Application of Graphic Methods in Karyotype Analysis

By Viera Černá and Ján Černý

Institute of Biology and Institute of Mathematical Analysis

Šafarik's University, Košice, Czechoslovakia

(Received for publication December 27, 1967)

Treating karyotype analysis the question of the corresponding chromosome pairs in the given diploid cell picture appears. The solution of this problem can be worked by mathematic methods and the number of pictures being large the digital computers can be also used. This possibility is described by P. M. LaBauve, R. J. LaBauve, D. F. Petersen (1965) and R. S. Ledley (1964). The application of computers in the case of the small number of the pictures, however, may be ineffective and some "manual" efficient methods for the above mentioned problem are to be found out.

The main purpose of the present paper is to suggest a graphic method, where one point graph corresponds to the every chromosome picture. The graph is treated by a simple nomogram. The application of nomography in biology is known already, we can cite from the recent papers, e. g. Z. Šesták (1966).

Method

Let us assume that the chromosomes on the diploid cell chromosome picture are preliminarily divided into morphologically different groups (e.g. according to the presence of satelites, or the approximate izobrachiality). Every group is treated independently and the chromosome pairs are formed only within the group (where every pair is considered to be able to be made).

Every chromosome is characterised by the two numbers — the lenghts of the short and the long arm. These two numbers can be considered as coordinates of a point in a plane and we obtain a group of plane points, corresponding to the group of chromosomes. As the chromosomes in one group are considered to be morphologically analogous (otherwise they could not belong to one group) the most presumably one chromosome pair will be formed by the chromosomes corresponding to the nearest points in the plane. If the number of the chromosomes in one group (and therefore the points in one plane) is larger, the formation of the corresponding point pairs may be ambiguous. Therefore we need:

1. criterion to enable us to formulate what is meant by the „better" set of point pairs from the given two sets,
2. the algorithm, leading to the best (= the most probable) set of pairs in our point graph.

The criterion

Two chromosomes forming a homologue pair are theoretically supposed to be of the same arm lengths. Measuring them we obtain the results x_1, y, for shorter and longer arm of the first chromosome and x_2, y, for the second one. x_1, y_1, x_2, y_2 are the values of some random variables X_1, Y_1, X_2, Y_2. When we put no conditions on these variables, then no method can be obtained. Conversely assuming any hypothesis concerning them, we can hardly perform any test before having the method for homologue pair finding which method is only to be designed according to these assumptions. The authors hope that the following suppositions which are usual in the other „measurement" problems, are in agreement with the reality:

(i) X_1, Y_1, X_2, Y, are mutually independent (and also independent on the measurements at other chromosomes of the cell).

(ii) X_1, Y_1, X_2, Y_2 are normally distributed and $MX_1 = MX_2$, $MY_1 = MY_2$.

(iii) X_1, Y_1, X_2, Y, have the same variances σ^2.

The assumption (i) is based on the fact that the measurement procedure is applied on every branch subsequently and independently and that the other influences could be also assumed independent.

(ii) follows from the great number of small influences causing the errors in measurement (see the paper by Matern and Simak, 1968).

The assumption (iii) is the weakest and can be fully correct only when the chance influences consist only in measuring procedure and in the grain of the photograph (neglecting the other reasons as the nonlinearity of a photograph mapping etc.).

Having (i) — (ii) fulfilled, we may suppose that X_2-X_1, Y_2-Y_1 are normal random variables with zero mean values and variances $\sigma^2_0 = 2\sigma^2$.

On the other hand, if these two chromosomes do not form one real homologue pair, then $MX_1 \neq MX_2$ or $MY_1 \neq MY_2$, and at least one from chance variables X_1-X_2, Y_2-Y_1 has a non-zero mean values.

Let us assume that in the considered group we have 2n chromosomes with arm lengths $x_1, y_1 \ldots\ldots, x_{2n}, y_{2n}$. Let us form the pairs $[(x_{i_1}, y_{i_1}); (x_{j_1}, y_{j_1})], \ldots, [(x_{i_n}, y_{i_n}); (x_{j_n}, y_{j_n})]$. There are many such sets of point pairs but we need the set corresponding to the real homologue pairs with the greatest probability, i. e. the set for which the values $x_{j_1}-x_{i_1}$; $y_{j_1}-y_{i_1}$; ...; $x_{j_n}-x_{i_n}$; $y_{j_n}-y_{i_n}$ are the most probable values of 2n independent normal chance variables with zero mean value and the equal variance a^2. According to the well known properties of multidimensional normal distributions will be the n-tuple of values more probable, if the total $(x_{j_1}-x_{i_1})^2 + (y_{j_1}-y_{i_1})^2 + \ldots + (x_{j_n}-x_{i_n})^2 + (y_{j_n}-y_{i_n})^2$ is lower.

Thus we have found the criterion. Translated to the geometrical language, the ordering of chromosome pairs will be the more probable the sum of the distance squares in corresponding point pairs will be the lower. There we have confirmed the fact, that

$$(x_{j_1} - x_{i_1})^2 + (y_{j_1} - y_{i_1})^2$$

is only the square of the distance between the points $(x_{i_1}; y_{i_1})$ and $(x_{j_1}; y_{j_1})$ in a plane.

The algorithm

First of all it is necessary to emphasize, that the algorithm, leading reliably to the optimum (i. e. to the minimum sum of the distance squares) is not suitable for

„manual" solution, for the usual number of chromosomes it can hardly be used even in digital computers. Therefore our algorithm must be more simple. It cannot ensure the optimum but only the comparative good results (in our experiments, however, we obtained mostly only optimum results!).

1. step of the algorithm: Trying to connect optionaly the nearest points in our graph we form the pairs from two points in such a way, that every point must be exactly in one pair.

2. step: Changing the points between two selected pairs we try to improve the solution. *Fig. 1* demonstrates it. The selected pairs are designated as AB and CD. We after replacing the points B and D in pairs. We consider, whether the connection AD and CB would not be better, i. e. whether

$$(x_D - x_A)^2 + (y_D - y_A)^2 + (x_B - x_C)^2 + (y_B - y_C)^2 < (x_B - x_A)^2 + (y_B - y_A)^2 + (x_D - x_C)^2 + (y_D - y_C)^2,$$

that means that

$$d^2(AD) + d^2(CB) < d^2(AB) + d^2(CB),$$

where e. g. $d(AB) = (x_B - x_A)^2 + (y_B - y_A)^2$

designates the distance of the points A and B in a plane.

In order to save time, we do not compute the above squares but we use the nomogram from the *fig. 2*. On its both sides there are two square scales for $d(X_1X_2)$, and $d(X_3X_4)$, where $X_1X_2X_3X_4$ are the considered points, and in the middle there is the linear scale for $h = d^2(X_1X_2) + d^2(X_3X_4)$. If we obtain d (AB) and d (CD) after measuring on the graph we find them on marginal scales and then we connect these two points by a stright line. The intersections of this line with the medial scale expresses the value $h_1 = d^2(AB) + d^2(CD)$. If we shall do the same with d (AD) and d (CD), we shall obtain $h_2 = d^2(AD) + d^2(CD)$ and we shall immediately see what value between h_1, h_2 is smaller and whether the pairs AB, CD are better than AD, CB.

The procedure of second step may start in the points with the lower subscripts and finish at the greatest one or conversely.

Finding no possibility for such a change, our work is finished.

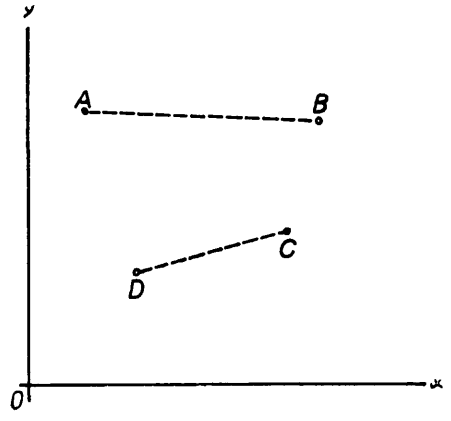

Fig. 1.

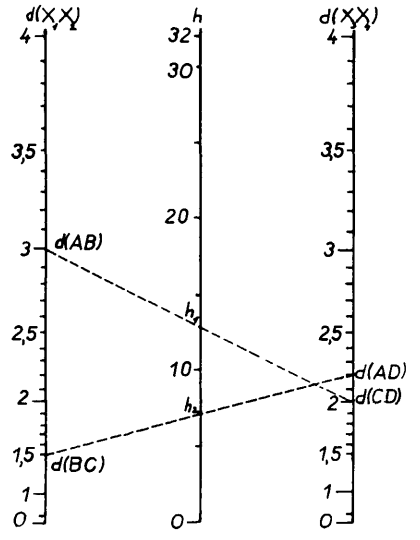

Fig. 2. — The nomogram for the value determination $h = d^2(x_1x_2) + d^2(x_3x_4)$. — The measured values are marked on the marginal scales, joined by the straight line and its intersection with the medium scale determines the value h.

Example

Let us have the group of izobrachial chromosomes *(Fig. 3)*. The corresponding points are on *Fig. 4* and the initial pairs after the first step of our algorithm are on the *Fig. 5*

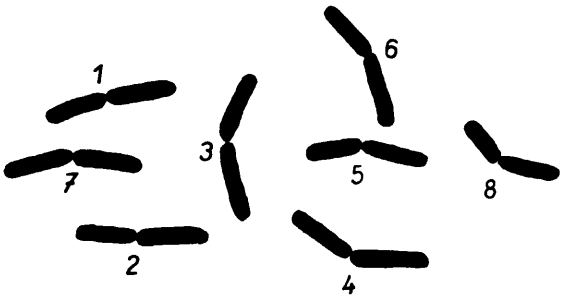

Fig. 3. — The group of eight izobrachial chromosomes (drawing for the example).

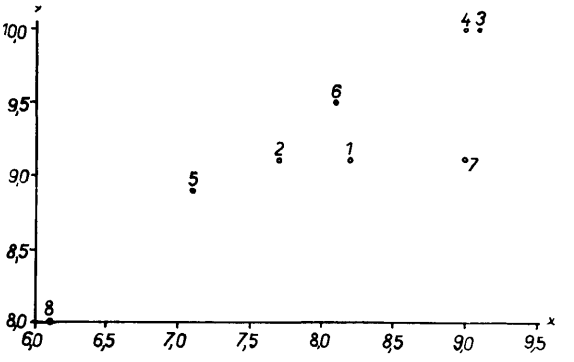

Fig. 4. — Points in a plane with the coordinates equal to the arm lengths of the corresponding chromosomes.

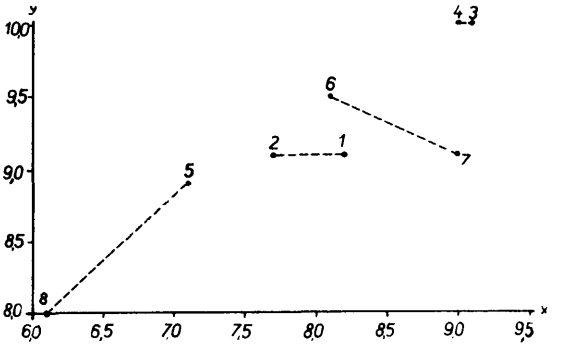

Fig. 5. — Pairs of chromosomes formed by judgement (are not in fact the best ones).

According to the second step we shall try to change the 4th and 7th chromosome, i. e. to replace the 3rd, 4th and 6th, 7th by 3rd, 7th and 6th, 4th. Out of our graph and the nomogram *(Fig. 2):*

$$d^2(3;4) + d^2(6;7) < d^2(3;7) + d^2(6;4)$$

resulted therefore the replacing $4 \longleftrightarrow 7$ is inconvenient. Then we try to replace $1 \longleftrightarrow 6$. This change will prove to be convenient, 1,7 and 2,6 will be better than 1,2 and 6,7. Than we can find no convenient change and therefore our algorithm finishes its work at the pairs 1,7, 2,6, 3,4, 5,8.

The case of more than one cell

Till now we have studied the applications of the graphic method in one cell photograph evaluation. In the sequel we shall suggest a method for many cells, using the abovementioned tools.

Let us consider that we have already evaluated k photographs of k different cells, i. e. we have formed the pairs of chromosomes assumed to be homologue.

1st step: Normalisation (Equalisation of measure scales). Every arm should be divided by the average value of the chromosome lengths in a photograph, i. e. by the value

$$\frac{1}{2n} \sum_{i=1}^{2n} (x_i + y_i).$$

2nd step: Reduction.

Every pair of chromosomes $[(x_i, y_i); (x_j, y_j)]$ should be replaced by a "average chromosome" $\left(\frac{x_i + x_j}{2}, \frac{y_i + y_j}{2}\right)$.

After having performed the steps 1 and 2 we obtain k groups of n "chromosomes":

$$(x_1^{(1)}, y_1^{(1)}; \ldots\ldots; (x_n^{(1)}, y_n^{(1)}),$$
$$(x_1^{(k)}, y_1^{(k)}; \ldots\ldots; (x_n^{(k)}, y_n^{(k)}).$$

The ordering in every group is immaterial. We wish to find the same chromosome on every photograph, i. e. in every of k groups. The simultaneous treatment of all groups is complicated even for computers. In manual method we have to treat them subsequently.

3rd step: Finding average idiogram.

Let us draw $(x_1^{(1)}, y_1^{(1)}; \ldots\ldots; (x_n^{(1)}, y_n^{(1)})$ in a graph as a points $(x_1^{(2)}, y_1^{(2)}; \ldots\ldots; (x_n^{(2)}, y_n^{(2)})$ as a small crosses *(Fig. 6* is an example). Let us search the pointcross pairs such that the sum of distance squares is minimum. The search can be done using the nomogram *(Fig. 2)* like in the one cell treatment. Then every point-cross pair is replaced by an average point in the center of the point-cross abscissa. These points should be drawn in a new graph and $[x_1^{(3)}, y_1^{(3)}]; \ldots\ldots; [x_n^{(3)}, y_n^{(3)}]$ added as crosses. The pair forming is to be repeated and every point-cross pair is to be again replaced by a new average point, not now a center but a 2 : 1 weighted average point in ⅓ distance from previous one (and ⅔ from a cross). Then $[x_1^{(4)}, y_1^{(4)}]; \ldots\ldots$ are added etc.

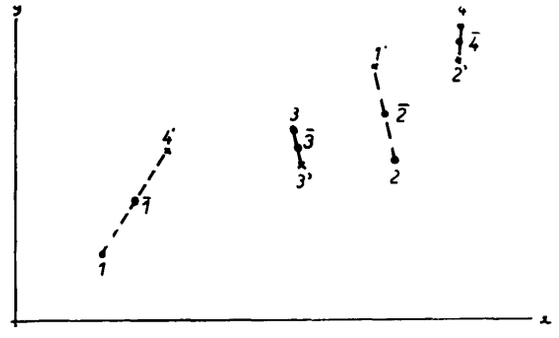

Fig. 6. — Average chromosomes from the first picture (1, 2, 3, 4), from the second (1', 2', 3', 4') and their average representatives ($\bar{1}$, $\bar{2}$, $\bar{3}$, $\bar{4}$).

Finally if we have treated first k—1 groups and have found the average point we add $(x_1^{(k)}, y_1^{(k)}); \ldots\ldots; (x_n^{(k)}, y_n^{(k)})$ as the crosses and form the point-cross pairs and replace them by (k—1) : 1 weighted average i. e. the point in 1/k distance from the previous one and (k—1)/k from the cross.

These new points $(\bar{x}_i, \bar{y}_i)$ represent the average quantitive idiogram of the treated species: $\bar{x}_i$, $\bar{y}_i$ are the arm-lengths of the i-th chromosome in it.

Complementary remarks

The authors hope that their "manual" method needing no computer may be useful for cytologists, although it is only an approximate method and its mathematical foundations are not fully prooved. Its possible weak points are following:

1. The assumption iii (and perhaps also i, ii).

2. The approximative algorithms, which do not guarantee the reading the optimum, namely in the case of more cells. In the last case the optimum is difficult not even for calculation but even for formulation (the autors did not do so).

3. There are no studies of the risk of error, because they are very complicated.

The cytologists have to estimate the "risks" before deciding to use this method, another one or none. The authors had used this method several times and obtained the satisfactory results namely in the case of one cell evaluation.

Summary

In the present paper the possibility of the graphic method application in the karyotype analysis is discussed. To form the pairs of corresponding chromosomes from the pictures of the diploid chromosomes is to be solved. It is assumed that the results of the arm lengths measurements are normally distributed with the equal variance. As the criterion we use the minimum sum of squares measured arm differences of the pairs (it is the analogue of the least-squares method). The above squares will be not calculated but obtained from the graph *(Fig. 4)* and the nomogram *(Fig. 2)*.

References

LaBauve, P. M., LaBauve, R. J., and Petersen, D. F.: A digitized comparator for karyotype analysis. Jour. Heredity **56**: 46—52 (1965). — Ledley, R. S.: High-speed automatic analysis of biomedical pictures. Science **146**: 216—223 (1964). — Matérn, B., and Simak, M.: Statistical problems in karyotype analysis. Hereditas **59**: 280—288 (1968). — Sesták, Z.: Construction of a simple nomogram for evaluating two-wawe-length spectrophotometric determination of chlorophylls. Biologia plantarum **8**: 97—109 (1966).