

SUR LA TAXINOMIE DU GENRE *EQUUS*

[ÉQUIDÉS]

Description et discrimination des espèces
d'après les données crâniométriques

par V. Eisenmann ⁽¹⁾

et J.-C. Turlot ⁽²⁾

Introduction

La paléontologie des animaux appartenant à des espèces relativement récentes (ère quaternaire) tente de placer les fossiles par rapport aux espèces actuelles décrites par les zoologues. Beaucoup de classifications de mammifères actuels reposent essentiellement sur la morphologie externe des animaux ; ainsi, peau, crinière et queue servent couramment de critères distinctifs entre les équidés actuels (Bourdelle, 1955, p. 1064 et suivantes). Le paléontologue, qui ne dispose que d'éléments fossilisés (dents, parties squelettiques), doit donc trouver des correspondances entre l'aspect extérieur de l'animal et son ostéologie et confirmer au niveau du squelette les classifications fondées sur la morphologie externe. Dans la mesure où les crânes sont considérés comme les éléments les plus caractéristiques du squelette, il est important de voir s'il existe des différences crâniométriques entre les espèces actuelles d'équidés et si ces différences débouchent sur une classification voisine de celles que proposent les zoologues (1.1.1).

Lorsque les nombres de variables et de spécimens sont élevés, il devient difficile de s'en faire une idée synthétique sans faire appel aux analyses statistiques multidimensionnelles. On sait depuis les recherches de L. Bellier (cf l'Analyse des Données, TI C n° 5, 6, 7) bientôt suivies par celles de J. Mahé (TI C n° 8) que l'analyse des correspondances réalise bien la synthèse des mesures crâniométriques. Dans la présente étude, entreprise sur les données recueillies par l'un d'entre nous (V.E.), c'est par cette méthode qu'on a d'abord traité l'ensemble des données (§ 2). A des tâtonnements et variantes près (§ 2.0 et aussi § 2.2 et § 5), l'étude repose sur un même tableau comprenant 25 variables et 349 crânes. D'emblée (§ 2.4) est apparue une bonne séparation des différents groupes de crânes. On a cherché à compléter cette vue globale en résolvant au niveau de chaque individu (crâne) le problème de son affectation à une classe (espèce) : d'une part (§ 3) avec des discriminations qui utilisent une information *a priori* (connaissance de l'espèce de chaque crâne) ; d'autre part (§ 4) avec la classification hiérarchique qui comme l'analyse des correspondances procède de façon purement inductive (en ne traitant que les seules données crâniométriques).

Chemin faisant, cette étude utilise avec succès de nombreuses méthodes d'analyse des données ; pondérations des groupes de variables (§ 2.2 : cf [Pond. Pr.] Cahiers, Vol II, n° 3) ; analyse discriminante (§ 3.1) et discriminante par boules (§ 3.2) ; classification hiérarchique avec calcul des contributions mutuelles (§ 4). Ainsi se trouve illustré l'exposé général placé en tête d'un précédent cahier (cf [Sep. Corr.] Vol II n° 4).

(1) V. Eisenmann, attachée de recherches au C.N.R.S. ; L A 12 du CNRS, Institut de Paléontologie, 8 rue Buffon. Paris.

(2) J.C. Turlot, Docteur 3° cycle, Laboratoire de statistique, Université Pierre et Marie Curie, Paris

(*) Le présent article expose les résultats de la thèse de J.C. Turlot (3° cycle Paris 1977). Les lecteurs intéressés par des résultats plus détaillés peuvent s'adresser au laboratoire de statistique pour se procurer cette thèse.

1 Matériel et données

1.1 Rappel zoologique

1.1.1 Systématique des équidés actuels

On distingue communément chez les équidés actuels 8 espèces sauvages et 2 espèces domestiques. Nous considérons qu'elles appartiennent toutes au même genre *Equus* à l'intérieur duquel nous distinguons 4 grands groupes : caballins (chevaux), hémioniens, asiniens, ânes et zébrins.

La nomenclature des espèces et sous-espèces est détaillée plus bas ; à côté du nom latin des divers taxons figure le sigle qui leur a été attribué dans les analyses et, entre parenthèses, le nombre de spécimens de ce taxon présents dans notre échantillon.

A. Caballins (45)

1. Cheval domestique : *Equus caballus*, C (27)

2. Cheval sauvage : *E. przewalskii*, P (17)

Un hybride a été noté CP.

B. Hémioniens (56)

3. Kiangs : *E. kiang*, K (17)

4. Hémiones : *E. hemionus*, H (39). Le matériel comprend des spécimens provenant de Syrie, HS (2) ; de Perse, HP (13) ; de Turkménie, HT (3) ; d'Inde, HI (10) ; de Mongolie, HM (2) et sans provenance connue, HO (9).

C. Asiniens (38)

5. Anes domestiques : *E. asinus*, AD (19)

6. Anes sauvages : *E. africanus* (18). On distingue les sous-espèces *africanus* AA (10) et *somaliensis*, AS (8). Un hybride domestique sauvage a été noté DA.

D. Zébrins (210)

7. Zèbres vrais : *E. zebra*, Z (34) avec deux sous-espèces qui n'ont pas été distinguées dans ce travail, le zèbre de montagne ou de Linné, *E. zebra zebra* et le zèbre de Hartmann, *E. zebra hartmannae*.

8. Zèbres de Grévy : *E. grevyi*, G (35)

9. Zèbres de Burchell : *E. burchelli*, B (132) où nous avons distingué six groupes suivant la provenance géographique : Kitengela (localité du Kenya), BK (34) ; autres localités du Kenya, BG (29) ; autres pays situés au Nord du Zambèze, BN (20) ; régions situées au Sud du Zambèze entre environ 15° et 21° de latitude sud, BS (14) ; régions situées encore plus au sud, BB (17) ; pas de provenance précise, BO (18).

10. Vrais couaggas : *E. quagga*, BQ (9)

Les couaggas sont considérés tantôt comme une espèce à part, tantôt comme une sous-espèce du zèbre de Burchell ; le terme de couagga au sens large s'applique à l'ensemble *E. quagga* + *E. burchelli*.

1.1.2 Répartition géographique du genre *Equus* (fig. 1)

La figure 1 montre que les équidés sauvages actuels sont répartis entre l'Asie et l'Afrique.

A. Caballins

Les chevaux de Przewalski (qui portent le nom de l'explorateur qui les découvrit vers la fin du siècle dernier) vivaient aux confins Nord et Ouest du désert de Gobi, dans les hautes steppes à climat continental et à végétation semi-désertique salée (Mohr, 1971, p 64). La dernière observation de membres de cette espèce en liberté semble dater de 1963 (Groves, 1974, p 57).

B. Hémioniens

Les hémioniens avaient une aire de répartition qui s'étendait de l'Asie Centrale au proche Orient. Les kiangs (*E. kiang*) vivent sur le plateau tibétain à des altitudes très élevées (plus de 4000 mètres). Les hémiones (*E. hemionus*) vivent aussi dans des conditions semi-désertiques ou désertiques mais non en altitude (Groves, 1974, p 102). Les sous-espèces *hemionus kulan*, *khur* et *onager* habitent respectivement la Mongolie, le Turkménie, l'Inde et l'Iran tandis que la sous-espèce *hemippus*, éteinte au siècle dernier, habitait la Syrie.

C. Asiniens

Les ânes sauvages, *E. africanus* sont localisés aujourd'hui au Nord-Est de l'Afrique : ils vivent dans un paysage plat, aride et pierreux (Groves, 1974, p 110). Les zoologues reconnaissent deux sous-espèces dont une seule subsiste à coup sûr :

- l'âne de Somalie habite la Somalie septentrionale, l'Erythrée, l'Ogaden et le pays Danaquil.

- l'âne de Nubie, peut-être éteint, vivait le long de la mer Rouge, au Nord du territoire précédent. Son aire de répartition s'étendait autrefois au-delà du Nil, jusqu'aux massifs du Tibesti et du Hoggar.

D. Zébrins

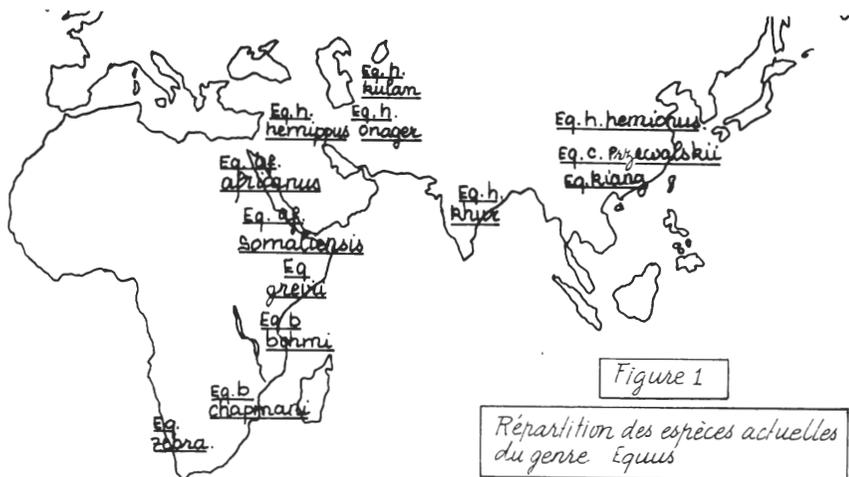
Les quatre espèces de zébrins sont africaines (Groves, 1974).

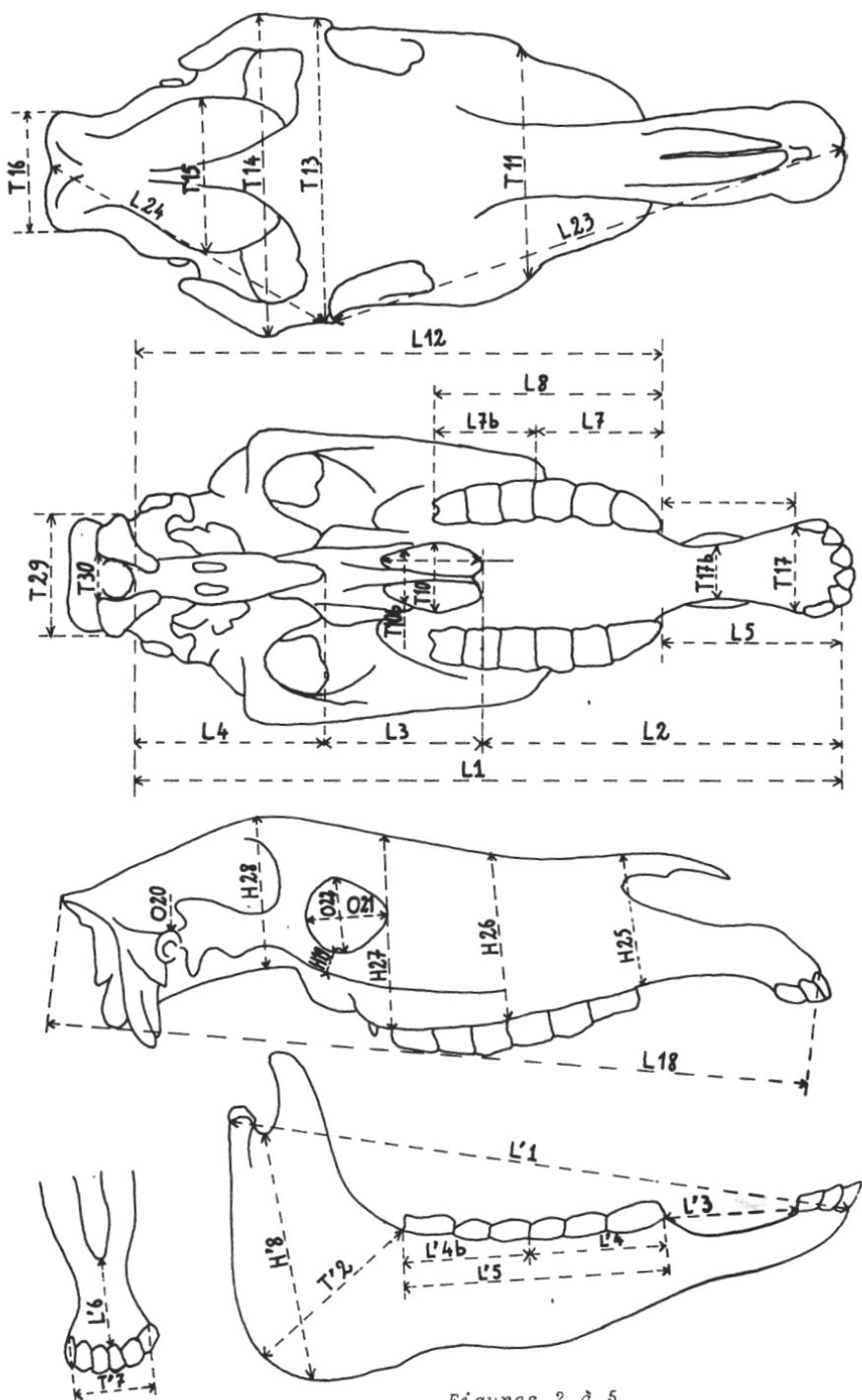
- *Equus grevyi* habite l'Ethiopie et le Nord du Kenya.

- *Equus zebra* occupe les régions montagneuses du Sud de l'Afrique. La sous-espèce *zebra* est à présent confinée dans les réserves du Cap tandis que la sous-espèce *hartmannae* habite le Sud-Ouest africain et une partie de l'Angola.

- Entre ces deux pôles de l'Afrique se place le vaste territoire des zèbres de Burchell, coupé en deux par le cours Est-Ouest du Zambèze. Les formes septentrionales (souvent appelées *E. burchelli boehmi*) vivent en Zambie, dans le Nord du Mozambique, en Tanzanie, au Ruanda, au Burundi, en Ouganda et au Kenya. Les formes méridionales (*E. burchelli chapmani* et *E. burchelli burchelli*) habitent le Sud du Mozambique, la Rhodésie, le Botswana, l'Afrique du Sud et le Sud de l'Angola.

- *Equus quagga*, exterminé au siècle dernier, habitait les plaines de l'Afrique du Sud.





Figures 2 à 5

L longueurs	L1 : longueur basiliaire.	L7b : longueur de la série molaire.
	L2 : longueur palatine.	L8 : longueur de la série jugale.
	L3 : longueur entre P et le milieu du bord postérieur du vomer V.	L9 : longueur des choanes.
	L4 : longueur entre V et le basion B.	L12 : longueur entre B et le milieu de la ligne unissant le bord antérieur des P ² .
	L5 : longueur du museau.	L18 : longueur totale du crâne.
	L6 : longueur du diastème.	L23 : longueur de la ligne oculaire antérieure.
	L7 : longueur de la série pré-molaire.	L24 : longueur de la ligne oculaire postérieure.
T mesures transverses	T10 : largeur maximale des choanes.	T16 : largeur maximale de la protubérance occipitale externe.
	T10b : largeur à l'arrière des choanes.	T17 : largeur du prémaxillaire.
	T11 : largeur de la face entre les points les plus antérieurs des crêtes faciales.	T17b : plus petite largeur du museau.
	T13 : largeur frontale.	T29 : largeur occipitale entre les bords externes des deux condyles occipitaux.
	T14 : largeur bizygomatique.	T30 : largeur du trou occipital.
H hauteurs	H19 : hauteur minimale de la portion infra-orbitaire de l'arcade zygomatique.	H26 : hauteur de la face entre P ⁴ et M ¹ .
	H25 : hauteur de la face en avant de P ² .	H27 : hauteur de la face en arrière de M ³ .
		H28 : hauteur du crâne immédiatement en arrière des orbites.
diamètres	O20 : hauteur du méat auditif externe.	
	O21 : diamètre orbitaire antéro-postérieur.	
	O22 : diamètre perpendiculaire au précédent.	
mandibule	L'1 : longueur de la mandibule.	L'4b : longueur de la série molaire.
	T'2 : largeur maximale de la mandibule.	L'5 : longueur de la série jugale.
	L'3 : longueur du diastème.	L'6 : longueur de la symphyse.
	L'4 : longueur de la série pré-molaire .	T'7 : largeur aux incisives.
		H'8 : hauteur de la mandibule.

1.2 Problèmes posés par le matériel, notations

Les têtes osseuses (crânes et mandibules) ont été étudiées dans divers musées européens et africains. Les principaux problèmes posés concernent la faible représentation de certains taxons, l'état de conservation des spécimens, les difficultés de détermination spécifique (et, ou de la provenance géographique) et les incertitudes concernant le sexe et l'âge des individus.

Dans notre échantillon, les zébrins sont en général mieux représentés que les autres équidés ; cependant les formes méridionales d'*E. burchelli* et les couaggas sont relativement rares. Les asiniens, surtout sauvages, sont très rares. Les diverses sous-espèces d'*E. hemionus* sont inégalement représentées ; les kiangs et les chevaux sauvages sont peu nombreux. La présence d'animaux domestiques, où les variations de taille dues à la sélection sont considérables, introduit une fluctuation, en quelque sorte artificielle.

La plupart des têtes d'hémioniens et de chevaux sauvages appartiennent à des animaux ayant longtemps vécu en parcs zoologiques et présentent diverses anomalies dentaires liées au grand âge et à la vie en captivité. La mauvaise conservation de plusieurs spécimens interdit de prendre certaines mesures ; les mandibules de quelques têtes manquent.

Les données fournies par les catalogues sur le sexe, l'âge, la provenance géographique et même l'appartenance spécifique des têtes osseuses peuvent être incomplètes ou inexactes. De façon générale, le sexe a été établi d'après le développement des canines, parfois, sur la forme du bassin. L'âge a été déterminé d'après la denture, en tenant compte du remplacement des dents lactéales par les définitives et du degré d'usure des incisives. Dans un certain nombre de cas la détermination spécifique proposée par le catalogue a été remplacée par une autre qui paraissait plus exacte ; on verra que les analyses ont souvent résolu les problèmes posés par ces cas "litigieux" et parfois mis en doute d'autres déterminations.

Au total, le matériel ayant servi de base à la plupart des analyses se compose de 349 crânes adultes (cet ensemble sera noté C) dont la répartition par taxon a été donnée dans un § précédent (1.1.1). Pour chaque spécimen on a noté l'espèce, l'origine géographique, le sexe et la classe d'âge (adulte jeune, âge moyen, vieux). Dans les analyses les individus sont repérés grâce à leur sigle suivi par un numéro d'ordre (à l'intérieur du taxon correspondant).

1.3 Variables

Les variables crâniométriques de base sont de 33 sur le crâne et de 9 sur la mandibule. Leur description est donnée par les figures 2 à 5.

Parmi les mesures prises, on note certains recouvrements, ce qui est contraire aux principes de l'analyse des correspondances (cf TI C n° 7, § 6.3 et TII B n° 5 § 2). Malgré les dangers que peuvent présenter des opérations arithmétiques sur les mesures, nous avons tenté de rendre les longueurs additives.

Les segments pris en considération seront : L5, L2-5, L3, L4 ; L2-5 étant obtenue par l'opération $(L2 - L5 + L12 - (L3 + L4)) / 2$.

Les petits segments L5-6 et L3-9 seront placés en variables supplémentaires. La mesure L1 a été prise directement, mais peut être également reconstituée par les opérations $L2 + L3 + L4 = L1$ d'une part et $L12 + L5 = L1$ d'autre part. Ces longueurs, ainsi que L2 et L12 seront placées en éléments supplémentaires. Il peut être intéressant de voir où se situent les segments $LAM = L12 - L8$ et $LMC = L3 + L4 - LAM$. La dernière longueur L18 est prise dans un plan différent et prend en compte la protubérance occipitale externe. La série jugale (L8) n'interviendra pas dans les analyses ($L8 \approx L7 + L7b$).

Nous indiquerons comme suit l'ensemble des mesures :

I_L l'ensemble des mesures longitudinales {L5, L2-5, L3, L4, L9, L18, L7, L7b, L6, L23, L24}; I_T les mesures transverses {T11, T13, T14, T15, T16, T17, T17b, T29, T30, T10, T10b}; I_H les hauteurs {H19, H25, H26, H27, H28}; I_O les orifices {O20, O21, O22}.

$$I = \{I_L\} \cup \{I_T\} \cup \{I_H\} \cup \{I_O\}.$$

2. Description des crânes

Après quelques tâtonnements (§ 2.0) on s'est vite arrêté à un choix satisfaisant des variables descriptives, permettant une bonne séparation des principaux groupes de crânes (§ 2.1). Cette représentation peut encore être améliorée si l'on affecte de pondérations différentes les divers groupes de mesures : longitudinales, transversales et hauteurs (§ 2.2). La considération des mandibules (§2.3) ne semble en revanche rien apporter d'interprétable. Au total l'on retrouve par la seule crâniométrie un panorama du genre *Equus* déjà proposé par un spécialiste, d'après des caractères qualitatifs externes inaccessibles sur le squelette et *a fortiori* sur les fossiles (§ 2.4).

2.0 Analyses préliminaires

Lors d'une première analyse - nous ne l'exposerons pas en détail ici - nous avons remarqué que les dents intervenaient (surtout par leur usure) comme facteur d'âge, ce qui s'est traduit par un allongement du nuage de chacune des espèces le long du premier axe. En plaçant les longueurs des séries molaires et prémolaires en éléments supplémentaires, nous pensons que ce facteur sera bien amoindri et la discrimination meilleure. La trop forte opposition entre L2-5 (molaires-palatin) et L5 (museau) sur les premiers axes de cette même analyse nous laisse penser que la longueur L2-5 n'est pas indifférente à l'usure des dents; nous préférons pour cette raison reprendre la largeur L2 (longueur totale du palais). La longueur totale du crâne L18 - proche du centre de gravité - ainsi que la hauteur de l'arcade zygomatique H19 - trop fluctuante - ont été retirées.

On désignera par I' l'ensemble des variables conservées :

$$I' = (I - \{L6, L7, L7b, L18, L2-5, H19\}) \cup L2$$

2.1 Analyse du tableau $I' \times C$ [25 x 349]

L'analyse principale porte donc sur le tableau [25 x 349] ramené ainsi à 25 variables. La suite des taux d'inertie et valeurs propres est:

$$\tau_1 = 20,8\%; \tau_2 = 20,0\%; \tau_3 = 9,8\%; \tau_4 = 8,4\%; \tau_5 = 5,7\%.$$

$$\lambda_1 = 62 \times 10^{-5}; \lambda_2 = 60 \times 10^{-5}; \lambda_3 = 29 \times 10^{-5}; \lambda_4 = 25 \times 10^{-5}; \lambda_5 = 17 \times 10^{-5}.$$

Par rapport à l'analyse citée au § 2.0, les inerties des deux premiers facteurs se rejoignent tandis que τ_3 diminue.

2.1.1 Le nuage des variables (fig 6 et 7)

De nouvelles variables apparaissent sur le premier axe. Si la longueur du museau y apporte toujours par les variables L2 et L5 une certaine contribution (cette partie du crâne est en effet discriminante), celle-ci ne s'oppose plus aux dents - placées en variables supplémentaires - mais davantage à la largeur frontale T13, aux longueurs L24 et L3-9 (qui signalons-le, semble peu liée à L9). Le rétrécissement de la partie médiane du museau (T17b) en fonction de sa longueur apparaît très nettement. Le second facteur concerne essentiellement les choanes (par L9 et T10, T10b qui s'opposent) et la largeur de la protubérance externe (T16). Le troisième facteur, par une opposition très nette entre les variables L23

et L24, traduit un décalage de l'orbite par rapport au segment L1 - puisqu'aucune longueur (mises à part ces deux-là) ni la largeur frontale T13 ne sont liées à ce facteur. Tout comme les axes 1 et 2, les axes 3 et 4 présentent des pourcentages d'inertie presque égaux, mais deux fois moindres que les premiers ; nous savons qu'ils sont alors peu stables et il est préférable d'examiner le plan (3-4) - stable - sur le nuage des individus.

2.1.2 Le nuage des individus

Considérons le nuage C dans le plan des axes 1 et 2 (Fig 6). Les espèces semblent se répartir sur une couronne décentrée (les zèbres Burchell étant fort nombreux) ; le facteur âge n'apparaît plus, ce qui diminue la dispersion de chacun des nuages. La seule zone de recouvrement partiel d'espèces concerne asiniens et hémioniens. L'axe principal d'inertie sépare bien zèbres de Burchell et de Grévy des hémioniens et asiniens. Les premiers sont caractérisés par un museau long (L5 forte) et effilé dans sa partie médiane (T17b faible), alors que les seconds l'ont plutôt court et plus massif et présentent des dents plus développées que les zèbres de Burchell. La région correspondant au pharynx (L3-9) est également plus grande chez les asiniens et hémioniens que chez les zèbres de Burchell.

Le second axe marque une opposition entre les caballins (Cet P) les hémioniens HetK, et les zèbres de Grévy (on appellera zèbres, les zèbres vrais Z) chez lesquels L3 et les choanes L9 surtout, mais aussi le diamètre du conduit auditif externe (O20) et la largeur de la protubérance externe (T16) prennent des valeurs plus élevées. Les choanes des caballins et hémioniens sont plus larges (T10, T10b) et le museau plus fort au niveau des incisives (T17 et L5-6 sont importantes). Quelques zèbres s'isolent [Z49, Z50, Z51] surtout (probablement des zèbres de Linné), mais aussi [Z52, Z55, Z56, Z57] ; ce sont des adultes de petite taille dont la largeur L3 et les choanes sont petites par rapport aux autres zèbres. Peut-être les deux sous-espèces se distinguent-elles sur ce plan? On remarque également l'isolement des ânes de Somalie (AS).

Le plan des axes 3 et 4 (fig 7) dessine l'opposition entre les asiniens et les hémioniens le long de la première bissectrice ; les caractéristiques opérant cette discrimination sont les suivantes : une face plus haute à l'avant chez les hémioniens (H25 et H26) et à l'opposé un crâne large à l'arrière chez les asiniens (T14, T15, T16). De plus, les asiniens se distinguent des autres espèces par la position de leurs orbites, très en avant. Les ânes sauvages prennent une position plus démarquée, intermédiaire entre les ânes domestiques et les hémioniens. Alors que la plupart des espèces ne présentent qu'une faible variation sur la 2° diagonale, les zèbres de Burchell se répartissent tout le long de l'axe correspondant. Si nous cherchons à distinguer les populations N = {BG, BK, BN} vivant au Nord du Zambèze et groupées sous le nom de zèbres de Böhm des populations S = {BS, BB, BQ} du Sud, nous observons que la plupart des formes S sont d'un même côté de la première bissectrice, alors que presque toutes les formes N (à l'exception de 5 individus sur les 83) sont de l'autre. Les populations méridionales présentent un crâne plus effilé (T14, H27 et H28 faibles) se rapprochant des caballins sur ce plan. Ainsi le couagga (BQ) apparaît beaucoup plus proche des caballins que des formes du Nord {BG, BK, BN}. Il faut remarquer enfin que si Z respecte assez bien le profil des populations méridionales G n'a pas du tout l'allure des espèces septentrionales. En ce qui concerne les zébrins, ce troisième facteur a donc un intérêt taxinomique propre.

2.2 Analyse des correspondances du tableau transformé par pondération

2.2.1 Le choix des pondérations

Les longueurs apportent une contribution prépondérante au plan principal. On peut se demander si, par un système de pondérations tendant à donner davantage d'importance aux longueurs (T) et hauteurs (H), un aspect nouveau de discrimination ne se dégagerait pas. Nous pouvons considérer le tableau des données comme la juxtaposition de quatre tableaux

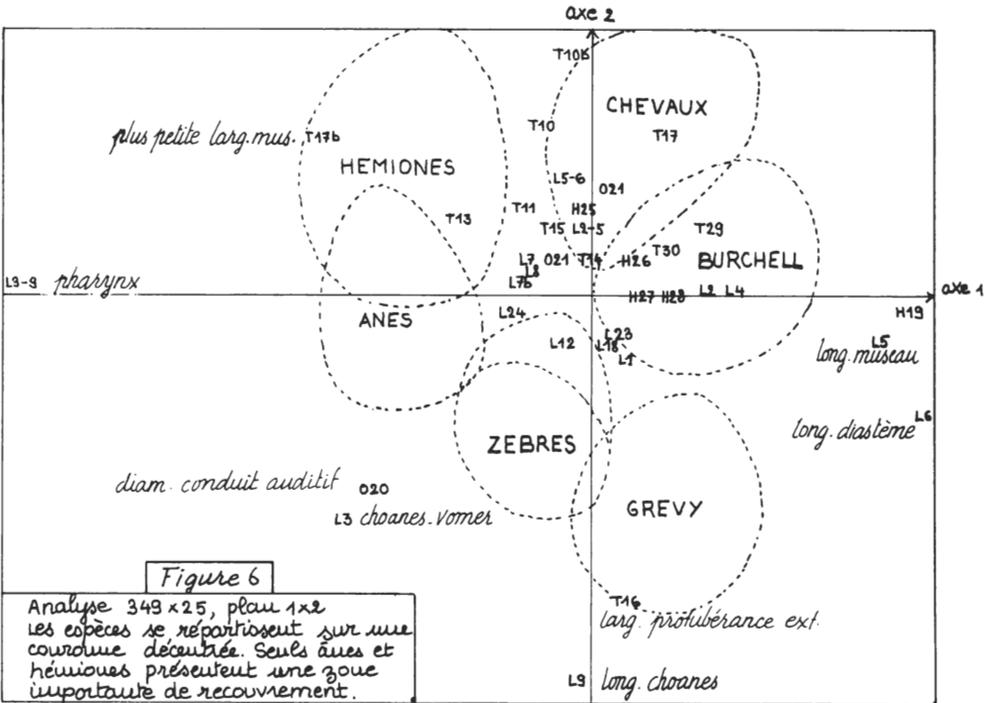


Figure 6

Analyse 349 x 25, plan 1 x 2
 Les espèces se répartissent sur une courbure décentrée. Seuls ânes et hémiions présentent une zone importante de recouvrement.

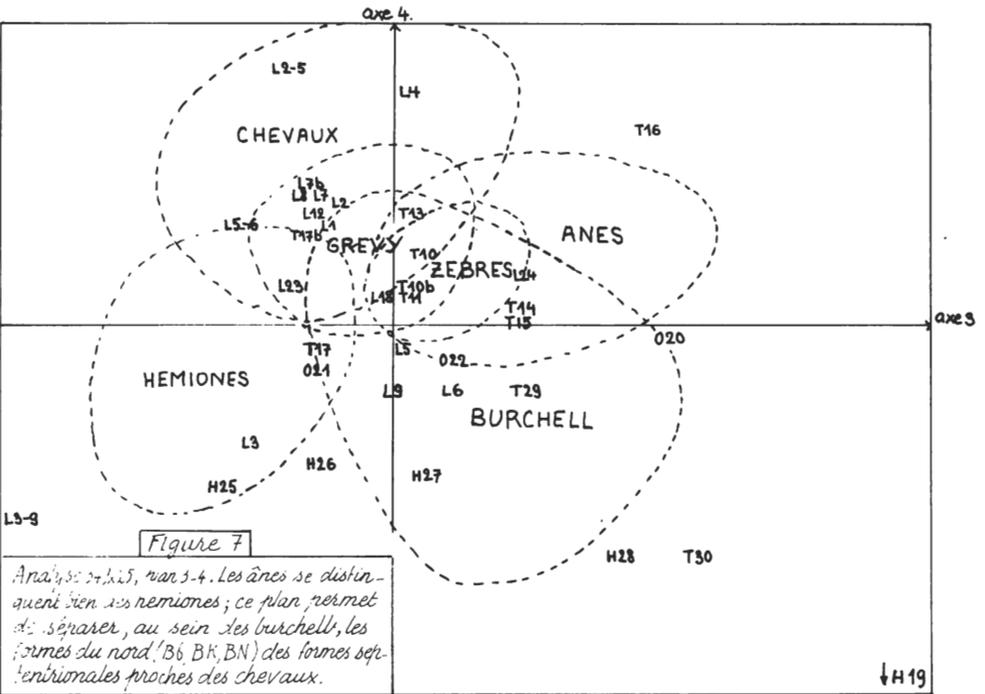


Figure 7

Analyse 34, x 25, plan 3-4. Les ânes se distinguent bien des hémiions; ce plan permet de séparer, au sein des burchells, les formes du nord (B6, BK, BN) des formes septentrionales proches des chevaux.

{C x L}, {C x T}, {C x H}, {C x O}. Plutôt que de vouloir donner à chaque variable une même importance (en les divisant par leur écart-type) parce que cela semble détruire les liens existant entre elles, nous préférons attribuer à chaque classe un coefficient multiplicateur. Ce système de coefficients est choisi de telle sorte que la contribution à l'inertie apportée par chacune des classes soit proportionnelle au nombre de variables qu'elle comprend. L'expression du terme général de la matrice d'inertie du nuage $N(I)$ par rapport au centre de gravité f_j est donnée par :

$$\tau_{jj'} = \sum \{ (f_j^i - f_{j'}) (f_{j'}^i - f_j) f_i \mid i \in I \}$$

posons : $S_L = \sum \{ \tau_{jj} \mid j \in L \}$; $S_T = \sum \{ \tau_{jj} \mid j \in T \}$; $S_H = \sum \{ \tau_{jj} \mid j \in H \}$;

$$S_O = \sum \{ \tau_{jj} \mid j \in O \}$$

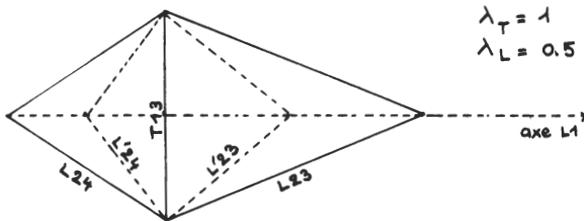
Le problème revient à chercher $\lambda_L, \lambda_T, \lambda_H, \lambda_O$ de telle sorte qu'en remplaçant le tableau initial P_{ij} par le tableau pondéré : $P'_{ij} = \lambda_L P_{ij}$ si $j \in L$, $\lambda_T P_{ij}$ si $j \in T$, $\lambda_H P_{ij}$ si $j \in H$, $\lambda_O P_{ij}$ si $j \in O$.

Les traces de la nouvelle matrice d'inertie τ'_{jj} , vérifient :

$$S'_L / |L| = S'_T / |T| = S'_H / |H| = S'_O / |O|.$$

Ce système de pondérations est obtenu à l'aide d'un algorithme programmé par A.W. Hamrouni (programme pondérations cf *Cahiers* Vol II n° 3 pp 353-359).

Au vu de l'analyse précédente, nous avons supprimé les variables T29 et T30 dont la qualité de représentation est mauvaise sur les premiers facteurs. Remarquons que les données ne sont pas à proprement parler hétérogènes : par notre recodage nous effectuons seulement une transformation géométrique du crâne consistant à multiplier par des coefficients différents les mesures prises dans des directions perpendiculaires L, H, T. Cependant pour les mesures obliques 23 et 24 aucun coefficient multiplicateur ne coïncidera avec la déformation (il faudrait le calculer pour chaque individu, et par le fait même que L23 et L24 font intervenir T13, ces deux longueurs ne nous permettent pas de donner la position exacte de l'orbite par rapport à l'axe L1), ces deux mesures ont donc été supprimées.

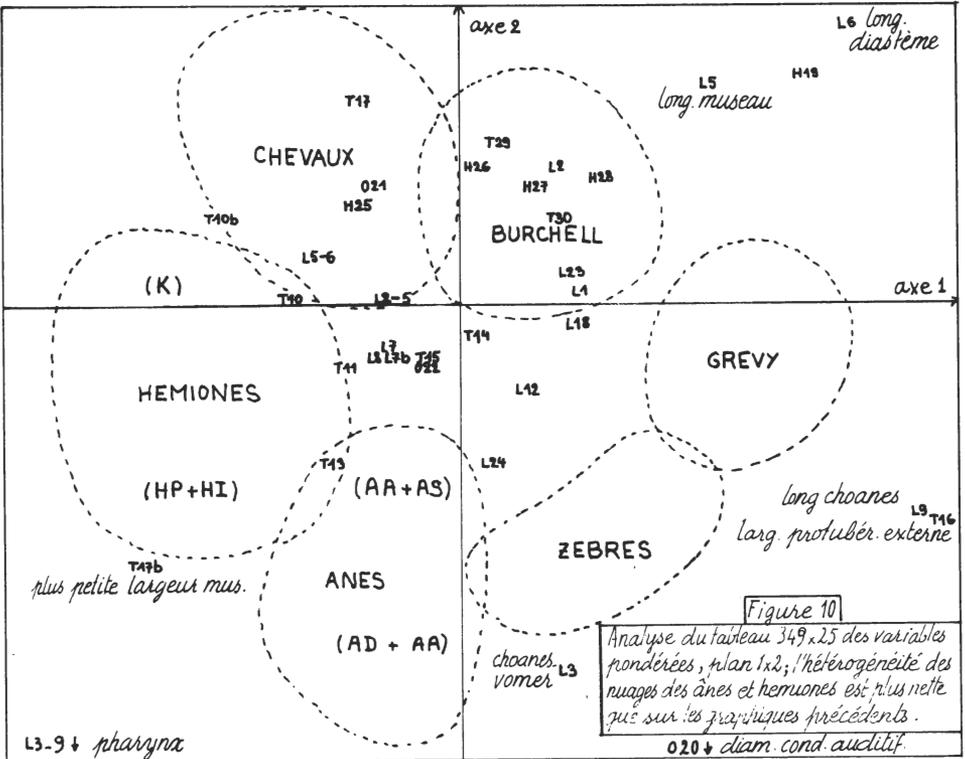
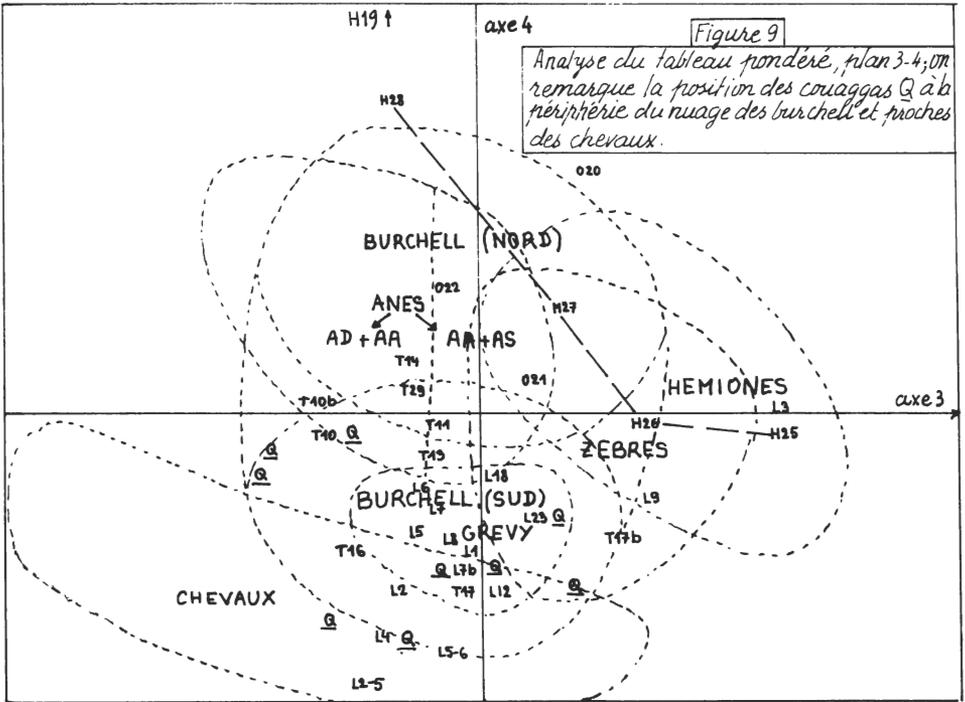


Ainsi nous prendrons en compte dans cette analyse 5 mesures longitudinales, 9 mesures transversales, 4 mesures de hauteurs et 3 diamètres.

Les pondérations obtenues sont respectivement : $\lambda_L = 0,54$, $\lambda_T = 1,15$, $\lambda_H = 1,33$, $\lambda_O = 1,80$. Ainsi les mesures maintenant analysées sont celles de crânes qu'on aurait déformés pour les rendre beaucoup plus courts, légèrement plus hauts, avec des orifices augmentés. Au point de vue géométrique, l'attribution aux orifices d'un facteur 1,8 ne se justifie pas, nous avons imposé $\lambda_O = 1$.

2.2.2 Les résultats de l'analyse (fig 9 et 10)

Les résultats sont peu modifiés ; les deux premiers axes ont été échangés, les longueurs L5 et L2 interviennent de manière plus discrète tandis qu'augmente la contribution du conduit auditif externe (O20).



Par la suppression des variables L23, L24, T29 et T30 (qui n'apportent qu'une faible contribution au premier plan), le nuage des individus s'écarte davantage du centre de gravité. Par l'importance accrue de O20, la séparation du nuage des hémioniens de celui des asiniens est meilleure. L'isolement des ânes de Somalie (AS) par rapport aux AD et AA est plus frappant. Quant aux kiangs, ils prennent une position intermédiaire entre les caballins et les autres hémioniens.

Le plan (3-4) est le même que précédemment à ceci près que les contributions des hauteurs sont plus importantes. Les nuages semblent s'allonger sur le troisième axe (surtout les espèces domestiques C et A). L'examen de ces deux nuages (où la variation de taille chez les adultes est notoire) révèle immédiatement qu'il s'agit du facteur de forme lié à la taille ; facteur qui est de plus discriminant du point de vue taxinomique (fig 8)

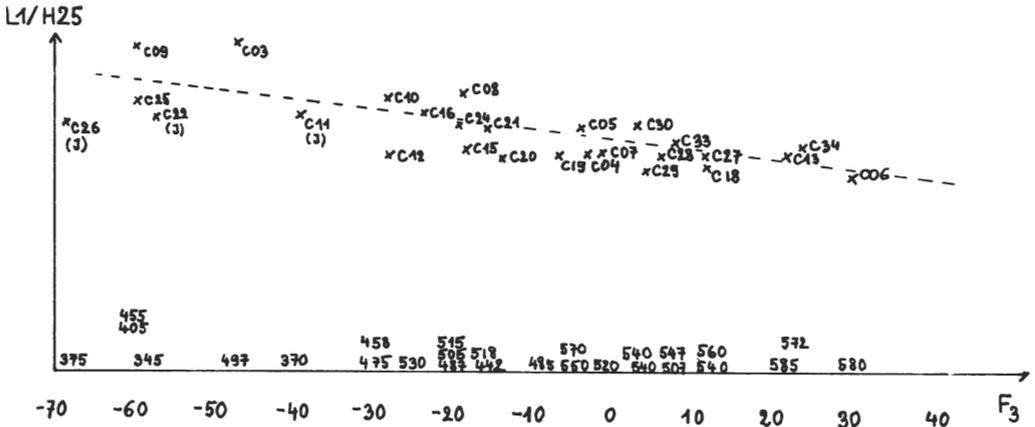


Figure 8 : variation de L1/H25 en fonction de F₃ chez *Eq. caballus* ; le long de l'axe des abscisses (F₃) on a porté les longueurs nasales des individus ; e.g 375 pour C26 ; 455 pour C09 etc... la lettre J signale un sujet jeune.

En abscisse figurent les valeurs que prennent les chevaux sur le troisième facteur (de -70 à +40), leur longueur nasale est également indiquée (variant de 370 à 585) ; en ordonnée nous trouvons le rapport L1/H25.

La figure 8 montre l'acroissement de L1 en fonction de F₃ et, simultanément, la décroissance du rapport (L1/H25) ; seuls les plus jeunes dont le museau n'est pas encore développé (J) semblent échapper à la règle. La mesure 25 est prise verticalement au niveau de la première prémolaire, le museau en se développant entraîne un avancement du nasal par rapport aux dents, d'où l'augmentation de H25.

2.3 Les mandibules

Sur un nombre importants d'équidés (262), les mesures sur la mandibule ont pu être prises (cf fig 4). Une analyse a été effectuée sur 29 variables : les 25 du crâne auxquelles ont été rajoutées 2', 6', 7', 8' prises sur la mâchoire inférieure (3', 4', 4'b, 5' relatives au diastème et aux dents ont été placées en variables supplémentaires). Plusieurs caractères du crâne et de la mandibule sont en rapport : L4' et L7, L4'b et L7b, L5' et L8, L3' et L6, T7' et T17 ; les liaisons correspondantes - très fortes - apparaissent sur les quatre premiers axes factoriels du tableau (262 x 29). Les trois premiers facteurs sont identiques à ceux de la première analyse décrite ; quant au quatrième, les mesures relatives à la mandibule y apportent une forte contribution. Ce facteur

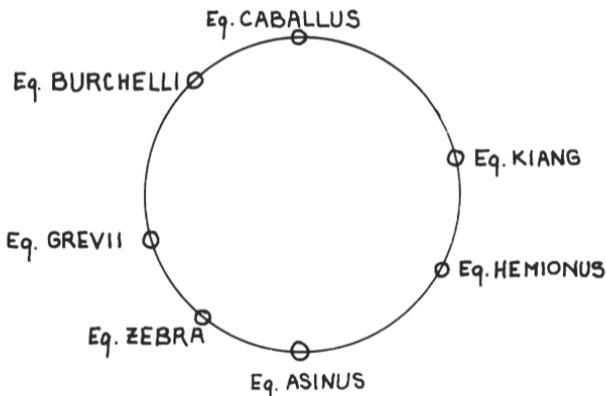
ne permet pas de distinguer les espèces ; une analyse partielle a confirmé cette constatation : les mesures prises sur la mandibule ne nous ont rien apporté pour la discrimination des espèces.

2.4 Conclusions

2.4.1 Représentation d'ensemble du genre Equus

A quelques détails près, toutes les analyses effectuées aboutissent aux mêmes résultats. Sur les 9 ou 10 espèces décrites par les zoologues, les analyses crâniométriques reconnaissent en général 6 espèces ou groupes d'espèces : caballins, hémioniens, asiniens, *E. zebra*, *E. grevyi*, *E. burchelli*. Sur le plan 1-2 des analyses factorielles, ces espèces sont disposées en cercle, dans l'ordre où elles ont été énumérées. La première diagonale sépare bien les zébrins des autres équidés. Le plan 3-4 joue un rôle important dans la discrimination des espèces dont les nuages étaient contigus sur le premier plan : il oppose bien les asiniens aux hémioniens et une partie des zèbres de Burchell aux caballins. Le même plan met aussi en évidence une composante géographique à l'intérieur des zèbres de Burchell en opposant les formes méridionales aux formes septentrionales ; seules les premières sont proches des *E. quagga* et des caballins. Les ânes sauvages, surtout de Somalie, ont un profil intermédiaire entre ceux des ânes domestiques et des hémioniens.

Le premier plan des analyses factorielles présente une grande ressemblance avec une représentation graphique imaginée par Bourdelle (1955, fig 1030) pour rendre compte des affinités morphologiques entre les espèces actuelles d'*Equus*. Dans les deux cas les espèces ou groupes d'espèces sont disposés en cercle (fig 6 et 10) suivant la même disposition. La coïncidence des deux représentations est d'autant plus frappante que nos résultats sont uniquement fondés sur les mensurations des crânes alors que Bourdelle a tenu compte de nombreux caractères anatomiques externes et internes. Il apparaît donc que la morphologie crânienne reflète bien la morphologie générale des espèces actuelles et que le système de mensurations choisi est suffisant pour la décrire.



Répartition des espèces du genre Equus d'après Bourdelle

La stabilité des plans (1-2) et (3-4) a été confirmée par l'analyse des variables pondérées ce qui n'a entraîné pratiquement qu'une modification des contributions, la discrimination se fait plus nette entre ânes et hémiones, la couronne que forment les nuages ressort mieux.

Les analyses des correspondances ont permis de porter un jugement sur le système de mensurations mandibulaires et crâniennes. Les mesures des mandibules ne suffisent pas à discriminer les espèces : lorsque ces mesures sont utilisées conjointement à celles des crânes, elles n'apportent aucune information supplémentaire. Parmi les mesures crâniennes, certaines

se sont montrées peu intéressantes ou même gênantes du fait de leur fluctuation excessive : longueur totale, hauteur de l'arcade zygomatique, longueurs dentaires, largeur des condyles et du trou occipital. En revanche les analyses ont montré ou confirmé l'importance de la région vomérienne et des choanes, du museau, de la protubérance occipitale externe, du méat auditif externe. Nous verrons que l'importance plus grande de certaines variables permet de réduire le nombre total de mesures utilisées en conservant une assez bonne discrimination des espèces.

Dans la suite nous précisons cette représentation globale du genre *Equus* en établissant entre les différentes espèces des cloisons aussi exactes que possible.

Note : Parallèlement à l'analyse du tableau [349 x 25] de base (§ 2.1), nous avons rendu les variables qualitatives par un découpage en 5 classes effectué sur le profil des mesures ; les résultats obtenus à partir de l'analyse des correspondances du nouveau tableau [349 x 125] à 125 colonnes sont tout à fait identiques à ceux décrits au § 2.1. Par là, se trouve confirmée d'une nouvelle façon la stabilité des plans (1-2) et (3-4), mais aussi la légitimité d'un tel découpage en classes des variables.

3. Discrimination des espèces

Nous considérons deux méthodes qui toutes deux se placent dans l'espace des premiers facteurs issus de l'analyse des correspondances : d'abord § 3.1 l'analyse discriminante, i.e. affectation de chaque individu à l'espèce du centre de laquelle il est le plus proche (ce qui revient à prendre pour cloisons les hyperplans médiateurs des segments joignant deux à deux les centres des classes) ; et § 3.2 la discrimination par boules ou affectation de chaque individu d'après les individus dont il est le plus proche) ; enfin au § 3.3 nous revenons sur le problème des mal classés ou limitrophes entre deux espèces.

3.1 Analyse discriminante

Par l'analyse factorielle des correspondances, sans faire usage d'hypothèses *a priori*, nous avons vu que la structure même du tableau des données fournissait une bonne séparation des espèces (à partir des 3 ou 4 premiers facteurs). Tout autre est le problème de la discrimination : une partition étant donnée *a priori*, on cherche sur les données une formule permettant de calculer la classe à laquelle doit être affecté un individu.

Soit I l'ensemble des individus, Q une partition de I (les espèces), l'élément $i \in I$ est affecté à une classe $q(i)$ unique. Chaque élément i de $I \in R_J$ (espace des variables) est muni d'une masse positive μ_i ($\mu_i = \sum \{x_{ij} \mid j \in J\}$), on note $\mu_q = \sum \{\mu_i \mid i \in q\}$ la masse de la classe $q \in Q$ et M la masse totale $M = \sum \{\mu_i \mid i \in I\} = \sum \{\mu_q \mid q \in Q\}$. Le vecteur de description de l'individu i est noté $i_J = \{i_j \mid j \in J\}$. Le centre de gravité de la classe q est $q_J = \frac{1}{\mu_q} \sum \{\mu_i i_J \mid i \in q\}$ et g_J celui du nuage total : $g_J = (1/M) \sum \{\mu_i i_J \mid i \in I\}$.

On définit les trois formes quadratiques d'inertie (avec les notations tensorielles ; cf [Sep. Corr.] § 2 ; Cahier Vol II n° 4, pp 370-372).

$$\sigma(I)_{JJ} = (1/M) \sum \{\mu_i (i_J - g_J) * (i_J - g_J) \mid i \in I\}$$

$$\sigma(Q)_{JJ} = (1/M) \sum \{\mu_q (q_J - g_J) * (q_J - g_J) \mid q \in Q\}$$

$$\sigma(I-Q)_{JJ} = (1/M) \sum \{\mu_i (i_J - q(i)_J) * (i_J - q(i)_J) \mid i \in I\}$$

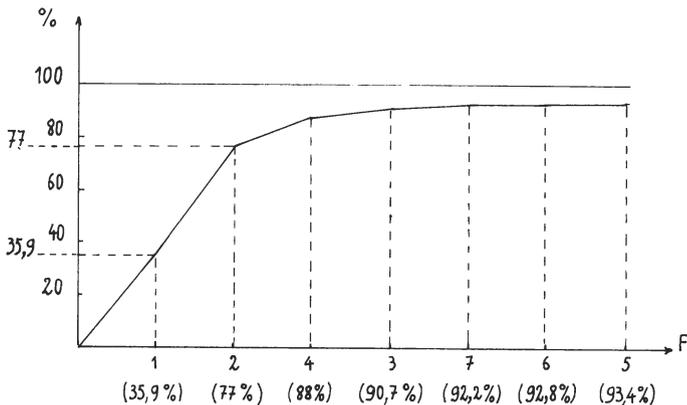
$\sigma(I)$ représente la variance totale du nuage.

$\sigma(Q)$ et $\sigma(I-Q)$ les variances inter et intra-classes.

Nous avons $\sigma(I) = \sigma(Q) + \sigma(I-Q)$, formule de décomposition de la variance de Huygens.

Considérons la métrique sur R^J définie par la forme quadratique $(\sigma(I)_{JJ})^{-1}$. Pour chaque individu i , on calcule sa distance aux différents centres de gravité : $d(i, q) = \|i - q\|_{\sigma(I)_{JJ}^{-1}}$; i sera affecté à la classe du centre à laquelle il est le plus proche. Pour se garder d'une discrimination illusoire, plutôt que de considérer les 33 variables de base, nous préférons utiliser comme variables discriminantes les premiers facteurs obtenus à partir de l'analyse factorielle qui, peu sensibles aux fluctuations d'erreurs, gardent l'essentiel de l'information que l'on possède. Les facteurs sont introduits successivement de manière à obtenir à chaque pas la meilleure discrimination au sens suivant : le rapport de la variance intraclasse à la variance totale $\sigma(I - Q)/\sigma(I)$ doit être minimum. Nous avons projeté le nuage des individus sur le sous-espace de R^J engendré par les deux premiers axes principaux d'inertie du nuage $\{q_J | q \in Q\}$ des centres des classes (axes discriminants principaux). Ce plan maximise le rapport de la variance interclasse à la variance totale du nuage projeté.

Les six classes considérées seront : B, C + P, A, H + K, Z et G (Card Q = 6). Les résultats ont été obtenus à partir des 7 premiers facteurs qui expliquent 73% de l'inertie totale du nuage des individus. Les facteurs introduits successivement sont le 1°, le 2°, le 4°, le 3°, le 7°, le 6° et le 5°. Le pourcentage des individus bien classés (si on les affecte, comme expliqué ci-dessus, à la classe du centre de laquelle ils sont les plus proches) croît de la manière suivante :



En abscisse figurent les facteurs introduits successivement ; en ordonnée, le pourcentage des biens classés (e.g. 90,7% après introduction successive des facteurs 1, 2, 4 et 3).

En se restreignant aux quatre premiers facteurs, la discrimination demeure pratiquement aussi bonne que si on en utilise sept. La matrice de confusion se présente de la manière suivante :

affectation obtenue

	C	A	H	Z	G	B
C	40		3	1		1
A		37	1			
H	4	2	50			
Z		1		32		1
G					35	
B-Q				5		130
Q	4			1		4

classe véritable

On notera en dehors de la diagonale principale 24 individus dont l'affectation est inexacte.

La distance moyenne des couaggas aux différents centres de gravité est $d(\overline{BQ}, \overline{C}) = 7.7$; $d(\overline{BQ}, \overline{A}) = 17$; $d(\overline{BQ}, \overline{H}) = 14,3$; $d(\overline{BQ}, \overline{Z}) = 10.7$, $d(\overline{BQ}, \overline{G}) = 16,4$; $d(\overline{BQ}, \overline{B}) = 8$.

Leur position intermédiaire entre Burchell et caballins que nous avons déjà notée est, par là même, mise en évidence.

La représentation du nuage des individus dans le plan des deux axes principaux de l'analyse factorielle discriminante (i.e. plan engendré par les axes principaux d'inertie du nuage des centres des classes) est assez semblable à celle de l'analyse des correspondances (plan 1-2), à ceci près que Z et G se retrouvent sur un même rayon, mais G plus éloigné du centre de gravité (fig 11).

La règle de discrimination ayant été basée sur l'ensemble des observations, nous voulons connaître ce qu'elle vaut pour séparer les classes potentielles (infinies : tous les caballins, tous les zèbres, etc.). Pour éprouver la stabilité de la méthode de discrimination, on simule son application à des éléments supplémentaires. Nous avons l'ensemble C des crânes, en deux parties d'égal effectif, par espèce, au hasard : C_1 échantillon de base, C_2 échantillon d'épreuve ou test. Seule C_1 intervient dans la règle de discrimination. Nous ne savons pas si les couaggas (BQ) doivent être classés parmi les Burchell ou s'ils constituent une espèce à part ; pour cette raison nous les avons placés en individus anonymes.

Les résultats (programme Mahal 3, Romeder) ont été les suivants :

$$C_1 = C_2$$

Facteurs introduits successivement	1	2	4	3	7	6	5
% bien classés éch. de base	35.4	73.8	88.4	92.2	93.9	94.5	95.1
% bien classés éch. d'épreuve	37.2	79.5	89.4	92.5	95	94.4	94.4

La valeur potentielle de la méthode de classement est ainsi mise en évidence. Nous voyons également confirmé l'apport suffisant des quatre ou cinq premiers facteurs.

Sur les 9 couaggas, 8 ont été classés parmi les caballins et un seulement (BQ06) parmi les zèbres de Burchell, encore est-il éloigné de tous

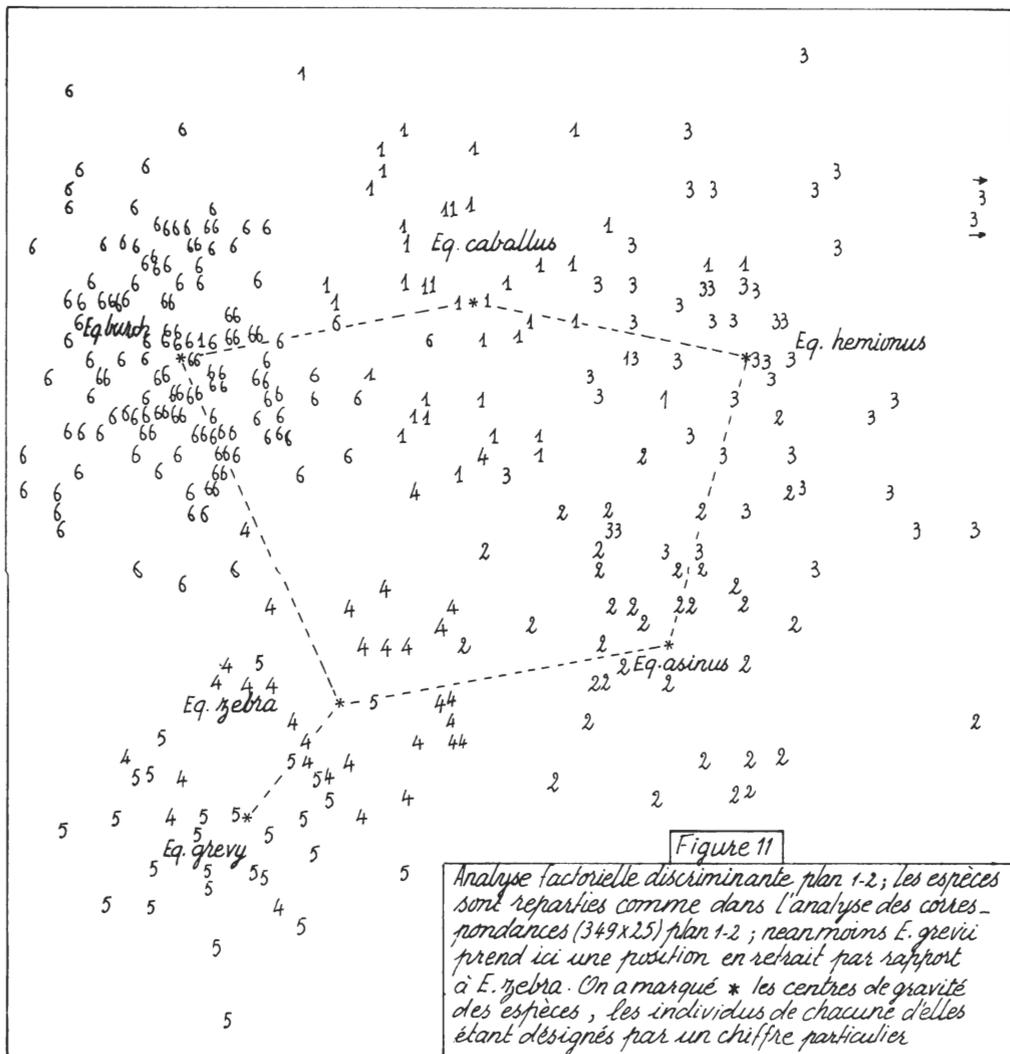


Figure 11

Analyse factorielle discriminante plan 1-2; les espèces sont réparties comme dans l'analyse des correspondances (349x25) plan 1-2; néanmoins *E. grevy* prend ici une position en retrait par rapport à *E. zebra*. On a marqué * les centres de gravité des espèces, les individus de chacune d'elles étant désignés par un chiffre particulier.

les 6 groupes. Bien que le nombre des couaggas soit faible (et il est peu probable qu'on puisse l'augmenter sensiblement), ils semblent bien se distinguer des zèbres de Burchell. L'analyse discriminante tend à créer une rupture entre BQ et les autres formes du Sud (BS, BB) (sur 18 crânes, deux BB seulement ont été mal classés et pris pour des zèbres).

Comme les facteurs de l'A.F.C. sont des facteurs de forme, nous avons cru bon d'y adjoindre la taille (la somme P_1 des mesures considérées dans l'analyse constitue la taille de l'individu) en vue d'améliorer la règle de discrimination des espèces. Les variables introduites successivement par la méthode du pas à pas sont les quatre premiers facteurs des correspondances - 2, 1, 4, 3 dans l'ordre - le 7^o, puis la taille des individus, enfin les nos 5 et 6. Ainsi la taille n'apparaît pas comme un élément essentiel de la distribution des espèces, les pourcentages de bien classés sont sensiblement identiques, $I_1 = 95$, $I_2 = 94,2\%$.

Cependant l'affectation de certains sujets a changé, mais il s'agit d'individus situés à la périphérie de leurs nuages respectifs et donc très sensibles à un faible déplacement du centre de gravité non nécessairement dû à l'introduction de cette nouvelle variable.

3.2 La discrimination par boules (ou par voisinages)

Un crâne peut être situé par rapport aux différentes populations soit en calculant sa distance au centre de gravité de chaque espèce (l'analyse discriminante ci-dessus), soit en examinant ses plus proches voisins. (Pour un programme de calcul dû à M.O. Lebeaux, Cahier Vol II n^o 4, pp 467-481).

En se restreignant à l'espace condensé des 7 premiers axes factoriels, l'expression de la distance entre deux individus est :

$$d_{\chi^2}^2(i, i') = \sum \{ (F_{\alpha}(i) - F_{\alpha}(i'))^2 | \alpha \in [1, 7] \}$$

La répartition en les différentes espèces des p plus proches voisins d'un crâne ainsi que le rayon de la boule qui les contient apportent des informations supplémentaires sur la position de chaque crâne. Le rayon de la boule R est en raison inverse de la densité du nuage au point considéré, tandis que l'examen des voisins fournit en quelque sorte une "probabilité" d'appartenance à chacun des groupes. Ainsi, en choisissant convenablement p , un spécimen pour lequel R est grand doit être considéré comme anormal (ou marginal) - il peut s'agir d'une erreur de mesure ; si au contraire R est faible, mais que le crâne ne correspond pas à l'espèce la mieux représentée parmi ses voisins, on peut pencher vers une erreur de détermination.

Supposons que les espèces aient une même distribution autour du centre de gravité G (i.e. aucune distinction des groupes n'est possible à partir des mesures effectuées), il est clair qu'alors les représentants de chaque classe se trouveront au voisinage de tout point en effectifs proportionnels à ceux des classes respectives. Nous proposons à partir de l'examen du voisinage v_i de chaque individu i la règle de discrimination suivante :

Soit P_l l'effectif de la classe Q_l ($1 \leq l \leq 6$), chaque individu i est affecté d'un poids $p_i = 1/P_l$, si $i \in Q_l$. Ainsi un individu $i_0 \in \mathcal{C}$ sera attribué à la classe l_0 maximisant $\sum \{ 1/P_{Q_l} | i \in v_{i_0} \cap Q_l \}$.

En prenant $p = 10$, 14 individus voient leur détermination remise en cause, dont sept étaient déjà mal classés précédemment.

La matrice de confusion se présente de la manière suivante :

réaffectation

classement a priori

	C	A	H	Z	G	B
C	42		1			2
A		35	1	2		
H	2	2	52			
Z				31		
G					35	
B	1			2		130

En prenant $p = 20$, les résultats obtenus sont très semblables; tous les individus mal classés ci-dessus le restent, quatre individus nouveaux présentent un voisinage problématique. Les couaggas sont classés avec les zèbres de Burchell à l'exception de BQ04 et BQ08 classés avec les cabalins, cependant le voisinage de chacun d'entre eux est essentiellement composé de formes méridionales et l'espèce *E. caballus* y figure systématiquement.

Par la méthode des voisinages, nous pouvons également nous faire une idée plus précise de l'homogénéité des différentes sous-espèces. Les effectifs de celles-ci étant généralement faibles, nous avons pris $p = 10$; nous ne pourrions cependant pas prendre en considération des populations représentées par moins de 10 sujets. Ainsi les crânes provenant de Kitengela présentent une réelle concentration; nous trouvons en moyenne près de 4 crânes de cette population au voisinage de chacun d'entre eux (3.75) alors que pour les autres cette proportion est voisine de 2. De façon plus précise, si les groupes considérés ont une même distribution sur les variables ou sur les facteurs, au voisinage de tout point les fréquences doivent être proportionnelles aux effectifs respectifs, à savoir :

BK = 2,3/10; BG = 2,1/10; BN = 1,5/10; BS = 0,9/10; BB = 1,2/10; BQ = 0,6/10.

Les quantités observées sont :

BK = 3,75/10; BG = 2,4/10; BN = 2,1/10; BS = 0,9/10; BB = 1,9/10; BQ = 1,5/10.

En réalité, ces quantités sont en rapport avec la population totale (d'autres espèces peuvent figurer dans les voisinages) et doivent donc être corrigées. De même, il faudrait considérer les boules sur l'ensemble des facteurs, c'est-à-dire conserver les distances initiales car les observations sur les facteurs ne sont pas indépendantes. Cependant dans les voisinages des zèbres de Böhm (BK, BG, BN) et de Chapman (BS) ne se présentent qu'exceptionnellement des individus d'autres espèces et leur qualité de représentation est excellente dans l'espace des 7 premiers facteurs (nous sommes ainsi proches de l'indépendance). La comparaison des deux fréquences nous fournit donc une bonne évaluation de la concentration de chacune de ces populations. Nous remarquons la grande dispersion des zèbres de Chapman par rapport aux sous-espèces *boehmi*, *burchelli* et à l'espèce *E. quagga*, la concentration des sujets provenant de Kitengela a déjà été signalée. Les kiangs et les ânes de Somalie se distinguent assez bien au sein des hémioniens et des asiniens. Par contre, nous ne voyons pas se dessiner de séparation entre *E. caballus* et *E. przewalskii*, en particulier l'individu hybride noté CP a pour voisins immédiats 5C et 5P.

Remarque 1

Les classements attribués par l'une et l'autre méthode ne doivent pas nécessairement coïncider. En effet, si l'on adopte la métrique du χ^2 par la méthode des boules les facteurs interviennent dans la proximité entre les crânes en proportion de leur variance, alors que par la métrique σ_{II}^{-1} utilisée en analyse discriminante tous les facteurs retenus ont la même importance. (on a respectivement : $d_{\chi^2}^2(i, i') = \Sigma \{ (F_{\alpha}(i) - F_{\alpha}(i'))^2 \mid \alpha = 1, 7 \}$;

$$d_{\sigma_{II}^{-1}}^2(i, i') = \Sigma \{ \lambda_{\alpha}^{-1} (F_{\alpha}(i) - F_{\alpha}(i'))^2 \mid \alpha = 1, 7 \}.$$

Ainsi en analyse discriminante l'influence des facteurs de rang élevé se trouve accrue, notamment celle du facteur géographique (plan 3-4) d'où l'affectation de la plupart des couaggas à l'espèce *E. caballus* par la première méthode. De plus l'analyse discriminante met en place des cloisons planes (hyperplans médiateurs) ; tandis que la méthode des boules permet de suivre, s'il y a lieu les sinuosités d'une frontière.

Remarque 2

Nous avons, toujours dans le cadre de la discrimination des espèces, tenté d'équilibrer par un système de pondération sur les individus le poids de chacun des groupes, mais cette fois-ci avant l'analyse factorielle. Cette opération a été effectuée dans l'espoir d'améliorer la règle d'affectation à l'aide de peu de facteurs. Le poids affecté à l'individu i de la classe j est $P'_{ij} = P_{ij} \cdot e/n_j$ où e désigne une constante, n_j l'effectif de la classe j et P_{ij} le poids usuel de l'individu i .

En prenant $e=140$, le poids des zèbres de Burchell reste identique, celui des zèbres de Grévy et vrais zèbres est multiplié par 4, celui des hémioniens par 2.5 et celui des asiniens et caballins par 3.5 et 4 ; cela revient, en vertu du principe d'équivalence distributionnelle, à remplacer chaque observation d'*E. grevii* ou *E. zebræ* par quatre sujets identiques à celle-ci etc...

Les quatre premiers facteurs sont peu modifiés, cependant que l'inertie du 1^o axe se voit augmentée (25%) et celle du quatrième diminuée (6%). La discrimination se fait légèrement mieux sur le plan (1-2) pour donner sur quatre ou cinq facteurs des résultats tout à fait identiques à ceux obtenus à partir des facteurs de l'analyse du tableau I x C.

Les crânes à détermination douteuse

Dès le premier abord, au moment de la constitution de notre échantillon, 27 têtes osseuses ont posé des problèmes de détermination. Certaines pièces incomplètes n'ont pu être soumises aux analyses et 19 cas litigieux seulement ont été examinés au travers de l'analyse des correspondances et des analyses discriminantes sur les facteurs associés. Dans 15 des cas, la détermination proposée par l'un d'entre nous (V.E.) , en contradiction avec celle du catalogue, s'est trouvée confirmée par la position des crânes sur les premiers plans factoriels et par l'étude de leur voisinage immédiat. Cependant, la discrimination par boules remet en question les déterminations de 11 nouveaux crânes et l'analyse discriminante semble confirmer le mauvais classement de 5 d'entre eux.

En bref, les analyses successives s'accordent à montrer la place marginale occupée par quelques spécimens. Il s'agit en général d'animaux séniles, ayant longtemps vécu en captivité et porteurs d'anomalies diverses, circonstances qui peuvent expliquer leur mauvaise détermination au départ ou leur mauvais classement par les analyses.

De façon générale, il semble que les résultats obtenus par les études de voisinages soient plus intéressants que ceux des analyses discriminantes. Il est en tout cas plus naturel de comparer un spécimen à ceux qui lui sont le plus semblables plutôt qu'à des "individus moyens" fictifs correspondant aux centres de gravité des nuages. Bertrand (19, p102), dans sa critique de "l'homme moyen" de Quetelet l'a déjà noté (cf Cahiers, Vol 1 n^o 2 p102)

4 Classification ascendante hiérarchique

La classification qui procède par agrégation suivant le moment centré d'ordre 2 (cf CAH TI B n° 4 § 2.5.1-2.5.3) a été effectuée dans le sous-espace constitué des 5 premiers axes (expliquant 65% de l'inertie totale du nuage) de la première analyse décrite, suivant le programme [CAH2.FACT] de M. Jambu (cf cahiers Vol 1 n° 1 pp 77-92). La suite des taux afférents aux noeuds les plus hauts est la suivante :

$$\tau_1 = 22\% , \tau_2 = 19\% , \tau_3 = 8,8\% , \tau_4 = 6,6\% , \tau_5 = 2,7\% .$$

Les espèces se reconnaissent bien si l'on considère la partition en six classes obtenues en coupant l'arbre au niveau $\tau = 30$ par exemple (fig 18). Mis à part le rattachement des ânes de Somalie au groupe des hémioniens, on ne distingue pas de sous-groupes homogènes au sein des différentes espèces. Puis on note les trois regroupements suivants : zèbres de Burchell et caballins ($\tau = 8,8$), vrais zèbres et zèbres de Grévy ($\tau = 3,7$), hémioniens et asiniens ($\tau = 6,6$) ; les trois classes sont très distinctes sur l'arbre. Ainsi, cette approche nouvelle confirme que les zèbres de Burchell ressemblent plus au caballins qu'aux autres zébrins (*E. greyvi* et *E. zebra*). Par ailleurs, les hémioniens sont plus proches des asiniens que des caballins ; l'âne de Somalie semble même constituer un intermédiaire entre hémioniens et ânes domestiques.

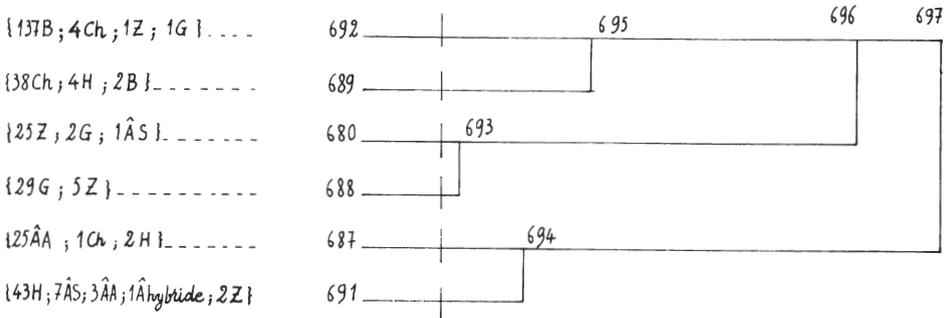


Figure 12 : Arbre de classification. Avec chaque noeud on a marqué son identificateur (N 697 etc...) et le taux d'inertie correspondant. La composition des principales classes est donnée entre accolades ; on a noté \hat{A} = âne ; $\hat{A}A$ = âne de Nubie ; $\hat{A}S$ = âne de Somalie ; B = Burchell ; Ch = cheval etc...

Les contributions relatives afférentes aux noeuds les plus élevés de la hiérarchie sont les suivantes :

	F ₁	F ₂	F ₃	F ₄	F ₅
N697	.86	.10	.2	.0	.0
N696	.14	.78	.0	.4	.2
N695	.0	.32	.11	.54	.0
N694	.11	.01	.73	.13	.0
N693	.0	.39	.16	.1	.42
N692	.52	.27	.11	.8	.0
N691	.01	.48	.34	.02	.12
N690	.55	.00	.10	.33	.0

Tableau des contributions relatives des facteurs aux noeuds.
(Le total de chaque ligne vaut 1 aux erreurs d'arrondi près)

Il est important de remarquer que les premières divisions - celles qui mènent aux espèces - sont le fait d'un facteur unique (N697, N696, N694) ou de deux facteurs (N695, N693). En particulier les deux premiers facteurs fournissent pour les trois groupes principaux une clé de détermination qui peut se résumer de la façon suivante :

1. Museau court et massif, forte largeur frontale, grandes longueurs de la ligne oculaire postérieure et du vomer (L3-9) : asiniens ou hémioniens.
2. Sinon, grande longueur palatin-vomer, choanes longues et étroites, conduit auditif externe et protubérance occipitale externe bien développés : *E. zebra* ou *E. grevyi*.
3. Sinon : *E. burchelli* ou *E. caballus*.

Dans la reconnaissance des sous-groupes, un facteur suffira pour la distinction entre asiniens et hémioniens mais deux facteurs seront nécessaires dans les autres cas. En résumé :

- Face haute en avant : hémioniens ; crâne large en arrière et orbites très antérieures : asiniens.
- Les différences entre *E. zebra* et *E. grevyi* sont plus faibles : conduit auditif externe très développé : *E. zebra* ; museau très allongé et étranglé : *E. grevyi*.
- Indice vomérien de Franck (L4/L3) faible : *E. burchelli* ; museau large et crâne bas : *caballins*.

Certaines autres caractéristiques marquent des tendances mais ne permettent pas de distinctions plus fines.

5 . Réduction des variables

Les analyses multidimensionnelles d'un matériel fossile, en général incomplet, nécessitent la reconstitution des variables qui n'ont pas pu être mesurées. La reconstitution des données manquantes sur un individu se fait par analogie avec un type moyen central et, de ce fait, entraîne l'individu vers le centre de gravité du nuage. Il importe donc de réduire le plus possible le nombre de données à reconstituer ce qui peut être fait si le nombre total de données est lui-même réduit. Il s'agit, pratiquement, de conserver une meilleure discrimination en analysant le plus petit nombre possible de données ; ce problème a été résolu en deux étapes.

Nous avons obtenu une bonne discrimination des espèces à partir des premiers facteurs - dans l'ordre 1, 2, 4, 3 puis 7 - de l'analyse des correspondances (§ III.I). En se restreignant aux mesures apportant une forte contribution à ces facteurs, on peut espérer conserver une bonne règle de discrimination des espèces. En procédant par étapes, nous avons été amenés à retrancher successivement les variables O21, O22 (diamètres orbitaux), O20, les largeurs T11, T14, T15, puis T10, T10b. Si les parts d'inertie augmentent sur les premiers axes, les interprétations des plans (1-2) et (3-4) ne sont pas sensiblement modifiées. Les résultats obtenus par la méthode de discrimination sur les 17 variables restantes sont donnés dans le tableau suivant :

Facteurs introduits successivement	2	1	3	4	6	5	7
% bien classés éch. de base	54.2	79.8	90.5	92.6	92.9	94.6	94
% bien classés éch. test	47.8	72.6	82.2	86.6	89.2	91.1	93.6

Poursuivant cet effort de réduction, nous avons abouti à ne retenir que 10 variables, les résultats sont alors les suivants :

Facteurs introduits successivement	1	2	4	3	6	5
% bien classés éch. de base	53.8	79.5	84.2	86.	88.9	88.3
% bien classés éch. test	48.5	71.6	81.0	87.	86.4	89.4

En permettant de bien reclasser près de 9 crânes sur 10, la règle de discrimination demeure intéressante.

Nous nous réservons d'exposer ultérieurement aux statisticiens ce que l'analyse multidimensionnelle a apporté à la paléontologie.

6. Conclusion

Les analyses de correspondances des données crâniométriques récoltées sur 349 *Equus* actuels permettent de reconnaître 6 grands groupes : caballins (chevaux domestiques et sauvages), hémioniens, (hémiones et kiangs), asiniens (ânes domestiques et sauvages), vrais zèbres, zèbres de Grévy et zèbres de Burchell, et de mettre en évidence leurs caractéristiques. Environ 60% de l'information est extraite par les deux premiers plans factoriels qui restent stables malgré certaines modifications apportées au tableau initial (pondérations des variables, division de variables continues en classes discontinues, réduction du nombre de variables utilisées).

Les analyses discriminantes permettent d'attribuer correctement plus de 90% des spécimens à l'un des 6 groupes précédemment cités. L'étude des voisinages semble apporter des résultats plus fiables que ceux des autres analyses discriminantes.

Les méthodes utilisées ont permis d'établir une hiérarchie parmi des variables retenues au départ : certaines mesures se sont montrées inutiles ou gênantes, d'autres sont au contraire particulièrement importantes. En conséquence, il a été possible de réduire de 25 à 10 le nombre des variables, tout en conservant une assez bonne description et discrimination des crânes.

La représentation graphique des affinités crâniennes entre les principaux groupes d'*Equus* sur le premier plan des analyses factorielles se fait suivant une couronne ; elle est presque identique à la représentation circulaire imaginée par Bourdelle à partir d'observations anatomiques variées. La ressemblance des deux représentations montre que l'anatomie crânienne concorde avec les autres données morphologiques externes et internes et qu'elle peut être utilisée pour caractériser certains taxons.

Par ailleurs ces relations circulaires se prêtent mal à des dichotomies : *a priori* n'importe quelles coupures du cercle peuvent être imaginées, associant divers groupes contigus d'espèces. La classification ascendante hiérarchique indique que certains regroupements sont plus probables que d'autres et permet ainsi de proposer à la discussion un schéma de relations phylétiques entre les espèces actuelles.

La mise en évidence de l'importance de certaines variables et les résultats concernant la parenté morphologique des espèces actuelles trouveront leur application dans l'étude des *Equus* fossiles.

BIBLIOGRAPHIE

- BOURDELLE E., 1955. Sous-ordre des *Hippomorpha*, famille des *Equidae*, sous-famille des *Equinae* in GRASSE P.P. : *Traité de zoologie*, tome XVII, fasc. 1 : 1006-1088, fig. 961-1030. Masson et Cie.
- GROVES C.P., 1974. *Horses, Asses and zebras in the wild* : 192 p, 12 fig, 16 pl. hors texte. Ed. David et Charles, Newton Abbot, London.
- MOHR E., 1971. *The asiatic wild Horse (translated by D.M. Goodhall)* : 124 p, 100 fig, J.A. Allen et co., London.