

Effect of Inversion Polymorphism on the Neutral Nucleotide Variability of Linked Chromosomal Regions in *Drosophila*

Arcadio Navarro,¹ Antonio Barbadilla and Alfredo Ruiz

Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Manuscript received August 15, 1999

Accepted for publication February 14, 2000

ABSTRACT

Recombination is a main factor determining nucleotide variability in different regions of the genome. Chromosomal inversions, which are ubiquitous in the genus *Drosophila*, are known to reduce and redistribute recombination, and thus their specific effect on nucleotide variation may be of major importance as an explanatory factor for levels of DNA variation. Here, we use the coalescent approach to study this effect. First, we develop analytical expressions to predict nucleotide variability in old inversion polymorphisms that have reached mutation-drift-flux equilibrium. The effects on nucleotide variability of a new arrangement appearing in the population and reaching a stable polymorphism are then studied by computer simulation. We show that inversions modulate nucleotide variability in a complex way. The establishment of an inversion polymorphism involves a partial selective sweep that eliminates part of the variability in the population. This is followed by a slow convergence to the equilibrium values. During this convergence, regions close to the breakpoints exhibit much lower variability than central regions. However, at equilibrium, regions close to the breakpoints have higher levels of variability and differentiation between arrangements than regions in the middle of the inverted segment. The implications of these findings for overall variability levels during the evolution of *Drosophila* species are discussed.

CHROMOSOMAL inversion polymorphisms have been a cornerstone in the study of evolution all through the history of population genetics. Since the establishment of the modern synthesis, inversions have been a privileged system to study such diverse subjects as phylogenies, geographical clines, temporal cycles, meiotic drive, and, of course, to look for evidence of natural selection (see Krimbas and Powell 1992 for a review). In fact, the first studies on the role of natural selection in the maintenance of genetic polymorphisms, either in nature or in experimental populations, used inversions because they could be detected by means of simple cytological techniques and their frequency changes could easily be followed through the generations (Dobzhansky 1970; Lewontin 1981). The onset of electrophoresis and the allozymes era was followed by an intense search for linkage disequilibria between allozyme loci associated with inversion polymorphisms and between allozyme loci and the inversions themselves, because they were thought to be generated by epistatic selection (Prakash and Lewontin 1968; Zapata and Álvarez 1987, 1992, 1993; Krimbas and Powell 1992). Now, in the DNA era, inversions may still be useful places to look for selection (Kreitman

and Wayne 1994; Depaulis *et al.* 1999), because, by reducing recombination, inversions can act as amplifiers of the effects on nucleotide polymorphism of selective phenomena such as hitchhiking with favorable mutations (Kaplan *et al.* 1989; Aquadro and Begun 1993; Aquadro *et al.* 1994) or deleterious background selection (Charlesworth *et al.* 1993; Charlesworth 1994; Hudson 1994; Hudson and Kaplan 1995).

Recombination affects levels of nucleotide polymorphism. In *Drosophila*, it accounts for one-quarter of the variance among genes in nucleotide diversity (see Moriyama and Powell 1996 for a review) and an increasing amount of evidence of the same trend is being gathered in other organisms, such as *Mus domesticus* (Nachman 1997), several species of the genus *Lycopersicon* (Stephan and Langley 1998), and humans (Nachman *et al.* 1998). It is also well established that recombination rates are strongly influenced by inversions (Sturtevant and Beadle 1936), the main reason being that, in the heterokaryotypes, crossing-over events within the inversion loop give rise to nonfunctional or nonviable aneuploid meiotic products, and recombination results only from multiple crossovers and gene conversion. Also, inversions are ubiquitous in *Drosophila*: more than three-quarters of all the species in the genus are polymorphic for paracentric inversions (Sperlich and Pfriem 1986; Krimbas and Powell 1992). It is, therefore, clear that inversions can strongly influence nucleotide polymorphism levels. This influence can take several forms. First, inversions reduce and redistribute recombination in heterokaryotypes (Navarro and

Corresponding author: Arcadio Navarro, Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Lab, King's Bldg's., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom.
E-mail: arcadi@holyrood.ed.ac.uk

¹ *Present address:* Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom.

Ruiz 1997; Navarro *et al.* 1997). Hence, the dynamics of selective sweeps and background selection in chromosomes segregating for different arrangements may be different—and their effects conceivably larger—than in chromosomes without inversions. Second, recombination is not uniformly distributed along chromosomes (Lindsley and Sandler 1977; Ashburner 1989; True *et al.* 1996) and when inversions change the position of genes, they are also changing their recombinational context in homokaryotypes. Third, because inversion polymorphisms are maintained by balancing selection (Dobzhansky 1970), they will increase the average life expectancy of nucleotide variability linked to them (Strobeck 1983; Hudson and Kaplan 1988; Kaplan *et al.* 1988). Finally, the latter effect can be just the opposite with new inversions. A recently appeared inversion increasing its frequency may produce a selective sweep that can potentially eliminate variability in large segments of the chromosome. All these different factors may have powerful and contradictory effects on variability. The aims of this work are to obtain theoretical predictions concerning the amount and pattern of nucleotide variability associated with inversion polymorphisms and to shed some light on the overall effect of inversions on the level of nucleotide variability within species.

The basic tools to carry out such studies have been developed in recent years using the coalescent approach. Theoretical and simulation studies concerning DNA variability under balancing selection (Hudson and Kaplan 1988; Kaplan *et al.* 1988; Hey 1991; Nordborg 1997) or under subdivision and migration (Slatkin 1987; Strobeck 1987; Tajima 1989a, 1993; Notohara 1990; Nordborg 1997) are providing a detailed picture of the properties of the amount of DNA polymorphism in a population. Although the analogy between inversion systems and balancing selection or subdivided populations is clear, because all of them produce a structured population, the coalescent approach has never been explicitly applied to inversions. One of the causes of this vacuum may be found in the scarcity and somewhat contradictory nature of empirical information about the degree of exchange of genetic information between arrangements along the inverted chromosome (Krimbas and Powell 1992; Navarro *et al.* 1997). Also, the lack of detailed theoretical predictions of the effect of inversions on recombination makes it difficult to obtain realistic recombination values for every position along the inverted chromosomal region. A recent theoretical study from Navarro *et al.* (1997) provides such results. Given the physical and genetic lengths of an inversion, theoretical predictions of recombination and gene flux caused by crossing over and gene conversion between arrangements can be obtained for every site along the chromosome for heterokaryotypes.

The work presented here deals with the effect on neutral nucleotide variability of both new and old inversion polymorphisms. That is, we exclusively consider

the effect on variability of inversions themselves. We focus on the two most common measures of DNA variability: the number of segregating sites and the average number of pairwise differences in a sample of DNA sequences (Watterson 1975; Tajima 1983, 1993). We study these two variability measures in a population of DNA sequences linked to a chromosome segregating for two arrangements. First, we develop analytically equations for the mutation-drift-flux equilibrium case, in which the inversion polymorphism is precisely balanced (*i.e.*, inversion frequencies do not change from one generation to another). It is assumed that the polymorphism was established enough time ago for the DNA variability in the population to have reached mutation-drift equilibrium. Second, we explore by means of computer simulation cases in which DNA variability is not at equilibrium because the inversion polymorphism was recently established. Changing the age of the polymorphism in the simulations allows us to study the approach to the equilibrium values of DNA variability previously derived using analytical methods.

MODELS AND METHODS

We study the properties of a sample of n DNA sequences at a locus linked to a chromosome segregating for two arrangements, *Standard* (St) and *Inversion* (In), at frequencies p and q , respectively. The two arrangements differ by a single paracentric inversion and St is the oldest one. We denote by N the population size and by Φ the per generation probability of gene exchange between arrangements, *i.e.*, the probability that a DNA sequence recombines with the inversion, which only happens in heterokaryotypes and that is referred to as the probability of gene flux (Navarro *et al.* 1997). It is assumed that karyotype frequencies are maintained approximately at Hardy-Weinberg equilibrium. Accordingly, the probability that a DNA sequence linked to a St chromosome descends from a sequence linked to an In chromosome in the previous generation is $q\Phi$; the converse probability for an In chromosome is $p\Phi$. Following the infinite-sites model (Kimura 1969), we assume that the DNA sequence is so large that every new mutation takes place in a previously unmutated site. The mutants are selectively neutral and μ is the mutation rate per sequence and per generation.

In the development of our analytical results, we make use of the analogy between inversion polymorphisms and subdivided populations. Thus, although results can be obtained in several other ways (see, for example, Hudson and Kaplan 1988; Hudson 1990; or Nordborg 1997), we follow Tajima's (1989a, 1993) approach to obtain equations giving the number of segregating sites in a sample of n alleles taken at random from a population that has reached mutation-drift-flux equilibrium. For the simulation studies, we use the standard principles of the coalescent process to construct the

genealogical tree of a sample and the associated time for each branch (Hudson 1990). The analogy between inversion polymorphisms and subdivided populations is also used to adapt the general coalescent process to a population segregating for two chromosomal arrangements.

We follow a method analogous to the one described in Strobeck (1987) to construct phylogenetic trees for samples taken from such a population. For every sample, the simulation starts by generating a maximum of four random numbers, each derived from the appropriate exponential distribution, which represent the time until one of the four possible events affecting the sample (no simultaneous events are allowed): the time until the most recent flux event (for each arrangement with one or more alleles in the sample, $t_{\Phi(St \rightarrow In)}$ and $t_{\Phi(In \rightarrow St)}$) and the time until the most recent coalescence event (within each arrangement with two or more alleles in the sample, $t_{C(St)}$ and $t_{C(In)}$). The smallest of these four times is chosen and the sample is modified by the creation of the corresponding branches and nodes. The chosen time is associated with the newly created branches and the process starts over. The simulation stops when the most recent common ancestor of the sample is reached.

To study nucleotide variability in a new arrangement appearing in the population and reaching a stable polymorphism, we use the simulation method outlined in Braverman *et al.* (1995) after adapting it to overdominant, instead of directional, selection. Let the three karyotypes St/St , St/In , and In/In have fitnesses $1 - s_1$, 1 , and $1 - s_2$, respectively. We start by constructing the tree in the way described in the previous paragraph, with the frequencies of the arrangements in the population being $\hat{p} = s_2/(s_1 + s_2)$ and $\hat{q} = s_1/(s_1 + s_2)$. At a given time, we make the simulation enter a selective phase as described in Braverman *et al.* (1995). At the beginning of this phase the frequency of arrangement In is changed to $q = s_1/(s_1 + s_2) - \epsilon$, where $\epsilon = 1/2N$ (Braverman *et al.* 1995), that is, we remove one of the In gametes from the population. Then, the deterministic equation for Δq under overdominant selection is used to change allele frequencies every generation. Because the simulation works backward in time, by removing a gamete we allow the frequency equilibrium to be broken and q to decrease deterministically. The process continues until $q \leq \epsilon$, *i.e.*, until only one In gamete is left. At this point, this In gamete is mutated to St and the selective phase is exited. The standard coalescent process starts over again in a population formed exclusively by St chromosomes.

The events of coalescence or gene flux between arrangements during the selective phase are simulated in the same way as in Braverman *et al.* (1995). We make time change in a per generation basis. For each generation the probabilities of the four possible events are computed. Subtracting the sum of the four probabilities from one gives us the probability of no events taking

place during that generation. The probabilities of zero events are multiplied every generation until the product is less than a random number drawn from a uniform distribution between zero and one. When that happens, one of the four events is chosen, taking into account the probability of that event in the current generation. Of course, the simulation is also exited if the most recent common ancestor of the sample is reached during the selective phase. The main difference between our algorithm and that used by Braverman *et al.* (1995), which considered directional selection, is that the differential equation describing the change of allele frequencies with time under overdominant selection lacks an analytical solution (Nei 1987; Nagylaki 1992) and, therefore, we compute the transition probabilities and change q on a per generation basis.

We illustrate and discuss our results using parameter values from *Drosophila* because most of the evidence on inversions and on nucleotide variability comes from this genus. The mutation rate per nucleotide per generation of *Drosophila melanogaster* ranges between 10^{-8} and 10^{-9} (Powell 1997). In the same species, the population size has been estimated to be of the order of 10^6 (Kreitman 1983; Powell 1997), the average $\theta (= 4N\mu)$ per nucleotide being ~ 0.005 (Hudson 1993). To avoid the use of many decimals we will focus all through this article on alleles of 100 nucleotides and, therefore, on a θ value of 0.5.

Inversions affect our model by modifying gene flux rates all along the inverted segment. According to Navarro *et al.* (1997) and Navarro and Ruiz (1997), the gene flux per nucleotide and per generation between arrangements will range between $\Phi = 10^{-2}$ in the center of a large inversion and $\Phi = 10^{-8}$ in regions close to the breakpoints of a short inversion. This predicted range includes most of the empirically estimated gene flux values available in the literature: 10^{-4} in the central region of inversion $In(3L)$ Payne of *D. melanogaster* (Payne 1924); 10^{-5} in the central region of inversion $In(3R)P_{18}$ of *D. melanogaster* (Chovnick 1973); and 10^{-7} near the breakpoints of O_{3+4}/O_{St} heterokaryotypes in *D. subobscura* (Rozas and Aguadé 1994). Details on how to obtain Φ values for any site along the chromosome can be found in Navarro *et al.* (1997).

ANALYTICAL RESULTS

We use the coalescent approach to study the variability of n DNA sequences, among which i sequences are randomly chosen from St chromosomes and $j (= n - i)$ sequences from In chromosomes. Let $Q(i, j)$ represent the state of the sample. In terms of the genealogical relationships of the sequences in the sample, that is, going back in the past, there are four possible adjacent states into which $Q(i, j)$ can move in a single generation, namely, $Q(i - 1, j)$, $Q(i, j - 1)$, $Q(i - 1, j + 1)$, and $Q(i + 1, j - 1)$. The first two changes represent common

ancestor events and the latter gene flux events. The probabilities of these events are (derived following Hudson 1983 and Tajima 1989a):

$$P\{Q(i, j) \text{ to } Q(i - 1, j)\} = \frac{i(1 - j)}{4Np} \quad (1a)$$

$$P\{Q(i, j) \text{ to } Q(i, j - 1)\} = \frac{j(1 - i)}{4Nq} \quad (1b)$$

$$P\{Q(i, j) \text{ to } Q(i - 1, j + 1)\} = iq\Phi \quad (1c)$$

$$P\{Q(i, j) \text{ to } Q(i + 1, j - 1)\} = jp\Phi. \quad (1d)$$

It follows from these equations that any $Q(i, j)$ will finally converge to $Q(1, 0)$ or $Q(0, 1)$ unless $\Phi = 0$.

Let $S(i, j)$ be the expected number of segregating sites in a sample in state $Q(i, j)$ taken at random from a population at mutation-drift-flux equilibrium. Given the infinite-sites model, the number of segregating sites is the number of mutations that take place while $Q(i, j)$ is converging to $Q(1, 0)$ or $Q(0, 1)$. To calculate this number we must first consider the sojourn time of the sample, *i.e.*, the expected number of generations during which $Q(i, j)$ does not change. The probability that $Q(i, j)$ changes to one of the four adjacent states in a single generation is

$$P(i, j) = iq\Phi + jp\Phi + \frac{i(i - 1)}{4Np} + \frac{j(j - 1)}{4Nq}. \quad (2)$$

Therefore, the sojourn time until the first change is geometrically distributed with mean $1/P(i, j)$ generations. While there are n alleles, $n\mu$ mutations take place every generation; hence, $C(i, j)$, the expected number of mutations taking place while $Q(i, j)$ remains the same, is

$$C(i, j) = \frac{n\mu}{P(i, j)} = \frac{n\theta pq}{pj(F + j - 1) + qi(F + i - 1)}, \quad (3)$$

where $F = 4N\Phi pq$ and $\theta = 4N\mu$.

Given that $Q(i, j)$ changes, the conditional probabilities that it changes to each one of the four adjacent states are easily obtained from Equation 1. With those probabilities and (3) we can readily obtain an iterative expression for $S(i, j)$,

$$S(i, j) = \frac{\theta(i + j)pq + pja_{ij} + qib_{ij}}{pj(F + j - 1) + qi(F + i - 1)}, \quad (4)$$

where

$$a_{ij} = (j - 1)S(i, j - 1) + FS(i + 1, j - 1)$$

and

$$b_{ij} = (i - 1)S(i - 1, j) + FS(i - 1, j + 1).$$

Of course, $S(1, 0) = S(0, 1) = 0$ by definition. Equation 4 can also be obtained from previous results on balancing selection (Hudson and Kaplan 1988; Kaplan *et al.* 1988; Hey 1991; Nordborg 1997).

From (4), $S(i, j)$ can be computed for every value of i and j . For instance, when $n = 2$,

$$S(2, 0) = \frac{\theta p + FS(1, 1)}{1 + F} \quad (5a)$$

$$S(1, 1) = \frac{2\theta pq + pFS(2, 0) + qFS(0, 2)}{F} S(0, 2) \quad (5b)$$

$$= \frac{\theta q + FS(1, 1)}{1 + F}. \quad (5c)$$

And solving these equations, we get

$$S(2, 0) = \theta \left(1 + \frac{q(p - q)}{1 + F} \right) \quad (6a)$$

$$S(1, 1) = \theta \left(1 + \frac{2pq}{F} \right) = \theta \left(1 + \frac{1}{2N\Phi} \right) \quad (6b)$$

$$S(0, 2) = \theta \left(1 + \frac{p(q - p)}{1 + F} \right). \quad (6c)$$

To ascertain the effect of inversions on nucleotide variability in the population as a whole, we must consider the expected number of segregating sites in a sample of n sequences taken at random from the entire population, $S(n)$. Because we are assuming that chromosome arrangements are in Hardy-Weinberg equilibrium frequencies, $S(n)$ can be easily obtained from

$$S(n) = \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} S(i, n - i). \quad (7)$$

In addition, Equation 6 (and 7 making $n = 2$) gives us the average number of pairwise differences between the alleles in our sample, $E(k)$, which is equal to the expected number of segregating sites among a sample of two alleles (Hudson 1990; Tajima 1993). $E(k)$ has an advantage over S : under a neutral model with no recombination, it gives a direct estimate of θ (Tajima 1993); that is, it is independent of n . On the other hand, when the neutral theory is correct and there is no population subdivision, we can obtain an estimate of θ from S in the following way (Watterson 1975):

$$\hat{\theta} = \frac{S}{\sum_{i=1}^{n-1} 1/i} \quad (8)$$

Both variability measures, $E(k)$ and $\hat{\theta}$, combine in Tajima's method for testing the neutral mutation hypothesis (Tajima 1989b) and are quite useful for studying the effect of inversions on DNA variability. Equations for the variance of $k(n)$ under some specific conditions can be found in the appendix.

The results presented so far allow us to study the effect on variability of a precisely balanced inversion polymorphism that reached mutation-drift-flux equilibrium a long time ago. Table 1 gives the values of $\hat{\theta}$ and $E(k)$ together with its standard deviation (when formulas are available; see appendix) under different arrangement frequencies and under different sample

TABLE 1
Theoretical expectations for mutation-flux-drift equilibrium

<i>In</i> frequency	Φ	<i>n</i>	<i>St</i>		<i>In</i>		Population		
			$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>	
0.5	10^{-8}	2	0.5000	0.5000 (3.6400)	0.5000	0.5000 (3.6400)	13.0000	13.0000	
		10	0.4994	0.5000 (2.3121)	0.4994	0.5000 (2.3121)	9.3500	13.0000	
		20	0.4989	0.5000 (2.1980)	0.4989	0.5000 (2.1980)	7.5800	13.0000	
	10^{-6}	2	0.5000	0.5000	0.5000	0.5000	0.6250	0.6250	
		10	0.4846	0.5000	0.4846	0.5000	0.6033	0.6250	
		20	0.4713	0.5000	0.4713	0.5000	0.5888	0.6250	
	10^{-4}	2	0.5000	0.5000	0.5000	0.5000	0.5013	0.5013	
		10	0.4991	0.5000	0.4991	0.5000	0.5012	0.5013	
		20	0.4979	0.5000	0.4979	0.5000	0.5012	0.5013	
	10^{-2}	2	0.5000	0.5000 (0.8603)	0.5000	0.5000 (0.8603)	0.5000	0.5000	
		10	0.5000	0.5000 (0.5229)	0.5000	0.5000 (0.5229)	0.5000	0.5000	
		20	0.5000	0.5000 (0.4960)	0.5000	0.5000 (0.4960)	0.5000	0.5000	
	0.8	10^{-8}	2	0.2615	0.2615 (2.8996)	0.5596	0.5596 (2.9755)	8.5286	8.5286
			10	0.2613	0.2615	0.5594	0.5596	8.3995	8.5286
			20	0.2611	0.2615	0.5592	0.5596	7.4761	8.5286
10^{-6}		2	0.3537	0.3537	0.5366	0.5366	0.5976	0.5976	
		10	0.3401	0.3537	0.5298	0.5366	0.5901	0.5976	
		20	0.3283	0.3537	0.5237	0.5366	0.5810	0.5976	
10^{-4}		2	0.4963	0.4963	0.5009	0.5009	0.5012	0.5012	
		10	0.4929	0.4963	0.5007	0.5009	0.5012	0.5012	
		20	0.4886	0.4963	0.5004	0.5009	0.5012	0.5012	
10^{-2}		2	0.5000	0.5000 (0.8660)	0.5000	0.5000 (0.8660)	0.5000	0.5000	
		10	0.4999	0.5000	0.5000	0.5000	0.5000	0.5000	
		20	0.4999	0.5000	0.5000	0.5000	0.5000	0.5000	
0.1		10^{-8}	2	0.5399	0.5399 (2.3074)	0.1415	0.1415 (2.1742)	5.0287	5.0287
			10	0.5398	0.5399	0.1412	0.1415	6.2680	5.0287
			20	0.5397	0.5399	0.1411	0.1415	6.6926	5.0287
	10^{-6}	2	0.5294	0.5294	0.2353	0.2353	0.5662	0.5662	
		10	0.5266	0.5294	0.2281	0.2353	0.5698	0.5662	
		20	0.5240	0.5294	0.2218	0.2353	0.5691	0.5662	
	10^{-4}	2	0.5011	0.5011	0.4903	0.4903	0.5012	0.5012	
		10	0.5010	0.5011	0.4832	0.4903	0.5012	0.5012	
		20	0.5008	0.5011	0.4748	0.4903	0.5012	0.5012	
	10^{-2}	2	0.5000	0.5000 (0.8660)	0.4999	0.4999 (0.8660)	0.5000	0.5000	
		10	0.5000	0.5000	0.4998	0.4999	0.5000	0.5000	
		20	0.5000	0.5000	0.4997	0.4999	0.5000	0.5000	

E(k) and $\hat{\theta}$ for *n* DNA sequences sampled at random from either *St* chromosomes, *In* chromosomes, or the entire population. When formulas are available (see appendix), standard deviations (σ) can be found within parentheses. $N = 10^6$, $\mu = 1.25 \times 10^{-7}$ ($\theta = 0.5$).

sizes for samples taken at random either from the entire population or from a given arrangement class. As expected, the behavior of $\hat{\theta}$ is dependent on the sample size, mainly for low flux rates, while *E(k)* does not depend on *n*. Of course, the standard deviation of *k* de-

creases with increasing sample size, although it remains exceedingly large, as always happens with pairwise measures, mainly when no intragenic recombination is allowed (Tajima 1983; Hudson 1990).

Flux rates affect the variability in the population as a

TABLE 2
Theoretical expectations for mutation-flux-drift equilibrium

Φ	Inversion equilibrium frequencies		
	0.5	0.8	0.1
10^{-8}	25.5000 (25.7536)	25.5000 (25.6667)	25.5000 (25.5984)
10^{-6}	0.7500 (1.0897)	0.7500 (1.0771)	0.7500 (1.0623)
10^{-4}	0.5025 (0.8682)	0.5025 (0.8682)	0.5025 (0.8682)
10^{-2}	0.5000 (0.8660)	0.5000 (0.8660)	0.5000 (0.8660)

Average number of nucleotide differences, $E(k)$, between a *St* allele and an *In* allele. Standard deviations (σ) are within parentheses. $N = 10^6$, $\mu = 1.25 \times 10^{-7}$ ($\theta = 0.5$).

whole, which increases as flux decreases. Flux rates of 10^{-2} or higher make $E(k)$ and $\hat{\theta}$ equal to their values in a population without inversions (in our case $\theta = 0.5$). On the other hand, flux rates $<10^{-4}$ produce a rapid departure of DNA variability from its state in a population without inversions. Given that heterokaryotypes always have large regions with gene flux rates $<10^{-4}$, for instance, regions around inversion breakpoints (Navarro *et al.* 1997), it is obvious that inversions will have a strong effect on nucleotide variability.

The frequency of the chromosome arrangements in the population also has a remarkable effect on variability. The maximum values of $E(k)$ and $\hat{\theta}$ for the whole population are reached with intermediate inversion frequencies (Table 1). Table 1 shows further effects of inversions. We can see that, with intermediate frequencies, variability is scarcely reduced within each arrangement when compared to variability in a population without inversions (in which $\theta = 0.5$). On the other hand, if frequencies are not intermediate, variability in the lowest frequency arrangement is notably reduced. This reduction is caused by the diminished population size of the gametes carrying each arrangement, which boosts the loss of genetic variability by drift. In contrast, the variability of the most frequent arrangement increases. This variability increase tends to balance the low variability in the less frequent arrangement because of gene flux acting as conservative migration, which agrees with the invariant property of structured populations noted by Mayurama (1971) and Nagylaki (1982). Also, it is worth noting that variability in the low frequency arrangement, though reduced, is still higher than that expected in an isolated population of the same size.

Both variability augments can be explained by the same mechanism. With low flux and a lot of drift (mainly in the low frequency arrangement), the two kinds of chromosomes are highly differentiated and, therefore, almost every allele coming by recombination from the other arrangement will be absent in the recipient arrangement. These new alleles add new variability at a higher rate than mutation. This effect overpowers drift and increases with decreasing flux. It only disappears with gene flux rates $\ll 10^{-8}$ (*i.e.*, very close to zero and

smaller than the mutation rate we assume). In that case $E(k)$ and $\hat{\theta}$, both in *St* and *In*, converge to their values in a neutral population of sizes Np and Nq , respectively.

Differentiation between arrangements can be measured by means of the number of pairwise differences between an *In* allele and a *St* allele (Equation 6b). As we can see in Table 2, equilibrium pairwise differences between arrangements do not depend on inversion frequencies and standard deviations are practically unaffected by them, which agrees with previous results (Nordborg 1997).

SIMULATION RESULTS

The simulation program described in models and methods allows us to obtain $E(k)$ and $\hat{\theta}$ for different values of the selection coefficients and the age of the inversion polymorphism. Using this tool, we are able to study the approach to mutation-drift-flux equilibrium for nucleotide variability in any given inversion polymorphism. Taking into account the size of the standard deviations we are dealing with (Tables 1 and 2), all the values given in this section are based on 100,000 runs of our program.

In Tables 3 and 4 we can see the values of $E(k)$ and $\hat{\theta}$ 100 generations after the stabilization of three different polymorphisms (with the frequencies of the new arrangement, *In*, being 0.5, 0.8, and 0.1), that is, 100 generations after the end of the selective phase that produced the stable inversion polymorphisms we are studying. During this short amount of time, no new variability has appeared in the population and gene flux has had no time to homogenize the variability between arrangements. Thus, the footprint left by the origin of the new inversion is still visible. As expected, it is quite similar to the trail generated by a selective sweep (Braverman *et al.* 1995). However, given that we are considering overdominant selection, our sweep is partial; *i.e.*, it stops before the inversion reaches fixation and, therefore, it does not eliminate all the genetic variability in the population. *St* chromosomes are left with an amount of variability proportional to their frequency in the population. Moreover, some of the nucleotide variability is

TABLE 3
Simulation results

<i>In</i> frequency	Φ	<i>n</i>	<i>St</i>		<i>In</i>		Population	
			$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>
0.5 (<i>s</i> ₁ = 0.1, <i>s</i> ₂ = 0.1)	10 ⁻⁸	2	0.2493	0.2493	0.0001	0.0001	0.1888	0.1888
		10	0.2496	0.2496	0.0007	0.0006	0.1988	0.1890
		20	0.2472	0.2473	0.0023	0.0013	0.2087	0.1909
	10 ⁻⁶	2	0.2501	0.2501	0.0001	0.0001	0.1894	0.1894
		10	0.2486	0.2482	0.0009	0.0007	0.1969	0.1867
		20	0.2495	0.2505	0.0025	0.0014	0.2074	0.1907
	10 ⁻⁴	2	0.2486	0.2486	0.0118	0.0118	0.1921	0.1921
		10	0.2481	0.2478	0.0197	0.0118	0.1997	0.1900
		20	0.2503	0.2500	0.0324	0.0130	0.2084	0.1919
10 ⁻²	2	0.2497	0.2497	0.2523	0.2523	0.2497	0.2497	
	10	0.2450	0.2460	0.2469	0.2478	0.2500	0.2519	
	20	0.2478	0.2475	0.2487	0.2483	0.2498	0.2509	
0.8 (<i>s</i> ₁ = 0.1, <i>s</i> ₂ = 0.025)	10 ⁻⁸	2	0.0993	0.0993	0.0002	0.0002	0.0355	0.0355
		10	0.0994	0.0989	0.0019	0.0014	0.0490	0.0370
		20	0.0993	0.0996	0.0051	0.0029	0.0618	0.0389
	10 ⁻⁶	2	0.0991	0.0991	0.0002	0.0002	0.0375	0.0375
		10	0.0992	0.0996	0.0017	0.0013	0.0498	0.0374
		20	0.0991	0.0989	0.0052	0.0015	0.0600	0.0377
	10 ⁻⁴	2	0.0987	0.0987	0.0049	0.0049	0.0371	0.0371
		10	0.0999	0.1001	0.0097	0.0060	0.0527	0.0397
		20	0.0991	0.0990	0.0172	0.0075	0.0618	0.0395
10 ⁻²	2	0.0962	0.0962	0.0926	0.0926	0.0938	0.0938	
	10	0.1000	0.0988	0.1016	0.1005	0.0999	0.0977	
	20	0.1104	0.1017	0.1106	0.1017	0.1107	0.1026	
0.1 (<i>s</i> ₁ = 0.0111, <i>s</i> ₂ = 0.1)	10 ⁻⁸	2	0.4476	0.4476	0.0004	0.0004	0.4434	0.4434
		10	0.4507	0.4506	0.0039	0.0030	0.4464	0.4502
		20	0.4490	0.4501	0.0110	0.0065	0.4453	0.4495
	10 ⁻⁶	2	0.4497	0.4497	0.0006	0.0006	0.4419	0.4419
		10	0.4474	0.4467	0.0039	0.0031	0.4468	0.4489
		20	0.4489	0.4490	0.0113	0.0065	0.0063	0.4504
	10 ⁻⁴	2	0.4517	0.4517	0.1242	0.1242	0.4461	0.4461
		10	0.4498	0.4507	0.1821	0.1303	0.4500	0.4507
		20	0.4503	0.4520	0.2389	0.1417	0.4486	0.4520
10 ⁻²	2	0.4510	0.4510	0.4537	0.4537	0.4524	0.4524	
	10	0.4493	0.4494	0.4477	0.4482	0.4508	0.4513	
	20	0.4626	0.4550	0.4644	0.4545	0.4632	0.4550	

E(k) and $\hat{\theta}$ for *n* DNA sequences sampled at random either from *St* chromosomes, *In* chromosomes, or the entire population. *N* = 10⁶, μ = 1.25 × 10⁻⁷ (θ = 0.5). The stabilization of arrangement frequencies at the values determined by *s*₁ and *s*₂ took place 100 generations ago.

rescued by gene flux, because it allows *In* chromosomes to receive variants that will otherwise be lost. The smaller the selection coefficients affecting the arrangements, the slower the selective phase and the larger the amount of variability maintained in the population.

However, the variability differences caused by selection coefficients differing by as much as an order of magnitude are not very important (compare Tables 3 and 4). The reason for that must be sought in the approach of arrangement frequencies to equilibrium. Under overdominant selection, *In* frequencies increase in a sigmoidal way and, therefore, for much of the time since the appearance of the inversion its frequency is either close to zero, which makes it irrelevant, or close

to the equilibrium point, $\hat{q} = s_1 / (s_1 + s_2)$, which does not change with the magnitude of the selection coefficients. The only relevant variability differences are built up during the lineal increment phase and the amount of time spent in that phase is always short when compared to the total length of the genealogy. For example, when *s*₁ = *s*₂ = 0.1, 57 generations are needed for an inversion to increase its frequency from 0.05 to 0.45. When *s*₁ = *s*₂ = 0.01, the same increment needs 600 generations; *i.e.*, selection coefficients make a difference of 543 generations only in a tree that can have >10⁵ generations. Hence, from now on we use the data in Table 3 as the starting point to study the approach to mutation-drift-flux equilibrium.

TABLE 4
Simulation results

<i>In</i> frequency	Φ	<i>n</i>	<i>St</i>		<i>In</i>		Population	
			$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>	$\hat{\theta}$	<i>E(k)</i>
0.5 (<i>s</i> ₁ = 0.01, <i>s</i> ₂ = 0.01)	10 ⁻⁸	2	0.2509	0.2509	0.0008	0.0008	0.1854	0.1854
		10	0.2479	0.2481	0.0078	0.0060	0.2037	0.1920
		20	0.2476	0.2485	0.0232	0.0131	0.2228	0.1964
	10 ⁻⁶	2	0.2484	0.2484	0.0009	0.0009	0.1863	0.1863
		10	0.2483	0.2487	0.0081	0.0063	0.2043	0.1931
		20	0.2488	0.2500	0.0235	0.0133	0.2237	0.1966
	10 ⁻⁴	2	0.2499	0.2499	0.0860	0.0860	0.2088	0.2088
		10	0.2490	0.2496	0.1234	0.0922	0.2150	0.2070
		20	0.2502	0.2498	0.1641	0.1000	0.2326	0.2111
	10 ⁻²	2	0.2486	0.2486	0.2499	0.2499	0.2478	0.2478
		10	0.3319	0.2965	0.3344	0.2979	0.3339	0.2984
	0.8 (<i>s</i> ₁ = 0.01, <i>s</i> ₂ = 0.0025)	10 ⁻⁸	2	0.0990	0.0990	0.0016	0.0016	0.0368
10			0.0980	0.0986	0.0167	0.0128	0.0646	0.0487
20			0.1019	0.1012	0.0501	0.0280	0.1059	0.0609
10 ⁻⁶		2	0.0968	0.0968	0.0018	0.0019	0.0377	0.0377
		10	0.0984	0.0997	0.0170	0.0131	0.0632	0.0473
		20	0.1008	0.1000	0.0503	0.0282	0.1050	0.0605
10 ⁻⁴		2	0.0998	0.0998	0.0328	0.0328	0.0448	0.0448
		10	0.1003	0.1002	0.0680	0.0500	0.0831	0.0649
		20	0.1003	0.0999	0.1211	0.0679	0.1309	0.0789
10 ⁻²		2	0.1261	0.1261	0.1339	0.1339	0.1307	0.1307
		10	0.4652	0.3067	0.4697	0.3095	0.4714	0.3120
0.1 (<i>s</i> ₁ = 0.00111, <i>s</i> ₂ = 0.01)		10 ⁻⁸	2	0.4447	0.4447	0.0040	0.0040	0.4670
	10		0.4447	0.4471	0.0272	0.0224	0.4438	0.4440
	20		0.4488	0.4479	0.0642	0.0427	0.4688	0.4580
	10 ⁻⁶	2	0.4435	0.4435	0.0042	0.0042	0.4457	0.4457
		10	0.4435	0.4466	0.0272	0.0224	0.4452	0.4470
		20	0.4511	0.4513	0.0643	0.0427	0.4626	0.4539
	10 ⁻⁴	2	0.4484	0.4484	0.4343	0.4343	0.4497	0.4497
		10	0.4495	0.4502	0.4685	0.4570	0.4543	0.4542
		20	0.4578	0.4542	0.5301	0.4731	0.4699	0.4600
	10 ⁻²	2	0.4527	0.4527	0.4496	0.4496	0.4524	0.4524
		10	0.4487	0.4490	0.4497	0.4500	0.4480	0.4484

Same as Table 3 with smaller selection coefficients.

The convergence to the equilibrium variability in the population as a whole is drawn in Figure 1a. During the first million generations, almost no new diversity is added to the population. Gene flux, having higher rates than mutation, plays a very important role during this phase because it homogenizes variabilities within the two arrangements. Only after the first several million generations has mutation added enough variability to reach the equilibrium. With high gene flux ($\Phi = 10^{-2}$) the equilibrium point is independent of the frequency of inversions. On the other hand, lower gene flux ($\Phi = 10^{-6}$) makes the equilibrium variabilities higher for intermediate arrangement frequencies. Note that the equilibrium points obtained by simulation are equivalent to those obtained analytically in the previous section.

Figure 1b shows the changes in *E(k)* between two *In* alleles during the convergence to equilibrium. The convergence process within an average inversion is plotted in Figure 2a. As we can see in these figures, gene

flux plays a key role in determining both the amount of variability that is lost during the origin of the inversion polymorphism and the speed at which this lost variability is recovered. With low gene flux, nucleotide variability within the newly appeared arrangement is zero, or very close to zero, during the first 10⁵-10⁶ generations. Convergence to mutation-drift-flux equilibrium is slow because of the scarce amount of variability incoming from *St* chromosomes. On the other hand, with high rates of gene flux the variability within *In* chromosomes is very close to the variability left in *St* chromosomes after the partial sweep and the convergence to equilibrium is faster. Moreover, during the first million generations, gene flux is the main cause of the increase of variability within inversion chromosomes because it adds new variability (imported from standard chromosomes) at higher rates than mutation.

The process that is meanwhile taking place within *St* chromosomes is represented in Figures 1c and 2b. In

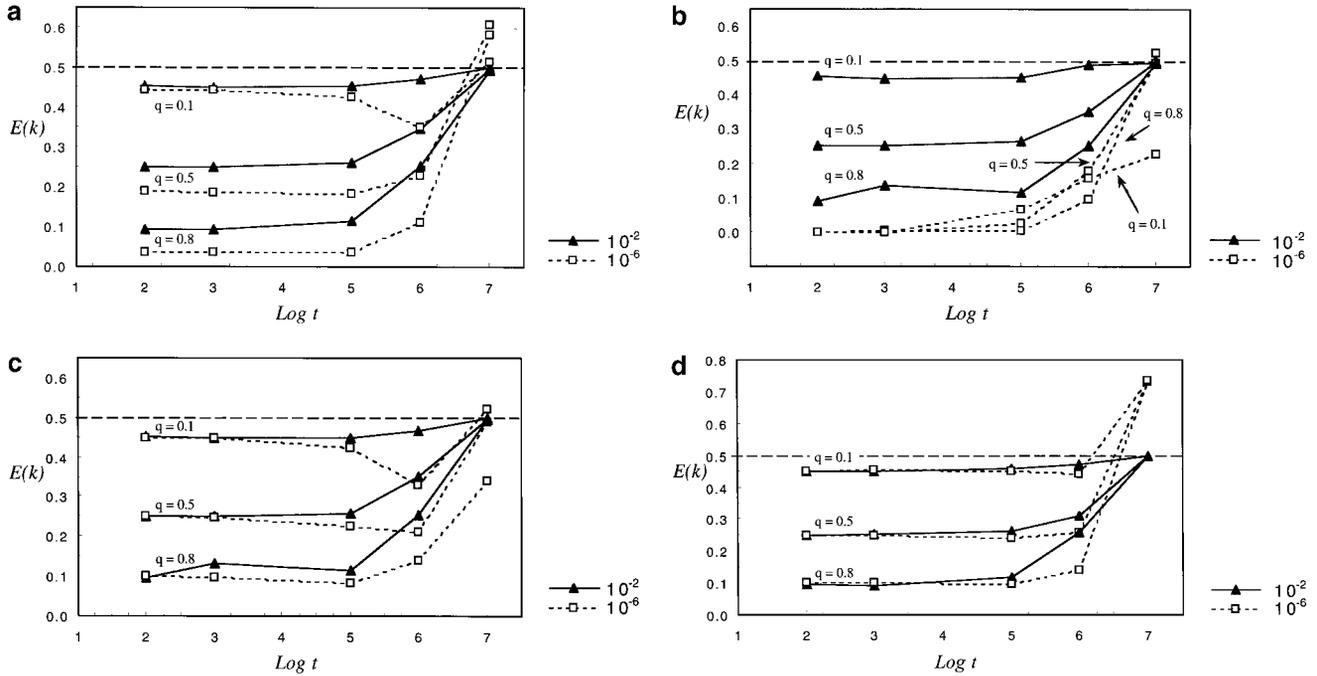


Figure 1.—Simulation results. Approach of nucleotide variability to mutation-drift-flux equilibrium. Abscissa: decimal logarithm of the number of generations since the stabilization of the polymorphism. Ordinate: (a) $E(k)$ for the entire population; (b) $E(k)$ for the *In* chromosomes; (c) $E(k)$ for the *St* chromosomes; (d) $E(k)$ for two alleles taken at random, one from the pool of *St* chromosomes and the other from the pool of *In* chromosomes. $\Phi = 10^{-2}$ stands for gene flux in the center of an average inversion and $\Phi = 10^{-6}$ for gene flux around the breakpoints. $N = 10^6$ and $\mu = 1.25 \times 10^{-7}$, so the equilibrium $E(k)$ for a population without inversions is 0.5 (dotted line).

this case, of course, gene flux has little influence on the initial variability. It does, however, affect the way in which variability changes, as well as the equilibrium points. We can see that with low flux rates, during the first 10^5 – 10^6 generations, variability within *St* chromosomes decreases. This can be explained by a sink-source mechanism: the relatively great allele diversity stored in *St* chromosomes is transferred by flux to *In* chromosomes, where low flux rates forced an initial elimination of variability. This process lasts until the homogenization of the two arrangements; hence, variability decreases within *St* chromosomes while increasing within *In* chromosomes. *In* chromosomes can undergo a similar temporary variability decrease if flux rates are high and the new inversion reaches a high frequency (Figure 1b).

In relation to the time dynamics of the pairwise differences between arrangements, Figures 1d and 2c show that, as proved in analytical results, the equilibrium values of $E(k)$ are dependent only on gene flux. On the other hand, during the first 10^5 – 10^6 generations of polymorphism, the pairwise differences between *In* and *St* chromosomes are dependent only on arrangement frequencies.

DISCUSSION

On its way to the establishment of a balanced polymorphism, a newly arisen inversion sweeps a lot of variability

from the population. Just after the stabilization of the arrangement frequencies, the chromosomes bearing the newly appeared arrangement will have almost no variability (Tables 3 and 4). As the polymorphism grows older, a slow convergence to mutation-drift-flux equilibrium starts. At this equilibrium, the level of DNA polymorphism in the population as a whole can be higher than in a population of the same size without segregating arrangements (Table 1 and Figure 1). Which of these two effects of an inversion, to reduce or to increment variability, will prevail depends on the time that it takes to reach equilibrium. The convergence to the equilibrium values proceeds at very different speeds, strongly depending on gene flux rates, and, thus, it proceeds differently in different regions of the inverted segment.

In regions close to the breakpoints, flux rates are very low (Figure 2) and, therefore (1) the strength of the partial sweep is greater, and almost no variability is left within the new arrangement; and (2) the linkage disequilibria generated by the frequency increment of the new inversion will persist for a long time (Navarro *et al.* 1996). Around the breakpoints, only new mutations supply variability to the new arrangement and it is unlikely that new mutants will be exchanged between arrangements. Therefore, in regions close to the breakpoints of *In* chromosomes, variability will be very low until enough new mutants are added, which will take $>10^6$

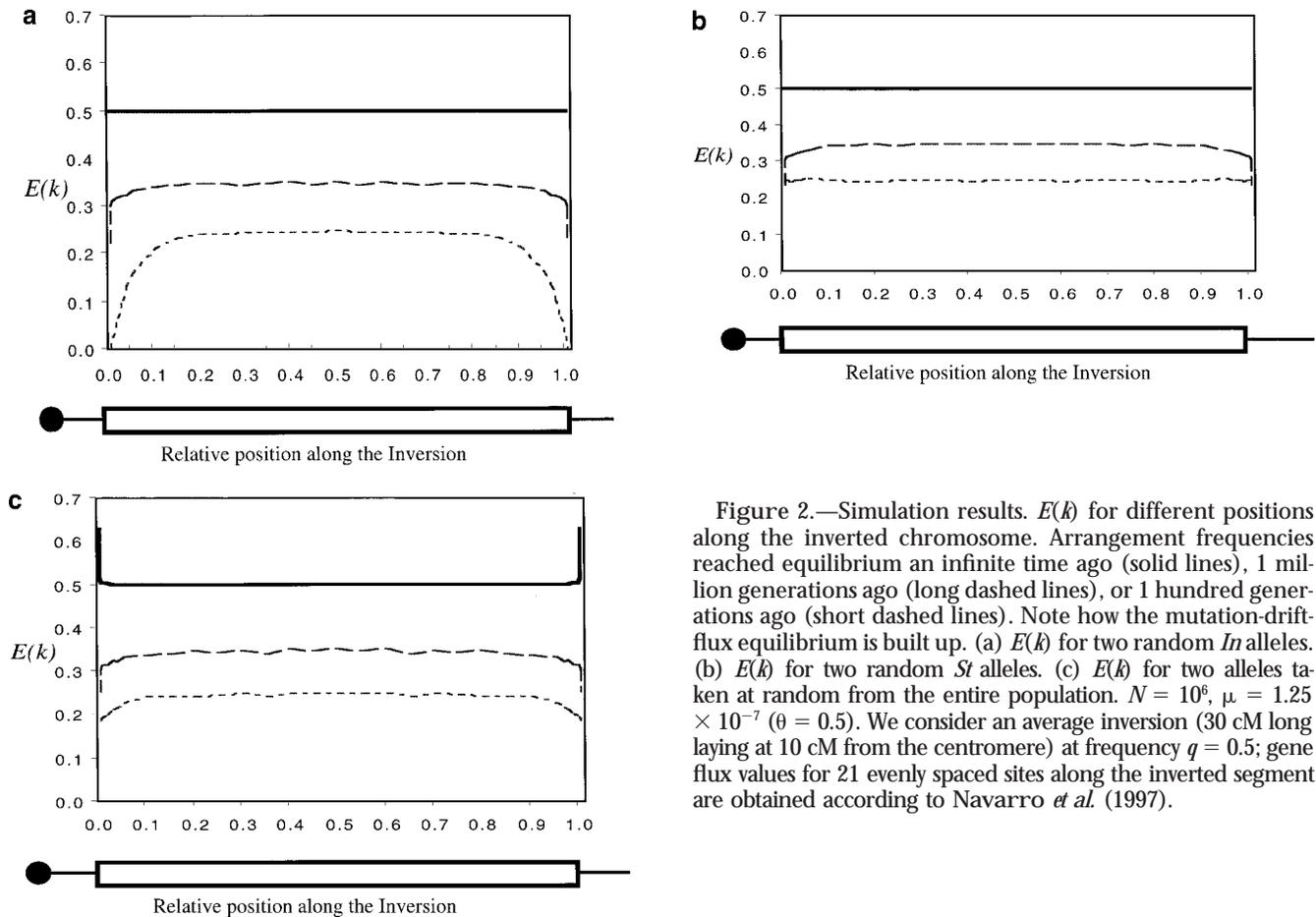


Figure 2.—Simulation results. $E(k)$ for different positions along the inverted chromosome. Arrangement frequencies reached equilibrium an infinite time ago (solid lines), 1 million generations ago (long dashed lines), or 1 hundred generations ago (short dashed lines). Note how the mutation-drift-flux equilibrium is built up. (a) $E(k)$ for two random *In* alleles. (b) $E(k)$ for two random *St* alleles. (c) $E(k)$ for two alleles taken at random from the entire population. $N = 10^6$, $\mu = 1.25 \times 10^{-7}$ ($\theta = 0.5$). We consider an average inversion (30 cM long laying at 10 cM from the centromere) at frequency $q = 0.5$; gene flux values for 21 evenly spaced sites along the inverted segment are obtained according to Navarro *et al.* (1997).

generations (Figures 1b and 2a). Around the breakpoints of *St* chromosomes, variability is initially greater than that of *In* chromosomes, but decreases afterward over several generations (Figure 1c) because gene flux tends to homogenize the two arrangements and no variability was left in the *In* chromosomes. Although nucleotide variability in relation to inversions has been studied by several authors (Aquadro *et al.* 1986; Aguadé 1988; Bénassi *et al.* 1993; Rozas and Aguadé 1993, 1994; Wesley and Eanes 1994; Popadić and Anderson 1995; Popadić *et al.* 1995; Andolfatto *et al.* 1999; Cáceres *et al.* 1999; Depaulis *et al.* 1999; Rozas *et al.* 1999), the only available study in which nucleotide variability was surveyed simultaneously at different positions of a chromosome segregating for two arrangements separated by a single inversion has been carried out by Hasson and Eanes (1996). Our theoretical expectations are consistent with their findings. First, the breakpoints of inversion *In*(3*L*)*Payne* of *D. melanogaster* host 20 times less polymorphism ($\pi = 0.0003$) than the breakpoints of *St* chromosomes ($\pi = 0.0060$). Furthermore, the *Hsp83* gene locus, which is close to, but not exactly at, the distal breakpoint, presents higher levels of variability ($\pi = 0.0067$) and lower levels of differentiation between arrangements (Nei's $d = 0.0053$) than the breakpoint itself ($\pi = 0.0058$, $d = 0.0068$), although the differences are not statistically significant (Hasson and Eanes 1996).

In the population as a whole, nucleotide variability around breakpoints is low during the first 10^5 – 10^6 generations. The higher the frequency of *In* chromosomes, the lower the levels of DNA polymorphism (Figure 1a). At equilibrium, on the other hand, low gene flux rates induce substantial differentiation between *St* and *In* chromosomes, which causes an increment in the level of variability of the whole population (Table 1, Figures 1d and 2c). This enhancement of polymorphism levels is due to the extension of the average lifetime of mutants caused by balancing selection (Hudson and Kaplan 1988; Kaplan *et al.* 1988) and it is greater with lower gene flux and intermediate arrangement frequencies.

Gene flux rates are higher in the center of the inverted regions and hence (1) gene flux preserves some of the starting variability from the initial sweep by sheltering it in the new arrangement; and (2) the differentiation between arrangements will decrease at a steady rate, as new mutations are exchanged and some of the variability stored in *St* chromosomes enters the inversion by gene flux. Inverted chromosomes, therefore, have a higher starting level of polymorphism with higher flux rates (Figure 1b). The smaller the frequency of inversions and the greater the flux, the higher the starting polymorphism level. On the contrary, neither the initial variability within *St* chromosomes (Figure 1c) nor the initial differentiation between *St* and *In* (Figure 1d) is

affected by gene flux rates. The main differences between the central region and the regions around the breakpoints arise from the buildup of the equilibrium in the central region of inversions and on the equilibrium state itself. When equilibrium has been achieved, higher flux rates make the amount of variability in the central zone smaller than that of the regions around breakpoints (Figure 2c). On the other hand, higher flux rates during the convergence to equilibrium allow for a rapid increase in polymorphism levels and a decrease in differences between arrangements, which starts at about generation 10^5 (Figure 1). Again this result is consistent with the finding by Hasson and Eanes (1996) of higher levels of nucleotide variability in the *Est-6* gene ($\pi = 0.0192$), which lies approximately in the middle of inversion *In(3L)Payne*, than at the breakpoints of the inversion ($\pi = 0.0058$). Also, the silent polymorphism levels for *Est-6* were roughly similar in both arrangements ($\pi = 0.0162$ in *St* and $\pi = 0.0200$ in *In*).

In this analysis we have focused on neutral variability without considering any explicit source for the overdominance of the inversion. The selective maintenance of inversion polymorphisms has been the subject of abundant theoretical work. Some models consider associations of the inversion with either a single gene or a group of genes with additive relationships (e.g., Nei *et al.* 1967; Ohta and Kojima 1968). Under those models, the establishment of an inversion polymorphism would be a rare event, because, eventually, the linkage disequilibrium between the selected loci and the inversion would break down, rendering the polymorphism unstable and allowing the inversion to drift away until fixation or loss. The most widely accepted models for the maintenance of inversion polymorphisms consider their association with a complex of genes where epistatic selection maintains gametic disequilibrium. Essentially, a new inversion can reach a stable polymorphism only if it occurs in chromosomes carrying an excess gametic type (Charlesworth and Charlesworth 1973; Charlesworth 1974; see Krimbas and Powell 1992 for a review). Under these conditions, a stable polymorphism can be reached because certain recombination events within heterokaryotypes generate unfavored gametic types that would be eliminated by selection. This would have the additional effect of reducing recombination even further and, thus, increasing differentiation between arrangements.

Inversions are the most common form of chromosomal change in the evolutionary history of *Drosophila*. More than 28,000 paracentric inversions are estimated to be currently segregating in natural populations of *Drosophila* and >42,000 paracentric inversions have become fixed during the evolution of the genus (Sorsa 1988). To what extent has all this continuous chromosomal reorganization affected DNA variability? Ranz *et al.* (1997) studied the divergence of chromosomal element E between *D. melanogaster* and *D. repleta*. Although evolutionary rates may vary between elements and lineages, a rate of fixation of inversions of approximately

one inversion per million years was estimated for the E element (Ranz *et al.* 1997, 1999). Because most inversions will never reach fixation and will be lost after segregating for some time, we can consider a million years as an overestimate of the average lifespan of an inversion. This time is equivalent to 5×10^6 – 10^7 generations if we take 5–10 generations per year as an average for the *Drosophila* genus (Ashburner 1989; Powell 1997). An examination of Figure 1 yields the conclusion that at least 10^7 generations are needed to reach mutation-drift-flux equilibrium. It follows that it is very unlikely to achieve equilibrium within the inverted segment. If gene flux rates are $\leq 10^{-4}$, inversions can increase variability levels when mutation-drift-flux equilibrium is reached (Table 1). However, at least 10^7 generations are needed to achieve equilibrium, and most of the time variability is lower in low gene flux regions (Figures 1 and 2). This fact may imply that, other things being equal, chromosomes and/or species having high levels of inversion polymorphism will have lower levels of DNA polymorphism. It has been pointed out by Akashi (1996), Begun (1996), and Moriyama and Powell (1996) that the autosomes of *D. melanogaster*, which are moderately polymorphic for inversions, have lower variability than those of *D. simulans*, which have no known polymorphic inversions. Stronger hitchhiking and background selection in *D. melanogaster* have been proposed as possible causes of this striking correlation (Begun 1996). Our results show that the presence of inversions in the evolutionary history of *D. melanogaster* may help to explain its lower levels of nucleotide polymorphism without appealing to other selective forces. Although some caution must be raised because *D. melanogaster* levels of variability may have been underestimated (Begun 1996; Labate *et al.* 1999), the extant data fit the expected pattern. Of course, had the inversions been maintained for a longer time, or had the mutation rates been higher, inversions would produce the opposite effect and increase variability. On the other hand, factors, like fluctuating selection, that cause oscillations in the frequencies of the arrangements will boost the loss of neutral variability. Further research concerning all these possibilities is currently under way.

We thank P. Andolfatto, N. Barton, A. Berry, E. Betrán, B. Charlesworth, A. Clark, F. Depaulis, J. Rozas, and two anonymous reviewers for valuable discussion and criticism. Work was supported by a Formació del Personal Investigador (FPI) fellowship from the DGU (Generalitat de Catalunya, Spain) to A.N. and grant PB95-0607 from the DGICYT (Ministerio de Educación y Ciencia, Spain) to A.R.

LITERATURE CITED

- Aguadé, M., 1988 Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and noninverted chromosomes. *Genetics* **119**: 135–140.
- Akashi, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**: 1297–1307.
- Andolfatto, P., J. D. Wall and M. Kreitman, 1999 Unusual haplo-

- type structure at the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. *Genetics* **153**: 1297–1311.
- Aquadro, C. F., and D. J. Begun, 1993 Evidence for an implication of genetic hitchhiking in the *Drosophila* genome, pp. 159–178 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.
- Aquadro, C. F., S. F. Desse, M. M. Bland, C. H. Langley and C. C. Laurie-Ahlberg, 1986 Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**: 1165–1190.
- Aquadro, C. F., D. J. Begun and E. C. Kindahl, 1994 Selection, recombination and DNA polymorphism in *Drosophila*, pp. 46–56 in *Non-Neutral Evolution*, edited by B. Golding. Chapman & Hall, New York.
- Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Begun, D. J., 1996 Population genetics of silent and replacement variation in *Drosophila simulans* and *D. melanogaster*: X/autosomes differences? *Mol. Biol. Evol.* **13**: 1405–1407.
- Bénassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. *Genetics* **134**: 789–799.
- Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- Cáceres, M., J. M. Ranz, A. Barbadilla, M. Long and A. Ruiz, 1999 Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.
- Charlesworth, B., 1974 Inversion polymorphism in a two-locus genetic system. *Genet. Res.* **23**: 259–280.
- Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**: 213–227.
- Charlesworth, B., and D. Charlesworth, 1973 Selection of new inversions in multilocus genetic systems. *Genet. Res.* **21**: 167–183.
- Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Chovnick, A., 1973 Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. *Genetics* **75**: 123–131.
- Depaulis, F., L. Brazier and M. Veuille, 1999 Selective sweep at the *Drosophila melanogaster* *Suppressor of Hairless* locus and its association with the *In(2L)t* inversion polymorphism. *Genetics* **152**: 1017–1024.
- Dobzhansky, Th., 1970 *Genetics of the Evolutionary Process*. Columbia University Press, New York.
- Hasson, E., and W. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**: 1565–1575.
- Hey, J., 1991 A multi-dimensional coalescent process applied to multiallelic selection models and migration models. *Theor. Popul. Biol.* **39**: 30–48.
- Hudson, R. R., 1983 Properties of a neutral alleles model with intragenic recombination. *Theor. Popul. Biol.* **23**: 183–201.
- Hudson, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**: 1–44.
- Hudson, R. R., 1993 The how and why of generating gene genealogies, pp. 23–36 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.
- Hudson, R. R., 1994 How can the low levels of *Drosophila* sequence variation in regions of the genome with low levels of recombination be explained. *Proc. Natl. Acad. Sci. USA* **91**: 6815–6818.
- Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. *Genetics* **120**: 831–840.
- Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. *Genetics* **141**: 1605–1617.
- Kaplan, N. L., T. Darden and R. R. Hudson, 1988 The coalescent process in models with selection. *Genetics* **120**: 841–848.
- Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- Kimura, M., 1969 The rate of molecular evolution considered from the standpoint of population genetics. *Proc. Natl. Acad. Sci. USA* **63**: 1181–1188.
- Kreitman, M., 1983 Nucleotide polymorphism at the *alcohol dehydrogenase* locus of *Drosophila melanogaster*. *Nature* **304**: 412–417.
- Kreitman, M., and M. L. Wayne, 1994 Organization of genetic variation at the molecular level: lessons from *Drosophila*, pp. 157–183 in *Molecular Ecology and Evolution: Approaches and Implications*, edited by D. Schierwater, B. Streit, G. P. Wagner and R. de Salle. Birkhäuser, Basel.
- Krimbas, C. B., and J. R. Powell, 1992 *Drosophila Inversion Polymorphism*. CRC Press, Boca Raton, FL.
- Labate, J. A., C. H. Biermann and W. F. Eanes, 1999 Nucleotide variation at the *runt* locus in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **16**: 724–731.
- Lewontin, R. C., 1981 The scientific work of Th. Dobzhansky, pp. 93–115 *Dobzhansky's Genetics of Natural Populations*, edited by R. C. Lewontin, J. A. Moore, W. B. Provine and B. Wallace. Columbia University Press, New York.
- Lindley, D. L., and L. Sandler, 1977 The genetic analysis of meiosis in female *Drosophila melanogaster*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **277**: 295–312.
- Mayurama, T., 1971 An invariant property of a structured population. *Genet. Res.* **18**: 81–84.
- Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261–277.
- Nachman, M. W., 1997 Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* **147**: 1303–1316.
- Nachman, M. W., V. L. Bauer, S. L. Crowell and C. F. Aquadro, 1998 DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**: 1133–1141.
- Nagylaki, T., 1982 Geographical invariance in population genetics. *J. Theor. Biol.* **99**: 159–172.
- Nagylaki, T., 1992 *Introduction to Theoretical Population Genetics*. Springer-Verlag, Berlin.
- Navarro, A., and A. Ruiz, 1997 On the fertility effects of pericentric inversions. *Genetics* **147**: 931–933.
- Navarro, A., E. Betrán, C. Zapata and A. Ruiz, 1996 Dynamics of gametic disequilibria between loci linked to chromosome inversions: the recombination redistributing effect of inversions. *Genet. Res.* **67**: 67–76.
- Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by crossing over and gene conversion in inversion heterokaryotypes. *Genetics* **146**: 695–709.
- Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei, M., K.-I. Kojima and H. E. Schaeffer, 1967 Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**: 741–750.
- Nordborg, M., 1997 Structured coalescent processes on different time scales. *Genetics* **146**: 1501–1514.
- Notohara, M., 1990 The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.* **29**: 59–75.
- Ohta, T., and K.-I. Kojima, 1968 Survival probabilities of new inversions in large populations. *Biometrics* **24**: 501–516.
- Payne, F., 1924 Crossover modifiers in the third chromosome of *Drosophila melanogaster*. *Genetics* **9**: 327–342.
- Popadić, A., and W. Anderson, 1995 Evidence of gene conversion in the Amylase multigene family of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**: 564–572.
- Popadić, A., D. Popadić and W. Anderson, 1995 Interchromosomal exchange of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**: 938–943.
- Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, Oxford.
- Prakash, S., and R. C. Lewontin, 1968 A molecular approach to the study of genetic heterozygosity in natural populations. III. Direct evidence of coadaptation in gene arrangements of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **59**: 398–405.
- Ranz, J. M., C. Segarra and A. Ruiz, 1997 Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* species group. *Genetics* **145**: 281–295.
- Ranz, J. M., M. Cáceres and A. Ruiz, 1999 Comparative mapping of cosmids and gene clones from a 1.6 Mb chromosomal region of *Drosophila melanogaster* in three species of the distantly related subgenus *Drosophila*. *Chromosoma* **108**: 32–43.
- Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the rp49 region of *Drosophila subobscura* between different chro-

- mosomal gene arrangements. *Proc. Natl. Acad. Sci. USA* **90**: 8083–8087.
- Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **91**: 11517–11521.
- Rozas, J., C. Segarra, G. Ribó and M. Aguadé, 1999 Molecular population genetics of the rp49 gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**: 189–202.
- Slatkin, M., 1987 The average number of sites separating DNA sequences drawn from a subdivided population. *Theor. Popul. Biol.* **32**: 42–49.
- Sorsa, V., 1988 *Polytene Chromosomes in Genetic Research*. John Wiley & Sons, New York.
- Sperlich, D., and P. Pfiem, 1986 Chromosomal polymorphism in natural and experimental populations, pp. 257–309 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner, H. L. Carson and J. N. Thompson Jr., Vol. 3e, Academic Press, London.
- Stephan, W., and C. H. Langley, 1998 DNA polymorphism in *Ly-copersicon* and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor. Popul. Biol.* **41**: 237–254.
- Strobeck, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**: 545–555.
- Strobeck, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117**: 149–153.
- Sturtevant, A. H., and G. W. Beadle, 1936 The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* **21**: 544–604.
- Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- Tajima, F., 1989a DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* **123**: 229–240.
- Tajima, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Tajima, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Sunderland, MA.
- True, J. R., J. M. Mercer and C. C. Laurie, 1996 Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**: 507–523.
- Wakeley, J., 1996 The variance of pairwise nucleotide differences in two populations with migration. *Theor. Popul. Biol.* **49**: 39–57.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- Wesley, C. S., and W. F. Eanes, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 3132–3136.
- Zapata, C., and G. Álvarez, 1987 Gametic disequilibrium in populations of *Drosophila subobscura*: a review of experimental evidence. *Genét. Ibér.* **39**: 593–616.
- Zapata, C., and G. Álvarez, 1992 The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. *Evolution* **46**: 1900–1917.
- Zapata, C., and G. Álvarez, 1993 On the detection of nonrandom associations between DNA polymorphism in natural populations. *Mol. Biol. Evol.* **10**: 823–841.

Communicating editor: A. G. Clark

APPENDIX

Variances for $k(n)$, the number of pairwise differences in a sample of n alleles, can be easily found for $n = 2$ by developing an expression for pairwise identities and using it as the moment generating function of the distribution of coalescence times (Hudson 1990). Doing so, we obtain the rather bulky expressions

$$Vk(2, 0) = \frac{\theta}{1+F} (F + p(1 + 2q)) + \frac{\theta^2}{(1+F)^2} \left(\frac{8p^2q^2}{F} + F^2 + 2F(1 + pq) + p(4 + 3p(3 - 4p(1 + q))) \right) \quad (A1a)$$

$$Vk(1, 1) = \frac{\theta}{1+F} \left(\frac{F + 2pq}{F} + F + 2pq \right) + \frac{\theta^2}{1+F} \left(\frac{4p^2q^2(1 + 3F)}{F^2} + 1 + F + 2pq \right) \quad (A1b)$$

$$Vk(0, 2) = \frac{\theta}{1+F} (F + q(1 + 2p)) + \frac{\theta^2}{(1+F)^2} \left(\frac{8p^2q^2}{F} + F^2 + 2F(1 + pq) + q(4 + 3q(3 - 4q(1 + p))) \right). \quad (A1c)$$

These equations can be simplified if we assume equal arrangement frequencies ($p = q$):

$$Vk(2, 0) = Vk(0, 2) = \theta \left(1 + \theta + \frac{\theta}{2F} \right) \quad (A2a)$$

$$Vk(1, 1) = \theta \left(1 + \theta + \frac{\theta}{2F} \right) + \frac{\theta}{2F} \left(1 + \frac{\theta}{2F} \right). \quad (A2b)$$

If n increases, the mean number of pairwise differences remains the same but, of course, the variance decreases. However, as variances decrease the expressions giving them increase in size and in number (for example, if $n = 10$ one has to obtain 11 different enormous expressions).

Variances for the simplest case ($p = q$) can be obtained with some pain following Wakeley (1996). We end up with two expressions equivalent to those in Wakeley (1996). The first one gives the variance of $k(n)$ when all the n alleles are linked to the same arrangement:

$$Vk(n, 0) = Vk(0, n) = \frac{\theta}{6n(n-1)(3 + 6F + 2F^2)} \times (6F + n(6 + 7F + 4F^2) + n^2(6 + 13F + 4F^2) + \frac{\theta}{6F(1+F)(1+2F)} \times (6F(1+F)(15 + 54F + 64F^2 + 13F^3) + n(18 + 123F + 327F^2 + 358F^3 + 160F^4 + 32F^5) + n^2(3 + 4F)(6 + 27F + 43F^2 + 34F^3 + 8F^4))). \quad (A3)$$

The second one gives the variance of $k(n)$ when i alleles are linked to St and j alleles to In :

$$\begin{aligned}
 Vk(i, j) = & \frac{\theta}{6jF(3 + 6F + 2F^2)} (F^2(9 + 4F) + (i + j)F(3 + F)(3 + 2F) \\
 & + ij(9 + 18F + 15F^2 + 4F^3) \\
 & + \frac{\theta}{6F(1 + F)^2(1 + 2F)} \\
 & \times (F^2(36 + 222F + 573F^2 + 706F^3
 \end{aligned}$$

$$\begin{aligned}
 & + 400F^4 + 80F^5) \\
 & + (i + j)F(27 + 153F + 390F^2 \\
 & \quad + 525F^3 + 374F^4 \\
 & \quad + 128F^5 + 16F^6) \\
 & + ij(27 + 162F + 459F^2 + 780F^3 \\
 & \quad + 843F^4 + 562F^5 \\
 & \quad + 208F^6 + 32F^7)). \quad (A4)
 \end{aligned}$$