Supplementary Information

Intron retention in the *Drosophila melanogaster* Rieske iron sulfur protein gene generates a new protein

Alisson M. Gontijo, Veronica Miguela, Michael F. Whiting, R. C. Woodruff, Maria Dominguez

Supplementary Figures S1-S8
Supplementary Tables S1-S6
Supplementary Discussion
Supplementary References
Supplementary Figure S1.

*Drosophila melanogaster*
1. EZ30 Ethiopia Zway
2. P4 Perysburgh OH recomb. allele 1
3. RVC-4 Riverside CA USA #3871
4. Tana12 Madagascar Antarasambolo
5. Kivu Uganda Kisoro
6. Cornmark NY USA #3850
7. EZ51 Ethiopia Zway
8. Kyoc70 Kenya Malindi New Market
9. EV Ellenville NY 3851
10. Hikone-A-S
11. P4 Perysburg OH allele 2
12. Kyoc6 Kenya Malindi New Market
13. RC1 Riverside CA USA #3855
14. Flores9 #3747. Gemza #4508
15. BG02 Bogota Colombia #3842
16. ZS11 Zimbabwe Senega
17. PB Perysburg OH USA
18. Oregon R4
19. Canton-S, Kornba Dam-3 #3853
20. P7 Perysburg OH USA
21. Reids-1 Madeira Portugal #3866

*Drosophila simulans*
Dam1. ZOM Malawi, Zomba
Dam2. Kib11 Uganda Kibale forest
Dam3. #4736
Dam4. Genome Reference (w1118)
Dam5. MZ48 Madagascar
Dam6. Tana10 Madagascar Antarasambolo
Supplementary Figure S1. Alignment of a segment of the RFeSP locus spanning intron2. Only one line per haplotype is shown for D. melanogaster (Dmel). There are 16 haplotypes for intron2a and 4 haplotypes for intron2b. The number (n) of times each haplotype was found is depicted. The 5' splice donor and 3' splice acceptor are shown as black arrows, while the putative branch point is shown with a black arrowhead. Nucleotides: G in magenta, C in dark blue, A in yellow, and T in cyan. Key deletions are depicted by dark bars on the left. Note that a phylogenetically informative polymorphic insertion/deletion (indel) in the D. simulans complex (Dsim compl.) was found on the 3' region of the intron2. This polymorphic indel has apparently become differentially fixed in either D. mauritiana (Dmau) or in D. sechelia (Dsech). Dsim1, Dsim2, and so on correspond to different D. simulans lines [Dsim1= ZOM4 Malawi. Zomba, Dsim2 = Kib11 Uganda Kibale forest, Dsim3 = #4736, Dsim4 = Genome Reference (Strain 14021-0251.195 Dsim[w[501]], Dsim5 = M258 Madagascar, Dsim6 = Tana10 Madagascar Anananarivo]. Dyak1 and Dyak2 correspond to the 14021-0261-00 genome reference and 14021-0261-01 D. yakuba lines, respectively. Other taxa: D. tessieri (Dtes), D. erecta (Dere), D. orena (Dore), D. santomea (Dsan). D. melanogaster lines (from left to right columns, group according to haplotype: intron2a alleles: 1st Column: EZ39 Ethiopya Ziway; 2nd Column: P4 Perrysburg Ohio recombinant allele 1 (P4 strain was heterozygote); 3rd Column: RVC-4 Riverside CA USA #3871; 4th Column: Tana12 Madagascar Antananarivo; 5th Column: Kis4 Uganda Kisoro; 6th Column: Commank NY USA #3850; 7th Column: EZ61 Ethiopia Ziway; 8th Column: Ky070 Kenya Malindi New Market; 9th Column: EV Ellenville NY 3851; 10th Column: Hikone-A-S, Hikone-R 4267, Hikone-A-W, PVM Madeira Portugal #3861 and Zarate Argentina; 11th Column: P4 Perrysburg Ohio allele 2; 12th Column: Ky036 Kenya Malindi New Market; 13th Column: RC1 Riverside CA USA #3865; 14th Column: Florida-9 #2374 and Crimea #4266; 15th Column: BOG2 Bogota Colombia #3842, w^1118^ Urbana-S #4272, ZS11 Zimbabwe Sengwa Wildlife Reserve, Kis3 Uganda Kisoro; 16th Column: ZS15 Zimbabwe Sengwa Wildlife Reserve, 17th Column: P8 Perrysburg Ohio USA; intron2b alleles: 18th column: Oregon #5, INPA7 Manaus (this strain was heterozygote for intron2 types but had predominant intron2b), Swedish-C #4271, TW3 Varna NY USA #3874, MO1 Monroe County NY USA #3857, Q12 Israel #3864, Ber2 Bermuda #3840, MWA Madeira Wisconsin USA #3859, CA1 Capetown South Africa #3846, BV1 Blacksburg Virginia #3845, VAG1 Athens #3875, Samarkand #4270, Harwich #4264, M2 Australia #3855, Amherst-3 #4265, pi2cP #2384, BS1 Barcelona Spain #3844, Lausanne-S #4268, PYR-3 Pyrenees Spain #3863, Berlin-K #8522, P1 Perryburg Ohio USA, Koriba Dam #3854, P10 Perryburg Ohio, USA (majority intron2b but has traces of intron2a), P7 Perryburg Ohio USA, RSVP10 Brasil Santa Maria, Tana11 Madagascar Antananarivo, SP Brasil Sao Jose do Rio Preto, #3884 wild 3B; 19th Column: Canton-2 and Koriba Dam-3 #3853; 20th Column: P7 Perryburg Ohio USA; 21th Column: Reids-1 Madeira Portugal #3866. Exact descriptions and origins of each line can be found in Supplementary Data S1.
Supplementary Figure S2. Rate of nonsynonymous (dN) and synonymous (dS) polymorphisms in a 191-bp fragment of the RFeSP locus of D. melanogaster lines. dN/dS ratios are shown for intron2a group alleles (a) and intron2b group alleles (b). dN/dS ratios can be a robust method of inferring evolutionary processes acting on coding regions, if one considers the limitations of applying these measurements to intraspecies polymorphism studies. Interestingly, all of the eight coding region polymorphisms that were found segregating in D. melanogaster RFeSP were found in intron2a group alleles, while none were found in RFeSP alleles of the intron2b group. For intron2a alleles, the average values obtained were: dN = 0.0069 ± 0.0001, dS = 0.0691 ± 0.0202, and dN/dS = 0.1121, which although not statistically significant (Fisher’s exact tests, p > 0.1, considering dNObserved = dSObserved as the null hypothesis), roughly indicated a 9-fold excess of possible synonymous polymorphisms over nonsynonymous ones.
Supplementary Figure S3. Lack of association between intron2 allelic groups with the chromosomal inversion In(2L)t. (a) Scheme of the Standard chromosome 2L and the breakpoints of the In(2L)t inversion. The localization of the RFeSP locus in the Standard chromosome is depicted. Primer pairs are depicted as “A-D”. The presence of the In(2L)t is diagnosed by the amplification of a ~250-bp fragment. (b) Screening of subset of intron2a lines for In(2L)t. (c) Screening of a subset of intron2b lines for In(2L)t.
Supplementary Figure S4. Association between intron2 allelic groups with the gene copy polymorphism at the Or22 locus. (a) Scheme of the Or22 gene copy polymorphisms in D. melanogaster and genetic distance from RFeSP on the Chr2L\textsuperscript{26}. In D. melanogaster, Or22 segregates either as a long or a short variant\textsuperscript{26}. The long variant has a duplicated Or22 (Or22a and Or22b), and occurred in 74% of the strains in our study. The short variant consists of a single chimeric Or22a/b, occurring in 26% of the lines. (b) Linkage disequilibrium between RFeSP and Or22 for all RFeSP groups (see text). $X^2 = 13.982$. (c) Linkage disequilibrium between RFeSP and Or22 for all RFeSP intron2a groups (see text). $X^2 = 9$. Shown is the normalized D value and the Chi-square goodness of fit test.
Supplementary Figure S5. Synteny conservation of the \textit{RFeSP} chromosomal region. Synteny conservation of the RFeSP chromosomal region between \textit{D. melanogaster} and \textit{D. grimshawi}. Magenta arrows indicate key conserved syntenic genes. The localization of the \textit{RFeSP} locus is depicted by a Cyan arrow. Genome sequences were obtained from the UCSC Genome Browser. http://genome.ucsc.edu/ The \textit{D. melanogaster} assembly used was the Apr. 2006 (BDGP R5/dm3) Assembly. The \textit{D. grimshawi} assembly used was the Aug. 2005 (Agencourt prelim/droGri1) Assembly.
Supplementary Figure S6. RFeSP-RA codons were biased by negative selection on RFeSP. An example of a correlation between codon usage bias and the fixation of PTC loss in the cryptic coding sequence of RFeSP-RA. Shown is the PTC loss that happened following the Tyr_{199} mutation described in Figure 3b (main text). Red shows a bias for the TAU codon and blue shows a preference for the TAC codon.
Supplementary Figure. S7. Targeted mutagenesis completely eliminates splicing from the intron2bΔ62 locus. Agarose gel showing products of an RT-PCR with cDNA isolated from S2 cells transiently transfected with Noble::TagRFP-T fusions and mutants. Lane 1: pUAST-Noble-TagRFP-T. Lane 2: pUAST-NobleW164STOP-TagRFP-T fusion. Lane 3: pUAST-NobleOPT-TagRFP-T, which has no splice sites, at the cost of a Val159Ile mutation. Note that the presence of a STOP codon within intron2 (Lane 2) visibly favors splicing in comparison with the wild-type locus.
Supplementary Figure S8. Noble is probably processed and requires an intact N-terminus to reach the mitochondria. (a) Gal-4 inducible transgenes used to study Noble function in vitro and in vivo. In pUAST-Noble-TagRFP-T, TagRFP-T is inserted posteriorly to the RFeSP termination codon, but in frame to Noble. Nevertheless, in addition to Noble-TagRFP-T, the production of non-tagged RFeSP is expected from spliced pUAST-Noble-TagRFP-T (since this construct carries the intron2bΔ62). The predicted MW of Noble::TagRFP-T is ≈56 kDa. In pUAST-VisGreen-RFeSP/Noble, VisGreen was inserted N-terminally following the Methionine. Thus both RFeSP and Noble should be produced concomitantly, with a predicted MW of ≈52 and ≈56 kDa, respectively. (b) Proteins from S2 cells transiently transfected with different plasmids were separated in a 10% SDS-Page gel and subject to Western Blotting with anti-GFP. Green arrows show two specific bands found after transient transfection with pUAST-VisGreen-RFeSP/Noble. Transient transfections were carried out with equimolar ratios of: 1) pMT-Gal4+ pUAST; 2) pMT-Gal4 + pUAST-TagRFP-T-CG9925-HA; 3) pMT-Gal4 + pUAST-Noble-TagRFP-T; and 4) pMT-Gal4 + pUAST-VisGreen-RFeSP/Noble. (c) Same as in b, but the Western Blot was probed with anti-TagRFP-T. Red arrows point to specific bands found after transient transfection with Noble-TagRFP-T or TagRFP-T-CG9925-HA. Transfection 2 was performed to demonstrate that the anti-TagRFP-T antibody indeed recognizes TagRFP-T. The predicted MW of the TagRFP-T-CG9925-HA fusion is 128 kDa. Noble-TagRFP-T was found as a single band migrating slightly above 50 kDa, which is nevertheless under the predicted size of the fusion protein (~55 kDa), strongly suggesting that Noble is N-terminally processed. (d) Protein alignment of a fragment of RFeSP homologues from diverse taxa (from top to bottom: Homo sapiens, Mus musculus, Gallus gallus, Danio rerio, Aedes aegypti, Anopheles gambiae, Culex pipiens, Drosophila melanogaster, D. yakuba, D. willistoni, D. virilis, D. grimshawi, and Caenorhabditis elegans). The putative MPP processing signal RX-XHXD55-56 (the hyphen stands at the putative cleavage site, depicted with a pair of scissors on the alignment) is depicted by a red box. (e-j) VisGreen-RFeSP/Noble do not associate with mitochondria in D. melanogaster salivary gland cells in vivo. We used the ey-Gal4 driver to activate pUAST-VisGreen-RFeSP/Noble and pUAST-Noble-TagRFP-T in salivary glands. (e,g,h,j) VisGreen-RFeSP/Noble in green. (f,g,i,j) Mitochondria were visualized indirectly via Noble-TagRFP-T (in red). (h-j) Same as in e-g but with an enhanced brightness and contrast. Bar = 30 μm in panels e-j.
**Supplementary Table S1.** Complete list of lines and stocks of *D. melanogaster* and closely related *Drosophila* species used in this study.

| Lines (REF# Codes) | Description | Species | Origin | Obtained from or Collected by (date) |
|-------------------|-------------|---------|--------|-------------------------------------|
| 1                 | Canton-S    | *D. melanogaster* | Bloomington |
| 3                 | Hikone-A-S  | *D. melanogaster* | Bloomington |
| 4                 | Hikone-A-W  | *D. melanogaster* | Bloomington |
| 2057              | y¹; Gr22b¹ Gr22d¹ cn¹ CG33964¹¹¹¹ bw¹ sp¹ LysC¹ MstProx¹ GsdD5¹ Rb6⁰ | *D. melanogaster* | Bloomington |
| 5                 | Oregon-R-C  | *D. melanogaster* | Bloomington |
| 2374              | Florida-9   | *D. melanogaster* | Bloomington |
| 2384              | pi; <P>     | *D. melanogaster* | Bloomington |
| 3840              | BER 2, Bermuda | *D. melanogaster* | Bloomington |
| 3842              | BOG 2, Bogota, Colombia | *D. melanogaster* | Bloomington |
| 3844              | BS-1, Barcelona, Spain | *D. melanogaster* | Bloomington |
| 3845              | BV1, Blacksburg, Virginia | *D. melanogaster* | Bloomington |
| 3846              | CA1, Capetown, South Africa | *D. melanogaster* | Bloomington |
| 3850              | CO 7, Commack, New York, USA, 1961 | *D. melanogaster* | Bloomington |
| 3851              | EV, Ellenville, New York, USA | *D. melanogaster* | Bloomington |
| 3853              | KSA 3, Koriba Dam, South Africa | *D. melanogaster* | Bloomington |
| 3854              | KSA 4, Koriba Dam, South Africa | *D. melanogaster* | Bloomington |
| 3855              | M2, Australia | *D. melanogaster* | Bloomington |
| 3857              | MO 1, Monroe County, New York, USA | *D. melanogaster* | Bloomington |
| 3859              | MWA 1, Madera, Wisconsin, USA | *D. melanogaster* | Bloomington |
| 3860              | NO 1, New Orleans, Louisiana, USA | *D. melanogaster* | Bloomington |
| 3861              | PVM, Madeira, Portugal | *D. melanogaster* | Bloomington |
| 3863              | PYR 3, Pyrenees, Spain | *D. melanogaster* | Bloomington |
| 3864              | Q2, Israel | *D. melanogaster* | Bloomington |
| 3865              | RC1, Riverside, California, USA | *D. melanogaster* | Bloomington |
| 3866              | Reids 1, Madeira, Portugal | *D. melanogaster* | Bloomington |
| 3871              | RVC 4, Riverside, California, USA | *D. melanogaster* | Bloomington |
| 3874              | TW 3, Varna, New York, USA | *D. melanogaster* | Bloomington |
| 3875              | VAG 1, Athens | *D. melanogaster* | Bloomington |
| 3880              | Wild 2A, Painesville, Ohio, USA | *D. melanogaster* | Bloomington |
| 3884              | Wild 3B, Mt. Sterling, Ohio, USA | *D. melanogaster* | Bloomington |
| Stock number | Description | Species | Stock center |
|-------------|-------------|---------|--------------|
| 4264        | Harwich     | D. melanogaster | Bloomington  |
| 4265        | Amherst 3   | D. melanogaster | Bloomington  |
| 4266        | Crimea      | D. melanogaster | Bloomington  |
| 4267        | Hikone-R, high DMN demethylase activity | D. melanogaster | Bloomington  |
| 4268        | Lausanne-S  | D. melanogaster | Bloomington  |
| 4270        | Samarkand   | D. melanogaster | Bloomington  |
| 4271        | Swedish-C   | D. melanogaster | Bloomington  |
| 4272        | Urbana-S    | D. melanogaster | Bloomington  |
| 4736        | D. simulans C147.4 | D. melanogaster | Bloomington  |
| 7193        | w[+]; P[w+mc]=sqh-EYFP-Golgi3 | D. melanogaster | Bloomington  |
| 7194        | w[+]; P[w+mc]=sqh-EYFP-Mito3 | D. melanogaster | Bloomington  |
| 7195        | w[+]; P[w+mc]=sqh-EYFP-ER3 | D. melanogaster | Bloomington  |
| 8442        | UAS-mitoGFP | D. melanogaster | Bloomington  |
| 8522        | Berlin-K    | D. melanogaster | Bloomington  |
| 14021-0241.01 | Dmau/wild-type | D. mauritiana | Drosophila species stock center. |
| 14021-0251.195 | Dsim/w[501] | D. simulans | USA Drosophila species stock center. |
| 14021-0248.25 | Dsec/wild-type | D. sechellia | Cousin Island, Seychelles Drosophila species stock center. |
| 14021-0245-01 | Dore/wild-type | D. orena | Bafut, Ngemba, Cameroon Drosophila species stock center. |
| 14021-0257-00 | Dtei/wild-type | D. teissieri | Drosophila species stock center. |
| 14011-0111-49 | Dper/wild-type | D. persimilis | Mount. St. Helena, California, U.S.A. Drosophila species stock center. |
| 14021-0271-00 | Dstol/wild-type | D. santomea | Obo Natural Reserve, submontane forest, San Tome and Príncipe Island (Africa). Drosophila species stock center. |
| 14024-0371-13 | Dana(In3R)A | D. ananassae | Hawaii, USA. Drosophila species stock center. |
| 14021-0261-00 | Dyak/wild-type | D. yakuba | Ivory Coast Drosophila species stock center. |
| 14021-0261-01 | Dyak/wild-type | D. yakuba | Between Liberia and Ivory Coast, Liberia. Drosophila species stock center. |
| 14021-0224-01 | Dere/wild-type | D. erecta. | Drosophila species stock center. |
| 14011-0121-94 | Dpse/wild-type | D. pseudoobscura | Anderson, Mesa Verde, Colorado, USA. Drosophila species stock center. |
| 14027-0461.02 | Dele/wild-type | D. elegans | Okinawa, Japan. Drosophila species |
| Stock Code | Stock Type | Stock Name | Country/Location |
|------------|------------|------------|-----------------|
| 14022-0271.00 | Dlutf/wild-type | D. lutescens | Mito, Honshu, Japan |
| 14022-0311.14 | Dtak/wild-type | D. takahashi | Yun Shui, Taiwan |
| 14022-0301.01 | Dpsh/wild-type | D. pseudotakahashi | Atherton Tableland, North Queensland, Australia |
| 14023-0311.00 | Dsz/wild-type | D. suzukii | Hachijo Island, Tokyo |
| 14026-0451.05 | Deug/wild-type | D. eugracilis | Kuala Belalong, Ulu Temburong National Park, Brunei |
| 14023-0011.00 | Dfuy/wild-type | D. fayumai | Kuala Belalong, Ulu Temburong Ntl. Park, Brunei |
| 14029-0011.00 | Dprs/wild-type | D. prostipennis | Kuala Belalong, Ulu Temburong National Park, Brunei |
| 14023-0331.01 | Dluc/wild-type | D. lucipennis | Wulai, Taiwan |
| 14023-0311.00 | Dsuz/wild-type | D. suzukii | Hachijo Island, Tokyo |
| 14028-0701.00 | Dtsa/wild-type | D. tsacasi | Ivory Coast |
| 14025-0441.00 | Dfic/wild-type | D. ficsphila, | stock center |

From the Dominguez Lab.

Upf mutant 25G

Upf mutant 14J

Canton-S-BU

P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16

C1, C2

Perrysburg, OH, USA.

Cleveland, OH, USA.

R.C. Woodruff (Fall 2009)

C. Schloterrer lab.

Jean David (March 2008); Donated from C. Schloterrer lab.

Jean David (March 2008); Donated from C. Schloterrer lab.

Artyom Kopp

C. Niessinger (March 2001); Donated from C. Schloterrer lab.

M. Imhof (March 2001); Donated from C. Schloterrer lab.

J. Pool (Winter 2008/9);
| Code   | Species       | Location                          | Date                  | Source                        |
|--------|---------------|-----------------------------------|-----------------------|-------------------------------|
| EZ61   | *D. melanogaster* | Ethiopia, Ziway                    | Winter 2008/10;       | Donated from C. Schloterrer lab. |
| Tana11 | *D. melanogaster* | Madagascar, Antananarivo           | March 2008;           | Donated from C. Schloterrer lab. |
| Tana12 | *D. melanogaster* | Madagascar, Antananarivo           | March 2008;           | Donated from C. Schloterrer lab. |
| Kyo36  | *D. melanogaster* | Kenya, Malindi New Market          | July 2001;            | Donated from C. Schloterrer lab. |
| Kyo70  | *D. melanogaster* | Kenya, Malindi New Market          | July 2001;            | Donated from C. Schloterrer lab. |
| Kis3   | *D. melanogaster* | Uganda, Kisoro                     | March 2001;           | Donated from C. Schloterrer lab. |
| Kis4   | *D. melanogaster* | Uganda, Kisoro                     | March 2001;           | Donated from C. Schloterrer lab. |
| ZS11   | *D. melanogaster* | Zimbabwe, Sengwa Wildlife Res.     | C. F. Aquadro (1990); | Donated from C. Schloterrer lab. |
| ZS15   | *D. melanogaster* | Zimbabwe, Sengwa Wildlife Res.     | C. F. Aquadro (1990); | Donated from C. Schloterrer lab. |
| INPA-7 | *D. melanogaster* | Manaus, AM, Brazil                 | E. Loreto (2006-2008) |                               |
| Sao Jose| *D. melanogaster* | São Jose Rio Preto, SP, Brazil     | E. Loreto (2006-2008) |                               |
| RS-V10 | *D. melanogaster* | Santa Maria, RS, Brazil            | E. Loreto (2006-2008) |                               |
| Zaraté | *D. melanogaster* | Argentina                          | E. Loreto (2006-2008) |                               |
**Supplementary Table S2.** Values of DNA polymorphism and frequency spectrum neutrality tests for all *D. melanogaster* (Dmel) *RFeSP intron2* alleles together, as well as *intron2a* and *intron2b* groups separately. *D. simulans* (Dsim) *RFeSP intron2* alleles are shown as comparison. The *D. tiessieri RFeSP intron2* was used as an outgroup.

|       | Neutrality Tests (T) |       |       |       |       |       |       |       |       |
|-------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|
|       | m\(^a\) | S\(^b\) | k\(^c\) | \(θ_π\)\(^d\) | \(θ_s\)\(^e\) | D\(^f\) | \(D_2\)\(^g\) | \(D_3\)\(^h\) | F\(^i\)\(^*\) | F\(^i\) | H\(^j\) | Y\(^k\)\(^*\) | Y\(^k\)\(^*\) |
| Dmel  |                  |       |       |       |       |       |       |       |       |
| All   | 57    | 28    | 21    | 5.78  | 6.05  | -0.14 | -1.44 | -1.64  | -1.10  | -1.21  | 0.26  | 0.69  | 0.75  |       |
|       | (0.516)\(^3\) | (0.072) | (0.059) | (0.137) | (0.130) | (0.441) | (0.804) | (0.819) |       |       |       |       |       |
| 2a    | 26    | 21    | 17    | 3.99  | 5.50  | -0.99 | -0.93 | -1.08  | -1.03  | -1.10  | 0.25  | -0.76 | -0.54 |       |
|       | (0.164) | (0.168) | (0.188) | (0.150) | (0.158) | (0.434) | (0.240) | (0.319) |       |       |       |       |       |
| 2b    | 31    | 3     | 4     | 0.25  | 0.74  | -1.55 | -1.51 | -1.60  | -1.64  | -1.71  | -0.30 | -1.37 | -1.33 |       |
|       | (0.033) | (0.033) | (0.026) | (0.033) | (0.031) | (0.389) | (0.078) | (0.090) |       |       |       |       |       |
| Dsim  | 6     | 18    | 6     | 7.73  | 7.88  | -0.12 | -0.10 | -0.41  | -0.10  | -0.38  | -0.61 | -0.15 | -0.04 |       |
|       | (0.478) | (0.470) | (0.421) | (0.485) | (0.446) | (0.180) | (0.545) | (0.317) |       |       |       |       |       |

\(^a\) m = number of sequences analyzed. A ~300-bp sequence around the *RFeSP intron2* was used (fig. S1).

\(^b\) S = number of segregating sites. Insertions and deletions were counted as one segregating site.

\(^c\) k = number of haplotypes

\(^d\) \(θ_π\) = diversity estimator, as in (41)

\(^e\) \(θ_s\) = diversity estimator, as in (41)

\(^f\) D = Tajima’s D (57)

\(^g\) \(D_2\) = Fu and Li’s \(D_2\) as described in (58)

\(^h\) \(D_3\) = Fu and Li’s \(D_3\) as described in (58)

\(^i\) F = Fay and Wu’s \(Y^*\) as described in (59)

\(^j\) \(Y^*\) = Achaz’s \(Y^*\) Y as described in (41)

\(^k\) For each Test (T) p value = P(Tobs > Ttheo) according to 100000 coalescence simulations (41)
**Supplementary Table S3.** Survey of 222 nucleotides from the *RFeSP-RA* and *RFeSP-RB* overlapping coding sequence. Sequence starts on the first Ala<sub>177</sub> codon of *RFeSP-RA* (which is about where the 3<sup>rd</sup> exon of *RFeSP-RB* begins) and ends in the Glu codon of *RFeSP-RA* (which is equivalent to the STOP codon of *RFeSP-RB*).

| aa # | aa<sup>a</sup> | codon | possible PTC mutation | RFeSP-RA | preferred codon<sup>b</sup> | resulting aa | resulting codon | RFeSP-RB | translat. change<sup>c</sup> | codon bias | no PTC<sup>d</sup> | neutral |
|------|----------------|-------|-----------------------|----------|-----------------------------|--------------|----------------|----------|--------------------------|------------|--------------|---------|
| Exon3 starts on “C” |
| 177  |               | GCA   |                       |          |                             |              |                |          |                          |            |              |         |
| 178  |               | ACG   |                       |          |                             |              |                |          |                          |            |              |         |
| 179  | Cys           | TGT   | TGA                   |          | Val                         | GTG          | GTG            | Glu      | GAG                      | 1          | 2            |         |
| 180  |               | GAT   |                       |          |                             |              |                |          |                          |            |              |         |
| 181  | Glu           | CAA   | TAA                   |          | Iso                         | ATC          | ATC            | Iso      | ATT                      | 1          | 2            |         |
| 182  |               | GCC   |                       |          |                             |              |                |          |                          |            |              |         |
| 183  | Arg           | CGA   | TGA                   |          | Pro                         | CCC          | CCC/CCG        | Pro      | CCT                      | 1          | 2            |         |
| 184  |               | GTG   |                       |          |                             |              |                |          |                          |            |              |         |
| 185  |               | GCT   |                       |          |                             |              |                |          |                          |            |              |         |
| 186  |               | GGT   |                       |          |                             |              |                |          |                          |            |              |         |
| 187  |               | GGT   |                       |          |                             |              |                |          |                          |            |              |         |
| 188  |               | CAT   |                       |          |                             |              |                |          |                          |            |              |         |
| 189  |               | CGG   |                       |          |                             |              |                |          |                          |            |              |         |
| 190  |               | AGT   |                       |          |                             |              |                |          |                          |            |              |         |
| 191  |               | GTC   |                       |          |                             |              |                |          |                          |            |              |         |
| 192  |               | CAC   |                       |          |                             |              |                |          |                          |            |              |         |
| 193  |               | GCA   |                       |          |                             |              |                |          |                          |            |              |         |
| 194  |               | TCT   |                       |          |                             |              |                |          |                          |            |              |         |
| 195  |               | GGG   |                       |          |                             |              |                |          |                          |            |              |         |
| 196  |               | CTG   |                       |          |                             |              |                |          |                          |            |              |         |
| 197  |               | TGT   |                       |          |                             |              |                |          |                          |            |              |         |
| 198  |               | GCC   |                       |          |                             |              |                |          |                          |            |              |         |
| 199  |               | CAT   |                       |          |                             |              |                |          |                          |            |              |         |
| 200  |               | CGC   |                       |          |                             |              |                |          |                          |            |              |         |
| 201  | Glu           | GAA   | TAA                   | Ala      | GCG                         | GCC/GCG      | Ala            | GCT      | 1                        | 2          |              |         |
| 202  |               | CGC   |                       |          |                             |              |                |          |                          |            |              |         |

<sup>a</sup> aa: amino acid; <sup>b</sup> codon: codon; <sup>c</sup> translat. change: translation change; <sup>d</sup> no PTC: no PTC (premature termination codon).
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 203 | CGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 204 | Arg | CGA | TGA | Gly | GGC | GGC | Gly | GGT |   |   |   |   |   |   |   |   |
| 205 |   | CTG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 206 |   | GGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 207 | Trp | TGG | TGA | Gly | GGC | GGC | Asp | GAC |   |   |   |   |   |   |   |   |
| 208 |   | CTA |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 209 |   | CTA |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 210 |   | CTG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 211 |   | CCC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 212 |   | CTG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 213 |   | CCA |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 214 |   | CGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 215 |   | CTC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 216 |   | CCA |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 217 |   | CTA |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 218 | Arg | CGA | TGA | Tyr | TAC | TAC | Tyr | TAT |   |   |   |   |   |   |   |   |
| 219 |   | CGC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 220 |   | CTC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 221 |   | CGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 222 | Lys | AAG | TAG | Gly | GGA | GGC | Gly | GGT |   |   |   |   |   |   |   |   |
| 223 |   | GAT |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 224 |   | CGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 225 | Glu | CAA | TAA | Arg | CGC | CGC/CGG | Arg | CGT |   |   |   |   |   |   |   |   |
| 226 |   | GGG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 227 |   | ACC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 228 |   | GCG |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 229 |   | GCC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 230 |   | CCT |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 231 | Glu | CAA | TAA | Leu | CTC | CGT/CTC | Leu | CTA |   |   |   |   |   |   |   |   |
| 232 |   | CTT |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 233 | Gly | GGA | TGA | Leu | TTG | CGT/CTC | Phe | TTT |   |   |   |   |   |   |   |   |
| 234 |   | GGT |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 235 |   | GCC |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| Exon3 ends on first “G” | 236 | CAC | 3 |
|-------------------------|-----|-----|---|
| 237                     | CCA | 3  |
| 238                     | Arg | CGA | TGA | His | CAC | CAC | His | CAT | 1 | 2 |
| 239                     | GTT | 3  |
| 240                     | CCC | 3  |
| 241                     | Glu | CAA | TAA | Pro | CCC | CCC/CCG | Pro | CCT | 1 | 2 |
| 242                     | Arg | CGA | TGA | Asn | AAC | AAC | Asn | AAT | 1 | 2 |
| 243                     | GGG | 3  |
| 244                     | TCT | 3  |
| 245                     | TCT | 3  |
| 246                     | CGT | 3  |
| 247                     | GGT | 3  |
| 248                     | CGG | 3  |
| 249                     | CTA | 3  |

| Glu | GAG | TAG | PTC | TAG/TAG/TAA | Tyr | TAT | 1 | 2 |

| TOTAL | 74 aa | 222 nt | 4 | 10 | 207 | 1 |
|-------|-------|--------|---|----|-----|---|
| P(N) | 0.0060 | 0.0150 | 0.9775 | 0.0015 |

a Number of nucleotides affected by the specific type of mutation.

b Only aa that can change to PTCs (premature translation termination codons) in RFeSP-RA are listed.

c According to (42).

d Mutations that would affect translation, such as inducing an aa change or removing the original STOP codon from RFeSP-RB.

e Mutations that do not induce a PTC in the reading frame of RFeSP-RA.

f P(N) is the probability of the respective type of mutation amongst possible nucleotide substitutions in this 222 nucleotide stretch.

For each set there is only one nucleotide substitution that would result in a PTC (e.g., GGA must mutate to GGT, and not GGC or GGG to convert Lys22, of RFeSP-PA into a induce a PTC), therefore, the probability of a “neutral” mutation to RFeSP-RB inducing a PTC on RFeSP-RA is $1/222 \times 1/3 = 0.0015$. 

20
**Supplementary Table S4.** List and origin of *Non-Drosophilinae* insects used in this study.

| Code      | Order     | Family       | Genus     | Species     |
|-----------|-----------|--------------|-----------|-------------|
| Nser1<sup>a</sup> | Megaloptera | Corydalidae  | Nigronia  | serricornis |
| Bcol1<sup>a</sup> | Mecoptera  | Boreidae     | Boreus    | coloradensis|
| Phel1<sup>a</sup> | Mecoptera  | Panorpidae   | Panorpa   | helena      |
| Hapi1<sup>a</sup> | Mecoptera  | Bittacidae   | Hylobittaus | apicalis   |
| DI43<sup>a</sup> | Diptera    | Asilidae     | Sarapogon | fletcheri   |
| Psycho1<sup>b</sup> | Diptera    | Psychodidae  |           |             |
| Tipula<sup>b</sup>  | Diptera    | Tipulidae    | Tipula    | sp.         |
| Sipho_Ctenocephalides<sup>c</sup> | Siphonaptera | Pulicidae  | Ctenocephalides | canis     |

<sup>a</sup> Collected and classified by the Whiting lab

<sup>b</sup> Collected and classified by A. Gontijo in El Campello, Spain.

<sup>c</sup> DNA Sample donated by Dr. Michael J. Raupach Forschungsmuseum Alexander Koenig DNA-Bank
### Supplementary Table S5. List of primers used in this study.

| Primer name                  | Primer sequence               |
|-----------------------------|--------------------------------|
| **Sequencing primers (RFeSP)** |                                |
| **Dmelanogaster:**          |                                |
| RFeSP alternative1 left     | AAACGATAAACAGCTCTCTCTAGCC      |
| RFeSP alternative1 right    | GAGGGCAAGTCGGTTACTTTCA         |
| **Dpersimilis**             |                                |
| RFeSP al1 persi left        | AAACGATAAACACTGTAATGT          |
| RPeSP al1 perypseudo right  | GAGGGCAAGTCGGTGACATTCA         |
| **Dananassae:**             |                                |
| RFeSP al1 anana left        | AAACGATAAACAGCTCTCT            |
| RFeSP al1 anana right       | GAGGGCAAGTCGGTGACATTCA         |
| **Dpseudoosbcura:**         |                                |
| RFeSP al1 pseudo left       | AAACGATAAACACACTGTAAAT         |
| RPeSP al1 perypseudo right  | GAGGGCAAGTCGGTGACATTCA         |
| **Degenerate primers:**     |                                |
| RFeSP deg forw1             | GAYAATHCCIGARGGNAA             |
| RFeSP deg rev2              | RTGRGAIGGRCARTARTA             |
| RFeSP deg rev2OK            | RTGRCAIGGRCARTARTA             |
| **RFeSP primers**           |                                |
| RFeSP left                  | ATCGGTGCACACTCCGATGA           |
| RFeSP right                 | GCGATCCGGAGGTAGTGAT            |
| RFeSP general left          | GTTCGTACGACTGTTCTC             |
| RFeSP general right         | TGACGTAAGCAGCCGAC             |
| RFeSP large-isoform left    | CTCTAGCCGACACGAGAAG            |
| RFeSP large-isoform right   | CGGAGGCTGATGATGAGGT            |
| RFeSP alternative1 left     | AAACGATAACAGCTCCTCTAGCC        |
| RFeSP alternative1 right    | GAGGGCAAGTCGGTTACTTTCA         |
| RFeSP alternative2 left     | CAACCTCCGATGACCA              |
| RFeSP alternative2 right    | CAATCGCCAGGCACACC             |
**Supplementary Table S6.** List of the Accession numbers or references for all sequences used in this study.

| Organism                          | Accession number reference                                      |
|-----------------------------------|-----------------------------------------------------------------|
| *Homo sapiens*                    | NP_005994.2                                                      |
| *Mus musculus*                    | NP_079986.1                                                      |
| *Gallus gallus*                   | NP_001005843.1                                                  |
| *Danio rerio*                     | NP_001096664.1                                                  |
| *Acrysthosiphon pism*              | NP_001156726.1                                                  |
| *Pediculus humanus corporis*      | XP_002430151.1                                                   |
| *Graphocephala atropunctata*      | ABD98746.1                                                       |
| *Apis melifera*                   | XP_394657.1                                                      |
| *Nosonia vitripennis*             | XP_001607363.1                                                   |
| *Tribolium castaneum*             | NP_001164310.1                                                  |
| *Bombyx mori*                     | NP_001106738.1                                                  |
| *Plutella xylostella*             | ACF21937.1                                                       |
| *Aedes aegypti*                   | CH477279.1                                                       |
| *Anopheles gambiae*               | XP_319708.4                                                      |
| *Culex pipiens quinquefasciatus*  | XP_001867379.1                                                  |
| *Armigeres subalbatus*            | EU205098 (ASAP ID ACN-0185249)                                  |
| *Phlebotomus papatasi*            | Jun09 >gnl.til1948740609 >gnl.til1949783636>gnl.til1948738511  |
| *Lutzomyia longipalpis*           | >gnl.til2009434506 name!272012802                                |
| *Glossina morsitans*              | ADD20226.1, Supercontig_0001640 GeneDB; cn1072 chr IGGG1       |
| *Rhagoletis pomonella*            | EZ137853 (contig21634)                                          |
| *Drosophila melanogaster*         | NP_722715.1                                                      |
| *D. sechellia*                    | XP_002041717.1                                                  |
| *D. simulans* (named herein Dsim4) | XP_002077770.1                                                 |
| *D. yakuba* (named herein Dyak1)  | XP_002087556.1                                                  |
| *D. erecta*                       | XP_001968304.1                                                  |
| *D. ananassae*                    | XP_001963250.1                                                  |
| *D. pseudoobscura*                | XP_002132913.1                                                  |
| *D. persimilis*                   | XP_002020736.1                                                  |
| *D. willistoni*                   | XP_002064944.1                                                  |
| *D. virilis*                      | XP_002051732.1                                                  |
| *D. mojavensis*                   | XP_002001879.1                                                  |
| *D. grimshawi*                    | XP_001988905.1                                                  |
| *Caenorhabditis elegans*          | NP_501361.1                                                      |
Supplementary Discussion

The relationship between Tipulidae and other Diptera and its possible implications on the intron2 evolutionary course

In our analyses of the evolution of *RFeSP intron2*, *Tipula* sp. was placed basal to both Culicomorpha and Psychodomorpha based on the apparent consensus between morphological and molecular data\(^{27, 28, 45-47}\). Tipulidae (crane flies) have been traditionally placed as sister to all other Diptera based on morphological characters\(^{27}\). Molecular studies do not disagree with this, but clearly show that lower Diptera diverged very quickly, so that support for the branching relationships between these families is generally low\(^{45}\). However, it has been suggested that Tipulomorpha might not be the earliest branching infra-order in Diptera. This position has been recently attributed to Deuterophlebiidae\(^{45}\), which has not been sampled in the current study. The establishment of a more solid relationship between early branching Diptera will allow the conclusion whether or not *intron2* at Asp\(_{158}\) is a synapomorphy of Diptera or not. The importance of these relationships for the correct tracing of the history of *intron2* is reflected in the following possibility: if Tipulomorpha are indeed basal to Culicomorpha and Psychodomorpha, the crane flies could have fixed an intermediary state between the loss of the 70-nt upstream intron that is found in the other Antliophora taxa and the gain of the dipteran *Asp\(_{158}\) intron2*. This type of change could provide evidence of a smoother transition than the severe *RFeSP* locus remodeling that we have proposed during the divergence of Diptera from other Antliophora. If on the other hand, Tipulomorpha are not basal to Culicomorpha, the absence in *Tipula* sp. could reflect a simple secondary loss. Additional sampling in Tipulomorpha could shed light into this question. That *intron2* at
Asp$_{158}$ is found in both Culicomorpha and Psychodomorpha strongly suggests that it was present in an early Dipteran ancestor. It will be interesting to increase sampling in lower Diptera to narrow down this problem. Actually, it might be that this intron helps resolve these polytomies, or strengthen poorly supported branches.
**Supplementary References**

55. Hartl, F.U., Schmidt, B., Wachter, E., Weiss, H., Neupert, W. Transport into mitochondria and intramitochondrial sorting of the Fe/S protein of ubiquinol-cytochrome c reductase. *Cell.* **47**, 939-951 (1986).

56. Brandt, U., Yu, L., Yu, C.A., Trumpower, B.L. The mitochondrial targeting presequence of the Rieske iron-sulfur protein is processed in a single step after insertion into the cytochrome bc1 complex in mammals and retained as a subunit in the complex. *J. Biol. Chem.* **268**, 8387-8390 (1993).

57. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics.* **123**, 585-595 (1989).

58. Fu, Y.X., Li, W.H. Statistical tests of neutrality of mutations. *Genetics.* **133**, 693-709 (1993).

59. Fay, J.C., Wu, C.I. Hitchhiking under positive Darwinian selection. *Genetics.* **155**, 1405-1413 (2000).