

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

**Variações no gene *yolk* em moscas das frutas
do grupo *fraterculus***

Gustavo Castro de Oliveira

**SÃO CARLOS
2009**

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

**Variações no gene *yolk* em moscas das frutas
do grupo *fraterculus***

Gustavo Castro de Oliveira*

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Evolução da Universidade Federal de São Carlos como parte dos requisitos para a obtenção do Título de Mestre em Genética e Evolução, Área de Concentração.

**Orientador: Prof. Dr. Reinaldo
Alves de Brito**

***Bolsista Capes**

SÃO CARLOS

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

O48vg

Oliveira, Gustavo Castro de.

Variações no gene *yolk* em moscas das frutas do grupo *fraterculus* / Gustavo Castro de Oliveira. -- São Carlos : UFSCar, 2009.
80 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2009.

1. Genética de populações. 2. Mosca-das-frutas. 3. Grupo *fraterculus*. 4. *Yolk*. 5. Filogeografia. I. Título.

CDD: 575.15 (20^a)

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

**Variações no gene *yolk* em moscas das frutas do grupo
*fraterculus***

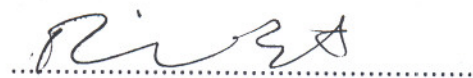
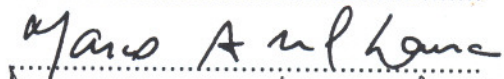
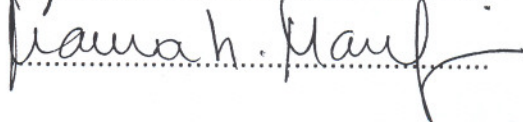
Dissertação de Mestrado de Gustavo Castro de Oliveira

Banca Examinadora

Prof. Dr. Reinaldo Alves de Brito

Prof. Dr. Marco Antonio Del Lama

Profa. Dra. Maura Helena Manfrin


.....

.....

.....

2009

Agradecimentos

Agradeço a todos que, direta ou indiretamente, contribuíram para o desenvolvimento deste trabalho.

Em especial:

A Deus, pela força em todos os momentos dessa árdua caminhada, por tudo de bom e especial que acontece em minha vida e pela oportunidade de ter pessoas maravilhosas ao meu redor.

Aos meus pais, pelo amor, carinho, educação e pela força e apoio incondicional que recebo em todos os momentos. Ao meu irmão, sempre disposto a ajudar, fortalecendo minha caminhada do dia a dia, amo vocês!!!

A toda minha família, pelo apoio, carinho, compreensão e orações a fim de que eu obtivesse essa conquista.

Ao meu orientador Reinaldo, pelo empenho, dedicação e paciência durante essa minha trajetória. Aprendi muita coisa nova e interessante e agradeço-lhe por acreditar em mim e em meu potencial.

Ao Programa de Pós-Graduação em Genética e Evolução da UFSCar.

À CAPES pelo apoio financeiro concedido.

Aos meus amigos(as) de graduação, que apesar do distanciamento, continuam sendo especiais e inesquecíveis. Nos momentos bons ou ruins a amizade sempre prevaleceu e vou guardar isso pra sempre em meu coração. Essa conquista também é de todos vocês.

À Andréa e ao Fernando pela amizade, incentivo, apoio, preocupação em todos os momentos do desenvolvimento de minha pesquisa.

Ao Ider, ao Daniel, por sempre me ajudarem nas horas de desespero, pela paciência e amizade.

Ao Bruno, Fer, Van, Marquinhos, André, Tati, Nelci, Carol, Aline, Victor, Clarissa, Simone, Isabela, Gizella, Nathália, Isabela (baratinha), Ider, Daniel, Éliton, Karina, Vanessa, João Paulo, entre todos... muito muito muito obrigado por tudo, pelas palavras de amizade, pelos conselhos, pela ajuda nas atividades de laboratório, pelos momentos de descontração (brilhantes), vocês são todos especiais e inesquecíveis, quero continuar mais um tempão com todos vocês. É uma alegria e um prazer muito grande fazer parte dessa turma, que é simplesmente sensacional!

Muito Obrigado! Sou uma pessoa muito feliz, pois estou cercado de pessoas maravilhosas e especiais!

1- Resumo

Espécies do grupo *fraterculus* estão relacionadas com os maiores danos a culturas de frutas carnosas por atacar frutos verdes e maduros indistintamente, o que as torna pragas de grande importância econômica. Contudo, estas espécies são de difícil distinção, com a existência potencial de diversas espécies crípticas. Dessa forma, o entendimento da biologia e, particularmente, da estrutura populacional desses insetos-praga tem grande importância para o desenvolvimento de novas estratégias de manejo. Neste trabalho, investigamos a variação no gene *yolk* em 37 indivíduos ao longo da distribuição do grupo no Brasil para conhecermos melhor o padrão estrutural das espécies do grupo *fraterculus*. Os dados indicam a existência de recombinação, de forma que analisamos a região amplificada separadamente por regiões distintas: 5', mid e 3'. Análises de polimorfismos nestas regiões apontaram para altos valores de diversidade nucleotídicas intra e interespecíficas para as três regiões gênicas em questão, sendo estes valores maiores para a região 3'. A metodologia da Análise dos Clados Aninhados (NCPA) foi utilizada para inferências de possíveis relações entre a configuração das redes haplotípicas com o padrão de distribuição geográfica considerando estas três regiões distintas. Através do uso desta metodologia, observamos dois eventos principais que podem estar influenciando a distribuição das espécies do grupo *fraterculus* no território nacional. A primeira inferência refere-se à expansão populacional no sentido Centro-Sul do Brasil, referente preferencialmente à espécie *Anastrepha fraterculus*. O segundo evento de inferência indica fluxo gênico restrito com isolamento por distância ao longo da distribuição global dos espécimes amostrados. A datação destes eventos indica uma congruência temporal, o que pode indicar que a não ocorrência de fluxo gênico em clados inferiores possa estar associado a restrições amostrais, uma vez que nossa amostragem é de fato restrita. Além disso, em populações mais próximas, o fluxo gênico pode estar ocorrendo com frequência suficiente para promover a homogeneização genética das populações. Os resultados obtidos mediante a análise das seqüências do gene *yolk* nos permitiu um melhor conhecimento acerca dos níveis de variação referentes a esse marcador, bem como determinar os processos que possam estar envolvidos no padrão de distribuição atual das espécies do grupo *fraterculus* no território nacional.

Palavras chave: *Anastrepha*, grupo *fraterculus*, *yolk*, Análise Filogeográfica dos clados aninhados

1- Abstract

Species of the *fraterculus* group are associated to the biggest damages to fruit crops due to the fact that they attack indiscriminately green and ripe fruits, making them pests of great economic importance. These species are hard to tell apart, even showing some potential cryptic species. The understanding of the populational structure and biology of this species group is of paramount importance to strategies of management and control of these pests. Here, we investigate variation on the gene *yolk* em 37 individuals sampled throughout the species distribution in Brazil seeking to better understand the populational structure of this species group in Brazil. Our data indicated existence of recombination, which lead us to analyze three different regions in the gene, 5', mid, and 3'. These regions show high levels of nucleotide diversity intra and interspecifically for all three gene regions investigated. The use of Nested Clade Phylogenetic Analysis independently used on these regions indicate two main results that occurred more than one throughout our analyses. The first is a range expansion from north-NE populations towards the south, mostly related to specimens of *Anastrepha fraterculus*. The second common event, detected eight times, is restricted gene flow with isolation by distance (IBD). Dating of such events indicated that they are temporarily congruent which might indicate that the lack of IBD in other levels of the analysis might be caused either by sampling limitations or an excess of local gene flow that tampers out as we move farther in space. Our results of the gene *yolk* have provided us with a better understanding of the levels of local variation of this marker of for the species that may help us determine the evolutionary processes that shaped the species group current distribution.

Key words: *Anastrepha*, group *fraterculus*, *yolk*, Nested Clade Phylogeographical Analysis

Lista de Figuras

Fig. 1 Representação dos enhancers envolvidos na ativação dos promotores dos genes <i>yk1</i> e <i>yk2</i>	8
Fig. 2 Representação esquemática da região amplificada do gene <i>yolk</i> , seus primers e regiões exônicas e intrônicas	18
Fig. 3 Representação dos clones positivos de diferentes indivíduos utilizados em reações de sequenciamento	25
Fig. 4 Eletroferograma mostrando polimorfismo na região marcada entre duas colônias do mesmo indivíduo (a) VAlta Cx4G4 colônia 1 (b) VAlta Cx4G4 colônia 2.....	26
Fig. 5 Eletroferograma mostrando polimorfismos na região marcada entre duas colônias de indivíduos distintos (a) V.Alta Cx4G4 colônia 2, (b) J. Pessoa Cx2G4 colônia 4)	26
Fig. 6- Representação esquemática dos genes <i>yolk</i> em <i>Drosophila</i> (<i>yk1</i> , <i>yk2</i> , <i>yk3</i>) e subdivisão em três partes do gene <i>yolk</i> em <i>Anastrepha</i> . As cores azul e verde representam respectivamente éxons e íntrons. É importante ressaltar que para <i>Anastrepha</i> não há evidência da presença de outras cópias do gene <i>yolk</i>	28
Fig. 7 Representação gráfica de população estável, referente às regiões 5', mid e 3', respectivamente	33
Fig. 8 Representação gráfica de população em expansão, referente às regiões 5', mid e 3', respectivamente	33
Fig. 9 Rede haplotípica referente à região 5'	37
Fig. 10 Rede de haplótipos referente à região mid.....	40
Fig. 11 Rede de haplótipos referente à região 3'	42
Fig. 12 Nível 1 de aninhamento para a rede 5'	44
Fig. 13 Nível 2 de aninhamento para a rede 5'	45
Fig. 14 Nível 3 de aninhamento para a rede 5'	46
Fig. 15 Nível 4 de aninhamento para a rede 5'	47
Fig. 16 Nível 1 de aninhamento para a rede mid.....	48
Fig. 17 Nível 2 de aninhamento para a rede mid.....	49
Fig. 18 Nível 3 de aninhamento para a rede mid.....	50
Fig. 19 Nível 4 de aninhamento para a rede mid.....	51
Fig. 20 Nível 1 de aninhamento para a rede 3'	52
Fig. 21 Nível 2 de aninhamento para a rede 3'	53
Fig. 22 Nível 3 de aninhamento para a rede 3'	54
Fig.23 a) Expansão contínua da distribuição b) Fluxo gênico Restrito com IBD.....	66

Lista de Tabelas

Tab. 1	Localidades de coleta, coordenadas, número de indivíduos amostrados e espécies .	16
Tab. 2	Condições específicas utilizadas para a amplificação do gene <i>yolk</i>	24
Tab. 3	Representação dos índices de diversidade nucleotídica, theta e número de sítios polimórficos para as três espécies de <i>Anastrepha</i> referentes às regiões particionadas do gene <i>yolk</i>	29
Tab. 4	Representação dos índices de diversidade nucleotídicas e ka/ks comparados para as três regiões do gene <i>yolk</i>	31
Tab. 5	Valores comparados de diversidade nucleotídica e theta referentes às partições do gene <i>yolk</i> , <i>fruitless</i> e <i>doublesex</i> para as três espécies de <i>Anastrepha</i>	31
Tab. 6	Representação dos valores de neutralidade (Tajima D, Fu & Li F e D, Fu & Li Fs), bem como os valores de distribuição mismatch (r e R2) para as três regiões do gene <i>yolk</i> .	34
Tab.7	Representação dos haplótipos e suas respectivas localidades referentes à região 5'	36
Tab. 8	Representação dos haplótipos e suas respectivas localidades referentes à região mid	38
Tab.9	Representação dos haplótipos e suas respectivas localidades referentes à região 3' .	41
Tab. 10	Representação dos valores de Dc, Dn e I-T para todos os clados referentes à região 5'	57
Tab. 11	Representação dos valores de Dc, Dn e I-T para todos os clados referentes à região mid.....	60
Tab. 12	Representação dos valores de Dc, Dn e I-T para todos os clados referentes à região 3'	62
Tab. 13	Clados que apresentaram valores significativos para as regiões 5', mid e 3' .	63
Tab.14	Representação dos valores de K, theta, T e Qui quadrado evidenciando para ocorrência de evento único relacionado à expansão contínua da distribuição para os três clados em questão (P<0,05).....	64
Tab. 15	Representação dos valores de K, theta e T. Teste de Takahata <i>et al.</i> (2002) de congruência temporal destes valores indica valor de $\chi^2=0.097$, n.s. Indicando não necessidade de mais de um evento de fluxo gênico restrito com isolamento por distância para os clados em questão (P<0,05)	65

Sumário

1- Resumo.....	3
1- Abstract.....	4
2. Introdução.....	1
2.1 Especiação.....	1
2.2 Espécies estudadas.....	3
2.3 As proteínas yolk.....	5
2.4 Proteínas do vitelo do ovo: Comparação entre diferentes filós.....	8
2.5 Filogeografia.....	11
3. Objetivos.....	14
4. Metodologia.....	15
4.1 Coleta.....	15
4.2 Locais de Coleta.....	15
4.3 Elaboração de primers.....	16
4.4 Extração de DNA.....	18
4.5 Amplificação por PCR e purificação do produto.....	19
4.6 Quantificação dos produtos de PCR.....	19
4.7 Clonagem.....	20
4.8 Sequenciamento.....	20
4.9 Análises das sequências de DNA.....	21
5 - Resultados e Discussão.....	24
5.1 Análises de Polimorfismo e Divergência entre as regiões gênicas.....	28
5.2 Testes de Neutralidade e Distribuição Mismatch.....	32
5.3 Análise das redes de haplótipos.....	34
5.4 Nested Clade Phylogeographical Analysis (Análise filogeográfica dos clados aninhados).....	43
6 - Conclusões.....	69
7. Referências Bibliográficas.....	70

2. Introdução

2.1 Especiação

O entendimento do processo de formação de novas espécies é essencial em estudos de biologia evolutiva. O conhecimento acerca do processo de especiação é fundamental para o entendimento dos padrões que moldam a distribuição dos organismos no tempo e no espaço. Apesar dos avanços recentes nos estudos referentes ao processo de especiação, ainda pouco se sabe como os eventos e forças evolutivas atuam na separação e formação de novas espécies, principalmente em organismos não pertencentes ao gênero *Drosophila* (Butlin e Ritchie, 2001; Wu e Ting, 2004).

Muitos consideram que o processo de especiação se concretiza a partir do momento em que indivíduos de uma população não conseguem mais trocar material genético entre si. Esse processo pode se iniciar através de uma barreira geográfica ou devido a fatores que promovam a diferenciação dos indivíduos mesmo que estes ocupem a mesma área. Muitas vezes, a ausência do fluxo gênico entre as populações se caracteriza como um fator importante para o processo de especiação, de modo que o isolamento reprodutivo eventual seja devido à divergência de genes, ao acaso ou por seleção natural, ocorrentes durante o processo de especiação.

Quando duas populações se tornam distintas e isoladas reprodutivamente fica mais difícil de se saber se a diferenciação entre elas é devido ao processo de especiação em si ou refere-se à diferenciação genética ocorrida após este processo. A diferença entre as espécies pode ter surgido posteriormente à ocorrência do processo de especiação, simplesmente como produto da ausência de fluxo gênico entre as populações. Sendo assim, a dificuldade em se estudar o processo de especiação está no fato de se determinar os possíveis genes realmente responsáveis pelo isolamento reprodutivo (Wu e Ting, 2004).

O estabelecimento de isolamento reprodutivo entre populações favorece, ao longo do tempo, o aumento da divergência entre as espécies e, em conseqüência, uma maior dificuldade em se distinguir genes realmente responsáveis pelo isolamento reprodutivo inicial de outros polimorfismos. Desse modo, o estudo dos componentes genéticos referentes às diferenças fenotípicas em espécies irmãs ou proximamente relacionadas

poderia auxiliar na busca dos genes realmente envolvidos no processo de especiação (Zeng *et al.*, 2000).

A dificuldade na determinação dos fatores envolvidos no processo de especiação pode ser exemplificado através da observação de eventos de especiação mesmo com a ausência de mudanças cariotípicas (Carson e Kaneshiro, 1976; Kornfield *et al.*, 1979), mudanças morfológicas (Mayr, 1963), mudança de nicho (Heed, 1971), e diferenciação significativa entre alozimas (Carson e Kaneshiro, 1976; Kornfield e cols., 1979; Coyne e Orr, 1997).

Dois estratégias distintas vêm sendo utilizadas para se estudar o processo genético de especiação. A primeira estratégia utiliza estudos filogeográficos de isozimas, DNA nuclear e mitocondrial (Avise, 1987) que vêm se mostrando uma ferramenta eficaz no estudo do processo de especiação. Através dessa estratégia é possível mapear o padrão de polimorfismo de regiões específicas do DNA de interesse, associando-o com o processo de diferenciação populacional. A caracterização do processo de especiação pode ser feita através do estabelecimento da arquitetura genética de fenótipos importantes aliada aos fatores históricos e demográficos envolvidos nesse processo, onde aí incluímos a filogeografia (Templeton *et al.*, 1995 ; Avise, 1987). Uma segunda estratégia envolve o mapeamento de características consideradas importantes nos processos de especiação, como aquelas relacionadas a adaptações específicas de cada espécie, diferenças nos traços sexuais, incluindo comportamento (Panhuis *et al.*, 2000) e elementos causadores de inviabilidade ou esterilidade do híbrido (Butlin e Ritchie, 2001).

A filogeografia é uma disciplina integrativa que se utiliza de dados paleontológicos, geografia histórica, evolução molecular e ecologia para explicar o padrão de distribuição geográfico das linhagens genéticas (Avise, 2000). Trata-se de uma área que vem sendo utilizada com sucesso para o estudo e entendimento dos fatores que são essenciais ao processo de especiação (Knowles, 2004). Através desses estudos é possível compreender e relacionar a evolução de diferentes linhagens com eventos históricos, e ou demográficos responsáveis pela sua diversificação. Além disso, os estudos filogeográficos podem ser utilizados na análise de variação genética e estrutura das populações animais e vegetais (Paetkau *et al.*, 1998; Lenk *et al.*, 1999; de Brito *et al.*, 2002; Caicedo e Schaal, 2004), bem como nos fornecer uma base importante para a compreensão de como eventos geológicos e

mesmo antrópicos influenciam as populações naturais (Avice, 2004).

A seleção sexual pode promover uma rápida divergência em traços que envolvam a escolha de parceiros para a reprodução (Andersson, 1994), sendo que muitos desses traços envolvem reconhecimento sexual, promovendo desta forma acasalamentos preferenciais entre os indivíduos (Boake 2000; Panhuis *et al.* 2001; Coyne & Orr 2004). Portanto, tem sido proposto que a seleção sexual, ao promover a rápida divergência em traços que envolvam reconhecimento sexual, possa estar diretamente relacionada ao processo de especiação (Andersson 1994; Gray & Cade 2000 ; Masta & Maddison 2002). Estudos envolvendo marcadores moleculares mostraram que genes associados com reconhecimento ou determinação sexual apresentam taxa de evolução mais rápida comparado a outros genes em geral (Tsaur e Wu, 1997; Graham, 2002). Quando duas populações estão sob efeito de seleção sexual, é possível que apresentarão maior divergência no que se refere à preferência sexual e, desta forma, o resultado apontaria para o isolamento sexual destas populações, concretizando assim o processo de especiação.

2.2 Espécies estudadas

A maioria das espécies do gênero *Anastrepha* é endêmica à Região Neotropical e algumas ocorrem no sul da região Neártica. O gênero está estabelecido no sul dos EUA, no México, na América Central e em toda a América do Sul, exceto no Chile onde ocorre, esporadicamente, ao norte do deserto de Atacama na fronteira com o Peru (Malavasi *et al.* 2000). A família Tephritidae, onde se encontra o gênero *Anastrepha*, possui aproximadamente 500 gêneros e 4448 espécies e sub-espécies (Norrbon, 2004). O gênero *Anastrepha* apresenta cerca de aproximadamente 197 espécies atualmente reconhecidas (Norrbon *et al.*, 2004). No Brasil, há o registro de 95 espécies distribuídas em, pelo menos, 13 grupos (Uramoto, *et al.*, 2004). A maioria dessas espécies são responsáveis por danos de grandes proporções às culturas de frutas carnosas, apresentando grande importância na economia mundial, sendo diretamente responsável por afetar a qualidade de muitos frutos, gerando prejuízo econômico em vários países. As larvas se desenvolvem no endocarpo e alimentam-se dos tecidos de plantas incluindo frutos, sementes, folhas, caules, ou flores, enquanto os adultos alimentam-se de líquidos como seiva, néctar, melado ou excrementos

(Malavasi *et al.*, 2000). Além disso, o processo mecânico de oviposição permite a invasão de outros organismos, como por exemplo os fungos, que danificam a parte carnosa do fruto (Carvalho, 2005). Dessa forma, percebemos que as espécies desse gênero estão relacionadas com os maiores danos em cultivares por atacar frutos verdes e maduros indistintamente (Zucchi, 2000).

No presente estudo daremos ênfase a três espécies do gênero *Anastrepha*, incluindo aquelas economicamente mais relevantes: *Anastrepha fraterculus*, *Anastrepha obliqua* e *Anastrepha sororcula* que são das principais pragas de frutos do país (Zucchi, 2000; Selivon, *et al.*, 2000). Essas espécies estão incluídas no grupo *fraterculus*, caracterizado por ser um grupo heterogêneo de espécies crípticas, de difícil distinção (Morgante *et al.*, 1980). No presente trabalho iremos nos referir ao grupo de espécies *fraterculus* para designar um conjunto de aproximadamente 29 espécies proximamente relacionadas, incluindo as três espécies envolvidas neste trabalho. Uma segunda nomenclatura é utilizada (complexo *fraterculus*) quando nos referimos a um grupo de espécies atualmente designado *A. fraterculus* proper (Smith-Caldas *et al.*, 2001). As espécies do grupo *fraterculus* são importantes não somente por sua ampla distribuição, mas também por suas preferências com relação aos frutos para oviposição.

As espécies citadas anteriormente devem ser consideradas oligotróficas, pois embora exista certa especificidade, em geral as espécies não estão restritas a apenas um hospedeiro (Solferini e Morgante, 1987). *A. fraterculus* infesta aproximadamente 67 espécies de plantas, 24 delas pertencentes à família Myrtaceae (onde podemos incluir pitanga e uvaia), *A. sororcula* infesta 19 espécies, sendo 15 delas Myrtaceae, e *A. obliqua* infesta 27 espécies, 16 Myrtaceae e 8 Anacardiaceae (onde podemos incluir manga e caju) (Zucchi, 2000).

As características marcantes que permitem a distinção de espécies dentro desse grupo são: as manchas no escutelo, o padrão de coloração da asa, a morfologia do ápice do ovipositor, a morfologia do ovo e o horário de cópula (Selivon, 2000; Santos *et al.*, 2001). Em *A. fraterculus* o comportamento de corte é matutino (Malavasi, 1984), ocorrendo preferencialmente entre as 7-13 horas, enquanto que em *A. sororcula* o comportamento de corte é vespertino, preferencialmente entre as 15-19 horas, não havendo portanto, sobreposição de horário de cruzamentos entre essas duas espécies (Selivon, 2000). Já *A.*

obliqua tem comportamento de corte geralmente, mas não exclusivamente, matutino (Malavasi, 1984). Embora não exista um marcador único que permita diferenciar as espécies do grupo *fraterculus*, o conjunto das características acima mencionadas permite, na grande maioria dos casos, a sua correta identificação.

Há evidências de que os padrões de variação e divergências genéticas nessas espécies não correspondam às diferenciações morfológicas (Morgante *et al.*, 1980; Selivon, 2000) dificultando a identificação e separação das diferentes espécies do grupo *fraterculus*. Dessa forma, existe grande variabilidade intraespecífica nas diferentes espécies, sendo inclusive sugerido que o grupo *fraterculus* seja de fato um complexo de espécies crípticas (Steck, 1991; Amaral, 1994; Selivon, 2000).

Outro aspecto importante é que vários cruzamentos entre algumas espécies deste grupo produzem híbridos interespecíficos viáveis (Selivon, *et al.*, 2005), permitindo o estudo de genes envolvidos na determinação das diferenças interespecíficas. A existência de polimorfismos intraespecíficos bem como de cruzamentos viáveis entre diferentes espécies do grupo sugere que estas espécies do grupo *fraterculus* divergiram há pouco tempo e, possivelmente, ainda não acumularam diferenças genéticas significativas independentemente das advindas do processo de especiação.

2.3 As proteínas yolk

Em *Drosophila*, alguns genes participam ativamente de uma cascata de eventos diretamente responsável pelo processo de determinação sexual. Três genes principais coordenam este processo: *sex-lethal (sxl)*, *transformer (tra)* e *doublesex (dsx)*, de modo que o último gene da cascata, o *doublesex*, é o que atua ativando ou não a síntese das proteínas yolk (Alvares, *et al.*, 2009), através de seus produtos, ativo (DSX f- proteína doublesex fêmea específico) ou inativo (DSXm- proteína doublesex macho específico) (Lebo *et al.*, 2009). As proteínas yolk tem papel importante na formação do produto nutritivo utilizado para a nutrição do embrião durante seu estágio de desenvolvimento.

Há 2 grandes famílias de proteínas envolvidas na formação do produto nutritivo (vitelo) utilizado para a nutrição do embrião em sua fase de desenvolvimento. A primeira grande família é a das vitelogeninas, que são encontradas em vertebrados ovíparos (sapos,

galinhas, peixes), em nematóides e alguns insetos. A outra grande família é denominada proteínas *yolk* e estão presentes em dípteros e moscas domésticas (Bownes, 1992; Sappington, 2002). Essas duas famílias de proteínas diferem grandemente em relação a seus sítios de síntese, tamanho e processamento (Chen *et al.*, 1997).

Três genes específicos estão envolvidos na formação das proteínas do ovo em *Drosophila melanogaster*, sendo que a síntese dessas proteínas a partir desses três genes se inicia a partir do momento em que a mosca libera-se de sua pupa. A transcrição desses genes é regulada pelo gene *doublesex* (*dsx*) que atua diretamente ativando ou não os genes *yolk* (Alvares, *et al.*, 2009), e também pelos genes *transformer* e *transformer 2*, que atuam ativando ou não o gene *doublesex* (Lebo, *et al.*, 2009). Além dos genes da cadeia de determinação sexual, a síntese das proteínas *yolk* é estimulada pela ação de dois hormônios do desenvolvimento: o hormônio juvenil e o 20 hydroecdisone (Bownes, *et al.* 1988, Andam, *et al.*, 2004; Rauschenbach, *et al.*, 2007)

Em *Drosophila*, o gene *doublesex* está no final da cadeia de determinação sexual, de modo que suas proteínas sexo específicas DSX (f) e DSX(m) com diferentes domínios C-terminais podem ativar ou inativar a síntese proteica dos três genes *yolks* (*yk1*, *yk2* e *yk3*) envolvidos na formação das proteínas do vitelo do ovo (Chen, *et al.*, 2008). As proteínas DSX (f) e DSX(m) se ligam ao mesmo sítio de transcrição dos genes *yk1*, *yk2* e *yk3* (Cho e Wensink, 1997), sendo que tais sítios se encontram no espaço intergênico entre *yk1* e *yk2* e na região adjacente ao gene *yk3* (Burtis *et al.*, 1991; An e Wensink, 1995; Erdman *et al.*, 1996), de modo que a proteína DSX(f) atua ativando e DSX(m) atua inibindo a transcrição desses genes.

Os genes *yolk* (*yk*) estão ligados ao cromossomo X e apresentam controles diferenciais em suas expressões (Hovemann e Galler, 1982). A síntese dos polipeptídeos *yolk* acontece em dois locais específicos, o primeiro local de síntese acontece nos tecidos gordurosos do corpo adulto da fêmea de onde os polipeptídeos são transportados para dentro da hemolinfa e em seguida “direcionados” ao ovário; o segundo local de síntese processa-se nas células do folículo ovariano (Bepas, *et al.*, 2008). Os genes *yk1* e *yk2* estão separados por 1225 pb e realizam a transcrição em direções divergentes, de forma que os elementos que controlam o início da transcrição desses genes são aparentemente

compartilhados (Saunders e Bownes, 1986; Bownes, 1994) enquanto que *yk3* dista aproximadamente 1Mb de *yk1* (Garabedian *et al*, 1987).

A produção de cada um desses polipeptídeos se dá por uma cópia simples de cada gene situado no cromossomo X, de modo que esses genes tornam-se ativos pela ação regulatória de outros genes envolvidos diretamente na diferenciação sexual, entre eles o *doublesex (dsx)*, o *transformer (tra)* e o *transformer 2 (tra-2)* (Alvares, *et al.*, 2009). A síntese dos polipeptídeos *yolk* inicia-se no momento em que a fêmea libera-se de sua pupa e aumenta drasticamente durante seu estágio de maturação. A síntese dos transcritos dos genes envolvidos com a formação do vitelo do ovo é também influenciada diretamente pela ação hormonal, com a ação do hormônio juvenil nas células do epitélio folicular, e em outras partes da cabeça, tórax e abdômen do corpo adulto das fêmeas (Richard, *et al.*, 1998;) e pela ação do hormônio esteróide 20 hydroyecdisona (Richard, *et al.*, 1998; Richard *et al.*, 2001).

Estudos envolvendo a linhagem germinativa de *Drosophila melanogaster* demonstraram a presença de dois enhancers tecido específicos envolvidos na síntese dos polipeptídeos *yolk 1* e *2*. Estes enhancers são comuns aos genes *yk1* e *yk2* estando localizados na região intergênica *yk1/yk2* (Garabedian *et al*, 1986). Um desses enhancer apresenta 125pb e atua bidirecionalmente na síntese dos polipeptídeos *yolk yk1* e *yk2* presentes na cabeça, tórax e abdômen da fêmea adulta, estando localizado entre os genes *yk1* e *yk2*, mais proximamente a *yk1*. O outro enhancer direciona a transcrição nos ovários e se encontra dentro de um fragmento de aproximadamente 2,8 kb, que começa a aproximadamente 886 kb de distância *upstream* de *yk1* e engloba totalmente *yk2* (Logan, *et al.* 1989).

Nos ovários, o enhancer ovariano atua sobre os promotores de *yk1* e *yk2*, promovendo a síntese dos polipeptídeos relacionados a estes genes, já o enhancer relacionado à síntese dos polipeptídeos presentes nas porções gordurosas do corpo adulto da fêmea encontra-se inativo. Nos ovários, temos a atuação do elemento 1, potencializando a ação dos promotores de *yk1* e *yk2*. De maneira oposta, nas partes gordurosas do corpo adulto da fêmea, o enhancer ovariano encontra-se inativo, e a síntese dos polipeptídeos *yk 1* e *yk2* se dá pela ação do enhancer das partes gordurosas atuando sobre os promotores *yk1* e *yk2*. Na síntese dos polipeptídeos *yk1* e *yk2* nas regiões gordurosas, verificou-se que o

elemento 1 está potencialmente inativo (Logan, *et al.* 1989).

A figura 1 representa a atuação dos diferentes enhancers envolvidos na ativação dos promotores dos genes *yk1* e *yk2* envolvidos na síntese dos polipeptídeos *yk1* e *yk2*.

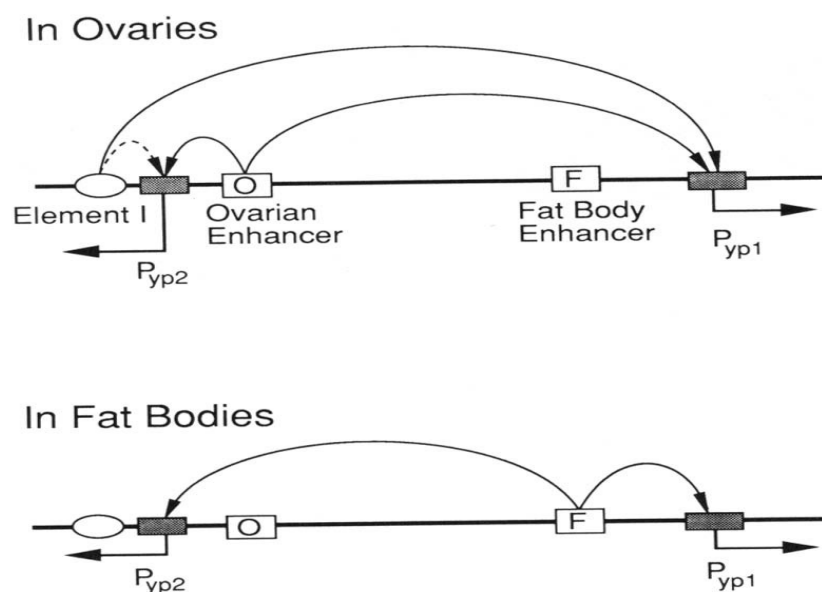


Fig. 1 Representação dos enhancers envolvidos na ativação dos promotores dos genes *yk1* e *yk2*

2.4 Proteínas do vitelo do ovo: Comparação entre diferentes filios

O uso do termo proteínas vitelogênicas ou vitelogeninas é freqüentemente aplicado a dípteros superiores como *D. melanogaster* e outros filios. No entanto, tais proteínas em *D. melanogaster* diferem amplamente daquelas encontradas nas diferentes classes de nematóides, vertebrados e também em outros dípteros, no que se refere a sítios de síntese, processamento e tamanho (Bownes, 1982). Assim, Bownes (1982) sugeriu o uso do termo proteínas *yolk* quando nos referimos à classe de proteínas do ovo em *Drosophila melanogaster*. Apesar da comparação entre as seqüências *yk* e *vg* mostrarem pouca relação entre os precursores destas duas classes de proteínas (Rina e Savakis 1991) a confusão de terminologia ainda persiste.

Em *Ceratitis capitata* (família Tephritidae) há duas proteínas importantes envolvidas na formação do vitelo do ovo, essas proteínas são denominadas vitelogeninas (*vg1* e *vg2*), com pesos moleculares de 49,000 e 46,000 dáltons, respectivamente. Esses

polipeptídeos são sintetizados a partir de 4 genes que estão localizados na mesma região do cromossomo 5 e estão organizados em pares, de modo que cada par codifica os dois polipeptídeos (Vg1 e Vg2) em sentidos divergentes (Rina e Savakis, 1991). Os genes *Vg1* e *Vg2* em *C. Capitata* e *yk1*, *yk2* e *yk3* em *D. melanogaster* apresentam algumas similaridades em suas estruturas, sendo que o gene *Vg1* apresenta dois íntrons na mesma posição daqueles encontrados no gene *yk1*, e o gene *Vg2* apresenta somente um íntron, assim como acontece nos genes *yk2* e *yk3*. Este padrão de similaridade engloba também a sequência primária, o padrão de hidrofobicidade e a estrutura secundária relacionada às proteínas desses genes (Rina e Savakis, 1991).

No gênero *Anastrepha* poucos trabalhos foram feitos relacionados à estrutura e ação do gene *yolk* e suas proteínas no desenvolvimento do embrião. Diferentemente de *Drosophila*, estudos com *Anastrepha suspensa* mostraram a presença de somente uma cópia do gene *yolk* apresentando três éxons e dois íntrons. Este gene é responsável pela síntese de dois polipeptídeos de, aproximadamente, 48 kda que são sintetizados principalmente no ovário, de forma que, a partir do quarto ou quinto dias de vida, tem-se um aumento significativo na síntese de tais polipeptídeos (Handler, 1997; Kendra et. Al., 2006). Além do ovário, a síntese acontece também nas partes gordurosas do corpo do adulto, porém, em escala menor comparado ao gênero *Drosophila* (Handler, 1997).

Em *Anastrepha suspensa* a presença dos hormônios juvenil e 20 hydroyecdisona parece não influenciar tão significativamente na síntese dos polipeptídeos *yk* comparado a *Drosophila melanogaster* (Handler, 1997). Os abdômens de fêmeas jovens (3-5 dias de vida) foram tratados com hormônio juvenil e 20 hydroyecdisona a fim de se verificar o possível aumento na síntese dos polipeptídeos comparado a abdômens que não receberam tais tratamentos. A taxa de síntese não diferiu significativamente entre os abdômens tratados e não tratados (Handler, 1997).

Em *Apis mellifera* as vitelogeninas são monômeros de aproximadamente 180kDa e estão presentes predominantemente na hemolinfa. A síntese desses monômeros aumenta significativamente entre o segundo e terceiro dias de vida e atinge seus picos de produção entre o quinto e décimo quinto dias (Amdam, et al. 2002). Em *Apis*, as vitelogeninas parece ter um papel importante no processo de remoção de radicais livres, reduzindo o

stress oxidativo, contribuindo assim para um aumento no tempo de vida desses insetos (Nelson, *et al.*, 2007).

No nematóide *C. elegans* duas grandes classes de proteínas são conhecidas. A primeira classe consiste de 2 polipeptídeos com 170,00 kda (vitelogeninas 170 A e 170 B) que são especificados por um conjunto de 4 genes, *vit 3* e *vit 4* que estão proximamente ligados no cromossomo X e são responsáveis pela síntese da vitelogenina 170A, *vit 5* que não está proximamente ligado a *vit 3* e *vit 4*, e *vit 2* relacionado à síntese da vitelogenina 170 B (Spiel e Blumenthal, 1985). A segunda classe é composta por dois polipeptídeos menores, de 88 e 115kda (Sharrock, 1984), clivados na cavidade corpórea a partir de um polipeptídeo maior denominado vg 180, com aproximadamente 180,00 kda. (Spiel e Blumenthal, 1985). Este polipeptídeo vg 180 é codificado por um único gene, *vg6*, que é distantemente relacionado ao outros 4 genes codificadores dos dois polipeptídeos da primeira classe de proteínas.

Nos vertebrados, a síntese das vitelogeninas ocorre preferencialmente no fígado através de uma cascata coordenada do sistema endócrino que envolve o coração, ovário, fígado e o sistema circulatório (Finn, 2007), sendo que o hormônio envolvido diretamente nesta síntese é o estrógeno (Wahli & Ryffel, 1985, Miracle *et al.*, 2006). O sistema circulatório é o responsável pelo transporte das vitelogeninas do fígado até o oócito, onde será fonte de nutrientes para o embrião durante seu estágio de desenvolvimento (Finn, 2007). A composição da gema do ovo nos vertebrados é bem conhecida, constituindo-se principalmente de proteínas, lípidos, fósforo e cálcio (Brawand *et al.*, 2008).

Em *Xenopus laevis* e *Gallus gallus* as vitelogeninas variam de 200- 250 kda, sendo clivadas dentro dos oócitos e produzindo lipovitelina 1 de aproximadamente 120kda, fosfovítina rica em serina com aproximadamente 35kd e lipovitelina 2, com peso aproximado de 30kb (Wahli, 1988; Brawand *et al.*, 2008).

Em algumas classes de mamíferos (placentários e marsupiais) a nutrição do embrião se dá através de uma nova estrutura denominada placenta, sendo esta responsável pelo contato direto entre mãe/embrião-feto. A placenta é vascularizada e tem um papel similar ao das vitelogeninas, sendo diretamente responsável pela transmissão dos nutrientes da mãe ao embrião/feto (Brawand *et al.*, 2008).

2.5 Filogeografia

A filogeografia refere-se ao estudo dos princípios e processos que governam a distribuição geográfica de linhagens genealógicas. Como a própria definição diz, a filogeografia não lida somente com as relações filogenéticas entre os taxa estudados, mas também com os componentes ou processos históricos que moldaram a distribuição desses taxas (Avice, 2000). Dessa forma, a análise e interpretação da distribuição das linhagens requer o processamento conjunto de informações de uma série de disciplinas, entre as quais podemos incluir a genética molecular, filogenética, genética de populações, etologia, geologia, paleontologia e geografia histórica. Assim, o aspecto multidisciplinar da filogeografia cria uma ponte entre os processos micro e macroevolutivos (Avice, 2000).

Dados provenientes de marcadores moleculares são especialmente úteis para investigar os processos de especiação (Terborgh, 1992; Avice 2000) sendo capazes de fornecer informações acerca de quantidade de fluxo gênico entre populações, detectar eventos de fundador e gargalos populacionais (Avice, 1994) e estimar expansão de território em relação à localização de supostos refúgios e zonas de expansão (Rogers e Harpending, 1992).

O DNA mitocondrial (DNAmít) é um marcador amplamente utilizado em estudos filogeográficos, pois apresenta características únicas – ausência de seqüências não codificantes e de recombinação, herança predominantemente materna e alta taxa de evolução, sendo portanto um marcador extremamente informativo em estudos evolutivos no que se refere à capacidade de caracterizar variabilidade intra e interespecífica, identificar espécies e populações e estimar relações filogenéticas (Avice 1994, 2000, 2009) . O DNA nuclear também é bastante utilizado em estudos filogeográficos, no entanto, a possível ocorrência de recombinação nas seqüências a serem analisadas pode gerar uma interpretação equivocada acerca dos dados obtidos, determinando assim inferências não correspondentes à verdadeira história evolutiva das espécies (Templeton *et al.*, 2000).

Várias metodologias podem ser utilizadas para explicar o padrão de distribuição e estruturação genética das populações; no entanto, muitas delas realizam inferências biológicas sem a utilização de um suporte estatístico eficaz (Creer, *et al* 2001). Dessa forma, Alan Templeton e colegas desenvolveram a metodologia Nested Clade

Phylogeographic Analysis (NCPA) utilizando-se de ferramentas estatísticas como suporte para a construção de cladogramas que determinem as relações genéticas entre diferentes haplótipos com o padrão de distribuição geográfica destes (Templeton, *et al.*, 1995; Templeton, 2008). O uso da estatística nesta metodologia está associada à construção de uma rede haplotípica com 95% de parcimônia, de modo que essa mesma rede é a que é utilizada para testar a possível associação genética dos haplótipos com a geografia (Weiss e Ferrand, 2007).

Nesta metodologia é testada a validade da hipótese nula referente à não associação entre haplótipos e geografia (Templeton, *et al.*, 1995; Templeton, 2008). Caso a hipótese nula seja rejeitada, três diferentes cenários são propostos para explicar o padrão estrutural das populações e eventos históricos; tais cenários referem-se ao fluxo gênico restrito, à expansão contínua da distribuição ou ainda a fragmentação alopátrica (Templeton, *et al.*, 1995).

Apesar de amplamente utilizada, a metodologia da NCPA é recebida com ceticismo por parte de alguns geneticistas de populações. Em trabalhos atuais, Knowles & Maddison, (2002) e Knowles (2008) criticaram o uso desta metodologia, argumentando que a NCPA tende a resultar em falsos positivos, inferindo um processo que pode não estar acontecendo ou não ser representativo da história evolutiva das espécies. O principal questionamento de Lacey Knowles refere-se à simulação de dados envolvendo um simples locus, onde a análise isolada de um gene pode não revelar de forma verdadeira a história evolutiva de uma espécie. Na visão de Lacey Knowles, se a NCPA falha em lidar com cenários evolutivos simples, os problemas do uso dessa metodologia serão ainda maiores quando relacionados a cenários múltiplos e complexos que ocorrem na natureza.

Knowles contrasta dois métodos distintos, um desses métodos estima parâmetros comumente utilizados para se fazer inferências a respeito do passado demográfico. Apesar de apresentar um suporte estatístico eficaz, este método falha em dar pouca ênfase a questão da história geográfica. Por outro lado, métodos que buscam a reconstrução da história filogeográfica levam em consideração muitos cenários geográficos alternativos, mas não apresentam um suporte estatístico eficaz que possa explicitar um determinado processo biológico associado a um cenário geográfico específico. Knowles acrescenta ainda sobre a importância da filogeografia estatística para a obtenção de resultados mais acurados

referentes principalmente à segunda metodologia, de forma que a utilização da ferramenta estatística pode contribuir amplamente para o melhor entendimento de eventos do passado, associando-os com os padrões atuais de distribuição e estruturas populacionais (Knowles, 2004, Knowles e Maddison 2008). Templeton reconhece as falhas da NCPA para inferências baseadas em estudos de somente um locus e sugere a utilização de múltiplas regiões de DNA como ferramenta capaz de minimizar os possíveis erros de inferência. Assim, o estudo conjunto de vários loci torna-se mais eficaz na determinação da verdadeira história evolutiva das espécies (Templeton, 2004).

Outras metodologias podem ser utilizadas em estudos filogeográficos. Os métodos de detecção de seleção visam determinar se um gene na população está sob efeito de seleção natural ou neutralismo. Testes de Neutralidade como os de Tajima, Fu e Li D e F, Macdonald-Kreitman entre outros, baseiam-se em modelos estatísticos para determinar se a seleção está ou não atuando. Uma vez detectada a atuação da seleção, podemos determinar o tipo de seleção predominante, podendo ser purificadora ou ainda positiva. A seleção purificadora tem como maior característica o aumento das mutações sinônimas em relação às não sinônimas, sendo que sua ação principal visa eliminar mutações deletérias que são desvantajosas na população. A seleção positiva, por sua vez, favorece o aumento de mutações não sinônimas e geralmente favorece um alelo que pode ter sido anteriormente neutro, mas que tenha se tornado vantajoso devido a mudanças ambientais, tornando-se importante para a evolução de uma nova função gênica (Evanovich, 2004).

Nesta dissertação, faremos uso de metodologias filogeográficas visando o conhecimento acerca do nível de variação do gene *yolk* para as espécies do gênero *Anastrepha*, inferindo desta forma o possível padrão atual de distribuição de *Anastrepha* ao longo do território nacional.

3. Objetivos

- A) Estabelecimento de primers que possam amplificar regiões específicas do gene *yolk* em espécies de *Anastrepha* do grupo *fraterculus*.
- B) Determinar, a partir do sequenciamento das regiões amplificadas do genoma de espécies de moscas-das-frutas do grupo *fraterculus*, o nível de variação genética intra e inter-específica.
- C) Proceder análises filogeográficas através do estudo da variação existente no gene *yolk*, determinando o possível papel deste gene no padrão de distribuição geográfica das espécies do grupo *fraterculus* ao longo do território brasileiro.

4. Metodologia

4.1 Coleta

As coletas de *Anastrepha* foram feitas pela coleta direta de adultos no campo ou pela retirada de frutos de regiões com pouca interferência do homem, sendo levados para o laboratório onde são mantidos em caixas acrílicas com vermiculita, onde as larvas se enterrarão para posterior empupação. Em uma etapa seguinte retiramos as pupas da vermiculita, transferindo-as para uma nova caixa acrílica com alimento e água. Após a eclosão das pupas, as moscas são separadas por sexo e mantidas em caixas acrílicas de 50x50x40 cm. Após uma semana as moscas atingem a maturidade sexual, quando então são identificadas. Algumas destas moscas são armazenadas em etanol 70% para posterior extração de DNA, enquanto outras são utilizadas em cruzamentos. As moscas adultas são tratadas com uma dieta de 10ml de mel, 30ml de hidrolisado de proteína de milho, 6g de levedo de cerveja, 0,5g de Sustagem sabor morango, 25g de açúcar mascavo e 45g de açúcar branco.

4.2 Locais de Coleta

Na tentativa de se obter indivíduos de várias localidades realizamos coletas nas regiões Sul, Sudeste, Centro-Oeste, Norte e Nordeste. Dessa forma, obtivemos uma amostragem homogênea e bem distribuída ao longo da área de dispersão destas moscas no Brasil.

No presente trabalho foram amostrados, em média, 1 indivíduo por localidade, sendo que em algumas delas amostramos 2, ou 3 indivíduos. Esta variação na amostragem dependeu da disponibilidade dos indivíduos em nosso laboratório referente à cada localidade. Os indivíduos amostrados de cada localidade poderiam pertencer à mesma ou a espécies distintas. No total foram amostrados 37 indivíduos de 24 localidades representadas na tabela abaixo.

Pontos de Coleta	Latitude	Longitude	N ^a de indivíduos amostrados	Espécie(s)
Araguaina TO	07°11'00''S	48°12'00''W	2	A. obliqua
Babaçulândia TO	07°12'17''S	47°45'25''W	2	A. obliqua
Bauru SP	22°19'32''S	49°05'08''W	1	A.fraterculus
Belém PA	01°26'00''S	48°38'00''W	2	A.obliqua
B.Vista de Goiás GO	17°58'28''S	48°57'10''W	1	A.obliqua
Bonito PE	08°28'13''S	35°43'43''W	3	A.sororcula;A.fraterculus; A.obliqua
Cabo Frio RJ	22°53'00''S	42°05'00''W	2	A. fraterculus
Campinas SP	22°54'20''S	47°03'39''W	2	A.fraterculus
Campo Grande MS	21°34'00''S	54°54'54''W	2	A.sororcula
Curitiba PR	25°25'48''S	49°16'15''W	1	A.fraterculus
Estrela D'Oeste SP	20°17'16''S	54°24'03''W	2	A.sororcula
Feira de Santana BA	12°17'00''S	38°55'00''W	1	A.obliqua
Florianópolis SC	27°35'36''S	48°35'36''W	1	A.fraterculus
Gurupi TO	11°33'45''S	49°04'07''W	1	A.obliqua
João Pessoa PB	07°06'57''S	34°53'14''W	1	A. fraterculus
Linhares ES	19°23'28''S	40°04'20''W	1	A. obliqua
Joinville SC	26°18'16''S	48°50'54''W	1	A. fraterculus
Marilac MG	18°30'29''S	42°05'02''W	2	A. fraterculus
Palmas TO	10°10'01''S	48°19'58''W	3	A.obliqua
Piracicaba SP	22°42'00''S	47°32'00''W	1	A. sororcula
Recife PE	08°10'52''S	34°54'47''W	1	A. sororcula
São Carlos SP	22°01'03''S	47°53'27''W	1	A. fraterculus
Três Lagoas MS	20°45'04''S	51°40'42''W	1	A.obliqua
Vargem Alta ES	20°40'00''S	35°48'00''W	2	A. fraterculus

Tab. 1 Localidades de coleta, coordenadas, número de indivíduos amostrados e espécies

4.3 Elaboração de primers

Não existem ainda seqüências disponíveis do gene *yolk* para as espécies estudadas. Sendo assim, desenvolvemos primers específicos para a amplificação destas regiões em *Anastrepha*, através do alinhamento das sequências de nucleotídeos ortólogos do gene *yolk* obtidos a partir de quatro gêneros filogeneticamente próximos ao gênero *Anastrepha* disponibilizados no GenBank; *Drosophila*, *Anopheles*, *Ceratitis* e *Bactrocera*. Tratamos

estas sequências com editores de texto padrão e passamos para o formato adequado para análise no programa BioEdit (Hall, 1999) e utilizamos o software ClustalW (Higgins, *et al.*, 1994) para o alinhamento das sequências homólogas. Tal software permite o alinhamento mesmo quando as diferentes sequências apresentam tamanhos distintos relacionados a processos de inserções ou deleções. A partir deste alinhamento foi possível inferir as regiões evolutivamente conservadas com potenciais para a construção de primers.

Realizamos o alinhamento visual destas sequências para se verificar a adequação do alinhamento e também para se detectar regiões mais conservadas que pudessem estar flanqueando regiões altamente polimórficas. As regiões mais conservadas no maior número de taxa filogeneticamente próximos de *Anastrepha* foram as escolhidas para o desenho de primers específicos conservados que pudessem também amplificar sequências do gene *yolk* em *Anastrepha*. Escolhemos os primers que seriam testados e, com o auxílio do programa *OligoAnalyzer3.0*(<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx/oligocalc.asp>) analisamos as temperaturas de desnaturação, valores de delta G e possibilidade de formação de estruturas secundárias, tanto em ligações entre os mesmos primers como entre os diferentes do conjunto. Assim, verificamos a estabilidade das estruturas de forma a tentar reduzir a produção de artefatos nas reações de amplificação, otimizando o processo de obtenção das sequências nucleotídicas do gene em questão. Uma vez testados e escolhidos, usamos o Basic Local Alignment Search Tool (*BLAST*) para verificar se tais regiões são únicas ou apresentam sequências similares em outras regiões não homólogas encontradas no “GenBank”. Sendo estas regiões específicas, tais primers foram sintetizados.

O par de primers utilizados nas reações de amplificação estão representados na figura 2. A representação inclui também a região amplificada, bem como os íntrons e éxons. Amplificamos uma região de aproximadamente 900 pb, incluindo 3 éxons e 2 íntrons.

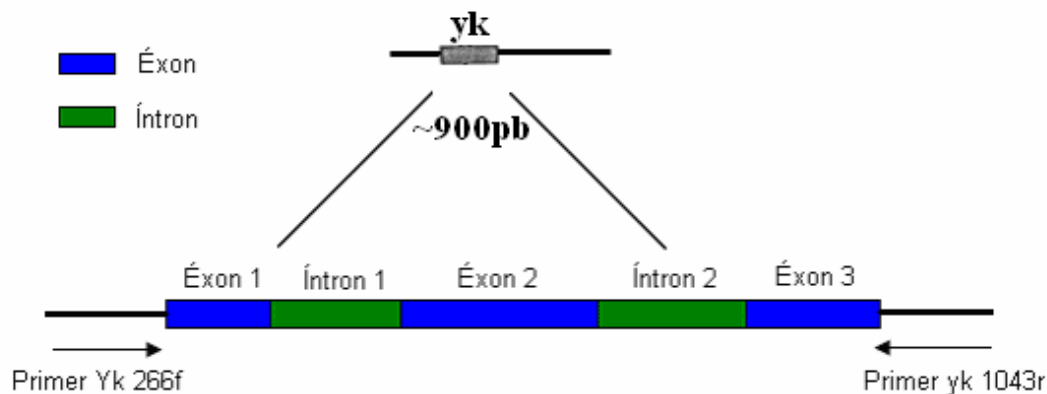


Fig. 2 Representação esquemática da região amplificada do gene *yolk*, seus primers e regiões exônicas e intrônicas

4.4 Extração de DNA

Perfuramos o corpo dos insetos, incubando-os em tubos de polipropileno contendo solução de 100 mM de NaCl, 100 mM Tris-Cl pH 8, 25 mM EDTA pH 8, 0,5% SDS e solução de Proteinase K (20mg/ml), para quebra das células e liberação do conteúdo intracelular, por 2 a 3 horas a 37°C. Após este tempo, adicionamos RNase (4mg/ml), incubando a solução a 37°C por 15 a 60 minutos. Posteriormente centrifugamos a solução a 3000 rpm por 1 minuto a fim de se retirar as impurezas insolúveis. Transferimos o sobrenadante para um outro tubo, adicionando à solução 4 M de tiocianato de guanidina e 0,1 M Tris-Cl pH 7,5, e, após homogeneização em vórtex por 5 a 7 segundos, o tubo é centrifugado a 13000 rpm por 10 minutos para a precipitação das proteínas. O sobrenadante é posto em outro tubo para posterior precipitação do DNA através de solução de isopropanol 100%. Com o DNA precipitado descartamos o sobrenadante e lavamos o precipitado por duas vezes com etanol 70%, invertendo gentilmente o tubo por várias vezes. Em seguida, centrifugamos a amostra a 13000 rpm por 5 minutos, descartando o sobrenadante. Após a lavagem final deixamos o etanol evaporar completamente a 37°C e ressuspendemos o DNA em 200 µl de água ultrapura.

4.5 Amplificação por PCR e purificação do produto

Testamos um total de 4 conjuntos de primers através de diferentes condições de temperatura e concentrações de reagentes e assim estipulamos as condições ótimas a serem utilizadas nas reações de amplificação. Os conjuntos de primers testados foram: yk266f--yk1043r, yk 266f--yk886r, yk601f--yk1043r e yk601f-yk1102r, sendo que o par de primers yk266f--yk1043r apresentou os melhores resultados de amplificação e por isso o escolhemos para ser utilizado neste trabalho.

Uma vez estabelecidas as condições ótimas de amplificação para o conjunto yk266f--yk1043r, as reações foram feitas utilizando-se DNA extraído de indivíduos de diferentes localidades do Brasil. As concentrações dos reagentes utilizados nas reações de amplificação foram as seguintes: Tampão 10X (100mM Tris-Cl pH 9; 500mM KCl; 25mM MgCl₂; 0,1-0,2 mg/ml BSA), BSA (0,1µg/µl), dNTPs (0,2mM), primers forward e reverse (0,2mM) e *Taq* polimerase (0.5U) com 1/100 de *Pfu* polimerase, responsável por reduzir a taxa de incorporação de erros, aumentando a fidelidade na PCR (Cline e cols., 1996).

Corremos os produtos de PCR em géis de agarose 1%, corados com brometo de etídeo e visualizados em luz ultravioleta para verificar a eficiência da reação, e assim estabelecemos as melhores temperaturas e condições de primers e dNTPs na amplificação da região de interesse. Marcadores de peso moleculares foram utilizados para confirmar se o tamanho do fragmento amplificado correspondia ao tamanho esperado.

Os produtos de PCR foram precipitados com uma solução de PEG 8000 (20% de Polietilenoglicol 8000 e 2,5 M NaCl) (Lis & Schleif, 1975). Este procedimento garante a precipitação do produto amplificado e, uma vez que não precipita eficientemente produtos com menos de 200 pares de bases, serve para eliminar artefatos da reação de amplificação como dímeros de primer e nucleotídeos não incorporados.

4.6 Quantificação dos produtos de PCR

O produto purificado é novamente submetido a eletroforese em gel de agarose 1% e corado com brometo de etídeo a fim de se determinar a pureza e concentração dos fragmentos obtidos. Esta concentração é também estabelecida através de Dot Blot (processo de quantificação em gota), onde comparamos a intensidade da fluorescência (proporcional a concentração de DNA aplicado) de cada amostra a padrões conhecidos (marcadores de 10, 20, 30, 40 e 50 ng).

4.7 Clonagem

Clonamos as amostras purificadas em reações de ligação com plasmídeos T&A cloning vector que apresentam resistência à ampicilina, usando o kit de clonagem T&A Cloning Kit Protocol Book (Real Biotech Corporation), de acordo com recomendações do fabricante. A reação de ligação é usada diretamente na transformação de bactérias competentes da cepa DH5 α utilizando o método de choque térmico.

Plaquaremos as colônias recombinantes em meio de cultura LB seletivo (contendo ampicilina (100 μ g/ μ l)) e coletamos colônias individuais após um tempo de aproximadamente 18h. Replaquaremos as colônias amostradas em novas placas seletivas para uso futuro. Amostramos aproximadamente 5 colônias para cada reação de ligação contendo um fragmento diferente de DNA amplificado.

Colocamos estas colônias em água ultrapura por 5 minutos a 90°C para lise das células e posterior liberação dos plasmídeos, e amplificamos este material com primers M13 forward e reverse (0,2mM), que ligam-se próximo ao sítio de clonagem do plasmídeo, para revelar as colônias que incorporaram o inserto. Dessa forma diferenciamos as colônias que incorporaram o inserto de falsos positivos. Os produtos de PCR no tamanho adequado foram novamente purificados através da precipitação com PEG 8000 e ressuspensos em 15 μ l de água ultrapura.

4.8 Sequenciamento

Os produtos amplificados destas clonagens foram purificados, quantificados, e em seguida utilizados em reações de sequenciamento, na concentração adequada para o tamanho do inserto. Fizemos as reações de sequenciamento utilizando os primers M13 forward e M13 reverse a 5mM, e o Kit Dynamic ET Terminator (GE) de acordo com recomendações do fabricante. Além disso, adicionamos à reação DMSO 5% a fim de se melhorar a eficiência de reações contendo insertos ricos em bases nitrogenadas G e C, este mesmo reagente não afeta negativamente regiões que não são ricas em G e C (Choi *et al.*, 1999), aumentando-se, deste modo, a especificidade da reação. Subsequente precipitamos estas reações com Acetato de Amônio 7,5 M e etanol 95% e em seguida as ressuspendemos em tampão de corrida adequado para a eletroforese em um sequenciador automático MEGABACE 4800 do laboratório de Imunogenética, DGE, UFSCar. Alternativamente, os produtos de PCR foram enviados à empresa de sequenciamento Macrogen na Coréia do Sul, onde tanto a reação quanto a corrida do sequenciamento foram realizadas.

4.9 Análises das sequências de DNA

As sequências obtidas foram analisadas com o software Chromas 2.31 (<http://en.bio-soft.net/dna/chromas.html>) que permite a visualização dos eletroferogramas e a verificação da qualidade visual do sequenciamento. Como ambas as fitas foram sequenciadas, esta verificação foi feita pelo alinhamento de ambos os eletroferogramas e obtenção da sequência consenso para o clone em questão. Posteriormente as sequências consenso de cada clone foram transferidas para o software BioEdit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>) para alinhamento e posterior análise filogeográfica.

Para verificar a presença de recombinação utilizamos o programa RPD 3 Beta 34 (<http://www.bioinf.manchester.ac.uk/recombination/programs.shtml>), que compara as sequências três a três, determinando duas mais proximamente relacionadas. Através dessa comparação, é calculado uma porcentagem de identidade nas diferentes posições das sequências, esperando-se encontrar maior proximidade entre as duas sequências que apresentam maior relação. Se em uma determinada região encontrarmos uma maior

porcentagem de identidade entre duas seqüências que apresentam menor relação é possível que esteja ocorrendo, nessa região, recombinação entre essas duas seqüências.

Uma vez efetuada a verificação da presença de recombinação, realizamos uma investigação do nível de polimorfismo, estimando o índice de diversidade de nucleotídeos, através da comparação par a par do número de nucleotídeos diferentes por sítio entre todas as seqüências estudadas e estimamos a média entre todas as combinações tanto dentro das espécies estudadas como entre elas pelo programa Dnasp v. 5.00.03. (Rozas *et al.*, 2003). O programa Dnasp também foi utilizado para a realização de testes de seleção feitos com Tajima D, Fu & Li F e D e Fs de Fu.

O programa computacional TCS versão 1.21 (Clemente *et al.*, 2000) foi utilizado para a obtenção de cladogramas não enraizados de haplótipos que revelam as relações genealógicas entre as seqüências. O TCS aplica a estatística de parcimônia pelo algoritmo desenvolvido por Templeton *et al.* (1992). Os “loops” presentes na rede foram resolvidos utilizando critérios baseados na teoria da coalescência (Crandall & Templeton, 1993) e podem ser resumidos da seguinte maneira: haplótipos são mais prováveis de estarem conectados com haplótipos de maior frequência do que com haplótipos raros, haplótipos são mais prováveis de estarem conectados com haplótipos do interior do que com haplótipos das extremidades e são mais prováveis de estarem conectados com haplótipos da mesma população ou região do que com haplótipos que ocorrem em regiões distantes.

Para análise filogeográfica utilizamos a análise cladística aninhada (Templeton *et al.* 1995) para determinar os principais fatores que possam estar afetando a estrutura populacional das espécies do grupo *fraterculus* no Brasil. Nesta metodologia, utilizamos a parcimônia estatística para inferir relações de parentesco entre os haplótipos amostrados por ser um método desenvolvido especialmente para lidar com filogenias intraspecíficas. Os haplótipos são considerados as menores unidades da análise e por isso considerados clados nível zero. O aninhamento inicia-se na extremidade da rede de haplótipos e move-se um passo mutacional para o interior, unindo os haplótipos conectados a este passo num subconjunto denominado clado de nível 1. O procedimento é repetido até a obtenção de um nível de aninhamento referente a somente um clado (clado final) (Templeton 2008).

Através do Programa Geodis (Posada *et al.* 2000) estimamos os seguintes índices: Distância do clado (Dc) e Distância aninhada (Dn). A análise combinada destes dois índices

permite a distinção entre fatores históricos, como por exemplo expansão da distribuição geográfica, de fatores ecológicos, como fluxo gênico restrito, que influenciaram a distribuição de um táxon. (de Brito *et al.*, 2002). A análise dos dados no programa Geodis evidencia interpretações filogeográficas mais ajustadas que permitem distinguir entre a história e estrutura das populações (Templeton e Sing, 1993; Templeton *et al.*, 1995; Templeton, 1998). Uma chave de inferência é utilizada para interpretar associações significativas entre haplótipos e geografia, fornecendo assim uma inferência biológica apropriada para os nossos dados.

Utilizamos ainda o programa Migrate para estimar taxas de migração que possam estar ocorrendo entre as diferentes espécies aqui estudadas (Beerli, 2001). O programa Migrate utiliza de máxima verossimilhança para estimativa conjunta de theta e M através do uso de uma estratégia de cadeias de Markov Monte Carlo.

5 - Resultados e Discussão

Testamos um total de quatro conjuntos de primers *yolk* (yk266f--yk1043r, yk266f--yk886r, yk601f--yk1043r, yk601f--yk1102r) utilizando pools de DNA de indivíduos de diferentes regiões do Brasil. Após diversos testes de temperatura e concentração de primers, verificamos que o conjunto yk266f--yk1043r amplificou uma única banda de forma consistente que pode ser visualizada em gel de agarose 1%. Além de ter apresentado o melhor resultado de amplificação, esse conjunto permitiu a obtenção de um fragmento longo, de aproximadamente 900pb, sendo assim, o escolhemos para amplificar uma região de interesse do gene *yolk*.

As concentrações de reagentes utilizados na amplificação desta região foram as seguintes: Tampão 10X (100mM Tris-Cl pH 9,0; 500mM KCl; 25mM MgCl₂; 0,1 mg/ml BSA), dNTPs (0,2mM), primers forward (yk 266) e reverse (yk 1043) (0,2mM) e *Taq* polimerase (0,5U) com 1/100 de *Pfu* polimerase. As condições utilizadas nas reações de amplificação estão representadas na tabela 2.

Etapas	T °C	Tempo	ciclos
1	94°	3 min	1
2	94°	30 segs	4
3	61°	30 segs	
4	72°	1 min	
5	94°	30 segs	33
6	58°	30 segs	
7	72°	1min	
8	72°	10 min	

Tab. 2 Condições específicas utilizadas para a amplificação do gene *yolk*

Algumas reações de amplificação continuaram inespecíficas, produzindo um certo borrão ao redor da banda, mesmo variando-se todas as condições de ajuste da reação. Nesse caso retiramos a banda de interesse do gel de agarose, ressuspendendo-a em água para, em seguida, proceder uma reação de reamplificação, com temperaturas maiores, que variavam de 62-65°C.

Amplificamos e/ou reamplificamos a região de interesse para o DNA de 37 indivíduos de diferentes localidades do Brasil, incluindo as três espécies envolvidas neste trabalho: *A. obliqua*, *A. fraterculus* e *A. sororcula*, sendo que, posteriormente, os produtos de amplificação foram clonados e seqüenciados. Para cada indivíduo clonado selecionamos inicialmente 5 colônias a fim de se verificar, através de amplificação com o primer M13, se essas colônias apresentavam o tamanho de fragmento esperado (colônias positivas) para o conjunto de primers em questão. Das 5 colônias selecionadas para cada indivíduo, verificamos os tamanhos dos produtos através de eletroforese em gel de agarose e com a utilização de marcador de peso molecular. Em seguida quantificamos as amostras que possuíam bandas de tamanho esperado e escolhemos duas colônias positivas de cada indivíduo para o sequenciamento, que foi feito utilizando-se primers M13 forward e M13 reverse. A figura 3 representa os diferentes clones positivos utilizados em reações de sequenciamento.

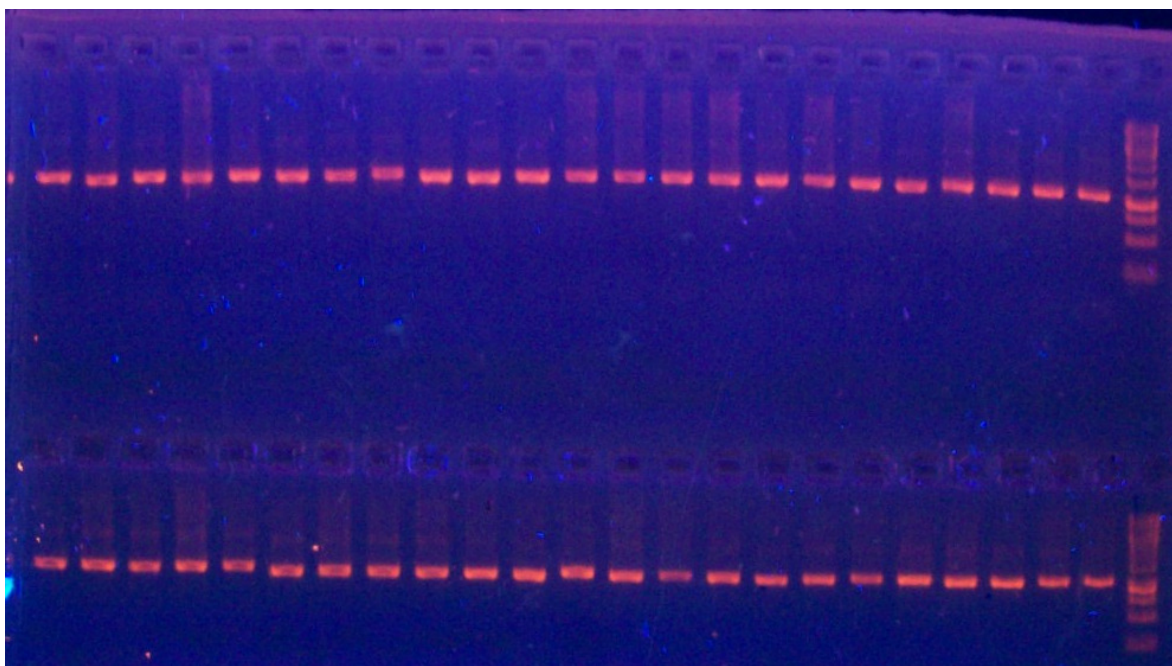


Fig. 3 Representação dos clones positivos de diferentes indivíduos utilizados em reações de sequenciamento

As seqüências foram obtidas através do sequenciamento utilizando primers M13 forward e reverse, sendo que os produtos de PCR purificados foram enviados à empresa de sequenciamento Macrogen na Coréia do Sul, onde tanto a reação quanto a corrida foram

realizadas. As seqüências assim obtidas foram conferidas no GENBANK (<http://www.ncbi.nlm.nih.gov>) através da ferramenta BLAST e correspondem à região esperada do genoma das espécies estudadas neste trabalho. Através da visualização das seqüências no programa Chromas 2.31 (<http://en.bio-soft.net/dna/chromas.html>) verificamos a qualidade das seqüências foward e reverse e inferimos uma seqüência consenso para o posterior alinhamento manual no programa Bioedit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>). Através do alinhamento das seqüências consenso, pudemos identificar polimorfismos nesta região de interesse tanto dentro quanto entre indivíduos. As figuras 4 e 5 representam respectivamente polimorfismos no mesmo indivíduo e entre indivíduos.

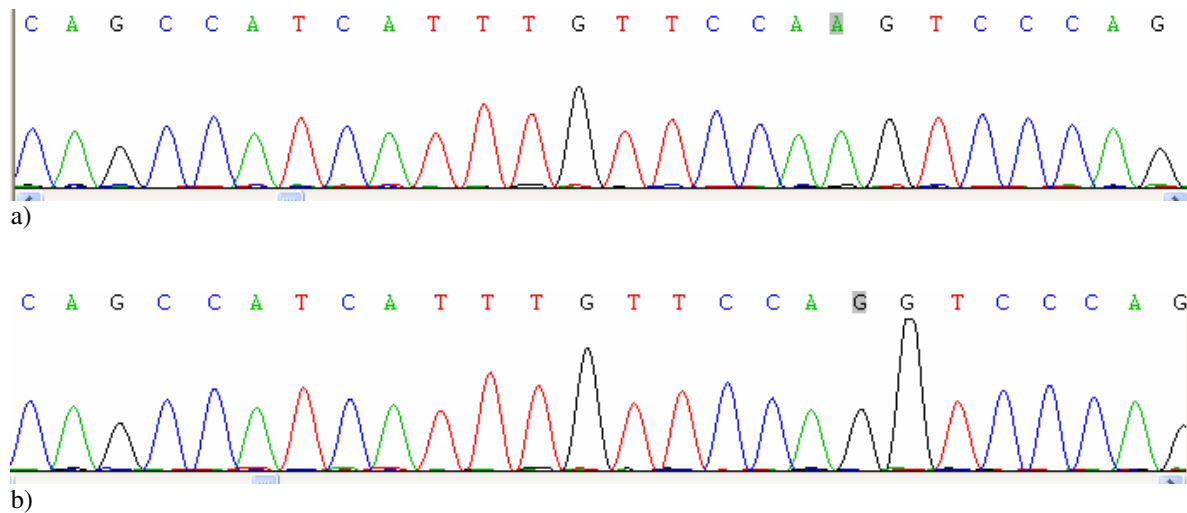


Fig. 4 Eletroferograma mostrando polimorfismo na região marcada entre duas colônias do mesmo indivíduo (a) VALta Cx4G4 colônia 1 (b) VALta Cx4G4 colônia 2

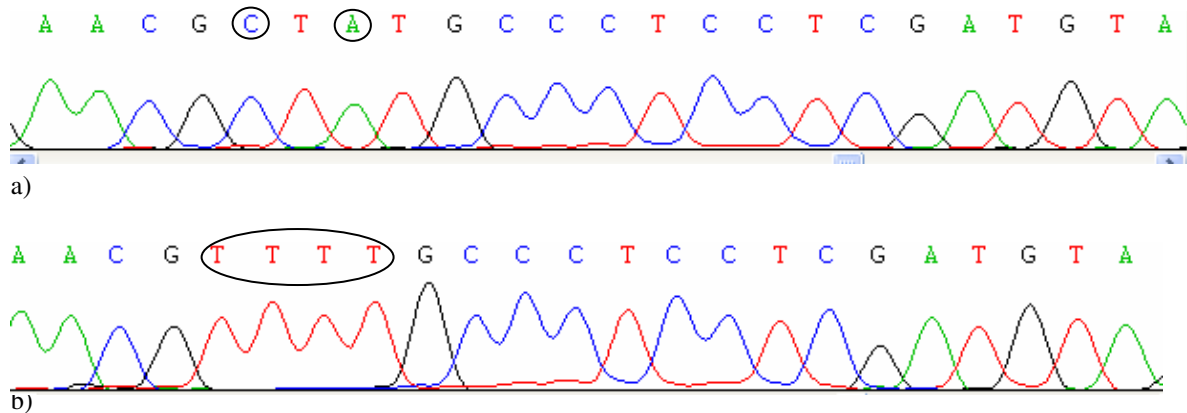


Fig. 5 Eletroferograma mostrando polimorfismos na região marcada entre duas colônias de indivíduos

distintos (a) V.Alta Cx4G4 colonia 2, (b) J. Pessoa Cx2G4 colonia 4)

Embora exista mais de uma cópia do gene *yolk* em diversos dípteros (Sappington, 2002), não existem evidências ainda da existência de mais de uma cópia do gene *yolk* em *Anastrepha*. Apenas um grupo restrito de Diptera foi analisado, e há exceções, como a mosca Tsé-tsé que apresenta apenas uma cópia (Hens *et al.*, 2004). Nossos resultados de sequenciamento indicam ao menos que não encontramos mais de duas cópias de *yolk* nos indivíduos que tiveram mais de duas colônias seqüenciadas (diversos) e não encontramos muita divergência entre as cópias seqüenciadas. Isso sugere que podemos ter apenas uma cópia do gene *yolk*, ou que estejamos amplificando apenas uma, uma vez que a outra cópia presente em outros Diptera apresenta já uma certa divergência molecular de forma a impedir sua amplificação com os primers aqui utilizados.

Através da utilização do programa RDP 3 Beta 34 (<http://www.bioinf.manchester.ac.uk/recombination/programs.shtml>), detectamos a presença de recombinações em algumas seqüências referentes aos seguintes clones: Belém Cx5A85, Babaçulândia Cx1E311, 3 Lagoas Cx1A25, 3 Lagoas Cx1A28, Babaçulândia Cx1E12, Campinas Cx3E23, Babaçulândia Cx1E14, Marilac Cx3G34, Estrela D'Oeste Cx2G117, Campinas Cx5E12, Piracicaba Cx4E81, Florianópolis Cx2H21, C.Grande Cx4F53, V. Alta Cx4G34, C.Frio Cx5B41, Florianópolis Cx2H23, V. Alta Cx4G42, João Pessoa Cx2G42, C. Frio Cx5B31, C. Frio Cx5B32, Bonito Cx2C86, sendo que essas recombinações localizavam-se na região de aproximadamente 170pb e outra na região de aproximadamente 530pb. Como a recombinação mistura sinais históricos de parentesco, temos que removê-las antes das análises filogenéticas. Dessa forma, Templeton *et al.*, (2000) sugerem duas alternativas para resolver o problema de recombinação nas seqüências. Em uma delas, removemos as seqüências recombinantes e mantemos a região inteira analisando apenas indivíduos que não sofreram recombinações. Alternativamente, fazemos a partição dos dados, ou seja, dividindo a seqüência toda em regiões que não sofreram recombinação, tratando-as como regiões distintas. No caso do gene *yolk*, tivemos que particionar os 908 pb em três partes. A primeira parte englobando o primeiro éxon, o primeiro íntron e parte do segundo éxon, caracterizando uma região de aproximadamente 247 pb. A segunda parte englobando somente o segundo éxon, com aproximadamente 337

pb e a terceira parte englobando o terceiro éxon, com cerca de 163 pares de base. A representação gráfica dos genes *yolk* em *Drosophila*, bem como a partição efetuada neste trabalho referente à cópia única do gene *yolk* em *Anastrepha* estão representados na figura 6. O par de primers utilizados neste trabalho amplificou grande parte do gene *yolk* em *Anastrepha*.

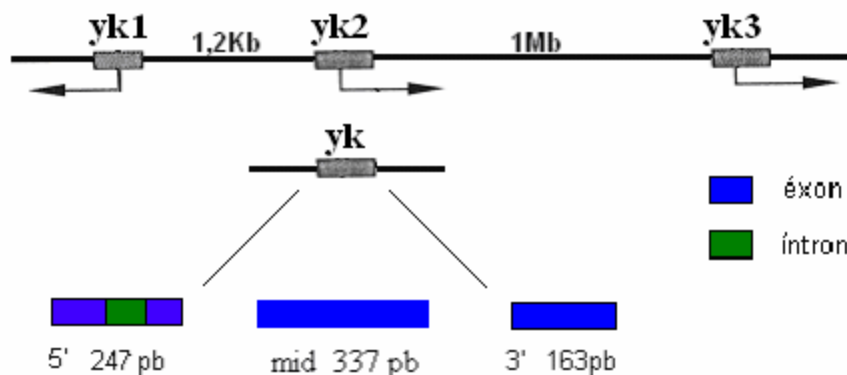


Fig. 6- Representação esquemática dos genes *yolk* em *Drosophila* (*yk1*, *yk2*, *yk3*) e subdivisão em três partes do gene *yolk* em *Anastrepha*. As cores azul e verde representam respectivamente éxons e íntrons. É importante ressaltar que para *Anastrepha* não há evidência da presença de outras cópias do gene *yolk*

A subdivisão do gene *yolk* em partes reduziu para a grande maioria das seqüências o efeito das recombinações, embora tenhamos detectado ainda efeitos de recombinação na seqüência Estrela D'Oeste Cx2G117 na região mid, e assim a removemos da análise para essa região.

5.1 Análises de Polimorfismo e Divergência entre as regiões gênicas

Como mencionado anteriormente, o gene *yolk* amplificado neste trabalho apresenta cerca de 900 pb e foi subdividido em três partes, sendo que somente a primeira delas apresenta um íntron. Através do programa DnaSP v. 5.00.03 (Rozas, *et al.*, 2003) analisamos o nível de polimorfismo e divergência para as três partes nas três espécies envolvidas no trabalho. A porcentagem média de nucleotídeos diferentes por sítio entre

seqüências define um parâmetro importante de diversidade molecular, a diversidade nucleotídica. Este parâmetro, bem como o número de sítios polimórficos e os valores de theta (indicativo da quantidade de polimorfismos mantido na população pelo balanço entre deriva genética e mutação) por seqüência foram utilizados para determinar o grau de polimorfismo dentro e entre as espécies.

Todas as regiões nas três espécies mostraram-se altamente polimórficas e isso ficou evidenciado pelos altos valores de diversidade nucleotídica e pelo número de sítios polimórficos. Na primeira parte do gene (região 5') encontramos um íntron altamente polimórfico, sendo este polimorfismo uma ferramenta eficaz no estudo comparativo entre populações.

Esses valores também mostraram que o índice de diversidade é ainda maior na região final (3') do gene, sendo que esta região é a que apresenta o menor número de nucleotídeos analisados. Os valores de theta também mostraram-se elevados e são sempre menores em *A. sororcula* possivelmente porque amostramos somente 7 indivíduos desta espécie. O número de sítios polimórficos, em grande parte do gene, é maior em *A. obliqua*, mesmo amostrando-se uma menor quantidade de indivíduos dessa espécie comparado à *A. fraterculus*.

A tabela 3 representa os níveis de polimorfismo através dos dados de diversidade nucleotídica (Pi), bem como o número de sítios polimórficos (S) e valores de theta por nucleotídeo (Θ) nas três espécies em questão para as três regiões do gene *yolk*.

	5'			Mid			3'		
	Pi	Θ	S	Pi	Θ	S	Pi	Θ	S
<i>A.obliqua</i>	0,0227	0,0495	48	0,0182	0,0355	47	0,0300	0,0408	26
<i>A.fraterculus</i>	0,0179	0,0536	40	0,0215	0,0449	60	0,0341	0,0361	23
<i>A.sororcula</i>	0,0169	0,0278	19	0,0215	0,0267	28	0,0321	0,0310	16

Tab. 3 Representação dos índices de diversidade nucleotídica, theta e número de sítios polimórficos para as três espécies de *Anastrepha* referentes às regiões particionadas do gene *yolk*

Os índices de diversidade nucleotídica também mostraram-se elevados quando comparamos as três espécies duas a duas. Nesse caso, verificamos os maiores valores de diversidade nucleotídica total entre as espécies *A. obliqua* e *A. fraterculus* para as três

regiões gênicas, e os menores valores comparados de diversidade nucleotídica total entre as espécies *A. fraterculus* e *A. sororcula* também para as três regiões em questão. Essas comparações revelam que pode haver uma menor diferenciação entre *A. fraterculus* e *A. sororcula*, mostrando que estas espécies podem estar mais próximas filogeneticamente quando comparamos uma dessas espécies à *A. obliqua*. Morfologicamente, a maior dificuldade em se distinguir um indivíduo *A. fraterculus* de um indivíduo *A. sororcula* também representa uma possível evidência a mostrar uma maior aproximação entre essas duas espécies.

Outro ponto importante da análise é que embora um número reduzido de indivíduos tenha sido seqüenciado, este número já foi suficiente para indicar que não existem sítios que estejam fixados em uma espécie e não estejam presentes em outra. Isso mostra a dificuldade que pode existir na descoberta de marcadores eficazes que possam estar distinguindo amplamente essas espécies.

Outro índice importante é o que mede a relação entre as taxas de substituições não sinônimas e sinônimas. Através da obtenção deste índice, é possível determinar se os trechos gênicos analisados estão sob influência de algum tipo de seleção ou neutralidade. Os resultados obtidos para essa análise comprovam que, apesar de encontrarmos substituições que promovam alterações a nível de aminoácidos, verificamos que, no geral, a maioria das substituições acontece na terceira base do códon, não levando à substituição do aminoácido. Todos os valores de Ka/Ks entre as espécies nos três trechos indicaram para uma possível ocorrência de seleção purificadora, sendo que sob o efeito deste tipo de seleção temos um processo de “filtragem” acontecendo a fim de remover alelos com efeitos deletérios na população. Os valores da relação entre as diversidades nucleotídicas não sinônimas e sinônimas foram menores entre *A. fraterculus* e *A. sorocula* para duas das três regiões gênicas analisadas, e maiores entre *A. fraterculus* e *A. obliqua* para as três regiões gênicas, evidenciando novamente uma possível maior aproximação entre as espécies *A. fraterculus* e *A. sororcula*. A tabela 3 representa os índices de diversidade nucleotídica comparados, bem como os valores de Ka/Ks comparados, para as regiões 5', mid e 3' respectivamente.

	5'		Mid		3'	
	Pi	Ka/Ks	Pi	Ka/Ks	Pi	Ka/Ks
<i>Afr-Aob</i>	0,02195	0,42	0,02272	0,42	0,03312	0,24
<i>Afr-Aso</i>	0,01831	0,30	0,02039	0,33	0,03225	0,23
<i>Aob-Aso</i>	0,02119	0,32	0,02153	0,35	0,03335	0,20

Tab. 4 Representação dos índices de diversidade nucleotídicas e ka/ks comparados para as três regiões do gene *yolk*

A análise dos dados mostram que o gene *yolk* apresentou elevado nível de variação tanto intra, quanto interespecificamente. A análise comparada destes valores com valores obtidos para outros genes nucleares, o gene *fruitless* e o gene *doublesex*, investigados nas mesmas espécies de *Anastrepha* mostrou um padrão similar de alto nível de polimorfismo. Estudos preliminares feitos em nosso laboratório apontaram para altos índices de diversidade nucleotídica e theta para as três espécies de *Anastrepha* quando analisamos os genes *fruitless* e *doublesex* (Sobrinho Júnior, 2009). No entanto, os valores de diversidade nucleotídica e theta para esses dois genes foram ligeiramente menores quando comparados ao gene *yolk*, o que pode ser reflexo da existência de dois íntrons nas sequências de *yolk* analisadas. A tabela 5 abaixo compara os valores de diversidade nucleotídica e theta entre as três espécies para as partições do gene *yolk*, para o gene *doublesex* e *fruitless*.

	<i>yolk</i>						<i>fru</i>		<i>dsx</i>	
	5'		Mid		3'					
	Pi	Θ	Pi	Θ	Pi	Θ	Pi	Θ	Pi	Θ
<i>A.obliqua</i>	0,0227	0,0495	0,0182	0,0355	0,030	0,0408	0,0126	0,0268	0,0106	0,0107
<i>A.fraterculus</i>	0,0179	0,0536	0,0215	0,0449	0,034	0,0361	0,0108	0,2774	0,0178	0,0345
<i>A.sororcula</i>	0,0169	0,0278	0,0215	0,0267	0,032	0,0310	0,0136	0,0252	0,0161	0,0245

Tab. 5 Valores comparados de diversidade nucleotídica e theta referentes às partições do gene *yolk*, *fruitless* e *doublesex* para as três espécies de *Anastrepha*

Outro ponto importante está no fato de que tanto dados mitocondriais (Smith-Caldas *et al.*, 2001), como estes dados obtidos de genes nucleares em nosso laboratório revelaram em geral uma ausência de marcadores específicos de espécie para o grupo *fraterculus*. Este padrão pode ser indicação de divergência recente entre as espécies deste grupo aqui

estudadas ou pode ainda refletir a existência de fluxo gênico na natureza entre estas espécies, ou mesmo um problema maior taxonômico na identificação de espécies do grupo.

5.2 Testes de Neutralidade e Distribuição Mismatch

Para a teoria neutra das mutações, a maioria das substituições de bases que segregam na população e se tornam fixadas em uma população é neutra com respeito ao sucesso reprodutivo (fitness) (Kimura, 1968).

Atualmente muitos testes são utilizados em genética de populações a fim de se determinar a ocorrência ou não de neutralidade seletiva em populações naturais. Para o presente trabalho utilizamos os testes de D de Tajima (Tajima, 1989), D e F de Fu & Li (Fu & Li, 1993) e Fs de Fu (Fu, 1996; Fu, 1997).

O Teste D de Tajima (Tajima, 1989) utiliza a variação genética dentro das populações. Esta variação é calculada a partir da relação entre o número de sítios segregantes e o número médio de diferenças nucleotídicas estimado pela comparação par-a-par. Os testes D e F de Fu & Li também utilizam-se de polimorfismos de DNA, porém, levam em consideração o fator temporal das mutações que geram sítios polimórficos, classificando-as como mutações antigas ou mais recentes (Fu & Li, 1993). Todos estes testes não têm o mesmo poder de detectar desvios da neutralidade causada por diferentes forças evolutivas. A dificuldade da análise destes testes de neutralidade vem do fato de todos eles serem também afetados por processos populacionais como expansão ou gargalos. Além disso, processos distintos, como “hitchhiking”, “background selection” ou “selective sweep” (Fu, 1997) afetam diferentemente a cada um destes testes, sendo que seu uso combinado pode permitir uma inferência dos padrões de seleção que afetam regiões específicas de DNA (de Brito, *et al* 2002).

Todos os testes de neutralidade realizados, tanto os de Tajima como os de Fu e Li mostraram-se significativos, com exceção ao teste de Tajima para a região 3', que apresentou um valor de -1,27042 ($P > 0,10$). A tabela 6 mostra os valores de significância para as três regiões gênicas em todos os testes efetuados. Esses valores de significância rejeitam a hipótese de evolução neutra para estas regiões, corroborando com o resultado

anterior sobre uma possível ocorrência de seleção, que de acordo com os valores de Ka/Ks , pode vir a ser do tipo purificadora.

Testamos também, através das três regiões gênicas, o padrão de distribuição de distâncias par-a-par (mismatch distribution), que refere-se ao número analisado de sítios diferentes par-a-par nas seqüências amostradas. Inicialmente consideramos para análise de nossos dados o padrão referente à população estável, em um segundo momento consideramos população em expansão. As Figuras 7 e 8 representam graficamente para população estável e outra para população em expansão, respectivamente, para as regiões 5', mid e 3'. A curva verde representa os valores esperados, enquanto a vermelha corresponde aos valores observados.

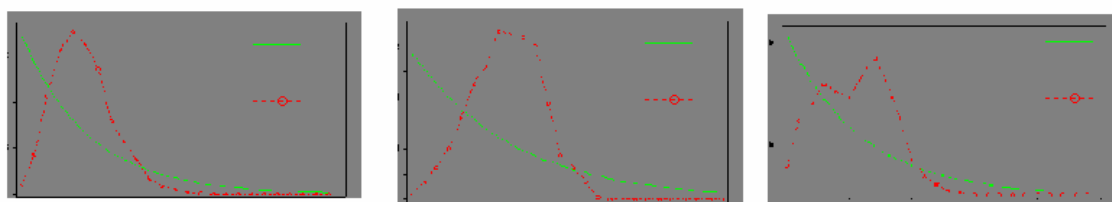


Fig. 7 Representação gráfica de população estável, referente às regiões 5', mid e 3', respectivamente

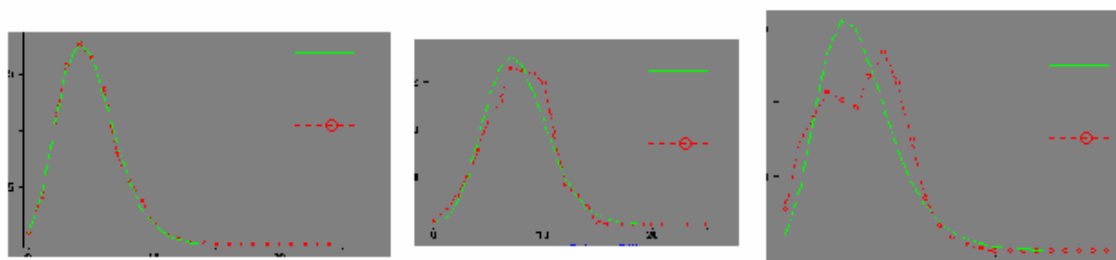


Fig. 8 Representação gráfica de população em expansão, referente às regiões 5', mid e 3', respectivamente

Em nossas análises não podemos eliminar eventos demográficos como o crescimento populacional tendo afetado nossas amostras, desta forma, utilizamos como contraste para análise em mismatch a população em crescimento e assim, observamos que os resultados obtidos para as regiões 5' e mid são condizentes com valores esperados para um crescimento do tamanho populacional. Os valores referentes aos dois principais parâmetros utilizados (r , $R2$) para determinar o padrão de distribuição mismatch estão

representados na tabela 6. Os valores significativos de r e $R2$ são indicativos de que possa ter ocorrido crescimento do tamanho populacional para as regiões 5' e mid.

	5'	Mid	3'
Tajima D	-2,610***	-2,1635*	-
Fu&Li F	-5,267**	-4,665**	-3,432**
Fu&Li D	-5,584**	-5,043**	-3,868**
Fu&Li Fs	-95,46	-85,71	-47,45
r	0,0136*	0,0069*	0,0072*
R2	0,0207*	0,0302*	0,0561*

P<0,05* P<0,02** P<0,001 ***

Tab. 6 Representação dos valores de neutralidade (Tajima D, Fu & Li F e D, Fu & Li Fs), bem como os valores de distribuição mismatch (r e $R2$) para as três regiões do gene *yolk*.

Deve-se ressaltar que o padrão de coleta aqui realizado também pode gerar os resultados descritos anteriormente referentes à ocorrência de possível expansão populacional ou ainda seleção do tipo purificadora, haja visto que coletamos poucos indivíduos de diversas partes do país, buscando fornecer uma primeira informação geral do status deste grupo de espécies no país. Desta forma, estamos potencialmente enviesando nossos resultados para mutantes raros de baixa frequência, o que é o padrão condizente também com o esperado em expansões populacionais ou na seleção purificadora.

5.3 Análise das redes de haplótipos

A genealogia entre os haplótipos foi inferida por meio da construção de 3 redes distintas, referentes às 3 regiões gênicas previamente indicadas. A relação entre os haplótipos foi estimada através dos métodos de parcimônia estatística, em que os agrupamentos são feitos baseados em mutações que ocorrem ao longo das seqüências. Por serem baseadas em cálculos referentes à frequência desses haplótipos nas populações amostradas utilizando-se um algoritmo específico descrito por Templeton *et al.*, (1992) as redes haplotípicas são responsáveis por produzir uma reconstrução evolutiva da história genealógica da variação encontrada nas amostras de DNA que experimentaram pouca ou nenhuma recombinação (Templeton, 2001). Daí a importância de um conhecimento

adequado sob a ocorrência de recombinação em um determinado trecho gênico, pois dessa forma podemos minimizar seus efeitos e inferir uma história genealógica mais coerente para um determinado grupo.

Com base na parcimônia foram geradas três diferentes redes haplotípicas referentes às regiões 5', mid e 3' como representado nas Figuras 9, 10 e 11 respectivamente. Nas redes geradas pelo programa TCS 1.21 (Clemente *et al.*, 2000) os haplótipos são representados por elipses, sendo que o tamanho das elipses é proporcional ao número de indivíduos encontrados para cada haplótipo. Não houve nenhum haplótipo com número muito grande de indivíduos e por isso as elipses nas três redes, em geral, não variam muito em tamanho. Os círculos menores, que ligam os haplótipos identificados, correspondem aos haplótipos não amostrados e são classificados como intermediários, e os haplótipos representados por retângulos correspondem aos haplótipos ancestrais. Nas três redes, os haplótipos identificados foram representados por elipses, os haplótipos inferidos (intermediários) pelo número 0, e os ancestrais, inferidos por coalescência, foram representados por retângulos.

A rede haplotípica referente à região 5' apresentou ao todo 61 haplótipos, sendo que o mais freqüente foi o IX, presente em 7 seqüências. Outros haplótipos apareceram mais de uma vez, o XLVII e o XXIX esteve representado em duas seqüências cada um, e o haplótipo XXII esteve representado em três seqüências. A tabela 7 representa os haplótipos e suas localizações referentes à região 5'. O símbolo * seguido de um número cardinal significa que um mesmo haplótipo foi representado mais de uma vez na mesma localidade.

Localidades	Haplótipos
Araguaina TO	VI, XVII, XXXIX, XL
Babaçulândia TO	XIII, XIV, XVI
Bauru SP	XXVI, LXI
Belém PA	IX, XII, XLI, XLIV
B.V. de Goiás GO	XXIX, XXX
Bonito PE	IX, VII, XXXII, LVII, LVIII, LX
Cabo Frio RJ	XXII, III, X, XXIII
Campinas SP	II, XXVIII, XLVIII, LIX
Campo Grande MS	XXII, XXXVI, L, LI
Curitiba PR	XX, XXIV
Estrela D'Oeste SP	V, VIII, XLV

F. de Santana BA	LII, LIII
Florianópolis SC	XXXVII, XXXVIII
Gurupi TO	XV, LV
João Pessoa PB	I, XXXIII
Linhares ES	XVIII
Joinville SC	XXVII, XXXV
Marilac MG	IX, XXV, XLIII, XLVI
Palmas TO	IX * 3, XI, XIX, XLIV
Piracicaba SP	IX, IV
Recife PE	XLVII * 2
São Carlos SP	XXI, XLIX
Três Lagoas MS	LIV, LVI
Vargem Alta ES	XXIX, XXII, XXXI, XXXIV

Tab.7 Representação dos haplótipos e suas respectivas localidades referentes à região 5'

A rede da região 5' é caracterizada por conter o ancestral (IX) central, sendo que este centro apresentou reticulação, com vários ramos curtos de poucos passos mutacionais gerando haplótipos mais recentes. Desse mesmo ancestral, três grandes ramos geraram três outros haplótipos também antigos (não ancestrais) e com reticulações, sendo que estes produziram haplótipos mais recentes, com poucos passos mutacionais entre eles. Esta rede, portanto, pode ser caracterizada pela presença de quatro grandes regiões: uma região central reticulada de onde saem três longos ramos gerando outras três regiões mais externas também reticuladas. Os haplótipos ancestrais pertencem em sua maioria à espécie *A. obliqua* oriundos da região Norte-Nordeste do país, gerando três ramos independentes, cada um deles com somente um passo mutacional. Um desses ramos dá origem somente a haplótipos *A. obliqua*, sendo que estes são pertencentes à região Norte-Nordeste do Brasil, excetuando-se o haplótipo pertencente à localidade Três Lagoas - MS. Os outros dois ramos oriundos dos mesmos ancestrais deram origem à maioria de haplótipos *A. fraterculus*, sendo que grande parte destes pertencem à região Centro-Sul do Brasil.

A rede de haplótipos da região 5' está representada na figura 9, com as cores amarelo, azul e vermelho representando respectivamente os haplótipos pertencentes às espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*.

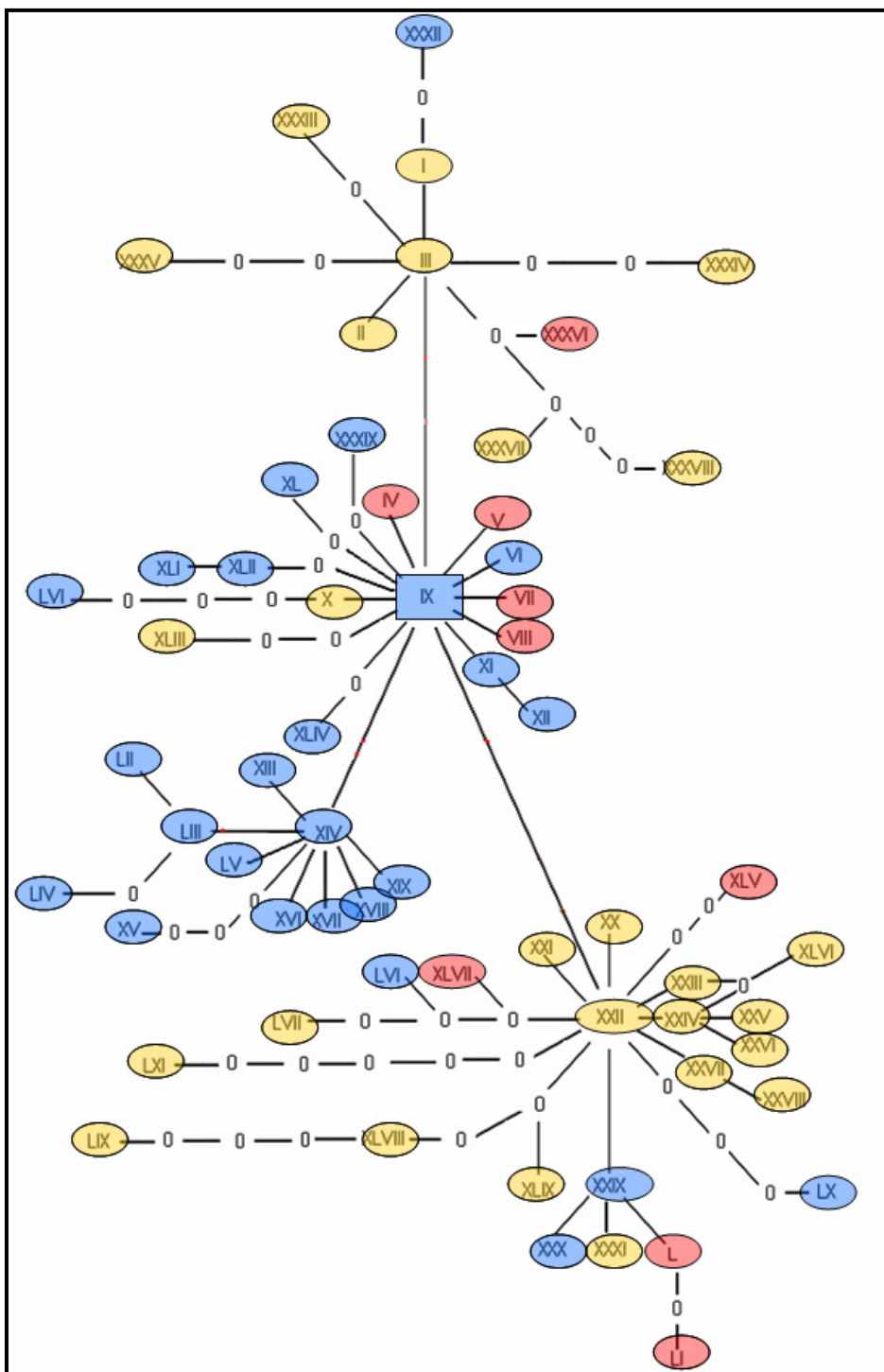


Fig. 9 Rede haplotípica referente à região 5'

A rede correspondente à região mid apresentou um total de 66 haplótipos, sendo que a frequência de haplótipos repetidos foi inferior comparado à região 5'. O haplótipo XIII

foi o mais freqüente, estando presente em quatro sequências. Outros haplótipos estiveram presentes em mais de uma sequência, o VII esteve presente em três sequências e o XXVI foi representado em duas sequências. A tabela 8 representa os haplótipos e suas localizações referentes à região mid. O símbolo * seguido de um número cardinal significa que um mesmo haplótipo foi representado mais de uma vez na mesma localidade. O haplótipo XIV não está representado e por isso temos o haplótipo LXVII (Babaçulândia).

Localidades	Haplótipos
Araguaina TO	VIII, XXIV, XXV, XXVI
Babaçulândia TO	III, XXXI, LXVI, LXVII
Bauru SP	XXII, XLIX
Belém PA	VII * 3, V
B.V. de Goiás GO	XLVII, XLVIII
Bonito PE	XIII* 2, XV, XXXII, XXXIV, XXXV
Cabo Frio RJ	XXI, LVI, LXIV, LXV
Campinas SP	XLIII, LIII, LIV, LVII
Campo Grande MS	XVIII, XIX, XX, XXXIII
Curitiba PR	L, LII
Estrela D'Oeste SP	XII, XLVI, LV
F. de Santana BA	IX, XXVII
Florianópolis SC	XLI, XLII
Gurupi TO	X, I
João Pessoa PB	LXI, XXIX
Linhares ES	LIX
Joinville SC	XI, XVII
Marilac MG	XVI, XXX, XXXVII, XXXVIII
Palmas TO	XIII, II, IV, XXXVI, XXXIX, XL
Piracicaba SP	XIII, LI
Recife PE	XXVIII, LX
São Carlos SP	LXII, LXIII
Três Lagoas MS	I, VI
Vargem Alta ES	XXIII, XLIV, XLV, LVIII

Tab. 8 Representação dos haplótipos e suas respectivas localidades referentes à região mid

A rede haplotípica mid apresenta maior complexidade comparada às outras redes.

O haplótipo ancestral (I) encontra-se na região superior da rede e origina haplótipos recentes, com poucos passos mutacionais entre eles. Através de um ramo interior é produzido o haplótipo X, sendo que deste haplótipo saem quatro ramos que vão determinar a estrutura restante da rede. Em dois desses ramos, temos a origem direta de dois haplótipos (XI e XXX) através de somente um passo mutacional, nos outros dois ramos tem-se uma estrutura mais complexa, sendo que em um deles temos a presença de grande quantidade de indivíduos referentes à espécie *A. fraterculus*. O haplótipo ancestral superior (I) pertence a espécie *A. obliqua*, originando na parte superior da rede somente haplótipos *A. obliqua* pertencentes às regiões Norte-Nordeste do Brasil. Os haplótipos pertencentes à espécie *A. fraterculus* concentram-se na parte inferior da rede e parecem ter se originado (em grande parte) através de um haplótipo da espécie *A. sororcula* (XX) pertencente à localidade Campo Grande- MS. A rede de haplótipos da região mid está representada na Figura 10, com as cores amarelo, azul e vermelho representando respectivamente os haplótipos pertencentes às espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*. O haplótipo XIV não está representado na rede mid e por isso temos o haplótipo LXVII.

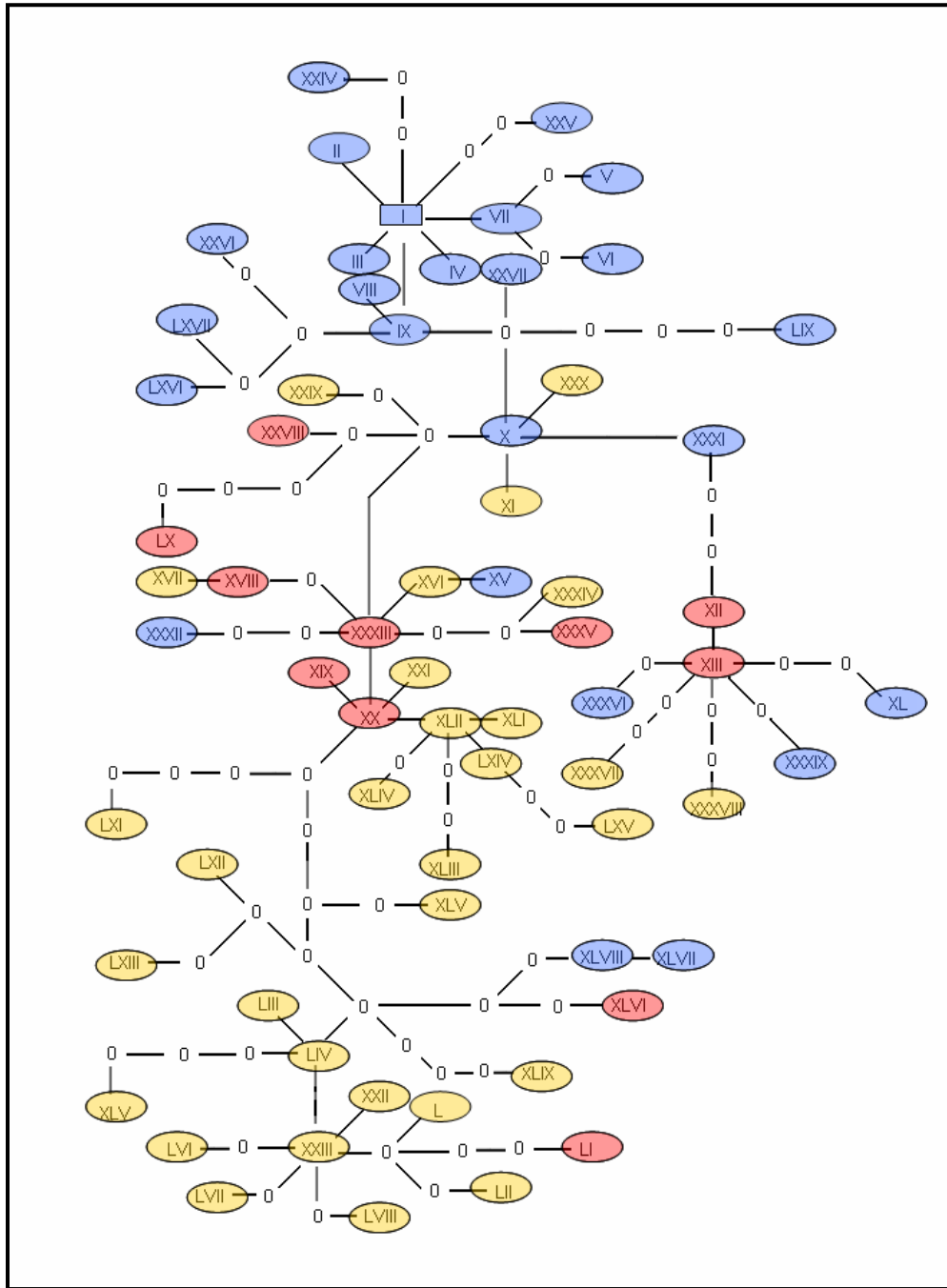


Fig. 10 Rede de haplótipos referente à região mid

A rede haplotípica 3' é a que apresenta o maior número de haplótipos repetidos e isso pode estar diretamente relacionado à menor quantidade de bases presentes nessa região. O haplótipo I é o mais freqüente e aparece em sete seqüências, o XXIX em cinco seqüências, o XIV em quatro seqüências, os haplótipos XVIII e XXI aparecem em três seqüências cada, e os haplótipos VII e XIX aparecem em duas seqüências cada. A tabela 9

representa os haplótipos e suas localizações referentes à região 3'. O símbolo * seguido de um número cardinal significa que um mesmo haplótipo foi representado mais de uma vez na mesma localidade.

Localidades	Haplótipos
Araguaina TO	IV, VIII, VII, I
Babaçulândia TO	I, XLVII, XLVIII, L
Bauru SP	III, XXXIII
Belém PA	I*3, II
B.V. de Goiás GO	XIII, LIV
Bonito PE	XXIX*2, XIV, XXIV, XXXV, XXXVIII
Cabo Frio RJ	V, XXII, XVIII, XXI
Campinas SP	XX, XXXII, XLIV, XXI
Campo Grande MS	XXXIX, XLV, XLVI, LI
Curitiba PR	XXXVI, XXXVII
Estrela D'Oeste SP	XVII, XVIII, XXVII, LIII
F. de Santana BA	IX, X
Florianópolis SC	XIX*2
Gurupi TO	I, VII
João Pessoa PB	XXI, XIV
Linhares ES	XXXI
Joinville SC	XXXII, XVI
Marilac MG	XIV, XXIII, XXV, XLIX
Palmas TO	XXIX*2, I, XXVIII, XXX, XLII
Piracicaba SP	LII, XXIX
Recife PE	XIV, XV
São Carlos SP	XLI, XLIII
Três Lagoas MS	XXVI, VI
Vargem Alta ES	XI, XII, XL, XVIII

Tab.9 Representação dos haplótipos e suas respectivas localidades referentes à região 3'

A rede haplotípica 3' é mais simples, com a presença de 54 haplótipos, sendo que o haplótipo ancestral (I) é representado somente por seqüências de *A.obliqua*. Na parte superior da rede a partir desse ancestral surgem vários haplótipos recentes, com poucos passos mutacionais entre eles. A região central da rede é dividida em quatro ramos a partir

de um haplótipo inferido, sendo que em dois desses ramos são gerados dois haplótipos através de somente um passo mutacional e em um terceiro ramo temos na parte inferior duas reticulações. Na rede 3' não observamos qualquer passo mutacional que origine somente haplótipos de uma espécie. Os haplótipos das três espécies misturam-se ao longo da rede, tanto na parte superior quanto na inferior, e a menor estruturação haplotípica referente à região 3' do gene *yolk* pode ter sido determinada pela menor quantidade de nucleotídeos amostrados para esta região, sendo tal fato observado pelo aumento no número de haplótipos semelhantes. A rede de haplótipos da região 3' está representada na figura 11, com as cores amarelo, azul e vermelho representando respectivamente os haplótipos pertencentes às espécies *A. fraterculus*, *A. obliqua* e *A. sororcula*

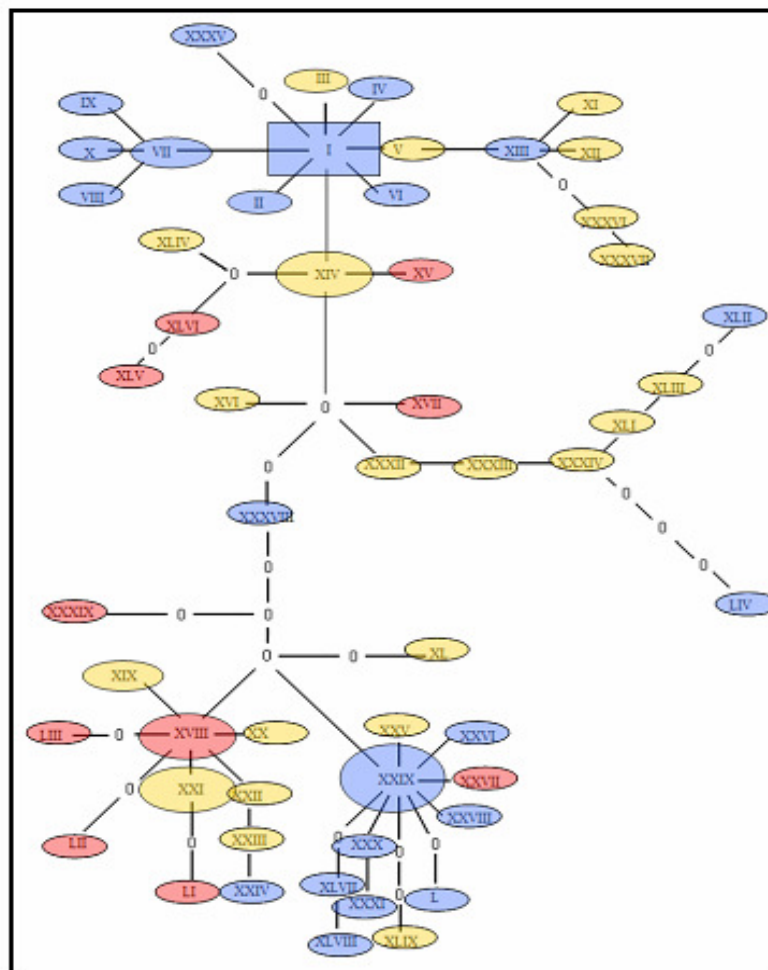


Fig. 11 Rede de haplótipos referente à região 3'

5.4 Nested Clade Phylogeographical Analysis (Análise filogeográfica dos clados aninhados)

A metodologia de Nested Clade Phylogeographical Analysis foi aplicada com o objetivo de se diferenciar eventos históricos da estrutura genética das populações (Templeton, 1998; Templeton, 2001). A análise de clado aninhado foi aplicada nas três redes haplotípicas de acordo com as regras de Templeton (1995). O aninhamento inicia-se nas extremidades da rede e movem-se um passo mutacional para o interior, conectando os haplótipos ligados por este passo, de modo que o subconjunto contendo esses haplótipos refere-se ao clado nível um. Os clados de nível dois são formados utilizando as mesmas normas de aninhamento, tendo os clados de nível um como unidades. O processo de aninhamento prossegue até se atingir o clado final, onde todos os haplótipos estarão incluídos.

As figuras de 12-22 representam os aninhamentos em seus diferentes níveis, separadamente, para as regiões 5', mid e 3', respectivamente.

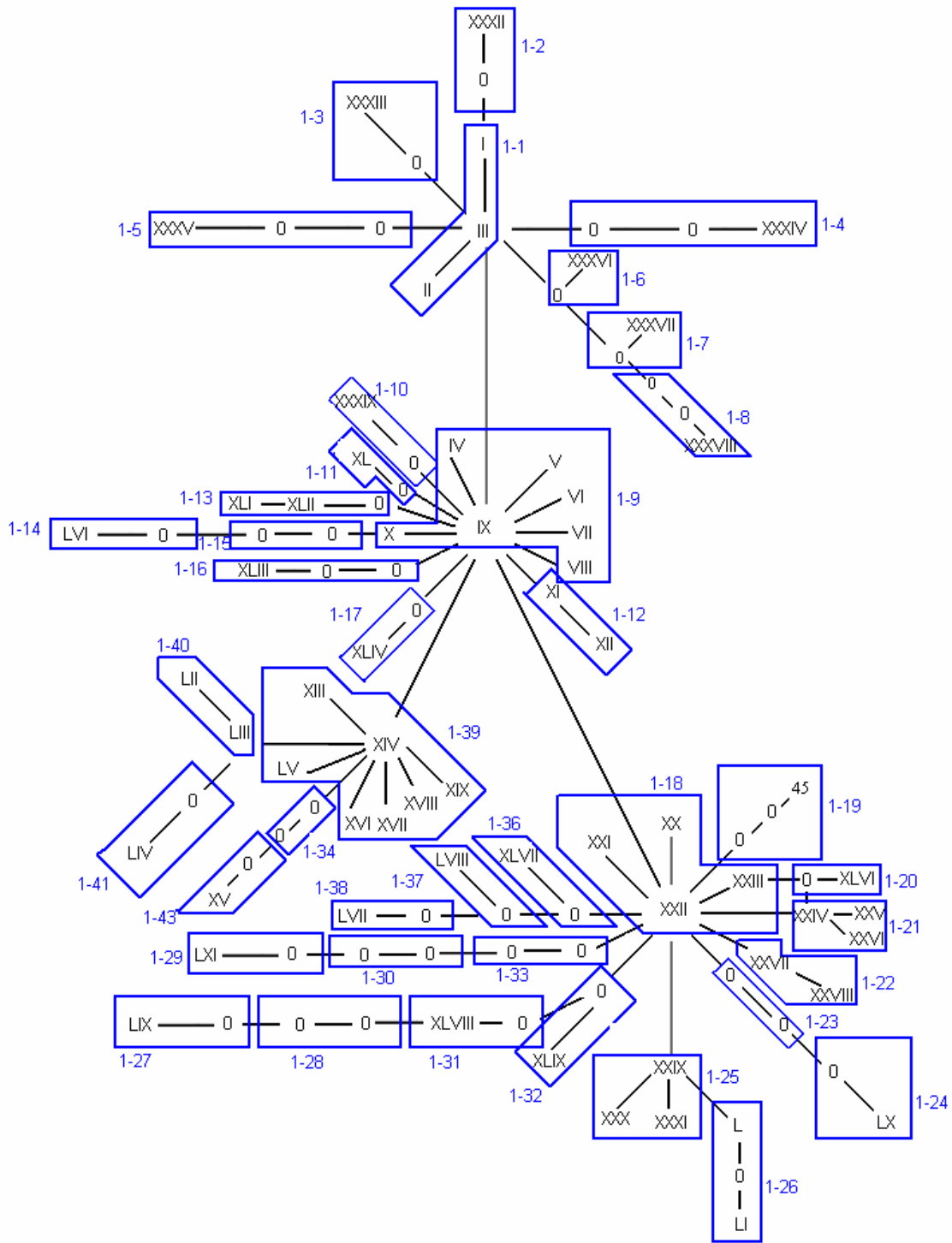


Fig. 12 Nível 1 de aninhamento para a rede 5'

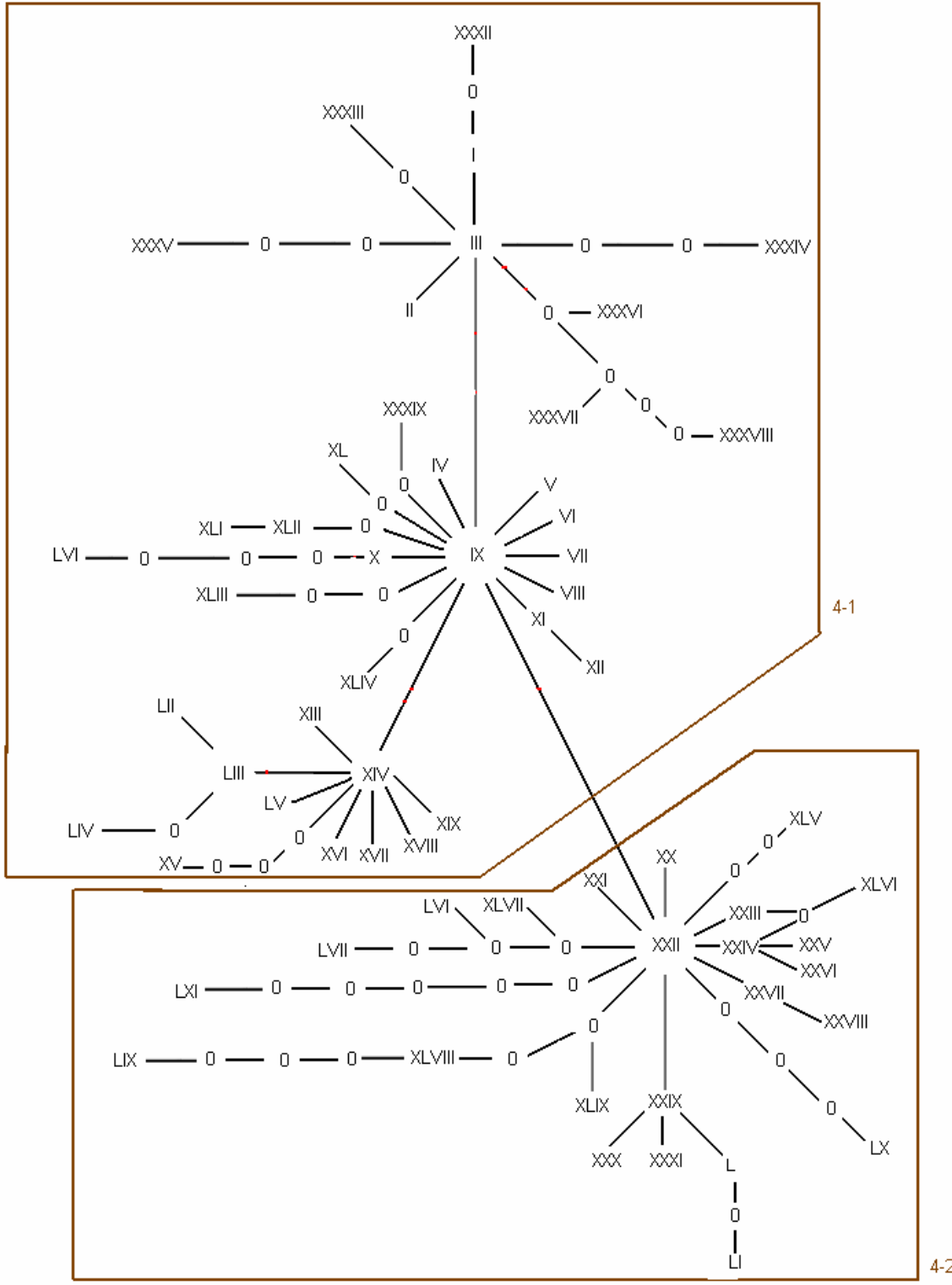


Fig. 15 Nível 4 de aninhamento para a rede 5'

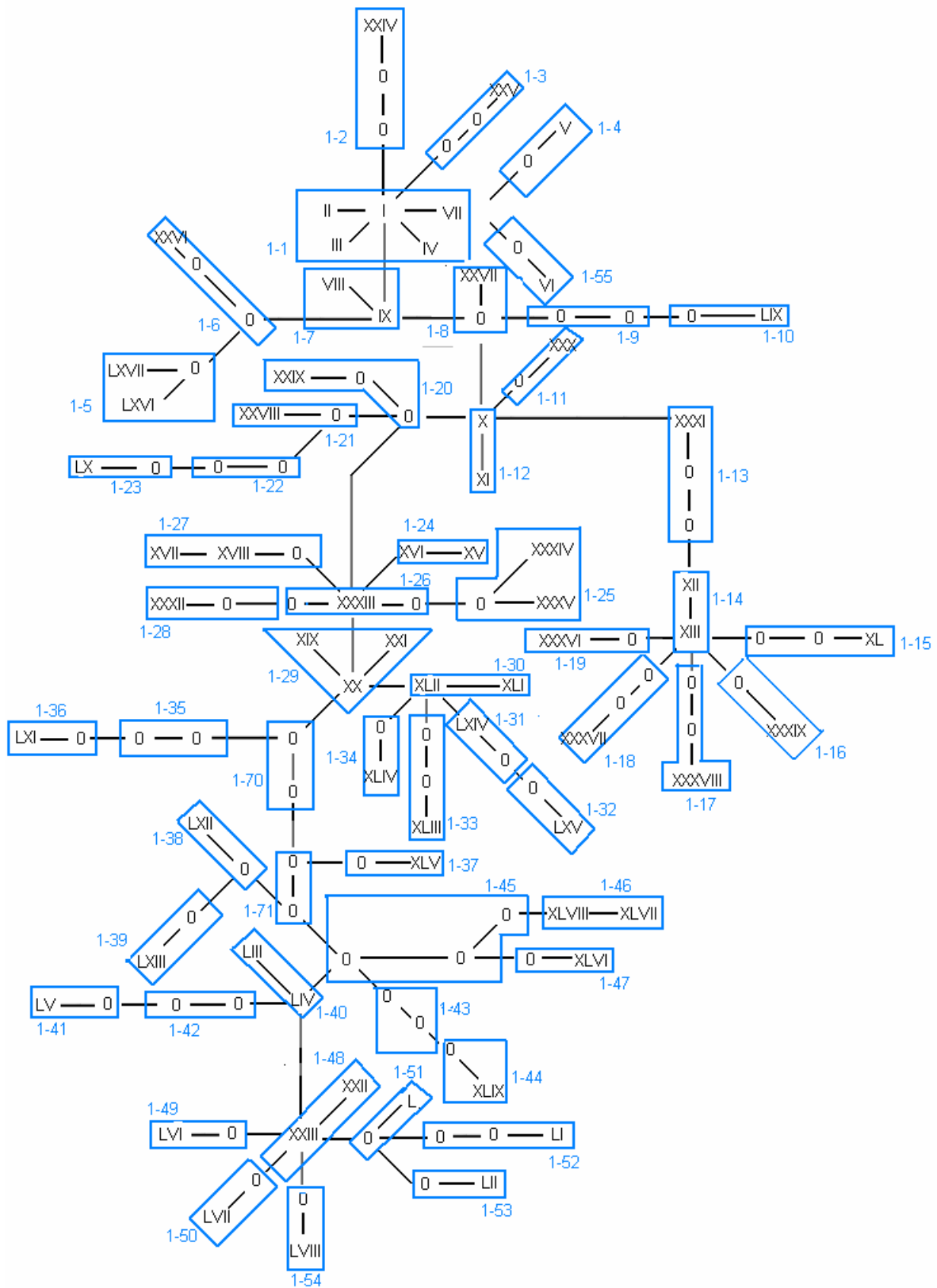


Fig. 16 Nível 1 de aninhamento para a rede mid

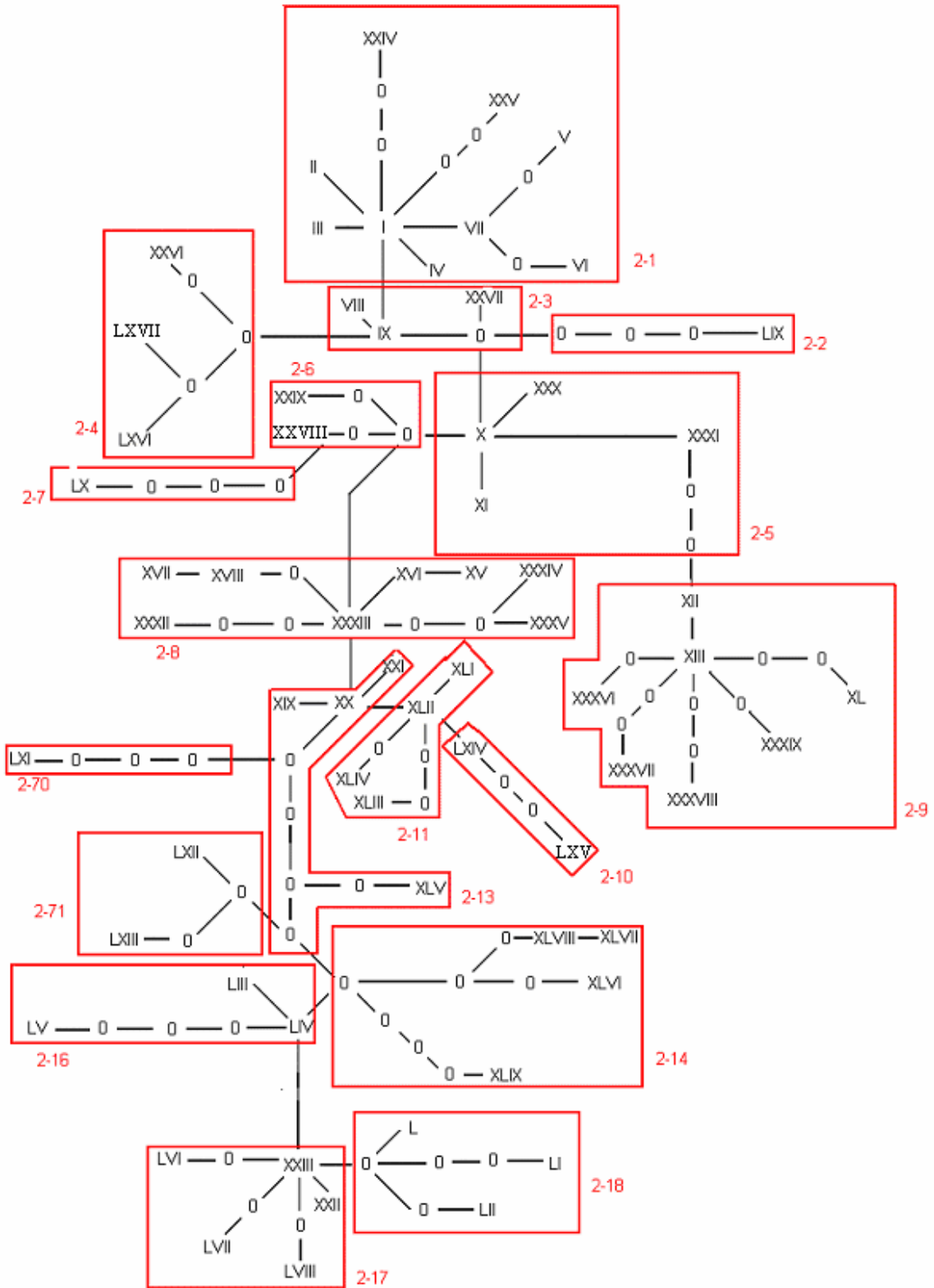


Fig. 17 Nível 2 de aninhamento para a rede mid

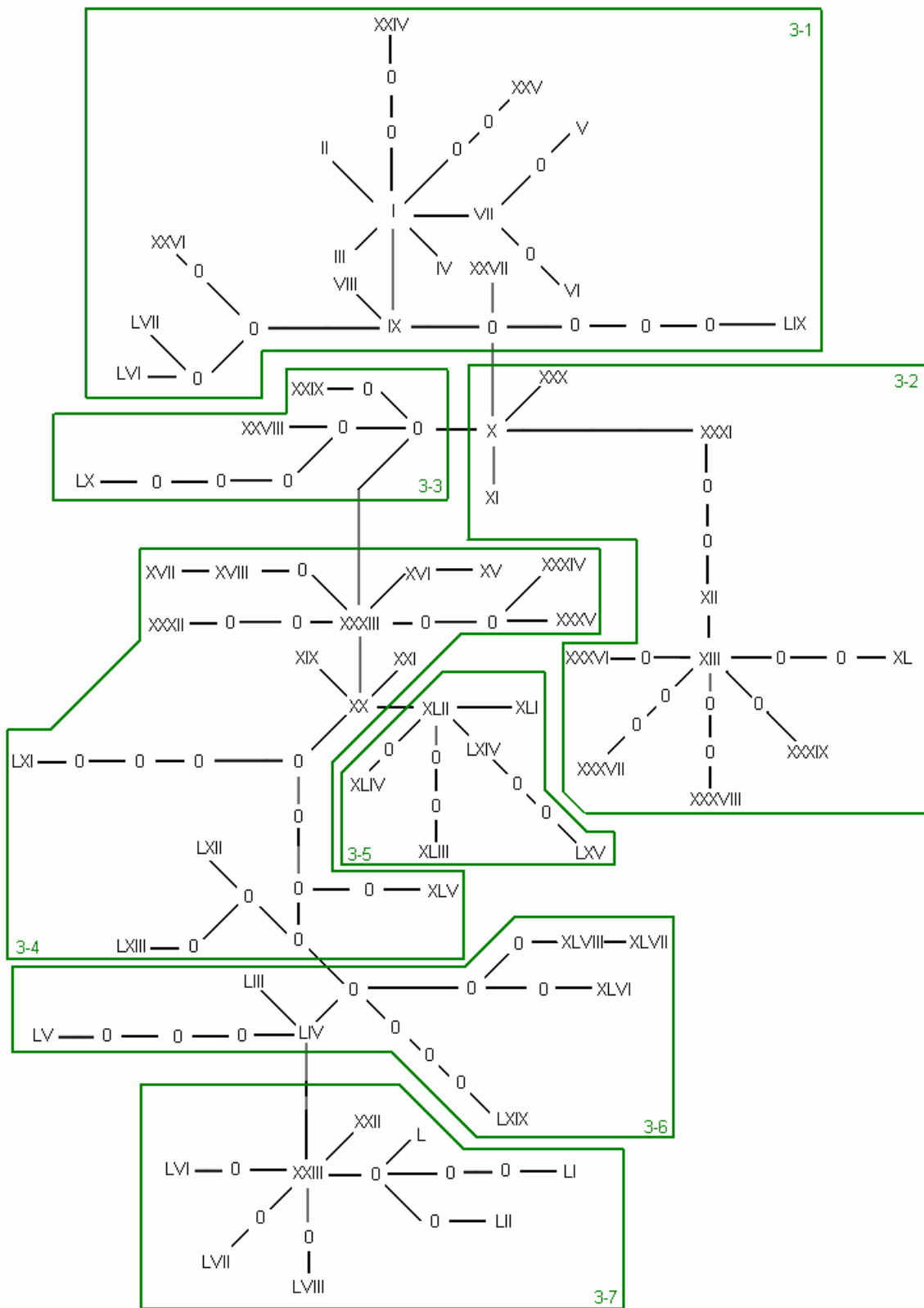


Fig. 18 Nível 3 de aninhamento para a rede mid

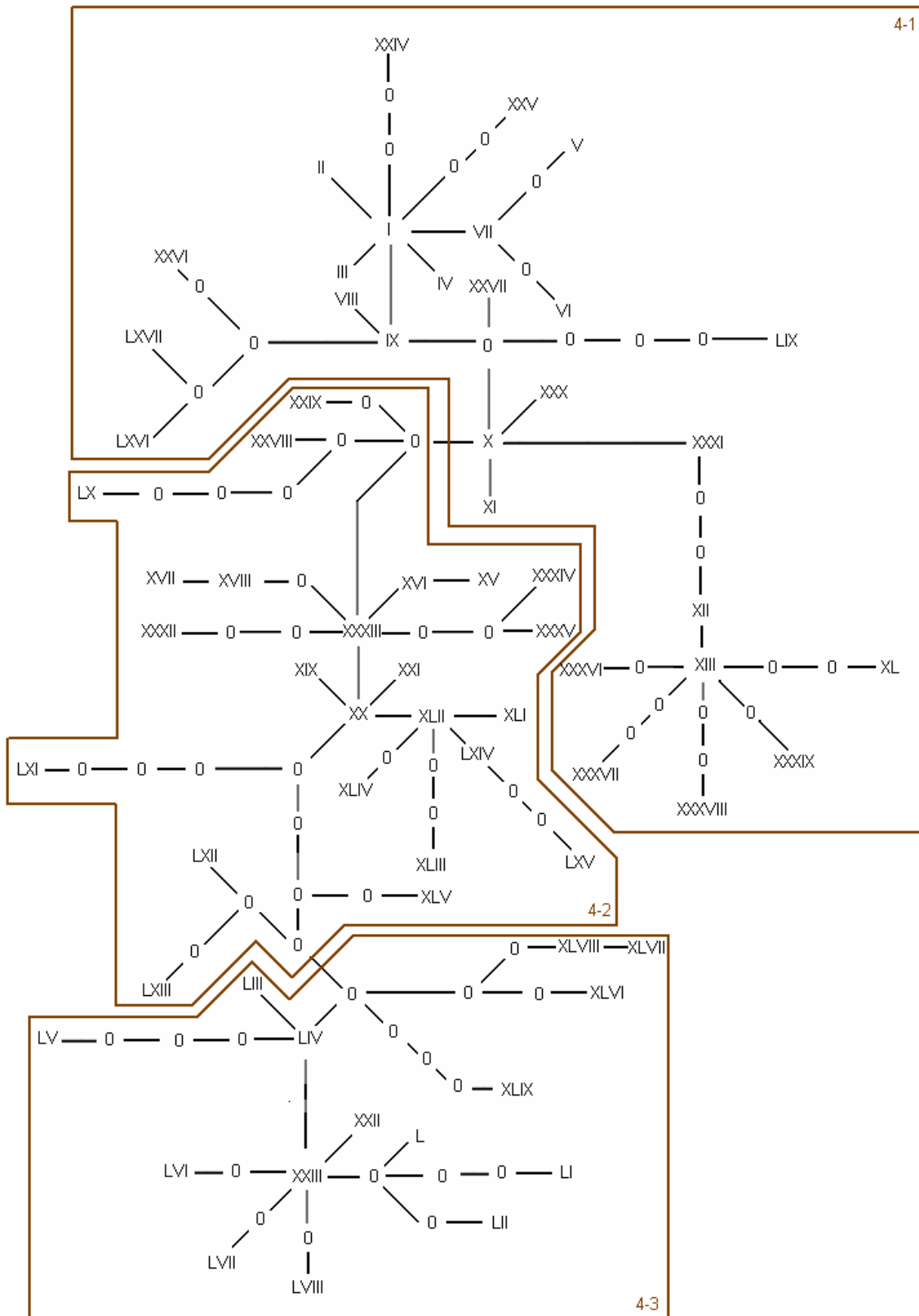


Fig. 19 Nível 4 de aninhamento para a rede mid

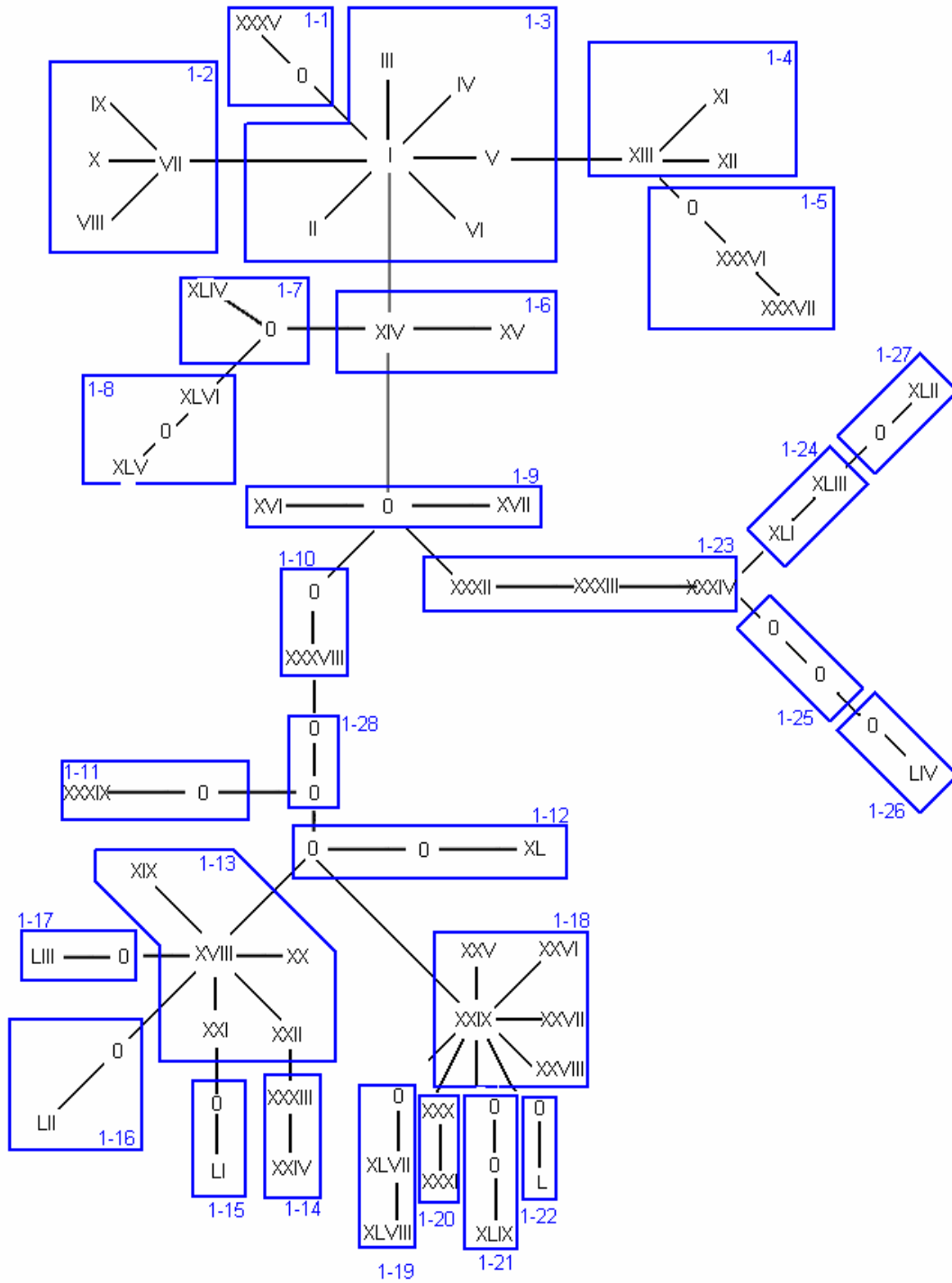


Fig. 20 Nível 1 de aninhamento para a rede 3'

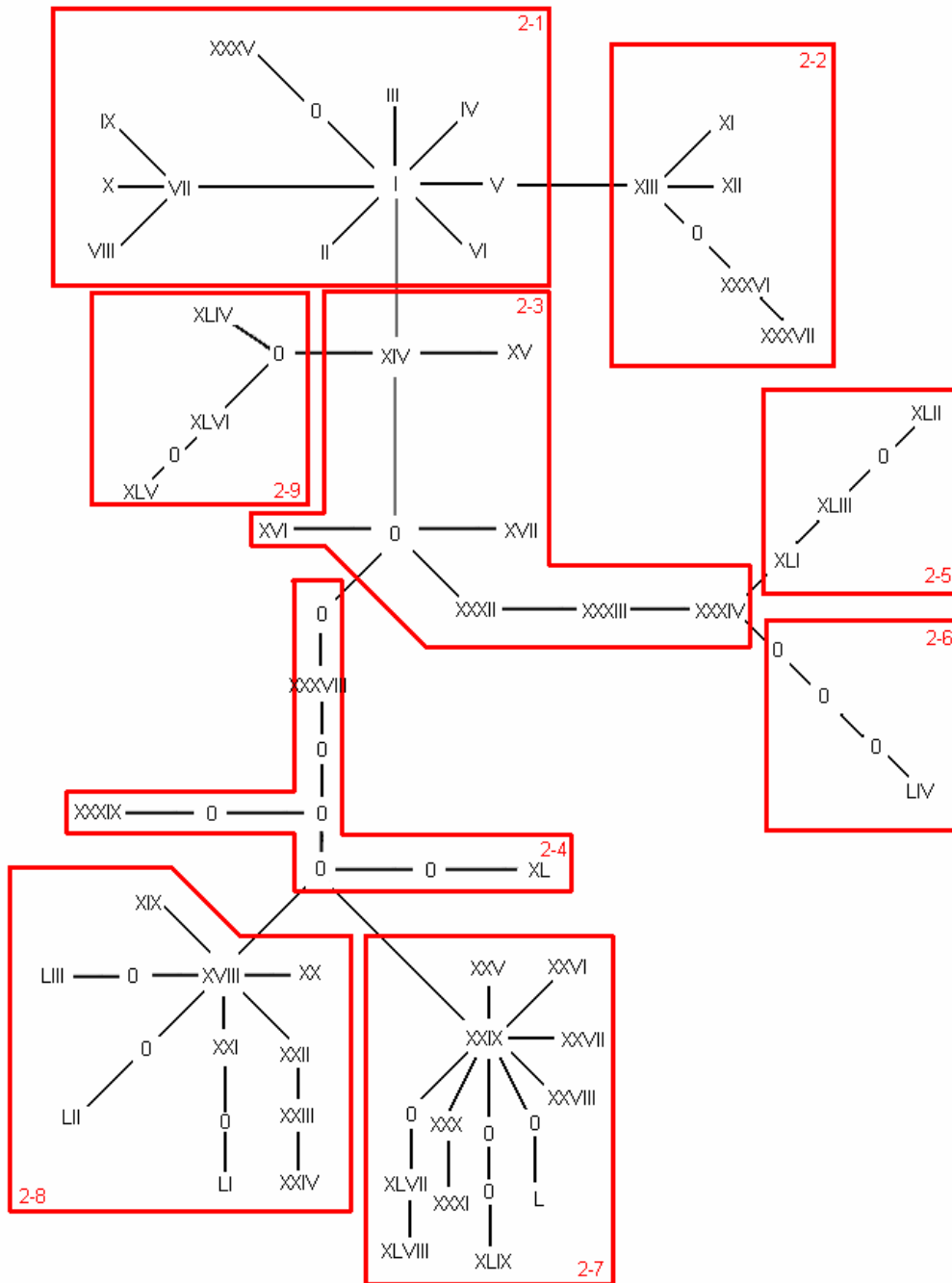


Fig. 21 Nível 2 de aninhamento para a rede 3'

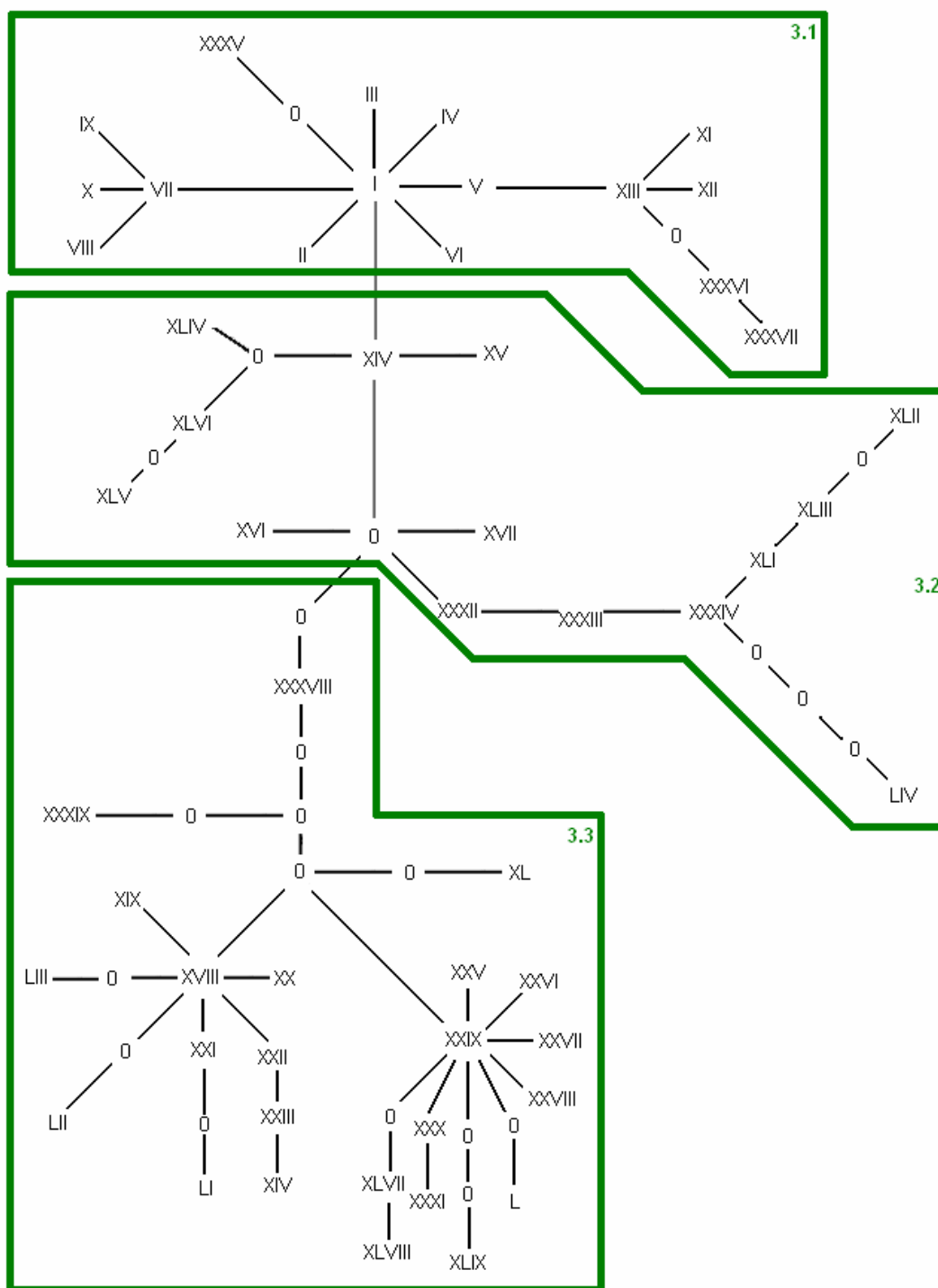


Fig. 22 Nível 3 de aninhamento para a rede 3'

A análise dos clados aninhados foi efetuada no programa Geodis (Posada *et al.* 2000), no qual pudemos comparar a hipótese nula de nenhuma associação entre a estrutura

das redes e a distribuição geográfica dos haplótipos, com uma hipótese alternativa que associa uma relação direta entre o agrupamento dos componentes da rede de acordo com suas localizações geográficas. A análise no Geodis incluiu estimativas de índices de distâncias geográficas (Dc, Dn e I-T) entre haplótipos ou clados que foram utilizados para interpretar processos históricos e contemporâneos, de acordo com a chave de Templeton (2008, atualizada). Os resultados obtidos para cada rede haplotípica mostraram que a hipótese nula não pode ser rejeitada para grande parte do clados, ou seja, a não rejeição da hipótese nula implica a falta de estruturação geográfica nos dados obtidos. Por outro lado, alguns haplótipos apresentaram níveis significantes para os valores de Dc, Dn ou I-T rejeitando assim a hipótese nula de não associação entre o agrupamento das redes haplotípicas e suas distribuições geográficas, mostrando nesse caso associações para as quais foi possível apontar relações filogeográficas.

A tabela 10 representa o resultado completo dos valores de Dc, Dn e I-T para todos os clados, significativos ou não, referentes à região 5'. Os haplótipos e clados marcados com coloração acinzentadas representam os interiores, os símbolos S e L representam valores de significância para significativamente mais restrito e mais disperso do que seria de esperar ao acaso.

Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade
I	0	1023										
II	0	1167	1_1	1031	1031							
III	0	909										
I-T	0	-201				2_1	1100	1158				
XXXII	-	-	1_2	0	938							
XXXIII	-	-	1_3	0	1114							
XXXIV	-	-	1_4	0	690							
XXXV	-	-	1_5	0	1483				3_1	1134 ^L	1327 ^L	
			I-T	1031	-25							
XXXVI	-	-	1_6	0	742							
XXXVII	-	-	1_7	0	183	2_2	295	1088				
XXXVIII	-	-	1_8	0	183							
			I-T	0	279							

acinzentada representa os clados interiores e os símbolos S e L representam que clado apresentou-se significativamente mais restrito e mais disperso, respectivamente, do que seria de esperar ao acaso.

Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn
I	529	741												
II	0	117												
III	0	313	1_1	661	661									
IV	0	117												
VII	0 ^s	911												
I-T	212	660				2_1	716	786						
XXIV	-	-	1_2	0	366									
XXV	-	-	1_3	0	366									
V	-	-	1_4	964	1122									
VI	-	-	1_55	422	852									
			I-T	178	-83									
LIX	-	-	1_10			2_2	0	1185						
LXVI	-	-	1_5	0	16				3_1	772	833			
LXVII	-	-				2_4	22 ^s	427						
XXVI	-	-	1_6	0	33									
			I-T	0	16									
VIII	0	777	1_7	517	465	2_3	372	775						
IX	0	386												
I-T	0	-391												
XXVII	-	-	1_8	0	232							4_1	837	948
						I-T	-170	31						
XXX	-	-	1_11	0	611									
X	0	819	1_12	819	833	2_5	837	836						
XI	0	819												
I-T	0	0							3_2	819	845			
XXXI	-	-	1_13	0	1078									
			I-T	546	303									
XII	0	981	1_14	955 ^L	982 ^L									
XIII	961	949												
XL	-	-	1_15	0	681	2_9	805	807						
XXXIX	-	-	1_16	0	681									
XXXVIII	-	-	1_17	0	486				I-T	46	12			
XXXVII	-	-	1_18	0	486									

XXXVI	-	-	1_19	0	681														
			I-T	955 ^L	379 ^L	I-T		-32	-29										
XXIX	-	-	1_20	0	59	2_6	59	59	3_3	53 ^S	1443 ^L								
XXVIII	-	-	1_21	0	59														
LX	-	-	1_23	-	-	2_7		0	39										
						I-T		59	20										
<hr/>																			
Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn					
<hr/>																			
XV	0	787	1_24	628	767														
XVI	0	522																	
I-T	0	-266																	
XXXIV	-	-	1_25	0	1444	2_8	1101	1094											
XXXV	-	-																	
XXXIII	-	-	1_26	0	1117														
XVII	0	268	1_27	359	1052														
XVIII	0	542							3_4	989	1022								
I-T	0	274																	
XXXII	-	-	1_28	0	1444														
			I-T	-282	-21														
LXI	-	-	1_36	-	-	2_70		0	1663										
LXIII	-	-	1_39	-	-														
LXII	-	-	1_38	-	-	2_71		0	502										
XIX	0	444										4_2	1135 ^L	1076 ^L					
XX	0	444	1_29	590	722	2_13	831	940											
XXI	0	884																	
I-T	0	-220																	
XLV	-	-	1_37	0	1158														
			I-T	590	-436	I-T	1011	153											
LXIV	-	-	1_31	-	-	2_10	0	390	3_5	493 ^S	921								
LXV	-	-	1_32	-	-														
XLI	-	-	1_30	0	320														
XLII	-	-				2_11	466	527											
XLIII	-	-	1_33	0	318														
XLIV	-	-	1_34	0	1198				I-T	340	171								
			I-T	0	-438	I-T	466	138											
XLIX	-	-	1_44	0	317														
XLVII	-	-	1_46	0	195	2_14	272	291	3_6	332	384								
XLVIII	-	-																	
XLVI	-	-	1_47	0	491														
			I-T	0	-23							4_3	419 ^S	771 ^S					

LIII	-	-	1_40	0	269	2_16	360	429				
LIV	-	-										
LV	-	-	1_41	0	543							
			I-T	0	-273							
L	-	-	1_51	0	117							
LI	-	-	1_52	0	234	2_18	156	337	3_7	448	447	
LII	-	-	1_53	0	117							
XXII	0	460	1_48	615	666							
XXIII	0	925										
I-T	0	465				2_17	556	560			I-T	370 ^L 242 ^L
LVI	-	-	1_49	0	150							
LVII	-	-	1_50	0	415							
LVIII	-	-	1_54	0	774				I-T	-116	-63	
			I-T	615	220	I-T	400	223				

Tab. 11 Representação dos valores de Dc, Dn e I-T para todos os clados referentes à região mid

A tabela 12 refere-se à tabela dos resultados completos do Geodis referentes à região 3', onde todos os clados estão representados, significativos ou não. A coloração acinzentada representa os clados interiores, e os símbolos S e L representam valores de significância para significativamente mais restrito e mais disperso do que seria de esperar ao acaso.

Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn	Clade	Dc	Dn
I	433 ^S	578 ^S									
II		0 1100									
III		0 1222									
IV	0 463		1_3		0 1222						
V	0 1456										
VI	0 1097										
I-T	433 ^S	-489 ^S				2_1	793 ^S	880			
XXXV	-	-	1_1	824	875						
VII	220	599									
VIII	0 632		1_2	580	608				3_1	942	990
IX	0 553										
X	0 553										
I-T	220	201	I-T	341	164						

XXIII	0	522	1_14	628	948				
XXIV	0	787							
I-T	0	-266	I-T	548	-6				
XXV	0	558							
XXVI	0	650							
XXVII	0	862	1_18	813	861				
XXVIII	0	756							
XXIX	929	935							
I-T	929	229							
XLVII			1_19	0	906	2_7	831	853	
XLVIII									
XXX	0	1019	1_20	507	743				
XXXI	0	336							
I-T	0	682							
XLIX			1_21	0	561				
L			1_22	0	906				I-T 15 13
			I-T	644	67	I-T	310	280	

Tab. 12 Representação dos valores de Dc, Dn e I-T para todos os clados referentes à região 3'

Uma vez determinado quais clados apresentaram uma estrutura genética espacial significativa, evidenciado pela significância de qualquer um dos três parâmetros de distância (Dc, Dn e I-T), utilizamos a chave de inferência de Templeton (2008) para mostrar os eventos que possam estar relacionados à esse valor de significância. Na tabela 13 estão representados, para as três regiões gênicas, os resultados de inferência dos clados que foram congruentes com seus padrões de distribuição geográfica.

	Clade	Chave do Clado	Inferência
5'	3_4	1,2no; 11(b)yes; 12no	Continuous Range Expansion
	3_3	1no; 2(a)yes, 3no; 4no	Restricted Gene Flow with Isolation by Distance
	4_1	1no; 2(a)yes, 3no; 4no	Restricted Gene Flow with Isolation by Distance
	4_2	1no; 2(a, d)yes; 3no; 4no	Restricted Gene Flow with Isolation by Distance
	C. Final	1no; 2(a, d)yes; 3no; 4no	Restricted Gene Flow with Isolation by Distance
mid	1_1	1,19yes; 20no	Inadequate Geographical Sampling
	2_9	1,19 yes; 20no	Inadequate Geographical Sampling
	3_1	1no; 2(a)yes; 3no; 4no	Restricted Gene Flow with Isolation by Distance

	4_2 C. Final	1no;2(b)yes;3(a)yes;5no;6yes;13 yes 1no; 2(a,c,d)yes; 3no; 4no	Long Distance Colonization Restricted Gene Flow with Isolation by Distance
3'	1_3	1,2no; 11(b)yes; 12no	Continuous Range Expansion
	2_3	1,19yes; 20no	Inadequate Geographical Sampling
	2_4	1,19yes; 20no	Inadequate Geographical Sampling
	3_1	1,2no; 11(b)yes; 12no	Continuous Range Expansion
	3_2	1no; 2(c,d)yes; 3no; 4no	Restricted Gene Flow with Isolation by Distance
	3_3	1no; 2(d)yes; 3no; 4no	Restricted Gene Flow with Isolation by Distance

Tab. 13 Clados que apresentaram valores significativos para as regiões 5', mid e 3'.

Baseado nos resultados de inferência verificamos que dois eventos principais ocorreram na área de distribuição dos indivíduos amostrados. As inferências para os clados 3' 1_3, 3' 3_1 e 5' 3_4 apontaram para a ocorrência de uma expansão contínua da distribuição no sentido Norte-Nordeste em direção à região Centro-Sul do Brasil. Interessantemente, todos estes eventos referem-se à grupos de *A. fraterculus*, indicando possivelmente uma dispersão desta espécie da região mais central e nordeste do Brasil em direção ao sul. Esta evidência pode ser verificada de forma clara quando observamos o clado 1_3 referente à região 3', em que os haplótipos mais recentes (tips) encontram-se mais ao Sul, enquanto os haplótipos mais antigos (interiores) estão representados na porção Norte-Nordeste. Através da metodologia de Takahata *et al.* (2001) podemos inferir um tempo estimado para estes eventos. Ao fazê-lo, isso permite-nos inferir se os eventos similares detectados são geográfica e temporalmente congruentes (o que seria de fato evidência de se tratar de apenas um evento) ou se não há tal concordância. Ao fazer tal teste, verificamos que os eventos de expansão contínua não são estatisticamente distintos, indicando que podem ser todos evidência de um mesmo evento único. A tabela 14 representa os valores referentes a k, theta, T inferido e qui-quadrado referentes à expansão contínua da distribuição, mostrando que tais eventos não são estatisticamente distintos.

range exp	k	theta	T	quiquadrado
3' 1_3	1,667	0,01343	561567,9	
3' 3_1	2,525	0,02219	847089,6	
5' 3_4	2,8	0,02115	673087,4	0,876963

Tab.14 Representação dos valores de K, theta, T e Qui quadrado evidenciando para ocorrência de evento único relacionado à expansão contínua da distribuição para os três clados em questão (P<0,05)

A interpretação de análise cladística aninhada para dados de DNA mitocondrial em *Anastrepha* revelou a presença de dois eventos relacionados à expansão contínua da distribuição no território nacional. O primeiro evento trata-se de uma expansão da região central do Brasil à Bahia e Belém. Uma amostragem mais detalhada é necessária para determinar se tal evento se dá em direção à ocupação desse trecho Norte-Nordeste ou acontece no sentido inverso (Santos, 1994; Selivon, 1996). Um outro evento de expansão contínua da distribuição refere-se à expansão da espécie *A. fraterculus* para a região de Belo Horizonte, que tem sido observada no interior de São Paulo e no Sul do país (Santos, 1994; Selivon, 1996). Nesses dois casos observamos uma certa diferença comparada à análise cladística do gene *yolk*, onde os dados apontam para expansão em direção ao centro-sul do Brasil. Como o estudo realizado com DNA mitocondrial não estimou datas de divergência, não podemos inferir se estes eventos são antagônicos ou simplesmente reflexo de taxas evolutivas e padrões de divergência diferentes para marcadores genéticos distintos do genoma.

Um segundo evento parece apresentar maior influência sobre o padrão de distribuição espacial dos haplótipos do gene *yolk* para as populações de *Anastrepha* ao longo do território brasileiro. As inferências para oito clados (3' - 3_2, 3_3; mid- 3_1, clado final e 5' - 3_3, 4_1, 4_2 e clado final) apontaram para a ocorrência de fluxo gênico restrito com isolamento por distância. Através da verificação dos valores significativos de Dc, Dn ou I-T para a inferência de fluxo gênico restrito com isolamento por distância, observamos tal resultado apenas para os clados superiores. Através da metodologia de Takahata *et al.* (2001), verificamos que os eventos de restrição ao fluxo gênico são todos razoavelmente recentes, datando da ordem de 1 milhão de gerações atrás, se considerarmos uma taxa de mutação de 1×10^{-8} /nucleotídeo/geração e considerando-se que devem haver entre 3-5 gerações de *Anastrepha* por ano, isto permite estimar tais eventos em torno de 250.000 anos atrás. O fato de termos encontrado estimativas temporais congruentes para diversos eventos de isolamento por distância é surpreendente, pois eventos recorrentes como estes têm o potencial de continuar a ocorrer mesmo nos dias de hoje, e não temos evidência de que isto

esteja acontecendo. Pode ser que não tenhamos detectado valores significativos em clados inferiores (mais recentes) por restrições amostrais, uma vez que nossa amostragem é de fato restrita, mas é também possível que nestes níveis, ou seja, em populações mais próximas, o fluxo gênico possa estar ocorrendo com frequência suficiente para promover a homogeneização genética das populações. Por sua vez, em clados superiores o fluxo gênico torna-se restrito, porque o nível de fluxo gênico entre as populações mais distantes não é suficiente para impedir um certo grau de diferenciação ao longo do tempo. A Tabela 15 representa os valores referentes a k, theta e T inferido referentes ao fluxo gênico restrito com isolamento por distância, mostrando que tais eventos não são estatisticamente distintos.

Clado	k	theta	T
5' 3_3	3,826	0,03367	911597,9
3' 3_2	3,571	0,02508	1126570
3' 3_3	3,895	0,04436	1319102
mid final	7,605	0,06183	1218240
5' 4_1	4,363	0,05618	1060720
5' 4_2	4,692	0,04858	1113526
5' clade f	4,437	0,08011	1098460

Tab. 15 Representação dos valores de K, theta e T. Teste de Takahata *et al.* (2002) de congruência temporal destes valores indica valor de $\chi^2=0.097$, n.s. Indicando não necessidade de mais de um evento de fluxo gênico restrito com isolamento por distância para os clados em questão ($P<0,05$)

Estimativas de fluxo gênico entre estas populações feitas através do software Migrate encontraram valores bastante altos, superiores a 60 e chegando a 200 quando estimamos fluxo gênico entre espécies, e superiores a 40 chegando a 300 quando artificialmente separamos as espécies em norte e sul. Tais valores são muito altos e de fato podem refletir aspectos distintos.

Podemos distinguir um certo agrupamento entre as espécies de *Anastrepha*, particularmente nos cladogramas das regiões 5' e mid. O clado 3_3 (5') apresenta um subclado (2_14) com haplótipos antigos (interior) com presença de grande parte de indivíduos *A. fraterculus*, um outro subclado (2_6) pertencente à esse mesmo clado infere a presença de indivíduos *A. obliqua* e *A. sororcula*, indicando um possível papel da redução do fluxo gênico promovendo os agrupamentos populacionais. Em grande parte dos clados que inferem fluxo gênico restrito com isolamento por distância temos um padrão parecido

ao clado 3_3, de modo que, invariavelmente, *A. fraterculus* se apresenta nos ramos mais internos dos cladogramas. O padrão de isolamento por distância foi também encontrado para dados de DNA mitocondrial de clados dispersos por todo o Brasil (de Brito e Matioli, dados não publicados).

As figuras abaixo representam esquematicamente os dois principais eventos que podem ter ocorrido na área de distribuição dos indivíduos amostrados. A figura 23a representa a direção da expansão contínua da distribuição, enquanto a figura 23b mostra a área total de ocorrência de isolamento por distância através do fluxo gênico restrito.

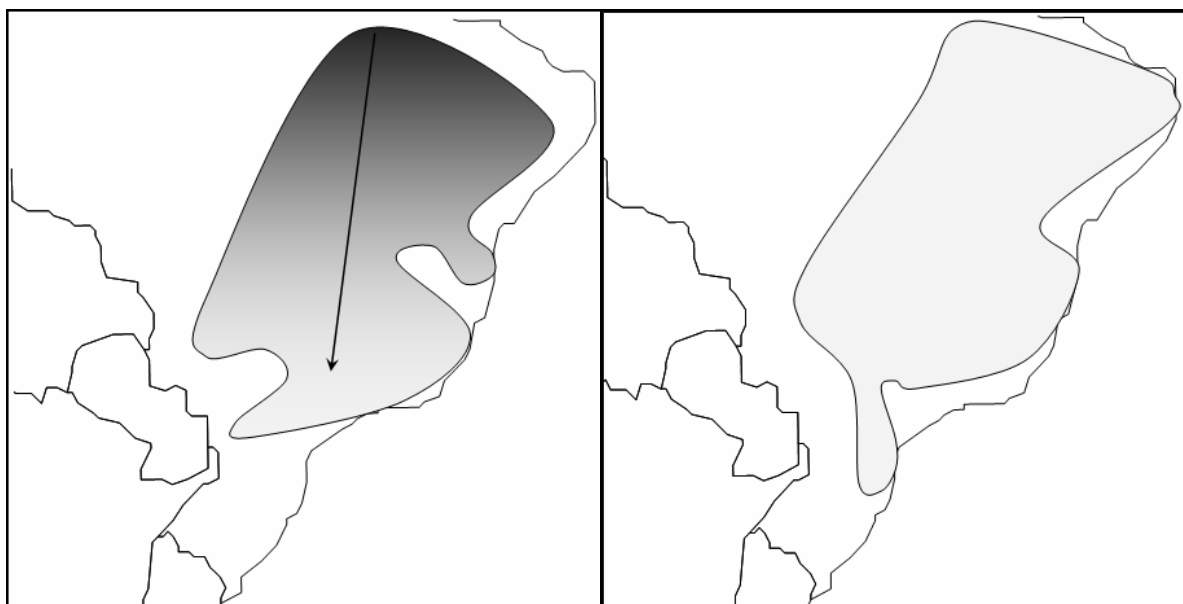


Fig.23 a) Expansão contínua da distribuição

b) Fluxo gênico Restrito com IBD

Deve ser ressaltado também que um padrão de restrição ao fluxo gênico por isolamento por distância não é adequado para estimativas de fluxo gênico como medimos, uma vez que o fluxo gênico entre populações próximas é maior do que entre populações mais distantes. Ao fazermos uma separação artificial norte-sul apenas para efeitos de estimativa geral deste fluxo, corremos o risco que nossas estimativas sejam superestimadas, como imaginamos que possam estar, por considerar este fluxo existente entre populações próximas e generalizá-lo para populações mais distantes. A capacidade de detecção de sinal da NCPA depende tanto do tipo de sinal quanto do tamanho amostral. Eventos mais radicais podem ser observados com tamanhos amostrais menores, enquanto que outros

eventos podem requerer um tamanho amostral significativo para serem observados. Como a topologia da árvore pode influenciar na análise resultante, e esta topologia depende de eventos estocásticos (o processo de mutação marcando os ramos) para sua existência, existe um componente estocástico na árvore do qual não podemos escapar. No entanto, quando consideramos esta análise, utilizamos a árvore como sendo um componente conhecido que define os contrastes que devemos fazer. Por desconsiderar o efeito estocástico do cladograma, diversas críticas têm sido feitas ao NCPA (Petit, 2008., Knowles & Maddison., 2002 e Knowles, 2008). Tais críticas ressaltam que uma análise filogeográfica feita apenas com um marcador molecular é invariavelmente incapaz de separar o componente estocástico do processo real de diferenciação populacional. Isto fez com que modelos probabilísticos tenham sido criados para tentar compensar esta estocasticidade (Carstens e Richards, 2007; Richards, *et al.*, 2007, Knowles e Maddison, 2008, Knowles, 2008). Tais modelos são interessantes por permitirem uma avaliação desta estocasticidade, por outro lado são limitados pois apenas permitem-nos comparar modelos e testar probabilidades de eventos que consideramos prováveis. Isto é uma grande limitação pois na maioria das vezes não sabemos o que aconteceu às populações, e portanto não testamos todas as possibilidades.

A necessidade de incorporação da estocasticidade da árvore ao processo de inferência filogeográfica é uma crítica válida que ressalta que os modelos mais adequados devem considerar diversas fontes de informação para a inferência populacional. Por este motivo, Templeton (2009) sugeriu que tal análise deva ser feita com diversos marcadores moleculares ao mesmo tempo. Como marcadores distintos nos mesmos organismos estarão passando por eventos demográficos e evolutivos similares, embora filtrados pelo processo estocástico de forma distinta, tal estratégia permitiria a correção do processo estocástico na inferência populacional. Para que tal estratégia seja eficiente, deve-se considerar que apenas eventos que tenham sido observados em mais de um marcador ao mesmo tempo devam ser considerados como prováveis. Tal proposta tem a vantagem de continuar permitindo a capacidade exploratória do NCPA, uma de suas grandes vantagens, e, ao mesmo tempo, ser capaz de considerar a estocasticidade das genealogias produzidas.

Na análise dos dados aqui obtidos, embora tenhamos investigado apenas um gene, temos de fato informações genéticas como que de três marcadores genéticos distintos. A

recombinação é um processo que, se não considerado na análise, é capaz de destruir as relações históricas entre os organismos. Como detectamos em nossa análise três regiões que apresentam significativo grau de recombinação entre elas, podemos considerar tais regiões quase como marcadores independentes, uma vez que para todos os indivíduos que sofreram recombinação tais marcadores serão de fato independentes. Dessa forma, consideramos como estatisticamente significantes apenas os eventos que foram detectados em mais de uma região gênica, algumas vezes em mais de um clado na mesma região.

Um bom entendimento àcerca dos eventos que têm afetado a evolução e distribuição das espécies do grupo *fraterculus* ainda dependerá da análise de diversos outros marcadores genéticos para que sejamos capazes de discernir eventos populacionais de processos seletivos específicos. Contudo, mediante a análise das seqüências do gene *yolk*, obtivemos um melhor conhecimento no que se refere aos níveis de variação desse marcador. Além disso, foi possível inferir o possível padrão de distribuição atual das espécies do grupo *fraterculus* no território nacional. Assim, uma primeira indicação nos foi dada a respeito dos processos que possam estar determinando a verdadeira história evolutiva dessas espécies.

6 - Conclusões

Análises das seqüências do gene *yolk* apontaram para ocorrência de recombinação, sendo que, para minimizar os efeitos deste evento, particionamos nossos dados em três partes, sendo que estas foram analisadas em separadamente.

Assim como outros genes estudados em nosso laboratório, o gene *yolk* mostrou-se altamente polimórfico nas três regiões gênicas analisadas para as três espécies estudadas nesse trabalho, tanto a nível intra como a nível interpopulacional.

Através da metodologia de Nested Clade Pylogeography Analysis (NCPA) verificamos que a maioria dos clados analisados não apresentou associação entre a configuração da rede haplotípica com a distribuição geográfica. No entanto, alguns mostraram tal associação, de modo que dois eventos principais foram inferidos para explicar o padrão de distribuição das populações de *Anastrepha* no território nacional. Um dos eventos refere-se à expansão contínua da distribuição do sentido Norte- Nordeste em direção à região Centro- Sul do Brasil. Um segundo evento refere-se ao fluxo gênico restrito com isolamento por distância. Este segundo evento foi detectado basicamente em clados superiores, sendo que o fluxo gênico restrito pode não ter sido detectado em clados inferiores devido à restrição amostral, ou ainda porque níveis inferiores considera populações mais próximas, onde o fluxo gênico possa estar ocorrendo com frequência suficiente para promover a homogeneização genética das populações.

Através da metodologia de Takahata *et al.*, (2001) verificamos que tanto os eventos de expansão contínua da distribuição como aqueles referentes ao fluxo gênico restrito não são estatisticamente distintos, indicando que ambos podem ser evidência de um mesmo evento. Para o caso de fluxo gênico restrito com isolamento por distância não encontramos sinais de ocorrências recentes, sendo tal resultado surpreendente por se tratar de um evento recorrente, que inclusive pode estar acontecendo nos dias atuais.

7. Referências Bibliográficas

ALVAREZ, M.; RUIZ, M. F.; SANCHEZ, L. Effect of the gene *doublesex* of *Anastrepha* on the somatic sexual development of *Drosophila*. **PLoS ONE**, v. 4, n. 4, p. 5141, 2009.

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

AMDAM, G.V. et al. Hormonal control of the yolk precursor vitellogenin regulates immune function and longevity in honeybees. **Exp. Gerontol.**, v. 39, n. 5, p. 767-773, 2004.

AN, W.; WENSINK, P. C. Integrating sex- and tissue-specific regulation within a single *Drosophila* enhancer. **Genes & Dev.**, v. 9, n. 5, p. 256-266, 1995.

ANDERSSON, M. **Sexual selection**. New York: Princeton University, 1994.

AVISE, J. C. Identification and interpretation of mitochondrial DNA stocks in marine species. In: PROCEEDINGS OF THE STOCK IDENTIFICATION WORKSHOP, 9., 1987, Panamá. Proceedings...Panamá: National Oceanographic and Atmospheric Administration, Panama City, 1987. p. 105-136.

AVISE, J. C. **Molecular markers, natural history and evolution**. New York: Chapman & Hall, 1994.

AVISE, J. C. **Phylogeography: The History and Formation of Species**. Cambridge, MA: Harvard University Press, 2000.

AVISE, J. C. Molecular markers in conservation genetics. (2004). In: Molecular markers, natural history and evolution. 2a ed. Sunderland: Sinauer Associates Inc. p. 541, 2004.

AVISE, J. C. Phylogeography: retrospect and prospect. **Journ. of Biog.**, v. 36, n. 5, p. 3-15, 2009.

BEBAS, P. et al. Yolk protein is expressed in the insect testis and interacts with sperm. **BMC Devel. Biology.**, v. 8, n. 64, p. 1-12, 2008.

BEERLI, P.; FELSESTEIN, J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. **Proc. Nat. Acad. Sci. USA.**, v. 98, n. 8, p. 4563-4568, 2001.

BOAKE, C. R. Flying apart: mating behavior and speciation. **BioScience**, v. 50, n. 6, p. 501-508, 2000.

BOWNES, M. The role of 20-hydroxyecdysone in yolk polypeptide synthesis by male and female fat bodies of *Drosophila melanogaster*. **J. Insect. Physiol.**, v. 28, p.317-328, 1982.

BOWNES, M. Hormonal and genetic regulation of vitellogenesis in *Drosophila* (yolk protein). **Quart. Rev. Biol.**, v. 57, n. 3, p. 247-274, 1982.

BOWNES, M.; SCOTT, A.; SHIRRAS, A. Dietary components modulate yolk protein gene transcription in *Drosophila melanogaster*. **Development**, v. 103, n.1, p. 119 -128, 1988.

BOWNES, M. The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster*. **BioEssays**, v. 16, n. 10, p. 745-752, 1994.

BRAWAND, D.; WAHLI, W.; KAESSMANN, H. Loss of Egg Yolk Genes in Mammals and the Origin of Lactation and Placentation. **PLoS Biol.**, v. 6, n. 3, p. 63, 2008.

BURTIS, K.C. et al. The doublesex proteins of *Drosophila melanogaster* bind directly to a sex-specific yolk protein gene enhancer. **The Embo Journal**, v. 10, n. 9, p. 2577-2582, 1991.

BUTLIN, R.; RITCHIE, M.G. Searching for speciation genes. **Nature**, v. 412, p. 31-33, 2001.

CAICEDO, A. L.; SCHAAL, B. A. Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. **Mol. Ecol.**, v. 13, n. 7, p. 1871-1882, 2004.

CARSON, H. L.; KANESHIRO, K. Y. *Drosophila* of Hawaii: systematics and ecological

genetics. **Ann. Rev. Ecol. System.**, v. 7, p. 311-346, 1976.

CARSTENS, B.C.; RICHARDS, C.L. Integrating coalescent and ecological niche modeling in comparative phylogeography. **Evolution**, v. 61, n. 6, p. 1439-1454, 2007.

CARVALHO, P.G. Efeitos do nitrogênio no metabolismo de frutanos em *Vernonia herbacea* (Vell.) Rusby. 2005. Dissertação (Mestrado em Botânica) - Escola Superior de Agronomia Luis de Queiróz, Piracicaba.

CHEN, J. S.; SAPPINGTON, T. W.; RAIKHEL, A .S. Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry. **J. Mol. Evol.**, v. 44, n. 4, p. 440-451, 1997

CHEN, S.L. et al. Female-specific *doublesex* dsRNA interrupts yolk protein gene expression and reproductive ability in oriental fruit fly, *Bactrocera dorsalis* (Hendel). **Insec. Biochem. and Molec. Biology**, v. 38, n. 2, p. 155-165, 2008

CHO, S.; WENSINCK, P. C. DNA binding by the male and female *doublesex* proteins of *Drosophila melanogaster*. **J. Biol. Chem.**, v. 272, n. 6, p. 3185-3189, 1997

CHOI, J. S. et al. Improved cycle sequencing of GC- rich DNA template. **Experimental and Molec. Medicine**, v. 31, n.1, p. 20-24, 1999

CLEMENTE, M.; POSADA, D.; CRANDALL, K. A. TCS: a computer program to estimate gene genealogies. **Mol. Ecol.**, v. 9, n.10, p. 1657-1660, 2000

CLINE, J.; BRAMAN, J. C.; HOGREFE, H. H. PCR fidelity of *Pfu* polymerase and other thermostable DNA polymerases. **Nucleic. Acids. Res.**, v. 24, n. 18, p. 3546-3551, 1996

COYNE, J. A; ORR, H. A. Patterns of speciation in *Drosophila* revisited. **Evolution**, v. 51, n.1, p. 295-303, 1997

COYNE, J.A.; ORR, H.A. **Speciation**. Sunderland (MA): Sinauer Associates, 2004

CRANDALL, K.A.; TEMPLETON, A.R. Empirical tests of some predictions

from coalescent theory with applications to intraspecific phylogeny reconstruction. **Genetics**. v. 134, n. 3, p. 959–969, 1993

CREER, S.A. et al. Multiple causation of phylogeographical pattern as revealed by nested clade analysis of the bamboo viper (*Trimeresurus stejnegeri*) within Taiwan. **Mol. Ecol.**, v. 10, n.8, p. 1967–1981, 2001.

De BRITO, R.A.; MANFRIN, M.H.; SENE, F.M. Nested cladistic analysis of Brazilian populations of *Drosophila serido*. **Mol. Phylogenet. Evol.** v. 22, n.1, p. 131-143, 2002

ERDMAN, S. E.; CHEN, H.J.; BURTIS, K. C. Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila* Doublesex proteins. **Genetics**. v. 144, n. 4, p. 1639-1652, 1996

EVANOVICH, E. Evolução molecular. Projeto Evoluindo - Biociência.org. Disponível em: <<http://www.evoluindo.biociencia.org>>. Acesso em 05 maio. 2004

FINN, R.N. The physiology and toxicology of salmonid eggs and larvae in relation to water quality criteria. **Aquat. Toxicol.** v. 81, n. 4, p. 337–354, 2007

FU, Y. X.; LI, W. H. Statistical tests of neutrality of mutations. **Genetics**, v. 133, n. 3, p. 693-709, 1993

FU, Y.X. Estimating the age of the common ancestor of a DNA sample using the number of segregating sites. **Genetics**, v.144, v. 2, p. 829-838, 1996

FU, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, v. 147, n. 2, p. 915-925, 1997

GARABEDIAN, M.J.; SHEPHERD, B.M.; WENSINK, P.C. A tissue-specific transcription enhancer from the *Drosophila* yolk protein 1 gene. **Cell**, v.45, n.6, p. 859–867, 1986

GARABEDIAN, M. J. et al. The nucleotide sequence of the gene coding for *Drosophila melanogaster* yolk protein 3. **Gene (Amst.)** v. 55, n.1, p. 1-8, 1987

GRAHAN, P.; PENN, J. K. M.; SCHEDLY, P. Masters change, slaves remain. **Bioassays**

v. 25, n.1, p.1-4, 2003

GRAY, D.A.; W.H. Cade. Sexual selection and speciation in field crickets. **Evolution**, v. 97, n. 26, p. 14449-14454, 2000

HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl. Acids. Symp. Ser.**, v. 41, p. 95-98, 1999

HANDLER, A. M. Developmental Regulation of Yolk Protein Gene Expression in *Anastrepha suspensa*. **Insect Biochem. and Physiol.**, v. 36, n.1, p. 25-35, 1997

HEED, W.B. Host plant specificity and speciation in Hawaiian *Drosophila*. **Taxon**, v. 20, n.1, p. 115-121, 1971

HENS, K. et al. Cloning and expression of the yolk protein of the tsetse fly *Glossina morsitans morsitans*. **Insect Biochem. Mol. Biol.** v. 34, n.12, p. 1281-1287, 2004

HOVEMANN, B.; GALLER, R. Vitellogenin in *Drosophila*: a comparison of *yk1* and *yk2* genes and their transcription products. **Nucleic Acids Res.** v.10, n.7, p. 2261-2274, 1982

KENDRA, P. E. et al. Assessment of female reproductive status in *Anastrepha suspensa* (Diptera: Tephritidae). **Fla. Entomol.** v. 89, n. 2, p.144–151, 2006

KIMURA, M. Evolutionary rate at the molecular level. **Nature**, v. 217, n. 5129, p. 624-626, 1968

KNOWLES, L.L.; MADDISON, W.P. Statistical phylogeography. **Mol. Ecol.**, v. 11, n.12, p. 2623-2635, 2002

KNOWLES, L.L. The burgeoning field of statistical phylogeography. **J. Evol. Biol.**, v. 17, n.1, p. 1-10, 2003

KNOWLES, L.L. Why does a method that fails continue to be used? **Evolution**, v. 62, n. 11, p. 2713-2717, 2008

KORNFIELD, V.; RICHER, C.; WARMAN, J. Biochemical and cytological differentiation among Cichlid fish of the Sea of Galilee. **Evolution**, v. 33, n.1, p. 1-14, 1979

LEBO, M.S. et al. Somatic, germline and sex hierarchy regulated gene expression during *Drosophila* metamorphosis. **BMC Genomics**. v. 10, p. 80, 2009

LENK, P. et al. Mitochondrial phylogeography of the european pond turtle, *Emys orbicularis* (Linnaeus 1758). **Molec. Ecol.**, v. 8, n. 11, p. 1911-1922, 1999

LIS, J.T.; SCHLEIF, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. **Nuc. Acids Research**. v. 2, n.3, p. 383-390, 1975

LOGAN, S.K.; GARABEDIAN, M.J.; WENSINK, P.C. DNA regions that regulate the ovarian transcriptional specificity of *Drosophila* yolk protein genes. **Genes Dev**. v. 3, p. 1453-1461, 1989

MALAVASI, A. Estudo de duas espécies crípticas do gênero *Anastrepha*. 1984 Tese – Instituto de Biociências do Departamento de Biologia, Universidade de São Paulo, São Paulo

MALAVASI, A.; ZUCCHI, R.A.; SUGAYAMA, R. L. Biogeografia, *In*: A. Malavasi; R. A. Zucchi (Ed.). Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado. Ribeirão Preto: Holos Editora, p. 93-98

MASTA, S.; W.P. MADDISON. Sexual selection driving diversification in jumping spiders. **Proceed. of the Nation. Acad. of Sciences of the USA**, v. 99, n. 7, p. 4442-4447, 2002

MAYR, E. **Animal species and evolution**. Cambridge, Massachusetts: Belknap Press of Harvard University Press, 1963

MOORE, W.S. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. **Evolution**, v. 49, n. 4, p. 718-726, 1995

MORGANTE, J.S.; MALAVASI, A.; BUSH, G.L. Biochemical systematics and evolutionary relationships of Neotropical *Anastrepha*. **Ann. Entomol. Soc. Am.**, v. 73, n. 6, p. 622–630, 1980

NELSON, C. M. et al. The gene vitellogenin has multiple coordinating effects on social organization. **PLoS Biol.**, v. 5, n. 3, p. 62, 2007

NORRBON, A. L.; ZUCCHI, R.A.; HERNANDEZ-ORTIS, V. Phylogeny of the genera *Anastrepha* and *Toxotrypana* (Trypetinae: Toxotripanini) based on morphology, p. 299-342. In: A. L. NORRBOM; M. ALUJA (eds.). Fruit flies (Tephritidae): phylogeny, 2004

PAETKAU, D.; SHIELDS, G.F; STROBECK, C. Gene flow between insular, coastal and interior populations of brown bears in Alaska. **Molec. Ecol.**, v. 7, p. 1283-1292, 1998

PANHUIS, T.M. et al. Sexual selection and speciation. **Trends Ecol. Evol.** v. 16, n.7, p.364-71, 2001

POSADA, D.; CRANDALL, K.A.; TEMPLETON, A. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. **Molec. Ecology** v. 9, p. 487- 488, 2000

RAIKHEL, A. S.; DHADIALLA, T. S. Accumulation of yolk proteins in insect oocytes. **Ann. Rev. Ent.** v. 37, p. 217 251, 1992

RAND, D.M.; DORFSMAN, M.; KANN, L.M. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. **Genetics** v. 138, p. 741–756, 1994

RAUSCHENBACH, I. Y. et al. Dopamine and octopamine regulate 20-hydroxyecdysone level in vivo in *Drosophila*. **J. Insect Physiol.**, v. 65, n. 2, p. 95-102, 2007

RICHARD, D.S. et al. Ecdysteroids regulate yolk protein uptake by *Drosophila melanogaster* Oocytes. **J. Insect Physiol.**, v. 44, n. 7, p.637-644, 1998

RICHARD, D.S.; JONES, J.M.; BARBARITO, M.R. Vitellogenesis in diapausing and mutant *Drosophila melanogaster*: Further evidence for the relative roles of ecdysteroids and juvenile hormones. **J. Insect Physiol.**, v. 47, n. 8, p. 905–913, 2001

RICHARDS, C.L.; CARSTENS, B.C.; KNOWLES, L. Distribution modeling and

statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. **Journ. of Biogeog.**, v.34, p. 1833-1845, 2007

RINA, M.; SAVAKIS, C. A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitidis capitata*: sequence and structural conservation in Dipteran yolk proteins and their genes. **Genetics**, v. 127, p. 769-780, 1991

ROGERS, A.R.; HARPENDING, H. Population growth makes waves in the distribution of pairwise genetic differences. **Mol. Biol. Evol.** v. 9, n. 3, p. 552–569, 1992.

ROZAS, J.; SANCHES-DEL BARRIO, J.C.; MESSENGUER, X. DnaSP, DNA polymorphism analyses by the coalescent and other methods. **Bioinformatics**, v. 19, n. 18, p. 2496-2497, 2003

SANTOS, P. Variabilidade do DNA mitocondrial em populações de três espécies de *Anastrepha* (Diptera: Tephritidae). 1994 Tese (Doutorado em Genética) – Instituto de Biociências do Departamento de Biologia, Universidade de São Paulo, São Paulo

SANTOS, P.; URAMOTO, K.; MATIOLI, S.R. Experimental hybridization among *Anastrepha* species (Diptera: Tephritidae): production and morphological characterization of F1 hybrids. **Ann. Ent. Soc. Am.**, v. 94, n. 5, p. 717-725, 2001

SAPPINGTON, T.W. The major yolk proteins of higher diptera are homologs of a class of minor yolk proteins in lepidoptera. **Journal of Molec. Evol.**, v. 55, n. 4, p. 470-475, 2002

SAUNDERS, R. D. C.; BOWNES, M. Sequence analysis of a yolk protein secretion mutant in *Drosophila melanogaster*. **Mol. & Gen. Genet.**, v. 205, p. 557-560, 1986

SCRIBNER, K.T.; AVISE, J.C. Cytonuclear genetics of experimental fish hybrid zones inside Biosphere 2. **Proc. Natl. Acad. Sci., USA** v. 91, n. 11, p. 5066–5069, 1994

SELIVON, D. Estudo sobre a diferenciação populacional em *Anastrepha fraterculus* (Wiedemann) (Diptera: Tephritidae). 1996 Tese (Doutorado em Genética) – Instituto de Biociências do Departamento de Biologia, Universidade de São Paulo, São Paulo.

SELIVON, D. Biologia e padrões de especiação. *In*: Malavasi, A. e Zucchi, R.A. (ed.). moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado. Ribeirão Preto: Holos Editora, 2000

SELIVON, D.; PERONDINI, A.L.P.; ROCHA, L.S. Karyotype characterization of *Anastrepha* fruit flies (Diptera : Tephritidae). **Neotrop. Entomol.**, v. 34, n. 2, p. 273-279, 2005

SHARROCK, W. J. Cleavage of two yolk proteins from a precursor in *C. elegans*. **J. Mol. Biol.**, v. 174, n. 3, p. 419-431, 1984.

SMITH-CALDAS, E.R.B. et al. Phylogenetic relationships among species of the *fraterculus* group (*Anastrepha*: Diptera:Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. **Neotrop. Entomol.**, v. 30, n. 4, p. 565 - 573, 2001.

SOBRINHO-JR, I.S.; DE BRITO, R.A. Evolução molecular dos genes doublesex e fruitless em moscas-das-frutas do grupo *Anastrepha fraterculus* (Diptera, Tephritidae). 2009. Tese (Doutorado em Genética e Evolução) – Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo.

SPIEL, J. et al. The *C. elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. **Nuc. Acid. Res.**, v. 13, n. 14, p. 5283-5295, 1985

STECK, G.J. Biochemical systematics and population genetic structure of *Anastrepha fraterculus* and related species (Diptera: Tephritidae). **Ann. Ent. Soc. Am.**, v. 84, p. 10-28, 1991

TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, p. 585–595, 1989

TEMPLETON, A.R. Human origins and analysis of mitochondrial DNA sequences. **Science**, v. 255, n. 5045, p. 737, 1992

TEMPLETON, A.R.; SING, C.F. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. **Genetics**, v. 132, p. 659–669, 1993

TEMPLETON, A.R.; ROUTMAN, E.; PHILLIPS, C.A. Separating population structure from population history: A cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger salamander, *Ambystoma tigrinum*. **Genetics**, v. 140, p.767-782, 1995

TEMPLETON, A.R. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. **Mol. Ecol.**, v.7, p.381- 397, 1998

TEMPLETON, A.R. et al. Cladistic structure within the human lipoprotein lipase gene and its implications for phenotypic association studies. **Genetics**, v. 156, p.1259-1275, 2000

TEMPLETON, A.R. Using phylogeographic analyses of gene trees to test species status and processes. **Mol. Ecol.**, v. 10, p. 779–791, 2001

TEMPLETON A. R. Statistical phylogeography: methods of evaluating and minimizing inference errors. **Mol. Ecol.**, v. 13, p.789-809, 2004

TEMPLETON, A.R. Haplotype trees and modern human origins. **Yearb. of Phys. Anthropol.**, v. 48, p. 33–59, 2005

TEMPLETON, A.R. Nested clade analysis: an extensively validated method for strong phylogeographic inference. **Mol. Ecol.**, v. 17, n. 8, p. 1877-1880, 2008

TERBORGH, J. Maintenance of diversity in tropical forests. **Biotropica**, v. 24, n. 2, p. 283-292, 1992

TSAUR, S.C.; WU ,C.I. Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. **Mol. Biol. Evol.** v. 14, n. 5, p. 544-549, 1997

URAMOTO, K.; WALDER, J. M. M.; ZUCCHI, R. A. Biodiversidade de moscas-das-frutas do gênero *Anastrepha* (Diptera, Tephritidae) no campus da ESALQUSP, Piracicaba, SãoPaulo. **Rev. Bras. de Entom.**, v. 48, n. 3, p. 409-414, 2004

VALE, D. Vitellogenesis in insects and other groups: a review. **Mem. Inst. Oswaldo Cruz**, v. 88, n. 1, p. 1-26, 1993

WAHLI, W.; RYFFEL, G.U. *Xenopus* vitellogenin genes. **Oxf. Surv. Eukaryot. Genes**, v. 2, p. 96-120, 1985

WAHLI, W. Evolution and expression of vitellogenin genes. **Trends Genet.**, v. 4, n. 8, p. 227-232, 1988

WEISS, S.; FERRAND, N. Phylogeography of Southern European Refugia. **Springer**, Netherlands, p. 390, 2007

WU, C.I.; TING, C.T. Genes and speciation. **Nature**, v. 5, p. 114-122, 2004

ZENG, Z.B. et al. Genetic architecture of a morphological shape difference between two *Drosophila* species. **Genetics**, v. 154, p. 299-310, 2000

ZUCCHI, R.A. Espécies de *Anastrepha*, sinónímias, plantas hospedeiras e parasitóides. *In*: Malavasi, A. e Zucchi, R.A. (Ed.). Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado. Ribeirão Preto: Holos Editora, p. 41-48, 2000