

Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

**INSTITUTO OSWALDO CRUZ**  
**Mestrado em Biologia Computacional e Sistemas**

**Organização e Evolução da Família Gênica Carboxilesterase  
em Artrópodes**

André Luiz Quintanilha Torres

**Orientadores: Dr<sup>a</sup> Renata Schama**

Laboratório de Biologia Computacional e Sistemas

IOC/Fiocruz

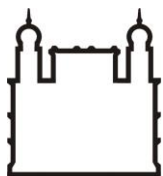
**Dr. Rafael Dias Mesquita**

Laboratório de Bioinformática

Instituto de Química/UFRJ

Rio de Janeiro

2013



Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

**INSTITUTO OSWALDO CRUZ**  
**Mestrado em Biologia Computacional e Sistemas**

**AUTOR:** André Luiz Quintanilha Torres

**Organização e Evolução da Família Gênica Carboxilesterase em  
Artrópodes**

**Orientadores: Dr<sup>a</sup> Renata Schama**

Laboratório de Biologia Computacional e Sistemas  
IOC/Fiocruz.

**Dr. Rafael Dias Mesquita**

Laboratório de Bioinformática  
Instituto de Química/UFRJ.

Aprovada em: 09/09/2013

**Examinadores:**

Dr<sup>a</sup> Cláudia Augusta de Moraes Russo

Dr<sup>a</sup> Denise Valle

Dr. Fabio Faria da Mota

Dr. Rodolpho Mattos Albano

Dr. Antônio Basílio de Miranda

Rio de Janeiro, 09 de Setembro 2013.

“Se queres prever o futuro, estude o passado.” - Confúcio

Dedico esse trabalho principalmente aos meus pais, Maria do Carmo Quintanilha Torres e Edemir Joaquim Torres, que tiveram papel crucial na minha educação. Aos meus orientadores, Dra. Renata Schama e Dr. Rafael Mesquita, que me conduziram atenciosamente durante o mestrado e sempre me estimularam a fazer o melhor pelo meu trabalho. E a pesquisadora Dra. Denise Valle que sempre trabalhou em prol da melhoria do combate a dengue no Brasil.

## **Agradecimentos**

Agradeço a todos do Laboratório de Fisiologia e Controle de Artrópodes Vetores (Fiocruz – IOC), especialmente a Dra. Renata Schama pela excelente orientação e estímulo para a conclusão deste trabalho. A Dra. Denise Valle, Priscila Fernandes, Diogo Belinato e Dr. Ademir Martins que me ajudaram no início do mestrado agregar conhecimento a respeito sobre os mecanismos de resistência a inseticidas e a importância do monitoramento e controle de insetos vetores.

A todos do Laboratório de Bioinformática (UFRJ), especialmente o Dr. Rafael Mesquita que não só disponibilizou o servidor que gerou grande parte dos resultados que estão neste trabalho, mas também me orientou a fazer sempre o melhor durante o desenvolvimento desta dissertação.

Ao Núcleo de Atendimento a Computação de Alto Desempenho (NACAD/COPPE-UFRJ) pela disponibilização do servidor para realização das reconstruções filogenéticas preliminares.

Ao Centro Nacional de Pesquisas (CNPq) pelo financiamento da bolsa de mestrado.

Agradeço também aos meus tios Érbio Faria e Lucilla Quintanilha, e minha prima Izabela Quintanilha Faria pelo suporte que me deram nas diversas vezes que participei de congressos localizados em Brasília. Além disso não posso deixar de agradecer as inesquecíveis idas às praias do Rio durante as férias de verão. Concluindo os merecidos agradecimentos, deixo aqui registrado minha eterna gratidão aos meus pais que foram as pessoas que mais contribuíram para o meu crescimento científico. Provavelmente, sem eles, eu não teria chegado até aqui.



## Índice de Imagens

<b>Figura 1.1</b> Esquema mostrando o aumento da resistência em uma população hipotética..	4
<b>Figura 1.2</b> Esquema dos três principais tipos de alterações que conferem resistência a inseticida.....	6
<b>Figura 1.3</b> Fases da metabolização de xenobióticos, desde a entrada no corpo do inseto até sua excreção.....	7
<b>Figura 1.4</b> Visão tridimensional da estrutura cristalizada de um membro da família CCE (PDB: 2FJ0) responsável pela degradação de hormônio juvenil na espécie <i>Manduca sexta</i> .....	10
<b>Figura 1.5</b> Esquema comparativo de regiões características de proteínas bem caracterizadas pertencentes a família das CCEs.....	11
<b>Figura 1.6</b> Cladograma esquemático representando a organização dos 13 clados da família CCE que foram identificados por Oakeshott (2005).....	15
<b>Figura 1.7</b> Esquema mostrando alguns eventos de rearranjo cromossômico.....	16
<b>Figura 1.8</b> Cladograma esquemático mostrando as relações entre dois genes hipotéticos.....	17
<b>Figura 1.9</b> Esquema de diferentes mecanismos de duplicação gênica.....	20
<b>Figura 1.10</b> Esquema mostrado os tipos de consequências que podem ocorrer para as cópias duplicadas.....	21
<b>Figura 1.11</b> Esquema resumido dos subfilos que constituem o filo Arthropoda. Em vermelho está destacado a subfilo Hexapoda (Adaptado de Edgecome, 2010).....	23
<b>Figura 1.12</b> Representação taxonômica atual da classe Insecta.....	24
<b>Figura 3.13</b> Esquema do pipeline usado para as buscas nos genomas como descrito para o programa FAT (Seabra-Junior et al., 2011).....	29
<b>Figura 3.14</b> Esquema do pipeline usado para executar os programas de predição e editar os resultados de forma codificada nos cabeçalhos dos arquivos em formato FASTA...	31
<b>Figura 4.15</b> Número total de sequências encontradas por espécie.....	33
<b>Figura 4.16</b> Distribuição de tamanho (em número de resíduos) das sequências encontradas pelo pipeline FAT usando o perfil COesterase (PF00135) para as 36 espécies.....	34
<b>Figura 4.17</b> Distribuição de tamanho das sequências de CCEs encontradas pelo pipeline FAT para 12 espécies de <i>Drosophila</i> .....	35
<b>Figura 4.18</b> Gráfico de dispersão mostrando a relação entre o Log10 do e-value e tamanho da sequência.....	36
<b>Figura 4.19</b> Gráfico de dispersão mostrando a relação entre o Log10 e-value e tamanho da sequência encontrada pelo pipeline FAT para as 12 espécies de <i>Drosophila</i> .....	38

<b>Figura 4.20</b> Tabela gerada pelo pipeline desenvolvido para organizar as informações relevantes do trabalho. ....	39
<b>Figura 4.21</b> Exemplo do arquivo FASTA das sequências após a codificação das informações dos programas de predição no cabeçalho. ....	40
<b>Figura 4.22</b> Ordenamento do tamanho das sequências antes e depois da execução do passo de exclusão feita pelo pipeline de processamento de informação. ....	40
<b>Figura 4.23</b> Número total de sequências por espécie após a exclusão feita pelo pipeline secundário. ....	41
<b>Figura 4.24</b> Padrão de dispersão de proteínas encontradas de acordo com a relação ao valor de e-value obtido na busca HMM e seu respectivo número de resíduos. ....	44
<b>Figura 4.25</b> Estatística dos resultados obtidos pelos programas de predição. ....	46
<b>Figura 4.26</b> Árvore filogenética não enraizada dos membros da família CCE. ....	48
<b>Figura 4.27</b> Neurologinas (Clado L). Apenas valores de bootstrap maiores que 60% estão mostrados. ....	52
<b>Figura 4.28</b> A – Neurotactina (Clado M) e B – Gliotactina (Clado K). ....	55
<b>Figura 4.29</b> A - Clado Não caracterizado (Clado I). B – Clado Não caracterizado 1 – Novo. ....	57
<b>Figura 4.30</b> Clado das glutactinas (Clado H). ....	59
<b>Figura 4.31</b> Clado das acetilcolinesterases (Clado J). ....	61
<b>Figura 4.32</b> Clado Tegumentares (Clado D). ....	63
<b>Figura 4.33</b> Clado Beta-esterases - <i>Drosophila</i> . ....	65
<b>Figura 4.34</b> Clado JHE – Hemiptera, Clado Beta-esterase – Conservada e Clado Beta-esterase - Multi ordem. ....	67
<b>Figura 4.35</b> Clado Não caracterizado – <i>T. castaneum</i> . ....	68
<b>Figura 4.36</b> Clado F ( Clado Não caracterizado – Hymenoptera e Clado JHE – Diptera). ....	71
<b>Figura 4.37</b> Clado JHE Lepidoptera (Clado G). ....	72
<b>Figura 4.38</b> Grupo Detoxicante com os principais clados colapsados. ....	73
<b>Figura 4.39</b> Clado Diptera – Detox 1, Diptera – Detox 2 e Coleoptera – Detox 1. Apenas bootstrap maiores que 60% estão mostrados. ....	78
<b>Figura 4.40</b> Clado Diptera – Detox 3, Hymenoptera – Detox 1, Hymenoptera – Detox 2, Lepidoptera – Detox 1, Lepidoptera – Detox 2, Phthiraptera – Detox 1 e Hemiptera – Detox 1. ....	81
<b>Figura 5.41</b> Organização do <i>cluster</i> ancestral de genes homólogos de JHE em quatro espécies de dípteros. ....	88
<b>Figura 5.42</b> Organização de um segundo <i>cluster</i> de genes homólogos de JHE duplicados em <i>Cu. quinquefasciatus</i> . ....	89

## Índice de Tabelas

<b>Tabela 1.1.</b> Lista resumida das principais classes de inseticidas usados no controle de insetos e seus respectivos alvos ou efeitos (para classes sem mecanismo de ação elucidado). .....	3
<b>Tabela 3.2</b> Lista de genomas com predição gênica utilizados neste trabalho. ....	28
<b>Tabela 4.3</b> Clados identificados no presente trabalho e por Oakeshot e colaboradores em 2010. ....	49
<b>Tabela 4.4</b> Número de genes por espécie nos clados neurologinas. ....	53
<b>Tabela 5.5.</b> Número de CCEs por espécie encontradas em cada clado encontrado. ....	86

## Listagem de Abreviaturas

ACh	Acetilcolina
AChE	Acetilcolinesterase
Asp	Ácido aspártico
Bti	<i>Bacillus thuringiensis var. israelensis</i>
CCE	Carboxilesterases
CSI	Chitin Synthase Inhibitors
CX	Carbamatos
CYP	Monooxigenase de função múltipla
DDT	Dicloro-Difenil-Tricloroetano
DNA	Ácido desoxiribonucleico
GABA	Ácido gama-aminobutírico
Gli	Glicina
Glu	Ácido glutâmico
GST	Glutathiona S-transferase
His	Histidina
HJ	Hormônio Juvenil
HMM	<i>hidden markov models</i>
IGR	<i>Insect Growth Regulator</i>
JHA	<i>Juvenile Hormone Analogs</i>
JHE	<i>Juvenile Hormone Esterase</i>
Mb	Megabase
mDNA	DNA mitocondrial
Nav	Canal de sódio regulado por voltagem
NCBI	National Center for Biotechnology Information
NLGN	Neuroliginas
OC	Organoclorados
OP	Organofosforados
PNCD	Programa Nacional de Controle da Dengue
PY	Piretróides
rDNA	DNA ribossomal
Ser	Serina
SNC	Sistema Nervoso Central

## Resumo

A família gênica das carboxilesterases (CCEs) é muito diversa e, em insetos, está dividida em três principais classes. Duas classes, Detoxificantes e Processamento de hormônios e feromônios, contêm genes majoritariamente catalíticos envolvidos com o metabolismo de xenobióticos e degradação de moléculas de hormônios que sinalizam processos importantes como a ecdise. No entanto, alguns destes genes já foram implicados na resistência a organofosforados (OP) e carbamatos (CX) em diversas espécies artrópodes. A classe Desenvolvimento neuronal contém genes não catalíticos, com exceção do codificante para a enzima acetilcolinesterase que é alvo de OP e CX. Este gene também já foi implicado na resistência a inseticidas por insensibilidade do sítio alvo. Isso faz com que essa família seja importante no monitoramento e controle de insetos vetores e pragas agrícolas. Todos os membros possuem o domínio COesterase que é caracterizado pela dobra alfa/beta hidrolase que posiciona os três resíduos da tríade catalítica (Ser-Glu-His ou Ser-Asp-His) no fundo da cavidade de ligação com o substrato. Estes genes evoluíram por duplicação, em sua maioria por recombinação desigual, e por isso as cópias tendem a ficar dispostas em *clusters* no genoma. Nesse trabalho foram identificados 1299 CCEs homólogas em 36 espécies (33 de insetos) e foi observada uma média de 36 sequências por espécie. *Ixodes scapularis* (71) e *P. humanus* (17) possuem o maior e o menor número de CCEs, respectivamente. Outros vetores, como *Cu. quinquefasciatus* e *Ae. aegypti*, também apresentaram número acima da média. Os himenópteros possuem um número menor de sequências em relação à média, exceto *N. vitripennis*. Sessenta e um por cento das sequências possuem os três resíduos catalíticos, 56% são secretadas, 34% intracelulares e 10% mitocondriais. Regiões transmembranares foram encontradas em 27% das sequências. No total foram identificados 39 clados e, como esperado, em alguns casos, com o aumento da amostragem não houve suporte estatístico corroborando a classificação proposta por Oakeshott e colaboradores em 2010. A maior diferença em relação à classificação anterior foi observada para a classe Detoxificante. Esta possui clados contendo proteínas de espécies de uma mesma ordem taxonômica e não relacionadas funcionalmente, indicando que as duplicações ocorreram dentro das diferentes linhagens. Nessa classe, a maioria dos genes possui a tríade Ser-Glu-His e é endereçado para as três localizações celulares. Na classe Processadoras de hormônios e feromônios, as JHE também parecem ter evoluído independentemente em cada linhagem de inseto pois em lepidópteros, dípteros e hemípteros estes genes ficam em clados distintos. Na classe Desenvolvimento neuronal, diferente das outras duas, um número menor de duplicações gênicas para cada espécie é encontrado. Nesta classe, a presença de regiões associadas com o ancoramento na membrana em alguns membros sugere que houve fusão entre domínios transmembranares e o domínio COesterase ao longo da história evolutiva. Os resultados mostram que a classificação funcional é aplicável apenas para a classe Desenvolvimento neuronal. Para as outras duas classes os genes dentro das diferentes linhagens taxonômicas são mais próximos filogeneticamente indicando que a evolução da funcionalidade ocorreu separadamente em cada linhagem.

## Abstract

The carboxylesterase (CCE) gene family is very diverse and is divided into three main classes in insects. Two classes, Dietary/detoxification and Hormone/semiochemical processing are catalytic and contain genes mainly involved with xenobiotic metabolism and degradation of hormone molecules. However, some of these genes have been implicated in the resistance to organophosphorus (OP) and carbamates (CX) insecticides in several arthropod species. The Neurodevelopmental class contains non-catalytic genes, with the exception of the gene encoding the enzyme acetylcholinesterase that is the target of OP and CX. This gene has also been implicated in insecticide resistance due to mutations that affect insecticide binding. This fact makes CCE an important family in insecticide resistance monitoring and control of agricultural pests and insect vectors. All CCE members have a COesterase domain, characterized by an alpha/beta hydrolase fold that is responsible for correctly positioning the catalytic triad (Ser-Glu-His or Ser-Asp-His). The genes in this family evolved by duplication, usually by unequal crossing-over, so the copies tend to be arranged in clusters in the genome. In this work 1299 CCEs were identified in 36 species (33 insects) and an average of 36 sequences per species was observed. *Ixodes scapularis* (71) and *P. humanus* (17) have the highest and lowest number of CCEs, respectively. Other insect vectors such as *Cu. quinquefasciatus* and *Ae. aegypti*, also showed an increased CCE number. Hymenoptera have a deficit in the number of sequences except *N. vitripennis*. Sixty one percent of the sequences have catalytic residues, 56% are secreted, 34% are intracellular and 10% are mitochondrial. Transmembrane regions were found in 27% of the sequences. In total, 39 clades were identified and, as expected, in some cases, inconsistencies when compared to Oakeshott et al classification in 2010 were found. In the dietary/detoxification class a taxonomic instead of functional classification seems to fit better the evolution of the genes found. In this class, most of the genes have the Ser-Glu-His triad and didn't show a specific cellular localization. In the hormone/semiochemical processing class, the JHE also appear to have evolved independently on each insect lineage, lepidoteran, dipteran and hemipteran species are grouped in different clades. The Neurodevelopmental class, contrasting with the other two classes, showed a smaller number of genomic duplications for each species and functional relationship among its genes. In this class, the presence of membrane anchoring regions for some members suggests that transmembrane and COesterase domains were fused during the evolutionary history. The results show that the functional classification works only for the Neurodevelopmental class. However, for the other two classes the taxonomic group appears to be more important than the function of these genes for their classification.

## Índice

Índice de Imagens .....	v
Índice de Tabelas .....	vii
Listagem de Abreviaturas .....	viii
Resumo .....	ix
Abstract.....	x
1. Introdução.....	1
1.1 Controle de insetos.....	1
1.2 Resistência metabólica em insetos.....	6
1.3 Carboxilesterases .....	8
<b>1.3.1 Caracterização</b> .....	8
<b>1.3.2 Classificação</b> .....	12
A. <i>Detoxificantes</i> .....	12
B. <i>Processadoras de hormônios e feromônios</i> .....	13
C. <i>Desenvolvimento neuronal</i> .....	14
1.4 Evolução de famílias gênicas.....	15
1.5 Taxonomia de Artrópodes .....	22
1.6 Bioinformática .....	25
2. Objetivos .....	26
2.1 Objetivo geral .....	26
2.2 Objetivos específicos .....	26
3. Material e métodos.....	27
3.1 Bases de dados .....	27
3.2 Busca e identificação das sequências proteicas .....	29
3.3 <i>Pipeline</i> de processamento de informação.....	30
3.4 Análises.....	31
4. Resultados.....	32
4.1 Sequências encontradas .....	32
4.2 <i>Pipeline</i> de processamento de informação.....	38
4.3 Predições.....	44
4.4 Análises filogenéticas de CCEs de artrópodes .....	46
<b>4.4.1 Desenvolvimento neuronal</b> .....	51
<b>4.4.2 Processadoras de Hormônios e feromônios</b> .....	62
<b>4.4.3 Detoxificantes</b> .....	73
5. Discussão.....	82
5.1 Conteúdo gênico nos genomas .....	82
5.2 Evolução da família Carboxilesterase em insetos.....	84
5.3 Classificação das Carboxilesterases de inseto .....	90
5.4 Evolução dos domínios .....	91
6. Conclusões.....	95
7. Referências Bibliográficas .....	96
Anexos.....	109
<b>Figura S1.</b> Árvore filogenética enraizada dos membros da família CCE.....	119
<b>Figura S2.</b> Árvore filogenética não enraizada dos membros da família CCE.....	120
<b>Figura S3.</b> Esquema resumido das ordens do subfilo Hexapoda representadas nesse trabalho.....	121

<b>Tabela S1.</b> Taxonomia das espécies estudadas no presente trabalho e a respectiva correlação entre abreviação usada no prefixo da identificação das sequências nas árvores filogenéticas e prefixo original do banco de dados.....	122
<b>Tabela S2.</b> Sequências com e-value menor que $1e-58$ não incluídas na filogenia devido à problemas de processamento. ....	123



# 1. Introdução

## 1.1 Controle de insetos

No Brasil e no mundo, diversas doenças com impacto importante na saúde pública como a dengue, febre amarela, malária, leishmaniose, entre outras, são transmitidas por insetos vetores (Hemingway & Ranson, 2000). Além disso, outras espécies de insetos prejudicam o comércio agropecuário, como no caso de pestes agrícolas (McKee et al., 2008). Devido a estes e outros problemas, diversos métodos foram desenvolvidos e podem ser empregados no controle de insetos como o uso de agentes biológicos, busca ativa e eliminação dos criadouros e utilização de inseticidas (Braga & Valle, 2007a). Para a saúde pública, doenças transmitidas por insetos vetores sempre foram um grande problema em países tropicais e subtropicais e com a expansão do habitat de diversas espécies, hoje isso tem se tornado um problema mundial. Por exemplo, no caso da dengue, o mosquito da espécie *Aedes aegypti* (Linnaeus, 1762) é o principal vetor (Reed, 1900) e acredita-se que no mundo cerca de 2 milhões de pessoas vivam em áreas de risco (WHO, 2006). Para o controle do *Ae. aegypti* no Brasil, mesmo com o recente crescimento de programas que incentivam a eliminação de criadouros domiciliares, a utilização de inseticidas é considerada um dos principais métodos para diminuir o tamanho populacional do vetor (Fundação Nacional de Saúde, 2002). Historicamente, inseticidas organoclorados (OC) foram usados mundialmente tanto para o controle de insetos vetores de doenças quanto de pestes agrícolas. No entanto, esta classe foi banida da maior parte dos países devido a seus efeitos cumulativos no ambiente (Waliszewski et al., 2003). Em 1967, com a reintrodução do vetor da febre amarela urbana e da dengue no Brasil, a utilização de organofosforados (OP) pelos programas de controle vetorial foi iniciada para o controle do mosquito *Ae. aegypti*. Sua utilização foi intensificada em 1986 devido aos frequentes surtos epidêmicos (Schatzmayr et al., 1986). Em 1999, o detecção de resistência a OPs em diversas populações do país catalisou a troca desse inseticida. Contra mosquitos na fase adulta, passaram a ser utilizados inseticidas da classe dos piretróides (PY) em todo o país e contra os estágios larvais, iniciou-se o controle biológico com o microorganismo *Bacillus thuringiensis* var. *israelensis* (Bti) onde havia sido detectada resistência a OP (Funasa, 2001). Atualmente, devido ao surgimento de populações resistentes tanto a inseticidas neurotóxicos quanto ao baixo efeito residual de Bti, muitos municípios

passaram a adotar reguladores do crescimento contra mosquitos no estágio larval (Braga et al., 2004; Silva & Mendes, 2007; Martins et al., 2008). Esse breve histórico representa as medidas que foram tomadas até o momento em relação ao controle do mosquito *Ae. aegypti* no Brasil. Para outros vetores e pragas agrícolas o panorama de controle é diferente, o que complica ainda mais o padrão de dispersão da resistência no país. Estas estratégias de controle, com alternância de classe de inseticida, visam dificultar o surgimento de resistência pois as diferentes classes possuem mecanismos de ação distintos e a alternância periódica diminui a pressão seletiva sobre as populações alvo (Georghiou, 1994; Zhao et al., 2010).

Em relação aos inseticidas, a maior parte deles são neurotóxicos e atuam no sistema nervoso central (SNC) dos insetos, prejudicando o funcionamento normal dos neurônios (Tabela 1.1). Esses inseticidas se ligam a alvos diferentes que se localizam na membrana celular do neurônio e assim interferem na propagação do impulso nervoso (Futuko et al., 1990; Davies et al., 2007). Os inseticidas da classe dos OCs possuem como principal representante o DDT, cujo mecanismo de ação ainda não é muito bem descrito. Sua atuação acontece no canal de sódio regulado por voltagem ( $Na_v$ ), que se localiza na membrana do axônio de neurônios (Davies et al., 2007). Outros inseticidas dessa classe atuam no receptor de ácido gama-aminobutírico (GABA) que é responsável por proporcionar o estado de hiperpolarização nos neurônios, por meio do aumento da permeabilidade a íons cloreto (Ware, 2000). Assim como o DDT, os inseticidas da classe PY atuam se ligando ao  $Na_v$ , causando um desequilíbrio das cargas positivas e negativas do meio intra e extracelular. Isso provoca repetidas despolarizações que causam super-estimulação e paralisia, podendo levar o inseto à morte (Davies et al., 2007). As classes dos OPs e dos carbamatos (CXs), também atuam no SNC, porém possuem alvo diferente dos OCs e PYs. Os inseticidas dessas duas classes inibem a enzima acetilcolinesterase (AChE), fazendo com que o neurotransmissor acetilcolina (ACh) se acumule na fenda sináptica e mantenha o estímulo para o impulso nervoso, também levando o inseto à morte (Fukuto, 1990).

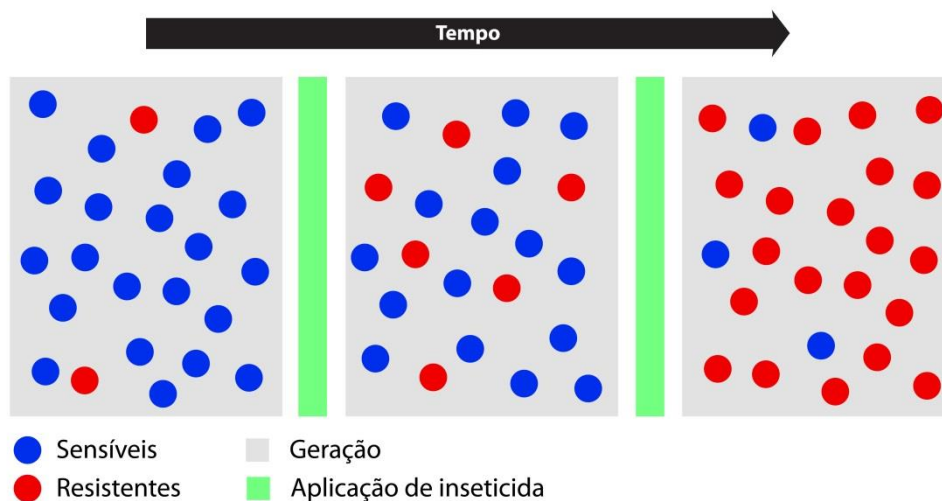
O grupo de inseticidas reguladores do crescimento de insetos, mais conhecidos como IGRs devido a sua denominação na língua inglesa (*Insect Growth Regulator*), atua modulando de maneira adversa processos fisiológicos essenciais para o desenvolvimento e crescimento do inseto como a troca do exoesqueleto (Graf, 1993). Em comparação à classe dos inseticidas que atuam no SNC, os IGRs são menos tóxicos para outros metazóários pois atuam no processo de troca do exoesqueleto que é uma

característica exclusiva de artrópodes (Tunaz & Uygun, 2004). Além disso, atualmente essa classe é vista como uma boa alternativa para o controle de populações resistentes aos inseticidas neurotóxicos devido à radical diferença em seus mecanismos de ação (Martins et al., 2008; Suman et al., 2010). Os análogos de hormônio juvenil (JHAs, do inglês *Juvenile Hormone Analogs*), por exemplo, são moléculas análogas ao hormônio encontrado na maioria dos artrópodes que está envolvido no processo de muda. A diminuição dos níveis de hormônio circulante na hemolinfa sinaliza o processo de muda além de regular outros processos fisiológicos importantes (Graf et al., 1993; Tunaz & Uygun, 2004). Moléculas de JHA, quando aplicadas em insetos durante o estágio embrionário, podem causar má formação do inseto e prejudicar a metamorfose. Isso já foi demonstrado para a espécie *Ae. aegypti*, onde a maioria dos insetos morre no estágio de pupa ou durante a emergência para a fase adulta (Braga et al., 2005). Outros IGRs também podem ser usados no controle de insetos como os inibidores da síntese de quitina (CSI, do inglês *Chitin Synthase Inhibitors*) e os agonistas de ecdisona (Tunaz & Uygun, 2004). O mecanismo de ação dos CSI ainda não foi bem descrito embora existam evidências de que sejam eficazes no controle de insetos, inclusive em populações resistentes a inseticidas neurotóxicos (Martins et al., 2008; Belinato et al., 2009; Fontoura et al., 2012).

**Tabela 1.1.** Lista resumida das principais classes de inseticidas usadas no controle de insetos e seus respectivos alvos ou efeitos (para classes sem mecanismo de ação elucidado).

Grupo	Classes	Alvo ou efeito
Neurotóxicos	Organoclorados (OC)	Canal de sódio ou Receptor de GABA
	Organofosforados (OP)	Acetilcolinesterase
	Carbamatos (CX)	Acetilcolinesterase
	Piretroides (PY)	Canal de sódio
IGRs	Agonista de Ecdisona, hormônio da muda	Receptores de ecdisona/Indução de mudas
	Análogos de hormônio juvenil	Inibe a metamorfose para estágio adulto
	Inibidor da síntese de quitina	Inibe a produção de quitina

O surgimento de resistência aos inseticidas usados atualmente no país e no mundo faz com que um programa de controle de vetores e pragas seja fundamental. Por exemplo, no caso da dengue, a Rede MoRENAa que faz parte do Programa Nacional de Controle da Dengue (PNCD) veio como uma forma de monitorar as populações do vetor para que a resistência aos inseticidas em uso seja detectada quando ainda estiver em baixa incidência (Fundação Nacional de Saúde, 2002). Em populações naturais, devido à variabilidade genética, parte dos indivíduos pode ser resistente a algum inseticida das diferentes classes. Devido às constantes aplicações esses indivíduos sobrevivem e passam essa característica para as próximas gerações o que provoca o surgimento de populações resistentes (Figura 1.1) (Chevillon et al., 1999; Lenormand et al., 1999).



**Figura 1.1** Esquema mostrando o aumento da resistência em uma população hipotética. Os círculos azuis representam indivíduos sensíveis a um determinado inseticida e os vermelhos os resistentes. Cada retângulo cinza representa uma fotografia da composição genética da população em cada geração, após a exposição a um inseticida. Os retângulos verdes representam as aplicações de um determinado inseticida. A seta preta demonstra o passar do tempo.

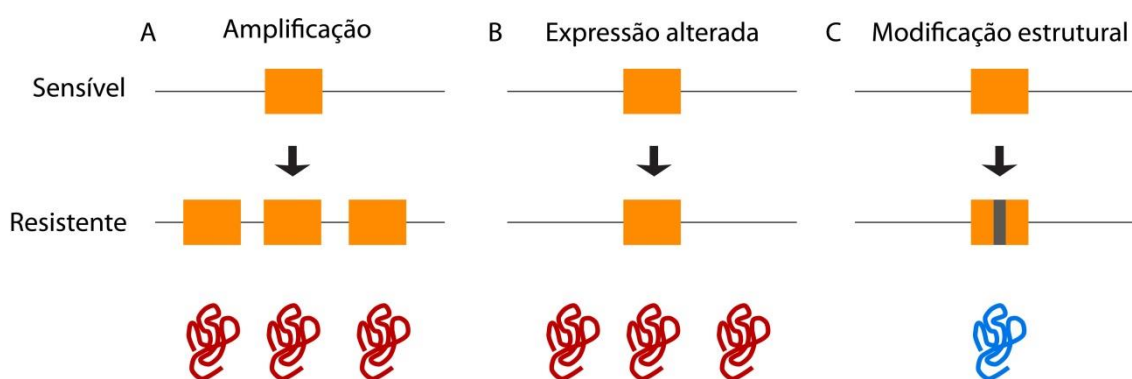
A resistência pode ser provocada por razões distintas e por isso, está dividida em três principais mecanismos que já foram descritos em populações naturais: A) redução da permeabilidade da cutícula (Medina et al., 2002); B) mutações pontuais e/ou mudanças de expressão nos alvos dos inseticidas (Hemingway et al., 2004; Russell et al., 2004) e C) detoxificação através de alteração seja do nível de expressão ou seja da especificidade enzimática, o que permite melhor degradação ou sequestro das moléculas de inseticida (Oakeshott et al., 2003; Vontas et al., 2001).

Os motivos da redução da penetrabilidade de inseticidas através da cutícula ainda não são muito bem caracterizados, embora existam evidências de que isso ocorra em insetos (Medina et al., 2002; Ahmad et al., 2006). Em populações de *Helicoverpa armigera* sensíveis e resistentes a piretróides, foi observado que a penetração do inseticida foi mais lenta em insetos resistentes do que nos sensíveis (Ahmad et al., 2006) indicando, portanto, que a diferença de penetração do inseticida através da cutícula pode ser uma das causas de resistência. A insensibilidade do sítio alvo é caracterizada pela ocorrência de mutações na sequência codificante do gene alvo de inseticidas que resultam em substituições na sequência de aminoácidos e com isso prejudicam a ligação do inseticida. Essas alterações podem ser identificadas através da comparação da sequência gênica da enzima alvo de indivíduos sensíveis e resistentes. No caso do canal de sódio, por exemplo, algumas mutações já foram identificadas em diversas espécies como, por exemplo, *Musca domestica* e *Haematobia irritans* e foram associadas ao surgimento da resistência a inseticidas da classe PY (Williamson et al., 1996; Guerrero et al., 1997). No caso da enzima AChE, três mutações pontuais (G119S; F290V; F331W) já foram associadas como causas da insensibilidade a OPs e CXs em espécies como *Anopheles gambiae*, *Culex pipiens* e *Tetranychus urticae* (Weill et al., 2003; Alout et al., 2007; Kwon et al., 2010). Entretanto, outras espécies como *Ae. aegypti*, por exemplo, não possuem modificações descritas na sequência nucleotídica do gene AChE associadas com a resistência e, portanto, acredita-se que as populações resistentes a OP e CX desta espécie possuam alterações em enzimas detoxificantes (Weill et al., 2004).

Existem evidências de que, além de *Ae. aegypti*, em outras espécies as enzimas envolvidas com o metabolismo de xenobióticos participam da detoxificação de inseticidas (para uma revisão ver Bass & Field, 2011). Este tipo de mecanismo é denominado resistência metabólica, onde membros de famílias gênicas participam metabolizando ou sequestrando moléculas de inseticida circulantes no corpo do inseto (Field et al., 1988; Vontas et al., 2001; Braga & Valle, 2007b; Puinean et al., 2010). Uma melhor metabolização do inseticida pode ser alcançada tanto por uma maior especificidade entre enzima detoxificante e inseticida quanto por uma maior produção de produto gênico. Estes dois tipos de alteração já foram descritos em populações naturais de insetos e são causados por diferentes mecanismos. No caso de um aumento na quantidade de produto gênico, tanto a amplificação gênica (Figura 1.2A) quanto alterações no promotor do gene (Figura 1.2B) provocam o aumento de sua transcrição.

Isso resulta em uma maior quantidade de enzimas disponíveis capazes de impedir o acesso das moléculas de inseticida ao seu alvo. Para algumas espécies de insetos, como *Cu. quinquefasciatus* e *Myzus persicae*, isso já é bem caracterizado e um número maior de transcritos destas enzimas é observado em indivíduos resistentes quando comparados com susceptíveis (Raymond et al., 1989; Devonshire et al., 1998; Field et al., 1999; Cui et al., 2006). No caso de maior especificidade, mutações pontuais na sequência codificante do gene podem causar alterações que melhoram a ligação entre enzima e um substrato, nesse caso, com o inseticida (Heidari et al., 2004a, 2004b) (Figura 1.2C).

Assim, diversos tipos de alterações nesses genes podem resultar no surgimento de fenótipos resistentes. Além disso, a ubiquidade com que os diferentes tipos de alterações acontecem sugere que esses eventos podem representar um mecanismo comum de adaptação evolutiva em artrópodes (Bass & Field, 2011). É importante levar em consideração que diferentes mecanismos de resistência podem ocorrer concomitantemente, conferindo caráter multifatorial à resistência, o que torna sua caracterização uma tarefa complexa (Saavedra-Rodriguez et al., 2011; Silva et al., 2012). Em *Musca domestica*, por exemplo, resistência a OPs foi associada a alterações tanto em genes envolvidos com a resistência metabólica quanto na sequência da enzima alvo, AChE (Syvanen et al., 1994; 1996; Russel et al., 2004).

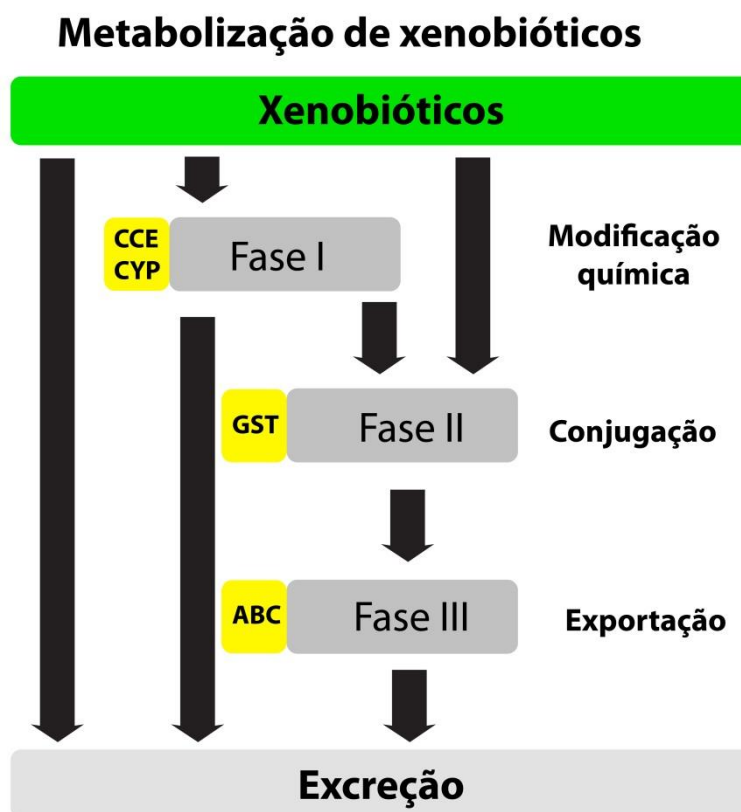


**Figura 1.2** Esquema dos três principais tipos de alterações que conferem resistência a inseticida. A) Amplificação gênica; B) Alteração da expressão gênica e C) Modificação estrutural causada por mutação.

## 1.2 Resistência metabólica em insetos

As principais famílias gênicas implicadas na resistência a inseticidas são representadas pelas enzimas glutatona S-transferases (GST), monooxigenases de função múltipla (CYP), carboxilesterases (CCE) e, recentemente, alguns indícios

também apontam que os transportadores ABC tem potencial envolvimento com a resistência (Gahan et al., 2010; Bass & Field, 2011). As enzimas dessas famílias gênicas atuam em diferentes fases da detoxificação de xenobióticos e proporcionam uma excreção mais eficiente das substâncias tóxicas (Xu et al., 2005). Membros das famílias CCEs e CYPs participam da fase de modificação estrutural do substrato (Fase I), onde atuam respectivamente hidrolisando e oxidando seus alvos. Membros da família das GSTs participam da fase de conjugação (Fase II), onde atuam adicionando uma molécula de glutatona aos substratos, tornando-os mais solúveis. Os transportadores ABC estão implicados na fase de exportação dos metabólitos (Fase III) transportando compostos tóxicos para fora das células (Xu et al., 2005) (Figura 1.3).



**Figura 1.3** Fases da metabolização de xenobióticos, desde a entrada no corpo do inseto até sua excreção. Em amarelo estão representadas as principais famílias gênicas envolvidas no processo de detoxificação. Carboxilesterases (CCE) e monooxigenases de função múltipla (CYP) são enzimas associadas à fase I ou fase de modificação química. Glutatonas S-transferase (GST) são associadas com a fase II ou fase de conjugação. Transportadores ABC estão associados com a fase III ou fase de exportação.

Por possuírem alta diversidade molecular, essas famílias foram organizadas de acordo com suas principais características funcionais e relacionamento filogenético (Xu

et al., 2005; Oakeshott et al., 2005; 2010; Friedman, 2011). Em insetos, a família das GSTs atualmente está dividida em seis subclasses: Delta, Epsilon, Sigma, Omega, Theta e Zeta (Friedman, 2011). Os membros dessa família são capazes de tornar moléculas naturais ou artificiais mais solúveis em água através da conjugação de uma molécula de glutationa ao substrato (Habig et al., 1974). A família das CYPs mostra uma grande expansão em insetos quando comparada com outros eucariotos (Hlavica et al., 2011). Seus membros estão diretamente envolvidos com a biotransformação da maioria dos xenobióticos através da reação de oxidação, o que os torna mais solúveis para serem excretados com maior facilidade (Feyereisen, 1999; Poupardin et al., 2010). Os transportadores ABC são complexos proteicos transmembranares que utilizam energia da hidrólise de uma molécula de adenosina trifosfato para transportar diversos tipos de substratos através da membrana celular como, por exemplo, componentes citotóxicos presentes na célula (Davidson et al., 2008; Gahan et al., 2010; Licht & Schneider, 2011). A família das CCEs atualmente está dividida em três grandes grupos: Detoxificantes, Processadoras de hormônios e feromônios e Desenvolvimento neuronal (Oakeshott et al., 2005, 2010; Montella et al., 2012). Nessa família, os membros do grupo detoxificante são os que majoritariamente estão envolvidos com a resistência metabólica (Oakeshott et al., 2005).

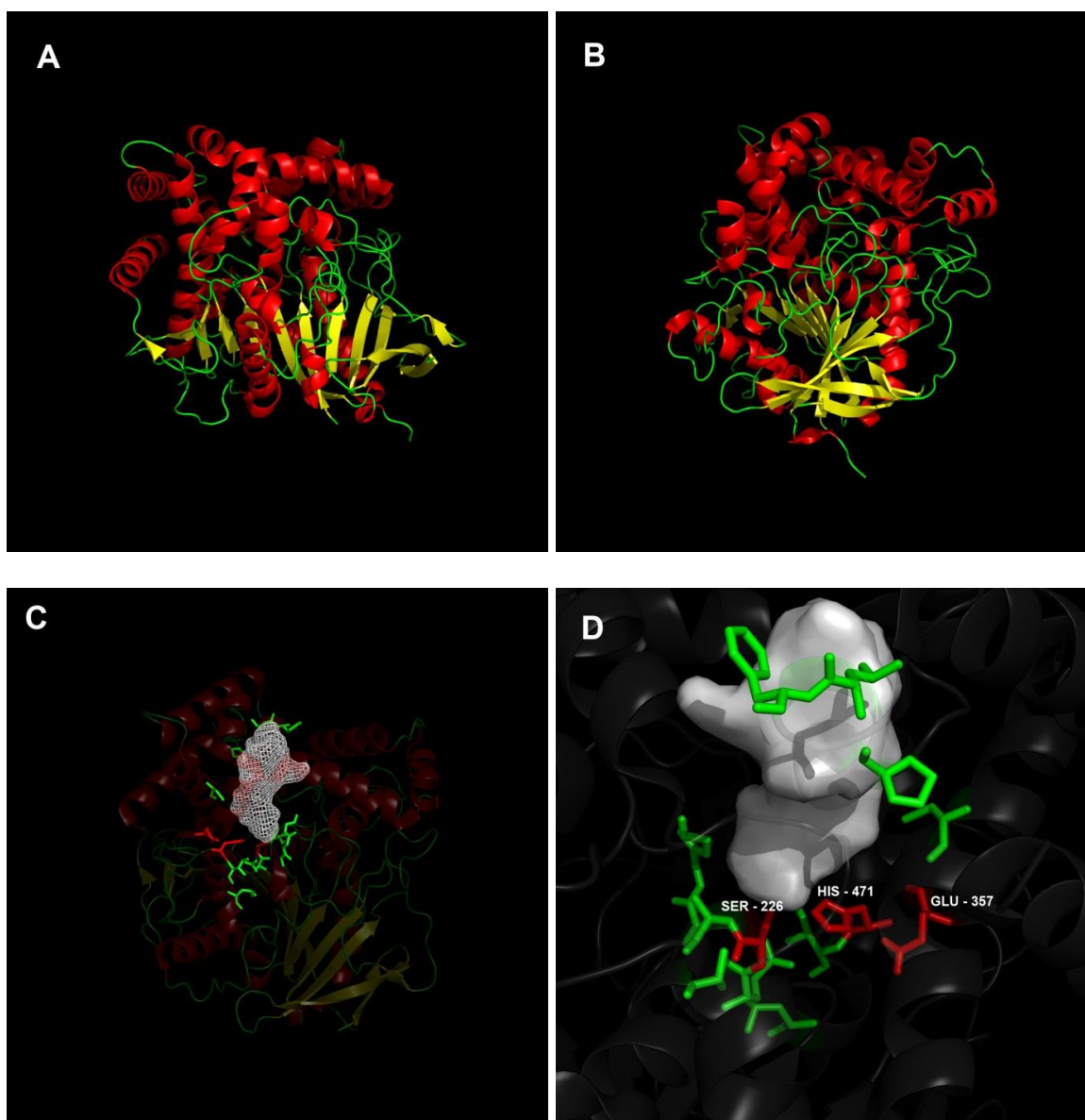
### 1.3 Carboxilesterases

#### *1.3.1 Caracterização*

A família das CCEs é diversa e pertence ao grupo das alfa/beta hidrolases (Lenfant et al., 2013) cujos membros possuem em comum o domínio COesterase (domínio Pfam PF00135). A enzima AChE, que é alvo de OPs e CXs, faz parte desta família, tornando-a importante não só para estudos de resistência metabólica como de resistência por insensibilidade de sítio alvo. Nas CCEs existem membros catalíticos e não catalíticos; os ativos participam da clivagem de substratos ésteres e, devido a sua alta diversidade molecular e baixa especificidade de substrato, são de difícil classificação. Seus membros apresentam sequência de aminoácidos com regiões conservadas, como o sítio catalítico, e regiões de baixa conservação, principalmente nas dobras entre as estruturas secundárias características da família (Figura 1.4A e B). Em relação à sua conformação tridimensional, as CCEs assim como as lipases (entre outras enzimas), apresentam a dobra alfa/beta hidrolase (PF00561) caracterizada basicamente

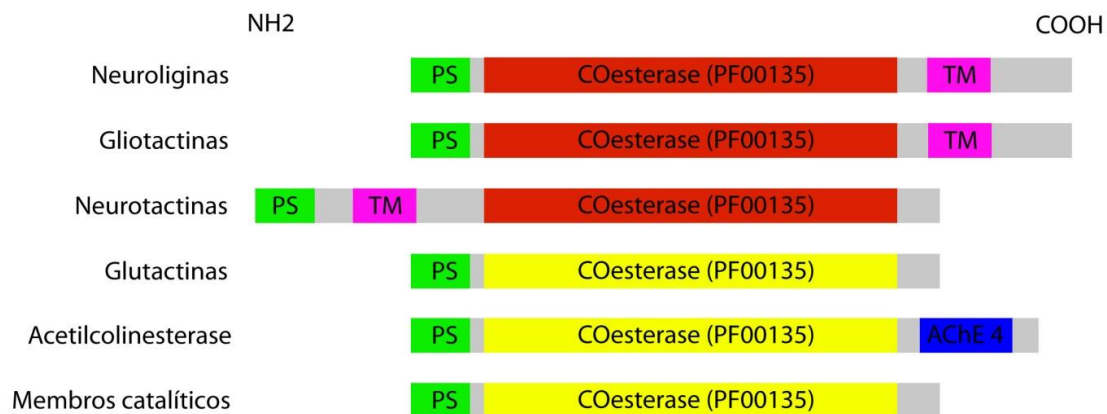


pela alternância entre fitas-beta e alfa-hélices (Figura 1.4A e B) que são estabilizadas por pontes dissulfeto estabelecidas entre resíduos de cisteína conservados ao longo da sequência proteica (Ollis et al., 1992). A estrutura formada por essa dobra se assemelha a uma fileira de costelas e funciona adequadamente para o posicionamento dos três resíduos catalíticos (tríade) dentro da cavidade enzimática (Figura 1.4C e D) (Nardini & Dijkstra, 1999). A tríade que corresponde ao sítio catalítico, característica da família, é a parte mais conservada da proteína e a ausência de algum desses três resíduos impossibilita a reação de hidrólise (Gibney et al., 1990). Tipicamente, a tríade é composta por um resíduo nucleofílico representado por uma serina ou cisteína, um resíduo ácido representado por um ácido glutâmico ou ácido aspártico e um resíduo de histidina que se apresenta conservado (Figura 1.4D) (Cygler et al., 1993). Esses três resíduos se localizam de forma espaçada ao longo da sequência proteica embora, quando a proteína assume a sua conformação nativa, os três ficam tridimensionalmente próximos para fazer a hidrólise do substrato (Cygler et al., 1993; Krejci et al., 1991).



**Figura 1.4** Visão tridimensional da estrutura cristalizada de um membro da família CCE (PDB: 2FJ0) responsável pela degradação de hormônio juvenil na espécie *Manduca sexta*. É possível observar na visão longitudinal (A) e transversal (B) as fitas-beta em amarelo e alfa-hélices em vermelho formando a estrutura conservada da dobra alfa/beta hidrolase. C) Destaque para o molde da cavidade e os resíduos catalíticos e de ligação com o substrato em relação à estrutura tridimensional. D) Visualização aproximada do centro ativo com detalhe para o molde da cavidade enzimática em branco obtido pela ferramenta online POCASA (<http://altair.sci.hokudai.ac.jp/g6/service/pocasa/>). Os aminoácidos ressaltados em vermelho situados no fundo da cavidade representam os resíduos catalíticos serina, ácido glutâmico e histidina que foram abreviados como “SER-226”, “GLU-357” e “HIS-471” respectivamente. Em verde os resíduos que constituem os sítios de ligação.

Apesar de o domínio conservado COesterase ser uma característica comum para os membros dessa família, alguns não possuem a tríade catalítica como, por exemplo, as neurotactinas, gliotactinas e neuroliginas (De la Escalera et al., 1990; Auld et al., 1995; Song et al., 1999). A ausência de um ou mais resíduos da tríade indicam que o membro não é cataliticamente funcional, o que é corroborado por trabalhos que descrevem essas proteínas. As neuroliginas, por exemplo, possuem papel importante na formação e manutenção das regiões de sinapse entre neurônios (Song et al., 1999; Biswas et al., 2008), as neurotactinas já foram associadas com processos de adesão celular (Darboux et al., 1996) e as gliotactinas têm papel fundamental no desenvolvimento e manutenção da barreira entre sistema nervoso e circulatório em insetos (Auld et al., 1995). Esses membros, com exceção das glutactinas, atuam presos à membrana citoplasmática das células neuronais e possuem regiões hidrofóbicas associadas ao ancoramento em membranas, o que não ocorre no grupo catalítico desta família (Figura 1.5). Um outro membro da família que possui domínio extra é a enzima AChE, na qual uma região na porção carboxi-terminal da proteína foi associada a formação de tetradímeros (PF08674; Figura 1.5) (Dvir et al., 2004).



**Figura 1.5** Esquema comparativo de regiões características de proteínas bem caracterizadas pertencentes a família das CCEs. Os retângulos na cor verde no início da porção N-terminal correspondem ao peptídeo sinal (PS). Os retângulos amarelos representam o domínio COesterase de membros catalíticos e os vermelhos o domínio COesterase de membros não catalíticos. Retângulos rosas representam a sequência de aminoácidos hidrofóbicos que está envolvida com o ancoramento da proteína na membrana celular (TM). O retângulo azul (AChE4) representa o sítio de tetradimerização da enzima acetilcolinesterase (PF08674) (adaptado de Oakeshott et al., 2005).

### 1.3.2 Classificação

Alguns trabalhos já foram feitos com o intuito de compreender e melhor classificar esta família gênica tão complexa (Ranson et al., 2002; Oakeshott et al., 2005; 2010). Oakeshott e colaboradores em 2005 propuseram a primeira classificação para a família das CCEs utilizando todas as proteínas homólogas encontradas nos genomas de *An. gambiae* (51) e *Drosophila melanogaster* (35). Adicionalmente, o autor utilizou outras 35 CCEs que tinham função já descrita com o intuito de embasar sua classificação. As CCEs foram organizadas em três grupos principais: Detoxificantes, Processadoras de hormônios e feromônios e Desenvolvimento neuronal. Nesses grupos foram identificados ao todo 14 clados, de acordo com seu relacionamento filogenético e evidências sobre sua possível função. Após a publicação desse trabalho e com o sequenciamento do genoma de novas espécies de insetos, outros trabalhos foram publicados utilizando a classificação de 2005 para organizar os homólogos dessa família (Claudianos et al., 2006; Strode et al., 2008; Yu et al., 2009; Tsubota & Shiotsuki, 2010a; Lee et al., 2010; Ramsey et al., 2010; Reddy et al., 2012). Em 2010, Oakeshott e colaboradores publicaram um novo trabalho organizando as três principais famílias gênicas detoxificantes (GST, CYP e CCE) para a espécie *Nasonia vitripennis* da ordem Hymenoptera. Nessa análise os autores utilizaram sequências preditas nos genomas das espécies *Apis mellifera*, *N. vitripennis* e *Tribolium castaneum*, além das utilizadas no trabalho de 2005 e identificaram 13 clados para as CCEs (Figura 1.6). Fazendo uma comparação entre as duas classificações fica claro que o aumento do número de sequências na análise causou grandes diferenças na classificação desta família proteica, sendo o grupo Detoxificante o que sofreu maior número de alterações (Oakeshott et al., 2010). Isso mostra a complexidade de se obter uma boa classificação para essa família que possui papel importante na resistência a inseticidas (Montella et al., 2012).

#### A. *Detoxificantes*

Esse grupo é constituído por proteínas catalíticas envolvidas na detoxificação de xenobióticos. Entretanto, em 2005, Oakeshott e colaboradores mostraram que esse grupo também abriga proteínas previamente associadas à resistência metabólica como, por exemplo, o membro estA2 de *Cu. pipiens* (Raymond et al., 1989). O grupo é monofilético e é constituído por três clados que foram identificados como: A – enzimas

detoxificantes de xenobióticos de himenópteros, B – alfa-esterases e enzimas similares geralmente microssomais, C – função desconhecida (Figura 1.6). Para os himenópteros a maior representação está no clado A, que contém genes envolvidos com o processo de detoxificação, sendo esta a razão do nome dado ao clado. O clado B contém as alfa-esterases de dípteros identificadas por Oakeshott e colaboradores (2005), além das de himenópteros e coleópteros. Um trabalho que antecedeu a classificação feita em 2005 descreve a organização de onze genes de alfa-esterases em *D. melanogaster* e mostra que eles estão dispostos *in tandem* e possivelmente surgiram devido à duplicação por recombinação desigual (Robin et al., 1996). No clado C existe uma representação maior de genes da espécie *T. castaneum*, da ordem Coleoptera (Oakeshott et al., 2010).

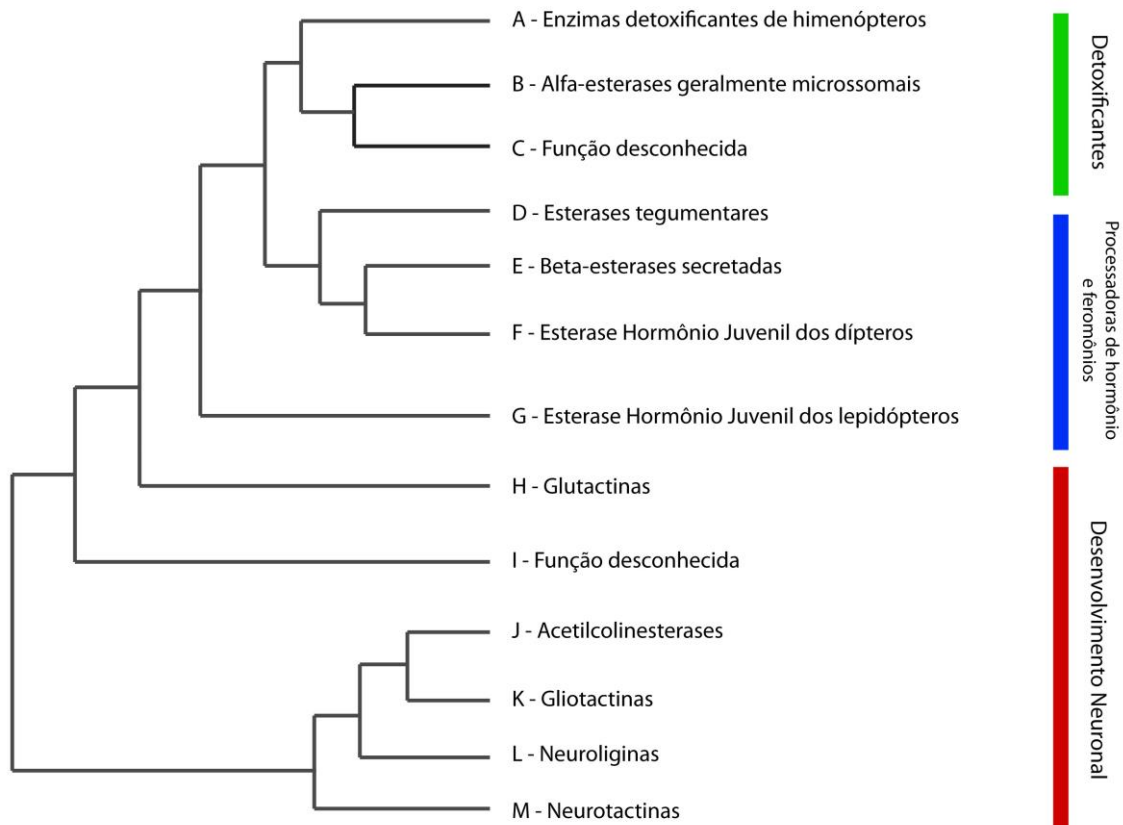
### B. *Processadoras de hormônios e feromônios*

As proteínas organizadas nesse grupo são majoritariamente catalíticas e estão envolvidas na degradação de substratos que sinalizam processos fisiológicos importantes em artrópodes como, por exemplo, o processo de ecdise (Hammock et al., 1985). Esse grupo é constituído por quatro cladogramas: D - Esterases tegumentares, E - Beta-esterases secretadas, F - Esterases de Hormônio Juvenil dos Dípteros, G – Esterases de Hormônio Juvenil dos Lepidópteros. Dois trabalhos que antecederam a classificação de 2005 mostraram indícios sobre a organização e evolução de alguns desses genes em *Drosophila*, como no caso das Esterases de Hormônio Juvenil dos Dípteros (Clado F) e Beta-esterases (Clado E). Esses trabalhos mostram que esses genes ficam organizados *in tandem* e possuem sintenia conservada para a maioria das espécies estudadas (Crone et al., 2007; Robin et al., 2009). Um maior número de genes da espécie *N. vitripennis* em relação à *Ap. mellifera* (ambas da ordem Hymenoptera) foi encontrado no clado D, o que indica uma expansão de genes relacionados com degradação de feromônio para essa espécie (três desses genes são candidatos a degradação de feromônios, Oakeshott et al., 2010). Por outro lado, a ausência de genes de *An. gambiae* e *Ae. aegypti* indica que eles podem ter sido perdidos em mosquitos. Além disso, no clado E foram encontrados membros candidatos à degradação de hormônio juvenil na espécie *N. vitripennis*. Ainda nesse clado, é importante citar a presença das proteínas E4 e FE4 que já foram associadas à resistência metabólica a organofosforados na espécie *My. persicae* (Field et al., 1998; 1999; Field & Blackman, 2003). O clado F é um dos cladogramas que possui sua função bem estabelecida na degradação de hormônio juvenil no que diz respeito à

ordem Diptera (Oakeshott et al., 2005; Bai et al., 2007; Crone et al., 2007). Em 2010, Oakeshott e colaboradores ressaltaram a existência de um membro candidato a Esterase de hormônio juvenil (JHE, do inglês *Juvenile Hormone Esterase*) na espécie *T. castaneum* devido à presença do motivo “GQSAG” próximo a serina nucleofílica e pela relação filogenética com outras sequências já caracterizadas como JHE em outras espécies da ordem Coleoptera. No mesmo ano, Tsubota e colaboradores caracterizaram esse gene como sendo o único que realiza esta função em *T. castaneum* (Tsubota et al., 2010). O clado G contém genes de JHE da ordem Lepidoptera e também contém as CCEs de *An. gambiae* sem função conhecida.

### C. *Desenvolvimento neuronal*

Esse grupo abriga proteínas com sequência proteica conservada e características diferentes dos dois grupos anteriores, como a ausência da tríade catalítica em muitos casos. Apenas os genes da enzima AChE, que atua na fenda sináptica hidrolisando o neurotransmissor acetilcolina (Futuko, 1990), sempre possuem a tríade completa. Essa enzima, alvo de OPs e CXs, quando sofre mutações pontuais na sequência codificante impedindo a ligação do inseticida, é associada à resistência por insensibilidade do sítio alvo. Outros membros são bem estudados como as neuroliginas que não são catalíticas e participam do desenvolvimento do SNC. Em *Drosophila*, as enzimas neuroliginas 1 e 2 foram descritas por participar do desenvolvimento, crescimento e diferenciação pós-sináptica de junções neuromusculares (Banovic et al., 2010; Sun et al., 2011). Esse grupo é constituído de seis clados: H – Glutactinas, I – Função desconhecida, J – Acetilcolinesterases, K – Gliotactinas, L – Neuroliginas e M – Neurotactinas. Na classificação de 2005, as glutactinas (clado H) faziam parte do grupo de processadoras de hormônio e feromônio e na classificação de 2010 foram reposicionadas nesse grupo (Oakeshott et al., 2010). Um dado interessante é que o número de genes em cada clado pode variar por espécie. Por exemplo, no clado M dípteros e coleópteros possuem dois genes e as duas espécies de himenópteros só possuem um (Oakeshott et al., 2010).

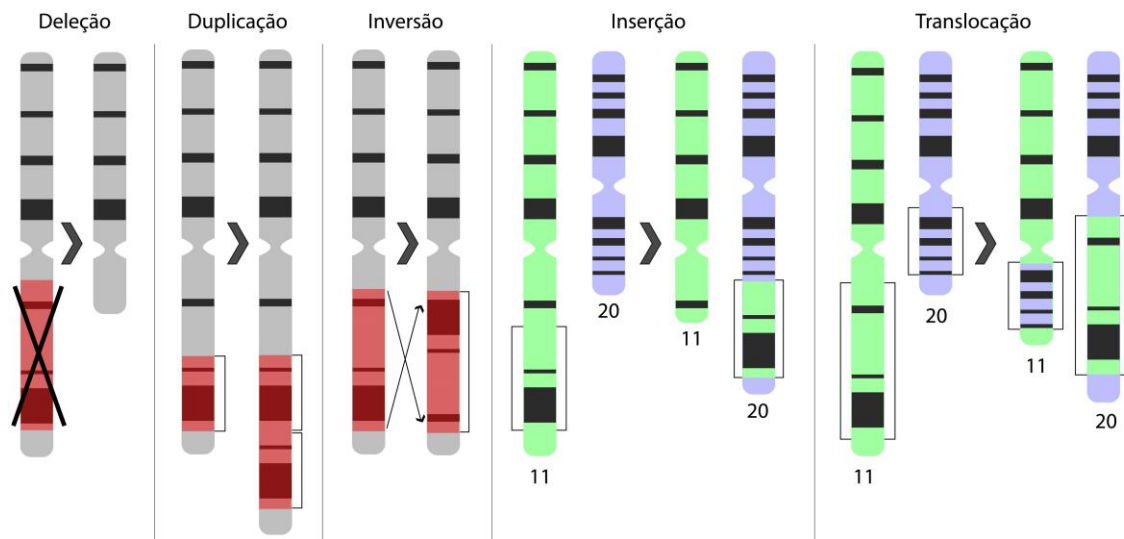


**Figura 1.6** Cladograma esquemático representando a organização dos 13 clados da família CCE que foram identificados por Oakeshott (2005). As barras laterais delimitam os clados que estão incluídos nos três grandes grupos descritos pelo autor, em verde – esterases envolvidas em processos fisiológicos de detoxificação, Azul – esterases envolvidas no processamento de hormônios e feromônios e vermelho – esterases envolvidas no desenvolvimento e funcionamento do sistema nervoso central (Adaptado de Montella et al., 2012).

#### 1.4 Evolução de famílias gênicas

Uma família gênica é um grupo de genes homólogos que compartilham características similares e evoluíram a partir de um ancestral comum através de duplicações. Em espécies próximas essas famílias podem apresentar composição muito similar, o que pode não acontecer entre espécies que divergiram há mais tempo. Isso ocorre pois esses genes estão sujeitos à dinâmica natural do genoma que, por sua vez, está sujeito a eventos de ganho e perda de conteúdo, como duplicações, deleções, translocações, inversões e inserções (Figura 1.7). Além disso, as características de cada gene serão importantes para a sua fixação ou perda em uma população (Lynch & Katju, 2004; Hahn et al., 2007). Embora já houvesse relatos sobre eventos de duplicação gênica desde 1930, Ohno foi um dos primeiros autores a tentar explicar os mecanismos

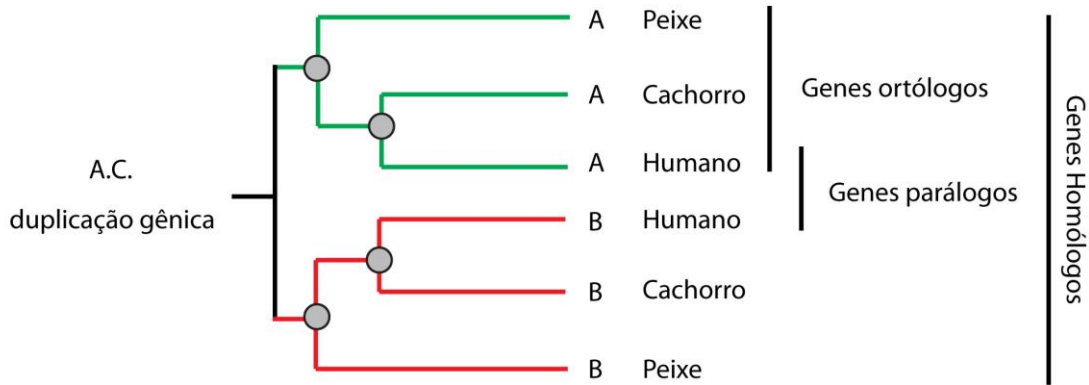
de duplicação e destacar sua importância na contribuição da diversidade e evolução das espécies (Zhang, 2003). As duplicações variam quanto ao seu tamanho e região do genoma, podendo englobar desde uma parte de um gene, um segmento com vários genes (duplicação segmentar; Stankiewicz et al., 2004; Sharp et al., 2005), cromossomos inteiros (aneuploidia; Hassold et al., 2007) ou até mesmo genomas completos (ploidias; Kellis et al., 2004).



**Figura 1.7** Esquema mostrando alguns eventos de rearranjo cromossômico. Em cinza está representado um cromossomo hipotético. Em vermelho estão representadas as regiões antes e depois de cada evento (deleção, duplicação e inversão). Em verde e azul estão representados cromossomos hipotéticos numerados como 11 e 20 respectivamente. As regiões delimitadas por colchetes estão envolvidas nos eventos de inserção e translocação entre os dois cromossomos.

Após um evento de duplicação, os genes ancestrais e suas cópias são chamados de genes parálogos enquanto genes que surgiram a partir de um evento de especiação são chamados de ortólogos. Quando comparamos famílias gênicas em genomas de espécies diferentes podemos encontrar genes ortólogos e parálogos, ilustrado na Figura 1.8 (Tatusov et al., 1997). Ao longo da evolução esses genes sofrem modificações nas diferentes espécies, tais como outras duplicações e deleções, o que dificulta a inferência de paralogia e ortologia para os diversos membros de uma família gênica (Durand & Hoberman, 2006). A inferência do relacionamento filogenético entre todos esses genes é a melhor maneira de tentar compreender como evoluíram.





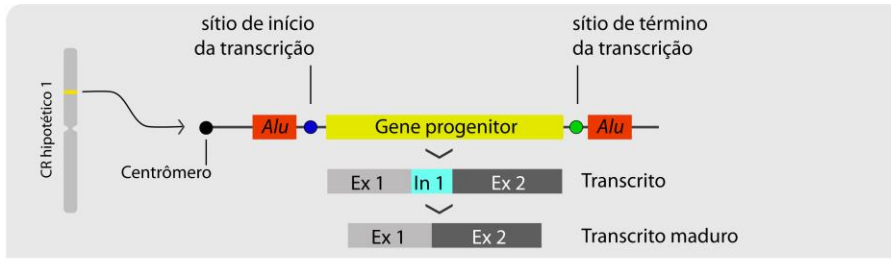
**Figura 1.8** Cladograma esquemático mostrando as relações entre dois genes hipotéticos (A e B) que foram duplicados no ancestral comum de três táxons (Humano, Cachorro e Peixe). As barras laterais agrupam os genes hipotéticos que de acordo com a relação mostrada na árvore esquemática são considerados ortólogos, parálogos e homólogos. Os círculos cinzas representam os eventos de especiação.

Nas diferentes espécies, as famílias gênicas podem apresentar variações no número de cópias, mostrando que os membros dessa família seguem caminhos evolutivos distintos. A diferença de nicho ecológico e outras características intrínsecas das espécies fazem com que estas, e seus genes, estejam sob pressões seletivas diferentes. Em muitos casos a mesma família gênica pode ter composições completamente distintas entre espécies próximas. Por exemplo, em uma comparação feita nos genomas de 12 espécies de *Drosophila* viu-se que cerca de 40% das famílias gênicas diferem em número de cópias (Hahn et al., 2007). Essas diferenças são ainda mais proeminentes para espécies que divergiram há mais tempo, como é descrito para a família gênica dos receptores nucleares de hormônio em *Homo sapiens*, *Caenorhabditis elegans* e *D. melanogaster* (International Human Genome Sequencing Consortium, 2001).

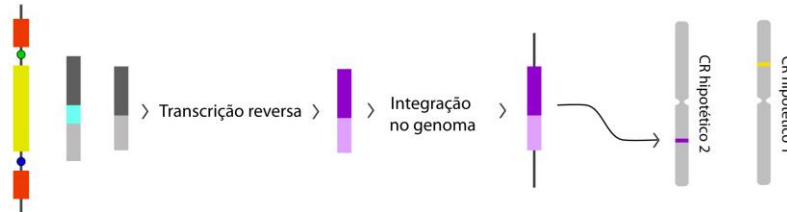
Alguns mecanismos de duplicação gênica já foram descritos, mas a importância de cada um deles para a evolução das diferentes famílias gênicas ainda é controversa (Nei & Hooney, 2005). Cada mecanismo de duplicação deixa assinaturas na cópia final e, em muitos casos, podemos inferir por qual mecanismo a cópia foi gerada. Na transcrição reversa, a duplicação ocorre através da retrotranscrição de uma fita de RNA transcrito em uma fita simples de DNA que depois é inserida aleatoriamente no genoma. A nova cópia é desprovida tanto da estrutura gênica exon-ítron quanto do promotor que são encontrados originalmente no gene copiado. Além disso, também é possível observar traços da sequência referente à cauda poliadenilada comum em transcritos

maduros (Zhang, 2003) (Figura 1.9A). Um trabalho realizado em humanos mostra que cerca de 70% dos pseudogenes foram originados por retrotransposição enquanto o resto surgiu por recombinação (Torrents et al., 2003).

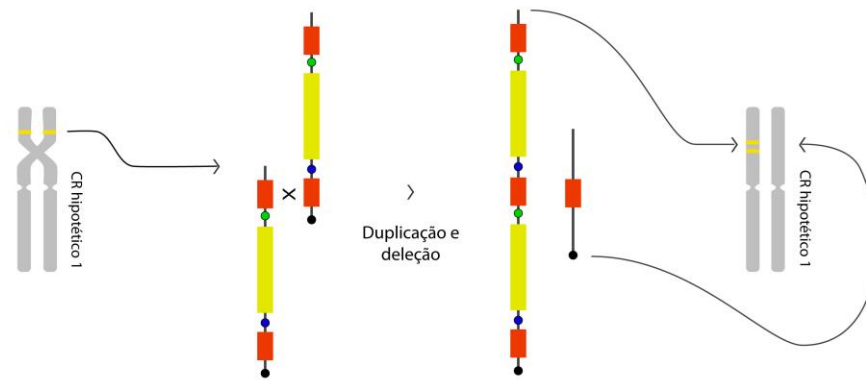
A recombinação desigual acontece durante a replicação do DNA através do pareamento de cromossomos homólogos e geralmente as cópias são dispostas em sequência (Figura 1.9B) (Zhang, 2003). Em alguns casos isso pode ocorrer diversas vezes formando um grupo de genes em uma região do genoma como é visto, por exemplo, em *Drosophila* para as beta-esterases (Robin et al., 2009). Alguns trabalhos mostram evidências sobre a existência de regiões do genoma mais susceptíveis para o acontecimento de duplicação gênica do que outras, como no caso das sequências *Alu*, que representam cerca de 10% do genoma humano (Bailey et al., 2003). Erros de escorregamento da DNA polimerase durante a replicação do DNA genômico também são comuns e podem causar o surgimento de pequenas regiões duplicadas. Além disso, durante a replicação, a presença de segmentos de DNA muito similares, muito comum quando outras duplicações já estão presentes *in tandem*, pode favorecer a formação de alças que posteriormente aumentam ainda mais o número de cópias do segmento de DNA (Figura 1.9C). Algumas doenças humanas são causadas por duplicações gênicas como, por exemplo, a doença de *Charcot-Marie-Tooth*. Esta doença é causada pelo aumento de dosagem do gene *myelin protein zero* que ocorre devido a eventos de duplicação do mesmo (Høyer et al., 2011; Maeda et al., 2012). Nos eventos de recombinação desigual, ao mesmo tempo que ocorre a duplicação de uma região o mesmo segmento pode ser deletado no cromossomo homólogo (Figura 1.9B). Deleções também podem ser mediadas pelo pareamento de dois genes homólogos vizinhos, resultando na formação de uma alça que leva à possível deleção de um segmento do genoma (Figura 1.9D). Em alguns casos, a deleção de genes importantes pode ocasionar o surgimento de fenótipos menos adaptados ao ambiente como em algumas doenças genéticas (Emanuel & Shaikh, 2001). A síndrome de ictiose, por exemplo, é causada pela deleção do gene *esteroide sulfatase* presente no cromossomo X (Mohandas et al., 1979; Bonifas et al., 1987). Efeitos deletérios também podem ser provocados por inversões como caracterizado para a Hemofilia A, onde modificações na estrutura gênica do fator VIII devido à inversão de uma região genômica prejudicam a integridade do gene, provocando o desenvolvimento da doença (Emanuel & Shaikh, 2001).



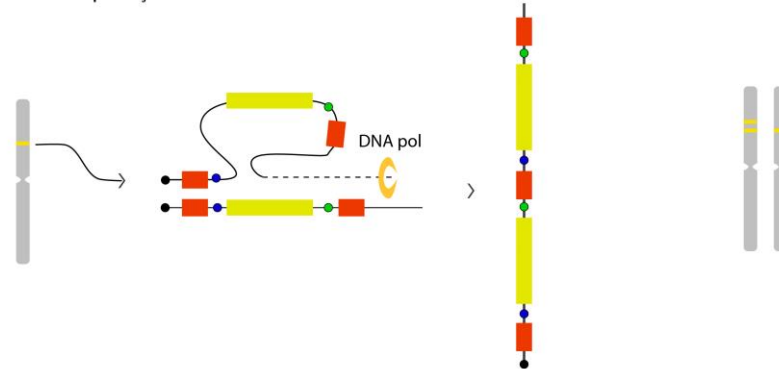
A Transcrição reversa e integração no genoma



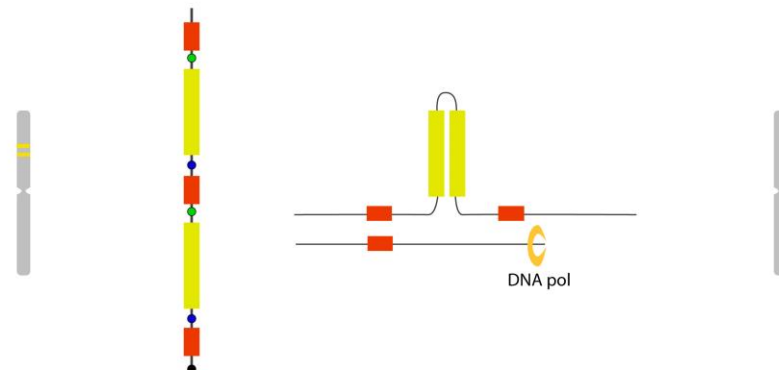
B Recombinação de desigual



C Erro de replicação



D Deleção por pareamento de parálogos



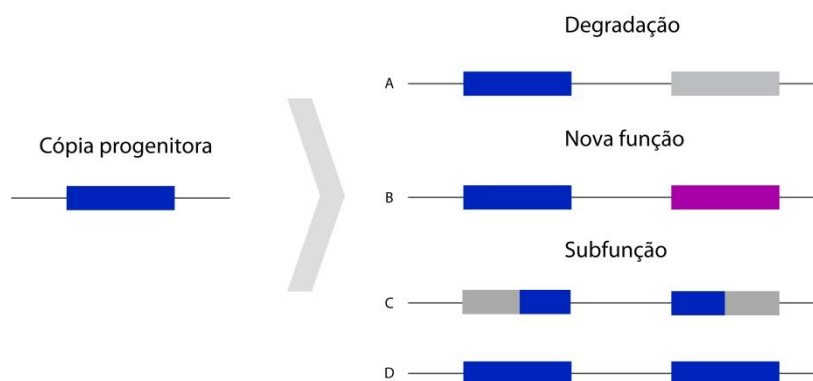
**Figura 1.9** Esquema de diferentes mecanismos de duplicação gênica. A – retrotransposição; B – duplicação e deleção de um segmento por recombinação gênica; C – duplicação de um segmento *in tandem* devido a um erro de replicação e D – deleção de um segmento mediada pela formação de uma alça entre dois genes parálogos. A imagem com fundo cinza representa um segmento genômico hipotético incluindo a estrutura de um gene e duas regiões *Alu* (adaptado de Hurler, 2004 e Emanuel & Shaikh, 2001).

Após a duplicação do gene progenitor, a fixação de sua cópia ocorre pelo aumento de sua frequência na população até um ponto em que faça parte permanentemente do *pool* gênico deste grupo. Depois da fixação, quatro possíveis caminhos podem ser seguidos pela cópia durante o processo evolutivo. Em um deles, a cópia pode ser degradada pelo acúmulo de mutações até que as características comuns com a cópia progenitora, tais como estrutura gênica e conservação da sequência, não sejam mais detectadas (Figura 1.10A). Em outro panorama, a cópia pode se manter funcional, com a mesma função ancestral ou desenvolver uma nova função (neofuncionalização; Figura 1.10B), ou ainda compartilhar a função ancestral com o gene progenitor (subfuncionalização; Figura 1.10C e D) (Hurler, 2004; Zhang, 2003; Conant & Wolfe, 2008). Ainda são poucos os exemplos que mostram claramente o trajeto de genes que evoluíram por alguma dessas maneiras, mas alguns estudos já apontam indícios consistentes.

Na neofuncionalização, tanto a cópia quanto o gene progenitor podem sofrer modificações e assim assumir novas funções (Conant & Wolfe, 2008). Um exemplo que aponta indícios do surgimento de uma nova função foi visto em peixes que habitam áreas frias e possuem uma proteína anticongelante essencial para sua sobrevivência. Um estudo mostrou que essa proteína pode ter sido originada através da duplicação do gene que codifica a enzima ácido salicílico sintetase que, em sua forma original, também presente no peixe, possui função completamente diferente (Deng et al., 2010). No caso da subfuncionalização, tanto a cópia quanto o gene ancestral permanecem com a mesma função inicialmente, mas alterações podem levar cada um deles a assumir parte da função ancestral. Por exemplo, dois genes envolvidos com a via metabólica de degradação da galactose (GAL1 e GAL3) possuem evidências de terem evoluído de um ancestral comum em *Saccharomyces cerevisiae*. GAL1 atua fosforilando moléculas de galactose transformando-as em galactose-1-fosfato enquanto GAL3 atua como co-indutor para a expressão dos genes da via. Entretanto, a espécie *Kluyveromyces lactis* não possui o gene GAL3, indicando que GAL1 deve exercer as duas funções. O mesmo

é visto para mutantes da espécie *S. cerevisiae* que não expressam GAL3 (Meyer et al., 1991).

Existe ainda muita controvérsia a respeito da manutenção da nova cópia por seleção quando ela mantém a mesma função do gene progenitor (Swanson, 2003). Isto porque em muitos genes o desbalanço da dosagem por si só já seria deletério e seria eliminado por seleção natural (Qian & Zhang, 2008; Birchler & Veitia, 2012). No entanto, em diversos casos, a duplicação da quantidade de transcrito pode ser favorável, como, por exemplo, no caso de genes relacionados à detoxificação de xenobióticos. Um bom exemplo de resposta adaptativa à pressão de seleção a inseticidas é o do gene de carboxilesterase em espécies do mosquito *Culex* que, em indivíduos resistentes, está amplificado no genoma (Yan et al., 2012). Além disso, quando a duplicação da dosagem não é deletéria, a presença de uma segunda cópia pode servir como um tamponamento contra a presença de alelos nulos (Rabinow et al., 1991). Foi observado que processos celulares mais básicos, como regulação da transcrição, transdução de sinais, genes de desenvolvimento, entre outros, acumulam menos duplicações, enquanto que os genes relacionados com o metabolismo e resposta a stress, especialmente stress ambiental, acumularam mais cópias durante a evolução (Vogel & Chothia, 2006). Além disso, em muitos casos, apesar de manter as funções originais do gene, existem diferenças importantes no padrão de expressão da nova cópia (Huminiecki & Wolfe, 2004). No caso dos genes *kindlin*, que têm função de adesão e são importantes no processo de migração celular, o padrão de expressão dos três genes parálogos em vertebrados é diferente nos tecidos avaliados. Um deles inclusive é expresso apenas em leucócitos, mostrando a especialização no padrão de expressão, possivelmente ocasionado pela redundância funcional após a duplicação (Khan et al., 2011).



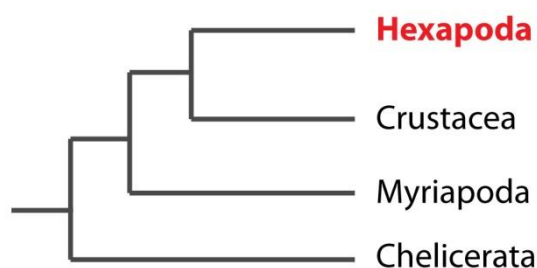
**Figura 1.10** Esquema mostrando os tipos de consequências que podem ocorrer para as cópias duplicadas: A – degradação (ou pseudogenização); B - desenvolvimento de uma

nova função (neofuncionalização); C – cada cópia passa a assumir uma parte da função ancestral (subfuncionalização) e D – ambas as cópias compartilham a função ancestral (adaptado de Hurles, 2004).

Ao longo da evolução, genes homólogos podem divergir ao ponto de assumirem funções diferentes e genes que não compartilham um ancestral comum podem assumir funções similares, por convergência evolutiva (Zhang, 2003; Hurles, 2004). No entanto, assinaturas nas sequências desses genes (ou sequência de aminoácidos de suas proteínas) e a avaliação de seus padrões evolutivos em árvores filogenéticas podem nos dar a direção quanto à origem de determinada função.

## 1.5 Taxonomia de Artrópodes

O filo Arthropoda é conhecido por ser o maior do Reino animal e é responsável por grande parte da biodiversidade conhecida no planeta (Edgecombe, 2010). As espécies que fazem parte desse filo compartilham características que sustentam o monofilismo do grupo como, por exemplo, o processo de ecdise, ou muda, do exoesqueleto. Além dos caracteres morfológicos, a monofilia do filo também é suportada por caracteres moleculares, porém, para alguns ramos internos ainda existe controvérsia (Edgecombe, 2010). Atualmente esse filo está dividido em quatro subfilos: Chelicerata, Myriapoda, Crustacea e Hexapoda (Edgecombe, 2010) (Figura 1.11). O subfilo Chelicerata abriga as classes Xiphosura (ex: caranguejo-ferradura) e Arachnida (ex: aranhas, escorpiões e carrapatos) e apresenta fortes indícios de ser monofilético. O subfilo Myriapoda abriga as classes Chilopoda, Diplopoda, Symphyla e Pauropoda. Os subfilos Crustacea e Hexapoda formam um grupo monofilético bem suportado por caracteres moleculares chamado de Pancrustacea (Edgecombe, 2010; Trautwein et al., 2012). Hexapoda possui duas classes: Entognatha e Insecta, sendo a última detentora de grande parte da diversidade de espécies do grupo (Figura 1.12).

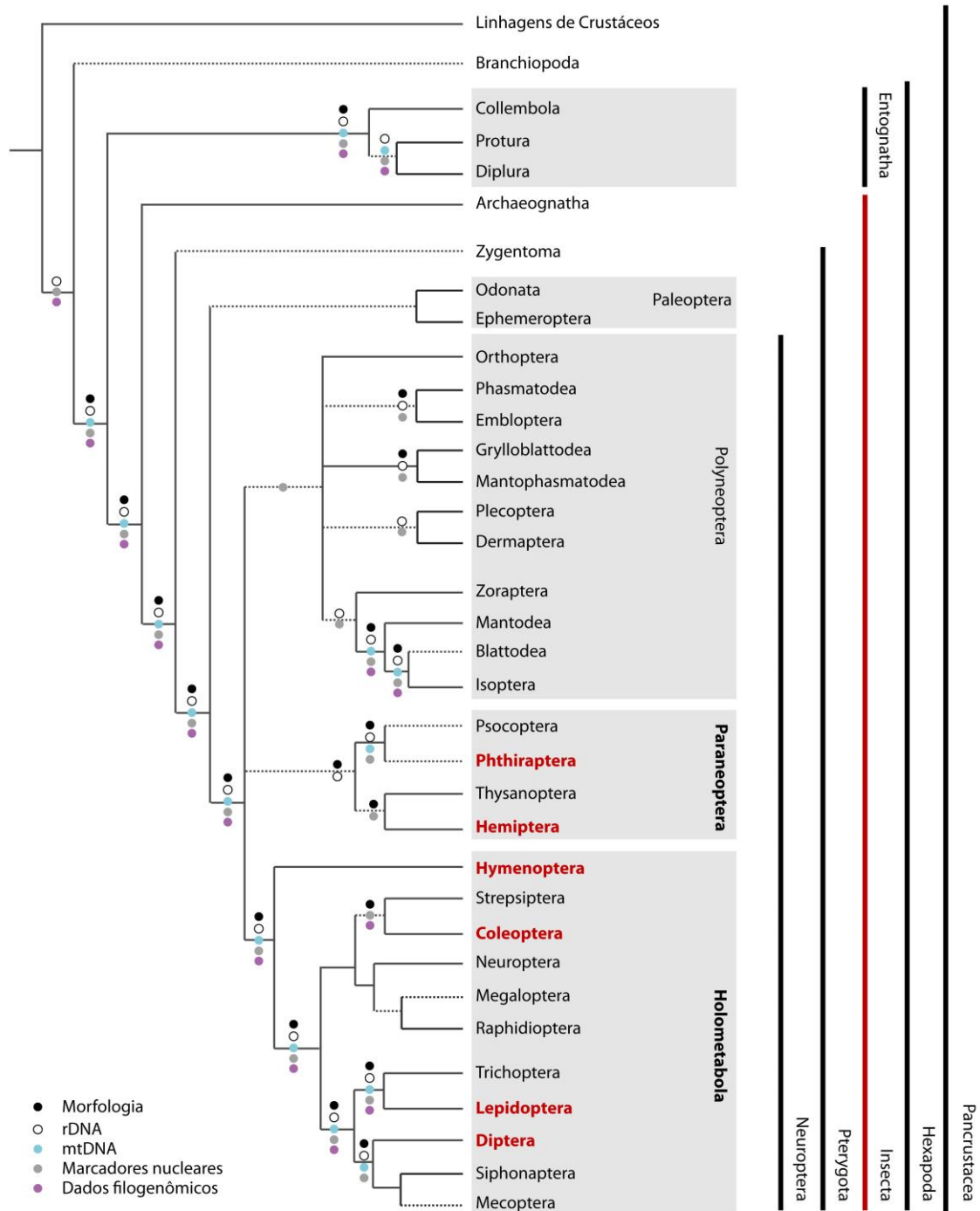


**Figura 1.11** Esquema resumido dos subfilos que constituem o filo Arthropoda. Em vermelho está destacado a subfilo Hexapoda (Adaptado de Edgecome, 2010).

A diversificação da classe Insecta ocorreu no permiano e jurássico, quando novos padrões morfológicos e moleculares surgiram em um curto período de tempo (Whitfield & Kjer, 2008). Para análises filogenéticas esse tipo de panorama torna difícil a compreensão dos padrões de divergência em ramos mais antigos devido à rápida diversificação (Whitfield & Kjer, 2008). O uso de dados moleculares ajudou a resolver algumas questões taxonômicas como, por exemplo, a relação entre as ordens do grupo Holometabola (Trautwein et al., 2012). No entanto, mesmo dados moleculares são fortemente influenciados pela rápida diversificação, já que a única informação a que temos acesso é a dos genes de organismos atuais que foram influenciados pelo curto período de evolução conjunta e longo período posterior de evolução das diferentes linhagens.

No presente trabalho foram utilizados dados de 33 espécies de seis ordens taxonômicas da classe Insecta. Para o grupo dos Holometabolas, estão representadas as ordens Diptera, Lepidoptera, Coleoptera e Hymenoptera e para os Paraneoptera, as ordens Hemiptera e Phthiraptera. Essas ordens estão ressaltadas em vermelho na Figura 1.12 onde está representada a taxonomia atual dos hexápodos (adaptada de Trautwein et al., 2012). O monofiletismo de holometábolos e de suas ordens é bem suportado tanto por caracteres moleculares (dados filogenômicos, rDNA, mDNA e marcadores nucleares) quanto caracteres morfológicos. Entretanto, para os Paraneoptera ainda existe discordância entre os caracteres avaliados (principalmente moleculares) e por isso, o posicionamento taxonômico desse grupo ainda não foi bem estabelecido. Além desses grupos taxonômicos, também foram incluídos duas espécies de artrópodes da classe

Arachnida que faz parte do subfilo Chelicerata e uma espécie do filo Nematoda, externo aos Artrópodes.



**Figura 1.12** Representação taxonômica atual da classe Insecta. Linhas pontilhadas indicam pouco suporte para a monofilia do grupo. Os círculos coloridos representam o tipo de dado que suporta as relações entre os nós (ver legenda no pé da figura). As ordens que possuem espécies incluídas no presente trabalho estão destacadas em vermelho (Adaptada de Trautwein et al., 2012).



## 1.6 Bioinformática

A bioinformática utiliza conhecimentos das áreas de ciência da computação e matemática para organizar e analisar informações biológicas em larga escala para melhor compreender problemas biológicos (Luscombe et al., 2001; Goodman, 2002). Com o início do projeto de sequenciamento do genoma humano na década de 1980, a bioinformática veio disponibilizar métodos automáticos capazes de lidar com o grande volume de dados gerados (Goodman, 2002). Em 1997 foi iniciada a criação de laboratórios dedicados a essa nova área no Brasil focados em estudar inicialmente organismos de interesse comercial. A bactéria *Xillela fastidiosa*, que ataca plantações de laranja causando grandes prejuízos na economia, por exemplo, foi a primeira a ter o genoma mapeado em território nacional (Neshich, 2007). Posteriormente, outros organismos de interesse médico e econômico começaram a ser estudados, permitindo o desenvolvimento e avanço da área no país (Neshich, 2007). Dados atuais publicados pelo *National Center for Biotechnology Information* (NCBI) mostram que no mundo inteiro mais de três mil genomas de eucariotos já foram ou estão em processo de ser sequenciados (NCBI, 2013). Em espécies de insetos, 15 genomas já foram completamente montados e 261 estão em andamento (NCBI, 2013). Em 2011, o projeto *The 5.000 insect genome project* aumenta a expectativa de crescimento desse número e destaca a importância da bioinformática para o sucesso do projeto (Levine, 2011).

Após o sequenciamento de um genoma, por exemplo, o processo de montagem dos fragmentos, predição e anotação demandam conhecimentos tanto da área biológica quanto da informática para que os dados possam ser organizados de forma coerente (Goodman, 2002). Além disso, apenas o mapeamento de um genoma isoladamente não consegue desvendar problemas biológicos relevantes. Para isso são utilizados processos automatizados para analisar dados de fontes importantes de informação como, por exemplo, os de expressão gênica para alcançar uma melhor compreensão dos sistemas biológicos. A criação de algoritmos para automatizar o processamento dessa informação é crucial frente ao constante aumento do volume de dados, porém, as etapas de decisão e execução devem ser baseadas no conhecimento biológico prévio (Goodman, 2002). A interação entre áreas é uma das principais características da bioinformática tornando possível a compreensão de problemas biológicos complexos através da manipulação e análise de dados em larga escala (Luscombe et al., 2001).

## 2. Objetivos

### 2.1 Objetivo geral

Identificar proteínas homólogas da família gênica das carboxilesterases (CCE) em 33 genomas de insetos anotados e classificá-las de acordo com a evolução dos diferentes membros.

### 2.2 Objetivos específicos

- Identificar proteínas homólogas da família CCE em 33 genomas já publicados de espécies da classe Insecta, 2 de Arachnida e 1 de Nematódeo;
- Predizer a localização celular dos membros identificados;
- Organizar as proteínas através de análises filogenéticas;
- Classificá-las de acordo com seu relacionamento filogenético, predição da localização celular e informações disponíveis na literatura e em bases de dados públicas;
- Compreender a evolução da família CCE.

### 3. Material e métodos

#### 3.1 Bases de dados

Os bancos de dados de proteínas preditas para os genomas das 36 espécies selecionadas foram baixados dos seguintes sítios públicos: VectorBase (<http://www.vectorbase.org/>), FlyBase (<http://flybase.org/>), Hymenoptera Genome Database (<http://hymenopteragenome.org/>), BeetleBase (<http://beetlebase.org/>), AphidBase (<http://www.aphidbase.com/>), Silkworm Genome Database (<http://silkworm.genomics.org.cn/>), WormBase (<http://www.wormbase.org/>) e NCBI (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) (Tabela 3.2). O banco de proteínas anotadas manualmente (UniProt) do Swiss-Prot (<http://www.uniprot.org/>) foi baixado e formatado para ser utilizado em buscas realizadas com a ferramenta BLAST, localmente.

**Tabela 3.2** Lista de genomas com predição gênica utilizados neste trabalho. Categoria taxonômica, nome da espécie, endereço de onde foi obtido o banco de dados, versão usada, tamanho do genoma (Mb) e número de gene preditos. Em alguns casos não existe uma predição final (-).

Ordem	Família	Espécie	FTP	Versão	Genoma	Genes
Diptera	Drosophilidae	<i>Drosophila ananassae</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	176	15,276
	Drosophilidae	<i>Drosophila erecta</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	134	15,324
	Drosophilidae	<i>Drosophila grimshawi</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	138	15,270
	Drosophilidae	<i>Drosophila melanogaster</i>	<a href="http://flybase.org/">http://flybase.org/</a>	5.44	118	13,703
	Drosophilidae	<i>Drosophila mojavensis</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	161	14,849
	Drosophilidae	<i>Drosophila persimilis</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	138	17,325
	Drosophilidae	<i>Drosophila pseudoobscura</i>	<a href="http://flybase.org/">http://flybase.org/</a>	2.27	127	16,363
	Drosophilidae	<i>Drosophila sechellia</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	115	16,884
	Drosophilidae	<i>Drosophila simulans</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.4	111	15,983
	Drosophilidae	<i>Drosophila virilis</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.2	172	14,680
	Drosophilidae	<i>Drosophila willistoni</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	187	15,816
	Drosophilidae	<i>Drosophila yakuba</i>	<a href="http://flybase.org/">http://flybase.org/</a>	1.3	127	16,884
	Culicidae	<i>Aedes aegypti</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	1.2	1376	15,419
	Culicidae	<i>Anopheles gambiae</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	3.6	278	13,683
	Culicidae	<i>Culex quinquefasciatus</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	1.2	579	18,883
Hymenoptera	Apidae	<i>Apis mellifera</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	1.1	236	10,157
	Apidae	<i>Apis florea</i>	<a href="ftp://ftp.ncbi.nlm.nih.gov/genomes/">ftp://ftp.ncbi.nlm.nih.gov/genomes/</a>	1.0	-	-
	Apidae	<i>Bombus impatiens</i>	<a href="ftp://ftp.ncbi.nlm.nih.gov/genomes/">ftp://ftp.ncbi.nlm.nih.gov/genomes/</a>	2.0	-	-
	Apidae	<i>Bombus terrestris</i>	<a href="ftp://ftp.ncbi.nlm.nih.gov/genomes/">ftp://ftp.ncbi.nlm.nih.gov/genomes/</a>	1.0	-	-
	Megachilidae	<i>Megachile rotundata</i>	<a href="ftp://ftp.ncbi.nlm.nih.gov/genomes/">ftp://ftp.ncbi.nlm.nih.gov/genomes/</a>	1.0	-	-
	Pteromalidae	<i>Nasonia vitripennis</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	1.2	295	18,000
	Formicidae	<i>Acromyrmex echinaior</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	3.8	300	17,278
	Formicidae	<i>Atta cephalotes</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	1.2	290	18,153
	Formicidae	<i>Camponotus floridanus</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	3.3	240	-
	Formicidae	<i>Herpegnathos saltator</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	3.3	330	-
	Formicidae	<i>Linepithema humile</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	1.2	251	-
	Formicidae	<i>Pogonomyrmex barbatus</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	1.2	250	17,177
Formicidae	<i>Solenopsis invicta</i>	<a href="http://hymenopteragenome.org/">http://hymenopteragenome.org/</a>	2.2.3	484	16,569	
Hemiptera	Aphididae	<i>Acyrtosiphon pisum</i>	<a href="http://www.aphidbase.com/">http://www.aphidbase.com/</a>	2.1	464	-
	Reduviidae	<i>Rodhnius prolixus</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	1.0	-	-
Coleoptera	Tenebrionidae	<i>Tribolium castaneum</i>	<a href="http://beetlebase.org/">http://beetlebase.org/</a>	3.0	204	16,404
Lepidoptera	Bombycidae	<i>Bombyx mori</i>	<a href="http://www.silkdb.org/">http://www.silkdb.org/</a>	glean	432	14,623
Phthiraptera	pediculidae	<i>Pediculus humanus</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	1.2	108	10,773
Ixodida	Ixodidae	<i>Ixodes scapularis</i>	<a href="http://www.vectorbase.org/">http://www.vectorbase.org/</a>	1.1	-	-
Mesostigmata	Phytoseiidae	<i>Metaseiulus occidentalis</i>	<a href="ftp://ftp.ncbi.nlm.nih.gov/genomes/">ftp://ftp.ncbi.nlm.nih.gov/genomes/</a>	1.0	88	-
Rhabditida	Rhabditidae	<i>Caenorhabditis elegans</i>	<a href="http://www.wormbase.org/">http://www.wormbase.org/</a>	8.1	100	19,735

## 3.2 Busca e identificação das sequências proteicas

A busca e identificação das sequências proteicas foi realizada usando o *pipeline* FAT (Functional Analysis Tool; Seabra-Junior et al., 2011), que utiliza os programas HMMER (Finn et al., 2011) e BLAST (Altschul et al., 1990) localmente. A busca pelos membros da família Carboxilesterase (CCE) foi realizada utilizando o domínio Pfam das COesterases (PF00135) disponível na base de dados do *Protein family database* (Pfam) (Punta et al., 2011) pelo programa HMMER. O programa BLAST foi usado para buscas utilizando as sequências obtidas como *query* contra o banco de dados UniProt (Swissprot) com o intuito de inferir similaridade e assim identificá-las preliminarmente (Figura 3.13). O *pipeline* inicia sua execução selecionando as proteínas através do programa HMMsearch com um perfil de domínios de interesse configurado. Os domínios conservados no banco Pfam são representados como *hidden markov models* (HMM), modelos estatísticos que, de maneira simplista, representam a probabilidade de se encontrar um determinado aminoácido em uma posição específica da sequência. Eles são construídos a partir de uma região conservada de um alinhamento múltiplo de proteínas homólogas conhecidas e bem caracterizadas (*seed*) que depois é enriquecido automaticamente por comparação com sequências disponíveis no banco SwissProt que é maior e mais diverso. Este HMM ainda poderá ser manualmente encurtado caso suas extremidades tragam muitos resultados falso positivos, o que não foi o caso nesta dissertação. O HMM possui informação de uma região conservada entre diversas proteínas homólogas, e por isso os algoritmos de busca por similaridade que o utilizam como, por exemplo, o HMMsearch e o HMMscan são considerados mais sensíveis na busca por homólogos distantes que o BLAST.



**Figura 3.13** Esquema do *pipeline* usado para as buscas nos genomas como descrito para o programa FAT (Seabra-Junior et al., 2011). Os retângulos cinza e verde simbolizam os dados de entrada para os programas que estão representados nos quadrados azuis no fluxo de dados.

### 3.3 *Pipeline* de processamento de informação

Devido ao grande número de sequências a serem analisadas, um *pipeline* secundário foi criado para processar de forma automática os resultados obtidos através das ferramentas *CD-search Batch* (disponível no sítio <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) (Marchler-Bauer et al., 2009), TargetP 1.1 (Emanuelsson et al., 2007) e TMHMM (Krogh et al., 2001) (Figura 3.14). A ferramenta *CD-search Batch* identifica os aminoácidos conservados dos domínios existentes na proteína sendo analisada. No caso das CCEs, queremos identificar os aminoácidos importantes para a função catalítica como, por exemplo, os que compõem o sítio catalítico e a cavidade de ligação com o substrato. O TargetP que é direcionado para prever a possível localização celular de proteínas levando em consideração a sequência do peptídeo sinal, foi utilizado para distinguir CCEs citosólicas, mitocondriais e secretadas. O programa TMHMM identifica sequências hidrofóbicas ao longo da proteína e por isso foi utilizado para caracterizar as hélices transmembranares presentes nos membros do grupo chamado de neuronal das CCEs (Oakeshott et al., 2005) como, por exemplo, as neurotactinas (De la Escalera et al., 1990). Todas essas informações foram extraídas e codificadas de forma padronizada nos cabeçalhos dos arquivos das sequências, em formato FASTA, para melhorar a visualização da árvore filogenética. Também foi incluído no *pipeline* um passo de exclusão de sequências de má qualidade que prejudicavam as análises filogenéticas devido à falta de sobreposição no alinhamento múltiplo.

Em uma análise do resultado do programa HMMsearch foi observado que havia uma relação direta do *e-value* obtido com a área de cobertura das CCEs selecionadas, pois o domínio conservado ocupa cerca de 85.47% (405.5 aminoácidos em média) (Pfam, 2013). Em muitos casos, apesar do *e-value* ser significativo e menor do que o ponto de corte normalmente aceito ( $P < 0,0001$ ), foi observado que a sequência era muito menor do que o tamanho médio dos membros dessa família (550 aminoácidos; Cygler et al., 1993), embora possuísse parte do domínio COesterase. Portanto, um novo ponto de corte foi calculado para o *e-value*. Para tal propósito e para que este método de cálculo pudesse ser validado, análises preliminares foram feitas avaliando as CCEs dos 12 genomas de *Drosophila* disponíveis. Estas sequências foram escolhidas por já terem sido extensamente avaliadas na literatura, tornando possível uma melhor identificação da presença de falsos positivos e negativos para relacioná-los com o *e-value* obtido. A

partir dessas análises, as CCEs que possuíam valor de *e-value* maior que  $1e-58$  na busca HMM foram excluídas do restante das análises.



**Figura 3.14** Esquema do *pipeline* usado para executar os programas de predição e editar os resultados de forma codificada nos cabeçalhos dos arquivos em formato FASTA. Os retângulos cinzas, azuis e verdes representam respectivamente os arquivos que são gerados pelo *pipeline* FAT, os scripts e programas executados e os arquivos gerados.

### 3.4 Análises

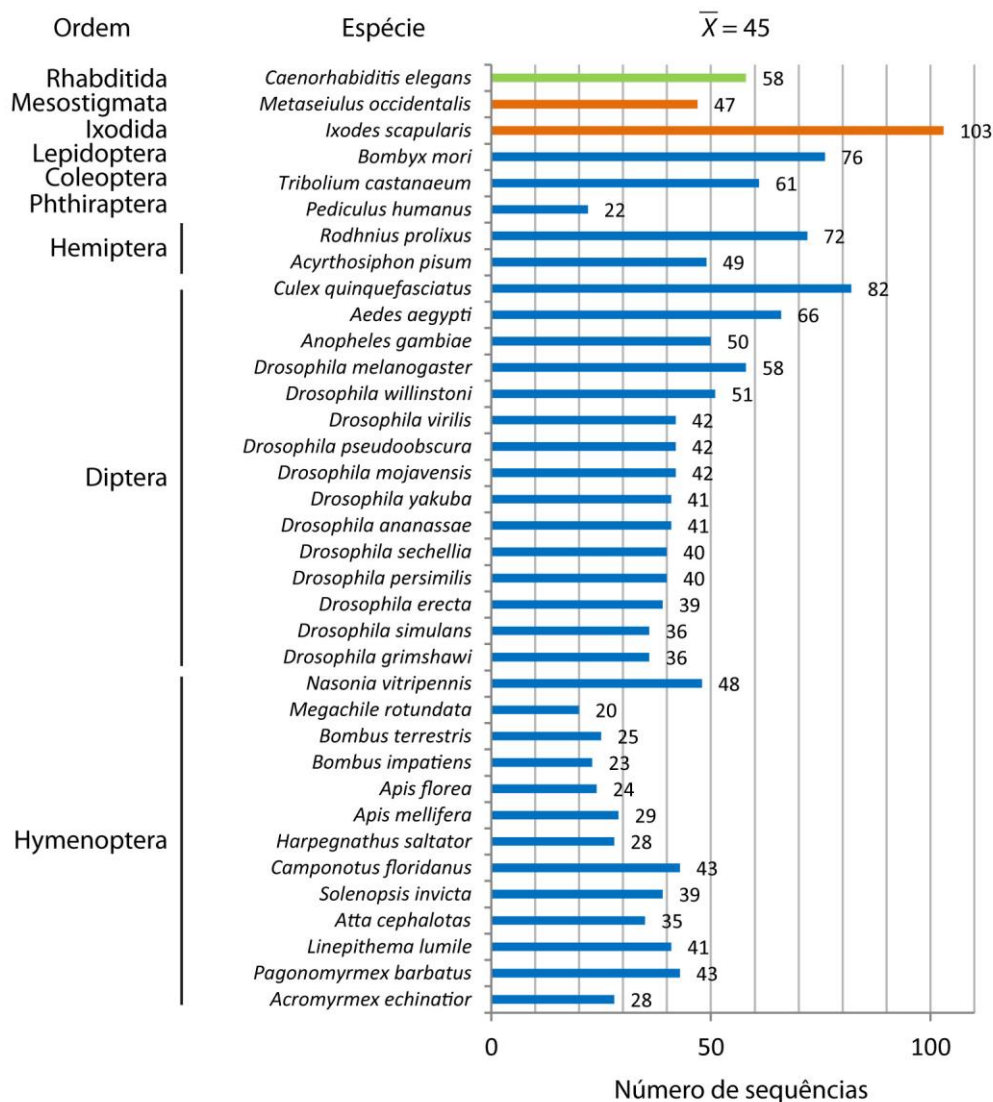
As sequências selecionadas foram alinhadas pelo programa MAFFT (Kato et al., 2002). A construção da árvore filogenética foi feita baseada em análises de máxima verossimilhança utilizando o programa RaxML (Stamatakis, 2006) usando a matriz de substituição *Jones Taylor Thornton* (JTT) (Jones et al., 1992). Para avaliar a confiança nos ramos da árvore filogenética, foi realizado uma análise *Bootstrap* com 500 réplicas. As análises estatísticas e gráficos desta dissertação foram feitos utilizando o pacote R (R Development Core Team).

## 4. Resultados

### 4.1 Sequências encontradas

Ao total 1620 sequências foram encontradas pelo *pipeline* FAT. A média foi de 45 sequências por espécie mas para algumas esse número variou bastante. O número de sequências encontradas nos genomas não parece ter relação com o fato de a espécie ser transmissora de doenças, apesar de algumas espécies de vetores como *Ixodes scapularis*, *Rhodnius prolixus*, *Culex quinquefasciatus* e *Aedes aegypti* apresentarem um número total de sequências superior à média observada. Em contraste, a espécie *Pediculus humanus* apresentou apenas 22 sequências. Além disso, outras espécies, como *Bombyx mori* e *Tribolium castanaeum*, que não transmitem doenças a humanos, também apresentaram um número expressivo de sequências (Figura 4.15).

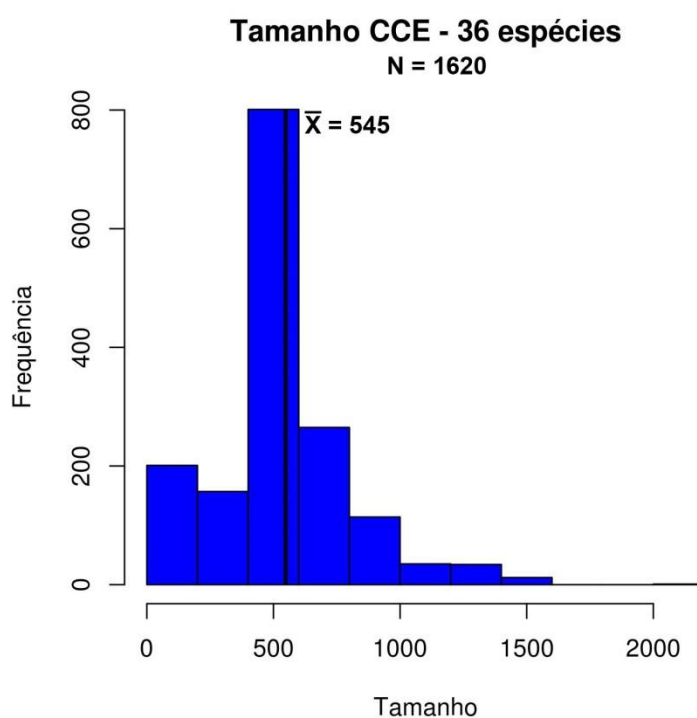




**Figura 4.15** Número total de seqüências encontradas por espécie. As barras em azul representam as espécies da classe Insecta, em laranja a classe Arachnida (ambas do filo Arthropoda) e em verde a classe Secernentea (Filo Nematoda). As ordens taxonômicas estão representadas nas barras laterais.

A média de aminoácidos por seqüência é de 545, o que corresponde aos resultados descritos por Cygler em 1993 para CCEs catalíticas. Nessa região da distribuição é onde se encontra a maioria dos membros que possuem o domínio COesterase completo e geralmente catalítico (Figura 4.16). O tamanho das seqüências variou de 42 a 2086 aminoácidos e, portanto, nas extremidades da distribuição estão presentes membros com características bem diferentes. As seqüências maiores que a média possuem outras regiões além do domínio COesterase como, por exemplo, as neurotactinas, gliotactinas e neuroliginas, que possuem uma região transmembranar já descrita (De la Escalera et al., 1990; Auld et al., 1995; Song et al., 1999). Além disso,

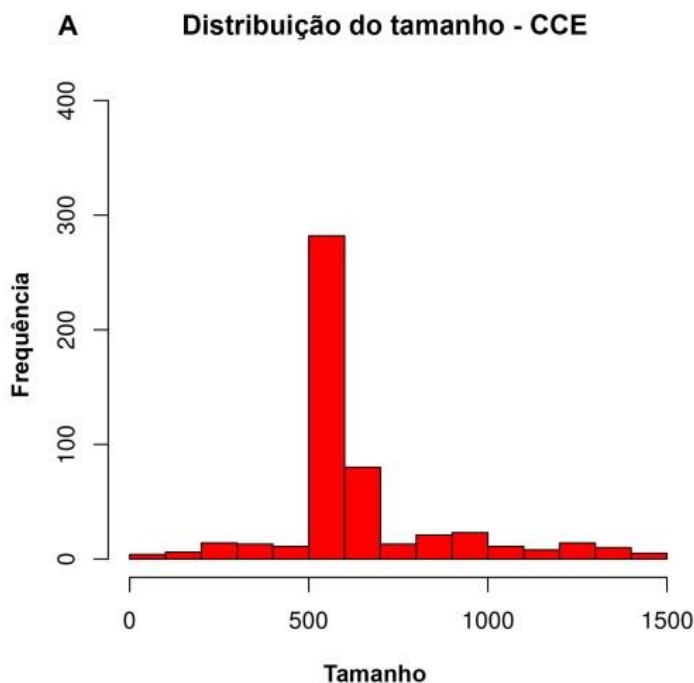
outros genes possuem mais de um domínio COesterase e por isso possuem tamanho acima da média observada (Figura 4.16). Não existem indícios na literatura de que CCEs quiméricos sejam funcionais, entretanto, para algumas CCEs de espécies de deuterostomados foram encontrados mais de um domínio COesterase (Johnson & Moore, 2012) e talvez isso possa ser resultado de eventos de duplicação (Zhang, 2003; Hurles, 2004) ou de erros na predição gênica (Nagy et al., 2008). O grupo de sequências pequenas que possui domínio parcial de COesterase fica na extremidade esquerda da distribuição (Figura 4.16) e, portanto, pode ser fruto de processos de pseudogenização (Zhang, 2003; Hurles, 2004) ou de problemas relacionados a predição gênica.



**Figura 4.16** Distribuição de tamanho (em número de resíduos) das sequências encontradas pelo *pipeline* FAT usando o perfil COesterase (PF00135) para as 36 espécies.

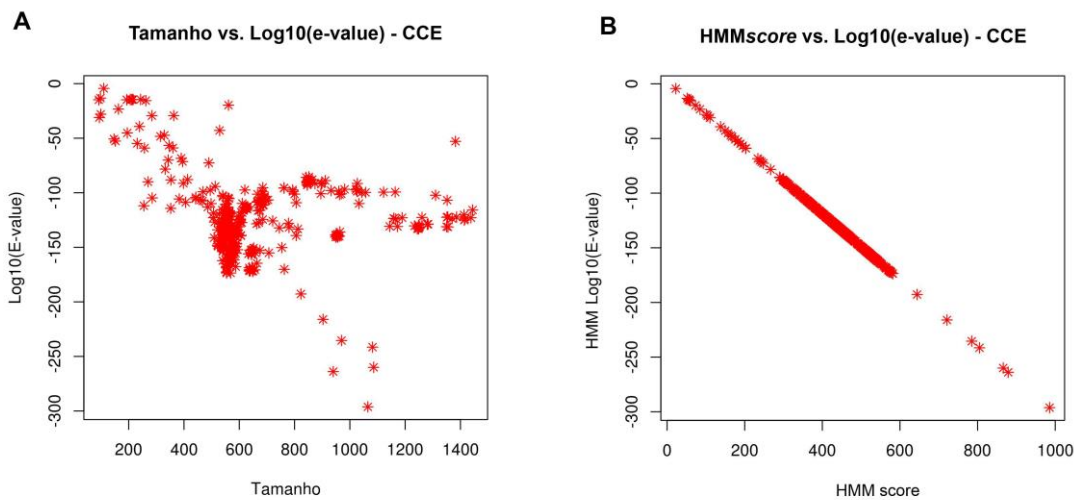
Devido à grande variação de tamanho encontrada entre as sequências, houve dificuldade de se obter um bom alinhamento múltiplo, pois as sequências menores muitas vezes não possuíam regiões que se sobrepusessem. Para resolver esse problema de forma sistemática, os valores de *score* e *e-value* obtidos pelo programa HMMsearch foram avaliados nos 12 genomas bem estudados de *Drosophila* disponíveis. O intuito dessa análise foi avaliar o comportamento dos valores de *e-value* para genes funcionais e não funcionais em um conjunto de sequências bem predito. Como pode ser observado

na Figura 4.17, a distribuição de tamanho das CCEs nas 12 espécies de *Drosophila* apresenta um perfil semelhante ao encontrado nas análises dos resultados de todos os genomas avaliados (observado na Figura 4.16).



**Figura 4.17** Distribuição de tamanho das sequências de CCEs encontradas pelo *pipeline* FAT para 12 espécies de *Drosophila*.

A grande variação de tamanho para os membros da família CCE impacta na variação dos *e-values* encontrados (Figura 4.18). Além disso, para as CCEs com menos de 500 resíduos, pode ser observada uma relação direta entre o tamanho e o valor do respectivo *e-value* obtido pelo programa HMMsearch (Figura 4.18A). O fato de o domínio conservado COesterase (PF00135) ocupar cerca de 85% (405 aa) da sequência proteica vai influenciar diretamente o valor de *e-value* já que o *score* obtido no alinhamento reflete de maneira direta o tamanho total da sequência e consequentemente seu *e-value* (Figura 4.18B) sendo isto bem visualizado nestas sequências menores. No entanto, para as sequências com mais de 500 resíduos esta relação não é tão clara, sendo que sequências de um mesmo tamanho podem ter *e-values* muito diferentes como mostra a Figura 4.18A.



**Figura 4.18** Gráfico de dispersão mostrando a relação entre o Log10 do *e-value* e tamanho da sequência (A) e entre o Log10 do *e-value* e o *score* (B) para todas as sequências encontradas pelo *pipeline* FAT. Os valores de *e-value* e *score* foram extraídos do resultado do HMMsearch para a família CCE (PF000135).

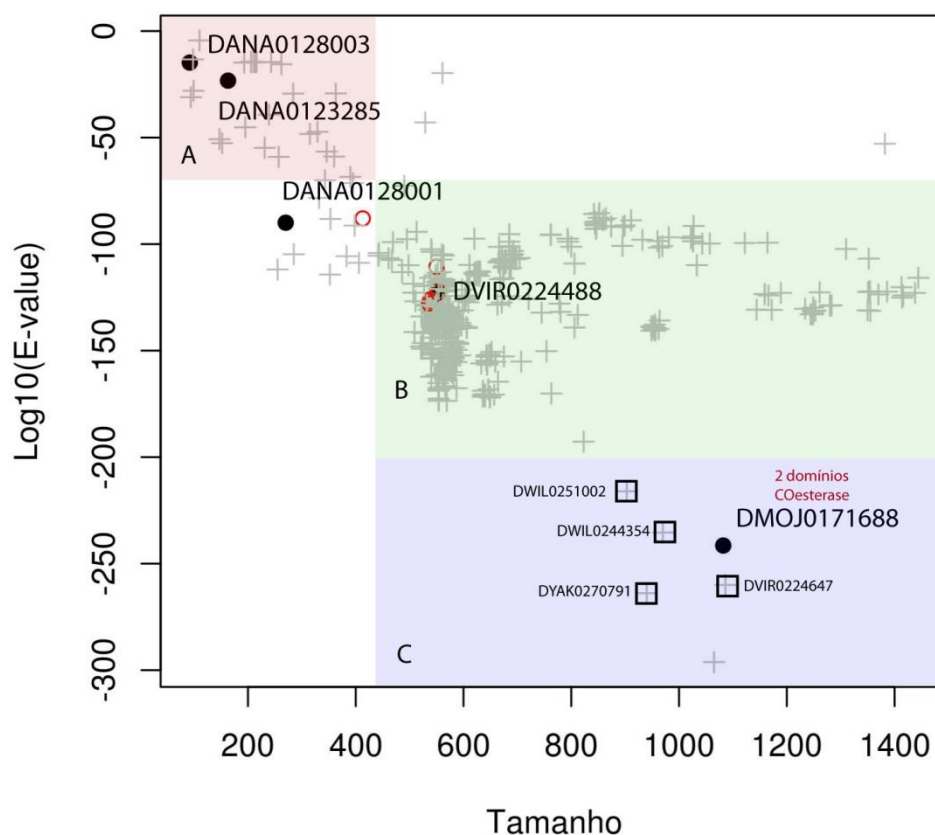
Para avaliar a relação do tamanho com o *e-value* de genes não funcionais foram utilizados dados de uma publicação que caracteriza a evolução do grupo de genes beta-esterase em *Drosophila* (Robin et al., 2009). Nesse trabalho o autor identifica nove genes não funcionais para as espécies *D. ananassae* (3), *D. virilis* (1), *D. mojavensis* (2), *D. buzzatii* (2) e *D. grimshawi* (1). Os três genes de *D. ananassae* foram considerados nulos, pois possuem o domínio COesterase fragmentado. Os das espécies *D. virilis*, *D. mojavensis* e *D. buzzatii* possuem substituições nos resíduos da tríade e por isso são caracterizados como não catalíticos. O de *D. grimshawi* foi considerado um pseudogene. Para confirmar a composição gênica dos *clusters* e acessar as sequências proteicas desses genes, o *cluster* de beta esterases de *D. melanogaster* foi visualizado no navegador de genomas do sítio FlyBase (<http://flybase.org/>) através da palavra-chave “est6” e a partir da aba de genes ortólogos as sequências das outras espécies foram acessadas.

Devido à contínua curagem das predições gênicas do flybase, a identificação dos genes determinados por Robin e colaboradores (2009) sofreu algumas alterações e nem todos os genes puderam ser encontrados. No conjunto de dados dessa dissertação, os dois genes não funcionais de *D. mojavensis* estão unidos em um único arquivo fasta (identificado como DMOJ0171688) e o de *D. grimshawi* não foi encontrado. A espécie *D. buzzatii* não possui genoma sequenciado e por isso também não está incluída neste

trabalho. Os genes de *D. ananassae* e *D. virilis* estão de acordo com a descrição feita por Robin e colaboradores em 2009.

Com essas informações a localização dos cinco genes foi ressaltada no gráfico de dispersão (círculos cheios pretos na Figura 4.19). A Figura 4.19 mostra o mesmo gráfico de dispersão das sequências em Log10 do *e-value* por tamanho que na Figura 4.18, mas agora dividido em três áreas importantes. Na área vermelha estão as sequências de menor tamanho e valores de *e-value* maiores do que o novo ponto de corte escolhido. Na área verde se localizam as sequências com o domínio COesterase íntegro e na área azul sequências com dois domínios COesterase. Neste gráfico, os dois genes de *D. ananassae* (DANA0128003 e DANA0123285) ocupam a área vermelha onde a maioria das sequências possui tamanho menor que a média descrita para os membros catalíticos e possuem *e-value* maior do que o ponto de corte escolhido. A sequência DANA0128001 que possui apenas um dos resíduos catalíticos fica na região intermediária entre as áreas vermelha e verde. O gene da espécie *D. virilis* (DVIR0224488) se localiza no aglomerado de genes junto com os 11 membros de beta-esterases catalíticos (identificados por Robin e colaboradores em 2009) que estão representados por círculos vermelhos na área verde. O gene da espécie *D. mojavensis* (DMOJ0171688; que possui 1.182 aminoácidos) possui dois domínios conservados COesterase e por isso está localizado na área azul. Ainda na região azul foram analisadas outras quatro sequências (DWIL0251002, DWIL0244354, DYAK0270791, DVIR0224647, quadrados sem preenchimento na Figura 4.19) que não foram estudadas por Robin e colaboradores em 2009. Para essas sequências também foi observada a presença de dois domínios COesterases, o que não é comum para os membros da família CCE, no entanto, Johnson e Moore em 2012 também descreveram a presença de genes com mais de um domínio COesterase em seus resultados.

### Tamanho vs. Log10(E-value) - Não funcionais ou pseudogenes - CCE



**Figura 4.19** Gráfico de dispersão mostrando a relação entre o Log10 *e-value* e tamanho da sequência encontrada pelo *pipeline* FAT para as 12 espécies de *Drosophila*. Os círculos pretos destacam os genes não funcionais do grupo das beta-esterases, os vermelhos os funcionais e as cruces cinza representam os demais membros da família das CCE para as 12 espécies de *Drosophila*. Os quadrados pretos indicam outros quatro genes com 2 domínios COesterase não estudados por Robin e colaboradores em 2009. A) Área vermelha onde estão sequências com tamanho menor que a média e *e-value* maiores do que o ponto de corte escolhido; B) área verde onde se concentram a maior parte das CCEs com um único domínio inteiro e C) área azul onde se encontram proteínas com dois domínios COesterase conservados.

#### 4.2 Pipeline de processamento de informação

O *pipeline* criado extrai as informações relevantes dos arquivos finais das análises (Multifasta, HMMsearch, TargetP, CD-search e TMHMM) e cria um arquivo texto cujas informações são delimitadas pelo caractere especial “[”]. As informações para cada sequência analisada são organizadas em linhas separadas, de acordo com o código de identificação de cada sequência. Como pode ser visto na Figura 4.20, a primeira

coluna do texto, representada pela cor cinza, lista os códigos de identificação das sequências. Nas duas colunas seguintes, representadas pela cor laranja, ficam organizados os dados oriundos do resultado do programa CD-search que correspondem à predição dos resíduos que compõem a tríade catalítica e a cavidade de ligação do substrato. Nas colunas seguintes, em azul, ficam organizados os resultados oriundos da predição da localização celular feita pelo programa TargetP e o seu nível de confiança (1 corresponde a alta probabilidade da localização estar correta, variando até 5 onde a probabilidade da ocorrência de falsos-positivo é alta) (Emanuelsson et al., 2007). A coluna em vermelho lista o tamanho das sequências proteicas. Nas duas colunas em verde ficam organizados os resultados do programa TMHMM, que informa a presença ou ausência de hélices hidrofóbicas e no caso de proteínas que sejam transmembranares, a segunda coluna verde informa a possível orientação da hélice através do código o (*outside*) e i (*inside*) junto da posição do resíduo inicial e final da hélice. Nas últimas duas colunas ficam listados os valores de *e-value* e *score* obtidos pelo programa HMMsearch executado pelo *pipeline* FAT.

```
DANA0120494|S202,E335,H456|G119,G120,G121,D201,S202,A203,A206,V360,L364,N365,V402,G457,L460||4|554|PredHel=0|Topology=o|1.6e-169|565.1
DANA0120489|S225,E357,H480|G142,G143,A144,E224,S225,A226,A229,V383,G387,L388,F426,A481,L484|M|2|574|PredHel=0|Topology=o|8.2e-144|480.2
DANA0118167|G337,E481,H596|G257,G258,E259,P336,G337,A338,A341,Q505,V511,F512,L551,D597,R600|S|1|951|PredHel=1|Topology=o724-746|2e-139|465.7
```

**Figura 4.20** Tabela gerada pelo *pipeline* desenvolvido para organizar as informações relevantes do trabalho. Código de identificação (cinza), resíduos catalíticos e de ligação ao substrato (CD-search - laranja), predição da localização celular (TargetP - azul), tamanho da sequência (vermelho), hélice transmembranar (TMHMM - verde) e *e-value* e *score* (HMMsearch - preto).

Tendo como base as informações extraídas do arquivo de resultados descrito acima, o *pipeline* também cria um arquivo em formato FASTA com a codificação dos resultados para cada sequência em seu cabeçalho. Assim, além do código de identificação da sequência as informações do CD-search, TargetP e TMHMM, de forma compacta e delimitada por underscores “\_” pode ser visualizada facilmente mesmo em análises posteriores. Como pode ser visto na Figura 4.21, o cabeçalho é iniciado pelo código de identificação da sequência em preto, seguido pelos aminoácidos preditos pelo CD-search em laranja (com as posições omitidas), a localização predita pelo TargetP em azul e a presença ou ausência (O ou Y respectivamente) de predição para hélice transmembranar em verde. A orientação da hélice transmembranar é omitida no cabeçalho.



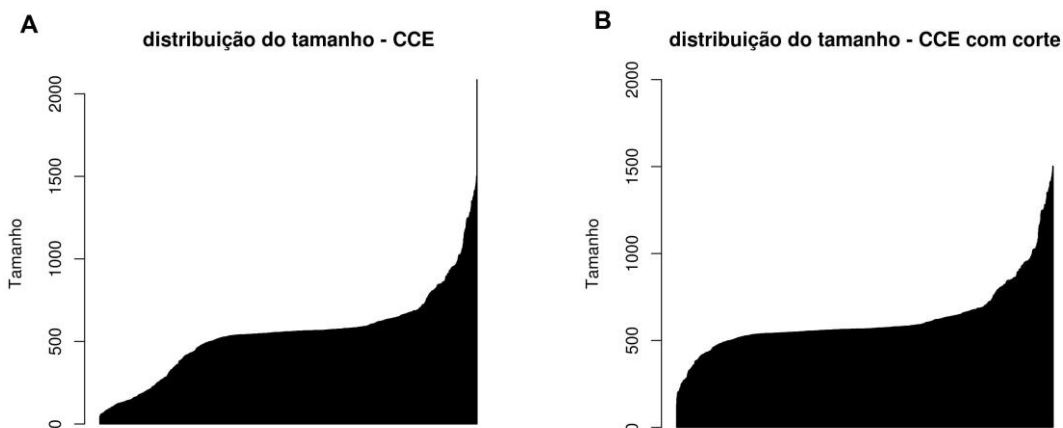
```

>DANA0120494_SEH_GGGDSAAVLNVGL_I4_Y
MVDKVKQRRRTSEKTVVSTTYGPIKGVKRKSIYQSYFSFERIPFAKPPVGELRYKAPQPPEVWTEVKSCTSQGGPKPLQKHVFEMTDGSEDCLYLNVTYKKNLYPT
KMPFVMWYIYGGFRFGEASRECYSPDYLLREDVVVISINYRLGPLGLCIDDPEFDVPGNAGLKDQVLALRWVKANCSRFGGDSGNITIFGDSAGSASVHYMMITEQ
TRGLFHKAICMSGNTLSPWAVTPQRNWPYRLAVQAGYTGENNVRDWEFLNNAKGSDIKANGELCIDEEKKERIGFSFGPVIPIYVTSHCVVPKPIEMMRTAWSNS
IPLILGGVSNELLLYSETKANPKCLNELDDCRFVPIELNMDRDSALCREYGEQLRQCYYGDKTPSLDTLHEYLMVSHYFVFPPIYRTVLSRLQYATGAPTYLYRFD
FDSKHFHNLRLILSCGKVRGTCHGDDL SYLFYNSLARKLKNHTREYKCIERLVGLWTHFATCGNPNFDEQADLWQPVDAAAIEKHQLKCLNISDELKVIDVPDMRKL
MVWESFFRRDELL

```

**Figura 4.21** Exemplo do arquivo FASTA das sequências após a codificação das informações dos programas de predição no cabeçalho. As posições dos resíduos identificados pelo programa CD-search e a topologia das hélices transmembranares foram omitidas para facilitar interpretação dos resultados em análises subsequentes.

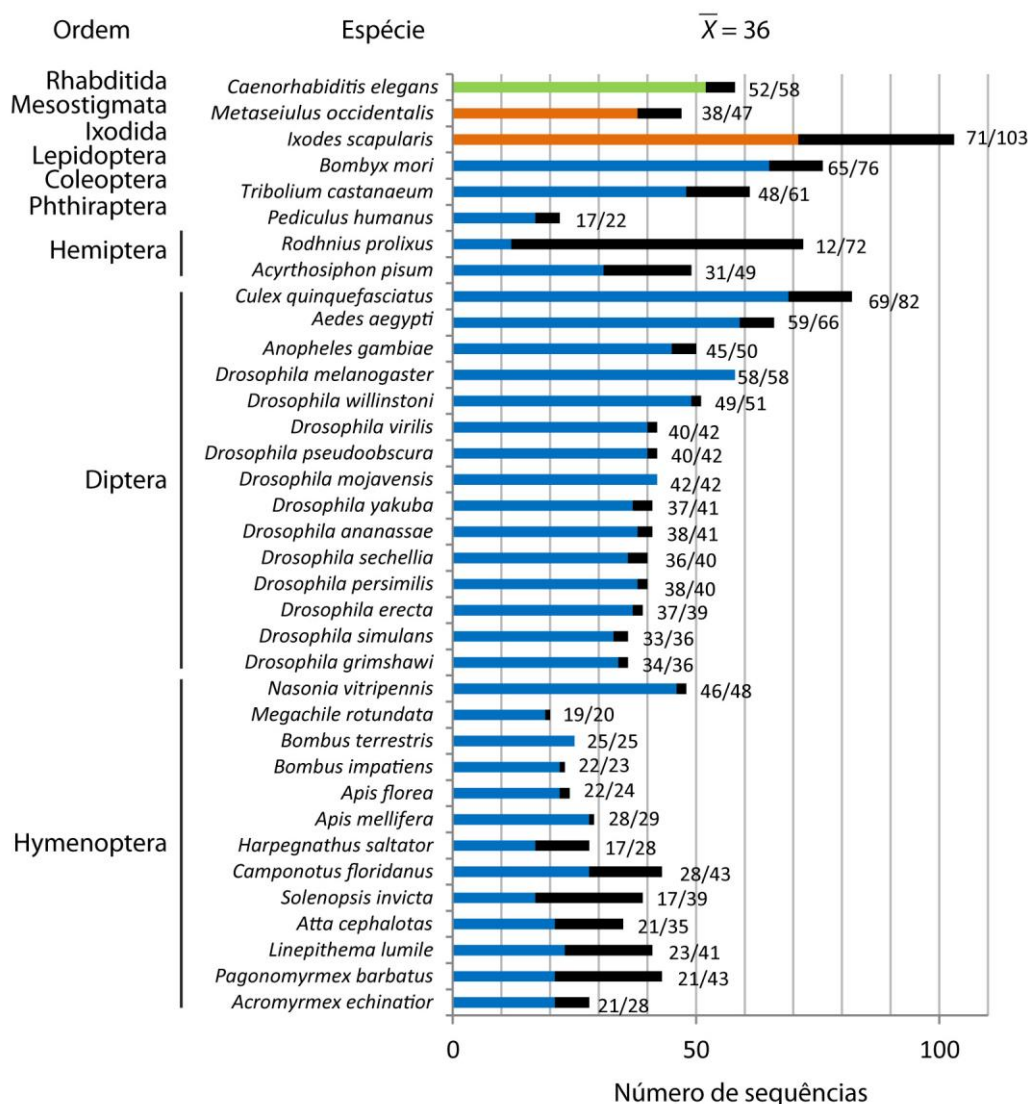
Além do processamento de informações, o *pipeline* também conta com uma etapa de exclusão de sequências de acordo com o *e-value*. Após as análises dos genomas de *Drosophila* descritas acima, levando em conta o tamanho das sequências e seus respectivos *e-values*, foi configurado o valor de  $1e-58$  para servir como corte de exclusão de fragmentos de CCEs. No universo de 1620 sequências, 1299 possuíam valores acima do corte e, portanto, passaram a compor o conjunto de sequências a serem analisadas. A exclusão de sequências pelo *e-value* eliminou sequências de tamanho muito pequeno, e que poderiam atrapalhar os alinhamentos múltiplos, como pode ser visto na comparação feita antes e depois da aplicação desse passo pelo *pipeline* (Figura 4.22). Na Figura 4.22A, o ordenamento do tamanho das CCEs mostra a presença de sequências com tamanho inferior a 405 resíduos (tamanho do domínio conservado COesterase). Já na Figura 4.22B, o ordenamento mostra que esse extrato de sequências não faz parte do novo conjunto de dados.



**Figura 4.22** Ordenamento do tamanho das sequências antes e depois da execução do passo de exclusão feita pelo *pipeline* de processamento de informação.



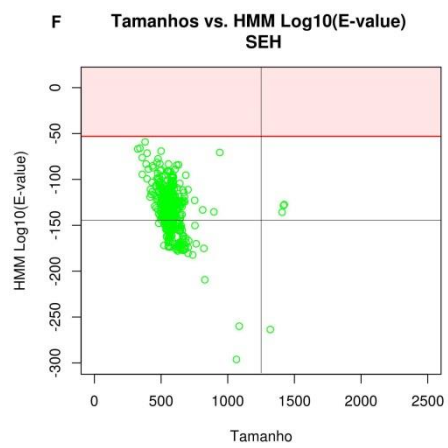
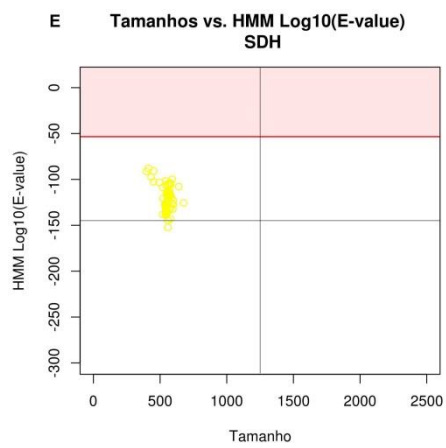
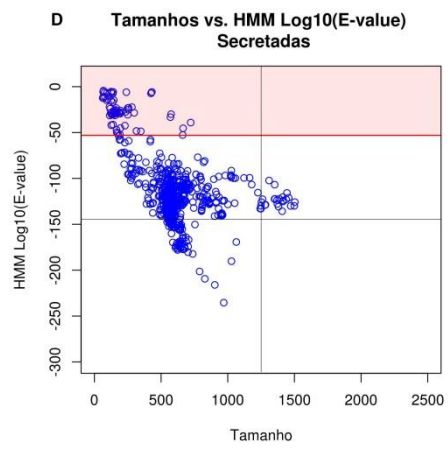
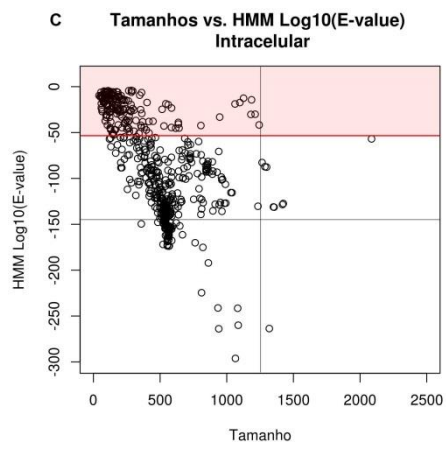
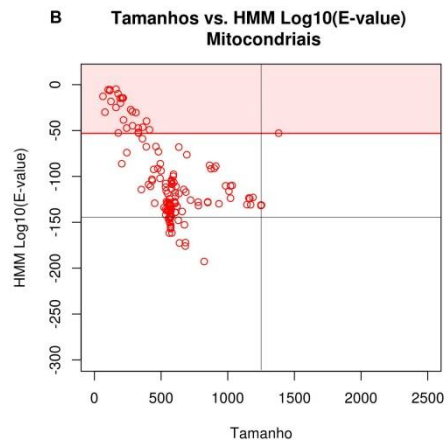
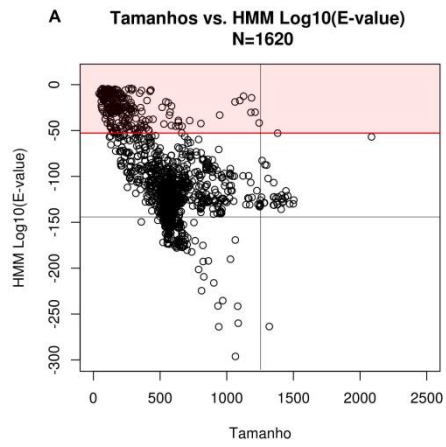
O impacto do corte no número total de esterases é relevante para algumas espécies como, por exemplo, *R. prolixus* que teve 83% (60) das sequências excluídas. Este resultado é compreensível, já que a predição gênica usada para este inseto ainda estava na versão “rascunho” e provavelmente possui problemas quanto a integridade dos genes. Somente durante a escrita desta dissertação a predição gênica estável (V1.1) foi liberada pelo VectorBase. Por outro lado, as espécies *D. melanogaster*, *D. mojavensis* e *Bombus terrestris* não tiveram nenhuma sequência excluída pelo corte. Após o corte, a média de sequência por espécie passou a ser 36, e não mais 45 como observado no conjunto original de sequências (Figura 4.23).

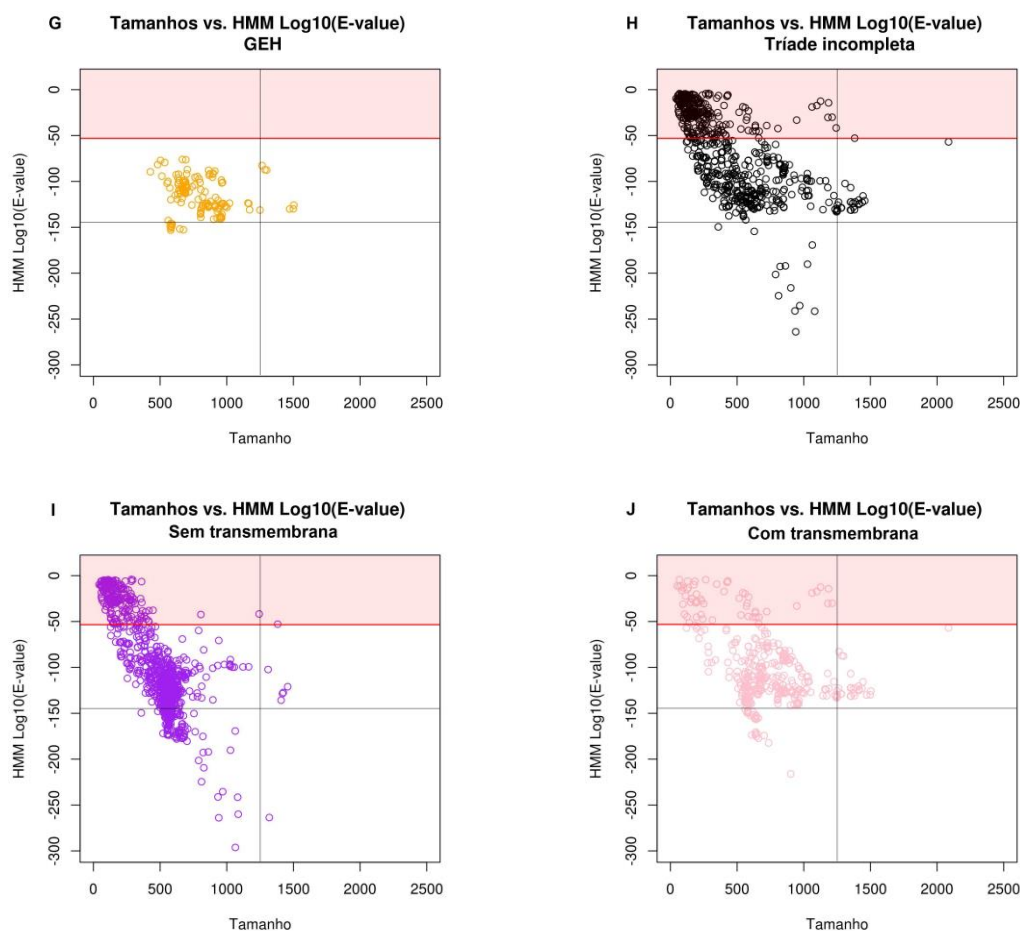


**Figura 4.23** Número total de sequências por espécie após a exclusão feita pelo *pipeline* secundário. Ao lado das barras estão representados o número total após o corte seguido pelo número inicial de sequências. As barras em azul representam as espécies da classe Insecta, em laranjas a classe Arachnida (ambas do filo Arthropoda) e em verde a classe

Secernentea (Filo Nematoda). As barras laterais na esquerda delimitam espécies das respectivas ordens taxonômicas.

Com o intuito de avaliar se, ao eliminar as sequências com *e-value* abaixo do novo ponto de corte estabelecido, estaríamos causando algum tipo de viés, diversos tipos de características das sequências e seus *e-values* foram avaliadas. Do total de proteínas identificadas (1620), é evidente que grande parte possui valores abaixo da média de tamanho que é de 545 aminoácidos e muitas delas foram retiradas da análise por possuírem *e-value* maior do que o ponto de corte (Figura 4.24A). Em relação à localização celular feita pelo programa TargetP, os três tipos de identificação (mitocondrial, intracelular e secretada) tiveram proteínas com valor de *e-value* acima do ponto de corte, sendo que no caso das intracelulares um número maior de sequências foi excluído (Figura 4.24B-D). Na identificação de proteínas transmembranares feita pelo programa TMHMM, os resultados também mostram que parte dessas sequências ficam acima do ponto de corte e foram excluídas (Figura 4.24I e J). Em relação à identificação da tríade catalítica feita pelo programa CD-search, os resultados apontam que os dois principais grupos de sequências que apresentaram tríade catalítica (Ser-Glu-His e Ser-Asp-His), não possuem proteínas acima da linha de corte e, portanto, nenhuma sequência que possuía a tríade catalítica completa foi excluída do trabalho (Figura 4.24E e F). Estes resultados mostram que não houve um viés importante com o novo ponto de corte estabelecido e se este viés aconteceu foi no sentido de deixar que as enzimas supostamente catalíticas (e mais prováveis de estarem envolvidas na resistência a inseticidas) continuassem na análise. Isso também acontece para o grupo de sequências que apresentaram um resíduo de glicina substituindo a serina nucleofílica da tríade (Figura 4.24G). Para o grupo de sequências que não apresentaram a tríade catalítica, com o esperado, fica evidente a presença de proteínas com tamanho abaixo da média (Figura 4.24H).





**Figura 4.24** Padrão de dispersão de proteínas encontradas de acordo com a relação ao valor de *e-value* obtido na busca HMM e seu respectivo número de resíduos. A linha vermelha no eixo y identifica o valor de corte ( $1e-58$ ). A área preenchida em vermelho claro representa o grupo de sequências excluídas com este ponto de corte. As linhas cinzas são guias para facilitar a visualização dos padrões de dispersão. Cada gráfico apresenta o padrão de dispersão para os seguintes grupos de sequências: A) universo de sequências encontradas pelo *pipeline* FAT; B) mitocondriais; C) intracelulares; D) secretadas; E) tríade SDH; F) tríade SEH; G) tríade GEH; H) tríade incompleta; I) sem transmembrana e J) com transmembrana.

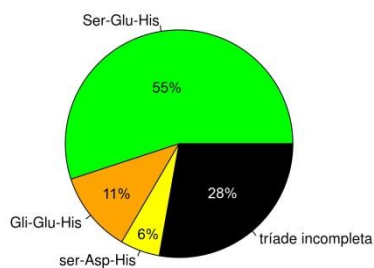
### 4.3 Predições

Do grupo de sequências selecionadas pelo novo ponto de corte (1299), 55% (715) tiveram os resíduos serina (Ser), ácido glutâmico (Glu) e histidina (His) preditos para a tríade catalítica. Apenas 6% (73) possuíam ácido aspártico (Asp) no lugar de ácido glutâmico (Glu). Em 11% (150) a tríade possuía um resíduo de glicina (Gli) na posição da serina catalítica caracterizando a tríade Gli-Glu-His. Quando nenhum dos resíduos acima foram encontrados nas posições esperadas para a tríade catalítica estas foram consideradas como incompletas, tendo ocorrido em 28% (361) das sequências.

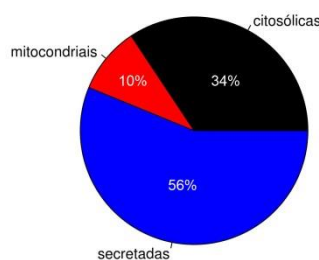
Nesse último grupo, na posição da serina nucleofílica foram identificadas substituições por valina, alanina, ácido aspártico e arginina. Na posição do resíduo ácido foram identificadas substituições por asparagina e alanina e na posição da histidina conservada foram observadas substituições por alanina, metionina, valina, arginina e lisina (Figura 4.25A).

O programa TargetP realiza a predição da localização celular das sequências através do peptídeo sinal localizado na porção amino-terminal das proteínas (Emanuelsson et al., 2007). As sequências com endereçamento predito para a via secretória representam 56% (730) das sequências enquanto que para a mitocôndria 10% (123). Proteínas que não tiveram peptídeo sinal identificado foram consideradas como citosólicas e representam 34% (446) das sequências (Figura 4.25B). Cerca de 27% (355) do conjunto de sequências selecionadas são potencialmente transmembranares, pois apresentaram predições de regiões hidrofóbicas que podem ter função de ancoramento na membrana. Ainda em relação ao programa TMHMM, é importante ressaltar que devido à característica hidrofóbica de peptídeos sinais, regiões transmembranares próximas a porção amino-terminal podem ser identificadas e por serem transitórias devem ser consideradas como falso-positivas (Figura 4.25C). Tendo em vista essa limitação, os dados da literatura que caracterizam as posições das alças transmembranares em neuroliginas, neurotactinas e gliotactinas, foram usados para avaliar a orientação das hélices e separá-las de sinais de endereçamento. Com isso foram identificadas 17% (226) de proteínas que potencialmente possuem regiões transmembranares e 10% (129) de peptídeos sinais que se encontravam na extremidade da porção amino-terminal (Figura 4.25D).

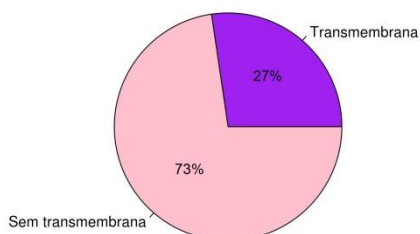
**A** predição dos resíduos catalíticos - CD search  
N=1299



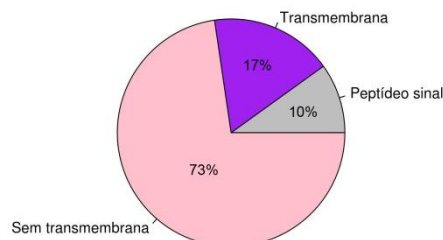
**B** Predição da localização celular - TargetP  
N=1299



**C** predição de regiões transmembranares - TMHMM  
N=1299



**D** Transmembrana e Peptídeo sinal hidrofóbico  
TMHMM N=1299

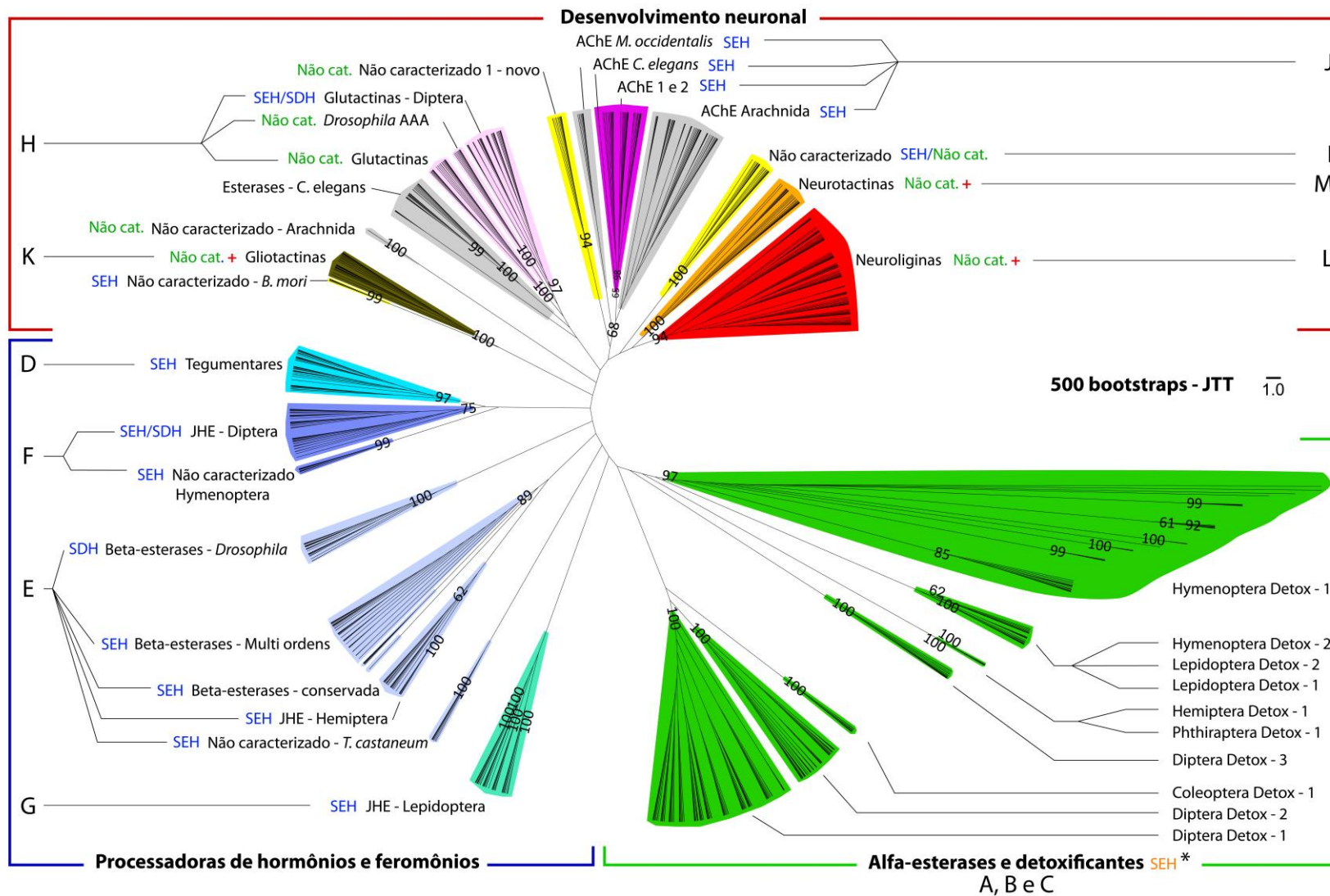


**Figura 4.25** Estatística dos resultados obtidos pelos programas de predição. A) Predição dos resíduos catalíticos usando CD-search; B) predição da localização celular usando TargetP; C) predição de hélices transmembranares usando TMHMM e D) reavaliação da predição de hélices transmembranares usando a literatura. (roxo: hélice transmembrana identificada, rosa: sem hélice transmembrana, e cinza: peptídeos sinal hidrofóbico erroneamente classificados como hélice transmembrana pelo TMHMM).

#### 4.4 Análises filogenéticas de CCEs de artrópodes

A análise filogenética dos genes da família CCE encontrada neste trabalho está representada na Figura 4.26. Trinta e oito sequências possuíam *e-value* menor que  $1e-58$ , mas não foram incluídas nessa análise devido ao modo de exclusão da primeira versão do *pipeline* desenvolvido (Tabela Suplementar 2). A Tabela 4.3 mostra a relação dos clados encontrados neste trabalho e os clados encontrados por Oakeshott e colaboradores em 2010. Na tabela podemos ver que em alguns casos os clados encontrados pelos autores também foram recuperados neste trabalho, como ocorre no clado das gliotactinas onde encontra-se genes ortólogos em uma relação 1:1 entre as

espécies. Em contraste, em outros clados, como o das beta-esterases de *Drosophila*, os genes encontrados pertencem a apenas uma única família taxonômica (Diptera). Ao total foram identificados 39 clados com valores de *bootstrap* acima de 60% (Figura Suplementar 1). Como era de se esperar com o aumento da amostragem, em alguns casos, não houve suporte estatístico corroborando a classificação proposta por Oakeshott e colaboradores. O exemplo mais drástico pode ser observado na classe Detoxicante, onde os três clados funcionais determinados por Oakeshott e colaboradores em 2010 na realidade se misturam nos diferentes clados táxon-específicos encontrados (Figura 4.26). Um melhor detalhamento dos principais clados e categorias funcionais de CCEs pode ser encontrado na Tabela 4.3.



**Figura 4.26** Árvore filogenética não enraizada dos membros da família CCE. Apenas os valores estatísticos (acima de 60%) são mostrados. Clados com proteínas transmembranares estão representadas por um sinal + vermelho. Clados com proteínas majoritariamente catalíticas secretadas estão representados pelo código que representa a tríade catalítica de três letras em azul (SEH ou SDH). Clados com proteínas majoritariamente endereçadas para diferentes locais estão representadas pelo código SEH na cor laranja.



**Tabela 4.3** Clados identificados no presente trabalho e por Oakeshot e colaboradores em 2010. Número de genes encontrados (número de espécies); Presença ou ausência de região transmembranar (TM); Endereçamento celular (S – secretadas, I – intracelulares e Sem padrão) e Tríade catalítica.

Clados Oakeshott et al., 2010	Clados encontrados neste trabalho	Nº de genes	TM	Endereçamento	Tríade	
L – Neurologina	L - Neurologina	137(35)	+	S	Não cat.	
K – Gliotactina	K - Gliotactina	37(32)	+	S	Não cat.	
M – Neurotactina	M - Neurotactina	31(28)	+	I	Não cat.	
J – Acetilcolinesterase	J - AChE 1	20(20)	-	S	SEH	
	J - AChE 2	33(31)	-	S	SEH	
	J - AChE <i>C. elegans</i>	3(1)	-	S	SEH	
	J - AChE - <i>M. occidentalis</i>	14(1)	-	S	SEH	
	J - AChE – Arachnida	81(2)	-	S	SEH	
	H – Glutactinas	H - Glutactinas	31(30)	-	S	Não cat.
		H - Glutactinas Diptera	54(15)	-	S	SEH/SDH
H - <i>Drosophila</i> AAA		15(12)	-	S	Não cat.	
I - Não caracterizado	I - Não caracterizado	31(31)	-	S	Não cat./SEH	
-	Não caracterizado 1 - Novo	19(19)	-	S	Não cat.	
-	Esterases - <i>C. elegans</i>	42(1)	-	Sem padrão	SEH	
-	Não caracterizado - Arachnida	2(2)	-	I	Não cat.	
-	Não caracterizado - <i>B. mori</i>	8(2)	-	S	SEH	
D – Tegumentares	D – Tegumentares	64(28)	-	S	SEH	
F - JHE – Diptera	F - JHE – Diptera	64(16)	-	S	SEH/SDH	
	F - Não caracterizado - Hymenoptera	13(11)	-	Sem padrão	SEH	

**Tabela 4.3** Continuação

Clados Oakeshott et al., 2010	Clados encontrados neste trabalho	Nº de genes	TM	Endereçamento	Tríade
E - Beta-esterase - <i>Drosophilas</i>       G – JHE Lepidoptera	E - Beta-esterase – <i>Drosophila</i>	32(12)	-	S	SDH
	E - Beta-esterase - multi Ordens	61(20)	-	S	SEH/SDH
	E - Beta-esterase conservada	30(28)	-	S	SEH/SDH
	E - JHE – Hemiptera	9(1)	-	S	SEH
	E - Não caracterizado - <i>T. castaneum</i>	12(1)	-	S	SEH
	G – JHE Lepidoptera	4(1)	-	S	SEH
	G – Lepidoptera 1	13(1)	-	S	SEH
	G – Diptera 1	12(3)	-	S	SEH
	G – Diptera 2	16(3)	-	S	SEH
	G – Hymenoptera 1	13(6)	-	S	SEH
A – Detoxificantes Hymenoptera	Hymenoptera Detox – 1	64(13)	-	I	SEH
B – Alfa-esterases	Hymenoptera Detox – 2	13(12)	-	I	SEH
C – Função desconhecida	Diptera Detox – 1	156(15)	-	I	SEH
	Diptera Detox – 2	62(15)	-	Sem padrão	SEH
	Diptera Detox – 3	18(15)	-	S	SEH
	Lepidoptera Detox – 1	13(1)	-	Sem padrão	SEH
	Lepidoptera Detox – 2	8(1)	-	I	SEH
	Coleoptera Detox – 1	14(1)	-	I	SEH
	Phthiraptera Detox – 1	2(1)	-	I	SEH
	Hemiptera Detox – 1	5(1)	-	Sem padrão	SEH

#### 4.4.1 Desenvolvimento neuronal

Nesse grupo estão organizados os seis clados (H-M; Figura 4.26) identificados previamente por Oakeshott e colaboradores (2005, 2010). Além desses, foram identificados outros três contendo proteínas de outros artrópodes (Não caracterizado 1 – Novo; Não caracterizado – Arachnida e Não caracterizado – *B. mori*). Os clados Gliotactina, Neurologina e Neurotactina (K, L e M), como já era esperado, são monofiléticos e os únicos que possuem proteínas não catalíticas com predição de região hidrofóbica associada ao ancoramento em membranas. O clado J, das Acetilcolinesterases, contém proteínas majoritariamente catalíticas e neste trabalho é possível ver que ocorreram expansões nas espécies da classe Arachnida. O clado I, ainda não caracterizado, é monofilético e anteriormente era considerado como não catalítico. Entretanto, neste trabalho detectamos a presença de sequências de himenópteros que possuem todos os três resíduos catalíticos da tríade (Ser-Glu-His). O clado H, das Glutactinas, como já descrito por Oakeshott e colaboradores em 2010, possui proteínas catalíticas e não catalíticas. O clado Não caracterizado 1 – Novo é monofilético e a presença de um gene do quelicerado *I. scapularis* indica que esse gene estava presente no ancestral dos artrópodes. Isso sugere que esse gene tenha sido perdido no ancestral dos himenópteros devido à ausência de genes de todas as treze espécies dessa ordem (Figura 4.29B). O clado Não caracterizado – Arachnida possui dois genes não catalíticos, um de *I. scapularis* e um de *M. occidentalis*. O clado Não caracterizado - *B. mori* possui cinco proteínas catalíticas e três não catalíticas de lepidóptero, além de um gene de *C. elegans*.

As neurologinas (Clado L) formam um clado monofilético com 94% de *bootstrap*. Este clado agrupa cinco clados monofiléticos que foram identificados em insetos como NLGN-1, 2, 3, 4 e 5 (Biswas et al., 2008). Além desses, dois outros clados foram recuperados. O clado aqui denominado de NLGN-CE, mais externo, contém um gene do quelicerado *M. occidentalis*, um de *B. mori* e os genes do nematódeo *C. elegans*. No outro clado encontramos quatro genes de *M. occidentalis* (Clado NLGN-MO), em uma posição mais interna (Figura 4.27). Nenhum gene de *I. scapularis* foi encontrado nesse grupo de proteínas podendo isto ser devido à falta de genes homólogos neste carrapato. Contudo, é mais provável ter acontecido por causa da exclusão de sequências de baixa qualidade feita no início das análises, já que em *M. occidentalis* o gene foi encontrado. Nem todas as espécies de insetos apresentaram todos os cinco clados de NLGN sendo que a representação do número de genes por espécie em cada clado está listada na Tabela 4.4. Como pode ser



existente para os ramos NLGN 1, 2, 3, 4 e 5 (Biswas et al., 2008) e os dois cladoss novos, NLGN-CE e NLGN-MO. Para a melhor identificaçãoo das espécies utilize o material suplementar (Tabela S1).

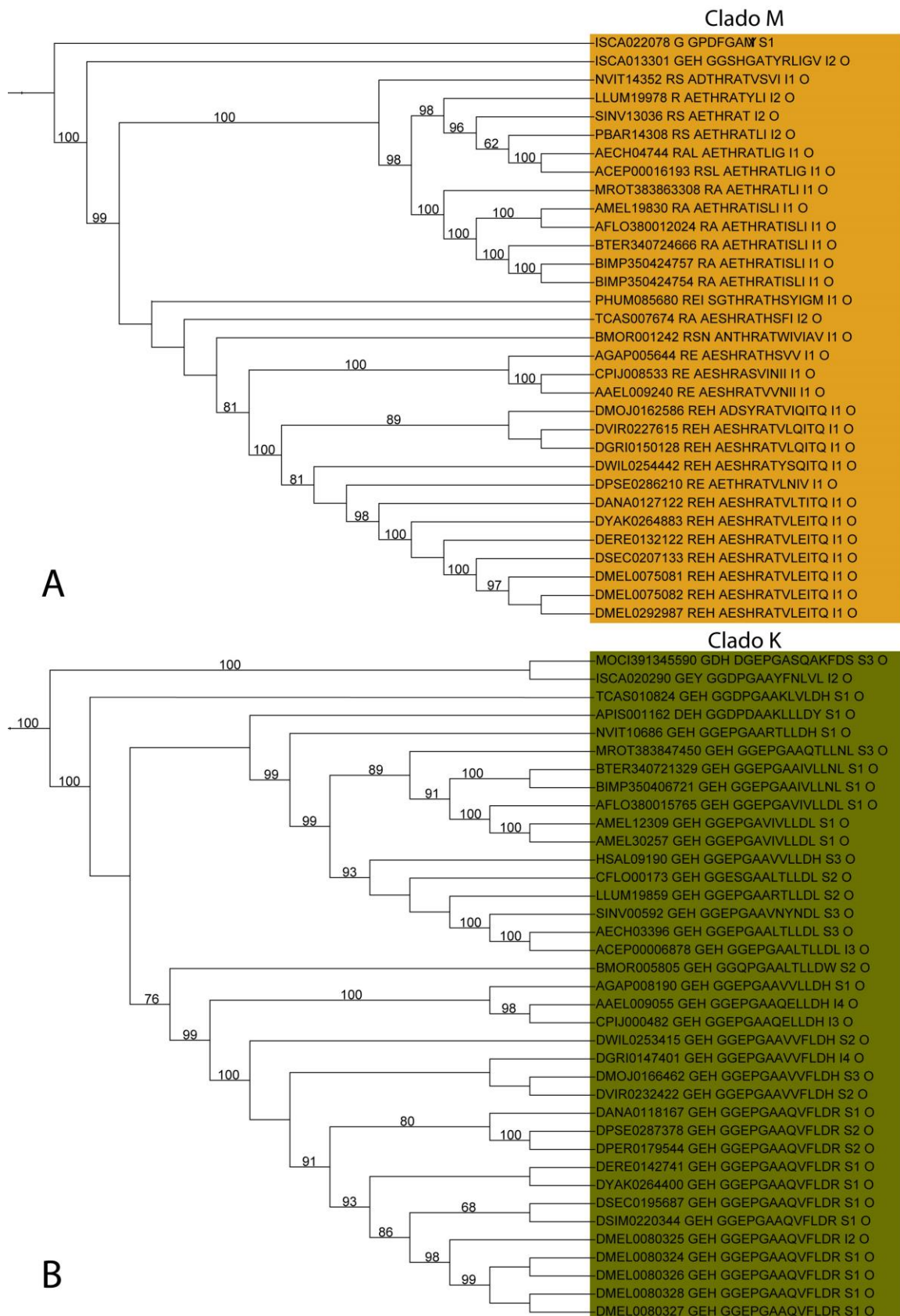
**Tabela 4.4** Número de genes por espécie nos cladoss neurologinas.

espécie	Neurologinas (NLGN)						
	1	2	3	4	5	CE	MO
<i>Acromyrmex echinator</i>	1	1	1		1		
<i>Pagonomyrmex barbatus</i>			1		1		
<i>Linepithema humile</i>	1	1	1	1	1		
<i>Atta cephalotes</i>		1		1	1		
<i>Solenopsis invicta</i>		1			1		
<i>Camponotus floridanus</i>	1	1		1	1		
<i>Harpegnathus saltator</i>			1		1		
<i>Apis mellifera</i>	1	1	1	1	1		
<i>Apis florea</i>		1			1		
<i>Bombus impatiens</i>	1	1	1	1	1		
<i>Bombus terrestris</i>	1	1	1	1	1		
<i>Megachile rotundata</i>	1	1	1	1	1		
<i>Nasonia vitripennis</i>	1	1	1	1	1		
<i>Drosophila grimshawi</i>		1	1	1			
<i>Drosophila simulans</i>	1	1	1	1			
<i>Drosophila erecta</i>	1	1	1	1			
<i>Drosophila persimilis</i>	1	1		1			
<i>Drosophila sechellia</i>	1	1	1	1			
<i>Drosophila ananassae</i>	1	1	1	1			
<i>Drosophila yakuba</i>	1	1	1	1			
<i>Drosophila mojavensis</i>	1	1	1	1			
<i>Drosophila pseudoobscura</i>	1	1	1	1			
<i>Drosophila virilis</i>	1	1	1	1			
<i>Drosophila willinstonii</i>	1	1	1	1			
<i>Drosophila melanogaster</i>	2	2	1	3			
<i>Anopheles gambiae</i>	1	1	1	1	1		
<i>Aedes aegypti</i>	1	1		1			
<i>Culex quinquefasciatus</i>	1	1	1	1	1		
<i>Acyrtosiphon pisum</i>	1		1				
<i>Rodhnius prolixus</i>			1				
<i>Pediculus humanus</i>	1	1	1	1	1		
<i>Tribolium castaneum</i>	1		1	2	1		
<i>Bombyx mori</i>		1	1		1	1	
<i>Ixodes scapularis</i>							
<i>Metaseiulus occidentalis</i>						1	4
<i>Caenorhabditis elegans</i>						5	

O gene da neurotactina (Clado M) é descrito como exclusivo da classe Insecta (Gilbert & Alud, 2005), entretanto, a detecçãoo do gene ISCA013301 de *I. scapularis* indica

que esse gene estava presente no ancestral dos artrópodes (Figura 4.28A). Os resultados das predições (TMHMM e CD-search) corroboram os dados da literatura, pois essas proteínas possuem região transmembranar na porção amino-terminal e não são catalíticas (Tabela 4.3). Nesse clado, não foram encontrados genes para as espécies *D. persimilis*, *D. simulans*, *Acy. pisum*, *R. prolixus*, *H. saltator*, *Ca. floridanus* e ainda para o quelicerado *M. occidentalis*. A falta destes genes pode ser devido a erros de predição, já que ele pode ser encontrado em espécies próximas de todas as espécies com o gene ausente. As espécies *D. melanogaster* e *B. impatiens* possuem 3 e 2 genes respectivamente (Figura 4.28A).

As gliotactinas (Clado K) formam um grupo monofilético com alto suporte (100% de *bootstrap*). Como as neurotactinas, estes genes apresentam o domínio COesterase e uma região transmembranar, porém, localizada na porção carboxi-terminal (Tabela 4.3). A presença bem suportada de dois genes de espécies da classe Arachnida, MOCI39135590 e ISCA020290, indica que as gliotactinas estavam presentes no ancestral do filo Artrópoda. Nesse clado não foram encontrados genes apenas para as espécies *R. prolixus* e *P. barbatus* do filo Arthropoda, enquanto *D. melanogaster* e *Ap. mellifera* possuem 5 e 2 genes respectivamente (Figura 4.28B). Um clado de genes não caracterizados de *B. mori* (com um de *C. elegans*) se agrupa com as gliotactinas, no entanto, sem suporte estatístico (Figura 4.26).

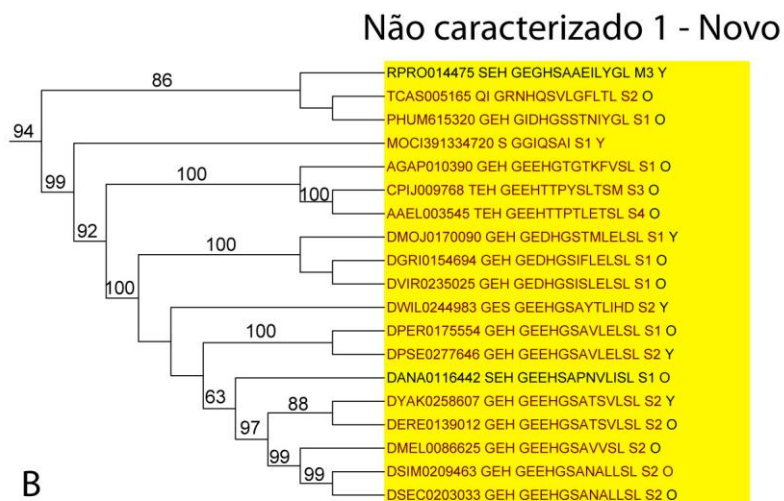
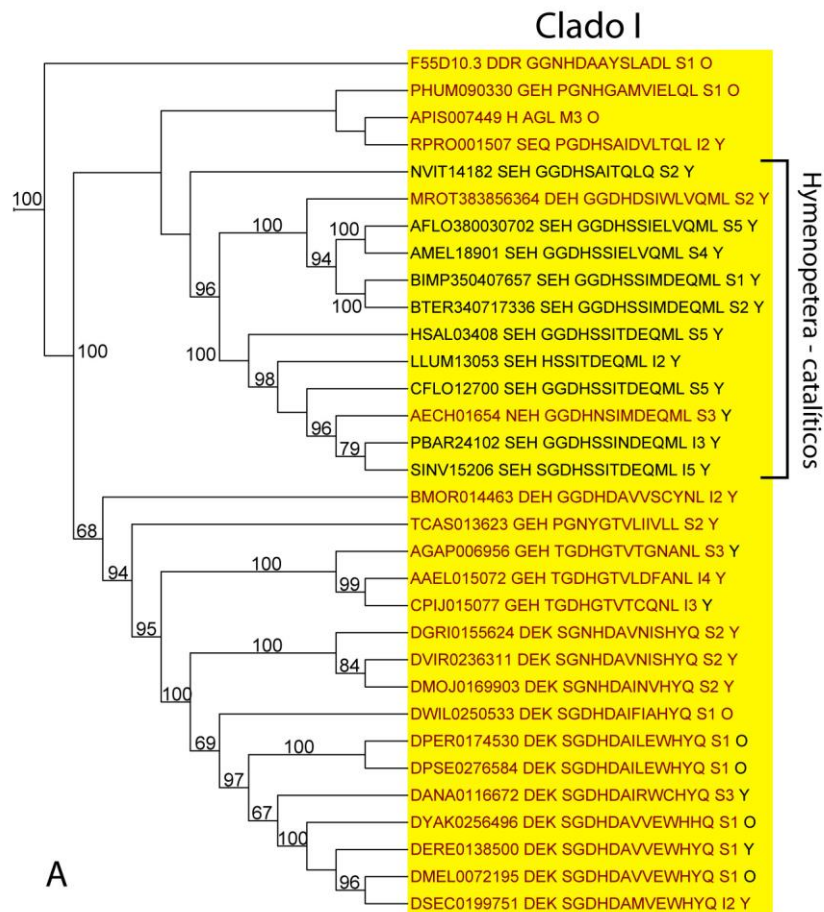


**Figura 4.28** A – Neurotactina (Clado M) e B – Gliotactina (Clado K). Apenas valores de *bootstrap* maiores que 60% estão mostrados. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

Diferente dos três clados citados acima (Neuroliginas – L, Neurotactinas – M e Gliotactinas – K), que já possuem indícios sobre sua função, o clado não caracterizado citado no trabalho de Oakeshott e colaboradores (Clado I) ainda permanece sem uma função conhecida (Oakeshott et al., 2005; 2010). Os resultados da presente análise apontam que esse gene já existia no ancestral comum dos insetos e nematódeos devido à presença do gene F55D10.3 de *C. elegans* (100% de *bootstrap*, Figura 4.29A). Nenhum gene foi encontrado em Arachnida o que pode indicar que houve perda desse gene nas espécies desse filo, ou que eles foram excluídos em etapas anteriores do trabalho. Com exceção de dez genes das espécies da ordem Hymenoptera que apresentaram a tríade catalítica predita como Ser-Glu-His, todos os outros genes apresentaram substituições em algum dos três resíduos catalíticos. Esses genes não possuem regiões hidrofóbicas que possam ser associadas ao ancoramento em membranas já que os resultados positivos obtidos pelo programa TMHMM se encontram na extremidade amino-terminal das sequências e por isso foram associados à presença de um peptídeo sinal de secreção. A maioria dos genes foram preditos como secretados ou intracelulares, tendo sinal mitocondrial apenas o gene da espécie *Acy. pisum* (Figura 4.29A). Um gene foi encontrado para cada espécie de inseto analisada e o relacionamento filogenético entre eles é igual ao esperado pelo relacionamento taxonômico conhecido para os grupos.

Algumas proteínas que haviam sido organizadas no clado neurotactina (Clado M) no trabalho publicado por Oakeshott e colaboradores em 2010 passaram a compor um novo clado (Não caracterizado 1 – Novo) com *bootstrap* de 94%. Nenhuma das treze espécies da ordem Hymenoptera e nem as espécies *I. scapularis*, *Acy. pisum*, *B. mori* e *C. elegans* estão presentes no clado. Esses resultados sugerem que esse gene possa ter sido perdido no ancestral da ordem Hymenoptera e também no da ordem Lepidoptera. No entanto, no caso de *I. scapularis* e *Acy. pisum* é mais provável que esse gene tenha sido excluído da análise em etapas anteriores, já que as espécies *M. occidentalis* e *R. prolixus* estão representadas nesse clado (Figura 4.29B). De acordo com os resultados obtidos pelos programas de predição, essas proteínas não são catalíticas, não possuem regiões transmembranares e são endereçadas para a via secretória. A exceção são os genes das espécies *R. prolixus* e *D. ananassae* (RPRO014475 e DANA0116442) que possuem a tríade Ser-Glu-His e, portanto, podem ser catalíticos (Figura 4.29B).



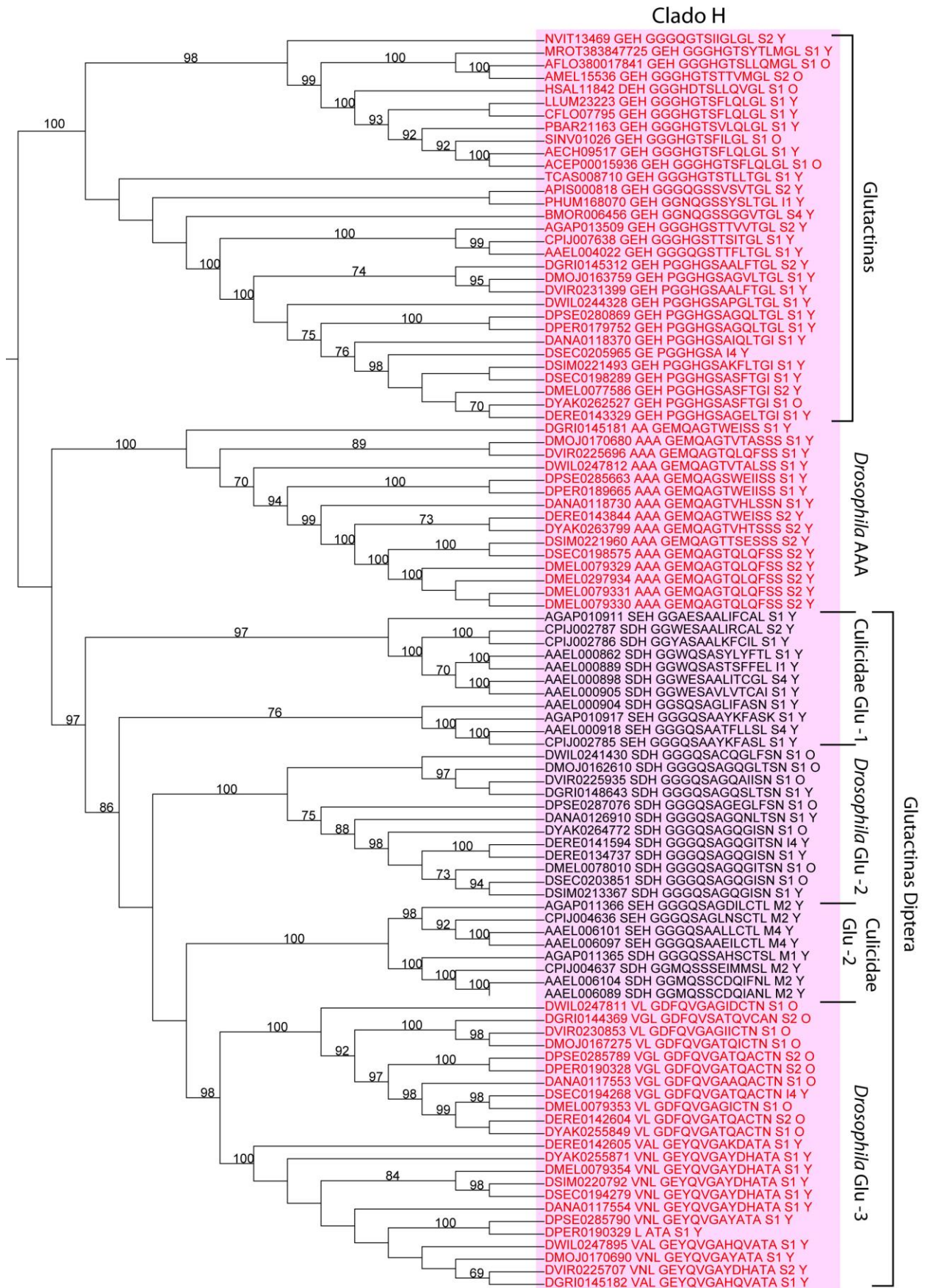


**Figura 4.29** A - Clado Não caracterizado (Clado I). B – Clado Não caracterizado 1 – Novo. Apenas valores de *bootstrap* maiores que 60% estão mostrados. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. A barra lateral delimita o grupo de genes que são catalíticos para as espécies da ordem Hymenoptera. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

A primeira proteína a ser identificada como glutactina foi caracterizada na espécie *D. melanogaster* em 1990. Ela foi descrita como uma proteína glicosilada semelhante as

entactinas de vertebrados e localizada na matriz extracelular de membranas basais (Olson et al., 1990). Seguindo a classificação mais recente, os membros deste grupo de genes são identificados como glutactinas (clado H) e estão divididos em dois grupos: Glutactinas e Glutactinas da ordem díptera (Oakeshott et al., 2010). Além desses, neste trabalho foi encontrado um clado que contém os ortólogos da glutactina caracterizada por Olson e colaboradores em 1990. Este clado aparece como grupo irmão das glutactinas de díptera, mas sem suporte estatístico (Figura 4.30). Neste clado as 12 espécies de *Drosophila* estão representadas (com quatro isoformas da espécie *D. melanogaster*). Todos os genes possuem substituições por alanina nos três resíduos da tríade catalítica e por isto o clado foi denominado *Drosophila* AAA. O clado Glutactina contém apenas genes de insetos que exibem substituições da serina nucleofílica por glicina, o que pode indicar a perda da atividade catalítica. A maioria dos genes é endereçada para a via secretória e não possui sequências hidrofóbicas associadas ao ancoramento em membranas. O clado das Glutactinas - Diptera contém genes de mosquito e *Drosophila* e é suportado por 97% de *bootstrap*. No grupo Culicidae Glu-1 foi possível verificar uma pequena expansão para *Ae. aegypti* (Figura 4.30) em relação às outras duas espécies de mosquito, o que mostra que duplicações deve ter ocorrido para essa espécie ou que houve perda desses genes em *Cu. quinquefasciatus* e *An. gambiae*. O clado Culicidae Glu-2 é irmão do clado *Drosophila* Glu-3 e contém dois genes de cada espécie com exceção de *Ae. aegypti*, que possui quatro. A previsão da tríade catalítica na maioria dos grupos (Culicidae Glu-1 e 2 e *Drosophila* Glu-2) indica que estes genes ainda possam ter algum tipo de função catalítica. No entanto, no clado *Drosophila* Glu-3 houve a perda dos resíduos catalíticos (Figura 4.30). Endereçamento para a mitocôndria foi identificado para o clado Culicidae Glu-2, entretanto, para os demais genes foi identificado sinal para a via secretória.

O clado denominado como Esterases de *C. elegans*, apesar de não possuir suporte estatístico, se agrupa, na árvore de máxima verossimilhança, com os clados de glutactina indicando que possam ter uma mesma origem (Figura 4.26, e detalhes na Figura S1).



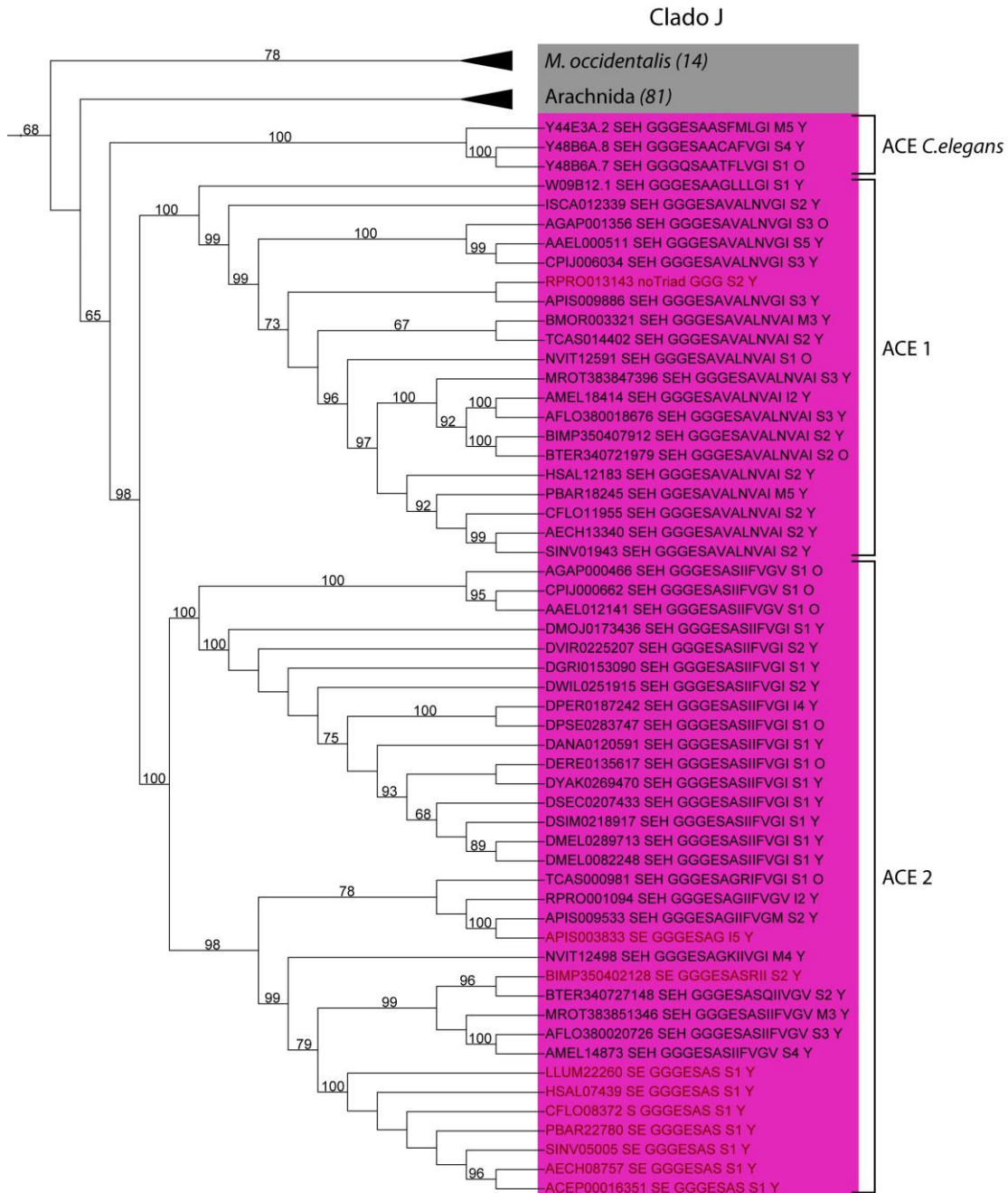
**Figura 4.30** Clado das glutactinas (Clado H). As barras laterais delimitam os diferentes clados já descritos na literatura. Os genes em vermelho apresentam substituições em algum

dos três resíduos da tríade catalítica ou ausência de algum deles. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado das acetilcolinesterases (AChE) (Clado J) contém proteínas que desempenham sua função no SNC degradando o neurotransmissor acetilcolina na maioria dos metazoários. Para tal função a preservação da tríade catalítica é fundamental (Gibney et al., 1990). Esse clado é monofilético e é suportado por valor de *bootstrap* de 68% (Figura 4.31). Os genes desse clado são majoritariamente catalíticos apresentando ácido glutâmico como resíduo ácido da tríade catalítica. A maioria apresenta sinal de endereçamento para a via secretória e nenhum apresentou região transmembranar que pudesse ser associada com o ancoramento em membranas. A ausência de região transmembranar já era esperada, pois existem indícios de que esta enzima fica ancorada na membrana, por causa de interação com uma *glicoproteína* de ancoramento (Icardona & Rosenberry, 1996). Foram identificados cinco ramos nesse clado, onde apenas o ramo denominado AChE - Arachnida não é bem suportado. Os cladogramas AChE - *C. elegans* e *M. occidentalis* possuem *bootstrap* de 100% e 78%, respectivamente, e contêm genes parálogos para cada uma das espécies (3 e 14 respectivamente). O clado AChE - Arachnida contém genes tanto da espécie *M. occidentalis* quanto *I. scapularis*. Em alguns dos ramos internos, é possível identificar genes ortólogos para as duas espécies além de genes parálogos, indicando que sucessivos eventos de duplicação gênica ocorreram, tanto no ancestral comum dessas duas espécies quanto após o processo de especiação de ambas. Em um trabalho recente, Temeyer e colaboradores usaram 52 sequências de AChE da espécie *I. scapularis* para mostrar a diversidade dessas enzimas e discutir o quão complexo pode ser o mecanismo de resistência nas espécies de carrapato (Temeyer et al., 2013). Na presente análise o clado AChE - Arachnida contém 15 genes da espécie *M. occidentalis* e 66 da espécie *I. scapularis*. O clado AChE 1 tem suporte estatístico de 100% e contém genes ortólogos das espécies das classes Insecta, Arachnida e de *C. elegans*, indicando que esses genes possivelmente são os mais antigos para o clado das acetilcolinesterases. Apenas o gene da espécie *R. prolixus* (RPRO013143) não apresenta o domínio COesterase inteiro, o que pode ser devido a problemas na predição gênica (a predição gênica usada para este inseto foi do tipo “rascunho” e somente durante a escrita desta dissertação a predição gênica estável (V1.1) foi liberada pelo VectorBase). Como era esperado, a ausência de genes das doze espécies de *Drosophila* nesse ramo reforça os dados da literatura que mostram que esse gene foi perdido no ancestral comum dos dípteros superiores (Huchard et al., 2006). O clado AChE 2 possui *bootstrap* de 100% e contém genes



ortólogos das espécies de insetos apenas. Os genes de todas as espécies de formiga, de *B. impatiens* e de *Acy. pisum* não possuem o domínio COesterase completo (também possivelmente devido a erros de predição gênica) e por isso não possuem todos os três resíduos catalíticos da tríade. O fato do clado AChE 2 conter apenas genes de insetos (Figura 4.31) evidencia que o mesmo foi originado de uma duplicação de AChE 1 no ancestral dos insetos.



**Figura 4.31** Clado das acetilcolinesterases (Clado J). Os ramos denominados AChE1, AChE2 e AChE *C. elegans* estão delimitados por barras laterais e sinalizados pela cor rosa. Os clados AChE - *M. occidentalis* e Arachnida foram colapsados e estão indicados pela cor cinza (número de genes encontrado entre parênteses). Os genes em vermelho apresentam

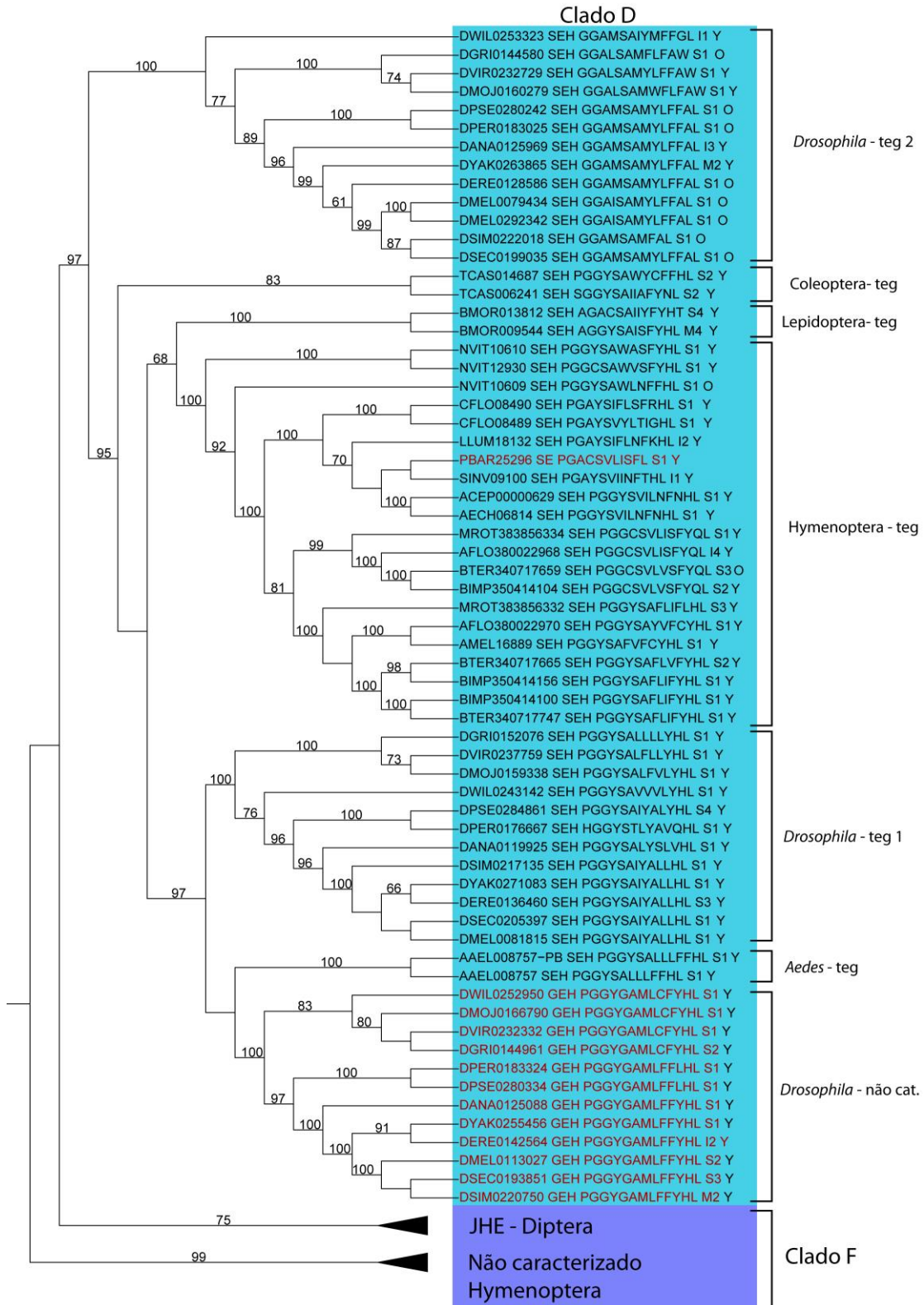
substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

#### 4.4.2 *Processadoras de Hormônios e feromônios*

Nesse grupo estão organizados quatro principais clados (D-G; Figura 4.26) identificados previamente por Oakeshott e colaboradores (2005; 2010). Os membros desses clados estão associados com a degradação de hormônios e feromônios que sinalizam processos fisiológicos importantes como, por exemplo, a ecdise. Os genes mais estudados nesse grupo são os das esterases de hormônio juvenil (JHE do inglês *Juvenile Hormone Esterase*) que degradam hormônio juvenil (HJ), possibilitando o processo de metamorfose para adulto. Todos os clados desse grupo contém majoritariamente proteínas catalíticas, secretadas e sem regiões relacionadas com o ancoramento em membranas. O clado das Tegumentares (clado D) é o único dos grandes clados determinados por Oakeshott e colaboradores (2005; 2010) que é monofilético com bom suporte estatístico (97% de *bootstrap*). O clado E não é monofilético, mas contém cinco clados monofiléticos, três com suporte estatístico (Beta-esterases – *Drosophila*, Beta-esterases – conservada e JHE – Hemiptera) e dois sem suporte estatístico (Beta-esterases – Multi ordens e Não caracterizado – *T. castaneum*). O clado F não é monofilético e organiza dois clados monofiléticos com bom suporte estatístico (JHE – Diptera e Não caracterizado – Hymenoptera). Nesta análise o clado JHE – Diptera parece estar mais relacionado com o clado Tegumentar (clado D) do que com o clado Não caracterizado – Hymenoptera. O clado G é monofilético (sem suporte estatístico) e agrupa cinco clados bem suportados de JHE de lepidópteros e um, não monofilético, que contém apenas genes da espécie *T. castaneum*.

O clado das Tegumentares (clado D; *bootstrap* de 97%) contém 64 genes apenas de espécies de holometábolos (Figura 4.32). Das três espécies de mosquito, apenas genes de *Ae. aegypti* foram encontrados. A maioria dos genes são catalíticos e apenas o gene PBAR25296, que não possui a histidina da tríade, e os genes contidos no ramo denominado de *Drosophila* – não cat. onde foi observado um resíduo de glicina no lugar da serina nucleofílica para as doze espécies de *Drosophila*, devem ter perdido a função catalítica. O clado *Drosophila* – teg 1 fica numa posição mais interna e é grupo irmão do clado *Drosophila* – não cat e *Aedes* – teg. O clado *Drosophila* – teg 2 fica em uma posição mais externa. As espécies das ordens Lepidoptera e Coleoptera possuem dois genes cada que ficam em dois ramos (Lepidoptera – teg e Coleoptera – teg) com suporte de 100% e 83% respectivamente. Os genes das treze

espécies da ordem Hymenoptera ficam no ramo denominado de Hymenoptera - teg que possui *bootstrap* de 100% (Figura 4.32).

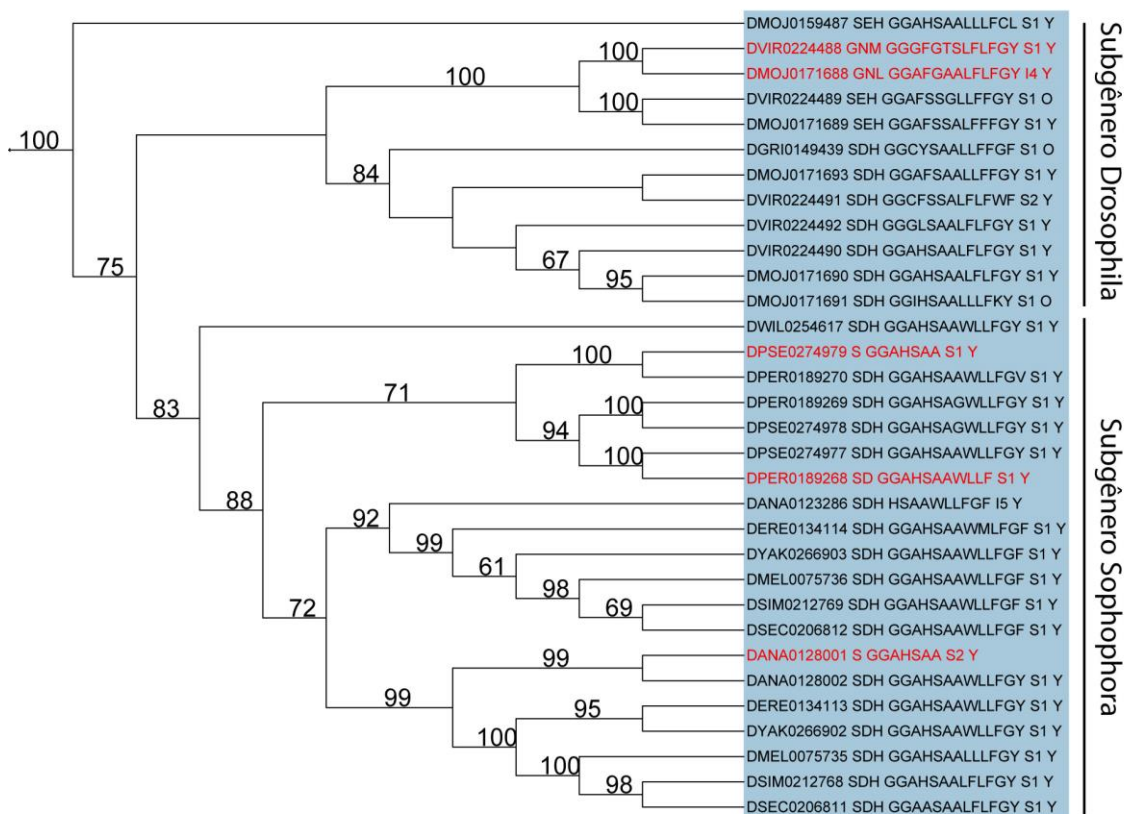


**Figura 4.32** Clado Tegumentares (Clado D). As barras laterais delimitam os clados dos grupos taxonômicos identificados. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. Os clados JHE –

Diptera e Não caracterizado – Hymenoptera (clado F) estão colapsados. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado das beta-esterases (clado E) foi um dos que mais sofreu alterações em relação a Oakeshott e colaboradores, 2010, como pode ser visto na Figura 4.26. Não existe suporte para corroborar a relação entre os cinco clados encontrados na presente análise com o que foi proposto pelos referidos autores. O clado Beta-esterases – *Drosophila* é monofilético e contém genes envolvidos no processamento de moléculas de feromônios no sistema sensorial de indivíduos machos (Chertemps et al. 2012). A maioria dos genes são secretados, não transmembranares e um dado interessante é que, ao contrário da maioria dos clados encontrados nesse trabalho, esse possui proteínas com ácido aspártico como resíduo ácido da tríade catalítica nos membros catalíticos, enquanto na maioria é encontrado o ácido glutâmico (Figura 4.33). Algumas diferenças de conteúdo gênico (principalmente referente ao número de genes catalíticos encontrados) foram observadas na presente análise quando comparadas com o trabalho de Robin e colaboradores em 2009. Os autores encontraram dois genes para a espécie *D. mojavensis* e no presente trabalho apenas um gene que possuía dois domínios COesterase não catalíticos (DMOJ0171688) foi encontrado. Para *D. persimilis* e *D. pseudoobscura* os autores encontraram três genes catalíticos, enquanto que apenas os genes DPSE0274979 e DPER0189268 foram considerados não catalíticos nesta tese. Além destes, para *D. ananassae*, o gene DANA128001 também não possui o domínio COesterase inteiro. Essas diferenças podem ser fruto do passo de exclusão feito no início das análises ou de modificações que ocorreram entre as versões de predição usadas para esta análise e o trabalho de Robin e colaboradores (2009).





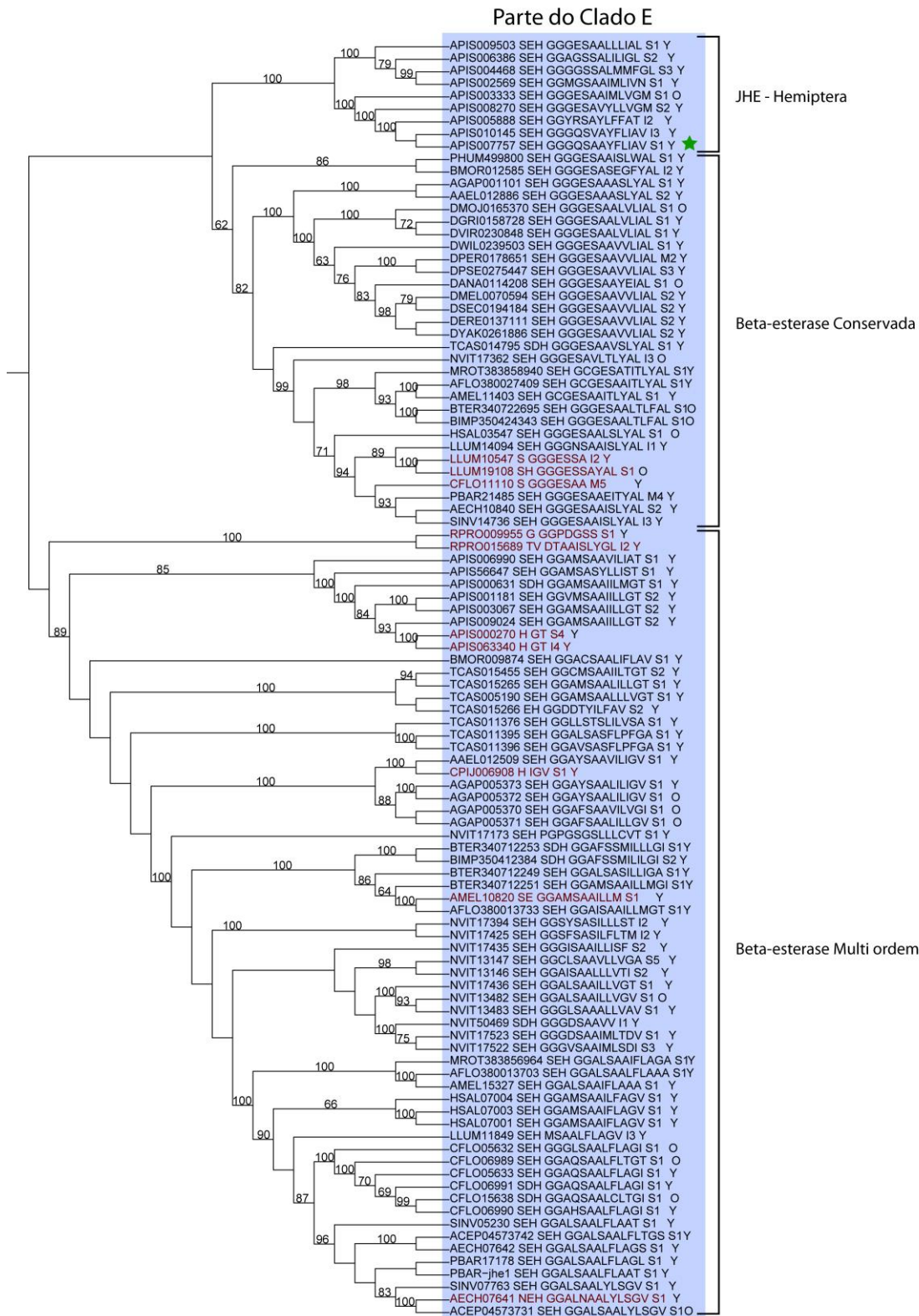
**Figura 4.33** Clado Beta-esterases - *Drosophila*. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. As barras laterais indicam os dois subgêneros de *Drosophila* citados no trabalho de Robin et al., 2009. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado JHE – Hemiptera é suportado por 100% de *bootstrap* e contém nove genes da espécie *Acy. pisum*. Esse clado se agrupa próximo às beta-esterases conservadas na árvore de máxima verossimilhança, mas sem suporte estatístico. Todos os genes são catalíticos e apresentam um resíduo de ácido glutâmico compondo a tríade catalítica. A maioria é endereçada para a via secretória, com exceção de APIS005888 e APIS10145, que são intracelulares. Nenhuma região hidrofóbica associada ao ancoramento em membranas foi encontrada. O nome do clado se deve ao trabalho de Ishikawa e colaboradores (2012) que reconstruíram a via regulatória de HJ na espécie *Acy. pisum* e identificaram os genes APIS007757 e APIS001447 como responsáveis pela degradação desse hormônio. A ausência de genes de *R. prolixus* nesse clado (Figura 4.34) pode ser reflexo da etapa inicial de exclusão.

O clado Beta-esterase conservada é monofilético e é suportado por um *bootstrap* de 62% (Figura 4.34). Esse clado contém, em sua maioria, genes catalíticos com ácido glutâmico compondo a tríade, exceto na espécie *T. castaneum* onde foi encontrado um

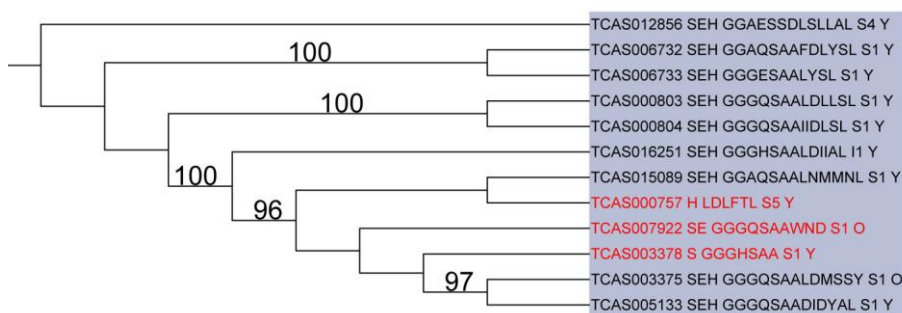
resíduo de ácido aspártico. A maioria dos genes possui sinal de secreção e não possui regiões hidrofóbicas associadas ao ancoramento em membranas. Nesse clado, quando presente no genoma, apenas um gene para cada espécie foi encontrado, com exceção da espécie *L. humile* que possui três genes, embora apenas um deles apresente o domínio COesterase completo. Não foram encontrados genes para as espécies *Atta cephalotes*, *Cu. quinquefasciatus*, *Acy. pisum* e *R. prolixus*. A espécie *Ca. floridanus* possui apenas um gene com domínio COesterase incompleto. A presença de apenas um gene para cada espécie indica que duplicações gênicas não são muito comuns para este gene, ao contrário do que ocorre em outros clados de beta esterases.

O clado Beta-esterase - Multi ordem possui *bootstrap* de 86% e contém genes de todas as espécies de insetos incluídas nesse trabalho com exceção das espécies de *Drosophila* e *P. humanus* (Figura 4.34). É possível que estes genes tenham sido perdidos nestas linhagens. Um clado com dois genes de *R. prolixus* (RPRO00995 e RPRO015889) é grupo irmão deste clado, e é possível que a falta do domínio COesterase completo nestes genes tenha dificultado a análise. Os genes APIS000270, APIS063340, CPIJ006908 e AMEL10820 também não possuem o domínio COesterase inteiro e o gene AECH07641 possui uma asparagina no lugar da serina nucleofílica. Para as espécies *N. vitripennis*, *Ca. floridanus*, *T. castaneum* e *An. gambiae* podem ser observadas expansões maiores quando comparadas com as demais espécies (Figura 4.34).



**Figura 4.34** Clado JHE – Hemiptera, Clado Beta-esterase – Conservada e Clado Beta-esterase - Multi ordem. A estrela verde indica o gene APIS007757 que foi caracterizado por participar da via de degradação de hormônio juvenil na espécie *Acy. pisum* (Ishikawa et al., 2012). Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. As barras laterais delimitam o nome dado a cada clado no presente trabalho. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado Não caracterizado – *T. castaneum* não possui bom suporte estatístico mas contém doze genes da espécie *T. castaneum* sem função conhecida. A maioria dos genes é catalítica e possui ácido glutâmico compondo a tríade catalítica. No entanto, os genes TCAS000757, TCAS007922 e TCAS003378 não possuem o domínio COesterase completo e por isso não apresentam todos os três resíduos da tríade. Apenas o gene TCAS16251, que foi predito como intracelular, não teve sinal de endereçamento para a via secretória. Nenhuma região hidrofóbica associada com o ancoramento em membranas foi encontrada (Figura 4.35).



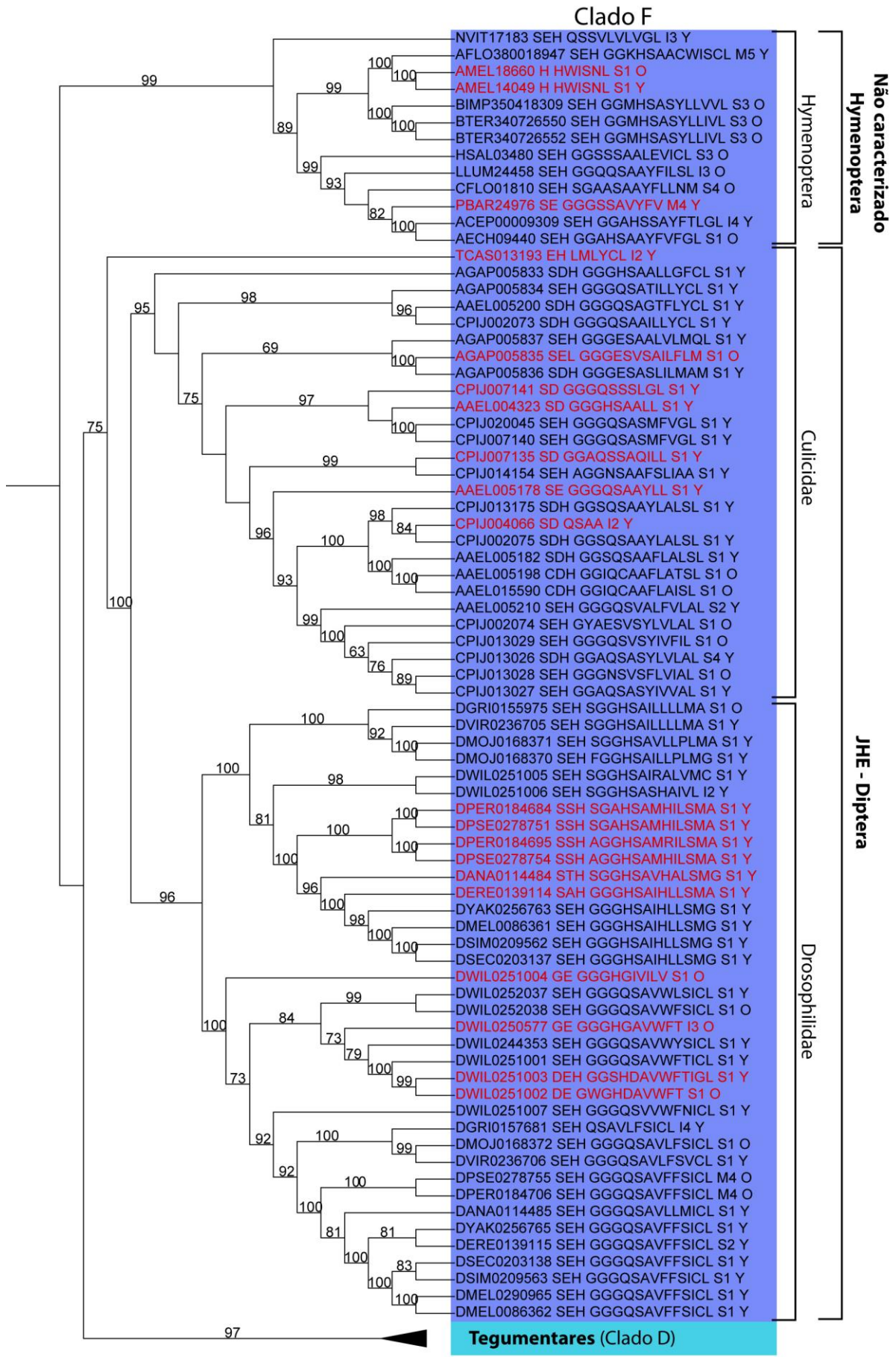
**Figura 4.35** Clado Não caracterizado – *T. castaneum*. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado F é formado por dois clados próximos (JHE – Diptera e Não caracterizado – Hymenoptera), mas sem suporte estatístico (Figura 4.36). O clado denominado de JHE – Diptera tem suporte estatístico de 75% e contém 56 genes das quinze espécies de dípteros incluídas nesse trabalho além de um gene de *T. castaneum* (TCAS13193) que se localiza na parte mais externa do clado. Esses genes estão envolvidos com a degradação de HJ que é o hormônio responsável pelo processo de metamorfose. A maioria dos genes é catalítica e possui ácido aspártico como resíduo ácido da tríade, o que difere do padrão visto para a maioria dos clados encontrados. Esses genes são endereçados para a via secretória e não possuem regiões associadas com o ancoramento em membranas. Os genes CPIJ007141, CPIJ004323, CPIJ004066 e CPIJ007135 de *Cu. quinquefasciatus*, AAEL005178 de *Ae. aegypti*, DWIL0251004, DWIL0250577 e DWIL0251002 de *D. willistoni* e TCAS013193 de *T. castaneum* não possuem o domínio COesterase inteiro. Os genes DPER0184684, DPSE0278751, DPSE0278754 e DPER084695 apresentaram uma serina no lugar do resíduo ácido da tríade; para o gene DERE0139114 foi encontrada uma alanina e para o gene DANA0114484 uma treonina. O gene DWIL0251003 apresentou uma arginina no lugar da

serina nucleofílica. Essas alterações podem indicar perda de função para esses genes, já que não possuem as características descritas por Cygler e colaboradores (1993) necessárias para possuir a capacidade catalítica. A topologia desse clado mostra que os genes de *Drosophila* e dos mosquitos ficam agrupados em ramos monofiléticos distintos, o que sugere que eventos de duplicação mais recentes ocorreram após a divergência dessas famílias (Figura 4.36). O gene AAEL005200 da espécie *Ae. aegypti* foi relatado como o principal envolvido na degradação de HJ (Bai et al., 2007). Genes das espécies *An. gambiae* e *Cu. quinquefasciatus* se agrupam com este gene de *Ae. aegypti* com 98% de *bootstrap* (Figura 4.36) indicando que a função descrita para o gene de *Ae. aegypti* pode ser conservada para as outras duas espécies de mosquito.

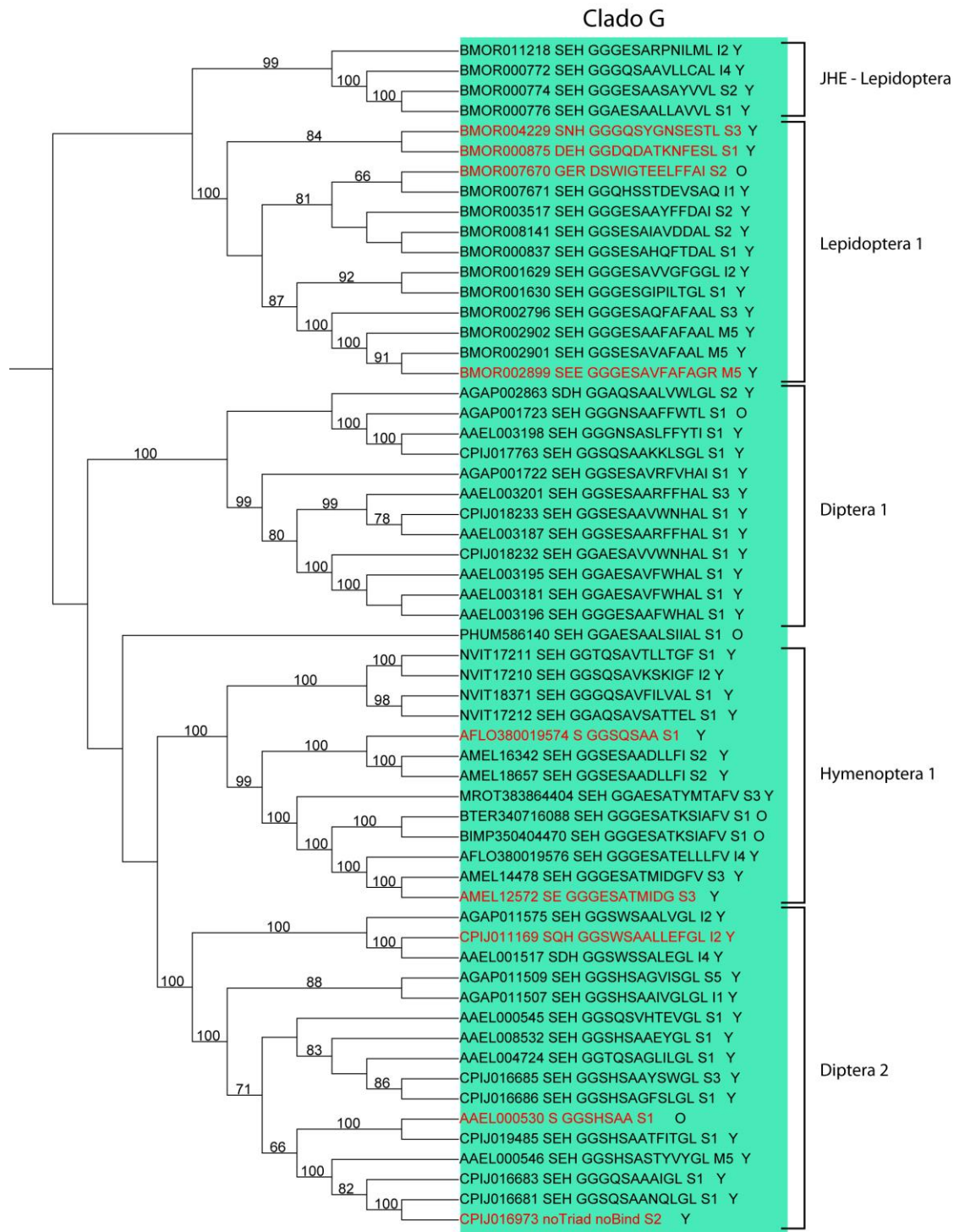
O clado Não caracterizado – Hymenoptera é suportado por *bootstrap* de 98% e possui 13 genes de onze espécies da ordem Hymenoptera. A maioria das espécies possui apenas um gene, exceto *Ap. mellifera* e *B. terrestris* que possuem dois e *M. rotundata* e *S. invicta* que não possuem nenhum membro. Dois genes de *Ap. mellifera* (AMEL18660 e AMEL14049) e um de *P. barbatus* (PBAR24976) não possuem todos os aminoácidos da tríade podendo isso indicar perda de função catalítica. A maioria dos genes são endereçados para a via secretória e não apresentam região hidrofóbica associada ao ancoramento em membranas (Figura 4.36).





**Figura 4.36** Clado F ( Clado Não caracterizado – Hymenoptera e Clado JHE – Diptera). As barras laterais indicam a representação do grupo taxonômico e o clado identificado no presente trabalho. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. O clado Tegumentar (Clado D) está colapsado. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado JHE Lepidoptera (Clado G) sofreu alterações em relação à classificação feita por Oakeshott e colaboradores (2010). Esse clado é constituído por cinco clados monofiléticos que possuem sequências de espécies das ordens Lepidoptera, Diptera e Hymenoptera, além de um gene de *P. humanus* (PHUM586140). O clado JHE Lepidoptera possui genes envolvidos na degradação de HJ em lepidópteros e foi usado como referência para nomear o grupo. A maioria dos genes são catalíticos e possui sinal para a via secretória. Nenhuma região hidrofóbica associada com o ancoramento em membranas foi encontrada. Os genes de Lepidoptera formam dois clados bem suportados que se agrupam na árvore de máxima verossimilhança sem suporte estatístico. O clado Diptera 1 é bem suportado com *bootstrap* de 100% e contém doze genes de mosquitos com uma representação maior para a espécie *Ae. aegypti*. O clado Diptera 2 é suportado por *bootstrap* de 100% e, assim como Diptera 1, possui apenas genes de mosquitos mas com representação proporcional para as três espécies. O ramo Hymenoptera 1 é suportado por *bootstrap* de 100% e não possui nenhum gene de formiga, o que indica perda desse gene no ancestral da família Formicidae ou que estes genes não foram recuperados nas análises. É possível ver expansões para *N. vitripennis* e *Ap. mellifera* (Figura 4.37).

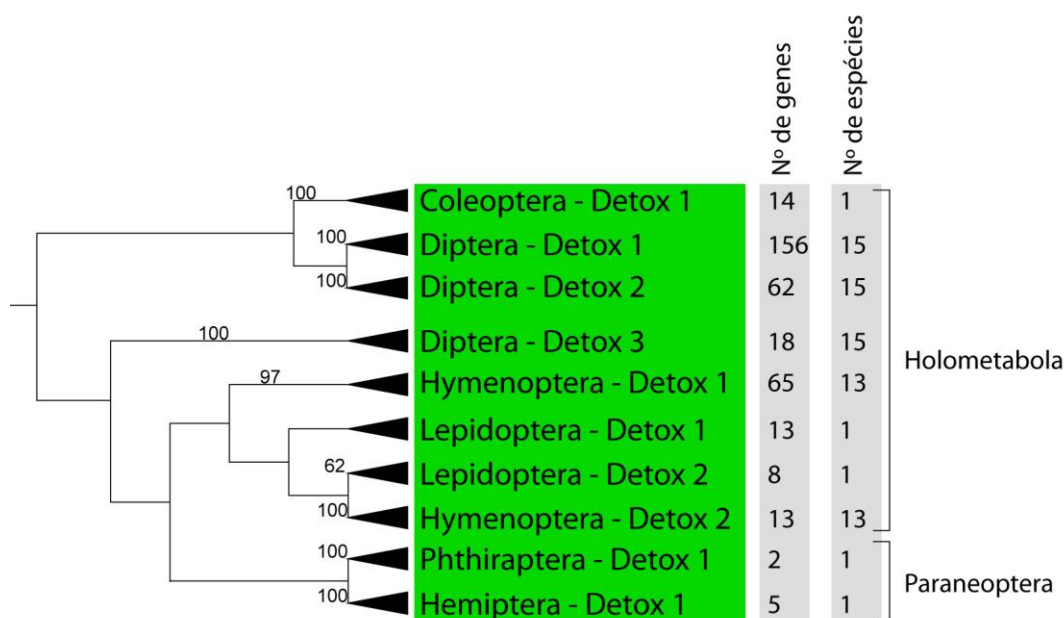


**Figura 4.37** Clado JHE Lepidoptera (Clado G). As barras laterais indicam os clados identificados de acordo com a representação do grupo taxonômico (JHE – Lepidoptera, Diptera 1 e 2, Lepidoptera 1 e Hymenoptera 1). Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).



### 4.4.3 Detoxificantes

Os genes que constituem esse grupo são catalíticos e possuem ácido glutâmico como resíduo ácido da tríade na maioria das sequências. Nenhuma região hidrofóbica associada com o ancoramento em membranas foi encontrada e a predição da localização celular foi majoritariamente intracelular. Na classificação feita em 2010 por Oakeshott e colaboradores, esse grupo foi dividido em 3 clados (Clado A, B e C), mas tal organização não parece ser a melhor para esse grupo, de acordo com a presente análise. Como pode ser visto na Figura 4.38, que mostra um panorama resumido da organização dos genes desse grupo, os ramos com bom suporte estatístico contêm genes de espécies que pertencem à mesma ordem taxonômica (Figura S3). Todos os genes contidos nesse grupo foram considerados como membros envolvidos em processos de detoxificação, o que corresponde aos clados A, B e C descrito por Oakeshott e colaboradores em 2005 e 2010. Os clados bem suportados foram nomeados de acordo com a ordem taxonômica das espécies representadas neles, seguidas do termo “Detox” (abreviatura para a palavra detoxificante) e o número do clado. A ordem Diptera possui três clados (Diptera – Detox 1, 2 e 3), a ordem Hymenoptera e Lepidoptera possuem dois cada (Hymenoptera – Detox 1 e 2 e Lepidoptera – Detox 1 e 2) e as ordens Coleoptera, Hemiptera e Phthiraptera possuem um cada (Coleoptera – Detox 1, Hemiptera – Detox 1 e Phthiraptera – Detox 1).



**Figura 4.38** Grupo Detoxificante com os principais clados colapsados. Ao lado dos clados o número de genes e o número de espécies para cada clado. As barras laterais mostram quais ordens pertencem aos grupos Holometabola e Paraneoptera.

O clado Diptera – Detox 1 é suportado por *bootstrap* de 100% e contém 156 genes das quinze espécies dessa ordem que são majoritariamente catalíticos e intracelulares, mas com alguns representantes mitocondriais. Esse clado possui um maior número (150) de sequências de *Drosophila* e os seis genes de mosquito formam um clado monofilético bem suportado (100%) mais externo (Figura 4.39). A disposição das espécies no clado dos genes de mosquito mostra que um evento de duplicação gênica ocorreu no ancestral do grupo já que os genes das diferentes espécies são mais próximos entre si. Os clados com genes de *Drosophila* são bem suportados e fica evidente que duplicações ocorreram tanto nos ancestrais de todas as espécies quanto dentro das diferentes linhagens (Figura 4.39). Alguns clados possuem apenas um gene de cada espécie, mas em sua maioria, outras duplicações com maiores ou menores expansões podem ser observadas para cada espécie. Em poucos casos não foram encontrados genes para algumas das espécies, o que pode ser devido a perdas durante a evolução destas linhagens ou simplesmente por dificuldade de se encontrar o gene no genoma da espécie.

O clado Diptera – Detox 2 é suportado por *bootstrap* de 100% e contém 62 genes das quinze espécies dessa ordem que são majoritariamente catalíticos mas não exibem nenhum padrão claro em relação à localização celular. Ao contrário do clado anterior, neste há uma maior representação de sequências de mosquito do que *Drosophila*. Os genes de mitocondriais formam um clado com suporte bom (100%) que se agrupa com 91% de *bootstrap* com um clado menor de mosquitos que possui duas sequências de cada uma das três espécies intracelulares de mosquito analisadas. Outros 41 genes de mosquitos ficam agrupados em um segundo clado com 82% de *bootstrap* (Figura 4.39). Entre os mosquitos, podemos perceber que a espécie *Cu. quinquefasciatus* possui um número maior de duplicações (23 genes contra 14 de *Ae. aegypti* e 10 de *An. gambiae*). A maioria dos genes possui ácido glutâmico (E) como resíduo ácido da tríade catalítica mas, no entanto, em três genes (um de cada espécie de mosquito (AGAP006700, AAEL011944 e CPIJ004752)) o resíduo de ácido aspártico (D) foi encontrado. Esses três genes formam um clado monofilético com 100% de *bootstrap* (Figura 4.39). O clado Coleoptera – Detox 1 possui apenas sequências de *T. castaneum*, o único representante desta classe amostrado. Este clado é suportado por *bootstrap* de 100% e contém quatorze genes majoritariamente catalíticos e intracelulares.

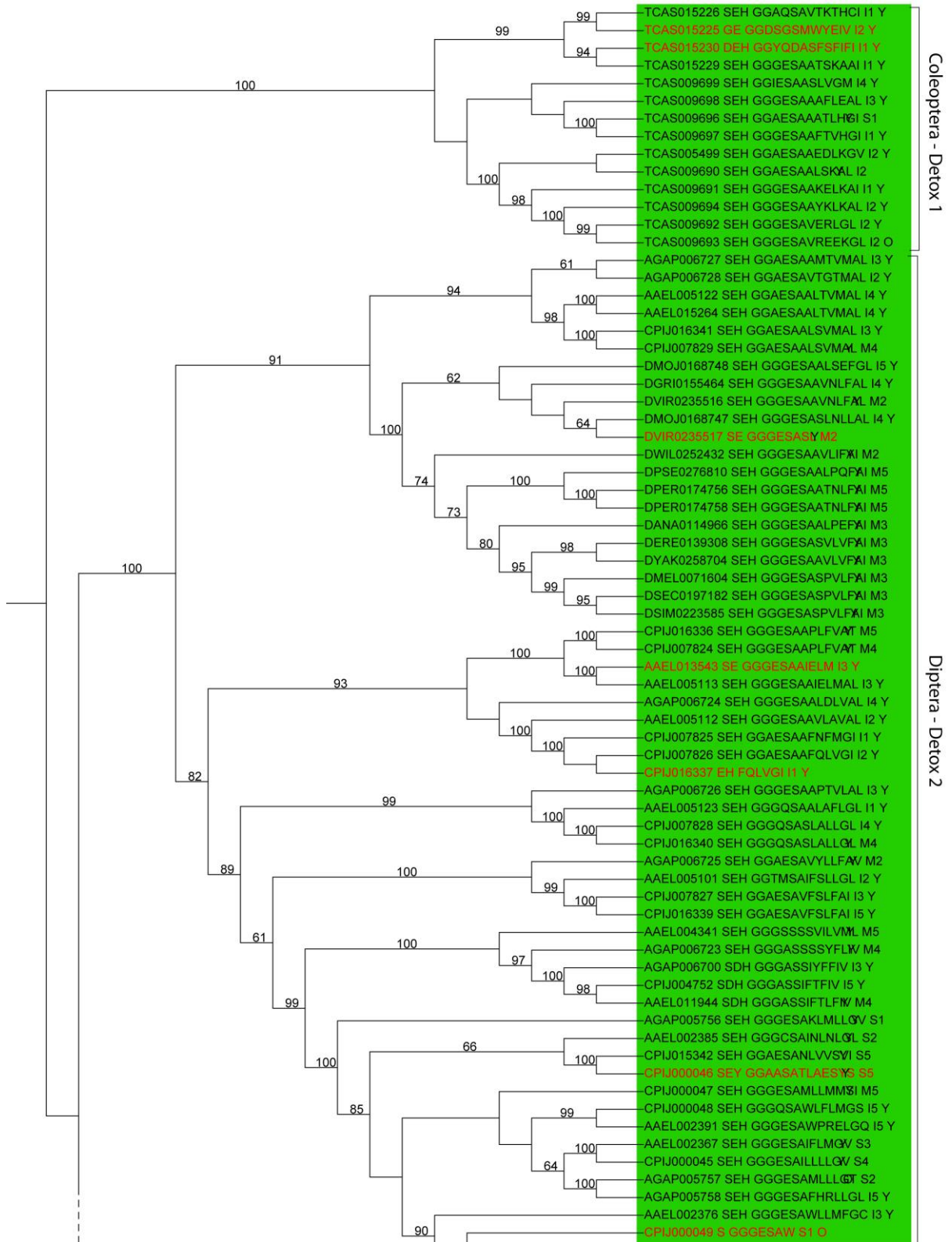
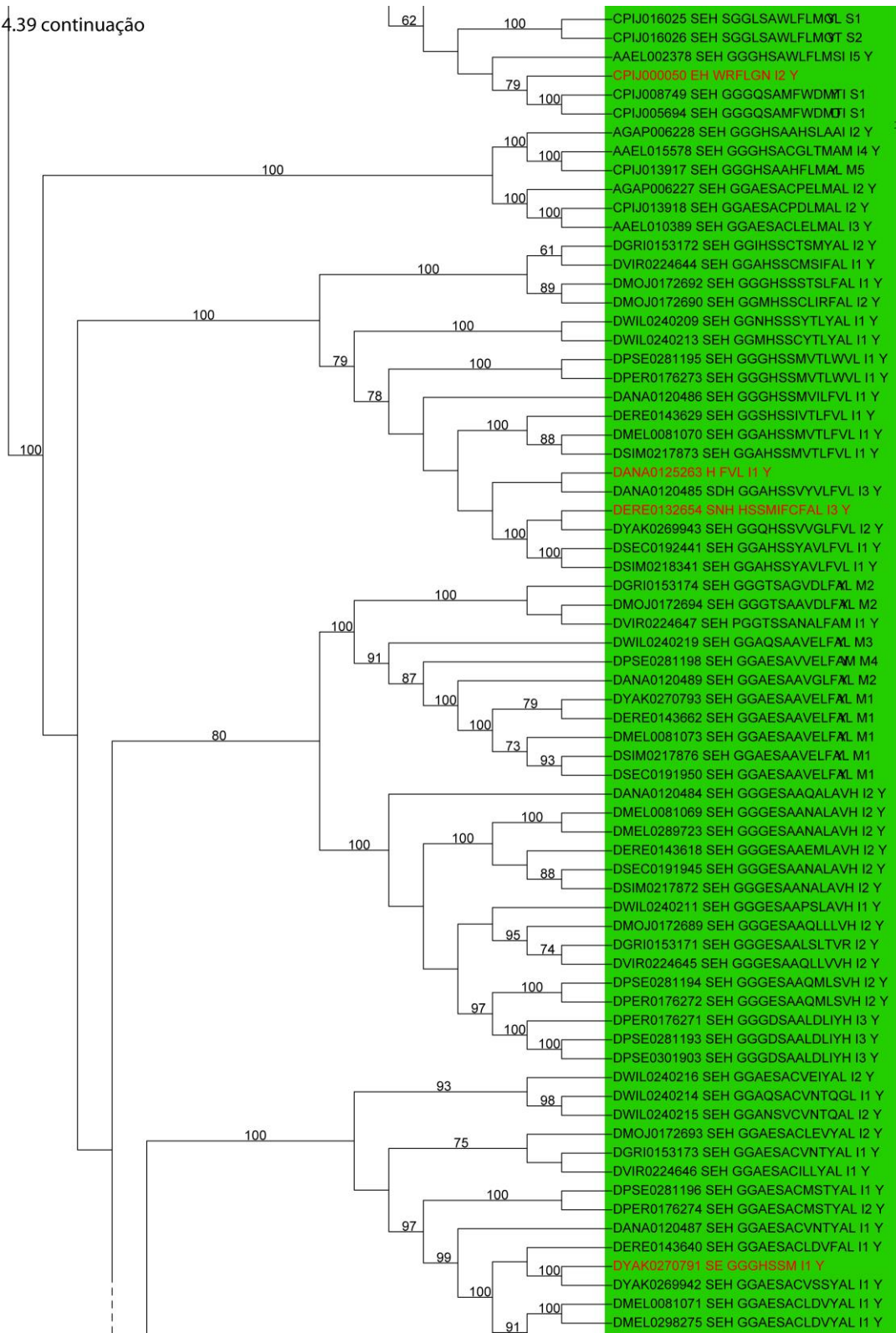


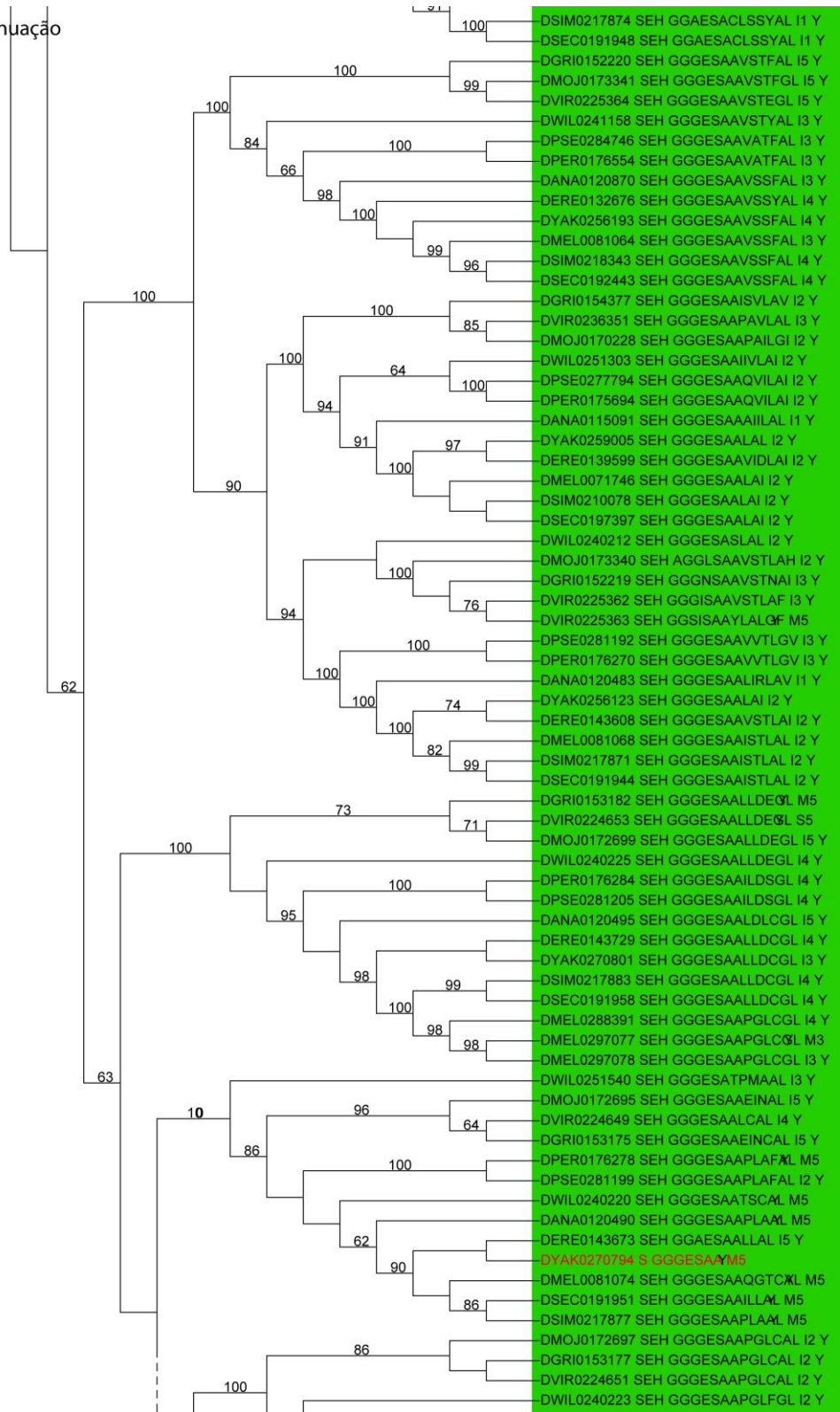
Figura 4.39 continuação



Diptera - Detox 1

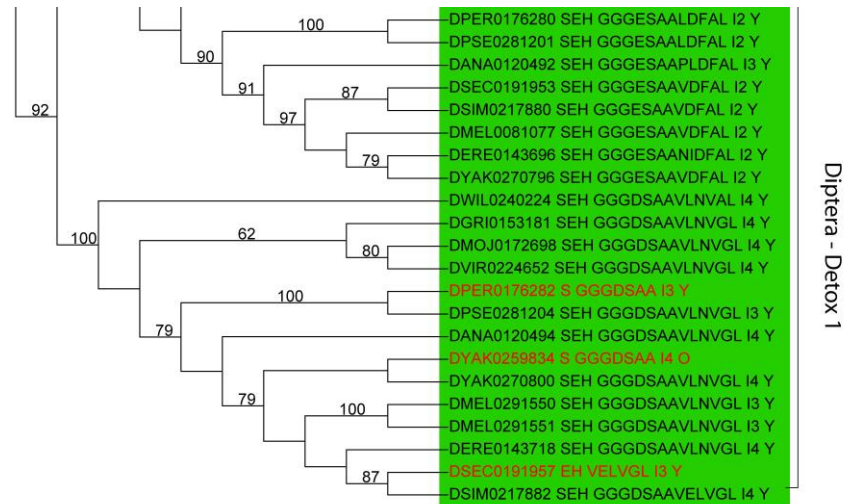


Figura 4.39 continuação



Diptera - Detox 1

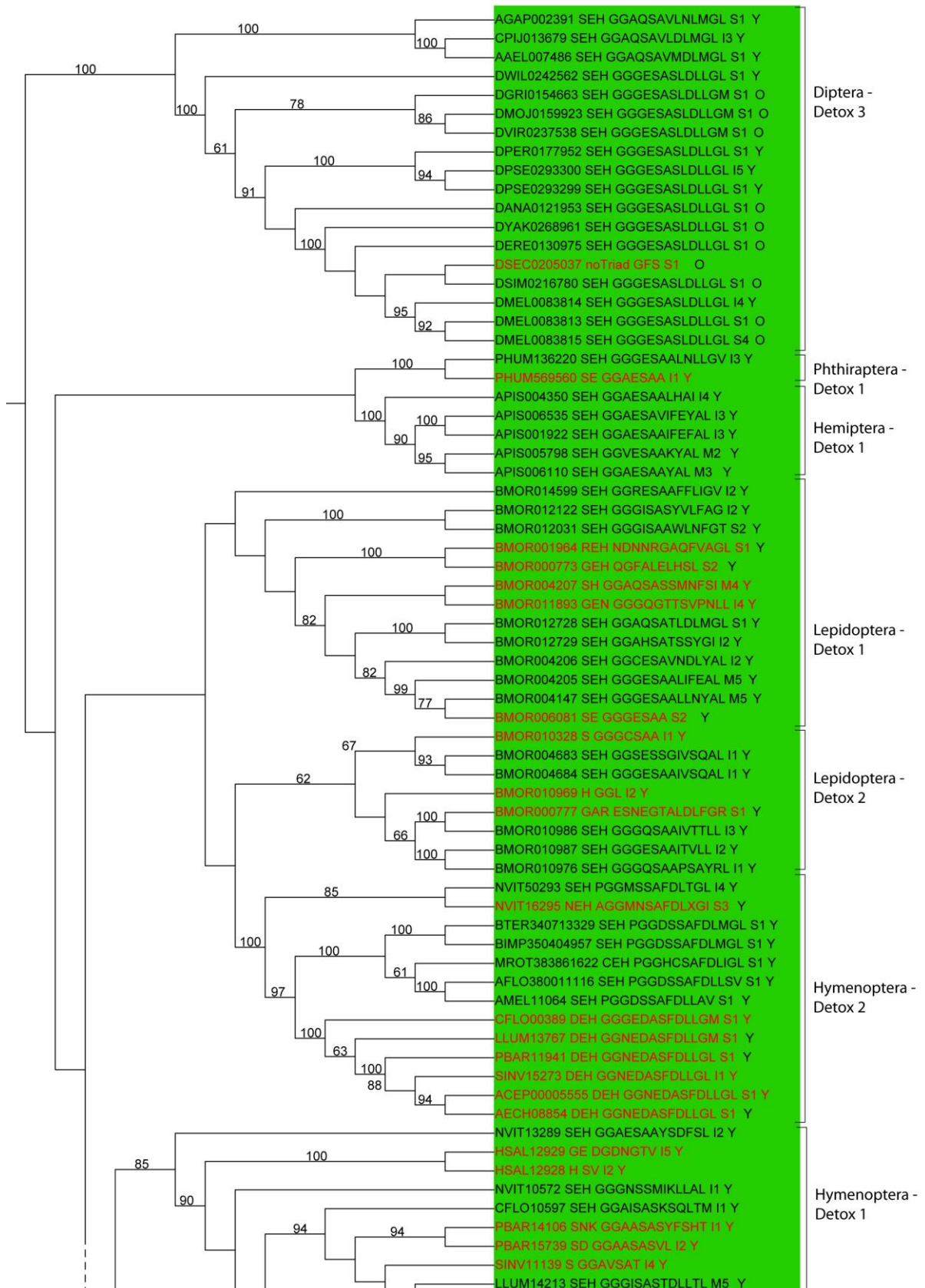
Figura 4.39 continuação



**Figura 4.39** Clado Diptera – Detox 1, Diptera – Detox 2 e Coleoptera – Detox 1. Apenas *bootstrap* maiores que 60% estão mostrados. As barras laterais delimitam os três clados. Os genes em vermelho apresentam substituições em algum dos três resíduos da tríade catalítica ou ausência de algum deles. Para a melhor identificação das espécies utilize o material suplementar (Tabela S1).

O clado Diptera – Detox 3 é suportado por *bootstrap* de 100% e contém um gene para cada uma das 15 espécies da ordem Diptera com exceção das espécies *D. melanogaster* e *D. pseudoobscura* que possuem três e dois genes respectivamente. Todos os genes são majoritariamente catalíticos e secretados com exceção da espécie *D. sechellia* (DSEC0205037) que não possui nenhum dos resíduos catalíticos (Figura 4.40). Novamente, neste caso os mosquitos aparecem como grupo externo ao de *Drosophila*. O clado Hymenoptera – Detox 1 é suportado por *bootstrap* de 97% e possui 65 genes das 13 espécies dessa ordem que são na maioria catalíticos e intracelulares. No entanto, neste clado diversos genes não continham todos os resíduos da tríade (Figura 4.40), sugerindo uma grande quantidade de pseudogenes ou genes com predição parcial. Diversas expansões podem ser encontradas em diferentes espécies, como é o caso de *N. vitripennis*. O clado Hymenoptera – Detox 2 é suportado por *bootstrap* de 100% e possui 13 genes de doze espécies da ordem Hymenoptera. Nenhum gene de *H. saltator* foi encontrado, enquanto *N. vitripennis* possui dois genes. Apenas os genes das espécies da família Apidae e um da espécie *N. vitripennis* apresentaram os três resíduos catalíticos conforme descrito por Cygler et al. (1993) (Figura 4.40). Um dado interessante é a conservação da tríade DEH (substituição da serina por um ácido aspártico) no clado que contém os genes de formiga. Esta tríade pode não ser catalítica, mas de alguma forma esse gene ficou conservado no genomas das espécies do grupo. O único clado não suportado por valor de *bootstrap* maior que 60% (ponto de corte escolhido) é o Lepidoptera – Detox 1. Este clado contém treze genes da espécie *B. mori* (única espécie do

grupo amostrada) que são majoritariamente catalíticos e não possuem um padrão de localização celular específico. Três genes (BMOR001964, BMOR000773 e BMOR011893) possuem substituições nos resíduos catalíticos e dois não possuem a tríade catalítica completa (BMOR004207 e BMOR006081), o que pode indicar que tenham perdido a função catalítica (Figura 4.40). O clado Lepidoptera – Detox 2 é suportado por *bootstrap* de 62% e contém oito genes majoritariamente catalíticos e intracelulares. O gene BMOR00077 possui substituições nos resíduos catalíticos e dois (BMOR010328 e BMOR010969) não possuem a tríade catalítica completa (Figura 4.40), o que pode indicar que tenham perdido a função catalítica. O clado Phthiraptera – Detox 1 é suportado por *bootstrap* de 100% e contém dois genes da espécie *P. humanus* (único representante deste grupo amostrado) que são intracelulares. O gene PHUM136220 é catalítico e PHUM569560 não possui a histidina da tríade catalítica (Figura 4.40), indicando perda de função. O clado Hemiptera – Detox 1 é suportado por *bootstrap* de 100% e contém cinco genes catalíticos da espécie *Acy. pisum*, onde três são intracelulares e dois mitocondriais (Figura 4.40).







## 5. Discussão

### 5.1 Conteúdo gênico nos genomas

O total de carboxilesterases (CCEs) extraídas nos 36 genomas foi de 1299 sequências, representando uma média de 36 sequências por espécie porém, este número variou bastante. Para os himenópteros é possível observar que o número total de CCEs é menor do que a média para a maioria das espécies, com exceção apenas da espécie *Nasonia vitripennis* que possui 46 sequências (Figura 4.23). Este dado está de acordo com a literatura, uma vez que Oakeshott e colaboradores em 2010 descrevem que existem expansões nos grupos detoxificante e de processadores de hormônio e feromônio em *N. vitripennis*. As espécies das famílias Megachilidae e Apidae seguem o repertório de enzimas descrito na literatura para *Ap. mellifera*, na qual foi observado um déficit no número de genes de CCEs (Claudianos et al., 2006). É possível que estas expansões não tenham ocorrido nestas duas famílias ou, se aconteceram, esses genes tenham sido perdidos ao longo da evolução. Entre as sete espécies da família Formicidae, foi observado um déficit maior para as espécies *Acromyrmex echinator* e *Harpegnathos saltator*, que possuem ao todo 21 e 17 sequências, respectivamente. Nygaard e colaboradores em 2011 observaram um déficit de CCEs no genoma de *Acr. echinator* que foi associado à sua dieta alimentar com menos toxinas devido ao hábito de cultivar fungos que degradam o material vegetal antes da sua ingestão. Esse tipo de associação também foi feito para *Ap. mellifera*, que possui uma dieta alimentar livre de toxinas (Claudianos et al., 2006). Espécies de Hymenoptera são, em muitos casos, sociais e a interação entre os membros se dá por processamento de moléculas químicas, que são secretadas (Tegoni et al., 2004). Já se sabe que as CCEs exercem um papel importante na degradação de hormônios (JHE) (Sparks et al., 1983) e também de outras moléculas importantes para a reprodução (Gilbert & Richmond, 1982; Chertemps et al. 2012). Apesar disso, estas enzimas não parecem ser importantes na troca de informação entre membros sociais, já que a maioria das espécies sociais analisadas neste trabalho não possui um grande repertório de CCEs. Realmente, análises comparativas de genomas de insetos sociais revelaram diversos outros genes importantes para a vida em uma sociedade complexa, sem indicar que CCEs estejam neste repertório (Fischman et al., 2011).

Outra espécie que apresentou déficit de CCEs em seu genoma foi *Pediculus humanus*, que é hematófaga e parasita de humanos. Essa espécie foi uma das que apresentou o menor número de CCEs (17) dentre todas as 36 espécies. Não só para essa família gênica, mas para

outras enzimas envolvidas no processo de detoxificação, essa tendência também foi observada (CYP, GST e ABC; Lee et al. 2010). Em outros insetos hematófagos analisados neste estudo, um número bem maior de CCEs foi encontrado (ver abaixo), indicando possivelmente que o nicho ecológico e o tamanho compacto de seu genoma tenham uma grande influência no número de genes de CCEs nesta espécie de piolho.

Diferente das espécies das ordens Hymenoptera e Phthiraptera, quatro espécies de interesse comercial apresentaram número de enzimas superior à média. *Metaseiulus occidentalis* e *Acyrtosiphon pisum*, que são conhecidos por prejudicar a produção agrícola, possuem 38 e 31 sequências respectivamente. *Tribolium castaneum*, que prejudica a produções de grãos apresentou 48 sequências e *Bombyx mori*, que produz a seda utilizada para confecção de roupas possui 65 CCEs. Além disso, os três mosquitos que são transmissores de doenças também possuem um número de CCEs acima da média. A espécie *Cu. quinquefasciatus* foi a que apresentou o maior número (69) entre as três espécies estudadas. A resistência a inseticidas nesta espécie é muito bem conhecida e já se observou que diversos genes de CCEs estão implicados na resistência a organofosforados (OP) em populações de mosquitos do gênero *Culex* (Raymond et al., 1989; Weill et al., 2004; Cui et al., 2006; Yan et al., 2012). Além do piolho (*P. humanus*) a única outra espécie de importância médica em que um número menor de CCEs foi encontrado foi no barbeiro (*Rhodnius prolixus*). Nessa espécie da ordem Hemiptera foram inicialmente encontradas 72 sequências, porém, após as análises preliminares foi observado que cerca de 83% dos genes estavam fragmentados. Isso provavelmente ocorreu devido à versão da predição gênica usada. Devido à recente liberação do genoma, a versão de suas predições ainda era muito preliminar e, portanto, boa parte dos genes ainda estava parcialmente predita. Realmente erros de predição gênica são comuns em anotações iniciais que ainda não foram curadas manualmente (Wang et al., 2003; Nagy et al., 2008). Em eucariotos, isso ocorre principalmente por conta da complexidade da estrutura gênica (presença de exons e introns), o que limita a automatização do processo de predição (Wang et al., 2008). No entanto, a análise comparativa de genomas e o conhecimento de características específicas do conjunto de dados ajuda a identificar sequências de baixa qualidade. Por isso, é possível que para *R. prolixus* um número maior de genes da família CCE ainda exista e que este trabalho não tenha avaliado seu repertório completo. O maior número de CCEs foi encontrado no genoma do carrapto *Ixodes scapularis* (71), um vetor de doenças humanas e de importância veterinária. Análises de transcriptoma de glândula salivar desta espécie mostram que duplicações gênicas, e possivelmente genômicas, tem um papel importante na sua evolução

(Mans & Neitz, 2004; Ribeiro et al., 2006). O número expressivo de genes dessa família gênica encontrado em artrópodes de importância médica e comercial pode ter relação com o controle químico. Muitos destes genes já foram implicados no surgimento de resistência a inseticidas em populações naturais desses artrópodes (Raymond et al., 1989; Devonshire et al., 1998; Field et al., 1999; Cui et al., 2006) e este grande repertório de genes facilita que tal tipo de evento ocorra devido à alta diversidade molecular. No entanto, quando programas de monitoramento dessas populações naturais são capazes de identificar o aumento da tolerância a inseticidas ainda no início (Macoris et al., 2005; WHO, 2013) é possível que alterações no método de controle sejam feitas para diminuir a pressão seletiva causada pelo controle químico (Georghiou, 1994; Zhao et al., 2010). Por isto, um melhor conhecimento do número de genes por espécie e seu relacionamento filogenético, que pode indicar sua função, é relevante.

## 5.2 Evolução da família Carboxilesterase em insetos

As análises filogenéticas mostram que esta família gênica é bastante diversificada. A falta de suporte estatístico nos ramos mais internos ocorre provavelmente devido à rápida diversificação dos insetos no cretáceo, muito associada à diversificação das plantas (Whitfield & Kjer, 2008). A formação de ramos internos curtos e ramos mais externos compridos dificultam a análise filogenética (Whitfield & Lockhart, 2007) e ainda hoje é um dos grandes problemas para a taxonomia de insetos (Trautwein et al., 2012). No entanto, a falta de suporte apenas dificulta a identificação do relacionamento filogenético entre os diferentes clados dentro da família gênica. Valores de *bootstrap* altos podem ser encontrados dando suporte aos diferentes grupos de genes identificados. Os três grandes grupos determinados por Oakeshott e colaboradores (2010), apesar de não suportados, identificam grupos de genes que parecem ter evoluído de forma diferente em insetos. Por exemplo, quando avaliamos o número de genes encontrados em cada espécie vemos que, para as glutactinas, gliotactinas, acetilcolinesterases e neurotactinas, todas originalmente colocadas dentro do grupo das CCEs neuronais, apenas um ou poucos genes de cada clado podem ser encontrados dentro dos diferentes genomas (Tabela 5.5). Isto indica que eventos de duplicação gênica são menos comuns para estes genes ou que, quando ocorrem, podem ter um efeito deletério dentro do genoma e são eliminados. Já para os genes classificados como Detoxificantes, ou mesmo para os Processadores de hormônios e feromônios, muitas cópias gênicas podem ser encontradas dentro dos genomas de cada espécie de inseto. Realmente, já

se observou que genes relacionados com detoxificação são encontrados em um número maior de cópias nos genomas de diversos organismos, um aumento de dosagem destes genes não é letal e em muitos casos pode ser benéfico (Field et al., 1988; 1999; Raymond et al., 1989; Devonshire et al., 1998; Vontas et al., 2000; Cui et al., 2006; Silva et al., 2012; Yan et al., 2012). Um bom exemplo é visto em populações de mosquito do gênero *Culex* resistentes a OP, nas quais foi observado um aumento na expressão de genes de esterases capazes de lidar com a exposição ao inseticida (Cui et al., 2006; Yan et al., 2012). Caso semelhante também já foi observado para a família das amilases em humanos onde um número diferente de cópias pode ser encontrado nas diferentes populações, de acordo com o consumo histórico de amido (Perry et al., 2007). O desbalanço da dosagem gênica em alguns casos pode causar efeitos deletérios ou mesmo letais. É possível que, para os diferentes genes classificados como neuronais o problema de dosagem possa ser uma limitação, pois na maioria dos clados observa-se apenas um ou poucos genes ortólogos de cada genoma. O contrário é observado nas classes Processadoras de hormônio e feromônios e Detoxificantes, que possuem indícios de terem evoluído por duplicações gênicas que ocorreram tanto dentro das diferentes espécies, quanto nos seus ancestrais compartilhados.

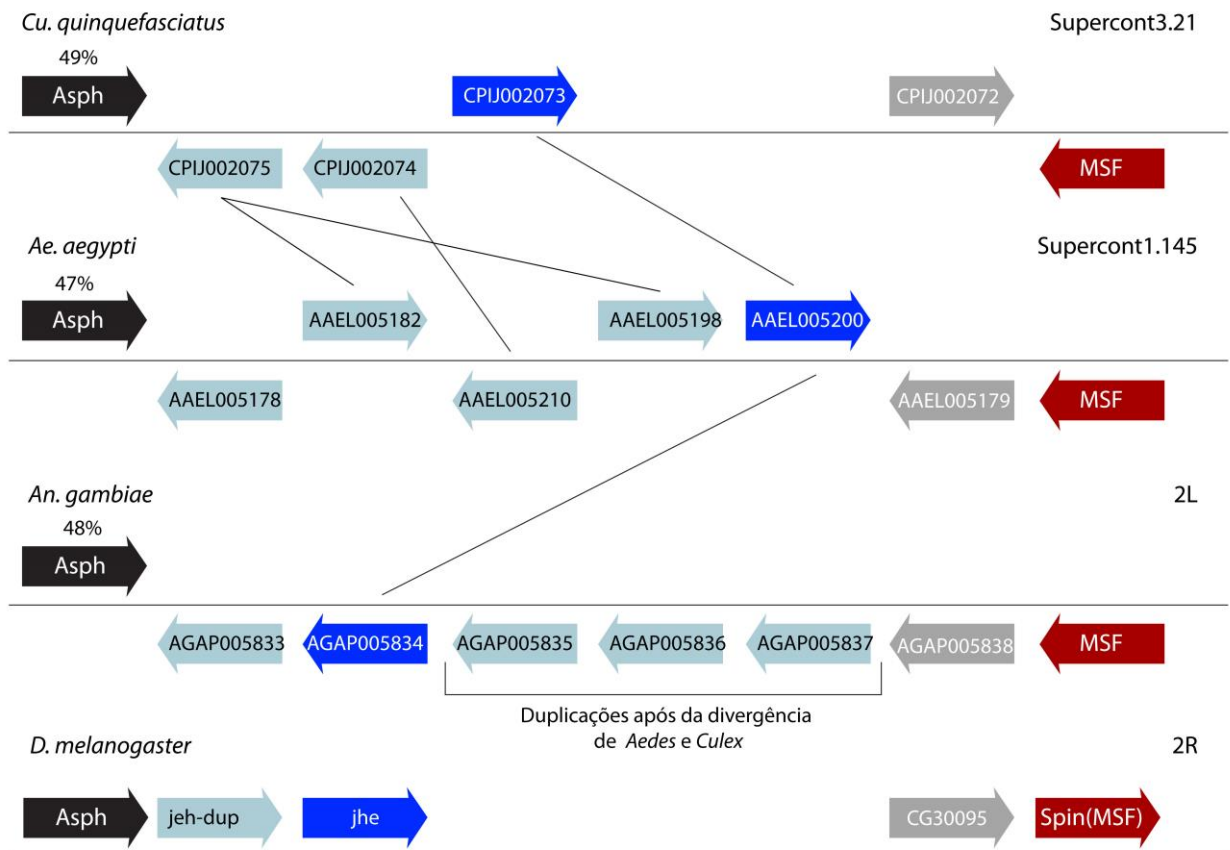
**Tabela 5.5.** Número de CCEs por espécie encontradas em cada clado encontrado.

Clados encontrado neste trabalho		Hymenoptera													Diptera													Hemiptera		Pht	Col	Lep	Ixo	Mês	Rha			
		AECH	PBAR	LLUM	ACEP	SINV	CFLO	HSAL	NVIT	MROT	BIMP	BTER	AFLO	AMEL	DGRI	DSIM	DERE	DPER	DSEC	DANA	DYAK	DMOJ	DPSE	DVIR	DWIL	DMEL	AGAM	AAEG	CQUI	APIS	RPRO	PHUM	TCAS	BMOR	ISCA	MOC	CELE	
Desenvolvimento Neuronal	L - Neurologina	4	2	5	3	1	3	2	5	5	5	2	5	3	4	4	3	4	4	4	4	4	4	4	8	5	3	5	2	1	5	4	4	5	5	5		
	K - Gliotactina	1		1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	5	1	1	1	1			1	1	1	1	1		
	M - Neurotactina	1	1	1	1	1			1	1	2	1	1	1			1	1	1	1	1	1	1	1	3	1	1	1			1	1	1	1				
	J - AChE 1	1	1			1	1	1	1	1	1	1	1													1	1	1	1	1		1	1	1				
	J - AChE 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1		1						
	J - AChE <i>C. elegans</i>																																					
	J - AChE - <i>M. occidentalis</i>																																			14		
	J - AChE - Arachnida																																		66	15		
	H - Glutactinas	1	1	1	1	1	1	1	1	1			1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1		1	1	1					
	H - Glutactinas Diptera														3	2	4	2	3	3	3	3	3	3	3	4	10	5										
	H - <i>Drosophila</i> AAA														1	1	1	1	1	1	1	1	1	1	4													
	I - Não caracterizado	1	1	1		1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1				
	Não caracterizado 1 - Novo														1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	
	Esterases - <i>C. elegans</i>																																					42
	Não caracterizado - Arachnida																																			1	1	
Não caracterizado - <i>B. mori</i>																																		7		1		
Processadoras de hormônios e feromônios	D - Tegumentares	1	1	1	1	1	2		3	2	3	3	2	1	3	3	3	3	3	3	3	3	3	3	4		1				2	2						
	F - JHE - Diptera														2	2	2	3	2	2	2	3	2	11	3	5	7	14				1						
	F - Não caracterizado - Hymenoptera	1	1	1	1		1	1	1		1	2	1	2																								
	E - Beta-esterase - <i>Drosophila</i>														1	2	2	3	2	3	2	6	3	5	1	2												
	E - Beta-esterase - multi Ordens	2	2	1	2	2	6	3	12	1	1	3	2	2												4	1	1	8				7	1				
	E - Beta-esterase conservada	1	1	3		1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1				1	1	1					
	E - JHE - Hemiptera																												9									
	E - Não caracterizado - <i>T. castaneum</i>																																	12				
	G - JHE Lepidoptera																																			4		
	G - Lepidoptera 1																																			13		
Detoxificantes	G - Diptera 1																										3	6	3									
	G - Diptera 2																										3	6	7									
	G - Hymenoptera 1								4	1	1	1	2	4																								
	Hymenoptera Detox - 1	4	7	5	5	4	9	3	9	2	2	4	5	5																								
	Hymenoptera Detox - 2	1	1	1	1	1	1		2	1	1	1	1	1																								
	Diptera Detox - 1														11	12	12	11	11	13	12	12	13	12	15	15	2	2	2									
	Diptera Detox - 2														1	1	1	2	1	1	1	2	1	2	1	10	14	23										
	Diptera Detox - 3														1	1	1		1	1	1	1	3	1	1	3	1	1	1									
	Lepidoptera Detox - 1																																			13		
	Lepidoptera Detox - 2																																			8		
Coleoptera Detox - 1																																		14				
Phthiraptera Detox - 1																																		2				
Hemiptera Detox - 1																																		5				
Total	20	20	22	17	16	28	15	43	19	21	25	22	28	33	32	37	34	36	38	36	42	41	40	47	57	45	59	68	30	5	12	48	58	70	37	52		

As duas classes catalíticas, Detoxificantes e Processadoras de hormônios e feromônios, contêm genes que possivelmente tiveram um ancestral comum, pois conservam um sítio catalítico ativo e não possuem outros domínios além do COesterase. Além disso, a presença dos genes de acetilcolinesterase, que possuem a tríade catalítica completa e ativa, e de pequenos grupos de genes que também possuem a tríade completa se agrupando com os genes neuronais (Clados *Glutactina – Diptera* e *Não Caracterizado*; Figura S1) indica que a função catalítica tenha surgido primeiro. É mais parcimonioso acreditar que a tríade foi perdida em alguns genes do que imaginar que ela tenha sido reconstruída mais de uma vez na história evolutiva desta família gênica.

Para as CCEs, a recombinação desigual parece ser o principal mecanismo de duplicação gênica. Este tem como principal característica a formação de agrupamentos de genes dentro dos genomas, e em espécies mais próximas filogeneticamente esses agrupamentos podem ser idênticos (Zhang, 2003; Hurles, 2004). Isso pode ser observado principalmente para os genes das classes detoxificantes (Robin et al., 1996; 2009) e para as esterases de hormônio juvenil de *Drosophila* (Crone et al., 2007) que ficam organizadas *in tandem* no genoma dessas espécies.

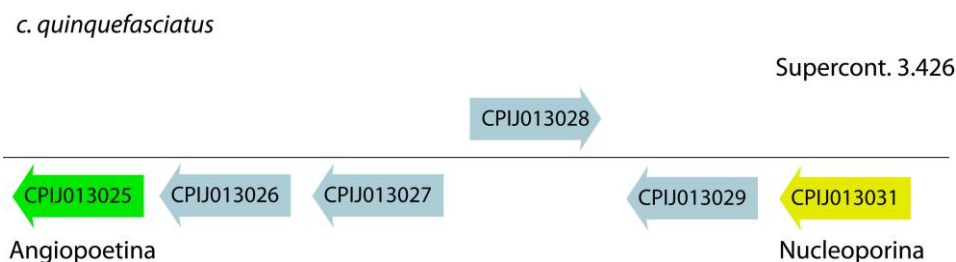
O grupo das Processadoras de hormônios e feromônios contém genes envolvidos com a degradação de substratos que sinalizam processos fisiológicos importantes como, por exemplo, o processo de metamorfose (Sparks et al., 1979; 1983). Em espécies de *Drosophila*, a evolução de JHE foi caracterizada por Crone e colaboradores em 2007, mostrando que nos genomas da maioria das espécies estudadas existiam genes funcionais e cópias com substituições que desabilitavam sua capacidade catalítica. Neste trabalho, o clado JHE *Diptera* (Clado F) é monofilético e, além dos genes de *Drosophila*, também contém genes de mosquito, indicando ancestralidade em comum. Uma análise de sintenia realmente mostra que um dos genes que flanqueia o agrupamento de genes de JHEs em mosquitos é homólogo ao gene *Aspartyl beta-hydroxylase* (Asph) que flanqueia este mesmo agrupamento de genes de JHEs em *D. melanogaster* (Crone et al., 2007). Do outro lado do agrupamento de genes de JHEs, o gene seguinte de *Drosophila* (CG30095) não é homólogo ao de mosquitos, entretanto, o próximo, um gene da super-família *Major Facilitator* (MSF), é homólogo, suportando a sintenia entre as quatro espécies (Figura 5.41).



**Figura 5.41** Organização do *cluster* ancestral de genes homólogos de JHE em quatro espécies de dípteros. Em preto e vermelho estão representados os genes homólogos ao de *D. melanogaster*, CG8421 – Aspartyl beta-hydroxylase e CG8428 – Major Facilitator Superfamily. Em cinza um gene que compartilha a mesma posição nas quatro espécies mas que não é relacionado evolutivamente. Em azul escuro estão representados os genes que possuem indícios de participarem da degradação de hormônio juvenil e, em azul claro, as cópias homólogas. As linhas mostram quais genes estão mais próximos filogeneticamente. Os valores percentuais mostram a identidade em relação ao gene respectivo de *D. melanogaster*.

Na espécie *Cu. quinquefasciatus* é possível identificar mais expansões nesse clado (Figura 4.36). Outros *clusters* de agrupamentos de homólogos de JHE também podem ser observados em outras regiões do genoma desta espécie (por exemplo, no *Supercontig* 3.426; Figura 5.42) indicando que realmente a recombinação desigual é uma força importante.





**Figura 5.42** Organização de um segundo *cluster* de genes homólogos de JHE duplicados em *Cu. quinquefasciatus*. Em verde e amarelo estão os genes flanqueadores e em azul claro estão as cópias de JHE.

Nesse trabalho, as alfa esterases de *Drosophila* e mosquitos ficaram no clado monofilético Diptera – Detox 1 (Figura 4.39). Alguns genes desse clado já foram identificados e relacionados com resistência a inseticidas em algumas populações naturais. As enzimas EstA (CPIJ013918) e EstB (CPIJ013917) são as que mais possuem indícios de estarem envolvidas no surgimento da resistência a OP em mosquitos do gênero *Culex* (Mouchès et al., 1987; Raymond et al., 1989; Cui et al., 2006; Yan et al., 2012). Em moscas existem indícios em *Haematobia irritans* (Guerrero et al., 2000) e *Musca domestica* (Claudianos et al., 1999; Scott & Zhang, 2003) da enzima Est7 estar envolvida com a resistência, pois sua super expressão foi detectada em indivíduos resistentes. Além desses, outros clados recuperados neste trabalho também possuem genes reconhecidamente relacionados com resistência a inseticidas em insetos. O clado Beta-esterase Multi ordem contém genes de *Acy. pisum* que, de acordo com as análises feitas por Ramsey e colaboradores (2010), se posicionam no mesmo clado das esterases E4 e EF4 do afídeo *My. persicae*, que já foram implicadas no aumento da resistência a OP (Field et al., 1998; 1999; Field & Blackman, 2003; Lan et al., 2005). Os resultados mostram que clados de diferentes classes podem estar implicados na resistência a inseticidas. O grande número de duplicações gênicas encontradas nestes dois grande grupos de genes Detoxificantes e Processadores de hormônios e feromônios podem ter implicações importantes para o surgimento da resistência a inseticidas em insetos vetores e pragas, especialmente quando consideramos que nestes insetos foi encontrado um número maior de genes nestes dois grupos de CCE.

### 5.3 Classificação das Carboxilesterases de inseto

A classificação dos genes da família das CCEs em insetos sempre foi importante. Desde muito cedo nos estudos de resistência a inseticidas se observou que esta é uma família importante, participando tanto da resistência metabólica quanto insensibilidade do sítio alvo (Bass & Field, 2011). A resistência metabólica é um assunto de grande complexidade e, no caso da família das CCEs, isto se dá por compreender genes de diferentes classes e, portanto, um grande número de espécies moleculares que atuam sobre diversos substratos. Tanto genes do grupo Detoxificante quanto do grupo das Processadoras de hormônios e feromônios, ambos envolvidos na metabolização de xenobióticos (Xu et al., 2005), já foram associados com a resistência a inseticidas em diversas populações naturais (Mouchès et al., 1987; Raymond et al., 1989; Cui et al., 2006; Carvalho et al., 2009; Yan et al., 2012). Talvez por esta razão a classificação das carboxilesterases em insetos seja controversa, e diversos métodos classificatórios já foram propostos, embora sem um consenso (Montella et al., 2012). A classificação que melhor conseguiu organizar a grande diversidade desta família gênica é a filogenética, primeiro estabelecida por Oakeshott e colaboradores em 2005. No entanto, Tsubota e Shiotsuki (2010b), analisando 425 CCEs de inseto, observaram que uma classificação funcional talvez não seja possível para alguns grupos de CCEs.

Realmente, os resultados obtidos neste trabalho mostram que, mesmo com o aumento da amostragem de sequências de CCEs de insetos, uma classificação baseada em dados filogenéticos não descreve a relação funcional entre os genes. Por exemplo, para a organização dos genes da classe Detoxificante, o grupo taxonômico e não a função do gene parece ser o aspecto mais importante (Figura 4.38). Neste grupo vemos expansões (duplicações) que ocorreram nas diversas linhagens separadamente. Esse tipo de organização mostra que tais genes evoluíram de forma independente nos diferentes grupos taxonômicos, muito provavelmente devido à conquista de novos nichos ecológicos (Després et al., 2007). Com isto, apenas em alguns poucos casos é possível determinar substratos compartilhados por estes genes nas diferentes espécies. Alguns membros do grupo das Processadoras de hormônios e feromônios como, por exemplo, os JHE, talvez possam ser classificados funcionalmente pois existem indícios de que atuam sobre o mesmo substrato em espécies diferentes (Bai et al., 2007; Ishikawa et al., 2012). No entanto, nesse trabalho e em outros já publicados vemos que mesmo esses genes evoluíram independentemente em cada linhagem de inseto (Oakeshott et al., 2005; Claudianos et al., 2006). Isto indica que a função de

degradação de HJ em insetos evoluiu mais de uma vez no grupo (Figura 4.26). O grupo neuronal parece ser o que mais se adequa a uma classificação funcional, pois os diferentes clados contêm apenas um ou poucos genes ortólogos para a maioria das espécies, além de desempenharem funções conservadas. No entanto, apenas a AChE, que é alvo de OP e CX, parece ser importante no surgimento da resistência a inseticidas (Weill et al., 2003; Alout et al., 2007; Kwon et al., 2010). Os outros genes deste grupo possuem funções não catalíticas importantes, mas não relacionadas à resistência a inseticidas.

O relacionamento filogenético entre os diferentes genes da família das CCEs e a avaliação de quais clados estariam mais relacionados com a resistência a inseticidas em artrópodes mostra que uma previsão de genes mais importantes para a resistência e uma classificação funcional que ajudasse neste aspecto não é possível. A maioria dos genes catalíticos desta família gênica são promíscuos quanto aos substratos que podem hidrolisar (Russel et al., 2010) e, como é possível observar na árvore filogenética, a especificidade em relação a um determinado substrato evoluiu diversas vezes na história evolutiva do grupo.

#### 5.4 Evolução dos domínios

Todos os genes contidos nesse trabalho possuem o domínio COesterase (PF00135), embora sua função seja bastante diferente entre alguns membros dessa família. Por exemplo, as neuroliginas possuem o domínio COesterase não catalítico e têm função de adesão celular durante a formação de sinapses neuronais (Banovic et al., 2010). Já os membros catalíticos hidrolizam diferentes substratos esterés, o que inclui moléculas bem diversas. Isso mostra a versatilidade desse domínio, já que proteínas que realizam reações enzimáticas contêm o mesmo domínio de proteínas sem função de catálise. A re-utilização de domínios para diferentes funções não é novidade (Vogel et al., 2004; Ekman et al., 2005, Bashton & Chothia, 2007). Por exemplo, na proteína naftaleno 1,2-dioxigenase o domínio *Bet v1-like* (PITP) é catalítico e atua junto com um domínio *Rieske* presente na região amino-terminal da proteína para realizar a catálise do naftaleno (Carredano et al., 2000). Entretanto, na proteína fosfatidilinositol este domínio não é catalítico e atua transferindo fosfolipídeos entre as superfícies das membranas (Yoder et al., 2001). Na realidade, hoje suspeita-se que mesmo para o surgimento de genes órfãos (sem homólogo próximo detectado) processos de duplicação gênica seguidos de rápida divergência devam ter ocorrido (Tautz & Damazet-Loso, 2011). É claro que novos genes podem surgir aleatoriamente, mas este é um processo muito mais demorado e bem menos eficiente (Carvunis et al., 2012).

Algumas CCEs não catalíticas (neuroliginas, gliotactinas e neurotactinas) possuem regiões hidrofóbicas conhecidas que são responsáveis por mantê-las ancoradas na membrana citoplasmática de células neuronais (de la Escalera et al., 1990; Auld et al., 1995; Ichtchenko et al., 1995). A presença e a conservação da posição da(s) hélice(s) transmembranar(es) na região carboxi-terminal nas neuroliginas indica que essa região se fusionou com um domínio COesterase para formar uma proteína funcional. Isso pode ter ocorrido por eventos de recombinação desigual ou por embaralhamento de exons (Bashton & Chothia, 2007) permitindo o surgimento da nova função. Estudos mostram que esta proteína surgiu no ancestral comum de cnidários e bilaterais (Ryan & Grant, 2009) e possui função importante na formação das sinapses nervosas (Banovic et al., 2010). As neurotactinas e gliotactinas, que também estão envolvidas com o desenvolvimento do SNC, também parecem ter sido originadas pelo fusão de um domínio COesterase com um domínio transmembranar. No entanto, estes parecem ter sido pelo menos dois eventos separados, ou novos rearranjos ocorreram, já que o posicionamento desses domínios nestas proteínas é bem diferente. Nas gliotactinas e neuroliginas o domínio transmembranar encontra-se na região carboxi-terminal enquanto que nas neurotactinas ele está presente antes do domínio COesterase na porção amino-terminal da proteína (Figura 1.5).

Para a enzima acetilcolinesterase, que é catalítica, uma fusão de domínios também parece ter sido importante para a função da enzima. No entanto, o domínio adquirido possibilita a interação da parte carboxi-terminal da proteína com uma molécula de glicofosfatidilinositol (GPI), que é responsável pelo ancoramento na membrana citoplasmática (Bütikofer et al., 1990). Isso indica que os eventos de fusão de domínios com ganho de nova função ocorreram mais de uma vez ao longo da história evolutiva desta família gênica, demonstrando a versatilidade do domínio COesterase. A fusão de domínios é bastante conhecida em diversas espécies (Vogel et al., 2004; Ekman et al., 2005; Han et al., 2007). Um bom exemplo de fusão é visto para os genes *dihidrofolato redutase (DHFR)* e  *timidilato sintetase (TS)* que são transcritos separadamente em animais e fungos (Philippe et al., 2000). No entanto, em plantas, alveolados e euglenoides esses dois genes estão fusionados, codificando uma proteína bifuncional (Ivanetich & Santi, 1990). Além da fusão de domínios diferentes, em alguns casos também vemos fusão de dois domínios COesterase em uma única CCE (Figura 4.19), o que pode ser fruto de recombinação desigual. A fusão entre domínios iguais já foi caracterizada, por exemplo, na  *proteína FokI* que está envolvida no reconhecimento de DNA e possui três domínios *winged helix* na porção amino-terminal (Wah et al., 1997). Para as CCEs dos deuterostomados também há evidências da ocorrência tanto de fusões entre

domínios COesterases quanto entre um domínio COesterase e outros (*RING finger*, *chitin-binding*, *immunoglobulin*, *fibronectin type 3*, entre outros). Nestes casos, os autores especulam que os eventos de fusão podem ter contribuído na evolução da complexidade morfológica, principalmente do sistema nervoso e da crista neural dos vertebrados (Johnson & Moore, 2012).

A identificação dos resíduos da tríade mostrou que 61% das sequências são potencialmente catalíticas, pois possuem a tríade predita como Ser-Glu-His ou Ser-Asp-His. No entanto, para as CCEs, além da serina, um resíduo de ácido aspártico ou cisteína pode ocupar o lugar do resíduo nucleofílico em membros catalíticos (Nardini & Dijkstra, 1999). Nesse trabalho, a ocorrência de um desses dois aminoácidos nessa posição foi rara, sendo observada apenas em todas as formigas e *Megachile rotundata* (Hymenoptera – Detox 2; Figura 4.40), *T. castaneum* (Coleoptera –Detox 1; Figura 4.39) e *Ae. aegypti* (JHE Diptera; Figura 4.36). O desenvolvimento de uma nova função causada por substituições no sítio catalítico já foi descrito para a enzima epimerase de *Escherichia coli*. Originalmente essa enzima é monofuncional e realiza uma reação de epimerização, mas devido à substituição de um ácido aspártico para glicina no sítio catalítico passa a desempenhar também a função de síntese de o-succinilbenzoato (Vick & Gerlt, 2007).

Em 11% das sequências foi observada uma glicina no lugar da serina nucleofílica, indicando perda de função catalítica (Gibney et al., 1990). No entanto, isto não impede que essas proteínas sejam funcionais, já que as gliotactinas possuem esse tipo de substituição e participam da formação da barreira entre o sistema nervoso e circulatório (Auld et al., 1995). O restante das sequências (28%) ficam dentro do conjunto nomeado de tríade incompleta, pois ou possuem tríade diferente dos dois conjuntos anteriores (Ser-Glu-His e Ser-Asp-His; Gli-Glu-His) ou não possuem todos os três resíduos catalíticos. Nesse conjunto estão genes das três classes, Desenvolvimento neuronal, Detoxificantes e Processadoras de hormônios e feromônios. Nas duas classes que possuem genes catalíticos, Detoxificantes e Processadoras de hormônios e feromônios, em todos os clados foram vistos genes com tríade incompleta, exceto nos clados JHE Hemiptera e Hemiptera – Detox 1. Estes são genes truncados que não foram excluídos pelo *pipeline* e podem estar seguindo o caminho de pseudogenização. É claro que indícios de pseudogenização e outras inferências sobre conteúdo da tríade devem ser vistos com cautela, já que podem também ser influenciados por problemas na predição gênica usada (Wang et al., 2003; Nagy et al., 2008).

Ao longo da evolução, os resíduos catalíticos foram conservados no domínio COesterase de membros que realizam hidrólise (Krejci et al., 1991; Cygler et al., 1993;

Nardini & Dijkstra, 1999), o que é observado para grande parte das sequências de inseto usadas nesse trabalho. No entanto, em membros não catalíticos da classe Desenvolvimento neuronal, são encontrados apenas vestígios dos resíduos do sítio catalítico, indicando que estes resíduos provavelmente não são importantes para a nova função e são apenas resquícios do domínio COesterase ancestral.

Além dos diferentes domínios e resíduos catalíticos encontrados neste trabalho, uma alta diversidade nos padrões de endereçamento celular também foi encontrada. É possível que estes novos endereçamentos surjam como uma forma de resolver a redundância funcional que ocorre após os eventos de duplicação, tão comuns nesta família gênica. No entanto, não é possível determinar, só pelo padrão de endereçamento, se essas proteínas compartilham funções similares em locais diferentes ou se já adquiriram novas funções.

## 6. Conclusões

Um deficit de genes foi identificado para os himenópteros, com exceção de *N. vitripennis*, o que provavelmente pode estar ligado ao estilo de vida e à dieta alimentar destes animais.

Um número maior de CCEs foi encontrado em insetos vetores e pragas agrícolas, indicando que esta família gênica pode ser realmente um fator importante no desenvolvimento de resistência a inseticidas em populações naturais.

As análises filogenéticas mostram que as duas classes catalíticas, Detoxificantes e Processadoras de hormônios e feromônios, apresentam um número maior de duplicações gênicas por espécie do que a classe Desenvolvimento neuronal. O desbalanço da dosagem gênica para genes envolvidos com o metabolismo e degradação de hormônios não parece ser deletério. Por outro lado, para genes responsáveis pelo desenvolvimento do sistema nervoso, pertencentes à classe Neuronal, o desbalanço gênico pode ser um fator limitante.

A mistura de padrões de endereçamento para os genes detoxificantes sugere que isso possa ser uma saída para a redundância funcional que geralmente se segue após eventos de duplicação. No entanto, não é possível determinar, só pelo padrão de endereçamento, se estas proteínas compartilham funções similares em locais diferentes ou se já adquiriram novas funções.

A identificação de regiões hidrofóbicas nas proteínas neuroliginas, gliotactinas e neurotactinas indica que esses genes foram originados a partir de eventos de fusão entre um domínio COesterase e um domínio transmembranar.

O relacionamento filogenético entre os diferentes genes da família das CCEs e a avaliação de quais clados estariam mais relacionados com a resistência a inseticidas em artrópodes mostra que uma previsão de genes mais importantes para a resistência e uma classificação funcional que ajudasse neste aspecto não é possível. O exemplo mais drástico pode ser observado na classe Detoxificante, onde os três clados funcionais determinados por Oakeshott e colaboradores em 2010 na realidade se misturam nos diferentes clados táxon-específicos encontrados.

## 7. Referências Bibliográficas

- Ahmad M, Denholm I, Bromilow RH. 2006. Delayed cuticular penetration and enhanced metabolism of deltamethrin in pyrethroid-resistant strains of *Helicoverpa armigera* from China and Pakistan. **Pest Management Science**, 62:805–10.
- Alout H, Berthomieu A, Hadjivassilis A, Weill M. 2007. A new amino-acid substitution in acetylcholinesterase 1 confers insecticide resistance to *Culex pipiens* mosquitoes from Cyprus. **Insect Biochemistry and Molecular Biology**, 37(1):41-7.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. **Journal Molecular Biology**, 215(3):403-10.
- Auld VJ, Fetter RD, Broadie K, Goodman CS. 1995. Gliotactin, a Novel Transmembrane Protein on Peripheral Gila, Is Required to Form the Blood-Nerve Barrier in *Drosophila*. **Cell**, 81:757-67.
- Bai H, Ramaseshadri P, Palli SR. 2007. Identification and characterization of juvenile hormone esterase gene from the yellow fever mosquito, *Aedes aegypti*. **Insect Biochemistry and Molecular Biology**, 37:829-37.
- Bailey JA, Liu G, Eichler EE. 2003. An Alu Transposition Model for the Origin and Expansion of Human Segmental Duplications. **The American Society of Human Genetics**, 73(4):823-34.
- Banovic D, Khorramshahi O, Oswald D, Wichmann C, Riedt T, Fouquet W, et al. 2010. *Drosophila* Neuroligin 1 Promotes Growth and Postsynaptic Differentiation at Glutamatergic Neuromuscular Junctions. **Cell press**, 66:724–38.
- Bashton M, Chothia C. 2007. The generation of new protein functions by the combination of domains. **Structure**, 15(1):85–99.
- Bass C, Field ML. 2011. Gene amplification and insecticide resistance. **Pest Management Science**, 67(8):886-90.
- Belinato TA, Martins AJ, Lima JBP, Lima-Camara TN, Peixoto AA, Valle D. 2009. Effect of the chitin synthesis inhibitor triflumuron on the development, viability and reproduction of *Aedes aegypti*. **Memórias do Instituto Oswaldo Cruz**, 104(1):43-47.
- Biswas S, Russell RJ, Jackson CJ, Vidovic M, Ganeshina O, Oakeshott JG, et al. 2008. Bridging the synaptic gap: neuroligins and neurexin I in *Apis mellifera*. **PLoS One**, 3:e3542.
- Birchler JA, Veitia RA. 2012. Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. **PNAS**, 109(37):14746–53.
- Bonifas JM, Morley BJ, Oakey RE, Kant YW, Epstein EH. 1987. Cloning of a cDNA for steroid sulfatase: Frequent occurrence of gene deletions in patients with recessive X chromosome-linked ichthyosis. **PNAS**, 84:9248-51.



- Braga IA, Lima JBP, Soares SS, Valle D. 2004. *Aedes aegypti* Resistance to Temephos during 2001 in Several Municipalities in the States of Rio de Janeiro, Sergipe, and Alagoas, Brazil. **Memórias do Instituto Oswaldo Cruz**, Vol. 99(2):199-203.
- Braga IA, Mello CB, Peixoto AA, Valle D. 2005. Evaluation of methoprene effect on *Aedes aegypti* (Diptera: Culicidae) development in laboratory conditions. **Memórias do Instituto Oswaldo Cruz**, 100(4):435-40.
- Braga IA, Valle D. 2007a. *Aedes aegypti*: history of control in Brazil. **Epidemiologia e Serviços de Saúde**, 16(2):113-18.
- Braga IA, Valle D. 2007b. *Aedes aegypti*: inseticidas, mecanismos de ação e resistência. **Epidemiologia e Serviços de Saúde**, 16(4):279-93.
- Bütikofer P, Kuypers FA, Shackleton C, Brodbeck U, Stieger S. 1990. Molecular Species Analysis of the Glycosylphosphatidylinositol Anchor of *Torpedo marmorata* Acetylcholinesterase. **The Journal of Biological Chemistry**, 265(5):18983-87.
- Carredano E, Karlsson A, Kauppi B, Choudhury D, Parales RE, Parales JV, et al. 2000. Substrate binding site of naphthalene 1,2-dioxygenase: functional implications of indole binding. **The Journal of Molecular Biology**, 296(2):701-12.
- Carvalho RA, Torres TT, Paniago MG, Azeredo-Espin AM. 2009. Molecular characterization of esterase E3 gene associated with organophosphorus insecticide resistance in the New World screwworm fly, *Cochliomyia hominivorax*. **Medical and Veterinary Entomology**, 23:86–91.
- Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, et al. 2012. Proto-genes and de novo gene birth. **Nature**, 487(7407):370-4.
- Chevillon C, Raymond M, Guillemaud T, Lenormand T, Pasteur N. 1999. Population genetics of insecticide resistance in the mosquito *Culex pipiens*. **Biological Journal of the Linnean Socieb**, 68:(347)157.
- Claudianos C, Russell RJ, Oakeshott JG. 1999. The same amino acid substitution in orthologous esterases confers organophosphate resistance on the house fly and a blowfly. **insect biochemistry and molecular biology**, 29:675–86.
- Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, et al. 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. **Insect Molecular Biology**, 15(5):615–36.
- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. **Nature Reviews Genetics**, 9:938-50.
- Crone EJ, Sutherland TD, Campbell PM, Coppin CW, Russell RJ, Oakeshott JG. 2007. Only one esterase of *Drosophila melanogaster* is likely to degrade juvenile hormone in vivo. **Insect Biochemistry and Molecular Biology**, 37:540–49.
- Cui F, Lin LF, Qiao CL, Xu Y, Marquine M, Weill M, et al. 2006. Insecticide resistance in Chinese populations of the *Culex pipiens* complex through esterase overproduction. **Entomologia Experimentalis et Applicata**, 120(3):211-20.

- Cygler M, Schrag JD, Sussman JL, Harel M, Silman I, Gentry MK, et al. 1993. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. **Protein Science**, 2(3):366-82.
- Darboux I, Barthalay Y, Piovant M, Hipeau-Jacquotte R. 1996. The structure-function relationships in *Drosophila* neurotactin show that cholinesterasic domains may have adhesive properties. **The EMBO Journal**, 15(18):4835-43.
- Davidson AL, Dassa E, Orelle C, Chen J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. **Microbiology and Molecular Biology Reviews**, 72(2):317-64.
- Davies TEG, Field LM, Usherwood PNR, Williamson MS. 2007. DDT, Pyrethrins, Pyrethroids and insect sodium channels. **Informa healthcare**, 59(3):151-62.
- De la Escalera, Bockamp E, Moya F, Piovant M, Jimenez F. 1990. Characterization and gene cloning of neurotactin, a *Drosophila* transmembrane protein related to cholinesterases. **The EMBO Journal**, 9(11):3593-3601.
- Deng C, Cheng CHC, Yea H, Heb X, Chen L. 2010. Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. **PNAS**, 107(50):21593-98.
- Després L, David J, Gallet C. 2007. The evolutionary ecology of insect resistance to plant chemicals. **TRENDS in Ecology and Evolution**, 22(6):298-307.
- Devonshire AL, Field LM, Foster SP, Moores GD, Williamson MS, Blackman RL. 1998. The evolution of insecticide resistance in the peachpotato aphid, *Myzus persicae*. **Philosophical Transactions of the Royal Society B**, 353:1677-84.
- Durand D, Hoberman R. 2006. Diagnosing duplications – can it be done? **TRENDS in Genetics**, 22(3):156-64.
- Dvir H, Harel M, Bon S, Liu WQ, Vidal M, Garbay C, et al. 2004. The synaptic acetylcholinesterase tetramer assembles around a polyproline II helix. **The EMBO Journal**, 23:4394-4405.
- Edgecombe GD. 2010. Arthropod phylogeny: An overview from the perspectives of morphology, molecular data and the fossil record. **Arthropod Structure & Development**, 39:74-87.
- Ekman D, Björklund AK, Frey-Skött J, Elofsson A. 2005. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. **Journal of Molecular Biology**, 348(1):231-43.
- Emanuel BS, Shaikh TH. 2001. Segmental duplications: an ‘expanding’ role in genomic instability and disease. **Nature**, 2:791-800.
- Emanuelsson O, Brunak S, Heijne GV, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. **Nature Protocols**, 2:953-71.
- Feyereisen R. 1999. Insect P450 enzymes. **Annual Review of Entomology**, 44(1):507-33.

- Field LM, Devonshire AL, Forde BG. 1988. Molecular evidence that insecticide resistance in peach-potato aphids (*Myzus persicae* Sulz.) results from amplification of an esterase gene. **Biochemical Journal**, 251:309–12.
- Field LM, Blackman RL, Tyler-Smith C, Devonshire AL. 1999. Relationship between amount of esterase and gene copy number in insecticideresistant *Myzus persicae* (Sulzer). **Biochemical Journal**, 339:737–42.
- Field ML, Blackman RL. 2003. Insecticide resistance in the aphid *Myzus persicae* (Sulzer): chromosome location and epigenetic effects on esterase gene expression in clonal lineages. **Biological Journal of the Linnean Society**, 79:107–13.
- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: Interactive sequence similarity searching. **Nucleic Acids Research**, 39(Web Server issue):W29-W37.
- Fischman BJ, Woodard SH, Robinson GE. 2011. Molecular evolutionary analyses of insect societies. **PNAS**, 108 Suppl 2, 10847–54.
- Fontoura NG, Bellinato DF, Valle D, Lima JBP. 2012. The efficacy of a chitin synthesis inhibitor against field populations of organophosphate-resistant *Aedes aegypti* in Brazil. **Memórias Instituto Oswaldo Cruz**, 107:387–95.
- Funasa 2001. Dengue: Instruções para pessoal de combate ao vetor - Manual de normas técnicas. **Fundação Nacional de Saúde**, Ministério da Saúde, Brasília.
- Fundação Nacional de Saúde. 2002. Programa Nacional de Controle da Dengue (PNCD). **Fundação Nacional de Saúde**, Ministério da Saúde, Brasília.
- Friedman R. 2011. Genomic organization of the glutathione S-transferase family in insects. **Molecular Phylogenetics and Evolution**, 61(3):924-32.
- Fukuto, 1990, Mechanism of Action of Organophosphorus and Carbamate Insecticides. **Environ Health Perspect**, 87:245–54.
- Graf JF. 1993. The role of insect growth regulators in arthropod control. **Parasitology Today**, 9(12):471–74.
- Gahan LJ, Pauchet Y, Vogel H, Heckel DG. 2010. An ABC Transporter Mutation Is Correlated with Insect Resistance to *Bacillus thuringiensis* Cry1Ac Toxin. **PLoS Genetics**, 6(12): e1001248.
- Georghiou GP. 1994. Principles of insecticide resistance management. **Phytoprotection**, 75(4):51-59.
- Gibney G, Camp S, Dionne M, Macphee-Quigley K, Taylor P. 1990. Mutagenesis of essential functional residues in acetylcholinesterase. **PNAS**, 87:7546-50.
- Gilbert DG, Richmond RC. 1982. Esterase 6 in *Drosophila melanogaster*. Reproductive function of active and null males at low temperature. **PNAS**, 79:2962-66
- Gilbert MM, Alud VJ. 2005. Evolution of clams (cholinesterase-like adhesion molecules): structure and function during development. **Frontiers in Bioscience**, 10: 2177-92.

- Goodman N. 2002. Biological data becomes computer literate: new advances in bioinformatics. **Current Opinion in Biotechnology**, 13:68–71.
- Guerrero FD, Jamroz RC, Kammlah D, Kunz SE. 1997. Toxicological and molecular characterization of pyrethroid-resistant horn flies, *Haematobia irritans*: identification of kdr and super-kdr point mutations. **Insect Biochemistry and Molecular Biology**, 27(8–9):745–55.
- Guerrero FD. 2000. Cloning of a horn fly cDNA, HiaE7, encoding an esterase whose transcript concentration is elevated in diazinon-resistant flies. **Insect Biochemistry and Molecular Biology**, 30:1107–15.
- Habig WH, Pabst MJ, Jakoby WB. 1974. Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. **Journal of Biological Chemistry**. 249(22):7130-9.
- Hahn MW, Han MV, Han SG. 2007. Gene family evolution across 12 *Drosophila* genomes. **PLoS Genetics**, 311:e197.
- Hammock BD. 1985. em *Comprehensive Insect Physiology, Biochemistry, and Pharmacology*, eds Kerkut GA., Gilbert LI (Pergamon Press, New York), pp 431–72.
- Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. 2007. The folding and evolution of multidomain proteins. **Nature Reviews Molecular Cell Biology**, 8:319-30.
- Hassold T, Hall H, Hunt P. 2007. The origin of human aneuploidy: where we have been, where we are going. **Human Molecular Genetics**, 2007, 16(2):R203–R208.
- Heidari R, Devonshire AL, Campbell BE, Bell KL, Dorrian SJ, Oakeshott JG, et al. 2004a. Hydrolysis of organophosphorus insecticides by in vitro modified carboxylesterase E3 from *Lucilia cuprina*. **Insect Biochemistry and Molecular Biology**, 34:353–63.
- Heidari R, Devonshire AL, Campbell BE, Dorrian SJ, Oakeshott JG, Russell RJ. 2004b. Hydrolysis of pyrethroids by carboxylesterases from *Lucilia cuprina* and *Drosophila melanogaster* with active sites modified by in vitro mutagenesis. **Insect Biochemistry and Molecular Biology**, 35:597–609.
- Hemingway J, Ranson H. 2000. Insecticide resistance in insect vectors of human disease. **Annual Review Entomology**, 45:371–91.
- Hemingway J, Hawkes NJ, McCarroll L, Ranson H. 2004. The molecular basis of insecticide resistance in mosquitoes, **Insect Biochemistry and Molecular Biology**, 34(7):653-65.
- Hlavica P. 2011. Insect cytochromes P450: Topology of structural elements predicted to govern catalytic versatility. **Journal of Inorganic Biochemistry**, 105(10):1354-64.
- Høyer H, Braathen GJ, Eek AK, Skjelbred CF, Russell MB. 2011. Charcot-Marie-Tooth caused by a copy number variation in myelin protein zero. **European Journal of Medical Genetics**, 54(6):e580–e583.

- Huchard E, Martinez M, Alout H, Douzery EJ, Lutfalla G, Berthomieu A, et al. 2006. Acetylcholinesterase genes within the Diptera: takeover and loss in true flies. **Proceedings of the Royal Society B**. 22;273(1601):2595-2604.
- Huminięcki L, Wolfe KH. 2004. Divergence of Spatial Gene Expression Profiles Following Species-Specific Gene Duplications in Human and Mouse. **Genome Research**,14:1870-79.
- Hurles M. 2004. Gene Duplication: The Genomic Trade in Spare Parts. **PLoS Biology**, 2(7):900-4.
- Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Messier M, Moomaw C, et al. 1995. Neuroligin 1: A Splice Site-Specific Ligand for I –Neurexins. **Cell**, 81(3):435-43.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. **Nature**, 409:860-921.
- Ishikawa A, Ogawa K, Gotoh H, Walsh TK, Tagu D, Brisson JA, et al. 2012. Juvenile hormone titre and related gene expression. during the change of reproductive modes in the pea aphid. **Insect Molecular Biology**, 21(1):49-60.
- Ivanetich KM, Santi DV. 1990. Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa. **The FASEB Journal**. 4(6):1591-7.
- Johnson G, Moore SW. 2012. The carboxylesterase/cholinesterase gene family in invertebrate deuterostomes. **Comparative Biochemistry and Physiology**, 7(2):83-93.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. **Computer Applications in the Biosciences**, 8(3):275-82.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. (describes the FFT-NS-1, FFT-NS-2 and FFT-NS-i strategies). **Nucleic Acids Research**. 30:3059-66.
- Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. **Nature**, 428(6983):617-24.
- Khan AA, Janke A, Shimokawa T, Zhang H. 2011. Phylogenetic Analysis of Kindlins Suggests Subfunctionalization of an Ancestral Unduplicated Kindlin into Three Paralogs in Vertebrates. **Evolutionary Bioinformatics**, 7:7–19.
- Krejci E, duval N, chatonnet A, vincensf P, massouliv J. 1991. Cholinesterase-like domains in enzymes and structural proteins: Functional and evolutionary relationships and identification of a catalytically essential aspartic acid. **PNAS**, 88:6647-51.
- Krogh A, Larsson B, Heijne GV, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. **Journal of Molecular Biology**, 305(3):567-80.
- Kwon DH, Im JS, Ahn JJ, Lee J, Clark JM, Lee SH. 2010. Acetylcholinesterase point mutations putatively associated with monocrotophos resistance in the two-spotted spider mite, **Pesticide Biochemistry and Physiology**, 96(1):36-42.

- Lan W, Cong J, Jiang H, Jiang S, Qiao C. 2005. Expression and Characterization of Carboxylesterase E4 Gene from Peach–Potato Aphid (*Myzus persicae*) for Degradation of Carbaryl and Malathion. **Biotechnology Letters**, 27(15):1141-46.
- Lee SH, Kang JS, Min JS, Yoon KS, Strycharz JP, Johnson R, et al. 2010. Decreased detoxification genes and genome size make the human body louse an efficient model to study xenobiotic metabolism. **Insect Molecular Biology**, 19(5):599-615.
- Lenfant N, Hotelier T, Bourne Y, Marchot P, Chatonnet A. 2013. Proteins with an alpha/beta hydrolase fold: Relationships between subfamilies in an ever-growing superfamily. **Chemico-Biological Interactions**, 203:266–68.
- Lenormand T, Bourguet D, Guillemaud T, Raymond M. 1999. Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. **Nature**, 400(6747):861-4.
- Levine R. 2011. The 5.000 insect genome project. **American Entomologist**, 57(2):111-3.
- Licht A, Schneider E. 2011. ATP binding cassette systems: structures, mechanisms, and functions. **Central European Journal of Biology**, 6(5):785-801.
- Luscombe NM, Greenbaum D, Gerstein M. 2001. What is bioinformatics? An introduction and overview. **Yearbook of Medical Informatics**, Yale University New Haven, USA, 83-100.
- Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. **Trends in Genetics**, 20(11):544-9.
- Maeda MH, Mitsui J, Soong B, Takahashi Y, Ishiura H, Hayashi S, et al. 2012. Increased Gene Dosage of Myelin Protein Zero Causes Charcot-Marie-Tooth Disease. **ANNALS of Neurology**, 71:84–92.
- Mans BJ, Neitz AW. 2004. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. **Insect Biochemistry and Molecular Biology**, 34(1):1-17.
- Marchler-Bauer A, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, et al. 2009. CDD: specific functional annotation with the Conserved Domain Database. **Nucleic Acids research**, 37(Database issue):D205–D210.
- Macoris MLG, Andrighetti MTM, Nalon KCR, Garbeloto VC, Júnior ALC. 2005. Standardization of Bioassays for Monitoring Resistance to Insecticides in *Aedes aegypti*. **Dengue Bulletin**, 29:176-82.
- Martins AJ, Belinato TA, Lima JBP, Valle D. 2008. Chitin synthesis inhibitor effect on *Aedes aegypti* populations susceptible and resistant to the organophosphate temephos. **Pest Management Science**, 64:676–80.
- McKee GJ, Carter CA, Chalfant JA, Goodhue RE, Zalom FG. 2008. Pesticide Resistance, Population Dynamics, and Invasive Species Management. **Contractor and Cooperator Report**, 43:21-31.

- Medina P, Smagghe G, Budia F, del Estal P, Tirry L, Viñuela E. 2002. Significance of penetration, excretion, and transovarial uptake to toxicity of three insect growth regulators in predatory lacewing adults. **Archives of Insect Biochemistry and Physiology**, 51: 91–101.
- Meyer J, Walker-Jonah A, Hollenberg CP. 1991. Galactokinase Encoded by GAL1 Is a Bifunctional Protein Required for Induction of the GAL Genes in *Kluyveromyces lactis* and Is Able To Suppress the gal3 Phenotype in *Saccharomyces cerevisiae*. **Molecular and Cellular Biology**, 11(11):5454-61.
- Mohandas T, Shapiro LJ, Sparkest RS, Sparkest MC. 1979. Regional assignment of the steroid sulfatase-X-linked ichthyosis locus: Implications for a noninactivated region on the short arm of human X chromosome. **PNAS**, 76(11):5779-83.
- Montella IR, Schama R, Valle D. 2012. Classification of Esterases: An important gene family involved in insecticide resistance. **Memórias Instituto Oswaldo Cruz**, 107(4): 437-49.
- Mouchès C, Magnin M, Bergo J, Silvestri M, Beyssat V, Pasteur N, et al. 1987. Overproduction of detoxifying esterases in organophosphateresistant *Culex* mosquitoes and their presence in other insects. **PNAS**. 84:2113-16.
- Nagy A, Hegyi H, Farkas K, Tordai H, Kozma E, Bányai L, et al. 2008. Identification and correction of abnormal, incomplete and mispredicted proteins in public databases. **BMC Bioinformatics**, 9:353.
- Nardini M, Dijkstra BW. 1999. Alpha/beta hydrolase fold enzymes: the family keeps growing. **Current Opinion in Structural Biology**, 9(6):732-7.
- Nei M, Rooney AP. 2005. Concerted and Birth-and-Death Evolution of Multigene Families. **Annual Review of Genetics**, 39:121–52.
- NCBI. Genome Information by organism. Bethesda, USA; Acesso em 18 de Maio de 2013. Disponível em: <http://www.ncbi.nlm.nih.gov/genome/browse/>.
- Neshich G. 2007. Computational Biology in Brazil. **PLoS Computational Biology**, 3(10):e185.
- Oakeshott JG, Horne I, Sutherland TD, Russell RJ. 2003. The genomics of insecticide resistance. **Genome Biology**, 4: 202.
- Oakeshott JG, Claudianos C, Campbell PM, Newcomb RD, Russell RJ. 2005. Biochemical Genetics and Genomics of Insect Esterases. **Comprehensive molecular insect science**, 5:309-61.
- Oakeshott JG, Johnson RM, BerenbaumMR, Ranson H, Cristino AS, Claudianos C. 2010. Metabolic Enzymes Associated with Xenobiotic and Chemosensory Responses in *Nasonia vitripennis*. **Insect Molecular Biology**, 19(Suppl. 1):147–63.
- Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, et al. 1992. The alpha/beta hydrolase fold. **Protein Engineering**, 5(3):197-11.

- Olson PF, Fessler LI, Nelson RE, Sterne1 RE, Campbell AG, Fessler JH. 1990. Glutactin, a novel *Drosophila* basement membranerelated glycoprotein with sequence similarity to serine esterases. **The EMBO Journal**, 9(4):1219-27.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. **Nature Genetics**, 39(10):1256–60.
- Philippe H, Germot A, Moreira D. 2000. The new phylogeny of eukaryotes. **Current Opinion in Genetics & Development**, 10(6):596-601.
- Poupardin R, Riaz MA, Vontas J, David JP, Reynaud S. 2010. Transcription profiling of eleven cytochrome P450s potentially involved in xenobiotic metabolism in the mosquito *Aedes aegypti*. **Insect Molecular Biology**, 19(2):185-93.
- Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, et al. 2010. Amplification of a Cytochrome P450 Gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. **Plos Genetics**, 6:e1000999.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. 2012. The Pfam protein families database. **Nucleic Acids Research**, 40(Database issue):D290-D301.
- Qian W, Zhang J. 2008. Gene Dosage and Gene Duplicability. **Genetics**, 179:2319–24.
- R Development Core Team. 2011. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria.
- Rabinow L, Nguyen-Huynh AT, Birchler JA. 1991. A trans-Acting Regulatory Gene That Inversely Affects the Expression of the white, brown and scarlet Loci in *Drosophila*. **Genetics**, 129:463-80.
- Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. 2010. Comparative analysis of detoxification enzymes in *Acyrtosiphon pisum* and *Myzus persicae*. **Insect Molecular Biology**, 19(2):155-64.
- Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. 2002. Evolution of Supergene Families Associated with Insecticide Resistance. **Science**, 298(5591):179-81.
- Raymond M, Beyssat-Arnaouty V, Sivasubramanian N, Mouchès C, Georghiou GP, Pasteur N. 1989. Amplification of various esterase B's responsible for organophosphate resistance in *Culex* mosquitoes. **Biochemical Genetics**, 27:417–23.
- Reddy BPN, Rao BP, Prasad GBKS, Raghavendra K. 2012. Identification and classification of detoxification enzymes from *Culex quinquefasciatus* (Diptera: Culicidae). **Bioinformatics**, 8(9):430-36.
- Reed W, Carroll J, Agramonte A, Lazear JW. 1900. The Etiology of Yellow Fever—A Preliminary Note. **Public Health Paper Report**, 26:37–53.



- Ribeiro JMC, Alarcon-Chaidez F, Francischetti IMB, Mans BJ, Mather TN, Valenzuela JG, et al. 2006. An annotated catalog of salivary gland transcripts from Ixodes scapularis ticks. **Insect Biochemistry and Molecular Biology**, 36:111–29.
- Robin C, Russell RJ, Medveczky KM, Oakeshott JG. 1996. Duplication and divergence of the genes of the  $\alpha$ -esterase cluster of *Drosophila melanogaster*. **Journal of Molecular Evolution**, 43(3):241-52.
- Robin C, Bardsley LMJ, Coppin C, Oakeshott, JG. 2009. Birth and death of genes and functions in the beta-esterase cluster of *Drosophila*. **Journal of Molecular evolution**, 69(1):10–21.
- Russell RJ, Scott C, Jackson CJ, Pandey R, Pandey G, Taylor MC, et al. 2011. The evolution of new enzyme function: lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects. **Evolutionary Applications**, 4(2):225-48.
- Russell RJ, Claudianos C, Campbell PM, Horne I, Sutherland TD, Oakeshott JG. 2004. Two major classes of target site insensitivity mutations confer resistance to organophosphate and carbamate insecticides. **Pesticide Biochemistry and Physiology**, 79(3):84-93.
- Ryan TJ, Grant SGN. 2009. The origin and evolution of synapses. **Nature Reviews Neuroscience**, 10(10):701-12.
- Saavedra-Rodriguez K, Suarez AF, Salas IF, Strode C, Ranson H, Hemingway J, et al. 2011. Transcription of detoxification genes after permethrin selection in the mosquito *Aedes aegypti*. **Insect Molecular Biology**, 21(1):61–77.
- Scott JG, Zhang L. 2003. The house fly aliesterase gene (M $\alpha$ E7) is not associated with insecticide resistance or P450 expression in three strains of house fly. **Insect Biochemistry and Molecular Biology**, 33:139–44.
- Seabra-Junior ES, Souza EM e Mesquita RD. 2011. FAT - Functional Analysis Tool. **Software n**, Brasil. 11083-6, INPI.
- Schatzmayr HG, Nogueira RMR, Rosa APAT. 1986. An outbreak of dengue virus at Rio de Janeiro. **Memórias do Instituto Oswaldo Cruz**, 81:245-46.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. 2005. Segmental Duplications and Copy-Number Variation in the Human Genome. **The American Journal of Human Genetics**, 77:78–88.
- Silva AX, Jander G, Samaniego H, Ramsey JS, Figueroa CC. 2012. Insecticide Resistance Mechanisms in the Green Peach Aphid *Myzus persicae* (Hemiptera: Aphididae) I: A Transcriptomic Survey. **PLoS ONE**, 7(6):e36366.
- Silva JJ, Mendes J. 2007. Susceptibility of *Aedes aegypti* (L) to the insect growth regulators diflubenzuron and methoprene in Uberlândia, State of Minas Gerais. **Revista da Sociedade Brasileira de Medicina Tropical**, 40(6):612-16.
- Song J, Ichtchenko K, Südhof TC, Brose N. 1999. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. **PNAS**, 96:1100–05.

- Sparks TC, Willis W, Shorey HH, Hammock BD. 1979. Haemolymph juvenile hormone esterase activity in synchronous last instar larvae of the cabbage looper, *Trichoplusia ni*. **Journal of Insect Physiology**, 25(2):125–32.
- Sparks TC, Hammock BD, Riddiford LM. 1983. The haemolymph juvenile hormone esterase of *Manduca sexta* (L.)—inhibition and regulation. **Insect Biochemistry**, 13(5):529–41.
- Stamatakis A. 2006. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. **Bioinformatics**, 22(21):2688-90.
- Stankiewicz P, Shaw CJ, Withers M, Inoue K, Lupski JR. 2004. Serial segmental duplications during primate evolution result in complex human genome architecture. **Genome Research**, 14:2209–20.
- Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, et al. 2008. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. **Insect Biochemistry and Molecular Biology**, 38:(1)113-23.
- Suman DS, Parashar BD, Prakash S. 2010. Efficacy of Various Insect Growth Regulators on Organophosphate Resistant Immatures of *Culex quinquefasciatus* (Diptera: Culicidae) from Different Geographical Areas of India. **Journal of Entomology**, 7: 33-43.
- Sun M, Xing G, Yuan L, Gan G, Knight D, With SI, et al. 2011. Neuroligin 2 Is Required for Synapse Development and Function at the *Drosophila* Neuromuscular Junction. **The Journal of Neuroscience**, 31(2):687– 99.
- Swanson WJ. 2003. Adaptive evolution of genes and gene families. **Current Opinion in Genetics & Development**, 13:617–22.
- Syvanen M, Zhou ZH, Wang JY. 1994. Glutathione transferase gene family from the housefly *Musca domestica*. **Molecular and General Genetics**, 245(1):25-31.
- Syvanen M, Zhou Z, Wharton J, Goldsbury C, Clark A. 1996. Heterogeneity of the glutathione transferase genes encoding enzymes responsible for insecticide degradation in the housefly. **Journal of Molecular Evolution**, 43(3):236-40.
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A Genomic Perspective on Protein Families. **Science**, 278(5338):631-37.
- Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. **Nature Reviews Genetics**, 12(10):692-702.
- Tegoni, M, Campanacci, V, Cambillau, C. (2004). Structural aspects of sexual attraction and chemical communication in insects. **Trends in Biochemical Sciences**, 29(5):257–64.
- Temeyer KB, Tuckow AP, Brake DK, Li AY, Pérez de León AA. 2013. Acetylcholinesterases of blood-feeding flies and ticks. **Chemico-Biological Interactions**, 203(1):319-22.

- Torrents D, Suyama M, Zdobnov E, Bork P. 2003. A Genome-Wide Survey of Human Pseudogenes. **Genome Research**, 13:2559–67.
- Trautwein MD, Wiegmann BM, Beutel R, Kjer KM, Yeates DK. 2012. Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. **Annual Reviews of Entomology**. 57:449–68.
- Tsubota T, Minakuchi C, Nakakura T, Shinoda T, Shiotsuki T. 2010. Molecular characterization of a gene encoding juvenile hormone esterase in the red flour beetle, *Tribolium castaneum*. **Insect Molecular Biology**, 19(4):527–35.
- Tsubota T, Shiotsuki T. 2010a. Genomic analysis of carboxyl/ cholinesterase genes in the silkworm *Bombyx mori*. **BMC Genomics**, 11:377.
- Tsubota T, Shiotsuki T. 2010b. Genomic and phylogenetic analysis of insect carboxyl/cholinesterase genes. *Journal of Pesticide Science*, 35(3):310–14.
- Tunaz H, Uygun N. 2004. Insect Growth Regulators for Insect Pest Control. **Turkish Journal of Agriculture and Forestry**, 28:377-387.
- Vick JE, Gerlt JA. 2007. Evolutionary potential of (beta/alpha)<sub>8</sub>-barrels: stepwise evolution of a "new" reaction in the enolase superfamily. **Biochemistry**, 46(50):14589-97.
- Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004. Structure, function and evolution of multidomain proteins. **Current Opinion in Structural Biology**, 14(2):208–16.
- Vogel C, Chothia C. 2006. Protein Family Expansions and Biological Complexity. **PLoS Computational Biology**, 2(5):e48.
- Vontas JG, Small GJ, Hemingway J. 2001. Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in *Nilaparvata lugens*. **Biochemical Journal**, 357:65–72.
- Wah DA, Hirsch JA, Dorner LF, Schildkraut I, Aggarwal AK. 1997. Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, 388(6637):97-100.
- Waliszewski SM, Gomez-Arroyo S, Infanzon RM, Villalobos-Pietrini R, Hart MM. 2003. Comparison of organochlorine pesticide levels between abdominal and breast adipose tissue. **Bulletin of Environmental Contamination and Toxicology**, 71(1):156-62.
- Wang J, Li S, Zhang Y, Zheng H, Xu Z, Ye J, et al. 2003. Vertebrate gene predictions and the problem of large genes. **Nature Reviews Genetics**, 4(9):741-9.
- Ware, GW. 2000. *The Pesticide Book*, 5th Ed, Thomson Publications, Fresno, California, 418p.
- Weill M, Lutfalla G, Mogensen K, Chandre F, Berthomieu A, Berticat C, et al. 2003. Comparative genomics: Insecticide resistance in mosquito vectors. **Nature**, 423(6936):136-37.

- Weill M, Berthomieu A, Berticat C, Lutfalla G, Negre V, Pasteur N, et al. 2004. Insecticide resistance: a silent base prediction. **Current Biology**, 14(14):552–3.
- Whitfield JB, Kjer KM. 2008. Ancient Rapid Radiations of Insects: Challenges for Phylogenetic Analysis. **Annual Review of Entomology**, 53:449–72.
- Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. 1996. Identification of mutations in the housefly para-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. **Molecular and General Genetics**, 252(1-2):51-60.
- World Health Organization. 2006. Report of the Scientific Working Group meeting on dengue. Available: <http://apps.who.int/tdr/svc/publications/tdrresearch-publications/swgreport-dengue> Accessed 26/04/2013.
- World Health Organization. 2013. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes. Available: <http://www.who.int/malaria/publications/atoz/9789241505154/en/index.html>
- Xu C, Li CY, Kong AT. 2005. Induction of Phase I, II and III Drug Metabolism/Transport by Xenobiotics. **Archives of Pharmacal Research**, 28(3):249-68.
- Yan L, Yang P, Jiang F, Cui N, Ma E, Qiao C, et al. 2012. Transcriptomic and phylogenetic analysis of *Culex pipiens quinquefasciatus* for three detoxification gene families. **BMC Genomics**, 13:609.
- Yoder MD, Thomas LM, Tremblay JM, Oliver RL, Yarbrough LR, Helmkamp GM Jr. 2001. Structure of a multifunctional protein. Mammalian phosphatidylinositol transfer protein complexed with phosphatidylcholine. **The Journal of Biological Chemistry**, 276(12):9246-52.
- Yu QY, Lu C, Li WL, Xiang ZH, Zhang Z. 2009, Annotation and expression of carboxylesterases in the silkworm, *Bombyx mori*. **BioMed Central Genomics**, 10:553.
- Zhao J, Collinsa HL, Sheltona AM. 2010. Testing insecticide resistance management strategies: mosaic versus rotations. **Pest Management Science**. 66:1101–05.
- Zhang J. 2003. Evolution by gene duplication: an update. **Trends in Ecology and Evolution**, 18(6):292-98.

# Anexos

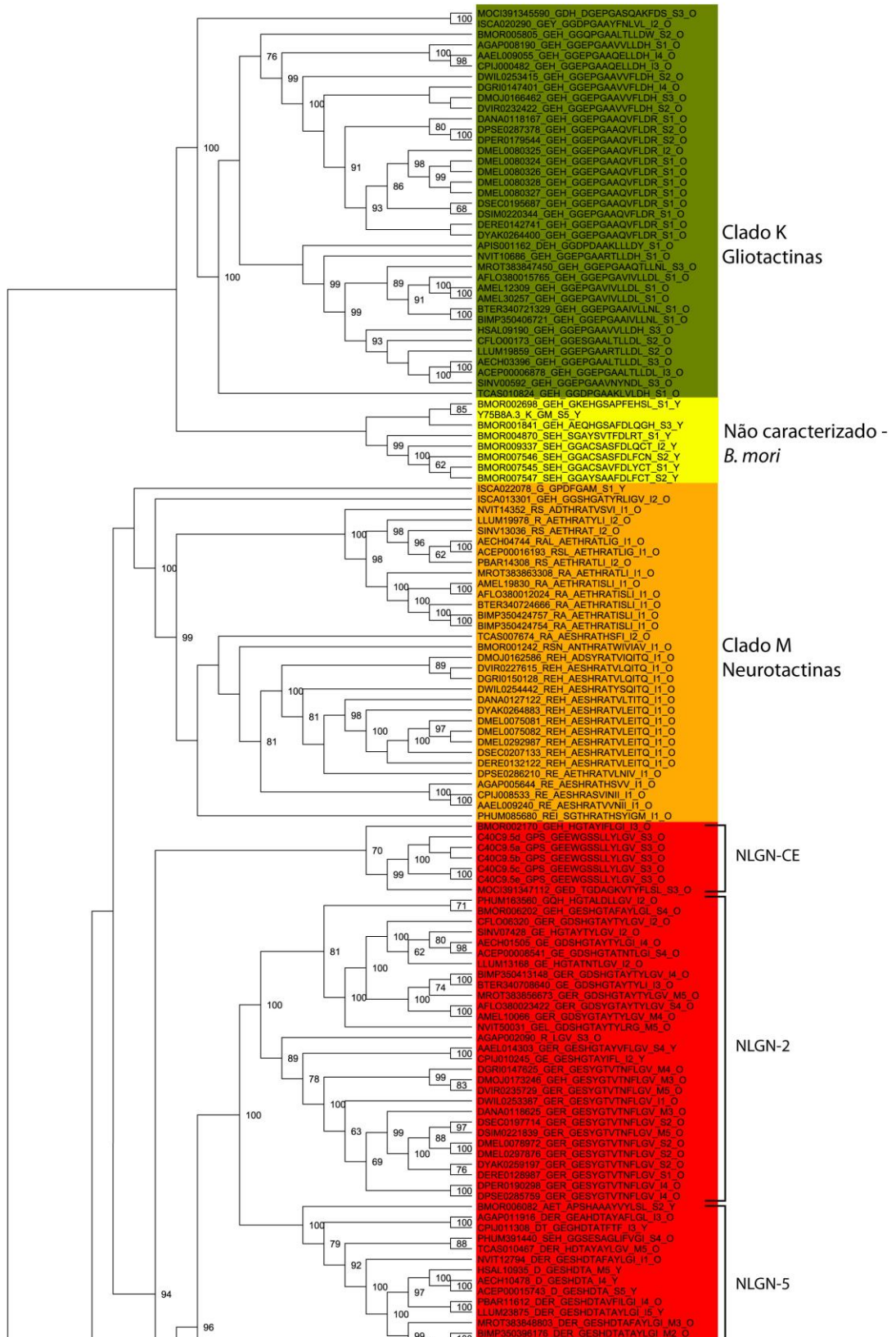


Figura S1 Continuação

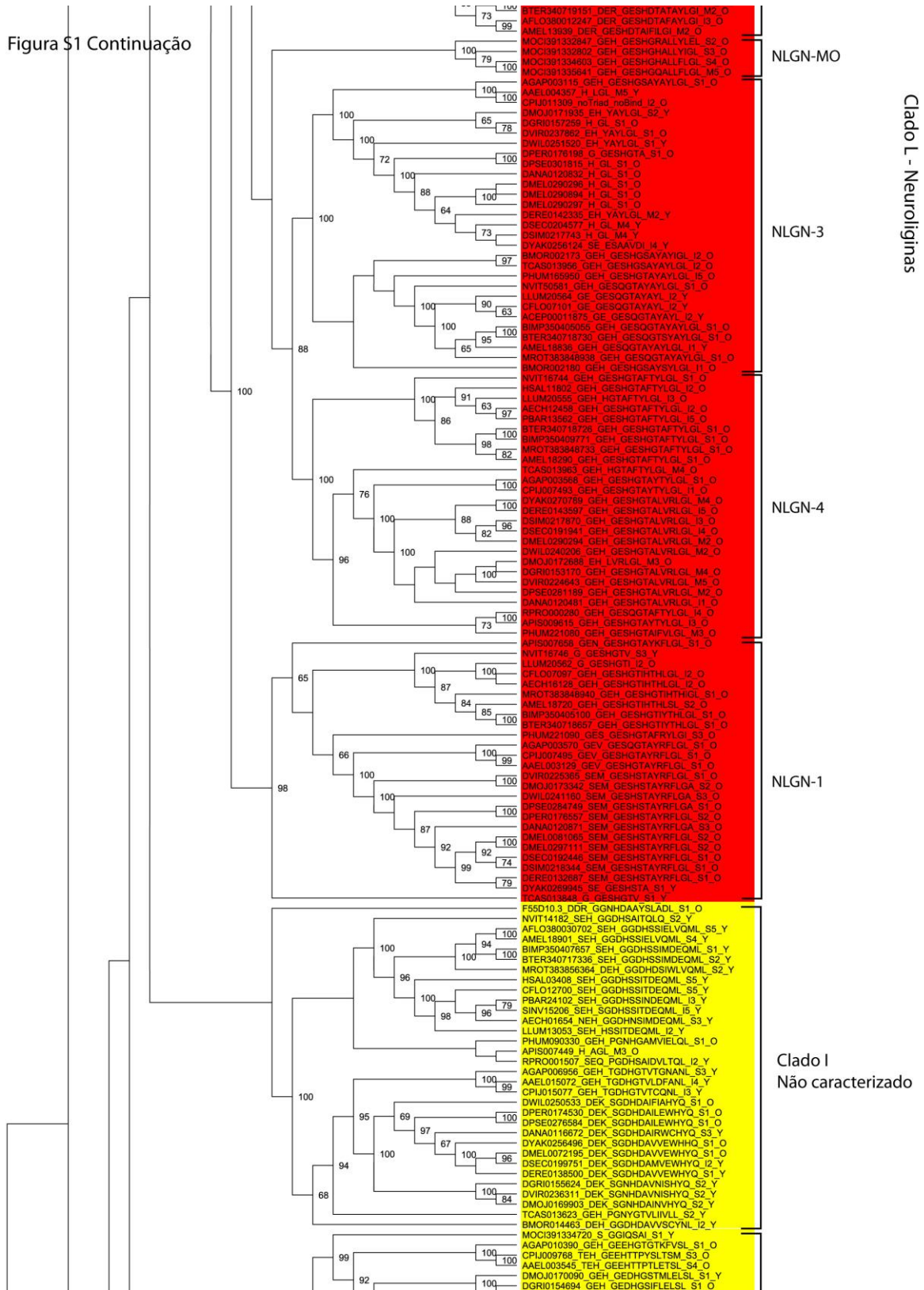




Figura S1 Continuação

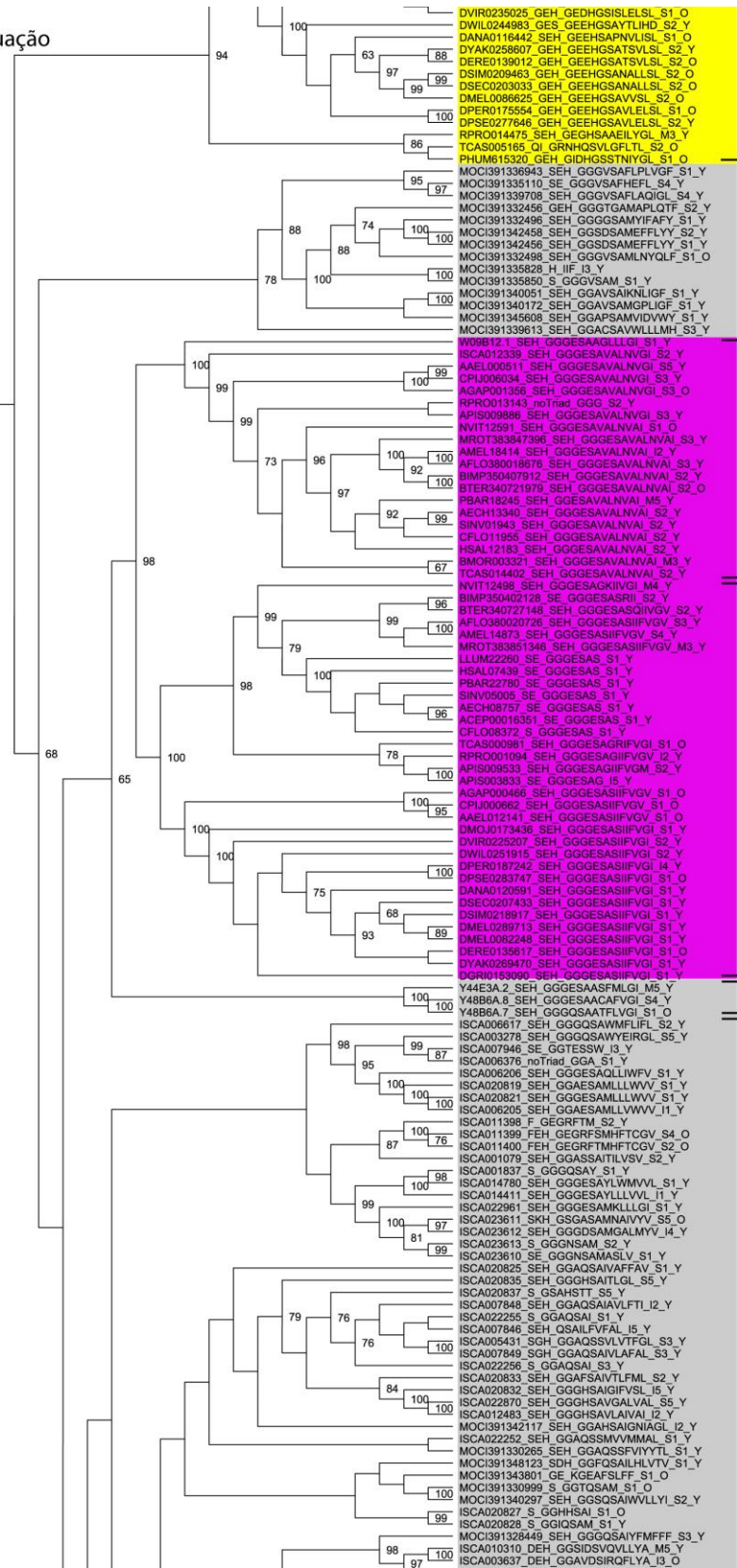


Figura S1 Continuação

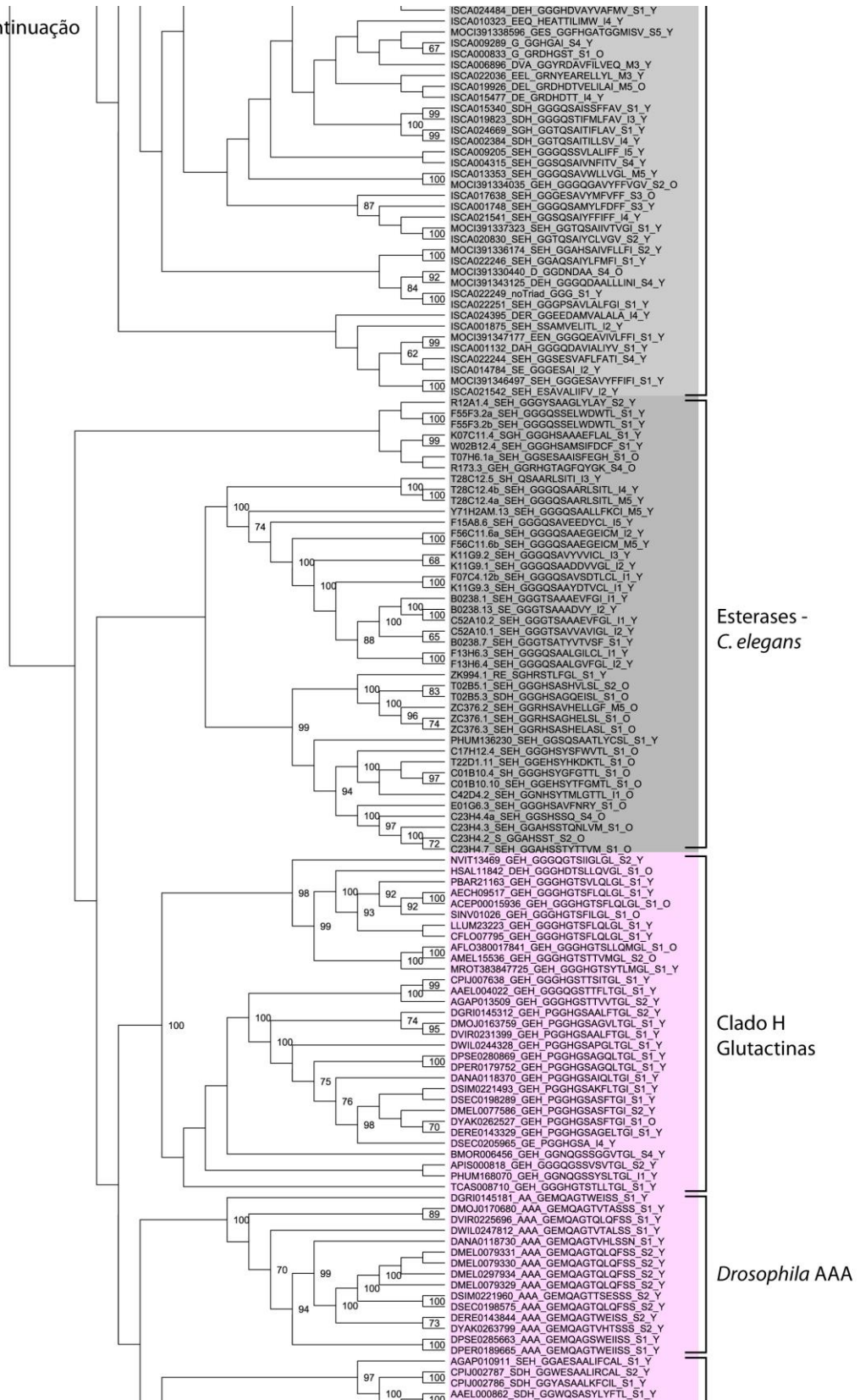
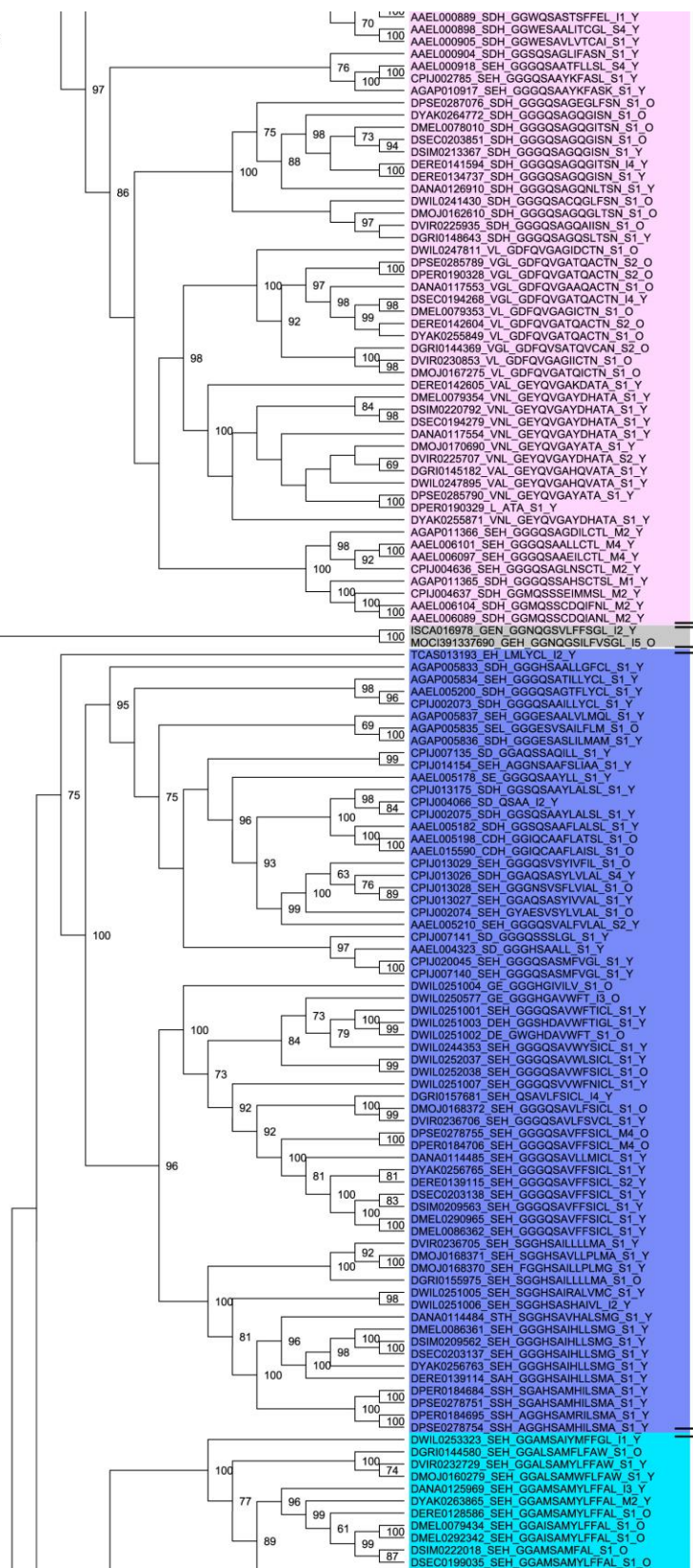




Figura S1 Continuação

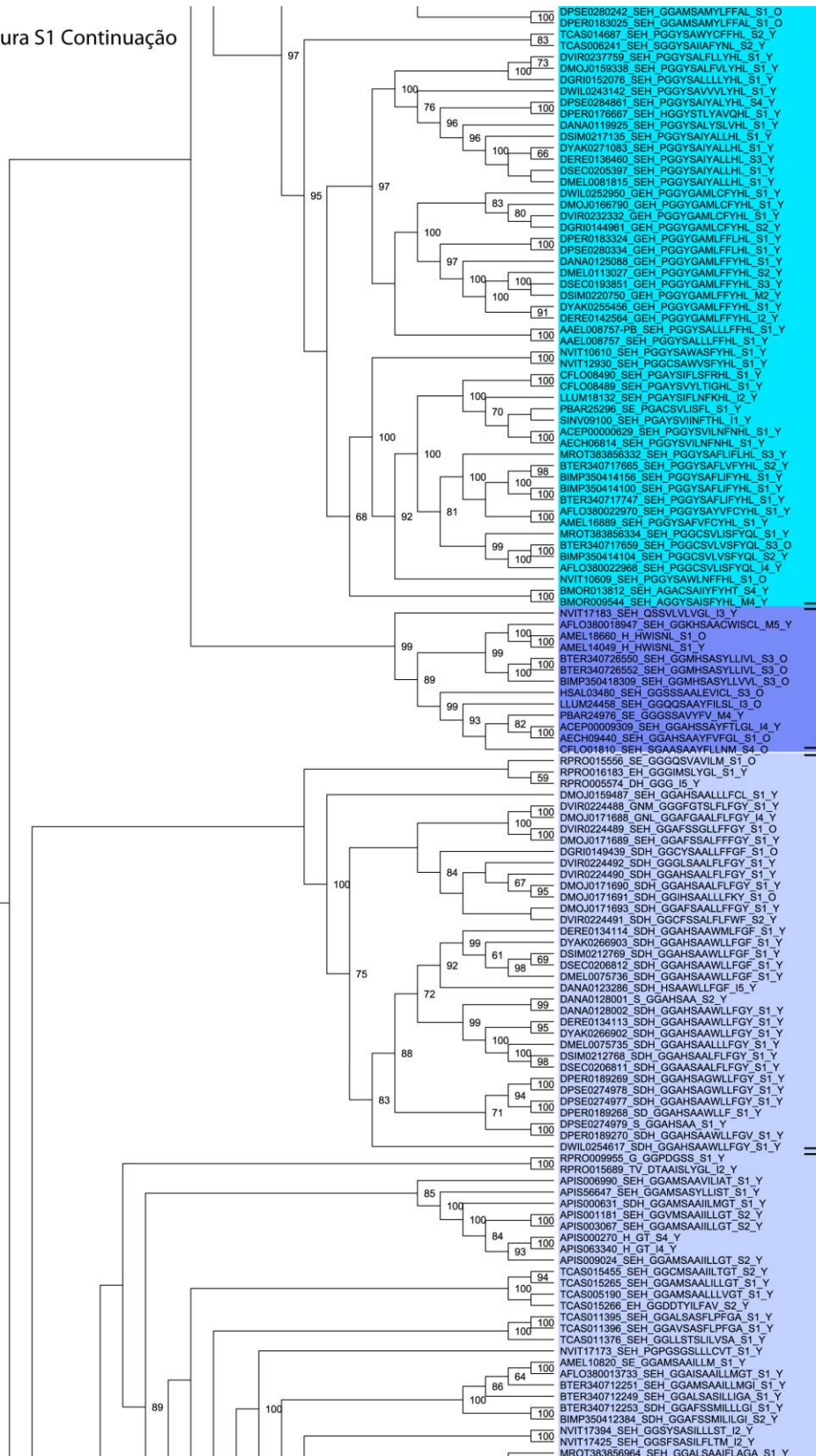


Glutactinas Diptera

Não caracterizado - Arachnida

Clado F JHE - Diptera

Figura S1 Continuação



Clado D  
Tegumentares

Clado F  
Não caracterizado -  
Hymenoptera

Clado E  
Beta-esterase -  
*Drosophila*

Figura S1 Continuação

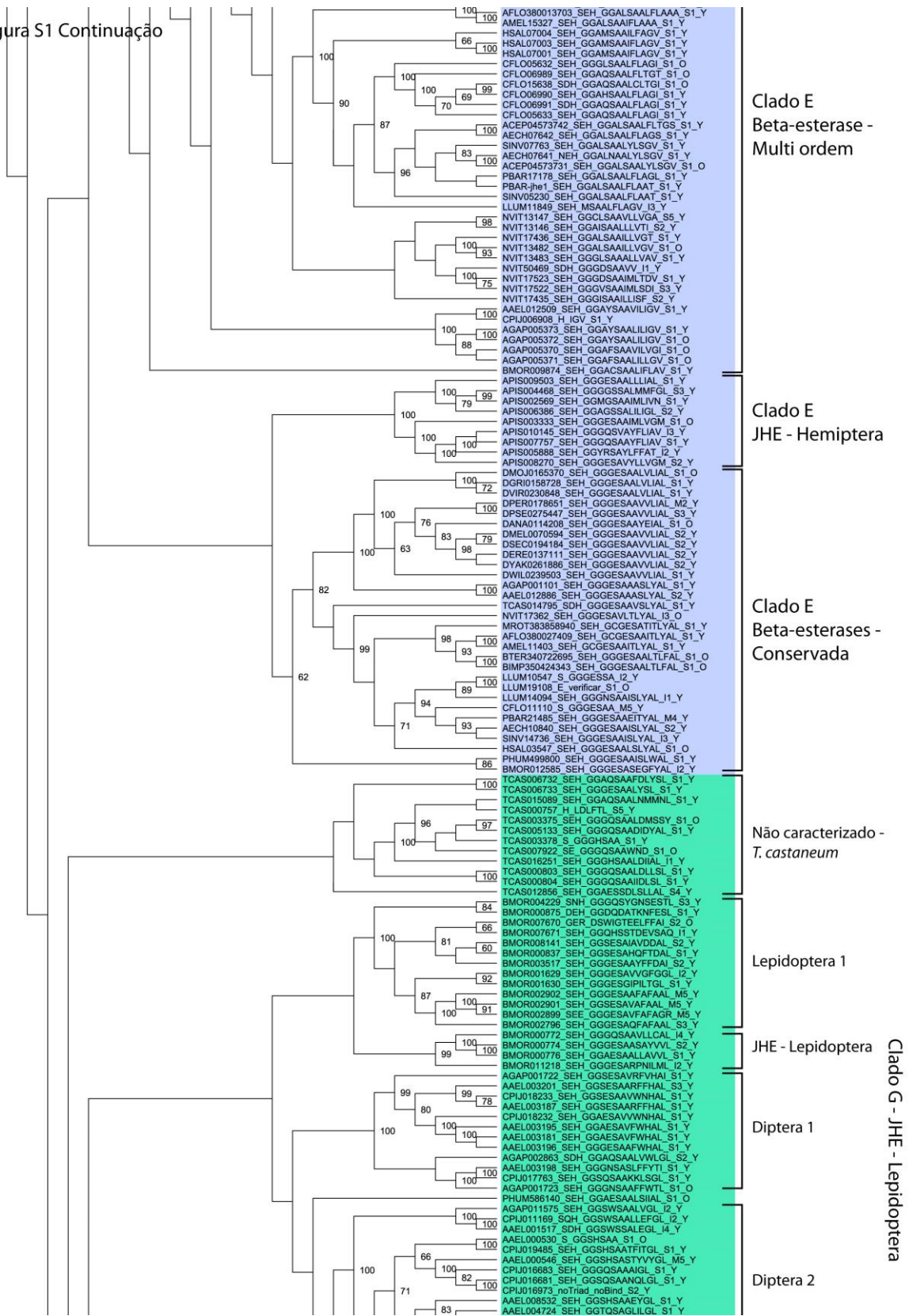




Figura S1 Continuação

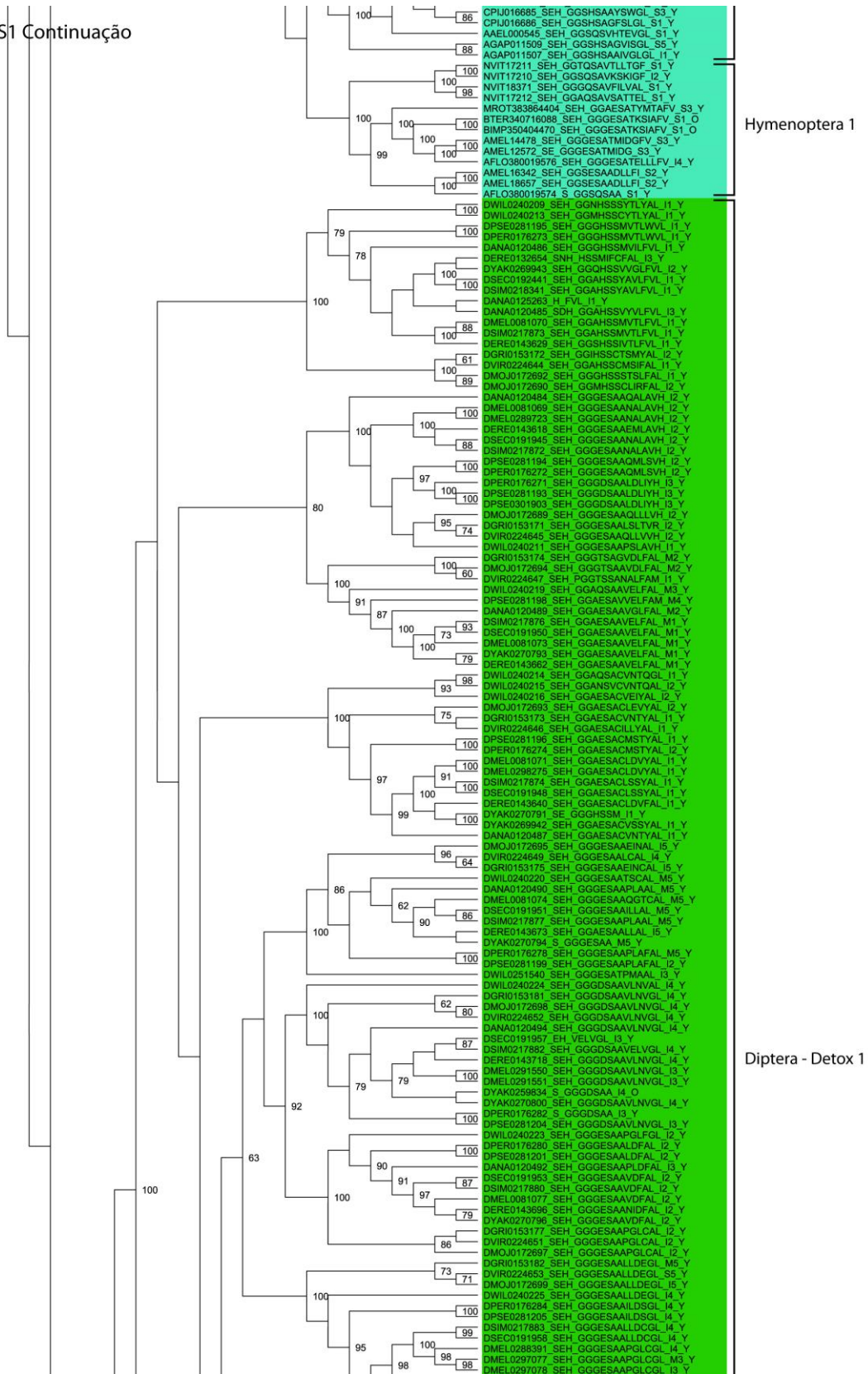
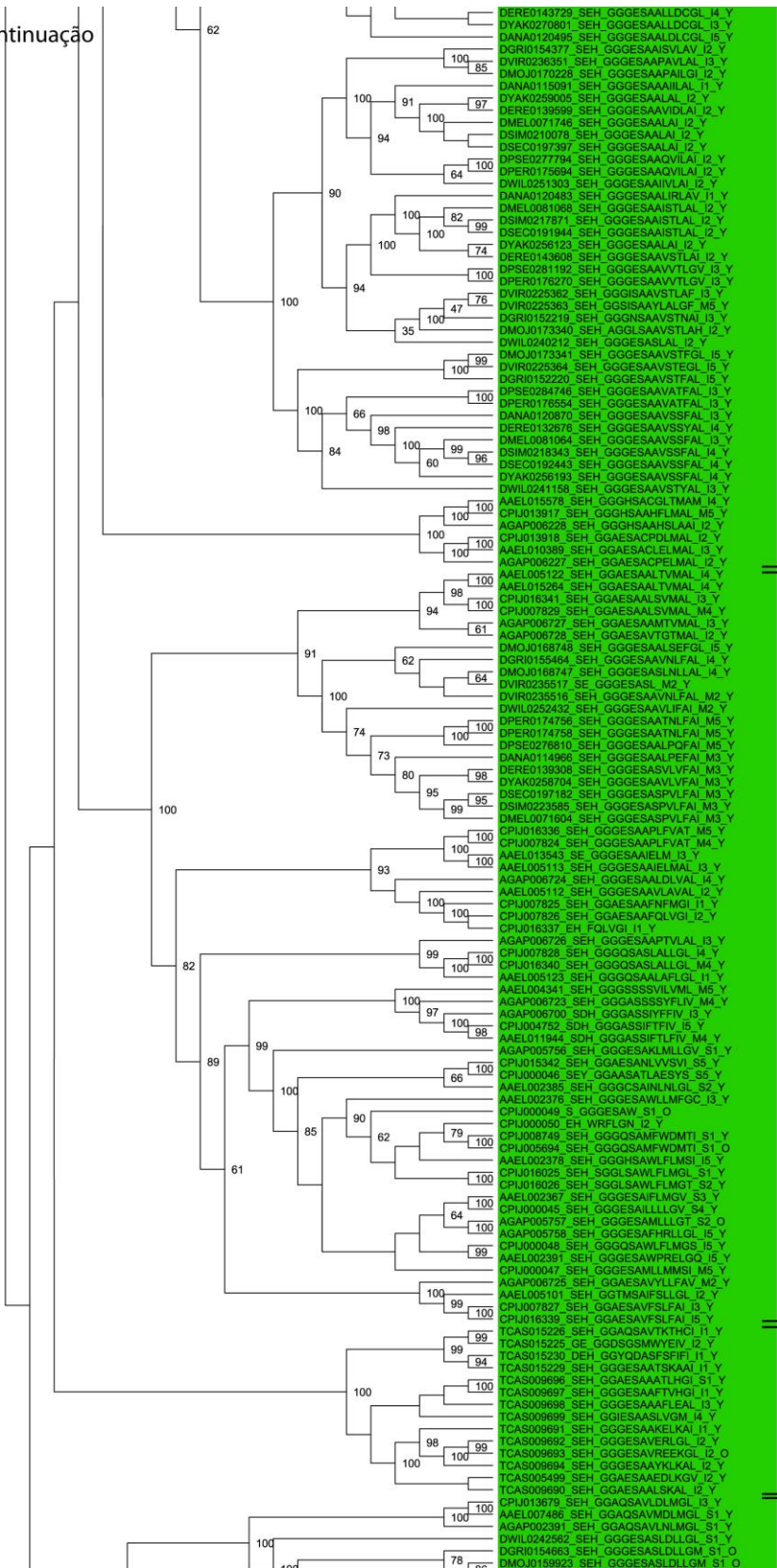


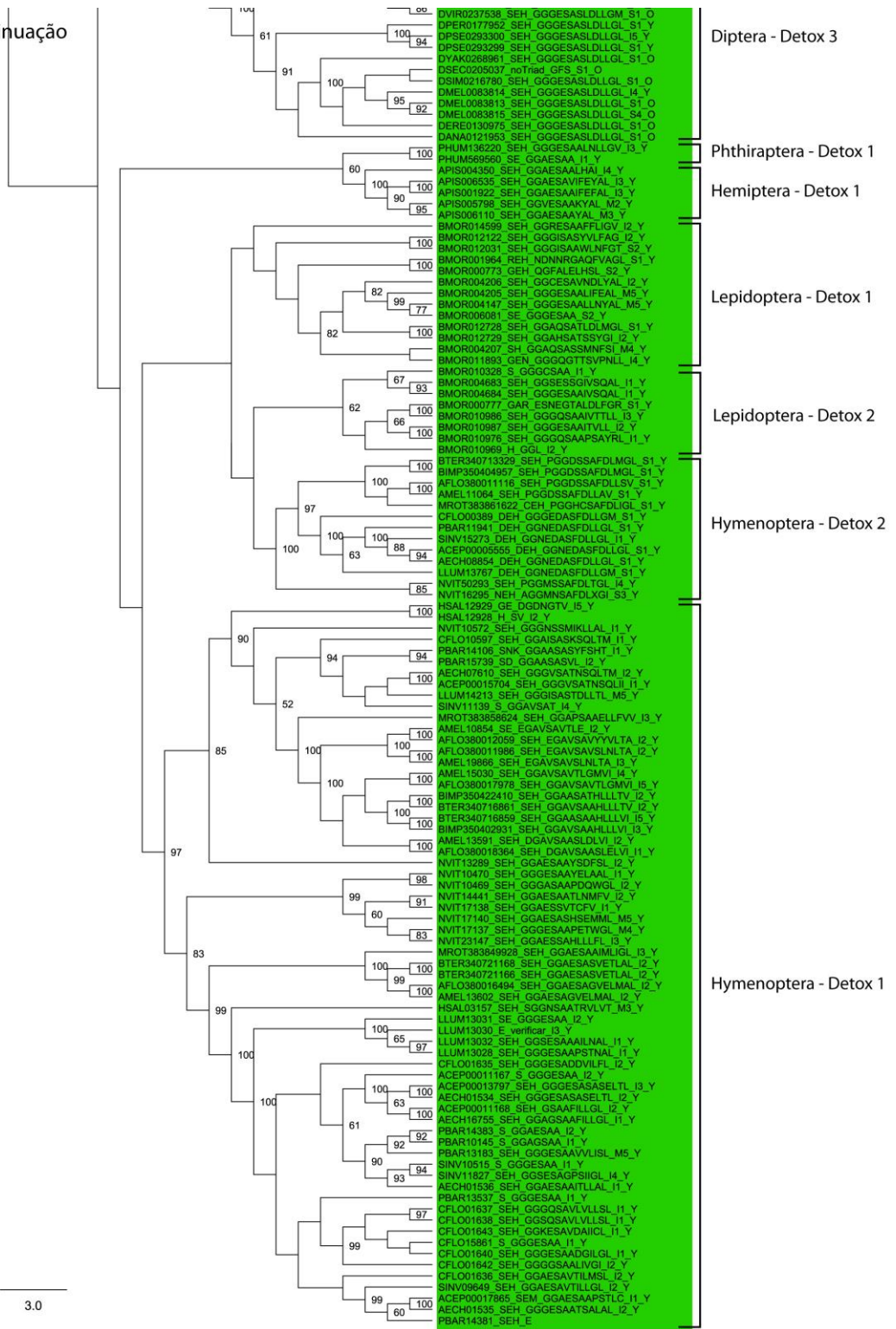
Figura S1 Continuação



Diptera - Detox 2

Coleoptera - Detox 1

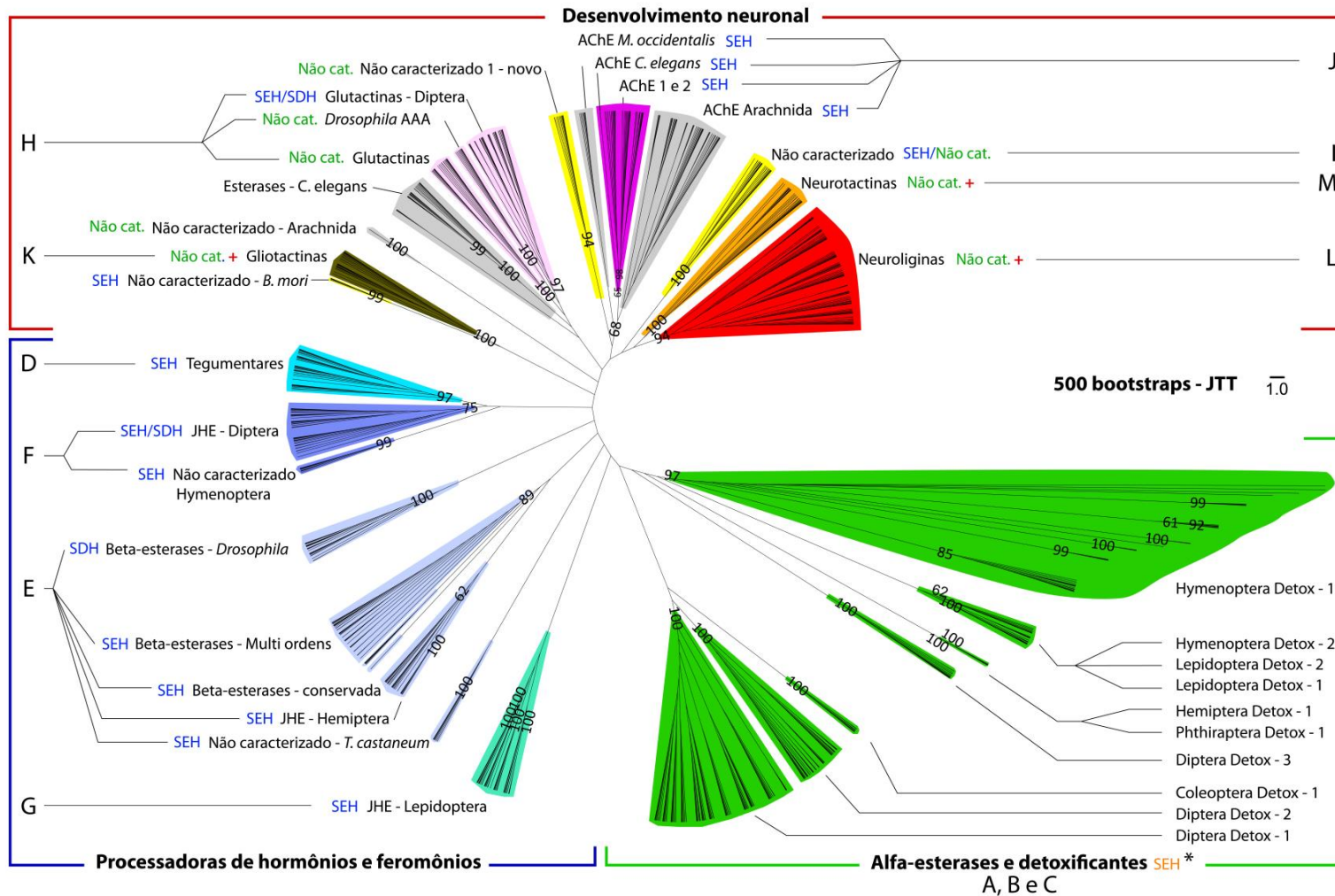
Figura S1 Continuação



3.0

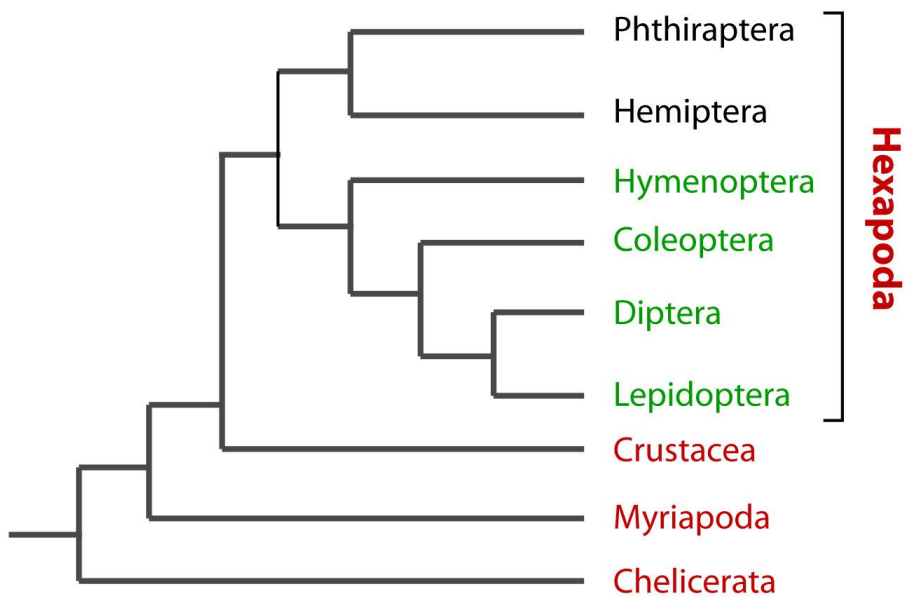
**Figura S1.** Árvore filogenética enraizada dos membros da família CCE. Apenas os valores estatísticos (acima de 60%) são mostrados. As barras laterais delimitam os clados encontrados.





**Figura S2.** Árvore filogenética não enraizada dos membros da família CCE. Apenas os valores estatísticos (acima de 60%) são mostrados. Clados com proteínas transmembranares estão representadas por um sinal + vermelho. Clados com proteínas majoritariamente catalíticas secretadas estão representados pelo código que representa a tríade catalítica de três letras em azul (SEH ou SDH). Clados com proteínas majoritariamente endereçadas para diferentes locais estão representadas pelo código SEH na cor laranja.





**Figura S3.** Esquema resumido das ordens do subfilo Hexapoda representadas nesse trabalho. Holometabolas estão representados em verde; Paraneoptera em preto; Subfilos estão mostrados em vermelho.

**Tabela S1.** Taxonomia das espécies estudadas no presente trabalho e a respectiva correlação entre abreviação usada no prefixo da identificação das sequências nas árvores filogenéticas e prefixo original do banco de dados.

Filo	Classe	Ordem	Família	Espécies	Abreviação	Prefixo BD
Arthropoda	Insecta	Hymenoptera	Formicidae	<i>Acromyrmex echinaior</i>	AECH	AECH
				<i>Pagonomyrmex barbatus</i>	PBAR	PB
				<i>Linepithema humile</i>	LLUM	LH
				<i>Atta cephalotes</i>	ACEP	ACEP
				<i>Solenopsis invicta</i>	SINV	SI
				<i>Camponotus floridanus</i>	CFLO	CFLO
				<i>Harpegnathus saltator</i>	HSAL	HSAL
		Apidae	<i>Apis mellifera</i>	AMEL	GB	
			<i>Apis florea</i>	AFLO	gi	
			<i>Bombus impatiens</i>	BIMP	gi	
			<i>Bombus terrestris</i>	BTER	gi	
		Megachilidae	<i>Megachile rotundata</i>	MROT	gi	
		Pteromalidae	<i>Nasonia vitripennis</i>	NVIT	NV	
		Diptera	Drosophilidae	<i>Drosophila grimshawi</i>	DGRI	FBpp
				<i>Drosophila simulans</i>	DSIM	FBpp
				<i>Drosophila erecta</i>	DERE	FBpp
				<i>Drosophila persimilis</i>	DPER	FBpp
				<i>Drosophila sechellia</i>	DSEC	FBpp
				<i>Drosophila ananassae</i>	DANA	FBpp
				<i>Drosophila yakuba</i>	DYAK	FBpp
				<i>Drosophila mojavensis</i>	DMOJ	FBpp
				<i>Drosophila pseudoobscura</i>	DPSE	FBpp
				<i>Drosophila virilis</i>	DVIR	FBpp
<i>Drosophila willinstoni</i>	DWIL			FBpp		
<i>Drosophila melanogaster</i>	DMEL			FBpp		
Culicidae	<i>Anopheles gambiae</i>			AGAP	AGAP	
	<i>Aedes aegypti</i>			AAEL	AAEL	
	<i>Culex quinquefasciatus</i>			CPIJ	CPIJ	
Hemiptera	Aphididae			<i>Acyrtosiphon pisum</i>	APIS	ACYPI
	Reduviidae	<i>Rodnius prolixus</i>	RPRO	RPRC		
Phthiraptera	Pediculidae	<i>Pediculus humanus</i>	PHUM	PHUM		
Coleoptera	Tenebrionidae	<i>Tribolium castaneum</i>	TCAS	TC		
Lepidoptera	Bombycidae	<i>Bombyx mori</i>	BMOR	BGIBMGA		
Arachnida	Mesostigmata	Ixodida	Ixodidae	<i>Ixodes scapularis</i>	ISCA	ISCW
				Phytoseiidae	<i>Metaseiulus occidentalis</i>	MOCI
Nematoda	Secernentea	Rhabditida	Rhabditidae	<i>Caenorhabditis elegans</i>	Nomenclatura wormbase	Nomenclatura wormbase

**Tabela S2.** Sequências com *e-value* menor que  $1e-58$  não incluídas na filogenia devido à problemas de processamento.

ID	Triade	Endereçamento	TM	<i>e-value</i>	Score
ACEP00002806	NEH	S	Y	3.8e-123	412.3
ACEP00014153	SEH	S	Y	8.6e-165	549.7
ACEP00014908	EH	I	Y	8.6e-78	262.6
ACEP00017864	E	I	Y	6.9e-60	203.5
APIS084883	SEY	S	O	2.2e-117	394.3
BIMP350412381	SEH	I	Y	3.7e-162	540.3
BMOR000775	EH	I	Y	1e-82	278.6
BMOR003724	SEN	I	Y	1.7e-60	205.2
BMOR004934	SEH	S	Y	2.3e-116	389.6
BMOR010505	SEG	S	Y	2.8e-111	372.8
BMOR010988	SEH	I	Y	1.2e-127	426.9
BMOR012722	SEH	S	O	2.9e-82	277.1
CPIJ005122	SEH	S	Y	5.2e-84	283.2
DGRI0152221	SEM	S	O	7.5e-126	420.9
DPER0176269	GEH	M	O	6.5e-132	441.1
DPER0176276	SEH	M	Y	1.6e-116	390.3
DPER0176892	DH	S	O	2.3e-73	247.9
DSIM0220791	VGL	S	O	3.4e-98	329.7
DWIL0240218	EH	M	Y	1.2e-59	202.6
DWIL0244354	NE	S	Y	3.8e-236	785
DYAK0270790	SEH	I	Y	8.9e-100	335
HSAL00833	DEH	S	Y	1e-104	351.5
HSAL07071	RS	I	O	1e-72	245.9
LLUM22629	SEH	S	Y	1.2e-167	558.9
LLUM25422	SEH	I	Y	4.3e-120	402
MOCI391326926	E	S	O	1.4e-60	205.2
NVIT10611	SEH	S	Y	1.7e-136	456.4
NVIT20077	SEH	S	Y	2.3e-111	373.5
NVIT50366	SEH	S	O	4.4e-105	352.7
PBAR14381	SEH	I	Y	7.4e-60	203.3
PBAR17618	SE	S	O	7.2e-90	302.4
PBAR19852	EH	S	O	4.1e-83	280.1
PHUM391440	SEH	S	O	1.1e-134	449.7
PHUM490200	SEH	I	Y	6.5e-176	585.7
PHUM617310	GEH	I	O	6.9e-129	430.5
RPRO009957	NS	I	Y	4.3e-69	234.7
RPRO009972	noTriad	I	Y	6.6e-64	217.6
SINV09844	G	I	Y	8e-62	209.8