

Une ressource internet pour le génome de la Paramécie

<http://paramecium.cgm.cnrs-gif.fr/db/index>

CONTEXTE

Le génome de la Paramécie, équivalent en taille au génome de la Drosophile, vient d'être séquencé, assemblé et annoté au Genoscope-CNS [1]. La paramécie est un organisme modèle étudié depuis plus de 50 ans, qui a besoin d'une base de données ayant la capacité d'intégrer l'ensemble des informations biologiques, séquences et annotations du génome.

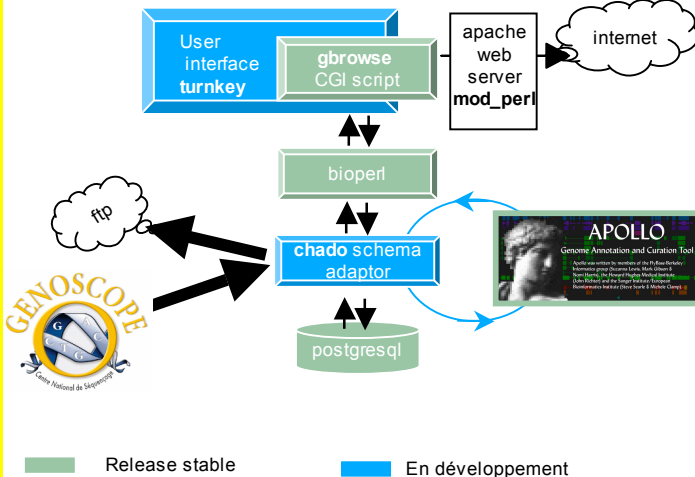
Le projet GMOD [2] fournit des outils génériques permettant la mise en place d'une nouvelle base de données pour organisme modèle. Au cœur du projet, le schéma relationnel chado, développé à FlyBase. Ce schéma est modulaire et intègre l'utilisation de vocabulaires contrôlés et structurés (ontologies).

OBJECTIFS

Nous nous sommes engagés dans le développement d'une base de données pour la paramécie. ParameciumDB, s'appuyant sur les outils fournis par le projet GMOD. ParameciumDB doit centraliser et assurer la qualité de l'annotation experte du génome et constituer une ressource pour tous les biologistes, intégrant les données génétiques, de cartographie des chromosomes, de structure et fonction des gènes, de transcriptome, de protéomique et de bibliographie.

Pour compléter le dispositif, nous souhaitons mettre en place un environnement d'outils d'analyse pour assurer la mise à disposition de ce génome et la qualité des données génomiques pour les analyses bio-informatiques. En effet, divers aspects de ce génome semblent très intéressants (très grande nombre de gènes prédits; duplication globale récente du génome; introns nombreux de seulement 25 nt; ARNs non codants impliqués dans des processus épigénétiques etc).

PROJET



CALENDRIER

- Janvier 2005 Embauche d'Olivier Amaiz (CDD IMPBio)
- Mars 2005 Visite de Scott Cain (coordinateur du projet GMOD)
- Avril 2005 Version 'pre-release' de ParameciumDB en ligne
- AOÛT 2005 Inauguration officielle de ParameciumDB, avec les données du génome. Soumission à 'Nucleic Acids Research Database Collection'.**
- 2006. Interfaçage d'Apollo. Formulaire pour la soumission de données. Publication de ParameciumDB.

LABORATOIRES IMPLIQUES

CGM / UPR 2167 / Gif-sur-Yvette
LGM/ ENS UMR8541 / Paris
IGR / UMR1599 / Villejuif
Genoscope / Evry
Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

MOTS CLES

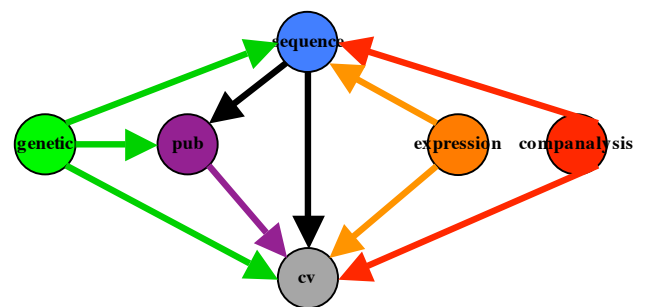
Paramecium; génome; Model Organism Database; mod_perl, chado, turnkey, Generic Genome Browser [3]; Apollo [4]; OBO (Open Biological Ontologies); SGBD

REFERENCES

- [1] <http://www.genoscope.cns.fr/paramecium>
- [2] <http://www.gmod.org>
- [3] Stein, L. D. et al. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res* 12, 1599-1610.
- [4] Lewis, S. E. et al. (2002). Apollo: a sequence annotation editor. *Genome Biol* 3, RESEARCH0082.

RESULTATS

MODULARITE DU SCHEMA CHADO



- CV Vocabulaires contrôlés (ontologies)
- expression Module pour les données transcriptome (MIAME compatible)