

## Tutorial SeedServer

Na aula de text-mining vocês selecionaram alguns símbolos que representam os genes (Ex: GLUD1 representa Glutamate dehydrogenase 1). Podemos utilizar um único gene para construir um grupo de homólogos com a ferramenta SeedServer.

Nossa ferramenta utiliza como dado de entrada identificadores da base de dados protéica UniProt (<http://www.uniprot.org>), logo, precisamos converter o símbolo gênico em um UniProtID (Ex GLUD1 ⇔ P00367).

Para isso, podemos utilizar a busca avançada do site UniProt dado acima, para buscar por “Gene Name” e “Organism” de interesse. Guarde um identificador UniProt resultante dessa pesquisa para utilizarmos no SeedServer.

Um serviço web está em desenvolvimento no nosso servidor biodados (<http://biodados.icb.ufmg.br:8080/SeedServer/>) onde em breve será possível criar e exportar os grupos de homólogos criados pela ferramenta.

No entanto, como estamos em um curso prático, vamos ver como esse site trabalharia para lhe entregar os resultados já processados:

Relembre passo a passo o funcionamento do SeedServer :  
[http://biodados.icb.ufmg.br/cbab2011\\_data/SS\\_fluxograma.pdf](http://biodados.icb.ufmg.br/cbab2011_data/SS_fluxograma.pdf)

Agora vamos rodar alguns deles por linhas de comando:

- Seed Linkage (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2277401/?tool=pubmed>)
  - Em seu /home/cbab2011/SEUNOME crie um diretório chamado SeedLinkage e mova-se para dentro dele.
  - Faça o download do pacote Seed Linkage aqui: ([http://biodados.icb.ufmg.br/cbab2011\\_data/](http://biodados.icb.ufmg.br/cbab2011_data/))

```
wget http://biodados.icb.ufmg.br/cbab2011_data/SL_package.tar
```

- Descompacte com: `tar -xvf SL_package.tar`
- Verifique o arquivo de entrada: `SL_input`
- Como as demais configurações já foram previamente definidas, rode o Seed Linkage com o comando:

```
php batch_seed_linkage.php -i SL_input -d /home/cbab2011/DATABASE/uniprot2010_10 -r 50 -s 50 -c 50 -l 0.3
```

```
php evaluate_similarity.php SL_input.clusters
```

- Consulte o resultado no arquivo chamado: `SL_input.clusters.disambiguated`

- UEKO
  - Como já vimos, os dados que formam o UEKO estão organizados em uma tabela MySQL. Nela é possível consultar a qual grupo KO pertence uma entrada UniProt e também os membros que compõem o enriquecimento desse KO. Essa consulta pode ser feita manualmente ou através de um script, como é o caso do SeedServer.
  
- PSI-BLAST
  - O processo de validação das proteínas recrutadas se dá por um método baseado no algoritmo PSI-BLAST e é centralizado nas seqüências seed (recrutadoras) para pontuação.
  
- Comparação de estrutura secundária (Secondary structure Overlap – SOV)
  - Para cada sequência recrutada, uma predição de estrutura secundária é feita (software Predator) e comparada com todas as seeds presentes no agrupamento, sendo a melhor delas selecionada para compor a tabela final.
  - Vejamos um exemplo de saída do preditor.
 

```
more predator_file
```
  - Vejamos uma tabela criada durante essa etapa.
 

```
select * from seedsov_guedes_1_clade limit 10;
```
  
- SeedServer (1 por dupla)
  - Agora que conhecemos todos os passos do processo, podemos utilizá-los de uma forma automatizada.
  - Em seu /home/cbab2011/SEUNOME, crie um diretório chamado SeedServer e mova-se para dentro dele.
  - Faça o download do pacote SeedServer aqui: ([http://biodados.icb.ufmg.br/cbab2011\\_data/](http://biodados.icb.ufmg.br/cbab2011_data/))
  - `wget http://biodados.icb.ufmg.br/cbab2011_data/SS_package.tar.gz`
  - Descompacte com: `tar -xzf SS_package.tar.gz`
  - Edite o arquivo de entrada “input\_file” com os parâmetros gerais SeedServer. Modifique somente o conteúdo dentro das flags. Assegure-se de que os identificadores UniProt existem na versão atual e que estejam separados por vírgula somente.

**IMPORTANTE:** Deve-se escolher um processID único, para evitar conflito entre as tabelas que serão geradas durante o processamento pelo MySQL! Consulte os outros grupos.

ex:

```
vi input_file
```

```
<file>test_input.txt</file>
<SLtable>allorg</SLtable>
<input>UniprotID</input>
<parameter_r>50</parameter_r>
<parameter_c>50</parameter_c>
<parameter_s>50</parameter_s>
<parameter_l>0.3</parameter_l>
<e_value>1e-10</e_value>
<clade_limit>yes</clade_limit>
<self_score>0.7</self_score>s
```

- Como as demais configurações já foram previamente definidas, rode o Seed Linkage com o comando:

```
screen perl roda input_file
```

OBS: para sair do screen, mantenha pressionada a tecla ctrl e depois tecla A e depois D. Com esse comando screen você pode fechar o terminal sem que o seu processo seja finalizado antes de terminar.

- Com sorte, teremos tempo de analisar os resultados, para isso precisamos realizar mais consultas no banco de dados MySQL.

```
- select * from temp_YOURCODE_1_clade where PSI = 1;
```

```
- select * from temp_YOURCODE_1_clade where PSI = 1 and SL != 0;
```

```
- select * from temp_YOURCODE_1_clade where PSI = 1 and maxsov > 0.8;
```

```
- select * from temp_YOURCODE_1_clade where PSI = 1 and KO = 2;
```

```
- select * from temp_YOURCODE_1_clade where PSI = 1 and phylum = 7711;
```

- Podemos alinhar as seqüências validadas com um alinhador múltiplo e então visualizar o alinhamento em uma árvore filogenética.

- ```
select uniprot_ID from temp_YOURCODE_1_clade where PSI = 1 and maxsov > 0.8 into outfile `~/tmp/YOURCODE_IDS`;
```

- ```
fastacmd -i /tmp/YOURCODE_IDS -d /home/cbab2011/DATABASE/uniprot2010_10 YOURCODE_IDS.fasta -o
```

- Agora que possui o arquivo FASTA, acesse o alinhador múltiplo PRANK online (<http://www.ebi.ac.uk/goldman-srv/webPRANK/>)

- Após obter o alinhamento, exporte a saída no formato NEXUS
- Utilize esse arquivo pra criar a árvore em (<http://itol.embl.de/upload.cgi>)