

OPTIMISATION DU SCHÉMA DE PARALLÉLISATION DU CODE BFAST

- **DATE DE DÉBUT :** Janvier/mai 2012
- **LIEU :** CS-SI – 22, avenue Galilée – 92350 Le Plessis Robinson
- **CONTACT :** benoist.gaston@c-s.fr
- **DURÉE :** 6 mois

CONTEXTE :

Le code BFAST est un code d'alignement de séquences ADN, développé à UCLA par Nils Homer ("<http://bfast.sourceforge.net/>"). BFAST repose sur les exécutions successives de trois parties du code : l'alignement global, l'alignement local et une phase de post-traitement.

Le code est écrit en C et intègre une parallélisation utilisant la mémoire partagée d'un noeud de calcul, cette parallélisation est faite par Pthreads. Le code est composé d'environ 25000 lignes de code source.

L'objectif est l'extension de la parallélisation du code BFAST à un contexte multi-noeuds afin d'accélérer les temps de restitution d'une procédure d'alignements de nombreuses séquences courtes sur un génome de référence.

STAGE PROPOSÉ :

Une première version MPI-pthread a été implémentée, cette version sature très rapidement en terme de performances.

Certaines pistes d'amélioration ont été identifiées, la 1ère d'entre elles consiste à introduire un mode de type maître-esclave de sorte que le maître distribue à la volée les données de calculs, et se charge d'ordonner les résultats au fur et à mesure qu'ils lui reviennent, avant de les écrire correctement.

- **NIVEAU REQUIS :** Master ou école d'ingénieur
- **COMPÉTENCES :**
 - environnement de développement Unix/Linux
 - programmation parallèle MPI et pthreads
 - langage C