

Colonization of America by *Drosophila subobscura*: Heterotic effect of chromosomal arrangements revealed by the persistence of lethal genes

F. Mestres*, J. Balanyà*, C. Arenas†, E. Solé*, and L. Serra**

Departaments de *Genètica and †d'Estadística, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain

Communicated by Ramon Margalef, University of Barcelona, Barcelona, Spain, May 25, 2001 (received for review February 20, 2001)

About 20 years ago *Drosophila subobscura*, a native Palearctic species, colonized both North and South America. In Palearctic populations lethal genes are not associated in general with particular chromosomal arrangements. In colonizing populations they are not randomly distributed and usually are associated to a different degree with chromosomal arrangements caused by the founder event. The persistence of two lethal genes in the colonizing populations, one completely associated with the O_5 inversion and the other partially associated with the O_{3+4+7} arrangement, has been analyzed. In all populations studied (five North American and six South American) the observed frequency of the lethal gene completely associated with the O_5 inversion is higher than expected, the difference being statistically significant in all South American and one North American populations. The observed frequency of the lethal gene partially associated with the O_{3+4+7} arrangement is also significantly higher than expected. Taking into account that the O_5 inversion exhibits significant latitudinal clines both in North and South America, an overdominant model favoring the heterokaryotypes seems to be in operation. From this model, a polynomial expression has been developed that allows us to estimate the relative fitness and the coefficient of selection against all karyotypes not carrying the O_5 inversion. The relative fitness of the O_5 heterokaryotypes is higher in South American than in North American populations. Furthermore, the observed frequencies of the lethal genes studied are in general very close to those of the equilibrium. This case is an outstanding demonstration in nature of an heterotic effect of chromosomal segments associated with lethal genes on a large geographic scale.

Drosophila subobscura was considered a native Palearctic species until it was first detected in Puerto Montt (Chile) in 1978 (1). In 1982 it also was found in Port Townsend (Washington) (2). This species spread rapidly in both colonized areas. It is now found throughout Chile, from La Serena to Punta Arenas, and along the West Coast of North America (3, 4). This evolutionary phenomenon of colonization has been studied from both ecological and genetic points of view. The study of lethal genes has helped us to understand certain aspects of the colonization of America by *D. subobscura*. The presence of allelic lethal genes in North and South American populations strongly suggests that both colonizations are related (5). It also has been estimated that between 10 and 150 individuals were involved in the initial colonization (6). Lethal genes are not randomly distributed among the chromosomal arrangements in colonizing populations (7). Some lethal genes are completely associated with particular chromosomal arrangements, as in the case of the O_5 ; in others, the association is partial as it occurs with the O_{3+4+7} arrangement (5, 6, 8). In Palearctic populations these associations are not found in general (6, 7).

In the present study the persistence of some lethal genes in North and South American colonizing populations of *D. subobscura* has been analyzed, and an heterotic model has been developed to explain this phenomenon. This analysis constitutes an outstanding example of heterosis on a large geographic scale

including several populations from two continents derived from the same colonizing event.

Materials and Methods

Because of the complete association of a lethal gene and the O_5 chromosomal inversion in colonizing populations, the frequency of this inversion corresponds to the frequency of this lethal gene, and thus it is only necessary to study the frequency of this inversion in different periods of time to monitor the persistence of this lethal gene. On the other hand, a partial association between a lethal gene and the O_{3+4+7} chromosomal arrangement has been found in American populations, too. In this case, as in the American populations there are O_{3+4+7} arrangements free of lethal genes, the study of the persistence of this lethal requires a knowledge of its frequency in different periods of time. In the present study, the following North and South American colonizing populations have been used: Gilroy (37° 00' N), Davis (38° 33' N), and Eureka (40° 49' N) (California); Medford (42° 20' N) and Salem (44° 57' N) (Oregon); Centralia (46° 43' N) and Bellingham (48° 45' N) (Washington); Santiago de Chile (33° 30' S), Laja (37° 10' S), Valdivia (39° 46' S), Puerto Montt (41° 28' S) and Coihaique (45° 35' S) (Chile); Bariloche (41° 11' S) (Argentina). The relevant data of these populations can be found in previous articles (8–12).

The persistence of lethal genes in the American populations was analyzed by using the following procedure: if q_0 (initial frequency of a lethal gene) is known, the expected frequency of this completely recessive lethal gene after t generations (q_{exp}) can be calculated by the well-known expression:

$$q_{exp} = \frac{q_0}{1 + tq_0}$$

The number of generations per year of *D. subobscura* in nature has been estimated as 4–6 (13, 14). Thus, t is obtained by multiplying the number of years that have elapsed since q_0 was estimated by 5 (the mean number of generations per year). To determine whether the observed frequency of the lethal gene (q_{obs}) is significantly higher than q_{exp} at the 0.05 level the following test was carried out:

$$H_0: q_{obs} = q_{exp}$$

$$H_1: q_{obs} > q_{exp}$$

Taking into account that the number of lethal genes follows a binomial distribution with parameters, under H_0 , the sample size n and q_{exp} , we obtain W as the critical region, defined by:

$$W = \{k: k \geq \theta\},$$

*To whom reprint requests should be addressed. E-mail: lluis@porthos.bio.ub.es.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

where k , the number of lethal genes, is greater or equal to a constant O . This constant is determined in such a way that the significant level of the test is the fixed value 0.05. The computation of the constant O is given by the expression:

$$\sum_{i=0}^{\theta-1} \binom{n}{i} (q_{\text{exp}})^i (1 - q_{\text{exp}})^{n-i} \geq 0.95.$$

According to this procedure, the persistence of the lethal gene completely associated with the O_5 inversion was analyzed for Gilroy, Davis, Eureka, Medford, Centralia, Santiago, Laja, Valdivia, Bariloche, Puerto Montt, and Coihaique populations. Most probably only one chromosome carrying the O_5 inversion associated with the lethal gene was present in the original sample of colonizers (15). Thus, under this assumption a general analysis of the persistence of this lethal gene in North and South American populations was carried out, as follows: q_0 was estimated assuming an initial sample size of colonizers of 10, 50, and 100 individuals, respectively (4, 6, 16). The value of q_{exp} was calculated by assuming that the colonization started around 1978 in South America (1) and around 1980 in North America (2). The value of q_{obs} was estimated as a weighted average by combining all localities within each continent of the observed frequency of the O_5 inversion in 1986 in South America and in 1994 in North America. Also, the persistence of a recessive lethal gene partially associated with the O_{3+4+7} chromosomal arrangement was analyzed in the population of Gilroy, in 1985 and 1991. In this case, where the association is not complete, lethal allelism tests had to be carried out in the population because O_{3+4+7} arrangements free of lethal genes are found, too. These were the only samples for which data on lethal allelism were available.

Theoretically, persistence of a lethal gene can be explained by an heterotic effect of other linked genes included on the same chromosomal arrangement. According to this hypothesis of

linked heterosis, we have developed the following model for the lethal gene completely associated with the O_5 inversion. If O_G represents any O chromosome arrangement different from O_5 , then the relative fitness of the three classes of genotypes are:

$$\begin{array}{l} \text{Genotype:} \quad O_G O_G \quad O_G O_5 \quad O_5 O_5 \\ \text{Relative fitness:} \quad w_1 = 1 - s \quad w_2 = 1 \quad w_3 = 0. \end{array}$$

The aim is to calculate, for each population, the value of s (selection coefficient against $O_G O_G$ karyotypes) when the number of generations, k , and the initial (q_0) and final (q_k) frequencies of the O_5 inversion are known.

Thus, after one generation, the expected frequency of the O_5 inversion is given by

$$q_1 = \frac{p_0 q_0}{p_0^2 (1-s) + 2p_0 q_0}.$$

The expected frequency of the O_5 inversion after two generations is given by:

$$q_2 = \frac{p_0 q_0}{p_0^2 (1-s)^2 + p_0 q_0 (1-s) + 2p_0 q_0}.$$

As can be shown, by mathematical induction, after k generations the frequency q_k is given by

$$q_k = \frac{p_0 q_0}{p_0^2 (1-s)^k + p_0 q_0 (1-s)^{k-1} + \dots + p_0 q_0 (1-s) + 2p_0 q_0},$$

so

$$q_k p_0^2 (1-s)^k + p_0 q_0 q_k (1-s)^{k-1} + \dots + p_0 q_0 q_k (1-s) + 2p_0 q_0 q_k - p_0 q_0 = 0,$$

Table 1. Analysis of the persistence of some lethal genes in colonizing populations of *D. subobscura*

Lethal	Population	Years	q_0	q_{exp}	q_{obs}	θ
South American populations						
O_5	Santiago	81–86	0.026 (5/192)	0.016	0.081 (11/135)	6*
O_5	Laja	81–86	0.051 (15/301)	0.022	0.111 (16/144)	7*
O_5	Valdivia	81–86	0.095 (18/186)	0.028	0.068 (11/161)	9*
O_5	Bariloche	81–86	0.090 (12/133)	0.028	0.124 (15/121)	8*
O_5	Puerto Montt	81–86	0.130 (19/146)	0.031	0.129 (19/147)	9*
O_5	Coihaique	81–86	0.057 (12/213)	0.023	0.142 (16/113)	7*
North American populations						
O_5	Gilroy	85–94	0.007 (1/141)	0.005	0.010 (1/96)	2
O_5	Davis	85–94	0.018 (4/227)	0.010	0.010 (1/100)	4
O_5	Eureka	85–94	0.077 (12/156)	0.017	0.032 (2/63)	5
O_5	Medford	86–94	0.064 (6/94)	0.018	0.020 (2/102)	5
O_5	Centralia	86–95	0.106 (18/170)	0.018	0.077 (5/65)	4*
O_{3+4+7}	Gilroy	85–91	0.027 (3/111)	0.015	0.052 (4/77)	4*
Global analysis for South America						
O_5	General (10 [†])	78–86	0.050	0.017	0.107 (88/821)	21*
O_5	General (50 [†])	78–86	0.010	0.007	0.107 (88/821)	11*
O_5	General (100 [†])	78–86	0.005	0.004	0.107 (88/821)	8*
Global analysis for North America						
O_5	General (10 [†])	80–94	0.050	0.011	0.036 (21/580)	12*
O_5	General (50 [†])	80–94	0.010	0.006	0.036 (21/580)	8*
O_5	General (100 [†])	80–94	0.005	0.004	0.036 (21/580)	6*

q_0 = initial frequency of a lethal gene; q_{exp} = expected frequency of this lethal gene after t generations; q_{obs} = observed frequency of this lethal gene after t generations; θ = constant for which the test has the fixed value of 0.05 as the significant level.

* q_{obs} significantly higher than q_{exp} at the 0.05 level.

[†]Estimated number of colonizers.

Table 2. Relative fitness and selection coefficient against $O_G O_G$ individuals estimated from the model:

Population	K	Genotype:	$O_G O_G$	$O_G O_5$	$O_5 O_5$
		Relative fitness:	$w_1 = 1 - s$	$w_2 = 1$	$w_3 = 0$
		Relative fitness (positive real root of the polynomial),	Coefficient of selection,		Equilibrium frequency,
		w_1	s		$q_e = s/(s + 1)$
Santiago	25	0.895	0.105		0.095
Laja	25	0.869	0.131		0.116
Valdivia	25	0.931	0.069		0.065
Bariloche	25	0.857	0.143		0.125
P. Montt	25	0.852	0.148		0.129
Coihaique	25	0.832	0.168		0.144
S. America (10*)	40	0.879	0.121		0.108
S. America (50*)	40	0.874	0.126		0.112
S. America (100*)	40	0.869	0.131		0.116
Gilroy	45	0.983	0.017		0.017
Davis	45	0.999	0.001		0.001
Eureka	45	0.973	0.027		0.026
Medford	40	0.996	0.004		0.004
Centralia	45	0.917	0.083		0.077
N. America (10*)	70	0.963	0.037		0.036
N. America (50*)	70	0.957	0.043		0.041
N. America (100*)	70	0.952	0.048		0.046
Gilroy (O_{3+4+7})	30	0.935	0.065		0.061

O_G = any O chromosomal arrangement different from O_5 . K = polynomial degree (number of generations). In the case of the population of Gilroy (O_{3+4+7}) the model corresponds to the lethal gene partially associated with the O_{3+4+7} arrangement (see text for further details). *Estimated number of colonizers.

thus, a polynomial of degree k in the indeterminate $(1 - s)$ is obtained. The degree of the polynomial, k , is the number of generations elapsed between q_0 and q_k and the relative fitness $(1 - s)$ of the $O_G O_G$ karyotypes is found as the positive real root of this polynomial. In the particular case of the lethal gene partially associated with the O_{3+4+7} , O_G stands for any chromosomal arrangement free of this lethal.

Once the s value has been estimated, the equilibrium frequency for this model of heterozygote advantage can be obtained as $q_e = s/(s + 1)$.

Results

The results of the analysis of the persistence of some lethal genes are presented in Table 1. With regard to the lethal gene completely associated with the O_5 inversion, its persistence in the period 1981–1986 is significantly higher than expected in all South American populations. Furthermore, despite being a lethal gene, in some of these populations (Santiago, Laja, Bariloche, and Coihaique) its frequency in the 1986 samples is even higher than that obtained in 1981. On the other hand, in North American populations, although the observed frequency of the lethal gene is almost always higher than expected, this difference is significant only in Centralia. Furthermore, in these populations, the observed frequency of the lethal gene is in general lower than its initial frequency (Table 1). The results corresponding to the global analysis show that the observed frequency of the lethal gene is significantly higher than expected in all cases, both in North and South America. The observed frequencies are also higher than the initial frequencies in all but one case (Table 1). Finally, for the lethal gene partially associated with the O_{3+4+7} arrangement, only data from the Gilroy sample could be used. Its observed frequency after a 6-year period is also significantly higher than expected and higher than its initial frequency (Table 1).

In Table 2 we present the relative fitness and selection coefficient against the $O_G O_G$ karyotypes according to the het-

erotic model used to explain the persistence of lethal genes. The relative fitness of these karyotypes is higher in North American than in South American samples, both when considering each population separately or in the general cases assuming different numbers of colonizers from which the initial frequency of O_5 has been deduced. Within each hemisphere, the fitness values obtained in the general cases are very similar irrespective of the number of colonizers. The equilibrium frequencies expected according to the heterotic fitness model are presented in the last column of Table 2. With regard to the O_5 inversion, these frequencies are in general very similar to those observed in North and South American populations. For the global analysis (both in North and South America) the observed frequencies practically coincide with the equilibrium frequencies. There is also a good agreement for the O_{3+4+7} chromosomal arrangement in the Gilroy sample.

Discussion

The dynamics of lethal genes in populations depends on their origin by mutation and their elimination by selection (17–19). If the frequency of a lethal gene does not decrease in time, some sort of positive selection for the heterozygotes should be assumed. Many cases of persistence of lethal genes in natural populations have been described (20–30). Frequently, the persistence of a lethal gene has been attributed to an heterotic effect in the heterozygotes (20, 21, 29, 31–33), either due to an heterotic effect of the lethal gene itself (34, 35) or by the effect of the genes tightly linked with the lethal, as it is the case for those that are included in the same chromosomal inversion (24, 25, 27, 29, 36). Another explanation for the persistence of the lethal gene could be its association with a segregation distortion gene (37).

However, the present study provides some clear advances: its scope is based on a large geographical scale, including two continents and several populations in each of them (six in South

America and five in North America), representing two natural replicates of the same colonizing event. Furthermore, the relevance of chromosomal inversions to explain the observed heterosis is emphasized in this case by the strong association between lethal genes and chromosomal arrangements due to the recent founder event. In the American populations of *D. subobscura* two lethal genes have persisted since the colonization started: one of them is associated with the O_{3+4+7} chromosomal arrangement and the other with the O_5 inversion. The first case has been studied only in the Gilroy population and the association is not complete (this lethal is always found within the O_{3+4+7} arrangements but some of them are free of this lethal gene). Furthermore, the frequency of this arrangement presents a significant latitudinal cline both in North and South America (9–11). The frequency of this lethal gene in the population of Gilroy in 1991 is significantly higher than expected and higher than its frequency in 1985 (Table 1). This finding could suggest that this arrangement, despite carrying a lethal gene, could have a heterotic effect on the heterokaryotypes.

In the second case, there is a complete association of the lethal gene with the O_5 inversion (5, 6, 8). In both hemispheres the observed frequency of this lethal gene is higher than expected, significant in the Centralia sample, in all South American populations, and in all global analyses (Table 1). Thus, a possible heterotic effect must be assumed. Some additional observations are strongly supporting this hypothesis: the frequencies of O_5 inversion are showing a significant latitudinal cline both in North and South America (9–11). Also, in a study of inbreeding of flies from a North American population, by nine generations of brother-sister mating, the observed frequency of the O_5 inversion remained higher than expected (38). Furthermore, in an analysis of the recombination in the heterokaryotypes O_5/O_{3+4} , a significant excess of the O_5 gametes was observed in the offspring (39). This is most probably not caused by a segregation distortion-like phenomenon. In *D. subobscura* a sex-ratio gene arrangement has been described, but it is located on the A (sex) chromosome ($A_{2+3+5+7}$) and was found only in populations from Tunisia, Morocco, and the Canary Islands (40, 41). In addition, autosomal meiotic drive phenomena—the so-called segregation distortion condition—has not been described in *D. subobscura* so far. In *D. melanogaster* the segregation distortion gene locus is

located on the second chromosome, which is not homologous to chromosome O of *D. subobscura* (the O chromosome is homologous to chromosome arm 3R of *D. melanogaster*) and the effect is limited to males (but not females) heterozygous for the segregation distortion gene condition (42). In the present case the analysis of the segregation of the O_5 inversion was carried out with heterokaryotypic O_5/O_{3+4} females, thus a segregation distortion gene condition seems not to be the reason for the significant excess of O_5 gametes. Finally, the clinal distribution of this inversion, both in North and South America, which emphasizes its adaptive value (41), makes it highly unlikely that the observed excess of O_5 gametes can be attributed to a segregation distortion phenomenon.

In South American populations there is a quite clear heterotic effect of the O_5 inversion, both when considering each population individually as well as in the three global cases (Table 2). In North America, although the persistence of the lethal gene completely associated with the O_5 inversion is significantly higher than expected only in the global cases and in the population of Centralia, the coefficient of selection against the $O_G O_G$ karyotypes is very low in all cases. Thus the heterotic effect, if any, is scarcely noticeable (Table 2). This phenomenon could be related to the fact that in South American populations the frequency of the O_5 inversion is 1 order of magnitude higher than in North America, making the heterotic effect more easily detectable. As the observed frequency of the O_5 inversion (q_{obs}) is very close to the equilibrium frequency expected under the heterozygote advantage model in all populations (q_{exp} in Table 2), there is a clear difference between the selection coefficients against $O_G O_G$ karyotypes in both colonized areas, perhaps reflecting the different environmental conditions. This is an outstanding example of an heterotic system in nature including a chromosomal inversion completely associated with a lethal gene, the relative fitness of this system being different in both colonized hemispheres of the New World probably because of the particular environmental conditions of the localities.

We thank Prof. R. B. Huey for the collection from Centralia and Profs. A. Fontdevila and M. Santos for their comments on the manuscript. This work was funded by Grants PB96-0793-C04-03 and BOS2000-0295-C02-02 from the Dirección General Enseñanza Superior (Spain) and Grant 1998SGR 00050 from the Generalitat de Catalunya.

- Brncic, D., Prevosti, A., Budnik, M., Monclús, M. & Ocaña, J. (1981) *Genetica* **56**, 3–9.
- Beckenbach, A. T. & Prevosti, A. (1986) *Am. Midl. Nat.* **115**, 10–18.
- Prevosti, A., Serra, L., Monclús, M., Mestres, F., Latorre, A., Ribó, G. & Aguadé, M. (1987) *Evol. Biol. (Col.)* **1**, 1–24.
- Prevosti, A., Serra, L., Aguadé, M., Ribó, G., Mestres, F., Balaña, J. & Monclús, M. (1989) in *Evolutionary Biology of Transient Unstable Populations*, ed. Fontdevila, A. (Springer, Berlin), pp. 114–129.
- Mestres, F., Balaña, J., Segarra, C., Prevosti, A. & Serra, L. (1992) *Evolution (Lawrence, Kans.)* **46**, 1564–1568.
- Mestres, F., Pegueroles, G., Prevosti, A. & Serra, L. (1990) *Evolution (Lawrence, Kans.)* **44**, 1823–1836.
- Mestres, F. & Serra, L. (1999) *J. Zool. Syst. Evol. Res.* **37**, 7–11.
- Mestres, F., Serra, L. & Ayala, F. J. (1995) *Genetics* **140**, 1297–1305.
- Prevosti, A., Serra, L., Ribó, G., Aguadé, M., Sagarra, E., Monclús, M. & García, M. P. (1985) *Evolution (Lawrence, Kans.)* **39**, 838–844.
- Prevosti, A., Ribó, G., Serra, L., Aguadé, M., Balaña, J., Monclús, M. & Mestres, F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5597–5600.
- Prevosti, A., Serra, L., Segarra, C., Aguadé, M., Ribó, G. & Monclús, M. (1990) *Evolution (Lawrence, Kans.)* **44**, 218–221.
- Solé, E., Mestres, F., Balanya, J., Arenas, C. & Serra, L. (2001) *Hereditas* **133**, 65–72.
- Begon, M. (1976) *J. Anim. Ecol.* **45**, 441–456.
- Begon, M. (1976) *Oecologia* **23**, 31–47.
- Mestres, F. & Serra, L. (1995) *Hereditas* **123**, 39–46.
- Ayala, F. J., Serra, L. & Prevosti, A. (1989) *Genome* **31**, 246–255.
- Dubinín, N. P. (1946) *Genetics* **31**, 21–38.
- Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper and Row, New York).
- Robertson, A. & Narain, P. (1971) *Theor. Popul. Biol.* **2**, 24–50.
- Oshima, C. (1961) *Annu. Rep. Natl. Inst. Genet. (Japan)* **12**, 11–12.
- Oshima, C. (1962) *Annu. Rep. Natl. Inst. Genet. (Japan)* **13**, 13.
- Oshima, C. (1962) *Proc. Jpn. Acad.* **38**, 278–283.
- Oshima, C. (1963) *Proc. Jpn. Acad.* **39**, 125–130.
- Oshima, C., Watanabe, T. & Watanabe, T. (1963) *Annu. Rep. Natl. Inst. Genet. (Japan)* **14**, 12.
- Oshima, C. (1967) *Ciència Cultura* **19**, 102–110.
- Golubovsky, M. D. & Victorova, G. V. (1968) *Genetika* **4**, 48–57.
- Oshima, C. (1969) *Jpn. J. Genet.* **44**, Suppl. 1, 209–216.
- Golubovsky, M. D. (1970) *Genetika* **6**, 78–91.
- Watanabe, T. K. & Oshima, C. (1970) *Genetics* **64**, 93–106.
- Choi, Y. & Paik, Y. K. (1983) *Korean J. Genet.* **5**, 35–44.
- Wallace, B. (1966) *Am. Nat.* **100**, 565–578.
- Ives, P. T. (1970) *Evolution (Lawrence, Kans.)* **24**, 507–518.
- Wallace, B. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1374–1378.
- Gustafsson, A. (1947) *Hereditas* **33**, 573–575.
- Mukai, T. & Burdick, A. B. (1959) *Genetics* **44**, 211–232.
- Bryant, S. H. (1976) *Genetics* **84**, 777–786.
- Watanabe, T. K. (1967) *Jpn. J. Genet.* **43**, 375–386.
- Pegueroles, G., Mestres, F. & Serra, L. (1996) *Genetica* **98**, 289–296.
- Mestres, F., Sanz, J. & Serra, L. (1998) *Hereditas* **128**, 105–113.
- Jungen, H. (1968) *Arch. Jul. Klaus-Stift.* **43**, 3–54.
- Krimbas, C. B. (1993) *Drosophila subobscura: Biology, Genetics, and Inversion Polymorphism* (Verlag Dr. Kovac, Hamburg, Germany).
- Powell, J. R. (1997) *Progress and Prospects in Evolutionary Biology* (Oxford Univ. Press, New York).