Selective Sweep in the *Flotillin-2* Region of European *Drosophila melanogaster*

Annegret Werzner\*a, Pavlos Pavlidis\*b, Lino Ometto\*c, Wolfgang Stephan, Stefan Laurent\*

Section of Evolutionary Biology, Department of Biology II, University of Munich, Planegg-Martinsried, Germany

Abstract

Localizing genes that are subject to recent positive selection is a major goal of evolutionary biology. In the model organism *Drosophila melanogaster* many attempts have been made in recent years to identify such genes by conducting so-called genome scans of selection. These analyses consisted in typing a large number of genetic markers along the genomes of a sample of individuals and then identifying those loci that harbor patterns of genetic variation, which are compatible with the ones generated by a selective sweep. In this study we conduct an in-depth analysis of a genomic region located on the X chromosome of *D. melanogaster* that was identified as a potential target of recent positive selection by a previous genome scan of selection. To this end we re-sequenced 20 kilobases around the *Flotillin-2* gene (*Flo-2*) and conducted a detailed analysis of the allele frequencies and linkage disequilibrium observed in this new dataset. The results of this analysis reveal eight genetic novelties that are specific to temperate populations of *D. melanogaster* and that may have arisen during the expansion of the species outside its ancestral sub-Saharan habitat since about 16,000 years ago.

Introduction

Localizing genes that are subject to recent positive selection is a major goal of evolutionary biology. Recent methodological and theoretical advances are facilitating their identification by allowing the shift from candidate-locus approaches to genome-wide analyses. This is particular relevant for studying the effects of recent positive selection, which may be hampered by the confounding effects of demography. Within a population an ecologically favored allele might rise in frequency and become fixed in the population. This process is often caused by environmental changes or colonization of new environments. In case the favored allele already exists in the population selection operates on standing genetic variation, otherwise a favored mutation must arise de novo. If selection is strong, the fixation of the beneficial mutation can have a joint effect on linked neutral sites, which are also expected to increase in frequency. This process, known as genetic hitchhiking [1], generates a signal of i) reduced genetic variation at the target of selection [2,3], ii) a skewed site frequency spectrum (SFS) due to an excess of rare and high-frequency derived alleles [4,5,6], and iii) increased linkage disequilibrium (LD) on both sides of the target of selection but reduced LD between them [7,8,9].

The search for adaptive signals is usually carried out using a large number of loci on a genome-wide scale to identify regions that show considerable deviation from neutral expectations. As generally known, however, standard neutrality defined by the Wright-Fisher model is rarely met in natural populations, thus cautioning on the use of the classical neutrality tests. In fact, population structure and demographic events such as population bottlenecks or population size expansion can mimic genetic footprints of selection and possibly lead to false positives [5,10,11]. However, a useful neutral expectation may be provided by a demographic null model that is fitted to the demographic history of a population [12].

This approach has been used by Glinka et al. [13] and Ometto et al. [14] who performed a chromosome-wide scan of DNA variation in a derived European population of *Drosophila melanogaster* and compared it to a putative ancestral population from Southeast Africa. More than 250 loci on the X chromosome were used to evaluate the contribution of adaptive evolution in the European population. The demographic null model of the European population was defined as a bottleneck caused by the post-glacial colonization of Eurasia around 10,000–15,000 years ago [14,15,16,17]. While the bottleneck could account for most of the reduction of variation observed in the European population, the analysis revealed several loci and regions whose very low level of genetic variation and highly skewed SFS were not compatible with the expectations under such a demographic scenario.

In this follow-up study, we concentrate on one of these low-variation regions comprising about 20 kb centered around the cytological position 13A1 between landmarks X:14810532 and X:14829908 (FlyBase, FB2012_06, released November 6th, 2012).
Overall, five genes are located within this region: CG9504, CG9305, CG32591, CG9509, and CG32593.

CG9504 and CG9305 are two of the 12 genes that were assigned to the GMC oxidoreductase family and are clustered on the X chromosome of D. melanogaster [18]. The functions of these GMC homologues are still largely unknown except for CG9504, which was described as ecdysone oxidase (Eo) [19]. An expression analysis showed an up-regulation of both Eo and CG9505 during embryonic and metamorphic development, possibly indicating a joint function in the ecdysteroid metabolism [18]. Remarkably, the GMC cluster is located within the second intron of the Flo-2 gene with opposing transcription orientations. This arrangement is conserved throughout four distantly related insect species: D. melanogaster, Anopheles gambiae, Apis mellifera, and Tribolium castaneum [18], consistent with the occurrence of strong purifying natural selection.

CG9509, named pudgy (pdg), encodes a long chain fatty acid CoA ligase and is up-regulated in the case of food deprivation [20]. As transcriptional target gene of the insulin pathway, the expression of pdg is crucial for several metabolic parameters, such as glycogen storage and lipid homeostasis [21]. In general, nutritionally regulated genes are expected to orchestrate the energy homeostasis control, including the ability to mobilize stored energy resources, and pdg is of vital importance in this process.

CG32593, flotilin-2 (Flo-2), is highly conserved among a wide range of species (orthologous to the vertebrate reggie I) and encodes for Flotilin-2, a scaffolding protein that is involved in the formation of non-caveolar lipid rafts [22]. In Drosophila, Flo-2 has been found to be required to delimit the spread of epidermal wound response [23] and is an important component of the morphogens Wnt and Hedgehog [24]. It is a single transmembrane protein that is characterized by a short membrane-anchoring segment at the N-terminal part of the protein and a large cytoplasmic C-terminal domain [25]. The exact mode of membrane association is dependent on co-translational protein modifications at highly conserved N-terminal regions. If these regions are altered by mutations or if the protein is truncated the ability of membrane anchoring is lost. Flo-2 is expressed predominantly in neuronal structures such as the optic lobes and the central brain during all developmental stages of D. melanogaster [26]. Overexpression of Flo-2 leads to detrimental effects during the development of eyes, ocelli, bristles, and wings [27], while knockout mutants surprisingly show no noticeable phenotypic abnormalities. As Flo-2 covers almost the entire genomic section of the candidate region we will refer to it as the ‘Flo-2 region’.

Using the full-length sequence of the Flo-2 region in both the European and African population samples mentioned above we applied two statistical tests for selection: 1) SweepFinder based on information of the SFS [28], and 2) the ω statistic based on measures of LD [7,29]. The demographic null-model was inferred by simulations from the colonization model suggested by Laurent et al. [30]. Similar approaches were successfully used in other case studies of selective sweeps in D. melanogaster in our laboratory [31,32,33].

Materials and Methods

Fly Strains and Conditions of Culture

Intraspecific data were collected from highly inbred D. melanogaster lines, 12 derived from a European population (Leiden, The Netherlands) and 12 from an African population (Lake Kariba, Zimbabwe). All stocks were kept at 23°C, 45% humidity, and under constant light conditions. Development took place on a high-nutrient killed yeast food medium (12 ml) in glass vials of 200 ml. For interspecific comparison, we used the annotated sequence of D. sechellia (http://flybase.org, Release 5.31) [34].

Sequence Data Collection and Analysis

Primers were designed based on the D. melanogaster genome (FlyBase, Release 5.1). We amplified and sequenced the complete genomic region X:14010552-X:14829906 using 36 partially overlapping DNA fragments (primers are given in Table S1). DNA sequences were obtained from individual male flies and generated as described in Glösa et al. [15]. Sequences were assembled into contigs using the program Seqman (DNASTar, Madison, WI, USA). Finally, sequences for all lines were aligned using the algorithm MUSCLE (Edgar 2004) with the online tool available at http://mobyle.pasteur.fr/cgi-bin/portal.py [35], creating a 20,011 bp alignment of 24 D. melanogaster lines and D. sechellia as outgroup. Sequences have been deposited in GenBank (accession numbers: KC460991-KC461010).

Basic population genetic parameters were estimated by a sliding window analysis (window size of 1,000 bp with 500 bp overlap) using the program DnaSP 5.0 [36]. We estimated nucleotide diversity using π [37] and θW [38]. The allele frequency distribution was measured with the summary statistic Tajima’s D [39] and based on the total number of segregating sites. The interspecific divergence to D. sechellia was determined for all 24 inbred lines of D. melanogaster. Furthermore, we used D. sechellia to polarize the state of the segregating sites in our population sample. A variant was considered ancestral if it was shared between both species and derived if it was present only in D. melanogaster.

A Wilcoxon rank sum test was performed to compare the nucleotide diversity of the Flo-2 region with the entire X chromosome [14]. To avoid multiple testing due to window overlaps, only non-overlapping neighboring windows were selected for the analysis.

Demographic Modeling

The selective sweep detection methods used in this study (i.e. SweepFinder and ω) requires the specification of a neutral demographic model. For this we used a slightly modified version of the demographic model of Laurent et al. [30], which describes our current understanding of the demographic history of African, European, and Asian natural populations of D. melanogaster. In this study we re-estimated the parameters of the Laurent et al. [30] model using an Approximate Bayesian Computation (ABC) approach [40], while adding to the summarized dataset the number of fragments that are monomorphic in the African, European, and Asian samples. These fragments are generally removed from ABC estimations because standard statistics like Tajima’s D [39] are undefined when the number of segregating sites is zero, which in turn causes technical problems in subsequent calculations. Nevertheless, the frequency of such monomorphic fragments within a genome scan does carry information about the past action of genetic drift within a population and should be taken into account in ABC model inference. This inclusion forces our demographic model to account for the fact that genomic regions lacking genetic variation, like the Flo-2 region, can arise due to genetic drift alone without invoking the action of positive selection. Times of divergence and effective population sizes of the three populations were estimated applying the ABC algorithm described in Laurent et al. [30] to a genome-wide nucleotide polymorphism dataset taken from previous studies [13,14,30,41]. The observed values for the number of monomorphic fragments within the African, European, and Asian populations are 0, 16, and 46, respectively. The prior distributions used for the ABC estimations are described in Table S2.
Selective Sweep Analysis

First, the European dataset was analyzed using the SweepFinder algorithm [28]. Typically, SweepFinder compares the SFS of a small region of the genome (‘window’) to the SFS of the rest of the chromosome, which is considered neutral. In this study, the neutral SFS has been estimated using simulations of the above-mentioned version of our demographic model for the European population. For data analysis, we defined 1,000 overlapping windows of variable sizes according to the strength of selection and recombination rate.

We calculated the composite likelihood ratio (CLR) for each of these 1,000 windows along our sequence alignment for two models: a model without selection based on the neutral SFS vs. a model of a recent selective sweep, as suggested by Kim and Stephan [6] and later modified in SweepFinder by Nielsen et al. [29]. Thus, we consider the spatial pattern of allele frequencies along the studied genomic sequence, as predicted by a selective sweep model given the background pattern of a bottleneck.

Incorporating the demographic history of the European population in the estimation of the null distribution controls the false-positive rate [29]. As the European sample lacks genetic variation for a large part of the sequence we included monomorphic sites in our analysis. We used the combined dataset of the European and African samples to determine the state of monomorphic European sites that were polymorphic in Africa. This is expected to increase the power for detecting the signature of a selective sweep.

Second, the European dataset was subjected to an analysis of LD using the \(\theta\) statistic developed by Kim and Nielsen [7]. The selective sweep model predicts elevated levels of LD within the two flanking regions of the selected site, while LD is not expected to extend across the two regions. This pattern is identified based on high values of \(\theta\). As before, the dataset was split into 1,000 overlapping windows spanning between 2,000 and 10,000 bp. The borders of the two flanking regions were allowed to vary, and the window size was chosen according to the size for which \(\theta\) assumes the maximum value [29].

Statistical significance of the maximum values assumed by both statistics, CLR\(\max\) and \(\theta\)\(\max\), was inferred from 10,000 neutral coalescent simulations using a slightly modified version of the demographic model of Laurent et al. [30]. To do these simulations we fixed \(\theta\) using the effective population sizes estimated/used by the Laurent et al. model and divergence-based mutation rate estimates. This approach was preferred over the one that fixed the number of segregating sites, which has been shown to be estimates. This approach was preferred over the one that fixed the

Results

Sequence Data Collection and Analysis of the Flo-2 Region

To detect positive selection in the European sample of D. melanogaster we sequenced one of the low-variation regions discovered by Ometto et al. [14] using 12 inbred lines from the Netherlands (Leiden) and 12 inbred lines from the ancestral African range (Zimbabwe). The full sequence comprises 20,011 bp in total and is referred to as the Flo-2 region. It is closely located to one of the fragments sequenced by Ometto et al. [14] that was added to Figure 1.

The sliding window analysis showed a conspicuous reduction of genetic variability for the European sample compared to the African one, with a valley of low variation of around 10 kb in size (6.7 kb–16.8 kb; see line with black square, Figure 1). In the European sample the mean value of nucleotide variation in the Flo-2 region is \(h = 0.0024\), which is significantly lower than the observed value for the entire X chromosome in European D. melanogaster (\(h = 0.0047\) [14]; fine dashed horizontal line in Figure 1; Wilcoxon rank sum test, \(P = 0.0026\)). The corresponding genetic region in the African sample has a nucleotide diversity of \(h = 0.0141\), which matches the mean value of the entire X chromosome (\(h = 0.0131\) [14]; wide dashed horizontal line, Figure 1; Wilcoxon rank sum test, \(P = 0.1766\)). Our fully sequenced fragment covers several genes including Flo, CG9503, CG32291, parts of pdg and eight different transcripts of Flo-2 (Figure 1, Gene map, from left to right). Nonetheless, functional constraint as the sole cause of this pattern of genetic variability in the European sample can be ruled out as the level of genetic variation of the African sample is intermediate to high, and divergence to the sibling species D. sechellia remains constantly high at around 6%.

Tajima’s \(D\) statistic is negative in the European sample for most parts of the Flo-2 region with an average value of \(D_k = -0.6416\) (\(D_X = -0.103\). [14]; Figure 1, D Europe). The African sample shows an average value of \(D_k = -0.2088\) in the Flo-2 region, a value that is above the European value and the African chromosomal average (\(D_X = -0.608\), a value that was considered to reflect population growth [14]; Figure 1, D Africa). Thus, in the Flo-2 region hallmarks of positive directional selection (including reduced variation and skews in the SFS) are only observed in the European population.

Demographic and Selective Sweep Analyses

The results of our demographic analysis are summarized in Table 1. Adding the number of monomorphic fragments to the vector of observed statistics in our ABC inference procedure did not change much the demographic estimates proposed by Laurent et al. [30]. The SweepFinder statistic for the Flo-2 region is significant at the 5% threshold obtained by neutral simulations (see Materials and Methods; Figure 2, CLR). The \(\theta\) statistic assumes relatively high values, but is not significant (Figure 2).

Private European Alleles

Geographically restricted genetic hitchhiking suggests that the target of selection should be fixed in the European population sample and absent in the African population sample. Overall there are 11 fixed different nucleotide substitutions between the African and the European sample within the Flo-2 region (1–11, Table 2; black diamonds in Figure 1). Four of them are located within exons, while the remaining are located within Flo-2 introns. In eight such cases the European population differs from the ancestral state of the African population (i.e., African D. melanogaster and D. sechellia share the same state), suggesting that they are derived private European alleles. The overrepresentation of eight derived alleles in the 10-kb segment of reduced variation within the 20-kb long Flo-2 region is highly significant (binomial distribution, \(P = 0.004\)). In the following we describe these fixed differences in detail according to their location across the genes present in the Flo-2 region.

CG9503. Three fixed differences are located within the gene CG9503: one within the first exon, and two within the second exon. Within the second exon the substitution that is private to the European sample leads to an alteration of the amino acid sequence. Namely, a serine (Ser = ancestral state) is substituted by an asparagin (Asn) at the amino acid position 473. This substitution is not expected to alter the gene function as both amino acids have equivalent characteristics [hydrophilic, aliphatic,
polar, neutral). The two other fixed differences are synonymous substitutions in the African sample while the European alleles are of ancestral state.

**Flotillin-2.** The intronic region of Flo-2 carries seven fixed differences between the European and the African populations, six of which are of derived state in the former (Table 2). One fixed difference is located at a synonymous site of Flo-2 without causing an amino acid substitution. Another fixed difference between the European and African samples is an indel polymorphism due to a dinucleotide microsatellite within the first exon of the transcript C of the Flo-2 gene (Figure 1, gene map, Flo-2-C), and which results in the exon being 20 bp longer in the European sample than the African sample. Since the African D. melanogaster lines share the same state with D. sechellia, the most parsimonious hypothesis is that European lines gained the additional nucleotides. Interestingly there are a total of 38 indels spread along the Flo-2 region (Figure 1, black triangles on the x-axis) but only this one is fixed between the African and European samples (Table 2, nr. 12). The annotated version of transcript form Flo-2-C at Flybase (version FB2010_05, released May 28th, 2010) is strongly supported by a full-length cDNA clone generated and sequenced by Rubin et al. [49]. The sequence has been subjected to integrity checks for accuracy, the presence of an open reading frame (ORF) as well as the start and stop codon. Interestingly, not all of our studied D. melanogaster lines have an ORF (Table S3, Flo-2-C alignment starts at transcription site +46). Specifically, seven African lines show premature termination codons and two lines have frame shifts due to microsatellite length polymorphism leading to a stop codon loss. The remaining three African lines have a gene version that forms the Flo-2-C splicing variant in the correct way (lines 186, 377, 384).

In contrast, in the European sample the majority of the lines – ten out of the twelve – have an integral Flo-2-C. Of the two European

![Figure 1. Nucleotide variation of the Flo-2 region in an African and European sample of D. melanogaster. A) The valley of reduced genetic variation in a European sample of D. melanogaster compared to an ancestral population sample from Africa (solid lines; first y-axis, \( \theta_w \)). The fully sequenced 20,011-bp fragment of the Flo-2 region is connected with an adjacent fragment of Ometto et al. [14], indicated by squares at coordinate 0 (white square-Africa; black square-Europe, distance to the fully sequenced Flo-2 region around 3,000 bp, \( \theta_w \) values were interpolated between the two fragments). The x-axis shows the relative position on the X-chromosome. The dashed horizontal lines show the average nucleotide variation of the X-chromosome for the African population sample (wide dashed line, \( \theta_{AX} = 0.0131 \)) and the European population sample (dashed line, \( \theta_{EX} = 0.0047 \)). Black triangles show the position of indels; black diamonds indicate fixed nucleotide differences between both samples. The second y-axis corresponds to divergence of D. melanogaster to D. sechellia (dashed line). The gene map gives the name and the position of the coding sequence of all genes located within the valley of reduced variation. These genes are Eo, CG9503 CG32591, pdgy (partly) and eight different transcripts of Flo-2. B) Tajima’s D statistic (solid lines) is shown for the African and European sample of D. melanogaster (white square-Africa; black square-Europe). The horizontal lines show the average value of Tajima’s D for the X chromosome of the African population sample (wide dashed line, \( D_{XA} = -0.608 \)) and the European population sample (dashed line, \( D_{XE} = -0.103 \)). Significance values are indicated as follows: \# \( P < 0.10 \); * \( P < 0.05 \); ** \( P < 0.01 \). doi:10.1371/journal.pone.0056629.g001
These candidates to characterize the detailed pattern of nucleotide variations. In this follow-up study, we used one of these candidates and found candidate loci that deviated from the neutral demographic null model. The exit out of Africa, the times of colonization (times of divergence) and the respective population sizes were estimated by means of ABC using data from the X chromosome and chromosome 3.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Chromosome X</th>
<th>Chromosome 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mode</td>
<td>Q 2.5%</td>
</tr>
<tr>
<td>Current African population size</td>
<td>4,635,114</td>
<td>2,392,436</td>
</tr>
<tr>
<td>Current European population size</td>
<td>1,586,481</td>
<td>760,318</td>
</tr>
<tr>
<td>Current Asian population size</td>
<td>338,810</td>
<td>91,417</td>
</tr>
<tr>
<td>Bottleneck size of the European population</td>
<td>22,975</td>
<td>10,898</td>
</tr>
<tr>
<td>Bottleneck size of the Asian population</td>
<td>10,741</td>
<td>3,554</td>
</tr>
<tr>
<td>Exit out of Africa</td>
<td>15,628</td>
<td>7,700</td>
</tr>
<tr>
<td>Colonization time of the South-Asian continent</td>
<td>3,264</td>
<td>1,194</td>
</tr>
<tr>
<td>Time of the expansion of the African population</td>
<td>27,643</td>
<td>2,383</td>
</tr>
<tr>
<td>Size of the ancestral African population</td>
<td>1,890,641</td>
<td>580,795</td>
</tr>
</tbody>
</table>

The parameters of the neutral demographic model that was used for the sweep analysis are given for the African, European and Asian population of D. melanogaster. The exit out of Africa, the times of colonization (times of divergence) and the respective population sizes were estimated by means of ABC using data from the X chromosome and chromosome 3.

**Discussion**

**Sequence Analysis Across the Flo-2 Region**

Ometto et al. [14] analyzed X-linked chromosome-wide patterns of DNA variation in the European population of D. melanogaster and found candidate loci that deviated from the demographic null model. In this follow-up study, we used one of these candidates to characterize the detailed pattern of nucleotide diversity and detect the possible action of positive selection.

Nucleotide diversity of the fully sequenced Flo-2 region in the European sample shows a selective sweep-like pattern: a valley of reduced genetic variation of around 10 kb in size and flanking parts that steadily increase to neutral levels of European variation. Tajima’s D statistic is strongly negative for the European population and, in some segments, significantly different from zero. The Flo-2 region is characterized by a high density of genes that are functionally and structurally well-conserved, suggesting strong evolutionary constraints. Divergence to D. sechellia remains constant at around 6% and thus excludes purifying selection and/or low mutation rates as major causes of low genetic diversity.

Furthermore, the corresponding region in the African sample agrees well with the average value for X-linked nucleotide diversity without any appreciable decline. Thus it is unlikely that the low genetic variability of the European population is a consequence of an ancestral low polymorphism.

**Selective Sweep Analysis**

We used two statistical tests for the sweep analysis of the Flo-2 region: ω and SweepFinder. Both statistics capture different aspects of the data: ω is based on LD, while the CLR of SweepFinder is dependent on the SFS. Statistical significance was inferred using neutral coalescent simulations that were based on a slightly modified version of the demographic model of Laurent et al. [30]. SweepFinder revealed a statistically significant departure from neutral expectations of the European sample, suggesting that the Flo-2 region has been a target of positive selection during the recent history of the European population. The results of the LD analysis were only marginally significant. In the following we discuss the genetic processes that have occurred in this part of the Flo-2 region and may have created novelties on which selection has possibly operated.

**European Genetic Novelties**

The European population sample has eight fixed nucleotide substitutions, all being private and derived, and the majority of which (six out of eight) are located in intronic regions of the Flo-2 gene. The results of previous studies indicate that nucleotide and indel changes in introns can affect gene expression [44,45,46]. Interestingly, the second intron of Flo-2 harbors the genes of the GMI oxidoreductase cluster (including Eo and CG9503), suggesting that they may indeed contain functional regulatory elements. Gene expression analysis of our European and African population samples revealed, however, no significant differences in the levels of Eo expression between populations for males [47], and only...
marginally for females [48]. Nonetheless, both Eo and CG9503 are highly expressed only during embryonic and metamorphic development and have tissue-specific expression [18]. Similarly, Flo-2 experiences a restricted tissue-specific expression in later developmental stages. As CG9503, Flo-2 is in fact expressed in the wing disc, where its over-expression leads to detrimental effects during the development of wings [26]. A closer examination of intergenic and intronic regions around the GMC oxidoreductase cluster and the study of the development- and tissue-specific expression patterns may ultimately reveal the existence of regulatory elements and their role in the adaptive history of Drosophila.

In D. melanogaster the Flo-2 gene has experienced a large transcript diversification, with isoforms that have distinct expression patterns. While the transcripts Flo-2-A and Flo-2-B are continuously expressed in larvae and adult flies, expression of Flo-2-C was shown to be restricted to larvae and pupae [49]. Interestingly, our results revealed that a functional Flo-2-C is shared by fewer African lines (three out of twelve) than European lines (ten out of twelve), although it is premature to associate this observation to adaptive processes. Since Flo-2-C is just one of the several functional isoforms of Flo-2, it does not suffer the genetic load typical of nonsense mRNAs. Under such relaxed conditions negative selection might be negligible and facilitates the diversifications of the Flo-2 gene. For instance, another premature termination codon was found in the Flo-2-G transcript of D. melanogaster (Figure 1, gene map) with a surprisingly high population frequency of 31.6%. Neumann-Giesen et al. [25] showed that a truncated Flotillin-2 has the ability to form homooligomers that enhance membrane association with full-length Flotillin-2. Thus, Flo-2-C might be a new gene variant whose gene product extends the function of the original Flo-2.

### Supporting Information

#### Table S1 Set of primers, which were used for DNA amplification and sequencing of the Flo-2 region.

Primer design was based on the D. melanogaster genome (FlyBase, Release 46).

#### Table S2 Prior distribution of the parameters of the neutral demographic model inferred by ABC estimation for X-linked and autosomal data.

#### Table S3 DNA sequence alignment of the upper part of the Flo-2-C transcript (146) in D. melanogaster and D. sechellia (D. sec).

African lines (A-number) show high levels of nonsense mutations (X) due to frame shifts (X) or premature stop codons (*) compared to the European lines (E-number) which mostly show intact transcripts (>). Codons (shaded in grey) specify an amino acid (below each codon, single letter code, not shaded).

#### Acknowledgments

We thank Nadia Singh and two reviewers for their helpful comments on the manuscript.

#### Author Contributions

Conceived and designed the experiments: WS. Performed the experiments: AW LO. Analyzed the data: AW PP SL. Contributed reagents/materials/analysis tools: PP SL. Wrote the paper: AW WS SL.

### References


---

The derived state is highlighted in grey.

doi:10.1371/journal.pone.0056629.t002

<table>
<thead>
<tr>
<th>nr.</th>
<th>abs. position</th>
<th>rel. position</th>
<th>location</th>
<th>Outgroup</th>
<th>Africa</th>
<th>Europe</th>
<th>consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14814106</td>
<td>6218</td>
<td>CG9503</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>Ser → Asn</td>
</tr>
<tr>
<td>2</td>
<td>14814438</td>
<td>6550</td>
<td>CG9503</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>syn. (Phe)</td>
</tr>
<tr>
<td>3</td>
<td>14815391</td>
<td>7504</td>
<td>CG9503</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>syn. (Asn)</td>
</tr>
<tr>
<td>4</td>
<td>14817203</td>
<td>9356</td>
<td>Intron</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>14817533</td>
<td>9713</td>
<td>Intron</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>14818608</td>
<td>10798</td>
<td>Intron</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>14818649</td>
<td>10839</td>
<td>Intron</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>14818898</td>
<td>11107</td>
<td>Intron</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>14821087</td>
<td>13429</td>
<td>Intron</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>14821196</td>
<td>13539</td>
<td>Intron</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>14824970</td>
<td>17399</td>
<td>Flo-2</td>
<td>C</td>
<td>C</td>
<td>A</td>
<td>syn. (Ala)</td>
</tr>
<tr>
<td>12</td>
<td>14825835</td>
<td>12986</td>
<td>Flo-2</td>
<td>–</td>
<td>–</td>
<td>Ins.</td>
<td>Additional AS: THTHTNT</td>
</tr>
</tbody>
</table>