



ANR Génomique

**Lettre d'information
de
SIGENAE**



N°23

du 29 avril 2008 – Rédacteur : CK

Cette lettre d'information est destinée aux membres des équipes de recherche participant aux projets AGENAE (Analyse des GENômes des Animaux d'Elevage) ou ANR Génomique. Elle a pour but de vous informer sur les évolutions de l'équipe SIGENAE, les nouveaux outils, services et formations mis en place.

L'équipe SIGENAE : Composition :

Prénom Nom	Localisation	Tel	Email
Laurence Amilhat	Toulouse	05 61 28 57 08	laurence.amilhat@toulouse.inra.fr
Philippe Bardou	Toulouse	05 61 28 57 09	philippe.bardou@toulouse.inra.fr
Cédric Cabau	Tours	02 47 42 75 42	cedric.cabau@tours.inra.fr
Pierrot Casel	Toulouse	05 61 28 57 09	pierrot.casel@toulouse.inra.fr
Patrice Dehais	Toulouse	05 61 28 57 08	patrice.dehais@toulouse.inra.fr
Christophe Klopp	Toulouse	05 61 28 50 36	christophe.klopp@toulouse.inra.fr
François Moreews	Rennes	02 99 84 75 95	francois.moreews@irisa.fr

Prolongation et modification des contrats :

Laurence Amilhat termine son contrat Aquagenome à la fin du mois d'avril 2008 et débute une nouvelle année sur le projet ANR Delisus. Les travaux prévus sont relatifs au stockage et à la mise disposition, dans l'environnement Ensembl, des SNP et des génotypes générés au cours du projet.

Pierrot Casel poursuit son contrat dans le cadre du projet EADGENE sur l'annotation des puces oligo mises à disposition des biologistes travaillant dans le réseau. Il a pour but de produire des fichiers d'annotation à jour pour les trois supports poulet, bovin et porc.

1/ PhyloSig – chaîne de traitement de phylogénomie à grande échelle. :

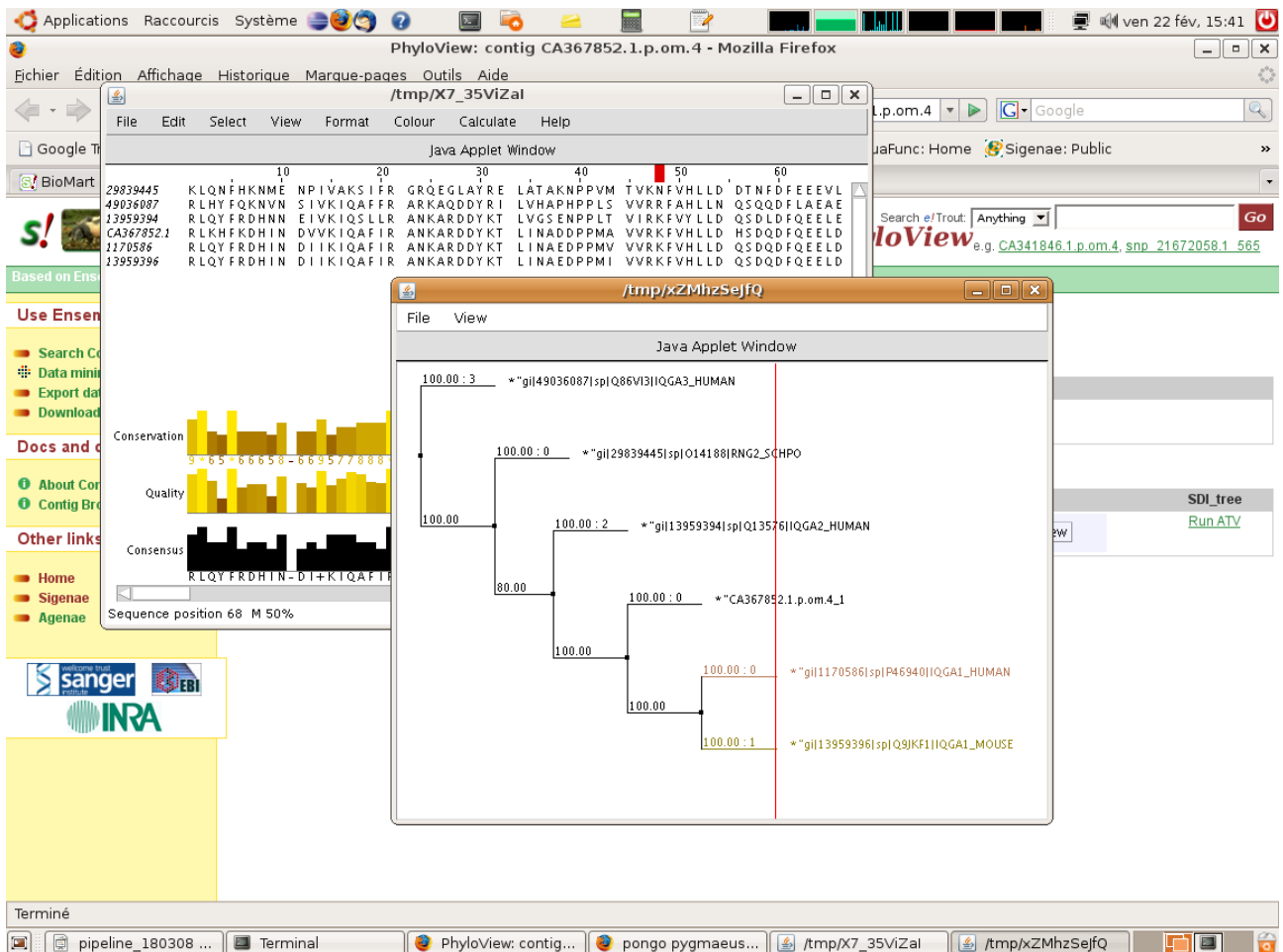
PhyloSig a été développé dans le cadre du projet Aquagenome. Cette chaîne de traitements permet d'automatiser les étapes d'une analyse de phylogénomie. Elle a été développée dans le but de réaliser des annotations à grande échelle par phylogénomie de clones déposés sur le support générique AGENAE truite.

PhyloSig se décompose en quatre étapes:

1. **l'identification de séquence homologues.** Cette étape correspond à une recherche par similarité dans des banques de séquences (programme BLAST). Les séquences homologues sont sélectionnées selon leur degré de similarité et la taxonomie.
2. **l'alignement multiple.** Il est construit avec le programme MUSCLE. Chaque colonne de l'alignement est supposée contenir des acides aminés ayant une évolution commune. C'est pourquoi les régions dont l'homologie est ambiguë sont retirées de l'alignement. L'outil GBLOCKS nous permet de ne garder que des blocs de l'alignement où l'homologie est conservée.
3. **la construction d'arbres phylogénétiques.** 3 méthodes sont utilisées pour construire les arbres de phylogénomie: distance, parcimonie et maximum de vraisemblance. Un arbre consensus (100 bootstrap) est construit en utilisant chacune de ces méthodes (programmes : Phylip et Phym). Puis, un arbre consensus est bâti à partir des 3 arbres précédents.
4. **Inférence des évènements de duplication ou de spéciation.** Cette étape est réalisée par l'outil SDI du logiciel ATV, en comparant l'arbre consensus final des séquences homologues avec l'arbre de la vie extrait du NCBI.

Les outils utilisés, les banques pour les recherches d'homologie, ainsi que les critères de sélection des similarités peuvent être paramétrés en fonction des besoins de l'utilisateur.

Les fichiers résultats (alignements multiples et arbres) sont mis à la disposition de l'utilisateur dans un environnement BioMart.



Nous avons utilisé PhyloSig sur les contigs publics de truites contenant un clone déposé sur les membranes génériques AGENAE avec des paramètres différents:

- Version 1: Base interrogée= Swissprot; e-value seuil pour le blast= 1e-50; paramètres de sélection Gblock=Stringent, pas de sélection des hits en fonction de leur taxonomie.
- Version 2: Base interrogée= Swissprot; e-value seuil pour le blast= 1e-10; paramètres de sélection Gblock=Souple, pas de sélection des hits en fonction de leur taxonomie.

Ces résultats sont consultables à l'adresse suivante:

<http://public-contigbrowser.sigenae.org:9090/index.html> (suivre le lien dans la marge de gauche: Data mining [BioMart], puis choisir l'espèce *Oncorhynchus mykiss* et *omykiss_phylogeny*).

PhyloSig a été développé en Shell et en Perl. Il peut être installé sur toute plateforme Unix/Linux. Il est disponible sous forme de fichier .tgz sur le site: <http://mulcyber.toulouse.inra.fr/>

2/ Report de la formation F08c :

La formation F08c décrite ci-dessous a été reportée au mois d'octobre prochain. Un nouveau bulletin d'inscription vous sera transmis début septembre 2008.

F08c : Automatiser vos traitements bio-informatiques (1 jour)

Vous avez l'habitude de récupérer des informations en naviguant sur différents site web, vous relancer souvent les mêmes traitements, cette formation est faite pour vous simplifier la vie :

- les concepts : web-services et workflow,
- les outils : taverna et les web-services mis à disposition,
- la pratique : découverte à l'aide d'exemples et d'exercices.

Pour toute demande d'information ou de travaux, veuillez envoyer un mail à sigenasupport@jouy.inra.fr en précisant vos nom et coordonnées.