

FRO TM 201

Les Cahiers de l'Analyse des Données
Vol XIV - 1989 - n°3 - pp. 385-392.

**TYPOLOGIE DES
VIRUS DE PLANTES EN BÂTONNET
D'APRÈS LA COMPOSITION EN ACIDES AMINÉS
DE LEUR PROTÉINE DE CAPSIDE**

[VIRUS]

D. DESBOIS*, G. VIDAL*
D. FARGETTE**, C. FAUQUET**

1 Aperçu de la taxinomie des phytovirus en bâtonnet

La nomenclature du CITV (Comité International pour la Taxinomie des Virus) reconnaît, dans l'ensemble des virus en bâtonnet rigides (RSV), trois groupes caractérisés par le nombre de molécules d'ARN que comporte le génome: les **tobamovirus** (9 virus à génome monopartite), les **tobravirus** (2 virus à génome bipartite) et les **hordeivirus** (3 virus à génome tripartite).

D'autres caractères devant être pris en compte pour s'assurer que l'on constitue des groupes homogènes, cette nomenclature laisse de nombreux virus non classés (au moins 21). On a donc récemment proposé de créer un nouveau groupe: les **furovirus**, ou virus en bâtonnet transmis par les champignons du sol. Le type en serait le SBMV (soil-born wheat mosaic virus), virus de la mosaïque du blé transmis par le sol; lequel, bien que doté d'un génome bipartite, n'a pu être rangé avec les tobavirus.

D'ailleurs, le mode de transmission apparaît comme un critère de bon rendement; car les tobamovirus sont tous transmis par contact; tandis que les tobavirus le sont par des nématodes; le mode de transmission des hordeivirus restant seul inconnu.

Cependant, la composition en acides aminés de la protéine de la capsid du virion est un autre critère que la taxinomie virale utilise depuis plus de 20 ans, avec des succès divers. Grâce à l'analyse des correspondances et à la classification ascendante hiérarchique, nous avons pu tirer de ce critère des schémas taxinomiques s'accordant, sur un vaste corpus, avec ceux adoptés par

(*) Assistants techniques au C.U.T.I., Université Nationale de Côte d'Ivoire.

(**) Chargés de recherche à l'ORSTOM.

خلال سنة او عدة س
سنوية و حلت هذه
النسب على درجات.

المقارنة

تشيريات مختلفة

يظهر كاتب

التجميعية التي تقابل

للتحليل المنطقي.

حالة البحوث

موجب المعطيات

ص : ٤٤٩-٤٦٦.

بعد تذكير

التحسينات العديدة ال

مقارنة

الأشخاص المهتمين

عند المعالج المهتم

ج. معايطي ، ت.

تبدي الدرا.

لتأمين حياة المرضى .

فهارس ال

جدول المهتم

رسالة للقر

Fonds Documentaire IRD



010022165

Fonds Documentaire IRD

Cote : B* 22165 Ex: Unique

le CITV (cf. Fauquet et coll., 1986 et 1987): l'objet de la présente étude est d'appliquer cette méthode aux données disponibles relatives aux RSV.

2 Tableau des données et enchaînement des analyses

Nous avons rassemblé, en un tableau 61×18 , l'ensemble des données publiées sur la composition en acides aminés de la protéine capsidaire des phytovirus RSV, en y adjoignant certaines compositions nouvelles obtenues sur des isolats de Côte d'Ivoire. Il importe de souligner le terme d'*isolats*: pour ce que l'on considère comme un même virus, on peut disposer de plusieurs profils, déterminés sur des souches distinctes, et légèrement différents entre eux. Ainsi, les 9 tobamovirus fournissent 39 lignes à notre tableau; le BSMV (barley stripe mosaic virus), type des hordeivirus, est dans 3 lignes.

Sans entrer dans des détails qui ne concernent pas le statisticien, il faut expliquer les sigles adoptés pour les 61 éléments de l'ensemble *I*. Chaque sigle est formé d'un nombre précédé d'un caractère relatif au mode de transmission: 'C', contact, pour les tobamovirus; 'N', nématodes, pour les tobavirus; 'F', pour la transmission fongique, caractère suggéré pour un éventuel groupe des furovirus; '?', pour les hordeivirus dont le mode de transmission est inconnu; 'f', enfin, pour le CCV (*Chara corallina* virus), virus non classé.

Sur le tableau de dépouillement de la classification automatique, on a, outre nos sigles, rappelé ceux par lesquels les spécialistes désignent les espèces, (généralement des abréviations de locutions anglaises: e.g. PCV pour "peanut clump virus").

L'ensemble *J* est formé des acides aminés, constituants des protéines animales et végétales, désignés par leurs sigles usuels (GLY pour glycine, ALA pour alanine, etc...). Il faut seulement noter que les diacides ne sont pas distingués de leurs monoamides: ASX désigne le cumul de l'acide aspartique avec l'aspartine; et GLX, celui de l'acide glutamique avec la glutamine.

Ceci posé, on a dans notre tableau, à l'intersection de la ligne *i* et de la colonne *j*, le nombre $k(i,j)$ de résidus acides aminés *j* de la protéine de capsid du virus *i*. Par exemple, $k(?182, VAL) = 10$, parce que, pour la souche 182 (de l'hordeivirus BSMV dont le mode de transmission est inconnu), le nombre de molécules de *valine* trouvées dans la protéine de la capsid est 10.

On a d'abord soumis le tableau $I \times J$, (61×18), tel quel, à l'analyse des correspondances. Puis on a effectué une classification ascendante hiérarchique (CAH) sur chacun des deux ensembles *I* et *J*. De la classification sur *I*, on a extrait une partition de l'ensemble des souches de virus en 9 classes. Afin d'interpréter cette partition (en terme de composition de la capsid), on a, d'une part, projeté les centres des classes sur les axes issus de l'analyse factorielle; et, d'autre part, considéré les profils moyens des classes sur l'ensemble des acides

aminés, ou plus précisément ensemble.

3 Résultats

3.1 Analyse

3.1.1 Tableau

trace : 1.097e-1
rang : 1
lambda : 416 20
taux : 3789 18
cumul : 3789 56

3.1.2 Représentation

SIGJ	QLT	PDS	INR
ASX	163	117	171
THR	908	79	1031
SER	678	89	681
GLX	746	103	441
PRO	473	54	261
GLY	980	57	1991
ALA	451	99	291
CYS	293	5	161
VAL	670	72	311
MET	875	8	871
ILE	703	43	261
LEU	96	81	141
TYR	374	32	261
PHE	790	45	341
HIS	895	8	861
LYS	929	28	1381
ARG	779	65	311
TRP	411	15	251

Nous publions
acides aminés contr

L'axe 1 est c
contribution princij
plus écarté est l'H
apporte une contrib

Sur le demi-ax

Mais c'est sur
l'axe est créé par l'
que la CAH range

On notera qu
MÉThionine et l'H
et 86%, respective

la présente étude est
s aux RSV.

analyses

semble des données
otéine capsidaire des
ouvelles obtenues sur
me d'isolats: pour ce
er de plusieurs profils,
rents entre eux. Ainsi,
BSMV (barley stripe

le statisticien, il faut
emble I. Chaque sigle
ode de transmission:
ur les tobnavirus; 'F',
n éventuel groupe des
mission est inconnu;
classé.

tomatique, on a, outre
ésignent les espèces,
g. PCV pour "peanut

tuants des protéines
Y pour glycine, ALA
iacides ne sont pas
de l'acide aspartique
glutamine.

de la ligne i et de la
protéine de capsid
er la souche 182 (de
anu), le nombre de
t 10.

el, à l'analyse des
adante hiérarchique
ification sur I, on a
en 9 classes. Afin
apsid), on a, d'une
alyse factorielle; et,
ensemble des acides

aminés, ou plus précisément sur une partition en 7 classes adoptée pour cet ensemble.

3 Résultats des analyses

3.1 Analyse factorielle

3.1.1 Tableau des valeurs propres et taux

trace :	1.097e-1										
rang :	1	2	3	4	5	6	7	8	9	10	
lambda :	416	206	135	114	61	39	27	23	15	14	e-4
taux :	3789	1875	1232	1037	552	353	248	213	140	124	e-4
cumul :	3789	5664	6897	7933	8486	8839	9086	9299	9440	9564	e-4

3.1.2 Représentation de l'ensemble J des acides aminés

SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
ASX	163	117	17	-29	53	2	16	17	1	-14	12	2	36	81	13
THR	908	79	103	299	628	171	20	3	2	11	1	1	199	277	275
SER	678	89	68	111	147	26	-143	242	88	-151	270	149	40	19	12
GLX	746	103	44	38	31	4	87	161	38	143	431	156	-77	123	53
PRO	473	54	26	21	9	1	-123	285	40	25	12	3	-94	167	42
GLY	980	57	199	-559	821	430	122	39	41	-56	8	13	207	112	215
ALA	451	99	29	2	0	0	-68	146	22	-1	0	0	-99	305	85
CYS	293	5	16	76	17	1	39	5	0	258	193	25	164	79	12
VAL	670	72	31	49	50	4	138	399	67	-102	215	55	-17	6	2
MET	875	8	87	240	48	11	-469	185	86	829	578	408	276	64	54
ILE	703	43	26	93	132	9	191	550	76	-37	21	4	8	1	0
LEU	96	81	14	17	16	1	14	10	1	25	33	4	-27	38	5
TYR	374	32	26	4	0	0	27	8	1	138	215	45	115	151	38
PHE	790	45	34	143	243	22	-133	208	38	-167	328	91	-30	10	3
HIS	895	8	86	-946	809	183	191	33	15	208	39	27	-127	15	12
LYS	929	28	138	-434	354	129	-552	572	421	-5	0	0	-47	4	5
ARG	779	65	31	-39	29	2	127	307	51	-4	0	0	-152	442	132
TRP	411	15	25	108	64	4	126	87	11	126	86	17	-179	174	41

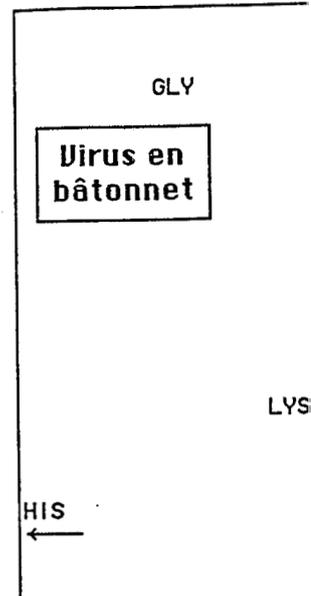
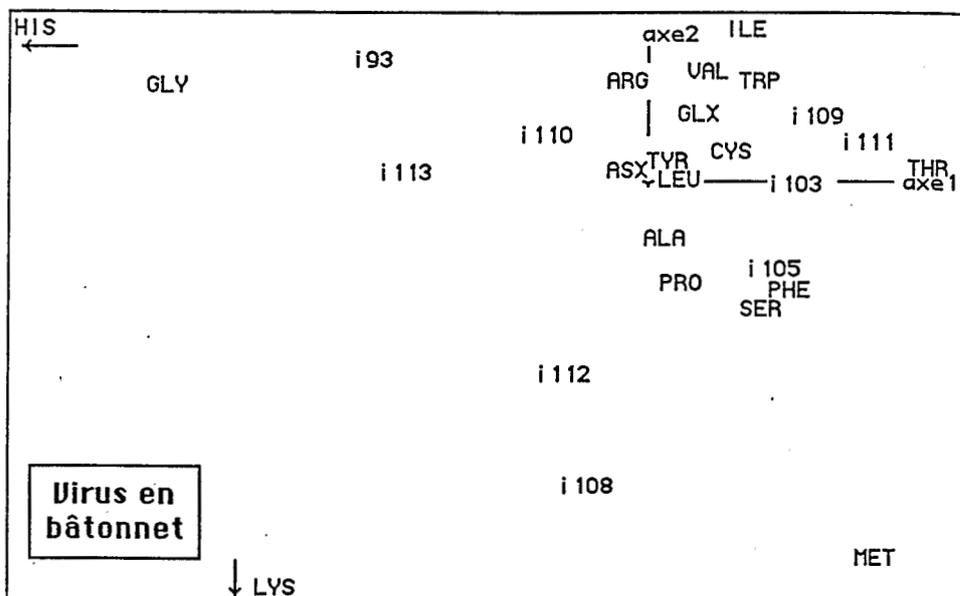
Nous publions ce listage afin de montrer avec précision comment les divers acides aminés contribuent à créer les 4 premiers axes.

L'axe 1 est créé par {GLY, HIS, LYS}, situés du côté (F1<0); la contribution principale vient de la GLYcine (CTR1 = 430%); mais le point le plus écarté est l'HISTidine (F1 = -.946) qui, malgré sa rareté (PDS = 8%), apporte une contribution de 183%.

Sur le demi-axe (F2<0) on trouve la LYSine associée à la MÉThionine.

Mais c'est sur l'axe 3 que MET est le mieux représenté (CO2 = .578); et l'axe est créé par l'association de MET et de quelques souches de tobamovirus que la CAH range dans la classe i103 (cf. infra, §3.2.2).

On notera que deux acides aminés peu fréquents (PDS = 8%), la MÉThionine et l'HISTidine contribuent notablement à l'inertie globale (INR = 87 et 86%, respectivement).



3.1.3 Représentation de l'ensemble I des souches de virus

Il n'y a pas lieu de marquer individuellement les 61 souches: on a seulement placé, sur les plans (1,2) et (1,4), les centres des 9 classes de la partition retenue. Le plan (1,3) n'est pas publié, parce que l'axe 3 est occupé par l'association déjà signalée entre MET et la classe i103.

3.2 Classification ascendante hiérarchique

3.2.1 Classification de l'ensemble des acides aminés

La première dichotomie, visible dans le plan (1,2) et déjà signalée comme créant l'axe 1, est celle qui oppose j33 = {LYS, HIS, GLY} à tout le reste, qui constitue la classe j34. Celle-ci se subdivise en j30 et j32.

Dans j32 s'agrègent à un niveau élevé MET, THR et la paire {PHE, SER}, notée dans la suite PHSR.

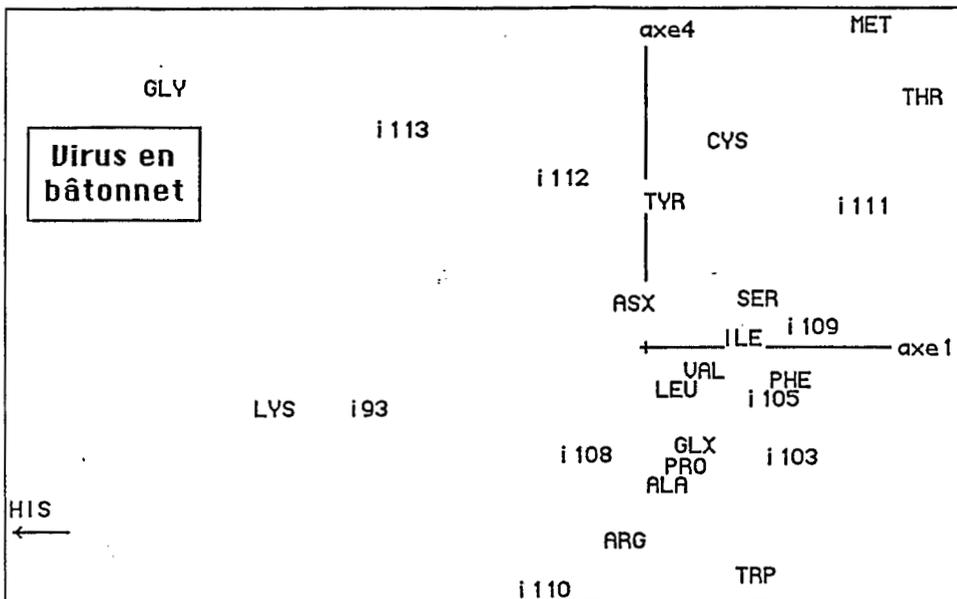
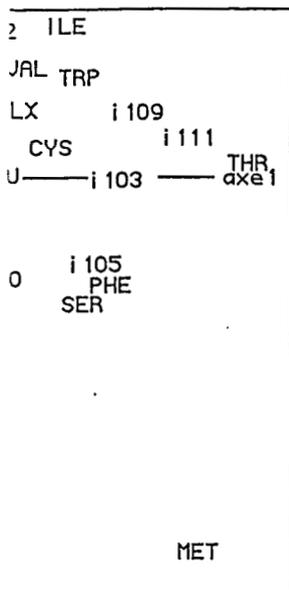
La majorité des acides aminés est comprise dans la classe j30 où s'agrègent à un bas niveau les classes j29 et j28; cette dernière est distribuée autour de l'origine des axes (ou centre de gravité du nuage). La classe j29 se voit dans le quadrant (F1>=0, F2>0).

On remarquera que la CYS téine, acide aminé soufré rare dans nos capsides, s'agrège au sein de j29 à GLX (acide GLutamique et GLutamine comptés ensemble)

Pour étiqueter l'a donne , sur les axes fa des souches de virus s de J qui comprend d

ASX	2125	28	30
LEU			=cdg
TYR			
ALA	23		
PRO			
CYS	20	29	
GLX			F2++++
TRP	2226		
ARG			
ILE	19		
VAL			
MET		F3++++	
THR		F1++++	
PHE	24		
SER			
LYS			F2---
HIS		27	
GLY			

d'effectifs plus élevé:
(LYS, MET)



souches de virus

on a seulement les classes de la partition d'axe 3 est occupé par

acides aminés

et déjà signalée comme {LY} à tout le reste, qui

et la paire {PHE, SER},

la classe j30 où s'agrègent est distribuée autour de la classe j29 se voit dans le

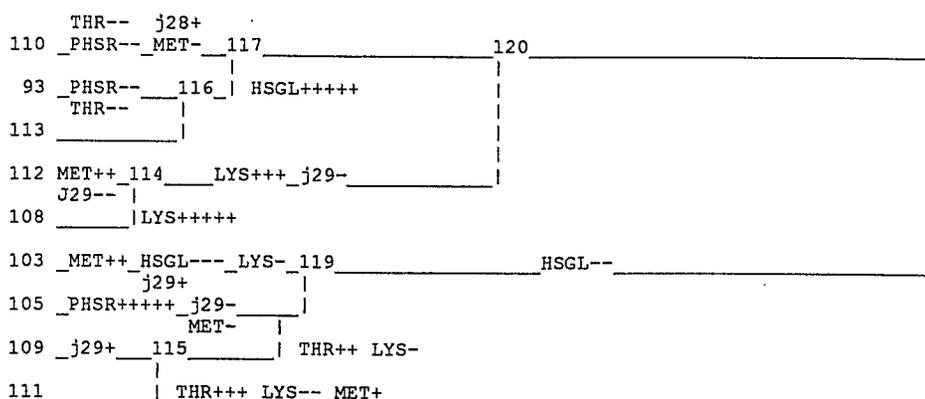
est rare dans nos capsides, et GLutamine comptés

Pour étiqueter l'arbre des acides aminés, on a utilisé le listage Factor qui donne, sur les axes factoriels, les coordonnées des centres des classes. L'arbre des souches de virus sera étiqueté, d'après le listage Vacor, suivant une partition de *J* qui comprend deux acides aminés isolés, deux paires et deux classes

ASX	2125	28	30	34
LEU			=cdg	
TYR				
ALA	23			
PRO				
CYS	20	29		
GLX			F2++++	
TRP	2226			
ARG				
ILE	19			
VAL				
MET	F3++++	32		F1++
THR	F1++++	31		
PHE	24		(PHSR) F2-- F3--	
SER				
LYS	F2----	33		F1----
HIS	27		(HSGL) F1----	
GLY				

d'effectifs plus élevé:

{LYS, MET, HSGL, PHSR, j28, j29}.



ci-dessus, l'arbre de la partition des phytovirus en 9 classes
ci-dessous, tableau du contenu des classes

c Partition en 9 classes : Sigles des individus de la classe numéro c		
110	?182 ?311 ?312 F251	hordeivirus & PCV India
93	F38 F34 F36 F37 F39 F35	PCV
113	F260 F262 F259 F261	RSNV

112	F46 f52 N328 N327	BNYVV CCV TRV
108	N310 N309 N331 N330	TRV

103	C321 C320 C322 C323 C317 C318 C51 C319 C53	HRV UMMV CRMV TOSNV
105	C55 C296 C58 C59 C295 C297	CV CGMMV
109	C315 C50 C67 C66 C48 C286 C60 C47 C291	TMV TOMV SOV FMV
	C65 C64 C63 C62 C49 C313 C314 C316	
111	C69 C54 C68 C292 C56 C293 C57	U2 O2 T2MV ORSV GTAMV

3.2.2 Classification de l'ensemble des souches de virus

Construite par voie ascendante (c'est-à-dire en agrégeant d'abord les individus le plus proches, puis les individus aux petites classes ainsi créées ou ces petites classes entre elles) une classification peut être lue aussi bien à partir du bas que du haut.

Au sommet de la hiérarchie, on a la dichotomie {i120 ≠ i119}, avec dans cette dernière classe l'ensemble des déterminations afférentes à des tobamovirus (virus transmis par contact, dont les sigles commencent par la lettre 'C').

La classe i120 se partage entre i114 et i117, elle-même subdivisée en i110 et i116. En bref, à des exceptions près sur lesquelles nous reviendrons, on peut poser: i114 = tobavirus = 'N', transmission par les nématodes; i116 = furovirus = 'F', transmission fongique; i110 = hordeivirus = '?', mode de transmission inconnu.

En partant de la b
agrégée à un niveau trè
que l'on a dans notre
souches de même n
correspondants sont m
d'après les noms en lett
les lignes du tableau du

À cette interprétat
seules exception 3 déter
priori de ranger dans l
sépare des souches de
BNYVV (beet necrotic
f52 (CCV), dont le mod

Reste à interpréter
aminés dans la protéin
factorielle suffit à sug
entrer dans les détails il
l'arbre.

S'opposent d'abor
toutes ses subdivisions
très faibles en H1Stidi
i120, plus hétérogène,
acides aminés nettemen

Dans le quadrant
associée à la LYSine, p

L'autre subdivisio
détache la classe i110,
notamment intervenir
(TRiPtothane, ARGini

Au sein de la class
notée pour un taux élev
i103 et i105, s'opposer
particulièrement hauts
ainsi qu'on l'a dit, par l
METHionine; acide am
est plutôt rare dans le r
le nombre maximum
capside).

Il conviendrait er
virus avec la structure s

4 Conclusion taxinomique

Selon les normes du CITV, un virus n'est affecté à un groupe que s'il en possède toutes les propriétés caractéristiques; si l'une des propriétés manque, ou seulement si l'étude du virus n'est pas achevée, on donne à celui-ci le statut de "non classé" ou de membre "putatif" d'une classe. En outre, faute d'une représentation spatiale globale, la classification du CITV ne permet pas de mettre en évidence des relations entre les groupes, ni de proposer l'existence de groupes intermédiaires.

À l'inverse, notre typologie fondée sur la composition en acides aminés de la capsidie, utilise des calculs de distance entre virus, et permet de calculer des distances entre centres de classes ou entre virus et centre de classe. L'inconvénient est qu'il reste à intégrer dans la représentation multidimensionnelle les autres données disponibles relatives aux virus; cela n'est d'ailleurs pas impossible.

Il faut également noter que certains virus en bâtonnet sont très labiles et ne se trouvent dans les plantes infectées qu'à des concentrations très faibles. En pareil cas, il est cependant possible d'obtenir de petites quantités de protéine capsidaire purifiée. L'analyse de celle-ci apparaît susceptible de fournir des critères relativement sûrs pour classer les virus et identifier de nouveaux groupes.

Références bibliographiques

C. Fauquet, J. Dejardin, J.-C. Thouvenel: Evidence that the amino-acid composition of the particle proteins of plant viruses is characteristic of the virus group; in *Intervirology*, Vol XXV (1986); I, pp. 1-13, II, pp. 190-200.

C. Fauquet, D. Desbois, D. Fargette, G. Vidal: Classification des virus de plantes par la composition en acides aminés de leur protéine capsidaire; in *Résumé des Rencontres de Virologie*, p.9, 1-5 février 1987, Aussois, France.

MARCHÉ PREMIÈRE RENTABILITÉ

1 Origine et st

Comme dans un
n°1) nous analysons
Immobilier: d'une par
villes de province ave
des logements Neufs,
deux-pièces; d'autre
arrondissements de F
mètre carré et de locat

Il n'y a pas lieu
IMMOB.] pour la co
Nous devons, en reva
de chacun de nos deu
mètre carré.

Si le premier tab
(ce que nous avons c
loyers est environ le d
on a donc créé et anal
par 10.

Dans le deuxièm
loyers et prix de ven
donnés sur 2 colonn
distingue 4 tailles de l

(*) Assistant de Mathém.
Étudiant en doctorat à l'U

En partant de la base de la hiérarchie édifée sur *I*, on trouve, chacune agrégée à un niveau très bas, les 9 classes de la partition retenue. Chaque fois que l'on a dans notre corpus des déterminations multiples portant sur des souches de même nom ou reconnues pour très proches, les profils correspondants sont mis dans une même classe: le lecteur en pourra juger d'après les noms en lettres capitales qui accompagnent les sigles numériques sur les lignes du tableau du contenu des classes.

À cette interprétation de la CAH en termes de contenu des classes font seules exception 3 déterminations afférentes à des souches qu'on envisageait *a priori* de ranger dans le nouveau groupe des furovirus: F251 (PCV-India) se sépare des souches de PCV pour rejoindre les hordevirus proches; F46, BNYVV (beet necrotic yellow vein virus), s'agrège aux tobavirus; en fait autant f52 (CCV), dont le mode de transmission est toutefois inconnu.

Reste à interpréter les classes en termes de proportions des divers acides aminés dans la protéine de la capsid. L'examen des plans issus de l'analyse factorielle suffit à suggérer les grandes lignes de l'interprétation; mais pour entrer dans les détails il faut recourir au listage Vacor qui a servi à l'étiquetage de l'arbre.

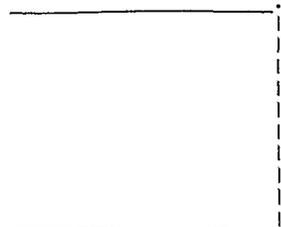
S'opposent d'abord sur l'axe 1 les classes i120 et i119; celle-ci est, dans toutes ses subdivisions retenues, caractérisée par des taux relativement faibles ou très faibles en HISTidine et GLYcine (HSGL). Les subdivisions de la classe i120, plus hétérogène, ont généralement des taux élevés en l'un ou l'autre des 3 acides aminés nettement situés du côté (F1<0): {HIS, GLY, LYS}.

Dans le quadrant (F1<0, F2<0), s'isole la classe i114 des tobavirus, associée à la LYSine, particulièrement dans sa subdivision i108.

L'autre subdivision, i117, de la classe i120, se partage sur l'axe 4, où se détache la classe i110, du côté (F4<0): l'étiquetage de i110 est complexe, faisant notamment intervenir, avec le signe +, les acides aminés de la classe j28 (TRiPtothane, ARGinine, ...).

Au sein de la classe i119 des tobavirus, la classe i115 est principalement notée pour un taux élevé de THRéonine; tandis que les deux autres subdivisions, i103 et i105, s'opposent sur l'axe 3; celle-ci (F3<0), caractérisée par des taux particulièrement hauts en PHÉnylalanine et SÉRine (PHSR); et celle-là (F3>0), ainsi qu'on l'a dit, par les nombres élevés, dans leurs capsides, de molécules de MÉThionine; acide aminé qui, ainsi qu'on l'a vérifié sur le tableau des données est plutôt rare dans le reste de notre corpus (mis à part le cas de BNYVV, qui a le nombre maximum de 7 molécules, parmi les 198 que compte au total sa capsid).

Il conviendrait encore de mettre en rapport le mode de transmission des virus avec la structure spatiale de la capsid édifée par les acides aminés...



classes

la classe numéro c

PCV India
PCV
RSNV

BNYVV CCV TRV
FRV

HRV UMMV CRMV TOSNV
CV CGMMV
FMV TOMV SOV FMV

J2 O2 T2MV ORSV GTAMV

es de virus

agrégant d'abord les classes ainsi créées ou lue aussi bien à partir

20 ≠ i119), avec dans ntes à des tobavirus ur la lettre 'C').

me subdivisée en i110 s reviendrons, on peut s nématodes; i116 = virus = '?', mode de