Glossaire

Acide aminé
Voir Protéine.

ADN
L'acide désoxyribonucléique (ADN) est la molécule support de l'information génétique. Celle-ci est codée par l'enchaînement, ou séquence, des maillons élémentaires de l'ADN, les nucléotides. Une molécule d'ADN est formée de deux chaînes aux séquences complémentaires, appariées et enroulées en une double hélice, structure élucidée par James Watson et Francis Crick en 1953.

ADN complémentaire (ADNc)
Molécule d'ADN obtenue par copie d'une molécule d'ARN messager (ARNm). La molécule d'ADNc, à la différence de l'ARN, peut être clonée et amplifiée, et donne ainsi accès à la séquence de l'ARNm. Son nom provient du fait que l'un des brins de l'ADN est complémentaire du brin d'ARN. (Voir la fiche correspondante)

Allèles
Différentes versions de la séquence d'un même gène existant dans une population.

Annotation
Une fois l'ADN séquencé, l'annotation consiste à définir *in silicio* les parties de la séquence qui correspondent à des gènes (instructions) et à obtenir une première idée de leur fonction. Elle utilise différentes ressources, dont des programmes informatiques de prédiction (on parle d'annotation *ab initio* lorsqu'ils sont seuls utilisés) à la fiabilité imparfaite, des comparaisons avec d'autres séquences génomiques, avec des séquences d'ADNc ou de protéines du même organisme ou d'organismes différents. *(Voir les fiches* Interpréter les séquences, Comparer les génomes, ADN complémentaires, Annotation du génome humain*)*

ARN
L'acide ribonucléique (ARN) est une molécule très voisine de l'ADN, mais qui est constituée d'une seule chaîne de nucléotides. Les instructions codées dans la séquence de l'ADN sont recopiées ("transcrites") sous la forme d'une catégorie d'ARN dits messagers pour être traduites en protéines, les véritables agents des fonctions cellulaires. On parle aussi de transcrits.

Base
La partie variable des nucléotides. Il existe 4 types de bases dans l'ADN : thymine (T), Cytosine (C), Adénine (A) et Guanine (G). Dans l'ARN, la thymine est remplacée par l'uracile (U). La succession des bases le long de la molécule d'ADN constitue la séquence. Les bases en vis-à-vis sur chaque brin d'ADN sont appariées selon les règles A–T et G–C. Ces règles permettent de définir la séquence complémentaire d'une séquence donnée.

Chromosome
Les génomes sont organisés en molécules d'ADN nommées chromosomes. Les cellules bactériennes contiennent chacune un unique chromosome, alors que les génomes des organismes supérieurs (plantes, animaux) comprennent plusieurs types de chromosomes ; chaque type est généralement présent en deux exemplaires non identiques par cellule, hérités de chacun des parents. Chez l'homme, il y a 23 paires de chromosomes, dont une paire de chromosomes sexuels de types différents chez l'homme (XY). Il existe donc 24 types de chromosomes humains, numérotés plus ou moins par ordre de taille. Durant la division cellulaire, les molécules d'ADN, associées à des protéines, se condensent et les chromosomes prennent alors la forme de bâtonnets visibles au microscope.

Clonage
En biologie moléculaire, ce mot désigne la reproduction à l'identique d'un fragment d'ADN, intégré dans une bactérie et recopié par celle-ci comme son ADN propre. Après étalement sur une boîte de Petri, chaque clone bactérien – ensemble de cellules issues de la division d'une cellule bactérienne initiale – forme une colonie. C'est le moyen d'isoler les fragments d'ADN d'un mélange, de les conserver et de les préparer en grande quantité. Par extension, le mot clone désigne le fragment d'ADN présent et reproduit à l'infini dans un clone bactérien.

Electrophorèse
Procédé permettant de séparer des molécules (d'ADN dans le cas du séquençage) selon leur taille. Le cœur du dispositif est un gel – un réseau qui emprisonne de l'eau en polymérisant, et au travers duquel les molécules se faufilent sous l'effet d'un champ électrique. Les petites molécules migrent plus vite que les grosses, qui sont freinées. On sait séparer ainsi des molécules d'ADN dont la longueur ne diffère que d'un seul nucléotide.

Enzymes de restriction
Enzymes capables de couper les molécules d'ADN au niveau de séquences précises. Il en existe plus de 400, qui reconnaissent des séquences de 4 à 8 bases de long. Les séquences les plus longues sont les plus rares dans le génome, et permettent donc de couper ce dernier en grands fragments. On se sert aussi des enzymes de restriction pour établir des "profils" de coupure des grands fragments (fingerprint). La comparaison des profils permet de repérer les fragments chevauchants.

Glossaire (suite)

Exon
La plupart des gènes des animaux et des plantes sont morcelés. Les parties biologiquement significatives, les exons, sont séparées par des séquences intercalaires sans signification, les introns, qui sont éliminées dans l'ARN messager. Le repérage des frontières des exons est l'un des problèmes majeurs de l'annotation.

Gel
Voir Electrophorèse.

Gène
Unité élémentaire fonctionnelle dans le génome. Un gène correspond à une instruction à effectuer par la cellule. Le génome d'une bactérie en contient quelques milliers, celui des plantes ou des vertébrés (et donc de l'homme), de l'ordre de 30 000. Certains gènes codent des ARN aux fonctions cruciales, mais la plupart codent la séquence de protéines, qui sont synthétisées via des ARN messagers (Voir ce mot).

Génome
Ensemble de l'information génétique d'un organisme, présente dans chacune de ces cellules. Le support matériel du génome est l'ADN (sauf chez certains virus, où il s'agit d'ARN), sous la forme de longues molécules nommées chromosomes.

Insert
Voir Vecteur.

In silicio
Qualifie une opération effectuée uniquement par ordinateur.

Intron
Voir Exon

Lecture
Ce terme désigne une opération de séquençage unitaire ainsi que l'information qu'elle livre, c'est-à-dire une séquence de 800 à 1000 bases de long, généralement lue à une des extrémités d'un fragment d'ADN plus long.

Matrice
Les matrices sont des molécules d'ADN contenant un insert que l'on veut séquencer, extraites et purifiées à partir d'un clone bactérien. Elles sont alors prêtes à servir de support (matrice) pour les réactions de séquence ou de PCR.

Nucléotide
Les nucléotides sont les molécules dont l'enchaînement constitue les brins d'ADN. Un nucléotide est formé d'une partie constante – un phosphate et un sucre – et d'une partie variable, une base azotée qui peut prendre quatre formes (voir Base).

PCR
La Réaction de Polymérisation en Chaîne (PCR) permet d'amplifier une séquence d'ADN pour laquelle on dispose d'amorces capables de s'apparier à ses extrémités. Un cycle de PCR multiplie par 2 le nombre de molécules, et l'enchaînement d'une trentaine de cycles de quelques minutes chacun livre des quantités d'ADN suffisantes pour le séquençage (2^{30} = 1 milliard de molécules identiques).

Protéine
Comme l'ADN et l'ARN, les protéines sont des macromolécules formées d'une chaîne de longueur variable. Les maillons de cette chaîne sont de petites molécules nommées acides aminés. Il en existe vingt types dans les protéines. La séquence des acides aminés dans une protéine est déterminée par la séquence des bases dans le gène correspondant selon un code simple, où un triplet de bases consécutives désigne un type d'acide aminé. Cette traduction est opérée à partir de la séquence des ARN messagers, copie de celle des gènes. Les protéines réalisent presque toutes les fonctions du vivant ; elles servent notamment de matériaux de construction dans la cellule, de récepteurs pour les signaux dans et entre les cellules, et de catalyseurs pour les réactions biochimiques. Dans nombre de maladies génétiques, une mutation dans un gène se traduit par la synthèse d'une protéine défectueuse, ou par son absence.

Séquence, séquençage
Le séquençage est l'opération qui consiste à déterminer l'ordre, ou séquence, dans lequel se succèdent les bases le long d'une molécule d'ADN. Une séquence se présente donc comme un texte plus ou moins long, écrit uniquement avec les quatre lettres A, T, G et C (voir Base). Le Genoscope a ainsi déterminé les 87 millions de lettres de la séquence ininterrompue du chromosome 14 humain.

Transformation
Opération qui consiste à faire pénétrer un vecteur (avec l'insert qu'il contient) dans une bactérie afin de réaliser le clonage de cet insert. On utilise aujourd'hui, avec une bonne efficacité, des chocs électriques très brefs.

Vecteur
Petites molécules d'ADN circulaires présentes dans les bactéries à côté du chromosome bactérien. Extraites des bactéries, elles sont utilisées pour le clonage moléculaire de fragments d'ADN à étudier. Après que le fragment y a été inséré (d'où son nom d'insert), le vecteur est réintroduit dans une cellule bactérienne par transformation.