

Zeitschrift: Saussurea : journal de la Société botanique de Genève
Herausgeber: Société botanique de Genève
Band: 6 (1975)

Artikel: Taxonomie numérique : chimie taxonomique et ordinateur au service de la systématique
Autor: Mascherpa, Jean-Michel
DOI: <https://doi.org/10.5169/seals-1099060>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 18.04.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Taxonomie numérique: chimie taxonomique et ordinateur au service de la systématique

JEAN-MICHEL MASCHERPA

Résumé

MASCHERPA, J.-M. (1975). Taxonomie numérique: chimie taxonomique et ordinateur au service de la systématique. *Saussurea* 6: 171-185.

La chimie taxonomique et la taxonomie numérique peuvent apporter une aide précieuse aux travaux de systématique classiques. Ces nouvelles techniques se révèlent puissantes et permettent de descendre à un niveau taxonomique que les analyses morphologiques atteignent difficilement. Une revue de ces techniques est effectuée et trois exemples de classification automatique sur des données biochimiques sont proposés.

Abstract

MASCHERPA J.-M. (1975). Numerical taxonomy: chemotaxonomy and computer at systematics service. *Saussurea* 6: 171-185. In French.

Chemical and numerical taxonomy can be of valuable help in classical works in systematics. These new techniques prove to be powerful and can reach a lower taxonomical level than morphological analysis. A review of these techniques is made and three examples of automatic classification based upon biochemical data are proposed.

Introduction

De même que selon un vieil adage "la Nature a horreur du vide" l'esprit humain a toujours eu horreur du désordre. La Science a tenté de mettre de l'ordre dans sa compréhension de l'Univers en essayant de mieux expliquer les phénomènes qui le régissent. Il fallait "donner une place à chaque chose, et mettre chaque chose à sa place". Mais, quand on se trouve devant un nombre d'individus aussi incalculable que celui formant le règne végétal, il faut un esprit aussi vaste que celui de Linné, De Candolle ou Boissier pour appréhender la complexité et la multitude des rapports liant les individus entre eux, et ainsi dégager les divers groupements taxonomiques. D'autre part, si Linné avait pu établir une classification de quelque 10 000 espèces sur des caractères uniquement floraux, on doit maintenant ranger près de 400 000 espèces sur une multitude de caractères. Une systématique d'un seul homme est devenue impossible.

Jusqu'à ces dernières années, bien des travaux ont donné lieu à de remarquables monographies, en essayant de réviser certains points, de reclasser certains groupes ambigus des grandes Flores classiques. Ceci est d'autant plus vrai pour les espèces

dites charnières et conduit souvent à des opinions divergentes selon les auteurs, les observations de type morphologique ou cytologique ne permettant souvent pas de dégager une position systématique précise. Deux techniques relativement récentes peuvent alors apporter une aide aux systématiciens: la chimie taxonomique et la taxonomie numérique (ordinateur).

1. La chimie taxonomique

L'étude des caractères biochimiques des individus que l'on veut classer est une idée déjà ancienne, 1920-1930 avec les travaux de l'école de Mez, dite de Koenigsberg (MIÈGE, 1975). Les chercheurs se sont d'abord intéressés aux produits du métabolisme secondaire (SWAIN, 1966). Mais ceux-ci présentent le désavantage de n'être souvent que des produits finals du métabolisme, donc de ne pas refléter directement le génôme des individus. Depuis quelques années, la recherche de caractères biochimiques constants a conduit les taxonomistes à prendre en considération les produits du métabolisme primaire des plantes, les protéines, surtout si on les extrait d'organes connus pour être stables du point de vue métabolique, les graines. En outre, l'étude de familles protéiques relativement stables que sont les protéines des graines — albumines, globulines, glutélines ou prolamines — conduit à de meilleures interprétations taxonomiques que celles basées sur les spectres protéiques totaux.

Les reproches le plus souvent adressés à cette technique sont de deux ordres:

- leur lourdeur ou la quantité d'efforts et de matériel à mettre en œuvre pour obtenir des résultats valables. Partir sur le terrain avec une centrifugeuse et un spectrophotomètre ne représente peut-être pas le moyen le plus commode de détermination;
- la somme, parfois rebutante, des résultats à analyser et intégrer.

Il paraît évident que cette technique ne peut trouver sa pleine justification que dans l'étude des groupes que les moyens classiques de détermination ne peuvent arriver à démêler, et que si une entière collaboration peut se développer entre les différents taxonomistes. De plus, pour regrouper les données fournies par le laboratoire, l'emploi de l'ordinateur est inévitable. Il permet, par des algorithmes mathématiques, de synthétiser les résultats et de dégager une vue d'ensemble des phénomènes observés.

2. Le traitement informatique des données

Tout travail systématique commence par la récolte des données, qu'elles soient de type morphologique, cytologique, génétique, biochimique ou écologique. Le cher-

cheur se trouve très rapidement en face d'une somme d'observations à analyser et intégrer, ce qu'il peut faire de deux manières: la technique dichotomique, clé de voûte de toutes les Flores classiques existantes, et la technique synthétique.

En dichotomie, on observe un caractère et on sépare les individus selon qu'ils possèdent ou non ce caractère. On arrive ainsi à la formation de groupements définis. Chaque caractère est observé pour lui-même, et non comme faisant partie d'un tout, d'une entité vivante. Avec cette méthode, la hiérarchisation des caractères observés est primordiale, car elle seule permet d'arriver rapidement à la définition des groupements biologiques.

Avec la technique synthétique, l'observateur essaie d'intégrer la totalité des caractères reconnus, de manière à tirer des informations sur les individus et non les caractères. Comme dans la majorité des cas, le nombre des caractères est important, seul l'ordinateur peut synthétiser les données. De nombreuses techniques ont été proposées pour retrouver les groupements naturels biologiques. Nous ne citerons que l'analyse factorielle, l'analyse en composantes principales, l'analyse en composantes multiples, l'analyse canonique, la sériation, etc. Ces techniques se trouvent souvent regroupées dans l'analyse de groupements (SNEATH & SOKAL, 1973). De nombreux auteurs ont défini les critères auxquels devraient répondre toute classification (DAVIS & HEYWOOD, 1963; DEUTSCH, 1966; LANCE & WILLIAMS, 1966, 1967). Retenons les trois principes cités par McNeill (MCNEILL & al., 1969). Toute classification biologique doit être *stable*, c'est-à-dire peu affectée par l'adjonction de nouvelles données, *objective*, des résultats similaires doivent pouvoir être obtenus par des observateurs indépendants travaillant sur les mêmes organismes, et *prévisionnelle*, c'est-à-dire prévoyant l'adjonction de nouveaux taxons.

De toutes ces techniques de classification, la plus utilisée est certainement l'analyse de groupement ou "cluster analysis". Dans son livre remarquable, EVERITT (1974) classe les techniques de groupement elles-mêmes, et décrit: les techniques hiérarchiques, dans lesquelles les individus sont rangés en classes, puis en groupes, pour arriver à définir un arbre de classification ou dendrogramme — les techniques de partition par optimisation, où les groupes sont formés par optimisation d'un critère de groupement, les classes étant mutuellement exclusives — les techniques de densité ou recherche de mode, où les groupes sont formés en cherchant les régions de concentration relativement dense d'individus — etc. En général, ces techniques commencent toutes par l'établissement d'une matrice de similitude entre les individus observés. Ici, la littérature fournit une liste impressionnante de coefficients de similitude: distance euclidienne, coefficient de corrélation, coefficient de distance métrique de Manhattan, coefficient de divergence, coefficient de Pearson, de Jaccard, de Gower, distance généralisée de Mahalanobis, etc. Cette liste n'est naturellement pas exhaustive, tellement les mathématiciens nous en ont apportés. Pour plus d'informations, nous ne saurions que trop conseiller le livre de Sneath & Sokal qui dresse la liste des coefficients de similitude connus, mais aussi en dissèque les avantages et les inconvénients.

2.1. L'analyse de groupement

La plus générale et la plus communément utilisée en taxonomie numérique. Elle débute par l'établissement d'une matrice de similitude entre les individus. A partir de cette matrice, on peut rechercher la paire d'individus présentant la plus

forte similarité, paire qui formera le premier groupe. On recalcule alors la matrice de similitude entre cette paire et les individus isolés restants, la prochaine plus forte similitude donnant la deuxième paire, etc., jusqu'à ce que tous les individus soient groupés. On arrive ainsi à la formation d'un dendrogramme représentant les types d'association des paires, puis des groupes entre eux. C'est une analyse en mode-Q portant sur la comparaison entre les individus, en opposition à l'analyse en mode-R portant sur la comparaison entre les variables. Définissons ici le terme de matrice.

Pour calculer la similitude entre les paires d'individus, on range, conventionnellement, les données originales sous forme d'un tableau, ou matrice, dont les lignes représentent les individus et les colonnes les données originales. Ainsi, si on doit classer N individus sur la base de p variables observées, on construira une matrice des données originales de la forme:

Variables	1	2	3	...	p
Individus					
A	a_1	a_2	a_3	...	a_p
B	b_1	b_2	b_3	...	b_p
.					
.					
.					
N	n_1	n_2	n_3	...	n_p

Dans une matrice de cette forme, on désignera par X_{ij} la valeur de la j ème variable mesurée sur le i ème individu.

On peut alors imaginer de représenter les individus dans un système de coordonnées orthogonales où chaque axe de coordonnée représente une variable. Il est évident que l'esprit humain refuse de concevoir un espace euclidien — espace où les axes de coordonnées sont orthogonaux — à plus de trois dimensions. Mais cette contrainte purement métaphysique est inconnue de l'ordinateur. Dans un tel espace à autant de dimensions que l'on a mesuré de variables, encore appelé hyperspace, on peut facilement visualiser les groupes, ou nuages, formés par les individus à classer. Cependant, pour rendre plus aisée la lecture des différents diagrammes, il est préférable d'essayer de réduire au minimum l'hyperspace original, tout en conservant la plus grande partie de l'information contenue dans les données de départ. Si on considère que cette information peut être matérialisée par la variance totale liée à la matrice des données originales, et que la contraction de l'hyperspace peut arriver à conserver 80 à 85% de la variance totale dans un espace à 5 ou 6 dimensions judicieusement choisies, le gain en compréhension et en temps est considérable. Ceci est rendu possible par l'utilisation d'un algorithme, succession logique de raisonnements et de calculs: la recherche des facteurs principaux, ou axes principaux d'étirement, ou axes d'inertie privilégiée des nuages de points. Cette idée est importante, car elle conduit, dans une certaine mesure, à tenir compte de la hiérarchisation des variables, chère à la méthode dichotomique. Dans l'analyse en facteurs principaux, il faut cependant garder à l'esprit que les nouveaux

axes formés ne représentent pas des variables privilégiées originales, mais les axes principaux d'étirement des nuages de points dans l'hyperespace original.

2.2. L'analyse factorielle des correspondances

Une technique plus nouvelle conduit à des résultats équivalents: l'analyse factorielle des correspondances. Citons ESCOFFIER-CORDIER (1969): "L'application de cette technique permet de dégager une structure des ensembles qui se trouvent en relation, considérée uniquement du point de vue de ses relations avec l'autre ou avec les autres."

Selon BENZECRI (1973), on peut définir l'analyse des correspondances comme la recherche simultanée des axes principaux d'inertie des deux nuages, celui des individus et celui des variables, associés à un tableau de contingence, la métrique euclidienne étant fournie ici par la distance du χ^2 . Ainsi, on peut projeter, sur un même graphe, l'espace réduit des individus et l'espace réduit des variables et déterminer la correspondance entre les deux. Le livre de Benzécri permettra d'avoir une vue complète de cette technique, difficile à résumer en quelques mots.

2.3. L'analyse discriminante

Une fois que le taxonomiste a établi une classification et qu'il en a vérifié la validité, l'analyse discriminante permettra d'ajuster les données au modèle (ROMEDER, 1973). Ici, on se base sur l'existence de groupes définis par une des techniques de classification mentionnées ci-dessus, et on recherche si certaines variables, parmi toutes celles étudiées, ne permettent pas, à elles seules, de retrouver l'appartenance d'un individu anonyme à un des groupes pré-établis. C'est une forme de hiérarchisation des données.

L'utilisation successive de ces trois techniques permet au taxonomiste d'obtenir une bonne vue d'ensemble de son problème, et ce sans introduire de subjectivité dans sa recherche, puisqu'entièrement automatique. A partir des données de base, l'analyse de groupement fournit la classification des individus, puis l'analyse des correspondances dégage les facteurs principaux de détermination et enfin l'analyse discriminante sort, parmi les variables originales, celles qui sont primordiales à observer pour déterminer l'appartenance d'un individu anonyme, pas encore classé, à un des groupes précédemment établis.

3. Quelques exemples de classification automatique

Pour illustrer ces techniques, nous allons présenter trois exemples de classifications de taxons différents obtenues à partir de données biochimiques.

1a R-MODE CLUSTER ANALYSIS ON DISTANCE FUNCTION WITH PRINTED DENDROGRAM.
 BY J.M.PARKS - KANSAS - (VERSION 197) - MODIFIED NOVEMBER 1973.
 CHEMICAL TAXONOMY COMPUTING SERIES - GENEVA

TRAVAUX PRATIQUES - CHIMIE TAXONOMIQUE

R-MODE PRINCIPAL COMPONENTS ANALYSIS ON DISTANCE FUNCTION MATRIX

ORIGINAL DATA INPUT MATRIX

PH.CONTENDER	7.2	10.2	4.2	14.6	17.2	34.8	49.3	65.2	030010001016000011010001010000010110000010100000100
PH.COCO	7.1	10.3	4.4	15.4	16.0	34.9	46.0	54.0	030010001010000001101000010100000101100000101000001000
PH.PROCESSOR	7.0	10.1	4.6	16.4	15.1	35.5	42.5	61.3	000010001010000001110000010100000101000001010000010
RADIS	3.0	13.5	5.6	16.4	5.4	35.6	18.2	8.2	010110010100110000110111100001110000011000000000101010000
BLE	3.0	9.3	2.5	1.5	12.0	6.0	205.0	11.3	00101100101001100100000000100110101010101010101001001
NAVET	3.5	12.2	6.2	25.3	5.7	40.7	13.8	19.3	0100001101101000101000010100001010010100000110110101001000
VIGNA	7.0	13.5	5.5	13.4	12.7	24.3	52.5	94.0	110110110000000010101000011000001100000100011000110
DOLIQUE	6.3	13.4	5.1	17.0	12.3	33.4	36.8	87.6	01000010101001000011010000110100001100000100100011011001001

1b PRINCIPAL FACTOR ANALYSIS OF VARIABLES

FACTOR	1	2	3	4	5
NUMBER OF ITERATIONS EQUALS	22.812	7.998	4.495	3.258	2.811
VARIANCE EXTRACTED EQUALS					
FACTOR	1	2	3	4	5
NUMBER OF ITERATIONS EQUALS	22.812	7.998	4.495	3.258	2.811
VARIANCE EXTRACTED EQUALS					

1c

NORMALIZED FACTOR MEASUREMENTS, UNWEIGHTED

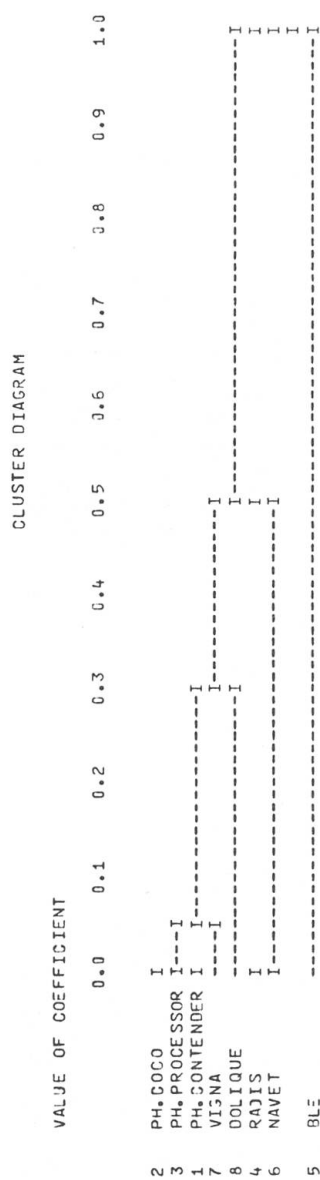
1	1.00000	0.63056	0.35409	0.04939
2	0.99976	0.63144	0.05405	0.08446
3	0.99973	0.63019	0.08215	0.08429
4	0.33047	0.00000	0.17271	0.00635
5	0.00000	1.00000	1.00000	1.00000
6	0.33540	0.04254	0.11442	0.01675
7	0.96601	0.82062	0.02597	0.00000
8	0.44698	0.68211	0.00000	0.05678

1d

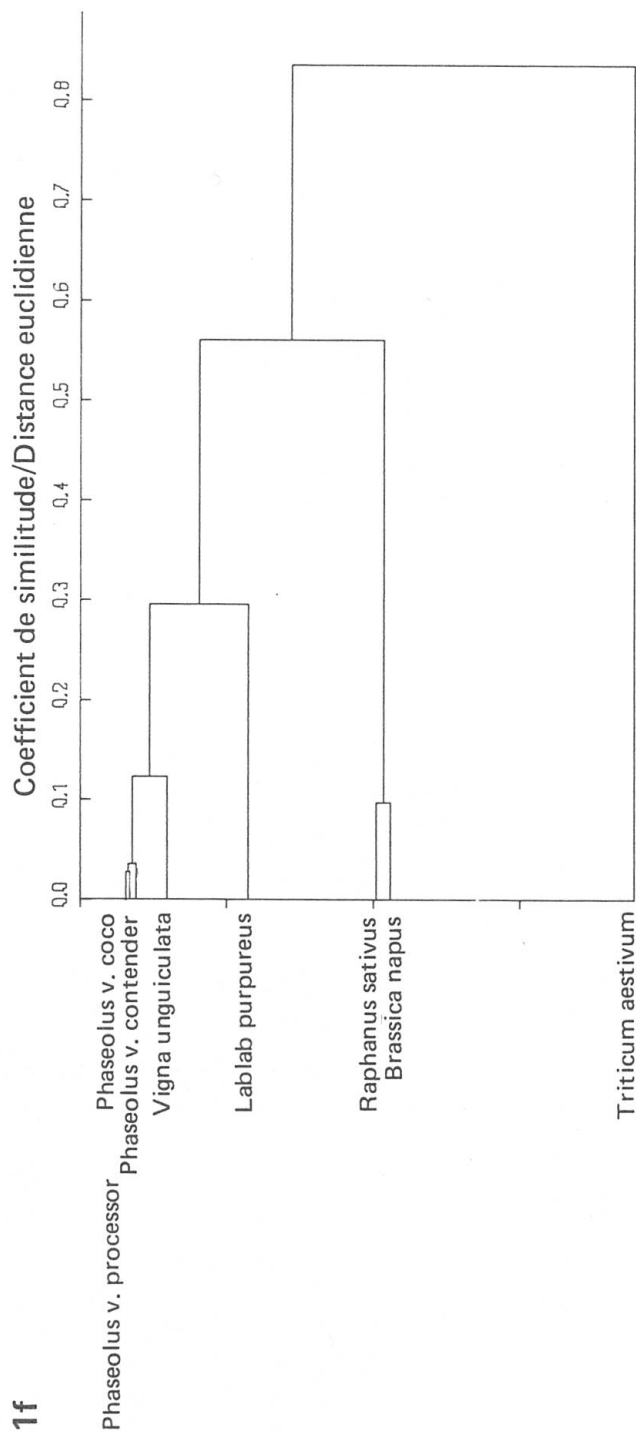
CLUSTER TYPE 6 ANALYSIS

LEFT SAMPLE NUMBER	2	2	4	2	2	2	2
RIGHT SAMPLE NUMBER	3	1	6	7	8	4	5
DISTANCE FUNCTION	0.014063	0.018844	0.048614	0.104762	0.273546	0.422637	0.802477
NO. IN LEFT	1	2	1	3	4	5	7
NO. IN RIGHT	1	1	1	1	1	2	1

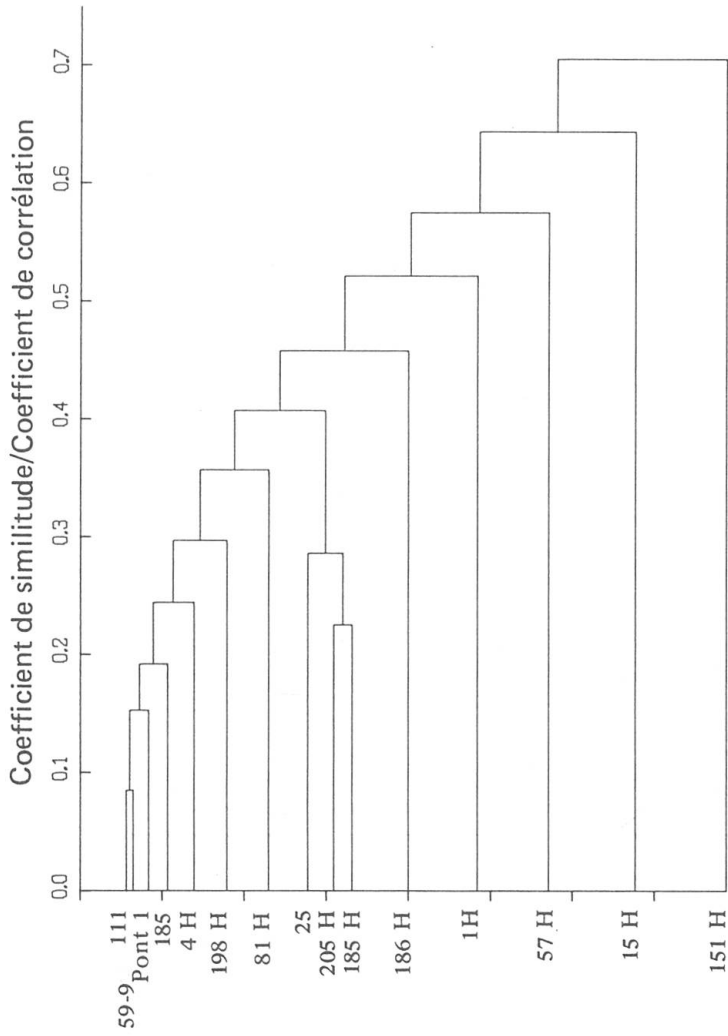
1e



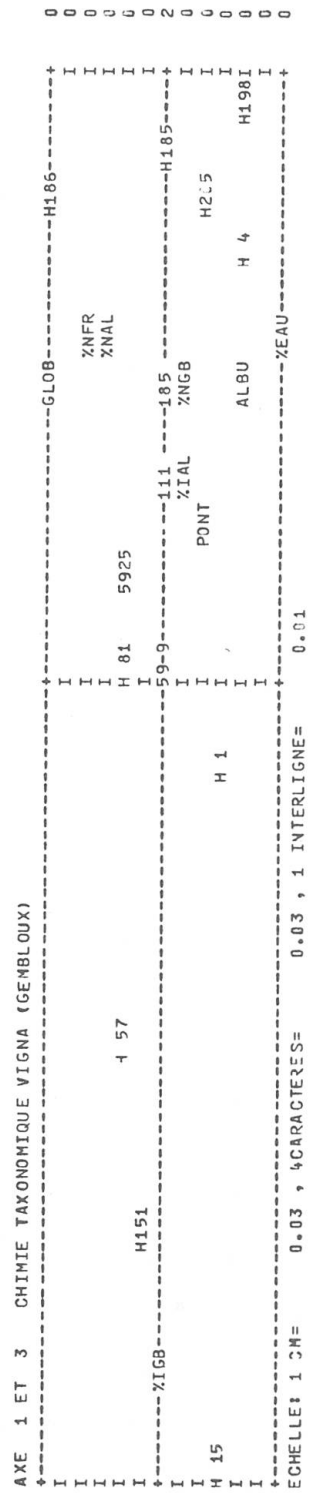
1f



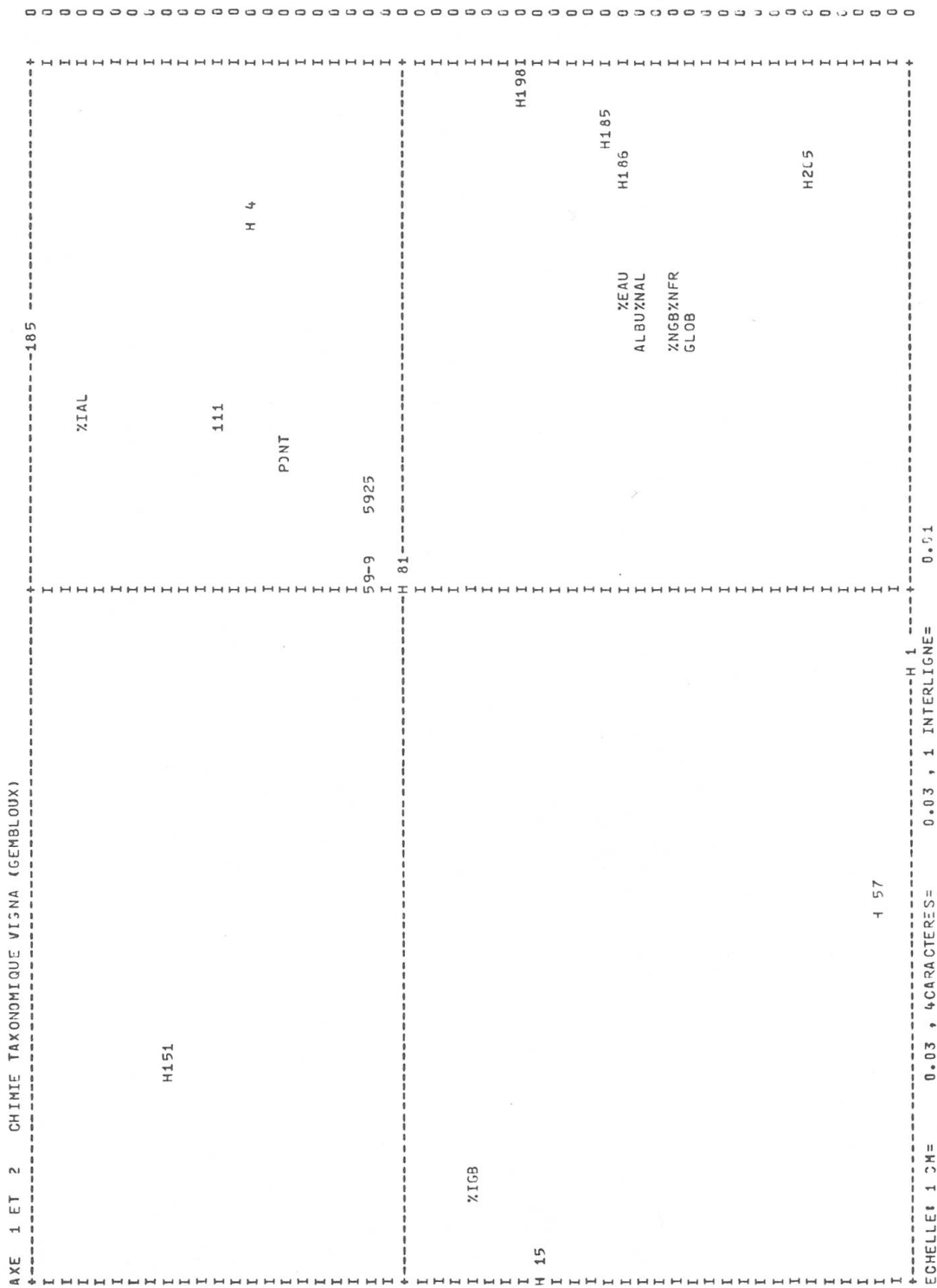
2a



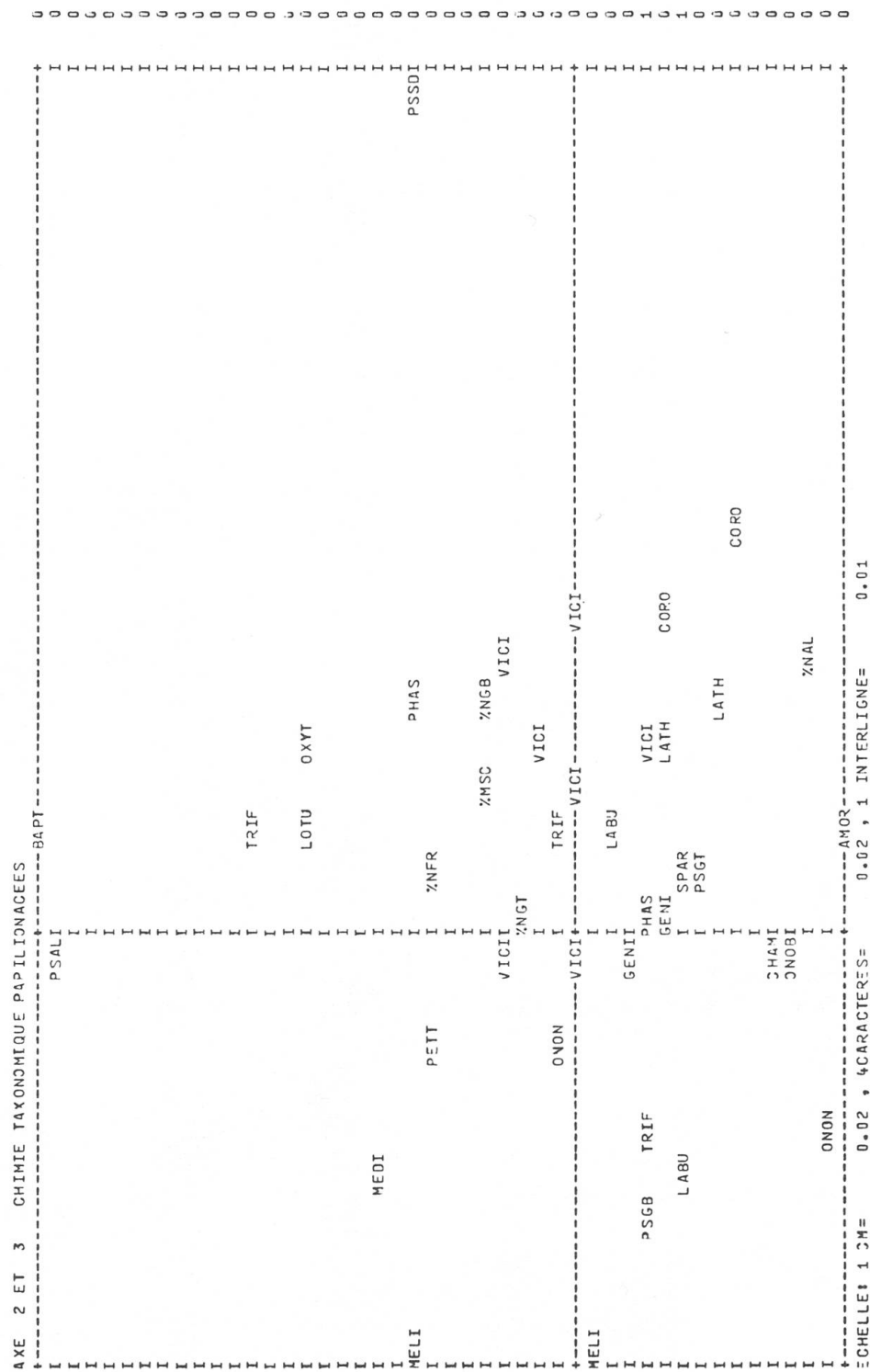
2b



2b



3a



3.1. Classification d'individus appartenant à des variétés, espèces et taxons différents

Cette étude a fait l'objet de séances de séminaires pour étudiants en biologie de l'Université de Genève. Les résultats des analyses biochimiques sont dus à A. Spierer-Royer et A. Grange, assistants au Laboratoire de chimie taxonomique de l'Université de Genève. L'analyse portait sur 3 variétés de *Phaseolus vulgaris*, les variétés Coco, Contender et Processor, un *Vigna unguiculata*, un *Lablab purpureus*, ces 5 individus appartenant aux *Phaseoleae*. Nous avons ajouté deux *Brassicaceae*, *Raphanus sativus* et *Brassica napus*, et une Monocotylédone, *Triticum aestivum*. Sur chaque individu, nous avons mesuré: le pourcentage d'azote contenu dans la farine, la proportion d'azote protéique (azote du lyophilisat de l'extrait salino-soluble dialysé), les rapports respectivement albuminiques et globuliniques à l'azote total de la farine, le rapport de l'azote albuminique à l'azote globulinique. A côté de ces résultats quantitatifs, nous avons ajouté des résultats qualitatifs, à savoir la présence ou l'absence de 50 bandes d'électrophorèse sur gel d'agarose, soit 22 bandes révélées par la technique du Noir amide sur les albumines, 15 sur les globulines et 13 bandes d'estérase appartenant aux albumines. Nous avons aussi mesuré l'activité anti-trypsine des albumines. La figure 1a représente la matrice originale des données.

Nous avons utilisé le programme *Dendrogram* de PARKS (1970) et l'ordinateur CDC 3800 de l'Etat et Université de Genève. *Dendrogram* est basé sur l'algorithme suivant: les variables originales sont en premier lieu standardisées de manière à ce que toutes les mesures soient comprises entre 0.0 et 1.0, pour donner un poids égal à toutes les variables. Pour éviter de travailler sur des variables pouvant être corrélées, le programme effectue une analyse en mode-R, sur les variables et fournit donc un espace orthogonal. Pour cette analyse, on calcule la matrice de similarité entre les variables au moyen de la distance euclidienne suivante:

$$D_{1,2} = 1.0 - \left[\frac{\sum_{i=1}^N (X_{1j} - X_{2j})^2}{N} \right]^{1/2}$$

Puis le programme calcule la matrice de similarité des individus entre eux. Le coefficient de similarité employé est la distance euclidienne en mode-Q selon la formule suivante:

$$D_{1,2} = \left[\frac{\sum_{i=1}^M (X_{i1} - X_{i2})^2}{M} \right]^{1/2}$$

Une analyse de groupement est alors effectuée et les groupements sont résumés en un dendrogramme. Les figures 1 b-d résument les pas principaux de ce programme. Le dendrogramme final est représenté sur la figure 1e. On retrouve les groupements classiques fournis par la taxonomie morphologique. La figure 1b montre que les 41.374% de la variance originale suffisent à expliquer le phénomène, au moyen de seulement 5 facteurs principaux. La figure 1f montre un autre dendrogramme obtenu avec un programme différent, le programme *Dendrograph* de MCCAMMON & WENNINGER (1970), sur les mêmes données originales. En plus de la variabilité entre les groupes, ce programme permet d'avoir une idée de la variabilité à l'intérieur des

groupes. Sur le dendrogramme, dendrogramme en deux dimensions, on voit bien que la variabilité entre les groupes de Légumineuses est plus élevée que celle du groupe des Crucifères.

3.2. Classification de 15 variétés de *Vigna*

Cette étude est tirée du travail de doctorat, actuellement en cours, de SPIERER-ROYER. 11 variétés de *Vigna unguiculata* (les variétés Pont-1, 111, 59-9, 185, 25, H4, H81, H185, H186, H198 et H205), une variété de *Vigna mungo* (H15), une variété de *Vigna radiata* (H57), une variété de *Vigna angularis* (H151) et une variété de *Vigna umbellata* (H1), ont été étudiées sur les caractères suivants: pourcentage d'azote dans la farine, dans les albumines et les globulines, le pourcentage d'humidité de la farine, l'activité anti-trypsine des albumines et des globulines, les taux d'albumines et de globulines extraits (en grammes pour 10 grammes de farine). Avec ces huit variables, nous avons effectué une analyse de groupement, au moyen des programmes *Dendrogram & Dendrograph*, le coefficient de similitude étant un coefficient de corrélation. Sur la figure 2a, on voit que les variétés de l'espèce *Vigna unguiculata* sont nettement séparées des variétés appartenant aux quatre autres espèces.

La figure 2b résume les résultats obtenus par l'analyse factorielle des correspondances. Les deux axes principaux contiennent 99.33% de la variance originale. On remarque que l'axe 1 présente un antagonisme %IGB-%NGB (pourcentage de l'activité anti-trypsine des globulines – pourcentage d'azote contenu dans les globulines) qui discrimine le groupe des *Vigna unguiculata* des autres espèces de *Vigna*. L'axe 2 présente un antagonisme %IAL-%NAL qui discrimine les *Vigna unguiculata* entre-eux. Le genre *Vigna* semble être très sensible aux deux axes généraux Albumines-Globulines. Cette classification est conforme à celle que nous avait proposée R. Maréchal, de la Faculté des sciences agronomiques de l'Etat, Gembloux (Belgique), dont les *Vigna* sont originaires (communication personnelle).

3b

Pas numéro 4

Variables entrées. 4 3 1 5

Pourcentage de bien classés. 96.43

TABLEAU DE CLASSEMENT DE L'ECHANTILLON DE BASE

		GROUPE D'AFFECTION					
		1	2	3	4	5	6
GROUPE ORIGINE	1	9	0	0	0	0	0
	2	0	7	0	0	0	0
	3	0	0	6	0	0	0
	4	0	0	0	2	0	0
	5	0	1*	0	0	1	0
	6	0	0	0	0	0	2

1* Individus mal classés.

3.3. Etude de 33 espèces de Papilionaceae

Cette étude est tirée du travail de doctorat, actuellement en cours, de MISSET. Elle porte sur 33 espèces de Légumineuses. On se réfèrera à l'article de M.-Th. Misset (p. 187-193) pour avoir la matrice originale des données. Dans cette étude, nous avons cependant ajouté deux variables nouvelles, le poids des graines sèches et le poids de matière sèche extraite de la farine. La classification présentée en page 192 a été obtenue au moyen des programmes *Dendrogram* & *Dendrograph*. L'analyse des correspondances, effectuée au moyen du programme Benzécrici, écrit par N. Tabet, est résumée sur la figure 3a. Les projections des deux espaces réduits des individus et des variables sur les trois premiers axes factoriels montrent que les *Vicieae* et les *Phaseoleae* sont très liées au pourcentage de l'azote globulinique, alors que les *Coronilleae* sont liées au poids sec des glutélines. Les 93.78% de la variance originale sont expliqués par les trois premiers axes factoriels, le premier présentant un antagonisme poids sec des glutélines-% d'azote globulinique, le second un antagonisme poids sec des globulines-% d'azote des albumines et le troisième un antagonisme poids sec des albumines-% d'azote des glutélines.

L'analyse discriminante, effectuée au moyen du programme *Mahal-3* écrit par J.-M. Romeder, fournit les mêmes résultats, résumés sur la figure 3b. Sur les 6 variables étudiées (les six premières de l'analyse précédente), quatre seulement suffisent à discriminer les tribus des *Vicieae*, *Genisteae*, *Trifolieae*, *Phaseoleae*, *Ononideae*, *Coronilleae*. Par ordre d'importance, ce sont: le % d'azote globulinique, le poids sec des glutélines, le poids sec des globulines et le % d'azote albuminique. On retrouve les deux premiers axes factoriels de l'analyse précédente.

Conclusion

Ces trois études, bien que très sommaires, démontrent l'apport appréciable que peuvent apporter à la systématique, l'utilisation de techniques nouvelles comme la chimie taxonomique et les techniques de classification automatique. Si le principe de leur emploi et le diagnostic d'utilisation sont bien posés, l'ambiguïté de certains taxons difficiles à classer pourra être résolue. Pour cela, il faut bien entendu que les soupçons qui pèsent sur elle soient levés. Nous pensons notamment aux difficultés de reproductibilité des résultats biochimiques, d'interprétation des diagrammes électrophorétiques, de compilation des données par l'ordinateur. Ce sont les tâches que s'est fixé notre laboratoire.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BENZÉCRI, J. P. (1973). *L'analyse des données*. Vol. I & II. Dunod, Paris.
- DAVIS, P. H. & V. H. HEYWOOD (1963). *Principles of Angiosperm Taxonomy*. Oliver & Boyd, Edinburg.
- DEUTSCH, K. W. (1966). On theories, taxonomies and models as communication codes for organising information. *Behav. Sci.* 11: 1-17.

- ESCOFIER-CORDIER, B. (1969). L'analyse factorielle des correspondances. Thèse, Rennes 1965. *Cahiers du Bureau universitaire des recherches opérationnelles* n° 13.
- EVERITT, B. (1974). *Cluster analysis*. Heinemann educational Books Ltd., London.
- LANCE, G. N. & W. T. WILLIAMS (1966). A generalized sorting strategy for computer classification. *Nature* 212: 218.
- & W. T. WILLIAMS (1967). A general theory of classificatory sorting strategies. 1. Hierarchical systems. *Comp. J.* 9: 373-380.
- MCCAMMON, R. B. & G. WENNINGER (1970). The dendrograph. *Computer Contribution* 48. Merriam D. F., Kansas.
- MCNEILL, J., P. F. PARKER & V. H. HEYWOOD (1969). A taximetric approach to the classification of the spiny-fruited members (tribe Caucalideae) of the flowering-plant family Umbellifereae. *Numerical Taxonomy*. Cole, A. J., Academic Press, London.
- MIÈGE, J. (1975). Protéines des graines en taxonomie et phylogénie. In J. Miège (ed.), *Les protéines des graines*: 305-365. Genève.
- MISSET, M.-TH. (1975). Travail de doctorat de l'Université de Genève (actuellement en cours).
- PARKS, J. M. (1970). Fortran IV program for Q-mode cluster analysis on distance function with printed dendrogram. *Computer contribution* 46. Merriam D. F., Kansas.
- ROMEDER, J.-M. (1973). *Méthodes et programmes d'analyse discriminante*. Dunod, Paris-Bruxelles-Montréal.
- SNEATH, P. H. A. & R. R. SOKAL (1973). *Numerical taxonomy. The principle and practice of numerical classification*. Freeman & Co., San Francisco.
- SPIERER-ROYER, A. (1975). Travail de doctorat de l'Université de Genève (actuellement en cours).
- SWAIN, T. (1966). *Comparative Phytochemistry*. London & New York.

