Présentation de la plate-forme BacTrans²

- Equipe ComBi du LINA
- Sylvain Demey

http://www.sciences.univ-nantes.fr/lina/bioserv/

Sylvain.Demey@univ-nantes.fr

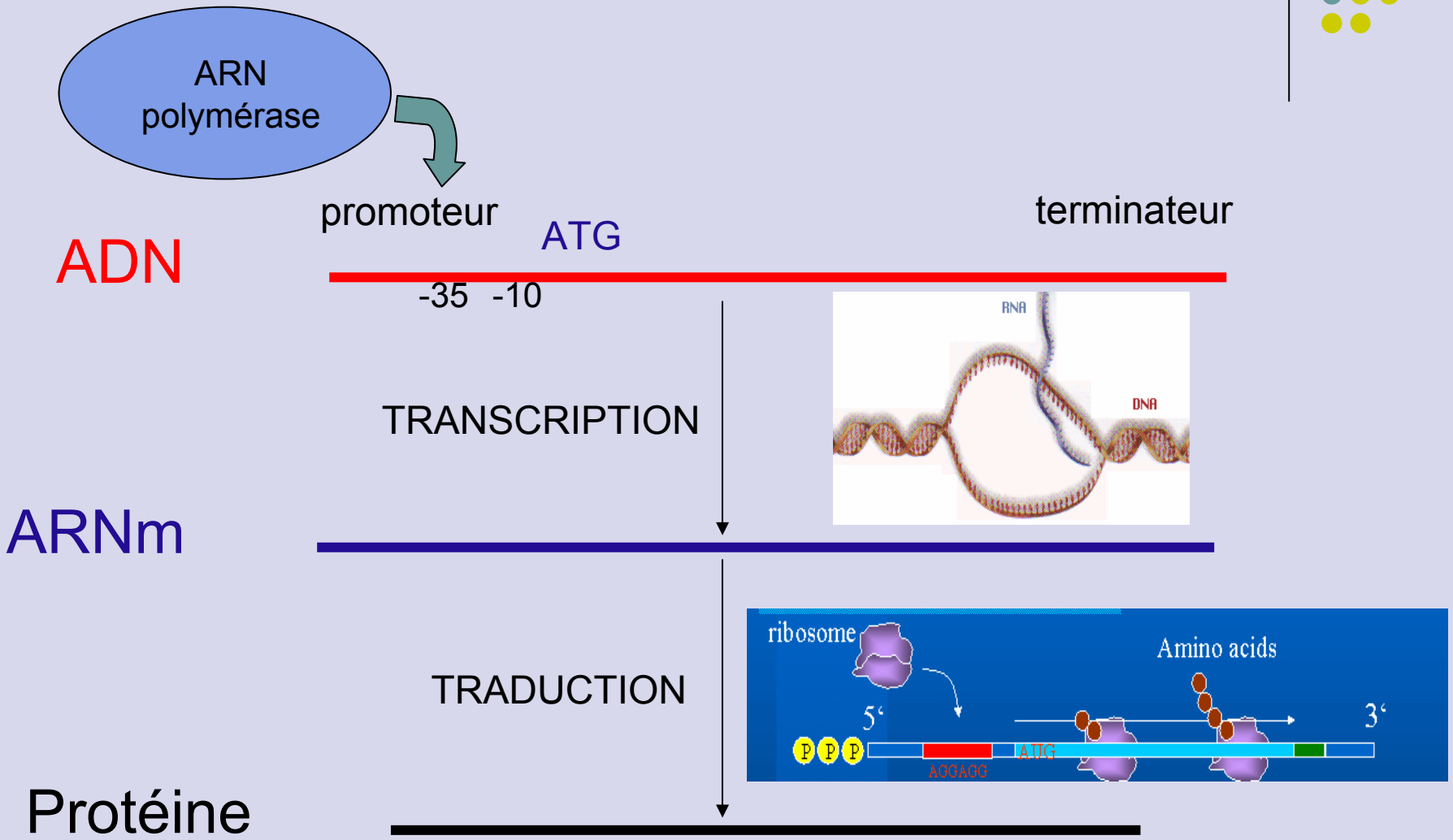
PLAN

- Introduction
- Le Contexte Biologique du Projet
- La Plate-forme et son utilisation
- La fonctionnalité ‘Statistiques’
- Conclusion-Perspectives

Introduction

- Idée née en Janvier 2002 d'une collaboration informelle entre l'UMR Recherche en Biocatalyse du Laboratoire des Biotechnologies de l'Université de Nantes et l'équipe ComBi du LINA
- Axée sur l'inférence de connaissances sur les promoteurs forts des génomes bactériens pour aider les biologistes à étudier ces promoteurs

Le Contexte Biologique du Projet



Le Contexte Biologique du Projet (suite)

- But : Donner aux biologistes les caractéristiques complètes des gènes à promoteurs forts des génomes bactériens et des outils d'analyses
- Travail d'abord réalisé sur *Thermotoga maritima* puis étendu à 30 autres procaryotes
- Caractéristiques → 23 critères relatifs :
  - aux signaux de transcription, de traduction
  - aux structures secondaires prédites, relatives à certaines régions des ARNm concernés
  - à la stabilité de ces molécules.

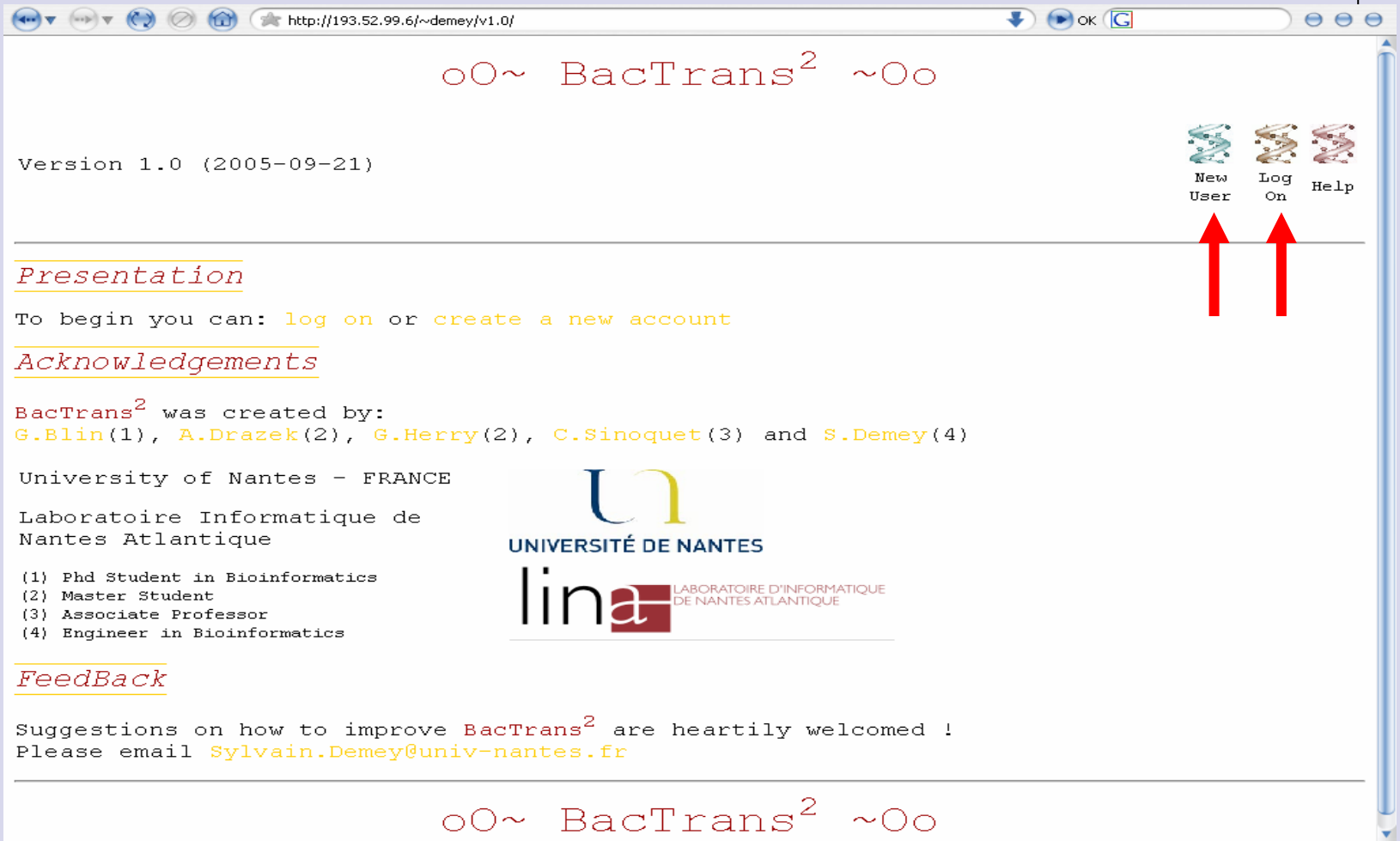
Le Contexte Biologique du Projet (suite)

- *Aquifex aeolicus----------80°c*
- *Bacillus subtilis----------37°c*
- *Borrelia burgdorferi----------37°c*
- *Brucella melitensis----------35°c*
- *Chlamydophila pneumoniae------35°c*
- *Clostridium perfringens-------37°c*
- *Deinococcus radiodurans-------30°c*
- *Escherichia coli k12----------37°c*
- *Haemophilus influenza---------37°c*
- *Helicobacter pylori----------37°c*
- *Mycobacterium leprae----------30°c*
- *Mycobacterium tuberculosis----37°c*
- *Neisseria meningitidis--------37°c*
- *Oceanobacillus iheyensis------<15°c*
- *Pseudomonas aeruginosa--------55°c*
- *Pyrococcus abyssi----------95°c*
- *Pyrococcus horikoshii---------98°c*
- *Rickettsia prowazekii---------35°c*
- *Salmonella typhimurium--------37°c*
- *Shewanella oneidensis---------22°c*
- *Sinorhizobium meliloti--------31°c*
- *Staphylococcus aureus---------37°c*
- *Streptococcus pneumoniae------37°c*
- *Streptomyces coelicolor-------30°c*
- *Thermoanaerobacter tengcongensis-75°c*
- *Thermotoga maritima----------80°c*
- *Treponema pallidum nichols----18°c*
- *Vibrio cholerae----------37°c*
- *Xanthomonas campestris--------28°c*
- *Yersinia pestis----------28°c*

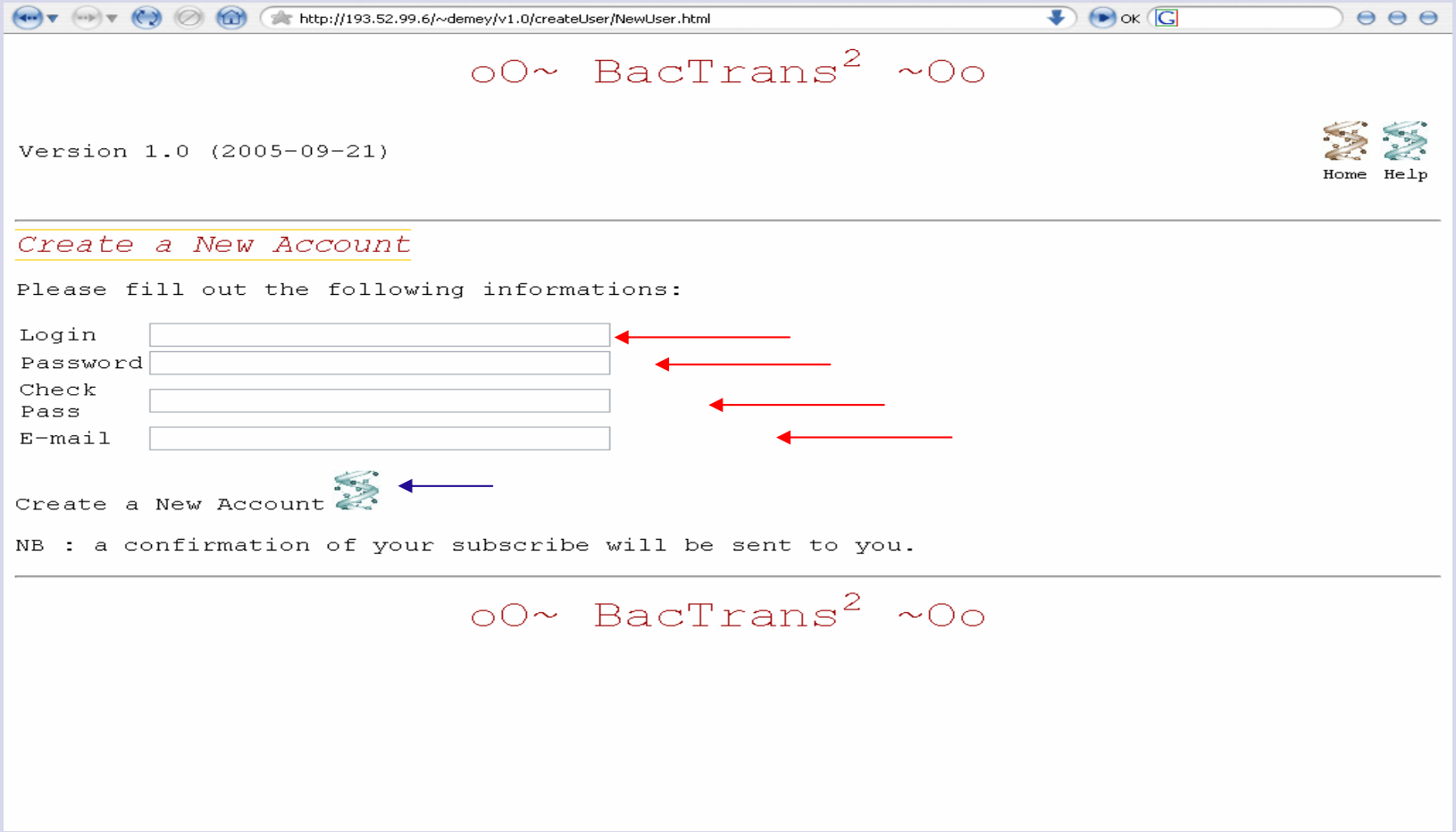
La Plate-forme

- Développée sur Bioserv (serveur financé par Ouest-Génopole)
- Réalisée dans le but d'une large utilisation par les biologistes étudiant les Procaryotes
- Fonctions:
  - Regrouper les connaissances sur les promoteurs forts
  - Mettre à disposition des outils d'analyses

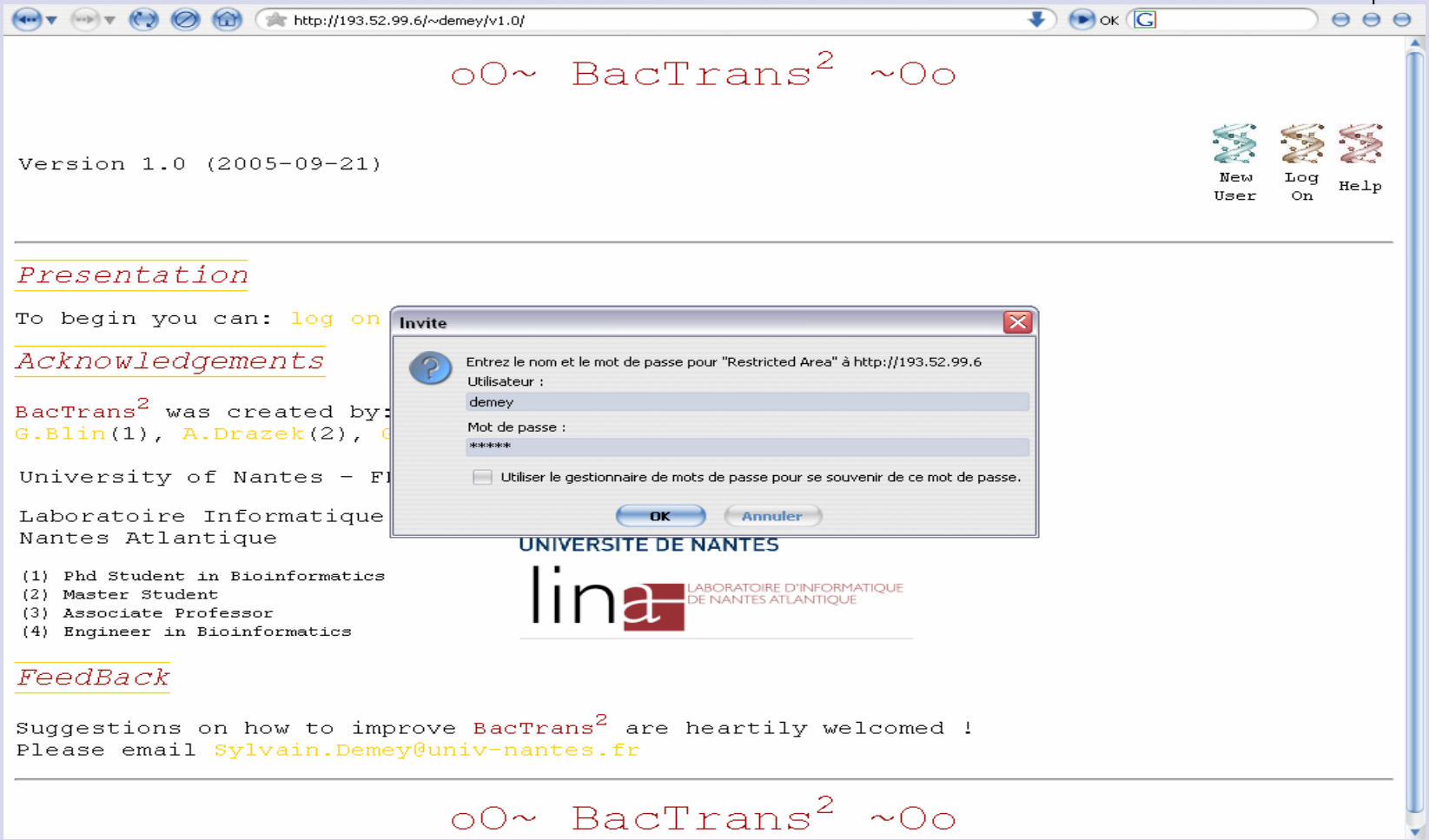
La Plate-forme et son utilisation



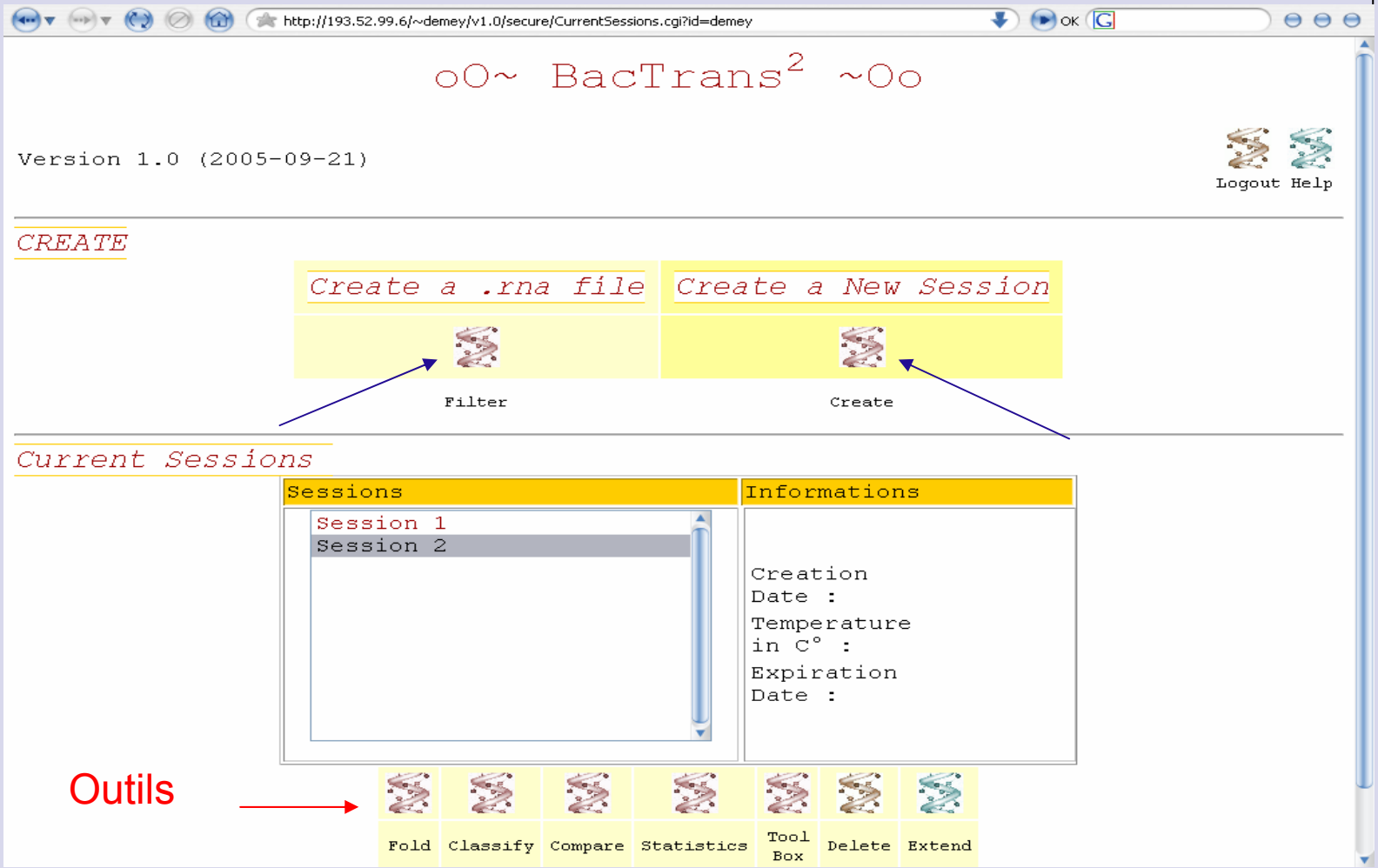
La Plate-forme: création d'un compte



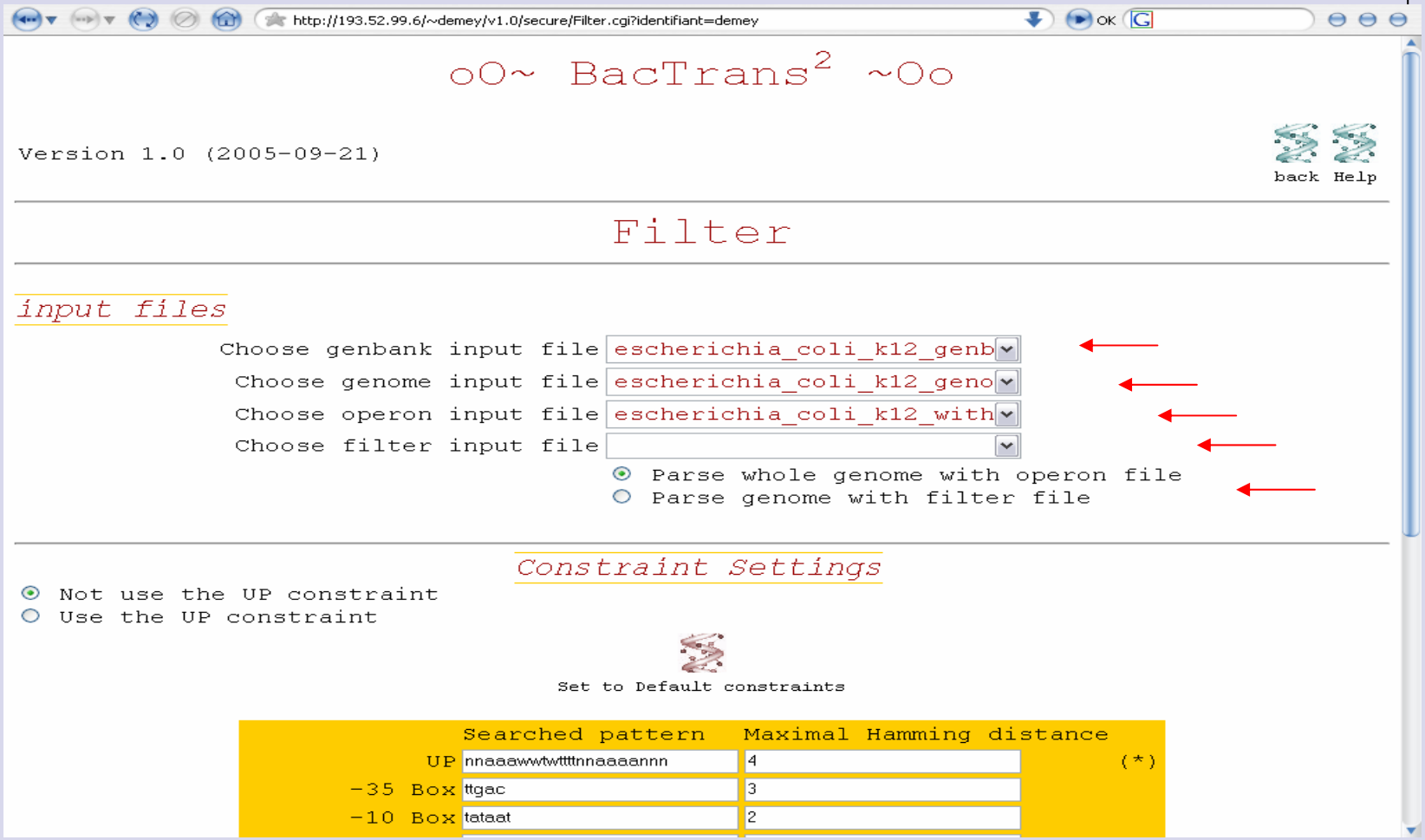
La Plate-forme: ‘log on’



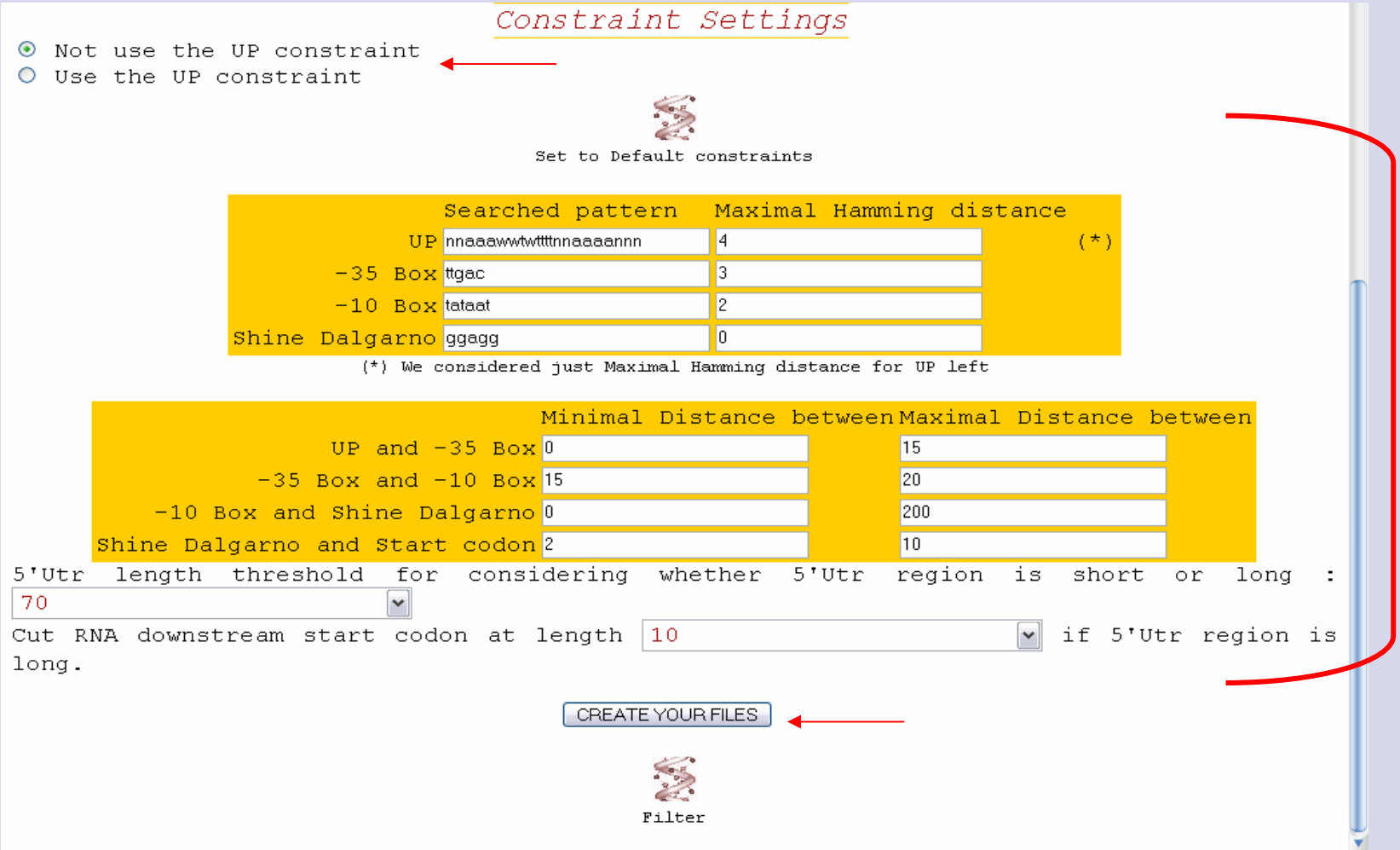
La Plate-forme: la page centrale



La Plate-forme: ‘Filter’



La Plate-forme: ‘Filter’(suite)



Constraint Settings

- (●) Not use the UP constraint
- ( ) Use the UP constraint

Set to Default constraints

	Searched pattern	Maximal Hamming distance	
UP	nnaaawwtwttttnnaaaannn	4	(*)
-35 Box	ttgac	3	
-10 Box	tataat	2	
Shine Dalgarno	ggagg	0	

(*) We considered just Maximal Hamming distance for UP left

	Minimal Distance between	Maximal Distance between
UP and -35 Box	0	15
-35 Box and -10 Box	15	20
-10 Box and Shine Dalgarno	0	200
Shine Dalgarno and Start codon	2	10

5'Utr length threshold for considering whether 5'Utr region is short or long : 70

Cut RNA downstream start codon at length 10 if 5'Utr region is long.

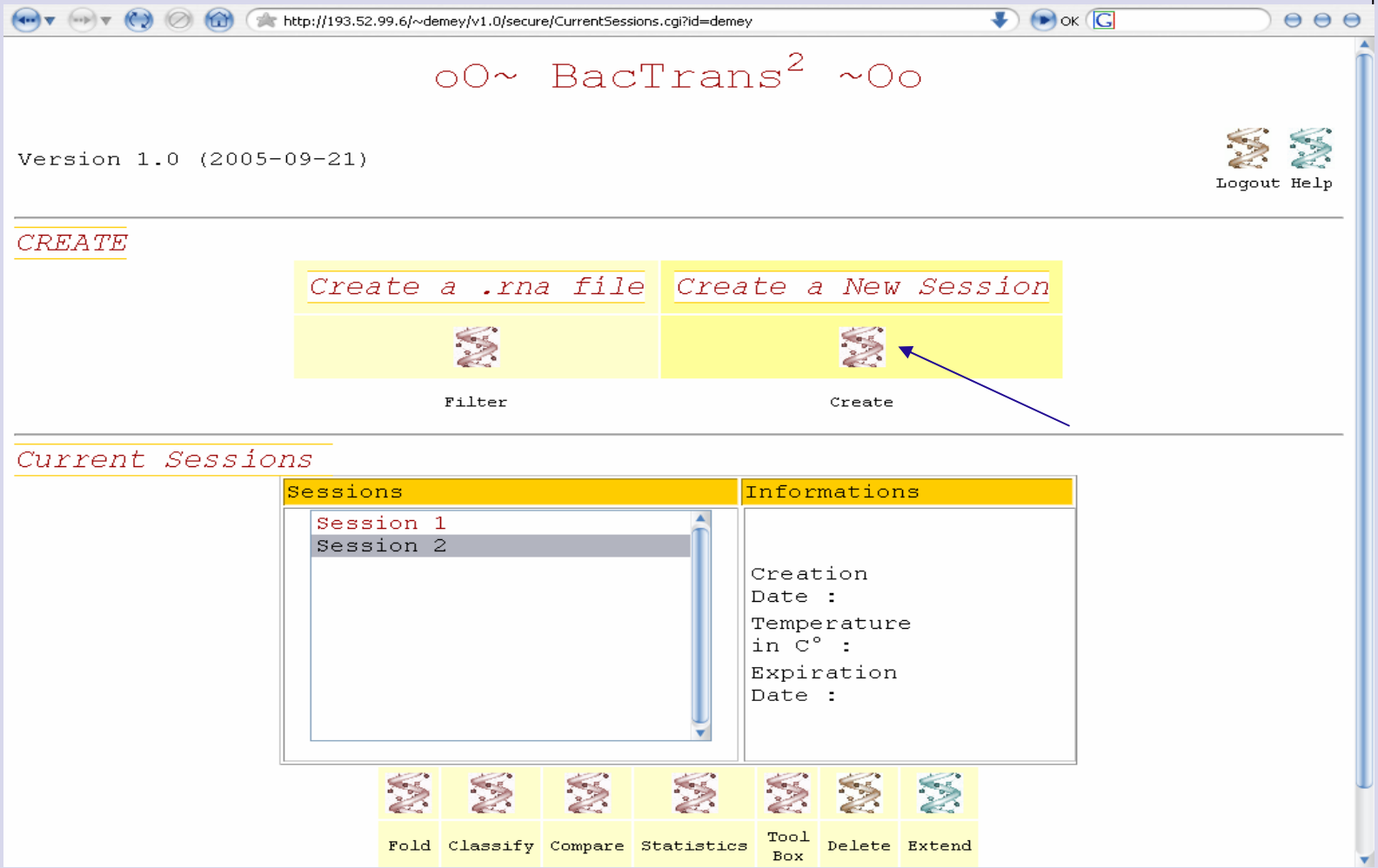
CREATE YOUR FILES

Filter

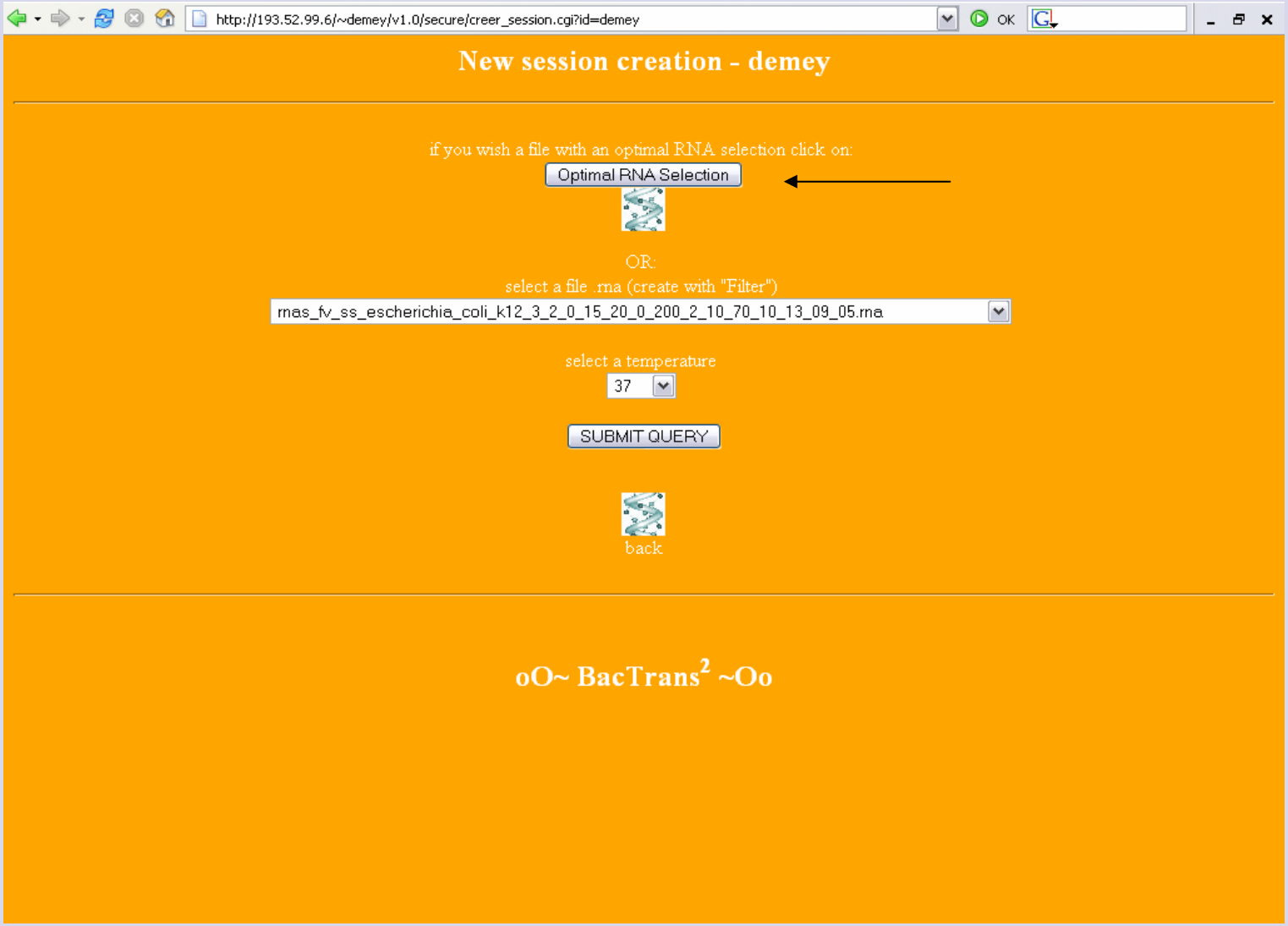
La Plate-forme: ‘Filter’(suite)

/gene/ TM1016 1
/5'UTR length/ 14
/global error score/ 5.0
/UP left errors/ 2
/UP right errors/ 2
/errors (-35 box)/ 1
/errors (-10 box)/ 1
/errors (SD box)/ 0
/distance UP,-35 box/ 3
/distance -35 box,-10 box/ 17
/distance -10 box, Shine-Dalgarno/ 8
/distance Shine-Dalgarno, start codon/ 9
/positions for UP w.r.t. +1 transcript./ -61 -40
/positions for -35 box w.r.t. +1 transcript./ -36 -32
/positions for -10 box w.r.t. +1 transcript./ -14 -9
/positions for Shine-Dalgarno box w.r.t. +1 transcript./ 1 5
/position for start codon w.r.t. +1 transcript./ 15
/-35 box/ ttgaa
/tataBox/ tttaat
/Shine-Dalgarno box/ ggagg
/start codon/ atg
/leftContext Shine-Dalgarno box/ tagcc
/rightContext Shine-Dalgarno box/ tgatg
/leftContextStartCodon/ gtgag
/rightContextStartCodon/ tatag
ggaggtgatgtgagatgtatagagataaaattgcaaaaattatcggtactgtcttttcggcattgacagtacttccactt

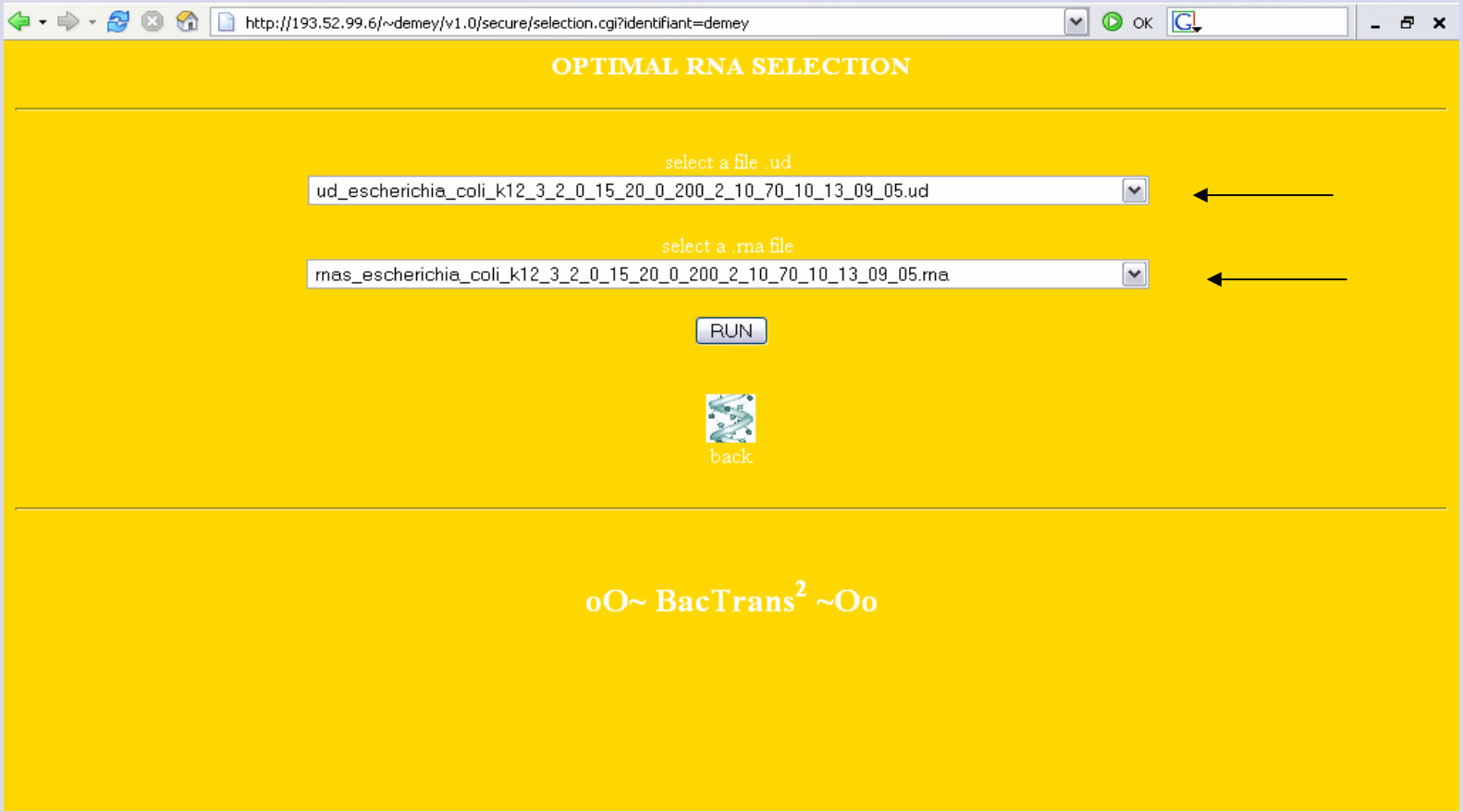
La Plate-forme : création de session



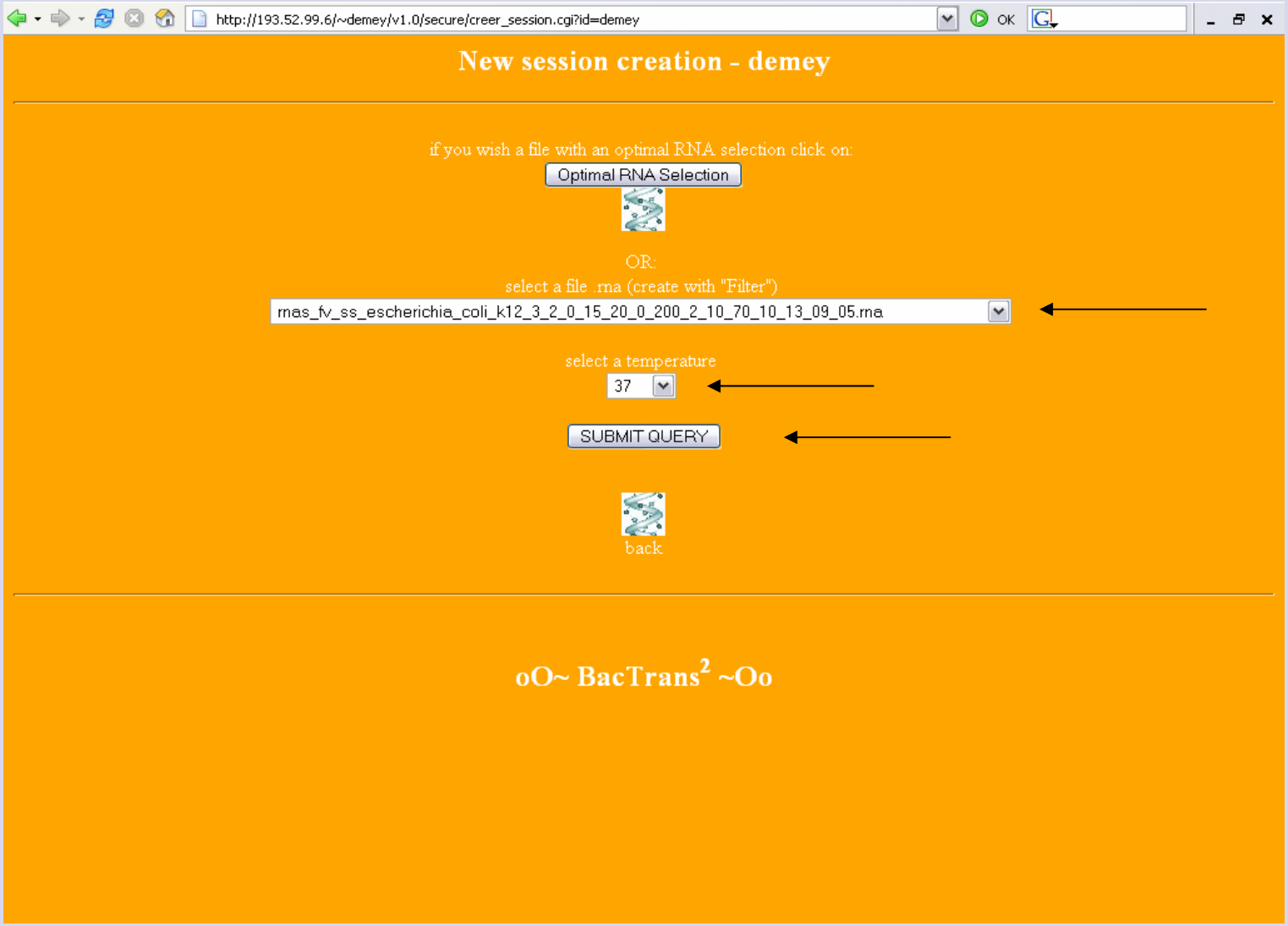
La Plate-forme : création de session (suite)



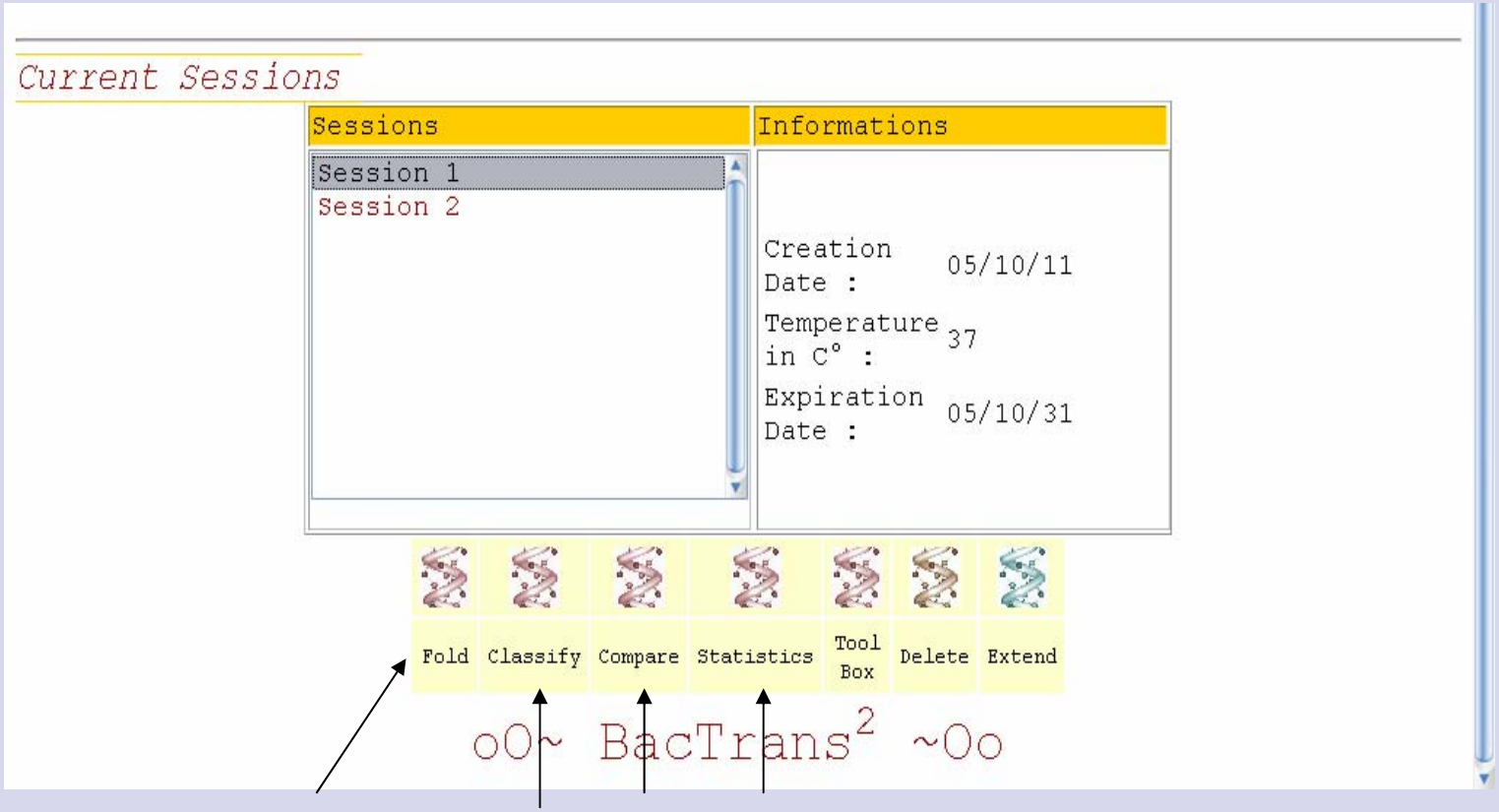
La Plate-forme : création de session (suite)



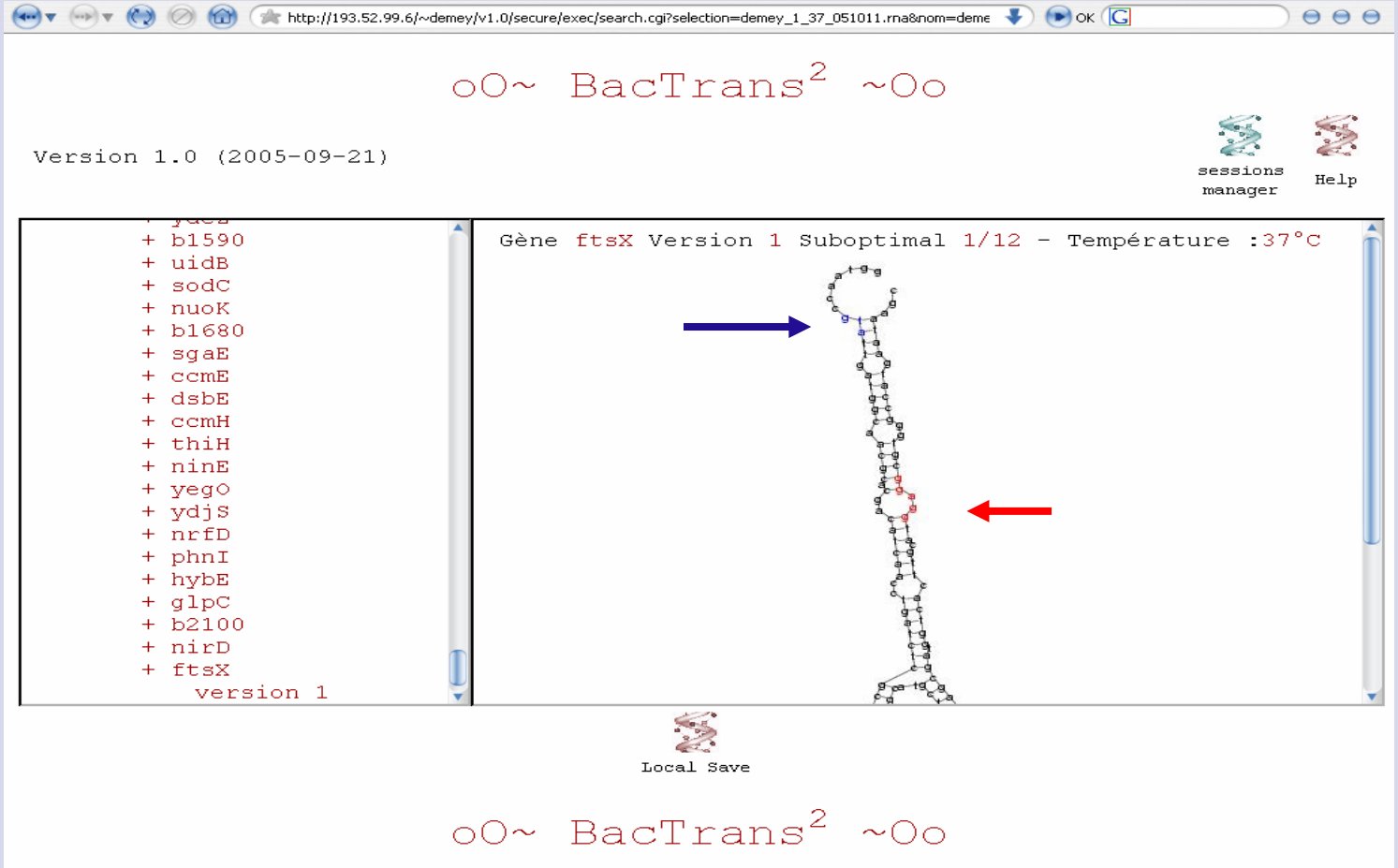
La Plate-forme : création de session (suite)



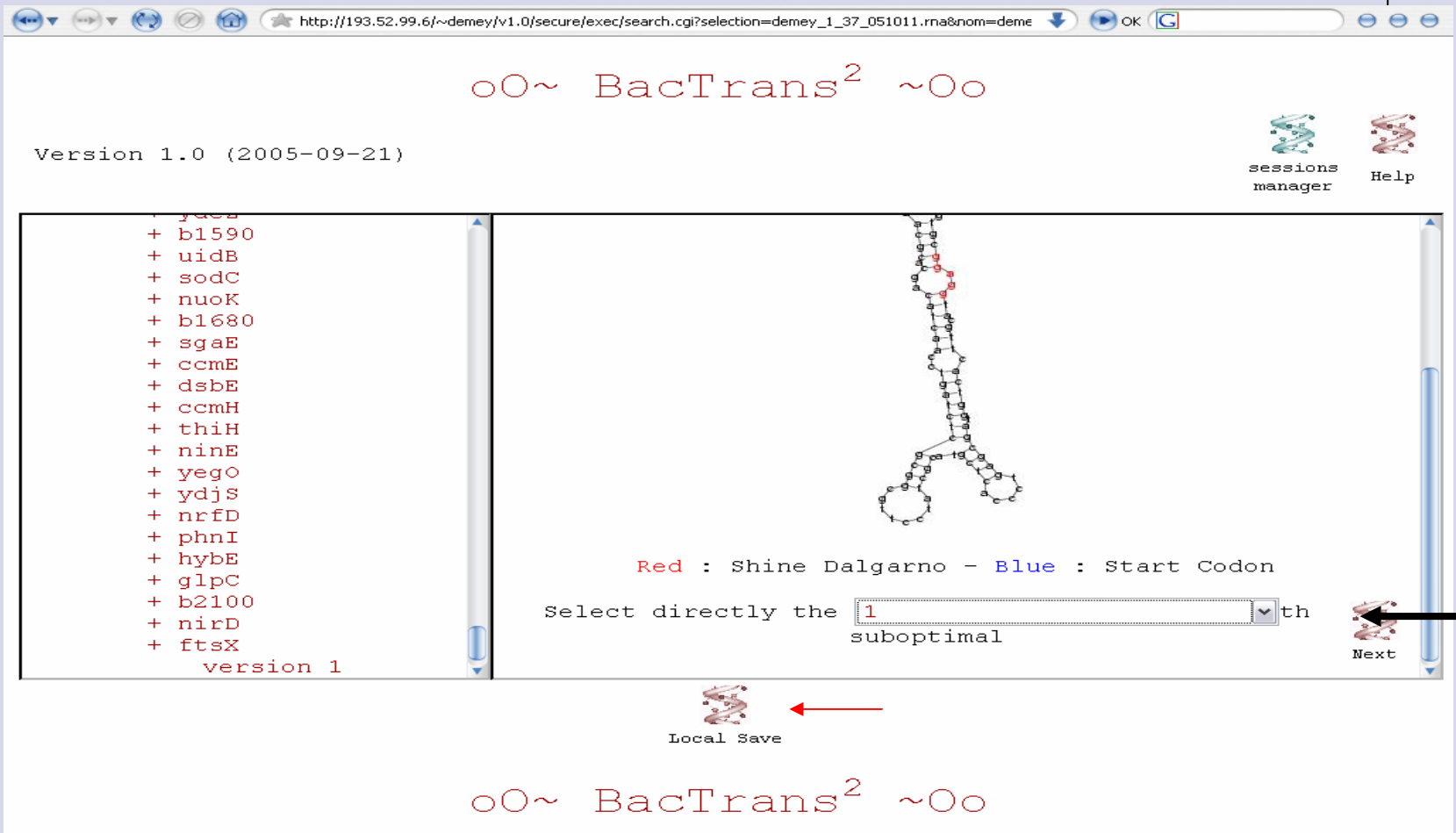
La Plate-forme : Les Outils



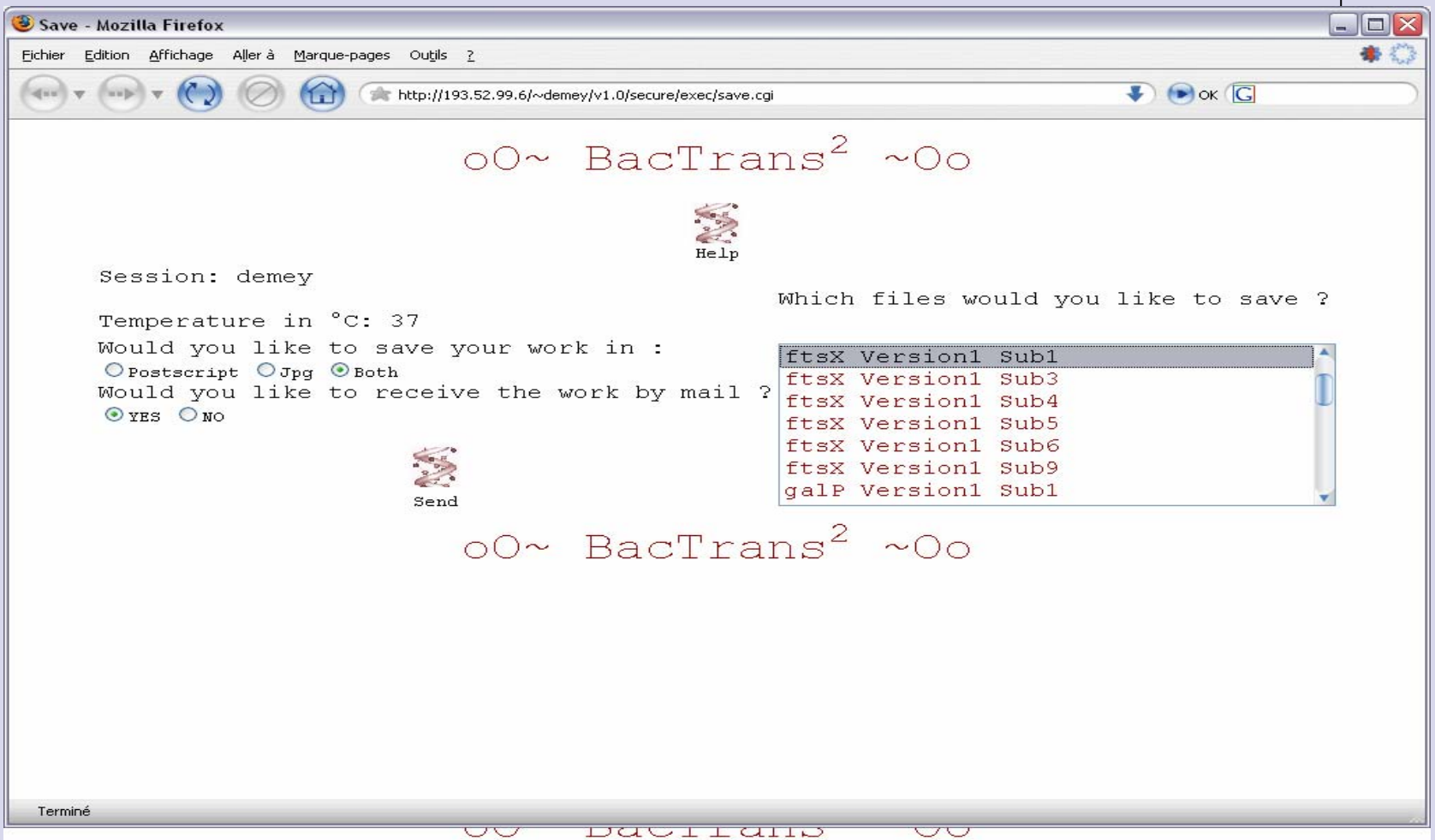
La Plate-forme : ‘Fold’



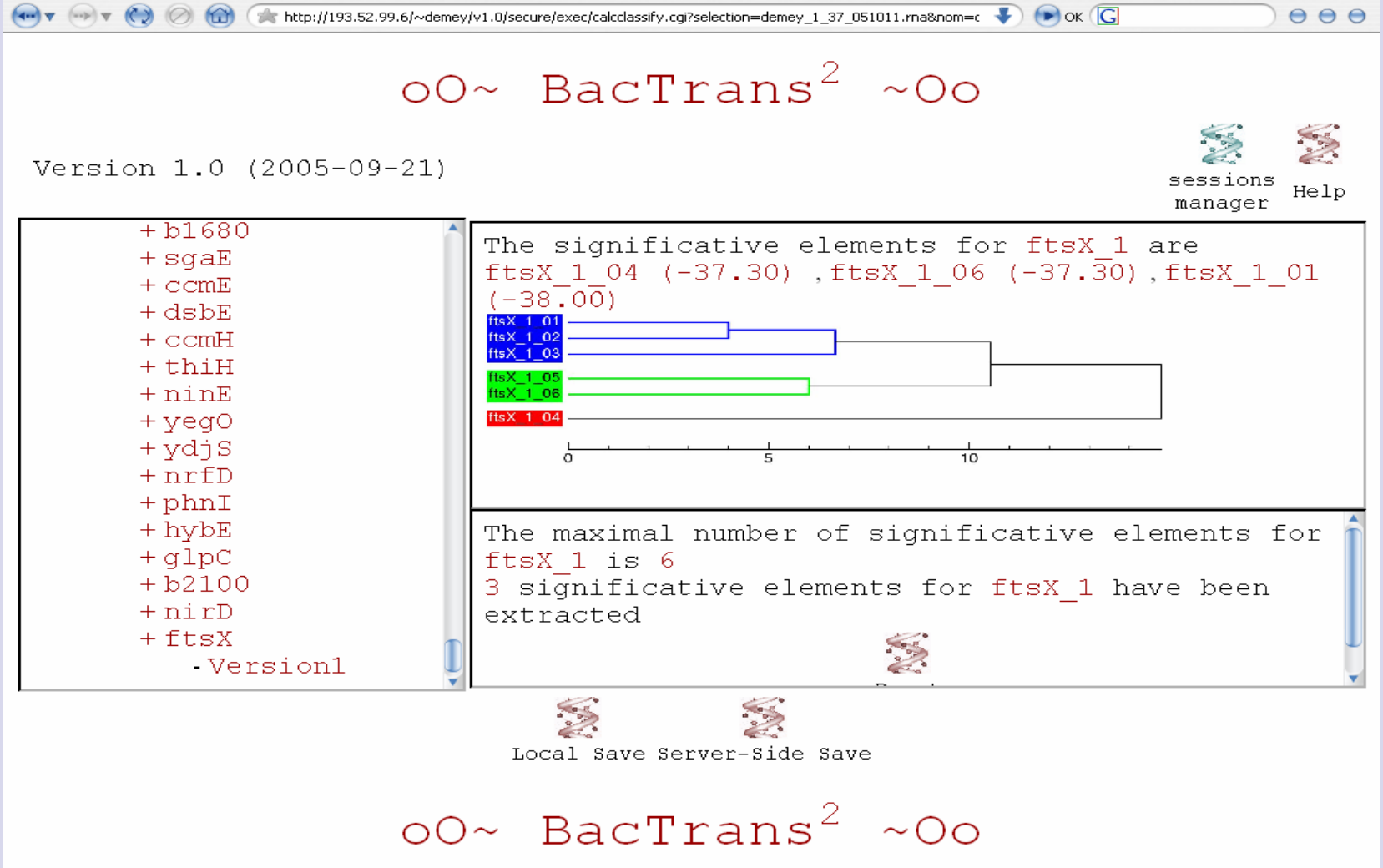
La Plate-forme : ‘Fold’(suite)



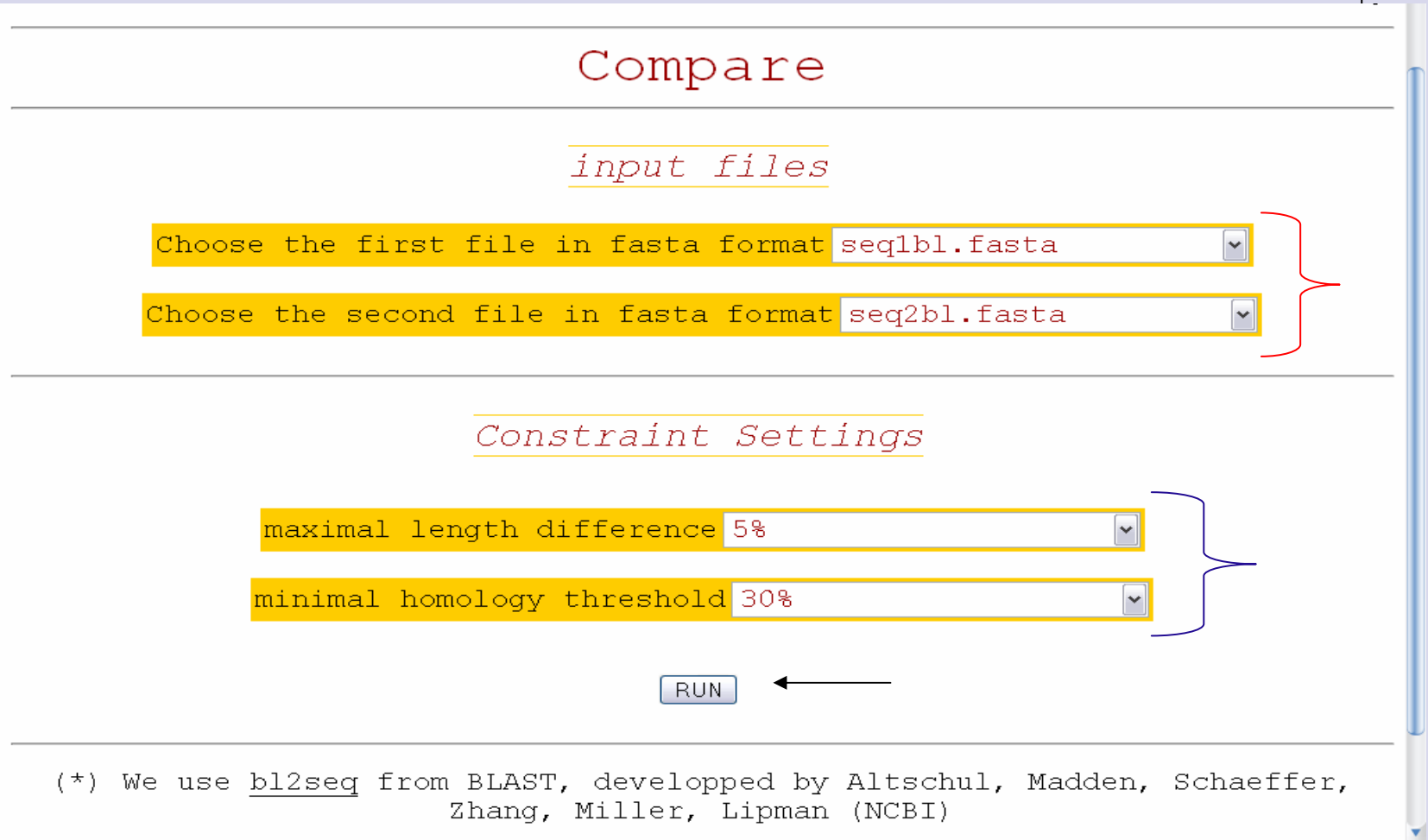
La Plate-forme : ‘Fold’(suite)



La Plate-forme : ‘Classify’



La Plate-forme : ‘Compare’



La Plate-forme : ‘Compare’(suite)

http://193.52.99.6/~demey/v1.0/sqce/result_demey.html?id=demey

oO~ BacTrans^2 ~Oo

Result of Comparison

Query= PAB2026 (387 letters)

```
>PAB2026
          Length = 387

 Score =  753 bits (1944), Expect = 0.0
 Identities = 387/387 (100%), Positives = 387/387 (100%)

Query: 1   MPTNVTAEYLAAEEEYRNAKTIPEKIRALEKMYALVPKHKGTEKLRLQIKRKLAELRKEL 60
           MPTNVTAEYLAAEEEYRNAKTIPEKIRALEKMYALVPKHKGTEKLRLQIKRKLAELRKEL
Sbjct: 1   MPTNVTAEYLAAEEEYRNAKTIPEKIRALEKMYALVPKHKGTEKLRLQIKRKLAELRKEL 60

Query: 61  EKQRQQRKGGGYSLAVKKEGAAQIVLVGLPNVGKSELLRALTGVDVESADYPFTTTEPIP 120
           EKQRQQRKGGGYSLAVKKEGAAQIVLVGLPNVGKSELLRALTGVDVESADYPFTTTEPIP
Sbjct: 61  EKQRQQRKGGGYSLAVKKEGAAQIVLVGLPNVGKSELLRALTGVDVESADYPFTTTEPIP 120

Query: 121 AMMNYKDVQIQLVEVPGLIEGAALGKGMGPQLLAVIRNADAIAIVIDLSEDPIKQMKILL 180
           AMMNYKDVQIQLVEVPGLIEGAALGKGMGPQLLAVIRNADAIAIVIDLSEDPIKQMKILL
Sbjct: 121 AMMNYKDVQIQLVEVPGLIEGAALGKGMGPQLLAVIRNADAIAIVIDLSEDPIKQMKILL 180

Query: 181 REFERAGIKVNKRRPRIEIRRTPSGGIVINGIENIKGDINEVMKMLREEKIHSAEITVRE 240
           REFERAGIKVNKRRPRIEIRRTPSGGIVINGIENIKGDINEVMKMLREEKIHSAEITVRE
Sbjct: 181 REFERAGIKVNKRRPRIEIRRTPSGGIVINGIENIKGDINEVMKMLREEKIHSAEITVRE 240

Query: 241 PVTLEDFADALDESLVWKKAIIIANKGDAPGSKDNYEKLVKAYGERFKIIPVSAKKRINL 300
           PVTLEDFADALDESLVWKKAIIIANKGDAPGSKDNYEKLVKAYGERFKIIPVSAKKRINL
Sbjct: 241 PVTLEDFADALDESLVWKKAIIIANKGDAPGSKDNYEKLVKAYGERFKIIPVSAKKRINL 300

Query: 301 DAVKEALYDVAEIIRVFTKSPGEEPAYPPIALKRGSTVLDVAERIHKDLVKNFRYARVWG 360
           DAVKEALYDVAEIIRVFTKSPGEEPAYPPIALKRGSTVLDVAERIHKDLVKNFRYARVWG
Sbjct: 301 DAVKEALYDVAEIIRVFTKSPGEEPAYPPIALKRGSTVLDVAERIHKDLVKNFRYARVWG 360

Query: 361 KSVKFPGQRVGPDHVLEDGDIVEIHAR 387
           KSVKFPGQRVGPDHVLEDGDIVEIHAR
Sbjct: 361 KSVKFPGQRVGPDHVLEDGDIVEIHAR 387
```

Lambda K H 0.316 0.136 0.378 Gapped Lambda K H 0.267 0.0410 0.140 Matrix: BLOSUM62 Gap Penalties: Existence: 11, Extension: 1 Number of Hits to

http://193.52.99.6/~demey/v1.0/sqce/result_demey.html?id=demey

```
           DAVKEALYDVAEIIRVFTKSPGEEPAYPPIALKRGSTVLDVAERIHKDLVKNFRYARVWG
Sbjct: 298 DAVKEALYDVAEIIRVFTKSPGEEPAYPPIALKRGSTVLDVAERIHKDLVKNFRYARVWG 357

Query: 361 KSVKFPGQRVGPDHVLEDGDI 381
           KSVKFPGQRVGPDHVLEDGDI
Sbjct: 358 KSVKFPGQRVGPDHVLEDGDI 378
```

Lambda K H 0.316 0.136 0.378 Gapped Lambda K H 0.267 0.0410 0.140 Matrix: BLOSUM62 Gap Penalties: Existence: 11, Extension: 1 Number of Hits to DB: 777 Number of Sequences: 0 Number of extensions: 13 Number of successful extensions: 6 Number of sequences better than 10.0: 1 Number of HSP's better than 10.0 without gapping: 1 Number of HSP's successfully gapped in prelim test: 0 Number of HSP's that attempted gapping in prelim test: 0 Number of HSP's gapped (non-prelim): 1 length of query: 387 length of database: 378 effective HSP length: 30 effective length of query: 357 effective length of database: 348 effective search space: 124236 effective search space used: 124236 T: 11 A: 40 X1: 16 (7.3 bits) X2: 38 (14.6 bits) X3: 64 (24.7 bits) S1: 24 (13.8 bits) S2: 24 (13.9 bits)

Query= PAB1296 (309 letters)

```
>PAB2026
          Length = 387

 Score = 20.8 bits (42), Expect = 0.055
 Identities = 11/47 (23%), Positives = 22/47 (46%)
```

Identities < threshold of minimal homology

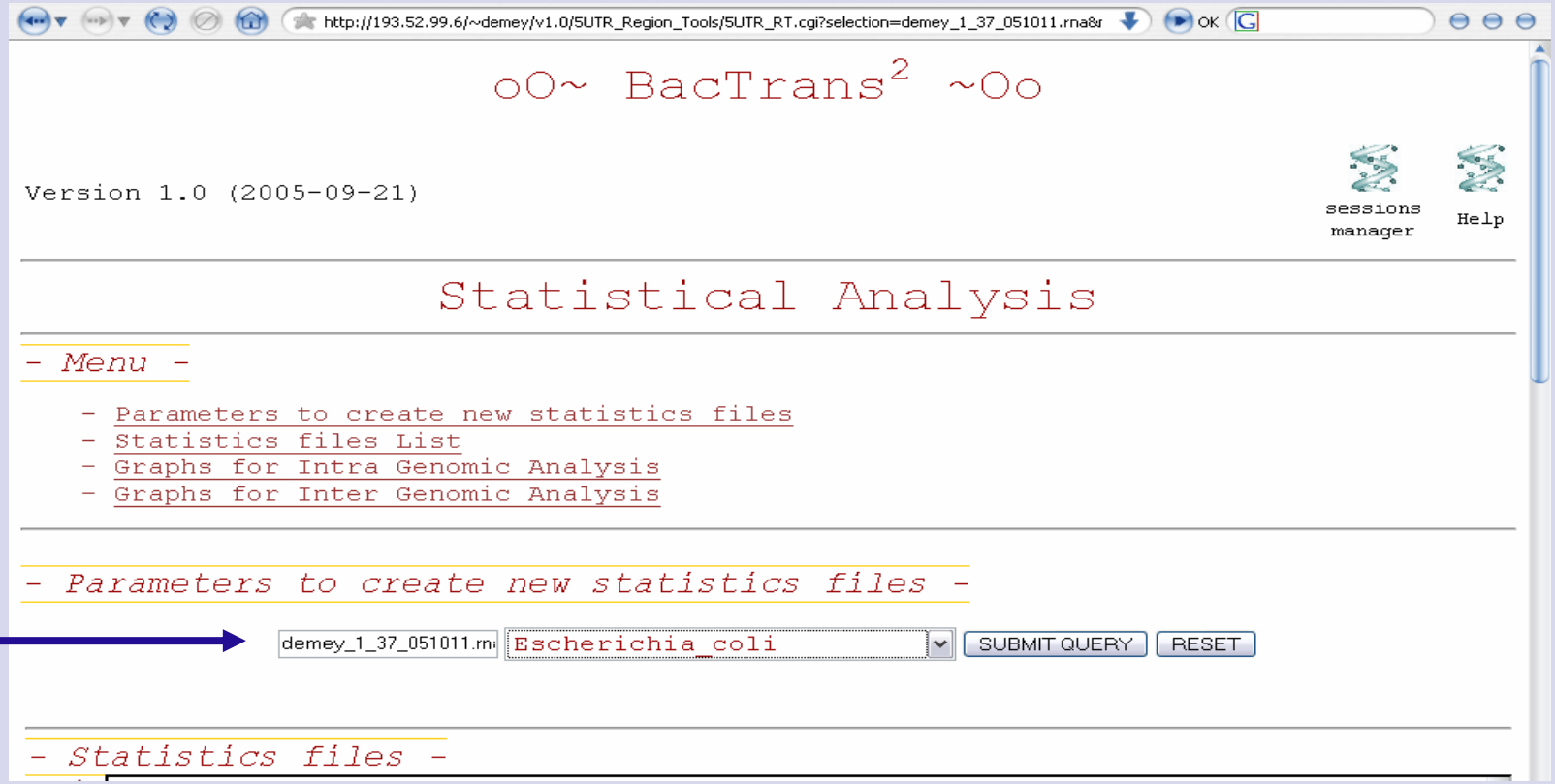
Query= PAB1296 (309 letters) Lambda K H 0.320 0.139 0.383 Gapped Lambda K H 0.267 0.0410 0.140 Matrix: BLOSUM62 Gap Penalties: Existence: 11, Extension: 1 Number of Hits to DB: 109 Number of Sequences: 0 Number of extensions: 4 Number of successful extensions: 0 Number of sequences better than 10.0: 0 Number of HSP's better than 10.0 without gapping: 0 Number of HSP's successfully gapped in prelim test: 0 Number of HSP's that attempted gapping in prelim test: 0 Number of HSP's gapped (non-prelim): 0 length of query: 309 length of database: 177 effective HSP length: 23 effective length of query: 286 effective length of database: 154 effective search space: 44044 effective search space used: 44044 T: 11 A: 40 X1: 16 (7.4 bits) X2: 38 (14.6 bits) X3: 64 (24.7 bits) S1: 20 (12.1 bits) S2: 20 (12.3 bits)

Query= PAB1296 (309 letters)

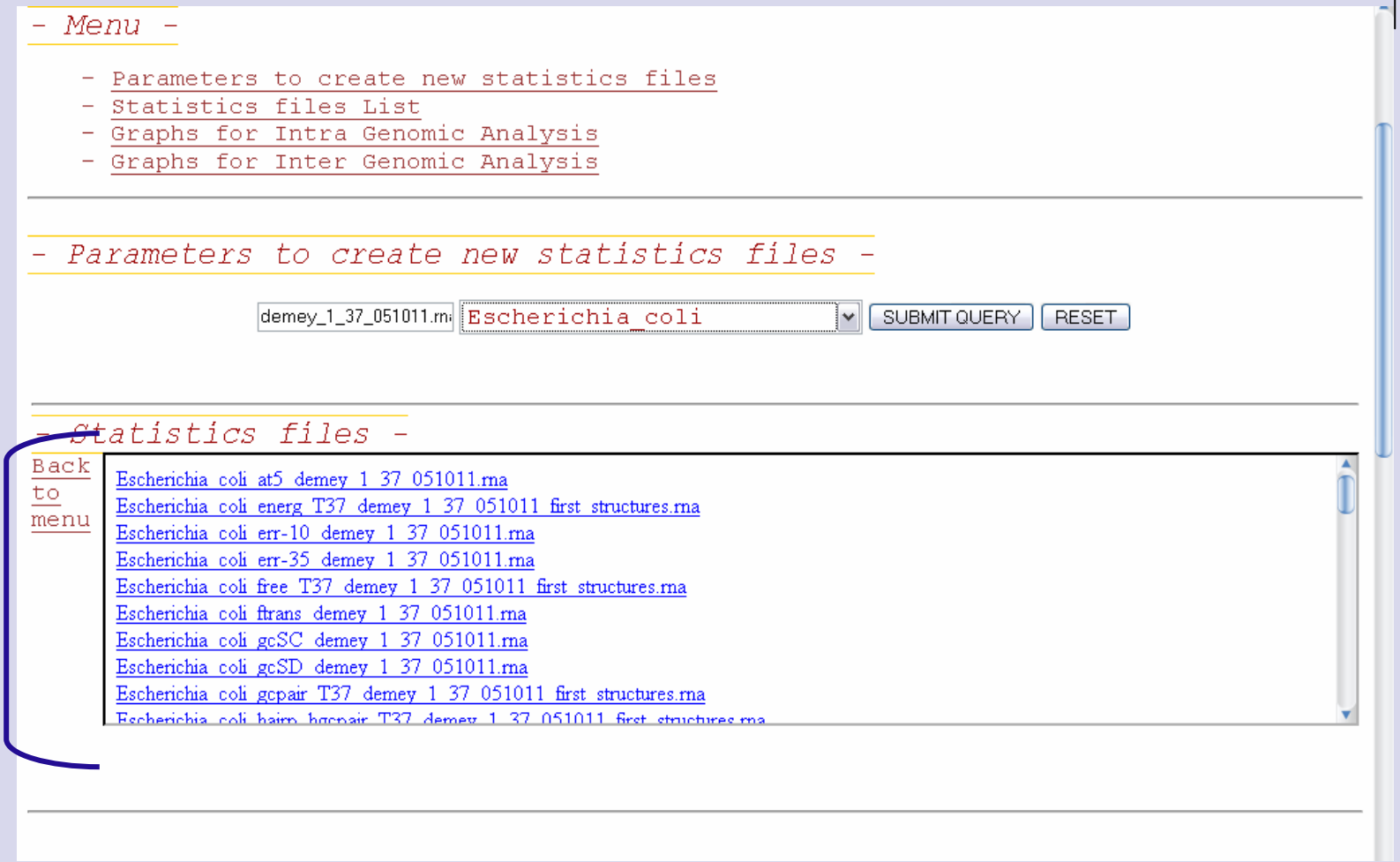
```
>PAB1296
          Length = 309

 Score =  596 bits (1537), Expect = e-175
 Identities = 309/309 (100%), Positives = 309/309 (100%)
```

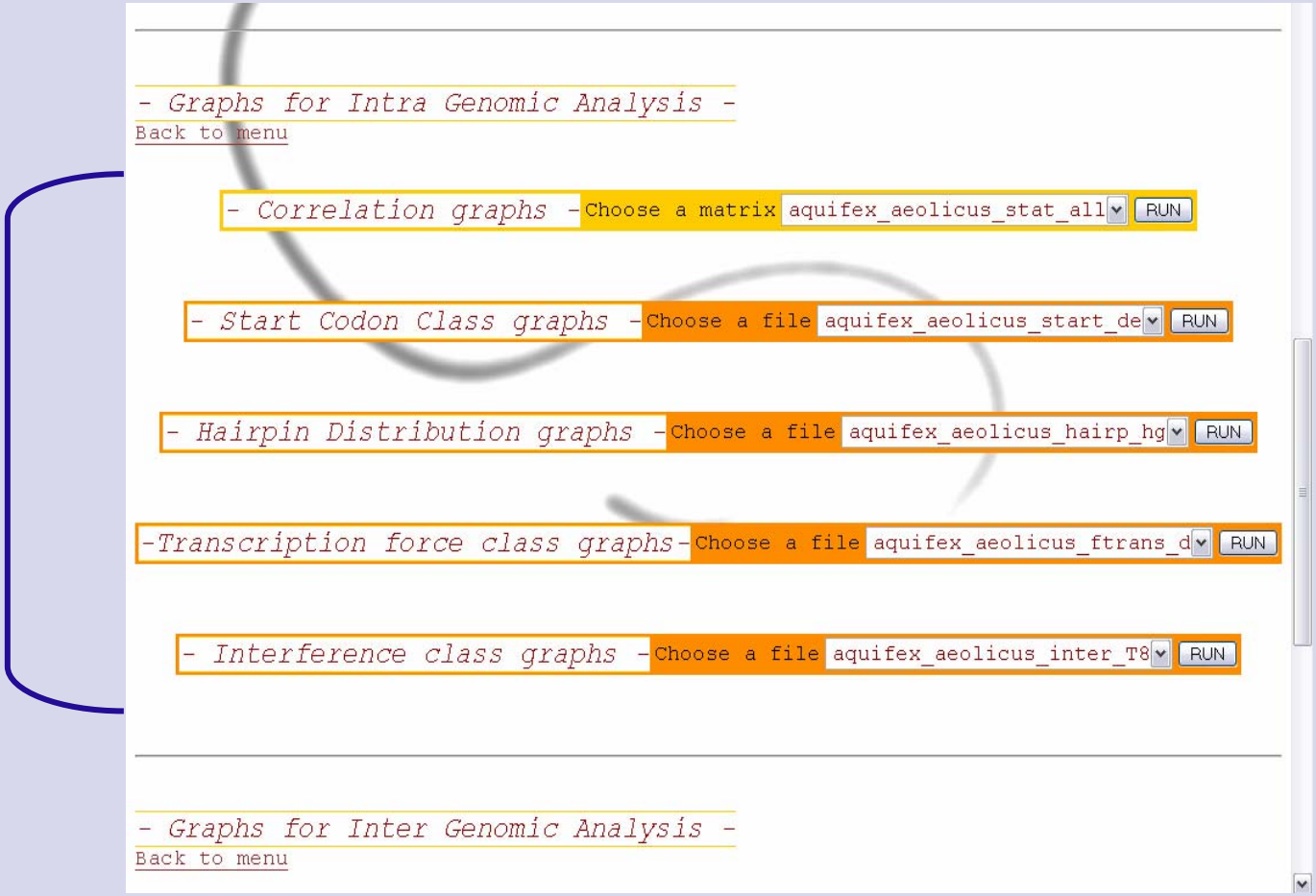
La fonctionnalité ‘Statistiques’



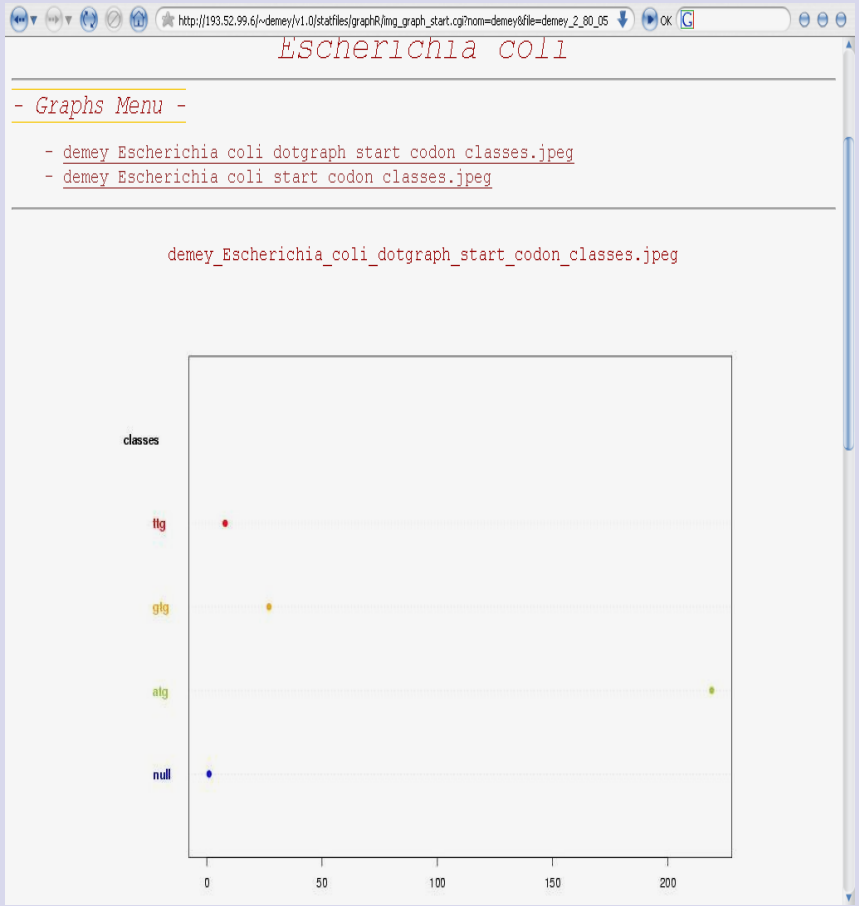
La fonctionnalité ‘Statistiques’ (suite)

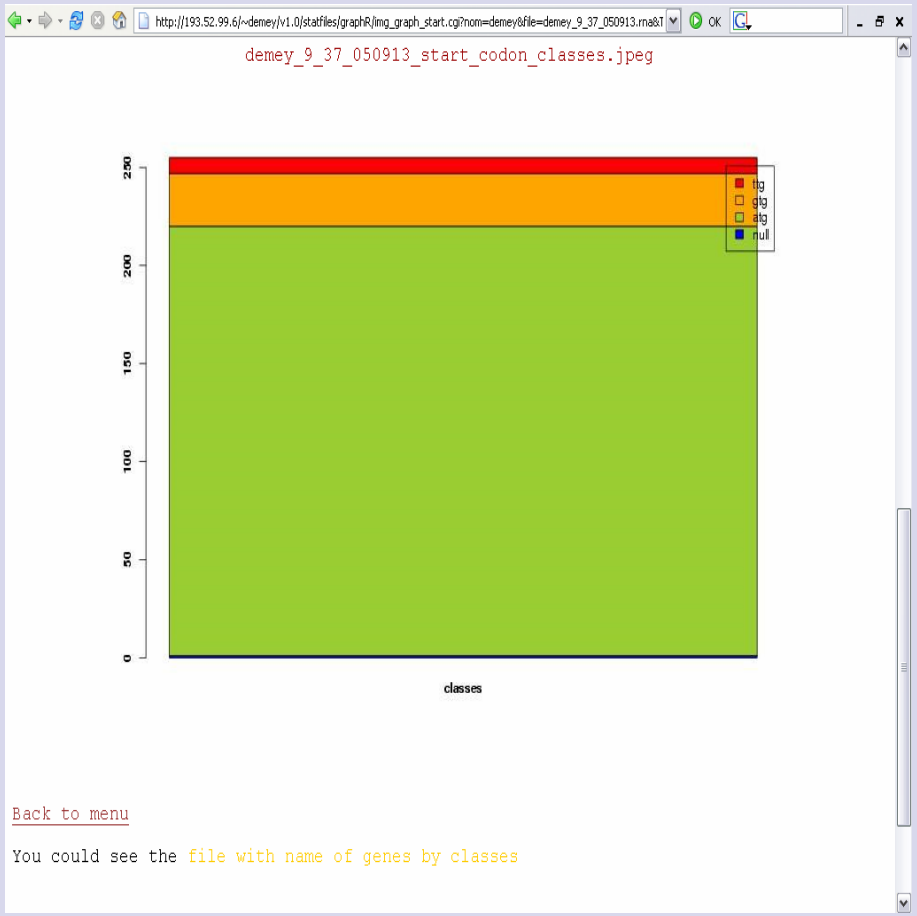


La fonctionnalité ‘Statistiques’ (suite)

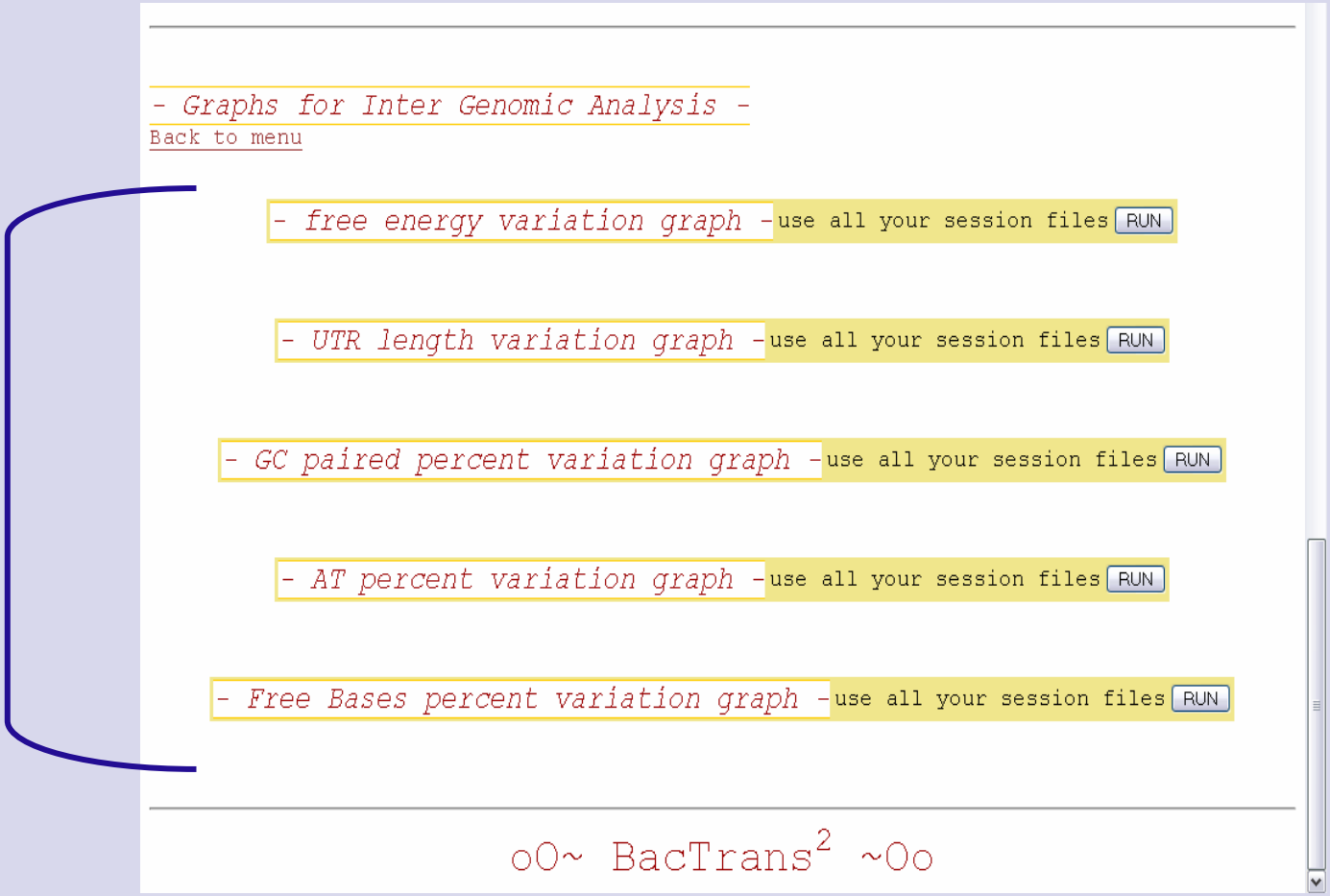


La fonctionnalité ‘Statistiques’ : exemple de graphiques du point de vue Intra Génomique

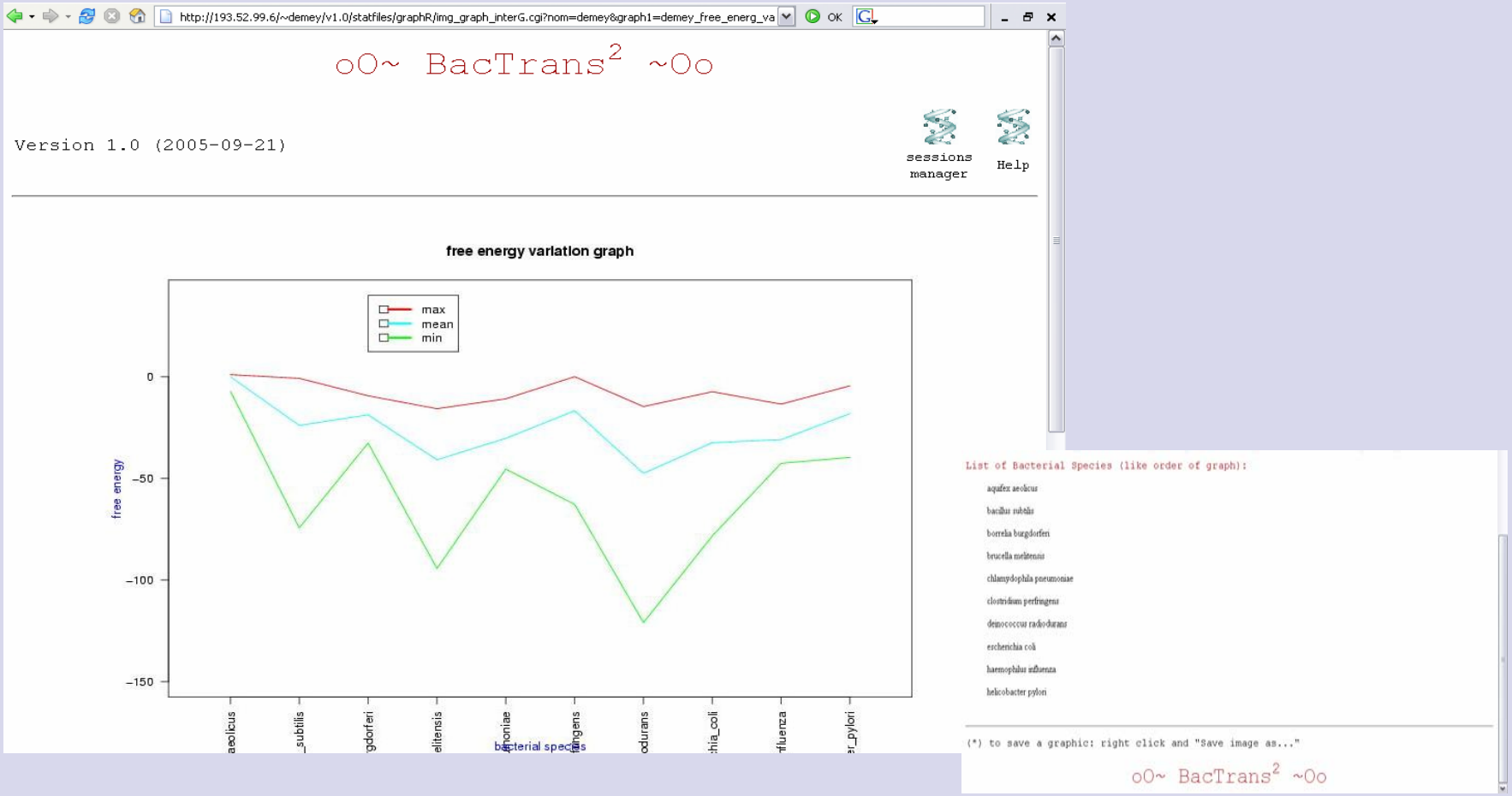




La fonctionnalité ‘Statistiques’ (suite)



La fonctionnalité ‘Statistiques’: exemple de graphiques du point de vue Inter Génomique

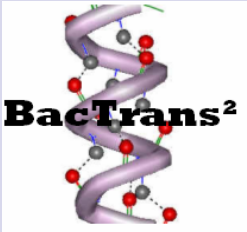


Conclusion-Perspectives

- Finaliser la plate-forme
  - Automatisation de certaines tâches
  - Etudes statistiques pour valider la plate-forme
  - La mettre à disposition
- La soumettre aux biologistes
  - Ajout de nouvelles espèces à leurs demandes
  - Développer de nouveaux outils suivant leurs besoins
- → Plate-forme évolutive

Merci de votre attention

~o0 BacTrans² 0o~



Sylvain.Demey@univ-nantes.fr