

Sortie du programme show_emfit d'estimation des paramètres quand votre fichier modèle ne contient plus d'erreur

Moyle

gwenlaere.richard@biocy.biotoul.fr (guest)
set email | sign-in | sign-out
refresh workspace

The screenshot shows the Moyle web interface. On the left is a navigation sidebar with sections: Programs (alignment, annotation, enrichment, genes, show_emfit, show_viterbi, sequence), Data Bookmarks, Text (sigmanew_model), Jobs (show_emfit - 10/14/16 10:06:58), and Tutorials. The main content area has a top navigation bar with 'Welcome', 'Forms', 'Data Bookmarks', 'Jobs', and 'Tutorials'. Below this is a job overview for 'show_emfit - 10/14/16 10:06:58'. A yellow warning banner states: 'Your job finished with an unusual status code (1), check your results carefully.' Below the warning are buttons for 'get help', 'back to form', 'remove job', and 'download'. The 'results' section contains two panels. The first panel, 'Learn model (Learnmodel)', shows the output of the 'sigma_learn.model' job, including state and transition information. The second panel, 'Learn select likelihoods (Text)', shows the output of the 'sigma_learn.select.likelihoods' job, listing three models with their loglikelihoods and identifying the best model. At the bottom of each panel are buttons for 'full screen view' and 'bookmark as [job name] or [other job name] further analysis'. Two blue arrows point from the 'sigma_learn.model' and 'sigma_learn.select.likelihoods' labels in the job overview to the corresponding panels in the results section.

Sauver le fichier sigma_learn.model qui contient votre modèle avec les probabilités estimées

Choisir show_viterbi et cliquer sur further analysis . Cela lancera automatiquement la prédiction avec l'algorithme de Viterbi en utilisant le modèle appris sur le jeu de séquences utilisés auparavant pour l'apprentissage

Après avoir cliquer sur further analysis, vous vous trouverez sur cette page, vous permettant de lancer le programme de Viterbi

Mobyle

set email | sign-in | sign-out
refresh workspace

Search [more]

Programs

- alignment
- annotation
- enrichment
- genes
 - show_emfit
 - show_viterbi
- sequence

Data Bookmarks [overview]

Learnmodel: sigma_learn.model
Text: sigmanew_model

Jobs [overview]

✓ show_emfit - 10/14/16 10:06:58

Tutorials

- How to use Mobyle? A step by step tutorial
- Registration information
- Sequence formats
- Alignment formats

Welcome Forms Data Bookmarks Jobs Tutorials

show_emfit x show_viterbi x

show_viterbi 2.0

SHOW: Structured HOMogeneities Watcher. Prediction Step

Run Reset

Inputs options

- Model

paste upload result EDIT CLEAR

select

```
BEGIN_STATE
state_id: bound
BEGIN_TRANSITIONS
  type: 1
  state: background_1
  ptrans: 1
```

References :

AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system

Author(s): K. Bryson, V. Loux, R. Bossy, P. Nicolas, S. Chaillou, M. van de Guchte, S. Penaud, E. Maguin, M. Hoebeke, P. Bessières and J-F Gibrat

Le programme Viterbi fournit en sortie un fichier résultat par séquence analysée. Pour récupérer simultanément tous les résultats, cliquer sur download et enregistrer l'archive.

Il faudra par la suite extraire les fichiers de cette archive. Vous obtiendrez une liste de fichiers avec l'extension **.vit**

Pour exécuter le parser, il faudra également extraire les fichiers de l'archive fournie dans le TD et qui contient les fichiers de séquences ADN ayant servies à l'apprentissage des probabilités et analysées ensuite par le Viterbi pour prédire la localisation des promoteurs dans chacune d'entre-elles. Vous obtiendrez un répertoire **seq** qui contiendra des fichiers avec l'extension **.dna**.

Le parser requiert une comparaison entre le fichier **.dna** d'une séquence et le fichier **.vit** qui lui est associé. Copier les fichiers **.vit** dans le répertoire **seq**

Mobyle

set email | sign-in | sign-out
refresh workspace

The screenshot displays the Mobyle web interface. On the left is a navigation sidebar with sections: Programs (alignment, annotation, enrichment, genes, show_emfit, show_viterbi, sequence), Data Bookmarks (sigma_learn.model, sigmanew_model), Jobs (show_emfit - 10/14/16 10:06:58, show_viterbi - 10/14/16 10:13), and Tutorials (How to use Mobyle?, Registration information, Sequence formats, Alignment formats). The main content area shows a job completion message: "Your job finished with an unusual status code (1), check your results carefully." Below this is a list of prediction files: alsR.vit, f29_G2.vit, f29_G3a.vit, and glpD.vit. Each file entry includes a "save" button, a "full screen view" button, a "bookmark" button, a text input field with the filename, an "or" separator, a dropdown menu, and a "further analysis" button. A yellow warning banner is present for each file, stating: "This file cannot be displayed because too many have been produced by this job. Click here to display this result in a separate window."

Modifier le programme pour indiquer le chemin où se trouve vos fichiers **.dna** et **.vit** (donc tous le chemin jusqu'au répertoire **seq**

```
1  #!/usr/bin/perl -w
2
3  use strict;
4
5  my $rep = '//home/fichant/M1MABS/BGPG/TD_HMM/2016/';
6
7  my @motif_liste = ( 'sigma-35+', 'sigma-10+' );
8
9  opendir(DIR, $rep) or warn "Cannot open $rep: $!";
10 my @liste_of_files = grep /\.vit$/, sort(readdir DIR);
11 if ( scalar @liste_of_files == 0 ) {
12     print "pas de fichier trouve!\n";
13     exit(1);
14 }
15
16 close(DIR);
17
18 my $stp_sigma_35 = 0;
19 my $stp_sigma_10 = 0;
20 my $stp_sigma_35alt = 0;
21 my $stp_sigma_10alt = 0;
22 my $tot = 0;
23
```

Ligne à modifier

Pour lancer votre parser et récupérer les résultats dans un fichier :

```
./parser_viterbi.pl > resultat_viterbi.txt
```

Vous obtiendrez les résultats sous cette forme :

Nom du gène position prédite de la boîte -35, sa position réelle dans la séquence et le motif dans la séquence puis la position prédite, réelle et le motif dans la séquence de la boîte -10 et finalement la taille prédite et réelle du linker entre les deux boîtes

aadK	29	16	ATCACA	52	40	TATAAG	17	18
ald	73	68	AAGAAT	96	91	TACACT	17	17
alsR	55	68	TCGATT	80	91	TGTAAT	19	17
amyLY	54	68	TTGTTT	77	91	TAATTT	17	17
bglu	75	67	TTGGCC	98	91	AATCAT	17	18
cad_P2	13	68	TTGAGA	36	93	TATTAT	17	19
cheR	77	68	TTGAAA	100	90	TAAAAC	17	16
citA_	68	68	TTGATT	96	91	TTTACT	22	17
citC	63	68	CAGACA	86	91	TTTAAA	17	17
...	--	--	-----	--	--	-----	--	--

Si vous ouvrez le fichier aadK.dna sur la première ligne du fichier vous aurez :

```
>aadk 143 ATCACA 16 TATAAG 40
```

En résultat final, le parser établit une synthèse entre vos prédictions et les positions réelles pour chacun des deux motifs.
On ne s'occupera pas de la ligne sigma_35alt

Ici 85.17% des motifs sigma_35 et 85.59% des motifs sigma_10 ont été correctement prédit en acceptant un écart de plus ou moins deux bases par rapport à la position réelle (ecart < 3)

```
Statistiques pour un ecart < 3
sigma_35 85.17 sigma_10 85.59
sigma_35alt 0.00 sigma_10alt 0.00
sigma_35tot 85.17 sigma_10tot 85.59
```