

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**Estudos Epigenéticos em *Drosophila*: Uma Análise Evolutiva Multinível da
Enigmática Enzima DNMT2**

Gilberto Cavalheiro Vieira

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências - Genética e Biologia Molecular.

Orientadora: Dra. Vera Lúcia da Silva Valente

Co-orientador: Dr. Gustavo Fioravanti Vieira

Porto Alegre

Abril de 2019.

Este trabalho foi desenvolvido no Laboratório de *Drosophila* e Núcleo de Bioinformática da Imunogenética (NBLI) do Departamento de Genética da Universidade Federal Do Rio Grande Do Sul (UFRGS), sendo financiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - Processo PRONEX 10/0028-7), PROPESQ-UFRGS (Pró-Reitoria de Pesquisa da Universidade do Rio Grande do Sul) e PPGBM (Programa de Pós-Graduação em Genética e Biologia Molecular).

*There are more things in **genotype** and **phenotype**,
Horatio, than are dreamt of in your philosophy. -
Hamlet Act 1, scene 5, 159–167.*

AGRADECIMENTOS

Primeiramente gostaria de agradecer meus amados pais, minha mãe, **Dagmar**, pela força e perseverança atávica que lhe forjou uma alma inabalável diante dos desafios e meu amado pai, **Rivadavia**, por sua inigualável fonte de ardor pelo trabalho diário e sua sanha pelo saber. Meu irmão **Riva**, pessoa de personalidade única, um ser humano permeado de profundo conhecimento e que, mesmo mais jovem do que eu, é um exemplo de retidão científica e de aprofundamento acadêmico, além de um companheiro inestimável.

É claro que não poderia deixar de agradecer à minha companheira (esposa, namorada, amiga, parceira, redentora...) de todos os momentos - bons e ruins -, **Camila**. “Amore mio”, só tenho a te agradecer pelo exemplo de pessoa que você é, tanto na vida pessoal como na profissional. Obrigado por todo esse amor que recebo de ti, que me alimenta a alma e me faz cada dia mais forte. Sou uma pessoa afortunada por ter você ao meu lado.

Não é segredo a minha gratidão à minha orientadora **Vera Valente**, que por esses mais de dez anos de LabDros depositou sua preciosa confiança em meu trabalho, sempre acreditando e incentivando meu crescimento não só no âmbito acadêmico, mas no humano também. Obrigado por me mostrar que ciência se faz com paixão, também.

Ao **Gustavo Fioravanti Vieira** que mais do que orientador, se tornou um grande amigo. Alguém que admiro enormemente. Conhecer tua história só me fez respeitar ainda mais o teu legado. Espero, de todo coração, que nossa parceria continue por muitos e muitos anos.

Existe aqui um agradecimento todo especial à **Marícia**, não só pela amizade, mas aqui há uma gratidão infinita por ter me escolhido pra ser teu bolsista de iniciação científica, lá atrás, em 2008. O teu jeito apaixonado pela Epigenética, e pelo problema científico que nos debruçávamos, era tanto que transbordou e inundou minha alma científica. Espero um dia ser um décimo do que és e ser alguém tão inspirador quanto você é. A palavra “obrigado” parece pequena.

Agradeço de todo o coração à **Maríndia** por sempre estar disposta a compartilhar todo seu conhecimento ao longo desses anos. Foram tantos trabalhos, sempre apoiando e

dando uma luz quando parecia não haver solução para um determinado problema, além de ter papel fundamental na “ontogenia” da presente tese.

Rebeca fostes e és fundamental nisso tudo. Saiba disso. Gratidão pela parceria. E que parceria! Escarafunchadores de resultados. Tirar leite de pedra é conosco. O que mais desejo nessa jornada acadêmica é dar continuidade nessa rica parceria científica e amizade. Esse trabalho não seria o que é sem a tua incomensurável ajuda. E por seres essa pessoa incrível, tenha certeza que coisas incríveis estão para acontecer na tua vida.

Agradeço imensamente à **Carol Flores** pelos anos de conversas e ideias a respeito da vida acadêmica e da vida, enquanto repicávamos nossas linhagens. Obrigado pelas pizzas de sardinha maravilhosas (risos). E claro, agradeço a parceria que resultou em um dos capítulos da tese e espero que possamos encaminhar outras parcerias como essa.

Agradecimentos à **Pâmela** pela amizade e parceria de empreitadas no laboratório, fazendo frente em momentos em que eu não conseguia mais dar conta de todo o obrado. Só sei de uma coisa: já és uma grande pesquisadora, além de uma pessoa incrível.

Ao **Elmo** por sua inexorável cordialidade e atenção dada frente às dúvidas recorrentes sobre o PPGBM, amizade e apoio ao projeto de extensão do Curso em Biologia Evolutiva, ao longo desses anos.

Agradeço imensamente aos amigos-companheiros de empreitada do **CURSO DE BIOLOGIA EVOLUTIVA**, nesses mais de cinco anos de estrada. Obrigado **Leonardo Luvison, André Klein, Carol Diegues, Claudio Reis, Voltaire Paes, Marcel Lacerda, Ronaldo Paesi, Josmael Corso, Fernando Bueno, Daniel Rockenbach, Lucas Viscardi, Tomaz Melo, Giovanna Giudicelli e Thamara Almeida**. Gratidão ao **Prof. Aldo Araújo** por ter acreditado e depositado confiança em nós ao longo dessa jornada. Aprendi e continuo aprendendo muito com cada um de vocês.

Por fim, meus mais sinceros agradecimentos aos amigos de laboratório pelos satisfatórios momentos compartilhados: **Dani** (dá-lhe Grêmio... e depois boa sorte ao Inter! Hehehe...); **Berê** (obrigado pelo carinho sempre!); **Helena** (obrigado pelo carinho e pelos pudins maravilhosos... risos). E fica aqui o meu agradecimento e desejo de todo o sucesso do mundo pra galerínea nova que está chegando ao LabDros: **Carol, Natasha, Henrique, Robson, Arthur, Mayara e Ane**.

SUMÁRIO

Agradecimentos.....	4
Listas de Abreviaturas e Símbolos	7
Resumo.....	11
Abstract	13
Capítulo I.....	15
Introdução	16
A Epigenética	16
A Metilação de Citosinas Genômicas e Não-Genômicas.....	17
As DNA Metiltransferases	18
DNMT1 e DNMT3: Manutenção e Metilação <i>de novo</i>	20
Enigmática DNMT2.....	22
A Epigenética em <i>Drosophila</i>	24
Peculiaridades Epigenéticas em <i>Drosophila willistoni</i>	28
A Importância da Bioinformática na Resolução de Estruturas e Sistemas de Proteínas-Alvo	30
Objetivos	32
Capítulo II	34
Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions.....	34
Capítulo III.....	64
Characterizing the transcriptional expression and <i>in situ</i> localization of the <i>Dnmt2</i> gene in <i>Drosophila willistoni</i>	64
Capítulo IV.....	89
Linking epigenetic function to electrostatics: The DNMT2 structural model example	89
Capítulo V	111
Dissecting the Dynamics between the DNA Methyltransferase 2 (DNMT2) and the DNA between Prokaryotes and Eukaryotes.....	111
Capítulo VI.....	152
Discussão Geral.....	153
Perspectivas.....	159
Referências Bibliográficas	162
Anexos.....	185

LISTAS DE ABREVIATURAS E SÍMBOLOS

1DCT – PDB do cristal da enzima DNMT2 de *Haemophilus influenzae*
1G55 – PDB do cristal da enzima DNMT2 humana
1MHT – PDB do cristal da enzima DNMT2 de *Haemophilus haemolyticus*
2COV_{GE} – Covariância entre genótipo e ambiente
3D - tridimensional
3D – Tridimensional
3QV2 – PDB do cristal da enzima DNMT2 de *Entamoeba histolytica*
4DKJ – PDB do cristal da enzima DNMT2 de *Mycoplasma penetrans*
4H0N – PDB do cristal da enzima DNMT2 de *Spodoptera frugiperda*
Adomet – S-adenosilmetionina
APBS – *Adaptive Poisson-Boltzmann Solver*
BEAF - *Boundary Element-Associated Factor*
BLAST - *Basic Local Alignment Search Tool*
bp - base pairs
cDNA - DNA complementar
CH₃ – Metil
CNS – Sistema Nervoso Central
CpA - citosina- fosfato-adenina
CpG - citosina- fosfato-guanina
CpT - citosina- fosfato-timina
Ct - *Cycle threshold*
D.F. – Graus de Liberdade
dN – Mutações não-sinônimas
DNA - Ácido Desoxirribonucleico
DNase - Desoxirribonuclase
DNMT – DNA metiltransferase
DREF - *DNA Replication-Related Element Factor*
dS – Mutações sinônimas
dsRNA - RNA de cadeia dupla
Ehmeth – Metiltransferase de *Entamoeba histolytica*
EHMT - Histona metiltransferase da eucromatina
ERC – Covariação de Taxa Evolucionária

ETs – Elementos transponíveis

GEFs - *Guanine Nucleotide Exchange Factors*

GsDnmt2 - *Geobacter sulfurreducens* DNMT2

GTR – Modelo General Time Reversible

H – Hidrogênio

H3K27 – Lisina-27 da Histona-3

H3K9 – Lisina-9 da Histona-3

H3K9me2 – Lisina-9 da Histona-3 dimetilada

H3K9me3 – Lisina-9 da Histona-3 trimetilada

HaeIII – Metiltransferase de *Haemophilus influenzae*

HDAC - Histona deacetilases

HhaI – Metiltransferase de *Haemophilus haemolyticus*

JAK - *Janus kinase*

JTT – Modelo Jones-Taylor-Thornton

kb – quilobase

Kcat - número de renovação, é o número máximo de mols de substrato que podem ser convertidos em produto por mol de enzima em uma unidade de tempo

Km - constante de Michaelis

Lnl - *log likelihood*

LRT – *Likelihood Ratio Test*

LTRs - longas repetições terminais

m4C-MTase – DNA metiltransferase com atividade de metilação do N4 de citosinas

m5C-MTase - DNA metiltransferase com atividade de metilação do C5 de citosinas

m6A-MTase – DNA metiltransferase com atividade de metilação do N6 de adeninas

MDBs – Proteínas de ligação ao DNA metilado

melDnmt2 – Enzima Dnmt2 de *D. melanogaster*

MgCl₂ - Cloreto de Magnésio

min - minuto

ML - Maximum Likelihood

mL - mililitro

mM – milimolar

mRNA - RNA mensageiro

MSRE - Endonucleases de Restrição Sensíveis à Metilação

MTases – Metiltransferases
ncRNA – RNA não-codificante
ng - nanograma
NuRD - *Nucleosome Remodelling and Deacetylase*
°C - graus Celsius
ORF - Fase de leitura aberta
PAML - *Phylogenetic Analysis by Maximum Likelihood*
PCNA - Antígeno Nuclear de Proliferação Celular
PCR – Reação em Cadeia de Polimerase
PDB – *Protein Data Bank*
PEP – Fosfoenolpiruvato
PEV - *Position-Effect Variegation*
PHYML – *Phylogenies Maximum Likelihood*
PIPSA – Análises de Similaridades de Propriedades das Interações de Proteína
pmol – pico mol
pmol - picomol
QH - Medida de similaridade estrutural
Rb – Retinoblastoma
RdDM - Metilação do DNA dirigida por RNA
rDNA - DNA ribossomal
rhoDnmt2 – Enzima Dnmt2 de *D. rhopaloa*
RISC - Complexo de silenciamento induzido por RNA
RMN - Ressonância Magnética Nuclear
RMN – Ressonância Magnética Nuclear
RMSD – Desvio da raiz quadrada média
RMSF – Flutuação da raiz quadrada média
RNA - Ácido Ribonucleico
RNAi - RNA de interferência
RNase - Ribonuclease
RPF1 - *Ribosome Production Factor 1*
rRNA - RNA ribossomal
RT-PCR - Reação em Cadeia da Polimerase com Transcrição Reversa
S.E. – Erro Padrão

SAM – S-adenosilmetionina
SAP - *Shrimp Alkaline Phosphatase*
sfDNMT2 – Metiltransferase de *S. frugiperda*
siRNA – Pequenos RNA de interferência
SMN - *Survival Motor Neuron*
Su(var) - *Suppressor of variegation*
TE – Elementos Transponíveis
TET - *Ten-Eleven Translocation*
T_m – Temperatura de fusão
TRD - Domínio de reconhecimento de alvo
tRNA - RNA transportador
Uniprot - Banco de dados *Universal Protein Resource*
V_E – Variância ambiental
V_G – Variância genética
V_{GxE} – Interação entre genótipo e ambiente
V_P – Variância fenotípica total
V_ε – Variância residual
wilDnmt2 – Enzima Dnmt2 de *D. willistoni*
WT – *Wild-type*
μg - micrograma
μL - microlitro
μM - micromolar
ω – seleção positiva

RESUMO

Fundamentalmente, o processo epigenético caracteriza-se por contribuir com diferentes fenótipos alternativos emergindo em um organismo a partir de seu genótipo. Os mecanismos envolvidos na modulação da expressão dos genes são amplamente estudados, especialmente em vertebrados (mamíferos). Podemos dividir esses mecanismos epigenéticos de três formas: (i) modificação de citosinas genômicas; (ii) modificação pós-traducional das caudas de histonas; e (iii) controle de transcritos mediados por RNAs não-codificantes. Desses três mecanismos descritos, a modificação de citosinas genômicas por meio da metilação é a mais bem estudada. Apesar do amplo conhecimento que se estabeleceu sobre as funções e impactos da metilação em mamíferos nos aspectos fisiológicos e ontogenéticos, em invertebrados ainda residem diversas questões a serem elucidadas. Como *Drosophila* possui apenas a DNA metiltransferase 2 (Dnmt2), é reconhecida como organismo “*Dnmt2-only*”. Os padrões de metilação de citosinas nesse grupo de organismos possui destacadas diferenças em relação aos demais organismos que possuem as DNA metiltransferases canônicas (Dnmt1 e Dnmt3).

Assim, a presente tese apresenta em um primeiro momento, extensivo estudo sobre a conservação e evolução da Dnmt2 em drosofilídeos. Para isso, utilizamos aproximadamente 70 espécies de Drosophilidae. Como resultado das análises filogenéticas, os principais clados de Drosophilidae foram recuperados, onde observamos claramente as relações evolutivas entre o subgênero *Drosophila* e *Sophophora* e suas espécies. Analisamos quais as forças evolutivas conduziram a história da Dnmt2 dentro do gênero *Drosophila* e observamos que em *Dnmt2* de drosofilídeos há forte ação de seleção purificadora. Entretanto, seis sítios apresentaram sinais de seleção positiva e doze outros sítios, seleção positiva desbalanceadora - favorecendo mudanças estruturais e funcionais na enzima Dnmt2. Destacamos a alta taxa de substituição nucleotídica por códon no clado das espécies do subgrupo *willistoni*. Além disso, aprofundamos os estudos sobre os aspectos evolutivos dentro do contexto epigenético, analisando possíveis proteínas que interajam, direta ou indiretamente, com a Dnmt2. Ao todo, quinze proteínas compartilham significativa taxa de covariação evolutiva com Dnmt2, sugerindo serem potenciais parceiras em diferentes redes de controle de expressão gênica.

A relevância de um gene pode ser aferida, entre diversas formas, por seu padrão de expressão ao longo da ontogenia do organismo. Como *D. willistoni* se destaca por suas peculiaridades no âmbito epigenético, em um segundo momento, procuramos analisar os

padrões dos níveis transcricionais e localizar o gene *Dnmt2* através de hibridação *in situ* em cromossomos politênicos. Verificamos que o gene *Dnmt2* de *D. willistoni* possui maior nível transcricional nos estágios iniciais, diminuindo o com o avançar do desenvolvimento. Interessantemente, os dados obtidos por qPCR em adultos de *D. willistoni*, sugerem que esses possuem níveis de expressão levemente maiores do que aqueles observados em estudos prévios com *D. melanogaster*. Também detectamos transcritos de *Dnmt2* já no período da oogênese, o que pode ser um indicativo que esse seja de origem materna. O gene *Dnmt2* se encontra em região subtelomérica do braço III (Elemento B de Muller, comprovando sintenia cromossômica com outras 12 espécies de *Drosophila*). A localização de *Dnmt2* é por si só interessante, pois genes localizados em regiões próximas aos telômeros (e centrômeros) possuem características evolutivas marcadamente distintas das demais regiões cromossômicas, o que pode estar relacionado com as diferenças quanto aos aspectos evolutivos moleculares descritos previamente em *D. willistoni*.

Por fim, nas duas últimas etapas analisamos as propriedades físico-químicas emergentes da estrutura terciária de diversas *Dnmt2*, com especial atenção à região responsável pelo reconhecimento da sequência alvo a ser metilada (TRD). Utilizamos para esse estudo as *Dnmt2* de procariotos HhaI e HaeIII, do gênero *Haemophilus* (reconhecidas por serem DNA metiltransferases), a *Dnmt2* de *Geobacter sulfurreducens*, de *Entamoeba histolytica*, de *Spodoptera frugiperda* e humana (todas com afinidade por tRNA) e vinte e seis modelos gerados por meio de modelagem molecular por homologia estrutural de drosofilídeos. Constatamos que mutações em sítios da TRD modificam o perfil de distribuição de cargas eletrostáticas de superfície (CES) e, conseqüentemente, a atividade cinética da enzima junto ao substrato. Observamos que mesmo entre espécies próximas (drosofilídeos), as CES são diferentes. *D. willistoni* apresenta características em seu perfil de CES que a diferencia das demais espécies de *Drosophila* e a aproxima dos valores de cinética enzimática encontrados experimentalmente em HhaI. Através de simulações de dinâmica molecular do complexo *Dnmt2*-DNA de HhaI, *E. histolytica*, *S. frugiperda*, *D. melanogaster*, *D. willistoni* e *Dnmt2* humana, constatamos que as diferenças encontradas nas CES das TRDs refletem diretamente mudanças nas dinâmicas entre enzima e ligante. A *Dnmt2* de *D. willistoni* é a enzima eucariótica que apresentou o maior número de ligações de Hidrogênio estáveis com DNA, comportamento similar ao encontrado no complexo HhaI-DNA, reconhecidamente uma DNA metiltransferase.

ABSTRACT

Fundamentally, the epigenetic process is characterized by contributing with alternative phenotypes emerging in an organism from its genotype. The mechanisms involved in modulating gene expression are extensively studied, especially in vertebrates (mammals). We can divide these epigenetic mechanisms in three ways: (i) modification of genomic cytosines; (ii) histone tails post-translational modification; and (iii) control of transcripts mediated by non-coding RNAs. Of these three mechanisms described, the modification of genomic cytosines through methylation is the most well studied. In spite of the wide knowledge that has been established on the functions and impacts of the methylation in mammals in the physiological and ontogenetic aspects, in invertebrates there still remain several questions to be elucidated. As *Drosophila* has only DNA methyltransferase 2 (Dnmt2), it is recognized as "*Dnmt2*-only" organisms. The methylation patterns of cytosines in this group of organisms have remarkable differences in relation to the other organisms that have the canonical DNA methyltransferases (Dnmt1 and Dnmt3).

Thus, the present thesis presents at first, an extensive study on the conservation and evolution of *Dnmt2* in drosophilids. For this, we use approximately 70 species of Drosophilidae. As a result of the phylogenetic analyzes, the main clades of Drosophilidae were recovered, where we clearly observe the evolutionary relationships between the subgenus *Drosophila* and *Sophophora* and their species. We analyzed which evolutionary forces led to the history of *Dnmt2* within the genus *Drosophila* and we observed that in drosophilids *Dnmt2* there is a strong action of purifying selection. However, six sites showed positive selection signs and twelve other sites, positive-destabilizing selection - favoring structural and functional changes in the enzyme Dnmt2. We highlight the high rate of nucleotide substitution per codon in the clade of the species of the subgroup *willistoni*. In addition, we study the evolutionary aspects within the epigenetic context, analyzing possible proteins that interact, directly or indirectly, with Dnmt2. In all, fifteen proteins share a significant evolutionary rate covariation with Dnmt2, suggesting to be potential partners in different networks of control of gene expression.

The relevance of a gene can be measured, among its various forms, by its pattern of expression throughout the organism ontogeny. As *D. willistoni* stands out for its peculiarities in the epigenetic scope, in this second moment, we try to analyze the

transcriptional levels and localization of the *Dnmt2* gene through *in situ* hybridization in polytene chromosomes. We verified that the *D. willistoni Dnmt2* has a higher transcriptional level in the initial stages, decreasing with the development progress. Interestingly, the data obtained by qPCR in adults of *D. willistoni* suggest that they have levels of expression slightly higher than those observed in previous studies with *D. melanogaster*. We also detected *Dnmt2* transcripts already in the oogenesis period, which may be indicative of maternal origin. The *Dnmt2* is found in the subtelomeric region of the arm IIL (Muller B Element, confirming chromosomal syntenia with 12 other *Drosophila* species). The localization of *Dnmt2* is interesting in itself, because genes located in regions close to telomeres (and centromeres) have markedly different evolutionary characteristics than the other chromosomal regions, which may be related to the differences in molecular evolution aspects previously described in *D. willistoni*.

Finally, in the last two steps we analyzed the emerging physicochemical properties of the tertiary structure of several *Dnmt2*, with special attention to the region responsible for the recognition of the target sequence to be methylated (TRD). We used the prokaryotes *Dnmt2* HhaI and HaeIII of the genus *Haemophilus* (recognized as DNA methyltransferases), the *Geobacter sulfurreducens*, *Entamoeba histolytica*, *Spodoptera frugiperda* and human (all with tRNA affinity) *Dnmt2* and twenty-six drosophilids models generated by molecular homology modeling. We found that site mutations in the TRD modify the electrostatic surface charge distribution profile (ESCD) and, consequently, the enzymatic kinetics with the substrate. We observed that even among nearby species (drosophilids), ESCD are different. *D. willistoni* presents characteristics in its ESCD profile that differentiates it from the other *Drosophila* species and approximate the values of enzymatic kinetics found experimentally in HhaI. Through molecular dynamics simulations of the HhaI, *E. histolytica*, *S. frugiperda*, *D. melanogaster*, *D. willistoni* and human *Dnmt2*, we found that the differences in the ESCD of the TRDs directly reflect changes in the dynamics between enzyme and ligand. *D. willistoni Dnmt2* is the eukaryotic enzyme that showed the highest number of hydrogen bonds with DNA, a behavior similar to that found in the complex HhaI-DNA, known as DNA methyltransferase.

CAPÍTULO I

Introdução

Objetivos

Introdução

A Epigenética

O termo epigenética faz referência aos mecanismos envolvidos no processo de controle de expressão gênica nos organismos. Em 1942, Conrad H. Waddington, em parceria com Ernst Hadorn, cunhou o termo epigenética que deriva da palavra grega *epigenesis* (*epi* – sobre; *genesis* – origem). Os estudos desenvolvidos por Waddington e Hadorn procuravam descrever e entender os processos genéticos envolvidos no desenvolvimento dos organismos, estabelecendo diálogo entre duas áreas que pareciam estar apartadas por conta de contextos histórico-científicos.

Se por um lado os estudos sobre o desenvolvimento dos organismos se mostraram incapazes de responder questões evolutivas, sendo permeados por ideias vitalistas e obscurantistas para explicar a ontogenia dos organismos, por outro lado, a genética trazia luz às questões sobre hereditariedade e os fatores mecanicistas por trás da variabilidade dos organismos. Além disso, a ideia do selecionismo estrito delegava ao ambiente um papel passivo nos processos evolutivos, sem ação na emergência de novos fenótipos. Entretanto, em seus estudos, Waddington notou que fatores ambientais estressores poderiam causar a emergência de fenótipos não esperados em *Drosophila*, e que esses padrões ontogenéticos não esperados poderiam ser fixados numa população através de assimilação genética (Waddington 1942). Porém, apenas quatro décadas após as publicações de suas observações é que os mecanismos moleculares por trás de tais fenômenos seriam elucidados.

Os mecanismos envolvidos nos fenômenos epigenéticos atuam, basicamente, por três vias: (i) na alteração da expressão gênica, mudando a acessibilidade à cromatina para a regulação transcricional por meio das modificações no DNA, mas sem haver alterações na sequência das bases do genoma; (ii) modificações de histonas com a remodelação do nucleossomo (Lo and Sukumar 2008); e (iii) ação de RNA não-codificantes (Ørom and Shiekhatar 2011; Spitale et al. 2011; Dozmorov et al. 2013). É por via da herança epigenética que ocorre transmissão de informações que não são codificadas pelos genes de uma célula-mãe para a célula-filha, possibilitando o processo de citodiferenciação e posterior morfogênese, ou até mesmo a transmissão de processos ontogenéticos entre diferentes gerações sem haver a necessidade de haver para isso mutações genéticas (Kiani

et al. 2013). Além disso, a modulação desses mecanismos epigenéticos pode contribuir para o surgimento de diferentes fenótipos em um organismo a partir de um mesmo genótipo. Esses fenótipos alternativos podem ser induzidos em resposta a diferentes cenários ambientais (internos e externos ao organismo), modificando os estados epigenéticos de um ou mais loci do genoma do organismo. Em contrapartida, a desregulação do *status* epigenético (epimutações) pode resultar em efeitos deletérios aos organismos, resultando em diversas doenças como transtornos mentais, dependência de drogas (Tsankova et al. 2007; Sweatt 2009; Kumar and Rao 2011) e câncer (Feinberg and Vogelstein 1983; Hardy and Tollefsbol 2011; Park et al. 2011), em humanos.

No âmbito evolutivo se discutiu por algum tempo a importância e a relevância da epigenética. Logo, se os mecanismos evolutivos atuam em última análise sobre os fenótipos, a epigenética acaba tendo importante papel dentro da Biologia Evolutiva. Até mesmo a genética quantitativa, que leva em conta os mecanismos que influenciam a variação quantitativa através da fórmula de variância fenotípica ($V_P = V_G + V_E + V_{G \times E} + 2COV_{GE} + V_\epsilon$ onde, V_G é a variância genética; V_E variância ambiental; $V_{G \times E}$ interação genética-ambiental; $2COV_{GE}$ covariância genética-ambiental e V_ϵ variância epigenética), pode haver influência da epigenética, levando por conseguinte a consequências evolutivas (Banta and Richards 2018). Devemos lembrar que entre genótipo e fenótipo, e conectando cada um desses níveis entre si, reside um apanhado complexo de processos ontogenéticos que podem sofrer influência ambiental (interna ou externa).

A Metilação de Citosinas Genômicas e Não-Genômicas

A metilação de citosinas é o mecanismo epigenético mais estudado e descrito (Vilkaitis et al. 2001; Marhold et al. 2004; Varriale 2014). Esse caracteriza-se pelo processo de substituição de um átomo de hidrogênio (H) por um grupo metil (CH₃) no carbono 5 do anel de citosina, sendo a reação catalisada por diferentes tipos de metiltransferases (MTases). A metilação de citosinas genômicas está ligada a diversos processos como o *imprinting* genômico, a estabilidade genômica e o silenciamento de transposons e de retrovírus endógenos (Bird 1992; Yoder et al. 1997b).

Quando ocorre o processo de modificação de citosinas por metilação há a alteração no padrão físico-químico do nucleotídeo, alterando os padrões de distribuição de cargas eletrostáticas (Prokhortchouk and Defossez 2008; Severin et al. 2011; Goel et

al. 2014; Sanchez and Mackenzie 2016). Essa mudança acaba interferindo na capacidade de fatores de transcrição reconhecerem e interagirem de maneira adequada em tais sítios. Além disso, tais modificações abrem a possibilidade para que fatores *trans* do processo transcricional, como as histonas deacetilases e proteínas de ligação ao DNA metilado (MDBs), interajam com o DNA - essas últimas parecem competir por sítios alvos de fatores de transcrição (Nan et al. 1998; Fuks et al. 2000; Kim et al. 2002).

Interessantemente o padrão de distribuição de citosinas metiladas não é homogêneo em posição e função ao longo do genoma de diferentes espécies. Em mamíferos são amplamente descritas metilações nos sítios de início de transcrição nas ilhas CpG (Okano et al. 1999). Contudo, em genes que não possuem sítios de início de transcrição ricos em CpG pode ocorrer metilação em sequências não-CpG, sendo o processo de silenciamento promovido primordialmente pelo nucleossomo (Farthing et al. 2008; Han et al. 2011). Outro aspecto interessante, é a metilação em citosinas no corpo dos genes (termo utilizado para metilação nas unidades de transcrição gênica). A metilação no corpo do gene está relacionada aos mecanismos de controle de processamento alternativo de RNA em vertebrados (Flores et al. 2012). Em invertebrados, apesar de no princípio se acreditar que houvessem metilações espúrias, atualmente estudos tem associado esse fenômeno a funções biológicas de alta complexidade, como a determinação de castas em abelhas e formigas (Elango et al. 2009; Bonasio et al. 2012).

Entretanto, a metilação de citosinas não-genômicas também desempenha papel fundamental nos processos de controle de expressão gênica. Nas últimas décadas foi descoberta uma série de modificações em RNA não-codificantes que apresentam importância na resposta a diversos cenários. Exemplificam-se eventos estressores por mobilização de elementos transponíveis, limitação de recursos alimentares, infecção viral, entre outros (Liebers et al. 2014; Matzke and Mosher 2014; Lyons et al. 2018).

As DNA Metiltransferases

O fenômeno de metilação de citosinas é mediado por diferentes classes de DNA metiltransferases (DNA MTase). As duas primeiras classes atuam convertendo os nitrogênios exocíclicos dos nucleotídeos adenina e citosina em N6-metiladenina (m6A-MTase) (Zhang et al. 2015) e N4-metilcitosina (m4C-MTase) (Ehrlich et al. 1987), respectivamente, sendo encontradas em procariotos. A terceira classe de DNA MTase são

denominadas de DNA (citosina-5) metiltransferases (m5C-MTase) (Kumar et al. 1994), que catalisam a reação de metilação inserindo um grupamento metil (CH₃) no Carbono cinco do anel pirimidínico de citosinas, convertendo-as em 5-metilcitosina (**Figura 1**), sendo estas amplamente distribuídas entre procariotos e eucariotos (Motorin et al. 2009).

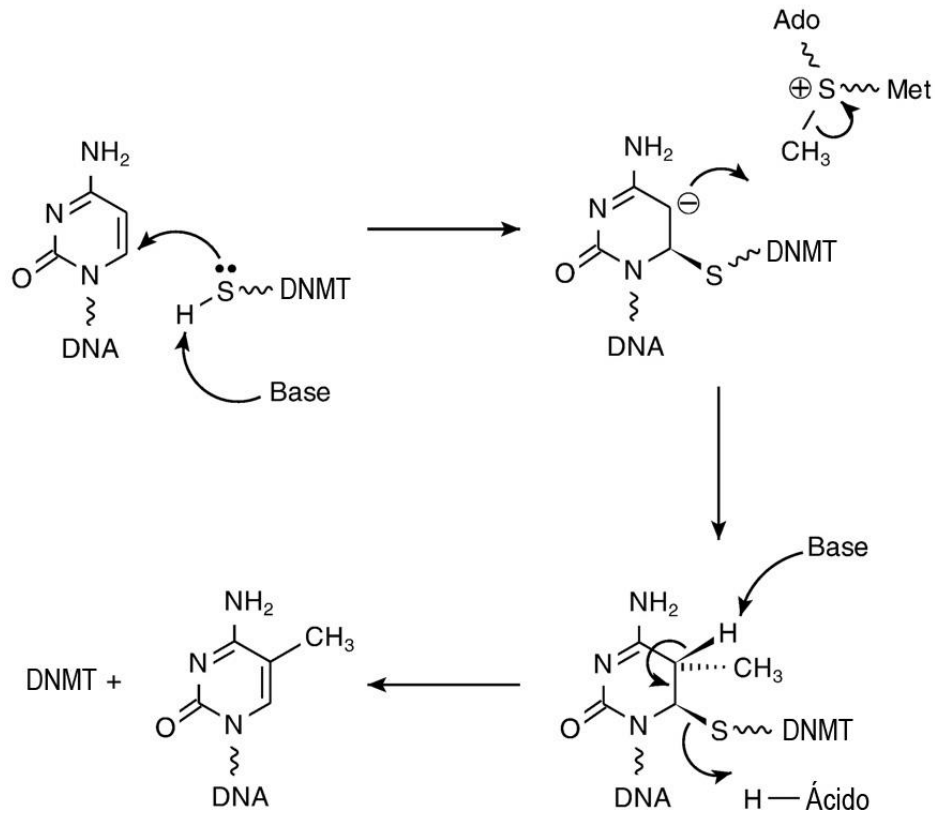


Figura 1. Representação dos processos de metilação de citosina mediada pelas DNA MTase. Num primeiro momento há um ataque nucleofílico na posição seis da citosina por parte da enzima (Cys81 em MHaI, por exemplo). Essa ação ativa a posição cinco para um ataque eletrofílico pelo cofator (AdoMet), quando o grupamento metil é transferido para a citosina alvo.

Na classe das m5C-MTase existem três diferentes famílias de enzimas: a DNMT1, a DNMT2 e as DNMT3 (**Figura 2**), todas com homologia entre si no domínio catalítico (Klimasauskas et al. 1994c; Jurkowska and Jeltsch 2016). Entretanto, cada uma delas desempenha diferentes papéis no processo de estabelecimento e manutenção dos padrões de metilação no desenvolvimento dos organismos.

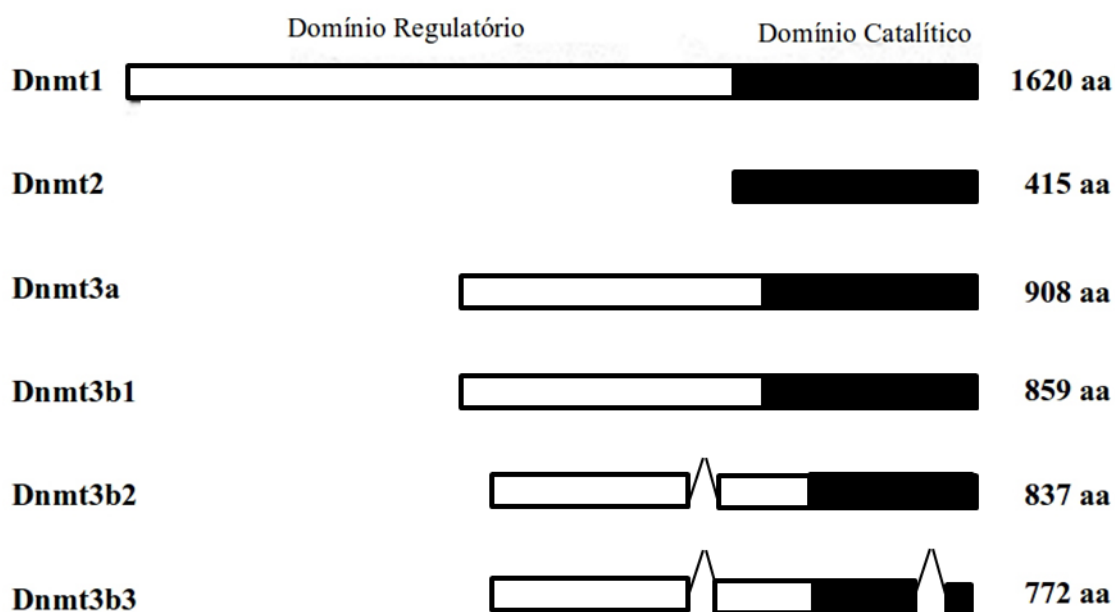


Figura 2. Ilustração esquemática de DNMTs conhecidas de vertebrados. A região em preto representa o domínio catalítico C-terminal conservado (imagem adaptada de Mund *et al.*, 2004).

DNMT1 e DNMT3: Manutenção e Metilação *de novo*

A DNMT1 é responsável pela manutenção das marcações epigenéticas. Durante o processo de replicação do DNA, a DNMT1 terá maior atividade sobre o DNA hemimetilado, visto que a replicação da molécula de DNA é semiconservativa. Assim, a fita molde irá carregar os padrões de metilação que serão copiados pela DNMT1 para as novas fitas sintetizadas (Yoder *et al.* 1997a).

Molecularmente, a DNMT1 é a maior das MTases, com massa molecular de 184 kDa, possuindo duas isoformas, DNMT1a e DNMT1b, em que a DNMT1b resulta do processamento de splicing alternativo do gene de DNMT1 (Bonfils *et al.* 2000). A DNMT1 quando silenciada, resulta em fenótipos letais pela perda da manutenção dos padrões de metilação no genoma replicado. Em estudos com camundongos, verificou-se que indivíduos homozigóticos DNMT1^{-/-} não sobrevivem à gestação. Processos tumorigênicos são amplamente descritos quando da atividade anormal das MTases. Elevados níveis de transcritos de DNMT1 e atividade MTase em associação com células cancerígenas *in vitro* e tumores *in vivo* já foram descritos (Kautiainen and Jones 1986; el-Deiry *et al.* 1991; Saito *et al.* 2003). Outro aspecto interessante é que as MTase atuam

conjuntamente com outros fatores envolvidos nos mecanismos de controle de expressão gênica, como é o caso da associação da DNMT2 com a histona deacetilase 2 (HDAC2) (Rountree et al. 2000). Em sua região N-terminal, tem-se o domínio regulatório (Pradhan and Esteve 2003), enquanto a região C-terminal comporta o domínio catalítico. Ambas as regiões estão conectadas por uma região de repetição Gli-Cis. Como a DNMT1 é comumente encontrada nas forquilhas de replicação de DNA acaba tendo interações com proteínas como a PCNA (*Proliferative Cell Nuclear Antigen*) (Chuang et al. 1997), histonas deacetilases (Fuks et al. 2000; Rountree et al. 2000), produtos do gene de retinoblastoma (Rb) (Robertson et al. 2000), DNMT3a e DNMT3b (Kim et al. 2002).

Em mamíferos, durante o estágio de pré-implantação embrionária, o genoma apresenta-se praticamente livre de metilação, contudo se observa nesse estágio a transcrição do gene *DNMT1* (Mertineit et al. 1998; La Salle et al. 2004). Posteriormente o transcrito de *DNMT1* é traduzido no citoplasma, porém a enzima DNMT1 não é direcionada para o núcleo, sendo mantida no citoplasma. Por outro lado, verifica-se durante esse período inicial de desenvolvimento uma intensa atividade catalítica das DNMT3a e DNMT3b, estabelecendo os padrões de metilação iniciais. Após esse período, os níveis e atividade da DNMT3a e DNMT3b diminuem gradativamente, passando, então, a atuar primordialmente a DNMT1 (Okano et al. 1999).

Enquanto a DNMT1 é responsável pela manutenção das marcações epigenéticas, as DNMT3a e DNMT3b são as principais moléculas envolvidas na metilação *de novo* em vertebrados. A metilação *de novo* nada mais é do que o fenômeno de estabilização dos padrões de metilação em estágios de desenvolvimento iniciais (Okano et al. 1999). A inativação de ambas DNMT3 é letal para os organismos, visto que a metilação *de novo* é essencial para que ocorra a correta compartimentalização do genoma durante o processo de citodiferenciação e posterior morfogênese tecidual (Okano et al. 1999). As DNMT3a e a DNMT3b diferenciam-se em suas afinidades espaço-temporais ao longo do estágio do desenvolvimento do organismo. A DNMT3b possui uma maior atividade nos estágios iniciais de desenvolvimento, metilando regiões centroméricas, enquanto que a DNMT3a atua nos estágios posteriores, metilando genes ou sequências fundamentais para o desenvolvimento pós-implantação (Okano et al. 1999). A princípio, DNMT3a e DNMT3b atuam redundantemente na metilação do genoma e dos elementos repetitivos, contudo, DNMT3b tem também um papel proeminente na metilação das ilhas CpG nos cromossomos autossômicos e X (Auclair et al. 2014).

Análises filogenéticas sugerem que os genes ancestrais de DNMT3a e DNMT3b surgiram depois da emergência dos urocordados e antes da radiação evolutiva dos gnatostomados, através de processo de duplicação gênica. Após o fenômeno de duplicação o gene da DNMT3b sofreu diversas substituições não-sinônimas. Em mamíferos a simples substituição de um aminoácido (uma isoleucina por uma asparagina na posição 662) da DNMT3b, é considerada um evento preponderante para conceber o aumento de sua atividade catalítica (Shen et al. 2010).

Enigmática DNMT2

A DNMT2 apresenta ampla distribuição nos mais diversos filos e em diferentes reinos, apresentando alta conservação de seus sítios catalíticos, mesmo quando se compara procariotos e eucariotos (Jurkowski and Jeltsch 2011; Schulz et al. 2012). Em alguns organismos a única MTase presente é a DNMT2, como é o caso dos drosofilídeos (Schaefer and Lyko 2010). Além da homologia e conservação de seus sítios catalíticos, a DNMT2 possui peculiaridades que acabam por atrair a atenção de diversos grupos de pesquisa. Como exemplo podemos citar sua plasticidade ao substrato em que atua (Dong et al. 2001; Ponger and Li 2005; Jeltsch et al. 2006; Schaefer and Lyko 2010; Durdevic 2013). Interessantemente, a DNMT2 parece carregar propriedades ambivalentes que permitem executar funções de metilação *de novo* e atuar sobre DNA hemimetilado. Por conta dessas características, a DNMT2 é considerada uma subfamília enigmática.

No que diz respeito à atividade em diferentes substratos, em mamíferos, invertebrados e plantas a DNMT2 é classificada *in vivo* como uma tRNA^{Asp} MTase (Goll et al. 2006a; Tuorto et al. 2012). Entretanto, em organismos que não possuem as MTases canônicas (DNMT1 e DNMT3), mas apresentam metilação genômica, a DNMT2 parece atuar como MTase em ambos os substratos: DNA e tRNA (Krauss and Reuter 2011). Em diversos trabalhos realizados há a descrição de atividade de metilação de DNA promovido pela DNMT2, mesmo que em níveis basais (Hermann et al. 2003; Kunert et al. 2003; Kuhlmann et al. 2005). O que se discute são as atividades preferenciais da DNMT2 e os mecanismos ecológicos e evolutivos que modulam sua atividade.

Devido a sua conservação e caracterização inicial como uma DNA MTase, a DNMT2 foi classificada como uma enzima nuclear, porém, a DNMT2 parece variar

quanto à sua citolocalização de espécie para espécie. Em humanos, por exemplo, ela se localiza em regiões citoplasmáticas (Goll et al. 2006a). Já em *Dyctiostelium discoideum* e *Entamoeba histolytica*, a DNMT2 está localizada unicamente na matriz nuclear (Kuhlmann et al. 2005; Schaefer et al. 2008). Essa característica contrasta com a alta especificidade local das enzimas DNMT1 e DNMT3 à região do núcleo celular. Em humanos, a DNMT2 possui uma maior taxa de expressão em tecidos como o cardíaco, o pulmonar e as gônadas (Yoder et al. 1997a; Okano et al. 1998), enquanto em organismos como *Drosophila melanogaster* e *Danio rerio* altos níveis de expressão de DNMT2 são encontrados nos estágios iniciais de desenvolvimento (Lyko et al. 1985; Roder et al. 2000; Kunert et al. 2003).

Tais características tornam a DNMT2 uma enzima peculiarmente diferente das demais metilases. O fato da DNMT2 atuar em diferentes substratos e também estar localizada em diferentes compartimentos celulares permite elaborar a hipótese de sua interação com diferentes tipos de moléculas e seu envolvimento em diferentes rotas metabólicas. Já foi descrita a interação entre DNMT2 e a enzima Enolase - característica da rota glicolítica – em *Entamoeba histolytica*, onde a Enolase parece modular a atividade da DNMT2 como tRNAAsp MTase, em cenários de escassez de glicose (Tovy et al. 2010). Quando *E. histolytica* encontra-se em situações de estresse, a DNMT2 é mobilizada para regiões de grânulos de estresse citoplasmático e corpos de processamento de RNA, envolvidos no silenciamento de retrotransposons, sugerindo que tal atividade esteja ligada a algum tipo de mecanismo de defesa celular primitivo contra infecções virais (Thiagarajan et al. 2011). O vírus HIV consegue aumentar sua sobrevivência na célula hospedeira utilizando a capacidade de metilação de RNA da DNMT2, potencializando a estabilidade do seu próprio genoma. Após a infecção, a DNMT2 se realoca do núcleo para os grânulos de estresse citoplasmático e metila o RNA do vírus HIV, promovendo a estabilidade pós-transcricional do RNA viral (Dev et al. 2017).

A DNMT2 também é alvo da atividade da Dicer-2 em resposta a tratamentos *heat shock* (Durdevic et al. 2013b), controlando a expressão e a regulação de diversas proteínas *Heat shock* (Fisher et al. 2006; Schaefer et al. 2010; Thiagarajan et al. 2011), além de estar envolvida na longevidade de *Drosophila* e em fibroblastos de mamíferos (Lin et al. 2005; Lewinska et al. 2017; Lewinska et al. 2018) e também na integridade estrutural da cromatina (Phalke et al. 2009).

Assim, apesar da DNMT2 possuir forte conservação de seus sítios catalíticos, sua principal funcionalidade e seus mecanismos de modulação ainda são uma incógnita. Estudos promovendo a depleção de DNMT2 em diversos organismos falharam em elucidar sua função primordial. Dessa forma, novos estudos estão procurando encontrar as respostas através de uma nova abordagem, identificando as capacidades funcionais da DNMT2 e sua relação com as proteínas com as quais interage.

A Epigenética em *Drosophila*

O organismo *Drosophila melanogaster* foi considerado livre de metilação genômica até final da década de 90 (Patel and Gopinathan 1987; Tweedie et al. 1999). Entretanto com o surgimento de novas técnicas de análises moleculares, em 2000 uma série de trabalhos acabou descrevendo a presença de citosinas metiladas em genoma de *D. melanogaster* (Gowher et al. 2000a; Lyko et al. 2000; Lyko 2001). Diferentemente do que ocorre em vertebrados, apenas aproximadamente 1 em cada 1000-2000 resíduos de citosina são metilados em moscas adultas, sendo que a metilação do DNA foi detectada em todos os estágios de desenvolvimento em *D. melanogaster*. Além disso, a metilação em *Drosophila* possui a peculiaridade de se apresentar em dinucleotídeos CpA, CpT e CpC, e não em CpG como é comumente encontrado em mamíferos (Lyko et al. 1985; Kunert et al. 2003; Panikar et al. 2015).



Figura 3. Níveis de metilação de citosinas ao longo dos estágios de desenvolvimento em *Drosophila melanogaster* (imagem modificada de Lyko et al., 2000).

O fenômeno de metilação em drosofilídeos é considerado pouco relevante por muitos pesquisadores, tal argumento está baseado no baixo percentual de citosinas metiladas no genoma dos indivíduos desse grupo e pela postura consensual de que a DNMT2 atua muito mais como uma tRNA^{asp} MTase do que propriamente uma DNA MTase (Hermann et al. 2003; Liu et al. 2003; Tang et al. 2003; Jeltsch et al. 2006). Dessa forma, as citosinas metiladas seriam resultantes de atividade residual da DNMT2, e não de um processo controlado. Assim, os mecanismos de controle de expressão gênica em drosofilídeos subjazem em outros níveis, que não a metilação de citosinas genômicas.

Basicamente, como descrito previamente, existem três mecanismos de controle de expressão gênica (i. metilação de citosinas, ii. modificações pós-traducionais de histonas; e iii. mecanismos mediados por RNAs). Em *Drosophila*, ao que as evidências fazem concluir, os mecanismos (ii) e (iii) são os principais mecanismos epigenéticos envolvidos no controle dos níveis de transcritos. A metilação de prolínas, lisinas e argininas nas caudas de histonas estão intimamente relacionadas às respostas oriundas de eventos estressores, como *heat shock*, onde existe um aumento no grau de metilação de prolínas N-terminal das histonas H2B e uma diminuição nos níveis de metilação na histona H3, promovendo assim a remodelação da cromatina e permitindo que fatores de transcrição acessem genes envolvidos na resposta ao evento estressor (Desrosiers and Tanguay 1988). A histona H1 (histona de ligação ou grampo) também desempenha importante papel, garantindo a estabilidade cromossômica e o silenciamento de Elementos Transponíveis (ETs) em *Drosophila*. Em particular ressaltam-se os retrotransposons R1/R2 que estão presentes no *locus* rDNA (Vujatovic et al. 2012).

Interessantemente, o estado da cromatina varia de acordo com as diferentes famílias de ETs, entre espécies intimamente relacionadas e também entre as linhagens do tipo selvagem. Esse é o caso descrito nas análises do *status* epigenético de quatro famílias ETs (*roo*, *tirant*, *412* e *F*) em sete linhagens selvagens de *D. melanogaster* e *D. simulans* (Rebollo et al. 2012). Sabe-se que em drosofilídeos diferentes retrotransposons estão associados com H3K9me3 e H3K9me2, tanto nas regiões promotoras quanto em suas ORFs (*open reading frames*) (Riddle et al. 2011). A associação de modificações em histonas para repressão (H3K9me2/3) e ativação gênica (H3K4me2/3) também foram observadas em retrotransposons nas regiões de eucromatina e heterocromatina, embora o enriquecimento para H3K4me2/3 seja fraco ou moderado no segundo cenário (Klenov et al. 2007; Gravio et al. 2018).

Mais recentemente, verificou-se que os níveis da proteína E(z), uma histona metiltransferase (H3K27), podem ser modulados em resposta ao estado nutricional de larvas em *D. melanogaster*, modificando a longevidade dos indivíduos (Xia et al. 2016). Quando as larvas foram submetidas a uma dieta de baixo valor proteico, os níveis de trimetilação da H3K27 eram aumentados por conta da atividade da proteína E(z), diminuindo a longevidade das moscas de F0 a F2. O interessante, é que temos aqui uma descrição de um fenômeno de herança transgeracional, visto que as linhagens posteriores não foram submetidas a dieta de baixo valor proteico, mas mantiveram a resposta fenotípica ao evento estressor anterior. Alterações dinâmicas de acetilações de H3K27 e H3K9 mediadas por histonas acetiltransferase (HAT) em regiões potencializadoras e promotoras reprimem sinergicamente genes associados à pluripotência relacionada com células-tronco ou progenitoras neurais, regulando positivamente genes associados à morfogênese neuronal, com diferenciação de neurônios em drosofilídeos (Ye et al. 2017). Por sinal, a utilização desses organismos modelos (drosófilas) em estudos epigenéticos tem se mostrado extremamente profícuo no entendimento de diversos fenômenos como a geração e manutenção de memórias. O papel fundamental da histona metiltransferase da eucromatina (EHMT), que faz parte de uma família de metilases H3K9, no desenvolvimento dos dendritos periféricos, comportamento locomotor, aprendizagem não-associativa e memória de comportamento de côrte, estabeleceu-se através de estudos com *Drosophila* (Kramer et al. 2011). Tais resultados são relevantes para o entendimento dos mecanismos fisiopatológicos subjacentes à Síndrome de Kleefstra, uma forma grave de incapacidade intelectual causada por mutações em *EHMT1* humana.

Por fim, os mecanismos de controle de expressão gênica mediados por RNAs não-codificantes vem recebendo especial atenção nos últimos tempos. Os RNAs de interferência (RNAi) possuem forte associação com os diferentes mecanismos epigenéticos, como metilação/acetilação de histonas e metilação de DNA genômico (Volpe et al. 2002). RNAs de cadeia dupla (dsRNA) desencadeiam o processo de interferência gênica pós-transcricional dependente de homologia (RNAi) intimamente relacionado com a co-supressão de transcritos exógenos virais, na maioria dos casos. Os dsRNA são processados por uma família de enzimas *Dicer* em pequenas sequências de dsRNA possuindo de 21-25 nucleotídeos, denominados pequenos RNAs de interferência (ou siRNA – *small interfering RNA*). Esses siRNA são agrupados a complexos proteicos chamados Complexos de Silenciamento Induzidos por RNA (RISC, *RNA-induced*

silencing complex), que irão atuar degradando RNA mensageiros homólogos. Apesar da ênfase dada sobre as diferentes formas de silenciamento por RNA em nível pós-transcricional, existe um aumento de evidências que dsRNA podem também induzir o silenciamento transcricional e promover mudanças epigenéticas ao nível genômico (Mette et al. 2000).

Um dos casos é a metilação direcionada por RNA (RdDM, *RNA-directed DNA methylation*) envolvendo metilação *de novo*, que foi primeiramente descrito em plantas e posteriormente em vertebrados (Wassenegger 2000; Matzke and Moshier 2014). Se porventura o dsRNA contiver em sua sequência homologia com sítios promotores, a metilação *de novo* e silenciamento gênico transcricional pode ocorrer *in trans* (Jones et al. 1999; Mette et al. 2000). As pistas de como tais RNAs atraem a maquinaria de metilação genômica vem da observação de que a RdDM frequentemente ocorre conjuntamente com a formação de RNAi, tanto em plantas, quanto em animais e alguns fungos. Em mamíferos, por exemplo, a maquinaria de formação de RNAi acaba por interagir com a enzima DNMT1 e DNMT3 (responsável pela manutenção e metilação *de novo*, respectivamente) (Matzke and Birchler 2005). Uma ligação semelhante ocorre em *Drosophila*, onde RNAi, H3mK9 e proteínas de heterocromatina atuam em conjunto para promover o silenciamento gênico (Pal-Bhadra et al. 2004). Mutações nos genes envolvidos nos mecanismos de processamento de RNAi, como *aubergine*, *homeless*, e *piwi* acabam causando perda de metilação em H3K9 e capacidade de ligação de proteínas heterocromáticas (Williams and Rubin 2002; Finnegan et al. 2003). Entretanto, até o momento nenhuma evidência do fenômeno de RdDM foi descrita em drosofilídeos.

Conforme descrito anteriormente, muitos autores consideram a metilação de citosinas em drosofilídeos como uma consequência de atividade espúria da única m5C-MTase presente nesses organismos, a DNMT2. Nesse sentido, a repressão transcricional mediada por proteínas de ligação a sítios CpG metilados já foram descritas (Roder et al. 2000). Essas proteínas possuem domínio específico de reconhecimento de sequências onde ocorre a presença de citosina metilada e são conhecidas pela sigla MBD (*methylated-DNA binding domain*). As proteínas com MBD induzem a condensação da cromatina através do recrutamento de HDAC, gerando um complexo proteico de repressão transcricional. Em *Drosophila melanogaster* encontra-se presente o gene *dMBD2/3* que codifica duas formas de proteína de ligação à citosinas metiladas, oriundas de processamento alternativo de RNA, gerando as isoformas: dMBD2/3 e dMBD2/3Δ

(Roder et al. 2000). Interessantemente, a isoforma dMBD2/3 Δ tem preferência por sítios 5mCpG enquanto a dMBD2/3, 5mCpT/A (Roder et al. 2000; Marhold 2004). Mutantes MBD2/3 apresentam altas taxas de defeitos de segregação cromossômica durante os estágios iniciais de desenvolvimento (Marhold et al. 2002; Marhold 2004). Além disso, machos que possuem redução na expressão de dMBD-R2, especificamente em neurônios de octopamina (equivalente a norepinefrina) exibem comportamento de corte voltada a interespecies divergentes, como a de *D. virilis* e *D. yakuba* e diminuição no sucesso de acasalamento coespecífico. Surpreendentemente, há o aumento do cortejo entre machos que possuem redução na expressão de dMBD-R2 e consequente diminuição da agressividade (Gupta et al. 2017). Esses dados sugerem a existência de estruturas de cromatina dependentes de metilação em *Drosophila*, ou seja, o papel da metilação de citosinas genômicas parece ser fundamental em diversos níveis nesses organismos.

Peculiaridades Epigenéticas em *Drosophila willistoni*

O subgênero *Sophophora* de *Drosophila* é classicamente dividido em dois grandes grupos: as espécies do Velho Mundo e as do Novo Mundo, conhecidas como Neotropicais. As espécies do Velho Mundo estão distribuídas dentro dos grupos *melanogaster*, *setifemur*, *populi*, *obscura*, *fima*, *díspar* e *dentissima*. Já as espécies Neotropicais estão subdivididas dentro dos grupos *saltans* e *willistoni*. Estudos biogeográficos indicam que a raiz dos clados *melanogaster* e *obscura* reside no sudeste asiático, tendo o ancestral comum vivido por volta de 30-27 milhões de anos atrás (Oligoceno) (Throckmorton 1975). Entretanto, estudos com marcadores genéticos indicam que a separação tenha ocorrido antes, há aproximadamente 45 milhões de anos (Robe et al. 2010), sendo que as espécies do Velho Mundo em sua maioria cosmopolitas (Lachaise et al. 1988). Já as espécies Neotropicais possuem sua origem a partir de uma ancestral *protomelanogaster* que teria se estabelecido nas regiões Neotropicais por volta de 25-22 milhões de anos (Throckmorton 1975).

Dentre os dois grupos constituintes das espécies Neotropicais, o subgrupo *willistoni* é o que chama mais a atenção por conta de suas diversas peculiaridades, as quais não são encontradas em qualquer outro grupo de drosofilídeos. Em *D. willistoni* podemos destacar o elevado grau de polimorfismo cromossômico, contando com mais de 50 inversões cromossômicas descritas em amostras coletadas do sul da Flórida, na

América do Norte, até o sul da América do Sul, na região do rio La Plata (Valente et al. 2003; Rohde and Valente 2012). Espera-se uma íntima relação entre recursos ecológicos presentes em determinado nicho e variabilidade genética maior e entre variabilidade ecológica com maior polimorfismo populacional. Geralmente regiões centrais de distribuição populacional apresentam maior diversidade de nicho para exploração do que as regiões limítrofes (da Cunha et al. 1959). Porém, os indivíduos de *D. willistoni* coletados no Uruguai (região limítrofe de sua distribuição) apresentaram alto grau de polimorfismo (Valente et al. 2003), indo contra ao que se esperava. Assim, o que se sugere para esse grupo de organismos, é que no caso das regiões limítrofes de distribuição, que possuem condições ambientais menos favoráveis, o polimorfismo cromossômico acaba se tornando uma ferramenta importante para sobrepujar as adversidades impostas pelo ambiente.

Contudo, as idiossincrasias do subgrupo *willistoni* vão além do polimorfismo cromossômico. *D. willistoni* é reconhecida por também possuir um alto grau de polimorfismo enzimático, 60% aproximadamente também (Ayala and Powell 1972). A redução no desvio de uso de códons – que está relacionado com a melhoria da síntese proteica - é outra característica pertencente às espécies do subgrupo *willistoni* (Anderson et al. 1993; Clark et al. 2007).

Através da técnica MSRE (*Methylation-Sensitive Restriction Endonuclease*) em estudo realizado por nosso grupo de pesquisa do Laboratório de *Drosophila* – UFRGS, detectou-se pela primeira vez o fenômeno de metilação de citosinas em *D. willistoni*. Porém, muito mais do que observar os padrões de clivagem nessa espécie, revelou-se padrões diferentes de clivagem entre machos e fêmeas, o que não ocorre em *D. melanogaster* (Garcia et al. 2007). A presença de metilação sexo-específica em espécies do subgrupo *willistoni*, assim, é outro fenômeno peculiar a esses organismos. As sequências provenientes desse processo de clivagem de DNA por MSRE eram compostas por diferentes elementos do genoma e de rDNA. Dando continuidade e aprofundando-se nessa emergente e instigante questão que se abria, em trabalho posterior, verificou-se que *D. tropicalis* e *D. insularis* também possuem os mesmos padrões de metilação sexo-específica de rDNA encontrados em *D. willistoni*. Entretanto, em *D. paulistorum* e *D. equinoxialis*, que também apresentaram padrões de metilação sexo-específica, esses padrões não estavam restritos aos genes ribossomais (D'Ávila et al. 2010). O interessante é que espécies do subgrupo *bocainensis* (grupo *willistoni*), considerado basal ao subgrupo

willistoni (Robe et al. 2010), não apresentaram padrões de metilação sexo-específica, sugerindo que tal fenômeno tenha emergido após a divergência evolutiva desses dois grupos (D'Ávila et al. 2010). Mais recentemente, constatou-se que o percentual de conteúdo do dinucleotídeos GC (% conteúdo GC) em *D. willistoni*, *D. mojavensis*, *D. virilis* e *D. grimshawi* são substancialmente menor do que nas demais espécies do gênero *Drosophila*, entretanto os níveis de citosinas genômicas metiladas (%mC) é destacadamente elevado em comparação às demais espécies de *Drosophila*, dentro da relação %5mC / %conteúdo GC (Deshmukh et al. 2018). Por esse cenário exposto até aqui, os estudos epigenéticos nesse específico grupo de drosofilídeos se faz por demais instigante.

A Importância da Bioinformática na Resolução de Estruturas e Sistemas de Proteínas-Alvo

Métodos computacionais podem melhorar as análises de triagem, revelando o papel das mutações individuais e seu impacto na função da proteína. O entendimento e a possibilidade do estudo das funções, estruturas e mecanismos pelas quais as proteínas executam suas tarefas é um passo importante fornecido pela bioinformática. A compreensão das estruturas tridimensionais de moléculas possibilitou avanços significativos nas áreas farmacológicas e de diagnose (Santos Filho and Bicca De Alencastro 2003).

O crescente número de estruturas de proteínas resolvidas pelas técnicas como a Cristalografia de Raios-X e Ressonância Magnética Nuclear (RMN) depositadas em bancos de dados, permitiu a sistematização desses dados em projetos proteômicos amplos. As informações depositadas propiciaram um arcabouço de conhecimento das mais diversas interações moleculares e funções biológicas de inúmeras proteínas, e a partir desse instante, se tornou possível afirmar que a função de uma proteína pode ser concebida a partir de sua estrutura tridimensional, o que leva à importância de se conhecer a estrutura 3D de sequências proteicas (Sánchez and Šali 1999).

Apesar da inquestionável importância dos métodos experimentais de resolução de estruturas tridimensionais de proteínas (Cristalografia de Raios-X e RMN), essas possuem diversas limitações, como a necessidade de se obter grandes quantidades de amostras, e mesmo assim não há garantias da qualidade dos cristais oriundos do trabalho

experimental. Outro revés pode ser encontrado em determinadas classes de proteínas como as proteínas de membrana, das quais dificilmente se obtém uma cristalização que possa ser utilizada satisfatoriamente por RMN. Por conta dessas dificuldades, não é de se estranhar a discrepância existente entre o número de sequências primárias descritas (559.228 sequências depositadas no Uniprot) em contrapartida das estruturas tridimensionais depositadas em banco de dados (149.424 estruturas cristalográficas proteicas depositadas no PDB – consulta feita em março de 2019). Dessa maneira, a utilização de métodos computacionais torna-se importante para a ampliação do conhecimento e para a resolução de estruturas proteicas ainda não conhecidas (Sánchez and Šali 1999; Kiefer et al. 2009).

As simulações de dinâmica molecular também são ferramentas de bioinformática que tem colaborado significativamente no entendimento das relações existentes entre as estruturas tridimensionais proteicas com seus cofatores e substratos. Utilizando os princípios de mecânica clássica, métodos mistos (QM/MM) e quânticos, a dinâmica molecular pode fornecer dados a respeito do comportamento das moléculas em meio aquoso ou no vácuo, como por exemplo, a estabilidade da conformação de uma enzima (Pikkemaat et al. 2002) e possíveis alterações das estruturas secundárias ao longo do tempo (Campbell et al. 2016), assim como as relações de interação intramoleculares e da própria molécula com o meio.

Atualmente pode-se contar com diversos programas capazes de modelagem molecular. Em dinâmica, os mais comumente empregados são o Amber (Pearlman et al. 1995) e o GROMACS (Van Der Spoel et al. 2005). Esses programas contam com pacotes de algoritmos de análises dos dados obtidos das simulações que, conjuntamente com as averiguações visuais, permitem quantificar as mudanças conformacionais, interações interatômicas e diferentes níveis de energia ao longo das simulações. Basicamente, o algoritmo empregado nos programas de simulação de dinâmica molecular assinala posições e velocidades de cada partícula envolvida no sistema simulado, além de especificar o modelo físico que governará as interações entre as partículas. As forças podem ser computadas, e dessa maneira, os valores utilizados para atualizar a posição e velocidade das partículas segundo a segunda lei de Newton, dentro de um determinado tempo. Os modelos físicos existentes podem ser descritos em dois grupos: as interações interligadas e interações não-interligadas, onde o segundo grupo é representado por comportamentos de partículas como forças de van der Waals, ou as leis de Coulomb.

Através de tais ferramentas computacionais podemos analisar evolução molecular em diferentes níveis. O que ocorre basicamente ao se estudar os processos de evolução molecular, é que essa é conduzida primordialmente por mutações, as quais podem – ou não – modificar a processividade de uma enzima ou estruturação de uma proteína necessária para a manutenção da vida de um organismo; sendo alvo da seleção natural ou mantendo-se na população por deriva genética. É importante notar que a análise de sequências primárias e terciárias de proteínas podem trazer interessantes sugestões a respeito da evolução e da função de proteínas (Bhabha et al. 2013). Assim, as pressões evolutivas podem moldar e influenciar a dinâmica estrutural proteica em um processo adaptativo de acordo com diferentes ambientes, facilitando dessa maneira o surgimento de novas funcionalidades. Por fim, as assinaturas de processos evolutivos se expandem além da análise de substituições sinônimas e não sinônimas, mas se encontram também nas “pistas” conformacionais de novas estruturas que possam representar modificações funcionais das proteínas. Assim, empregaremos tais abordagens, em conjunto com análises filogenéticas, de hibridização *in situ* e quantificação dos níveis transcricionais de *DNMT2* na investigação dos fenômenos acima descritos.

Objetivos

Geral:

O objetivo geral do presente trabalho foi estudar os aspectos moleculares e evolutivos do gene e da enzima DNA metiltransferase 2 (DNMT2), além do fenômeno de metilação em âmbito funcional e genômico em diferentes espécies do gênero *Drosophila*, com especial atenção à espécie *Drosophila willistoni*, com vistas a contribuir para a melhor compreensão do complexo padrão de metilação sexo-específico e sua relação com a importância ontogenética e o histórico evolutivo envolvido nos organismos que possuem a referida peculiaridade.

Específicos:

- Avaliar a conservação e evolução molecular de *Dnmt2*, seu contexto evolutivo, fatores preponderantes e potenciais sistemas genéticos interagentes (**Capítulo II**);

- Analisar e quantificar os níveis relativos de expressão transcricional em diferentes estágios de desenvolvimento de *Dnmt2* e sua localização física em cromossomos politênicos de *Drosophila willistoni* (**Capítulo III**);
- Relacionar aspectos evolutivos e funcionais da família DNMT2 de drosofilídeos através de estudos comparativos das propriedades físico-químicas com demais estruturas cristalográficas resolvidas de DNMT2 de outros organismos (**Capítulo IV**);
- Analisar e caracterizar o comportamento das DNMT2 em simulação de dinâmica molecular quando na presença dos ligantes DNA (**Capítulo V**).

CAPÍTULO II

Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions

Gilberto Cavalheiro Vieira^{1,3}, Marícia Fantinel D'Ávila², Rebeca Zanini⁴, Maríndia Deprá^{1,4}, Vera Lúcia da Silva Valente^{1, 2, 4}

¹Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

²Departamento de Zoologia e Ciências Biológicas, Universidade Federal de Santa Maria (UFSM), Palmeira das Missões, RS, Brasil.


³ Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

⁴ Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

Artigo publicado no periódico *Genetics and Molecular Biology* (2018) 41(1 suppl 1):215-234



Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions

Gilberto Cavalheiro Vieira^{1,3} , Marícia Fantinel D'Ávila², Rebeca Zanini³, Maríndia Deprá^{1,4} and Vera Lúcia da Silva Valente^{1,2,4}

¹Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

²Departamento de Zoologia e Ciências Biológicas, Universidade Federal de Santa Maria (UFSM), Palmeira das Missões, RS, Brazil.

³Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

⁴Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil.

Abstract

The DNA methyltransferase 2 (DNMT2) protein is the most conserved member of the DNA methyltransferase family. Nevertheless, its substrate specificity is still controversial and elusive. The genomic role and determinants of DNA methylation are poorly understood in invertebrates, and several mechanisms and associations are suggested. In *Drosophila*, the only known DNMT gene is *Dnmt2*. Here we present our findings from a wide search for *Dnmt2* homologs in 68 species of Drosophilidae. We investigated its molecular evolution, and in our phylogenetic analyses the main clades of Drosophilidae species were recovered. We tested whether the *Dnmt2* has evolved neutrally or under positive selection along the subgenera *Drosophila* and *Sophophora* and investigated positive selection in relation to several physicochemical properties. Despite of a major selective constraint on *Dnmt2*, we detected six sites under positive selection. Regarding the DNMT2 protein, 12 sites under positive-destabilizing selection were found, which suggests a selection that favors structural and functional shifts in the protein. The search for new potential protein partners with DNMT2 revealed 15 proteins with high evolutionary rate covariation (ERC), indicating a plurality of DNMT2 functions in different pathways. These events might represent signs of molecular adaptation, with molecular peculiarities arising from the diversity of evolutionary histories experienced by drosophilids.

Keywords: *Drosophila*; Dnmt2; positive selection; epigenetic; positive-destabilizing selection.

Received: February 26, 2017; Accepted: June 18, 2017.

Introduction

Methylation of cytosine to form 5-methylcytosine is one of the most important epigenetic marks acting in the control of gene expression without altering the DNA nucleotide sequence. Cytosine methylation plays a critical role in the regulation of gene expression in higher eukaryotes. It is established by DNA methyltransferases (DNMTs), classified into three subfamilies: DNMT1, DNMT2 e DNMT3. The smallest eukaryotic methyltransferase, DNMT2, is most widely distributed in animals, fungi, protists, and plants (Ponger and Li, 2005). Dnmt2 was first identified in

mice and humans and appears to be well conserved among eukaryotes (Okano *et al.*, 1998; Yoder and Bestor, 1998). This enzyme is the only DNMT found in dipterans, including *Drosophila* (organisms “Dnmt2-only”) (Kucharski *et al.*, 2008). In agreement with this structural conservation, different methods in various organisms have shown DNMT2 to have DNA methyltransferase activity (Hermann *et al.*, 2003; Tang *et al.*, 2003; Fisher *et al.*, 2004; Katoh *et al.*, 2006). However, according to analyses of human and *Entamoeba* enzymes, its catalytic activity on DNA is very weak (Hermann *et al.*, 2003; Fisher *et al.*, 2004). In addition, cytosine methylation of non-coding RNA (ncRNA) plays an important role in the epigenetic landscape. The functions of some tRNA modifications remain obscure. However, Tuorto *et al.* (2012) have shown that cytosine-C5 methylation of tRNAs is associated with their structural stability and the rates of protein synthesis in

Send correspondence to Gilberto Cavalheiro Vieira, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500 - Building 43323, room 210, postal code 15053, Porto Alegre, RS, Brazil. E-mail: cavalheiro.rs@gmail.com.

mammals. Cytosine-C5 methylation is involved in protecting tRNA against degradation induced by cellular stress events (Saikia et al., 2012).

Furthermore, studies have associated DNMT2 with RNA interference in *Dictyostelium* (Kuhlmann et al., 2005) and covalent histone modification in *Drosophila* (Kunert et al., 2003), suggesting a role of DNMT2 in epigenetic regulation. Several studies have shown the occurrence of DNA methylation phenomena in *Drosophila* species (Lyko et al., 2000; Kunert et al., 2003; Marhold et al., 2004). However, recently, Takayama et al. (2014) showed that methylation in the genome of *Drosophila melanogaster* probably is independent of DNMT2 activity. These findings brought new questions about the epigenetic mechanisms involved in the methylation process in evolutionarily related species of drosophilids.

Garcia et al. (2007) compared the DNMT2 protein sequences of *D. willistoni* and *D. melanogaster* and found higher conservation at the domains putatively responsible for methyl transfer catalysis and variability in the region containing the specific target recognition domain (TRD). These findings may be indicative of variation in DNMT2 function among organisms, suggesting that the targets – or modularity – of methylation can also vary among species of the same genus. Furthermore, Garcia et al. (2007) described sex-specific methylation patterns in *D. willistoni*, not present in *D. melanogaster*. Using the Methylation-Sensitive Restriction Endonuclease (MSRE) technique and Southern blot analysis with specific probes, the results suggested that selection for different targets of methylation might occur between different, but closely related species (Garcia et al., 2007). Furthermore, D'Ávila et al. (2010) found phylogenetic correlations in the sex-specific methylation patterns in the species of the *willistoni* subgroup, where *D. willistoni*, *D. tropicalis* and *D. insularis* (closer related species) shared methylation patterns in ribosomal genes, whereas *D. equinoxialis* and *D. paulistorum* patterns apparently are not restricted to rDNA.

The presence of DNMT2 enzymes in Diptera was described by Marhold et al. (2004), revealing that DNMT2 protein sequences are highly conserved in *D. virilis*, *D. hydei*, *D. simulans*, *D. melanogaster* and *D. pseudoobscura*, primarily within the catalytic DNA methyltransferase motifs. The Drosophilidae family is among the most diverse of the Diptera, encompassing more than 4,200 species (Bächli, 2016). Species of this family, especially of the genus *Drosophila*, are widely used in many areas of contemporary biological research. However, only few have been investigated with respect to the occurrence of DNA methylation and the presence of the *Dnmt2* gene. Thus, the present study objectives are: (i) improve the previous search for *Dnmt2* (Marhold et al., 2004; Garcia et al., 2007), including a large number of *Drosophila* species and other Drosophilidae, to evaluate the conservation, or not, of *Dnmt2* in the genus; (ii) test whether the gene and protein

are evolving under relaxed selective constraint or positive selection; and (iii), given the current controversial and enigmatic scenario involving the role of DNMT2 among epigenetic mechanisms of drosophilids, an attempt to find potential protein interaction partners of DNMT2 by database searching for protein-protein interactions and via evolutionary rate covariation (ERC) analysis.

Our results indicate, as expected, substantial conservation of DNMT2 catalytic motifs. Nevertheless, the TRD and the connecting region of the two main domains (catalytic and TRD) show some variability among the species examined, including closely related species. We also detected that several sites are under positive selection. These are located in potential regions of protein-protein interaction. The multiplicity of proteins with high ERC values found in the present work supports the hypothesis of DNMT2 can be involved in several networks, through control of gene expression, genomic stability, and in response to stressor events in “Dnmt2-only” organisms, like drosophilids.

Material and Methods

Fly stocks

The conservation of DNMT2 in the family Drosophilidae was analyzed in 68 species of *Drosophila*, along with *Zaprionus indianus*, *Z. tuberculatus*, *Scaptodrosophila latifasciaeformis* and *S. lebanonensis* (Table S1). Most strains were maintained in the laboratory by mass crosses and reared in corn flour culture medium (Marques et al., 1966) in a controlled environment chamber (17 ± 1 °C, 60% r.h.), except for those species for which sequences were obtained directly from GenBank.

PCR, cloning and sequencing

Genomic DNA was extracted from adult flies following Sassi et al. (2005). Primers initially used are described in Marhold et al. (2004) (*5' D. melanogaster Dnmt2-F: 5' GTGGCATTGGCGGCATGCATTATGCCT 3'* and *D. melanogaster Dnmt2-R: 5' CGATACTTTTGTTCGATTCGTTGTTCTGGC 3'*). This pair of primers was designed directed against conserved catalytic motifs of *D. melanogaster* and used to amplify *Dnmt2* sequences from *D. simulans*, *D. hydei* and *D. virilis* (Marhold et al., 2004). In this work, specific primers were designed to *D. willistoni Dnmt2* genes (*wDnmt2A-F: 5' TCACCCACAACCTTGACATT 3'* and *wDnmt2C-R: 5' ACCTTCTCGCAGACACCAA 3'*). Both pairs of primers align in similar regions of the gene *Dnmt2*. PCRs were performed in 25 µL volumes containing 20 ng of genomic DNA, 1 U of *Taq* DNA polymerase (Invitrogen, Carlsberg, CA, USA), 1X reaction buffer, 200 µM of each nucleotide, 20 pmol of each primer and 1.5 mM MgCl₂. The amplification sets consisted of a denaturation step of 95 °C for 5 min, followed by 30 cycles at 95 °C for 40 s, 55 °C to 60 °C for

40 s and 72 °C for 1 min, and then a final extension cycle at 72 °C for 5 min. *Dnmt2* amplicons were directly purified by incubation at 37 °C for 30 min with Exonuclease I and Shrimp Alkaline Phosphatase (SAP) (both from USB, Cleveland, OH, USA) followed by a 15 min inactivation step at 80 °C. For the *Dnmt2* amplicons of *D. teissieri*, *D. ornatifrons*, *D. ornatipennis* and *D. tropicalis*, fragments were excised of the agarose gel and purified using Illustra GFX PCR DNA kit (GE Healthcare, Pittsburgh, PA, USA). The purified fragments were cloned into pCR4-TOPO plasmids (Invitrogen). DNA sequencing was performed by Macrogen Inc. (South Korea) using the appropriate primers (forward and reverse). The sequences generated by PCR and sequencing were assembled using the GAP 4 software of the Staden Package (Staden, 1996) (Table S2).

Dot blot analyses

For Dot blot hybridizations, samples of denatured DNA (1 µg) were transferred onto a nylon membrane (Hybond-N+; GE Healthcare Biosciences). The AlkPhos Direct Labelling and Detection System and the CDP-Star kit (GE Healthcare) were used to label and detect nucleic acids according to the manufacturer's instructions. The *Dnmt2* amplicon from *D. melanogaster* was used as a probe at the stringency temperature of 55 °C.

Data collection from public databases

In silico searches were performed to identify the complete sequence homologs of *Dnmt2* among 24 sequenced *Drosophila* genomes available in the FlyBase database, using the *Dnmt2* of *D. melanogaster* (Accession number: AAF53163.2) as query (Table S2). The *D. buzzatii* and *D. suzukii* *Dnmt2* genes were obtained from the *Drosophila buzzatii* Genome Project server and Spotted Wing Fly Base, a dedicated online resource for *D. suzukii* genomics (Table S2).

Evolutionary analysis

All sequences were aligned using the Muscle tool (Edgar, 2004). The evolutionary relationships among the *Dnmt2* sequences were estimated using Bayesian analysis, which is implemented in MrBayes (Ronquist and Huelsenbeck, 2003), with the evaluation of at least 1,000,000 generations and a burn-in region of 2,500 trees. Each nucleotide sequence was individually translated into its corresponding proteins and aligned using the Muscle tool with default parameter values. For the evolutionary analysis of the amino acid sequences, the Jones-Taylor-Thornton (JTT+G) model was used, as suggested by ProtTest 2.2 (Abascal *et al.*, 2005), in accordance with the Akaike information criterion (Akaike, 1974). A Bayesian analysis of the nucleotide sequences was performed with the general time reversible (GTR) model using the ratio of invariable sites (I) and the gamma distribution of the variable sites (G) model, as suggested in MrModel Test 2.3 (Nylander,

2004). The sequence of *Spodoptera frugiperda* was used as outgroup. For the nucleotide and amino acid divergence analyses, sequences were clustered within species groups to perform a p-distance analysis using the MEGA 7 package (Kumar *et al.*, 2016).

Analysis of positive selection

To investigate probable selective pressures that shaped the evolution of drosophilid *Dnmt2* genes, we performed a relaxed branch-site test and a strict branch-site test (Yang and Nielsen, 2002; Zhang *et al.*, 2005) using the software CODEML (Yang, 2007) - PAML package. This software tests for positive selection by comparing a series of alternative hypothesis that differ in how variable dN/dS ratio can change in different branches and codons, in which dN/dS > 1 would indicate positive selection and dN/dS < 1 would indicate a purifying selection, due to a selective constraint at the codon level.

The complete sequences of *Dnmt2* genes were used, obtained from *in silico* search, as previously described. The initial maximum likelihood (ML) phylogenetic trees were constructed using the 24 complete sequences (Table S1) by the software PHYML 3.1 (Guindon and Gascuel, 2003). The analysis of the nucleotide sequences was performed with the general time reversible (GTR) model using the ratio of invariable sites (I) and the gamma distribution of the variable sites (G) model, as suggested by MrModel Test 2.3 software (Nylander, 2004).

With the relaxed branch-site test or strict branch-site test, phylogenetic trees are separated into *foreground* branches, at which positive selection is tested, and *background* branches, represented by the other lineages. Both tests use the alternative model (MA), in which the codons in *foreground* are allowed to have a dN/dS > 1, and the *background* codons, dN/dS ≤ 1. The relaxed branch-site test null model (M1a) assumes that evolutionary rates are the same for all sites and branches, with all sites varying dN/dS from 0 and 1. In the null model (restricted MA), the dN/dS > 1 category is fixed to 1, so all sites with dN/dS > 1 are forced to evolve neutrally (dN/dS = 1). We used a log likelihood ratio test (LRT) to infer the positive selection when these values result in a significant value. The significance of the LTR was verified by a χ^2_2 null distribution, with critical values of 2.71 for 5% and 5.41 for 1% significance levels, respectively, originated from a null distribution composed of a 50:50 mixture of point mass 0 and χ^2_1 (Zhang *et al.*, 2005).

Additionally, we investigated positive selection with respect to several physicochemical properties of the datasets. The MM01 method of McClellan *et al.* (2005) evaluates whether nonsynonymous substitutions favored changes in protein, either structural or functional. The analyses were carried out by *TreeSAAP* 3.2 (McClellan and McCracken, 2001; Woolley *et al.*, 2003; McClellan *et al.*, 2005). First, global deviation from neutrality is verified by

a goodness-of-fit test, in which a comparison of neutral expected distribution and observed distribution of the selected physicochemical properties is made. Positive selection is detected in *TreeSAAP* software when the number of inferred amino acid replacements significantly exceeds the number of expected replacements caused by chance alone, given positive z-scores. Stabilizing-selection can be visualized when the magnitude of change is low (categories 1, 2 and 3), meaning it is a conservative process, while positive-destabilizing selection is represented as a high magnitude of change (categories 6, 7 and 8) (McClellan et al., 2005). Stabilizing selection is defined by McClellan et al. (2005) as a selection that tends to maintain the original biochemical attributes of the protein, and positive-destabilizing selection as a selection that favors structural and functional shifts in a region of a protein. In other words, positive-destabilizing selection represents signs of molecular adaptation. To verify which regions were under positive selection (stabilizing and destabilizing) we performed a sliding window analysis using the amino acid properties significant for this type of change (McClellan et al., 2005).

Potential protein-protein interaction partners

Using the database STRING (Szklarczyk et al., 2015) we conducted a search for predicted protein-protein interactions with DNMT2. To perform the search, the *D. melanogaster* DNMT2 sequence was used as query and data were collected from *D. ananassae*, *D. grimshawi*, *D. pseudoobscura*, *D. virilis*, *D. willistoni* and *D. yakuba*. We also attempted to evaluate the predicted protein-protein partners from the STRING search and find protein partners with DNMT2 by evolutionary rate covariation (ERC) using the ERC Analysis Webserver (Clark et al., 2012, 2013, Findlay et al., 2014).

Results

Detection of *Dnmt2* sequences in Drosophilidae species

In a preliminary screen of the presence of *Dnmt2* sequence homologs within the *Drosophila* genomes, we tested 56 species by dot blot analysis (Table S1 and Figure S1). Of these, 54 showed a positive signal for the *Dnmt2* probe (amplicon from *D. melanogaster*), two of which had a weak signal (*D. oreana* and *D. polymorpha*). In this assay we observed a strong hybridization signal, primarily in the *melanogaster* group. Nevertheless, hybridization was also detected in the other species, indicating that the *Dnmt2* gene has related sequences in all species groups analyzed.

To verify this homology, we also performed a *Dnmt2* homolog search by PCR amplification in a large number of Drosophilidae species from different *Drosophila* groups (Table S1). Altogether, 61 species were tested by PCR for presence of a *Dnmt2* gene. Thirty species tested positive for *Dnmt2* by PCR. We achieved 20 amplicons for several spe-

cies of the *Drosophila* genus, including the *Drosophila* subgenus (*guarani*, *guaramunu*, *tripunctata*, *calloptera*, *immigrans*, *mesophragmatica*, *flavopilosa* and *repleta* groups) and the *Sophophora* subgenus (*melanogaster* and *willistoni* groups). The representative species of the *Dorsilopha* subgenus did not show *Dnmt2* amplification with the primers used. The annealing regions of the primers correspond to motif I (forward primer) and motif X (reverse primer) of *Dnmt2*, both belonging to the catalytic domain of the enzyme (Figure S2). The length of the amplified fragments was around 800 bp.

DNMT2 conservation

The availability of several species maintained in culture chambers in our laboratory and genomic sequences in gene banks allowed us to search for open reading frames encoding DNMT2 in several genomes. For the analyses of gene conservation, all cloned sequences (obtained by direct PCR or cloned) (Table S3) and those obtained by *in silico* search were used in the phylogenetic analysis to investigate the evolutionary pattern and conservation of *Dnmt2* among Drosophilidae species. The sequences were aligned to build a phylogenetic matrix from the 44 analyzed species (Figure S2).

To perform the analyses of *Dnmt2* nucleotide and amino acid divergence, pairwise comparisons of the sequences were conducted (with species clustered by taxonomic group). The highest nucleotide divergence found was between the *willistoni* and *tripunctata* groups, with a p-distance value of 34.47%; whereas the lowest divergence (p-distance value of 10.70%) was detected between the *guarani* and *calloptera* groups (represented only by *D. ornatipennis*) (Table 1, standard error presented at Table S4). The comparison of all nucleotide sequences between species showed that the sequences of *D. willistoni* and *D. suzukii* are more divergent than any other, with a p-distance value of 35.3% (data not shown). When we estimated the average divergence of amino acid sequence pairs within subgenus *Drosophila* and *Sophophora*, it revealed the almost the same internal divergence: 17.8% (S.E. 1.65) and 18.7% (S.E. 1.62), respectively. Pairwise alignment of the amino acid sequences between groups (Table 1) revealed that *willistoni* and *mesophragmatica* groups were more divergent (p-distance value of 31.84%).

We used the MEGA7 package to compute the mean evolutionary rates at each DNMT2 site. Figure 1 shows the plots, which evolutionary rate are shown to each site for all Drosophilidae complete sequences and for each subgenus separately. Several DNMT2 sites are under evolutionary restrictions, and these sites are mainly part of the catalytic motifs of the enzyme. Nevertheless, 10 regions with sites having high evolutionary rates distributed throughout DNMT2 are evident (Figure 1). This means that sites showing a rate < 1 are evolving slower than average and those with a rate > 1 are evolving faster than average evolutionary

Table 1 - Estimates of evolutionary divergence between sequence pairs of different Drosophilidae species groups. The numbers of amino acid differences per site from the average over all sequence pairs between groups are given below the diagonal. The measures of nucleotide evolutionary divergence are provided above the diagonal. The p-distances are given in percentages.

Groups		1	2	3	4	5	6	7	8	9	10	11	12	13
1	<i>calloptera</i>		25.31	26.39	12.89	10.70	26.08	30.53	23.66	29.58	23.87	15.66	22.22	33.33
2	<i>flavopilosa</i>	23.32		23.30	24.23	25.77	26.77	30.83	15.12	29.32	16.62	25.62	16.20	32.18
3	<i>grimshawi</i>	24.22	18.83		25.31	26.13	26.23	28.62	22.79	27.11	23.15	26.39	20.37	33.95
4	<i>guaramunu</i>	10.09	21.08	23.54		15.15	25.23	29.27	22.97	26.72	23.46	13.87	21.14	34.07
5	<i>guarani</i>	7.92	24.81	23.92	11.29		25.36	30.54	23.56	29.15	24.74	17.16	22.89	32.95
6	<i>immigrans</i>	17.94	20.63	21.08	17.60	18.76		29.89	25.69	28.50	27.93	25.31	25.08	30.83
7	<i>melanogaster</i>	25.16	25.41	24.22	24.63	25.95	23.56		29.87	23.72	29.78	29.93	29.24	32.72
8	<i>mesophragmatica</i>	20.78	15.70	19.13	18.68	21.72	19.58	25.29		28.46	15.35	24.34	15.64	33.64
9	<i>obscura</i>	25.56	26.01	22.57	24.29	24.81	23.02	19.11	25.26		28.91	28.10	26.23	32.61
10	<i>repleta</i>	21.82	14.95	18.09	19.88	22.77	20.40	25.15	11.61	25.66		25.31	18.57	33.02
11	<i>tripunctata</i>	12.00	23.09	23.99	10.48	12.86	18.89	25.26	20.67	24.96	21.56		22.69	34.47
12	<i>virilis</i>	18.39	11.66	15.25	15.92	18.98	15.70	22.22	10.31	21.52	12.41	17.94		30.56
13	<i>willistoni</i>	31.17	31.17	28.48	30.72	31.76	27.91	29.55	31.84	27.20	30.19	31.22	28.92	

Standard error values are shown in Table S5.

rates. Evolutionary rates were estimated under the Jones-Taylor-Thornton (Jones *et al.*, 1992) model (+G).

The phylogeny obtained by amino acid analysis recovered the evolutionary relationship between the subgenera *Drosophila* and *Sophophora* (Figure 2). The *quinaria* section (*Drosophila* subgenus) composed of *guarani*, *guaramunu*, *calloptera* and *tripunctata* species groups appears as basal radiation. A second cluster in *Drosophila* subgenus includes the *virilis*, *repleta*, *flavopilosa* and *mesophragmatica* species groups (*virilis-repleta* radiation). The *Sophophora* subgenus is composed of the *melanogaster*, *willistoni* and *obscura* groups, where the *willistoni* group appears as basal to the *obscura-melanogaster* radiation. *D. ananassae* and *D. bipunctata* appear more externally positioned within the *melanogaster* group, likely reflecting the ancestral condition of the *ananassae* subgroup (to which both species belong) within this group (Kopp, 2006; Clark *et al.*, 2007). This placement is confirmed by the correct positioning of the remaining species of the *melanogaster* group, belonging to the *melanogaster*, *takahashii*, *rhopaloa* and *elegans* subgroups, consistent with previous studies (Lewis *et al.*, 2005; Kopp, 2006).

The nucleotide Bayesian inference tree of *Dnmt2* (Figure S3) showed congruence with the tree previously obtained for the Drosophilidae, which was based on nuclear genes and produced using similar methods (Gailey *et al.*, 2000; Tarrío *et al.*, 2001; Da Lage *et al.*, 2007). However, *D. immigrans* and *D. albicans* sequences were incorrectly positioned, grouping with the *Sophophora* subgenus, but with low statistical support. Also, the *quinaria* section presented polytomy. There is no strong evidence to sustain this finding, and the data from amino acids analysis confirm the correct positioning of this species, grouped with the rest

of the *Drosophila* subgenus species, according to previous studies (Throckmorton, 1975; Russo *et al.*, 1995; Kwiatkowski and Ayala, 1999), however, again some clades showed low statistical support and the *quinaria* section species grouped with polytomy.

Long-term evolutionary analysis in drosophilids *Dnmt2*

To evaluate positive selection in *Dnmt2*, we selected representative species of the main groups of drosophilids whose complete sequences could be obtained (Figure 3). Using the PAML package we obtained an initial tree that was later analyzed for positive selection in codeML. The initial ML tree recovers the evolutionary relationship between *Drosophila* and *Sophophora* subgenera (Figure 4), as well as the amino acid tree obtained from the Bayesian analysis (Figure 2). We used the results of the divergence analysis (Table 1) to establish which branch to use as *foreground*. When analyzing the results of divergence, we found that the divergence within the *Sophophora* subgenus is lower than the divergence between the *Sophophora* and *Drosophila* subgenera, as expected (Table S5). Moreover, the *Sophophora* and *Drosophila* subgenera show different evolutionary rates, with the *Sophophora* subgenus having more sites with fast evolution (Figure 1). Therefore, we asked whether the divergence observed between the groups is random or driven by selection. Thus, we established the *Sophophora* subgenus as *foreground* and the *Drosophila* subgenus as *background*.

The selective pressures over the *Dnmt2* sequence were investigated by the ratio of nonsynonymous to synonymous substitutions. Table 2 shows the parameters inferred for the null models (M1a and MA) and for the alternative

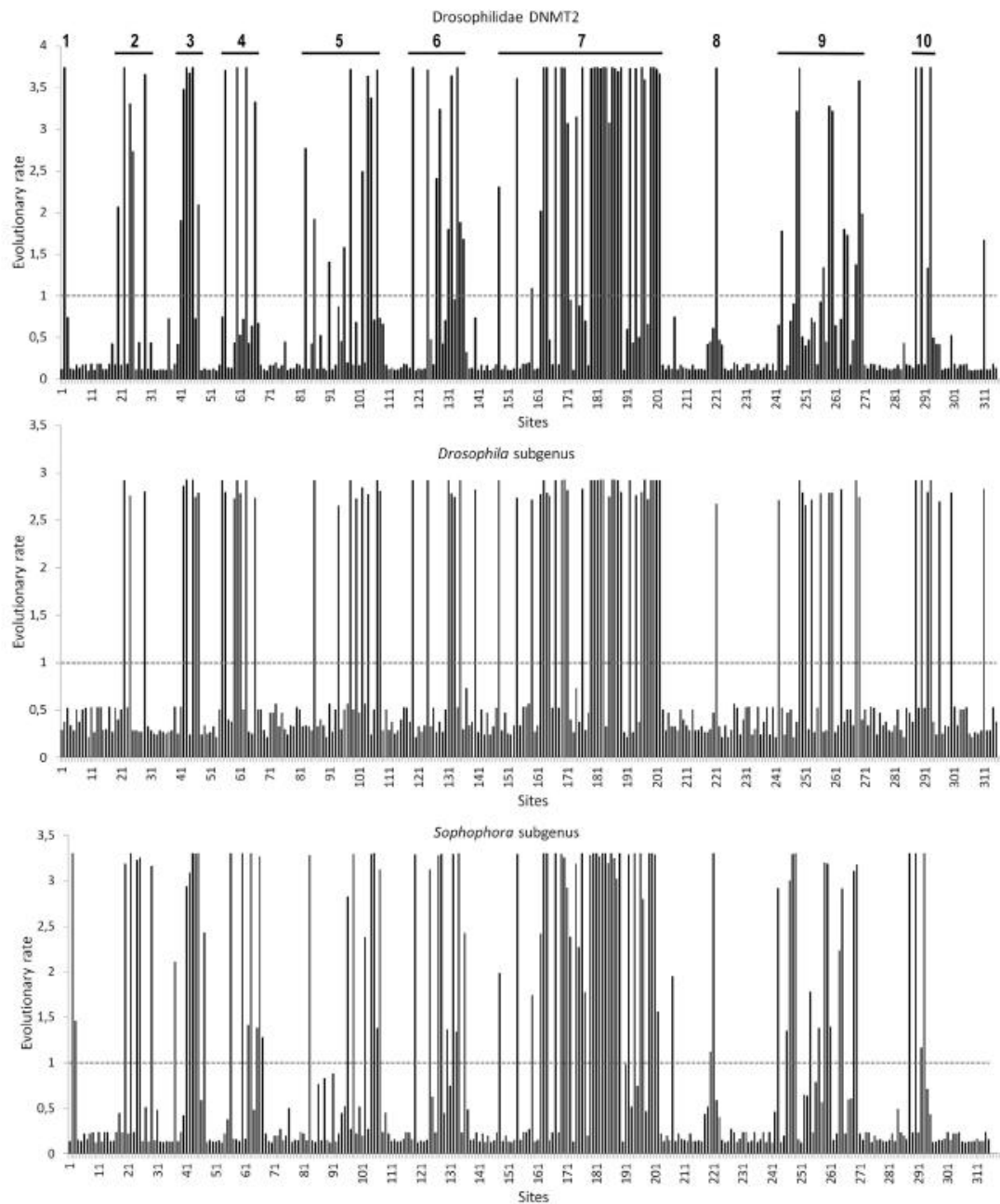


Figure 1 - Evolutionary rate from all Drosophilidae DNMT2 complete sequences, *Sophophora* and *Drosophila* subgenus separately. The rates are scaled such that the average evolutionary rate across all sites is 1. All positions containing gaps and missing data are eliminated.

MA model. The null model of selective constraint ($0 < \omega \leq 1$) was rejected by the relaxed branch-site test, indicating that the *foreground* branch (Figure 3) has diverged by re-

laxed selective constraint or by positive selection (Table 3). By the contrast of the restricted-MA versus MA models (*strict branch-site*), we could discriminate between the two

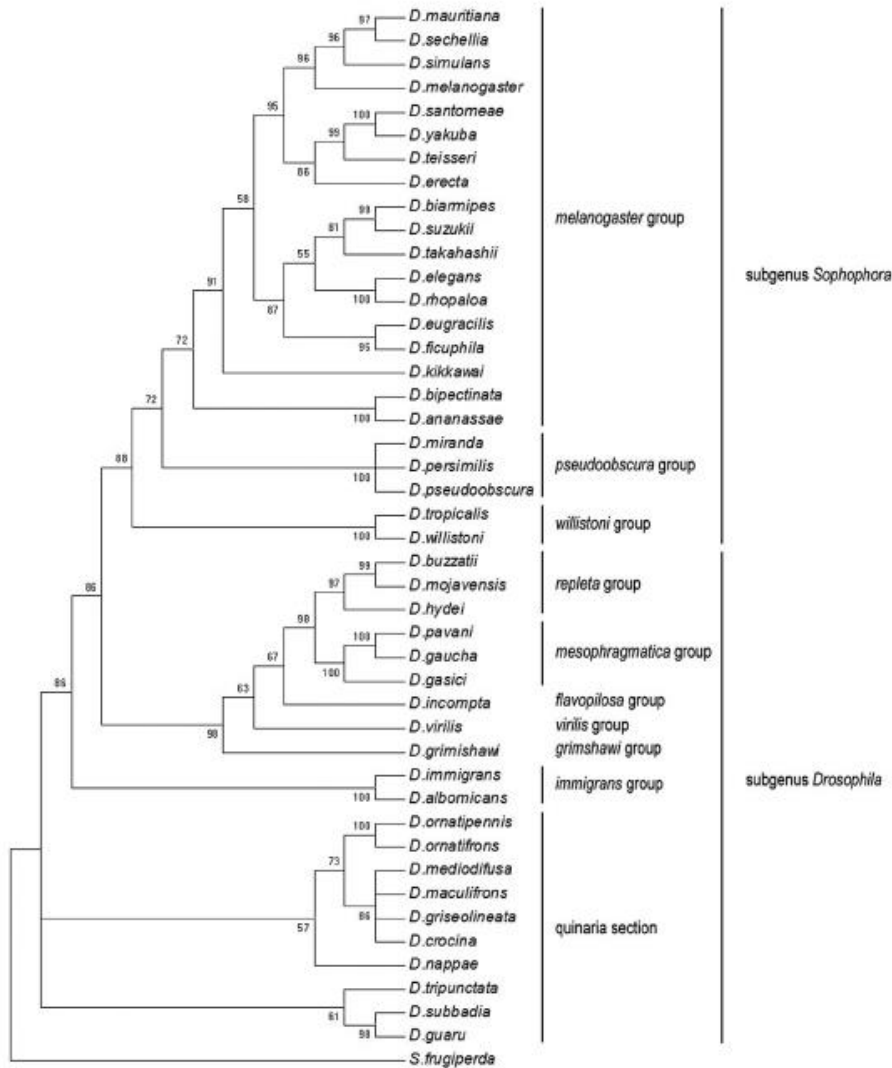


Figure 2 - Bayesian phylogenetic inference of the *Dnm2* gene in Drosophilidae species based on amino acid sequences alignment. Tree generated using the JTT model with a gamma distribution. Sequence of *Spodoptera frugiperda* was used as outgroup.

hypotheses, where the null model of selective constraint was rejected (Table 3), in favor of the hypothesis that several sites of *Sophophora* DNMT2 (Figure 3 and Figure 4) differentiated by positive selection. We estimated that six sites (position 85, 94, 109, 192, 257 and 311) were evolving under positive selection (Table 3) (Figure 3).

To investigate selection on amino acid properties we used the software *TreeSAAP*, based on the global goodness-of-fit statistics calculated by the *MM01* method. All physicochemical properties examined are significant (cut-off = 0.05) (Table 4). Seven properties demonstrated significantly positive z-scores under a trait of radical changes

category between 6 and 8: α -helical tendencies ($P\alpha$), Equilibrium constant (ionization of COOH) (pK^*), Polar requirement (Pr), Power to be at the C-terminal (αC), Power to be at the middle of alpha-helix (αm), Power to be at the N-terminal (αn) and Turn tendencies (Pt). Specific analysis with the sliding window in *TreeSAAP* showed that 12 amino acids were under positive-destabilizing selection (Figure 5a); the properties of most of these are related to the alpha-helix structures, mainly located in the catalytic domain.

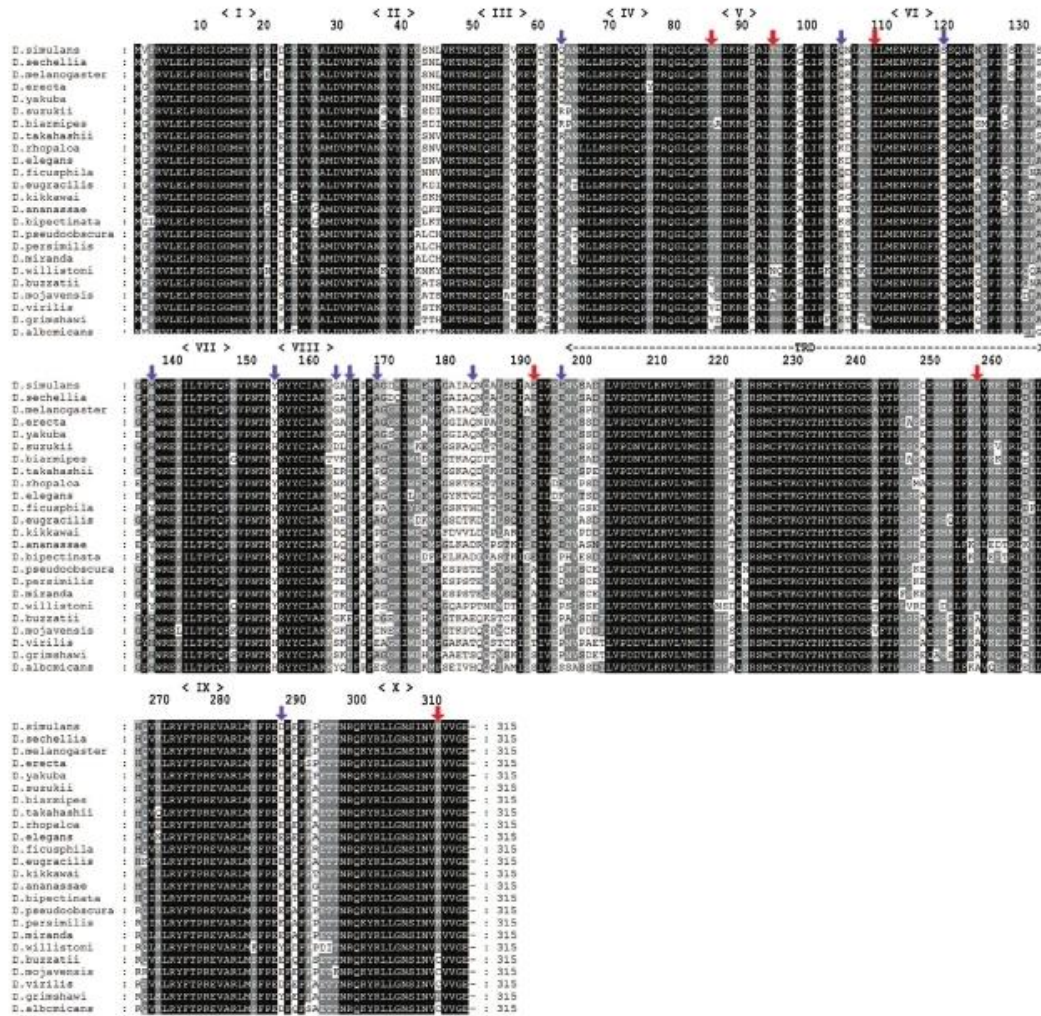


Figure 3 – Multiple alignment of DNMT2 sequences used in the positive selection analysis with no gaps. Black boxes represent conserved 100% in all sequences, dark grey 80% and light grey 60%. Red arrows indicate sites under positive selection and purple arrows sites under positive-destabilizing selection. Roman letters indicate the conserved (cytosine-5) MT2 motifs. TRD indicates the target recognition domain.

The DNMT2 sites that are under positive selection and positive-destabilizing selection (as indicated by the CodeML and TreeSAAP analyses, respectively) are shown in Figure 5b. Most sites under selection are located in the catalytic domain and especially located on the molecular surface. The only crystallographic model of DNMT2 available for arthropods is from *S. frugiperda* and it is used just as a representative model only (Li et al., 2012).

Protein-protein interactions and co-evolutionary predictions

We accessed the predicted genes as interacting with *D. melanogaster*, *D. ananassae*, *D. grimshawi*, *D. pseudo-*

obscura, *D. virilis*, *D. willistoni* and *D. yakuba Dnmt2* through the STRING database (Table 5). These were concatenated and the *D. melanogaster* gene annotations to ERC analysis were used, since the taxonomic group utilized in ERC Analysis Webserver is *D. melanogaster*. We also made an extensive search for genes that have significant ERC values with *D. melanogaster Dnmt2* by ERC Analysis Webserver Top Genes tool, which retrieves the genes with the highest ERC values for a given query gene from the entire genome (Clark et al., 2012., 2013, Findlay et al., 2014).

Altogether, 551 genes were obtained with an ERC value > 0.400 (p-value < 0.05) as result of the search. We

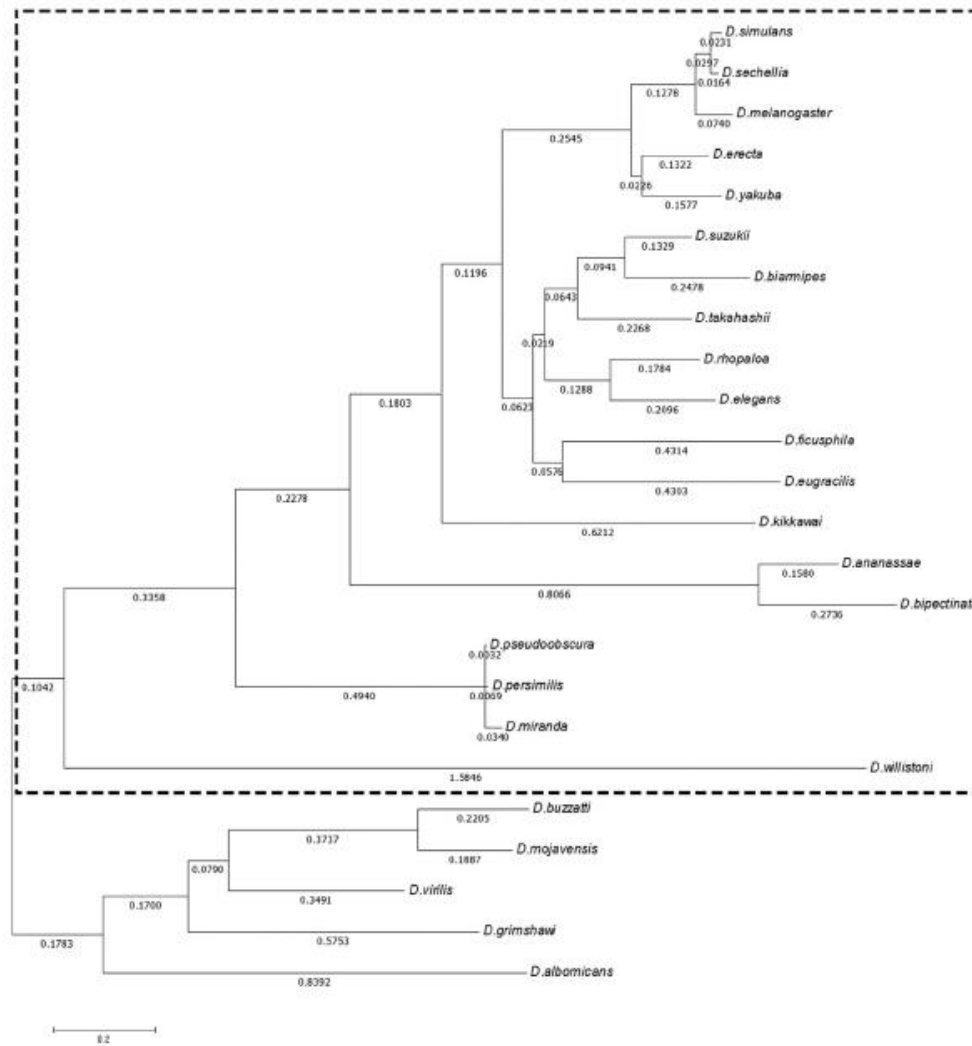


Figure 4 - ML phylogenetic tree of *Dnmt2* used in analysis for detection of selection. The dotted frame highlights the *foreground* branch. The genetic distances are in nucleotide substitution per codon (number below the branches).

Table 2 - Parameters estimates and log likelihood values for branch-site M1a, MA and restricted MA model.

Model	Parameters			lnL
M1a	$\omega_0 = 0.04651$	$p_0 = 0.86743$		-11 046.650097
Relaxed branch-site test null model	$\omega_1 = 1.00000$	$p_1 = 0.13257$		
Restricted MA ($\omega_2 = 1$ fixed)	$\omega_0 = 0.04520$	$\omega_2 = 1.000$		-11 039.638028
Strict branch-site test null model	$p_0 = 0.82646$	$p_1 = 0.12587$	$(p_{2a} + p_{2b}) = 0.04767$	
MA	$\omega_0 = 0.04596$	$\omega_2 = 50.90099$		-11 032.575669
Alternative model	$p_0 = 0.84965$	$p_1 = 0.12542$	$(p_{2a} + p_{2b}) = 0.02493$	

lnL = log likelihood; ω_0 = dN/dS values for sites with $0 < \omega < 1$; ω_1 = dN/dS values fixed to 1; ω_2 = dN/dS values for sites with $\omega > 1$, which corresponds only to sites in the foreground branch.

Table 3 - Comparison of null and alternative models by LRT and positively selected sites estimated by Bayes Empirical Bayes.

Test	Contrast	LRT	D.F.	χ^2 - Probability	Positively selected sites
Relaxed branch-site	M1a MA	25.20	2	$p < 7.7 \times 10^{-7}$	
Strict branch-site	Restricted MA MA	12.41	1	$p < 0.001$	85, 94, 109, 192, 257, 311

LRT - likelihood ratio test; D.F - Degrees of freedom.

Table 4 - Amino acid physicochemical properties under positive destabilizing selection in DNMT2.

Physicochemical property	Goodness-of-fit (neutral expectation)	radical change category (6, 7 and 8)	z-score
Alpha-helical tendencies (P_α)	33.339***	8	4.292***
Average number of surrounding residues (N_s)	119.136***		6.442***
Beta-structure tendencies (P_β)	30.909***		2.804**
Bulkiness (B_i)	35.052***		3.104***
Buriedness (B_r)	54.407***		2.580**
Chromatographic index (R_f)	107.109***		3.190***
Coil tendencies (C_i)	18.839**		-
Composition (C)	43.179***		3.898***
Compressibility (K^b)	22.548**		-
Equilibrium constant (ionization of COOH) (pK')	79.39***	8	2.828**
Helical contact area (C_s)	81.079***		3.732***
Hydropathy (H)	83.046***		3.885***
Isoelectric point (pH_i)	45.283***		4.261***
Long-range non-bonded energy (E_l)	78.459***		5.712***
Mean r.m.s. fluctuation displacement (F)	117.708***		5.736***
Molecular volume (M_v)	77.479***		3.151***
Molecular weight (M_w)	69.846***		3.577***
Normalized consensus hydrophobicity (H_{nc})	57.459***		2.062*
Partial specific volume (V^p)	61.97***		2.986**
Polar requirement (P_r)	29.801***	7	2.341**
Polarity (P)	59.713***		2.522**
Power to be at the C-terminal (α_c)	118.516***	6	5.758***
Power to be at the middle of alpha-helix (α_m)	49.271***	7	3.657***
Power to be at the N-terminal (α_n)	32.449***	7	3.159***
Refractive index (μ)	47.548***		3.632***
Short and medium range non-bonded energy (E_{sm})	67.785***		3.743***
Solvent accessible reduction ratio (R_s)	83.159***		2.942**
Surrounding hydrophobicity (H_p)	69.397***		2.345**
Thermodynamic transfer hydrophobicity (H_t)	50.917***		2.785**
Total non-bonded energy (E_t)	109.548***		5.654***
Turn tendencies (P_t)	109.136***	6	6.198***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

filtered the genes found according to the expression period from 00-12 hour of the embryonic stage, which corresponds to the expression of *Dnmt2* in *D. melanogaster* according to Lyko et al. (2000). The biological role played by genes according to the biological process described for DNMT2 was also considered (i.e., epigenetic functions,

transcriptional control, stress response, tRNA methylation, response to heat, positive regulation of innate immune response, telomere maintenance). At the end, we obtained a list of 35 genes and added the enolase enzyme, since this was the first DNMT2 interacting protein described (Tovy et al., 2010) (Table 5).

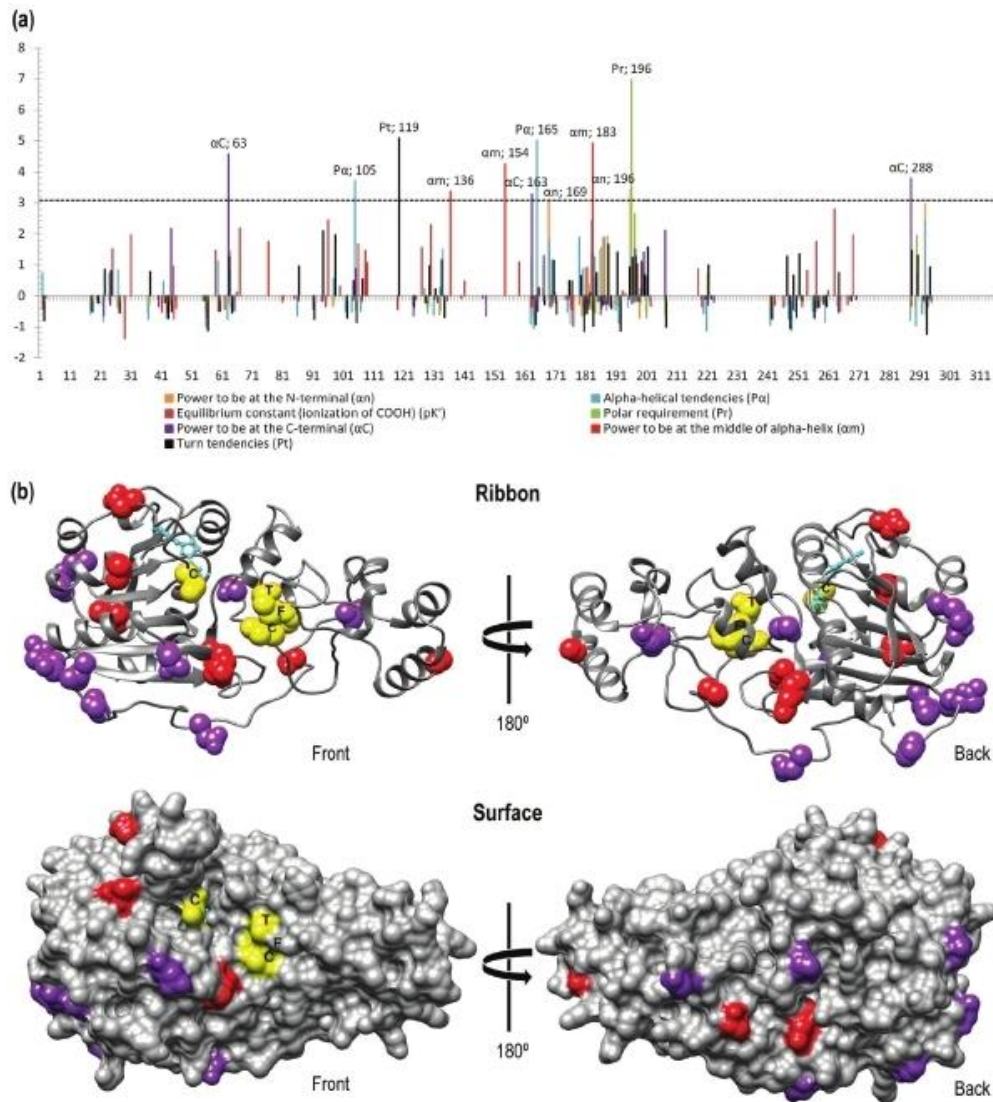


Figure 5 - DNMT2 structure predictions. (a) Sliding window plots of the z-scores of radically changed properties showing regions under positive-destabilizing selection in *Dnmt2*. Dashed horizontal line indicates the Bonferroni-corrected significant limit ($z\text{-score} = 3.09, p < 0.001$). Alpha-helical tendencies (P_{α}) – red; Equilibrium constant (ionization of COOH) (pK') – orange; Polar requirement (P_r) – green; Power to be at the C-terminal (α_c) – blue; Power to be at the middle of alpha-helix (α_m) – pink; Power to be at the N-terminal (α_n) – black; Turn tendencies (P_t) – purple. (b) Structure of *sfDNMT2* (PDVB: 4H0N) (Li et al., 2012), ribbon (up) and surface (down) representations. The respective sites under positive selection in drosophilids are represented in *Spodoptera frugiperda* structure. Sites under positive selection colored red, positive-destabilizing selection purple, motif CFT from TRD and the catalytic cysteine yellow and cofactor s-adenosyl methionine (SAM) cyan.

ERC values for the 36 genes were obtained by the Group ERC Analysis tool, which returns the ERC values between a group of genes and statistics for the strength of ERC between them (Table S6). Interestingly, *Dnmt2* presents high ERC values related with 10 genes ($p < 0.05$) (Table 6), in which four are related to transcriptional control

and three have a chromatin-remodeling function. However, genes with protein interaction prediction coming from the STRING database showed low ERC values (Table 6). The 15 genes with an ERC value > 0.400 ($p\text{-value} < 0.05$) are distributed between the Muller's elements A, B, C, D and E.

Table 5 - Gene annotation and sequence location from STRING and ERC webservice searching from *D. melanogaster* genome. Genes with high ERC value (> 0.400) and p-value < 0.05 are colored gray.

Gene	Annotation	Description
Atac3	CG32343	Contributes to histone acetyltransferase activity. Involved chromatin remodeling.
Cap	CG18408	Interacts selectively and non-covalently with vinculin, a protein found in muscle, fibroblasts, and epithelial cells that binds actin and appears to mediate attachment of actin filaments to integral proteins of the plasma membrane.
CG10262	CG10262	Proliferating Cell Nuclear Antigen (PCNA) domain. These polymerase processivity factors play a role in DNA replication and repair.
CG13035	CG44836	Uncharacterized protein involved in sensorial perception of pain.
CG14618	CG14618	Belongs to the class IV-like SAM-binding methyltransferase superfamily. tRNA (guanine(9)-N(1))-methyltransferase TRM10. Enzyme catalyzes the conversion of a guanosine residue to N1-methylguanine in position 37, next to the anticodon, in tRNA.
CG14906	CG14906	N-6 adenine-specific DNA methylases. Specifically methylate the amino group at the C-6 position of adenines in DNA.
CG16863	CG16863	The protein has zinc finger, BED-type, which is thought to be involved in chromatin insulation at BEAF (boundary element-associated factor) and DREF, a transcriptional regulator.
CG17124	CG17124	PKC-activated protein phosphatase-1 inhibitor. Stops, prevents or reduces the activity of a protein phosphatase, an enzyme that hydrolyzes phosphate groups from phosphorylated proteins.
CG6712	CG6712	Probably RNA binding inferred from sequence or structural similarity with <i>Saccharomyces</i> RPF1, involved in ribosome biogenesis. Belongs to Brix superfamily.
CG7470	CG7470	Predicted gamma-glutamyl phosphate reductase with delta1-pyrroline-5-carboxylate synthetase activity, glutamate 5-kinase activity. An inner mitochondrial membrane enzyme, is essential to the de novo synthesis of the amino acids proline and arginine. Involved in epithelium development and gerarium-derived egg chamber formation.
CTCF	CG8591	CTCF is a ubiquitous transcription factor that binds to insulators and domain boundaries. It mediates insulator function and blocks enhancers by binding to Cp190. It contributes to long-range chromatin interaction, organizes chromatin domain boundaries and coordinates nuclear architecture.
Dnmt2	CG10692	Methyltransferase 2 is a (cytosine-5) DNA/tRNA methyltransferase. It is involved in regulation of gene expression by cytosine-5 methylation. The major protein role is the modifications that protects tRNAs against endonucleolytic cleavage and contributes to stress resistance, protein translation and small RNA-mediated gene regulation.
Eggless	CG12196	Belongs to the class V-like SAM-binding methyltransferase superfamily. Histone-lysine methyltransferase family. Involved in negative regulation of gene expression.
Eno	CG17654	Responsible for the catalysis of the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP), the ninth and penultimate step of glycolysis
Haywire	CG8019	Helicase Ercc3, core RNA polymerase binding transcription factor activity. Involved in regulation of alternative mRNA splicing, via spliceosome, cell proliferation and growth.
Homer	CG11324	Involved in the positive regulation of circadian sleep/wake cycle, sleep; response to ethanol, behavioral response to ethanol, regulation of locomotion and adult behavior. Activity in stress response.
hop	CG1594	Signal transduction-non receptor tyrosine kinase. Members of the Janus kinase (JAK) family of cytoplasmic protein tyrosine kinases physically associate with ligand-bound receptors. The JAK-STAT pathway regulates the expression of pair rule gene even-skipped (a transcriptional repressor of a number of genes) early in embryogenesis.
Hsp22	CG4460	A key player in cell-protection mechanisms against oxidative injuries and aging in <i>Drosophila</i> . Activated in late third-instar larvae of <i>Drosophila melanogaster</i> in the absence of heat stress
MBD-like	CG8208	Methyl Binding Protein 2/3, a co-repressor and an integral component of the nucleosome remodelling and deacetylase (NuRD) complex. Negative regulation of transcription, involved in organism development
mei-S332	CG5303	Acts to maintain sister-chromatid cohesion before anaphase II of meiosis in both males and females.
mms4	CG12936	Methyl methanesulfonate sensitivity 4 is the non-catalytic subunit of the mus81-mms4 structure-selective endonuclease that functions in DNA repair.
mus209	CG9193	Belongs to the PCNA family. Involved in eggshell chorion gene amplification, DNA replication, mismatch repair, neurogenesis and antimicrobial humoral response.
Nsun2	CG6133	tRNA (cytosine-5-)-methyltransferase activity.
Orc2	CG3041	Origin recognition complex subunit 2. It is involved in eggshell chorion gene amplification, chromatin silencing, cell proliferation, DNA replication initiation, chromosome condensation and neurogenesis.

Table 5 - Cont.

Gene	Annotation	Description
Pnt	CG17077	It is a sequence-specific DNA binding transcription factor activity, repressing transcription factor binding, involved in positive regulation of transcription. It is involved in organism developmental process, open tracheal system development, post-embryonic organ morphogenesis, regulation of developmental process, sensory organ development, cardiovascular system development, regulation of RNA metabolic process, positive regulation of cell proliferation, anterior/posterior axis specification, multi-organism reproductive process, muscle cell differentiation, compound eye photoreceptor development, regulation of neurogenesis.
Rel	CG11992	Relish is a transcription factor and the downstream component of the Immune Deficiency pathway, which regulates the antibacterial response and other less characterized cellular processes.
RhoGEF4	CG8606	Rho family small GTPases act as molecular switches that regulate neuronal morphogenesis, including axon growth and guidance, dendritic spine formation, and synapse formation. These proteins are positively regulated by guanine nucleotide exchange factors (GEFs) of the Dbl family. Findings suggest that DRhoGEF4 may participate in cytoskeleton-related cellular events by specifically activating RhoA in neuronal morphogenesis.
Rpd3	CG7471	Histone deacetylase 1. Catalyzes the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4). Histone deacetylation may constitute a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events. For instance, deacetylation of histone H3 may be a prerequisite for the subsequent recruitment of the histone methyltransferase Su(var)3-9 to histones. Involved in position-effect variegation (PEV).
Rpp20	CG33931	A subunit of the RNase P and RNase MRP holoenzymes, has interaction with the Drosophila SMN protein. Immunofluorescence results indicate that Rpp20 is diffusely distributed throughout the cytoplasm with higher concentration observed in the nucleus. However, in response to stress, SMN forms aggregates and redistributes Rpp20 into punctuated cytoplasmic SMN granules.
Sna	CG3956	Snail is a transcription factor that contributes to embryonic mesoderm development, epithelial to mesenchymal transition and asymmetric cell division.
Su(var)2-5	CG8409	Suppressor of variegation 205 is a heterochromatin protein associated with the pericentric heterochromatin and telomeres in Drosophila. It is involved in the positive autoregulatory expression and can bind directly to nucleosomes.
Su(var)2-10	CG8068	It is responsible for establishing and maintaining chromosome organization in interphase nuclei, promoting chromosome structure and function.
Su(var)3-3	CG17149	Probable histone demethylase that specifically demethylates 'Lys-4' of histone H3, a specific tag for epigenetic transcriptional activation, thereby acting as a corepressor. Required for heterochromatic gene silencing
Su(var)3-9	CG43664	Histone methyltransferase that specifically trimethylates 'Lys-9' of histone H3 using monomethylated H3 'Lys-9' as substrate. H3 'Lys-9' trimethylation represents a specific tag for epigenetic transcriptional repression by recruiting Su(var)205/HP1 to methylated histones. Mainly functions in heterochromatin regions, thereby playing a central role in the establishment of constitutive heterochromatin at pericentric regions. Involved in heterochromatic gene silencing including the modification of position-effect-variegation.
Tet	CG9973	Ten-Eleven Translocation (TET) family protein. Involved in positive regulation of DNA demethylation, inducing positive regulation of transcription from RNA polymerase II promoter.
Thor	CG8846	Eukaryotic translation initiation factor 4E binding protein, controlled by tor. It contributes to translation regulation, response to environmental stress and cell growth regulation.

Discussion

Evolutionary scenario of drosophilids DNMT2

In our analyses of *Dnmt2* homologs among species of Drosophilidae (Dot blot, Figure S1), we detected a hybridization signal in the majority of the analyzed species. Not all species amplified using the *D. melanogaster* and *D. willistoni* primers, indicating that the degree of *Dnmt2* similarity among Drosophilidae species varies (Table S1). The representative species of the *Dorsilopha* subgenus did not present *Dnmt2* amplification with the primers used, which could indicate that the 5' and 3' regions of the *Dnmt2* gene must have certain variation in the species that did not show amplicons (Table S1). However, the average divergence value did not exceed 22.5% for the amino acid sequences, or 26.57% for the nucleotide sequences (data not shown).

Functionally important genes are often evolutionarily constrained because the amino acid sequence must be preserved for a protein's catalytic or structural role to be maintained (Figures 3 and S2).

Overall, the Neotropical *willistoni* species group presents the highest differences when compared to all other groups, both in the nucleotide and amino acid analysis (Table 1). Interestingly, the difference is reflected in the increased presence of basic amino acids in the DNMT2 of the *willistoni* subgroup. Such differences may result in changes in physicochemical properties of the enzyme, giving modulations and differential affinities between proteins of different groups. Shifts in codon preferences are described in *D. willistoni* (Singh *et al.*, 2006; Vicario *et al.*, 2007), and being a lineage-specific feature, it is suggested that the differential preference may influence the evolution of DNMT2 in

Table 6 - ERC values between potential DNMT2 protein-protein partners. The matrix shows all pairwise ERC values between genes below the diagonal and respectively p-values above the diagonal. Cells are shaded red according to the intensity of their deviation from the null expectation.

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 pnt	N/A	0.0010	0.1445	0.0781	0.4435	0.1082	0.0950	0.1429	0.3571	0.1486	0.1924	0.3492	0.0145	0.4435	0.0207	0.0170	0.0241	0.0760
2 Su(var)2-10	0.0311	N/A	0.0578	0.0168	0.0143	0.0868	0.0891	0.0912	0.1791	0.1679	0.3720	0.5330	0.0185	0.0912	0.0493	0.0572	0.0383	0.2213
3 CG14618	0.375	0.480	N/A	0.0093	0.0398	0.0541	0.0117	0.0249	0.0220	0.1015	0.1111	0.1780	0.0123	0.0732	0.0489	0.1203	0.0775	0.4843
4 CG14906	0.473	0.597	0.731	N/A	0.0089	0.0125	0.0025	0.0008	0.0387	0.0459	0.0879	0.2154	0.0266	0.1378	0.1770	0.1251	0.0431	0.4492
5 CG32343	0.000	0.608	0.593	0.782	N/A	0.0032	0.0070	0.0057	0.1500	0.0492	0.1193	0.3912	0.0180	0.0818	0.2100	0.1325	0.1437	0.3912
6 CTCF	0.426	0.429	0.555	0.758	0.801	N/A	0.0004	0.0032	0.0435	0.0238	0.1094	0.1260	0.0641	0.1074	0.1074	0.1087	0.0153	0.1688
7 CG16863	0.447	0.424	0.713	0.844	0.763	0.932	N/A	0.0009	0.0304	0.0262	0.0448	0.0905	0.0267	0.0693	0.0804	0.0794	0.0165	0.2525
8 sna	0.377	0.421	0.650	0.874	0.773	0.813	0.912	N/A	0.0149	0.1107	0.0146	0.0721	0.0278	0.0655	0.0518	0.0411	0.0127	0.3520
9 mei-S332	0.087	0.311	0.662	0.650	0.386	0.643	0.698	0.796	N/A	0.1072	0.0903	0.0312	0.0248	0.4685	0.0943	0.1754	0.0643	0.5163
10 homer	0.370	0.323	0.457	0.602	0.587	0.708	0.713	0.524	0.457	N/A	0.0026	0.0788	0.0196	0.1536	0.1112	0.0661	0.0397	0.0903
11 Rpp20	0.307	0.109	0.441	0.513	0.440	0.506	0.651	0.797	0.489	0.777	N/A	0.0127	0.0186	0.1018	0.0495	0.0950	0.0703	0.1915
12 RhoGIEF4	0.097	-0.025	0.337	0.313	0.000	0.480	0.541	0.602	0.638	0.498	0.716	N/A	0.0322	0.0241	0.0250	0.1788	0.1488	0.1462
13 M62	0.648	0.590	0.711	0.693	0.699	0.589	0.712	0.741	0.663	0.650	0.680	0.577	N/A	0.0179	0.0049	0.0339	0.0392	0.1800
14 Orc2	0.000	0.421	0.512	0.425	0.509	0.508	0.585	0.616	0.000	0.375	0.470	0.609	0.686	N/A	0.0044	0.0428	0.4939	0.1165
15 hay	0.609	0.497	0.572	0.369	0.297	0.508	0.563	0.653	0.481	0.440	0.582	0.605	0.771	0.741	N/A	0.0028	0.0082	0.0493
16 mms4	0.631	0.481	0.426	0.445	0.418	0.507	0.565	0.686	0.344	0.521	0.483	0.313	0.622	0.542	0.831	N/A	0.0041	0.0273
17 hop	0.596	0.524	0.503	0.629	0.399	0.747	0.762	0.805	0.540	0.588	0.528	0.351	0.602	0.000	0.774	0.845	N/A	0.0940
18 mms209	0.476	0.264	0.020	0.052	0.000	0.409	0.295	0.185	-0.010	0.479	0.336	0.355	0.349	0.397	0.621	0.703	0.551	N/A

the species of the *willistoni* group. Also, molecular evolution is atypical in this species of the *willistoni* group (and its sister group *saltans*), characterized by high rates of nucleotide substitution and low portions of G/C (DeSalle, 1992; Powell and DeSalle, 1995; Remsen and O'Grady, 2002).

Our phylogenetic analysis further showed that the DNMT2 relationships in Drosophilidae comprised three main clades: a *virilis-repleta* section, a *quinaria-tripunctata* section and the *Sophophora* subgenus (Figures 1 and S2), consistent with previous studies regarding the phylogeny of the genus (Lewis et al., 2005; Robe et al., 2005, 2010; Wang et al., 2006; Hatadani et al., 2009), although for the *quinaria-tripunctata* section, the evolutionary relationships between the species did not have a strong support value and presented, in some cases, polytomy. Monophyly of the subgenus *Sophophora* has been confirmed as well by Tataronov et al. (1999) and Robe et al. (2005). These findings suggest that the DNMT2 sequences exhibit similar patterns to the species evolution.

Despite the conservation of drosophilid DNMT2s, at least 10 regions have a high evolutionary rate along the protein (Figure 1). Noteworthy, region 7 has a long sequence with high evolutionary rate. The structural location of this region corresponds to the connecting handle between the catalytic domain and TRD. When analyzed separately, the evolutionary rates of the *Drosophila* and *Sophophora* subgenera showed very similar patterns with regard to the distribution of sites that evolve faster. Nonetheless, evolutionary rate values in the *Drosophila* subgenus seem to be smaller than those of *Sophophora* (Figure 1). Thus, we question whether some of these sites with faster evolutionary rates may be under some lineage-specific adaptive selection event, or just correspond to a relaxed selection pressure.

Positively selected sites in DNMT2 suggest adaptation for protein-protein interactions

The importance of *Dnmt2* for development in drosophilid species remains unclear. Marhold et al. (2004) compared DNMT2 sequences from different dipterans, showing high evolutionary conservation mainly in the catalytic domain. The occurrence of DNA methylation in various dipteran species was also demonstrated, suggesting that DNMT2-mediated DNA methylation has a deep evolutionary relationship of at least 250 million years. In any case, DNMT2-mediated DNA methylation remains an open issue.

By depletion of *Dnmt2* with RNA interference, Kurnert et al. (2003) demonstrated that DNMT2 is both necessary and sufficient for DNA methylation in *D. melanogaster* and suggest a different target sequence for DNA methylation: CpT/A. However, depletion of *Dnmt2* had no detectable effect on embryonic development, despite complete loss of DNA methylation. On the other hand, DNMT2 has low catalytic activity on DNA (Hermann et al., 2003;

Fisher *et al.*, 2004), and it seems to have a preference for tRNA-Asp as a methylation target (Goll *et al.*, 2006; Jurkowski *et al.*, 2008; Krauss and Reuter 2011; Tuorto *et al.*, 2012; Durdevic *et al.*, 2013a; Shanmugam *et al.*, 2014). More recently, Takayama *et al.* (2014) have brought another piece to the controversial and elusive puzzle of DNA methylation in the genome of *D. melanogaster*: lines deficient for DNMT2 retain genomic methylation, although with altered patterns. Adding up to this peculiar field, there is the phenomenon of sex-specific DNA methylation in species of the subgroup *willistoni* (Garcia *et al.*, 2007; D'Ávila *et al.*, 2010), which raises questions about the mechanisms that are directly involved in the sex-specific DNA methylation process in this drosophilid group.

These scenarios raise issues about the selective pressures acting on *Dnmt2* and on how the DNMT2 protein may have evolved along the several Drosophilidae lineages. Considering that *Dnmt2* was described as unnecessary for embryological development (Kunert *et al.*, 2003) and no loss of fitness with depletion of DNMT2 was detected (Takayama *et al.*, 2014) in *D. melanogaster*, one might expect a relaxed selective constraint. However, the present study demonstrated that *Dnmt2* is evolutionarily conserved (Figures 2 and S2), as previous studies have already shown (Marhold *et al.*, 2004), and therefore, it could be suggested that the entire gene has evolved under purifying selection and that positive selection has played only a minor role in the evolution of Drosophilidae *Dnmt2*.

Since the catalytic domain (residues from the cofactor-binding pocket and catalytic residues in the motifs ENV and PPC) and the motif CFT in TRD are generally conserved (Figures 3 and 5b), the 10 sites with high evolutionary rates in some motifs in the catalytic domain and most of the TRD motif would be due to relaxed selective constraint or neutral evolution (Figure 1). On the other hand, this variation would be better explained by an adaptive selection. Nevertheless, when we employed analyses which considered variation in ω rates in branches and sites, we found several positively selected sites in the catalytic domain and in the TRD from *Dnmt2* between *foreground* (subgenus *Sophophora*) and *background* (subgenus *Drosophila*) lineages (Table 3 and Figure 5b). Positive selection has been shown to act directly on amino acid residues exposed on the surface of proteins, while those in the core remain stable (Casewell *et al.*, 2011), favoring the maintenance of catalytic function of the protein.

The majority of sites under positive-destabilizing selection were located at the protein surface, which suggests that these sites are potentially involved in interactions with the surrounding environment. The sites 63, 163 and 288 (Figure 4a,b) were noted as having radically changed properties, namely Power to be at the C-terminal (α C). This property relates to the ability of residues to interact with other molecules, especially protein-protein and subunit interactions. Also, sites 136 to 183 are located in the connect-

ing region of the catalytic domain and TRD and could be a potential region of protein interaction.

DNMT2 has new potential partners that have not been considered yet

Tovy *et al.* (2010) established enolase as the first DNMT2 interacting protein and highlighted an unexpected role of a glycolytic enzyme in the modulation of DNMT2 activity. Other potential protein partners of DNMT2 have been described, such as EGGLESS (dSETDB1), which is a member of the family of SET/MBD proteins and methylates lysine 9 in histone H3, that mediates DNA methylation and is involved in silencing genes and retrotransposons (Gou *et al.*, 2010). Another probable candidate is MBD-like, a methyl-DNA binding protein that keeps certain genes epigenetically silenced during genome activation (Marhold *et al.*, 2004; Gutierrez and Sommer, 2004). Since DNMT2 has tRNA as a preferential target, NSUN2 (a tRNA methyltransferase which has the function of cytosine-C5 methylation) may also be a possible protein partner, cooperating with tRNA stability and protein synthesis (Tuorto *et al.*, 2012), together with DNMT2.

ERC values are typically elevated between interacting proteins and can be used to establish molecular and functional interactions between a pair and/or a group of proteins. Therefore, here we tested 36 proteins retrieved from searches through the STRING database and the tool ERC Analysis Webserver Top Genes. Interestingly, we found that the protein-interactions predicted with DNMT2 (STRING database) have low ERC values, mostly with negative values (Table 6). We had expected that high ERC values would be found in the protein pairs that are known to belong to epigenetic mechanisms predicted to interact with DNMT2. However, only the CG6712, MDB-like, RPD3 and SU(VAR)2-5 proteins showed positive ERC values, but these were very low ($ERC < 0.3$). DNMT2 and mus209 showed a low ERC (0.369) but this was not significant ($p < 0.18$). Even enolase, which was described as the first DNMT2 interacting protein (Tovy *et al.*, 2010), showed a low ERC value. Nonetheless, Clark *et al.* (2012) demonstrated that direct physical interaction is not required to establish a high evolutionary rate covariation, but that the two major components associated with ERC are cofunctionality (functional and physical interactions annotations) and the coevolution of expression levels. When we looked at the proteins with high ERC values, we observed a variety of proteins with different levels of gene expression regulation (Table 6). This was not surprising, given the previously known subtract duality and the diversity of biological processes involving DNMT2 (Kunert *et al.*, 2003; Tang *et al.*, 2003; Lin *et al.*, 2005; Goll *et al.*, 2006; Phalke *et al.*, 2009; Schaefer *et al.*, 2010; Durdevic *et al.*, 2013b). This said, we can rank the proteins with high ERC values into six groups: (1) chromatin remodeling, (2) transcription factors, (3) ex-

pression regulation, (4) DNA replication, (5) stress response, and (6) RNA editing.

The first group shows that DNMT2 has a high ERC value with proteins related to chromatin remodeling, like ATAC3 that is an essential *D. melanogaster* histone acetyl transferase (HAT) complex, together with transcriptional cofactors GCN5 (KAT2), ADA3, ADA2A, ATAC1 and HCF. This complex does not work in nucleosome remodeling itself, but it stimulates nucleosome sliding by the ISWI, SWI-SNF and RSC complexes (Suganuma *et al.*, 2008). The nucleosome remodeling process is required to expose sequences that may be a target for gene silencing.

In their review, Klose and Bird (2006) describe that in targeting *de novo* DNA methylation, transcription factors (TF) have the capacity to interact with DNMT enzymes and to promote cytosine-target methylation as a part of the molecular silencing repertoire. This specific DNMT-TF network may be responsible for promoting specific differentiation stages in different organs. The significant ERC values found between DNMT2 and TFs like PNT and SNAIL (Table 6) reinforce the DNMT-TF interaction findings.

The involvement of DNMT2 in gene expression regulation appears in the interaction with histone methyltransferases enzymes (HMT) with high ERC values between the heterochromatin protein (HP) SU(VAR)2-10. SU(VAR)2-10 encodes a member of the PIAS protein family that controls diverse functions and is involved in different aspects of chromosome structure and function by establishing/maintaining chromosome organization (Hari *et al.*, 2001). Furthermore, DNMT-TF interaction and DNMT-HP interaction can produce a ternary complex composed of a DNMT, a MBD and an HP, promoting the recruitment of histone H3K9 methyltransferases (HMTs) (Rai *et al.*, 2010). The CG16863 protein has a BED-type zinc finger domain and is not characterized yet. BED fingers are able to bind DNA and are present in some proteins like *Drosophila* boundary element-associated factor (BEAF), responsible by chromatin insulation and also required during early development (Gilbert *et al.*, 2006). DREF (DNA replication-related element-binding factor) is another protein that contains BED fingers; it is a transcription regulatory factor and it interacts genetically and physically with regulatory factors related to chromatin structures. Matsukage *et al.* (2008) identified more than 150 genes carrying DRE sequences in their promoter regions, most of them related to DNA replication, transcriptional regulation, cell cycle regulation, growth signal transduction and protein metabolism. The high ERC value between DNMT2 and BED-finger domain (0.712) is very interesting (Table 6), because the DREF target sequence is 5'-TATCGATA-3', carrying a CpG motif that can be a DNMT2 target for cytosine methylation.

Another protein group that shares a high ERC with DNMT2 is related with expression regulation: HOPSCOTCH and HAYWIRE. HOPSCOTCH is a protein

pseudokinase and is involved in many biological processes, like cell proliferation, structure morphogenesis, and others. Hou *et al.* (1996) have described Hopscotch/JAK kinase as an invertebrate JAK/STAT system. HOPSCOTCH regulates the transcription of target genes, such as the pair-rule gene *even-skipped*. *even-skipped* is a transcriptional repressor of a number of genes during early embryogenesis. Thus, having its role in the epigenetic system may DNMT2 have coevolved with Hopscotch/JAK to rearrange the epigenetic marks during the early development of drosophilids.

We also found significant evolutionary rate covariation values with proteins involved in DNA replication like ORC2, MMS4 and even MUS209 (a Proliferating cell nuclear antigen – PCNA) (Table 6). This may be linked to DNMT2 maintenance methylation of DNA in the replication event, or to *de novo* methylation during embryonic development.

Usually present in the nucleus, under conditions of stress, DNMT2 reallocated to cytoplasmic stress granules and RNA processing bodies (P-bodies). Thiagarajan *et al.* (2011) describe DNMT2 as part of the RNA processing machinery during cellular stress. During heat shock conditions, DICER-2 degrades tRNA and tRNA fragments, so DNMT2 can limit the extent of tRNA fragmentation during a stress event, since long double-stranded RNAs (dsRNAs) inhibit DICER-2 activity. Hence, DNMT2 is essential for DICER-2 processing in *Drosophila* (Durdevic *et al.*, 2013a).

Another role of DNMT2 against stress events is the silencing of retrotransposons and the control of RNA viruses in *Drosophila* (Phalke *et al.*, 2009; Durdevic *et al.*, 2013b). So, DNMT2 seems to be important as a control tool to various forms of stress that involve RNA, which can be triggered by an excess of endogenous (retrotransposons) or exogenous (viruses) RNAs. The presence of the proteins HOMER and Rpp20, with high ERC values, may contribute to the presence and function of DNMT2 involved in the response to stress events and opens new possible partner proteins. HOMER protein acts in response to ethanol (Urizar *et al.*, 2007), controlling circadian cycles (Naidoo *et al.*, 2012), and acts during the stress response. Rpp20 is a subunit of RNase P and RNase MRP that is involved in precursor rRNA processing (Li *et al.*, 2002) and interacts with SMN protein in response to stress (Hua and Zhou 2004).

Finally, we denoted an evolutionary rate relationship between DNMT2 and two nucleotide modification enzymes: CG14906 e CG14618. The CG14906 gene product corresponds to a TM-A70-like factor, an mRNA (2-O-methyladenosine-N(6)-methyltransferase (Bujnicki *et al.*, 2002), and CG14618 belongs to the class IV-like SAM-binding methyltransferase superfamily, tRNA methyltransferase TRM10 family. TRM10 catalyzes all known instances of m¹G9 modification and, according to Swinehart *et al.* (2013), it is involved in different pathways

beyond tRNA processing. The plurality and importance of RNA and tRNA modifications found in the last decade are well described (Chiari *et al.*, 2010; Chan *et al.*, 2012; Gu *et al.*, 2014; Hori, 2014; Swinehart and Jackman, 2015; Tuorto *et al.*, 2015). The pathway diversity and response plasticity to environmental modifications makes the tRNA/RNA processing enzymes a major link between the genome and environment.

Conclusions

Our study shows that, although *Dnmt2* is highly conserved within the Drosophilidae family, it carries several nonsynonymous changes in some domains, which were shown to be maintained by positive and destabilizing selection. Purifying selection remains the major maintenance process of gene function(s), but positive selection appears to act on certain domains potentially involved with environment interactions. Thus, the findings suggest that the residues affected by positive selection may be involved in an interaction-driven co-evolution and the connection of regions of catalytic domains and TRD that probably would interact – direct or indirectly – with other proteins.

We suggest that the episodes of adaptive evolution in *Dnmt2* could be related to the wide diversity of niches, behaviors, amplitude distribution of drosophilids, as well as with and other peculiarities, such as the presence of transposable elements, chromosomal inversions, chromosomal stability, sex-specific DNA methylation, responsive modulation of RNA methylation (coding and noncoding), and endosymbiotic interactions. The multiplicity of proteins having strong evolutionary rate covariation found in the present work supports the hypothesis of a plurality of DNMT2 functions in *Dnmt2*-only organisms, like drosophilids. Since epigenetic systems are dynamic and change with the environment and along the evolutionary history of the organisms, we think that a wide scenario was opened, and the next step will be to analyze probable *Dnmt2* interactions with other genes along the evolution of different lineages in their ecological and evolutionary contexts.

Acknowledgments

We would like to thank two anonymous referees for helpful comments and suggestions. This study was supported by grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

- Abascal F, Zardoya R and Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104-2105.

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716-723.
- Bujnicki JM, Feder M, Radlinska M and Blumenthal RM (2002) Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m6A methyltransferase. *J Mol Evol* 55:431-444.
- Casewell NR, Wagstaff SC, Harrison RA, Renjifo CL and Wüster W (2011) Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol* 28:2637-2649.
- Chan CTY, Pang YLJ, Deng W, Babu IR, Dyavaiah M, Begley TJ and Dedon PC (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun* 3:937.
- Chiari Y, Dion K, Colborn J, Parmakelis A and Powell JR (2010) On the possible role of tRNA base modifications in the evolution of codon usage: queuosine and *Drosophila*. *J Mol Evol* 70:339-345.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Clark NL, Alani E and Aquadro CF (2012) Evolutionary rate covariation reveals shared functionality and co-expression of genes. *Genome Res* 22:714-720.
- Clark NL, Alani E and Aquadro CF (2013) Evolutionary rate covariation involving meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics* 193:529-538.
- D'Ávila MF, Garcia RN, Panzera Y and Valente VLS (2010) Sex-specific methylation in *Drosophila*: an investigation of the *Sophophora* subgenus. *Genetica* 138:907-913.
- Da Lage JL, Kergoat GJ, Maczkowiak F, Silvain JF, Cariou ML and Lachaise D (2007) A phylogeny of Drosophilidae using the *Amyrel* gene: Questioning the *Drosophila melanogaster* species group boundaries. *J Zool Syst Evol Res* 45:47-63.
- DeSalle R (1992) The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. *Mol Phylogenet Evol* 1:31-40.
- Durdevic Z, Mobin MB, Hanna K, Lyko F and Schaefer M (2013a) The RNA methyltransferase *Dnmt2* is required for efficient Dicer-2-dependent siRNA pathway activity in *Drosophila*. *Cell Rep* 4:931-937.
- Durdevic Z, Hanna K, Gold B, Pollex T, Cherry S, Lyko F and Schaefer M (2013b) Efficient RNA virus control in *Drosophila* requires the RNA methyltransferase *Dnmt2*. *EMBO Rep* 14:269-275.
- Edgar RC (2004) MUSCLE: Mmultiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
- Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL and Wolfner MF (2014) Evolutionary rate covariation identifies new members of a protein network required for *Drosophila* female post-mating responses. *PLoS Genetics* 10:e1004108.
- Fisher O, Siman-Tov R and Ankril S (2004) Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmet) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res* 32:287-297.
- Gailey DA, Ho SK, Ohshima S, Liu JH, Eyassu M, Washington MA, Yamamoto D and Davis T (2000) A phylogeny of the

- Drosophilidae using the sex-behaviour gene fruitless. *Hereditas* 133:81-83.
- Garcia RN, D'Ávila MF, Loreto ELS, Panzera Y, Heredia FO and Valente VLS (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica* 131:91-105.
- Gilbert MK, Tan YY and Hart CM (2006) The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics* 173:1365-1375.
- Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, Golic KG, Jacobsen SE and Bestor TH (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311:395-398.
- Gou D, Rubalcava M, Sauer S, Mora-Bermúdez F, Erdjument-Bromage H, Tempst P, Kremmer E and Sauer F (2010) SETDB1 is involved in postembryonic DNA methylation and gene silencing in *Drosophila*. *PLoS One* 5:e10581.
- Gu C, Begley TJ and Dedon PC (2014) tRNA modifications regulate translation during cellular stress. *FEBS Lett* 28:4287-4296.
- Guindon S and Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704.
- Gutierrez A and Sommer RJ (2004) Evolution of dnmt-2 and mbd-2-like genes in the free-living nematodes *Pristionchus pacificus*, *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res* 32:6388-6396.
- Hari KL, Cook KR and Karpen GH (2001) The *Drosophila* Su(var)2-10 locus regulates chromosome structure and function and encodes a member of the PIAS protein family. *Genes Dev* 15:1334-1348.
- Hatadani LM, McInerney JO, de Medeiros HF, Junqueira AC, de Azeredo-Espin AM and Klaczko LB (2009) Molecular phylogeny of the *Drosophila tripunctata* and closely related species groups (Diptera: Drosophilidae). *Mol Phylogenet Evol* 51:595-600.
- Hermann A, Schmitt S and Jeltsch A (2003) The human Dnmt2 has residual DNA (cytosine-C5) methyltransferase activity. *J Biol Chem* 278:31717-31721.
- Hori H (2014) Methylated nucleosides in tRNA and tRNA methyltransferases. *Front Genet* 5:1-26.
- Hou XS, Melnick MB and Perrimon N (1996) Marelle acts downstream of the *Drosophila* HOP/JAK kinase and encodes a protein similar to the mammalian STATs. *Cell* 84:411-419.
- Hua Y and Zhou J (2004) Survival motor neuron protein facilitates assembly of stress granules. *FEBS Lett* 572:69-74.
- Jones DT, Taylor WR and Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275-282.
- Jurkowski TP, Meusburger M, Phalke S, Helm M, Nellen W, Reuter G and Jeltsch A (2008) Human DNMT2 methylates tRNA^{Asp} molecules using a DNA methyltransferase-like catalytic mechanism. *Rna* 14:1663-1670.
- Katoh M, Curk T, Xu Q, Zupan B, Kuspa A and Shaulsky G (2006) Developmentally regulated DNA methylation in *Dictyostelium discoideum*. *Eukaryot Cell* 5:18-25.
- Klose RJ and Bird AP (2006) Genomic DNA methylation: The mark and its mediators. *Trends Biochem Sci* 31:89-97.
- Kopp A (2006) Basal relationships in the *Drosophila melanogaster* species group. *Mol Phylogenet Evol* 39:787-798.
- Krauss V and Reuter G (2011) DNA methylation in *Drosophila* - a critical evaluation. *Progr Mol Biol Transl Sci* 110:177-191.
- Kucharski R, Maleszka J, Foret S and Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827-1830.
- Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, Söderbom F, et al. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res* 33:6405-6417.
- Kumar S, Tamura and Stecher G (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol Biol Evol*. *Mol Biol Evol* 33:1870-1874.
- Kunert N, Marhold J, Stanke J, Stach D and Lyko F (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* 130:5083-5090.
- Kwiatowski J and Ayala FJ (1999) Phylogeny of *Drosophila* and related genera: Conflict between molecular and anatomical analyses. *Mol Phylogenet Evol* 13:319-328.
- Lewis RL, Beckenbach AT and Mooers A (2005) The phylogeny of the subgroups within the *melanogaster* species group: Likelihood tests on COI and COII sequences and a Bayesian estimate of phylogeny. *Mol Phylogenet Evol* 37:15-24.
- Li S, Du J, Yang H, Yin J, Zhong J and Ding J (2012) Functional and structural characterization of DNMT2 from *Spodoptera frugiperda*. *J Mol Cell Biol* 5:64-66.
- Li X, Frank DN, Pace N, Zengel JM and Lindahl L (2002) Phylogenetic analysis of the structure of RNase MRP RNA in yeasts. *RNA* 8:740-751.
- Lin MJ, Tang LY, Reddy MN and Shen CKJ (2005) DNA methyltransferase gene dDnmt2 and longevity of *Drosophila*. *J Biol Chem* 280:861-864.
- Lyko F, Ramsahoye BH and Jaenisch R (2000) DNA methylation in *Drosophila melanogaster*. *Nature* 408:538-540.
- Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B and Lyko F (2004) Conservation of methylation in dipteran insects. *Insect Mol Biol* 13:117-123.
- Marques EK, Napp M and Winge H (1966) A corn meal, soybean flour, wheat germ medium for *Drosophila*. *Drosoph Inf Serv* 41:187.
- Matsukage A, Hirose F, Yoo MA and Yamaguchi M (2008) The DRE/DREF transcriptional regulatory system: A master key for cell proliferation. *Biochim Biophys Acta Gene Regul Mech* 1779:81-89.
- McClellan DA and McCracken KG (2001) Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. *Mol Biol Evol* 18:917-925.
- McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG and Sailsbery JK (2005) Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol Biol Evol* 22:437-455.
- Naidoo N, Ferber M, Galante RJ, McShane B, Hu JH, Zimmerman J, Maislin G, Cater J, Wyner A, Worley P, et al. (2012) Role of homer proteins in the maintenance of sleep-wake states. *PLoS One* 7:e35174.
- Nylander JAA (2004) MrModeltest v2. Software distributed by the author. Evolutionary Biology Center, Uppsala University.
- Okano M, Xie S and Li E (1998) Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* 26:2536-40.
- Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC and Reuter G (2009) Retrotransposon silencing and telomere in-

- tegrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* 41:696-702.
- Ponger L and Li W (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22:1119-1128.
- Powell JR and DeSalle R (1995) *Drosophila* molecular phylogenies and their uses In: Hecht MK, MacIntyre RJ and Clegg MT (eds) *Evolutionary Biology*. Plenum Press, New York, pp 87-138.
- Rai K, Jafrí IF, Chidester S, James SR, Karpf AR, Cairns BR and Jones DA (2010) Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. *J Biol Chem* 285:4110-4121.
- Remsen J and O'Grady PM (2002) Phylogeny of Drosophilidae (Diptera), with comments on combined analysis and character support. *Mol Phylogenet Evol* 24:248-263.
- Robe LJ, Cordeiro J, Loreto ELS and Valente VLS (2010) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genetica* 138:601-617.
- Robe LJ, Valente VL, Budnik M and Loreto EL (2005) Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: A nuclear versus mitochondrial gene approach. *Mol Phylogenet Evol* 36:623-640.
- Ronquist F and Huelsenbeck JP (2003) MrBayes version 3.0: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
- Russo CA, Takezaki N and Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* 12:391-404.
- Saikia M, Krokowski D, Guan B-J, Ivanov P, Parisien M, Hu G, Anderson P, Pan T and Hatzoglou M (2012) Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. *J Biol Chem* 287:42708-42725.
- Sassi AK, Herédia FO, Loreto ELS, Valente VLS and Rohde C (2005) Transposable elements P and gypsy in natural populations of *Drosophila willistoni*. *Genet Mol Biol* 4:734-739.
- Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M and Lyko F (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* 24:1590-1595.
- Shanmugam R, Aklujkar M, Schäfer M, Reinhardt R, Nickel O, Reuter G, Lovley DR, Ehrenhofer-Murray A, Nellen W, Ancri S, et al. (2014) The Dnmt2 RNA methyltransferase homolog of *Geobacter sulfurreducens* specifically methylates tRNA-Glu. *Nucleic Acids Res* 42:6487-6496.
- Singh ND, Arndt PF and Petrov DA (2006) Minor shift in background substitutional patterns in the *Drosophila saltans* and *willistoni* lineages is insufficient to explain GC content of coding sequences. *BMC Biology* 4:37.
- Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* 5:233-241.
- Suganuma T, Gutiérrez JL, Li B, Florens L, Swanson SK, Washburn MP, Abmayr SM and Workman JL (2008) ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol* 15:364-372.
- Swinehart W and Jackman J (2015) Diversity in mechanism and function of tRNA methyltransferases. *RNA Biol* 6286:37-41.
- Swinehart WE, Henderson JC and Jackman JE (2013) Unexpected expansion of tRNA substrate recognition by the yeast m1G9 methyltransferase Trm10. *RNA* 19:1137-46.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447-D452.
- Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L, Martin DIK and Boffelli D (2014) Genome methylation in *Drosophila melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res* 24:821-30.
- Tang LY, Reddy MN, Rasheva V, Lee TL, Lin MJ, Hung MS and Shen CK (2003) The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. *J Biol Chem* 278:33613-33616.
- Tarrio R, Rodriguez-Trelles F and Ayala FJ (2001) Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Mol Biol Evol* 18:1464-1473.
- Tatarenkov A, Kwiatowski J, Skarecky D, Barrio E and Ayala FJ (1999) On the evolution of Dopa decarboxylase (Ddc) and *Drosophila* systematics. *J Mol Evol* 48:445-462.
- Thiagarajan D, Dev RR and Khosla S (2011) The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics* 6:103-113.
- Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King RC (ed) *Handbook of Genetics*, Vol. 3. Plenum Press, New York, pp 421-469.
- Tovy A, Hofmann B, Helm M and Ancri S (2010) *In vitro* tRNA methylation assay with the Entamoeba histolytica DNA and tRNA methyltransferase Dnmt2 (Ehmet) enzyme. *J Vis Exp* 2:2-7.
- Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G and Lyko F (2012) RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol* 19:900-905.
- Tuorto F, Herbst F, Alerasool N, Bender S, Popp O, Federico G, Reitter S, Liebers R, Stoecklin G, Gröne H-J, et al. (2015) The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *EMBO J* 34:2350-2362.
- Urizar NL, Yang Z, Edenberg HJ and Davis RL (2007) *Drosophila* Homer is required in a small set of neurons including the ellipsoid body for normal ethanol sensitivity and tolerance. *J Neurosci* 27:4541-4551.
- Vicario S, Moriyama EN and Powell JR (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol Biol* 7:226.
- Wang BC, Park J, Watabe HA, Gao JJ, Xiangyu JG, Aotsuka T, Chen HW and Zhang YP (2006) Molecular phylogeny of the *Drosophila virilis* section (Diptera: Drosophilidae) based on mitochondrial and nuclear sequences. *Mol Phylogenet Evol* 40:484-500.
- Woolley S, Johnson J, Smith MJ, Crandall KA and McClellan DA (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19:671-672.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-1591.

Yang Z and Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908-917.

Yoder JA and Bestor TH (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet* 7:279-284.

Zhang J, Nielsen R and Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472-2479.

Internet Resources Section

Bächli G (2016) TaxoDros: The database on taxonomy of Drosophilidae, v.1.04, Database 2017/2. Available via <http://taxodros.unizh.ch/>. (accessed 26 February 2016).

Flybase - <http://flybase.bio.indiana.edu/blast/>

Drosophila buzzatii Genome Project - <http://dbuz.uab.cat/welcome.php>

SpottedWingFlyBase - <http://spottedwingflybase.oregonstate.edu/>

Evolutionary Rate Covariation (ERC) Analysis - http://csb.pitt.edu/erc_analysis/

Supplementary material

The following online material is available for this article:

Table S1 – Drosophilidae species used in the present study.

Table S2 - Complete MT2 homologous sequences annotation.

Table S3 - GenBank accession numbers for nucleotide sequences obtained by DNA sequencing.

Table S4 - Standard error estimate(s) from estimates of evolutionary divergence between sequence pairs of different Drosophilidae species groups.

Table S5 - Estimates of Average Percentual Divergence over Sequence Pairs Groups.

Table S6 - ERC values for 36 genes obtained by Group ERC Analysis tool and STRING database.

Figure S1 - Dot blot screening for Dnmt2 sequences homologous to the *Drosophila melanogaster Dnmt2* gene.

Figure S2 - Multiple sequence alignment of DNMT2 showing the high conservation among *Drosophila* species.

Figure S3 - Bayesian phylogenetic analysis of *Dnmt2* using nucleotide sequences.

Associate Editor: Loreta B. Freitas

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S1 - Drosophilidae species used in the present study. Results from Dot blot and PCR screening are shown, as well as the species when the sequences used in this work were obtained.

Genus <i>Drosophila</i>					Genus <i>Sophophora</i>				
Group	Species	Dot	PCR	Sequence	Group	Species	Dot	PCR	Sequence
guarani	<i>D. ornathrons</i>	+	+	√	melanogaster	<i>D. melanogaster</i>	+	+	DB
	<i>D. subbadia</i>	+	+	√		<i>D. simulans</i>	+	+	DB
	<i>D. guaru</i>	+	+	√		<i>D. sechellia</i>	NT	+	DB
guaramunu	<i>D. griseolineata</i>	+	+	√		<i>D. mauritiana</i>	+	+	√
	<i>D. maculifrons</i>	+	+	√		<i>D. teissieri</i>	+	+	√
tripunctata	<i>D. nappae</i>	+	+	√		<i>D. santomea</i>	+	+	√
	<i>D. paraguayensis</i>	NT	-			<i>D. erecta</i>	+	+	DB
	<i>D. crocina</i>	+	+	√		<i>D. yakuba</i>	+	+	DB
	<i>D. paramediostrata</i>	+	-			<i>D. kikawai</i>	+	-	DB
caradani	<i>D. tripunctata</i>	+	+	√		<i>D. ananassae</i>	+	+	DB
	<i>D. mediodiffusa</i>	NT	+	√	<i>D. bipectinata</i>	NT	NT	DB	
	<i>D. mediotictoides</i>	+	-		<i>D. malerkotliana</i>	+	-		
	<i>D. neocaradani</i>	+	-		<i>D. orena</i>	+	-		
	<i>D. polymorpha</i>	+	-		<i>D. ficusphila</i>	NT	NT	DB	
pallidipennis	<i>D. pallidipennis</i>	NT	-		<i>D. takahashi</i>	NT	NT	DB	
calloptera	<i>D. ornathrons</i>	+	+	√	<i>D. rhopaloa</i>	NT	NT	DB	
	<i>D. immigrans</i>	+	+	√	<i>D. sukiki</i>	NT	NT	DB	
immigrans	<i>D. albomicans</i>	NT	NT	DB	<i>D. biarmipes</i>	NT	NT	DB	
	<i>D. funebris</i>	+	-		<i>D. eugracilis</i>	NT	NT	DB	
funebris	<i>D. funebris</i>	+	-		<i>D. elegans</i>	NT	NT	DB	
	<i>D. gasci</i>	+	+	√	obscura	<i>D. pseudoobscura</i>	+	+	DB
	<i>D. brncici</i>	NT	-			<i>D. miranda</i>	NT	NT	DB
	<i>D. gaucha</i>	+	+	√		<i>D. persimilis</i>	NT	NT	DB
<i>D. pavani</i>	NT	+	√	saltans		<i>D. prosaltans</i>	+	-	
mesophragmatica	<i>D. hydei</i>	+	+		√	<i>D. saltans</i>	+	-	
	<i>D. mercatorum</i>	+	-			<i>D. neoleptica</i>	+	-	
	<i>D. mojavensis</i>	+	+	DB	<i>D. sturtevantii</i>	+	-		
canalineia	<i>D. buzzatii</i>	NT	-	DB	wilsoni	<i>D. sucinea</i>	+	NT	
	<i>D. canalinea</i>	+	-			<i>D. nebulosa</i>	+	-	
flavopilosa	<i>D. cestrus</i>	?	-			<i>D. willistonii (Gd-H4)</i>	+	+	DB
	<i>D. incompta</i>	+	+	√		<i>D. willistonii (Wp-4)</i>	+	NT	
wilis	<i>D. wilis</i>	+	+	DB	<i>D. wilistonii (17A2)</i>	+	NT		
grimshawi	<i>D. grimshawi</i>	NT	NT	DB	<i>D. paulistorum</i>	+	-		
robusta	<i>D. robusta</i>	+	-		<i>D. insularis</i>	+	-		
busckii***	<i>D. busckii***</i>	+	-		<i>D. tropicalis</i>	+	+	√	
	<i>Z. indianus*</i>	+	-		<i>D. equinoxialis</i>	+	-		
	<i>Z. tuberculatus*</i>	+	-		<i>D. capri corni</i>	+	-		
	<i>S. latifasciaeformis**</i>	+	-		<i>D. fumipennis</i>	+	-		
	<i>S. lebanonensis**</i>	+	-						

(+) signal detected in Dot blot or PCR amplification; (-) no PCR-specific amplification; (NT) not tested; (√) Dnmt2 homologous sequence obtained by PCR; (DB) Dnmt2 homologous sequence obtained via data bank (Genome/GenBank)

*Zaprionus genus: *Z. indianus* and *Z. tuberculatus*

**Scaptodrosophila genus: *S. latifasciaeformis* and *S. lebanonensis*

***Drosilopa subgenus: *D. busckii*

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S2 - Complete MT2 homologous sequences annotation.

	<i>Annotation symbol</i>
<i>Drosophila albomicans</i>	gi 405988621 gb JH859534.1 <i>Drosophila albomicans</i> strain KM55-5 unplaced genomic scaffold Dalb_scaffold_67169, whole genome shotgun sequence
<i>Drosophila ananassae</i>	gn dana scaffold_12943 type=golden_path_region; loc=scaffold_12943.1..5039921; ID=scaffold_12943; dbxref=GB:CH902624
<i>Drosophila biarmipes</i>	gi 459197679 gb KB462833.1 <i>Drosophila biarmipes</i> unplaced genomic scaffold scf7180000302422, whole genome shotgun sequence
<i>Drosophila bipectinata</i>	gi 459198809 gb KB464388.1 <i>Drosophila bipectinata</i> unplaced genomic scaffold scf7180000396728, whole genome shotgun sequence
<i>Drosophila buzzatii</i>	lcl scaffold27 [847237 - 848191] and [847126 - 847185]
<i>Drosophila elegans</i>	gi 459200570 gb KB458413.1 <i>Drosophila elegans</i> unplaced genomic scaffold scf7180000491028, whole genome shotgun sequence
<i>Drosophila erecta</i>	gn dere scaffold_4929 type=golden_path_region; loc=scaffold_4929.1..26641161; ID=scaffold_4929; dbxref=GB:CH954177
<i>Drosophila eugracilis</i>	gi 459206169 gb KB464972.1 <i>Drosophila eugracilis</i> unplaced genomic scaffold scf7180000409122, whole genome shotgun sequence
<i>Drosophila ficusphila</i>	gi 459201587 gb KB457400.1 <i>Drosophila ficusphila</i> unplaced genomic scaffold scf7180000453936, whole genome shotgun sequence
<i>Drosophila grimshawi</i>	gn dgr scaffold_15252 type=golden_path_region; loc=scaffold_15252.1..17193109; ID=scaffold_15252; dbxref=GB:CH916368
<i>Drosophila kikkawai</i>	gi 459202993 gb KB459629.1 <i>Drosophila kikkawai</i> unplaced genomic scaffold scf7180000302408, whole genome shotgun sequence
<i>Drosophila melanogaster</i>	FlyBase_Annotation_IDs:CG10692-PC. Accession number: AAF53163.2
<i>Drosophila miranda</i>	gi 480995217 gb CMD01520.2 <i>Drosophila miranda</i> strain MSH22 chromosome 4, whole genome shotgun sequence
<i>Drosophila mojavensis</i>	gn dmoj scaffold_6500 type=golden_path_region; loc=scaffold_6500.1..32352404; ID=scaffold_6500; dbxref=GB:CH933807
<i>Drosophila persimilis</i>	gn dper scaffold_8 type=golden_path_region; loc=scaffold_8.1..3966273; ID=scaffold_8; dbxref=GB:CH479187
<i>Drosophila pseudoobscura</i>	gn dpse 4_group2 type=golden_path_region; loc=4_group2.1..1235759; ID=4_group2; dbxref=GB:CH379059
<i>Drosophila rhopaloa</i>	gi 452191607 gb KB448653.1 <i>Drosophila rhopaloa</i> unplaced genomic scaffold scf7180000761302, whole genome shotgun sequence
<i>Drosophila sechellia</i>	gn dsec scaffold_16 type=golden_path_region; loc=scaffold_16.1..1878335; ID=scaffold_16; dbxref=GB:CH480831
<i>Drosophila similans</i>	gn dsim 2L type=golden_path_region; loc=2L.1..22036055; ID=2L; dbxref=GB:CM000361
<i>Drosophila sukui</i>	scaffold1 [22179055 - 22179114] - [22179153 - 22180145]
<i>Drosophila takahashii</i>	gi 459204042 gb KB461661.1 <i>Drosophila takahashii</i> unplaced genomic scaffold scf7180000415863, whole genome shotgun sequence
<i>Drosophila virilis</i>	gn dvir scaffold_12963 type=golden_path_region; loc=scaffold_12963.1..20206255; ID=scaffold_12963; dbxref=GB:CH940649
<i>Drosophila willistoni</i>	Dwil\GK21086-PB type=CDS; loc=scf2_1100000004851.join(40528.40546,40598.41559); name=Dwil\GK21086-RB; dbxref=FlyBase:FBpp0378827 / Garcia et al. 2007
<i>Drosophila yakuba</i>	gn dyak 2L type=golden_path_region; loc=2L.1..22324452; ID=2L; dbxref=GB:CM000157

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S3 - GenBank accession numbers for nucleotide sequences obtained by DNA sequencing performed by MacroGen Inc.

Accession numbers		
<i>Drosophila tropicalis</i> (clone 1)	Seq1	KJ873136
<i>Drosophila tropicalis</i> (clone 2)	Seq2	KJ873137
<i>Drosophila tripunctata</i>	Seq3	KJ873138
<i>Drosophila teisaien</i> (clone 4)	Seq4	KJ873139
<i>Drosophila teisaien</i> (clone 3)	Seq5	KJ873140
<i>Drosophila teisaien</i> (clone 2)	Seq6	KJ873141
<i>Drosophila subbadia</i>	Seq7	KJ873142
<i>Drosophila pavani</i>	Seq8	KJ873143
<i>Drosophila ornaiyennis</i> (clone 1)	Seq9	KJ873144
<i>Drosophila ornaiyennis</i> (clone 2)	Seq10	KJ873145
<i>Drosophila ornaiyfrons</i> (clone 3)	Seq11	KJ873146
<i>Drosophila ornaiyfrons</i> (clone 2)	Seq12	KJ873147
<i>Drosophila ornaiyfrons</i> (clone 1)	Seq13	KJ873148
<i>Drosophila nappae</i>	Seq14	KJ873149
<i>Drosophila mediodifusa</i>	Seq15	KJ873150
<i>Drosophila maculifrons</i>	Seq16	KJ873151
<i>Drosophila incompta</i>	Seq17	KJ873152
<i>Drosophila immigrans</i>	Seq18	KJ873153
<i>Drosophila hydei</i>	Seq19	KJ873154
<i>Drosophila guaru</i>	Seq20	KJ873155
<i>Drosophila griseolineata</i>	Seq21	KJ873156
<i>Drosophila gaucha</i>	Seq22	KJ873157
<i>Drosophila gaisi</i>	Seq23	KJ873158
<i>Drosophila crocina</i>	Seq24	KJ873159

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S4 - Standard error estimate(s) from estimates of evolutionary divergence between sequence pairs of different *Drosophilidae* species groups. Amino acid standard error is given below the diagonal. The nucleotide standard error is provided above the diagonal. The values are given in percentage.

Groups	1	2	3	4	5	6	7	8	9	10	11	12	13
1 calloptera		1.62	1.51	1.11	0.84	1.38	1.31	1.51	1.63	1.39	0.90	1.50	1.63
2 flavopilosa	2.82		1.52	1.53	1.45	1.33	1.31	1.25	1.56	1.14	1.32	1.31	1.55
3 grimshawi	2.79	2.57		1.55	1.43	1.36	1.37	1.40	1.56	1.34	1.34	1.37	1.63
4 guaramunu	1.75	2.64	2.69		0.96	1.29	1.28	1.42	1.60	1.37	0.76	1.40	1.58
5 guarani	1.38	2.73	2.65	1.59		1.22	1.19	1.33	1.44	1.24	0.82	1.32	1.49
6 immigrans	2.31	2.44	2.48	2.23	2.21		1.18	1.34	1.38	1.21	1.09	1.29	1.48
7 melanogaster	2.53	2.57	2.48	2.48	2.44	2.36		1.30	1.11	1.25	1.13	1.21	1.29
8 mesophragmatica	2.62	2.31	2.55	2.42	2.52	2.41	2.52		1.57	1.07	1.23	1.24	1.53
9 obscura	2.84	2.89	2.66	2.73	2.65	2.55	2.18	2.78		1.47	1.38	1.49	1.54
10 repleta	2.54	2.10	2.34	2.33	2.44	2.31	2.44	1.70	2.68		1.18	1.14	1.46
11 tripunctata	1.66	2.57	2.55	1.34	1.55	2.12	2.37	2.33	2.60	2.24		1.20	1.42
12 virilis	2.56	2.13	2.33	2.32	2.46	2.14	2.36	1.92	2.68	1.93	2.24		1.55
13 willistoni	2.94	2.91	2.84	2.87	2.84	2.64	2.66	2.90	2.71	2.75	2.79	2.82	

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S5 - Estimates of Average Percentual Divergence over Sequence Pairs Groups. The values are given in percentage.

Subgenus	Nucleotides	Within Groups		S.E.
		S.E.	Amino Acids	
<i>Drosophila</i>	21.94	0.94	17.89	1.65
<i>Sophophora</i>	22.65	0.89	18.17	1.62
		Between Groups		
<i>Drosophila vs Sophophora</i>	31.37	1.17	25.35	2.14

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

Table S6 - ERC values for 36 genes obtained by Group ERC Analysis tool and STRING database. The matrix shows all pairwise ERC values between genes below the diagonal and respectively p-values above the diagonal. Cells are shaded red according to the intensity of their deviation from the null expectation.

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36		
CG10	N/A	0.0	0.0	0.0	0.3	0.1	0.8	0.1	0.3	0.1	0.3	0.0	0.4	0.8	0.3	0.8	0.9	0.9	0.5	0.9	0.9	0.7	0.8	0.4	0.6	0.8	0.8	0.8	0.9	0.8	0.7	0.9	0.8	0.6	0.5	0.4		
1 262	0.0	0.0	0.0	0.3	0.1	0.8	0.1	0.3	0.1	0.3	0.0	0.4	0.8	0.3	0.8	0.9	0.9	0.5	0.9	0.9	0.7	0.8	0.4	0.6	0.8	0.8	0.8	0.9	0.8	0.7	0.9	0.8	0.6	0.5	0.4	0.4		
CG161	0.0	0.0	0.2	0.7	0.4	0.9	0.4	0.6	0.1	0.4	0.1	0.2	0.3	0.2	0.4	0.8	0.4	0.8	0.8	0.8	0.3	0.7	0.6	0.6	0.8	0.7	0.7	0.4	0.5	0.7	0.4	0.4	0.2	0.4	0.4	0.4	0.4	
2 33	0.0	0.0	0.2	0.7	0.4	0.9	0.4	0.6	0.1	0.4	0.1	0.2	0.3	0.2	0.4	0.8	0.4	0.8	0.8	0.8	0.3	0.7	0.6	0.6	0.8	0.7	0.7	0.4	0.5	0.7	0.4	0.4	0.2	0.4	0.4	0.4	0.4	
3 Thor	0.0	0.6	N/A	0.0	0.3	0.3	0.9	0.3	0.5	0.3	0.9	0.6	0.7	0.7	0.2	0.6	0.7	0.4	0.6	0.9	0.7	0.1	0.5	0.3	0.6	0.5	0.8	0.9	0.5	0.1	0.8	0.4	0.3	0.0	0.7	0.3	0.4	
CG17	0.0	0.1	0.6	N/A	0.1	0.5	0.7	0.3	0.5	0.5	0.9	0.7	0.9	0.6	0.2	0.5	0.5	0.6	0.7	0.9	0.9	0.5	0.7	0.5	0.6	0.5	0.7	0.7	0.9	0.8	0.9	0.8	0.7	0.3	0.6	0.2	0.6	
4 134	0.0	0.1	0.6	N/A	0.1	0.5	0.7	0.3	0.5	0.5	0.9	0.7	0.9	0.6	0.2	0.5	0.5	0.6	0.7	0.9	0.9	0.5	0.7	0.5	0.6	0.5	0.7	0.7	0.9	0.8	0.9	0.8	0.7	0.3	0.6	0.2	0.6	
5 Rpd3	0.0	0.1	0.4	N/A	0.0	0.0	0.9	0.9	0.4	0.8	0.9	0.4	0.7	0.9	0.3	0.7	0.7	0.9	0.8	0.9	0.8	0.5	0.5	0.2	0.2	0.4	0.5	0.2	0.7	0.4	0.2	0.6	0.8	0.5	0.7	0.2	0.2	
Su(vn	0.0	0.0	0.1	0.0	0.0	0.0	0.9	0.9	0.2	0.5	0.4	0.0	0.2	0.5	0.5	0.9	0.7	0.9	0.8	0.8	0.8	0.7	0.7	0.5	0.4	0.6	0.6	0.1	0.6	0.3	0.1	0.2	0.5	0.4	0.4	0.4	0.4	
1 rj3-9	0.0	0.0	0.1	0.0	0.0	0.0	0.9	0.9	0.2	0.5	0.4	0.0	0.2	0.5	0.5	0.9	0.7	0.9	0.8	0.8	0.8	0.7	0.7	0.5	0.4	0.6	0.6	0.1	0.6	0.3	0.1	0.2	0.5	0.4	0.4	0.4	0.4	
7 MBD-like	0.0	0.3	0.5	0.2	0.4	0.6	N/A	0.0	0.7	0.4	0.1	0.5	0.4	0.2	0.5	0.1	0.2	0.0	0.0	0.0	0.1	0.3	0.4	0.6	0.5	0.3	0.4	0.6	0.4	0.8	0.7	0.5	0.3	0.7	0.2	0.6		
8 Eno	0.0	0.0	0.1	0.1	0.4	0.6	0.4	N/A	0.4	0.0	0.1	0.4	0.4	0.2	0.1	0.1	0.5	0.1	0.1	0.1	0.2	0.2	0.3	0.4	0.6	0.5	0.5	0.6	0.8	0.6	0.7	0.9	0.8	0.5	0.4	0.8	0.8	
CG74	0.0	0.1	0.1	0.1	0.4	0.6	0.4	N/A	0.4	0.0	0.1	0.4	0.4	0.2	0.1	0.1	0.5	0.1	0.1	0.1	0.2	0.2	0.3	0.4	0.6	0.5	0.5	0.6	0.8	0.6	0.7	0.9	0.8	0.5	0.4	0.8	0.8	
1 egg	0.0	0.2	0.1	0.0	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.4	0.4	0.8	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	
Su(vn	0.0	0.0	0.5	0.6	0.0	0.0	0.4	0.3	0.2	0.5	N/A	0.0	0.0	0.0	0.4	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
1 rj3-3	0.0	0.0	0.5	0.6	0.0	0.0	0.4	0.3	0.2	0.5	N/A	0.0	0.0	0.0	0.4	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1 36	0.0	0.2	-	-	0.0	0.4	-	-	-	-	-	-	0.0	0.1	0.2	0.8	0.9	0.9	0.6	0.7	0.7	0.8	0.9	0.8	0.4	0.7	0.4	0.4	0.7	0.6	0.2	0.5	0.7	0.7	0.4	0.6
2 Rel	0.0	0.0	0.1	0.2	0.0	0.4	0.0	0.8	0.3	0.0	0.4	0.4	0.0	0.1	0.2	0.8	0.9	0.9	0.6	0.7	0.7	0.8	0.9	0.8	0.4	0.7	0.4	0.4	0.7	0.6	0.2	0.5	0.7	0.7	0.4	0.6
1 CG99	0.0	0.2	0.2	0.5	0.3	0.1	0.0	0.0	0.0	0.8	0.4	0.4	0.0	0.4	0.3	0.6	0.5	0.4	0.1	0.1	0.1	0.2	0.7	0.5	0.0	0.5	0.2	0.2	0.3	0.4	0.3	0.5	0.7	0.7	0.4	0.9
3 73	0.0	0.2	0.2	0.5	0.3	0.1	0.0	0.0	0.0	0.8	0.4	0.4	0.0	0.4	0.3	0.6	0.5	0.4	0.1	0.1	0.1	0.2	0.7	0.5	0.0	0.5	0.2	0.2	0.3	0.4	0.3	0.5	0.7	0.7	0.4	0.9
1 CG67	0.0	0.1	0.2	0.2	0.4	0.0	0.2	0.2	0.0	0.3	0.5	0.2	0.4	N/A	0.0	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.6	0.6	0.1	0.1	0.0	0.1	0.2	0.4	0.5	0.2	0.0	0.4	0.3	0.6
4 12	0.0	0.1	0.2	0.2	0.4	0.0	0.2	0.2	0.0	0.3	0.5	0.2	0.4	N/A	0.0	0.0	0.1	0.1	0.2	0.1	0.1	0.1	0.6	0.6	0.1	0.1	0.0	0.1	0.2	0.4	0.5	0.2	0.0	0.4	0.3	0.6
1 Su(vn	0.0	0.2	0.2	0.3	0.1	0.0	0.3	0.3	0.2	0.0	0.1	0.0	0.4	N/A	0.0	0.4	0.4	0.2	0.4	0.4	0.0	0.6	0.4	0.1	0.2	0.1	0.2	0.3	0.3	0.5	0.6	0.4	0.2	0.9	0.7	0.5
5 rj205	0.0	0.0	0.4	0.9	0.1	0.0	0.3	0.3	0.2	0.1	0.0	0.4	N/A	0.0	0.4	0.4	0.2	0.4	0.4	0.0	0.6	0.4	0.1	0.2	0.1	0.2	0.3	0.3	0.5	0.6	0.4	0.2	0.9	0.7	0.5	0.5
1 mei-	0.0	0.0	0.1	0.0	0.2	0.4	0.3	0.3	0.5	0.2	0.0	0.2	0.5	0.6	N/A	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.0	0.0	0.1	0.5	0.3	0.1	0.4	0.8	0.8	
6 S332	0.0	0.0	0.1	0.0	0.2	0.4	0.3	0.3	0.5	0.2	0.0	0.2	0.5	0.6	N/A	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.4	0.0	0.0	0.1	0.5	0.3	0.1	0.4	0.8	0.8	
1 CG14	0.0	0.3	0.2	0.0	0.2	0.2	0.2	0.0	0.4	0.5	0.0	0.4	0.0	0.3	0.0	0.6	0.7	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.4	0.1	0.0	0.4	0.3	0.6	
7 618	0.0	0.3	0.2	0.0	0.2	0.2	0.2	0.0	0.4	0.5	0.0	0.4	0.0	0.3	0.0	0.6	0.7	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.4	0.1	0.0	0.4	0.3	0.6	
1 CG14	0.0	0.0	0.0	0.1	0.4	0.5	0.4	0.2	0.0	0.0	0.4	0.0	0.2	0.0	0.0	0.6	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.1	0.0	0.1	0.4	0.1	0.0	0.2	0.5	0.8
8 906	0.0	0.0	0.0	0.1	0.4	0.5	0.4	0.2	0.0	0.0	0.4	0.0	0.2	0.0	0.0	0.6	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.1	0.0	0.1	0.4	0.1	0.0	0.2	0.5	0.8
1 Atac3	0.0	0.2	0.1	0.2	0.3	0.3	0.4	0.3	0.0	0.1	0.1	0.0	0.2	0.2	0.3	0.5	0.7	0.8	0.7	0.8	0.8	0.0	0.1	0.3	0.0	0.0	0.0	0.2	0.1	0.1	0.3	0.3	0.0	0.2	0.3	0.8
2 CTCF	0.0	0.3	0.5	0.6	0.4	0.3	0.4	0.2	0.3	0.2	0.3	0.1	0.2	0.3	0.0	0.6	0.5	0.7	0.8	0.8	0.8	0.0	0.1	0.3	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.6	0.9
2 CG16	0.0	0.2	0.3	0.4	0.4	0.3	0.3	0.1	0.5	0.4	0.1	0.2	0.2	0.3	0.0	0.6	0.7	0.8	0.7	0.9	0.9	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.6	0.9
1 S63	0.0	0.2	0.3	0.4	0.4	0.3	0.3	0.1	0.5	0.4	0.1	0.2	0.2	0.3	0.0	0.6																				

Genetics and Molecular Biology

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
2 Rpp240	0.02	0.00	0.180	0.050	0.230	0.015	0.007	0.095	0.551	0.698	0.449	0.316	0.028	0.029	0.070	0.489	0.441	0.513	0.440	0.506	0.651	0.797	0.777	N/A	0.012	0.018	0.014	0.049	0.070	0.095	0.091	0.159	0.299	0.347	0.564	0.842	
2 RhoG5EF4	0.073	0.000	0.273	0.128	0.167	0.019	0.000	0.458	0.445	0.447	0.140	0.222	0.939	0.847	0.793	0.638	0.337	0.313	0.000	0.480	0.541	0.642	0.498	0.716	N/A	0.002	0.024	0.025	0.048	0.188	0.178	0.468	0.952	0.525	0.509	0.790	0.873
2 Dant62	0.225	0.555	0.114	0.088	0.033	0.103	0.071	0.034	0.499	0.484	0.290	0.226	0.067	0.278	0.224	0.663	0.711	0.693	0.699	0.589	0.712	0.741	0.650	0.657	N/A	0.017	0.049	0.039	0.042	0.080	0.283	0.433	0.355	0.667	0.593	0.836	
2 Orc21	0.204	0.145	0.384	0.249	0.089	0.090	0.060	0.133	0.430	0.190	0.036	0.000	0.210	0.485	0.333	0.000	0.512	0.425	0.509	0.508	0.585	0.616	0.375	0.470	0.686	N/A	0.044	0.093	0.042	0.169	0.930	0.027	0.777	0.678	0.883	0.644	
2 hay6	0.216	0.186	0.794	0.247	0.261	0.372	0.168	0.384	0.378	0.451	0.107	0.059	0.334	0.262	0.481	0.572	0.369	0.297	0.208	0.508	0.563	0.653	0.440	0.582	0.605	0.771	0.741	N/A	0.008	0.049	0.543	0.098	0.099	0.429	0.429	0.933	0.822
2 hop9	0.420	0.000	0.102	0.557	0.231	0.040	0.028	0.131	0.387	0.252	0.028	0.110	0.186	0.126	0.540	0.503	0.629	0.399	0.747	0.762	0.805	0.588	0.528	0.351	0.602	0.600	0.774	N/A	0.003	0.940	0.033	0.003	0.133	0.208	0.393	0.963	0.844
3 mus40	0.300	0.009	0.336	0.387	0.225	0.020	0.275	0.280	0.400	0.309	0.236	0.388	0.061	0.039	0.106	0.344	0.426	0.445	0.418	0.507	0.565	0.686	0.521	0.483	0.313	0.622	0.542	0.831	0.845	N/A	0.027	0.506	0.273	0.239	0.861	0.179	
3 mus2109	0.168	0.176	0.450	0.578	0.218	0.373	0.233	0.432	0.009	0.134	0.049	0.143	0.136	0.051	0.085	0.010	0.202	0.052	0.000	0.409	0.295	0.185	0.479	0.336	0.355	0.349	0.397	0.621	0.551	0.703	N/A	0.923	0.352	0.943	0.759	0.986	
3 pmt9	0.350	0.000	0.064	0.204	0.153	0.247	0.002	0.380	0.157	0.220	0.053	0.052	0.176	0.151	0.087	0.075	0.373	0.400	0.026	0.447	0.477	0.370	0.307	0.097	0.644	0.000	0.609	0.596	0.631	0.476	N/A	0.010	0.602	0.829	0.397	0.179	
3 Stu(vn3)2-104	0.263	0.000	0.116	0.220	0.402	0.157	0.060	0.029	0.111	0.225	0.050	0.169	0.434	0.016	0.311	0.480	0.597	0.608	0.429	0.424	0.421	0.323	0.099	0.125	0.590	0.421	0.497	0.524	0.481	0.264	0.811	N/A	0.068	0.069	0.294	0.494	
3 Cap5	0.094	0.142	0.512	0.168	0.016	0.044	0.246	0.045	0.404	0.274	0.239	0.217	0.047	0.035	0.041	0.056	0.267	0.239	0.002	0.011	0.422	0.031	0.031	0.209	0.172	0.018	0.343	0.359	0.488	0.193	0.352	0.461	N/A	0.746	0.505	0.433	

Genetics and Molecular Biology

Protein	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
3 CG133035	0.004	0.018	0.328	0.153	0.279	0.053	0.233	0.293	0.274	0.101	0.237	0.016	0.148	0.136	0.490	0.278	0.170	0.037	0.000	0.122	0.155	0.506	0.232	0.383	0.381	0.043	0.009	0.126	0.206	0.197	0.072	0.321	0.399	0.228	N/A	0.022	0.225
3 Hsp262	0.000	0.000	0.089	0.316	0.171	0.064	0.120	0.406	0.293	0.195	0.175	0.128	0.427	0.135	0.235	0.385	0.110	0.408	0.381	0.625	0.634	0.515	0.557	0.533	0.567	0.364	0.330	0.000	0.398	0.303	0.300	0.127	0.057	0.008	0.552	N/A	0.522

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

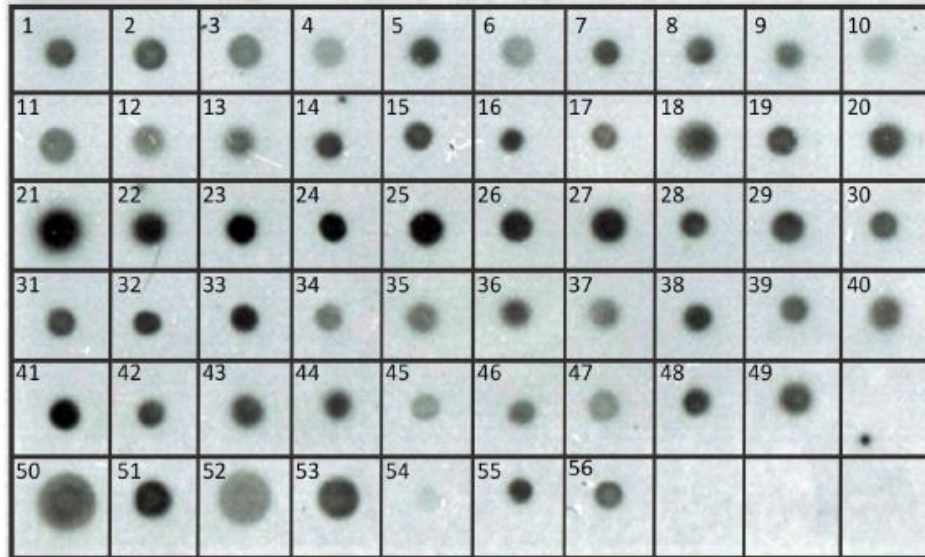


Figure S1 - Dot blot screening for *Dnmt2* sequences homologous to the *Drosophila melanogaster* *Dnmt2* gene. The species tested are the following: 1- *Drosophila ornatafrons*; 2- *D. subbada*; 3- *D. guaru*; 4- *D. griseolineata*; 5- *D. nappae*; 6- *D. parameãoestriata*; 7- *D. tripunctata*; 8- *D. medipictoides*; 9- *D. neocarãni*; 10- *D. polymorpha*; 11- *D. ornatepennis*; 12- *D. immitigrans*; 13- *D. funebris*; 14- *D. gasici*; 15- *D. gaucha*; 16- *D. mercatorum*; 17- *D. mojavenãis*; 18- *D. incompta*; 19- *D. virilis*; 20- *D. robusta*; 21- *D. melanogaster*; 22- *D. simulans*; 23- *D. mauritiana*; 24- *D. teissieri*; 25- *D. santomea*; 26- *D. erecta*; 27- *D. yakuba*; 28- *D. kikkawai*; 29- *D. ananassae*; 30- *D. malerkotliana*; 31- *D. pseudoobscura*; 32- *D. prosaltans*; 33- *D. saltans*; 34- *D. neoelliptica*; 35- *D. sturtevantii*; 36- *D. sucinea*; 37- *D. nebulosa*; 38- *D. willistoni* (ww strain); 39- *D. paulistorum* Orinocã; 40- *D. insularis*; 41- *D. tropicalis*; 42- *D. equinoxialis*; 43- *D. capricorni*; 44- *D. funipennis*; 45- *D. busckii*; 46- *Zaprionus indianus*; 47- *Z. tuberculatus*; 48- *S. laifasciaeformis*; 49- *S. lebanonensis*; 50- *D. maculifrons*; 51- *D. crocina*; 52- *D. hydei*; 53- *D. canalinea*; 54- *D. orena*; 55- *D. willistoni* (Wip-4 strain); 56- *D. willistoni* (17A2 strain).

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

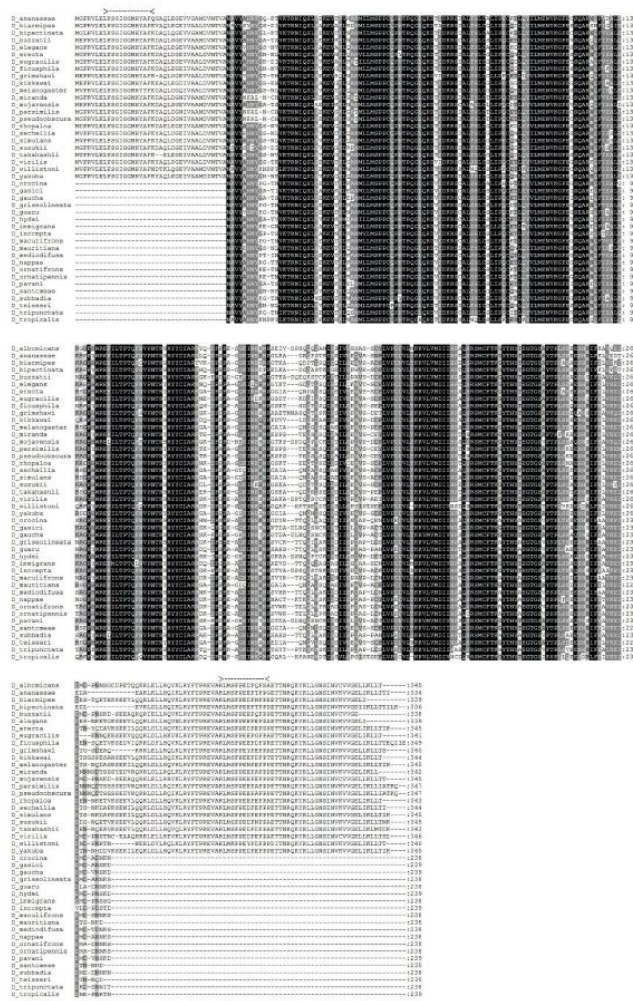


Figure S2 - Multiple sequence alignment of DNMT2 showing the high conservation among *Drosophila* species. Dashes indicate unknown amino acids. Black boxes indicate conserved in all sequences. The positions of PCR primers are indicated by dashed lines.

Supplementary Material to “Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions”

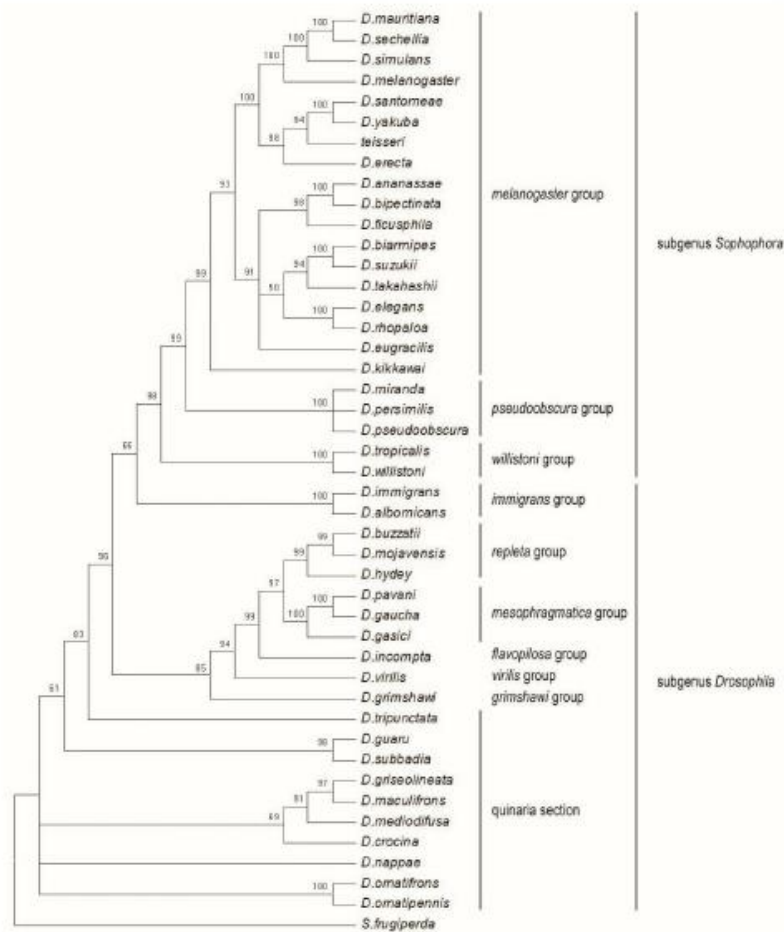


Figure S3 - Bayesian phylogenetic analysis of *Dmmt2* using nucleotide sequences alignment with the GTR + I + G model. Sequence of *Spodoptera frugiperda* was used as outgroup.

CAPÍTULO III

Characterizing the transcriptional expression and *in situ* localization of the *Dnmt2* gene in *Drosophila willistoni*

Gilberto Cavalheiro Vieira^{1,3}, Marícia Fantinel D'Ávila², Rebeca Zanini^{1,3}, Pâmela Silva de Oliveira^{1,3}, Maríndia Deprá^{1,3,4}, Carolina Flores Garcia^{1,3}, Vera Lúcia da Silva Valente^{1, 3, 4}

¹Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

²Departamento de Zoologia e Ciências Biológicas, Universidade Federal do Pampa (Unipampa), São Gabriel, RS, Brasil.

³ Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

⁴Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil.

Manuscrito em preparação para ser submetido ao periódico BMC Genomics

Abstract

Organisms that have only the DNA methyltransferase 2 (Dnmt2) to mediate the DNA methylation are called "Dnmt2-only" and they have been investigated in recent surveys. *Drosophila* is one of the "Dnmt2-only" organisms and is also an ideal model for Dnmt2 research. Homologs for this enzyme are found in virtually all organisms already investigated for the DNA methylation, including *Drosophila melanogaster*, but the biological function of the Dnmt2 protein is still uncertain. Some studies have pointed to a putative role during the early stages of invertebrate development. In this work, we present our findings on the *Dnmt2* expression in *D. willistoni*, a neotropical species of large ecological versatility with peculiar molecular features, when in comparison to other species of the same genus. We demonstrate here the presence of transcripts not only in the early stages of development, as described for other species, but also during the oogenesis. Our findings give insights on the possible biological function of *Dnmt2*-related processes associated with the development and differentiation of oocytes since germinative tissue formation seems to require a higher expression of *Dnmt2*. Additionally, we mapped the *Dnmt2* probe on the IIL chromosome arm (Muller's B element) of *D. willistoni*, near at the end of the singular telomeric region.

Keywords: Dnmt2, DNA methylation, *Drosophila willistoni*, embryogenesis, oogenesis

Introduction

Cytosine-5 RNA methylation is one among more than 100 distinct RNA modifications that has been already detected in tRNA, rRNA and mRNA (Li and Mason 2014), representing an important epigenetic modification that regulates gene expression in eukaryotes, with a standing role on their development and for the etiology of human diseases and mediating chromatin organization (Klose and Bird 2006; Zhao et al. 2016; Cheng et al. 2018). Elucidation of the molecular mechanisms mediating RNA and DNA methylation is crucial to understanding the roles that diverse nucleic acids play in the regulation of genetic information.

Although DNA methylation appears to be a widespread epigenetic regulatory mechanism, genomes are methylated in different ways in diverse organisms. DNA methylation in plant genomes, for instance, can occur symmetrically at cytosine nucleotides in both CG and CHG (H= A, T or C) contexts, whereas DNA methylation systems are well characterized in vertebrates (Bird 2002) and occur mostly symmetrically at the cytosine of a CG dinucleotide (Lee et al. 2010; Jin et al. 2011). The cytosine methylation is established and maintained by a family of conserved methyltransferases. In eukaryotes, there are three distinct families of DNA methyltransferases: Dnmt1, Dnmt2, and Dnmt3. Whereas Dnmt3 enzymes seem to be responsible for establishing DNA methylation patterns (*de novo* methyltransferases), Dnmt1 enzymes are involved in the maintenance of the methylation patterns (Goll and Bestor 2005). On the other hand, the biochemical activity and the biological function of Dnmt2 enzymes are still an open case.

Evidence for DNA methylation has been described and its importance demonstrated in different insect orders. The honeybee genome contains genes that encode orthologues of all vertebrate proteins required for DNA methylation (Schaefer and Lyko 2007; Elango et al. 2009). In addition to Dnmt2 (also found in Diptera), two CpG-specific DNMT family genes were identified: Dnmt1 and Dnmt3a/b genes (Wang et al. 2006). The specificity for CpT and CpA nucleotide residues distinguishes Dnmt2 from all other known animal DNA methyltransferases and confirms the previous suggestion of predominant non-CpG methylation in *Drosophila* (Lyko 2001; Kunert et al. 2003). The *Apis mellifera* genome shows greater similarities to vertebrate genomes than *Drosophila* and *Anopheles* genomes for the genes involved in DNA methylation, among others (The

Honeybee Genome Sequencing Consortium, 2006). For these reasons, methylation in *D. melanogaster* and other invertebrates remains a controversial research area.

Regarding the genus *Drosophila*, the species of this genus are so-called “*Dnmt2*-only” organisms, not containing any of canonical vertebrates DNA methyltransferases homologs (*Dnmt1* and *Dnmt3*). For the species of this genus studied so far, methylation was demonstrated at non-CpG nucleotides (Lyko et al. 2000; Kunert et al. 2003; Garcia et al. 2007), in contrast to most organisms already analyzed. Furthermore, the specificity for CpT, CpA, and CpC nucleotide residues distinguishes *Dnmt2* from all other known animal DNA methyltransferases (Kunert et al. 2003; Panikar et al. 2015), and the functional role of DNA methylation in *Drosophila* remains unclear. However, adding to that discussion, a recent study describes the importance of the methyl-CpG binding domain proteins to inhibit interspecies courtship and promoting aggression in *Drosophila* (Gupta et al. 2017). Another important difference between DNA methylation in mammals and *Drosophila* is the level of the cytosine nucleotides methylated. Whereas mammals have 2 to 10% of all cytosine residues modified to 5-methylcytosine (Li and Zhang 2014; Breiling and Lyko 2015), *Drosophila melanogaster* has only about 0.1 - 0.6% (Gowher et al. 2000b; Lyko et al. 2000), making it experimentally difficult to demonstrate unambiguously DNA methylation in this organism. Moreover, in contrast to the pattern of genome-wide DNA methylation in vertebrates, DNA methylation is relatively scattered in invertebrates (Suzuki and Bird 2008).

Several attempts were made to demonstrate the DNA methylation activity of *Dnmt2* in *D. melanogaster* (Kunert et al. 2003; Mund et al. 2004) revealing a low, but significant activity by distinct experimental methods. Overexpression of *Dnmt2* in *Drosophila* species appears to enhance genome-wide DNA methylation from 0.2% to 0.4 – 0.7%, as determined by capillary electrophoresis (Kunert et al. 2003; Mund et al. 2004). Nevertheless, in a recent study reported that *Dnmt2* controls DNA methylation in early *D. melanogaster* embryos and provide insight into its function in control of retrotransposon silencing and telomere integrity in somatic cells (Phalke et al. 2009). This report showed a clear-cut difference of methylation within *Invader4* elements LTRs between wild type and *Dnmt2*^{-/-} flies (Phalke et al. 2009). *Dnmt2* appears to mediate methylation on *D. melanogaster* embryos genome, even though both this activity and its functional consequences remaining poorly understood. However, a recent study showed that cytosine methylation in the genome of *Drosophila melanogaster* probably is

independent of Dnmt2 activity (Takayama et al. 2014). These findings show how elusive is the genomic methylation context in drosophilids.

Previous studies reported by our research group in the neotropical *D. willistoni* and its related species (*willistoni* subgroup) shows a distinct scenario. Adult flies show sex-specific patterns of rDNA genes investigated by Methylation Restriction Sensitive Endonucleases (Garcia et al. 2007; D'Ávila et al. 2010). Comparisons of *D. willistoni* and *D. melanogaster* Dnmt2 protein sequences indicated higher primary structure conservation on motifs responsible for the catalysis of methyl transfer and great variability in the region related to specific recognition of target DNA sequences. These outcomes from the *willistoni* subgroup species are encouraging due to the previously reported peculiarities of this species when compared to other species of the *Drosophila* genus that have their genomes sequenced (Clark et al. 2007). *D. willistoni* singularities correspond to overall genome size, distribution of transposable element classes, patterns of codon usage, dot chromosome lacking and unclear phylogenetic clustering placement, among others (Clark et al. 2007; Vicario et al. 2007; Schaefer et al. 2008).

In addition to the peculiarity of having sex-specific methylation (Garcia et al. 2007; D'Ávila et al. 2010), *D. willistoni* is known to have extensive gene arrangement polymorphisms on all chromosomes (da Cunha and Dobzhansky 1950; da Cunha and Dobzhansky 1954; Valente and Araújo 1986; Rohde and Valente 2012) as observed from chromosomal variability in natural populations. As a result of its higher rate of intraspecific polymorphism and significantly large independent evolutionary time compared to other species of *Drosophila* 12 Genomes Project, computational methods seem to demonstrate some ambiguity in its phylogenetic placement (Schaeffer et al. 2008). The ambiguity arises as a consequence of the elevated level of sequence and gene-order evolution that leads to a significantly long lineage leading to *D. willistoni*. Most phylogenetic reconstruction software tends to force *D. willistoni* as an outgroup, based on the high evolutionary rates. In light of these reasons, studies are needed to map more genes in the genome of *D. willistoni* in order to elucidate the discrepancies found in the previous reports. New approaches, using an additional marker will not only identify more genes to confirm the ordering of scaffolds assembled so far but will also facilitate further comparative studies by increasing the database from this species.

To improve the Dnmt2 knowledge about its biological function and importance, more assays for characterization and detection of Dnmt2-dependent DNA/RNA methylation will have to be established. Genomic DNA methylation patterns need to be characterized mostly in “*Dnmt2*-only” model systems, like *Drosophila*. In the present study, we detected and quantified the expression of *D. willistoni Dnmt2* in different development stages. Accordingly, we also detected the expression of Dnmt2 on oogenesis and embryogenesis, indicating a possible contribution of *Dnmt2* expression during development. In addition, we localized *Dnmt2* homologs on polytene chromosomes of *D. willistoni* and we attempted to establish cause and/or effect relationships for the exact location of the gene with its adjacent sites and possible activity modulated by position.

Material and Methods

Detection of Dnmt2 transcripts by Reverse Transcription PCR (RT-PCR)

RNA was obtained from *D. willistoni* adult males, adult fertilized females, adult females without ovaries, pupae, larvae and 0-3h dechorionated embryos using Trizol, according to the manufacturer's protocol (Invitrogen). Extracted RNA was further treated with DNaseI (Promega) to eliminate DNA contamination. cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher) with random primers. After cDNA synthesis, *Dnmt2* fragment was amplified by PCRs with the following primers wDnmt2AF: 5'-CAGGCTCGCCAACAGTTTAT-3' and wDnmt2BR: 5'-CCTTGTCCTTGCGTGCTAAA-3'. Primers were designed based on the *D. willistoni Dnmt2* sequence (Garcia et al. 2007). The PCR reaction was performed in 25 μ L reactions using 20 ng of cDNA, 1 U Platinum Taq DNA Polymerase (Invitrogen), 1x reaction buffer, 200 μ M dNTPs, 20 pmol of each primer and 1.5 mM MgCl₂. The amplification conditions were 95 °C for 5 min and 30 cycles of 95 °C for 40 s, 55 °C for 40 s and 72°C for 1 min, followed by a final extension cycle at 72 °C for 5 min. A fragment of 280 bp was expected as a result of the amplification reactions. The *β -actin* housekeeping gene was used as a control for amplification. PCR products were verified by electrophoresis on a 1% agarose gel and stained with GelRed GelRed™ (Sigma-Aldrich).

Expression analysis by Quantitative Real-Time PCR

The relative abundance of *D. willistoni* Dnmt2 mRNA transcripts was measured by quantitative real-time PCR (qPCR) using an Applied Biosystems (ABI) 7500 Real-Time PCR System with the same primers of conventional RT-PCR (*Dnmt2* and β -actin gene). Samples of cDNA from *D. willistoni*, produced as previously described, were used. The qPCR conditions were: 94 °C for 5 min followed by 40 cycles at 95 °C for 15 s, 60 °C for 10 s, 72 °C for 15 s and 35 s at 60 °C to measure fluorescence. Next, samples were heated from 55°C to 99 °C at a 0.1 °C/s temperature gradient to construct the denaturation curve of the amplified products. Relative quantifications of amplified products were made by the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001) and Ct values were obtained in the SDS software. SYBR-green (Molecular Probes) was used to detect amplification and to estimate Ct values, as well as to determine specificity of the amplicons by denaturing curves and melting temperatures (Tm). The β -actin gene was used as the internal control gene for all relative expression calculations. All assays were conducted with technical and biological triplicates.

Inspecting Dnmt2 mRNA transcripts by *in situ* hybridization in embryos

For *Dnmt2* riboprobe synthesis, the fragment of 949-bp of the gene was obtained from genomic DNA of adult flies by PCR using the primers wDnmt2C-F: 5' TCACCCACAACCTTGACATT 3' and wDnmt2D-R: 5' ACCTTCTCGCAGACACCAA 3'. Resulting PCR products were cloned into pCR4-TOPO vector (Invitrogen) and submitted to automatic nucleotide sequencing performed by Macrogen Inc. (Korea), to assign insert orientation. Resulting plasmids were then linearized according to the vector map and digoxigenin-labeled using a DIG-labeled dNTP mix (Roche). The antisense riboprobe was labeled by T3 RNA Polymerase *in vitro* transcription of the *NotI*-linearized plasmid. *In situ* hybridization assays were performed in embryos collected in different developmental stages (Deprá et al. 2009). Prior to *in situ* hybridization, probes (200 ng in 50µL) were mixed with *in situ* hybridization solution as follows: 50% formamide, 5x SSC, 100µg/mL of herring sperm DNA, 50µg/mL of heparin and 1% of Tween 20. The hybridization mixture was added to the embryos at 55 °C and incubated overnight. After hybridization embryos were washed several times with PBS and color development was performed with BCIP/NBT 75 (Promega). Embryonic

stages were identified according to the outlined criteria (Campos-Ortega and Hartenstein 1985).

Physical mapping of Dnmt2 gene

Non-fluorescent *in situ* hybridization technique for the precise location of *Dnmt2* sequences in polytene chromosomes of *D. willistoni* was used since it allows better visualization of the banding patterns and is recommended for precise physical mapping (Clark et al. 2007). DNA probe of *Dnmt2* of *D. willistoni* was constructed from a fragment of 949 bp cloned into the vector TOPO PCR-4 (Invitrogen). The probe was obtained by PCR, from the sequenced Gd-H4-1 line, as well as the photos of the polytene chromosomes. The probe was labeled with a Biotin PCR Labeling Core Kit (Jena Bioscience) following the manufacturer instructions, with minor adjustments. The labeled products were purified with a NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-NAGEL). Vectastain Elite ABC (Vector Laboratories) and DAB Substrate (Roche) performed the detection and revelation of hybridization signals, respectively. Slides were stained with Giemsa solution (5%) and coverslips were assembled with Entellan (Merck). Polytene chromosomes images were acquired in phase contrast photomicroscope and analyzed according to the photomap of *D. willistoni* (Schaeffer et al. 2008).

Results

Transcriptional expression of *D. willistoni Dnmt2* is developmental delimited

We performed RT-PCR analysis to verify the levels of *Dnmt2* transcriptional expression. The expression of *Dnmt2* transcripts was detected for all cDNA samples used: *D. willistoni* adult males, females, females without ovaries, pupae, larvae, and 0-3h dechorionated embryos. The amplification of a single fragment with an expected size of 280 bp, corresponding to processed mRNA, confirmed the absence of genomic DNA contaminants in all samples (**Figure 1**). In order to infer the amount of *Dnmt2* transcripts at different stages of development and samples analyzed, further investigations were carried out by qPCR.

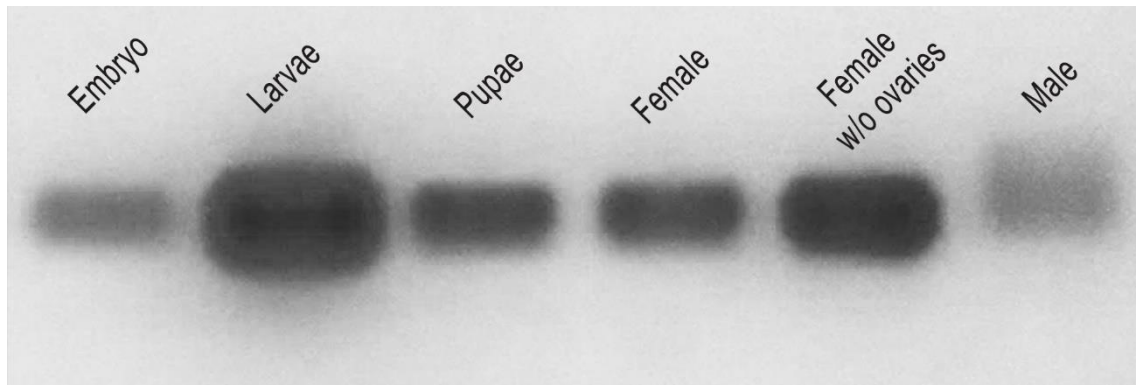


Figure 1. Detection of *Dnmt2* by conventional RT-PCR. All the RT-PCR products had the expected size (280 bp) on 1% agarose gel and stained with GelRed™ (Sigma-Aldrich).

The qPCR showed that *Dnmt2* expression has different levels along with *D. willistoni* development (**Figure 2**). The analysis showed a pattern of *Dnmt2* expression where embryos rank higher than the other developmental stages (approximately 2 to 3-fold). The *Dnmt2* transcription decreases along with the development, although the adult stages showed a slight increase. Females, females without ovaries and males showed no substantial differences between the levels of *Dnmt2* expression.

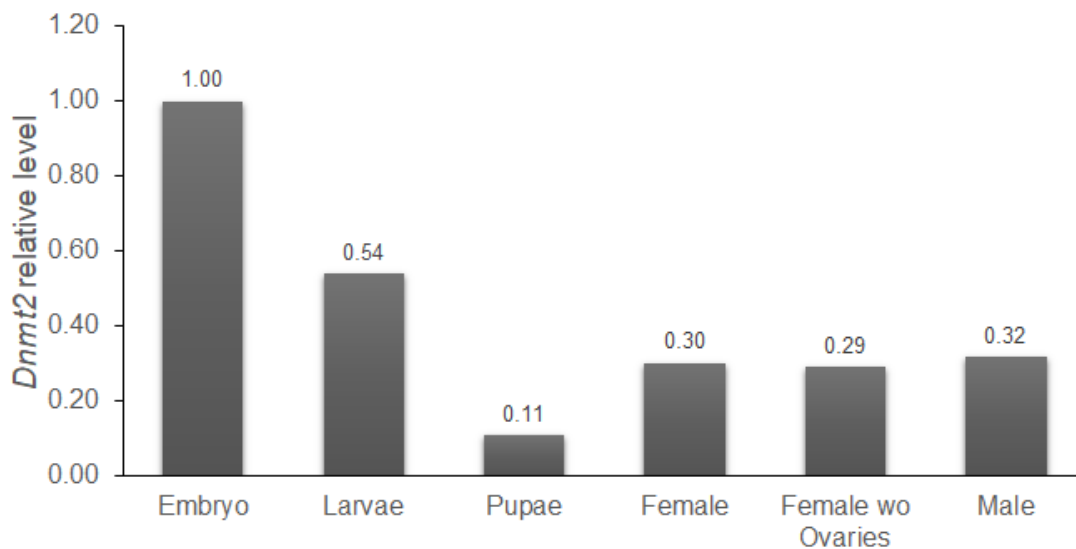


Figure 2. *Dnmt2* transcriptional expression levels in the *D. willistoni* samples. The results were normalized with β -actin expression. All data were obtained from triplicates.

***Dnmt2* transcripts follow the dynamics of embryonic development**

In situ hybridization experiments were conducted at different developmental oogenesis and embryogenesis stages to determine the spatial and temporal pattern of the *Dnmt2* transcriptional expression of the *D. willistoni*. Initially, the *Dnmt2* transcripts were detected as uniform and widespread in all stages of oogenesis (**Figure 3**). The staining pattern indicates early expression at the anterior end and in the ventral position of the oocyte (**Figure 3A-3H**).

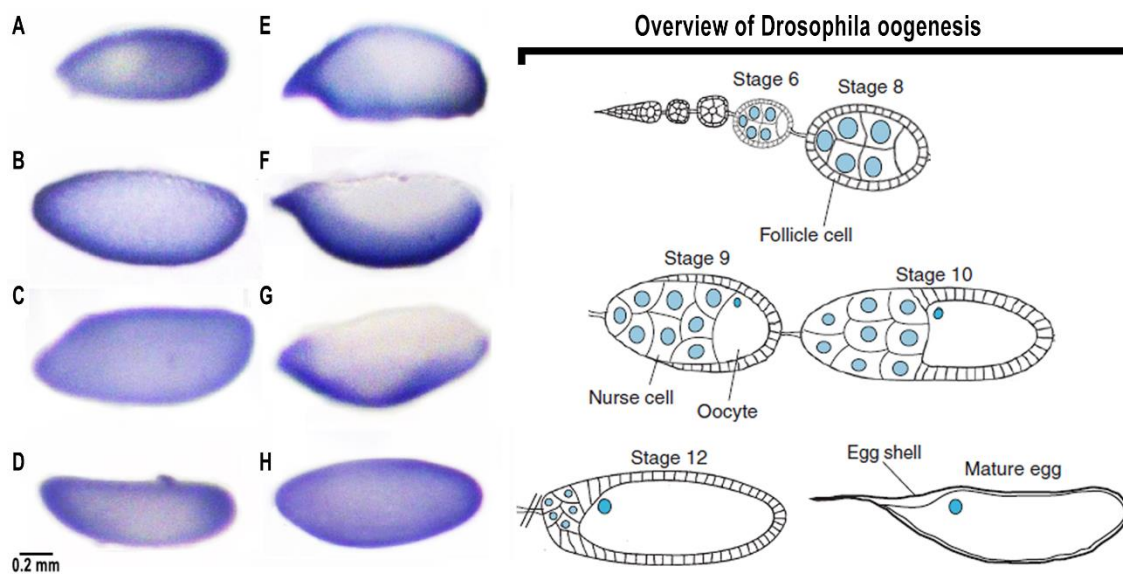


Figure 3. Spatial expression patterns of *Dnmt2* transcripts on *D. willistoni* during oogenesis, showing a widespread pattern during differentiation, which gradually reaches an evident pattern on its ventral surface; to the right, we have an overview of *Drosophila* oogenesis. The orientation of oocytes: anterior is to the left. Bar: 0.2mm.

In the initial embryonic developmental stages, *Dnmt2* transcripts appear as superficial and later as diffuse internal granules, a pattern similar to those found in maternal transcripts (**Figure 4A-4C**). This pattern persists until the cellular blastoderm stage. Throughout development, *Dnmt2* transcripts accumulate at the embryonic ventral periphery (**Figure 4D**), and subsequent expression can be detected on the anterior portion, coinciding with the moment when the cephalic furrow is formed (**Figure 4E**). The expression of *Dnmt2* was detected from the cephalic region expanding through the ventral region to the embryo's posterior pole (**Figure 4F-4H**). The anterior and posterior expression levels of staining decreased and became a basal signal from then on (**Figure**

4I). During late embryogenesis, at the germ-band retraction stage, transcript accumulation persisted as a basal expression in the ectoderm layer.

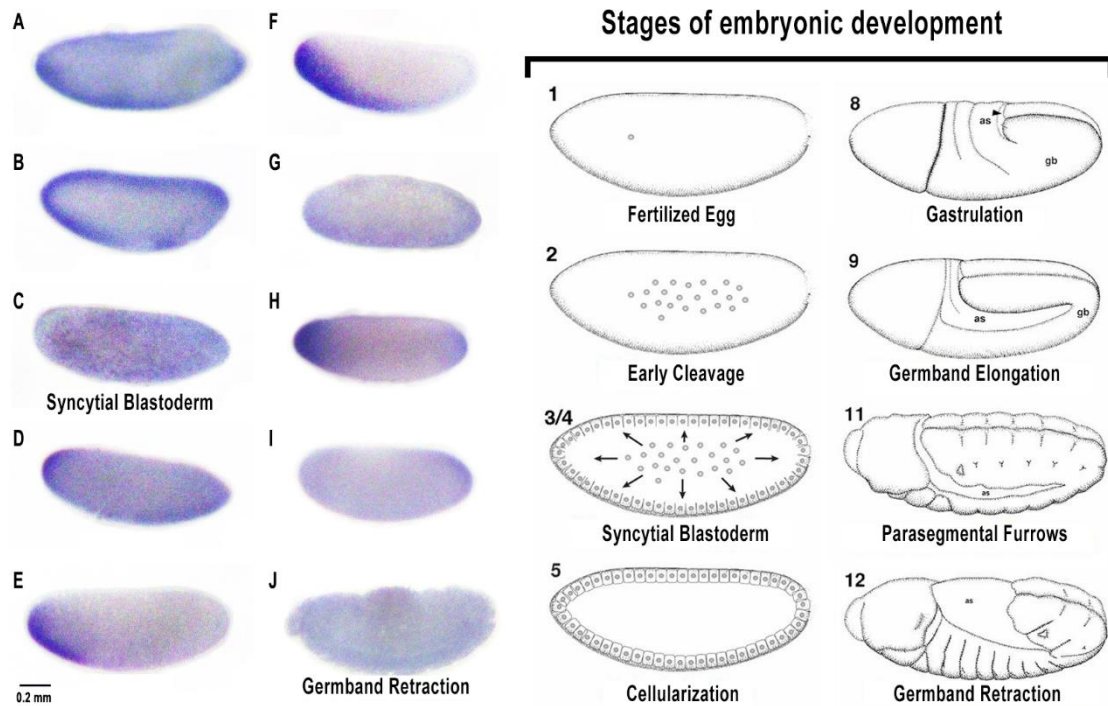


Figure 4. Spatial expression patterns of *Dnmt2* transcripts on *D. willistoni* during embryogenesis showing. **A-D**: first embryogenesis stages showing global staining patterns. **E-I**: signal expansion from anterior to posterior embryo pole; **J**: late embryogenesis showing germ-band retraction and weak staining. Images adapted from the Atlas of *Drosophila* Development (Campos-Ortega and Hartenstein 1985). The orientation of embryos: anterior is to the left. Bar: 0.2 mm.

The *Dnmt2* gene resides in the arm IIL of the *D. willistoni* chromosome

The location of the probe hybridization signal was determined to be at the distal region of the chromosome II left arm (IIL), specifically in section 55C (**Figure 5A and 5B**). The IIL arm of *D. willistoni* corresponds to the Muller B element, according to a previous study (Garcia et al. 2015). On FlyBase search tool (Wilson et al. 2008), we could verify that in the other eleven species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. erecta*, *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. virilis*, *D. mojavensis*, and *D. grimshawi*) whose genomes are available, the sequences homologs to *Dnmt2* are also localized at the B element.

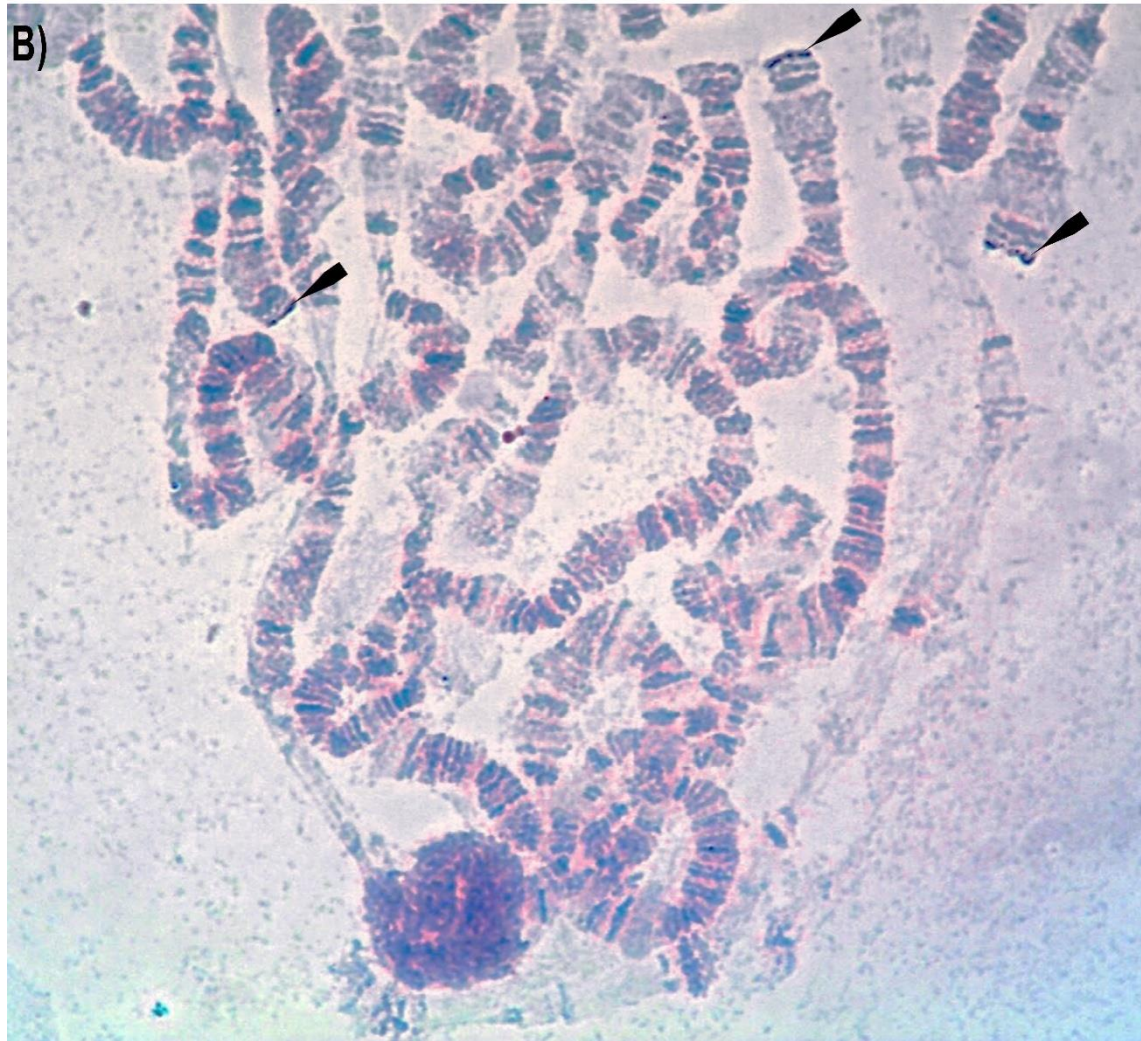


Figure 5. Physical mapping of *Dnmt2* gene in polytene chromosomes of *D. willistoni*. **A)** IIL arm with the site of gene *Dnmt2* hybridizations. **B)** *In situ* hybridization signal of *Dnmt2* gene in chromosomal arm IIL in Gd-H4-1 strain of *D. willistoni* (black arrows).

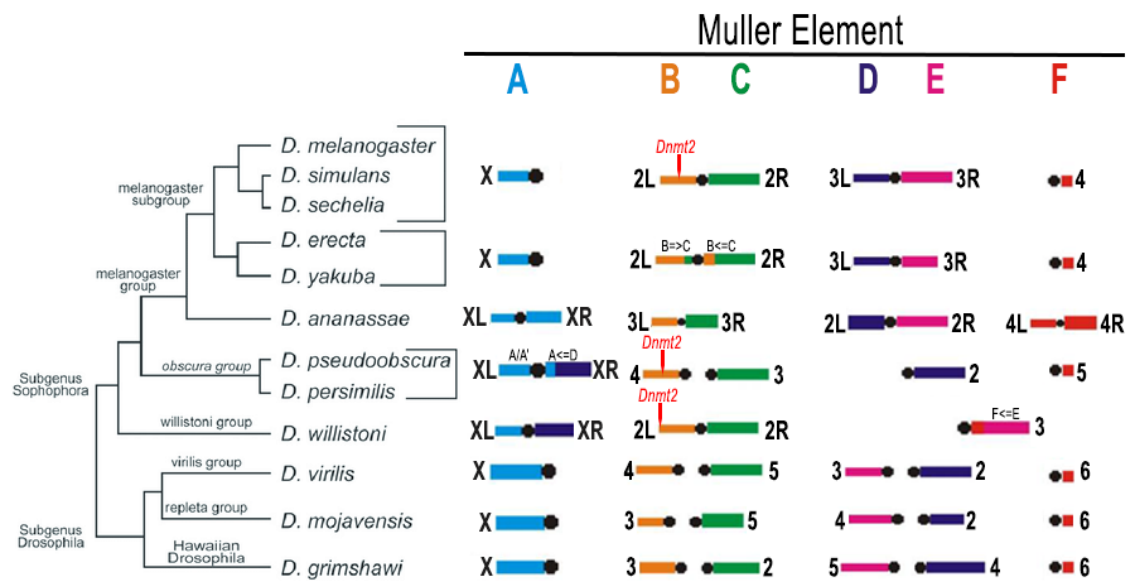


Figure 6. Schematic correspondence of rearrangement for the six Muller elements among twelve species of *Drosophila*. The centromeres are indicated with solid black circles. The approximate localization of the *Dnmt2* in the arms is indicated by a red arrow. Modified from Schaeffer *et al.* (2008) with corrections by Garcia *et al.* (2015).

Discussion

Drosophila are so-called “*Dnmt2*-only” organisms, that is, they do not have any of the canonical DNA methyltransferases (*Dnmt1* and *Dnmt3*). DNA methylation was also observed in other “*Dnmt2*-only” organisms, especially in other dipterans, though direct evidence of *Dnmt2* involvement is still enigmatic. Interestingly, there is a diverse group of animal species that have retained *Dnmt2* as their only DNA methyltransferase candidate, like *Schizosaccharomyces pombe*, *Dictyostelium discoideum*, *Entamoeba histolytica*, *Schistosoma mansoni*, and *D. melanogaster*. For instance, about 1 in 600 (0.17%) cytosine nucleotides are methylated in the DNA of the *Aedes albopictus* mosquito (Adams *et al.* 1979). However, although such species have genomic methylation, this is still at low global levels and *Dnmt2* biological function has been fervently discussed (Jeltsch *et al.* 2006; Krauss and Reuter 2011).

There are accumulating findings that instigate investigations in “*Dnmt2*-only” organisms, such as female-specific DNA methylation that was reported for *D. willistoni* and related species (Garcia *et al.* 2007; D’Ávila *et al.* 2010). Recently, males that have reduced expression of dMBD-R2 (a methylated cytosine binding protein), specifically

octopamine (equivalent to norepinephrine) neurons, have exhibited mating behavior toward divergent interspecies such as *D. virilis* and *D. yakuba* and a decrease in the success of co-specific mating (Gupta et al. 2017). These data strongly suggest the existence of methylation-dependent chromatin structures in *Drosophila*, that is, the role of methylation of genomic cytosines appears to be fundamental at various levels in these organisms. Thus, *Dnmt2* studies are of great importance towards a better understanding of DNA and RNA methylation in “*Dnmt2*-only” systems, whose data are still vastly controversial and uncertain.

Our quantitative analyses on different development stages showed differences in *Dnmt2* expression (**Figure 2**), with the highest expression being observed in embryos, in the *D. willistoni*, in which *Dnmt2* expression was 2-fold higher than in the larvae stage. Essentially, on *Drosophila* genus, DNA methylation is described as predominant during embryonic development. In *D. melanogaster*, the DNA methylation is prevalent in young (1–2 h) embryos, but less marked in older (15–16h) embryos and only trace amounts of 5-methylcytosine were found in isolated ovaries (oocytes), and *Dnmt2* transcripts present similar levels corresponding to DNA methylation stages (Lyko 2001). The data herein showed a very close pattern of *Dnmt2* expression in *D. willistoni*, similar to previously founded in *D. melanogaster*. Interestingly, the *Dnmt2* expression in females and males of *D. willistoni* revealed a slightly higher value when compared with the pupae stage, and this relative expression levels in the *D. willistoni* adults are higher than what is found in *D. melanogaster* adults (**Figure 2**) (Lyko 2001). Thus, this *Dnmt2* expression can be relevant, since recently we had reported evidence of DNA methylation in adult flies of *D. willistoni* and closely relates species (Garcia et al. 2007; D'Ávila et al. 2010).

Employing whole *D. willistoni* embryos hybridization with *Dnmt2* riboprobe, we could observe transcriptional expression along different embryonic stages. The oocytes showed the most prominent hybridization signals, suggesting a higher activity of the enzyme on this stage of oogenesis (**Figure 3**). The early stages of embryogenesis (**Figure 4**) showed hybridization staining with distinct patterns, in which the expression pattern seems to decrease from syncytial blastoderm to late development. Interestingly, the expression patterns ranged from a granular composition throughout the whole embryo on syncytial blastoderm to a peripheral occurrence on cellular blastoderm. During gastrulation, we observed that the *Dnmt2* expression seems to suffer compartmentalization from the anterior portion in the ventral region, reaching the

posterior portion. Moreover, it showed only a fading basal signal in the head region and in the position of the developing central nervous system (CNS). All the presented results with embryos of *D. willistoni* are in agreement with those previously found concerning the genomic DNA methylation that predominates during early embryonic development and decreases at later stages (Lyko 2001), presumably as a result of reduced methyltransferase expression.

The importance of DNA methylation in *Drosophila* still is an elusive case, and there are some attempts to understand the biological function in *Drosophila* genomes. However, when we are talking about a genus with more than 2.800 species, peculiarities may arise, like the sex-specific methylation described in *D. willistoni* (Garcia et al. 2007; D'Ávila et al. 2010), so generalization must be avoided. It was observed that overexpression of *D. melanogaster* *Dnmt2* results in an extended fly life span and in overexpression of several genes (Lin et al. 2005). Also, it is discussed whether *Dnmt2* acts as DNA or tRNA methyltransferase. If *Dnmt2* was described associated to a nuclear matrix, and if *Drosophila* is a “*Dnmt2*-only” organism, and if *D. willistoni* has sex-specific methylation in its genome, we cannot deny that the evidence suggests a duality in the activity of that enzyme in these organisms. The peculiarities of *Dnmt2* enzyme rise in different organisms; in humans, for example, it is located in cytoplasmic regions (Goll et al. 2006a), whereas in *Dyctiostelium discoideum* and *Entamoeba histolytica*, *Dnmt2* is located only in the nuclear matrix (Kuhlmann et al. 2005; Schaefer et al. 2008), and guaranteeing the structural integrity of chromatin and silencing retrotransposons in *Drosophila* (Phalke et al. 2009). Generating *Dnmt2* null mutations allowed to correlate loss of DNA methylation, identified in early *D. melanogaster* embryos by bisulfite sequencing and restriction enzyme analysis. Besides that, these results suggest the *Dnmt2*-dependent DNA methylation during early embryonic development of *D. melanogaster* has a key function in control of retrotransposon silencing in somatic cells, specifically as an initial step of the process (Phalke et al. 2009). This feature contrasts with the high local specificity of the *Dnmt1* and *Dnmt3* to the cell nucleus region.

On the other hand, the HIV virus is able to increase its survival in the host cell using the RNA methylation activity of *Dnmt2*, increasing the stability of its own genome. Through infection, *Dnmt2* relocates from the nucleus to the cytoplasmic stress granules and methylates the RNA virus, promoting the post-transcriptional stability of viral RNA (Dev et al. 2017). Interference RNAs (RNAi) have a strong association with different

epigenetic mechanisms, such as histone methylation/acetylation and genomic DNA methylation (Volpe et al. 2002). Double-stranded RNAs (dsRNA) trigger the process of post-transcriptional homology-dependent gene interference (RNAi) closely related to the co-suppression of viral exogenous transcripts. The dsRNAs are processed by a Dicer family of enzymes into small dsRNA sequences having 21-25 nucleotides, termed small interfering RNAs (or siRNAs). Interestingly, it is known that Dnmt2 also interacts with Dicer-2 in response to heat shock treatments (Durdevic et al. 2013b; Durdevic and Schaefer 2013), controlling the expression and regulation of various Heat Shock proteins (Fisher et al. 2006; Schaefer et al. 2010; Thiagarajan et al. 2011). Finally, the activity of Dnmt2 in conjunction with NSun2 is important to promote the stability of tRNAs and protein synthesis (Tuorto et al. 2012). That is, we can assume that the Dnmt2 has a multilevel biological role, inserted in an extensive network of interactions in the genetic machinery of the gene expression control.

In the present study, we also performed the *in situ* localization in the chromosomes of the *D. willistoni* *Dnmt2* (**Figure 5**). The gene is located in the arm IIL (Muller B element), like in *D. melanogaster* and *D. pseudoobscura*, where it was possible to ascertain the position of the gene, it is present in B element. This is a very important aspect because we present here the determination of another gene marker in *D. willistoni*, following previous work (Garcia et al. 2015), which will help in the organization of the *D. willistoni* scaffolds regarding the physical position of the genes in the chromosomes.

Surprisingly, *D. willistoni* *Dnmt2* is found in the subtelomeric region of the chromosome IIL (Muller B element), whereas *D. melanogaster* and *D. pseudoobscura* *Dnmt2* are found, approximately, in the central region of the chromosome IIL (Muller B element) (**Figure 6**). Most likely, the differences found for the Muller element and the location of the *Dnmt2* in the chromosome are related to translocation events followed by breakage and inversion of the region where the gene is found.

It is known that heterochromatin regions (centromeric and telomeric) have different patterns of evolution than those found in the euchromatin regions (gene-rich). Comparative analyses between telomeric regions of humans and chimpanzees have shown that both are very distinct, in that the human chromosomes present a large scale of rearrangements and differences in the repetitive elements present (Trask et al. 1998). Interestingly, the chromosome ends in *Drosophila melanogaster* have the peculiarity of

presenting transposable elements (*HeT-A* and *TART*) (Rubin 1978; Young et al. 1983; Renkawitz-Pohl and Bialojan 1984; Traverse and Pardue 1988), contrasting with the human chromosome ends that are structured from the activity of the enzyme telomerase. Another interesting aspect is that most-distal regions of *Drosophila* subtelomeres seem to evolve rapidly between *Drosophila melanogaster* and its close relatives, *D. simulans* and *D. yakuba* (Anderson et al. 2009). In this region, the mutation rate of the most-distal portion is so high that the structure is likely to be different among cells of the same individual (Roberts 1979). However, what is observed in terms of recombination in the telomeric regions of *Drosophila melanogaster* is that there is a great suppression of crossing over events, within and in regions proximal to the telomeres (Lindsley et al. 1977). These aspects are in agreement with previous findings of our research group, in which *D. willistoni Dnmt2* shows the highest rate of nucleotide substitution per codon in relation to the other drosophilids and different selective pressures at specific sites involved in an interaction-driven co-evolution with other genes (Vieira et al. 2018).

Conclusion

Epigenetic mechanisms are fundamental in the coordination of the ontogeny of organisms and in the intermediation of information from the environment to the nucleus and from the nucleus to the environment. The understanding of such mechanisms is of paramount importance in order to understand the ecological, physiological and evolutionary aspects of the different species. Among the epigenetic phenomena, cytosine methylation is the most widely studied and known. However, its role is still elusive in organisms known as "*Dnmt2*-only", among them *Drosophila*.

In the present work, we verified that the expression of *Dnmt2* in *D. willistoni* is closely related to its ontogenetic stages. In addition, it has been found that expression levels in adults of *D. willistoni Dnmt2* are apparently slightly higher than in *D. melanogaster*. It is suggested, therefore, that this difference may be closely related to the phenomenon of sex-specific methylation, a peculiar and restricted phenomenon to the species of the subgroup *willistoni*. We also conduct the *in situ* localization of the *Dnmt2* gene, revealing its presence in the arm IIL (Muller B element). The *Dnmt2* is located in the subtelomeric region, and this brings up a series of issues that involve the peculiar characteristics of *D. willistoni Dnmt2* enzyme, like evolutionary pattern in nucleotide

substitutions rates (Vieira et al. 2018), the high prevalence of basic amino acids residues in its target recognition domain surface (Vieira et al. 2017) and the epigenetic phenomena of sex-specific methylation (Garcia et al. 2007; D'Ávila et al. 2010).

With that, fortunately, further questions are being raised about such peculiarities in this unique group of Neotropical species which are *D. willistoni*. Future studies are needed for a more in-depth understanding of the importance of sex-specific methylation in development and survival, as well as the impact related to the subtelomeric position occupied by the *Dnmt2*.

References

Adams RLP, McKay EL, Craig LM and Burdon RH (1979) Methylation of mosquito DNA. *Biochim Biophys Acta - Nucleic Acids Protein Synth* 563:72–81. doi: 10.1016/0005-2787(79)90008-X

Anderson JA, Gilliland WD and Langley CH (2009) Molecular population genetics and evolution of *Drosophila* meiosis genes. *Genetics* 181:177–185. doi: 10.1534/genetics.108.093807

Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16:6–21. doi: 10.1101/gad.947102

Breiling A and Lyko F (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin* 8:1–9. doi: 10.1186/s13072-015-0016-6

Campos-Ortega JA and Hartenstein V (1985) The Embryonic Development of *Drosophila melanogaster*. doi: 10.1007/978-3-662-02454-6

Cheng JX, Chen L, Li Y, Cloe A, Yue M, Wei J, Watanabe KA, Shammo JM, Anastasi J, Shen QJ et al. (2018) RNA cytosine methylation and methyltransferases mediate

chromatin organization and 5-azacytidine response and resistance in leukaemia. *Nat Commun* 9:1–16. doi: 10.1038/s41467-018-03513-4

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218. doi: 10.1038/nature06341

D'Ávila MF, Garcia RN, Panzera Y and Valente VLS (2010) Sex-specific methylation in *Drosophila*: An investigation of the *Sophophora* subgenus. *Genetica* 138:907–913. doi: 10.1007/s10709-010-9473-9

da Cunha B and Dobzhansky T (1954) A Further Study Of Chromosomal Polymorphism In *Drosophila willistoni* In Its Relation To The Environment. *Evolution* (N Y) 8:119–134.

da Cunha AB and Dobzhansky T (1950) Adaptive Chromosomal Polymorphism In *Drosophila willistoni*. 212–235.

Deprá M, Valente VLS, Margis R and Loreto ELS (2009) The *hobo* transposon and *hobo*-related elements are expressed as developmental genes in *Drosophila*. *Gene* 448:57–63. doi: 10.1016/J.GENE.2009.08.012

Dev RR, Ganji R, Singh SP, Mahalingam S, Banerjee S and Khosla S (2017) Cytosine methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during infection. *Biochem J* 474:2009–2026. doi: 10.1042/BCJ20170258

Durdevic Z, Mobin MB, Hanna K, Lyko F and Schaefer M (2013) The RNA methyltransferase *dnmt2* is required for efficient *dicer-2*-dependent siRNA pathway activity in *Drosophila*. *Cell Rep* 4:931–937. doi: 10.1016/j.celrep.2013.07.046

Durdevic Z and Schaefer M (2013) TRNA modifications: Necessary for correct tRNA-

derived fragments during the recovery from stress? *BioEssays* 35:323–327. doi: 10.1002/bies.201200158

Elango N, Hunt BG, Goodisman MAD and Yi S V. (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *PNAS* 106:11206–11211.

Fisher O, Siman-Tov R and Ankri S (2006) Pleiotropic phenotype in *Entamoeba histolytica* overexpressing DNA methyltransferase (Ehmeth). *Mol Biochem Parasitol* 147:48–54. doi: 10.1016/j.molbiopara.2006.01.007

Garcia CF, Delprat A, Ruiz A and Valente VLS (2015) Reassignment of *Drosophila willistoni* genome scaffolds to chromosome II arms. *G3 Gene Genomes Genet* 5:2559–2566. doi: 10.1534/g3.115.021311

Garcia RN, D'Ávila MF, Robe LJ, Loreto ELDS, Panzera Y, De Heredia FO and Valente VLS (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica* 131:91–105. doi: 10.1007/s10709-006-9116-3

Goll MG and Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-LL, Zhang X, Golic KG, Jacobsen SE and Bestor TH (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* (80-) 311:395–398. doi: 10.1126/science.1120976

Gowher H, Leismann O and Jeltsch A (2000) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J* 19:6918–6923. doi: 10.1093/emboj/19.24.6918

Gupta T, Morgan HR, Andrews JC, Brewer ER and Certel SJ (2017) Methyl-CpG binding domain proteins inhibit interspecies courtship and promote aggression in *Drosophila*. *Sci*

Rep 7:1–12. doi: 10.1038/s41598-017-05844-6

Jeltsch A, Nellen W and Lyko F (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 31:306–308. doi: 10.1016/j.tibs.2006.04.005

Jin B, Li Y and Robertson KD (2011) DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes and Cancer* 2:607–617. doi: 10.1177/1947601910393957

Klose RJ and Bird AP (2006) Genomic DNA methylation: The mark and its mediators. *Trends Biochem Sci* 31:89–97. doi: 10.1016/j.tibs.2005.12.008

Krauss V and Reuter G (2011) DNA Methylation in *Drosophila*-a critical evaluation. *Prog Mol Biol Transl Sci* 101:177–191. doi: 10.1016/B978-0-12-387685-0.00003-2

Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, Söderbom F et al. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res* 33:6405–6417. doi: 10.1093/nar/gki952

Kunert N, Marhold J, Stanke J, Stach D and Lyko F (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* 130:5083–5090. doi: 10.1242/dev.00716

Lee T -f., Zhai J and Meyers BC (2010) Conservation and divergence in eukaryotic DNA methylation. *Proc Natl Acad Sci* 107:9027–9028. doi: 10.1073/pnas.1005440107

Li E and Zhang Y (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6:a019133. doi: 10.1101/cshperspect.a019133

Li S and Mason CE (2014) The Pivotal Regulatory Landscape of RNA Modifications. *Annu Rev Genomics Hum Genet* 15:127–150. doi: 10.1146/annurev-genom-090413-025405

- Lin M-JJ, Tang L-YY, Reddy MN and Shen C-KKJJ (2005) DNA methyltransferase gene *dDnmt2* and longevity of *Drosophila*. *J Biol Chem* 280:861–864. doi: 10.1074/jbc.C400477200
- Lindsley DL, Sandler L, Counce SJ, Chandley AC and Lewis KR (1977) The Genetic Analysis of Meiosis in Female *Drosophila melanogaster*. *Philos Trans R Soc Lond B Biol Sci* 277:295–312. doi: 10.2307/2417716
- Livak KJ and Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25:402–408. doi: 10.1006/METH.2001.1262
- Lyko F (2001) DNA methylation learns to fly. *Trends Genet* 17:169–172. doi: 10.1016/S0168-9525(01)02234-X
- Lyko F, Whittaker AJ, Orr-Weaver TL and Jaenisch R (2000) The putative *Drosophila* methyltransferase gene *dDnmt2* is contained in a transposon-like element and is expressed specifically in ovaries. *Mech Dev* 95:215–217. doi: 10.1016/S0925-4773(00)00325-7
- Mund C, Musch T, Strödicke M, Assmann B, Li E and Lyko F (2004) Comparative analysis of DNA methylation patterns in transgenic *Drosophila* overexpressing mouse DNA methyltransferases. *Biochem J* 378:763–8. doi: 10.1042/BJ20031567
- Panikar CS, Rajpathak SN, Abhyankar V, Deshmukh S and Deobagkar DD (2015) Presence of DNA methyltransferase activity and CpC methylation in *Drosophila melanogaster*. *Mol Biol Rep* 42:1615–1621. doi: 10.1007/s11033-015-3931-5
- Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC and Reuter G (2009) Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends

on the cytosine-5 methyltransferase DNMT2. *Nat Genet* 41:696–702. doi: 10.1038/ng.360

Renkawitz-Pohl R and Bialojan S (1984) A DNA sequence of *Drosophila melanogaster* with a differential telomeric distribution. *Chromosoma* 89:206–11.

Roberts PA (1979) Rapid change of chromomeric and pairing patterns of polytene chromosome tips in *D. melanogaster*: Migration of polytene-non-polytene transition zone? *Genetics* 92:

Rohde C and Valente VLS (2012) Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. *Genet Mol Biol* 35:966–979. doi: 10.1590/S1415-47572012000600012

Rubin GM (1978) Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. *Cold Spring Harb Symp Quant Biol* 42 Pt 2:1041–6.

Schaefer M and Lyko F (2007) DNA methylation with a sting: An active DNA methylation system in the honeybee. *BioEssays* 29:208–211. doi: 10.1002/bies.20548

Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M and Lyko F (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* 24:1590–1595. doi: 10.1101/gad.586710

Schaefer M, Steringer JP and Lyko F (2008) The *Drosophila* cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS One*. doi: 10.1371/journal.pone.0001414

Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, Rohde C, Valente VLS, Aguadé M, Anderson WW et al. (2008) Polytene chromosomal maps of 11 *Drosophila* species: The order of genomic scaffolds inferred from genetic and physical

maps. *Genetics* 179:1601–1655. doi: 10.1534/genetics.107.086074

Suzuki MM and Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465–476. doi: 10.1038/nrg2341

Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L, Martin DIK and Boffelli D (2014) Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res* 24:821–830. doi: 10.1101/gr.162412.113

Thiagarajan D, Dev RR and Khosla S (2011) The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics* 6:103–113. doi: 10.4161/epi.6.1.13418

Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, Collins C, Giorgi D, Iadonato S, Johnson F et al. (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* 7:13–26.

Traverse KL and Pardue ML (1988) A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc Natl Acad Sci U S A* 85:8116–20.

Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G and Lyko F (2012) RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol* 19:900–905. doi: 10.1038/nsmb.2357

Valente VLS and Araújo AM (1986) Chromosomal polymorphism, climatic factors, and variation in population size of *Drosophila willistoni* in southern Brazil. *Heredity (Edinb)* 57:149–159. doi: 10.1038/hdy.1986.105

Vicario S, Moriyama EN and Powell JR (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol Biol* 7:1–17. doi: 10.1186/1471-2148-7-226

Vieira GC, D'Ávila MF, Zanini R, Deprá M and Valente VLS (2018) Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions. *Genet Mol Biol*. doi: 10.1590/1678-4685-gmb-2017-0056

Vieira GC, Sinigaglia M, Vieira GF and Valente VLS (2017) Linking epigenetic function to electrostatics: The DNMT2 structural model example. *PLoS One* 12:1–19. doi: 10.1371/journal.pone.0178643

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS and Martienssen RA (2002) Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* (80-) 297:1833–1837. doi: 10.1038/098448b0

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA and Robinson GE (2006) Functional CpG methylation system in a social insect. *Science* 314:645–7. doi: 10.1126/science.1135213

Wilson RJ, Goodman JI and Strelts VB (2008) FlyBase: Integration and improvements to query tools. *Nucleic Acids Res* 36:588–593. doi: 10.1093/nar/gkm930

Young BS, Pession A, Traverse KL, French C and Pardue ML (1983) Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell* 34:85–94.

Zhao BS, Roundtree IA and He C (2016) Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* 18:31–42. doi: 10.1038/nrm.2016.132

CAPÍTULO IV

Linking epigenetic function to electrostatics: The DNMT2 structural model example

Gilberto Cavalheiro Vieira^{1,3}, Gustavo Fioravanti Vieira^{1,4}, Marialva Sinigaglia^{1,3,5},
Vera Lúcia da Silva Valente^{1,2,3}

¹ Programa de Pós Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

² Programa de Pós Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

³ Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

⁴ Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

⁵ Instituto do Câncer Infantil, Porto Alegre, RS, Brazil.

RESEARCH ARTICLE

Linking epigenetic function to electrostatics: The DNMT2 structural model example

Gilberto Cavalheiro Vieira^{1,2*}, Gustavo Fioravanti Vieira^{1,3}, Marialva Sinigaglia^{1,2,4}, Vera Lúcia da Silva Valente^{1,2,5}

1 Programa de Pós Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil, **2** Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil, **3** Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, **4** Instituto do Câncer Infantil, Porto Alegre, RS, Brazil, **5** Programa de Pós Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

* cavalheiro.rs@gmail.com



OPEN ACCESS

Citation: Vieira GC, Vieira GF, Sinigaglia M, Silva Valente VLd (2017) Linking epigenetic function to electrostatics: The DNMT2 structural model example. PLoS ONE 12(6): e0178643. <https://doi.org/10.1371/journal.pone.0178643>

Editor: Patrizio Dimitri, Università degli Studi di Roma La Sapienza, ITALY

Received: January 16, 2017

Accepted: May 16, 2017

Published: June 2, 2017

Copyright: © 2017 Vieira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS no 10/0028-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD-CAPES).

Competing interests: The authors have declared that no competing interests exist.

Abstract

The amino acid sequence of DNMT2 is very similar to the catalytic domains of bacterial and eukaryotic proteins. However, there is great variability in the region of recognition of the target sequence. While bacterial DNMT2 acts as a DNA methyltransferase, previous studies have indicated low DNA methylation activity in eukaryotic DNMT2, with preference by tRNA methylation. Drosophilids are known as DNMT2-only species and the DNA methylation phenomenon is a not elucidated case yet, as well as the ontogenetic and physiologic importance of DNMT2 for this species group. In addition, more recently study showed that methylation in the genome in *Drosophila melanogaster* is independent in relation to DNMT2. Despite these findings, Drosophilidae family has more than 4,200 species with great ecological diversity and historical evolution, thus we, therefore, aimed to examine the drosophilids DNMT2 in order to verify its conservation at the physicochemical and structural levels in a functional context. We examined the twenty-six DNMT2 models generated by molecular modelling and five crystallographic structures deposited in the Protein Data Bank (PDB) using different approaches. Our results showed that despite sequence and structural similarity between species close related, we found outstanding differences when they are analyzed in the context of surface distribution of electrostatic properties. The differences found in the electrostatic potentials may be linked with different affinities and processivity of DNMT2 for its different substrates (DNA, RNA or tRNA) and even for interactions with other proteins involved in the epigenetic mechanisms.

Introduction

Cytosine methylation plays a critical role in the regulation of gene expression in higher eukaryotes. Cytosine methylation is performed by DNA methyltransferases (DNMTs), which are classified into three subfamilies: DNMT1, DNMT2 and DNMT3a and 3b. DNMT2 is the smallest eukaryotic methyltransferase (approximately 400 amino acids) and the most widely

distributed in animals, fungi, protists, and plants [1]. In agreement with structural conservation, different methods in various systems have shown that DNMT2 has DNA methyltransferase activity [2–5]. DNMT2 was first identified in mice and humans and is likely conserved among eukaryotes [6,7]. This enzyme is the only DNMT found in dipterans, including *Drosophila* [8].

Nevertheless, according to analyses of human and *Entamoeba* enzymes, its catalytic activity on DNA is very weak, suggesting that in some species have an alternative role [2,4]. For example, studies have associated DNMT2 with RNA interference in *Dictyostelium* and covalent histone modification in *Drosophila* [9,10]. Additionally, several studies have shown the occurrence of DNA methylation in *Drosophila* species [11–13]. Recent results reported showed methylation in the genome in *Drosophila melanogaster*, even in deficient-DNMT2 strains, indicating a probable independence of this epigenetic phenomenon in relation to DNMT2 [14]. However, the family Drosophilidae is among the most diverse of the Diptera, including more than 4,200 species [15], each with its diversified ecological, historical and evolutionary context. Species of this family, especially of the genus *Drosophila*, are widely used in many areas of contemporary biological research, and few of them have been investigated with respect to the occurrence of DNA methylation.

The distribution of DNMT2 enzymes in the Diptera reveals that DNMT2 protein sequences are highly conserved in *Drosophila virilis*, *D. hydei*, *D. simulans*, *D. melanogaster* and *D. pseudoobscura* [13]. This conservation occurs primarily within the catalytic DNA methyltransferase motifs [13]. Focusing on the *Drosophila* genus, the comparison between the DNMT2 protein sequences of *D. willistoni* and *D. melanogaster* revealed higher conservation at the domains putatively responsible for methyl transfer catalysis and great variability in the region containing the specific DNA target recognition domain (TRD) [16].

Another interesting aspect is the presence of differential cleavage patterns between males and females in the *Drosophila willistoni*, presenting a DNA methylation phenomenon that does not occur in *D. melanogaster* [16]. Further investigations with more species of the same *Sophophora* subgenus of *Drosophila* with respect to putative differences between sexes demonstrated that this phenomenon only occurs in sibling species of the *willistoni* subgroup [17]. The studies also established phylogenetic correlations in the sex-specific methylation patterns in the subgroup *willistoni* species, where *D. willistoni*, *D. tropicalis* and *D. insularis* (closer species) share methylation patterns in ribosomal genes, and the *D. equinoxialis* and *D. paulistorum* patterns are not restricted to rDNA [16,17]. Accordingly, these results suggest that selection for different targets of methylation may even occur between closely related species.

The diversity of action of DNMT2 can be verified by the findings where *Geobacter sulfurreducens* DNMT2 (*Gs*DNMT2) shows low methylation activity in tRNA-Asp and it has preferably methylates of the cytosine 28 of tRNA-Glu [18]. The trade in the specificity of *Gs*DNMT2 from tRNA-Asp to tRNA-Glu is related to changes within the *Gs*DNMT2 protein and the *Gs*-tRNAs. However, these modifications are only efficient when combined, because *Gs*DNMT2 still methylates tRNA-Asp from other species.

In addition to evolutionary studies based on molecular markers, another powerful tool that can substantially contribute to evolutionary studies is comparative-homology modelling, which is an approach that has great success for the prediction of protein 3D structures [19,20]. This process derives from some assumptions that protein structures have a higher degree of conservation than their amino acid sequences [21]. Significant progress in the prediction algorithms of the tertiary protein structure has been made in recent decades [22]. With the increased accuracy of these methods, a new dimension in the evolutionary analysis of proteins is envisioned.

Furthermore, the analysis of primary and tertiary protein sequences provides interesting suggestions regarding the evolution of protein function [23]. Clearly, the protein spatial organization tends to be more conserved than nucleotide and the amino acid sequence during evolution. However, the increasing information about three-dimensional protein structures has shown that homologous proteins might be structurally different, despite the high conservation of their primary sequence and functional similarities [24]. Tertiary structure protein analyses contribute to a better understanding of the evolution of protein functionality. From the organization's knowledge of the different domains that compose a protein, it is possible to compute molecular surfaces with an electrostatic potential. The electrostatic interactions are the fundamental driving force underlying biological processes, playing an important role in protein-ligand interaction and protein-protein molecular recognition [25].

Our objectives in present work were to analyse the structural and physicochemical characteristics evolution of the DNMT2 in Drosophilidae comparing to *Spodoptera frugiperda* [26], human [27], *Entamoeba histolytica* [28], *Haemophilus haemolyticus* [29], [30], *Haemophilus influenzae* [31] DNMT2 and *Geobacter sulfurreducens* DNMT2 [18]. For this, we perform molecular homology modelling and determination of the electrostatic potential molecular surface of the DNMT2 enzymes. The analysis from 3D structures would launch a new layer for an understanding of the variations in the function and the recognition mechanisms presented in the DNMT2 enzyme and for predicting the impact of mutations through its evolution.

Materials and methods

Sequences and crystallographic structures retrieving

In silico searches were performed to identify DNMT2 homologous sequences among the 24 sequenced *Drosophila* genomes (S2 Table) available in the FlyBase database (<http://flybase.bio.indiana.edu/blast/>). In most cases, the sequence annotated as "DNMT2" was directly recovered. In the other genomes, the *D. melanogaster* sequence (Flybase Annotation symbol: Dmel/CG10692) was used as query on BLASTn. The *D. willistoni* sequence used in this study was the isoform B, described by Garcia et al. (2007), which has 341 amino acids. The *D. buzzatii* and *D. suzukii* DNMT2 were obtained from the *Drosophila buzzatii* Genome Project server (<http://dbuz.uab.cat/welcome.php>) and SpottedWingFlyBase (<http://spottedwingflybase.oregonstate.edu/>), a dedicated online resource for *Drosophila suzukii* genomics. All nucleotide and amino acid sequences were then aligned using the Muscle tool [32]. The *Mus musculus* DNMT2 sequence (NP_034197) was obtained on BLASTp using as query the human DNMT2 sequence and the *Geobacter sulfurreducens* DNMT2 directly from accession number (GSU0227) from Uniprot (<http://www.uniprot.org/>).

Using as query sequence of *D. melanogaster* and human DNMT2, we performed a search through the program BLAST (Basic Local Alignment Search Tool) by homologous structures in the Protein Data Bank (PDB).

Molecular modelling of dnmt2

The DNMT2 models were generated using a homology modelling approach. The Modeller 9.14 [33] program was applied using the *Spodoptera frugiperda* (PDB 4H0N) [26] and human DNMT2 structures as templates (PDB 1G55) [27], according to the best alignment between the template and the target sequence. We used the complete DNMT2 sequences obtained by searching *in silico*. The alignment was further verified manually and adjusted, considering the location of insertion/deletion in loops. The homology modelling was performed with a semi-automated approach, using Python scripts previously developed by our group. The modelling protocol followed the default optimization and refinement protocol, as described in the

Modeller online manual (available at <http://salilab.org/modeller/9.13/manual/node19.html>). One hundred models were generated, and the best model was selected using the DOPE score and Procheck [34].

The generated models were evaluated by Ramachandran plot analysis, which is a well-known evaluation tool to assess the stereochemical quality of a model through the analysis of *phi* and *psi* angles for all protein residues and an overall model quality evaluation was performed using Qmean6 [35] and Verify3D [36]. The DNMT2 models were used for subsequent analysis.

Structural conservation analyses

Sequence and structure analyses are often split into two separate approaches in evolutionary research. In this study, we conducted analyses on the structures of the DNMT2 models generated and their sequences using the Multiseq [37] program incorporated in the VMD 1.9.2 [38] as a plugin tool. The structural homology measure is based on the structural similarity measure, *QH*, which was designed to include the effects of the gaps on the aligned portion [39].

Electrostatic potential molecular surface analyses

For the calculation of the Poisson-Boltzmann equation, the *Adaptive Poisson-Boltzmann Solver* software (APBS) [40] was used, which is able to describe the electrostatic interactions between the molecule and solvent. The APBS was run as a plugin in 1.9 programs Chimera [41] and VMD 1.9.1.

Comparative analysis of the electrostatic properties of proteins was performed using the PIPSA algorithm (Protein Similarity Property Interaction Analysis) [42]. The program permits the classification of proteins according to the properties of molecular interactions fields, and the electrostatic potential molecular surface is the most informative molecular interaction field in these cases.

Results

Molecular modelling DNMT2

The search in the PDB database returned seventeen crystallographic structures of DNMT2 (S1 Table). In the present study, we used *S. frugiperda* [26], human [27], *E. histolytica* [28], *H. haemolyticus* [29] and *H. influenzae* [31] crystallographic models.

To analyze the structural conservation and emerging physicochemical properties of the tertiary structure of DNMT2 we generated twenty-six DNMT2 models of species whose genomes are deposited in databases from their amino acid residue sequences. The models of the DNMT2 protein structure were approved by Verify 3D analysis and presented a high percentage (>86%) of residues with an average 3D-1D score > 0.2. In the Ramachandran plot, all models presented around 90% of residues the most favoured region. A small percentage of residues of the models were in the disallowed regions (in the region of 1%, in every models), and most of these belonged to loop structures of the molecules (S2 Table).

The overall model evaluations were performed by Qmean6. The QMEAN6 score is composite by a linear combination of 6 terms (estimated model reliability between 0–1), with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography. The models were concluded to be of good quality, according to the z-scores obtained by Qmean6 (S2 Table). The crystallographic structure of *S. frugiperda* DNMT2 (PDB: 4H0N) [26] was used as a quality parameter of the models.

Differences in amino acid composition of DNMT2

Despite to be considered an enzyme with high conservation along the various taxa, even the identity between related species is not high. Between *H. haemolyticus* (HhaI) and *H. influenzae* (HaeIII) there is a divergence of 0.7370 (S3 Table). The highest divergence is found between *M. musculus* and HaeIII with 0.8121. The lowest divergence values are found between *D. sechellia*—*D. simulans* (0.0116) and *D. persimilis*—*D. pseudoobscura* (0.0086). In drosophilids the highest divergence found is between *D. willistoni*—*D. biarmipes* (0.3021).

Interestingly, *D. willistoni* are in a position more ancestral in the phylogenetic relationship based on identity, grouping externally to *Drosophila* subgenus (Fig 1A). *D. bipectinata* and *D. ananassae* are placed out of *melanogaster* group when was expected a basal positioning. The other evolutionary relationships between species are established as expected.

Comparatively analyzing the distribution of basic residues on the surface of eukaryotes DNMT2 we can observe a higher prevalence of basic residues in drosophilids DNMT2 in relation to vertebrate (Fig 1B). In contrast, GsDNMT2 differs both as structural conformations as in the presence of basic amino acids in a different pattern from other eukaryotes DNMT2 (Fig 1B).

We assess the conservation of key residues involved in the substrate interaction according to the experimental data [43]. The residues that interfere strongly in the catalytic action of human DNMT2 are present in DNMT2 of drosophilids with synonymous mutations of physicochemical properties in two sites: R275>K and K367>R (Fig 2). However, in GsDNMT2 are not present the key residues K122, R289, K295 and R371. In Ehmeth there is no match in three keys residues: R84, K122 e R371 (Fig 2).

The DNMT2 present structural conservation of catalytic domain, but differs in TRD

The structural homology measure of the DNMT2 models was conducted by Multiseq and Chimera programs (Fig 3). The molecules were coloured according to the sequence identity, and

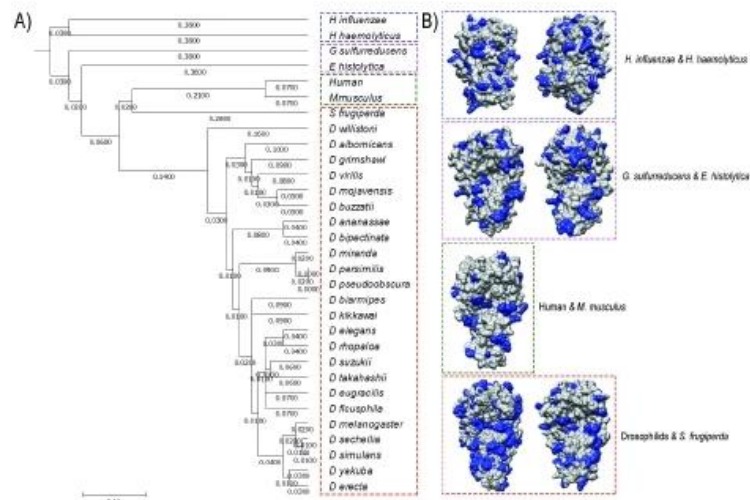


Fig 1. (A) Phylogenetic relationships based on the identity of DNMT2 sequences. The numbers above the branches are the percentage identity distance. **(B)** Representation of eukaryotes DNMT2 frontal structures with surface colored gray. All drosophilids structures are superimposed. The positively charged residues (histidine, lysine, and arginine) are colored blue.

<https://doi.org/10.1371/journal.pone.0178643.g001>

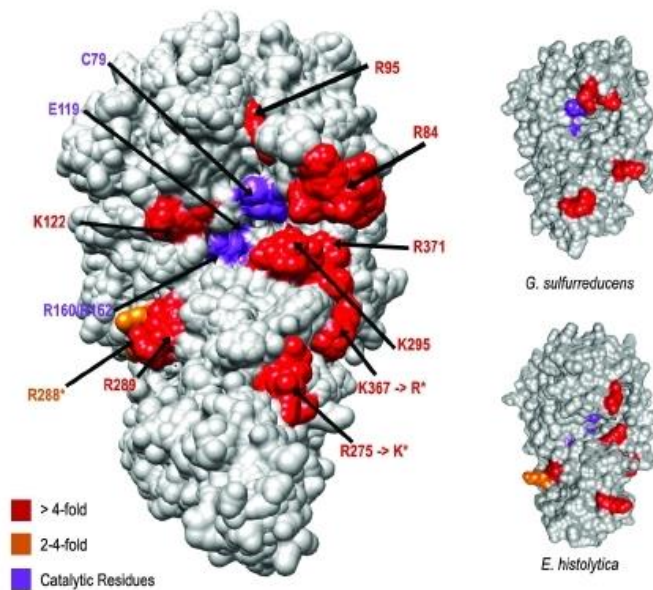


Fig 2. Superposition of drosophilids and human DNMT2 surface representations showing the conservation of residues that have a strong influence in the binding site, according to Jurkowski et al. 2012 where residues that strongly interfere with catalysis activity (> 4-fold) are colored red, 2-4-fold orange and catalytic residues colored purple. The residues K367 and R275 from human DNMT2 correspond to arginine (R) and lysine (K) in drosophilids, respectively. The *G. sulfurreducens* and *E. histolytica* DNMT2 surfaces are presented separately.

<https://doi.org/10.1371/journal.pone.0178643.g002>

we can measure the charge variation per site from the structures. The conserved residues are located mostly in the catalytic region, whereas the low identity residues are located in the connection of the catalytic domains and TRD. Additionally, TRD has a lower amino acid identity, despite the preservation of signature peptides in the area (motifs CFT and E/DGTS) (Fig 3B).

There is also a structural difference in the substrate recognition domain: an insertion (approximately 9 residues) in various drosophilids between the characteristic α -helix/loop/ α -helix motif of this region, except in *D. ananassae*, *D. bipectinata*, *D. grimshawi*, *D. tropicalis* and *D. willistoni*. As expected, structural variation can be observed between closed species, as the genus *Haemophilus* (Fig 3A), since these proteins have different target-sites methylation. Throughout the evolution, the C-terminal region of TRD gradually assumed a more structured architecture with the appearance of the characteristic structural motif α -helix/loop/ α -helix.

Analyzing the differences in surface electrostatic potentials of the mutated MT2

Cohen et al. (2004) altered the sequence specificity of HaeIII by directed evolution *in vitro*. The mutant variations showed three new target sites efficiently: AGCC, CGCC and the original GGCC, which the methylation in CGCC site is slightly more efficiently than AGCC [44]. Thus we promoted the respective mutations in HaeIII, according to Cohen et al. experiment (Table 1) for the comparative analysis of the electrostatic properties (Fig 4A and 4B).

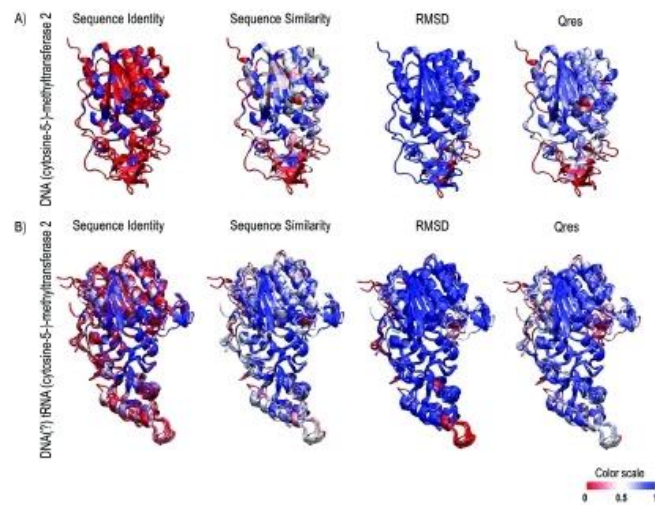


Fig 3. Multiple structural alignments of all DNMT2 by STAMP in VMD-multiseq. The structures are colored by sequence identity, similarity, RMSD and Qres values (where 1 indicates that the structures are identical and 0 more dissimilar). **(A)** Overriding DNMT2 of HhaI (PDB 1MHT) and HaeIII (PDB 1DCT) structures. **(B)** Superposition of remaining DNA/RNA MTase2 structures, including GsDNMT2.

<https://doi.org/10.1371/journal.pone.0178643.g003>

HaeIII-wt appears distantly from all other HaeIII mutated (Fig 4A). Interestingly the change of sequence specificity in HaeIII occurs by only the changing of four residues (HaeIII-T29), which impacts the electrostatic potentials surface, causing a decrease in the positive electrostatic potential in the mutated region (Fig 4B).

We also elaborate eight human DNMT2 structures with the six respective mutations that strongly interfere with catalysis activity (Table 2). In this case, the analyses of differences in surface electrostatic potentials for each mutant type appear to be more drastic in response to mutated residues (Fig 4C and 4D).

Extensive comparative analysis of the electrostatic properties of DNMT2 surface

As the interaction between the substrate (DNA, RNA or tRNA) occurs distributed around a region, with multiples residues interacting with ligand we posed the analysis probe in a central position in DNMT2 TRD. This position for TRD is the classical CFT motif present in eukaryotes MT2 (not present in prokaryotes, corresponding spatially to TLS and VQA motifs from HhaI and HaeIII, respectively). The region is assigned with a radius of 35Å around the correspondent TRD.

The comparison of the electrostatic interaction properties of proteins analysis resulted in four main clusters: I, II, III and IV (Fig 5A). The cluster I is mostly for species of the *melanogaster* group, sharing high similarity in the electrostatic surface profile (Fig 5A and 5B). A sub-cluster I-1 can be found, which *D. sechellia*, *D. simulans* and *D. melanogaster* show the lowest electrostatic distance values: 0.3715 (*D. simulans*-*D. melanogaster*), 0.4243 (*D. simulans*-*D. sechellia*) and 0.4604 (*D. sechellia*-*D. melanogaster*) (Fig 5A). The subcluster I-2 also shares high similarity between the species and the distribution of charges on its surface (Fig 5B).

Table 1. Sequence of HaeIII mutants from Cohen et al. 2004.

Type	Residue				AGCC activity	GGCC activity
	225	260	261	262		
Wild-type	R	N	L	N	<0.05	50.0
R225A	A	N	L	N	0.96	1.2
T29	A	L	M	W	33.0	5.1
T7	A	L	F	W	3.9	0.12
T28	A	L	S	W	3.0	0.41
T26	A	L	T	W	2.9	0.32
T45	A	L	W	W	2.3	0.32
T1	A	L	R	W	2.0	0.10
T40	A	L	C	W	1.8	0.071

<https://doi.org/10.1371/journal.pone.0178643.t001>

The cluster II is very heterogeneous through the similarities and distribution of charges on its surface, being composed of species from *obscura* subgroup, *Drosophila* subgenus and *D. ananassae* (*Sophophora* subgenus) (Fig 5A). Here we can recognize two distribution of surface electrostatic charges patterns: one represented by *D. ananassae* electrostatic potential surface (cluster II-1) and other represented by *D. pseudoobscura* pattern (cluster II-2) (Fig 5B). Clusters III and IV also behave proteins with heterogeneous surface properties. In cluster III we find the *Entamoeba histolytica*, *D. albomicans*, *D. virilis* and *D. willistoni* structures, however only *D. virilis* and *D. willistoni* share similarity in its electrostatic potential surfaces, with a distance value of 0.6387. The cluster IV-1 groups Human, *M. musculus* and *S. frugiperda*, but the similarity between *S. frugiperda*-*M. musculus* and *S. frugiperda*-Human MT2 are low, distance values 0.9044 and 0.9154 respectively (Fig 5A and 5B). At least, in the cluster IV-2 we find de DNA methyltransferase from HhaI and HaeIII. The *G. sulfurreducens* is also grouping together with HhaI and HaeIII but does not share similarities with prokaryotes DNMT2.

Interestingly, the human and *M. musculus* DNMT2 stand out in comparison with other MT2 molecules analyzed the smaller area of positive electrostatic potential fields in their TRD. On the other hand, the electrostatic surface of prokaryote HaeIII e HhaI is characterized by a prominent prevalence of positive charges in the TRD (Fig 5B). The lower prevalence of positive potential in the interaction interface with the substrate appears to be a feature of tRNA (cytosine-5)-methyltransferase.

The drosophilids electrostatic potential surface profile is somehow heterogeneous between them. It is clear that the distribution pattern of surface charges in drosophilids DNMT2 differs from Ehmeth and GsDNMT2, the same occurring with human and *M. musculus* DNMT2 (Fig 5B).

Predicting kinetic constants K_{cat}/K_m from DNMT2

The reaction mechanisms of DNMT2 have not been studied extensively, except for the prokaryotes HhaI and HaeIII, which have consistency measure for its kinetic constants (Table 3). Molecular electrostatic potential differences seem to correlate with kinetic rates and this correlation can be used to predict enzyme kinetic parameters [45]. The parameters K_{cat}/K_m are linked with the region responsible for the interaction between the substrate and the enzyme, so the electrostatic potential closer to the active site presents significant for the kinetic parameter values.

An excellent linear correlation between calculated differences in electrostatic potentials and kinetic values could be achieved ($R^2 = 0.9512$) and this can be used to predict $\ln K_{cat}/K_m$ values for the other MT2 (Fig 6A). Thus, we used the kinetic constants K_{cat}/K_m rate from HhaI and

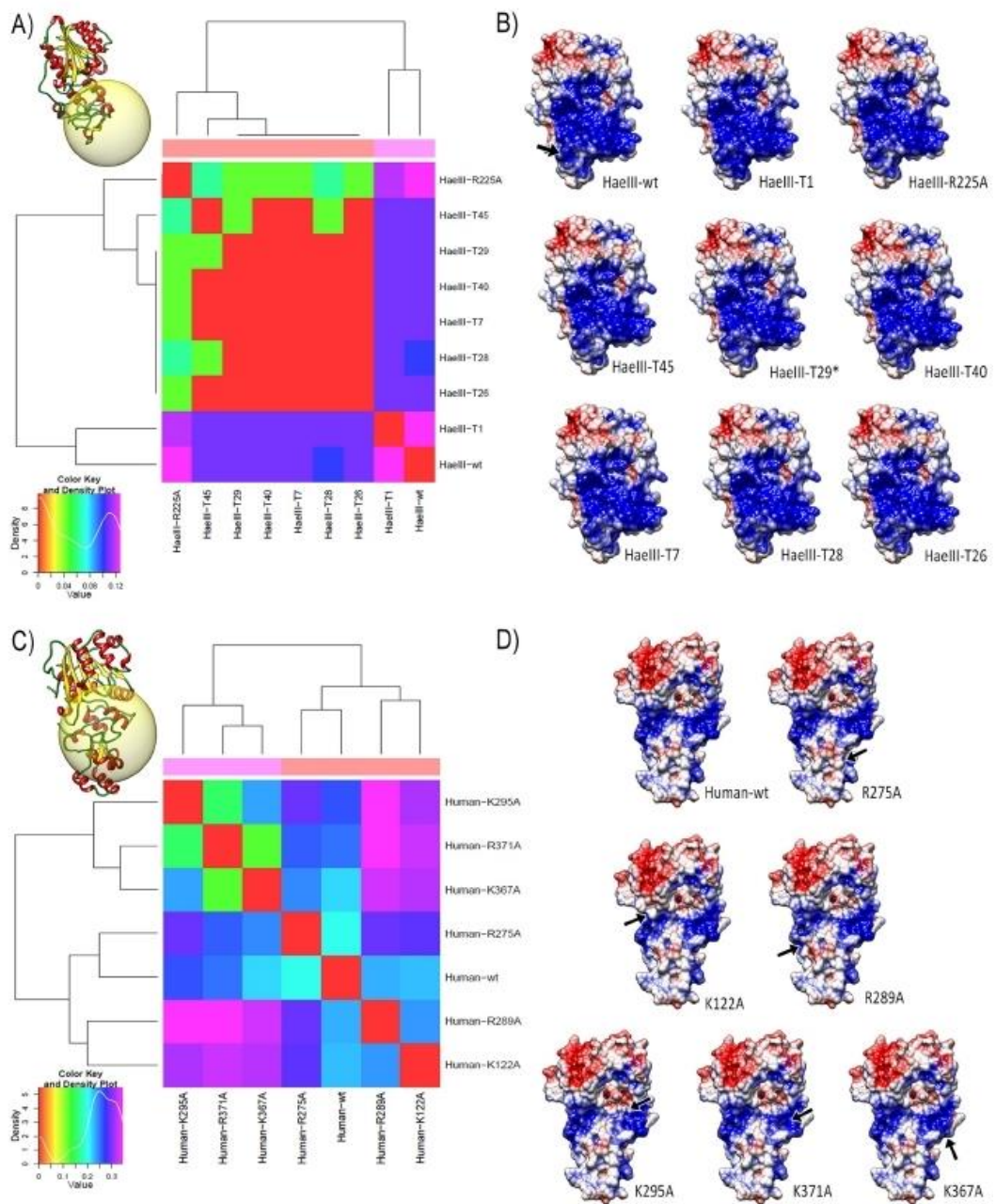


Fig 4. Epogram of MT2 structures. The molecules are grouped according to the similarity of the electrostatic potentials from the distance matrix constructed by PIPSA tool. The colour key, above left, indicates the relationship of similarity: where red indicates maximum similarity.

whereas violet indicates low similarity. The surfaces are coloured according to the electrostatic potential: Blue for positive potential (5 kT), red for negative (-5 kT) and white for neutral, where kT = kilotesla. The black arrow in the top left structure indicates the target region of mutated residues. **(A)** Epogram showing the clusters of the comparison of HaeIII electrostatic potentials, at up left are shown the ribbon representation of the region for comparison of the electrostatic potential. **(B)** Electrostatic potential surface of HaeIII structures (substrate interface). **(C)** Human MT2 epogram from comparison of electrostatic potentials, at up left are shown the ribbon representation of the region for comparison of the electrostatic potential. **(D)** Electrostatic potential surface of human MT2 structures (substrate interface) and respective mutations.

<https://doi.org/10.1371/journal.pone.0178643.g004>

HaeIII to infer the measure the kinetic values for the others DNMT2 analyzed in present work. For the K_{cat}/K_m parameter, we find that an increase of K_{cat}/K_m of 1 ln unit is related to a decrease of 0.1046 kcal/mol/e in electrostatic potential of surface residues, approximately (Fig 6A). In the extreme points, we have the differences between the pairs HaeIII-*D. yakuba* (-1.08E+01; 1.20E+00) and *D. rhopaloa*- HaeIII (1.37E+01; -1.53E+00).

HaeIII has the lowest $\ln K_{cat}/K_m$ value of all (1.02E+01), while HhaI has $\ln K_{cat}/K_m$ value 1.63E+01. Interestingly, *D. willistoni* (1.59E+01), *D. virilis* (1.59E+01) and *S. frugiperda* (1.30E+01) show a predicted $\ln K_{cat}/K_m$ values between HaeIII and HhaI. Instead, *D. rhopaloa* show the highest $\ln K_{cat}/K_m$ (2.39E+01) (Fig 6B). When we clustered the values from differences in $\ln K_{cat}/K_m$ and electrostatic potentials, setting the drosophilids vs all other DNMT2, we can observe that the most drosophilids DNMT2 shares similarity with vertebrates DNMT2, Ehmeth and GsDNMT2, while depart from DNA-MT2 (HaeIII and HhaI) and from S β DNMT2 (Fig 6C). Analyzing separately *D. willistoni*, *D. melanogaster* and *D. rhopaloa*, we can see more clearly the differences and similarities pair to pair with theses MT2. *D. melanogaster* seems to be more similar to the other drosophilids regarding to the differences in the electrostatic potentials surface with average differences 0.181 Kcal/mol/e (Fig 6D), while *D. willistoni* and *D. rhopaloa* seems to be more dissimilar with the other drosophilids MT2, averages 0.366 and 0.322 Kcal/mol/e, respectively (Fig 6E and 6F). However, *D. willistoni* MT2 shows greater similarity with HhaI (0.045 Kcal/mol/e) and HaeIII (0.637 Kcal/mol/e), whereas *D. rhopaloa* has the largest difference, 0.870 Kcal/mol/e and 1.529 Kcal/mol/e, respectively, a more similar profile to *D. melanogaster* in relation to these prokaryotic MT2

Discussion

In general, the DNMT2 is considered a tRNA-ASP methyltransferase than precisely a DNA methyltransferase [2,3,47]. However, the discovery of sex-specific methylation phenomenon reestablishes the importance of maintaining a critical view of the importance and the multifunctionality of DNMT2, precisely due to the peculiarities within groups of species evolutionarily close and described as DNMT2-only.

By the end of the 90s the genome of *Drosophila melanogaster* was considered methylation-free, as described in several studies that analyzed the stages of pupa and adult *Drosophila* [48,49]. However, studies have been presented evidences that there would be lower levels of

Table 2. Sequence of human DNMT2 mutants from Jurkowski et al. 2012.

Type	Residue	WT % activity	Deviation (%)
wt		100.00	-
Mutant-I	K295A	20.24	±1.71
Mutant-II	K122A	18.54	±6.71
Mutant-III	R289A	19.27	±9.02
Mutant-IV	R371A	14.15	±2.44
Mutant-V	K367A	10.37	±2.07
Mutant-VI	R275A	6.46	±1.71

<https://doi.org/10.1371/journal.pone.0178643.t002>

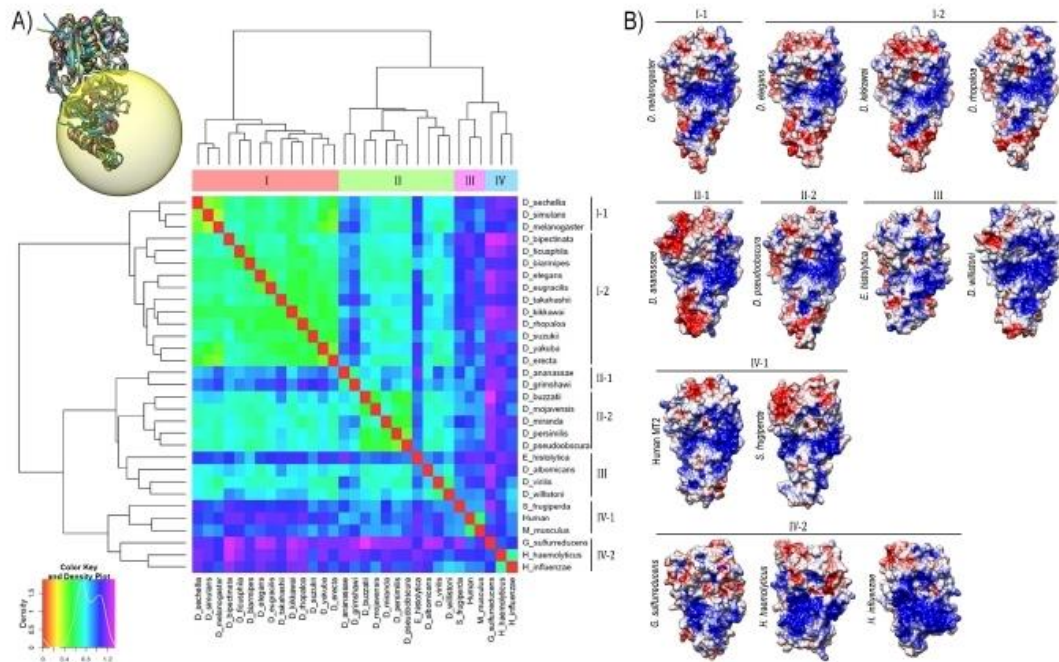


Fig 5. (A) Epogram of MT2 structures. The molecules are grouped according to the similarity of the electrostatic potentials from the distance matrix constructed by the program PIPSA. The colour key, above left, indicates the relationship of similarity: red indicates maximum similarity, whereas violet indicates low similarity. **(B)** Electrostatic potential surface of DNMT2 structures (substrate interface). The surfaces are coloured according to the electrostatic potential: blue for positive potential (5 kT), red for negative (-5 kT) and white for neutral, where kT = kilotesla. The structures are grouped according to the similarity of the distribution profile of electrostatic potentials.

<https://doi.org/10.1371/journal.pone.0178643.g005>

methylation in *Drosophila* embryos and adults. The analysis employed confirmed the existence of methylation in these organisms, but with the peculiarity of being sites for CpA, CpT, and CpC in embryos of *D. melanogaster*, and not CpG as is commonly found in mammalian [11]. Our research group which has focused on the study of Neotropical species (as *willistoni* subgroup) showed for the first time the phenomenon sex-specific methylation in *D. willistoni* genome [16]. Using the Methylation-Sensitive Restriction Endonuclease (MSRE) technique and Southern blot with specific probes, the results suggested that selection for different targets of methylation may occur between different, but closely related species [16]. In this case, species from *willistoni* subgroup seems to have a different methylation sites pattern to *D. melanogaster*, since the restriction sites of enzymes used are 5'-GGCC-3' and 5'-AGCT-3', respectively [16,17]. Given this scenario, in the present work, we approach the evolution of DNMT2 from

Table 3. Kinetic constants for DNA methyltransferase from HaeIII and HhaI.

Organism	K_{cat}/K_m	K_{cat}/K_m (ln)	Reference
<i>H. influenzae</i> (HaeIII)	2.60E+04	1.02E+01	[44]
<i>H. haemolyticus</i> (HhaI)	1.20E+07	1.63E+01	[46]

<https://doi.org/10.1371/journal.pone.0178643.t003>

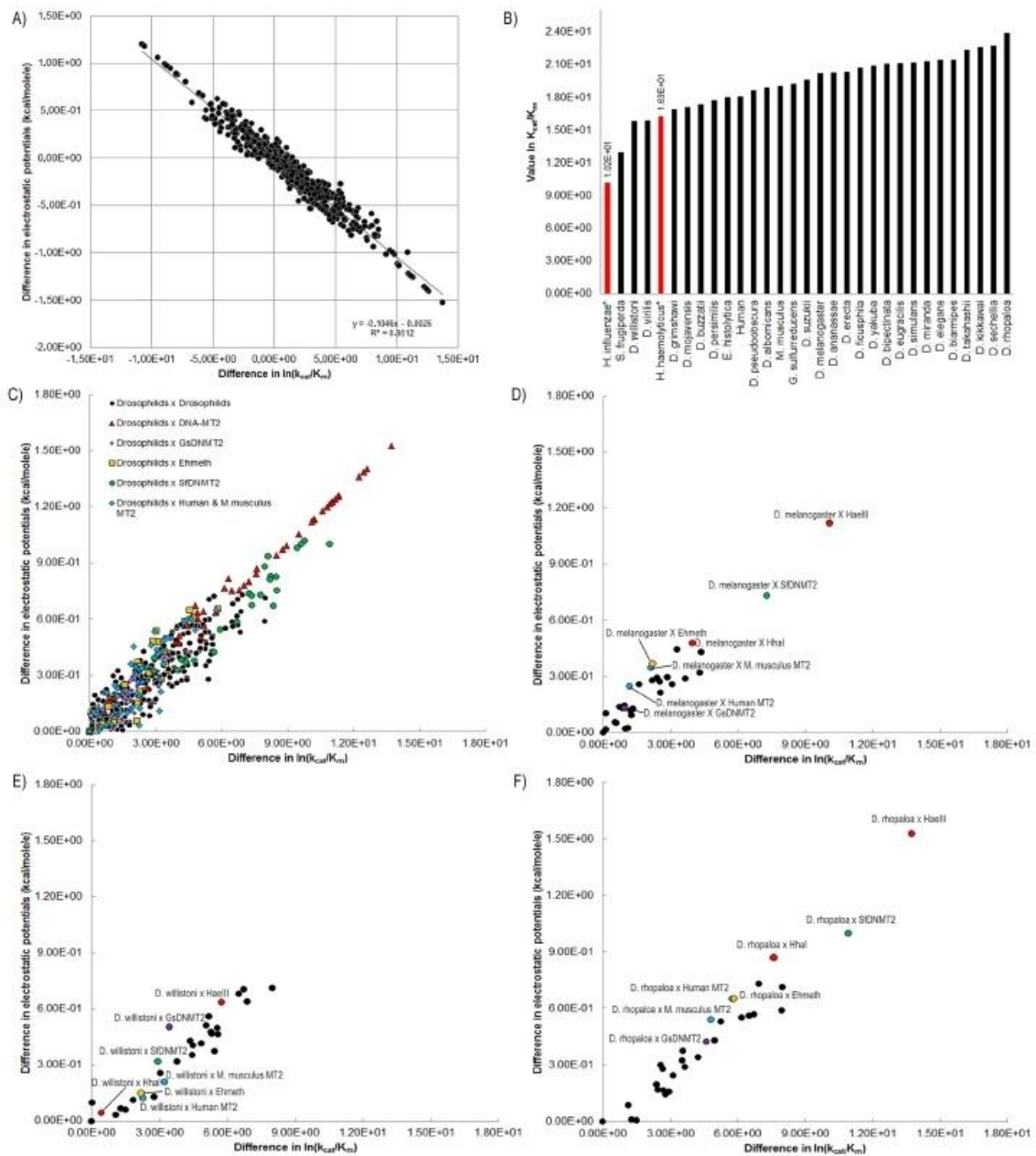


Fig 6. (A) Correlation between experimental in K_{cat}/K_m values and the electrostatic potential differences. **(B)** Predicted K_{cat}/K_m from methyltransferase 2 structures. The experimental values from HhaI and HaeIII are highlighted in red. **(C)** Clusters from correlation in K_{cat}/K_m and electrostatic potential differences setting drosophilids correlations with all others. The values were transformed to the same quadrant. **(D)** Correlation between in (K_{cat}/K_m) for *D.*

melanogaster DNMT2 and electrostatic potential differences for each DNMT2 protein pair. (E) Correlation between $\ln(K_{cat}/K_m)$ for *D. willistoni* DNMT2 and electrostatic potential differences for each DNMT2 protein pair. (F) Correlation between $\ln(K_{cat}/K_m)$ for *D. rhopaloea* DNMT2 and electrostatic potential differences for each DNMT2 protein pair.

<https://doi.org/10.1371/journal.pone.0178643.g006>

structural and physical chemistry perspective in search of features that shed light on the functional plurality of the enzyme, with special attention to drosophilids.

Although there maintenance of certain catalytic motifs, generally DNMT2 cannot be described as being high conserved among different organisms (Fig 1A and S1 Table), average divergence among the studied species are 40% (S1 Table). Even between drosophilids, it finds great sequences divergence. It is interesting to note that the signatures of evolutionary processes expand beyond the analysis of synonymous substitutions and not synonymous, they are also found in the conformational "clues" of new structures that may represent functional changes of proteins. Thus, 3D similarities in protein structures may exist in the absence of sequence identity [19,50]. HaeIII and HhaI share only 30% identity, but have high structural conservation demonstrated by the RMSD values and the structural differences between HaeIII and HhaI are found in TRD (Figs 1A and 4A), which is consistent with the fact that both having the same function, however, with different target-sequence recognition [29,31].

Clearly, it is observed that the DNMT2 spatial organization tends to be more conserved than their sequence during evolution (Fig 3A and 3B), majority into the catalytic domain, as expected, since the maintenance of catalytic function in different DNMT2. However, structural differences of TRD are more pronounced when comparing the DNA-DNMT2 to tRNA-(DNA)DNMT2. The Ehmeth TRD assumes an intermediate architecture between the two DNMT2 types but has a more juxtaposed overlap with tRNA-(DNA)DNMT2. The differences both in composition and architecture of DNMT2 here studied are linked to differences in biological functions and characteristics of action of TRD on the substrate observed in the DNMT2 of different species, which even in close species (like *H. haemolyticus* and *H. influenzae*) present a high difference in composition, structure and recognition mechanisms. GsDNMT2 shows a structural composition more closed to Ehmeth than others eukaryote DNMT2, remembering that it has a preference for tRNA-Glu than the tRNA-asp substrate. The changes in the target substrate recognition are closely linked with enzyme and substrate architecture at the same time in this case [18].

Electrostatic interactions of a protein are characterized as long-range attractive forces, capable of interfering with the association rate between two molecules, as a cofactor-protein and protein-substrate and it is another interesting level to analyzing when we study the evolutionary modifications. For being a strong and long-range force, the knowledge of how the electrostatic potential is distributed on the surface of a protein ends up being critical to understanding the behavior and function of a molecule.

The alteration of sequence specificity or the processivity of an enzyme can be obtained by the changing of only one or a couple of residues, sometimes. Cohen et al. (2004) altered the target-sequence of MT2 HaeIII from GGCC to AGCC by the mutation of residues R225, N260, L261 and N262. These mutations also modified the electrostatic potentials surface, decreasing substantially the positive charge in the mutated region (Fig 4A and 4B). In another experiment, Jurkowski et al. (2012) systematically mutated lysine and arginine residues to alanine to study the methylation activity and binding of the variants and found eight residues that caused a strong decrease in catalytic activity. Despite many variants shown strong reduce of catalytic activity, only a weak loss of tRNA binding or even bound better to tRNA wild-type DNMT2. In this case, probably the DNMT2 can induce conformational changes in tRNA before the transfer reaction of a methyl group to target cytosine. As expected, the analysis of differences

in surface electrostatic potentials for each mutant correlates with the experimentally observed loss of activity (Fig 4C and 4D).

Interestingly, the extensive analysis of electrostatic potential surfaces shown that the evolutionarily related groups have striking differences in how the charges are distributed on the surface of their molecules when the expectation was for drosophilids been clustering with low difference distance (Fig 5A and 5B). However, drosophilids were grouped into two main clusters, with the high distance between the species from these two clusters. The cluster-II is very heterogeneous with regard to the distribution profile of electrostatic potentials (Fig 5B). It draws attention to the electrostatic potential surface of *D. willistoni* in relation to other drosophilids. Visual analysis of the distribution of the *D. willistoni* DNMT2 surface charges reveals a larger area with a positive electrostatic profile in TRD of this Neotropical species (Fig 5B), in addition to group separately from the other species of the subgenus *Sophophora*.

For the prediction of kinetic constants from DNMT2 was taken into account that the kinetic parameters are linked with the region responsible for the interaction between the substrate and the enzyme [45]. From linear correlation calculated differences in electrostatic potentials the predicted kinetic values could be achieved (Fig 6A). The average values $\ln K_{cat}/K_m$ for prokaryote DNMT2 are $1.32E+01$ kcal/mol/e, when for drosophilids the average show higher predicted values, $2.04E+01$ kcal/mol/e. However, *D. willistoni*, *D. virilis* and *S. frugiperda* have predicted kinetic values very close to the HaeIII and HhaI, which can suggest a differential reactivity for the same substrate of prokaryotes DNMT2, in the case, DNA (Fig 6B). It is clear the heterogeneity of drosophilids MT2 (Fig 6C–6F).

The kinetic parameter K_{cat}/K_m measures the affinity to the transition state. Often a very reactive substrate will have a high K_m value since it will react faster than it is released from the enzyme unreacted. Together with the pattern of distribution of positive and negative net charges of the molecular surface, these data are in agreement with the fact that surface residues can influence kinetic values not only via the potential in the active site region but also influence the substrate affinity as it approaches the active site [45].

The differences found here may be related to the various peculiarities present in the evolutionary history of each of the studied strains. Even among drosophilids, which we would expect similarity of potentials between species, their evolutionary histories differ greatly in several aspects. For example, the ways the transpositions of TEs are regulated are mainly associated with epigenetic mechanisms [51,52]. The distribution of TEs and repetitive sequences in many *Drosophila* genomes vary greatly in relation to the percentage found: *D. simulans* 2.7%, *D. willistoni* 16% and *D. ananassae* 25% [53]. The events of TEs mobilization may be related to the occurrence of breaks and rearrangements of chromosomal arms, modifications to regulatory gene sequences, or even the complete inactivation of a gene, fulfilling an important role in molecular evolution of organisms, being able to promote the emergence of new genes, as well as new pathways in cellular signaling network and *D. willistoni* is known as a specie with high chromosomal polymorphism [54].

On the other hand, the differences between the *D. willistoni* and other species DNMT2 may be related to the processivity of the enzyme with the RNA substrate. Besides the genomic cytosine methylation as a process to silencing TEs, the silencing of retrotransposons and the control of RNA viruses in *Drosophila* can be mediate by DNMT2, playing an important role as a control mechanism to stress that involves RNA, like retrotransposons [55,56]. DNMT2 is an essential part of the RNA processing machinery during cellular stress [57]. In *Drosophila*, when in heat stress conditions, DICER-2 enzyme degrades tRNA and tRNA fragments, so DNMT2 then goes into action to limits the extent of tRNA fragmentation during the stress event, because too long double-stranded RNAs inhibit DICER-2 activity [56,58]. More recently, it was surprisingly found that DNMT2 methylates more efficiently DNA fragments

when presented as covalent DNA-RNA hybrids in the structural context of a tRNA [59]. Such context containing both RNA and DNA covalently linked would be found in replication forks, transcription sites, DNA repair events, or retroviral replication. Moreover, it seems that the DNMT2 substrate recognition process is linked to the recognition of at least two levels: structural and primary sequence [59].

Whereas *willistoni* subgroup have sex-specific methylation of cytosine in its genome (present only in drosophilids belonging to this subgroup) [16], a high degree of chromosomal polymorphism [54,60], a wide distribution area—where many people are exposed to different environmental scenarios—, may favor a selection toward some degree of differential modulation of the DNMT2 function. *D. willistoni* also has a high percentage of enzyme polymorphism, around 60% [61]. Among the species of *Drosophila*, *D. willistoni* also shows a reduction in the diversion of use of codons—which is related to the improvement of protein synthesis—in relation to other drosophilids [62,63]. Thus, it is expected in the subgroup *willistoni* mechanisms ensuring adequate genomic stability but at the same time have sufficient plasticity to allow rapid adaptive capacity in various environments.

Conclusions

Studies about the functional diversities of macromolecules can lead to a better understanding of the biological diversity that exists in nature. The macro-aspects (molecule as a whole and interactions with other molecules) and micro (physicochemical characteristics, distribution of electrostatic potentials, behavior in physiological environment and conservation of residues) are fundamental for the understanding of which and how different evolutionary mechanisms can be crucial in preserving or acquiring new functions of a molecule. In the present work, it was verified that DNMT2 of evolutionarily close species have different characteristics in the distribution of surface potentials, mainly in the region responsible for the recognition of the target sequence. This feature may be involved with different affinities between the enzymes for its different substrates (DNA, RNA or tRNA) and even for the epigenetic mechanisms involved throughout the evolutionary history of the species.

In the case of the Neotropical species of the *willistoni* subgroup, we suggest that there may be an intermingled system of intrinsic factors (presence of TEs and chromosomal polymorphism, for example) and extrinsic (geographic distribution and variable niches) that would be acting as evolutionary forces in the emergence and maintenance of differences observed in DNMT2. These aspects raise new questions about the importance of DNMT2 in the ontological, ecological and evolutionary aspects of this peculiar group of drosophilids.

Supporting information

S1 Table. DNMT2 crystallographic models deposited in the database Protein Data Bank used to analyze and select the template structure for homology modeling process and evolutionary studies.

(DOCX)

S2 Table. Percentage (%) of protein residues in each region of the Ramachandran Plot and additional evaluation tools for model viability assessment.

(DOCX)

S3 Table. Estimates of evolutionary divergence between sequences. The number of amino acid differences per site from between sequences are shown. Standard error estimate(s) are shown above the diagonal. The analysis involved 31 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 385 positions in the final

dataset. Evolutionary analyses were conducted in MEGA7. (DOCX)

Acknowledgments

This study was supported by grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS no 10/0028-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD-CAPEs).

Author Contributions

Conceptualization: GCV GFV MS VLSV.

Formal analysis: GCV.

Funding acquisition: VLSV.

Investigation: GCV.

Methodology: GCV GFV MS.

Project administration: VLSV.

Resources: VLSV.

Supervision: GFV VLSV.

Validation: GFV MS VLSV.

Writing – original draft: GCV.

Writing – review & editing: GCV GFV.

References

1. Ponger L, Li W. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol.* 2005; 22(4):1119–1128. <https://doi.org/10.1093/molbev/msi098> PMID: 15689527
2. Hermann A, Schmitt S, Jeltsch A. The human Dnmt2 has residual DNA (cytosine-C5) methyltransferase activity. *J Biol Chem.* 2003; 278:31717–31721. <https://doi.org/10.1074/jbc.M305448200> PMID: 12794065
3. Tang L-Y, Reddy MN, Rasheva V, Lee T-L, Lin M-J, Hung M-S, et al. The eukaryotic DNMT2 genes encode a new class of cytosine-5 DNA methyltransferases. *J Biol Chem.* 2003; 278:33613–6. <https://doi.org/10.1074/jbc.C300255200> PMID: 12819212
4. Fisher O, Siman-Tov R, Anki S. Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (EhMeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res.* 2004; 32:287–297. <https://doi.org/10.1093/nar/gkh161> PMID: 14715927
5. Katoh M, Curk T, Xu Q, Zupan B, Kuspa A, Shaulsky G. Developmentally regulated DNA methylation in *Dictyostelium discoideum*. *Eukaryot Cell.* 2006; 5:18–25. <https://doi.org/10.1128/EC.5.1.18-25.2006> PMID: 16400165
6. Okano M, Xie S, Li E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 1998; 26:2536–40. PMID: 9592134
7. Yoder JA, Bestor TH. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet.* 1998; 7:279–284. PMID: 9425235
8. Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. *Science.* 2008; 319:1827–30. <https://doi.org/10.1126/science.1153069> PMID: 18339900
9. Kunert N, Marhold J, Stanke J, Stach D, Lyko F. A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development.* 2003; 130:5083–90. <https://doi.org/10.1242/dev.00716> PMID: 12944428

10. Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, et al. Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res.* 2005; 33:6405–17. <https://doi.org/10.1093/nar/gki952> PMID: 16282589
11. Lyko F, Ramsahoye BH, Jaenisch R. DNA methylation in *Drosophila melanogaster*. *Nature.* 2000; 408:538–540. <https://doi.org/10.1038/35046205> PMID: 11117732
12. Gowher H, Leismann O and Jeltsch, A. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J.* 2000; 19, 6918–6923 <https://doi.org/10.1093/emboj/19.24.6918> PMID: 11118227
13. Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B, et al. Conservation of DNA methylation in dipteran insects. *Insect Mol Biol.* 2004; 13:117–23. <https://doi.org/10.1111/j.0962-1075.2004.00466.x> PMID: 15056358
14. Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L, et al. Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res.* 2014; 24:821–30. <https://doi.org/10.1101/gr.162412.113> PMID: 24558263
15. Bächli G. TaxoDros: the database on taxonomy of Drosophilidae, v.1.04, Database 2016/01. Available via <http://taxodros.unizh.ch/>. Accessed on 20/11/2016.
16. Garcia RN, D'Ávila MF, Robe LJ, Loreto ELDS, Panzera Y, Heredia FO, et al. First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica.* 2007; 131:91–105. <https://doi.org/10.1007/s10709-006-9116-3> PMID: 17205375
17. D'Ávila MF, Garcia RN, Panzera Y, Valente VLS. Sex-specific methylation in *Drosophila*: an investigation of the *Sophophora* subgenus. *Genetica.* 2010; 138:907–913. <https://doi.org/10.1007/s10709-010-9473-9> PMID: 20640488
18. Shanmugam R, Aklujkar M, Schäfer M, Reinhardt R, Nickel O, Reuter G, et al. The Dnmt2 RNA methyltransferase homolog of *Geobacter sulfurreducens* specifically methylates tRNA-Glu. *Nucleic Acids Res.* 2014; 42: 6487–6496. <https://doi.org/10.1093/nar/gku256> PMID: 24711368
19. Kinch LN, Grishin NV. Evolution of protein structures and functions. *Curr Opin Struct Biol.* 2002; 12:400–8. PMID: 12127461
20. Deane C, Blundell T. Protein comparative modeling and drug discovery. In: WERMUTH C. *The practice of medical chemistry.* London: Elsevier Academic Press. 2003.
21. Branden C and Tooze J. *Introduction to protein structure.* 2nd Edition, Garland Publishing, Inc., 1999; New York.
22. Damborsky J, Brezovsky J. ScienceDirect Computational tools for designing and engineering enzymes. *Curr Opin Chem Biol.* 2013; 19:8–16. <https://doi.org/10.1016/j.cbpa.2013.12.003> PMID: 24780274
23. Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nature Structural & Molecular Biology.* 2013; 20:1243–1249.
24. Grishin NV. Fold change in evolution of protein structures. *J Struct Biol.* 2001; 134:167–85. <https://doi.org/10.1006/jsbi.2001.4335> PMID: 11551177
25. Hildebrandt A, Blosser R, Rjasanow S, Kohlbacher O, Lenhof HP. Electrostatic potentials of proteins in water: A structured continuum approach. *Bioinformatics.* 2007; 23:99–103. <https://doi.org/10.1093/bioinformatics/btl312> PMID: 17237112
26. Li S, Du J, Yang H, Yin J, Ding J, Zhong J. Functional and structural characterization of DNMT2 from *Spodoptera frugiperda*. *J Mol Cell Biol.* 2013; 5:64–66. <https://doi.org/10.1093/jmcb/mjs057> PMID: 23103599
27. Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH, Cheng X. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res.* 2001; 29:439–48. PMID: 11139614
28. Schulz EC, Roth HM, Anki S, Ficner R. Structure analysis of *Entamoeba histolytica* DNMT2 (EhMeth). *PLoS One.* 2012; 7:e38728. <https://doi.org/10.1371/journal.pone.0038728> PMID: 22737219
29. Klimasauskas S, Kumar S, Roberts RJ and Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell.* 1994; 76:357–69. PMID: 8293469
30. O'Gara M, Roberts RJ, Cheng X. A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase. *J Mol Biol.* 1996; 263:597–606. <https://doi.org/10.1006/jmbi.1996.0601> PMID: 8918941
31. Reinisch KM, Chen L, Verdine GL, Lipscomb WN. The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell.* 1995; 82:143–53. PMID: 7606780
32. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32:1792–7. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147

33. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* 2000; 29:291–325. <https://doi.org/10.1146/annurev.biophys.29.1.291> PMID: 10940251
34. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol.* 1996; 8:477–486.
35. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011; 27(3):343–50. <https://doi.org/10.1093/bioinformatics/btq662> PMID: 21134891
36. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991; 253:164–70. PMID: 1853201
37. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics.* 2006; 7:382. <https://doi.org/10.1186/1471-2105-7-382> PMID: 16914055
38. Humphrey W, Dalke A, Schulten K. VMD—Visual Molecular Dynamics. *J Molec Graphics.* 1996; 14:33–38.
39. Eastwood MP, Hardin C, Luthey-Schulten Z, Wolynes PG. Evaluating protein structure-prediction schemes using energy landscape theory. *IBM J Res Dev.* 2001; 45: 475–497.
40. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A.* 2001; 98:10037–41. <https://doi.org/10.1073/pnas.181342398> PMID: 11517324
41. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera: a visualization system for exploratory research and analysis. *Comput Chem.* 2004; 25(13):1605–12.
42. Blomberg N, Gabdoulhine RR, Nilges M, Wade RC. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins Struct Funct Genet.* 1999; 37:379–87. PMID: 10591098
43. Jurkowski TP, Shanmugam R, Helm M, Jeltsch A. Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry.* 2012; 51: 4438–44. <https://doi.org/10.1021/bi3002659> PMID: 22591353
44. Cohen HM, Tawfik DS and Griffiths AD. Altering the sequence specificity of HaellII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Eng Des Sel.* 2004; 17:3–11. <https://doi.org/10.1093/protein/gzh001> PMID: 14985532
45. Gabdoulhine RR, Stein M and Wade RC. qPIPSA: relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics.* 2007; 8:373. <https://doi.org/10.1186/1471-2105-8-373> PMID: 17919319
46. Lindstrom WM, Flynn J and Reich NO. Reconciling structure and function in HhaI DNA cytosine-C-5 methyltransferase. *J Biol Chem.* 2000; 275:4912–4919. PMID: 10671528
47. Jeltsch A, Nellen W, Lyko F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci.* 2006; 31: 306–8. <https://doi.org/10.1016/j.tibs.2006.04.005> PMID: 16679017
48. Patel CV and Gopinathan KP. Determination of trace amounts of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* 1987; 164, 164–169. PMID: 2960248
49. Tweedie S, Ng HH, Barlow AL, Turner BM, Hendrich B and Bird A. Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat. Genet.* 1999; 23, 389–390 <https://doi.org/10.1038/70490> PMID: 10581020
50. Lupas AN, Ponting CP, Russell RB. On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? *J Struct Biol.* 2001; 134:191–203. <https://doi.org/10.1006/jsbi.2001.4393> PMID: 11551179
51. Schaefer M, Lyko F. Lack of evidence for DNA methylation of Invader4 retroelements in *Drosophila* and implications for DNMT2-mediated epigenetic regulation. *Nat Genet.* 2009; 42: 920–921.
52. Lerat E, Bulet N, Biémont C, Vieira C. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene.* 2011; 473:100–9. <https://doi.org/10.1016/j.gene.2010.11.009> PMID: 21156200
53. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007; 450:203–18. <https://doi.org/10.1038/nature06341> PMID: 17994087
54. Rohde C, Valente VLS. Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. *Genet Mol Biol.* 2012; 35: 966–79. PMID: 23411997

55. Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G. Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet.* 2009; 41: 696–702. <https://doi.org/10.1038/ng.360> PMID: 19412177
56. Durdevic Z, Hanna K, Gold B, Pollex T, Cherry S, Lyko F, et al. Efficient RNA virus control in *Drosophila* requires the RNA methyltransferase DNMT2. *EMBO Rep.* 2013; 14: 269–275. <https://doi.org/10.1038/embor.2013.3> PMID: 23370384
57. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, et al. RNA methylation by DNMT2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* 2010; 24: 1590–1595. <https://doi.org/10.1101/gad.586710> PMID: 20679393
58. Thiagarajan D, Dev RR, Khosla S. The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics.* 2011; 6: 103–113. <https://doi.org/10.4161/epl.6.1.13418> PMID: 20864816
59. Kaiser S, Jurkowski TP, Kellner S, Schneider D, Helm M, Kaiser S, et al. The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biol.* Taylor & Francis; 2016; 0: 1–11.
60. Da Cunha AB, Buria H, Dobzhansky T. Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution.* 1950; 4:212–235.
61. Ayala FJ and Powell JR. Allozymes as diagnostic characters of sibling species of *Drosophila*. *Proc Natl Acad Sci USA.* 1972; 69:1094–1096. PMID: 4504321
62. Vicario S, Moriyama EN, Powell JR. Codon usage in twelve species of *Drosophila*. *BMC Evol Biol.* 2007; 7:226. <https://doi.org/10.1186/1471-2148-7-226> PMID: 18005411
63. Heger A, Ponting CP: Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Research.* 2007; 17(12):1837–1849. <https://doi.org/10.1101/gr.6249707> PMID: 17989258

S1 Table. DNMT2 crystallographic models deposited in the database Protein Data Bank used to analyze and select the template structure for homology modeling process and evolutionary studies.

DNMT2	<i>D. melanogaster</i> vs. crystal sequence Identity (%)	<i>G. sulfurreducens</i> vs. crystal sequence Identity (%)	<i>M. musculus</i> vs. crystal sequence Identity (%)	PDB ID
SfDNMT2 (<i>S. frugiperda</i>)*	43.0	26.9	44.1	4H0N
Human DNMT2*	41.0	35.1	82.8	1G55
EhMeth (<i>E. histolytica</i>)*	30.0	24.9	34.6	3QV2
HhaI (<i>H. haemolyticus</i>)*	24.0	19.7	21.7	1MHT
HaeIII (<i>H. influenzae</i>)*	25.0	22.0	20.4	1DCT
M.MpeI (<i>M. penetrans</i>)	22.0	21.3	16.0	4DKJ

* Structures used in the analyzes

S2 Table. Percentage (%) of protein residues in each region of the Ramachandran Plot and additional evaluation tools for model viability assessment.

Species	PROCHECK				Qmean6	Verify 3D	
	Most favored	Additional allowed	Generously allowed	Disallowed	Z-score	Residues Averaged 3D-1D > 0.2a	Results
<i>D. albomicans</i>	91.8	7.5	0.3	0.3	-1.80	87.25	Aproved
<i>D. ananassae</i>	91.1	7.8	0.7	0.4	-1.89	92.51	Aproved
<i>D. biarmipes</i>	90.9	7.7	0.7	0.7	-1.34	89.15	Aproved
<i>D. bipectinata</i>	86.9	11.3	0.7	1.1	-1.93	97.02	Aproved
<i>D. buzzatii</i>	90.9	7.3	1.0	0.7	-2.07	86.98	Aproved
<i>D. elegans</i>	90.3	8.3	1.0	0.3	-2.10	88.01	Aproved
<i>D. erecta</i>	91.8	7.2	0.3	0.7	-1.48	89.31	Aproved
<i>D. eugracilis</i>	91.0	7.6	1.0	0.3	-1.33	89.47	Aproved
<i>D. ficusphila</i>	90.6	8.4	0.7	0.3	-1.83	89.11	Aproved

<i>D. grimshawi</i>	90.3	8.3	0.0	1.4	-1.79	91.76	Aproved
<i>D. kikkawai</i>	89.5	9.1	1.0	0.3	-2.07	93.33	Aproved
<i>D. melanogaster</i>	90.6	8.7	0.0	0.7	-1.39	91.01	Aproved
<i>D. miranda</i>	89.1	9.2	1.0	0.7	-1.84	90.43	Aproved
<i>D. mojavensis</i>	91.7	7.6	0.3	0.3	-2.01	89.57	Aproved
<i>D. persimilis</i>	90.2	8.1	1.0	0.7	-1.71	90.20	Aproved
<i>D. pseudoobscura</i>	90.5	7.8	1.0	0.7	-1.67	93.08	Aproved
<i>D. rhopaloa</i>	89.5	9.2	1.0	0.3	-2.00	94.51	Aproved
<i>D. sechellia</i>	90.7	7.6	1.0	0.7	-2.01	90.70	Aproved
<i>D. simulans</i>	91.0	7.6	0.7	0.7	-1.26	89.28	Aproved
<i>D. suzukii</i>	90.2	7.8	1.0	1.0	-1.71	93.73	Aproved
<i>D. takahashii</i>	90.7	7.6	1.0	0.7	-1.93	92.44	Aproved
<i>D. virilis</i>	90.0	9.0	0.7	0.3	-2.09	92.20	Aproved
<i>D. willistoni</i>	88.6	10.7	0.7	0.0	-1.89	90.32	Aproved
<i>D. yakuba</i>	91.8	7.2	0.3	0.7	-1.20	89.28	Aproved
<i>M. musculus</i>	92.0	6.6	1.0	0.3	-0.37	92.54	Aproved
<i>G. sulfurreducens</i>	90.2	8.6	0.8	0.4	-1.20	92.79	Aproved
<i>S. frugiperda</i> (4H0N)	91.5	8.5	0.0	0.0	-1.06	98.19	Aproved

^a Values given in %;

CAPÍTULO V

Dissecting the Dynamics between the DNA Methyltransferase 2 (DNMT2) and the DNA between Prokaryotes and Eukaryotes

Gilberto Cavalheiro Vieira^{1,2,3}, Marialva Sinigaglia⁵, Vera Lúcia da Silva Valente^{1,2,3},
Gustavo Fioravanti Vieira^{1,3,4,6}

¹ Programa de Pós Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

² Programa de Pós Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

³ Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

⁴ Núcleo de Bioinformática do Laboratório de Imunogenética, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

⁵ Instituto do Câncer Infantil, Porto Alegre, RS, Brazil.

⁶ Programa de Pós Graduação em Saúde e Desenvolvimento Humano, Universidade La Salle, Canoas, RS, Brazil

Manuscrito em preparação para ser submetido ao periódico Scientific Reports

Abstract

DNMT2 enzyme is the most conserved and widely presents methyltransferase among the various branches of life, but its biological function is still unknown. In prokaryotes, DNMT2 has a clear biological role acting as a DNA methyltransferase. However, in eukaryotes several studies described the involvement of DNMT2 in different biological functions, acting on DNA and/or RNA. The DNMT2 catalytic process is well studied and described, however, its substrate recognition mechanisms still poorly understood, and only some studies have mapped the crucial residues involved in the ligand binding. Thus, the aim of the present study was to understand the peculiarities and mechanisms involved in target sequence recognition and the affinity of the different DNMT2, when complexed with DNA. Here we analyze the DNMT2 of six different species by molecular dynamics simulation. As results we establish an interaction-model, confirming the crucial residues involved in the recognition and substrate interaction by DNMT2, and we also identified new key-residues in protein-substrate interaction.

Keywords

Dnmt2; epigenetics; DNA/tRNA methyltransferase; molecular dynamics; ligand-protein interactions

Introduction

DNA methylation is one of the most well-studied DNA modifications involved in gene regulation. This epigenetic mechanism is essential for development, genome defense, genome imprinting, chromatin stability and elements transposable silencing. The enzyme responsible for cytosine methylation is the DNA-(cytosine-C5)-methyltransferase (DCMtase), classified into three subfamilies: DNMT1, DNMT2 and DNMT3a, and 3b. The DNMT1 is the major enzyme in maintenance of the pattern of DNA methylation after DNA replication, whereas DNMT3 (-A, -B and 3L) proteins are primarily responsible for the establishment of *de novo* genomic DNA methylation patterns and, as such, play an important role in human developmental, reproductive, and mental health (Chédin 2011). The DNMT2 does not participate in genome methylation in mammals but seems to be an RNA methyltransferase.

DNMT enzyme family is divided into an N-terminal (regulatory region) and C-terminal, which it shares ten conserved residues motifs, where the catalytic domain, target-recognizing domain (TRD), and the coenzyme (S-adenosyl-l-methionine – SAM) binding site are found. DNMT2 does not have an N-terminal regulatory region, which is present in the other DNMTs (Yoder and Bestor 1998; Hermann et al. 2004). It is the smallest methyltransferase (approximately 400 amino acids) and the most conserved among animals, fungi, protists, and plants (Jeltsch 2002; Ponger and Li 2005) and its biological function has not been clarified yet. Furthermore, some organisms like *Drosophila*, *Entamoeba* and *Dictyostelium* seem to lose the DNMT1 and DNMT3a/b during its evolution, and the only DCMtase found is the DNMT2, so-called “*Dnmt2*-only” organisms. With respect to their biological function, studies associated DNMT2 with RNA interference in *Dictyostelium* and covalent histones modification, longevity in *Drosophila* and DNA activity (Kunert et al. 2003; Kuhlmann et al. 2005; Lin et al. 2005; Katoh et al. 2006), also acting in C5-methylation of C38 in the tRNA anticodon-loop to protect tRNAs from cleavage (Schaefer 2018).

Basically, the reaction mechanism of the DNMTs is the same among the different families. The catalytic process is mediated by a nucleophilic attack on the target cytosine carbon-6. This attack is promoted by the catalytic cysteine present in the conserved motif PCQ in the DCMtase active site (Santi et al. 1983; Wu and Santi 1987). From the protonation of the target cytosine in the N3 site mediated by the glutamic acid of the ENV

motif, covalent bonding occurs that exposes the C5 position of cytosine to the nucleophilic attack, where the methyl group of the co-enzyme SAM will be transferred to the carbon-5. The intermediate state of covalent binding between DNMTase in complex with DNA is widely described (Klimasauskas et al. 1994a; Reinisch et al. 1995), highlighting the fundamental role of the PCQ motif, more specifically the cysteine catalysis (Everett 1990; Chen et al. 1991; Hanck et al. 1993). In prokaryotic DNMTase, this same motif has been demonstrated crucial to DNMT2 activity (Wyszynski et al. 1991; Wyszynski et al. 1993; Hurd et al. 1999). However, this event is only possible after the so-called base flipping process, in which the target cytosine is withdrawn from the double-helix context and positioned close to the coenzyme binding site. For this, there are several crucial residues for the stabilization of the transient structure, such as PCQ and ENV motifs (Klimasauskas et al. 1994a; Reinisch et al. 1995) (**Figure 1**).

If on the one hand the mechanisms and catalytic processes of DNMT2 are widely described and known, on the other hand, the reasons why eukaryotic DNMT2 have low DNA activity are still unknown, despite the structural conservation of prokaryotic and eukaryotic catalytic domains. Some “hints” as to the reason for this affinity difference can be obtained by analyzing the tertiary structures of the TRD of different DNMT2. For example, HhaI and HaeIII (prokaryotic DNMT2 with DNA-MTase activity) are known to exhibit structural conservation in their catalytic domains but differ in TRD structure and amino-acid composition, and still acting on DNA as substrate. Nevertheless, they eventually have different methylation target sequences from each other, where HhaI recognizes 5'-GCGC-3', while HaeIII 5'-GGCC-3' sites (Klimasauskas et al. 1994b; O’Gara et al. 1996b). When comparing DNMT2 tertiary structures and the distribution of molecular electrostatic surface in different DNMT2, marked differences emerge within latter aspect, mainly in TRD region (Vieira et al. 2017), even in evolutionarily close species, as *D. melanogaster* that has preference for DNA methylation in CT, CA and CC context (Kunert et al. 2003) and *D. willistoni*, that seems to methylate CG sites (Garcia et al. 2007) - the same context preference found in human DNMT2 (Hermann et al. 2003).

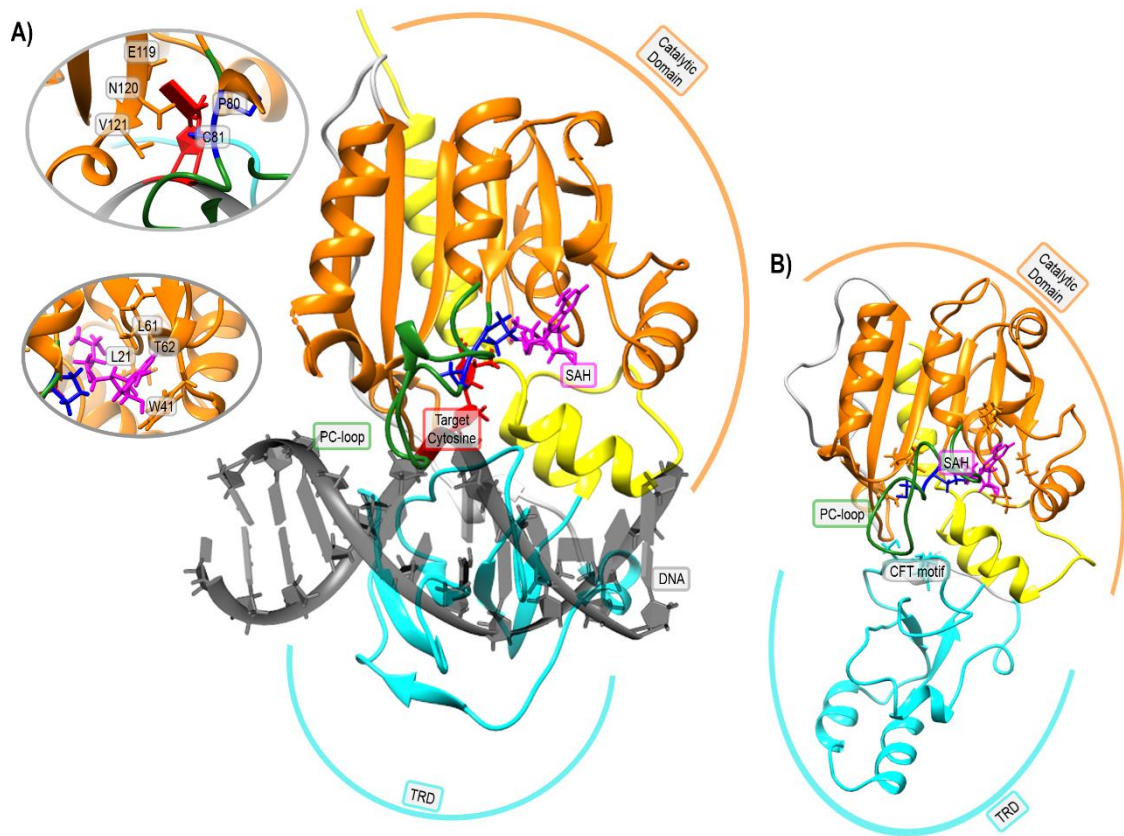


Figure 1. DNMT2 structures. (A) HhaI (PDB: 1MHT), and a structural model of (B) human DNMT2. The enzyme elements are colored as follow: N-terminal catalytic domain (orange) and C-terminal catalytic domain (yellow); PC-loop in green, Proline and catalytic Cysteine in blue; TRD in cyan; S-adenosylhomocysteine (SAH) in magenta; target-cytosine in red, and DNA in gray. Key-residues from catalytic domain showed as sticks and labeled. Structural conservation of the catalytic domain between prokaryotes and eukaryotes can be noted, whereas TRD varies substantially.

It is interesting to note that according to previous studies, Ehmeth (*Entamoeba histolytica* DNMT2) and human MT2 (also Known as TRDMT1) presented weak activity on DNA substrate and shown preference to RNA substrate, suggesting that in some species rose an alternative role to DNMT2: the tRNA methylation (Hermann et al. 2003; Fisher et al. 2004). More recently, it was found that DNMT2 can methylates more efficiently DNA fragments in a covalent DNA-RNA hybrids context (Kaiser et al. 2016). Systematic mutations in surface-exposed lysines and arginines to alanine were able to

elucidate some key-residues for the catalytic activity of the DNMT2 in the tRNA methylation (Jurkowski et al. 2012).

The function performed by an enzyme underlies its structure, electrostatic potential surface, and dynamics. As described previously, the structure of the catalytic domain of DNMT2 is very well known and described, but with respect to TRD, its aspects are still neglected. Since diverse crystallographic structures of different DNMT2 have been resolved, like HhaI (PDB: 1MHT), HaeIII (PDB: 1DCT), *Spodoptera frugiperda* (PDB: 4H0N) and human MT2 (PDB: 1G55), the elaboration of predicted structural models of other species has become possible with great accuracy through homology modeling. However, studies focusing on molecular structures fail to take into account the subtle aspects of the dynamics between the different enzymes, especially in the enzyme-substrate context.

Molecular dynamics (MD) simulations can provide a series of detailed information about a protein's behavior, its motions, conformation modifications and interactions with substrates. This information from the MD is supported by the laws of thermodynamics, which offers reliability in the results obtained when adequate force field and simulation time are provided and can be compared to experimental data. Thus, the present study aims to look more closely at the interactions that emerge within the DNMT2-DNA complex of several DNMT2, through analyses of protein-ligand all-atoms from MD simulations, seeking to shed light on the mechanisms involved in target substrate recognition and how the enzyme still has remnants of DNA methylation activity.

Material and Methods

Molecular modeling of DNMT2

The Modeller 9.14 (Marti-Renom et al. 2000) program was applied to model the DNMT2 from *D. melanogaster* and *D. willistoni* obtained by searching *in silico*, using the *Spodoptera frugiperda* (PDB 4H0N) (Sisi Li, Jiamu Du, Hui Yang, Juan Yin, Jianping Ding 2012) Ehmeth (PDB 3QV2) (Schulz et al. 2012) and human DNMT2 structures as templates (PDB 1G55) (Dong et al. 2001), according to the best alignment between the template and the target sequence. The alignment was further verified manually and

adjusted, considering the location of insertion/deletion in loops. The homology modeling was performed using Python scripts. The modeling protocol followed the default optimization and refinement protocol, as described in the Modeller online manual (available at <http://salilab.org/modeller/9.13/manual/node19.html>). We generated one hundred models, which the best model was selected using the DOPE score and Procheck (Laskowski et al. 1996). The generated models were evaluated by Ramachandran plot analysis, an overall model quality evaluation was performed using Qmean6 (Benkert et al. 2011) and Verify3D (Bowie JU, Luthey R 1991).

The DNA oligomer used in MD simulations is the same from the crystallographic structure of HhaI DNMT2 (PDB: 1MHT). It was decided to use the HhaI DNA structure due to the fact of the site methylation target be the sequence (5'-GCGC-3') is very similar to the settings found in *D. willistoni* (Garcia et al. 2007) and human MT2 target sites methylation and have a close contest to that found in the target regions of tRNA-Asp methylation, where the C38 (cytosine target) is inserted in a context 5'-PuCGC-3'. The MT2-DNA complex composition was made fitting the structures template (HhaI) and the target-model in UCSF Chimera program (Pettersen et al. 2004). Soon afterward, the structures of the target model and DNA were assembled and brief energy minimization was performed to adjust the side chains of the residues.

Molecular dynamics simulations

The crystals structures from *H. haemolyticus* (1MHT), *E. histolytica* (3QV2), *Spodoptera frugiperda* (4H0N), human (1G55), and the DNMT2 models from *D. melanogaster* (melDNMT2) and *D. willistoni* (wilDNMT2) were subjected to replicates 50ns of a MD simulation using the GROMACS v4.5.1 package (Pronk et al. 2013) on a Linux platform using the AMBER99SB force field. An appropriate number of sodium (Na⁺) and chloride (Cl⁻) counter-ions was added to neutralize the system. The cubic box was defined with at least 15 °A of solvation layer around the protein, using an SPC water model and periodic boundary conditions. The v-rescale (tau t = 0.1 ps) and Parrinello-Rahman (tau p = 2 ps) algorithms were used for temperature and pressure coupling, respectively. Cutoff values of 1.2 nm were used both for van der Waals and Coulomb interactions, with Fast Particle-Mesh Ewald electrostatics (PME).

The MD simulations were divided into four main stages: energy minimization (EM), solvation, thermalization, and production (Figueiredo et al. 2014). After, MD plots were generated with the programs from GROMACS v4.5.1 package (Van Der Spoel et al. 2005) and visualized with xmgrace (<http://plasma-gate.weizmann.ac.il/Grace/>). Visual inspection of the MD trajectories was performed with VMD 1.9.1 (Humphrey et al. 1996), PyMOL 1.0 (Delanno, 2002) and UCSF Chimera (Pettersen et al. 2004).

Generalized order parameters (S^2) and Gibbs free energy analysis

Protein dynamics make important but poorly understood contributions to molecular recognition phenomena. To understand the implications of thermodynamics that emerge from dynamic conformational changes, we access generalized order parameters (Lipari and Szabo 1982) and entropy values of apo and holo structures from the MD simulation. For this, we use the relations between Lipari-Szabo generalized order parameter (S^2) and the conformational entropies derived from Yang and Kay (Yang and Kay 1996; MacRaid et al. 2007). This relationship is essentially independent of motional model for $S_{LZ}^2 < 0.95$, so only residues consistent with this criterion were included in considerations of entropy change. Generalized order parameters are calculated from the MD trajectory of individual N-H bond vectors. RMS deviations, atomic fluctuations, and structure averaging were calculated directly, using programs from GROMACS v4.5.1 package (Van Der Spoel et al. 2005).

Free energy landscape (FEL) was done using the GROMACS utility *g_sham*. The FEL shows all possible conformations a molecule assumed during an MD simulation, together with their corresponding energy typically reported as the Gibbs Free Energy.

Results

DNMT2 Residues fluctuation from Molecular Dynamic Simulation

The differences in atomic fluctuations of the holo and apoenzyme found to indicate a loss of dynamics when it is linked to DNA, least in Ehmeth, which it presents subtle dynamics gain. However, the observed differences were not uniform along with the same enzyme (**Figure 2**).

Interesting, Ehmeth shown a gain of dynamics when linked to DNA. The biggest differences [Holo-Apo] were found in HhaI (-9.6376) and wilDNMT2 (-9.822), indicating the accentuated loss of dynamics in the structure when DNA is present (**Table S1**). Even if the structures and phylogenetic relationships are close, there are differences in atomic fluctuation between related MT2, like meIDNMT2 and wilDNMT2.

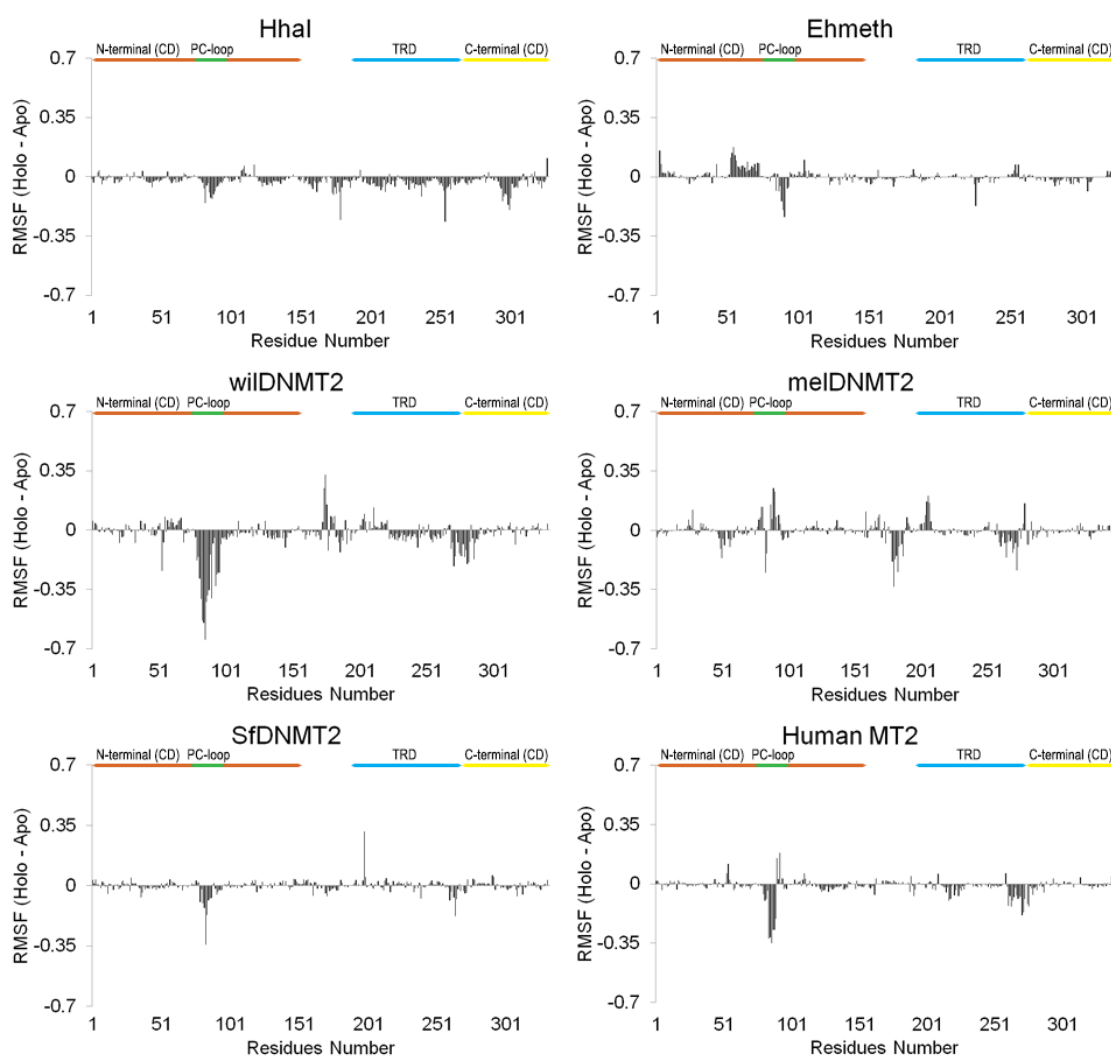


Figure 2. Differences in per-residue plots of atomic fluctuations (RMSF) between holo-protein and apo-protein (Δ RMSF=holo-apo). The main domains are represented above the RMSF values. In orange is represented the catalytic domain (CD), PC-loop in the green and target-recognizing domain (TRD) in cyan. In general, the values of Δ RMSF between the catalytic domain of prokaryotes and eukaryotes are similar, while in the TRD there are more differences between the fluctuations. It is observed that between

eukaryotes, melDNMT2 present different behavior to the pattern found in the other DNMT2.

There are two main regions in DNMT2 which show substantial variations when compared holo- and apoenzyme structures: the first is between residues 71-99, which correspond to PC-loop (with the catalytic cysteine); and the second above residues 260-290, that belong to TRD C-terminal region. The hinge loop connection of melDNMT2 (between residues 165-205, approximately) seems to present a lower fluctuation when in the presence of DNA, while wilDNMT2 presented gain of dynamics in the presence of substrate (**Figure 2**). Regarding the PC-loop region, between the eukaryotic DNMT2, wilDNMT2 presented the lowest dynamics in the holoenzyme.

Thermodynamic contribution measured by per-residue entropy

Through Lipari-Szabo generalized order parameters (S^2) we can measure the behavior of amide bond vectors by the MD simulations. Despite the structure conservation, the DNMT2s display different dynamic behavior. Except for the PC-loop, that shares the same behavior between all the different DNMT2, each DNMT2 present several peculiar residues which become more, or less, dynamics in the presence of DNA (**Figure 3**).

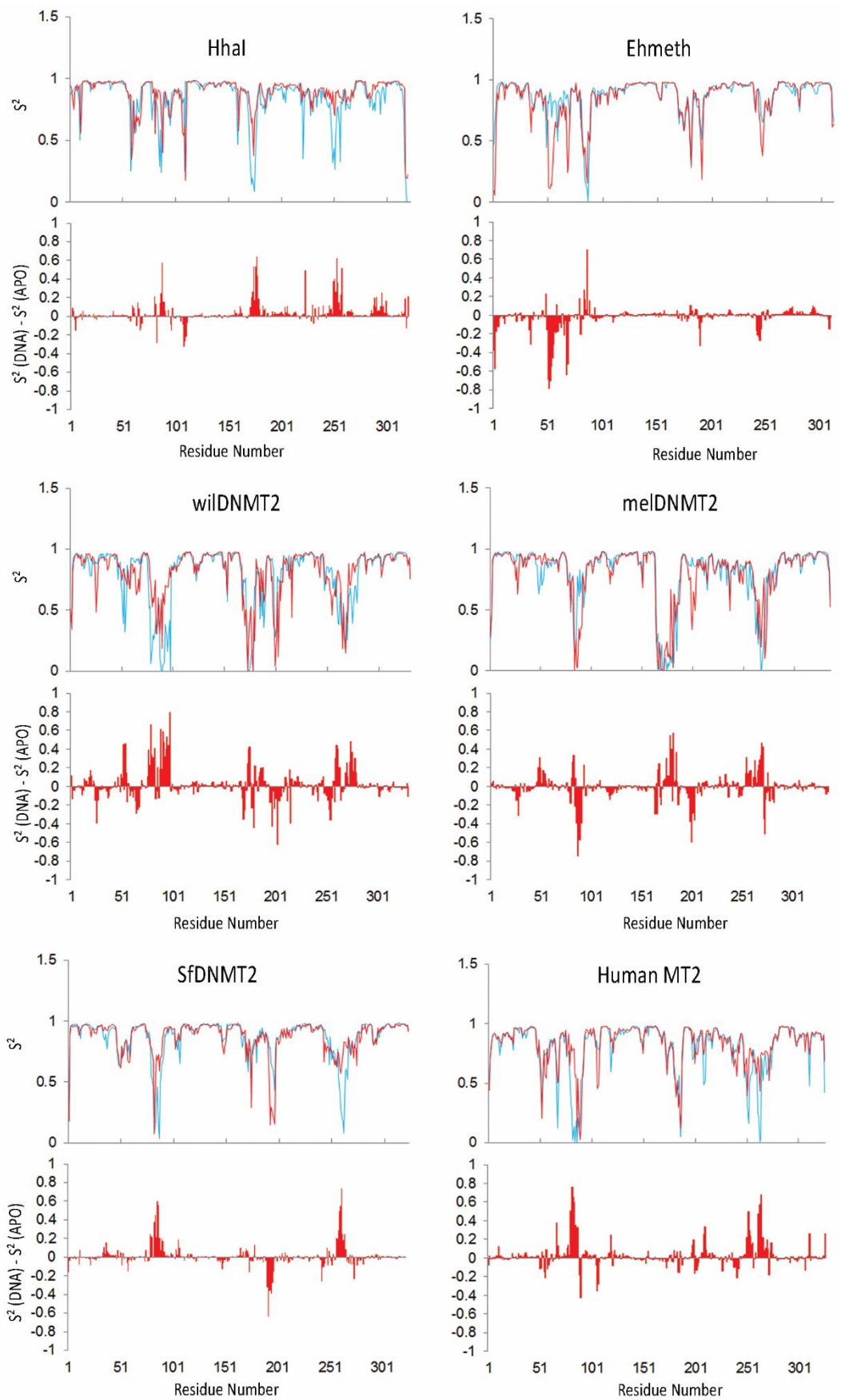


Figure 3. Generalized order parameters (S^2) and differences in S^2 parameters between apo-protein (blue) and holoprotein (red). Scheme presenting the main domains of DNMT2 and the residues that approximately delimit them presented below. If the bond is absolutely rigid, showing no fluctuation, then that value is 1 or close to 1. If the amide bond shows high fluctuation along the MD, then the ultimate value is 0 – or close too.

At TRD region, HhaI presents several residues with lost in dynamics when in the presence of its substrate (**Figure 3**). The melDNMT2 has the most anomalous behavior between the DNMT2 studied, showing a dynamic gain in regions that are normally more stable in other DNMT2, both in the catalytic domain and in the TRD (e.g. residues site 1-51 and 101-151, that below to PC-loop region) (**Figure 3 and 4A**). Ehmeth also presented a different dynamic pattern than the other DNMT2 here analyzed. The Ehmeth catalytic domain presented a remarkable gain of dynamics in the holoenzyme (in the region comprising the residues 50-70, the PC-loop), while in the other DNMT2 this region leans towards to have a loss of dynamics. In the Ehmeth TRD region, the dynamic behavior from apo and holoenzyme are very similar to each other (**Figure 3**).

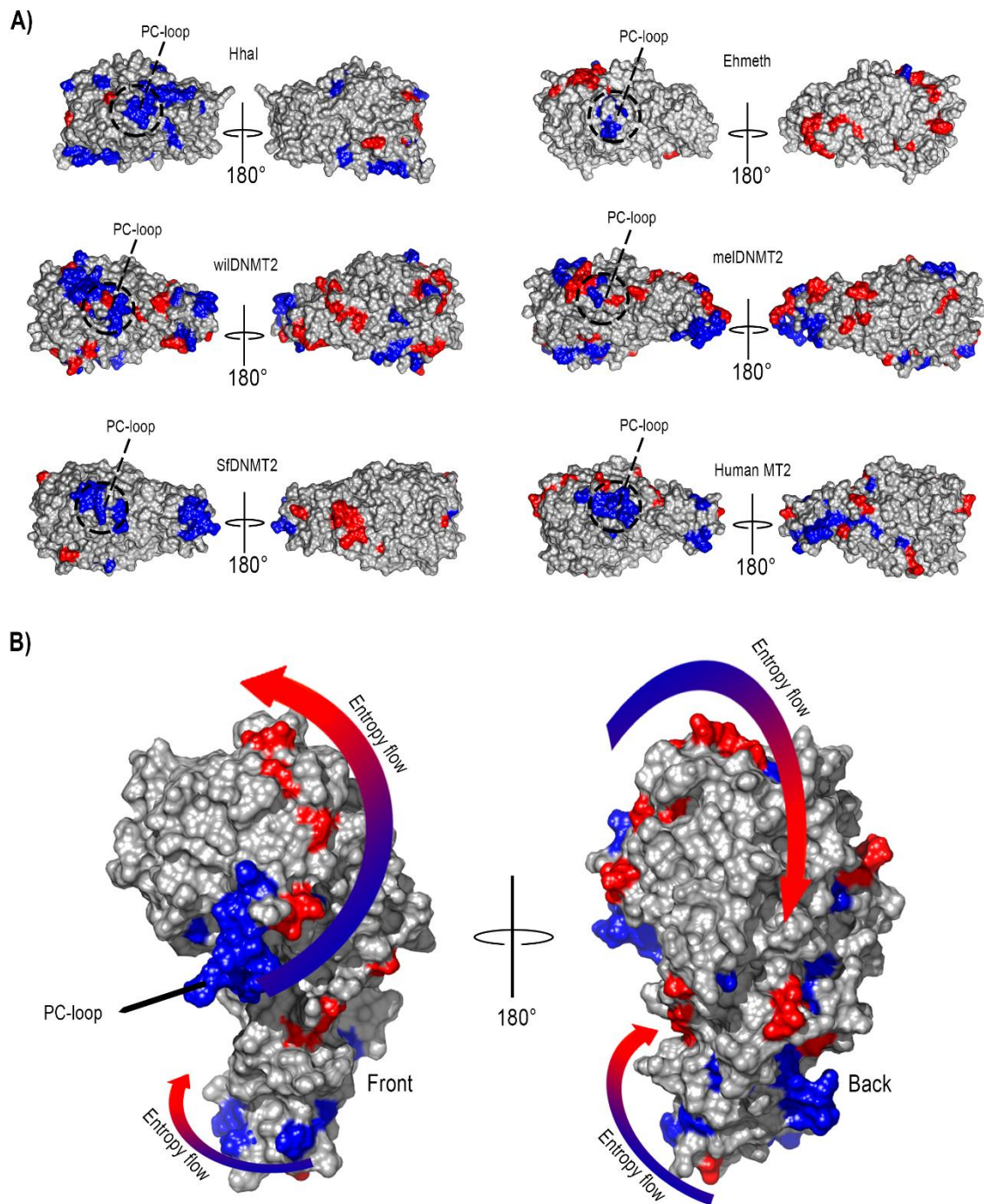


Figure 4. DNMT2 surface representation with residues annotation from differences in S^2 parameters between holo and apoenzyme, which $[S^2(\text{Holo}) - S^2(\text{Apo})] > 0.113$ (blue) or < -0.113 (red). (A) On the left side, a front view of the enzyme is displayed, showing the cleft where DNA bound. On the right side, a back view is displayed, obtained by a 180° rotation about the vertical axis. (B) Representation of the trend of entropy flow on the surface of DNMT2, in which the arrows indicate the direction (blue color, smaller entropy and red, greater entropy). Positive values indicate a lower fluctuation when DNA was present, whereas negative values indicate that the residues had a greater fluctuation.

On the other hand, wilDNMT2, SfDNMT2, and human DNMT2 showed somehow similarities in the dynamics. Clearly, we can distinguish three main regions which show accentuated entropy, approximately from residues (i) 75-100, (ii) 160-200 and (iii) 251-270 (**Figure 3**). The first region is located in the catalytic domain of DNMT2, mainly comprising the PC-loop. In this region, wilDNMT2 presents a loss of dynamics, while SfDNMT2 and human DNMT2 also present a more rigid behavior in the presence of DNA, but not so pronounced (**Figure 3** and **4A**). The second region refers to the hinge-loop that links the catalytic domain and TRD. SfDNMT2 presented the smallest difference between apo and holoenzyme. We do not take into account human DNMT2 in this region because it has a large insertion with no resolution crystallographic, so this region was excluded in the molecular modeling.

In order, the third region refers to the TRD core, comprising the conserved residues: CFTxxYxxY/I (where "x" may be any other residue), E/DGTGS and DIV/I/C (residues 230 to 260). The five DNMT2 (melDNMT2, wilDNMT2, SfDNMT2, and human MT2) showed a heterogeneous behavior upon DNA binding. SfDNMT2 and human presented the greatest difference between apo and holoprotein states (**Figure 3**). In general, there seems to be a system of compensation and distribution of entropy through the protein structure, in which residues facing the protein-substrate interface (front) lose entropy when in the presence of the ligand, whereas in the "back" of the enzyme there is an increased entropy (**Figure 4B**).

Through Free Energy Landscape (FEL) analysis we can determine the conformational pattern which it is more favorable of each DNMT2 from apo and holoenzyme states, based on the Gibbs free energy. Basically, conformers with lower Gibbs free energy favor a folded state, and an increase in the Gibbs free energy favor unfolded state in the FEL analysis. We also calculated the variance of the ΔG values obtained through the FEL, since the variance measures the deviations from the mean, that is, the difference between each data and the mean and evaluates the degree of dispersion of a dataset.

HhaI showed more states by the simulation with lower Gibbs free energy in the presence of DNA, while in the apo-protein populated a larger area in the FEL landscape (**Figure 5** and **S1**). All eukaryotic DNMT2, when associated with the DNA, presented a

wide occupation of the landscape with the exception of wilDNMT2, and human MT2. The differences are observed more easily when we plot the ΔG values (**Figure 6**). We can note that HhaI and wilDNMT2 shown a similar pattern of dispersion of Gibbs free energy values.

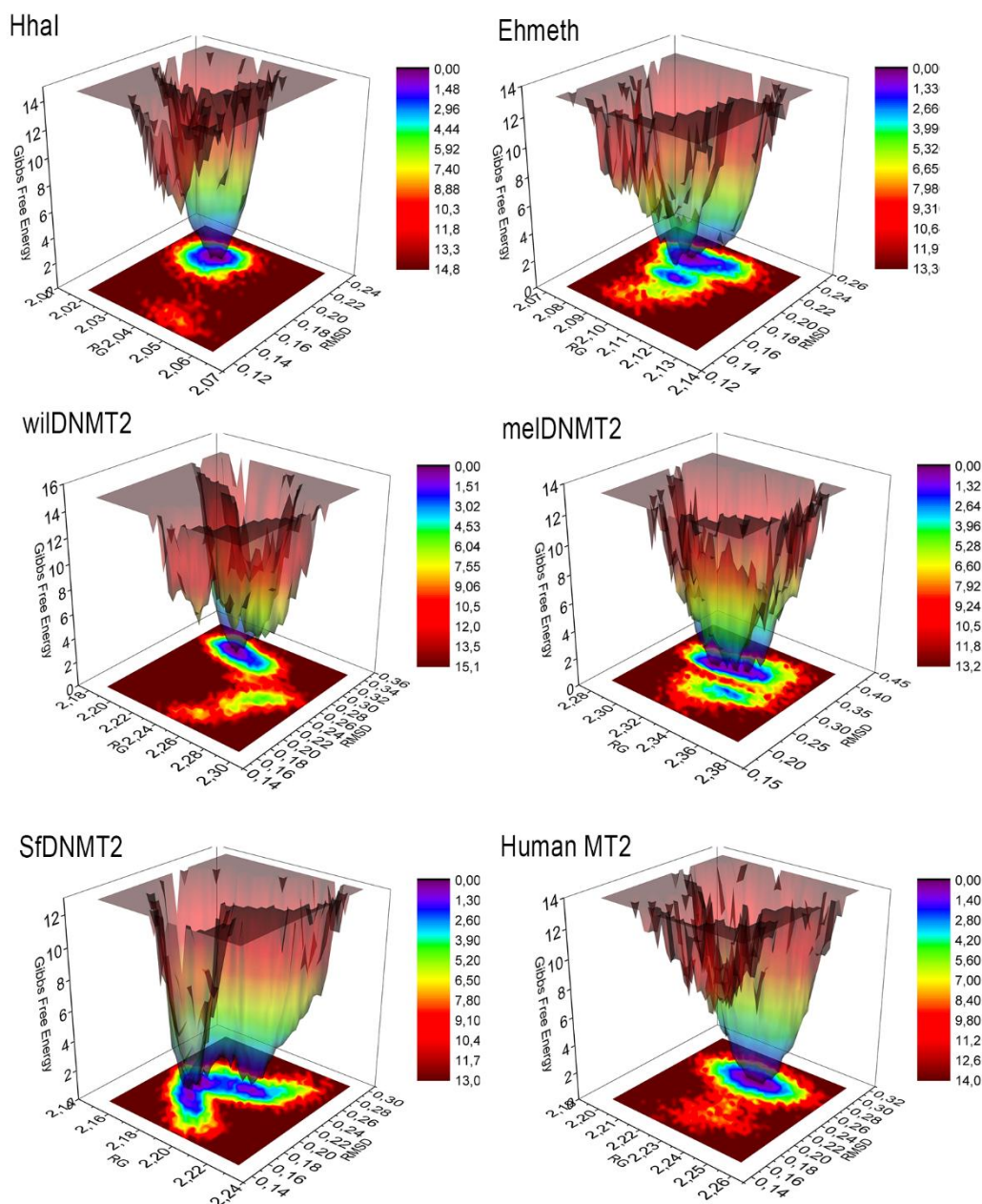


Figure 5. 3D free energy landscape of complex DNMT2-DNA as a function of the radius of gyration (RG) and the root-mean-square deviation (RMSD). All enzymes also were analyzed in apo state (**Figure S1**). The free energy values are shown next to the graphs. HhaI seems to populate a state able to accommodate DNA with great affinity, and the same can be observed in wilDNMT2 and Human MT2.

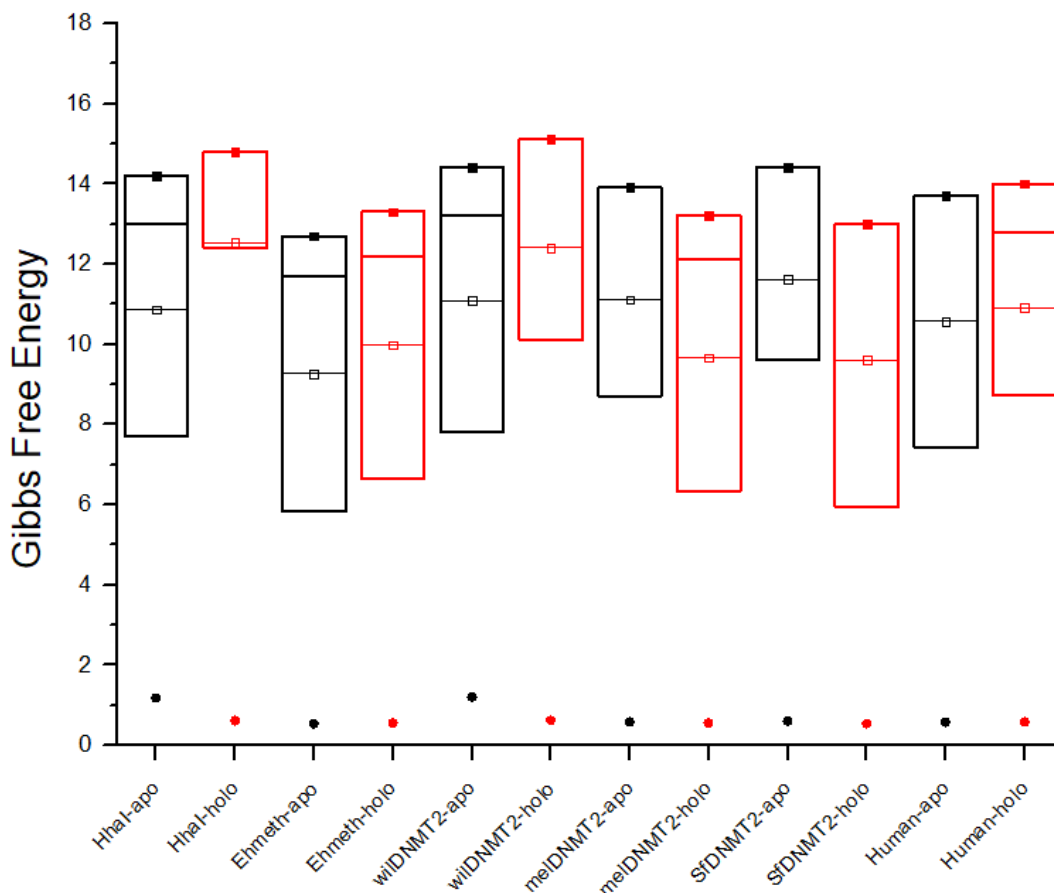


Figure 6. Gibbs free energy plot from apo- and holoenzymes along molecular dynamic simulations. Dots show the 1% minimum occupancy values. Filled squares, the maximum values. Empty squares represent averages values. The boxes denote the 25-75% occupancy values.

The H-bonds interactions in the DNMT2-DNA complex

We asked whether the distributions of different molecular potential energy surface of DNMT2 of this study are linked to interactions with the substrate. Thus, we decided to evaluate the behavior of molecules in an aqueous medium by molecular dynamics simulation. HhaI DNMT2 showed an average of H-bonds 33.51. The human MT2 have fewer H-bonds with the substrate over the dynamic simulations (H-bonds average 13.45), while Ehmeth appears then with an H-bond average of 16.98 (**Figure 7**). Interestingly,

wilDNMT2 has a greater number and stable pattern of H-bonds throughout the dynamic simulations (**Figure 7 and 8**).

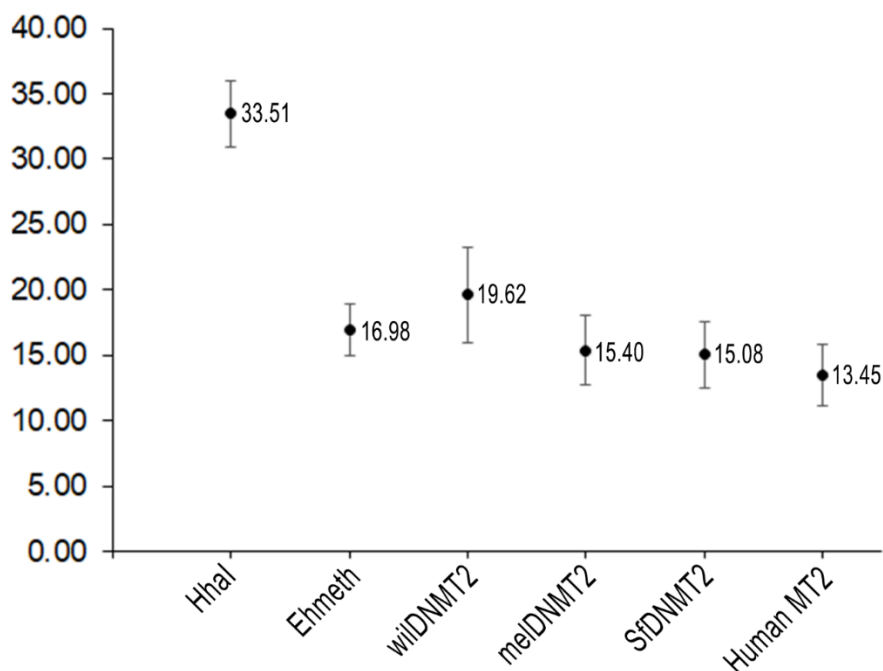


Figure 7. Average H-bonds between protein-DNA complex from DNMT2 over the molecular dynamics simulations. HhaI acts preferentially on DNA and present averages of H-bonds detachably larger than the other DNMT2, as expected. Among eukaryotic DNMT2, the human enzyme has the lowest average H-bonds, while wilDNMT2 has the highest mean values.

From the snapshots obtained from different moments of the molecular dynamics simulations, it was possible to analyze the structural behavior of MT2 and DNA, as well as to observe the residues that act primarily in the protein-substrate interaction (**Figure 9A and 9B**). The residues described as having protein-DNA interaction in HhaI in previous studies are present among those that formed H-bonds with the substrate during the molecular dynamics simulations. However, during the MDS other residues, besides those already annotated, also showed interactions with the DNA (**Table S2**).

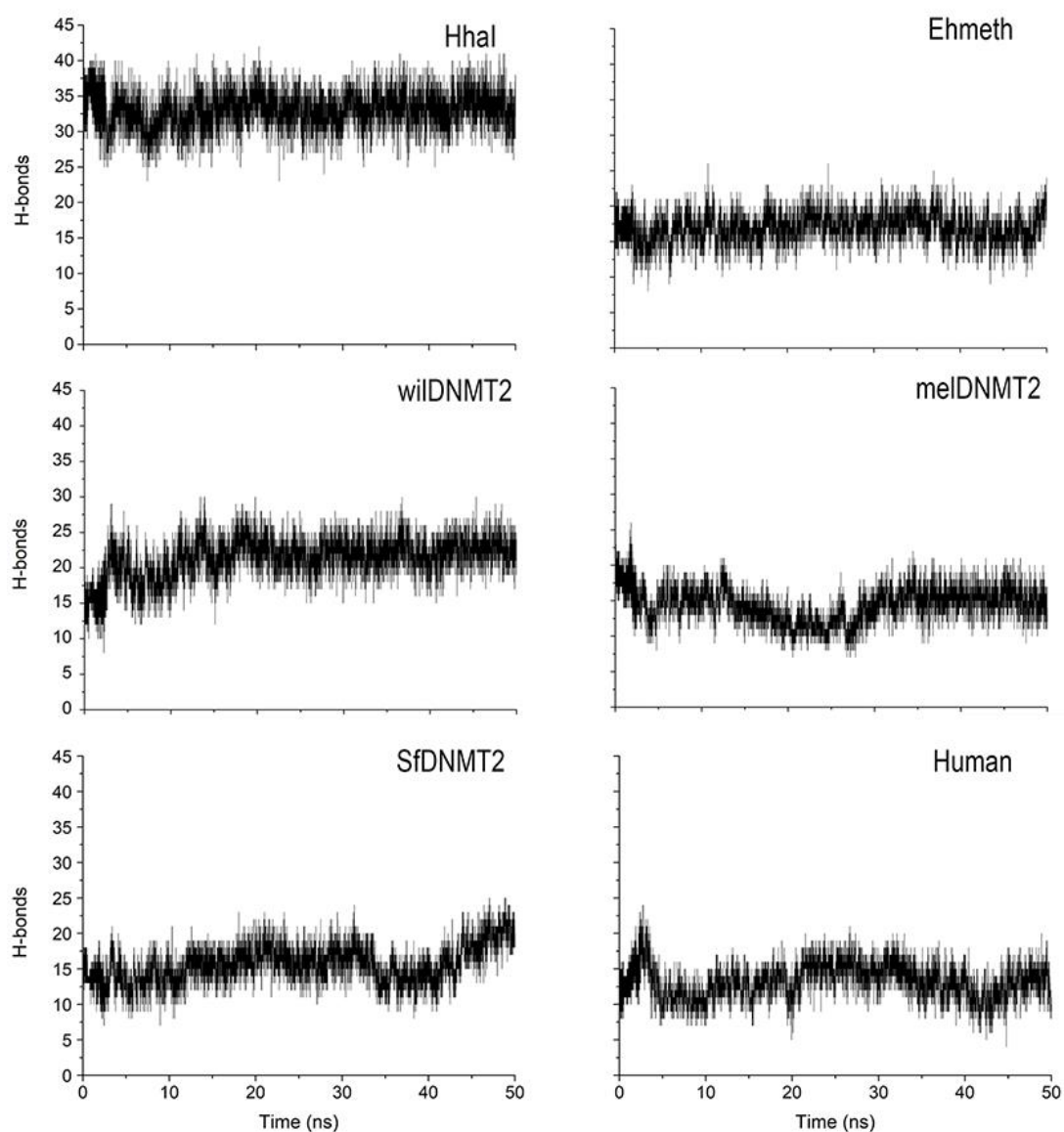


Figure 8. Number of H-bonds between protein and DNA of each simulated structure, throughout a 50 ns molecular dynamics simulation. It can be seen that the number of H-bonds of HhaI remains stable throughout the MD. Among the other DNMT2, only Ehmeth and wilDNMT2 maintain a more stable H-bond number, however, Ehmeth makes fewer connections than wilDNMT2.

The DNA molecule showed a more stable behavior linked to HhaI enzyme (**Figure 9A and S2**). In most eukaryotic MT2 there are a series of accentuated deformations and torsions in the DNA molecule during molecular dynamics (**Figure 9A**). Interestingly, wilDNMT2 appears to exhibit a behavior similar to that observed in HhaI,

with low DNA fluctuation, low dynamics, and maintenance of interactions mediated by more stable H-bonds along MD (Figure 8 and 9).

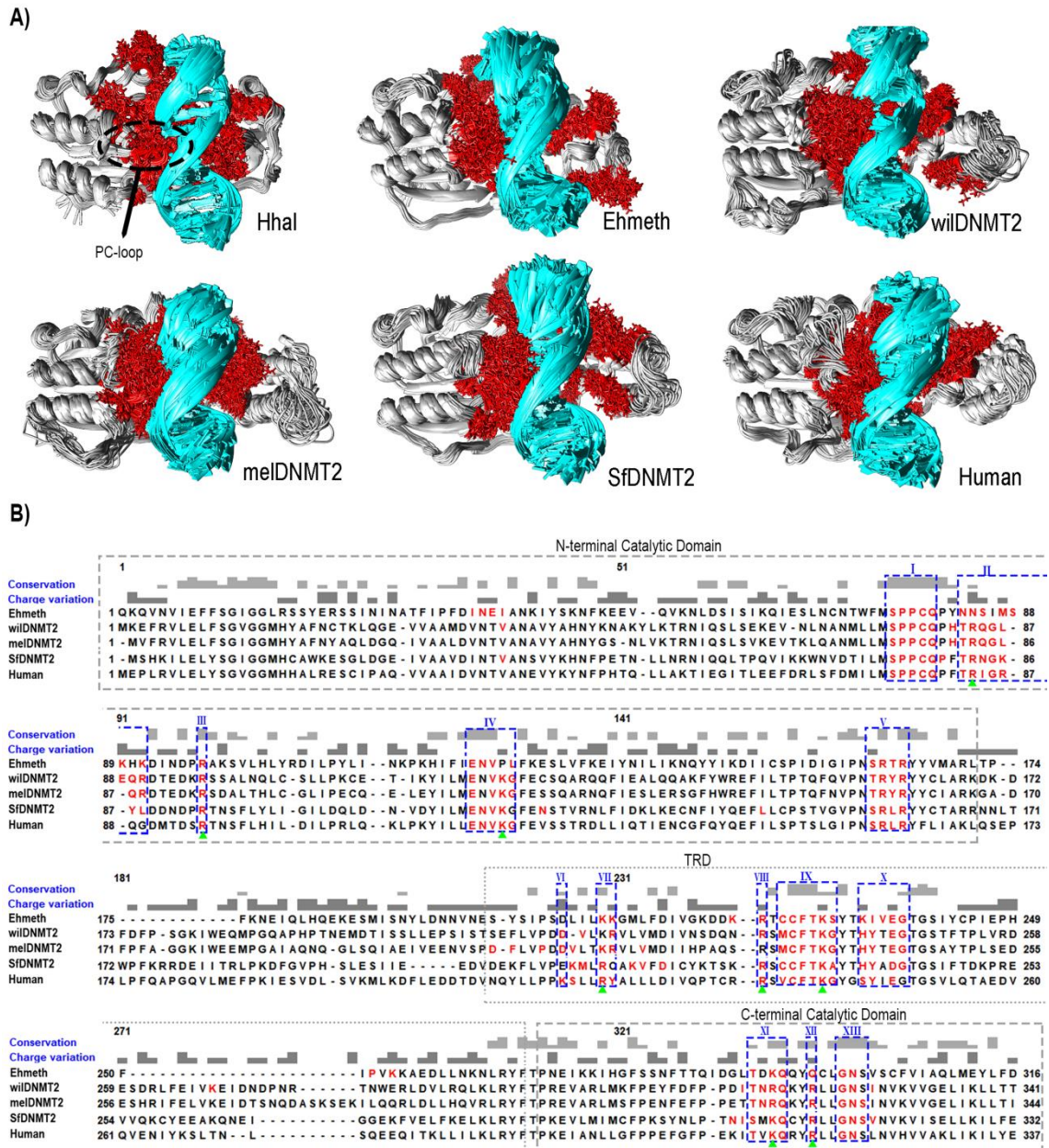


Figure 9. DNMT2-DNA complex representation and sequences. (A) Each DNMT2 is represented by several structures obtained at different molecular dynamics simulation times, from the time 5 to 50 ns. Residues with hydrogen bonds during MD simulation are colored red; a molecule of DNA in cyan. Residues were taken with hydrogen bonds that were within the parameters of 3.5 Å with an angle of 30°. (B) Eukaryotic DNMT2 alignment. The residues that showed H-bond interaction with DNA are highlighted in red. In blue, roman letters indicate the motifs that show conservation of interactions with the

substrate. Residues described as pivotal for human DNMT2 catalytic activity are marked with a green triangle.

Among eukaryotic enzymes, we detected thirteen motifs that showed common protein-DNA interactions (**Figure 9B**). Motifs I-V and XI-XIII belong to the catalytic domain of DNMT2, while motifs VI-X to TRD. Some motifs are characterized by high conservation of the residues and small charge variation (I, III, IV, VIII, IX and XIII), while other motifs end up with greater residues and charge variation (II, VI, VII, X, and XI). The first motifs group, for the most part, belong to the catalytic domain, while the second group belongs mostly to the TRD.

As for the behavior of apo and holoproteins, the enzymes HhaI, meDNMT2, wilDNMT2, and Human MT2 presented higher stability and lower RMSD values when complexed with DNA (**Figure 10**). When the RMSD values of the two major domains (catalytic and TRD) are analyzed separately, it is seen that MT2 work in a distinct modular manner between the domains.

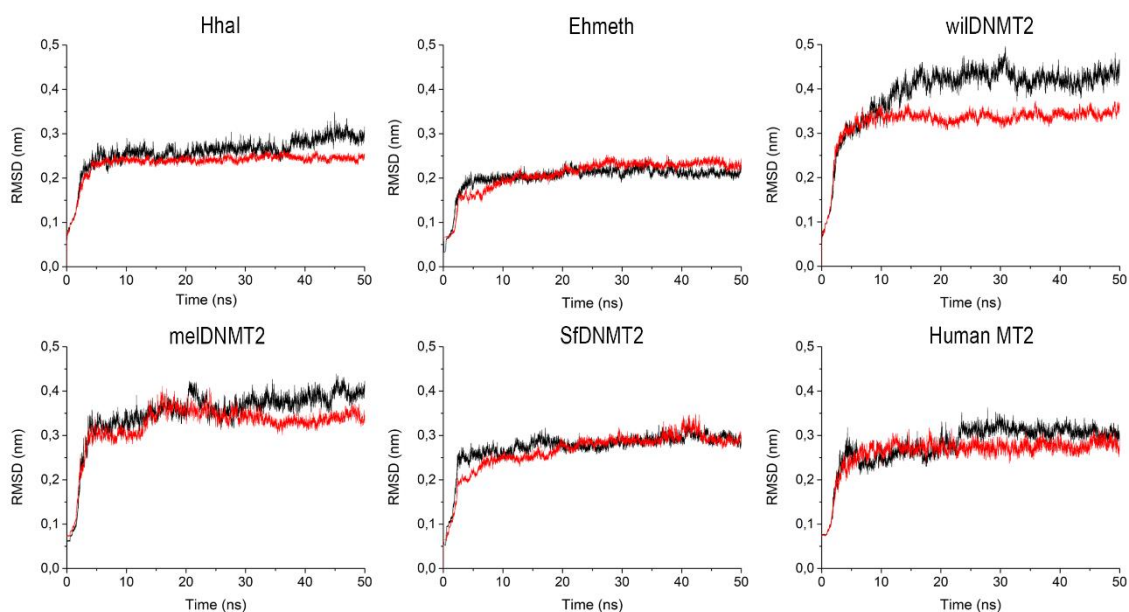


Figure 10. RMSD of the apo- (black) and holo-protein (red) over the 50 ns molecular dynamics simulations at 300 K. All structures stabilize from the 10 ns of simulation. The fluctuations found during the simulations of apoproteins refer mainly to the mobility of loops that make contact with the substrate when it is present.

Discussion

The key mechanisms of recognition of target sequences and the catalytic process of the DNMT2 family are well described from studies with HhaI and HaeIII enzymes. However, the prokaryotic DNMT2 differ substantially from eukaryotic DNMT2 regarding substrate preference, while HhaI and HaeIII are DNA methyltransferases (Klimasauskas et al. 1994c; Reinisch et al. 1995), the eukaryotic DNMT2 seem to have lost affinity for this substrate, turning to the action on RNA molecules (Kuhlmann et al. 2005; Durdevic et al. 2013a; Müller et al. 2013; Kaiser et al. 2016). Still, enshrined the differences of targeted substrates, these two groups of DNMT2 have deep conservation in their catalytic domains (**Figure 1**), suggesting similarities in mechanisms involved in the methylation at cytosine-target, an event known as base-flipping (Chen et al. 1991; Klimasauskas et al. 1994a; Reinisch et al. 1995). In order to better understand how the process of this change of substrate affinity was promoted, here we present an extensive analysis by means of MD simulation of five eukaryotic DNMT2, comparing to DNMT2 HhaI.

Since conformational changes are likely to play a fundamental role in regulating the biological activity of DNMTs, the substrate specificity cannot be restricted from the conformational changes through MD simulations (Evans and Bronowska 2010). So, other factors may be involved in the mechanisms inherent in the recognition of the sequence and/or target substrate of DNMT2s. We cannot deny the importance of structure-related factors such as amino acid substitution at key positions that may interfere with substrate binding affinity; so, the dynamics of the enzyme may be closely related to the specificity with a given substrate.

Although the DNMT2 possess strong structural conservation in their catalytic domains, prokaryotic and eukaryotic enzymes present great variability in the TRD residues composition and architecture (Jeltsch 2002; Ponger and Li 2005; Jeltsch et al. 2016; Vieira et al. 2017). This aspect is reflected in the global dynamics of each of the proteins studied here. It is clear that there is a pattern in certain motifs, especially those related to catalytic motifs. This same pattern can be observed in prokaryotic and eukaryotic enzymes. However, eukaryotic DNMT2 exhibit distinct behavior compared to HhaI. Even among evolutionarily closely related species differences can be noted (i.e. wildDNMT2 and melDNMT2). The dynamics differences found in the present study

between *D.willistoni* and *D.melanogaster* are consistent with our findings published in previous work, where the difference in the distribution of the molecular electrical potential surfaces between both species differ substantially (Vieira et al. 2017).

Differences in the composition of the residues induced changes in the flexibility of the proteins, indicating that it may affect the function of the proteins by altering enzyme-DNA interactions (Elengoe et al. 2014; Nagasundaram et al. 2015). When compared to the tertiary structures of the different DNMT2 TRDs, we can observe the fact that the prokaryotic enzymes exhibit a low structural organization, being formed basically of loops and most of the interactions with the target DNA sequence being mediated by these several loops, whereas, the eukaryotic enzymes have a structural organization more elaborated. This modification in the tertiary structure directly impacts the enzyme interactions with the substrate, as we can observe (**Figure 7, 8 and 9**).

Concerning the thermodynamics involving the studied DNMT2, it is evident that a loss of entropy is expected in the formation of a stable complex between protein and substrate, associated with the free diffusion of one component with respect to the other (**Figure 3 and 4**). Conformational entropy is considering an important component of the free energy of binding, generally, the conformational entropy is linearly correlated with total binding entropy (Marlow et al. 2010). As expected, HhaI showed a large entropy loss when the DNA was present, while the entropies losses were not so pronounced in the DNMT2 eukaryotic. These results indicate an enthalpy-entropy compensation network intrinsic to each DNMT2 and that directly impacts the free energy of the system and its variation between the apo and holoenzyme states (**Figure 4, 5 and 6**).

Differences in the structure, surface electrostatic potentials, and thermodynamic properties can influence the waters surrounding the protein and bound ligands, being an important contributor to the observed enthalpy-entropy compensation, and this is closely related to the hydrophilic and hydrophobic interactions of the physicochemical profile of residues that compose the protein and ligand (Swaminathan et al. 2002; Fiscaro et al. 2004; Breiten et al. 2013). Assuming that the electrostatic interactions have an influence on binding free energy, and since H-bonds are essentially electrostatic interaction, we investigated the H-bonds between enzyme and substrate.

Most DNA contacts are realized into the major groove in the protein-substrate interface. HhaI have a greater number of interacting residues, which are arranged in their TRD, which is composed mainly for loops (**Figure 1, 9A and 9B**). In addition to the residues described in previous works, new residues are presented here as important in the stabilization of the protein-DNA complex (**Table S2**). Another interesting fact is the contribution of several residues from the loop containing the catalytic cysteine (~ residues 75-90) not only with the target cytosine but also interacting with the DNA minor groove, within phosphate backbone, orphan guanine and the adjacent guanine and thymine (**Figure 9A**). The involvement of this loop in the recognition of cognate DNA has been described in previous work, with MTase HgiDII (Castelán-Vega. et al. 2010). Still, in the catalytic domain, prokaryotic and eukaryotic DNMT2 share conserved residues-key as interacting with the substrate (**Figure 9B**). This fact is in agreement with previous studies that demonstrated, although the target substrate of the two MTase groups is different, the mechanisms involved in the methylation catalysis are the same (Dong et al. 2001; Cheng and Blumenthal 2008; Evans and Bronowska 2010; Swinehart and Jackman 2015). Thus, the differences emerge primarily in the H-bonds made by TRD.

The TRD composition differs in physicochemical nature, making them establish different contact points, restricting contacts between protein and DNA (**Figure 9A and 9B**). Despite the differences in the number and frequencies of interacting residues between eukaryotic enzymes, we can establish the consensus residues that seems to be fundamental in contact and substrate recognition by DNMT2 (**Figure 9B**), such as Cys238 and Thr240 from conserved CFT motif (motif IX) (reference: Human MT2), Lys241, Gly237 and the E/DGT motif (motif X). It is important to note that residues Cys238 and Phe239 of the conserved CFT motif of eukaryotes are in the same spatial position of the TL motif of HhaI, which interacts with thymine adjacent to the target cytosine. In addition, in HhaI, the Arg240 is crucial for the recognition of the target sequence, loop closure and catalysis (Estabrook et al. 2009), but this residue has no positional homolog in the eukaryotic DNMT2, there is only some arginines and lysines that still appear in the TRD making H-bonds, in regions where there is no sequence conservation and with a large variation in the electrostatic charges. It is interesting to highlight the fact that the eight residues previously described as strongly linked to the affinity and catalytic activity of DNMT2 (Jurkowski et al. 2012) presented constant H-bonds throughout the MD simulations (**Figure 9B**).

The HhaI accommodate the DNA molecule more adequately, this is reflected in the little distortion of the oligomer represented in the different MD simulation snapshots (**Figure 9A**). This same accommodation aspect is observed in the snapshots of wilDNMT2 and Human MT2, however, DNA of the former has less structural distortions than the DNA of the latter. All other enzymes seem to wriggle, stretch, twist the DNA in their interaction framework, as in an attempt to find a more favorable position.

From REBASE (Roberts et al. 2015) we made a sequence searching, using the sequences of human MT2, wilDNMT2, and Ehmeth. As results, only three putative enzymes shown assigned recognition: the RsaI, an m4C-MTase, which it has recognition sequence 5'-GTAC-3'; and two 6mA-MTase, M.GmeI and M.GmeII, recognition motif GGATC and TCCAGG, respectively. Surprisingly, M.GmeI and M.GmeII share conservation with residues from eukaryotic TRD, which make H-bond with DNA. These results may indicate the eukaryotic DNMT2 promiscuity in relation not only to the type of molecule (DNA or RNA) but also to the target sequence. More recently, to get insight into the structural basis to DNMT2- tRNA complex, the crystal structure of *Saccharomyces pombe* DNMT2 was determined. The interaction with tRNA was analyzed by means of mass-spectrometry using UV cross-linked Dnmt2-tRNA complex, and the results showed the residues cross-linked to tRNAAsp are Lys91, Trp221, His223 and Cys303 (Johannsson et al. 2018). These residues are very close to the residues found in the present study forming H-bonds with the DNA.

The results obtained clearly show a global trend during dynamics in the DNMT2. Several studies have identified changes in dynamics that favor the ligand binding and they are similarly dispersed throughout the entire protein (Okazaki and Takada 2008; Stockmann et al. 2008; Daniels et al. 2015; Kaplan et al. 2016). It is evident, however, that the change in dynamics associated with the DNA binding is linked to an enthalpy-entropy compensation, which contributes to the binding free energy. Here, the cofactor binding (SAM or SAH) was not take in the count, but it is known that SAM seems to play a dual role, acting not merely as a donor of the methyl group for the target cytosine methylation but also facilitating the molecular association between DNMT and substrate (Evans and Bronowska 2010).

Even considering a global trend, there are differences found here that are congruent with the various results from previous studies that denote plasticity or duality

with respect to recognition sites and action substrates. For example, human MT2 appears to be adapted almost exclusively to acting on tRNA (Hermann et al. 2003; Goll et al. 2006a; Müller et al. 2013; Shanmugam et al. 2015), which gives it characteristics described herein of lowest affinity and interactions with DNA throughout the dynamics. On the other hand, drosophilids having only DNMT2 as a DCM methyltransferase (Lyko et al. 2000; Marhold et al. 2004), it seems to act on the two substrates (DNA and/or RNA)(Jeltsch et al. 2006), which justifies the highest affinity and interaction with DNA during the MD. Nevertheless, it is possible to note substantial differences when comparing *D. melanogaster*, which has methylation sites CpA, CpC and CpT (Lyko et al. 1985; Kunert et al. 2003; Panikar et al. 2015), and *D. willistoni*, which presents methylation mainly in context CpG, besides the sex-specific methylation - a unique phenomenon, not recurring in any other group of species of drosophilids (Garcia et al. 2007; D'Ávila et al. 2010). Thus, our results suggest that protein dynamics, associated with structural conformation, surface electrostatic potentials, and TRD composition seems to contribute to the differences in substrate recognition and binding specificity found in DNMT2 family.

Summarily, we can infer that the peculiarities existing in DNMT2 are closely related to the evolutionary history, the ecological and epigenetic context of the different species, since we have eukaryotic organisms that carry the three families of DNMTs (DNMT1, DNMT2, and DNMT3), others that lost DNMT2 in their evolutionary journey and those that only carry the enzyme DNMT2. The fact that the enzyme is strongly conserved among the most different eukaryotic lineages is evidence to its evolutionary importance; on the other hand, the fact that its target recognition domain encompasses great diversity (in various molecular levels: primary structure, sub-structure, secondary structure, molecular electrical potential surfaces) suggests that DNMT2 has great adaptive potential.

Disclosure of potential conflicts of interest

The authors have no potential conflicts of interest to disclose.

Acknowledgments

This study was supported by grants and fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS n° 10/0028-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PNPD-CAPES).

References

- Benkert P, Biasini M and Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350. doi: 10.1093/bioinformatics/btq662
- Bowie JU, Lu "thy R ED (1991) A method to identify protein sequences that fold into a known three- dimensional structure. *Science* (80-) 253:164–70.
- Breiten B, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, Bowers CM, Heroux A, Krilov G and Whitesides GM (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* 135:15579–15584. doi: 10.1021/ja4075776
- Castelán-Vega. JA, Jiménez-Alberto A and Ribas-Aparicio RM (2010) Homology modeling and molecular dynamics simulations of HgiDII methyltransferase in complex with DNA and S-adenosyl-methionine: Catalytic mechanism and interactions with DNA. *J Mol Model* 16:1213–1222. doi: 10.1007/s00894-009-0632-9
- Chédin F (2011) The DNMT3 family of mammalian de novo DNA methyltransferases, 1st ed. *Prog Mol Biol Transl Sci*. doi: 10.1016/B978-0-12-387685-0.00007-X
- Chen L, MacMillan AM, Chang W, Ezaz-Nikpay K, Verdine GL and Lane WS (1991) Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase. *Biochemistry* 30:11018–11025. doi: 10.1021/bi00110a002
- Cheng X and Blumenthal RM (2008) Mammalian DNA Methyltransferases: A Structural Perspective. *Structure* 16:341–350. doi: 10.1016/j.str.2008.01.004
- D'Ávila MF, Garcia RN, Panzera Y and Valente VLS (2010) Sex-specific methylation in

Drosophila: An investigation of the *Sophophora* subgenus. *Genetica* 138:907–913. doi: 10.1007/s10709-010-9473-9

Daniels KG, Suo Y and Oas TG (2015) Conformational kinetics reveals affinities of protein conformational states. *Proc Natl Acad Sci* 112:9352–9357. doi: 10.1073/pnas.1502084112

Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH and Cheng X (2001) Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res* 29:439–448. doi: 10.1093/nar/29.2.439

Durdevic Z, Hanna K, Gold B, Pollex T, Cherry S, Lyko F and Schaefer M (2013) Efficient RNA virus control in *Drosophila* requires the RNA methyltransferase Dnmt2. *EMBO Rep* 14:269–275. doi: 10.1038/embor.2013.3

Elengoe A, Abu Naser M and Hamdan S (2014) Modeling and docking studies on novel mutants (K71L and T204V) of the ATPase domain of human heat shock 70 kDa protein 1. *Int J Mol Sci* 15:6797–6814. doi: 10.3390/ijms15046797

Estabrook RA, Nguyen TT, Fera N and Reich NO (2009) Coupling sequence-specific recognition to DNA modification. *J Biol Chem* 284:22690–22696. doi: 10.1074/jbc.M109.015966

Evans DA and Bronowska AK (2010) Implications of fast-time scale dynamics of human DNA/RNA cytosine methyltransferases (DNMTs) for protein function. *Theor Chem Acc* 125:407–418. doi: 10.1007/s00214-009-0681-2

Everett A (1990) Identification of a Critical by Mass Spectrometry * Cysteine in EcoRI DNA Methyltransferase. 265:17713–17719.

Figueiredo DF, Antunes DA, Rigo MM, Mendes MFA, Silva JP, Mayer FQ, Matte U, Giugliani R, Vieira GF and Sinigaglia M (2014) Lessons from molecular modeling human α -l-iduronidase. *J Mol Graph Model* 54:107–113. doi: 10.1016/j.jmgm.2014.10.004

Fisher O, Siman-Tov R and Ankri S (2004) Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res* 32:287–297. doi: 10.1093/nar/gkh161

- Fisicaro E, Compari C and Braibanti a. (2004) Entropy/enthalpy compensation: hydrophobic effect, micelles and protein complexes. *Phys Chem Chem Phys* 6:4156. doi: 10.1039/b404327h
- Garcia RN, D'Ávila MF, Robe LJ, Loreto ELDS, Panzera Y, De Heredia FO and Valente VLS (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica* 131:91–105. doi: 10.1007/s10709-006-9116-3
- Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-LL, Zhang X, Golic KG, Jacobsen SE and Bestor TH (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* (80-) 311:395–398. doi: 10.1126/science.1120976
- Hanck T, Schmidt S, Fritz H, Genetik M and Gottingen G (1993) Sequence-specific and mechanism-based crosslinking of Dem DNA cytosine-C methyltransferase of *E. coli* K-12 to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine. *21:303–309*.
- Hermann A, Gowher H and Jeltsch A (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci* 61:2571–2587. doi: 10.1007/s00018-004-4201-1
- Hermann A, Schmitt S and Jeltsch A (2003) The human Dnmt2 has residual DNA-(Cytosine-C5) methyltransferase activity. *J Biol Chem* 278:31717–31721. doi: 10.1074/jbc.M305448200
- Humphrey W, Dalke A and Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38. doi: 10.1016/0263-7855(96)00018-5
- Hurd PJ, Whitmarsh AJ, Baldwin GS, Kelly SM, Waltho JP, Price NC, Connolly BA and Hornby DP (1999) Mechanism-based inhibition of C5-cytosine DNA methyltransferases by 2-H pyrimidinone. *J Mol Biol* 286:389–401. doi: 10.1006/jmbi.1998.2491
- Jeltsch A (2002) Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem* 3:274–293. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S
- Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M and Helm M (2016) Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol* 6286:00–00. doi: 10.1080/15476286.2016.1191737

Jeltsch A, Nellen W and Lyko F (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 31:306–308. doi: 10.1016/j.tibs.2006.04.005

Johannsson S, Neumann P, Wulf A, Welp LM, Gerber H-D, Krull M, Diederichsen U, Urlaub H and Ficner R (2018) Structural insights into the stimulation of *S. pombe* Dnmt2 catalytic efficiency by the tRNA nucleoside queuosine Dnmt2 methylates cytosine at position 38 of tRNA. 1–10. doi: 10.1038/s41598-018-27118-5

Jurkowski TP, Shanmugam R, Helm M and Jeltsch A (2012) Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry* 51:4438–4444. doi: 10.1021/bi3002659

Kaiser S, Jurkowski TP, Kellner S, Schneider D, Helm M, Kaiser S, Jurkowski TP, Kellner S, Schneider D, Kaiser S et al. (2016) The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biol* 0:1–11. doi: 10.1080/15476286.2016.1236170

Kaplan M, Narasimhan S, de Heus C, Mance D, van Doorn S, Houben K, Popov-Čeleketić D, Damman R, Katrukha EA, Jain P et al. (2016) EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* 167:1241–1251.e11. doi: 10.1016/j.cell.2016.10.038

Katoh M, Curk T, Xu Q, Zupan B, Kuspa A and Shaulsky G (2006) Developmentally Regulated DNA Methylation in *Dictyostelium discoideum*. *Am Soc Microbiol* 5:18–25. doi: 10.1128/EC.5.1.18

Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994a) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–69.

Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994b) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–369. doi: 10.1016/0092-8674(94)90342-5

Klimasauskas S, Kumar S, Roberts RJ, Cheng X, Klimasauskas, Salius Kumar, Sanjay Roberts, Richard Cheng X, Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994c) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–369. doi: 10.1016/0092-8674(94)90342-5

- Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, Söderbom F et al. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res* 33:6405–6417. doi: 10.1093/nar/gki952
- Kunert N, Marhold J, Stanke J, Stach D and Lyko F (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* 130:5083–5090. doi: 10.1242/dev.00716
- Laskowski R, Rullmann JA, MacArthur M, Kaptein R and Thornton J (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486. doi: 10.1007/BF00228148
- Lin M-JJ, Tang L-YY, Reddy MN and Shen C-KKJJ (2005) DNA methyltransferase gene dDnmt2 and longevity of *Drosophila*. *J Biol Chem* 280:861–864. doi: 10.1074/jbc.C400477200
- Lipari G and Szabo A (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J Am Chem Soc* 104:4559–4570. doi: 10.1021/ja00381a010
- Lyko F, Ramsahoye BH and Jaenisch R (1985) DNA methylation in *Drosophila melanogaster*. *Nature* 19:1065–1068. doi: 10.1002/(SICI)1097-0177(199909)216:1<1::AID-DVDY1>3.0.CO;2-T
- Lyko F, Whittaker AJ, Orr-Weaver TL and Jaenisch R (2000) The putative *Drosophila* methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. *Mech Dev* 95:215–217. doi: 10.1016/S0925-4773(00)00325-7
- MacRaild CA, Daranas AH, Bronowska A and Homans SW (2007) Global Changes in Local Protein Dynamics Reduce the Entropic Cost of Carbohydrate Binding in the Arabinose-binding Protein. *J Mol Biol* 368:822–832. doi: 10.1016/j.jmb.2007.02.055
- Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B and Lyko F (2004) Conservation of DNA methylation in dipteran insects. *Insect Mol Biol* 13:117–123. doi: 10.1111/j.0962-1075.2004.00466.x
- Marlow MS, Dogan J, Frederick KK, Valentine KG and Wand AJ (2010) The role of

conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6:352–358. doi: 10.1038/nchembio.347

Marti-Renom MA, Stuart AC, Sanchez R, Melo F and Sali A (2000) Comparative Protein Structure Modeling of Genes and Genomes. *Annu Rev Biophys Biomol Struct* 29:291–325.

Müller S, Windhof IM, Maximov V, Jurkowski T, Jeltsch A, Förstner KU, Sharma CM, Gräf R and Nellen W (2013) Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). *Nucleic Acids Res* 41:8615–8627. doi: 10.1093/nar/gkt634

Nagasundaram N, Zhu H, Liu J, Karthick V, George Priya Doss C, Chakraborty C and Chen L (2015) Analysing the effect of mutation on protein function and discovering potential inhibitors of CDK4: Molecular modelling and dynamics studies. *PLoS One* 10:1–21. doi: 10.1371/journal.pone.0133969

O’Gara M, Roberts RJ and Cheng X (1996) A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase. *J Mol Biol* 263:597–606. doi: 10.1006/jmbi.1996.0601

Okazaki K -i. and Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci* 105:11182–11187. doi: 10.1073/pnas.0802524105

Panikar CS, Rajpathak SN, Abhyankar V, Deshmukh S and Deobagkar DD (2015) Presence of DNA methyltransferase activity and CpC methylation in *Drosophila melanogaster*. *Mol Biol Rep* 42:1615–1621. doi: 10.1007/s11033-015-3931-5

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi: 10.1002/jcc.20084

Ponger L and Li W-H (2005) Evolutionary Diversification of DNA Methyltransferases in Eukaryotic Genomes. *Mol Biol Evol* 22:1119–1128. doi: 10.1093/molbev/msi098

Pronk S, Phillips S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D et al. (2013) GROMACS 4.5: A high-throughput and

highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854. doi: 10.1093/bioinformatics/btt055

Reinisch KM, Chen L, Verdine GL and Lipscomb WN (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* 82:143–153. doi: 10.1016/0092-8674(95)90060-8

Roberts RJ, Vincze T, Posfai J and Macelis D (2015) REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299. doi: 10.1093/nar/gku1046

Santi D, Garrett C and Barr P (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* 33:9–10. doi: 10.1016/0092-8674(83)90327-6

Schaefer MR (2018) tRNA modifications: Necessary for correct tRNA-derived fragments during the recovery from stress? doi: 10.1002/bies.201200158

Schulz EC, Roth HM, Ankri S and Ficner R (2012) Structure analysis of *Entamoeba histolytica* DNMT2 (EhMeth). *PLoS One*. doi: 10.1371/journal.pone.0038728

Shanmugam R, Fierer J, Kaiser S, Helm M, Jurkowski TP and Jeltsch A (2015) Cytosine methylation of tRNA-Asp by DNMT2 has a role in translation of proteins containing poly-Asp sequences. *Cell Discov* 1:1–10. doi: 10.1038/celldisc.2015.10

Sisi Li, Jiamu Du, Hui Yang, Juan Yin, Jianping Ding JZ (2012) Functional and structural characterization of DNMT 2 from *Spodoptera frugiperda*. *J Mol Cell Biol* 5:64–66.

Stockmann H, Bronowska A, Syme NR, Thompson GS, Kalverda AP, Warriner SL and Homans SW (2008) Residual Ligand Entropy in the Binding of p-Substituted Benzenesulfonamide Ligands to Bovine Carbonic Anhydrase II. *J Am Chem Soc* 130:12420–12426. doi: 10.1021/ja803755m

Swaminathan CP, Sankpal UT, Rao DN and Surolia A (2002) Water-assisted dual mode cofactor recognition by HhaI DNA methyltransferase. *J Biol Chem* 277:4042–4049. doi: 10.1074/jbc.M109237200

Swinehart W and Jackman J (2015) Diversity in mechanism and function of tRNA methyltransferases. *RNA Biol* 6286:37–41. doi: 10.1080/15476286.2015.1008358

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJC (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718. doi: 10.1002/jcc.20291

Vieira GC, Sinigaglia M, Vieira GF and Valente VLS (2017) Linking epigenetic function to electrostatics: The DNMT2 structural model example. *PLoS One* 12:1–19. doi: 10.1371/journal.pone.0178643

Wu JC and Santi D V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J Biol Chem* 262:4778–4786. doi: 10.1016/S0959-440X(97)80013-9

Wyszynski MW, Gabbara S and Bhagwat AS (1991) Substitutions of cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. *20:319–326.*

Wyszynski MW, Gabbara S, Kubareva EA, Romanova EA, Oretskaya TS, Gromova ES, Shabarova ZA and Bhagwat AS (1993) The cysteine conserved among DNA cytosine methylases required for methyl transfer, but not for specific DNA binding. *Nucleic Acids Res* 21:295–301. doi: 10.1093/nar/21.2.295

Yang D and Kay LE (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J Mol Biol* 263:369–382. doi: 10.1006/jmbi.1996.0581

Yoder JA and Bestor TH (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet* 7:279–284. doi: 10.1093/hmg/7.2.279

Supporting information

Table S1. Sum of the differences between RMSFs [Holo - Apo].

Table S2. Residues involved in Protein-DNA interactions.

Table S3. Interaction energy protein-DNA from structures although 50ns molecular dynamic simulation.

Figure S1. 3D free energy landscape of DNMT2 apoenzyme as a function of the radius of gyration (RG) and the root-mean-square deviation (RMSD).

Figure S2. DNA RMSD over the 50 ns molecular dynamics simulations at 300 K.

Figure S3. RMSD of the catalytic domain (apo-, black, and holo-, red) over the 50 ns molecular dynamics simulations at 300 K.

Figure S4. RMSD of the TRD (apo-, black, and holo-, red) over the 50 ns molecular dynamics simulations at 300 K.

Supplementary Material

Table S1. Sum of the differences between RMSFs [Holo - Apo].

HhaI	HaeIII	Ehmeth	meIDNMT2	wilDNMT2	SfDNMT2	Human DNMT2
-9.6376	-7.9525	0.106	-1.6562	-9.8229	-0.8807	-5.528

Table S2. Residues involved in Protein-DNA interactions.

HhaI				
Residues	N°	Reference	MD	N°
			TYR	44
			GLY	78
PHE	79	[3][6]	PHE	79
CYS	81	[3][6]	CYS	81
			GLN	82
SER	85	[3][6]	SER	85
ILE	86	[3][6]	ILE	86
			SER	87
			GLY	88
LYS	89	[3][6]	LYS	89
			GLN	90
ARG	97	[3][6]	ARG	97
GLU	119	[1][3][5][6]	GLU	119
			ASN	120
			LYS	122
			ASN	123
SER	126	[3][6]	SER	126
LYS	162	[3][6]	LYS	162
			ARG	163
ARG	165	[1][3][4][6]	ARG	165
ARG	209	[3][6]	ARG	209
THR	226	[3][6]	THR	226
ARG	228	[3][6]	ARG	228
LYS	234	[3][6]	LYS	234
GLN	237	[2][3][6]	GLN	237
GLU	239	[2]	GLU	239
ARG	240	[2][3][6]	ARG	240
TYR	242	[2][3][6]	TYR	242
THR	250	[1][3][6]	THR	250
LEU	251	[1]		
SER	252	[3][6]		
ALA	253	[3][6]	ALA	253
TYR	254	[2][3][6]	TYR	254
			GLY	255
			GLY	256
GLY	257	[3][6]	GLY	257
			ILE	258
			LYS	261
SER	294	[3][6]	SER	294
			SER	296
GLN	297	[3][6]	GLN	297

Table S3. Interaction energy protein-DNA from structures although 50ns molecular dynamic simulation.

DNMT2						
	HhaI	Ehmeth	wilDNMT2	meDNMT2	SfDNMT2	Human
E_{int} (kJ/mol)	-10762.19	-6324.38	-8260.36	-5577.38	-6427.03	-6001.59

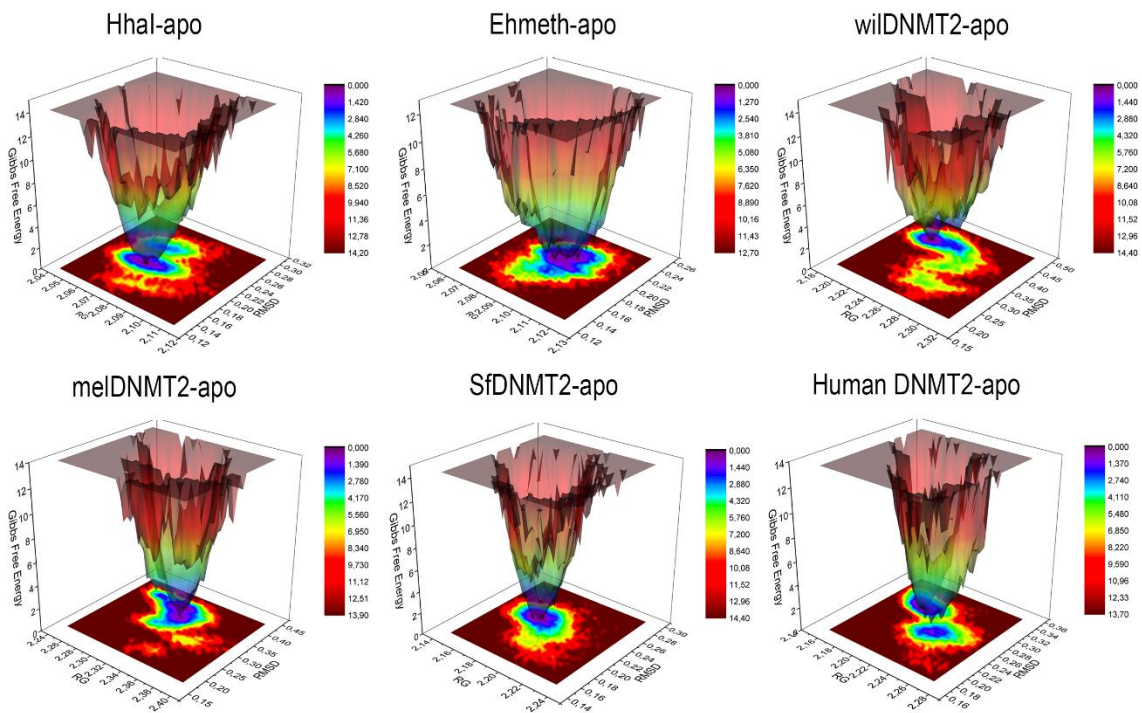


Figure S1. 3D free energy landscape of DNMT2 apoenzyme as a function of the radius of gyration (RG) and the root-mean-square deviation (RMSD).

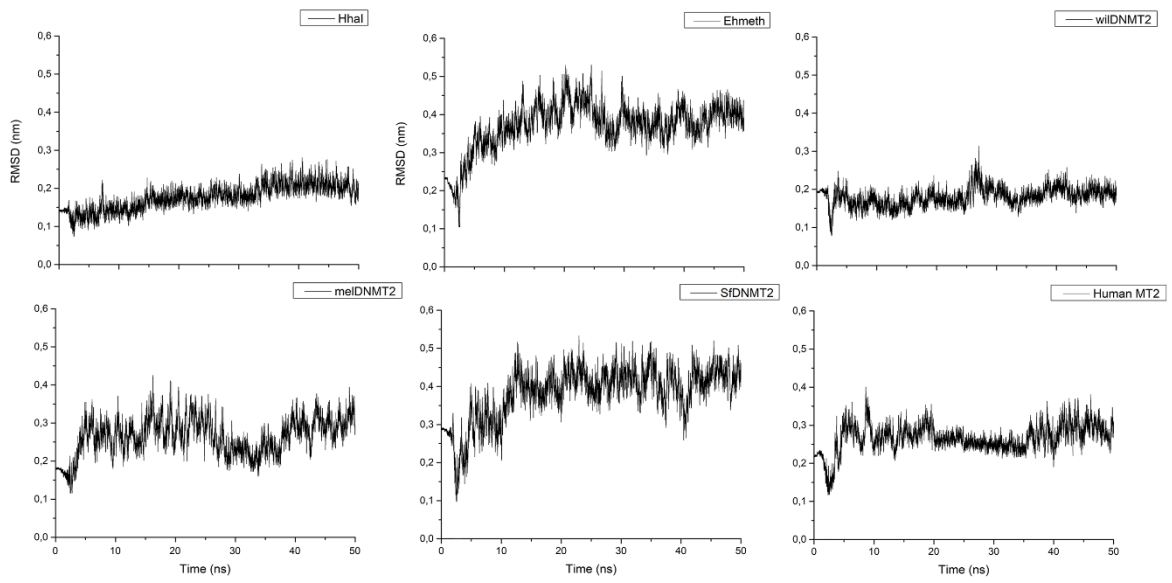


Figure S2. DNA RMSD over the 50 ns molecular dynamics simulations at 300 K.

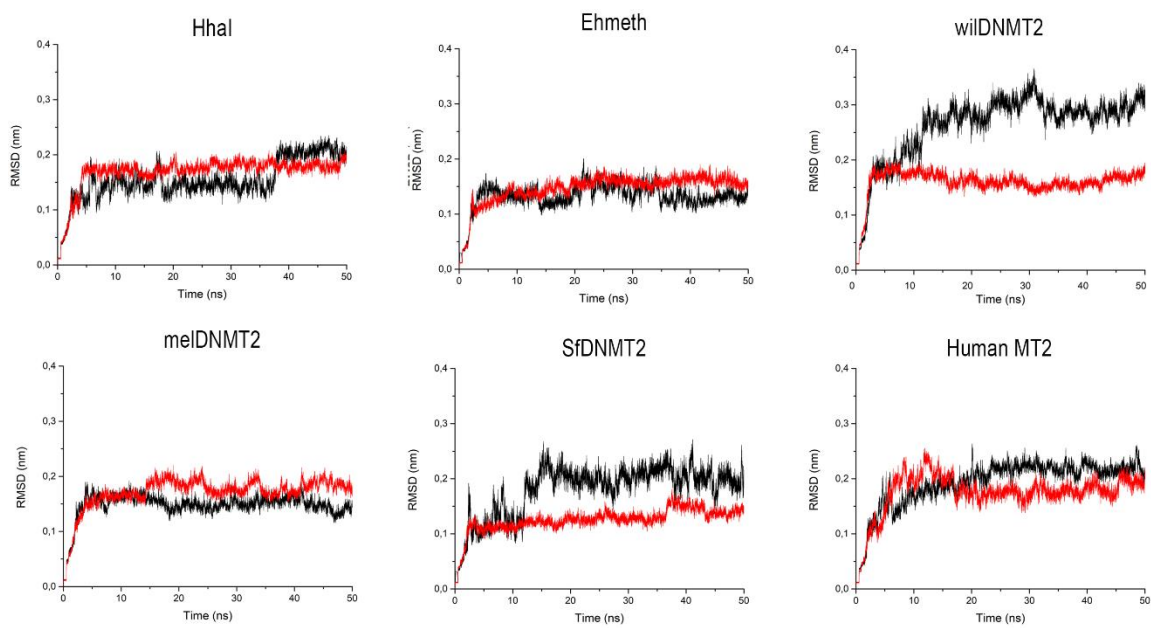


Figure S3. RMSD of the catalytic domain (apo-, black, and holo-, red) over the 50 ns molecular dynamics simulations at 300 K.

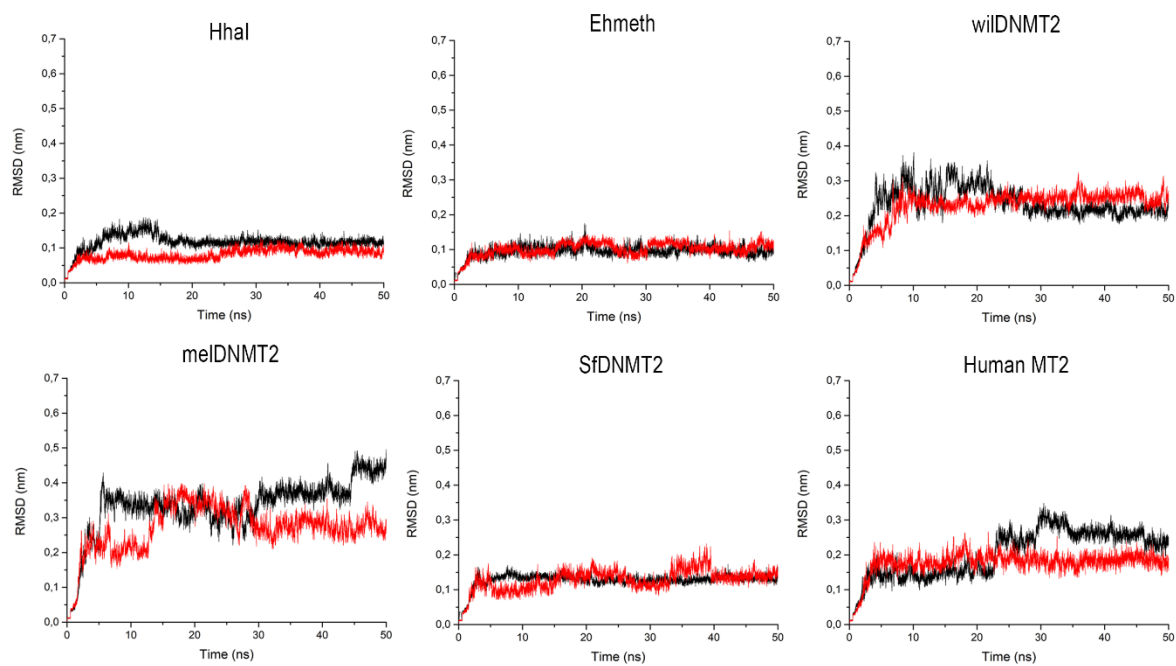


Figure S4. RMSD of the TRD (apo-, black, and holo-, red) over the 50 ns molecular dynamics simulations at 300 K.

CAPÍTULO VI

Discussão Geral

Perspectivas

Referências Bibliográficas

Anexos

Discussão Geral

Desde que foi publicada a primeira versão do genoma humano, isso em 26 de junho de 2000, o termo “epigenética” aos poucos foi tomando espaço – o que em um primeiro momento parece ser um contrassenso. A ideia inicial de que com o acesso a todos os genes e sequências que compõem o genoma dos organismos seria a chave para desvendar os mistérios por de trás da Biologia dos organismos, foi aos poucos se apagando frente ao fato que o genoma (a molécula de DNA propriamente) por si só não produz estruturas. A partir dessa realidade que se levantava frente aos olhos dos pesquisadores, a epigenética surge com a esperança de mostrar que os organismos são muito mais do que a soma de seus genes. A “redescoberta” da epigenética abriu as portas para o entendimento de como ambiente (interno e externo dos organismos) e genoma se comunicam.

E tudo isso tem início nos trabalhos seminais conduzidos por Conrad Hal Waddington, em 1942, com *Drosophila*, demonstrando como o ambiente pode induzir alterações na programação ontogenética dos organismos. Porém, seus achados, assim como as pesquisas conduzidas por Richard Benedict Goldschmidt (1878-1958), se mostravam pouco ortodoxos para o *ethos* de seu tempo, onde a Teoria Sintética Evolutiva tomava forma, erguida sobre as bases da Seleção Natural, Genética de Populações e Paleontologia. Ao longo de mais de sete décadas a ideia do genocentrismo entremeou os principais estudos nas mais diversas áreas da Biologia, enquanto a Ecologia e o Desenvolvimento eram relegados a campos secundários e tomando apenas papéis passivos dentro do contexto da Biologia Evolutiva. Contudo, esse cenário passa a se modificar a partir da década de 80, e mais fortemente com a chegada dos anos 90 com o advento de novas técnicas e tecnologias que possibilitaram a abertura da “caixa-preta”, que até então parecia indecifrável, de como fenótipo e genótipo se comunicam. O interessante, e não menos irônico, é que o surgimento, ostracismo e ressurgimento do campo de estudos epigenéticos estão amparados em ensaios feitos com *Drosophila*. Tal fato só demonstra a grande importância desses organismos e a incrível contribuição que vêm dando à ciência.

A presente tese tem como foco a compreensão da função biológica e histórico evolutivo envolvendo a enzima m5C-MTase Dnmt2 de *Drosophila*. Isso se faz interessante visto que, até final da década de 90, *Drosophila* era considerada como organismo livre de metilação de citosinas genômicas ou com níveis indetectáveis (Lyko

et al. 2000). Entretanto com o advento de novas análises que capacitaram o aprofundamento dos fenômenos epigenéticos de metilação, verificou-se que *Drosophila* possui citosinas metiladas (mesmo que em níveis menores que vertebrados). O interessante é que ao se buscar a enzima mediadora desse fenômeno, constatou-se que *Drosophila* não possui as MTase canônicas (Dnmt1 e Dnmt3), estando presente apenas a MTase Dnmt2 (Kunert et al. 2003). Dessa maneira, o gênero *Drosophila* passa a ser classificado como organismos “*Dnmt2-only*”. Logo, tanto o fenômeno de metilação genômica em drosofilídeos, como a função biológica de sua única MTase (Dnmt2), são objetos de estudos que atraem substancial curiosidade científica.

As investigações realizadas têm como ponto de partida os resultados obtidos em estudos prévios realizados por nosso grupo de pesquisa, onde a abordagem sobre elementos transponíveis em espécies neotropicais de *Drosophila willistoni* demonstrou que a regulação desses elementos é substancialmente diferente entre *D. melanogaster* (Sassi et al. 2005). Em 1997, Garcia, em sua dissertação de Mestrado, utilizando enzimas sensíveis às citosinas metiladas, verificou que os padrões de bandeamento nos cromossomos politênicos de *D. melanogaster* e *D. willistoni* se distinguiam. Para procurar compreender tais padrões encontrados, nosso grupo de pesquisa deu continuidade nas análises, o que possibilitou evidenciar o fenômeno de metilação no genoma de *D. willistoni*, mas não só isso, os padrões de metilação entre machos e fêmeas dessa espécie apresentavam diferenças em genes de DNA ribossomal (rDNA), ou seja, havia se descoberto o fenômeno de metilação sexo-específica em *D. willistoni* (Garcia et al. 2007). Procurando ampliar o entendimento desse fenômeno, e averiguar se o mesmo seria recorrente nas demais espécies de drosofilídeos, estabeleceu-se uma extensa análise sobre os padrões de metilação (D’Ávila et al. 2010). Analisando diversas linhagens do subgênero *Sophophora* de *Drosophila*, D’Ávila et al. (2010) verificaram que o fenômeno de metilação sexo-específica se restringe às espécies do subgrupo *willistoni*.

Garcia et al. (2007) lançou olhar para a Dnmt2 e analisou a conservação da enzima, comparando as sequências de *D. willistoni* e *D. melanogaster*. O que se verificou é que, apesar de substancial conservação dos domínios catalíticos das enzimas das duas espécies analisadas, essas diferenciam-se especialmente na região responsável por fazer o reconhecimento da sequência alvo a ser metilada (TRD). De fato, a Dnmt2 é considerada uma MTase peculiar e enigmática por conta da sua capacidade de atuar sobre diferentes substratos – DNA ou RNA – e ao mesmo tempo ser a mais amplamente

conservada das metiltransferases (Lyko et al. 2000; Lin et al. 2005; Goll et al. 2006b; Rai et al. 2007; Schaefer et al. 2008; Krauss and Reuter 2011; Kaiser et al. 2016; Genencher et al. 2018).

Por conta disso, no **Capítulo II** da presente tese procuramos fazer uma análise ampla da Dnmt2, quanto à sua conservação e prováveis forças evolutivas envolvidas. Verificamos que a maioria das 68 espécies analisadas apresentaram sinais positivos de hibridização e *amplicons* com sondas e *primers* de *Dnmt2*. Apenas espécies do subgênero *Dorsilopha* não apresentaram sinais ou *amplicons*, o que também não é suficiente para excluirmos o fato delas possuírem o gene da *Dnmt2*, mas que provavelmente exista destacada variabilidade que não tenha permitido a hibridização ou amplificação em PCR com as sondas e *primers* utilizados no presente estudo. Observamos que o gene *Dnmt2* possui considerável conservação entre 44 sequências analisadas. As análises filogenéticas com *Dnmt2* resgatam as relações evolutivas entre os clados dos subgêneros *Drosophila* e *Sophophora*. A enzima Dnmt2 possui dez motivos catalíticos (Lyko et al. 2000), e entre os motivos VIII e XIX encontra-se a TRD. Os motivos catalíticos apresentam grande conservação, porém com algumas regiões apresentando alta taxa evolutiva, enquanto a TRD apresenta uma taxa evolutiva maior entre as espécies. Tal fato está de acordo com o esperado tomando como base os achados descritos no trabalho de Garcia et al. (2007).

Logo, se os resultados obtidos apresentam, de modo geral, uma ampla conservação da Dnmt2 entre drosofilídeos, questionamos qual seriam os mecanismos evolutivos envolvidos nas diferenças encontradas dentro dos motivos catalíticos e TRD. Demonstramos que a seleção purificadora é preponderante na manutenção e conservação da Dnmt2 entre drosofilídeos, entretanto seis sítios apresentam sinais de seleção positiva e doze sítios evoluíram sob seleção positiva desestabilizadora, favorecendo mudanças estruturais e funcionais entre as Dnmt2 de drosofilídeos. A maioria desses sítios estão localizados em regiões de potencial interação com outras moléculas.

O que sabemos é que se determinadas enzimas estão interligadas e pertencem a uma mesma rede metabólica, logo esperamos que suas taxas evolutivas acabem sendo parecidas. Partindo desse princípio, efetuamos uma extensiva análise de genes possuindo covariação das taxas evolutivas com a Dnmt2 de *Drosophila*, nos revelando quinze proteínas que estão intimamente relacionadas à controles de expressão gênica *cis*, *trans*, remodelação de cromatina, modificações de RNAs e DNA. Tais achados destacam ainda

mais as peculiaridades inerentes da *Dnmt2*, ainda mais em organismos “*Dnmt2-only*”, como é o caso de *Drosophila*. Por fim, outro aspecto muito interessante que encontramos no estudo conduzido no **Capítulo II**, foi a alta taxa de variação nucleotídica por códon do gene *Dnmt2* em *D. willistoni*. Esse resultado discutiremos um pouco mais adiante.

A metilação de citosinas genômicas durante a ontogenia dos organismos é importante para o estabelecimento da identidade das células, promovendo assim a citodiferenciação e morfogênese ao longo do desenvolvimento (Serman et al. 2006; Reik 2007). Estudos anteriores procuraram caracterizar os padrões de expressão gênica da *Dnmt2* em *Drosophila*. O que se observou é que existe uma maior taxa de transcrição de *Dnmt2* nos estágios iniciais de desenvolvimento de *Drosophila* (Lyko 2001; Kunert et al. 2003). Em *Drosophila* mutante para *Dnmt2* verificou-se a perda de citosinas genômicas metiladas em embriões, perda da integridade da cromatina e aumento da atividade de retrotransposons (Phalke et al. 2009). Diante desse cenário e somando-se o fato do peculiar fenômeno de metilação sexo-específica em espécies do subgrupo *willistoni*, passamos a investigar os padrões de expressão gênica de *Dnmt2*, no âmbito espaço-temporal e sua localização *in situ* em cromossomos politênicos de *D. willistoni* (**Capítulo III**).

Nossas análises quantitativas dos níveis de transcritos de *Dnmt2* (**Capítulo III**), em um primeiro momento, não parecem diferir dos padrões encontrados por Lyko et al. (2001), no qual se observa elevados níveis de expressão em estágios iniciais de vida (embrião), diminuindo substancialmente ao longo das demais fases do desenvolvimento (larva, pupa e adultos). Entretanto, em uma observação mais atenta, verificamos que os níveis de transcritos de *Dnmt2* em adultos de *D. willistoni* são levemente maiores daqueles encontrados em estudos prévios (Lyko 2001; Kunert et al. 2003). Esse resultado é bem sugestivo, ainda mais quando estamos tratando de uma espécie que possui padrões de metilação sexo-específica. A hibridização *in situ* com ribosondas de *Dnmt2* em oócitos e embriões dão suporte aos resultados obtidos por qPCR.

Quanto à localização física em cromossomos politênicos (**Capítulo III**), a hibridação *in situ* revelou que o gene da *Dnmt2* de *D. willistoni* encontra-se no braço III, correspondente ao elemento B de Muller. Sua localização corrobora a proposta de preservação de sintenia dos elementos cromossômicos como sendo adaptativos (Schaeffer et al. 2008). Entretanto, a surpresa da localização da *Dnmt2* reside em sua

posição específica no braço III: o gene se encontra em região subtelomérica. Esse resultado acaba ligando os dados obtidos no **Capítulo II** com os dados do **Capítulo III**. A alta taxa de variação nucleotídica por códon do gene *Dnmt2* em *D. willistoni* pode ser explicada por conta da sua posição subtelomérica. Sabe-se que regiões heterocromáticas (centroméricas e teloméricas) e regiões eucromáticas possuem padrões evolutivos diferentes. Os padrões evolutivos dos telômeros (e de regiões subteloméricas) distinguem-se até mesmo entre espécies evolutivamente próximas (Trask et al. 1998). Em *D. melanogaster* a região telomérica é formada por elementos transponíveis *Het-A* e *TART*, porém através de análises por bandamento de cromossomos politênicos, utilizando diferentes marcadores, demonstrou-se que a região telomérica de *D. willistoni* é rica em dinucleotídeos CpG, o que contrasta com os telômeros de *D. melanogaster* (Garcia 1997, dados não publicados). Além disso, regiões teloméricas possuem baixas taxas de recombinação (Lindsley et al. 1977), mas em contrapartida possuem altas taxas de mutações (Roberts 1979).

Sabemos a partir dos resultados obtidos no **Capítulo II** que a *Dnmt2* de drosofilídeos é, de modo geral, conservada em seus motivos catalíticos, mas varia na TRD. Entretanto, analisar a estrutura primária de enzimas/proteínas acaba não permitindo que entendamos as interações e características físico-químicas que emergem a partir do dobramento proteico. Para aprofundarmos o entendimento a respeito das diferenças existentes entre as *Dnmt2* de drosofilídeos modelamos computacionalmente 26 estruturas terciárias (**Capítulo IV**). Além disso, encaminhamos uma análise evolutiva estrutural das *Dnmt2*, comparando os conformêros de 26 drosofilídeos com as *Dnmt2* de *Haemophilus influenzae*, *Haemophilus haemolyticus*, *Geobacter sulfurreducens* (procariotos), *Entamoeba histolytica* (protozoa), *Spodoptera frugiperda* (Lepdoptera) e de *Mus musculus* e *Homo sapiens* (vertebrados). De fato, constatamos conservação no domínio catalítico das *Dnmt2*, porém destacadas diferenças na região TRD entre procariotos e eucariotos. Essa diferença implica na preferência de substratos, na qual procariotos tem alta processividade em DNA (O’Gara et al. 1996a), enquanto a *Dnmt2* de eucariotos tem maior afinidade por RNA (Durdevic et al. 2013b).

Outra característica interessante que nossas análises apresentaram são as diferenças nos perfis de distribuição de potencial eletrostático de superfície entre as *Dnmt2* com afinidade por DNA e aquelas com ação sobre RNA (**Capítulo IV**). Claramente as enzimas com afinidade por DNA possuem uma distribuição acentuada de

cargas positivas na superfície da região TRD, o que não ocorre nas Dnmt2 de eucariotos. Curiosamente, a extensa análise dos potenciais eletrostáticos de superfície mostrou que grupos relacionados evolutivamente possuem diferenças marcantes na forma como as cargas estão distribuídas na superfície de suas moléculas, quando o esperado era que os drosofilídeos estivessem agrupados com pequenas diferenças. E aqui destacamos o perfil de distribuição de cargas na TRD de *D. willistoni*, onde essa possui prevalência de cargas positivas. Tanto que a correlação entre as constantes cinéticas preditas K_{cat}/K_m de *D. willistoni* se aproximam aos valores da Dnmt2 de *Haemophilus influenzae*, uma DNA MTase. Esse dado é interessante, pois se *D. willistoni* possui metilação sexo-específica, esse fenômeno é de algum modo bem regulado para promover o padrão, logo a Dnmt2 deve trazer consigo algum grau diferenciado de afinidade ao substrato (DNA) que não seja compartilhado com as demais espécies de drosofilídeos.

A questão que fica é: será que as diferenças de distribuição de cargas eletrostáticas de superfície na TRD das diferentes Dnmt2, associada às diferenças estruturais, de fato influenciam a afinidade entre o complexo Dnmt2-DNA?

Para responder a essa questão, elaboramos uma série de simulações de dinâmica molecular para avaliar o comportamento de diferentes Dnmt2 (**Capítulo V**). As análises oriundas das dinâmicas de HhaI (*H. influenzae*), Ehmeth (*E. histolytica*), SfDnmt2 (*S. frugiperda*), wilDnmt2 (*D. willistoni*), melDnmt2 (*D. melanogaster*) e a MT2 humana revelaram que realmente aquelas Dnmt2 que apresentam um perfil de cargas eletrostáticas de superfície positiva mais elevadas na TRD apresentam maior grau de interação com o substrato, no caso, o DNA. A wilDnmt2 possui um comportamento de alguma forma semelhante ao observado em HhaI, com baixa flutuação estrutural da molécula DNA, baixa dinâmica e manutenção de interações ligações de Hidrogênio mais estáveis ao longo de simulação de dinâmica molecular. Em termos gerais, os resultados obtidos no **Capítulo V** também trazem informações interessantes de como se processam as relações energéticas entre os estados apo (sem substrato ligado à enzima) e holo (substrato e enzima complexados). O que percebemos é que há uma rede de compensação entalpia-entropia intrínseca a cada Dnmt2 e que impacta diretamente a energia livre do sistema e sua variação entre os estados apo e holo.

Em síntese, os dados apresentados na presente tese (i) estão de acordo com o que diversos autores têm descrito em sistemas epigenéticos de drosofilídeos quanto aos níveis

de expressão; (ii) entretanto, o subgrupo *willistoni* de *Drosophila* aparece como um grupo de espécies que merece especial atenção diante das peculiaridades emergentes em níveis moleculares, epigenéticos e citogenéticos. Assim, não é ousado concluir que, apesar de *Drosophila melanogaster* ser um objeto de estudo valioso e ser amplamente empregado na Biologia, em seus mais diversificados campos, que a generalização de eventos genéticos, moleculares, fisiológicos e ecológicos para todos os demais drosofilídeos parece ser forçosa, justamente pelo fato de estarmos lidando com um gênero (*Drosophila*) constituído por uma imensa diversidade de espécies que carregam consigo marcas características de suas peculiares histórias evolutivas.

Perspectivas

Os trabalhos realizados anteriormente por nosso grupo de pesquisa do Laboratório de *Drosophila* da UFRGS abriram vasto campo de investigação a ser explorado. Desde a caracterização do fenômeno de metilação sexo-específica em *D. willistoni* (Garcia et al. 2007); passando pelo aprofundamento das buscas por entender a recorrência de tal fenômeno em drosofilídeos (D'Ávila et al. 2010) e, com a presente tese, revelando a conservação, história evolutiva e forças que moldaram as Dnmt2 em drosofilídeos (**Capítulo II**); analisando o perfil de expressão de Dnmt2 no âmbito espaço-temporal e sua localização cromossômica (**Capítulo III**); avaliando as características físico-químicas que emergem da estrutura terciária das Dnmt2 de drosofilídeos, revelando suas peculiaridades (**Capítulo IV**); e como essas características moleculares atuam nas interações da Dnmt2 com o substrato de DNA (**Capítulo V**). Entretanto há muito a ser esclarecido sobre os mecanismos epigenéticos envolvendo a Dnmt2 e a metilação de DNA e RNA, além do peculiar fenômeno de metilação sexo-específica.

Dessa maneira, listo aqui algumas questões pertinentes para a ampliação do fenômeno epigenético até aqui abordado e possíveis abordagens:

- Quais são os genes próximos à Dnmt2 de *D. willistoni*? Por estar em região subtelomérica qual o impacto desses no contexto evolutivo e fisiológico? Que forças evolutivas resguardam tal região cromossômica de *D. willistoni* e que diferenças existem (se existem) em relação aos demais drosofilídeos? O trabalho realizado no (Capítulo II) em conjunto com os

dados obtidos por Garcia et al. (2015) servem como ponto de partida para uma análise citológica por meio de experimentos de *genome-walking*.

- Como estão distribuídas as citosinas metiladas fisicamente no genoma de *D. willistoni*? Para isso, podemos empregar a técnica de FISH em cromossomos metafásicos, utilizando anticorpos anti-5metilcitosina (Pendina et al. 2016).
- Qual o *status* do metiloma de *D. willistoni*? Ele se diferencia em relação à *D. melanogaster*? Para tais questões, análises podem ser feitas por eletroforese capilar (Stach et al. 2003), ou através de espectrometria de massas (Reichetzeder et al. 2014).
- Quais são os genes ou sequências metiladas em *D. willistoni*? Através de cromatografia por afinidade poderemos acessar as sequências metiladas, utilizando para isso colunas de ligação ao DNA com anticorpos anti-5metilcitosinas (Salzberg et al. 2004).
- No trabalho realizado por Garcia et al. (2007) diversos fragmentos de DNA de *D. willistoni* oriundos da digestão pela técnica MSRE foram isolados e sequenciados. O que são essas sequências? Quais as características dessas sequências quanto ao seu conteúdo CpG? Existem genes próximos a esses fragmentos? Para essas questões uma busca *in silico* pode ser conduzida para localizar essas sequências junto ao genoma de *D. willistoni* e avaliar o que se encontra em regiões próximas.
- Qual a importância fisiológica e ontogenética da Dnmt2 para *D. willistoni*? Como os mecanismos epigenéticos fazem a ponte entre genótipo e fenótipo, entre ambiente e genoma, parece interessante avaliar os impactos ocasionados pelo composto orgânico sintético Bisfenol A (BPA), que já se mostrou um potente disruptor estrogênico, interferindo na rota metabólica de reciclagem de S-adenosilmetionina (grupo doador de metil intermediado pela Dnmt2). Para isso, pode se estabelecer protocolo com diferentes concentrações de BPA em meio de cultura onde as linhagens vivem e avaliar diversos parâmetros em nível macro: como taxa de mortalidade, fecundidade, desenvolvimento de fenótipos aberrantes e modificação de comportamento. E também avaliar em nível molecular os impactos oriundos da exposição de *D. willistoni* ao BPA, como: níveis

expressão de Dnmt2, níveis de citosinas metiladas e modificação nos padrões de metilação sexo-específico.

Referências Bibliográficas

- Adams RLP, McKay EL, Craig LM and Burdon RH (1979) Methylation of mosquito DNA. *Biochim Biophys Acta - Nucleic Acids Protein Synth* 563:72–81. doi: 10.1016/0005-2787(79)90008-X
- Anderson CL, Anderson CL, Carew EA and Powell JR (1993) Evolution of the Adh locus in the *Drosophila willistoni* group: the loss of an intron, and shift in codon usage. *Mol Biol Evol.* doi: 10.1093/oxfordjournals.molbev.a040027
- Anderson JA, Gilliland WD and Langley CH (2009) Molecular population genetics and evolution of *Drosophila* meiosis genes. *Genetics* 181:177–185. doi: 10.1534/genetics.108.093807
- Auclair G, Guibert S, Bender A and Weber M (2014) Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. 1–16. doi: 10.1186/s13059-014-0545-5
- Ayala FJ and Powell JR (1972) Allozymes as Diagnostic Characters of Sibling Species of *Drosophila*. *Proc Natl Acad Sci* 69:1094–1096. doi: 10.1073/pnas.69.5.1094
- Banta JA and Richards CL (2018) Quantitative epigenetics and evolution. *Heredity (Edinb)* 121:210–224. doi: 10.1038/s41437-018-0114-x
- Benkert P, Biasini M and Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350. doi: 10.1093/bioinformatics/btq662
- Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A, Wilson IA and Wright PE (2013) Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat Struct Mol Biol* 20:1243–1249. doi: 10.1038/nsmb.2676
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16:6–21. doi: 10.1101/gad.947102
- Bird A (1992) The essentials of DNA methylation. *Cell* 70:5–8. doi: 10.1016/0092-8674(92)90526-I
- Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y et

- al. (2012) Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol* 22:1755–1764. doi: 10.1016/j.cub.2012.07.042
- Bonfils C, Beaulieu N, Chan E, Cotton-Montpetit J and MacLeod AR (2000) Characterization of the human DNA methyltransferase splice variant Dnmt1b. *J Biol Chem* 275:10754–10760. doi: 10.1074/jbc.275.15.10754
- Bowie JU, Luthey-Schulten R (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* (80-) 253:164–70.
- Breiling A and Lyko F (2015) Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics and Chromatin* 8:1–9. doi: 10.1186/s13072-015-0016-6
- Breiten B, Lockett MR, Sherman W, Fujita S, Al-Sayah M, Lange H, Bowers CM, Heroux A, Krilov G and Whitesides GM (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* 135:15579–15584. doi: 10.1021/ja4075776
- Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F et al. (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol* 12:944–950. doi: 10.1038/nchembio.2175
- Campos-Ortega JA and Hartenstein V (1985) *The Embryonic Development of Drosophila melanogaster*. doi: 10.1007/978-3-662-02454-6
- Castelán-Vega JA, Jiménez-Alberto A and Ribas-Aparicio RM (2010) Homology modeling and molecular dynamics simulations of HgiDII methyltransferase in complex with DNA and S-adenosyl-methionine: Catalytic mechanism and interactions with DNA. *J Mol Model* 16:1213–1222. doi: 10.1007/s00894-009-0632-9
- Chédin F (2011) *The DNMT3 family of mammalian de novo DNA methyltransferases*, 1st ed. *Prog Mol Biol Transl Sci*. doi: 10.1016/B978-0-12-387685-0.00007-X
- Chen L, MacMillan AM, Chang W, Ezaz-Nikpay K, Verdine GL and Lane WS (1991) Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase. *Biochemistry* 30:11018–11025. doi: 10.1021/bi00110a002

Cheng JX, Chen L, Li Y, Cloe A, Yue M, Wei J, Watanabe KA, Shammo JM, Anastasi J, Shen QJ et al. (2018) RNA cytosine methylation and methyltransferases mediate chromatin organization and 5-azacytidine response and resistance in leukaemia. *Nat Commun* 9:1–16. doi: 10.1038/s41467-018-03513-4

Cheng X and Blumenthal RM (2008) Mammalian DNA Methyltransferases: A Structural Perspective. *Structure* 16:341–350. doi: 10.1016/j.str.2008.01.004

Chuang LSH, Ian HI, Koh TW, Ng HH, Xu G and Li BFL (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21(WAF1). *Science* (80-) 277:1996–2000. doi: 10.1126/science.277.5334.1996

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218. doi: 10.1038/nature06341

D'Ávila MF, Garcia RN, Panzera Y and Valente VLS (2010) Sex-specific methylation in *Drosophila*: An investigation of the *Sophophora* subgenus. *Genetica* 138:907–913. doi: 10.1007/s10709-010-9473-9

da Cunha B and Dobzhansky T (1954) A Further Study Of Chromosomal Polymorphism In *Drosophila willistoni* In Its Relation To The Environment. *Evolution* (N Y) 8:119–134.

da Cunha AB and Dobzhansky T (1950) Adaptive Chromosomal Polymorphism In *Drosophila willistoni*. 212–235.

da Cunha AB, Dobzhansky T, Pavlovsky O and Spassky B (1959) Genetics Of Natural Populations. Xxviii. Supplementary Data On The Chromosomal Polymorphism In *Drosophila willistoni* In Its Relation To The Environment. *Evolution* (N Y) 13:389–404. doi: 10.1111/j.1558-5646.1959.tb03025.x

Daniels KG, Suo Y and Oas TG (2015) Conformational kinetics reveals affinities of protein conformational states. *Proc Natl Acad Sci* 112:9352–9357. doi: 10.1073/pnas.1502084112

Deprá M, Valente VLS, Margis R and Loreto ELS (2009) The hobo transposon and hobo-related elements are expressed as developmental genes in *Drosophila*. *Gene* 448:57–63. doi: 10.1016/J.GENE.2009.08.012

Deshmukh S, Ponnaluri VC, Dai N, Pradhan S and Deobagkar D (2018) Levels of DNA

cytosine methylation in the *Drosophila* genome. PeerJ 6:e5119. doi: 10.7717/peerj.5119

Desrosiers R and Tanguay RM (1988) Methylation of *Drosophila* histones at proline, lysine, and arginine residues during heat shock. J Biol Chem 263:4686–4692.

Dev RR, Ganji R, Singh SP, Mahalingam S, Banerjee S and Khosla S (2017) Cytosine methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during infection. Biochem J 474:2009–2026. doi: 10.1042/BCJ20170258

Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH and Cheng X (2001) Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res 29:439–448. doi: 10.1093/nar/29.2.439

Dozmorov MG, Giles CB, Koelsch KA and Wren JD (2013) Systematic classification of non-coding RNAs by epigenomic similarity. BMC Bioinformatics 14:S2. doi: 10.1186/1471-2105-14-S14-S2

Durdevic Z (2013) Characterization of the Biological Function of Dnmt2 in *Drosophila melanogaster*.

Durdevic Z, Hanna K, Gold B, Pollex T, Cherry S, Lyko F and Schaefer M (2013a) Efficient RNA virus control in *Drosophila* requires the RNA methyltransferase Dnmt2. EMBO Rep 14:269–275. doi: 10.1038/embor.2013.3

Durdevic Z, Mobin MB, Hanna K, Lyko F and Schaefer M (2013b) The RNA methyltransferase dnmt2 is required for efficient dicer-2-dependent siRNA pathway activity in *Drosophila*. Cell Rep 4:931–937. doi: 10.1016/j.celrep.2013.07.046

Durdevic Z and Schaefer M (2013) tRNA modifications: Necessary for correct tRNA-derived fragments during the recovery from stress? BioEssays 35:323–327. doi: 10.1002/bies.201200158

Ehrlich M, Wilson GG, Kuo KC and Gehrke CW (1987) N4-methylcytosine as a minor base in bacterial DNA. J Bacteriol 939–943.

el-Deiry WS, Nelkin BD, Celano P, Yen RW, Falco JP, Hamilton SR and Baylin SB (1991) High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. Proc Natl Acad Sci U S A 88:3470–4. doi: 10.1073/pnas.88.8.3470

Elango N, Hunt BG, Goodisman MAD and Yi S V. (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. PNAS 106:11206–11211.

Elengoe A, Abu Naser M and Hamdan S (2014) Modeling and docking studies on novel mutants (K71L and T204V) of the ATPase domain of human heat shock 70 kDa protein 1. Int J Mol Sci 15:6797–6814. doi: 10.3390/ijms15046797

Estabrook RA, Nguyen TT, Fera N and Reich NO (2009) Coupling sequence-specific recognition to DNA modification. J Biol Chem 284:22690–22696. doi: 10.1074/jbc.M109.015966

Evans DA and Bronowska AK (2010) Implications of fast-time scale dynamics of human DNA/RNA cytosine methyltransferases (DNMTs) for protein function. Theor Chem Acc 125:407–418. doi: 10.1007/s00214-009-0681-2

Everett A (1990) Identification of a Critical by Mass Spectrometry * Cysteine in EcoRI DNA Methyltransferase. 265:17713–17719.

Farthing CR, Ficiz G, Ng RK, Chan CF, Andrews S, Dean W, Hemberger M and Reik W (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genet. doi: 10.1371/journal.pgen.1000116

Feinberg AP and Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301:89–92. doi: 10.1038/301089a0

Figueiredo DF, Antunes DA, Rigo MM, Mendes MFA, Silva JP, Mayer FQ, Matte U, Giugliani R, Vieira GF and Sinigaglia M (2014) Lessons from molecular modeling human α -l-iduronidase. J Mol Graph Model 54:107–113. doi: 10.1016/j.jmglm.2014.10.004

Finnegan EJ, Margis R and Waterhouse PM (2003) Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of dicer-1 from Drosophila. Curr Biol 13:236–240. doi: 10.1016/S0960-9822(03)00010-1

Fisher O, Siman-Tov R and Ankri S (2006) Pleiotropic phenotype in *Entamoeba histolytica* overexpressing DNA methyltransferase (Ehmeth). Mol Biochem Parasitol

147:48–54. doi: 10.1016/j.molbiopara.2006.01.007

Fisher O, Siman-Tov R and Ankri S (2004) Characterization of cytosine methylated regions and 5-cytosine DNA methyltransferase (Ehmeth) in the protozoan parasite *Entamoeba histolytica*. *Nucleic Acids Res* 32:287–297. doi: 10.1093/nar/gkh161

Fisicaro E, Compari C and Braibanti a. (2004) Entropy/enthalpy compensation: hydrophobic effect, micelles and protein complexes. *Phys Chem Chem Phys* 6:4156. doi: 10.1039/b404327h

Flores K, Wolschin F, Corneveaux JJ, Allen AN, Huentelman MJ and Amdam G V. (2012) Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*. doi: 10.1186/1471-2164-13-480

Fuks F, Burgers WA, Brehm A, Hughes-Davies L and Kouzarides T (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet* 24:88–91. doi: 10.1038/71750

Garcia CF, Delprat A, Ruiz A and Valente VLS (2015) Reassignment of *Drosophila willistoni* genome scaffolds to chromosome II arms. *G3 Gene Genomes Genet* 5:2559–2566. doi: 10.1534/g3.115.021311

Garcia RN, D'Ávila MF, Robe LJ, Loreto ELDS, Panzera Y, De Heredia FO and Valente VLS (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica* 131:91–105. doi: 10.1007/s10709-006-9116-3

Genencher B, Durdevic Z, Hanna K, Zinkl D, Mobin MB, Senturk N, Da Silva B, Legrand C, Carré C, Lyko F et al. (2018) Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. *Cell Rep* 22:1861–1874. doi: 10.1016/j.celrep.2018.01.061

Goel P, Rohs R, Lazarovici A, Dantas Machado AC, Rao S, Rastogi C, Bussemaker HJ and Zhou T (2014) Evolving insights on how cytosine methylation affects protein-DNA binding. *Brief Funct Genomics* 14:61–73. doi: 10.1093/bfpg/elu040

Goll MG and Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481–514. doi: 10.1146/annurev.biochem.74.010904.153721

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-LL, Zhang X, Golic KG, Jacobsen SE and Bestor TH (2006a) Methylation of tRNA^{Asp} by the DNA methyltransferase

homolog Dnmt2. *Science* (80-) 311:395–398. doi: 10.1126/science.1120976

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C, Zhang X, Golic KG, Jacobsen SE and Bestor TH (2006b) Methylation of tRNA Asp by the DNA Methyltransferase Homolog Dnmt2. *Science* 311:395–398. doi: 10.1126/science.1120976

Gowher H, Leismann O and Jeltsch a. (2000a) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *Embo* 19:6918–6923.

Gowher H, Leismann O and Jeltsch A (2000b) DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J* 19:6918–6923. doi: 10.1093/emboj/19.24.6918

Gravio D, James P, Sajjadi S, Tomar AS, Shrestha S, Issarapu P and Yadav DK (2018) Original article Candidate genes linking maternal nutrient exposure to offspring health via DNA methylation : a review of existing evidence in humans with specific focus on one-carbon metabolism. doi: 10.1093/ije/dyy153

Gupta T, Morgan HR, Andrews JC, Brewer ER and Certel SJ (2017) Methyl-CpG binding domain proteins inhibit interspecies courtship and promote aggression in *Drosophila*. *Sci Rep* 7:1–12. doi: 10.1038/s41598-017-05844-6

Han H, Cortez CC, Yang X, Nichols PW, Jones PA and Liang G (2011) DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum Mol Genet* 20:4299–4310. doi: 10.1093/hmg/ddr356

Hanck T, Schmidt S, Fritz H, Genetik M and Gottingen G (1993) Sequence-specific and mechanism-based crosslinking of Dem DNA cytosine-C methyltransferase of *E. coli* K-12 to synthetic oligonucleotides containing 5-fluoro-2'-deoxycytidine. 21:303–309.

Hardy TM and Tollefsbol TO (2011) Epigenetic diet: impact on the epigenome and cancer. *Epigenomics* 3:503–518. doi: 10.2217/epi.11.71

Hermann A, Gowher H and Jeltsch A (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol Life Sci* 61:2571–2587. doi: 10.1007/s00018-004-4201-1

Hermann A, Schmitt S and Jeltsch A (2003) The human Dnmt2 has residual DNA-(Cytosine-C5) methyltransferase activity. *J Biol Chem* 278:31717–31721. doi: 10.1074/jbc.M305448200

- Humphrey W, Dalke A and Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38. doi: 10.1016/0263-7855(96)00018-5
- Hurd PJ, Whitmarsh AJ, Baldwin GS, Kelly SM, Waltho JP, Price NC, Connolly BA and Hornby DP (1999) Mechanism-based inhibition of C5-cytosine DNA methyltransferases by 2-H pyrimidinone. *J Mol Biol* 286:389–401. doi: 10.1006/jmbi.1998.2491
- Jeltsch A (2002) Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem* 3:274–293. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S
- Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M and Helm M (2016) Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol* 6286:00–00. doi: 10.1080/15476286.2016.1191737
- Jeltsch A, Nellen W and Lyko F (2006) Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 31:306–308. doi: 10.1016/j.tibs.2006.04.005
- Jin B, Li Y and Robertson KD (2011) DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes and Cancer* 2:607–617. doi: 10.1177/1947601910393957
- Johannsson S, Neumann P, Wulf A, Welp LM, Gerber H-D, Krull M, Diederichsen U, Urlaub H and Ficner R (2018) Structural insights into the stimulation of *S. pombe* Dnmt2 catalytic efficiency by the tRNA nucleoside queuosine Dnmt2 methylates cytosine at position 38 of tRNA. 1–10. doi: 10.1038/s41598-018-27118-5
- Jones L, Hamilton AJ, Voinnet O, Thomas CL, Andrew J, Baulcombe DC and Baulcombe DC (1999) RNA-DNA Interactions and DNA Methylation in Post-Transcriptional Gene Silencing. *Plant Cell* 11:2291–2301.
- Jurkowska RZ and Jeltsch A (2016) Enzymology of mammalian DNA methyltransferases. *Adv Exp Med Biol* 945:87–122. doi: 10.1007/978-3-319-43624-1_5
- Jurkowski TP and Jeltsch A (2011) On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. *PLoS One* 6:1–9. doi: 10.1371/journal.pone.0028104
- Jurkowski TP, Shanmugam R, Helm M and Jeltsch A (2012) Mapping the tRNA binding site on the surface of human DNMT2 methyltransferase. *Biochemistry* 51:4438–4444. doi: 10.1021/bi3002659

Kaiser S, Jurkowski TP, Kellner S, Schneider D, Helm M, Kaiser S, Jurkowski TP, Kellner S, Schneider D, Kaiser S et al. (2016) The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biol* 0:1–11. doi: 10.1080/15476286.2016.1236170

Kaplan M, Narasimhan S, de Heus C, Mance D, van Doorn S, Houben K, Popov-Čeleketić D, Damman R, Katrukha EA, Jain P et al. (2016) EGFR Dynamics Change during Activation in Native Membranes as Revealed by NMR. *Cell* 167:1241-1251.e11. doi: 10.1016/j.cell.2016.10.038

Katoh M, Curk T, Xu Q, Zupan B, Kuspa A and Shaulsky G (2006) Developmentally Regulated DNA Methylation in *Dictyostelium discoideum*. *Am Soc Microbiol* 5:18–25. doi: 10.1128/EC.5.1.18

Kautiainen TL and Jones PA (1986) DNA methyltransferase levels in tumorigenic and nontumorigenic cells in culture. *J Biol Chem* 261:1594–1598.

Kiani J, Grandjean V, Liebers R, Tuorto F, Ghanbarian H, Lyko F, Cuzin F and Rassoulzadegan M (2013) RNA-Mediated Epigenetic Heredity Requires the Cytosine Methyltransferase Dnmt2. *PLoS Genet*. doi: 10.1371/journal.pgen.1003498

Kiefer F, Arnold K, Künzli M, Bordoli L and Schwede T (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* 37:387–392. doi: 10.1093/nar/gkn750

Kim G Do, Ni J, Kelesoglu N, Roberts RJ and Pradhan S (2002) Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J* 21:4183–4195. doi: 10.1093/emboj/cdf401

Klenov MS, Lavrov SA, Stolyarenko AD, Sergey S, Aravin AA, Tuschl T and Gvozdev VA (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. *35:5430–5438*. doi: 10.1093/nar/gkm576

Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994a) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–69.

Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994b) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–369. doi: 10.1016/0092-8674(94)90342-

Klimasauskas S, Kumar S, Roberts RJ, Cheng X, Klimasauskas S, Salius Kumar, Sanjay Roberts, Richard Cheng X, Klimasauskas S, Kumar S, Roberts RJ and Cheng X (1994c) HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76:357–369. doi: 10.1016/0092-8674(94)90342-5

Klose RJ and Bird AP (2006) Genomic DNA methylation: The mark and its mediators. *Trends Biochem Sci* 31:89–97. doi: 10.1016/j.tibs.2005.12.008

Kramer JM, Kochinke K, Oortveld MAW, Marks H, Kramer D, de Jong EK, Asztalos Z, Westwood JT, Stunnenberg HG, Sokolowski MB et al. (2011) Epigenetic regulation of learning and memory by *Drosophila* EHMT/G9a. *PLoS Biol.* doi: 10.1371/journal.pbio.1000569

Krauss V and Reuter G (2011) DNA Methylation in *Drosophila*-a critical evaluation. *Prog Mol Biol Transl Sci* 101:177–191. doi: 10.1016/B978-0-12-387685-0.00003-2

Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, Söderbom F et al. (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. *Nucleic Acids Res* 33:6405–6417. doi: 10.1093/nar/gki952

Kumar R and Rao DN (2011) A nucleotide insertion between two adjacent methyltransferases in *Helicobacter pylori* results in a bifunctional DNA methyltransferase. *Biochem J* 433:487–495. doi: 10.1042/BJ20101668

Kumar S, Cheng X, Klimasauskas S, Sha M, Posfai J, Roberts RJ and Wilson GG (1994) The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res* 22:1–10. doi: 10.1093/nar/22.1.1

Kunert N, Marhold J, Stanke J, Stach D and Lyko F (2003) A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development* 130:5083–5090. doi: 10.1242/dev.00716

La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH and Trasler JM (2004) Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev Biol* 268:403–415. doi: 10.1016/j.ydbio.2003.12.031

Lachaise D, Cariou M, David JR, Lemeunier F, Tsacas L and Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol Biol* 22:159–225. doi: 10.1007/978-1-4613-0931-4_4

- Laskowski R, Rullmann JA, MacArthur M, Kaptein R and Thornton J (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486. doi: 10.1007/BF00228148
- Lee T -f., Zhai J and Meyers BC (2010) Conservation and divergence in eukaryotic DNA methylation. *Proc Natl Acad Sci* 107:9027–9028. doi: 10.1073/pnas.1005440107
- Lewinska A, Adamczyk-Grochala J, Kwasniewicz E, Deregowska A, Semik E, Zabek T and Wnuk M (2018) Reduced levels of methyltransferase DNMT2 sensitize human fibroblasts to oxidative stress and DNA damage that is accompanied by changes in proliferation-related miRNA expression. *Redox Biol* 14:20–34. doi: 10.1016/j.redox.2017.08.012
- Lewinska A, Adamczyk-Grochala J, Kwasniewicz E and Wnuk M (2017) Downregulation of methyltransferase Dnmt2 results in condition-dependent telomere shortening and senescence or apoptosis in mouse fibroblasts. *J Cell Physiol* 232:3714–3726. doi: 10.1002/jcp.25848
- Li E and Zhang Y (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6:a019133. doi: 10.1101/cshperspect.a019133
- Li S and Mason CE (2014) The Pivotal Regulatory Landscape of RNA Modifications. *Annu Rev Genomics Hum Genet* 15:127–150. doi: 10.1146/annurev-genom-090413-025405
- Liebers R, Rassoulzadegan M and Lyko F (2014) Epigenetic Regulation by Heritable RNA. *PLoS Genet*. doi: 10.1371/journal.pgen.1004296
- Lin M-JJ, Tang L-YY, Reddy MN and Shen C-KKJJ (2005) DNA methyltransferase gene *dDnmt2* and longevity of *Drosophila*. *J Biol Chem* 280:861–864. doi: 10.1074/jbc.C400477200
- Lindsley DL, Sandler L, Counce SJ, Chandley AC and Lewis KR (1977) The Genetic Analysis of Meiosis in Female *Drosophila melanogaster*. *Philos Trans R Soc Lond B Biol Sci* 277:295–312. doi: 10.2307/2417716
- Lipari G and Szabo A (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J Am Chem Soc* 104:4559–4570. doi: 10.1021/ja00381a010

- Liu K, Wang YF, Cantemir C and Muller MT (2003) Endogenous Assays of DNA Methyltransferases: Evidence for Differential Activities of DNMT1, DNMT2, and DNMT3 in Mammalian Cells In Vivo Endogenous Assays of DNA Methyltransferases: Evidence for Differential Activities of DNMT1, DNMT2, and DNMT3 in M. *Mol Cell Biol* 23:2709. doi: 10.1128/MCB.23.8.2709
- Livak KJ and Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25:402–408. doi: 10.1006/METH.2001.1262
- Lo P-K and Sukumar S (2008) Epigenomics and breast cancer. *Pharmacogenomics* 9:1879–902. doi: 10.2217/14622416.9.12.1879
- Lyko F (2001) DNA methylation learns to fly. *Trends Genet* 17:169–172. doi: 10.1016/S0168-9525(01)02234-X
- Lyko F, Ramsahoye BH and Jaenisch R (1985) DNA methylation in *Drosophila melanogaster*. *Nature* 19:1065–1068. doi: 10.1002/(SICI)1097-0177(199909)216:1<1::AID-DVDY1>3.0.CO;2-T
- Lyko F, Whittaker AJ, Orr-Weaver TL and Jaenisch R (2000) The putative *Drosophila* methyltransferase gene *dDnmt2* is contained in a transposon-like element and is expressed specifically in ovaries. *Mech Dev* 95:215–217. doi: 10.1016/S0925-4773(00)00325-7
- Lyons SM, Fay MM and Ivanov P (2018) The role of RNA modifications in the regulation of tRNA cleavage. *FEBS Lett*. doi: 10.1002/1873-3468.13205
- MacRaidl CA, Daranas AH, Bronowska A and Homans SW (2007) Global Changes in Local Protein Dynamics Reduce the Entropic Cost of Carbohydrate Binding in the Arabinose-binding Protein. *J Mol Biol* 368:822–832. doi: 10.1016/j.jmb.2007.02.055
- Marhold J (2004) The *Drosophila* MBD2/3 protein mediates interactions between the MI-2 chromatin complex and CpT/A-methylated DNA. *Development* 131:6033–6039. doi: 10.1242/dev.01531
- Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B and Lyko F (2004) Conservation of DNA methylation in dipteran insects. *Insect Mol Biol* 13:117–123. doi: 10.1111/j.0962-1075.2004.00466.x

- Marhold J, Zbylut M, Lankenau DH, Li M, Gerlich D, Ballestar E, Mechler BM and Lyko F (2002) Stage-specific chromosomal association of *Drosophila* dMBD2/3 during genome activation. *Chromosoma* 111:13–21. doi: 10.1007/s00412-002-0188-2
- Marlow MS, Dogan J, Frederick KK, Valentine KG and Wand AJ (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6:352–358. doi: 10.1038/nchembio.347
- Marti-Renom MA, Stuart AC, Sanchez R, Melo F and Sali A (2000) Comparative Protein Structure Modeling of Genes and Genomes. *Annu Rev Biophys Biomol Struct* 29:291–325.
- Matzke MA and Birchler JA (2005) RNAi-mediated pathways in the nucleus. *Nat Rev Genet* 6:24–35. doi: 10.1038/nrg1500
- Matzke MA and Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat Rev Genet* 15:394–408. doi: 10.1038/nrg3683
- Mertineit C, Yoder J a, Taketo T, Laird DW, Trasler JM and Bestor TH (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* 125:889–897.
- Mette MFF, Aufsatz W, van der Winden J, Matzke MAMA and Matzke AJMAJM (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* 19:5194–5201. doi: 10.1093/emboj/19.19.5194
- Motorin Y, Lyko F and Helm M (2009) 5-methylcytosine in RNA: Detection, enzymatic formation and biological functions. *Nucleic Acids Res* 38:1415–1430. doi: 10.1093/nar/gkp1117
- Müller S, Windhof IM, Maximov V, Jurkowski T, Jeltsch A, Förstner KU, Sharma CM, Gräf R and Nellen W (2013) Target recognition, RNA methylation activity and transcriptional regulation of the *Dictyostelium discoideum* Dnmt2-homologue (DnmA). *Nucleic Acids Res* 41:8615–8627. doi: 10.1093/nar/gkt634
- Mund C, Musch T, Strödicke M, Assmann B, Li E and Lyko F (2004) Comparative analysis of DNA methylation patterns in transgenic *Drosophila* overexpressing mouse DNA methyltransferases. *Biochem J* 378:763–8. doi: 10.1042/BJ20031567
- Nagasundaram N, Zhu H, Liu J, Karthick V, George Priya Doss C, Chakraborty C and

- Chen L (2015) Analysing the effect of mutation on protein function and discovering potential inhibitors of CDK4: Molecular modelling and dynamics studies. *PLoS One* 10:1–21. doi: 10.1371/journal.pone.0133969
- Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN and Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393:386–389. doi: 10.1038/30764
- O’Gara M, Klimasauskas S, Roberts RJ and Cheng X (1996a) Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes. *J Mol Biol* 261:634–645. doi: 10.1006/jmbi.1996.0489
- O’Gara M, Roberts RJ and Cheng X (1996b) A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase. *J Mol Biol* 263:597–606. doi: 10.1006/jmbi.1996.0601
- Okano M, Bell DW, Haber DA and Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257. doi: 10.1016/S0092-8674(00)81656-6
- Okano M, Xie S and Li E (1998) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* 26:2536–2540. doi: 10.1093/nar/26.11.2536
- Okazaki K -i. and Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci* 105:11182–11187. doi: 10.1073/pnas.0802524105
- Ørom UA and Shiekhattar R (2011) Long non-coding RNAs and enhancers. *Curr Opin Genet Dev* 21:194–198. doi: 10.1016/j.gde.2011.01.020
- Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA and Elgin SC. (2004) Heterochromatic Silencing and HP1 Localization in *Drosophila* Are Dependent on the RNAi Machinery. *Science* (80-) 303:669–672. doi: 10.1038/098448b0
- Panikar CS, Rajpathak SN, Abhyankar V, Deshmukh S and Deobagkar DD (2015) Presence of DNA methyltransferase activity and CpC methylation in *Drosophila melanogaster*. *Mol Biol Rep* 42:1615–1621. doi: 10.1007/s11033-015-3931-5

- Park YJ, Claus R, Weichenhan D and Plass C (2011) Genome-wide epigenetic modifications in cancer. *Prog Drug Res* 67:25–49. doi: 10.1007/978-3-7643-8989-5_2
- Patel C V. and Gopinathan KP (1987) Determination of trace amounts of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal Biochem* 164:164–169. doi: 10.1016/0003-2697(87)90381-2
- Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, Ferguson D, Seibel G and Kollman P (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* 91:1–41. doi: 10.1016/0010-4655(95)00041-D
- Pendina AA, Efimova OA, Tkhonov A V., Chiryaeva OG, Fedorova ID, Koltsova AS, Krapivin MI, Parfenyev SE, Kuznetzova T V. and Baranov VS (2016) Immunofluorescence Staining for Cytosine Modifications Like 5-Methylcytosine and Its Oxidative Derivatives and FISH. *Encycl Plant Crop Sci* 468–471. doi: 10.1081/e-epcs-120005681
- Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi: 10.1002/jcc.20084
- Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC and Reuter G (2009) Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* 41:696–702. doi: 10.1038/ng.360
- Pikkemaat MG, Linssen ABM, Berendsen HJC and Janssen DB (2002) Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng Des Sel* 15:185–192. doi: 10.1093/protein/15.3.185
- Ponger L and Li W-H (2005) Evolutionary Diversification of DNA Methyltransferases in Eukaryotic Genomes. *Mol Biol Evol* 22:1119–1128. doi: 10.1093/molbev/msi098
- Pradhan S and Esteve PO (2003) Mammalian DNA (cytosine-5) methyltransferases and their expression. *Clin Immunol* 109:6–16. doi: 10.1016/S1521-6616(03)00204-3
- Prokhortchouk E and Defossez PA (2008) The cell biology of DNA methylation in

mammals. *Biochim Biophys Acta - Mol Cell Res* 1783:2167–2173. doi: 10.1016/j.bbamcr.2008.07.015

Pronk S, P??il S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29:845–854. doi: 10.1093/bioinformatics/btt055

Rai K, Chidester S, Zavala C V., Manos EJ, James SR, Karpf AR, Jones DA and Cairns BR (2007) Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. *Genes Dev* 21:261–266. doi: 10.1101/gad.1472907

Rebollo R, Horard B, Begeot F, Delattre M, Gilson E and Vieira C (2012) A Snapshot of Histone Modifications within Transposable Elements in *Drosophila* Wild Type Strains. *PLoS One* 7:1–7. doi: 10.1371/journal.pone.0044253

Reichetzeder C, Kleuser B, Dwi Putra SE, Neuber C and Hocher B (2014) Analysis of Genomic DNA Methylation Levels in Human Placenta using Liquid Chromatography-Electrospray Ionization Tandem Mass Spectrometry. *Cell Physiol Biochem* 33:945–952. doi: 10.1159/000358666

Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447:425–432. doi: 10.1038/nature05918

Reinisch KM, Chen L, Verdine GL and Lipscomb WN (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* 82:143–153. doi: 10.1016/0092-8674(95)90060-8

Renkawitz-Pohl R and Bialojan S (1984) A DNA sequence of *Drosophila melanogaster* with a differential telomeric distribution. *Chromosoma* 89:206–11.

Riddle NC, Minoda A, Kharchenko P V, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-basso D et al. (2011) Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. 4:147–163. doi: 10.1101/gr.110098.110.due

Robe LJ, Cordeiro J, Loreto ELS and Valente VLS (2010) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genetica* 138:601–617. doi: 10.1007/s10709-009-9432-5

Roberts PA (1979) Rapid change of chromomeric and pairing patterns of polytene chromosome tips in *D. melanogaster*: Migration of polytene-non-polytene transition zone? *Genetics* 92:

Roberts RJ, Vincze T, Posfai J and Macelis D (2015) REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299. doi: 10.1093/nar/gku1046

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL and Wolffe AP (2000) DNMT1 forms a complex with RB, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat Genet* 25:338–342. doi: 10.1038/77124

Roder K, Hung MS, Lee TL, Lin TY, Xiao H, Isobe KI, Juang JL and Shen CJ (2000) Transcriptional repression by *Drosophila* methyl-CpG-binding proteins. *Mol Cell Biol* 20:7401–7409. doi: 10.1128/MCB.20.19.7401-7409.2000

Rohde C and Valente VLS (2012) Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. *Genet Mol Biol* 35:966–979. doi: 10.1590/S1415-47572012000600012

Rountree MR, Bachman KE and Baylin SB (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat Genet* 25:269–277. doi: 10.1038/77023

Rubin GM (1978) Isolation of a telomeric DNA sequence from *Drosophila melanogaster*. *Cold Spring Harb Symp Quant Biol* 42 Pt 2:1041–6.

Saito Y, Kanai Y, Nakagawa T, Sakamoto M, Saito H, Ishi H and Hirohashi S (2003) Increased protein expression of DNA methyltransferase (DNMT) 1 is significantly correlated with the malignant potential and poor prognosis of human hepatocellular carcinomas. *Int J Cancer* 105:527–532. doi: 10.1002/ijc.11127

Salzberg A, Fisher O, Siman-Tov R and Ankri S (2004) Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochem Biophys Res Commun* 322:465–469. doi: 10.1016/j.bbrc.2004.07.134

Sanchez R and Mackenzie SA (2016) Information thermodynamics of cytosine DNA methylation. *PLoS One* 11:1–20. doi: 10.1371/journal.pone.0150427

Sánchez R and Šali A (1999) Comparative Protein Structure Modeling in Genomics. *J*

Comput Phys 151:388–401. doi: 10.1006/JCPH.1999.6200

Santi D, Garrett C and Barr P (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* 33:9–10. doi: 10.1016/0092-8674(83)90327-6

Santos Filho OA and Bicca De Alencastro R (2003) Modelagem de proteínas por homologia. *Quim Nova* 26:253–259.

Sassi AK, Herédia F, Loreto ÉL da S, Valente VLS and Rohde C (2005) Transposable elements P and gypsy in natural populations of *Drosophila willistoni*. *Genet Mol Biol* 28:734–739. doi: 10.1590/S1415-47572005000500013

Schaefer M and Lyko F (2010) Solving the Dnmt2 enigma. *Chromosoma* 119:35–40. doi: 10.1007/s00412-009-0240-6

Schaefer M and Lyko F (2007) DNA methylation with a sting: An active DNA methylation system in the honeybee. *BioEssays* 29:208–211. doi: 10.1002/bies.20548

Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M and Lyko F (2010) RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev* 24:1590–1595. doi: 10.1101/gad.586710

Schaefer M, Steringer JP and Lyko F (2008) The *Drosophila* cytosine-5 methyltransferase Dnmt2 is associated with the nuclear matrix and can access DNA during mitosis. *PLoS One*. doi: 10.1371/journal.pone.0001414

Schaefer MR (2018) tRNA modifications: Necessary for correct tRNA-derived fragments during the recovery from stress? doi: 10.1002/bies.201200158

Schaeffer SW, Bhutkar A, McAllister BF, Matsuda M, Matzkin LM, O'Grady PM, Rohde C, Valente VLS, Aguadé M, Anderson WW et al. (2008) Polytene chromosomal maps of 11 *Drosophila* species: The order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179:1601–1655. doi: 10.1534/genetics.107.086074

Schulz EC, Roth HM, Ankri S and Ficner R (2012) Structure analysis of *Entamoeba histolytica* DNMT2 (EhMeth). *PLoS One*. doi: 10.1371/journal.pone.0038728

Serman A, Vlahović M, Serman L and Bulić-Jakus F (2006) DNA methylation as a regulatory mechanism for gene expression in mammals. *Coll Antropol* 30:665–71.

- Severin PMD, Zou X, Gaub HE and Schulten K (2011) Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res* 39:8740–8751. doi: 10.1093/nar/gkr578
- Shanmugam R, Fierer J, Kaiser S, Helm M, Jurkowski TP and Jeltsch A (2015) Cytosine methylation of tRNA-Asp by DNMT2 has a role in translation of proteins containing poly-Asp sequences. *Cell Discov* 1:1–10. doi: 10.1038/celldisc.2015.10
- Shen L, Gao G, Zhang Y, Zhang H, Ye Z, Huang S, Huang J and Kang J (2010) A single amino acid substitution confers enhanced methylation activity of mammalian Dnmt3b on chromatin DNA. *Nucleic Acids Res* 38:6054–6064. doi: 10.1093/nar/gkq456
- Sisi Li, Jiamu Du, Hui Yang, Juan Yin, Jianping Ding JZ (2012) Functional and structural characterization of DNMT 2 from *Spodoptera frugiperda*. *J Mol Cell Biol* 5:64–66.
- Spitale RC, Tsai MC and Chang HY (2011) RNA templating the epigenome: Long noncoding RNAs as molecular scaffolds. *Epigenetics* 6:539–543. doi: 10.4161/epi.6.5.15221
- Stach D, Schmitz OJ, Stilgenbauer S, Benner A, Döhner H, Wiessler M and Lyko F (2003) Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res* 31:E2. doi: 10.1093/nar/gng002
- Stockmann H, Bronowska A, Syme NR, Thompson GS, Kalverda AP, Warriner SL and Homans SW (2008) Residual Ligand Entropy in the Binding of p-Substituted Benzenesulfonamide Ligands to Bovine Carbonic Anhydrase II. *J Am Chem Soc* 130:12420–12426. doi: 10.1021/ja803755m
- Suzuki MM and Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465–476. doi: 10.1038/nrg2341
- Swaminathan CP, Sankpal UT, Rao DN and Surolia A (2002) Water-assisted dual mode cofactor recognition by HhaI DNA methyltransferase. *J Biol Chem* 277:4042–4049. doi: 10.1074/jbc.M109237200
- Sweatt JD (2009) Experience-Dependent Epigenetic Modifications in the Central Nervous System. *Biol Psychiatry* 65:191–197. doi: 10.1016/j.biopsych.2008.09.002
- Swinehart W and Jackman J (2015) Diversity in mechanism and function of tRNA methyltransferases. *RNA Biol* 6286:37–41. doi: 10.1080/15476286.2015.1008358

- Takayama S, Dhabhi J, Roberts A, Mao G, Heo SJ, Pachter L, Martin DIK and Boffelli D (2014) Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res* 24:821–830. doi: 10.1101/gr.162412.113
- Tang LY, Reddy MN, Rasheva V, Lee TL, Lin MJ, Hung MS and Shen CKJ (2003) The Eukaryotic DNMT2 Genes Encode a New Class of Cytosine-5 DNA Methyltransferases. *J Biol Chem* 278:33613–33616. doi: 10.1074/jbc.C300255200
- Thiagarajan D, Dev RR and Khosla S (2011) The DNA methyltransferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics* 6:103–113. doi: 10.4161/epi.6.1.13418
- Throckmorton L (1975) *The phylogeny, ecology and geography of Drosophila*, 3ed ed. Plenum Press, New York
- Tovy A, Hofmann B, Helm M and Ankri S (2010) In vitro tRNA methylation assay with the *Entamoeba histolytica* DNA and tRNA methyltransferase Dnmt2 (Ehmeth) enzyme. *J Vis Exp* 2:2–7. doi: 10.3791/2390
- Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, Blankenship J, Collins C, Giorgi D, Iadonato S, Johnson F et al. (1998) Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* 7:13–26.
- Traverse KL and Pardue ML (1988) A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc Natl Acad Sci U S A* 85:8116–20.
- Tsankova N, Renthal W, Kumar A and Nestler EJ (2007) Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci* 8:355–367. doi: 10.1038/nrn2132
- Tuorto F, Liebers R, Musch T, Schaefer M, Hofmann S, Kellner S, Frye M, Helm M, Stoecklin G and Lyko F (2012) RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat Struct Mol Biol* 19:900–905. doi: 10.1038/nsmb.2357
- Tweedie S, Ng HH, Barlow AL, Turner BM, Hendrich B and Bird A (1999) Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat Genet* 23:389–390. doi: 10.1038/70490

- Valente VLS and Araújo AM (1986) Chromosomal polymorphism, climatic factors, and variation in population size of *Drosophila willistoni* in southern Brazil. *Heredity* (Edinb) 57:149–159. doi: 10.1038/hdy.1986.105
- Valente VLS, Goñi B, Valiati VH, Rohde C and Basilio Morales N (2003) Chromosomal polymorphism in *Drosophila willistoni* populations from Uruguay. *Genet Mol Biol* 26:163–173. doi: 10.1590/S1415-47572003000200009
- Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJC (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718. doi: 10.1002/jcc.20291
- Varriale A (2014) DNA Methylation, Epigenetics, and Evolution in Vertebrates: Facts and Challenges. 2014:1–7. doi: 10.1155/2014/475981
- Vicario S, Moriyama EN and Powell JR (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol Biol* 7:1–17. doi: 10.1186/1471-2148-7-226
- Vieira GC, D'Ávila MF, Zanini R, Deprá M and Valente VLS (2018) Evolution of DNMT2 in drosophilids: Evidence for positive and purifying selection and insights into new protein (pathways) interactions. *Genet Mol Biol*. doi: 10.1590/1678-4685-gmb-2017-0056
- Vieira GC, Sinigaglia M, Vieira GF and Valente VLS (2017) Linking epigenetic function to electrostatics: The DNMT2 structural model example. *PLoS One* 12:1–19. doi: 10.1371/journal.pone.0178643
- Vilkaitis G, Merkiene E, Serva S, Weinhold E and Klimašauskas S (2001) The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of HhaI methyltransferase. *J Biol Chem* 276:20924–20934. doi: 10.1074/jbc.M101429200
- Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS and Martienssen RA (2002) Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* (80-) 297:1833–1837. doi: 10.1038/098448b0
- Vujatovic O, Zaragoza K, Vaquero A, Reina O, Bernués J and Azorín F (2012) *Drosophila melanogaster* linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic Acids Res* 40:5402–5414. doi: 10.1093/nar/gks224
- Waddington CH (1942) The Epigenotype. *Endeavour* 1:18–20. doi: 10.1093/ije/dyr184

- Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA and Robinson GE (2006) Functional CpG methylation system in a social insect. *Science* 314:645–7. doi: 10.1126/science.1135213
- Wassenegger M (2004) RNA-directed DNA methylation. *Plant Mol Biol* 43:203–220. doi: 10.1016/j.pbi.2011.02.003
- Williams RW and Rubin GM (2002) ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc Natl Acad Sci USA* 99:6889–6894. doi: 10.1073/pnas.072190799
- Wilson RJ, Goodman JI and Strelets VB (2008) FlyBase: Integration and improvements to query tools. *Nucleic Acids Res* 36:588–593. doi: 10.1093/nar/gkm930
- Wu JC and Santi DV (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J Biol Chem* 262:4778–4786. doi: 10.1016/S0959-440X(97)80013-9
- Wyszynski MW, Gabbara S and Bhagwat AS (1991) Substitutions of cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. *Nucleic Acids Res* 20:319–326.
- Wyszynski MW, Gabbara S, Kubareva EA, Romanova EA, Oretskaya TS, Gromova ES, Shabarova ZA and Bhagwat AS (1993) The cysteine conserved among DNA cytosine methylases required for methyl transfer, but not for specific DNA binding. *Nucleic Acids Res* 21:295–301. doi: 10.1093/nar/21.2.295
- Xia B, Gerstin E, Schones DE, Huang W and de Belle JS (2016) Transgenerational programming of longevity through E(z)-mediated histone H3K27 trimethylation in *Drosophila*. *Aging (Albany NY)* 8:2988–3008. doi: 10.18632/aging.101107
- Yang D and Kay LE (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J Mol Biol* 263:369–382. doi: 10.1006/jmbi.1996.0581
- Ye Y, Li M, Gu L, Chen X, Shi J, Zhang X and Jiang C (2017) Chromatin remodeling during in vivo neural stem cells differentiating to neurons in early *Drosophila* embryos. *Cell Death Differ* 24:409–420. doi: 10.1038/cdd.2016.135
- Yoder JA and Bestor TH (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet* 7:279–284. doi: 10.1093/hmg/7.2.279

Yoder JA, Soman NS, Verdine GL and Bestor TH (1997a) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol* 270:385–395. doi: 10.1006/jmbi.1997.1125

Yoder JA, Walsh CP and Bestor TH (1997b) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335–340. doi: 10.1016/S0168-9525(97)01181-5

Young BS, Pession A, Traverse KL, French C and Pardue ML (1983) Telomere regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. *Cell* 34:85–94.

Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, Yin R, Zhang D, Zhang P, Liu J et al. (2015) N6-methyladenine DNA modification in *Drosophila*. *Cell* 161:893–906. doi: 10.1016/j.cell.2015.04.018

Zhao BS, Roundtree IA and He C (2016) Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* 18:31–42. doi: 10.1038/nrm.2016.132

Evolução Biológica: da pesquisa ao ensino

Leonardo Augusto Luvison Araújo (Org.)

φ editora fi

Diagramação e capa: Lucas Fontella Margoni

Ilustração de capa: Internet

Revisão Ortográfica e Textual: Leticia Santos Rodrigues.



Todos os livros publicados pela Editora Fi estão sob os direitos da Creative Commons 4.0 https://creativecommons.org/licenses/by/4.0/deed.pt_BR



<http://www.abecbrasil.org.br>

Dados Internacionais de Catalogação na Publicação (CIP)

ARAUJO, Leonardo Augusto Luvison (Org.).

Evolução Biológica: da pesquisa ao ensino [recurso eletrônico] / Leonardo Augusto Luvison Araújo (Org.) - Porto Alegre, RS: Editora Fi, 2017.

519 p.

ISBN - 978-85-5696-240-9

Disponível em: <http://www.editorafi.org>

1. Evolução Biológica; 2. Ensino; 3. Divulgação Científica. 4. História da Ciência. I: Curso de Biologia Evolutiva da UFRGS. I. Título.

CDD-570

Índices para catálogo sistemático:

1. Biologia 570

SUMÁRIO

PREFÁCIO	11
Russel Teresinha Dutra da Rosa	
INTRODUÇÃO	23
A COMPREENSÃO DE EVOLUÇÃO BIOLÓGICA NO BRASIL: O TRÍPLIO PROBLEMA	
Leonardo Augusto Luvison Amújo	
SEÇÃO I EVOLUÇÃO BIOLÓGICA: CONCEITOS, DISCIPLINAS E DEBATES	
1	35
PALEONTOLOGIA E EVOLUÇÃO NO TEMPO PROFUNDO	
Voltaire Dutra Poes Neto; Marcel Brêta Lucinda Santos; Tomaz Poncetti Melo	
2	61
A HISTÓRIA EVOLUTIVA DOS HOMINÍNEOS: DO BIPEDALISMO AO SIMBOLISMO	
Luís Henrique Visconti	
3	85
PSICOLOGIA EVOLUTIVA E GENÉTICA: ASPECTOS EVOLUTIVOS DA DEPRESSÃO, ESQUIZOFRENIA E SEXUALIDADE	
Luís Henrique Visconti	
4	103
SELEÇÃO SEXUAL: OUTRO MECANISMO EVOLUTIVO?	
André Luis Klein	
5	139
EVOLUÇÃO DA FOSSORIALIDADE EM UMA LINHAGEM DE LAGARTOS DAS CAATINGAS BRASILEIRAS: ESPECIAÇÃO, ADAPTAÇÃO, ÓRGÃOS VESTIGIAIS E DESENVOLVIMENTO EMBRIONÁRIO	
Miguel Tereza Rodrigues	
6	147
AS FRONTEIRAS DA DIVERSIDADE: ESPECIAÇÃO E HIBRIDAÇÃO	
André Luis Klein	
7	177
ADMIRÁVEL MUNDO NOVO: A EPIGENÉTICA	
Gilberto Cavallheiro Vieira	

8	213
<hr/>	
EVO-DEVO E A EXPANSÃO DAS FRONTEIRAS DA BIOLOGIA EVOLUTIVA Gilberto Cavallheiro Vieira	
SEÇÃO 2 ENSINO DE EVOLUÇÃO, HISTÓRIA E FILOSOFIA	
9	245
<hr/>	
SISTEMÁTICA FILOGENÉTICA: ABORDANDO A EVOLUÇÃO EM SALA DE AULA Valtierre Dutra Paes Neto; Ronaldo Antonio Paesi	
10	263
<hr/>	
UM CONTO DA CAROCHINHA: POR QUE DARWIN NÃO É PAI SOLTEIRO DA EVOLUÇÃO? Carolina de Souza Diegues	
11	271
<hr/>	
A IMPORTÂNCIA DO TRIBUTO A ALFRED WALLACE PARA O ENSINO DE CIÊNCIAS Marcel Antonio Corrêa Varella; Josmael Coiso	
12	297
<hr/>	
FRITZ MÜLLER, O NATURALISTA DO SÉCULO XIX E POSSÍVEIS RELAÇÕES COM A EDUCAÇÃO NA CONTEMPORANEIDADE Flávia Pacheco Alves de Souza; Andrea Paula dos Santos Oliveira Kamensky	
13	339
<hr/>	
FILOSOFIA, CIÊNCIA E BIOLOGIA: ONDE MORA A RACIONALIDADE? Claudio Ricardo Martins dos Reis	
14	369
<hr/>	
TENTATIVA E ERRO: DA PRÁTICA CIENTÍFICA À PRÁTICA PEDAGÓGICA Daniel Longo Rockenbach; André Luis Klein; Leonardo Augusto Luvison Amêijo	
15	381
<hr/>	
ECOLOGIA E EVOLUÇÃO: A ABORDAGEM EVOLUTIVA EM LIVROS DE ECOLOGIA Claudio Ricardo Martins dos Reis	
16	397
<hr/>	
A TEORIA DA EVOLUÇÃO É UM FIO CONDUTOR NOS LIVROS DIDÁTICOS DE BIOLOGIA DO ENSINO MÉDIO? Douglas Senna Engelke	
17	413
<hr/>	
A VISÃO ANTROPOCÊNTRICA EM UMA REVISTA DE DIVULGAÇÃO CIENTÍFICA PARA CRIANÇAS Camila Leão	

18	445
<hr/>	
UM MISTÉRIO REVELADO: COMO OS CONHECIMENTOS EVOLUTIVOS VÊM SENDO EXPLORADOS NAS AULAS DE BIOLOGIA DO ENSINO MÉDIO? Carolina de Souza Diegues	
19	455
<hr/>	
APRENDENDO EVOLUÇÃO BIOLÓGICA A PARTIR DAS ÁRVORES EVOLUTIVAS Leonardo Augusto Luvison Amújo	
20	473
<hr/>	
SIMULAÇÃO DE MECANISMOS EVOLUTIVOS NO POPG: UM ROTEIRO DE AULA André Luis Klein	
21	481
<hr/>	
A BIOLOGIA EVOLUTIVA E A COMPREENSÃO DO MUNDO MICROBIANO Fernando Bueno Ferreira Fonseca de Fraga	
SOBRE OS AUTORES E AUTORAS	513
<hr/>	
AGRADECIMENTOS	517
<hr/>	

ADMIRÁVEL MUNDO NOVO: A EPIGENÉTICA

Gilberto Cavalheiro Vieira

Introdução

Quando observamos um organismo, como o ser humano, notamos que há uma grande diversidade de tecidos, que são compostos por outra grande diversidade de tipos celulares. Em um indivíduo, as hemácias, os neurônios, as células do fígado ou da pele apresentam formas e funções diferenciadas, apesar de carregarem em seu interior a mesma informação genética. Então, como a partir de uma única célula primordial (o zigoto), temos o desenvolvimento de um organismo portando seus mais de 130 tipos celulares, em que cada tipo possui sua forma e função específica (Figura 1)?

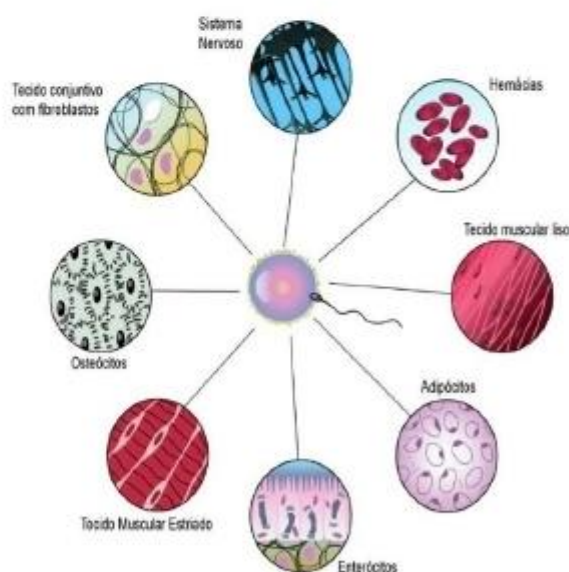


Figura 1. A partir da célula primordial, zigoto, temos o processo de desenvolvimento do organismo, desencadeado por eventos de citodiferenciação e consequente morfogênese. Todas as informações estão contidas no genoma, porém a forma como essas informações são utilizadas (quando, como e quanto) são controladas de maneira diferencial ao longo da ontogenia do organismo.

Salvo raras exceções onde ocorre edição do material genético, as especialidades dessa variedade de tipos celulares que

compõem o corpo humano têm como origem as diferenças **epigenéticas**, e não genética. O termo “epigenética” foi cunhado por Conrad H. Waddington, em 1942, a partir da palavra grega *epigenesis*. Em parceria com Ernst Hadorn, Waddington tinha seus estudos focados em uma área que combinava a genética e o desenvolvimento biológico, procurando descrever e entender os processos genéticos envolvidos no desenvolvimento dos organismos (Figura 2).

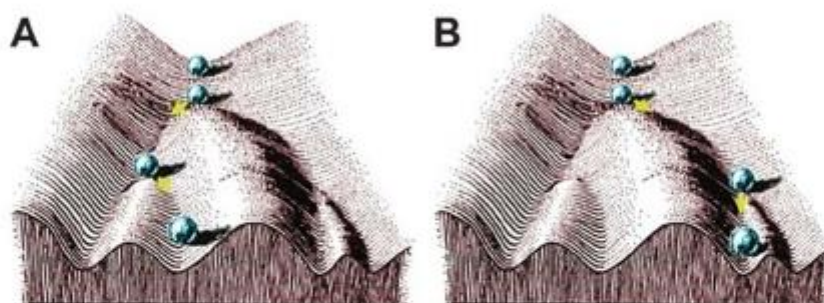


Figura 2. Paisagem de Waddington ou Cenário Epigenético. Essa ilustração representa os caminhos (picos e fendas) percorridos ao longo do processo de diferenciação celular (esfera no topo) até seu destino final de citodiferenciação (esfera abaixo). Modificado de K. Mitchell.

O interessante é que através de suas observações, Waddington notou que fatores ambientais estressores poderiam causar a emergência de fenótipos não esperados em *Drosophila* (Quadro 1), como um processo de assimilação genética (Waddington, 1942), e esses novos fenótipos eram transmitidos para as gerações seguintes, se fixando, mesmo quando cessavam os fatores ambientais estressores – fenômeno esse conhecido como **herança transgeracional**, o qual discutiremos mais adiante. Contudo, apenas no início da década de 1990 foi possível a elucidação dos mecanismos moleculares envolvidos nas observações descritas por Waddington (Figura 3).

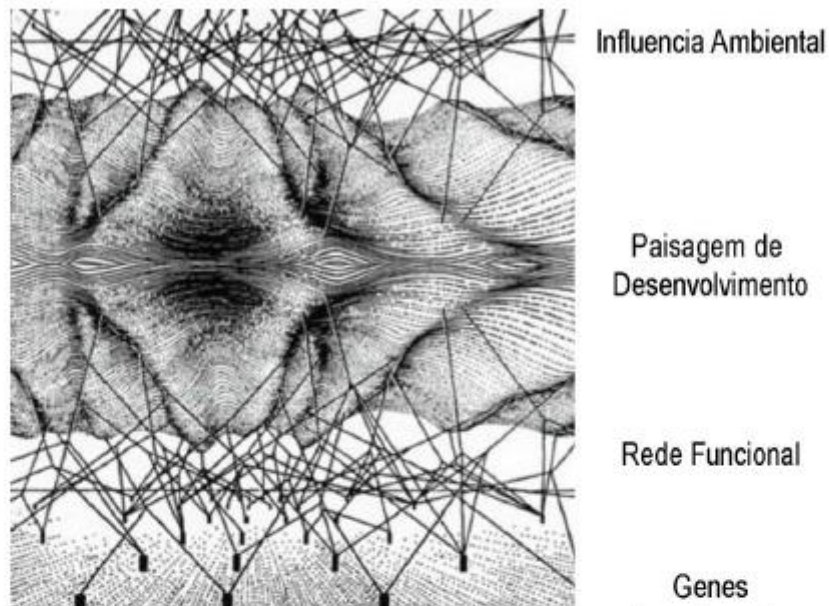


Figura 3. Redes de interações entre ambiente-genes e genes-genes, formando a paisagem de desenvolvimento. Modificado de Waddington, 1942.

Quadro 1. Moscas-da-fruta (*Drosophila*), essas pequenas notáveis das Ciências Biológicas



Drosophila melanogaster. Créditos: Shutterstock

Na ciência, a experimentação é um passo importante (para não dizer fundamental) quando procuramos estabelecer relações e implicações de determinados fenômenos observados. Entretanto, muitos experimentos na área das Ciências Biológicas são inviáveis ou muito complicados de serem observados quando tratamos da questão de tempo de desenvolvimento do organismo, da taxa de

reprodução, do tamanho, da necessidade de controlar variáveis ambientais em que esse organismo estará exposto, etc. Em pesquisas médicas, ficam mais claros os obstáculos éticos e práticos em se testar novos fármacos ou terapias usando seres humanos. Dessa maneira, diversos estudos focam-se no uso de organismos modelos.

Um dos organismos clássicos em estudos nas áreas da genética, hereditariedade, desenvolvimento embrionário, aprendizado, comportamento e envelhecimento são as moscas-da-fruta, *Drosophila*. Apesar de humanos e drosófilas não parecerem tão similares, sabe-se que os mecanismos biológicos mais fundamentais e redes de controle de desenvolvimento e sobrevivência são conservados ao longo da evolução entre essas duas espécies.

O primeiro registro do uso desses organismos em laboratório foi feito por William Castle, em 1901, porém o “pai” das pesquisas com *Drosophila* é inquestionavelmente Thomas Hunt Morgan. Morgan é responsável por “refinar” a teoria da hereditariedade proposta pela primeira vez por Gregor Mendel. Com o uso desses organismos ele foi capaz de definir que os genes estavam de alguma maneira estabelecidos nos cromossomos – e isso muito tempo antes de se demonstrar que o material genético é o DNA.

Nessas cinco décadas, as drosófilas passaram a ser organismo modelo predominante em pesquisas que visavam o entendimento de como os genes dirigem os processos de desenvolvimento embrionário, a partir de uma única célula até chegar ao estágio de um indivíduo maduro multicelular. Em 1995, Christiane Nüsslein-Volhard, Eric Wieschaus e Ed Lewis descobriram os genes-mestres responsáveis pelo controle do desenvolvimento embrionário. O mais interessante é que esses genes-mestres, que desempenham importante papel no desenvolvimento de drosófilas, também são críticos para o desenvolvimento de todos os animais, incluindo a espécie humana.

O que se descobriu é que, apesar da arquitetura corporal de humanos e drosófilas diferenciarem-se enormemente, diversos blocos de construção e processos de engenharia celular que controlam o desenvolvimento embrionário são conservados ao longo da evolução e são incrivelmente similares em suas sequências de DNA entre espécies evolutivamente distantes. Contudo, isso será assunto para um próximo capítulo.

Mas que características conferem a esses pequenos insetos tamanha presença nos estudos científicos?

As drosófilas são excelentes organismos modelos, pois possuem um ciclo de vida complexo, mas de curta duração, possibilitando a análise de diversas gerações em pouco tempo, visto que em um período de 8-14 dias temos novas gerações surgindo. Cada fêmea pode colocar em média de 40 a 60 ovos por dia durante sua vida. A manutenção delas é relativamente simples e barata, se alimentando de frutas fermentadas ou meio de cultura elaborado com farinha de milho e fermento biológico colocados em pequenos vidros, podendo ser estocadas em câmaras com temperatura controlada. Além disso, apesar de serem pequenas, a visualização de fenótipos e mutações é muito simples, bastando uma lupa.

Em questão molecular, a similaridade entre os genes de drosófilas e humanos é alta, pois diversos genes envolvidos em doenças em humanos estão presentes no genoma de drosófilas – aproximadamente 75% desses genes. Outro aspecto de grande interesse é a simplicidade com que se pode criar animais transgênicos que expressam certas proteínas, como a proteína fluorescente verde da água-viva, por exemplo.

Fundamentalmente, o processo epigenético caracteriza-se por contribuir com diferentes fenótipos celulares alternativos emergindo em um organismo a partir de seu genótipo. Esses

fenótipos alternativos podem ser induzidos por determinados cenários ambientais (interno ou externo ao organismo) que, por sua vez, influenciam diferentes estados epigenéticos, de um ou mais loci do genoma do organismo. Sabe-se ainda que a herança epigenética viabiliza a transmissão de informações que não são codificadas pelos genes de uma célula-mãe para a célula-filha, ou de geração para geração. Esse fenômeno ocorre, principalmente, quando observamos uma célula do fígado se dividir e dar origem a duas outras células idênticas. Isso ocorre porque as marcações epigenéticas são herdadas de uma geração de linhagem celular para a próxima geração de células.

Os mecanismos envolvidos nos fenômenos epigenéticos atuam na alteração da expressão gênica, mudando a acessibilidade à cromatina para a regulação transcricional por meio das modificações do DNA, mas sem haver alterações na sequência das bases do genoma. Esses processos são críticos no desenvolvimento normal e na especialização de funções das diferentes células, dos diferentes tecidos de um organismo. Entretanto, as modificações epigenéticas desreguladas podem resultar em efeitos danosos ao organismo, com a emergência de diversas doenças como transtornos mentais, dependência de drogas (Tsankova et al., 2007; Sweatt et al., 2009; Kumar et al., 2008) e câncer (Feinberg et al., 1983; Hardy et al., 2011; Park et al., 2011).

De maneira resumida, as regulações epigenéticas podem ser descritas como as interações mútuas entre três eventos: a) a metilação de DNA; b) as modificações de histonas com a remodelação do nucleossomo (Lo et al., 2008); e c) a ação de RNA não codificantes (Dozmorov et al., 2013; Spitale et al., 2011; Ørom and Shiekhattar, 2011) (Figura 4). Nas próximas seções iremos abordar cada um desses mecanismos epigenéticos.

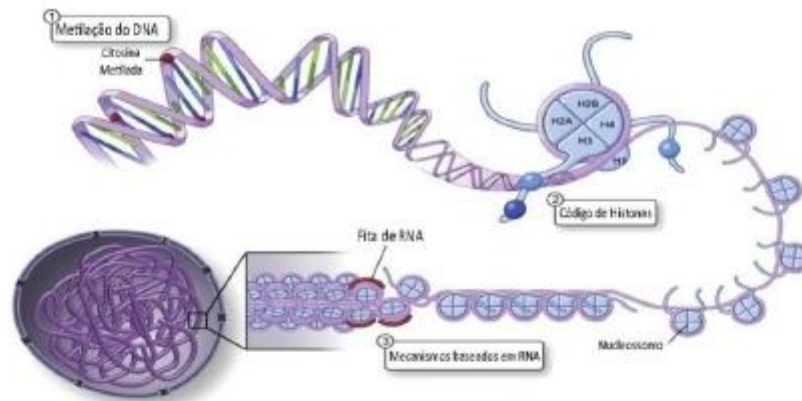


Figura 4. Mecanismos epigenéticos envolvidos no controle de expressão gênica. Modificado de Yan, 2010.

O fenômeno de metilação

A metilação do DNA genômico é um dos principais mecanismos de regulação epigenética nos organismos. O fenômeno de metilação do DNA consiste no processo de substituição de um átomo de hidrogênio (H) por um grupo metil (CH_3), sendo a reação catalisada por diversos tipos de metiltransferases (MTases). O controle de expressão gênica, de *imprinting* genômico, de estabilidade genômica e do silenciamento de transposons e de retrovírus endógenos, por exemplo, são mediados por metilação (Bird, 2002; Yoder et al., 1997) (Figura 5).

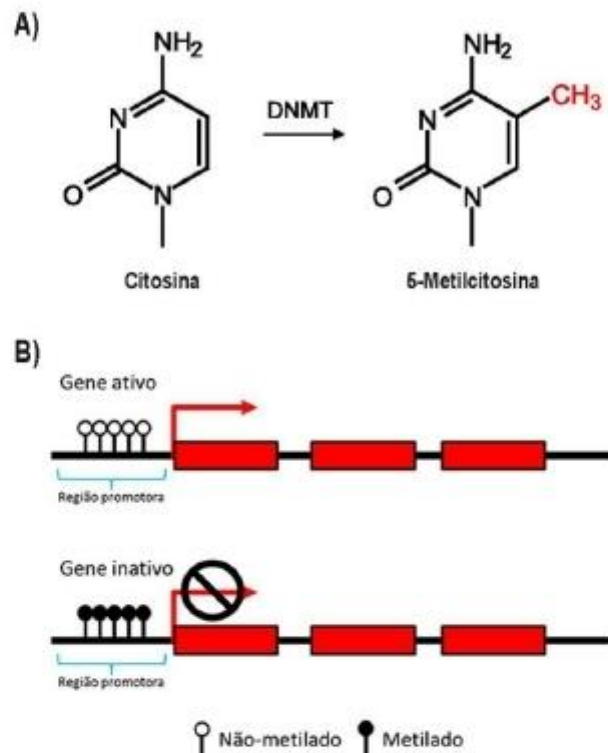


Figura 5. Modificações de citosinas: A) metilação de citosina genômica, modificando o carbono-5 do anel aromático da base nitrogenada pela ação de enzimas DNA-metiltransferases, transformando citosina em 5-metilcitosina (5mC); B) controle de expressão gênica por ação de metilação de citosinas em regiões promotoras de genes, impedindo que fatores de transcrição e a RNA-polimerase possa ancorar de maneira adequada ao DNA e, assim, iniciar a transcrição gênica.

Porém, os mecanismos de controle de expressão gênica envolvendo a metilação não atuam somente de maneira direta na atividade transcricional do gene, podendo atuar por vias indiretas, impedindo fatores de transcrição de acessarem genes metilados ou interagindo com fatores *trans* do processo transcricional, como as histonas deacetilases e proteínas de ligação ao DNA metilado (MDBs) – essas últimas parecem competir por sítios alvos de fatores de transcrição (Nan et al., 1998; Fuks et al., 2000).

Outro aspecto interessante no fenômeno de metilação é que o mesmo não é homogêneo em posição e função ao longo do genoma. Em mamíferos são amplamente descritas metilações nas ilhas CpG nos sítios de início de transcrição como sendo um mecanismo de silenciamento ou diminuição de expressão gênica (Okano et al., 1999); contudo, em genes que não possuem sítios de início de transcrição ricos em CpG pode ocorrer metilação em sequências não CpG (Farthing et al., 2008; Han et al., 2011). As ilhas CpG são regiões do genoma compostas por citosinas e guaninas em sequência como, por exemplo, na sequência: 5'-ATTTCGCGCGCGCGCGCGCATC-3'.

Além disso, a localização das citosinas metiladas ao longo das diversas regiões que compõem um gene pode resultar em diferentes fenômenos, não estando somente envolvidas no processo de silenciamento gênico. Regiões a jusante do sítio promotor geralmente possuem depleção em dinucleotídeos CpG e possuem metilação em elementos repetitivos e no corpo do gene (Weber et al., 2005). O que se observa é que o processo de alongamento transcricional não é impedido pelas citosinas metiladas ao longo do gene, ao contrário do que ocorre no processo de início de transcrição, além delas poderem estar envolvidas no controle de *splicing* (Sharma et al., 2011; Laurent et al., 2010). Dessa forma, os estudos de epigenômica (mais especificamente o metiloma), favorecidos pelo crescente acúmulo de informações advindas dos mais diversos organismos que possuem seus genomas sequenciados, demonstram que os padrões de metilação se estendem para além das regiões promotoras dos genes, sendo mais heterogêneas em sua distribuição do que se supunha anteriormente.

Desvendando a linguagem das histonas

A cromatina é o estado em que o DNA está empacotado dentro da célula, onde o nucleossomo é a unidade fundamental,

sendo composto por um octâmero de quatro histonas centrais (H3, H4, H2A e H2B) envolvidas por aproximadamente 147 pares de bases do DNA (Figura 6). Estruturalmente, as histonas possuem forma globular com caudas nas porções N-terminais. São nessas caudas que residem a chave do entendimento da linguagem das histonas. Existem ao menos oito diferentes modificações na cauda das histonas até então descritas (Tabela 1). A lista de modificações que as histonas podem receber continua em crescimento.

Modificações da cromatina	Resíduos modificados			Função regulatória
Acetilação	K-ac			Transcrição, reparo, replicação, condensação
Metilação (lisinas)	K-me1	K-me2	K-me3	Transcrição, reparo
	R-me1	R-me2	R-me3	
Metilação (argininas)				Transcrição
Fosforilação	S-ph	T-ph		Transcrição, reparo, condensação
Ubiquitilação	K-ub			Transcrição, reparo
Sumoilação	K-su			Transcrição
ADP ribosilação	E-ar			Transcrição
Deiminação	R > Cit			Transcrição
Isomerização de prolina	P-cis	P-trans		Transcrição

As histonas podem ser modificadas em diversos sítios ao longo de suas caudas. Existem mais de 60 diferentes resíduos nas histonas onde ocorrem essas modificações (Figura 6). Entretanto, tudo indica que esses dados estão subestimados. Para termos uma ideia da complexidade dessas modificações, a metilação de lisinas ou argininas pode ser feita de três diferentes formas: (mono-), (di-) ou trimetiladas para lisinas e (mono-) ou (di-) para argininas. Esse arranjo vasto de combinações permite um enorme potencial para respostas funcionais, além disso, cabe lembrar que nem todas essas modificações irão ocorrer na mesma histona ao mesmo

tempo. O momento que essas marcações ocorrem dependerá dos sinais e condições dentro da célula.

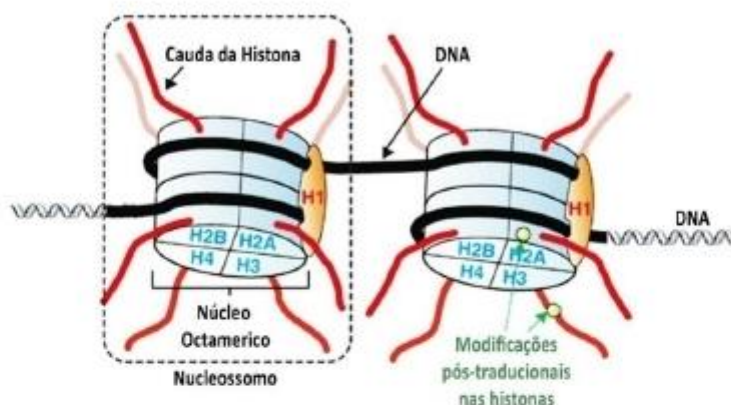


Figura 6. Octâmero de Histonas. Modificado de Carey e Workman, 2007.

As duas modificações mais bem descritas são a acetilação e a metilação de histonas. A acetilação de histonas é catalisada pela enzima histona acetiltransferase (HAT, em inglês) e essa marcação pode ser removida pela histona deacetilases (HDAC). A acetilação de histonas está associada com o aumento da atividade gênica, aumentando a acessibilidade dos fatores transcricionais ao DNA. A acetilação neutraliza a característica básica da lisina, diminuindo a interação entre histona-DNA. Quando da modificação promovida pelas histonas deacetilases, a cromatina adquire uma nova estrutura conformacional que desfavorece a iniciação transcricional. Já a metilação de lisinas é catalisada pela metiltransferase de lisinas de histonas (HKMT) e a remoção dessa marcação é realizada pelas histonas demetilases. A trimetilação da lisina 4 da histona 3 (H3me3K4) está associada com a região promotora de genes ativos e a monometilação com regiões potenciadoras (*enhancers*) dos genes. A metilação de histonas pode estar envolvida com a supressão de sinais de expressão gênica como, por exemplo, H3me2K9 e H3meK27. Nesse contexto, os grupos metila interagem com as histonas formando nucleossomos

que dificultam o acesso dos fatores transcripcionais a determinadas regiões do DNA. A fosforilação é outra modificação que acarreta importantes modificações na compactação da cromatina via modificações de cargas. Entretanto, a função dessa modificação ainda não está bem descrita (Figura 7).

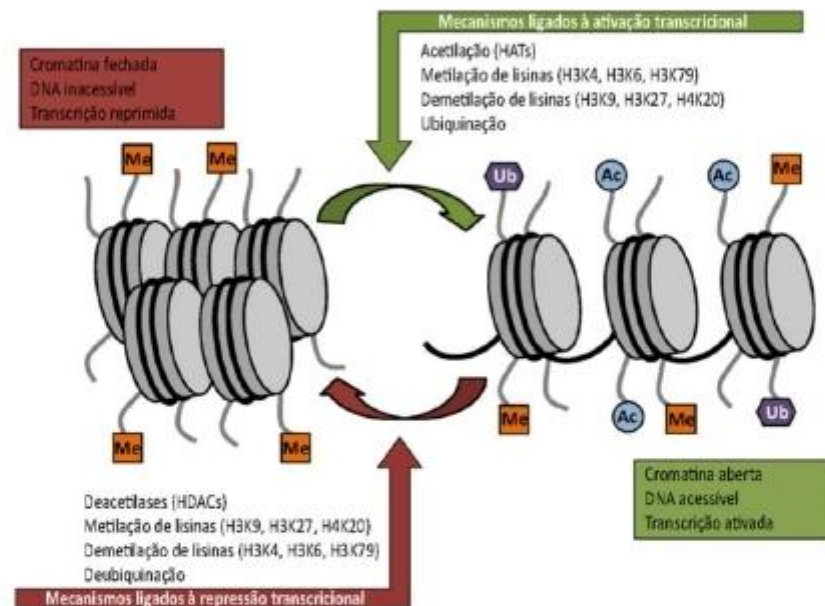


Figura 7. Principais modificações de histonas envolvidos na ativação ou repressão transcricional. Modificado de Carey e Workman, 2007.

As modificações podem afetar a organização global da cromatina, afetando o contato entre diferentes histonas adjacentes no nucleossomo ou a interação das histonas com o DNA. De maneira simples, podemos dizer que as funções das modificações nas histonas podem ser divididas em duas categorias: a) estabelecimento do ambiente global da cromatina; e b) orquestração das demandas biológicas baseadas nas informações contidas no DNA. Para estabelecer o ambiente global da cromatina, as modificações atuam no particionamento do genoma em domínios distintos, como a eucromatina e heterocromatina. Na

euromatina, o DNA se mantém acessível aos fatores de transcrição, enquanto na heterocromatina o DNA se torna inacessível aos mecanismos de transcrição. Para facilitar as funções baseadas nas informações contidas no DNA, as modificações de histonas orquestram a cromatina de tal maneira a ajudar a executar uma determinada função. Pode ser uma função bem localizada, como a transcrição de um gene ou o reparo de DNA, ou algo mais amplo em questão genômica, como a replicação do DNA ou a condensação do cromossomo. Todas essas atividades biológicas requerem recrutamento ordenado do maquinário para deixar o DNA acessível, manipular e então novamente empacotar de maneira correta a cromatina em seu estado anterior.

No que diz respeito ao entendimento de todos esses mecanismos envolvidos nas modificações de histonas e suas funcionalidades, pode se afirmar que ainda há muito que ser desvendado e analisado. Cada uma das etapas das modificações é precedida de uma série de eventos de sinalização. Essa intensa troca de informações dentro e fora do núcleo e da célula tem que ser traduzida de alguma forma para que o maquinário nuclear possibilite a ativação ou desativação de determinados genes de maneira finamente coordenada.

Mecanismos baseados em RNA – Modulação da expressão gênica mediada por RNA não codificante (ncRNA)

Silenciamento gênico pós-transcricional mediado por RNA (ou TGS, em inglês) foi descrito pela primeira vez em 1989, em estudos com a planta do tabaco (Matzke et al., 1989). Posteriormente, estudos demonstraram que pequenas moléculas de RNA não codificante (small nuclear RNA – snRNA) podem ser direcionadas às regiões de sítios promotores de genes-alvos promovendo a modulação do estado de silenciamento gênico e indução de modificação da cromatina (Kim et al., 2006) (Figura 8). Além de estar envolvido no processo de silenciamento gênico,

existem evidências sugerindo o envolvimento das snRNA em aspectos ligados à regulação transcricional. Essas moléculas poderiam, assim, ativar determinados genes quando ligadas aos sítios promotores com baixo conteúdo GC (Janowski et al., 2007).

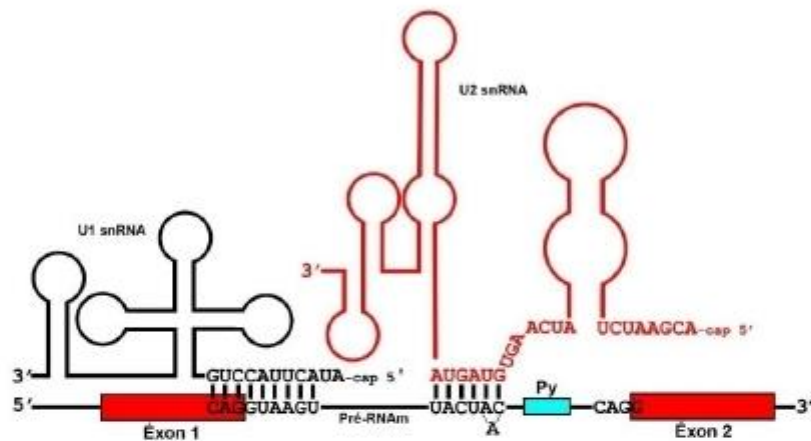


Figura 8. Interações entre pré-RNA mensageiro (pré-RNAm) e os RNA não codificantes U1 e U2 envolvidos no processo de *splicing* do RNAm. De acordo com as interações entre o complexo de *splicing* e o pré-RNAm alvo, poderá haver a geração de transcritos truncados, que não serão traduzidos, promovendo, assim, o controle de expressão gênica pós-transcricional.

Outro tipo de snRNA descrito e que carrega importante função nos mecanismos de controle de expressão gênica pós-transcricional são os microRNAs. Os microRNAs atuam influenciando a estabilidade, compartimentalização e tradução dos RNAs mensageiros (mRNAs). Dessa maneira, a expressão de diversos genes é regulada de acordo com o estabelecimento de diversos processos biológicos-chave, como a proliferação, morfogênese, apoptose e diferenciação (Carthew et al., 2006).

Os RNA transportadores (tRNAs) são melhor conhecidos por sua função de adaptadores durante o processo de tradução do código genético, porém em procariotos e eucariotos também estão envolvidos na regulação de expressão gênica (Figura 9). Essas moléculas de tRNA também possuem implicação como substrato

na formação de ligação peptídica não ribossomal, marcação pós-traducional de proteínas, modificação de fosfolipídios em membranas celulares e na síntese de antibióticos. Mais recentemente, os fragmentos de tRNA, ou tRFs, também foram reconhecidos por possuírem funções regulatórias. Inicialmente, esses produtos de clivagem foram considerados como resultado de degradação aleatória, no entanto, vários estudos revelaram a sua produção como resultado de clivagens específicas e possivelmente regulada em resposta a eventos estressores, como fome ou invasão viral (Raina & Ibba, 2014).

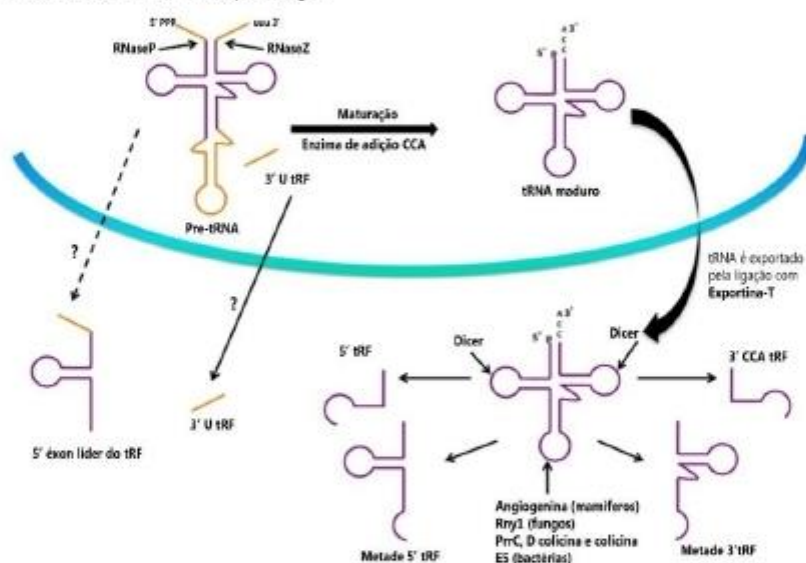


Figura 9. Processos envolvidos na formação dos fragmentos de tRNA (tRFs). Os precursores são processados pelas enzimas RNaseZ e RNaseP, por uma endonuclease de *splicing* e pela adição da terminação CCA, formando, assim, o tRNA maduro no núcleo. Tanto o processamento do pre-tRNA como o tRNA maduro podem dar origem aos pequenos fragmentos de RNA. As prováveis rotas são apresentadas na figura. As linhas pontilhadas e os pontos de interrogação indicam que os mecanismos de formação ou transporte desses tRFs são desconhecidos ainda. Modificado de Raina & Ibba, 2014.

Em estudos recentes, foi demonstrado que os fragmentos de tRNA sensibilizam as células para a ativação induzida por estresse oxidativo da via supressora de tumor p53 (Hanada et al.,

2013). Isso sugere que a clivagem de tRNAs ativa a apoptose via ativação do gene p53 e, portanto, atua como mecanismo de proteção contra câncer, enquanto a não clivagem do tRNA o induz a se ligar ao citocromo-c, evitando a apoptose e auxiliando, assim, o desenvolvimento do câncer. Claramente, ainda há muito a ser descoberto sobre os vários papéis reguladores do tRNA (Figura 10).

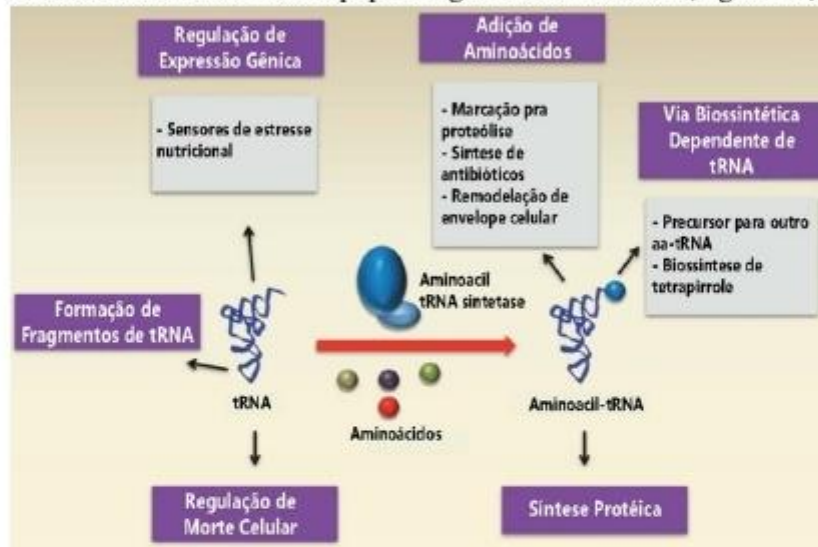


Figura 10. As várias funções desempenhadas pelos tRNAs (carregados ou não com aminoácidos). Modificado de Raina & Ibba, 2014.

O pecado de nossos pais: a herança transgeracional

Quando Lamarck, sendo o bom cientista que era, postulou sua hipótese sobre a transformação das espécies, ele tomou o cuidado de levantar o máximo possível de dados que dessem suporte à sua hipótese. Um dos exemplos mais famosos foi o recordado por Lamarck a respeito dos filhos de ferreiros (um labor que exigia demais do físico) que tendiam a ter braços mais musculosos do que os filhos de tecelões (um trabalho que exigia bem menos do físico). Lamarck interpretou essa observação como se os filhos de ferreiros herdassem as características adquiridas pelos pais – no caso, o de ter braços musculosos.

É claro que nossa interpretação contemporânea a respeito desses fatos é diferente. Reconhecemos que um homem que possui genes que lhe confirmam a habilidade de desenvolver mais massa muscular poderia ter algum tipo de vantagem em uma profissão como a de ferreiro. Assim, essa profissão atrairia e destacaria aqueles que estivessem geneticamente melhor amparados para o exercício do labor. Dessa maneira, nossa interpretação assumiria que os filhos desses ferreiros herdariam a sua tendência genética a ter uma maior massa muscular.

Seria um erro olharmos para trás e pintarmos Lamarck como um bobo, ou alguém digno de chacotas. Infelizmente, o que observamos na maioria das vezes em sala de aula é a caracterização do trabalho de Lamarck como digno de piadas frente ao trabalho desenvolvido por Darwin. É claro, também, que as hipóteses lamarckianas não são mais aceitas cientificamente, mas há de se ressaltar que Lamarck foi o primeiro a querer responder de maneira genuína os mecanismos envolvidos na transformação das espécies. O modelo explicativo para a evolução das espécies proposto por Darwin, via seleção natural, é um dos conceitos mais poderosos já descritos no campo das ciências biológicas. O poder de seu modelo teórico se torna maior quando, em torno de 1930, ocorre o casamento entre a sua teoria de evolução por seleção natural e os trabalhos de Mendel (redescobertos em 1900), conferindo um poder explicativo ímpar sobre hereditariedade e o nosso entendimento molecular sobre o DNA como matéria-prima da hereditariedade.

Se pudéssemos resumir o último século e meio a respeito da teoria evolutiva em um parágrafo, diríamos que: A variação ao acaso em genes cria a variação fenotípica nos indivíduos. Alguns indivíduos sobrevivem diferencialmente em relação aos demais, em um dado ambiente, e esses indivíduos deixarão para seus herdeiros suas aptidões genéticas. Essa linhagem, assim, terá um maior sucesso reprodutivo. Eventualmente, com o passar de diversas gerações, espécies distintas irão evoluir.

A matéria-prima para a variação ao acaso é a mutação nas sequências do DNA dos indivíduos de uma população. É nesse ponto que o modelo lamarckiano de características adquiridas realmente sucumbe frente ao modelo darwiniano de evolução. Uma mudança adquirida no fenótipo deveria de alguma forma ser refletida sobre as sequências do DNA, mudando-o dramaticamente. Dessa maneira, a característica adquirida poderia ser transmitida para a próxima geração, dos pais para os filhos. Entretanto, há pouquíssimas evidências de que esse fenômeno possa realmente vir a ocorrer, exceto em casos de respostas a elementos químicos ou irradiação que possam danificar o DNA (mutagênicos), causando uma mudança na sequência original dos nucleotídeos.

August Weismann, em 1893, formulou sua hipótese sobre hereditariedade, denominada “germoplasma”. A partir dessa hipótese, o autor ficou conhecido como responsável por derrubar de vez o modelo lamarckiano, afirmando que somente os gametas são as células responsáveis pela transmissão de informações genéticas para a geração seguinte, em que as células somáticas não assumem qualquer função hereditária. Assim, a **Barreira de Weismann** estabelece a premissa de que informações hereditárias não podem passar do soma para a linhagem germinativa. O que temos é um amplo corpo de dados e evidências que se colocam contra a herança lamarckiana e, dessa forma, há poucas razões para que os cientistas trabalhem dentro dessa linha de pensamento.

Realmente é difícil imaginar como diferentes fatores ambientais seriam capazes de influenciar as células de um indivíduo, podendo agir em um gene específico, fazendo com que este mude a sua sequência de pares de base. Porém, é extremamente óbvio que modificações epigenéticas – metilação de DNA, modificações de histonas, RNAs não codificantes – possam, de fato, ocorrer em genes específicos como resposta a influências ambientais em uma célula, ou conjunto de células. Mas aqui

novamente encontramos uma dificuldade de estabelecer uma ligação de modificações entre gerações.

É sabido que as marcações epigenéticas adquiridas no desenvolvimento e também ao longo da vida são “resetadas” em dois momentos – algo análogo ao ato de formatar um disco-rígido de computador, preparando para receber um novo sistema operacional –, o que seria o programa de desenvolvimento do novo organismo. O primeiro momento de “formatação” ou reprogramação ocorre durante a fase de proliferação e migração da linhagem de células germinativas nos órgãos reprodutivos dos progenitores. Em seguida, ao longo da espermatogênese e da oogênese o genoma recebe marcações epigenéticas específicas (espermatozoides recebem uma assinatura epigenética diferente dos oócitos). O segundo momento de reprogramação ocorrerá após a fecundação. Assim que ocorre a cariogamia (fusão dos pró-núcleos espermáticos e do oócito), as assinaturas epigenéticas depositadas no genoma do espermatozoide e do oócito são novamente apagadas para que sejam estabelecidas novas marcações pertencentes ao programa de desenvolvimento do zigoto (Figura 11).

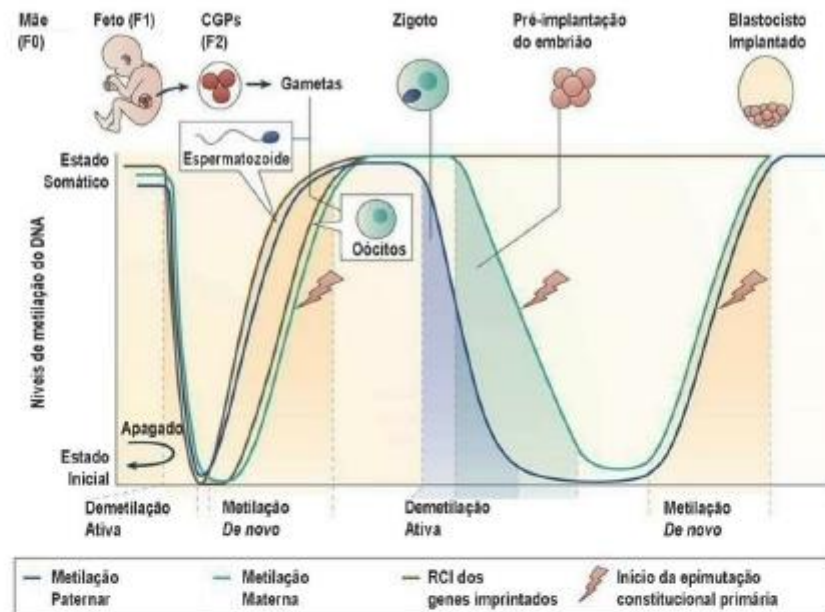


Figura 11. Reprogramação epigenética. Os estádios mais vulneráveis em que é mais provável que uma epimutação constitucional primária seja estabelecida, devido às perturbações ambientais ou celulares, são mostradas pelas áreas marcadas em laranja. O estágio em que é maior a chance de ocorrer o apagamento de uma epimutação parental existente, reduzindo a probabilidade de herança epigenética transgeracional, está representado pela seta preta curva. Modificado de Hitchins, 2015.

Entretanto, o que vem se observando é que esse mecanismo de reformatação das marcações epigenéticas não são a prova de erros. Simplificadamente, o que se observa é que há três momentos-chave onde modificações nos padrões de metilação podem ocorrer: a) durante a gametogênese, onde fatores ambientais podem alterar os padrões de metilação dos gametas (epimutação); b) na fase de proliferação celular após a fecundação, quando há a demetilação, ou seja, retirada das marcas epigenéticas oriundas dos pró-núcleos maternos e paternos, nesse caso, as marcações epigenéticas podem persistir e, assim, modificar os padrões posteriores de desenvolvimento; e c) no momento de implantação do embrião, que é quando ocorre o processo de metilação *de novo*, necessário para o desenvolvimento

embrionário, dando identidade às células, promovendo, dessa forma, a citodiferenciação e a morfogênese (Figura 11).

Mas de que maneira fatores ambientais poderiam modificar as marcações epigenéticas? Podemos usar como exemplo a resposta a uma determinada rota de sinalização hormonal. Tipicamente, um hormônio como o estrogênio irá se ligar a um receptor na membrana plasmática da célula, como da glândula mamária. O estrogênio e o seu receptor são direcionados ao núcleo da célula, onde irão se ligar a regiões específicas no DNA, o sítio promotor de determinados genes. Esse evento ajuda o processo de ativação dos genes. Quando o estrogênio se liga ao sítio promotor, ele atrai várias enzimas do maquinário epigenético. Essas enzimas, por sua vez, irão implementar modificações nas histonas (como vimos anteriormente), removendo marcações que promoviam a repressão da expressão gênica e colocando marcas que irão manter a expressão ativa do gene. Assim, o ambiente agindo via hormonal poderá induzir marcações epigenéticas diferenciais em genes específicos.

Nota-se que as modificações epigenéticas não alteram a sequência do DNA, mas alteram como os genes são expressos. Esse processo, de modo geral, é a base da programação genética do desenvolvimento de um organismo, podendo desencadear uma série de doenças quando desregulado. Sabe-se, ainda, que as modificações epigenéticas são transmitidas de células-mãe para células-filhas. É por isso que não nascem dentes no globo ocular, por exemplo, pois os padrões epigenéticos são mantidos, conferindo identidade celular e, por conseguinte, tecidual a esses órgãos. Se eventos estressores modificarem os padrões de assinaturas epigenéticas em células gaméticas, teremos um mecanismo no qual modificações epigenéticas induzidas pelo ambiente poderão ser transmitidas para as próximas gerações. Teremos assim uma espécie de mecanismo similar à herança lamarckiana. Ou seja, uma modificação epigenética (ao contrário de genética) poderia ser passada dos pais para os filhos.

Antes de continuarmos, é interessante deixarmos claro algumas terminologias, visto que na literatura científica podemos

encontrar algumas confusões referentes aos termos empregados. Artigos científicos mais antigos referiam-se à transmissão epigenética de algum traço adquirido, mas sem descrever qualquer evidência de mudanças na metilação de DNA genômico ou alterações nas histonas. Entretanto, isso não demonstrava negligência por parte dos autores, mas era reflexo das diferentes formas com que a palavra “epigenética” era empregada. Por muito tempo, o termo “transmissão epigenética” referia-se à herança de alguma característica que não pudesse ser explicada pela genética. Nesse caso, a palavra epigenética foi utilizada para descrever um fenômeno, e não um mecanismo molecular. Na tentativa de manter a terminologia mais clara possível, utilizaremos o termo “herança transgeracional” para descrever o fenômeno de transmissão de uma característica adquirida e somente utilizaremos “epigenética” para descrever os eventos moleculares envolvidos.

Terminologia esclarecida, passaremos a analisar se existe alguma evidência de herança transgeracional. Sim, há fortes evidências que ocorre herança transgeracional em humanos. Uma das mais fortes evidências vem dos sobreviventes do período conhecido como “Fome do Inverno Holandês”. Graças à excelente infraestrutura médica holandesa, os altos padrões de dados coletados e preservados, foi possível aos epidemiologistas seguirem e traçarem os sobreviventes de um período de fome que perdurou por muitos anos na Holanda no período da II Grande Guerra. Resumidamente, os epidemiologistas puderam monitorar não somente aquela geração que viveu nesse período, mas também os filhos e netos.

Esse monitoramento trouxe à tona um extraordinário efeito. Quando uma mulher grávida passa por má-nutrição durante os três primeiros meses de gestação, seus filhos irão nascer com peso dentro da média (peso normal), mas quando adultos terão maiores chances de desenvolverem obesidade e outras desordens fisiológicas, como diabetes e problemas cardíacos. Estranhamente, quando uma mulher dessa linhagem de crianças se tornam mães, seus primogênitos tendem a ser mais pesados do que aqueles de grupos controles (mais

pesados e maiores que a média) (Lumey et al., 2007; Lumey et al., 2008) (Figura 12).

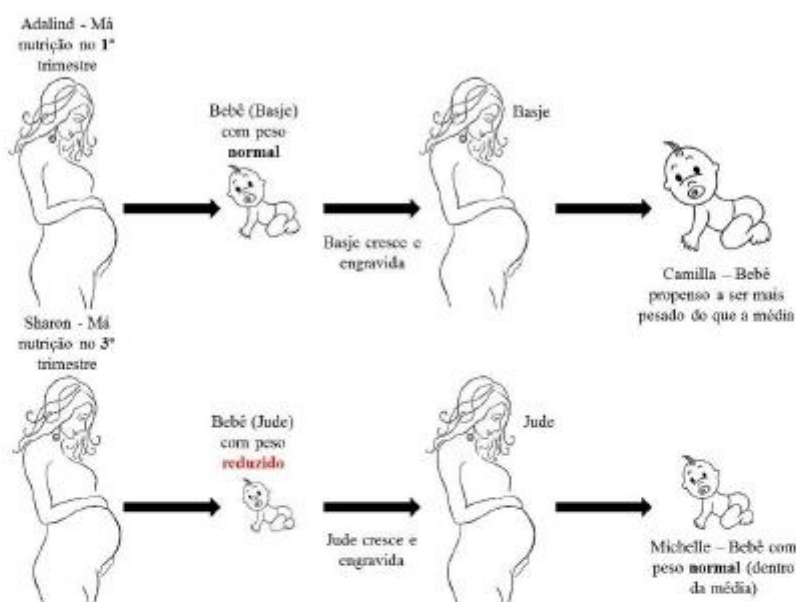


Figura 12. Efeitos da má nutrição através de duas gerações de filhas e netas de uma mulher que engravidou durante a “Fome do Inverno Holandês”. O período em que ocorre a má nutrição é crítico para os efeitos subsequentes no peso corporal. Os nomes são meramente ilustrativos. Modificado de Carey, 2012

Os efeitos no peso no nascimento do bebê Camilla são realmente surpreendentes. Quando Camilla estava se desenvolvendo, sua mãe Basje presumivelmente era saudável. O único período pela qual Basje passou por má nutrição foi quando ela estava em seus primeiros estágios de desenvolvimento no útero materno (Adalind). O que parece é que houve algum efeito sobre sua filha (Camilla), mesmo que Camilla nunca tenha sido exposta a algum período de má nutrição durante os estágios iniciais de desenvolvimento.

Estariamos observando alguma forma de herança transgeracional (lamarckiana), mas sendo causada por mecanismos epigenéticos? Será que as mudanças nos padrões de metilação, ou modificações nas histonas, que ocorreram em Basje como resultado

da má nutrição no período inicial de suas 12 primeiras semanas de desenvolvimento no útero de sua mãe (Adalind) passaram via núcleo de seus oócitos para sua filha (Camilla)? Talvez, mas também não podemos excluir outras explicações potenciais para essas observações.

Poderia, por exemplo, algum efeito não identificado nos estágios iniciais de má nutrição de quando Adalind estava grávida de Basje ter feito com que Basje passasse mais nutrientes do que o normal através da placenta para sua filha em desenvolvimento, Camilla. Isso poderia ter criado um efeito transgeracional, mas não seria um efeito causado pela Basje passando alguma modificação epigenética para Camilla. Ou seja, seria uma mudança causada pelas condições da vida intrauterina quando Camilla estava em desenvolvimento e crescendo. Também devemos lembrar que o oócito é muito grande, carregando em seu citoplasma uma série de funções para quando o oócito é fertilizado. Talvez algo tenha sido modificado durante os estágios iniciais de desenvolvimento em Basje, que tenha resultado em conteúdo citoplasmático não usual em seus oócitos, fazendo com que Camilla tivesse um padrão de crescimento não usual. Cabe ressaltar que a produção dos gametas femininos se inicia logo cedo no desenvolvimento embrionário.

Assim, percebemos que existem vários mecanismos que poderiam ser utilizados para explicar os padrões de herança observados na linha materna nos sobreviventes da “Fome do Inverno Holandês”. Facilitaria o entendimento de como a epigenética atuaria na herança adquirida se pudéssemos estudar uma situação humana menos complicada. O ideal seria a observação em um cenário onde não tivéssemos a preocupação sobre os efeitos do ambiente intrauterino ou influência do conteúdo citoplasmático do oócito. Ou seja, seria interessante ouvirmos os pais, pois homens não engravidam e os espermatozoides colaboram quase nada em questão de conteúdo citoplasmático.

Outro estudo histórico lança luz no caso de herança transgeracional por via paterna (Kaati et al., 2002). Em uma região isolada do norte da Suécia, chamada Överkalix, entre o final do século

XIX e início do século XX, houve um período de forte estiagem causada por diversos fatores (problemas na colheita, ações militares, problemas com o armazenamento adequado dos alimentos), entremeada com um período de grande fartura. Os pesquisadores estudaram os padrões de mortalidade para os descendentes de pessoas que viverem durante esses períodos, em particular, analisaram a quantidade de calorias ingeridas durante o estágio da infância conhecido como **Período de Crescimento Lento** (PCL). Em um cenário onde todos os fatores se mantenham iguais e estáveis, o crescimento infantil é mais devagar nesse período da vida, que antecede a puberdade. É algo completamente normal, sendo registrado na maioria da população humana.

Ao analisarem os dados, os pesquisadores verificaram que se a demanda de comida era escassa durante a fase em que os pais estavam na PCL, seus filhos tinham menos risco de morrerem por causa de doenças cardiovasculares (derrame, pressão alta, doenças coronarianas). Se, por outro lado, um homem tivesse acesso à comida em excesso durante a PCL, seus netos tiveram aumentado os riscos de morte em decorrência de problemas ocasionados pela diabetes. Assim como o caso da Camilla na “Fome do Inverno Holandês”, os filhos e netos tiveram seus fenótipos alterados em resposta às mudanças ambientais, as quais eles não vivenciaram.

Esses dados, claramente, não podem ser em decorrência da vida intrauterina, nem efeitos citoplasmáticos, pelas razões anteriormente e obviamente descritas. Dessa maneira, parece que a hipótese de herança transgeracional seja mais plausível, em decorrência da disponibilidade de comida na geração dos avôs, sendo mediado via epigenética. Esses dados são mais impressionantes quando consideramos o fato de que o efeito nutricional original ocorreu quando os garotos estavam na pré-puberdade e, assim, não começaram a produção de gametas. Mesmo assim, esse fato foi capaz de ser passado para os seus filhos e netos.

Entretanto, existem diversas ressalvas que devem ser levadas em consideração nesse trabalho sobre herança transgeracional

através da linhagem parental. Em particular, existem riscos envolvendo os registros de óbitos antigos, extrapolando os dados históricos. Além disso, alguns efeitos que foram observados não eram demasiadamente amplos, em questão de frequência. Eis um problema de estudos envolvendo humanos, pois frequentemente lidamos com problemas de tamanho amostral, variação genética e impossibilidade de controlar os fatores ambientais. Assim, sempre haverá o risco de se estabelecer conclusões equivocadas dos dados obtidos, da mesma forma como acreditamos que Lamarck fez com seus estudos envolvendo as famílias de ferreiros.

Diante dessa problemática e, principalmente, com o intuito de averiguar se a herança transgeracional é realmente um processo existente, os cientistas voltaram seus olhos para organismos modelos, como camundongos e até mesmo drosófilas (moscas-da-fruta). Um dos estudos clássicos que teve como objetivo compreender os mecanismos epigenéticos envolvidos na herança transgeracional foi o trabalho elaborado por Dolinoy e colaboradores em 2006. Nesse estudo, Dolinoy e colaboradores avaliaram a ação de genisteína em camundongos com pelagem agouti amarela. A genisteína é um fitoestrógeno presente na soja, que diminui a fertilidade feminina, atua na quimioprevenção de câncer e na diminuição da deposição adiposa. O que foi observado é que a suplementação materna de genisteína durante a gestação, em níveis comparáveis aos seres humanos, que possuem dietas com consumo elevado de soja, alterou a cor da pelagem de animais agouti heterozigotos (A^{vy} / a) para pseudoagouti. Outros dois estudos também foram realizados com essa mesma linhagem de camundongos: Morgan (1999) e Blewitt (2006) com Bisfenol-A (BPA) e ácido fólico.

O BPA é utilizado na produção do policarbonato de bisfenol A, policarbonato mais comum presente em materiais plásticos e reconhecidamente um disruptor estrogênico. O ácido fólico, ou folato, é fonte de grupamento metil, matéria-prima no processo de produção de S-Adenosilmetionina (SAM), cofator presente nas enzimas DNA

metiltransferases, que irá atuar como doador do grupamento metil à citosina alvo (Figura 13).

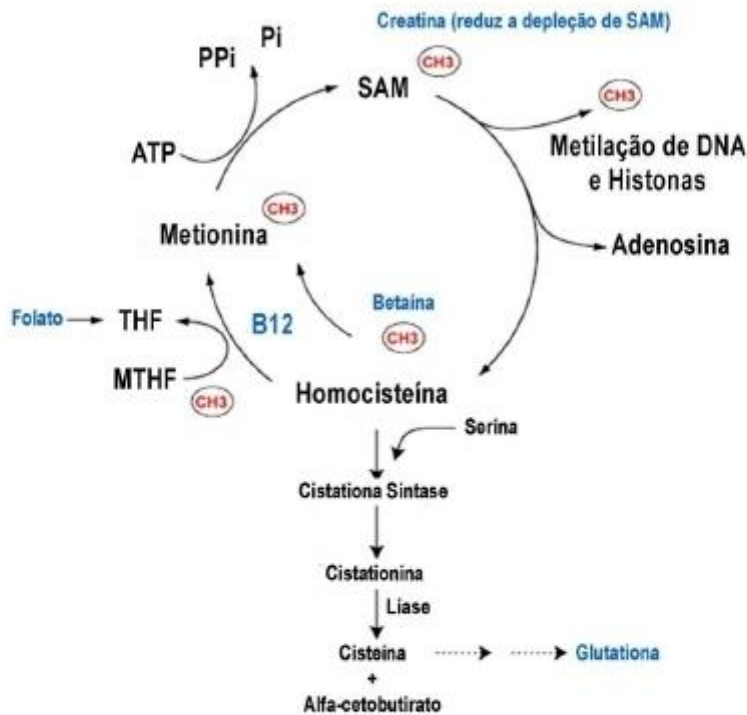


Figura 13. Diagrama de várias rotas que teoricamente aumentam a disponibilidade de grupos metil em sistemas biológicos. Alguns destes são usados como suplementos alimentares. Por exemplo, betaína é uma fonte de grupos metil e, teoricamente, está disponível para metionina. A creatina pode ajudar a reduzir a depleção de SAM, uma vez que a síntese *de novo* de creatina consome SAM. O ácido fólico pode ser tomado por via oral e, claro, a vitamina B12 pode ser tomada por via oral ou por injeção.

O que Morgan e Blewitt observaram foram mudanças semelhantes às observadas por Dolinoy na pelagem de camundongos quando esses eram expostos ao BPA, ou suplementação de folato em sua alimentação durante o período de gestação, sendo que essas mudanças persistiram ao longo de diversas gerações, mesmo quando cessado o contato com o BPA ou ácido fólico (Figura 14).

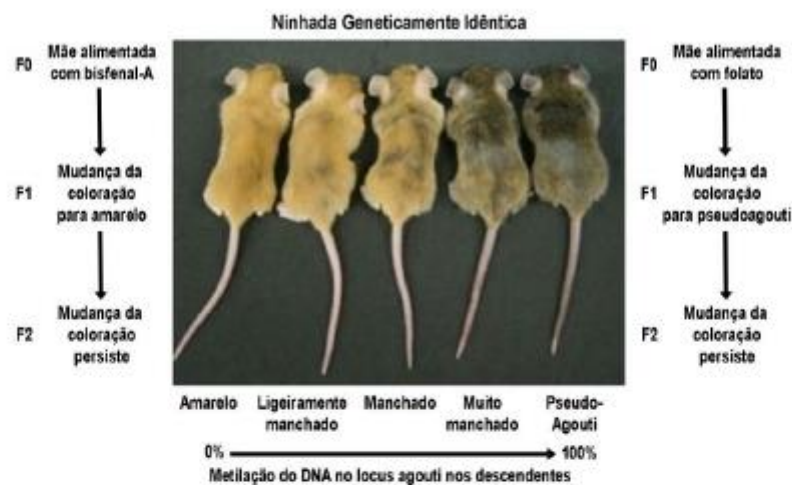


Figura 14. As alterações de metilação do DNA no *locus* Agouti amarelo (A^Y) são fenotipicamente detectáveis e influenciadas pelo meio ambiente. A cor da pelagem varia nesses camundongos em um espectro de amarelo para agouti. As diferenças na cor da pelagem se devem a uma alteração na metilação do DNA no *locus* de Agouti. Os camundongos que possuem *loci* Agouti com baixo nível de metilação (por exemplo, 0%) são amarelos, enquanto que aqueles que possuem alto nível de metilação do DNA (por exemplo, 100%) aparecem com pelagem agouti (marrom). O tratamento com BPA ou ácido fólico pode mudar a distribuição da cor da pelagem para amarelo ou pseudoagouti, respectivamente. Essa mudança de cor da pelagem persiste nas gerações subsequentes indicando que a mudança epigenética que ocorre no *locus* agouti é hereditária.

Molecularmente falando, o que ocorre é que no *locus* Agouti está presente um retrotransposon próximo ao sítio promotor do gene responsável pela coloração agouti amarelo (retrotransposon é um tipo de elemento móvel autônomo de DNA, capaz de se mobilizar e se inserir em novos locais do genoma do hospedeiro). Nos camundongos que possuem baixos níveis de metilação do retrotransposon, o gene é continuamente expresso, fazendo com que tenham pelagem mais clara. Entretanto, quando há alterações no metabolismo envolvido com os mecanismos epigenéticos, temos maiores níveis de metilação no retrotransposon, que conseqüentemente diminui os níveis de expressão do gene agouti, fazendo com que a pelagem fique escura, no caso da suplementação com folato (quando normalmente seria

amarelo claro). Já quando temos camundongos com pelagem escura, e estes são submetidos a uma dieta com a presença de BPA, a coloração se torna amarela, quando deveria ser mais escura, pois o BPA desregula o metabolismo de SAM, diminuindo a disponibilidade de doadores de grupamento metil e, conseqüentemente, diminuindo os níveis de metilação na região onde está presente o retrotransposons (Figura 15).

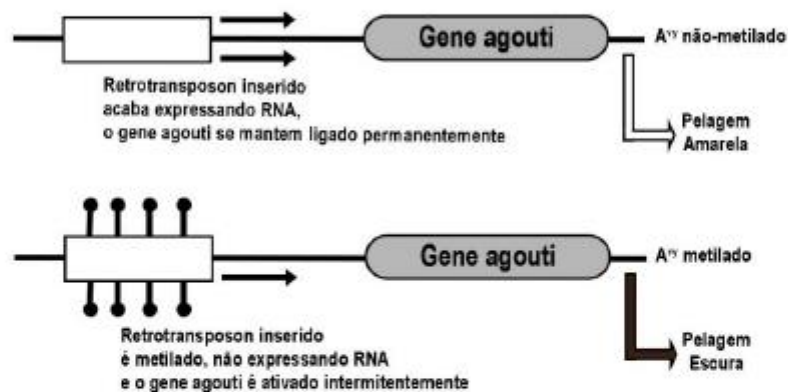


Figura 15. Variações nos padrões de metilação de DNA (representado pelos círculos pretos) influenciam a expressão de um retrotransposons inserido junto ao sítio promotor do gene *agouti*. A variação na expressão do retrotransposon afeta a expressão do gene *agouti*, em contrapartida, fazendo com que a coloração da pelagem varie entre animais geneticamente idênticos.

Atualmente, diversos estudos apresentam de forma robusta os mecanismos envolvidos nos fenômenos de herança transgeracional. Além das modificações dos padrões de metilação de DNA ou modificações de histonas, um novo candidato a intermediar essas relações inter e transgeracionais são os RNA não codificantes (apresentados anteriormente). A herança transgeracional é mais frequentemente transmitida através da linhagem germinal em consequência de eventos primários que ocorrem em células somáticas, o que implica a transmissão de informações do soma para linhagem germinal. O interessante é que até pouco tempo nenhum exemplo de transmissão de

informações mediada por DNA ou RNA da linhagem celular somática para as células germinativas foi relatado. Porém, esse panorama passou a mudar recentemente. Em trabalho realizado por Cossetti e colaboradores (2014), a equipe gerou um camundongo modelo xenoenxertado com células de melanoma humano, que expressavam de forma estável o gene EGFP (uma proteína fluorescente). O que eles encontraram foi incrível. Cossetti encontrou RNA do gene EGFP liberado das células humanas xenoenxertadas na corrente sanguínea e nos espermatozoides dos camundongos. A forma como esse RNA chega às células germinativas é através de exossomas.

De fato, os snRNA apresentam-se como vetores potenciais na interface entre genes e meio ambiente. Um exemplo notável emerge da recente descoberta de que o estresse traumático no início da vida é capaz de alterar a expressão do microRNAs (miRNAs) em camundongos, e as respostas comportamentais e metabólicas na progênie. Se injetarmos esses RNAs de espermatozoides de machos que sofreram algum tipo de evento traumático em oócitos de fêmeas que não passaram por esse evento estressante, os descendentes acabam reproduzindo as alterações comportamentais e metabólicas, tal qual a geração paterna. Assim como eventos ambientais estressores podem modificar os padrões da programação epigenética no ambiente intrauterino, o mesmo pode ser observado na linhagem paterna. De modo geral, a janela para que esses exossomos contendo snRNAs possam ser passados da linhagem somática para a germinativa ocorre um pouco antes da primeira divisão meiótica do espermatócito, quando é estabelecido o estoque de RNAs, proteínas de ligação ao RNA e sistemas de repressão traducional (Figura 16). Uma vez que ocorreu a compactação de DNA paterno, que envolve a substituição da maioria das histonas por protaminas, não existe um mecanismo claro pelo qual as marcações epigenéticas possam ser alteradas. O tempo entre a compactação da cromatina e o trânsito do esperma maduro para o epidídimo é de pelo menos 10 dias, período no qual

a maquinaria epigenética e transcricional não está mais ativa. No entanto, a secreção de exossomos e outros fatores que interagem com os espermatozoides em amadurecimento no epidídimo podem conferir marcas epigenéticas duradouras, como os microRNAs.

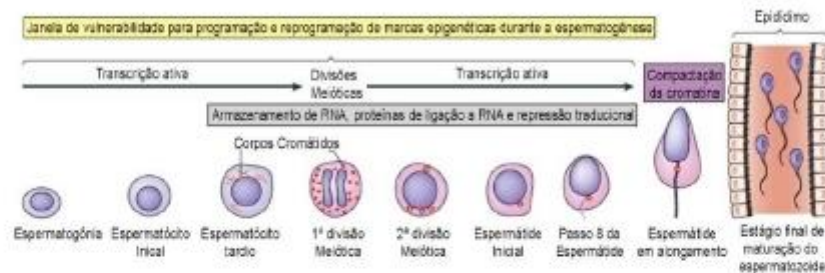


Figura 16. Janelas de vulnerabilidade à reprogramação ambiental na espermatogênese. A figura mostra os estágios da espermatogênese e maturação do espermatozoide, destacando os períodos em que exposições ambientais podem reprogramar as marcações epigenéticas. Modificado de Bale, 2014.

Os dados acumulados ao longo das décadas de estudos epigenéticos acabaram propiciando um campo fértil de teste de hipóteses e explicações a respeito de fenômenos transgeracionais que possuem padrões de herança não mendeliana. Diversos estudos começaram a se preocupar em questionar os efeitos de disruptores hormonais, efeitos da alimentação, hormônios artificiais que são lançados no ambiente (desde o efeito de pílulas contraceptivas até a ação de pesticidas) na tarefa de avaliar potenciais causas sutis, mas com potenciais efeitos transgeracionais na população humana também na vida dos demais organismos na natureza. Por exemplo, temos pesquisas que demonstram claramente os efeitos danosos do BPA nas populações de peixes, diminuindo a fertilidade desses animais – o mesmo já se observa em populações humanas. Como vimos anteriormente, questões nutricionais também podem afetar os padrões de programação epigenética, levando as gerações futuras a desenvolverem uma série de problemas cardiovasculares, de diabetes, de câncer, problemas renais, entre outros. É provável que

uma série de doenças que se alastram dentro de nossa população seja em decorrência do pecado de nossos avós e pais, mesmo que não tenhamos sido expostos a esses fatores geradores. Da mesma forma, podemos nos questionar sobre o legado que deixaremos para nossos filhos, netos e bisnetos.

Assim, percebemos que as informações contidas em nosso genoma não estão isoladas do ambiente que nos cerca. Há uma intensa troca de informações, numa via de mão dupla. A epigenética justamente nos traz as ferramentas moleculares que fazem o papel de mediador entre o ambiente e a genótipo. Modificações nos padrões epigenéticos – as epimutações – são capazes de modificar os padrões de expressão dos genes, alterando o transcriptoma ao longo do desenvolvimento do organismo. Essas alterações, por sua vez, dão à luz a fenótipos não esperados, com respostas fisiológicas diferentes, o que, por conseguinte, alimentará o *pool* fenotípico de uma população com novas variantes fenotípicas, aumentando ou diminuindo a aptidão dos indivíduos. Essas variantes poderão ser alvo dos mecanismos evolutivos, como a seleção natural, modificando o rumo da evolução da espécie. Um admirável mundo novo de possibilidades se abriu frente aos nossos olhos entre o final do século XX e esse início de século XXI. Ainda temos muito a compreender. A cada dia novas peças nesse quebra-cabeça são acrescentadas, e isso torna a epigenética um campo por demais fascinante.

Referências

- BIRD, A. DNA methylation patterns and epigenetic memory. *Genes & development*, 1 jan. 2002. v. 16, n. 1, p. 6-21. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11782440>>. Acesso em: 22 maio 2013.
- BLEWITT, M. et al. The Use of Mouse Models to Study Epigenetics The Use of Mouse Models to Study Epigenetics. *Cold Spring Harb Perspect Biol*, 2014. p. 1-28.

- CARTHEW, R. W. Gene regulation by microRNAs. *Current Opinion in Genetics and Development*, 2006. v. 16, n. 2, p. 203-208.
- COSSETTI, C. et al. Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: Possible transport by exosomes. *PLoS ONE*, 2014. v. 9, n. 7.
- DOLINOY, D. C. et al. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environmental Health Perspectives*, 2006. v. 114, n. 4, p. 567-572.
- DOZMOROV, M. G. et al. Systematic classification of non-coding RNAs by epigenomic similarity. *BMC bioinformatics*, 2013. v. 14 Suppl 1, n. Suppl 14, p. S2. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3851203&tool=pmcentrez&rendertype=abstract>>.
- FARTHING, C. R. et al. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genetics*, 2008. v. 4, n. 6.
- FEINBERG, A. P.; VOGELSTEIN, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 1983. v. 301, n. 5895, p. 89-92. Disponível em: <<http://www.nature.com/doi/10.1038/301089a0>>.
- FRANÇOIS FUKS, WENDY A. BURGERS, ALEXANDER BREHM, L. H.-D. & T. K. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. 2000. v. 24, n. *Nature*, p. 1-4.
- HAN, H. et al. DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Human Molecular Genetics*, 2011. v. 20, n. 22, p. 4299-4310.
- HANADA, T. et al. CLP1 links tRNA metabolism to progressive motor-neuron loss. *Nature*, 2013. v. 495, n. 7442, p. 474-480.
- HARDY, T. M.; TOLLEFSBOL, T. O. Epigenetic diet: impact on the epigenome and cancer. *Epigenomics*, 2011. v. 3, n. 4, p. 503-518.
- HITCHINS, M. P. Constitutional epimutation as a mechanism for cancer causality and heritability? *Nature Reviews Cancer*, 2015. v. 15, n. 10, p. 625-634. Disponível em: <<http://www.nature.com/doi/10.1038/nrc4001>>.

- JANOWSKI, B. A. et al. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nature Chemical Biology*, 2007. v. 3, n. 3, p. 166-173. Disponível em: <<http://www.nature.com/doi/10.1038/nchembio860>>.
- KAATI, G.; BYGREN, L. O.; EDVINSSON, S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *European Journal of Human Genetics*, 2002. v. 10, n. 11, p. 682-688. Disponível em: <<http://www.nature.com/doi/10.1038/sj.ejhg.5200859>>.
- KIM, D. H. et al. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature Structural & Molecular Biology*, 2006. v. 13, n. 9, p. 793-797. Disponível em: <<http://www.nature.com/doi/10.1038/nsmb1142>>.
- LAURENT, L. et al. Dynamic changes in the human methylome during differentiation. 2010. n. *Genome Res*, p. 320-331.
- LI, B.; CAREY, M.; WORKMAN, J. L. The Role of Chromatin during Transcription. *Cell*, 2007. v. 128, n. 4, p. 707-719.
- LO, P.-K.; SUKUMAR, S. Epigenomics and breast cancer. *Pharmacogenomics*, 2009. v. 9, n. 12, p. 1879-1902.
- LUMEY, L. H. et al. Cohort profile: The Dutch Hunger Winter families study. *International Journal of Epidemiology*, 2007. v. 36, n. 6, p. 1196-1204.
- LUMEY, L. H. et al. Lipid profiles in middle-aged men and women after famine exposure during gestation: The Dutch Hunger Winter Families Study. *American Journal of Clinical Nutrition*, 2009. v. 89, n. 6, p. 1737-1743.
- MATZKE, M. A. et al. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *The EMBO journal*, 1989. v. 8, n. 3, p. 643-649. Disponível em: <[/pmc/articles/PMC400855/?report=abstract](http://pmc/articles/PMC400855/?report=abstract)>.
- MORGAN, H. D. et al. Epigenetic inheritance at the agouti locus in the mouse. *Nature genetics*, 1999. v. 23, n. 3, p. 314-318.
- MULLER, H. J. The Production of Mutations by X-Rays. *Proc Natl Acad Sci U S A*, 1928. v. 14, n. 9, p. 714-726.

- NAN, X. et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 28 maio. 1998. v. 393, n. 6683, p. 386-389. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9620804>>.
- OKANO, M. et al. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 29 out. 1999. v. 99, n. 3, p. 247-257. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/10555141>>.
- ØROM, U. A.; SHIEKHATTAR, R. Long non-coding RNAs and enhancers. *Current opinion in genetics & development*, 2011. v. 21, p. 194-198.
- PARK, Y. J. et al. Genome-wide epigenetic modifications in cancer. *Progress in Drug Research*, 2011. v. 67, p. 25-49.
- prizes/medicine/laureates/1995/
- RAINA, M.; IBBA, M. TRNAs as regulators of biological processes. *NULL: Frontiers in Genetics*, 2014. v. 5, n. JUN, p. 1-14.
- RODGERS, A. B. et al. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences*, 2015. v. 112, n. 44, p. 13699-13704.
- SHARMA, S. et al. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genetics*, 2011. v. 7, n. 2.
- SPITALE, R. C.; TSAI, M. C.; CHANG, H. Y. RNA templating the epigenome: Long noncoding RNAs as molecular scaffolds. *Epigenetics*, 2011. v. 6, p. 539-543.
- SWEATT, J. D. Experience-Dependent Epigenetic Modifications in the Central Nervous System. *Biological Psychiatry*, 2009. v. 65, n. 3, p. 191-197. Disponível em: <<http://dx.doi.org/10.1016/j.biopsych.2008.09.002>>.
- The Nobel Prize in Physiology or Medicine 1995. http://nobelprize.org/nobel_
- TSANKOVA, N. et al. Epigenetic regulation in psychiatric disorders. *Nature reviews. Neuroscience*, 2007. v. 8, n. May, p. 355-367.
- WADDINGTON, C. H. The epigenotype. *Endeavour*, 1942. v. 1, p. 18-20.

- WEBER, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature genetics*, 2005. v. 37, n. 8, p. 853-862.
- YAN, M. S.-C.; MATOUK, C. C.; MARSDEN, P. A. Epigenetics of the vascular endothelium. *Journal of Applied Physiology*, 2010. v. 109, n. 3, p. 916-926. Disponível em: <<http://jap.physiology.org/cgi/doi/10.1152/japphysiol.00131.2010>>.
- YODER, J. A.; WALSH, C. P.; BESTOR, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, 1997. v. 13, n. 8, p. 335-340.

EVO-DEVO E A EXPANSÃO DAS FRONTEIRAS DA BIOLOGIA EVOLUTIVA

Gilberto Cavalheiro Vieira

Apresentação

Evo-Devo é uma das designações dadas à disciplina conhecida como Biologia Evolutiva do Desenvolvimento, ou Evolução do Desenvolvimento. Relativamente nova no campo da Biologia Evolutiva, a Evo-Devo tem como objetivo o entendimento dos mecanismos responsáveis pelas mudanças evolutivas dos organismos, através de estudos comparativos entre sequências de genes e processos que regulam o desenvolvimento embrionário dos mais diversos organismos, de modo a compreender como novas formas, funções e comportamentos poderiam emergir. O termo “Biologia Evolutiva do Desenvolvimento” foi utilizado pela primeira vez no livro de Peter Calow, em 1983, chamado *Princípios Evolutivos*, em que descreveu a Evo-Devo como uma área relativamente nova e de grande complexidade, que unificava os campos da evolução e do desenvolvimento (embriologia). Contudo, as raízes da Evo-Devo são bem mais profundas, como iremos perceber a seguir. Ao longo do presente capítulo, iremos passear um pouco pela origem, desafios, questões abordadas, mecanismos e as consequências advindas desse novo campo da Biologia Evolutiva – sem qualquer pretensão de exaurir o tema, que por si só é complexo e encantador.

Os primeiros passos da embriologia na evolução

Quando Charles Darwin publicou *A origem das espécies*, em 1859, destacou a importância dos estudos de embriologia (desenvolvimento dos organismos) para o entendimento dos

mecanismos e processos evolutivos das espécies. A partir de então, diversos morfologistas/embriologistas passaram a abordar o desenvolvimento dos organismos sobre uma nova ótica – à luz da evolução. A teoria da recapitulação de Ernst Haeckel (1866) desempenhou forte influência sobre a área: o autor propunha que a ontogenia (desenvolvimento) recapitula a filogenia (evolução) dos organismos. O interessante desse contexto histórico é que o primeiro estudo publicado testando a teoria evolutiva de Darwin, utilizando embriões e larvas, foi o de Fritz Müller (1864). Através da observação da ontogenia de crustáceos, Müller demonstrou que a ontogenia poderia ser utilizada para entender os padrões da história evolutiva das espécies (filogenia); ou seja, mecanismos evolutivos podem ser encontrados na ontogenia (desenvolvimento).



Figura 1. Johann Friedrich Theodor Müller (1821-1897) – formado em Filosofia na Universidade de Berlim, Alemanha. Em 1852 deixou a Alemanha e firmou residência em Blumenau, Santa Catarina. Lá desenvolveu um intenso trabalho com observações de diversos organismos, com destaque aos crustáceos. Fritz Müller foi o primeiro a testar experimentalmente as ideias de Darwin, realizando estudos comparativos de embriologia, ontogenia, ecologia, fisiologia e morfologia. Era chamado carinhosamente por Darwin de “Príncipe dos Observadores”. O capítulo 12 deste livro é dedicado ao autor.

As estratégias de vida dos crustáceos observados por Müller eram tão variadas que o autor percebeu que poderia usar os detalhes e as variedades dos estágios da vida desses organismos para estabelecer as relações evolutivas entre as diversas espécies.

Entretanto, Haeckel tomou o caminho completamente oposto ao de Müller em sua abordagem, teorizando que a filogenia (evolução) explicaria a ontogenia (desenvolvimento), como descrita nas bases de sua Lei Biogenética. Por sinal, para desgosto de Darwin, Haeckel considerava a evolução das espécies por seleção natural apenas um fragmento de sua teoria universal do desenvolvimento das espécies.

Dessa maneira, a embriologia providenciou uma forma de estudar e entender a evolução das espécies. Ao final de 1860 até a metade de 1880, a embriologia evolutiva foi o campo que mais atraiu pesquisadores, destacando-se grandes zoólogos, como William Bateson, pesquisador inglês que cunhou o termo “genética”, mas que começou sua carreira no campo da embriologia evolutiva. Seu entusiasmo com a área era tal que chegou a comentar que a embriologia seria a quintessência da morfologia, sendo que todo zoólogo deveria ser um embriologista, e o tema central de toda e qualquer conversa profissional deveria ser a evolução (Bateson 1922). Porém, alguma coisa fez com que esse encanto se desfizesse. A tal ponto que, mesmo nos dias de hoje, a embriologia seja ensinada nas escolas (e no ensino superior) como uma área praticamente restrita ao campo descritivo-morfológico, como que desconectada com as bases evolutivas da Biologia. Mas em que momento essa fogaosa paixão se apagou?



Figura 2. Theodosius Dobzhansky (1900-1975) geneticista e evolucionista russo (nasceu na Ucrânia, que fazia parte do Império Russo na época), sendo um dos responsáveis pela unificação da Biologia Evolutiva na Síntese Moderna da Evolução. Aos 27 anos migrou para os Estados Unidos a convite da Fundação Rockefeller, onde trabalhou ao lado de Thomas Hunt Morgan (pioneiro no uso de drosófilas em experimentos genéticos). Em sua obra, *Genética e a Origem das Espécies*, definiu evolução como um processo de mudanças nas frequências alélicas dentro do conjunto de uma população ao longo das gerações.

Até 1880, diversos trabalhos procurando abordar a reconstrução das relações evolutiva das espécies utilizando as sequências do desenvolvimento embrionário comparado com diversos organismos se mostraram infrutíferos; além disso, diferentes problemáticas com abordagens vitalistas e obscuras para a explicação de como se dava o desenvolvimento dos organismos e, principalmente, a redescoberta da genética mendeliana, em 1900, acabaram colocando toda a embriologia evolutiva de lado, posição que demorou mais de um século para ser mudada. A genética mendeliana, apresentando a segregação dos fatores hereditários (genes), suportada por estudos com moscas-da-fruta (*Drosophila*), estabeleceram as fundações onde a ciência genética seria erguida poderosamente.

A publicação do livro seminal *Genética e a Origem das Espécies*, de Theodosius Dobzhansky (1937), sentenciou ao ostracismo a embriologia evolutiva, estabelecendo as bases do entendimento da evolução das espécies através da genética de populações, fundamentada em modelos matemáticos desenvolvidos na década de 1920. A partir da metade do século XX

temos um cenário evolutivo baseado na fusão das leis mendelianas e a genética de populações, em conjunto com a paleontologia, que trazia os registros historiográficos da evolução das espécies.

O reencontro entre a filogenia (evo) e a ontogenia (devo)

Em 1922, William Bateson – o mesmo biólogo entusiasmado com a embriologia apresentado anteriormente – lança um ensaio intitulado “Evolutionary Faith and Modern Doubts”, onde sentenciou à morte a embriologia no campo evolutivo, exaltando a genética ao afirmar que os geneticistas seriam os sucessores dos morfologistas/embriologista no século XX. Nesse princípio de século, os experimentos com *Drosophila*, organismo-modelo para genética, acabaram eliminando o desenvolvimento embriológico e o ambiente de papéis significativos dentro da discussão da teoria evolutiva.

Diversos pesquisadores tentaram ao longo das primeiras sete décadas do século XX estabelecer a ligação entre a embriologia e a evolução, como foi o caso do geneticista Richard Benedict Goldschmidt (1878-1958). Goldschmidt se destacou por seus trabalhos visando ao entendimento de processos como a determinação do sexo, metamorfose (questionando como um mesmo genótipo pode dar origem a uma lagarta e a uma mariposa) e o surgimento de fenocópias (como um mesmo genótipo pode dar origem a formas alternativas em resposta às mudanças ambientais), sendo pioneiro no entendimento dos eventos de heterocronia (mudança do tempo na ocorrência de certos eventos do desenvolvimento embrionário, que podem levar a mudanças de tamanho e de forma final do organismo) (Goldschmidt 1918). Assim como Conrad Waddington (1940), Goldschmidt também se debruçou em pesquisas com o intuito de compreender fenômenos como a norma de reação (descreve o padrão de expressão fenotípica a partir de um dado genótipo em resposta às variações ambientais) e assimilação genética (processo

pelo qual um fenótipo originalmente produzido em resposta a uma condição ambiental, como a exposição a um teratógeno, torna-se posteriormente geneticamente codificado por seleção artificial ou seleção natural).

Vale destacar a criação do termo “Monstro Esperançoso”, por Goldschmidt, apresentado em seu livro *The Material Basis of Evolution* (1940). Em seu livro, Goldschmidt defende a ideia de que as mudanças evolutivas não são sempre graduais e contínuas, mas eventualmente grandes mudanças macroevolutivas podem ser geradas via “reações sistêmicas” e macromutações. Ou seja, Goldschmidt argumentava que as grandes mudanças evolutivas eram consequências de macromutações e tal hipótese foi conhecida como “Hipótese do Monstro Esperançoso”. O pressuposto da Síntese Moderna de evolução por acumulação lenta e gradual de pequenas mutações era um fator importante, porém, só poderia dar conta de explicar as variações dentro das espécies (microevolução), não sendo robusta o suficiente para explicar a evolução de novidades morfológicas de novas espécies. Assim, o autor defendia a ideia de que as diferenças genéticas profundas entre as espécies seriam necessárias (macromutações) como recurso para grandes mudanças genéticas (macroevolução) que, por conseguinte, em alguma ocasião poderia originar os “monstros esperançosos”, sendo que essas mudanças poderiam ocorrer nos programas genéticos de desenvolvimento dos organismos.

Não é difícil imaginar que hipóteses como as de Goldschmidt não encontrariam eco diante do cenário em que a Teoria Evolutiva se erguia (genética populacional, modelos matemáticos, seleção natural, modificações cumulativas por mutações ocorrendo de maneira lenta e gradual). Suas ideias não eram nada ortodoxas diante do paradigma dominante das décadas de 1940 e 1950.

A reunião entre a embriologia e evolução só se daria no final da década de 1970 e início da década de 1980, mais precisamente – se é que podemos estabelecer um marco histórico –

com a publicação do livro *Ontogeny and Phylogeny*, de Stephen J. Gould (1977). Nessa incrível obra de mais de 500 páginas, intensa, densa e rica em detalhes, Gould nos leva de volta ao passado, até 450 a.C., relatando como se deu, ao longo de mais de dois milênios, a relação entre embriologia e evolução. Um trabalho de fato hercúleo, que nos apresenta que desde muitos séculos viemos buscando estabelecer as relações envolvidas entre desenvolvimento e evolução. Gould revisa o conceito elaborado por Haeckel, em 1860, e por Gavin de Beer, em 1930, conhecido como “heterocronia” – mudanças nos padrões temporais dos processos de desenvolvimento entre um descendente e seus ancestrais.

Inspirado nos trabalhos de Goldschmidt (1918), em suas discussões sobre como genes podem afetar os padrões fisiológicos em insetos a partir de resposta às mudanças no ambiente, de Beer sugeriu que tais mudanças poderiam ter implicações na evolução das espécies. Temos um exemplo claro de como os estudos investigativos da embriologia comparada podem lançar luz às questões evolutivas. A heterocronia, como um mecanismo genuíno do desenvolvimento, opera durante determinado estágio ontogenético, podendo ser alvo da seleção natural e, assim, importante para as mudanças evolutivas.

Interessante notar que Julian Huxley, um dos responsáveis pela Síntese Moderna, destacava a importância dos estudos no campo do desenvolvimento, mais precisamente em relação aos genes envolvidos na ontogenia dos organismos – “o estudo de genes durante o desenvolvimento é tão essencial para uma compreensão da evolução quanto o estudo da mutação e seleção” (Huxley 1942, p. 8) –, porém assim como os demais proponentes da Síntese, Huxley não advogou a favor de incorporar a embriologia ao corpo da teoria evolutiva.

Já Ernst Mayr utilizou de outra abordagem para retirar a importância da embriologia da evolução, argumentando que a biologia das causas “próximas” (no qual o desenvolvimento está inserido) e a biologia das causas “últimas” (evolução e interações

ecológicas) seriam dois campos distintos e que não teria sobreposição ou interação uma com a outra (Mayr 1982, 1997). De fato, a Evo-Devo ganharia destaque somente quando a heterocronia foi se mostrando cada vez mais presente em diversos estudos comparativos de tecidos ao longo do desenvolvimento de espécies de diferentes classes e até mesmo filos, a partir da década de 1980 (MacDonald and Hall, 2001; Tills et al., 2011).

A “caixa de ferramentas” e a inesperada falta de diferenças

Foi em 1978, através do trabalho publicado por Lewis, *A gene complex controlling segmentation in Drosophila*, que a Evo-Devo teve seu maior impulso. Não foi uma forma de ressurgimento das cinzas do que previamente se sabia sobre os mecanismos evolutivos de desenvolvimento, mas uma descoberta desconcertante em um cenário baseado na ideia de diferenças genéticas sustentando a diversidade biológica ao longo da evolução: os genes **homeóticos**. Genes homeóticos, também conhecidos como “Hox”, são genes responsáveis pela produção de fatores de transcrição que apresentam uma sequência conservada de 180 pares de bases, conhecida como “homeobox”. Esses genes estão presentes em todos os animais, plantas e fungos, com alto grau de conservação. Nos animais, os genes homeóticos são responsáveis pela determinação dos eixos antero-posterior, dorso-ventral e na especificação de regiões (como número de segmentos) ao longo do corpo do organismo – cabeça em uma extremidade, cauda na outra, tórax antes do abdômen, asas presentes em um par específico de segmentos e assim por diante (Lewis 1978; Gehring 1985, 1998; Averof 1997; Grenier et al. 1997; Carroll 2008).

Ironicamente (ou não), o mesmo organismo modelo que lançou os estudos da evolução do desenvolvimento ao limbo na teoria evolutiva foi o ser vivo que a resgatou: a *Drosophila*.

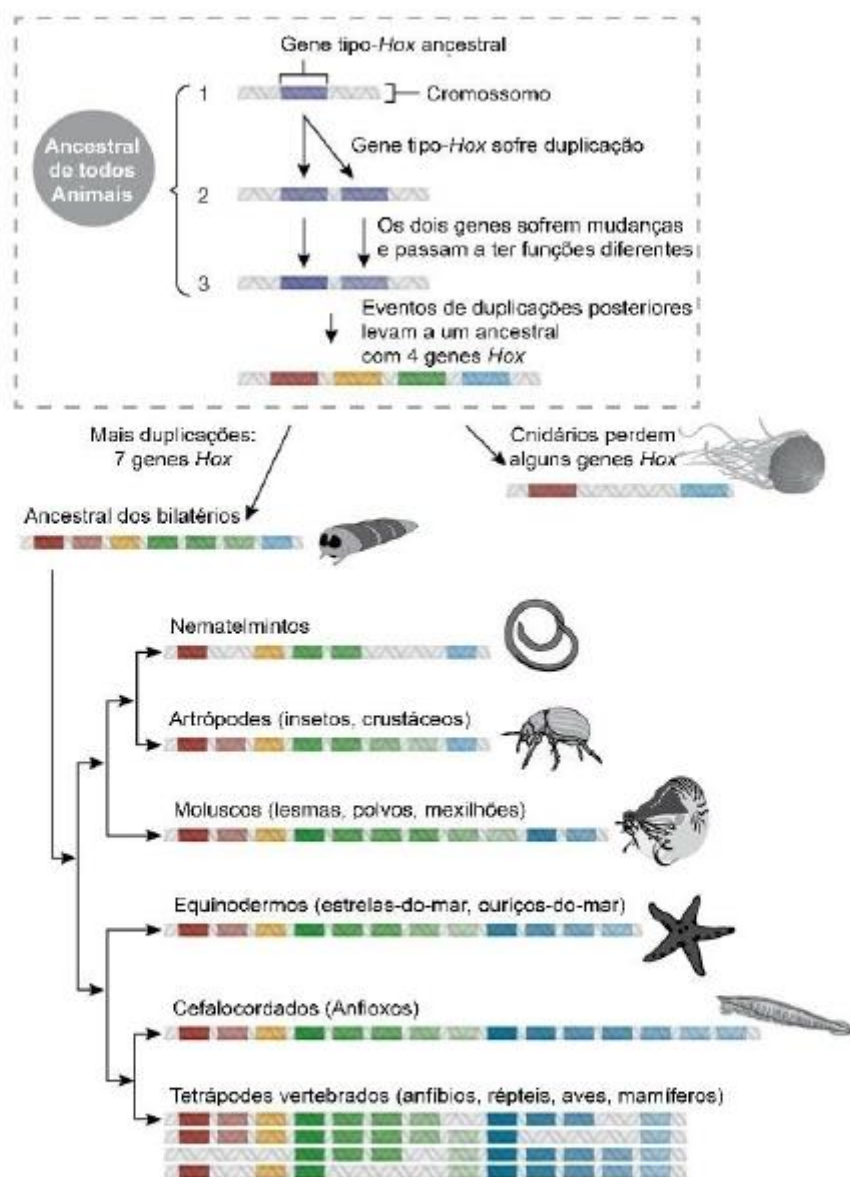


Figura 3. Os genes Hox são mais similares entre espécies próximas evolutivamente. Através da comparação das sequências dos genes Hox, os cientistas podem determinar a história evolutiva das espécies, verificando quando ocorreram os eventos de duplicação e em quais ramos ocorreram eventos de perda de um ou outro gene Hox (adaptado de <http://learn.genetics.utah.edu>).

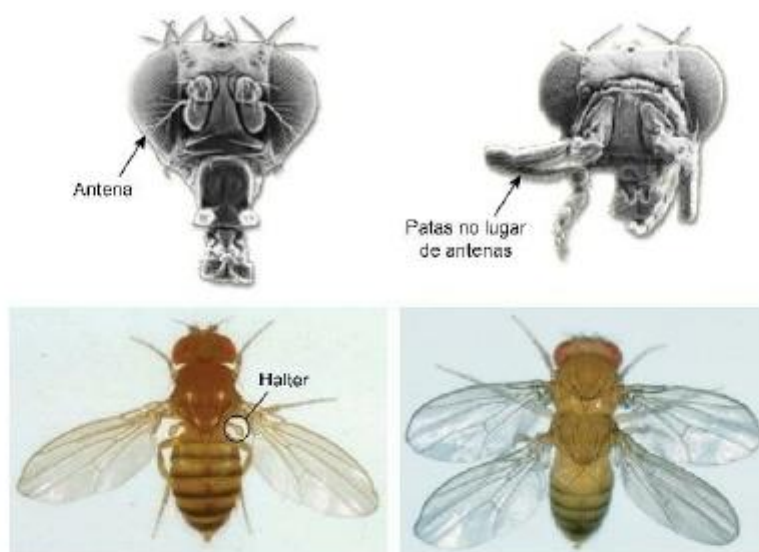


Figura 3. Alguns fenótipos oriundos de mutações em genes Hox. Acima, o fenótipo resultante de mutação no gene *antennapedia*, onde a drosófila acaba apresentando patas no lugar que deveriam ser as antenas. Abaixo, o fenótipo correspondente à mutação no gene *ultrabitorax*, fazendo com que a o indivíduo se desenvolva com dois tórax. Fotografia: F. R. Turner, Indiana University.

Quanto mais os estudos avançavam na descoberta de mais genes homeóticos e sua profunda conservação entre os mais diversos organismos (mesmo aqueles muito distantes filogeneticamente), mais surpresos os pesquisadores na área da Biologia Evolutiva ficavam: **quando nós mais esperávamos encontrar variação, encontramos conservação, uma falta de mudança** (Gerhart e Kirschner, 1997). A descoberta dos “genes-mestres”, também conhecidos como “genes do desenvolvimento” ou “genes regulatórios”, abriu campo para novas formas de observar os mecanismos pelo qual novas formas poderiam emergir através da modulação desses genes. Um dos genes mais bem estudados de genes-mestres são os conhecidos como *paired-box*, mais especificamente o *Pax-6* (em vertebrados) ou *eyeless* (em *Drosophila*). O gene *Pax-6* está presente em todos os organismos do reino animal e sua função está ligada na “anteriorização” do corpo do embrião e formação das estruturas e órgãos sensoriais

dos animais. Esse gene é reconhecido, principalmente, pelo controle do desenvolvimento dos olhos.

O mais interessante é que os genes homeóticos possuem equivalência funcional entre genes ortólogos distantes (genes que divergiram por especiação); ou seja, genes provenientes de organismos cujo ancestral comum viveu há mais de 500 milhões de anos atrás podem ter funções semelhantes.

Voltemos ao gene homeóticos *Pax-6*, que é responsável pelo desenvolvimento dos olhos nos organismos: de fato, o gene *eyeless* (que o correspondente ao *Pax-6* em vertebrados) de moscas-da-fruta pode iniciar olhos em sapos e o *Pax-6* de sapos podem iniciar olhos em *Drosophila* (Dahl et al., 1997; Kammermeier et al., 2001; Suga et al., 2010). Isso é uma conservação genética surpreendente e que jamais foi cogitada, até então, que pudesse ocorrer em animais cuja morfologia varia tão enormemente. É claro que nesse caso as moscas não terão olhos de sapos e nem sapos terão olhos de moscas, pois os genes homeóticos apenas endereçam as regiões e planos corporais, a estrutura final será dada pela ação de diversos outros genes.

Entre as décadas de 1980 e 1990, temos o campo da Biologia Evolutiva fervilhante com o advento de novas tecnologias de análise molecular, em conjunto com os estudos de mecanismos de desenvolvimento como a heterocronia, genes homeóticos, estudos no campo da ecologia com as interações entre ambientais e o ressurgimento da epigenética, levando a Evo-Devo a tomar forma como uma nova disciplina robusta dentro do campo de estudos evolutivos.



Figura 4. Diferentes fenótipos resultantes de mutações ocorridas no gene homeótico *Pax-6*. Os fenótipos tipo selvagem (TS) estão representados acima e as mutações (MT) ocorridas nos ortólogos de *Pax-6* em humanos, camundongos, peixe-zebra e *Drosophila*, abaixo. As descrições dos fenótipos mutantes (FT) estão listadas abaixo das imagens de cada uma das espécies. Interessante notar a correspondência fenotípica entre os mutantes de espécies tão distantes filogeneticamente. Adaptado de Washington et al., 2009.

Os estudos de Evo-Devo passaram a abordar questões como:

- A origem e evolução do desenvolvimento embrionário;
- O papel exercido pelas modificações nos processos de desenvolvimento e suas consequências na produção de novas características (fenótipos);
- Como a plasticidade adaptativa do desenvolvimento facilita a origem e manutenção de ciclos de vida complexos, com embriões, larvas e adultos (de certa forma, um resgate dos trabalhos de Waddington e Goldschmidt);
- De que maneira a interação dos organismos com o ambiente pode fomentar mudanças evolutivas;
- Integração entre os mecanismos genéticos e epigenéticos;
- Manutenção das novidades e variações fenotípicas em resposta às mudanças ambientais e nos programas de desenvolvimento embrionário.

Expandindo as fronteiras da Biologia Evolutiva

Assim como o que ocorreu com a Epigenética, os avanços tecnológicos na área da Biologia Molecular promoveram o que podemos denominar como uma revolução da Evo-Devo. Em questão de duas décadas, diversos trabalhos detalharam os mecanismos moleculares que operam à nível celular e tecidual ao longo do desenvolvimento dos organismos. É como que se a caixa-preta, antes inacessível, que ligava o genótipo e fenótipo fosse aberta. O mais interessante dentro desse novo cenário é que, como já dito anteriormente, acabou-se revelando que os genes envolvidos no desenvolvimento dos organismos se apresentam com alto grau de conservação, dentre os mais diversos filos e até mesmo entre reinos. Dessa forma, os olhares passaram a vasculhar os sistemas de regulação desses genes, as mudanças nos sistemas de regulação-cis dos genes seriam mais significantes que mudanças no número de genes ou função proteica (Carroll 2000).

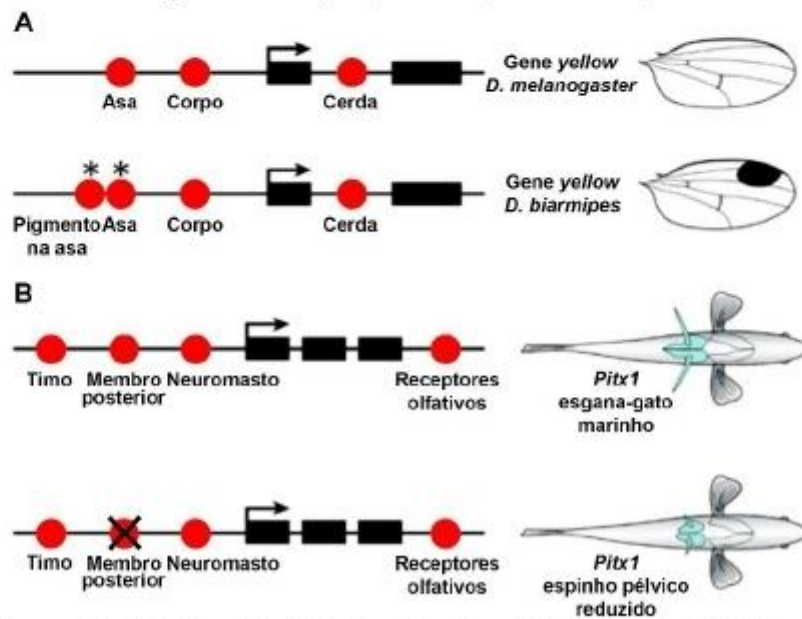


Figura 5. A relação de modularidade de regiões cis-regulatórias de genes pleiotrópicos permite que diferentes partes do corpo possam evoluir de maneira independente: a) a

expressão do gene de pigmentação *yellow* em *Drosophila* está sob o controle de diversas regiões cis-regulatórias (círculos vermelhos). As diferenças existentes no aparecimento de pigmentação nas asas de *D. biarmipes* (espécie irmã de *D. melanogaster*) resultam do aparecimento de uma nova região cis-regulatória na linhagem de *D. biarmipes*; b) da mesma maneira, a expressão do gene *Pitx1* em vertebrados é controlado a partir de múltiplas regiões cis-regulatórias (círculos vermelhos). Na região pélvica de peixes esgana-gato (*Gasterosteus aculeatus*) há espinhos pélvicos. Em algumas linhagens de água-doce, esses espinhos são reduzidos em decorrência de uma perda seletiva de atividade do elemento regulador do membro posterior. Adaptado de Carroll, 2005.

Ao longo das últimas três décadas, a Biologia Evolutiva vem ampliando seu arcabouço, integrando novos campos, antes deixados de lado pela Síntese Moderna na década de 1930. Biologia do desenvolvimento e paleontologia, além de áreas que surgiram depois, como a Biologia Molecular e Genética Molecular, passaram a integrar e transformar o cenário de estudos evolutivos, revelando questões nunca antes imaginadas. As concepções de como os processos evolutivos ocorrem acabaram se transformando, pois as mudanças evolutivas não se dão diretamente da transformação de um adulto ancestral para um descendente adulto, mas sim quando os processos de desenvolvimento produzem novas características a cada geração ao longo da evolução de uma dada linhagem. Darwin (1859) já havia descrito que o desenvolvimento é responsável por introduzir variações ao nível do indivíduo. Se por um lado, o desenvolvimento pode ser descrito como um processo geneticamente programado e cíclico, a evolução é não programada e contingente. A Evo-Devo, dessa forma, delega uma base teórica para explicar a origem de novidades fenotípicas (Pigliucci e Müller, 2010) como as penas das aves, a origem das flores, o desenvolvimento de asas em insetos, a origem das carapaças em tartarugas, entre outros exemplos (Prum e Brush 2002; Xu et al., 2011; Leiarert et al., 2011; Carrol et al., 1995; Gilbert et al., 2001).

No campo da Biologia Evolutiva, passamos a entender como pequenas mudanças no genótipo levam ao surgimento de novos fenótipos durante a evolução, levando à compreensão de como a microevolução pode contribuir com a macroevolução, com

o surgimento de novas classes e ordens de organismos. Os pesquisadores, então, passam a questionar dentro da Evo-Devo como os processos de desenvolvimento podem impor um viés e restrições no direcionamento da evolução e no surgimento de novas variantes fenotípicas. Assim, qualquer limitação imposta pelos programas de desenvolvimento sobre os fenótipos irá afetar os tipos morfológicos possíveis de surgirem e suas respostas à seleção natural. É justamente essa limitação nos programas de desenvolvimento dos organismos que, por exemplo, não faz possível o surgimento de asas nas costas de mamíferos ou rodas no lugar de patas.

O surgimento de novas variantes fenotípicas pode ocorrer de maneira rápida ou muito lentamente, dependendo de como se dá as modificações entre as interações em diversos níveis ao longo do desenvolvimento dos organismos. Os mecanismos de desenvolvimento evolutivo são multiníveis, podendo ser genético, celular, ontogenético, fisiológico, hormonal ou qualquer combinação entre esses. Como o desenvolvimento é cíclico e programado, percebe-se que há uma hierarquia, com novas propriedades e mecanismos emergindo ao longo do processo de desenvolvimento embrionário, sendo cada passo dependente do processo ou estágio anterior. Por exemplo, quando ocorrem sucessivas divisões celulares, temos uma crescente rede de interações célula-célula, formando o que podemos chamar de “ambiente interno”. Modificações das interações dentro desse ambiente interno podem desencadear alterações nas redes de interações genéticas, respondendo, assim, com o surgimento de variantes fenotípicas a cada estágio, fornecendo a ligação regulatória entre o genótipo e os módulos celulares (Davidson 2006; Wagner et al. 2007; Wilkins 2007).

As restrições impostas aos sistemas interativos ao longo do desenvolvimento estão ligadas às limitações mecanicistas definidas pelo próprio desenvolvimento, em decorrência da modularidade dentro da ontogenia. Em insetos, o crescimento das características

do corpo, como os chifres (Stern & Emlen 1999), está ligado ao tamanho do corpo, através de uma regulação comum mediada pelo hormônio juvenil, sugerindo mecanismos para a variação evolutiva. A variação na alocação de recursos de uma parte do corpo tem potencial de afetar a forma ou tamanho de outras partes do corpo, indicando interações que controlam o crescimento corporal do organismo que podem impor limitações no desenvolvimento (Figura 6). A escala de partes do corpo também pode ser modificada em resposta à seleção artificial, o que fornece uma ligação entre a microevolução e o desenvolvimento.

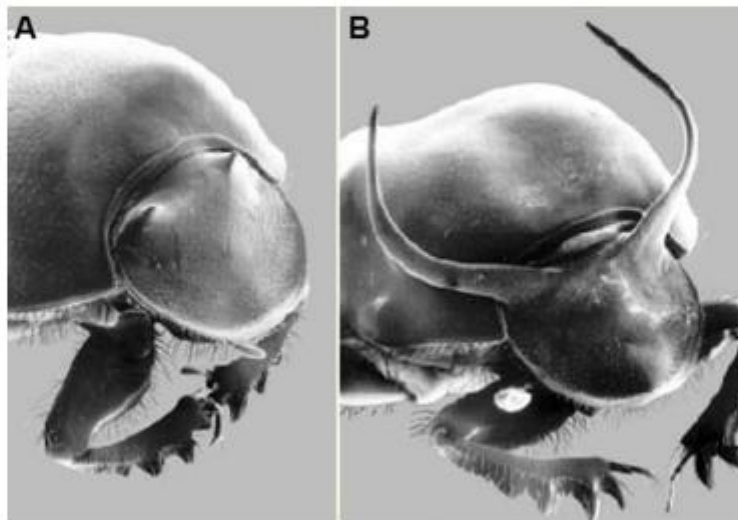


Figura 6. Polifenismo em besouros machos da espécie *Onthophagus taurus*.

Uma estruturação modular acaba por gerar restrições evolutivas porque algumas interações entre módulos podem ser difíceis de desacoplar. No entanto, de maneira paradoxal, a modularidade também pode permitir uma mudança evolutiva acentuada, porque muitas interações inter-modulares podem ser dissociadas no tempo (heterocronia) ou em outras formas que permitem padrões de desenvolvimento viáveis, embora alterados (Raff, 1996; Wagner, 1996). Novamente, o que podemos observar é

que os mecanismos genéticos que permitem tais dissociações modulares provavelmente se encontram em uma combinação de regiões promotoras dos genes (ao invés dos genes propriamente ditos), que permitem a modificação de expressão gênica de várias maneiras, e podendo ser prontamente cooptado a novas funções. Dessa maneira, podemos ter redes de interações modulares com caminhos de desenvolvimento diferentes que podem convergir para estruturas similares.

Para resumir, de forma geral, podemos dizer que a Evo-Devo está amparada sobre oito princípios básicos:

- **Pleiotropia:** a maior parte das proteínas envolvidas na regulação do desenvolvimento participam de múltiplos e independentes processos na formação e padronização de estruturas morfológicas bem distintas;
- **Complexidade da Ancestralidade Genética:** animais morfológicamente distintos (principalmente quando adultos) e que divergiram a partir de um ancestral comum há muitos milhões de anos atrás possuem um constituinte genético (*toolkit* – caixa de ferramentas) muito similar de genes responsáveis pela padronização do corpo;
- **Equivalência funcional de ortólogos e parálogos distantes filogeneticamente:** genes provenientes de organismos cujo ancestral comum viveu há mais de 500 milhões de anos atrás podem ter funções semelhantes e estar envolvidos em rotas similares ou distintas no desenvolvimento;
- **Homologia profunda:** geralmente homoplasias (diferentes ambientes ou pressões seletivas que possibilitam o aparecimento ou o reaparecimento de características semelhantes em organismos que não compartilham um ancestral comum mais recente) são descartadas imediatamente em cladogramas, pois podem provocar problemas nas relações evolutivas entre grupos distantes filogeneticamente. Entretanto, tal ortodoxia é desafiada pela Evo-Devo. Por exemplo, para o desenvolvimento do coração de um humano, assim como para o desenvolvimento das bombas do sistema circulatório de um artrópode, o mesmo gene-mestre é requisitado: o *Tinman* (“homem de lata”, em referência ao personagem do Mágico de Oz que procura por um coração). Poderíamos assumir que o coração de invertebrados e vertebrados surgiram por convergência, sendo assim uma característica homoplásica, porém, com as evidências que temos agora,

sabemos que o mesmo gene é necessário para a formação desse órgão em grupos tão distintos e distantes filogeneticamente. A presença de coração nesses organismos passa a ser descrita como uma homologia profunda;

- **A queda do principal fator conhecido como “motor evolutivo”:** diferentemente do que se pensava até então, as duplicações gênicas passam a não ser mais consideradas o principal mecanismo fomentador de novidades evolutivas. Ao invés disso, temos que as modificações nas sequências cis e trans regulatórias dos genes envolvidos no desenvolvimento é que passam a ser os motores evolutivos, fomentando o aparecimento de novos padrões morfológicos;
- **Os mecanismos centrais da Biologia do Desenvolvimento: heterotopia, heterocronia, heterometria e heterotipia:** a **heterotopia** é a mudança na localização da expressão dos fatores de transcrição e proteínas de sinalização parácrinas, ou seja, uma modificação espacial da expressão gênica. A **heterocronia**, já descrita anteriormente, é a mudança no *timing* dos processos de desenvolvimento. Na heterocronia, um módulo pode ter seu tempo de expressão ou de crescimento alterado em relação a outros módulos do embrião em desenvolvimento, sendo amplamente observada nos vertebrados, por exemplo, no desenvolvimento das estruturas ósseas do organismo. A **heterometria** é a modificação da quantidade de um determinado produto gênico ou de uma estrutura. Por fim, a **heterotipia** está relacionada com a modificação de regiões codificantes de um determinado gene, fazendo com que ocorra a mudança de suas propriedades funcionais da proteína sintetizada. Cabe notar que diferentemente da heterotipia, heterocronia e heterometria em que as mutações nas regiões regulatórias dos genes é que fomentam as alterações, na heterotipia as mutações ocorrem no próprio gene;
- **Modularidade das regiões cis-regulatórias:** como descrito anteriormente, com a emergência da Evo-Devo, os olhares dos biólogos evolutivos voltam-se muito mais às modificações nas regiões regulatórias dos genes do que propriamente às mutações das regiões codificantes. A modulação da expressão de um dado gene é capaz de, associado com as redes de interações do programa de desenvolvimento, sustentar a emergência de variantes fenotípicas;
- **Fatores de transcrição:** mutações em genes responsáveis pela codificação de fatores de transcrição podem fomentar modificações nas interações das sequências-alvos que tais fatores atuavam anteriormente. Por consequência, podemos ter a modificação espaço-

temporal e na quantidade de determinados genes, alterando o programa de desenvolvimento.

Em um cenário de entendimento mais amplo no contexto evolutivo, vemos que os genes não fazem estruturas, mas sim, os processos de desenvolvimento é que são responsáveis pela estruturação do organismo – através de forças físicas como estímulos mecânicos, temperatura do ambiente, interação com produtos químicos produzidos por outras espécies, entre outros. Dois pontos aparecem intimamente relacionados nesse contexto: a **citodiferenciação** (nível celular) e a **morfogênese** (organização 3D dos diferentes tipos celulares) – nível supracelular, interações célula-célula e célula-ambiente. Ambos os fenômenos (citodiferenciação e morfogênese) são passíveis das ações de modulação dos mecanismos de Evo-Devo, promovendo o surgimento de novos padrões fenotípicos.

A plasticidade fenotípica

A plasticidade fenotípica nada mais é do que a capacidade de um organismo modificar seu fenótipo em resposta às mudanças do ambiente, devido à interação no desenvolvimento biológico (West-Eberhard, 2003). Vale lembrar que quando nos referimos às mudanças ambientais, fazemos referência aos estímulos interno e externos ao longo do desenvolvimento. Diferentemente dos demais mecanismos da Evo-Devo, a plasticidade do desenvolvimento possui uma abordagem de contexto populacional, enquanto os demais estão inseridos em um cenário do organismo individual.

Por diversas décadas, a plasticidade fenotípica foi estudada, sendo abordada como consequência de modificações ao longo do desenvolvimento e não como causa e matriz de emergência de novos fenótipos (Levins 1968; Via e Lande 1985; Schlichting 1986; Stearns 1989; Pigliucci, 2001). A principal contribuição dos

estudos, levando-se em conta a plasticidade fenotípica no campo da evolução, está ligada à forma como a **acomodação genética e fenotípica** proporcionam a oportunidade de fixação de novidades oriundas da mudança de programas do desenvolvimento.

Por **acomodação fenotípica** entende-se como ajustes funcionais de partes de um organismo ao longo do desenvolvimento embrionário que não envolvem mutações genéticas. As acomodações fenotípicas podem promover o que chamamos de **acomodação genética** se esses novos fenótipos forem estabilizados e ajustados ao longo das gerações por seleção de variações genéticas existentes, por variações genéticas que se eram crípticas ou por mutações recentes que fixem o novo programa de desenvolvimento (West-Eberhard 2003; Lande 2009; Baldwin 1902; Schmalhausen, 1949). A acomodação genética, dessa maneira, pode proporcionar um mecanismo de rápida adaptação em resposta às mudanças ambientais, no qual o ambiente induz e seleciona fenótipos alternativos simultaneamente. Dentro desse campo, verifica-se que a plasticidade dentro das espécies pode gerar diferenças fenotípicas paralelas às exibidas por espécies estreitamente relacionadas e a divergência evolutiva entre linhagens descendentes também tem sido associada à plasticidade fenotípica ancestral.

Outro conceito central relacionado à plasticidade fenotípica é o da **modularidade**, um conceito em que o plano corporal (*Bauplan*) dos organismos está organizado em diversas unidades fenotípicas (módulos). Os módulos fenotípicos estão dispostos de maneira que não sejam completamente dependentes, nem totalmente independentes, dos demais módulos que compõem o corpo de um organismo. Dessa maneira, as partes de um organismo possuem propriedades de dissociabilidade e integração ao longo do desenvolvimento. Para West-Eberhard, o processo fundamental que define a organização modular é o desenvolvimento, sendo que uma unidade modular deve possuir três critérios:

- Ter uma função principal;
- Estar integrado em seus diversos caracteres por efeitos pleiotrópicos; e
- Ser semi-independente de outras unidades modulares.

A plasticidade fenotípica se faz presente e observável quando temos, por ação de fatores ambientais ou mutações, a modificação da regulação de genes pleiotrópicos que resultam na expressão de diferentes subunidades modulares (fenótipos alternativos). Quando efeitos ambientais ou mutações, de maneira cumulativa, extrapolam o limiar de determinação de um dado padrão (via de desenvolvimento) de uma subunidade fenotípica A, temos a mudança de “chave” dessa via, que resulta na coexpressão de uma subunidade fenotípica B. Um exemplo clássico desse mecanismo pode ser encontrado na modulação hormonal na formação dos bicos dos diferentes tentilhões de galápagos (Figura 7).

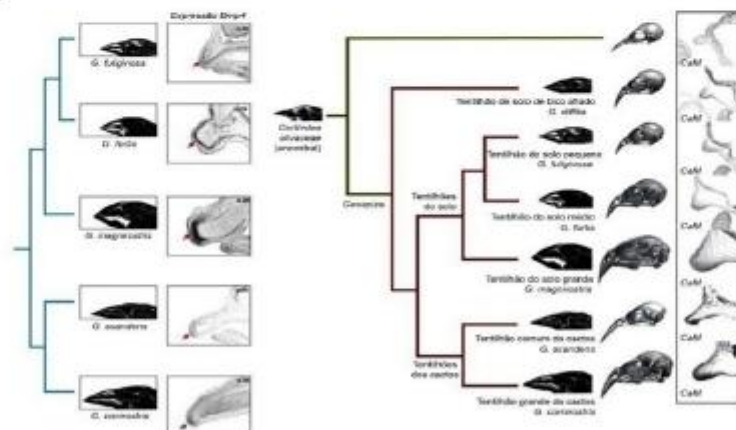


Figura 7. A expressão dos genes *Bmp4* e *Calmodulina* (*CaM*) podem lançar luz à respeito das variações encontradas nos Tentilhões de Darwin. A proteína BMP4 acaba por ser regulada por heterocronia, assim como por heterometria. Já a calmodulina é regulada por heterometria. Ao longo do desenvolvimento, as modulações desses dois genes promovem a manifestação de diferentes formas de bicos, ao passo que a seleção natural irá permitir que determinadas morfologias sobrevivam diferencialmente de acordo com as pressões seletivas envolvidas. Imagens modificadas de Abzhanov et al., 2006 e 2004.

Os módulos de desenvolvimento parecem ter emergido através da integração de processos genéticos que regularam eventos separados. Mais tarde, à medida que as ontogenias mais complexas evoluíram, módulos mais individualizados podem ter surgido através de elementos de unidades integradas maiores em entidades modulares separadas, cada um sendo alvo independente de seleção. Como a plasticidade fenotípica está intimamente ligada às vias de desenvolvimento (e às interações modulares), temos que diferentes genótipos podem convergir em suas vias de desenvolvimento em fenótipos similares em resposta às mudanças ambientais – uma **norma de reação**. É interessante analisar tal fato, pois organismos evolutivamente distantes podem convergir em estruturas ou planos corporais através da modulação de suas vias de desenvolvimento.

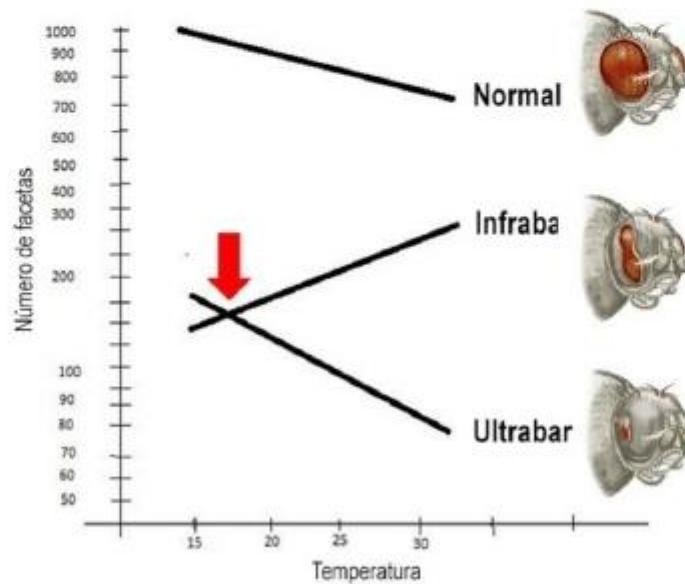


Figura 8. A norma de reação fica aparente quando diferentes genótipos, em resposta às variações ambientais, acabam por apresentar fenótipos semelhantes. Esse fenômeno pode ser observado quando as retas das moscas infrabar e ultrabar se cruzam numa determinada temperatura, o que significa que dois genótipos distintos podem resultar num mesmo fenótipo (número de facetas) dependendo do ambiente (temperatura).

Um exemplo clássico desse processo, onde genótipos diferentes originam fenótipos similares, pode ser observado através da exposição de diferentes mutantes de moscas-da-fruta (*Drosophila melanogaster*) a diferentes temperaturas. Os olhos desses organismos são compostos, sendo formados por diversas facetas. Existem dois tipos de mutantes que acabam por ter diferentes números de facetas em relação ao tipo-selvagem: ultrabar e infrabar. *Drosophila* ultrabar possui por volta de 200 facetas compondo os olhos, enquanto a infrabar possui ao redor de 100 facetas (o tipo-selvagem possui mais de 1000 facetas). Em resposta às mudanças de temperaturas, essas três linhagens acabam apresentando fenótipos diferentes. É nesse instante que surge um fenômeno interessante: nas moscas do tipo-selvagem, o número de facetas decresce quando a temperatura aumenta, o mesmo acontecendo para as moscas de genótipo ultrabar, que é ainda mais acentuado. Já as moscas infrabar respondem de maneira diferente, aumentando o número de facetas quando a temperatura aumenta. Nesse instante, há um momento, em uma determinada temperatura, em que o fenótipo das moscas infrabar e ultrabar acaba sendo o mesmo. **Temos dois genótipos distintos com um mesmo fenótipo.** Isso significa que uma variação do fenótipo ocorre em resposta às mudanças ambientais (plasticidade fenotípica), promovendo o aparecimento de fenocópias.

Outro exemplo muito interessante que compreende as questões de plasticidade fenotípica e acomodação genética é a evolução do dígito oposito (hálux – corresponde ao dedão das mãos ou pés nos humanos) nas aves, permitindo com que se segurem firmemente aos galhos ou na captura de presas.

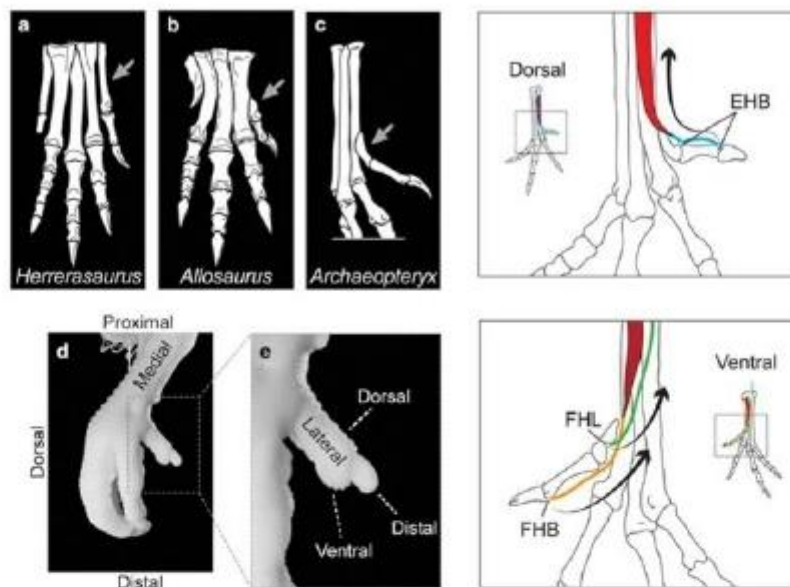


Figura 9. Principais transformações do *Mt1* na evolução do hallux oposto nas aves. Imagens de Botelho et al., 2015.

O hálux das aves evoluiu a partir do hálux não opositor dos primeiros dinossauros terópodes. Através da administração de fármacos que causa paralisia parcial nas atividades musculares em embriões de galinhas, Botelho e colaboradores (2015) verificaram que o hálux não retrovertia, apresentando uma fenocópia da morfologia das patas dos primeiros dinossauros tetanuranos (como os alossauros). O que se verifica é que as interações de diferentes módulos, como a musculatura embrionária, bem como o início de sua atividade, são necessárias para a torção do metatarso 1 e a retroversão do hálux. Além disso, a diferença no *timing* (heterocronia) na maturação e ossificação é crucial para que ocorra a torção do metatarso 1, logo, o atraso na maturação óssea acaba por fornecer uma maior plasticidade, facilitando a torção muscular. Tais experimentos demonstram que as interações ao nível dos sistemas de órgãos podem influenciar os rumos da

evolução dos organismos, através da mecânica do desenvolvimento e posterior seleção desse novo padrão ontogenético (Figura 9).

A Evo-Devo emerge como uma disciplina sistêmica, redefinindo as prioridades no campo da própria Biologia Evolutiva com um enfoque que abarca a complexidade e a integração entre os diversos níveis biológicos, desde os níveis físicos, onde o desenvolvimento pode ser visto em termos de sequências de DNA, ao nível celular e histológico, passando pelo nível sistêmico, com teorias que promovem a complementariedade do fenômeno ontogenético. Essa abordagem lança uma visão onde o desenvolvimento é tratado dentro de um contexto de intenso fluxo de informação entre os diversos níveis. De fato, o que observamos é que o fluxo de informação é o processo que torna viável a unificação desses diversos níveis (do molecular ao ambiental), proporcionando a chance de estabelecer equações e modelos para o fenômeno do desenvolvimento e de uma visão ampliada de hereditariedade.

Ao longo da história da Biologia Evolutiva, o desenvolvimento passou de “quintessência” ao ostracismo absoluto; passou de uma abordagem vitalista, proposta por Driesch (que defendia a ideia de que existiriam fatores não físicos direcionando a ontogenia), ao reducionismo radical de Roux (que sustentava a ideia de que o embrião nada mais seria uma máquina que seguia uma programação pré-estabelecida). Porém, através de Oskar Hertwig (1894), emergiu uma proposta capaz de contemplar essa relação multinível do desenvolvimento, a qual se denominou **organicismo absoluto**. Mais tarde, Paul Weiss (1926) e Ludwig von Bertalanffy (1932) formalizariam tal filosofia que embasa a ideia de que **as propriedades do conjunto não podem ser previstas apenas a partir das propriedades das partes de seus componentes**. Assim, temos que as propriedades das partes são estabelecidas pelo seu relacionamento com o todo (Drack et al., 2007).

A Biologia Evolutiva do Desenvolvimento, como nova disciplina, não trata de desmontar ou atomizar as partes de um sistema, mas propicia a integração em múltiplos níveis. Esse novo cenário que se abre exige que a Biologia (seus pesquisadores e filósofos) desenvolvam novas maneiras de pensar sobre o organicismo que temos hoje em dia, em contraponto aos programas reducionistas que perduraram por tantos anos. Talvez seja necessário pensarmos em uma mudança na filosofia das Ciências Biológicas, no sentido pleno do termo.

Referências

- ABZHANOV, A. et al. Bmp4 and Morphological Variation of Beaks in Darwin ' s Finches. *Science*, 2004. v. 305, September, p. 1462-1465.
- ABZHANOV, A. et al. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, 2006. v. 442, p. 563-567.
- AVEROF, M. Arthropod evolution: same Hox genes, different body plans. *Current Biology*, 1997.
- BALDWIN, J.M. *Development and evolution*. New York, NY: MacMillan. 1902.
- BATESON, W. Evolutionary faith and modern doubts. *Nature*, 1922.
- CALOW, P. *Evolutionary Principles*. Springer US, 1983.
- CARROLL, S. B. Evolution at two levels: On genes and form. *PLoS Biology*, 2005. v. 3, n. 7, p. 1159-1166.
- CARROLL, S.; WEATHERBEE, S.; LANGELAND, J. Homeotic genes and the regulation and evolution of insect wing number. *Nature*, 1995.
- CREWS, D.; GORE, A. C. Transgenerational Epigenetics. Current Controversies and Debates. *Transgenerational Epigenetics*, 2014. p. 371-390.
- DAHL, E.; KOSEKI, H.; BALLING, R. Pax genes and organogenesis. *Bioessays*, 1997.
- DARWIN, C. *On the Origin of Species by Means of Natural Selection*. 1969.

- DAVIDSON, E. *The regulatory genome: gene regulatory networks in development and evolution*. 2010.
- DE BEER, G.R. *Embryos and evolution*. Oxford: Clarendon; 1930.
- DESPLAN, C. Master control genes in development and evolution. *Science*, 1999.
- DOBZHANSKY, T. *Genetics and the origin of species*. [S.l.]: Columbia University Press, 1982.
- DRACK, M.; APFALTER, W. Is Paul A. Weiss' and Ludwig von Bertalanffy's System Thinking Still Valid Today? *Systems Research and Behavioral Science*, 2008. v. 8, n. 3, p. 27-42.
- GEHRING, W. The homeo box: a key to the understanding of development? *Cell*, 1985
- GERHART, J.; KIRSCHNER, M. *Cells, embryos, and evolution: Toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability*. Wiley, 1997.
- GILBERT, S.; LOREDO, G.; BRUKMAN, A. Morphogenesis of the turtle shell: the development of a novel structure in tetrapod evolution. *Evol Dev*, 3(2), p. 47-58, 2001.
- GOLDSCHMIDT, R. *The Material Basis of Evolution*. New Jersey: Pageant Books, 1960.
- GOLDSCHMIDT, R. A preliminary report on some genetic experiments concerning evolution. *The American Naturalist*, 1918.
- GOULD, S. *Ontogeny and phylogeny*. The Belknap Press of Harvard University Press. Cambridge, Massachusetts. London, 1977.
- GRENIER, J. et al. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Current Biology*, 1997.
- HAECKEL, E. *Generelle Morphologie der Organismen allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles*. 1866.
- HERTWIG, O. *Zeit-und Streitfragen der Biologie I. Präformation oder Epigenese Grundzüge einer Entwicklungstheorie der Organismen*. Gustav Fischer,

Jena. Translated as *The Biological Problem of Today: Preformation or Epigenesis?* (P. C. Mitchell, transl.). Macmillan, New York. 1894.

HUXLEY, J. *Evolution: The Modern Synthesis*. London George Allen & Unwin, 1942.

KAMMERMEIER, L. et al. Differential expression and function of the *Drosophila* Pax6 genes *eyeless* and *twin of eyeless* in embryonic central nervous system development. *Mech Dev*. 103(1-2), p. 71-8, 2001.

LANDE, R. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of evolutionary biology*, 2009.

LELIAERT, F.; VERBRUGGEN, H.; ZECHMAN, F. Into the deep: new discoveries at the base of the green plant phylogeny. *Bioessays*, 2011.

LEVINS, R. *Evolution in changing environments: some theoretical explorations*, 1968.

LEWIS, E. A gene complex controlling segmentation in *Drosophila*. *Nature*, 1978.

MACDONALD, M.; HALL, B. Altered timing of the extracellular-matrix-mediated epithelial-mesenchymal interaction that initiates mandibular skeletogenesis in three inbred strains of mice. *Journal of Experimental Zoology*, 2001.

MAIENSCHEIN, J. Stem cell research: a target article collection: Part II--what's in a name: embryos, clones, and stem cells. *The American journal of bioethics: AJOB*, v. 2, n. 1, p. 12-19, 2002.

MAYR, E. The establishment of evolutionary biology as a discrete biological discipline. *BioEssays*, 1997.

MAYR, E. *The growth of biological thought: Diversity, evolution, and inheritance*. Belknap Press, 1982.

MÜLLER, F. *Facts and arguments for Darwin*. London: John Murray, Albemarle Street. 1869.

PIGLIUCCI, M; MÜLLER, G. *Evolution - the extended synthesis*. Cambridge, Massachusetts. Massachusetts Institute of Technology. 2010.

PIGLIUCCI, M. *Phenotypic plasticity: beyond nature and nurture*. The Johns Hopkins University Press. Baltimore, Maryland. 2001.

- PIGLIUCCI, M.; MURREN, C. Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental*, 2006.
- PRUM, R.; BRUSH, A. The evolutionary origin and diversification of feathers. *The Quarterly review of biology*, 2002.
- RAFF, R. *The shape of life: genes, development, and the evolution of animal form*. University of Chicago Press, 2012.
- SCHLICHTING, C. The evolution of phenotypic plasticity in plants. *Annual review of ecology and systematics*, 1986.
- STEARNS, S. The evolutionary significance of phenotypic plasticity. *Bioscience*, 1989.
- SCHMALHAUSEN, I. *Factors of evolution: the theory of stabilizing selection*. 1949.
- STERN, D. L.; EMLEN, D. J. The developmental basis for allometry in insects. *Development*, v. 126, n. 6, p. 1091-1101, 1999.
- SUGA, H.; TSCHOPP, P.; GRAZIUSI, D. Flexibly deployed Pax genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish. *Proceedings of the national academy of sciences*, 2010.
- TILLS, O. et al. A genetic basis for intraspecific differences in developmental timing? *Evolution & Development*, 2011.
- VIA, S.; LANDE, R. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 1985.
- WADDINGTON, C. Organizers and Genes. *Price*, 1940.
- WAGNER, G. Homologues, natural kinds and the evolution of modularity. *American Zoologist*, 1996.
- WAGNER, G.; PAVLICEV, M.; CHEVERUD, J. The road to modularity. *Nature Reviews Genetics*, 2007.

SOBRE OS AUTORES E AUTORAS

André Luis Klein

Biólogo e doutor em Ciências (Genética e Biologia Molecular) pela Universidade Federal do Rio Grande do Sul. Possui experiência de pesquisa em Genética Ecológica, Etologia e Evolução, com ênfase em mimetismo, coloração protetiva e seleção sexual em lepidópteros e aracnídeos. Atualmente, é professor na Universidade Regional de Blumenau (FURB).

Andrea Paula dos Santos Oliveira Kamensky

Professora da Universidade Federal do ABC, pós-doutora em História da Ciência pelo Centro Simão Mathias da Pontifícia Universidade Católica de São Paulo. Doutora em História Econômica, mestre em História Social e graduada em História pela Universidade de São Paulo.

Camila Leão

Bióloga, graduada pela Universidade Federal do Rio Grande do Sul (2014) e aluna de mestrado em Ecologia na mesma universidade. Atualmente, trabalha com ecologia de populações de aves amazônicas de terra firme. *Site* do grupo de pesquisa: <http://ferrazlab.org/>.

Carolina de Souza Diegues

Bióloga e mestra em Ecologia pela Universidade Federal do Rio Grande do Sul (2012). Tem interesse em Ecologia, especialmente em Ecologia de Populações, com ênfase em Mastozoologia e Biologia da Conservação. Também se interessa por Educação, História e Filosofia da Ciência, especificamente àquelas relacionadas à Teoria da Evolução. Atualmente é professora de Ciências no Ensino Fundamental e de Biologia em curso preparatório pré-vestibular/ENEM, além disso, desenvolve pesquisa no campo de Genética Ecológica.

Claudio Ricardo Martins dos Reis

Biólogo, mestre em Ecologia e doutorando em Filosofia na Universidade Federal do Rio Grande do Sul (UFRGS). Publicou trabalhos em temas como Ecologia, Filosofia da Ciência, Filosofia e História da Biologia e Filosofia Política. Organizou e escreveu um capítulo para o livro “Flora da Bacia do Rio Pelotas: uso e conservação de espécies” (Rolim et al., 2016).

Participou de diversos projetos de extensão com temáticas socioambientais e foi ministrante nas diferentes edições do Curso de Biologia Evolutiva na UFRGS. Suas pesquisas atualmente focam em Filosofia, especialmente Filosofia da Ciência, incluindo a dimensão social do conhecimento científico e o papel apropriado dos valores na produção desse conhecimento.

Daniel Longo Rockenbach

Possui graduação em Ciências Biológicas, licenciatura pela Universidade Federal do Rio Grande do Sul (2010). É professor do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul. Biólogo formal e apaixonado, antropólogo informal e apaixonado, movido pela curiosidade e pela eterna transformação do ser ao seu redor.

Douglas Senna Engelke

Doutor em Neurociências pelo Programa de Pós-Graduação em Neurologia/Neurociências da Universidade Federal de São Paulo. Licenciado em Ciências Biológicas pela Universidade Federal do Rio Grande do Sul (UFRGS), em 2009, e técnico em Biotecnologia formado pela Escola Técnica da UFRGS, atual IFRS. Atua principalmente nos seguintes temas: Fisiologia, Psicobiologia, memória e aprendizagem, Neuropsicofarmacologia, dependência química, sistema endocanabinoide, manutenção da memória via PKMzeta, generalização da memória, Ayahuasca, Neuroinflamação, Neuroimunologia, Eletrofisiologia, Patch Clamp.

Fernando Bueno Ferreira Fonseca de Fraga

Licenciado e bacharel em Ciências Biológicas pela Universidade Federal do Rio Grande do Sul (UFRGS) e técnico em Biotecnologia pela mesma instituição. Atuou como técnico em Biologia na Faculdade de Agronomia da UFRGS e, atualmente, ocupa o cargo de docente do ensino básico técnico e tecnológico no Instituto Federal de Santa Catarina (IFSC). Desde 2008 desenvolve pesquisas e atividades nos campos de Microbiologia e divulgação das ciências.

Flavia Pacheco Alves de Souza

Doutoranda no programa de Evolução e Diversidade da Universidade Federal do ABC, mestra em Ensino, História e Filosofia das Ciências e Matemática pela Universidade Federal do ABC, especialista em educação ambiental pelo Centro Universitário Senac, Rio de Janeiro, e graduada em Ciências Biológicas pelo Instituto Federal de Educação, Ciência e Tecnologia de São Paulo.

Gilberto Cavalheiro Vieira

Possui graduação em Ciências Biológicas (licenciatura) e é mestre em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (UFRGS). Atualmente, cursa doutorado no Programa de Pós-Graduação de Genética e Biologia Molecular na UFRGS. Desenvolve trabalhos nas áreas de Evolução, Genética e Biologia Molecular de invertebrados, com ênfase em Epigenética. Atua, também, como professor de Biologia em cursos preparatórios pré-vestibular/ENEM.

Josmael Corso

Doutorado em Ciências e mestrado em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (UFRGS). Tem experiência em Biologia Evolutiva, com ênfase em evolução molecular da produção e percepção da coloração em vertebrados. Possui interesse em ensino de ciências, formação de professores da educação básica, divulgação e história da ciência. Atualmente, é professor substituto junto ao Projeto Amora do Colégio de Aplicação da UFRGS (CAp/UFRGS).

Leonardo Augusto Luvison Araújo

Graduado em Ciências Biológicas e mestre em Genética e Biologia Molecular na Universidade Federal do Rio Grande do Sul (UFRGS). Foi professor substituto de Biologia e Ciências no Colégio de Aplicação da UFRGS, entre 2015-2016. Atualmente, é doutorando em Educação e graduando em Filosofia na UFRGS. Desenvolve pesquisas sobre História e Filosofia da Biologia, Evolução Biológica e Ensino de Evolução.

Lucas Henriques Viscardi

Possui graduação em Arqueologia (bacharelado) pela Universidade Federal de Rio Grande e é mestre em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Sul (UFRGS). Atualmente, cursa doutorado no Programa de Pós-Graduação de Genética e Biologia Molecular na UFRGS. Desenvolve trabalhos nas áreas de Arqueologia e Genética, com ênfase em temas envolvendo evolução do comportamento humano e de outros primatas, considerando tanto aspectos biológicos como culturais.

Marco Antonio Corrêa Varela

Doutor em Psicologia Experimental. Tem experiência em Etologia, Psicologia Experimental e Biologia Evolutiva, com ênfase em Psicologia Evolucionista, Musicologia Evolucionista e Estética Evolucionista, abordando: mal-entendidos e o ensino de Evolução Humana, seleção sexual, escolha de parceiros amorosos, estratégias sexuais, diferenças individuais, paleoarte e evolução das propensões artísticas. Possui, no ScienceBlogs Brasil, o *blog* de divulgação científica Marco Evolutivo – Evolução e Comportamento Humano. Foi repórter de Ciência e Saúde freelancer para empresas como Folha de S.Paulo e Natura.

Marcel Baêta Lacerda Santos

Bacharel e licenciado em Ciências Biológicas pelo Centro Universitário Positivo (2007). Mestre e doutor pelo Programa de Pós-Graduação em Geociências pela Universidade Federal do Rio Grande do Sul, na área de concentração de Paleontologia. Tem interesse em Paleontologia e Evolução, com escopo de pesquisa voltado para anatomia comparada, taxonomia, bioestratigrafia, evolução e paleoecologia de vertebrados fósseis. Apresenta também treinamento e experiência didática em ilustração científica, com ênfase em trabalhos para artigos e materiais didáticos.

Miguel Trefaut Urbano Rodrigues.

Graduou-se em Ciências Biológicas pela Université de Paris VII (1978), obtendo seu doutorado em Zoologia (1984) e sua livre-docência (1995) pelo Instituto de Biociências da Universidade de São Paulo. É professor titular do Instituto de Biociências da Universidade de São Paulo desde

1996. É Research Associate do American Museum of Natural History, membro titular da Academia Brasileira de Ciências e da Academia de Ciências do Estado de São Paulo. Tem experiência na área de Herpetologia, com ênfase em Sistemática, Zoogeografia e Evolução.

Ronaldo Antonio Paesi

Licenciado e Bacharel em Ciências Biológicas e mestre em Biologia Animal pela Universidade Federal do Rio Grande do Sul (UFRGS). É professor de Biologia no Estado do Rio Grande do Sul.

Russel Teresinha Dutra da Rosa

Licenciada em Ciências Biológicas (1989), mestre em Educação (1996) e doutora em Educação pela Universidade Federal do Rio Grande do Sul (UFRGS) (2007). Atualmente, é professora associada da UFRGS. Tem experiência na área de Educação, atuando principalmente nos seguintes temas: formação de professores, ensino de ciências biológicas e divulgação científica.

Tomaz Panceri Melo

Graduado em Ciências Biológicas na Universidade Federal do Rio Grande do Sul (UFRGS) e doutorando no Laboratório de Paleovertebrados da mesma instituição. Tem experiência na área de Paleontologia de Vertebrados, com interesse na microestrutura óssea, ecologia, evolução, sistemática e biogeografia de amniotas fósseis, em especial pararrépteis e cinodontes não-mammaliaformes.

Voltaire Dutra Paes Neto

Doutorando pelo Programa de Pós-Graduação em Geociências do Instituto de Geociências da Universidade Federal do Rio Grande do Sul (IGEO/UFRGS). Mestre em Ciências pelo IGEO/UFRGS e graduado em Ciências Biológicas pela UFRGS. Tem interesse em paleobiologia, sistemática e evolução de arcossaumorfos, além de tafonomia, anatomia comparada de vertebrados e icnofósseis. Atua também como ilustrador científico e paleoartista.

AGRADECIMENTOS

A concretização deste projeto só foi possível devido ao apoio, a colaboração e a confiança depositada das diversas pessoas que contribuíram ao longo de todo o processo de maturação do presente livro. É de essencial importância agradecer a cada um que colaborou pessoalmente e/ou através das contribuições do projeto via financiamento coletivo. Também agradecemos o apoio institucional do Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS e a cada um dos participantes do Curso de Biologia Evolutiva, que desde 2014 nos brindam com a audiência e com a troca de ideias e saberes. Sem o apoio de cada um de vocês, nada disso seria possível.

Dessa maneira, fazemos questão de deixar aqui nosso especial agradecimento para as seguintes pessoas:

Adriana Ludwig	Dimitri Reynier Dib Ferreira
Agustin Martinelli	Edenilson de Sousa
Alcemar Martello	Eduardo Bayon Britz
Aldo Mellender de Araújo	Eduardo Londero
Alex Rozenquanz Schutz	Edvaldo Lima Coimbra
Alice Hirschmann	Emerson Salles Neves
Aline de Godoy	Fabiana Schneck
Ana Carolina Biacchi Brust	Fabiano Carvalho Brito
Ana Emilia Quezado de Figueiredo	Fábio Alves Rodrigues
Ana Paula Moraes Goetz	Fabricio Cappelari
Andre Bianco	Felipe Zilio
André Luis Klein	Gabriel Perotoni
Andrea Marrero	Gilberto Cavalheiro Vieira
Bianca Villavicencio	Guga Eidt
Caetano Filho	Gustavo Medina Tavares
Caio Cerqueira	Gustavo Scaron
Carolina Diegues	Helida Cunha
Clarice Sumi Kawasaki	Helier Balbinotti da Silveira
Claudio Reis	Ignacio Benites Moreno
Clediana Amaral Matzembaker	Igor Stelmach Pessi
Cristiano Feldens Schwertner	Ismael Verrastro Brack
Daniel Angelo Sganzerla Graichen	Ivano Casagrande Jr.
Diego Hepp	Ivo Pedro Klein

Jair Moisés de Sousa
Jean Lucas Poppe
Jeferson Jacob Arenzon
João Carlos Coimbra
Jonathas Jackson
Josiane Alves da Silva e Silva
Julia Lima
Juliana Cordeiro
Juliana Schmidt da Silva
Juliano Vaz de Abreu
Katia Janaina Zanini
Kétini Mafalda Sacon Baccin
Kim Barão
Lara Cristiani Rocha
Laura Verrastro
Leonardo Bergamo
Letícia Rech Bolzan
Letícia Ruiz Sueiro
Luana Vasconcellos
Lucas Sulzbach Rilho
Luiz Alberto Cobucci
Luiz Alfredo Jung
Luiz Fernando Biazus
Maico Fiedler
Manoel Ronaldo Carvalho Paiva
Marcelo Krause
Marco Andrei Klein
Marina Bento Soares
Mateus Santos de Souza
Mauricio Hoffmann
Maurício Tavares
Moisés Alves Lopes Filho
Nelio Marco Vincenzo Bizzo
Nelson Casagrande Junior
Patricia Kunt
Paulo Jota Pacheco
Priscila Teixeira Losekann
Rafael Bisso Machado
Ricardo Nery
Russel Dutra da Rosa
Sabrina de Azevedo Reschke

Samir de Deus Elian Andrade
Sandra Müller
Soeli Drecksler Klein
Thais Castro
Tuane L. Carvalho
Vanessa Jorge Gasparini
Victor Hugo Valiati
William Tavares



Fly



ISSN: 1933-6934 (Print) 1933-6942 (Online) Journal homepage: <https://www.tandfonline.com/loi/kfly20>


Combining morphology and molecular data to improve *Drosophila paulistorum* (Diptera, Drosophilidae) taxonomic status

Rebeca Zanini, Mário Josias Müller, Gilberto Cavalheiro Vieira, Víctor Hugo Valiati, Maríndia Deprá & Vera Lúcia da Silva Valente

To cite this article: Rebeca Zanini, Mário Josias Müller, Gilberto Cavalheiro Vieira, Víctor Hugo Valiati, Maríndia Deprá & Vera Lúcia da Silva Valente (2018) Combining morphology and molecular data to improve *Drosophila paulistorum* (Diptera, Drosophilidae) taxonomic status, *Fly*, 12:2, 81-94, DOI: [10.1080/19336934.2018.1429859](https://doi.org/10.1080/19336934.2018.1429859)

To link to this article: <https://doi.org/10.1080/19336934.2018.1429859>

 View supplementary material [↗](#)

 Accepted author version posted online: 22 Jan 2018.
Published online: 07 Feb 2018.

 Submit your article to this journal [↗](#)

 Article views: 108

 View Crossmark data [↗](#)

Full Terms & Conditions of access and use can be found at
<https://www.tandfonline.com/action/journalInformation?journalCode=kfly20>

RESEARCH PAPER



Combining morphology and molecular data to improve *Drosophila paulistorum* (Diptera, Drosophilidae) taxonomic status

Rebeca Zanini^{a,b}, Mário Josias Müller^c, Gilberto Cavalheiro Vieira^b, Victor Hugo Valiati^c, Maríndia Deprá^{a,b} and Vera Lúcia da Silva Valente^{a,b}

^aDepartamento de Zoologia, Programa de Pós-Graduação em Biologia Animal (PPGBAN), Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil; ^bLaboratório de Drosophila, Departamento de Genética, Programa de Pós-Graduação em Genética e Biologia Molecular (PPGBM), Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil; ^cLaboratório de Biologia Molecular, Programa de Pós-Graduação em Biologia, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brazil

ABSTRACT

The *willistoni* species subgroup has been the subject of several studies since the latter half of the past century and is considered a Neotropical model for evolutionary studies, given the many levels of reproductive isolation and different evolutionary stages occurring within them. Here we present for the first time a phylogenetic reconstruction combining morphological characters and molecular data obtained from 8 gene fragments (*COI*, *COII*, *Cytb*, *Adh*, *Ddc*, *Hb*, *kl-3* and *per*). Some relationships were incongruent when comparing morphological and molecular data. Also, morphological data presented some unresolved polytomies, which could reflect the very recent divergence of the subgroup. The total evidence phylogenetic reconstruction presented well-supported relationships and summarized the results of all analyses. The diversification of the *willistoni* subgroup began about 7.3 Ma with the split of *D. insularis* while *D. paulistorum* complex has a much more recent diversification history, which began about 2.1 Ma and apparently has not completed the speciation process, since the average time to sister species separation is one million years, and some entities of the *D. paulistorum* complex diverge between 0.3 and 1 Ma. Based on the obtained data, we propose the categorization of the former “semispecies” of *D. paulistorum* as a subspecies and describe the subspecies *D. paulistorum amazonian*, *D. paulistorum andeanbrazilian*, *D. paulistorum centroamerican*, *D. paulistorum interior*, *D. paulistorum orinocan* and *D. paulistorum transitional*.

ARTICLE HISTORY

Received 10 October 2017
Revised 7 December 2017
Accepted 12 January 2018

KEYWORDS

divergence time estimation;
D. paulistorum; morphology;
phylogeny; subspecies; total
evidence; *willistoni* subgroup

Introduction

The *Drosophila paulistorum* species complex belongs to the *willistoni* subgroup, a Neotropical representative of the *Sophophora* subgenus. The *willistoni* subgroup also includes other five siblings – *D. willistoni* Sturtevant 1916, *D. equinoxialis* Dobzhansky 1946, *D. tropicalis* Burla and Da Cunha 1949, *D. insularis* Dobzhansky *et al.* 1957 and *D. pavlovskiana* Kastritsis and Dobzhansky 1967. Some species of the *willistoni* subgroup present differentiation at the subspecific level: *D. tropicalis tropicalis* and *D. t. cubana* [1], *D. willistoni willistoni* [2], *D. w. quechua* [2], *D. w. winge* [3], *D. equinoxialis equinoxialis* and *D. e. caribbensis* [4].

Drosophila paulistorum has traditionally been defined as a “superspecies” [5] which is composed of six “semispecies”: Amazonian, Andean-Brazilian,

Centroamerican, Interior, Orinocan and Transitional [6]. The identity of these taxa was defined based on crossing tests. Dobzhansky and Spassky [6] crossed several lineages of *D. paulistorum* from different localities and observed that some of the crossings did not produce viable or fertile progeny and then divided *D. paulistorum* into six distinct geographic races – Amazonian, Andean-Brazilian, Centroamerican, Guyanan, Orinocan and Transitional. Later, the Guyanan race was upgraded to *D. pavlovskiana* species while a new Interior “semispecies” was posteriorly described by Pérez-Salas *et al.* [7]. The six entities which compose *D. paulistorum* live in sympatry at some level (review in Zanini *et al.* [8]).

The relationship between *D. paulistorum* “semispecies” varies from complete reproductive isolation

CONTACT Rebeca Zanini  rebeca.zanini@gmail.com; Vera Lúcia da Silva Valente  vera.valente@pq.cnpq.br  Av. Bento Gonçalves, 9500 – Prédio 43312 M, Zipcode: 91501–970 P. O. box: 15053, Porto Alegre, RS Brazil.
Supplemental data for this article can be accessed on the  publisher's website.

© 2018 Informa UK Limited, trading as Taylor & Francis Group

without hybrid progeny (crossings between Andean-Brazilian and Interior; Amazonian and Orinocan; Amazonian and Transitional), hybrids who died in eggs or pre-adults stages, male sterility only and fertile F_1 . As previously observed by Dobzhansky e Spassky [6], the Transitional race is the only one to produce fertile descendents when crossed with Andean-Brazilian and Centroamerican [9].

The *willistoni* subgroup has been a model for evolutionary studies since the latter half of the past century. This subgroup is a perfect model for evolutionary studies, since it encompasses species in various taxonomic levels and in distinct evolutionary stages, characterized by the different degrees of reproductive isolation. Some attempts to establish the evolutionary relationships within the *willistoni* subgroup were made, with Spassky et al. [10]. providing a schematic diagram representing the evolutionary relationships based on cytological, biogeographical, genetic and biochemical data. Ayala et al. [4]. constructed a dendrogram based on the genetic differentiation of the species at 36 allozymes loci. Gleason et al. [11]. presented the first DNA-based phylogeny of the *willistoni* subgroup, using three marker genes – *Adh* (Alcohol Dehydrogenase), *per* (Period) and *COI* (Cytochrome c Oxidase subunit I), and included all siblings and semi-species as well as some species of *D. bocainensis* subgroup as an outgroup. Later, Tarrío et al. [12]. used the *Xdh* (Xanthine Dehydrogenase) gene fragment to clarify some aspects of the evolutionary relationships of the *D. willistoni* group. O’Grady and Kidwell [13] proposed a phylogeny of the *Sophophora* subgenus and included some species of the *D. willistoni* group. In a recent study, Robe et al [14]. presented a phylogenetic analysis including two mitochondrial gene fragments – *COI* and *COII* (Cytochrome c Oxidase subunit II) and four nuclear genes – *Adh*, *Amd* (Alpha methyl dopa), *Hb* (Hunchback) and *per*.

Despite all these studies, some questions still remained, mainly those concerning the *D. paulistorum* complex. In the preceding studies, some relationships were incongruent between data partitions and barely supported. Additionally, despite morphological studies [15-21], there is no phylogenetic reconstruction of the *willistoni* subgroup based on morphological characters or a strong time estimation of the cladistic events. Although past studies tried to differentiate *D. paulistorum* semispecies, some authors believe that there are no sufficient differences in the

morphological, physiological and ecological traits [19,22]. The taxonomic status of *D. paulistorum* is also a major problem, since “semispecies” and “super-species” are not valid taxonomic categories, according to the International Code of Zoological Nomenclature.

Drosophila willistoni species subgroup is inserted in a complex evolutionary scenario where neither speciation or diversification of species was completely clarified. This background makes *D. willistoni* subgroup a suitable model for evolutionary and biogeographical studies. Accordingly, in the present study we attempt to contribute to a better understanding of the evolutionary history of the *willistoni* species subgroup, with emphasis on the *D. paulistorum* species complex. The main goal of this study is to present a phylogenetic hypothesis for *D. willistoni* subgroup, employing for the first time a total evidence dataset (morphological and molecular characters) and discuss the evolutionary events associated with this subgroup, especially concerning the *D. paulistorum* species complex and, based on the results, suggest a new categorization for the *D. paulistorum*.

Results

The genetic distances were obtained for each of the eight molecular data partitions and combined datasets (Figure 1, Supplementary material 1). A low nucleotidic divergence was observed among the species of the *willistoni* subgroup, especially among the species of *D. paulistorum* cluster. In addition, there is haplotype sharing among some members of the subgroup: *COI* (*D. paulistorum* Centroamerican and Transitional); *COII* (*D. paulistorum* Centroamerican and Transitional; *D. paulistorum* Interior and Orinocan); *Cytb* (*D. paulistorum* Centroamerican and Transitional); *Adh* (*D. paulistorum* Orinocan and Transitional); *Ddc* (*D. paulistorum* Interior and Orinocan); *Hb* (*D. paulistorum* Amazonian and Andean-Brazilian; *D. equinoxialis* and *D. insularis*); *kl-3* (*D. paulistorum* Interior, Andean-Brazilian and Orinocan; *D. paulistorum* Amazonian and Transitional).

The mitochondrial genes *COI* and *Cytb* presented the highest divergence levels among species, although some species have shared haplotypes. The chi-square test did not find any evidence of base composition heterogeneity between taxa, despite the significant nucleotide frequency bias in some partitions, especially in the 3rd codon position and a deviation in the

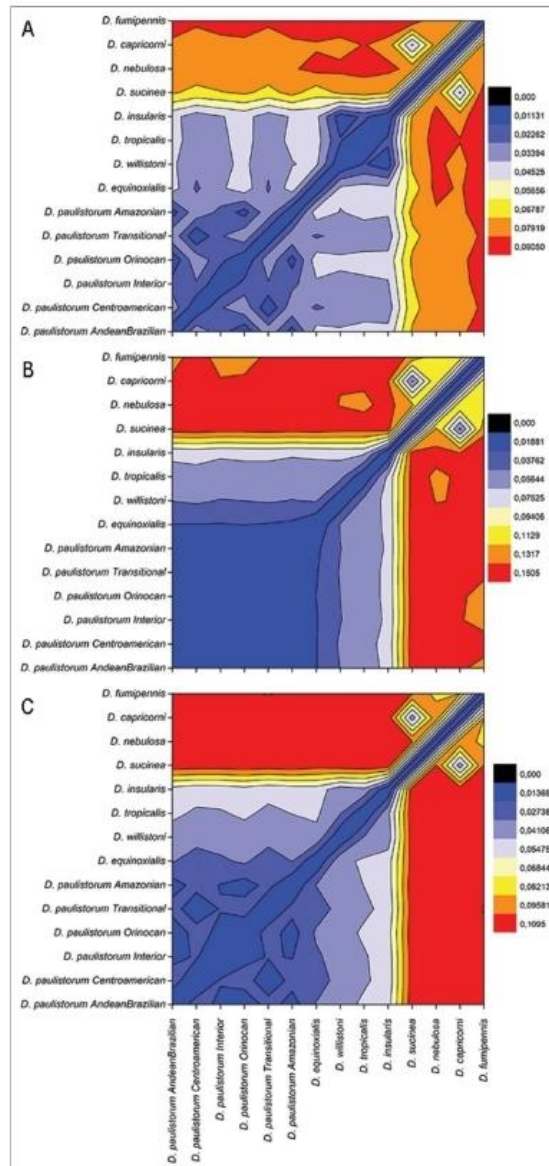


Figure 1. Graphic representation of evolutionary divergence between sequences using Maximum Composite Likelihood model. A. Mitochondrial dataset; B. Nuclear dataset; C. Molecular total evidence dataset.

transitions versus transversions ratio (Supplementary material 2). All matrices exhibited significant levels of heterogeneity in substitution rates among sites, which were adjusted through implementation of continuous gamma distribution (G) and/or proportion of variable sites (I) (Supplementary material 2).

Regarding the phylogenetic reconstructions, the topologies resulting from morphological and molecular datasets are partially incongruent, but *D. paulistorum* is recovered as monophyletic in both approaches (Figures 2 and 3). The morphological reconstruction result in a tree that recovered the three major clades (Figure 2). The *bocainensis* subgroup branched first and recovered two clades – one composed of *D. sucinea* and *D. capricorni* and the other composed of *D. nebulosa* and *D. fumipennis*. The sibling species were the next to branch out, in a polytomic clade formed by *D. insularis*, *D. equinoxialis*, *D. willistoni* and *D. tropicalis*, with the latter two grouped as sister-species. The *D. paulistorum* cluster was recovered with a polytomy between Amazonian and Transitional and a possible common ancestor of the remaining semispecies, which branched out in two distinct clades: Interior and Andean-Brazilian, and

Orinocan and Centroamerican. However, all these relationships were poorly supported.

The Bayesian Inference and Maximum Likelihood phylogenies (Figure 4 and Supplementary material 3, respectively) obtained from the total evidence matrix resulted in a very similar topology, with minor discrepancies in the position of *D. paulistorum* Centroamerican and *D. paulistorum* Interior. Nodes were strongly supported by posterior probability and bootstrap.

The total evidence tree (Figure 4) presented well-supported relationships without unresolved polytomies and summarized the results of all analysis (Figures 2 and 3, Supplementary material 3). The *bocainensis* subgroup is ancestral to *willistoni* subgroup and divided in two clades – *D. sucinea* and *D. capricorni*, and *D. nebulosa* and *D. fumipennis* (Figure 4). Regarding *willistoni* subgroup, the divergence sequence is as follows: *D. insularis*, *D. tropicalis*, *D. willistoni* and *D. equinoxialis*, followed by *D. paulistorum* Transitional, Centroamerican, Interior, Amazonian, and the sister-species Orinocan and Andean-Brazilian. Concerning the *D. paulistorum* cluster, only the split between *D. paulistorum* Interior and remaining species showed a low support value in

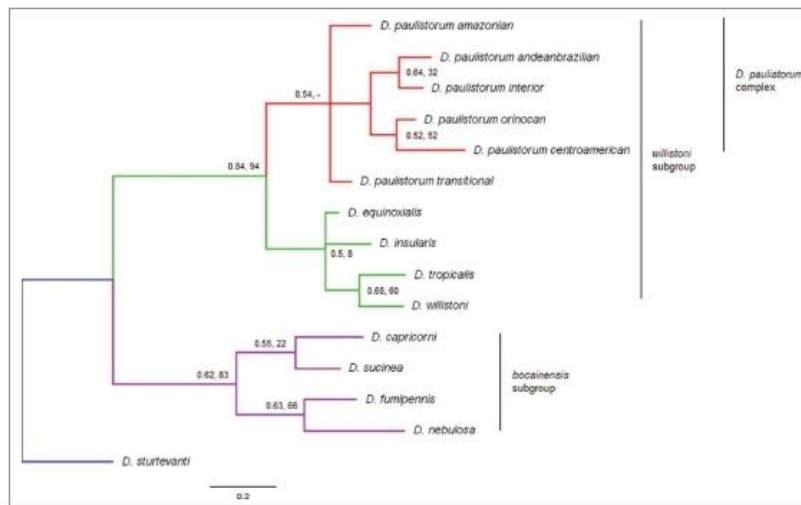


Figure 2. Bayesian Inference tree obtained with morphological dataset. The numbers in the nodes are Bayesian Posterior Probabilities (PP) and maximum likelihood bootstrap support (BS). (-) indicates that there is no support for the node.

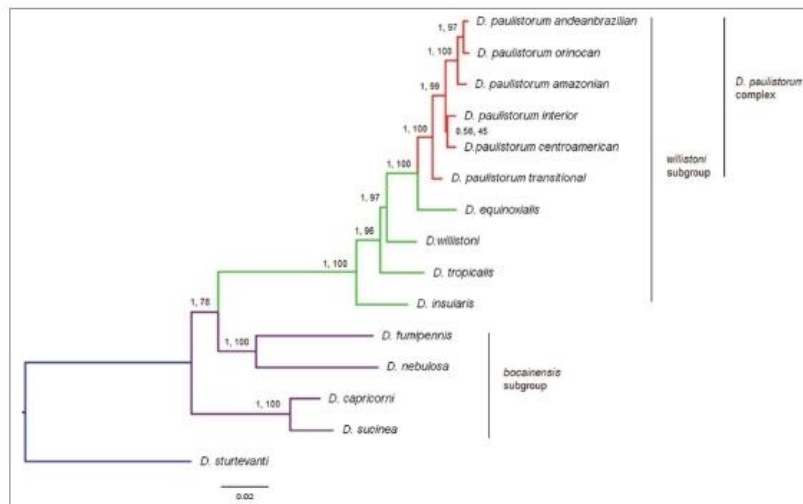


Figure 3. Bayesian Inference tree obtained with the molecular dataset. The numbers in the nodes are Bayesian Posterior Probabilities (PP) and maximum likelihood bootstrap support (BS). (-) indicates that there is no support for the node.

the Bayesian Inference ($PP < 0.6$) and is not supported in Maximum Likelihood (Supplementary material 3).

The diversification in the *willistoni* subgroup began around 7.3 Ma, with the branching of *D. insularis*. *D. tropicalis* branched out around 5.5 Ma; *D. willistoni* around 4.8 Ma; *D. equinoxialis* branched out at approximately 3.1 Ma and the *D. paulistorum* complex began their divergence process around 2.1 Ma, with the *D. paulistorum* Transitional drift. The split between *D. paulistorum* Interior and *D. paulistorum*

Centroamerican occurred around 1.1 Ma, followed by *D. paulistorum* Amazonian around 0.7 Ma. The latest split occurred around 0.33 Ma, between *D. paulistorum* Orinocan e *D. paulistorum* Andean-Brazilian (Figure 5).

Discussion

The entire *willistoni* group, including non-sibling species, is in distinct evolutionary levels. The *bocainensis*

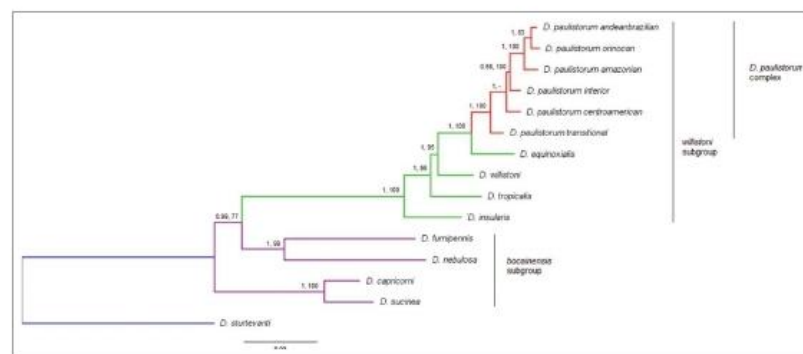


Figure 4. Bayesian Inference tree based on total evidence dataset – morphological characters, nuclear genes dataset (*Adh*, *Kl-3*, *Hb*, *Ddc*, *per*) and mitochondrial genes dataset (*COI*, *COII*, *Cytb*). The numbers in the nodes are Bayesian Posterior Probabilities (PP) and maximum likelihood bootstrap support (BS). (-) indicates that there is no support for the node.

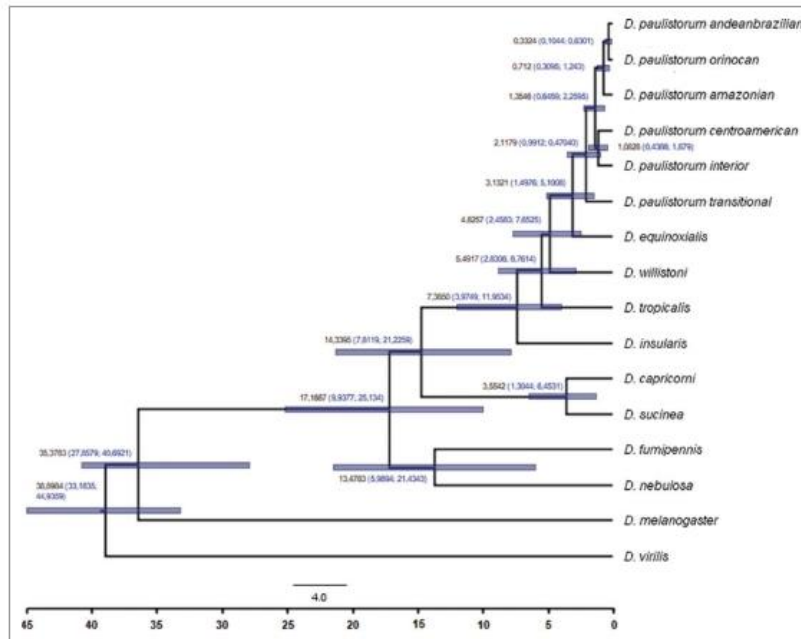


Figure 5. Divergence time estimation with two calibration points based on total evidence data partition. The values above the branches are the ages of the nodes. The 95% HPD interval is represented by the blue bars.

subgroup, presents high divergence levels between species and the relationships were consistently recovered in the different approaches. Considering just the species analyzed in this study, *bocainensis* subgroup seems to have had a polyphyletic origin. *D. sucinea* and *D. capricorni* seems closely related, since this relationship was recovered using different gene markers and morphology. This relationship corroborates O'Grady and Kidwell [13] but, nevertheless, more members of *bocainensis* subgroup must be included to accurately estimate evolutionary relationships within this subgroup.

Some relationships found in our study were previously described. We found that *D. insularis* was the first sibling to diverge from a common ancestor. This species has been considered basal in relation to the remaining subgroup. Spassky et al. [18], considered *D. insularis* a probable insular offshoot of the continental *D. tropicalis*. Ayala et al. [4], considered *D. insularis* as the oldest lineage of the *willistoni* subgroup. Gleason et al. [11], also found that *D. insularis* seems to be the first species of this subgroup to diverge, with the same

hypothesis also presented by Robe et al. [14]. Still, Rhode et al. [23], established that *D. insularis* II-R chromosomes are the most rearranged, corroborating the molecular findings. The sequence in which *D. tropicalis*, *D. willistoni*, *D. equinoxialis* and *D. paulistorum* split was previously found by Gleason et al. [11], with *per* and *Adh* genes and by Robe et al. [14], with *Adh*, *Amd* and *per*.

Considering the relatively recent divergence time of the *willistoni* subgroup, the low levels of morphological difference are justified. The most evident morphological differences between the *willistoni* subgroup siblings are in the male genitalia, which can be explained by the distinct evolutionary rate of genital characters or morphological characters in general. Masly and Presgraves [24], stated that the characters of the external genitalia evolve faster than other morphological character and the rapid divergence of the male genitalia in insects is so striking that recently diverged species can exhibit pronounced differences in the male genitalia [25–27]. Richmond et al. [28], resuming McPeck et al. [29,30], claims that sexual

characteristics go through substantial shifts during speciation events, due to alternating selective pressures coming from directional and stabilizing selection on mate recognition systems. According to the authors, a corresponding shift in female genital morphology is not necessary, since those are not necessarily impaired by structural variations that give copulatory advantages to males. The lack of definition in the phylogenetic relationships obtained through morphological characters analysis can be a reflection of the recent events of speciation. It is possible that these characters did not accumulate enough differences for a strong phylogenetic signal, although the differences are enough to identify sibling species and even the *D. paulistorum* semispecies [21].

Diversification of the *willistoni* subgroup, according to our estimation, began around 7.3 (3.9749 – 11.9534) Ma, during the upper Miocene. During this period, there were changes in the landscape of Central and South America, like marine incursions forming wetlands in western Amazonia [31] and the closure of the Isthmus of Panama [32]. *D. paulistorum* complex diversification began around 2.1 (0.9912 – 3.4704) Ma, during the Plio-Pleistocene with the split of *D. paulistorum* Transitional. The remaining taxa differentiated during the Pleistocene, characterized by global climatic oscillations and expansion-contraction of the South American biomes [33]. Based on our results, it is suggested that the diversification in the *willistoni* subgroup and in the *D. paulistorum* complex began a little earlier than suggested by Robe et al. [14]. These authors put the beginning of *D. willistoni* diversification around 5.7 Ma and the *D. paulistorum* complex around 0.9 Ma. Even though that estimation was performed based only on an *Adh* gene fragment, the values fit our 95% confidence interval.

We hypothesize that the origin and further diversification of the *willistoni* subgroup took place in the North region of South America, at the Amazonian area. Most of the species of the subgroup live in sympatry in this area, except for *D. insularis*, which occurs in the Antilles and *D. paulistorum* Centroamerican, in Central America [8,10]. The geomorphological events during the Miocene, especially the closure of the Isthmus of Panamá, allowed ancestral lineages from South America to expand its distribution to Central America. This pattern is also observed in the *nannoptera* group [34] and in *fasciola* group [35], although in the latter the diversification began in Central America and

expanded to South America. The geographic distribution of the *willistoni* group corroborates our hypothesis of diversification. Moreover, Haffer [36] proposed some hypotheses to explain the vertebrate speciation in Amazonia and this explanation could be extrapolated to the *willistoni* subgroup speciation process, as already suggest by Spassky et al. [10]. and reviewed in Ehrmann and Powell [9].

Haffer [36,37] theorized a historical explanation to geographical distribution patterns of some closely related bird species in Tropical America. According to Spassky et al. [10], there is a resemblance between those bird species and *D. paulistorum* semispecies. During dry periods of the Pleistocene, the Amazonian forest was divided into smaller patches of forest, surrounded by areas of non-forest open vegetation [36,37]. These forest areas acted as “refuge areas” for taxa dependent on the humidity, promoting a rapid differentiation of the Amazonian fauna in a very recent geological time [36,37]. Spassky et al. [10]. related *D. paulistorum* semispecies distribution with known Haffer’s refuges. Transitional semispecies occur alone in the Chocó refuge and in the mountains near the Caribbean coast of Colombia and Venezuela. Centroamerican semispecies distribution coincides with Haffer’s Centroamerican refuge. Orinocan and Amazonian semispecies may be derived from Catatumbo, Imeri, Guyana or Belém refuges. Interior distribution corresponds to Napo or Imori refuges. Andean-Brazilian may have expanded from the East Peruvian refuge [10]. Corroborating this hypothesis, the diversification of *Drosophila fasciola* subgroup also began during the Miocene, and part of this subgroup divergence began in the Plio-Pleistocene and Pleistocene [35], very similar to the *D. willistoni* subgroup diversification.

Speciation and subspeciation in the *willistoni* subgroup have been controversial issues. Several research groups have disagreed on some aspects since Theodosius Dobzhansky choose this subgroup as a model for evolutionary studies. Thus, the results obtained in our study lead us to the question: Is *D. paulistorum* a single species?

Based on the male genitalia morphology [21] and molecular analyses, in addition to previous chromosomal studies [23], it could be proposed that the former “semispecies” are, in fact six distinct species, supporting the elevation of the “semispecies” Guyanensis to *D. pavlovskiana*, which was based only in

chromosomal studies and crossing tests [38]. However, even if the six components of *D. paulistorum* complex could be considered species, they are probably still in the process of reproductive isolation, and for this reason, we suggest a more conservative approach. Still regarding the classification of *D. paulistorum* as a “superspecies” composed of “semispecies”, Mayr [39] claimed that “semispecies of a superspecies should be mostly allopatric”. This is not true for the *D. paulistorum* complex, since most of them live in sympatry in the North of South America and in the Amazonia area [8,10].

Regarding the time necessary to complete the speciation process, about 1 million years [40] is suggested, which leads us to believe that some entities of the *D. paulistorum* complex still did not complete the process. These authors highlight that this is the required time for divergence between sister-species, which is not the same time required for the complete reproductive isolation, which could be more than 4 million years. This could explain the interspecific crossings between *Drosophila* species. Thus, we suggest the former semispecies to be classified as subspecies of *D. paulistorum*, putting them into a valid taxonomic category according to the International Code of Zoological Nomenclature [41].

Drosophila paulistorum general description can be found in Burla et al. [15]. We are describing here the subspecies of *D. paulistorum*:

D. paulistorum amazonian

Description

Arista with 8–9 branches; light ocellar triangle; tan first flagellomere, with long and sparse pilosity; prominent carina; 2 oral bristles. Tan mesonotum and scutellum. Yellow pleura and legs. Eight irregular rows of acrostichal hairs. Convergent scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 15 preniseta, nine longer and six shorter and one setae in the ventral hook. Square shaped hypandrium, dome-shaped lobes with a slightly convergent seta. The lobes almost touch each other. Large and slightly separated teeth, inserted in the lobe line. Lateral extensions prominent toward the top. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a small round expansion at the distal end.

Distribution

Panama, Trinidad, Colombia, Venezuela, Guyana and Northern Brazil.

D. paulistorum andeanbrazilian

Description

Arista with 7 branches; light ocellar triangle; tan first flagellomere, with short dense pilosity; short carina; 2–3 oral bristles. Yellow mesonotum and scutellum. Yellow pleura and legs. Seven irregular rows of acrostichal hairs. Convergent scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 15 preniseta, eight longer and seven shorter and two seta in the ventral hook. Square shaped hypandrium, slightly square shaped lobes, convergent paramedian seta. Very close, medium sized thin teeth that are twice the height of the lobes, inserted in the lobe line. Lateral extensions almost continuous with the lobes. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a small round expansion at the distal end.

Distribution

Brazil, Ecuador, Peru, Colombia and Venezuela.

D. paulistorum centroamerican

Description

Arista with 8 branches; dark ocellar triangle; dark first flagellomere, with short dense pilosity; prominent carina; 2 oral bristles. Tan mesonotum and scutellum. Yellow pleura and legs. Six regular rows of acrostichal hairs. Parallel scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 15 preniseta, eight longer and seven shorter and two seta in the ventral hook. Rectangular hypandrium, irregular shaped lobes with convergent seta. Very close medium sized thin teeth, twice the height of the lobes, inserted below the lobe line. Lateral extensions adjacent, very close to the lobes. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a fan-like expansion at the distal end.

Distribution

Tical (Colombia), Lancetilla (Honduras), San Salvador (El Salvador), Turrialba (Costa Rica) and Boquete, Almirante and the Central area of Panama.

D. paulistorum interior**Description**

Arista with 8 branches; light ocellar triangle; tan first flagellomere, with short dense pilosity; prominent carina; 2 oral bristles. Yellow mesonotum and scutellum. Yellow pleura and legs. Six regular rows of acrostichal hairs. Divergent scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 12 crescent sized preniseta and two seta in the ventral hook. Hypandrium like *D. paulistorum andeanbrazilian*, although the lateral extensions are not continuous with the lobes and there is no gap between teeth and lobes. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a small round expansion at the distal end.

Distribution

Colombia, Venezuela and Northern Brazil.

D. paulistorum orinocan**Description**

Arista with 8 branches; light ocellar triangle; yellow first flagellomere, with long dense pilosity; short carina; 2 oral bristles. Yellow mesonotum and scutellum. Yellow pleura and legs. Five regular rows of acrostichal hairs. Convergent scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 12 preniseta, approximately the same size and two seta in the ventral hook. Smallest hypandrium among *D. paulistorum* subspecies, square shaped with small round lobes, convergent paramedian seta, medium sized thin teeth, three times the size of the lobes, inserted a little below the lobe line. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a fan-like expansion at the distal end.

Distribution

Panama, Trinidad, Colombia, Venezuela and Guyana, mainly in the Caribbean coast.

D. paulistorum transitional**Description**

Arista with 8 branches; light ocellar triangle; yellow first flagellomere, with long dense pilosity; prominent carina; 2 oral bristles. Yellow mesonotum and scutellum. Yellow pleura and legs. Eight regular rows of acrostichal hairs. Convergent scutellar bristles. Hyaline wings. Yellow abdomen, with narrow continuous darker bands. Epandrium with 12 bristles. Surstylus with 13 preniseta, 12 of the same size and one larger in the middle and two seta in the ventral hook. Square shaped hypandrium with small round lobes very close to the teeth. Convergent paramedian seta and large thick teeth, almost three times the size of the lobes, inserted in the lobe line. Lateral extensions prominent toward the top, higher than the lobes. Beak-like distyphallus. Rod shaped aedeagal apodeme, with a small round expansion at the distal end.

Distribution

Santa Marta, Chocó Condoto and Valle, in Colombia; Perija, Vigía, Barinas, Sarare, Rancho Grande and Guatopo, in Venezuela.

Material and methods**Taxon sampling**

Seventeen species were included in the analyzed datasets (Table 1). The ingroup comprises 10 species: *D. willistoni*, *D. equinoxialis*, *D. tropicalis*, *D. insularis* and the six representatives of the *D. paulistorum* complex. The outgroup includes 4 species of the *bocainensis* subgroup and *D. sturtevantii* (*saltans* group). The obtained topologies were rooted with *D. sturtevantii*. For the divergence time estimation, we also included *D. melanogaster* (*Sophophora* subgenus) and *D. virilis* (*Drosophila* subgenus).

Morphological characters

The morphological dataset is composed of 45 characters (Supplementary material 4). The morphological characters were chosen based on previous studies with

Drosophila species [42,43]. For general morphology characters, 20 individuals of each species of *willistoni* and *bocainensis* subgroups were studied. We analyzed different strains of *D. willistoni*, *D. equinoxialis*, *D. insularis*, *D. paulistorum* Amazonian, *D. paulistorum* Andean-Brazilian and *D. nebulosa* from several localities (Supplementary material 5). Male genitalia characters of *willistoni* subgroup were obtained from Zanini et al. [21]. For the *bocainensis* subgroup, we observed at least 20 individuals of each species, using scanning electron microscopy and light microscopy preparations following Zanini et al. [21].

Morphological data were coded in Mesquite 3.31 [44] and the matrix was exported as a Nexus file. Inapplicable states were coded as “-”. Characters were treated as nonadditive. Morphological characters were treated under the Mkv evolutionary model [45], when using Maximum Likelihood probabilistic method.

DNA extraction and amplification

DNA was isolated following Sassi et al. [46]. Phylogenetic information was evaluated with sequences obtained from five nuclear loci – *Alcohol Dehydrogenase (Adh)*, *Dopa-Decarboxylase (Ddc)*, *Hunchback (Hb)*, *kl-3* (male fertility factors *kl-3* located in the short arm of the Y chromosome) and *Period (per)*. Additionally, three mitochondrial markers were employed – *Cytochrome oxidase subunit I (COI)*, *Cytochrome Oxidase subunit II (COII)*, *Cytochrome Oxidase b (Cytb)*. We generated new sequences for *COI*, *Cytb*, *Ddc*, *kl-3* for all species analyzed, and *Adh* and *COII* for some species. *Hb*, *per*, *Adh* (in part) and *COII* (in part) were downloaded from GenBank (Table 1).

The amplification conditions as well as the primers used to amplify the sequences of molecular markers are in Supplementary material 6. Following the PCR, amplicons were purified with Exonuclease I (10 U/ μ l) and Shrimp Alkaline Phosphatase (1 U/ μ l). MacroGen (Seoul, South Korea) carried out the sequencing, using BigDye technology, in both forward and reverse directions, using the same primers used in the amplification. All generated amplicons were deposited in GenBank (Table 1).

Sequences analyses

The electropherograms obtained were analyzed, assembled and the consensus sequences generated were edited and visualized using Staden Package [47].

Consensus sequences were aligned using the Clustal W algorithm, implemented in MEGA version 7.0 [48].

Alignments were concatenated in a single matrix using DNAsp v. 5.10.1 [49]. Morphological and molecular datasets were merged in Mesquite 3.31 [44] and exported in NEXUS (for MrBayes analyses) and Phylip (for RAxML analyses).

Patterns of nucleotide substitution for each locus and combined loci were evaluated using the evolutionary models suggested by the Akaike Information Criterion [50] performed in JmodelTest [51]. Descriptive parameters were obtained in MEGA 7.0 [48] and in DAMBE 5.3.109 [52], and can be visualized in Supplementary material 2.

Each loci were tested for substitution saturation in DAMBE 5.3.109 (Supplementary material 7).

Pairwise genetic divergences were estimated in MEGA 7.0, considering each loci and total evidence dataset using Maximum Composite Likelihood; variance was estimated with 1000 bootstrap replications (Figure 1, Supplementary material 1).

Phylogenetic analyses

Preliminary phylogenetic analyses were made for each molecular marker, and for different combined datasets – nuclear markers, mitochondrial markers, morphological and nuclear markers combined, morphological and mitochondrial markers combined, nuclear and mitochondrial markers combined (Supplementary material 3).

Gene trees and combined molecular trees were constructed using two methods: Bayesian Inference performed in MrBayes 3.1.2 [53] and Maximum Likelihood (ML) performed in RAxML v.8.0.0 [54].

For the Bayesian Inference with combined datasets, the matrix was partitioned into morphological data and into different gene sets, and a nucleotide substitution model was used for each partition. The Markov Chain Monte Carlo (MCMC) was run for at least 1,000,000 generations, sampling trees every 100 generations and using four chains with heating default values. The runs were stopped when the convergence values declined below 0.01 and applied a burn-in of 25% of the generations before summarizing trees and parameters. Node support reflected the posterior probabilities of each clade.

For Maximum Likelihood, the GTR-CAT model was used for the molecular markers and the Multi-

state-CAT+MK was used for the morphological dataset. The node support was based on 1000 bootstrap replications.

Divergence time estimation

Divergence time was estimated with BEAST 2 [55] with parameters defined in BEAUTi [55]. The MCMC chains were run independently three times for 20,000,000 generations, with sampling every 1000 generations, under a log-normal relaxed-clock model using a Yule process as the tree prior for speciation model and a random start tree, assuming the models GTR+I+G. The convergence was analyzed in Tracer [55]. Trees were summarized in Tree Annotator [55], with 25% burn-in and visualized and edited with Fig-Tree [56].

Two calibration points were established based on previously dated speciation events. The first calibration point was the split between *D. willistoni* and *D. melanogaster* groups (36 MA) and the second point was the split of subgenus *Drosophila* and *Sophophora* (40 MA), both based on Powell and De Salle [57] and Russo et al. [58]. Combined nuclear and mitochondrial datasets were employed for divergence time estimation analysis.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.


Acknowledgments

We thank to CNPq (process number 142645/2011-7) and FAPERGS (process number 10/0028-7) for financial support. We thank all the collectors for the provided strains. We also thank Guilherme Atencio and Rachael Shevin for the English review.

Funding

Conselho Nacional de Desenvolvimento Científico e Tecnológico (142645/2011-7). Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (10/0028-7).

ORCID

Rebeca Zanini  <http://orcid.org/0000-0001-5474-1574>

References

- [1] Townsend JI. Cryptic sub speciation in *Drosophila* belonging to the subgenus *Sophophora*. *Am Nat.* 1954;88:339–351. doi:10.1086/281842
- [2] Ayala FJ, Tracey ML. Enzyme variability in the *Drosophila willistoni* group. VIII. Genetic differentiation and reproductive isolation between two subspecies. *J Hered.* 1973;64:120–124. doi:10.1093/oxfordjournals.jhered.a108367
- [3] Mardiros XB, Park R, Clifton B, et al. Postmating Reproductive isolation between strains of *Drosophila willistoni*. *Fly (Austin)*. 2016;10(4):162–171. doi:10.1080/19336934.2016.1197448
- [4] Ayala FJ, Tracey M, Barr LG, et al. Genetic and reproductive differentiation of the subspecies *D. equinoxialis caribbensis*. *Evolution.* 1974;28:24–41. doi:10.1111/j.1558-5646.1974.tb00723.x
- [5] Dobzhansky T, Ehrman L, Pavlovsky O, et al. The super-species *D. paulistorum*. *Proc Nat Acad Sci USA.* 1964;51:3–9. doi:10.1073/pnas.51.1.3
- [6] Dobzhansky T, Spassky B. *Drosophila paulistorum* a cluster of species in statu nascendi. *Proc Nat Acad Sci USA.* 1959;45:419–428. doi:10.1073/pnas.45.3.419
- [7] Perez-Salas S, Richmond RC, Pavlovsky OA, et al. The interior semispecies of *Drosophila paulistorum*. *Evolution.* 1970;24:519–527. doi:10.1111/j.1558-5646.1970.tb01786.x
- [8] Zanini R, Deprá M, Valente VLS. On the geographic distribution of the *Drosophila willistoni* group (Diptera, Drosophilidae) – updated distribution of *willistoni* subgroup. *Dros Inf Serv.* 2015;98:39–43
- [9] Ehrmann L, Powell JR. *The Genetics and Biology of Drosophila Vol. 3b*. New York: Academic Press Inc; c1982. The *Drosophila willistoni* species group; p. 193–220
- [10] Spassky B, Richmond RC, Pérez-Salas S, et al. Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. *Evolution.* 1971;25:129–143
- [11] Gleason JM, Griffith EC, Powell JR. A molecular phylogeny of the *Drosophila willistoni* group: conflict between species concepts? *Evolution.* 1998;52:1093–1103. doi:10.1111/j.1558-5646.1998.tb01836.x
- [12] Tarrío R, Rodríguez-Trelles F, Ayala FJ. Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila saltans* and *willistoni* groups, a case study. *Mol Phylogenetics Evolution.* 2000;16:344–349. doi:10.1006/mpev.2000.0813
- [13] O'Grady PM, Kidwell MG. Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Molec Phyl Evol.* 2002;22:442–453. doi:10.1006/mpev.2001.1053
- [14] Robe LJ, Cordeiro J, Loreto ELS, et al. Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genetica.* 2010;138:601–617. doi:10.1007/s10709-009-9432-5
- [15] Burla H, DaCunha AB, Cordeiro AR, et al. The *willistoni* group of sibling species of *Drosophila*. *Evolution.* 1949;3:300–314. doi:10.1111/j.1558-5646.1949.tb00031.x
- [16] Hsu TC. The external genital apparatus of male Drosophilidae in relation to systematics. In: *Studies in Genetics of Drosophila*. Univ Texas Publ; 1949;4920:80–142

- [17] Malogowkin C. Sobre a genitália dos Drosophilidae (Diptera) III. Grupo *willistoni* do gênero *Drosophila*. Rev Bras Biol. 1952;12:79–96
- [18] Spassky B. Morphological differences between sibling species of *Drosophila*. Univ Texas Publ. 1957;5721:48–61
- [19] Pasteur G. Biometrical data on the semispecies of the *Drosophila paulistorum* complex. Evolution. 1970;24:156–168. doi:10.1111/j.1558-5646.1970.tb01749.x.
- [20] Rohde C, Monteiro AGF, Cabral WBM, et al. The importance of the identification of the *willistoni* subgroup of *Drosophila* at the species level: the first evidence of *D. equinoxialis* in the Northeast region of Brazil. Dros Inf Serv. 2010;93:118–122
- [21] Zanini R, Deprá M, Valente VLS. Can sibling species of the *Drosophila willistoni* subgroup be recognized through combined microscopy techniques? Rev Bras Ent. 2015b;59:323–331. doi:10.1016/j.rbe.2015.09.006
- [22] Perez-Salas S, Ehrmann L. Mechanisms of male sterility in hybrids of *D. paulistorum* group. Genetics. 1971;69:63–70
- [23] Rhode C, Garcia ACL, Valiati VH, et al. Chromosomal evolution of sibling species of the *Drosophila willistoni* group. I. Chromosomal arm IIR (Muller's element B). Genetica. 2006;126:77–88. doi:10.1007/s10709-005-1433-4
- [24] Masly JP, Presgraves DC. High-resolution genome wide dissection of the two rules of speciation in *Drosophila*. PLoS Biol. 2007;5:e243. doi:10.1371/journal.pbio.0050243
- [25] Richards O. The specific characters of the British bumblebees (Hymenoptera). Trans Ent Soc Lond. 1927;75:233–268. doi:10.1111/j.1365-2311.1927.tb00073.x
- [26] Liu J, Hercher JM, Stam LF, et al. Genetic analysis of a morphological shape difference in the male genitalia of *Drosophila simulans* and *mauritiana*. Genetics. 1996;142:1129–1145
- [27] Song H. Species-specificity of male genitalia characterized by shape, size and complexity. Ins Syst Evol. 2009;40:159. doi:10.1163/187631209X424571
- [28] Richmond MP, Johnson S, Markow TA. Evolution of reproductive morphology among recently diverged taxa in the *Drosophila mojavensis* species cluster. Ecol Evol. 2012;2:397–408. doi:10.1002/ece3.93
- [29] McPeck M, Shen AL, Torrey JZ, et al. The tempo and mode of three-dimensional morphological evolution in male reproductive structures. Am Nat. 2008;171:E158–E178. doi:10.1086/587076
- [30] McPeck MA, Shen L, Farid H. The correlated evolution of three-dimensional reproductive structures between male and female damselflies. Evolution. 2009;63:73–83. doi:10.1111/j.1558-5646.2008.00527.x
- [31] Gross M, Ramos MIF, Piller WE. A minute ostracod (Crustacea: Cytheromatidae) from the Miocene Solimões Formation (western Amazonia, Brazil): evidence for marine incursions? J Syst Palaeontol. 2015;14:1–22. doi:10.1080/14772019.2015.1078850.
- [32] Montes C, Cardona A, McFadden R, et al. Evidence for middle Eocene and younger land emergence in Central Panama: implications for isthmus closure. Geol Soc Am Bull. 2012;124:780–799. doi:10.1130/B30528.1
- [33] Werneck FP. The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. Quat Sci Rev. 2011;30:1630–1648. doi:10.1016/j.quascirev.2011.03.009
- [34] Lang M, Polihronakis RM, Acurio AE, et al. Radiation of the *Drosophila nanoptera* species group in Mexico. J Evol Biol. 2014;27:575–584. doi:10.1111/jeb.12325
- [35] Franco FF, Silva ECC, Barrios-Leal DY, et al. The Calibrated Phylogeny of the *Drosophila fasciola* Subgroup (*D. repleta* Group Wasserman) Indicates Neogene Diversification of Its Internal Branches. Neotrop Entomol. 2017;46:537–545. doi:10.1007/s13744-017-0484-4
- [36] Haffer J. Speciation on Amazonian forest birds. Science. 1969;165:131–137. doi:10.1126/science.165.3889.131
- [37] Haffer J. Amazonia: an overview. Biod Cons. 1997;6:451–476. doi:10.1023/A:1018320925954
- [38] Kastritis CD, Dobzhansky T. *Drosophila pavlovskiana*, a race or a species? Am Mid Nat. 1966;78:244–248. doi:10.2307/2423385
- [39] Mayr E. Populations, species and evolution: An abridgment of animal species and evolution. Cambridge: Harvard University Press; 1970
- [40] Etienne RS, Morlon E, Lambert A. Estimating the duration of speciation from phylogenies. Evolution. 2014;68:2430–2440. doi:10.1111/evo.12433
- [41] <http://www.nhm.ac.uk/hosted-sites/iczn/code/>
- [42] Yassin A. Phylogenetic relationships among species subgroups in the *Drosophila saltans* group (Diptera: Drosophilidae): Can morphology solve a molecular conflict? Zool Res. 2009;30(3):225–232. doi:10.3724/SP.J.1141.2009.03225
- [43] Souza TAJ, Noll FB, Bicudo HEMC, et al. Scanning electron microscopy of male terminalia and its application to species recognition and phylogenetic reconstruction in the *Drosophila saltans* group. Plos One. 2014;9(6):1–12. doi:10.1371/journal.pone.0097156
- [44] Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version 3.3. <http://mesquiteproject.org>
- [45] Lewis PO. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol. 2001;50:913–925. doi:10.1080/106351501753462876
- [46] Sassi AK, Heredia F, Loreto ELS, et al. Transposable elements P and gypsy in natural populations of *Drosophila willistoni*. Gen Molec Biol. 2005;28:734–739. doi:10.1590/S1415-47572005000500013
- [47] Staden R. The Staden sequence analysis package. Molec Biotech. 1996;5:233–24. doi:10.1007/BF02900361
- [48] Kumar S, Stecher G, Tamura K. MEGA 7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–1874. doi:10.1093/molbev/msw054
- [49] Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–1452. doi:10.1093/bioinformatics/btp187
- [50] Akaike H. A new look at the statistical model identification. IEEE Trans Automat. Control. 1974;19:716–723. doi:10.1109/TAC.1974.1100705

- [51] Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 2008;25(7):1253–1256. doi:10.1093/molbev/msn083
- [52] Xia X. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2013;30:1720–1728. doi:10.1093/molbev/mst064
- [53] Huelsenbeck JP, Ronquist F. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17:754–755. doi:10.1093/bioinformatics/17.8.754
- [54] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–1313. doi:10.1093/bioinformatics/btu033
- [55] Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biol.* 2014;10(4):e1003537. doi:10.1371/journal.pcbi.1003537
- [56] Rambaut A, Suchard MA, Xie D, et al. Computer Program and documentation distributed by the Author [WWW document]. 2014. URL <http://beast.bio.ed.ac.uk/Tracer>.
- [57] Powell JR, De Salle R. *Drosophila* molecular phylogenies and their use. In: Max H. Hecht, et al., editors. *Evolutionary biology*. Vol. 28. New York: Plenum Press; 1995.
- [58] Russo CAM, Takezaki N, Nei M. Molecular phylogeny and divergence times of Drosophilid species. *Mol Biol Evol.* 1995;12:391–404