Orloci, L. 1968. A model for the analysis of structure in taxonomic collections. Can. J. Bot. 46:1093-1097.

Orloci, L. 1971. An information theory model for pattern analysis. J. Ecol. (in press).

Seal, H. 1964. Multivariate statistical analysis for biologists. Wiley. New York.

Sokal, R. R. and Rohlf, F. J. 1969. Biometry. W. H. Freeman and Co., San Francisco.

Sokal, R. R. and Sneath, P. H. A. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco.

Wirth, M., Estabrook, George F. and Rogers, David J. 1966. A graph theory model for systematic biology, with an example from the Oncidiinae (Orchidaceae), Systematic Zoology 15:59-69.

R. C. Jancey

3. Reanalysis of Longley's data: translocations in maize.

A. E. Longley (1958, 1961) has provided the most extensive cytological data on translocations available from any organism (Lindsley and Grell (1968) list approximately 800 translocations in Drosophila). Whereas the Drosophila data are based entirely on salivary analysis, Longley's data are determined entirely from pachytene analysis. Longley's data provide assignment of a 'breakage point' to a specific unit within a chromosome arm. He has discussed the limits of his data, e.g. sampling error, and some corrections have been listed in previous Newsletters (e.g. Burnham, 1969).

We have become interested in the distribution of genetic events and undertook to reanalyze Longley's data. Initially, we have attempted to describe the distribution, and in particular to establish any areas within chromosomes in which there can be detected a higher frequency of breakpoints than might be expected on a hypothesis of 'random distribution'. Our present report, which should be considered preliminary, consists largely of a summary of these distribution analyses. Next, however, we are reexamining the distribution in terms of probability statements for particular types of rearrangements. While not yet finished, it is clear that centromere-translocation probabilities far exceed that expected by chance alone. Lastly, designed but not yet processed, is a production of likelihood statements for different length-interstial and interchange-

segment aberrations. In turn, we propose to extent these analyses to the Drosophila data.

A retrospective study of the type reported below must necessarily be viewed with some reservations. While most of Longley's data are derived from induced translocations, it may be argued that the different treatments, plus differential survival, are responsible for any deviation from randomness rather than assigning such a property to the genetic material itself. Nevertheless, particularly in relation to Drosophila, we feel that Longley's data are in fact the least susceptible to the criticisms above of any data of this type which will become available in the near future.

Histograms per .05 arm length have been plotted and are summarized per arm in Table 1. In some later analyses length units have been converted to 'real lengths' by formulae derived from the data of Rhoades (1950) and Neuffer, Jones and Zuber (1968). These estimates are of mid-pachytene lengths. From our own data we assigned an estimate of 2μ as the average length of centromeres.

We set out to test two basic hypotheses concerning the data:

1. That the chromosome breaks did not depart from a random distribution.
2. That the association of chromosome segments did not depart from a random distribution.

Apart from testing these hypotheses for the chromosome complement as a whole, it was obviously desirable to see if departures from the hypothesis were characteristic of sub-sets of the whole.

A number of statistical tests of these hypotheses would have been appropriate, the most obvious being the X^2 test and the G test (Sokal and Rohlf, 1969). Because of general familiarity with the test, it was decided to employ X^2, though certain computational and theoretical advantages are attached to the use of the G test. The possibility of applying other analytical tools to the data has been mentioned above.

Clearly, in order to test the hypotheses, it was necessary to create a number of classes within the chromosome complement, to which break or association frequencies could be assigned. It would have been most desirable to divide the total chromosome lengths into small, say

2μ, segments; however, given the amount of data available, this would have resulted in very low expected frequencies per cell. Such a situation would greatly reduce the validity of the X^2 test, or indeed of any test serving the same purpose. The larger the class size, however, the less information would be yielded concerning localised deviations from randomness. It was concluded that taking chromosome arms as classes would provide the most useful compromise between class size and expected number of observations per class.

There was no logical basis for assigning centromere breaks to one arm in preference to the other. Consequently, the two hypotheses were tested first on arm data alone, then the first hypothesis was tested for centromere data, and finally the second hypothesis was tested between the two classes of data.

The computation of expected values for each cell requires some explanation:

Expected Cell Frequencies for 1st Hypothesis.

Since the hypothesis postulates a random distribution of chromosome breaks, the expected frequency, F, per chromosome arm is given by $F = [n/L]\cdot\ell$, where L = Total Complement Length, ℓ = Segment Length of Arm, and n = Total Number of observed breaks for chromosome complement.

Expected Cell Frequencies for 2nd Hypothesis.

If the number of translocations involving segment 1 is F_1, and the translocations involve random reassociation of breaks, then the number of translocation pairs involving segment 1 and segment 2 will be:

$$[F_1/L\text{-}L']\ell_2$$

where L = Total complement length

L' = Chromosome length of which segment 1 is a part (because no inversion data was included)

ℓ_2 = Chromosome length of which segment 2 is a part

Similarly, if the number of translocations involving segment 2 is F_2, then the number of translocation pairs involving segments 2 and 1 will be:

$$[F_2/L\text{-}L'']\ell_1$$

where L'' = Chromosome length of which segment 2 is a part

ℓ_1 = length of segment 1

Thus the total number of different translocation pairs involving

both segments is:

$$[(F_1/L\text{-}L')\ell_2] + [(F_2/L\text{-}L'')\ell_1]$$

Results

Table 1 shows the distribution of chromosome breaks by arm. It will be noticed that the most striking deviations from expected values occur in the arms 2S, 5L and 9L. Over all, the data depart from the hypothesis of a random distribution of breaks such that the probability of occurrence within the hypothesis is considerably less than 0.005.

Table 1

Distribution of breakage frequencies by arm

Chr. segment	Obs.	Expected*	$\frac{(f - F)^2}{F}$
1S	101	125.3	4.7
1L	141	163.1	3.0
2S	75	103.2	7.7
2L	135	129.5	0.2
3S	71	72.4	0.0
3L	127	144.5	2.1
4S	87	79.1	0.8
4L	115	126.7	1.1
5S	109	99.7	0.9
5L	145	109.9	11.2
6S	36	41.6	0.8
6L	148	129.1	2.8
7S	35	44.1	1.9
7L	111	119.7	0.6
8S	42	40.2	0.1
8L	151	126.0	5.0
9S	50	54.2	0.3
9L	133	97.3	13.1
10S	36	33.9	0.1
10L	88	95.20	0.5
	1936	1934.7	56.9
D.F. = 19		$X^2 = 56.9$ $p < 0.005$	

*See text. Length data from Rhoades (1950) and Neuffer, Jones and Zuber (1968).

The hypothesis of random distribution of breaks stands a little firmer in the case of breaks occurring within centromeres (see Table 2).

Departures from expected values are most noticeable in the centromeres of chromosomes 6, 7 and 8. The overall probability of the data occurring within the hypothesis is < 0.05.

Table 2

Distribution of breakage frequencies in centromeres

Centromere of Chr. No.	Obs. breaks	$\frac{(f - F)^2}{F}$
1	8	0.21
2	10	0.04
3	12	0.72
4	5	2.06
5	10	0.04
6	2	5.82
7	15	3.34
8	16	4.63
9	7	0.61
10	9	0.02
		17.48
Expected: F = 9.4	D.F. = 9	$p < 0.05$

On the basis of the arbitrary assignment of a length of 2μ to the centromeres, the break frequency within centromeres is somewhat higher than that found in the rest of the chromosome. Since this length (2μ) is probably an over-estimate, the frequency per unit-length within centromeres is likely to be even more different from the arm frequency.

The observed values for translocation pairs between all segments are contained in Table 3. (It should be emphasized that these pairs are, of course, evidenced cytologically as two translocations). Departures from expected values are contained in Table 4. Pairs which show striking departures from expectation are 5L-8S, 5L-9L, 5L-10S and 10S-8L. It will be noticed that the deviations are on an arm basis only; other pairwise values within the same chromosomes are unremarkable. The X^2 value for

Table 3

Observed values for breakage frequencies* ordered by pair associations. Data from Longley (1958, 1961). Σ observed frequencies = 962.

		1		2		3		4		5		6		7		8		9		10	
		S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L
1	S			6	5	1	8	4	9	6	13	2	13	6	6	2	5	1	12	0	3
	L			6	9	5	12	8	9	11	9	2	8	2	14	4	15	4	7	2	11
2	S					3	10	2	7	4	5	3	8	0	4	1	4	1	6	0	5
	L					7	11	11	11	12	8	2	11	4	8	5	14	3	3	2	7
3	S							3	5	4	1	3	5	0	6	3	8	1	7	1	5
	L							2	4	2	13	1	13	2	8	4	13	3	14	0	6
4	S									8	6	4	7	2	4	0	15	0	5	0	3
	L									7	11	2	11	1	6	2	4	5	12	2	7
5	S											4	9	5	4	4	13	2	6	2	5
	L											2	12	1	8	8	11	5	19	8	6
6	S													0	4	1	1	0	5	1	2
	L													4	11	1	8	5	9	5	10
7	S															1	1	2	4	0	1
	L															2	10	5	6	1	2
8	S																	0	3	1	0
	L																	8	7	9	4
9	S																			1	3
	L																			1	6

*Excluding inversions and translocations involving centromeres.

Table 4

Within cell contributions to Chi Square Test for pair-wise associations of chromosome arms. The sum of the four components represents Chi Square values for pair-wise associations between chromosomes.

		1		2		3		4		5		6		7		8		9		10	
		S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L	S	L
1	S			0.2	1.7	3.1	0.4	0.4	4.9	0.1	3.9	0.2	1.9	3.1	0.6	0.2	1.5	1.9	4.4	2.3	1.9
	L			1.2	0.6	0.3	0.1	0.1	0.5	0.4	0.1	0.8	1.2	0.9	0.9	0.1	1.2	0.1	0.4	0.3	0.7
2	S					0.3	0.4	1.3	0.0	0.5	0.2	0.2	0.1	2.4	1.1	0.7	1.2	1.3	0.1	1.9	0.0
	L					0.7	0.0	5.3	0.4	3.4	0.0	0.3	0.3	0.2	0.0	1.7	3.2	0.2	2.2	0.1	0.0
3	S							0.0	0.0	0.0	2.5	1.2	0.0	1.7	0.4	1.4	1.9	0.5	2.7	0.1	0.4
	L							2.9	3.7	4.4	2.1	1.6	0.7	0.6	0.3	0.2	0.8	0.4	5.1	2.7	0.3
4	S									3.3	0.4	2.9	0.4	0.0	0.3	1.7	16.9	2.3	0.2	1.4	0.3
	L									0.0	1.5	0.2	0.4	1.3	0.7	0.2	2.6	0.4	4.1	0.0	0.0
5	S											1.4	0.5	3.0	1.0	1.6	5.5	0.3	0.1	0.0	0.0
	L											0.1	2.3	1.0	0.1	13.4	1.5	1.0	29.9	18.1	0.0
6	S													1.0	0.7	0.9	1.7	1.2	3.7	0.1	0.0
	L													0.3	0.7	1.7	0.1	0.4	0.7	3.0	1.6
7	S															0.0	1.3	0.4	1.3	0.8	0.7
	L															0.1	0.4	0.6	0.0	0.6	2.9
8	S																	1.2	0.4	0.1	2.0
	L																	2.8	0.0	19.5	1.0
9	S																			0.0	0.0
	L																			0.3	0.2

the whole table is significant at well below the 0.005 level.

Translocation pairs involving centromere breaks are shown as observed values in Table 5. Because of the very small size of the centromeres, the expected frequencies for pairwise translocations would have been too low for satisfactory use of the X^2 test. It will be noticed, however, that the majority of translocations involving a centromere also involve a second one. If the hypothesis of random reassociation of breaks is correct, such a situation is highly unlikely in view of the small size of the centromeres, even taking into account their greater propensity for breakage. This is shown quite clearly in Table 6, where the X^2 value of 978.77 would be significant at probability levels much more stringent then 0.005.

Table 5

Translocation pairs involving centromeres - observed values

	C2	2L	C3	C4	5S	C5	C6	C7	C8	8L	9S	C9	C10
C1	2	1				1			1				1
C2			1	3	1				2				
3S													1
C3				1		3		2	2			1	
4S													1
C4						1							
C5								2	1			1	1
C6										1	1		
6L												1	
C7									6			3	2
7L													1
C8												1	2

C = centromere S = short arm L = long arm

Table 6

Distribution of reassociation pairs, including centromeres, recognising the difference in break frequency between arms and centromeres*

	Obs	Expected	$\frac{(f - F)^2}{F}$
Between arms	962	928.28	1.22
Between arms and centromeres	8	82.84	67.61
Between centromeres	43	1.86	909.94
	1013	1012.98	978.77
D.F. = 2		$p < 0.005$	

*Expected values are on the basis of pachytene length (see text) and corrected for the demonstrated higher breakage frequencies in the centromeres.

It has been demonstrated that in maize the hypotheses of random breakage and random reassociation are not supported by the data. Inasmuch as these data represent the most extensive collection, further analyses may reveal new concepts concerning the composition and organisation of nuclear chromatin.

R. C. Jancey
D. B. Walden

4. Principal Components Analysis applied to addition segments in chromosome 9.

In an attempt to quantify differences within chromosomal parameters on a biometrical basis, several types of tests, both parametric and non-parametric, were performed. These types of repetitive testing lead ultimately to the acceptance or rejection of the null hypothesis of no difference. Other types of analyses are available (see Jancey, this News Letter). Of particular noteworthiness for our ongoing interest in karyotype analysis are principal components analysis (PCA) and graph theory.

The source material for our study consisted of six reciprocal interchange stocks. The meiotic post-interchange length of 9S in these