Mdg-1 mobile element heterozygosity in *Drosophila melanogaster*

C. Biémont and
C. Gautier

Laboratoire de Biologie des Populations, Université Lyon 1, 69622 Villeurbanne, France.

Using *in situ* hybridisation to the giant chromosomes of salivary glands, mdg-1 mobile element polymorphism was studied in individuals of *Drosophila melanogaster* from a mass-mated population and from inbred lines established from this population (analysed at the 15th, 27th and 35th generations). The number of mdg-1 copies and the number of heterozygous insertion sites were estimated in the expected hybrids, which resulted from theoretically crossing the inbred lines. The mean number of labelling sites in hybrids that originated from lines from the 35th generation was equal to 30·5 and was similar to the value observed in the initial population. An average proportion of heterozygous sites of 0·85 (with a range from 0·62 to 0·96), and an average number of copies of 35 (29 to 44) were found in these hybrids. These values may be considered as valid estimates of the values in the initial population. Other measurements of genetic heterogeneity, determined on the mass-mated population and the inbred lines, are proposed and discussed.

INTRODUCTION

It is now well established that many species contain families of mobile DNA sequences that play an important role in the construction of an individual's genome. However, experimental data on the distribution of such mobile elements in the chromosomes of various populations is just beginning to appear (for *Drosophila melanogaster* see Gvozdev, 1981; Montgomery and Langley, 1983; Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Biémont, 1986; Biémont and Aouar, 1986). The main result is that a constant average copy number of elements is set up in mass-mated *Drosophila* populations, while the position of the elements on the chromosomes shows great variability.

According to some theoretical models of the population dynamics of transposable elements (Brookfield, 1982; Charlesworth, 1983; Charlesworth and Charlesworth, 1983; Langley, Brookfield and Kaplan, 1983, Charlesworth and Langley, 1986), and recent experimental observations (Biémont and Aouar, 1986, in *Drosophila*; Alleman and Freeling, 1986, in maize), such an equilibrium may be maintained either by the regulation of transposition and excision in response to copy number by genome, or by a selective process on individuals as a complex function of the copy number of elements in the genome. However, we have only a few ideas on the level of heterozygosity of such mobile elements in populations. Indeed, DNA blotting and *in situ* hybridisation techniques do not allow heterozygous insertion points of mobile elements to be detected, except in the latter technique when the chromatids are separated.

Here, we checked for heterozygosity level by an indirect way. We analysed highly inbred lines of *Drosophila melanogaster* for their insertion pattern of the mobile element mdg-1 (a copia-like element: Ilyin *et al.*, 1980; Tchurikov *et al.*, 1981) by means of *in situ* hybridisation with a biotinylated-DNA probe (Rigby *et al.*, 1977; Langer-Safer *et al.*, 1982) on the giant chromosomes of salivary glands. By considering the lines as a random sample of gametes from individuals of the initial population, we reconstructed this initial population as the ensemble of the expected hybrids obtained from theoretical crosses of the inbred lines. Different estimates of genetic diversity were calculated and discussed.

MATERIALS AND METHODS

A population of *Drosophila melanogaster* was established from 50 flies captured in Lerik

(Azerbaïdjan, USSR). The population so formed was maintained in the laboratory by mass culture (around 500 flies per generation), at 25°C, for 12 months before the experiment was started. The insertion sites of the mdg-1 elements were determined on the whole genome in two samples of 11 and 13 larvae (analysed at a 1-year interval) from the population, and on 17 unselected inbred lines established from this population by sib mating every generation. The mdg-1 insertion pattern of the lines were determined at the 15th, 27th and 35th generations.

Salivary glands of third instar larvae reared at 18°C were squashed in 45 per cent acetic acid, dehydrated in 96 per cent ethanol and air dried. The slides were incubated in $2 \times$ SSC solution for 30 minutes at 70°C, dehydrated through ethanol and air dried. They were then treated with 0·07 NaOH for 2 minutes, dehydrated and then air dried. A solution of biotinylated-denatured DNA probe (mdg-1 included in pBR322 that contained $4 \times$ SSC, dextran and 50 per cent formamide was added to the preparations. Annealing was carried out overnight in a humid chamber at 37°C. The slides were then washed in successive changes of $2 \times$ SSC and PBS. The hybridisation sites were made visible by reacting the chromosomes to a biotinylated-streptavidin complex labeled with horseradish peroxidase and finally with diaminobenzidine and H_2O_2 as stains for peroxidase. The slides were then Giemsa stained. The hydbridisation sites were analysed using coloured pictures obtained from a phase contrast microscope. Four to six nuclei were analysed per larva. It was observed that no polymorphism of insertion exists between different nuclei of the same salivary gland. For a precise location of the insertion sites, the pictures of chromosomes from all the larvae were compared two by two. In case of doubt about the precise location of the site on the chromosomal bands, hybrids were done between lines; the location of the insertion sites was determined again and compared to that of the inbred lines. For a detailed analysis of the patterns of mdg-1 insertion sites in individuals from the mass-mated population and the inbred lines, see Biémont (1986) and Biémont and Aouar (1986).

RESULTS

Mdg-1 polymorphism

The number of mdg-1 element insertions (the labelling sites) on the genome ranged from 11 to 29 for the 17 inbred lines and from 24 to 40 in the two samples of individuals from the mass-mated population. The low copy number of insertions in the inbred lines reflects the segregation in the lines of the sites that were heterozygous in the individuals of the initial population.

From the 17 inbred lines, we observed 108, 114 and 106 different insertion sites for the generations 15, 27 and 35, respectively. This is low compared to 175 and 149 sites observed on the 11 and 13 larvae from the two samples of the mass-mated population (Biémont, 1986); this may reflect the fact that 17 inbred lines is not enough to give a complete picture of the parental insertion pattern. The smaller number of insertion sites in the second sample of larvae from the mass-mated population may be caused by drift in this laboratory population.

Since homozygous inbred lines may be considered as a random sample of gametes from individuals of the initial population, we analysed the expected mdg-1 insertion patterns of the $n(n-1)/2$ hybrids obtained by theoretically crossing the n inbred lines. Note that the probability to have n gametes i in a sample of N gametes is:

$$C_N^n \, p_i^n (1-p_i)^{N-n}$$

and the probability of an homozygous pair (i, i) among the n gametes of the sample is $n(n-1)/N(N-1)$. Hence, the probability to obtain an homozygous pair is:

$$p(i, i) = \sum_{n=2}^{N} C_N^n \, p_i^n (1-p_i)^{N-n} \times n(n-1)/N(N-1)$$

$$= \sum_{n=2}^{N} \frac{N!}{n!(N-n)!} \frac{n(n-1)}{N(N-1)} p_i^n (1-p_i)^{N-n}$$

$$= p_i^2 \sum_{n=2}^{N} \frac{(N-2)!}{(n-2)!(N-n)!} p_i^{n-2} (1-p_i)^{N-n}$$

$$= p_i^2 (p_i + (1-p_i))^{N-2} = p_i^2.$$

So, considering the hybrids without the crosses between the lines is equivalent as taking 2 gametes in a binomial way from the ensemble of gametes.

We then calculated the number of labelling sites, the number of copies in the genome, and the number of heterozygous insertion sites. We considered that in first approximation these hybrids may represent individuals from a mass-mated population. Overall, 34–38 copies of the element in the diploid genome corresponded to an average number of labelling sites around 30–33; among the sites, 26–28 were heterozygous (see table 1). This leads to a proportion of heterozygous sites per

Table 1 Measurements of genetic variability due to mdg-1 insertions

Samples	Genetic variability indexes				
	A	H	P_H	$P_{H,400}$	NH
Generation 15	0·14	5·3	0·26	0·07	27·9
Generation 27	0·13	4·8	0·24	0·07	27·7
Generation 35	0·12	4·1	0·25	0·07	26·3
Lerik 1	0·09	2·9	0·16	0·07	28·1
Lerik 2	0·10	3·1	0·17	0·06	25·5

A: allelism as defined by Ohta (1985, 1986).
H: Langley's index (Langley et al., 1983).
P_H: probability of heterozygous sites among all the sites observed.
$P_{H,400}$: the same as P_H but relative to 400 theoretical possible insertion sites.
NH: heterozygous site number.

hybrid (No of heterozygous sites/No of labelled sites) equal to 0·85 on average with a range from 0·62 to 0·96 (all generations compounded). A decrease in variance of the number of labelling sites in the hybrids, from the 15th to the 35th generations, was caused by a loss of insertion sites in the inbred lines that led to a decrease in variance of this insertion number in these lines (fig. 1). Note also, that the average per individual number of labelling sites in the mass-mated population was 31·6 and 29·6 for the 11 and the 13 larvae samples, repectively (it was equal to 30·5 for the pooled data); this number was closer to the value calculated from data on hybrids from the 35th generation (30·5) than to the values obtained from earlier generations (33·21 and 32·46 for the 15th and 27th generations, respectively).

Note that a functional relationship exists between the number of labelled sites, NS, the number of heterozygous sites, NH, and the number of copies, NC, so that:

$$NC = 2NS - NH. \tag{1}$$

So,

$$2NS = NC + NH$$

with

$$NC = a_1 NS$$
$$NH = a_2 NS.$$

Hence,

$$NC + NH = (a_1 + a_2) NS$$
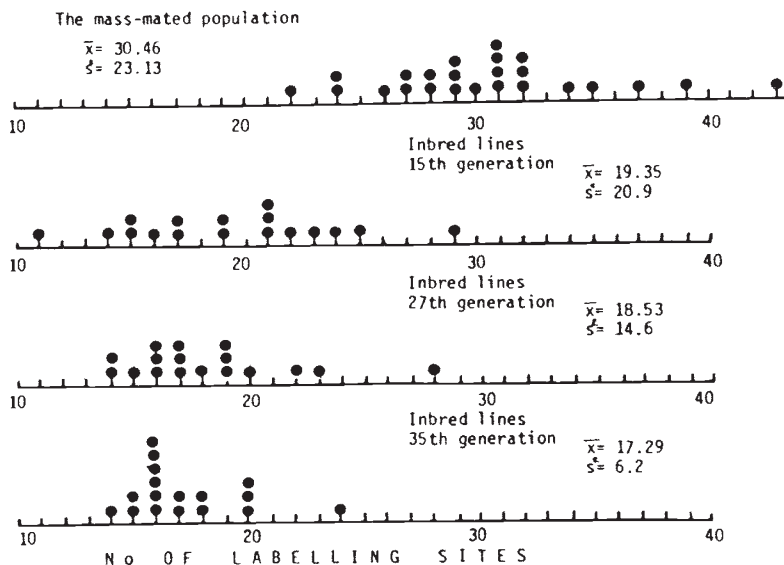
and

$$a_1 + a_2 = 2.$$

Figure 1. Distribution of the number of labelling sites in individuals from the mass-mated population (the data from the two samples are pooled), and from the inbred lines for the 15th, 27th and 35th generations.

Of course, the question remains whether the estimates of copy number and proportion of heterozygous sites are good approximations of these numbers in the initial mass-mated population. The problem with the *in situ* hybridisation technique is that there is no way to check for such parameters in individuals from natural populations, except when the chromatids are separated; but this rarely happens. So we need other estimates of genetic variability that could be used for the hybrid and the mass-mated population data to be compared.

Indexes of genetic variability.

Different measurements of genetic variability were calculated:

.P_H = the probability that a random site will be heterozygous in a random zygote. This index is similar to classical heterozygosity measurements from electrophoresis data. However, it depends on the number of insertion sites that is considered. In this calculation we considered each site of insertion as a classical Mendelian locus as done in some theoretical approaches of the dynamics of transposable elements (Charlesworth, 1983). The formula used was:

$$P_H = (1/r)(2 \sum f_i(1-f_i))$$

where r is the "total number of sites". If r is equal to the number of detected sites in the population and if the population size is large, then the detection of a new rare site leads to a decrease of P_H. This is not the variation expected for a heterozygosity measure. One solution, however, consists of taking r as the maximum number of sites that can be detected by the *in situ* hybridisation technique (here 400). This, however, supposes that all these sites may bear mdg-1 insertions in a given population; but, as discussed below, this is not certain (see also Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Biémont, 1986). We present here three numbers: P_H where r is taken as the number of sites detected in the population, $P_{H,400}$ where r equals 400, and NH, which is the estimated number of heterozygous sites calculated previously ($NH = 2 \sum f_i(1-f_i)$).

.$H = \sum f_i^2$ = The mean number of sites at which the two genomes of a zygote are occupied (Langley *et al.*, 1983). This statistic has the important property of being independent of the number of different sites detected in the population; detection of a new rare site scarcely modifies H. However, this parameter is not really suitable for measuring genetic variability if considered alone, as shown by the

following simple example: in population 1 there exists only one site that is present in all genomes; in population 2 there exist 100 sites, each of them having a probability of 1/10. Clearly, population 2 has a higher genetic variability than population 1; but in both cases the value of H is 1. Note, however, that NH and H taken altogether give a complete information on the genetic variability. Indeed, in the above example, the population 1 has $NH = 0$ and the population 2 has $NH = 18$.

.A = The probability that a random copy of a genome finds a homologous copy at the same chromosomal site of another genome of the population (Ohta, 1985, 1986). This parameter A is a good index of genetic variability. Indeed, in the preceeding example population 1 gives $A = 1$ and population 2 gives $A = 0.1$. The main drawback of A is the difficulty of its estimation. We simply used $(\sum f_i^2)/(\sum f_i) = H/(\sum f_i)$.

Note also that we can write

$$A = (NC - NH)/NC = 1 - (NH/NC).$$

But, from the equations (1), it appears that $NH/NC = a_2/a_1$; so, $A = 1 - (a_2/a_1)$.

These different indexes were computed in the two different experimental situations: the two samples from the mass-mated population, and the individuals from the 17 inbred lines in generation 15, 27 and 35. Since the experimental method gave no direct access to the value of the frequency f_i of sites in an haploid genome, we admitted here that the frequency of zygotes with no copy in two homologous sites i was $a_i = (1-f_i)^2$ and we then estimated f_i by $1 - \sqrt{a_i}$. To compare the mass-mated population with the inbred lines, we considered each line as a gamete of the population. We then estimated f_i as the frequency of lines with site i. Since lines were considered as random gametes, we used hypergeometric models so that:

$$H = \sum (m_i(m_i-1))/(n(n-1))$$
$$A = H/\sum f_i$$
$$NH = 2 \sum m_i(n-m_i)/(n(n-1))$$
$$P_H = NH/r$$

where m_i is the number of inbred lines with an insertion in the site i; n is the total number of lines; r is the total number of labelled bands in the n inbred lines.

The main feature of table 1 is the constancy of NH opposed to the lower values of A, H and P_H in the mass-mated population samples. This may be due to a greater mobility of mdg-1 in inbred

lines. The appearance of new sites, in inbred lines, can be counterbalanced, by a random loss process leading to a decrease in the number of sites simultaneously present in two lines. Note also that the values of P_H (from 0·16 to 0·24) were low because of the high homozygosity for element absence, which results from insertions seen only in one or two lines. Hence, the new mdg-1 insertions or excisions that appeared in some inbred lines (Biémont and Aouar, 1986) should have affected the values of this parameter. Indeed, an insertion in a site not already present in a line defines a new site almost completely homozygous for the absence of the element; hence, the heterozygosity may go down. However, from the 15th to the 35th generations, no significant change was observed for either P_H or the number of heterozygous sites in hybrid individuals, NH.

We have already noted that P_H is sensitive to the total number of insertion sites; it also depends on the number of individuals or lines analysed, so that the minimum value relative to rare insertion sites is attained for a site carried by only one individual or line. In this case, the minimum value is $2[((N-1)/N)^{1/2} - (N-1)/N]$, when calculated for N individuals, and $2/N$ when determined from N inbred lines. Moreover, this parameter is determined only from sites that were really observed. However, if the total number of possible sites of insertion is limited on a population, and if all these sites are observed in an experiment, then, this parameter may really express the degree of heterozygosity for the mobile element that the population is able to maintain.

Since the various estimates of genetic variability were similar whatever the set of data used (from the mass-mated population or the inbred lines), we can assume that an average proportion of heterozygous insertion sites per genome of around 0·85 is a good approximation of this value in the mass-mated population.

DISCUSSION

The purpose of this paper is to analyse genetic variability caused by localization of mdg-1 element insertions. We must first recognise that results depend on the experimental method. The method used, *in situ* hybridisation, is able to define about 400 different hybridisation bands in the polytene chromomosomes. It is clear that each band does not correspond to only one insertion site. In the case of retroviruses, it is known that several insertion sites, inducing different types of tumors, can

be less distant than 1·5 kb (Silver and Kozak, 1986; Shen-Ong *et al.*, 1986), and the *in situ* hybridisation only allows mdg-1 to be localised to within a 200 kb subdivision in the Bridges map. This does not imply, however, that any nucleotide of the genome is a suitable insertion site (see Gvozdev, 1981; Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Biémont, 1986; Pasyukova *et al.*, 1986). Note that the appearance of new sites in inbred lines is far from random; indeed, among 84 transpositions detected in the generations 27 and 35, only 28 were on new bands and 56 in bands that have been already detected (unpublished results); this concurs with a limited number of possible insertion sites and does not corroborate the results of Montgomery and Langley (1983), which showed very great variation between individuals in transposable element position on the X chromosomes; it was shown that their data were consistent with all sites being uniquely occupied (Kaplan and Brookfield, 1983). Previous data (Belyaeva *et al.*, 1984; Biémont *et al.*, 1985; Biémont, 1986) are more in agreement with a finite number of possible sites of insertion in such a way that different individuals may share some sites in common, and not only because of homology by descent. The fact that many new insertions over generations were in sites already seen in some lines, explains the observed constancy of P_H, NH and A over generations, and is compatible with the observation that mdg elements often insert into bands where they had previously been located (Gerasimova *et al.*, 1985; Pasyukova *et al.*, 1986); of course, one also could incriminate cross-contamination between lines; but, since the lines were maintained by only one sib pair every generation, a contamination by alien flies should have resulted in a drastic change in the mdg pattern on all the chromosome arms. Only minor changes were observed, and not always simultaneously on all the chromosome arms (Biémont and Aouar, 1986). Moreover, how can contamination explain the mdg-1 excisions observed in some lines?

Note, however, that impact on genetic variability measurement seems to be similar when using electrophoresis analysis for which the precision in allele separation greatly depends on the method used. If the nucleotide level is considered, then an infinity of alleles is possible (there exist 3×10^{73} possible messengers that could be translated into a perfect human beta globin). Hence, the genetic variability measurements computed here must not be considered as absolute numbers but as a mean of comparison between different populations or experimental situations.

One problem that remains is how to explain the maintenance of the high heterozygosity level in individuals (the proportion of heterozygous insertion sites in a genome is 0·85). Of course, transpositions may be involved in this variation but more information on various mobile elements and on different populations is needed before a complete understanding of this phenomenon can be elaborated. This recalls the classical observation of a substantial level of enzymatic heterozygosity found in many species. No consensus has been attained as to how this enzymatic heterozygosity is maintained in natural populations; and the question whether most evolutionary change is governed by natural selection or by random fluctuations of adaptatively equivalent variants is still widely debated (see Lewontin, 1974).

Mobile elements such as mdg-1, by their high number of insertions in the genome, participate in the maintenance of the pool of mutations and, hence, the genetic load (Biémont *et al.*, 1985; Mukaï *et al.*, 1983). The way such elements are regulated and controlled should then help explain the maintenance of polymorphism in experimental and natural populations.

REFERENCES

ALLEMAN, M. AND FREELING, M. 1986. The Mu transposable elements in maize: evidence for transposition and copy number regulation during development. *Genetics, 112,* 107–119.

BELYAEVA, E. SP., ANANIEV, E. V. AND GVOZDEV, V. A. 1984. Distribution of mobile dispersed genes (mdg-1 and mdg-3) in the chromosomes of *Drosophila melanogaster. Chromosoma, 90,* 16–19.

BIÉMONT, C. 1986. Polymorphism of the Mdg-1 and I mobile elements in *Drosophila melanogaster. Chromosoma, 93,* 393–397.

BIÉMONT, C. AND AOUAR, A. 1986. Copy-number dependent transpositions and excisions of the mdg-1 mobile element in inbred lines of *Drosophila melanogaster. Heredity* (In press).

BIÉMONT, C., BELYAEVA, E. SP., PASYUKOVA, E. G. AND KOGAN, G. 1985. Mobile gene localization and viability in *Drosophila melanogaster. Experientia, 41,* 1474–1476.

BROOKFIELD, J. F. Y. 1982. Interspersed repetitive DNA sequences are unlikely to be parasitic. *J. Theor. Biol., 94,* 281–299.

CHARLESWORTH, D. 1983. Models of the evolution of some genetic systems. *Proc. R. Soc. London, 219,* 265–279.

CHARLESWORTH, B. AND CHARLESWORTH, D. 1983. The population dynamics of transposable elements. *Genet. Res., 42,* 1–27.

CHARLESWORTH, B. AND LANGLEY, C. H. 1986. The evolution of self-regulated transposition of transposable elements. *Genetics, 112,* 359–383.

GERASIMOVA, T. I., MATJUNINA, L. V., MIZROKHI, L. J. AND GEORGIEV, G. P. 1985. Successive transposition explosions in *Drosophila melanogaster* and reverse transpositions of mobile dispersed genetic elements. *EMBO J., 4,* 3773–3779.

GVOZDEV, V. A. 1981. The nature and functions of intercalary heterochromatin in *Drosophila melanogaster.* In *Molecular bases of genetic processes.* Proc. XIV Int. Congr., Genet. 3, pp. 257–271.

ILYIN, Y. V., CHMELIAUSKAITE, V. G. AND GEORGIEV, G. P. 1980. Double-stranded sequences in RNA of *Drosophila melanogaster:* relation to mobile dispersed genes. *Nucleic Acid Res., 8,* 3439–3457.

KAPLAN, N. AND BROOKFIELD, J. F. Y. 1983. Transposable elements in Mendelian population. III. Statistical results. *Genetics, 104,* 485–495.

LANGER-SAFER, P. R., LEVINE, M. AND WARD, D. C. 1982. Immunological method for mapping genes on *Drosophila* polytene chromosomes. *Proc. Natl. Acad. Sci., 79,* 4381–4385.

LANGLEY, C. H., BROOKFIELD, J. F. Y. AND KAPLAN, N. 1983. Transposable elements in Mendelian populations. I. A theory. *Genetics, 104,* 457–471.

LEWONTIN, R. C. 1974. *The Genetic Basis of Evolutionary Changes.* Columbia University Press, New York.

MONTGOMERY, E. A. AND LANGLEY, C. H. 1983. Transposable elements in Mendelian populations. II. Distribution of copia-like elements in natural populations. *Genetics, 104,* 473–483.

MUKAÏ, T. AND YUKUHIRO, K. 1983. An extremely high rate of deleterious viability mutations in *Drosophila* possibly caused by transposons in non-coding regions. *Proc. Jap. Acad. Sci., 59,* 316–319.

OHTA, T. 1985. A model of duplicative transposition and gene conversion for repetitive DNA families. *Genetics, 110,* 513–524.

OHTA, T. 1986. Population genetics of an expanding family of mobile genetic elements. *Genetics, 113,* 145–158.

PASYUKOVA, E. G., BELYAEVA, E. SP., KOGAN, G. L., KAÏDANOV, L. Z. AND GVOZDEV, V. A. 1986. Concerted transpositions of mobile genetic elements coupled with fitness changes in *Drosophila melanogaster. Mol. Biol. Evol., 3,* 299–312.

RIGBY, P. W., DIECKMANN, M., RHODES, C. AND BERG, P. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol., 113,* 237–251.

SHEN-ONG, G. L. C., MORSE, H. C., POTTER, M. AND MUSKINSKI, J. F. 1986. Two modes of c-myb activation in virus-induced mouse myeloid tumors. *Mol. Cell. Biol., 6,* 380–392.

SILVER, J. AND KOZAK, C. 1986. Common proviral integration region on mouse chromosome 7 in lymphomas and myelogenous leukemias induced by Friend Murine Leukemia virus. *J. Virology, 57,* 526–533.

TCHURIKOV, N. A., ILYIN, Y. V., SKRYABIN, K. G., ANANIEV, E. V., BAYEV, A. A., KRAYEV, A. S., ZELENTSOVA, E. S., KULGUSKIN, V. V., LYUBOMIRSKAYA, N. V. AND GEORGIEV, G. P. 1981. General properties of mobile dispersed genetic elements in *Drosophila. Cold Spring Harbor Sympos. quant. Biol., 45,* 655–665.