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

EVOLUÇÃO MOLECULAR DA FAMÍLIA GÊNICA DOS RECEPTORES DE ODORES E PROTEÍNAS LIGANTES A FEROMÔNIOS E GENÉTICA DE POPULAÇÕES DE GENES QUIMIOSSENSORIAIS EM ESPÉCIES DE *ANASTREPHA* DO GRUPO FRATERCULUS.

São Carlos
2016

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

Evolução molecular da família gênica dos receptores de odores e proteínas ligantes a feromônios e genética de populações de genes quimiossensoriais em espécies de *Anastrepha* do grupo fraterculus.

Candidata: Diana Marcela Rojas Gallardo

Dissertação apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular do Centro de Ciências Biológicas e da Saúde da Universidad Federal de São Carlos para obtenção do título de Mestre em Genética e Evolução.

Orientação: Dr. Reinaldo Alves de Brito

São Carlos
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

R741e

Rojas Gallardo, Diana Marcela
Evolução molecular da família gênica dos receptores de odores e proteínas ligantes a feromônios e genética de populações de genes quimiossensoriais em espécies de Anastrepha do grupo fraterculus / Diana Marcela Rojas Gallardo. -- São Carlos : UFSCar, 2016.
95 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2016.

1. Grupo fraterculus. 2. Receptores olfativos. 3. Proteínas ligantes a odores. 4. Seleção positiva. I. Título.

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Diana Marcela Rojas Gallardo, realizada em 02/05/2016:

Prof. Dr. Reinaldo Otavio Alvarenga Alves de Brito
UFSCar

Prof. Dr. Marco Antonio Del Lama
UFSCar

Profa. Dra. Maura Helena Manfrin
USP

Agradecimentos

Agradeço à CAPES, CNPq e FAPESP pelo apoio financeiro.

Agradeço especialmente ao Prof. Dr. Reinaldo Alves de Brito pela oportunidade, confiança e orientação. Obrigada por acreditar em min e me permitir fazer parte do seu grupo e pelo respeito que ele sempre demonstrou por mim e pelo meu trabalho.

Aos meus amigos do Laboratório de Genética de Populações e Evolução: André, Emeline, Felipe, Carlos, Fernando, Manu, Samira, Nancy, Isabela, Isabel e Paty pela amizade, carinho e ajuda.

A minha bela família e meu amado chiquito por todo o apoio e amor na distância.

A Mayra e todos os amigos que fiz durante esse tempo por aqueles momentos em que me fizeram sentir como se estivesse na minha terrinha.

Resumo

A presente dissertação encontra-se dividida em três capítulos. O primeiro capítulo apresenta uma concisa revisão bibliográfica que aborda os principais conceitos teóricos, a justificativa e os objetivos delineados para este estudo. O segundo capítulo apresenta um estudo da evolução molecular da família gênica dos receptores de odores (ORs) identificados nos transcriptomas de duas espécies de moscas-das-frutas de grande importância econômica: *Anastrepha fraterculus* e *A.obliqua*. Os resultados mostraram uma alta porcentagem de identidade média entre os ORs destas espécies, assim como expansões gênicas recentes com sinal de seleção positiva. Quando comparamos as taxas de substituições sinônimas e não-sinônimas entre as espécies de *Anastrepha* encontramos evidências de seleção positiva no gene *Or7c*, que está associado em *D. melanogaster* a um potencial importante papel nos comportamentos de agregação e escolha de frutos para oviposição. No terceiro capítulo apresentamos um estudo do padrão de evolução molecular dos genes que codificam para proteínas ligantes aos feromônios (PBPs), também identificados em *A. fraterculus* e *A. obliqua*, assim como também estudamos o padrão de polimorfismos, divergência e estrutura dos genes quimiossensoriais *Obp28a*, *Obp84a, Or7c* e *Or7d* os quais foram amplificados em quatro espécies de moscas-das-frutas do grupo fraterculus, *A. fraterculus*, *A. obliqua*, *A. sororcula* e *A. turpiniae*. Este estudo foi realizado contrastando genes identificados com sinais de seleção positiva e seleção purificadora com o intuito de investigar se eles estão contribuindo para a diferenciação entre algumas das espécies desse grupo. Não encontramos evidências de seleção positiva nas PBPs estudadas em uma comparação mais global, embora tenhamos encontrado sinais de seleção positiva em alguns dos genes e linhagens estudadas. A análise populacional de genes quimiossensoriais em diferentes espécies de *Anastrepha* detectou níveis altos de diversidade nucleotídica e haplotípica dentro das espécies. Os testes de divergência mostraram que a espécie *A. obliqua* é a espécie mais diferenciada, apresentando, em geral, altos níveis de substituições nucleotídicas, divergência não-sinônima, assim como diferenças fixadas quando comparada com as outras espécies. Os genes *Obp28a*, *Or7c* e *Or7d* mostraram-se diferenciados em *A. obliqua*, indicando um potencial papel na diferenciação desta espécie com respeito às outras espécies estudadas.

Palavras chave: Grupo fraterculus, receptores olfativos, proteínas ligantes a odores, seleção positiva, divergência.

Abstract

This dissertation is divided into three chapters. In the first chapter, we provide a concise literature review that discusses key theoretical concepts, the rationale, and main objectives outlined for this study. The second chapter investigates the molecular evolution of the gene family of odor receptors (ORs) identified in the transcriptomes of two species of fruit flies of great economic importance: *Anastrepha fraterculus* and *A. obliqua*. The results showed a high percentage of average identities between *OR*s from these species, as well as recent gene expansions with signs of positive selection. A comparison of rates of synonymous and non-synonymous substitutions among *Anastrepha* species detected evidence of positive selection in the gene *Or7c*, which is associated to an important potential role in aggregation behavior and host choice for oviposition in *D. melanogaster*. The third chapter investigates patterns of molecular evolution in pheromone binding proteins (*PBP*s), also identified in *A. fraterculus* and *A. obliqua*, as well as studied pattern of polymorphisms, divergence and populational structure of four chemosensory genes amplified in four species of tephritid flies of fraterculus group: *A. fraterculus*, *A. obliqua*, *A. sororcula* and *A. turpiniae*. This study contrasted previously identified genes with evidence of positive and purifying selection in order to investigate whether they are contributing to the differentiation among some of the species of this group. We found no evidence of positive selection in PBPs studied in a more global comparison, although we found positive selection signals in some of the genes and studied strains. Population analysis of chemosensory genes in different species of *Anastrepha* detected high levels of intraspecific nucleotide and haplotype diversity. Divergence tests showed that *A. obliqua* is the most different species of the ones here investigated, having, in general, high levels of nucleotide substitutions, non-synonymous divergence, as well as fixed species specific differences, whereas we failed to find similar differences amongst the other species here studied. The genes *Obp28a*, *Or7c* and *Or7d* were differentiated in *A. obliqua*, indicating a potential role in the differentiation of other species in the group, or in this species' diversification and adaptation.

Key words: Fraterculus group, odorant receptors, pheromone binding proteins, positive selection, divergence.

Sumário

1. Revisão bibliográfica 8

1.1 Moscas-das-frutas do gênero *Anastrepha.* 8

1.2 O sistema olfatório dos insetos 10

1.3 Receptores olfativos 11

1.4 Proteínas de ligação aos Feromônios 13

1.5 Justificativa 15

1.6 Objetivos 16

1.7 Referências 17

2. Molecular evolution of Odorant Receptors proteins gene family in two closely related species of fruit flies from the Anastrepha fraterculus group 20

2.1 Resumo 20

2.2 Abstract 21

2.3 Introduction 22

2.4 Materials and methods 23

2.4.1 Identification of A. fraterculus and A. obliqua OR *genes* 23

2.4.2 Phylogenetic analysis 24

2.4.3 Molecular evolution analysis 25

2.5 Results and Discussion 26

2.5.1 Olfactory receptors genes in Anastrepha fraterculus and Anastrepha obliqua 26

2.5.2 Phylogenetic relationships among OR genes 29

2.5.3 Analysis of positive selection in ORs 34

2.6 Conclusions 44

2.7 References 45

3. Divergence and variation in two pheromone binding proteins and two odorant receptors in four *Anastrepha* species of the fraterculus group. 48

3.1. Resumo 48

3.2. Abstract 49

3.3. Introduction 50

3.4. Materials and methods 52

3.4.1. Identification of A. fraterculus and A. obliqua PBPs genes 52

3.4.2. Phylogenetic analysis of PBPs 52

3.4.3. Molecular evolution analysis and candidate genes 53

3.4.4. Species and populations sampled 54

3.4.5. Molecular procedures 56

3.4.6. Diversity, neutrality and divergence analysis among Anastrepha species 56

3.5. Results and Discussion 59

3.5.1. Pheromone binding proteins genes in Anastrepha fraterculus and Anastrepha obliqua 59
3.5.2. Molecular evolution and positive selection in PBPs 61
3.5.3. Diversity, neutrality and divergence among Anastrepha species 62
3.6. Conclusions 75
4. Considerações finais 80
4.1. Referências 81
5. Anexos 82
5.1. Supplementary Material Chapter 2. 82
5.2. Supplementary Material Chapter 3. 89

1. Revisão bibliográfica

1.1 Moscas-das-frutas do gênero *Anastrepha.*

Moscas-das-frutas da família *Tephritidae* são de importância econômica uma vez que algumas das espécies infestam a maioria das frutas produzidas no país (ZUCCHI, 2000). Esta família tem cerca de 500 gêneros, sendo *Anastrepha* (Schiner, 1868) o mais representativo e diversificado com 230 espécies conhecidas até agora (HERNÁNDEZ-ORTIZ et al., 2012). A importância econômica das moscas deste gênero reside em que são responsáveis por danos às culturas carnosas, uma vez que as larvas se desenvolvem no endocarpo da fruta e o processo de oviposição permite a invasão de outros organismos como fungos, provocando o amadurecimento precoce dos frutos (DUARTE; MALAVASI, 2000).

Economicamente as espécies mais importantes dentro deste gênero são *A. grandis* (Macquart, 1846), *A. fraterculus* (Wiedemann, 1830), *A. obliqua* (Macquart, 1835), *A. pseudoparallela* (Loew, 1873), *A. sororcula* (Zucchi, 1979), *A. striata* (Schiner, 1868) e *A. zenildae* (Zucchi, 1979) (Zucchi, 2000). A maioria destas pertence ao grupo de espécies fraterculus, que é composto por 29 espécies (SMITH-CALDAS et al., 2001). Neste trabalho, daremos foco em quatro espécies do grupo; *Anastrepha fraterculus*, *Anastrepha obliqua*, *Anastrepha sororcula* e *Anastrepha turpiniae*, as quais são importantes não apenas por sua ampla distribuição, mas também por suas preferências de frutos para oviposição.

As espécies deste grupo são consideradas polífagas, pois embora haja alguma preferência por determinados grupos de plantas, em geral elas não estão restritas a um único hospedeiro (SOLFERINI; MORGANTE, 1987). No Brasil existem 38 famílias de plantas hospedeiras que são parasitadas por *Anastrepha*, dentro dessas famílias as principais são *Myrtaceae*, *Anacardiaceae*, *Sapotaceae* e *Passifloraceae*. As espécies que apresentam maior número de hospedeiros associados são *A. fraterculus* com 81, *A. obliqua* com 71 e *A. sororcula* com 21. Para *A. fraterculus* os hospedeiros mais frequentes (40% das espécies) são representantes da família *Myrtaceae*, assim como para *A. sororcula* (71%), en quanto que *A. obliqua* infesta preferencialmente frutos da família *Anacardiaceae* (SELIVON, 2000).

As moscas das frutas completam seu desenvolvimento passando por quatro estágios: ovo, larva, pupa e adulto. O ciclo de vida consiste na oviposição em frutos carnosos, os ovos eclodem em larvas que se alimentam e consomem parte do fruto o qual apodrece e cai no solo. As larvas saem do fruto e empupam enterradas no solo, depois de entre oito e 14 dias, as pupas dão origem a moscas que serão maduras sexualmente após cinco a sete dias, recomeçando o ciclo (SALLES, 2000).

Alguns marcadores morfológicos têm sido utilizados para a diferenciação das espécies do grupo fraterculus, o qual engloba diversas espécies proximamente relacionadas e que são de difícil diferenciação (SMITH-CALDAS et al., 2001). Entre estes marcadores estão: a morfologia do ápice do ovipositor (principal marcador morfológico utilizado), presença ou ausência de manchas no subescutelo e o padrão de venação e coloração das asas (PERRE et al., 2014). Este conjunto de características é capaz, na maioria das vezes, de distinguir entre as espécies. Porém alguns marcadores, como a morfologia do ápice do ovipositor, apresentam sobreposição interespecífica tornando difícil a diferenciação das espécies mais intimamente associadas dentro do grupo (ARAUJO; ZUCCHI, 2006).

Em função da dificuldade da identificação de espécies do grupo fraterculus usando marcadores morfológicos, também têm sido avaliados marcadores comportamentais (MORGANTE; MALAVASI; PROKOPY, 1983), alozímicos (MATIOLI; MORGANTE; MALAVASI, 1986), cariotípicos (SOLFERINI; MORGANTE, 1987) e moleculares do tipo mitocondrial e nuclear (SMITH-CALDAS et al., 2001; SOBRINHO, 2009), mas nenhum ainda se mostrou completamente eficaz para a diferenciação das espécies (BARR; CUI; MCPHERON, 2005).

A similaridade entre as espécies do grupo fraterculus possivelmente se deve a uma divergência recente na qual as espécies ainda não acumularam diferenças genéticas significativas, além daquelas oriundas do processo da especiação, sendo obtidos híbridos viáveis em condições de laboratório (DOS SANTOS; URAMOTO; MATIOLI, 2001; SELIVON; PERONDINI; MORGANTE, 2005). Isto faz que o grupo fraterculus seja um ótimo modelo para o estudo dos processos envolvidos em eventos de especiação e diferenciação populacional, como por exemplo, o estudo da evolução de famílias gênicas, que geralmente desempenham um importante papel na construção de diferenças morfológicas, fisiológicas e metabólicas entre as espécies (HAHN; HAN; HAN, 2007).

A comparação de genes em distintas espécies tem sido a melhor estratégia para a compreensão dos processos envolvidos na geração da diversidade genômica (SIEGLAFF et al., 2009). Esses estudos são ainda mais relevantes em espécies que divergiram há pouco tempo, como é o caso do grupo fraterculus. A análise comparativa de genes que possam estar envolvidos nos processos de diferenciação ecológica entre espécies desse grupo, como a escolha de frutos hospedeiros ou parceiros, pode contribuir para uma maior compreensão acerca dos mecanismos que estão envolvidos no processo de isolamento especiação das moscas deste grupo.

1.2 O sistema olfatório dos insetos

A percepção sensorial é fundamental para a sobrevivência e reprodução de qualquer organismo. Especificamente identificar odores é fundamental para executar funções básicas, tais como a detecção de alimentos e substratos de oviposição, detectar a presença de possíveis parceiros ou predadores, e também para a comunicação e coordenação social (VIEIRA; SÁNCHEZ-GRACIA; ROZAS, 2007).

O sistema sensorial dos insetos é caracterizado pela sua elevada especificidade e sensibilidade. Nos dípteros o sistema olfativo é composto por dois órgãos principais: as antenas e os palpos maxilares. Na superfície desses órgãos há estruturas porosas chamadas cerdas ou sensilas sensoriais, que geralmente possuem dois dendritos de neurônios olfativos. Cada antena contém cerca de 1200 neurônios olfativos, e o terceiro segmento da antena é o maior órgão olfativo com três tipos de cerdas: basicônica, tricóides e coelônicas (SU; MENUZ; CARLSON, 2009).

O passo inicial para a codificação da informação olfativa em insetos ocorre nos neurônios receptores das cerdas olfativas. Esta etapa envolve a recepção de moléculas provenientes do ambiente externo, o transporte através das sensilas e a interação com o receptor neuronal (Figura 1) (SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009). As principais proteínas envolvidas nesta primeira etapa são: Proteínas ligantes aos odores *(Odorant Binding Proteins)*, enzimas degradantes de odores *(Odorant Degrading Enzimes)*, receptores de odores *(Odorant Receptors)* e proteínas de membrana de neurônio sensorial.

No segundo passo a ligação da molécula de odor com o receptor gera uma mudança no potencial transmembrana das dendrites conhecido como potencial receptor. A despolarização da membrana abre o canal iônico o qual está envolvido na geração do potencial de ação cuja frequência de disparo constitui o código neural do sinal químico (RENOU, 2014). A transformação neural do código que define a percepção do odor ocorre no cérebro onde a percepção é processada e leva a uma resposta comportamental, de acordo com o estado interno do animal e com a integração de outras modalidades sensoriais (BRUYNE, M., 2003).

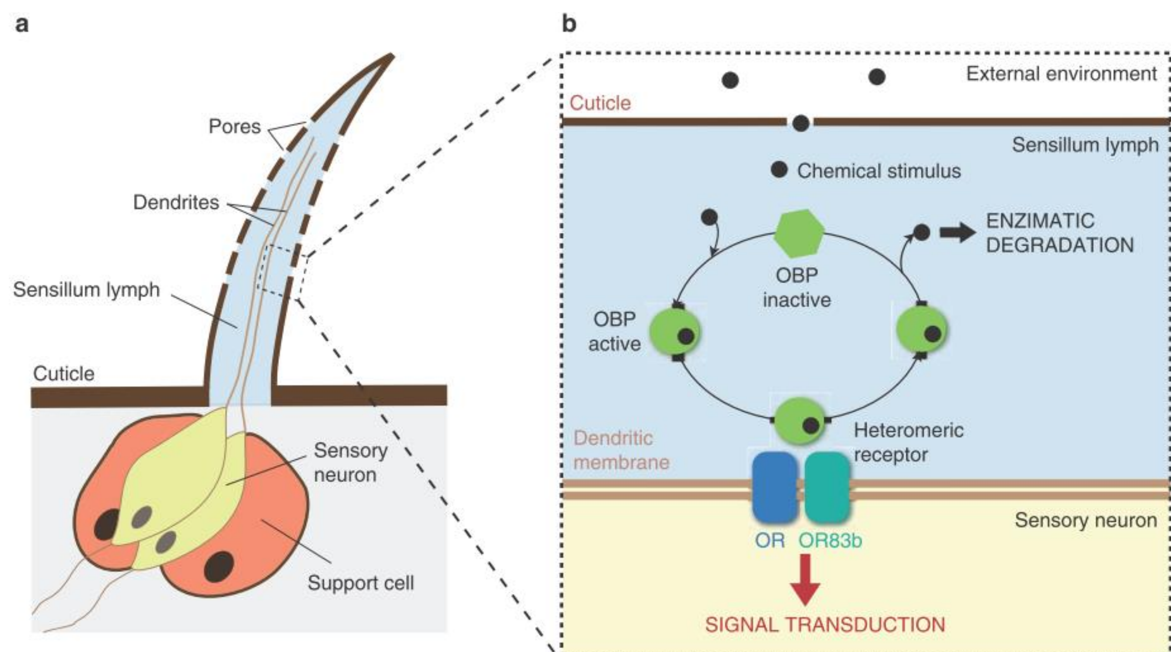


Figura 1. a) Representação esquemática da estrutura geral de uma sensila olfatória. b) Primeiro passo molecular na transdução de sinais quimiosensoriais em insetos. Fonte: Sanchez et al., 2009.

1.3 Receptores olfativos

Os receptores olfativos (ORs) são proteínas com sete domínios trans-membrana e cerca de 400 aminoácidos, que se ligam às moléculas de odor provenientes do ambiente, e convertem o sinal químico para a ativação dos neurônios sensoriais. Nos insetos os ORs têm uma topologia de membrana invertida em que o C-terminal está localizado na superfície extracelular formando canais de cátions dependentes de odores (SATO et al., 2008; WICHER et al., 2008). A maioria dos neurônios olfativos expressa dois receptores que atuam em

conjunto: um OR específico e o co-receptor denominado *Orco* (Or83b). A conformação deste complexo é essencial para a localização celular e a função do canal iônico (Figura 1) (BENTON et al., 2006; LARSSON et al., 2004). O Orco é altamente conservado entre as espécies de insetos e o único gene ortólogo que tem sido descrito nas diferentes ordens de insetos, mostrando o papel essencial desse receptor na percepção de odores (JONES et al., 2005).

Os genes que codificam os ORs têm sido estudados em várias espécies de insetos e os resultados destes estudos mostram que os repertórios podem ser altamente variáveis, sendo descritos desde zero na traça (MISSBACH et al., 2014), 60 em *Drosophila melanogaster*, 80 em *Anopheles gambiae,* 160 genes em *Apis mellifera* (NOZAWA; NEI, 2007) até 400 em formigas eussociais (ZHOU et al., 2012). Estes genes são membros de uma família gênica de rápida divergência, em *D. melanogaster*, por exemplo, o par mais divergente tem uma similaridade de apenas 10% em suas sequências de aminoácidos (GUO; KIM, 2007)

Nos repertórios de genes ORs dos insetos é possível encontrar notáveis expansões das diferentes subfamílias que são específicas para cada espécie. Por exemplo, os ORs em *Drosophila* apresentam 15 grupos filogenéticos diferentes, cada um com um número variável de membros parálogos (HILL et al., 2002). O principal mecanismo envolvido na evolução desta família gênica é o “birth and death”, no qual novos membros da família são gerados por duplicação gênica em tandem, progressivamente divergem em sequência e função e eventualmente podem ser perdidos por uma deleção ou por um evento de pseudogenização (SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009).

Em *Drosophila* os genes *ORs* têm padrões de evolução multimodal com mudanças na estrutura genética (perda ou ganho de íntrons), na localização cromossômica, duplicação seguida de ganho ou perda de nova função. Alguns desses genes estão sujeitos à seleção positiva, indicando que a duplicação de genes pode ser importante para moscas adquirirem nova função olfativa (ENGSONTIA et al., 2014; GARDINER et al., 2009; GUO; KIM, 2007; SMADJA et al., 2009).

Tem sido sugerido que a perda ou ganho funcional dos receptores olfativos teve um papel importante nas transições ecológicas e tróficas de algumas espécies (BOHBOT; PITTS, 2015; CAREY et al., 2010; GUO; KIM, 2007; MCBRIDE, 2007). Por exemplo, GOLDMAN-HUERTAS et al.,2015 estudaram o repertório de genes ORs de *Scaptomyza flava*

(*Drosophilidae*) uma espécie herbívora especialista em folhas de plantas da família *Brassicaceae* e que não é atraída por voláteis de levedura. Quando comparam o repertório de *S. flava* com o repertório de *D. melanogaster* encontraram que em *S. flava* foram suprimidos os receptores olfativos que detectam os voláteis de leveduras em *D. melanogaster* e medeiam o comportamento na seleção de hospedeiro.

Existem também exemplos que sugerem que mudanças discretas que alteram a especificidade nos receptores olfativos podem fornecer um mecanismo que contribui para o isolamento reprodutivo entre espécies. Por exemplo, o OR responsável pelo reconhecimento do feromônio sexual entre duas espécies proximamente relacionadas do gênero *Ostrinia* (Lepidóptera), as quais estão isoladas reprodutivamente por suas respostas a feromônios sutilmente diferenciados, apresentam uma mutação no receptor de *Ostrinia furnacalis* que altera o padrão de reconhecimento do feromônio reduzindo a resposta desta espécie ao feromônio produzido por *Ostrinia nubilalis* (LEARY et al., 2012).

Em função das potenciais pressões seletivas e rápida adaptação, os genes da família dos ORs apresentam uma alta taxa evolutiva (VIEIRA; SÁNCHEZ-GRACIA; ROZAS, 2007), o que os tornam um objeto de estudo muito promissor para avaliar os processos evolutivos envolvidos durante e depois da especiação do gênero *Anastrepha*.

1.4 Proteínas de ligação aos Feromônios

As proteínas ligantes a odores (OBPs) são pequenas proteínas globulares solúveis (10 a 30 kDa) que são sintetizadas e segregadas por células que rodeiam os neurônios olfativos auxiliares. Possuem de 120 a 150 aminoácidos, e são caracterizadas por terem um domínio específico com seis resíduos de cisteína em posições altamente conservadas que formam três pontes dissulfeto (HONSON; GONG; PLETTNER, 2005). Estas proteínas atuam como veículos moleculares para transportar e distribuir o odor para os receptores olfativos (LI et al., 2014). Além disso, acredita-se que elas possam ter outras funções fisiológicas, dado que, análises de expressão mostraram que as OBPs não estão restritas aos tecidos olfatórios e gustatórios (PELOSI et al., 2006).

As OBPs pertencem a uma família gênica que tem sido identificada em uma grande variedade de insetos, apresentando em geral membros parálogos de uma espécie extremamente divergentes, assim como as comparações entre ortólogos e parálogos entre espécies (SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009). A evolução desta família gênica segue o modelo “birth and death” e é altamente dinâmica com um número elevado de perdas e ganhos de genes e surgimento de diferentes linhagens ou subfamílias (VIEIRA; ROZAS, 2011). A família das OBPs apresenta um número variável de membros que vão desde 21 em *Apis melífera* (FORET; MALESZKA, 2006) até 66 em *Anopheles gambiae* (MANOHARAN et al., 2013).

Em insetos, a família das OBPs pode ser dividida em duas classes principais: proteínas ligantes a odores gerais (GOBP1 e GOBP2) e proteínas ligantes a feromônios (PBPs), sendo esta última o foco de estudo neste trabalho. As PBPs participam no transporte de feromônios sexuais hidrofóbicos (PRESTWICH; DU; LAFOREST, 1995) e facilitam a remoção dos metabolitos de feromônios após ter ocorrido a transdução do sinal (VOGT; RIDDIFORD, 1981). As PBPs exibem especificidade na ligação aos seus respectivos feromônios, embora às vezes podem se ligar a outras moléculas semelhantes, cada PBP interage com um feromônio específico de uma maneira única (MOHANTY; ZUBKOV; GRONENBORN, 2004).

Estudos de comparações entre genes ortólogos de PBPs e GOBPs entre espécies de mariposas mostram que as PBPs têm taxas de substituição evolutivas mais altas do que a GOBPs, o que foi sugerido como sendo causado pela seleção imposta pelas mudanças relativamente frequentes de feromônios entre as espécies (VOGT; RYBCZYNSKI; LERNER, 1991). Estes estudos também mostram que a divergência funcional presente na diferenciação das copias geradas por duplicação gênica, provavelmente, está associada à seleção positiva (SÁNCHEZ-GRACIA; ROZAS, 2008).

Uma análise da variação nas sequências de duas PBPs em quatro espécies do subgrupo melanogaster e em uma amostra populacional de *D. melanogaster* revelou que o gene *Obp83a* apresentava, na sequência de aminoácidos, um número alto de substituições fixadas para diferentes espécies, sugerindo a ação da seleção natural na diferenciação deste gene (SÁNCHEZ-GRACIA; AGUADÉ; ROZAS, 2003). Outro estudo encontrou evidência de seleção positiva favorecendo fixação de alterações de aminoácidos nas PBP de mariposas do gênero *Choristoneura*, particularmente em uma linhagem de *Choristoneura fumiferana* na qual aconteceu uma mudança no uso um feromônio de aldeído para um de acetato

(WILLETT, 2000). Estes estudos revelam que a investigação de genes PBPs tem um importante potencial para o entendimento da diferenciação de espécies filogeneticamente próximas.

Assim, o estudo de PBPs em espécies de *Anastrepha* é de grande importância, uma vez que um melhor entendimento dos genes envolvidos na recepção de feromônios entre estas espécies proximamente relacionadas pode ajudar a determinar se estes genes podem estar potencialmente contribuindo para o seu isolamento reprodutivo.

1.5 Justificativa

Têm sido observadas alterações na evolução das famílias quimiosensoriais dos insetos e provavelmente envolvem importantes aspectos reprodutivos, ecológicos e/ou comportamentais. Estudar o padrão global de evolução das famílias PBPs e ORs pode oferecer uma perspectiva sobre as forças seletivas que estão impulsionando a evolução desses genes olfativos e ter uma compreensão do papel potencial destas famílias na divergência das moscas-das-frutas do grupo *fraterculus*. O estudo destas duas famílias de genes pode também revelar marcadores genéticos que podem ser úteis na diferenciação de espécies. Estes atributos são particularmente importantes pela dificuldade na separação de espécies do grupo fraterculus, quer pela sua recente divergência, quer pela manutenção de fluxo gênico recorrente entre espécies do grupo, e pela importância do entendimento dos processos subjacentes envolvidos em sua diferenciação.

Diante do exposto, o presente trabalho irá avaliar a diversidade e o padrão de evolução da família gênica dos ORs e das PBPs em espécies do grupo *Anastrepha* com o intuito de identificar genes com altas taxas de evolução e sinais de seleção positiva. Este estudo também irá avaliar a variabilidade intra e interespecífica de alguns genes quimiossensoriais, buscando, a determinação do quanto dessa diversidade genética pode ter implicações na diferenciação das espécies estudadas, estabelecendo o quanto dessa diversidade é compartilhada entre espécies e quanto é específica da espécie.

1.6 Objetivos

Objetivo Geral

Avaliar o padrão de evolução da família dos ORs *(Olfactory Receptors - Receptores de Odor)* dos genes PBPs *(Pheromone Binding Protein - Proteínas ligantes a feromônios)* em espécies do grupo fraterculus e avaliar o papel dos genes ORs e dos genes PBPs na diferenciação das espécies deste grupo.

Objetivos específicos

- Isolar e caracterizar *in silico* genes das famílias *PBPs e ORs* em bibliotecas de *cDNA* de tecidos de cabeça e reprodutivos de machos e fêmeas de *A. fraterculus* e *A. obliqua*, previamente gerados em nosso laboratório.
- Investigar a história e padrões evolutivos dos genes *ORs e PBPs* encontrados, de forma a permitir a identificação de genes com taxas evolutivas altas e sinais de seleção positiva ou purificadora influenciando sua evolução.
- Avaliar diferenciação intra e interespecífica em espécies do grupo fraterculus pela investigação mais detalhada de genes *Obp28a*, *Obp84a*, *Or7c* e *Or7d* em diferentes populações de *Anastrepha fraterculus*, *A. obliqua*, *A. sororcula* e *A. turpinae*.

1.7 Referências

ARAUJO, E. L.; ZUCCHI, R. A. Medidas do acúleo na caracterização de cinco espécies de Anastrepha do grupo fraterculus (Diptera: Tephritidae). **Neotropical Entomology**, v. 35, n. 3, p. 329–337, jun. 2006.

BARR, N. B.; CUI, L.; MCPHERON, B. A. Molecular Systematics of Nuclear Gene period; in Genus Anastrepha (Tephritidae). **Annals of the Entomological Society of America**, v. 98, n. 2, p. 173–180, 1 mar. 2005.

BENTON, R. et al. Atypical Membrane Topology and Heteromeric Function of Drosophila Odorant Receptors In Vivo. **PLoS Biology**, v. 4, n. 2, p. e20, 17 jan. 2006.

BOHBOT, J. D.; PITTS, R. J. The narrowing olfactory landscape of insect odorant receptors. **Frontiers in Ecology and Evolution**, v. 3, 10 abr. 2015.

BRUYNE, M. Physiology and genetics of odor perception in Drosophila. In: **Insect Pheromone Biochemistry and Molecular Biology**. Blomquist, G., Vogt, R ed. Elsevier Academic Press, 2003. p. 475.

CAREY, A. F. et al. Odorant reception in the malaria mosquito Anopheles gambiae. **Nature**, v. 464, n. 7285, p. 66–71, 4 mar. 2010.

DOS SANTOS, P.; URAMOTO, K.; MATIOLI, S. R. Experimental Hybridization Among Anastrepha Species (Diptera: Tephritidae): Production and Morphological Characterization of Hybrids. **Annals of the Entomological Society of America**, v. 94, n. 5, p. 717–725, 1 set. 2001.

DUARTE, A. L.; MALAVASI, A. Tratamentos quarenterários. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado.** Malavasi, A. Zucchi, RA ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

ENGSONTIA, P. et al. Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. **Journal of Molecular Evolution**, v. 79, n. 1-2, p. 21–39, ago. 2014.

FORET, S.; MALESZKA, R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). **Genome Research**, v. 16, n. 11, p. 1404–1413, 25 out. 2006.

GARDINER, A. et al. Sites of evolutionary divergence differ between olfactory and gustatory receptors of Drosophila. **Biology Letters**, v. 5, n. 2, p. 244–247, 23 abr. 2009.

GOLDMAN-HUERTAS, B. et al. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 10, p. 3026–3031, 10 mar. 2015.

GUO, S.; KIM, J. Molecular Evolution of Drosophila Odorant Receptor Genes. **Molecular Biology and Evolution**, v. 24, n. 5, p. 1198–1207, 13 fev. 2007.

HAHN, M. W.; HAN, M. V.; HAN, S.-G. Gene Family Evolution across 12 Drosophila Genomes. **PLoS Genetics**, v. 3, n. 11, p. e197, 2007.

HERNÁNDEZ-ORTIZ, V. et al. Cryptic Species of the Anastrepha fraterculus Complex (Diptera: Tephritidae): A Multivariate Approach for the Recognition of South American Morphotypes. **Annals of the Entomological Society of America**, v. 105, n. 2, p. 305–318, 1 mar. 2012.

HILL, C. A. et al. G protein-coupled receptors in Anopheles gambiae. **Science (New York, N.Y.)**, v. 298, n. 5591, p. 176–178, 4 out. 2002.

HONSON, N. S.; GONG, Y.; PLETTNER, E. Structure and Function of Insect Odorant and Pheromone-Binding Proteins (OBPs and PBPs) and Chemosensory-Specific Proteins (CSPs). In: **Recent Advances in Phytochemistry**. Elsevier, 2005. v. 39p. 227–268.

JONES, W. D. et al. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. **Current biology: CB**, v. 15, n. 4, p. R119–121, 22 fev. 2005.

LARSSON, M. C. et al. Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. **Neuron**, v. 43, n. 5, p. 703–714, 2 set. 2004.

LEARY, G. P. et al. Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 35, p. 14081–14086, 28 ago. 2012.

LI, H. et al. Construction and analysis of cDNA libraries from the antennae of Batocera horsfieldi and expression pattern of putative odorant binding proteins. **Journal of Insect Science (Online)**, v. 14, p. 57, 2014.

MANOHARAN, M. et al. Comparative genomics of odorant binding proteins in Anopheles gambiae, Aedes aegypti, and Culex quinquefasciatus. **Genome Biology and Evolution**, v. 5, n. 1, p. 163–180, 2013.

MATIOLI, S. R.; MORGANTE, J. S.; MALAVASI, A. Genetical and biochemical comparisons of alcohol dehydrogenase isozymes from Anastrepha fraterculus and A. obliqua (Diptera: Tephritidae): Evidence for gene duplication. **Biochemical Genetics**, v. 24, n. 1, p. 13–24, 1986.

MCBRIDE, C. S. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 12, p. 4996–5001, 20 mar. 2007.

MISSBACH, C. et al. Evolution of insect olfactory receptors. **eLife**, v. 3, 26 mar. 2014.

MOHANTY, S.; ZUBKOV, S.; GRONENBORN, A. M. The solution NMR structure of Antheraea polyphemus PBP provides new insight into pheromone recognition by pheromone-binding proteins. **Journal of Molecular Biology**, v. 337, n. 2, p. 443–451, 19 mar. 2004.

MORGANTE, J. S.; MALAVASI, A.; PROKOPY, R. J. Mating Behavior of Wild Anastrepha fraterculus (Diptera: Tephritidae) on a Caged Host Tree. **The Florida Entomologist**, v. 66, n. 2, p. 234, jun. 1983.

NOZAWA, M.; NEI, M. Evolutionary dynamics of olfactory receptor genes in Drosophila species. **Proceedings of the National Academy of Sciences**, v. 104, n. 17, p. 7122–7127, 24 abr. 2007.

PELOSI, P. et al. Soluble proteins in insect chemical communication. **Cellular and molecular life sciences: CMLS**, v. 63, n. 14, p. 1658–1676, jul. 2006.

PERRE, P. et al. Morphometric Differentiation of Fruit Fly Pest Species of the Anastrepha fraterculus Group (Diptera: Tephritidae). **Annals of the Entomological Society of America**, v. 107, n. 2, p. 490–495, 1 mar. 2014.

PRESTWICH, G. D.; DU, G.; LAFOREST, S. How is pheromone specificity encoded in proteins **Chemical Senses**, v. 20, n. 4, p. 461–469, ago. 1995.

SALLES, L., A. Biología e ciclo de vida de Anastrepha fraterculcus (Wied.). In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Malavasi, A.; Zucchi, R.A ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

SÁNCHEZ-GRACIA, A.; AGUADÉ, M.; ROZAS, J. Patterns of nucleotide polymorphism and divergence in the odorant-binding protein genes OS-E and OS-F: analysis in the melanogaster species subgroup of Drosophila. **Genetics**, v. 165, n. 3, p. 1279–1288, nov. 2003.

SÁNCHEZ-GRACIA, A.; ROZAS, J. Divergent evolution and molecular adaptation in the Drosophila odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. **BMC Evolutionary Biology**, v. 8, n. 1, p. 323, 2008.

SÁNCHEZ-GRACIA, A.; VIEIRA, F. G.; ROZAS, J. Molecular evolution of the major chemosensory gene families in insects. **Heredity**, v. 103, n. 3, p. 208–216, set. 2009.

SATO, K. et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. **Nature**, v. 452, n. 7190, p. 1002–1006, 24 abr. 2008.

SELIVON, D. Relações com as plantas hospedeiras. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado.** Malavasi, A.; Zucchi, R.A ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

SELIVON, D.; PERONDINI, A. L. P.; MORGANTE, J. S. A Genetic–Morphological Characterization of Two Cryptic Species of the Anastrepha fraterculus Complex (Diptera: Tephritidae). **Annals of the Entomological Society of America**, v. 98, n. 3, p. 367–381, 1 maio 2005.

SIEGLAFF, D. H. et al. Comparative genomics allows the discovery of cis-regulatory elements in mosquitoes. **Proceedings of the National Academy of Sciences**, v. 106, n. 9, p. 3053–3058, 3 mar. 2009.

SMADJA, C. et al. Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, Acyrthosiphon pisum. **Molecular Biology and Evolution**, v. 26, n. 9, p. 2073–2086, 1 set. 2009.

SMITH-CALDAS, M. R. B. et al. Phylogenetic Relationships Among Species of the fraterculus Group (Anastrepha: Diptera: Tephritidae) Inferred from DNA Sequences of Mitochondrial Cytochrome Oxidase I. **Neotropical Entomology**, v. 30, n. 4, p. 565–573, dez. 2001.

SOBRINHO, I. **Evolução molecular dos genes doublesex e fruitless em moscas-das-frutas do grupo Anastrepha fraterculus (Diptera, Tephritidae) Tese (Doutorado). São Carlos. 140p**Tese (outorado), 2009.

SOLFERINI, V. N.; MORGANTE, J. S. Karyotype Study of Eight Species of *Anastrepha* (Diptera: Tephritidae). **Caryologia**, v. 40, n. 3, p. 229–241, jan. 1987.

SU, C.-Y.; MENUZ, K.; CARLSON, J. R. Olfactory Perception: Receptors, Cells, and Circuits. **Cell**, v. 139, n. 1, p. 45–59, out. 2009.

VIEIRA, F. G.; ROZAS, J. Comparative Genomics of the Odorant-Binding and Chemosensory Protein Gene Families across the Arthropoda: Origin and Evolutionary History of the Chemosensory System. **Genome Biology and Evolution**, v. 3, n. 0, p. 476–490, 13 set. 2011.

VIEIRA, F. G.; SÁNCHEZ-GRACIA, A.; ROZAS, J. Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. **Genome Biology**, v. 8, n. 11, p. R235, 2007.

VOGT, R. G.; RIDDIFORD, L. M. Pheromone binding and inactivation by moth antennae. **Nature**, v. 293, n. 5828, p. 161–163, 10 set. 1981.

VOGT, R. G.; RYBCZYNSKI, R.; LERNER, M. R. Molecular cloning and sequencing of general odorant-binding proteins GOBP1 and GOBP2 from the tobacco hawk moth Manduca sexta: comparisons with other insect OBPs and their signal peptides. **The Journal of Neuroscience: The Official Journal of the Society for Neuroscience**, v. 11, n. 10, p. 2972–2984, out. 1991.

WICHER, D. et al. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. **Nature**, v. 452, n. 7190, p. 1007–1011, 24 abr. 2008.

WILLETT, C. S. Evidence for directional selection acting on pheromone-binding proteins in the genus Choristoneura. **Molecular Biology and Evolution**, v. 17, n. 4, p. 553–562, abr. 2000.

ZHOU, X. et al. Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent Ant Species Reveals Sex-Specific Signatures of Odor Coding. **PLoS Genetics**, v. 8, n. 8, p. e1002930, 30 ago. 2012.

ZUCCHI, R. A. Taxonomia. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto: Holos-FAPESP, 2000. p. 327p.

2. Molecular evolution of Odorant Receptors proteins gene family in two closely related species of fruit flies from the Anastrepha fraterculus group

2.1 Resumo

Os receptores de odores (ORs) são das proteínas mais importantes no reconhecimento dos sinais químicos nos insetos. A percepção das moléculas de odor é um processo essencial para a sobrevivência dos indivíduos uma vez que permite a detecção de fontes de alimentos, parceiros, substratos de oviposição e predadores. Este trabalho é uma investigação sobre as proteínas envolvidas no sistema olfativo de duas espécies proximamente relacionadas de moscas-das-frutas do gênero *Anastrepha* pertencentes ao grupo fraterculus. Buscamos por cópias dos genes ORs em transcriptomas de tecidos reprodutivos e de cabeça das espécies *Anastrepha fraterculu*s e *A. obliqua*, previamente gerados, e estudamos o padrão de evolução molecular destas proteínas com o intuito de identificar genes e regiões sujeitas à seleção positiva. Encontramos um total de 38 genes OR em *A. fraterculus* e 35 em *A. obliqua*. Estes genes apresentaram um alto valor de similaridade entre ortológos destas espécies com uma identidade média de 98.15%. Encontramos também expansões gênicas nas espécies de tefrítideos estudados, as quais correspondem aos genes *Or7* e *Or33* de *Drosophila melanogaster* e evidência de seleção positiva em alguns genes, principalmente nos genes das expansões recentes. A investigação do efeito das mutações indicou que apenas uma fração destes sítios sob seleção positiva estava associada a mudanças radicais para as propriedades dos aminoácidos aqui investigadas. Quando comparamos as taxas de substituições sinônimas e não sinônimas entre as espécies de tefrítideos não houve indicação de seleção positiva, porém, nas comparações entre as espécies de *Anastrepha* encontramos evidências de seleção positiva no gene *Or7c* para o qual tem sido relatado um papel importante nos comportamentos de agregação e escolha de fruto para oviposição. O alto nível de similaridade encontrado entre as cópias supostamente homólogas de genes *OR* nestas duas espécies pode ser explicado pela recente divergência deste grupo. Os resultados de seleção positiva nas expansões podem indicar que a seleção está relacionada com a divergência em função dos diferentes membros parálogos, o que ressalta um potencial papel relevante para o gene *Or7c* para estudos de diversificação entre espécies de *Anastrepha* em função do maior nível de divergência e evidência de seleção positiva entre as duas espécies de *Anastrepha* estudadas.

Palavras chave: Receptores de odor, espécies irmãs, família gênica, analises filogenéticos, seleção positiva.

2.2 Abstract

Odorant Receptors (ORs) are one of the most important proteins involved in insect recognition of chemicals cues. This biological process is essential for survival, since allows individuals the detection of food, mates, oviposition sites and predators. This work represents a first approach to the knowledge of proteins involved in chemical perception in two closely related species fruits flies of the *Anastrepha* genus. Here we searched for odorants receptors in the transcriptomes of *Anastrepha fraterculus* and *Anastrepha obliqua* and investigate the patterns of molecular evolution and selective pressures in proteins, towards to identify genes potentially involved in species differentiation. We found a total of 38 genes for *A. fraterculus* and 35 for *A. obliqua* which corresponded with 29 putative orthologous between these two species. We found high levels of similarity among those orthologous genes with an average identity of 98.15%. We also found recent genic expansions in tephritid species, which correspond to *Or7* and *Or33* genes in *Drosophila melanogaster*. We found evidence of positive selection mainly in genes of these recent expansions, which could be associated with functional divergence in the evolution of paralogous copies on this gene family. The majority of sites with evidence of positive selection did not show radical changes in some amino acid properties here tested. When we compared synonymous to nonsynonymous substitution rates between these two species and other tephritids, we failed to find evidence of positive selection using the Ka/Ks ratio, which could be a consequence of purifying selection associated with a limitation of this method to perform more detailed analyses of individual changes amidst several conserved sites. The comparison between *Anastrepha* species reflected the high similarity between these two species, which could a consequence of their recent divergence. Despite this similarity, we found evidence of positive selection in the *Or7c* gene, which has been reported to have an important role in aggregation and oviposition site selection behavior in *D. melanogaster*, making it a very interesting gene for further studies in the differentiation not only of *A. fraterculus* and *A. obliqua*, but possibly among other species in the *Anastrepha* genus..

Key words: Odorant receptors, sister species, gene family, phylogenetic analysis, positive selection.

2.3 Introduction

Chemical perception is fundamental for survival and reproduction of any organism. In insects, the identification of odorant volatiles is essential for detecting food, oviposition sites, potential mates and predators. The first step of chemical stimuli recognition occurs in olfactory receptor neurons (ORNs) housed in the antenna and the maxillary palps. These ORNs express Odorant receptors (ORs) which are the front line of odorant detection and their task is to convert chemical signals into electrical outputs (CHESLER; FIRESTEIN, 2008).

Insect ORs are proteins with seven trans-membrane domains with about 400 amino acids which have an inverted membrane topology where the C-termini are located on the extracellular portion. ORs function as ligand-gated ion channels but the assembly of a heterodimeric receptor (the protein ORCO plus a conventional OR) is needed to confer odor detection (SATO et al., 2008; WICHER et al., 2008). These proteins are encoded by a highly diverse gene family which has a variable number of paralogous members and species-specific lineages of gene expansions. Functional gains and losses of odorant receptors have an important role in ecological and trophic shifts for some species (BOHBOT; PITTS, 2015; CAREY et al., 2010; GUO; KIM, 2007a; MCBRIDE, 2007). Likewise, discrete changes which alters the olfactory receptors specificity could be involved with reproductive isolation between closely related species (LEARY et al., 2012).

Here, we study the OR gene family in two closely related fruit flies species; *Anastrepha fraterculus* and *Anastrepha obliqua*, which are important fruit pests in South America. These species belong to the fraterculus group, a complex of closely related species, some of which that diverged recently and show high morphological similarity. *A. fraterculus* and *A. obliqua* differ in a number of morphological and behavioral traits, including host preference (CRUZ-LÓPEZ et al., 2006; GREGORIO; SANT'ANA; REDAELLI, 2010; LÓPEZ-GUILLÉN; TOLEDO; ROJAS, 2010). Even though they are both considered polyfagous, *A. fraterculus* has been associated with a wide number of host fruits mainly of the Myrtaceae family, while *A. obliqua* has been associated to a smaller number of hosts, several of those of Anacardiaceae family (SELIVON, 2000). The study of ORs genes in these two closely related species may provide clues about the differentiation and host preference of these species, which is why we seeked OR genes in the transcriptomes of head and reproductive tissues of *A. fraterculus* and *A. obliqua* and investigated patterns of molecular evolution which would

enable us to identify differentiated genes with high rates of evolution and signal of positive selection.

2.4 Materials and methods

2.4.1 Identification of A. fraterculus and A. obliqua **OR** *genes*

We searched for *OR* gene sequences in contigs inferred from transcriptomes made from reproductive and head tissues of individuals from populations of *A. fraterculus* from Southeastern (22° 01' 03" S, 47° 53' 27" 138 W) and of *A. obliqua* from Midwest (16° 41' 58" S, 49 ^{o} 16' 35" W) regions of Brazil. Details on sample preparation, molecular procedures, assembly and annotation of the RNA-seq data are described in elsewhere (REZENDE et al., 2016) but, briefly, total RNA was extracted using the TRizol/chloroform protocol (Chomczynski and Mackey 1995). RNA-seq libraries were constructed from four µg of total RNA using the TruSeq Stranded Total RNA Sample Prep Kit (Illumina) protocol, according to the manufacturer's instructions. Pools of 12 libraries were run on an Illumina HiSeq2000 on a lane with runs of 2 x 100 bp paired-end reads. All reads were trimmed for 5 quality and length with SeqyClean (available at https://bitbucket.org/izhbannikov/seqyclean), keeping only reads with a minimum sequence length of 50 pb, a minimum of 0.01 for the parameter 'max-avg-error' and 0.05 for 'max-error-at-ends', and an average Phred quality score ≥ 20. Processed reads were assembled in two independent transcriptomes, one for each species, using the Trinity short read assembler (release 2013-02-25) (GRABHERR et al. 2011), using default parameters.

We conducted BLASTp searches (ALTSCHUL et al., 1990) in the assemblies using the Gene Ontology and the *Drosophila melanogaster* ORs as a query, since it is the best curated insect genome. We also performed tBLASTn searches in the assemblies using the tephritid *Ceratitis capitata* OR genes available on GenBank (National Center for Biotechnology Information, NCBI).

We retained only the first match of each search and contigs larger than 100 amino acids with an E ≤ 10^{-5}. Open reading frames (ORFs) were inferred using the software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). To translate the nucleic acid sequences to their corresponding peptide sequences we used EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). The annotated sequences were named as *Afra* and *Aobl*, followed by the respective OR name in accordance with their greatest similarity with *D. melanogaster´s* ORs. When two or more sequences were associated with the same OR, those sequences were differentiated with an alphabetical postscript.

2.4.2 Phylogenetic analysis

The amino acid sequences of the identified ORs for *Anastrepha* species, *Ceratitis capitata*, *Bactrocera dorsalis* and the sequences for *D. melanogaster* were aligned using MAFFT v7.21(KATOH et al., 2005) and some manual adjustments were performed using BioEdit v7.2. (HALL, 1999). The alignment of the nucleotide sequences used protein alignments as guide with the program PAL2NAL (SUYAMA; TORRENTS; BORK, 2006). The best-fitted model of nucleotide substitution for the alignment, inferred with jModelTest (POSADA, 2008) using the Akaike information criteria (AIC) as model selection strategy was considered to estimate maximum likelihood phylogenetic relationships with PhyML v3.0 (GUINDON et al., 2010) and the ML tree was visualized using the program FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

We first inferred phylogenetic relationships among the identified sequences from *A. fraterculus* and *A. obliqua* and all sequences reported for *D. melanogaster* (obtained from GeneBank, Supplementary material 1), because this species has the best curated genome which served as a yardstick to corroborate our annotation and help us name the OR sequences from the species here sequenced. Subsequently, we inferred a new phylogenetic tree using the OR sequences identified for *Anastrepha* as well as the OR sequences from the tephritids *Ceratitis capitata* and *Bactrocera dorsalis* (Supplementary material 1).

2.4.3 Molecular evolution analysis

Because of the great divergence, we performed the evolutionary analyses among paralogous copies of ORs from *A. fraterculcus*, *A. obliqua* and at least one copy from *C. capitata* and/or *B. dorsalis* separating specific lineages (or subfamilies) of OR genes (supplementary material 2). We investigated patterns of molecular evolution and positive selection in each of these subfamilies through a strict branch-site test comparing model A vs. Anull model (ZHANG, J.; NIELSEN, R.; YANG, Z., 2005). For that, we used the software CODEML, implemented in PAML v.4 (YANG, 2007) to calculate the log-likelihood value (lnL) for every model. The comparison of those models was made using a Likelihood ratio test (LRT) (ANISIMOVA; BIELAWSKI; YANG, 2001). Significant results were determined with X^2 tests and Bonferroni correction for the instances in which several branches were tested. When there was significant evidence of positive selection, we used the Bayes empirical Bayes (BEB) method (YANG; WONG; NIELSEN, 2005) to calculate the posterior probabilities for positive selection on each individual codon in the alignment under the A model. Sites with evidence positive selection were mapped onto the receptor's topology predicted with the software TOPCONS (TSIRIGOS et al., 2015) and then diagrams of the 2D receptors structure were created using TOPO2 (http://www.sacs.ucsf.edu/TOPO2/).

We also investigated, on the gene expansions here identified, whether changes in the amino acid level were associated with conserved or altered biochemical properties with the PRIME (PRoperty Informed Model of Evolution) analysis available on the Datamonkey server (http://www.datamonkey.org/), which considered the neighbor-joining trees described in Supplementary Material 3. Positively selected sites were once again mapped onto the protein´s topology to evaluate their distribution.

With the goal of identifying genes with high evolutionary rates and potential evidence of positive selection between *A .fraterculus* and *A. obliqua*, we used KaKs_Calculator (ZHANG et al., 2006) with the MS model (POSADA, 2003) to estimate pairwise Ka, Ks rates and the Ka/Ks ratio for the putatively orthologous OR genes that were identified in full length from these two species. We also estimate the same parameters from alignments of *Anastrepha* OR sequences to other putatively orthologous Tephritid copies from *C. capitata* and *B. dorsalis.*

2.5 Results and Discussion

*2.5.1 Olfactory receptors genes in **Anastrepha fraterculus** and **Anastrepha obliqua***

We identified a similar number of OR sequences in the two species: 38 sequences in *A. fraterculus* and 35 in *A. obliqua* (Supplementary Material 3), which corresponded to 17 *OR* genes in *D. melanogaster*. Among the 38 *AfraOr* genes, there were 18 intact genes and 20 partial genes, whereas among the 35 *AoblOr* genes, there were 13 intact genes and 22 partial genes. The average lengths (amino acids) of the fragments found in *A. fraterculus* were 248 (±90.8) while in *A. obliqua* were 197 (±96). That indicates that in most cases we were able to recover more than half of the total length of the OR when the sequence was not complete (Table1). For the two species the lengths of the complete OR genes varied from 385 to 474 amino acids being *Orco* the biggest identified gene.

If we compare with other species of dipterans, the number of OR genes identified in the *Anastrepha* species was smaller. The repertoire of OR genes ranged from 55 genes in *Drosophila persimilis* to 72 in *Drosophila grimshawi* (SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009), whereas other species of *Tephritidae*, such as *B. dorsalis* and *C. capitata* were reported to have 64 and 62 OR genes respectively (GenBank-National Center for Biotechnology Information, NCBI). The identification of a smaller number of sequences in *Anastrepha* was expected, not only because we searched for these genes only in transcriptome libraries of reproductive and head tissues, but particularly considering the normally low levels of expression of these genes (CLYNE et al., 1999).

Table 1. Fragment lengths in amino acids of ORs genes found in *A. fraterculus* and *A. obliqua* transcriptomes and the size reported for their corresponding putative orthologous gene in *D. melanogaster* and/or *C. capitata.*

	A. fraterculus	***A. obliqua***	***D. melanogaster***	***C. capitata***
Or3a	414	414	+	385
Or4a	296†	+	375	+
Or5a	325†	277†	+	336
Or7b	172†	126†	+	376
Or7c	393	393	+	394
Or7d	396	396	+	396
Or7e	451	451	+	454
Or7f	456	456	+	451
Or10a	402	402	406	395
Or13a	270†	315†	418	446
Or19a	100†	+	387	+
Or22c	246†	+	402	+
Or23a	392	392	+	+
Or24a	364†	314†	398	398
Or33e	385†	208†	+	378
Or33f	+	198†	+	+
Or33g	389	389	+	388
Or33h	385	385	+	385
Or43a	113†	123†	376	379
Or45a	115†	142†	378	389
Or49a	270†	125†	396	394
Or49b	273†	191†	375	+
Or53a	112†	+	+	+
Or54a	297†	155†	+	+
Or59d	394	+	+	336
Or63a	415	+	420	+
Or63b	+	138†	+	+
Or63c	417†	160†	+	+
Or67c	+	185†	+	+
Or67e	386	386	+	+
Or67f	356†	387	+	+
Or69c	424	319†	+	+
Or82a	204†	186†	385	+
Or83a	+	129†	453	+
Or83c	418	278†	397	+
Or83e	276†	192†	+	471
Or85d	172†	+	412	390
Or85e	232†	+	417	449
Or87a	405	293†	+	+
Or89a	+	100†	401	410
Or94a	382†	+	+	+

Or95a	+	183[†]		
Or95b	384	384	+	+
Orco	473	473	486	473
Average(Sd)	**325.1 (±107.5)**	**275.5 (±118.9)**		

[†]Incomplete genes. [+]Orthologous gene not identified for the species. The lengths reported in this table do not consider stop codons.

We identified a total of 29 putative orthologous genes between the *Anastrepha* species here studied (Figure 1 and Table 1). We believe that there may be more orthologous genes between these two closely related species since they have diverged recently and for the reason previously considered we probably failed to identify the complete repertoire of OR genes for each species. Other analyses that compared species of the same genus have identified a high number of orthologous genes; for example, a total of 59 orthologous genes were identified amongst the species in the OR repertoires of 12 *Drosophila* species (GUO; KIM, 2007b).

The average identity percentage among the intact genes within species was 17.2% in *A. fraterculus* and 19.4% in *A. obliqua*; values that ranged from 9.9% to 50.64% in *A. fraterculus* and from 11.2% to 51.91% in *A. obliqua*. The high divergence found among the paralogous copies is a characteristic hallmark of this gene family and is related with a functional diversification and ecological specialization of each gene in the family (MCBRIDE; ARGUELLO; O'MEARA, 2007). This high level of divergence also suggests that this gene family is old, or at least some of their genes, since for protein families sequence similarity is expected to decrease exponentially in time (LUZ; STAUB; VINGRON, 2006).

A comparison of the set of complete orthologous genes between *A. fraterculus* and *A. obliqua* reveals a high percent of identity between the species (Table 2). The average identity between putatively orthologous copies from these different species was 98.15%, with the *Or7c* being the most divergent gene (94.1%), while *Orco* was the most conserved, showing the same sequence in both species. The high levels of genetic identity between the odorant receptors of these species may reflect their recent divergence (VANÍČKOVÁ et al., 2015) and the potentially homologous function of these genes.

Table 2. Percent identity between complete orthologous genes identified in *A. fraterculus* and *A. obliqua* transcriptomes.

Orthologous	Identity (%)
Or3a	96.62
Or7c	94.91
Or7d	98.23
Or7e	98.45
Or7f	98.68
Or10a	99.25
Or23a	99.48
Or33g	99.49
Or33h	96.88
Or67e	97.41
Or95b	98.44
Orco	100
Average	**98.15**

2.5.2 Phylogenetic relationships among OR genes

The phylogenetic analysis of *D. melanogaster* and *Anastrepha* OR genes (Figure 1) produced a tree in which we can observe five main lineages that have variable number of sub-lineages or subfamilies, and genes. The deepest, possibly the most ancient lineage connects all Orco genes, which are the most conservative gene in the family and essential for all OR proteins ability to connect to odors. The second more ancient lineage also connects only a few related *OR* genes, two copies per species. The other lineages show that more recent lineages have increasing number of members, suggesting that the most recent lineages show a stronger pattern the evolutionary dynamic of the birth-and-death process in which a large number of gene gains, gene losses and pseudogenization events are involved in the evolution of this gene family (EIRÍN-LÓPEZ et al., 2012). This analysis also showed 10 putatively orthologous genes among species of *Anastrepha* and *D. melanogaster* (Figure 1); a number that could be higher if we consider that there are some genes that were found in *A. fraterculus* but not in *A. obliqua* that have orthologous genes with *D melanogaster* (i.e., *AfraOr19a*, *AfraOr22c*, *AfraOr59d*).

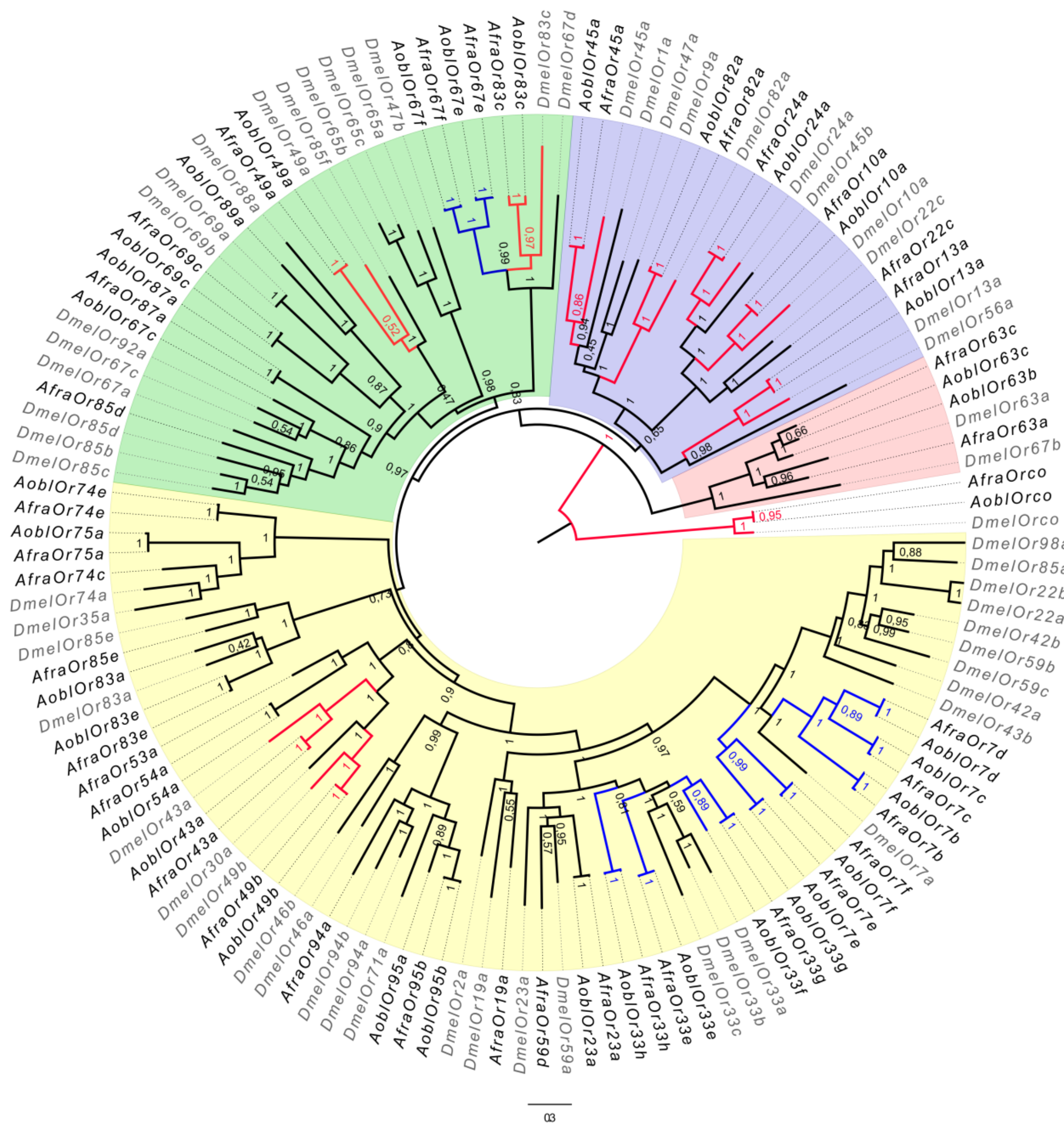


Figure 1. Phylogenetic relationships between *D. melanogaste*r (DmelOr) OR genes and the genes identified in *A. fraterculus* (*AfraOr*) and *A. obliqua* (*AoblOr*). Branch supports (Bayes aLRT) are shown in each node. Red branches represent orthologous genes between three species and blue branches represent genes in *A. fraterculus* and *A. Obliqua* that do not have associated genes with *D. melanogaste*r.

The most recent lineage is the largest and shows, among the sub-lineages, some gene expansions in *Anastrepha* species that do not have equivalent orthologous genes in *D. melanogaster*, such as the expansion of *Or7* genes (*Or7b, Or7c, Or7d, Or7e* and *Or7f*), Or33 genes (*Or33e, Or33f, Or33g* and *Or33h*). These expansions could represent important genes in the *Anastrepha* group and could be related to the detection of specific chemical stimuli relevant in the ecological niche of this group as is, for example, host plant selection. This is even more relevant if we consider that we failed to identify several copies of OR that we believe would be present in *Anastrepha* because our analysis was restricted to head and reproductive transcriptomes.

We performed a more comprehensive phylogenetic analysis of the OR genes in Tephritids encompassing the OR genes identified in *A. fraterculus* and *A. obliqua* with the reported OR genes of *C. capitata* and *B. dorsalis*. Such analysis showed a total of 15 putative orthologous genes shared among the four species (Figure 2). Among these 15 genes there are seven genes that also have orthologous copies in *D. melanogaster*. This result supports the hypothesis that, there seems to be a core set of receptors that would be more or less conserved across several different lineages and might respond to general odors required for said level, whereas there are other genes that are idiosyncratic to a particular subgroup and even species. This hypothesis is supported by the observation that the number of orthologous groups including representatives of all surveyed species gradually decreases with increasing divergence time, product of the birth-and-death model of evolution (EIRÍN-LÓPEZ et al., 2012).

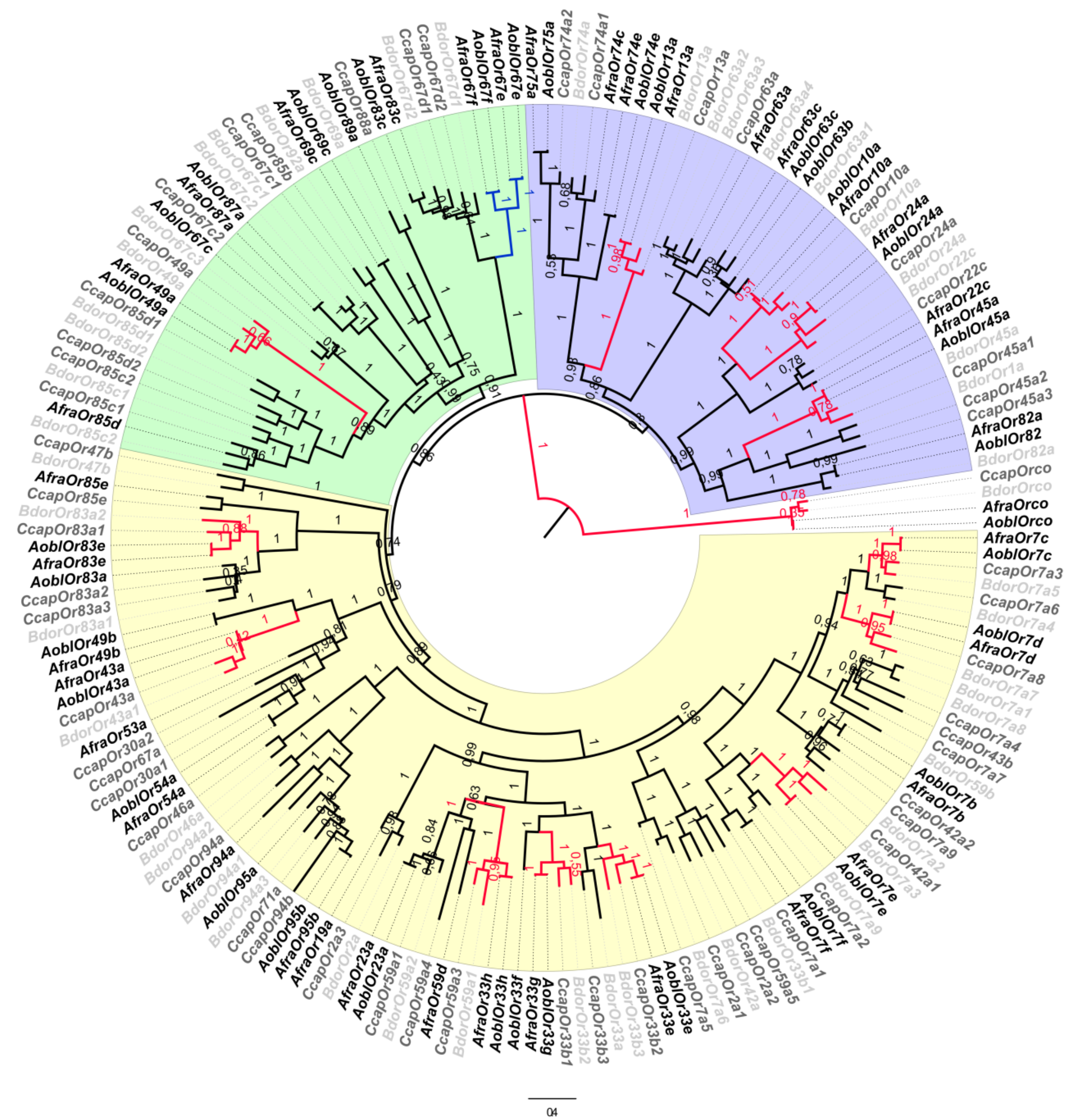


Figure 2. Phylogenetic relationships between *C. capitata* (*CcapOr*), *B. dorsalis* (*BdorOr*) *A. fraterculus* (*AfraOr*) and *A. obliqua* (*AoblOr*) OR genes. Branch supports (Bayes aLRT) are shown in each node. Red branches represent orthologous genes between the four species and the blue branch represent genes with no orthologous in *C. capitata* and *B. dorsalis*.

Since we found 9 genes in *A. fraterculus* that were not found in *A. obliqua* and 6 genes in *A. obliqua* that were not found in *A. fraterculus*, we looked for potential orthologous genes in other tephritids species used in this study, with the purpose of identifying potential species-specific genes in the *Anastrepha* species. It would be important to do so particularly because, as discussed previously, we used transcriptome data, which would provide an incomplete source of information about the total number of ORs. So, if the gene is absent in one species of *Anastrepha*, but present in its ancestors, it makes it more likely that said absence might be more a question of differences in expression level, rather than the actual gene being absent. We found that among the genes identified only in *A. fraterculus*, only the gene *AfraOr53a* does not seem to have a putatively orthologous in other *Tephritidae* species (*C. capitata* and *B. dorsalis*) or *D. melanogaster*, whereas in *A. obliqua*, we failed to find clear putatively orthologous for the genes *AoblOr63b* and *AoblOr95a* on other species, though the latter may be related to *Or94*. We consider that these genes (*AfraOr53a*, *AoblOr63b* and *AoblOr95a*) actually do not represent species-specific genes since they were not found in their full length, which could be a consequence of differential expression between the species.

In the phylogenetic tree of tephritid ORs (Figure 2) we found four main lineages with variable number of subfamilies and number of genes as was found in the previous inference that only compared *Anastrepha*'s ORs with *D. melanogaster*'s. Indeed, the topologies of both trees follow a similar pattern with the exception that the second deepest lineage found in the *D. melanogaster* and *Anastrepha* phylogeny is located within the second deepest lineage of the tephritids tree. Likewise, we detected a general trend of increased number of genes with the more recent lineages, but there is an exception to that pattern in the second oldest lineage. The number of genes in these four lineages varied from 35 in the third lineage (green) to 96 in the fourth (yellow), which is the largest lineage. This lineage has putative homologous copies of the *Or7* and *Or33* set of genes in *Anastrepha*, *C. capitata* and *B. dorsalis*. In general, it is noteworthy that even though our analysis was based only on transcriptome data, it has captured the great majority of lineages represented in *Drosophila* and other Tephritidae. Nonetheless, we also found some other genes associated with these expansions that were not identified in Anastrepha species, for example the orthologous groups *BdorOr7a9*/*CcapOr7a2*, *BdorOr7a4*/*CcapOr7a6* and *BdorOr33a*/*CcapOr33b3*, indicating that there probably are more genes of these expansions that were not identified in the *Anastrepha* species, possibly as a consequence of the incomplete representation of the transcriptomes.

The *Or7* gene expansion found in *Tephritidae* species is an interesting group of genes. The gene *Or7a*, found as a single copy in different species of *Drosophila*, codes for a OR which responds to many alcohols and aldehydes odors (HALLEM; CARLSON, 2006). These chemical complexes are some of the main components of fruit volatile compounds (EL HADI et al., 2013). Even though *A. fraterculus* and *A. obliqua* are considered oligotrophic, they have some host specificity. A study of chemical perception in *A. fraterculus* in which were measure the electrophysiological responses to several fruits extracts, shows that for this species, the largest electrophysiological responses were recorded for *Campomanesia xanthocarpa* (Myrtaceae) and *Psidium cattleianum* (Myrtaceae) (GREGORIO; SANT'ANA; REDAELLI, 2010). Whereas studies of attraction of volatiles from fruits shows that *A. obliqua* is strongly attracted to extracts from Anacardiaceae fruits *Spondias mombin* and *Spondias purpurea* L (CRUZ-LÓPEZ et al., 2006; LÓPEZ-GUILLÉN; TOLEDO; ROJAS, 2010), so these species' diversification may be associated with host preference and its molecular cues. Furthermore recent data has indicated that pheromone activation of these receptor (*Or7c*) in *Drosophila* promotes aggregation and modulates female oviposition site selection (LIN et al., 2015) making these genes interesting candidates for further studies which that would assess their role in the behavioral differences in host preference in *Anastrepha* flies.

2.5.3 Analysis of positive selection in ORs

We used the *Tephritidae* phylogenetic tree as a guideline to divide the ORs in subfamilies which are separated by distinct evolutionary branches to investigate rates and patterns of evolution in theses clusters of complete *AfraOr* and *AoblOr* genes. This analysis allowed us to evaluate the clusters of orthologous genes with higher evolutionary rates to investigate the occurrence of positive selection. We tested a total 11 groups of orthologous genes which were divided in six subfamilies, since we further subdivided two of the main lineages previously identified among Tephritidae because of the large genetic distance among them as well. Eight of these tested orthologous groups were in the largest and most recent lineage of the tephritid OR phylogeny. In supplementary Material 2 are shown the subdivisions and the lineages tested for positive selection.

We used the software codeml, which calculates the statistical distribution of the nonsynonymous to synonymous substitution rate ratio (dN/dS or ω) among sites and branches in the phylogeny, to detect evidence of positive selection affecting some lineages or sites. After the LRT comparison of the null and alternative models, we detected positive selection in

the lineages associated with *Or3a*, *Or10a*, *Or7d*, *Or7e Or33*g and the *Or33h* sequences (Table 3). As it has been shown by other studies which reported ORs genes evolving under positive selection (ENGSONTIA et al., 2014; GUO; KIM, 2007b; SMADJA et al., 2009), most of them are genes of recent expansions. Our results indicate that selection is acting mainly in recently duplicated copies of tephritids ORs, supporting the hypothesis raised in other studies that genic duplications are important to provide new olfactory functions (ENGSONTIA et al., 2014, 2015; GUO; KIM, 2007b; ROBERTSON; WANNER, 2006; SMADJA et al., 2009; ZHOU et al., 2015) since these new copies undergone faster evolution driven by positive selection accelerating the fixation of advantageous mutations that may have enhances the activity of the new function (KONDRASHOV et al., 2002; SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009; TAYLOR; RAES, 2004; ZHANG, 2003).

Table 3. Positively selection detected on groups of paralogous genes among different *Tephritidae* species.

Gene	LTR[1]	Positively selected sites[2]
Or3a	10.780742*	58, 68, 169, 199, 208, 270, 285, 451
Or10a	6.655330*	68, 72, 144, 223, 250, 283, 368, 392, 409
Or33g	7.399754*	171, 322
Or33h	7.522716*	11, 28, 41, 48, 53, 81, 85, 88, 90
Or7d	15.364004**	119, 129, 135, 170, 212, 261, 294, 310, 339, 505
Or7e	19. 939050**	144, 219, 377, 458

[1]Likelihood Ratio Test. [2]Corresponds with the site in the general alignment used for the test. *$p<0.01$; **$p<0.001$.

After detecting branches with evidence of positive selection, we used the Bayes Empirical Bayes (BEB) analysis, which calculates the posterior probabilities for site classes under the alternative model, to detect sites that experience positive selection ($\omega>1$) in the alignment. We consider as positively selected the sites with posterior probabilities ≥ 95%. This analysis showed that most of the sites in the evaluated OR genes evolve under strong purifying selection, nevertheless, we identified several sites evolving under positive selection (Figure 3). When we investigated the distribution of the positively selected sites on the

predicted protein structure, we found that these sites are found in extracellular, trans-membrane and intracellular regions (Figure 4).

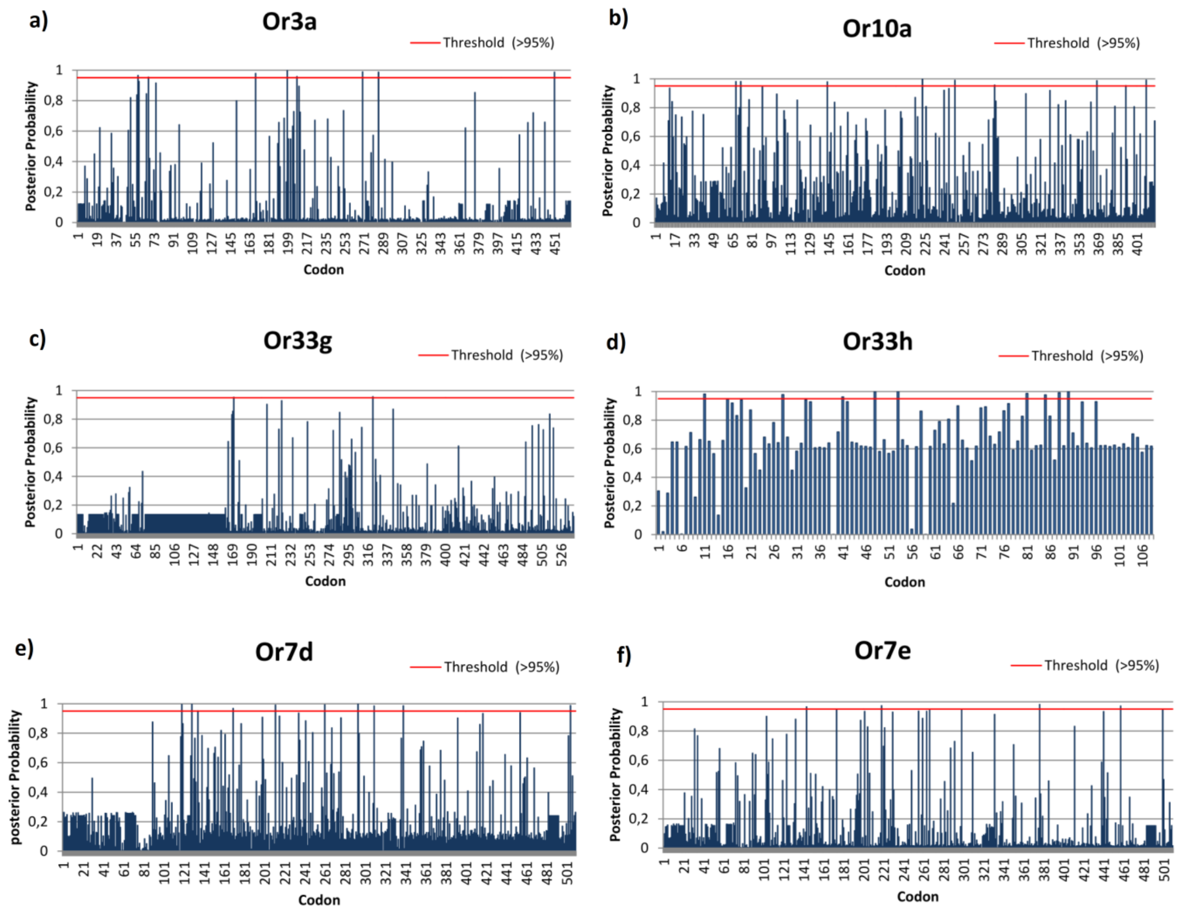


Figure 3. Results of Bayes Empirical Bayes (BEB) showing the posterior probabilities of every site to belong to the positive selection site class (ω>1). Dashed red line indicates the posterior probability threshold of 0.95. BEB results for the a) *Or3a* orthologous cluster; b) *Or10a* orthologous cluster; c) *Or33g* orthologous cluster; d) *Or33h* orthologous cluster; e) *Or7d* orthologous cluster; e) *Or7e* orthologous cluster.

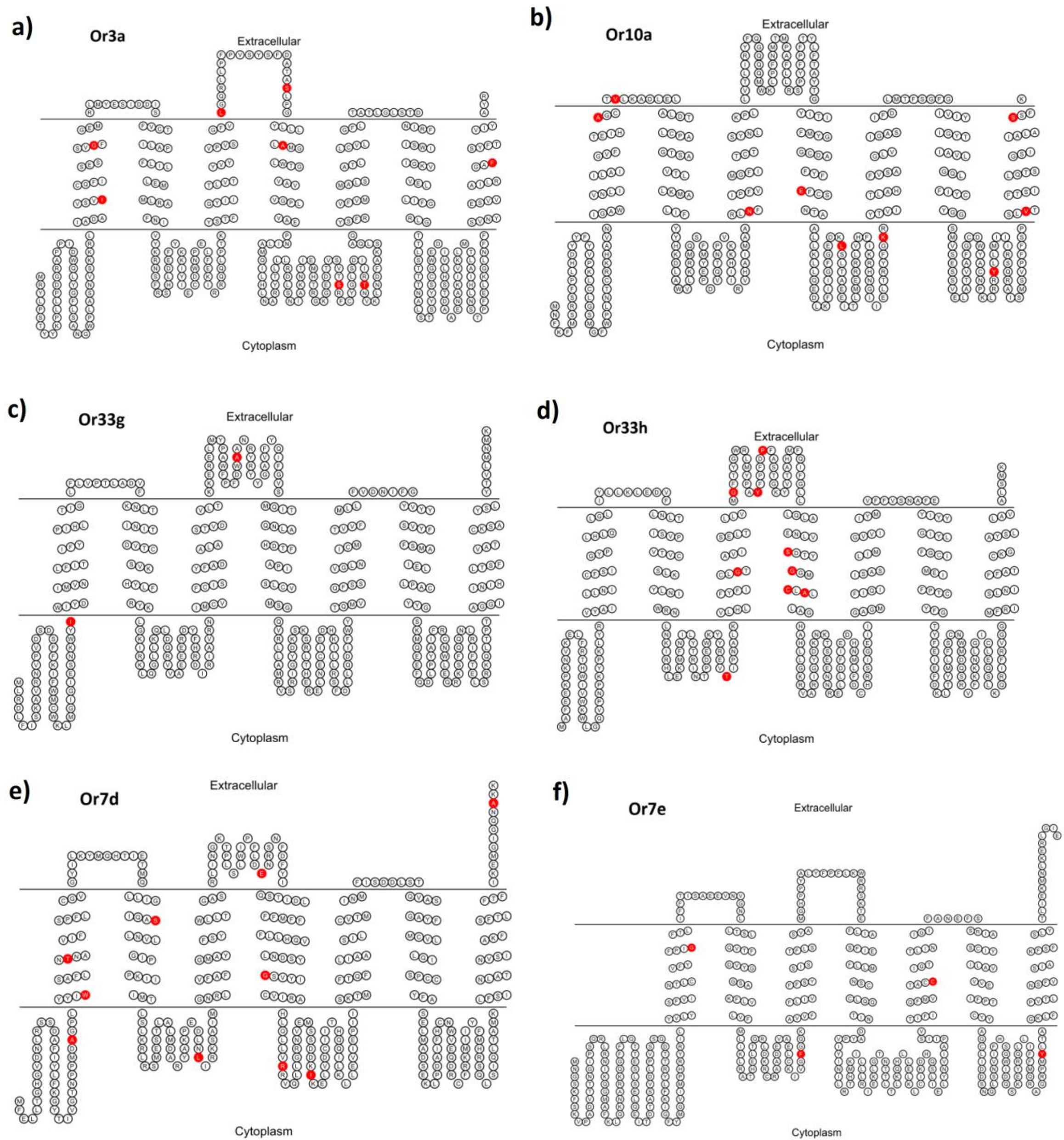


Figure 4. Predicted topology of OR proteins and distribution of positively selected sites. Red dots indicate positively selected sites in trans-membrane topology predicted by TOPCONS software. Proteins have 7 trans-membrane domains and an intracellular N terminus characteristic of these receptors. a) *Or3a* predicted topology; b) *Or10a* predicted topology; c) *Or33g* predicted topology; d) *Or33h* predicted topology; e) *Or7d* predicted topology; e) *Or7e* predicted topology.

For the gene *Or3a* (Figure 4a) we found eight positively selected sites, two of them located on the second extracellular loop (2EL), two on the first transmembrane domain (1TMD), one on the fourth and seventh TMD and three on the third intracellular loop (3IL). On the gene *Or10a* (Figure 4b), we found nine positive selected sites, being one on the first EL, one on the first, third and fourth TMD, two on 3IL and one on 4IL. The *Or33g* (Figure 4c) was the gene with the least amount of positive selected sites, having only one on the NH_2-tail and one on 2EL. The gene *Or33h* (Figure 4d) had nine positively selected sites, three on 2EL, one on 3TMD, four on 4TMD and one on 2IL. The gene with most positively selected sites (10) was *Or7d*, which had one each on 2EL,COOH-tail, 2TMD, 4TMD, NH2-tail, 2IL and two on 3EL and 1TMD. In the gene *Or7e* (Figure 4f) we found four positively selected sites distributed one each on 1TMD, 5TMD, 2IL and 4IL.

We also evaluated if, in general, the distribution of positively selected sites follows a pattern or were over represented in some of the main protein regions. The proportion of positively selected sites in trans-membrane regions was 47.6%, as well as, the proportion of positive selected sites in extracellular and intracellular regions was 21.4% and 30.9%, respectively. The number of candidate sites was higher in the trans-membrane region but not significantly different than expected in a homogenous model of distribution (taking account the size of every section of the protein) (X^2 test P = 0.106).

Insect olfaction requires the expression in the olfactory sensory neurons of two *OR* genes; one variable odorant binding subunit (OrX) and one constant subunit named *Orco* (BENTON, 2006). These proteins are seven transmembrane domains receptors whose topology have an orientation in which the N-terminus is located in the intracellular portion and the C-terminus is located in the extracellular portion (BENTON et al., 2006). Functional insect OR proteins consist of a heterodimeric complex (unknown stoichiometry) of these two expressed ORs which form a ligand-activated non selective ion channel (Sato et al., 2008; Wicher et al., 2008). Even though the precise mechanisms of signal transduction of these receptors remain unclear (JONES et al., 2011), the evolution of gene copies in this gene family may be explained by exploring their dual function as a receptor, as well as, an ion channel. Recent data on evolutionary couplings across the large insect OR family was used to build *de novo* the first three-dimensional models of these proteins and predict odor binding and ion conducting domains, providing the best structure-activity dissection of the OR proteins to date (HOPF et al., 2015).

These results suggest that the external half of the trans-membrane domains (TM) 2, 3 and 4 comprises a part of the ligand-binding pocket of these proteins and we believe that positively selected sites found in these regions, observed on genes *Or3a*, *Or10*a, *Or33h* and *Or7d*, could influence the detention of chemical signals promoting changes in the protein leading to the detections of new chemicals signs in the evolution of these groups of orthologous receptors. This furthermore suggests that the intracellular loop 3 (IL3) forms at least part of the interface for the assembly of the *Orco* and the ligand specific *Or* into the heteromeric complex and the formation of the ion channel, positively selected sites found in this region, which we detected on genes *Or3a*, *Or10a* and *Or7d*, could be involved in the interactions between the two subunits in the complex.

Upon mapping point mutations that may have direct or indirect effects on channel functions in OR proteins (HOPF et al., 2015), we also found that the trans-membrane helices 5, 6 and 7 are involved in the conformation of the central part of the ion conducting channel. There are other examples of heterodimeric ligand-gated ion channels in which trans-membrane helices of each subunit contribute to formation of the core of the ion channel (KARAKAS; FURUKAWA, 2014; TRAYNELIS et al., 2010) and has been found that mutations in this core could cause changes in membrane conductance and the depolarized resting potential of the cells (ZUO et al., 1997). Therefore, positively selected sites on these trans-membrane regions (*Or3a*, *Or7e* and *Or10a*) could be related with roles in relaying the signal generated upon ligand binding to the receptor (LEE, 1997).

Though we identified sites evolving under positive selection, we also wanted to investigate whether the changes brought forth by these changes also led to radical changes in amino acid attributes that might radically change protein’s structure. Results of the PRIME analysis, in which we evaluated if changes at the amino acid level have been associated with any radical change of biochemical properties in the evolution of the two main gene expansions, shows that the 36 positively selected sites identified in the *Or7* expansion was associated with 36 conserved and 8 changing properties. Similar results were found for the *Or33* expansion which shows 32 positively selected sites with 30 conserved and 8 changing properties (Table 4). The reason why there are more properties than sites is because some sites were associated with more than one property change. Among the conserved property categories, volume (19 sites for the *Or7a* expansion and 8 sites for the *Or33* expansion) and isoelectric point (10 sites for the *Or7a* expansion and 11 sites for the *Or33* expansion) were the most overrepresented.

Table 4. Positively selected sites identified by PRIME.

	Properties	Or7 expansion	Or33 expansion
Conserved	Chemical Composition	142†, 250, 265, 301	65, 89, 256, 291, 318
	Polarity	330	122, 162, 236, 273†, 415
	Volume	41, 46, 135†, 136†, 142†, 178, 186, 187, 208†, 254, 309, 310, 329, 356†, 396, 397, 399, 451, 459	55, 71, 79, 93†, 128†, 171, 191, 380
	Iso-electric point	97, 104, 166, 208†, 231, 278, 320, 356†, 385, 438	58, 69, 113, 137, 163, 166†, 209, 233, 247, 269, 299,
	Hydropathy	87, 337	93†
Changing	Chemical Composition		79†
	Polarity	136†, 208†	
	Volume	142†	166†, 234
	Iso-electric point	27, 46†, 92	93†, 128†, 143
	Hydropathy	135†, 263	117, 273†

†Sites with more than one change in biochemical properties.

This tendency to conserve biochemical characteristics indicates that proteins expressed from genes of the two expansions change under these biochemical constraints. The volume each residue occupies in a protein contributes with its folding and, therefore, its final structure, leading to steric constraints in the evolution of proteins (GERSTEIN; SONNHAMMER; CHOTHIA, 1994). On the other hand, the isoelectric point determines the protein charge, being important for solubility and protein interactions. These, along with the other properties here investigated, provide an array of properties that influence protein structure. As we see in Table 4, the proportion of sites with radical changes (changing properties) is only a fraction of the positively selected sites, showing that the majority of positively selected sites have a tendency to conserve chemical properties but in some sites these changes led to radical changes in protein properties. Because of that, some of the

conserved changes here considered may actually be radical to a different property not here investigated, so we cannot state that these changes have not selective impact. In fact, even synonymous changes may have a selective impact, when we consider the effect of codon bias on levels of expression (KURLAND, 1991; QUAX et al., 2015). Interestingly, the property with most radical changes in both expansions was isoelectric point, which also showed the highest number of conserved changed, which is also the case for the volume. Even though not too many of the positively selected sites are associated with radical changes, these changes may be related with important changes in protein charge and interaction which may enable the detection of new odor molecules.

Although the nature of conserved and radical amino acid properties is similar between the expansions, the distribution of the affected sites in the topology is different in these two categories. For example, the distribution of the sites with changing properties on the *Or7* expansion differs from the distribution of sites with changing properties on the *Or33* expansion. On Figure 5 we can observe that sites with changing properties on the evolution of the *Or7* expansion are distributed mainly on trans-membrane domains, while in the *Or33* expansion, these sites are more widely distributed in the intracellular, trans-membrane and extracellular regions. These different patterns of distribution in property changing sites could indicate that every expansion of the OR genes family has its own evolutionary dynamic despite sharing some common biochemical constraints.

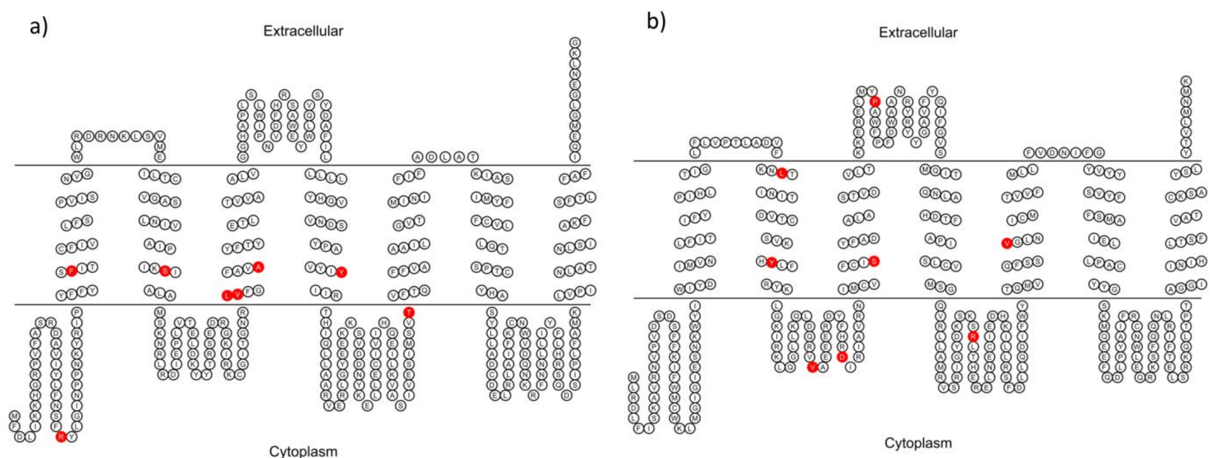


Figure 5. Positively selected sites with changing properties on the OR gene expansions. a) Positively selected sites on the *Or7* expansion plotted on the topology of the *Or7c* gene. b) Positively selected sites on the *Or33* expansion plotted on the topology of the *Or33g* gene.

We also investigated rates of synonymous and non-synonymous substitutions between complete putative orthologous genes identified in *A. fraterculus* and *A. obliqua* to help us identify genes with high rates of differentiation and signs of positive selection between these two closely related species. The Ka/Ks analysis revealed just one pair of homologous copies with Ka/Ks values above 0.5 (Table 5), *Or7c* with a Ka/Ks value of 0.9. The gene *Or7c* may represent an interesting gene for further studies to evaluate whether it could be involved in the species' differentiation since it is part of the most recent gene expansion in the evolution of ORs, being one of the most recent genes, as we can infer from the phylogenetic tree, was the most differentiated gene in this family among the two species and had signs of positive selection, suggesting that their evolution may have been a consequence of a recent duplication and differentiation shaped by selection. Furthermore, its putative orthologous copy from *D. melanogaster* is related with the reception of plant volatiles and female host plant selection behavior.

Table 5. Pairwise estimates of Ka, Ks, and Ka/Ks ratio between *A. fraterculus* and *A. obliqua* complete orthologous *OR* genes identified in transcriptomes.

Compared orthologous	Ka	Ks	Pairwise Ka/Ks	P-Value(Fisher)
AoblOrco-AfraOrco	2.39E+0	0.02386	0.00100	2.13E+00
AoblOr33g-AfraOr33g	0.00268	0.07106	0.03774	2.96E-05
AoblOr23a-AfraOr23a	0.00236	0.04318	0.05481	2.04E-02
AfraOr95b-AoblOr95b	0.00787	0.08524	0.09238	2.22E-04
AoblOr7f-AfraOr7f	0.00686	0.04576	0.14992	2.72E-01
AfraOr7e-AoblOr7e	0.00771	0.05008	0.15412	2.10E-01
AfraOr33h-AoblOr33h	0.01461	0.06473	0.22580	6.04E+00
AfraOr10a-AoblOr10a	0.00329	0.01442	0.22866	0.012036
AfraOr67e-AoblOr67e	0.01445	0.05323	0.27156	0.000290
AoblOr3a-AfraOr3a	0.01625	0.04556	0.35668	0.003782
AfraOr7d-AoblOr7d	0.01043	0.02738	0.38100	0.050479
AfraOr7c-AoblOr7c	0.02580	0.02838	0.90884	0.670886

We failed to find high Ka/Ks ratio values in all pairwise comparisons among *Anastrepha* species, *C. capitata* and *B. dorsalis* (Table 6). This result may be a consequence of strong purifying selection in the evolution of some regions of the genes, as it seems to be the case for the evolutionarily conserved *Orco*, but it might be a consequence of the reduced number of changes as a consequence of the recent time of divergence among these species. Even though

we detected higher Ks values in most pairwise comparisons between more distantly related species, an analysis of substitution saturation suggests that there might be saturation only on the most divergent contrasts to the fastest evolving genes (Supplementary Material 4). On the other hand, it is also possible that this may reflect the limitations of the Ka/Ks ratio, which is a general measure of the average impact of selection over the whole gene, to detect individual, or a limited number of, sites under selection in the midst of an evolutionarily conserved gene, which has been suggested by our results, even when we investigate more distantly related taxa.

Table 6. Paired estimates of Ka, Ks, and Ka/Ks ratio for the complete orthologous genes of *A. fraterculus* and *A. obliqua* compared with *C. capitata* and *B. dorsalis* orthologous.

		C. capitata				***B. dorsalis***			
		Ka	**Ks**	**Pairwise Ka/Ks**	**P-Value (Fisher)**	**Ka**	**Ks**	**Pairwise Ka/Ks**	**P-Value (Fisher)**
A. fraterculus	***Orco***	0.02	0.91	0.026	9.6E-113	0.02	0.91	0.026	1.5E-110
	Or33g	0.13	2.85	0.046	1.3E-187	0.16	1.26	0.126	6.7E-94
	Or33h	0.15	2.74	0.056	0.0E+00	0.30	1.47	**0.207**	6.1E-89
	Or10a	0.07	1.34	0.058	8.7E-115	0.10	0.82	0.126	7.0E-69
	Or7c	0.11	1.68	0.068	2.7E-133	0.11	1.50	0.075	9.3E-123
	Or7e	0.32	2.77	0.116	0.0E+00	0.19	1.76	0.107	4.7E-107
	Or23a	0.12	0.73	0.166	1.9E-30	0.13	1.02	0.124	6.0E-49
	Or7d	0.21	1.29	**0.168**	6.8E-81	0.27	1.69	0.158	3.1E-110
A. obliqua	***Orco***	0.02	1.11	0.018	1.8E-122	0.02	0.88	0.027	1.8E-108
	Or33g	0.13	2.86	0.046	9.5E-183	0.16	1.19	0.132	6.7E-87
	Or33h	0.15	2.69	0.055	4.4E-192	0.30	1.52	**0.199**	6.0E-95
	Or10a	0.08	1.30	0.061	1.2E-109	0.10	0.80	0.131	5.6E-67
	Or7c	0.10	1.37	0.076	2.4E-110	0.12	1.35	0.086	3.4E-111
	Or7e	0.32	2.82	0.114	0.E+00	0.20	2.00	0.099	2.6E-111
	Or23a	0.11	0.79	0.150	1.5E-34	0.14	1.05	0.129	9.8E-49
	Or7d	0.21	1.33	**0.160**	2.3E-85	0.27	1.93	0.138	3.3E-130

2.6 Conclusions

In this study we identified OR genes in *A. fraterculus* and *A. obliqua* transcriptomes to analyze for the first time the molecular evolution of the Tephritid odorant receptors. For *A. fraterculus* we identified 38 OR genes while we identified 35 for *A. obliqua* and the 29 putatively orthologous copies among these species showed high levels of similarity reflecting their recent divergence. These copies correspond to 16 copies putatively homologous to *D. melanogaster* and another 33 to other Tephritidae. A phylogenetic analysis indicated that there are a couple of recent gene expansions in Tephritidae affecting the genes *Or7* and *Or33*, and detected presence of some sites showing evidence of positive selection in these families, as well as a few others. We detected positive selection acting mainly on recently duplicated OR genes of tephritids. Sites that have evidence of positive selection were distributed homogeneously in extracellular, transmembrane and intracellular regions of the proteins. Among the *Anastrepha* OR genes here identified, the most differentiated gene was the *Or7c* which also shows evidence of positive selection indicating that this gene could be a potentially interesting gene to be investigated in the study of species differences of the species here studied, *A. fraterculus* and *A. obliqua*. These results provide a detailed picture of the general patterns influencing the evolution of this important gene family, as well as indicate several potentially interesting genes that could be involved with species differentiation not only between *A. fraterculus* and *A. obliqua*, which were here investigated, but among other species in the *Anastrepha* genus as well.

2.7 References

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

ANISIMOVA, M.; BIELAWSKI, J. P.; YANG, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. **Molecular Biology and Evolution**, v. 18, n. 8, p. 1585–1592, ago. 2001.

BENTON, R. et al. Atypical Membrane Topology and Heteromeric Function of Drosophila Odorant Receptors In Vivo. **PLoS Biology**, v. 4, n. 2, p. e20, 17 jan. 2006.

BENTON, R. On the ORigin of smell: odorant receptors in insects. **Cellular and molecular life sciences: CMLS**, v. 63, n. 14, p. 1579–1585, jul. 2006.

BOHBOT, J. D.; PITTS, R. J. The narrowing olfactory landscape of insect odorant receptors. **Frontiers in Ecology and Evolution**, v. 3, 10 abr. 2015.

CAREY, A. F. et al. Odorant reception in the malaria mosquito Anopheles gambiae. **Nature**, v. 464, n. 7285, p. 66–71, 4 mar. 2010.

CHESLER, A.; FIRESTEIN, S. Neuroscience: Current views on odour receptors. **Nature**, v. 452, n. 7190, p. 944–944, 24 abr. 2008.

CLYNE, P. J. et al. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in Drosophila. **Neuron**, v. 22, n. 2, p. 327–338, fev. 1999.

CRUZ-LÓPEZ, L. et al. A new potential attractant for Anastrepha obliqua from Spondias mombin fruits. **Journal of Chemical Ecology**, v. 32, n. 2, p. 351–365, fev. 2006.

EIRÍN-LÓPEZ, J. M. et al. The birth-and-death evolution of multigene families revisited. **Genome Dynamics**, v. 7, p. 170–196, 2012.

EL HADI, M. et al. Advances in Fruit Aroma Volatile Research. **Molecules**, v. 18, n. 7, p. 8200–8229, 11 jul. 2013.

ENGSONTIA, P. et al. Molecular Evolution of the Odorant and Gustatory Receptor Genes in Lepidopteran Insects: Implications for Their Adaptation and Speciation. **Journal of Molecular Evolution**, v. 79, n. 1-2, p. 21–39, ago. 2014.

ENGSONTIA, P. et al. Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. **BMC Research Notes**, v. 8, n. 1, dez. 2015.

GERSTEIN, M.; SONNHAMMER, E. L.; CHOTHIA, C. Volume changes in protein evolution. **Journal of Molecular Biology**, v. 236, n. 4, p. 1067–1078, 4 mar. 1994.

GREGORIO, P. L. F.; SANT'ANA, J.; REDAELLI, L. R. Percepção química e visual de Anastrepha fraterculus (Diptera, Tephritidae) em laboratório. **Iheringia. Série Zoologia**, v. 100, n. 2, jun. 2010.

GUINDON, S. et al. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. **Systematic Biology**, v. 59, n. 3, p. 307–321, 1 maio 2010.

GUO, S.; KIM, J. Molecular Evolution of Drosophila Odorant Receptor Genes. **Molecular Biology and Evolution**, v. 24, n. 5, p. 1198–1207, 13 fev. 2007a.

GUO, S.; KIM, J. Molecular Evolution of Drosophila Odorant Receptor Genes. **Molecular Biology and Evolution**, v. 24, n. 5, p. 1198–1207, 13 fev. 2007b.

HALL. BioEdit: a user-friendly biologicalsequence alignment editor and analysis programfor Windows 95/98/NT. **Nucleic Acids Symposium Series**, p. 95–98, 3 fev. 1999.

HALLEM, E. A.; CARLSON, J. R. Coding of Odors by a Receptor Repertoire. **Cell**, v. 125, n. 1, p. 143–160, abr. 2006.

HOPF, T. A. et al. Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. **Nature Communications**, v. 6, p. 6077, 13 jan. 2015.

JONES, P. L. et al. Functional agonism of insect odorant receptor ion channels. **Proceedings of the National Academy of Sciences**, v. 108, n. 21, p. 8821–8825, 24 maio 2011.

KARAKAS, E.; FURUKAWA, H. Crystal structure of a heterotetrameric NMDA receptor ion channel. **Science**, v. 344, n. 6187, p. 992–997, 30 maio 2014.

KATOH, K. et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. **Nucleic Acids Research**, v. 33, n. 2, p. 511–518, 19 jan. 2005.

KONDRASHOV, F. A. et al. Selection in the evolution of gene duplications. **Genome Biology**, v. 3, n. 2, p. RESEARCH0008, 2002.

KURLAND, C. G. Codon bias and gene expression. **FEBS letters**, v. 285, n. 2, p. 165–169, 22 jul. 1991.

LEARY, G. P. et al. Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 35, p. 14081–14086, 28 ago. 2012.

LEE, A. G. **Transmembrane Receptors and Channels**. [s.l.] Elsevier, 1997.

LIN, C.-C. et al. Food odors trigger *Drosophila* males to deposit a pheromone that guides aggregation and female oviposition decisions. **eLife**, v. 4, 30 set. 2015.

LÓPEZ-GUILLÉN, G.; TOLEDO, J.; ROJAS, J. C. Response of Anastrepha Obliqua (Diptera: Tephritidae) to Fruit Odors and Protein-Based Lures in Field Trials. **Florida Entomologist**, v. 93, n. 2, p. 317–318, jun. 2010.

LUZ, H.; STAUB, E.; VINGRON, M. About the interrelation of evolutionary rate and protein age. **Genome Informatics. International Conference on Genome Informatics**, v. 17, n. 1, p. 240–250, 2006.

MCBRIDE, C. S. Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 12, p. 4996–5001, 20 mar. 2007.

MCBRIDE, C. S.; ARGUELLO, J. R.; O'MEARA, B. C. Five Drosophila Genomes Reveal Nonneutral Evolution and the Signature of Host Specialization in the Chemoreceptor Superfamily. **Genetics**, v. 177, n. 3, p. 1395–1416, 1 nov. 2007.

POSADA, D. Using MODELTEST and PAUP* to select a model of nucleotide substitution. **Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]**, v. Chapter 6, p. Unit 6.5, fev. 2003.

POSADA, D. jModelTest: Phylogenetic Model Averaging. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1253–1256, 3 abr. 2008.

QUAX, T. E. F. et al. Codon Bias as a Means to Fine-Tune Gene Expression. **Molecular Cell**, v. 59, n. 2, p. 149–161, jul. 2015.

REZENDE, V. et al. **Candidate genes involved in differentiation between two non-model species of fruit flies (Anastrepha: Tephritidae) screened from head transcriptomes. Unpublished manuscript**. 2016.

ROBERTSON, H. M.; WANNER, K. W. The chemoreceptor superfamily in the honey bee, Apis mellifera: Expansion of the odorant, but not gustatory, receptor family. **Genome Research**, v. 16, n. 11, p. 1395–1403, 25 out. 2006.

SÁNCHEZ-GRACIA, A.; VIEIRA, F. G.; ROZAS, J. Molecular evolution of the major chemosensory gene families in insects. **Heredity**, v. 103, n. 3, p. 208–216, set. 2009.

SATO, K. et al. Insect olfactory receptors are heteromeric ligand-gated ion channels. **Nature**, v. 452, n. 7190, p. 1002–1006, 24 abr. 2008.

SELIVON, D. Relações com as plantas hospedeiras. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado.** Malavasi, A.; Zucchi, R.A ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

SMADJA, C. et al. Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, Acyrthosiphon pisum. **Molecular Biology and Evolution**, v. 26, n. 9, p. 2073–2086, 1 set. 2009.

SUYAMA, M.; TORRENTS, D.; BORK, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. **Nucleic Acids Research**, v. 34, n. Web Server, p. W609–W612, 1 jul. 2006.

TAYLOR, J. S.; RAES, J. Duplication and Divergence: The Evolution of New Genes and Old Ideas. **Annual Review of Genetics**, v. 38, n. 1, p. 615–643, dez. 2004.

TRAYNELIS, S. F. et al. Glutamate Receptor Ion Channels: Structure, Regulation, and Function. **Pharmacological Reviews**, v. 62, n. 3, p. 405–496, 1 set. 2010.

TSIRIGOS, K. D. et al. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. **Nucleic Acids Research**, v. 43, n. W1, p. W401–W407, 1 jul. 2015.

VANÍČKOVÁ, L. et al. Current knowledge of the species complex Anastrepha fraterculus (Diptera, Tephritidae) in Brazil. **ZooKeys**, v. 540, p. 211–237, 26 nov. 2015.

WICHER, D. et al. Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. **Nature**, v. 452, n. 7190, p. 1007–1011, 24 abr. 2008.

YANG, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586–1591, 18 abr. 2007.

YANG, Z.; WONG, W. S. W.; NIELSEN, R. Bayes empirical bayes inference of amino acid sites under positive selection. **Molecular Biology and Evolution**, v. 22, n. 4, p. 1107–1118, abr. 2005.

ZHANG, J. Evolution by gene duplication: an update. **Trends in Ecology & Evolution**, v. 18, n. 6, p. 292–298, jun. 2003.

ZHANG, J.; NIELSEN, R.; YANG, Z. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. **Molecular Biology and Evolution**, v. 22, n. 12, p. 2472–2479, 24 ago. 2005.

ZHANG, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. **Genomics, Proteomics & Bioinformatics**, v. 4, n. 4, p. 259–263, nov. 2006.

ZHOU, X. et al. Chemoreceptor Evolution in Hymenoptera and Its Implications for the Evolution of Eusociality. **Genome Biology and Evolution**, v. 7, n. 8, p. 2407–2416, ago. 2015.

ZUO, J. et al. Neurodegeneration in Lurcher mice caused by mutation in delta2 glutamate receptor gene. **Nature**, v. 388, n. 6644, p. 769–773, 21 ago. 1997.

3. Divergence and variation in two pheromone binding proteins and two odorant receptors in four *Anastrepha* species of the fraterculus group.

3.1. Resumo

As moscas-das-frutas do gênero *Anastrepha* são consideradas pragas importantes das culturas de frutas na América do Sul infestando uma ampla gama de hospedeiros. Os receptores olfativos e as proteínas ligantes a feromônios são de grande importância para sua sobrevivência e reprodução uma vez que estão envolvidos no passo inicial do mecanismo de detecção de odores no ambiente. Com o intuito de estudar o papel potencial de genes quimiossensoriais na divergência de espécies proximamente relacionadas do gênero *Anastrepha* identificamos genes *PBPs* em transcriptomas de *A. fraterculus* e *A. obliqua* previamente gerados e investigamos padrões de evolução molecular destes genes em comparação a outros Tephritidae e Drosophilidae. Escolhemos alguns destes genes além de outros OR previamente identificados para estudar padrões de polimorfismo e divergência em populações de quatro espécies do grupo fraterculus (*A. fraterculus*, *A. obliqua*, *A. sororcula*, *A. turpiniae*). Identificamos cinco genes *PBP* nas espécies *A. fraterculus* e *A. obliqua*, que apresentaram níveis altos de similaridade a suas respectivas cópias homólogas de outros tefritídeos, sendo que o gene *Obp28a* foi o mais diferenciado. Resultados de *strict branch site* test indicam seleção purificadora agindo na evolução das PBPs, mas análises de Ka/Ks sugerem seleção positiva no gene *Obp28a*. A análise de diversidade nucleotídica dos quatro genes quimiossensoriais amplificados (dois *PBPs* e dois *OR*) em populações das espécies aqui estudadas indicou ausência de desvios da neutralidade nos diferentes testes realizados e altos níveis de diversidade nucleotídica e haplotípica. Análises de divergência nucleotídica, de F_{st} e filogenéticas sugerem que *A. obliqua* é a espécie mais diferenciada enquanto que *A. fraterculus*, *A. sororcula* e *A. turpiniae* são mais próximas, podendo até estar apresentando fluxo gênico entre as mesmas. Os genes investigados apresentaram diferenciação em *A. obliqua* e essa diferencição poderia estar relacionada com a adaptação a diferentes hospedeiros e com a diferenciação nos feromônios entre as espécies.

Palavras chave: Espécies irmãs, divergência, receptores olfativos, Proteínas ligantes a feromônios.

3.2. Abstract

Fruit flies of the *Anastrepha* genus are considered one of the main pests to infest fruit crops in South America infecting a wide range of host plants. Odorant pheromone binding proteins (*PBPs*) and odorant receptors are of great importance for survival and reproduction of these flies since they participate in the initial steps of odor perception. In order to evaluate the potential role of chemosensory genes in the divergence of closely related species of the *Anastrepha fraterculus* group, we searched for *PBPs* in transcriptomes of *A. fraterculus* and *A. obliqua* to study their evolutionary patterns, when compared to *PBPs* from other Tephritidae and investigated patterns of polymorphism and divergence in two *PBP* genes and two *OR* genes in populations of four *Anastrepha* species. We identified five *PBPs* in the transcriptomes of *A. fraterculus* and *A. obliqua* that corresponded to putative homologous genes from other tephritids as well as drosophilids. The putative proteins expressed from these *PBP* genes show high levels of similarity when comparing orthologous copies between the two species. Among these copies *Obp28a* gene was the most differentiated between the two species. We found purifying selection acting in the evolution of *Anastrepha PBP* genes when we performed a strict branch site test but Ka/Ks analyses show evidence of positive selection in the *Obp28a* gene. When we analyzed the nucleotide diversity of four chemosensory genes amplified in populations of four species of the *Anastrepha fraterculus* group, we found high levels of nucleotide and haplotype diversity. We failed to find significant departures from neutrality. Analysis of nucleotide divergence shows that *A. obliqua* is the most differentiated species, so much so that we detected fixed differences in *A. obliqua* for all genes here studied and the *A. fraterculus*, *A. sororcula* and *A. turpiniae*, whereas these species did not show fixed differences for any gene. F_{st} and phylogenetic analyses suggest that *A. obliqua* is a species phylogenetically more differentiated and *A. fraterculus*, *A. sororcula* and *A. turpiniae* are more closely relates species or are experiencing gene flow that prevent the differentiation of the genes here studied. The differentiation of these four chemosensory genes in *A. obliqua* may be related with adaptation to different host or even with differences in pheromone blends.

Key Words: Sister Species, divergence, odorant receptors, pheromone binding proteins.

3.3. Introduction

Fruit flies of the genus *Anastrepha* are considered the most important pest of various cultivated fruits in central and South America since they are responsible for damage in the major fruits crops. The most economically important species in the genus are: *A. grandis* (Macquart, 1846), *A. fraterculus* (Wiedemann, 1830), *A. obliqua* (Macquart, 1835), *A. pseudoparallela* (Loew, 1873), *A. sororcula* (Zucchi, 1979), *A. striata* (Schiner, 1868) and *A. zenildae* (Zucchi, 1979) (ZUCCHI, 2000). Most of these species belong to the fraterculus group which includes closely related species that are morphologically difficult to differentiate due to the presence of overlapping interspecific in morphological markers (ARAUJO; ZUCCHI, 2006).

Approximately 38 host plant families have been associated with *Anastrepha* in Brazil, of which Myrtaceae, Anacardiaceae, Sapotaceae and Passifloraceae are the most important. *A. fraterculus*, with 110, followed by *A. obliqua* with 48, are the *Anastrepha* species associated with the higher number of host species, whereas *A. sororcula* has been found on 21 and *A. turpiniae* on 13 different host species. Even though several of these species have been considered polyphagous, there seems to be some host plant preferences, such that *A. fraterculus*, *A. turpiniae*, and *A. sororcula* are preferentially associated with species that belong to the Myrtaceae family, whereas *A. obliqua* has been reported to preferentially infest members of Anacardiaceae family (SELIVON, 2000). A wind tunnel test showed that *A. fraterculus* was more attracted to extract plants from Myrtaceae (such as guava and Surinam cherry), in which the oviposition rate was bigger than in plants from other families (GREGORIO; SANT'ANA; REDAELLI, 2010), whereas *A. obliqua* was more attracted to extracts from Anacardiaceae plants (such as mangoes and hog plum) (CRUZ-LÓPEZ et al., 2006; LÓPEZ-GUILLÉN et al., 2011).

Even though the majority of studies have focused on the attraction of fruit flies to lures or food attractants (TAN et al., 2014), fruit flies are attracted to more than fruit odors, since pheromones play an important role in their biology, particularly in several species of *Anastrepha* in which males produce sexual pheromones to attract conspecific males and females to form leks. This aggregation serves for localization and mate selection and generally occurs on foliage of host plants in which females gain access to both, males and sites for oviposition (VILELA; KOVALESKI, 2000). Pheromone blends of males of *Anastrepha spp* may have a wide number of compounds which are species-specific despite

some common compounds which are shared across several species (LIMA-MENDONÇA et al., 2014).

The detection of host plant volatiles as well as mate pheromones is carried out in the olfactory neurons housed in the sensilla which are specialized hair-like structures arrayed along the insect antennae (TEGONI; CAMPANACCI; CAMBILLAU, 2004). The main proteins involved in this recognition process are the odorant receptor proteins (OR), which are embedded in the dendrites' membrane, and the odorant binding proteins (OBP), which are expressed in the sensillary lymph (HANSSON; STENSMYR, 2011). Among OBP proteins there are some that specifically recognize and bind to pheromones and are named pheromone binding proteins (PBP). These olfactory proteins provide a mechanism of odor detection with extreme sensitivity and specificity discriminating a large number of olfactory signals (SU; MENUZ; CARLSON, 2009).

Few studies have investigated molecular evolution and levels of genetic polymorphism and divergence of these chemosensory genes at intraspecific and interspecific levels. Nevertheless, these studies provide important insight of the role of genetic diversity and patterns of selection that could be affecting population and species differentiation. Interspecific studies in *D. melanogaster* detected that some OR genes were influenced by selection, showing some sequence polymorphisms that were associated with variation in responses to molecular cues (ROLLMANN et al., 2010). At the interspecific level, high levels of genetic divergence and fixed differences were found among cryptic species of the *Anopheles gambiae* complex in the *Or39a* locus (ARNAL et al., 2014).

A study of a PBP locus in two populations of the Lepidoptera *Agrotis segetum* which the same range of volatiles, but in different mean ratios, found high nucleotide variation between these populations, some of which leading to amino acids replacements. In that study a phylogenetic analysis provide evidence that the two populations are in process of genetic isolation (LaForest et al., 1999). Another study that focused on two PBPs in four species of the melanogaster group found accumulation of fixed amino acids replacements in *D. melanogaster*, suggesting the action of natural selection in the evolution of the two PBPs in this group (Sánchez-Gracia et al., 2003).

In order to better understand the selective forces driving the evolution of olfactory genes and their potential role in the divergence of four fruit fly species of the fraterculus group, we searched for PBP genes in transcriptomes of head and reproductive tissues from fruit flies *A. fraterculus* and *A. obliqua,* investigated patterns of molecular evolution on these genes. To further investigate their potential role in species differences, we also used two *PBP* genes, here identified, and two OR genes, previously described, to study intra and interspecific patterns of nucleotide diversity and divergence in four closely related species of *Anastrepha* fruits flies (*A. fraterculus*, *A. obliqua*, *A. sororcula*, and *A. turpiniae*).

3.4. Materials and methods

3.4.1. Identification of A. fraterculus and A. obliqua PBPs genes

We searched for *PBP* genes in reproductive and head tissues transcriptomes of individuals from populations of *A. fraterculus* from Southeastern (22° 01' 03" S, 47° 53' 27" W) and of *A. obliqua* from Midwest (16° 41' 58" S, 49° 16' 35" W) regions of Brazil. Details on sample preparation, molecular procedures, assembly and annotation of the RNA-seq data are described elsewhere (REZENDE et al., 2016).

We conducted BLASTx (ALTSCHUL et al., 1990) searches on head and reproductive transcriptomes against the Gene Ontology and the *Drosophila melanogaster* database as a reference, since it is the best curated insect genome. We retained only the first match of each search and complete contigs with an e-value threshold of at least 10^{-5}. Open reading frames (ORFs) were inferred using the software ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and the signal peptide of every ORF was predict using PrediSi (PREDIction of SIgnal peptides) software tool (http://www.predisi.de). The annotated sequences were named as *Afra* and *Aobl* followed by the respective *Obp* name in concordance with their similarity with *D. melanogaster PBPs*

3.4.2. Phylogenetic analysis of PBPs

We performed a phylogenetic analysis using the identified sequences of *A. fraterculus* and *A. obliqua* and the sequences reported in GenBank (National Center for Biotechnology Information, NCBI) for *D. melanogaster*, *Ceratitis capitata*, *Bactrocera dorsalis* and *Bactrocera cucurbitae* (Supplementary Material 1). The amino acid sequences were aligned using MAFFT v7.21 (KATOH et al., 2005), and visually inspected with BioEdit v7.2.

(HALL, 1999). The nucleotide sequences were aligned using the amino acid alignment as a guide with PAL2NAL software (SUYAMA; TORRENTS; BORK, 2006). We inferred the best-fitted model of substitution for the alignment using jModelTest (POSADA, 2008) and used the Akaike information criteria (AIC) as model selection strategy. The Maximum likelihood phylogenetic relationships were inferred with PhyML v3.0 (GUINDON et al., 2010) and the ML tree was visualized using the program FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

3.4.3. Molecular evolution analysis and candidate genes

To identify genes with high evolutionary rates and evidence of positive selection, we investigated patterns of molecular evolution of *PBP* genes by separating the *PBP* gene tree into five lineages, considering that comparisons across these different lineages included *PBPs* that were too divergent, indicated by saturation at deeper evolutionary distance (Supplementary Material 2), which would hamper the analyses of positive selection. After that, we tested each lineage considering a strict branch-site test that compared models A vs. Anull (Zhang et al, 2005) using the software CODEML implemented in PAML v.4 (YANG, 2007). The comparison of those models was made using a Likelihood ratio test (LRT) (ANISIMOVA; BIELAWSKI; YANG, 2001) and significant results were determined with X^2 tests. The phylogenetic trees and tested branches are available in Supplementary Material 3. We also estimated pairwise Ka, Ks and the Ka/Ks ratio for the complete putative orthologous *PBP* genes between the two species here studied, *A. fraterculus* and *A. obliqua* using KaKs_Calculator (ZHANG et al., 2006) and the MS model (Posada, 2003).

We chose two of the *PBP* genes here identified, as well as two *OR* genes previously identified using the same procedures here described, on the same species (Chapter 2) to perform a more detailed analysis to investigate whether we would detect species specific fixed differences in these markers across different populations. In both cases we chose one gene that had evidence of positive selection and one that was evolving under purifying selection.

3.4.4. Species and populations sampled

Adult individuals that emerged from different host fruit plants collected from several localities in Brazil (Table 1 and Figure 1) were identified at the species level. We considered four species in this study: *Anastrepha fraterculus*, *A. obliqua*, *A. turpiniae* and *A. sororcula.* We sampled five individuals from each population and considered a total of three populations per species.

Table 1. Geographical coordinates of the sampled localities for each species.

Species	Location	S	W	Code
A. fraterculus	Bonito PE	8° 29' 21.768"	35° 43' 7.824"	AFB
	Florianópolis SC	27° 35' 41.496"	48° 32' 53.556"	AFF
	João Pessoa PB	7° 7' 10.164"	34° 50' 41.496"	AFJ
A. obliqua	Capanema PR	25° 39' 55.404"	53° 48' 29.844"	AOP
	Conceição do Jacareí RJ	22° 59' 10.752"	44° 5' 7.188"	AOJ
	Marialva PR	23° 29' 4.632"	51° 47' 35.124"	AOM
A. sororcula	Bonito PE	8° 29' 21.768"	35° 43' 7.824"	ASB
	João Pessoa PB	7° 7' 10.164"	34° 50' 41.496"	ASJ
	Recife PE	8° 2' 51.216"	34° 52' 37.056"	ASR
A. turpiniae	Dois Irmãos RS	29° 35' 23.424"	51° 5' 7.404"	ATD
	Guarapuava PR	25° 23' 22.128"	51° 27' 46.080"	ATG
	Três Marias MG	18° 12' 19.296"	45° 13' 55.092"	ATT

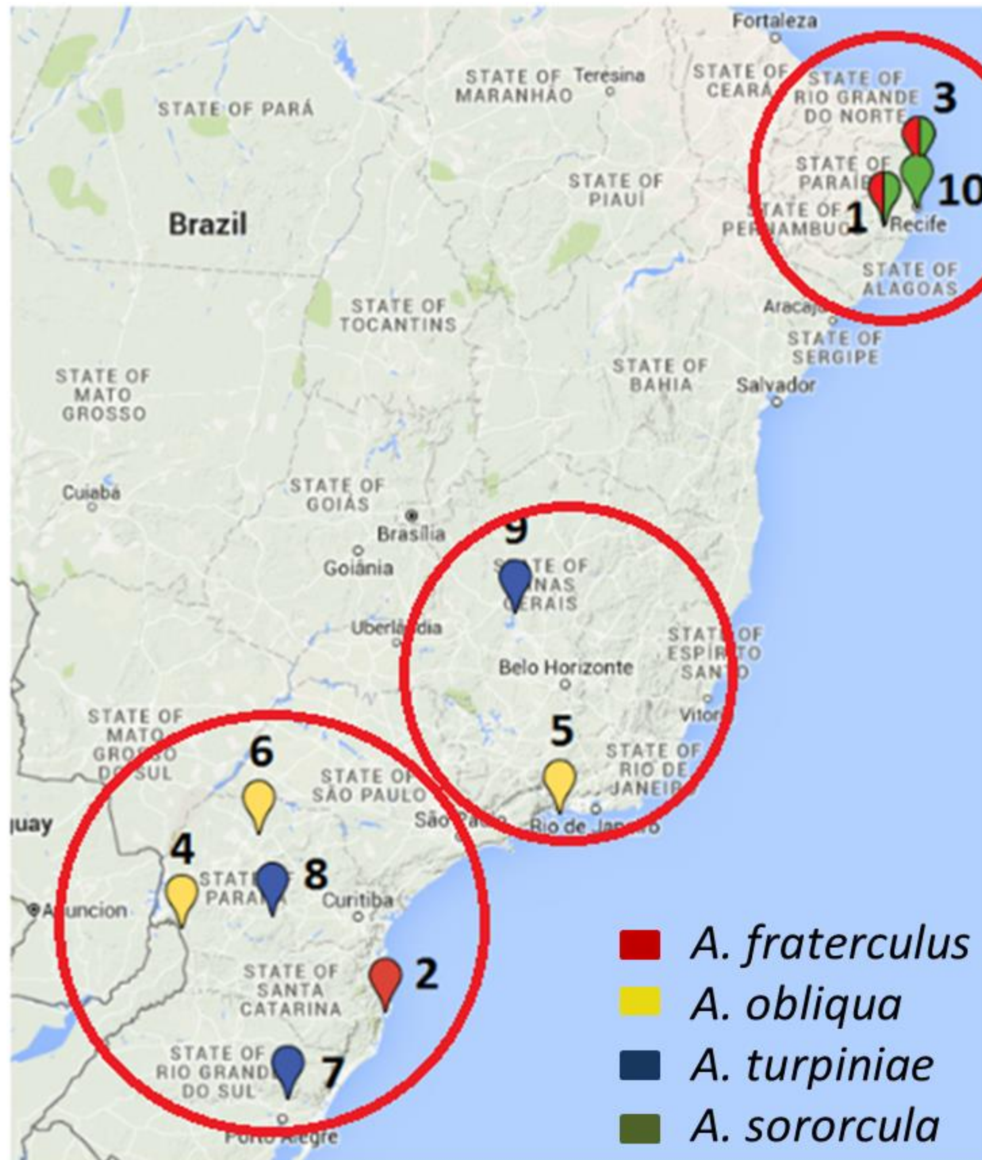


Figure 1. Localities Sampled. Red points indicate the localities of *A. fraterculus* populations; (1) Bonito, (2) Florianópolis, (3) João Pessoa. Yellow points indicate the localities of *A. obliqua* populations; (4) Capanema, (5) Conceição do Jacareí, (6) Marialva. Blue points represent the localities of *A. turpiniae* populations; (7) Dois Irmãos, (8) Guarapuava, (9) Três Marias. Green points indicate the localities of *A. sororcula* populations; (1) Bonito, (3) João Pessoa, (10) Recife. Red circles represent regions divison.

3.4.5. Molecular procedures

DNA was extracted from individual flies following the guanidinium thiocyanate-phenol-chloroform method (CHOMCZYNSKI; SACCHI, 1987) with slight modifications. PCR amplifications of each gene was performed with primers manually designed from the alignments of the *A. fraterculus* and *A. obliqua* sequences, considering the stability of their structures, examined with OligoAnalyzer 3.1 (https://www.idtdna.com/calc/analyzer).

The *PBP* genes were amplified as a unique fragment, while OR genes were divided in two fragments since they were longer using the primer sequences indicated on Table 2. PCR reactions used the following conditions: 1X *Taq* polymerase buffer (100mM Tris-HCl pH 8.8, 500mM KCl and 0.8% (v/v) Nonidet P40), 2mM $MgCl_2$, 0.2mM dNTPs, 0.25µM of primer, 1.5 U *Taq* Polymerase and 1.5µl of genomic DNA in a total volume of 25µ. Reactions consisted of 30 cycles of 94°C for 0.5 min, the corresponding annealing temperature for every gene (showed on Table 2), and 72°C extension for 1.3 min, followed by a final extension at 72°C for 5 min. Positive amplifications were purified by PEG 8000 precipitation (LIS; SCHLEIF, 1975), quantified in 1% agarose gel with ethidium bromide and sequenced with forward and reverse primers at Macrogen Inc, Korea. Sequences were visually inspected for quality with Bioedit v7.2 and sequences from the same individual were aligned using Clustal W (THOMPSON; HIGGINS; GIBSON, 1994) implemented on Bioedit v7.2. Heterozygous individuals had their genes amplified and PEG purified again and cloned with InsTAclone PCR cloning Kit (Thermo Scientific) following the manufacturer's recommendations. At least three colonies per individual were then purified, quantified and sequenced using the forward M13 primer.

3.4.6. Diversity, neutrality and divergence analysis among Anastrepha species

Sequences from each individual were aligned using Clustal on Bioedit v7.2, and visually inspected. Their structure is available on Supplementary Material 4. We used the aligned sequences to perform several neutrality and divergence tests using the complete CDS of the genes *Obp28a*, *Obp84a* and *Or7d*. In the case of the *Or7c* gene we used a fragment of 795 nucleotides witch corresponded to 67.4% of the CDS of this gene (265 amino acids of 393). This was done because of the inability to correctly assemble, for most of the samples, the two amplified fragments across an intron that was longer than 500pb.

Table 2. Primer sequences used for the amplification of the genes here studied

Gene	Fragment	Forward Primer (5'-3')	Reverse Primer (5'-3')	AT (°C)	Product (pb)
Obp28a	1	TGTATTTACACATTTACAATGGC	CTTGGTTCTAACTGTGGGATTC	50	601
Obp84a	1	CGCACTTCCGAGTTCTAAAG	GCTGACATTCAAACAGTTGTG	50	783
Or7d	1	TCCACAGTAAGTTGTCTAGC	GCAGTAACACGCACATAAAGTA	54	940
	2	ATTTTTGATCGCCGCTTCA	GCTTCTTTAAGTCCAATGCTG	52	434
Or7c	1	GCTCTCGCACCCATTATA	GTGACGCCAAGTATAGCC	60	750
	2	ATTTTTGATCGCCGCTTCA	GCTTCTTTAAGTCCAATGCTG	58	847

AT: Annealing temperature.

For each gene and species we calculated the following diversity estimates: number of haplotypes (h), haplotype diversity (Hd), nucleotide diversity (Pi), and the number of segregating sites. We performed the following neutrality tests for every gene in each species; Fu & Li's D, Fu & Li's F, Tajima´s D and Fu's Fs using DnaSP v5.10.01 software (LIBRADO; ROZAS, 2009). To investigate whether geographical distribution or species affiliation were responsible for the observed diversity distribution, we performed two AMOVA analyses, one grouping the population of the same species and other grouping the population of the same region independently of the species. We also performed a gene flow and differentiation analysis comparing every gene between the species, for that, we calculated paired F_{st} and AMOVA tests with the Arlequin 3.5.2.2 (EXCOFFIER; LAVAL; SCHNEIDER, 2005).

To estimate the total divergence of every gene per species, we calculated the average number of nucleotide substitutions per site (K(JC)) for every species against the other species. To measure the pairwise divergence between species we calculated K(JC), the synonymous nucleotide divergence (Ks), non-synonymous nucleotide divergence (Ka), Ka/Ks ratio and the number of fixed differences using DnaSP v5.10.01 software (LIBRADO; ROZAS, 2009).

Finally we performed a phylogenetic analysis for every gene to test whether the genic tree corresponds with species tree segregation. For that, we first identified and aligned all unique sequences per species. We performed a substitution saturation test on these aligned sequences through plotting transitions and transversions against the genetic distance with the best fitted model for every alignment with the software DAMBE 5.5.24 (Xia and Xie, 2001) (XIA; XIE, 2001) considering the substitution models and the maximum likelihood phylogenetic relationships for each gene calculated as previously indicated for the *PBP* genes using sequences from the same gene from one individual of *Anastrepha bistrigata* (Bezzi) that we sequenced following the same procedures previously indicated.

3.5. Results and Discussion

3.5.1. Pheromone binding proteins genes in Anastrepha fraterculus and Anastrepha obliqua

We identified five complete sequences associated with pheromone binding proteins in each of the *Anastrepha* species here investigated (Supplementary Material 5); which corresponded to five genes in *D. melanogaster*. Every sequence found in *A. fraterculus* had orthologous gene in *A. obliqua* (Table 3) and also in *C. capitata*, *B. dorsalis* and *B. cucurbitae* (Figure 2). In fact, we identified putative orthologous to all sequences from these species. The length of the ORFs ranged from 142 to 162 amino acids (Table 3), which are in accordance with the reported length for other insects like dipterans (HEKMAT-SCAFE et al., 2002) and lepidopterans (ALLEN; WANNER, 2011; FORSTNER et al., 2006; ZHANG et al., 2012).

Table 3. *PBP*s identified in *A. fraterculus* and *A. obliqua* transcriptomes

A. fraterculus	ORF*	Signal peptide	*A. obliqua*	ORF*	Signal peptide
AfraObp19d	142	1-18	*AoblObp19d*	142	1-18
AfraObp28a	146	1-18	*AoblObp28a*	146	1-18
AfraObp69a	147	1-23	*AoblObp69a*	147	1-23
AfraObp83a	155	1-31	*AoblObp83a*	155	1-31
AfraObp84a	165	1-27	*AoblObp84a*	165	1-27

*Open reading frames length in amino acids

The average percent identity at the amino acid level among paralogous copies was 19.81% in *A. fraterculus* and 19.15% in *A. obliqua*. The reduced similarity values among the different *PBP*s within species are related with the function of each one as carrier of different pheromones (DU; PRESTWICH, 1995; MAÏBÈCHE-COISNE et al., 1997; PLETTNER et al., 2000) and a reflection of their divergence. On the other hand, the comparisons between the orthologous genes of *A. fraterculus* and *A. obliqua* shows a high percent identity (95.764%) which ranged from 86.9% to 100%, being *Obp28a* the most differentiated PBP amongst the two species (Table 4). These high values of identity may reflect the recent divergence between these two species and could be an indication of function retention, such as may be the case of the gene *Obp69a* which has the same sequence for both species.

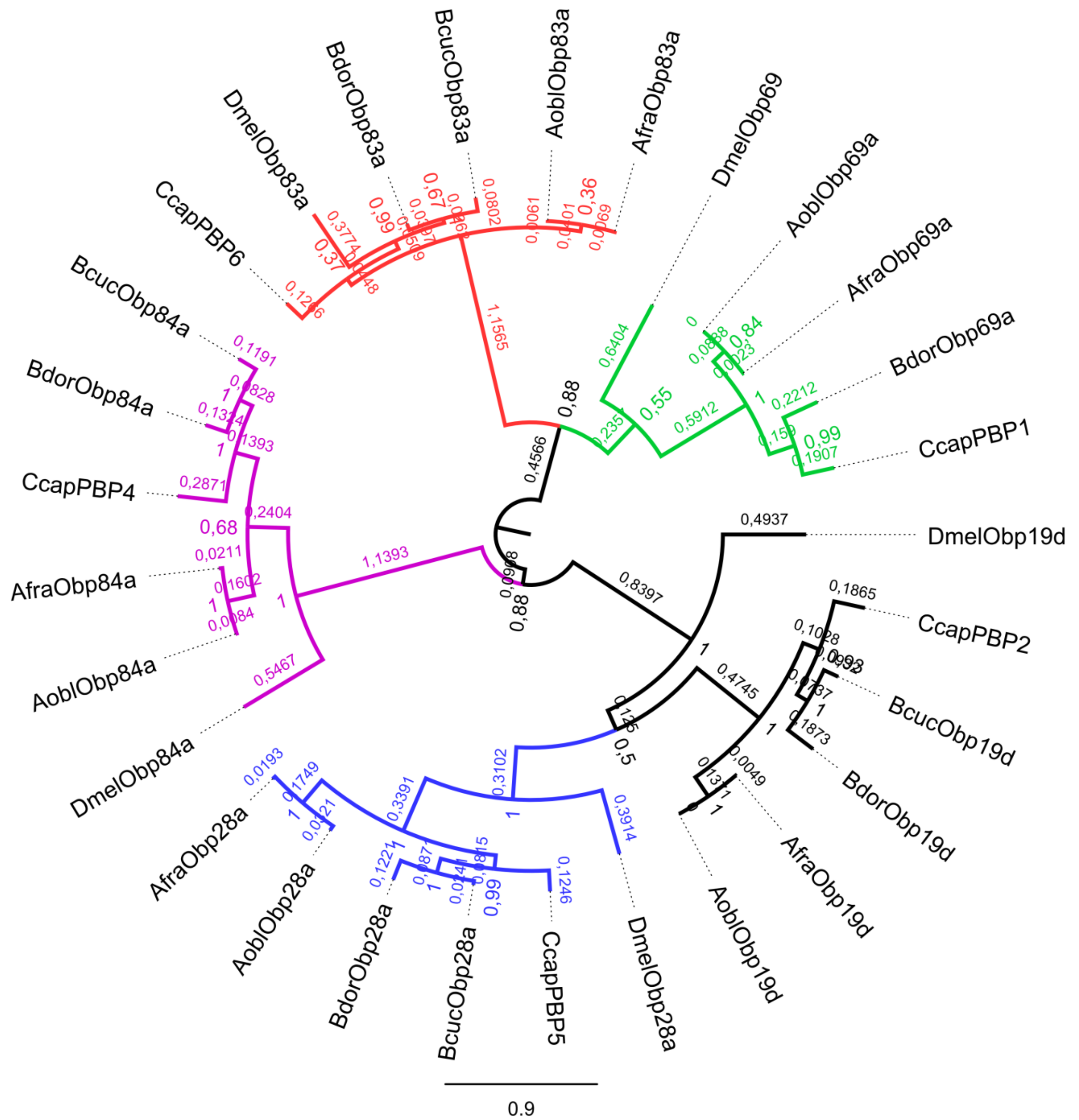


Figure 2. Maximum likelihood phylogenetic relationships among *A. fraterculus* (*Afra*), *A. obliqua* (*Aobl*), *C. capitata* (*Ccap*), *B. dorsalis* (*Bdor*), *B.cucurbitae* (*Bcuc*) and *D. melanogaster* (*Dmel*) *PBP* genes. Node labels represent branch support (Bayes aLRT) and branch labels represent branch lengths estimated as nucleotide substitutions per site.

Table 4. Percent identity at the amino acid level between orthologous PBP genes identified in *A. fraterculus* and *A. obliqua.*

Orthologous	% Identity
Obp19d	99.30
Obp28a	86.99
Obp69a	100.0
Obp83a	96.77
Obp84a	95.76
Average	**95.764**

The gene *Obp28a* is the most differentiated gene between these two species. Functional divergence associated with sequence divergence has been described for other genes in the *OBP* family in *D. melanogaster* (SÁNCHEZ-GRACIA; ROZAS, 2008) and other insects (ALLEN; WANNER, 2011; HARADA et al., 2012; LIU et al., 2015; PLETTNER et al., 2000). It is possible that the divergence found amongst *A. fraterculus* and *A. obliqua* sequences could be related with a functional differentiation that would be a reflection of pheromone induced behaviors like male aggregation and/or female attraction.

3.5.2. Molecular evolution and positive selection in PBPs

We investigated for positive selection comparing all *Anastrepha* lineages of *PBPs* using the strict branch-site test and we failed find significant evidence of positive selection at this level, indicating, instead, that these genes are evolving under purifying selection, possibly due to restrictions related with the significant role of these genes in individual survival and reproduction. Nonetheless, due to the great divergence across different *PBP* copies, it is possible that signal of positive selection have been erased by saturation, particularity because positive selection has been observed in other studies among orthologous groups of OBP genes in the *melanogaste*r group (SÁNCHEZ-GRACIA; VIEIRA; ROZAS, 2009), and also in different PBPs and GOBPs (FORET; MALESZKA, 2006; SÁNCHEZ-GRACIA; ROZAS, 2008; WILLETT, 2000). In spite of evidence of saturation when comparing more diverging taxa, such as drosophilids and tephritids, there is no indication that there is saturation influencing more closely related taxa such as within family comparisons.

A closer investigation of signals of positive selection acting on *Anastrepha PBP* genes compared synonymous and non-synonymous substitution rates between the complete putative orthologous genes identified in *A. fraterculus* and *A. obliqua*. The Ka/Ks analysis revealed two pairs of copies with Ka/Ks values above 1.0 (Table 5), *Obp28a* with a Ka/Ks value of 3.04 and *Obp83a* with value of 2.04, indicating that they may be under positive selection. Because of that, here we compared sequences of *PBPs* from populations of *A. fraterculus* and *A. obliqua*, which show lower levels of differentiation between orthologous genes and do not suffer the effect of saturation that may have affected deeper divergence.

Table 5. Pairwise estimates of Ka, Ks, and Ka/Ks ratio between *A. fraterculus* and *A. obliqua* complete orthologous *PBP* genes identified in transcriptomes.

Compared orthologous	Ka	Ks	Pairwise Ka/Ks	P-Value(Fisher)
AfraObp69a-AoblObp69a	9.86E-06	0.0098	0.001	0.0843
AfraObp19d-AoblObp19d	0.0029	0.0118	0.251	0.2209
AfraObp84a-AoblObp84a	0.0204	0.0641	0.318	0.0149
AfraObp83a-AoblObp83a	0.0155	0.0758	**2.043**	0.6005
AfraObp28a-AobliObp28a	0.0599	0.0197	**3.041**	0.1201

3.5.3. Diversity, neutrality and divergence among Anastrepha species

We chose four genes to investigate patterns of nucleotide segregation across different *Anastrepha* species, one *OR* and one *PBP* gene that was supposed to be evolving under positive selection, and an *OR* and a *PBP* gene that was potentially evolving under purifying selection. We chose these genes from the results here produced, but also from data generated elsewhere (Chapter 2). The selected *OR* genes are members of a recent gene expansion found in tephritids, such that the unique ortholog found in *D. melanogaster* (*DmelOr7a*) is related with detection of several fruit volatiles and with the process of aggregation and oviposition site selection. These two genes were chosen because *Or7c* represents the gene with the highest Ka/Ks ratio (0.9) in a comparison between *A. fraterculus* and *A. obliqua* whereas *Or7d* had a lowest Ka/Ks ratio (0.3), with the exception of *Orco*, which is conserved across Diptera, enabling us to contrast potentially different patterns of selection.

We amplified and sequenced individuals from three populations of four different species of *Anastrepha* for these genes in order to investigate patterns of nucleotide intra and interspecific divergence. Even though we only had sequences from *A. fraterculus* and *A. obliqua*, the evolutionary proximity among species in the fraterculus group enabled us to create conserved primers that would amplify different species in the genus, so much so that we managed to amplify and sequence all genes here investigated for *A. sororcula*, *A. turpiniae* and even for the outgroup *A. bistrigata*.

We obtained an average of 109 sequences for each gene across the four species investigated: *A. fraterculus*, *A. obliqua*, *A. sororcula* and *A. turpiniae*. We found an average number of haplotypes of 65.8±12.8 (Table 6) per gene and an average of 75.7 (±17.6) and 126.2 (±9.2) segregating sites in *OBP* and *OR* genes, respectively across all species. The values of haplotype and nucleotide diversity were similar between genes; the *Obp28a* gene had the highest average value of nucleotide diversity, while the *Obp84a* gene had the lowest average haplotype diversity. Across all genes, *A. sororcula* showed the highest nucleotide diversity, whereas *A. obliqua* had the lowest, though these values are not equally represented in haplotypic diversity. In general, all species had large number of different haplotypes, which could be a consequence of large effective population sizes.

We performed several neutrality tests and failed to find significant departures from neutrality for all genes using Fu & Li’s D and F and for Tajima´s D (Table 7). Nevertheless, we found significant negative values for Fu's Fs test for *AoblObp28a* and *AoblOr7d*, for *AsorObp84a* and *AsorOr7c* and for *AturObp28a*. These negative values for the Fu's Fs test indicate an excess of rare alleles, which is more likely caused by a recent population expansion or genetic hitchhiking (FU, 1997). Because this pattern did not influence all genes here investigated for the species affected, it may be more likely that it might be caused by genetic hitchhiking, though we cannot rule out population expansion.

Table 6. Haplotype and nucleotide diversity found in four olfactory genes.

Gene	Specie	h	Hd	Pi	S	Ns
Obp28a	*A. fraterculus*	15	0.891	0.013	18	28
	A. obliqua	21	0.951	0.009	27	30
	A. sororcula	16	0.896	0.017	24	30
	A. turpiniae	21	0.977	0.012	31	30
	Average		**0.929**	**0.013**		**118**
Obp84a	*A. fraterculus*	13	0.845	0.010	24	30
	A. obliqua	10	0.641	0.004	11	30
	A. sororcula	20	0.967	0.011	24	30
	A. turpiniae	7	0.673	0.010	18	30
	Average		**0.782**	**0.009**		**120**
Or7c	*A. fraterculus*	16	0.963	0.008	35	25
	A. obliqua	11	0.874	0.006	29	22
	A. sororcula	22	0.992	0.014	54	24
	A. turpiniae	20	0.985	0.013	50	24
	Average		**0.954**	**0.010**		**95**
Or7d	*A. fraterculus*	14	0.948	0.011	49	22
	A. obliqua	27	0.993	0.007	39	30
	A. sororcula	18	0.974	0.010	42	24
	A. turpiniae	17	0.962	0.008	37	28
	Average		**0.969**	**0.009**		**104**

h: Number of haplotypes; **Hd:** Haplotype diversity; **Pi:** Nucleotide diversity; **S:** Number of segregating sites. **Ns:** Number of sequences. Values reported in Average column for **Ns** correspond to Total values.

Table 7. Neutrality tests on four olfactory genes.

Gene	Species	Fu & Li's *D*	Fu & Li's *F*	Tajima´s *D*	Fu's *Fs*
Obp28a	*A. fraterculus*	1.212	1.231	0.529	-3.112
	A. obliqua	-1.181	-1.526	-1.400	-13.516*
	A. sororcula	0.454	0.619	0.615	-1.978
	A. turpiniae	-1.053	-1.302	-1.104	-10.475*
Obp84a	*A. fraterculus*	-2.464	-2.169	-0.519	-1.537
	A. obliqua	-0.679	-0.779	-0.593	-2.603
	A. sororcula	-0.383	-0.305	-0.077	-8.043*
	A. turpiniae	-0.565	-0.230	0.539	3.177
Or7d	*A. fraterculus*	0.887	0.683	-0.069	-0.022
	A. obliqua	-0.528	-0.616	-0.464	-16.607*
	A. sororcula	0.801	0.807	0.396	-2.981
	A. turpiniae	1.449	1.176	0.037	-1.654
Or7c	*A. fraterculus*	-0.660	-0.905	-0.952	-3.726
	A. obliqua	-1.076	-1.458	-1.355	-1.093
	A. sororcula	-1.674	-1.709	-0.914	-9.931*
	A. turpiniae	-0.405	-0.631	-0.759	-6.385

*$p<0.01$ After Bonferroni correction for multiple tests

To investigate patterns of intra and interspecific divergence, we calculated the average number of nucleotide substitutions per site comparing sequences across species in a pairwise manner, per gene. Table 8 shows that *A. obliqua* has the highest values of K(JC) across all genes, being the most differentiated species. The other three species had similar values of K(JC) showing lower levels of nucleotide substitutions, consequence, possibly, of a closer relationship.

Table 8. Divergence values for four olfactory genes per species.

Gene	Species	K(JC)	Gene	Species	K(JC)
Obp28a	*A. fraterculus*	0.03085	***Obp84a***	*A. fraterculus*	0.01720
	A. obliqua	0.05827		*A. obliqua*	0.02230
	A. sororcula	0.03438		*A. sororcula*	0.01665
	A. turpiniae	0.03188		*A. turpiniae*	0.01986
	Average	**0.03885**		**Average**	**0.01900**
Or7c	*A. fraterculus*	0.02014	***Or7d***	*A. fraterculus*	0.01949
	A. obliqua	0.03473		*A. obliqua*	0.02287
	A. sororcula	0.02170		*A. sororcula*	0.01753
	A. turpiniae	0.02133		*A. turpiniae*	0.01827
	Average	**0.02447**		**Average**	**0.01954**

K(JC): Average number of nucleotide substitutions per site for every species with Jukes and Cantor correction.

We performed pairwise analyses of nucleotide substitutions between different species in order to evaluate nucleotide divergence and the nature of these substitutions considering synonymous, non-synonymous nucleotide divergence and its ratio. All pairwise comparisons which involved *A. obliqua*, for every gene here investigated, had the highest values of average substitutions (K(JC)) showing a higher level of differentiation and divergence of this species compared to others (Table 9). On the other hand, the majority of genes (*Obp28a*, *Obp84a*, *and Or7c*) indicated that the comparisons between *A. fraterculus* and *A. turpiniae* had the lowest substitution levels followed by the comparisons between *A. fraterculus* and *A. sororcula* indicating that these species may be more closely related with *A. fraterculus* than *A. obliqua*.

All *Obp28a* pairwise analyses that included *A. obliqua* had Ka/Ks values greater than 1. An investigation of these values show that this is caused by an increase in non-synonymous substitutions, rather than a reduction of synonymous substitutions, suggesting that the differentiation of *AoblObp28a* could have been driven by positive selection (Table 8). In the case of the gene *Or7c*, Ka/Ks ratio values were close to 1 for most pairwise comparisons that involved *A. obliqua*, which could also indicate an important role for positive selection on its differentiation.

Table 9. Divergence and differentiation values for four olfactory genes between species pairs.

Gene		FD	K(JC)	Ka(JC)	Ks(JC)	Ka/Ks ratio
Obp28a	fraterculus-obliqua	13	**0.05810**	0.06394	0.03743	**1.708**
	fraterculus-sororcula	0	0.01810	0.01682	0.02275	0.739
	fraterculus-turpiniae	0	0.01708	0.01534	0.02339	0.656
	obliqua-sororcula	13	**0.06163**	0.06749	0.04087	**1.651**
	obliqua- turpiniae	12	**0.05507**	0.06053	0.03568	**1.696**
	sororcula- turpiniae	0	0.02305	0.02093	0.03075	0.681
Obp84a	fraterculus-obliqua	4	**0.02409**	0.01400	0.06493	0.216
	fraterculus-sororcula	0	0.01512	0.01282	0.02420	0.530
	fraterculus-turpiniae	0	0.01243	0.01296	0.01034	**1.253**
	obliqua-sororcula	0	0.01522	0.00787	0.04464	0.176
	obliqua- turpiniae	5	**0.02764**	0.01710	0.07037	0.243
	sororcula- turpiniae	0	0.01960	0.01713	0.02934	0.584
Or7c	fraterculus-obliqua	12	**0.03431**	0.03245	0.04019	**0.808**
	fraterculus-sororcula	0	0.01412	0.00949	0.02869	0.331
	fraterculus-turpiniae	0	0.01334	0.00951	0.02540	0.374
	obliqua-sororcula	12	**0.03510**	0.03276	0.04251	**0.771**
	obliqua- turpiniae	6	**0.03481**	0.03258	0.04185	**0.778**
	sororcula- turpiniae	0	0.01747	0.01208	0.03448	0.350
Or7d	fraterculus-obliqua	6	**0.02606**	0.02115	0.04317	0.490
	fraterculus-sororcula	0	0.01263	0.01058	0.01972	0.537
	fraterculus-turpiniae	0	0.01838	0.01646	0.02502	0.658
	obliqua-sororcula	6	**0.02294**	0.01769	0.04127	0.429
	obliqua- turpiniae	8	**0.02031**	0.01248	0.04788	0.261
	sororcula- turpiniae	0	0.01563	0.01340	0.02333	0.575

FD: Number of fixed differences. **K(JC):** Average number of nucleotide substitutions per site between species with Jukes and Cantor correction. **Ka(JC):** Non-synonymous Nucleotide Divergence; **Ks(JC):** Synonymous Nucleotide Divergence.

In general, pairwise comparisons in all genes that involved *A. fraterculus*, *A. turpiniae* and *A. sororcula* had low values of Ka/Ks ratio, with the exception of the comparison between *A. fraterculus* and *A. turpiniae* for the gene *Obp84a*, which had Ka/Ks ratio over 1. More interestingly, for all genes, we only detected fixed differences between species when *A. obliqua* was involved. The number of fixed differences was higher in *Obp28a*, which is not surprising, since this gene seems to have higher rates of non-synonymous substitutions, though we identified fixed differences separating *A. obliqua* from other species here investigated for all genes investigated and for all but a single pairwise comparison.

The analyses of molecular variance (AMOVA) we performed to investigate whether species separation and/or geographical distribution would explain the genetic variance observed in our data indicate (Table 10) that species separation explains better the interpopulational variation than geographical distribution, since we failed to find significant values for the geographic division among regions. Even though we detected a significant effect of species aggregation on the genetic diversity, there is still a significant amount of polymorphism that is still segregating among the species, so much so that *Obp28a* was the loci in which species differences explained the most variation, and it could only account for less than 60% of the total variation, suggesting that there is a percentage of the diversity that is not consistent with the species separation. It is interesting that the two genes that showed signs of positive selection were the ones that explained most of the interspecific variation, whereas the two genes that were considered to be under purifying selection explaining less variation, though still significant.

For any of the genes here investigated, we still have about 50% of the variation still segregating across different species, which could be a consequence of gene flow between the species or even segregation of ancestral polymorphisms given the recent divergence of the species. Recently, a study conducted in our laboratory with 20 different genes found evidence of introgressive hybridization occurring in different directions among *A. fraterculus*, *A. sororcula* and *A. obliqua* (unpublished data). These events of introgression among species of *Anastrepha* could explain a significant fraction of variation still segregating across different species here study.

Table 10. AMOVA analysis results for four olfactory genes contrasting the division of the populations in accordance with species and the geographic division.

SV	by species			
	Obp28a	***Obp84a***	***Or7c***	***Or7d***
Among species	58.21*	38.36*	47.39*	42.90*
Among populations within species	23.34*	35.48*	22.33*	24.80*
Within populations	18.45*	26.16*	30.28*	32.30*
	by regions			
Among regions	17.60	7.68	7.45	14.65
Among populations within regions	62.92*	64.89*	60.22*	51.90*
Within populations	19.49*	27.42*	32.33*	33.46*

SV: Source of variation. *p<0.01

Corroborating previous results, we found high *Fst* values for all pairwise comparisons between *A. obliqua* and the other species (Table 11), with the highest *Fst* value per gene being observed in the *Obp28a*. Once again, this great differentiation found between *A. obliqua* and the other three species suggests that *A. obliqua* is phylogenetically more distinct. Pairwise comparisons between *A. fraterculus* and *A. sororcula* had the lowest average *Fst* values (0.173±0.06), followed by *A. fraterculus* and *A. turpiniae* (0.246±0.14). These results may be an indication of recurrent gene flow still occurring among these species, or it may still be a reflex of the recent divergence, since *Fst* analyses fail to distinguish between historical and recurrent processes.

Table 11. Pairwise *Fst* values of four olfactory genes between species.

		F_{st}				
Species 1	**Species 2**	***Obp28a***	***Obp84a***	***Or7c***	***Or7d***	**Average**
A. fraterculus	*A. obliqua*	0.801*	0.683*	0.768*	0.633*	0.721± 0.08
A. fraterculus	*A. sororcula*	0.149*	0.251*	0.182*	0.111*	0.173±0.06
A. fraterculus	*A. turpiniae*	0.252*	0.145*	0.148*	0.439*	0.246±0.14
A. obliqua	*A. sororcula*	0.771*	0.451*	0.692*	0.599*	0.628±0.14
A. obliqua	*A. turpiniae*	0.793*	0.720*	0.694*	0.587*	0.699±0.09
A. sororcula	*A. turpiniae*	0.339*	0.416*	0.189*	0.366*	0.328±0.10
Overall		0.652	0.498	0.554	0.490	

*P<0.01

The ML phylogenetic inferences of different copies across all species indicate that for the genes *Obp28a* (Figure 3), *Or7c* (Figure 5) and *Or7d* (Figure 6) we can observe the separation of all *A. obliqua* sequences in an unique lineage while there is no separation in exclusive lineages for *A. fraterculus*, *A. sororcula* and *A. turpiniae* for any gene. On the other hand, we failed to find a separate lineage of *A. obliqua* sequences for the gene *Obp84a* (Figure 4), since *A. obliqua* was found along with some sequences of *A. sororcula*. Because other genes showed separated lineages for *A. obliqua* it is possible that these result could be explained by the presence of ancestral polymorphisms in these genes among those species.

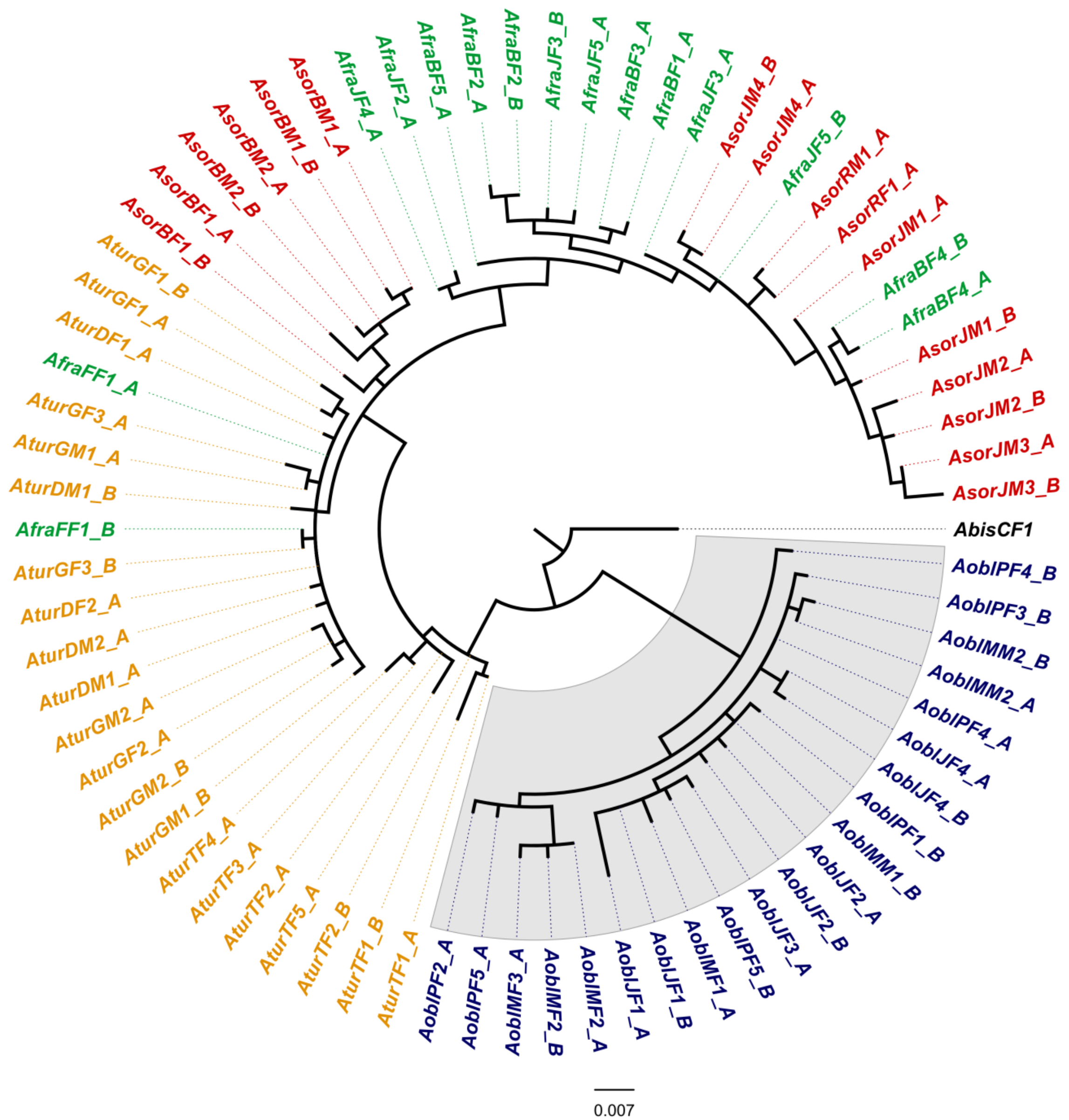


Figure 3. Maximum likelihood phylogenetic relationship among *Obp28a* gene sequences amplified in four species of *Anastrepha* flies. **Abis**: *Anastrepha bistrigata Obp28a*. **Afra** represents *A. fraterculus* sequences. **Aobli** represents *A. obliqua* sequences, **Asor** *A. sororcula* sequences and **Atur** *A. turpinae* sequences. Gray clade indicates the lineage with all *A. obliqua* sequences.

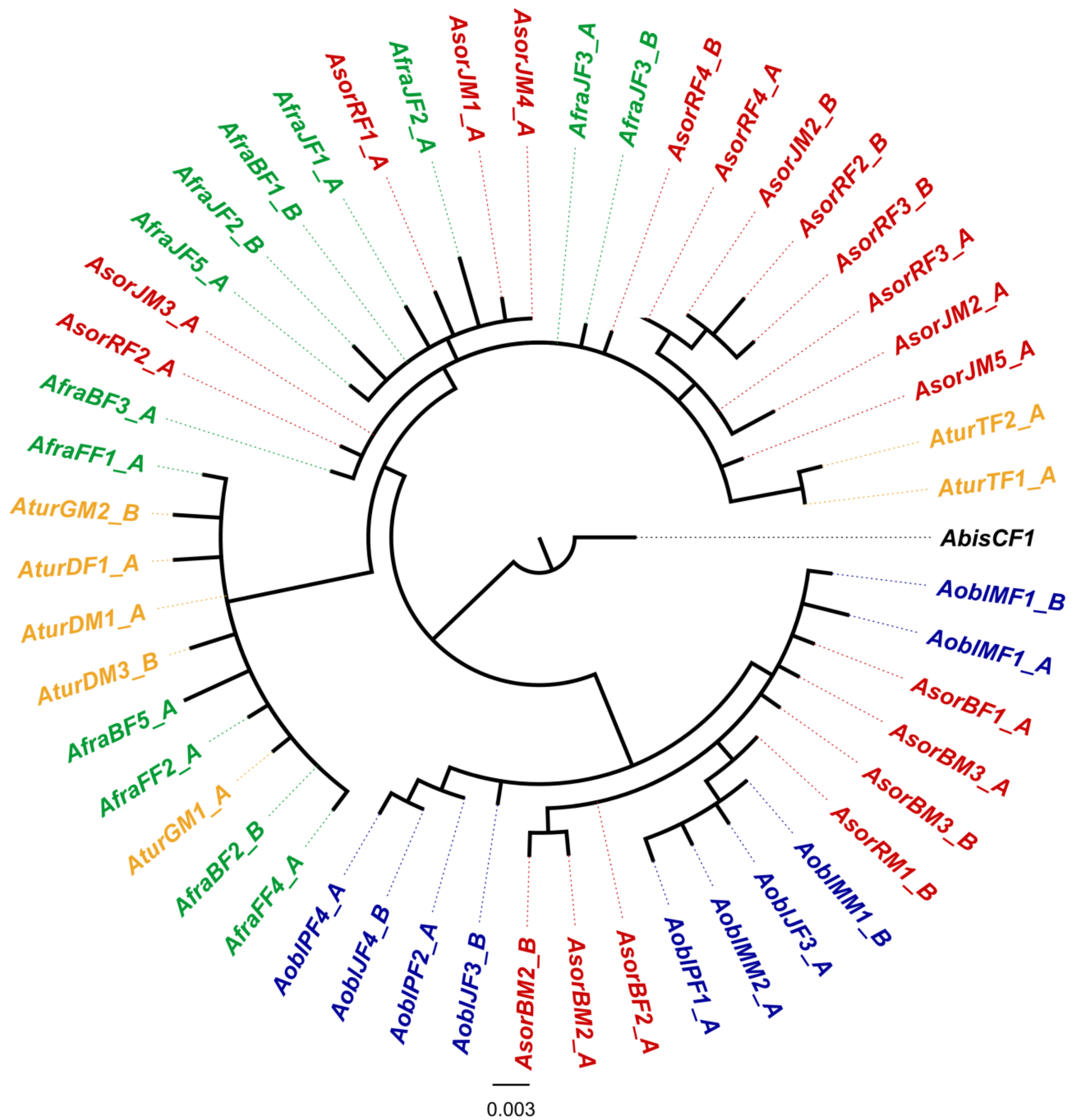


Figure 4. Maximum likelihood phylogenetic relationship among *Obp84a* gene sequences amplified in four species of *Anastrepha* flies. **Abis**: *Anastrepha bistrigata Obp84a*. **Afra** represents *A. fraterculus* sequences. **Aobli** represents *A. obliqua* sequences, **Asor** *A. sororcula* sequences and **Atur** *A. turpiniae* sequences. Gray clade indicates the lineage with all *A. obliqua* sequences.

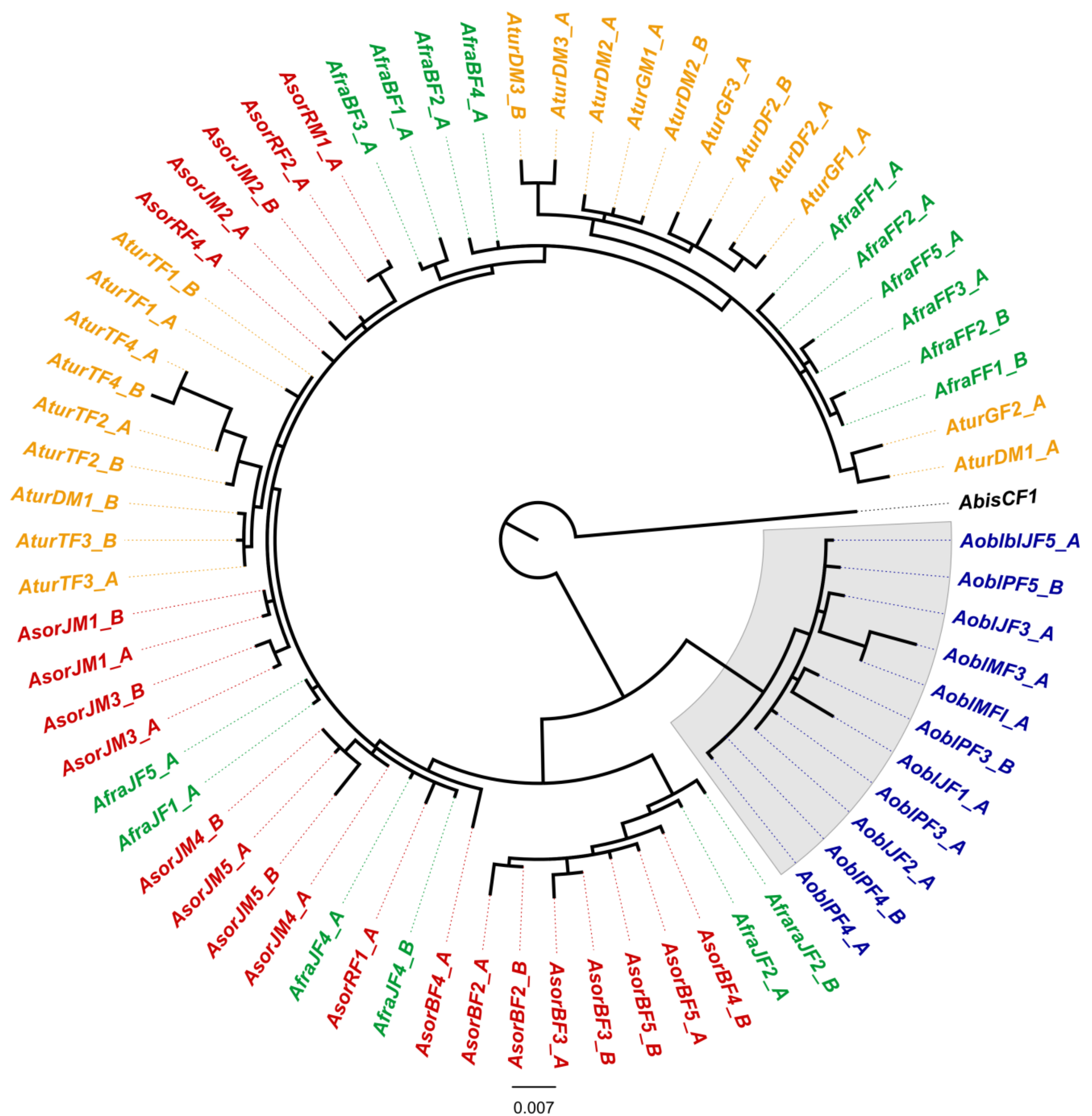


Figure 5. Maximum likelihood phylogenetic relationship among *Or7c* gene sequences amplified in four species of *Anastrepha* flies. **Abis**: *Anastrepha bistrigata Or7c*. **Afra** represents *A. fraterculus* sequences. **Aobli** represents *A. obliqua* sequences, **Asor** *A. sororcula* sequences and **Atur** *A. turpinae* sequences. Gray clade indicates the lineage with all *A. obliqua* sequences.

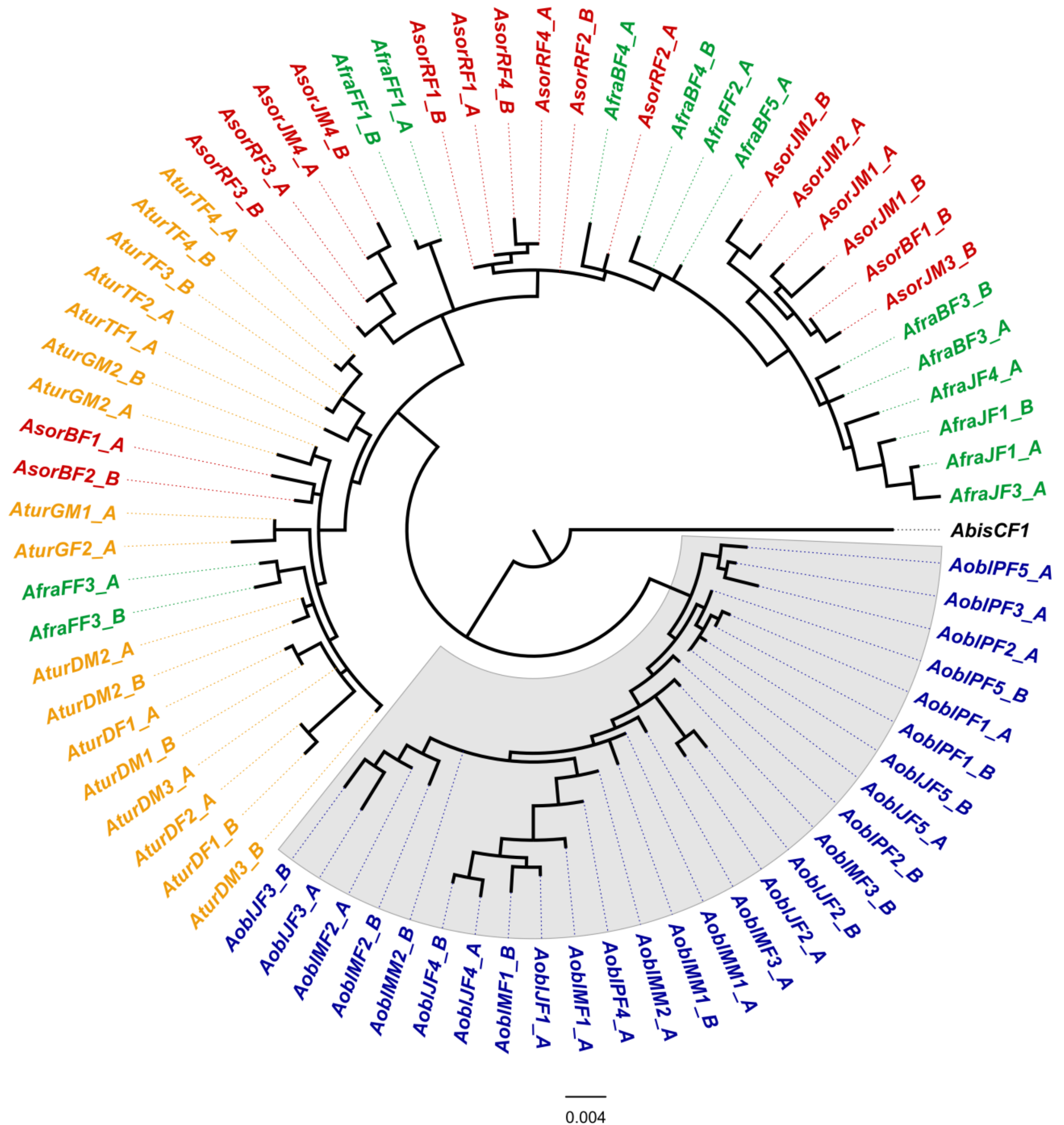


Figure 6. Maximum likelihood phylogenetic relationship among *Or7d* gene sequences amplified in four species of *Anastrepha* flies. **Abis**: *Anastrepha bistrigata Or7d*. **Afra** represents *A. fraterculus* sequences. **Aobli** represents *A. obliqua* sequences, **Asor**, *A. sororcula* sequences and **Atur**, *A. turpinae* sequences. Gray clade indicates the lineage with all *A. obliqua* sequences.

These results taken together, suggest that *A. obliqua* is a more differentiated species, while *A. fraterculus*, *A. sororcula* and *A. turpiniae* are more closely related, or are experiencing higher levels of gene flow that do not allow the total differentiation in genes here investigated. This pattern is consistent with results that suggest that *A. obliqua* is indeed phylogenetically more distantly related, whereas these other species would be phylogenetically closer (BARR; CUI; MCPHERON, 2005; SELIVON; PERONDINI; ROCHA, 2005). This differentiation may also be a consequence of adaptation to different hosts, since even though these species may use a broad range of hosts, *A. obliqua* is preferentially associated with fruits from Anacardiaceae trees, whereas the other species here investigated are more frequently associated with Myrtaceae (BOMFIM; UCHÔA-FERNANDES; BRAGANÇA, 2007; SELIVON, 2000).

Host plants may have special importance in the pheromone production, since there is a similarity among the volatiles emitted by males for female attraction and host plants volatiles and it has been hypothesized that these volatiles may be obtained at first larval instar of males (LIMA-MENDONÇA et al., 2014) providing a mechanism in which host plants can be also influencing the divergence of chemosensory genes. Furthermore, our results revealed several SNPs and haplotypes that distinguish *A. obliqua* from the other species of the group. These data need to be corroborated with a broader sampling, but if confirmed they may have important evolutionary and practical consequences, since it may enable the identification of this species by the use of a set of genetic markers

3.6. Conclusions

In this study we identified a total of five PBP genes in the transcriptomes of *A. fraterculus* and *A. obliqua*, which, despite their high levels of similarity, still showed at least one differentiated gene (*Obp28a*) that had significant signal of positive selection and may be involved in the differences in pheromone induced behaviors between these two species. Divergence and gene flow analysis of *PBPs* and *OR* genes in four closely related species of *Anastrepha* showed that *Obp28a*, *Or7c* and *Or7d* are differentiated in *A. obliqua* but not for *A. fraterculus*, *A. sororcula* and *A. turpiniae* suggesting that the latter are more closely related while *A. obliqua* is the most distant amongst the four species here studied. Phylogenetic analyses were consistent with divergence analysis showing that *A. obliqua* is phylogenetically more differentiated from the other three species for the genes here investigated. Finally the

approach here used, in which we investigated population and species levels of divergence and selection in homologous genes with different rates of evolution, some of which with evidence of positive selection, allowed us to identify genes with fixed differences for at least one species. It is possible that this is a reflection of drift affecting segregation of variation on the gene, but the evidence of positive selection may indicate that these genes could be involved in these species' divergence, and possibly be related with preferences to different host plants or pheromone perception.

3.7. References

ALLEN, J. E.; WANNER, K. W. Asian corn borer pheromone binding protein 3, a candidate for evolving specificity to the 12-tetradecenyl acetate sex pheromone. **Insect Biochemistry and Molecular Biology**, v. 41, n. 3, p. 141–149, mar. 2011.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

ANISIMOVA, M.; BIELAWSKI, J. P.; YANG, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. **Molecular Biology and Evolution**, v. 18, n. 8, p. 1585–1592, ago. 2001.

ARAUJO, E. L.; ZUCCHI, R. A. Medidas do acúleo na caracterização de cinco espécies de Anastrepha do grupo fraterculus (Diptera: Tephritidae). **Neotropical Entomology**, v. 35, n. 3, p. 329–337, jun. 2006.

ARNAL, A. et al. Genetic polymorphism at an odorant receptor gene (Or39) among mosquitoes of the Anopheles gambiae complex in Senegal (West Africa). **BMC Research Notes**, v. 7, n. 1, p. 321, 2014.

BARR, N. B.; CUI, L.; MCPHERON, B. A. Molecular Systematics of Nuclear Gene in Genus Anastrepha (Tephritidae). **Annals of the Entomological Society of America**, v. 98, n. 2, p. 173–180, 1 mar. 2005.

BOMFIM, D. A.; UCHÔA-FERNANDES, M. A.; BRAGANÇA, M. A. L. Hosts and parasitoids of fruit flies (Diptera: Tephritoidea) in the State of Tocantins, Brazil. **Neotropical Entomology**, v. 36, n. 6, p. 984–986, dez. 2007.

CHOMCZYNSKI, P.; SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry**, v. 162, n. 1, p. 156–159, abr. 1987.

DU, G.; PRESTWICH, G. D. Protein Structure Encodes the Ligand Binding Specificity in Pheromone Binding Proteins. **Biochemistry**, v. 34, n. 27, p. 8726–8732, jul. 1995.

EXCOFFIER, L.; LAVAL, G.; SCHNEIDER, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. **Evolutionary Bioinformatics Online**, v. 1, p. 47–50, 2005.

FORET, S.; MALESZKA, R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (Apis mellifera). **Genome Research**, v. 16, n. 11, p. 1404–1413, 25 out. 2006.

FORSTNER, M. et al. Candidate pheromone binding proteins of the silkmoth Bombyx mori. **Invertebrate Neuroscience**, v. 6, n. 4, p. 177–187, 14 nov. 2006.

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

GUINDON, S. et al. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. **Systematic Biology**, v. 59, n. 3, p. 307–321, 1 maio 2010.

HALL. BioEdit: a user-friendly biologicalsequence alignment editor and analysis programfor Windows 95/98/NT. **Nucleic Acids Symposium Series**, 41. p. 95–98, 2 mar. 1999.

HANSSON, B. S.; STENSMYR, M. C. Evolution of Insect Olfaction. **Neuron**, v. 72, n. 5, p. 698–711, dez. 2011.

HARADA, E. et al. Functional Evolution of Duplicated Odorant-Binding Protein Genes, Obp57d and Obp57e, in Drosophila. **PLoS ONE**, v. 7, n. 1, p. e29710, 6 jan. 2012.

HEKMAT-SCAFE, D. S. et al. Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster. **Genome Research**, v. 12, n. 9, p. 1357–1369, set. 2002.

KATOH, K. et al. MAFFT version 5: improvement in accuracy of multiple sequence alignment. **Nucleic Acids Research**, v. 33, n. 2, p. 511–518, 19 jan. 2005.

LAFOREST, S. M.; PRESTWICH, G. D.; LÖFSTEDT, C. Intraspecific nucleotide variation at the pheromone binding protein locus in the turnip moth, Agrotis segetum. **Insect Molecular Biology**, v. 8, n. 4, p. 481–490, nov. 1999.

LIBRADO, P.; ROZAS, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, n. 11, p. 1451–1452, 1 jun. 2009.

LIMA-MENDONÇA, A. et al. Semiochemicals of fruit flies of the genus *Anastrepha* . **Química Nova**, v. 37, n. 2, 2014.

LIS, J. T.; SCHLEIF, R. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. **Nucleic Acids Research**, v. 2, n. 3, p. 383–389, mar. 1975.

LIU, N.-Y. et al. Two subclasses of odorant-binding proteins in Spodoptera exigua display structural conservation and functional divergence. **Insect Molecular Biology**, v. 24, n. 2, p. 167–182, abr. 2015.

MAÏBÈCHE-COISNE, M. et al. Pheromone binding proteins of the moth Mamestra brassicae: Specificity of ligand binding. **Insect Biochemistry and Molecular Biology**, v. 27, n. 3, p. 213–221, mar. 1997.

PLETTNER, E. et al. Discrimination of pheromone enantiomers by two pheromone binding proteins from the gypsy moth Lymantria dispar. **Biochemistry**, v. 39, n. 30, p. 8953–8962, 1 ago. 2000.

POSADA, D. Using MODELTEST and PAUP* to select a model of nucleotide substitution. **Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]**, v. Chapter 6, p. Unit 6.5, fev. 2003.

POSADA, D. jModelTest: Phylogenetic Model Averaging. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1253–1256, 3 abr. 2008.

REZENDE, V. et al. **Candidate genes involved in differentiation between two non-model species of fruit flies (Anastrepha: Tephritidae) screened from head transcriptomes. Unpublished manuscript**. 2016.

ROLLMANN, S. M. et al. Odorant Receptor Polymorphisms and Natural Variation in Olfactory Behavior in Drosophila melanogaster. **Genetics**, v. 186, n. 2, p. 687–697, 1 out. 2010.

SÁNCHEZ-GRACIA, A.; AGUADÉ, M.; ROZAS, J. Patterns of nucleotide polymorphism and divergence in the odorant-binding protein genes OS-E and OS-F: analysis in the melanogaster species subgroup of Drosophila. **Genetics**, v. 165, n. 3, p. 1279–1288, nov. 2003.

SÁNCHEZ-GRACIA, A.; ROZAS, J. Divergent evolution and molecular adaptation in the Drosophila odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. **BMC Evolutionary Biology**, v. 8, n. 1, p. 323, 2008.

SÁNCHEZ-GRACIA, A.; VIEIRA, F. G.; ROZAS, J. Molecular evolution of the major chemosensory gene families in insects. **Heredity**, v. 103, n. 3, p. 208–216, set. 2009.

SELIVON, D. Relações com as plantas hospedeiras. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado.** Malavasi, A.; Zucchi, R.A ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

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

SU, C.-Y.; MENUZ, K.; CARLSON, J. R. Olfactory Perception: Receptors, Cells, and Circuits. **Cell**, v. 139, n. 1, p. 45–59, out. 2009.

SUYAMA, M.; TORRENTS, D.; BORK, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. **Nucleic Acids Research**, v. 34, n. Web Server, p. W609–W612, 1 jul. 2006.

TAN, K. H. et al. Pheromones, Male Lures, and Trapping of Tephritid Fruit Flies. In: SHELLY, T. et al. (Eds.). . **Trapping and the Detection, Control, and Regulation of Tephritid Fruit Flies**. Dordrecht: Springer Netherlands, 2014. p. 15–74.

TEGONI, M.; CAMPANACCI, V.; CAMBILLAU, C. Structural aspects of sexual attraction and chemical communication in insects. **Trends in Biochemical Sciences**, v. 29, n. 5, p. 257–264, maio 2004.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 11 nov. 1994.

VILELA, E.; KOVALESKI, A. Feromônios. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Malavasi, A.; Zucchi, R.A ed. Ribeirão Preto: Holos-FAPESP, 2000. p. 327.

WILLETT, C. S. Evidence for directional selection acting on pheromone-binding proteins in the genus Choristoneura. **Molecular Biology and Evolution**, v. 17, n. 4, p. 553–562, abr. 2000.

XIA, X.; XIE, Z. DAMBE: software package for data analysis in molecular biology and evolution. **The Journal of Heredity**, v. 92, n. 4, p. 371–373, ago. 2001.

YANG, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586–1591, 18 abr. 2007.

ZHANG, J.; NIELSEN, R.; YANG, Z. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. **Molecular Biology and Evolution**, v. 22, n. 12, p. 2472–2479, 24 ago. 2005.

ZHANG, T.-T. et al. Characterization of three pheromone-binding proteins (PBPs) of Helicoverpa armigera (Hübner) and their binding properties. **Journal of Insect Physiology**, v. 58, n. 7, p. 941–948, jul. 2012.

ZHANG, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. **Genomics, Proteomics & Bioinformatics**, v. 4, n. 4, p. 259–263, nov. 2006.

ZUCCHI, R. A. Taxonomia. In: **Moscas-das-frutas de importância econômica no Brasil: conhecimento básico e aplicado**. Ribeirão Preto: Holos-FAPESP, 2000. p. 327p.

4. Considerações finais

Neste trabalho foram usadas duas abordagens principais para alcançar os objetivos propostos. A primeira abordagem consistiu em identificar os genes das famílias gênicas de OR e PBPs nos transcriptomas de *Anastrepha fraterculus* e *A. obliqua* e investigar seus padrões filogenéticos e de evolução molecular com o intuito de compreender como a seleção atua na evolução destes genes assim como identificar genes com altas taxas evolutivas. A utilização de dados dos transcriptomas de alguns tecidos em certos estádios do desenvolvimento, e o fato de alguns desses genes estudados apresentarem baixos níveis de expressão limitou a quantidade de genes que poderiam ser identificados no caso da família gênica dos ORs. Essas famílias gênicas têm taxas evolutivas muito altas associadas a um padrão de nascimento e perda (*birth and death*) de genes que torna inclusive comum a existência de genes específicos para uma espécie. A não identificação completa do repertório de ORs para cada espécie pode ter limitado a determinação da existência de genes específicos para as espécies aqui estudadas, mas não podemos deixar de mencionar também a grande proximidade evolutiva entre as mesmas. Alguns dos genes ortólogos entre as duas espécies foram encontrados incompletos o que não permitiu realizar comparações de identidade e de Ka/Ks desse genes e também não foram considerados nos analises de PAML, por isto, é possível que o número de genes com menor percentagem de identidade e com evidência de seleção positiva seja maior entre as duas espécies estudadas. Contudo, usando os dados dos transcriptomas, foi possível a identificação de um número de PBPs e ORs que representou de forma completa as principais linhagens dos tefrítideos, a detecção de genes ORs com evidência de seleção positiva e também o encontro de genes PBPs e ORs com diferenciação entre as espécies *A. fraterculus* e *A. obliqua*. Estudos baseados no genoma destas espécies poderiam aprofundar e complementar os resultados encontrados neste trabalho.

A segunda abordagem consistiu no estudo da variação genética inter e intraespecífica de quatro genes quimiossensoriais, os quais foram identificados na primeira abordagem, em quatro espécies de moscas do gênero *Anastrepha*. Este estudo teve como objetivo avaliar quanto da variação presente nesses genes é compartilhada entre as espécies e quanto é específica para cada espécie. Os valores mais altos de divergência foram encontrados em *A. obliqua* os quais corresponderam com a presença de diferenças fixas para os quatro genes entre *A. obliqua* e as outras espécies. Estes resultados mostram que os genes *Obp28a*, *Or7c* e *Or7d* apresentam grande diferenciação em *A.obliqua* e no caso dos genes *Obp28* e *Or7c* a diferenciação parece estar influenciada por seleção positiva. Nós propomos que esses

resultados têm que ser corroborados com uma amostragem maior uma vez que as diferenças fixas encontradas também podem ser um reflexo da deriva genética agindo nas populações aqui estudadas. Seria interessante também estudar a variação de outros genes da família dos ORs uma vez que poderiam ser encontrados resultados similares para outros genes. Tem sido encontrado que padrões de diversidade genética nos genes quimiossensoriais podem estar afetando a diferenciação das populações e das espécies uma vez que alguns polimorfismos têm sido relacionados com variação em respostas aos estímulos químicos (LAFOREST; PRESTWICH; LÖFSTEDT, 1999; ROLLMANN et al., 2010). Porém, estudos funcionais que avaliem se a divergência e a variação encontrada nos genes de *A. obliqua* estão relacionadas com mudanças na proteína que de fato levem a mudanças na recepção de moléculas de odor nesta espécie têm que realizados.

4.1. Referências

LAFOREST, S. M.; PRESTWICH, G. D.; LÖFSTEDT, C. Intraspecific nucleotide variation at the pheromone binding protein locus in the turnip moth, Agrotis segetum. **Insect Molecular Biology**, v. 8, n. 4, p. 481–490, nov. 1999.

ROLLMANN, S. M. et al. Odorant Receptor Polymorphisms and Natural Variation in Olfactory Behavior in Drosophila melanogaster. **Genetics**, v. 186, n. 2, p. 687–697, 1 out. 2010.

5. Anexos

5.1. Supplementary Material Chapter 2.

Supplementary Material 1. Odorant receptors genes from *Drosophila melanogaster, Ceratitis capitata* and *Bactrocera dorsalis* used in the phylogenetic analysis.

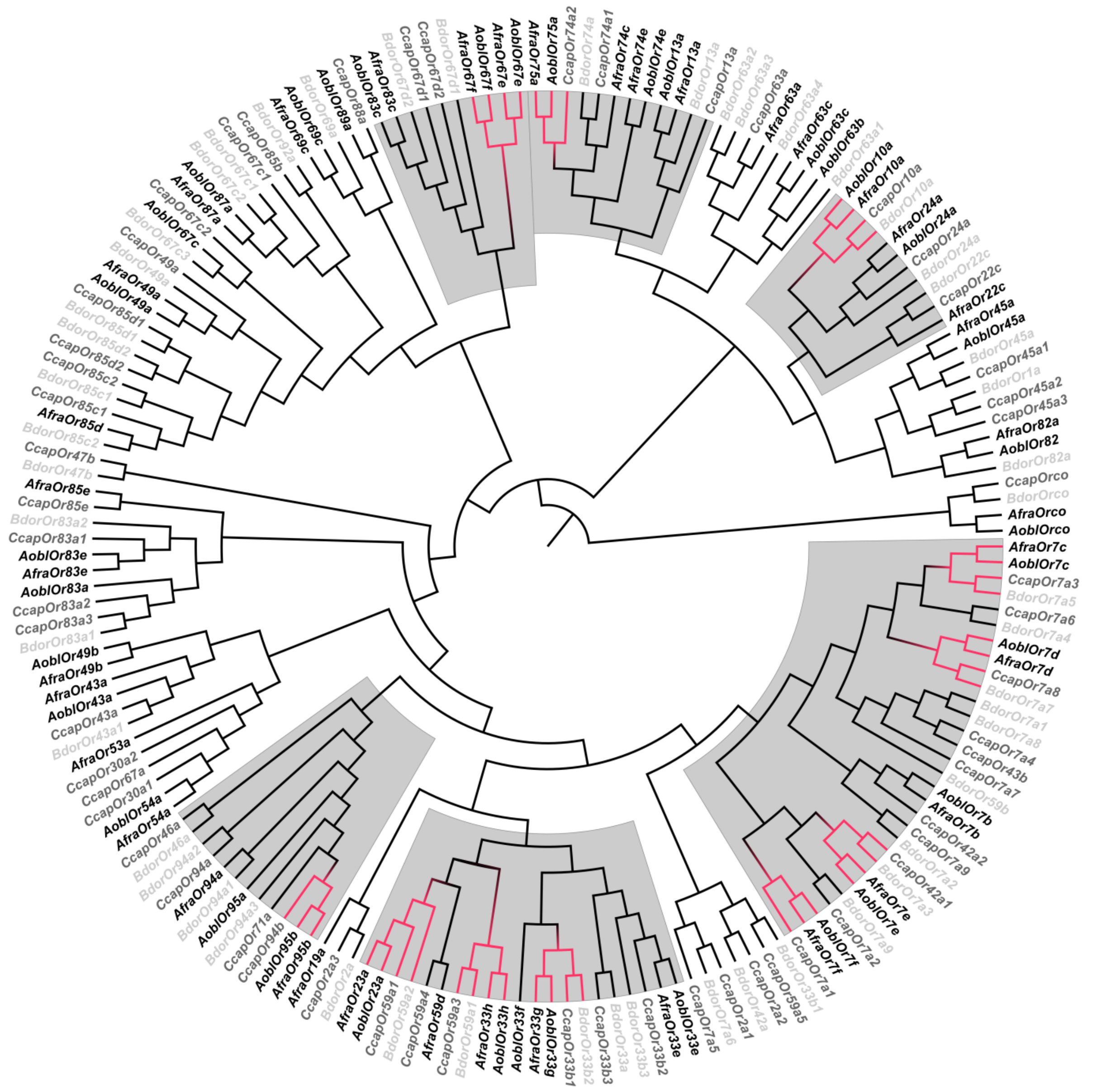
Specie	Gene	Accession Number
D. melanogaster	*DmelOr1a*	CG17885
D. melanogaster	*DmelOr2a*	CG3206
D. melanogaster	*DmelOr7a*	CG10759
D. melanogaster	*DmelOr9a*	CG15302
D. melanogaster	*DmelOr10a*	CG17867
D. melanogaster	*DmelOr13a*	CG12697
D. melanogaster	*DmelOr19a*	CG18859
D. melanogaster	*DmelOr22a*	CG12193
D. melanogaster	*DmelOr22b*	CG4231
D. melanogaster	*DmelOr22c*	CG15377
D. melanogaster	*DmelOr23a*	CG9880
D. melanogaster	*DmelOr24a*	CG11767
D. melanogaster	*DmelOr30a*	CG13106
D. melanogaster	*DmelOr33a*	CG16960
D. melanogaster	*DmelOr33b*	CG16961
D. melanogaster	*DmelOr33c*	CG5006
D. melanogaster	*DmelOr35a*	CG17868
D. melanogaster	*DmelOr42a*	CG17250
D. melanogaster	*DmelOr42b*	CG12754
D. melanogaster	*DmelOr43a*	CG1854
D. melanogaster	*DmelOr43b*	CG17853
D. melanogaster	*DmelOr45a*	CG1978
D. melanogaster	*DmelOr45b*	CG12931
D. melanogaster	*DmelOr46a*	CG17849
D. melanogaster	*DmelOr46b*	CG17848
D. melanogaster	*DmelOr47a*	CG13225
D. melanogaster	*DmelOr47b*	CG13206
D. melanogaster	*DmelOr49a*	CG13158
D. melanogaster	*DmelOr49b*	CG17584
D. melanogaster	*DmelOr56a*	CG12501
D. melanogaster	*DmelOr59a*	CG9820
D. melanogaster	*DmelOr59b*	CG3569
D. melanogaster	*DmelOr59c*	CG17226
D. melanogaster	*DmelOr63a*	CG9969
D. melanogaster	*DmelOr65a*	CG32401
D. melanogaster	*DmelOr65b*	CG32402
D. melanogaster	*DmelOr65c*	CG32403
D. melanogaster	*DmelOr67a*	CG1252
D. melanogaster	*DmelOr67b*	CG1417
D. melanogaster	*DmelOr67c*	CG1415
D. melanogaster	*DmelOr67d*	CG1415

D. melanogaster	*DmelOr69a*	CG17902
D. melanogaster	*DmelOr69b*	CG17902
D. melanogaster	*DmelOr71a*	CG17871
D. melanogaster	*DmelOr74a*	CG13726
D. melanogaster	*DmelOr82a*	CG31519
D. melanogaster	*DmelOr83a*	CG10612
D. melanogaster	*DmelOrco*	CG10609
D. melanogaster	*DmelOr83c*	CG15581
D. melanogaster	*DmelOr85a*	CG7454
D. melanogaster	*DmelOr85b*	CG11735
D. melanogaster	*DmelOr85c*	CG17911
D. melanogaster	*DmelOr85d*	CG11742
D. melanogaster	*DmelOr85e*	CG9700
D. melanogaster	*DmelOr85f*	CG16755
D. melanogaster	*DmelOr88a*	CG14360
D. melanogaster	*DmelOr92a*	CG17916
D. melanogaster	*DmelOr94a*	CG17241
D. melanogaster	*DmelOr94b*	CG6679
D. melanogaster	*DmelOr98a*	CG5540
C. capitata	*CcapOr10a*	XM_004534819
C. capitata	*CcapOr13a*	XM_004529762
C. capitata	*CcapOr22c*	XM_004531471.1
C. capitata	*CcapOr24a*	XM_004518525
C. capitata	*CcapOr2a1*	XM_004537593.1
C. capitata	*CcapOr2a2*	XM_004537591.1
C. capitata	*CcapOr2a3*	XM_004536724.1
C. capitata	*CcapOr30a1*	XM_004534137
C. capitata	*CcapOr30a2*	XM_004526987.1
C. capitata	*CcapOr33b1*	XM_004530334
C. capitata	*CcapOr33b2*	XM_004530333.1
C. capitata	*CcapOr33b3*	XM_004530332.1
C. capitata	*CcapOr42a1*	XM_004531114
C. capitata	*CcapOr42a2*	XM_004521270.1
C. capitata	*CcapOr43a*	XM_004533760
C. capitata	*CcapOr43b*	XM_004535380.1
C. capitata	*CcapOr45a1*	XM_004534410
C. capitata	*CcapOr45a2*	XM_004534409.1
C. capitata	*CcapOr45a3*	XM_004534408.1
C. capitata	*CcapOr46a*	XM_004526279.1
C. capitata	*CcapOr47b*	XM_004522668.1
C. capitata	*CcapOr49a*	XM_004520820.1
C. capitata	*CcapOr59a1*	XM_004520758
C. capitata	*CcapOr59a3*	XM_004537916.1
C. capitata	*CcapOr59a4*	XM_004526514.1
C. capitata	*CcapOr59a5*	XM_004537590.1
C. capitata	*CcapOr63a*	XM_004527036.1
C. capitata	*CcapOr67a*	XM_004534136.1
C. capitata	*CcapOr67c1*	XM_004535510
C. capitata	*CcapOr67c2*	XM_004521019
C. capitata	*CcapOr67d1*	XM_004533380

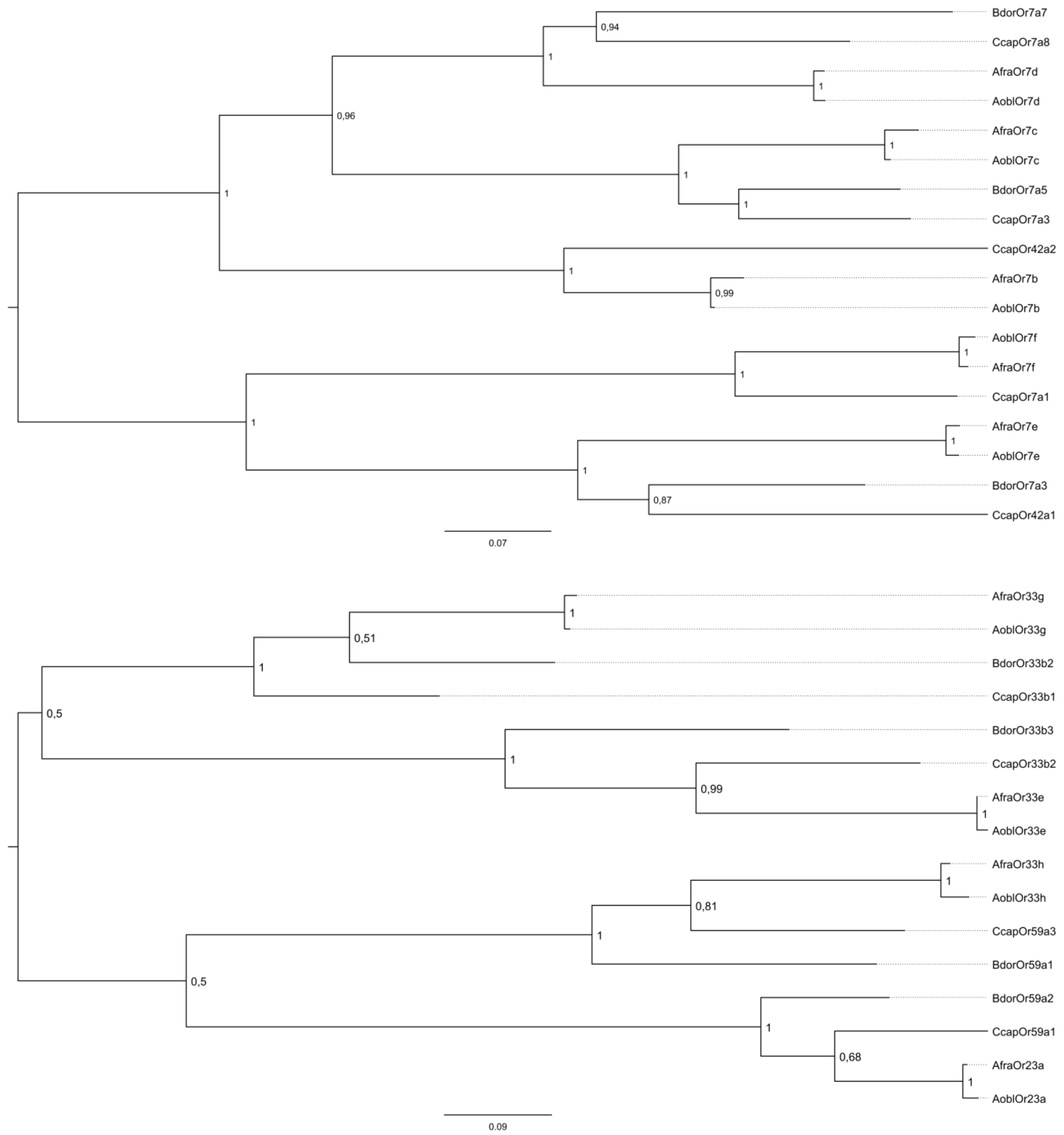
C. capitata	*CcapOr67d2*	XM_004533381
C. capitata	*CcapOr71a*	XM_004520766
C. capitata	*CcapOr74a1*	XM_004531456
C. capitata	*CcapOr74a2*	XM_004520297
C. capitata	*CcapOr7a1*	XM_004531116
C. capitata	*CcapOr7a2*	XM_004531115
C. capitata	*CcapOr7a3*	XM_004536451
C. capitata	*CcapOr7a4*	XM_004518243.1
C. capitata	*CcapOr7a5*	XM_004537592.1
C. capitata	*CcapOr7a6*	XM_004520851.1
C. capitata	*CcapOr7a7*	XM_004535381.1
C. capitata	*CcapOr7a8*	XM_004521261.1
C. capitata	*CcapOr7a9*	XM_004530090.1
C. capitata	*CcapOr83a1*	XM_004518257
C. capitata	*CcapOr83a2*	XM_004518258
C. capitata	*CcapOr83a3*	XM_004518259.1
C. capitata	*CcapOr85b*	XM_004537588.1
C. capitata	*CcapOr85c1*	XM_004518292.1
C. capitata	*CcapOr85c2*	XM_004518291.1
C. capitata	*CcapOr85d1*	XM_004518294.1
C. capitata	*CcapOr85d2*	XM_004518293.1
C. capitata	*CcapOr85e*	XM_004519791
C. capitata	*CcapOr88a*	XM_004535957
C. capitata	*CcapOr94b*	XM_004520765
C. capitata	*CcapOr94a*	XM_004531523
C. capitata	*CcapOrco*	XM_004518006
B. dorsalis	*BdorOr10a*	XM_011207927.1
B. dorsalis	*BdorOr13a*	XM_011206127.1
B. dorsalis	*BdorOr1a*	XM_011214144.1
B. dorsalis	*BdorOr22c*	XM_011211808.1
B. dorsalis	*BdorOr24a*	XM_011201220.1
B. dorsalis	*BdorOr2a*	XM_011200088.1
B. dorsalis	*BdorOr33a*	XM_011209636.1
B. dorsalis	*BdorOr33b1*	XM_011210598.1
B. dorsalis	*BdorOr33b2*	XM_011209638.1
B. dorsalis	*BdorOr33b3*	XM_011209637.1
B. dorsalis	*BdorOr42a*	XM_011210597.1
B. dorsalis	*BdorOr43a1*	XM_011215810.1
B. dorsalis	*BdorOr45a*	XM_011214145.1
B. dorsalis	*BdorOr46a*	XM_011202211.1
B. dorsalis	*BdorOr47b*	XM_011211685.1
B. dorsalis	*BdorOr49a*	XM_011214129.1
B. dorsalis	*BdorOr59a1*	XM_011201169.1
B. dorsalis	*BdorOr59a2*	XM_011199384.1
B. dorsalis	*BdorOr59b*	XM_011210599.1
B. dorsalis	*BdorOr63a1*	XM_011206774.1
B. dorsalis	*BdorOr63a2*	XM_011203515.1
B. dorsalis	*BdorOr63a3*	XM_011203514.1
B. dorsalis	*BdorOr63a4*	XM_011203513.1
B. dorsalis	*BdorOr67c1*	XM_011202099.1

B. dorsalis	*BdorOr67c2*	XM_011202098.1
B. dorsalis	*BdorOr67c3*	XM_011200850.1
B. dorsalis	*BdorOr67d1*	XM_011205402.1
B. dorsalis	*BdorOr67d2*	XM_011205401.1
B. dorsalis	*BdorOr69a*	XM_011211067.1
B. dorsalis	*BdorOr74a*	XM_011201817.1
B. dorsalis	*BdorOr7a1*	XM_011214679.1
B. dorsalis	*BdorOr7a2*	XM_011212210.1
B. dorsalis	*BdorOr7a3*	XM_011211270.1
B. dorsalis	*BdorOr7a4*	XM_011200916.1
B. dorsalis	*BdorOr7a5*	XM_011200418.1
B. dorsalis	*BdorOr7a6*	XM_011210596.1
B. dorsalis	*BdorOr7a7*	XM_011210425.1
B. dorsalis	*BdorOr7a8*	XM_011205865.1
B. dorsalis	*BdorOr7a9*	XM_011211139.1
B. dorsalis	*BdorOr82a*	XM_011210430.1
B. dorsalis	*BdorOr83a1*	XM_011205570.1
B. dorsalis	*BdorOr83a2*	XM_011205479.1
B. dorsalis	*BdorOr85c1*	XM_011211273.1
B. dorsalis	*BdorOr85c2*	XM_011211274.1
B. dorsalis	*BdorOr85d1*	XM_011211276.1
B. dorsalis	*BdorOr85d2*	XM_011211275.1
B. dorsalis	*BdorOr92a*	XM_011210517.1
B. dorsalis	*BdorOr94a1*	XM_011213450.1
B. dorsalis	*BdorOr94a2*	XM_011203454.1
B. dorsalis	*BdorOr94a3*	XM_011207197.1
B. dorsalis	*BdorOrco*	XM_011205476.1

Supplementary Material 2. Phylogenetic tree showing relationships between *C. capitata* (*CcapOr*), *B. dorsalis* (*BdorOr*) *A. fraterculus* (*AfraOr*) and *A. obliqua* (*AoblOr*) OR genes. Highlighted clades represents lineage-specific (sub-lineages) subdivisions used in PALM for the evolutionary analysis. Red branches represent the paralogous groups tested for positive selection in the different branch-site test analysis.

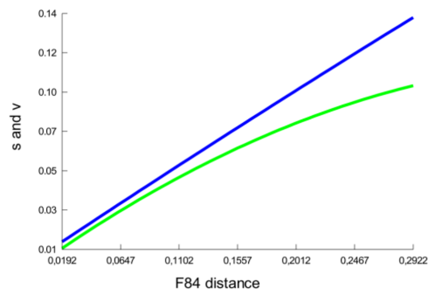


Supplementary Material 3. Neighbor joining phylogenetic trees for the two genic expansions used in PRIME analysis. Node labels represent branch support (Bootstrapp)

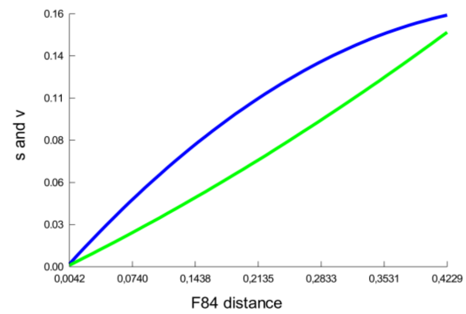


Supplementary Material 4. Plot of estimate numbers of transitions and transversions against F84 distance for the pairwise comparisons in the alignment used for Ka/Ks analysis of *OR* genes of *A. fraterculus*, *A. obliqua*, *C. capitata* and *B. dorsalis* performed in DAMBE 5.5.24

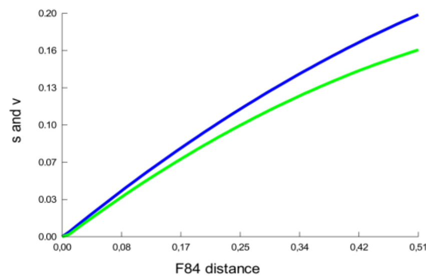
a) ***Or7c***



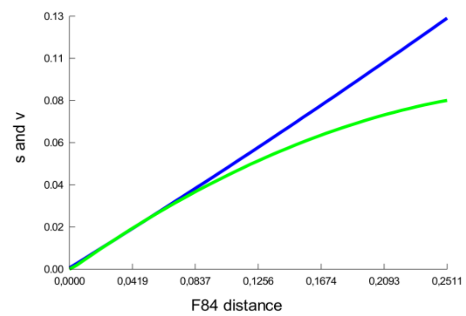
b) ***Or7c***



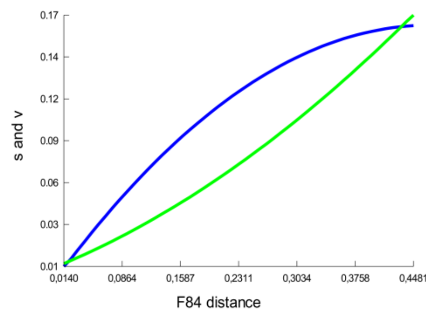
c) ***Or7e***



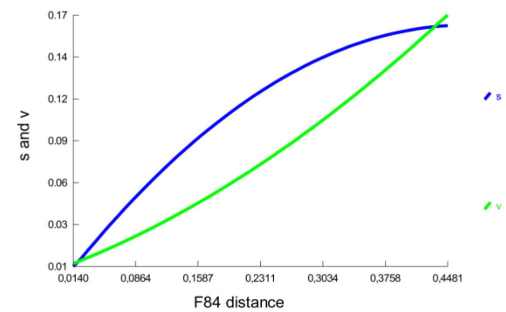
d) ***Or10a***



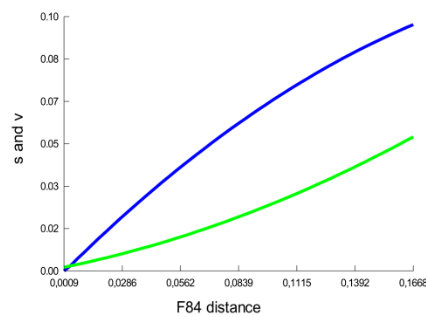
e) ***Or33g***



f) ***Or33h***



g) ***Orco***

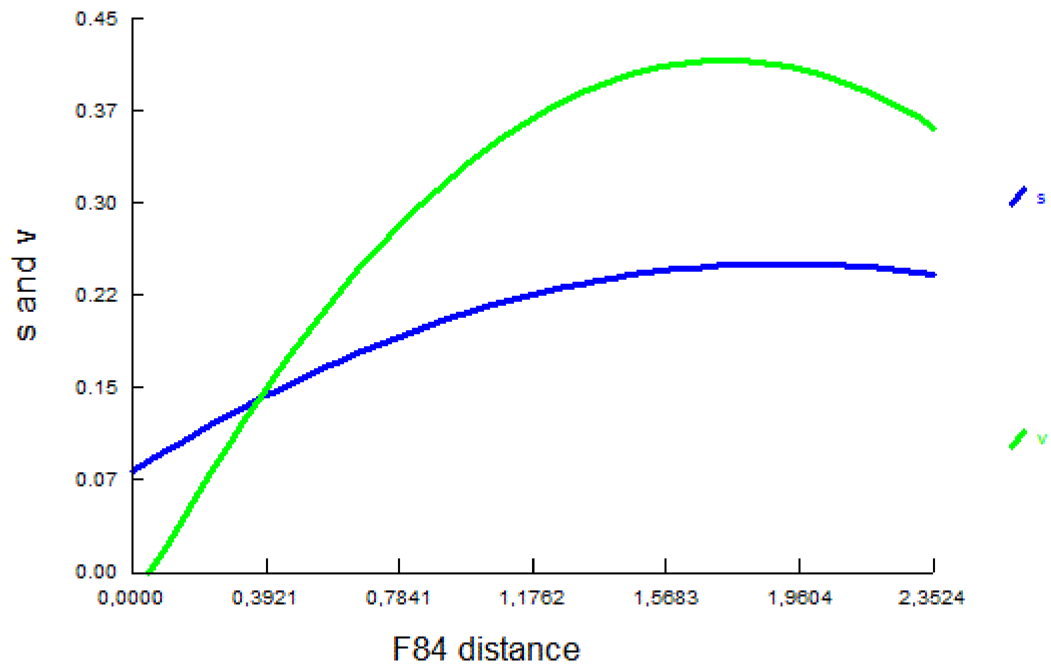


5.2. Supplementary Material Chapter 3.

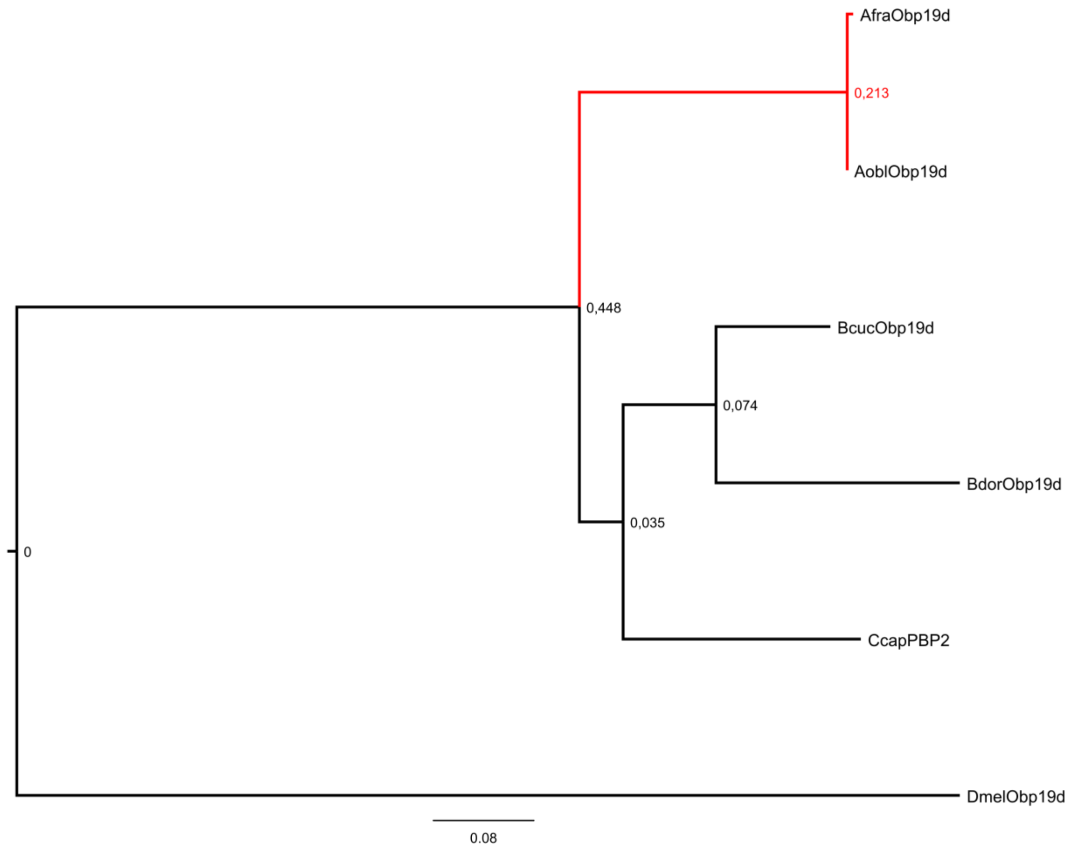
Supplementary Material 1. Pheromone binding proteins sequences from *Drosophila melanogaster*, *Ceratitis capitata*, *Bactrocera dorsalis* and *Bactrocera cucubirtae* used in phylogenetic analysis.

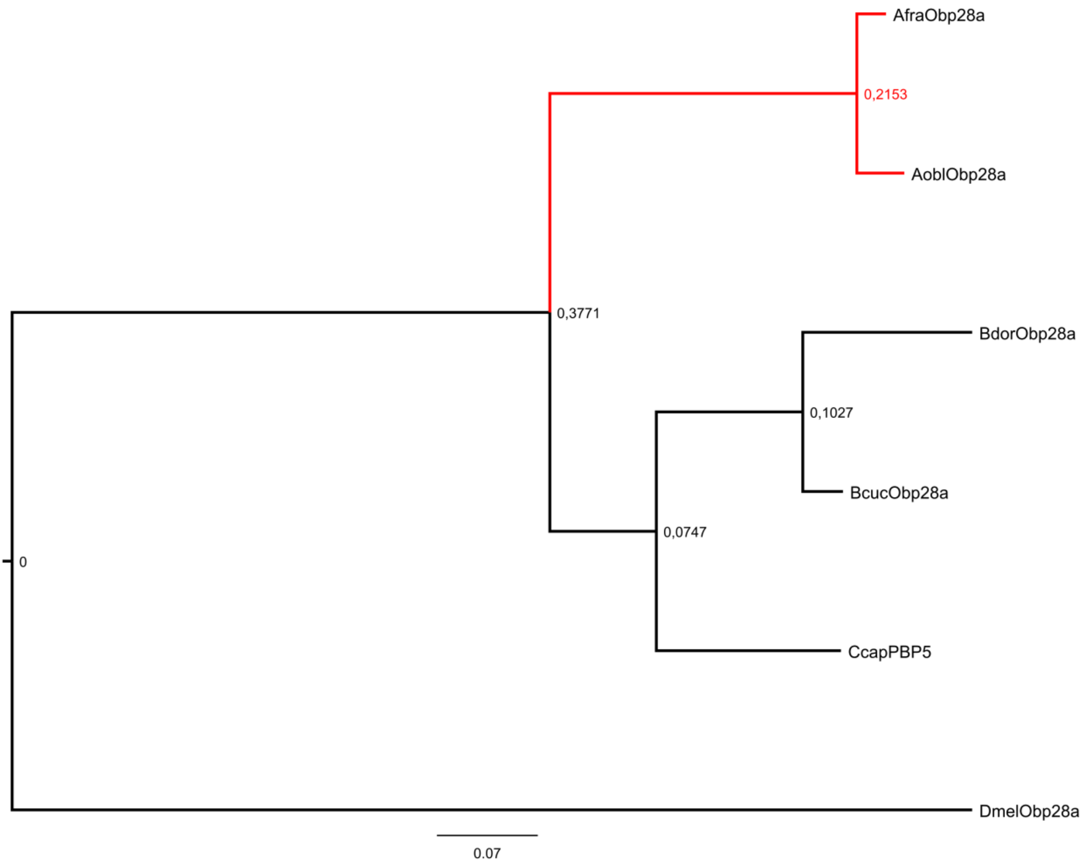
Specie	Gene	Accession Number
D. melanogaster	*DmelObp19d*	NM_078697.3
D. melanogaster	*DmelObp28a*	NM_078781.3
D. melanogaster	*DmelObp69a*	NM_079315.2
D. melanogaster	*DmelObp83a*	NM_079517.3
D. melanogaster	*DmelObp84a*	NM_057642.3
C. capitata	*CcapPBP2*	XM_004524978.2
C. capitata	*CcapPBP5*	XM_004524959.2
C. capitata	*CcapPBP1*	NM_001308406.1
C. capitata	*CcapPBP6*	XM_004523388.2
C. capitata	*CcapPBP4*	XM_004529312.2
B. dorsalis	*BdorObp19d*	XM_011213059.1
B. dorsalis	*BdorObp28a*	XM_011200518.1
B. dorsalis	*BdorObp69a*	KP743698.1
B. dorsalis	*BdorObp83a*	XM_011214170.1
B. dorsalis	*BdorObp84a*	XM_011200114.1
B. cucurbitae	*BcucObp19d*	XM_011188942.1
B. cucurbitae	*BcucObp28a*	XM_011194845.1
B. cucurbitae	*BcucObp83a*	XM_011193369.1
B. cucurbitae	*BcucObp84a*	XM_011186495.1

Supplementary Material 2. Plot of estimate numbers of transitions and transversions against F84 distance for the pairwise comparisons in the alignment of de *PBP* genes of *A. fraterculus*, *A. obliqua*, *C. capitata*, *B. dorsalis*, *B.cucurbitae* and *D. melanogaster* performed in DAMBE 5.5.24.

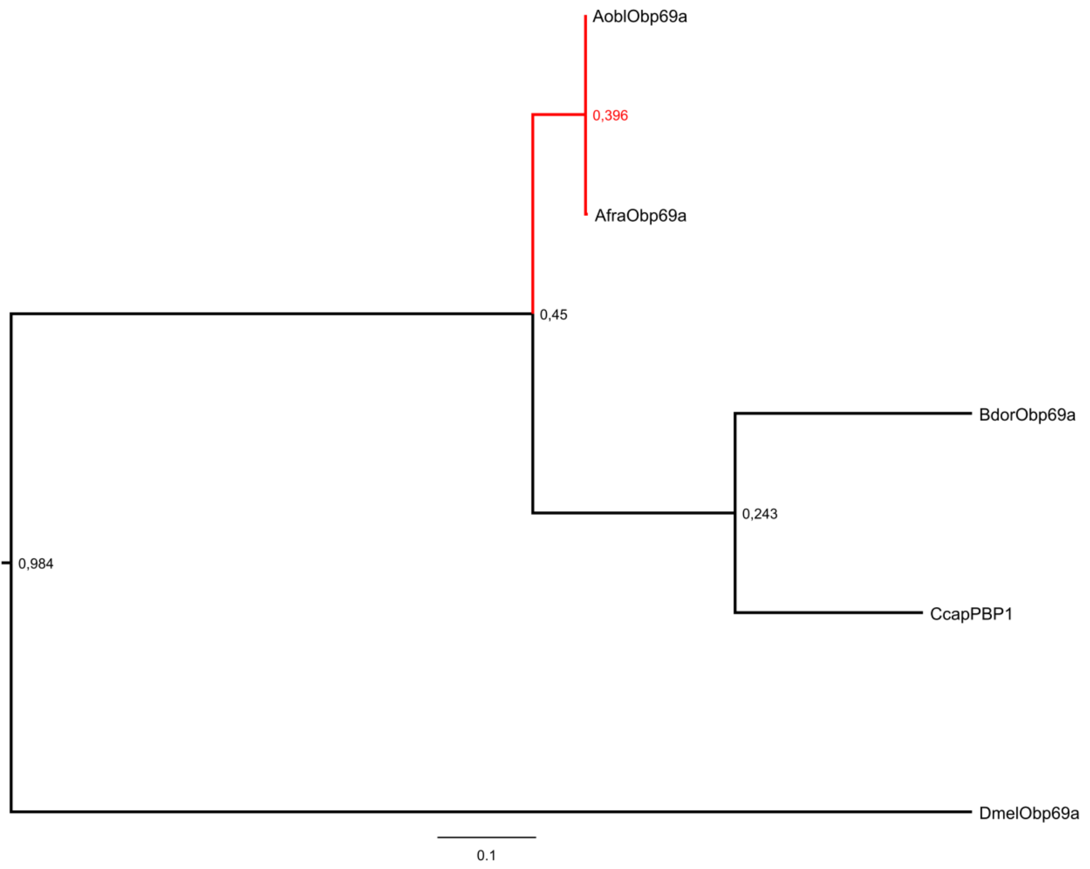


Supplementary Material 3. Lineage-specific non-rooted phylogenetic trees used in PALM for the evolutionary analysis. Red branches represent the Anastrepha homologous genes tested for positive selection in the different branch-site test analysis. Branch labels represent branch lengths estimated as amino acid substitutions per site.

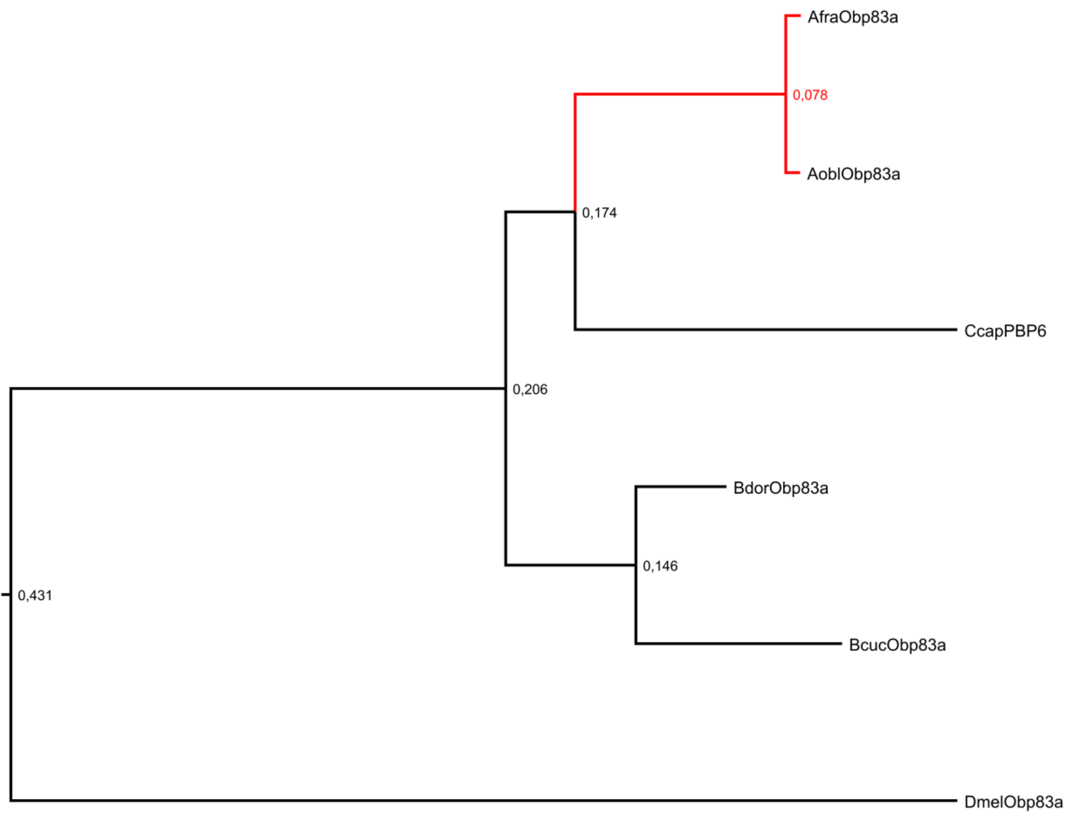


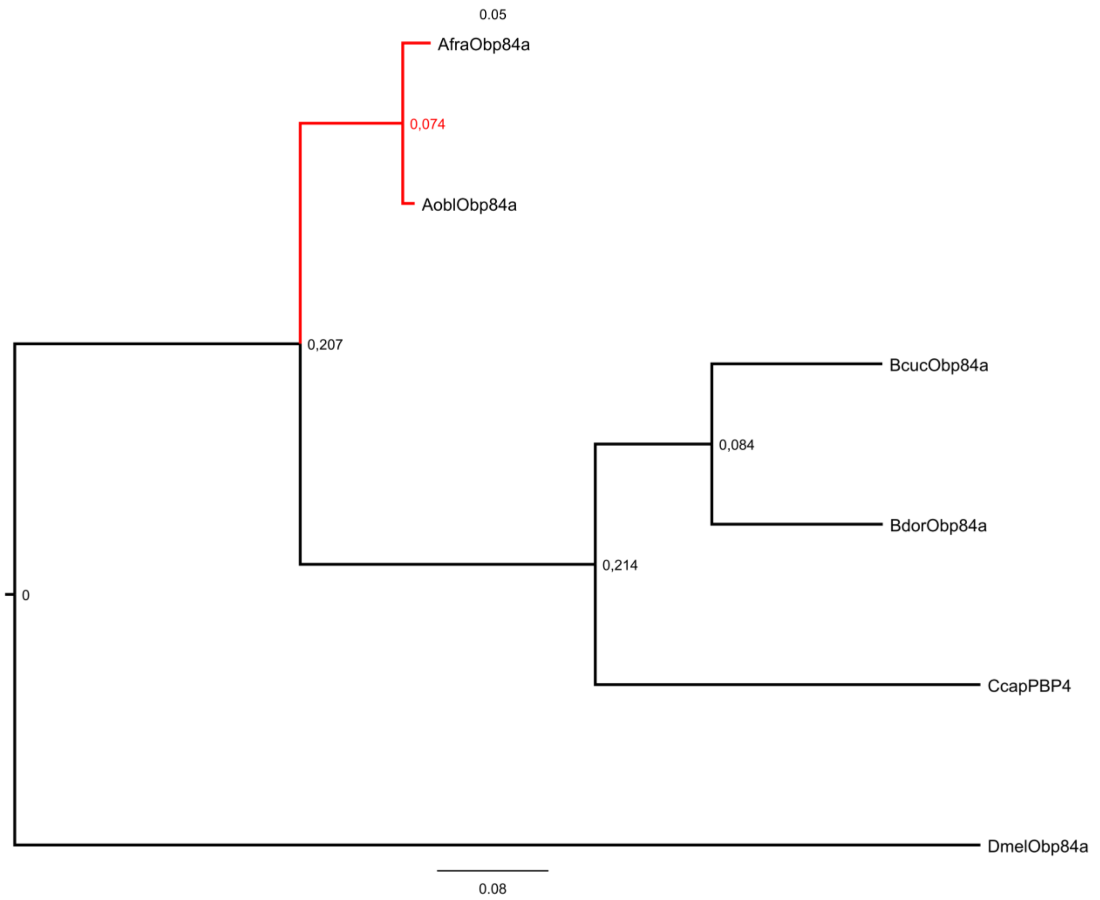


AfraObp28a
0,2153
AoblObp28a
0,3771
BdorObp28a
0,1027
BcucObp28a
0,0747
0
CcapPBP5
DmelObp28a
0.07



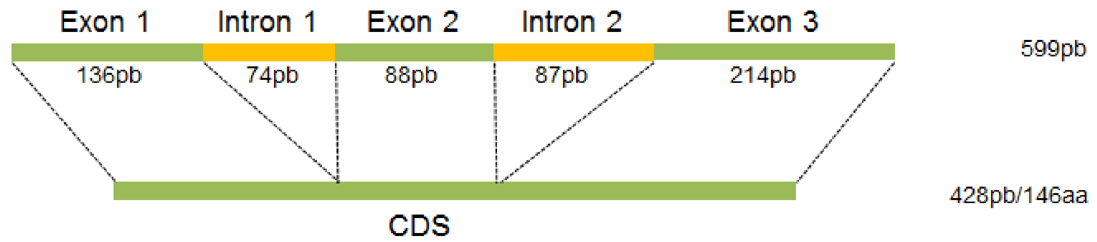
AoblObp69a
0,396
AfraObp69a
0,45
BdorObp69a
0,243
0,984
CcapPBP1
DmelObp69a
0.1


AfraObp83a
0,078
AoblObp83a
0,174
CcapPBP6
0,206
BdorObp83a
0,146
BcucObp83a
0,431
DmelObp83a
0.05

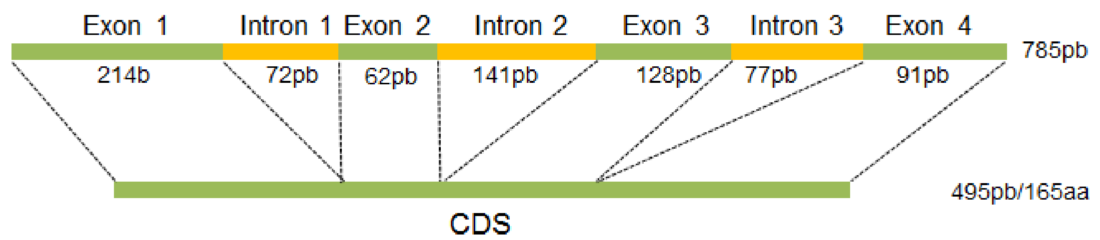

AfraObp84a
0,074
AoblObp84a
0,207
BcucObp84a
0,084
BdorObp84a
0,214
0
CcapPBP4
DmelObp84a
0.08

Supplementary Material 4. Structure and lengths of chemosensory genes amplified in *Anastrepha* species.

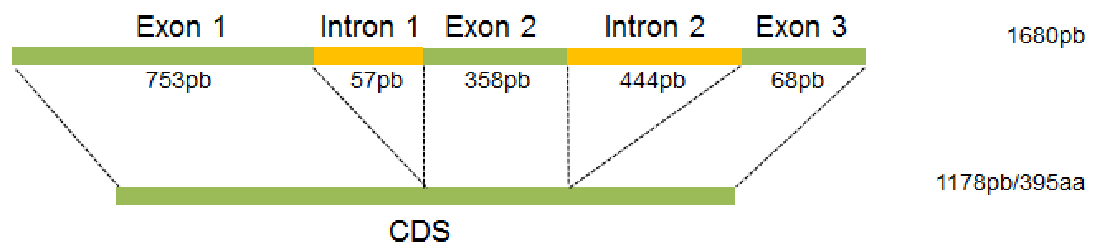
a) Obp28a



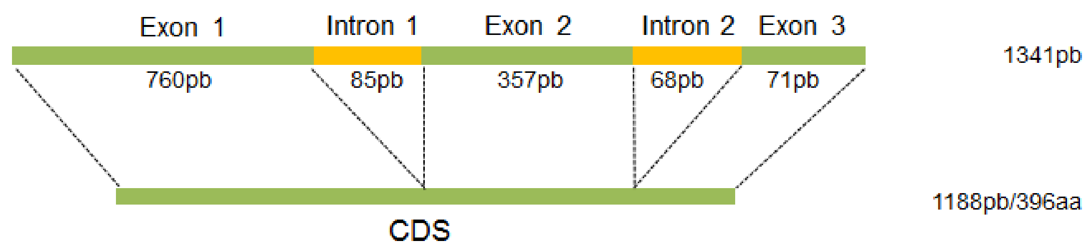
b) Obp84a



c) Or7c



d) Or7d



Supplementary Material 5. Pheromone binding proteins found in *A. farterculus* and *A. obliqua* transcriptomes.

```
                        10         20         30         40         50         60         70         80
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp19d  1   MRFLNIFLIL CAAFISYAEC HDSEKARAVA NECKDEVGAT DADVDSMFNH EPAGSSEAKC LHACVMKRFG LLNDEGKMDK
AoblObp19d  1   .......... .......... .......... .......... .......... .......... .......... ..........

                        90        100        110        120        130        140
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ..
AfraObp19d  81  EKALDILEKI HGDDEEQQNL GKEVVEACGD IEVDEDHCEA AEEYRMCIHG KADENGFKLG RV
AoblObp19d  81  .......... .......... .......... .......... .......... ..E....... ..


                        10         20         30         40         50         60         70         80
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp28a  1   MAKLILIAIF CILSGTLSKA FNKEEAIKQF ITRMEECREE VGAASSDIEE LVKKVPAAGK EGKCLRACLM KKYGVMNDGG
AoblObp28a  1   .......... ......I.E. ..R.....K. M..LG..... ....P..... .I..L..... .......... ........D.

                        90        100        110        120        130        140
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|.
AfraObp28a  81  KFIKAVALEH AATYADGDET KMKTATEIID ACGGTAVPDD PCEAAEVYGH CFMEQAKAHG IDKFEF
AoblObp28a  81  .......... ....S..... .....N.... .......... .....A..S. ..I...I... LE....


                        10         20         30         40         50         60         70         80
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp69a  1   MNYKQTLFLL LVYQCYFLAR VKALEIPPHM KSGAKKLSNI CMKEIGLTED MFQEVRKTGQ LPSDSRFKCF LHCMLDKIGL
AoblObp69a  1   .......... .......... .......... .......... .......... .......... .......... ..........

                        90        100        110        120        130        140
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|..
AfraObp69a  81  IDKENIVHLD NILELLPPEF LPIIEQLHTT CGTKSGADGC ETAFLTVECY INTNPVILKL VFVTFSD
AoblObp69a  81  .......... .......... .......... .......... .......... .......... .......


                        10         20         30         40         50         60         70         80
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp83a  1   MVSNGIWRAQ AFHVFLIVAL SSSLTLMHVQ AQEPRRDDKW PPPAVLKMAK VFHDICVGKT GVSEEAIKEF SDGQIHDDEA
AoblObp83a  1   ..L....... ..Y.L..... .......... .......... .......... .......... ..T....... ..........

                        90        100        110        120        130        140        150
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
AfraObp83a  81  LKCYMNCLFH EIDVVDDNGD VHLETLYDTV PGTVRDKLIN MAKDCVHPEG DTLCHKAWWF HQCWKKADPV HYFLP
AoblObp83a  81  .......... .......... .......N.. .......... .......... .......... .......... .....


                        10         20         30         40         50         60         70         80
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp84a  1   MKMAKSNAFN RLIFLLVLCC CKNLVATQKR AKDNGDIYVQ HKEKVAIGNN SSVIDVEEVM QICNASFSVP MEYIVQFNTT
AoblObp84a  1   .......... .......... .......... .......... ...Q...... .......... .......... ..........

                        90        100        110        120        130        140        150        160
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
AfraObp84a  81  SELSDTTDKT GMCFIRCFFE KSGLIKDWQL NTDLIMQNIM PIPKSSVQFC ESEAKNEVNA CVRTYAIAKC LMKSALENAR
AoblObp84a  81  G...N..... .......... .......... .......... ..TT...... .......M.. .......... ....V.....

                ....|
AfraObp84a  161 NNTTV
AoblObp84a  161 .....
```