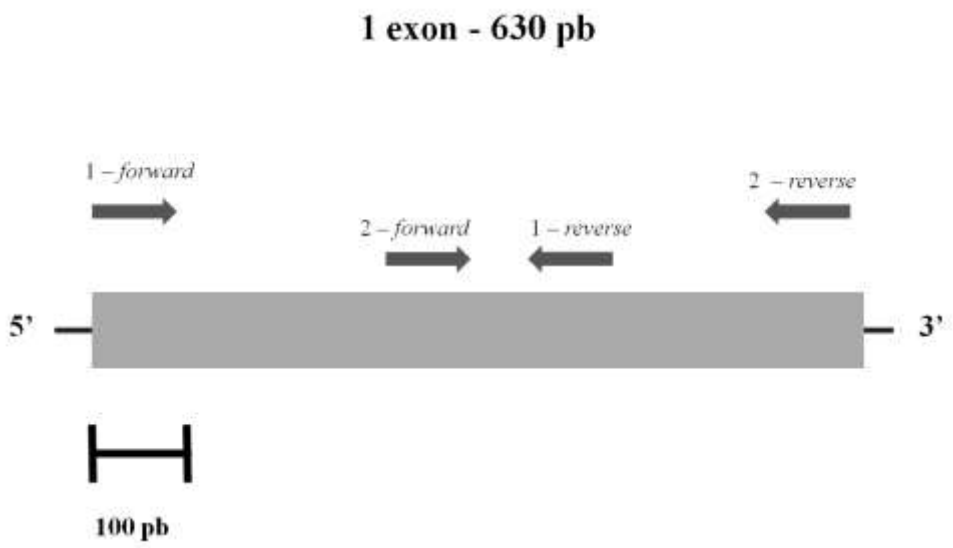


Fig. 2: Esquema da extensão total do gene *GSTD1* em *Drosophila mojavensis*, representado pelo retângulo cinza. Para a amplificação do gene *GSTD1* foram utilizados os conjuntos de *primers* 1 e 2 - *forward* e *reverse*, representados por setas de cor negra.

A qualidade dos produtos de PCR amplificados foi verificada por eletroforese em gel de agarose a 1%, corado com SYBR® safe (*DNA gel stain*). A visualização dos produtos de PCR foi realizada por exposição em luz UV em transiluminador e estimada o tamanho dos

fragmentos com o uso de marcador de peso molecular *Low DNA Mass Ladder* (©Invitrogen Corporation).

A purificação dos produtos de PCR foi realizada com o auxílio do conjunto de reagentes da *Illustra*™ (*Ge Healthcare*), seguindo as instruções do fabricante. A reação de sequenciamento foi realizada com o conjunto de reagentes *ABI PRISM BigDye*™ *Terminator Cycle Sequencing Ready Reaction*. Para a precipitação dos produtos de PCR, foram adicionados aos tubos 80 µl de isopropanol 65%. Os tubos foram incubados por 15 minutos no escuro, em temperatura ambiente, seguidos de centrifugação por 45 minutos a 4000 rpm. Posteriormente, o sequenciamento de *Sanger* automatizado de DNA foi realizado no Centro de Recursos Biológicos e Biologia Genômica - CREBIO, da FCAV-UNESP, do *campus* de Jaboticabal, no equipamento *ABI Prism*™ *377 sequencer* (Applied Biosystems, Foster City, CA, USA).

A qualidade das sequências foi avaliada por meio do programa Phred (Ewing & Green, 1998). As sequências *forward* e *reverse* foram comparadas, corrigidas e editadas com o auxílio dos programas Chromas Lite v.2.0 e BioEdit v.7.0. (Hall, 1999). No programa BLAST (*Basic Local Alignment Search Tool*) (Disponível em: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Acesso em: 25 ago 2017), foram avaliadas a similaridade das sequências de *GSTD1*, por meio dos critérios de *E-value* e *Ident* (Altschul et al. 1990; Benson et al. 2013). No programa BioEdit v.7.0, com o auxílio do algoritmo *ClustalW Multiple Alignment 1.8* (Thompson et al. 1994), foi realizado o alinhamento das sequências das espécies do cluster *D. buzzatii* e do grupo externo.

3.3 Análise da região codificadora (CD) e de resíduos de aminoácidos de GSTD1

No programa BioEdit v.7.0, foram realizadas as análises da composição dos conteúdos CG e AT das sequências codificadoras de *GSTD1*. Posteriormente, foram traduzidas em resíduos de aminoácidos no programa EMBOSS transeq (Disponível: <https://www.ebi.ac.uk/Tools/st/emboss_transeq/>. Acesso em: 12 set. 2017). No programa BioEdit v.7.0, foram executados os alinhamentos das sequências de resíduos de aminoácidos, entre as espécies do cluster *D. buzzatii* e do grupo externo. Foram realizados no programa BLASTP (Disponível: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Acesso em 14 set. 2017), buscas de similaridade entre as sequências de resíduos de aminoácidos e identificação dos domínios proteicos de GSTD1. No programa PIR (*Protein Information Resource*) (Disponível em: <https://pir.georgetown.edu/pirwww/search/comp_mw.shtml> Acesso em: 20 set. 2017)

(Baker et al. 2000), foram analisadas as proporções de resíduos de aminoácidos e o peso molecular mensurado pela massa em Dalton da proteína GSTD1. O programa PIR é associado aos bancos de dados do PIR-PSD, SWISS-PROT, TrEMBL, RefSeq, GenPept e PDB, que fornecem importantes dados sobre sequências de proteínas, como, por exemplo, as inferências de peso molecular de proteínas (Baker et al. 2000; Wu et al. 2003). Para essa análise, foram selecionadas as espécies do cluster *Drosophila buzzatii* e alguns representantes do grupo externo, a saber, *D. mojavensis*, *D. arizonae*, *D. navojoa* e *D. virilis*, com o objetivo de analisar variações na composição de GSTD1 entre as espécies.

3.4 Inferências filogenéticas para as espécies do cluster *D. buzzatii*

Foram realizadas inferências filogenéticas para as espécies do cluster *D. buzzatii*, usando a metodologia Bayesiana (BI), no programa MrBayes v.3.1.2 (Huelsenbeck, 2001, 2004). Essa análise, foi realizada para sequências do gene nuclear *GSTD1* com 630pb, gene mitocondrial *COI* com 549 pb e um agrupamento desses dois conjuntos de dados com 1179 pb. O objetivo dessas análises foi inferir as relações de parentesco e comparar os sinais filogenéticos entre o marcador sob seleção *GSTD1* e um "neutro" *COI*, para as espécies do cluster *D. buzzatii* e outros representantes do grupo *D. repleta*.

Inicialmente, no programa jModelTest v.2.1.2 (Posada, 1998), foram obtidos modelos de substituição nucleotídica que melhor explicam o processo de evolução nas sequências de *GSTD1* e *COI*, com o critério *Bayesian Information Criterion* (BIC) (Posada & Buckley, 2004; Luo et al. 2010).

Para isso, foram utilizadas as sequências do gene *GSTD1* obtidas no presente trabalho para as espécies do cluster *D. buzzatii* e para o grupo externo foram extraídas do GenBank (ver seção Amostras Biológicas). Para o gene *COI* foram extraídas sequências no GenBank, para as espécies do cluster: *D. buzzatii* (acesso GenBank: KX275224), *D. koepferae* (acesso GenBank: KP404433), *D. antonietae* (acesso GenBank: HM371179), *D. serido* (acesso GenBank: JN124578), *D. gouveai* (acesso GenBank: JN124677), *D. borborema* (acesso GenBank: KX275226) e *D. seriema* (acesso GenBank: JN124805). Além disso, foi extraído para espécies do grupo externo, a saber, os representantes foram *D. mojavensis* (acesso GenBank: DQ383719.1), *D. arizonae* (acesso GenBank: DQ383684.1), *D. navojoa* (acesso GenBank: DQ383666.1) e *D. virilis* (acesso GenBank: DQ471535.1). Por outro lado, foram retiradas das análises as espécies *D. aldrichi*, *D. hexastigma*, *D. eremophila* e *D. anceps*,

devido à ausência de dados sobre a extensão total das sequências nucleotídicas entre os marcadores *GSTD1* e *COI*.

As inferências filogenéticas Bayesianas foram realizadas com buscas de árvores em quatro corridas independentes, cada uma com quatro cadeias simultâneas, sendo uma "cold chain" e três "heated chains". As cadeias de Markov foram iniciadas com uma árvore aleatória, com o número de interações de 10.000.000 MCMC, amostrada a cada 1.000 gerações. Ao término da análise, com o número total de gerações, foram monitorados visualmente se os dados tinham atingido a convergência, por meio da verificação dos valores de ESS (*Effective Sample Size*) >200, no programa Tracer v.1.6 (Rambaut et al. 2018). Após o descarte de 10 % das árvores, correspondentes ao "burnin", uma árvore consenso foi calculada a partir das "cold chains" de cada uma das quatro corridas (Nascimento et al. 2017). Os valores da Probabilidade Posteriores (PP) foram calculados durante a análise de suporte dos ramos. As árvores filogenéticas consenso foram visualizadas e editadas no programa FigTree (Disponível em: <http://tree.bio.ed.ac.uk/software/figtree>. Acesso em: 20 mai 2018).

3.5 Teste de Seleção

Foi realizado o teste de seleção nas sequências de *GSTD1*, com os parâmetros das estimativas da razão dN/dS (=ω, dN representa o número de substituições nucleotídicas não sinônimas e dS o número de substituições sinônimas), para estimar distintos códons de resíduos de aminoácidos sob diferentes pressões de seleção (Nielsen & Yang et al. 1998; Yang et al. 2000; Doron-Faigenboim et al. 2005). O teste de seleção foi realizado com as sequências completas da região codificadora de *GSTD1*, que abrangeram os seguintes representantes: *D. buzzatii*, *D. koepferae*, *D. antonietae*, *D. serido*, *D. gouveai*, *D. borborema*, *D. seriema*, *D. mojavensis*, *D. arizonae*, *D. navojoa* e *D. virilis*. Para cada espécie de *Drosophila*, as taxas de dN e dS foram comparadas com o método de Nei-Gojobori (Nei & Gojobori, 1986). As diferentes taxas foram comparadas e testadas com o teste Z-Selection, realizado no programa MEGA 6 (Tamura et al. 2013), com nível de confiabilidade de 95%, em que o valor de *p* é a probabilidade de rejeitar a hipótese de neutralidade (H0: dN=dS) (Nei & Kumar, 2000). Os testes de seleção foram realizados no servidor SELECTON (disponível em: <http://selecton.tau.ac.il/>), com investigação dos sinais em códons de resíduos de aminoácidos específicos calculados com a razão dN/dS (ω), utilizando a metodologia Bayesiana (Doron-Faigenboim et al. 2005; Stern et al. 2007). Para isso, no programa MEGA

v. 6.0 (Tamura et al. 2013), foi obtido uma árvore filogenética por máxima verossimilhança (ML) para o gene *GSTD1*, que foi utilizada na análise de seleção em sítios específicos. Foram avaliados quatro modelos de seleção, estimados por máxima verossimilhança, no conjunto de dados: M8 (beta, $\omega \geq 1$), permite a seleção positiva operando em códons da proteína, com as proporções de p1 (1-p0) e valores de $\omega S \geq 1$ (Yang et al. 2000); M8a (beta, $\omega S=1$), permite detectar sinais de seleção purificadora ou neutras em códons da proteína, não permitindo sinais de seleção positiva, com as proporções de p1 (1-p0) (Swanson et al. 2003); M7 (beta) é semelhante ao modelo M8a, pela distribuição beta, mas com adição dos parâmetros de p e q, com ω estimado em 1 (Yang et al. 2000). Os valores gerados nos modelos são demonstrados por $\omega<1$, $\omega=1$ e $\omega>1$, que indicam processos de seleção purificadora, neutra e positiva, respectivamente.

Outro modelo testado foi o MEC (Doron-Faigenboim & Pupko, 2007), que permite avaliar uma maior probabilidade de matriz de substituição com 61 códons de senso. As taxas de transição (tr) e transversão (tv) são parâmetros integrados na matriz. Em MEC são considerados diferentes forças de seleção e intensidades, com as substituições não sinônimas multiplicadas por ω. Os valores gerados para ω indicam a intensidade de seleção (purificadora ou positiva). Os escores gerados por *Akaike Information Content* (AICc) para o modelo de MEC serviram para comparações com M8a.

Os nucleotídeos com sinais de seleção positiva foram mapeados na estrutura da proteína GSTD1, com buscas em regiões do sítio ativo e na superfície de interação proteína-proteína (Doron-Faigenboim et al. 2005). Essa análise foi realizada no servidor SELECTON, com o objetivo de identificar variantes em resíduos aminoácidos que melhoram a função proteica de GSTD1 no metabolismo de substratos tóxicos, como por exemplo, os alcaloides.

Foi avaliado a significância estatística dos resultados obtidos para dN/dS em cada modelo de seleção, por meio do *Likelihood Ratio Test* (LRT), calculado manualmente com a fórmula $2\Delta l=l_1-l_0$ (l_1 é o *likelihood* do modelo 1, l_0 é o *likelihood* do modelo 0 - hipótese nula e $2\Delta l$ significa a diferença entre os *likelihoods* dos modelos em comparação). Foram realizadas comparações entre os pares de modelos M8 *versus* M8a, M8 *versus* M7 e MEC *versus* M8a. Os resultados da significância de LRT entre os modelos M8 *versus* M8a foram avaliados com testes de X^2, assumido 1 grau de liberdade (um parâmetro adicional - p0), com valores críticos de qui-quadrado de 3,84 a 5%, 6,64 a 1%, 10,83 a 0,1%. Os modelos M8 *versus* M7 e MEC *versus* M8a. foram avaliados, assumindo 2 graus de liberdade (dois parâmetros adicionais - p0 e ωS), com os valores de 5,99 a 5%, 9,21 a 1%, 13,82 a 0,1%

(Yang, 2007; Yang & Reis, 2011). Os valores estatísticos serviram para avaliar a significância das variações em códons de resíduos de aminoácidos da proteína GSTD1, que foram evidenciados nos modelos de seleção.

3.6 Modelagem e análise *in silico* da proteína GSTD1

As análises "*in silico"* da proteína GSTD1 foram coordenadas pela profª. Drª. Silvana Giuliatti, responsável pelo Laboratório do Grupo de Bioinformática (GBI), do Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, USP. Inicialmente, foi obtido no *Protein Data Bank* (PDB), o modelo da proteína de GSTD1 de *Drosophila melanogaster* (acesso PDB: 3MAK), que foi usada para a modelagem "*in silico*" da proteína GSTD1 de *D. mojavensis* (acesso GenBank: ABW76256.1), realizada por meio do método de homologia comparativa, no programa Modeller 9 v.7 (Sali, 1995). O modelo da proteína GSTD1 de *D. mojavensis* serviu de referência para a modelagem proteica nas espécies do cluster *D. buzzatii*. Foram obtidos sete modelos da estrutura proteica de GSTD1, uma para cada espécie do cluster *D. buzzatii*, que passaram por validação no programa *Ramachandran Plot Analysis* (Disponível: <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>. Acesso em: 28 ago. 2017) (Lovell et al. 2003). Essa análise foi realizada com o objetivo de gerar e conhecer os modelos estruturais da proteína GSTD1 para todas as espécies do cluster.

No programa Chimera (Pettersen et al. 2004), foram realizados alinhamentos múltiplos das estruturas proteicas terciárias de GSTD1 entre as espécies em estudo, com o objetivo de visualizar modificações estruturais. Para essa análise, foram consideradas as modelagens obtidas para as espécies do cluster *D. buzzatii* e para o grupo externo representado por *D. mojavensis*, *D. navojoa* e *D. arizonae*.

Sabe-se que a proteína GSTD1 é constituída por dois sítios ativos, os sítios-G e H (Low et al. 2010; Lópes-Olmos et al. 2017). Para o sítio-G, os resíduos de aminoácidos nas posições 51-53 e 65-67 constituem regiões de ligação com glutationa, reconhecida como GSH (glutationa redutase), com função de aumentar a eficiência da proteína na conjugação com subtratos químicos durante a desintoxicação metabólica (Huber et al. 2008). O resíduo serina 10 é indicado como um estabilizador de ânion tiolato do sítio-G (Low et al. 2010). Nesse sentido, no programa MetaPocket, foram estimados os resíduos de aminoácidos que compõem o sítio–G, com o objetivo de evidenciar variações no sítio-G das proteínas de GSTD1 para cada espécie do cluster *D. buzzatii*. Para essa análise, foi utilizado como referência, um mapa de resíduos de aminoácidos do sítio-G da proteína GSTD1 de *D. melanogaster* (Low et al.

2010). Paralelamente, no programa PyRx (Disponível em: <https://pyrx.sourceforge.io/>) (Wolf, 2009), foram realizados *docking* da proteína GSTD1 com o ligante alcaloide mescalina ($C_{11}H_{17}NO_3$), para cada espécie do cluster, com o objetivo de estimar as mudanças nas posições de interações e afinidades químicas entre os modelos de GSTD1 e o alcaloide mescalina. O ligante mescalina foi selecionado para essa análise por estar presente em um grande número de cactáceas (Pardanani et al. 1978; Rätsch & Baker, 1993; Panis et al. 2016) e por possuir estrutura química definida (acesso PubChem CID: 4076). No programa PyMOL (DeLano, 2018), foram realizadas as medidas de distância entre o resíduo serina 10 do sítio-G e a posição do ligante na estrutura da proteína, com o intuito de mensurar as diferenças no posicionamento do alcaloide no sítio ativo. Posteriormente, no programa PyRx, foi estimado o gasto energético gerado no *docking* entre cada modelo da proteína GSTD1 com o ligante mescalina.

No programa PyMOL, foi realizada a análise de superfície hidrofóbica global da proteína GSTD1. O principal objetivo dessa análise foi evidenciar modificações bioquímicas em regiões da proteína, com possíveis relações no processo de interação química entre a proteína e o alcaloide. A análise de hidrofobicidade é realizada pelo critério de escala de cores, representada por vermelho, rosa claro e branco que corresponde às regiões hidrofóbicas, regiões com menos propriedades hidrofóbicas e regiões hidrofílicas, respectivamente.

4. RESULTADOS

4.1 Análise das sequências de DNA do gene *GSTD1*

Foram sequenciados 630 pb do gene *GSTD1* (*Glutationa S- transferase D1*), que correspondem a extensão total da região codificadora (CD), com um único exon. Os valores de PHRED calculados para as sequências do gene *GSTD1* ficaram entre 20 e 50, indicando alta confiabilidade de leitura. As informações obtidas no programa BLAST evidenciam grande similaridade entre as sequências de *GSTD1*, apresentando *scores E-value*=0.0 e identidade das sequências variando entre 92 a 95%, entre as espécies do cluster *Drosophila buzzatii* com os outros representantes de *Drosophila*.

Nos alinhamentos das sequências CDs de *GSTD1*, foram observadas mutações pontuais entre as espécies do cluster *D. buzzatii*, nas regiões N e C-terminal (Fig. 3). Para comparações sítios-específicos, foram realizados alinhamentos das sequências de *GSTD1* das espécies do cluster *D. buzzatii* com outras espécies de *Drosophila* cactofílicas, sendo estás: *D. aldrichi*, *D. mojavensis*, *D. arizonae*, *D. navojoa*, *D. hexastigma* e *D. anceps*. Além disso, foram alinhadas sequências de *GSTD1* de *D. eremophila* de hábito de solo e *D. virilis*, que utiliza outros recursos alimentares. Os alinhamentos foram realizados com base no parentesco filogenético entre as espécies (Oliveira et al.2012). Um conjunto de mutações é compartilhado entre as espécies do cluster *D. buzzatii*, a saber, nas posições: 94, 234, 243, 417, 444, 456, 531, 546 e 561 (Fig. 3). Em três indivíduos de *D. seriema* foram evidenciadas posições com sinal de heterozigoto, sendo elas: 204, 345, 351 e 354.

4.2 Composição da região codificadora (CD) e resíduos de aminoácidos de GSTD1

A composição das sequências nucleotídicas da região codificadora do gene *GSTD1*, entre as espécies do cluster, demonstra as seguintes proporções de conteúdo CG e AT: *D. buzzatii* com 53,81% e 46,19%, *D. koepferae* com 53,49% e 45,87%, *D. gouveai* com 53,81% e 46,19%, *D. seriema* com 53,65% e 45,56%, *D. serido* com 54,29% e 45,56%, *D. borborema* com 53,65% e 46,19%, e *D. antonietae* com 54,13% e 45,71%, respectivamente. Para as espécies do grupo externo foram observadas as seguintes proporções de conteúdo CG e AT: *D. mojavensis* com 55,87% e 44,13%, e *D. virilis* com 56,67% e 43,33%, respectivamente. Desta forma, as sequências gênicas de *GSTD1* demonstram maiores proporções em bases CG do que AT.

Nos alinhamentos das sequências de *GSTD1*, foram observadas 10 substituições não sinônimas entre as espécies do cluster *D. buzzatii*, nas posições: 56, 117, 349, 397, 408, 443, 585, 588, 624 e 629 (Fig. 3).

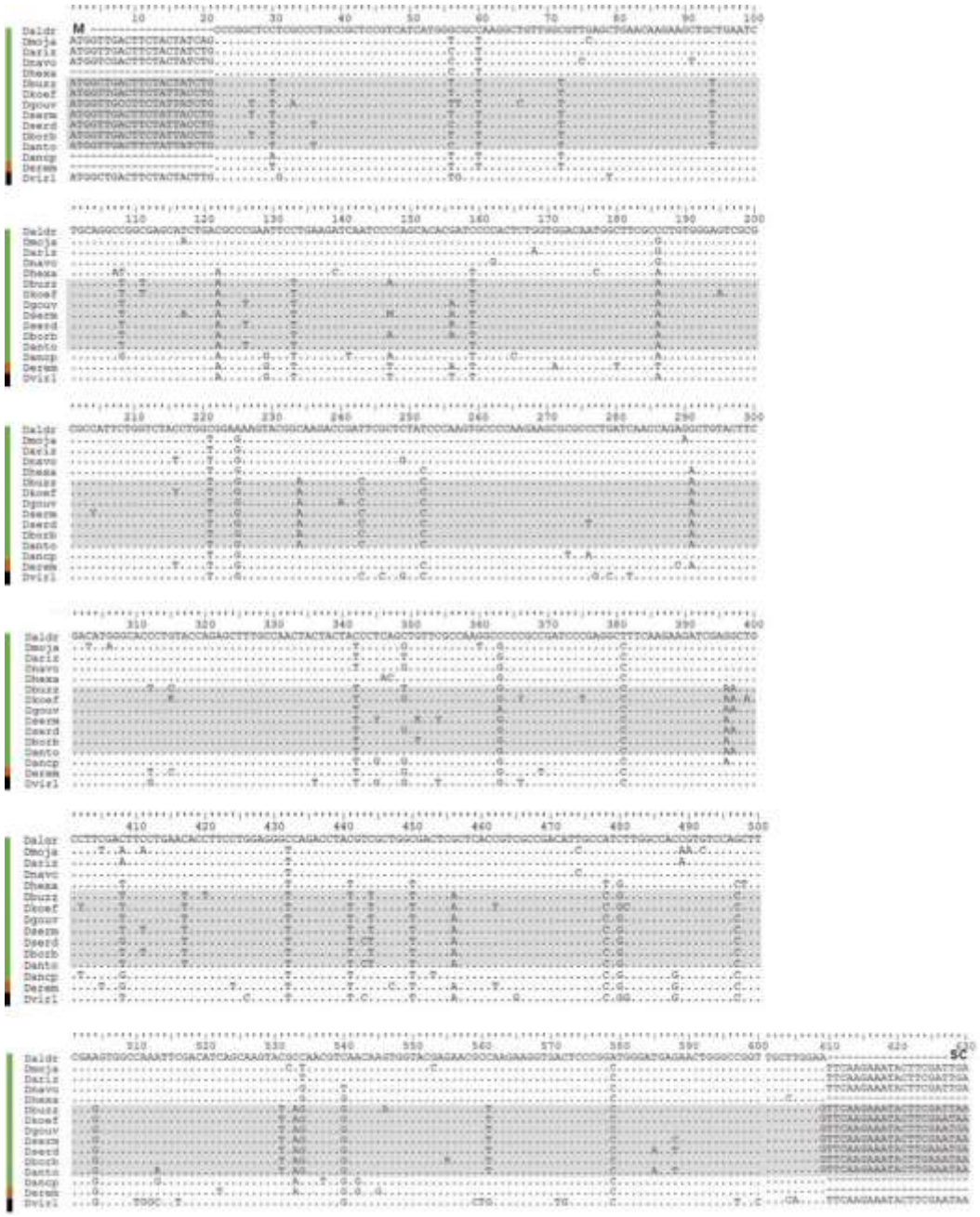


Fig. 3: Alinhamento das sequências codificadoras de *GSTD1* entre as espécies do cluster *Drosophila buzzatii*, representadas por Dbuzz - *D. buzzatii*, Dkoef - *D. koepferae*, Dgouv - *D. gouveai*, Dserm - *D. seriema*, Dserd - *D. serido*, Dborb - *D. borborema* e Danto - *D. antonietae*. O grupo externo está representado por Daldr - *D. aldrichi*, Dmoja - *D. mojavensis*, Dariz - *D. arizonae*, Dnavo - *D. navojoa*, Dhexa - *D. hexastigma*, Dancp - *D. anceps*, Derem - *D. eremophila* e Dvirl - *D. virilis*. Espécies de *Drosophila* estão representadas pela barra de cor verde para cactofílicas, de cor laranja para hábitos de solo e de cor preta para outros hábitos, que não inclui extratos de cactos. M - indica o primeiro resíduo de metionina da proteína GSTD1 e SC o Stop códon. Realce em cinza claro para as espécies do cluster. Sinais de heterozigotos representados por: M (A,C), Y (C,T), K (G,T) e R (A,G).

As sequências CDs completas do gene *GSTD1* resultam em uma proteína constituída por 209 resíduos de aminoácidos (Fig. 4). Uma vez que foi reconhecida a constituição e posição dos resíduos de aminoácidos na cadeia primária da proteína GSTD1, foram reconhecidas as regiões N-terminal, que compreende os resíduos 1-79, e C-terminal, com os resíduos 86-208. Além disso, foram identificados os resíduos de aminoácidos que constituem o sítio-G e o sítio-H (Fig. 5).

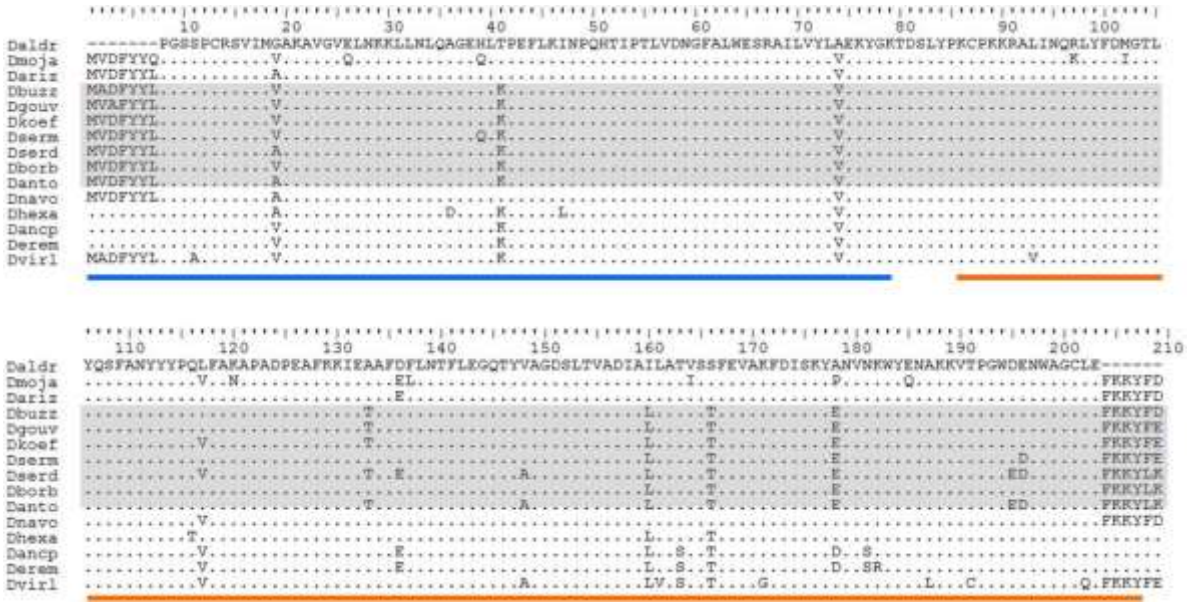


Figura 4: Alinhamento das sequências primárias da proteína GSTD1 entre as espécies do cluster *D. buzzatii*, representadas por Dbuzz - *D. buzzatii*, Dkoef - *D. koepferae*, Dgouv - *D. gouveai*, Dserm - *D. seriema*, Dserd - *D. serido*, Dborb - *D. borborema* e Danto - *D. antonietae*. O grupo externo está representado por Daldr - *D. aldrichi*, Dmoja - *D. mojavensis*, Dariz - *D. arizonae*, Dnavo - *D. navojoa*, Dhexa - *D. hexastigma*, Dancp - *D. anceps*, Derem - *D. eremophila* e Dvirl - *D. virilis*. Realce em cinza claro representa as sequências das espécies do cluster *D. buzzatii*. A região sublinhada em azul representa o domínio N-terminal e em laranja corresponde ao domínio C-terminal da proteína.

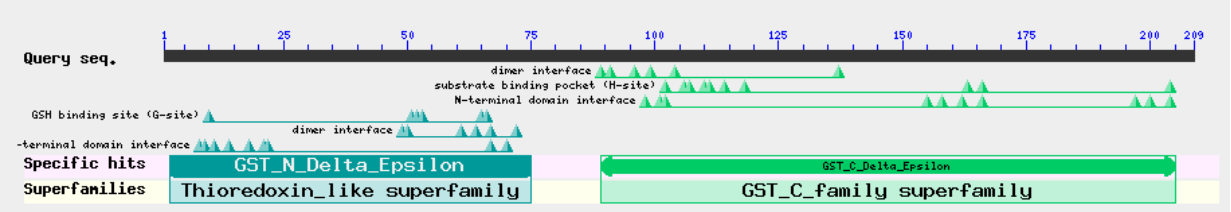


Fig. 5: Elementos estruturais da proteína GSTD1, realizado no programa BLASTP, usando as sequências primárias das proteínas de cada espécie do cluster. A escala em linha negra corresponde à composição em resíduos de aminoácidos da proteína (1-209). Em evidência as regiões de ligação ao GSH, sitio-G e o sítio-H. Caixas em azul piscina representam a região N-terminal e em verde corresponde à região C-terminal.

Nas sequências primárias da proteína GSTD1 foram evidenciadas substituições únicas em aminoácidos para *D. buzzatii* no resíduo 2, *D. gouveai* no resíduo 3, *D. seriema* nos resíduos 39 e 136, e em *D. antonietae* nos resíduos 160 e 166. Sítios com variações compartilhadas foram observadas entre: *D. serido* e *D. antonietae* no resíduo 19, *D. koepferae* e *D. serido* no resíduo 117, *D. koepferae*, *D. serido* e *D. antonietae* no resíduo 133, *D. seriema* e *D. borborema* nos resíduos 148 e 195, *D. koepferae*, *D. seriema* e *D. borborema* no resíduo 196 e para as espécies *D. seriema*, *D. serido* e *D. borborema* nos resíduos 208 e 209 (Tabela 2).

Substituições em aminoácidos da proteína GSTD1 que resultaram em variações em cargas energéticas estão representadas da seguinte forma: *D. gouveai* na posição 3 Asp/Ala, em *D. seriema* na posição 39 His/Gln, para *D. antonietae*, *D. serido* e *D. koepferae* na posição 133 Thr/Ala, e para *D. borborema*, *D. serido* e *D. seriema* na posição 209 D/K (Tabela 2).

Foi observada variação em proporção entre resíduos de aminoácidos da proteína GSTD1 para as espécies do cluster sendo: 0.9% para alanina, 1.0% para ácido aspártico, 0.5% para ácido glutâmico, 0.5% para histidina e 1.0% para lisina. Para *D. arizonae*, *D. navojoa* e *D. virilis* foram observadas maiores proporções de alanina (0.5%). Para *D. antonietae*, *D. borborema*, *D. serido* e *D. mojavensis* foram evidenciadas menores proporções de resíduos de fenilalanina (6.2%) do que nas demais espécies estudadas (6.7%) (Fig. 6). As variações em resíduos de aminoácidos foram estatisticamente significantes, a saber, entre os resíduos de ácido glutâmico e ácido aspártico (p=0,00005), histidina e lisina (p=0,00000) e alanina e ácido aspártico (p=0,00000), no entanto, não foi significante entre alanina e lisina (p=0,28456). Essas variações em resíduos influenciaram a estrutura química e no peso molecular da proteína GSTD1.

Os pesos moleculares da proteína GSTD1, inferidos com os conjuntos de resíduos de aminoácidos, foram: 23881.24Da para *D. buzzatii*, 23879.32Da para *D. gouveai*, 23909.31Da para *D. koepferae*, 23870.26Da para *D. seriema*, 23832.26Da para *D. serido*, 23858.34Da para *D. borborema* e 23832.25Da para *D. antonietae*. Os pesos das proteínas para as espécies do grupo externo foram: 23730.07Da para *D. mojavensis*, 23766.10Da para *D. arizonae*, 23738.05Da para *D. navojoa* e 23695.02Da para *D. virilis*. Estes valores sugerem uma semelhança entre as estimativas de peso molecular das proteínas GSTD1 para as espécies do cluster *D. buzzatii*, com evidência de sutis diferenças com os representantes do grupo externo.

Tabela 2: Sítios variantes no alinhamento dos resíduos aminoácidos da proteína primária de GSTD1 entre as espécies do cluster *Drosophila buzzatii* e grupo externo. Os resíduos aminoácidos das espécies do cluster estão realçados em cinza e as variações evidenciadas em negrito. (*) representa variações em categorias bioquímicas. (-) indica ausência/lacuna de sequência na região.

		*						*										*			
Espécie	**2**	**3**	**7**	**11**	**19**	**26**	**36**	**39**	**41**	**47**	**74**	**93**	**97**	**102**	**116**	**117**	**120**	**133**	**136**	**137**	**148**
D. aldrichi	-	-	-	S	G	E	A	H	T	I	A	L	K	I	Q	L	N	A	E	L	V
D. mojavensis	V	D	Q	S	V	Q	A	Q	T	I	V	L	R	M	Q	V	K	A	E	F	V
D. arizonae	V	D	L	S	A	E	A	H	T	I	V	L	R	M	Q	L	K	T	D	F	V
D. buzzatii	**A**	D	L	S	V	E	A	H	K	I	V	L	R	M	Q	L	K	**T**	D	F	V
D. gouveai	V	**A**	L	S	V	E	A	H	K	I	V	L	R	M	Q	L	K	**T**	D	F	V
D. koepferae	V	D	L	S	V	E	A	H	K	I	V	L	R	M	Q	**V**	K	A	D	F	V
D. seriema	V	D	L	S	V	E	A	**Q**	K	I	V	L	R	M	Q	L	K	**T**	**E**	F	**A**
D. serido	V	D	L	S	**A**	E	A	H	K	I	V	L	R	M	Q	**V**	K	A	D	F	V
D. borborema	V	D	L	S	V	E	A	H	K	I	V	L	R	M	Q	L	K	**T**	D	F	**A**
D. antonitae	V	D	L	S	**A**	E	A	H	K	I	V	L	R	M	Q	L	K	A	D	F	V
D. navojoa	V	D	L	S	A	E	A	H	T	I	V	L	R	M	Q	V	K	A	D	F	V
D. hexastigma	-	-	-	S	A	E	D	H	K	L	V	L	R	M	T	L	K	A	E	F	V
D. anceps	-	-	-	S	V	E	A	H	K	I	V	L	R	M	Q	V	K	A	E	F	V
D. eremophila	-	-	-	S	V	E	A	H	K	I	V	L	R	M	Q	V	K	T	D	F	V
D. virilis	A	D	L	A	V	E	A	H	K	I	V	V	R	M	Q	V	K	T	D	F	V

																*
Espécie	**160**	**161**	**163**	**164**	**166**	**171**	**181**	**182**	**185**	**187**	**191**	**195**	**196**	**202**	**208**	**209**
D. aldrichi	I	L	T	I	S	K	N	K	-	-	-	-	-	-	-	-
D. mojavensis	I	L	T	V	S	K	N	K	Q	A	T	D	E	L	F	D
D. arizonae	L	L	T	V	T	K	N	K	E	A	T	D	E	L	F	D
D. buzzatii	L	L	T	V	T	K	N	K	E	A	T	D	E	L	F	**E**
D. gouveai	L	L	T	V	T	K	N	K	E	A	T	D	E	L	F	**E**
D. koepferae	L	L	T	V	T	K	N	K	E	A	T	D	**D**	L	F	**E**
D. seriema	L	L	T	V	T	K	N	K	E	A	T	**E**	**D**	L	**L**	**K**
D. serido	L	L	T	V	T	K	N	K	E	A	T	D	E	L	**L**	**K**
D. borborema	L	L	T	V	T	K	N	K	E	A	T	**E**	**D**	L	**L**	**K**
D. antonitae	I	L	T	V	**S**	K	N	K	E	A	T	D	E	L	F	D
D. navojoa	L	L	T	V	T	K	N	K	E	A	T	D	E	L	F	D
D. hexastigma	L	L	S	V	T	K	N	K	E	A	T	D	E	L	-	-
D. anceps	L	L	S	V	T	K	S	K	E	A	T	D	E	L	-	-
D. eremophila	L	V	S	V	T	K	S	R	E	A	T	D	E	L	-	-
D. virilis	I	L	T	V	S	G	N	K	E	L	C	D	E	Q	F	E

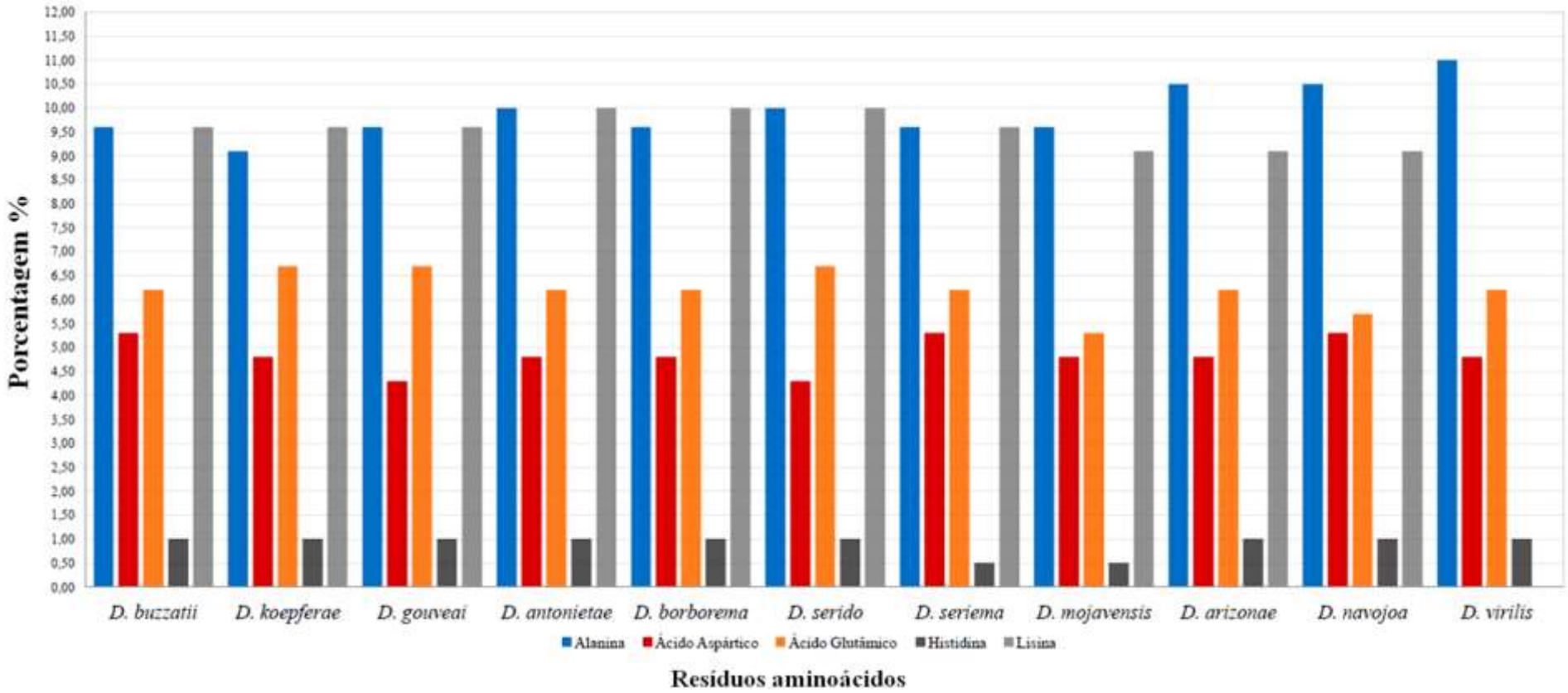


Fig. 6: Representação de variações em proporções de resíduos de aminoácidos da proteína GSTD1 entre as espécies do cluster *Drosophila buzzatii* e grupo externo representado por *D. mojavensis*, *D. arizonae*, *D. navojoa* e *D. virilis*. As variações em resíduos aminoácidos e categorias bioquímicas foram para alanina (apolar), ácido aspártico e ácido glutâmico (polar negativo), histidina e lisina (polar positivo).

4.3 Inferências filogenéticas para as espécies do cluster *Drosophila buzzatii*

Gene nuclear *GSTD1*

Para as inferências filogenéticas Bayesianas (BI) com o gene *GSTD1*, foram considerados 11 espécies e 630 caracteres nucleotídeos. Em virtude da ausência de sequências completas da região codificadora do gene *GSTD1* para espécies externas ao cluster, foram retirados das análises os representantes de *D. aldrichi*, *D. hexastigma*, *D. anceps* e *D. eremophila*.

Para a árvore BI consenso das sequências do gene *GSTD1*, o valor de log *LnL* (máxima probabilidade da melhor árvore) calculado foi - 1737.2559. Os principais resultados da análise filogenética por BI das sequências completas do gene *GSTD1* (Fig. 7) foram os seguintes:

- Monofilia das espécies cactofílicas *D. mojavensis*, *D. arizonae* e *D. navojoa*, com alto valor de probabilidade a posteriori (PP=1);
- Monofilia das espécies do cluster *D. buzzatii*, com alto valor de probabilidade a posteriori (PP=1);
- Entre as espécies do cluster *D. buzzatii* foi observado uma politomia para *D. koepferae* e *D. buzzatii* e dois agrupamentos principais: o primeiro composto por *D. antonietae* e *D. serido*, com alto valor de probabilidade a posteriori (PP=1); e o segundo agrupamento é composto pelas espécies *D. gouveai*, *D. borborema* e *D. seriema*, sendo que *D. gouveai* forma um clado, com alto valor de probabilidade a posteriori (PP=0,94), e uma ramificação composta por *D. borborema* e *D. seriema*, com alto valor de probabilidade a posteriori (PP=1).

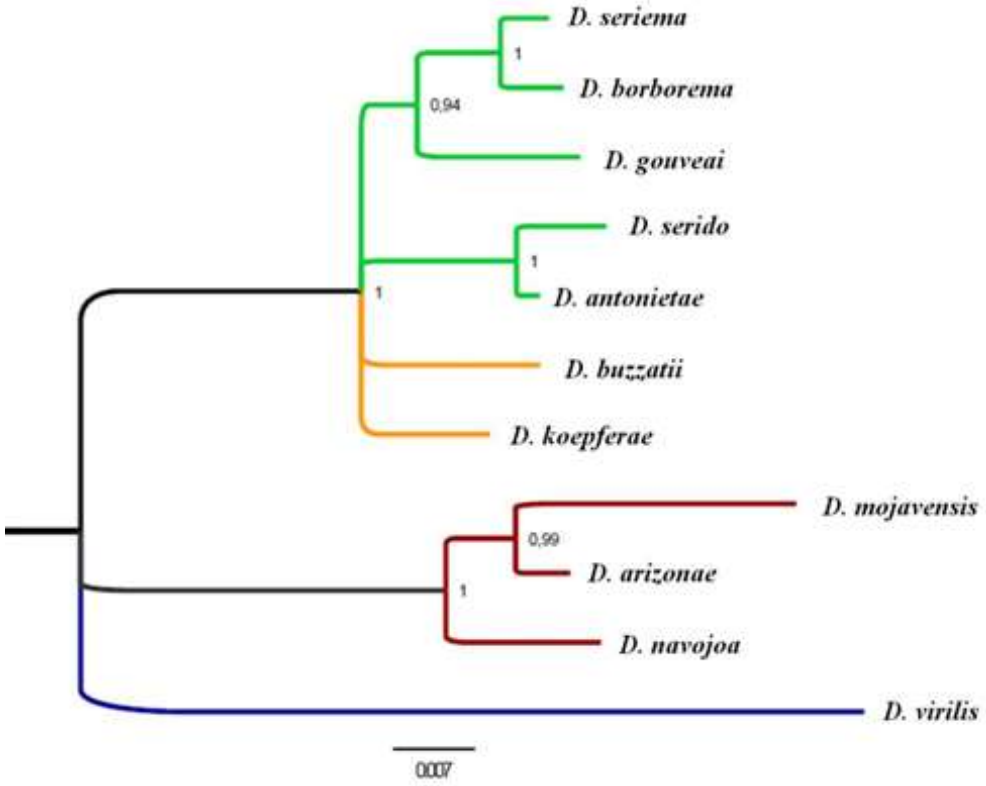


Fig. 7: Relações filogenéticas entre as espécies do cluster *D. buzzatii* inferidas com sequências completas do gene *GSTD1*, usando estatística Bayesiana, com o modelo GTR+G (BIC: 3790.5951 e deltaBIC: 15.9815). *Drosophila virilis* foi usada como grupo externo. A probabilidade posterior (PP) está localizada ao lado de cada ramo. A escala indica o número de substituições/sítio.

Gene mitocondrial *COI*

Nas inferências filogenéticas BI usando o gene mitocondrial *COI*, foram analisados 549 pb de 11 espécies, representadas por um indivíduo cada. Foram retiradas das análises *D. aldrichi*, *D. hexastigma*, *D. anceps* e *D. eremophila* devido ao fato de não haver sequências deste gene para estas espécies.

Para a árvore BI consenso definida a partir de sequências do gene *COI*, o valor de log *LnL* calculado foi -1809.982 (Fig. 8). Os principais resultados da análise foram:

- Monofilia das espécies cactofílicas *D. mojavensis*, *D. arizonae* e *D. navojoa*, com alto valor de probabilidade a posteriori (PP=1);
- Monofilia das espécies do cluster *D. buzzatii*, com alto valor de probabilidade a posteriori (PP=1);
- As espécies do cluster foram agrupadas em dois clados: o primeiro composto por *D. koepferae* e *D. buzzatii*, com baixo suporte de probabilidade a posteriori (PP= 0.55); o segundo composto por *D. antonietae*, *D. seriema*, *D. gouveai*, *D. borborema* e *D. serido*, com alto valor de probabilidade a posteriori (PP=1), sendo uma ramificação com *D. antonietae*, com alto valor probabilidade a posteriori (PP=1), uma ramificação

com *D. seriema* e *D. gouveai* formando um agrupamento, com alto valor de probabilidade a posteriori (PP=1), e por último um ramo com *D. borborema* e *D. serido*, com alto valor de probabilidade a posteriori (PP=0,98).

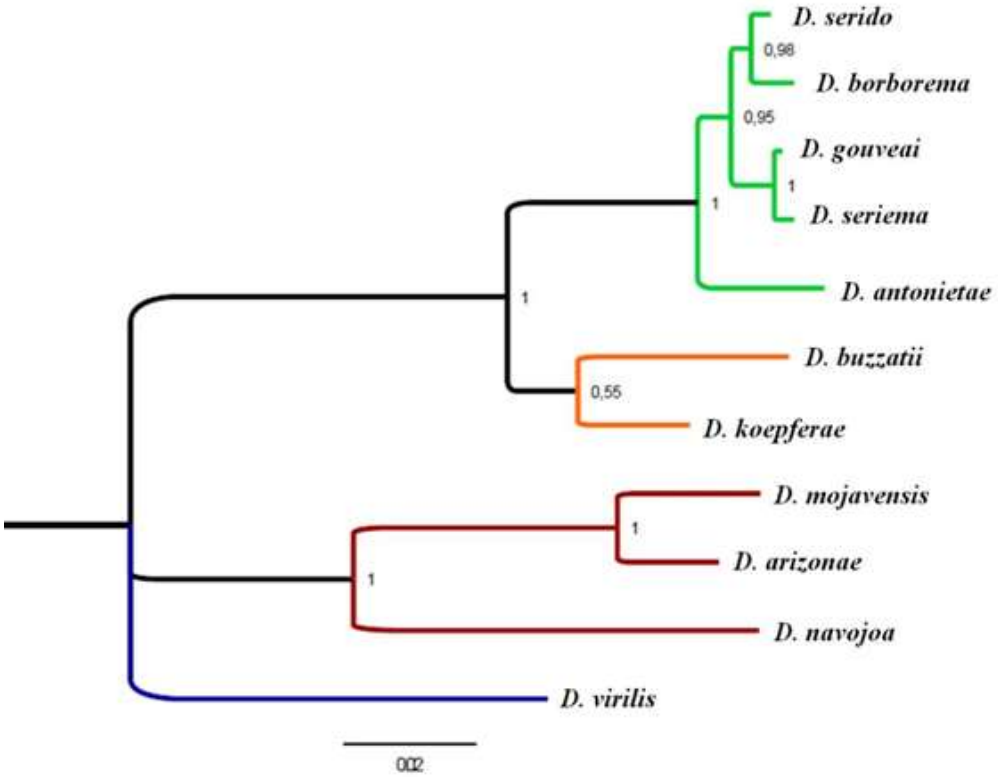


Fig. 8: Relações filogenéticas entre as espécies do cluster *D. buzzatii* inferidas com sequências parciais do gene *COI*, usando a metodologia Bayesiana, com o modelo GTR+G (BIC: 4824.193 e deltaBIC: 0.00). *Drosophila virilis* foi usada como grupo externo. A probabilidade posterior (PP) está localizada ao lado de cada ramo. A escala indica o número de substituições/sítio.

Genes *COI* e *GSTD1*

Nas inferências filogenéticas BI com os dados concatenados entre as sequências parciais do gene mitocondrial *COI* e total de *GSTD1*, foram consideradas 11 espécies e sequências de 1179 pb nucleotídicos. Para a árvore BI consenso das sequências *COI* e *GSTD1* foi calculado o valor de log *LnL* de -3623.0135 (Fig. 9).

A árvore BI consenso obtida teve como principais resultados:

- Monofilia das espécies cactofílicas *D. navojoa*, *D. mojavensis* e *D. arizonae*, com alto valor de probabilidade a posteriori (PP=1);
- Monofilia das espécies do cluster *D. buzzatii*, com alto valor probabilidade a posteriori (PP=1);

- As espécies do cluster *D. buzzatii*, estão organizadas em dois agrupamentos: o primeiro é composto por *D. koepferae* e *D. buzzatii*, com valor de probabilidade a posteriori (PP=0,893) e o segundo composto por *D. seriema*, *D. gouveai*, *D. borborema*, *D. serido* e *D. antonietae*, com alto valor de probabilidade a posteriori (PP=1); neste clado *D. seriema* e *D. gouveai* foram agrupadas em um ramo, com alto valor de probabilidade a posteriori (PP=0,99); *D. borborema* formou um clado único, com alto valor de probabilidade a posteriori (PP=0,93) e *D. serido* e *D. antonietae* foram agrupadas em um clado, com alto valor de probabilidade a posteriori (PP=1).

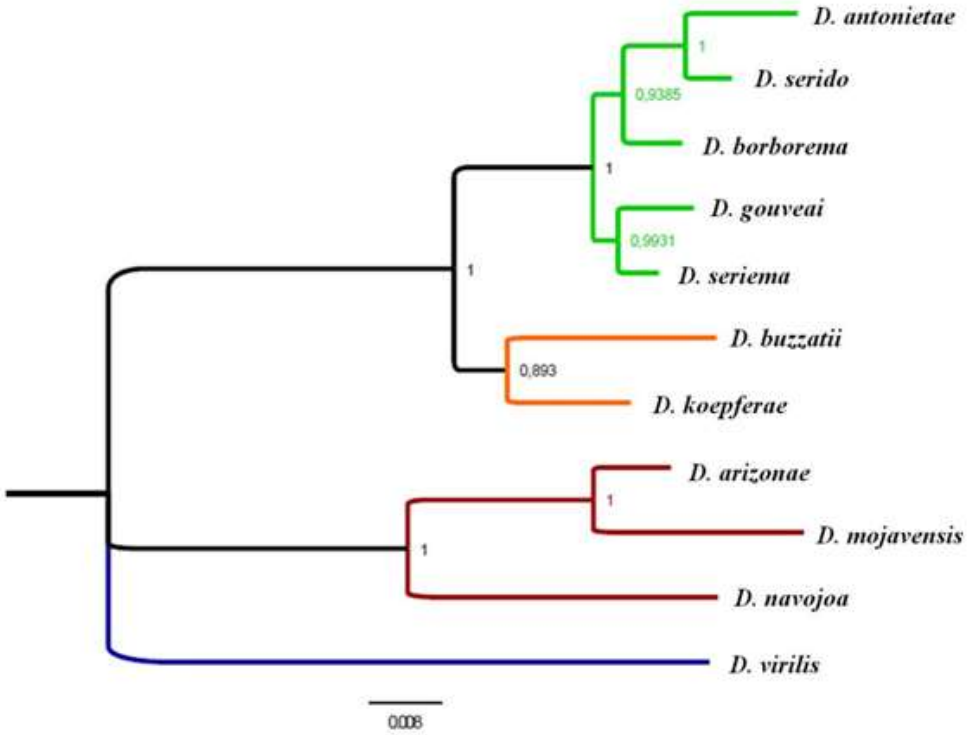


Fig. 9: Relações filogenéticas entre as espécies do cluster *D. buzzatii* inferidas com dados combinados entre as sequências parciais do gene *COI* e completas de *GSTD1*, usando a metodologia Bayesiana, com o modelo GTR. *Drosophila virilis* foi usada como grupo externo. A probabilidade posterior (PP) está localizada ao lado de cada ramo. A escala indica o número de substituições/sítio.

4.4 Testes de Seleção

Teste de seleção para o gene *GSTD1* com o modelo M8 sugere seleção positiva para os códons nas posições 39 (ω=1.5, Intervalo de Confiança - IC: 0.076, 4.9), 117 (ω=4.9, IC: 4.9, 4.9), 133 (ω=3.4, IC: 0.35, 4.9), 136 (ω=1.6, IC: 0.076, 4.9) e 209 (ω=4.5, IC: 0.35, 4.9), nos respectivos resíduos de aminoácidos (Fig. 10). O modelo M7 (beta), aplicado as sequências de *GSTD1* indica sinais de seleção purificadora para os códons nas posições 2 (ω=0.64 – IC: 0.033, 1), 7 (ω=0.51 – IC: 0.033, 1) e 19 (ω=0.77 - IC: 0.16, 1), na região N-

terminal, e nos códons das posições 160 (ω=0.49 - IC: 0.033, 1), 195 (ω=0.46 – IC: 0.033, 1), 196 (ω=0.52 – IC: 0.033, 1), da região C-terminal e por último o códon 208 (ω=0.51 – IC: 0.033, 1), na região da cauda do C-terminal. A análise comparativa entre os modelos M8 *versus* M7, com normalização dos valores de ω em 1, evidência sinal de seleção positiva para os códons nas posições 117 e 209 (ω=1.2) e nas posições 19, 39, 133, 136, 196 e 208 (ω=1.1), nos respectivos resíduos de aminoácidos. O teste de razão de verossimilhança (LRT) indica significância estatística de $p<0,01$ para os sinais de seleção positiva em códons de *GSTD1*, de acordo com o modelo M8 *vs* M7 (Tabela 3).

O resíduo de histidina, na posição 39, possui sinal de seleção positiva em *D. buzzatii*, *D. gouveai*, *D. koepferae*, *D. serido*, *D. borborema* e *D. antonietae*. O resíduo de ácido aspártico, na posição 209, apresenta sinal de seleção positiva, em *D. buzzatii*. Os resíduos 39 e 209, apresentam localização interna no sítio-G e na cauda C-terminal, da proteína GSTD1, respectivamente. Os resíduos de alanina e ácido aspártico, nas posições 133 e 136, respectivamente, apresentam sinais de seleção positiva em *D. borborema* e *D. seriema*. O resíduo de valina, na posição 117, apresenta sinal de seleção positiva, em *D. koepferae* e *D. serido*. Esses últimos resíduos possuem localização interna, na estrutura da proteína GSTD1.

Todos os resíduos de aminoácidos restantes apresentaram ω<1 e um intervalo de confiança que não é incluído um, que pode ser um indicativo de seleção purificadora, estando de acordo com o teste "Z-selection" aplicado nas sequências completas de *GSTD1* (Tabela 4). Teste de probabilidade Z-selection foi aplicado em pares de sequências em um total de 197 posições de códons de resíduos de aminoácidos, sendo que cerca de 3% dos códons foram retirados da análise, por erros de bases ambíguas. Os resultados obtidos sugerem sinal de seleção positiva significante (P= 0.023), entre as espécies *D. borborema* e *D. seriema* (Tabela 5).

Fig. 10: Mapa de resíduos de aminoácidos com sinal de seleção positiva, realizado no servidor SELECTON, com o modelo M8, para as sequências primárias da proteína GSTD1 das espécies estudadas. Em amarelo, resíduos com sinal de seleção positiva e em magenta resíduos com sinal de seleção purificadora ($p<0,005$).

Tabela 3: Testes de seleção com estimativas dos parâmetros para cada modelo, valores de ω, sítios em seleção, *likelihood* (lnL) e teste razão de verossimilhança LRT (2Δl) entre os modelos de seleção no gene *GSTD1*.

Modelo	Parâmetros	Sítios em Seleção	lnL	2Δl
M8 (beta, ω>1)	$p_1(1-p_0)$; $\omega_S>1$	His 39, **Val 117**, Ala133, Asp 136, Asp 209.	-1446.8	M8 *vs* M8a -168,12
M8a (beta, ω<1)	$p_1(1-p_0)$; $\omega_S=1$	Val 2, Leu 7, Ala 19, Val 148, Glu 196, Phe 208.	-1530.86	
M7 (beta)	p + q	Val 2, Leu 7, Ala 19, Ile 160, Asp 195, Glu 196, Phe 208.	-1413.5	M8 *vs* M7 66,60**
MEC	tr, tv, ω, f	Ala 19, His 39, **Val 117**, Ala 133, Asp 136, Glu 196, Phe 208, **Asp 209**.	-1389.57	MEC *vs* M8a -282,58

Em negrito, estão realçados os resíduos aminoácidos com sinal de seleção significativo (*P*< 0,05). Estimativa LRT entre M8 *vs* M8a, considerando 1 grau de liberdade (gl) e valores de qui-quadrado de 3,84, 6,64 e 10,83 em 5%*, 1%** e 0,1%*** respectivamente. Para comparações entre os modelos M8 *vs* M7 e MEC *vs* M8a, foram considerados 2 gl e valores de qui-quadrado em 5,99, 9,21, 13,82 em 5%*, 1%** e 0,1%***.

Tabela 4: Teste de hipótese Z-Selection, com o método de Nei e Gobojori (Nei & Gobojori, 1986) aplicado em sequências completas do gene *GSTD1*.

Teste-Z de seleção (H0: dn=ds)	dn-ds	Significância
Teste de Neutralidade (HA:dN≠dS)	-8.1	*P*< 0.001 (rejeito H0)
Seleção Positiva (HA: dN>dS)	-8.5	*P*= 1.000 (aceito H0)
Seleção Purificadora (HA: dN<dS)	-8.6	*P*< 0.001 (rejeito H0)

Tabela 5: Teste Z–Selection, com o método Nei-Gojobori, aplicado em pares de sequências do gene *GSTD1*.

Espécie	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1-*D. navojoa*		-2,122	-3,165	-6,106	-6,640	-6,571	-6,036	-5,754	-6,033	-5,686
2-*D. mojavensis*	1,000		-0,548	-5,134	-5,691	-5,624	-5,061	-5,079	-4,919	-4,577
3-*D. arizonae*	1,000	1,000		-5,380	-5,926	-5,855	-5,307	-5,440	-5,305	-4,944
4-*D. buzzatii*	1,000	1,000	1,000		-3,051	-3,305	-2,288	-2,054	-2,287	-1,675
5-*D. koepferae*	1,000	1,000	1,000	1,000		-3,845	-2,532	-2,275	-2,531	-2,389
6-*D. gouveai*	1,000	1,000	1,000	1,000	1,000		-2,408	-2,162	-2,407	-1,796
7-*D. seriema*	1,000	1,000	1,000	1,000	1,000	1,000		-1,396	2,009	-1,526
8-*D. serido*	1,000	1,000	1,000	1,000	1,000	1,000	1,000		-1,363	-1,178
9-*D. borborema*	1,000	1,000	1,000	1,000	1,000	1,000	**0,023**	1,000		-1,508
10-*D. antonietae*	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	

Probabilidade de rejeitar a hipótese nula de neutralidade estrita (dN = dS) é mostrada abaixo na diagonal. Valores de $P < 0,05$ são considerados estatisticamente significativos, com nível de confiança 95%, realçado em negrito. A estatística do teste (dN - dS) é mostrada acima da diagonal.

4.5 Modelagem e análise *in silico* da proteína GSTD1

No cluster *D. buzzatii*, a proteína primária de GSTD1 é composta por 209 resíduos de aminoácidos, apresentando estrutura terciária que compreende nove α-hélices e quatro β-folhas (Fig. 11). O domínio N-terminal é menor, compreende três α-hélices e quatro β-folhas e o domínio C-terminal é maior, abrangendo seis α-hélices. A α7 é representada por um pequena estrutura helicoidal.

Os alinhamentos múltiplos realizados com a estrutura terciária da proteína GSTD1 não mostraram modificações estruturais significantes entre as espécies do cluster *Drosophila buzzatii* e o grupo externo representado por *D. mojavensis*, *D. arizonae* e *D. navojoa* (Fig. 12). Além disso, foram identificados os resíduos de aminoácidos do sítio de ligação GSH (glutationa redutase), do sítio-G (resíduos Ser10, Pro12, Leu34, His39, Leu40, His51, Thr52, Ile53, Pro54, Glu65, Ser66, Arg67, Met102 e Tyr114) e do sítio hidrofóbico H (resíduos 68, 117–121, 123, 159, 160, 170 e 174) (Fig.14), conforme estimado para *D. melanogaster* (Low et al. 2010).

O resíduo de lisina na posição 171 está posicionado próximo do sítio-H (Fig.13), em todos os modelos da proteína GSTD1 entre as espécies do cluster *D. buzzatii*, com função relacionada a desintoxicação ao pesticida DDT (dicloro-difenil-tricloroetano) (Low et al. 2007).

Diferenças estruturais observadas nos alinhamentos múltiplos foram analisadas na composição de resíduos do sítio-G, devido à sua função de ligação aos substratos químicos.

Para as espécies do cluster *D. buzzatii*, no sitio G, foram observadas substituições em *D. seriema* no resíduos 39 (His/Glu), em *D. serido* e *D. koepferae* no resíduo 117 (Lys/Val), e por fim, em *D. buzzatii* foi observado em resíduos nas posições 40 (L/H) e 51 (His/Leu). Substituições fora dos sítios-G e H foram identificadas em resíduos de aminoácidos nas seguintes posições: 195 (Asp/Glu) para *D. borborema* e *D. seriema*; 196 (Glu/Asp) para *D. borborema*, *D. seriema* e *D. koepferae*; e 208 (Phe/Leu) para *D. borborema*, *D. serido* e *D. seriema*. Por último, foram observadas várias substituições no resíduo 209, para *D. borborema*, *D. seriema* e *D. serido* ocorreu substituições de Asp/Lys, e para *D. koepferae*, *D. gouveai* e *D. buzzatii* mudanças de Asp/Glu (Tabela 2). As substituições no resíduo 196 estão localizadas na α-hélice 8, na região C-terminal, que forma uma ponte de sal com o resíduos de Lys na posição 29, da região N-terminal (Fig. 11).

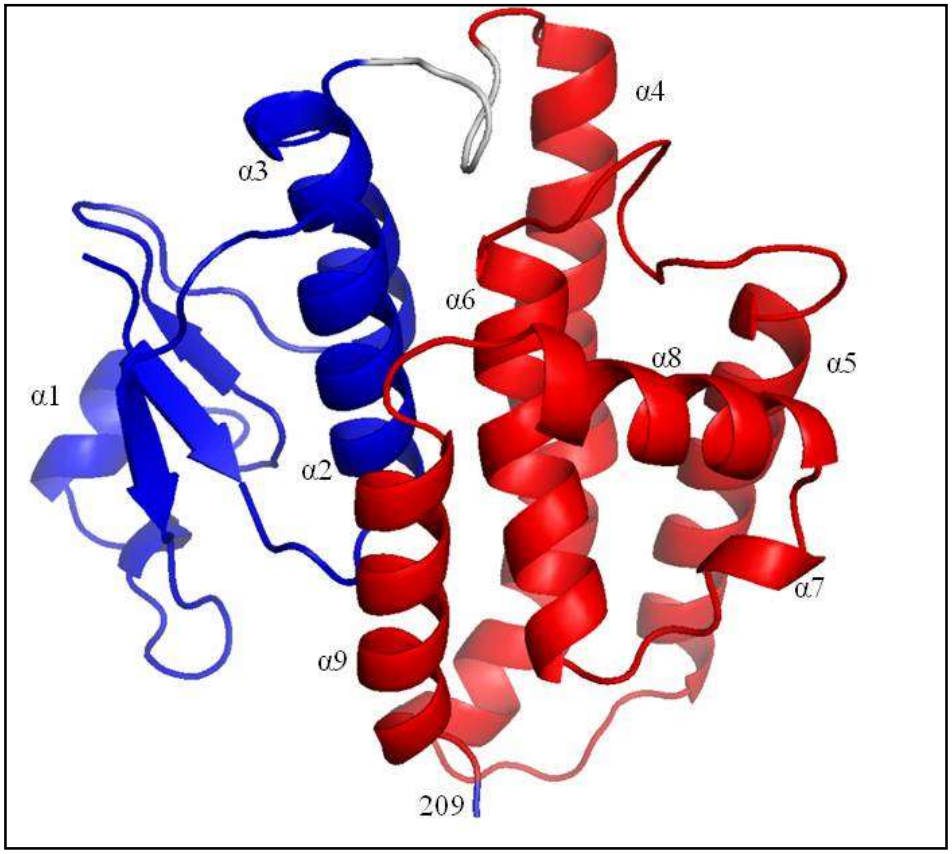


Fig. 11: Modelo representativo da estrutura terciária da proteína GSTD1 para as espécies do cluster *D. buzzatii*, estilo em *cartoon ribbon*, com a identificação dos elementos estruturais terciários de α-helice e β-folha da proteína GSTD1, representados nas regiões N-terminal em azul e C-terminal em vermelho. Resíduo 209, em cor azul, representa a cauda da região C-terminal. Identificações dos elementos estruturais seguiram informações prévias publicadas para *D. melanogaster* (Low et al. 2010).

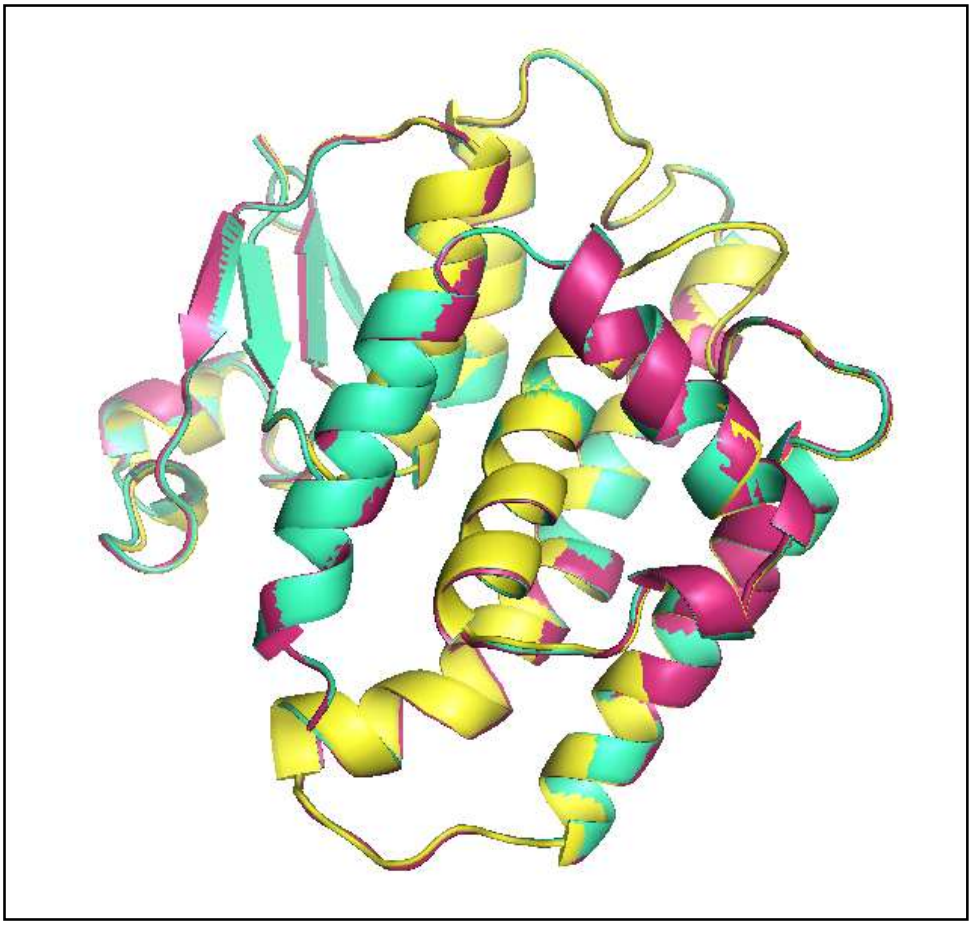


Fig. 12: Alinhamentos múltiplos dos modelos de estrutura terciária da proteína GSTD1, realizado no programa Chimera, para as espécies do cluster *D. buzzatii* e do grupo externo representado por *D. mojavensis*, *D. navojoa* e *D. arizonae*. Os modelos demonstram os elementos de α-helice e β-folha, em estilo *cartoon ribbon*.

4.6 *Docking* molecular da proteína GSTD1 com o ligante mescalina

Para as espécies do cluster *Drosophila buzzatii*, foram realizados *docking* da proteína *GSTD1* com o ligante mescalina ($C_{11}H_{17}NO_3$), usando as modelagens *in silico* de cada proteína, que evidenciam alterações em regiões de afinidade de ligação com o substrato alcaloide (Fig. 14-20). Nas Figuras 14 a 20 estão indicadas as diferenças calculadas na distância do posicionamento do alcaloide com o resíduo serina 10 (sítio-G), responsável por estabilidade e de ligação principal do sítio ativo. Com efeito, as substituições em resíduos aminoácidos observados na composição do sítio-G influenciam na afinidade química de ligação da proteína com o ligante mescalina. Comparações entre o *docking* realizado para cada espécie do cluster evidenciam variações em gastos energéticos no processo de interação do sítio-G com o alcaloide. Foram obtidos os menores valores de gastos energéticos, para as seguintes espécies: *D. seriema* e *D. gouveai* com gasto energético de -5,9 Kcal/mol, *D. koepferae* com gasto energético de -5,6 Kcal/mol e com distâncias do resíduo de serina 10,

correspondendo a 8,9Å, 8,8Å e 8,7Å, respectivamente. Para *Drosophila antonietae* o gasto energético foi -5,3 Kcal/mol e a distância de 8,8Å. *Drosophila borborema* apresentou gasto energético de -5,3 Kcal/mol e a distância de 4,1Å. Para *D. serido* e *D. buzzatii* foram observados gastos energéticos de -5,1 e -4,8 Kcal/mol, e distâncias do resíduo de serina 10 em 4,1Å e 6,5Å, respectivamente. Em conclusão, *D. seriema*, *D. gouveai* e *D. koepferae* demonstram os melhores modelos da proteína GSTD1 em relação a interação molecular com o ligante mescalina, em termos de gastos energéticos, entre as espécies do cluster *D. buzzatii*.

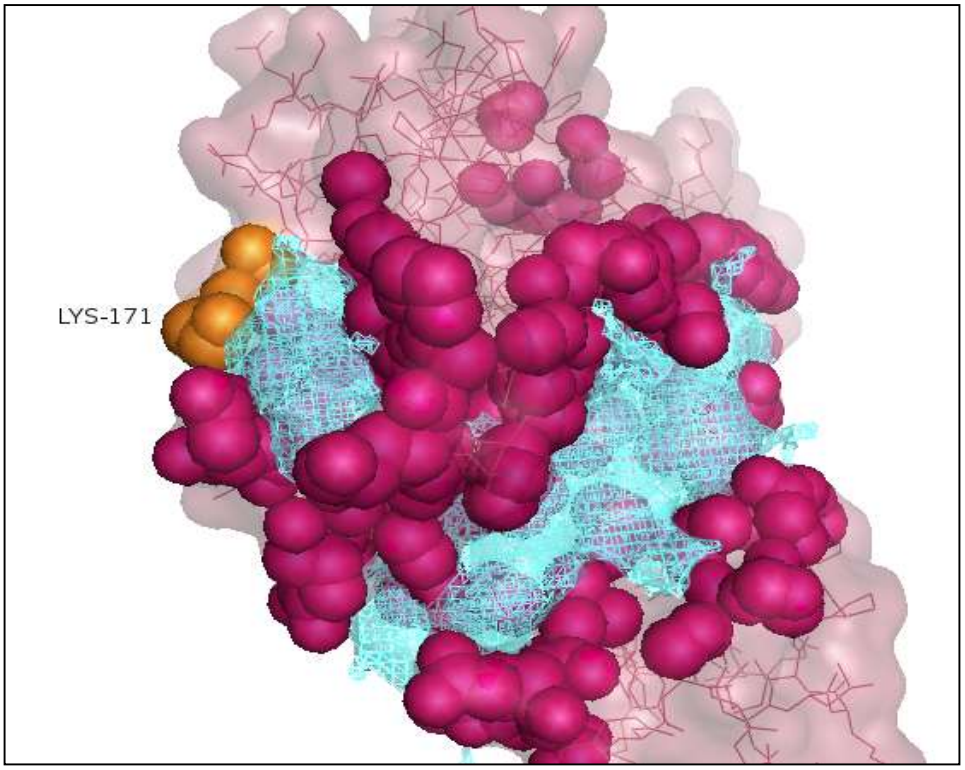


Fig. 13: Superfície proteica de GSTD1 inferida *in silico*. Em cor margenta estão representados os sítios G e H, representando o maior e menor *pocket*, respectivamente. O resíduo do aminoácido de lisina 171, está representado pela região destacada em amarelo.

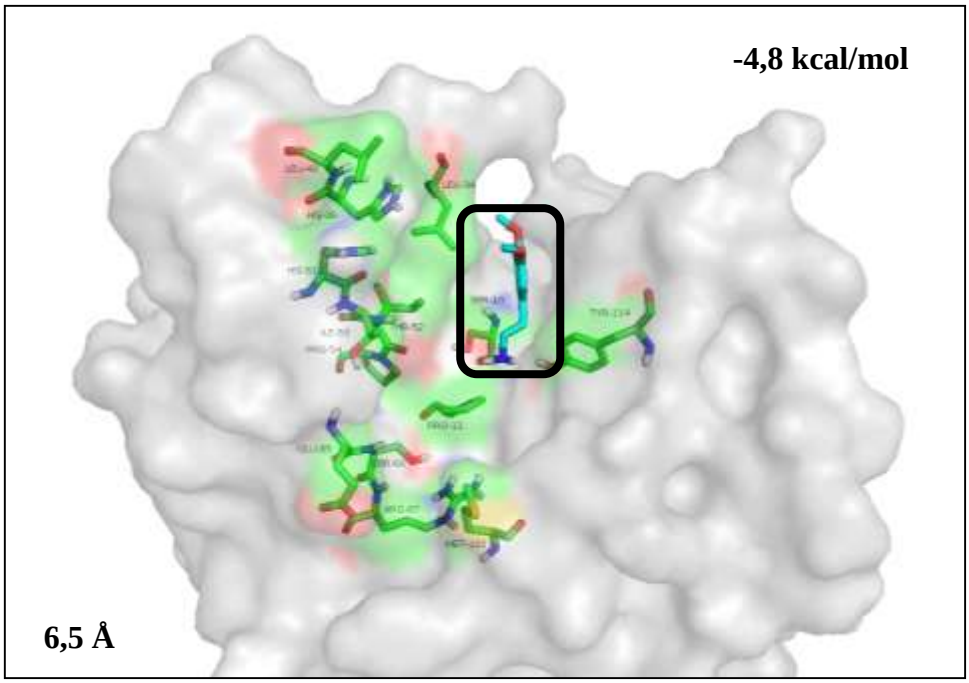


Fig. 14: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. buzzatii*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

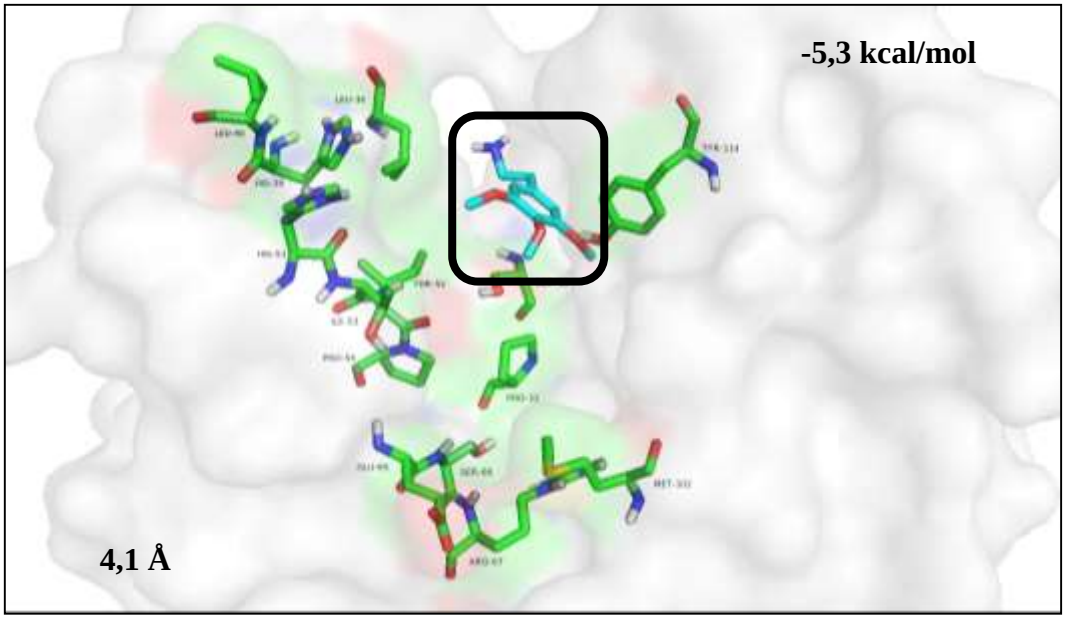


Fig. 15: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. borborema*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

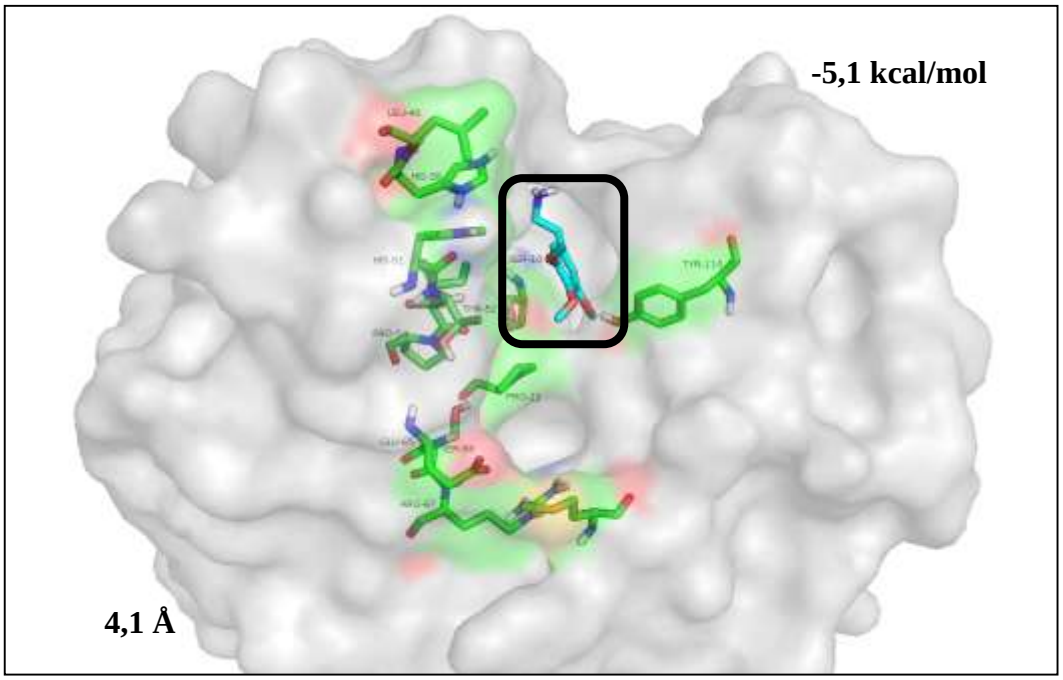


Fig. 16: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. serido*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

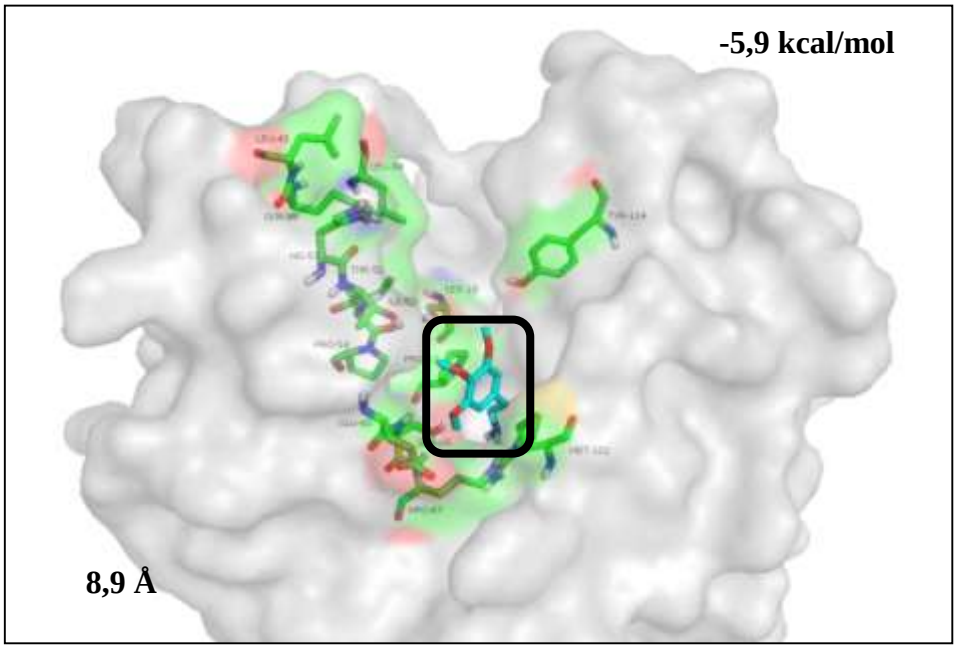


Fig. 17: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. seriema*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

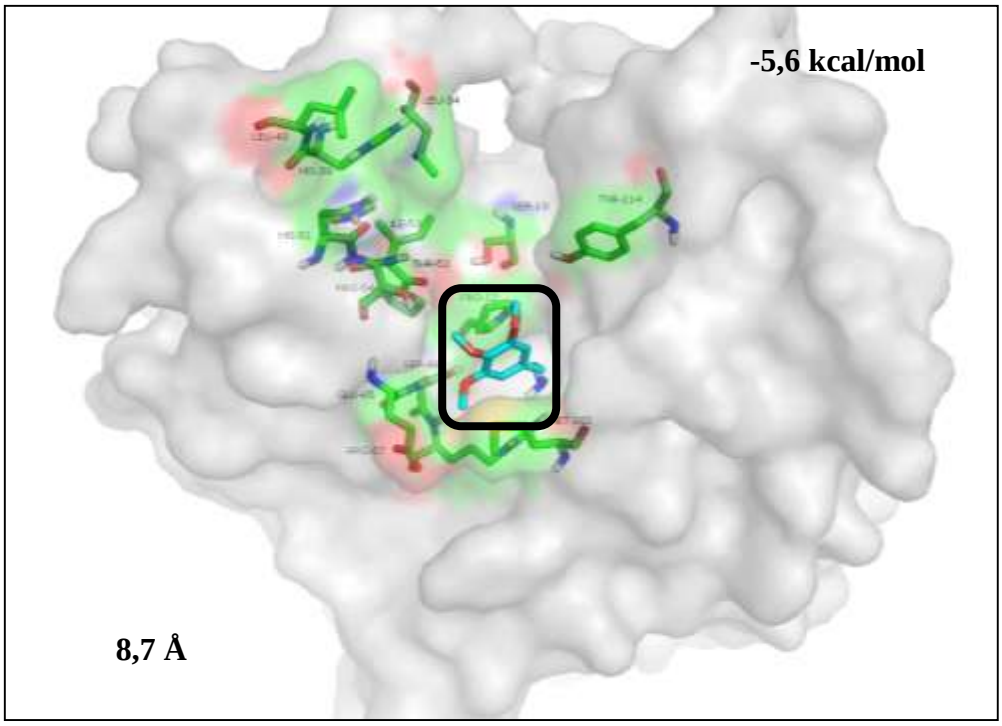


Fig. 18: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. koepferae*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

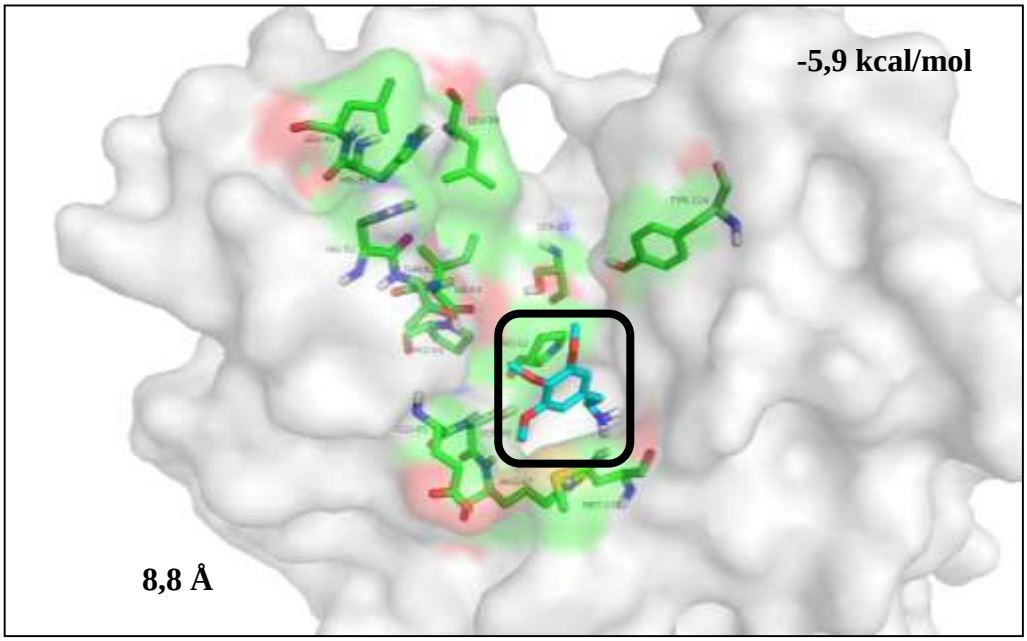


Fig. 19: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. gouveai*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

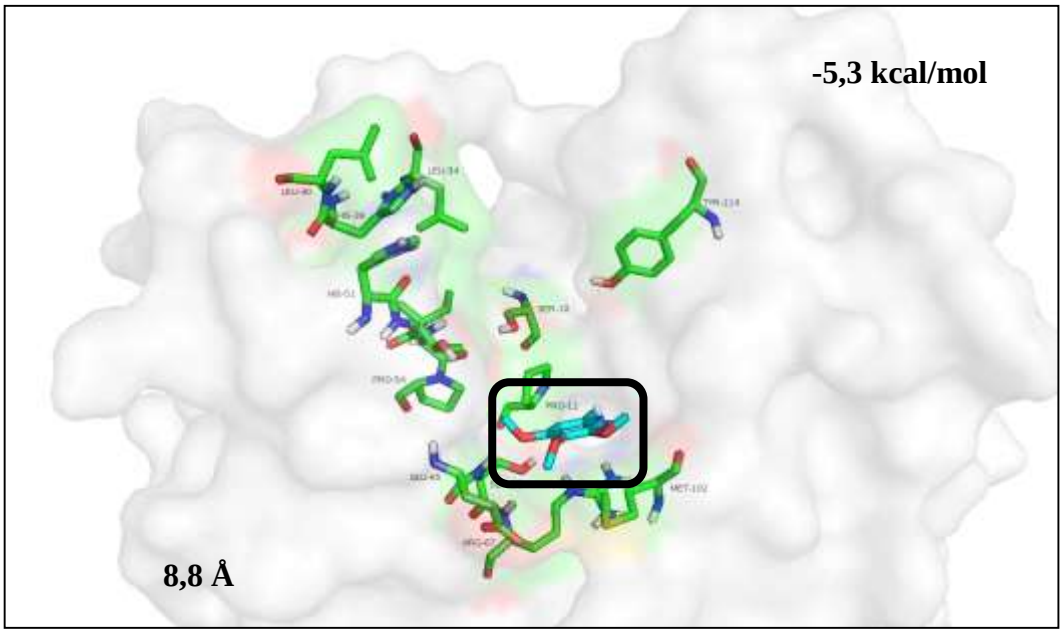


Fig. 20: Estrutura proteica monomérica de GSTD1 *docked* com mescalina representado por *sticks* em azul piscina e vermelho, para a espécie *D. antonietae*. O retângulo de cor preto evidência a estrutura do ligante mescalina. Os resíduos de aminoácidos do sítio-G estão representados por *sticks* em verde, vermelho e azul. O valor mensurado sobre o gasto energético está localizado no canto superior direito e distância entre alcaloide e serina 10, no canto inferior esquerdo.

4.7 Hidrofobicidade

A análise *in silico* de hidrofobicidade da proteína GSTD1 para as espécies do cluster *D. buzzatii* apresenta porções semelhantes de regiões hidrofóbicas e hidrofílicas, com evidência de maior composição de resíduos com propriedades hidrofóbicas no entorno do resíduo serina 10 (sítio-G) (Fig. 21 A - C). Para *D. seriema*, *D. gouveai*, *D. koepferae* e *D. antonietae* (Fig. 21 D - G), é observado um sutil aumento em proporções de regiões hidrofílicas próximo do resíduo de serina 10.

Para *D. seriema*, *D. gouveai*, *D. koepferae* e *D. antonietae* foi observado aumento de regiões com propriedades menos hidrofóbicas na parte basal, próximo do bolsão do sítio-G, da proteína GSTD1.

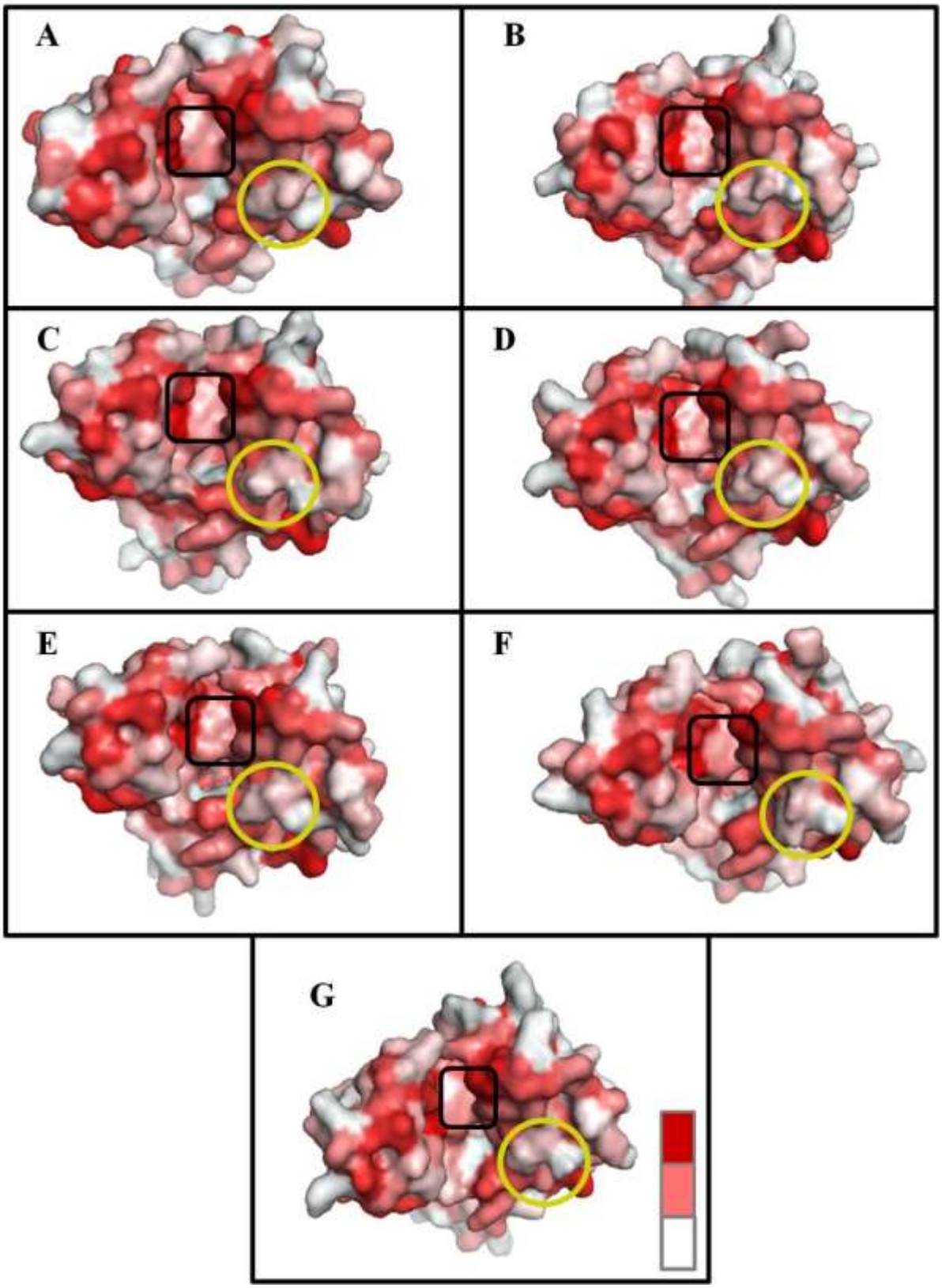


Fig. 21: Modelo *in silico* da superfície global da proteína GSTD1, representando regiões hidrofóbicas em vermelho, regiões com propriedades menos hidrofóbicas em rosa e regiões hidrofílicas em branco. As espécies do cluster *D. buzzatii* estão representadas em: A - *D. buzzatii*; B - *D. borborema*; C - *D. serido*; D - *D. seriema*; E - *D. koepferae*; F - *D. gouveai* e G - *D. antonietae*. O retângulo de cor preto corresponde a região do resíduo de serina 10. O círculo de cor amarelo corresponde a região com diferenças em propriedades hidrofóbicas, próximo do sítio-G, entre as espécies do cluster *D. buzzatii*.

4.8 Seleção Positiva e Alterações Estruturais Proteicas

Sinais de seleção positiva foram evidenciados em trincas de códons, nas posições 19, 39, 117, 133, 136, 196, 208 e 209, nos respectivos resíduos de aminoácidos da proteína GSTD1. Os sinais de seleção positiva na posição 39 são estatisticamente significantes, com $p<0,05$, entre *D. buzzatii*, *D. gouveai*, *D. koepferae*, *D. serido*, *D. borborema* e *D. antonietae*. Para *D. buzzatii* foi observado sinal de seleção positiva no códon que corresponde ao resíduo 209. As substituições em resíduos de aminoácidos nas posições 39 e 209 resultaram em alterações da estrutura química e das propriedades bioquímicas das proteínas de GSTD1 (Fig. 22). Essas informações sugerem que mutações nas sequências de DNA sob seleção positiva estão relacionadas com modificações em estruturas químicas da proteína GSTD1, entre as espécies cactofílicas do cluster *D. buzzatii*. Foram observadas mudanças nos resíduos de aminoácidos 19, 117, 133, 136 e 196 com assinatura de seleção positiva, entretanto estes não alteram as estruturas terciárias da proteína.

Para *D. koepferae*, foi observado sinal de seleção positiva no códon que corresponde ao resíduo de aminoácido 117, que apresentou uma substituição de Leu/Val, com modificação da estrutura química no sítio-H (Fig. 22 C). Para *D. koepferae*, o resíduo de valina na posição 117 representa um possível ganho nas ligações químicas com grupo de lipídios dentro do sítio-H.

O resíduo de aminoácido 209 foi o que demonstrou o maior número de alterações em grupos bioquímicos (Asp/Lys/Glu) na proteína GSTD1, entre as espécies do cluster *D. buzzatii* (Fig. 22). Este resíduo tem sido descrito como de suma importância para a estabilidade proteica durante as interações química (Low et al. 2010). Para o resíduo 39 não há relação com variações estruturais químicas para as espécies do cluster (Fig. 22 D).

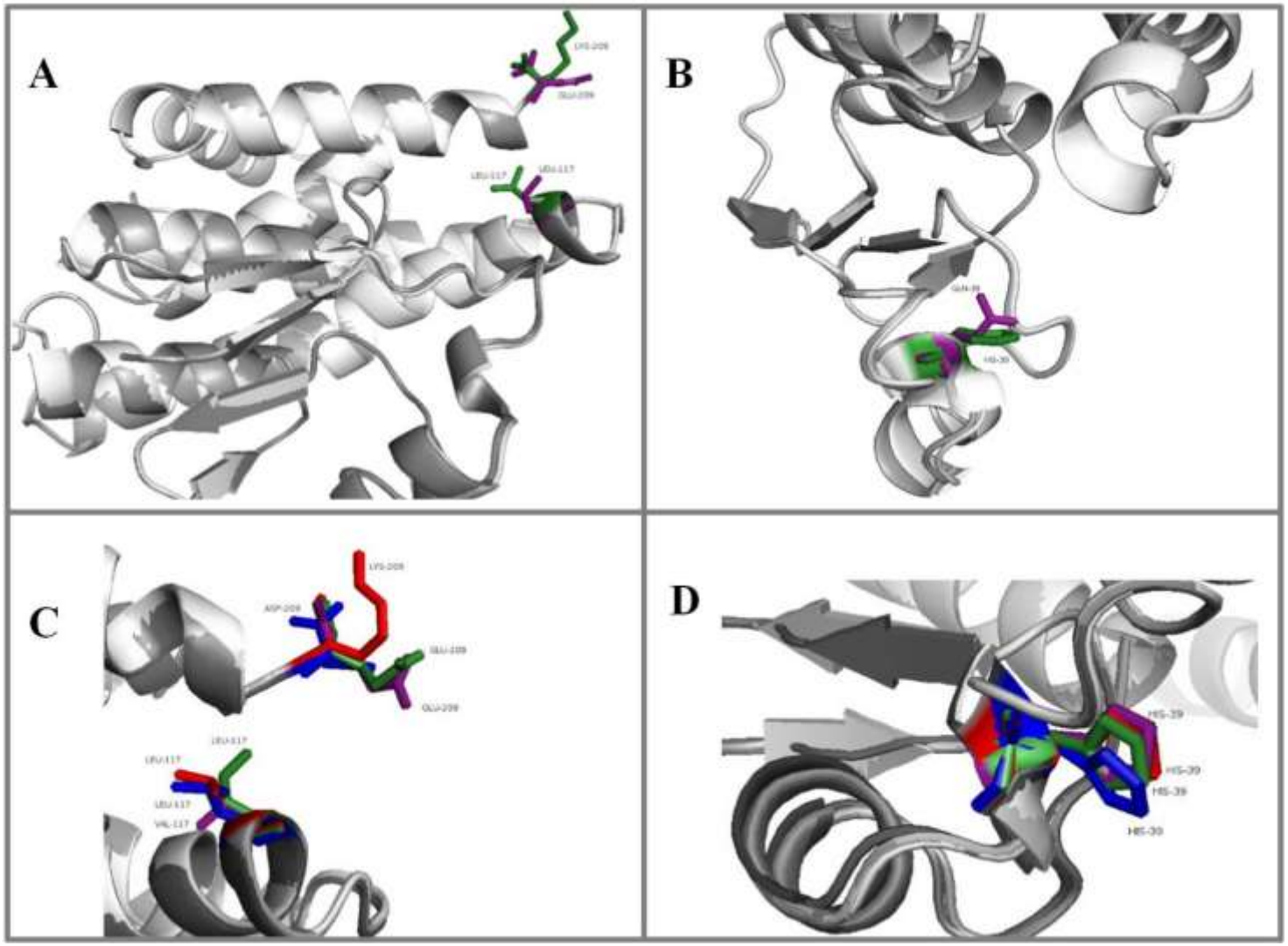


Fig. 22: Alinhamento das estruturas terciárias da proteína GSTD1 entre as espécies do cluster *D. buzzatii*. Elementos estruturais proteicos de *D. borborema*, estão representados em branco e verde, e para *D. seriema*, em cinza e violeta, com evidência para os resíduos de aminoácidos 117 e 209 (A) e 39 (B). Para *D. antonietae*, *D. gouveai*, *D. koepferae* e *D. buzzatii*, representados em vermelho, verde, violeta e azul, respectivamente, com evidência para os resíduos 117 e 209 (C) e 39 (D).

5. DISCUSSÃO

O gene *GSTD1* é reconhecido como um importante alvo de seleção e com funções relevantes no processo de adaptação ao uso de diferentes hospedeiros em espécies de *Drosophila* (Matzkin, 2008; 2014; Gloss et al. 2014; Matzkin, 2014; Panis et al. 2016). Com isso, diversos estudos têm buscado conhecer variantes em sequência de DNA e em resíduos de aminoácidos, com capacidade de melhorar a eficiência da proteína GSTD1 (Toung et al. 1993; Low et al. 2007, 2010; Saisawang et al. 2012; Matzkin, 2008, 2014; Vorojeikina et al. 2017). Nesse sentido, em *D. melanogaster* foram observados elementos regulatórios nas posições -674 a -667 (TGATACAG) e -2173 a -2166 (TGATTCAG) próximos do códon de iniciação de *GSTD1*, indicados como sinais específicos para respostas transcricionais ao composto de 12-0-tetradecanoilforbol-13-acetato (Toung et al. 1993). Por outro lado, substituição de glicina por lisina, no resíduo do aminoácido 171, resulta em aumento da eficiência da ligação da proteína GSTD1 com o pesticida dicloro-difenil-tricloroetano, em *D. melanogaster* (Low et al. 2007). *Drosophila arizonae* e *D. mojavensis* apresentam substituições de leucina por glutamina, no resíduo de aminoácido 7, posicionado na região do sítio-H, que pode ter consequências significativas na atividade e especificidade de GSTD1 ao reconhecimento de substratos (Matzkin, 2008). Neste trabalho, foi observado que as trocas de aminoácidos alteraram a estrutura química e energética na estrutura primária da proteína GSTD1, entre as espécies do cluster *D. buzzatii*. O gene *GSTD1* é relacionado a eventos de diversificação, com evolução semelhante aos padrões filogenéticos propostos com dados de inversões cromossômicas, entre as espécies do cluster. As substituições nas bases nucleotídicas 117 e 629 apresentam sinais de seleção positiva, com alterações em resíduos de aminoácidos, que resultam em alterações químicas e estruturais nas regiões do sítio ativo G e na cauda C-terminal. A estrutura terciária da proteína de GSTD1 de *D. seriema*, *D. gouveai* e *D. koepferae* apresenta melhor afinidade de ligação química entre o sítio-G de GSTD1 e o alcaloide mescalina.

5.1 Variações em sequências codificadoras e em resíduos de aminoácidos de GSTD1

As sequências do gene *GSTD1* são constituídas por 630pb, compreendendo a extensão total do único éxon, compostas por maiores proporções de bases CG do que AT, entre as espécies do cluster *D. buzzatii*. Alta proporção de conteúdo CG tem sido observada em outros genes nucleares, como, por exemplo, em *period* (Franco et al. 2010) e *Adh* (McDonal & Kreitman, 1991) e também na composição do DNA genômico de diversos organismos, com pequenas variações em proporções de bases CG relacionadas a fatores ambientais, tamanho do genoma (Foerstner et al. 2005), viés no uso de códons (revisão Castro-Chavez, 2011) e, por último, a tripla ligação química na interação CG, que proporciona maior estabilidade. Com efeito, as sequências de *GSTD1* são semelhantes entre as espécies do cluster, com pequenas variações em proporções de bases CG.

Na região C-terminal, foram observadas maiores quantidades de substituições não sinônimas do que sinônimas, entre as espécies do cluster *D. buzzatii*. Os dados são concordantes com as informações obtidas para *D. melanogaster*, em que há um maior número substituições não sinônimas na região C-terminal, que compreende parte do sítio ativo H, possuindo papel de seletividade no reconhecimento de substrato químico, e a região N-terminal compreende o sítio ativo G de GSTD1, que é indicada conservada (Saisawang et al. 2012). Em *Anopheles gambiae*, a região C-terminal possui alta divergência em sequências do gene *GSTD1*, com papel na especificidade na interação química com substratos (Rason et al. 1997). Portanto, a região C-terminal é mais variável em *Drosophila* e em outras espécies, com função de ligação do sítio-H com substratos químicos, por outro lado, pequenas alterações em resíduos na região N-terminal podem alterar a eficiência do sítio-G da proteína GSTD1, no processo de desintoxicação.

Para as sequências de *GSTD1*, no genoma sequenciado de *D. buzzatii* (Disponível em: <http://dbuz.uab.cat/welcome.php>), foi obtido um fragmento parcial de 275pb (código de acesso: lcl|scaffold_1_2:1214675-1214959) com alta similaridade em sequência nucleotídica com o gene *GSTD1*, que pode indicar um possível evento de duplicação do gene *GSTD1* em *D. buzzatii*. Os dados concordam com as evidências de duplicações gênica de *GSTD1* em *D. mojavensis* (Guillén & Ruiz, 2012) e *Sophophora nigrita* (Gloss et al. 2014). Contudo, em *S. nigrita* a duplicação de *GSTD1* não resulta em produto proteico funcional (Gloss et al. 2014).

O gene *GSTD1* está localizado no cromossomo 3 braço R, no meio do complexo Muller E (Consórcio *Drosophila* 12 Genomas), em *D. mojavensis* (Clarck et al. 2007), região com

provável recombinação gênica. No presente estudo, não foi possível analisar recombinação gênica, pelo fato de não ter sido amostrados representantes ao nível populacional. Por outro lado, estudo com populações de *D. mojavensis* (Sonora/México) e *D. arizonae* (Baja/México) evidencia maiores frequências de polimorfismos do que diferenças fixas em sequências do gene *GSTD1* (Matzkin, 2008). Isso dificulta a observação de um padrão de variação de ancestralidade, porém os dados populacionais rejeitam a hipótese de neutralidade e as diferenças fixas entre *D. arizonae* e *D. mojavensis* localizadas em regiões do sítio-H (Leu-7-Gln) e no sítio G (His-39-Gln), sugerem evolução adaptativa na proteína GSTD1 (Matzkin, 2008). Deste modo, as mudanças evolutivas em sequências de *GSTD1* envolvem uma combinação de alterações transcricional e codificadora, durante o processo de adaptação das populações (Matzkin, 2008).

Para as espécies de *Drosophila*, as sequências de mRNA do gene *GSTD1* são constituídas com maiores proporções de bases guaninas na 1° posição, adeninas na 2° posição e citosinas na 3° posição. Diante disso, no presente trabalho, algumas possibilidades de códons para *GSTD1* foram: GAN, N = representa posição do códon com alterações de bases, para resíduos de ácido aspártico e ácido glutâmico; GNC para resíduos de valina, ácido aspártico e glicina. Outras combinações encontradas em trincas de códons de *GSTD1* foram as seguintes: GUN para resíduos de valina, ANA para resíduos de isoleucina, treonina, lisina, arginina e serina. Contudo, a leitura de códons é sujeita ao reconhecimento químicos pelos tRNAs e podem ocorrer viés no processo tradução. A primeira e segunda posição dos códons são importantes para o reconhecimento pelos tRNAs e consequentemente na identificação dos resíduos de aminoácidos, porém a terceira base da trinca é mais suscetível a mudanças evolutivas e geralmente não interfere no reconhecimento da trinca de códons pelo tRNA específico, no processo de tradução e formação da proteína primária (revisão Castro-Chavez, 2011; Hoffmann et al. 2018). Em *D. melanogaster* foi observado maior quantidade de tRNAs para o reconhecimento do códon CGC que específica resíduos de arginina do que para os códons: CGU, CGA e CGG (revisão Castro-Chavez, 2011). Em vista disso, as substituições sinônimas podem influenciar na eficiência no reconhecimento de códons. Com isso, a maior proporção de substituições sinônimas encontradas em *GSTD1* pode influenciar na intensidade de leitura dos códons e na tradução das estruturas primárias da proteína no meio celular.

Na cadeia primária de *GSTD1* são evidenciadas variações em proporções de resíduos de aminoácidos, entre as espécies do cluster *D. buzzatii*. Fato este explicado por substituições não

sinônimas nas regiões C e N-terminal, que resultam de mudanças em trincas de códons e alteram a identificação dos resíduos de aminoácidos. Essas substituições em resíduos de aminoácidos na cadeia primária são importantes pelos seguintes motivos: capacidade de alterar as interações físicas e químicas entre átomos, dobramento das proteínas em fase terciária e, por fim, podem interferir na ligação química entre a enzima e os substratos (Grantham, 1974; Betts & Russell, 2003).

Para uma melhor compreensão da importância dessas variações, são necessárias destacar características importantes da proteína, como por exemplo, região celular, estrutura tridimensional e evolução (Betts & Russell, 2003; Huber et al. 2008). Sabe se que a proteína GSTD1 é encontrada em regiões intracelulares (Huber et al. 2008). Neste estudo, foram evidenciadas variações na composição de resíduos de aminoácidos de alanina, ácido aspártico, ácido glutâmico, histidina e lisina, da estrutura primária de GSTD1, entre os representantes do cluster *D. buzzatii*. Estas variações são importantes pois podem alterar a proteína em termos de estrutura e funções, na proteína GSTD1, devido as seguintes características dos resíduos de aminoácidos: (1) alanina possui carga apolar, com capacidade de especificidade na ligação com substrato, pela interação com outros átomos não reativos; (2) ácido aspártico e ácido glutâmico, carga polar negativa, em regiões do sítio ativo ou de interação entre proteínas possuem capacidade de ligação com outros átomos carregados positivamente; (3) histidina possui carga polar positiva, quando localizada em regiões do sítio ativo ou na superfície de proteínas apresenta propriedades de afinidade química com átomos de metais; (4) lisina apresenta carga polar positiva, com propriedade anfipática, com localização no sítio ativo, internamente na formação de pontes de sal e na superfície de proteínas (Betts & Russell, 2003). Deste modo, as variações encontradas em proporções de resíduos de aminoácidos da proteína GSTD1, entre as espécies do cluster *D. buzzatii*, podem influenciar na estrutura química e participam de processos essenciais de ligação e estabilidade da proteína GSTD1, em meio celular.

Por meio dos alinhamentos, foram observadas substituições em resíduos de aminoácidos, com mudança em propriedades bioquímicas, a saber, em *D. gouveai* (3 Asp/Ala) e *D. seriema* (39 His/Gln). Para *D. seriema*, a variação no resíduo 39 está localizada no sítio-G, que sugere alteração na função da proteína. Nesse sentido, dentre as variações encontradas em resíduos de aminoácidos, entre as espécies do cluster, a mudança no resíduo 39 de *D. seriema* é interessante, por ocorrer substituição entre histidina de carga polar positiva ($-CH_2-(N_2H_2C_3)-H$) para glutamina

de carga polar não-carregada (-$(CH_2)_2$-$CONH_2$). Isso implica em mudanças estruturais químicas dentro do sítio-G, com capacidade de alterar a ligação com substratos químicos. Além disso, substituições entre histidina e outros resíduos de aminoácidos são pouco frequentes, devido a suas propriedades bioquímicas, com flexibilidade de mover prótons para dentro ou fora da cadeia lateral, no processo de interação química (Betts & Russell, 2003).

Substituições nos resíduos 133 e 209 refletem em mudanças de categorias bioquímicas, compartilhadas entre algumas espécies do cluster *D. buzzatii* (ver Tabela 2). Curiosamente, o resíduo 209 apresentou variações entre resíduos de aminoácidos de carga positiva e negativa, com formação de três grupos, um constituído por *D. borborema*, *D. serido* e *D. seriema* (Asp/Lys), outro composto por *D. antonietae*, *D. koepferae* e *D. gouveai* (Asp/Glu) e, por último, *D. buzzatii* (Asp). Esse resíduo é interessante pelo fato de ser o último aminoácido da cauda C-terminal, com função relacionada à estabilidade da proteína e acessibilidade do sítio-G, que influência no processo de ligação a substratos (Low et al. 2010). Com isso, as mudanças em cargas energéticas do resíduo de aminoácido 209 podem modificar a eficiência da proteína GSTD1, entre as espécies do cluster *D. buzzatii*, no processo de desintoxicação, no uso de diferentes cactos hospedeiros.

Desta forma, o gene *GSTD1* apresenta importantes substituições únicas e compartilhadas em sequências de DNA e em resíduos de aminoácidos, com modificações em propriedades enzimáticas, entre as espécies do cluster *D. buzzatii*.

5.2 Inferências filogenéticas no cluster *D. buzzatii*

Variação genética na enzima *IDH* (*Isocitrato desidrogenase*) tem sido indicada como marcadores taxonômicos, entre as espécies do cluster *D. buzzatii* (Mateus et al. 2010). Com isso, mudanças em caracteres de dados proteicos são importantes para estabelecer relações de parentesco filogenético, pelo fato de ocorrer evolução relativamente lenta na constituição de proteínas em comparação com mudanças em sequências de nucleotídeos (Kumar & Singh, 2016), como foi observado no presente trabalho.

Nas relações filogenéticas inferidas a partir de análise individual dos genes *GSTD1* e *COI*, é sugerido o monofiletismo do cluster *D. buzzatii*, em concordância com o proposto em estudos citogenéticos e moleculares prévios (Ruiz & Wasserman, 1993; Rodríguez-Trelles et al. 2000; Manfrin et al. 2001; 2006; Santos et al. 2009; Franco et al. 2010). A topologia das árvores obtidas

pela análise do gene *GSTD1* e *COI*, com inferência Bayesiana, demonstra uma clara relação de parentesco entre as espécies cactofílicas *D. navojoa*, *D. arizonae*, *D. mojavensis* e o cluster *D. buzzatii*, como sugerido em hipóteses prévias (Oliveira et al. 2012).

Entre as espécies do cluster *D. buzzatii*, a hipótese de relações definidas pelo marcador *GSTD1* é semelhante à hipótese proposta com o gene mitocondrial *COI* (Manfrin et al. 2001). Para o marcador *GSTD1*, *D. koepferae* e *D. buzzatii* são indicadas espécies irmãs e mais próximas do ancestral do grupo, mas com baixo suporte estatístico. Por outro lado, o agrupamento entre *D. koepferae* e *D. buzzatii* está em concordância com o proposto por diversos marcadores moleculares, que reforçam o parentesco filogenético (Rodríguez-Trelles et al. 2000; revisão Manfrin et al. 2006; Santos et al. 2009; Franco et al. 2010).

Na análise Bayesiana, com os dados concatenados entre *GSTD1* e *COI*, *D. serido* e *D. antonietae* são agrupadas, indicadas espécies irmãs, em concordância com a inferência filogenética sugerida com *GSTD1*, porém *D. seriema*, *D. gouveai* e *D. borborema* apresentaram os mesmos padrões de ramificações do obtido com o marcador *COI*. As ramificações obtidas entre *D. gouveai*, *D. borborema* e *D. seriema* são discordantes entre o marcador nuclear *GSTD1* e mitocondrial *COI*. Essa discordância no estabelecimento das relações de parentesco entre essas espécies do cluster *D. buzzatii* se deve, principalmente, ao período recente de divergência, estimado ter ocorrido no período do Pleistoceno, cerca de 2,5 milhões de anos (Oliveira et al. 2012). Salienta-se ainda que, os marcadores nucleares e mitocondriais possuem diferentes características: (1) genes mitocondriais possuem maiores taxas de mutações do que os genes nucleares; (2) os mecanismos de reparo são menos eficientes em corrigir alteração no DNA mitocondrial do que no DNA nuclear; (3) alta taxa de fixação de mutações no mtDNA; (4) herança uni-parental do mtDNA e bi-parental do DNA nuclear (Wan et al. 2004; Mandal et al. 2014; Patwardhan et al. 2014). Deste modo, os marcadores *GSTD1* e *COI* apresentam diferentes padrões de evolução entre as espécies *D. gouveai*, *D. borborema* e *D. seriema*, indicadas com divergência mais recente dentro do cluster *D. buzzatii*, sem uma clara relação de parentesco.

Por outro lado, a organização interna de ramos na árvore filogenética BI obtida com *GSTD1* é corroborada pela hipótese filogenética proposta com dados de inversões cromossômicas ($2x^7$, $2e^8$, $2j^9$, 5g) (Ruiz et al. 2000; revisão Manfrin et al. 2006). No padrão de ramificação filogenético obtido com inversões cromossômicas, *D. serido* e *D. antonietae* são agrupadas, pelo compartilhamento da inversão fixa $2x^7$. *Drosophila borborema*, *D. seriema* e *D. gouveai* forma

um agrupamento interno, pela presença da inversão fixa de 2e[8]. *Drosophila koepferae* e *D. buzzatii* apresentam inversões fixas particulares, representadas por 2j[9] e 5g, respectivamente (Ruiz et al. 2000).

O padrão de ramificação obtido com os marcadores *GSTD1* e as inversões cromossômicas demonstram uma curiosa semelhança na estrutura de ramificação das árvores filogenéticas. Paralelamente, em *D. mojavensis* foi observado uma duplicação no gene *GSTD1*, com as formas *GSTD1a* e *GSTD1b*. *GSTD1a* foi reposicionada de sua localização na inversão cromossômica 2c para uma região do cromossomo 3, com implicação em mudanças na regulação da expressão gênica e na dinâmica evolutiva do gene (Guillén & Ruiz, 2012). *GSTD1a* apresenta maior expressão gênica do que *GSTD1b*, com consequência significativa na adaptação de linhagens de *D. mojavensis* e em espécies do complexo *mulleri*, ao uso de cactos hospedeiros (Guillén & Ruiz, 2012). No presente trabalho, as sequências do gene *GSTD1* obtidas para as espécies do cluster *D. buzzatii* correspondem a forma gênica *GSTD1*a de *D. mojavensis*. Para *D. buzzatii* ocorrem eventos de translocações, inversões e duplicações em regiões dos cromossomos autossômicos 2 e 3 (Naveira & Fontdevila, 1985), que compreende regiões genômicas de *GSTD1*. Embora especulativo, uma explicação para a semelhança no padrão das inferências filogenéticas com os dados dos marcadores *GSTD1* e inversões cromossômicas, pode ser elucidado pelos seguintes fatores: posição genômica do gene *GSTD1a*, localizada no cromossomo 3; compartilhamento de sinais moleculares presentes no ancestral do cluster *D. buzzatii*; e sob diferentes sinais de pressões seletivas, durante os processos de adaptação e evolução das espécies de *Drosophila* (Naveira & Fontdevila, 1985; Guillén & Ruiz, 2012), e no cluster *D. buzzatii*.

Uma segunda explicação sobre a dificuldade no estabelecimento da relação de parentesco filogenético entre *D. gouveai*, *D. borborema* e *D. seriema*, com os marcadores *GSTD1* (presente trabalho), *COI* (Manfrin et al. 2001; presente estudo), *period* (Franco et al. 2010) e DNA satélite (Kuhn, 2003), pode estar relacionado a processos de introgressão entre as espécies do cluster. De forma geral, os eventos de introgressão podem levar a elaboração de hipóteses filogenéticas distintas, devido ao mecanismo de herança genética entre diferentes linhagens evolutivas, com diferentes conjuntos gênicos, sob diferentes pressões evolutivas. No cluster *D. buzzatii*, tem sido observado eventos de introgressão entre *D. buzzatii* e *D. koepferae* (Soto et al. 2007), *D. serido* e *D. antonietae* (Kokudai et al. 2011) e *D. gouveai* e *D. antonietae* (Soto et al. 2010).

Por último, porém não menos importante, tem sido sugerido que a divergência e evolução das espécies cactofílicas do grupo *Drosophila repleta* ocorreu com a capacidade de explorar diferentes hospedeiros (Oliveira et al. 2012). As espécies cactofílicas de *Drosophila* passaram por eventos de adaptação, sob diferentes processos de evolução, ao uso de cactos *Opuntia*, considerados com composição química menos complexo do que os cactos colunares (Oliveira et al. 2012). De forma geral, os padrões de divergências entre as espécies do cluster *D. buzzatii*, envolvem fatores de rearranjos genômicos, eventos de introgressão e processos de adaptação ao uso de cactos.

5.3 Teste de seleção em sequências do gene *GSTD1*

Sinais de seleção positiva sugeridos para as bases nucleotídicas 117 e 625 resultam em alterações dos resíduos de aminoácidos 39 e 209, que levam a alterações químicas e estruturais nas regiões do sítio-G em *D. buzzatii*, *D. gouveai*, *D. koepferae*, *D. serido*, *D. borborema* e *D. antonietae*, e na cauda C-terminal, em *D. buzzatii*. Populações de *D. mojavensis* e *D. arizonae* (Baja e Sonora/México) apresentaram mudanças em resíduos de aminoácidos da proteína GSTD1, com substituições no resíduo 39, entre histidina e glutamina (Matzkin, 2008), em concordância com o presente estudo. Essa alteração no resíduo de aminoácido modifica a carga energética dentro do sítio-G, de polar positivo (histidina) para polar não carregado (glutamina). As substituições alteram o peso molecular e o volume médio dos resíduos, entre histidina (155Da; 166.3Å) e glutamina (146Da; 148.0Å) (Richards, 1977). Além disso, o ponto isoelétrico é diferente entre histidina (7.4) e glutamina (5.6), em um pH fisiológico quase neutro. Em síntese, o sinal de seleção positiva na base nucleotídica que muda o resíduo 39 altera a estrutura bioquímica, a polaridade química e pode melhorar a especificidade no reconhecimento do substrato no sítio-G da proteína GSTD1, entre as espécies do cluster *D. buzzatii*.

Em trabalho com enfoque em biologia estrutural e funcional da proteína GSTD1 em *D. melanogaster*, o resíduo 39 é indicado como importante para ligações polares com porções de glicina, na formação de pontes de sal e interações químicas com regiões de GSH (glutationa redutase) (Low et al. 2010). Com isso, a substituição nesse resíduo pode influenciar a estabilidade de ligação química interna na proteína GSTD1.

Foi observado em *D. melanogaster*, que o resíduo de glutamina, na posição 209 da cauda C-terminal, realiza interações com GSH, o que é necessário para a estabilidade da cauda (Low et

al. 2010). Por outro lado, a composição da cauda C-terminal nas espécies do cluster *D. buzzatii* difere do observado para *D. melanogaster* (209 - Glu), apresentando resíduos de ácido aspártico e ácido glutâmico, polares negativos, para *D. buzzatii*, *D. gouveai*, *D. koepferae* e *D. seriema*. Porém, para *D. serido*, *D. borborema* e *D. antonietae*, o resíduo 209 é representado por lisina, de carga polar positiva. Com isso, as mudanças nesses dois grupos de carga energética podem refletir na interação química com GSH, na ligação com glutationa, nas posições 51-53 e 65-67. Portanto, o sinal de seleção positiva no nucleotídeo 627 que está relacionado à mudança no aminoácido 209, em *D. buzzatii*, pode estar favorecendo a estabilidade da cauda C-terminal e possivelmente aumenta a capacidade de interação entre o sítio-G e o substrato químico.

Para *D. koepferae* e *D. serido* ocorreram alterações na estrutura química do resíduo do aminoácido 117 (Fig. 4; Tabela 3), com substituição de Leu/Val. Em *Drosophila hexastigma* foram observadas substituições no resíduo 116 de Glu/Tre e no resíduo 117 de Met/Leu (López-Olmos et al. 2017). A substituição no resíduo 116 está posicionado próximo ao sítio-H, e o resíduo 117 está posicionado dentro do sítio-H. As mudanças em resíduos de aminoácidos da proteína GSTD1 em *D. hexastigma* sugerem funções relacionadas a adaptação ao uso de extratos de diferentes espécies de cactos (López-Olmos et al. 2017). Para *D. koepferae* e *D. serido*, o resíduo de valina na posição 117 representa um possível ganho nas ligações químicas com grupo de lipídios dentro do sítio-H.

Por outro lado, sinais de seleção positiva no resíduo 171 (Gly/Lys), entre as espécies de *Drosophila*, têm sido sugeridos como uma resposta adaptativa a toxina ambiental DDT (dicloro-difenil-tricloroetano) (Low et al. 2007). Todavia, em *D. melanogaster*, por ressonância magnética, tem sido observado mudanças e deslocamentos estruturais nos resíduos de aminoácidos 170 e 172, sugeridos como importantes para a ligação química de GSTD1 com DDT (Low et al. 2010). Nos dados obtidos para as espécies do cluster *D. buzzatii*, não foram sugeridos sinais de seleção positiva nos códons que resultam nos resíduos de aminoácidos nas posições 170, 171 e 172. Salienta-se que substituições de glicina por lisina no resíduo 171 (Gly/Lys), presente em *D. melanogaster*, *D. pseudoobscura*, *D. persimilis* e *D. mojavenis* podem ter ocorrido pelo menos três vezes na radiação dentro do gênero *Drosophila* (Low et al. 2007). Essas substituições requerem duas alterações em posições de códons (GGN/AAR), o que resulta em mudança de resíduos de aminoácidos em categoria bioquímica. Com isso, a função de ligação do resíduo do aminoácido 171 a substratos químicos, no processo de desintoxicação, apresenta caráter

conservado entre as linhagens evolutivas mais recentes, que abrange os representantes do cluster *D. buzzatii*.

Sinais de seleção purificadora foram detectadas em nucleotídeos nas posições 587 e 623 que refletem em mudanças nos resíduos de aminoácidos nas posições 196 e 208, localizados na região C-terminal, na hélice α8 e na cauda da proteína, respectivamente, próximos do sítio-G. Os dados sugerem que essas substituições em resíduos de aminoácidos estão sendo eliminados, devido a capacidade de diminuir a eficiência da proteína GSTD1, no processo de adaptação proteica entre as espécies de cluster. As substituições nos resíduos 196 (Glu/Asp) e 208 (Phe/Leu) ocorrem dentro das mesmas categorias energéticas, sendo polar negativa e apolar, respectivamente. Porém, as substituições em aminoácidos alteram o peso e o volume médio de área, entre os resíduos de ácido aspártico (133Da; 122.0Å) e ácido glutâmico (147Da; 143.9Å) (Richards, 1977), e entre fenilalanina (165Da; 189.9Å) e leucina (131Da; 166.7Å) (Kim lab, 2004). O ponto isoelétrico é diferente entre ácido glutâmico (3.22) e ácido aspártico (2.77), e fenilalanina (5.48) e leucina (5.98), em um pH fisiológico quase neutro. Com isso, essas alterações refletem em variações na estrutura química, da cadeia lateral na α8, próximo do sítio-G. Em síntese, a proteína GSTD1 está em processo de dinâmica evolutiva, com alterações nos resíduos dos aminoácidos 196 e 208, com reajustes que podem não favorecer o processo de adaptação da proteica no reconhecimento de substratos químicos.

Os sinais de seleção em nucleotídeos que estão relacionados a aminoácidos em posições localizados no sítio-G, na cauda C-terminal e na estrutura interna (ver Fig. 10) formam um conjunto de alterações, em composição de resíduos de aminoácidos, estrutura química, carga energética e propriedades bioquímicas, que podem alterar as interações químicas de GSTD1 com um substrato químico e aumentar sua eficiência, no processo de desintoxicação. As espécies do cluster *D. buzzatii* possuem desenvolvimento larval em tecidos necrosados de cactos, que são constituídos por um ambiente químico complexo, como por exemplo, alcaloides, hidrocarbonetos, esteroides, ácidos graxos, glicosídeos triterpenos (Stintzing & Carle, 2005; Brito-Filho, 2015) e produtos de interação com microrganismos (Fogleman & Danielson, 2001). A exploração de recursos no gênero deve ter implicado em ajustes adaptativos diferenciais pelas espécies de *Drosophila* (Kircher et al. 1967; Matzkin,2008, 2014; Panis et al. 2016; Lópes-Olmos et al. 2017). Entre os compostos potencialmente tóxicos, os alcaloides específicos presentes em cada espécie de cacto têm sido indicados como um dos fatores que moldam o processo de

adaptação das espécies de *Drosophila* (Kircher et al. 1967; Matzkin, 2014; Panis et al. 2016; Lópes-Olmos et al. 2017). Com isso, essas variações na proteína GSTD1, com sua função de desintoxicação, podem representar elementos importantes na dinâmica de interações químicas da proteína com os substratos em que as larvas de *Drosophila* se desenvolvem.

Nesse sentido, é de grande importância os dados de interações moleculares da proteína GSTD1 com alcaloides, com o objetivo de entender os mecanismos da biologia estrutural da proteína e evolução, entre as espécies cactofílicas de *Drosophila*.

5.4 Modelagem *in silico* e análise da proteína GSTD1

No presente estudo, as estruturas primárias, secundárias e terciárias das proteínas de GSTD1 foram obtidas por modelagem *in silico*, por meio de análises de bioinformática, devido à grande disponibilidade de ferramentas e flexibilidade em análise de estrutura, predição de funções proteicas e interações entre moléculas químicas (Gupta et al. 2014; Singh et al. 2016). Por outro lado, técnicas de determinação da estrutura 3D de proteínas nativas, como por exemplo, ressonância magnética e cristalografia de raios-X, fornecem dados importantes sobre macromoléculas proteicas, porém são muito caras e trabalhosas (Schmidt & Lamzin, 2002). Porém, as metodologias de ressonância magnética e cristalografia de raios-X possuem limitações no processo de purificação, concentração em miligrama do material e manutenção de complexos proteicos nativos para a obtenção de cristais. Nesse sentido, tem sido empregado o uso de previsões computacionais para elucidar estruturas de proteínas (Aloy & Russell, 2006).

Foram realizadas modelagens *in silico* por meio do método de homologia, usando as sequências da estrutura primária de GSTD1 de cada espécie do cluster *D. buzzatii*. Para as modelagens por homologia, foi usado um modelo prévio 3D da proteína GSTD1 de *D. melanogaster*, disponível no acesso PDB: 3MAK, devido a sua estrutura apresentar homologia ≥ 50 % (Sali & Blundell, 1993) com as espécies estudadas.

O método por homologia ou comparativa possui grande confiabilidade na predição de estruturas proteicas, com erros limitados em cadeias laterais e posicionamento de *loops* (Gupta et al. 2014). No presente estudo, as modelagens por homologia da proteína GSTD1 foram realizadas com o modelo de *D. melanogaster* para modelar a proteína de *D. mojavensis* e, posteriormente, foi usado o modelo dessa última espécie para obter os modelos proteicos dos representantes do cluster *D. buzzatii*.

Foram obtidos os modelos de proteínas terciárias de GSTD1 para cada uma das espécies do cluster *D. buzzatii* (Fig. 14-20). As proteínas de GSTD1 obtidas são semelhantes à estrutura cristalizada descrita para *D. melanogaster* (Low et al. 2010; Wongsantichon et al. 2012). Porém, a proteína GSTD1 das espécies do cluster apresenta nove α-hélices e em *D. melanogaster* é constituída por oito α-hélices. Por outro lado, nos alinhamentos da estrutura terciária de GSTD1 obtidos entre as espécies do cluster *D. buzzatii* com *D. mojavensis*, *D. arizonae* e *D. navojoa*, foi observada grande similaridade entre os elementos estruturais (Fig. 12). Isso indica um estado conservado na composição de elementos estruturais da proteína GSTD1 entre as espécies de *Drosophila* cactofílicas, que pode sugerir uma manutenção de sua função e interações moleculares, no processo de desintoxicação. Contudo, as mudanças na proteína GSTD1 estão localizadas em resíduos de aminoácidos particulares nas regiões do sítio-G, sítio-H e cauda C-terminal, que representam regiões importantes para a atividade e eficiência da proteína, em sua função enzimática.

No *docking* da proteína GSTD1 com o ligante mescalina, foi observado diferenças em posições de ligação, entre as espécies do cluster (Fig. 14-20). As proteínas terciárias de GSTD1 de *D. seriema*, *D. gouveai* e *D. koepferae* apresentaram os melhores valores de afinidade de ligação química com o ligante mescalina, em termos de gasto energético (Fig. 14-20). A composição de resíduos de aminoácidos internos do sítio-G é idêntica entre as espécies, nas posições 65-67, 102, 12, 39, 40, 34, 10, 114 e 51-54, exceto para *D. seriema*, que apresentou uma substituição na posição 39, de histidina para glutamina, com alteração de carga polar positivo para polar não carregado. De modo geral, a carga energética dentro do sítio-G é conservada entre as espécies do cluster *D. buzzatii*.

O *docking* molecular da proteína GSTD1 de *D. koepferae* com o ligante mescalina, é indicado como um dos melhores modelos de afinidade química, entre as espécies do cluster *D. buzzatii*. *Drosophila koepferae* se desenvolve em tecidos necróticos de *Trichocereus terscheckii* (revisão Manfrin & Sene, 2006), com extratos compostos por grandes frações de alcaloides tricocereina e mescalina (Panis et al. 2016). Esses alcaloides têm capacidade de afetar caracteres de desempenho, viabilidade de larvas para adultos e tamanho do corpo em *D. buzzatii* (Soto et al. 2014). Desta forma, os resultados do *docking* molecular apresentam relações com os dados biológicos, em que *D. koepferae* apresenta melhor afinidade na interação química entre GSTD1 e o ligante mescalina, composto presente em extratos de *T. terscheckii*, cacto indicado como

hospedeiro primário para a espécie (Soto et al. 2014). Para *D. buzzatii* foi observado maior gasto energético, no processo de ligação da proteína GSTD1 ao ligante mescalina. Esse dado indica que a proteína GSTD1 de *D. buzzatii* apresenta menor capacidade de interação química com o alcaloide mescalina e possivelmente menor eficiência no processo de desintoxicação. Nesse sentido, os dados podem indicar que a proteína das espécies *D. seriema*, *D. gouveai* e *D. koepferae* possuem melhor ligação química com mescalina e podem favorecer os processos de desintoxicação.

O ligante mescalina é posicionado na parte basal do bolsão da proteína GSTD1, distante do resíduo de serina 10, em *D. seriema*, *D. gouveai*, *D. koepferae* e *D. antonietae*. No entanto, o ligante está posicionado entre dois sítios de ligação GSH, nas posições 51-53 e 65-67. Em síntese, os modelos das proteínas de GSTD1 apresentaram os melhores valores de afinidade química com o ligante mescalina quando posicionado entre os dois sítios GSH, entre as espécies de *Drosophila*.

O ângulo de posicionamento do ligante mescalina no bolsão do sítio-G, da proteína GSTD1, entre as espécies do cluster (Fig. 14-20), pode ser explicado pelo tamanho da estrutura química de mescalina, flexibilidade do ligante, capacidade de ligações químicas com as pontes de hidrogênio (Forli et al. 2016) ou de interação com forças de Van der Waals, entre a estrutura química de mescalina e o sítio-G.

Outro fator que influência nas interações moleculares, entre proteínas e substratos químicos, é a localização de regiões hidrofóbicas e hidrofílicas na superfície da estrutura proteica. O mapeamento de manchas em regiões da superfície da proteína GSTD1 possibilitou identificar as regiões de acessibilidade a solvente e o seu papel no processo de reconhecimento intra e intermolecular (Eisenhaber, 1996, 1999). Em vista disso, uma explicação sobre as diferenças de ligação química do sítio-G, por meio do *docking*, pode ser explicada por diferenças em porções de regiões hidrofóbicas e hidrofílicas, próximo ao sítio-G. No presente estudo, foi observado maior proporção de regiões com propriedades menos hidrofóbicas próximo do sítio-G em *D. seriema*, *D. koepferae*, *D. gouveai* e *D. antonietae*. Por outro lado, em *D. buzzatii*, *D. borborema* e *D. serido* foi observado maior intensidade de regiões hidrofóbicas próximo do resíduo de serina 10 e na parte basal do sítio-G. Portanto, a proporção de regiões hidrofóbicas no sítio-G é um dos fatores que influência a ligação química com o ligante mescalina.

A superfície de acesso a solvente na proteína GSTD1 apresenta manchas hidrofílicas isoladas no entorno do sítio-G, entre as espécies do cluster *D. buzzatii*. Uma exceção é a proteína terciária de *D. antonietae* que apresentou maior proporção de manchas hidrofílicas (Fig. 21-G). Portanto, a proteína GSTD1 de *D. antonietae* possui maior capacidade de interações com átomos de água, disponíveis em um substrato químico.

Em resumo, as proporções de manchas hidrofóbicas e hidrofílicas na proteína GSTD1, entre as espécies do cluster *D. buzzatii*, indicam diferenças nas interações químicas das proteínas com substâncias químicas que constituem os substratos e, possivelmente, aos alcaloides no processo de desintoxicação.

Paralelamente, as mudanças no resíduo 209 podem tornar mais disponíveis o sítio-G, nas interações químicas. Salienta-se ainda que, para uma maior compreensão das interações químicas do ligante mescalina ao sítio-G e a estabilidade da proteína GSTD1, faz-se necessário realizar análises bioinformáticas de dinâmica molecular da proteína GSTD1 com o alcaloide mescalina.

Uma sugestão alternativa, sobre a eficiência da proteína GSTD1, é que a quantidade de proteínas disponíveis nas células refletem na eficiência do processo de desintoxicação metabólica. Com isso, os fatores de regulação de expressão gênica como, por exemplo, as marcas epigenéticas, podem ajudar a elucidar sobre a quantidade de transcritos e proteínas disponíveis no meio celular. Um vez que para *D. buzzatii* foi evidenciado expressão gênica diferencial de *GSTD1* em larvas cultivadas em diferentes substratos de cactos (Panis et al. 2016).

Em síntese, o processo de desintoxicação realizado pela proteína GSTD1 envolve vários fatores moleculares. Substituições não sinônimas ocorrem em menor proporção em resíduos de aminoácidos da proteína GSTD1, entre as espécies do cluster *D. buzzatii*, porém alteram propriedades bioquímicas, volume, área, carga energética e porções de regiões hidrofóbicas, que são importantes nas interações moleculares da proteína GSTD1 com alcaloides.

6. CONCLUSÕES

A hipótese formulada foi testada e todos os objetivos propostos foram atingidos.

No presente trabalho, com base nos resultados obtidos, podemos tirar algumas importantes conclusões sobre a evolução de GSTD1, nas espécies do cluster *D. buzzatii*:

- Os códons do gene *GSTD1* apresentaram heterogeneidade de pressões seletivas, com acentuada pressão de seleção positiva em códons correspondentes a região C-terminal da proteína, que pode sugerir um ganho de interações com substratos químicos.
- O marcador *GSTD1* possui sinal filogenético que sugerem a monofilia das espécies do cluster *D. buzzatii*. *Drosophila antonietae* e *D. serido* são indicadas espécies irmãs e próximas do agrupamento composto por *D. gouveai*, *D. borborema* e *D. seriema*.
- Sinal de seleção positiva em códons do resíduo do aminoácido 39 de GSTD1, em *D. koepferae*, *D. serido*, *D. borborema* e *D. antonietae* sugerem ganho de função de ligação química do sítio-G com o alcaloide mescalina. Para *D. koepferae* e *D. serido*, a substituição de leucina por valina, na posição 117, pode indicar um ganho na interação com lipídios dentro do sítio-H da proteína GSTD1.
- Proteínas terciárias de GSTD1 em *D. seriema*, *D. gouveai* e *D. koepferae* apresentaram as melhores capacidades de ligação química com o ligante mescalina, que pode estar relacionado a posição entre duas regiões de ligação GSH (glutationa redutase) e próximo de uma porção proteica com propriedades menos hidrofóbicas.
- O processo de desintoxicação realizado pela proteína GSTD1, entre as espécies do cluster *D. buzzatii*, envolve mudanças bioquímicas, em estruturas químicas e energéticas em resíduos de aminoácidos, posicionados dentro e próximo dos sítios-G e H.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ALLOCATI, N.; FEDERICI, L.; MASULLI, M.; DI-LLIO, C. Glutathione transferases in bacteria. *FEBS Journal*, v.276, n.1, p.58-75, 2009.

ALOY, P.; RUSSELL, R.B. Structural systems biology: modelling protein interactions. *Nat. Rev. Mol. Cell Biol.*, v.7, p.188-197, 2006.

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, v.215, p.401-410, 1990.

ARIVALAGAN, P.; DHANARANI, S.; SHANKAR, C.; PIRUTHIVIRAJ, P.; RANGANATHAN, K.; SARATALE, R.G.; THAMARAISELVI, K. Electrophoretic pattern of glutathione S-transferase (GST) in antibiotic resistance gram-positive bacteria from poultry litter. *Microbial Pathogenesis*, v.10, n.1, p.285-290, 2017.

BAKER, W.C.; GARAVELLI, J.S.; HUANG, H.; MCGARVEY, P.B.; ORCUTT, B.C.; SRINIVASARAO, G.Y.; XIAO, C.; YEH, L.S.; LEDLEY, R.S.; JANDA, J.F.; PFEIFFER, F.; MEWES, H.W.; TSUGITA, A.; WU, C. The protein information resource (PIR). *Nucleic Acids Research*, v.28, n.1, p.41-44, 2000.

BARKER, J.S.F.; SENE, F.M.; EAST, P.D.; PEREIRA, M.A.Q.R. Allozyme and chromosomal polymorphism of *Drosophila buzzatii* in Brazil and Argentina. *Genetica*, v.67, n.1, p.161-170, 1985.

BENSON, D.A.; CAVANAUGH, M.; CLARCK, K.; KARSCH-MIZRACHI, I.; LIPMAN, D.J.; OSTELL, J.; SAYERS, E.W. GenBank. *Nucleic Acids Research*, v.8, n.3, p.186-194, 2013.

BETTS, M.J.; RUSSELL, R.B. Amino acid properties and consequences of substitutions. In: BARNERS, M.R.; GRAY, I.C. (eds) **Bioinformatics for Geneticists.** Wiley, New York, 2003, cap.14, p.289-316.

BIANCHI, A.; SAMORINI, G. **Plants in association with Ayahuasca**. Berlim: Yearbook f. Ethnomedicine, 1993. 24p.

BRITO-FILHO, S.G.DE. **Estudo fitoquímico e atividade antimicrobiana de *Pilosocereus pachycladus* F.Ritter (Cactaceae).** João Pessoa, 2015. 151p. Tese de Doutorado – Departamento de Produtos naturais e sintéticos bioativos – Universidade Federal da Paraíba.

CASTRO-CHAVEZ, F. Most used codons per amino acid and per genome in the code of man compared to other organisms according to the rotating circular genetic code. *Neuroquantology*, v.9, n.4, p.1-30, 2011.

CLARCK, A. G. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, v. 450, n.8, p.203-2018, 2007.

DAVET, A. **Estudo fitoquímico e biológico do cacto – *Cereus jamacaru* De Candolle, Cactaceae.** Curitiba, 2005. 100p. Dissertação de Mestrado – Departamento de Ciências Farmacêuticas – Universidade Federal do Paraná.

DEBRITO, R.A.; MANFRIN, M.H.; SENE, F.M. Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genetics and Molecular Biology*, v.25, n.2, p.161-171, 2002.

DELANO, W. The PyMOL Molecular Graphics System. Disponível em: <http://www.pymol.org> Acesso em: 20 jan 2018.

DIXON, D.P.; LAPTHORN, A.; EDWARDS, R. Plant glutathione transferases. *Genome Biology*, v.3, n.3, p.1-10, 2002.

DONNARUMMA, L.; DE-ANGELIS, G.; GRAMENZI, F.; VITTOZZI, L. Xenobiotic metabolizing enzyme systems in test fish: III. Comparative studies of liver cytosolic Glutathione S-transferases. *Ecotoxicology and environmental safety*, v.16, n.1, p.180-186, 1988.

DORON-FAIGENBOIM, A.; PUPKO, T.A combined empirical and mechanistic codon model. *Mol. Biol. Evol.*, v.24, p.388-397, 2007.

DORON-FAIGENBOIM, A.; STERN, A.; MAYROSE, I.; BACHARACH, E.; PUPKO, T. Selection: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics*, v.21, n.9, p.2101-2103, 2005.

EISENHABER, F. Hydrophobic regions on protein surfaces. Derivation of the salvation energy from their area distribution in crystallographic protein structures. *Protein Sci*., v.5, n.8, p.1676-1686, 1996.

EISENHABER, F. Hydrophobic regions on protein surfaces. *Perspectives in Drug Discovery and Desing*, v.17, p.27-42, 1999.

EWING, B.; GREEN, P. Base-Calling of automated sequencer traces using *Phred*. II. Error probabilities. *Genome Research*, n.8, p. 186-194, 1998.

FOERSTNER, K.U.; VON-MERING, C.; HOOPER, S.D.; BORK, P. Environments shape the nucleotide composition of genomes. *EMBO Rep.*, v.6, n.12, p.1208-1213, 2005.

FONTDEVILA, A.; RUIZ, A.; ALONSO, G.; OCAÑA, J. Evolutionary history of *Drosophila buzzatii*. I. Natural chromosomal polymorphism in colonized populations of the Old World. *Evolution*, v.35, n.1, p.148-157, 1981.

FOGLEMAN, J.C.; DANIELSON, P.B. Chemical interactions in the cactos – microorganism – *Drosophila* model system of the Sonoran desert. *Amer. Zool*., v.41, p.877-889, 2001.

FORLI, S.; HUEY, R.; PIQUE, M.E.; SANNER, M.; GOODSELL, D.S.; OLSON, A.J. Computational protein-ligand docking and virtual drug screening with the AutoDock suíte. *Nat. Protoc*., v.11, n.5, p.905-919, 2016.

FRANCO, F.DE.F. **História evolutiva do "cluster" *Drosophila buzzatii* (grupo *D. repleta*): eventos históricos e diversificação de espécies no Brasil.** Ribeirão Preto, 2009. 197p. Tese de Doutorado – Faculdade de Medicina de Ribeirão Preto – Universidade de São Paulo.

FRANCO, F.F., PRADO, P.R.R., SENE, F.M., COSTA, L.F.; MANFRIN, M.H. Aedeagus morphology as a discriminant marker in two closely related cactophilic species of *Drosophila* (Diptera; Drosophilidae) in South America. *Anais da Academia Brasileira de Ciências*, v.78, n.2, p.203-212, 2006.

FRANCO, F.F., SILVA-BERNARDI, E., SENE, F., HASSON, E.; MANFRIN, M.H. Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the *Drosophila buzzatii* cluster. *Journal of Zoological Systematics and Evolutionary Research*, v.48, n.4, p.322-331, 2010.

GRANTHAM, R. Amino acid difference formula to help explain protein evolution. *Science*, v.185, p.862-864, 1974.

GUILLÉN, Y.; RUIZ, A. Gene alterations at *Drosophila* inversion breakpoints provide *prima facie* evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics*, v.13, n.53, p.1-18, 2012.

GHOSH, R.; UPADHAYAY, A.D.; ROY, A.K. In silico analysis, structure modeling and phosphorylation site prediction of vitellogenin protein from *Gibelion catla*. *Journal of Applied Biotechonology & Bioengineering*, v.3, n.1, p.1-6, 2017.

GLOSS, A.D.; VASSÃO, D.G.; HAILEY, A.L.; NELSON,-DITTRICH, A.C.; SCHRAMM, K.; REICHELT, M.; RAST, T.J.; WEICHSEL, A.; CRAVENS, M.G.; GERSHENZON, J.; MONTFORT, W.R.; WHITEMAN, N.K. Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol. Biol. Evol.*, v.31, n.9, p.2441-2456, 2014.

GUPTA, C.L.; AKHTAR, S.; BAJPAI, P. *In silico* protein modeling: possibilities and limitations. *EXCLI Jornal*, v.13, p.513-515, 2014.

HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, n.41, p. 95-98, 1999.

HARTL, F.U.; HAYER-HARTL, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Structural & Molecular Biology*, v.16, n.6, p.574-581, 2009.

HOFFMANN, A.; FALLMANN, J.; VILARDO, E.; MÖRL, M.; STADLER, P.F.; AMMAN, F. Accurate mapping of tRNA reads. *Bioinformatics*, v.34, n.7, p.1116-1124, 2018.

HUBER, P.C.; ALMEIDA, W.P.; FÁTIMA, A.DE Glutationa e enzimas relacionadas: papel biológico e importância em processos patológicos. *Quim. Nova*, v.31, n.5, p.1170-1179, 2008.

HUELSENBECK, J.P.; LARGET, B.; ALFARO, M.E. Bayesian phylogenetic model selection using reversible jump Markov Chain Monte Carlo. *Molecular Biology and evolution*, v.21, n.6, p.1123-1133, 2004.

HUELSENBECK, J.P.; RONQUIST, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, v.17, n.8, p.754-755, 2001.

KING, S.A.; ONAYIFEKE, B.; AKORLI, J.; SIBOMONA, I.; CHABI, J.; MANFUL-GWIRA, T.; DADZIE, S.; SUZUKI, T.; WILSON, M.D.; BOAKYE, D.A.; SOUZA, D.K.DE The role of detoxification enzymes in the adaptation of the major malaria vector *Anopheles gambiae* (Giles; Diptera:Culicidae) to polluted water. *Journal of Medical Entomology*, v.54, n.1, p.1674-1683, 2017.

KIM LAB – Computational and Integrative Biology. Disponível em: <http://www.kimlab.org/>. Acesso em: 20 març 2018.

KIRCHER, H.W.; HEED, W.B.; RUSSELL, J.S.; GROVE, J. Senita Cactus Alkaloids: Their Significance to sonorant desert *Drosophila* ecology. *J. Insect Physiol*., v.13, p.1869-1874, 1967.

KOKUDAI, C.B.S.; SENE, F.M.; MANFRIN, M.H. Sympatry and asympatric introgression between the cactophilic species *Drosophila serido* and *Drosophila antonietae* (Diptera: Drosophilidae). *Entomological Society of America*, v.104, n.3, p.434-442, 2011.

KUMAR, S.; SINGH, A.K. Allozyme polymorphism in *Drosophila*. *Proc. Zool. Soc.*, v.69, n.1, p.22-31, 2016.

KUHN, G.C.S. **Dinâmica evolutiva de DNAs satélites em populações e espécies de *Drosophila* do cluster *buzzatii*.** Ribeirão Preto, 2003. 144p. Tese de Doutorado – Faculdade de medicina de Preto – Universidade de São Paulo.

LEE, D.Y.; KIM, K-A.; YU, Y.G.; KIM, K-S. Substitution of aspartic acid with glutamic acid increases the unfolding translation temperature of a protein. *Biochemical and Biophysical Research Communications*, v. 320, p.900-906, 2004.

LÓPEZ-OLMOS, K.; MARKOW, T.A.; MACHADO, C.A. Evolution of *GSTD1* in cactophilic *Drosophila*. *J. Mol. Evol.*, v.84, n.1, p.285-294, 2017.

LOVELL, S.C.; DAVIS, I.W.; ARENDALL III, W.B.; DE BAKKER, P.I.W.; WORD, J.M.; PRISANT, M.G.; RICHARDSON, J.S.; RICHARDSON D.C. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*, v.50, n.3, p.437-450, 2003.

LOW, W.Y.; FEIL, S.C.; NG., H.L.; GORMAN, M.A.; MORTON, C.J.; PYKE, J.; MCCONVILLE, M.J.; BIERI, M.; MOK, Y.F.; ROBIN, C.; GOOLEY, P.R.; PARKER, M.W.; BATTERHAM, P. Recognition and detoxification of the insecticide DDT by *Drosophila melanogaster* Glutathione S-Transferase D1. *J. Mol. Biol.*, v.399, n.1, p.358-366, 2010.

LOW, W.Y.; NG, H.L.; MORTON, C.J.; PARKER, M.W.; BATTERHAM, P.; ROBIN, C. Molecular Evolution of Glutathione *S*-Transferase in the genus *Drosophila*. *Genetics*, v.177, n.1, p.1363-1375, 2007.

LUO, A.; QIAO, H.; ZHANG, Y.; SHI, W.; HO, Y.; XU, W.; ZHANG, A.; ZHU, C. Performace of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evolutionary Biology*, v.10, n.242, p.1-13, 2010.

MACHADO, C.A.; HEY, J. The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proc. R. Soc. Lond. Biol.*, v.270, n.1, p.1193-1202, 2003.

MANDAL, S. DE; CHHAKCHHUAK, L.; GURUSUBRAMANIAN, G.; KUMAR, N.S. Mitochondrial markes for identification and phylogenetic studies in insects – A review. *DE GRUYTER*, v.2, p.1-9, 2014.

MANFRIN, M.H.; DEBRITO, R.A.; SENE, F.M. Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. *Annals of the Entomological Society of America*, v.94, n.3, p.333-346, 2001.

MANFRIN, M.H.; SENE, F.M. Cactophilic *Drosophila* in South America: A model for evolutionary studies. *Genetica*, v.126, n.1, p.57-75, 2006.

MATEUS, R.P.; MACHADO, L.P.B.; MORAES, E.M.; SENE, F.M. Allozymatic divergence between border populations of two cryptic species of the *Drosophila buzzatii* cluster species (Diptera: Drosophilidae). *Biochemical Systematics and Ecology*, v.38, p.410-415, 210.

MATZKIN, L.M. The molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a Glutathione S-Transferase gene (*GstD1*) in *Drosophila mojavensis*. *Genetics*, v.178, n.1, p.1073-1083, 2008.

MATZKIN, L.M. Ecological Genomics of host shifts in *Drosophila mojavensis*. *Adv. Exp. Med. Biol.*, v.781, p.233-247, 2014.

MCDONALD, J.H.; KREITMAN, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, v.351, n.20, p.652-654, 1991.

MORAES, E.M.; SENE, F.M. Relationships between necrotic cactus availability and population size in a cactophilic *Drosophila* (Diptera, Drosophilidae) located on a sandstone table hill in Brazil. *Rev. Biol. Trop*., v.51, n.1, p.205-212, 2003.

MORTON, R.A. Evolution of *Drosophila* insecticide resistance. *Genome*, v.36, n.1, p.1-7, 1993.

NASCIMENTO, F.F.; REIS, M.DOS; YANG, Z. A biologist's to Bayesian phylogenetic analysis. *Nature ecology & evolution*, v.1, n.10, p.1446-1454, 2017.

NAVEIRA, H.; FONTDEVILA, A. The evolutionary history of *Drosophila buzzati*i. IX. High frequencies of new chromosome rearrangements induced by introgressive hybridization. *Chromosoma*, v.91, n.2, p.87-94, 1985.

NEI, M.; GOJOBORI, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, v.3, n.5, p.418-426, 1986.

NEI, M.; KUMAR, S. **Molecular evolution and phylogenetics**. New York: Oxford University Press, 2000. 395p.

NIELSEN, R.; YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, v.148, n.3, p.929-936, 1998.

OLIVEIRA, D.C.S.G.; ALMEIDA, F.C.; O'GRADY, P.; ARMELLA, M.A.; DESALLE, R.; ETGES, W.J. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila replete* species goup. *Molecular Phylogenetics and Evolution*, v.64, n.3, p.533-544, 2012.

PADRÓ, J.; PANIS, D.N.De; VRDOLJAK, J.; CARMONA, P.M.; COLINES, B.; HASSON, E.; SOTO, I.M. Experimental evolution of alkaloid tolerance in sibling *Drosophila* species with different degrees of specialization. *Evolutionary Biology*, v.45, n. 170, p.170-181, 2018.

PÁL, C.; PAPP, B.; LERCHER, M.J. An integrated view of protein evolution. *Nature Reviews Genetics*, v.7, p.337-348, 2006.

PANIS, D.N.DE; PADRÓ, J.; FURIÓ-TARÍ, P.; TARAZONA, S.; CARMONA, P.S.M.; SOTO, I.M.; DOPAZO, H.; CONESA, A.; HASSON, E. Transcriptome modulation during host shift is driven by secondary metabolites in desert *Drosophila*. *Molecular Ecology*, v.25, n.18, p. 4534-4550, 2016.

PARDANANI, J.H.; MCLAUGHLIN, J.L.; KONDRAT, R.W.; COOKS, R.G. Cactus alkaloids. XXXVI. Mescaline and related compounds from *Trichocereus peruvianus*. *Lloydia*, v.40, n.6, p.585-590, 1978.

PATWARDHAN, A.; RAY, S.; ROY, A. Molecular markes in phylogenetic studies – A review. *Phylogenetics & Evolutionary Biology*, v.2, n.131, p.1-9, 2014.

PDB (*Protein Data Bank*). Disponível em: <https://www.rcsb.org/pdb/home/home.do>. Acesso em: 10 out 2017.

PEREIRA, M.A.Q.R., VILELA, C.R.; SENE, F.M. Notes on breeding and feeding sites of some species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae). *Ciência e Cultura*, v. 35, n.1, p.1313-1319, 1983.

PETTERSEN, E.F; GODDARD, T.D.; HUANG, C.C.; COUCH, G.S.; GREENBLATT, D.M.; MENG, E.C.; FERRIN, T.E. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*., v.25, n.13, p.1605-1612, 2004.

PIR (*Protein Information Resource*) Disponível em: <http://pir.georgetown.edu/pirwww/search/comp_mw.shtml>. Acesso em: 20 set 2017.

POSADA, D.; BUCKLEY, T.R. Model selection and model averaging in phylogetics: advantages of Akaike Information Critterion and Bayesian Approaches over Likelihood Ratio Tests. *Systematic Biology*, v.53, n.5, p.793-808, 2004.

POSADA, D.; CRANDALL, K.A. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, v.14, n.9, p.817-818, 1998.

PubChem. Disponível em: <https://pubchem.ncbi.nlm.nih.gov/>. Acesso em: 15 fev 2018.

RAMBAUT, A.; DRUMMOND, A.J.; XIE, D.; BAELE, G.; SUCHARD, M.A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, p.1-5, 2018. Disponível em:10.1093/sysbio/syy032. Acesso em: 11 jul 2018.

RASON, H.; PRAPANTHADARA, L.; HEMINGWAY, J. Cloning and chracterization of two glutathione S-transferases from a DDT-resistant strain of *Anopheles gambiae*. *Biochem. J.*, v.324, n1, p.97-102, 1997.

RICHARDS, F.M. Areas, volumes, packing, and protein structure. *Ann. Rev. Biophys. Bioeng*., v.6, p.151-176, 1977.

RODRÍGUEZ-TRELLES, F., ALARCÓN, L.; FONTDEVILA, A. Molecular evolution and phylogeny of the *buzzatii* complex (*Drosophila repleta* group): a maximum-likelihood approach. *Molecular Biology and Evolution*, v.17, n.7, p.1112-1122, 2000.

RUIZ, A.; WASSERMAN, M. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity*, v.70, n.1, p.582-596, 1993.

RUIZ, A.; CANSIAN, A.M.; KHUN, G.C.S.; ALVES, M.A.R.; SENE, F.M. The *Drosophila serido* puzzle: putting new pieces together. *Genetica*, v.108, n.217, 2000.

SAISAWANG, C.; WONGSANTICHON, J.; KETTERMAN, A.J. A preliminary characterization of the cytosolic glutathione transferase proteome from *Drosophila melanogaster*. *Biochem. J.*, v. 442, n.1, p.181-190, 2012.

SAISAWANG, C.S.; KETTERMAN, A.J. Micro-plasticity of genomes as illustrated by the evolution of glutathione transferases in 12 *Drosophila* species. *PLOS ONE*, v.9, n.10, p.1-16, 2014.

SALI, A.; BLUNDELL, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, v.234, n.3, 779-815, 1993.

SALI, A. Comparative protein modeling by satisfaction of spatial restraints. *Molecular Medicine Today*, v.1, p.270-277, 1995.

SANTOS, M.H.; HASSON, E.; SENE, F.M.; MANFRIN, M.H. *Esterase-5* gene sequences from the *Drosophila buzzatii* cluster. *Dros. Inf. Serv.*, v.92, n.1, p. 44-56, 2009.

SCHMIDT, A.; LAMZIN, V.S. *Veni, vidi, vici* – atomic resolution unravelling the mysteries of protein function. *Curr Opin Struct Biol.*, v.12, n.6, p.698-703, 2002.

SILVA, A.; SENE, F. Morphological geographic variability in *Drosophila serido* (Diptera, Drosophilidae). *Rev. Bras. Entomol*, v.35, n.1, p. 455-468, 1991.

SIKOSEK, T.; CHAN, H.S. Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface*, v.11, n.100, p.1-35, 2014.

SINGH, N.; UPADHYAY, S.; JAISWAR, A.; MISHRA, N. *In silico* Analysis of protein. *J. Bioinform, Genomics, Proteomics*, v.1, n.2, p.1000-1007, 2016.

SOTO, I.M.; CARREIRA, V.P.; FANARA, J.J.; HASSON, E. Evolution of male genitalia: environmental and genetic factors affect genital morphology in two *Drosophila* sibling species and their hybrids. *BMC Evolutionary Biology*, v.7, n.77, p.1-11, 2007.

SOTO, I.M.; SOTO, E.M.; CORIO, C.; CARREIRA, V.P.; MANFRIN, M.; HASSON, E. Male genital and wing morphology in the cactophilic sibling species *Drosophila gouveai* and *Drosophila antonietae* and their hybrids reared in different host plants. *Environmental Entomology*, v.39, n.3, p.865-873, 2010.

SOTO, I.M.; CARREIRA, V.P.; CORIO, C.; PADRÓ, J.; SOTO, E.M.; HASSON, E. Differences in tolerance to host cactus alkaloids in *Drosophila koepferae* and *D. buzzatii*. *PLOS ONE*, v.9, n.2, p.1-9, 2014.

STERN, A.; DORON-FAIGENBOIM, A.; EREZ, E.; MARTZ, E.; BACHARACH, E.; PUPKO, T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Research*, v.35, p.506-511, 2007.

STINTZING, F.C.; CARLE, R. Cactus stems (*Opuntia* spp.): A review on their chemistry, technology, and uses. *Mol. Nutr. FoodRes.*, n.49, p.175-194, 2005.

SWANSON, W.J.; NIELSEN, R.; YANG, Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.*, v.20, n.1, p.18-20, 2003.

TAMURA, K.; STECHER, G.; PETERSON, D.; FILIPSKI, A.; KUMAR, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, v.30, n.12, p.2725-2729, 2013.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, v.22, n.22, p.4673-80, 1994.

TIDON-SKLORZ, R.; SENE, F.M. *Drosophila antonietae* sp. n. and *Drosophila gouveai* sp. n.: Two new species of the *Drosophila serido* superspecies taxon (Diptera, Drosophilidae). *Iheringia, série Zoologia*, n.90, p.141-146, 2001.

TOUNG, Y.P.S; HSIEH, T.S.; TU, C.P.D. The Glutathione S- Transferase D genes: a divergently organized, introless gene family in *Drosophila melanogaster*. *The Journal of Biological Chemistry*, v.268, n.13, p.9737-9746, 1993.

VILELA, C.R.; SENE, F.M. Two new neotropical species of the *repleta* group of the Genus *Drosophila* (Diptera, Drosophilidae). *Rev. Bras. Entomol.*, v.30, n.1, p.295-299, 1983.

VOROJEIKINA, D.; BROBERG, K.; LOVE, T.M.; DAVIDSON, P.W.; WIJNGAARDEN, E.V.; RAND, M.D. Glutathione S-Transferase activity moderates methylmercury toxicity during development in *Drosophila*. *Toxicological science*, v.157, n.1, p.211-221, 2017.

YANG, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, v.22, n.8, p. 1586-1591, 2007.

YANG, Z.; NIELSEN, R.; GOLDMAN, N.; PEDERSEN, A.M.K. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, v.155, p.431-449, 2000.

YANG, Z.; REIS, M.DOS. Statistical Properties of the Branch-Site Test of Positive Selection. *Mol. Biol. Evol.*, v.28, n.3, p.1217-1228, 2011.

ZELDOVICH, K.B.; SHAKHNOVICH, E.I. Understanding protein evolution: from protein physics to Darwinian selection. *Annual Review Physical Chemistry*, v.59, p.105-127, 2008.

ZHANG, J.; YANG, J.R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet*., v.16, n.7, p.409-420, 2015.

WAN, Q.H.; WU, H.; FUJIHARA, T.; FANG, S.G. Which genetic marker for which conservation genetics issue? *Electrophoresis*, v.25, n.14, p.2165-2176, 2004.

WOLF, L.K. PyRx. *Chemical & Engineering News*, v.87, n.31, 2009.

WONGSANTICHON, J.; ROBINSON, R.C.; KETTERMAN, A.J. Structural evidence for conformational changes of Delta class glutathione transferases after ligand binding. *Archives of Biochemistry and Biophysics*, v.521, p.77-83, 2012.

WU, C.H.; YEH, L.S.; HUANG, H.; ARMINSKI, L.; CASTRO-ALVEAR, J.; CHEN, Y.; HU, Z.; KOURTESIS, P.; LEDLEY, R.S.; SUZEK, B.E.; VINAYAKA, C.R.; ZHANG, J.; BARKER, W.C. The protein information Resource. *Nucleic Acids Research*, v.31, n.1, p.345-347, 2003.