

**Classification des modalités d'une variable
qualitative :
approche classique, approche symbolique
et application en Ecotoxicologie**

Marie Chavent, Chantal Lacomblez

Laboratoire de Mathématiques Appliquées de Bordeaux (MAB),

Alain Boudou, Régine Maury-Brachet

Laboratoire d'Ecophysiologie et d'Ecotoxicologie des Systèmes aquatiques
(LEESA)

Université Bordeaux 1

Le problème :

	X	$Y_1 \dots Y_p$
1	Variable qualitative à M modalités	P variables quantitatives
...		
...		
N		

**Tableau de données classique :
approche classique**

	$Y_1 \dots Y_p$
1	P variables quantitatives intervalles
...	
...	
M	

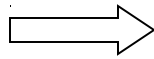
**Tableau de données intervalles :
approche symbolique**

⇒ **Trouver une partition des M modalités de la variable qualitative X**

⇒ **Méthode divisive de classification de type segmentation non supervisée**

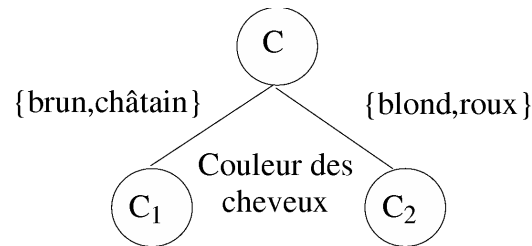
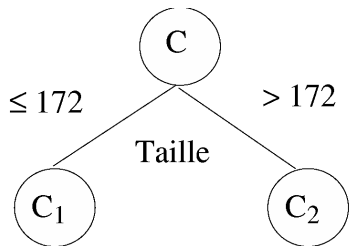
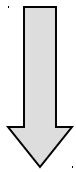
⇒ **Méthode divisive de classification de type segmentation non supervisée**

	$X_1 \dots X_q$	$Y_1 \dots Y_p$
1	q variables de coupure	p variables pour le critère
...		
...		
N		



$$D = \begin{pmatrix} d_{11} & \dots & d_{n1} \\ \vdots & & \vdots \\ d_{1n} & \dots & d_{nn} \end{pmatrix}$$

Critère d'évaluation des bi-partitions basé sur le tableau de distances



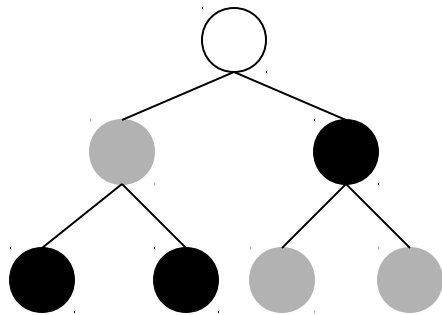
Division en fonction d'une variable et de sa transformation binaire

⇒ **Méthode divisive de classification de type segmentation non supervisée**

- Le critère d'évaluation des partitions :

$$Q(C_k) = \frac{1}{2nn_k} \sum_{j \in C_k} \sum_{j' \in C_k} d_{jj'}^2 \quad \text{et} \quad W(P) = \sum_{k=1}^K Q(C_k)$$

- choix de la classe à diviser : maximiser $\Delta(C_k) = Q(C_k) - Q(C_k^1) - Q(C_k^2)$



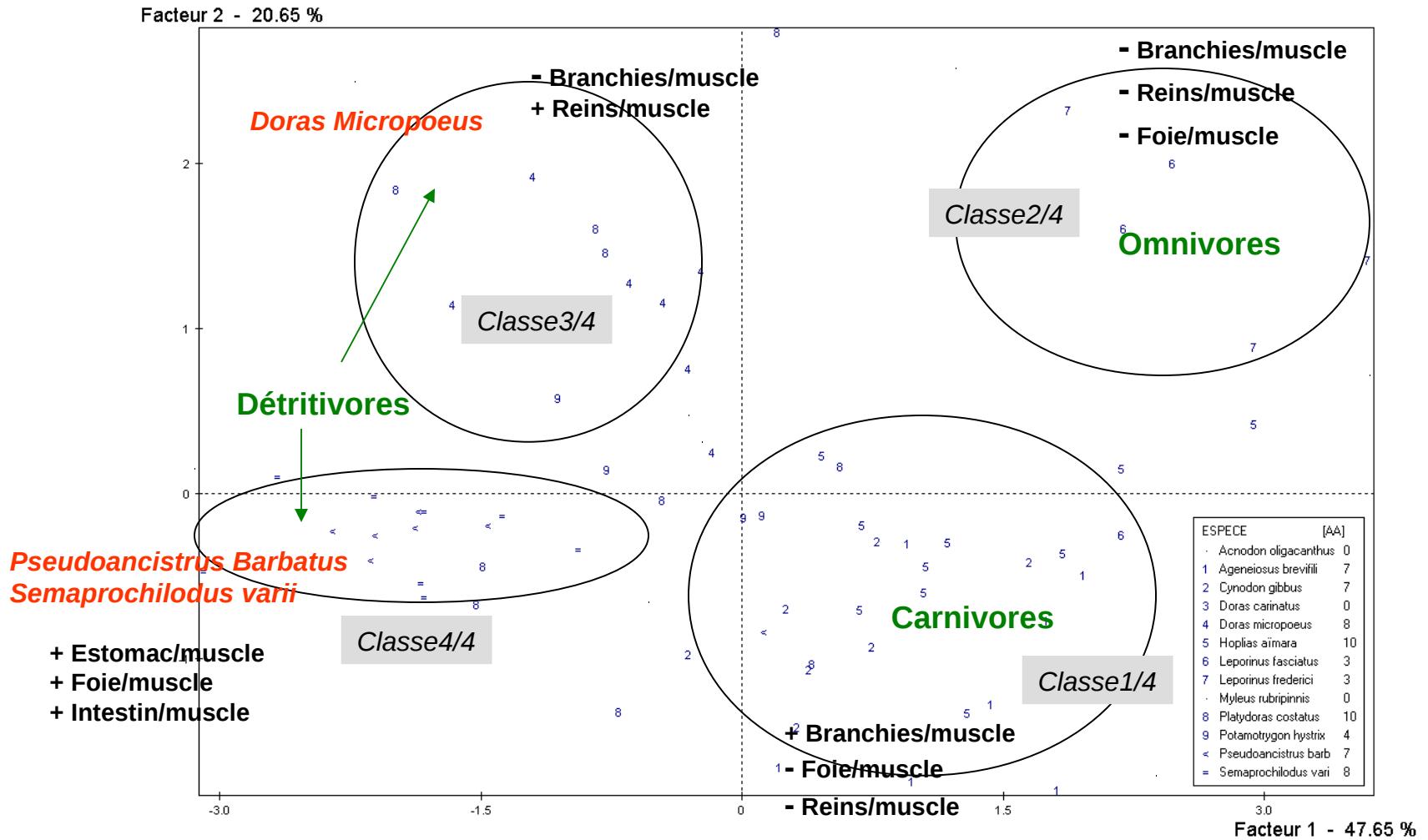
L 'application :

67 poissons appartenant à 10 espèces et 3 régimes alimentaires :

Carnivores	Omnivores	Détritivores
Ageneiosus brevifilis (7) Cynodon gibbus (7) Hoplias aimara (10) Potamotrygon hystrix (4)	Leporinus fasciatus (3) Leporinus frederici (3)	Doras micropoeus (8) Platydoras costatus (10) Pseudoancistrus barbatus (7) Semaprochilodus varii (8)

	ESPECE	REGIME	LN(Foie/Muscle)	...	LN(Estomac/Muslce)
1	Ageneiosus brevifili	Carnivore	-0,12	...	NA
	Cynodon gibbus	Carnivore	1,59	...	0,22
∴	∴	∴	∴	...	∴
	Leporinus frederici	Omnivore	-0,04	...	-1,77
∴	∴	∴	∴	...	∴
	Doras micropoeus	Détritivores	0,8	...	-0,89
67	Doras micropoeus	Détritivores	1,34	...	-1,45

Approche classique : étude des 67 poissons

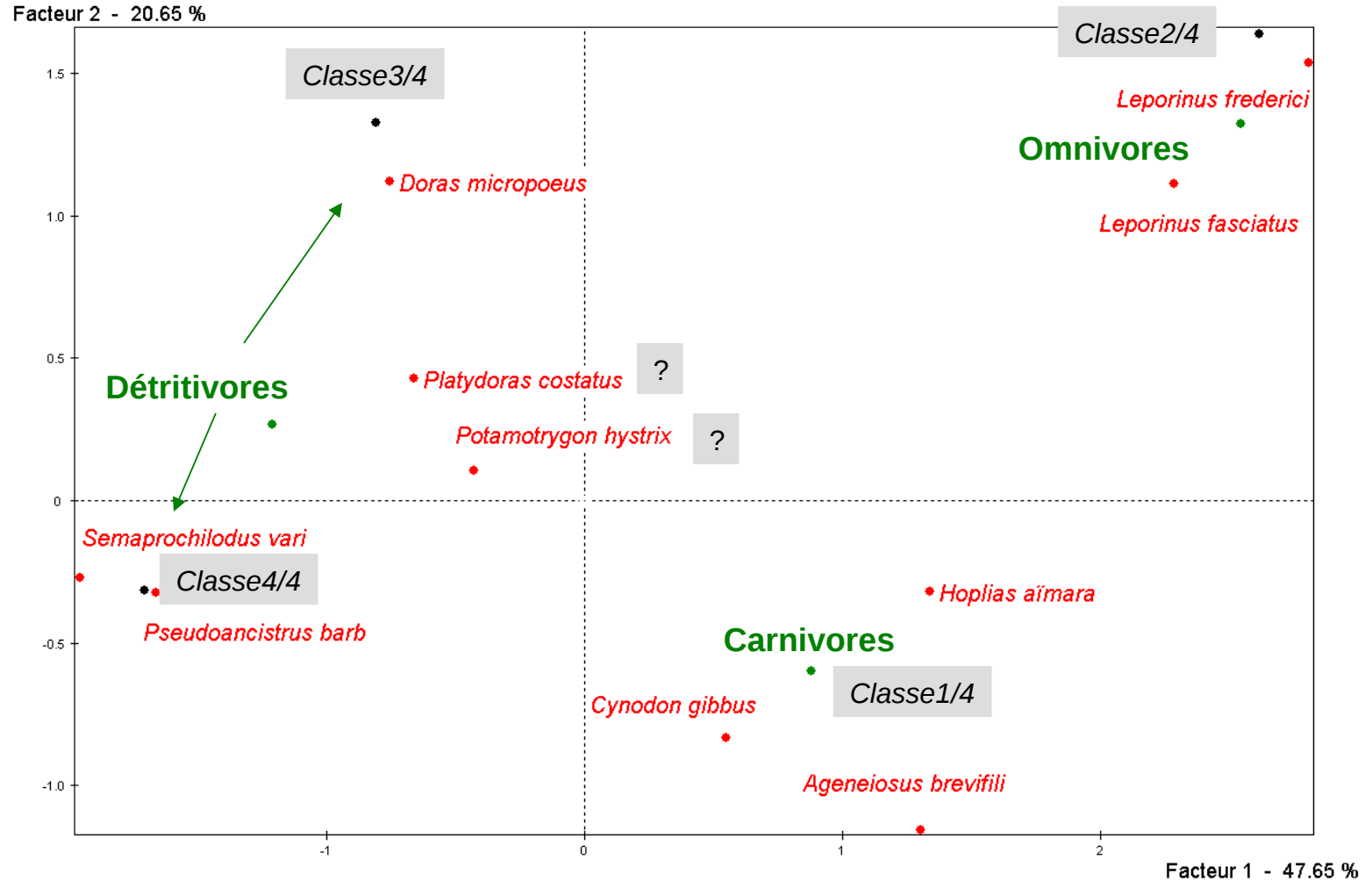


Classification hiérarchique en 4 classes, sur coordonnées factorielles

Approche classique : étude des 67 poissons

- **classe 1/4** : constituée à 86% de carnivores (41% de carnivores dans la population totale) et contient 92% des carnivores. Elle contient 100% des *Ageneiosus brevifilis*, des *Cynodon Gibbus* et des *Hoplias aîmara*.
- **Classe 2/4** : constituée à 100% d'omnivores et contient 83% des omnivores. Elle contient 100% des *Leporinus fasciatus* (3 sur 3) et 2 *Leporinus frederici* sur 3.
- **Classe 3/4** : constituée à 92% de détritivores et contient 36% des détritivores. Elle contient 100% des *Doras micropoeus*, ceux ci formant 61% de la classe.
- **Classe 4/4** : constituée à 94% de détritivores et contient 54% des détritivores. Elle contient 100% des *Semaprochilodus varii* et 85% des *Pseudoancistrus barbatus*.

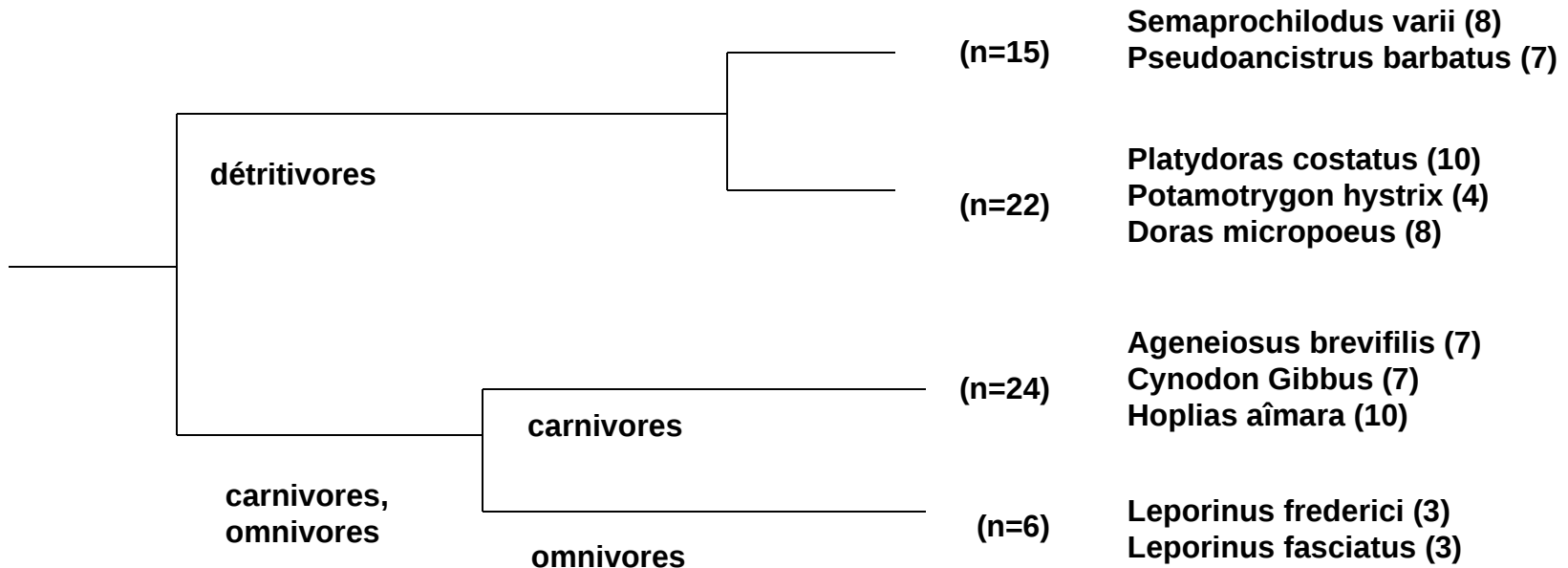
Approche classique : étude des 67 poissons



Classification hiérarchique en 4 classes, sur coordonnées factorielles

Approche classique : classification divisive

	ESPECE	LN(Foie/Muscle)	...	LN(Estomac/Muscle)
1	Ageneiosus brevifili	-0,12	...	NA
	Cynodon gibbus	1,59	...	0,22
⋮	⋮	⋮	...	⋮
	Doras micropoeus	0,8	...	-0,89
67	Doras micropoeus	1,34	...	-1,45



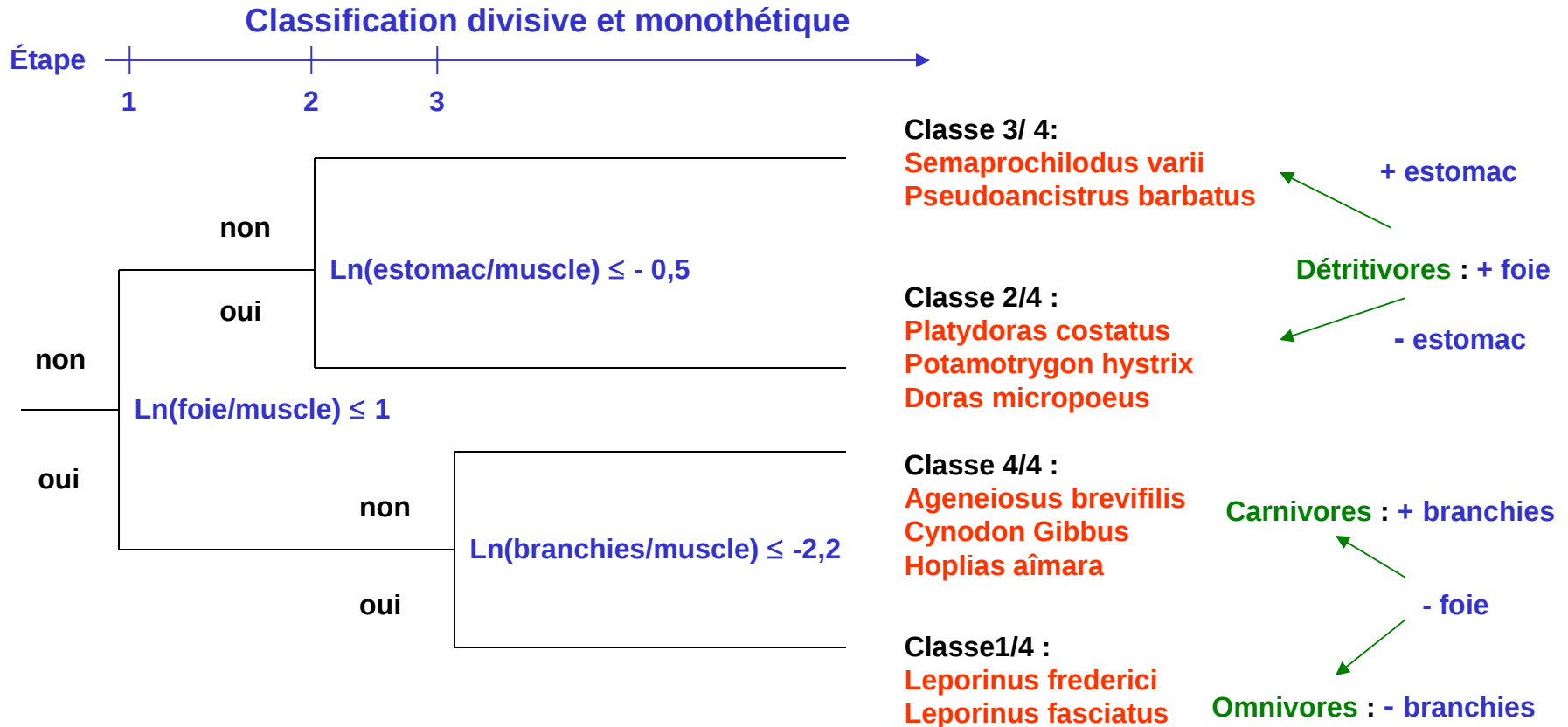
Approche symbolique : étude des 10 espèces

	LN(Foie/muscle)	...	LN(Intestin/muscle)	LN(estomac/muscle)
Ageneiosus brevifilis	[-0.8 ; -0.08]	...	[-1.45 ; -0.46]	[-1.48 ; -0.59]
⋮	⋮	⋮	⋮	⋮
Cynodon Gibbus	[0.12 ; 1.42]	...	[-1.67 ; -0.68]	[-1.61 ; -0.1]

⇒ Méthode DB2SO pour la construction des intervalles

⇒ Les variables de coupure et les variables utilisées pour calculer le critère sont les mêmes

Approche symbolique : classification divisive



- Étape 1 :** séparation des espèces **carnivores** (sauf **Potamotrygon hystrix**) et **omnivores** des espèces **détritivores** en fonction de la concentration dans le **foie** (par rapport au muscle).
- Étape 2 :** séparation des espèces **détritivores** en deux classes en fonction de la concentration dans l' **estomac**
- Étape 3 :** séparation des espèces **carnivores** et **omnivores** en fonction de la concentration dans les **branchies**