

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E EVOLUÇÃO

Evolução molecular dos genes *doublesex* e *fruitless* em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae)

Iderval da Silva Jr. Sobrinho

São Carlos - SP

2012

Iderval da Silva Jr. Sobrinho

Evolução molecular dos genes *doublesex* e *fruitless* em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae)

Orientador:

Prof. Dr. Reinaldo O. A. Alves de Brito

Tese de doutorado apresentada ao Programa de Pós-Graduação em Genética e Evolução do Centro de Ciências Biológicas e da Saúde da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de doutor em Genética e Evolução, área de concentração Genética e Evolução.

São Carlos - SP

2012

Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária/UFSCar

S677em

Sobrinho, Iderval da Silva Júnior.

Evolução molecular dos genes *doublesex* e *fruitless* em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae) / Iderval da Silva Júnior Sobrinho. -- São Carlos : UFSCar, 2012.

135 f.

Tese (Doutorado) -- Universidade Federal de São Carlos, 2009.

1. Genética e evolução. 2. Seleção positiva. 3. Determinação sexual. 4. Seleção natural. I. Título.

CDD: 575 (20^a)

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E EVOLUÇÃO

Evolução molecular nos genes *doublesex* e *fruitless* em moscas-das-frutas do complexo *Anastrepha fraterculus* (Diptera, Tephritidae)

Tese de Doutorado de Iderval da Silva Júnior Sobrinho

Banca Examinadora

Prof. Dr. Reinaldo Alves de Brito

Prof. Dr. Otávio Henrique Thiemann

Prof. Dr. Frederico Guillermo Hoffmann Jauje

Prof. Dr. Carlos Eduardo Guerra Schrago

Prof. Dr. Giancarlo Conde Xavier Oliveira

.....
237

.....
Thiemann

.....
F. Hoffmann

.....
C. E. Guerra

.....
Giancarlo Conde Xavier Oliveira

Dedicatória

Dedico este trabalho à todos que tentaram e tentam tornar este mundo um lugar melhor.

... mais vale a ave que voa sem asas ...

Último sopro de Fernando Perillo e Bororó e C.Ribeiro

Agradecimentos

Agradeço à FAPESP, CNPq e CAPES pelo apoio financeiro e a seus funcionários pelo respeito ao pós-graduando.

Aos funcionários do Departamento de Genética e Evolução por tornarem este departamento um ambiente agradável de se trabalhar.

Aos professores Flávio e Pedro Galleti por disponibilizarem o uso do laboratório e equipamentos nos primeiros anos do doutorado, sem o seu apoio este trabalho jamais existiria.

Aos amigos do Laboratório de Genética de Populações e Evolução (vulgo LabEvol, leia-se *labivel*): Nelci (gel"o"), Gustavo, Bruno, Vanessa (Ban), Fernanda (o MAL condensado), Marquinho, Daniel (hacker do Linux), André (ACR, eu sei que é você), Aline, Victor, Cíntia (e a Camila também) e Carol Kaminsky pela amizade e por sempre darem o apoio e ajuda na hora certa.

Aos meus grandes amigos Débora, Pablo, Rosângela, Andréa Peripato e Fernando Pacheco, pela confiança, pelo apoio nas horas mais desesperadoras e por me mostrarem o sentido da verdadeira amizade.

A todos meus amigos presentes e ausentes do Departamento de Genética e Evolução por toda ajuda que me deram de tão boa vontade.

Por fim agradeço ao Prof. Dr. Reinaldo Brito pelo aceite em me orientar quando muitos haviam me negado esta oportunidade, por me ensinar de tão bom grado e entusiasmo esta maravilhosa área que é a evolução. Em especial estou muito grato pela confiança e respeito que ele sempre demonstrou por mim e por meu trabalho. Digo por fim que para mim foi uma honra ser orientado por ele e ter feito parte de sua equipe.

Resumo

As espécies que formam grupo *Anastrepha fraterculus* estão entre as principais pragas agrícolas da América do Sul e compõem um grupo de espécies intimamente relacionadas de difícil identificação. Devido à grande variabilidade e sobreposição encontrada nos marcadores utilizados em sua identificação, ainda é difícil o estabelecimento da monofilia recíproca das espécies deste grupo. Muitas evidências sugerem que genes relacionados às características sexuais seriam bons candidatos para discriminar espécies com baixa divergência, graças à possibilidade de estarem envolvidos no próprio processo de separação das mesmas e apresentarem uma elevada taxa de diferenciação. A maior parte destes estudos têm se concentrado em genes relacionados à interação entre machos e fêmeas, porém poucos em genes da diferenciação sexual. Dessa forma, estudamos neste trabalho dois genes, *doublesex (dsx)* e *fruitless (fru)*, envolvidos com a determinação sexual com o intuito inicial de avaliar seu possível papel na diferenciação do complexo de espécies *Anastrepha fraterculus*. Analisamos também os padrões de evolução molecular destes genes, examinando desde dados interespecíficos de alta divergência e dados intraespecíficos de baixa divergência. Neste processo avaliamos a capacidade de diferentes metodologias em detectar sinais de padrões seletivos específicos nas regiões estudadas, bem como testamos uma nova abordagem para estudo de evolução molecular em dados populacionais. Nesta nova abordagem conciliamos em uma mesma análise dados de frequência de mutações sinônimas e não-sinônimas, da sua posição na rede haplotípica, bem como da capacidade das mutações não-sinônimas em mudar as propriedades físico-químicas de aminoácidos. Diferentemente do encontrado em estudos anteriores, tanto *dsx* quanto *fru* mostraram sinais inequívocos de seleção positiva atuando em sua diferenciação, embora revelando processos diferentes. Além disso, detectamos um evento de *selective sweep* na região conectora de *fru* em *A. obliqua*, possivelmente envolvido na separação da rede haplotípica em dois grupos, com *A. obliqua* isolada das outras duas espécies. Apesar dos dois genes participarem da mesma cascata de diferenciação sexual, *fru* estabeleceu de forma mais clara a separação, porém ainda não completa, de *A. obliqua* das outras espécies aqui estudadas.

Abstract

Species of the *Anastrepha fraterculus* group are one of the most important agricultural pest from South America and form a group of closely related species whose identification is very difficult. It is problematic to establish reciprocal monophyly of species from this group, because of the great variability and superimposition of morphological traits used in their identification. Many evidences suggest that sex-related genes would be good candidates to discriminate species with low divergence, because they could be involved in the speciation of such species and show high level of differentiation rate. Most of such studies are on genes related to male-female interaction, however few are focused on sex-differentiation genes. We originally aimed to study the role of two sex-determination genes, *doublesex* (*dsx*) and *fruitless* (*fru*), in the differentiation of the species of *Anastrepha fraterculus* complex. We also studied the molecular evolutionary pattern of these genes, from high divergence interspecific to low divergence intraspecific data. In the process we assessed the hability of different methodologies in detecting signals of particular selective patterns in the studied regions, and tested a new approach to investigate patterns of positive selection using population data. In this new approach we combined in the same analysis synonymous and nonsynonymous mutations frequencies, their position in the haplotype network, and nonsynonymous mutations capacity to change amino acid physicochemical properties. Differently of what was proposed in previous works, both *dsx* and *fru* showed distinct signal of positive selection promoting their differentiation. Furthermore, we detected a selective sweep in the *fru* connecting region of *A. obliqua*, possibly involved in the split of haplotype network in two groups, with *A. obliqua* isolated from the other two species. Although both genes are involved in the sex-differentiation cascade, *fru* established in clearer way the separation, although not yet completely, of *A. obliqua* from the other two species here studied.

Lista de Figuras

- Figura 1. Características morfológicas do tórax (A), asa (B) e extremidade do acúleo (ovipositor) (C) utilizadas na identificação de espécies de *Anastrepha*. B e C modificado de (Aluja & Norrbom 2000)..... 16
- Figura A1. Rede haplotípica separada por duas mutações não-sinônimas que levam a alterações radicais nas propriedades físico-químicas dos aminoácidos (8 e 22). Os números sobre os ramos indicam as mutações sinônimas e não-sinônimas internas..... 28
- Figura A2. Árvore filogenética estimada por Máxima Verossimilhança para os dados de contrastes macroevolutivos mostrando o ramo foreground destacado por uma área cinza. Os tamanhos dos ramos estão estimados em substituições nucleotídicas por códon..... 30
- Figura A3. *Bayes Empirical Bayes* mostrando probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) para o ramo *foreground*. A área sombreada representa o domínio de ligação ao DNA. A linha tracejada indica a probabilidade posterior de 0,95. A numeração dos códons não indica a posição real do mesmo no *dsx*. 32
- Figura A4. *Bayes Empirical Bayes* mostrando probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) para o ramo *foreground*. A área sombreada representa o domínio de ligação ao DNA. A linha tracejada indica a probabilidade posterior de 0,95. A numeração dos códons não indica a posição real do mesmo no *dsx*. Os valores de ω médio estão representados no eixo da direita.....34
- Figura A5. Intervalo de confiança de valores de ω estimado para *doublesex* COImt 35
- Figura A6. Distribuição de mutações não-sinônimas que levam a alterações radicais nas propriedades dos aminoácidos. O número presente na denominação da mutação indica a posição do códon onde ela ocorreu e a letra corresponde ao aminoácido codificado pelo códon. Quando ambas as mutações ocorreram num mesmo ponto de coleta foi indicada a presença de homozigotos ou heterozigotos para tais mutações. 36

Figura B1. Árvore filogenética gerada por Máxima Verossimilhança no software Phylip ver 3.68. O ramo <i>foreground</i> está destacado em cinza.....	55
Figura B2. Probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) na região conectora de <i>fru</i> . A linha tracejada indica o limite de 0,95.....	57
Figura B3. Estimativa de áreas da região conectora de <i>fru</i> sob seleção positiva desestabilizadora estimados por janelas deslizantes. Os valores da posição dos códons representam os valores iniciais dos intervalos de 20 códons. A linha horizontal representa o limite de significância de 0,001 (com correção de Bonferroni)	58
Figura B4. Rede de haplótipos da região conectora de <i>fru</i> inferida por parcimônia estatística no TCS ver. 1.21. As linhas espessas representam mutações não-sinônimas e as mais finas sinônimas. Os círculos vazios correspondem a haplótipos inferidos.....	60
Figura B5. Rede de haplótipos do íntron adjacente à região conectora inferida por parcimônia estatística no TCS ver. 1.21.....	61
Figura B6. Probabilidades posteriores de sítios em <i>fru</i> nas oito categorias de w estimadas pela análise Bayes Empirical Bayes sob o modelo M8. Os valores no interior do gráfico indicam os ω médios dos respectivos códons. O sítio com maior valor de ω foi o 202 com $\omega = 1,20$	66
Figura B7. Estimativa de áreas da região conectora de <i>fru</i> sob seleção positiva desestabilizadora estimadas por janelas deslizantes. Os valores das posições de códons representam os valores iniciais dos intervalos de 20 códons. A linha tracejada representa o limite de significância de 0,001 com correção de Bonferroni.....	68

Lista de Tabelas

Tabela A1. Locais amostrados com informação geográfica, haplotípica e de espécies.....	29
Tabela A2. Estimativa de diversidade genética na região de <i>doublesex</i>	30
Tabela A3. Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito nos dados macroevolutivos.....	31
Tabela A4. Contraste dos modelos M1a, MA restrito e MA por LRT nos dados macroevolutivos.	31
Tabela A5. Teste de heterogeneidade das taxas evolutivas entre <i>dsx</i> e COImt.....	33
Tabela A6. Contrastes dos modelos para teste de heterogeneidade de taxas evolutivas entre <i>dsx</i> e COImt.....	35
Tabela A7 . Teste de contingência de mutações sinônimas e não-sinônimas vs ponta e interior.	37
Tabela A8. Contagem de indivíduos gerados por mutações sinônimas e não-sinônimas internas	38
Tabela A9. Análise de heterogeneidade nas distribuições de frequências entre número de indivíduos formados por uma mutação interna e tipo de mutação.....	39
Tabela B1. Locais amostrados com informação geográfica, haplotípica e de espécies.....	49
Tabela B2. Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito.	54
Tabela B3. Comparação por LRT dos modelos M1a, MA e MA restrito.....	54

Tabela B4. Detecção de sítios sob seleção positiva desestabilizadora nas propriedades físico-químicas dos aminoácidos do éxon da região conectora de <i>fru</i>	56
Tabela B5. Estimativa da diversidade genética do primeiro éxon da região conectora e íntron.	61
Tabela B6. Testes de neutralidade: região conectora e íntron.	62
Tabela B7. Análise de contingência de mutações sinônimas e não-sinônimas versus posição na rede de haplótipos (interior ou de exterior).	62
Tabela B8. Contagem de indivíduos gerados por mutações sinônimas e não-sinônimas internas	63
Tabela B9. Teste de heterogeneidade de ω em sítios. Valores de máxima verossimilhança, estimativa de parâmetros e LRTs para a região conectora de <i>fru</i>	64
Tabela B10. Valores da análise de desvios em propriedades físico-químicas estimados no TreeSAAP.....	67

Sumário

Capítulo I - Introdução Geral.....	11
Introdução.....	12
1. Revisão geral das espécies.....	12
2. Cascata de diferenciação sexual em insetos	13
2.1. <i>doublesex</i> : função e organização genômica.....	14
2.2. <i>fruitless</i> : função e organização genômica.....	15
3. Busca de marcadores para o estudo da diferenciação do complexo <i>fraterculus</i>	17
Capítulo II - Evolução molecular do gene <i>doublesex</i> no grupo <i>Anastrepha fraterculus</i> (Diptera, Tephritidae).....	18
Introdução.....	19
Material e Métodos.....	21
1. Extração do DNA e seqüenciamento.....	21
2. Análise da evolução molecular de <i>doublesex</i>	21
3. Análises populacionais de seleção.....	23
3.1 Comparação de taxas evolutivas entre <i>dsx</i> e COI.....	24
3.2 Diversidade genética da região de <i>dsx</i> comum aos sexos	25
Resultados.....	27
1. Diversidade em <i>doublesex</i> e análise filogenética.....	27
2. Comparação das taxas evolutivas e seleção em <i>dsx</i>	30
3. Estudo de taxas evolutivas em dados intrapopulacionais.....	33
3.1 Testes populacionais de neutralidade	37
Discussão.....	40
Capítulo III - Evolução molecular do gene <i>fruitless</i> no grupo <i>Anastrepha fraterculus</i> (Diptera, Tephritidae)	45
Introdução.....	46
Material e Métodos.....	48
1. Extração do DNA e seqüenciamento.....	48
2. Análise da diversidade genética e do padrão evolutivo.....	50
3. Diversidade genética da região de <i>dsx</i> comum aos sexos e testes de seleção	53
Resultados.....	54
1. Seleção positiva em <i>fru</i>	54
2. Detecção de seleção natural em dados populacionais de baixa divergência em <i>fru</i>	59

2.1 Relação filogenética entre os haplótipos e diversidade genética no íntron e região conectora de <i>fru</i>	59
2.2 Espectro de frequência de sítios e teste de contingência	62
2.3 Detecção de seleção por máxima verossimilhança.....	64
2.4 Detecção de seleção por desvios em propriedades físico-químicas dos aminoácidos	64
Discussão	69
Padrão seletivo sobre <i>fru</i> usando dados de baixa divergência.....	711
Capítulo IV - Considerações finais.....	76
Considerações finais	77
Referências Bibliográficas.....	80
Referência Bibliográfica.....	81
Anexos	89

Capítulo I

Introdução Geral

Introdução

1. Revisão geral das espécies

O gênero *Anastrepha* pertence à família Tephritidae e à tribo Toxotrypanini da subfamília Trypetinae (Malavasi & Zucchi 2000). Este gênero é endêmico à região Neotropical com distribuição que vai desde parte da América do Norte até a América do Sul, à exceção do Chile e Sul da Argentina, e contém mais de 180 espécies (Steck 1991), das quais 94 ocorrem no Brasil (Zucchi 2000). Muitas espécies deste gênero são relativamente jovens e sugere-se que sua especiação foi rápida e recente (Morgante *et al.* 1980). Um conjunto de aproximadamente 29 espécies proximamente relacionadas forma o **grupo de espécies *fraterculus*** (Norrbon *et al.* 2000), o qual contém um complexo de espécies crípticas atualmente designado **complexo *A. fraterculus*** proper (Smith-Caldas *et al.* 2001).

O ciclo de vida básico é muito similar entre todas as espécies de *Anastrepha*. Na maioria das espécies, a fêmea deposita seus ovos na região do epicarpo ou mesocarpo da fruta. As larvas passam por três instar no interior da fruta, consumindo-a, depois saem do fruto para o solo onde se enterram para empupar (Aluja 1994; Malavasi & Zucchi 2000). Após um período de aproximadamente 14 dias os adultos eclodem das pupas e amadurecem sexualmente depois de aproximadamente 5 a 7 dias e após a corte a cópula e a oviposição o ciclo reinicia-se (Malavasi & Zucchi 2000). As espécies desse grupo devem ser consideradas oligotróficas, pois embora exista certa especificidade a plantas hospedeiras, em geral as espécies não são restritas a apenas um hospedeiro. *A. fraterculus* infesta cerca de 67 espécies de plantas, 24 sendo Myrtaceae, *A. sororcula* infesta cerca de 19 espécies, sendo 15 delas Myrtaceae, enquanto *A. obliqua* infesta 27 espécies, sendo 16 Myrtaceae e 8 Anacardiaceae (Zucchi 2000).

As espécies de *Anastrepha* são identificadas pelo uso de uma gama de marcadores morfológicos, incluindo o padrão de cor do corpo, principalmente de manchas presentes no subescutelo e/ou mediotergito, coloração e padrão de manchas na asa e na forma, denticção e tamanho das extremidades do acúleo, o qual é uma estrutura de oviposição presente apenas em fêmeas (Figura 1) (Malavasi & Zucchi 2000; Aluja & Norrbom 2000). Estas marcas variam entre algumas espécies e são utilizadas em conjunto com outros marcadores morfológicos para sua identificação, embora sejam utilizadas principalmente as variações no ápice do acúleo na identificação a nível de espécie (Norrbon *et al.* 2000). Por causa disso, a taxonomia de *Anastrepha* é baseada quase que exclusivamente em

fêmeas adultas (Aluja 1994). Das espécies deste grupo, *A. fraterculus* apresenta maior variação intraespecífica, sendo considerada um complexo de espécie crípticas (Steck 1991; Amaral 1994; Selivon 2000). Embora não exista um marcador único que permita diferenciar as espécies do grupo *fraterculus*, o conjunto das características acima mencionadas permite, na maioria dos casos, a sua correta identificação. Apesar disso, a alta plasticidade fenotípica destes marcadores, a qual leva à sobreposição de diversas medidas quantitativas, dificulta seu uso na separação de espécies muito próximas, principalmente as do grupo *fraterculus* (Malavasi & Zucchi 2000; Araujo & Zucchi 2006).

A análise conjunta dos dados de diferentes marcadores genéticos (Smith-Caldas *et al.* 2001; Ruiz *et al.* 2007a; Ruiz *et al.* 2007b) indicam um padrão complexo para as espécies do grupo *fraterculus* e confirmam a inabilidade das análises através do mtDNA e genes nucleares em distinguir diferentes espécies do complexo *fraterculus* e a necessidade de um estudo mais detalhado acerca do status das espécies.

2. Cascata de diferenciação sexual em insetos

Nos insetos a cascata de diferenciação é melhor compreendida no gênero *Drosophila*; o sinal inicial da cascata é fornecido pela razão entre os cromossomos sexuais X e os autossômicos. Nas fêmeas, por serem XX, esta razão é de 2:2, enquanto que nos machos a razão é de 1:2; a dose dupla de fatores de transcrição no cromossomo X ativa um promotor de *slx* (Schütt & Nöthiger 2000). O *slx* possui dois éxons iniciais que são eliminados durante o processamento de seu mRNA. Um destes éxons possui um códon de parada que precisa ser eliminado para que ocorra a produção de uma proteína funcional. Uma vez iniciada a produção da proteína ativa, estabelece-se um ciclo de auto-regulação, no qual a própria proteína SLX coordena o processamento do mRNA de *slx* eliminando o éxon com o códon de parada. Em machos o processamento alternativo com eliminação do éxon com códon de parada nunca ocorre, assim, eles não produzem a proteína funcional e consequentemente não estabelecem o ciclo de auto-regulação (Pomiankowski & Wilkins A. 2004).

Além do mecanismo de auto-regulação, a proteína SLX também coordena o processamento alternativo dos transcritos do gene *transformer (tra)*, cujo produto atua logo abaixo na cascata de diferenciação sexual. Da mesma forma que em *sxl*, *tra* produz transcritos que contêm códons de parada logo no início de sua sequência. Em fêmeas, a SXL elimina o segmento que contém o códon de parada do transcrito de *tra* levando à produção de sua forma ativa TRA. Esta forma ativa feminina juntamente com uma forma constitutiva TRA2 atua no processamento alternativo dos mRNAs de dois genes do final da cascata de diferenciação sexual: *doublesex (dsx)* e *fruitless (fru)*

(Coschigano & Wensink 1993; Pomiankowski & Wilkins A. 2004). Em fêmeas, o heterodímero TRA/TRA2 promoverá o processamento alternativo para a forma feminina ativa de DSX^F , que por sua vez estimulará ou reprimirá a expressão de genes responsáveis pelo desenvolvimento de características sexuais na direção do desenvolvimento de traços feminino. Em machos, na ausência do complexo TRA/TRA2 o transcrito de *dsx* será processado de acordo com o *default* da cascata de diferenciação levando à formação do produto masculino ativo DSX^M , o qual promoverá o desenvolvimento de características masculinas (Cline & Meyer 1996; Pomiankowski & Wilkins A. 2004). Da mesma forma, o heterodímero TRA/TRA2 presente em fêmeas processará os transcritos de *fru*. Entretanto, neste caso, o processamento para a forma feminina leva à tradução de uma proteína inativa, visto que em seu éxon inicial específico de fêmea há um códon de parada. Novamente a ausência de TRA/TRA2 leva à produção da forma ativa de *fru*, FRU^M , pelo processamento alternativo que elimina o éxon específico feminino que contém o códon de parada. Uma vez formado, FRU^M controlará o desenvolvimento do comportamento de corte masculino (Demir & Dickson 2005). Wilkins (1995) propôs que as vias de determinação sexual evoluíram numa ordem reversa, iniciando a partir de seus elementos basais e incorporando novos elementos nas categorias superiores. Corroborando esta hipótese, a cascata de diferenciação parece ser conservada apenas em seus elementos finais, *tra*, *dsx* e *fru*, uma vez que *sxl* funciona como iniciador apenas em *Drosophila* e em mais nenhuma linhagem de insetos (Hediger *et al.* 2004; Burghardt 2005; Cho *et al.* 2007).

2.1. *doublesex*: função e organização genômica

O produto do gene *dsx* é um fator de transcrição responsável pela ativação ou repressão de genes responsáveis pelo desenvolvimento de características masculinas e femininas. Suas formas masculinas e femininas atuam de forma antagônica. Se em machos DSX^M estimula a expressão de um determinado gene, em fêmeas o mesmo gene é reprimido por DSX^F (Coschigano & Wensink 1993; Cho & Wensink 1997; Chen *et al.* 2008) Como exemplo temos a regulação do gene *yolk*, que é responsável pela formação do vitelo. Em fêmeas a expressão deste gene é estimulada pela forma feminina de DSX , ao passo que em machos sua expressão é reprimida (Chen *et al.* 2008).

A organização genômica de *dsx* é variável na linhagem dos insetos (Suzuki *et al.* 2001; Ruiz *et al.* 2005; Lagos *et al.* 2005), porém uma organização geral pode ser delimitada. Ele é composto por quatro elementos principais: (1) um domínio de ligação ao DNA (DM), o qual foi descrito como uma nova classe de motivos dedo de zinco (Erdman & Burtis 1993; Zhu *et al.* 2000), (2) um domínio de oligomerização (OD2), (3) um segmento específico de fêmea e (4) um segmento específico de

macho (Suzuki *et al.* 2001; Ruiz *et al.* 2005; Lagos *et al.* 2005). Em *Anastrepha*, o gene está dividido em 4 éxons que se organizam na seguinte ordem partindo da extremidade 5' para a 3': O primeiro éxon é comum a ambos os sexos e em sua extremidade 5' encontra-se o DM; o segundo éxon também é comum aos sexos e codifica o OD2; parte desse domínio está presente no éxon seguinte, o qual é específico a fêmeas e por último o quarto éxon codifica a região específica de macho (Ruiz *et al.* 2005). Os domínios DM e OD2 são bastante conservados na linhagem dos insetos, porém a região que os conecta é bastante variável (Kuhn *et al.* 2000; Ruiz *et al.* 2005).

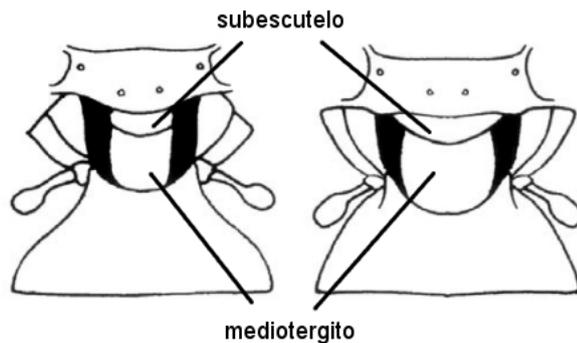
2.2. *fruitless*: função e organização genômica

Da mesma forma que em *dsx*, o mRNA de *fru* é processado alternativamente produzindo proteínas específicas para cada sexo, sendo que apenas a masculina é ativa (Demir & Dickson 2005). Este é um gene funcionalmente complexo, sendo responsável principalmente pela formação de grupos neuronais que estabelecerão o comportamento de corte masculino. Além disso, seu produto coordena a conectividade de neurônios a estruturas importantes à fertilidade, pela inervação de órgãos reprodutivos (Lee *et al.* 2001), ao reconhecimento do sexo do outro indivíduo, através da inervação de receptores de feromônios (Datta *et al.* 2008) e à realização da cópula, pela formação da junção neuromuscular responsável pelo desenvolvimento do Músculo de Lawrence presente apenas em machos (Ito *et al.* 1996; Billeter *et al.* 2006).

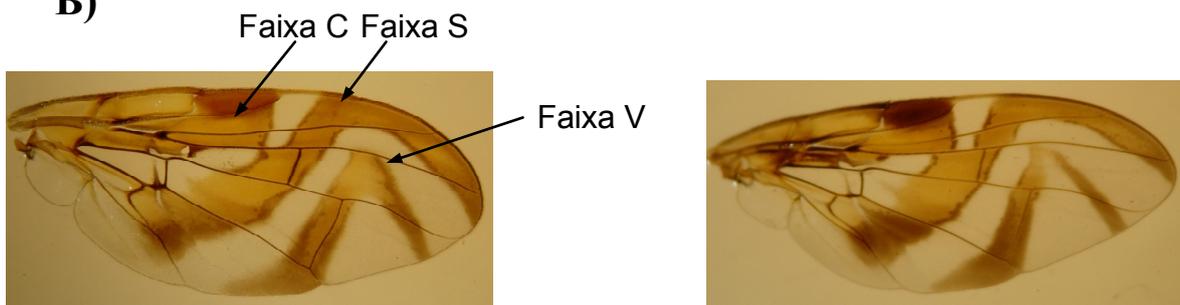
Entre os genes da cascata de diferenciação sexual, o *fru* é o que apresenta uma organização genômica e um padrão de processamento de mRNA mais complexos. Como no caso do *dsx*, pode-se estabelecer um padrão geral de organização. O *fru* pode ser dividido em quatro regiões principais: (1) um domínio específico ao sexo, (2) um domínio de dimerização (BTB), (3) uma região conectora ligando o BTB ao próximo domínio e (4) um domínio de ligação a DNA (*Zinc-finger*) (Demir & Dickson 2005). O primeiro domínio é composto por éxons específicos a cada um dos sexos. Após o processamento do transcrito, somente a forma masculina é ativa, visto que o mRNA para FRU^F mantém o éxon feminino que contém um códon de parada, o qual interrompe a tradução completa da proteína. O BTB é formado por dois éxons que se unem no processamento do mRNA, da mesma forma, a região conectora é formada pela união de três ou quatro éxons, dependendo do grupo taxonômico. O domínio de ligação ao DNA na extremidade 3' é composto por três ou quatro éxons (Demir & Dickson 2005; Gailey *et al.* 2006). Estes éxons são processados de forma independente e cada um produz uma isoforma específica de *fru* (Billeter *et al.* 2006). A presença de quatro promotores diferentes, um na extremidade 3' anterior aos éxons específicos a cada sexo e outros três a jusante

destes éxons, permite um padrão complexo de processamento do mRNA, o qual promove a expressão tanto de formas específicas de cada sexo pelo primeiro promotor, quanto daquelas formas independentes do sexo pelos outros três promotores restantes. Assim, apesar das formas sexuais de fêmea serem inativas, suas formas não relacionadas ao sexo não o são, e desempenham um papel importante no desenvolvimento do sistema nervoso central destas moscas (Song *et al.* 2002; Demir & Dickson 2005).

A)



B)



C)

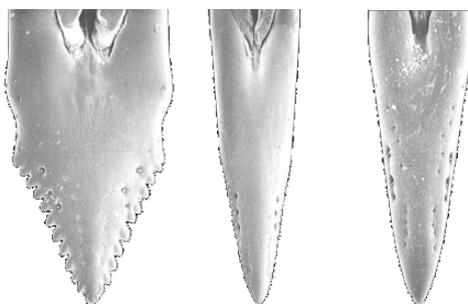


Figura 1. Características morfológicas do tórax (A), asa (B) e extremidade do acúleo (ovipositor) (C) utilizadas na identificação de espécies de *Anastrepha*. B e C modificado de (Aluja & Norrbom 2000).

3. Busca de marcadores para o estudo da diferenciação do complexo *fraterculus*

A ênfase deste estudo será no complexo de espécies *fraterculus*, mais precisamente nas espécies *A. fraterculus*, *A. obliqua* e *A. sororcula* que são das principais pragas de frutos no país (Solferini e Morgante, 1987) e espécies mais comumente encontradas neste complexo. Considerando-se a dificuldade em se encontrar marcadores que nos permitam identificar tais espécies, buscamos um entendimento maior das forças evolutivas envolvidas na diferenciação deste complexo. Para tanto, neste trabalho estudamos genes envolvidos com a determinação sexual de forma a avaliar inicialmente seu eventual papel na diferenciação deste complexo, bem como investigar mais detalhadamente a dinâmica das forças evolutivas envolvidas com os padrões de evolução molecular destes genes. Com este intuito, avaliamos também a capacidade das metodologias atuais de estudo de seleção positiva, como os softwares PAML (Yang 2007) e o TreeSAAP (Woolley *et al.* 2003) de detectar tais padrões em dados populacionais que apresentam uma baixa divergência e que sofrem com a estocasticidade evolutiva que limita a interpretação de metodologias baseadas no estudo das razões das taxas dN/dS (Kryazhimskiy & Plotkin 2008).

Capítulo II

Evolução molecular do gene *doublesex* no grupo *Anastrepha fraterculus* (Diptera, Tephritidae)

Iderval S. Sobrinho-Jr. e Reinaldo A. de Brito

Introdução

Diversos genes relacionados a reprodução em animais e plantas estão entre os mais divergentes, evoluindo mais rapidamente do que outros não relacionados a características sexuais (Clark *et al.* 2006), possivelmente em virtude da seleção e/ou conflito sexual (Jagadeeshan & Singh 2005; Clark *et al.* 2007). Em mamíferos, por exemplo, diversas proteínas relacionadas a fertilização foram identificadas como proteínas que evoluem rapidamente (Swanson *et al.* 2003). Da mesma forma, genes de auto-incompatibilidade em plantas (Castric & Vekemans 2007), proteínas de reconhecimento de gametas em gastrópodos marinhos (Galindo *et al.* 2003) e proteínas reprodutivas masculinas e femininas de *Drosophila* (Swanson *et al.* 2001; Panhuis & Swanson 2006) têm mostrado o mesmo padrão de diversificação rápida. A maioria dos estudos moleculares têm se concentrado recentemente naqueles genes envolvidos em fertilização ou interação entre macho e fêmea e apenas poucos nos genes responsáveis pela diferenciação sexual. O gene *transformer* (*tra*), um dos genes principais da cascata de diferenciação sexual em Diptera (McAllister & McVean 2000; Kulathinal *et al.* 2003), e o gene *complementary sex determiner* (*csd*) da cascata de diferenciação da abelha (Cho *et al.* 2006) são alguns dos exemplos de genes de determinação sexual cuja evolução molecular foi estudada.

Um outro gene que desempenha um papel central na cascata de determinação sexual é o *doublesex* (*dsx*). *dsx* é um fator de transcrição que age ativando ou reprimindo genes na cascata que são responsáveis por caracteres masculinos ou femininos em insetos (Zhu *et al.* 2000). Este gene é funcionalmente conservado entre insetos e outros animais menos relacionados e, diferentemente de outros genes da cascata de determinação sexual (Saccone *et al.* 1998; Lagos *et al.* 2005); (Schütt & Nöthiger 2000), ocupa uma posição conservada no fim da cascata (Hediger *et al.* 2004; Cho *et al.* 2007). Apesar de sua conservação funcional, *dsx* varia tanto em sua organização genômica quanto em seu padrão de *splicing* entre insetos (Suzuki *et al.* 2001; Scali *et al.* 2005; Lagos *et al.* 2005). Em *Anastrepha*, *dsx* é composto de quatro éxons: os dois primeiros são comuns a ambos os sexos e contêm o domínio de ligação ao DNA próximo à região 5' e o domínio de oligomerização na região 3' do éxon 2. O terceiro éxon é expresso exclusivamente em fêmeas e o quarto é específico de machos (Ruiz *et al.* 2005). O domínio de ligação ao DNA foi descrito como uma nova classe de motivos DM dedos de zinco (*zinc-fingers*), os quais são estruturalmente conservados entre metazoários (Zhu *et al.* 2000).

Diferentes estratégias têm sido usadas para se investigar a evolução adaptativa molecular. Os métodos mais gerais são baseados nas expectativas neutras do espectro de frequências de haplótipos e fornecem uma idéia geral das forças seletivas agindo sobre a região de DNA usando

parâmetros populacionais (Tajima 1989a; McDonald & Kreitman 1991; Fu & Li 1993). Contudo, tais métodos são sensíveis a mudanças demográficas e sua interpretação é raramente inequívoca (Tajima 1989a; Yang 2002). Por isto, métodos não-enviesados, que consideram a razão entre mudanças não-sinônimas e sinônimas (ω) para se inferir padrões de seleção, têm sido comumente usados para se investigarem mudanças positivamente selecionadas (Galindo *et al.* 2003; Clark & Swanson 2005; Berlin & Smith 2005; Castric & Vekemans 2007). Apesar de sua especificidade para se detectar mudanças adaptativas, estratégias de ω apenas levam em consideração a substituição nucleotídica em um códon, e não consideram o efeito de substituições de aminoácidos nas propriedades físico-químicas das proteínas. Porque a evolução adaptativa está diretamente relacionada a modificações na estrutura ou função das proteínas, mesmo se o sítio apresentar $\omega > 1$, mudanças no códon poderiam ainda provocar alterações estruturalmente conservadas se tais mudanças não interferirem com a estrutura ou função protéica. Isso poderia ser o caso se mudanças de aminoácido fossem quimicamente similares. Por outro lado, a evolução adaptativa poderia ser observada mesmo em códons com $\omega < 1$ se tais mudanças promoverem alterações para a aminoácidos com propriedades físico-químicas radicalmente diferentes (McClellan *et al.* 2005).

Se considerarmos apenas a razão da taxa ω , podemos perder sinais importantes de evolução adaptativa, particularmente em genes que têm extensas parcelas das seqüências conservadas e sinais muito recentes de seleção positiva (Nickel *et al.* 2008). Em virtude da conservação estrutural e funcional do motivo DM e do papel central que *dsx* desempenha na diferenciação sexual, tal gene deve ter uma influência direta no valor adaptativo e deveria ser em geral razoavelmente conservado. Por outro lado, *dsx* também está envolvido com o comportamento sexual (Rideout *et al.* 2007) e pode estar sujeito a seleção positiva. Estudos anteriores feitos com este gene em *Anastrepha* indicaram que este gene estaria evoluindo apenas de acordo com a seleção purificadora (Ruiz *et al.* 2007b). Contudo, tal estudo investigou apenas um casal de espécimes por espécie e não realizou uma análise extensa dos padrões de seleção. Neste trabalho, investigamos dados populacionais para diversos espécimes do grupo *fraterculus*, com o intuito de investigar padrões de adaptação molecular na região comum do *dsx* e testar a hipótese deste gene ser evolutivamente conservado e evoluir na região sob seleção purificadora, usando de diversas estratégias, tanto baseadas na utilização de dados de alta divergência quanto de dados de baixa divergência (dados populacionais).

Material e Métodos

Frutos de diversas espécies de planta infestadas por *Anastrepha* foram coletadas de 32 localidades no Brasil (Tabela 1) e dispostas em vermiculita por 14 dias quando pupas eram separadas. Após a emergência em maturação, as moscas eram identificadas e, se pertencentes ao grupo *A. fraterculus*, na maioria das vezes *A. fraterculus*, *A. obliqua* e *A. sororcula*, eram separadas e imediatamente processadas ou preservadas em etanol 95% até a extração do DNA.

1. Extração do DNA e seqüenciamento

A extração do DNA foi feita usando-se um protocolo modificado de Chomeczynski & Sacchi (1987) para se garantir a manutenção do exo-esqueleto intacto para análises morfológicas futuras. Amplificamos uma região próxima ao início do gene *doublesex* no éxon comum a ambos os sexos usando primers degenerados criados de seqüências homólogas de espécies proximamente relacionadas (5'– ATGGTTTCNGAGGATAATTG –3' and 5'– GCGNCCNACNACYGANATNGGCAA –3'). Esta região foi amplificada por PCR a partir do DNA genômico em um termociclador PTC-200 (BioRad) usando uma mistura de *Taq* e *Pfu* polimerase para se reduzirem erros de incorporação (Cline *et al.* 1996). Os produtos de PCR foram purificados por precipitação com PEG 8000 (Lis & Schleif 1975) e clonados usando o kit InsTAclone (Fermentas). Ao menos duas colônias recombinantes foram seqüenciadas usando-se primers M13F (*forward*) e M13R (*reverse*) com o kit DYEnamic™ ET dye terminator (GE Healthcare) e corridos em um MegaBace 1000 (GE Healthcare). A qualidade do seqüenciamento foi inspecionada visualmente no software Chromas ver. 2.31 (<http://www.technelysium.com.au>). Quando seqüências iguais ou muito similares (diferindo em uma ou duas bases) foram obtidas, seqüências adicionais foram realizadas em colônias recombinantes diferentes (até o total de 5) para se confirmar a eventual homozigosidade dos indivíduos.

2. Análise da evolução molecular de *doublesex*

A análise de seleção na região amplificada do gene *dsx* foi feita considerando uma estratégia hierárquica. Inicialmente avaliamos este padrão de evolução em um contexto macroevolutivo, contrastando apenas três seqüências de *Anastrepha* (uma de cada das espécies aqui

estudadas) com outras seqüências de Muscomorpha disponíveis no GenBank. Em uma segunda instância, as análises foram feitas considerando apenas o conjunto total de seqüências obtidas para *Anastrepha*. Como os testes de detecção de seleção positiva, como o PAML, são contingentes de um ramo evolutivo longo para que tenhamos um número razoável de mutações de forma a prover graus de liberdade suficientes para lidar adequadamente com a estocasticidade evolutiva (Kryazhimskiy & Plotkin 2008), a análise macroevolutiva buscou inicialmente estabelecer se de fato podemos detectar seleção positiva para esta região quando consideramos longos ramos evolutivos. Em um segundo instante, fizemos uso destas mesmas metodologias considerando dessa vez nossos dados populacionais, de forma a verificar a adequação e capacidade dos métodos previamente mencionados em analisar tais dados.

Utilizamos para a análise macroevolutiva as seguintes espécies para as quais o gene *dsx* estava disponível no GenBank (<http://www.ncbi.nlm.nih.gov>) ou FlyBase (<http://flybase.org>): *Bactrocera tryoni*, *B. dorsalis*, *B. oleae*, *Drosophila pseudoobscura*, *D. melanogaster*, *D. silvestris*, *D. heteroneura*, *Ceratitidis capitata*, *D. ananassae*, *D. mojavensis*, *D. persimilis* (número de acesso no GenBank AF040077.1, AY669317.1, AJ715415.1, XM_002137347.1, AE014297.2, AF051672.1, AF051668.1, AF124047.1, GF16562, GI10222, GI23549, respectivamente) e *D. yakuba*, *D. virilis*, *D. grimshawi*, *D. simulans*, *D. sechellia*, e *D. erecta* (FBtr0271707, FBtr0226750, FBtr0154631, FBtr0220077, FBtr0201684, respectivamente, no FlyBase). Todas as seqüências foram alinhadas usando o software Clustal a partir do alinhamento de aminoácidos no software BioEdit Sequence Alignment Editor (Hall 1999) e visualmente inspecionadas. Utilizamos para a análise populacional, além das seqüências aqui obtidas, também seqüências de outros espécimes de machos e fêmeas de *Anastrepha* disponíveis no GenBank: *A. serpentina*, *A. bistrigata*, *A. grandis*, *A. striata*, *A. fraterculus sp1*, *fraterculus sp2*, *fraterculus sp3*, *fraterculus sp4*, *A. obliqua* e *A. sororcula* (números de acesso no GenBank: DQ494329, DQ494338, DQ494332, DQ494341, DQ494328, DQ494337, DQ494331, DQ494340, DQ494344, DQ494334, DQ494325, DQ494335, DQ494326, DQ494336, DQ494327, DQ494343, AY948420, AY948421, DQ494330 and DQ494339, respectivamente). Estas seqüências foram alinhadas usando o software Clustal no software BioEdit Sequence Alignment Editor (Hall 1999) e visualmente inspecionadas.

Como a recombinação interfere tanto com as inferências filogenéticas quanto com a capacidade de detecção de seleção, realizamos três métodos diferentes para se detectar eventos de recombinação em nossos dados populacionais: o teste *Maximum Chisquare* de Maynard Smith (1992), realizado no programa MaxChi, e os testes GENECONV (Sawyer 1989) e RDP, implementados no programa RDP ver. 3b14 (Martin & Rybicki 2000). Um limite de $p < 0,05$ foi estabelecido para todos os três testes e, quando aplicável, foi realizada uma aleatorização de 1000 réplicas.

As razões das taxas de substituição sinônimas/não-sinônimas (dN/dS, também denominado ω) foram medidas no PAML para se inferir as pressões seletivas na proteína. Uma razão de taxas $\omega > 1$ em um sítio específico indicaria seleção positiva, porque as substituições não-sinônimas teriam taxas de fixação mais altas do que mutações sinônimas devido a vantagens seletivas. Por outro lado, um $\omega < 1$ indicaria seleção purificadora, causada por restrições funcionais no códon. Os dados macroevolutivos foram analisados através do *Relaxed Branch-site test* e *Strict Branch-site test* (Yang & Nielsen 2002; Zhang *et al.* 2005). Para tanto, o ramo das *Anastrepha* foi denominado de ramo *foreground*, no qual seria testada a ocorrência de seleção positiva, e o restante dos ramos foram denominados de *background*. No primeiro teste contrasta-se o modelo M1a (modelo nulo) com o modelo MA (modelo alternativo) através de um teste de razão de verossimilhanças (LRT - *likelihood ratio test*). Neste contraste, o modelo nulo admite a mesma taxa evolutiva com todos os sítios variando seus valores de ω entre 0 e 1. Por outro lado, no segundo teste o modelo MA é restrito de forma que a categoria que assume $\omega > 0$ seja fixada a 1, e é contrastado contra o modelo MA.

Para os dados populacionais o núcleo central da análise é o cálculo dos valores de ω usando a máxima verossimilhança sob modelos diferentes de substituições de códon, denominadas M1a, M2a, M7, M8 (Yang *et al.* 2005) e M8a (Swanson *et al.* 2003; Wong *et al.* 2004), as quais definem premissas diferentes para as restrições evolutivas dos códons. A seleção positiva é estimada por LRT entre modelos que permitem $\omega > 1$ (M2a e M8) e outros que não o fazem (M1a, M7 e M8a). Realizamos LRTs contrastando os seguintes pares de modelos: M1a-M2a, M7-M8 e M8-M8a. A assinatura de seleção positiva em códons individuais foi estimada por análises Bayesianas em que permitiu-se que a razão ω variasse entre sítios. O método de *Bayes Empirical Bayes* (Yang *et al.* 2005) foi utilizado para se estimar quais sítios estavam sob a influência de seleção positiva. Porque alguns modelos podem apresentar problemas de falta de convergência nas estimativas de verossimilhança, realizamos as análises duas vezes com valores iniciais diferentes de ω .

3. Análises populacionais de seleção

Além das análises realizadas com os dados macroevolutivos coletados, realizamos algumas estratégias com dados populacionais buscando detectar sinais de seleção positiva e contrastar tais resultados com a metodologia previamente mencionada.

3.1 Comparação de taxas evolutivas entre *dsx* e COI

Com o intuito de se determinar a diferença nas taxas de evolução sinônimas e não-sinônimas para *dsx* e da porção do início do gene mitocondrial citocromo oxidase I (estudada para o mesmo grupo de espécies em diversos dos mesmos indivíduos (Smith-Caldas *et al.* 2001); de Brito, R.A. e Matioli, S.R., dados não-publicados), contrastamos suas taxas usando dois métodos independentes.

O primeiro método baseia-se na metodologia de Yang & Swanson (2002) na qual os dados são unidos e definidos como partições diferentes. Uma série de modelos é ajustada aos dados, sendo o teste de hipótese baseado no contraste entre cada modelo. O modelo mais simples, A, assume que ambas partições formam um único conjunto de dados e estima para elas os mesmos valores de ω , taxas de transição/transversão (κ), frequência de sítios (π) e comprimentos de ramos (r), o qual indica a taxa de substituição da partição. O último e mais complexo modelo (E) assume que os dois conjuntos de dados formam partições completamente independentes, cada uma com sua estimativa de parâmetros particular. Entre estes dois modelos, temos aqueles intermediários, como por exemplo, o modelo B, o qual assume que as partições possuem comprimentos de ramos diferentes e proporcionais, mas compartilham os mesmo valores de ω , κ e π . Um outro modelo intermediário usado nos contrastes foi o modelo C, que estima as mesmas taxas ω e κ para as duas partições e determina independentes valores de r e π para as mesmas. Dessa forma, o contraste entre o modelo A (nulo) contra o modelo E (alternativo) testa a hipótese de homogeneidade de todas as taxas entre as partições, no caso COI e *dsx*. O contraste entre A *versus* B testa a hipótese de homogeneidade nas taxas de substituição e por último, C *versus* E testa a hipótese de homogeneidade apenas nos parâmetros ω e κ entre as partições. Estes modelos são aninhados e o grau de liberdade é obtido pela diferença entre o número de parâmetros de cada um. No caso dos contrastes descritos acima, o valores para A x E, A x B e C x E, são respectivamente, 12, 1 e 2.

Visto que a metodologia de Yang & Swanson (2002) não permite a separação das estimativas de ω dos κ , as razões ω para esses dois genes foram estatisticamente comparadas de acordo com o método de Lu & Rausher (2003). A significância dessas diferenças é obtida pelo estabelecimento de um intervalo de confiança para os valores de ω estimados pelo modelo M0 a partir do CODEML implementado no PAML ver. 4 (Yang 2007). O intervalo de confiança calculado por um teste LRT do *log-likelihood* calculado para um modelo não-restrito M0, em que o ω pode variar livremente, e o *log-likelihood* para um modelo restrito M0, que assume um valor de ω arbitrário. Os valores arbitrários variam dos valores não-restritos de ω de um gene até os valores não-restritos do

outro gene. Os cálculos continuam até que o LRT indique uma diferença significativa entre os modelos testados (restritos e não-restritos). Os modelos M0 restritos e não-restritos são aninhados e têm 1 grau de liberdade para o LRT. Os valores de ω são considerados estatisticamente diferentes quando os dois intervalos de confiança nos genes diferentes não se sobrepõem.

3.2 Diversidade genética da região de *dsx* comum aos sexos

Índices de diversidade geral, tais como a diversidade haplotípica (H_d) e nucleotídica (π) (Nei 1987), o número de sítios polimórficos, as diversidades nucleotídicas sinônima (π_s) e não-sinônima (π_a) foram calculadas em nossos dados populacionais usando o programa DnaSP ver. 4.10.9 (Rozas *et al.* 2003). Redes de haplótipos foram inferidas no programa TCS 2.1 (Clement *et al.* 2000) e foram manualmente convertidas a formatos de árvore Newick para algumas das análises de neutralidade. Os testes de neutralidade D de Tajima (Tajima 1989b) e D e F de Fu e Li (Fu & Li 1993) foram realizados no DnaSP usando seqüências de *Bactrocera dorsalis* como grupo externo, por ser a espécie mais próxima de *Anastrepha* cujo *dsx* se apresentava completo. Para a análise de dados populacionais, buscamos testar se mutações que levam a alterações de aminoácidos têm maior ou menor probabilidade de deixar descendentes na população. Se esta região está evoluindo neutramente, não esperamos encontrar diferenças entre tais probabilidades quando comparamos mutações sinônimas e não-sinônimas. Caso a região esteja sujeita à seleção purificadora, esperamos que a maioria das mutações não-sinônimas deixe menos descendentes do que mutações sinônimas (Yang 2002). Por outro lado, caso esta região esteja sujeita à seleção positiva, podemos encontrar mutações vantajosas que estejam em alta freqüência, e com isso, podemos ter que tais mutações deixariam mais descendentes do que outras mutações (Fay & Wu 2005). Realizamos dois tipos de testes estatísticos para tentar detectar tais sinais, em que consideramos apenas mutações que estavam presentes pelo menos em dois ou mais haplótipos - uma vez que estamos buscando contrastar probabilidades diferenciadas de sobrevivência, devemos considerar apenas mutações que já sobreviveram tempo suficiente para se reproduzir. Mutações que estão presentes apenas em *singletons* não passaram ainda pelo teste evolutivo da sobrevivência e reprodução no tempo (Templeton 1996). Um dos testes estatísticos fez uso de testes exatos de Fisher em tabelas de contingência 2 x 2 (Templeton 1996), que é uma versão mais refinada do teste de McDonald & Kreitman (1991), uma vez que avalia também polimorfismo intraspecíficos e não apenas faz um contraste entre mutações polimórficas e fixadas entre linhagens. Estas tabelas foram feitas para se contrastar a distribuição de substituições sinônimas e não-sinônimas em haplótipos encontrados nas pontas ou no interior (haplótipos jovens ou antigos).

Contudo, consideramos isso de forma hierárquica, contrastando classes cada vez maiores de mutações, por exemplo, mutações que deixaram dois, três, quatro descendentes e assim sucessivamente. Alternativamente, realizamos o teste não-paramétrico de Mann-Whitmann, que neste caso testa se mutações sinônimas e não-sinônimas apresentam um número distinto de descendentes em nossas amostras através de uma análise de ranqueamento.

Resultados

1. Diversidade em *doublesex* e análise filogenética

Uma região de 543 pb do *doublesex*, incluindo o domínio de ligação ao DNA e uma grande parte do primeiro éxon, foi clonada e sequenciada para 51 indivíduos de diversas localidades no Brasil (Tabela A1) que, combinadas a outras seqüências disponíveis no GenBank (Ruiz *et al.* 2007b), revelaram 98 haplótipos diferentes, a grande maioria composta de *singletons*. Apenas 6 haplótipos foram encontrados em mais de uma cópia, três dos quais compartilhados entre espécies diferentes, incluindo o haplótipo mais comum, 1, o qual foi encontrado nas três espécies do grupo aqui estudadas, *A. fraterculus*, *A. obliqua* e *A. sororcula*.

Os três testes para recombinação diferentes não detectaram recombinação entre as seqüências aqui estudadas. A rede de haplótipos inferida no TCS (Figura A1) não é significativamente diferente de uma árvore filogenética de máxima verossimilhança inferida no PHYLIP ver 3.67 (dados não apresentados). Esta rede mostra que, embora a maior parte da variação seja específica para cada espécie, o que é trivial considerando-se que quase todos os haplótipos são *singletons*, não há nenhuma mutação que seja fixada para qualquer espécie aqui estudada. Estimativas de diversidades haplotípicas e nucleotídicas, e substituições sinônimas e não-sinônimas são similares independentemente se *A. fraterculus*, *A. obliqua* e *A. sororcula* são consideradas em separado ou como um único grupo em conjunto com outras espécies do grupo *fraterculus* (Tabela A2).

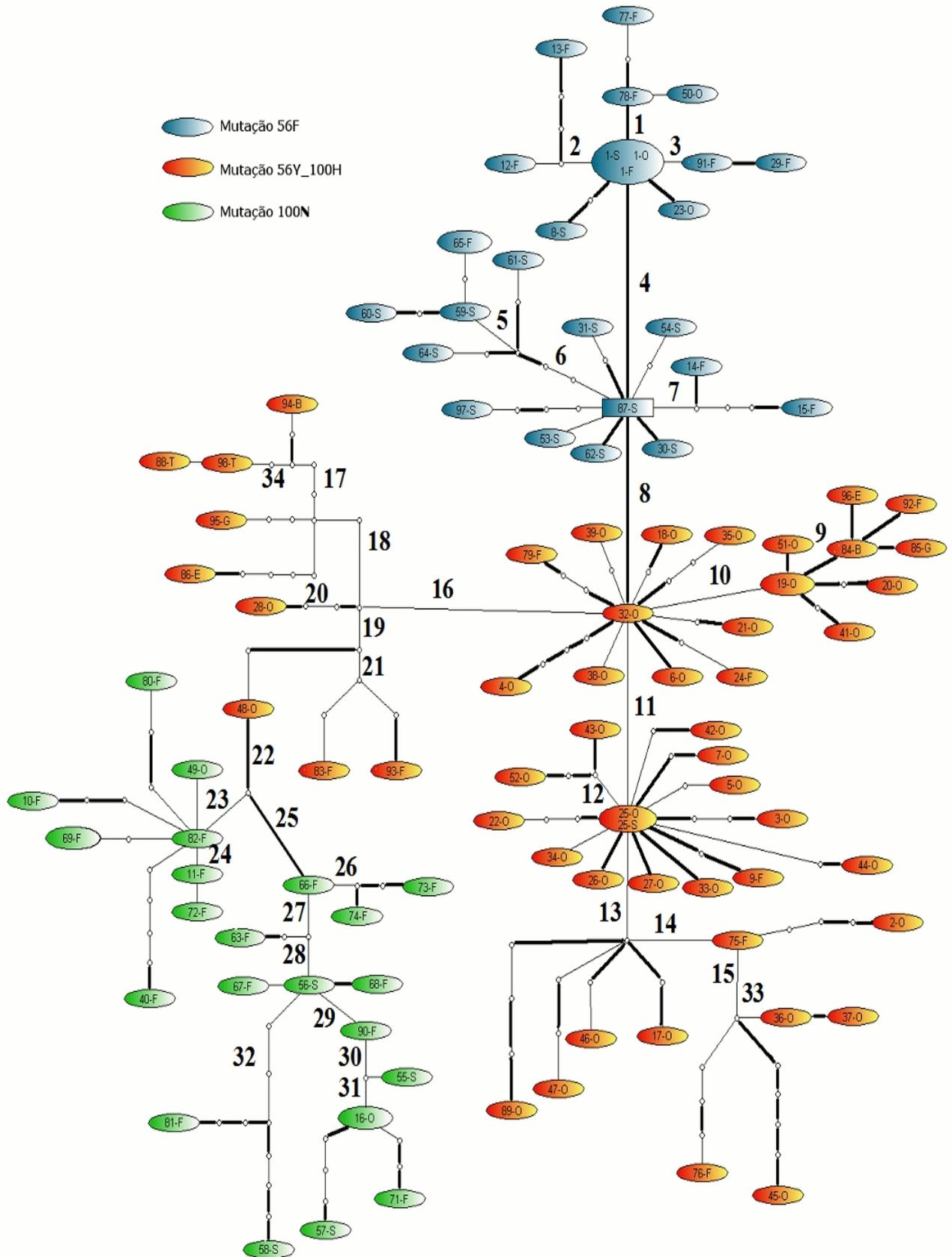


Figura A1. Rede haplotípica separada por duas mutações não-sinônimas que levam a alterações radicais nas propriedades físico-químicas dos aminoácidos (8 e 22). Os números sobre os ramos indicam as mutações sinônimas e não-sinônimas internas.

Tabela A1. Locais amostrados com informação geográfica, haplotípica e de espécies.

código	Localidade	Latitude (S)	Longitude (W)	Haplótipos
01	São Carlos - SP	22° 01' 03"	47° 53' 27"	9-F
02	Vitória - ES	20° 19' 10"	40° 20' 16"	2-O, 3-O, 25-O, 26-O, 45-O
03	Feira do Santana -BA	12° 16' 00"	38° 58' 00"	4-O, 5-O
04	Gurupi - TO	11° 43' 45"	49° 04' 07"	6-O, 7-O, 41-O, 42-O
05	São Sebastião – SP (Santiago)	23° 45' 36"	45° 24' 35"	8-S, 1-S, 77-F, 78-F, 82-F, 92-F, 86-E, 96-E
06	Bertioga - SP	23° 51' 16"	46° 08' 19"	10-F, 11-F
07	Conceição do Almeida - BA	12° 46' 46"	39° 10' 12"	12-F, 13-F, 14-F, 15-F
08	Araguaína – TO	07° 11' 28"	48° 12' 26"	16-O, 17-O, 18-O, 19-O, 20-O, 21-O, 22-O
09	Santo Amaro – BA	12° 32' 48"	38° 42' 43"	1-O, 23-O, 24-F
10	Goiânia - GO	16° 40' 43"	49° 15' 14"	27-O, 28-O
11	Bonito – PE	08° 28' 13"	35° 43' 43"	29-F, 1-F, 30-S, 31-S
12	Redenção – PA	08° 01' 43"	50° 01' 53"	32-S, 33-O, 34-O, 35-O
13	Babaçulândia – TO	07° 12' 17"	47° 45' 25"	36-O, 37-O
14	Pirenópolis – GO	15° 51' 09"	48° 57' 33"	38-O, 39-O, 40-F
15	Três Lagoas – MS	20° 45' 04"	51° 40' 42"	43-O, 44-O
16	Belém -PA	01° 27' 21"	48° 30' 16"	46-O, 47-O
17	Porto Franco - MA	06° 20' 18"	47° 23' 57"	48-O, 49-O
18	Belo Horizonte - MG	19° 55' 15"	43° 56' 16"	50-O
19	Bela Vista de Goiás - GO	16° 58' 22"	48° 57' 12"	51-O
20	Linhares – ES	19° 23' 28"	40° 04' 20"	52-O, 53-S, 54-S
21	Piracicaba – SP	22° 43' 31"	47° 38' 57"	55-S, 56-S, 57-S, 58-S
22	Nova Souré – BA	11° 14' 00"	38° 29' 00"	59-S, 60-S
23	Bauru – SP	22° 18' 53"	49° 03' 38"	25-S, 61-S, 62-S, 63-F
24	Porto Seguro – BA	16° 26' 59"	39° 03' 53"	64-S
25	Boracéia – SP	22° 11' 35"	48° 46' 44"	1-F, 65-F
26	Bariri – SP	22° 04' 28"	48° 44' 25"	66-F, 65-F
27	Itabira – MG	19° 37' 09"	43° 13' 37"	67-F, 68-F
28	Moji das Cruzes – SP	23° 31' 22"	46° 11' 18"	69-F
29	Vacaria – RS	28° 30' 44"	50° 56' 02"	70-F, 71-F, 72-F
30	Vargem Alta – ES	20° 40' 17"	41° 00' 25"	73-F, 74-F, 75-F, 76-F
31	Natal – RN	05° 47' 42"	35° 12' 34"	79-F
32	Caçador - SC	26° 46' 31"	51° 00' 54"	80-F
33	Santa Isabel - SP	23°18' 0"	46°13' 0"	81-F, 90-F, 1-F, 91-F
34	Guayaquil - Ecuador	00°02' 0"	80°01' 0"	83-F, 93-F
35	Parati - RJ	23°10' 0"	44°12' 0"	87-S, 97-S
36	Indaiatuba - SP	22°42' 0"	47°37' 0"	19-O, 89-O
37	Laranjal Paulista - SP	23°05' 0"	47°45' 0"	85-G, 95-G
38	Ribeirão Preto - SP	21°05' 0"	47°50' 0"	84-B, 94-B
39	Uberlândia - MG	18°56' 0"	48°13' 0"	88-T, 98-T

O número nos haplótipos representa sua identificação e a letra corresponde às espécies nas quais o haplótipo foi encontrado. O - *A. obliqua*; F - *A. fraterculus*; S - *A. sororcula*; B - *A. bistrigata*; G - *A. grandis*; E - *A. serpentina*; T - *A. striata*. Haplótipos em negrito foram obtidos a partir do GenBank (<http://www.ncbi.nlm.nih.gov>).

Tabela A2. Estimativa de diversidade genética na região de *doublesex*.

Espécies	N	h	Hd	S	S _y	N _{Sy}	π	π _s	π _a
<i>A. fraterculus</i>	37	34	0,995 ± 0,007	77	42	33	0,0178	0,0403	0,0102
<i>A. obliqua</i>	46	41	0,992 ± 0,008	77	34	47	0,0106	0,0212	0,0068
<i>A. sororcula</i>	18	18	1,000 ± 0,019	45	25	21	0,0161	0,0376	0,0092
<i>Anastrepha</i> sp.	109	97	0,995 ± 0,003	156	71	68	0,0159	0,0371	0,0089

N = número de sequências; h = número de haplótipos; Hd = diversidade haplotípica; S = Número de sítios polimórficos; S_y = número de mudanças sinônimas; N_{Sy} = número de mudanças não-sinônimas; π = diversidade nucleotídica; π_s = diversidade nucleotídica sinônima; π_a = diversidade nucleotídica não-sinônima;

2. Comparação das taxas evolutivas e seleção em *dsx*

A Figura A2 mostra a árvore filogenética gerada pela análise de dados macroevolutivos considerada para o teste de heterogeneidade de valores de ω entre ramos e códon. Nela estão mostrados os ramos *foreground* e *background* contrastados na análise. A Tabela A3 mostra os parâmetros estimados pelos modelos nulos M1a e MA restrito e modelo alternativo MA.

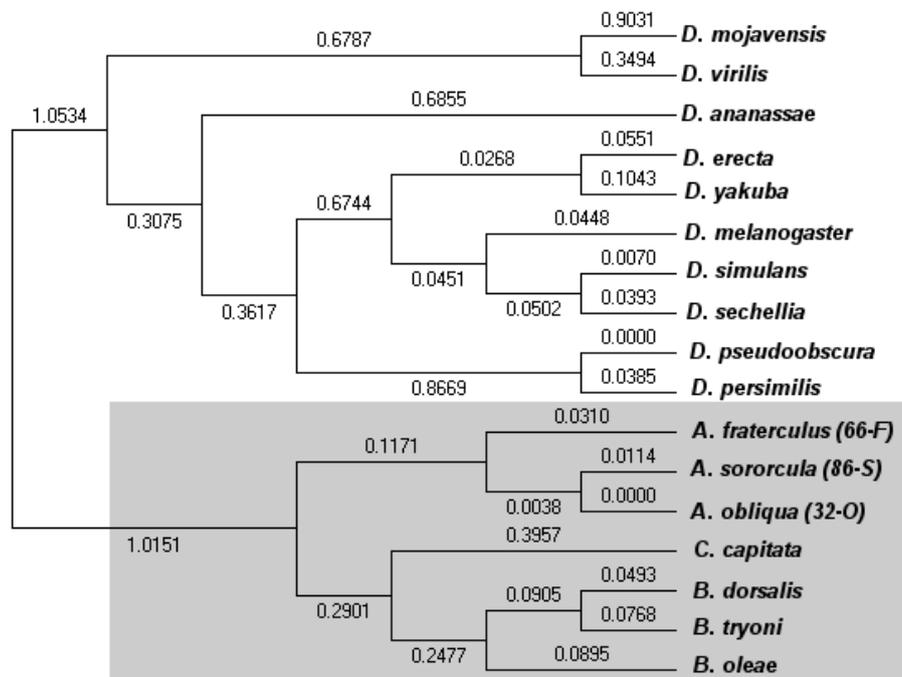


Figura A2. Árvore filogenética estimada por Máxima Verossimilhança para os dados de contrastes macroevolutivos mostrando o ramo *foreground* destacado por uma área cinza. Os tamanhos dos ramos estão estimados em substituições nucleotídicas por códon.

Tabela A3. Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito nos dados macroevolutivos.

Modelo	Parâmetros	LnL
M1a <i>Modelo nulo</i>	$\omega_0 = 0,090$ $p_0 = 0,828$ $\omega_1 = 1,000$ $p_1 = 0,172$	-5700,245
MA restrito ($\omega_2 = 1$, fixado) <i>Modelo nulo</i>	$\omega_0 = 0,080$; $\omega_2 = 1,000$ $p_0 = 0,704$; $p_1 = 0,149$; ($p_2 + p_3$) = 0,147	-5686,851
MA <i>Modelo alternativo</i>	$\omega_0 = 0,083$; $\omega_2 = 15,244$ $p_0 = 0,707$; $p_1 = 0,146$; ($p_2 + p_3$) = 0,147	-5679,637

LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS para sítios com $0 < \omega < 1$; ω_1 = valores de dN/dS fixados a 1; ω_2 = valores de dN/dS para sítios com $\omega > 1$ e correspondem apenas a sítios no ramo *foreground*.

Os contrastes entre o modelo para seleção positiva no *foreground* e os modelos nulos comprovam que parte da diferenciação ocorrida em *dsx* deveu-se à ação de seleção positiva sobre a região estudada (Tabela A4), particularmente entre os códons 139 e 203, onde o teste de BEB revelou a maior parte dos sítios selecionados positivamente (Figura A3).

Tabela A4. Contraste dos modelos M1a, MA restrito e MA por LRT nos dados macroevolutivos.

Contraste	G.L.	LRT	Sítios selecionados positivamente
M1a x MA	2	41,216**	11, 20, 141, 144, 146, 147, 151, 152, 197, 200, 201, 323, 327, 329, 331
MA restrito x MA	1	14,427*	

* $p < 0,01$; ** $p < 10^{-8}$

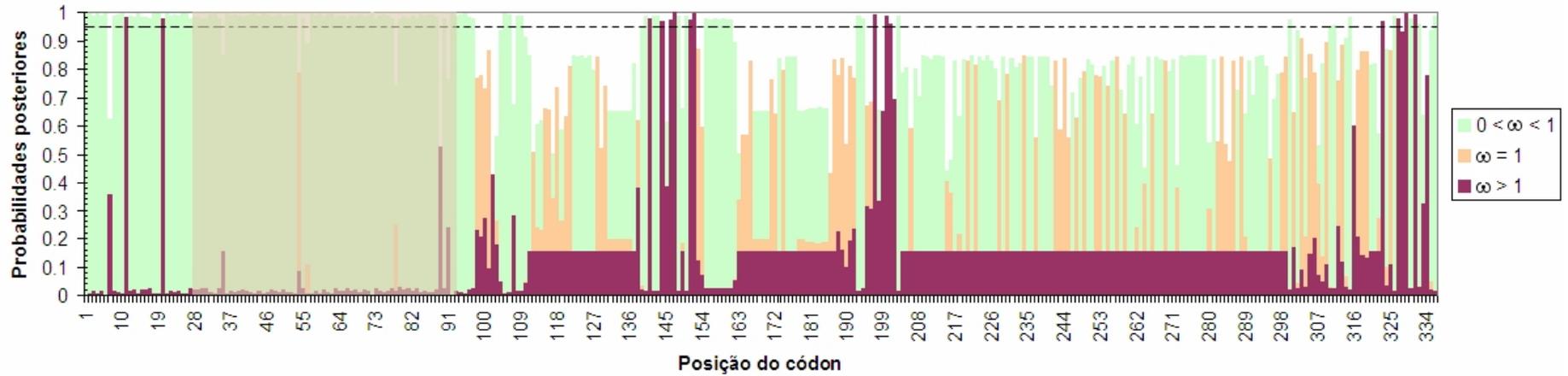


Figura A3. *Bayes Empirical Bayes* mostrando probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) para o ramo *foreground*. A área sombreada representa o domínio de ligação ao DNA. A linha tracejada indica a probabilidade posterior de 0,95. A numeração dos códons não indica a posição real do mesmo no *dsx*. Os valores de ω médio estão representados no eixo da direita.

3. Estudo de taxas evolutivas em dados intrapopulacionais

O modelo de substituição homogênea M0 do CODEML (PAML ver. 4) estimou a razão média do ω como sendo 0,450 para o *dsx*. A comparação dos modelos de seleção (M2a e M8) contra os modelos neutros no PAML (M1a, M7 e M8a) não detectou diferenças significantes entre modelos que permitiam seleção positiva e aqueles que restringiam os valores de ω a valores ≤ 1 . Da mesma forma, nenhum sítio sob seleção positiva foi detectado pela análise de *Bayes Empirical Bayes* (Figura A4). Estas análises mostram que a maioria dos sítios está sob seleção purificadora, com 58% dos códonos apresentando ω entre 0,3 e 0,4 e outros 23% apresentando ω entre 0,5 e 0,6 no geral. Tais valores são virtualmente os mesmos quando apenas consideramos a região de ligação de DNA (Figura A4).

Para termos uma idéia da pressão seletiva *relativa* sobre a região estudada de *dsx*, comparamos sua taxa ω com a de um gene notadamente conservado, COI (Tabela A5). O primeiro teste de heterogeneidade (A x E) rejeitou a hipótese nula de homogeneidade nas taxas totais. Num segundo teste (A x B) a hipótese de homogeneidade nas taxas de substituição foi rejeitada e novamente houve a rejeição da hipótese de homogeneidade de ω e κ entre COI e *dsx*. (Tabela A6). De fato, as taxas observadas para COI e *dsx*, respectivamente 0,09 e 0,45, podem ser consideradas significativamente diferentes pela estimativa de seus intervalos de confiança (Figura A5).

Tabela A5. Teste de heterogeneidade das taxas evolutivas entre *dsx* e COI_{mt}.

Modelo	LnL	ω	κ	taxas
A	-2484,478	0,256	3,743	1
B	-2478,264	0,281	3,733	COI - 1 <i>dsx</i> - 1,953
C	-2408,038	0,258	3,893	COI - 1 <i>dsx</i> - 1,863
E	-2394,283	COI - 0,079 <i>dsx</i> - 0,578	COI - 4,530 <i>dsx</i> - 3,676	COI - 1 <i>dsx</i> - 1,786

LnL = Log da máxima verossimilhança

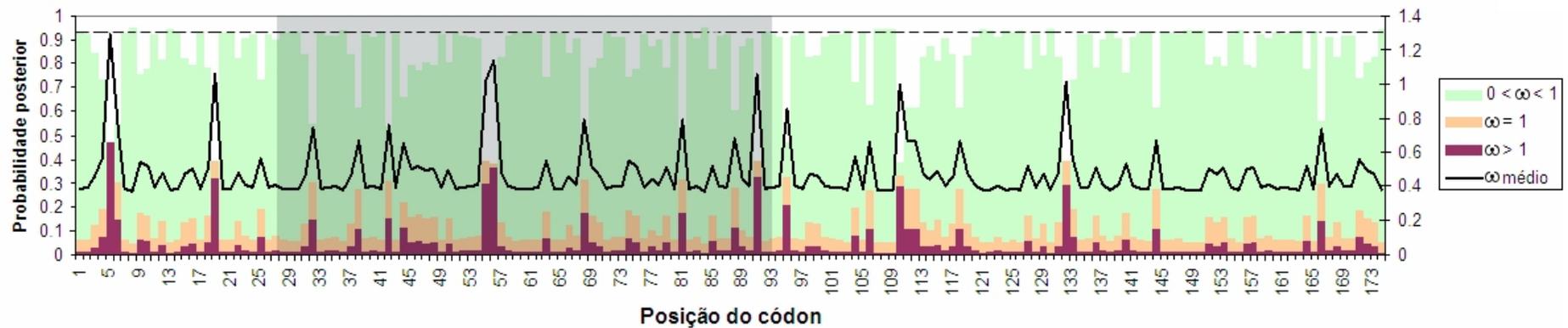


Figura A4. *Bayes Empirical Bayes* mostrando probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) para o ramo *foreground*. A área sombreada representa o domínio de ligação ao DNA. A linha tracejada indica a probabilidade posterior de 0,95. A numeração dos códons não indica a posição real do mesmo no *dsx*. Os valores de ω médio estão representados no eixo da direita.

Tabela A6. Contrastes dos modelos para teste de heterogeneidade de taxas evolutivas entre *dsx* e COInt

Contraste	LRT	G.L.	$p \chi^2_{GL}$
Modelo A x Modelo E	214,326	12	$3,549 \cdot 10^{-32}$
Modelo A x Modelo B	15,219	1	$4,000 \cdot 10^{-4}$
Modelo C x Modelo E	8,691	2	$1,909 \cdot 10^{-7}$

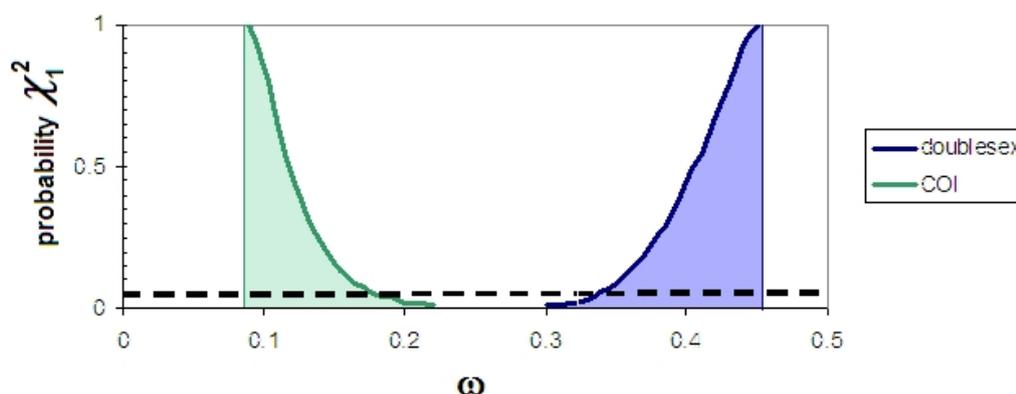


Figura A5. Intervalo de confiança de valores de ω estimado para *doublesex* COInt

As mutações não-sinônimas internas foram avaliadas quanto à possibilidade de alteração radical de uma propriedade físico-química (Figura A1) (Sainudiin *et al.* 2005). Neste caso, apenas duas mutações não-sinônimas internas, 8 e 22, levaram a mudanças radicais no formato ou na carga da molécula. Considerando a mutação *56-Y_100-H* como sendo a ancestral (inferida pela probabilidade estimada no TCS para este grupo), a mutação 8 correspondeu à mudança de um aminoácido polar (Y - Tirosina) para um apolar (F - Fenilalanina), enquanto a mutação 22 resultou na mudança de um aminoácido polar carregado positivamente (H - Histidina) para um polar sem carga (N - Asparagina). Tais mutações separaram a rede em grandes 3 grupos mutacionais (Figura A1), e foram analisadas individualmente quanto à sua distribuição geográfica e homoziguidade nos indivíduos (Figura A6). A maior parte dos indivíduos encontrados eram homozigotos para cada uma das mutações apesar de suas áreas de ocorrência se sobreporem em alguns casos. Alguns indivíduos heterozigotos foram encontrados, porém, numa frequência baixa e com localização restrita.

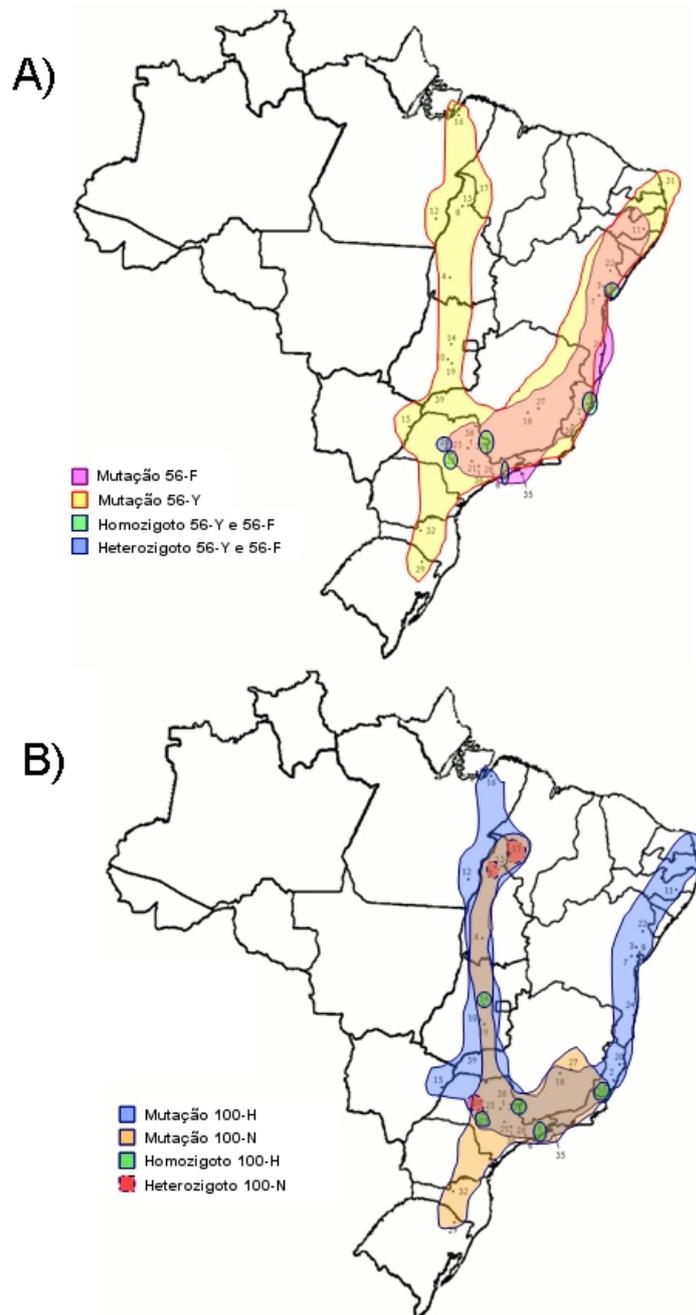


Figura A6. Distribuição de mutações não-sinônimas que levam a alterações radicais nas propriedades dos aminoácidos. O número presente na denominação da mutação indica a posição do códon onde ela ocorreu e a letra corresponde ao aminoácido codificado pelo códon. Quando ambas as mutações ocorreram num mesmo ponto de coleta foi indicada a presença de homozigotos ou heterozigotos para tais mutações.

3.1 Testes populacionais de neutralidade

Os testes D de Tajima e D e F de Fu e Li foram inicialmente usados nos dados de *doublesex*. Todos estes testes mostraram desvios significantes da neutralidade (Tajima's D = - 2,36, $p < 0,01$, Fu e Li's D = -6,39 e F = -5,50, $p < 0,02$). O teste de contingência de Templeton apresenta desvio de neutralidade revelando um significativo excesso de substituições não-sinônimas nas pontas e substituições sinônimas nos ramos interiores (Tabela A7 e Figura A1).

Para avaliar o impacto das mutações não-sinônimas na dinâmica evolutiva da região do *dsx* estudada, consideramos o número de indivíduos gerados por tais mutações em relação às mutações sinônimas (Tabela A8).

Tabela A7 . Teste de contingência de mutações sinônimas e não-sinônimas vs ponta e interior.

	Não-sinônima	Sinônima
Ponta	92	71
Interior	8	32

Probabilidade do teste exato de Fisher's sob a hipótese nula de homogeneidade: $p < 0,000025^{***}$

Os testes de contingência contrastando mutações, sinônimas e não-sinônimas, nas pontas ou nos interiores (Tabela A9), contextualizados pelo número de indivíduos produzidos por estas mutações, rejeitou a hipótese nula de homogeneidade nas frequências de indivíduos originados por estas mutações.

Da mesma forma, o teste não-paramétrico de Mann-Whitney revelou que substituições não-sinônimas geram mais descendentes do que substituições sinônimas (ranking mediano de 31,4 e 19,8, respectivamente, $p < 0,02$). Desta forma, para ambos testes detectamos uma heterogeneidade nas frequências em direção a um excesso de mutações sinônimas internas originando poucos indivíduos enquanto mutações não-sinônimas em geral deram origem a um número maior de descendentes.

Tabela A8. Contagem de indivíduos gerados por mutações sinônimas e não-sinônimas internas

Mutação interna	Tipo de mutação interna Sinônima ou Não-sinônima	Número de indivíduos originados a partir da mutação interna
2	Sinônima	2
3	Sinônima	2
7	Sinônima	2
12	Sinônima	2
21	Sinônima	2
24	Sinônima	2
26	Sinônima	2
32	Sinônima	2
32	Sinônima	2
32	Sinônima	2
33	Sinônima	2
34	Sinônima	2
34	Sinônima	2
17	Sinônima	3
17	Sinônima	3
17	Sinônima	3
1	Não-sinônima	3
5	Sinônima	4
15	Sinônima	4
31	Sinônima	4
9	Não-sinônima	4
18	Sinônima	5
18	Sinônima	5
30	Sinônima	5
6	Sinônima	6
6	Sinônima	6
14	Sinônima	6
29	Sinônima	6
6	Não-sinônima	6
10	Sinônima	9
23	Sinônima	9
13	Sinônima	10
28	Sinônima	11
27	Sinônima	12
25	Não-sinônima	15
4	Não-sinônima	16
22	Não-sinônima	24
20	Sinônima	25
20	Não-sinônima	25
11	Sinônima	26
19	Sinônima	27
16	Sinônima	33
8	Não-sinônima	78

* Quando os dois tipos de mutações estão presentes em um mesmo ramo interno, cada uma é contada independentemente.

Tabela A9. Análise de heterogeneidade nas distribuições de frequências entre número de indivíduos formados por uma mutação interna e tipo de mutação.

	2 a 11 indivíduos	> 11 indivíduos	Total
Sinônima	21	5	26
Não-sinônima	3	5	8
Total	24	10	34

*Probabilidade pelo teste permutacional exato usando o algoritmo de (Roff & Bentzen 1989): $p < 0,05$

	2 a 12 indivíduos	> 12 indivíduos	Total
Sinônima	22	4	26
Não-sinônima	3	5	8
Total	25	9	34

*Probabilidade pelo teste permutacional exato usando o algoritmo de (Roff & Bentzen 1989): $p < 0,05$

Discussão

Nossa investigação populacional da região conectora do gene *dsx* indicou altos níveis de polimorfismo intra e interespecífico no grupo *Anastrepha fraterculus*. A diversidade haplotípica (Hd) no *dsx* foi muito alta, revelando um excesso de haplótipos raros e únicos, com apenas um conjunto de haplótipos que foram encontrados mais de uma vez. Nosso esquema amostral, que favoreceu indivíduos de localidades diferentes, pode explicar em parte o excesso de haplótipos raros observados, mas não é capaz de explicar a alta proporção de heterozigotos em geral observada. Portanto, é possível que a diversidade haplotípica observada deva ser em grande parte representativa da população como um todo. Apesar disso, os polimorfismos detectados não estão associados com diferenças fixadas que sejam específicas para cada espécie, um resultado consistente com outros estudos do grupo de espécies *fraterculus* usando dados de COI mitocondriais e genes nucleares (Smith-Caldas *et al.* 2001; Ruiz *et al.* 2005; Ruiz *et al.* 2007b). Como não fomos capazes de detectar monofilia recíproca ou mesmo parafilia entre espécies diferentes neste grupo de espécies, alguns dos parâmetros populacionais estimados também foram estimados para o grupo de espécies como um todo, considerando o clado como uma unidade evolutiva de interesse. Pelo mesmo motivo, o teste de McDonald e Kreitman não pode ser realizado, uma vez que tal teste requer grupos de espécies bem definidos para se comparar a razão de mudanças sinônimas e não-sinônimas fixadas e polimórficas entre os indivíduos. Esta alta diversidade de alelos raros resulta em testes de neutralidade, como o D de Tajima e o D e o F de Fu e Li, significativamente negativos que são compatíveis com um padrão geral de seleção purificadora. Tais resultados concordam com os dados de diversidade haplotípica que indicam um excesso global de alelos raros que se encaixa nas expectativas de seleção de fundo, em que o balanço entre seleção contra substituições deletérias e mutações recorrentes mantém a frequência de tais variantes a um nível baixo. Embora tais testes sejam usados para se detectar os efeitos gerais da seleção na região do DNA, resultados similares podem também ser obtidos por alguns fatores demográficos, tais como expansão do tamanho populacional, mesmo na ausência da seleção (Charlesworth *et al.* 2003; Zeng *et al.* 2007).

Para se investigar os efeitos de seleção nesta região de DNA, e tentar separar os efeitos demográficos, usamos do método provavelmente mais considerado para a determinação de regiões específicas influenciadas por seleção, a investigação da razão de taxas ω (Yang *et al.* 2005).

A análise da razão de ω com dados populacionais no PAML confirmou os padrões gerais de seleção purificadora, evidenciados pela inabilidade em se rejeitar modelos que não permitem sítios sob seleção positiva ($\omega > 1$) por testes de LRT. Ou seja, de acordo com este teste,

não há evidência de sítios sujeitos à seleção positiva. Apesar disso, uma análise mais detalhada do ω em códons individuais com o BEB mostra que, embora a região comum do *dsx* seja conservada, esta região apresenta níveis mais altos de substituições não-sinônimas do que a região do COInt, mesmo em sua região mais restrita, o domínio de ligação de DNA. A estimativa da taxa de evolução da região amostrada do *dsx* foi aproximadamente oito vezes maior do que a do COI mitocondrial (0,450 e 0,057, respectivamente) para o mesmo grupo de espécies (de fato, para diversas das mesmas populações e indivíduos). Isso pode sugerir que a seleção purificadora agindo na região *dsx* seja relativamente mais fraca em comparação ao COInt, ou que talvez os padrões de seleção no *dsx* sejam mais complexos e que exista mais do que apenas seleção purificadora agindo na variação genética desta região. Isto pode ser em parte explicado pela ausência, ou baixa frequência, de recombinação nos genes mitocondriais, o que pode não ser necessariamente o caso (Tsaousis *et al.* 2005), é possível que a seleção purificadora seja de fato mais forte nos genes mitocondriais do que em alguns genes nucleares, como o *dsx*. De qualquer forma, esta taxa evolutiva mais alta em *dsx* não seria exatamente compatível com um modelo exclusivo de seleção purificadora para esta região como anteriormente proposto (Ruiz *et al.* 2007b) e podem sugerir a existência de uma força evolutiva promovendo esta alta taxa evolutiva, e talvez até mesmo a manutenção de ao menos parte destes polimorfismos, o que seria em geral incompatível com um modelo puro de seleção purificadora.

Embora a razão das taxas ω avaliada em códons individuais possa ser uma metodologia poderosa para se detectar sítios sob seleção positiva, o teste é bastante conservador para se detectar evolução adaptativa em genes geralmente conservados (McClellan *et al.* 2005; Porter *et al.* 2006) e pode ainda ter pouco poder para lidar com seqüências com baixa divergência ou mesmo sob seleção positiva fraca (Nickel *et al.* 2008), requerendo em geral um nível razoável de divergência entre os táxons em estudo. Por este motivo, utilizamos inicialmente esta metodologia para o contraste de espécies que apresentam um tempo de divergência significativo de forma a nos permitir detectar de fato, em situações mais apropriadas para o uso dos testes de razão de ω , se esta região do *dsx* que estamos estudando apresenta qualquer evidência de seleção positiva influenciando sua evolução. Este contraste, a que nos referimos como macroevolutivo, indicou claramente que embora a região de ligação com o DNA esteja evoluindo em geral conservadoramente, a região conectora apresenta claramente diversos sítios que estão evoluindo sob a ação da seleção positiva.

Apesar desta evidência de seleção positiva no contraste entre espécies distinta de Diptera, referida como testes macroevolutivos, o uso dos testes de razão de ω em nossos dados populacionais não foi capaz de detectar evidências significantes de seleção positiva. Como o uso destas metodologias em dados populacionais é muito mais limitado e problemático (Kryazhimskiy

& Plotkin 2008), devemos considerar se de fato a seleção positiva detectada no teste anterior indica um processo que aconteceu na separação entre os Drosophilidae e os Tephritidae, e hoje não é detectado entre as espécies de *Anastrepha* pela inexistência destes sinais, ou simplesmente a seleção positiva continua a ocorrer, mas não temos sinais suficientes para sua detecção utilizando os programas previamente mencionados nestas condições não adequadas para seu uso.

Portanto, buscamos estratégias diferentes para tentar detectar sinais de mudanças adaptativas no *dsx*, uma vez que as seqüências de diferentes espécies de *Anastrepha* mostraram baixa divergência e a região estudada foi em geral conservada. Um primeiro aspecto relevante seria considerar testes que levam em consideração as mudanças em propriedades físico-químicas de aminoácidos em sítios não-sinônimos (por exemplo, como os de McClellan *et al.* 2005). Quando considerando as propriedades físico-químicas no estudo da evolução adaptativa, devemos considerar dois possíveis cenários: (1) as substituições entre aminoácidos radicalmente diferentes têm maior probabilidade de promover mudanças estruturais ou funcionais em uma proteína. Neste caso, podemos invocar a seleção positiva e mudanças ocorreriam mais freqüentemente em um sítio sob seleção do que esperado por acaso, mesmo se os valores de ω forem mais baixos do que 1; (2) substituições entre aminoácidos quimicamente similares levariam a pouca ou nenhuma mudança na estrutura e função de uma proteína, e portanto, mesmo se um sítio apresentar ω maior do que 1, ele poderia ainda assim ser funcional ou estruturalmente conservado e estaria evoluindo conservativamente.

Considerando a terminologia de Sainudiin *et al.* (2005), para a identificação de mutações radicais, que utiliza um número bem menor de características físico-químicas do que o TreeSAAP, encontramos apenas duas mutações radicais presentes em nossa análise populacional. Estas mutações definem três grandes grupos de haplótipos no cladograma inferido pela parcimônia estatística no TCS 1.21. Curiosamente, ao avaliarmos os indivíduos amostrados para estas mutações radicalmente distintas, observamos que, embora quase todos os haplótipos sejam distintos, a grande maioria dos indivíduos amostrados foi homozigota para estas mutações radicais. Isto é ainda mais relevante em se considerando nossa estratégia amostral de coletar poucos indivíduos por região em diversos pontos da distribuição das espécies. Ainda mais interessantemente, quando estudamos a distribuição geográfica destas mudanças radicais, encontramos uma congruência razoável na distribuição destas mudanças radicais, que definem grupos distintos que estão em geral distribuídos em regiões distintas com um número muito reduzido de heterozigotos entre estas mutações radicais, geralmente localizados em regiões intermediárias entre estes grupos. Estes resultados são em grande contraste com outros dados de genes mitocondriais e nucleares para estas espécies que não encontraram uma fragmentação histórica significativa entre amostras coletadas de regiões distintas do Brasil. Tal padrão pode ser condizente com processos de seleção positiva favorecendo a

manutenção de polimorfismos distintos em regiões diferentes, um aspecto fascinante que merece ser ainda bem mais explorado.

Estes dados indicam, contudo, que apesar dos testes de ω não terem indicado uma seleção positiva significativa, existem padrões na região do *dsx* aqui estudada que indicam que esta região esteja evoluindo diferentemente de um modelo de seleção eminentemente purificadora, e buscamos métodos que a pudessem detectar na população.

Um dos sinais mais relevantes de seleção positiva que se mantém por um longo período, basicamente até que a mutação seja completamente fixada em um *selective sweep*, é que esperamos encontrar alelos sob seleção com uma frequência média maior do que seria de se esperar se a região estivesse evoluindo neutramente (Halliburton 2004). Por outro lado, regiões sujeitas à seleção purificadora apresentam em geral um excesso de alelos de baixas frequências (Fay & Wu 2000). Ambos padrões podem ser explicados em geral por processos demográficos específicos, mas o contraste de regiões distintas pode permitir a distinção destes parâmetros.

As duas estratégias que criamos, a análise hierárquica do teste de contigência de Templeton e o teste de Mann-Whitney, buscam avaliar se mutações sinônimas e não-sinônimas apresentam a mesma distribuição de probabilidades de sobrevivência na população. O que ambos os testes nos revelam é que de fato substituições não-sinônimas, embora menos frequentes do que substituições sinônimas na população, apresentam-se de fato em classes mais frequentes do que seria de se esperar pelo acaso. Isso significa que estes testes revelam que tais mutações estão sendo ativamente mantidas na população por um processo de seleção positiva nesta região do *dsx*, apesar de um padrão geral de seleção purificadora.

Estudos têm demonstrado que embora a maioria dos genes reguladores, tais como os genes HOX, tem sido extremamente conservados evolutivamente (Ferrier & Minguillón 2003), diversos genes que estejam sob seleção positiva, como os genes que distinguem chimpanzés de humanos (Johnson *et al.* 2001; Bakewell *et al.* 2007) e genes na cascata de determinação sexual, tais como o *transformer*, podem apresentar níveis surpreendentemente altos de variabilidade (Clark *et al.* 2006). Tais estudos têm demonstrado que estes genes podem apresentar altas taxas de evolução que tem sido consideradas como neutras (Kulathinal *et al.* 2003) ou direcionadas por seleção (O'Neil & Belote 1992). Considerando que o gene *dsx* é crítico para a diferenciação de insetos em machos e fêmeas, as mutações no *dsx* que promovem efeitos altamente deletérios podem fornecer sérias restrições à reprodução. Portanto, esperaríamos níveis de variação não-sinônimas mais baixas neste gene em que a razão das taxas ω fosse em geral mais baixa. Na realidade, diversas mutações que estão segregando nessa população desviam destas expectativas e apresentam mudanças radicais para a proteína.

Uma explicação possível para o padrão de variação encontrada na região comum do *dsx* poderia ser a natureza dualista dos efeitos do *dsx*. Por um lado, o *dsx* é essencial para uma função altamente conservada, a regulação da cascata de determinação sexual, por outro lado, *dsx*, assim como o *fruitless*, também está envolvido no comportamento de corte masculino (Rideout *et al.* 2007), o qual pode ser sujeito à uma seleção sexual positiva. Conseqüentemente, mudanças na DSX podem ser seletivamente desvantajosas para uma função, mas não para a outra. É mesmo possível que por causa de tais interações complexas, as substituições no *dsx* poderiam se comportar como mutações ligeiramente desvantajosas ou vantajosas dependendo do arcabouço genético com o qual está interagindo, um processo que se assemelha a epistasia de sinal (Weinreich *et al.* 2005). Isto é ainda mais provável se considerarmos a possibilidade de não-aditividade e epistasia que tem sido comumente detectadas em caracteres complexos (Carlborg & Haley 2004).

Uma combinação de seleção fortemente purificadora por um lado, e seleção positiva em sítios específicos pode produzir o padrão geral detectado de restrição seletiva relaxada em certos testes de seleção aqui investigados, particularmente na região comum entre os sexos. Tal padrão de restrição seletiva relaxada parece ser comum a outros genes de determinação sexual, tais como o *transformer* (*tra*), que mostrou-se evoluir relativamente não restrito, com uma razão de taxas ω geralmente superior ($\omega = 0,2$ to $0,3$) àquelas encontradas em genes não relacionados à diferenciação sexual (McAllister & McVean 2000; Kulathinal *et al.* 2003). Tais dados indicam, contudo, que o *tra* em *Drosophila* ainda esteja sob seleção purificadora, embora moderada em grande extensão do gene (McAllister & McVean 2000; Kulathinal *et al.* 2003). Por outro lado, *tra* foi considerado sob intensa seleção purificadora para algumas espécies do grupo *fraterculus* (Ruiz *et al.* 2007b).

Em conclusão, nosso estudo mostra que embora a região comum do *dsx* seja relativamente conservada, apresenta uma taxa não-usual de mudanças não-sinônimas, que pode indicar que seleção purificadora não seja tão forte quanto esperado para um gene diretamente relacionado ao valor adaptativa do inseto. Parte desta variação pode ser explicada por mudanças adaptativas na regiões fora do domínio de ligação do DNA promovidas por seleção positiva, reveladas por testes populacionais aqui desenvolvidos. Além disso, o relaxamento da restrição seletiva mediada por interações epistáticas entre *dsx* e os genes por este regulados podem explicar a segregação das substituições não-sinônimas no resto da região comum do *dsx*.

Capítulo III

Evolução molecular do gene *fruitless* no grupo *Anastrepha fraterculus* (Diptera, Tephritidae)

Iderval S. Sobrinho-Jr. e Reinaldo A. de Brito

Introdução

As espécies de *Anastrepha* são identificadas pelo uso de uma gama de marcadores morfológicos, mas principalmente pela forma e tamanho do ápice do acúleo e em alguns grupos, como no grupo *fraterculus*, são separadas por diferenças sutis neste marcador (Malavasi & Zucchi 2000; Araujo & Zucchi 2006). Além de restringir a identificação às fêmeas, este marcador morfológico apresenta alta plasticidade fenotípica, a ponto de haver sobreposição de diversas medidas quantitativas, o que restringe seu uso para a separação de espécies muito próximas deste grupo, notadamente *A. fraterculus*, *A. obliqua*, *A. sororcula*, *A. zenildae* e *A. turpinae* (Araujo & Zucchi 2006). Mesmo marcadores genéticos como DNA mitocondrial e genes nucleares como *doublesex* (*dsx*) e *transformer* (*tra*) mostraram baixa resolução filogenética no gênero *Anastrepha* (Smith-Caldas *et al.* 2001; Ruiz *et al.* 2007a; Ruiz *et al.* 2007b).

Apesar de *dsx* e *tra* terem apresentado baixa resolução, em geral estudos de evolução molecular têm mostrado que muitos genes relacionados à reprodução apresentam uma taxa mais elevada de divergência (Swanson & Vacquier 2002b; Gerrard & Meyer 2007) devido principalmente à ação de seleção positiva (Swanson *et al.* 2004; Clark & Swanson 2005). Por estarem envolvidos com características reprodutivas, sugere-se a seleção sexual e/ou conflito sexual como mecanismos para esta rápida diferenciação (Swanson & Vacquier 2002a). Por causa dessa alta divergência e da possibilidade de estarem envolvidos diretamente no isolamento reprodutivo, tais genes seriam os candidatos mais indicados para se distinguir espécies recentemente separadas e de difícil distinção por outros marcadores genéticos, bioquímicos e/ou morfológicos (Ting *et al.* 2000).

Genes relacionados a características reprodutivas seriam então, bons candidatos a marcadores genéticos para estudar espécies crípticas como as do grupo de espécies de *Anastrepha*, especialmente aqueles envolvidos diretamente em características passíveis de serem selecionadas por seleção sexual. Um dos genes desta cascata de diferenciação sexual, *fruitless* (*fru*), está envolvido diretamente no estabelecimento e desenvolvimento de circuitos neuronais específicos para o desenvolvimento do comportamento de corte de machos em Diptera (Demir & Dickson 2005; Kyriacou 2005; Kimura *et al.* 2008), na resposta a feromônios (Datta *et al.* 2008) e na formação da junção neuromuscular responsável pelo desenvolvimento de fibras musculares específicas de machos chamadas de Músculo de Lawrence (Ito *et al.* 1996; Billeter *et al.* 2006).

O *fru* pode ser dividido em quatro regiões principais: domínio específico ao sexo, domínio de dimerização (BTB), região conectora e domínio de ligação a DNA (*Zinc-finger*) (Demir

& Dickson 2005). O primeiro domínio é composto por dois éxons diferentes, cada um específico a um sexo. Somente a forma masculina é ativa, visto que o mRNA de *fru* é processado alternativamente de forma dependente ao sexo e no éxon feminino um códon de parada interrompe a tradução completa da proteína. O BTB é formado por dois éxons que se unem no processamento do mRNA, da mesma forma a região conectora é formada pela união de três ou quatro éxons, dependendo do grupo taxonômico. O domínio de ligação ao DNA na extremidade 3' é composto por três ou quatro éxons (Demir & Dickson 2005; Gailey *et al.* 2006). Estes éxons são processados de forma independente e cada um forma uma isoforma específica de *fru* (Billeter *et al.* 2006). A presença de quatro promotores diferentes, um na extremidade 3' anterior aos éxons específicos a cada sexo e outros três a jusante destes éxons, permite um padrão complexo de processamento do mRNA, o qual promove a expressão tanto de formas específicas de cada sexo pelo primeiro promotor, quanto daquelas independentes do sexo do indivíduo pelos outros três promotores restantes. Assim, apesar das formas sexuais de fêmea serem inativas, suas formas não relacionadas ao sexo não o são (Demir & Dickson 2005).

Apesar de ser um gene envolvido na reprodução e ter um potencial para uma rápida diversificação por seleção adaptativa, este gene apresenta uma certa conservação em seus dois domínios principais: domínios de dimerização (BTB) e de ligação ao DNA (*Zinc-finger*) (Gailey *et al.* 2006). Em contrapartida, como a região conectora e a extremidade amino da proteína mostraram elevada divergência entre as espécies estudadas, foi sugerido que estas regiões não desempenhariam funções primordiais para a proteína e se divergiriam por um relaxamento da restrição seletiva ou poderiam conter informações para o desenvolvimento do comportamento de corte de machos específico para cada espécie, daí sua alta divergência entre os grupos estudados (Gailey *et al.* 2006).

Portanto, um dos objetivos deste trabalho foi o de estudar o nível de variabilidade genética do gene *fru* e verificar se a alta variabilidade na região conectora era devido à ação de seleção positiva ou se a região está evoluindo por um relaxamento de restrição seletiva. Por último procuramos verificar o potencial uso desta região como um marcador para diferenciar as espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*.

Material e Métodos

1. Extração do DNA e seqüenciamento

Frutos de diversas espécies de planta infestadas por *Anastrepha* foram coletadas de 32 localidades no Brasil (Tabela B1) e dispostas em vermiculita por 14 dias quando pupas eram separadas. Após a emergência em maturação, as moscas eram identificadas e se pertencentes ao grupo *A. fraterculus* group, na maioria das vezes *A. fraterculus*, *A. obliqua* e *A. sororcula*, eram separadas e imediatamente processadas ou preservadas em etanol 95% até a extração do DNA.

A extração do DNA foi feita usando-se um protocolo modificado de (Chomczynski & Sacchi 1987) para se garantir a manutenção intacta do exo-esqueleto para análises morfológicas futuras. Amplificamos uma região do extremo final do domínio de dimerização (BTB) até próximo ao final do primeiro éxon da região conectora do gene *fru*, usando primers degenerados criados de seqüências homólogas de espécies proximamente relacionadas (5'-AGTTCGCTGCCGATGTTYCTCAA-3' e 5'-GACAGRCACTAYCCGCAGGACTCTCAG-3'). Esta região foi amplificada por PCR a partir do DNA genômico em um termociclador PTC-200 (BioRad) usando uma mistura de *Taq* e *Pfu* polimerase para se reduzir erros de incorporação (Cline *et al.* 1996). Os produtos de PCR foram purificados por precipitação com PEG 8000 (Lis & Schleif 1975) de clonados usando o kit InsTAclone (Fermentas). Ao menos duas colônias recombinantes foram seqüenciadas usando-se primers M13F (forward) e M13R (reverse) com o kit DYEnamic™ ET dye terminator (GE Healthcare) em corridos em um MegaBace 1000 (GE Healthcare). A qualidade do seqüenciamento foi inspecionado visualmente no software Chromas ver. 2.31 (<http://www.technelysium.com.au>). Tais seqüências foram alinhadas usando o software Clustal no software BioEdit Sequence Alignment Editor (Hall 1999) e visualmente inspecionadas. Quando seqüências iguais ou muito similares (diferindo de uma ou duas bases) foram obtidas, seqüências adicionais foram realizadas em colônias recombinantes diferentes (até o total de 5 colônias) para se confirmar a eventual homozigosidade dos indivíduos ou possíveis incorporações errôneas de base.

Tabela B1. Locais amostrados com informação geográfica, haplotípica e de espécies.

código	Localidade	Latitude (S)	Longitude (W)	Haplótipos
01	São Carlos - SP	22° 01' 03"	47° 53' 27"	1-f.1, 2-f.1
02	Vitória - ES	20° 19' 10"	40° 20' 16"	38-o.2, 39-o.2, 72-o.2, 73-o.2
03	Feira do Santana -BA	12° 16' 00"	38° 58' 00"	69-o.3, 70-o.3
04	Gurupi - TO	11° 43' 45"	49° 04' 07"	57-o.4, 58-o.4
05	São Sebastião – SP (Santiago)	23° 45' 36"	45° 24' 35"	8-f.5, 9-f.5, 82-s.5, 83-s.5
06	Bertioga - SP	23° 51' 16"	46° 08' 19"	20-f.6, 21-f.6
07	Conceição do Almeida - BA	12° 46' 46"	39° 10' 12"	28-f.7, 29-f.7, 30-f.7, 31-f.7
08	Araguaína – TO	07° 11' 28"	48° 12' 26"	40-o.8, 41-o.8, 42-o.8, 43-o.8, 44-o.8, 45-o.8, 46-o.8, 47-o.8
09	Santo Amaro – BA	12° 32' 48"	38° 42' 43"	1-f.9, 27-f.9, 65-o.9, 66-o.9, 67-o.9, 68-o.9
10	Goiânia - GO	16° 40' 43"	49° 15' 14"	50-o.10, 51-o.10
11	Bonito – PE	08° 28' 13"	35° 43' 43"	3-f.11, 4-f.11, 5-f.11, 77-s.11, 78-s.11, 79-s.11
12	Redenção – PA	08° 01' 43"	50° 01' 53"	52-o.12, 53-o.12, 54-o.12
13	Babaçulândia – TO	07° 12' 17"	47° 45' 25"	38-o.13, 3-o.13
14	Pirenópolis – GO	15° 51' 09"	48° 57' 33"	55-o.14, 56-o.14
15	Três Lagoas – MS	20° 45' 04"	51° 40' 42"	61-o.15, 62-o.15
16	Belém -PA	01° 27' 21"	48° 30' 16"	63-o.16, 64-o.16
18	Belo Horizonte - MG	19° 55' 15"	43° 56' 16"	71-o.18
19	Bela Vista de Goiás - GO	16° 58' 22"	48° 57' 12"	48-o.19, 49-o.19
20	Linhares – ES	19° 23' 28"	40° 04' 20"	74-o.20, 75-o.20, 94-s.20, 95-s.20
21	Piracicaba – SP	22° 43' 31"	47° 38' 57"	84-s.21, 85-s.21, 86-s.21, 87-s.21
22	Nova Souré – BA	11° 14' 00"	38° 29' 00"	92-s.22, 93-s.22
23	Bauru – SP	22° 18' 53"	49° 03' 38"	24-f.23, 25-f.23, 88-s.23, 89-s.23, 90-s.23, 91-s.23
24	Porto Seguro – BA	16° 26' 59"	39° 03' 53"	80-s.24, 81-s.24
25	Boracéia – SP	22° 11' 35"	48° 46' 44"	10-f.25, 11-f.25
26	Bariri – SP	22° 04' 28"	48° 44' 25"	12-f.26, 13-f.26
27	Itabira – MG	19° 37' 09"	43° 13' 37"	6-f.27, 7-f.27
28	Moji das Cruzes – SP	23° 31' 22"	46° 11' 18"	22-f.28, 23-f.28
29	Vacaria – RS	28° 30' 44"	50° 56' 02"	32-f.29, 33-f.29, 34-f.29, 35-f.29
30	Vargem Alta – ES	20° 40' 17"	41° 00' 25"	16-f.30, 17-f.30, 18-f.30, 19-f.30
31	Natal – RN	05° 47' 42"	35° 12' 34"	14-f.31, 15-f.31
32	Caçador - SC	26° 46' 31"	51° 00' 54"	36-f.32, 37-f.32
33	Santa Isabel - SP	23°18' 0"	46°13' 0"	76-o.33

O primeiro número nos haplótipos representa sua identificação, a letra corresponde às espécies nas quais os haplótipos foram encontrados e o último número corresponde ao código da localização. O - *A. obliqua*; F - *A. fraterculus*; S - *A. sororcula*.

2. Análise da diversidade genética e do padrão evolutivo

Índices de diversidade geral, tais como a diversidade haplotípica (H_d) e nucleotídica (π) (Nei 1987), o número de sítios polimórficos, as diversidades nucleotídicas sinônima (π_s) e não-sinônima (π_a) foram calculados usando o programa DnaSP ver. 4.10.9 (Rozas *et al.* 2003). Como a recombinação interfere com as inferências filogenéticas, realizamos três métodos diferentes para se detectar eventos de recombinação: o teste *Maximum Chisquare* de Maynard Smith (1992) realizados no programa MaxChi, e os testes GENECONV (Sawyer 1989) e RDP implementados no programa RDP ver. 3b14 (Martin & Rybicki 2000). Um limite de $p < 0,05$ foi estabelecido para todos os três testes e, quando aplicável, foi realizada uma aleatorização de 1000 réplicas.

Os testes de neutralidade D de Tajima (Tajima 1989b), D e F de Fu e Li (Fu & Li 1993) e H de Fay e Wu (Fay & Wu 2000) foram realizados no DnaSP usando seqüências de *Drosophila melanogaster* (número de acesso no GenBank: D84437.1) como grupo externo nas análises da região conectora de *fru* e de *Ceratitis capitata* (número de acesso no GenBank: AF124047.1) nas análises do íntron. Não foram utilizadas as mesmas seqüências para os testes entre íntron e região conectora, visto que a região amplificada de *C. capitata* abrangia apenas uma pequena porção da região conectora e as seqüências do íntron de *Drosophila melanogaster* apresentava divergência muito grande em relação às obtidas neste trabalho, impossibilitando seu correto alinhamento. Desvios de neutralidade também foram testados usando a distribuição de substituições sinônimas e não-sinônimas em haplótipos encontrado nas pontas ou no interior (haplótipos jovens ou antigos) por um teste exato de Fisher em uma tabela de contingência 2 x 2 (Templeton 1996).

Investigamos a possibilidade de seleção positiva em dados macroevolutivos e populacionais através de duas metodologias diferentes: uma metodologia que faz uso do CODEML, implementada no programa PAML ver. 4 (Yang 2007), e outra implementada no programa TreeSAAP ver. 3.2 (Woolley *et al.* 2003).

Para a investigação da ocorrência de seleção positiva em dados macroevolutivos analisamos a razão da taxa de substituições não-sinônimas por sítio não-sinônimo pela taxa de substituições sinônimas por sítio sinônimo (dN/dS , também denominado ω). As razões das taxas de substituição sinônimas/não-sinônimas foram medidas no programa *codeml*, implementado no pacote PAML ver. 4 (Yang 2007), para se inferir as pressões seletivas na proteína. Uma razão de $\omega > 1$ em um sítio específico indicaria seleção positiva, porque as substituições não-sinônimas teriam taxas de fixação mais altas do que mutações sinônimas devido a vantagens seletivas. Por outro lado, um $\omega < 1$ indica seleção purificadora, causada por restrições funcionais no códon. A análise leva

em consideração a relação entre os haplótipos estudados. Dessa forma, estimamos inicialmente a relação filogenética entre Tephritidae, representado por três seqüências das três espécies estudadas neste trabalho (*A. fraterculus*, *A. obliqua* e *A. sororcula*), e Drosophilidae representado por 10 espécies diferentes de *Drosophila* (*D. simulans*, *D. sechellia*, *D. melanogaster*, *D. erecta*, *D. yakuba*, *D. heteroneura*, *D. silvestris*, *D. grimshawi*, *D. virilis* e *D. pseudoobscura*). As seqüências de *Drosophila* foram obtidas a partir do GenBank (<http://www.ncbi.nlm.nih.gov>) com os respectivos números de acesso: AF297054.1, AF297055.1, D84437.1, AF298222.1, AF297056.1, AF051668.1, AF051665.1, AF105124.1, AY028967.1 e AF297059.1. A estimativa da árvore filogenética foi realizada por máxima verossimilhança no programa *dnaml* implementado no pacote Phylip ver 3.68 (Felsenstein 1989; Felsenstein 2004), assumindo uma taxa de transição/transversão (κ) igual a 2 e uma taxa de variação constante entre os sítios.

Os dados macroevolutivos foram analisados através do *Relaxed Branch-site test* (Teste Relaxado de sítios e ramos) e *Strict Branch-site test* (Teste Estrito de sítios e ramos) (Yang & Nielsen 2002; Zhang *et al.* 2005). Nestes testes, o ramo das *Anastrepha* foi denominado de ramo *foreground*, no qual seria testada a ocorrência de seleção positiva, e foi contrastado contra as outras espécies de *Drosophila*, cujos ramos foram denominados de *background*. Ambos os testes utilizam como modelo alternativo o modelo MA, no qual somente o ramo *foreground* pode conter sítios com $\omega > 1$ e este modelo alternativo é contrastado com os modelos nulos através um Teste de Razão de Verossimilhanças (LRT - *log-likelihood ratio test*) (Zhang *et al.* 2005). No primeiro teste contrasta-se o modelo MA contra o modelo M1a (modelo nulo). Neste contraste, o modelo nulo M1a admite a mesma taxa evolutiva para todos os sítios e ramos, com todos os sítios variando seus valores de ω entre 0 e 1 apenas. Por outro lado, no segundo teste, o modelo MA é restrito de forma que a categoria que assume $\omega > 1$ seja fixada a 1, isto é, todos os sítios com valores de $\omega > 1$ são forçados a evoluírem neutramente ($\omega = 1$). Então, este modelo nulo é contrastado contra o modelo MA.

Para os dados populacionais, a presença de seleção positiva foi estimada pelo cálculo dos valores de ω usando a máxima verossimilhança sob modelos diferentes de substituições de códon, denominados M1a, M2a, M7, M8 (Yang *et al.* 2005) e M8a (Swanson *et al.* 2003; Wong *et al.* 2004), os quais definem premissas diferentes para as restrições evolutivas dos códons. A seleção positiva é estimada por LRT entre modelos que permitem $\omega > 1$ (M2a e M8) e outros que não o fazem (M1a, M7 e M8a). Realizamos LRTs contrastando os seguintes pares de modelos: M1a-M2a, M7-M8 e M8-M8a. Neste caso, utilizamos a rede haplotípica de todos os haplótipos de *Anastrepha* obtidos neste trabalho, inferida por parcimônia estatística no TCS ver 1.21 (Clement *et al.* 2000).

Tanto para os dados de alta divergência (macroevolutivos) quando para os de baixa divergência (populacionais), a assinatura de seleção positiva em códons individuais foi estimada

por análises Bayesianas em que se permitiu que ω variasse entre sítios. Utilizamos o método de *Bayes Empirical Bayes* (BEB) (Yang *et al.* 2005) para estimar quais sítios estavam sob a influência de seleção positiva. No *Branch-site test* estimamos as probabilidades posteriores pela BEB através dos parâmetros do modelo MA, enquanto que no teste de seleção para dados populacionais tais probabilidades foram obtidas pelos parâmetros do modelo M2a.

As análises realizadas no TreeSAAP consideram as mudanças em diversas propriedades físico-químicas induzidas por cada substituição não-sinônimas. Um desvio global da neutralidade é verificado por um teste de ajustamento entre uma distribuição neutra esperada e a distribuição observada nas propriedades físico-químicas selecionadas (McClellan & McCracken 2001; Woolley *et al.* 2003; McClellan *et al.* 2005). Além disso, TreeSAAP também separa a magnitude das mudanças não-sinônimas em um gradiente que vai desde substituições mais conservadas até mais radicais, de acordo com mudanças em propriedades físico-químicas específicas. As classes mais baixas (1 a 3) representam as mudanças mais conservadoras e as classes mais altas (6 a 8) as mudanças mais radicais (McClellan *et al.* 2005). Para cada categoria, um escore z é calculado, o qual é então comparado entre todas as classes. Escores z significativamente positivos indicam que a região selecionada está sob a influência de seleção positiva, enquanto valores significativamente negativos indicam sinal de seleção purificadora. Quando um escore z significativo é detectado em categorias com magnitudes mais conservadoras (1 a 3), a propriedade é considerada sob seleção estabilizadora. Por outro lado, a propriedade é considerada sob seleção desestabilizadora quando um escore z significativo é detectado em uma categoria com magnitude mais radical (6 a 8) (McClellan *et al.* 2005). McClellan *et al.* (2005) definem seleção estabilizadora como a seleção que tende a manter os atributos bioquímicos da proteína, mesmo que tenha havido uma inferência de seleção positiva, e seleção desestabilizadora como uma seleção que favorece mudanças estruturais e funcionais em uma região da proteína. Dessa forma, a seleção positiva-desestabilizadora representa uma assinatura de adaptação molecular.

Consideramos oito categorias de magnitude de mudança para as análises e seguimos a categorização apresentada por McClellan *et al.* (2005). Apenas propriedades de aminoácidos identificadas por um escore z significativamente positivo nas categorias de magnitude 6, 7 ou 8 foram consideradas como sendo afetadas por seleção positiva-desestabilizadora. Para se verificar quais regiões específicas eram afetadas por tal seleção, realizamos uma análise da janela deslizante usando as propriedades de aminoácidos cujos resultados foram significativos para as mudanças. Janelas deslizantes de comprimento de 20 aminoácidos com um passo de um códon foram selecionadas por apresentar um melhor sinal, de acordo com Porter *et al.* (2006).

3. Diversidade genética da região de *fru* comum aos sexos e testes de seleção

Índices de diversidade geral, tais como a diversidade haplotípica (H_d) e nucleotídica (π) (Nei 1987), o número de sítios polimórficos, as diversidades nucleotídicas sinônima (π_s) e não-sinônima (π_a) foram calculadas em nossos dados populacionais usando o programa DnaSP ver. 4.10.9 (Rozas *et al.* 2003). Redes de haplótipos foram inferidas no programa TCS 2.1 (Clement *et al.* 2000) e foram manualmente convertidas a formatos de árvore Newick para algumas das análises de detecção de seleção. Os testes de neutralidade D de Tajima (Tajima 1989b) e D e F de Fu e Li (Fu & Li 1993) foram realizados no DnaSP usando seqüências de *Bactrocera dorsalis* como grupo externo, por ser a espécie mais próxima de *Anastrepha* cujo *dsx* se apresentava completo. Para a análise de dados populacionais, buscamos testar se mutações que levam a alterações de aminoácidos tem maior ou menor probabilidade de deixar descendentes na população. Se esta região está evoluindo neutramente, não esperamos encontrar diferenças entre tais probabilidades quando comparamos mutações sinônimas e não-sinônimas. Caso a região esteja sujeita à seleção purificadora, esperamos que a maioria das mutações não-sinônimas deixe menos descendentes do que mutações sinônimas (Yang 2002). Por outro lado, caso esta região esteja sujeita à seleção positiva, podemos encontrar mutações vantajosas que estejam em alta frequência, e com isso, podemos ter que tais mutações deixariam mais descendentes do que outras mutações (Fay & Wu 2005). Realizamos dois tipos de testes estatísticos para tentar detectar tais sinais, em que consideramos apenas mutações que estavam presentes pelo menos em dois ou mais haplótipos, uma vez que estamos buscando contrastar probabilidades diferenciadas de sobrevivência, devemos considerar apenas mutações que já sobreviveram tempo suficiente para se reproduzir. Mutações que estão presentes apenas em *singletons* não passaram ainda pelo teste evolutivo da sobrevivência e reprodução no tempo (Templeton 1996). Um dos testes estatísticos fez uso de testes exatos de Fisher em tabelas de contingência 2 x 2 (Templeton 1996), que é uma versão mais refinada do teste de McDonald & Kreitman (1991), uma vez que avalia também polimorfismos intraspecíficos e não apenas faz um contraste entre mutações polimórficas e fixadas entre linhagens. Estas tabelas foram feitas para se contrastar a distribuição de substituições sinônimas e não-sinônimas em haplótipos encontrados nas pontas ou no interior (haplótipos jovens ou antigos). Contudo, consideramos isso de forma hierárquica, contrastando classes cada vez maiores de mutações, por exemplo, mutações que deixaram dois, três, quatro descendentes e assim sucessivamente. Alternativamente, realizamos o teste não-paramétrico de Mann-Whitmann, que neste caso testa se mutações sinônimas e não-sinônimas apresentam um número distinto de descendentes em nossas amostras através de uma análise de ranqueamento.

Resultados

1. Seleção positiva em *fru*

Amplificamos aproximadamente 802 pb abrangendo um íntron de 81 pb e 721 pb do primeiro éxon da região conectora do *fru* para 50 indivíduos de três espécies do gênero *Anastrepha* (*A. fraterculus*, *A. obliqua* e *A. sororcula*). Selecionamos três destes indivíduos, sendo um de cada uma das espécies acima mencionadas, para o estudo macroevolutivo da seleção na região conectora de *fru*.

A variação no regime seletivo sobre a região conectora de *fru* foi investigada pela análise da razão das taxas de substituições não-sinônimas por sinônimas (ω) (Yang *et al.* 2005; Zhang *et al.* 2005). Os parâmetros dos modelos nulos M1a e MA restrito e do modelo alternativo MA estão apresentados na Tabela B2. O teste que contrasta os modelos M1a-MA rejeitou o modelo nulo de restrição seletiva ($0 < \omega \leq 1$) indicando que os ramos *background* e *foreground* (Figura B1) divergiram nesta região por meio de uma restrição seletiva mais branda ou por seleção positiva (Tabela B3). Para discriminar as duas hipóteses, contrastamos os modelos MA restrito e MA, em cujo contraste o modelo nulo foi novamente rejeitado (Tabela B3).

Tabela B2. Parâmetros estimados e valores de logaritmo da máxima verossimilhança para os modelos M1a, MA e MA restrito.

Modelo	Parâmetros			LnL
M1a <i>Modelo nulo</i>	$\omega_0 = 0,065$ $\omega_1 = 1,000$	$p_0 = 0,870$ $p_1 = 0,130$		-3647,925
MA restrito ($\omega_2 = 1$, fixado) <i>Modelo nulo</i>	$\omega_0 = 0,050$ $p_0 = 0,653$	$\omega_2 = 1,000$ $p_1 = 0,096$	$(p_2 + p_3) = 0,252$	-3629,574
MA <i>Modelo alternativo</i>	$\omega_0 = 0,053$ $p_0 = 0,679$	$\omega_2 = 82,666$ $p_1 = 0,093$	$(p_2 + p_3) = 0,228$	-3627,665

LnL = logaritmo da máxima verossimilhança; ω_0 = valores de dN/dS para sítios com $0 < \omega < 1$; ω_1 = valores de dN/dS fixados a 1; ω_2 = valores de dN/dS para sítios $\omega > 1$ e correspondem apenas a sítios no ramo *foreground*.

Tabela B3. Comparação por LRT dos modelos M1a, MA e MA restrito.

Contraste	LRT	G.L.	Probabilidade - χ^2	Sítios selecionados positivamente
M1a x MA	40,52021	2	$p < 10^{-8}$	16, 25, 31, 49, 61, 64, 65,
MA restrito x MA	3,816756	1	$p < 0.05$	83, 105, 111, 131, 164, 167, 188, 194, 195,

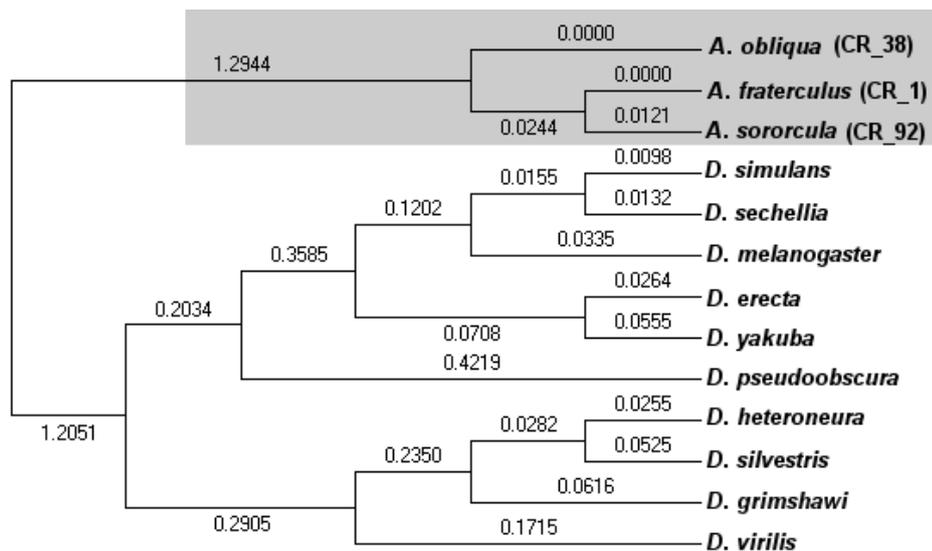


Figura B1. Árvore filogenética gerada por Máxima Verossimilhança no software Phylip ver 3.68. O ramo *foreground* está destacado em cinza.

A rejeição do modelo MA restrito indica que parte da região conectora de *fru* se diferenciou entre as linhagens mediante seleção positiva. Assim, por meio da análise Bayesiana (BEB) foram estimados 16 sítios com probabilidades posteriores de possuírem $\omega > 1$ de 0,95 (Tabela B3 e Figura B2) que se concentram em um trecho que vai da extremidade 5' até a região mediana do fragmento estudado (códon 16 a 195), sendo que na porção final há predominância de sítios sob restrição seletiva. Os dados do programa TreeSAAP (McClellan *et al.* 2005) indicam que todas as 31 propriedades físico-químicas analisadas mostraram desvios significativos da neutralidade a um limite de 0,05 no teste de ajustamento (Tabela B4). Todavia, apenas três destas propriedades mostraram escores *z* significativamente positivos nas categorias para mudanças radicais (6 a 8): Tendências para α -hélices (P_α), Hidropatia (*H*) e Peso molecular (M_w). A análise por janelas deslizantes mostrou que os aminoácidos selecionados positivamente concentraram-se na extremidade 5' da região conectora de *fru* (Figura B3).

Tabela B4. Detecção de sítios sob seleção positiva desestabilizadora nas propriedades físico-químicas dos aminoácidos do éxon da região conectora de *fru*.

Propriedade físico-química	teste de ajustamento (expectativa neutra)	categoria desestabilizadora (6, 7 e 8)	Escore z
Tendências para a-hélices ($P\alpha$)	36,498***	8	1,975*
Average number of surrounding residues (N_s)	76,820***	-	-
Tendências de estrutura beta ($P\beta$)	67,631***	-	-
<i>Bulkiness</i> (B_l)	40,169***	-	-
<i>Buriedness</i> (B_r)	47,688***	-	-
Índice cromatográfico (R_f)	65,981***	-	-
Tendências de helicoidização (P_c)	82,908***	-	-
Composição (C)	52,020***	-	-
Compressibilidade (K^0)	36,874***	-	-
Constante de equilíbrio (ionização do COOH) (pK')	33,814***	-	-
Área de contato helicoidal (C_a)	41,651***	-	-
Hidropatia (H)	92,128***	6	4,972***
Ponto isoeletrico (pH_i)	56,298***	-	-
<i>Long-range non-bonded energy</i> (E_l)	38,021***	-	-
<i>Mean r.m.s. fluctuation displacement</i> (F)	57,826***	-	-
Volume molecular (M_v)	48,275***	-	-
Peso molecular (M_w)	53,230***	6	3,928***
Hidrofobicidade de consenso normalizada (H_{nc})	29,042***	-	-
Volume específico parcial (V^0)	22,173**	-	-
Requerimento polar (P_r)	77,854***	-	-
Polarity (P)	117,046***	-	-
Capacidade de estar na extremidade carboxi de uma alfa-hélice (α_c)	57,177***	-	-
Capacidade de estar no meio de uma alfa-hélice (α_m)	40,309***	-	-
Capacidade de estar na extremidade amino de uma alfa-hélice (α_n)	35,782***	-	-
Índice refrativo (μ)	17,263*	-	-
<i>Short and medium range non-bonded energy</i> (E_{sm})	73,460***	-	-
Solvent accessible reduction ratio (R_a)	16,175*	-	-
Hidrofobicidade circundante (H_p)	90,284***	-	-
Hidrofobicidade de transferência termodinâmica (H_t)	90,401***	-	-
Energia total não ligada (E_t)	32,042***	-	-
Turn tendencies (P_t)	45,706***	-	-

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

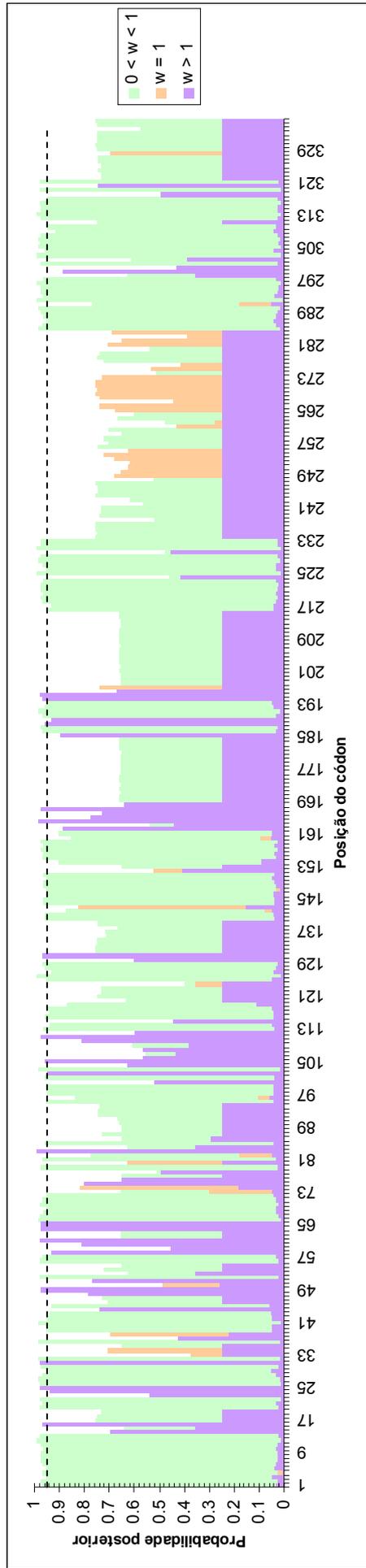


Figura B2. Probabilidades posteriores de sítios sob seleção purificadora ($0 < \omega < 1$), positiva ($\omega > 1$) e neutros ($\omega = 1$) na região conectora de *fru*. A linha tracejada indica o limite de 0,95.

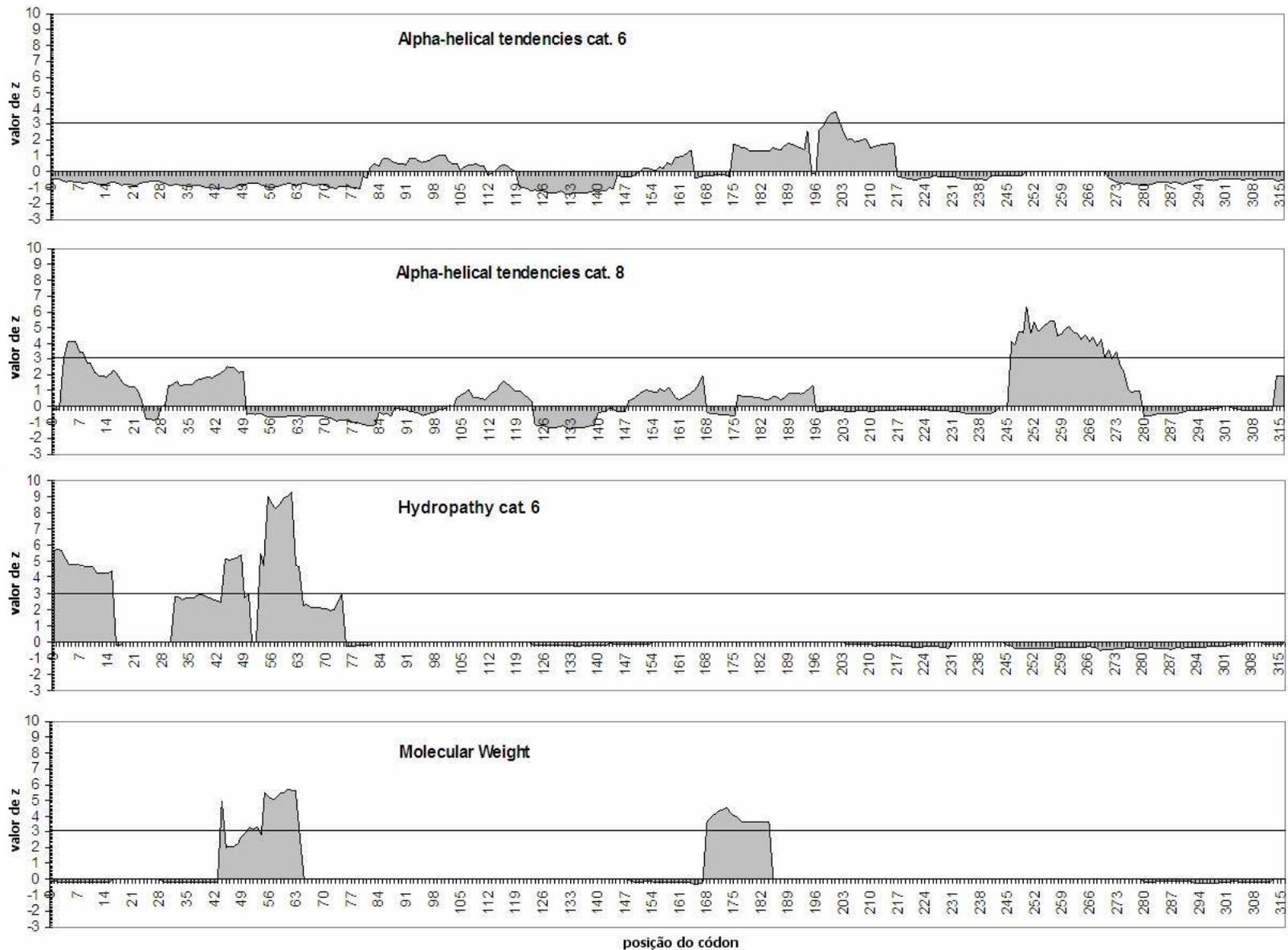


Figura B3. Estimativa de áreas da região conectora de *fru* sob seleção positiva destabilizadora estimados por janelas deslizantes. Os valores da posição dos códons representam os valores iniciais dos intervalos de 20 códons. A linha horizontal representa o limite de significância de 0,001 (com correção de Bonferroni)

2. Detecção de seleção natural em dados populacionais de baixa divergência em *fru*

2.1 Relação filogenética entre os haplótipos e diversidade genética no íntron e região conectora de *fru*

Utilizamos a parcimônia estatística para estabelecer uma rede de haplótipos entre as seqüências populacionais obtidas para a região intrônica e conectora (Figura B4 e B5). Estão resumidas na Tabela B5 as medidas de variabilidade nucleotídica para toda a região seqüenciada e cada domínio funcional. Considerando apenas a região conectora das três espécies, foram encontrados 167 sítios polimórficos, sendo que nove dos quais tiveram mais de um variante nucleotídico presente entre os haplótipos. Por outro lado, no íntron foram encontrados 28 sítios polimórficos e seis com mais de um variante nucleotídico (Tabela B5). A comparação entre as seqüências também revelou a existência de 38 haplótipos diferentes para o íntron e 94 haplótipos para a região conectora, sendo a maioria composta por haplótipos únicos promovendo, dessa forma, elevados índices de diversidade haplotípica tanto para o íntron quanto para a região conectora. Os índices de diversidade tanto para o íntron quanto para a região conectora foram relativamente elevados, com estes últimos mostrando os níveis mais altos (Tabela B5). Apesar da grande variabilidade encontrada, não encontramos sinais significativos de recombinação através dos testes implementados no RDP, e a topologia da árvore tampouco é congruente com tal inferência.

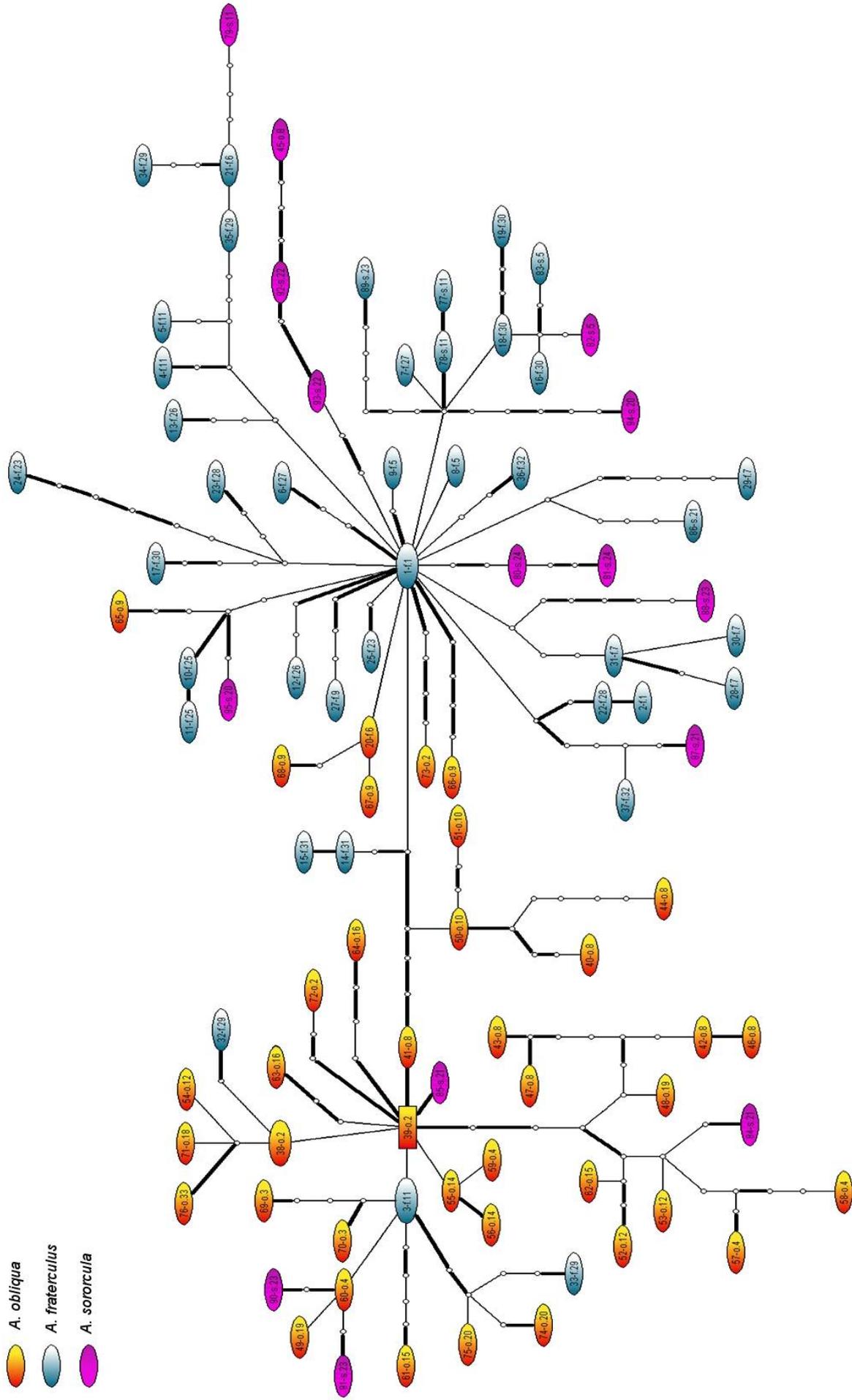


Figura B4. Rede de haplótipos da região conectora de *fru* inferida por parcimônia estatística no TCS ver. 1.21. As linhas espessas representam mutações não-sinônimas e as mais finas sinônimas. Os círculos vazios correspondem a haplótipos inferidos.

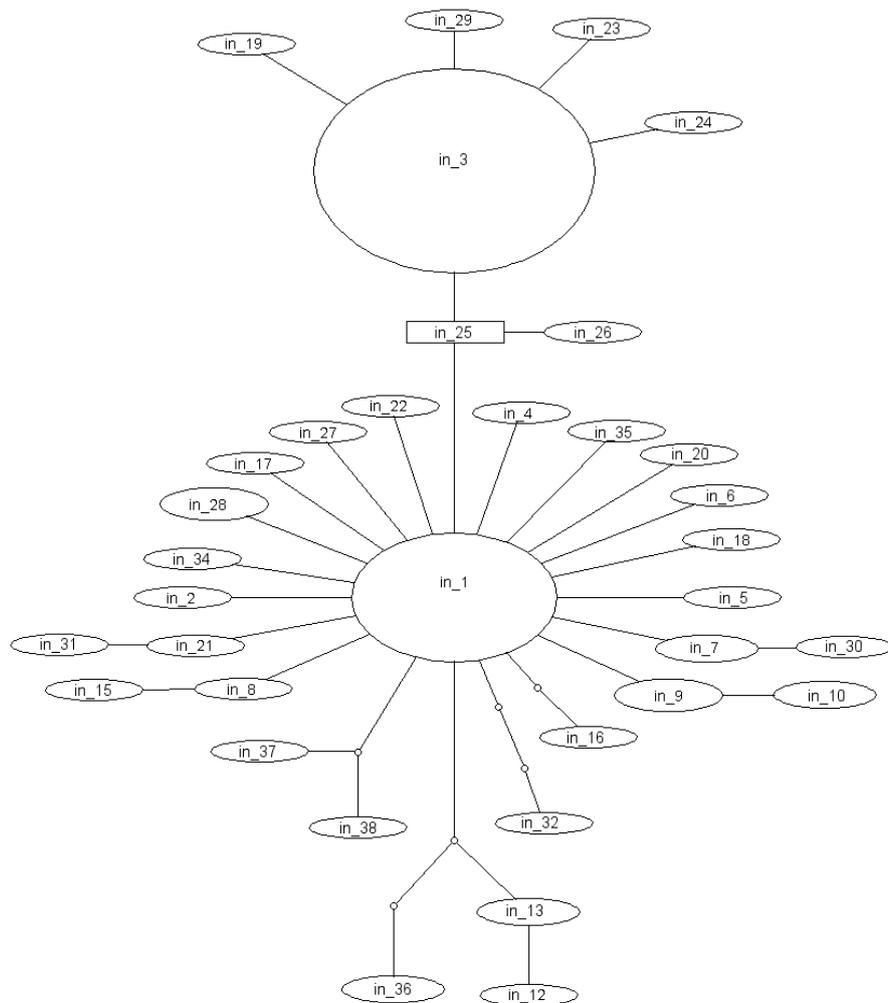


Figura B5. Rede de haplótipos do íntron adjacente à região conectora inferida por parcimônia estatística no TCS ver. 1.21.

Tabela B5. Estimativa da diversidade genética do primeiro éxon da região conectora e íntron.

Espécies	N	h	Hd	S	S _y	N _{Sy}	sg	m ₃	π	π _s	π _a	θ
<i>A. fraterculus</i>	37	36	0,998 ± 0,007	82	42	34	78	4	0,011	0,0269	0,006	0,028 ± 0,009
<i>A. obliqua</i>	41	40	0,999 ± 0,006	80	41	39	77	3	0,013	0,0260	0,008	0,027 ± 0,008
<i>A. sororcula</i>	19	19	1,000 ± 0,017	62	39	26	60	2	0,014	0,0338	0,007	0,025 ± 0,009
<i>Anastrepha</i> sp.	97	94	0,999 ± 0,002	167	83	78	157	9	0,013	0,0299	0,007	0,047 ± 0,009
<i>A. fraterculus</i>	37	22	0,922 ± 0,030	17	-	-	17	0	0,027	-	-	0,057 ± 0,021
<i>A. obliqua</i>	41	11	0,429 ± 0,097	8	-	-	8	0	0,009	-	-	0,026 ± 0,012
<i>A. sororcula</i>	19	13	0,936 ± 0,037	16	-	-	15	1	0,039	-	-	0,064 ± 0,026
<i>Anastrepha</i> sp.	97	38	0,834 ± 0,032	28	-	-	22	6	0,026	-	-	0,077 ± 0,023

As informações na área sombreada referem-se ao íntron e na outra ao éxon (Região conectora). N = número de seqüências; h = número de haplótipos; Hd = diversidade haplotípica; S = Número de sítios polimórficos; S_y = Número de mudanças sinônimas; N_{Sy} = número de mudanças não-sinônimas; sg = singletons; m₃ = sítios com mais de 3 variantes; π = diversidade nucleotídica; π_s = diversidade nucleotídica sinônima; π_a = diversidade nucleotídica não-sinônima; θ = teta de Waterson.

2.2 Espectro de frequência de sítios e teste de contingência.

Considerando todas as seqüências em único conjunto, todos os testes de seleção apresentaram valores significativamente negativos, tanto para a região conectora quanto para o íntron (Tabela B6). Os mesmos testes foram realizados com conjuntos separados de *A. fraterculus*, *A. obliqua* e *A. sororcula*. Para a região conectora, todos os testes, à exceção do D de Fu e Li, continuaram mostrando valores significativamente negativos, ao passo que para o íntron, somente o D de Tajima para *A. obliqua* e o H de Fay e Wu para *A. fraterculus* e *A. sororcula*, continuaram significativamente negativos. (Tabela B6). Note-se, porém, que apesar de continuarem negativos para a região do íntron, os valores de H de Fay e Wu foram bem menores em relação aos encontrados para a região conectora. Contrário ao observado nos testes de neutralidade, a hipótese de evolução neutra nesta região do *fru* não pôde ser rejeitada para este grupo de espécies quando o teste de contingência de Templeton foi utilizado (Tabela B7), nem tampouco no teste de Mann-Whitney considerando todas as substituições presentes em mais de um indivíduo (Tabela B8). Por outro lado, se ao invés de considerarmos apenas as mutações presentes em mais de um indivíduo, considerarmos as presentes em pelo menos três, este resultado é significativo ($p < 0,03$).

Tabela B6. Testes de neutralidade: região conectora e íntron.

	Espécies	Fu and Li D	Fu and Li F	Fay and Wu H	Fay and Wu normalized H	Tajima D
<i>fruitless</i> Reg. conectora	<i>Anastrepha</i> sp.	-4,578*	-4,334*	-81,514**	-4,266	-2,435**
	<i>A. fraterculus</i>	-1,584 ^a	-2,197	-46,276**	-4,079	-2,244**
	<i>A. obliqua</i>	-2,355*	-2,590*	-30,660**	-2,818	-1,924*
	<i>A. sororcula</i>	-1,596 ^a	-2,038 ^a	-24,883**	-2,555	-1,882*
<i>fruitless</i> Íntron	<i>Anastrepha</i> sp.	-3,528*	-3,412*	-8,766*	-2,862	-1,989*
	<i>A. fraterculus</i>	-1,951 ^a	-2,157 ^a	-4,532*	-1,944	-1,739 ^a
	<i>A. obliqua</i>	-1,181	-1,389	-2,813*	-2,104	-1,839*
	<i>A. sororcula</i>	-1,205	-1,497	-2,164	-0,836	-1,476

* $P < 0,05$; ** $P < 0,01$; # $P < 0,001$
a não significativo $0,10 > P > 0,05$

Tabela B7. Análise de contingência de mutações sinônimas e não-sinônimas versus posição na rede de haplótipos (interior ou de exterior).

	Não-sinônima	Sinônima
Exterior	86	93
Interior	23	44

Tabela B8. Contagem de indivíduos gerados por mutações sinônimas e não-sinônimas internas

Mutação interna	Tipo de mutação interna Sinônima ou Não-sinônima	Número de indivíduos originados a partir da mutação interna
2	Sinônima	2
3	Sinônima	2
7	Sinônima	2
12	Sinônima	2
21	Sinônima	2
24	Sinônima	2
26	Sinônima	2
32	Sinônima	2
32	Sinônima	2
32	Sinônima	2
33	Sinônima	2
34	Sinônima	2
34	Sinônima	2
17	Sinônima	3
17	Sinônima	3
17	Sinônima	3
1	Não-sinônima	3
5	Sinônima	4
15	Sinônima	4
31	Sinônima	4
9	Não-sinônima	4
18	Sinônima	5
18	Sinônima	5
30	Sinônima	5
6	Sinônima	6
6	Sinônima	6
14	Sinônima	6
29	Sinônima	6
6	Não-sinônima	6
10	Sinônima	9
23	Sinônima	9
13	Sinônima	10
28	Sinônima	11
27	Sinônima	12
25	Não-sinônima	15
4	Não-sinônima	16
22	Não-sinônima	24
20	Sinônima	25
20	Não-sinônima	25
11	Sinônima	26
19	Sinônima	27
16	Sinônima	33
8	Não-sinônima	78

* Quando os dois tipos de mutações estão presentes em um mesmo ramo interno, cada uma é contada independentemente.

2.3 Detecção de seleção por máxima verossimilhança

Como analisado nos dados interespecíficos, a ação da seleção sobre a região conectora de *fru* foi investigada novamente nos dados populacionais pela análise da razão das taxas de substituição não-sinônima pela sinônima (Yang *et al.* 2005). A Tabela B9 mostra que os modelos nulos de restrição seletiva (sítios com $\omega < 1$) não puderam ser rejeitados. Em concordância a este resultado, a análise Bayesiana (BEB) indicou a ação preponderante de seleção purificadora nesta região (Figura B6). Por esta figura percebe-se que o ω médio nesta região varia em torno de 0,3 e são observados poucos sítios neutros ($\omega \sim 1$). Apenas os códons 74 e 202 apresentaram $\omega > 1$, com $\omega = 1,18$ e $1,45$, respectivamente. Todavia a ação de seleção positiva não pôde ser confirmada, visto que as probabilidades posteriores destes códons possuírem $\omega > 1$ estão abaixo de 0,95.

Tabela B9. Teste de heterogeneidade de ω em sítios. Valores de máxima verossimilhança, estimativa de parâmetros e LRTs para a região conectora de *fru*.

modelo nulo	parâmetros	$\ln L_0$	modelo alternativo	parâmetros	$\ln L_1$	LRT (G.L., p)
M1a (neutro)	$\omega_0 = 0,27$ $\omega_1 = 1$ (fixado) $p_0 = 0,924$ $p_1 = 0,076$	-2619,996	M2a (seleção positiva)	$\omega_0 = 0,27$ $\omega_1 = 1$ $\omega_2 = 1$ $p_0 = 0,924$ $p_1 = 0,056$ $p_3 = 0,020$	-2619,408	$1,4 \times 10^{-5}$ (2, p = 0,99)
M7 (β) 10 categorias	$p = 2,030$ $q = 4,332$	-2619,334	M8 (β e $\omega > 1$) 11 categorias	$p_0 = 1$ $p = 2,030$ $q = 4,332$ $p_1 = 0$ $\omega_2 = 1$	-2619,334	0,00 (2, p = 1)

ω_0 = estimativa de $dN/dS < 1$; ω_1 = estimativa de $dN/dS = 1$; ω_2 = estimativa de $dN/dS > 1$; p_0 = proporção de sítios com $0 < \omega < 1$; p_1 = proporção de sítios com $\omega = 1$; p e q são parâmetros da distribuição β .

2.4 Detecção de seleção por desvios em propriedades físico-químicas dos aminoácidos

Por esta estratégia as substituições de aminoácido encontradas foram analisadas considerando-se suas alterações em 31 propriedades físico-químicas, as quais estão resumidas na Tabela B10. Destas propriedades, 20 apresentaram um desvio significativo da expectativa neutra e destas propriedades pré-selecionadas, quatro desviaram-se significativamente na direção de uma mudança radical com excesso de diversificação, ou seja, apresentaram sinais de seleção positiva desestabilizadora. Dentre estas propriedades sob seleção desestabilizadora, três estão envolvidas na conformação estrutural da proteína: Compressibilidade (K^0), poder de estar numa extremidade amino (α_n) e carboxi (α_c) de uma α -hélice. Apenas uma destas propriedades está envolvida em

mudanças químicas: hidrofobicidade local (H_p). A maior parte das mudanças em propriedades estruturais ocorreram na extremidade 3' do segmento, como revelado pela análise por janelas deslizantes (Figura B7). Apesar de a hidrofobicidade mostrar valores de z significativos, tais valores tornaram-se limítrofes ao limite de significância de 0.01 na análise por janelas deslizantes. Assim, não foi possível estimar com confiança qual região apresenta desvios para esta propriedade.

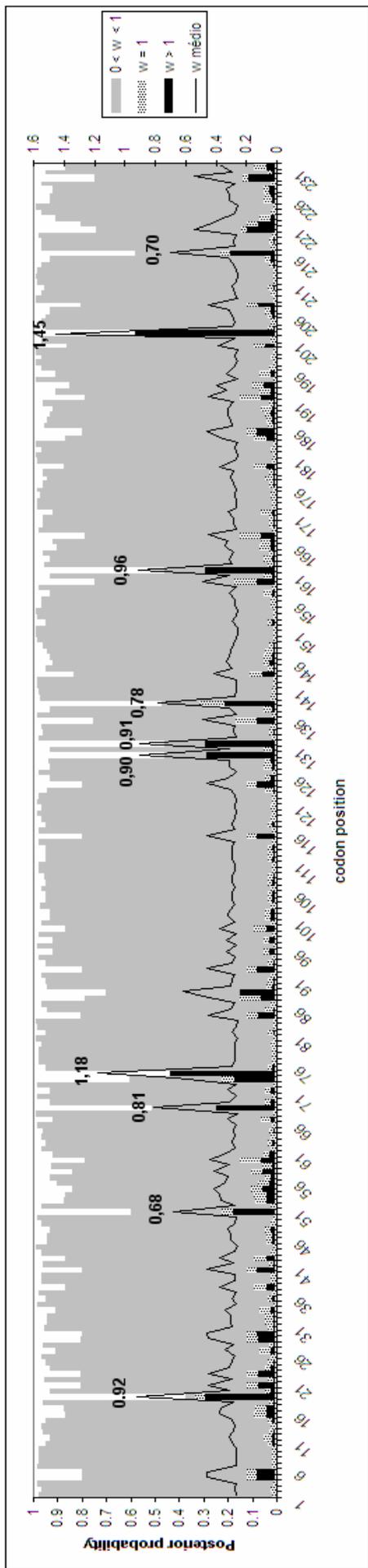


Figura B6. Probabilidades posteriores de sítios em f_{ru} nas oito categorias de ω estimadas pela análise Bayes Empirical Bayes sob o modelo M8. Os valores no interior do gráfico indicam os médios dos respectivos códon. O sítio com maior valor de ω foi o 202 com $\omega = 1,20$.

Tabela B10. Valores da análise de desvios em propriedades físico-químicas estimados no TreeSAAP.

Propriedades físico-químicas	Teste de ajustamento (expectativa neutra)	Categoria desestabilizadora (6,7 e 8)	Escore z
Tendências de estrutura beta (P_β)	17,615*	-	
Índice cromatográfico (R_f)	17,984*	-	
Tendências de helicoidização (P_c)	18,748*	-	
Composição (C)	18,811**	-	
Compressibilidade (K^0)	18,238*	7	2,277*
Constante de equilíbrio (ionização do COOH) (pK')	20,619**	-	
Área de contato helicoidal (C_a)	32,312***	-	
Hidropatia (H)	14,393*	-	
Ponto isoelétrico(pH_i)	24,670***	-	
<i>Mean r.m.s. fluctuation displacement</i> (F)	20,331**	-	
Volume molecular (M_v)	15,734*	-	
Peso molecular (M_w)	26,338***	-	
Volume específico parcial(V^0)	17,734*	-	
Capacidade de estar na extremidade carboxi de uma alfa-hélice (α_c)	21,462**	6, 7	2,502**, 1,729*
Capacidade de estar no meio de uma alfa-hélice (α_m)	27,203***	-	
Capacidade de estar na extremidade amino de uma alfa-hélice (α_n)	25,639***	8	2,971**
<i>Short and medium range non-bonded energy</i> (E_{sm})	17,239*	-	
Hidrofobicidade circundante (H_p)	16,851*	6	2,160*
Hidrofobicidade de transferência termodinâmica (H_t)	17,076*	-	
<i>Turn tendencies</i> (P_t)	39,276***	-	

* p< 0,05; ** p< 0,01; *** p< 0,001

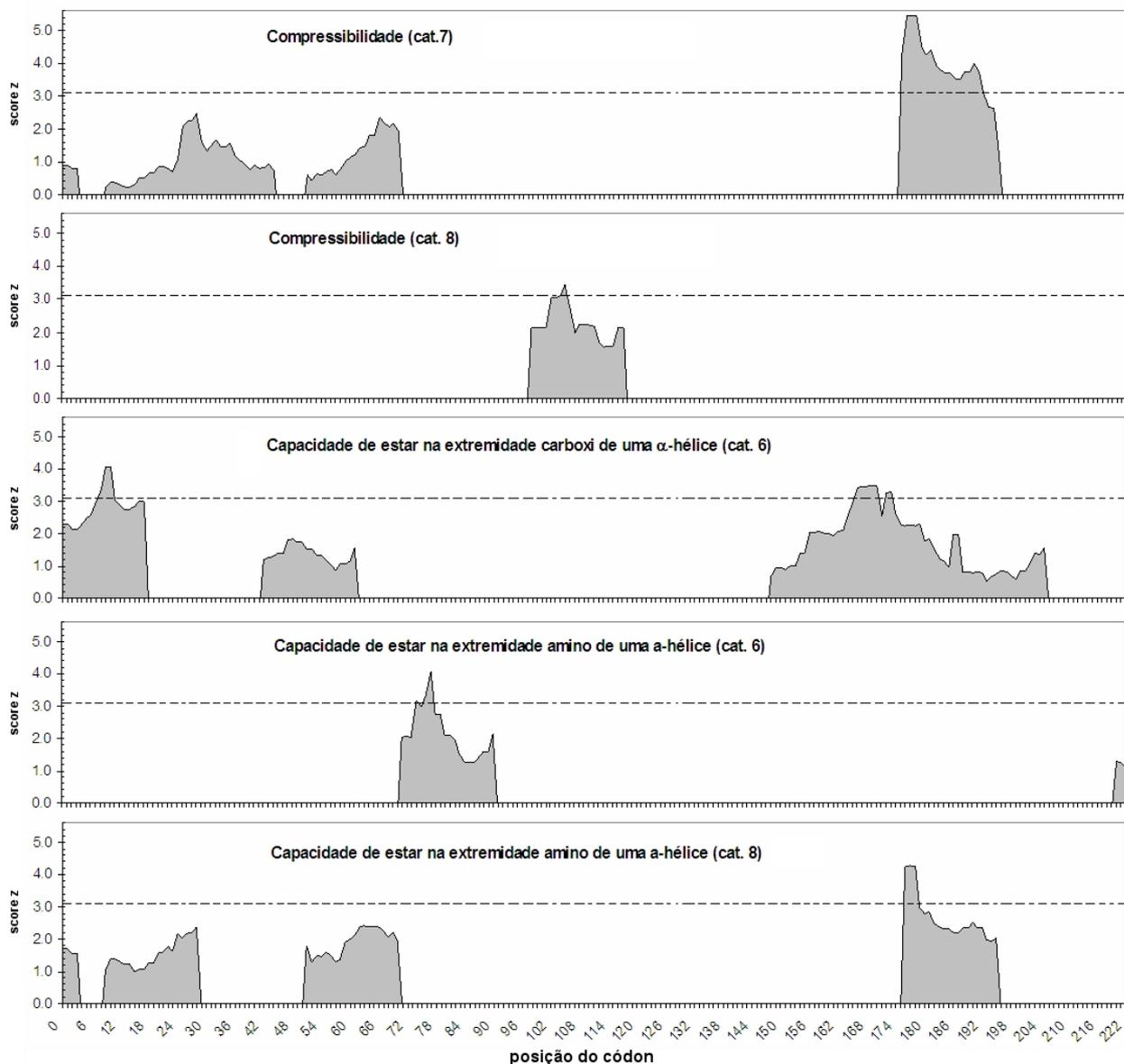


Figura B7. Estimativa de áreas da região conectora de *fru* sob seleção positiva desestabilizadora estimadas por janelas deslizantes. Os valores das posições de códon representam os valores iniciais dos intervalos de 20 códon. A linha tracejada representa o limite de significância de 0,001 com correção de Bonferroni.

Discussão

Os genes relacionados à reprodução estão entre os genes que apresentam as mais elevadas taxas de divergência entre espécies. (Swanson & Vacquier 2002b; Cho *et al.* 2006; Chapman & . 2008). Uma comparação de sequências do gene *fru* de insetos de linhagens evolutivas bem distintas demonstrou que embora os domínios de dimerização (BTB) e de ligação ao DNA (*Zinc-finger*) sejam muito conservados ao longo da linhagem dos insetos (Gailey *et al.* 2006), a região conectora e a extremidade amino da proteína mostraram elevada divergência entre as espécies estudadas. Tais autores sugeriram que tais regiões ou não desempenhariam funções primordiais para a proteína e estariam divergindo por um relaxamento da restrição seletiva, ou tal região conteria informações para o desenvolvimento do comportamento de corte de machos que seriam específicas para cada espécie, daí sua alta divergência entre os grupos estudados.

No presente trabalho realizamos um teste direto do contraste entre a hipótese de evolução por restrição seletiva mais branda contra a de evolução por seleção positiva nesta região conectora analisando modelos evolutivos distintos no PAML para dados macroevolutivos, em que contrastamos moscas dos grupos Drosophilidae e Tephritidae, este último representado por três espécies do gênero *Anastrepha*. Apesar da informação prévia da região conectora ser bastante divergente, a primeira hipótese a ser testada era a de que esta região estivesse realmente evoluindo de uma forma menos restritiva (ou seja, mais neutramente). De fato, o primeiro contraste de modelos evolutivos no teste de heterogeneidade de valores de ω entre ramos e sítios (modelos M1a e MA) indicou que esta região do gene *fru* está sendo afetada por seleção, embora tal teste não seja capaz de discriminar entre seleção positiva ou uma seleção purificadora mais relaxada. Já o segundo teste, contrastando os modelos MA restrito e MA, avalia diretamente a presença de regiões que tenham evoluído por seleção positiva e foi utilizado para contrastar as duas hipóteses de evolução de sítios na região conectora. Tal teste demonstrou, inequivocamente, que parte da diferenciação entre as linhagens *background* (Drosophilidae) e *foreground* (Tephritidae) deu-se por seleção positiva agindo ao menos em 16 sítios ao longo desta região conectora (Figura B2).

Dos 16 sítios que foram detectados pelo uso do PAML, 14 estavam inseridos em regiões onde havia sido detectada seleção positiva desestabilizadora pelo TreeSAAP utilizando estes mesmos dados. Os códons 16 e 25, por exemplo, estão localizados na região onde se estimou um desvio significativo para diversificação na propriedade *Tendência para α -hélice*. Provavelmente o desvio detectado nesta propriedade deveu-se à fixação de aminoácidos físico-quimicamente diferentes nestas duas posições. O mesmo valendo para outros sítios como 31, 49, 61, 64, 65 e 83 para a propriedade *hidropatia*. Dependendo do aminoácido alterado, uma mudança

radicalmente diferente em uma propriedade físico-química pode levar à mudança em outra propriedade, visto que tais propriedades podem estar associadas. Por exemplo, o aminoácido tirosina pode ser classificado como um aminoácido *polar* e ao mesmo tempo um aminoácido de grande *volume* (Sainudiin *et al.* 2005). Por outro lado, os sítios 16, 25 e 31, situaram-se na região onde *Hidropatia* apresentou desvios, porém tais sítios se envolveram em desvios em *peso molecular*. Além disso, os sítios 188, 194 e 195 associaram-se apenas a desvios em *peso molecular*.

Apesar da abordagem por estudo da razão dN/dS ser uma estatística muito utilizada para a detecção de sinais de seleção positiva, há algumas limitações em sua capacidade de detectar tais sinais. Muito de sua habilidade em detectar tal sinal pode ser perdida ou mesmo conclusões equivocadas podem ser tiradas se esta técnica for aplicada a regiões consideravelmente conservadas ou particularmente se não considerarmos os efeitos que as mudanças de aminoácidos promoverão nas propriedades físico-químicas das proteínas (McClellan *et al.* 2005; Porter *et al.* 2006). Por exemplo, considerando-se apenas o excesso de substituições não-sinônimas pode-se concluir que a região estudada apresenta sinais de seleção positiva. Contudo, se as substituições não-sinônimas observadas levarem a mudanças para aminoácidos quimicamente similares, o que se observa na realidade é a ação de seleção purificadora para manutenção da estrutura e/ou função química da região que apresenta tais mutações. Isso pode ser observado nos códons 164 e 167, os quais foram indicados como sítios que divergiram por seleção positiva, porém não promoveram desvios significativos em propriedades físico-químicas radicalmente diferentes. Logo, a fixação de substituições não-sinônimas nestes sítios indica muito mais uma restrição seletiva favorecendo substituições conservativas.

Por outro lado, ao considerar as mudanças nas propriedades físico-químicas o método de McClellan *et al.* (2005) permitiu a identificação de sítios sob seleção positiva mesmo em regiões relativamente conservadas. Isso pode ser notado no trecho entre os códons 249 e 284, no meio da BEB, onde observou-se um agrupamento de sítios preponderantemente neutros ($\omega = 1$) em que o TreeSAAP mostrou desvios significativos para mudanças radicais na *Tendência para α -hélice*.

Outra restrição das análises da razão ω , que na verdade é partilhada pelo TreeSAAP, é que muito de sua habilidade em detectar sinais de seleção positiva pode ser perdida e conclusões errôneas podem ser inferidas se esta técnica for aplicada a regiões consideravelmente conservadas ou particularmente se estivermos considerando dados populacionais em que pouca diferenciação existe entre os ramos (McClellan *et al.* 2005; Porter *et al.* 2006; Kryazhimskiy & Plotkin 2008). Tal limitação é consequência da estocasticidade do processo evolutivo, e portanto, linhagens que tiveram pouco tempo de divergência têm também, como consequência, uma pequena amostragem das mutações que ocorreram para se determinar a existência de seleção positiva ou purificadora. Por

este motivo, o uso de ω para a análise de seleção positiva em dados populacionais é em geral desaconselhado (Kryazhimskiy & Plotkin 2008) e devemos utilizar de outras estratégias para se tentar identificar tais padrões de seleção quando há baixos níveis de divergência, como o detectado para dados populacionais.

Padrão seletivo sobre *fru* usando dados de baixa divergência

Mediante a estimativa de ω e considerando apenas os dados populacionais das três espécies de *Anastrepha*, nossos resultados fornecem pouco suporte para a hipótese de seleção positiva na região conectora do *fru*. Muito provavelmente, a baixa divergência entre as espécies aqui analisadas foi responsável por este resultado, uma vez que o poder de detecção destes testes diminui com a redução da divergência entre os taxa (Anisimova *et al.* 2001; Nickel *et al.* 2008), na mesma proporção em que aumenta o problema com a estocasticidade evolutiva. Mesmo o valor global de ω estimado para esta região deve ser considerado com certa cautela, pois para dados de baixa divergência e que ainda contêm polimorfismos ancestrais segregantes, os testes baseados na variação de ω são pouco sensíveis aos sinais deixados pela seleção (Kryazhimskiy & Plotkin 2008). Neste contexto, um valor de $\omega = 0,27$ para esta região pode representar na realidade uma pressão de seleção purificadora bem mais intensa do que o indicado. Isto ocorre pela violação do pressuposto de que toda a variação divergente encontrada entre as seqüências é considerada uma substituição que passou pelo processo de fixação na linhagem por deriva ou seleção (Yang *et al.* 2000). Esta condição é, de fato, violada na comparação de *A. fraterculus*, *A. obliqua* e *A. sororcula*, nas quais polimorfismos ancestrais ainda segregam nas populações.

Ao contrário do observado nas análises por códons, encontramos certa similaridade entre as propriedades físico-químicas detectadas em dados divergentes e pouco divergentes. Apesar de não serem exatamente as mesmas propriedades, elas representam alterações similares na estrutura ou condição química da proteína: P_α versus $\alpha_c + \alpha_n$; H versus H_p ; K^0 versus M_w . Interessantemente, as três propriedades estruturais afetaram áreas aproximadamente congruentes na extremidade 3' da região estudada, indicando uma predisposição desta região para diversificação em sua estrutura.

Novamente os resultados observados em dados de baixa divergência devem ser considerados com cautela, uma vez que ainda não foi desenvolvido um trabalho rigoroso sobre a robustez deste método frente a dados que ainda contêm polimorfismos segregantes, e as restrições condicionadas à estocasticidade evolutiva determinadas aos testes de razão ω aplicam-se igualmente a esta estratégia, uma vez que ela baseia-se na estimativa inicial de sítios ancestrais para

determinar posteriormente se a mudança na propriedade encaixa-se numa categoria de substituição radical ou não. Quando polimorfismos ainda segregam nas populações torna-se difícil estimar qual o estado ancestral através do *baseml*, o qual é utilizado pelo TreeSAAP para realizar esta inferência.

Um dos testes que realizamos considerou as mutações sinônimas e não-sinônimas no contexto da topologia do cladograma através do teste de contingência de Templeton (Templeton 1996). A não rejeição da hipótese nula neste teste indica que esta região ou está evoluindo de forma preponderantemente neutra, ou as informações contidas nos dados não foram suficientes para a detecção de sinais de seleção sobre esta região. Entretanto, deve-se considerar que o poder deste tipo de teste é dependente do número de contrastes definidos pela topologia da árvore, e portanto diminui quando a topologia da árvore apresenta poucas ramificações internas como a encontrada na rede haplotípica da região conectora de *fru*. Este resultado é corroborado pela estratégia não-paramétrica de Mann-Whitney para investigação do padrão de sobrevivência diferencial entre mutações sinônimas e não-sinônimas no tempo evolutivo estabelecida por Sobrinho Jr. e de Brito (*subm.*), que também não foi significativa. Interessantemente, quando consideramos apenas mutações que deixaram mais de dois descendentes (ou seja, 3 ou mais), ao invés de dois ou mais avaliados em *dsx* por Sobrinho Jr. e de Brito (*subm.*), encontramos evidências significantes de que mutações não-sinônimas apresentam-se em mais descendentes do que mutações sinônimas também em *fru* ($p < 0,03$). Desta forma, este teste mais uma vez foi capaz de detectar sinais positivos de seleção.

A comparação conjunta dos testes de neutralidade permitiu-nos avaliar melhor a influência da seleção sobre estas regiões, uma vez que cada um deles é sensível a um aspecto particular do espectro de frequência (Zeng *et al.* 2007). As estatísticas D de Fu e Li e D de Tajima indicaram um excesso de variantes de baixa frequência, porém, a estatística significativamente negativa de H de Fay e Wu mostrou um excesso de mutações derivadas de alta frequência. De acordo com Fay & Wu (2000) um sinal evidente da ocorrência de efeito carona por seleção positiva é o excesso de variantes de baixa e de alta frequência coexistindo em uma população, ao passo que o esperado pela seleção purificadora seria apenas um excesso de variantes de baixa frequência, que foi exatamente o padrão detectado pelo teste de Mann-Whitney de Sobrinho Jr. e de Brito (*subm.*) nestes dados populacionais de *fru*. Diversos processos diferentes, inclusive parâmetros populacionais, podem causar excesso de alelos raros na população, contudo, apenas seleção positiva e modelos demográficos específicos são capazes de explicar um excesso de alelos de alta frequência (Fay & Wu 2000). Assim, a análise conjunta destes testes de neutralidade indica que a região conectora de *fru* de *Anastrepha* apresenta sinais da ação de seleção positiva, se não na própria região em si, em uma região muito próxima.

A comparação dos índices de diversidade entre a região codificante do *dsx* estudada por Sobrinho e de Brito (2009a) e da região conectora de *fru* analisada neste trabalho indica valores similares de diversidade haplotípica em π e θ . Uma única exceção foi a comparação entre π_{dsx} e π_{fru} para *A. fraterculus*, na qual o valor de π_{dsx} foi quase 2 vezes maior que o π_{fru} ; $1,8 \times 10^{-2}$ e $1,1 \times 10^{-2}$, respectivamente. A análise das seqüências da região conectora de *fru* mostrou um elevado nível de diversidade haplotípica, comparável ao encontrado em *dsx* (Sobrinho e de Brito, 2009a). Tal nível elevado de diversidade pode ter como causa a estratégia amostral em que poucos indivíduos foram coletados de diversas localidades diferentes adotada neste trabalho que favorece a ocorrência de haplótipos únicos, embora isto não explique o excesso de heterozigotos encontrados.

A recombinação intragênica também poderia levar a um aumento da variabilidade nesta região, particularmente na diversidade haplotípica e conseqüentemente na heterozigosidade encontrada na amostra. Todavia, os diversos testes realizados para detecção de recombinação não indicaram sua ocorrência na região estudada. Apesar destes testes não terem detectado eventos de recombinação, isso pode também significar que recombinações teriam ocorrido a uma frequência abaixo daquela detectável por estes métodos (Posada & Crandall 2001). É particularmente relevante ressaltar que tais metodologias são também sensíveis à divergência entre as regiões sujeitas à recombinação, de forma que a detecção de recombinação entre haplótipos filogeneticamente próximos é muito mais difícil do que entre haplótipos mais distintos, nos quais muitos eventos de recombinação podem ter ocorrido, facilitando assim sua detecção (Posada & Crandall 2001).

Embora a região conectora tenha apresentado níveis relativamente elevados de diversidade, estes ficaram abaixo daqueles encontrados no íntron adjacente à sua extremidade 5'. Isso sugere que as duas regiões possuem dinâmicas evolutivas diferentes, com o íntron evoluindo de forma menos restringida. Este padrão já era esperado, pois a expectativa para regiões não-codificantes como íntrons é que elas evoluam neutramente, apesar de terem sido descritas algumas regiões intrônicas mais conservadas que regiões codificantes adjacentes (McAllister & McVean 2000). Interessantemente, observamos uma redução nos níveis de diversidade da região intrônica de *A. obliqua* em comparação com as outras espécies estudadas. Naquele grupo os valores de π e θ chegaram a ser 3 vezes menores que os estimados para *A. fraterculus* e *A. sororcula*. Mesmo a Hd mostrou valores 2 vezes menores que o encontrado para as outras duas espécies. Diferentemente do observado nestas outras espécies, o íntron de *A. obliqua* apresentou índices de π e θ similares aos encontrados na região conectora desta mesma espécie.

Se considerarmos que estas regiões intrônicas estariam evoluindo neutramente, níveis de diversidade similares deveriam ser encontrados entre as três espécies aqui estudadas. Um desvio desta expectativa sugeriria que as populações de onde as amostras se originaram passaram por eventos demográficos distintos, ou que esta região está sob diferentes regimes seletivos entre as

espécies. Eventos demográficos deveriam alterar a variação observada no genoma como um todo (Templeton 2006), dessa forma, esta redução na diversidade do íntron em *A. obliqua* deveria ser observada também em regiões próximas a esta no genoma. Neste caso, devemos descartar discrepâncias associadas a desvios causados pela amostragem e por diferentes eventos demográficos, pois as seqüências provêm dos mesmos indivíduos e tal redução na diversidade não foi observada nas respectivas regiões conectoras ligadas aos íntrons e mesmo na região codificante do *dsx* proveniente destes mesmos indivíduos (Sobrinho e de Brito, 2009a).

Devemos, entretanto, considerar as conseqüências de um potencial efeito carona afetando esta região. Esta é uma possibilidade factível, considerando-se a existência de um ramo maior que distingue a maior parte das *A. obliqua* das outras espécies aqui estudadas. Tal ramo pode indicar um *selective sweep* que aconteceu naquela espécie e a diferenciou das outras. Um evento de *selective sweep* leva à redução da variação genética local com o aumento da freqüência de polimorfismos que estejam a ele associados e dependendo de sua intensidade pode até levar à fixação total na região (Fay & Wu 2005). Após tal redução, leva um certo tempo até que a variação genética se recupere, pelo acúmulo de mutações aleatórias. Quando apenas a deriva e a mutação contribuem para o aumento desta diversidade, levam em média $4N$ gerações para que esta diversidade seja recuperada aos níveis do equilíbrio de Fisher-Wright, enquanto que se existe seleção purificadora levará um tempo maior. Contudo, quando existe seleção positiva agindo na região, alelos que são favorecidos irão mais rapidamente recuperar suas altas freqüências alélicas, o que irá também afetar o nível médio de heterozigosidade da região sob seleção positiva. Portanto, a existência de um *selective sweep* terá um efeito mais prolongado no nível de variação genética em regiões sob seleção purificadora, menos prolongado em regiões sob neutralidade e ainda menor em regiões sujeitas a seleção positiva (desde que esta não seja forte o suficiente para levar a novos eventos de fixação). Dessa forma, o fato de termos encontrado uma variabilidade genética similar na região codificadora nas três espécies e uma redução na variação do íntron de *A. obliqua*, sugere que um *selective sweep* afetou tal espécie de forma que a variação intrônica ainda não está recuperada, enquanto a variação na região codificadora, sujeita à seleção positiva, já recuperou, ou manteve, valores similares aos das outras espécies. A topologia do cladograma gerado para os dados de íntron é condizente com esta hipótese ao apresentar um excesso de alelos raros em uma estrutura do tipo estrela. Os valores de H de Fay e Wu muito menores na região codificadora que os estimados para o íntron, somente confirmam esta hipótese.

A Figura B4 revela claramente a presença de diversos haplótipos compartilhados entre as espécies, principalmente entre *A. fraterculus* e *A. sororcula*, com conexão entre si por poucos passos mutacionais. Isto sugere a inexistência de marcadores específicos para cada espécie, apesar de haver uma separação por alguns passos mutacionais entre um grande número de *A.*

obliqua das demais espécies, a qual praticamente polariza a rede haplotípica em dois grandes grupos. Do lado esquerdo da rede temos um aglomerado de haplótipos oriundos principalmente de *A. obliqua* e do lado direito o grupo é formado principalmente por *A. fraterculus* e *A. sororcula*. Apenas 7 haplótipos não pertencentes ao grupo de *A. obliqua* foram encontrados neste agrupamento e 5 haplótipos de *A. obliqua* participam do agrupamento de *A. fraterculus* e *A. sororcula*. Este é um padrão muito semelhante ao encontrado na resolução da árvore filogenética entre duas espécies intimamente relacionadas de *Gryllus* (Andrés *et al.* 2008). Tanto neste trabalho quanto em nosso, uma proteína relacionada a características reprodutivas, apresentou uma resolução filogenética maior que a de outros genes. Por exemplo, no caso do grupo de espécies *fraterculus*, marcadores mitocondriais foram incapazes de estabelecer uma monofilia, mesmo que tênue entre os membros de deste grupo (Smith-Caldas *et al.* 2001). Uma das possíveis explicações para a resolução mais fina trazida pelo *fru* reside no fato desta região estar se divergindo adaptativamente e talvez participando de um evento de isolamento reprodutivo. Se esta hipótese provar-se verdadeira, o favorecimento de poucos variantes neste gene eliminaria mais rapidamente polimorfismos ancestrais ainda presentes entre estas espécies favorecendo a detecção de monofilia, pelo menos deste grupo.

Ao contrário da conclusão de Gailey *et al.* (2006) de que a região conectora de *fru* tenha evoluído por relaxamento na restrição seletiva, concluímos que esta região se diversificou em parte por episódios de seleção positiva. Por este gene estar envolvido no desenvolvimento do comportamento de corte masculino, tais episódios de evolução adaptativa em *fru* pode, muito provavelmente ter se dado por seleção sexual. As conclusões aqui tiradas referem-se principalmente à comparação de táxons que se divergiram há mais tempo, embora tenhamos evidência de um possível evento de *selective sweep*, provavelmente associado à separação de *A. obliqua* de *A. fraterculus* e *A. sororcula*. Apesar de termos encontrado uma associação significativa entre a variação no gene *fru* e a separação de *A. obliqua*, tais eventos podem estar associados apenas por terem ocorrido historicamente em uma mesma época. Dessa forma, para uma resposta conclusiva a respeito do papel efetivo do gene *fru* na separação de *A. obliqua*, possivelmente por seleção sexual, será necessário o estudo de todo o gene *fru* e até mesmo da sua interação com outros genes da cascata de diferenciação sexual de forma a estabelecer de fato uma relação causal.

Capítulo IV

Evolução molecular do gene *doublesex* e *fruitless* :

Considerações finais

Considerações finais

Estudos comparativos entre linhagens divergentes têm mostrado que genes da cascata de diferenciação sexual possuem domínios extremamente conservados e que, por causa disso, estão sob uma elevada restrição seletiva (Gailey *et al.* 2006; Cho *et al.* 2007; Ruiz *et al.* 2007a; Ruiz *et al.* 2007b). Isso é compreensível, visto que tais genes são críticos para o desenvolvimento do organismo e de características intimamente relacionadas ao valor adaptativo do indivíduo (Lee *et al.* 2001; Song *et al.* 2002; Zhang *et al.* 2006). Assim, espera-se que mutações em regiões importantes destes genes sejam em sua maioria deletérios e possam diminuir ou mesmo eliminar a capacidade de reprodução do indivíduo, ou até mesmo impedir o seu desenvolvimento adequado. Por exemplo, mutações em *dsx*, especialmente em seus domínios conservados, têm sido associadas a fenótipos intersexuais inférteis (Zhang *et al.* 2006) e o mesmo vale para o gene *fru*, no qual têm sido descritas mutações em seu domínios mais conservados, BTB e de ligação ao DNA, que levam à infertilidade (Lee *et al.* 2001). Dessa forma espera-se uma frequência menor na população de mutações não-sinônimas segregando nestas regiões e, em comparações interespecíficas, uma taxa de ω reduzida indicando a ação de seleção purificadora sobre estes domínios. Implícita nas considerações acima está a suposição de que regiões importantes do genoma deveriam estar sob uma pressão de seleção purificadora constante. Entretanto, a dinâmica evolutiva dos genes é mais complexa. Regiões com um padrão de diversificação próximo ao neutro e que à primeira vista mostram-se "não-funcionais", podem na realidade revelar importantes traços de seleção adaptativa. Em diversas situações, a seleção agindo em regiões distintas no genoma pode alterar nossa expectativa sobre a evolução de certos genes, particularmente em genes física ou geneticamente ligados.

Mostramos neste trabalho que apesar de serem considerados "genes conservados", tanto *fru* quanto *dsx* apresentam regiões que diversificaram-se via seleção positiva, especialmente nos segmentos codificantes que conectam os dois domínios mais conservados destas moléculas. Isto indica que, mesmo para dados populacionais, a grande proximidade existente entre estas regiões e as regiões conservadas sob seleção purificadora não é suficiente em geral para reduzir os seus níveis de variação a não ser em circunstâncias especiais, como um *selective sweep* aqui detectado. Isso é de certa forma interessante, uma vez que estas regiões conectoras, notadamente em *fru*, eram consideradas de pouca importância para o funcionamento da proteína e deveriam evoluir neutramente (Gailey *et al.* 2006).

Embora tenhamos detectado assinaturas de seleção positiva nos dois genes, o padrão observado em *dsx* diferenciou-se daquele verificado em *fru*. Neste último gene, a ocorrência de

selective sweep na região conectora para o grupo de *A. obliqua* parece estar envolvida na estruturação da rede em dois grupos distintos, com o isolamento geral de *A. obliqua* das outras espécies. Este padrão é ainda mais relevante em se considerando que nossos dados, tanto destes quanto de outros marcadores genéticos, têm indicado que as espécies do grupo *fraterculus* apresentam baixa diferenciação e pouca subestrutura populacional mesmo em nível continental. Por este motivo, a existência de um ramo em *fru* com diversas mutações, a grande maioria não-sinônima, que separa a grande maioria de espécimes de *A. obliqua* dos espécimes de *A. fraterculus* e *A. sororcula* é o padrão esperado para os genes chamados de especiação, ou seja, genes que estejam envolvidos com o sistema de reconhecimento e determinação das espécies. Um padrão semelhante foi encontrado em genes expressos em tecidos reprodutivos em uma busca para marcadores genéticos que discriminassem duas espécies crípticas de grilos que tinham grande dificuldade de diferenciação pelo uso de diversos marcadores (Andrés *et al.* 2008), reforçando a hipótese de que genes envolvidos com características reprodutivas são candidatos favoráveis para se detectar a monofilia de espécies recentemente divergidas que ainda retêm muito polimorfismo ancestral.

Por outro lado, os dados obtidos para *dsx* são também interessantes e evidenciam algo diferente, uma vez que os dados de *dsx* revelam mutações radicais que estão distribuídas em um padrão específico continental de forma que a grande maioria dos espécimes coletados apresentam-se homozigotos e com uma distribuição específica no Brasil, não obstante a baixa diferenciação continental para outros marcadores genéticos.

Vale ressaltar que a detecção de sinais de seleção positiva não é uma tarefa trivial, principalmente quando estudamos genes que estão sob razoável restrição seletiva. Além disso, as estratégias geralmente usadas para sua detecção são mais adequadas para o contraste de seqüências que apresentam uma grande diferenciação, e não dados populacionais, uma vez que estas metodologias têm problemas para lidar com a estocasticidade evolutiva (Kryazhimskiy & Plotkin 2008). Por causa disso, tentamos uma abordagem que não só avaliasse se a mutação levaria ou não a uma alteração de amino-ácido, mas também sua “importância” na molécula, considerando, para isso, as mudanças nas propriedades físico-químicas dos aminoácidos mutados e sua persistência na população, medida através do número de cópias deixadas por ela ao longo do processo evolutivo. É claro que esta metodologia por si só não é suficiente para entender por completo a dinâmica evolutiva do gene estudado. Para tanto, devem ser utilizadas diferentes abordagens para detecção de seleção, principalmente quando estudamos dados intrapopulacionais, que apresentam baixa divergência e estão sujeitos às diferentes dinâmicas populacionais. Apesar de já existirem inúmeros testes para detecção de seleção em dados populacionais que fazem uso da frequência de alelos raros ou comuns e sua distribuição na população, esta metodologia representa uma abordagem inovadora

ainda não utilizada em estudos de genética evolutiva, por considerar simultaneamente a frequência de uma mutação, a mudança causada por ela nas propriedades dos aminoácidos mutados e sua posição numa rede haplotípica. Utilizando-a pudemos inferir pelo menos duas mutações sob seleção adaptativa, estando uma delas localizada no próprio domínio de ligação ao DNA de *dsx*. Contudo, tais conclusões devem ser tomadas com cautela, uma vez que não foram realizados estudos sobre a robustez da análise frente à detecção de falsos positivos. Estudos estes que merecem e serão realizados no futuro através de simulações computacionais.

Referências Bibliográficas

Referências Bibliográficas

ALUJA, M.. Bionomics and Management of Anastrepha. *Annual Review of Entomology*, v 39, p. 155-178, 1994.

ALUJA, M.; A. L. NORRBOM. *Fruit flies (Tephritidae) : phylogeny and evolution of behavior*. 1 ed. Boca Raton: CRC Press, 2000. 967 p.

AMARAL, P. M. Estudo da variabilidade isozímica de sete locos em larvas de populações naturais de *Anastrepha* (Diptera, Tephritidae). 1993. 183 f. Tese (Doutorado em Genética) - Dept. Biologia Instituto de Biociências da Universidade de São Paulo, São Paulo, 1994.

ANDRÉS, J. A.; MAROJA, L. S.; HARRISON, R. G.. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proceedings of the Royal Society B: Biological Sciences*, v. 275, p. 1975-1983, 2008.

ANISIMOVA, M.; BIELAWSKI, J. P.; YANG, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution*, v. 18, p. 1585-1592, 2001.

ARAUJO, E. L.; ZUCCHI, R. A. Medidas do acúleo na caracterização de cinco espécies de *Anastrepha* do grupo *fraterculus* (Diptera: Tephritidae). *Neotropical Entomology*, v. 35, p. 329-337, 2006.

BAKEWELL, M. A.; SHI, P.; ZHANG, J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proceedings of the National Academy of Sciences*, v. 104, p. 7489-7494, 2007.

BERLIN, S.; SMITH, N. Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evolutionary Biology*, v. 5, p. 65-74, 2005.

BILLETER, J. C. *et al.* Isoform-specific control of male neuronal bifferentiation and behavior in *Drosophila* by the *fruitless* gene. *Current Biology*, v. 16, p. 1063-1076, 2006.

BURGHARDT, G. *et al.* The *transformer2* gene in *Musca domestica* is required for selecting and maintaining the female pathway of development. *Development Genes and Evolution*, v. 215, p. 165-176, 2005.

CARLBORG, O.; HALEY, C. S. Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics*, v. 5, p. 618-625, 2004.

CASTRIC, V.; VEKEMANS, X. Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *BMC Evolutionary Biology*, v. 7; p. 132, 2007.

- CHAPMAN, T. The soup in my fly: Evolution, form and function of seminal fluid proteins. *Plos Biology*, v. 6, p. 1379-1382, 2008.
- CHARLESWORTH, B.; CHARLESWORTH, D.; BARTON, N. H. The effects of genetic and geographic structure on neutral variation. *Annual Review of Ecology Evolution and Systematics*, v. 34, p. 99-125, 2003.
- CHEN, S. L. *et al.* Female-specific *doublesex* dsRNA interrupts yolk protein gene expression and reproductive ability in oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Insect Biochemistry and Molecular Biology*, v. 38, p. 155-165, 2008.
- CHO, S.; WENSINK, P. C. DNA Binding by the Male and Female Doublesex Proteins of *Drosophila melanogaster*. *Journal of Biological Chemistry*, v. 272, p. 3185-3189, 1997.
- CHO, S. *et al.* Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research*, v. 16, p. 1366-1375, 2006.
- CHO, S.; HUANG, Z. Y.; ZHANG, J. Sex-specific splicing of the honeybee *doublesex* gene reveals 300 Million years of evolution at the bottom of the insect sex-determination pathway. *Genetics*, v. 177, p. 1733-1741, 2007.
- CHOMCZYNSKI, P.; SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, v. 162, p. 156-159, 1987.
- CLARK, N. L., AAGAARD, J. E.; SWANSON, W. J. Evolution of reproductive proteins from animals and plants. *Reproduction*, v. 131, p. 11-22, 2006.
- CLARK, N. L. *et al.* Duplication and selection on abalone sperm lysin in an allopatric population. *Molecular Biology and Evolution*, v. 24, p. 2081-2090, 2007.
- CLARK, N. L.; SWANSON, W. J. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genetics*, v. 1, p. e35, 2005.
- CLEMENT, M.; POSADA, D.; CRANDALL, K. A. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, v. 9, p. 1657-1659, 2000.
- CLINE, J.; BRAMAN, J. C.; HOGREFE, H. H. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, v. 24, p. 3546-3551, 1996.
- CLINE, T. W.; MEYER, B. J. Vive la difference: males vs females in flies vs worms. *Annual Review of Genetics*, v. 30, p. 637-702, 1996.
- COSCHIGANO, K. T.; WENSINK, P. C. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes and Development*, v. 7, p. 42-54, 1993.
- DATTA, S. R. *et al.* The *Drosophila* pheromone cVA activates a sexually dimorphic neural circuit. *Nature*, v. 452, p. 473-477, 2008.

- DEMIR, E.; DICKSON, B. J. *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell*, v. 121, p. 785–794, 2005.
- ERDMAN, S. E.; BURTIS, K. C. The *Drosophila doublesex* proteins share a novel zinc finger related DNA binding domain. *The EMBO Journal*, v. 12, p. 527-535, 1993.
- FAY, J.C.; WU, C.-I. Detecting hitchhiking from patterns of DNA polymorphism. In: Nurminsky, D. (ed.) *Selective Sweep*. New York and Gerogetown: Eureka.com and Kluwer Academic / Plenum Publishers, 2005. p. 65-77.
- FAY, J. C.; WU, C.-I. Hitchhiking Under Positive Darwinian Selection. *Genetics*, v. 155, p. 1405-1413, 2000.
- FELSENSTEIN, J. PHYLIP (Phylogeny Inference Package) version 3.6. [3.68]. Distribuído pelo autor. Department of Genome Sciences, University of Washington, Seattle, 2004. *Programa de computador*.
- FELSENSTEIN, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, v. 5, p. 164-166, 1989.
- FERRIER, D. E. K.; MINGUILLÓN, C. Evolution of the *Hox/ParaHox* gene clusters. *The International Journal of Developmental Biology*, v. 47, p. 605-611, 2003.
- FU, Y.-X.; LI, W.-H. Statistical tests of neutrality of mutations. *Genetics*, v. 133, p. 693-709, 1993.
- GAILEY, D. A. *et al.* Functional conservation of the *fruitless* male sex-determination gene across 250 Myr of insect evolution. *Molecular Biology and Evolution*, v. 23, p. 633-643, 2006.
- GALINDO, B. E.; VACQUIER, V. D.; SWANSON, W. J. Positive selection in the egg receptor for abalone sperm lysin. *Proceedings of the National Academy of Sciences*, v. 100, p. 4639-4643, 2003.
- GERRARD, D. T.; MEYER, A. Positive selection and gene conversion in SPP120, a fertilization-related gene, during the east african cichlid fish radiation. *Molecular Biology and Evolution*, v. 24, p. 2286-2297, 2007.
- HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, v. 41, p. 95-98, 1999.
- HALLIBURTON, R. Natural selection I: Basic models. In: _____. *Introduction to population genetics*. 1. ed. Upper Saddle River: Pearson Prentice Hall, 2004. p. 129-182.
- HEDIGER, M. *et al.* Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator doublesex. *Development Genes and Evolution*, v. 214, p. 29-42, 2004.
- ITO, H. *et al.* Sexual orientation in *Drosophila* is altered by the satori mutation in the sex-determination gene *fruitless* that encodes a zinc finger protein with a BTB domain. *Proceedings of the National Academy of Sciences of the United States of America*, v. 93, p. 9687-9692, 1996.

- JAGADEESHAN, S.; SINGH, R. S. Rapidly evolving genes of *Drosophila*: Differing levels of selective pressure in testis, ovary, and head tissues between sibling species. *Molecular Biology and Evolution*, v. 22, p. 1793-1801, 2005.
- JOHNSON, M. E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature*, v. 413, p. 514-519, 2001.
- KIMURA, K. I., *et al.* Fruitless and doublesex coordinate to generate male-specific neurons that can initiate courtship. *Neuron*, v. 59, p. 759-769, 2008.
- KRYAZHIMSKIY, S.; PLOTKIN, J. B. The population genetics of dN/dS. *PLoS Genetics*, v. 4, p. 1-10, 2008.
- KUHN, S.; SIEVERT, V.; TRAUT, W. The sex-determining gene doublesex in the fly *Megaselia scalaris*: Conserved structure and sexspecific splicing. *Genome*, v. 43, p. 1011-1020, 2000.
- KULATHINAL, R. J. *et al.* Rapid evolution of the sex-determining gene, *transformer*: Structural diversity and rate heterogeneity among sibling species of *Drosophila*. *Molecular Biology and Evolution*, v. 20, p. 441-452, 2003.
- KYRIACOU, C. P. Sex in fruitflies is fruitless. *Nature*, v. 436, p. 334-335, 2005.
- LAGOS, D. *et al.* Isolation and characterization of the *Bactrocera oleae* genes orthologous to the sex determining *Sex-lethal* and *doublesex* genes of *Drosophila melanogaster*. *Gene*, v. 348, p. 111-121, 2005.
- LEE, G. *et al.* New reproductive anomalies in *fruitless*-mutant *Drosophila* males: extreme lengthening of mating durations and infertility correlated with defective serotonergic innervation of reproductive organs. *Journal of Neurobiology*, v. 47, p. 121-149, 2001.
- LIS, J. T.; SCHLEIF, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Research*, v. 2, p. 383-390, 1975.
- LU, Y.; RAUSHER, M. D. Evolutionary rate variation in anthocyanin pathway genes. *Molecular Biology and Evolution*, v. 20, p. 1844-1853, 2003.
- MALAVASI, A.; R. A. ZUCCHI. *Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado*. 1 ed. Ribeirão Preto: Holos Editora, 2000. 327 p.
- MARTIN, D.; RYBICKI, E. RDP: Detection of recombination amongst aligned sequences. *Bioinformatics*, v. 16, p. 562-563, 2000.
- MCALLISTER, B. F.; MCVEAN, G. A. T. Neutral evolution of the sex-determining gene *transformer* in *Drosophila*. *Genetics*, v. 154, p. 1711-1720, 2000.
- MCCLELLAN, D. A. *et al.* Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Molecular Biology and Evolution*, v. 22, p. 437-455, 2005.

- MCCLELLAN, D. A.; MCCRACKEN, K. G. Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. *Molecular Biology and Evolution*, v. 18, p. 917-925, 2001.
- MCDONALD, J. H.; KREITMAN, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, v. 351, p. 652-654, 1991.
- MORGANTE, J. S.; MALAVASI, A.; BUSH, G. L. Biochemical systematics and evolutionary relationships of neotropical *Anastrepha*. *Annals of the Entomological Society of America (USA)*, v. 73, p. 622-630, 1980.
- NEI, M. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press, 1987. 512 p.
- NICKEL, G. C. *et al.* An empirical test for branch-specific positive selection. *Genetics*, v. 179, p. 2183-2193, 2008.
- NORRBOM, A. L.; ZUCCHI, R. A.; HERNÁNDEZ-ORTIZ, V. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotrypanini) based on morphology. In: Aluja, M.; Norrbom, A. L. (eds.) *Fruit flies (Tephritidae): Phylogeny and evolution of behavior*. Boca Raton: CRC Press, 2000. p. 299-342.
- O'NEIL, M. T.; BELOTE, J. M. Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics*, v. 131, p. 113-128, 1992.
- PANHUIS, T. M.; SWANSON, W. J. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. *Genetics*, v. 173, p. 2039-2047, 2006.
- POMIANKOWSKI, A. N. R.; WILKINS A. The Evolution of the *Drosophila* Sex-Determination Pathway. *Genetics*, v. 166, p. 1761-1773, 2004.
- PORTER, M. L. *et al.* Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins. *Molecular Biology and Evolution*, v. 24, p. 253-268, 2006.
- POSADA, D.; CRANDALL, K. A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, p. 13757-13762, 2001.
- RIDEOUT, E. J.; BILLETER, J. C.; GOODWIN, S. F. The sex-determination genes *fruitless* and *doublesex* specify a neural substrate required for courtship song. *Current Biology*, v. 17, p. 1473-1478, 2007.
- ROFF, D. A.; BENTZEN, P. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Molecular Biology and Evolution*, v. 6, p. 539-545, 1989.
- ROZAS, J. *et al.* DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, v. 19, p. 2496-2497, 2003.

- RUIZ, M. F., *et al.* The gene *transformer* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *PLoS ONE*, v.2, p. e1239, 2007a.
- RUIZ, M. F. *et al.* The gene *doublesex* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. *Development Genes and Evolution*, v. 217, p. 675-731, 2007b.
- RUIZ, M. F. *et al.* The gene *doublesex* of the fruit fly *Anastrepha obliqua* (Diptera, Tephritidae). *Genetics*, v. 171, p. 849-854, 2005.
- SACCONI, G. *et al.* The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene *sex-lethal* is structurally conserved, but not sex-specifically regulated. *Development*, v. 125, p. 1495-1500, 1998.
- SAINUDIIN, R. *et al.* Detecting Site-Specific Physicochemical Selective Pressures: Applications to the Class I HLA of the Human Major Histocompatibility Complex and the SRK of the Plant Sporophytic Self-Incompatibility System. *Journal of Molecular Evolution*, v. 60, p. 315-326, 2005.
- SAWYER, S. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution*, 6: 526-538, 1989.
- SCALI, C.; CATTERUCCIA, F.; LI, Q.; CRISANTI, A. Identification of sex-specific transcripts of the *Anopheles gambiae doublesex* gene. *Journal of Experimental Biology*, v. 208, p. 3701-3709, 2005.
- SCHÜTT, C.; NÖTHIGER, R. Structure, function and evolution of sex-determining systems in Dipteran insects. *Development*, v. 127, p. 667-677, 2000.
- SELIVON, D. Biologia e padrões de especiação. In: Malavasi, A.; Zucchi, R. A. (eds.) *Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado*. Ribeirão Preto: Holos Editora, 2000. p. 25-28.
- SMITH, J. M. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, v. 34, p. 126-129, 1992.
- SMITH-CALDAS, M. R. B. *et al.* Phylogenetic relationships among species of the *fraterculus* group (*Anastrepha*: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial *cytochrome oxidase I*. *Neotropical Entomology*, v. 30, p. 565-573, 2001.
- SOBRINHO JR., I.S.; DE BRITO, R.A. Capítulo I - Evolução molecular do gene *doublesex* no grupo *Anastrepha fraterculus* (Diptera, Tephritidae) in Sobrinho Jr., I.S. e de Brito, R.A. Evolução molecular dos genes *doublesex* e *fruitless* em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae) Dept. Genética e Evolução - CCBS, Universidade Federal de São Carlos, São Paulo, 2009. *Presente tese de doutorado*.
- SONG, H. J. *et al.* The *fruitless* Gene Is Required for the Proper Formation of Axonal Tracts in the Embryonic Central Nervous System of *Drosophila*. *Genetics*, v. 162, p. 1703-1724, 2002.

- STECK, G. J. Biochemical systematics and population genetic structure of *Anastrepha fraterculus* and related species (Diptera: Tephritidae). *Annals of the Entomological Society of America*, v. 84, p. 10-28, 1991.
- SUZUKI, M. G. *et al.* The mechanism of sex-specific splicing at the *doublesex* gene is different between *Drosophila melanogaster* and *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, v. 31, p. 1201-1211, 2001.
- SWANSON, W. J.; VACQUIER, V. D. Reproductive Protein Evolution. *Annual Review of Ecology and Systematics*, v. 33, p. 161-179, 2002a.
- SWANSON, W. J. *et al.* Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences*, v. 98, p. 7375-7379, 2001.
- SWANSON, W. J.; NIELSEN, R.; YANG, Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, v. 20, p. 18-20, 2003.
- SWANSON, W. J.; VACQUIER, V. D. The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, v. 3, p. 137-144, 2002b.
- SWANSON, W. J. *et al.* Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics*, v. 168, p. 1457-1465, 2004.
- TAJIMA, F. The effect of change in population size on DNA polymorphism. *Genetics*, v. 123, p. 597-601, 1989b.
- TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA Polymorphism. *Genetics*, v. 123, p. 585-595, 1989a.
- TEMPLETON, A. R. Contingency tests of neutrality using intra/interspecific gene trees: The rejection of neutrality for the evolution of the mitochondrial *cytochrome oxidase II* gene in the hominoid primates. *Genetics*, v. 144, p. 1263-1270, 1996.
- TEMPLETON, A. R. *Population genetics and microevolutionary theory*. 1 ed. Hoboken: John Wiley and Sons, Inc., 2006. 705 p.
- TING, C. T.; TSAUR, S. C.; WU, C. I. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proceedings of the National Academy of Sciences of the United States of America*, v. 97, p. 5313-5316, 2000.
- TSAOUSIS, A. D. *et al.* Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution*, v. 22, p. 925-933, 2005.
- WEINREICH, D. M.; WATSON, R. A.; CHAO, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, v. 59, p. 1165-1174, 2005.

- WILKINS, A. S. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays*, v. 17, p. 71-77, 1995.
- WONG, W. S. W. *et al.* Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, v. 168, p. 1041-1051, 2004.
- WOOLLEY, S. *et al.* TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, v. 19, p. 671-672, 2003.
- YANG, Z. Inference of selection from multiple species alignments. *Current Opinion in Genetics & Development*, v. 12, p. 688-694, 2002.
- YANG, Z.; NIELSEN, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, v. 19, p. 908-917, 2002.
- YANG, Z.; SWANSON, W. J. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution*, v. 19, p. 49-57, 2002.
- YANG, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, v. 24, p. 1586-1591, 2007.
- YANG, Z. *et al.* Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*, v. 155, p. 431-449, 2000.
- YANG, Z.; WONG, W. S. W.; NIELSEN, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, v. 22, p. 1107-1118, 2005.
- ZENG, K.; SHI, S.; WU, C.-I. Compound tests for the detection of hitchhiking under positive selection. *Molecular Biology and Evolution*, v. 24, p. 1898-1908, 2007.
- ZHANG, J.; NIELSEN, R.; YANG, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, v. 22, p. 2472-2479, 2005.
- ZHANG, W. *et al.* Regulation of sexual dimorphism: mutational and chemogenetic analysis of the *doublesex* DM domain. *Molecular and Cellular Biology*, v. 26, p. 535-547, 2006.
- ZHU, L. *et al.* Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes and Development*, v. 14, p. 1750-1764, 2000.
- ZUCCHI, R. A. Espécies de *Anastrepha*, sinónimas, plantas hospedeiras e parasitóides. In; Malavasi, A.; Zucchi, R. A. (eds.) *Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado*. Ribeirão Preto: Holos Editora, 2000. p. 41-48.

Anexos

Anexo 1. Alinhamento das sequências de *doublesex*, mostrando os haplótipos obtidos neste trabalho. Os pontos “.” representam base idênticas às da sequência de referência; traços “-“ indicam ausência de bases.

	10	20	30	40	50	60	70	80	90	100	
1-F	ATGTCGGATTTCGGACATGCTCGATTCAAAGGCAGACGTTTGTGGTGGCGCCTCCAGCAGCAGCGGCAGTTCGATCTCGCCCCGAACCCCTCCTAATTGTG										
9-F										G.....
10-F										A.....
11-F										A.....
12-F										
13-F										T.....
14-F										
15-F										
24-F										
29-F										
40-F										A.....
63-F										A.....
66-F										A.....
65-F										
67-F										A.....
68-F										A.....
69-F										A.....
69-F										A.....
16-F										A.....
70-F										A.....
71-F										A.....
72-F										A.....
73-F										A.....
74-F										A.....
75-F										G.....
76-F										A.....
77-F										
78-F										
79-F										C.....
81-F										A.....
80-F										A.....
89-F										A.....
90-F										
91-F										
82-F										A.....
92-F										A.....
1-S										
8-S										G.....
25-S										
30-S										

31-S A A
 53-S
 54-S
 55-S A C
 56-S A
 57-S A GT T C
 58-S A T T
 59-S
 60-S G
 61-S
 62-S
 64-S
 86-S
 96-S C
 1-O
 2-O T G T
 3-O ---
 4-O
 5-O C
 6-O T
 7-O
 16-O A T C
 17-O T C
 18-O A
 19-O
 20-O A
 21-O
 22-O
 23-O
 25-O
 26-O
 27-O A
 28-O
 32-O
 33-O
 34-O
 35-O T T
 36-O A T
 37-O A T C
 38-O A
 39-O
 41-O C
 42-O
 43-O A
 44-O A
 25-O
 45-O A TT T

46-OT.....T.....
 47-OT.....T.....
 48-OA.....
 49-OA.....
 50-OT.....
 51-OC.....
 52-O
 88-OT.....
 83-B
 93-B
 84-G
 94-G
 87-T
 97-T
 85-E
 95-E

.....110.....120.....130.....140.....150.....160.....170.....180.....190.....200.....
 1-F CCAGATGCCGTAATCATGGACTAAAAATCACATTGAAAGGACACAAACGCTATTGTAAGTTCCGGTTTTGCACTTGCGAAAAATGTCGCCTAACCGCCGA
 9-FA.....
 10-FA.....
 11-FA.....
 12-FT.....
 13-FC.....
 14-F
 15-FG.....
 24-FA.....
 29-F
 40-FA.A.....
 63-FA.....
 66-FA.....
 65-FT.....
 67-FA.....
 68-FA.....
 69-FC.....A.....
 69-FC.....A.....
 16-FG.....A.....
 70-FG.....G.....A.....
 71-FA.....
 72-FA.....G.....
 73-FA.....
 74-FA.....
 75-FA.....
 76-F

77-F	A
78-F	A
79-F	A
81-F	A
80-F	A
89-F	A
90-F	A
91-F G	A
82-F	A T
92-F	A T T
1-S
8-S C
25-S	A
30-S
31-S
53-S T
54-S
55-S	A T
56-S	A
57-S G	A
58-S G	A
59-S T
60-S T
61-S T
62-S
64-S
86-S
96-S	C
1-O
2-O	A
3-O	A
4-O	A A
5-O C
6-O	A
7-O	A
16-O G	A
17-O	G
18-O	A
19-O	A
20-O A	A
21-O	G
22-O T A
23-O A	
25-O	A
26-O	A
27-O	A
28-O	A

32-O	A
33-O	A
34-OC	A
35-O	A
36-O	A
37-O	A
38-O	A
39-O	A
41-O	A
42-O	A
43-O	A
44-O	A
25-O	A
45-OG	T
46-O	A
47-O	A
48-O	A
49-O	AT
50-O	A
51-O	A
52-O	A
88-OC	A
83-BG	A
93-B	A
84-GG	A
94-GC	T
87-T	AT
97-T	AT
85-EC	CA
95-EG	AG

	210	220	230	240	250	260	270	280	290	300
1-F	TCGGCAACGCGTTATGGCGCTACAGACAGCATTGCGACGCGCCCAAGCCCAGGATGAGCAACGCGTTTTGCAGATGCACGAGGTGCCGCCAGTCGTTTCAT									
9-F									
10-FG.....GA.....A..A..									
11-FGA.....A..									
12-F									
13-F									
14-FG.....									
15-FG.....									
24-F									
29-FG.....									
40-FGA.....A..									
63-FC.....GA.....A..									
66-FGA.....A..									
65-FC.....									
67-FC.....GA.....A..									
68-FCC.....GA.....A..									
69-FGA.....A..									
69-FGA.....A..									
16-FC.....A.....GA.....A..									
70-FC.....A.....GA.....T.....A..									
71-FGA.....A..									
72-FGA.....A..									
73-FGA.....A..									
74-F									
75-FA.....									
76-FT.....									
77-F									
78-FT.....									
79-FGA.....A..									
81-FGA.....A..									
80-FC.....GA.....T.....A..									
89-FC.....GA.....A..									
90-F									
91-F									
82-FG.....									
92-F									
1-S									
8-S									
25-S									
30-SG.....									
31-S									
53-S									
54-SA.....A.....GA.....A..									
55-SC.....A.....GA.....A..									

56-S C GA A
57-S C A GA A
58-S C G GA A
59-S
60-S
61-S
62-S
64-S
86-S
96-S
1-O
2-O
3-O T
4-O
5-O
6-O
7-O
16-O C A GA A
17-O
18-O T
19-O
20-O
21-O
22-O A
23-O
25-O
26-O G
27-O
28-O G
32-O
33-O
34-O
35-O
36-O
37-O
38-O
39-O
41-O G
42-O
43-O G
44-O
25-O
45-O
46-O T
47-O A
48-O GA
49-O GA A

50-O
51-O
52-OA.....G.....
88-O
83-B
93-BC.....
84-GC.....
94-GA.....
87-TC.....
97-TC.....
85-EA.....C.....
95-E

	310	320	330	340	350	360	370	380	390	400
1-F	GCGCCCACAGCTCTGCTCAATCATCATCATTACATCATCATCAT---CACTTGAACCAGAATCATCATGCGACTGCTGCCGGCTGCCGCTGCTGCCGCTG									
9-FT.....CAT.C..A.....									
10-FT.....									
11-FT.....									
12-F									
13-F									
14-FC.....G.....									
15-FG.....C.....									
24-FC.....C.....									
29-FC.....									
40-FC.....C.....									
63-F									
66-F									
65-FC.....T.....									
67-FC.....C.....									
68-FC.....									
69-F									
69-F									
16-FC.....									
70-FC.....									
71-FT.....									
72-FT.....									
73-FT.....									
74-FC..A.....									
75-F									
76-FC.....									
77-F									
78-FC.....C.....									
79-F									
81-F									
80-FC.....									
89-FC.....									
90-FC.....									
91-FG.....C.....									
82-F									
92-F									
1-S									
8-S									
25-SC..A.....									
30-SC.....									
31-SC.....									
53-SC.....									
54-SC.....									
55-SC.....									

56-S C. ---
57-S C. C. ---
58-S C. A. --- T.
59-S --- C.
60-S --- C. G.
61-S --- C. G.
62-S T. --- C.
64-S T. --- C.
86-S --- C.
96-S --- C.
1-O ---
2-O --- C. A.
3-O --- C. A.
4-O A. --- G.
5-O --- C. A.
6-O ---
7-O --- C. A. A.
16-O C. G. ---
17-O --- C. A.
18-O ---
19-O --- C.
20-O --- C.
21-O ---
22-O T. --- C. A.
23-O ---
25-O --- C. A.
26-O --- C. A.
27-O --- C. A.
28-O C ---
32-O ---
33-O --- C. A.
34-O --- C. A.
35-O --- C.
36-O --- C.
37-O --- C.
38-O ---
39-O ---
41-O --- C.
42-O C. --- T. A.
43-O --- C. A.
44-O --- C. A.
25-O --- C. A.
45-O --- G.
46-O --- C. A.
47-O C. --- C. A.
48-O ---
49-O ---

```

50-O ..... --- .....
51-O ..... --- ..... C .....
52-O ..... --- C A .....
88-O ..... G C G ..... --- C A .....
83-B ..... --- ..... C .....
93-B ..... C G ..... --- T .....
84-G ..... --- ..... C .....
94-G ..... C ..... --- .....
87-T ..... C G ..... --- T .....
97-T ..... C G ..... --- T .....
85-E ..... G ..... --- .....
95-E ..... --- ..... C .....

```

```

          410      420      430      440      450      460      470      480      490      500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
1-F  CAGCACATCATCATATTTTCGACAGCAGTCCGATCTCCGCCACAGACCGAACATGGCAGCGGAGGTGGTGGTGGCGGTATGGTAGGTGGCACAGTACCGAC
9-F  ..... A .....
10-F ..... A ..... A ..... T .....
11-F ..... A ..... A .....
12-F ..... C .....
13-F ..... C .....
14-F ..... A .....
15-F ..... A G .....
24-F ..... A .....
29-F .....
40-F ..... TA ..... A .....
63-F ..... G ..... A ..... A ..... A ..... T .....
66-F ..... G ..... A ..... A .....
65-F ..... G ..... A A .....
67-F ..... G ..... A ..... A .....
68-F ..... G ..... A ..... A .....
69-F ..... A ..... A .....
69-F ..... A ..... A .....
16-F ..... G ..... A ..... A .....
70-F ..... G ..... A ..... A ..... T .....
71-F ..... A ..... A ..... A ..... T .....
72-F ..... G ..... A ..... A .....
73-F ..... G ..... A ..... A .....
74-F ..... A .....
75-F ..... A .....
76-F ..... T .....
77-F ..... T .....
78-F ..... A .....
79-F ..... A ..... G ..... A .....

```

81-F	A	A
80-F	G	A
89-F	G	A
90-F				
91-F	A		
82-F	A	A
92-F	A	A
1-S				
8-S				
25-S	A		
30-S	A		
31-S	A		
53-S	A		
54-S	A		
55-S	G	A
56-S	G	A
57-S	G	A
58-S	G	A
59-S	G	A	A
60-S	G	A	A
61-S	G	A	A
62-S			A
64-S	G	A	A
86-S			A
96-S			A
1-O				
2-O			A
3-O	T	A
4-O			A
5-O			A
6-O			A
7-O			A
16-O	G	A
17-O			A
18-O			A
19-O			A
20-O			A
21-O	A	A
22-O			A
23-O				
25-O			A
26-O			A
27-O			A
28-O			A
32-O			A
33-O			A
34-O			A

35-O G A
 36-O A
 37-O A
 38-O A
 39-O A C
 41-O A
 42-O A G
 43-O A
 44-O A
 25-O A
 45-O A
 46-O A
 47-O A
 48-O A A
 49-O A A
 50-O T
 51-O A
 52-O C A
 88-O A
 83-B A
 93-B C A A G
 84-G A
 94-G A A
 87-T C A G A
 97-T C A A
 85-E G A A
 95-E A

	510	520	530	540
			
1-F	CATCACCTCTGTACCTGTCTCAGCTCCACCACCGGAACATCAT			
9-F			
10-FC...			
11-FC...			
12-F			
13-F			
14-F			
15-F			
24-FC.....			
29-F			
40-F	A			C.....
63-F			
66-F			
65-FC.....			
67-F			
68-F			
69-FT.....C...			
69-FT.....C...			
16-F			
70-F			
71-FC.....			
72-F			
73-F			
74-F			
75-F			
76-F			
77-F			
78-FA.....			
79-FT.....C.....			
81-FC.....			
80-F			
89-F			
90-F			
91-F			
82-F			
92-F			
1-S			
8-S			
25-S			
30-S			
31-S			
53-S			
54-S			
55-S			

56-S
57-S
58-S
59-SC.....
60-SC.....
61-SC.....
62-S
64-SC.....G.....
86-S
96-S
1-O
2-OG.....
3-O
4-O
5-O
6-O
7-OC.....
16-O
17-O
18-O
19-O
20-O
21-O
22-O
23-O
25-O
26-O
27-O
28-O
32-O
33-OG.....
34-O
35-O
36-O
37-O
38-O
39-OG.....
41-O
42-O
43-O
44-OG.....
25-O
45-O
46-O
47-O
48-O
49-OC.....

50-O
51-O
52-O
88-O
83-B
93-B T C
84-G
94-G C
87-T C
97-T C
85-E C
95-E

Anexo 2. Alinhamento de *doublesex*. Dados macroevolutivos. Os pontos “.” representam base idênticas às da seqüência de referência; traços “-” indicam ausência de bases.

	10	20	30	40	50	60	70	80	90	100
<i>D. yakuba</i>	ATG	TCCGACTCGG	ACATGATCGACTCAA	AAGACGACGATG	TGGCGGTGCCTCC	AGTTCAGTGGCAGCT	CGATTTCGCCGAGG	ACGCCACCCAACTGG		
<i>D. melanogaster</i>				C	A				A	G
<i>D. simulans</i>				C	A				A	
<i>D. sechellia</i>				C	A		A	C	A	
<i>D. erecta</i>				C	A	A		C		
<i>D. pseudoobscura</i>	T	T	G	G	C	G	C	G	T	A
<i>D. persimilis</i>	T	T	G	G	C	G	C	G	T	A
<i>D. virilis</i>			G	T	G	C	G	C	A	AC
<i>D. mojavensis</i>			G	T	G	C	C	C	T	T
<i>D. ananassae</i>		T	T	C	G	C	A	G		
<i>B. dorsalis</i>		T	CAT	GCA	T	C	T	G	CAG	C
<i>B. tryoni</i>	TG	T	T	CGT	GCA	T	C	T	G	CAG
<i>B. oleae</i>		T	CAT	GCA	T	T	T	T	CAG	C
<i>C. capitata</i>		T	ACA	T	GC	C	G	A	T	AG
<i>A. sororcula</i> (86-S)		T	C	T	GCA	T	T	C	CAG	C
<i>A. obliqua</i> (32-O)		T	C	T	GCA	T	T	C	CAG	C
<i>A. fraterculus</i> (66-F)		T	C	T	AGCA	T	T	C	CAG	C
	110	120	130	140	150	160	170	180	190	200
<i>D. yakuba</i>	CCC	GCTGCCCAATCAT	GGCCTGAAGATTAC	ACTGAAGGGACACA	AAGAGGTTACTGCA	AGTTCGCTATTGTAC	GTGCGAAAAGTGCC	GGCTGACGGCGGA		
<i>D. melanogaster</i>			A	C	C	A		C	C	G
<i>D. simulans</i>			C	C		C		C	C	G
<i>D. sechellia</i>			C	C		C		C	C	G
<i>D. erecta</i>			A	C	C			C	C	G
<i>D. pseudoobscura</i>	T	A	C	C		C		C	A	A
<i>D. persimilis</i>	T	A	C	C		C		C	C	A
<i>D. virilis</i>	G	T	C	C	G	C	T	C	C	A
<i>D. mojavensis</i>	G	T	C	C	G	C	CT	T	C	CGGACAGGCAGCG
<i>D. ananassae</i>		C	C	T	C	C		C	C	A
<i>B. dorsalis</i>	A	G	T	G	C	A	C	A	A	AC
<i>B. tryoni</i>	A	G	T	A	C	A	C	A	A	AC
<i>B. oleae</i>	A	G	A	A	C	A	C	A	A	AC
<i>C. capitata</i>	TA	G	T	A	C	C	A	A	AC	C
<i>A. sororcula</i> (86-S)	A	A	T	A	A	A	C	T	A	AC
<i>A. obliqua</i> (32-O)	A	A	T	A	A	A	C	T	A	AC
<i>A. fraterculus</i> (66-F)	A	A	T	A	A	A	C	T	A	AC
	210	220	230	240	250	260	270	280	290	300
<i>D. yakuba</i>	CCG	CAGCGGGTGTGG	CTCTGCAAACGGC	CTTGAGGCGAGCCC	AGGCGCAGGACGAG	CAGCGGGCAGTGC	ACATGCATGAGGTG	CCGCTGCCAATCCG		
<i>D. melanogaster</i>						T		G		
<i>D. simulans</i>						T		C		
<i>D. sechellia</i>						C		T		
<i>D. erecta</i>								G		
<i>D. pseudoobscura</i>		C	C	G	C	C	C			C

D. persimilisC.C.G.....C.C.C.....A.....C.C.....C.....C.GTCGA..
D. virilis TA.G....C.....G....G....GC..C...T..T.....AT.G.....C.....GC.GGCGGGC
D. mojaviensis -----G.GCTC.T.C.....A..CT.G.C.....C.....C.GGCGA.A
D. ananassae ..T..AA...A....C....G.....C.....C.....G..C.....C.....GA.T.CAA..
B. dorsalis T..T....C.....A....G..T..T.....G..A..A.....AA..TT...G..T..C...A...A.TGGTA.AT
B. tryoni T..T....C..C...A....G..T..T...A..G..A..A.....AA..TT...G..A..C...A..T..A.TGGTA.AC
B. oleae T..T..A..C..C...A....G....A....A..G..A..A.....T..A..AA..TT...G..T..C...A...A.TGGTC.AT
C. capitata T..T..A..C..A....A..T..G..C..A.....T..G.....T..A..T...A...TTT...G..A..C...A...A.GTGT..AC
A. sororcula (86-S) T..G..A..C..T...G..A..G..A..A..C..A..C...A..C...T...A..C.TTT...G...C.....A..T.GT..AT
A. obliqua (32-O) T..G..A..C..T...G..A..G..A..A..C..A..C...A..C...T...A..C.TTT...G...C.....A..T.GT..AT
A. fraterculus (66-F) T..G..A..C..T...G..A..G..A..A..C..A..C...A..C...T...A..ATTT...G...C.....A..T.GT.AAT

310 320 330 340 350 360 370 380 390 400

D. yakuba GCGGCCACAACCTTTGCTGAGTCATCACCATCATGTGCGCAGCTCCCGCCCATGTCCACGCCACCATGTGCATGCCCATCACGCACACGGA-----
D. melanogasterG..CC.C.....C.....T.....C.....
D. simulansG..CC..T.....T.....T.....
D. sechelliaG..CC..T.....T.....T.....
D. erecta
D. pseudoobscura T.AA.T..TG.CC....G.C..C....C..C....T..CA.....C..T..A..T..T..C..C...C.G..C...AGC..T..C-----
D. persimilis T.AA.T..TG.CC....G.C..C....C..C....T..CA.....C..T..A..T..T..C..C...C.G..C...AGC..T..C-----
D. virilis ..TA.AG.GG.CC....C.....TGGG...CATCATCA.GTGCATGCGCA.GC.CA.AG..C.CC...G.....GG..T..C-----
D. mojaviensis C...AG.GG.GC...TG.C.C....TGCC...CATCATCA.GTGCATGCC--GCTCA.AG...C.CC...G.....GC..T..CC-----
D. ananassae .GCA..G.C...C...C..C..C....C..C.C...C..AG.A.T.GCC.C.GC.C..GCT..C..C..C...GC.CACATGCACGGGCCCATC
B. dorsalis .GAC.A..GG.A..A..T.A....T..C-----
B. tryoni .GAC.A..GG.A..A..T.A....T..C-----
B. oleae .G.C.A..GG.A..A..T.A....T..C-----
C. capitata ..AC.AG.GG.A..A..T.A...C..T...-----
A. sororcula (86-S) ..C.....G..C...C.A.....T...-----
A. obliqua (32-O) ..C.....G..C...C.A.....T...-----
A. fraterculus (66-F) ..C.....G..C...C.A.....T...-----

410 420 430 440 450 460 470 480 490 500

D. yakuba -----CACCACTCGCATCAGGACATGTCCTGCAC---CACCAGCAGGCAGCGGCGGCCGAGCAGCAGCTCCGTCG-----
D. melanogaster -----GGA.....
D. simulans -----GGA.....
D. sechellia -----GGA.....A.....T-----
D. erecta -----
D. pseudoobscura -----CACCACTCG.....C.....G.....CAG.....AT..T..TT.A.GC..G.....T.....
D. persimilis -----CACCACTCG.....C..T..C.....G.....CAG.....AT..T..TT.A.GC..G..T.....T.....
D. virilis -----T.C.....GT.....CAGCA..C..A..GA.G...TGG.AG.TGCACCGCCGCCA--
D. mojaviensisC.....G.....T.....G..A..A..A..G..C...AG.AG..GCAGCAGCAGGAGC
D. ananassae ACGCCGCCACGGA...T.....C..C..C..G.....A..G..A..T..A..C...G..GG.AC.AC-----
B. dorsalis -----TTG...TCAT...TTGA..CAAAAT..T---TGC.AGT..T..A..T..A.....T..G..G..A-----
B. tryoni -----TTG...TCAT...TTGA..CAAAAT..T---TGC.AGT..T..A..T..A.....T..T..G..G..T-----
B. oleae -----TTG...TCAT...TTGA..CAGAAT..T---TGC.AGC..T..A..T..A.....T..T..G..G..A-----
C. capitata -----TTA..T..TCAC...TTT.A..CCAAAT..T---TGC.ACT...C..T..A.....T..T..G..TG..T-----
A. sororcula (86-S) -----TTA..T..TCAT...C..TTGA..CCAGAAT..T---TGC.ACT..T...T...T..T..C..G..AG..A-----
A. obliqua (32-O) -----TTA..T..TCAT...TTGA..CCAGAAT..T---TGC.ACT..T...T...T..T..C..G..AG..A-----
A. fraterculus (66-F) -----TTA..T..TCAT...TTGA..CCAGAAT..T---TGC.ACT..T...T...T..T..C..G..AG..A-----

510 520 530 540 550 560 570 580 590 600

B. oleae -----
C. capitata -----
A. sororcula (86-S) -----
A. obliqua (32-O) -----
A. fraterculus (66-F) -----

	810	820	830	840	850	860	870	880	890	900				
<i>D. yakuba</i>	CAGTGGG	---GGCCC	GGCCAGC	CAGCAGCT	CTGGCGGT	GGAGCCCC	CAGTTC	CAGCAAT	CGGGCAG	CGGCCACT	TCGAGCAG	CGGATCC	CAGCAGT	GGAGGA
<i>D. melanogaster</i>T	---.TCCACACACTCTT
<i>D. simulans</i>T	---.TCCACTAACTCTTT
<i>D. sechellia</i>T	---.TGCCACTACTCTTTT
<i>D. erecta</i>T	---.TTCACTCCCCCCC
<i>D. pseudoobscura</i>	A..CATT	---.A	..A..T	..C..A	..A..C	..AG	..TCGAG	..TGGC	..C	..TT	---.A	..A	..C	..A
<i>D. persimilis</i>	A..CATT	---.A	..A..T	..C..A	..A..C	..AG	..TCGAG	..TGGC	..C	..TT	---.A	..A	..C	..A
<i>D. virilis</i>	.G.CA	.CGTGA	.G.A	..CG	..C	..T	..C	---.GAG	..TCCAG	..AT	---.TG	..AA	..TG	..C
<i>D. mojavensis</i>	.G.CA	.CGTG	.G.A	..C	..C	..---	..G	..GGGT	..CGGTGT	..---	..TT	..GT	..GC	..C
<i>D. ananassae</i>	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>B. dorsalis</i>	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>B. tryoni</i>	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>B. oleae</i>	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>C. capitata</i>	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>A. sororcula</i> (86-S)	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>A. obliqua</i> (32-O)	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT
<i>A. fraterculus</i> (66-F)	-----CC	..T	..G	..CG	..AG	..TCCAGCCA	..C	..TCCT

	910	920	930	940	950	960	970	980	990	1000				
<i>D. yakuba</i>	GGAGGAGG	AGGCGCAG	GGGGGAGT	TCGGGAGG	CGGAGT	AGGAGGT	GGTAGAT	CCTCG	-----	GGGACAT	CGGTGAT	CACTAG	CGCCGAT	-----
<i>D. melanogaster</i>GC	..C	..T	..G	..C	..G	..A	-----	..AAAA	-----
<i>D. simulans</i>GC	..C	..T	..G	..C	..G	..A	-----	..AAAA	-----
<i>D. sechellia</i>GC	..C	..T	..G	..C	..G	..A	-----	..AAAA	-----
<i>D. erecta</i>C	..AC	..C	..C	..C	..G	..A	-----	..AAAA	-----
<i>D. pseudoobscura</i>	..T	..TT	..T	..A	..GC	..AA	..CG	..AAGC	..CA	..CAGC	..C	..T	..G	..CGGG
<i>D. persimilis</i>	..T	..TT	..T	..AAAC	..A	-----	..AGT	..A	..CAGC	..C	..T	..G	..CGGG	..CGGCAGTGGC
<i>D. virilis</i>	..CC	..CA	..C	..T	-----	..A	..CA	..CT	..GC	..CT	..CA	..C	..CGGAGGC	-----
<i>D. mojavensis</i>	A	..C	..CTCCA	..CA	..CT	..GC	..CT	..CA	..C	..CGGAGGC	-----	..C
<i>D. ananassae</i>CT	..CCC	..G	..TCA	-----	..GGT	..C	..ATCC	..CT	..A	..ACCCCGGG	..AGGCAGCAGC	..C	..C
<i>B. dorsalis</i>	..C	..TAAT	..TTAGCA	..TA	..TG	..AAC	..C	..TAT	..CC	..T	..AATAG	..T	..TG	..C
<i>B. tryoni</i>	..C	..AAT	..TTAGCA	..TA	..T	-----	..T	..TAT	..CC	..T	..AATAG	..T	..TG	..T
<i>B. oleae</i>	..C	..AAT	..TTAGCA	..TA	..TG	..AAC	..C	..TAT	..CC	..T	..AATAG	..T	..TG	..T
<i>C. capitata</i>	A	..C	..G	..T	..T	-----	..TTT	..CC	..T	..AT	..G	..T	..TG	..C
<i>A. sororcula</i> (86-S)	A	..C	..G	..T	..T	-----	..GGT	..C	..TATGT	..CACAGT	..C	..GA	..C
<i>A. obliqua</i> (32-O)	A	..C	..G	..T	..T	-----	..GGT	..C	..TATGT	..CACAGT	..C	..GA	..C
<i>A. fraterculus</i> (66-F)	A	..C	..G	..T	..T	-----	..GGT	..C	..TATGT	..CACAGT	..C	..GA	..C

D. yakuba ----CATCAC
D. melanogaster ---.C...
D. simulans ---.C...
D. sechellia ---.C...
D. erecta ---.....

<i>D. pseudoobscura</i>	TGTCACG
<i>D. persimilis</i>	TGTCACG
<i>D. virilis</i>	--..A...
<i>D. mojavensis</i>	--.....
<i>D. ananassae</i>	--..C...
<i>B. dorsalis</i>	AA.....
<i>B. tryoni</i>	AA.....
<i>B. oleae</i>	AA.....
<i>C. capitata</i>	AA.....
<i>A. sororcula</i> (86-S)	AA.....T
<i>A. obliqua</i> (32-O)	AA.....T
<i>A. fraterculus</i> (66-F)	AA.....T


```

in_3-o.18 ....G.....-----..--.....-..C.....-.....
in_3-o.20 ....G.....-----..--.....-..C.....-.....
in_3-o.20 ....G.....-----..--.....-..C.....-.....
in_3-o.33 ....G.....-----..--.....-..C.....-.....
in_3-s.21 ....G.....-----..--.....-..C.....-.....
in_3-s.23 ....G.....-----..--.....-..C.....-.....
in_3-s.23 ....G.....-----..--.....-..C.....-.....
in_4-f.11 ....G.....-----..--.....-..G.....-.....
in_5-f.11 ....G.....-----..--.....-..C.....-.....
in_6-f.27 ....G.....-----..--.....-..A.....-.....
in_7-f.5 ....G.....-----..--.....-..T.....-.....
in_7-s.5 ....G.....-----..--.....-..T.....-.....
in_8-f.5 ....G.....-----..--.....-..T.....-.....
in_9-f.25 ....G.....-----..--.....-..T.....-.....
in_9-f.25 ....G.....-----..--.....-..T.....-.....
in_10-f.31 ....G.....-----..--.....-..T.....-..G.....-.....
in_11-f.31 ....G.....-----..--.....-..T.....-..G.....-.....
in_12-f.30 ....G.....-----..--.....-..TATAT.....-..G.....-..A.....-.....
in_13-f.6 ....G.....-----..--.....-..TATAT.....-..A.....-.....
in_13-f.6 ....G.....-----..--.....-..TATAT.....-..A.....-.....
in_14-f.28 ....G.....-----..--.....-.....-.....-.....
in_14-f.28 ....G.....-----..--.....-.....-.....-.....
in_14-f.23 ....G.....-----..--.....-.....-.....-.....
in_14-f.7 ....G.....-----..--.....-.....-.....-.....
in_14-f.7 ....G.....-----..--.....-.....-.....-.....
in_14-f.7 ....G.....-----..--.....-.....-.....-.....
in_14-o.8 ....G.....-----..--.....-.....-.....-.....
in_14-o.9 ....G.....-----..--.....-.....-.....-.....
in_14-s.11 ....G.....-----..--.....-.....-.....-.....
in_14-s.24 ....G.....-----..--.....-.....-.....-.....
in_14-s.24 ....G.....-----..--.....-.....-.....-.....
in_15-f.23 ....G.....-----..--.....-..T.....-T.....-.....
in_16-f.9 ....G.....-----..--.....-.....-..A.....C.....-.....
in_17-f.9 ....G.....-----..--.....-.....-T.....-.....-.....
in_18-f.7 ....G.....-----..--.....-.....-G.....-.....-.....
in_19-f.29 ....G.....-----..--.....-..C.....-..T.....-.....
in_20-f.29 ....G.T.....-----..--.....-.....-.....-.....
in_21-f.29 ....G.....-----..--.....-.....-.....-..C.....-.....
in_22-f.32 ....G.....-----..--.....-.....-.....-.....G.....-.....
in_23-o.2 ....G.....-----..--.....-..C.....-..C.....-.....
in_24-o.19 ....G.....-----..--.....-..C.....-.....-.....
in_25-o.14 ....G.....-----..--.....-.....-.....-.....
in_26-o.16 ....G.....-----..--.....-..C.....-.....-.....
in_27-o.9 ....G.....-----..--.....-..C.....-.....-.....
in_28-o.9 ....G.....-----..--.....-.....-T.....-.....
in_28-s.22 ....G.....-----..--.....-.....-T.....-.....
in_28-s.22 ....G.....-----..--.....-.....-T.....-.....

```

```

in_29-o.3   ....G.....-----..--.....-G.C.....-.....
in_30-o.2   ....G.G.....-----.....T.....-.....
in_31-s.11  .....G.....G.....-----.....-.....A.....C.....
in_32-s.11  .....G.....G.....-----.....T.....-.....
in_33-s.5    ....G.....-----.....T.....-.....
in_34-s.21  .....C.....-----.....-.....C.....
in_35-s.21  .....C.....-----.....-.....
in_36-s.23  ....G....C.....-----.....-.....T-.....A.....
in_36-s.23  ....G....C.....-----.....-.....T-.....A.....
in_37-s.20  .....C.....-----.....C-..A.....-.....
in_38-s.20  .....C.....-----.....-..A.....-.....

```


41-o.8
 42-o.8
 43-o.8G.....
 44-o.8
 45-o.8
 46-o.8
 47-o.8
 48-o.19
 49-o.19
 50-o.10
 51-o.10
 52-o.12
 53-o.12G.....
 54-o.12
 38-o.13
 55-o.14
 56-o.14
 57-o.4A.....G.....
 58-o.4A.....G.....
 59-o.4G.....
 60-o.4
 61-o.15A.....
 62-o.15
 63-o.16A.....
 64-o.16T.....
 65-o.9
 66-o.9T.....
 67-o.9
 68-o.9
 69-o.3G.....
 70-o.3G.....
 71-o.18
 72-o.2
 73-o.2
 74-o.20A.....T.....
 75-o.20T.....
 76-o.33
 77-s.11
 78-s.11
 79-s.11C.....
 80-s.24C.....
 81-s.24C.....
 82-s.5
 83-s.5
 84-s.21G.....
 85-s.21
 86-s.21

32-f.29 T. T.
 33-f.29 T. T.
 34-f.29
 35-f.29
 36-f.32 G.
 37-f.32
 38-o.2 T. T.
 39-o.2 T. T.
 40-o.8 T. T.
 41-o.8 T. T.
 42-o.8 T. T.
 43-o.8 T. T.
 44-o.8 T. T.
 45-o.8 G.
 46-o.8 T. T.
 47-o.8 T. T.
 48-o.19 T. T.
 49-o.19 T. T.
 50-o.10 T. T.
 51-o.10 T. T.
 52-o.12 T. T.
 53-o.12 T. T.
 54-o.12 T. T.
 38-o.13 T. T.
 55-o.14 T. T.
 56-o.14 T. TA.
 57-o.4 T. T.
 58-o.4 T. T.
 59-o.4 T. T.
 60-o.4 T. T.
 61-o.15 T. T. C.
 62-o.15 T. T.
 63-o.16 T. T.
 64-o.16 A. T. T.
 65-o.9 C. A.
 66-o.9 G.
 67-o.9 T.
 68-o.9 A.
 69-o.3 T. T.
 70-o.3 T. T.
 71-o.18 G. T. T.
 72-o.2 T.
 73-o.2
 74-o.20 C. T. T.
 75-o.20 T. T. C.
 76-o.33 T. T.
 77-s.11

25-f.23
 27-f.9T.....
 28-f.7
 29-f.7A.....
 30-f.7
 31-f.7
 32-f.29G.....C.....
 33-f.29C.....C.....C.....C.....
 34-f.29C.....
 35-f.29
 36-f.32T.....
 37-f.32A.T.....
 38-o.2C.....
 39-o.2C.....
 40-o.8C.....
 41-o.8
 42-o.8CT.....
 43-o.8C.....
 44-o.8C.....
 45-o.8
 46-o.8CT.....
 47-o.8C.....
 48-o.19C.....
 49-o.19C.....
 50-o.10
 51-o.10C.....T.....
 52-o.12G.....C.....
 53-o.12C.....
 54-o.12C.....C.....
 38-o.13C.....
 55-o.14C.....
 56-o.14C.....
 57-o.4C.....C.....
 58-o.4C.....G.....
 59-o.4C.....
 60-o.4C.....
 61-o.15C.....G.....
 62-o.15C.....C.....
 63-o.16C.....
 64-o.16C.....C.....
 65-o.9T.....
 66-o.9
 67-o.9
 68-o.9
 69-o.3C.....G.....
 70-o.3C.....
 71-o.18C.....C.....

72-o.2C.....
 73-o.2
 74-o.20C.....
 75-o.20C.....
 76-o.33C.....C.....
 77-s.11
 78-s.11
 79-s.11C.....
 80-s.24
 81-s.24
 82-s.5C.....
 83-s.5
 84-s.21C.....
 85-s.21C.....
 86-s.21C.....
 87-s.21A.T.....
 88-s.23
 89-s.23
 90-s.23C.....G.....
 91-s.23C.....
 92-s.22C.....
 93-s.22C.....
 94-s.20
 95-s.20

310 320 330 340 350 360 370 380 390 400

1-f.1 GGCGAACGTTACCGATTACAGCGGCCCAATGGCTGCTGCAGTGGCGGCTGCGGCTACACGTAGCGCCAGCGCTGATGCGCTAGGCGGTTTCGACGAATA
 1-f.9
 2-f.1
 3-f.11
 3-o.13
 4-f.11T.....
 5-f.11T.....
 6-f.27
 7-f.27
 8-f.5
 9-f.5C.....
 10-f.25
 11-f.25
 12-f.26
 13-f.26
 14-f.31
 15-f.31
 16-f.30A.....
 17-f.30A.....
 18-f.30

19-f.30
 20-f.6
 21-f.6T.....
 22-f.28
 23-f.28
 24-f.23G.....A.....
 25-f.23G.....
 27-f.9A.....
 28-f.7T.....C.....
 29-f.7T.T.....T.....a.....
 30-f.7T.....C.....
 31-f.7T.....C.....
 32-f.29A.....
 33-f.29
 34-f.29T.....
 35-f.29T.....
 36-f.32
 37-f.32
 38-o.2A.....
 39-o.2
 40-o.8
 41-o.8
 42-o.8T.A.....
 43-o.8A.....T.....
 44-o.8G.....C.....A.....
 45-o.8
 46-o.8T.A.....
 47-o.8A.....T.....G.....
 48-o.19A.....T.....
 49-o.19
 50-o.10
 51-o.10G.....
 52-o.12A.....
 53-o.12
 54-o.12T.A.....
 38-o.13A.....
 55-o.14C.....
 56-o.14C.....
 57-o.4A.....
 58-o.4A.....C.....
 59-o.4C.....
 60-o.4
 61-o.15A.....
 62-o.15A.....
 63-o.16GC.....
 64-o.16
 65-o.9

66-o.9
 67-o.9
 68-o.9
 69-o.3A.....
 70-o.3
 71-o.18A.....
 72-o.2
 73-o.2T.....
 74-o.20
 75-o.20
 76-o.33A.....
 77-s.11
 78-s.11
 79-s.11T.....
 80-s.24
 81-s.24T.....
 82-s.5T.....
 83-s.5
 84-s.21
 85-s.21
 86-s.21T.....A.....
 87-s.21
 88-s.23C.....
 89-s.23G.....G.....
 90-s.23
 91-s.23
 92-s.22G.....
 93-s.22G.....
 94-s.20A.....T.....A.....
 95-s.20

.....410.....420.....430.....440.....450.....460.....470.....480.....490.....500.....

1-f.1 CAGGCTGCACCGGTAGCGGTAGTACCGTTGGTGGTGGAGTGTTAAATAGTCGCAATAGTGATGCGAATAGTGAGCGTGGTGGCGGTGGTGGTGGTGCGGG
 1-f.9
 2-f.1C.....A.....
 3-f.11AA.....
 3-o.13AA.....
 4-f.11C.....
 5-f.11C.....C.....
 6-f.27A.....
 7-f.27C.....
 8-f.5
 9-f.5
 10-f.25C.....
 11-f.25A.....C.....
 12-f.26T.....

13-f.26C.....T.....
14-f.31
15-f.31
16-f.30C.....
17-f.30C.....
18-f.30
19-f.30G.....
20-f.6
21-f.6A.....C.....
22-f.28C.....A.....
23-f.28C.....C.....
24-f.23C.....G.....C.....
25-f.23C.....
27-f.9
28-f.7C.....C.....
29-f.7C.....
30-f.7C.....
31-f.7C.....
32-f.29AA.....
33-f.29C.....AA.....
34-f.29A.....C.....
35-f.29A.....C.....
36-f.32
37-f.32C.....C.....A.....
38-o.2AA.....
39-o.2AA.....
40-o.8
41-o.8AA.....
42-o.8A.....T.....
43-o.8A.....A.....T.....
44-o.8C.....
45-o.8---.....
46-o.8A.....T.....
47-o.8A.....A.....T.....
48-o.19
49-o.19T.....AA.....
50-o.10
51-o.10
52-o.12T.....
53-o.12T.....
54-o.12AA.....
38-o.13AA.....
55-o.14AA.....
56-o.14AA.....
57-o.4C.....T.....
58-o.4C.....T.....
59-o.4AA.....

60-o.4T.....AA.....
61-o.15CAA.....
62-o.15T.....
63-o.16AA.....
64-o.16AA.....
65-o.9C.....
66-o.9T.....
67-o.9
68-o.9C.....
69-o.3AA.....
70-o.3AA.....
71-o.18AA.....
72-o.2A.....AA.....
73-o.2 T.....T.....
74-o.20C.....AA.....
75-o.20C.....AA.....
76-o.33AA.....
77-s.11T.....G.....
78-s.11T.....
79-s.11C.....
80-s.24-----
81-s.24-----
82-s.5C.....
83-s.5C.....T.....-----
84-s.21A.....T.....C.....
85-s.21AA.....
86-s.21C.....
87-s.21A.....C.....A.....C.....
88-s.23C.....C.....C.....
89-s.23A.....
90-s.23T.....AA.....
91-s.23T.....AA.....
92-s.22
93-s.22
94-s.20C.....
95-s.20C.....

510 520 530 540 550 560 570 580 590 600
1-f.1 AAATGGAGGCAGCAGTGGCGGGCGCAGCAGCAGC-----ATGGTCGCGATCGCGCCGACAGTCGTGATGAACTCATGCAGCTAGAC
1-f.9
2-f.1A.....
3-f.11T.....
3-o.13T.....
4-f.11G.....
5-f.11
6-f.27

7-f.27T.....	-----
8-f.5	-----
9-f.5	-----
10-f.25T.....	-----
11-f.25T.....	-----
12-f.26	-----
13-f.26	-----
14-f.31	-----
15-f.31	-----
16-f.30T.....T.....	-----
17-f.30	-----
18-f.30T.....T.....	-----
19-f.30T.....T.....	-----
20-f.6	-----C.....
21-f.6	-----
22-f.28A.....	-----
23-f.28	-----
24-f.23	-----
25-f.23	-----
27-f.9	-----
28-f.7	-----
29-f.7	-----
30-f.7	-----
31-f.7	-----
32-f.29	-----
33-f.29T.....	-----
34-f.29	-----
35-f.29	-----
36-f.32	-----
37-f.32	-----
38-o.2	-----
39-o.2	-----
40-o.8G.....A.....	-----
41-o.8	-----
42-o.8T.....	-----
43-o.8T.....	-----A.....
44-o.8	-----
45-o.8	-----C.....
46-o.8T.....	-----
47-o.8T.....	-----A.....
48-o.19T.....	-----
49-o.19T.....T.....	-----
50-o.10A.....	-----
51-o.10A.....	-----
52-o.12T.....	-----
53-o.12T.....	-----
54-o.12	-----

```

38-o.13 .....-----.....
55-o.14 .....-----.....
56-o.14 .....-----.....
57-o.4 .....T.....
58-o.4 .....T.....
59-o.4 .....-----.....
60-o.4 .....T.....
61-o.15 .....T.....
62-o.15 .....T.....
63-o.16 .....-----.....
64-o.16 .....-----.....
65-o.9 .....T.....
66-o.9 .....G.....G.....
67-o.9 .....C.....
68-o.9 .....C.....
69-o.3 .....T.....A.....
70-o.3 .....T.....
71-o.18 .....-----.....
72-o.2 .....-----.....
73-o.2 .....TGGCGGGCGCAGTAGCAGC.....
74-o.20 .....T.....
75-o.20 .....T.....
76-o.33 .....G.....
77-s.11 .....T.....
78-s.11 .....T.....
79-s.11 .....T.....
80-s.24 .....A.....
81-s.24 .....A.....A.....G.....
82-s.5 .....T.....T.....
83-s.5 .....T.....T.....
84-s.21 .....T.....
85-s.21 .....-----.....
86-s.21 .....T.....
87-s.21 .....-----.....
88-s.23 .....A.....
89-s.23 .....T.....
90-s.23 .....T.....
91-s.23 .....T.....
92-s.22 .....-----.....C.....
93-s.22 .....-----.....
94-s.20 .....T.....
95-s.20 .....T.....

```

```

610 620 630 640 650 660 670 680 690 700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

```

1-f.1 TATAGTAACAAGGATAACAGAGATCGTGATAGAGAGATGTCCACCACGCCGGTGGATCATATAGGTAGTAATAAGCGAAGACGTAAGAACTCATCCAAC
1-f.9 .....

```

2-f.1
 3-f.11 T.....
 3-o.13 T.....
 4-f.11 G.....
 5-f.11
 6-f.27
 7-f.27
 8-f.5
 9-f.5 A.....
 10-f.25
 11-f.25
 12-f.26 T.G.....
 13-f.26
 14-f.31
 15-f.31
 16-f.30
 17-f.30 C.....
 18-f.30
 19-f.30
 20-f.6
 21-f.6
 22-f.28
 23-f.28 G.....
 24-f.23
 25-f.23
 27-f.9 G.....
 28-f.7 G.....
 29-f.7 G.....
 30-f.7
 31-f.7
 32-f.29 T.....
 33-f.29 AT.....
 34-f.29
 35-f.29
 36-f.32 C.....
 37-f.32
 38-o.2 T.....
 39-o.2 T.....
 40-o.8
 41-o.8 T.....
 42-o.8 T.....
 43-o.8 T.....
 44-o.8
 45-o.8 T.G.....
 46-o.8
 47-o.8 T.....
 48-o.19 T.....

49-o.19 T
 50-o.10
 51-o.10
 52-o.12 T
 53-o.12 T A
 54-o.12 T
 38-o.13 T
 55-o.14 T
 56-o.14 T
 57-o.4 T
 58-o.4 T
 59-o.4 T
 60-o.4 T
 61-o.15 T
 62-o.15 T
 63-o.16 T
 64-o.16 T
 65-o.9
 66-o.9
 67-o.9
 68-o.9
 69-o.3 T
 70-o.3 G T
 71-o.18 T
 72-o.2
 73-o.2
 74-o.20 T
 75-o.20 T
 76-o.33 T
 77-s.11
 78-s.11
 79-s.11
 80-s.24
 81-s.24
 82-s.5
 83-s.5
 84-s.21 T
 85-s.21 T
 86-s.21
 87-s.21
 88-s.23 GT
 89-s.23 G
 90-s.23 C T
 91-s.23 T
 92-s.22 T G
 93-s.22 T
 94-s.20

95-s.20

710 720

1-f.1 GTGATAATTCGTTGACCTCGACACAT
1-f.9
2-f.1
3-f.11
3-o.13
4-f.11
5-f.11
6-f.27
7-f.27
8-f.5
9-f.5
10-f.25
11-f.25
12-f.26
13-f.26
14-f.31
15-f.31
16-f.30
17-f.30T.....
18-f.30
19-f.30
20-f.6
21-f.6
22-f.28
23-f.28
24-f.23
25-f.23
27-f.9
28-f.7
29-f.7
30-f.7C.....
31-f.7
32-f.29
33-f.29
34-f.29
35-f.29
36-f.32
37-f.32
38-o.2
39-o.2
40-o.8
41-o.8
42-o.8

43-o.8
44-o.8
45-o.8
46-o.8
47-o.8
48-o.19
49-o.19
50-o.10
51-o.10
52-o.12
53-o.12
54-o.12
38-o.13
55-o.14
56-o.14
57-o.4
58-o.4
59-o.4
60-o.4
61-o.15
62-o.15
63-o.16
64-o.16G.....
65-o.9
66-o.9
67-o.9
68-o.9
69-o.3
70-o.3
71-o.18
72-o.2
73-o.2
74-o.20
75-o.20
76-o.33
77-s.11
78-s.11
79-s.11
80-s.24
81-s.24
82-s.5
83-s.5
84-s.21
85-s.21
86-s.21
87-s.21
88-s.23

89-s.23
90-s.23
91-s.23
92-s.22
93-s.22
94-s.20A
95-s.20

Anexo 5. Alinhamento de *fruitless*. Dados macroevolutivos. Os pontos “.” representam base idênticas às da sequência de referência; traços “-” indicam ausência de bases.

```

          10      20      30      40      50      60      70      80      90      100
D. pseudoobscura  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gtgcgcggtctcacagataacaacaatctgaactatcggtccgactgcgacaagctgcgcgactcggccgccagctccccaccggccgaggaccctcct
D. melanogaster   ....t.....c.c.....t....g....t.g.g....a..t.g.gagta
D. yakuba         ....t.....c.c.....t.t.....g....t.g.g....a..t.g.gagta
D. virilis       ....c.....c.....c.....t....t....t.....g....t.g.g....a..t.g.gag-
D. grimshawi     ..t.....c.....t....t.....g....t.g.g....t.g.t..t..ag-
D. simulans      ....t..t..t.....c.c.....t....g....t.g.g....a..t.g.gagta
D. sechellia     ....t.....c.c.....t....g....t.g.g....a..t.g.gagta
D. erecta        ....t.....c.c.....t....t.....g....t.g.g....a..t.g.gagta
D. silvestris    ..t.....c.....t....t.....t....g....t.g....t.g.t..t..ag-
D. heteroneura   ..t.....c.....t....t.....t....g....t.g....t.g.t..c..ag-
A. fraterculus (1-f.1) ..T.T.....G.C.....C.C.T....TA.G.GCAT-----C.ATATC..G.....AC.....
A. obliqua (39-o.2) ..T.T.....G.C.....C.C.T....TA.G.GCAT-----C.ATATC..G.....AC.....
A. sororcula (93-s.22) ..T.T.....G.C.....C.C.T....TA.G.GCAT-----C.ATATC..G.....AC.....

          110     120     130     140     150     160     170     180     190     200
D. pseudoobscura  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
ccagctatggggggcgccctcggg-----gtggcgacgggtgcgagactca--cgcgactccctgctcc-----cgctgcgaaacg
D. melanogaster   at---.c---act.....g.cggcgtgggggc...c..t.c.a...c.a.c-----
D. yakuba         at---.c---ac.....g.cggcgtgggggc...c..t.c.a...c...c-----
D. virilis       -----t.....g.c-----c..t--.....t.....
D. grimshawi     -----c-----c..t--.....c.t.....t.....
D. simulans      at---.c---ac.....g.cggcgtgggggc...c..t.c.a...c.a.c-----a.....
D. sechellia     at---.c---ac.....g.cggcgtgggggc...c..t.c.a...c.a.c-----a.....
D. erecta        at---.c---ac.....g.cggcgtgggggc...c..t.c.a...c...c-----a.t.....
D. silvestris    -----c-----c..t--.....c.....t.....
D. heteroneura   -----c-----c..t--.....c.....t.....
A. fraterculus (1-f.1) -----CA.TA.T....T....CGGCATCGGC-----AATGT.A.ACA..C..A-----T..TCAA.A..TACGTGGTTCGTGC.AC.....
A. obliqua (39-o.2) -----CA.TA.T....T....CGGCATCGGC-----AATGT.A.ACA..C..A-----T..TCAA.A..TACGTGGTTCGTGC.AC.....
A. sororcula (93-s.22) -----CA.TA.T....T....CGGCATCGGC-----AATGT.A.ACA..C..A-----T..TCAA.A..TACGTGGTTCGTGC.AC.....

          210     220     230     240     250     260     270     280     290     300
D. pseudoobscura  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
cgacctgcgcgacgagctgtcgcagcgc---agcagcagc-----
D. melanogaster   g..t.....a.....---.....agcatgagcgaacgcagctcg-----gcggcagcagcggcggcggcggcagca
D. yakuba         g..t.....a.....---.....agcatgagcgaacgcagctcg-----gcggcagcagcggcggcggcggcagca
D. virilis       .....t..t..ca.gca.agc.....ct.agcctaagtgaacgtagctcc-----gccgccggcggcagctgctgcccg
D. grimshawi     -.t.a.....tt.ca.gca.t---.....ttaagcctgagcgaacgcaactcg-----gcagcagctgccgccgcgagcagcagcggct
D. simulans      g..t.....a.....---.....agcatgagcgaacgcagctcg-----gcggcagcagcggcggcggcggcagca
D. sechellia     g..t.....a.....---.....agcatgagcgaacgcagctcg-----gcggcagcagcggcggcggcggcagca
D. erecta        g..t.....a.....---.....agcatgagcgaacgcagctcg-----gcggcagcagcggcggcggcggcagca
D. silvestris    -.t.....ta.ca.gca.---.....ttaagcctgagcgaacgcaactcggaagcagctgccgccgccgcgagcagcagcggct

```



```

D. silvestris      ....-----g.....a.a.t.g.g...-----
D. heteroneura    ....-----g.....a.a.t.g.g...-----
A. fraterculus (1-f.1) A.GCTGCACCGGTAGCGGTAGTACCGTTGGTGGTGGAGTGTTAAATAGTCGCAAT..T..GCGAAT..T....T..TGG.GGT.GTGGTGGTGC GGGA
A. obliqua (39-o.2)  A.GCTGCACCGGTAGCGGTAGTACCGTTGGTGGTGGAGTGTTAAATAGTCGCAAT..T..GCGAAT..T....T..TGG.GGT.GTAATGGTGC GGGA
A. sororcula (93-s.22) A.GCTGCACCGGTAGCGGTAGTACCGTTGGTGGTGGAGTGTTAAATAGTCGCAAT..T..GCGAAT..T....T..TGG.GGT.GTGGTGGTGC GGGA

                610      620      630      640      650      660      670      680      690      700
D. pseudoobscura  -----cgggcccagcagtcgcgacgagctgttcagctcgattacagcaccaggacaaca
D. melanogaster   -----a.g.t.....t.t.a.....g....t...a.....t...
D. yakuba         -----a.g.t.....t.t.....g.....a.....t...
D. virilis        -----t.t.....g.t.....c.....g.c.....t.....t.t.
D. grimshawi      -----t.t.....g.t.....c.....g.c.....t.....t...
D. simulans       -----a.g.t.....t.t.....g....t...a.....t...
D. sechellia      -----a.g.t.....t.t.....g....t...a.....t...
D. erecta         -----a.g.....t.t.....g....t...a.....t...
D. silvestris     -----t.t...g.t....c....t.g.c....t.t....t...
D. heteroneura    -----t.t...g.t....c....t.g.c....t.t....t...
A. fraterculus (1-f.1) AATGGAGGCAGCAGTGGCGGGCGCAGCAGCAGCATTGGTCGCGAT..C.....T.T.A.CA.....A.C.T.T.A.....T...-
A. obliqua (39-o.2)  AATGGAGGCAGCAGTGGCGGGCGCAGCAGCAGCATTGGTCGCGAT..C.....T.T.A.CA.....A.C.T.T.A.....T...-
A. sororcula (93-s.22) AATGGAGGCAGCAGTGGCGGGCGCAGCAGCAGCATTGGTCGCGAT..C.....T.T.A.CA.....A.C.T.T.A.....T.T-

                710      720      730      740      750      760      770      780      790      800
D. pseudoobscura  acaacagcaacatcagcaacaccagcaacaacaataacaacaacagcagcaacaacaacaacat-----
D. melanogaster   ....t.....g..t.gt..g.gg.....c.....t.at.at.a.....t.g..gcagcaac---aataacaacagcagcagcaatag
D. yakuba         ....t.....g..t.gt..g.gg.....c.....t.t.a.a.at.g.g.g..t..caacaac-----agcagcaatagggagcg
D. virilis        .t.t.t...g..a....a.a.....c.....t.at.....g..gt.g.g..caataac-----aacaacaacaacagc
D. grimshawi      ....t.t...a..a....a.a.....t.gc.c....g..t...a.....t.....caacagc-----
D. simulans       ....t.....g..t.gt..g.gg.....c.....t.at.a.a.....t.g..gcagcaac---aacaacaacagcagcagcaatag
D. sechellia      ....t.t...g..t.gt..g.gg.....c.....t.a.a.a.....t.g..gcagcaac---aacaacaacagcagcagcaatag
D. erecta         ....t.....g.....gt..g.g.....c.....t.t.a.a.a...t.g.g.g..caacaacaacaacagcagcagcaatagggagcg
D. silvestris     ....t.t...g..a....a.a.....c.....t.g.c.a...gt.g.....aacaac-----
D. heteroneura    ....t.t...g..a....a.a.....c.....t.c.a.g.gt.g.....caataac-----
A. fraterculus (1-f.1) -----
A. obliqua (39-o.2) -----
A. sororcula (93-s.22) -----

                810      820      830      840      850      860      870      880      890      900
D. pseudoobscura  ----cgggagcgggaacgggagcgggaaagagaacgagagcgggacagcagcagggaccgagagggagctctccaccaccccggtggaccagctgagt
D. melanogaster   ggag..ca.caacagcg.c..a..t..gc.g.g....aa.a..gc.tgagc....a.g..c.....g.....g.....g.....
D. yakuba         caacaacagcg.a....t.....gc.g..a.....t..gc.tgagc.a..a.g..c.....a.....g.....g.....
D. virilis        cagca.ca.caaca.caaca.caaca.tc...g..ca.cgatag...aga...a..t..g..c.....g..g.c....t....t..
D. grimshawi      -----a.cagcaaca.caaca.tc...g..ca.cgatag...agat..a..t..t..c.....g.....g.g.c.a..t....t..
D. simulans       ggag..ca.caacagcg.c..a..t..gc.g.g.g...a.a..gc.tgagc....a.g..c.....g.....g.....g.....
D. sechellia      ggag..ca.caacagcg.c..a..t..gc.g.g.g...a.a..gc.tgagc....a.g..c.....g.....g.....g.....

```

D. erecta caacaacagcg.a.....t.....gc.g...a.....t.gc.tgagc.....a.g..c.....g.....g.....g.....g.....
D. silvestris ----aacagca.cagcaaca.caaca.tc....g..ca.cgatag...agat..a..t..g..c.....g.....g..a..c..t..t.....t..
D. heteroneura -aacaacagca.cagcaaca.caaca.tc....g..ca.cgatag...agat..a..t..g..t.....g.....g..g..c..t..t.....t..
A. fraterculus (1-f.1) ----- .A..T..T..T..A..A.G.....G.....T..TA.AG..
A. obliqua (39-o.2) ----- .A..T..T..T..A..T.G.....G.....T..TA.AG..
A. sororcula (93-s.22) ----- .A..T..T..T..A..A.G.....G.....T..TA.AG..

910 920 930 940 950 960 970 980 990 1000

D. pseudoobscura agtagtaagcgcagacgtaagaactcatcatccaactgtgataactcgtctgtcctcgagccaccatgCGAACGCGGCCCATGCCAGGACAGGCCTACC
D. melanogaster
D. yakuba
D. virilist..t.....t.....cata.....a.....g
D. grimshawig.....t..t.....t.....t.....cata.....a.....g
D. simulans
D. sechellia
D. erecta
D. silvestrisg.....t..t.....t.....t.....cata.....a.....g
D. heteroneurag.....t..t.....t.....t.....cata.....a.....g
A. fraterculus (1-f.1) ...A.....A.....C---.....T..T..A.....CA..T-----
A. obliqua (39-o.2) ...A.....A.....C---.....T..T..A.....CA..T-----
A. sororcula (93-s.22) ...A.....A.....C---.....T..T..A.....CA..T-----

....|...
D. pseudoobscura cgcaggac
D. melanogaster
D. yakuba
D. virilis
D. grimshawi
D. simulans
D. sechellia
D. erecta
D. silvestris
D. heteroneura
A. fraterculus (1-f.1) -----
A. obliqua (39-o.2) -----
A. sororcula (93-s.22) -----