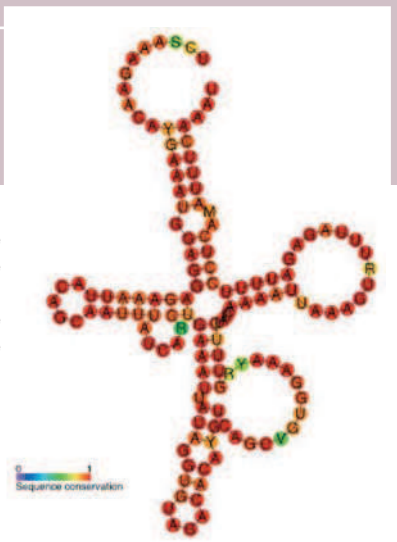
Programme « Systèmes embarqués et grandes infrastructures », édition 2008

Projet BIOWIC

La puissance des cartes graphiques au service de la bioinformatique



Représentation graphique du repliement 2D d'une séquence d'ARN obtenu à l'aide d'algorithmes de prédiction de structure

Le projet « BIOWIC Workflow pour les traitements intensifs en bioinformatique » est un projet de recherche industrielle coordonné par l'EPI INRIA Symbiose (IRISA Rennes). Il associe aussi l'EPI INRIA Cairn (IRISA Rennes), la Plateforme de Bioinformatique GenOuest à Rennes, le MIG INRA à Jouy en Josas et ELIAUS (Université de Perpignan). Le projet a commencé en janvier 2009 pour une durée de 36 mois : il bénéficie d'une aide ANR de 633 k€ pour un coût global de l'ordre de 2,2 M€.

Des données génomiques qui croissent plus vite que la puissance des machines

La bioinformatique se caractérise par le traitement de très grandes masses de données issues de l'information génétique contenue dans les cellules de chaque être vivant. Au cours des deux dernières décennies, la production de données a plus ou moins suivi l'évolution des performances des processeurs. Aujourd'hui, la situation est bouleversée. D'une part, les avancées spectaculaires des biotechnologies, et les nouvelles techniques de séquençage associées, décuplent la production de données génomiques. D'autre part, les performances des processeurs n'évoluent plus aussi rapidement : les contraintes thermiques empêchent la fréquence d'horloge de croitre au même rythme que par le passé. Les bioinformaticiens sont donc confrontés au problème de traiter une masse croissante de données génomiques en un temps toujours plus bref avec des ressources en calcul qui n'évoluent pas en conséquence. Le projet BioWIC à pour objectif de répondre aux besoins actuels et futurs pour accélérer les traitements des données génomiques des grands projets de bioinformatiques. Le principal levier est l'usage d'accélérateurs matériels avec, notamment l'exploitation de cartes graphiques de dernière génération.

Les cartes graphiques : des machines parallèles puissantes mais pas si faciles à programmer

Les cartes graphiques sont des machines parallèles très puissantes. Elles contiennent une ou plusieurs puces intégrant un très grand nombre d'unités spécialisées dans le traitement des images. Cependant, depuis quelques années, leur domaine applicatif évolue progressivement vers d'autres horizons grâce à une flexibilité de programmation accrue. En effet, des langages de haut niveau, non nécessairement ciblés vers l'image, permettent maintenant de programmer cette ressource pour d'autres applications gourmandes en calcul. Ceci dit, toutes les applications ne sont pas systématiquement aptes à une mise en œuvre sur ce matériel. Les algorithmes doivent posséder de bonnes propriétés de parallélisme pour que leurs exécutions puissent être dispatchées efficacement sur les nombreuses unités de calcul. La difficulté, mais aussi le challenge, réside donc dans la manière d'exprimer les calculs sous une forme compatible avec l'architecture des puces graphiques. A l'heure actuelle, cela suppose à la fois une connaissance fine de leur structure interne et une programmation exigeante pour tirer parti de toute leur puissance.

IMPACTS

Résultats majeurs

Un programme de bioinformatique particulièrement coûteux en calculs, UnaFold (prédiction de repliement d'ADN), a été parallélisé sur carte graphique, puis validé par une application bioinformatique conséquente. Sommairement, il s'agissait d'estimer la pertinence du repliement 2D d'une famille de séquences d'ADN par rapport à un échantillon de plusieurs milliers de séquences aléatoires de même composition. L'intégration de deux cartes graphiques dans une même machine a fournit une accélération d'un facteur 50. Ainsi, le temps de calcul pour traiter une séquence d'ADN de 5 000 nucléotides passe de 15 jours sur un serveur standard (dual-core Intel Xeon 2.66 GHz) à quelques heures sur le même serveur équipé de deux cartes graphiques.

Production scientifique et brevets

Un papier décrivant la parallélisation de l'algorithme UnaFold sur cartes graphiques a été publié dans les actes de l'International Conference on Computational Science en mai 2009. Il relate, entre autre, une expérimentation menée avec une équipe de biologistes de l'INRA pour détecter les microARNs sur la base de leur structure 2D dans le génome du puceron. Il montre également qu'un cluster de 32 nœuds est un peu plus rapide qu'un serveur de base équipé d'une carte graphique de dernière génération.

Contact : Dominique Lavenier • lavenier@irisa.fr • http://biowic.inria