MOLECULAR EVOLUTION OF REGULATORY GENES OF DEVELOPMENT IN *Drosophila*: THE *daughterless* GENE.

J. Antonio BOUZADA, Manuel R. GANDARELA and Emilio VALADÉ
Dpto. Bioloxía Fundamental (Xenética), Univ. Santiago de Compostela. 15706 A Coruña. Spain.

daughterless (*da*) is a very promiscuous gene of *Drosophila*, whose functions can be separed in two groups. In the first are the functions done by the maternal transcripts: sex determination and dosage compensation (Cline, 1980). In the second, are the functions done by the embryonic transcripts: determination of the peripheral nervous system fate (Caudy et al., 1988) and differentiation of cells that form the adult cuticle (Cronmiller and Cline, 1987). A predicted protein product of this transcription unit has sequence similarities with the oncogene *myc*, with the gene *MyoD1* which is involved in myoblast determination, and with the *Drosophila achaete-scute* complex, which is involved in neuronal precursor determination. This gene product act as heterodimer with other gene products which are responsible of the functional specificity. Formation of these heterodimers is possible due to the presence of two dimerization domains in the *da* structure: the basic-Helix-Loop-Helix (bHLH) and the leucin zipper. Our work is a continuation of the studies of the evolution of genes implicated in sex determination of *Drosophila* by O'Neil and Belote (1992) and Scott Whalthour and Schaffer (1994) using the gene *transformer* (*tra*). These studies also are interesting because genes implied in sex determination or embryonary development could be involved in speciation if their coevolution leads to the emergence of reproductive barriers.

We have studied the evolution of *da* in the *melanogaster* complex. In order to do it we amplified (by PCR) , cloned and sequenced a 315 bp region of this gene, including the bHLH and Leucin zipper domains in the four species of this complex: strains Hawaii and Florida of *D. melanogaster*; strains Colombia, Australia and XX;Y yw of *D. simulans*; strain David 83 of *D. mauritiana*; and strain David 85 of *D. sechellia*, supplied by the National *Drosophila* Species Resource Center at Bowling Green. We also used the already published sequence of this gene in the strain *Canton-S* of *melanogaster* (Caudy et al., 1988).

The obtained sequences were aligned using the multiple alignment option from the CLUSTAL V program (Higgins et al., 1992). Then, the polymorphism (allelic variation within a species) in the *D. melanogaster* and *D. simulans* sequences was calculated using the method of Nei and Li (1979). The estimation of divergence (variation between species in homologous sequences) was calculated using only a random sequence by species.

Finally, we applied the HKA test (Hudson et al., 1987) to determine whether the locus *daughterless* appears to be evolving according to a neutral model (Table 3). The neutral theory of molecular evolution predicts that regions of the genome that evolve at high rates, reveled by interspecific DNA sequence comparations, will also exhibit high levels of polymorphism within species.

In *D. melanogaster*, only 4 nucleotide polymorphisms exist among the three sequences. Two of the variants are silent substitutions, while the other two produce aminoacid replacement. This polymorphism in the *da* sequences is only lightly elevated in comparation with data of *zeste* sequences. In contrast, there were 17 polymorphic nucleotide sites among the three *D. simulans* sequences, ten of which cause aminoacid replacements. Thus the level of polimorphism as synonymous as nonsynonymous in *Drosophila simulans* is ten fold higher for *da* than the polymorphism found in *zeste* (table 1).

There are 15 nucleotide differences among the four species, 8 are synonymous and 7 nonsynonymous (table 2). The amount of divergence at *daughterless,* between the four species, is lightly higher than the levels seen at other loci We performed HKA test (Hudson et al., 1987) using the *zeste* locus as a control, whose patterns of polymorphism and divergence indicate neutral evolution (Table 3). We made two group of data: total sites and synonymous sites. In the first group, the value of χ^2 of 14.96 in the *melanogaster-simulans* comparison and of 6.772 in the *simulans-mauritiana* were statistically significant. In the second group, the χ^2 value of 6.740 in the *melanogaster-simulans* comparison and of 11.594 in the *simulans-mauritiana* were also statistically significant (table 3). Therefore, we can reject, with a high level of signification, a neutral model of evolution on the basis of our current sample of *daughterless* alleles.

Species	Polimorphism	*daughterless*	*zeste*
D. melanogaster	Total	0.0084	0.0061
	Synonymous	0.0235	0.0293
	Nonsynonymous	0.0054	0.0000
D. simulans	Total	0.0372	0.0031
	Synonymous	0.0903	0.0150
	Nonsynonymous	0.0204	0.0000

Table 1. DNA polymorphism at the *Drosophila daughterless* and *zeste* locus, for the *D. malanogaster* and *D . simulans* species. Data of *zeste* are of Hey and Kliman (1993).

LOCUS	synonymous changes	nonsynonymous changes	S	R
da	8	7	0.113	0.028
zeste	13.5	3.5	0.081	0.005

Table 2. DNA Divergence at the *daughterless* and *zeste* locus in the *melanogaster* subgroup of *Drosophila*.. In the table, S is the number of synonymous changes per synonymous site and R is the number of nonsynonymous changes per nonsynonymous site. Data of *zeste* are of Hey and Kliman (1993).

These values so high of χ^2 indicate a strong tendency for nucleotide polymorphism to accumulate within some species but not between species. The reason of the high χ^2 values is the high level of polymorphism in the *Drosophila simulans* strains, both synonymous and nonsynonymous. The amount of divergence in *daughterless,* between the four species, is lightly elevated comparated to levels seen at other loci, but we find a polymorphism in *D. simulans* ten fold higher in *daughterless* than the levels found in *zeste,* indicating that neither low levels of functional constraint on the locus or elevated levels of neutral mutation can be invoked to explain the high level of nucleotide polymorphism within *D. simulans*.

Total Sites

	θ^a					
Sp 1-Sp 2	*da*	*zeste*	T^b	f^c	χ^{2d}	P^e
mel-sim	2.844	4.830	4.110	2.5	14.96	<0.001
mel-mau	1.896	5.635	3.613	1	2.094	0.5-0.3
mel-sech	2.212	4.830	4.272	1	1.219	0.7-0.5
sim-mau	8.532	9.660	0.048	1	6.772	0.05-0.02
sim-sech	9.796	8.855	0.025	1	3.645	0.2-0.1

Synonymous Sites

	θ^a					
Sp 1-Sp 2	*da*	*zeste*	T^b	f^c	χ^{2d}	P^e
mel-sim	2.310	6.011	2.226	1.4	6.740	0.05-0.02
mel-mau	1.750	6.680	2.057	1	0.218	0.9-0.8
mel-sech	1.820	6.513	2.408	1	0.277	0.9-0.8
sim-mau	2.750	4.676	0.790	1	11.59	0.01-0.001
sim-sech	4.690	7.014	0.262	1	1.442	0.3-0.2

Table 3. HKA test for *daughterless* and *zeste* loci and two species. Results for total sites and synonymous sites are shown separately.
[a] Estimate of $4N\mu$ for species 1.
[b] Estimate of the time since the common ancestor of the species, in units of 2N generations, where N is the effective population size of species 1.
[c] Estimate of the scalar by which estimates of $4N\mu$ for species 1 are multiplied to get those of species 2.
[d] Goodness-of-fit statistic.
[e] Probability of observing an χ^2 greater than or equal to the actual value, when a χ^2 distribution with 2 degrees of freedon is assumed)

This pattern could be due to balancing selection. The presence of balanced polymorphism in the *daughterless* gene could explain the high level of polymorphism observed in *D. simulans*. The existence of a balanced polymorphism at a single site can lead to higher levels of neutral polymorphism at linked sites (Strobeck, 1983). The reason this can occur is that during the time that the balanced polymorphism is maintened by selection, new mutations will tend to accumulate in the region tightly linked to the selective site. Balancing selection is likely to be important in genes as *Adh* (Hudson et al., 1987) or *prune* (Simmons et al., 1994).

References:

Caudy, M., Vässin, H., Brand, M., Tuma, R., Jan, L-Y. and Jan, Y-N. (1988). *daughterless,* a *Drosophila* gene essential for both meurogenesis an sex determination, has sequence similarities to *myc* and the *achaete-scute complex*.. Cell 55: 1061-1067.

Cline, T. W. (1980). Maternal and zigotic sex-specific gene interations in *Drosophila melanogaster.* Genetics 90: 683-698.

Cronmiller, C. and Cline, T. W. (1987). The *Drosophila* sex determination gene *daughterless* has defferent funtions in the germ line versus soma. Cell 48: 479-487.

Hey, J. and Kliman R. M. (1993). Populations genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol Biol. Evol. 10: 804-822.

Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. 8: 189-191.

Hudson, R. R., Kreitman, M. and Aguadé, M. (1987). A test of neutral molecualr evolution based on nucleotide data. Genetics 116: 153-159.

Nei, M. and Li, W-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Acad. Sci. USA 76: 5269-5273.

O'Neil, M. T. and Belote, J. (1992). Interespecific comparation of the *transformer* gene of *Drosophila* revels an unusually high degree of evolutionary divergence. Genetics 131: 113-128.

Simmons, G. M., Kwork, W., Matulonis, P. and Venkatesh, T. (1994). Polymorphism and divergence at the *prune* locus in *Drosphila melanogaster* and *simulans.* Mol. Biol. Evol. 11; 666-671.

Strobeck, C. (1983). Expected linkaje disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics 103;:545-555.

Whalthour, C. S. and Schaffer, S. W. (1994). Molecular population genetics of sex determination genes: the *transformer* gene of *Drosophila melanogaster.* Genetics 136: 1367-1372.