

Axe transversal : méthodologie

Laboratoire de Biométrie et Biologie Evolutive (UMR 5558);
CNRS; Univ. Lyon 1, 43 bd 11 nov, 69622, Villeurbanne Cedex,
France.

24 Nov. 2005



Motivations

Augmentation de la dispersion spatiale du laboratoire.

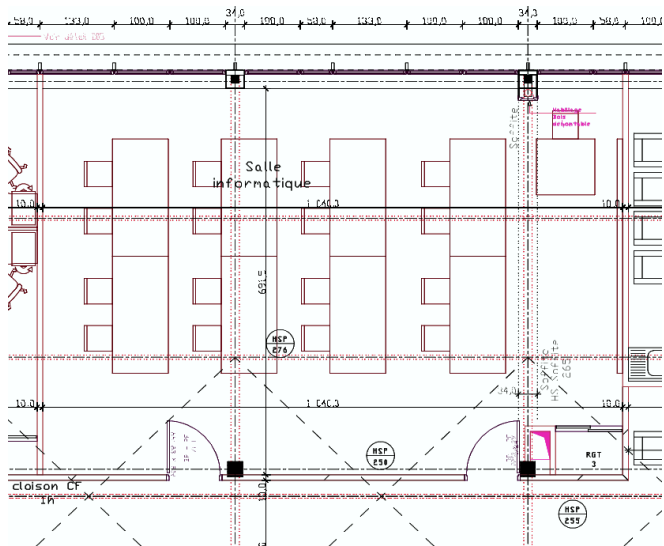
- 1 Favoriser les transferts méthodologiques d'un domaine d'application à l'autre
- 2 Favoriser l'émergence de nouvelles méthodes par confrontation d'idées
- 3 Favoriser une meilleure intégration de la recherche, de l'enseignement et de la formation continue
- 4 Favoriser une démarche de qualité (analyse de données, logiciels)

Mise en œuvre

Les moyens d'interaction

- 1 Réunions de travail
- 2 Séminaires
- 3 Liste de diffusion
- 4 Environnement de travail collaboratif
- 5 Codirections de thèses
- 6 Enseignement

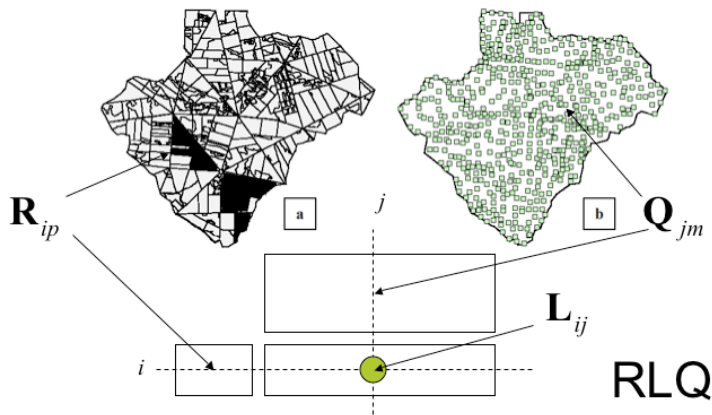
La salle de formation du PRABI



Perspectives scientifiques

- 1 Les statistiques bayésiennes
- 2 Les systèmes dynamiques stochastiques
- 3 L'analyse des données vectorielles
- 4 L'algorithmique
- 5 Les bases de données et de connaissances

Innovations méthodologiques en écologie

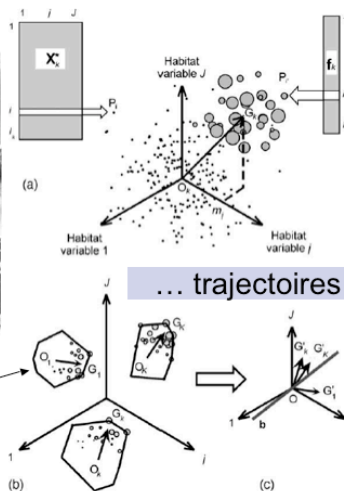
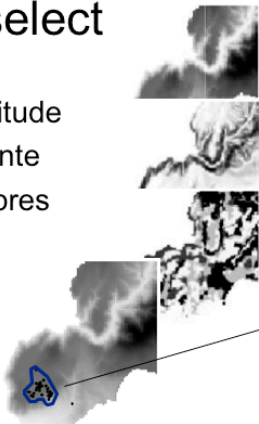


Dray, S., Pettorelli, N., Chessel, D. (2002) : *Journal of Vegetation Science* **13** : 867-874

Innovations méthodologiques en écologie

K-select

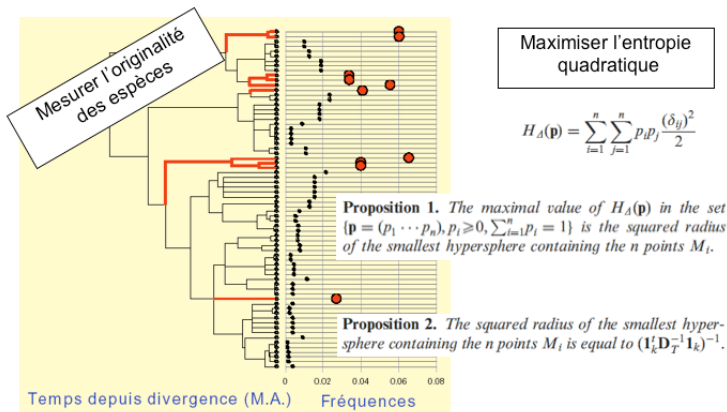
- Altitude
- Pente
- Arbres
- ...



Calenge, C., Dufour, A.-B. (2005). *Ecological modelling* **186** : 143-153

Innovations méthodologiques en écologie

Méthodes = théorèmes

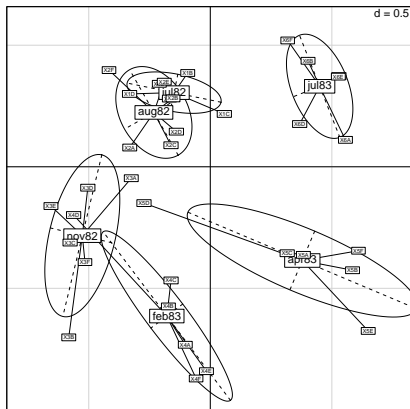


Pavoine, S., Ollier, S., Dufour, A.-B. (2005) *Ecology Letters* **8** : 579-586

Pavoine, S., Ollier, S., Pontier, D. (2005) *Theor. Pop. Biol.* **67** : 231-239

Transferts méthodologiques

```
library(ade4)
data(ardeche)
exemple(ardeche)
```




Exemple **ACI** : l'analyse des correspondances internes

1988 développée en hydrobiologie
Cazes, P., Chessel, D., and Dolédec, S. (1988)
Rev. Stat. Appl., **36**:
39-54.

2005 utilisée pour l'étude de l'usage du code synonyme des gènes tissus-spécifiques
Sémon *et al.* (2005) *Mol. Biol. Evol.* *in press*

Diffusion transparente des méthodes

Passage de ade4 en tant que package . Un utilisateur, Campo Elías Pardo, a écrit à propos du code source de l'**ACI** :

I look for the reason of this result. I found the causa:

```
tabinit <- data.frame(tabinit+wrmat)
```

why is wrmat summed? Is this an error?

The correct sentence is:

```
tabinit <- data.frame(tabinit)
```

Le logiciel **libre** favorise la **qualité** des logiciels diffusés.

Paradoxe de la reine rouge



CRAN Daily Package Check Results

Last updated on 2005-11-20 11:45:03

	Package	Version	Priority	Maintainer	r-devel	r-patched	r-release	Comment
1	AMORE	0.2-0		Manuel Castejón Limas	OK	OK	OK	

...

146	adapt	1.0-3		Thomas Lumley	OK	OK	OK	
147	ade4	1.4-0		Simon Penel	OK	OK	OK	
148	adehabitat	1.3		Clément Calenge	OK	OK	OK	

...

643	zicounts	1.1.4		S M Mwalili	WARN	WARN	WARN	
644	zoo	1.0-2		Achim Zeileis	OK	OK	OK	



Des modèles formels a tous les niveaux d'organisation

Des modèles. . .

- Mathématiques
- D'abstraction des connaissances
- De simulation d'algorithmes avec des programmes informatiques réputés justes

Au niveau des. . .

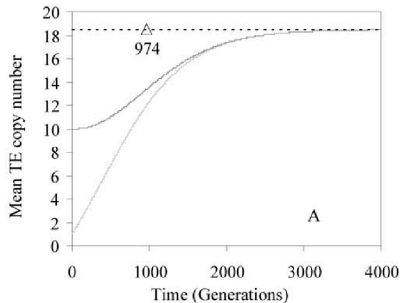
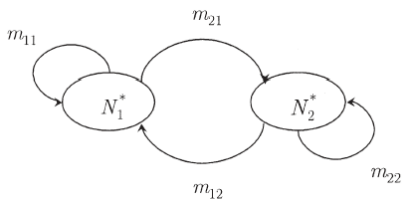
- populations
- groupes
- individus
- tissus
- cellules
- subcellulaire
- molécules

Quelques exemples publiés récemment :

Populations de drosophiles

$$\Delta \bar{n}_1 \approx \bar{n}_1((u - v) - (a + b\bar{n}_1)) + m(\bar{n}_2 - \bar{n}_1)$$

$$\Delta \bar{n}_2 \approx \bar{n}_2((u - v) - (a + b\bar{n}_2)) + m(\bar{n}_1 - \bar{n}_2)$$



Deceliere, G., Charles, S., Biémont, C. (2005) *Genetics* **169**:467-474

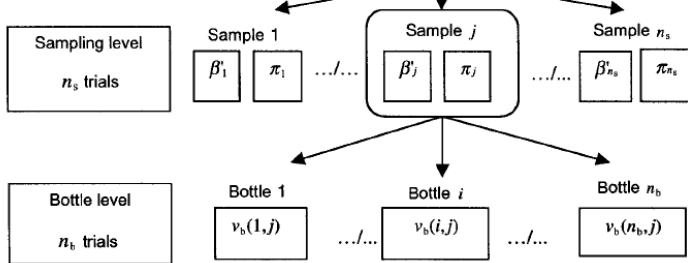
Populations de patients

Optimisation d'hémocultures.

Patient-to-patient variability

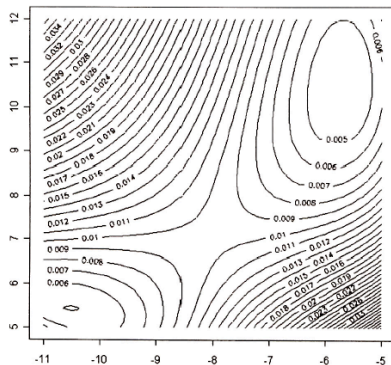
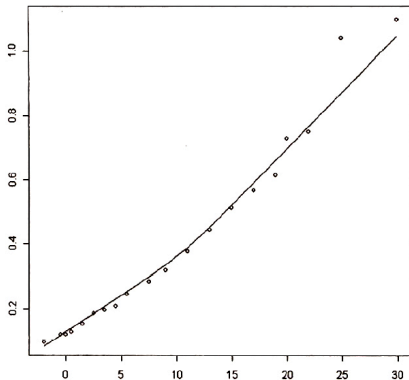


Patient k: Within-patient variability



Populations de *Listeria monocytogenes*

$$\mu_{\max} = \theta_4 \left[\frac{T - \theta_1}{\theta_3 - \theta_2} + \left(\frac{\theta_2 - \theta_1}{(\theta_3 - \theta_1)(\theta_3 - \theta_2)} \right) (1 + e^{(\theta_1 - \theta_3)}) \ln \left(\frac{1 + e^{(\theta_3 - T)}}{1 + e^{(\theta_3 - \theta_1)}} \right) \right]^2$$



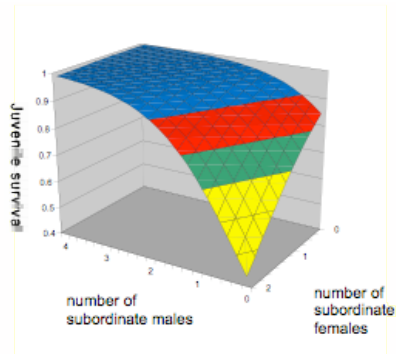
Charles-Bajard, S., Flandrois, J.-P., Tomassone, R. (2003) *Rev. Stat. Appl.* **LI**:59-71

Groupes de marmottes

Survie des marmottons en fonction du nombre de subordonnés mâles et femelles

$$w(h) = \sum_{i=0}^{\infty} p(h)^i * r * c * s(h)$$

$$f(h) = \sum_{i=0}^{\infty} r^i * p(h) * c * s(h)$$

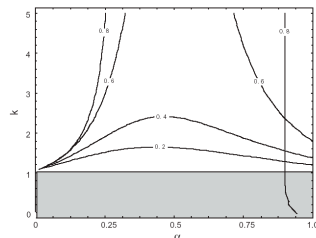


Allainé, D., Theuriau, F. (2004) *Behav. Ecol.* **15**:916-924

Individus de deux espèces

Diapause prolongée et stabilité des systèmes hôtes-parasites

$$\begin{aligned}
 N_t &= (1 - \alpha)\lambda N_{t-1}(1 + aP_{t-1}/k)^{-k} + \sigma(1 - \beta)\lambda ND_{t-1}, \\
 ND_t &= \alpha N_{t-1}(1 + aP_{t-1}/k)^{-k} + \sigma\beta ND_{t-1}, \\
 P_t &= (1 - \alpha)N_{t-1}\{1 - (1 + aP_{t-1}/k)^{-k}\} \\
 &\quad + \sigma(1 - \beta)PD_{t-1}, \\
 PD_t &= \alpha N_{t-1}\{1 - (1 + aP_{t-1}/k)^{-k}\} + \sigma\beta PD_{t-1}, \quad (C.2)
 \end{aligned}$$



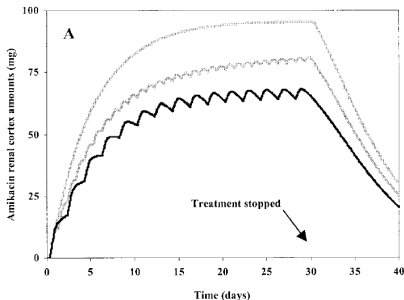
Corley, J.C., Capurro, A.F., Bernstein, C. (2004) *Theor. Pop. Biol.* **65**:193-203

Niveau tissulaire

Toxicité des aminoglycosides pour les reins.

$$\frac{dQ_{S(t)}}{dt} = -\{[K_i + K_s \cdot (1 - k_{reabs}) \cdot CL_{CR(t)}] + K_{cp}\} \cdot Q_{S(t)} + K_{pc} \cdot Q_{P(t)} + i(t)$$

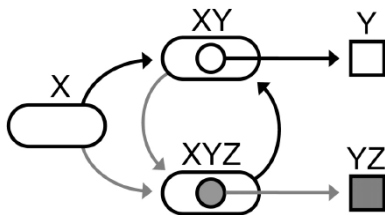
$$CL_{CR(t)} = CL_{CR0} - \frac{CL_{CRMAX} \cdot E_{(t)}^{\delta}}{E_{50}^{\delta} + E_{(t)}^{\delta}} + \eta \cdot CL_{CR(t)} \cdot \sin(\omega \cdot t + \phi)$$



Rougier, F. et al. (2003) *Antimicrobial Agents and Chemotherapy* **47**:1010-1016

Niveau subcellulaire

Un virus manipulateur du comportement.



Gandon, S., Rivero, A., Varaldi, J. (2006)
 Superparasitism evolution: adaptation or
 manipulation ? *Am. Nat.* in press

$$[x] = N - [xy] - [xyz],$$

$$[xy] = \{b[y] + \beta[yz](1 - \tau_v)\} \\ \times [x] - (d + e)[xy] + S_{xyz} - S_{xy},$$

$$[xyz] = \beta[yz][x]\tau_v - (d + e)[xyz] + S_{xy} - S_{xyz},$$

$$\frac{\partial y(n)}{\partial t} = -\frac{\partial y(n)}{\partial n} \frac{dn}{dt} - my(n),$$

$$\frac{\partial yz(n_z)}{\partial t} = -\frac{\partial yz(n_z)}{\partial n_z} \frac{dn_z}{dt} - myz(n_z),$$

$$S_{xyz} = \{sb[y] + \sigma\beta[yz](1 - \tau_v)\}\{[xyz]c(1 - \tau_h)\},$$

$$S_{xy} = \sigma\beta[yz][xy]\tau_v\{c + (1 - c)\tau_h\}.$$

Niveau subcellulaire

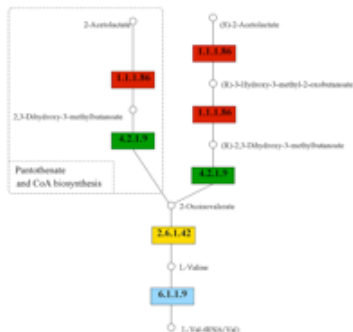
Modèles d'abstraction des voies métaboliques

Motif : ■ ■ ■ ■ ■

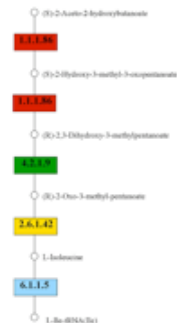
Biosynthèse de la Leucine



Biosynthèse de la Valine



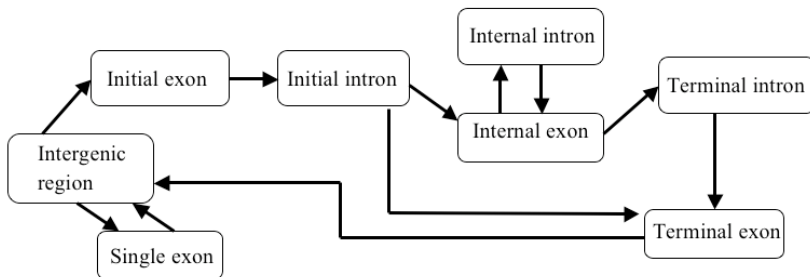
Biosynthèse de l'Isoleucine



Lacroix, V., Fernandes, C.G., Sagot, M.-F. (WABI 2005)

Niveau subcellulaire

HMM et analyse des isochores



Melo de Lima C., Guéguen L., Piau D., et Gautier C. (JOBIM 2005)

Niveau moléculaire

number of
parameters

12

9

8

6

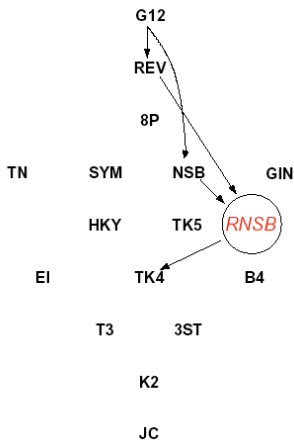
5

4

3

2

1

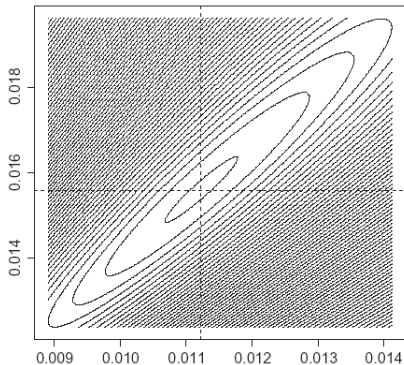


$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t)\mathbf{R}$$

$$\frac{dp_{ij}(t)}{dt} = \sum_{k=1}^4 p_{ik}(t)r_{kj}$$




$$\mathbf{X}(t) = \mathbf{P}'(t)\mathbf{X}(t=0)\mathbf{P}^T(t)$$

$$x_{ij}(t) = \sum_{k=1}^4 p'_{ik}(t)f_k p_{jk}(t)$$

Zagordi, O., Lobry, J.R. (2005) *Gene* 347:175-182

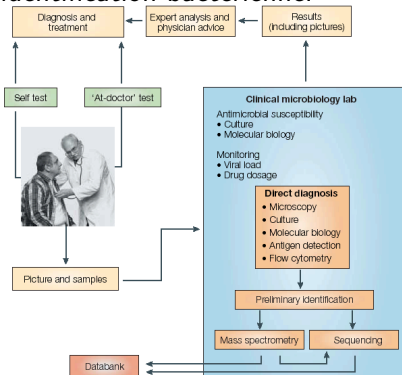
Productions informatiques

Bases de données, Logiciels et Webiciels

Nom	Nature	Remarques
ACNUC	Base de données	Base de données à la base de la plupart des bases génomiques du laboratoire, diffusion Web + Client serveur
ADE 4	Package 	logiciel implémentant les méthodes d'analyse de données dites vectorielles avec une référence particulière à l'écologie
ADEHabitat	Package 	Sélection d'habitat
BIBI	BD + identification bactérienne	Outil d'identification bactérienne à partir de séquences, application médicale, Web
CpGProD	Logiciel	Identification des îlots CpG
GEM	Base de données	Cartographie génomique comparée
Hogenome, Hovergen, Homolens	Base de données	Familles de gènes et phylogénies
Hoppsigen	Base de données	Rétropseudogènes
Lalnview	Webiciel+local	Comparaison de séquences
Oriloc	Webiciel+local	Origine de réplication
Philowin	C	Phylogénie moléculaire
RAP	Java	Réconciliation d'arbres phylogéniques
Sarment	Bibliothèque Python	HMM
SeqinR	Package 	Client ACNUC, analyse de séquences
SMILE	Logiciel C	Recherche de motifs complexes dans les séquences

Bases de données

BIBI : base données et identification bactérienne.



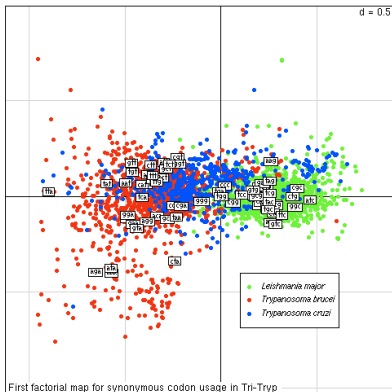
ProDom : base de familles de domaines protéiques.



Projet Genomicro : base de données d'alignement multiples de génomes complet.

CC-IN2P3

Logiciels : vers plus de vigueur hybride ?



Modes de diffusion :

- ① Données via accès Web
- ② Méthodes via logiciel à télécharger
- ③ Hybride : serveur de données + méthodes

Charif, D. *et al.* (2005)

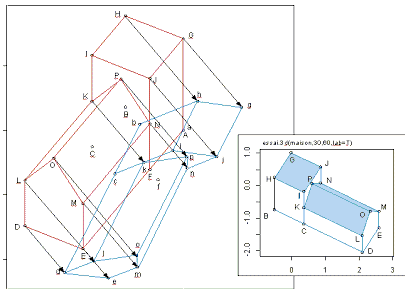
Bioinformatics, 21:545-547.

```
#####
library(ade4)
library(seqinr)
#####
```

Do it again

Forte interaction avec l'enseignement

Développement et diffusion de matériel pédagogique adapté à la biométrie.



DYNATICA®V1.0



Forte interaction avec l'enseignement

① Des thèmes communs.

Filière	Nom
BIM	Bioinformatique et Modélisation
BMC	Biologie Moléculaire et Cellulaire
EEME	Evolution Ecosystèmes Microbiologie Modélisation
MIV	Mathématiques et Informatique du Vivant

② Des outils de développement en commun.

Outil	Exemple	Recherche	Enseignement
Travail collaboratif	CVS	Gestion du code source des programmes	Gestion du code source des documents pédagogiques
Programmation littéraire	Sweave	Articles reproductibles	Documents pédagogiques à jour
Contrôles d'intégrité	Cron-jobs	Contrôle des packages	Contrôle automatique des documents pédagogiques

Axe transversal : méthodologie

Laboratoire de Biométrie et Biologie Evolutive (UMR 5558);
CNRS; Univ. Lyon 1, 43 bd 11 nov, 69622, Villeurbanne Cedex,
France.

24 Nov. 2005

