

EMBL-EBI crée un nouveau support de stockage de données: l'ADN

Tous les films et émissions télévisées jamais produits – dans un verre d'eau

Hinxton, le 23 janvier 2013 – Des chercheurs du LEBM-Institut européen de bioinformatique (EMBL-EBI en anglais) ont créé une nouvelle technique pour stocker des données sous forme d'ADN – un support qui résiste pendant des dizaines de milliers d'années. Cette nouvelle méthode, publiée aujourd'hui dans la revue *Nature*, permet de stocker plus de 100 millions d'heures de vidéo en haute définition dans un volume équivalent à une tasse d'ADN.

Le monde regorge d'informations numériques – environ trois zettabytes (3000 milliards de milliards de bytes) – et l'afflux constant de nouvelles données constitue un véritable défi pour les archivistes. Les disques durs sont coûteux et nécessitent beaucoup d'électricité, et les meilleurs matériaux d'archivage 'non-électriques', tels que la bande magnétique, se dégradent en moins d'une décennie. Ce problème devient de plus en plus pressant en sciences de la vie où la recherche repose sur des quantités de données toujours plus énormes (dont des séquences d'ADN).

"Nous savons déjà que l'ADN est un outil de stockage d'information fiable car celui que nous pouvons extraire d'os de mammoths laineux vieux de plusieurs dizaines de milliers d'années est toujours exploitable", explique Nick Goldman de l'EMBL-EBI. "L'ADN est aussi étonnamment petit et dense et n'a pas besoin d'électricité pour être stocké, donc il est facile à conserver et à échanger."

Lire de l'ADN est assez simple ; le principal obstacle à son utilisation comme support de stockage réside dans son écriture. Deux défis se posent : tout d'abord les méthodes actuelles permettent seulement la synthèse de chaînes courtes d'ADN. Ensuite, la lecture et la synthèse d'ADN sont sujettes à erreurs, en particulier quand la même lettre d'ADN est répétée. Les co-auteurs, Nick Goldman et Ewan Birney (directeur adjoint de l'EMBL-EBI), se sont lancés dans la création d'un code permettant de surmonter ces deux problèmes.

"Nous savons que nous devons créer un code n'utilisant que des chaînes courtes d'ADN, et où la répétition en série d'une même lettre serait impossible. Nous avons donc décidé de décomposer le code en de multiples fragments se chevauchant dans les deux sens, et indexés de façon à indiquer leur position précise dans l'ensemble, le tout avec un schéma interdisant les répétitions. Ainsi, pour qu'une erreur apparaisse dans le code global, il faudrait avoir la même erreur sur quatre fragments différents – ce qui serait très rare", théorétise Ewan Birney.

La nouvelle méthode nécessite que l'ADN soit produit à partir d'information digitale: c'est là qu'entre en jeu Agilent Technologies, Inc., une société basée en Californie qui a offert



Nick Goldman de l'EMBL-EBI, observant de l'ADN nouvellement synthétisé.
Crédit: EMBL Photolab

bénévolement ses services. Ewan Birney et Nick Goldman leur ont envoyé les versions digitales du fichier .mp3 du discours de Martin Luther King "I Have a Dream"; d'une photo .jpg de l'EMBL-EBI; d'un .pdf du papier séminal de Watson et Crick "Structure moléculaire des acides nucléiques"; d'un fichier .txt de l'ensemble des sonnets de Shakespeare, et d'un fichier qui décrit le codage.

"Nous avons téléchargé les fichiers d'internet et les avons utilisés pour synthétiser des centaines de milliers de morceaux d'ADN – le résultat ressemble à un petit flocon de poussière", explique Emily Leproust d'Agilent. Agilent a envoyé l'échantillon à l'EMBL-EBI, où les chercheurs ont réussi à séquencer l'ADN et à décoder les fichiers sans erreurs.

"Nous avons créé un code qui tolère les erreurs grâce à une structure moléculaire dont nous savons qu'elle durera pendant 10 000 ans, voire plus, dans de bonnes conditions," explique Nick Goldman. "Tant que quelqu'un connaîtra le code, on pourra le lire en sens inverse à condition d'avoir un lecteur approprié."

Bien qu'il existe de nombreux aspects pratiques à résoudre, la densité et la longévité de l'ADN en font un moyen de stockage intéressant. La prochaine étape pour les chercheurs est de parfaire le système de codage et d'en explorer les aspects pratiques, pour ouvrir la voie à une solution commercialement viable de stockage à base d'ADN. ●

Article source

Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. Advanced online publication in *Nature* on 23 January 2013; DOI: 10.1038/nature11875

Contacts: contactpress@ebi.ac.uk

Mary Todd-Bergman, Outreach Programme project Leader, Hinxton, UK, Tel: +44 1223 494665, www.ebi.ac.uk

Isabelle Kling, Chef de projet EMBL, Heidelberg, Germany, Tel.: +49 6221 387 8355, www.embl.org, isabelle.kling@embl.de

A propos du LEBM-IEB

L'Institut Européen de Bioinformatique (IEB) fait partie du Laboratoire Européen de Biologie Moléculaire (LEBM). Il se trouve sur le campus de génomique du Wellcome Trust, à Hinxton près de Cambridge (Royaume-Uni). L'IEB a été créé à partir du travail d'avant-garde du LEBM pour la création de bases de données biologiques publiques pour la communauté des chercheurs. Il abrite certaines des banques de données biologiques les plus importantes, y compris de séquences d'ADN (ENA), de séquences protéiques (UniProt), de génomes animaux (Ensembl), de structures tridimensionnelles (la Protein Databank européenne), des données issues d'expériences d'expression génique (ArrayExpress), d'interactions protéine-protéine (IntAct), et de voies de signalisation (Reactome). L'IEB abrite plusieurs groupes de chercheurs et ses scientifiques développent continuellement de nouveaux outils pour la communauté des bioinformaticiens. www.ebi.ac.uk/information

A propos du LEBM

Le Laboratoire Européen de Biologie Moléculaire est un institut de recherche fondamentale financé par 20 pays membres (l'Allemagne, l'Autriche, la Belgique, la Croatie, le Danemark, l'Espagne, la Finlande, la France, la Grèce, l'Irlande, l'Islande, Israël, l'Italie, le Luxembourg, la Norvège, les Pays-Bas, le Portugal, le Royaume-Uni, la Suède et la Suisse) et un membre associé l'Australie. La recherche au LEBM est menée par quelque 85 groupes indépendants qui couvrent tout le champ de la biologie moléculaire. Le laboratoire est géographiquement réparti de la façon suivante : le siège se situe à Heidelberg, et il y a quatre antennes, une à Hinxton (le Laboratoire Européen pour la Bio-informatique), une à Grenoble, une à Hambourg et une à Monterotondo près de Rome. La mission du LEBM est diversifiée : recherche fondamentale en biologie moléculaire ; formation des scientifiques, étudiants et visiteurs quel que soit leur niveau ; service aux chercheurs des pays membres ; développement de nouveaux instruments et de nouvelles méthodes dans le domaine des sciences du vivant ainsi que transfert de technologie. Le Programme International d'Études Doctorales du LEBM accueille environ 190 étudiants. Enfin, le laboratoire offre une plateforme de dialogue avec le public grâce à des activités de communication scientifiques telles que des séries de conférences, les programmes de visiteurs et la diffusion des réussites scientifiques. www.embl.org

Informations pour l'utilisation

Les relevés de presse et de photo du LEBM sont protégés par les droits d'auteur. Il est permis de les imprimer gratuitement et de les distribuer pour l'utilisation non-commerciale par la presse écrite, la radio et la télévision ainsi que par les médias électroniques, à condition que des références claires soient faites concernant les auteurs, les photographes et les créateurs. Des copies des photos en haute résolution sont disponibles sur le site web du LEBM au www.embl.org.