Davis, Terence. 2002. Comparison of the sexual behaviour gene *fruitless* between *D. melanogaster* and two sympatric Hawaiian species, *D. heteroneura* and *D. silvestris. Dros. Inf. Serv.* 85: 25-28.

Comparison of the sexual behaviour gene *fruitless* between *D. melanogaster* and two sympatric Hawaiian species, *D. heteroneura* and *D. silvestris.*

Davis, Terence. Department of Pathology, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN, Wales. e.mail davist2@cardiff.ac.uk

The Hawaiian *Drosophila* complex probably consists of more than 1,000 species (Kaneshiro, 2000), including some of the most morphologically diverse species in the genus, yet current evidence suggests they arose from one or two introductions to the Hawaiian chain roughly 25-40 million years ago (Kambysellis *et al*., 1995; Carson, 1997). An example of recent speciation is the split between the sympatric species *D. heteroneura* and *D. silvestris*. Both these species are in the *planitibia* species group, are found only on the Island of Hawaii, and are believed to have diverged from a common ancestor within the last 400,000 years, the estimated age of the island of Hawaii (Carson, 1970). This would entail a single founder event from an older island. An alternative theory is that these species diverged from different parental species of the *planitibia* group before migrating to Hawaii, a situation requiring two migration events. In either case the common ancestor(s) presumably migrated from one of the neighbouring Islands to the northwest. The species with the most recent common ancestor(s) are most probably *D. planitibia* from Maui or *D. differens* from Molokai. For more information on the possible origins of *D. heteroneura* and *D. silvestris* see Carson (1970, 1997), Ahearn *et al*. (1974), and Kaneshiro (2000). Kaneshiro (1976) has proposed a role for sexual behaviour and sexual selection in the evolution of the Hawaiian *Drosophila*, particularly the role of female mate choice (Ohta, 1978; Kaneshiro, 1980). Many of the Hawaiian species have evolved elaborate sexually dimorphic characteristics, *e.g*., the broadened head observed in *D. heteroneura*: females of this species are believed to prefer males with wide heads (Boake *et al*., 1997).

Table 1: Comparison of intron-exon sizes (bp) between the three species.

	exon	*D. heteroneura*		*D.silvestris*		*D. melanogaster*[b]	
#[c]	type[a]	exon	intron	exon	intron	exon	intron
III	c	148	147	148	146	148	72
IV	c	201	173	201	173	201	307
V	c	816	2310	816	>2100	828	1416
VI	c	273	675	273	663	255	529
VII	C	875	161	887	159	802	268
VIII	c	133	3708[d]	133	>3600[d]	133	3665[d]
VIIIa	D	141[e]		141[e]		141[e]	
IX	A	1435	≥3085	1435	NK[g]	1224	2760
X	E	548[e]	NK[g]	NK[g]	NK[g]	1621	4261
XI	B	561[f]		NK[g]		1824	

[a]Type: c is an exon common to most or all transcipts, A, B, C, D, E are 3' terminal exons
[b]From sequence AE003722
[c]Exon enumeration according to Davis and Ito, 2001.
[d]Calculated using exon VIII (not VIIIa)
[e]To stop codon
[f]cDNA sequence
[g]not known

As part of a study into the evolution of sexual behaviour in the Hawaiian *Drosophila*, I have extensively sequenced the sexual behaviour gene *fruitless* from *D. heteroneura* and *D. silvestris*. In this paper I have compared the genomic sequences and the structure of the *fruitless* gene in the Hawaiian species with the complete *D. melano-*

gaster sequence (GenBank Accession AE003722).

A total of 21,524bp of the *fru* locus from *D. heteroneura* have been sequenced in three pieces. The largest contig is 17,532bp in length (AF051662) and includes exons III to IX as enumerated by Davis and Ito (2001). Exon X is included in 3,431bp of adjacent genomic sequence (AF051664) and exon XI is a cDNA sequence of 561bp (AF051669). The proteins encoded by these sequences begin at the BTB domain and include the type A, B, and C protein types (Davis *et al*., 2000a). In addition protein types D and E have been deduced by comparison with the *D. melanogaster* sequence (Table 1); however, these have not been found as cDNAs. For the definition of the various protein types see Usui-Aoki *et al*., (2000) and Davis and Ito (2001). In *D. silvestris* 8,030bp of this locus have been sequenced in three pieces (Davis *et al*., 2000b). The first piece of 2,404bp includes exons III to V (AF051665), the

Table 2. Nucleotide changes in the *fru* gene between *D. heteroneura* and *D. silvestris*

[a]nucleotide position					
No.	change	No.	change	No	change
4617	C>T	8177	C>G	10626	ins TATA
4788	A>G(E)	8181	del T	10724	del TA
4926	del G	8195	del AAATGC	10739	C>A
4929	GG>TT	8269	C>G	10817	ins T
5145	C>T(S)	8414	T>C	10968	del T
5154	G>A	8423	TA>GG	11166	ins T
5156	G>T	8428	del AT	11170	C>G
5192	C>A	8432	A>G	11171	G>A
5203	C>G	8435	del TT		
5264	del GTT	8588	G>A(G)	13388	CT>GC
5286	G>T	8708	T>C(S>P)	13410	C>G
5288	del ATAGTA	8753	T>C	13415	T>G
5313	T>C	8806	C>T	13421	ins T
5352	G>C(V>L)	8936	ins T	13447	C>G
5417	C>T(G)	8947	TT>CC	13452	C>G
5534	T>C(N)	9028	ins A	13474	T>G
5558	ins CGCCGC (ins AA)	9155	A>G	13502	T>C
5891	del CAACAA (del NN)	9160	A>C	13505	C>T
6078	T>C(S>P)	9238	A>T	13512	T>C
6213	A>T	9239	del AAA	13452	C>G
6236	ins ATAT	9247	del TGA_9	13749	del $ACAACA_{10}$
6239	C>T	9442	T>C(V>A)	13998	del A
6324	C>T	9603	Ins $CACCAG(CAA)_2$	14005	ins A
6402	A>G		(ins HQQQ)	14167	del A
6431	G>A	9620	G>A	14181	ins AA
6443	C>T	9623	G>T(Q>H)	14211	G>A
6523	del TT	9956	A>C	14216	T>A
6572	G>A	9986	T>A(N>K)	14369	G>A(G>D)
6616	del CATT	10081	ins CC	14457	T>C(G)
6633	ins TTAGTAAAACTATA	10107	C>G	14892	C>T(S)
	ATCAACTGAGTAATGC	10112	del AA	15069	T>G(R)
6661	A>G	10288	C>T	15408	CC>TT
6693	T>G	10322	del GC	15449	T>A
6714	del T	10324	C>T	15833	AA>TT
		10328	T>C		
8150	del G	10335	del CTCT		

[a]nucleotide position refers to the position in the *D. heteroneura* sequence Accession number AF051662. Amino acids are indicated in brackets for coding sequence changes (when only a single amino acid is given the nucleotide change is a synonymous one). The gaps in the numeration indicate the three different *D. silvestris* sequences.

second piece of 3,039bp includes exons VI to VIII (AF051666) and the third piece of 2,587bp includes exon IX (AF051667). The proteins encoded begin at the BTB domain and include the type A, C and D protein types (Table 1), although only type A has so far been found as a cDNA (Davis *et al.*, 2000b).

The male specific peptide encoded by exons I and II in *D. melanogaster* has not yet been found in the Hawaiian *Drosophila*, and none of the putative promoter sequences are known. The majority of the sequence for each species is intronic. The intron and exon sizes and the transcripts for the Hawaiian species and *D. melanogaster* are summarised in Table 1.

On a gross level the *fru* genes in the three species are precisely conserved in that there are the same number of exons in the same order (Table 1), and the intron-exon boundaries are the same (not including exons I and II). The known and deduced transcripts are also well conserved. The actual exon sizes have some small differences: when compared to *D. heteroneura* exon VII is longer in *D. silvestris*, and all exons except III and IV are slightly different lengths in *D. melanogaster* (Table 1). Exon VII encodes the type C terminal exon. The other full length terminal exon known for the Hawaiian species is exon IX (type A) and this is longer than that in *D. melanogaster*.

At the nucleotide sequence level the genes (introns and exons) have numerous differences that are summed up for the two Hawaiian species in Table 2. Counting deletions and insertions as base pair changes (*i.e.*, a deletion or insertion of 1 base is equivalent to a single change) there are a total of 216 changes in the 8,030 bases of syntenic sequence, *i.e.*, 2.7% difference. The vast majority (183) of these are non-coding, and there are only 15 amino acid differences. In contrast the intronic sequence of *D. melanogaster* has little, if any, conservation with the Hawaiian *Drosophila* (not shown). The exception to this is a short section 5' to exon III (bases 3,950-4,034 in *D. heteroneura* AF051662). This sequence does not correspond to any of the six known promoters in this gene (Davis and Ito, 2001) and is possibly some sort of enhancer sequence. It has 82% nucleotide identity over 85 bases (not shown). This lack of conservation of non-coding sequence extends to the 3' UTR sequences of each transcript.

The exons in *D. melanogaster* vary considerably in the level of amino acid conservation with the Hawaiian species. The conservation varies from 100% for exons III and IV that encode the BTB domain, to 46% for exon IX. The combined conservation for the exons common to the various transcripts (exons III, IV, V, VI and VIII) is 78%. The 3' terminal exons show amino acid conservation of 72%, 63%, and 86% (for types B, C and E respectively). The 3' end of the type D transcript (exon VIIIa) has only two amino acids for each species (Gly and Glu).

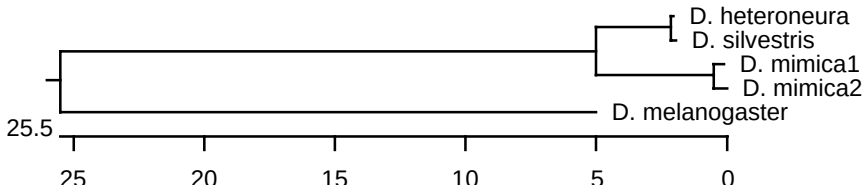

Figure 1. ClustalW analysis of the exon IX coding sequences. D. mimica1 and D. mimica2 are different alleles for this species.

A striking observation is that there are only four coding sequence changes (one amino acid change) for exon IX (975bp coding) between the Hawaiian species. This is one fifteenth of the amino acid changes in one third of the protein. Exon IX encodes the type A Zinc finger sequence. The first third of this exon is partially conserved in *D. melanogaster* and the Zn finger region is highly conserved (Davis *et al*., 2000a,b). The rest of the exon (approximately half the length), however, is completely unconserved. The overall conservation is 46%. I have looked at this exon in a related Hawaiian species, *D. mimica* (AF051673). The 1011bp of coding sequence for this species has 110 nucleotide differences (11%) compared to the *D. heteroneura* sequence (not shown), and the protein homol-ogy is high with only 36 amino acid changes. The majority

of these changes are in the region of high variability seen between *D. melanogaster* and *D. heteroneura*. Thus the majority of this exon appears to be a rapidly diverging sequence and would appear to be a useful region for future phylogenetic studies in the Hawaiian *Drosophila*. A phylogenetic representation of the exon IX sequences using ClustalW is given in Figure 1. Interestingly, the two different alleles from *D. mimica* are less well conserved than the *D. heteroneura* and *D. silvestris* sequences.

Unfortunately the close similarity of the *fru* sequences between *D. heteroneura* and *D. silvestris* may not shed light on the two alternate theories of the origins of these species. Although at first glance the data suggest that these species are the result of a very recent speciation event (presumably after the migration to Hawaii of the parental species), the sequences may have converged due to sequence introgression through the natural hybridisation known in these species (Carson *et al.*, 1989; Kaneshiro, 2000). However, the data do indicate the close relationship between the two species.

References: Ahearn, J.N., H.L. Carson, T. Dobzhansky, and K.Y. Kaneshiro 1974, Proc. Natl. Acad. Sci. USA. 71: 901-903; Boake, C.R.B., M.P. DeAngelis, and D.K. Andreadis 1997, Proc. Natl. Acad. Sci. USA. 94: 12442-12445; Carson, H.L., 1997, J. Hered. 88: 343-352; Carson, H.L., K.Y. Kaneshiro, and F.C. Val 1989, Evolution 43: 190-203; Davis, T., J. Kurihara, and D. Yamamoto 2000a, Gene 246: 143-149; Davis, T., J. Kurihara, E. Yoshino, and D. Yamamoto 2000b, Hereditas 132: 67-78; Kambysellis, M.K., K. Ho, E. Craddock, F. Piano, M. Parisi, and J. Cohen 1995, Curr. Biol. 5: 1129-1139; Kaneshiro, K.Y., 1974, Evolution 30: 740-745; Kaneshiro, K.Y., 1980, Evolution 34: 437-444; Kaneshiro, K.Y., 2000, Sexual selection and speciation in Hawaiian *Drosophila* (Drosophilidae): A model system for research in Tephritidae, In: *Fruit Flies (Tephritidae): Phylogeny and Evolution of Behaviour* (Aluja, M., and A.L. Norrbom eds.) pp. 861-77, CRC Press; Ohta, A., 1978, Evolution 32: 485-492; Usui-Aoki, K., H. Ito, K. Ui-Tei, K. Takahashi, T. Lukacsovich, W. Awano, H. Nakata, Z.F. Piao, E.E. Nilsson, J. Tomida, and D. Yamamoto 2000, Nat. Cell Biol. 2: 500-506.