formerly Typeset

Open access • Posted Content • DOI:10.1101/510586

Genomics and transcriptomics to unravel sex determination pathway and its evolution in sand flies — Source link

Petrella, Serena Aceto, Colonna, Giuseppe Saccone ...+6 more authors

Institutions: University of Naples Federico II, National Research Council, Stazione Zoologica Anton Dohrn, Charles University in Prague ...+1 more institutions

Published on: 04 Jan 2019 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Nematocera

Related papers:

- The evolution of sex determination systems in dipteran insects other than Drosophila.
- Sex Determination Cascade in Insects: A Great Treasure House of Alternative Splicing
- Mechanism of sex determination in mammals
- The transformer gene of Ceratitis capitata : a paradigm for a conserved epigenetic master regulator of sex determination in insects
- Comparative Genomic Analysis of Drosophila melanogaster and Vector Mosquito Developmental Genes

Share this paper:

View more about this paper here: https://typeset.io/papers/genomics-and-transcriptomics-to-unravel-sex-determination-3uw98fnsyi

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Genomics and transcriptomics to unravel sex determination pathway and its evolution in sand flies**

2

3 Valeria Petrella[1], Serena Aceto[1], Vincenza Colonna[2], Giuseppe Saccone[1], Remo Sanges[3], Nikola Polanska[4], Petr Volf[4],

4 Luigi Gradoni[5], Gioia Bongiorno[5], Marco Salvemini[1,]*

5

6 Valeria Petrella, valeria.petrella@unina.it; Department of Biology, University of Naples Federico II, Naples, Italy.

7 Serena Aceto, serena.aceto@unina.it; Department of Biology, University of Naples Federico II, Naples, Italy.

8 Giuseppe Saccone, giuseppe.saccone@unina.it; Department of Biology, University of Naples Federico II, Naples, Italy.

9 Remo Sanges, remo.sanges@szn.it, Stazione Zoologica "Anton Dohrn", Naples, Italy.

10 Vincenza Colonna, vincenza.colonna@igb.cnr.it; National Research Council, Institute of Genetics and Biophysics, Naples, Italy.

11 Nikola Polanska, nikola.polanska@natur.cuni.cz; Department of Parasitology, Charles University, Prague, Czech Republic.

12 Petr Volf, volf@cesnet.cz; Department of Parasitology, Charles University, Prague, Czech Republic.

13 Luigi Gradoni, luigi.gradoni@iss.it; Unit of Vector-borne Diseases, Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy.

14 Gioia Bongiorno, gioia.bongiorno@iss.it; Unit of Vector-borne Diseases, Department of Infectious Diseases, Istituto Superiore di Sanità, Rome, Italy.

15 Marco Salvemini*, marco.salvemini@unina.it; Department of Biology, University of Naples Federico II, Naples, Italy.

16

17 *Corresponding author.

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

42 **Abstract**

43

44 **Background**

45 Phlebotomine sand flies (Diptera, Nematocera) are important vectors of several pathogens, including
46 *Leishmania* parasites, causing serious diseases of humans and dogs. Despite their importance as disease
47 vectors, most aspects of sand fly biology remain unknown including the molecular bases of their
48 reproduction and sex determination, aspects also relevant for the development of novel vector control
49 strategies.

50

51 **Results**: Using a comparative genomics/transcriptomics approach, we identified the sex determining genes in
52 phlebotomine sand flies and proposed the first model for the sex determination cascade of these insects. For
53 all the genes identified, we produced manually curated gene models, developmental gene expression profile
54 and performed evolutionary molecular analysis. We identified and characterized, for the first time in a
55 Nematocera species, the *transformer* (*tra*) homolog which exhibits both conserved and novel features. The
56 analysis of the *tra* locus in sand flies and its expression pattern suggest that this gene is able to autoregulate
57 its own splicing, as observed in the fruit fly *Ceratitis capitata* and several other insect species.

58

59 **Conclusions**: Our results permit to fill the gap about sex determination in sand flies, contribute to a better
60 understanding of this developmental pathway in Nematocera and open the way for the identification of sex
61 determining orthologs in other species of this important Diptera sub-order. Furthermore, the sex
62 determination genes identified in our work also provide the opportunity of future biotech applications to
63 control natural population of sand flies, reducing their impact on public health.

64

65

66 **Keywords**: Sex determination, sand fly, Nematocera, genomics, transcriptomics, alternative splicing.

67

68

69

70

71

72

73

74

75

76

77

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

78 **Background**

79 In animals, sex determination is the process by which early embryos of metazoan species with sexual
80 reproduction operate a binary decision between two conditions: male or female development. This key
81 decision results in individuals that can be identified as males, females, or in some cases hermaphrodites and,
82 in species with a genetic sex determination system, underlies genomic differences between sexes. In most
83 cases, the presence of heteromorphic sexual chromosomes represents the primary signal for sex
84 determination. According to the initial decision, the primary signal is then transduced, through a genetic
85 pathway organized in a cascade of regulatory genes, to downstream regulators responsible for sexual
86 differentiation [1–3].

87 Insects are among the largest taxonomic animal groups on Earth and, not surprisingly, they exhibit a wide
88 variety of sex determining systems, with highly variable primary signals and widely conserved genetic
89 transduction mechanisms to downstream regulators [4–6]. *Drosophila melanogaster* (Diptera, Drosophilidae)
90 is the model species where sex determination is known at the higher level of molecular resolution (Fig. 1). In
91 this species, sex determination is controlled by five main genes, *Sex-lethal* (*Sxl*), *transformer* (*tra*),
92 *transformer-2* (*tra-2*), *doublesex* (*dsx*) and *fruitless* (*fru*), hierarchically organized in a regulative cascade: *Sxl*
93 -> *tra*+*tra-2* -> *dsx*, *fru*. This cascade is activated by a primary signal represented by the number of X
94 chromosomes [7,8]. In the last 20 years, homology-based approaches in species belonging to various insects
95 orders (Diptera, Coleoptera, Lepidoptera, Hymenoptera) led to discover only partial conservation of the
96 *Drosophila* sex determination genetic pathway: in all species studied the *Sxl* ortholog was not involved in
97 sex determination while the *tra* ortholog is able to control the female-splicing of its own pre-mRNA as well
98 as to control, similarly to *Drosophila*, the female-specific splicing of the *dsx* and *fru* downstream genes
99 [6,9,10]. In female embryos, the maternal *tra* contribution establishes the female-specific autoregulatory
100 splicing of *tra* and leads to female development, which is epigenetically maintained during development in
101 the absence of the initial positive signal. In male embryos, the establishment of *tra* autoregulatory feedback
102 loop is impaired by the presence of a masculinizing factor able to interfere with the maternal and/or the
103 zygotic *tra* function, blocking its positive autoregulation and leading to male development, as shown recently
104 in *Musca domestica* (Sharma et al., 2017). Hence, the *tra*+*tra2*->*dsx*/fru sex determination module with an
105 autoregulating *tra*, firstly discovered in the Mediterranean fruit fly *Ceratitis capitata* (Pane et al., 2002;
106 Salvemini et al., 2009a), represents the core pathway of insect sex determination [13]. The only remarkable
107 exception is represented by the Lepidoptera order, where a different sex determination system exists with the
108 primary signal constituted by a small RNA, the absence of the *tra* ortholog and the *dsx* splicing controlled by
109 different splicing regulators [14].

110 Within Diptera, the insect order where sex determination has been studied in the largest number of species,
111 the *tra* ortholog has been identified only in species belonging to the Brachycera suborder [15–21]. For the
112 basal suborder Nematocera, which includes very important hematophagous vector species such as
113 mosquitoes, sand flies and black flies, the *tra* ortholog or its functional analog has not yet been found in any

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

114 species and limited knowledge is available in general about sex determination, mainly restricted to mosquito
115 species [22–29] (Fig. 1).
116 Within Nematocera, phlebotomine sand flies are second only to mosquitoes in importance as a vector of
117 pathogens that cause diseases to humans and animals worldwide, including leishmaniases, sand fly fever,
118 meningitis, vesicular stomatitis and Chandipura virus encephalitis [30]. Among the over 800 species of sand
119 fly described to date, 98 are proven or suspected vectors of human leishmaniases; these include 42
120 *Phlebotomus* species in the Old World and 56 *Lutzomyia* species (*sensu*) in the New World [31].
121 Leishmaniasis are diseases of great public health concern, being endemic in over 98 countries, with more
122 than 350 million people at risk and 2,357,000 disability-adjusted life years lost [32]. It is estimated that about
123 1.3 million new cases of leishmaniasis (0.2-0.4 million visceral and 0.7-1.2 million cutaneous leishmaniasis)
124 occur every year, with 20,000-40,000 deaths caused by the visceral form. With expanding endemicity,
125 leishmaniasis is becoming a worldwide re-emerging public health problem [33].
126 Despite their importance as disease vectors, most aspects of sand fly biology remain unknown, including sex
127 determination and sexual differentiation. To fill this gap and contribute to a better understanding of the
128 evolution of sex determination mechanisms in insects, in the present study we applied a comparative
129 genomic/transcriptomic approach to identify and characterize sex determining genes in sand fly species. For
130 the first time we present in a unique study the analysis of the key components of the sex determining
131 cascade, also identifying, the first *transformer* homolog in a Nematocera species.
132
133
134 **Results and Discussion**
135
136 **Identification of *PpeSxl*, *Ppetra*, *Ppetra-2*, *Ppedsx, and Ppefru* sex determining genes in the sand fly**
137 ***Phlebotomus perniciosus*.** In the Old World, the sand fly *Phlebotomus perniciosus* (Diptera, Nematocera) is
138 the main vector of *Leishmania infantum* (Kinetoplastida: Trypanosomatidae), the parasitic protozoan that
139 causes visceral and cutaneous leishmaniasis in humans and canine reservoir host, as well as of various
140 known and emerging arboviruses considered relevant from an European public health perspective (Toscana
141 Virus, Naples Virus, Sicilian Virus) [34]. Proteins encoded by insect sex determining genes are characterized
142 by domains very well conserved across insect orders and distinctive of each gene family: the DNA-binding
143 DM (Doublesex Mab3) domain for the DSX proteins [35], the protein-protein BTB (Broad-Complex,
144 Tramtrack and Bric a brac) binding domain for the FRU proteins [36] and the RNA-binding RRM (RNA
145 Recognition Motif) domain for SXL and TRA-2 proteins [37]. Conversely, the female-specific serine-
146 arginine rich TRA protein exhibits a general low conservation of its primary sequence and the absence of
147 functional characterized domains. In most of the insect species analyzed to date, the only conserved parts of
148 the TRA protein are the TRACAM (*Ceratitis*-*Apis*-*Musca*) domain, putatively involved in the autoregulation

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

149 of the *tra* gene, the DIPTERA domain, and the HYMENOPTERA domain, the last two with unknown
150 function [13,18,38].
151 We performed a TBLASTN search against the available *P. perniciosus* adult transcriptome database
152 (http://pernibase.evosexdevo.eu) [39] to identify transcripts encoding for sex determining proteins, using
153 other insects sex determining protein sequences as query terms (Additional file 1: Table S1). In *P.*
154 *perniciosus* we identified the complete open reading frames (ORF) of the transcripts encoding for the
155 putative SXL, TRA-2 and male- and female-specific isoforms of DSX (Additional file 2: Figures S1-S5). We
156 named the corresponding genes as *PpeSxl*, *Ppetra-2*, and *Ppedsx*. In addition, we identified partial ORFs
157 encoding for FRU, and we named the gene as *Ppefru*. The incomplete transcripts encoding for FRU proteins
158 lack their 3' ends and therefore complete ORFs were obtained by 3' RACE (Additional file 2: Figures S6-
159 S7), as described in supplementary methods. Using the TBLASTN approach no *tra* ortholog was found in
160 the *P. perniciosus* transcriptome. This result was expected due to the low level of nucleotide and protein
161 sequence conservation of the *tra* gene, also in closely related insect species and considering that the cloning
162 of *tra* in *Ceratitis* was performed by synteny rather than by direct homology [15,16,40,41].
163 We validated the transcription and the splicing pattern of *PpeSxl*, *Ppetra-2*, *Ppedsx* and *Ppefru* by RT-PCR
164 on mRNAs extracted from adult *P. perniciosus* males and females, using the *Ppesod* gene as endogenous
165 positive control and to exclude genomic DNA contamination of the cDNAs (Fig. 2A). The RT-PCR primer
166 pairs for the *PpeSxl* transcript amplified in both sexes multiple non-sex-specific transcripts probably
167 produced by alternative splicing (Additional file 2: Figure S8), as observed in other insect species [42,43].
168 Functional analyses of *Sxl* in several dipteran species [44,45] show that *Sxl* is a master switch gene of sex
169 determination only in Drosophilidae [9,46]. Therefore, we supposed that *Sxl* is probably not essential for the
170 sex determination in *P. perniciosus* and decided to exclude it from further analyses.
171 The RT-PCR analysis of the *Ppetra-2* transcript showed a non-sex-specific expression at adult stage and
172 revealed the existence of a second isoform (*Ppetra-2B*) expressed in both sexes (Fig. 2B). Cloning and
173 sequencing of *Ppetra-2B* showed that it encodes for a putative TRA-2 protein with slight amino acid (aa)
174 differences in the N-terminus respect to PpeTRA-2A. A similar *tra-2* non-sex-specific splicing pattern was
175 reported in the whitefly *Bemisia tabaci*, where the two encoded TRA-2 isoforms differ at their N-terminus
176 for a wider region of 123 aa [47].
177 The RT-PCR analysis of the *Ppedsx* and *Ppefru* transcripts revealed that both genes are regulated by sex-
178 specific alternative splicing as in other insect species (Fig. 2C-2D). Notably, in both *Ppedsx* and *Ppefru*
179 female-specific transcripts we identified a cluster of putative TRA/TRA-2 binding sites (Fig. 2E).
180 In Diptera, the presence of a conserved TRA/TRA-2 binding site cluster in *dsx* and *fru* genes is always
181 associated to the presence of the TRA active protein [15]. Encouraged by finding conserved TRA/TRA-2
182 binding sites in *Ppedsx* and *Ppefru* and by the presence of a PpeTRA-2 with a highly conserved RRM
183 domain (Additional file 2: Figure S2), we pursued a strategy to identify the ortholog of *tra* in *P. perniciosus.*
184 This approach was based on the hypothesis that also in sand flies the *tra* gene could regulate its own sex-

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

185 specific alternative splicing binding a cluster of TRA/TRA-2 binding sites. Therefore, we analyzed the *P.*
186 *perniciosus* adult transcriptome with the DREG tool of the Emboss Suite (http://emboss.sourceforge.net/) to
187 detect transcripts containing putative TRA/TRA-2 binding sites. We identified an assembled transcript
188 (c23543.g1.i2, 3858 bp-long) containing the highest number of TRA/TRA-2 binding sites, with six elements
189 clustered in a 324bp-long sequence (Fig. 3A) and located between two putative exons encoding for a serine-
190 arginine rich sequence. Using RT-PCR primer pairs spanning the region containing the TRA/TRA-2 binding
191 sites we were able to amplify two male-specific (M1 and M2) and three female-specific (F1, F2 and F3)
192 cDNA fragments (Fig. 3B), demonstrating that the c23543.g1.i2 transcript undergoes sex-specific alternative
193 splicing regulation, as expected for a *tra* ortholog. The five full-length cDNAs were cloned and sequenced
194 after 5' and 3' RACE experiments, performed as described in Methods. The virtual translation of the five
195 cDNAs revealed that M1, M2, F2 and F3 encode for very short polypeptides due to premature stop codons.
196 Only the female-specific F1 cDNA has a full ORF and encodes for a SR rich sequence (282 aa) containing a
197 short region similar to the TRA Diptera domain (Additional file 2: Figure S9). We named this putative
198 protein as PpeTRA and the corresponding gene as *Ppetra*. PpeTRA is missing the putative autoregulation
199 TRACAM domain, which is present in all the known autoregulative TRA proteins of insects. To date,
200 PpeTRA represents the shortest insect TRA protein, excluding the non-autoregulating TRA of *D.*
201 *melanogaster* (197 aa) (Additional file 2: Figure S9).
202
203 **Developmental expression analysis of sex determining genes in *P. perniciosus*.** We performed an RT-
204 PCR analysis on total RNA extracted from samples of mixed sexes from different developmental stages
205 (embryos, larvae of 1st, 2nd and 4th instar and pupae) to analyze the developmental expression pattern of the
206 sex determining genes newly identified in *P. perniciosus* (Fig. 4). We used the *Ppesod* gene, constitutively
207 expressed in *P. perniciosus* (Petrella et al., 2015), as endogenous positive control, and the same primer pairs
208 of the RT-PCR analyses performed on adult samples, spanning the alternatively spliced regions of *tra*, *tra-2*,
209 *dsx* and *fru* genes, as reported in figure 2 and 3. We found that *Ppetra* is expressed since embryonic stage as
210 observed for other dipteran species [16,18,48], producing sex-specific transcripts. We amplified, in all
211 developmental stages, fragments of 446bp, 678 bp and 1065 bp corresponding to *Ppetra* F1, M1 and F3
212 transcripts, respectively (Fig. 4). *Ppetra-2* is expressed from the first instar larval stage until adulthood,
213 differently from other dipteran species, such as *Ceratitis capitata* and *Musca domestica*, where it is
214 expressed also at embryonic stage [49,50]. Both the *Ppetra-2A* and *B* transcripts were detected in all stages
215 but embryos (Fig. 4). *Ppedsx* and *Ppefru* are expressed from first-instar larval stage and second-instar larval
216 stages, respectively, until adulthood, both producing sex-specific transcripts by alternative-splicing during
217 development (Fig. 4). *Ppedsx* developmental expression pattern seems to be different respect to other
218 dipteran species, including *Drosophila*, *C. capitata* and the tiger mosquito *Aedes aegypti*, where *dsx* is
219 expressed also at the embryonic stage [23,51,52]. Conversely, the *Ppefru* developmental expression pattern

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

220 is conserved respect to *fru*-P1 promoter expression pattern observed in *Drosophila* and in *A. aegypti*
221 (Salvemini et al., 2010; 2013) with expression starting at late larval stage until adulthood.
222

223 **Evolution of *tra* genomic organization and of alternative splicing regulation in Phlebotominae.**

224 The *Ppetra* gene is the first *tra* ortholog isolated in a Nematocera species and the shortest *tra* gene (1.7Kb)
225 isolated to date in insects. To study its genomic organization we amplified, cloned and sequenced the 1725
226 bp fragment corresponding to the *Ppetra* locus, using a primer pair located in the 5' and 3' UTR of *Ppetra*
227 transcripts and adult genomic DNA. Aligning genomic *Ppetra* against the five *Ppetra* cDNA sequences, we
228 reconstruct the exon-intron organization of the *Ppetra* gene and identified the alternative splicing events
229 producing the *Ppetra* transcript isoforms (Fig. 5A). The *Ppetra* gene has four exons and three introns, all
230 with conserved GT-AG boundaries (Additional file 3: Figure S10). In females, *Ppetra* produces three
231 transcripts. Exon 1, 2, 3 and 4 are used to produce a mature mRNA corresponding to the F1 transcripts, with
232 an ORF encoding for the 282 aa-long PpeTRA protein. In addition to this, distinct parts of intron 2 are
233 retained in two other transcripts, one by an alternative 3' acceptor splicing site (transcript F2) and the other
234 by an intron retention mechanism (transcript F3). In both the F2 and F3 transcripts the presence of in-frame
235 stop codons causes short truncated PpeTRA isoforms. In males, *Ppetra* produces two transcripts: the M2
236 transcript is an unspliced transcript because it retains all the introns, while the M1 transcript is produced
237 through an alternative 5' donor splicing site choice. In the two *Ppetra* males-specific transcripts, the
238 introduction of premature stop codons leads to short truncated PpeTRA isoforms.
239 To study the evolution of the genomic organization and alternative splicing regulation in sand flies, we
240 searched the *tra* orthologs in seven other Phlebotominae species by TBLASTN using PpeTRA as query. For
241 two species, *P. papatasi* and *L. longipalpis*, genome and transcriptome assemblies were available (PpapI1,
242 PpapI1.4, LlonJ1 and LlonJ1.4; https://www.vectorbase.org/). For the other two Old World sand fly species,
243 *P. bergeroti* and *P. duboscqi*, we assembled a draft genome using available sequencing data and the MINIA
244 genome assembler [53]. In addition, we produced *de novo* transcriptome assemblies by using all the available
245 sequencing data (up to March 2018) for *L. longipalpis* and for two New World species, *L.* (*Nyssomyia*)
246 *umbratilis* and *L.* (*Nyssomyia*) *neivai* using the Trinity *de novo* assembler [54,55] (see Additional file 15:
247 supplementary methods).
248 By querying PpeTRA against the genomes and transcriptomes of the Phlebotominae species, we identified
249 the *tra* ortholog in *P. papatasi* (*Ppatra*), *P. bergeroti* (*Pbetra*) and *P. duboscqi* (*Pdutra*) but this approach
250 could not identify any ortholog in the genomes/transcriptomes of the three New World sand fly species.
251 Furthermore, neither the TRA/TRA-2 binding sites *in silico* approach, that led to the identification of *Ppetra*,
252 nor a molecular approach in *L. longipalpis* by touch down RT-PCR with degenerated primers designed on
253 the alignment of *Ppetra* and *Ppatra* sequences (data not shown) could identify *tra* in New World sand flies.
254 Similarly, Geuverink and Beukeboom [38] *in silico* identified a putative *tra* gene in the Old World sand fly
255 species *P. papatasi* and reported its apparent absence in the New World sand fly *L. longipalpis.* As for

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

256 *Ppetra*, the putative *P. papatasi* TRA missed the most consistent recognition motif of a TRA protein, i.e. the
257 putative autoregulation TRACAM domain, leading the authors to be cautious about the true nature of their
258 identified sand fly *tra* ortholog [38].
259 We reconstructed gene models for the *P. papatasi*, *P. bergeroti* and *P. duboscqi tra* orthologs (Additional
260 file 3: Figures S11-S13), which encode for a 311 aa-long SR-protein with 61% identity respect to the
261 PpeTRA and missing, as in *P. perniciosus*, a conserved TRACAM domain (Additional file 4: Figure S14).
262 The four *Phlebotomus tra* genes revealed a conserved genomic organization with four exons and three
263 introns, with small differences in exons/introns lengths (Fig. 5B). In the intron 2 of *P. papatasi*, *P. bergeroti*
264 and *P. duboscqi*, we identified, as in *P. perniciosus*, six conserved TRA/TRA-2 binding sites (Fig. 5C). To
265 study the alternative splicing regulation of the *tra* gene in *Phlebotomus* species, we compared the intronic
266 sequences of the four species (Additional file 5: Figure S15). As in *P. perniciosus*, all *tra* introns exhibit
267 conserved GT-AG terminal dinucleotides. Intron 2, which is regulated by sex-specific alternative splicing in
268 *P. perniciosus*, has a putative conserved alternative splicing sites (SS) also in *P. papatasi*, *P. bergeroti* and *P.*
269 *duboscqi*. In the four species, the 5' donor SS of intron 2 seems to be weak and suboptimal, while the 3'
270 acceptor SS is a canonical strong splicing site. Finally, all the four species have a strong canonical male-
271 specific alternative 5' donor SS at about 230 bp downstream of exon 2 (Additional file 5: Figure S15). These
272 findings led us to suppose that in *P. perniciosus*, as well as in the other three *Phlebotomus* species, the male-
273 specific splicing of the *tra* pre-mRNA represents the default splicing mode. In contrast, in females, the
274 repression of the male-specific 5' donor SS of intron 2 is most probably due to the binding of TRA and
275 TRA-2 proteins on the TRA/TRA-2 binding site cluster, leading to the usage of the upstream 5' donor SS to
276 form to the female-specific *tra* transcript, thus producing a functional TRA only in females. This hypothesis
277 on the conserved splicing regulation was confirmed in *P. papatasi* by RT-PCR on adult RNA from males
278 and females (Additional file 6: Figure S16).
279 Figure S17 shows a comparison of the *tra* genomic locus among insect species (Additional file 6: Figure
280 S17). Despite differences in exon number and intron length, the sex-specific splicing regulation of the *tra*
281 gene exhibits a striking conservation. In all the considered species, including *P. perniciosus*, an alternative 5'
282 donor SS choice leads to a full TRA protein only in the female sex. To study the protein organization, we
283 compared PpeTRA with other arthropod TRA proteins (Fig. 6A). PpeTRA exhibits similar domain
284 organization respect to insect TRAs, with a DIPTERA domain located within the RS domain, as observed
285 also in TRA of *Lucilia cuprina* (LcTRA), *Cochliomyia hominivorax* (ChTRA) and *Glossina morsitans*
286 (GmTRA). At the same time, PpeTRA misses the TRACAM domain; the *Ppetra* regions corresponding to
287 the last 31 nucleotides of exon 2 and to the first 45 nucleotides of exon 3 (upstream and downstream of the
288 *Ppetra* sex-specifically regulated intron, respectively) encode for a PpeTRA protein portion that exhibits
289 only 8 out of 25 conserved amino acids respect to the insect TRACAM domain (Fig. 6B).
290 In conclusion, the conserved *tra* structure, the conserved sex-specific alternative splicing regulation and the
291 presence of a conserved TRA/TRA-2 binding site cluster in the sex-specifically regulated *tra* intron strongly

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

292 support the hypothesis of the autoregulation of the *tra* gene in sand flies, as observed for other dipteran and
293 non-dipteran species. At the same time, the absence in sand fly TRAs of a putative TRACAM domain
294 (supposedly involved in the *tra* autoregulation) [18], led us to hypothesize a still unknown molecular
295 mechanism of the TRA autoregulation specific to sand flies.
296
297 **Evolution of *tra-2* genomic organization in Phlebotominae.** *tra-2* is a single-copy gene that has been
298 characterized in *D. melanogaster* [56,57] and in several other dipteran species such as *D. virilis* [58], the
299 house fly *M. domestica* [50], the tephritids *C. capitata* [49,59] and twelve *Anastrepha* species [60], the
300 calliphorid *Lucilia cuprina* [19], and the Nematocera sciarid species *Sciara ocellaris* and *Bradysia*
301 *coprophila* [61]. In these species, *tra-2* is transcribed during development in both sexes, producing an RNA-
302 binding protein with two RS domains flanking an RRM domain. TRA-2 RRM is followed by a 19 aa-long
303 linker region, which is a distinctive and unique feature of the TRA-2 proteins [62]. Within Brachycera
304 suborder, TRA-2 is required for the sex-specific splicing regulation of the *dsx* and *fru* genes and, outside
305 Drosophilidae, it is also involved in the autoregulation of female-specific alternative splicing of the *tra* gene
306 [18,49,50,60,63].
307 Using the available genomic resources of *P. papatasi*, the assembled draft genomes of *P. bergeroti* and *P.*
308 *duboscqi* and the identified putative TRA-2 proteins of *P. perniciosus*, we reconstructed the partial putative
309 exon-intron structure of *tra-2* of the Old World sand flies consisting of 4 exons and 3 introns (Additional file
310 7: Figures S18-20). In addition, we identified in *P. papatasi*, *P. bergeroti* and *P. duboscqi* a putative
311 alternative 5' donor splicing site located downstream of the 5' donor splicing site of the exon 1 of *tra-2*,
312 which is conserved in *P. perniciosus* where it leads to the production of the PpeTRA-2B isoform. As these
313 species belong to different subgenera (*Phlebotomus* and *Larroussius*) this suggests that a similar non-sex-
314 specific alternative splicing event could be conserved also in other Old World sand flies (Additional file 7:
315 Figures S18-20). More in general, among dipteran, *tra-2* shows an overall conservation of exons encoding
316 for functional domains and both RRM and RS1 domains are coded by several exons. In Nematocera Old
317 World sand flies, the RS1 domain is encoded by a unique exon, while the RRM domain and the linker region
318 are organized in two exons (Additional file 8: Fig. S21).
319 As observed for the *tra* ortholog, *tra-2* seems to be absent in transcriptome and genome assembly of the New
320 World sand fly *L. longipalpis*. However, we found well conserved TRA-2 encoding transcripts missing the
321 N-terminus coding region in the *L. umbratilis* and *L. neivai*. This finding suggests that the *tra-2* ortholog
322 could be present also in *Lutzomyia* but not correctly assembled in the *L. longipalpis* released
323 transcriptome/genome assemblies. In figure S22 the multiple alignment of sand fly putative TRA-2 protein is
324 reported. A very well conserved RRM+linker region and RS1 region are present in all the species analyzed.
325 A RS2 region was detected only in *P. perniciosus*, *L. umbratilis* and *L. neivai*. The high percentage of
326 conserved residues of the RS2 region (22/51) suggests its conservation also in other *Phlebotomus* species
327 (Additional file 8: Fig. S22).

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

328 In summary, with our work we identified for the first time the *tra-2* gene in sand flies. Previously, the *tra-2*
329 ortholog of Nematocera was characterized only in the sciarid species *S. ocellaris* and *B. coprophila* and in
330 the mosquito *An. gambiae* and *Ae. aegypti*, where two and four orthologs were found, respectively. In *S.*
331 *ocellaris* and *B. coprophila* TRA-2 is highly conserved and shows conserved sex-determination function
332 when expressed in *Drosophila* [61]. Conversely, putative TRA-2 identified in mosquitoes seem to be
333 divergent respect to other dipteran TRA-2 and possibly not involved in the control of sex-specific splicing of
334 *dsx* and *fru* targets [23,64]. Recent functional tests by transgene-mediated RNAi against *Ae. aegypti tra-2*
335 orthologs have shown no female-to-male sex reversion, as obtained in *tra-2* RNAi functional studies in
336 Brachycera species, but a novel female-specific zygotic lethality. This finding supports the hypothesis that
337 *tra-2* does not play a conserved role in *Ae. aegypti* sex determination while it controls a novel female-
338 specific vital functions which need to be clarified [65]. Here we show that, as for *tra-2* of sciarid species, in
339 sand flies *tra-2* encodes for a protein conserved in its structure and domains, suggesting a conserved role in
340 the sex determination through sex-specific alternative splicing regulation of both *dsx* and *fru* downstream
341 target genes. In addition, we propose that *tra-2* could be involved in the autoregulation of the *tra* gene also in
342 Old World sand flies. The absence of a *tra* ortholog in New World sand flies poses a very interesting
343 problem about the function of *tra-2* in these species and about the evolution of the alternative splicing
344 regulation of *dsx* and *fru* genes.
345

346 **Evolution of *dsx* and *fru* genomic organization and alternative splicing regulation in Phlebotominae.**

347 To study the evolution of the genomic organization and of the alternative splicing regulation of *dsx* and *fru*
348 genes in the sand flies, we aligned DSXs and FRUs of *P. perniciosus*, *P. papatasi* and *L. longipalpis* against
349 the genome sequences of *P. papatasi* and *L. longipalpis* using TBLASTN (Additional file 9: Figures S23-
350 S26, Additional file 15: Supplementary Methods). By manually refining the exon-intron junctions, we
351 obtained the structure of the genes. Compared with the orthologs in *D. melanogaster*, *An. gambiae* and *Ae.*
352 *aegypti,* we observed an overall conservation of the exon/intron organization and of the alternative splicing
353 regulation in sand flies (Additional file 9: Figures S27-S28).
354 In particular, as observed in other dipteran species [23,25,51,52,66,67], in sand flies *dsx* is organized in 4
355 exons spread over a large genomic region (146 Kb in *P. papatasi* and at least 191 Kb in *L. longipalpis*). Exon
356 one, which contains the ATG signal, encodes for the DSX OD1 domain and is followed by the second exon
357 encoding for the non-sex-specific part of the DSX OD2 domain. Exon three is female-specific and encodes
358 for the female-specific DSX C-terminus. Exon four is present in transcripts of both sexes as 3'untranslated
359 region in females and encoding for male-specific DSX C-terminus in males (Fig. S27). Interestingly, the
360 nucleotide sequence of the region surrounding the 3' acceptor female-specific splicing site of the *dsx* gene is
361 strictly conserved among *Phlebotomus* species (Additional file 9: Figures S29). A similar observation was
362 recently reported by Kyrou and colleagues [68] for *Anopheles* mosquito. This region was utilized to develop
363 a gene drive-based population suppression strategy resulted very effective in small scale caged experiments

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

364 [68]. This finding suggests that also the *dsx* gene of *Phlebotomus* species could be an ideal target to develop
365 future similar strategies for sand fly control in field.
366 The *fru* gene in sand flies is organized in eight exons distributed over a very large genomic region (at least
367 125 Kb in *P. papatasi* and 213 Kb in *L. longipalpis*). Exons one and two (named S1 and S2, respectively) are
368 common and female-specific respectively, with exon S1 encoding the male-specific N-terminus of FRU and
369 exon S2 utilized only in females as 5' untranslated region. Exons three and four (named C1 and C2) encode
370 for the BTB domain, while exons five to seven (named C3, C4 and C5) encode the poorly conserved
371 Connector region. The terminal exon eight encodes for a zinc-finger domain of type C (Additional file 9:
372 Figure S28). Using the *P. perniciosus fru* ZnF-A and the *D. melanogaster* protein sequence of ZnF-B as
373 queries, TBLASTN analysis of the genomic scaffold 549 of *P. papatasi* PpapI1 assembly, containing the *fru*
374 exon eight, revealed the presence of putative exons encoding very well conserved ZnF domains. This finding
375 suggests that also in sand flies the *fru* gene could encodes for multiple FRU isoforms by alternative splicing
376 at the 3'end of the primary transcripts (data not shown).
377 Figure 7A shows a schematic representation of the sex-specifically regulated regions of both *dsx* and *fru*
378 genes in *D. melanogaster*, *An. gambiae*, *Ae. aegypti*, *P. papatasi* and *L. longipalpis*. As for most of the
379 Brachycera species, in *Drosophila dsx* and *fru* sex-specific alternative splicing is achieved through two
380 different mechanisms. For *dsx*, a 3' alternative acceptor splicing site choice coupled with alternative
381 polyadenylation leads to sex-specific transcripts with different 3' ends encoding for sex-specific DSX C-
382 termini (Fig. 7A) [69]. For *fru*, a 5' alternative donor splicing site choice leads to sex-specific transcripts
383 with different 5' ends. In males, a male-specific FRU, with a unique N-terminus is obtained through the
384 usage of an ATG signal present in the *fru* male-specific exon (Fig. 7A) [70]. In females, a stop codon in the
385 female-specific exon produces a transcript with a very short open reading frame, probably not translated
386 (Fig. 7A). For both the genes, the male-specific splicing represents the default mode. In female, the presence
387 of TRA and the consequent formation of the TRA/TRA-2 complex which binds the TRA/TRA-2 binding
388 sites in *dsx* and *fru* female-specific exons, promotes female specific splicing [69–71].
389 In Nematocera, *dsx* and *fru* orthologs have been characterized in very few species including the mosquito *An.*
390 *gambiae* and *Ae. aegypti* [22–25,72–74]. While sex-specific splicing regulation of the *fru* orthologs in both
391 mosquito species is very well conserved respect to *Drosophila* (Fig. 7A) [22,24], for *dsx* a different
392 mechanism was described in each species. In *An. gambiae*, male-specific DSX is obtained by skipping the
393 female-specific *dsx* exon; instead the male-specific exon sequence is used in females as 3' untranslated
394 region due to the absence of an alternative polyadenylation signal [25]. In *Ae. aegypti*, *dsx* presents two
395 female-specific exons, like in Sciaridae [73], that are escaped in males. In females, inclusion of both or only
396 the second female-specific exon results in two isoforms. In both *Ae. aegypti* and *An. gambiae*, due to the
397 absence of an alternative polyadenylation signal in the female-specific *dsx* exons, male-specific exons are
398 used as 3' untranslated region [75].

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

399 In sand flies, *fru* has a very well conserved alternative splicing regulation, identical to *D. melanogaster* and
400 mosquitoes, based on a 3' alternative acceptor splicing site choice mechanism. The *dsx* gene alternative
401 splicing regulation is instead similar to *An. gambiae* regulation, with an exon-skipping of a female-specific
402 cassette exon only in males and with the males-specific exonic sequence, present also in female-specific
403 transcripts, used as untranslated region (Fig. 7A).
404 The analysis of *dsx* and *fru* female-specific exons in *P. perniciosus*, *P. papatasi* and *L. longipalpis* revealed
405 the presence of clusters of the *cis*-acting regulatory element named TRA/TRA-2 binding sites. In particular,
406 we identified nine elements in *P. perniciosus* (six located in the *PpedsxF* and three in the *PpefruF*
407 transcripts), six elements in *P. papatasi* (five located in the *PpadsxF* and one in the *PpafruF* transcripts) and
408 eleven elements in *L. longipalpis* (five located in the *LlodsxF* and six in the *LloefruF* transcripts) (Additional
409 file 12: Table S2). The identified TRA/TRA-2 binding sites are organized in clusters of at least three
410 elements except for the single element identified in the *PpafruF* female-specific exon (Additional file 9:
411 Figure S25A).
412 As for *P. perniciosus* (subgenus *Lariossus*), in both *P. bergeroti* and *P. duboscqi* the *fru* S1 exon, encoding
413 for the putative male-specific FRUM N-terminus, is followed by a putative female-specific S2 exon
414 containing three conserved TRA/TRA-2 binding sites (Additional file 11: Figures S30-S33). Similarly, the
415 *dsx* female-specific exon in *P. bergeroti* and *P. duboscqi* shows six clustered TRA/TRA-2 binding sites, as
416 observed in the other sand fly species (Additional file 11: Figures S32-S33). The absence of a cluster of
417 TRA/TRA-2 binding sites in *P. papatasi fru* could be also due to an incorrect assembly of the corresponding
418 *fru* genomic region.
419 Intra-species alignment of the TRA/TRA-2 binding sites in sand flies revealed high sequence conservation.
420 In figure 6B, the WebLogo (http://weblogo.berkeley.edu/) consensus sequences for TRA/TRA-2 binding
421 sites of various dipteran species are reported. Differently from other Nematocera species, such as the
422 mosquitoes *An. gambiae* and *Ae. aegypti* and the sciarid fly *S. ocellaris*, within each 13-bp long TRA/TRA-2
423 binding sites of sand flies we observed an invariable "core" of 8 bp (CAATCAAC) and a low variability, as
424 observed in *Drosophila*, in the first four bases and in the terminal base of the element. In a previous work we
425 proposed that in mosquitoes, the degeneration of the putative TRA/TRA-2 binding sites is related with the
426 absence of the *tra* ortholog and with the low level of TRA-2 conservation, suggesting that different upstream
427 regulators are involved in the control of *dsx* and *fru* genes in this Nematocera species [22]. Conversely, the
428 high conservation of the TRA/TRA-2 binding sites in *Phlebotomus*, which resembles the sequence
429 conservation level of the TRA/TRA-2 binding sites observed in *dsx* and *fru* genes of Brachycera, indicates
430 that these elements, located in untranslated regions of both genes, are under strong selective pressure.
431 Overall, our findings suggest that also in sand flies TRA and TRA-2 are involved in the regulation of the
432 sex-specific alternative splicing of *dsx* and *fru* genes, as observed in Brachycera.
433

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

434 **Phylogeny and selection at sex determination genes in sand flies.** Figure 8 shows the Neighbor-Joining
435 trees obtained from amino acid alignments of selected domains of the TRA (Fig. 8A), TRA2 (Fig. 8B), DSX
436 (Fig. 8C) and FRU (Fig. 8D) of *P. perniciosus* and other species (see Methods and Additional file 13: Table
437 S3). For all proteins, phylogenies segregates sequences in general agreement with the species phylogeny.
438 We investigated natural selection at molecular level as the ratio between the mean nonsynonymous and
439 synonymous substitution rates (ω) of the examined coding regions. To check if the ω ratios differed
440 significantly among the tree branches, we compared one-, two- and three-ratio models [76] for each gene.
441 The one-ratio model assumes an equal ω for all the branches, whereas the two- and three-ratio models
442 consider two and three different ω values, respectively. In addition, we tested the branch-site model that
443 assumes positive selection at specific sites within specific the tree branches [77,78]. The results obtained,
444 and the statistical significance of each comparison, are shown in supplementary Table S4 (Additional file 14:
445 Table S4). Overall ω is always lower than 1, showing that purifying selection acts on these genes (Additional
446 file 14: Table S4).
447 The evolutionary analysis of the TRACAM and DIPTERA *tra* domains shows that the one-ratio model best
448 fits the data (ω = 0.0864) and the absence of positive selection.
449 Within the *tra-2* RMM and linker domains, the two-ratio model is better supported than the one- and three-
450 ratio models, with the mosquito branch showing the lowest ω value (0.0407) when compared to the other
451 branches (ω = 0.0738). The branch-site model identifies two positively selected sites within the branch that
452 does not include mosquitoes; however, the comparison with its null model is not statistically supported.
453 Within the *dsx* OD1 and OD2 domains, the one-ratio model can be excluded in favor of the two- and three-
454 ratio models. The two-ratio model fits the data better than the three-ratio model, showing more relaxed
455 selective constraints of the Phlebotominae branch (ω = 0.0732) when compared to the other branches of the
456 tree (ω = 0.0367). The branch-site model that assumes positive selection at specific sites within the
457 Phlebotominae branch identifies three sites with ω significantly higher than 1 (Additional file 14: Table S4);
458 however, the comparison with the null model that assumes absence of positive selection is not statistically
459 significant.
460 Finally, within the *fru* male-specific domain, the three-ratio model is supported better than the one- and two-
461 ratio models, showing a relaxation of the selective constraints within the Phlebotominae branch (ω = 0.192)
462 when compared to the mosquito branch (ω = 0.1033) and to the branch including *Drosophila*, *Ceratitis* and
463 *Musca* (ω = 0.0288). Site and branch-site models do not show evidence of positive selection.
464 In conclusion, the analysis of the evolutionary pressure acting on the examined sex-determination genes
465 shows evidence of strong purifying selection. However, different selective constraints act on specific
466 branches of the *dsx* and *fru* and *tra-2* genes, whereas the evolutionary rates of the *tra* genes appear more
467 uniform.
468

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

469 **Conclusions**

470 Our results permit to hypothesize a model for the sex determination cascade of Phlebotominae sand flies as
471 shown in figure 9, which represents the first complete and conserved sex determination cascade observed in
472 Nematocera species. In particular, we identified all the key sex determining genes, that in figure 1 are
473 represented by question marks and, for the first time in a Nematocera species, we identified the homolog of
474 the *transformer* gene. In addition, our data strongly suggest the conservation of the autoregulation of the
475 sand fly *tra* gene as observed in Brachycera and in other insect orders. The availability of the sequence of
476 this *tra* gene will help to identify its homologs in other Nematocera species, many of them representing
477 important vectors of human diseases. Our model needs to be confirmed by functional analyses and verified
478 also in New World sand fly species, where the *tra* gene seems to be absent.
479 A further interesting question to be addressed in future is relative to the molecular nature of the primary
480 signals of sex determination in sand flies, to date completely unknown. To this aim the *P. perniciosus*
481 species could be an optimal starting point considering that it is the only Old World sand fly species with
482 described heteromorphic sexual chromosomes [79]. The identification of male determining factors and sex-
483 specific genomic loci in sand fly species could not only help to complete the understanding of sex
484 determination mechanisms in Nematocera but also to shed light on chromosome evolution in insects [80–83].
485 Finally, our results open the possibility of future biotech applications to control natural populations of sand
486 flies to reduce their impact on public health by using technologies available for other insect pests [84–86]. In
487 particular, the *tra* gene could be utilized to produce sexing strains to be implemented for SIT-based control
488 program [87], still missing for sand flies, while the *dsx* gene could be used to develop gene drive systems for
489 population suppression, as recently proposed for *Anopheles* mosquitoes [68].
490
491 **Methods**
492
493 **Sand flies sex determination genes cloning.** The samples of *P. perniciosus* used in this study were from
494 laboratory colonies held at the PV laboratory (Charles University, Department of Parasitology, Prague –
495 Czech Republic) and at the LG laboratory (Istituto Superiore di Sanità, Rome – Italy). The samples of *P.*
496 *papatasi* and *L. longipalpis* used in this study were from laboratory colonies held at the PV laboratory
497 (Charles University, Department of Parasitology, Prague – Czech Republic). The sand fly colonies were
498 reared under standard conditions as previously described [89]. Total RNA was extracted from pools of virgin
499 males and sugar-fed females (7–10 days old) of adult *P. perniciosus*, *P. papatasi* and *L. longipalpis* using the
500 PureLink® RNA Mini kit (Life Technologies) according to manufacturer's instruction, followed by on-
501 column PureLink® DNase (Ambion) treatment. Total RNA was resuspended in 100 µl of ddH_2O and
502 quantified using the NanoDrop 2000c spectrophotometer. The protein coding sequences of insect sex
503 determining genes were used as query to perform TBLASTN search against the PERNI data set (Additional
504 file 1: Table S1) [39]. The transcripts corresponding to the putative *P. perniciosus* orthologues (Additional

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

505 file 1: Table S1) were utilized to design PCR primer pairs (see Additional file 15: supplementary methods).
506 First-strand cDNA was synthesized from 200 ng of male and female total RNA using the EuroScript Reverse
507 Transcriptase kit (Euroclone) with oligo-dT, in a final volume of 20 µl. To amplify the orthologue of the
508 *fruitless* gene, cDNA was synthesized with the gene-specific primers. PCR amplifications were performed on
509 1 µl of 1:20 dilution of the cDNA template from adult males and females, in a final volume of 50 µl, using
510 the Dreamtaq DNA polymerase (Thermo Fisher Scientific) or the PfuUltra HF DNA polymerase (Agilent
511 Technologies). Appropriate annealing temperatures were adjusted to individual primer pairs (Additional file
512 15: Supplementary Methods). The 3' end of the *Ppetra* cDNAs were determined with the 3´ RACE System
513 for Rapid Amplification of cDNA Ends (Invitrogen); the 5' end of the *Ppetra* cDNA was determined with
514 the 5'/3' RACE kit 2nd generation (Roche). Reverse transcription was performed as recommended by the
515 suppliers. The obtained cDNA fragments were cloned using the Strataclone PCR cloning Kit (Agilent
516 Technologies), and the positive clones were sequenced on an ABI 310 Automated Sequencer (Applied
517 Biosystems). cDNA sequences were deposited at the GenBank database with the following accession
518 numbers: PpedsxM MK286442; PpedsxF MK286443; PpetraM1 MK286444; PpetraM2 MK286445;
519 PpetraF1 MK286446; PpetraF2 MK286447; PpetraF3 MK286448; Ppetra-2A MK286449; Ppetra-2B
520 MK286450; PpefruMA MK286451; PpefruMC MK286452; PpefruFA MK286453; PpefruFC MK286454;
521 PpatraF MK286455.
522
523 **Developmental expression analysis.** Total RNA was extracted from the different developmental stages of
524 *P. perniciosus* (embryos, 1st, 2nd and 4th instar, and pupae) using the High Pure RNA Tissue Kit (Roche)
525 according to manufacturer's instruction, followed by on-column DNase treatment. First-strand cDNA was
526 synthesized from 0.5 µg of total RNA using the First Strand cDNA Synthesis Kit for RT-PCR with both
527 oligo-dT primers and random examers, or with the fruC-nested gene-specific primer. PCR amplifications
528 were performed on 1 µl of 1:20 dilution of the cDNA template in a final volume of 50 µl using the
529 EmeraldAmp PCR Master Mix (Clontech). Appropriate annealing temperatures and cycle conditions were
530 adjusted to individual primer pairs (see supplemental methods).
531
532 ***Ppetra* genomic organization.** To identify the intronic region sequence of the *P. perniciosus transformer*
533 gene, genomic DNA was extracted from a single adult female using the NucleoSpin Tissue XS (Macherey-
534 Nagel) according to manufacturer. PCR amplification was conducted on 10 ng of genomic DNA in a final
535 volume of 50 µl using the primers Ppetra5utr/Ppetrastop3utr and the following thermal cycle: 95 °C for 3
536 min, 35 cycles of 94 °C for 30 sec, 56 °C for 30 sec, 72 °C for 2.30 min, final extension of 10 min at 72 °C.
537 The amplification product was cloned and sequenced as described above. The *Ppetra* genomic locus
538 sequence was deposited at the GenBank database with the following accession number: MK286466.
539

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

540 **Phylogeny and evolutionary analysis.** Nucleotide and encoded amino acid sequence of homologs of the
541 *Ppedsx*, *Ppefru*, *Ppetra* and *Ppetra-2* genes were downloaded from GenBank and the relative accession
542 numbers are listed in Table S3. Amino acid sequences were aligned using MUSCLE [90]. Due to high
543 sequence divergence, for each gene the alignments were restricted to the encoded protein regions whose
544 alignment is not ambiguous, as follow: TRA (TRACAM domain and DIPTERA domain), TRA-2 (RRM
545 domain and linker region), DSXF (OD1 and OD2 domains), FRUM (Male-specific N-terminal region and
546 BTB domain). Based on their amino acid alignments, nucleotide sequences were aligned using PAL2NAL
547 [91]. Neighbor-Joining trees were constructed on the amino acid alignments using MEGA7 [92], with 1,000
548 bootstrap replicates. The coding sequences of the *dsx*, *fru*, *tra* and *tra-2* homolog genes were analyzed with
549 the CODEML program from PAML v.4.8 [93] to evaluate their evolutionary rates. Different evolutionary
550 models were compared (branch, sites and branch-sites) to test for variation of the ratio between non-
551 synonymous and synonymous substitution rate (ω) at specific codons in the sequences and among the
552 branches of the trees. For each comparison, a likelihood ratio test was applied to establish which model best
553 fits the data.

554

555 **List of abbreviations**

556 DM (Doublesex Mab3); BTB (Broad-Complex, Tramtrack and Bric a brac); RRM (RNA Recognition
557 Motif); TRACAM (TRA *Ceratitis-Apis-Musca*); ORF (Open Reading Frame); RACE (Rapid Amplification
558 of cDNA END).

559

560 **Declarations**

561

562 **Ethics approval and consent to participate**

563 Not applicable.

564

565 **Consent for publication**

566 Not applicable.

567

568 **Availability of data and material**

569 All the sequencing data produced in this work are deposited in the GenBank public database or present in the
570 supplementary methods. *P. perniciosus* transcriptome assembly utilized in this work is freely available at
571 http://pernibase.evosexdevo.eu while the corresponding RNA-seq raw data are available at the SRA NCBI
572 database under the accession number PRJNA287743. Genome or transcriptome assemblies produced in this
573 study are available upon request or reproducible using instructions present in the supplementary methods.

574

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported the grant STAR2013_25 to MS from University of Naples Federico II and Compagnia di San Paolo, Naples, ITALY, in the frame of Programme STAR2013 (Sostegno Territoriale alle Attività di Ricerca).

Authors' contributions

MS conceived the study. MS and VP planned the experiments. VP performed all the molecular analyses. NP helped with DNA and RNA extractions and RT-PCR analyses. MS performed all bioinformatic analysis with additional contribution of SA, VC and RS. GS suggested the search for *tra* ortholog by using TRA/TRA-2 binding site sequences. MS performed the manual curation of sex determination genes and comparative genomics analyses. SA performed the phylogeny and evolutionary analyses. PV contributed with reagents and biological samples. GB and LG maintained the *P. perniciosus* colony and collected samples. MS, SA and VC wrote the manuscript with inputs by GS, RS, PV and LG. All authors read and approved the final manuscript.

Acknowledgements

The authors are deeply grateful to Riccardo Bianchi and Marta Marchili, Istituto Superiore di Sanità, Roma, Italy, for technical support in rearing sand flies. We deeply acknowledge the Sand Fly Genome Sequencing Consortium and the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) for releasing their unpublished data prior to project completion.

References

1. Haag ES, Doty A V. Sex determination across evolution: Connecting the dots. PLoS Biol. 2005.

2. Zarkower D. Establishing sexual dimorphism: Conservation amidst diversity? Nat. Rev. Genet. 2001. p. 175–85.

3. Marshall Graves JA. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu Rev Genet [Internet]. 2008;42:565–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18983263

4. Verhulst EC, Van De Zande L. Insect sex determination: A cascade of mechanisms. Sex Dev. 2014;8:5–6.

5. Sánchez L. Sex-determining mechanisms in insects. Int J Dev Biol [Internet]. 2008;52:837–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18956315

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

611 6. Gempe T, Beye M. Function and evolution of sex determination mechanisms, genes and pathways in
612 insects. BioEssays. 2011. p. 52–60.
613 7. Schu□tt C, Nöthiger R. Structure, function and evolution of sex-determining systems in Dipteran insects.
614 Development. 2000;127:667–77.
615 8. Erickson JW, Quintero JJ. Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio,
616 signals sex in Drosophila. PLoS Biol. 2007;5:2821–30.
617 9. Saccone G, Pane A, Polito LC. Sex determination in flies, fruitflies and butterflies. Genetica.
618 2002;116:15–23.
619 10. Bopp D, Saccone G, Beye M. Sex determination in insects: Variations on a common theme. Sex Dev.
620 2014;8:20–8.
621 11. Pane A, Salvemini M, Delli Bovi P, Polito C, Saccone G. The transformer gene in Ceratitis capitata
622 provides a genetic basis for selecting and remembering the sexual fate. Development. 2002;129:3715–25.
623 12. Salvemini M, Robertson M, Aronson B, Atkinson P, Polito LC, Saccone G. Ceratitis capitata
624 transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for
625 female sex determination. Int J Dev Biol [Internet]. 2009 [cited 2015 Nov 2];53:109–20. Available from:
626 http://www.ncbi.nlm.nih.gov/pubmed/19123132
627 13. Verhulst EC, van de Zande L, Beukeboom LW. Insect sex determination: It all evolves around
628 transformer. Curr. Opin. Genet. Dev. 2010. p. 376–83.
629 14. Katsuma S, Kiuchi T, Kawamoto M, Fujimoto T, Sahara K. Unique sex determination system in the
630 silkworm, Bombyx mori: current status and beyond. Proc Jpn Acad, Ser B [Internet]. 2018;94:205. Available
631 from: https://www.jstage.jst.go.jp/article/pjab/94/5/94_PJA9405B-01/_pdf/-char/ja
632 15. Saccone G, Salvemini M, Polito LC. The transformer gene of Ceratitis capitata: A paradigm for a
633 conserved epigenetic master regulator of sex determination in insects. Genetica. 2011;139:99–111.
634 16. Ruiz MF, Milano A, Salvemini M, Eirín-López JM, Perondini ALP, Selivon D, et al. The gene
635 transformer of anastrepha fruit flies (Diptera, tephritidae) and its evolution in insects. PLoS One [Internet].
636 2007 [cited 2015 Nov 2];2:e1239. Available from:
637 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2080774&tool=pmcentrez&rendertype=abstract
638 17. Li F, Vensko SP, Belikoff EJ, Scott MJ. Conservation and Sex-Specific Splicing of the transformer Gene
639 in the Calliphorids Cochliomyia hominivorax, Cochliomyia macellaria and Lucilia sericata. PLoS One.
640 2013;8:1–14.
641 18. Hediger M, Henggeler C, Meier N, Perez R, Saccone G, Bopp D. Molecular characterization of the key
642 switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the
643 housefly. Genetics. 2010;184:155–70.
644 19. Concha C, Scott MJ. Sexual development in Lucilia cuprina (Diptera, Calliphoridae) is controlled by the
645 transformer gene. Genetics. 2009;182:785–98.
646 20. Luo Y, Zhao S, Li J, Li P, Yan R. Isolation and Molecular Characterization of the Transformer Gene

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

647 From Bactrocera cucurbitae (Diptera: Tephritidae). J Insect Sci. 2017;17.
648 21. Laohakieat K, Aketarawong N, Isasawin S, Thitamadee S, Thanaphum S. The study of the transformer
649 gene from Bactrocera dorsalis and B. correcta with putative core promoter regions. BMC Genet. 2016;17.
650 22. Salvemini M, D'Amato R, Petrella V, Aceto S, Nimmo D, Neira M, et al. The Orthologue of the Fruitfly
651 Sex Behaviour Gene Fruitless in the Mosquito Aedes aegypti: Evolution of Genomic Organisation and
652 Alternative Splicing. PLoS One. 2013;8.
653 23. Salvemini M, Mauro U, Lombardo F, Milano A, Zazzaro V, Arcà B, et al. Genomic organization and
654 splicing evolution of the doublesex gene, a Drosophila regulator of sexual differentiation, in the dengue and
655 yellow fever mosquito Aedes aegypti. BMC Evol Biol. 2011;11:41.
656 24. Gailey DA, Billeter JC, Liu JH, Bauzon F, Allendorfer JB, Goodwin SF. Functional conservation of the
657 fruitless male sex-determination gene across 250 Myr of insect evolution. Mol Biol Evol. 2006;23:633–43.
658 25. Scali C. Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol
659 [Internet]. 2005;208:3701–9. Available from: http://jeb.biologists.org/cgi/doi/10.1242/jeb.01819
660 26. Criscione F, Qi Y, Tu Z. GUY1 confers complete female lethality and is a strong candidate for a male-
661 determining factor in anopheles Stephensi. Elife. 2016;5.
662 27. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria mosquito
663 Anopheles gambiae. Science (80-). 2016;353:67–9.
664 28. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. A male-determining factor in the
665 mosquito Aedes aegypti. Science (80-). 2015;348:1268–70.
666 29. Biedler JK, Tu Z. Sex Determination in Mosquitoes. Adv In Insect Phys. 2016. p. 37–66.
667 30. Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L. Phlebotomine sandflies and the
668 spreading of leishmaniases and other diseases of public health concern. Med. Vet. Entomol. 2013. p. 123–47.
669 31. Young DG, Duncan MA. Guide of the identification and geografic distribution of Lutzomyia Sand Flies
670 in Mexico, West Indies, Central and South America (Diptera: Psychodidae). Mem Am Entomol Inst.
671 1994;54:1–881.
672 32. WHO WHO. Control of the Leishmaniases. World Health Organization, Geneva. Tech Rep Ser. 2010.
673 33. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, et al. Leishmaniasis worldwide and global
674 estimates of its incidence. PLoS One. 2012.
675 34. Depaquit J, Grandadam M, Fouque F, Andry PE, Peyrefitte C. Arthropod-borne viruses transmitted by
676 Phlebotomine sandflies in Europe: a review. Euro Surveill. 2010. p. 19507.
677 35. Volff JN, Zarkower D, Bardwell VJ, Schartl M. Evolutionary Dynamics of the DM Domain Gene Family
678 in Metazoans. J Mol Evol. 2003.
679 36. Perez-Torrado R, Yamada D, Defossez PA. Born to bind: The BTB protein-protein interaction domain.
680 BioEssays. 2006. p. 1194–202.
681 37. Cléry A, Blatter M, Allain FHT. RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol.
682 2008. p. 290–8.

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

683 38. Geuverink E, Beukeboom LW. Phylogenetic distribution and evolutionary dynamics of the sex
684 determination genes doublesex and transformer in insects. Sex Dev. 2014;8:38–49.
685 39. Petrella V, Aceto S, Musacchia F, Colonna V, Robinson M, Benes V, et al. De novo assembly and sex-
686 specific transcriptome profiling in the sand fly Phlebotomus perniciosus (Diptera, Phlebotominae), a major
687 Old World vector of Leishmania infantum. BMC Genomics [Internet]. 2015 [cited 2015 Nov 2];16:847.
688 Available from:
689 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4619268&tool=pmcentrez&rendertype=abstract
690 40. McAllister BF, McVean GAT. Neutral evolution of the sex-determining gene transformer in Drosophila.
691 Genetics. 2000;154:1711–20.
692 41. Kulathinal RJ, Skwarek L, Morton RA, Singh RS. Rapid evolution of the sex-determining gene,
693 transformer: Structural diversity and rate heterogeneity among sibling species of drosophila. Mol Biol Evol.
694 2003;20:441–52.
695 42. Serna E, Gorab E, Ruiz MF, Goday C, Eirín-López JM, Sánchez L. The gene Sex-lethal of the sciaridae
696 family (order diptera, suborder nematocera) and its phylogeny in dipteran insects. Genetics. 2004;168:907–
697 21.
698 43. Traut W, Niimi T, Ikeo K, Sahara K. Phylogeny of the sex-determining gene <i>Sex-
699 lethal</i> in insects. Genome [Internet]. 2006;49:254–62. Available from: http://article.pubs.nrc-
700 cnrc.gc.ca/ppv/RPViewDoc?issn=1480-3321&volume=49&issue=3&startPage=254&ab=y
701 44. Meise M, Hilfiker-Kleiner D, Dübendorfer a, Brunner C, Nöthiger R, Bopp D. Sex-lethal, the master
702 sex-determining gene in Drosophila, is not sex-specifically regulated in Musca domestica. Development.
703 1998;125:1487–94.
704 45. Saccone G, Peluso I, Artiaco D, Giordano E, Bopp D, Polito LC. The Ceratitis capitata homologue of the
705 Drosophila sex-determining gene sex-lethal is structurally conserved, but not sex-specifically regulated.
706 Development. 1998;125:1495–500.
707 46. Graham P, Penn JKM, Schedl P. Masters change, slaves remain. BioEssays. 2003. p. 1–4.
708 47. Xie W, Guo L, Jiao X, Yang N, Yang X, Wu Q, et al. Transcriptomic dissection of sexual differences in
709 Bemisia tabaci, an invasive agricultural pest worldwide. Sci Rep [Internet]. 2014;4:4088. Available from:
710 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3924218&tool=pmcentrez&rendertype=abstract
711 48. Pane A, Salvemini M, Delli Bovi P, Polito C, Saccone G. The transformer gene in Ceratitis capitata
712 provides a genetic basis for selecting and remembering the sexual fate. Development. 2002;129.
713 49. Salvemini M, Robertson M, Aronson B, Atkinson P, Polito LC, Saccone G. Ceratitis capitata
714 transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for
715 female sex determination. Int J Dev Biol. 2009;53:109–20.
716 50. Burghardt G, Hediger M, Siegenthaler C, Moser M, Dübendorfer A, Bopp D. The transformer2 gene in
717 Musca domestica is required for selecting and maintaining the female pathway of development. Dev Genes
718 Evol. 2005;215:165–76.

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

719 51. Burtis KC, Baker BS. Drosophila doublesex gene controls somatic sexual differentiation by producing
720 alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell. 1989;56:997–1010.
721 52. Saccone G, Salvemini M, Pane A, Polito LC. Masculinization of XX Drosophila transgenic flies
722 expressing the Ceratitis capitata DoublesexM isoform. Int J Dev Biol [Internet]. 2008 [cited 2015 Nov
723 2];52:1051–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18956338
724 53. Salikhov K, Sacomoto G, Kucherov G. Using cascading bloom filters to improve the memory usage for
725 de Brujin graphs. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes
726 Bioinformatics). 2013. p. 364–76.
727 54. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome
728 assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
729 55. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, et al. reference generation and
730 analysis with Trinity. Nat Protoc [Internet]. 2014;8:1–43. Available from:
731 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3875132&tool=pmcentrez&rendertype=abstract
732 56. Belote JM, Baker BS. Sex determination in Drosophila melanogaster: analysis of transformer-2, a sex-
733 transforming locus. Proc Natl Acad Sci U S A [Internet]. 1982;79:1568–72. Available from:
734 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=346016&tool=pmcentrez&rendertype=abstract
735 57. Goralski TJ, Edström JE, Baker BS. The sex determination locus transformer-2 of Drosophila encodes a
736 polypeptide with similarity to RNA binding proteins. Cell. 1989;56:1011–8.
737 58. Chandler D, McGuffin ME, Piskur J, Yao J, Baker BS, Mattox W. Evolutionary conservation of
738 regulatory strategies for the sex determination factor transformer-2. Mol Cell Biol [Internet]. 1997;17:2908–
739 19. Available from:
740 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=232143&tool=pmcentrez&rendertype=abstract
741 59. Gomulski LM, Dimopoulos G, Xi Z, Soares MB, Bonaldo MF, Malacrida AR, et al. Gene discovery in
742 an invasive tephritid model pest species, the Mediterranean fruit fly, Ceratitis capitata. BMC Genomics.
743 2008;9:243.
744 60. Sarno F, Ruiz MF, Eirín-López JM, Perondini AL, Selivon D, Sánchez L. The gene transformer-2 of
745 Anastrepha fruit flies (Diptera, Tephritidae) and its evolution in insects. BMC Evol Biol [Internet].
746 2010;10:140. Available from: http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-10-140
747 61. Martín I, Ruiz MF, Sánchez L. The gene transformer-2 of Sciara (Diptera, Nematocera) and its effect on
748 Drosophila sexual development. BMC Dev Biol [Internet]. 2011;11:19. Available from:
749 http://bmcdevbiol.biomedcentral.com/articles/10.1186/1471-213X-11-19
750 62. Dauwalder B, Amaya-Manzanares F, Mattox W. A human homologue of the Drosophila sex
751 determination factor transformer-2 has conserved splicing regulatory functions. Proc Natl Acad Sci
752 [Internet]. 1996;93:9004–9. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.93.17.9004
753 63. Liu G, Wu Q, Li J, Zhang G, Wan F. RNAi-mediated knock-down of transformer and transformer 2 to
754 generate male-only progeny in the oriental fruit fly, Bactrocera dorsalis (Hendel). PLoS One. 2015;10.

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

755 64. Salvemini M, D'Amato R, Petrella V, Aceto S, Nimmo D, Neira M, et al. The Orthologue of the Fruitfly
756 Sex Behaviour Gene Fruitless in the Mosquito Aedes aegypti: Evolution of Genomic Organisation and
757 Alternative Splicing. PLoS One. 2013;8.
758 65. Hoang KP, Teo TM, Ho TX, Le VS. Mechanisms of sex determination and transmission ratio distortion
759 in Aedes aegypti. Parasit Vectors [Internet]. 2016;9:49. Available from:
760 http://www.parasitesandvectors.com/content/9/1/49
761 66. Hediger M, Burghardt G, Siegenthaler C, Buser N, Hilfiker-Kleiner D, Dübendorfer A, et al. Sex
762 determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal
763 regulator *doublesex*. Dev Genes Evol. 2004;214:29–42.
764 67. Kuhn S, Sievert V, Traut W. The sex-determining gene *doublesex* in the fly *Megaselia scalaris*:
765 Conserved structure and sex-specific splicing. Genome [Internet]. 2000;43:1011–20. Available from:
766 http://www.nrc.ca/cgi-bin/cisti/journals/rp/rp2_abst_e?gen_g00-078_43_ns_nf_gen43-00
767 68. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, et al. A CRISPR-Cas9 gene drive
768 targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. Nat
769 Biotechnol. 2018;36:1062–6.
770 69. Ryner LC, Baker BS. Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation.
771 Genes Dev. 1991;5:2071–85.
772 70. Baker BS, Heinrichs V, Ryner LC. Regulation of Sex-Specific Selection of fruitless 5' Splice Sites by
773 transformer and transformer-2. Mol Cell Biol. 1998;18:450–8.
774 71. Hoshijima K, Inoue K, Higuchi I, Sakamoto H, Shimura Y. Control of doublesex alternative splicing by
775 transformer and transformer-2 in Drosophila. Science (80-). 1991;252:833–6.
776 72. Salvemini M, Polito C, Saccone G. Fruitless alternative splicing and sex behaviour in insects: an ancient
777 and unforgettable love story? J Genet [Internet]. 2010 [cited 2015 Nov 2];89:287–99. Available from:
778 http://www.ncbi.nlm.nih.gov/pubmed/20876995
779 73. Ruiz MF, Alvarez M, Eirín-López JM, Sarno F, Kremer L, Barbero JL, et al. An unusual role for
780 doublesex in sex determination in the dipteran Sciara. Genetics. 2015;200:1181–99.
781 74. Price DC, Egizi A, Fonseca DM. Characterization of the doublesex gene within the Culex pipiens
782 complex suggests regulatory plasticity at the base of the mosquito sex determination cascade. BMC Evol
783 Biol. 2015;15.
784 75. Salvemini M, Mauro U, Lombardo F, Milano A, Zazzaro V, Arcà B, et al. Genomic organization and
785 splicing evolution of the doublesex gene, a Drosophila regulator of sexual differentiation, in the dengue and
786 yellow fever mosquito Aedes aegypti. BMC Evol Biol. 2011;11.
787 76. Yang Z, Nielsen R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol
788 Evol. 1998;46:409–18.
789 77. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive
790 selection. Mol Biol Evol. 2005;22:1107–18.

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

791 78. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting
792 positive selection at the molecular level. Mol Biol Evol. 2005;22:2472–9.
793 79. Kreutzer RD, Modi GB, Tesh RB, Young DG. Brain cell karyotypes of six species of New and Old
794 World sand flies (Diptera: Psychodidae). J Med Entomol. 1987;24:609–12.
795 80. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, et al. Origins and functional
796 evolution of Y chromosomes across mammals. Nature [Internet]. 2014;508:488–93. Available from:
797 http://www.ncbi.nlm.nih.gov/pubmed/24759410
798 81. Blackmon H, Ross L, Bachtrog D. Sex Determination, Sex Chromosomes, and Karyotype Evolution in
799 Insects. J Hered [Internet]. 2016;esw047. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27543823
800 82. Kaiser VB, Bachtrog D. Evolution of Sex Chromosomes in Insects. Annu Rev Genet [Internet].
801 2010;44:91–112. Available from: http://www.annualreviews.org/doi/10.1146/annurev-genet-102209-163600
802 83. Vyskot B. Y Chromosome Evolution. Brenner's Encycl Genet Second Ed. 2013. p. 372–5.
803 84. Papathanos P a, Bossin HC, Benedict MQ, Catteruccia F, Malcolm C a, Alphey L, et al. Sex separation
804 strategies: past experience and new approaches. Malar J. 2009;8 Suppl 2:S5.
805 85. Hendrichs J, Robinson A. Sterile Insect Technique. Encycl Insects. 2009. p. 953–7.
806 86. Hendrichs JP. Use of the sterile insect technique against key insect pests. Sustain Dev Int. 2000;75–9.
807 87. Saccone G, Pane A, De Simone A, Salvemini M, Milano A, Annunziata L, et al. New sexing strains for
808 Mediterranean fruit fly Ceratitis capitata: Transforming females into males. Area-Wide Control Insect Pests
809 From Res. to F. Implement. 2007.
810 88. De Murtas ID, Cirio U, Guerrieri G, Enkerlin D. An experiment to control the Mediterranean fruit fly on
811 the Island of Procida by the sterile-insect technique. Sterile-male Tech Control fruit flies STI/PUB/276
812 [Internet]. 1970. p. 59–70. Available from: file:///Y:/3772.pdf
813 89. Volf P, Volfova V. Establishment and maintenance of sand fly olonies. J Vector Ecol. 2011;36.
814 90. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic
815 Acids Res. 2004;32:1792–7.
816 91. Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the
817 corresponding codon alignments. Nucleic Acids Res. 2006;34.
818 92. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for
819 Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.
820 93. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl
821 Biosci [Internet]. 1997;13:555–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9367129
822
823
824 **Figure legends**
825

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

826 **Fig. 1** Sex determination in insect species. Orders, suborders and families of species with identified sex
827 determining genes reported in figure are indicated. In the model system *Drosophila melanogaster,* the
828 presence of two X chromosomes in the female embryo activates the *Sex-lethal* gene (*Sxl*) which, acting as a
829 gene-specific splicing regulator, promotes the female-specific splicing of its own pre-mRNA and of the pre-
830 mRNA of the gene downstream in the regulative cascade, *transformer* (*tra*). *tra* and the non-sex-specific
831 auxiliary factor *transformer*-2 (*tra-2*) encode for splicing factors (TRA and TRA-2 proteins) able to control
832 the splicing of at least two downstream target genes, responsible of sexual differentiation and courtship
833 behaviour: *doublesex* (*dsx*) and *fruitless* (*fru*), respectively. Both genes encode for sex-specific transcription
834 factors that potentially binds to multiple genome loci, leading to sex-specific gene expression and subsequent
835 sexual differentiation. In male embryo, the absence of the functional SXL protein leads to the male-specific
836 splicing of *tra*, *dsx* and *fru* pre-mRNAs resulting in the activation of the male development program. In
837 Hymenoptera, Coleoptera, and Diptera (Brachycera), different primary signals set the activity state of the *tra*
838 homolog able to autoregulate its own splicing in the female sex and to determine female development. In
839 mosquitoes (Diptera, Nematocera) *dsx* and *fru* genes exhibit, as for Brachycera species, a conserved
840 alternative splicing regulation, producing sex-specific protein isoforms. Recently, genomic/transcriptomic
841 studies of sex determination led to the discovery of novel primary signals including the Y-linked genes *Yob*
842 and *Guy-1* in the malaria vectors *Anopheles gambiae* and *An. stephensi*, respectively, and the putative
843 splicing factors *Nix* in the dengue vector *Aedes aegypti*. These primary signals are supposed to act upstream
844 of *dsx* and *fru* genes in the sex determination cascade. However, their mechanism of action, direct or indirect,
845 and the possible presence of an intermediate upstream regulator of *dsx* and *fru* splicing, is still an open
846 question.
847
848 **Fig. 2** Sex determining genes expression at adult stage in *P. perniciosus*. A) Positive RT-PCR control with
849 Ppesod+/Ppesod-primer pairs. These PCR primers span a 112-bp long intron of *Ppesod* gene and permit to
850 exclude genomic DNA contamination of cDNAs. B) *Ppetra-2* RT-PCR amplification. C) *Ppedsx* RT-PCR
851 amplification. D) *Ppefru* RT-PCR amplification. Light green boxes represent untranslated regions. Dark
852 green boxes represent non-sex-specific coding regions. Azure boxes represent male-specific untranslated
853 regions. Pink boxes represent female-specific untranslated regions. Blue and red boxes represent male-
854 specific and female-specific coding regions, respectively. The position of primers utilized for each gene are
855 indicated by short red arrows. Yellow vertical bars indicate the position of the putative TRA/TRA-2 binding
856 sites. E) Putative TRA/TRA-2 binding sites identified in *Ppedsx* and *Ppefru* female-specific transcripts.
857
858 **Fig. 3** *Ppetra* transcripts and expression at adult stage. A) *In-silico* identified *Ppetra* transcript, containing
859 six putative TRA/TRA-2 binding sites, indicated by yellow vertical bars. B) *Ppetra* RT-PCR amplification.
860 Azure boxes represent male-specific untranslated regions. Pink boxes represent female-specific untranslated

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

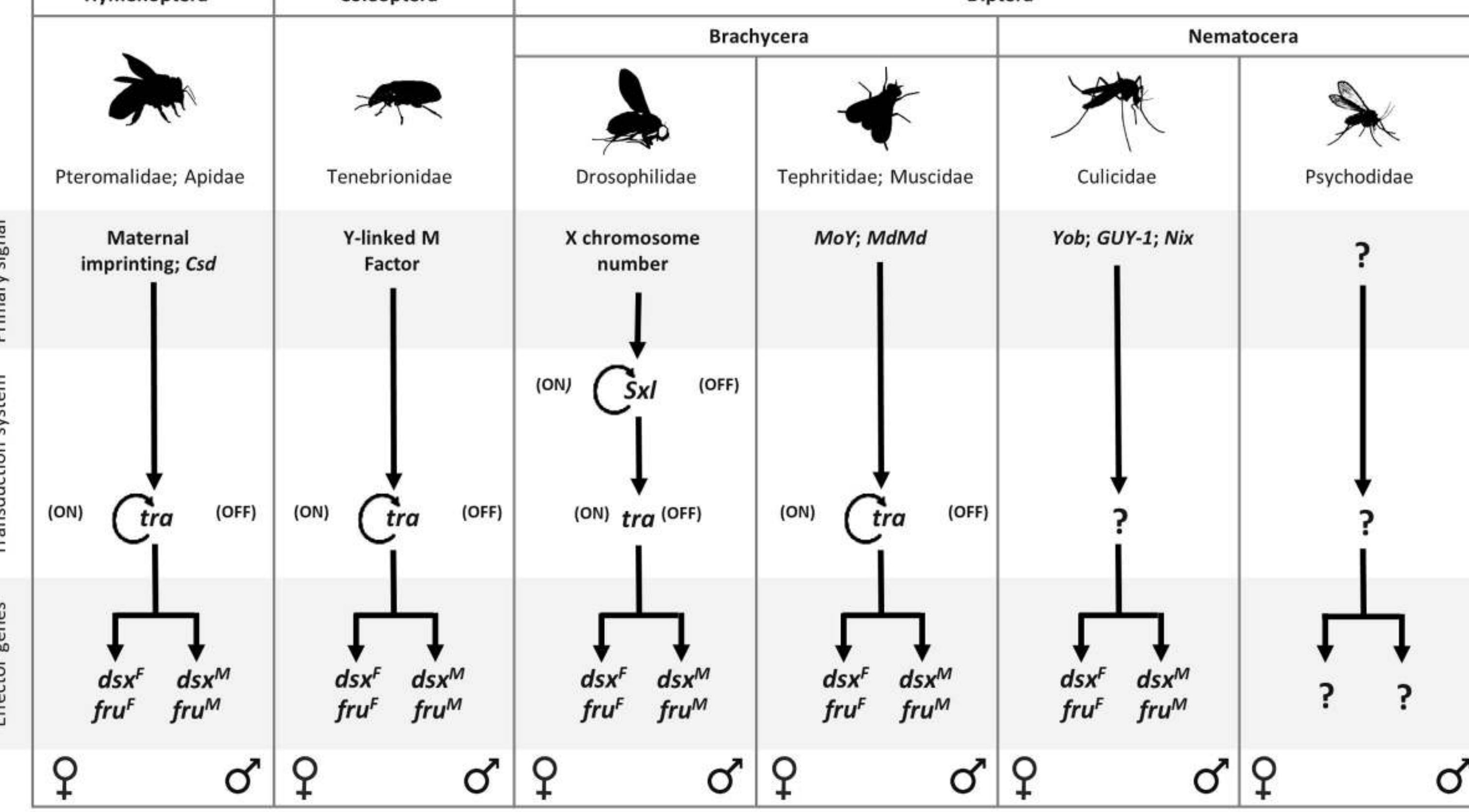
861 regions. Blue and red boxes represent male-specific and female-specific coding regions, respectively. The
862 positions of primers utilized are indicated by short red arrows.
863
864 **Fig. 4** Sand fly life cycle and developmental expression analyses of sex determining genes in *P. perniciosus*.
865 RT-PCR amplifications of *Ppesod* (A), *Ppetra* (B), *Ppetra-2* (C), *Ppedsx* (D) and *Ppefru* (E) were performed
866 on the following samples: E = 0-36h embryos; L1 = first instar larvae; L2 = second instar larvae; L4 = fourth
867 instar larvae; P = pupae; all samples are composed of mixed sexes. The *P. perniciosus sod* gene, utilized as
868 positive control, is constitutively expressed throughout development. The coloured bars indicate the
869 presence/absence of expression at each developmental stage of *Ppetra* (blue), *Ppetra-2* (green), *Ppedsx*
870 (green) and *Ppefru* (yellow) transcripts.
871
872 **Fig. 5** *P. perniciosus tra* genomic organization and evolution. A) *Ppetra* gene locus and sex-specific
873 transcripts. Azure boxes represent male-specific untranslated regions. Pink boxes represent female-specific
874 untranslated regions. Blue and red boxes represent male-specific and female-specific coding regions,
875 respectively. Striped blue-red boxes represent coding regions utilized both in the male and female sex.
876 Yellow vertical bars indicate the position of the putative TRA/TRA-2 binding sites. B) Comparison of *tra*
877 gene structures in *Phlebotomus* species. C) WebLogo consensus sequence of the putative TRA/TRA-2
878 binding sites identified in *Phlebotomus tra* species and of TRA/TRA-2 binding sites of Brachycera *tra* genes.
879
880 **Fig. 6** Phylogenetic relationship and protein sequence comparison of TRA/FEM proteins. A) TRA/FEM
881 protein schemes were aligned using the conserved sex-specific splicing site located within the TRACAM
882 domain encoding region as reference point (indicated by the red triangle). This sex-specific splicing site is
883 conserved in all the autoregulating *tra* genes. In sand flies TRA, the TRACAM domain is absent. Striped
884 yellow-grey box represents the position of the homologous sex-specific splicing site in *P. perniciosus* TRA.
885 *D. melanogaster* TRA protein was aligned using the position of the non-conserved sex-specific splicing site.
886 The crustacean *Daphnia magna* TRA was aligned using the position of the conserved TRACAM domain.
887 Percentages within red and blue boxes indicates the percentage of R and S residues and of P residues in the
888 RS and Proline-rich domains, respectively. To define the boundaries of the RS domain, we considered the
889 position of the first RS or SR couple of residues till the final RS or SR couple of residues and we considered
890 the selected region an RS domain only if its percentage of R and S is > of 25%. B) Multiple alignment of
891 insect TRACAM domains and the *P. perniciosus* TRA homologous region. Amino acids conserved in at
892 least two species are highlighted in black. The conserved sex-specific splicing site is indicated by red
893 triangle.
894
895 **Fig. 7** Evolution of sex-specific alternative splicing regulation of *dsx* and *fru* gene. A) Comparative
896 schematic representation of sex-specifically regulated regions of dsx and fru genes in *D. melanogaster*,

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

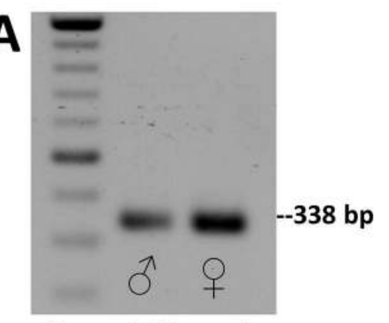
897 mosquito and sand fly species. Green boxes represent non-sex-specific coding regions. Azure boxes
898 represent male-specific untranslated regions. Pink boxes represent female-specific untranslated regions. Blue
899 and red boxes represent male-specific and female-specific coding regions, respectively. Yellow vertical bars
900 indicate the position of the putative TRA/TRA-2 binding sites. White vertical bars indicate the degenerated
901 mosquito putative TRA/TRA-2 binding sites. B) WebLogo consensus sequence of the putative TRA/TRA-2
902 binding sites of Brachycera and Nematocera species. Within sand flies, *L. longipalpis* exhibits the less
903 conserved TRA/TRA-2 binding sites, as expected for a species with upstream regulator/s of the alternative
904 splicing of *dsx* and *fru* genes, different from *tra*.
905
906 **Fig. 8** Neighbor-Joining trees obtained from the amino acid alignment of selected domains of the TRA,
907 TRA-2, FRU and DSX proteins. For the TRA alignment (A) TRACAM and DIPTERA domains were
908 utilized. For the TRA-2 alignment (B) we utilized the RRM domain and the linker region. For the DSX
909 alignment (C) OD1 and OD2 domains of the DSXF isoform were utilized. For FRU alignment (D) we
910 utilized the male-specific N-terminal region and the BTB domain. Brachycera species are highlighted in blue
911 and Nematocera species are highlighted in green. The protein IDs of the species belonging to Phlebotominae
912 are in red.
913
914 **Fig. 9** Model for sex determination in sand flies. In female embryos, a maternal *tra* mRNA or TRA protein
915 and a maternal auxiliary TRA-2 protein led to the activation of a positive feedback autoregulative loop. The
916 early TRA and TRA-2 proteins drive the female-specific splicing of the zygotically transcribed *tra* pre-
917 mRNA so that new TRA protein can be produced. The newly synthesized protein controls the maintenance
918 of *tra* autoregulation and the female-specific splicing of *dsx* and *fru* pre-mRNAs leading to female
919 development. In male embryos, *tra* autoregulation is impaired by a male-specific factor, resulting in absence
920 of the TRA protein, determining the male-specific splicing of the *dsx* and *fru* genes and thus inducing male
921 development.
922
923
924 **Additional files**
925
926 **Additional file 1**: **Table S1**. TBLASTN search of sex determination orthologs in the perniBASE dataset.
927 (XLSX 11.6 kb)
928 **Additional file 2**: **Figure S1**. Multiple sequence alignment of SXL proteins. **Figure S2**. Multiple sequence
929 alignment of TRA-2 proteins. **Figure S3**. Multiple sequence alignment of DSX amino-terminal regions.
930 **Figure S4**. Multiple sequence alignment of DSXF carboxy-terminal regions. **Figure S5**. Multiple sequence
931 alignment of DSXM carboxy-terminal regions. **Figure S6.** Multiple sequence alignment of FRUM amino-
932 terminal regions. **Figure S7**. Multiple sequence alignment of FRU proteins. **Figure S8.** Sex-lethal gene

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

933 expression at adult stage in *P. perniciosus*. **Figure S9**. Multiple sequence alignment of TRA proteins. (PDF
934 896 kb)
935 **Additional file 3**: **Figure S10**. Manually-curated *P. perniciosus transformer* gene model. **Figure S11**.
936 Manually-curated *P. papatasi transformer* gene model. **Figure S12**. Manually-curated *P. bergeroti*
937 *transformer* gene model. **Figure S13**. Manually-curated *P. duboscqi transformer* gene model. (PDF 350 kb)
938 **Additional file 4**: **Figure S14**. Multiple sequence alignment of TRA proteins in *Phlebotomus* spp. (PDF 252
939 kb)
940 **Additional file 5**: **Figure S15**. Multiple alignment of tra introns in *Phlebotomus* spp. (PDF 573 kb)
941 **Additional file 6**: **Figure S16**. *tra* gene expression at adult stage in *P. papatasi*. **Figure S17**. Phylogenetic
942 relationship, genomic structure and sex-specific splicing regulation of *transformer* orthologues in insects.
943 (PDF 877 kb)
944 **Additional file 7**: **Figure S18**. Manually-curated P. *papatasi tra-2* partial gene model. **Figure S19**.
945 Manually-curated *P. bergeroti tra-2* partial gene model. **Figure S20**. Manually-curated *P. duboscqi tra-2*
946 partial gene model. (PDF 142 kb)
947 **Additional file 8**: **Figure S21**. Comparison of genomic structures of dipteran *tra-2* genes. **Figure S22**.
948 Multiple sequence alignment of sand fly TRA-2 proteins. (PDF 188 kb)
949 **Additional file 9**: **Figure S23**. Manually-curated *P. papatasi doublesex* gene model. **Figure S24**. Manually-
950 curated *L. longipalpis doublesex* gene model. **Figure S25**. Manually-curated *P. papatasi fruitless* gene
951 model. **Figure S26**. Manually-curated *L. longipalpis fruitless* partial gene model. **Figure S27**. Comparison
952 of genomic structures of dipteran *dsx* genes. **Figure S28**. Comparison of genomic structures of dipteran *fru*
953 genes. (PDF 284 kb)
954 **Additional file 10**: **Figure S29**. Crispr/Cas9 target sites in sand fly *dsx* genes (PDF 156 kb)
955 **Additional file 11**: **Figure S30**. Manually-curated *P. bergeroti fruitless* partial gene model. **Figure S31**.
956 Manually-curated *P. duboscqi fruitless* partial gene model. **Figure S32**. Manually-curated *P. bergeroti*
957 *doublesex* partial gene model. **Figure S33**. Manually-curated *P. duboscqi doublesex* partial gene model (PDF
958 143 kb)
959 **Additional file 12**: TRA/TRA-2 binding sites of Brachycera and Nematocera species. (XLSX 17.0 kb)
960 **Additional file 13**: Accession numbers and ID of the sequences used in phylogenetic and evolutionary
961 analyses. (XLSX 17.0 kb)
962 **Additional file 14**: Statistics of the evolutionary analyses and comparison of different evolutionary models
963 (XLSX 17.0 kb)
964 **Additional file 15**: Supplementary Methods (PDF 284 kb)

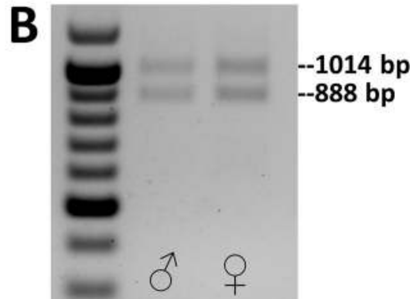

Hymenoptera
Coleoptera
Diptera
Brachycera
Nematocera
Pteromalidae; Apidae
Tenebrionidae
Drosophilidae
Tephritidae; Muscidae
Culicidae
Psychodidae
Primary signal
Maternal imprinting; Csd
Y-linked M Factor
X chromosome number
MoY; MdMd
Yob; GUY-1; Nix
?
Transduction system
(ON) Sxl (OFF)
(ON) tra (OFF)
?
Effector genes
dsxF fruF
dsxM fruM
?
♀
♂

A

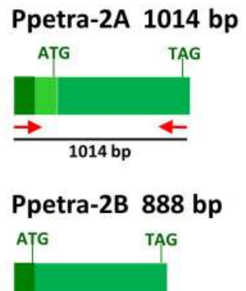


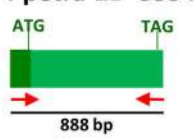
Ppesod+/Ppesod-

B



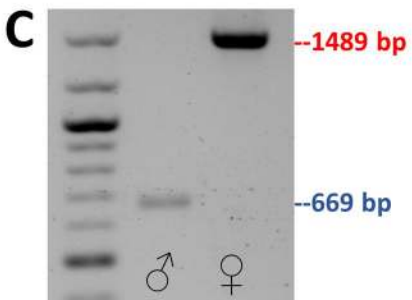
Ppetra-2+/Ppetra-2-



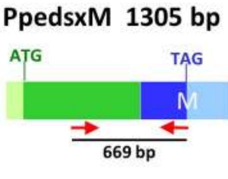


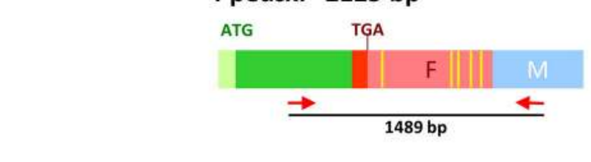
100 bp

C

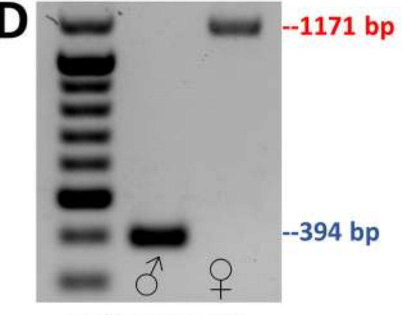


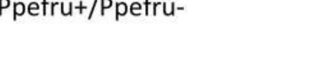
Ppedsx+/Ppedsx-



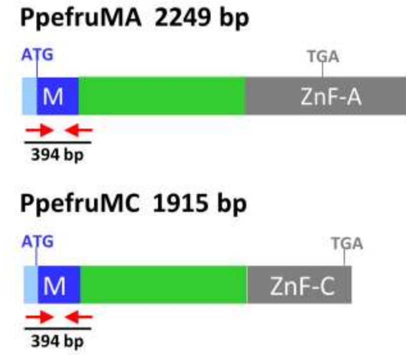


D





Ppefru+/Ppefru-

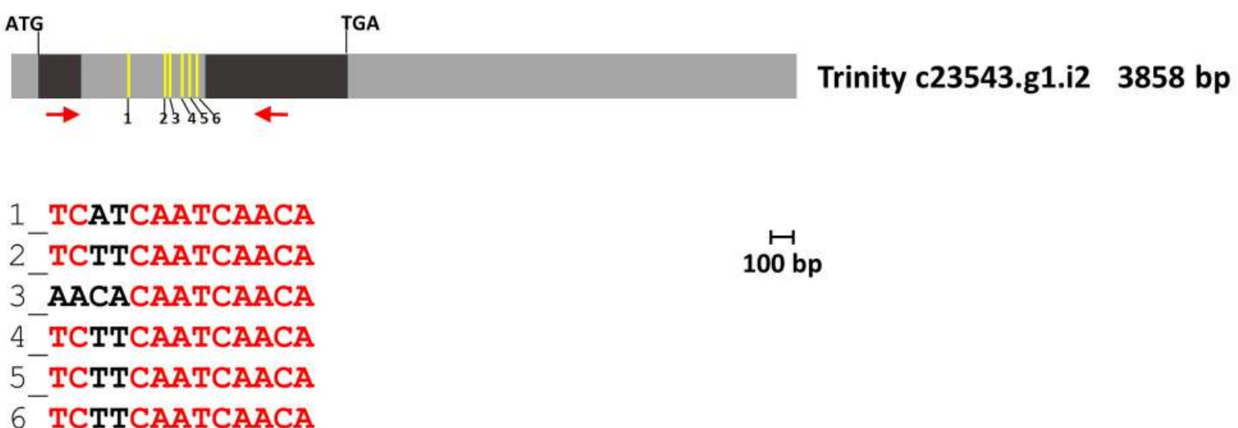


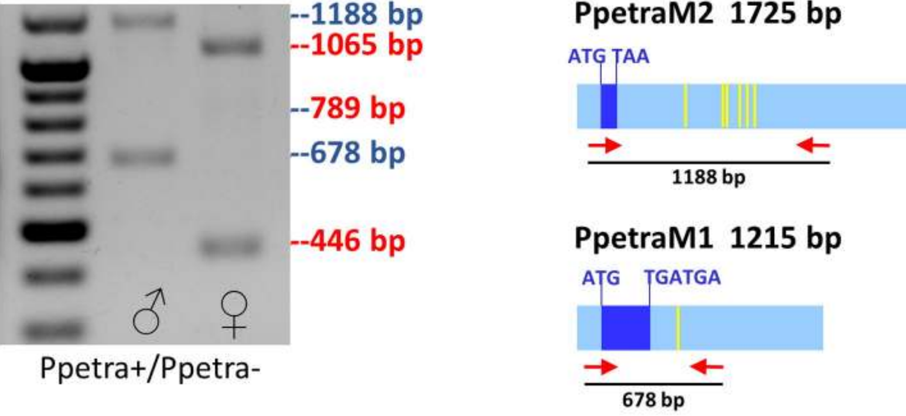
E ***PpedsxF*** **TRA/TRA-2 binding sites**

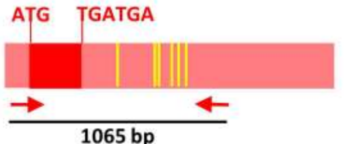
TCATCAATCAACA
TCTTCAATCAACA
TCTACAATCAACA
TCTACAATCAACA
TCTTCAATCAACA
TCTTCAATCAACT

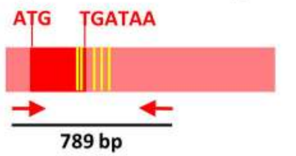
PpefruF **TRA/TRA-2 binding sites**

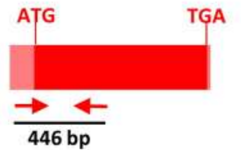
TCTGCAATCAACA
TCTTCAATCAACA
TCTTCAATCAACA











100 bp

E L1 L2 L4 P A

tra tra-2 dsx fru

A

E L1 L2 L4 P

--338 bp

Ppesod+/Ppesod-

B

E L1 L2 L4 P

--1065 bp
--678 bp
--446 bp

Ppetra+/Ppetra-

C

E L1 L2 L4 P

--1014 bp
--888 bp

Ppetra-2+/Ppetra-2-

D

E L1 L2 L4 P

--1489 bp
--669 bp

Ppedsx+/Ppedsx-

E

E L1 L2 L4 P

--1171 bp
--394 bp

Ppefru+/Ppefru-

A

ATG TAA TGATGA TGATAA TGA

1 2 ms1 fs1 3 4 *Ppetra* gene

int1 int2 int3

bioRxiv preprint doi: https://doi.org/10.1101/510586; this version posted January 4, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PpetraF1

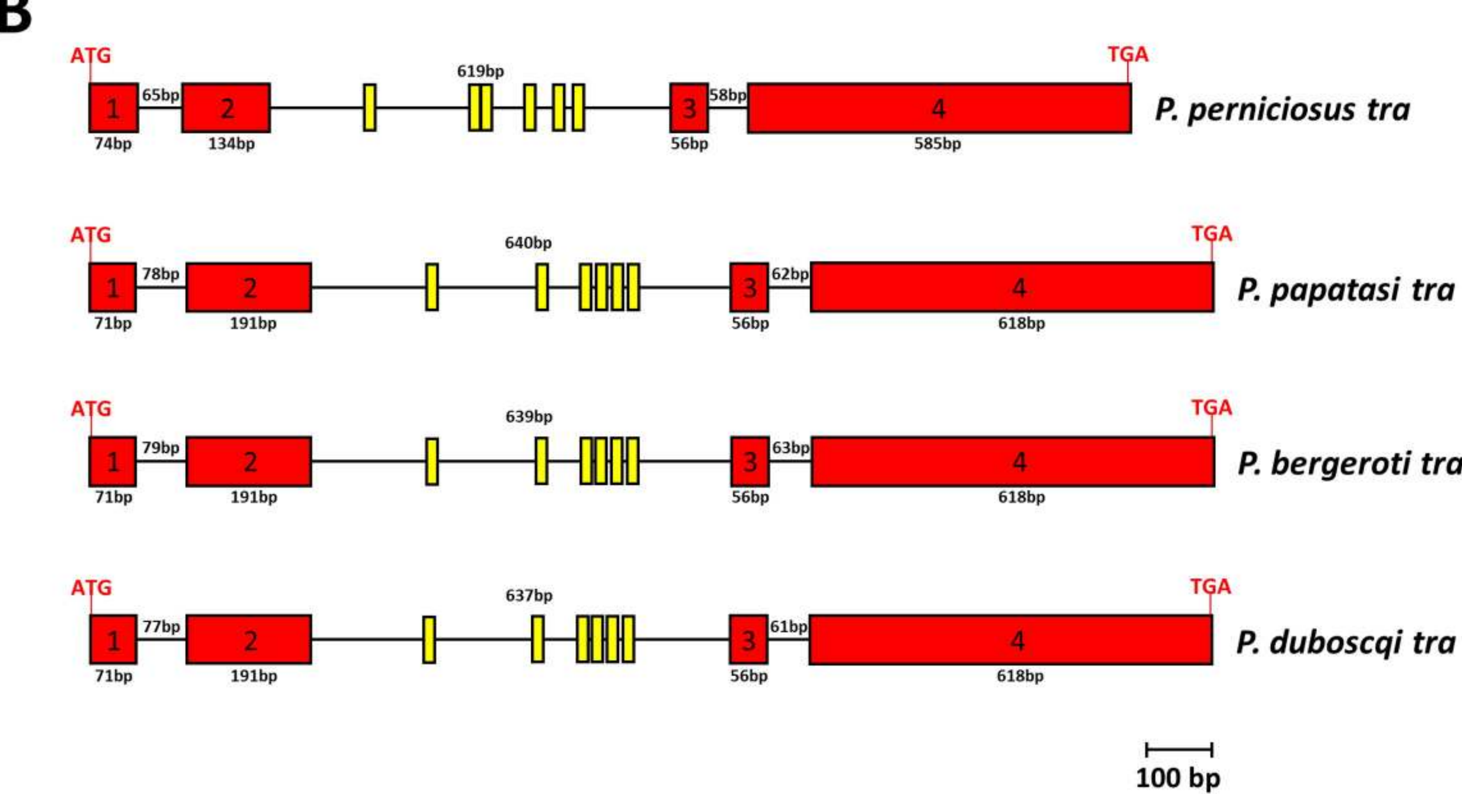
PpetraF2

PpetraF3

PpetraM1

PpetraM2

100 bp



C

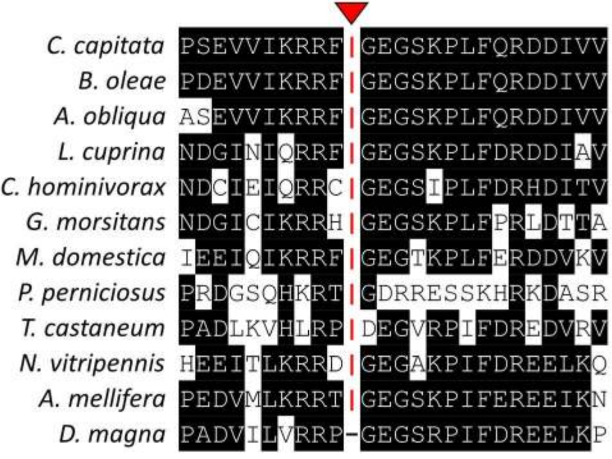
A

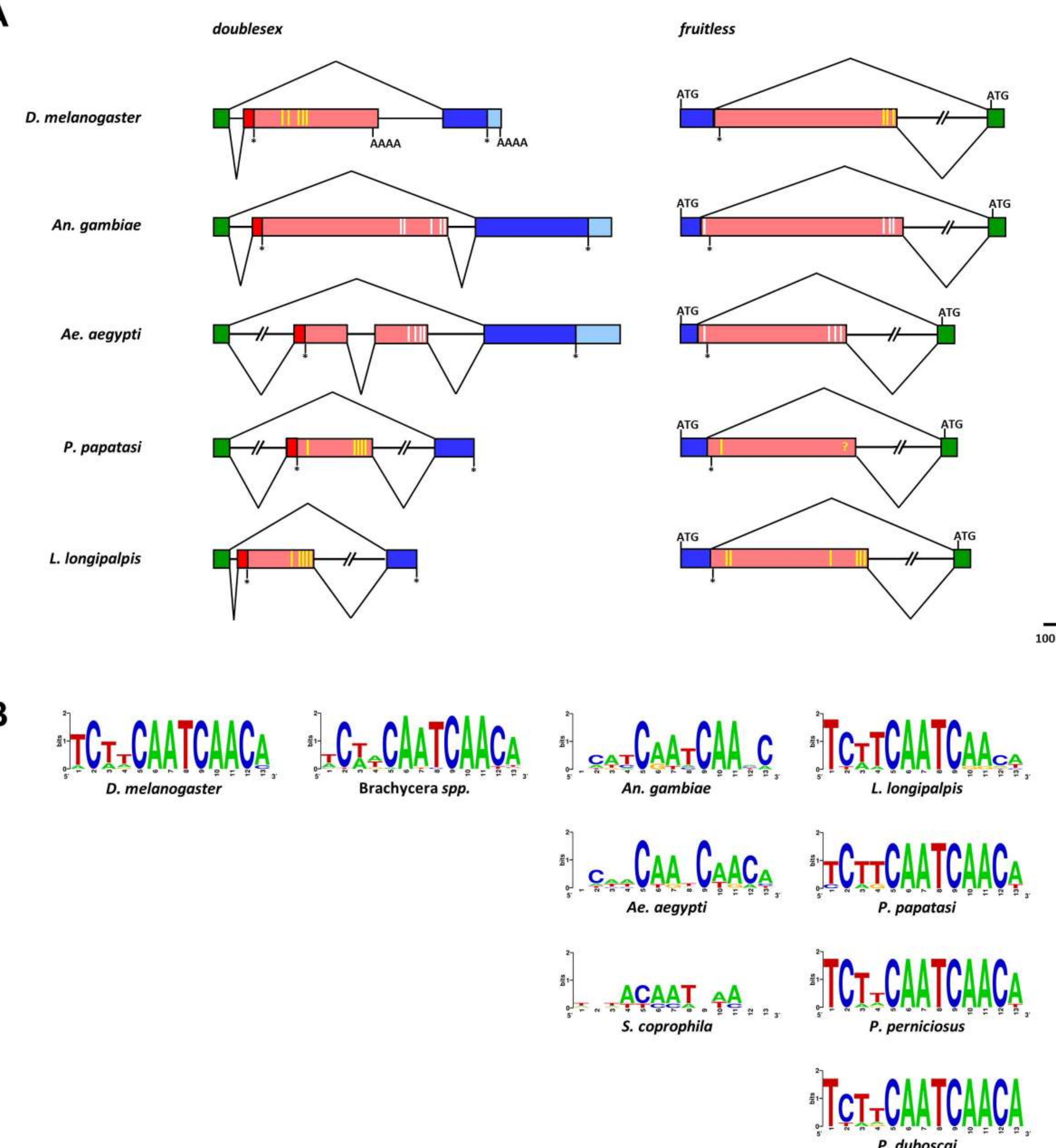
Drosophilidae
Acalyptratae
Tephritidae
Brachycera
Calliphoridae
Glossinidae
Calyptratae
Muscidae
Diptera
Nematocera
Coleoptera
Insecta
Hymenoptera
Crustacea

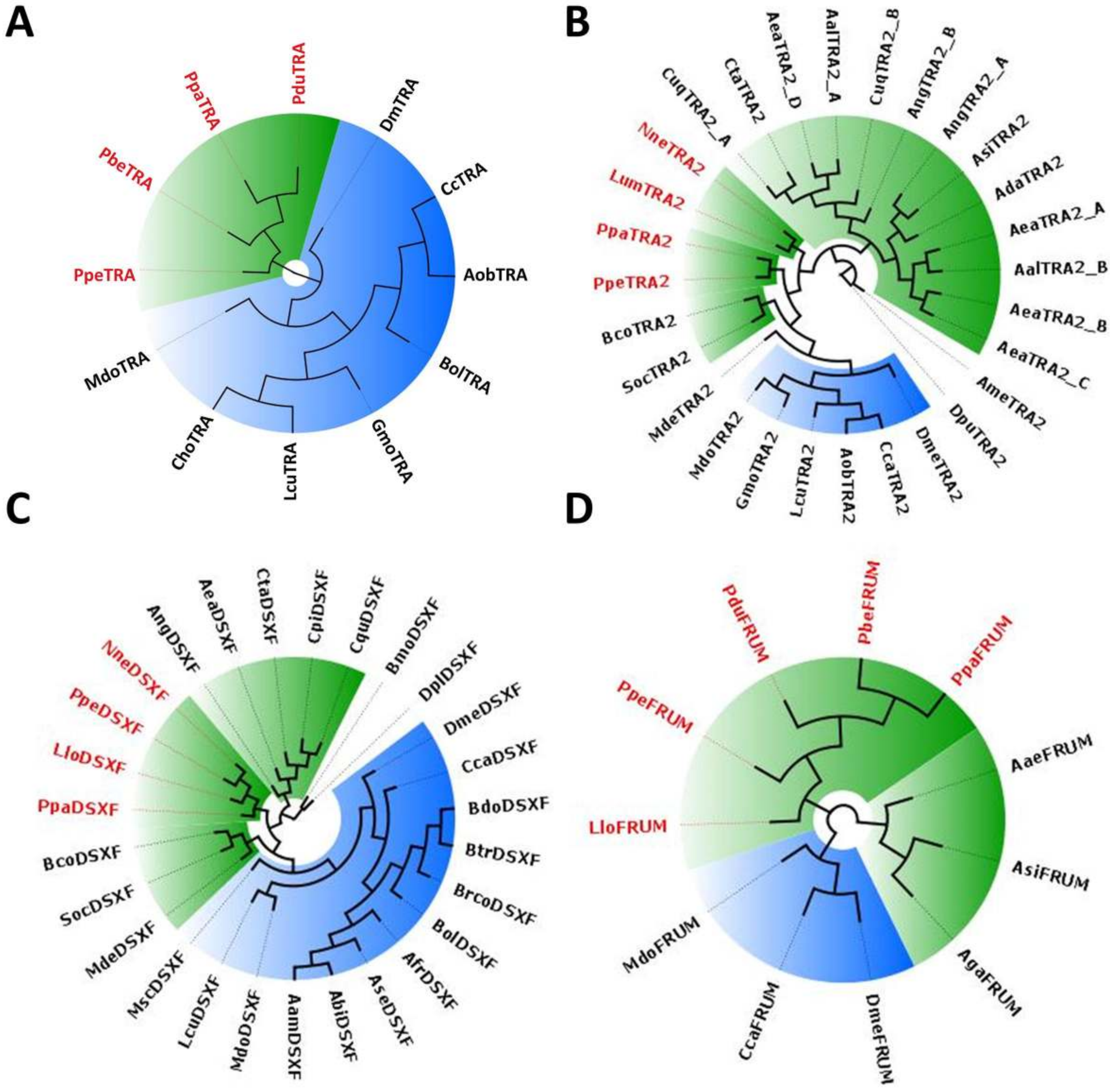
D. melanogaster NH$_2$– 58.5% 25.3% –COOH
C. capitata NH$_2$– 51.3% 26.3% –COOH
B. oleae NH$_2$– 44.4% 29.8% –COOH
A. obliqua NH$_2$– 55.1% 30.1% –COOH
L. cuprina NH$_2$– 46.0% 31.9% –COOH
C. hominivorax NH$_2$– 42.7% 26.4% –COOH
G. morsitans NH$_2$– 42.5% 21.6% –COOH
M. domestica NH$_2$– 58.9% 23.0% –COOH
P. perniciosus NH$_2$– 57.4% 27.7% –COOH
T. castaneum NH$_2$– 57.7% 23.4% –COOH
N. vitripennis NH$_2$– 40.2% 46.4% 26.7% –COOH
A. mellifera NH$_2$– 39.4% 46.3% 29.4% –COOH
D. magna NH$_2$– 32.1% 27.3% –COOH

100 aa

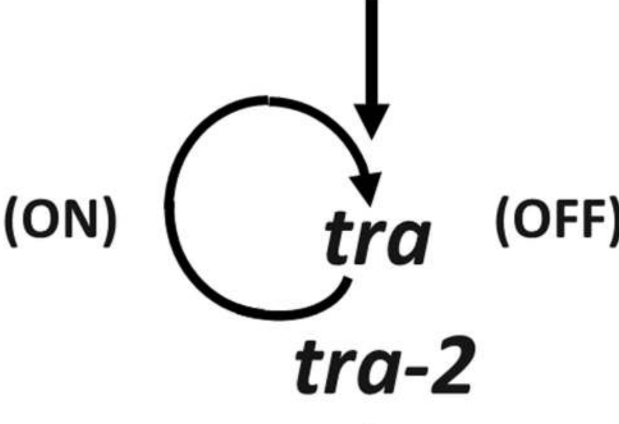
TRACAM domain
DIPTERA domain
HYMENOPTERA domain
RS domain
Proline-rich domain




A
doublesex
fruitless
D. melanogaster
An. gambiae
Ae. aegypti
P. papatasi
L. longipalpis
ATG
AAAA
100 Bp
B
D. melanogaster
Brachycera spp.
An. gambiae
L. longipalpis
Ae. aegypti
P. papatasi
S. coprophila
P. perniciosus
P. duboscqi
P. bergeroti


A
PduTRA
PpaTRA
PbeTRA
PpeTRA
DmTRA
CcTRA
AobTRA
BolTRA
GmoTRA
LcuTRA
ChoTRA
MdoTRA
B
NneTRA2
LumTRA2
PpaTRA2
PpeTRA2
BcoTRA2
SocTRA2
MdeTRA2
MdoTRA2
GmoTRA2
LcuTRA2
AobTRA2
CcaTRA2
DmeTRA2
DpuTRA2
AmeTRA2
AeaTRA2_C
AeaTRA2_B
AalTRA2_B
AeaTRA2_A
AdaTRA2
AsiTRA2
AngTRA2_A
AngTRA2_B
CuqTRA2_B
AalTRA2_A
AeaTRA2_D
CtaTRA2
CuqTRA2_A
C
NneDSXF
PpeDSXF
LloDSXF
PpaDSXF
BcoDSXF
SocDSXF
MdeDSXF
MscDSXF
LcuDSXF
MdoDSXF
AamDSXF
AbiDSXF
AseDSXF
AfrDSXF
BolDSXF
BrcoDSXF
BtrDSXF
BdoDSXF
CcaDSXF
DmeDSXF
DplDSXF
BmoDSXF
CquDSXF
CpiDSXF
CtaDSXF
AeaDSXF
AngDSXF
D
PduFRUM
PbeFRUM
PpaFRUM
PpeFRUM
LloFRUM
AaeFRUM
AsiFRUM
AgaFRUM
DmeFRUM
CcaFRUM
MdoFRUM

Primary Signal?



dsx^F, fru^F ***dsx^M, fru^M***