

Les puces à ADN : outils pour une nouvelle vision de la diversité et des ressources génétiques

Pierre J.L. Lagoda, Farid Regad

Les efforts de séquençage systématique des génomes de virus, bactéries et organismes supérieurs, dont l'homme, ont conduit les techniques d'acquisition de données moléculaires à se développer. Le coût d'acquisition de ces données a été divisé par 10, les instruments sont devenus plus performants et un raz-de-marée de données doit être traité. Autrement dit, les grands projets de séquençage du génome furent les premiers reflets d'un iceberg qui se profilait à l'horizon des nouvelles technologies moléculaires. Aujourd'hui, la partie visible en est la génomique structurale et fonctionnelle. Demain, on découvrira la partie immergée qui cache un changement de paradigme sur la manière d'appréhender les aspects fondamentaux et appliqués de la biologie. Il s'agit, en fait, d'une nouvelle vision du vivant dans ses aspects les plus intimes, dont les techniques moléculaires et informatiques à grand débit (membranes denses, *macro* et *microarrays* et finalement puces à ADN) seront les instruments. De l'analyse statique du génome on s'oriente vers une analyse dynamique du polymorphisme des niveaux d'expression du transcriptome.

Dans un premier temps, les puces à ADN nous permettront un changement d'échelle au niveau du génotypage : la

quantité de locus et le nombre d'échantillons à analyser, et donc de données à acquérir et à traiter, augmenteront considérablement. Bien que classique, l'application des puces à ADN aux problèmes des ressources génétiques permettra un gain en efficacité appréciable. Mais cette application des méthodes à grand débit n'épuise pas toutes les possibilités de ces techniques émergentes. Dans un deuxième temps, le niveau d'expression de tous les gènes transcrits devient mesurable, rendant le transcriptome accessible à l'analyse. Les profils d'expression des gènes en fonction des stress biotiques et abiotiques peuvent devenir des instruments de gestion de la biodiversité comme le sont déjà les profils des empreintes génétiques. C'est la genèse des méthodes à grand débit qui a ouvert de nouvelles perspectives et c'est leur développement qui est le moteur de nouvelles découvertes. Dans cet article, pour mieux appréhender leur potentiel utile à l'étude de la biodiversité et la gestion des ressources génétiques, nous nous efforcerons de décrire les puces à ADN, sans pour autant oublier de mentionner d'autres techniques apparentées, développées en parallèle pour l'étude de l'expression et de la structure du génome.

Changement d'échelle

Dès lors qu'*Homo sapiens* passa du stade de chasseur-cueilleur nomade au stade de cultivateur sédentaire, il tenta d'identifier des plantes et des animaux vigoureux,

résistants et productifs pour transmettre intuitivement ces caractères à la génération suivante. La pression démographique le poussa à augmenter régulièrement la quantité et la qualité de sa production agricole. Localement, des systèmes empiriques, de plus en plus fins et complexes, se sont construits sur le savoir des générations passées. C'est avec Mendel, en 1866 [1], que cet art des cultivateurs et éleveurs a été formalisé en une science, la génétique. La capacité à distinguer ou identifier un individu ou un groupe d'individus a des applications dans beaucoup de domaines et a récemment été rendue plus efficace par l'utilisation de techniques moléculaires. Les marqueurs génétiques permettent une description des espèces cultivées ou spontanées (par ex. identification de variétés). Pour la sélection, ils facilitent le choix de parents et aident dans le développement de nouvelles stratégies d'amélioration (sélection assistée par marqueurs). En particulier pour les plantes cultivées, le génotypage individuel peut être employé pour étudier et pour comparer les accessions différentes ou groupes d'accessions dans des collections. Des polymorphismes ténus, induits par les manipulations *in vitro* de plantes pour la propagation en masse, peuvent ainsi être identifiés et mesurés (c'est-à-dire pour l'étiquetage de hots-types, variants somaclonaux).

Un réservoir pratiquement inépuisable de polymorphismes est le génome lui-même, soumis aux effets mutagènes de l'environnement. De 0,5 à 10 nucléotides par 1 000 bases différent d'un individu à un

P.J.L. Lagoda, F. Regad : CIRAD-AMIS BIOTROP TA40/03, F-34398 Montpellier cedex 5, France.

Tirés à part : P. Lagoda

autre non apparenté, en fonction de l'espèce et de la zone du génome considéré [2]. La majorité de ces mutations ne se trouvent pas dans des gènes, mais dans des régions intergéniques non codantes. En fonction de l'espèce, l'ADN codant ne représente que 5 à 40 % du génome des organismes supérieurs.

Dans le passé, les caractères ou traits génétiques complexes ont été identifiés par leur phénotype visible. Les marqueurs moléculaires se rapportent directement au génotype des organismes plutôt qu'à leur phénotype et sont fondés sur la détection de différences dans la séquence nucléotidique de l'ADN entre des individus différents (apparentés ou non). Les marqueurs moléculaires ont une résolution plus grande et peuvent être employés aux premiers stades du développement de la plante.

Il n'est pas facile de couvrir le génome d'un organisme supérieur d'un réseau de marqueurs moléculaires, véritables balises ou étiquettes génétiques, pour avoir une bonne chance statistique de trouver une liaison avec un gène quelconque. Les méthodes moléculaires actuelles (RFLP, VNTR, AFLP, RAPD) puisent dans un sous-ensemble de tous les polymorphismes nucléotidiques disponibles (cf. article de S. Santoni). Le développement des techniques, induit par les projets de séquençage de grande envergure (projets de séquençage complet de génomes : nématode, *Arabidopsis*, riz, humain et autres), permet d'accéder à la totalité des polymorphismes mononucléotidiques d'un génome (SNP) [2] par séquençage comparatif de différents individus ou écotypes.

Les SNP sont des variations de séquences codantes (cSNP) ou non, intra ou extragéniques. Les SNP sont beaucoup plus nombreux que les RFLP et les microsatellites. Sur la base d'un SNP toutes les 1 000 paires de bases, ils pourraient représenter de 0,1 à 16 Mb de différences cumulées pour un génome de plante haploïde de 0,1 à 16 Gb. La caractérisation systématique de ce genre de marqueurs génétiques va permettre des études génétiques à grande échelle. Une base de données, dbSNP [3], a été créée au NCBI (National Center of Biotechnology Information, États-Unis) en collaboration avec le NHGRI (National Human Genome Research Institute, États-Unis) pour stocker et rendre accessible les 18 164 SNP mis en évidence à ce jour. Mise en place en octobre 1998, le NHGRI, où plusieurs équipes sont

impliquées, a financé un programme intensif pour collecter 50 000 SNP sur 3 ans [4]. Actuellement, la base de données ne contient que des SNP humains mais elle a été conçue pour stocker les informations de toute espèce [5, 6]. L'apport des méthodes à grand débit est d'abord une optimisation des outils moléculaires (encadré 1). En général, elles permettent de cribler plus d'individus par plus de marqueurs en un temps réduit. Dans une région du génome, le

nombre de marqueurs est maximisé en puisant dans les SNP locaux, ce qui pallie le problème des marqueurs rendus non informatifs par manque de polymorphisme. Ces technologies émergentes sont donc plus productives, mais il reste à étudier leur fiabilité et à en diminuer les coûts. La miniaturisation des supports permet de réduire la quantité d'ADN nécessaire et les volumes à manipuler. Ceci permet de travailler sur des plantules dont on conserve l'intégrité

Encadré 1

Marqueurs moléculaires et ressources génétiques

Jusqu'à présent, la gestion des ressources génétiques pour l'amélioration des plantes cultivées pouvait être qualifiée d'essentiellement descriptive. En effet, la méthodologie de base de cette approche est la génétique. Elle repose sur l'identification et l'analyse de variants présents dans les ressources génétiques naturelles (biodiversité), ou induits par mutagenèse artificielle – physique (rayonnements ionisants), chimique (agents alkylants par ex.) et biologique (T-ADN par ex.). L'élaboration de croisements et l'étude de la transmission des variations identifiées dans leurs descendance permettent d'analyser le déterminisme génétique des variations observées ainsi que de dénombrer les facteurs génétiques responsables et leur emplacement relatif sur des groupes de liaison représentant les chromosomes. Ces variations peuvent affecter des caractères discrets (par ex. gènes de résistance) mais aussi quantitatifs. Ces caractères quantitatifs sont plus complexes, généralement contrôlés par plusieurs gènes. La plupart des caractères d'importance agronomique (comme la qualité, l'arôme, le rendement, la hauteur de la plante par ex.) sont d'ailleurs déterminés par l'action, et l'interaction, d'un grand nombre de gènes dont il s'avère difficile, voire impossible, d'isoler des gènes majeurs. Les techniques utilisées sont moléculaires et se fondent sur l'utilisation de marqueurs qui détectent un sous-ensemble de mutations du génome, liées statistiquement aux gènes intéressants car co-ségrégant avec les caractères (QTL) intéressants. Le code à barre génétique ainsi constitué et l'étude de ses variations peuvent être corrélés aux variations phénotypiques afin d'identifier et d'évaluer des populations et des individus utiles. Ces accessions utiles sont conservées dans des *core collections* ou par d'autres méthodes *in vitro* du génie biotechnologique (par ex. la cryopréservation). Au niveau de la conservation, il s'agit de contrôler la stabilité génétique de ces accessions. Les marqueurs les plus proches de caractères intéressants servent d'étiquettes génétiques. On peut contrôler, par le biais de la détection de la présence et/ou de l'absence de ces étiquettes, l'introgession de régions génomiques portant des caractères intéressants d'une espèce à une autre, et la stabilité génétique de cette introgession.

Typiquement, dans cette approche, le génome est couvert par un réseau de balises moléculaires polymorphes (empreinte moléculaire du génome) qui, localement, peuvent approcher un maillage fin de l'ordre du 0,1 cM en fonction de la nature du marqueur (dominant ou co-dominant), de la nature du croisement (F2, BC par ex.) et de la taille de la population de descendants étudiés. Pour augmenter la résolution des marqueurs, il faut augmenter la taille des populations (nombre de méioses analysées). Mais la fréquence de recombinaison n'est pas uniforme : il existe des régions plus stables que d'autres, qui recombinent moins. Dans ces cas, il faut créer une nouvelle population de croisement pour pouvoir cartographier ces régions réticentes.

Molecular markers and genetic resources

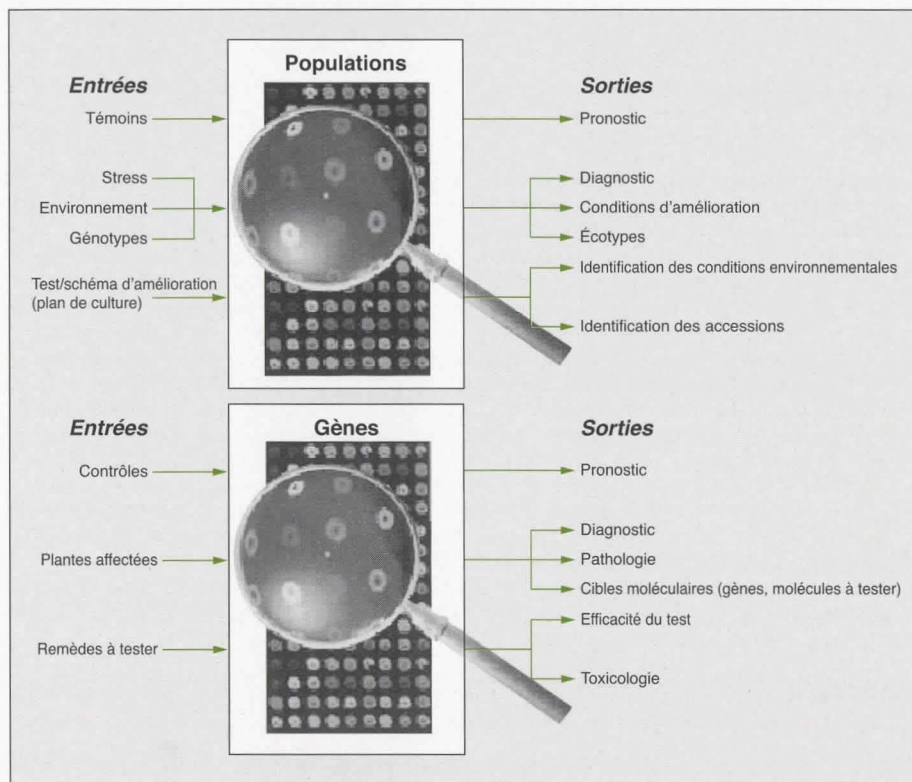


Figure 1. Paradigme. Apport méthodologique des techniques à grand débit (voir encadré 2).

Figure 1. Paradigm. Methodological input of high throughput techniques.

grâce à des biopsies moins destructrices, augmentant ainsi la précocité des marqueurs.

Changement de paradigme

Par leur productivité, rapidité et fiabilité accrues, les technologies émergentes à grand débit sont les outils de choix. Les données accumulées depuis le lancement des grands projets de séquençage systématique sur des organismes modèles ont donné naissance à un nouveau domaine de recherche : la génomique. Se sont alors associées à l'étude du génome (génomique structurale) deux facettes étudiées par la génomique fonctionnelle : le transcriptome (l'ensemble des gènes transcrits) et le protéome (l'ensemble des gènes traduits). Ces trois niveaux moléculaires qui font le lien entre génotype et phénotype sont en interactions complexes formant un véritable « généticien » [7] de boucles rétroactives, positives et négatives, qui

contrôlent l'expression (transcription) et l'activité (épissage régulier et alternatif, modifications post-transcriptionnelles et post-traductionnelles) des gènes.

L'analyse comparative des données accumulées sur les génomes de référence a produit des résultats intermédiaires qui orienteront les développements méthodologiques des recherches à venir [8]. L'arrangement des gènes à l'intérieur d'un génome est semblable chez des espèces voisines et plus ou moins conservé chez des espèces plus éloignées (synténie). D'où l'émergence de l'idée des espèces modèles et charnières ou relais : les possibilités de transfert des résultats d'une espèce vers d'autres, plus complexes, ou vers des espèces moins étudiées (orphelines). Ainsi, on a pu vérifier qu'approximativement un tiers des gènes (toutes espèces animales et végétales confondues) sont fortement conservés au niveau de leur séquence primaire chez tous les organismes étudiés. Cet ensemble de non-variants constitue un stock de fonctions primordiales, essentielles au développement des organismes et conservées au cours de l'évolution.

Génomique fonctionnelle et protéomique ont pris de l'importance grâce aux méthodes à grand débit. Les premières analyses structurales des génomes des eucaryotes ont montré chez la majorité des organismes supérieurs que la majorité du génome est non codant. L'analyse fonctionnelle comparative a permis de vérifier que la plupart des gènes, à un moment donné, sont dormants. Des mécanismes génétiques et épigénétiques modulent l'expression des gènes, dont les différents facteurs commencent à être identifiés. Ainsi l'ADN répété non codant, souvent décrit comme ballast moléculaire ou rebut génique (*junk DNA*), voire charge génétique (*selfish DNA*), est impliqué dans des processus de répression chromosomique régionale. Ils affectent des bras entiers par des modifications chimiques (méthylation, acétylation, phosphorylation et leurs contraires) ou peuvent avoir des cibles plus localisées (un gène unique ou un groupe de gènes adjacents). Cette adaptabilité dynamique de la chromatine aux sollicitations externes permet la désactivation ou au contraire l'activation de gènes. Elle est régulée par des gènes impliqués dans le cycle cellulaire (facteurs de transcription, récepteurs, cyclines, kinases) et le développement (embryogenèse, différenciation cellulaire, régulation de l'organogenèse), qui sont affectés par les signaux extérieurs. L'observation et l'analyse comparative du transcriptome de différents organismes dans différents stades de développement et sous différentes sollicitations (situations de stress biotiques et abiotiques) permettront l'émergence d'une véritable physiologie moléculaire intégrée [9]. Différents profils de transcription de l'organisme dans différents états peuvent être comparés dans un grand nombre de conditions expérimentales grâce aux méthodes à grand débit. Le quotient d'expression entre conditions expérimentales et témoins est calculé gène à gène, produisant des données sur des sous-ensembles de gènes en fonction des similitudes de leur profil d'expression. En d'autres termes, le marqueur n'est plus muet, mais il se confond avec le ou les gènes/allèles étudiés. Il est donc possible de constituer des groupes de corégulation, par exemple, d'en comparer les promoteurs et d'identifier des facteurs de régulation transcriptionnelle. L'attribution d'une fonction potentielle à un gène inconnu est facilitée par son appartenance à un groupe de gènes dont les fonc-

tions ont déjà été élucidées. L'étude d'un mutant déficient de ce gène permettra de valider ou d'infirmer l'hypothèse (figure 1, encadré 2).

Définitions et explications

En 1989, une équipe yougoslave et une équipe russe [10, 11] ont l'idée du concept de séquençage par hybridation. Le résultat d'une hybridation est aisément codé de façon binaire : s'il y a hybridation, les brins d'ADN sont complémentaires, bit = 1, s'il n'y a pas d'hybridation, les brins d'ADN ne sont pas complémentaires, bit = 0. D'où l'idée des bio-puces (*Gene Chips*, *DNA chip technology*, *microarrays* ; en français : bio-puces, puces à ADN) qui a été rapidement reprise par le secteur privé (Affymétrie) avec la participation active d'Eric Lander (Whitehead Institute, États-Unis).

Le concept repose sur la mise en contact d'un fragment d'ADN connu et/ou synthétisé avec un substrat. Cette méthode est très complexe et longue, puisqu'il faut des dizaines de milliers d'oligonucléotides (octamères). Au départ, le support était une membrane de nylon, puis un petit support solide qui ne ressemble pas vraiment à une puce électronique mais qui en reprend le principe. On sait aujourd'hui déposer 25 000 fragments d'ADN sur une surface de 5,3 cm x 5,3 cm, soit 1 dépôt par 300 µm. Le nombre d'échantillons d'ADN gravés sur une puce est passé de 20 000 à près de 1 million en seulement 3 ans, l'équivalent d'une loi de Moore pour les outils de la génomique ! En 1965, Gordon Moore, le cofondateur de la société Intel, observe que le nombre de transistors par inch au carré double tous les ans depuis l'invention des circuits intégrés. En 1985, la densité doublait tous les 18 mois, aujourd'hui elle double tous les 2,5 ans. Moore prédit que cette tendance devrait encore continuer jusqu'en 2017 avant de rencontrer des limites physiques incontournables. (<http://www.physics.udel.edu/wwwusers/watson/scen103/intel.html>). Grâce aux microprocesseurs, à la micro-fluidique (nano-PCR) [12] et aux bases de données, les spécialistes de la biologie moléculaire, tout en conservant le même savoir-faire, peuvent délaissier leurs tubes Eppendorf pour créer des systèmes informatiques capables d'effectuer l'analyse de centaines de fragments d'ADN en quelques secondes.

Encadré 2

Paradigme

Dans le domaine de la génétique des populations, l'acquisition de la séquence complète de génomes entiers ainsi que le développement de technologies pour déterminer la variation de séquence pan-génomique engendreront des études exhaustives de la diversité génomique comparative entre populations. Ces études, tout en apportant une nouvelle jeunesse à la génétique des populations, construiront les bases d'une connaissance de la genèse des variations génétiques et de leur maintien, ce qui mènera finalement à la compréhension des caractères complexes. Les techniques à grand débit, au niveau du génome et du transcriptome, donnent une représentation systémique de l'élément étudié. Par exemple, des changements dans l'état physiologique de cellules et/ou d'organes se reflètent au niveau des profils d'expression des gènes (voir figure 1). L'ARNm d'échantillons d'intérêt (entrées) est extrait, étiqueté et analysé par hybridation sur des puces à ADN. Le résultat de cette expérience produit de l'information quantitative sur l'expression de milliers de gènes cellulaires. Ces données d'expression seront corrélées à d'autres types d'informations (données agro-morphologiques et/ou phytopathologiques) qui auront une influence orientante et déterminante sur un éventail de domaines de recherche (sorties).

Paradigm

Dans la suite, nous nous proposons de présenter une liste non exhaustive de ces méthodes à grand débit, applicables aussi bien à l'étude de la structure du génome qu'à l'analyse des niveaux d'expression du transcriptome.

Macroarray

Les *macroarrays* sont des membranes de nylon sur lesquelles sont déposées de l'ADN (oligomères, ADNc, ADNg) ou des bactéries. Les différences avec les *dot-blots* traditionnels sont la densité de dépôt (36 000 à 50 000 sur une surface de 22 cm de côté) et le robot qui permet de produire ces membranes à haute densité. De l'ADN, des produits PCR, ou des bactéries sont ordonnés, en rangs serrés, sur une membrane de nylon selon un canevas qui permet de reconnaître, au niveau du signal (dupliqué), l'identité du clone et la légitimité du signal observé en éliminant les faux positifs qui donnent des signaux simples asymétriques (figure 2A). En ce qui concerne les membranes à ADNc, elles sont hybridées avec des sondes complexes dans des conditions d'excès de cible par rapport à la sonde. Ces sondes complexes sont issues d'une rétrotranscription d'ARN totaux ou d'ARNm. La quantité minimale de

molécules permettant la détection, eu égard à la sensibilité des différentes méthodes de détection (anticorps/enzymatique, radioactivité), est de l'ordre de 10^7 molécules. La limite de détection d'une espèce d'ARN messenger est de 10^{-5} . Elle est inversement proportionnelle à la concentration de la sonde complexe marquée. En d'autres termes : une molécule d'un ARN messenger spécifique par cellule faisant partie de l'échantillon de tissu source de l'ARN extrait est détectable [13]. La saisie des données peut être automatisée via un écran phosphorescent par la numérisation de la quantité de signal. Comme la cible est en excès par rapport à la sonde, l'intensité du signal, proportionnelle à la quantité de sonde fixée, est directement liée à la quantité relative de chacun des ARNm dans l'échantillon préparé. Ainsi, un spectre de niveaux d'expression peut être observé. Pour la cartographie physique, des banques entières de bactéries, intégrant de larges fragments génomiques, peuvent être déposées d'une façon ordonnée sur une seule membrane. Les sondes sont dans ce cas des marqueurs de cartographie génétique classiques. Le marquage des sondes est essentiellement radioactif, faute de pouvoir utiliser une détection fluorimétrique à cause de

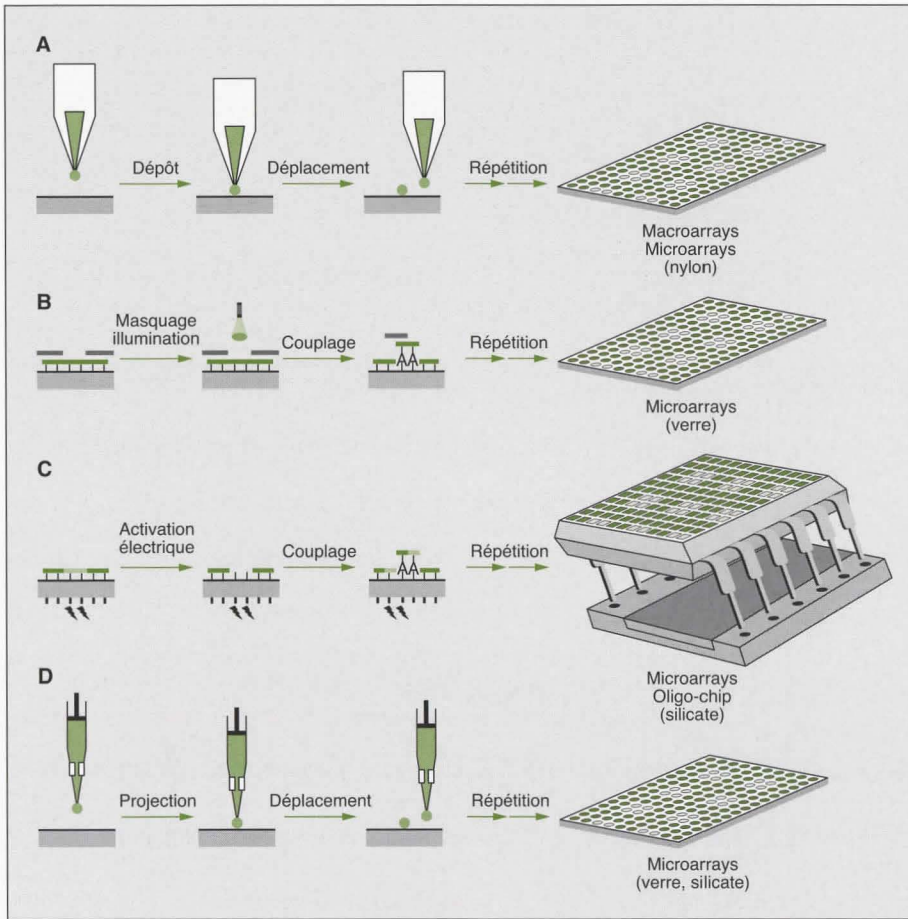


Figure 2. Fabrication de *microarrays*. Quatre méthodes de fabrication de *microarrays* sont décrites.

A. Dépôts robotisés : pour la fabrication de *macro* et de *microarrays*, un volume déterminé d'échantillon est chargé par capillarité sur des aiguilles pour être transféré par simple contact sur le substrat. Après le premier cycle de dépôts les aiguilles sont nettoyées et un second échantillon est chargé pour être adressé sur un site adjacent. L'automatisation des dépôts permet de faciliter la fabrication de ces *microarrays*.

B. Photolithogravure : après avoir déposé un substrat photosensible (-) sur le verre, on éclaire la puce vierge avec de la lumière ultraviolette à travers un masque. Les zones où l'on souhaite déposer un nucléotide spécifique (A : adénine) sont transparentes, les autres restant opaques à la lumière. Les zones éclairées de la puce sont activées et prêtes à recevoir des nucléotides. On verse ensuite une solution du nucléotide choisi (par ex. adénine), marquée d'une molécule photosensible. Ces chimères viennent se fixer sur les zones activées lors de l'étape précédente. L'étape de masquage est à nouveau réalisée, mais cette fois avec un schéma comportant les emplacements où l'on déposera la deuxième base. Le cycle masquage/fixation est alors renouvelé avec une deuxième puis une troisième base et ainsi de suite.

C. Synthèse électrochimique : les sites où vont être déposées des molécules d'ADN sont chargés électriquement. Les zones de la puce sont ainsi activées sont prêtes à recevoir des nucléotides. On verse ensuite une solution du nucléotide choisi (par ex. adénine), qui se lie électrochimiquement au substrat. Ensuite, un schéma de mise sous tension, comportant des adresses différentes, où l'on déposera une deuxième base, est appliqué. L'élongation des oligonucléotides se fait par répétition dirigée de l'application de champs électriques localisés. La réaction chimique d'élongation par polymérisation est activée par ce même champs.

D. Jet d'encre : l'échantillon est chargé dans une buse miniature équipée d'un système piézo-électrique (rectangles) puis un courant est appliqué pour projeter une quantité déterminée de liquide sur le substrat. Après le premier dépôt, la buse est nettoyée et le second échantillon est chargé puis déposé sur un site adjacent.

Figure 2. Microarray production. Four microarrays production methods are described.

A. Mechanical microspotting : a biochemical sample is loaded into a spotting pin by capillary action, and a small volume is transferred onto a solid surface by a physical contact between the pin and the solid substrate. After the first spotting cycle, the pin is washed and the second sample is loaded and deposited to an adjacent address. Robotic control systems and multiplexed printheads allow automated macro and microarray production

B. Photolithography : a glass wafer modified with photolabile protecting groups (-) is selectively activated for DNA synthesis by shining light through a photomask. The wafer is then flooded with a photoprotected DNA base (A : Adenine), resulting in spatially defined coupling on the chip surface. A second photomask is used and the procedure is repeated until the oligonucleotide synthesis is completed.

C. Electrochemical synthesis : DNA deposit sites are electrically charged. The silicate wafer is then flooded with a DNA base (A : Adenine) who is electrochemically linked to the activated sites. The electrical tension pattern is changed and the procedure is repeated until the oligonucleotide synthesis is completed.

D. Ink jetting : a biochemical sample is loaded into a miniature nozzle equipped with a piezoelectric fitting (rectangles) and an electrical current is used to expel a precise amount of liquid from the jet onto the substrate. After the first jetting step, the nozzle is washed and a second sample is loaded and deposited to an adjacent address.

l'auto fluorescence des supports (nylon). Cependant une détection colorimétrique à haute résolution a été développée [14], permettant d'atteindre des densités appréciables avec des espacements de l'ordre des 300 microns à la limite des *microarrays*. Les filtres à haute densité semblent faire l'unanimité pour leur sensibilité, leur relative facilité de manipulation et leur prix de revient pour le moment inférieur à celui des *microarrays* [15, 16].

Microarray et bio-puce

Microarray

Les *microarrays* reposent sur le développement des membranes à haute densité à dépôts d'ADN (oligomères, ADNc) en changeant de support : du nylon vers le verre ou une platine microélectronique de silicate (*chip*) [17]. Le passage du nylon vers le verre permet l'utilisation du marquage fluorescent et réduit les

volumes à manipuler de l'ordre des dizaines de millilitre vers l'ordre du microlitre. Les conditions d'hybridation sont les mêmes que celles décrites pour les *macroarrays* d'ADNc.

Bio-puce

Pour les dépôts sur *chip*, les circuits microélectroniques de la platine de silicium permettent de créer des champs électriques, localisés sur le support, qui servent à orienter et focaliser les dépôts

individuels (bio-puce). Lors de l'hybridation, l'application d'un champ électrique permet de contrôler localement la stringence. Il semble même possible de mesurer la résistivité en chaque point, dépôt à dépôt, et ainsi de distinguer un simple brin d'ADN (la sonde ne s'est pas attachée à la cible) d'un double brin (homologie de séquence avec la sonde). La taille de ces *microarrays* et bio-puces varie entre 5 cm de côté pour les plus larges à 1,5 cm de côté [18]. En fonction de la conception et de la nature du support, on peut distinguer des puces à faible densité (de l'ordre de centaines de dépôts), moyenne densité (de l'ordre des milliers de dépôts) et haute densité (de l'ordre de dizaines de milliers de dépôts ordonnés).

Oligo-chip

Au lieu de déposer des ADNc mécaniquement sur des supports, il est possible de synthétiser des oligomères (10 à 25 mères) directement sur le support. On parle alors plutôt d'*oligo-chip*. Les oligo-puces font partie des puces les plus petites et les plus denses. Les capacités actuelles de dépôt atteignent 250 000 oligonucléotides sur une surface de 1,28 x 1,28 cm [6]. Il existe essentiellement trois méthodes concurrentes pour produire un *oligo-chip*.

Photolithographie

Responsable de la miniaturisation des composants électroniques (semi-conducteur), cette technique, développée par Affymetrix, permet une précision extrême dans la fabrication des bio-puces. On procède pour cela en plusieurs étapes (figure 2B).

- Après avoir déposé un substrat photosensible sur le verre, on éclaire à travers un masque la puce vierge avec de la lumière ultraviolette. Ce masque est une plaque de quartz, sur laquelle a été dessiné le schéma de la puce. Les zones où l'on souhaite déposer un nucléotide spécifique (par ex. de l'adénine) sont transparentes, les autres restant opaques à la lumière. Les zones éclairées de la puce sont activées et prêtes à recevoir les nucléotides.

- On verse ensuite une solution du nucléotide choisi (par ex. adénine), que l'on a préalablement marquée de molécules photosensibles. Ces chimères viennent se fixer sur les zones activées lors de l'étape précédente. Dix millions de molécules identiques sont fixées sur un site de détection.

- L'étape de masquage est à nouveau répétée, mais cette fois avec un schéma comportant les emplacements où l'on déposera une deuxième base (par ex. cytosine).

- Les molécules de la deuxième base (par ex. cytosine), elles aussi flanquées d'une substance photosensible, sont déposées sur la puce.

- Le cycle masquage/fixation est alors renouvelé avec une troisième et une quatrième base et ainsi de suite.

- À force de construire les brins d'ADN brique à brique, on dispose enfin d'une puce dotée de milliers de sites comportant chacun un type d'ADN. Avec 64 sites, on peut représenter toutes les variations de molécules à trois bases d'ADN (4 à la puissance 3). Avec 1 024 sites, on construira donc toutes les combinaisons

de 5 bases, et ainsi de suite. Cette méthode est pour le moment très onéreuse. De plus, il est à noter que l'efficacité de synthèse avoisine 92-94 % à chaque étape de couplage [9], ce qui peut engendrer un bruit de fond d'autant plus important que l'oligonucléotide à synthétiser est long.

Synthèse électrochimique

En remplaçant le support en verre par du silicium, et les photons par des électrons, les propriétés électriques du semi-conducteur et des nucléotides permettent de contrôler avec précision la polymérisation des oligonucléotides. Chaque site du support est équipé d'une broche métallique, à la manière de celles des microprocesseurs (figure 2C); la réaction de polymérisation n'a lieu que sur les

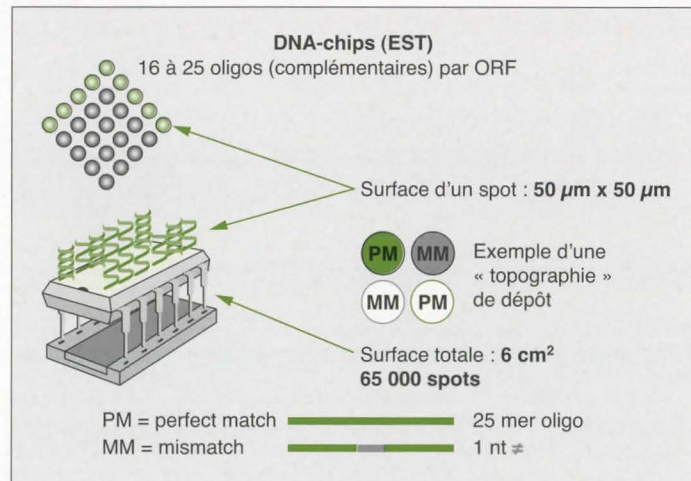


Figure 3. Tests de validation interne. Membranes à haute densité : nylon, 22 cm x 22 cm, 36 000 à 50 000 dépôts.

Microarray : verre, 1,8 cm x 1,8 cm, 6 400 dépôts.

Oligo-chips : verre ou silicium, entre 2,5 cm x 2,5 cm et 1,28 cm x 1,28 cm de surface, 250 000 dépôts.

La construction de *macro-arrays*, *micro-arrays* et *oligo-chips* est organisée par le dépôt raisonné de contrôles internes. Par exemple : pour valider un signal positif sur un *oligo-chip* d'expression, on a pris soin de déposer les oligonucléotides (25 mer) représentant le brin sens ainsi que le brin anti-sens une fois en homologie parfaite (PM, *perfect match*, disques pleins) et une seconde fois avec une mutation ponctuelle et centrale (MM, *mismatch*, cercles). L'absence de signal sur les sites de mutation ponctuelle, idéalement, est un test de validité pour tout signal PM observé.

Figure 3. Internal validity tests. High density filters : nylon, 22 cm x 22 cm, 36 000 to 50 000 spots.

Microarray : glass wafer, 1,8 cm x 1,8 cm, 6 400 spots.

Oligo-chips : glass or silicate wafer, from 2,5 cm x 2,5 cm to 1,28 cm x 1,28 cm, 250 000 spots.

Macroarrays, microarrays et oligochips are organised by internal control spotting patterns.

As an example : in order to identify false positive signals on an oligo expression chip, sense and antisense 25mer oligonucleotides representing an open reading frame have been spotted in duplicates as perfect matches (PM, filled circles) and central point mutation causing mismatches (MM, open circles). Thus, a positive signal at a PM site is confirmed by the absence of a signal at the corresponding MM site.

sites chargés électriquement. La production d'une puce de silicium équipée d'une broche métallique coûte pour le moment 7 FF (1,1 €) [19].

Impression à jet d'ADN

Pour accélérer la fabrication des bio-puces, certains chercheurs ont imaginé de détourner la technique d'impression à jet d'encre. Le principe : une tête d'impression contient plusieurs solutions de fragments d'ADN qu'elle dépose ensuite avec une grande précision (0,5 µm) sur le support [2, 20, 21] (figure 2D).

Hybridation et détection d'un signal sur une bio-puce

Quelle que soit la méthode de fabrication de bio-puces employée, on utilise généralement l'hybridation de type Southern où la sonde est marquée par un fluorochrome. La fluorescence sert à détecter les zones où l'hybridation de l'ADN s'est produite en présence de l'échantillon à analyser. Sur des lamelles de verre, le contrôle de la stringence se fait classiquement *via* la température de *melting* (T_m). Sur une lamelle de silicium équipée de broches, la stringence de l'hybridation est contrôlée par l'application d'une différence de potentiel, localisée à chaque site de dépôt. En effet, l'application d'une charge électrique au voisinage de chaque site comportant des fragments d'ADN accélère fortement le mécanisme d'hybridation tout en augmentant sa précision.

Pour la lecture des signaux, on éclaire avec un laser approprié les brins d'ADN de l'échantillon ayant été marqués d'une molécule fluorescente. Les sites hybridés émettent alors une lumière que peut capter une caméra vidéo. Reste à transmettre les images à l'ordinateur, lequel sera chargé de traiter l'image pour donner les résultats de l'expérience. Une autre méthode serait d'utiliser la variation de la résistivité électrique locale au niveau des spots où une hybridation a eu lieu (quotient des constantes diélectriques simple brin/double brin), et de mettre au point une lecture électrique des puces, grâce aux propriétés du silicium. On peut en effet observer de faibles variations de charge au voisinage des zones où l'ADN a hybridé. Elles sont ensuite directement transmises à l'ordinateur et le résultat de l'expérience est

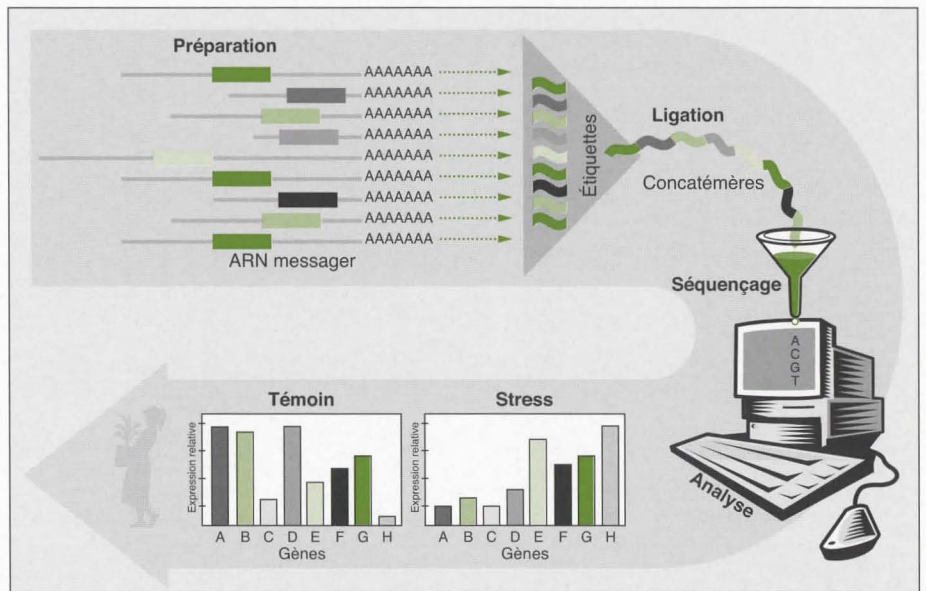


Figure 4. SAGE. Après la synthèse du brin complémentaire (ADNc) à l'aide d'une amorce oligo(dT) sur les ARN messagers (ARNm) à étudier, l'ADNc est digéré à l'aide d'une enzyme de restriction à coupures fréquentes dite enzyme d'ancrage qui va permettre la ligation d'un adaptateur contenant le site pour une endonucléase de restriction de type IIS possédant la particularité de couper l'ADN jusqu'à 20 pb du site reconnu. Les étiquettes ainsi générées sont ensuite liées entre elles puis amplifiées par PCR à l'aide d'amorces spécifiques aux adaptateurs. Les produits de cette amplification contiennent deux étiquettes appelées *ditag* qui sont ensuite digérées par une enzyme de restriction pour être mises bout à bout par ligation (concatamérisation) puis clonées en vue de leur séquençage. Le niveau d'expression d'un ARN messenger est donné grâce à une analyse informatique par la fréquence d'apparition de l'étiquette correspondante. SAGE semble être un outil de choix dans l'analyse de l'expression de gènes faiblement transcrits en comparaison des autres techniques développées à ce jour. Cependant, cette méthode nécessite la réalisation au préalable d'une base de données d'au moins 10 000 EST.

Figure 4. SAGE. After synthesis of the mRNA to be studied using an oligo(dT), a cDNA is restricted using a frequently cutting endonuclease. This « anchor enzyme » allows the ligation of adaptors containing a type IIS restriction endonuclease site. This enzyme cuts DNA up to 20 bp from the recognition site, thus creating tags which can be ligated and PCR amplified using adaptor specific primers. The amplification product then containing two tags is called « ditag ». The ditags are cut by an enzyme prior to ligation. The resulting concatamers are cloned into a vector in order to be sequenced. The sequence data are transferred into a database. The number of times each tag occurs measures the number of copies of the mRNA originally present in the biological sample. SAGE seems to be a tool of choice for the analysis of low transcription level genes compared to other present day techniques. However, a pre-existing database of at least 10 000 ESTs is needed for the successful application of the SAGE method.

obtenu en quelques fractions de secondes. Par construction, les puces à ADN peuvent aussi incorporer des tests internes de détection de faux positifs et/ou faux négatifs (figure 3). Il est possible de calibrer une puce par rapport au bruit de fond résiduel et d'incorporer cette donnée, caractéristique individuelle de la puce, aux logiciels de saisie [2]. L'augmentation de la fiabilité, l'élimination du bruit de fond (faux positifs) et la diminution des coûts de construction sont les défis des développements présents des puces à ADN. Les bio-puces, à l'instar de la révolution PCR, sont en passe de révolutionner tous les domaines où la biologie moléculaire est impliquée.

Autres méthodes à haut débit

Il existe d'autres méthodes à haut débit qui ne reposent pas sur une méthode de détection par hybridation. Trois de ces méthodes sont décrites ci-dessous.

SAGE : Serial Analysis of Gene Expression

La méthode SAGE [22] permet une analyse de la fréquence d'un ARN messenger, quel que soit son niveau d'expression,

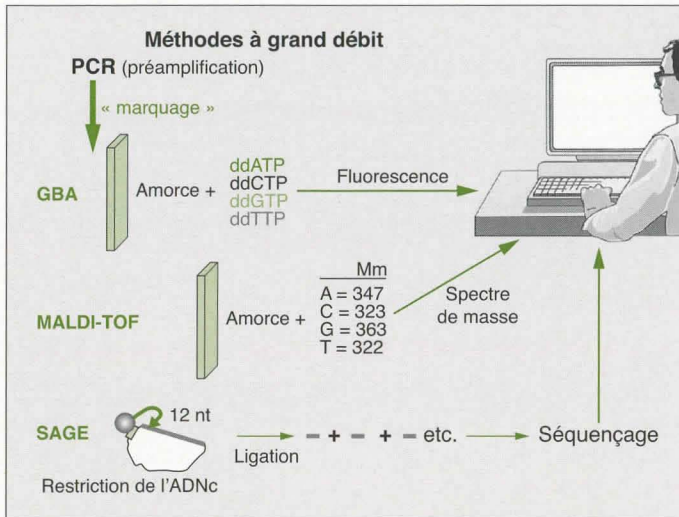


Figure 5. Méthodes à grand débit. Représentation schématique de trois méthodes à grand débit ayant pour point commun l'utilisation de la PCR. GBA, Genetic Bit Analysis, est une modification de la technique de séquençage du di-déoxy-terminateur de Sanger. MALDI-TOF, cf. figure 6 ; SAGE, cf. figure 4.

Figure 5. High throughput techniques. Three PCR based methods are depicted: GBA, Genetic Bit Analysis, is a modification of the Sanger di-deoxy-terminator sequencing technique. MALDI-TOF, see figure 6 ; SAGE, see figure 4.

parmi les milliers produits dans une cellule à un moment donné.

Cette technique est fondée sur trois principes.

Premièrement, sur le fait qu'une séquence d'une dizaine de nucléotides suffit pour constituer une étiquette caractéristique de chacun des ARN messagers d'une cellule. Une séquence de 9 nucléotides permet, statistiquement, de distinguer 262 144 transcrits alors que le génome humain, par exemple, pourrait coder environ 80 000 messagers.

Deuxièmement, les étiquettes peuvent être liées entre elles pour former une longue molécule composée d'étiquettes en série qui peut être clonée puis séquencée.

Troisièmement, la détermination du nombre d'apparitions d'une étiquette permet de connaître le niveau d'expression du transcrit correspondant (figures 4 et 5). Par contre, pour être efficace, la base de données servant à l'analyse doit contenir de l'ordre de 10^4 EST de l'organisme à étudier [23, 24]. Une application de cette méthode pour comparer des profils de transcription chez le riz a été récemment décrite par Matsumura *et al.* [25].

MALDI-TOF : Matrix-Assisted Laser Desorption Ionization/ Time Of Flight

La mesure de la masse moléculaire par spectrométrie de masse (MALDI-TOF) est devenue un outil de choix pour l'analyse des protéines, des acides nucléiques, de la structure, de la pureté, des modifications post-traductionnelles ainsi qu'un éventail d'autres caractéristiques moléculaires dont l'étude n'était pas facile par d'autres moyens. Les dernières machines disponibles sur le marché sont capables d'analyser des molécules de l'ordre du mégadalton avec une sensibilité de l'ordre de la femtomole avec un temps de détection proche de la microseconde (figures 5 et 6). De plus, la spectrométrie de masse représente une amélioration par rapport aux techniques fondées sur l'utilisation de gels et d'hybridation car on peut obtenir des informations précises sur l'élément étudié et l'analyse peut être complètement automatisée. L'ensemble de ces caractéristiques font du MALDI-TOF un instrument de choix dans l'analyse du protéome et du génome [26]. Récemment le MALDI-TOF a été utilisé pour la détection de polymorphismes de site ou polymorphismes ponctuels de

Encadré 3

Les enjeux économiques

Toutes les techniques décrites sont déjà sous brevets (il n'y a même pas un temps de grâce comme pour la PCR ou les RAPD) ! Il y a des batailles féroces aux États-Unis pour des parts de marché.

Les sociétés de biotechnologies s'installent en masse dans le sanctuaire de l'informatique (Silicon Valley).

– Bill Gates finance depuis plusieurs années les travaux de Leroy Hood, spécialiste du génome.

– Paul Allen, cofondateur de Microsoft, a investi dans Rosetta Inpharmatics, la firme qui commercialisera les produits développés par Leroy Hood (L'équipe de Leroy Hood, Rosetta, États-Unis, transformation de têtes d'impression à jet d'encre en imprimantes de fragments d'ADN).

– Microfab, une firme texane spécialiste de l'impression en qualité photo travaille aussi sur le jet d'ADN.

– Hyseq et Affymetrix, deux firmes pionnières dans la conception de détecteurs de gènes, ont réussi à lever près de 1 milliard de FF (plus de 150 millions €) en bourse (Affymetrix, Santa Clara, bio-puce en verre, étude des mutations des souches du virus du sida et de l'oncogène p53 impliqué dans l'oncogenèse).

– Nanogen, bio-puce au silicium. Construction d'une puce PCR à flux continu, la matrice de la PCR provenant d'une cellule unique.

– Motorola devrait lui aussi y venir.

Société	URL
Affymetrix	www.affymetrix.com
General Scanning	www.genscan.com
Genetic Microsystems	www.geneticmicr.com
Genometrix	www.genometrix.com
Genomic Solutions	www.genomicsolutions.com
Imaging Research	www.imaging.brocku.ca

DNA chips and business : the stakes

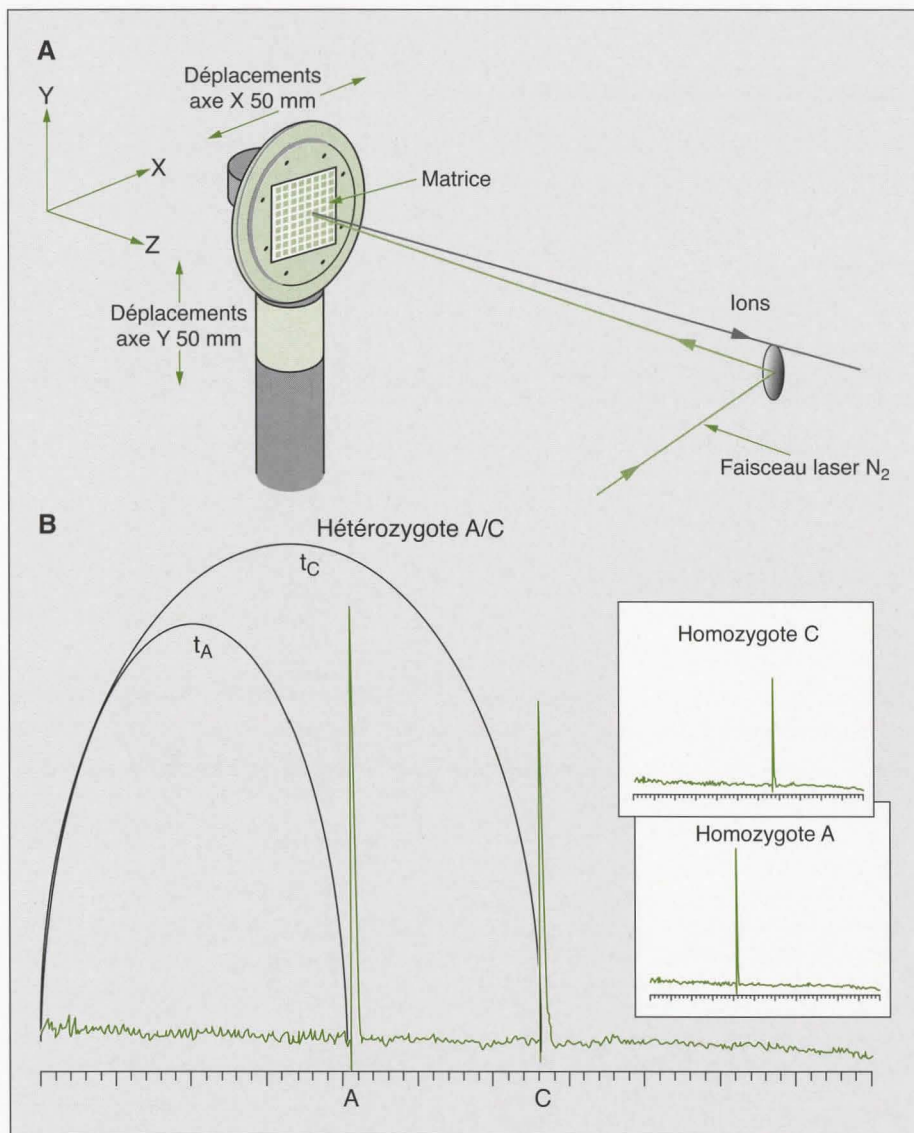


Figure 6. MALDI-TOF.

A. Description : L'échantillon à analyser, préalablement déposé sur une matrice de silicone (pouvant recevoir 96 ou 384 échantillons), est ionisé à l'aide d'un rayon laser. Par la suite, les ions sont accélérés dans un champs électrostatique ou électromagnétique vers l'analyseur de masses, puis séparés par temps de vol vers un détecteur. La matrice peut être déplacée grâce à un bras de type x, y et z, permettant au rayon laser de balayer d'une façon automatisée l'ensemble des échantillons déposés.

B. Utilisation du MALDI-TOF pour l'analyse de polymorphisme ponctuel (SNP)

Compte tenu de la sensibilité de la spectrométrie de masse, la présence d'une base mutée dans un oligonucléotide peut être aisément mise en évidence. La présence d'une adénine (A) comparée à une cytosine (C) dans l'oligonucléotide fera que le temps de vol $t_{A/C}$ de la molécule analysée sera différent. Il est donc facile, selon l'individu étudié, de déterminer s'il est homozygote ou hétérozygote pour tel SNP.

Figure 6. MALDI-TOF.

A. Description : Prior to LASER beam ionization, the samples to be analysed are put onto a silicon matrix (96 or 384 deposits). The ions are accelerated by an electrostatic or electromagnetic field in the mass analyzer and separated by time of flight to the detector. The sample matrix is then moved by a robot arm in the x, y and z dimensions, allowing the laser beam to cover the surface of the matrix by sweeping automatically over the whole set of spotted samples.

B. Using MALDI-TOF for SNP analysis

Considering the high sensitivity of mass spectrography, the presence of a mutated base in an oligonucleotide is easily detected. The difference in mass between an adenine (A) compared to a cytosine (C) in the oligonucleotide changes the time of flight $t_{A/C}$ of the molecule to be analyzed. Thus, it will be easy to determine whether the individual to be studied is homozygous or heterozygous at the corresponding SNP locus.

séquence ou SNP [27]. Cette méthode est fondée sur l'utilisation de la PCR pour produire des fragments alléliques à l'aide d'amorces spécifiques des locus à étudier. Les amorces sont fixées sur un support. L'ADN génomique des individus à tester est hybridé aux amorces. L'ADN polymérase rallonge l'amorce d'un di-déoxy-nucléotide sur le site du SNP préalablement identifié. Après lavage pour éliminer l'ADN génomique, les amorces allongées sont prêtes pour l'analyse. La spectrométrie de masse permet de repérer avec précision une différence dans la nature d'une base (A, C, G ou T) en comparant les masses moléculaires des produits PCR.

GBA : Genetic Bit Analysis

L'analyse génétique digitale (GBA) est une technique relativement nouvelle pour comptabiliser l'allélisme de régions génomiques en polymorphismes mono-nucléotidiques. Faisant partie de la « boîte à outils » du diagnostic de paternité humain, elle s'est répandue comme technique à grand débit pour le génotypage et l'analyse de descendance dans des élevages. Récemment, vu sa capacité à traiter les grands nombres, elle a trouvé des applications au niveau de l'amélioration des plantes [28]. La GBA est fondée sur deux opérations automatisables : la PCR et une réaction colorimétrique enzymatique pour différencier des génotypes (figure 5).

Conclusion

La naissance d'une nouvelle méthodologie, la génomique (structurelle et fonctionnelle), engendrée par le développement technologique à partir de la biologie moléculaire, parrainée par les projets de séquençage à grande envergure, ouvre de nouvelles perspectives. La génomique est née de la génétique moléculaire et de l'informatique. Ce qui a permis de la mettre au monde c'est la robotisation et la taylorisation des différentes procédures utilisées couramment dans des laboratoires de biologie moléculaire, ainsi que la miniaturisation des supports due au développement des nanotechnologies. Cette miniaturisation aidant, des laboratoires d'analyses moléculaire portables sont concevables. Ils permettront de génotyper des plantes aussi aisément aux champs que de noter leur phénotype ou d'évaluer les caractères agro-morphologiques.

Acronymes

ADN	Acide désoxyribonucléique (ADNc, codant ; ADNg, génomique)
AFLP	<i>Amplified Fragment Length Polymorphism</i> Polymorphisme de longueur de fragments amplifiés
ARN	Acide ribonucléique (ARNm, messenger)
BAC	<i>Bacterial Artificial Chromosome</i> Chromosome artificiel bactérien
EST	<i>Expressed Sequence Tag</i> Étiquette de séquence exprimée
GBA	<i>Genetic Bit Analysis</i> Analyse génétique digitale
MALDI-TOF	<i>Matrix-Assisted Laser-Desorption Ionization Time-Of-Flight</i> Mesure du temps de vol par spectrométrie de masse de désorption-ionisation par photons laser d'échantillons fixés sur matrice
QTL	<i>Quantitative Trait Locus</i> Locus de caractère quantitatif
RAPD	<i>Random Amplified Polymorphic DNA</i> Polymorphisme d'ADN amplifié aléatoirement
RFLP	<i>Restriction Fragment Length Polymorphism</i> Polymorphisme de longueur de fragment de restriction
SAGE	<i>Serial Analysis of Genome Expression</i> Analyse sérielle de l'expression du génome
SNP	<i>Single Nucleotide Polymorphism (cSNP, codant)</i> Polymorphisme mononucléotidique
VNTR	<i>Variable Number of Tandem Repeats</i> Nombre variable de répétitions en tandem

Acronyms

Au niveau du génome, les puces à ADN permettent d'effectuer un changement d'échelle, ce qui, au niveau du transcriptome, fait plus que potentialiser l'application des marqueurs moléculaires. Si la génétique permet d'isoler un gène à fonction connue et de l'étudier au niveau moléculaire, la génomique livre une pléthore de gènes (réels et potentiels) pour lesquels il faut déterminer la fonction. Les deux approches ne sont pas antinomiques, mais complémentaires. Prenant l'exemple de l'étude des résistances/tolérances aux maladies et ravageurs, (agro)physiologie, physiopathologie, phytopathologie, génétique, génétique des populations, génétique moléculaire et physiologie moléculaire trouveront une unification dans le domaine de la physiologie moléculaire intégrée (figure 1). L'étude du transcriptome se fait *via* des puces (ou d'autres méthodes à grand débit comme le SAGE) contenant la totalité (ou la majorité) des gènes (connus et/ou potentiels c'est-à-dire ADNc et EST) d'un organisme à analyser. Les profils de transcription se font par une hybridation unique avec une sonde complexe (l'ARNm de l'organisme dans un état donné). Pour utiliser une image, tous les gènes jouent un concert en fonction d'une

partition interprétée différemment selon les conditions génétiques (composition allélique) et épigénétiques (environnement, stress abiotique et biotique, par ex.). La génomique promet l'observation de l'expression de la totalité des gènes (sinon de la majorité du moins d'une très grande partie représentative) en même temps et avec les niveaux d'expression de chaque gène interdépendants. Au lieu d'isoler un gène (une voix de l'orchestre) on pourra écouter l'orchestre jouer le concert. Un gène mineur dans un état donné (jouant *pianissimo*) peut devenir un gène majeur (jouant *fortissimo*) dans un autre état de la plante ou sous différentes conditions de culture. Ed Coe s'est exclamé un jour : « Il y a tellement de gènes, et si peu de temps ! » C'était à la fin des années 80. Étudier l'expression d'un gène unique était une tâche qui occupait une période appréciable de la carrière d'un chercheur. Et chaque gène à étudier demandait une approche adaptée : le scientifique était occupé à développer, adapter et optimiser les techniques, il créait son instrument. À l'instar de la Chimie et de la Physique, où le progrès de la métrologie (science des mesures et du développement des instruments de mesure) a tou-

Summary

DNA chips: diversity and genetic resources revisited

P.J.L. Lagoda, F. Regad

At the brink of the third millenium, biology and informatics merged – giving rise to bioinformatics and genomics as a result of the massive inflow of data from international whole genome sequencing initiatives. High throughput techniques had to be developed to optimize existing molecular biology tools. More molecular markers can be tested on more individuals in less time. Comparative sequencing of related ecotypes is unravelling a bottomless source of molecular polymorphisms (SNPs) that could be tapped by geneticists and breeders. Moreover, the same techniques markedly increase functional genome analysis prospects (Figure 1). Changes in the physiological state of cells and tissues used for comparative analysis lead to specific changes in gene-expression patterns. Messenger RNA from samples of interest is isolated, labelled and analysed to yield quantitative expression information for thousands of cellular genes. Expression data, coupled with other types of information (physiological, agromorphological, genetic, etc.) can have an impact on a wide range of research areas (prognosis, diagnosis, pathology, etc.). Advances in genetic diversity and genetic resource management are linked with the development of high throughput techniques – the main ones are described here: i) techniques based on Southern detection: macroarrays, microarrays and oligochips (Figures 2 and 3). ii) PCR-based techniques: SAGE (Figure 4), MALDI-TOF (Figure 5) and GBA (Figure 6). Besides the obvious shift in the data acquisition scale, the overall fields of biodiversity analysis and genetic resource management are evolving into a new paradigm – from the analytic to the systemic and from the descriptive to the predictive.

Cahiers Agricultures 2000 ; 9 : 329-40.

jours entraîné une évolution de la pensée scientifique (et vice versa), les technologies émergentes promettent de libérer les biologistes de cette partie (plus ou

Utilisations envisageables pour les bio-puces

- Séquençage de *novu*.
- Étude de la variabilité génétique (étude de diversité, ressources génétiques) :
 - identification de gènes mutants : grâce aux lois de complémentarité des nucléotides, les brins d'ADN de l'échantillon analysé viennent se fixer (hybrider) sur les séquences complémentaires fixées sur la bio-puce. Une détection appropriée permet alors de savoir où l'hybridation s'est produite, et donc de déterminer l'éventuelle mutation d'un gène ;
 - reséquençage allélique systématique pour déterminer un polymorphisme.
- La cartographie génétique et l'étude de diversité de troisième génération. En ce moment, on travaille avec des microsatellites (quelques milliers de marqueurs pour l'homme. Par exemple, pour la localisation d'un gène morbide, on utilise un sous-ensemble de ces marqueurs, 800 à 1 000 VNTR, puis analyse de tous ces marqueurs sur tous les individus qui ont le pedigree cible). La notion de locus se réduit à l'unité génétique variable par excellence, le nucléotide : le marqueur est le changement de 1 nucléotide, SNP). Avec 5 000 oligonucléotides sur une puce, en une hybridation on obtient le génotype.
- Transcriptome (ARNm) : connaître, en fonction des états physiologiques particuliers de la cellule, quels gènes sont induits ou silencieux et, si induits, à quel niveau ils s'expriment.
- Environnement : connaître les états bactériologiques du sol, remonter jusqu'aux virus ou microbes présents dans l'échantillon analysé.
- Détecter les maladies : aller de l'analyse monogénique vers une analyse multi-paramétrique de la dynamique transcriptionnelle. Ceci permettra d'analyser d'une façon différentielle les transcriptomes induits par les différentes phases de la maladie et/ou du traitement dans différents organes. Un autre exemple d'application : en quelques secondes, détecter dans le sang des patients la présence de dizaines de variétés de virus et de microbes.

DNA-chips, applications and prospects

moins consciente) de leur préoccupation qui était du domaine de la métrologie pour pouvoir se consacrer à nouveau davantage à la méthodologie.

Les techniques moléculaires à grand débit et l'intégration des puces à ADN dans la panoplie du chercheur ont changé le contexte actuel des projets de recherche au niveau des sciences médicales et de la biologie moléculaire humaine. Néanmoins, ces techniques restent onéreuses : pour être efficaces, l'application des SNP au niveau du génome ou celle du SAGE au niveau du transcriptome demandent un investissement en amont, au niveau de l'acquisition de données de séquence, et en aval, pour l'acquisition d'instruments de lecture. Les lois du marché et de la concurrence font que les prix chutent (encadré 3). En 1980, Frederick Sanger a reçu son deuxième prix Nobel en Chimie pour sa contribution au développement d'une méthode qui a permis de séquencer une centaine de bases par jour. Les performances des grands centres de séquençage actuels se mesurent en Mb par jour et un débit de l'ordre de la Mb

par heure n'est plus un défi technique et est annoncé par Craig Venter [29]. La démocratisation de ces méthodes va fondamentalement changer notre vision des problèmes étudiables (encadré 4).

À côté des *core-collections ex situ*, des techniques de conservation *in situ* et des techniques de préservation *in vitro*

des ressources génétiques, la puissance des nouvelles méthodes fondées sur les techniques à grand débit et la génomique implique la création et la conservation de nouvelles ressources, des ressources moléculaires : ces ressources moléculaires seront conservées *in vivo* (banques de mutants naturels et/ou artificiels), *in vitro* (banques de grands fragments génomiques, par ex. BAC et banques de ADNc) et *in silico* (bases de données de séquences, par ex. EST).

Ces méthodes à grand débit n'impliquent donc pas seulement une optimisation des méthodes moléculaires, mais aussi un changement de paradigme du domaine de recherche des ressources génétiques : du descriptif vers le prédictif et d'une approche analytique vers une approche systémique ■

Références

1. Mendel G. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereines Brünn* 1866 ; 4 : 3-47.
2. Gilles PN, Wu DJ, Foster CB, et al. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nature Biotechnology* 1999 ; 17 : 365-70.
3. <http://ncbi.nlm.nih.gov/SNP/index.html>
4. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998 ; 280 : 1077-82.
5. Jakubowski J, Kornfeld K. A local, high-density, single-nucleotide polymorphism map used to clone *Caenorhabditis elegans cdf-1*. *Genetics* 1999 ; 153 : 743-52.
6. Winzeler EA, Richards DR, Conway AR, et al. Direct allelic variation scanning of the yeast genome. *Science* 1998 ; 281 : 1194-7.

Résumé

À l'aube du troisième millénaire, on assiste à la fusion de différents domaines du savoir pour créer de nouvelles disciplines de recherche. Ainsi, biologie et informatique se sont mariées pour produire la bio-informatique du côté conceptuel et, entre autres, les bio-puces au niveau de l'instrumentation. On assiste à un changement d'échelle au niveau de la quantité de locus et du nombre d'échantillons à analyser, et donc de données à acquérir et à traiter. Plus qu'optimiser les techniques de biologie moléculaire (sélection assistée par marqueurs, par ex.), instruments de la génétique, les technologies à grand débit sur lesquelles est fondée la génomique ont dévoilé de nouvelles perspectives et entraînent un changement de paradigme, ouvrant l'ère de la post-génomique. D'analytique, l'approche scientifique évoluera vers une méthodologie systémique. Notre vision, essentiellement descriptive, de la diversité et des ressources génétiques promet de devenir réellement prédictive.

7. Thomas R. Laws for the dynamics of regulatory networks. *Int J Dev Biol* 1998 ; 42 : 479-85.
8. Chakravarti A. Population genetics-making sense out of the sequence. *Nature Genetics* 1998 ; 21 : 56-60.
9. Groupe de réflexion Science & Technologie de l'Académie des Sciences. *Développement et applications de la génomique : L'après génome*. Paris : Editions Tec & Doc, 1999 ; 231 p.
10. Drmanac R, Labat I, Brukner I, et al. Sequencing of megabase plus DNA by hybridization : theory of the method. *Genomics* 1989 ; 2 : 114-28.
11. Khrapko KR, Lysov YuP, Khorlyn AA, et al. An oligonucleotide hybridization approach to DNA sequencing. *FEBS Letters* 1989 ; 256 : 118-22.
12. Soper SA, Ford SM, Xu Y, et al. Nanoliter-scale sample preparation methods directly coupled to polymethylmethacrylate-based microchips and gel-filled capillaries for the analysis of oligonucleotides. *J Chromatogr A* 1999 ; 853 : 107-20.
13. Jordan BR. Large-scale expression measurement by hybridization methods : from high-density membranes to DNA chips. *J Biochem* 1998 ; 124 : 251-8.
14. Chen JJ, Wu R, Yang PC, et al. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* 1998 ; 51 : 313-24.
15. Bertucci F, Bernard K, Loriod B, et al. Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 1999 ; 8 : 1715-22.
16. Desprez T, Amselem J, Caboche M, et al. Differential gene expression in *Arabidopsis* monitored using cDNA arrays. *Plant J* 1998 ; 14 : 643-52.
17. Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995 ; 270 : 467-70.
18. Southern EM, Maskos U, Elder JK. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides : evaluation using experimental models. *Genomics* 1992 ; 4 : 1008-17.
19. Livache T. *Communication personnelle*.
20. Cargill JF, Lebl M. New methods in combinatorial chemistry-robotics and parallel synthesis. *Curr Opin Chem Biol* 1997 ; 1 : 67-71.
21. Soane DS, Soane ZM. Method and device for moving molecules by the application of a plurality of electrical fields. United States Patent 5.750.015 : 1998.
22. Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science* 1995 ; 270 : 484-7.
23. <http://ncbi.nlm.nih.gov/SAGE/index.html>
24. Powell J. Enhanced concatemer cloning-a modification to the SAGE (Serial Analysis of Gene Expression) technique. *Nucleic Acids Res* 1998 ; 14 : 3445-6.
25. Matsumura H, Nirasawa S, Terauchi R. Technical advance : transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression. *Plant J* 1999 ; 20 : 719-26.
26. Beavis RC, Chait BT. Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol* 1996 ; 270 : 519-51.
27. Laken SJ, Jackson PE, Kinzler K, et al. Genotyping by mass spectrometric analysis of short DNA fragments. *Nature Biotechnology* 1998 ; 16 : 1352-6.
28. Alcalá J, Giovannoni JJ, Pike LM, et al. Application of Genetic Bit Analysis (GBA™) for allelic selection in plant breeding. *Molecular Breeding* 1997 ; 3 : 495-502.
29. Marshall E, Pennisi E. Hubris and the Human Genome. *Science* 1998 ; 280 : 994-5.