

COLLOQUE NATIONAL SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS



NICE du 26 au 30 AVRIL 1977

SYSTEME INTERACTIF DE CARYOTYPIE

J.L. BASILLE - S. CASTAN

LABORATOIRE C. E. R. F. I. A. I.U.T. INFORMATIQUE
Avenue de Rangueil 31077 TOULOUSE CEDEX

RESUME

Les chromosomes d'un caryotype sont habituellement classés selon la taille et l'indice centromérique, chacun de ces critères pouvant être mesuré à partir de la longueur, de la surface, du poids ou d'une combinaison de ces trois grandeurs.

En ce qui concerne l'indice centromérique, l'application des méthodes classiques nous a montré le manque de fiabilité des mesures obtenues et la non complémentarité de ces méthodes.

Nous avons alors choisi de ne considérer que la taille mesurée à partir de la surface et à partir du poids après normalisation de celui-ci par un recadrage statistique de la dynamique effectué globalement au niveau de chaque composante.

La méthode de classification employée est une méthode procédant en deux temps :

- affectation à chaque composante d'une classe candidate et éventuellement d'une classe suppléante.
- vérification de la cohérence des affectations avec d'éventuelles réaffectations.

Afin de faciliter l'étude de faisabilité et l'enchaînement des divers traitements mis en oeuvre (séparation des composantes connexes, diagramme de densité, rotation, cadrage de la dynamique, visualisations de résultats, etc...) nous avons réalisé un moniteur conversationnel qui permet d'aiguiller l'opérateur dans le choix des processeurs à activer, grâce à un dialogue sur console de visualisation, et gère l'exécution de ceux-ci.

SUMMARY

Chromosomes are generally classified from two criteria : length and centromeric index, each of which can be measured from length, area or any combination of these three variables.

As regards the centromeric index, experiments showed us classical methods were not reliable enough and could not be used complementarily.

So we chose to consider only the size measured from area and the weight after this has been normalized by a statistical dynamic stretching in a global way for each component.

The method we used to classify the chromosomes proceeds in two steps :

- first we assign each component a candidate class and a substitute one if possible.
- then we check the coherency of those assignments, reassign a component its substitute class if necessary (if it exists) and so may detect numerical abnormalities.

In order to test easily the feasibility of the different procedures we have carried out (component isolation, density profile, rotation, dynamic stretching, display of results etc...) and to chain them together, we achieved an interactive monitor that enables the operator to choose the procedures he wants to be executed by a dialogue on display and controls their execution.



SYSTEME INTERACTIF DE CARYOTYPIE

J.L. BASILLE - S. CASTAN

1. INTRODUCTION

Les cytogénéticiens ont l'habitude de classer les chromosomes selon deux critères : la taille et l'indice centromérique. Chacun de ces critères peut être mesuré à partir de la longueur, de la surface, du poids ou d'une fonction de ces trois variables^{1,2,3}

Nous allons voir pourquoi nous avons choisi de ne prendre en compte que la taille mesurée à partir de la surface et de la densité lumineuse et comment nous avons classé les chromosomes à partir de cette seule taille.

II METHODOLOGIE

La procédure globale se déroule en trois étapes.

II.1. SEPARATION DES COMPOSANTES CONNEXES

La première partie est une procédure générale qui consiste à séparer les composantes connexes. La méthode employée opère ligne par ligne selon l'automate représenté sur la figure 1. Elle peut donc être facilement implémentée sur un mini-calculateur. Pour chaque ligne un balayage permet de déterminer des "segments" dont les points appartiennent à une même forme. Le même seuil est choisi pour toute la photo. (ce fait sera utilisé dans l'étape suivante). Puis l'automate, par comparaison des segments des deux lignes consécutives, chaîne les segments appartenant à une même composante connexe. Quand toutes les lignes ont été traitées, une composante connexe est un ensemble de segments ainsi chaînés.

Une marge de largeur constante entourant chaque composante permet de donner le contexte. (Elle sera utilisée dans l'étape suivante).

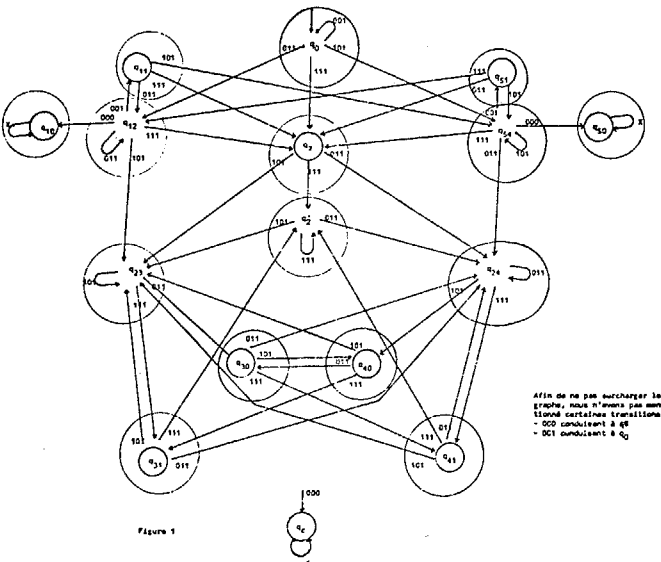


Figure 1

II.2. CRITERES DE CLASSIFICATION

Dans la deuxième partie, il s'agit de mesurer les critères qui serviront à la classification.

II.2.1. TAILLE. La longueur est relativement difficile à mesurer car les chromosomes peuvent être courbes ou tortueux. Nous avons appliqué des algorithmes de squelettisation afin d'obtenir des formes de ligne. Mais aucun ne nous a fourni, pour tous les chromosomes d'un caryotype, des squelettes topologiquement identiques aux formes classiques "X", "Y", etc... : apparition de boucles, branches parasites, bras inégaux, etc.

Notons enfin que la longueur, indépendamment de la façon dont elle est mesurée, est sujette à des variations difficiles à prendre en compte (contraction due à la colchicine par exemple).

La surface permet justement de compenser ces variations mais elle reste néanmoins sujette aux différences de luminosité puisque, qu'il soit absolu, relatif ou variable, il faut se fixer un seuil.

Le poids donne plus d'information que la surface mais il est nécessaire de le normaliser pour corriger la non-homogénéité des clichés comme nous le verrons en III.4.

II.2.2. Indice centromérique. En ce qui concerne l'indice centromérique, l'application des méthodes classiques^{4,5,6,7,8,9,10} nous a montré le manque de fiabilité des mesures obtenues compte tenu des photos utilisées, du capteur employé et du fait que nous travaillons sur un mini-calculateur. Nous avons mis en oeuvre les méthodes suivantes :

- profil de densité.
- squelettisation.
- balayage orthogonal à l'axe principal (ce qui nous a conduit à opérer une rotation des chromosomes).

D'autre part les résultats obtenus nous ont montré que de telles méthodes échouent avec les mêmes chromosomes (généralement les acrocentriques et les petits chromosomes) et qu'elles ne peuvent donc pas être utilisées de façon complémentaire.

II.2.3. Critères retenus. Compte tenu des différents résultats obtenus nous avons choisi de ne considérer que la surface et la poids.

La surface est normalement calculée d'après le nombre total de points situés à l'intérieur du contour du chromosome.

Habituellement le poids d'une forme est déterminé en sommant les valeurs de la fonction niveau-de-gris en chacun des points de la forme. Comme les clichés à partir desquels nous travaillons ne sont pas homogènes nous avons dû normaliser les résultats.

II.2.4. Recadrage de la dynamique. Nous avons défini un algorithme en trois étapes, basé sur les remarques suivantes :

1. Les composantes sont séparées avec le même seuil.
2. Une marge de largeur constante entoure chaque composante.
3. La largeur des chromosomes peut être considérée comme relativement constante.
4. Le niveau noir, de même que le blanc, n'est pas significatif.
5. Le cliché peut être considéré homogène sur des zones de la taille d'un chromosome.

La première étape consiste à redéfinir le niveau noir N . Nous prendrons comme niveau noir le maximum local correspondant au niveau de gris le plus élevé dans l'histogramme des occurrences des niveaux de gris.

La deuxième étape consiste à redéfinir le seuil. Les remarques 2 et 3 impliquent que la surface délimitée par la marge est quasi-proportionnelle à la surface de la composante. Donc, d'après les remarques 1 et 5, le poids total des points en lesquels la fonction niveau-de-gris est inférieure au seuil de séparation, devrait être quasi-proportionnel au poids de tous les points de la forme. Le rapport de proportion utilisé est le rapport moyen (différent du rapport fourni par l'histogramme moyen). Nous prendrons comme nouveau seuil le niveau N_S qui satisfait au mieux la condition ci dessus.



SYSTEME INTERACTIF DE CARYOTYPIE

J.L. BASILLE - S. CASTAN

Ces deux recadrages doivent donc répartir un ensemble de valeurs selon une distribution connue par points. Ils utilisent pour cela une même routine qui, à partir d'une fonction $f(x)$ en escalier, définie sur $[x_0, x_{n+1}[$ par :

$$f(x) = f(x_i) \text{ pour } x_i \leq x < x_{i+1} \quad (i \text{ de } 0 \text{ à } n),$$

et par :

$$f(x) = 0 \text{ ailleurs.}$$

définit une autre fonction en escalier $g(x)$ sur

$[y_0, y_{p+1}[$ par :

$$g(x) = \int_{y_j}^{y_{j+1}} f(x) dx \text{ pour } y_j \leq x < y_{j+1} \quad (j \text{ de } 0 \text{ à } p)$$

et par :

$$g(x) = 0 \text{ ailleurs.}$$

A partir de l'histogramme $h(x)$ défini comme une fonction en escalier, ces deux recadrages fournissent une fonction en escalier \bar{g} .

La dernière étape consiste à réajuster le rapport Noir / Gris calculé à partir de \bar{g} par :

$$\sigma = \frac{\bar{g}(N_{\text{Max}})}{N_{\text{Max}} - 1} \sum_{i=1}^{N_{\text{Max}} - 1} \bar{g}(i)$$

D'après les remarques 3 et 4, ce rapport devrait être le même pour tous les chromosomes. Si nous prenons le rapport moyen $\bar{\sigma}$ nous pouvons calculer le poids normalisé de chaque chromosome :

$$W = \left(\sum_{i=1}^{N_{\text{Max}} - 1} i \cdot \bar{g}(i) \right) \cdot \frac{T + \sum_{i=1}^{N_{\text{Max}} - 1} \bar{g}(i)}{\sum_{i=1}^{N_{\text{Max}} - 1} \bar{g}(i)} + N_{\text{Max}} \cdot (\bar{g}(N_{\text{Max}}) - T)$$

$$\text{où } T = \frac{\bar{\sigma} \cdot \bar{g}(N_{\text{Max}}) - \sum_{i=1}^{N_{\text{Max}} - 1} \bar{g}(i)}{\bar{\sigma} + 1}$$

II.3. CLASSIFICATION

Nous classons les chromosomes à partir des deux critères surface et poids ainsi normalisés. La méthode employée est une méthode générale que nous allons décrire brièvement.

II.3.1. Affectation des classes.

Considérons un chromosome k .

Pour chaque critère γ , deux classes, I_j et I_{j+1} , lui sont affectées avec les coefficients $\delta_{I_j}^{\gamma}(k)$ et $\delta_{I_{j+1}}^{\gamma}(k)$ d'après la mesure, $V^{\gamma}(k)$, du critère γ pour le chromosome k et selon les valeurs moyennes ordonnées $\bar{V}^{\gamma}(I)$, du critère γ pour les classes I . Les autres classes sont affectées avec des coefficients nuls.

Les coefficients $\delta_{I_j}^{\gamma}(k)$ peuvent être considérées comme des probabilités d'appartenance.

Pour l'ensemble des critères, chaque classe I est alors affectée au chromosome k avec le coefficient :

$$\delta_I(k) = \prod_{\gamma \in \Gamma} \delta_I^{\gamma}(k) \quad \Gamma: \text{ensemble des critères}$$

Ainsi pour le chromosome k ; la classe candidate, si elle existe, sera la classe J_1 telle que :

$$\delta_{J_1}(k) = \text{Max}_{I \in \mathcal{J}} \delta_I(k) \quad \mathcal{J}: \text{ensemble des classes.}$$

La classe suppléante, si elle existe, sera la classe J_2 telle que :

$$\delta_{J_2}(k) = \text{Max}_{I \in \mathcal{J}, J_1} \delta_I(k)$$

Si $\delta_I(k) = 0$ pour toute classe I de \mathcal{J} , cela signifie que les mesures des différents critères ne sont pas compatibles; dans ce cas le chromosome ne peut pas être classé.

III.3.2 Cohérence des affectations.

Lors de l'affectation des classes, nous pouvons vérifier la compatibilité des mesures correspondant aux différents critères. Néanmoins, il se peut que des composantes soient mal classées. Or nous connaissons l'effectif théorique de chaque classe pour un caryotype normal. La comparaison de cet effectif théorique avec l'effectif obtenu nous permet donc d'une part de contrôler la cohérence des affectations et de modifier éventuellement celles-ci et d'autre part de déceler les anomalies numériques éventuelles du caryotype.

Dans chaque classe J les composantes k sont classées par valeur décroissante du coefficient $\delta_J(k)$. Si une classe comporte plus de composantes qu'elle ne devrait théoriquement en comporter, on cherche à réaffecter la composante k_p dont le coefficient $\delta_J(k_p)$ est le plus faible conformément à la règle suivante :

Une composante est réaffectée si :

- d'une part il existe pour cette composante une classe suppléante qui lui a été affectée.
- d'autre part, cette classe suppléante comporte moins de composantes que l'effectif théorique.

Si la composante k_p ne peut être réaffectée, on cherche à réaffecter la composante k_{p-1} et ainsi tant que la classe J comporte des composantes en trop.

III MISE EN OEUVRE ET FAISABILITE

Pour des raisons d'opportunité et de facilité nous avons mis en oeuvre la plupart des méthodes citées ici, sur le calculateur Burroughs B3500 du département Informatique de l'Institut Universitaire de Technologie de Toulouse, le logiciel d'acquisition des données et de synthèse ayant été réalisé sur le calculateur T 1600 TE.

Afin de permettre une étude de faisabilité plus aisée et plus souple, nous avons réalisé un moniteur conversationnel pour enchaîner les divers traitements et gérer les fichiers de données utilisés. Ce moniteur utilise un terminal alphanumérique Burroughs TD 800 pour le dialogue avec l'opérateur. Certains résultats peuvent être visualisés sur un autre terminal du même type, mais la plupart sont sortis sur imprimante.

Le moniteur a deux rôles essentiels :

- Aiguiller l'opérateur vers les traitements désirés, en présentant ceux-ci.
- Lancer l'exécution des processeurs correspondants, leur transmettre les paramètres fournis par



SYSTEME INTERACTIF DE CARYOTYPIE

J.L. BASILLE - S. CASTAN

l'opérateur, assurer le dialogue avec ce dernier et contrôler la bonne marche du processeur.

Un processeur peut, par l'intermédiaire du moniteur, proposer à l'opérateur de lancer un autre processeur (en général une visualisation de résultats). Quand il rend la main au moniteur, celui-ci revient à l'aiguillage initial (voir figure 2).

La figure 3 montre l'enchaînement des principaux processeurs mis en œuvre avec les fichiers utilisés.

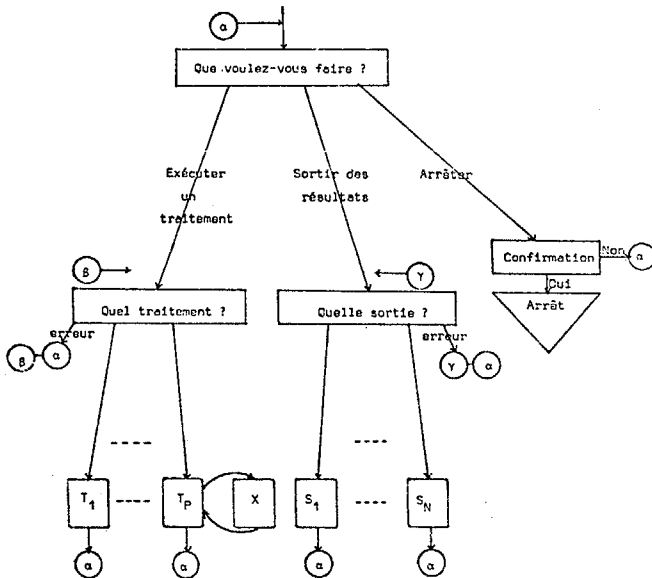


Figure 2

Schéma de principe de fonctionnement du moniteur

VI CONCLUSION

Nous nous sommes attachés à réaliser un système simple, adaptable à un mini-ordinateur, capable d'effectuer de manière la plus automatique possible, sous contrôle permanent d'un opérateur par le biais d'un moniteur interactif, la classification de caryotypes.

Nous avons été amenés à étudier diverses méthodes de traitement des images numériques et à les appliquer au problème de la caryotypie, en particulier pour la localisation du centromère. Toutes ces méthodes donnent des résultats convenables et comparables pour la plupart des chromosomes, en particulier les grands et les médiocentriques s'ils ne sont pas trop déformés. Par contre la localisation du centromère de chromosomes, plus spécialement les acrocentriques et les petits s'avère incertaine quelle que soit la méthode employée. D'autre part, la considération de la seule taille mesurée essentiellement à partir de la surface et / ou du poids après un recadrage global de la dynamique, permet d'obtenir des résultats satisfaisants en employant une méthode de classification qui vérifie la cohérence du classement.

Le résultat de la classification est en général donné de façon résumée en fournissant la formule caryotypique. Au cas où celle-ci montre la présence d'anomalies, l'opérateur peut obtenir une visualisation des chromosomes redressés afin d'examiner ceux-ci plus en détails et détecter ainsi des anomalies de structure telles que la délétion.

La non-prise en compte de l'indice centromérique limite à 7 le nombre de classes significatives, ce qui ne permet pas toujours de préciser les éventuelles anomalies. Il serait donc intéressant de définir un indice plus général et plus souple que l'indice centromérique, fournissant des renseignements sur la forme et l'allure générale du chromosome et pouvant caractériser en particulier les acrocentriques.

BIBLIOGRAPHY

- 1 - CASTELMAN K. R.
Automatic karyotyping with pictorial output.
J.P.L. Février 1974
- 2 - NEURATH P. W., ENSLEIN K.
Human chromosome analysis as computed from arm length measurements.
Cytogenetics 1969. 8. pp. 337-354
- 3 - HUMAN CHROMOSOME STUDY GROUP
A proposed standard system of nomenclature of human mitotic chromosomes.
J. of Heredity. 1960. pp. 214-218
- 4 - GALLUS G.
A decisional model of recognition applied to the chromosome boundaries.
J. Histoch. Cyto. 1974 - 22 - n° 7 pp. 546-553
- 5 - GRANLUND G. H.
The use of distribution functions to describe integrated density profiles of human chromosomes.
J. Theor. Biol. 1973. 40. n°3 -p. 573-589
- 6 - MORTIMER L.
Computer-oriented analysis of human chromosomes.
J. of. histoch. and cytoch. 1974. 66. n°7 pp. 554-560

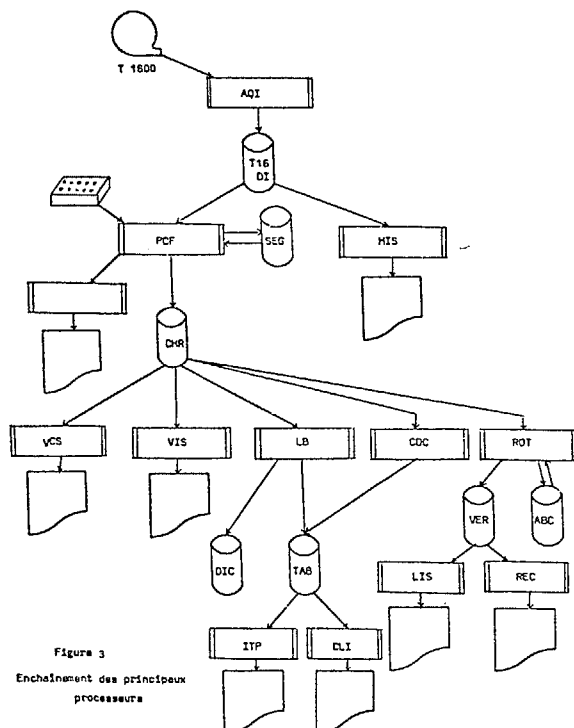


Figure 3

Enchaînement des principaux processeurs

SYSTEME INTERACTIF DE CARYOTYPIE

J.L. BASILLE - S. CASTAN

- 7 - CASTAN S., NABONNE A.
 Un algorithme de squelettisation opérant par tests locaux sur des suites des voisinales successifs.
 Colloque national sur le traitement du signal et ses applications. Nice - Juin 1975
 T 1600
 LATIL : acquisition d'une image
 TRIMHX : } remise en ordre
 TRIMDK : }
- 8 - CASTAN S., NABONNE A.
 Traitement automatique de plans de réseaux.
 Automatismes 1976-21 n°3 pp. 106-111
 COV164 : conversion 16 niveaux - 4 niveaux
 SYNTH : synthèse d'une image.
- 9 - ROSENFELD A.
 Picture processing by computer.
 Academic Press. 1969
- 10 - LELOUP J.
 Traitement interactif d'images. Application à l'amélioration et à la squelettisation d'images binaires.
 D.E.A. Informatique. Juin 1976
- 11 - CASTAN S., LATIL J.Y.
 Capteur d'Images de Toulouse.
 Automatismes 1976 - 21 n°3 pp. 90-93
- 12 - BASILLE J.L.
 Traitement d'images numériques. Application : Système inter-actif de caryotypie.
 Thèse de Science. Toulouse 1976. n° 540
- 13 - BASILLE J.L., CASTAN S.
 NATO Advanced Study Institute
 Digital Image processing and analysis
 1976 June 14 th - 25 th
- 14 - CASTAN S., LELOUP J.
 Analyse d'images. Recherche de paramètres.
 Rapport final Sept 1976 Contrat D.R.M.E.
 n° 75 / 124

ANNEXE

LISTE DES PROCESSEURS DISPONIBLES

B 3500

- AQI : simulation de l'acquisition d'une image
- PDF : séparation des composantes connexes
- JLB : diagrammes de densité - mesures
- CLI : classement
- HIS : histogramme
- ROT : rotation
- CDC : cadrage de la dynamique
- DIS : calcul des paramètres de dispersion
- IOH : moniteur inter-actif
- CRE, HIF, GRI, PSD : utilitaires divers
- IMP : impression de l'image : restitution visuelle
- HEX : impression de l'image en 16 niveaux (hexadécimal)
- VIS : visualisation des composantes d'une image
- VCS : impression en 16 niveaux des composantes
- ITP : sortie du tableau des mesures
- LIS : impression en 16 niveaux des composantes redressées
- IDC : sortie des diagrammes de densité

