

Mobyle

Catherine Letondal, Bertrand Néron, Pierre Tuffery

`letondal@pasteur.fr, bneron@pasteur.fr, tuffery@ebgm.jussieu.fr`

Institut Pasteur, EBM INSERM U726

Plan

- pourquoi Mobyle ?
- problèmes traités :
 - chargement et formattage de données
 - navigation et recherche dans le portail
 - enchaînement d'analyses
 - historique des analyses
- interface avec d'autres projets

bioweb.pasteur.fr

Logiciels en ligne sur le serveur Web ([carte d'orientation](#))

- Recherches dans les banques.

- Conversions de formats de séquences.

- Alignements et comparaisons de séquences.

- Recherche de séquences similaires dans les banques ([BLAST](#) et [FASTA](#))
- Comparaison et alignement de deux séquences
- Alignements multiples
 - [CLUSTALW](#)
 - [autres programmes](#)
- Alignements de structures
- HMM (Hidden Markov Models)

- Analyses phylogénétiques.

- [PHYLIP](#)
- [autres programmes](#)

- Recherche de gènes et régions codantes.

- [GENSCAN](#)
- [autres programmes](#)

- Recherche et extraction de motifs, HMM.

- Recherche de motifs ou de domaines connus
- Recherche d'un motif libre
- Extraction de motifs à partir d'un ensemble de séquences
- Recherche de répétitions
- HMM (Hidden Markov Models)
- [autres programmes](#)

- Outils pour les séquences

- Analyses de séquences de protéines.

- Recherche de motifs
- Composants de protéines
- Propriétés de protéines
- Cinétique Enzymatique

- Analyses de séquences d'ADN.

- Recherche de motifs
- Sites de restriction
- Facteurs de transcription
- Répétitions
- Usage des codons, composition
- Îlots CpG
- Primers
- Propriétés de l'ADN
- [autres programmes](#)

- Analyses d'ARN.

- [MFOLD](#)
- [autres programmes](#)

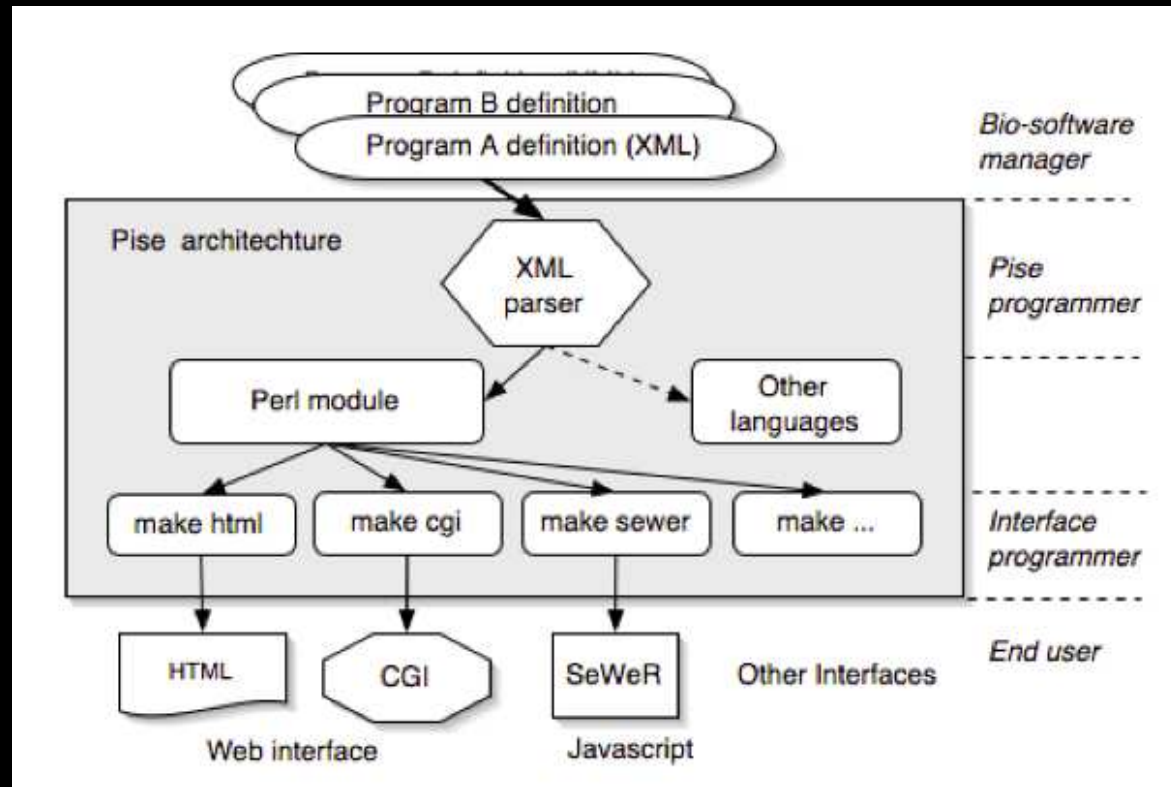
- Analyses de structure.

- Structure secondaire
- Structure tertiaire

- Nouveautés **NEW**

- [SMILE](#)
- [PDBSEARCH](#)
- [CHECKTREE](#)

Pise



C. Letondal (2001), "A Web interface generator for molecular biology programs in Unix", Bioinformatics, Oxford University Press, 17(1), 2001

Problèmes traités

*... sur quoi avons-nous le plus porté
notre attention ...?*

Chargement et formattage de données


- boîte de données commune (SeWeR, M. Basu)
- fichier, DB, copier-coller
- fichiers d'analyses précédentes
- plusieurs fichiers chargés pour la même analyse
- répétition de l'analyse pour plusieurs séquences

Chargement et formattage de données

- formats :
- assistance

Enter or Paste a DNA Sequence in any [supported format](#): Help on: [sequence formats](#).

GGAGGACGCT	TGCAGTGGCG	GCCCCTGACT	TGCTGGATCC	CAAGTCTGCT	
GGAGGACGCT	TGCAGTGGCG	GCCCCTGACT	TGCTGGATCC	CAAGTCTGCT	60
CCAAACCGAG	GCTCTCATT	TCCTCAAAAC	CCACCGTGCT	TGCTTCCCGG	120
ACTCGGCCAT	TAATGTTATG	AAATGGAAGA	CAGTCTCCAC	GATTTTCCTG	180

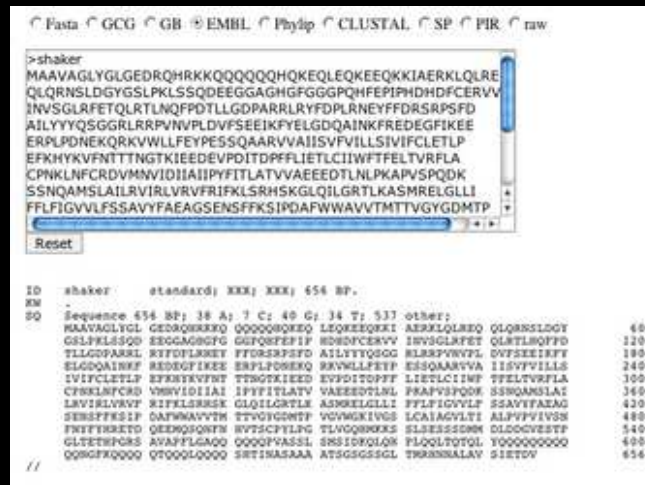
 ~~10~~ ATTGTCGIATT
TGTCCTGTTA

Undo

or [Upload](#) a file:

Chargement et formattage de données

- formats :
 - formattage dynamique

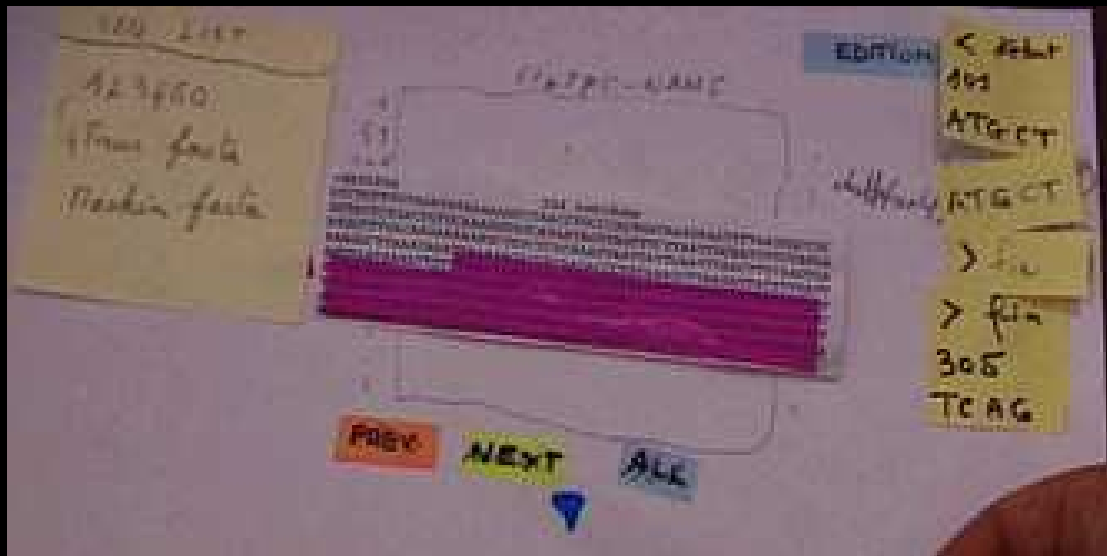


```
< Fasta < GCG < GB < EMBL < Phylo < CLUSTAL < SP < PIR < raw
>shaker
MAAVAGLYGLGEDRQHRKKQQQQQHQKEQLEQKEEQKKAERKLQRE
QLQRNSLDGYGSLPKLSSQDEEGGAGHGFGGQPMFEPHMDHDFCERVV
INVSGLRFETQLRTLQFPDILLGDPARRLRVDFPLRNEYFFDRSRPSFD
AILYYQSGGRLRPVNVPLDVFSEIKFYELGDQAINKFREDEGFKEE
ERLPDNEKQKRVWLLFEYPESSQAARVVAIISVFILLSVIFCLETLP
EFKHYKVFNTTTNGTKIEEDVPDITDPFFLIETLCIHWFTFELTVRFLA
CPNKLNFCRDVMNVIDIIAIIPIYFILTATVVAEEEDTLNLPKAPVSPQDK
SSNQAMSLAILRVIRLVRFRIKLSRHSGKLGLGATLKASMRGLLLI
FFLFIGVVLFSASVYFAEGSENSFFKSIPDAFWWAVVTMTTVGYGDMTP
Reset

ID shaker standard; XXX; XXX; 654 BP.
XX
50 Sequence 654 BP: 18 A; 7 C; 40 G; 34 T; 537 other:
60 MAAVAGLYGL GEDRQHRKKQ QQQQHQKEQ LEQKEEQKKA ERKLQREQ QLQRNSLDGY
120 GSLPKLSSQD EGGAGHGFG GFGQPMFEP HMDHDFCERV INVSGLRFET QLTRLHDFC
180 TLLGDPARRL RITDFLAKET FFGSRPSFD AILYYTQSGG ELARPVWVTL DVFSEIKAFY
240 ELGDAIRHF REDEGFKEE ENFLPQGGQ RNVMLLFETP EESQAARVA IISVFVILL
300 TVIFCLETLP EFKHYKVFNT TNGTKIEED VPDITDPFF LIETLCIHW FTFELTVRFLA
360 CPNKLNFCRD VMNVIDIIAI IFYFILTATV VAEEDTLNL PRAFPVSPQD SSNQAMSLAI
420 LRVIRLVRFRI FELSRRSEK GLIILGATLE ASKRELGLLI FFLFIGVVLFS SAVITFAZAG
480 SEHSFFESIP DAFWNAVVTM ITVGYGDMTP VGVNKRIVES LCAIAGVLT IALPFPVIVSS
540 PRTFHSKETO GEDRQDQNTM RYTCPIYLG TLVQGRKES SLSESSDMM GLOOGVEETP
600 GLTETHFGKS AVAFPLAGDQ QQQQPVASLL SRSIDVQLQV FLQDLTOTQL YQQQQQQQQQ
654 QQQGFKQQQQ QQQQLOQQQ SRTINASAAA ATSGSGSSGL TWRNNALAV SIETOV
```


Chargement et formattage de données

- filtrer les données :



Navigation et recherche dans le portail

Mobyle

?

» Programs

+

2d structure

+

3d structure

+

composition

+

mutation

+

profiles

+

phylogeny

bambe

dnapars

phyloquart

» Files

» Jobs

toppred 19-8-05 10:15

blast2>>clustalw>>dnapars 1-8-05 18:01

blast2 23-7-05 11:18

DB

Sequence

Sequences

DNA

Protein

Alignment

Structure

Report

Data

DB

File

History

Repeat

your e-mail:

More Data

Reset

dnapars

fastdnaml

fastDNAm! : construction of phylogenetic trees of DNA sequences using maximum likelihood ([Olsen, M](#)
[Hagstrom, Overbeek](#))

Parameters

Sequence Alignment File

Input Options

Instructs the program to use
empirical base frequencies derived

Navigation et recherche dans le portail

- utilisation en mode tutorial

This form enables you to choose a sequence database. Genbank contains sequences databases. Nrl3d contains sequences which 3D structure is known.

Now, select the *SwissProt* database and hit Select.

Please select a database

GenBank	
SwissProt	
PIR	
nrl3d	

Select

Navigation et recherche dans le portail

- aide pour remplir les champs d'un formulaire

Enter a sequence

You can load a sequence by following these steps: ChooseDB\EnterID

--

Enchaînement d'analyses

Mobyle

DB Sequence Sequences DNA Protein Alignment Structure Report

Data DB File History Repeat

your e-mail:

More Data Run
Reset Pipe

blast2>>clustalw>>dnapars

BLAST2 : with gaps (Altschul, Madden, Schaeffer, Zhang, Miller, Lipman)

Parameters

Blast program: blastp

Start of required region in query sequence (-L)

End of required region in query sequence (-L)

toppred 19-8-05 10:15
blast2>>clustalw>>dnapars 1-8-05 18:01
blast2 23-7-05 11:18

Tutorials

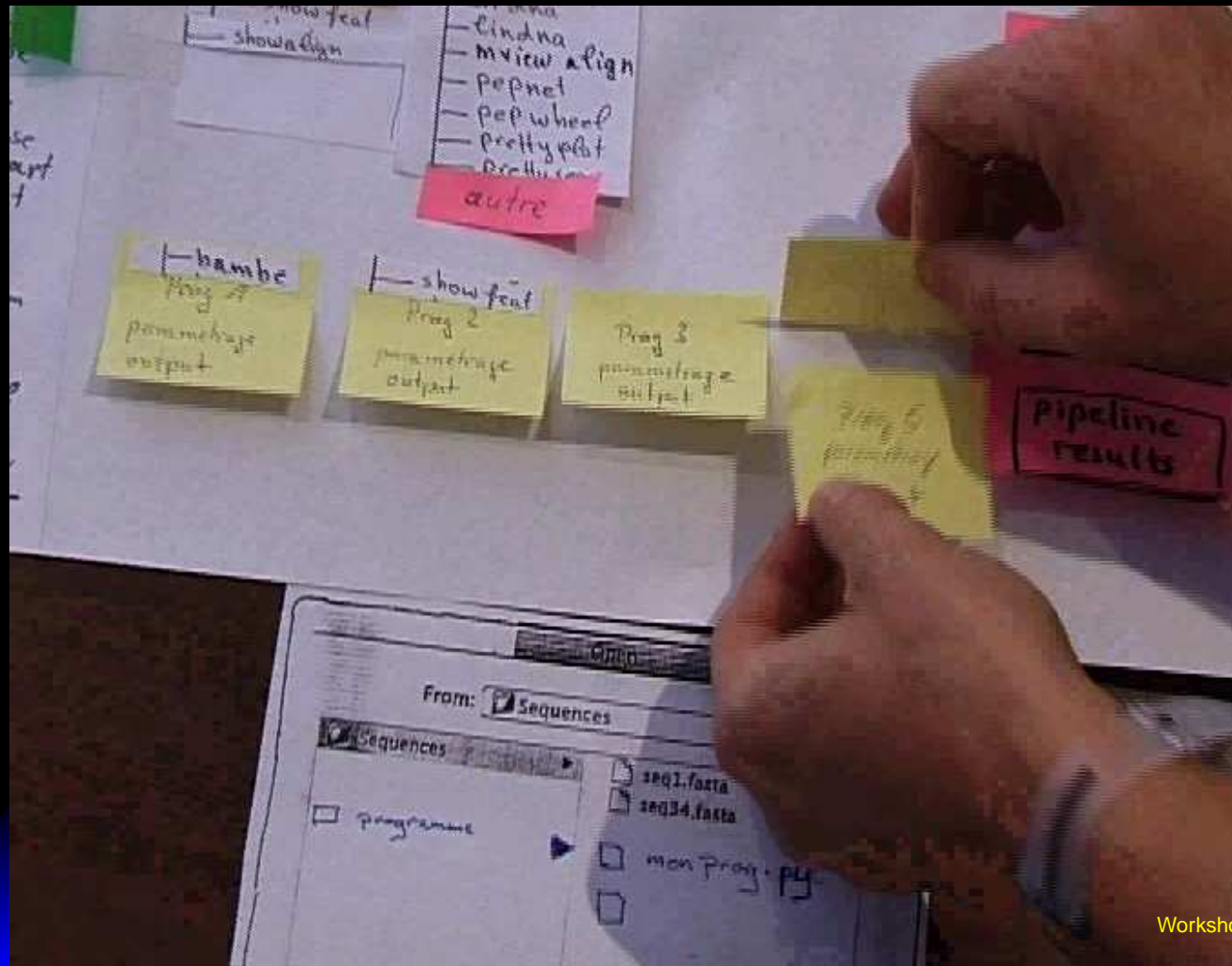
Enchaînement d'analyses

- des urls bizarres...

```
pipeline=ChooseDB.process?db=sprot|EnterID.process?id=malk_ecoli|DisplayEntry.render
```

Enchaînement d'analyses

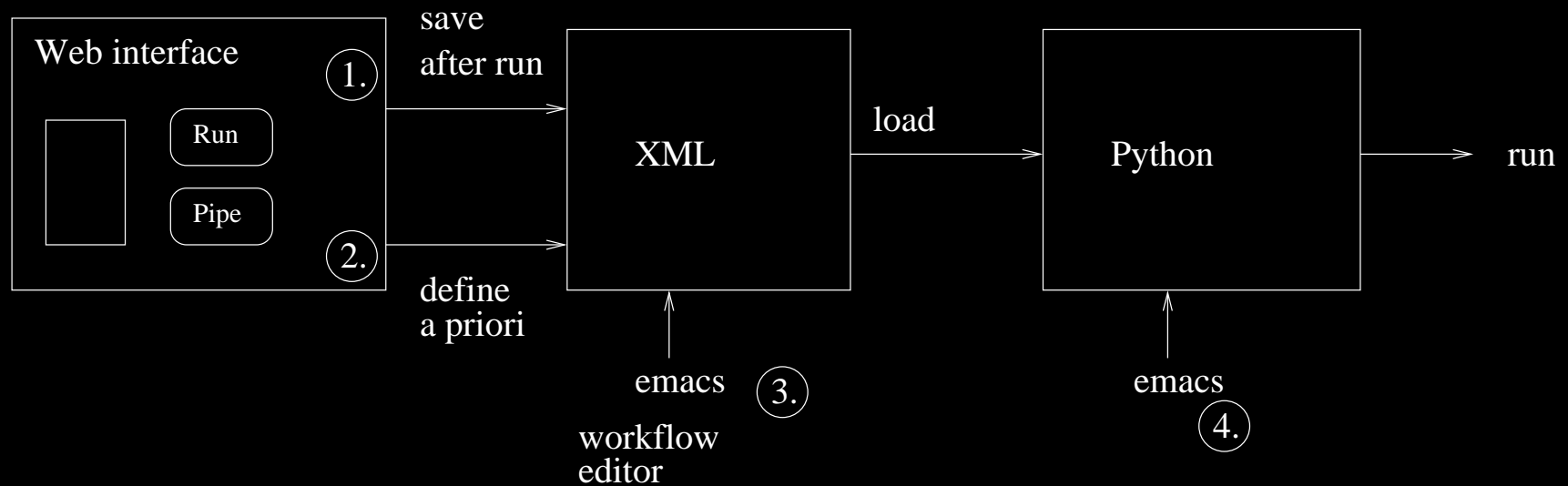
- zig-zag entre machines (y compris la vôtre ?)



Scripting

```
f = MobyFactory()  
dnadist = f.program('dnadist',  
                    infile=align)  
neighbor = f.program('neighbor')  
drawtree = f.program('drawtree')  
  
wk = PiseWorkflow()  
wk.addpipe(method=dnadist,  
            tomethod=neighbor, pipetype="dist")  
wk.addpipe(method=neighbor,  
            tomethod=drawtree, pipetype="treefile")  
wk.run()
```


Pipelines : summary



Historique et accès aux résultats

- voir les jobs lancés à partir de tel ou tel formulaire

cluster W

► form

▼ results

CLUSTAL W (1.9.2) multiple sequence alignment

```

SAY3956  GTCTCTTTGAGAGATGTAAGAGTTCTAAGTCCGCACTGATAAATAMTTTAAAGGCGG
SAY3957  GTCTCTTTGAGAGATGTAAGAGTTCTAAGTCCGCACTGATAAATAMTTTAAAGGCGG
SAY3952  GTCTCTTTGAGAGATGTAAGAGTTCTAAGTCCGCACTGATAAATAMTTTAAAGGCGG
*****

SAY3956  GGTATTTCTGACTGTCAGAGGTTACCATATATGTTTCTTAAATGGAATCTTATATG
SAY3957  GGTATTTCTGACTGTCAGAGGTTACCATATATGTTTCTTAAATGGAATCTTATATG
SAY3952  GGTATTTCTGACTGTCAGAGGTTACCATATATGTTTCTTAAATGGAATCTTATATG
*****

SAY3956  ATGTTTTCACAGAGAGAGAGTTGTTCTGCAAAATTAATTATTGAATTTACTTTAAGTG
SAY3957  ATGTTTTCACAGAGAGAGAGTTGTTCTGCAAAATTAATTATTGAATTTACTTTAAGTG
SAY3952  ATGTTTTCACAGAGAGAGAGTTGTTCTGCAAAATTAATTATTGAATTTACTTTAAGTG
*****

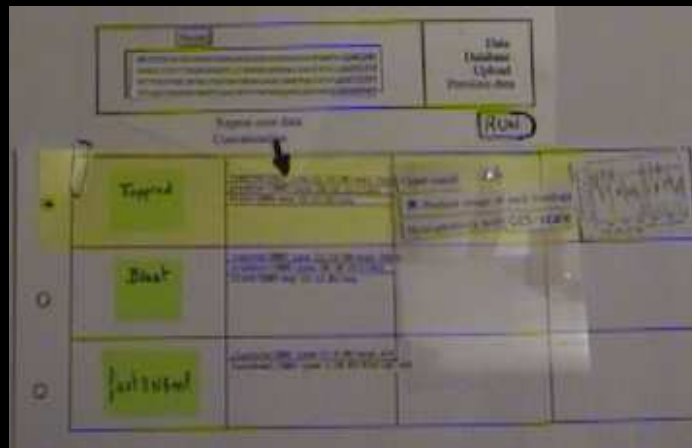
```

▼ jobs

- A 410120 1/3/04 15:06
- A 410140 6/9/04 11:00
- A 410145 6/9/04 11:07

Historique et accès aux résultats

- relancer une analyse

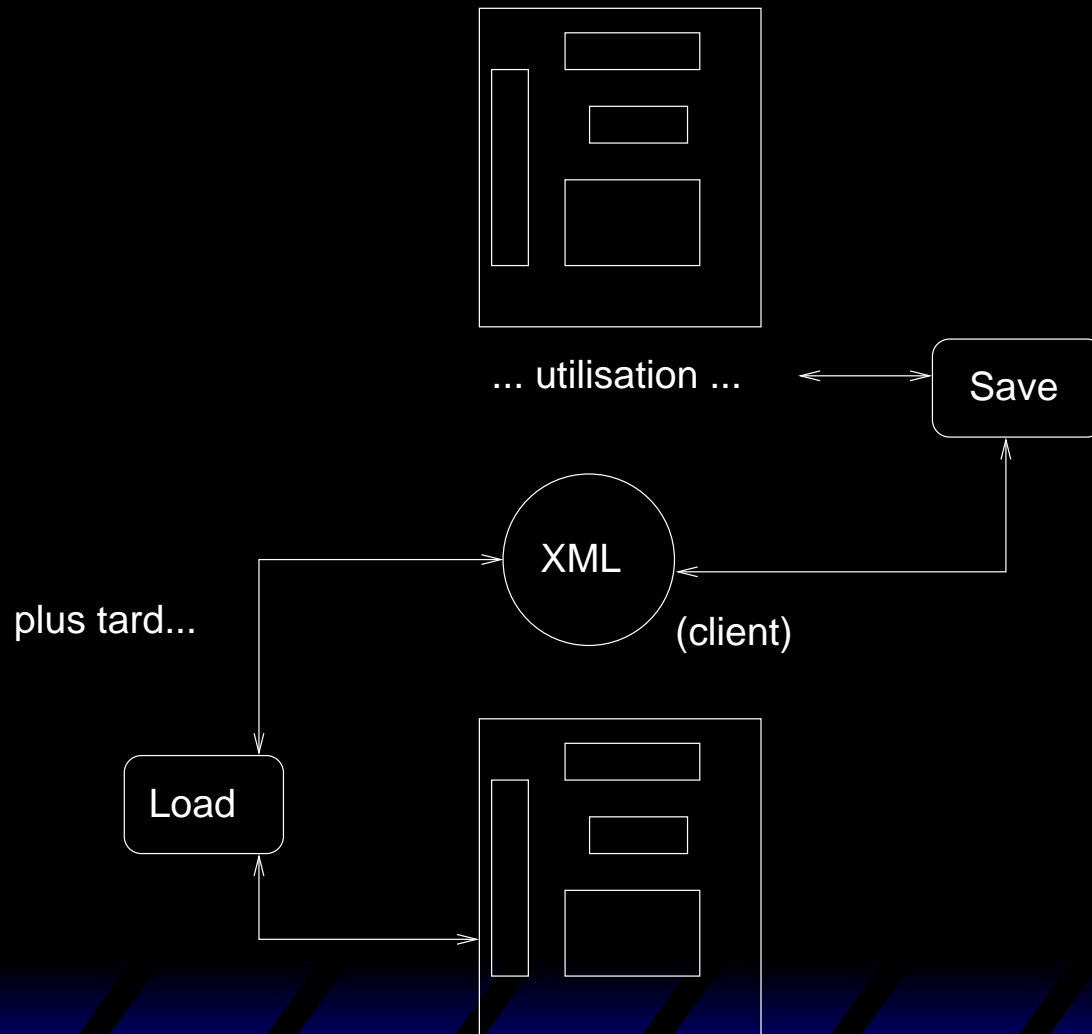


- sauver un paramétrage et le recharger

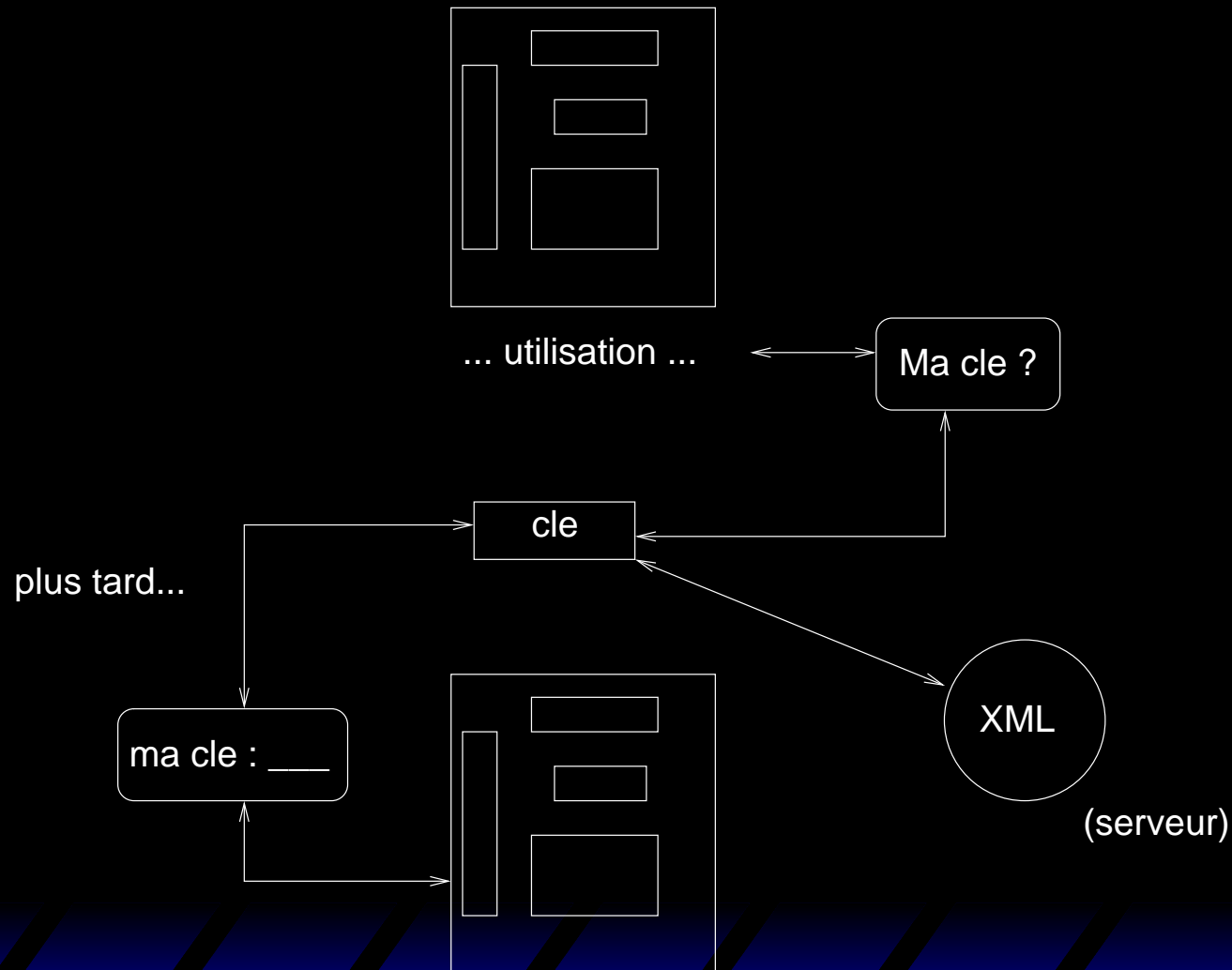
Profils d'utilisation

- login + mot de passe ...?
- cookies, etc... ?
- plutôt : profils d'utilisation
- 2 solutions : coté serveur / coté client

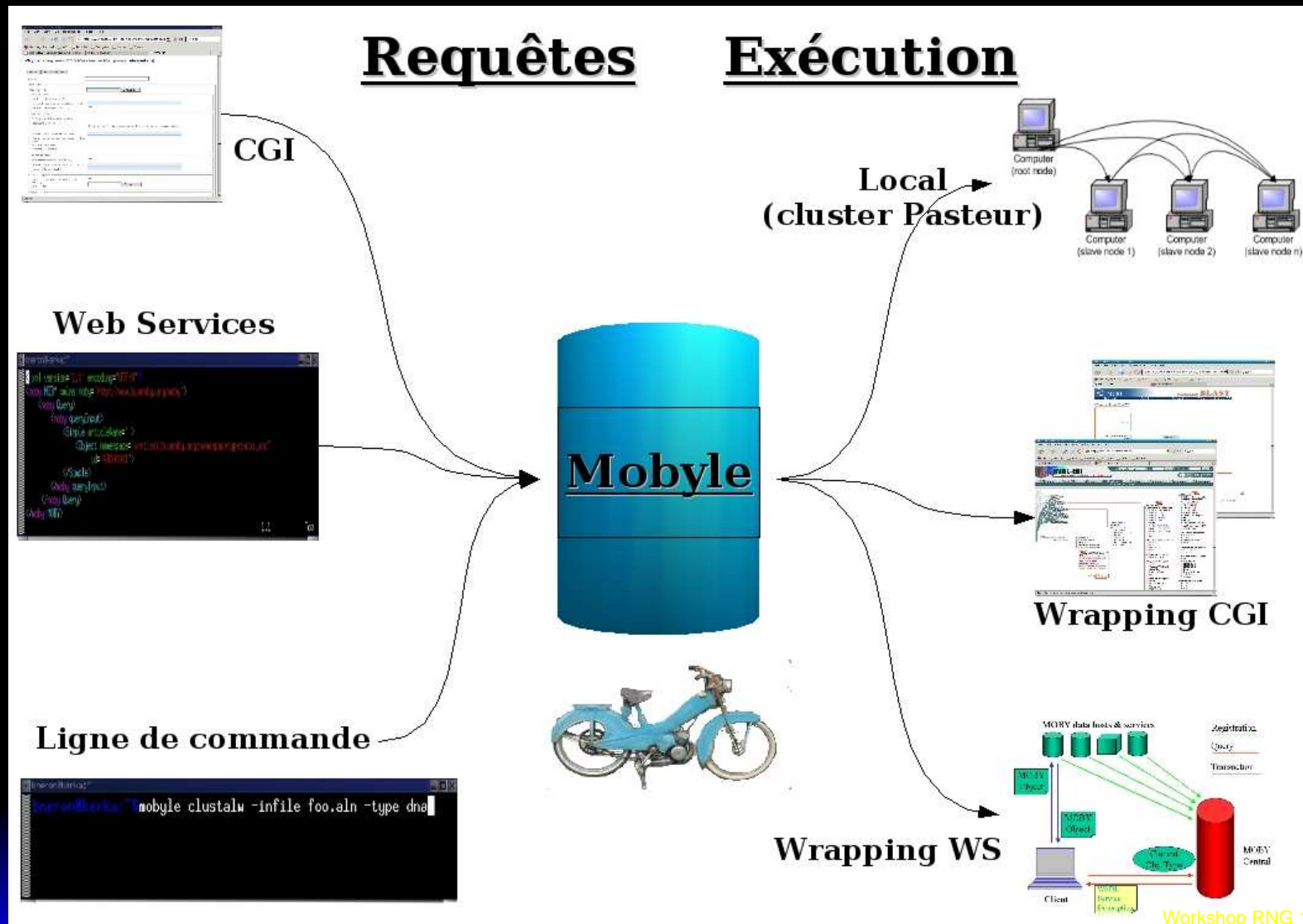
Profils d'utilisation



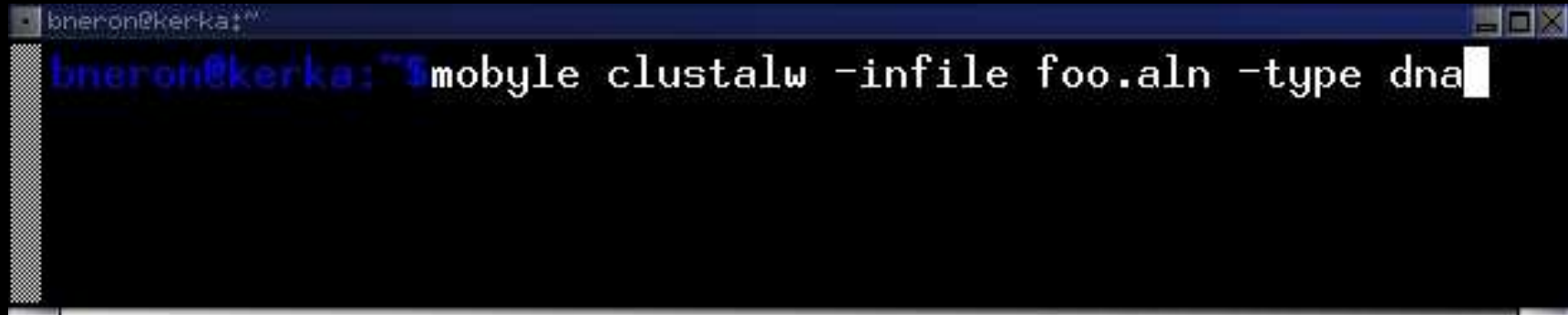
Profils d'utilisation



Interfaces



Ligne de commande

A screenshot of a terminal window with a dark blue title bar and standard window controls. The terminal has a black background with white text. The prompt 'bneron@kerka:' is shown in white. The command 'mobyle clustalw -infile foo.aln -type dna' is entered in white text, followed by a white cursor. The terminal window is positioned in the upper half of the slide, above a blue and black diagonal striped pattern.

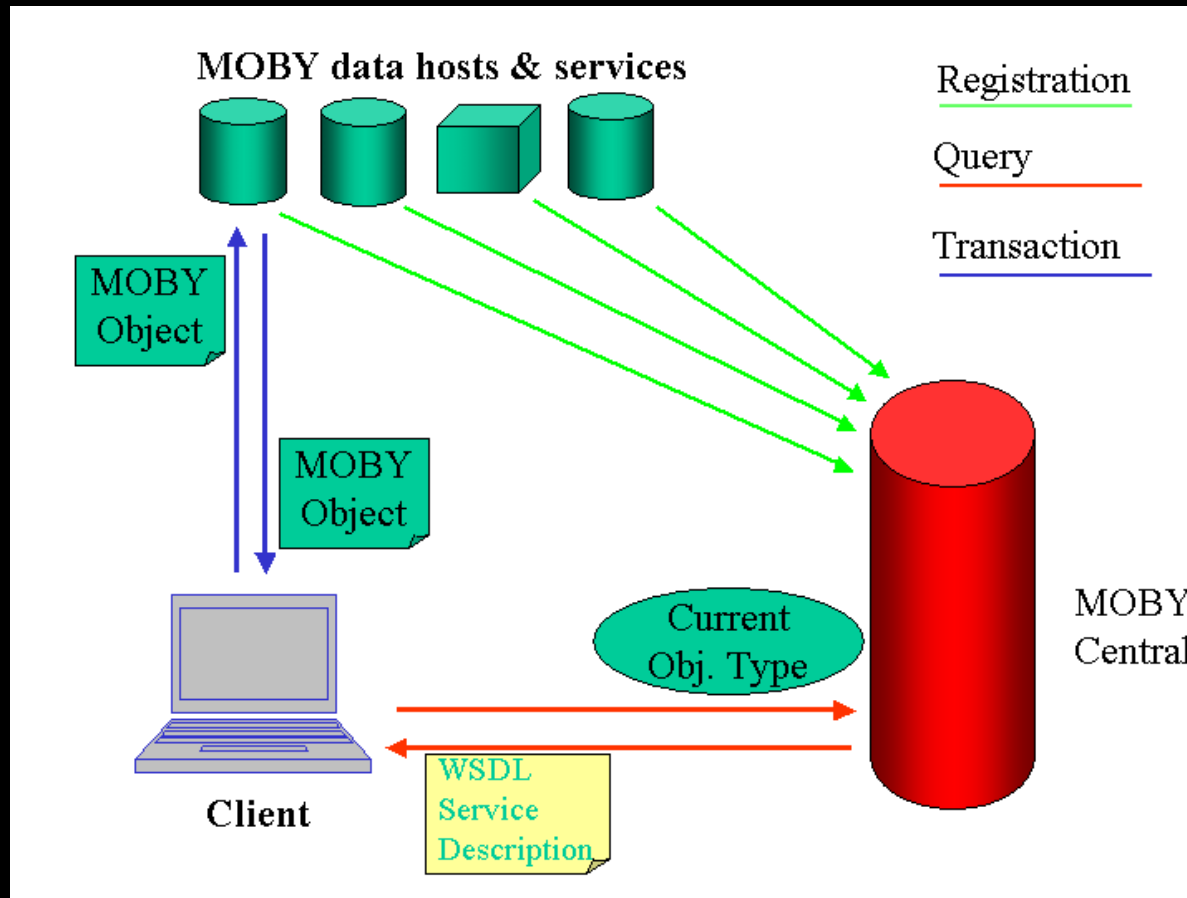
```
bneron@kerka:~$ mobyle clustalw -infile foo.aln -type dna
```


Utilisation des programmes comme services Web

```
bneron@kerka:~$  
<?xml version="1.0" encoding="UTF-8"?>  
<moby:MOBY xmlns:moby="http://www.biomoby.org/moby">  
  <moby:Query>  
    <moby:queryInput>  
      <Simple articleName="">  
        <Object namespace="urn:ledi:biomoby.org:namespace:type:nobi_acc"  
          id="AB0000001">  
        </Object>  
      </Simple>  
    </moby:queryInput>  
  </moby:Query>  
</moby:MOBY>
```

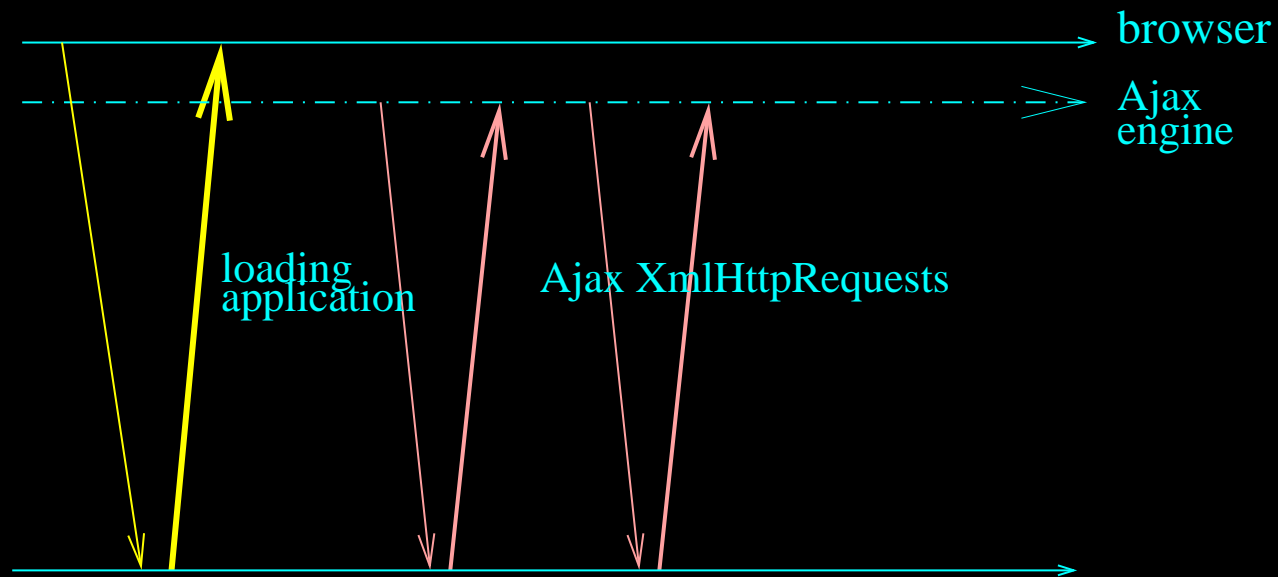
1.1 Top

Biomoby



AJAX : le Web plus interactif

- requêtes dynamiques
- multiplexage de documents



Méthodologie de spécification...

- conception participative : pas seulement des réunions avec les utilisateurs...
- ... des entretiens dans le laboratoire
(*qu'est-ce qu'il vous arrive vraiment devant votre écran ?*)
- ... des ateliers de brainstorming et de conception (*et vous, comment voyez-vous les choses ?*)

Méthodologie de spécification...

- *C. Letondal and W. E. Mackay (2004), Participatory Programming and the Scope of Mutual Responsibility: Balancing scientific, design and software commitment, in Proceedings of PDC 2004 (Participatory Design Conference), July 27 -31, 2004 - Toronto, Canada*
- *C. Letondal, O. Amanatian Participatory Design of Pipeline tools and Web services in bioinformatics, at Requirements Capture for Collaboration in eScience Workshop, January 2004, NESC*

Interfaces avec d'autres projets

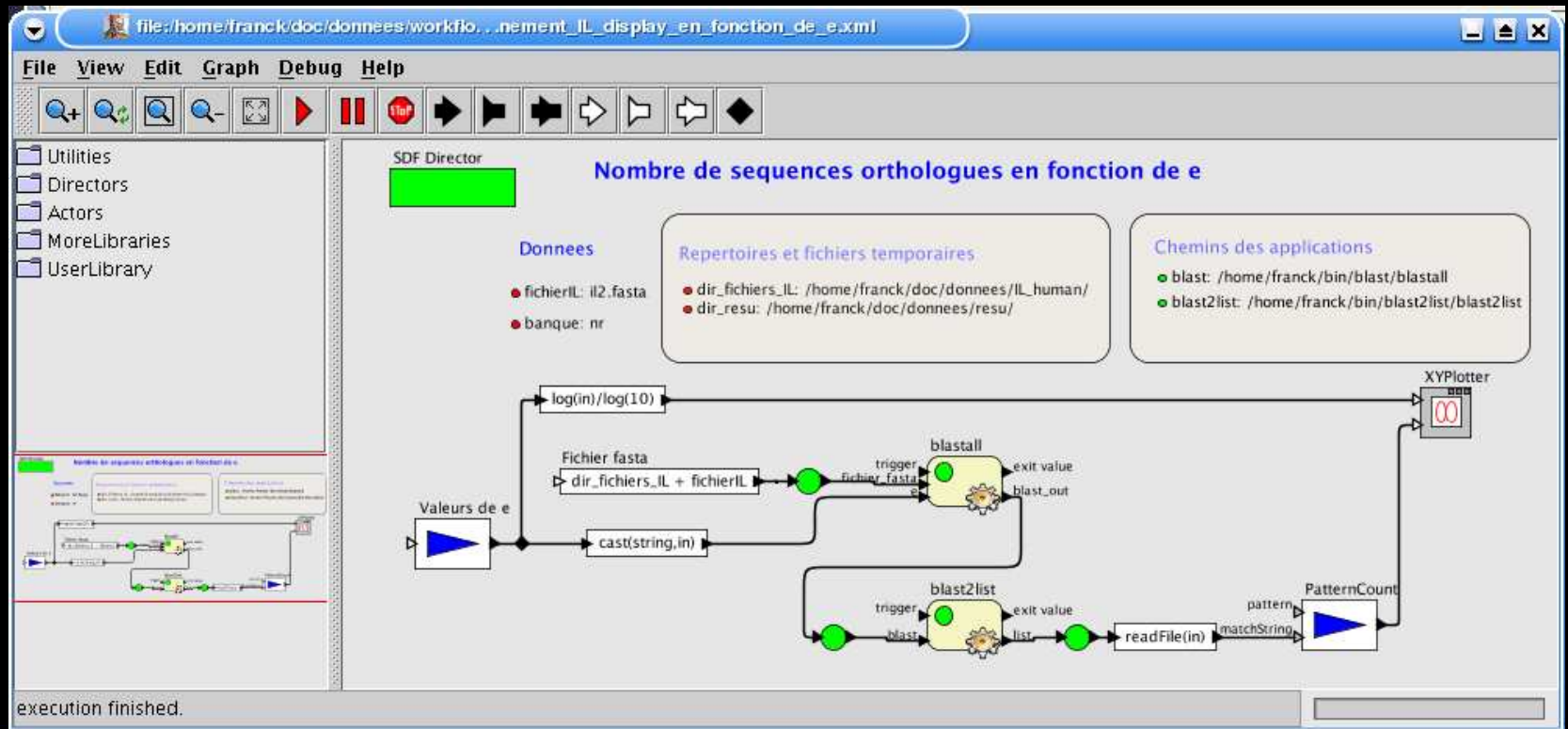
- portail RPBS (Pierre Tuffery)
- Panoramic (Thierry Rose)
- G-Pipe (Samuel Thoraval, Alexander Garcia)
- biomoby (www.biomoby.org)
- a-book (Wendy Mackay, Pascal Costa-Cunha)

RPBS : Ressource Parisienne pour la Bioinformatique Structurale

- Pierre Tuffery, Yan Wong, Francois Moreews, Cynthia Alland
- <http://bioserv.rpbs.jussieu.fr/RPBS>

Panoramic

- Thierry Rose, Franck Valentin

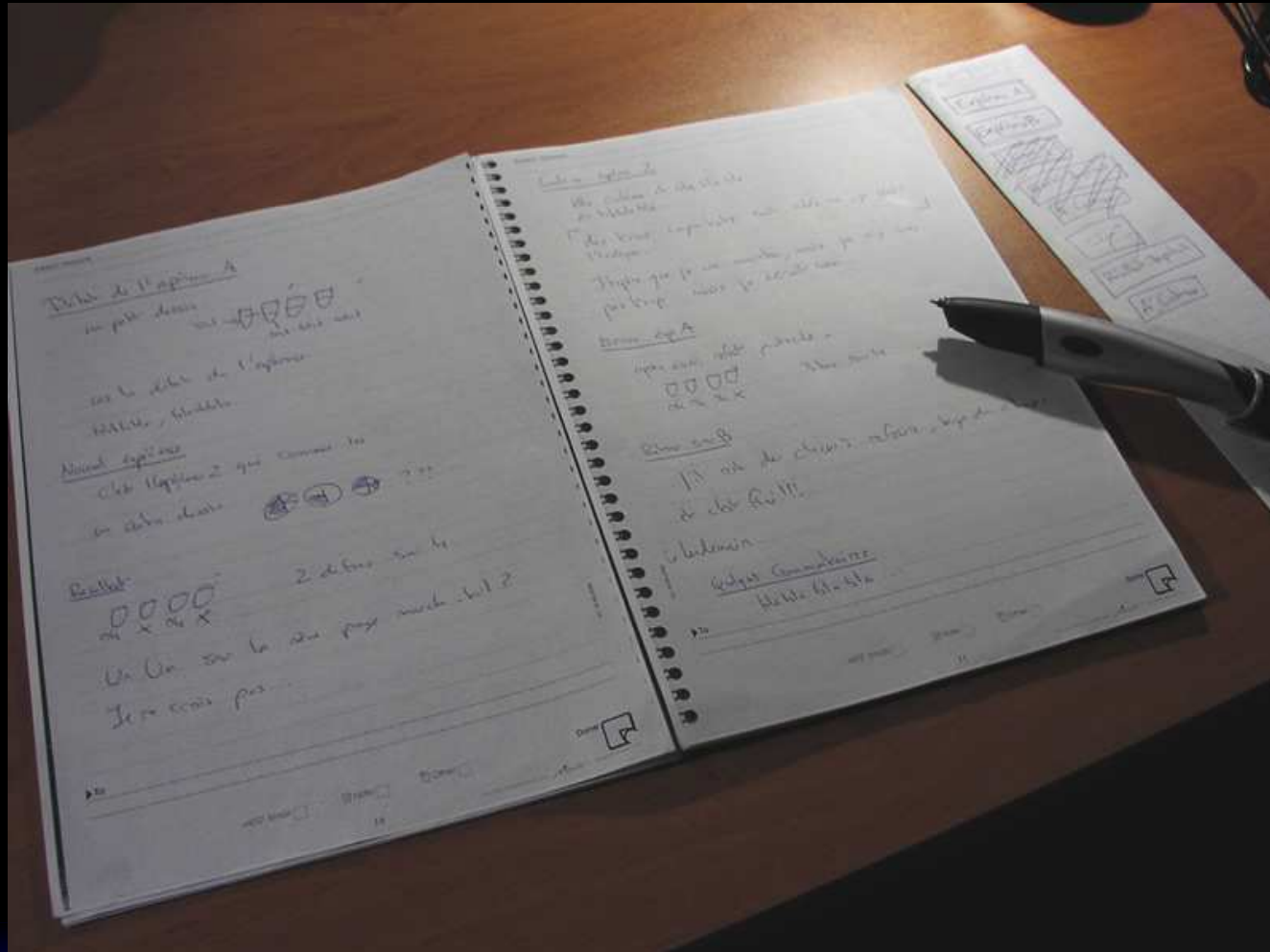


a-book

- Wendy Mackay, Guillaume Pothier, Pascal Costa-Cunha



a-book



a-book



Contributions

- Bertrand Néron : conception, maquettage, développement
- Pierre Tuffery (EBGM) : conception, maquettage, développement
- Thierry Rose (Unité d'Immunogénétique Cellulaire) : philosophie, conception, maquettage
- Yan Wong (Infobiogen) : python / biomoby

Contributions

- Olivier Amanatian : entretiens au laboratoire
- Lina Tanguy : maquettage
- Samira Laribi : entretiens, maquettage, recherche dans la documentation
- Aurélien Tabard : entretiens, maquettage, navigateur Web pour la biologie

Contributions

- Wendy Mackay (INRIA Futurs) : cahiers de laboratoire augmentés
- Samuel Thoraval, Alexander Garcia : G-Pipe
- Nicolas Roussel (LRI) : navigateur Web pour la biologie
- Emmanuel Pietriga (INRIA Futurs) : outils de navigation et de visualisation

Contributions

V. Dominguez, L. Frangeul, J. Allignet , B. Regnault, A-M Gilles, G. Guigon, B. Caudron, C. Bourseaux-Eude, I. Iteman, C. Baurgouin, JP.Carlier, D. Clermont, L. Debarbieux, B. Jacquelin, JF Chartes, H.O.Nghiem, S. Moreira, F.Valentin, S. Gomez, P. Dehoux, A. Taly, B. Neron, C. Roth, K. Eiglmier , C. Dauga, G. Reysset, O. Amanatian, L. Tanguy, P. Tuffery, J-P Corre, C. Weber, F. Hantraye, B. Callendret, S. Garbay, C. Dauga, A. Taly, S. Thoraval, T. Planchenaut, T. Rose, T. Garnier, B. Lakowski, M. Vergassola, J. Guijarro, C. Maufrais, K. Wecker, L. Simon, S. Correia, V. Jan, P. Cansonnet, O. Poupel, R. Cheynier, C. Saveanu, L. Damier, N. Joly, E. Deveaud, L. Mancio, P. David, W. Mackay, P. Costa-Cunha

Contributions

B. Neron, P. Tuffery, C. Letondal (2005) Mobyle: a Web portal framework for bioinformatics analyses, poster presented at NETTAB 2005.

Merci à vous !

