

Extension des services proposés par le PBIL

<http://pbil.ibcp.fr/html/pbil-ext-impbio85.html>

CONTEXTE

La biologie à grande échelle (séquençages de génomes complets, génomique structurale, puces à ADN, protéomique, bibliographie...) produit une quantité de données hétérogènes toujours croissante qui s'accompagne d'un développement très rapide de la demande en analyses de la part de la communauté des biologistes. La Bioinformatique a donc pour missions d'assurer la gestion, l'accès et l'analyse de ces données biologiques pour une meilleure compréhension du vivant par le développement de base de données, de méthodes d'analyses pertinentes et de logiciels et webiciels à destination de la communauté des biologistes.

OBJECTIFS

Le Pôle Bioinformatique Lyonnais (PBIL, <http://pbil.univ-lyon1.fr/>) a été créé en 1998 par les équipes Bioinformatique et RMN Structurales (Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS) et Bioinformatique et Génomique Évolutive (Laboratoire Biométrie, Biologie Évolutive, UMR 5558 CNRS). Le PBIL a obtenu une reconnaissance internationale grâce à ses activités de recherches (bases de données et méthodologies) et de services offerts à la communauté des biologistes au travers de son serveur Web et de ses nombreux logiciels. L'interface WWW mise en place assure la visibilité du PBIL comme une suite logique et cohérente avec le souci de fournir le meilleur service à l'utilisateur.

Le PBIL est aussi impliqué dans des développements de divers projets bioinformatiques en collaboration avec d'autres équipes. Pour ce projet, le PBIL est associé à l'équipe Structure et Evolution des Récepteurs Nucléaires d'Hormones (LBMC-CNRS-UMR5665) qui développe la base de données des récepteurs nucléaires NuReBase.

L'objectif du projet PBIL-EXTENSION est d'améliorer et d'étendre les services proposés aux utilisateurs du PBIL par:

- 1) le renforcement du serveur **NPS@** par le déploiement sur une architecture distribuée de type cluster de calcul et l'addition d'un système de gestion d'espace virtuel de travail pour les utilisateurs,
- 2) l'ajout de fonctionnalités de phylogénie moléculaire, en mode client/serveur (**PHYLO_JAVA**), et d'outils assurant l'interface entre le **système R** d'analyse statistique et les séquences,
- 3) l'extension de la base **NuReBase** à des données hétérogènes : patrons d'expression des données SAGE et biopuces, et intégration d'outils de gestion de ces données et de navigation entre celles-ci.

PROJET

Extension des services proposés par le PBIL : distribution sur cluster de calcul du serveur **NPS@** et addition d'un espace virtuel de travail utilisateur; déploiement d'un système client/serveur de phylogénie moléculaire et d'outils d'analyse statistique; extension à des données hétérogènes de la banque NuReBase de récepteurs nucléaires.

RESULTATS

- Mise en place du cluster de calcul pour le serveur **NPS@**.
- Déploiement du serveur **NPS@** sur le cluster et mise-à-jour distribuée des DB.
- Développement du gestionnaire de l'espace virtuel de **NPS@**.
- Extension de **NuReBase**: intégration d'informations sur les ligands et perturbateurs endocriniens.
- Développement de la base **XpressDB** pour les données d'expression (implémentation sous MySQL).
- Développement d'une interface entre les outils d'analyse statistique de l'environnement R et les données de séquences (**sequinR**).
- Architecture client/serveur d'interrogation des banques de séquences au format ACNUC.

CALENDRIER

- 2004** **NPS@**: installation du cluster, mise-à-jour distribuée des DB. **PHYLO_JAVA**: diffusion du client et addition de méthodes phylogénétiques, interface publique de parallélisation des bootstraps. **NuReBase**: modélisation patrons d'expression et ontologies, normalisation des données d'EST.
- 2005** **NPS@**: déployé et optimisé pour le cluster.
- 2006** **NPS@** espace virtuel. **PHYLO_JAVA**: diffusion du serveur. **NuReBase**: intégration et accès des données biologiques.

LABORATOIRES IMPLIQUES

Equipe Bioinformatique et RMN Structurales - Institut de Biologie et Chimie des Protéines - UMR 5086 CNRS/UCBL, Lyon

Equipe Bioinformatique et Génomique Évolutive - Laboratoire de Biométrie, Biologie Évolutive - UMR 5558 CNRS/UCBL, Villeurbanne

Equipe Structure et Evolution des Récepteurs Nucléaires - Laboratoire de Biologie Moléculaire et Cellulaire - UMR 5665 CNRS/ENS, Lyon

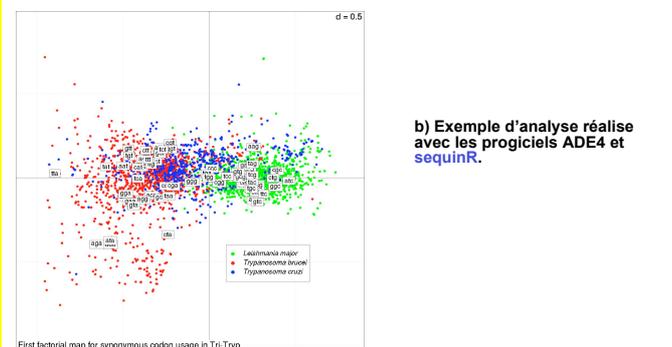
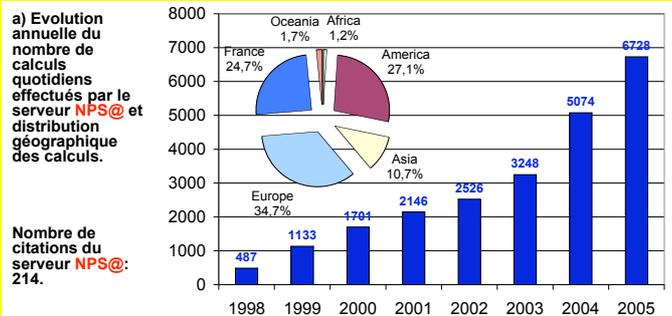
MOTS CLES

Calcul distribué; cluster de calcul; grille de ressources informatiques; GRID computing; bases de données; espace virtuel de travail; récepteurs nucléaires; client/serveur; phylogénie moléculaire; analyse de séquences; statistiques; serveurs web; services.

PUBLICATIONS PRINCIPALES

- [1] Combet C, Blanchet C, Geourjon C, Deleage G. NPS@: network protein sequence analysis. Trends Biochem Sci., 2000 25(3):147-50.
- [2] Perrière G, Combet C, Penel S, Blanchet C, Thioulouse J, Geourjon C, Grassot J, Charavay C, Gouy M, Duret L, Deleage G. Integrated databanks access and sequence/structure analysis services at the PBIL. Nucleic Acids Res., 2003, 31(13):3393-9.
- [3] Ruau D, Duarte J, Ourjadal T, Perrière G, Laudet V, Robinson-Rechavi M. Update of NUREBASE: Nuclear hormone receptor functional genomics. Nucleic Acids Res., 2004, Vol. 32, Database issue D165-D167
- [4] Charif D, Thioulouse J, Lobry JR, Perrière G. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics, 2005, 21(4):545-547.

SERVICES DU PBIL



c) Page d'accueil de la base NuReBase

d) Page d'accueil de la base XpressDB

