

# Analyse de génomes: des avancées pour la biologie et le monde médical

François Rechenmann

► **To cite this version:**

François Rechenmann. Analyse de génomes: des avancées pour la biologie et le monde médical. Collection "20 ans d'avancées et de perspectives en sciences du numérique", INRIA, 2012, 2 p. <hal-00820542>

**HAL Id: hal-00820542**

**<https://hal.inria.fr/hal-00820542>**

Submitted on 6 May 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analyse de génomes: des avancées pour la biologie et le monde médical



Les premières séquences génomiques et les premières méthodes d'analyse de gènes sont apparues dans les années 70. Mais les progrès spectaculaires des techniques de séquençage des 20 dernières années, associés aux performances croissantes des technologies informatiques, ont ouvert un immense champ d'exploration de ces données aux bioinformaticiens. Un défi toujours d'actualité pour l'avancée des connaissances en biologie et en médecine.

Témoignage de **François Rechenmann**, directeur de recherche, équipe-projet Ibis

Dans les vingt ans qui ont suivi le lancement du projet « Génome Humain » en 1990, les technologies de séquençage des génomes ont progressé de six ordres de grandeur. Ainsi, là où plus de 40 laboratoires se sont mobilisés pendant plusieurs années pour obtenir en 1997 le génome complet de la bactérie *B. subtilis*, un résultat similaire peut aujourd'hui être obtenu en quelques jours à un coût inférieur à 1000 €.

## Localiser les gènes, déterminer la fonction d'une protéine et interpréter les grandes quantités d'informations générées

Connaître la séquence de génomes n'est qu'une première étape, l'interpréter en est une autre. C'est là qu'intervient la bioinformatique.

Il s'agissait tout d'abord de prédire la localisation des gènes, c'est-à-dire les régions sur la séquence génomique qui portent l'information utilisée par la cellule pour synthétiser une ou plusieurs protéines. Le problème a trouvé de bonnes solutions algorithmiques pour les organismes bactériens, mais reste très imparfaitement résolu pour les génomes eucaryotes, tels que ceux des mammifères, dont les gènes, perdus dans la séquence génomique, sont de plus morcelés en sous-régions codantes et non-codantes.

La prédiction des fonctions d'une protéine, dont la séquence en acides aminés a été déterminée à partir de la localisation et la structure d'un gène, ne possède pas de solution algorithmique. C'est essentiellement par comparaison directe ou indirecte avec des protéines dont la fonction est connue que les fonctions d'une protéine peuvent être inférées. Cette comparaison, du fait de la taille des séquences et des bases de référence utilisées, met en œuvre des heuristiques qui restent très gourmandes en puissance de calcul et sollicitent les infrastructures distribuées.

Disposer de bases de données de référence est devenu un enjeu pour l'interprétation des séquences génomiques. Au sein de mon équipe de recherche Helix, en collaboration avec l'Université Joseph Fourier à Grenoble et l'Institut Suisse de Bioinformatique à Genève, nous avons conçu la base de données MicroB qui intègre les connaissances génomiques, protéiques et métaboliques sur plusieurs centaines d'organismes microbiens.

Ces différents outils d'analyse, de gestion, mais aussi de visualisation, doivent être aisément accessibles et utilisables par les biologistes dont l'objectif est une meilleure compréhension des organismes dont le génome est séquençé. Le consortium Genostar, qui a associé de 1999 à fin 2003, Inria, Institut Pasteur et les deux sociétés de biotechnologies Genome Express et Hybrigenics, a ainsi développé un environnement intégré d'analyse et de gestion des génomes microbiens, qui est maintenant valorisé, étendu et distribué par la société Genostar.

“ Connaître et analyser les génomes ont des applications très diverses ”

Connaître le génome de l'homme permet de mieux connaître l'organisme humain et ses pathologies ; connaître les génomes des bactéries et des virus est tout aussi fondamental. Outre la contribution à la conception de vaccins, d'antibiotiques ou d'antiviraux, la connaissance des micro-organismes a des retombées importantes dans le domaine des biotechnologies. Il s'agit là d'utiliser des bactéries, des levures ou encore des microalgues, adéquatement sélectionnées ou modifiées, pour leur faire produire des molécules d'intérêt, par exemple des précurseurs de biocarburants.

**Les progrès du séquençage et des méthodes d'analyse sont en train de modifier**

## profondément le monde médical

Ainsi, on sait que tous les cancers trouvent leur origine dans des modifications du génome. Plusieurs projets, typiquement menés au sein de l'ICGG (International Cancer Genome Consortium), visent donc à séquencer, pour un même patient, les génomes de cellules cancéreuses et de cellules normales, et ce pour différents types de cancers. En comparant les deux génomes, et en utilisant également un génome de référence où les gènes ont été localisés et caractérisés, l'identification des différences et leur association à la pathologie devraient permettre de construire des « cartes d'identité » des tumeurs, ouvrant la voie à des diagnostics fins et des thérapies ciblées. À Lyon, la Fondation Synergie Lyon Cancer mène notamment un projet très ambitieux de cette nature.

### ET DANS 20 ANS ?

**François Rechenmann**, fondateur en 1999 de l'équipe-projet Inria Helix pour développer des outils informatiques d'analyse de séquences génomiques puis, en 2004, de la société Genostar pour en promouvoir l'usage au sein des laboratoires tant académiques qu'industriels.



© INRIA / Jim Wallace

"Il faut s'attendre à ce que séquenceurs et logiciels bioinformatiques se multiplient en milieu clinique puis, à terme, dans de nombreux laboratoires d'analyses médicales. Il serait alors possible, par exemple, de suivre l'évolution de la résistance de l'agent pathogène chez des patients en cours de thérapies antivirales ou antibactériennes. Pour cela, les bioinformaticiens devront proposer des logiciels, simples d'usage, mais extrêmement fiables et pertinents, susceptibles d'être certifiés.

Du côté des biotechnologies, la modification de souches par ajout de gènes et de leurs signaux de régulation nécessite que soient développés de véritables systèmes de conception assistée par ordinateur pour aider le microbiologiste à choisir et assembler des composants sous la forme de séquences. La suite du processus, à savoir la réalisation des vecteurs et leur évaluation, sera toujours plus automatisée, permettant de tester et de cribler rapidement de très nombreuses constructions."

## Numérique & société

**Idée reçue :** Il est vrai que le génome de l'Homme a été (presque) intégralement séquencé, c'est-à-dire que la quasi-totalité de la succession des nucléotides qui constituent les macromolécules d'ADN humaines (le génome) est connue et disponible sous la forme d'un texte long de près de 3 milliards de lettres. Ces lettres – A, C, G, T – sont les initiales des motifs chimiques qui distinguent les quatre types de nucléotides. Mais connaître un texte est une chose, l'interpréter en est une autre.

Source : [interstices.info](http://interstices.info)

**Février 2001** a marqué une étape importante dans la mise à jour du patrimoine génétique humain. Publication est faite de la quasi totalité (94%) de la carte génétique, qui est le résultat de plus de 10 années de travail et de mise en commun de ressources humaines et techniques de scientifiques de 18 pays.

Source : [chu-tours.fr](http://chu-tours.fr)

**2008**, les scientifiques annoncent que "La séquence du génome humain aujourd'hui accessible dans les bases de données est aussi complète que les techniques actuelles le permettent. Elle correspond essentiellement à la partie « séquençable » du génome, celle qui contient la quasi totalité des gènes. Cette fraction du génome, nommée euchromatine, représente 2,9 milliards de nucléotides, soit 90% des 3,2 milliards de nucléotides de l'ensemble du génome humain. Elle a été séquencée à 99% (le 1% restant correspondant aux quelques centaines de trous que l'on n'est pas parvenu à boucher)."

Source : [cns.fr](http://cns.fr)

## 1992 - 2012



- Collection "20 ans d'avancées et de perspectives en sciences du numérique" par les chercheurs d'équipes Inria de Grenoble et Lyon.
- [www.inria.fr/20ansgrenoble](http://www.inria.fr/20ansgrenoble)

© Inria - Editions Victoria