Theor Appl Genet (1987) 73:630–634

Computerized analysis of chromosomal parameters in karyotype studies

J. L. Oud [1], P. Kakes [2] and J. H. De Jong [3]

[1] Hugo de Vries Laboratory, Cytogenetics Unit, University of Amsterdam, Kruislaan 318, NL-1098 SM Amsterdam, The Netherlands
[2] Department of Ecology, Laboratory of Biology, Free University, De Boelelaan 1087, NL-1081 HV Amsterdam, The Netherlands
[3] Department of Genetics, Agricultural University, Gen. Foulkesweg 53, NL-6703 BM Wageningen, The Netherlands

Received June 8, 1986; Accepted October 21, 1986
Communicated by R. Riley

Summary. A study is presented of the possibilities and limitations of semi-automated karyotype analysis on the basis of chromosome length and centromere index. A number of computer programs have been developed for 1) quick and precise measurements of chromosome arm length with the help of a graphics tablet, 2) computing (relative) length and centromere index and statistical analyses of the data, and 3) representation of these chromosomal parameters in two-dimensional scattergrams. An ellipse representing 95% of the probability mass is drawn around the bivariate mean of each chromosome. The size and orientation of the axes are calculated from repeated measurements of the chromosomes of one metaphase plate. If there is a correlation between length and centromere index, which is often the case, the axes of the ellipse are tilted. Incorporation of such a covariance analysis proved to be of great importance for an accurate karyotype analysis. The "Computer Aided Karyotyping" package does not contain routines for an automated classification of the chromosomes. The main reason is that the variation in length and centromere index of a given chromosome in different cells is often much larger than the variation between nonhomologous chromosomes. In addition, it was our aim to develop universal karyotyping aids which can be used regardless of the species studied.

Key words: Karyotype – Automated analysis – Chromosome measurement – Computer – Cytogenetics

Introduction

Until now, devices for automated karyotype analyses have been mainly restricted to human cytogenetics, since a prerequisite for successful automation is the availability of large amounts of high quality G-banded metaphases of one, well-defined karyotype. The considerable differences between plant species and technical problems related to plant chromosomes have limited the use of computers in plant cytogenetics.

Examples of such problems are:

1) General applicable techniques for producing high quality preparations with large numbers of well spread metaphase plates are scarcely available. This is partly caused by the cell wall which prevents protoplasts from swelling in a hypotonic solution, so that spreading techniques are not directly usable. Exposure of the cell material to pectolytic enzymes can overcome this problem only to some extent although their use is becoming more and more popular in plant cytogenetics, especially in chromosome banding techniques. The enzymatic maceration also enables air-dry techniques, such as these being used in mammalian cytogenetics (Mouras et al. 1978; Pijnacker and Ferwerda 1984).
2) The field of plant cytogenetics involves many different species with karyotypes showing large variations in chromosome size, centromere position, chromosome number and ploidy level.
3) It is difficult (and for the moment often impossible) to produce discriminating banding patterns in most plant species. G-banding cannot be obtained in plant chromosomes (Greilhuber 1977), whereas C-banding in plants is mostly restricted to centromere and/or telomere regions.

Notwithstanding the above mentioned limitations, intrinsic to plant cytogenetics, there are several steps in karyotype analysis where computer systems can be used very well. There is a general interest in precise chromosome arm measurements and statistical analysis of the data. A current method of karyotype analysis is the visualization of mean- and l.s.d. (least significant difference)-values of the relative length and the centromere index of each chromosome in a two-dimensional scattergram. The procedure is sound and straigthforward when clear differences exist between all the chromosomes of the haploid karyotype. Often the researcher has to deal with chromosome complements in which at least some of the chromosomes show overlap with respect to relative length

and centromere index, or both. When additional information, such as chromosome banding is lacking, various errors may occur.

The first two are "reversal of order" and "reversal of arms" in the karyogram (Matern and Simak 1968). The following two errors have to do with the graphic representation of the data from repeated measurements of the length of chromosome arms. In several publications chromosomes are considered different if the rectangles drawn around the l.s.d.-values on both axes of two chromosomal parameters do not overlap. Such a procedure is clearly too conservative, as the vertices of the rectangle correspond with extreme values for both estimators, which will occur with a much lower probability than that for which the l.s.d. is calculated. A solution for this problem is: assume that relative length and centromere index are normally distributed, and calculate the joint distribution of the two estimators. The actual parameters will, with a predefined probability, lie within an ellipse around the bivariate mean. The properties of such a confidence ellipse (principal axes and slope of the major axis) are calculated from repeated measurements of the chromosomes of one cell. If there is no correlation between the values of relative length and centromere index of a chromosome, the axes of the ellipse will overlap the "l.s.d. cross" (note that they will not coincide as the ellipse refers to a bivariate distribution and the l.s.d. to an univariate one). However, if there is a correlation, the axes of the ellipse will be tilted and the resulting figure can be quite different from the simple ellipse. The calculations needed to determine the properties of the confidence ellipses and their construction by hand are tedious and therefore error prone.

Our aim is to develop computer procedures to take over the computation, storage and presentation of data which are relevant for karyotype analysis, whereas the crucial decisions about chromosome classification remain the investigator's duty. These procedures have to work irrespective the karyotype of the plant (or animal) species. The present paper demonstrates the possibilities and limitations of our approach to Computer Aided Karyotyping (CAK). Two plant species, *Petunia hybrida* and *Pyrrhopappus carolinianus*, whose chromosome portraits are well known to the authors, have been used for testing and demonstrating the computer programs. Since these programs intend to be universal in design we also used the karyotype of 'Chinese' hamster.

Materials and methods

A) Hardware

The system consists of a Hewlett-Packard 9816S personal computer (now renumbered to HP 9000 type 216), HP 9121D dual disk drive, HP 9111A graphics tablet, HP printer and a HP 7470A plotter.

B) Software

The CAK package consists of five programs. The CHROMDIGIT program comprises the whole measuring procedure. Each chromosome arm is measured six or ten times, i.e. in alternating order three or five times one chromatid and three or five times the other. The centromere regions are excluded from the actual measurements. The CHROMDIGIT program offers at any time during a measuring session the choice between measuring continuously, or point by point. The first method implies that the graphics tablet checks the position of the stylus ten times per second, as long as the stylus follows the chromatid. With the second method the tablet digitizes a single point when the stylus is pressed down. The CHROMosome CALCulation program CHROMCALC computes the mean relative or absolute or length and centromere index ((s/l+s) * 100) of each chromosome (Cf. Nomenclature rules of the Denver Study Group 1960; Chicago Conference 1966), and standard deviations and 95% confidence intervals of both parameters. On the basis of these data, the CONFELLIPS (CONFidence ELLIPSe) program plots a scattergram of the mean length and centromere index. A covariance analysis (Sokal and Rohlf 1981) is incorporated to calculate the degree of correlation between chromosome length and centromere index. Around the bivariate mean of each chromosome a 90, 95 or 99% equal frequency ellipse (also known as confidence or probability ellipse) is drawn. Further, the CAK package consists of a routine to enter data of measurements made by hand (CHROmosome MEASURe program CHROMEASUR) and to draw an idiogram (program IDIOGRAM).

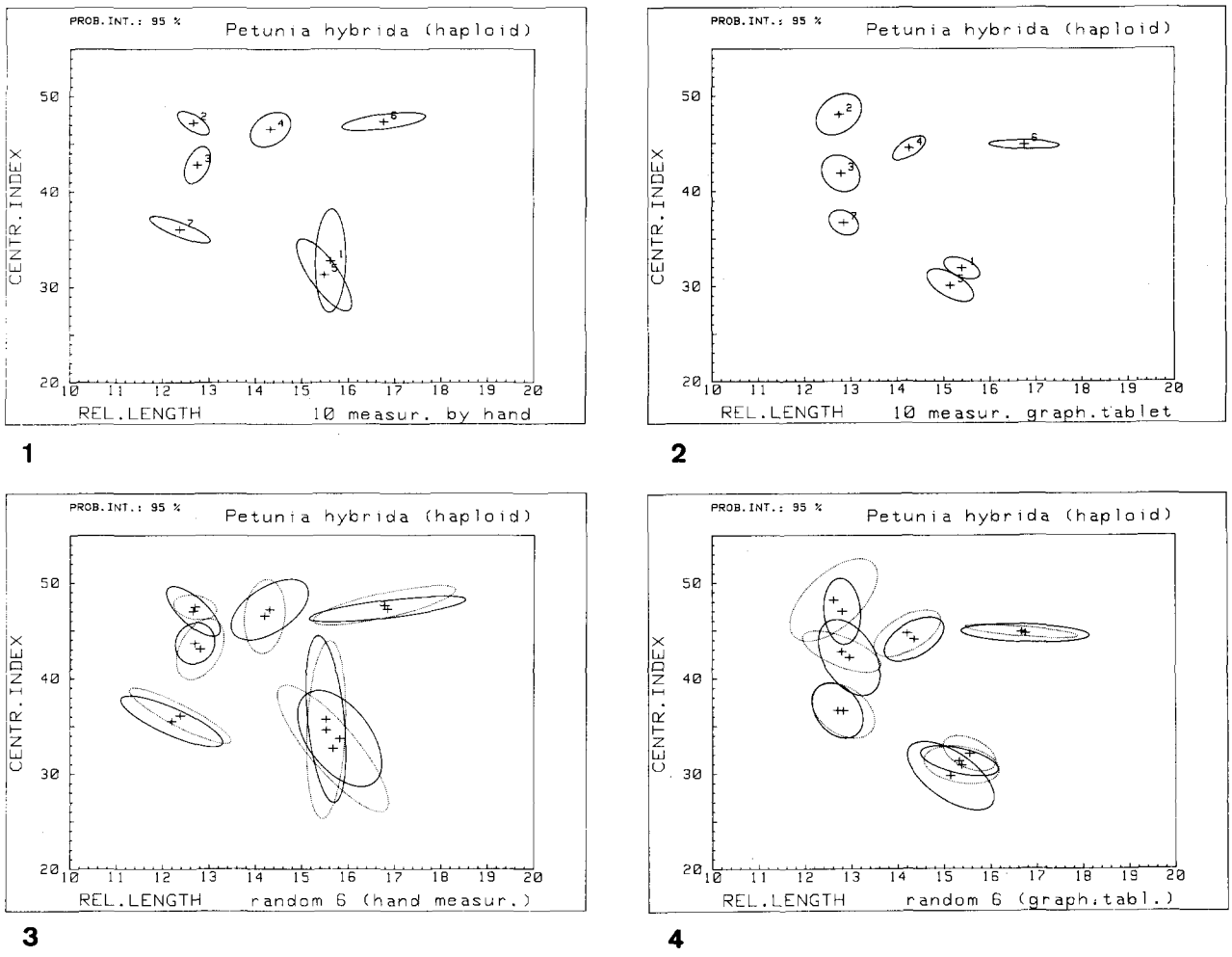
C) Bioware

Karyotype analyses were made from (1) α-bromonaphtalene or colchicine pretreated root tip metaphases of haploid *Petunia hybrida* ($2n = x = 7$) and diploid *Pyrrhopappus carolinianus* (Asteraceae) ($2n = 12$), and (2) colcimid pretreated bone marrow metaphases of 'Chinese' hamsters (*Cricetulus griceus*) ($2n = 20 + XY$). Chromosome preparations were made according to the appropriate standard methods and stained with Feulgen, lactopropionic orcein or Giemsa.

Results and discussion

Figures 1–7 show the results of CAK analyses of *Petunia hybrida* (haploid), *Pyrrhopappus carolinianus* and the 'Chinese' hamster.

First we performed an analysis to see if we could strike a reasonable balance between the number of cells which have to be measured and the number of repetitive measurements of a cell. Therefore, data were used of measurements at photographs, made with a ruler of three *Petunia* metaphases (Fig. 1 shows the CONFELLIPS plot of one cell). Ten repetitions were made by the same observer. The chromosomes were measured in random order. Next, five subsets of the data of each cell were constructed by random sampling six measurements from each set of ten. Two examples are shown in Fig. 3. Comparison of the subsets with each other and with the parent set shows that, although the equal frequency ellipses of the former are larger, as was expected, there is neither a substantial difference between subsets nor between subsets and the parent set. In all photographs two chromosomes, the longest one and the satellite chromosome, clearly stand out. An analysis of variance of both length and centromere index is conducted on the 30 (3×10) observations made on each of these chromosomes. As shown in



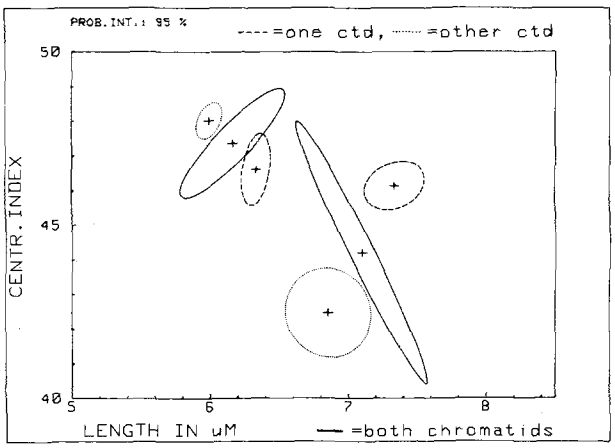
Figs. 1–4. CONFELLIPS plots of a haploid *Petunia hybrida* metaphase on the basis of ten measurements (**1** and **2**) made by hand with a ruler (**1**) or with a graphics tablet (**2**). The numbers 1–7 refer to the order in which the chromosomes are measured; the satellite chromosome is no. 5. Figs. **3** and **4** are two subsets of six measurements each, randomly chosen from the data of Figs. **1** and **2**, respectively. Further explanation in the text

Table 1, the variance due to measurement errors is negligible compared to the variance between cells. We therefore argued that many repetitions do not contribute much to the accuracy of a karyotype analysis, and settled for a number of six repetitions (to accomodate another number, only some simple changes have to be made in the program).

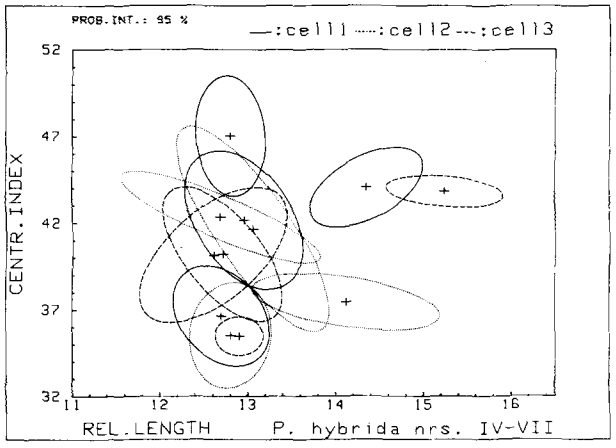
All further length measurements of chromosome arms were made with the graphics tablet. To analyse the precision of this method, the same cell as shown in Fig. 1 is measured ten times by another observer (Fig. 2), whereafter five subsets of six measurements per chromosome are also made (Fig. 4). Comparison of the hand made and the graphics tablet data (Figs. 1, 3 and 2, 4) indicates that in general the two methods are equally accurate. Probably the difference in observer is more important than the difference between measuring methods. However, in the case of crooked chromosomes, the use of a graphics tablet is preferred.

CHROMCALC assumes that relative length and centromere index are normally distributed. However, the assumption of normality is formally incorrect, as relative length and centromere index are percentages, respectively, proportions. The deviation will always be small for the relative length, whereas for the centromere index only near acrocentric chromosomes will deviate from normality.

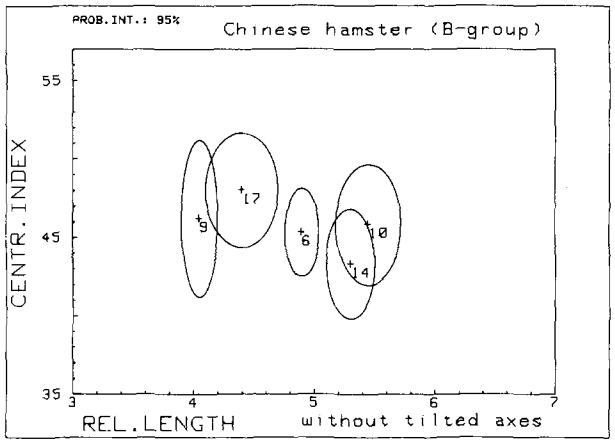
The construction of confidence ellipses and the introduction of covariance analyses are, as far as we know, unique properties of the CAK package. The various CONFELLIPS plots demonstrate that a correlation between chromosome length and centromere index is a rule rather than the exception. Although we have no conclusive explanation for this correlation,



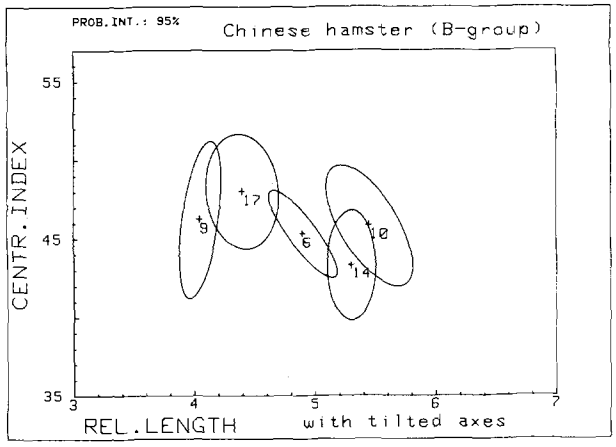
5



7



6a



6b

Fig. 5. CONFELLIPS plot of two *Pyrrhopappus carolinianus* chromosomes. Of both chromosomes the two sister chromatids are measured separately six times, and a mixture of the two sets of data is drawn. Further explanation in the text. **Fig. 6.** CONFELLIPS plots of the 'Chinese' hamster chromosome pair III and IV and the X-chromosome, with the axes of the equal frequency ellipses set perpendicular to the X- and Y-axis (**a**) and with tilted axes because of a co-variance between relative length and centromere index (**b**). **Fig. 7.** CONFELLIPS plot of the chromosomes IV–VII of three *Petunia hybrida* cells from the same plant

what often occurs is that variation in total chromosome length is due to variation in one arm only. Such a variation could be the result of an ambiguity in the determination of the end of one chromosome arm, or it has to do with a difference in arm length between the sister chromatids. The latter is the case in the example of the two *Pyrrhopappus carolinianus* chromosomes shown in Fig. 5. For both chromosomes, ellipses were plotted on the basis of data from (1) six measurements of one chromatid, (2) six measurements of the other chromatid, and (3) a combination of the measuring series 1, 3 and 5 of the first chromatid and 2, 4 and 6 of the other chromatid. The slope of the major axis of the ellipse representing both chromatids can be explained for the most part by a relative large difference in length between the sister chromatids of one arm. The chromatids of the short arm of the longest chromosome showed a difference of 15.6%, whereas a difference of 9.4% exists in the long arm chromatids of the other chromosome. Note, however, that there is also some correlation between length and centromere index when only one chromatid is measured.

To show the importance of being aware of the possible correlation between chromosome length and centromere index, the male 'Chinese' hamster "B-group" chromosomes are depicted in Fig. 6a (with the axes of the confidence ellipses set perpendicular to the X- and Y-axis) and Fig. 6b (with tilted axes). The B-group consists of the third and fourth pair of autosomes and the X-chromosome. On the basis of Fig. 6a, it is plausible to identify the numbers 10 and 14 as chromosome pair III, the numbers 9 and 17 as chromosome

Table 1. Four analyses of variance (ANOVA) on 3×10 data each, of chromosome length and centromere index of chromosome I and the satellite chromosome in three metaphase plates of *Petunia hybrida* (haploid)

ANOVA: length chromosome I

Source	Df	SS	MS	F-ratio	F-prob.
Total	29	53.0387			
Between cells	2	25.1207	12.5603	12.1473	0.00017
Within cells	27	27.9180	1.0340		

ANOVA: centromere index chromosome I

Source	Df	SS	MS	F-ratio	F-prob.
Total	29	67.5074			
Between cells	2	31.9063	15.9532	12.0989	0.00018
Within cells	27	35.6011	1.3186		

ANOVA: length satellite chromosome

Source	Df	SS	MS	F-ratio	F-prob.
Total	29	20.1297			
Between cells	2	12.5527	6.2763	22.3652	<0.00001
Within cells	27	7.5770	0.2806		

ANOVA: centromere index satellite chromosome

Source	Df	SS	MS	F-ratio	F-prob.
Total	29	247.4931			
Between cells	2	71.1166	35.5583	5.4433	0.01032
Within cells	27	176.3765	6.5325		

pair IV, whereas number 6 is likely to be the X-chromosome. However, Fig. 6 makes clear that it can only be concluded that the individual chromosomes within the B-group of this metaphase plate cannot be identified. Comparable situations may occur in karyotype analyses of aneuploids or cells with structural chromosome mutations.

The system has some important and intentional limitations so as to avoid the risk of reversal of order and reversal of arms in the karyogram. In this respect there is a striking difference between our CAK approach and that of others (McGurk and Rivlin 1983; Green et al. 1984). First, we have not tried to incorporate chromosome classification procedures, since the variation between comparable chromosomes in different metaphases is often much larger than the difference between non-homologous chromosomes within one cell. Fig. 7 demonstrates this phenomenon. Chromosomes IV to VII of *Petunia hybrida* are approximately, 2.5 μm long, and have a centromere index between 35 and 50. Even when three haploid cells are compared, a classification of the chromosomes IV to VII is not possible on the basis of length and centromere index only. Second, we have not incorporated a routine for automated correction if in some of the repeated measurements the length of the chromosome arm is shorter than the length of the short arm.

In view of all possible pitfalls concerning chromosome classification and our aim to develop aids for karyotype analyses which can be used irrespective of the species studied, we believe that decisions about chromosome classification have to remain the task of the investigator. As long as chromosome length and centromere index are the only available parameters for karyotype analysis, which is often the case in plant cytogenetics, the CONFELLIPS plots give all relevant information. It is our experience that a careful analysis of the CONFELLIPS plots more often forces the investigator to be reserved about the possibilities of chromosome identification, than that they provide additional information in favour of a positive identification.

A copy on a 3.5″ floppy disc (at nominal cost) of the CAK programs written for Hewlett-Packard series 200 (former 9800) PCs is available to cytogeneticists. With some minor changes the programs can also be used on HP 9000 series 300 computers.

Acknowledgements. We acknowledge Mr. J. H. Mowitz (Technical Centre FSW, University of Amsterdam) and Dr. P. C. Diegenbach (Dept. of Zoology, University of Amsterdam) for their indispensable contribution to the statistical part of the work.

References

Chicago Conference (1966) Standardization in human cytogenetics. Birth defects: original article series, vol II, no 2. The National Science Foundation, New York

Denver Study Group (1960) A proposed standard system of nomenclature of human mitotic chromosomes. Acta Genet 10:322–328

Green DM, Myers PZ, Ryena, DL (1984) CHROMPAC III: an improved package for microcomputer-assisted analysis of karyotypes. J Hered 75:143

Greilhuber J (1977) Why plant chromosomes do not show G-bands. Theor Appl Genet 50:121–124

Matern B, Simak M (1968) Statistical problems in karyotype analysis, Hereditas 59:280–288

McGurk J, Rivlin K (1983) A BASIC computer program for chromosome measurements and analysis. J Hered 74:304

Mouras A, Salesses G, Lutz A (1978) Sur l'Utilisation des Protoplastes en Cytologie: Amélioration d'une Methode recente en vue de l'Identification des Chromosomes mitotique des genres Nicotiana et Prunus. Caryologia 31: 117–127

Pijnacker LP, Ferwerda MA (1984) Giemsa C-banding of potato chromosomes. Can J Genet Cytol 26:415–419

Sokal RR, Rohlf FJ (1981) Biometry, 2nd edn. Freeman, San Fransisco