Multivariate Data Analysis in Microbial Ecology
New Skin for the old Ceremony

Jean Thioulouse

UMR 5558 CNRS « Biométrie, Biologie Évolutive »
CNRS – University of Lyon - France

Headlines

- Topic: Environmetrics and Ecology
  - ✓ descriptive exploratory multivariate data analysis ("Geometric Data Analysis")
  - ✓ ade4 and ade4TkGUI packages
  - ✓ case studies
- The EcoMic – RMQS project
- Mycorrhizal symbiosis in tropical soils

The EcoMic – RMQS project

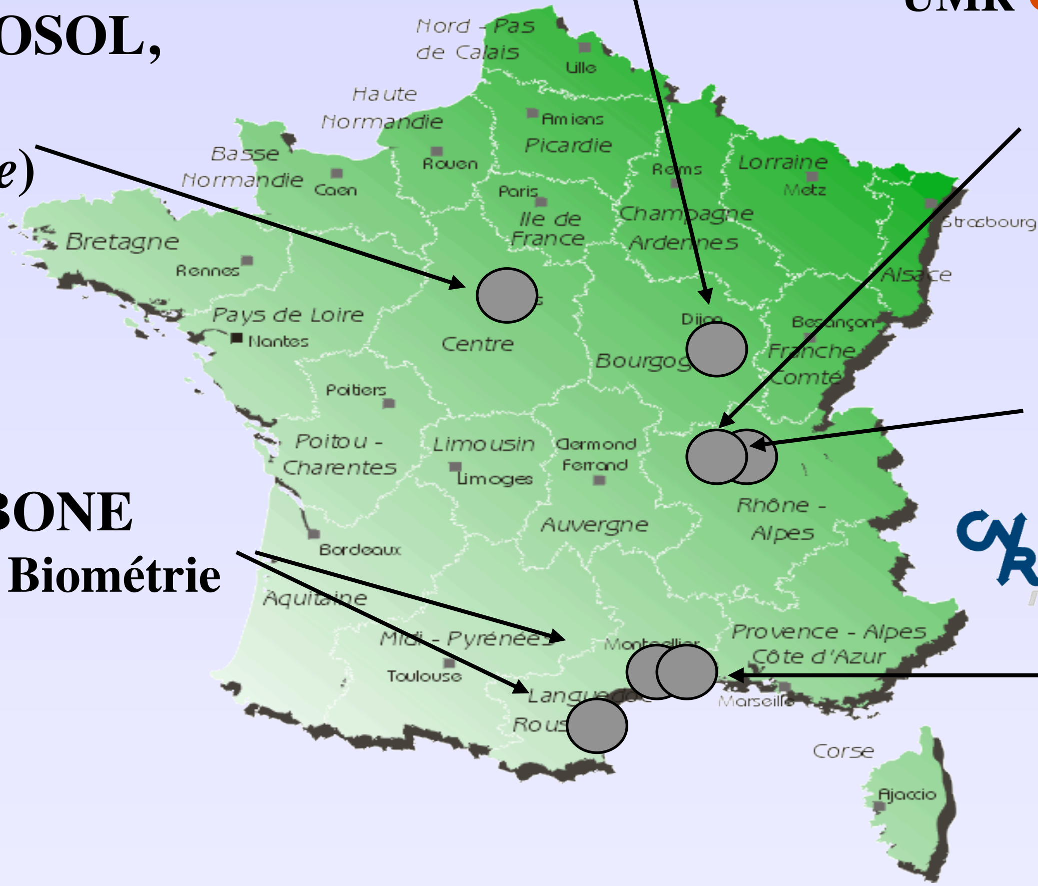
- Analyse the relationships between soil microbial molecular diversity and environmental factors at the regional and national scales in France.

Molecular diversity of soil bacterial communities ⟷ Environmental factors

Coordinator : L Ranjard, UMR INRA/ U-Bourgogne
Microbiologie et Géochimie des Sols, Dijon
(*Microbial Ecology*)

Unité INRA INFOSOL, Orléans
(*Soil Science*)

UMR CNRS Génomique Microbienne Environnementale, Lyon
(*Microbial Ecology*)



UMR CNRS/UCBL Biométrie Biologie Evolutive, Lyon
(*Data analysis*)

LBE INRA-NARBONE
Analyse des Systèmes et Biométrie
(*Modelling*)

DREAM Unit, CEFE-CNRS Montpellier
(*Soil Science*)

Multidisciplinary

(Soil science, Microbial ecology, Modelling, Data analysis)

Multi-institutionnal

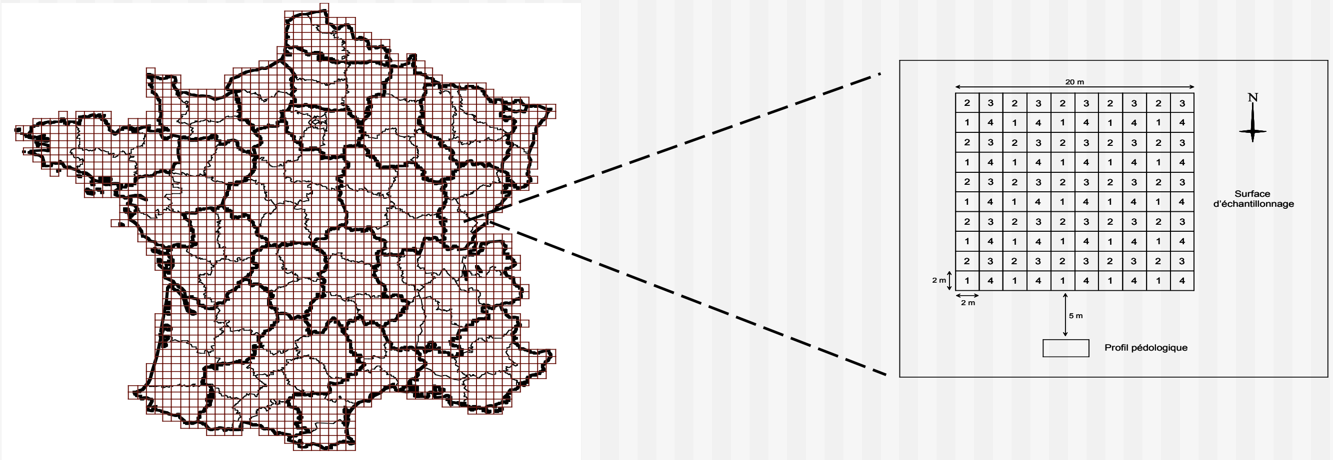
(INRA, CNRS, Universities)

The EcoMic – RMQS project

- Large (2M€) ANR project on Microbial Ecology of French soils
- Microbial diversity in soil
  - ✓ Evaluate beta diversity
  - ✓ Processes generating and maintaining this diversity
  - ✓ Large spatial scale (France)
  - ✓ Molecular tools (PCR, DNA fingerprints, DNA μarrays)
- Based on the **RMQS soil library**

The RMQS

- Soil Quality Measure Network
- Started in 2002 by Infosol - INRA Orleans
- Square sampling grid over all France 16 x 16 km
- 2200 sampling points, finished in 2009
- Renewed every 10 years.



The RMQS

Many parameters are measured:

- Physico-chemical parameters (pedology)
    - ✓ granulometry, pH
    - ✓ C, N, Ca, Na, heavy metals, etc.
- Vegetation cover, lanscape, agricultural practices, etc.
- Molecular data (DNA extraction from raw soil samples)

Six regions

- Based on vegetation, landscape, climate, pedology, and *available samples* *(578)*

Region 1: North, intermediate

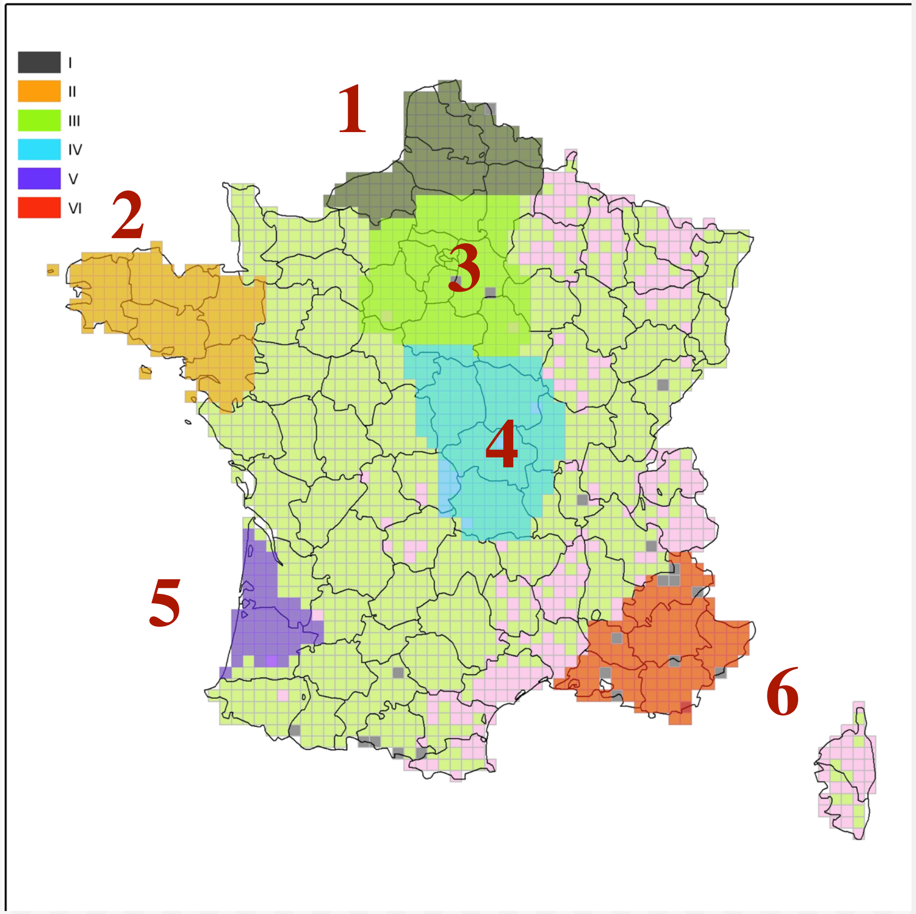
Region 2: Brittany, low diversity

Region 3: grand Paris, highly urbanized

Region 4: Center, intermediate

Region 5: Landes, very low diversity, sand dunes and pine forests

Region 6: South Alps, highest diversity, mountains, contrasted climate



Molecular data

- RISA: Ribosomal Intergene Spacer Analysis
  - ✓ length polymorphism of the intergene sequence between the large and small ribosomal subunit genes
  - ✓ no information on taxonomic level (OTU: Operational Taxonomical Unit)
  - ✓ data already available for almost 1000 sites
- Sequencer data must be processed before analysis
  - ✓ prepRISA package: rectangular data tables (sites x OTU)
  - ✓ hundreds of OTU
  - ✓ typical data table size ≈ 2200 x 500

Molecular data

- DNA *μ*arrays
  - ✓ probes can be specific of particular taxonomic levels
  - ✓ data not yet available
  - ✓ thousands of probes
  - ✓ typical data table size ≈ 2200 x 10 000

Microbiogeography questions

- "Everything is everywhere, *but*, the environment selects." Baas Becking, 1934.
- "Are microbial communities a black box with no spatial structure or, like macroorganisms do they exhibit a particular distribution with predictable aggregates patterns from local to regional scales ? " Horner-Devine et al., *Nature*, 2004.
- "Identify the environmental factors (edaphic, climatic, anthropogenic…) which exert the strongest influences on microbial communities in nature." Martiny et al., *Nature Reviews*, 2006.

Biological objectives

- Inventory of bacterial diversity in French soils
- ***Spatial components*** of this diversity and ecological models to explain it (***species-area relationship***)
- Mechanisms determining this diversity (pedology, climate, vegetation, etc.)
- Quantify the impact of human activities (agriculture, industrial sites, wastes)
- Microbiological markers of soil evolution in various ecological situations

Species - area relationship

- S = number of species, A = area under study

$$S = CA^{z}$$

- When using RISA, OTU ≠ species

  ✓ OTU_A and OTU_a number of OTU in areas A and a

$$OTU_a = OTU_A \left(\frac{a}{A}\right)^{z}$$

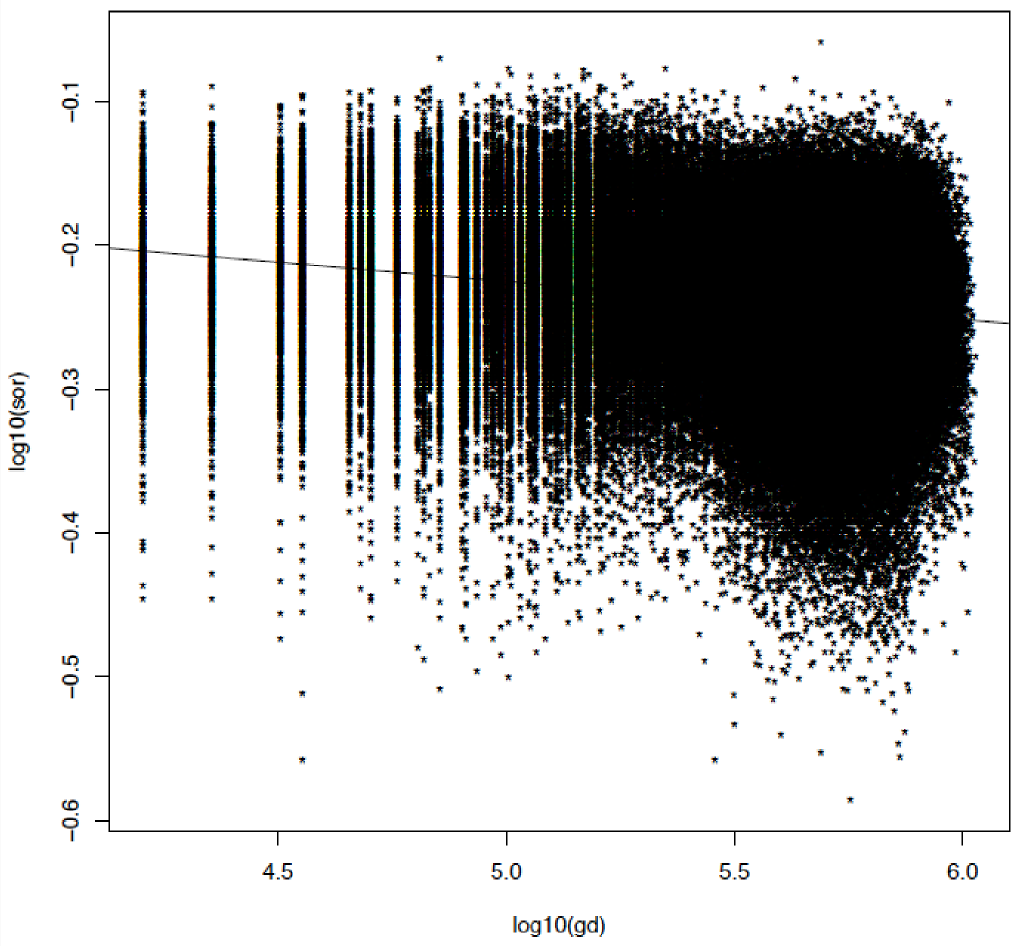
  ✓ Sørensen index for 2 samples at distances D and d

$$\chi_d = \chi_D \left(\frac{d}{D}\right)^{-2z}$$

  ✓ Green *et al.* 2004, *Nature*, **432**, 747-750

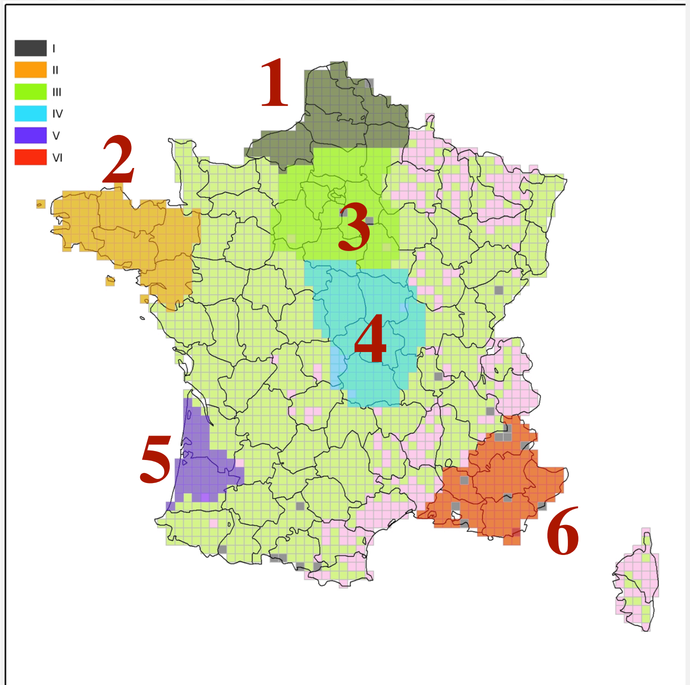
Species - area relationship

- Computations: R (packages: vegan, labdsv)



Species - area relationship

Region 1	Slope = **-0.00590**	p = **0.031**
Region 2	Slope = **-0.00022**	p = **0.002**
Region 3	Slope = **+0.00494**	p = **0.018**
Region 4	Slope = **-0.00683**	p = **0.008**
Region 5	Slope = **-0.00416**	p = **0.690**
Region 6	Slope = **-0.01280**	p = **0.000**
Total	Slope = **-0.02657**	p = **0.000**



Slopes are negative: **diversity increases**

Region 5: very homogeneous (p-value NS)

Region 3: slope is positive (urbanized)

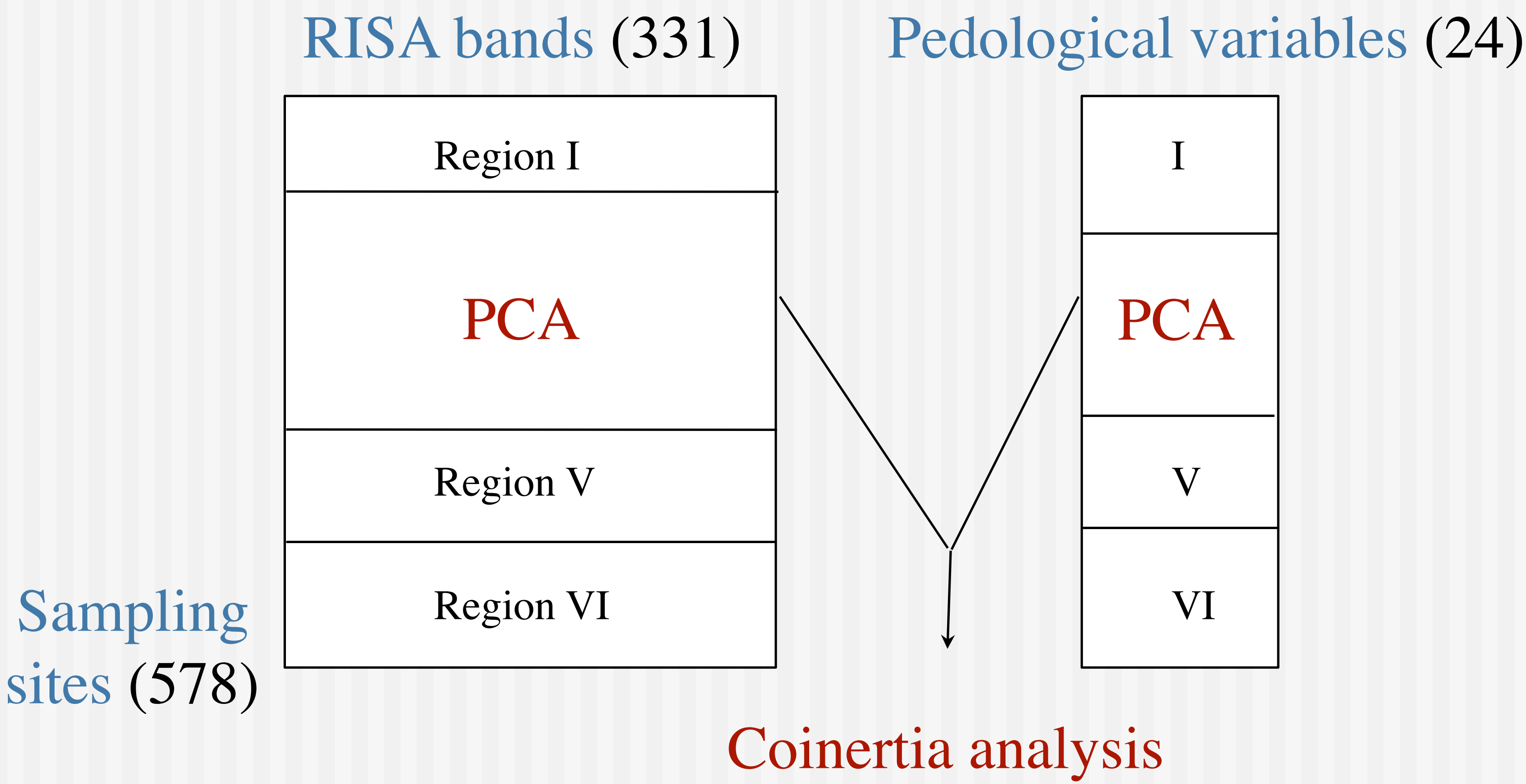
Region 6: high landscape diversity

Region 2: low diversity

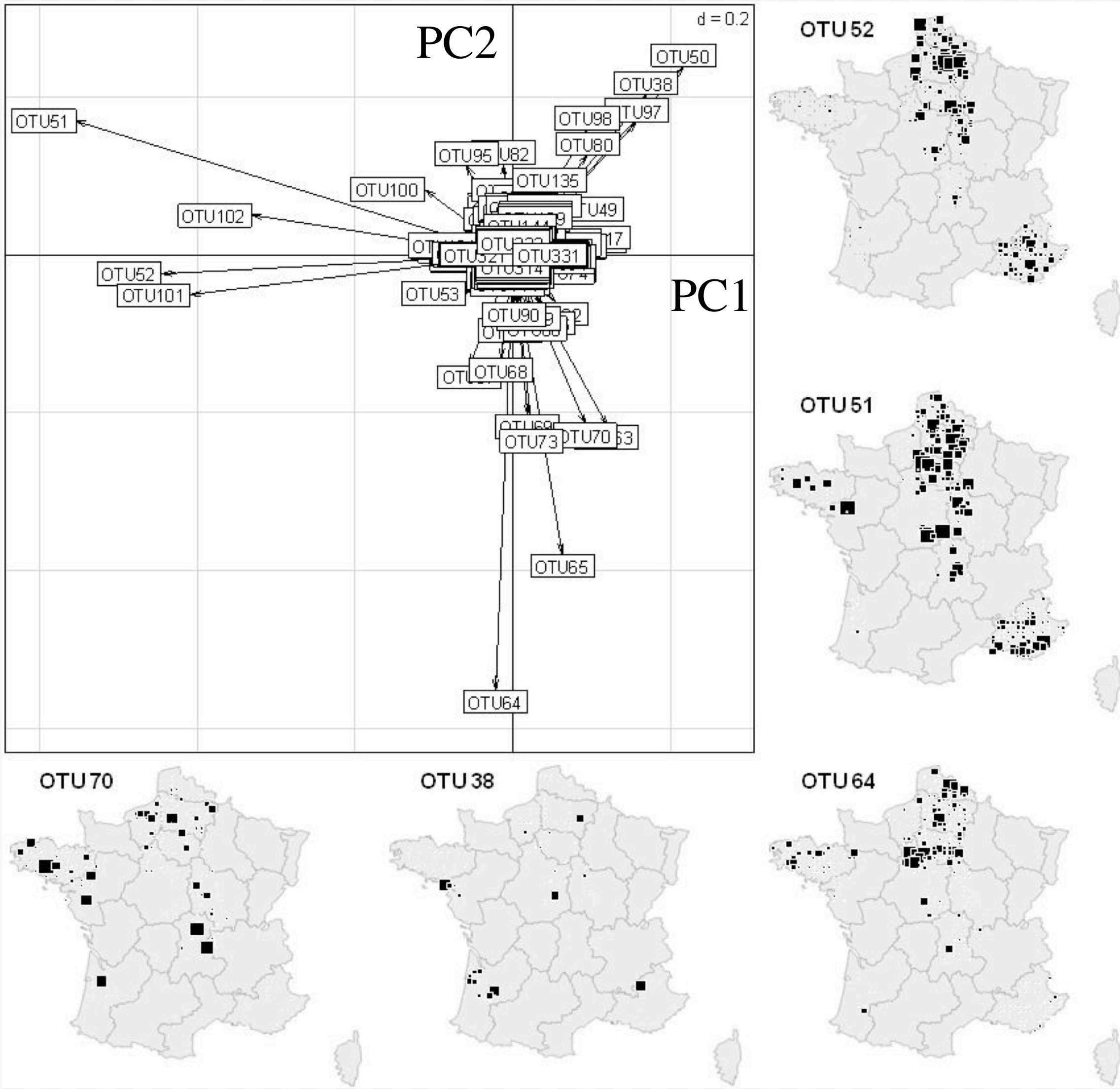
Regions 1 and 4: intermediate

Multivariate analysis: data sets

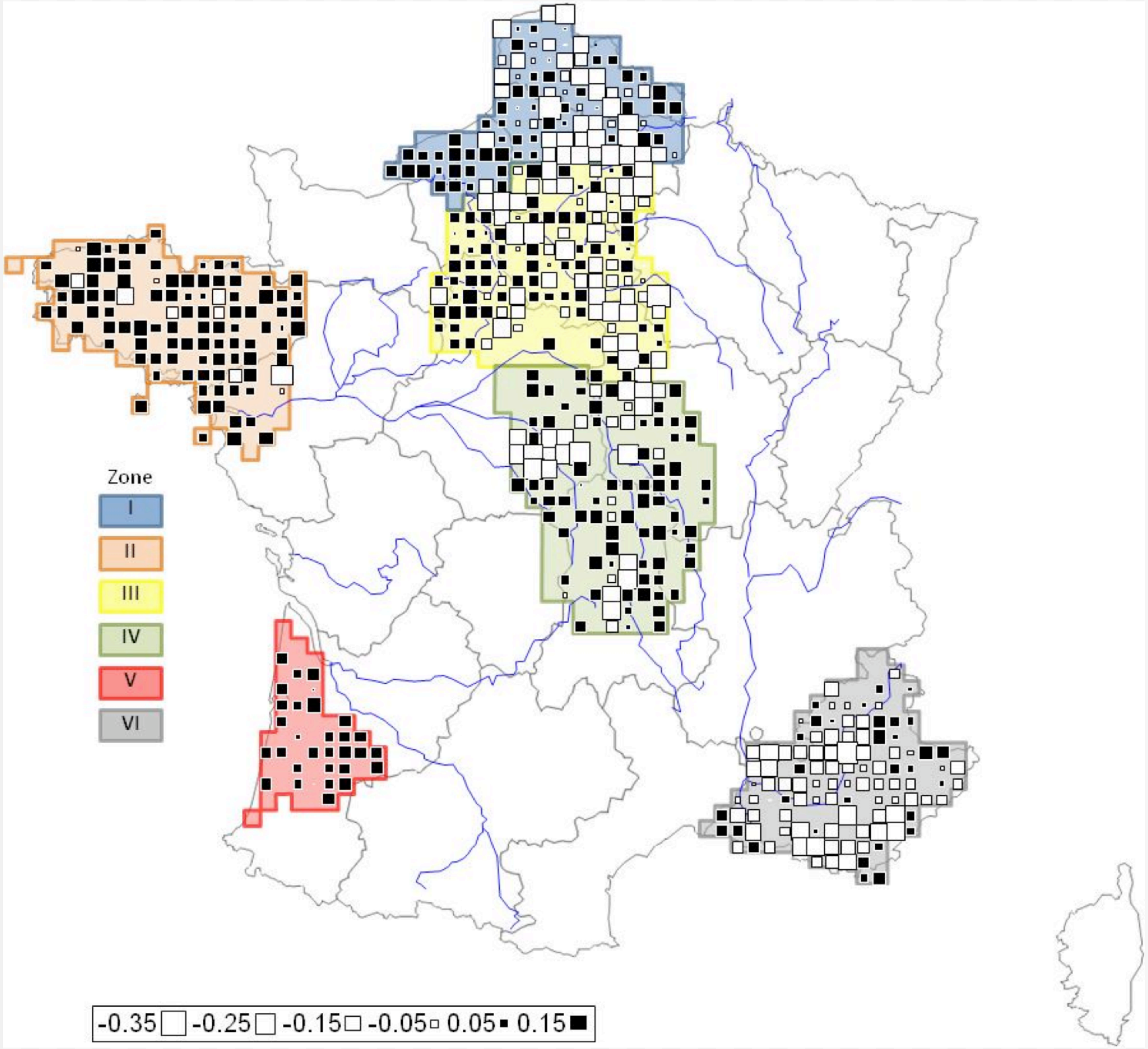
- Computations: R (prepRISA, ade4 + ade4TkGUI)



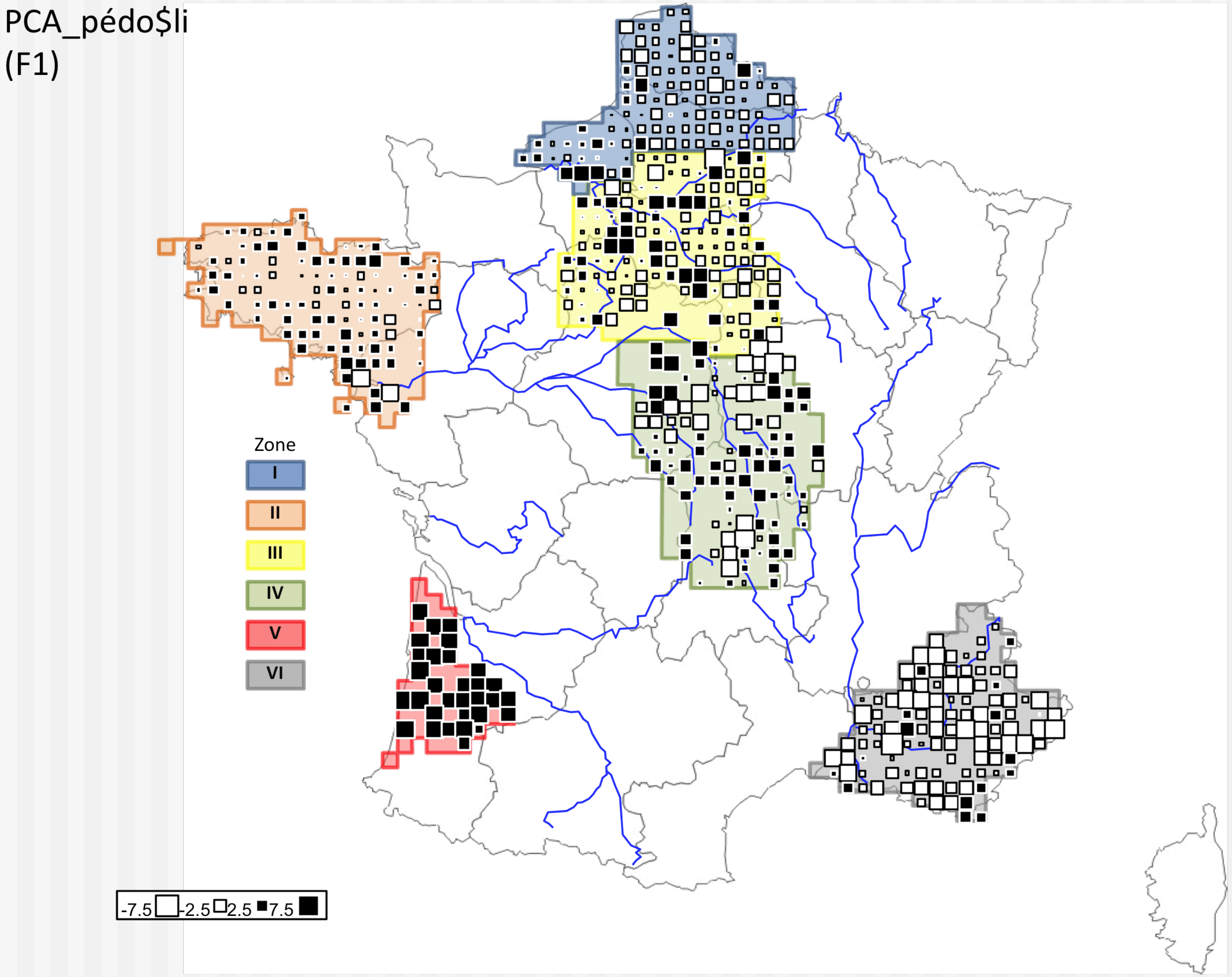
PCA of RISA data: OTU distrib.



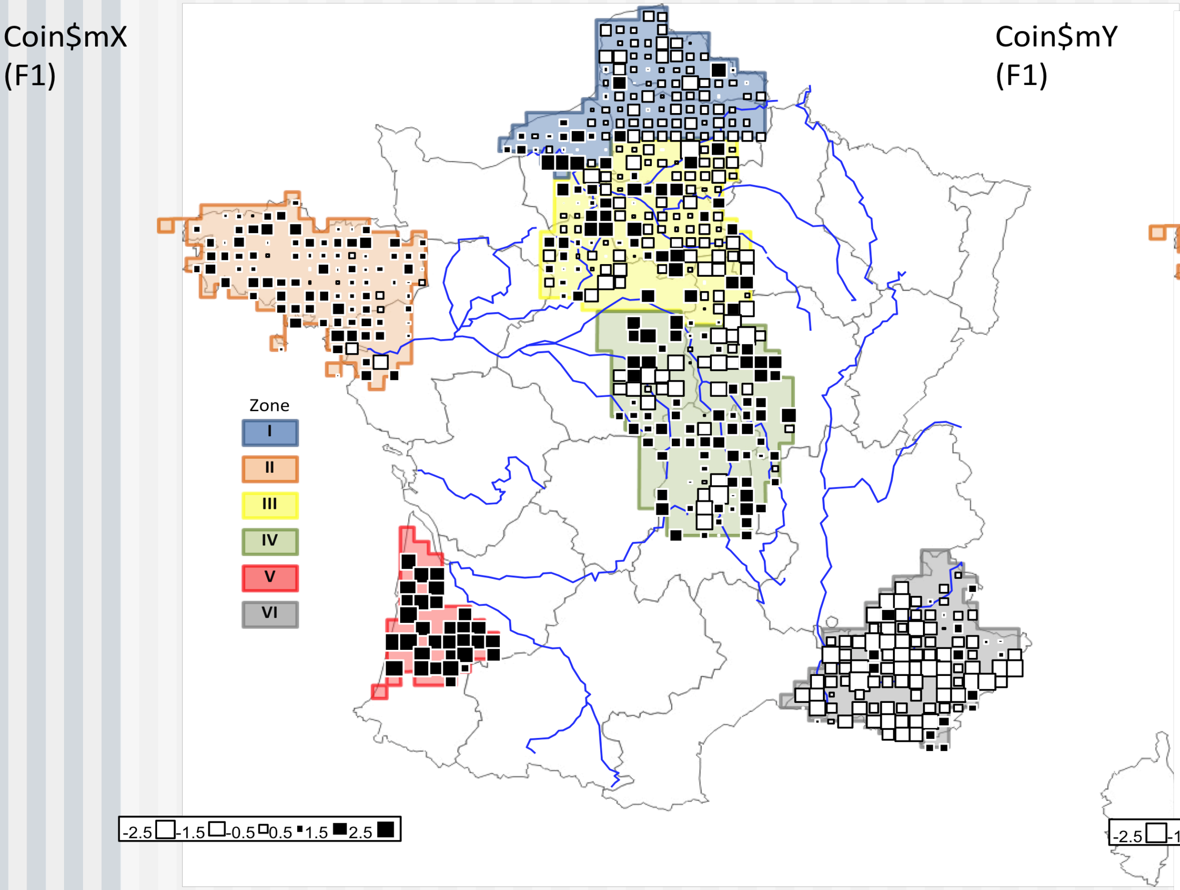
PCA of RISA data (PC1)

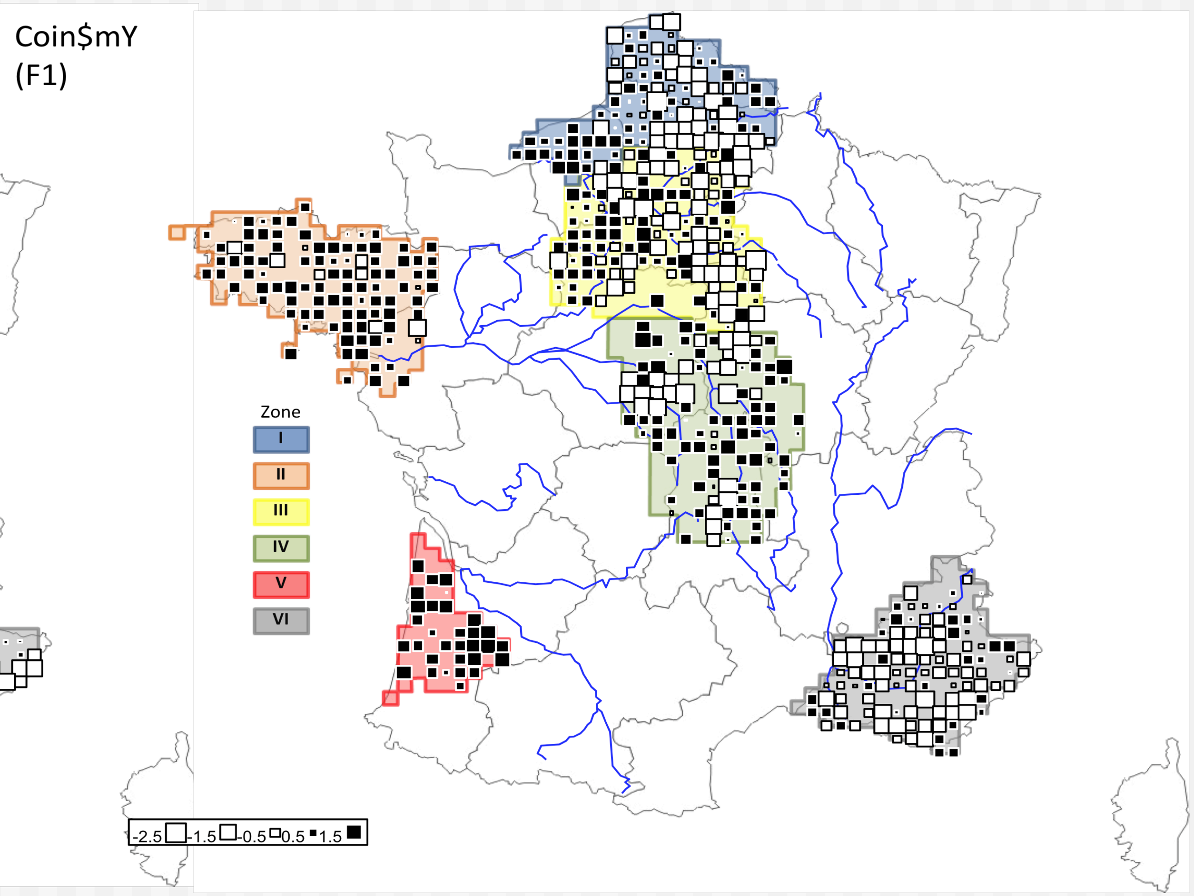


PCA of pedological data (PC1)



Coinertia analysis (pedo/RISA)

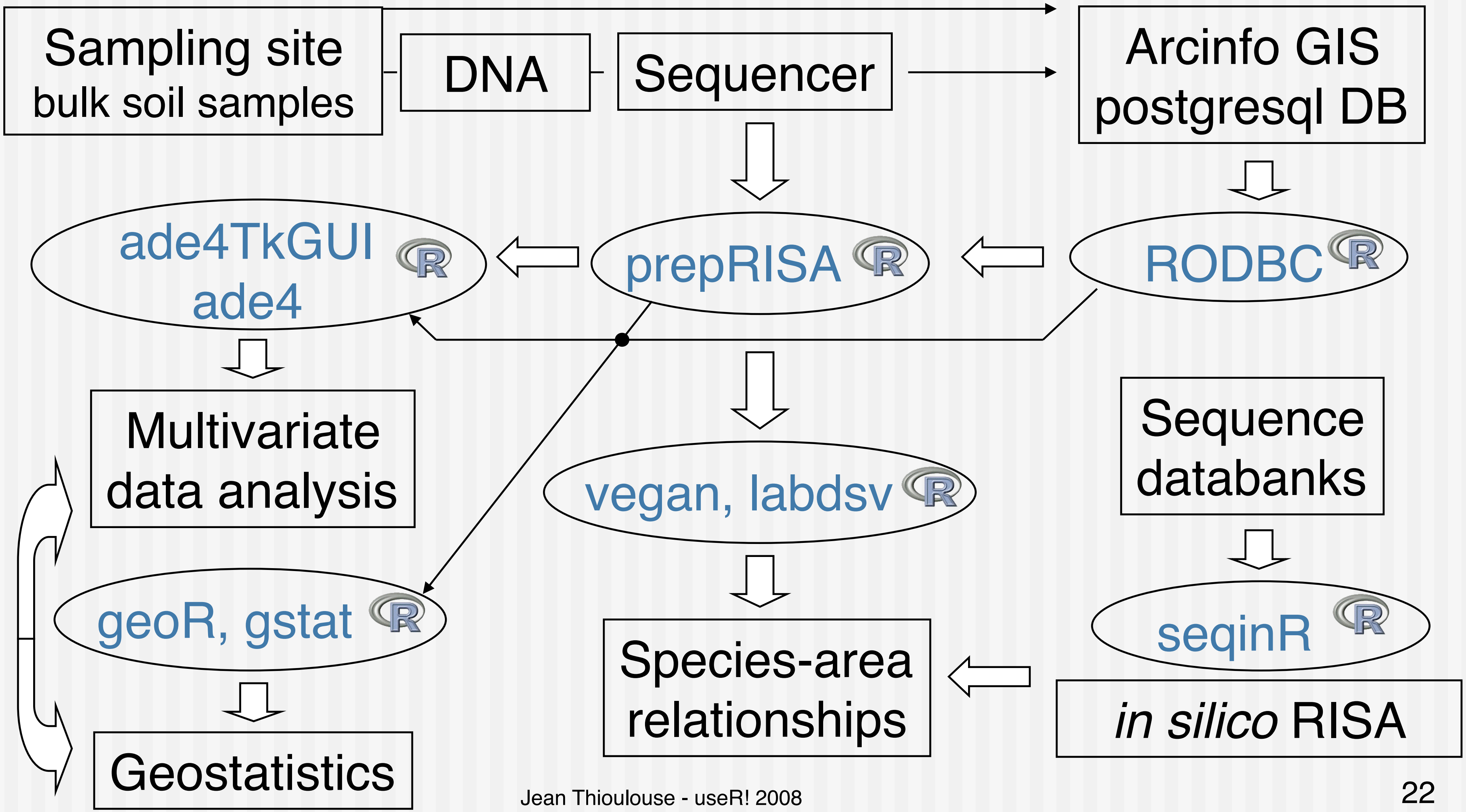




Biological interpretations

- The spatial structures of physico-chemical pedological variables and RISA data are very similar, and this similarity exists at both regional and national scales (within and between regions)
- Ecological processes responsible of spatial structures in animals and plants (differenciation, extinction, endemism) either do not exist for bacterias, or may be masked by bacteria characteristics, such as dispersion capacities, stress resistance, or short generation time.
- "everything is everywhere, but, the environment selects"
- These results are for RISA OTU, which are mostly unknown. But (hopefully) more to come with DNA μarrays...

RISA methodological flow chart



Mycorrhizal symbiosis in tropical soils

- Soil is a very complex environment, with many interactions between plants, mycorrhizal fungi, soil bacterial communities, and abiotic factors
- Research project started in 2000 in collaboration with several IRD research labs in Africa:
  - ✓ Dakar (Senegal)
  - ✓ Ouagadougou (Burkina Faso)
  - ✓ Marrakech (Morocco)
  - ✓ Antananarivo (Madagascar)

Mycorrhizal symbiosis in tropical soils

- Relationships between plants, mycorrhizal symbiosis, and the soil bacterial microflora
- Review of 15 papers published since 2005 on this topic in microbial ecology journals
  - ✓ Effects of mycorrhizal symbiosis
  - ✓ Nurse plants
  - ✓ Termite mound powder amendment

Mycorrhizal symbiosis in tropical soils

- Data used to study the diversity of soil bacteria
  - ✓ RISA and DNA *μ*arrays are too complex to be used
  - ✓ Catabolic profiles / SIR (substrate induced response)

    profile = measure of CO_2 production for ≈ 30 substrates

    functional diversity *vs.* taxonomic diversity
  - ✓ DNA fingerprint: DGGE

Mycorrhizal symbiosis in tropical soils

- Statistical data analysis methods (ade4 R)

  ✓ BGA (between group analysis): robust alternative to discriminant analysis to separate groups. PCA on group means, with projection of original data

  ✓ Coinertia analysis: analyse the relationships between two data tables. Robust alternative to Canonical Analysis or Canonical Correspondence Analysis

  ✓ both methods allow the use of low numbers of samples as compared to the number of variables

  ✓ permutation test

Mycorrhizal symbiosis in tropical soils

- Statistical data analysis methods
  - ✓ Cluster analysis on DGGE fingerprints (hclust R)
  - ✓ Linear model for various hypothesis test (lm, nlme R)

Results

- Effects of mycorrhizal symbiosis
  - ✓ improves plant growth (*Cupressus atlantica, Acacia holosericea, Acacia seyal, Uapaca bojeri*)
  - ✓ improves P solubilization and assimilation
  - ✓ improves seedling survival (*C.a.*) in degraded soils
  - ✓ improves the functional diversity of soil microflora
  - ✓ counterbalance the negative effect of an exotic plant species on the structure and functioning of soil bacterial communities

Results

- Nurse plant effects
  - ✓ *Lavandula* species (*L. stoechas, L. dentata, L. multifida*) act as nurse plants for *C. Atlantica*
  - ✓ improve plant growth
  - ✓ improve mycorrizal soil infectivity
  - ✓ interactions with the functional diversity of soil bacteria
  - ✓ *Thymus satureioides* also improves plant growth and functional diversity

Results

- Termite mound powder amendment effect
  - ✓ improves mycorrhizal symbiosis and plant growth of *A. holosericea* and *A. seyal* (MHB: mycorriza helper bacteria)
  - ✓ improves nutrient supplies (effect on soil microflora)
  - ✓ can be used for biological control of phytoparasite *Striga hermontica* (effect on soil microflora)
  - ✓ decreases Cd toxicity and improves its accumulation in plants (sorghum)

Result examples: BGA on SIR

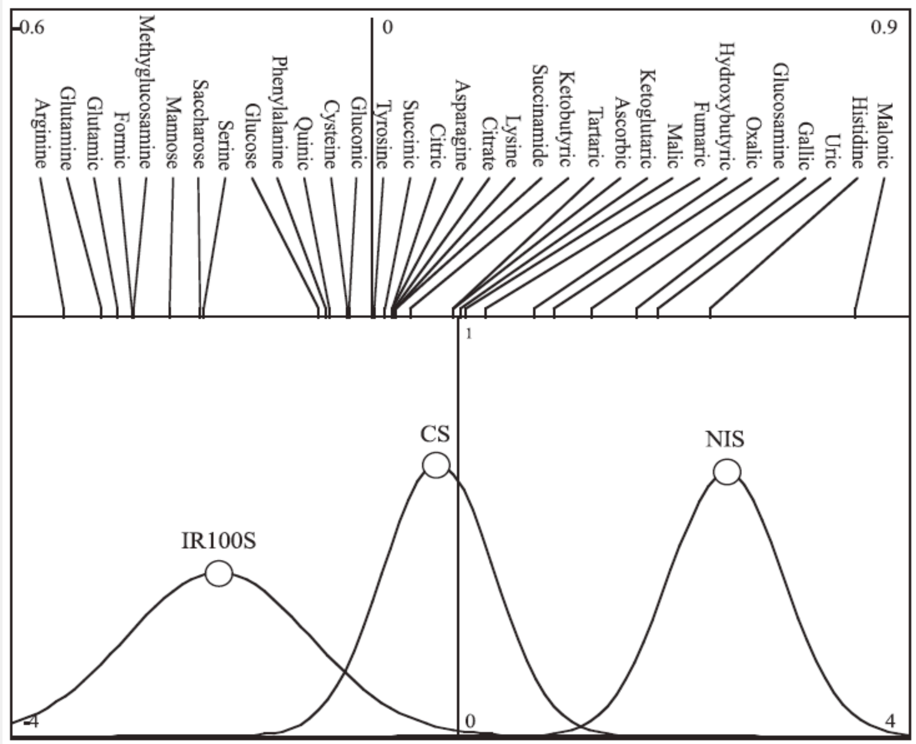
Graphical display (PC2 biplot) of BGA on catabolic profiles.

Up: scores of the 33 substrates. Down: the three Gauss curves represent the mean and variance of the scores of soil samples (three treatments).

CS: crop soil

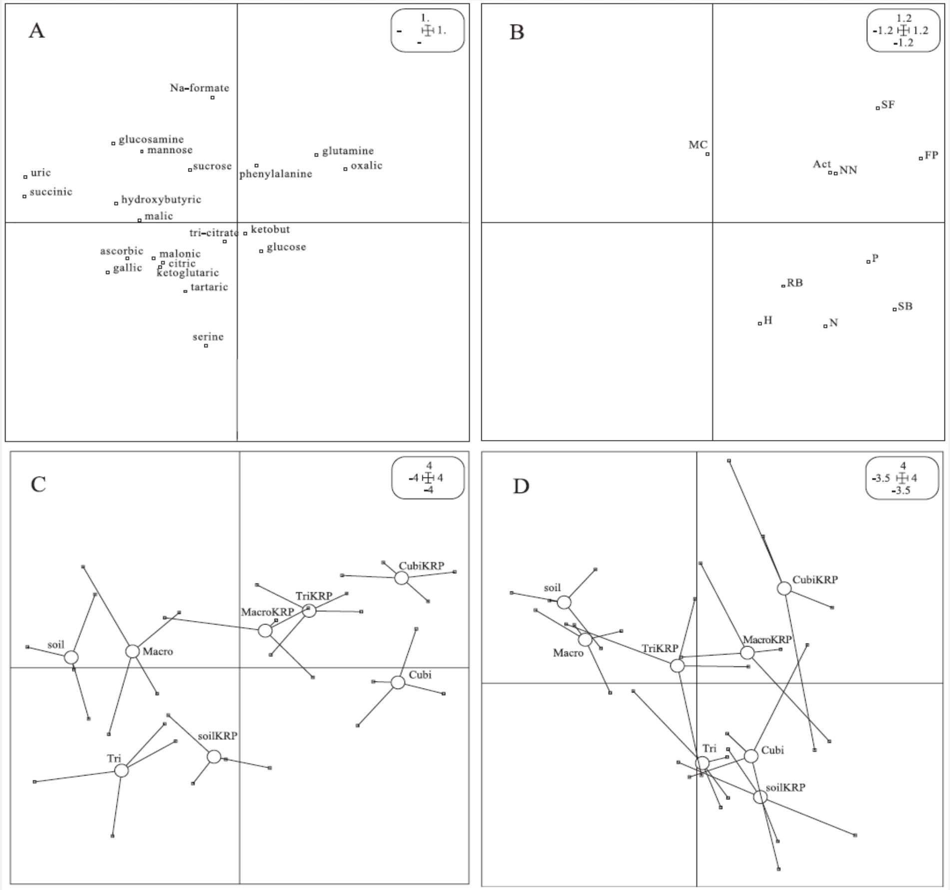
NIS: soil from uninoculated trees

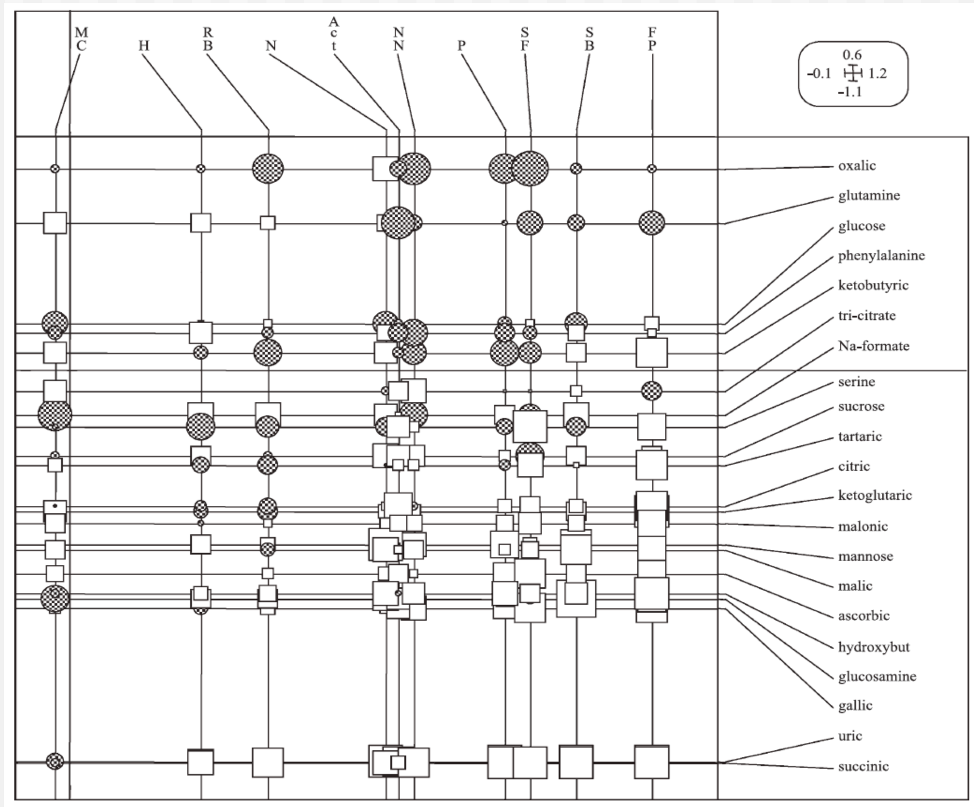
IR100S: soil from *P. albus* IR100 fungi inoculated trees



Result examples: Coinertia on SIR

Graphical display of Coinertia analysis between catabolic profiles and plant growth parameters. Left: F1xF2 factor maps. Right: cross-covariances





Conclusion

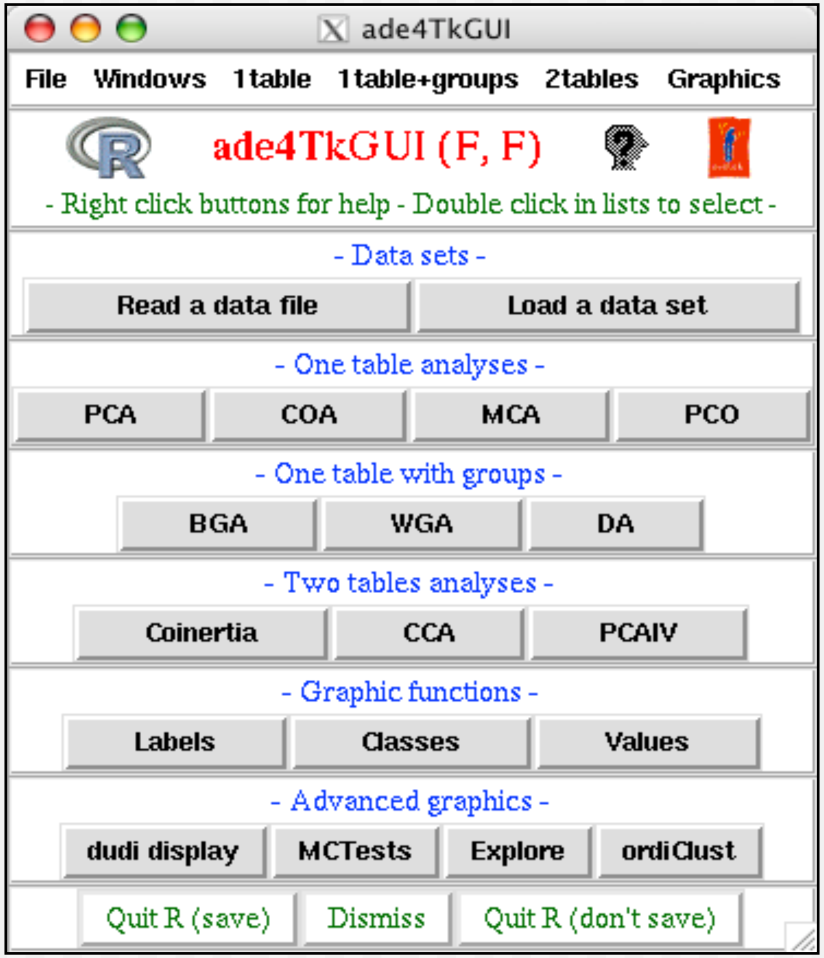
- R in the Biometry lab
  - ✓ Has become an indispensable tool in just a few years
  - ✓ Teaching (all levels)
  - ✓ 2 mailing lists: « rpackdev » and « rpourlesnuls »
  - ✓ 6 packages (on CRAN): ade4 and spin-off packages
  - ✓ ade4TkGUI, adehabitat, adegenet
  - ✓ seqinR, nlstools

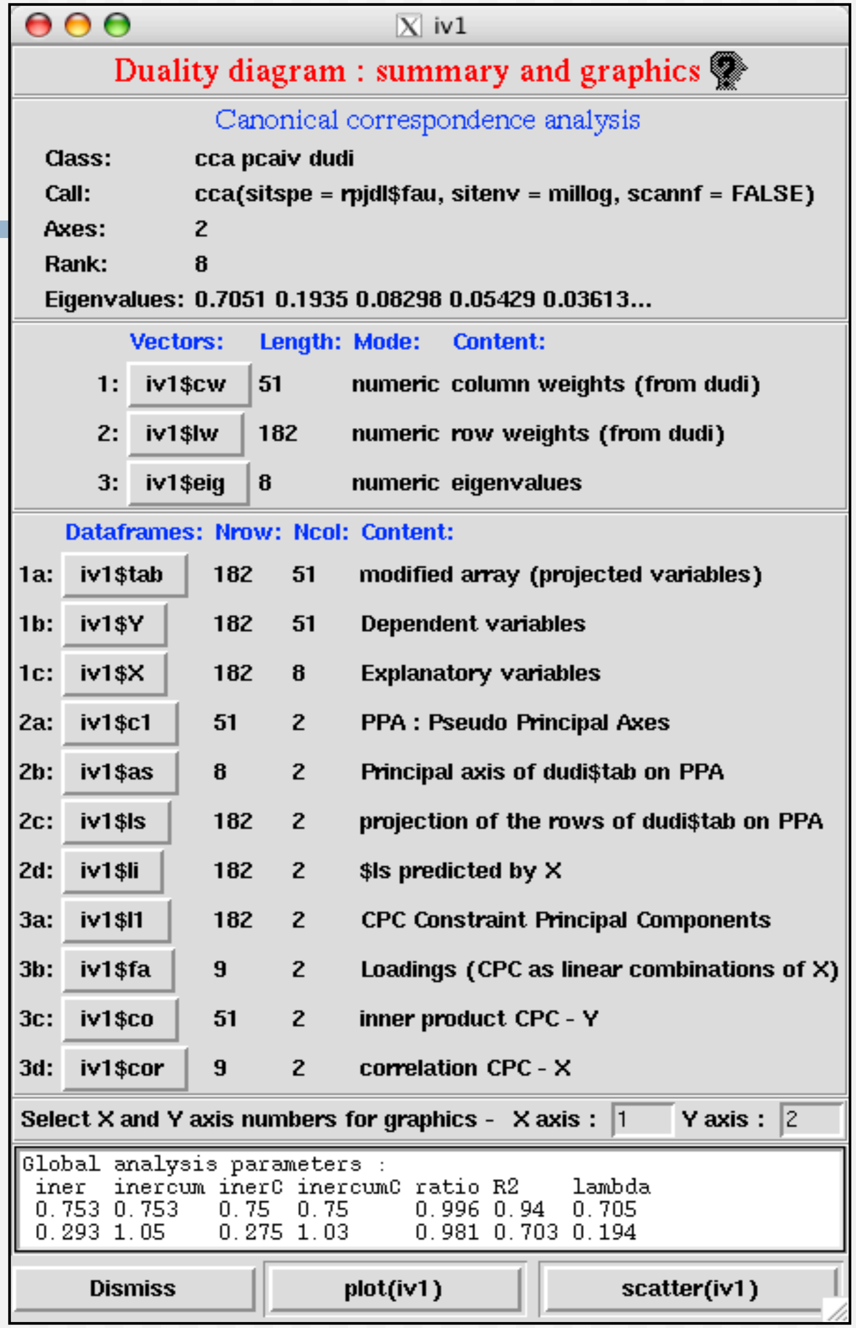
Conclusion

- Need for GUIs
  - ✓ For teaching
  - ✓ For biologists
  - ✓ ade4TkGUI screenshots

Conclusion

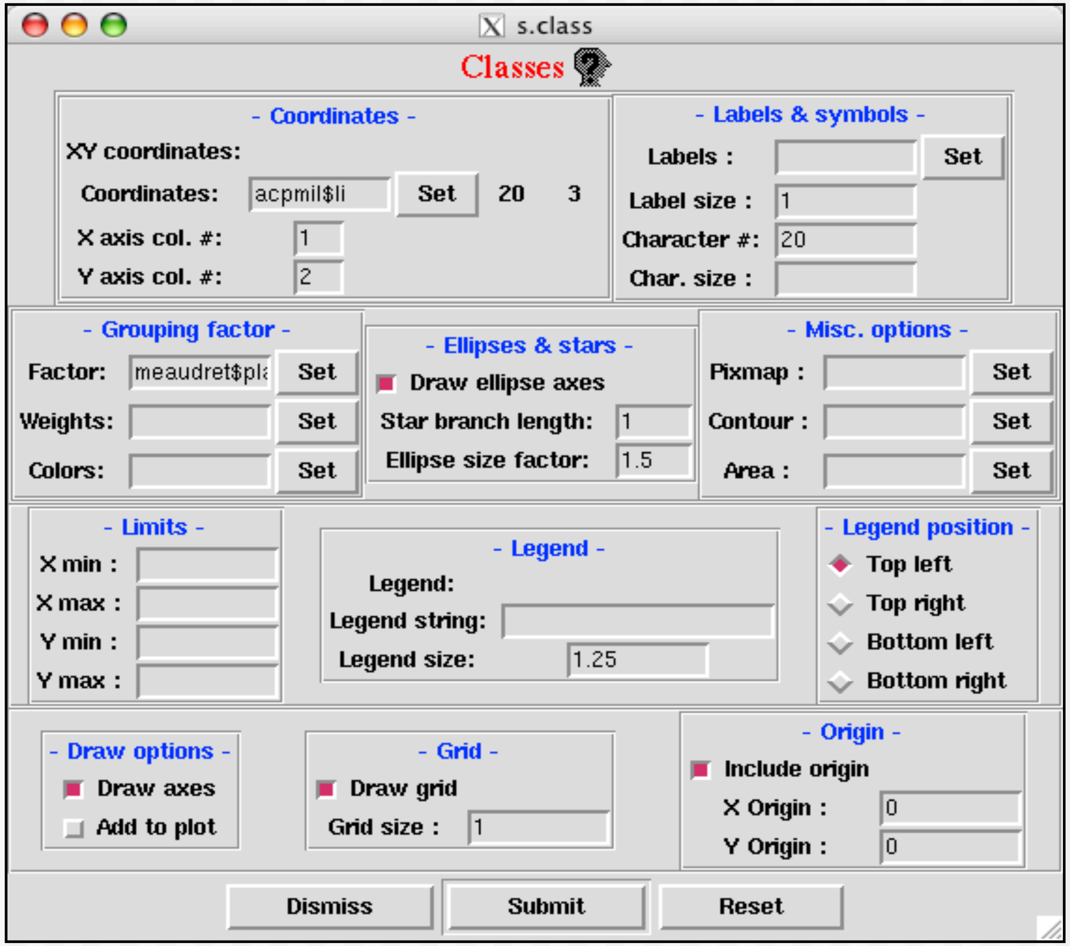
Synthetic display

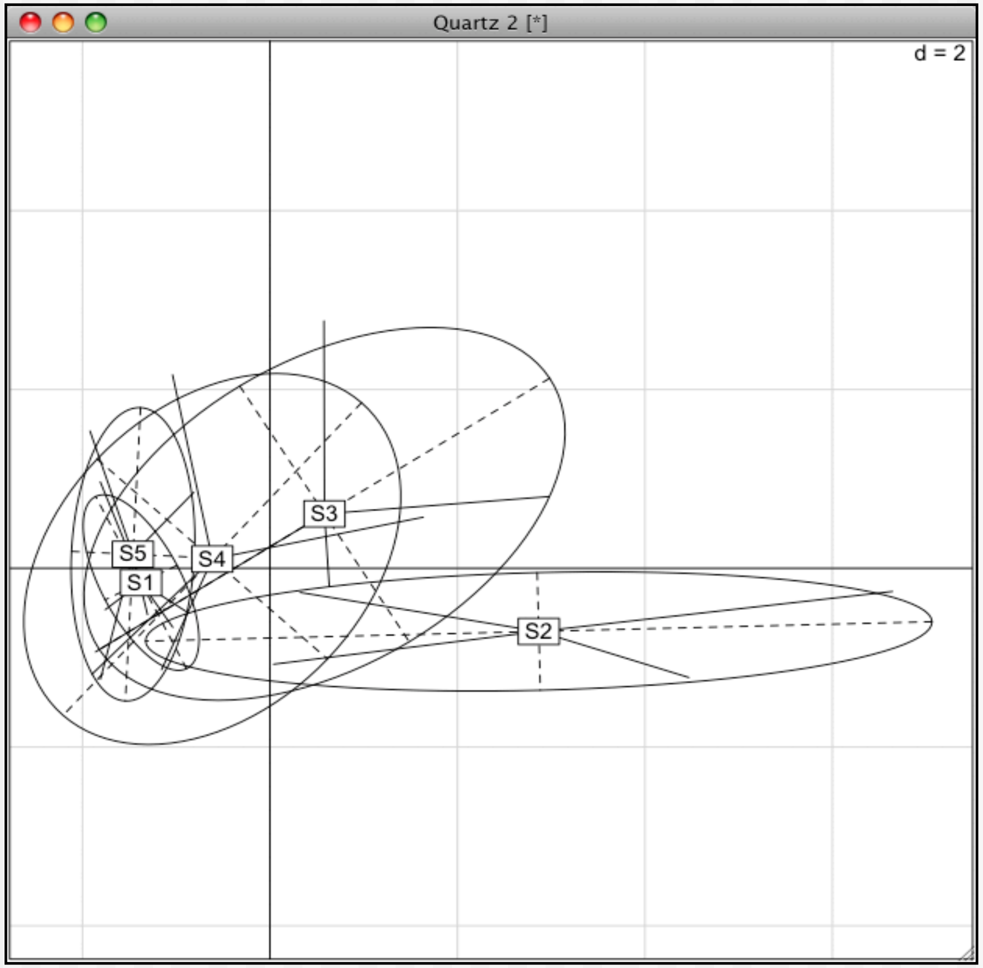




Conclusion

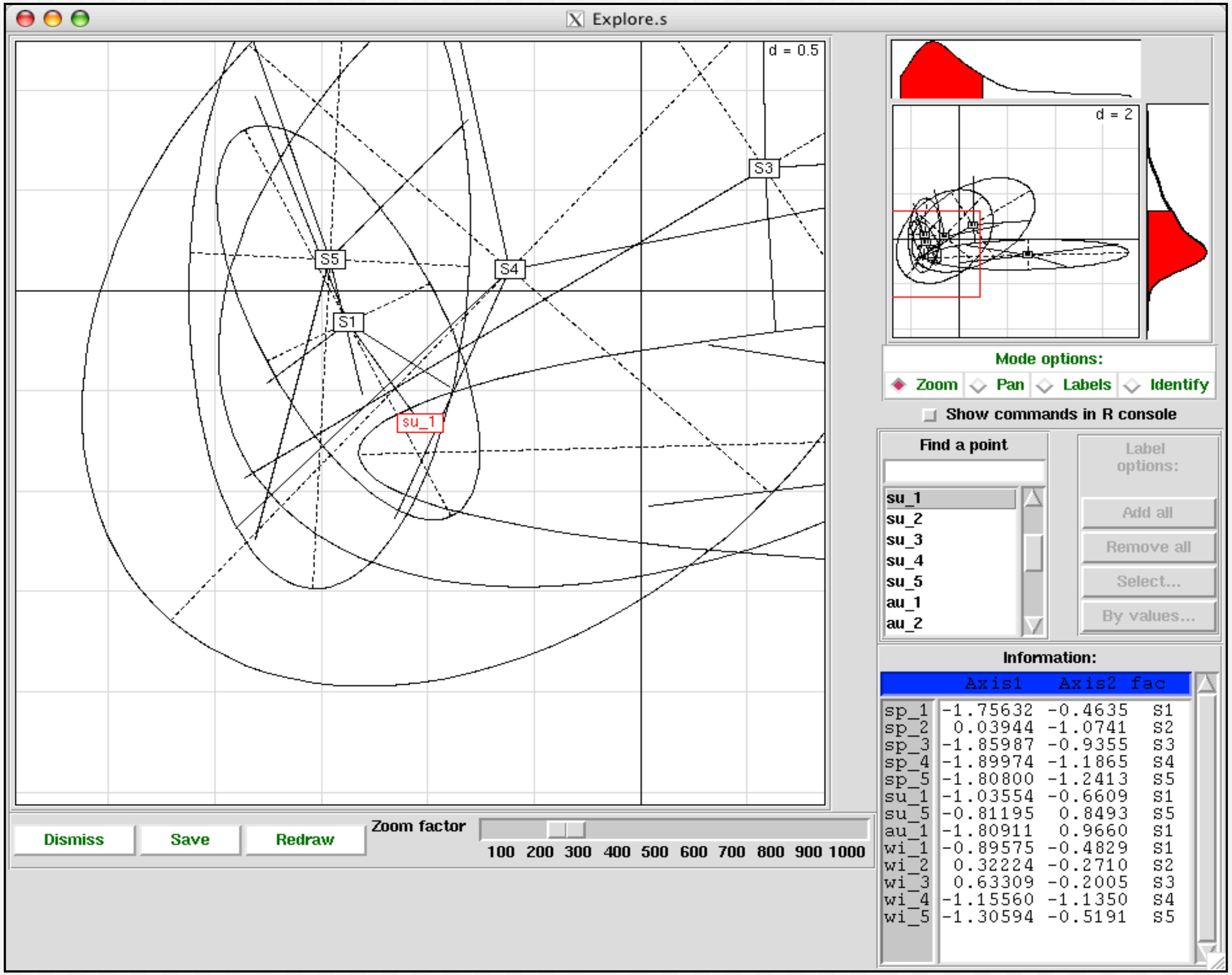
Complex graphs





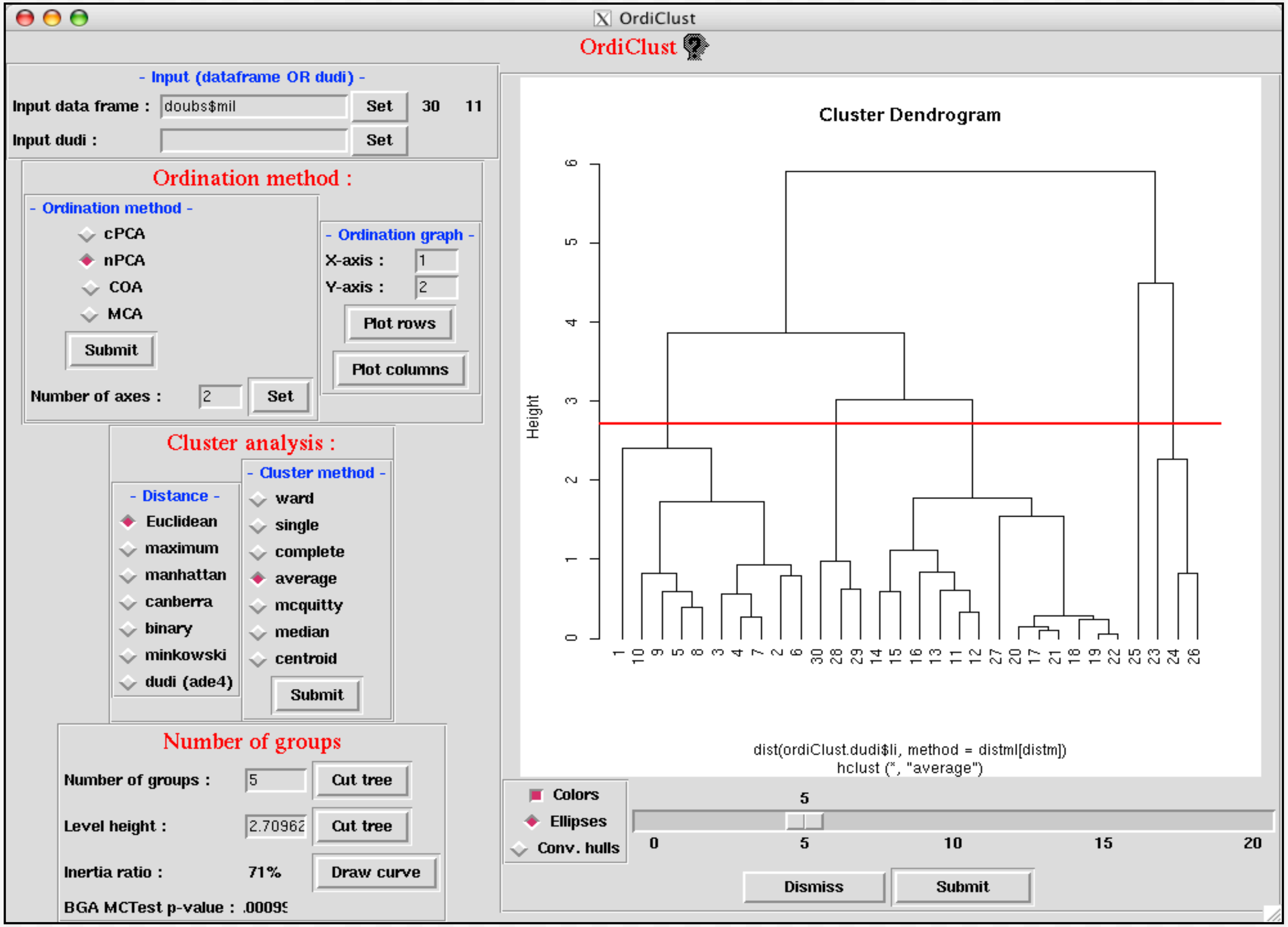
Conclusion

Interactive factor map



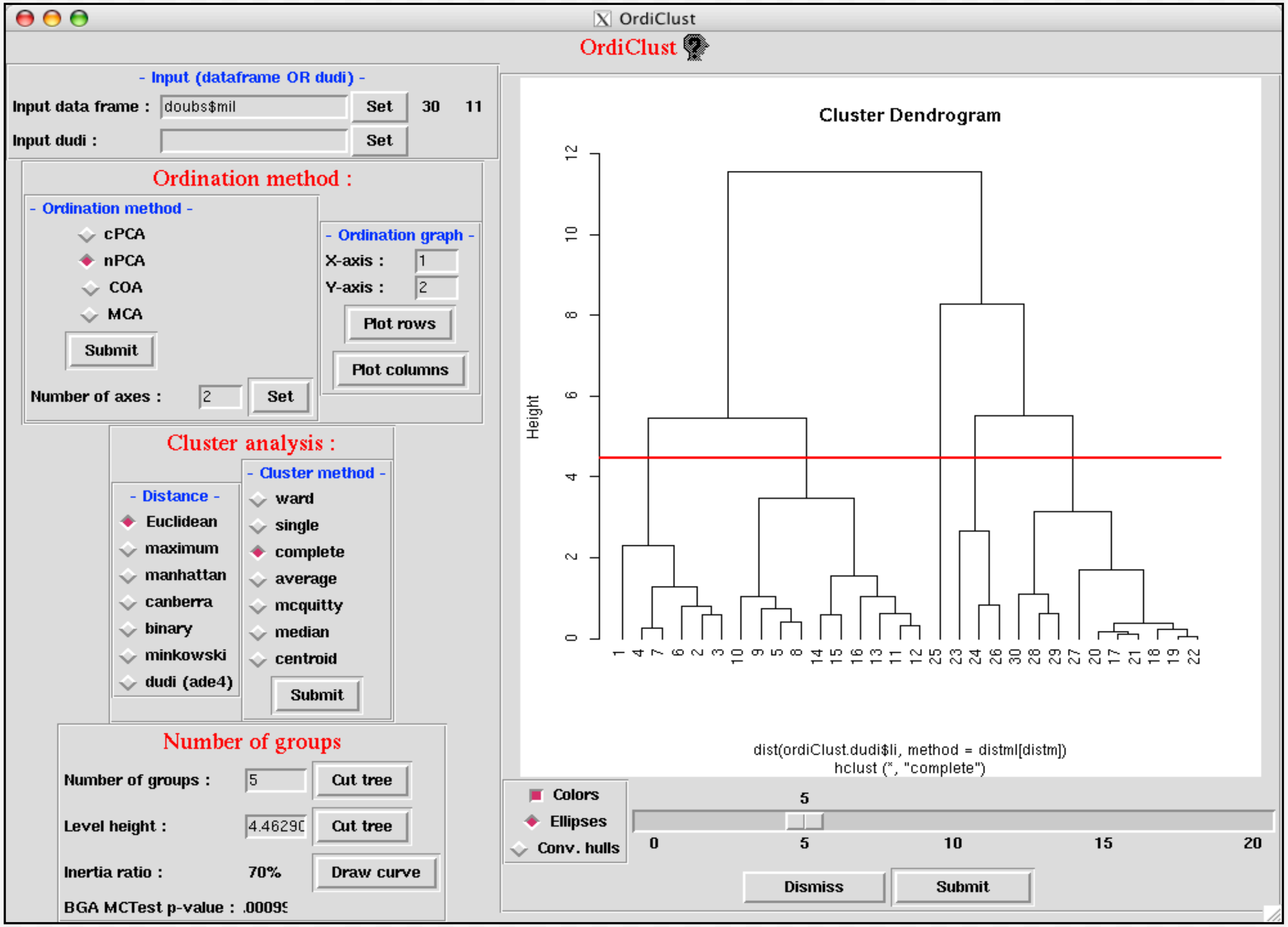
Conclusion

Interactive data analysis



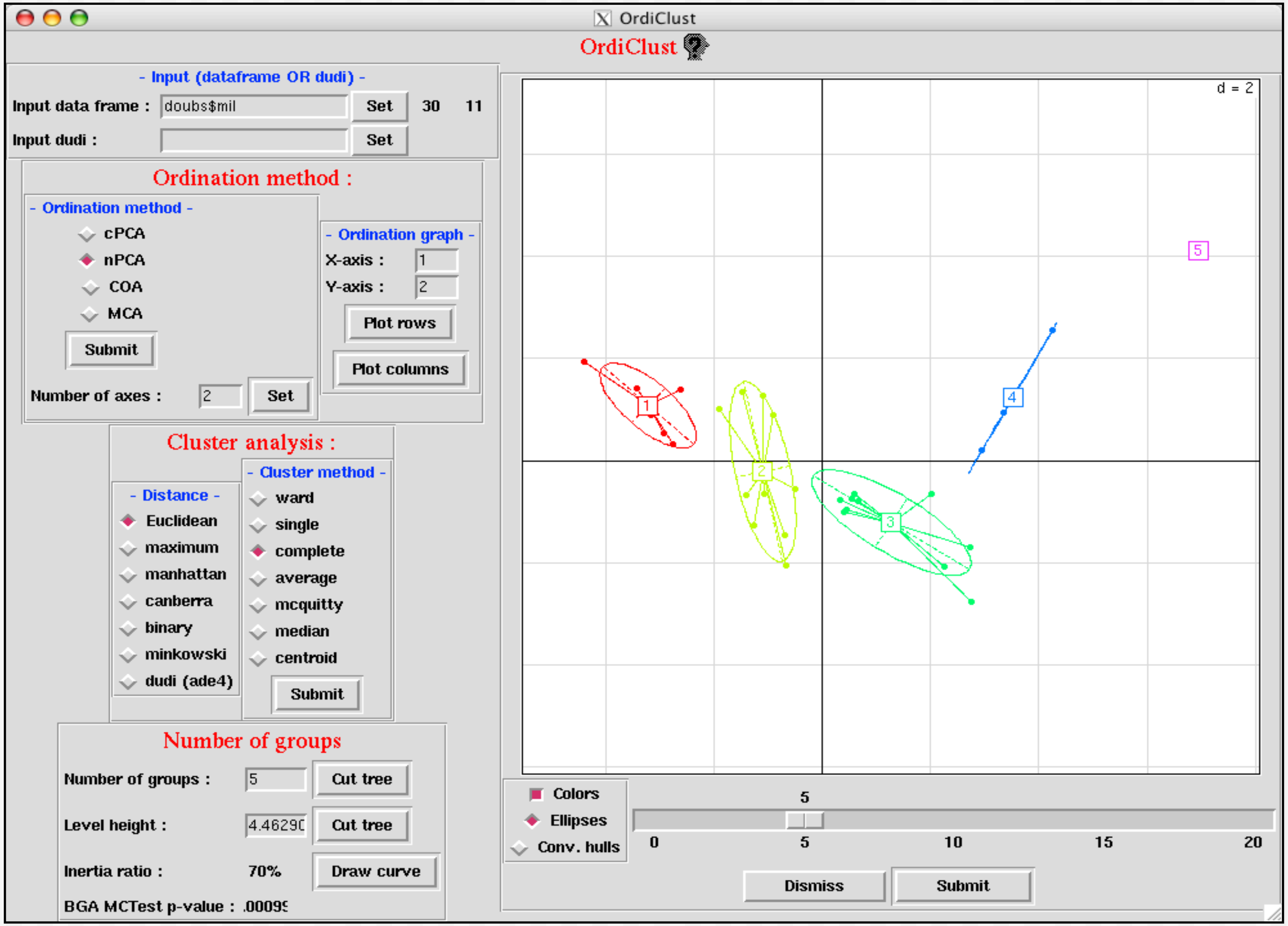
Conclusion

Interactive data analysis



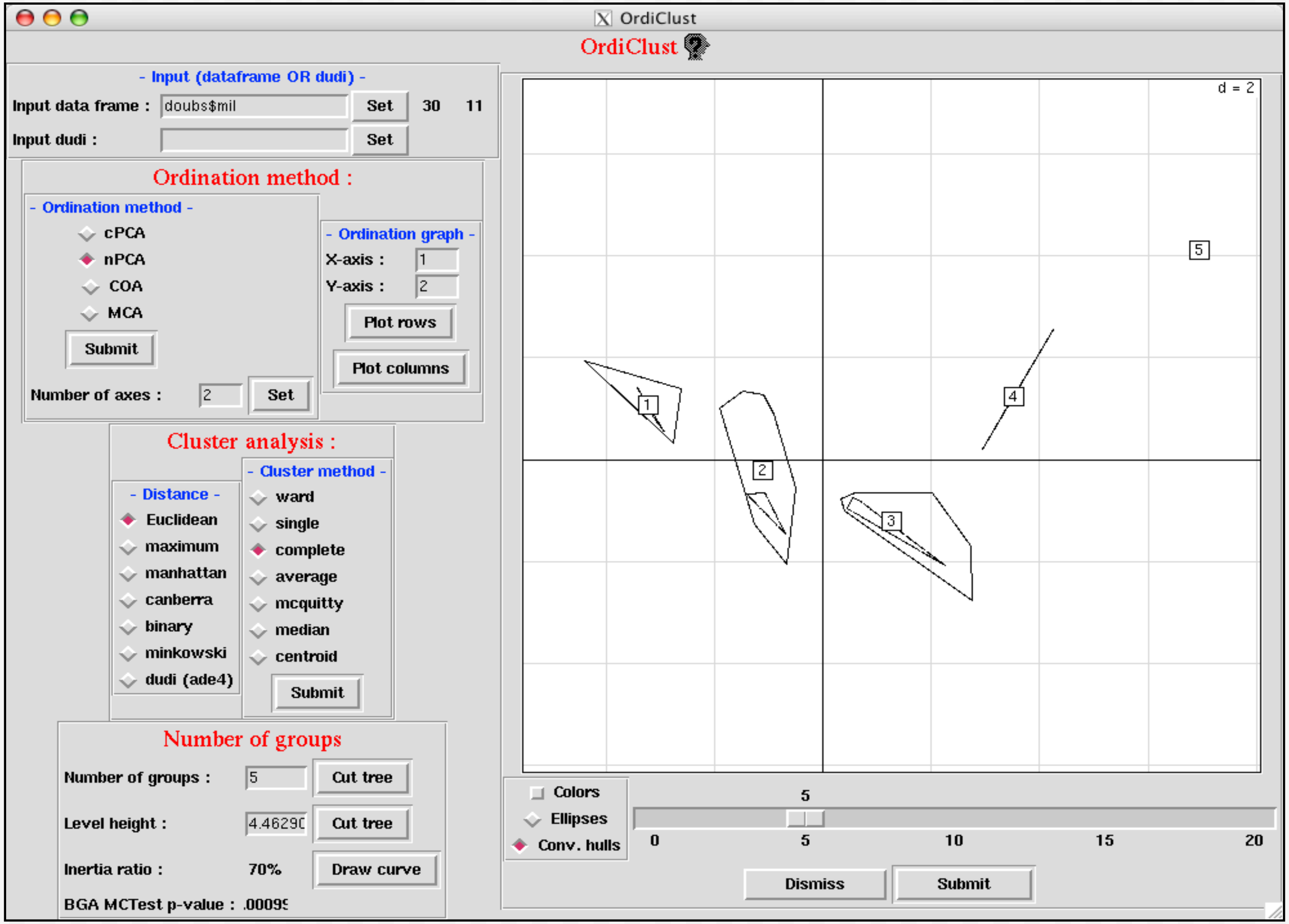
Conclusion

Interactive data analysis



Conclusion

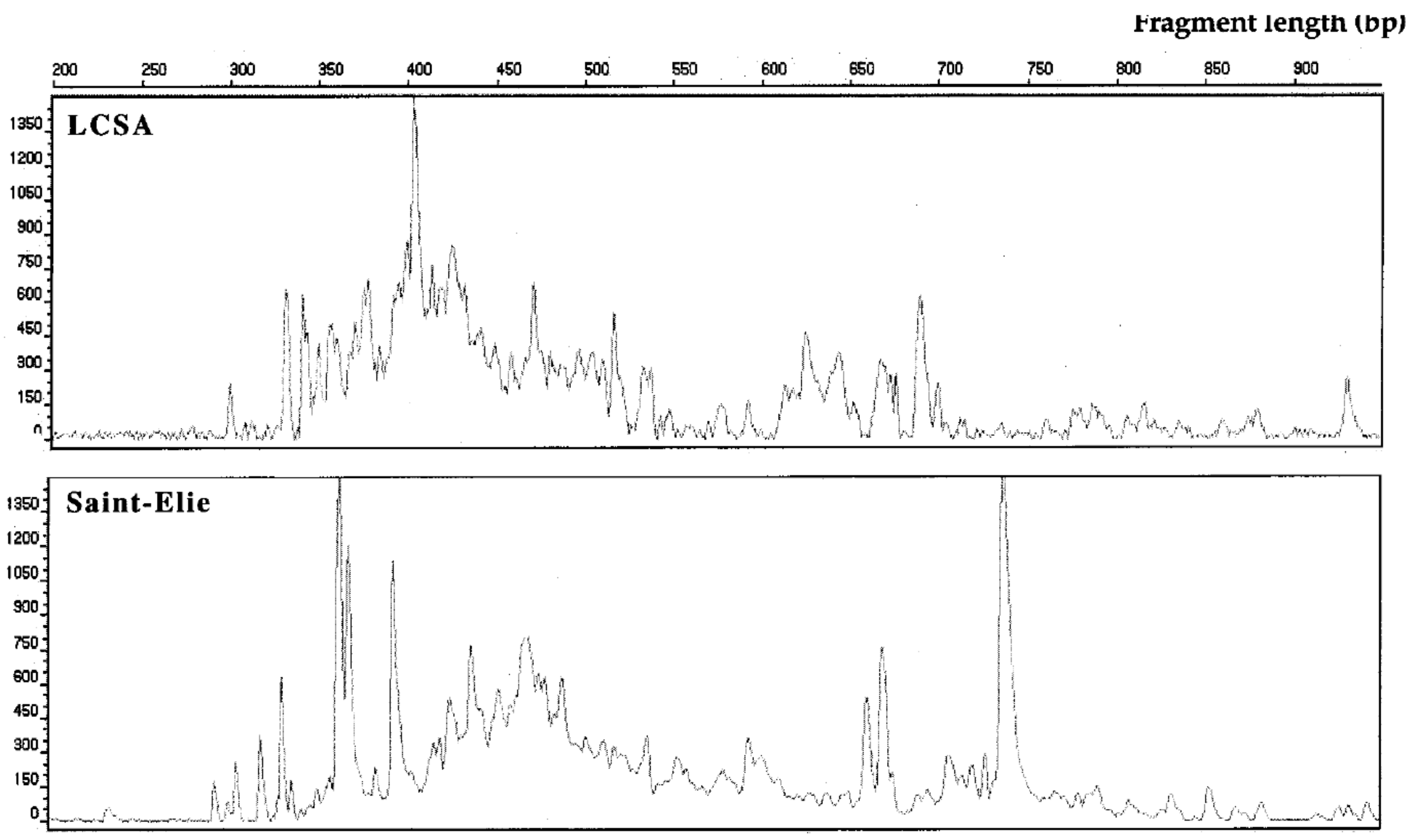
Interactive data analysis



Acknowledgments

- EcoMic - RMQS
  - ✓ Lionel Ranjard, Samuel Dequiedt - INRA Lab in Dijon
  - ✓ Bruno Saby, Manuel Martin - INRA Lab in Orleans
- Mycorrhizal symbiosis
  - ✓ Robin Duponnois - IRD Lab in Dakar & Ouagadougou
- ade4 package
  - ✓ Stéphane Dray, Anne-Béatrice Dufour - Biometry Lab in Lyon

RISA profiles



Remigi P., Faye A., Kane A., Deruaz M., Thioulouse J., Cissoko M., Prin Y., Galiana A., Dreyfus B., & Duponnois R. (2008). Applied and Environmental Microbiology 74, 5, 1485-1493.

Andrianjaka Z., Bally R., Lepage M., Thioulouse J., Comte G., & Duponnois, R. (2007). Applied Soil Ecology 37, 175-183.

Kisa M., Sanon A., Thioulouse J., Assigbetse K., Sylla S., Dieng L., Berthelin J., Prin Y., Galiana A., Lepage M. & Duponnois R. (2007). FEMS Microbiology Ecology, 62, 32-44.

Ouahmane L., Hafidi M., Thioulouse J., Ducousso M., Kisa M., Prin Y., Galiana A., Boumezzough A., & Duponnois R. (2007). Journal of Applied Microbiology 103, 683-690.

Ouahmane L., Thioulouse J., Hafidi M., Prin Y., Galiana A., Plenchette C., Kisa M., & Duponnois R. (2007). Forest Ecology and Management 241, 200-208.

Ramanankierana N., Ducousso M., Rakotoarimanga N., Prin Y., Thioulouse J., Randrianjohany E., Ramaroson L., Kisa M., Galiana A. & Duponnois, R. (2007). Mycorrhiza, 17, 3, 195-208.

Duponnois R., Assigbetse K., Ramanankierana H., Kisa M., Thioulouse, J. & Lepage, M. (2006). FEMS Microbiology Ecology, 56, 292-303.

Duponnois R., Kisa M., Assigbetse K., Prin Y., Thioulouse J., Issartel M., Moulin P., & Lepage M. (2006). Science of the Total Environment, 370, 391-400.

Ouahmane L., Hafidi M., Plenchette C., Kisa M., Boumezouch A., Thioulouse J. & Duponnois R. (2006). Applied Soil Ecology, 34, 190-199.

Ouahmane L., Duponnois R., Hafidi M., Kisa M., Boumezouch A., Thioulouse J. & Plenchette C. (2006). Plant Ecology, 185, 123-134.

Ramanankierana N., Rakotoarimanga N., Thioulouse J., Kisa M., Randrianjohany E., Ramaroson L. & Duponnois R. (2006). International Journal of Soil Science, 1, 8-19.

Sanon A., Martin P., Thioulouse J., Plenchette C., Spichiger R., Lepage M. & Duponnois R. (2006). Mycorrhiza, 16, 125-132.

Assigbetse K., Gueye M., Thioulouse, J. & Duponnois R. (2005). Microbial Ecology, 50, 350-359.

Duponnois, R., Colombet, A., Hien, V., & Thioulouse, J. (2005). Soil Biology and Biochemistry, 37,1460-1468.

Duponnois, R., Paugy, M., Thioulouse J., Masse, D. & Lepage, M. (2005). Geoderma, 124, 349-361.