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

Manuel R. GANDARELA, J. Antonio BOUZADA and Emilio VALADÉ
Dep. de Bioloxía Fundamental, Facultade Bioloxía, Universidade de Santiago de Compostela, 15706, A Coruña, Spain

The study of intra and interespecific variation at some genes of development is interesting in regard to a possible relation between speciation and development, specially, the genes that interact to carry out its function. The postzigotic reproductive isolation may be considered as the result of the incompatibility of two developmental patterns in the hybrid. Hence, the interaction between complementary factors postulated to explain the isolation could correspond to a disrupted interaction between developmental proteins in the hybrid. It should be necessary to look for candidate genes and to analyze its variation in related species.

Then, we decided to focus our study on the *achete-scute* complex, involved in the development of sensory organs and the central nervous system of *Drosophila*. One member of this complex is the *T4* transcription unit, also known as *scute* or *sisterless-b*. This gene acts as a pleiotropic gene taking part in both the diferentation of peripheral nervous system (*scute* function) and the establishment of X:A ratio (*sisterless-b* function). In the development of sensory organs, *scute* (*sc*) and other member of the complex, *achaete*, are expressed in the imaginal discs confering on ectodermal cell clusters the ability to become neural elements. Both *sc* and *ac* yield two *basic-Helix-Loop-Helix* proteins (*bHLH*) which carry out their functions by interacting with other *bHLH* protein encoded by the *daugtherless* gene. On the other hand, as we have mentioned, sc contributes to the establishment of the X:A ratio (the number of X chromosomes relative to the number of sets of autosomes) which is determinig very important process in the development of *Drosophila*, namely dosage compensation and sex determination. The X:A ratio is established by the balance between genes contained in the X chromosomes (numerator elements) and genes contained in the autosomes (denominator elements).The *scute* gene acts as numerator element, and again, the protein encoded by *daughterless* mediates its action.

		LOCUS		
Species	Polymorphism	*scute*	*asense*	*period*
melanogaster	total	0.0019	0.0021	0.0062
	synonymous	0.0000	0.0043	0.0400
	non synonymous	0.0025	0.0015	0.0003
simulans	total	0.0093	0.0000	0.0115
	synonymous	0.0152	0.0000	0.0400
	non synonymous	0.0074	0.0000	0.0020
mauritiana	total	0.0056	0.0023	0.0118
	synonymous	0.0000	0.0048	0.0456
	non synonymous	0.0074	0.0012	0.0015

Table 1. Levels of polymorphism estimated as the average number of pairwise differences per base pair within species (based on all base-pair differences). The *period* and *asense* data are obtained from Kliman and Hey (1992) and Hilton *et al.* (1994), respectively.

Given the significance of *sc* in the development of *Drosophila,* it would be important to understand the evolutionary forces that are affecting this gene. Hence, it is necessary to study its levels of interespecific and intraespecific variation.

Our study was done in the *melanogaster* complex. Three strains of *D. melanogaster* (CantonS; Oahu, Hawaii; Toonda, Australia), three strains of *D. simulans* (Leticia, Colombia; Australia; X^XY y w), two strains of *D. mauritiana* (David, J. 1974; David, J. 1983) and one strain of *D. sechellia* (David, J. 1985) were included. From each strain, a region of the *sc* gene was amplified by PCR , cloned and sequenced.

From the alignment of sequences we estimated levels of polymorphism and divergence and compared these estimates with the values observed in *asense* and *period*, taken from literature. Intraespecific variation is summarized in Table 1, and divergence is summarized in Table 2. As it happened in *asense, sc* exhibits a pattern of very reduced levels of total polymorphism, within *D. melanogaster*, *D. simulans* and *D. mauritiana*, relative to other loci studied in the *melanogaster* complex (Table 1). When we consider synonymous and non synonymous interespecific variation separately, we observe that the reduction is stronger in the first case and that the non synonymous polymorphism is slightly increased. About the divergence we find similar results, a slight reduction in the synonymous variation and an increment in the non synonymous variation (Table 2).

LOCUS	Analyzed length (coding region)	Silent sites	S	R
scute	360	88.3	0.136	0.015
asense	1067	208	0.072	0.016
period	1679	386	0.200	0.003

Table 2. Levels of divergence of *scute* , *asense* and *period*. The values of *asense* and *period* were obtained from Kliman and Hey (1992) and Hilton *et al.* (1994), respectively. Changes were determined for a data subset with one sequence randomly drawn from each species. Silent sites were calculated by considering, for each base position of a *D. simulans* sequence, the fraction of possible base changes ($^1/_3$, $^2/_3$, $^3/_3$) that would not affect the amino-acid sequence. S is the no. of synonymous changes divided by the no. of silent sites. R is the no. of replacement changes divided by the no. of replacement sites.

The *sc* region was tested for departures from an equilibrium neutral model of evolution. We used the HKA test that allows us to reject a neutral model if the ratio of polymorphism to divergence differs significatly among independent loci, provided that one locus evolves according to the predictions of the neutral theory. The tests were performed on synonymous and total variation and we used the *period* locus in the comparisons because this gene evolves according to the neutral model. The results of HKA test (Table 3) indicate a departure from the neutral model of evolution. The deviation was significant (α=0.05)

in the *simulans/mauritiana* comparison when total variation was considered. If we consider synonymous variation, the deviations were significant both *simulans/mauritiana* and *melanogaster/simulans* comparisons.

Our results agree with previous studies on the tip of the X chromosome where the *achaete-scute* complex is located. This region is characterized by low rates of recombination. There are several RFLP studies that informed of a reduced variation within the *yellow-achaete-scute* region in *D. melanogaster* and *D. simulans*. These results were confirmed by a later study of sequence analysis about the *asense* gene that also detected a reduced variation in *D. mauritiana*. A process known as 'genetic hitchhiking' was used to explain these observations.

Genetic hitchhiking occurs when an advantageous mutation arises in the population. If this change confers a higher reproductive success upon carriers, it may increase in frequency and replace all other variants of the gene. Simultaneously, these adaptations are expected to lead to the fixation of variants in sites linked to the adventageous mutation. This process would give rise to a region of zero or reduced variation among DNA sequences around the site of the original mutation. The size of the affected region depends on how much recombination occurred near the site during the fixation process (size increases as the rate of recombination decreases).

A)

	θ^a					
Sp. 1-Sp. 2	*scute*	*period*	T^b	f^c	X^{2d}	P^e
mel-sim	1.79	8.41	6.29	3.15	4.10	0.20-0.10
mel-mau	1.43	8.41	7.09	2.69	4.49	0.20-0.10
sim-mau	3.60	20.15	1.28	1.11	6.69	0.05-0.025

B)

	θ^a					
Sp.1-Sp. 2	*scute*	*period*	T^b	f^c	X^{2d}	P^e
mel-sim	1.14	7.72	6.69	2.87	5.89	0.10-0.05
mel-mau	0.79	8.11	7.62	2.67	6.68	0.05-0.025
sim-mau	1.67	17.76	1.50	1.21	6.27	0.05-0.025

Table 3. Results of HKA test for *scute* and *period* loci considering total substitutions (A) and synonymous substitutions (B). [a] Estimate of $3N\mu$ for species 1. [b] Estimate of the time since the common ancestor of the species, in units of $^3/_2$ generations, where N is the effective population size of species 1. [c] Estimate of the scalar by which estimates of $3N\mu$ for species1 are multiplied to get those of species 2. [d] Probability of observing an X^2 greater than or equal to the actual value, when an χ^2 distribution with 2 degrees of freedom is assumed.

Our results suggest that natural selection is acting on the variation of the *scute* locus in *D. melanogaster, D. simulans* and *D. mauritiana*. As we showed above, the test HKA (Table 3) reveals a departure from the neutral model of molecular evolution in this gene and this is more evident when we consider synonymous variation. This result and the pattern of variation described above are in accordance with the hitchhiking model. If only purifying selection is acting we would expect that synonymous variation were the least affected. But if a region is subjected to the action of genetic hitchhiking, the selective sweep of a new beneficial allele will also lead to the fixation of silent variants that in other conditions would not be affected by purifying selection. The increase of non synonymous polymorphism observed in the three species agree to genetic hitchhiking predictions. In general natural selection is expected to be less effective in removing deleterious variation from regions of low recombination. This effect should be specially marked for a region undergoing selective sweeps where the selection against slightly deleterious mutations will be dominated by the fixation process of linked favorable mutations. Therefore, if most of the detected non synonymous polymorphism corresponds to slightly deleterious mutations, its increase will be in accordance with the hitchhiking model.

As we explained above, the levels of variation are affected in the three species analyzed. The most parsimonious mode to explain these observations is to consider that the selective sweep started in the ancestral species prior to speciation. In adition to the differences accumulated after the species splitting, the divergence presents a component due to genetic polymorphism witihin the ancestral species. In consequence, if the selective sweep started in the ancestral species, a reduction of synonymous variation and an increase of non synonymous variation were produced at that time. Therefore, the divergence would also be affected in the same way. Our results in *scute* locus agree with these predictions.

References

Campuzano, S. and Modolell, J. (1992). Patterning of the *Drosophila* nervous system: the *achaete-scute* gene complex. TIG 8: 202-208.

Cline, T.W. (1993). The *Drosophila* sex determination signal: how do flies count two?. TIG 9: 385-390.

Hilton, H., Kliman, R.M. and Hey, J. (1994). Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution 48: 1900-1913.

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

Kliman R.M. and Hey, J. (1992). DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics 133: 375-387.

Martín-Campos, J.M., Comerón, J.M., Miyashita, N. and Aguadé, M. (1992). Intraespecific and Interespecific variation at the y-ac-sc region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics 130: 805-816.