Development Genes and Evolution (2024) 234:153–171
https://doi.org/10.1007/s00427-024-00723-3

RESEARCH

# Molecular evolution and genetic diversity of *defective chorion 1* in *Anastrepha fraterculus* and *Anastrepha obliqua* (Diptera, Tephritidae)

**Mariana de Cássia Bisio[1] · Edyane Moraes dos Santos[1] · Camilla Alves Santos[2] · Samira Chahad-Ehlers[1] · Reinaldo Alves de Brito[1]**

Received: 26 April 2024 / Accepted: 17 October 2024 / Published online: 7 November 2024

## Abstract

The family Tephritidae comprises numerous fruit fly species, some of which are economically significant, such as several in the genus *Anastrepha*. Most pest species in this genus belong to the *fraterculus* group, characterized by closely related species that are difficult to differentiate due to recent divergence and gene flow. Identifying genetic markers for their study is paramount for understanding the group's evolution and eventual phytosanitary control. Because there is variation in eggshell morphology among species in the genus, the study of the rapidly evolving *defective chorion 1* (*dec-1*) gene, which is crucial for chorion formation and reproduction, could provide relevant information for *Anastrepha* differentiation. We compared transcriptome sequences of *dec-1* from two of the most important pest species in the genus, *Anastrepha fraterculus* and *Anastrepha obliqua* to *dec-1* sequences from *Anastrepha ludens*, which was used for structure prediction. Furthermore, we amplified a conserved exon across populations of these species. These data revealed three alternative transcripts in *A. fraterculus* and *A. obliqua*, consistent with patterns found in other Tephritidae; we obtained orthologous sequences for these other tephritids from NCBI to investigate patterns of selection affecting this gene at different hierarchical levels using different methods. These analyses show a general pattern of purifying selection across the whole gene and throughout its history at different hierarchical levels, from populations to more distantly related species. That notwithstanding, we still found evidence of positive and episodic diversifying selection at different levels. Different parts of the gene have shown distinct evolutionary rates, which were associated with the diverse proproteins produced by posttranslational changes of DEC-1, with proproteins that are incorporated in the chorion earlier in egg formation being in general more conserved than others that are incorporated later. This correlation appears more evident in certain lineages, including the branch that separates *Anastrepha*, as well as other internal branches that differentiate species within the genus. Our data showed that this gene shows remarkable variation across its different exons, which has proven to be informative at different evolutionary levels. These changes hold promise not only for studying differentiation in *Anastrepha* but also for the eventual management of selected pest species.

**Keywords** *dec-1* · *Anastrepha* · Chorionic genes · Differentiation · Molecular evolution

---

Communicated by Siegfried Roth

Mariana de Cássia Bisio and Edyane Moraes dos Santos contributed equally to this work.

✉ Reinaldo Alves de Brito
brito@ufscar.br

Mariana de Cássia Bisio
decassiabisio@gmail.com

Edyane Moraes dos Santos
edyanemoraes@hotmail.com

Camilla Alves Santos
csantos.alvess@gmail.com

Samira Chahad-Ehlers
schahad@gmail.com

[1] Departamento de Genética e Evolução, Universidade Federal de São Carlos, Via Washington Luis Km 235, São Carlos, SP 13565-905, Brazil

[2] Departamento de Genética e Biologia Evolutiva, Instituto de Biociências - Universidade de São Paulo., São Paulo, SP 05508-090, Brazil

## Introduction

The dipteran family Tephritidae harbors approximately 4930 described species divided into 500 genera distributed worldwide (Norrbom 2004; Schutze et al. 2017). Among these, a few fruit fly species, particularly in the genera *Anastrepha*, *Bactrocera*, *Ceratitis*, and *Rhagoletis*, are of significant interest because they are among the most important agricultural pests (White 2006; Raza et al. 2020) (Zucchi et al. 2004). *Anastrepha*, the largest genus in the tribe Toxotrypanini (Norrbom et al. 2012), is composed of 330 species in 23 taxonomic groups distributed across the tropics and subtropics of the Americas (Norrbom et al. 2018, 2021). The *fraterculus* species group in *Anastrepha*, which comprises 34 species formally described to date (Norrbom et al. 2012), includes two widely distributed and closely related species: the South American fruit fly *Anastrepha fraterculus* (Wiedmann 1830) and the West Indies fruit fly *Anastrepha obliqua* (Macquart 1835). These species are known for infesting several host fruits across a broad geographical range, though the former tends to prefer Myrtaceae, whereas the latter favors Anacardiaceae (Solferini and Morgante 1987; Zucchi 2000).

Due to their host diversity and extensive spatial distribution, *A. fraterculus* and *A. obliqua* may experience genetic and ecological forces that promote differentiation across their distribution. Understanding these forces is paramount for investigating the processes behind species formation in this group, but it has also become increasingly important for planning and implementing population control strategies (Ruiz-Montoya et al. 2020). It has been suggested that reproductive isolation may be an essential factor leading to speciation (Mayr 1959; Rabosky 2016), so much so that it has been commonly employed as a yardstick for defining and delineating species. Others have suggested that species are formed by a combination of forces that define separately evolving metapopulation lineages (De Queiroz 2007). While there is evidence of pre- and post-zygotic isolation among some sympatric and allopatric populations of *Anastrepha* species (Selivon et al. 2005; Vera et al. 2006; Rull et al. 2018), hybridization between *A. fraterculus* and *A. obliqua* has successfully produced viable offspring in the laboratory, though following Haldane's rule (Santos et al. 2001). This finding has been corroborated by population studies that indicate gene flow between these species, despite the presence of reproductive barriers (Díaz et al. 2018; Scully et al. 2018; Congrains et al. 2021, 2023). In this scenario, it is possible that different parts of the genome may introgress differently, so genes with specific reproductive biases may play a crucial role in differentiation and speciation, making their study of particular relevance for understanding the underlying forces separating *Anastrepha* species.

Genes associated with egg formation have been suggested to play important roles in the differentiation of some groups, such as Drosophilidae (Eickbush and Izzo 1995) and even *Anastrepha* (Gonçalves et al. 2013). In the *Drosophila melanogaster* subgroup, these genes have been implicated in adaptive processes to the environment, with chorionic proteins evolving more rapidly, especially those in the outer membrane (chorion), showing strong positive selection compared to the inner (vitelline) membrane proteins (Jagadeeshan and Singh 2007), a pattern that was also found among *Anastrepha* species (Gonçalves et al. 2013). This selective pressure may foster rapid gene diversification and reproductive incompatibilities (Findlay and Swanson 2010), similar to sexual selection, sexual conflict, immunity, and protein recognition (Swanson and Vacquier 2002). The importance of studying egg formation genes among *Anastrepha* species is reinforced by the fact that egg morphology exhibits significant variability and may be used as diagnostic to distinguish eggs from several *Anastrepha* species, even among different morphotypes of *A. fraterculus* (Selivon and Perondini 1998; Selivon et al. 2004; Figueiredo et al. 2013, 2017).

The *defective chorion 1* (*dec-1*) gene has been extensively studied in *Drosophila melanogaster*, where it encodes follicular proteins (*fcl*) (Andersson and Lambertsson 1991) and plays a crucial role in several stages of choriogenesis (Waring et al. 1990). The DEC-1 protein and its derivatives are essential for the formation of insect eggshells, and mutations in the *dec-1* gene can result in gross eggshell deformities during morphogenesis, as well as in sterile females (Bauer and Waring 1987). This gene produces at least three alternatively spliced mRNAs with different isoforms in *Drosophila* that encode the post-translationally modified proproteins FC106, FC125, and FC177 (Waring et al. 1990; Noguerón et al. 2000; Badciong et al. 2001; Mauzy-Melitz and Waring 2003; Fakhouri et al. 2006). FC106 is the most abundant proprotein (Mauzy-Melitz and Waring 2003) and is present in all eggshell layers, with a protein derivative located in the inner chorionic layer (ICL) (Noguerón et al. 2000). This process is related to several mechanisms of oogenesis, such as organization (Noguerón et al. 2000; Badciong et al. 2001), stability (Waring 2000), and physicochemical protection of the egg (Badciong et al. 2001).

Together with FC125 and FC177, the proprotein FC106 has been shown to be essential for functional egg formation and female fertility (Noguerón et al. 2000; Mauzy-Melitz and Waring 2003). The "FC" in their names denotes that they are synthetized and secreted by follicular cells and subsequently stored in the vitelline membrane and chorion layers, where they are cleaved into at least five polypeptide products (Noguerón et al. 2000). In *Drosophila*, *dec-1* contains five tandem copies of a 78 bp-long sequence followed by seven repeats of varying lengths and identities

(Waring et al. 1990; Escher and Lambertsson 1996). The structural complexity and distinct spatial distributions of *dec-1* products, which are involved in egg structure and function (Mauzy-Melitz and Waring 2003), are a reflection of several gene pre- and post-translational changes that lead to different DEC-1 protein variants (Hawley and Waring 1988), providing evidence that this gene is rapidly evolving (Badciong et al. 2001). Even though phylogenetic studies demonstrate a high degree of conservation of *dec-1* general organization among *Drosophila* species, there is extensive divergence in part of its coding regions (Badciong et al. 2001). Mutagenesis experiments targeting conserved regions have demonstrated the importance of this gene in comparative evolutionary analyses (Hanna-Rose et al. 1997; Takahashi et al. 1999), supported by the presence of conserved transcribed regions both within and between species, as observed in *Drosophila* (Andersson and Lambertsson 1991).

Despite its general conservation, understanding the underlying genetic and evolutionary mechanisms that control the presence of multiple expressed DEC-1 isoforms within and between individuals, as observed in *Drosophila* species, remains a significant challenge. The primary objective of this study was to expand the knowledge on the *dec-1* gene in *Anastrepha* and Tephritidae, particularly regarding its role in their differentiation. To address this knowledge gap, we investigated whether there is a variation in this gene in *Anastrepha*, and our findings provide insight into *Anastrepha* diversification and systematics. In this context, we investigated the *dec-1* structure and phylogenetic signals by investigating rates of evolution and checking for potential sites under positive and purifying selection across different evolutionary levels. This investigation extends beyond the scope of the studied Tephritidae to include Diptera taxa for which there is information on the *dec-1* gene. By exploring potentially influential genes involved in fundamental biological processes, such as oogenesis and reproduction, in a phylogenetic framework, this research provides relevant insights not only into *Anastrepha* differentiation but also into broader-scale evolutionary processes affecting this gene.

## Materials and methods

### Identification of the *defective chorion 1* gene in *Anastrepha*

*The Drosophila melanogaster dec-1* sequence (*Dmdec-1*) (Flybase FBgn0000427) was used as a query to identify putatively homologous transcripts in the *A. fraterculus* and *A. obliqua* reproductive and head transcriptomes previously produced in our laboratory (Gonçalves et al. 2013; Rezende et al. 2016; Congrains et al. 2018, 2021).

The genomic sequence of *dec-1* was obtained by mining a previous draft of the *A. ludens* genome, generously shared by Dr. Scott Geib prior to its publication, which was masked for repeats and low-complexity regions (Dupuis et al. 2019); this sequence is now published (Sim et al. 2024) and available in GenBank (GCA_028408465.1). Genome prediction and annotation were performed using Funannotate (v. 1.8.9) (Palmer and Stajich 2019), which performs ab initio prediction with Augustus (v. 3.3.3) (Keller et al. 2011), Snap (v. 2006–07-28) (Korf 2004), and GlimmerHMM (v. 3.0.4) (Majoros et al. 2004). Augustus was trained using protein sequences from *Rhagoletis zephyria*, another Tephritidae species, along with Diptera BUSCO models (Odb10) (Manni et al. 2021) to enhance the identification of protein-coding regions. Following gene model prediction, protein-coding sequences in this pipeline were annotated by querying curated databases such as UniProt (UniProt_Consortium 2019), with a cutoff value of $1e10^{-3}$, and Gene Ontology (Ashburner et al. 2000) for protein function and domain identification. These putative gene models derived from the *A. ludens* genome were then used to identify and manually mine genomic sequences putatively orthologous to *dec-1* (*Aluddec-1*). The *A. obliqua* genome has also recently been made available in GenBank (NCBI accession number GCA_027943255.1) (Sim et al., 2024), and its annotation indicated two *dec-1* (*GenAoblidec-1*) isoforms, fc106 and fc125 (XM_054883939.1 and XM_054883939.1, respectively).

### Characterization and molecular structure of dec-1

Putative *dec-1* sequences and potential splicing variants were identified using *A. ludens* gene sequences as queries in BLAST searches (Altschul and Koonin 1998) against the NCBI nr and sra databases. We used *R. zephyria dec-1* sequences (NCBI gene ID: 108,362,233) as templates to identify *dec-1* exons in *A. fraterculus* and *A. obliqua* obtained from transcriptomes produced in our laboratory. The inferred nucleotide coding sequences of *A. ludens*, *A. fraterculus*, and *A. obliqua* (*Aluddec-1*, *Afradec-1*, and *Aobldec-1*, respectively) were translated into amino acid sequences using the ExPASy Translate tool (https://web.expasy.org/translate/). DNA sequences were aligned using MAFFT (Katoh and Standley 2013) implemented on BioEdit (Hall 2004). Motif identification was performed using the MOTIF server (http://www.genome.jp/tools/motif/) with Pfam domain settings. Finally, motif similarity between *Afradec-1* and its putative orthologs was estimated using the SMS tool (https://www.bioinformatics.org/sms2/ident_sim.html).

### Molecular procedures

#### Population samples and fly stocks

Specimens of *A. fraterculus* and *A. obliqua* were derived from fruits collected at eight sampling points distributed across Brazil (Table 1), which were maintained in plastic cages (21 × 13 × 13 cm) with vermiculite until pupation occurred. The cages were maintained under laboratory-controlled conditions of temperature (26 °C ± 5 °C), 65% relative humidity, and natural lighting. Adult flies were allowed to emerge and mature sexually for at least 10–15 days in a new cage with food (yeast, sugar, honey, and hydrolyzed protein) and water ad libitum. Mature adults were identified following the criteria outlined by Malavasi et al. (2000). Some specimens were subsequently stored at −80 °C for molecular analyses. Total DNA extraction was performed with five females from each original population (Table 1) following the standard acid guanidinium thiocyanate-phenol−chloroform protocol (Nelson and Krawetz 1992).

## Sequence analysis and phylogenetic sets

Given the length of *dec-1*, and the fact that several isoforms are produced by alternative splicing across different insect species, we selected a specific region on exon 3 that is common to all isoforms to investigate intraspecific variation and patterns of selection across different individuals and populations. This approach allowed us to investigate intraspecific variation and patterns of selection on a wider sample in a more amenable faction. To do so, primers (dc1_2371F: 5′-ATG TCT ATG GAT AGG CAA ATG -3′/dc1_3236R: 5′-GGA GYG GCC AAG CCT AGC ATA TC-3′) were used to amplify this region in 27 flies from four populations of *A. fraterculus* and *A. obliqua* (Table 1). Amplifications were carried out using a Bio-Rad T100™ Thermocycler programmed for 35 cycles, including an initial denaturation step of 3′ at 94 °C, followed by 34 cycles of 1′ at 94 °C, 90″ at 54.2 °C, and 1′ at 72 °C, with a final extension step of 72 °C for 3′. The amplification products were checked on a 1% agarose gel containing ethidium bromide under UV illumination. Positive PCR products were PEG-purified (polyethylene glycol 8000) (Lis and Schleif 1975) and cloned and inserted into *Escherichia coli* DH5α using an InsTAclone kit (Fermentas) according to the manufacturer's instructions. Positive clones were then amplified with standard M13 primers (M13F: 5′-GTA AAA CGA CGG CCA GT-3′/M13R: 5′-GGA AAC AGC TAT GAC CAT G-3′) following a protocol involving 1′ at 94 °C, 34 cycles of 30″ at 94 °C, 30″ at 56 °C, and 60″ at 72 °C, with a final extension step of 3′ at 72 °C. The products were subsequently subjected to PEG purification, and the resulting clones were Sanger sequenced (approximately 3 sequenced clones per sample) at Macrogen, Inc. (Seoul, Korea). To improve the reliability of our sequence data, we performed multiple independent amplifications and sequencing of each sample and removed singletons.

High-quality sequences derived from amplified clones were visually inspected for each sample. Because of the short size of the amplified product, we sequenced only the M13

**Table 1** Samples of *Anastrepha obliqua* and *Anastrepha fraterculus* collected from different locations and used for population analysis

| Species | Locality | Coordinates | | Samples |
|---|---|---|---|---|
| *A. obliqua* | Capanema/PR | 53°49′ 04″W | 25°39′49"S | Ao_Cap1<br>Ao_Cap2 |
| | Marialva/PR | 51°46′ 53″W | 23°29′48″S | Ao_Mari1<br>Ao_Mari2<br>Ao_Mari3 |
| | Caetité/BA | 42°29′33″W | 14°03′50″S | Ao_Cae1<br>Ao_Cae2<br>Ao_Cae3<br>Ao_Cae4<br>Ao_Cae5 |
| | Passos/MG | 46°37′04″W | 20°43′45″S | Ao_Pas1<br>Ao_Pas2<br>Ao_Pas3<br>Ao_Pas4 |
| *A. fraterculus* | São Thomé das Letras/MG | 45°01′27″W | 21°43′13″S | Af_Tho1<br>Af_Tho2<br>Af_Tho3<br>Af_Tho4<br>Af_Tho5 |
| | Pinheiros/ES | 40°12′47″W | 18°24′28″S | Af_Pin1<br>Af_Pin2<br>Af_Pin3<br>Af_Pin4 |
| | Espírito Santo do Turvo/SP | 49°25′54″W | 22°41′30″S | Af_Tur1<br>Af_Tur2 |
| | Criciúma/SC | 49°23′47″W | 28°40′17″S | Af_Cri1 |

forward primer. When there was variation among clones per sample (most often point changes), we retained only the one with the highest overall chromatographic quality.

To investigate molecular evolution and selection at different levels of taxonomic divergence, we analyzed two separate datasets, one that investigated levels of intra- and interspecific variation in two of the most important *Anastrepha fraterculus* group species and another that investigated patterns of molecular evolution of *dec-1* across Cyclorrhapha (Diptera: Muscomorpha). Both datasets were independently aligned using MAFFT software, employing the L-INS-i method (Katoh et al. 2019).

The "*Anastrepha*" *dataset* was composed of 31 sequences derived from fragments of the third exon of *dec-1* amplified from natural populations of *A. obliqua* and *A. fraterculus* (Table 1), which belong to the two most important branches in the *fraterculus* group (Congrains et al. 2023), as well as *dec-1* sequences from *A. ludens* and *A. grandis*. The latter was included as an outgroup. The reason we included *A. ludens* here is not only because it was derived from the best curated genome available at the time but also because *A. fraterculus* has been considered a species complex, some of which are more related to *A. ludens* than to other *A. fraterculus* (Hernández-Ortiz et al. 2012, 2015; Congrains et al. 2021, 2023).

The "Cyclorrhapha" *dataset* includes complete *dec-1* sequences derived from the transcriptomes of *A. obliqua*, *A. fraterculus*, and *A. grandis*, as well as the *dec-1* genomic sequences of *A. ludens* annotated and curated here. The shorter *dec-1* isoform (referred to as X3 here but representing the FC106 proprotein) was used in this dataset because it was the predominant isoform obtained from GenBank, since several species indicated below lack alternative isoforms. This dataset included not only putative *dec-1* orthologs from *Anastrepha* and other Tephritidae available in GenBank, including *Rhagoletis zephyria*, *Rhagoletis pomonella*, *Zeugodacus curcubitae*, *Bactrocera tryoni*, and *Ceratitis capitata* but also from a few Drosophilidae, such as *Drosophila melanogaster*, *D. pseudoobscura*, and *D. albomicans,* as well as from the Diopsidae *Teleopsis dalmanni*, treated as an outgroup within Cyclorrhapha (Table S1).

## Phylogenetic and selection tests

The use of two distinct datasets enables the examination of diverse selection patterns over different evolutionary periods and divergence levels. These analyses started by investigating saturation levels among putative homologous sequences of each dataset using Xia's saturation test (Xia and Li 1998) on DAMBE 7.3.11, using the GTR evolutionary model to assess genetic distances by comparing transition and transversion rates. This test takes into account saturation levels and transition/transversion rates, aiming to determine whether there was an underestimation of synonymous substitution rates (*dS*) with increasing evolutionary distance.

Phylogenetic relationships using maximum likelihood were inferred in PhyML v.3.0 (Guindon 2010), with the best evolutionary model of nucleotide substitution for each set estimated using ModelTest-ng (Darriba et al. 2019), considering the Bayesian Information Criterion (BIC) (Schwarz, 1978). The topologies of each dataset were visualized with FigTree v.1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/). These phylogenetic inferences were used to investigate patterns of molecular evolution in *dec-1* under two frameworks: CODEML and BEB implemented in PAML software (Yang 1997; Chen et al. 2009) and the HyPhy package in DataMonkey (Weaver et al. 2018; Kosakovsky Pond et al. 2019). For all tested models (CODEML, BEB, and HyPhy), two categories of analysis were employed: a global estimate of gene evolution (i.e., across all taxa within the phylogeny) and a branch-site test, both of which investigate the ratio of non-synonymous (*dN*) to synonymous (*dS*) substitutions. If this *dN/dS* ratio (or $\omega$) is less than 1, the region is under purifying selection; if it is close to 1, it is neutral, and it is considered to be under positive selection when it is greater than 1.

The test for positive selection using the CODEML model considered the framework of the inferred ML topology for each set to contrast different evolutionary models using a hierarchical series of likelihood ratio tests (LRTs). The first test involves contrasting a null model (M0 null) of evolution that does not allow for sites to have $\omega > 1$ (i.e., those that are experiencing positive selection) against an alternative model (MA) that allows for such occurrences. Other tests that make more explicit inferences about the model are hierarchically contrasted: M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8. These comparisons explore different assumptions about the $\omega$ ratio, including constant $\omega$ ratios in M0 (Goldman and Yang 1994; Yang et al. 1998), discrete site classes in M3 (Yang et al. 2000), and positive selection in M2a (Nielsen and Yang 1998; Yang et al. 2005). Further analyses were performed with M7 assuming a beta distribution of $\omega$ below 1 and M8 adding a category with $\omega > 1$, suggesting potential positive selection (Anisimova et al. 2001). In the final comparison, M8a vs. M8 contrasted the presence of an extra codon class with $\omega = 1$ against an additional category with $\omega > 1$, providing a comprehensive assessment of positive selection likelihoods (Swanson et al. 2003; Wong et al. 2004).

Posterior probabilities that sites are under positive or purifying selection were estimated using the Bayes empirical Bayes (BEB) method to provide additional evidence regarding potential regions and codons of *dec-1* evolving in response to selection (Yang et al. 2005). A Likelihood Ratio Test (LRT) estimated the posterior probability that a nucleotide site evolved under positive selection (significant if $p > 0.95$).

The HyPhy package also identifies signatures of positive selection or purifying selection by investigating the *dN/dS* ratio but allows different branches to have distinct $\omega$ values. Multiple methods were employed, specifically FEL (Kosakovsky Pond and Frost 2005) and FUBAR (Murrell et al. 2013), which assume constant selection pressure per site across the phylogeny to infer $\omega$; MEME, which investigates $\omega$ across individual sites on a proportion of branches using a mixed-effects maximum likelihood approach (Murrell et al. 2012); BUSTED, which tests whether at least one branch has experienced positive selection in specific sites (Murrell et al. 2015); and RELAX, which tests for the relaxation or strengthening of natural selection in specific branches (Wertheim et al. 2014). Amino acid replacements were further analyzed by TreeSAAP v.3.2, which investigates whether amino acid changes lead to radical physicochemical changes at the protein level (Woolley et al. 2003) on topologies generated using baseml on PAML (Yang 1997). Methods by Xia and Li (1998) and McClellan and McCracken (2001) were applied to infer amino acid substitution patterns, focusing on significant positive $z$ scores in categories six to eight, indicating the most extreme changes in structure or function. All conceptual references on physicochemical properties were extracted from Gromiha and Selvaraj (1997), Gromiha et al. (1999), Woolley et al. (2003), Novotny et al. (2007), and Beckstead et al. (2009). Departures from neutrality impact the distribution of radical/non-radical changes in tip and interior branches (younger and older lineages, respectively, following the expectations under neutrality and coalescence) the same way as it is for synonymous and non-synonymous substitutions (Takahata 1990; Yang 2002). We tested the difference in this distribution using Fisher's exact test on a $2 \times 2$ contingency table (Templeton 1996), which is a more refined version of the McDonald and Kreitman (1991) used to study intra- and interspecific polymorphisms. We expected that under neutrality the ratio of radical to non-radical changes should be the same in tip haplotypes (younger haplotypes) as in interior haplotypes (older haplotypes). Under purifying selection, we would expect to see fewer radical changes in interior branches, whereas when there were recurrent events of directional positive selection or balancing selection, we would expect more radical changes in the interior rather than on younger nodes, similar to patterns expected for synonymous/non-synonymous substitutions (Fay and Wu 2005; Sobrinho and de Brito 2012). The same contingency tests were used to contrast rates across different portions of the *dec-1* gene to investigate whether they would show different rates of radical to non-radical amino acid changes.

To investigate how eventual putatively selected regions would affect DEC-1 proteins, and whether they would produce a significant change in its secondary and tertiary structure, first, we tested whether this protein is ordered using AIUPred (Mészáros et al. 2018; Erdős and Dosztányi 2024), which estimates IUPred and Anchor values that infer the probability that regions in the protein are disordered or conditionally disordered by interactions and then modelled DEC-1 from *Drosophila melanogaster* (Isoform A) to use as model to infer patterns of other Tephritidae species (isoform X3) using Phyre2 (Kelley et al. 2015), Swiss-Prot (Bienert et al. 2017; Waterhouse et al. 2018), AlphaFold (Jumper et al. 2021), and Biovia Discovery Studio Visualizer software (Dassault_Systèmes 2021).

## Results

### Structural characterization of *A. fraterculus* and *A. obliqua dec-1*

Genomic and transcriptomic data, as well as direct sequencing of PCR products, were employed to determine the *dec-1* genomic sequence in *A. fraterculus* and *A. obliqua*. In the absence of genomic information for these species at the time these experiments were performed, we used genomic sequences from *A. ludens* to validate the transcriptome and sequencing data of *dec-1* exons in *A. obliqua* and *A. fraterculus*. Comparative analysis and alignment of the *A. ludens* genomic data with those of the other two species enabled inference of the presence of six exons in these species, one more than that described for *D. melanogaster*.

Figure 1 provides a graphical representation of the *dec-1* structure, including exon lengths for *Anastrepha fraterculus*. Exon lengths, coding sequences (CDSs), and putative proteins show slight size differences among the three *Anastrepha* species (Table 2). We identified three distinct *dec-1* isoforms in *A. fraterculus* and *A. obliqua*, designated as X1, X2, and X3, following the designations for putatively homologous isoforms of this gene from other Tephritidae flies, such as *R. zephyria* (Gene ID 108362233), *R. pomonella* (Gene ID: 118,735,351), and *C. capitata* (Gene ID: 101,457,983). These three isoforms are closely aligned with isoforms A, B, and C of the six currently described for *D. melanogaster* (Flybase: FBgn0000427).

The first four exons in the 5′-upstream region, along with the canonical AUG initiation codon on exon-1, are shared by all *Anastrepha* isoforms. Differences arise in the arrangement of the last two exons in the 3′ downstream region, particularly at exon 5, which exhibits variations in length due to alternative internal splicing. The putative isoforms generated by alternative splicing in *A. fraterculus* measure 1472 aa for X1, 1192 aa for X2, and 955 aa for the shorter X3 isoform and correspond to the fc125, fc177, and fc106 proproteins, respectively. In *A. obliqua*,

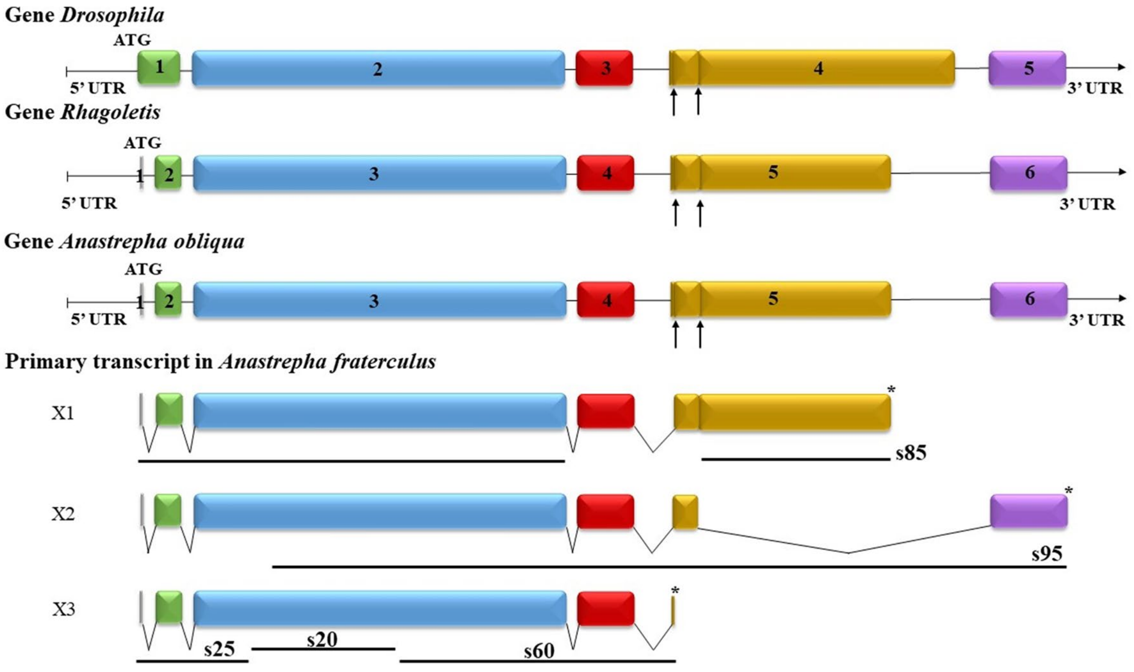

**Fig. 1** Structure of the *dec-1* gene in *Drosophila melanogaster* (NCBI: NM_167133.2), *Rhagoletis zephyria* (NCBI: XM_017615135.1), and *Anastrepha spp*. (this work), as well as characterization of three isoforms in the fruit fly *Anastrepha fraterculus*. The six detected exons are represented proportionally to their sizes in bp. Exon 5 (in yellow) shows internal alternative splicing sites (arrows). Intron positions were predicted based on the *Anastrepha ludens* genome. The inferred locations of the protein derivatives s85, s95, s25, s20, and s60, represented below each *Anastrepha* isoform, were predicted based on data from *D. melanogaster* (Waring et al. 1990; Noguerón and Waring 1995; Noguerón et al. 2000; Mauzy-Melitz and Waring 2003). UTR, untranslated region: 5′ and 3′; initiation codon (ATG); termination codon (*)

**Table 2** Arrangement of exons and CDS (Coding DNA Sequence), and respective inferred proteins (amino acid sequence) of the three isoforms (X3, X2, and X1) detected in *dec-1* of *Anastrepha obliqua* and *Anastrepha fraterculus* fruit flies. The letters a, b, and c in exon 5 represent different splice points

| | Isoforms | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *A. obliqua* | | | | *A. fraterculus* | | | |
| | | X3 | X2 | X1 | | X3 | X2 | X1 |
| Exon 1 | 9 | ✓ | ✓ | ✓ | 9 | ✓ | ✓ | ✓ |
| Exon 2 | 189 | ✓ | ✓ | ✓ | 189 | ✓ | ✓ | ✓ |
| Exon 3 | 2286 | ✓ | ✓ | ✓ | 2340 | ✓ | ✓ | ✓ |
| Exon 4 | 348 | ✓ | ✓ | ✓ | 309 | ✓ | ✓ | ✓ |
| Exon 5(a) | 18 | ✓ | | | 18 | ✓ | | |
| Exon 5(b) | 194 | | ✓ | ✓ | 197 | | ✓ | ✓ |
| Exon 5(c) | 1552 | | | ✓ | 1372 | | | ✓ |
| Exon 6 | 517 | | ✓ | | 532 | | ✓ | |
| CDS size | - | 2850 | 3543 | 4578 | - | 2865 | 3576 | 4416 |
| Protein size | - | 950 | 1181 | 1526 | - | 955 | 1192 | 1472 |

the corresponding values were 1526 aa for X1, 1181 aa for X2, and 950 aa for X3 (Table 2). These isoform labels lack a clear rule, and we followed those of *R. zephyria* (Gene ID 108362233) and *R. pomonella* (Gene ID 118735351). It should be noted that the *A. obliqua dec-1* isoforms in GenBank are named differently since their X1 is what we name here X2, with 1191 aa, whereas the GenBank X2 isoform has 954 aa and probably corresponds to what we refer to as the X3 isoform (fc106).

Two primary motifs were identified in the putative DEC-1 proteins using the Pfam database (with $p$ value $< 0.05$), denoted as the N-terminal region (Pfam ID: DEC-1 N; PF04625) and the C-terminal region (Pfam ID: DEC-1 C; PF04626) in *D. melanogaster*. In contrast to *D. melanogaster*

DEC-1, *Anastrepha* species did not have a motif in nine repeat regions (Pfam ID: Dec-1, Description: PF04624), whereas a novel RNAse P Rpr2/Rpp21/SNM1 subunit domain motif (Pfam ID: Rpr2, Description: PF04032) was identified on the longer isoform (X1) of *A. obliqua* (Figure S1). A comparative analysis of motifs detected in *A. fraterculus* and putative orthologs from various species revealed different degrees of conservation. The N-terminal motif displayed identity percentages ranging from ~ 27% (*Drosophila melanogaster*) to ~ 97% (*Anastrepha ludens*), while the C-terminal motif showed percentages ranging from ~ 17% (*Teleopsis dalmanni*) to 75% (*A. obliqua*) (Table 3.). Both motifs are highly conserved in the genus *Anastrepha*, with identities close to or above 90% in the N-terminal motif. However, their similarity to motifs from other flies is much lower, including those from other Tephritidae, a pattern that is more pronounced in the C-terminal region. The repeat region motif, exclusive to *Drosophila*, varied in the number of copies across different species, with *D. pseudoobscura* being more similar to *D. melanogaster* but having only two repeat regions, whereas *D. albomicans* lacks the repeat region altogether.

## Phylogenetic tests

The results from Xia's saturation test suggest that there is no significant underestimation of *dS* rates with increasing evolutionary distance at the levels of the contrasts investigated here. The maximum likelihood phylogenetic inference of the "*Anastrepha dataset*" limited to the amplified portion of the 3$^{rd}$ exon using the HKY + G4 substitution model placed all the *A. obliqua* samples together, while the *A. fraterculus* samples were divided into three clusters (Fig. 2A). The most internal branch included all *A. obliqua* and two samples of *A. fraterculus* from central Brazil, which we will refer to as the *A. fraterculus Clade VII branch*. We will refer to these different *A. fraterculus* clades by names that loosely correspond to branches that seem to match the phylogenetic and geographic location of said branches in samples from another study (Congrains et al. 2023). This group was sister to other specimens that form the *A. obliqua branch*, which is generally distributed in southern and central Brazil and is related to *A. ludens*. There is also an external branch with samples from the Brazilian coast that forms the *A. fraterculus Clade VI branch* and a branch that connects different species in the *fraterculus* species group, dubbed *Multispecies branch*. The maximum likelihood tree for the "*Cyclorrhapha dataset*" using the TPM3uf + I + G4 substitution model was based on the whole *dec-1* gene and confirmed the monophyly of *Anastrepha*, with *A. grandis* as basal to the other species in the *fraterculus* group (Fig. 2B). This topology places *Anastrepha* closer to *Rhagoletis* than to the other Tephritidae investigated here, *Ceratitis* and *Bactrocera* (Fig. 2B); however, in this inference, *Ceratitis* is basal to Tephritidae, which is different from the standard phylogeny for these taxa (Segura et al. 2006).

## Selection tests

### Global analysis of *dec-1*

ML phylogenetic trees were inferred to investigate patterns of molecular evolution and positive selection on the *dec-1* gene. The phylogenetic tree previously inferred allows for the proper contrast of the likelihood of different evolutionary models. The significant contrast between the M0 model (which assumes a uniform rate of evolution) and the M3 model (which incorporates discrete substitution rates and allows for positive selection) (M0 vs M3) across both the "*Anastrepha dataset*" and "*Cyclorrhapha dataset*" suggests that there is evolutionary rate heterogeneity along the *dec-1* gene ($p$ values < 0.0001). On the other hand, we failed to detect significant differences between the M1a and M2a evolutionary models for either set, which indicates insufficient

**Table 3.** Identity of putative *dec-1* domains (fc106, which corresponds to the X3 isoform) of *Anastrepha fraterculus* sequences and other putative orthologs in Cyclorrhapha

| Species | Family | Protein size | Motif N-terminal | Motif C-terminal |
|---|---|---|---|---|
| *Anastrepha obliqua* | Tephritidae | 950 | 88.69 | 75.00 |
| *Anastrepha ludens* | Tephritidae | 984 | 97.86 | 71.26 |
| *Anastrepha grandis* | Tephritidae | 972 | 91.42 | 64.20 |
| *Rhagoletis zephyria* | Tephritidae | 1071 | 64.82 | 48.68 |
| *Rhagoletis pomonella* | Tephritidae | 1071 | 64.82 | 48.68 |
| *Zeugodacus cucurbitae* | Tephritidae | 1106 | 64.74 | 46.24 |
| *Ceratitis capitata* | Tephritidae | 1120 | 63.86 | 46.11 |
| *Bactrocera tryoni* | Tephritidae | 1125 | 64.88 | 39.43 |
| *Drosophila melanogaster* | Drosophilidae | 950 | 27.07 | 29.01 |
| *Drosophila albomicans* | Drosophilidae | 921 | 30.57 | 27.10 |
| *Drosophila pseudoobscura* | Drosophilidae | 949 | 29.32 | 24.84 |
| *Teleopsis dalmanni* | Diopsidae | 1347 | 29.27 | 17.71 |

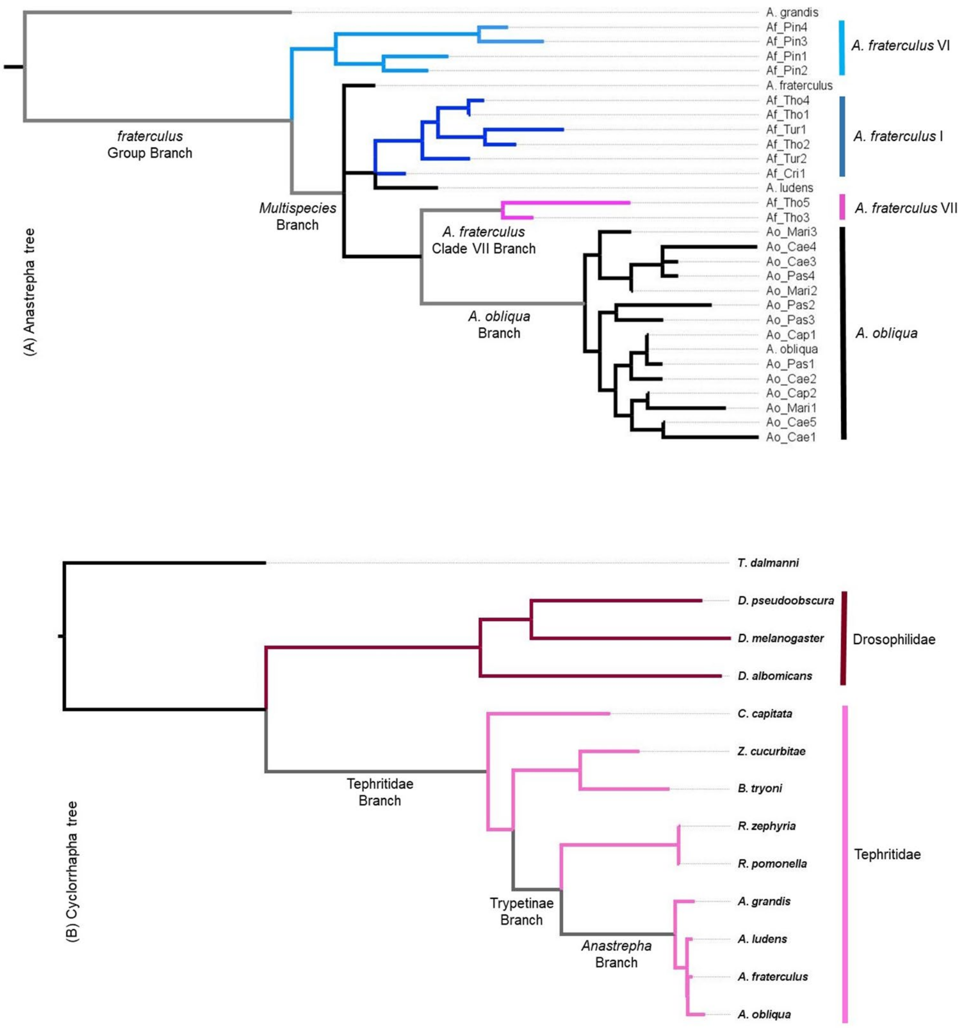

**Fig. 2** Phylogenetic trees of the "*Anastrepha dataset*" (top) and "*Cyclorrhapha dataset*" (bottom), which are composed of different sets of species and parts of the *dec-1* gene. The gray-marked branches were subjected to branch selection analysis. In the "*Anastrepha*" tree, *Anastrepha obliqua* sequences and three distinct evolutionary lineages identified in *Anastrepha fraterculus* are highlighted by different colors (**A**)

evidence of positive selection affecting the whole gene. Although the contrast between M7 and M8 for the more closely related "*Anastrepha*" dataset was borderline significant (*p* value = 0.075), it was highly significant for the "*Cyclorrhapha dataset*" (*p* value = 0.00), suggesting that there are specific sites under positive selection in *dec-1* at that level of divergence.

The results of the BEB analysis revealed 7 possible codons with greater probabilities of being under positive selection in the "*Anastrepha dataset*" (Fig. 3), although a single codon, 523G ($\omega = 1.27$), retained a significant 95% posterior probability.

As revealed by the BEB analysis, the global analysis performed on the "*Anastrepha dataset*" conducted by the FUBAR test in HyPhy, which estimates d*N*/d*S* substitution rates per site, identified the same codons under positive selection (523, 560, 656, 743, and 761), with posterior probabilities of 97.6%, 91.3%, 93.9%, 92.5%, and 94.2%, respectively, but it also identified codon 617 (with 93.5% posterior probability). The BUSTED test suggested that 12.7% of the branches were evolving under neutral selection, while 87.3% were under purifying selection. Furthermore, this analysis indicated that the majority of codons on exon 3 were under purifying selection (Fig. 3), particularly those associated with the proprotein s20, which was confirmed by FEL results that suggested no constant positive selection pressure along the gene. Despite these results, the MEME test suggested that codons 617, 656, and 724 underwent episodic or diversifying selection, and the two former codons were also identified as being under positive selection by FUBAR. This suggests that positive selection may have impacted these codons at specific evolutionary times and episodically without affecting the entire gene region.

A global analysis by PAML on the "*Cyclorrhapha dataset*", which investigates more distantly related taxa, identified codons 1309G and 1314V as being under positive selection, with posterior probabilities of 96% and 98.2%, respectively. These findings were further supported by FUBAR analysis, which showed a probability of 96.7% for codon 1314V, and by FEL, which identified an additional seven sites under

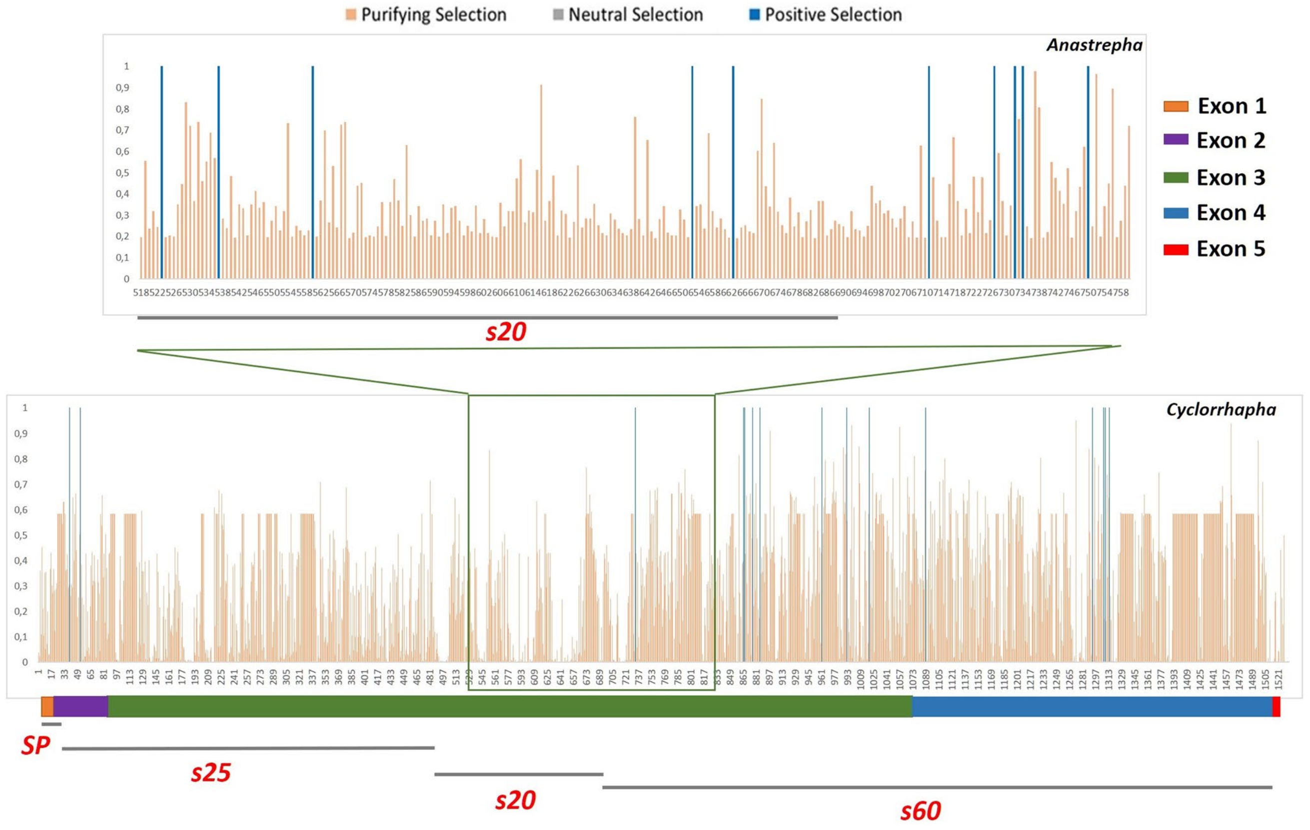

**Fig. 3** Naïve empirical Bayes (NEB) graphical analysis showing Bayes empirical Bayes (BEB) posterior probabilities of sites under purifying ($0<\omega<1$), positive ($\omega>1$) and neutral ($\omega=1$) selection for the *Cyclorrhapha* and *Anastrepha* datasets. The codons of the analyzed sequences are represented on the $x$-axis, and posterior probability values are represented on the $y$-axis. The values on the $y$-axis indicate the posterior probability of the $\omega$ value being above one at each analyzed site. The graphical representation is derived from NEB data, but only the BEB posterior probability values are considered. Exons 1, 2, 3, 4, and 5 indicate the positions and sizes of the exons in the *dec-1* gene represented as colored lines below the $x$-axis of the graph

diversifying selection. Additionally, MEME revealed 13 sites under episodic diversifying selection, affecting either the coding regions for s25 or s60 proproteins but not s20, which are basically the same sites shown in the BEB graph (Fig. 3); however, in this test, these sites are possibly evolving under positive selection but fail to reach posterior probabilities greater than 95%.

## Branch selection analysis

The branch-site selection analyses of the "*Anastrepha dataset*" dataset identified selection episodes in the *A. obliqua branch*, the *A. fraterculus Clade VII branch*, and in the branch that separates *A. fraterculus Clade VI* and the other *A. fraterculus* species (*Multispecies branch*). It should be noted that *A. fraterculus* was found paraphyletic in our analysis, which corroborates other analyses that suggest that *A. fraterculus* is a species complex and possibly not monophyletic (Selivon and Perondini 1998; Selivon et al. 2004, 2022; Hernández-Ortiz et al. 2012; Congrains et al. 2021, 2023). The overall $\omega$ value of 0.1280 indicated that most branches in the "*Anastrepha dataset*" dataset were under purifying selection with a probability of 97%. The RELAX test indicated that the *A. fraterculus Clade VII branch* was under intensified selection ($K=2.03$), while the *Multispecies branch* was under relaxed selection ($K=0.49$), whereas the *A. obliqua branch* had $K=0$, indicating an absence of evidence for relaxation or intensification of selection. Despite that, this branch showed a significant ΔLRT (4.27685; $p=0.03863$) on the branch-site test, with codon 724 having a 95% posterior probability of being under positive selection. On the other hand, the *A. fraterculus Clade VII branch* and the *Multispecies branch* showed no signs of positive selection.

We investigated patterns of selection on three branches in the "*Cyclorrhapha dataset*": one that defines the genus *Anastrepha* (*Anastrepha branch*), which showed a single codon, 526N (96.8%), with a significant probability of being under positive selection ($\Delta LRT=9.3923$, $p=0.0022$). There was no evidence of selection on the *Trypetinae Branch*, whereas 20% of the codons showed evidence of positive selection ($\Delta LRT = 36.539$; $p = 0.0000$) in the *Tephritidae Branch*, including codons 4A (98.6%), 52L (99.6%), 100 T (95.7%), 127Q (96.3%), 172N (99.5%), 182L (99.6%), 211 H (95.7%), 277S (99.4%), 297 (96.7%), 310I (99.2%), 318P (95.0%), 419 V (98.4%), 448 M (99.8%), and 523 K (98.7%), although the RELAX test failed to find significant evidence of relaxation or intensification of selection in any of the three branches analyzed at this level.

### Measurement of the physicochemical magnitude of amino acid substitutions

We found evidence that radical amino acid changes potentially altered some physicochemical properties (global $z$ values equal to or greater than six) at different evolutionary levels. In the "*Anastrepha dataset*," 22 codons exhibited significant $z$ scores (with $z = 6$ and 7, $p < 0.01$), most of which affected the *dec-1* region that codes for s20 and only 7 affecting the regions that code for s60. Furthermore, radical mutations on s20 are more frequent on interior nodes than on tips, contrary to non-radical mutations on s20 (Fisher's exact test $p = 0.0214$) and any mutations on s60 (Table 4.). The difference in the distribution of radical mutations in tips and interiors between s20 and s60 was also significant (Fisher's exact test $p = 0.0030$), since the latter showed no radical mutations on internal branches (Table 4.). This pattern is different from what is observed in the "*Cyclorrhapha dataset*", which showed more than 400 sites with radical physicochemical alterations and more than 1100 radical mutations (Table 5.). Although the latter dataset showed a greater percentage of radical mutations on tips than on internal branches, there was no significant difference in these relative rates across different regions of *dec-1*. However, there are different rates of mutation across these regions, with pro-protein s60 showing significant radical physicochemical alterations in close to 60% of the codons (Table 5.). However, there was a significant contrast in the pattern of radical mutations for the s20 proprotein region when we compared the "*Anastrepha dataset*" and the "*Cyclorrhapha dataset*," since the latter has many more radical changes on interior branches (chi-square 15.1, $p = 0.001733$), which was not observed on s60.

**Table 4.** Distribution of radical and non-radical changes defined on TreeSAAP across different parts of *dec-1* using the *Anastrepha dataset*, separated by interior branches and tip branches

| | Radical | | Non-radical | |
|---|---|---|---|---|
| | Tip | Interior | Tip | Interior |
| s20 | 10 | 13 | 7 | 0 |
| s60 | 24 | 7 | 16 | 11 |
| Total | 34 | 20 | 23 | 11 |

**Table 5.** Distribution of radical and non-radical changes defined on TreeSAAP across different parts of *dec-1* using the Cyclorrhapha *dataset*, separated by interior branches and tip branches

| | Radical | | Non-radical | |
|---|---|---|---|---|
| | Tip | Interior | Tip | Interior |
| Signal peptide | 11 | 11 | 34 | 30 |
| s25 | 86 | 62 | 239 | 199 |
| s20 | 149 | 138 | 363 | 293 |
| s60 | 378 | 275 | 776 | 580 |
| Total | 624 | 486 | 1412 | 1102 |

We attempted to model DEC-1 3D proteins from the Cyclorrhapha dataset using *D. melanogaster* Dec-1 as reference (Fig. 2S), but the results yielded low confidence scores (49.4 to 43.8%), rendering these models speculative. The proteins modeled, even from *D. melanogaster*, lacked coherent tertiary structures and displayed only secondary elements, such as alpha-helices and loops. The results from AIUPred suggest that Dec-1 is, mostly, composed of disordered regions with high IUPred values (Fig. 3S), although it also indicates that the positive selected sites seem to be on disordered regions positioned externally on the protein, which might be relevant for its function.

## Discussion

Protein-coding genes experience a myriad of selection pressures. Depending on the protein's importance, non-synonymous mutations may be more likely to be deleterious, whereas those enhancing fitness can become fixed and traced through evolutionary signals of positive selection (Sémon and Wolfe 2007; Conant and Wolfe 2008; Studer and Robinson-Rechavi 2009). Therefore, investigating patterns of selection on a gene is an important foray into its past and present physiological relevance. The characterization of the chorionic gene, *defective chorion-1* (*dec-1*), in *Anastrepha* and its comparison to other Tephritidae and Diptera allowed us to investigate which forces have shaped its evolution at different evolutionary scales and identify conserved gene regions, as well as some possibly subjected to positive selection events.

Structural analysis of the *dec-1* gene revealed 6 exons and three isoforms in *Anastrepha fraterculus* and *A. obliqua*, a pattern similar to what has been shown for Drosophilidae, such as *D. melanogaster*, *D. yakuba*, and *D. virilis*. Transcriptomic and sequencing data from laboratory samples of *A. fraterculus* and *A. obliqua* provided a more detailed characterization of the *dec-1* gene, improving our understanding of gene boundaries and expressed isoforms in this genus. Previous studies have shown the structural conservation of *dec-1*, so much so that transgenic studies have indicated its functional interchangeability across distant Drosophilidae species with regard to fertility and wild-type eggshell morphology (Badciong et al. 2001). Here, we show that despite amino acid sequence divergence, especially among more distantly related species, such as Drosophilidae and Tephritidae, there is evidence of preserved protein functionality throughout evolution across Cyclorrhapha (Badciong et al. 2001). This is supported by evidence of purifying selection acting across its length at different evolutionary levels. Despite the overall evidence of selective constraint, especially in some portions of the gene associated with some proproteins, occasional episodes of positive selection suggest the impact of dynamic evolutionary processes, shedding new light on the adaptive mechanisms of *dec-1*.

The overall pattern detected by global tests of selection on *dec-1* across different evolutionary hierarchies indicates that purifying selection seems to be the driving force along its five protein-coding exons throughout most of its history, although there is evidence of positive selection in some portions of the gene and at different points in time. This was shown in PAML, as well as on the global BUSTED test, which indicates that overall synonymous substitution rates are higher than non-synonymous rates ($dN/dS < 1$) and that most sites are evolving under purifying selection, although some sites have evidence of evolving under neutral conditions. Considering the functional role of *dec-1*, its isoforms, and protein derivatives in oogenesis (Noguerón et al. 2000), particularly in the formation of the chorion and vitelline membrane (Bauer and Waring 1987), it is reasonable to expect a prevalence of purifying selection across its length and history. This becomes even more apparent when we consider that mutations in exon 3, which may affect all three isoforms, can lead to sterility or egg anomalies (Waring et al. 1990; Kim et al. 2002; Spangenberg and Waring 2007), underscoring the biological significance of the products of this gene. Nonetheless, the results from MEME, as well as the contrast between the M7 and M8 models in PAML, suggest that some sites are evolving under positive selection.

An investigation of different rates of evolution across the gene using the Bayes empirical Bayes model (BEB) in PAML (Yang 1997), along with various tests conducted using HyPhy (Kosakovsky Pond et al. 2019), revealed that even though the majority of codons are evolving under purifying or neutral evolution, there are codons that experienced positive selection or episodic diversifying selection. A better understanding of the potential impact of changes across *dec-1* should consider the different proproteins and derivatives produced by the three isoforms normally found at different developmental times, which may have different functions. For instance, fc106 is much more abundant than are fc125 and fc177 and is found in oogenesis stages 9 and 10. On the other hand, fc177 expression increases in stages 11 and 12 (Hawley and Waring 1988; Mauzy-Melitz, and Waring 2003). The fc106 product is cleaved into the s25 and s80 proproteins during late oogenesis stage 10, but s80 is cleaved later into s60 and s20. That may be why s25, s60, and s85 are deposited in the egg layers, whereas s20 does not seem to be a structural component of the eggshell, although it is detected in vesicles within the oocyte (Spangenberg and Waring 2007). This might be one of the reasons why s20 seems to be more conserved than the other *dec-1* proproteins.

Analyses such as BEB (Fig. 3) suggest that evidence of positive selection between more distantly related taxa is restricted to the beginning of proprotein s25, corresponding to exon 2, and the middle portion of exon 3, which putatively encodes for the proprotein s60 but is also part of the s95 proprotein. Significant positively selected sites on the putative s20 proprotein were detected only among the more closely related *Anastrepha dataset*. Similar results, which in general point to the same sites under selection, were detected by the FUBAR and MEME tests. A similar pattern was revealed by TreeSAAP, which identified the same codons, 523 and 656, in the "*Anastrepha dataset*" and codon 1314 in the "*Cyclorrhapha dataset*" group, among other codons that showed evidence of positive selection.

Physicochemical changes in codons 523 and 656 could affect all three *dec-1* isoforms since they all share sequences from exon 3, but each derivative proprotein would be differentially affected since only the s20 derivative (produced from FC106), s85, and the initial regions of the s95 (FC125) derivatives would retain these domains. In contrast, the 1314 V codon found in the "*Cyclorrhapha dataset*," located at the end of exon 5 (equivalent to exon 4 in *D. melanogaster*), impacts the C-terminal region of DEC-1, a part of the s80 derivative, s60, and s85 proproteins. Changes in the s20 derivative could disrupt the organization of early protein interactions in the egg assembly process (Badciong et al. 2001) and induce female sterility, similar to alterations in the s95 derivative (Mauzy-Melitz and Waring 2003). Furthermore, changes in the function of the s85 derivative may destabilize the vitelline membrane (Waring et al. 1990) and the tripartite endochorion, as proper aggregation of chorion proteins to form the pillars, floor, and reticulated roof does not occur (Mauzy-Melitz and Waring 2003). Overall, the identified mutations occur in regions that might compromise egg formation and viability (Hawley and Waring 1988).

Different rates of synonymous and non-synonymous changes potentially indicate regions subject to different selection pressures, but it is not known whether these changes impact fitness. Another way to look at these changes is to evaluate whether inferred amino acid substitutions produce potentially radical physicochemical changes, which are more likely to have a fitness impact, since they are more likely to be selected against (Smith 2003), particularly in large populations (Weber and Whelan 2019). In general, regions that are under purifying selection are less likely to undergo radical amino acid changes since they are more likely to alter protein structure and, potentially, function, whereas positive selection might favor radical amino acid changes. Therefore, in conserved regions, radical changes are more likely to be found on the tips of trees, which would indicate that they are evolutionarily recent, whereas non-radical changes could be found in internal branches just as likely as on the tips (Pupko et al. 2003).

Overall, the *dec-1* results indicate that radical changes are more likely to be found on tree tips than on the interior, and this finding is consistent across the whole gene. The one distinction we observed is related to the part of the gene that codes for the s20 proprotein, since there are several internal amino acid changes segregating in the "*Anastrepha dataset*," but not for s60, whereas that pattern is not observed when we compare more distantly related taxa. To explain these differences, we should consider the impact of selection on the underlying genetic variation. Positive selection may lead to the sweep of other variation in the population, but the speed with which that happens depends on whether it is a hard or a soft sweep, and it is not necessarily trivial to distinguish them (Schneider et al. 2021). On the other hand, purifying or background selection does not remove neutral variation or even deleterious mutations at the same rate as with positive selection, showing larger variance (Cvijović et al. 2018). In this case, we would expect that regions that are subject to positive selection might not have any segregating radical changes when we compare closely related specimens but rather would show greater rates of evolution when comparing more distantly related specimens. In regions under purifying selection, however, we would still see segregating variation between closely related species but that would not be reflected in higher rates of evolution; rather, we would not see that reflected when comparing more distantly related taxa because that would not be favored by selection.

Another important aspect to consider when using rates of synonymous and non-synonymous changes to investigate selection in recently diverged lineages is that *dN*/*dS* is a measure of selective pressure that contrasts rates of fixed substitution between two homologous sequences (Kimura 1977), and as such, it was designed for analyzing more divergent lineages; otherwise, it might suffer from the stochasticity of the mutation/selection process. The comparison of these changes between more closely related taxa, generally by *Ka/Ks*, may be influenced by other factors that are not directly associated with selection intensity, such as population expansion or contraction, and other evolutionary forces that may even affect neighboring regions (Kryazhimskiy and Plotkin 2008). Therefore, it is relevant to contrast patterns of evolution considering adequate evolutionary models that could differentiate evidence of selection from its relaxation. In this framework, the RELAX analysis indicated that *A. fraterculus* Clade VII branch in the "*Anastrepha dataset*" has an intensification of selection ($K = 2.03$), and no clades in the "*Cyclorrhapha dataset*" showed evidence of intensification or relaxation of selection. Intensified selection pressures can push $\omega$ values away from neutrality ($\omega = 1$), while relaxed selection allows for the accumulation of neutral or slightly deleterious genetic variations, promoting genetic diversity within populations. This relaxation of selection potentially influences differentiation by providing raw material for adaptation to new environments and conditions (Templeton 2008).

Global analysis of selection detected potential sites experiencing positive and/or episodic diversifying selection, but because the average $\omega$ ratio is rarely above 1, detecting positive selection in specific lineages considering a phylogenetic framework may be more effective, as it targets those lineages specifically (Álvarez-Carretero et al. 2023). Branch-site tests in PAML identified specific branches with evidence of positive selection, such as the lineage separating *A. obliqua* and *A. fraterculus* in the "*Anastrepha dataset*" and all three tested branches in the *Cyclorrhapha dataset*. In all cases, we have evidence of a few lineages experiencing episodic positive selection in a background of purifying selection.

The pattern of purifying selection with a few events of positive or diversifying selection identified in the *dec-1* gene is also observed in genes directly and indirectly related to reproduction and sex determination, such as *transformer* (*tra*) and *fruitless* (*fru*) in *Drosophila* (McAllister and McVean 2000; Kulathinal et al. 2003; Parker et al. 2014) and even in *Anastrepha* (Sobrinho and de Brito 2010, 2012). More importantly, this trend has also been observed for several choriogenic genes investigated in *Anastrepha*, such as *Cp15*, *Cp16*, *Cp19*, and *Cp38*, whereas vitelline membrane genes, in general, have been experiencing only purifying selection, with the possible exception of a recently duplicated gene (*Vm26Ab* and *Vm26Aa′*) in the genus *Anastrepha* (Gonçalves et al. 2013).

Vitelline membrane genes under purifying selection and chorionic genes under positive selection were also described for *Drosophila* (Jagadeeshan and Singh 2007). These results have been attributed to the fact that vitelline membranes are internal to the chorion and therefore more subject to purifying selection, whereas chorionic proteins may have important roles with regard to desiccation and oxygen transport that could be subject to positive selection depending on environmental cues (Jagadeeshan and Singh 2007). Notably, portions of *dec-1* associated with s25, which are expressed earlier in egg development, seem to be more conserved than portions of the gene related to the s60 proprotein, which is produced later and seems to be deposited in more outer chorionic layers. However, most of the positive sites detected in the branches that define Tephritidae, as well as the genus *Anastrepha*, are associated with the s25 proprotein, which might be relevant when we consider the great variation across species of *Anastrepha* in terms of chorion morphology (Selivon and Perondini 1998; Selivon et al. 2004; Figueiredo et al. 2013). Unfortunately, the disordered patterns generally found in DEC-1, indicated by high IUPred values (Fig. 3S) limited our ability to establish how these potentially positively selected sites would directly affect DEC-1 protein structure, but seem to indicate two interesting patterns. First, the level of disorder seems to change across the DEC-1 protein, being unsurprisingly lower in the signal peptide, but also lower in the beginning of proprotein s25 and proprotein s20 and much higher in the proprotein s60. It is possible that DEC-1 does not require a stable 3D structure to function, instead relying on flexibility and multiple conformations during post-translational modifications, as an intrinsically disordered protein, which are very common on eukaryotes (Trivedi and Nagarajaram 2022). That notwithstanding, the positive selection sites seem to be on disordered regions positioned externally on the protein (Fig. 2S). This external positioning may be relevant, as it could influence protein interactions, accessibility to ligands, or regulatory mechanisms; however, their significance remains uncertain.

Despite the general pattern of purifying selection, the evidence of positive selection on *dec-1* in Tephritidae, particularly among different *Anastrepha* species, is relevant because of the great diversity of the inner chorion patterns among *Anastrepha* species, which are much more diverse than those found in other Tephritidae (Murillo and Jirón, 1994; Selivon and Perondini 1998; Dutra et al. 2011; Figueiredo et al. 2017). Although rapidly evolving genes in general have been associated with sexually selected traits or sexual conflict across different taxa (Swanson and Vacquier 2002; Clark et al. 2006; Panhuis and Swanson 2006; Turner and Hoekstra 2006), adaptive responses to environmental cues can also play an important role (Jagadeeshan and Singh 2007). This should not be overlooked, when we consider that several of the species here studied have different host preferences.

It is important to emphasize that phylogenetic inferences derived from *dec-1* are not meant to help elucidate relationships among the taxa studied here, which are generally well established, at least at higher levels. Rather, they are used to establish a framework that enables the investigation of molecular evolutionary patterns. Nonetheless, we should mention that the relationships depicted among the more diverse taxa in general reflect what has been indicated in the literature, with a closer relationship between *Anastrepha* and *Rhagoletis*, although the phylogenetic tree here produced placed *Ceratitis* basal to *Tephritidae* rather than on a clade with *Bactrocera* and *Zeugodacus* (Virgilio et al. 2015), a more accepted relationship among Toxotrypanini (Mengual et al. 2017).

The overall phylogenetic pattern observed, with short internal branches even between more distant groups, provides a crucial backdrop for understanding the broader evolutionary context of the *dec-1* gene and its potential functional implications, especially for more closely related lineages, such as the inferences among *Anastrepha* specimens. Our results indicated that the samples of *A. fraterculus* we investigated seemed to be separated into three different lineages. It has been suggested that *A. fraterculus* is a species complex with several cryptic species, with at least three in Brazil that can be identified via morphometric methods (Hernández-Ortiz et al. 2012, 2015). Although we did not investigate different morphotypes in this study that would enable us to compare these taxa to the ones described elsewhere (Hernández-Ortiz et al. 2015; Prezotto et al. 2019), the topology we found is remarkably similar to what has been described using transcriptomic and genomic data (Congrains et al. 2021, 2023), which identified three distinct branches for *A. fraterculus* while supporting a single *A. obliqua* branch. These diverse *A. fraterculus* lineages may be subject to distinct selective pressures (Smith-Caldas et al. 2001; Manni et al. 2015), despite their recent divergence. This detailed phylogenetic context sets the stage for understanding the evolutionary dynamics of the *dec-1* gene within specific *Anastrepha* lineages, although evidence of intraspecific gene flow among these lineages (Díaz et al. 2018; Congrains et al. 2023) may limit our ability to investigate that adequately.

Shifting focus to broader implications, a quest for mutations fixed by positive selection events can help identify novel functions and differentiate species (Sémon and Wolfe 2007; Conant and Wolfe 2008). In the context of a structural gene related to egg and chorion development, these changes may not be readily discernible or confer adaptive advantages, especially due to the complex interactions between genotype, phenotype, and environmental influences. Therefore, comprehending how reproductive genes are influenced by selective forces through selection analyses enables the identification of this genotypic and potentially phenotypic variation in the genome. Furthermore, these genes can serve as markers in the context of species differentiation amid phylogenetic uncertainties. Remarkably, *dec-1* produces adequate phylogenetic signals at different evolutionary levels across Tephritidae, but it seems to distinguish among different lineages of *A. fraterculus* for which there are only a handful of informative markers, despite the plethora of genes investigated (Congrains et al. 2021, 2023). Understanding its structure, isoforms, and evolutionary patterns across broader sampling also sets the stage for its establishment as a viable adequate marker to study variation across *Anastrepha*. It also provides insight into reproductive biology, which could foster the development of genetically modified females through transgenes, similar to what has been done in *Drosophila*, which could be important for population management, particularly in pest insects such as those of the Tephritidae family.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00427-024-00723-3.

**Acknowledgements** We are grateful to Dr. Scott Geib for the data shared with us and with people from the Lab of Population Genetics and Evolution, UFSCar and Dr. Julian Dupuis, Univ. of Kentucky, for invaluable comments on a previous version of this manuscript.

**Author contributions** Bisio, M. de C. and de Brito, R. A. designed the experiments. Bisio, M. de C., Santos, E. M. dos, Chahad-Ehlers, S., Santos. C. A., and de Brito, R. A. wrote the manuscript. Bisio, M. de C., de Brito, R. A., Santos, E. M. dos and Santos, C. A. performed different data analyses. All the authors have read and agreed to the final version of the manuscript. Bisio, M. de C. and Santos, E. M. dos, share the first authorship of this manuscript.

**Funding** The authors have been supported by grants and fellowships from the following Brazilian research funding agencies: Coordination for the Improvement of Higher Education Personnel — CAPES to M. C. B. (88887.497759/2020-00), S. C.-E.(88887.466869/2019-00) and E. M. S (88882.426711/2019-01) and R. A. B. (PVE 056/13); State of São Paulo Research Foundation — FAPESP (2018/06611–5 and 2022/12583–0) and CNPq (313504/2017–2) grants to R. A. B. and CNPq fellowships to E. M. S. and R. A. B.

**Data availability** Sequence data that support the findings of this study has been deposited in the NCBI and will be available upon manuscript acceptance. A version of these data has been made available in this submission.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

Altschul SF, Koonin EV (1998) Iterated profile searches with PSI-BLAST–a tool for discovery in protein databases. Trends Biochem Sci 23:444–447. https://doi.org/10.1016/s0968-0004(98)01298-5

Álvarez-Carretero S, Kapli P, Yang Z (2023) Beginner's guide on the use of paml to detect positive selection. Mol Biol Evol 40:msad041. https://doi.org/10.1093/molbev/msad041

Andersson S, Lambertsson A (1991) Evolution of the dec-1 eggshell locus in Drosophila. I. Restriction site mapping and limited sequence comparison in the melanogaster species subgroup. J Mol Evol 33:321–331. https://doi.org/10.1007/bf02102863

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Badciong JC, Otto JM, Waring GL (2001) The functions of the multiproduct and rapidly evolving dec-1 eggshell gene are conserved between evolutionarily distant species of Drosophila. Genetics 159:1089–1102. https://doi.org/10.1093/genetics/159.3.1089

Bauer BJ, Waring GL (1987) 7C female sterile mutants fail to accumulate early eggshell proteins necessary for later chorion morphogenesis in Drosophila. Dev Biol 121:349–358. https://doi.org/10.1016/0012-1606(87)90171-0

Beckstead WA, Ebbert MT, Rowe MJ, McClellan DA (2009) Evolutionary pressure on mitochondrial cytochrome b is consistent with a role of CytbI7T affecting longevity during caloric restriction. PLoS ONE 4:e5836. https://doi.org/10.1371/journal.pone.0005836

Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T (2017) The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res 45:D313-d319. https://doi.org/10.1093/nar/gkw1132

Chen S-H, Su S-Y, Lo C-Z, Chen K-H, Huang T-J, Kuo B-H, Lin C-Y (2009) PALM: a paralleled and integrated framework for phylogenetic inference with automatic likelihood model selectors. PLoS ONE 4:e8116. https://doi.org/10.1371/journal.pone.0008116

Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. Reproduction 131:11–22

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9:938–950. https://doi.org/10.1038/nrg2482

Congrains C, Campanini EB, Torres FR, Rezende VB, Nakamura AM, de Oliveira JL, Lima ALA, Chahad-Ehlers S, Sobrinho JIS, de Brito RA (2018) Evidence of adaptive evolution and relaxed constraints in sex-biased genes of South American and West Indies fruit flies (Diptera: Tephritidae). Genome Biol Evol 10:380–395. https://doi.org/10.1093/gbe/evy009

Congrains C, Zucchi RA, de Brito RA (2021) Phylogenomic approach reveals strong signatures of introgression in the rapid diversification of neotropical true fruit flies (Anastrepha: Tephritidae). Mol Phylogen Evol 162:107200. https://doi.org/10.1016/j.ympev.2021.107200

Congrains C, Dupuis JR, Steck GJ, Rodriguez EJ, Sutton BD, Branham MA, De Brito RA, Geib SM (2023) Phylogenomic Analysis Provides Diagnostic Tools for Identification of Species Complex. Evol Appl 16:1598–1618. https://doi.org/10.1111/eva.13589

Cvijović I, Good BH, Desai MM (2018) The effect of strong purifying selection on genetic diversity. Genetics 209:1235–1278. https://doi.org/10.1534/genetics.118.301058

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2019) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol 37:291–294. https://doi.org/10.1093/molbev/msz189

Dassault Systèmes (2021) BIOVIA Discovery studio visualizer. San Diego

De Queiroz K (2007) Species concepts and species delimitation. Syst Biol 56:879–886. https://doi.org/10.1080/10635150701701083

Díaz F, Luís A, Lima A, Nakamura AM, Fernandes F, Sobrinho I, de Brito RA (2018) Evidence for introgression among three species of the *Anastrepha fraterculus* group, a radiating species complex of fruit flies. Front Genet 10(9):359. https://doi.org/10.3389/fgene.2018.00359

Dupuis JR, Ruiz-Arce R, Barr NB, Thomas DB, Geib SM (2019) Range-wide population genomics of the Mexican fruit fly: toward development of pathway analysis tools. Evol Appl 12:1641–1660. https://doi.org/10.1111/eva.12824

Dutra VS, Ronchi-Teles B, Steck GJ, Silva JG (2011) Egg morphology of Anastrepha spp. (Diptera: Tephritidae) in the fraterculus group using scanning electron microscopy, Ann Entomol Soc

Am. Entomol Soc Am 104(1):16–24. https://doi.org/10.1603/AN10105

Eickbush TH and Izzo JA (1995) Chorion genes: molecular models of evolution, in: Goldsmith MR and Wilkins AS (Eds.), Molecular model systems in the Lepidoptera. Cambridge University Press, Cambridge, pp. 217–248. https://doi.org/10.1017/CBO9780511529931.009

Erdős G, Dosztányi Z (2024) AIUPred: combining energy estimation with deep learning for the enhanced prediction of protein disorder. Nucleic Acids Res 52:W176–W181. https://doi.org/10.1093/nar/gkae385

Escher SA, Lambertsson A (1996) Evolution of the dec-1 eggshell locus in Drosophila III. Sequence comparisons of the simulans complex repeated domain reveal non-concerted evolution. Hereditas 124:111–120

Fakhouri M, Elalayli M, Sherling D, Hall JD, Miller E, Sun X, Wells L, LeMosy EK (2006) Minor proteins and enzymes of the Drosophila eggshell matrix. Dev Biol 293:127–141

Fay JC, Wu CI (2005) Detecting hitchhiking from patterns of DNA polymorphism. In: Nurminsky D (ed) Selective sweep. Eurekah.com and Kluwer Academic / Plenum Publishers, New York and Gerogetown, pp 65–77

Figueiredo JVA, Perondini ALP, Ruggiro EM, Prezotto LF, Selivon D (2013) External eggshell morphology of Anastrepha fruit flies (Diptera: Tephritidae). Acta Zoologica 94:125–133. https://doi.org/10.1111/j.1463-6395.2011.00533.x

Figueiredo JV, Perondini AL, Selivon D (2017) Patterns of inner chorion structure in Anastrepha (Diptera: Tephritidae) eggs. Arthropod Struct Dev 46:236–245. https://doi.org/10.1016/j.asd.2016.11.004

Findlay GD, Swanson WJ (2010) Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays 32:26–36

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Gonçalves VR, Sobrinho IS, Malagó W, Henrique-Silva F, de Brito RA (2013) Transcriptome analysis of female reproductive tissues of Anastrepha obliqua and molecular evolution of eggshell proteins in the fraterculus group. Insect Mol Biol 22:551–561

Gromiha MM, Selvaraj S (1997) Influence of medium and long range interactions in different structural classes of globular proteins. J Biol Phys 23:151–162. https://doi.org/10.1023/a:1004981409616

Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A (1999) Relationship between amino acid properties and protein stability: buried mutations. J Protein Chem 18:565–578. https://doi.org/10.1023/a:1020603401001

Guindon S (2010) Bayesian estimation of divergence times from large sequence alignments. Mol Biol Evol 27:1768-81. https://doi.org/10.1093/molbev/msq060

Hall BG (2004) Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. Mol Biol Evol 22:792–802

Hanna-Rose W, Licht JD, Hansen U (1997) Two evolutionarily conserved repression domains in the Drosophila Kruppel protein differ in activator specificity. Mol Cell Biol 17:4820–4829. https://doi.org/10.1128/mcb.17.8.4820

Hawley RJ, Waring GL (1988) Cloning and analysis of the dec-1 female-sterile locus, a gene required for proper assembly of the Drosophila eggshell. Genes Dev 2:341–349. https://doi.org/10.1101/gad.2.3.341

Hernández-Ortiz V, Bartolucci AF, Morales-Valles P, Frías D, Selivon D (2012) Cryptic species of the Anastrepha fraterculus complex (Diptera: Tephritidae): a multivariate approach for the recognition of South American morphotypes. Ann Entomol Soc Am 105:305–318. https://doi.org/10.1603/AN11123

Hernández-Ortiz V, A. Canal N, Tigrero Salas JO, Ruíz-Hurtado FM and Dzul-Cauich JF (2015) Taxonomy and phenotypic relationships of the *Anastrepha fraterculus* complex in the Mesoamerican and Pacific Neotropical dominions (Diptera, Tephritidae). ZooKeys 540:95-124. https://doi.org/10.3897/zookeys.540.6027

Jagadeeshan S, Singh RS (2007) Rapid evolution of outer egg membrane proteins in the Drosophila melanogaster subgroup: a case of ecologically driven evolution of female reproductive traits. Mol Biol Evol 24:929–938

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–80. https://doi.org/10.1093/molbev/mst010

Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27:757–763. https://doi.org/10.1093/bioinformatics/btr010

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053

Kim C, Han K, Kim J, Yi JS, Kim C, Yim J, Kim YJ, Kim-Ha J (2002) Femcoat, a novel eggshell protein in Drosophila: functional analysis by double stranded RNA interference. Mech Dev 110:61–70. https://doi.org/10.1016/s0925-4773(01)00559-7

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276. https://doi.org/10.1038/267275a0

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59

Kosakovsky Pond SL, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, Wisotsky S, Spielman SJ, Frost SDW, Muse SV (2019) HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol 37:295–299. https://doi.org/10.1093/molbev/msz197

Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. PLoS Genet 4:e1000304–e1000304

Kulathinal RJ, Skwarek L, Morton RA, Singh RS (2003) Rapid evolution of the sex-determining gene, transformer: structural diversity and rate heterogeneity among sibling species of Drosophila. Mol Biol Evol 20:441–452

Lis JT, Schleif R (1975) Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. Nucleic Acids Res 2:383–390

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38:4647–4654. https://doi.org/10.1093/molbev/msab199

Manni M, Lima KM, Guglielmino CR, Lanzavecchia SB, Juri M, Vera T, Cladera J, Scolari F, Gomulski L, Bonizzoni M, Gasperi G, Silva JG and Malacrida AR (2015) Relevant genetic differentiation among Brazilian populations of Anastrepha fraterculus (Diptera, Tephritidae). ZooKeys 540:157–173

Mauzy-Melitz D, Waring GL (2003) fc177, a minor dec-1 proprotein, is necessary to prevent ectopic aggregation of the endochorion during eggshell assembly in Drosophila. Dev Biol 255:193–205. https://doi.org/10.1016/s0012-1606(02)00084-2

Mayr E (1959) Isolation as an evolutionary factor. Proc Am Phil Soc 103:221–230

McAllister BF, McVean GAT (2000) Neutral evolution of the sex-determining gene transformer in Drosophila. Genetics 154:1711–1720

McClellan DA, McCracken KG (2001) Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. Mol Biol Evol 18:917–925

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–652

Mengual X, Kerr P, Norrbom AL, Barr NB, Lewis ML, Stapelfeldt AM, Scheffer SJ, Woods P, Islam M-S, Korytkowski CA, Uramoto K, Rodriguez EJ, Sutton BD, Nolazco N, Steck GJ, Gaimari S (2017) Phylogenetic relationships of the tribe Toxotrypanini (Diptera: Tephritidae) based on molecular characters. Mol Phylogen Evol 113:84–112. https://doi.org/10.1016/j.ympev.2017.05.011

Mészáros B, Erdős G, Dosztányi Z (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res 46:W329–W337. https://doi.org/10.1093/nar/gky384

Murillo T, Jirón LF (1994) Egg morphology of Anastrepha obliqua and some comparative aspects with eggs of Anastrepha fraterculus (Diptera:Tephritidae). Fla Entomol 77:342–348

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764. https://doi.org/10.1371/journal.pgen.1002764

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K (2013) FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. Mol Biol Evol 30:1196–1205

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SL (2015) Gene-wide identification of episodic selection. Mol Biol Evol 32:1365–1371. https://doi.org/10.1093/molbev/msv035

Nelson JE, Krawetz SA (1992) Purification of cloned and genomic Dna by guanidine thiocyanate isobutyl alcohol fractionation. Anal Biochem 207:197–201

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Noguerón MI, Waring GL (1995) Regulated processing of dec-1 eggshell proteins in Drosophila. Dev Biol 172:272–279. https://doi.org/10.1006/dbio.1995.0022

Noguerón MI, Mauzy-Melitz D, Waring GL (2000) Drosophila dec-1 eggshell proteins are differentially distributed via a multistep extracellular processing and localization pathway. Dev Biol 225:459–470. https://doi.org/10.1006/dbio.2000.9805

Norrbom AL, Barr NB, Kerr P, Mengual X, Nolazco N, Rodriguez EJ, Steck GJ, Sutton BD, Uramoto K, Zucchi RA (2018) Synonymy of Toxotrypana Gerstaecker with Anastrepha Schiner (Diptera: Tephritidae). Proc Entomol Soc Washington 120:834–841. https://doi.org/10.4289/0013-8797.120.4.834

Norrbom AL, Muller A, Gangadin A, Sutton BD, Rodriguez EJ, Savaris M, Lampert S, Clavijo PAR, Steck GJ, Moore MR, Nolazco N, Troya H, Keil CB, Padilla A, Wiegmann BM, Cassel B, Branham M, Ruiz-Arce R (2021) New species and host plants of Anastrepha (Diptera: Tephritidae) primarily from Suriname and Pará, Brazil. Zootaxa 5044:1–74. https://doi.org/10.11646/zootaxa.5044.1.1

Norrbom AL, Korytkowski CA, Zucchi RA, Uramoto K, Venable GL, McCormick J and Dallwitz M (2012) Anastrepha and Toxotrypana: descriptions, illustrations, and interactive keys

Norrbom AL (2004) Updates to biosystematic database of world Diptera for Tephritidae through 1999, Diptera Data Dissemination Disk

Novotny M, Seibert M, Kleywegt GJ (2007) On the precision of calculated solvent-accessible surface areas. Acta Crystallogr D Biol Crystallogr 63:270–274. https://doi.org/10.1107/s0907444906044118

Palmer J and Stajich JE (2019) nextgenusfs/funannotate: funannotate v1.5.3 (1.5.3). Zenodo. https://doi.org/10.5281/zenodo.2604804

Panhuis TM, Swanson WJ (2006) Molecular evolution and population genetic analysis of candidate female reproductive genes in Drosophila. Genetics 173:2039–2047. https://doi.org/10.1534/genetics.105.053611

Parker DJ, Gardiner A, Neville MC, Ritchie MG, Goodwin SF (2014) The evolution of novelty in conserved genes; evidence of positive selection in the Drosophila fruitless gene is localised to alternatively spliced exons. Heredity 112:300–306. https://doi.org/10.1038/hdy.2013.106

Prezotto LF, Perondini ALP, Hernández-Ortiz V, Frías D, Selivon D (2019) What can integrated analysis of morphological and genetic data still reveal about the *Anastrepha fraterculus* (Diptera: Tephritidae) cryptic species complex? Insects 10:408. https://doi.org/10.3390/insects10110408

Pupko T, Sharan R, Hasegawa M, Shamir R, Graur D (2003) Detecting excess radical replacements in phylogenetic trees. Gene 319:127–135. https://doi.org/10.1016/s0378-1119(03)00802-3

Rabosky DL (2016) Reproductive isolation and the causes of speciation rate variation in nature. Biol J Linn Soc 118:13–25. https://doi.org/10.1111/bij.12703

Raza MF, Yao Z, Bai S, Cai Z, Zhang H (2020) Tephritidae fruit fly gut microbiome diversity, function and potential for applications. Bull Entomol Res 110:423–437. https://doi.org/10.1017/s0007485319000853

Rezende VB, Congrains C, Lima ALA, Campanini EB, Nakamura AM, Oliveira JL, Chahad-Ehlers S, Sobrinho I, de Brito RA (2016) Head transcriptomes of two closely related species of fruit flies of the Anastrepha fraterculus group reveals divergent genes in species with extensive gene flow. G3 6:3283–3295. https://doi.org/10.1534/g3.116.030486

Ruiz-Montoya L, Vallejo RV, Haymer D, Liedo P (2020) Genetic and ecological relationships of Anastrepha ludens (Diptera: Tephritidae) populations in Southern Mexico. Insects 11:815. https://doi.org/10.3390/insects11110815

Rull J, Tadeo E, Lasa R, Rodriguez CL, Altuzar-Molina A, Aluja M (2018) Experimental hybridization and reproductive isolation between two sympatric species of tephritid fruit flies in the Anastrepha fraterculus species group. Insect Sci 25:1045–1055. https://doi.org/10.1111/1744-7917.12489

Santos Pd, Uramoto K, Matioli SR (2001) Experimental hybridization among Anastrepha species (Diptera: Tephritidae): production and morphological characterization of F1 hybrids. Ann Entomol Soc Am 94:717–725

Schneider K, White TJ, Mitchell S, Adams CE, Reeve R, Elmer KR (2021) The pitfalls and virtues of population genetic summary statistics: detecting selective sweeps in recent divergences. J Evol Biol 34:893–909. https://doi.org/10.1111/jeb.13738

Schutze MK, Virgilio M, Norrbom A, Clarke AR (2017) Tephritid integrative taxonomy: where we are now, with a focus on the resolution of three tropical fruit fly species complexes. Annu Rev Entomol 62:147–164. https://doi.org/10.1146/annurev-ento-031616-035518

Scully ED, Geib SM, Mason CJ, Carlson JE, Tien M, Chen H-Y, Harding S, Tsai C-J, Hoover K (2018) Host-plant induced changes in microbial community structure and midgut gene expression in an invasive polyphage (Anoplophora glabripennis). Sci Rep 8:9620. https://doi.org/10.1038/s41598-018-27476-0

Segura MD, Callejas C, Fernández MP, Ochando MD (2006) New contributions towards the understanding of the phylogenetic relationships among economically important fruit flies (Diptera: Tephritidae). Bull Entomol Res 96:279–288. https://doi.org/10.1079/ber2006425

Selivon D, Perondini ALP (1998) Eggshell morphology in two cryptic species of the Anastrepha fraterculus complex (Diptera: Tephritidae). Ann Entomol Soc Am 91:473–478

Selivon D, Perondini ALP, Morgante JS (2005) A genetic-morphological characterization of two cryptic species of the Anastrepha fraterculus complex (Diptera: Tephritidae). Ann Entomol Soc Am 98:367–381

Selivon D, Vretos C, Fontes L and Perondini ALP (2004) New variant forms in the *Anastrepha fraterculus* complex (Diptera: Tephritidae), in: Barnes BN (Ed.), Proceedings of the 6th international symposium on fruit flies of economic importance, Stellenbosch, South Africa, 6–10 May 2002. Isteg Scientific Publications, Johannesburg, pp 253–258

Selivon D, Perondini ALP, Hernández-Ortiz V, doVal FC, Camacho A, Gomes FR and Prezotto LF (2022) Genetical, morphological, behavioral, and ecological traits support the existence of three Brazilian species of the Anastrepha fraterculus complex of cryptic species. Front Ecol Evol 10:836608. https://doi.org/10.3389/fevo.2022.836608

Sémon M, Wolfe KH (2007) Consequences of genome duplication. Curr Opin Genet Dev 17:505–512. https://doi.org/10.1016/j.gde.2007.09.007

Sim SB, Congrains C, Velasco-Cuervo SM, Corpuz RL, Kauwe AN, Scheffler B, Geib SM (2024) Genome report: chromosome-scale genome assembly of the West Indian fruit fly Anastrepha obliqua (Diptera: Tephritidae). G3 (Bethesda) 14(4):jkae024. https://doi.org/10.1093/g3journal/jkae024

Smith N (2003) Are radical and conservative substitution rates useful statistics in molecular evolution? J Mol Evol 57:467–478

Smith-Caldas MRB, Mcpheron BA, Silva JG, Zucchi RA (2001) Phylogenetic relationship among species of the fraterculus group (Anastrepha: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I gene. Neotrop Entomol 30:565–573

Sobrinho I, de Brito R (2010) Evidence for positive selection in the gene fruitless in Anastrepha fruit flies. BMC Evol Biol 10:293–293

Sobrinho IS, de Brito RA (2012) Positive and purifying selection influence the evolution of Doublesex in the Anastrepha fraterculus species group. PLoS ONE 7:e33446. https://doi.org/10.1371/journal.pone.0033446

Solferini VN, Morgante JS (1987) Karyotype study of eight species of Anastrepha (Diptera, Tephritidae). Caryologya 40:229–241

Spangenberg DK, Waring GL (2007) A mutant dec-1 transgene induces dominant female sterility in Drosophila melanogaster. Genetics 177:1595–1608. https://doi.org/10.1534/genetics.107.080168

Studer RA, Robinson-Rechavi M (2009) Evidence for an episodic model of protein sequence evolution. Biochem Soc Trans 37:783–786. https://doi.org/10.1042/bst0370783

Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. Nat Rev Genet 3:137–144

Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20

Takahashi Y, Hirose F, Matsukage A, Yamaguchi M (1999) Identification of three conserved regions in the DREF transcription factors from Drosophila melanogaster and Drosophila virilis. Nucleic Acids Res 27:510–516. https://doi.org/10.1093/nar/27.2.510

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc Natl Acad Sci USA 87:2419–2423

Templeton AR (1996) Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics 144:1263–1270

Templeton AR (2008) The reality and importance of founder speciation in evolution. BioEssays 30:470–479. https://doi.org/10.1002/bies.20745

Trivedi R, Nagarajaram HA (2022) Intrinsically disordered proteins: an overview. Int J Mol Sci 23:14050. https://doi.org/10.3390/ijms232214050

Turner LM, Hoekstra HE (2006) Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol Biol Evol 23:1656–1669

UniProt_Consortium (2019) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506-d515. https://doi.org/10.1093/nar/gky1049

Vera MT, Cáceres C, Wornoayporn V, Islam A, Robinson AS, De La Vega MH, Hendrichs J, Cayol J-P (2006) Mating incompatibility among populations of the South American fruit fly Anastrepha fraterculus (Diptera: Tephritidae). Ann Entomol Soc Am 99:387–397. https://doi.org/10.1603/0013-8746(2006)099[0387:MIAPOT]2.0.CO;2

Virgilio M, Jordaens K, Verwimp C, White IM, De Meyer M (2015) Higher phylogeny of frugivorous flies (Diptera, Tephritidae, Dacini): localised partition conflicts and a novel generic classification. Mol Phylogen Evol 85:171–179. https://doi.org/10.1016/j.ympev.2015.01.007

Waring GL (2000) Morphogenesis of the eggshell in Drosophila. Int Rev Cytol 198:67–108. https://doi.org/10.1016/s0074-7696(00)98003-3

Waring GL, Hawley RJ, Schoenfeld T (1990) Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by alternative RNA splicing and proteolytic cleavage events. Dev Biol 142:1–12. https://doi.org/10.1016/0012-1606(90)90146-A

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296-w303. https://doi.org/10.1093/nar/gky427

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL (2018) Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol 35:773–777. https://doi.org/10.1093/molbev/msx335

Weber CC, Whelan S (2019) Physicochemical amino acid properties better describe substitution rates in large populations. Mol Biol Evol 36:679–690. https://doi.org/10.1093/molbev/msz003

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2014) RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol 32:820–832. https://doi.org/10.1093/molbev/msu400

White IM (2006) Taxonomy of the Dacina (Diptera: Tephritidae) of Africa and the Middle East. African Entomol 2:1–156

Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA (2003) TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19:671–672

Xia X, Li WH (1998) What amino acid properties affect protein evolution? J Mol Evol 47:551–556

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13:555–556

Yang Z (2002) Inference of selection from multiple species alignments. Curr Opin Genet Dev 12:688–694

Yang Z, Nielsen R, Hasegawa M (1998) Models of amino acid substitution and applications to mitochondrial protein evolution. Mol Biol Evol 15:1600–1611

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Zucchi RA, Malavasi A, Nascimento AS, Walder JMM (2004) Prejuízos das moscas-das-frutas na exportação de citros. Visão Agrícola 2:73–77

Zucchi RA (2000) Espécies de Anastrepha, sinonímias, plantas hospedeiras e parasitóides, in: Malavasi A and Zucchi RA (Eds.), Moscas-das-frutas de importância econômica no Brasil: Conhecimento básico e aplicado. Holos Editora, Ribeirão Preto, pp. 41–48

Authorship: Bisio, M. de C. and Santos, E. M. dos, share the first authorship of this manuscript.