This is the accepted manuscript of an article published as:
The evolution of novelty in conserved genes; evidence of positive selection in the Drosophila fruitless gene is localised to alternatively spliced exons
Parker, D. J., Gardiner, A., Neville, M. C., Ritchie, M. G. & Goodwin, S. F. Mar 2014 In : Heredity. 112, 3, p. 300-306

Title: The evolution of novelty in conserved genes; evidence of positive selection in the *Drosophila fruitless* gene is localised to alternatively spliced exons

Authors: Darren J. Parker[1*], Anastasia Gardiner[2], Megan C. Neville[3], Michael G. Ritchie[1] and Stephen F. Goodwin[3*]

1. Centre for Biological Diversity, School of Biology, University of St Andrews, Fife. Scotland KY16 9TH UK
2. Université de Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France
3. Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, University of Oxford, Oxford, OX1 3PT

* Joint corresponding authors

Corresponding author information:
Darren J. Parker, Centre for Biological Diversity, School of Biology, University of St Andrews, Fife. Scotland KY16 9TH UK, Email: djp39@st-andrews.ac.uk, Tel: +44 1334 463371
and
Stephen F. Goodwin, Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, University of Oxford,

22 Oxford, OX1 3PT, Email: stephen.goodwin@dpag.ox.ac.uk , Tel: +44

23 1865 272454

24 Running title: Positively selected changes in *fruitless*

25 Word count for main text: 5251

Abstract

There has been much debate concerning whether *cis*-regulatory or coding changes are more likely to produce evolutionary innovation or adaptation in gene function, but an additional complication is that some genes can dramatically diverge through alternative splicing, increasing the diversity of gene function within a locus. The *fruitless* gene is a major transcription factor with a wide range of pleiotropic functions, including a fundamental conserved role in sexual differentiation, species-specific morphology, and an important influence on male sexual behaviour. Here we examine the structure of *fruitless* in multiple species of *Drosophila*, and determine the patterns of selective constraint acting across the coding region. We found that the pattern of selection, estimated from the ratio of non-synonymous to synonymous substitutions, varied considerably across the gene, with most regions of the gene evolutionarily conserved but with several regions showing evidence of divergence as a result of positive selection. The regions which showed evidence of positive selection were found to be localised to relatively consistent regions across multiple speciation events, and are associated with alternative splicing. Alternative splicing may thus provide a route to gene diversification in key regulatory loci.

Key words: *Drosophila, fru*, gene diversity, alternative splicing, positive selection

Introduction

The nature of the genes that cause important evolutionary change is much debated (Carroll, 2005; Hoekstra and Coyne, 2007; Stern, 2000; Stern and Orgogozo, 2008; Stern and Orgogozo, 2009). Recently this has often focused on whether *cis*-regulatory or coding changes are more likely to produce evolutionary innovation or adaptation. Currently the data to test this are not conclusive either way (Hoekstra and Coyne, 2007; Stern and Orgogozo, 2008) however it does appear that *cis*-regulatory changes may be more likely to underlie differences above the species level (Stern and Orgogozo, 2008). Despite the debate, it is clear that both coding and non-coding changes can cause species differences. For example, the evolution of key odorant receptor loci may underlie ecological speciation in *Drosophila sechellia* (Matsuo et al., 2007) whereas changes in the expression of genes involved in sexually dimorphic pheromonal production may influence sexual isolation in the same species group (Shirangi et al., 2009).

The argument in favour of *cis*-regulatory changes is based primarily on the idea that changes in *cis*-regulatory regions are less likely to suffer from the negative effects of pleiotropy, due to their modular

nature (Carroll, 2005; Stern and Orgogozo, 2008). However, there are alternative genetic mechanisms that may ameliorate the constraint imposed by the pleiotropy associated with coding changes, for example neofunctionalism resulting from gene duplication (Innan and Kondrashov, 2010; Lynch et al., 2001). Another, much less well studied mechanism, is alternative splicing (Long et al., 2003). Gene duplication and alternative splicing allow gene diversification by reducing the functional constraint on a gene (Chothia et al., 2003; Graveley, 2001). Alternative splicing and gene duplication appear to be negatively correlated at a genomic level (Jin et al., 2008; Kopelman et al., 2005; Talavera et al., 2007) suggesting that gene duplication and alternative splicing may be alternative evolutionary mechanisms influencing gene diversity (Kopelman et al., 2005). Although both processes reduce the amount of functional constraint on a sequence, allowing changes in gene product and expression, the location and type of the changes involved have been found to be different. Substitutions occurring within alternatively spliced genes are both more localised (mainly in those exons being alternatively spliced) and less conservative, than those in genes that have been duplicated (Talavera et al., 2007).

fruitless (*fru*) is an alternatively-spliced transcription factor that has

been identified in a broad range of insect groups (Salvemini et al., 2010) including Orthoptera (Boerjan et al., 2011; Ustinova and Mayer, 2006), Blattodea (Clynen et al., 2011), Hymenoptera (Bertossa et al., 2009) and Diptera (Gailey et al., 2006; Ryner et al., 1996; Salvemini et al., 2013; Salvemini et al., 2009; Sobrinho and de Brito, 2010). *fru* is a pleiotropic gene with at least two major functions: one that controls male sexual behaviour and another that is essential for viability in both sexes (Fig. 1). All Fru proteins are putative transcription factors containing a common BTB (protein:protein interaction) N-terminal domain, a connector region and, through alternative splicing, one of four C-terminal Zn finger DNA binding domains (A, B, C and D). Transcripts from the most distal *fru* promoter, P1, undergo sex-specific alternative splicing and encode the male-specific Fru^M proteins that only differ from the common isoforms by the addition of 101 amino acids at the N-terminus. These male-specific putative transcription factors determine many of the neuronal substrates for sexual behaviour in the male central nervous system (CNS) (Fig. 1) (Ito et al., 1996; Ryner et al., 1996).

The high level of pleiotropy associated with *fru* suggests it should be evolutionarily conserved (Billeter et al., 2006a; Wilkins, 1995). Such

conservation was shown by the ability of the *Anopheles gambiae* ortholog of *fru* to function when ectopically expressed in *D. melanogaster* resulting in the production of the fru-dependent male-specific muscle of Lawrence (MOL) (Gailey et al., 2006). As *A. gambiae* and *D. melanogaster* have been separated for approximately 250 mya (Gaunt and Miles, 2002; Zdobnov et al., 2002), Gailey et al. (2006) concluded that *fru* has been functionally conserved across this time period. This has been further emphasised with the finding that RNAi-mediated knockdown of *fru* extinguishes male courtship in the cockroach *Blattella germanica*, suggesting that the large role *fru* plays in the production of male sexual behaviours has been conserved for at least a large portion of insect evolution (Clynen et al., 2011). Despite this, many of the courtship behaviours influenced by *fru* are known to be species-specific, and *fru* has been implicated as a potential candidate gene for species-specific divergence in QTL studies (Gleason and Ritchie, 2004; Lagisz et al., 2012). Furthermore, a recent study of the *fru* connector region using three species of fruit fly (Genus: *Anastrepha*) found evidence of positive selection based on both sequence differences and population gene frequencies, suggesting that *fru* may contribute to species-specific differences in male courtship behaviour of *Anastrepha* species (Sobrinho and de Brito, 2010).

This highlights an intriguing conundrum about the widespread use of candidate genes in evolutionary biology; important genes would be expected to be under selective constraint, yet to be important to adaptation, such genes must evolve rapidly between species. The candidate gene approach has proven very successful in numerous studies of species differences (Martin and Orgogozo, 2013), including studies of behaviour (Fitzpatrick et al., 2005). *fru* provides an example of such a gene: on the one hand *fru* is known to be a highly pleiotropic essential gene for both sexes, suggesting it should be highly conserved. On the other hand *fru* has been implicated in the production of behaviour, which is typically species-specific. One possible resolution to this is that the alternative splicing of the exons in *fru* may allow some exons to accumulate changes that alter species-specific behaviour whilst other exons are conserved to maintain their essential functions. To address this we have conducted an analysis of the *fru*-coding region from 18 species of sequenced *Drosophila*. We examine i) the pattern of sequence variability across exons of *fru* between *Drosophila* species, ii) what proportion, if any, of such variability is due to positive selection and iii) if divergently selected regions of *fru* specifically occur in the alternatively spliced exons.

MATERIALS AND METHODS

Drosophila species

The *Drosophila* genome assemblies used in this paper were downloaded from the following websites in July 2012:

1. *D. melanogaster* (v. 5.47) from FlyBase (http://flybase.org/).
2. *D. simulans, D. sechellia D. yakuba, D. erecta, D. ananassae, D. persimilis, D. pseudoobscura, D. willistoni, D. virilis, D. mojavensis* and *D. grimshawi* from http://rana.lbl.gov/drosophila/assemblies.html (CAF1, comparative analysis freeze 1). Further information on these genome assemblies is available from Drosophila 12 Genomes Consortium (2007). In addition, the B exon for *D. simulans* was not available from the CAF1 assembly due to sequence failure in this region, and so the sequence for this exon was obtained from Genbank (accession number: GI: 111258132). We also re-sequenced the C exon for *D. simulans* and *D. sechellia* (see below) as these regions were also unavailable from the CAF1 assembly.
3. *D. bipectinata, D. kikkawai, D. elegans, D. eugracilis, D. ficusphila, D. rhopaloa, D. biarmipes, and D. takahashii* from https://www.hgsc.bcm.edu/content/drosophila-modencode-

project. The sequencing was provided by Baylor College of Medicine Human Genome Sequencing Center.

Re-sequencing assembly gaps in the *fru* locus

To obtain the sequence C exon of *fru* for *D. simulans* and *D. sechellia* genomic DNA was extracted from inbred lines of *D. simulans* (*f^2;nt, pm; st, e,* kindly provided by Jerry Coyne) and *D. sechellia* (David4A, kindly provided by Jean R. David) (see Gleason and Ritchie, 2004) using the single fly prep method developed by Gloor et al. (1993). The resulting extractions were then amplified via PCR using the following primers designed from the orthologous region in *D. melanogaster*: 5'-GACGGGCTGTTGTGTGTTC-3' and 5'-CACGCCCTTAAATGGATGA-3'. The PCR products from these reactions were then Sanger sequenced using Dundee Sequencing Services (www.dnaseq.co.uk), the consensus sequences of which were then submitted to Genbank (accession numbers: KF005597 and KF005598 for *D. simulans* and *D. sechellia* respectively).

Annotation of the *fru* orthologs in *Drosophila* species

Annotation of the orthologs of *D. melanogaster fruitless* (*fru*, CG14307) gene was performed for the other *Drosophila* species using a combination of BLAST (Altschul et al., 1990) , GeneWise

(Birney et al., 2004) and manual curation. Available amino acid sequences of the proteins encoded by the *fruitless* (FlyBase, FBpp0083060–67 and FBpp00839355-59) of *D*. *melanogaster* were used as the queries in TBLASTN search of each of the other *Drosophila* species' genomic DNA in turn. The worst scoring alignments were discounted. For the remainder, the genomic DNA involved in the alignment, with flanking regions, was extracted using a simple BioPerl script (Stajich et al., 2002). Provisional gene structures were predicted automatically by realigning the *D. melanogaster* proteins and the genomic region using GeneWise. Finally, coordinates of exons in the GeneWise predictions were corrected manually. This was necessary to obtain a realistic gene structure where the protein sequence diverged from that of the *D. melanogaster* protein in the region of a start, stop or splice site, causing the GeneWise model to truncate the exon. Thus, the loci structure and protein-coding exons were identified across 18 species of *Drosophila*. The *D. persimilis* and *D. rhopaloa* genome assemblies were found to have poor coverage of the region that includes *fruitless*, so we excluded these species from our analysis. The size of *fru* orthologs was defined as the sequence from the transcription start site in promoter 1 (P1) to the end of the C exon (Fig 1, Supplementary Table 1).

Sequence analysis

The protein coding sequences of *fru* were multiply aligned using *ClustalW* (Thompson et al., 1994) on translations, followed by *Protal2dna* (K. Schuerer, C. Letondal; http://bioweb.pasteur.fr) to obtain a codon alignment for use in PAML (below). Pairwise nucleotide identity values for the codon aligned sequences were obtained using the Geneious program (version 5.6.6. available from www.geneious.com).

The M0 model of *codeml* in the PAML computer package (Yang, 1997) was used to determine overall selective constraint acting on the *fru* protein coding exons through estimation of the ratio of the normalised nonsynonymous substitution rate (d_N) to normalised synonymous substitution rate (d_S), or $\omega=d_N/d_S$. $\omega > 1$ is considered to be strong evidence of positive selection for amino acid replacements, whereas $\omega \approx 0$ indicates purifying selection (Yang and Bielawski, 2000).

The alternative splicing of *fruitless* produces a number of well-defined transcripts in *D. melanogaster* of which the following were tested for evidence of positive selection across all of the species: the set of

transcripts that consist of C1-C5 exons and one of the 3' alternatively spliced exon ends (either A (Fru-RI, FBtr0083648), B (Fru-RK, FBtr0083650) or C (Fru-RF, FBtr0083644)), the transcript that includes exon C1-C4 and exon D (Fru-RD, FBtr0083647), the C1-C5 exons alone (Fru-RA, FBtr0083646), and the three male specific *fru* transcripts which include the C1-C5 exons, sex-specific N-terminus (S) and one of the 3' alternatively spliced exon ends (either A (Fru^{MA}), B (Fru^{MB}) or C (Fru^{MC})) (Fig. 1). In addition we also tested exon S separately.

To test for evidence of positive selection on the *fru* products, we used M7 *vs*. M8 and M8a *vs*. M8 site-based model comparisons in PAML (Yang, 1997). Models M7 and M8a are null models which do not allow any sites to have $\omega > 1$. M8 has the additional parameter of a class of sites (p_1) which allow $\omega > 1$. Models are compared by a log-likelihood ratio test, LRT (LRT = -2 times the difference in log-likelihood tested against a chi-squared (X^2) distribution with the number of degrees of freedom equal to the number of additional random effects). It should be noted that the use of two degrees of freedom for the M8 vs. M7 comparisons and one degree of freedom for the M8a vs. M8 comparisons is considered conservative (Swanson et al., 2003; Wong et al., 2004).

Site-based models average the value of ω over all of the branches in the tree meaning such tests lack power if selection has been concentrated on only a few branches. One could apply branch-based or branch-site-based models of selection, which allow the value of ω to vary between linages. A problem with this method is that any such divisions must be applied *a priori* and it is unclear why we would expect selection on *fru* to differ amongst *Drosophila* linages. As a result we did not apply any branch or branch site models to our data. The tree provided to PAML for selection analyses was produced using trees from Da Lage et al. 2007 and *Drosophila* 12 Genomes Consortium 2007 (Supplemental Fig. 1).

In order to obtain a visual indication of the regions of *fru* showing the highest values of ω, pairwise comparison of the values of ω along the *fru* coding regions was conducted between *D. melanogaster* and the other sequenced melanogaster group species (*D. elegans, D. eugracilis, D. ficusphila, D. biarmipes, D. takahashii, D. yakuba, D. erecta, D. sechellia and D. simulans*) using a sliding window. The size of the window for calculating ω for comparisons using *D. elegans, D. eugracilis, D. ficusphila, D. biarmipes, D. takahashii, D. yakuba and D. erecta* was 102bp (i.e. the *fru* alignment was split into

102bp ‘windows’, from which a value of ω was calculated). Windows that did not show any synonymous changes were combined with the following window to allow calculation of ω. For comparisons using the more closely related *D. sechellia and D. simulans* a 408bp window was used because there were a large number of regions with no synonymous changes. This 408bp window was then moved by 102bp to allow the regions of *fru* with the highest values of ω to be visualised. To avoid analysing any chimeric sequences, values of ω for each of the alternatively spliced exons (S, A, B, C and D) were calculated separately before concatenation to produce figures 2 and 3.

Results

Genomic location of the *fru* locus

fruitless is located on the right arm of the third chromosome (3R) in the *D. melanogaster* genome, spanning nearly 130 kbp, in cytological position 91A7-91B3 with genes *CG31122* and *CG7691* located up and downstream of *fru* respectively. We identified single copy orthologs of *fru* in 17 other *Drosophila* species. Only *D. simulans, D. yakuba* and *D. pseudoobscura* genomes are localised to chromosomes, the remainder are only available as scaffolds. We identified the location of the *fru* locus in each species and an approximate length of the region encompassing the *fru* exons (Supplementary Table 1). In *D. simulans* and *D. yakuba*, *fru* is located on the right arm of the third chromosome (as in *D. melanogaster*), and on the second Muller element in *D. pseudoobscura* (homologous to the 3R of *D. melanogaster*) (Powell, 1997). The total length of the *fru* locus varies between species from 117 kbp in *D. bipectinata* and 167 kbp in *D. mojavensis* (Supplementary Table 1). Local synteny of genes appears to be conserved as all but one of the *fru* orthologs identified in this study are flanked by the orthologs of *CG31122* and *CG7691*. The *fru*

ortholog of *D. kikkawai* is flanked by *CG31122* but not *CG7691*. This however is unlikely to represent a change in local synteny but rather is a result of *fru* occurring near the end of the assembled scaffold.

Organisation and structure of *fru*

Common exons

Across *Drosophila* species, we identified exons C1 – C5 and reconstructed the exon-intron structure of this region. Putative splice donor and acceptor sites are in agreement with the consensus motifs (Mount et al., 1992). The exons C1, C2 and part of C3 encode for BTB/POZ domains and the remainder of C3, C4 and C5 encode for the “connector” that joins BTB and 3’ zinc finger domains. The Fru BTB domain is a highly conserved ~120 amino acid long domain, found in many other *D. melanogaster* transcription factors (Bonchuk et al., 2011; Zollman et al., 1994). Across the species we found that the C1 and C2 exons are highly conserved, with pairwise nucleotide identity of 94% and few amino acid substitutions across all species (2 sites in C1 and 1 site in C2). The nucleotide and amino acid similarity is reduced in the C3, C4 and C5 exons with pairwise nucleotide identity values of 79%, 84% and 83% respectively.

Alternative 3' ends - zinc finger domains

A schematic of alternative splicing of the *fru* exons is presented in Figure 1. There are four main alternative 3' exons: A, B, C and D. Exons A, B and C each contain two C_2H_2 zinc finger binding domains (Ito et al., 1996; Ryner et al., 1996; Usui-Aoki et al., 2000). Manual inspection of the exon D alignment identified a pair of conserved cysteine and histidine residues separated by a motif of 28 amino acids (consensus sequence: CRHC RKWSGELADIRTSFVEGNSNFRLEIVNH HNKCKSH - cysteine and histidine motifs underlined). This is a significant departure from the consensus "finger" sequences (Wolfe et al., 2000) suggesting that exon D encodes for either an atypical zinc finger domain, a non-functional domain or a domain with novel structure. The zinc finger motifs of exons A, B and C have no amino acid substitutions across all species and the proposed zinc finger motif of the D exon has only 2 amino acid sites which vary between these species. Pairwise nucleotide identity values vary for the four alternative 3' exons, with exons A and D showing less sequence conservation across species than exons B and C (pairwise nucleotide identity values for exons A, B, C and D: 62%, 82%, 76% and 71% respectively).

371 **Alterative 5' sex-specific exon**

372

373 The alternatively spliced exon S was found to be similar across
374 species with a pairwise nucleotide identity value of 77%. In addition
375 the three transformer (tra/tra2) binding sites in the S exon UTR were
376 also found to be highly conserved (pairwise nucleotide identity value
377 of sites 96.4%, 97.2% and 88.6% respectively) (see Supplemental
378 Fig. 2 and 3 for alignments).

379

380

381 **Selection analysis**

382

383 Across the whole coding region of *fru* the value of ω was 0.107,
384 implying purifying selection is acting; however the value of ω varies
385 widely across the gene. Selective constraints on the region coding for
386 BTB domain are very strong (ω^{BTB} = 0.013), while the strength of
387 purifying selection acting on the C3-C5 exons encoding the "domains
388 connector" is weaker, with an average ω = 0.064. Purifying selection
389 on 80 amino acids that include the zinc finger motifs on exons A, B,
390 C and D is very strong (ω^{ZnF-A} = 0.00184; ω^{ZnF-B} = 0.00010; ω^{ZnF-C} =
391 0.00375; ω^{ZnFD} = 0.01805) with weaker constraint acting on the rest
392 of the exon (ω^{A} = 0.219; ω^{B} = 0.077; ω^{C} = 0.186; ω^{D} = 0.145).

393 Selective constraint across the region coding for the 5' sex-
394 specifically spliced exon S was also found to be mainly purifying (ω^S
395 = 0.074).
396
397 Comparison of the nested models M7 and M8 across the whole
398 coding region of *fru* found M8 to be a significantly better fit (p =
399 0.00001) with 3.4% of sites (p_1 = 0.03414, ω = 1.46311) under
400 positive selection. The more stringent test for positive selection (the
401 comparison of the M8a and M8 models) also found M8 to be a better
402 fit (p = 0.005). Comparison of M7 and M8 found M8 to be a
403 significantly better fit for most of the known transcripts (Table 1)
404 however M8 was a better fit for only 3 of transcripts when compared
405 to M8a. These contained either exon A (Fru-RI and Fru^{MA}) or exon D
406 (Fru-RD) indicating positive selection on these regions (Table 1). For
407 those transcripts the proportion of sites under positive selection (p_1)
408 was around 4% (Fru-RI: p_1 = 0.0383, ω = 1.412; Fru^{MA}: p_1 = 0.0382,
409 ω = 1.454; Fru-RD: p_1 = 0.0357, ω = 1.683) (Table 1). Transcripts
410 containing other exons either showed the M8 model to be a better fit
411 than M7 but not M8a (exons B and C) or M8 was not a better fit than
412 M7 (C1-C5, containing only the BTB domain and the connector),
413 implying these regions are evolving neutrally or under purifying
414 selection respectively. The M8 model was also found to not be a

415 better fit than M7 for exon S (p = 0.656, Table 1) implying this exon is
416 also evolving under purifying selection.
417
418 Pairwise sliding window comparisons of *fru* across the *melanogaster*
419 group species (Figs 2 and 3) shows values of ω are elevated in
420 similar areas in each of the pairwise comparisons: around the 5' end
421 of the A exon, in line with the finding that transcripts containing the A
422 exon are under positive selection (Table 1). There is evidence for
423 saturation, because values of ω for species more distant to *D.*
424 *melanogaster* have lower peaks of ω, probably as a result of a large
425 number of synonymous changes rather than a lack of non-
426 synonymous changes (Fig. 4). The pairwise sliding window
427 comparisons however did not show peaks in the region containing
428 exon D, despite evidence for positive selection on transcripts
429 containing this exon. An explanation for this may be that the
430 positively selected changes in exon D are less localised than in exon
431 A, and that, unlike exon A, the putative zinc-finger for exon D is in the
432 middle of this exon, which may make sites of diversifying selection
433 more difficult to visualise.
434

435 **Discussion**

Divergence during speciation is thought to be driven by strong selection (Coyne and Orr, 2004; Rundle and Nosil, 2005), thus such divergence would be expected to leave a signature of an excess of non-synonymous substitutions (dN) between closely related species. However, the increasing availability of genome projects and focussed studies of gene families are finding that relatively few genes show elevated dN in genomic comparisons (*Drosophila* 12 Genomes Consortium, 2007; Ellegren et al., 2012). Relaxed selection, especially following gene duplication, is undoubtedly also important to the evolution of new gene functions and species differences. *fru* is a gene with highly pleiotropic functions, some of which are essential for viability in both sexes (Anand et al., 2001; Song et al., 2002; Song and Taylor, 2003). Previous studies have suggested that *fru* should be evolutionarily conserved (Clynen et al., 2011; Gailey et al., 2006; Salvemini et al., 2009; Wilkins, 1995), yet it has also been implicated in the production of sexually dimorphic behaviour, which is known to change rapidly between species (Kraaijeveld et al., 2011; Mendelson and Shaw, 2005). In addition *fru* has also been implicated as a potential candidate gene for the production of species-specific behaviour differences (Gleason and Ritchie, 2004; Sobrinho and de Brito, 2010). The alternative splicing of *fru* may offer a resolution of this apparent contradiction, if some exons accumulate changes that

458 alter species-specific behaviour whilst other exons remain conserved
459 to maintain their essential functions. This predicts that different
460 transcripts of the same gene should have rather different
461 evolutionary rates and show variation in the relative rate of non-
462 synonymous substitutions.

463 **Positive selection is restricted to alternatively spliced exons**

464 We found evidence of positive selection acting on a small but
465 significant number of sites in the *fru* coding region (Table 1). These
466 sites are restricted to transcripts containing alternatively spliced
467 exons A or D. In contrast, alternatively spliced exons B and C did not
468 show evidence of positive selection, and appear to be governed
469 primarily by purifying selection with a small proportion of neutrally
470 evolving sites (Table 1). The male-specific alternatively spliced exon
471 S and common coding regions of *fru* transcripts also showed no
472 evidence of positive selection and appear to be under strong
473 selective constraints.

474 These findings raise clear predictions concerning the functional
475 importance of different transcripts, which, for example, could be
476 tested by mutagenesis or selective introgression experiments. As
477 transcripts containing exons B and C were found to be conserved,
478 we hypothesise that they are responsible for the essential functions

of *fru*, whereas transcripts containing exons A and D are more likely be involved in non-essential functions, which may contribute to phenotypic differences between species. As exon D does not appear to be included in *fru* isoforms controlling male sexual behaviour (Billeter et al., 2006b), we further hypothesise that sequence variation in isoforms containing exon A, could influence species-specific differences in male sexual behaviour. We know from molecular genetic studies, that *fru* exploits these multiple isoforms through spatial and temporal expression of either a single, or a combination of isoforms enabling specific phenotypic outcomes. For instance, the production of serotonergic neurons in the central nervous system that innervate the male reproductive system depends on the expression of Fru^{MB} and Fru^{MC} isoforms and not the Fru^{MA} isoform (Billeter et al., 2006b).

Our finding of positive selection in alternatively spliced exons at the 3' end of *fru* raises the question of why no positive selection was found in alternatively spliced exon S towards the 5' end of *fru*. A potential solution is that, although exon S is alternatively spliced, it is either present or absent in *fru* transcripts (i.e. there is no alternative exon to S, isoforms vary only in the presence or absence of exon S). This means that, unlike at the 3' end of *fru*, the alternative splicing of

exon S does not provide redundancy at the 5' end of *fru*, and thus does not provide any reduction in selective constraint for this exon.

Our finding, that positively selected changes are localised to alternatively spliced exons, is in broad agreement with previous studies which have shown that typically there are a greater number of positively selected changes in alternatively spliced exons than in constitutively spliced exons (Ermakova et al., 2006; Hughes, 2011; Ramensky et al., 2008). This suggests that alternative splicing may provide a general mechanism for the evoultion of noveliy in otherwise conserved genes. In contrast, a previous study looking at the patterns of selection on *fru* in *Anastrepha* fruit flies (Sobrinho and de Brito, 2010) found evidence for positive selection on constitutively spliced exon C3. We did not find evidence of positive selection in this region, however it is not known if positive selection also occurs in the alternatively spliced regions of *Anastrepha fru* as these regions are not currently available for study, making direct comparisons with our study difficult.

Positive selection on alternatively spliced exons presumably arises due to changes in protein structure. However, splicing regulation occurs via changes in exonic splicing regulators (ESRs), which are presumably themselves under selection. ESRs are typically short

sequences (usually hexmers) within coding regions which enhance or suppress splicing. As ESR motifs are regulatory in function, functional changes will not necessarily be detected by dN/dS style analyses. Selected changes in ESRs should not favour nonsynonymous changes over synonymous changes (synonymous changes, in fact, should be more likely to avoid potentially deleterious changes in protein sequence). This combined with the fact that ESRs are typically quite short, means that selection for changes in splicing regulation via ESRs is unlikely to be found by this analysis, so the evidence for positive selection found in this study is more likely to reflect selection for changes in the protein sequence.

How could the positive selection detected in some transcripts of *fru* act to alter traits, including distinct behaviours? Because *fru* is a transcription factor, sequence changes could either cause change in the target loci it binds to, or it could alter the expression of a similar suite of downstream loci. Our data perhaps suggest that the latter is more likely; the zinc finger motifs of all the 3' alternatively spliced exons (A, B, C and D) are highly conserved. This suggests that the positive selection detected is unlikely to be changing the sites the transcription factor binds to between species. As transcription factors typically interact with several proteins whilst binding DNA, changes to

544 the amino acid sequence outside the zinc finger may affect the
545 efficiency with which the transcription factor is able to bind to the
546 target DNA and/or influence the way the transcription factor interacts
547 with other proteins (Locker, 2001). As such, the changes in exon A
548 and D may influence the regulation of downstream genes to which
549 the zinc-finger binds. Currently the genes directly regulated by *fru* are
550 unknown (Villella and Hall, 2008), however as *fru* is known to be a
551 major gene in the sex determination cascade, the changes in *fru*
552 found by this study may influence the expression of a large number
553 of downstream targets (Baker et al., 2007).

554 Due to *fru's* position in the sex determination pathway and the role it
555 plays in the shaping of male sexual behaviour, these results suggest
556 that *fru* may be acting as a 'hotspot gene' for the evolution of male
557 sexual traits. Hotspot genes are those genes which are able to incur
558 a disproportionate number of evolutionary important mutations for a
559 trait: mutations which cause a large enough phenotypic change for
560 selection to act upon and that are able to be positive selected due to
561 limited negative pleiotropy (Martin and Orgogozo, 2013; Stern, 2000;
562 Stern and Orgogozo, 2009). Stern and Orgogozo (2009) suggest that
563 such hotspot genes will contribute disproportionally to the evolution
564 of differences between species. Of course, numerous high resolution
565 QTL studies of species differences will be required to assess the

likelihood of a disproportionate role of individual loci in species differences. Stern and Orgogozo (2009) also suggest that regions of a gene which experience less pleiotropy would be more likely to accumulate evolutionary relevant mutations. They suggested this in the context of *cis*-regulatory vs. coding mutations whereby cis-regulatory mutations would be more likely to accumulate changes (Carroll, 2005; Hoekstra and Coyne, 2007; Stern, 2000; Stern and Orgogozo, 2008; Stern and Orgogozo, 2009). The same might be true for alternatively spliced regions, likely to experience less pleiotropy than common coding regions due to the functional redundancy the production of alternative transcripts provides. Our findings are consistent with this: we found that positively selected changes in *fru* had accumulated in two of the alternatively spliced exons, showing that alternative splicing may impact a gene's ability to accumulate evolutionary relevant mutations. In many ways this is similar to the role of neofunctionalisation of recent duplicate loci in the generation of evolutionary novelty (Lynch and Conery, 2000). The widespread incidence of alternative splicing in plasticity, gene function and adaptation is starting to be understood, but how this will contribute to adaptive divergence and ultimately speciation is only beginning to be explored (Ast, 2004; Harr and Turner, 2010).

Acknowledgements

588 We thank Andy James and four reviewers for helpful advice on this
589 study. This work was funded by the NERC, UK (studentship to DJP
590 and grants to MGR & SFG).

Titles and legends to figures

Figure 1. The structure and splicing pattern of the *fruitless* gene in *D. melanogaster*. P1 promoter mRNA transcripts are sex-specifically spliced at the 5' end, resulting in the inclusion of the S exon and the addition of 101 amino acids (yellow) to male specific isoforms (Fru^M) and the inclusion of a premature stop codon in females (UAA). Alternative splicing at the 3' end of transcripts produced from the sex-specific P1 promoter and non-sex-specific P2-4 promoters results in the inclusion of alternative DNA-binding domains A (purple), B (orange), C (green) or D (brown). All isoforms contain the BTB domain (blue) and connector region (grey). Common exons C1-5 are included in $fru^{A/B/C}$ isoforms whereas the fru^D isoform includes exons C1-C4. Untranslated regions (UTRs) are shown in white and translation start codons are indicted (ATG)

Fig. 2 Values of dN/dS (ω) between *D. melanogaster* and *D. simulans*, *D. sechellia*, *D. erecta*, and *D. yakuba* across the coding region of *fruitless.* Values for each point represent the average dN/dS value for either a 102bp window for *D. erecta* and *D. yakuba* or a 408bp window for *D. sechellia and D. simulans*.

611 Fig. 3 Values of dN/dS (ω) between *D. melanogaster* and *D.*
612 *takahashi*, *D. biarmipes*, *D. eugracilis*, *D. fisusphila* and *D. elegans*
613 across the coding region of *fruitless.* Values for each point represent
614 the average dN/dS value in a 102bp window.
615

616 Fig 4. Values of dN and dS for melanogaster group species from
617 pairwise comparisons to *D. melanogaster.*
618

Transcripts/Exons	$2*(\ln_{M7}-\ln_{M8})$	*P*, d.f.=2	$2*(\ln_{M8A}-\ln_{M8})$	*P*, d.f.=1	p_1 M8	ω_1 M8
S exon/ S	0.84	0.655	0.000	1.000	-	-
Fru-RA/ C1-C5	1.11	0.757	0.000	1.000	-	-
Fru-RI/ C1-C5+A	11.86	**0.003**	4.149	**0.042**	0.038	1.412
Fru-RK/ C1-C5+B	7.47	**0.024**	0.112	0.946	0.010	1.126
Fru-RF/ C1-C5+C	14.23	**0.001**	0.080	0.778	0.023	1.068
Fru-RD/ C1-C4+D	47.84	**0.000**	9.303	**0.002**	0.036	1.683
FruMA/ S+C1-C5+A	14.19	**0.001**	5.193	**0.023**	0.038	1.454
FruMB/ S+C1-C5+B	5.52	0.063	0.000	1.000	0.013	1.000
FruMC/ S+C1-C5+C	12.17	**0.002**	0.010	0.921	0.026	1.026

Table 1. The results of the tests for positive selection on the *fruitless* transcripts.

$2*(\ln_{M7}-\ln_{M8})$ and $2*(\ln_{M8A}-\ln_{M8})$ are twice the difference of log likelihood between two models that was compared to the χ^2 distribution with the given degree of freedom. The exact *P* values and degree of freedom are shown. The p_1 is the proportion of positively selected sites with ω_1, calculated applying the M8 model.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Anand, A., A. Villella, L.C. Ryner, T. Carlo, S.F. Goodwin, H.J. Song, D.A. Gailey, A. Morales, J.C. Hall, B.S. Baker, and B.J. Taylor. 2001. Molecular genetic dissection of the sex-specific and vital functions of the *Drosophila melanogaster* sex determination gene *fruitless*. *Genetics*. **158**:1569-1595.

Ast, G. 2004. How did alternative splicing evolve? *Nat. Rev. Genet.* **5**:773-782.

Baker, D.A., L.A. Meadows, J. Wang, J.A. Dow, and S. Russell. 2007. Variable sexually dimorphic gene expression in laboratory strains of *Drosophila melanogaster*. *BMC Genomics*. **8**:454.

Bertossa, R.C., L. van de Zande, and L.W. Beukeboom. 2009. The *fruitless* gene in *Nasonia* displays complex sex-specific splicing and contains new zinc finger domains. *Mol. Biol. Evol.* **26**:1557-1569.

Billeter, J.C., E.J. Rideout, A.J. Dornan, and S.F. Goodwin. 2006a. Control of male sexual behavior in *Drosophila* by the sex determination pathway. *Current Biology*. **16**:R766-R776.

650 Billeter, J.C., A. Villella, J.B. Allendorfer, A.J. Dornan, M. Richardson,
651 D.A. Gailey, and S.F. Goodwin. 2006b. Isoform-specific
652 control of male neuronal differentiation and behavior in
653 *Drosophila* by the *fruitless* gene. *Current Biology*. **16**:1063-
654 1076.
655 Birney, E., M. Clamp, and R. Durbin. 2004. GeneWise and
656 genomewise. *Genome Research*. **14**:988-995.
657 Boerjan, B., J. Tobback, A. De Loof, L. Schoofs, and R. Huybrechts.
658 2011. *fruitless* RNAi knockdown in males interferes with
659 copulation success in *Schistocerca gregaria*. *Insect Biochem.*
660 *Mol. Biol.* **41**:340-347.
661 Bonchuk, A., S. Denisov, P. Georgiev, and O. Maksimenko. 2011.
662 *Drosophila* BTB/POZ domains of "ttk group" can form
663 multimers and selectively interact with each other. *J. Mol. Biol.*
664 **412**:423-436.
665 Carroll, S.B. 2005. Evolution at two levels: On genes and form.
666 *PLoS. Biol.* **3**:1159-1166.
667 Chothia, C., J. Gough, C. Vogel, and S.A. Teichmann. 2003.
668 Evolution of the protein repertoire. *Science*. **300**:1701-1703.
669 Clynen, E., L. Ciudad, X. Belles, and M.D. Piulachs. 2011.
670 Conservation of *fruitless*' role as master regulator of male

courtship behaviour from cockroaches to flies. *Dev. Genes Evol.* **221**:43-48.

Coyne, J.A., and H.A. Orr. 2004. Speciation Sinauer, Sunderland, MA, USA.

Da Lage, J.L., G.J. Kergoat, F. Maczkowiak, J.F. Silvain, M.L. Cariou, and D. Lachaise. 2007. A phylogeny of Drosophilidae using the *Amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. *Journal of Zoological Systematics and Evolutionary Research*. **45**:47-63.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. **450**:203-218.

Ellegren, H., L. Smeds, R. Burri, P.I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J.B.W. Wolf. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. **491**:756-760.

Ermakova, E.O., R.N. Nurtdinov, and M.S. Gelfand. 2006. Fast rate of evolution in alternatively spliced coding regions of mammalian genes. *BMC Genomics*. **7**:84.

Fitzpatrick, M.J., Y. Ben-Shahar, H.M. Smid, L.E.M. Vet, G.E. Robinson, and M.B. Sokolowski. 2005. Candidate genes for behavioural ecology. *Trends Ecol. Evol.* **20**:96-104.

693 Gailey, D.A., J.C. Billeter, J.H. Liu, F. Bauzon, J.B. Allendorfer, and
694 S.F. Goodwin. 2006. Functional conservation of the *fruitless*
695 male sex-determination gene across 250 Myr of insect
696 evolution. *Mol. Biol. Evol.* **23**:633-643.
697 Gaunt, M.W., and M.A. Miles. 2002. An insect molecular clock dates
698 the origin of the insects and accords with palaeontological and
699 biogeographic landmarks. *Mol. Biol. Evol.* **19**:748-761.
700 Gleason, J.M., and M.G. Ritchie. 2004. Do quantitative trait loci
701 (QTL) for a courtship song difference between *Drosophila*
702 *simulans* and *D. sechellia* coincide with candidate genes and
703 intraspecific QTL? *Genetics*. **166**:1303-1311.
704 Gloor, G.B., C.R. Preston, D.M. Johnsonschlitz, N.A. Nassif, R.W.
705 Phillis, W.K. Benz, H.M. Robertson, and W.R. Engels. 1993.
706 Type-I repressors of P-element mobility. *Genetics*. **135**:81-95.
707 Graveley, B.R. 2001. Alternative splicing: increasing diversity in the
708 proteomic world. *Trends in Genetics*. **17**:100-107.
709 Harr, B., and L.M. Turner. 2010. Genome-wide analysis of alternative
710 splicing evolution among *Mus* subspecies. *Mol. Ecol.* **19**:228-
711 239.
712 Hoekstra, H.E., and J.A. Coyne. 2007. The locus of evolution: Evo
713 devo and the genetics of adaptation. *Evolution*. **61**:995-1016.

714 Hughes, A.L. 2011. Runaway evolution of the male-specific exon of
715 the doublesex gene in Diptera. *Gene*. **472**:1-6.
716 Innan, H., and F. Kondrashov. 2010. The evolution of gene
717 duplications: classifying and distinguishing between models.
718 *Nat. Rev. Genet.* **11**:97-108.
719 Ito, H., K. Fujitani, K. Usui, K. ShimizuNishikawa, S. Tanaka, and D.
720 Yamamoto. 1996. Sexual orientation in *Drosophila* is altered
721 by the satori mutation in the sex-determination gene *fruitless*
722 that encodes a zinc finger protein with a BTB domain. *Proc.*
723 *Natl. Acad. Sci. U. S. A.* **93**:9687-9692.
724 Jin, L., K. Kryukov, J.C. Clemente, T. Komiyama, Y. Suzuki, T.
725 Imanishi, K. Ikeo, and T. Gojobori. 2008. The evolutionary
726 relationship between gene duplication and alternative splicing.
727 *Gene*. **427**:19-31.
728 Kopelman, N.M., D. Lancet, and I. Yanai. 2005. Alternative splicing
729 and gene duplication are inversely correlated evolutionary
730 mechanisms. *Nature Genet.* **37**:588-589.
731 Kraaijeveld, K., F.J.L. Kraaijeveld-Smit, and M.E. Maan. 2011.
732 Sexual selection and speciation: the comparative evidence
733 revisited. *Biological Reviews*. **86**:367-377.
734 Lagisz, M., S.Y. Wen, J. Routtu, K. Klappert, D. Mazzi, R. Morales-
735 Hojas, M.A. Schafer, J. Vieira, A. Hoikkala, M.G. Ritchie, and

R.K. Butlin. 2012. Two distinct genomic regions, harbouring the *period* and *fruitless* genes, affect male courtship song in *Drosophila montana*. *Heredity*. **108**:602–608.

Locker, J. 2001. Transcription factors. Academic Press, San Diego, California.

Long, M., E. Betran, K. Thornton, and W. Wang. 2003. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**:865-875.

Lynch, M., and J.S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science*. **290**:1151-1155.

Lynch, M., M. O'Hely, B. Walsh, and A. Force. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics*. **159**:1789-1804.

Martin, A., and V. Orgogozo. 2013. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution; international journal of organic evolution*. **67**:1235-1250.

Matsuo, T., S. Sugaya, J. Yasukawa, T. Aigaki, and Y. Fuyama. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol*. **5**:e118.

Mendelson, T.C., and K.L. Shaw. 2005. Sexual behaviour: Rapid speciation in an arthropod. *Nature*. **433**:375-376.

758 Mount, S.M., C. Burks, G. Hertz, G.D. Stormo, O. White, and C.
759 Fields. 1992. Splicing signals in *Drosophila* - intron size,
760 information-content, and consensus sequences. *Nucleic Acids*
761 *Res.* **20**:4255-4262.
762 Powell, J.R. 1997. Progress and prospects in evolutionary biology:
763 the *Drosophila* model. Oxford University Press.
764 Ramensky, V.E., R.N. Nurtdinov, A.D. Neverov, A.A. Mironov, and
765 M.S. Gelfand. 2008. Positive selection in alternatively spliced
766 exons of human genes. *American Journal of Human Genetics*.
767 **83**:94-98.
768 Rundle, H.D., and P. Nosil. 2005. Ecological speciation. *Ecol. Lett.*
769 **8**:336-352.
770 Ryner, L.C., S.F. Goodwin, D.H. Castrillon, A. Anand, A. Villella, B.S.
771 Baker, J.C. Hall, B.J. Taylor, and S.A. Wasserman. 1996.
772 Control of male sexual behavior and sexual orientation in
773 *Drosophila* by the *fruitless* gene. *Cell*. **87**:1079-1089.
774 Salvemini, M., R. D'Amato, V. Petrella, S. Aceto, D. Nimmo, M.
775 Neira, L. Alphey, L.C. Polito, and G. Saccone. 2013. The
776 orthologue of the fruitfly sex behaviour gene *fruitless* in the
777 mosquito *Aedes aegypti*: evolution of genomic organisation
778 and alternative splicing. *PLoS One*. **8**: e48554.

779 Salvemini, M., C. Polito, and G. Saccone. 2010. *fruitless* alternative
780 splicing and sex behaviour in insects: an ancient and
781 unforgettable love story? *J. Genet.* **89**:287-299.
782 Salvemini, M., M. Robertson, B. Aronson, P. Atkinson, L.C. Polito,
783 and G. Saccone. 2009. *Ceratitis capitata transformer-2* gene
784 is required to establish and maintain the autoregulation of
785 *Cctra*, the master gene for female sex determination.
786 *International Journal of Developmental Biology*. **53**:109-120.
787 Shirangi, T.R., H.D. Dufour, T.M. Williams, and S.B. Carroll. 2009.
788 Rapid evolution of sex pheromone-producing enzyme
789 expression in *Drosophila*. *PLoS Biol*. **7**:e1000168.
790 Sobrinho, I., and R. de Brito. 2010. Evidence for positive selection in
791 the gene *fruitless* in *Anastrepha* fruit flies. *BMC Evolutionary*
792 *Biology*. **10**:293.
793 Song, H.J., J.C. Billeter, E. Reynaud, T. Carlo, E.P. Spana, N.
794 Perrimon, S.F. Goodwin, B.S. Baker, and B.J. Taylor. 2002.
795 The *fruitless* gene is required for the proper formation of
796 axonal tracts in the embryonic central nervous system of
797 *Drosophila*. *Genetics*. **162**:1703-1724.
798 Song, H.J., and B.J. Taylor. 2003. *fruitless* gene is required to
799 maintain neuronal identity in evenskipped-expressing neurons

in the embryonic CNS of *Drosophila*. *Journal of Neurobiology*. **55**:115-133.

Stajich, J.E., D. Block, K. Boulez, S.E. Brenner, S.A. Chervitz, C. Dagdigian, G. Fuellen, J.G.R. Gilbert, I. Korf, H. Lapp, H. Lehvaslaiho, C. Matsalla, C.J. Mungall, B.I. Osborne, M.R. Pocock, P. Schattner, M. Senger, L.D. Stein, E. Stupka, M.D. Wilkinson, and E. Birney. 2002. The bioperl toolkit: Perl modules for the life sciences. *Genome Research*. **12**:1611-1618.

Stern, D.L. 2000. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution*. **54**:1079-1091.

Stern, D.L., and V. Orgogozo. 2008. The loci of evolution: How predictable is genetic evolution ? *Evolution*. **62**:2155-2177.

Stern, D.L., and V. Orgogozo. 2009. Is genetic evolution predictable? *Science*. **323**:746-751.

Swanson, W.J., R. Nielsen, and Q.F. Yang. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**:18-20.

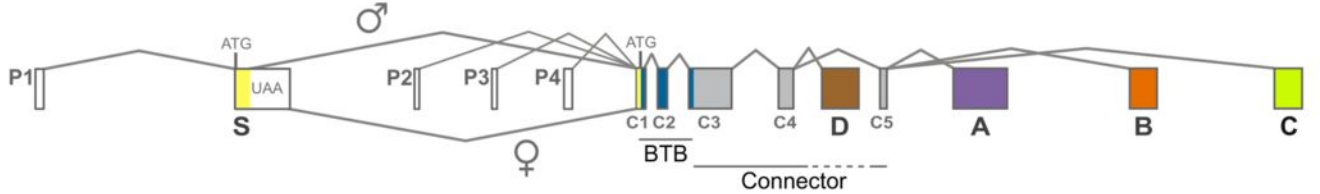
Talavera, D., C. Vogel, M. Orozco, S.A. Teichmann, and X. de la Cruz. 2007. The (In) dependence of alternative splicing and gene duplication. *PLoS Comput. Biol.* **3**:375-388.

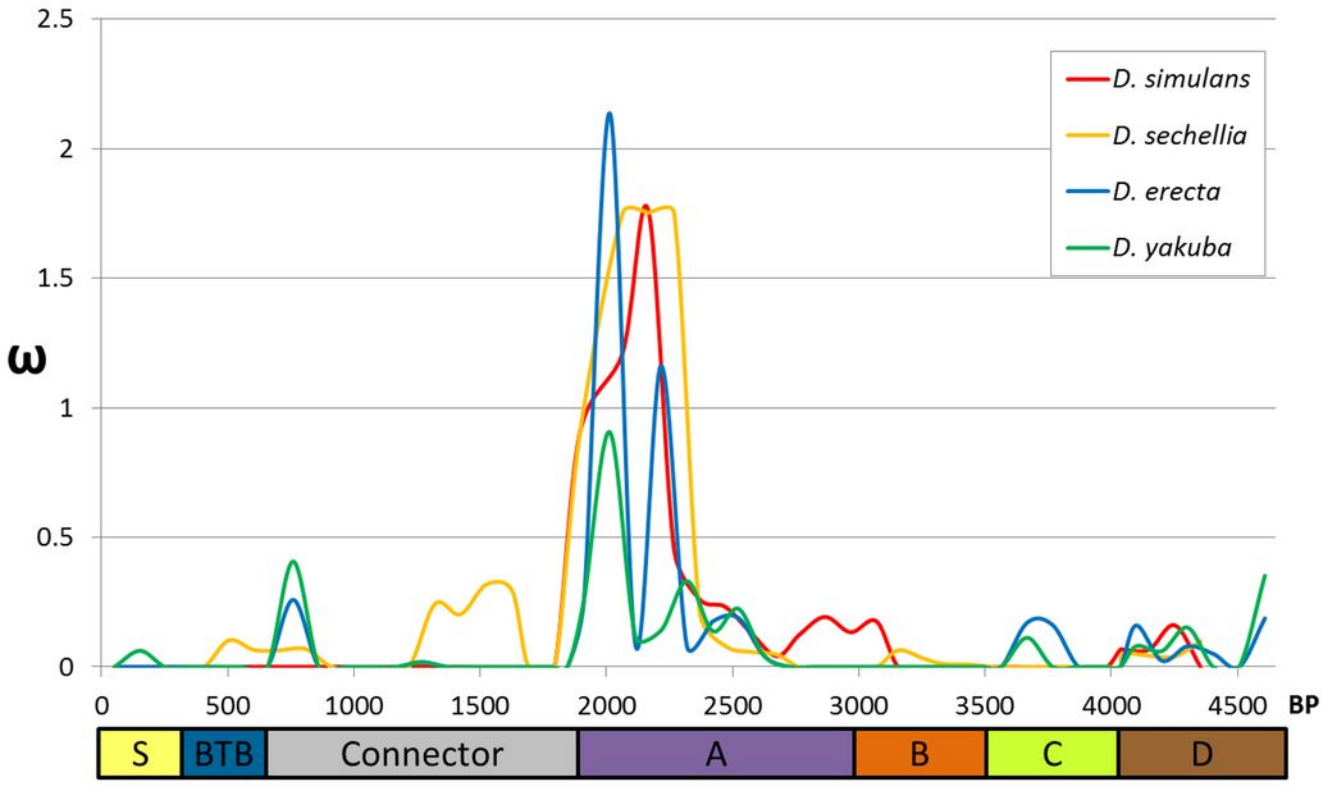
821 Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. ClustalW:
822 Improving the sensitivity of progressive multiple sequence
823 alignment through sequence weighting, position-specific gap
824 penalties and weight matrix choice. *Nucleic Acids Res.*
825 **22**:4673-4680.
826 Ustinova, J., and F. Mayer. 2006. Alternative starts of transcription,
827 several paralogues, and almost-fixed interspecific differences
828 of the gene *fruitless* in a hemimetabolous insect. *J. Mol. Evol.*
829 **63**:788-800.
830 Usui-Aoki, K., H. Ito, K. Ui-Tei, K. Takahashi, T. Lukacsovich, W.
831 Awano, H. Nakata, Z.F. Piao, E.E. Nilsson, J. Tomida, and D.
832 Yamamoto. 2000. Formation of the male-specific muscle in
833 female *Drosophila* by ectopic *fruitless* expression. *Nature Cell*
834 *Biology*. **2**:500-506.
835 Villella, A., and J.C. Hall. 2008. Neurogenetics of Courtship and
836 Mating in *Drosophila*. *In* Advances in Genetics, Vol 62. Vol.
837 62. J.C. Hall, editor. 67-184.
838 Wilkins, A.S. 1995. Moving up the hierarchy - A hypothesis on the
839 evolution of a genetic sex determination pathway. *Bioessays*.
840 **17**:71-77.

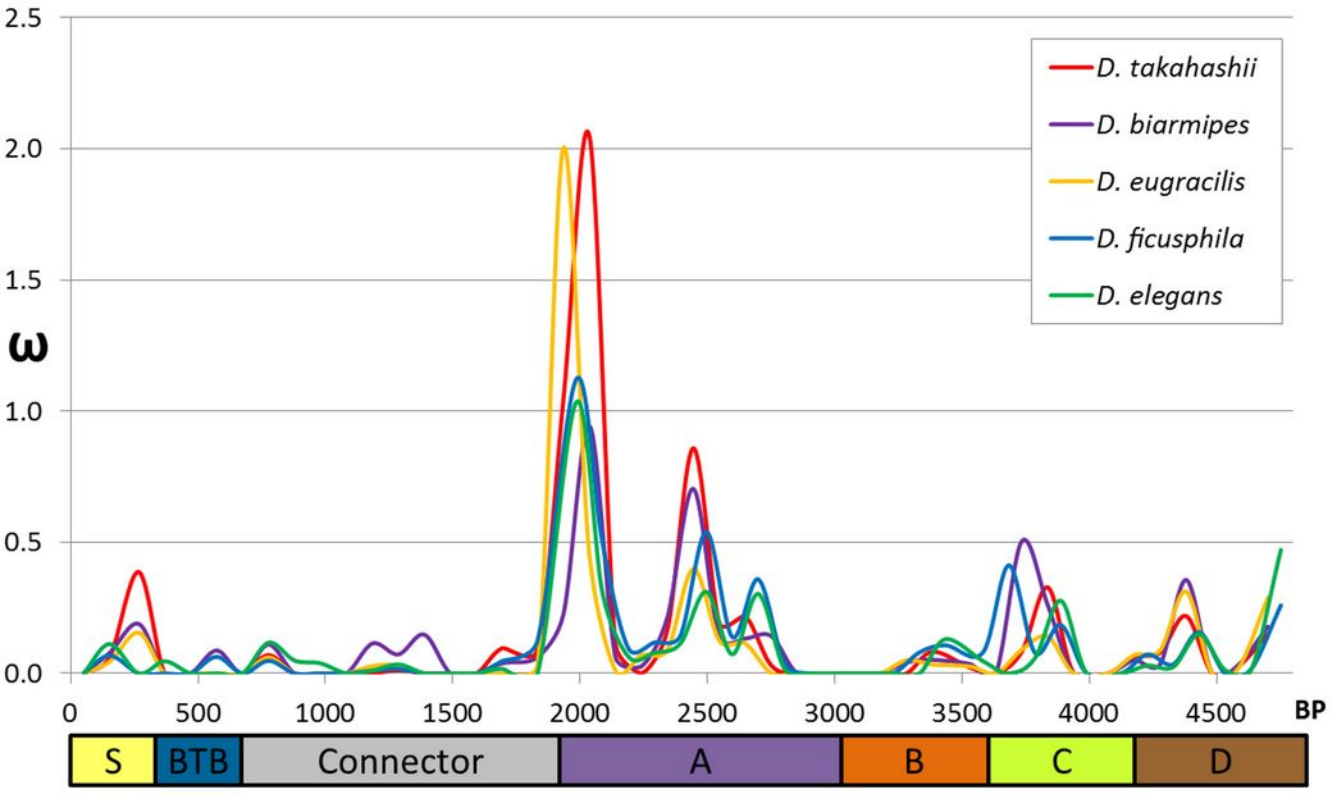
841 Wolfe, S.A., L. Nekludova, and C.O. Pabo. 2000. DNA recognition by
842 Cys(2)His(2) zinc finger proteins. *Annu. Rev. Biophys.*
843 *Biomolec. Struct.* **29**:183-212.
844 Wong, W.S.W., Z.H. Yang, N. Goldman, and R. Nielsen. 2004.
845 Accuracy and power of statistical methods for detecting
846 adaptive evolution in protein coding sequences and for
847 identifying positively selected sites. *Genetics*. **168**:1041-1051.
848 Yang, Z.H. 1997. PAML: a program package for phylogenetic
849 analysis by maximum likelihood. *Comput. Appl. Biosci.*
850 **13**:555-556.
851 Yang, Z.H., and J.P. Bielawski. 2000. Statistical methods for
852 detecting molecular adaptation. *Trends Ecol. Evol.* **15**:496-
853 503.
854 Zdobnov, E.M., C. von Mering, I. Letunic, D. Torrents, M. Suyama,
855 R.R. Copley, G.K. Christophides, D. Thomasova, R.A. Holt,
856 G.M. Subramanian, H.M. Mueller, G. Dimopoulos, J.H. Law,
857 M.A. Wells, E. Birney, R. Charlab, A.L. Halpern, E. Kokoza,
858 C.L. Kraft, Z.W. Lai, S. Lewis, C. Louis, C. Barillas-Mury, D.
859 Nusskern, G.M. Rubin, S.L. Salzberg, G.G. Sutton, P. Topalis,
860 R. Wides, P. Wincker, M. Yandell, F.H. Collins, J. Ribeiro,
861 W.M. Gelbart, F.C. Kafatos, and P. Bork. 2002. Comparative

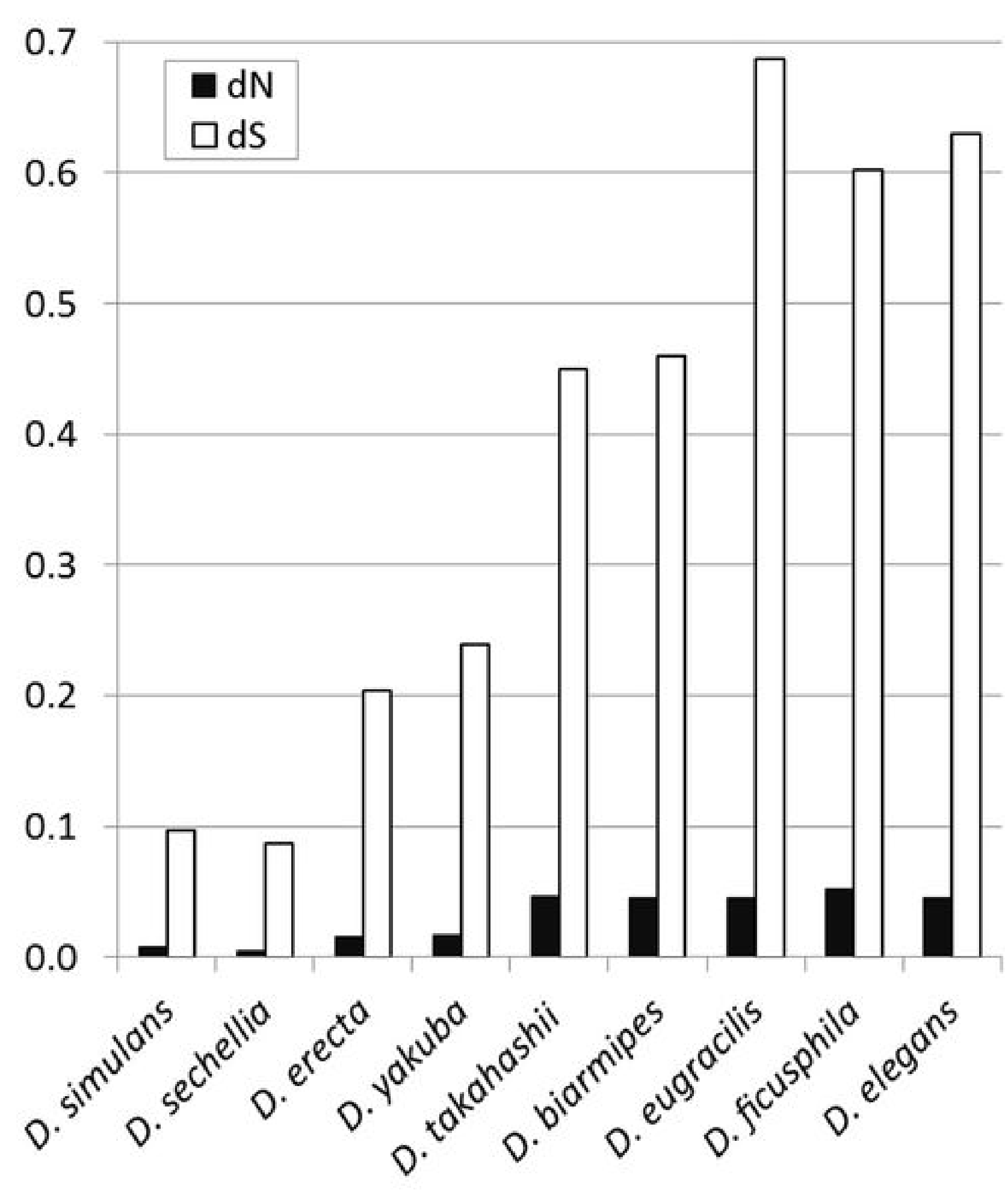
genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science*. **298**:149-159.

Zollman, S., D. Godt, G.G. Prive, J.L. Couderc, and F.A. Laski. 1994. The BTB domain, found primarily in zinc-finger proteins, defines an evolutionarily conserved family that includes several developmentally-regulated genes in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **91**:10717-10721.


♂
ATG
P1
UAA
S
P2
P3
P4
♀
ATG
C1 C2
C3
C4
D
C5
A
B
C
BTB
Connector


2.5
2
1.5
ω
1
0.5
0
D. simulans
D. sechellia
D. erecta
D. yakuba
0
500
1000
1500
2000
2500
3000
3500
4000
4500
BP
S
BTB
Connector
A
B
C
D


2.5
2.0
1.5
ω
1.0
0.5
0.0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
BP
D. takahashii
D. biarmipes
D. eugracilis
D. ficusphila
D. elegans
S
BTB
Connector
A
B
C
D


0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
dN
dS
D. simulans
D. sechellia
D. erecta
D. yakuba
D. takahashii
D. biarmipes
D. eugracilis
D. ficusphila
D. elegans