

L'ANALYSE CANONIQUE APPLIQUÉE A L'ÉTUDE
DE LA SYSTÉMATIQUE ÉVOLUTIVE
CHEZ L'ISOPODE *JAERA (ALBIFRONS)*
ALBIFRONS FORSMAN

par

G. PRUNUS et J. LEFEBVRE

Laboratoire de Zoologie, Université de Caen - 14,
Laboratoire de Génétique Factorielle C.N.R.Z., 78 - Jouy-en-Josas.

Manuscrit reçu le 6 mai 1971.

RÉSUMÉ

L'étude de systématique évolutive de G. PRUNUS a permis de mettre en évidence l'importante variation intra-spécifique de nombreux caractères quantitatifs chez *Jaera (albifrons) albifrons*. Cependant la méthode statistique utilisée ne permettait, ni la comparaison globale de toutes les populations en fonction d'un caractère donné, ni la prise en considération simultanée de tous les caractères mesurés.

Parmi les diverses techniques d'analyses multivariates qui pouvaient permettre de résoudre cette question, l'analyse canonique consiste à rechercher de nouvelles variables telles que les ellipsoïdes de dispersion des différentes populations soient transformées en sphères de rayon unité.

Par programmation sur I.B.M. 1620, cette technique a permis de calculer les trois premières variables canoniques, et ainsi de construire un diagramme tridimensionnel.

Ce diagramme traduit visuellement les distances « spatiales » qui séparent, à un moment donné, les différentes populations en cours d'évolution. La très grande variation de l'espèce se trouve ainsi confirmée et certains problèmes particuliers ont pu être précisés.

SUMMARY

A study in evolutive systematics by G. PRUNUS has pointed out the important intra-specific variation of numerous measurable characters in *Jaera (albifrons)*

albifrons. Yet, the statistical method here used did not make it possible to set up a general comparative scheme of all the studied populations for one character, nor to deal simultaneously with all the biometric features considered.

Among the varied possible techniques of multivariate analysis, the canonical analysis consists in the research of new variables so that the dispersion ellipsoids of the different populations may be transformed into spheres of one unit radius.

Programming according to that technique on I.B.M. 1620, the three first canonical variables were carried out and a tri-dimensionnel diagramm could be drawn.

This diagramm visually shows « spatial distances » which separate, at a given time, the different evolutive populations from one another. The conspicuous variability of the species was thus confirmed and some particular points could be settled with more precision.

Basés essentiellement sur des méthodes morphologiques, génétiques et biométriques, les travaux de systématique évolutive de PRUNUS (1969) ont permis de mettre en évidence un exemple de variation de plusieurs caractères de l'espèce *Jaera (albifrons) albifrons*.

Des divers caractères morphologiques analysés, à savoir des dimensions du corps et des périopodes et la sétosité des lobes sexuels portés par les sixième et septième périopodes des ♂♂, se sont dégagées des différences quantitatives significatives, entre les nombreuses population étudiées, dont la nature génétique est indiscutable.

Les classements des diverses populations analysées, en fonction de chaque caractère quantitatif étudié, se révèlent d'une part, indépendants des facteurs écogéographiques et cytogénétiques de variation auxquels on pouvait songer, d'autre part, indépendants l'un de l'autre, ce qui conduit à supposer l'existence, chez cette espèce, de plusieurs systèmes polygéniques indépendants.

Cependant, la méthode statistique utilisée (REEVE, 1940) n'a permis que des comparaisons deux à deux des croissances allométriques des populations; elle n'a permis ni la comparaison globale de toutes les populations en fonction d'un caractère donné, ni la prise en considération simultanée de tous les caractères mesurés.

Si aucune méthode mathématique actuellement connue ne permet la comparaison multiple de plusieurs coefficients d'allométrie, il existe, par contre, plusieurs techniques d'analyse statistique à plu-

sieurs variables permettant de comparer plusieurs populations d'organismes à partir de plusieurs caractères mesurés.

Il nous a donc semblé souhaitable d'abandonner, pour l'instant, l'aspect allométrique du problème et d'essayer d'établir des « distances » séparant les diverses populations analysées par l'application d'une de ces nombreuses méthodes de comparaisons multiples de moyennes (SOKAL, 1961).

Ces distances, véritables résultantes des différentes variations quantitatives, doivent être en relation directe avec l'état évolutif divergent de l'espèce.

I. TECHNIQUES

Les calculs statistiques à plusieurs variables, dits multivariates ou multidimensionnels, reposent sur des données théoriques (à base de calculs matriciels), datant parfois de plusieurs décennies (H. HOTTELING, 1935, 1936; M.S. BARTLETT, 1938, 1947, C.R. RAO, 1952; S. WRIGHT, 1954), qui n'ont reçu que peu d'applications jusqu'à l'avènement des calculateurs rapides et puissants. Actuellement, ce domaine est en pleine évolution, les applications faisant progresser la théorie, car ces analyses multivariates permettent d'examiner une même réalité, trop complexe, sous des points de vue nouveaux et simplifiés.

Ainsi divers problèmes biologiques ont été traités par l'analyse factorielle (G. TEISSIER, 1956; J. MATSAKIS, 1957; H. MORISHIMA et H.I. OKA, 1960), la méthode du D² de MAHALANOBIS (R.E. BLACKITH, 1957; R.E. BLACKITH et F.O. ALBRECHT, 1959; J. LOUIS et J. LEFEBVRE, 1968), l'analyse en composantes principales (R.E. BLACKITH, 1960; P. JOLICŒUR et J.E. MOSIMAN, 1960; R. ROUVIER, 1966) et l'analyse canonique (P. JOLICŒUR, 1959).

Sous le terme d'analyse canonique (canonical analysis) on rencontre dans la littérature deux types de traitements (J. LEFEBVRE, 1967) : d'une part, la recherche, à partir de deux groupes de variables, de deux nouvelles variables synthétiques, déduites des premières

par deux combinaisons linéaires et telles que leur corrélation dite « canonique » soit maximum, d'autre part, un changement d'axes et d'unités qui transforment les ellipsoïdes de dispersion « semblables » de plusieurs populations, de telle sorte que les nouvelles variables « canoniques » aient toutes une variance égale à l'unité, et qu'elles donnent la meilleure discrimination de ces populations.

C'est cette deuxième technique, dont les bases sont maintenant très bien établies (P. JOLICÉUR, 1959; H. SEAL, 1964) que nous avons utilisée.

Cette méthode d'analyse multidimensionnelle a permis à P. JOLICÉUR (1959) de mettre en évidence, à partir de nombreuses mesures craniométriques, des différences entre les populations de loups d'Amérique du Nord en corrélation avec les facteurs géographiques de dispersion. Cependant, si, à notre connaissance cette méthode n'a été appliquée que très rarement, c'est essentiellement par suite des difficultés de programmation.

Ces difficultés ont été résolues par les travaux de J. LEFEBVRE, M. BOITARD et M. WIMITZKY (1969) qui ont établi une programmation APB transposée sur I.B.M. 1620. Certaines modifications ultérieures ont encore amélioré les conditions d'exploitation, et ainsi notre problème a-t-il pu être traité, après quelques difficultés résultant du transfert des données, sur l'ordinateur I.B.M. 1620 de l'I.N.R.A. à Jouy-en-Josas.

Le principe de l'analyse canonique programmée peut être très rapidement résumé ainsi :

— Soit 10 populations de N_1, N_2, \dots, N_{10} individus ayant subi chacun la mesure de 17 variables supposées normales et centrées : X_1, X_2, \dots, X_{10} ; l'origine des populations étudiées et la nature des variables mesurées ont été définies antérieurement (G. PRUNUS, 1969).

— Les calculs préliminaires déterminent les matrices de dispersion P_i (somme des carrés et des produits des écarts).

— Ces dispersions étant supposées homogènes, les 10 matrices sont sommées :

$$P = \sum_{i=1}^{10} P_i,$$

et une matrice moyenne de covariance en est déduite :

$$V = \frac{I}{\sum N_i - 10} \times P$$

— Soit A la matrice de covariance des centres des ellipsoïdes de dispersion des 10 populations, c'est-à-dire des $X_{i_1}, X_{i_2}, \dots, X_{i_{17}}$ ($i = 1, 2, \dots, 10$).

— L'analyse canonique consiste alors à rechercher de nouvelles variables $y = C'x$ telles que $C'VC = 1$, c'est-à-dire telles qu'elles aient une variance égale à 1.

— Les vecteurs c composant la matrice C sont solutions de : $(A - \lambda V) c = 0$.

— L'équation $|A - \lambda V| = 0$ donne n racines $\lambda_1, \lambda_2, \dots, \lambda_n$; à chacune correspond un système linéaire homogène.

— Le plan Y_1, Y_2 défini par les deux premiers vecteurs canoniques correspondant aux deux plus grandes valeurs propres λ_1 , et λ_2 est tel que les points représentatifs des 10 populations y aient la dispersion maximum.

La prise en considération de la troisième variable canonique, Y_3 , sert à distinguer les populations dont les projections sur le plan Y_1, Y_2 sont trop proches; ainsi se trouve réalisé un diagramme tri-dimensionnel où les images ont des formes sphériques.

— Les rayons des sphères de confiance des centres des populations sont égaux à $2 \sqrt{N_j}$ pour le seuil habituel de 5 %.

— Les programmes successifs établis, ont permis à partir du calcul des matrices, d'obtenir les coordonnées des centres des populations et les rayons des sphères de confiance.

II. RÉSULTATS

Nous avons utilisé lors de cette analyse les dix populations dont les données avaient été inscrites sur cartes perforées pour des calculs précédents.

Sur le tableau I, sont indiquées, pour les 10 populations, les valeurs calculées des trois variables canoniques, Y_1 , Y_2 , Y_3 , qui représentent respectivement les coordonnées x , y , z des centres de populations dans l'espace à trois dimensions, et le rayon des sphères de confiance.

TABLEAU I

Valeurs calculées des trois premières variables canoniques propres à chaque population.

Populations	effectif N	1 ^e variable X	2 ^e variable Y	3 ^e variable Z	rayon de la sphère
Oslo	125	2,40	5,91	5,21	0,14
Askö	145	4,92	4,08	3,83	0,13
Wilhelmshaven	148	3,74	5,77	4,14	0,14
Den Helder	137	4,59	3,70	3,09	0,13
Lucs/mer	160	3,28	5,58	2,65	0,11
Erquy	140	1,54	5,24	3,83	0,13
Grève S ^t Michel	139	0,40	5,79	3,19	0,13
Penvins	142	1,75	5,17	2,91	0,13
Biarritz	199	0,65	2,14	4,28	0,12
Newport	205	3,22	5,12	5,27	0,11

Ces trois variables canoniques, Y_1 , Y_2 , Y_3 , représentent respectivement 44,07 %, 26,14 % et 14,99 % de la variance, soit au total 85,20 %; les deux premières variables ont donc une valeur prédominante, c'est pourquoi nous avons construit en premier lieu un diagramme à deux dimensions (Fig. 1) qui permet déjà de distinguer les populations en fonction de leur projection sur ce plan Y_1 Y_2 .

En second lieu, l'utilisation de la troisième variable canonique Y_3 nous a permis de construire un diagramme tridimensionnel (Fig. 2) où chaque population est représentée par une sphère.

Un calcul très simple, en application du théorème de Pythagore, a permis ensuite d'établir les distances spatiales séparant chaque population de ses voisines (Tableau II).

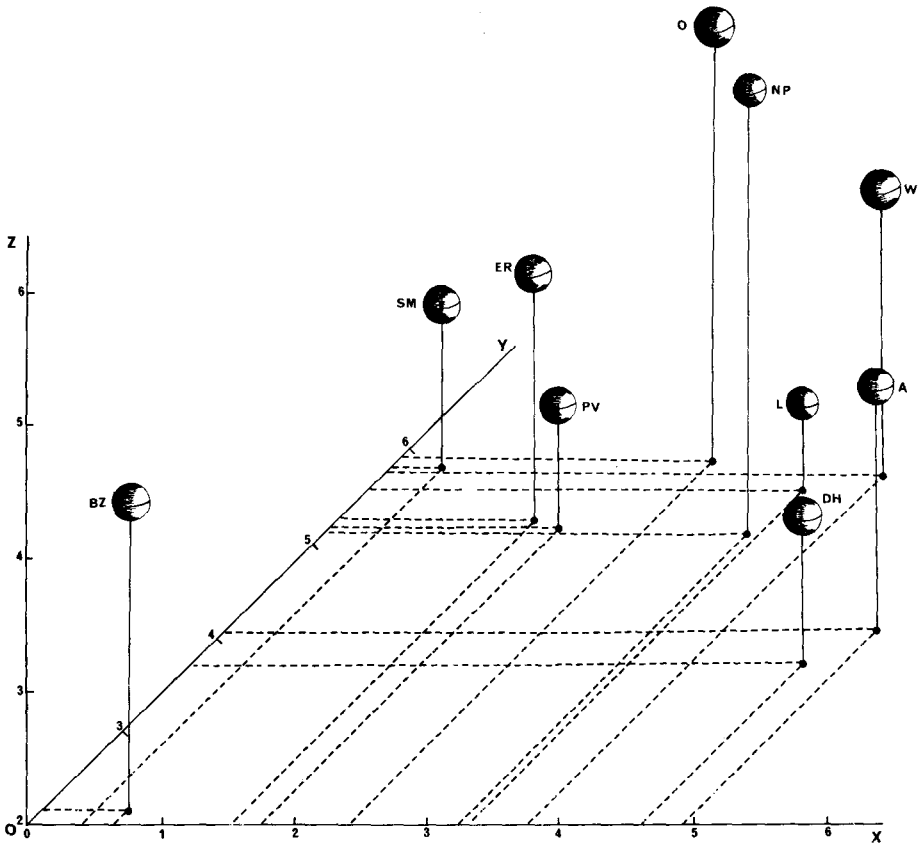


FIGURE 2

Diagramme tri-dimensionnel construit à partir des trois variables canoniques calculées.

III. INTERPRÉTATION

L'observation du diagramme nous permet de constater immédiatement la grande dispersion des populations dans l'espace, qui traduit donc indiscutablement la très grande variation de l'espèce étudiée.

La distance maximale correspondant à un chevauchement de deux populations étant égale à $0,28$ ($0,14 \times 2$), nous constatons que toutes les distances calculées sont nettement supérieures et traduisent donc l'indépendance de chaque population vis-à-vis de ses plus

proches voisines. A cet égard, l'utilisation de la troisième variable canonique est particulièrement intéressante car elle a permis de dissocier les populations d'Erquy et Penvins qui se chevauchaient sur le plan Y_1, Y_2 (Fig. 1).

Par rapport aux comparaisons fragmentaires effectuées précédemment par groupes de variables (G. PRUNUS, 1969) d'une part, aux variations écogéographiques et cytogénétiques qui affectent ces populations (P. LÉCHER, 1967; P. LÉCHER, 1968; G. PRUNUS, 1969; P. LÉCHER et G. PRUNUS, 1970) d'autre part, certains points, résultant de l'observation détaillée du diagramme, méritent d'être discutés :

a) La position très excentrique de la population BZ représente, à notre avis, le type même de phénomène ayant échappé aux analyses fragmentaires, alors qu'il est très compatible avec la position extrême qu'occupe cette population tant au point de vue géographique que cytogénétique (caryotype le plus faible observé : $2n\delta = 18$);

b) la distance assez faible entre Askö et Den Helder confirme le rapprochement très important entre ces deux populations en ce qui concerne le gradient d'allométrie du corps, et surtout le développement de la sétosité des lobes sexuels. Cette confirmation renforce notre opposition à la thèse de HARVEY et NAYLOR (1968), tendant à établir une corrélation entre la croissance du nombre de soies des lobes sexuels et la salinité de l'eau dans laquelle vivent les populations.

c) La position apparemment groupée des trois populations bretonnes (Penvins, Grève Saint-Michel et Erquy) nous ramène au problème le plus difficile à interpréter de cette étude de systématique et génétique évolutives :

Les études biométriques et caryologiques effectuées jusqu'à présent interdisent totalement de considérer l'ensemble des nombreuses populations bretonnes analysées comme une entité taxinomique particulière.

En ce qui concerne la sétosité moyenne, PV, SM, et ER, présentent les moyennes les plus faibles du lot analysé ici, mais d'une part, elles présentent entre elles des différences très significatives et d'autre part, certaines populations scandinaves (ex. Kristineberg) ont aussi peu de soies.

D'un point de vue caryologique, cet ensemble breton est encore plus hétérogène car, outre la répartition entre deux formules chromo-

somiques homogènes ($2n \delta = 22$ et $2n \delta = 20$), on y observe une hétérogénéité intra-population (Roscoff, Grève Saint-Michel, Penvins) qui a été interprétée comme le résultat d'une hybridation introgressive (P. LÉCHER et G. PRUNUS, 1970).

L'analyse présente ne peut donc pas être déterminante sur ce point, mais elle souligne l'étroite parenté entre ces populations compatible avec leur faible isolement géographique relatif.

IV. CONCLUSIONS

La disposition tout à fait anarchique des sphères sur le diagramme nous permet de confirmer que la variation intra-spécifique analysée est du type « random variation » (J.P. SAVAGE, 1963), c'est-à-dire sans relation directe avec certains facteurs écologiques ou géographiques.

Jusqu'à présent, l'aspect divergent du phénomène de micro-évolution que subit cette espèce ne pouvait être traduit clairement; il semble que la technique d'analyse canonique, et plus particulièrement le diagramme tri-dimensionnel qu'elle permet de construire, soit assez bien approprié à la résolution d'un tel problème car, même pour les biologistes non avertis des techniques statistiques à plusieurs variables, se trouvent visualisés, de façon spectaculaire, les distances « spatiales » qui séparent, à un temps donné, les différentes populations d'une même espèce en cours d'évolution.

BIBLIOGRAPHIE

1938. BARTLETT (M.S.). — Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.*, **34**, 33-40.
1947. BARTLETT (M.S.). — Multivariate analysis. *J. Roy. Statist. Soc. B*, **9**, 176-197.
1957. BLACKITH (R.E.). — Polymorphism in some Australian locusts and grasshoppers. *Biometrics*, **13**, 183-196.

SYSTÉMATIQUE ÉVOLUTIVE DE JAERA ALBIFRONS 803

1960. BLACKITH (R.E.). — A synthesis of multivariate techniques to distinguish patterns of growth in grasshoppers. *Biometrics*, **16**, 28-40.
1959. BLACKITH (R.E.), ALBRECHT (F.O.). — Morphometric differences between the eye-stripe polymorphs of the red locust. *Sci. J. R. Coll. Sci.*, **27**, 13-27.
1968. HARVEY (C.E.), NAYLOR (E.). — The taxonomic status of *Jaera albifrons syei* Bocquet (Isopoda Asellota). *Crustaceana*, **14**, 185-193.
1935. HOTELLING (H.). — The most predictable criterion. *J. Educ. Psychol.*, **26**, 139-142.
1936. HOTELLING (H.). — Relations between two sets of variables. *Biometrika*, **23**, 129-149.
1959. JOLICŒUR (P.). — Multivariate geographical variation in the wolf *Canis lupus*. *Evolution*, **13**, 283-299.
1960. JOLICŒUR (P.), MOSIMAN (J.E.). — Size and shape variation in the painted turtle. A principal component analysis. *Growth*, **24**, 339-354.
1968. LÉCHER (P.). — Cytogénétique de l'hybridation expérimentale et naturelle chez l'Isopode *Jaera (albifrons) syei* Bocquet. *Arch. Zool. exp. gén.*, **108**, 633-698.
1970. LÉCHER (P.), PRUNUS (G.). — Caryologie et taxinomie de *Jaera (albifrons) albifrons* (Crustacé Isopode). Populations des côtes bretonnes. (en cours de publication aux Arch. Zool. exp. gén.).
1967. LEFEBVRE (J.). — L'analyse canonique (séminaire de biométrie, Nancy). Editions du C.N.R.Z.
1969. LEFEBVRE (J.), BOITARD (M.), WIMITZKY (M.). — Programmes de calculs statistiques multidimensionnels — STAPB — Editions du C.N.R.Z.
1968. LOUIS (J.), LEFEBVRE (J.). — Etude quantitative de la divergence dans l'évolution morphologique de certaines entités intraspécifiques d'abeilles domestiques. *C. R. Ac. Sc.*, **266**, 1131-1133.
1957. MATSAKIS (J.). — Etude par l'analyse factorielle des dimensions du corps des mâles, des femelles et des indifférenciés d'une population d'*Idotea viridis*, Isopode valvifère. *C. R. Ac. Sc.*, **244**, 1082-1084.
1960. MORISHIMA (H.), OKA (H.I.). — The pattern of intraspecific variation in the genus *Oryza*: its quantitative representation by statistical methods. *Evolution*, **14**, 153-165.
1969. PRUNUS (G.). — Etude de systématique des populations chez l'Isopode *Jaera (albifrons) albifrons* Forsman. *Arch. Zool. exp. gén.*, **109**, 643-702.
1952. RAO (C.R.). — Advanced statistical methods in Biometric Research, New York.
1940. REEVE (E.C.R.). — Relative growth in the snout of anteaters. A study in the application of quantitative methods to systematics. *Proc. Zool. Soc. London, A*, **110**, 47-80.
1966. ROUVIER (R.). — L'analyse en composantes principales : son utilisation en génétique et ses rapports avec l'analyse discriminatoire. *Biometrics*, **22**, 343-357.
1963. SAVAGE (J.M.). — Evolution. Modern Biology series.

1964. SEAL (H.). — Multivariate statistical analysis for biologists. Methuen and Co Ltd, London.
1961. SOKAL (R.R.). — Distance as a measure of taxonomic similarity. *Syst. Zool.*, **10**, 70-79.
1956. TEISSIER (G.). — Analyse factorielle de la variabilité de *Dixippus morosus* aux différents stades de son développement. *XIV^e Int. Cong. Zool. Proc.*, 250-252.
1954. WRIGHT (S.). — The interpretation of multivariate systems in : *Statistics and mathematics in Biology*, Kempthorne, Ed., Iowa.