

# 4

---

## LES GÉNOMES ET SES ENJEUX

Alain BUCHETON

Giuseppe Baldacci

Didier Boichard

Michel Boulétreau

Michel Caboche

Michael Chandler

Claude Chevalet

Michel Delseny

Bernard Dujon

Annie Jacq

Pierre Legrain

Arnold Munnich

Alain Nicolas

Véronique Paquis

William Saurin

Hélène Richard-Foy

Jean Weissenbach

### **1 – LA GÉNÉTIQUE : AU CŒUR DES RÉVOLUTIONS SUCCESSIVES DES SCIENCES DE LA VIE**

#### **1.1 LE GÉNOME, LA GÉNÉTIQUE ET LES SCIENCES DU VIVANT**

L'aventure des sciences de la vie depuis un siècle a été profondément marquée par la génétique. Sans doute plus qu'aucune autre discipline, la génétique a été à l'origine des grandes révolutions biologiques successives. Dans le public, elle est souvent devenue le symbole même du développement des sciences de la vie dans leur ensemble. Elle suscite, par les immenses perspectives qu'elle offre, espoir et inquiétude à la fois. Elle est en effet devenue aujourd'hui l'une des voies d'approche privilégiées du vivant. Le véritable bond en avant dans l'étude du développement permis par l'intrusion de cette science est sans doute l'une des illustrations les plus marquantes de la puissance de l'approche génétique. Celle-ci offre maintenant d'extraordinaires possibilités de comprendre les mécanismes intimes de la vie et de modifier le vivant de façon durable. Ses implications dans les connaissances fondamentales et ses

applications dans les domaines de la santé et de l'agronomie sont considérables, avec leurs retombées évidentes sur le plan économique. Cela suffit à expliquer les moyens considérables qui lui sont consacrés dans la plupart des grands pays.

La génétique est une science jeune, mais elle est probablement celle qui, plus qu'aucune autre, a marqué de façon déterminante le développement de la biologie. 2003 a été l'année du cinquantenaire de la découverte de la structure de l'ADN, l'une des acquisitions majeures de la science du xx<sup>e</sup> siècle. Cette découverte, permettant d'expliquer la pérennité et les variations des caractères héréditaires, leur transmission, et même d'entrevoir les mécanismes de leur expression, allait dans les années suivantes permettre l'accès à la connaissance précise des mécanismes intimes de la vie. L'ensemble des sciences du vivant allait en être bouleversé. Avec la découverte des enzymes de restriction et de la transcriptase inverse et la mise au point des techniques de séquençage de l'ADN, la biologie moléculaire devenait opératoire dans les années soixante-dix. Les premiers vecteurs porteurs de gènes étrangers datent de cette période, signalant l'avènement de l'ingénierie génétique. La plus grande partie de ce qui est enseigné aujourd'hui en génétique était inconnu il y a 20 ans. On mesure le chemin parcouru, etc. La génétique a apporté une méthode et des moyens dont aucune discipline biologique ne peut se passer. Les progrès de cette discipline liés à l'avènement de la bioinformatique et à l'essor de la technologie ont abouti à la génomique, qui bouleverse maintenant les manières d'appréhender l'étude du vivant sous tous ses aspects. La séquence du génome d'un nombre croissant d'organismes est disponible, offrant une source d'information inestimable. Les progrès de la génétique et de la génomique permettent désormais d'aborder les questions biologiques de façon globale. Les perspectives ainsi offertes ne concernent pas que le développement des connaissances. La génétique et la génomique offrent des ouvertures d'application toujours plus proches et surtout plus importantes. Les domaines de la santé, du médicament et de l'agro-alimentaire

sont révolutionnés par les connaissances sur le génome et les possibilités de le manipuler. Cela explique l'intérêt croissant des grands groupes industriels pour ces problèmes.

## 1.2 LA GÉNÉTIQUE EN FRANCE ET DANS LE MONDE

La France dispose d'atouts en génétique. Le travail pionnier d'un certain nombre de grands scientifiques a permis de former un nombre conséquent de chercheurs dans des domaines comme la génétique des levures, des champignons filamenteux, de la drosophile, ou encore en génétique humaine.

Malgré l'importance du génome et l'existence d'un potentiel humain important, il semble qu'aujourd'hui, dans notre pays, on ne s'intéresse plus à la génétique, et que la génomique ne soit considérée que comme un outil ou un moyen d'aborder certains problèmes (séquençage, étude des transcriptomes et protéomes, etc.). L'étude des génomes elle-même ne semble plus susciter grand intérêt. On parle même d'aire post-génomique ! Pourtant, le problème majeur de la vie, celui de la relation entre le génotype d'un individu et son phénotype, reste une question d'une brûlante actualité plus de cent ans après la naissance de la génétique. Il est encore illusoire, par exemple, de prédire l'ensemble des changements résultant de modifications des génomes. Un développement bien compris de la génomique et de la plupart des disciplines biologiques, ne peut se concevoir sans l'apport décisif de l'approche génétique. On ne peut que s'inquiéter très sérieusement de l'état du financement de la génétique en France, précisément au moment où la génomique est en plein essor.

D'énormes programmes sont en cours dans le monde pour déterminer les domaines et la période d'expression de l'ensemble des gènes ainsi que pour déterminer systématiquement les résultats de leur inactivation dans

des espèces de plus en plus nombreuses. La France, à l'exception notable des programmes sur les levures, Arabidopsis et la souris, est le plus souvent absente de ces efforts. Il convient toutefois de rappeler que ces programmes, pour indispensables qu'ils soient, ne permettront pas à eux seuls de déterminer avec précision la fonction des gènes. Celle-ci requiert l'étude de nombreux mutants et des interactions avec d'autres gènes. Mais il convient surtout de rappeler qu'un génome ne se résume pas à ses phases ouvertes de lecture et que, s'il est déjà difficile d'identifier les gènes à la suite du séquençage, il est infiniment plus difficile encore de préciser le rôle des séquences non codantes, et que les analyses systématiques, dans l'état actuel des connaissances, ne permettront pas de résoudre ce problème, ou seulement de façon très partielle. C'est pourtant là que se trouve la plus grande partie des signaux de régulation et des séquences impliquées dans l'architecture du chromosome. Des questions comme l'organisation fonctionnelle et la compartimentation du noyau, la stabilité et les variations du génome, la régulation de l'expression des gènes, le contrôle épigénétique, restent déterminantes et sont l'objet d'activités très intenses. Dans le cadre du soutien à la génomique, le GREG, puis les ACC-SV, ou le programme génome du CNRS, ont apporté un appui très significatif à la génétique qui a permis aux équipes françaises de tenir leur place dans la compétition internationale. Il faut bien constater que, en dehors du programme Génoplante (dont, malgré son succès, la survie n'est pas assurée au-delà de 2005), la France ne dispose plus de programme pour soutenir cette recherche. Les grandes difficultés auxquelles le développement de la génétique s'est heurté dans notre pays jusqu'aux années soixante se retrouvent finalement aujourd'hui.

Pourtant, les problèmes de l'organisation, l'expression et l'évolution des génomes font l'objet d'études intenses et nombreuses dans toutes les grandes nations scientifiques. Des progrès considérables ont été réalisés, des questions fondamentales sont en voie d'être résolues, et de nouveaux thèmes ont émergé. Ainsi, on entrevoit maintenant la possibilité de

déterminer les origines de réplication chez les mammifères ou les points chauds de recombinaison. Les mécanismes assurant la plasticité des génomes et leurs conséquences sur le plan évolutif sont de mieux en mieux appréhendés. La régulation de l'activité du génome est aussi beaucoup mieux comprise, et, surtout, ces dernières années ont été marquées par la reconnaissance de l'importance des régulations épigénétiques que l'on croyait marginales. Ainsi, on commence maintenant à comprendre les relations entre le fonctionnement du génome et les modifications de la chromatine comme la méthylation de l'ADN ou les modifications des histones, soulignant une nouvelle fois qu'un génome ne se résume pas à ses phases de lecture. À un autre niveau, on commence tout juste à mettre en évidence les relations entre la localisation nucléaire des gènes et leur activité transcriptionnelle. Enfin, l'une des découvertes majeures de la génétique a sans aucun doute été la mise en évidence du phénomène « d'ARN interférence » (RNAi), du fait de son importance comme mécanisme de régulation, mais aussi par la puissance de l'outil qu'il constitue dans d'innombrables études et par les perspectives qu'il offre en thérapie. La découverte de ce phénomène fondamental et universel, totalement inconnu il y a seulement quelques années, illustre à quel point la recherche fondamentale en génétique doit être encouragée, de même qu'il illustre la portée des retombées de cette recherche. Dans tous ces domaines, il existe en France de nombreuses équipes souvent performantes et internationalement reconnues. Elles sont dans beaucoup de cas à la limite de leur survie et, si rien n'est fait, notre pays pourrait bien rapidement perdre pied dans ce secteur décisif de la recherche.

### 1.3 L'EXPRESSION DES GÉNOMES

Un bon nombre de laboratoires français se concentrent sur la régulation génétique et épigénétique de l'expression des génomes. Le rôle de la structure chromatinienne, la méthylation de

l'ADN, les modifications des histones (acétylation, phosphorylation, méthylation, etc.) dans la régulation de l'expression des gènes constitue un domaine dans lequel nombre d'équipes tiennent largement leur place dans la compétition internationale.

Le rôle des ARN dans les régulations épigénétiques a été reconnu plus récemment. Les petits ARNs non codants apparaissent de plus en plus comme ayant un rôle universel ayant été conservé au cours de l'évolution, ce qui souligne leur importance. Ils ont des fonctions « classiques » dans la biogenèse des ribosomes ou le fonctionnement du spliceosome par exemple. Ils contrôlent aussi l'expression génique, d'une part via le système d'ARN interférence qui cible la stabilité des ARN messagers, mais aussi par des mécanismes ciblant la traduction. Ils sont également capables d'induire, au moins dans le monde végétal, des modifications covalentes de la structure chromatinienne, en particulier la méthylation de l'ADN. Ils sont impliqués dans des processus aussi variés que le développement embryonnaire, le contrôle de virus et des séquences parasites comme les éléments transposables, et le contrôle des réarrangements du génome dans certaines espèces. La France n'est pas en retard dans ce domaine, mais il est important qu'un grand effort soit fait pour maintenir cette position.

## 1.4 L'ÉVOLUTION DES GÉNOMES

L'étude des mécanismes moléculaires assurant la stabilité et la plasticité des génomes, d'une part dans les organismes modèles (bactéries, levures, nématode, drosophile, souris, Arabidopsis) et d'autre part dans les lignées cellulaires animales, connaît actuellement un essor important. Les résultats montrant que la transformation tumorale est associée à l'instabilité génomique, et la mise en place par le gouvernement en 2003 d'un plan de lutte contre le cancer, expliquent en partie le développement rapide de ces thèmes de recherche. Pour étudier en détail les mécanismes de la

stabilité et de la plasticité génomique l'utilisation des systèmes modèles est indispensable. La concentration exclusive du soutien financier sur les cellules animales ou sur les mammifères serait une grave erreur. L'étude de la réplication, recombinaison et réparation des génomes se poursuit dans un certain nombre de laboratoires et a permis la constitution en France d'une communauté qui se réunit régulièrement et interagit par des collaborations et des échanges d'informations et de matériels. Cependant, cet axe aurait besoin d'être renforcé pour éviter que la France soit distancée.

L'accumulation des données de séquençage a mis en relief la place massive des éléments transposables dans les génomes, dont ils constituent une source majeure de plasticité. Cela est particulièrement évident chez l'homme puisque près de la moitié de son génome résulte de transposition (notamment de rétrotransposition). Les éléments transposables jouent un rôle moteur dans la plasticité des génomes et la structuration de la chromatine. Du fait de leur aptitude à transposer, ils peuvent induire différents types de réarrangements chromosomiques (insertions, délétions, inversions, translocations) et contribuent au brassage génétique. Certains d'entre eux permettent même la rétrotransposition d'ARN messagers cellulaires. Des résultats récents obtenus chez plusieurs eucaryotes (drosophile, plantes) suggèrent également qu'il existe une corrélation entre la présence d'éléments transposables et d'autres séquences répétées dans une région, et sa structuration sous forme d'hétérochromatine. Chez certains organismes (ciliés, ascaris), ils pourraient être à l'origine de l'élimination de régions du génome somatique. D'autre part, les éléments transposables participent à la modulation de l'expression des génomes en apportant des séquences régulatrices (promoteurs, enhanceurs, etc.) ou suivant des mécanismes qui restent à découvrir. Il faut enfin noter que les éléments transposables constituent une voie d'entrée originale et privilégiée pour aborder l'étude des régulations épigénétiques.

La France possède la seconde communauté (après les États-Unis) étudiant la transposition sous ses différents aspects (étude

moléculaire des mécanismes de transposition et rétrotransposition, régulation, populations, évolution). Celle-ci est organisée, tient des réunions régulières, interagit fortement, et est en grande partie regroupée au sein d'un GDR. Le manque actuel de moyens des laboratoires impliqués compromet fortement l'avenir de cette communauté.

## **1.5 L'APPORT DE LA GÉNÉTIQUE AUX GRANDES QUESTIONS DE LA BIOLOGIE**

La génétique constitue une méthode d'approche essentielle pour aborder pratiquement toutes les grandes questions de la biologie. Il est ainsi indéniable que la génétique a été déterminante pour l'étude du développement. Les travaux pionniers réalisés dans ce domaine sur la drosophile en Europe et aux États-Unis ont véritablement révolutionné la biologie du développement. Les gènes clés ont été caractérisés. Cette approche est devenue rapidement celle qui a fait le succès des organismes modèles. C'est aussi ce travail qui a permis de prendre conscience de la conservation du vivant, non plus seulement au niveau des mécanismes moléculaires de base de la cellule, mais à un niveau supérieur d'organisation avec la conservation des gènes, de leur fonction, et de la mise en place des plans d'organisation des organismes. Cette conservation justifie la notion d'organisme modèle puisque la fonction d'un gène chez l'homme peut être recherchée, en première approximation, chez l'un de ces organismes facilement utilisable pour l'analyse génétique.

De la même manière, l'approche génétique est de plus en plus largement utilisée dans nombre de domaines, des neurosciences à la biologie cellulaire.

Enfin l'introduction de la génétique dans les sciences de l'évolution, de la biodiversité et de l'écologie leur a permis de dépasser le stade de la description ou des reconstitutions historiques, pour les enrichir désormais d'un

fort contenu théorique, proposer de scénarios argumentés, avancer des hypothèses sur les mécanismes, qui ont contribué à en faire de véritables sciences expérimentales.

Il existe en France une très importante et très active communauté travaillant sur la génétique du développement, en utilisant divers organismes modèles (drosophile, nématode, souris, Arabidopsis, etc.), ou abordant par la génétique des problèmes tels que le développement et le fonctionnement du système nerveux (incluant la mémoire et l'apprentissage). Ces communautés sont généralement aidées par les actions spécifiques destinées à soutenir par exemple la biologie du développement, les neurosciences ou la biologie cellulaire.

## **1.6 QUELQUES ÉLÉMENTS SUR L'ÉTAT DES LIEUX DE LA RECHERCHE EN GÉNÉTIQUE EN FRANCE**

Faire un état des lieux de la recherche en génétique est une tâche complexe qui implique :

- de définir le contour de cette recherche ;
- de pouvoir faire un état des lieux précis des forces en jeu ;
- de mesurer la production des groupes de recherche travaillant dans ce domaine ;
- de pouvoir comparer la situation de la France à celle d'autres pays européens (comme l'Allemagne ou la Grande Bretagne) et aux États-Unis.

### **La recherche en génétique**

On inclut sous cette terminologie de nombreux types de travaux dans les domaines suivants : génétique, génétique des populations, évolution, structure des génomes, expression des génomes et son contrôle (avec

en particulier le contrôle épigénétique de l'expression génique), modifications des acides nucléiques, réplication et réparation de l'ADN, transcription, traduction, mutagenèse, transgenèse, etc. La recherche en génétique s'adresse à tous les organismes vivants depuis le virus jusqu'à l'homme. Chez l'homme elle a pour champ d'étude de nombreuses pathologies : maladies génétiques, prédispositions à des maladies, cancer, etc. Les méthodologies mises en œuvre sont variées : biologie moléculaire, biochimie, biologie cellulaire, cytogénétique, génétique classique, bioinformatique-biostatistique, épidémiologie, évolution, biodiversité, etc.

## L'état des lieux

En France, ces travaux sont menés principalement au CNRS, à l'INSERM, à l'INRA, au CEA, à l'IRD, au CIRAD et dans les universités. Les chercheurs du CNRS et du CEA se consacrent à des études fondamentales dans l'ensemble des champs disciplinaires mentionnés ci-dessus. L'INSERM est plus orienté sur des études génétiques dans les pathologies humaines. L'INRA s'intéresse aux génomes des espèces animales et végétales, principalement celles d'intérêt agronomique.

Au CNRS, 161 formations affichent le mot clé génétique. À l'INSERM, 87 unités et plus d'une centaine de formations diverses affichent ce mot clé. Ces laboratoires peuvent consacrer la totalité de leurs travaux à la génétique, mais plus souvent avoir la génétique comme champ d'activité périphérique.

## La production scientifique des groupes

Il est possible de trouver dans les bases de données les publications par organisme de recherche et par mot clé. On constate que les laboratoires CNRS et INSERM publient beaucoup. Le pourcentage de publications de fort impact est sensiblement le même au CNRS et à l'INSERM, avec des spécificités : forte implication

de l'INSERM dans la génétique humaine et les aspects génétiques de pathologies telles que les pathologies vasculaires, le diabète et les maladies métaboliques, l'immunologie. Le CNRS est plus généraliste, avec une forte implication dans l'étude des ARN, l'analyse et l'évolution des génomes, la génétique classique. Les deux organismes contribuent à la rédaction de revues didactiques dans des périodiques internationaux de premier plan, mais le CNRS est beaucoup plus actif que l'INSERM dans ce domaine.

## La situation au niveau international

La participation aux projets européens est souvent prise en compte pour évaluer la productivité des groupes de recherche. C'est un mauvais indicateur dans le domaine de la recherche fondamentale, le pré requis étant le plus souvent « l'impact sociétal » qui souvent ne fait que refléter la participation d'industriels au projet. Il y a une forte participation de laboratoires français à de tels projets soit dans les disciplines agro-alimentaires, soit dans le domaine médical sur des pathologies particulières. Par contre les laboratoires très fondamentalistes sont moins représentés.

Une comparaison de la production scientifique du CNRS et de l'INSERM en termes de publications dans des domaines comparables fournit un meilleur indicateur. Celle-ci montre une bonne activité de publication des laboratoires CNRS et INSERM, aussi bien au niveau de la qualité que de la quantité.

## Recommandations

Depuis les années soixante-dix, des incitations et des soutiens ministériels et des organismes de recherche à la génétique et à la biologie moléculaire ont permis la constitution progressive d'une large communauté scientifique en France. Celle-ci tient généralement bien sa place dans une compétition internationale de plus en plus vive. Ces efforts sont en voie d'anéantissement du fait de l'absence totale d'actions dans ce domaine.

Cela est d'autant plus incompréhensible que des moyens gigantesques, tant publics que privés, sont mobilisés aux États-Unis pour déterminer l'organisation et comprendre le fonctionnement des génomes. Ces moyens sont consacrés à la fois à la génomique et à la génétique. Des moyens considérables sont aussi engagés dans des pays de taille comparable à la France (notamment le Royaume-Uni, l'Allemagne, le Japon). Les dernières subventions octroyées en France à la génétique l'ont été par le programme génome du CNRS, et encore ces moyens étaient-ils très limités. Ce type de recherche, pourtant décisif, ne bénéficie plus d'AUCUN SOUTIEN SPECIFIQUE, hormis les efforts en génomique végétale soutenus par le programme Génoplante. La plupart des équipes engagées dans cette voie ne disposent plus que des crédits récurrents de plus en plus maigres des organismes de recherche, et des quelques associations caritatives qui acceptent encore de financer ce genre de recherche. Si cette situation devait durer, il y a fort à craindre que ces équipes s'éteignent ou réorienteront leurs activités dans d'autres domaines susceptibles d'attirer plus facilement des financements. C'est tout un pan de la recherche française, sans doute l'un des plus centraux, qui est menacé de disparition.

Il est urgent que des initiatives soient prises pour poursuivre les efforts auparavant engagés dans ce domaine vital de recherche. Une initiative majeure devrait prendre la forme d'un grand programme sur l'étude des génomes, lancé par le ministère de la recherche et les organismes de recherche, notamment le CNRS, dans lequel trouverait place un soutien fort à la génétique et aux aspects les plus innovants de la génomique.

## **2 – LA GÉNOMIQUE : DISCIPLINE CENTRALE DE LA BIOLOGIE**

### **2.1 QU'EST-CE QUE LA GÉNOMIQUE ?**

Définir la génomique est un exercice délicat, peut-être même inutile. Une approche génomique d'une question biologique vise à prendre en compte simultanément la totalité de l'information dont une cellule (ou un organisme) peut disposer, à observer la manière dont elle utilise cette information et les conséquences de cette utilisation. La cellule est un calculateur chimique qui évalue à chaque instant son état interne et son environnement immédiat, et produit des réponses appropriées à cette situation en sollicitant le stock d'instructions de son génome. Au sens large, la génomique fait appel à un ensemble d'approches expérimentales qui s'appuient sur la connaissance d'une séquence génomique. En bonne logique une approche génomique devrait donc commencer par l'établissement de la séquence du génome ou, à tout le moins, d'une grande collection d'ADNc de l'organisme considéré. Cela n'est apparu qu'avec les premières séquences génomiques au milieu des années 90 et a bénéficié du développement de techniques expérimentales qui sont apparues à peu près à la même époque ou antérieurement (hybridation fluorescente différentielle sur des réseaux de sondes d'ADN, double hybride, électrophorèse bidimensionnelle de protéines, etc.).

Ces approches se caractérisent par une importante « parallélisation » et nécessitent des savoir-faire particuliers. Le développement de technologies à l'échelle génomique s'accompagne aussi de la nécessité de pouvoir acquérir et gérer ces quantités massives de données. Tant pour les équipements que pour le fonctionnement, de telles plates-formes ont des besoins de moyens financiers sans commune mesure avec ceux des laboratoires académiques. La nécessité de mutualiser et de centraliser les ressources au niveau de plates-formes

technologiques conduites par des personnels spécialisés s'est donc, au moins en partie, imposée. Sont donc considérées dans le champ de la génomique toutes les approches visant à obtenir une image globale des processus biologiques se déroulant au sein d'une cellule. Ce rapport n'a cependant pas l'ambition de faire une description même résumée de ces approches, mais de se limiter à passer en revue les efforts entrepris au niveau national et de les replacer dans le contexte international.

## 2.2 SÉQUENÇAGE

Au niveau international, le séquençage des génomes s'est développé, depuis le milieu des années 80, avec l'idée forte de décrypter le génome humain. Cette idée a façonné une discipline naissante qui, depuis, a largement dépassé ses enjeux initiaux et inclut maintenant les domaines de l'agronomie, de l'environnement, des maladies infectieuses, de la biodiversité ou, simplement, ceux de la connaissance fondamentale.

Après le court prélude cartographique de l'homme, où l'exemple de Généthon a démontré à l'échelle internationale l'intérêt de la concentration des moyens, cette concentration s'est étendue au domaine du séquençage. Celui-ci a ensuite connu une accélération unique dans l'histoire de la biologie en raison de la compétition consécutive à l'initiative de la compagnie Celera et à la réponse publique à ce défi pour le séquençage du génome humain. Une course aux équipements et à la robotisation s'est accompagnée d'une diminution spectaculaire des coûts. Cette diminution se poursuit avec l'introduction d'une deuxième génération de séquenceurs multicapillaires. Les coûts sont à présent de l'ordre de l'euro par lecture dans les grands centres. Cette diminution des coûts renforce encore l'intérêt de concentrer l'activité de séquençage pour pouvoir minimiser la fraction des coûts fixes indirects, qui varient en raison inverse du nombre d'opérations unitaires.

Une première série de grands projets portant sur l'homme, les espèces modèles principales (souris, drosophile, nématode, arabette, etc.) et les pathogènes majeurs a été réalisée par de grands centres qui se sont dotés au moment de la compétition avec Celera de très importantes capacités de séquençage. Or, ces grands centres sont pour l'essentiel aux États-Unis (Voir à la fin de ce chapitre les capacités mondiales de séquençage). Les centres américains cumulent entre eux une capacité qui engendre chaque année une masse de données équivalentes à trois projets de type génome humain à une couverture de 10 équivalents génome.

Du fait de l'importante capacité installée, de nombreux projets de séquençage de grands génomes sont en cours ou près d'être achevés. Ils portent principalement sur le séquençage de génomes de vertébrés ou cordés (*Fugu*, *Tetraodon* et cione sous forme d'ébauche ; chimpanzé, chien, rat, bovin, poulet, xénope, poisson zèbre, medaka, amphioxus, *Oikopleura*, en cours), de protostomiens (anophèle, ébauche ; abeille en cours ainsi que d'autres branches de protostomiens notamment les lophotrochozoaires) et de plantes (riz, *Medicago*, maïs). Il est à noter que très rares sont les projets qui ont abouti à une séquence où la couverture est complète. La finition d'une séquence représente aujourd'hui un coût sensiblement supérieur à celui de l'établissement d'une séquence ébauche et suppose la mise en place d'une équipe spécialisée qui se consacre au projet. D'où découle l'intérêt de collaborations efficaces entre centres de séquençage et communauté scientifique traditionnelle, qui devrait être la force du système français.

La coordination à l'échelle internationale s'établit autour de l'ISC (International Sequencing Consortium), qui est une sorte de continuation du consortium génome humain avec quelques autres partenaires, notamment TIGR. Sous la pression des bases de données et des gros groupes de bioinformatique, l'ISC a adopté une charte dans laquelle les centres s'engagent à déposer leurs données concernant les génomes de « community interest » non assemblées dans les « Trace Repositories ». Il y a cependant un désaccord sur la définition de « community interest ».



Malgré la limitation de ses moyens relativement au reste du monde, la France, outre sa contribution au génome humain, a contribué de manière significative aux projets de séquençage génomique dès les années 80 avec la levure et *Bacillus subtilis*. Ces deux projets pionniers étaient collaboratifs. Ils ont contribué à développer l'idée de coopération au sein de chaque communauté scientifique (parfois pourtant de taille réduite) pour aborder les projets génome (on l'a vu avec *Arabidopsis*, la paramécie, etc.). Les scientifiques français ont ainsi réussi à mettre à profit les moyens dont ils disposaient pour réaliser des projets bien visibles au niveau international. Dans le monde des microorganismes, on peut citer, outre la levure et *B. subtilis* déjà mentionnés, le bacille de Koch (*Mycobacterium tuberculosis*, même si la séquence elle-même était réalisée en Angleterre) et celui de la lèpre et de mycobactéries apparentées, plusieurs bactéries pathogènes, quelques archae et cyanobactéries, des bactéries lactiques, la première microsporidie (*Encephalitozoon cuniculi*), les *Listeria* (dans un projet partiellement européen), et plus récemment *Ostreococcus tauri* (une algue unicellulaire, l'un des plus petits génomes eucaryotes), la paramécie, *Photobacterium luminescens*, d'autres levures et champignons, etc. Le Génoscope a joué un rôle clé dans la majorité de ces projets. C'est de l'articulation efficace entre ce dernier et les membres de la communauté scientifique que dépend le succès des projets de séquençage en France.

Il est vraisemblable que les très grands projets, notamment ceux qui sont financés par le NIH, vont connaître un certain ralentissement dans les années à venir. Néanmoins, du fait de l'importante diminution des coûts, le séquençage prendra une autre priorité avec les génomes d'espèces d'intérêt économique, ou d'espèces clés pour l'étude de l'évolution, ou encore de nouvelles espèces devenant des systèmes modèles. En effet, l'état de nos connaissances sur la diversité du monde vivant et les écosystèmes est très rudimentaire. L'étude des génomes est la voie de choix pour ces problèmes. L'exploration par la séquence des branches majeures de l'arbre de l'évolution est

en cours. Le travail d'inventaire des gènes est assez avancé chez les eucaryotes multicellulaires (plantes et animaux) et les champignons, mais reste inexistant chez les micro-eucaryotes, et très incomplet chez les procaryotes. Avec l'inquiétude croissante sur les effets dévastateurs des pollutions humaines sur la biosphère, il est vraisemblable que les études de génomique seront de plus en plus importantes. Le monde marin devrait aussi offrir un vaste champ d'exploration au niveau des génomes. Les Français devraient pouvoir être bien placés dans ces deux derniers domaines. De même, le re-séquençage de gènes ou de régions génomiques chez de nombreux individus d'une même espèce est une activité qui devrait augmenter.

## 2.3 ANALYSE DES SÉQUENCES GÉNOMIQUES

L'établissement de la séquence d'un génome s'accompagne systématiquement de son analyse à l'aide d'outils informatiques. Ces analyses systématiques portent avant tout sur l'inventaire des gènes. Les inventaires sont faits par des méthodes informatiques plus ou moins automatisées et se poursuivent par une « curation » humaine. La qualité de ces inventaires automatisés est largement dépendante du niveau de l'intervention humaine réalisée en aval. Faute de moyens, celle-ci n'est souvent plus réalisée ou au plus une seule fois. À l'inverse, les méthodes automatisées peuvent être appliquées de manière récurrente et autorisent des mises à jour périodiques des annotations. La qualité des annotations des génomes séquencés proposées par les plateformes d'annotation internationales (Ensembl, NCBI, TIGR, MIPS, etc.) est très inégale. Pour les génomes d'organismes multicellulaires, l'inventaire reste très dépendant de l'existence de ressources annexes (collection de séquences d'ADNc, comparaison avec des génomes situés à des distances évolutives appropriées, d'où l'accroissement du besoin de séquençage). L'apport de la génomique comparative dans

les annotations prend une importance croissante avec l'accumulation des génomes séquencés. Ces comparaisons s'appuient souvent sur des outils comme tBLASTx et nécessitent d'importantes capacités de calcul. Elles s'orientent aussi vers l'identification de séquences conservées non codantes ayant une fonction biologique. L'identification informatique des régions régulatrices n'a pas atteint le degré de fiabilité de l'établissement de modèles de gènes. La génomique comparative entre espèces très proches est essentielle pour progresser dans ce domaine (comme cela a été montré par l'étude des levures), d'où le besoin de plus de séquences.

## 2.4 GÉNOTYPAGE

La première manière d'établir un lien entre un génotype et un phénotype repose sur une cartographie génétique du locus responsable du phénotype. Cette question prend une importance considérable en génétique humaine, où l'analyse génétique se déroule dans des conditions particulières. Mais quel que soit l'organisme considéré, la cartographie génétique s'appuie sur des méthodes de génotypage de variants de séquence permettant de localiser la région. La nature des variants utilisés a évolué au cours du temps. Le développement de grandes collections de marqueurs multialléliques (microsatellites) cartographiés a transformé la cartographie de phénotypes mendéliens monogéniques qui est à présent devenue une activité routinière. En revanche, la recherche des locus de prédisposition génétique aux maladies multifactorielles nécessite de recourir à un génotypage massif auquel les marqueurs multialléliques se prêtent mal. C'est pourquoi d'importantes collections de variants bialléliques de SNP (« single nucleotide polymorphism ») ont été initiées en parallèle du séquençage du génome humain et continuent à être développées.

Ce génotypage massif se réalise lui aussi sur des plates-formes de génotypage à grande échelle. Quelques grandes plates-formes de

génotypage ont vu le jour aux États-Unis, au Royaume-Uni, en Allemagne, au Japon. Cette dernière est sans doute la plus importante à ce jour avec 300 millions de génotypes par an. Contrairement au séquençage, il existe aujourd'hui de multiples technologies de génotypage de SNP. Les coûts sont très variables d'une méthodologie à l'autre. Elles sont souvent développées par des entreprises privées qui les commercialisent avec les réactifs associés. Les mêmes approches sont à considérer pour des espèces modèles ou d'intérêt économique. Les besoins de génotypage ne font que croître. Des développements récents permettent à présent d'identifier des procédés à coûts fortement réduits (de l'ordre du centime par génotype) qui apparaîtront dans un délai de un à deux ans. Il faudra être capable d'acheter et de maîtriser ces nouvelles technologies en temps opportun.

Chez l'homme, des analyses préliminaires de déséquilibre de liaison à l'échelle du génome suggèrent que l'essentiel des variations de séquences serait regroupé sous forme de blocs haplotypiques à l'intérieur desquels il n'y aurait que très peu de recombinaisons et qui correspondraient pour l'essentiel à des haplotypes ancestraux conservés qui se retrouvent dans les grands groupes ethniques. Ces observations réorientent le travail cartographique aussi bien chez l'homme que dans d'autres espèces. L'établissement d'une carte de ces haplotypes (Hap Map) est en cours à l'échelle mondiale. Une première version de 600000 marqueurs est prévue pour fin 2004. On peut regretter que la France ne soit pas partie prenante de ce projet, faute de financement approprié.

## 2.5 TRANSCRIPTOME

Les séquences de génome ont permis d'appliquer à grande échelle la technique d'hybridation différentielle fluorescente développée par Pat Brown ou d'autres méthodes d'analyse globale des transcriptomes avec des « arrays » de sondes prédéterminées. Les premiers génomes séquencés comme la levure

ont été les premiers à être étudiés à grande échelle (mise en évidence de réseaux de régulation surtout bien avancé chez la levure, cycle cellulaire, voies métaboliques, réponses aux différents stress, etc.). Les collections de mutants aident ces études globales, surtout quand, comme chez la levure, les étiquetages moléculaires des mutants permettent de trier les gènes mutés *a posteriori* à partir de mélanges complexes de nombreux mutants. Les applications humaines en matière de diagnostic, dans le domaine du cancer en particulier, ont aussi été un élément moteur dans la diffusion des techniques d'analyse du transcriptome sur des « arrays » de sondes. Des sociétés commerciales proposent des réactifs (puces pangénome) et des plates-formes d'analyse clé en main (Affymetrix, Agilent, etc.). Ces solutions commerciales apportent une standardisation qui a beaucoup fait défaut dans la phase de développement de cette technologie. La pratique de cette technologie peut être envisagée à l'échelle du laboratoire, notamment pour l'appliquer à des questions biologiques spécifiques. Mais les coûts très élevés pour les puces et les plates-formes d'analyse ainsi que l'indispensable standardisation plaident pour une concentration et une mutualisation des moyens et des choix de priorité délicats. De nombreux logiciels d'analyse de données ont aussi été développés. Le développement des puces pangénome est aussi largement dépendant de la qualité des annotations des génomes séquencés. L'intérêt diagnostic de ces outils notamment dans le domaine du cancer a suscité le développement de grands programmes nationaux, tels qu'en France, le programme systématique d'exploration des transcriptomes de tumeurs, Carte d'Identité des Tumeurs, financé par la Ligue Nationale Contre le Cancer. Mais les « arrays » génomiques touchent tous les domaines de recherche et tous les organismes. Par rapport aux principaux systèmes modèles pour lesquels existent des solutions commerciales, la construction d'arrays pour les nouveaux organismes séquencés qui intéressent une communauté scientifique spécifique peut rapidement devenir limitante au niveau technique comme financier.

## 2.6 PROTÉOMES, INTERACTOMES, ÉTUDES FONCTIONNELLES

Les techniques de protéomique sont très nombreuses. Elles cherchent d'une part à établir un inventaire des protéines dans un type cellulaire ou un compartiment cellulaire particulier et d'autre part à mettre en évidence des interactions entre protéines. Les finalités de ces techniques sont multiples, la plus importante sur le plan fondamental visant à comprendre le fonctionnement des nombreux complexes et machines à multiples composants protéiques. Elles se heurtent encore à de grandes difficultés techniques et, en plus de problèmes propres à chacune, n'échappent pas aux problèmes généraux déjà mentionnés ci-dessus (reproductibilité, standardisation, etc.). Leur pratique à grande échelle reste limitée à quelques plates-formes ayant acquis une expertise encore peu répandue à l'échelle internationale. Les informations produites par ces approches sont utiles, mais sont surtout des points de départ pour d'autres études plus ciblées qui devraient être menées au niveau des laboratoires et non plus des plates-formes.

Beaucoup de nouvelles techniques et stratégies récentes d'analyse fonctionnelle des génomes ont utilisé la levure comme plate-forme de validation à cause du nombre limité de ses gènes et des outils génétiques disponibles (dont la collection de mutants et les collections de gènes clonés sous différents promoteurs dans différents vecteurs). C'est avec la levure que furent validés la technologie de double marquage des protéines pour la purification de complexes natifs (Tap-TAG), les premières « chips » à protéines et la recherche de fixation aux ligands, le marquage systématique des protéines par des épitopes artificiels ou par la GFP, etc., le but étant évidemment de transférer rapidement ces techniques à des systèmes plus complexes, dont les cellules humaines. C'est là précisément que les compagnies privées ont pris immédiatement le relais, limitant ainsi malheureusement la diffusion des connaissances dans ces domaines.

Les collections de mutants par délétion de gènes ont été construites chez la levure et chez *B. subtilis* et se révèlent des ressources très utiles. Chez la levure, des cartes d'interactions génétiques basées sur des phénotypes synthétiques sont en cours d'établissement. Générales ou plus spécifiques, elles font intervenir des stratégies expérimentales très sophistiquées dans lesquelles l'expertise en génétique joue le rôle essentiel. Cependant de telles collections de mutants resteront limitées aux organismes pour lesquels le remplacement de gènes par recombinaison homologue est efficace (il faut aussi que la proportion de gènes essentiels à la vie soit faible mais cela est probablement le cas pour la majorité des organismes). Pour d'autres organismes comme *Arabidopsis thaliana* (mais aussi et encore la levure), l'inactivation de gènes par transposons ou mutagenèse aléatoire par insertion d'ADN offre la possibilité d'établir des collections quasi-complètes de mutants. Selon la difficulté de caractérisation et de maintenance de ces mutants, ces collections doivent être centralisées (*A. thaliana*) ou peuvent ne pas l'être (levure). Mais pour la plupart des organismes, l'inactivation fonctionnelle des gènes par interférence ARN (RNAi) semble la voie privilégiée. C'est le cas pour *Caenorhabditis elegans*, par exemple.

## 2.7 GÉNÉTIQUE À L'ÉCHELLE GÉNOMIQUE, ET « SYSTEMS BIOLOGY »

Outre les projets d'étude fonctionnelle déjà décrits et portant souvent sur des micro-organismes, de gros programmes de mutagenèse sur la souris sont en cours au Royaume-Uni et en Allemagne. Comme pour le « knock-out » systématique des gènes, le phénotypage des animaux est crucial. Mais l'inventaire des protéines à partir des séquences de génome permet de mesurer l'ampleur de la tâche à réaliser. Une initiative visant à connaître la fonction des 30000 gènes humains en pratiquant une série d'approches systématiques pour chacun des gènes a été

lancée aux États-Unis. Pour l'instant seuls les États-Unis et le Royaume-Uni (Wellcome Trust) semblent concernés. Y aura-t-il une division du travail comme pour le séquençage ?

L'inventaire des gènes pose d'emblée deux questions liées qui vont devenir les enjeux majeurs des décennies à venir. Peut-on aujourd'hui prédire le phénotype résultant de l'altération d'un seul élément génomique ? Plus largement, peut-on décrire l'ensemble des caractéristiques d'un système biologique en vue de sa modélisation ? Il faut, à de rares exceptions près, répondre modestement par la négative à la première question. Pour certains, cette prédiction passe par la résolution préalable de la deuxième question. Nous devons devenir capable de prendre en compte l'ensemble des éléments d'un système. Les premières tentatives de descriptions exhaustives de systèmes biologiques ont été amorcées. À nouveau, les systèmes modèles les plus complets, comme la levure, servent de tremplins d'essais. On ne peut pas encore prédire les développements de ce domaine en émergence. Il est probable que les hésitations, essais, erreurs et recentrages seront nombreux avant que le domaine émerge réellement. L'une des difficultés est la fiabilité parfois limitée des données expérimentales à grande échelle, les erreurs engendrées pouvant masquer des phénomènes pourtant réels mais de faible amplitude. En attendant que la capacité de prédiction dont nous rêvons avec la « biologie des systèmes » devienne une réalité, les approches réductionnistes traditionnelles conserveront leur pleine utilité.

Les questions fonctionnelles continueront donc largement à bénéficier des approches plus classiques, réductionnistes, pratiquées de manière séquentielle, gène par gène ou protéine par protéine, mais sans toutefois revenir à étudier un seul système étroit et souvent artificiellement isolé de l'ensemble par un réductionnisme trop poussé. Ainsi l'établissement de la structure tridimensionnelle de l'ensemble des protéines d'un génome, l'inactivation par « gene targeting » (KO), les marquages des protéines observées à l'échelle intracellulaire, s'orientent vers le grand nombre et non plus le particulier.

## 2.8 CONCLUSION

Ces dernières années, la génomique en France s'est, en grande partie, structurée autour du CNRG (Consortium National de Recherche en Génomique), doté d'un budget modeste comparé aux efforts américains, anglais et japonais, mais qui n'est pas insignifiant. Le CNRG comprend d'une part des plates-formes nationales pour le séquençage (Genoscope, Centre National de Séquençage), pour le génotypage (Centre National de Génotypage) et, dans l'avenir, pour la construction et le phénotypage de souris dont les gènes sont inactivés individuellement (Institut Clinique de la Souris, capacité actuelle de 300 gènes par an). Les centres nationaux poursuivent d'une part des projets propres et d'autre part des projets en collaboration avec des équipes académiques, soumis dans le cadre d'appels à propositions. Le CNRG coordonne, dans le cadre du Réseau National des Génopoles (RNG), les plates-formes technologiques de génomique concentrées dans les grands centres académiques nationaux. Il contribue au financement des équipements des Génopoles et au fonctionnement partiel de certains grands programmes de génomique. Les Génopoles françaises sont pour l'essentiel des plates-formes de proximité pour les travaux de génomique des équipes académiques locales. Il était souhaité que se dégagent des plates-formes d'envergure nationale à l'image de l'Institut Clinique de la Souris.

Comme on l'a vu à plusieurs reprises, c'est de la capacité à intégrer harmonieusement les activités entre centres nationaux, plates-formes techniques et laboratoires classiques, que dépendront les succès de la recherche française dans tous les domaines de la génomique.

## 2.9 CAPACITÉS DE SÉQUENÇAGE ET DE GÉNOTYPAGE

### Séquençage

– États-Unis : quatre grands centres publics (dont trois financés par le NIH et un par le DOE) totalisant environ 350 à 400 séquenceurs ABI 3730 représentant une capacité annuelle de l'ordre de 200 à 300 millions de lectures. À cela s'ajoute le « Joint Technology Center » de la fondation Craig Venter pour la Science (environ 200 machines plus une énorme capacité de calcul informatique) ;

– Royaume-Uni : séquençage très centralisé au Sanger Institute (80 séquenceurs ABI 3730 et une capacité annuelle de 40 à 60 millions de lecture) ;

– Chine : 80 machines sur deux centres ;

– Japon : situation très éparpillée ; plusieurs centres avec 20 à 30 séquenceurs multicapillaires ;

– Allemagne : très peu de séquençage à grande échelle ;

– France : Genoscope, 16 séquenceurs ABI 3730 (8 à 12 millions de lectures).

### Génotypage

À la différence du séquençage où l'on observe une technologie pratiquement monolithique, plusieurs technologies restent en compétition (ParAllele, Illumina, Invader/Third Wave, etc.). Ne sont mentionnés ici que les plus gros centres qui, pour la plupart, participent au projet HapMap en plus de leur implication dans des projets ciblés sur des pathologies.

– États-Unis : trois centres publics (Baylor, Broad, NIH-Baltimore) bien identifiés, implication de fonds privés ;

– Baylor et Broad Institute (anciennement Whitehead Genome Center) sont les contributeurs majeurs aux États-Unis au projet HapMap (participation américaine de 33 %) ;

– Illumina fournit sa technologie aux gros centres mais participe aussi directement à HapMap ;

– Baltimore est impliqué dans des projets sur des maladies et fait du génotypage pour des laboratoires publics ;

– Baylor utilise la technologie « ParAllele » et Broad et Baltimore « Illumina » ;

– capacité estimée à 400 millions de génotypes par an (100 millions de dollars) ;

– Japon : un gros centre, Riken SNP Research Center ; 300 millions de génotypes par an (65 millions de dollars). Responsable d'un quart de HapMap. A aussi un gros financement (55 millions de dollars) pour des collections de malades. Utilise les technologies « Invader » et « Illumina » ;

– Canada : Université McGill, 10 millions de génotypes par an. Participe au projet HapMap (10 %). Utilise Illumina ;

– Royaume-Uni : Sanger Institute. Responsable de d'un quart de HapMap. Utilise Illumina ;

– France : Centre National de Génotypage, 7 millions de génotypes par an.

### **3 – GÉNÉTIQUE HUMAINE**

Le projet Génome humain, la recherche biomédicale et les progrès réalisés dans le domaine de la génétique humaine dans les dix dernières années ont un impact important sur le diagnostic, la prévention et le traitement de maladies généralement incurables. La rapidité de développement des techniques recourant à la biologie moléculaire est source de grands espoirs mais soulève également de nombreux problèmes de par notre incapacité à en évaluer toutes les conséquences possibles dans le futur. Parmi toutes les questions découlant des progrès réalisés en génétique

humaine, celles de nature éthique, juridique et économique sont au premier plan.

#### **3.1 À LA RECHERCHE DES GÈNES RESPONSABLES DE MALADIES GÉNÉTIQUES**

Avec l'identification des premiers gènes responsables de maladies monogéniques au cours des années 80, la génétique humaine a connu un développement sans précédent. La démarche de clonage positionnel est devenue routinière et a grandement bénéficié des résultats du projet génome humain. On connaît à présent environ 1500 gènes impliqués dans une (ou quelquefois plusieurs) des quelque 5000 pathologies pouvant présenter un mode de transmission monogénique, recensés au catalogue de McKusick (<http://www.ncbi.nlm.gov/Omim/>). La plupart des gènes responsables restent donc à identifier, mais il s'agit de pathologies de plus en plus rares (ou de formes minoritaires de maladies à forte hétérogénéité de locus) pour lesquelles il devient difficile de rassembler des familles en nombre suffisant. La recherche d'effets fondateurs dans des populations isolées et de familles consanguines permet souvent de réduire le nombre de familles nécessaires aux analyses. Même si les pathologies concernées sont rares, l'identification de ces gènes est cruciale, non seulement pour les patients, mais aussi puisqu'elle révèle les protéines qui participent aux processus physiologiques, parfois d'une grande complexité. Ainsi, du fait de l'hétérogénéité de locus, l'analyse génétique distingue aujourd'hui une centaine de formes différentes de rétinites pigmentaires (ou de surdités non syndromiques) qui, en première analyse, présentent souvent une même symptomatologie. La déconstruction des mécanismes physiologiques hautement sophistiqués par l'identification des gènes/protéines impliqués renouvellera profondément la connaissance de ces processus

cellulaires, tissulaires ou systémiques. Il demeure donc crucial de poursuivre l'effort d'indentification de gènes de maladies monogéniques, notamment en collaborant avec des pays dont les populations sont caractérisées par des familles nombreuses, une forte consanguinité, permettant de révéler des effets fondateurs.

Alors que de nombreux gènes de maladies génétiques ont déjà été identifiés, seuls quelques très rares facteurs de prédisposition à des maladies communes ont été découverts. Le fait que ces recherches n'aient que faiblement progressé, malgré les efforts considérables déjà engagés, souligne les difficultés inhérentes à ces travaux. Trois difficultés majeures viennent compliquer l'analyse génétique :

- l'hétérogénéité de locus ;
- la pluralité des locus pouvant contribuer au phénotype chez un même individu, situation fréquente pour les caractères quantitatifs tels que la glycémie, la pression artérielle, etc. ;
- l'importance souvent considérable des facteurs non génétiques.

Ces difficultés rendent l'analyse familiale particulièrement ardue. D'autres approches sont donc mises en œuvre pour les maladies multifactorielles. Plusieurs stratégies ont été proposées et utilisées. Sur le plan technique, les espoirs continuent de reposer sur la disposition d'un nombre quasi illimité de marqueurs génétiques (SNP ou « single nucleotide polymorphism »), soit pour observer directement des associations statistiquement significatives entre SNP et un phénotype, soit pour définir des haplotypes renfermant les allèles de prédisposition.

### 3.2 LES TESTS GÉNÉTIQUES

Une des premières conséquences de « l'explosion » des connaissances et des techniques en génétique moléculaire a été le développement de tests génétiques à visée diagnostique. Ces tests remplacent, dans un

nombre croissant de pathologies, des examens invasifs et non spécifiques. Actuellement, ils sont réalisés dans différentes situations :

– **le diagnostic post-natal** : utilisé pour diagnostiquer une maladie chez un enfant ou un adulte atteint. De plus, dans un certain nombre de cas, ces tests permettent de prédire l'apparition future d'une pathologie chez un individu sain au moment de la réalisation de l'examen (diagnostic présymptomatique). L'intérêt de cette médecine prédictive est évident lorsqu'il existe des moyens de prévenir l'apparition de la maladie (certains cancers familiaux, par exemple). Par contre, en l'absence de moyen préventif, les problèmes éthiques et sociaux se retrouvent au premier plan. Un autre problème lié au diagnostic présymptomatique concerne l'utilisation de la notion de prédisposition génétique dans un but discriminatoire. La réglementation concernant la confidentialité des données est primordiale si l'on veut éviter leur utilisation par un employeur ou par une compagnie d'assurance ;

– **le diagnostic prénatal (DPN)** : réalisé sur l'embryon ou le fœtus. Il consiste à prévenir la naissance d'un enfant malade. Là encore, les questions éthiques sont nombreuses : quel niveau de gravité nécessite le recours au DPN qui se soldera par une interruption de grossesse en cas d'enfant atteint ? Comment garantir que ce type d'examen ne sera pas réalisé dans le but de déterminer le sexe de l'enfant à naître pour des motifs autres que médicaux ou pour identifier des caractéristiques sans retentissement sur sa santé ?

Si le développement de tests génétiques précoces sur cellules embryonnaires circulantes dans le sang maternel constitue un espoir important pour le diagnostic prénatal dans les années à venir, le diagnostic préimplantatoire (DPI) représente aujourd'hui une alternative au DPN. Il permet de déterminer le statut sain ou malade d'un embryon, au stade 8 cellules, obtenu par fécondation in vitro. Le transfert d'embryons sains sélectionnés évite ainsi les interruptions de grossesse, qui font éventuellement suite à un DPN classique. Là encore, le développement de cette nouvelle technique

pose d'emblée un problème éthique lié au possible « détournement » de son indication initiale. Est-il souhaitable d'utiliser le DPI pour concevoir un enfant indemne d'une maladie génétique, mais qui pourra également être sélectionné sur ses capacités de futur donneur de cellules ou d'organes pour son frère ou sa sœur malade ? De même, que répondre à des parents ayant un enfant atteint d'une maladie non génétique et qui désirent concevoir un autre enfant HLA compatible, également dans un but thérapeutique ?

La réalisation de tests génétiques par un nombre grandissant de laboratoires publics et privés, notamment en Europe, pose un certain nombre de problèmes en santé publique :

– **La fiabilité des résultats** : Elle diffère d'un laboratoire à l'autre et s'explique notamment par la difficulté à mettre en place un contrôle qualité, la complexité de certaines pathologies, les limites des techniques utilisées et la frontière entre recherche et diagnostic souvent difficile à établir.

– **L'accessibilité à tous** : Si pour les maladies les plus courantes, chaque pays dispose au moins d'un laboratoire capable de réaliser le test adéquat, dans le cas de pathologies particulièrement rares, le seul laboratoire disponible est souvent situé dans un pays étranger.

Se pose alors, pour la famille, le problème de la prise en charge du coût, parfois élevé, de l'examen et pour le clinicien, celui de l'évaluation de la qualité du résultat.

La solution passe par l'organisation d'un réseau européen de laboratoires de référence. Cette organisation en réseaux, dont les bases ont déjà été posées sur le territoire français, est indispensable si l'on veut pouvoir, dans le futur, disposer de tests génétiques pour les 5 000 maladies rares<sup>1</sup> recensées à ce jour.

### 3.3 MALADIES GÉNÉTIQUES ET POSSIBILITÉS THÉRAPEUTIQUES

À l'heure actuelle, les espoirs thérapeutiques reposent sur plusieurs types d'approches qui doivent être développées parallèlement, telles que la production de protéines recombinantes, les transplantations d'organes, la thérapie génique ou la pharmacologie conventionnelle. Si la thérapie génique a représenté l'approche la plus « médiatisée », force est de constater, à l'heure actuelle, que les résultats ne sont pas à la hauteur des espoirs suscités.

#### Les protéines produites par génie génétique

Elles ont fait la preuve de leur efficacité et de leur innocuité dans plusieurs pathologies (les facteurs VIII et IX dans l'hémophilie, l'insuline dans le traitement du diabète ou l'hormone de croissance dans les retards staturaux). Leur production repose sur la construction d'animaux transgéniques qui produisent la protéine humaine d'intérêt, par exemple dans le lait. Le clonage de tels animaux s'avère intéressant pour pallier aux difficultés rencontrées pour obtenir et maintenir des lignées transgéniques capables de produire ce type de protéines de manière reproductible.

#### Les transplantations d'organes

Ultime solution thérapeutique pour un grand nombre de pathologies (mucoviscidose et greffe pulmonaire, cardiomyopathie et greffe cardiaque, déficits immunitaires et greffe de moelle osseuse, etc.), elle est soumise à deux limitations majeures : un nombre de donneurs insuffisant et le risque de rejet qui impose l'utilisation d'immunosuppresseurs tout au long

---

1. Une maladie est dite rare lorsqu'elle touche moins d'une personne sur deux mille dans la population générale. Environ trois millions de personnes en France sont concernées par ces affections.



de la vie du patient. L'utilisation de cellules souches est potentiellement extrêmement intéressante pour remplacer des cellules détruites ou lésées car elle permet notamment de pallier à ces limitations. On pourrait ainsi disposer à volonté de cellules ou de tissus différenciés, susceptibles d'être réimplantés chez un malade atteint d'un infarctus du myocarde ou d'une chorée de Huntington. Des cellules souches sont présentes chez un individu tout au long de sa vie, depuis la période embryonnaire jusqu'à l'âge adulte. Mais c'est pendant la période embryonnaire qu'elles sont les plus nombreuses et donc les plus faciles à isoler. L'utilisation potentielle de cellules souches embryonnaires obtenues à partir d'embryons humains surnuméraires (produits par fécondation *in vitro* et pour lesquels il n'y a plus de projet parental) suscite de vives controverses d'ordre éthique. L'obtention de cellules souches embryonnaires par clonage thérapeutique (issues d'embryons créés par transfert de noyau) permettrait de régler le problème d'immunocompatibilité entre donneur et receveur. En effet, le noyau transféré proviendrait d'une cellule somatique du receveur et les cellules implantées seraient alors compatibles pour le système HLA. Néanmoins, l'utilisation d'une telle technique est également source de vives polémiques, notamment en raison d'une dérive possible vers le clonage reproductif humain, unanimement condamné. Pour toutes ces raisons, les cellules souches adultes, plus difficiles à isoler que les cellules souches embryonnaires et dont le spectre de différenciation est plus limité, représenteront peut-être la solution dans les années à venir. La possibilité pour les chercheurs d'utiliser ces cellules souches embryonnaires varie d'un pays européen à l'autre. En France, dans le cadre de la révision des lois de bioéthique, le projet de loi adopté en janvier 2003 par le Sénat interdit le clonage thérapeutique et reproductif et autorise la recherche sur l'embryon humain de manière dérogatoire et provisoire. Néanmoins, l'immense espoir suscité par l'utilisation de ces cellules souches, quelle que soit leur origine, justifie l'importance des moyens humains et financiers qui seront investis dans les années à venir.

## La thérapie génique

Là encore, il faut distinguer la thérapie génique germinale, non applicable à l'homme, qui consiste en une modification du patrimoine génétique transmissible à la descendance, et la thérapie génique somatique, qui vise à corriger le fonctionnement anormal d'un gène dans une cellule somatique. Dans ce dernier cas, envisagé pour le traitement de maladies génétiques ou de cancers par exemple, la modification génétique ne sera pas transmise à la descendance. Il est clair que dans ce domaine, il existe aujourd'hui un fossé entre les espoirs suscités et les résultats obtenus. Plusieurs facteurs sont à l'origine de cette déconvenue : de nombreux problèmes techniques non résolus (choix du vecteur, ciblage des cellules d'intérêt, etc.), des indications plus limitées que celles annoncées au départ et surtout des risques réels de toxicité ou de mutagenèse insertionnelle. Il est clair que si le constat actuel ne doit pas faire abandonner l'espoir de guérir certaines maladies grâce à la thérapie génique, il convient d'être plus prudent à la fois sur la nature et le nombre de maladies potentiellement curables, mais aussi sur les délais nécessaires.

## La pharmacogénétique

Dans le cas d'un nombre croissant de maladies, l'identification du gène responsable et de la protéine défectueuse a permis de décortiquer les mécanismes à l'origine du déficit. Ces études fonctionnelles permettent d'envisager de nouvelles approches pharmacologiques adaptées à chaque affection, voire à chaque patient. On peut maintenant envisager d'utiliser des molécules qui vont rétablir un épissage correct, modifier les conséquences d'une mutation non sens sur la traduction, activer ou verrouiller une voie métabolique, etc. Les enjeux de la pharmacogénétique, basée sur la définition de profils génétiques (SNP) qui devraient permettre de prévoir la réaction d'un malade à un médicament et d'adapter un traitement « à la carte », sont également énormes sur le plan médical et économique.

### 3.4 CONCLUSION

Les questions soulevées par les progrès de la génétique médicale sont à la hauteur des résultats obtenus, même si ces derniers sont encore nettement insuffisants sur le plan thérapeutique. Si les questions éthiques sont aujourd'hui au premier plan, la mise en place d'un cadre législatif clair et réfléchi devrait aider à y répondre. De plus, un large consensus existe sur l'objectif principal de la recherche médicale qui vise à soigner le plus efficacement possible et à guérir des maladies humaines. Ce consensus devrait également aider à aplanir les divergences liées à l'émergence de ces nouvelles technologies. Dans le domaine biomédical, il convient de préserver une recherche fondamentale de qualité si l'on veut garantir celle de la recherche appliquée dans les années futures. Dans ce cadre, le renforcement des interactions entre chercheurs et cliniciens semble fondamental et le rôle joué par les Sciences de la Vie du CNRS et l'INSERM devrait être de tout premier plan.

## 4 – LA BIOINFORMATIQUE ET LES GÉNOMES

### 4.1 INTRODUCTION

La bioinformatique a l'ambition de développer et d'utiliser l'ensemble des techniques, outils, concepts et théories permettant d'analyser de manière systématique et automatique l'information des systèmes biologiques afin d'en prédire les comportements. À ce titre, si elle construit les moyens informatiques de traitement des données génétiques et génomiques, elle développe aussi ceux qui permettent d'analyser des données de structure des macromolécules ou de modéliser les systèmes biologiques plus complexes comme

les réseaux métaboliques, les réseaux de signalisation cellulaire, les mécanismes de régulation physiologique.

Outre-atlantique, on distingue volontiers « bioinformatics » et « computational biology ». Le premier champ recouvrant l'activité de recherche, de développement et d'application d'outils en vue d'étendre l'utilisation de données biologiques ; ce champ de recherche inclut l'acquisition, le stockage, l'organisation, l'archivage, l'analyse et la visualisation de ces données. Le second champ consiste dans le développement et l'application de théories, de méthodes d'analyse de données, de modélisations mathématiques et de techniques de simulation informatique en vue d'étudier des systèmes biologiques. La bioinformatique telle que nous l'entendons généralement en France recouvre ces deux champs de recherche.

Les produits de cette discipline sont des méthodes et des outils d'analyse de l'information, des bases de données permettant de structurer et de retrouver de la connaissance biologique, et des modèles décrivant des systèmes biologiques. Historiquement, le champ d'application majeur de la bioinformatique a été la biologie moléculaire et bien que l'on puisse faire remonter les premiers travaux qui relèvent de bioinformatique au milieu des années 60, les numéros spéciaux de la revue *Nucleic Acids Research* de 1982 et 1983 marquent avec éclat le développement mondial de ce champ. Le terme de bioinformatique est relativement récent : dans les années 80 les instances de la recherche française lui ont souvent préféré celui de « interface informatique-biologie ». Une telle dénomination a l'avantage de l'exactitude au moment où émerge une nouvelle discipline. Elle présente pourtant un inconvénient : l'épaisseur d'une interface est si faible qu'on n'y trouve probablement pas de place pour quiconque, et que les financements qui devraient lui être accordés sont généralement octroyés en fait à l'un ou l'autre des domaines que cette interface met en contact. Il est significatif que seuls sept articles émanent de laboratoires français parmi les 401 publications que citent Ouzounis et Valencia (*Bioinformatics*, 2003, 19 :2176-2190)

dans leur revue sur les premiers temps de la bioinformatique (des origines au début des années 90).

Les rapports de la bioinformatique d'une part et de la biologie moléculaire, de la génétique et de la génomique d'autre part sont importants : sans les méthodes et les programmes de la bioinformatique, on n'aurait pas su reconstruire les séquences des génomes que les grands programmes de séquençage ont produites ; sans les bases de données développées par la bioinformatique, on ne saurait organiser le sens que l'on sait extraire de ces données ; enfin, sans les services que cette discipline a su développer au niveau mondial sur le web, on n'aurait pas pu organiser les grandes collaborations internationales qu'ont été les grands programmes de séquençage et que seront de plus en plus les programmes d'analyse fonctionnelle de génomes.

## 4.1 DÉVELOPPEMENTS ALGORITHMIQUES

La prise en compte de la spécificité des données biologiques donne lieu à une importante activité de développement algorithmique. L'augmentation des flux de données, qui se produit depuis la décennie précédente, exige que ces algorithmes soient de plus en plus rapides. L'apparition de données nouvelles (données de transcriptomes ou de protéomes) motive des développements algorithmiques nouveaux. Enfin, les publications scientifiques elles-mêmes, par la croissance de leur volume, deviennent sujet d'analyse systématique et nécessitent des développements nouveaux.

Une grande partie de l'algorithmique des mots et du texte doit son développement aux données de séquences. Ainsi, l'opération qui permet de reconstruire les longues séquences des génomes complets à partir des fragments que l'on peut obtenir expérimentalement pose un problème compliqué : il s'agit d'identifier toutes les ressemblances que ces fragments

présentent entre eux de manière à déterminer lesquels recouvrent les mêmes régions d'un génome et ainsi les assembler. L'assemblage des séquences génomique repose sur l'application systématique de l'algorithme de Smith et Waterman (J. Mol. Biol., 1981, 147 : 195-197) ou de versions optimisées de celui-ci. Le programme Phrap, qui a été systématiquement employé pendant le projet génome humain, s'appuie sur une version optimisée de cet algorithme. De la même façon ce sont des développements algorithmiques nouveaux qui permettent de prendre en compte les données de séquençage de grands génomes par « shotgun » intégral.

Le séquençage de l'ADN fournit la structure chimique du génome. Cela ne suffit évidemment pas : pour faire sens de cette information il faut identifier les gènes. Cela a motivé le développement de nouvelles méthodes dont les premières versions ont été empruntées à la linguistique formelle. Les programmes comme Glimmer, Genmark, Genscan, correspondent à de telles techniques, fondés sur des modèles de Markov. Ces méthodes, qui incorporent des connaissances biologiques, sont de théories actives : elles conduisent à des prédictions que l'on peut ensuite soumettre à l'expérience.

La recherche de similarité entre les séquences nouvellement déterminées et celles pour lesquelles on dispose déjà d'une information fonctionnelle est une tâche lourde. La suite de logiciels Blast développée à cette fin repose également sur la mise en œuvre d'algorithmes reposant sur des heuristiques rapides. Les programmes de génomique récents ont motivé le développement de nouvelles méthodes comme le logiciel Blat développé à l'Université de Santa Cruz aux États-Unis.

Il est important de bien comprendre que le développement de méthodes bioinformatiques ne s'est pas fait uniquement en empruntant des algorithmes développés par ailleurs, au contraire cette discipline est une source de problèmes qui nécessitent des développements algorithmiques nouveaux.

## 4.2 DÉVELOPPEMENT DE BASE DE DONNÉES

Le développement de l'acquisition systématique d'informations biologiques a motivé le développement de bases de données pour conserver ces informations. Genbank est la base de données qui, avec la banque de donnée de l'EMBL et sa contrepartie japonaise DDBJ, centralise les informations sur les séquences nucléotidiques. En 1981, les 270 séquences de Genbank à Los Alamos contenaient 370 000 nucléotides. Genbank représente maintenant une quantité d'information 200 000 fois plus importante. Swiss-Prot, créée en 1986, contient à présent plus de 100 000 protéines annotées par des experts. On recense à présent 386 bases de données (Baxevanis, Nucl. Acids Res, 2003, 31 :1-12) dont les contenus sont aussi variés que l'ensemble des données de séquences nucléotidiques connues (par exemple Genbank ou la banque de nucléotides de l'EMBL), des données de génomique comparative (par exemple COG <http://www.ncbi.nlm.nih.gov/COG>), des données d'expression génétique (par exemple ArrayExpress <http://www.ebi.ac.uk/arrayexpress> ou BodyMap <http://bodymap.ims.u-tokyo.ac.jp>), des voies métaboliques (par exemple Kyoto Encyclopedia of Genes and Genomes <http://www.genome.ad.jp/kegg>), ou encore des catalogues de gènes associés à des phénotypes ou à des maladies (par exemple Online Mendelian Inheritance in Man <http://www.ncbi.nlm.nih.gov/omim/>).

La croissance en nombre de ces bases de données pose deux questions majeures : celle de leur entretien et celle de leur intégration.

L'entretien des bases de données est souvent sous-estimé : les instances scientifiques ont souvent tendance à considérer que le travail de recherche s'arrête au moment où la structure d'une telle base a été conçue et après qu'elle a été alimentée une première fois, on connaît alors les différents éléments qu'on veut y intégrer et on en a une première version. Une telle attitude a débouché sur ce qui a été

connu comme la « Swiss-prot funding crisis » de 1996. Swiss-Prot est une base de données de séquences protéiques annotées par des experts. Elle est reconnue par l'ensemble de la communauté des chercheurs de la biologie moléculaire comme une ressource unique, mais, en 1996, le renouvellement d'un financement européen ne put avoir lieu : le maintien de la base de données (et extension) n'avait alors pas été considéré comme un développement nouveau. Ceci conduisit à rechercher un mode de financement original, et une compagnie privée (Genebio) fut constituée pour la valoriser auprès des industriels. Ceci conduisit de fait à ce que l'utilisation de la base de donnée ne soit plus totalement libre, bien qu'elle le restât pour la communauté académique. Cette crise est maintenant résolue : le NIH a en effet annoncé qu'il accorde un financement sur trois ans de 15 millions de dollars pour réaliser l'intégration des trois banques majeures collectant des données de protéines, Swiss-prot, trEMBL et la banque PIR. La banque de données Uniprot sera le produit de cette intégration. Plus grave est le cas de YPD, la banque de données des protéines de levure : rachetée par la société Incyte, et elle n'est maintenant plus librement accessible.

L'intégration des banques de données est le second problème que pose l'accroissement du nombre de ces banques : en effet les informations disponibles sur un objet ou un processus biologiques sont dispersées dans plusieurs banques de données différentes. Deux approches sont maintenant adoptées : la constitution de banques de données spécialisées et le développement de systèmes d'intégration des bases de données. Les banques de données spécialisées s'intéressent à un organisme, un type d'objet biologique ou un processus particulier. Le travail consiste à ré-extraire les données contenues dans des banques de données générales et à y ajouter de nouvelles données expérimentales. Il existe un très grand nombre de telles bases de données spécialisées. Elles présentent l'intérêt de contenir une information validée par les experts d'un domaine biologique, mais il n'est pas rare qu'elles présentent un schéma très spécifique et des interfaces qui

les rendent peu inter-opérantes. La seconde approche consiste à développer des systèmes d'intégration de banques de données. On peut citer par exemple SRS (*Etzold et al.*, *Methods Enzymol.*, 1996, 266 :114-28), Kleisli (*Davidson et al.*, *International Journal of Data Libraries*, 1997, 1 :36-53), ou Discovery link développé par la société IBM. Ces systèmes reposent sur des techniques de médiation qui sont encore du domaine de la recherche en informatique.

Là encore, la bioinformatique affronte des problèmes originaux qui sont dus à la grande richesse des données de biologie et à l'évolution relativement rapide des concepts qu'il s'agit de représenter.

### 4.3 LE DÉVELOPPEMENT DES SERVICES WEB

De plus en plus, ces bases de données sont interrogeables sur le web. À travers l'interface de son navigateur, on peut lancer une requête. Les services web permettent aussi de mettre une méthode bioinformatique à disposition d'utilisateurs. Les utilisateurs de ces services qui peuvent être très éloignés géographiquement et les auteurs de méthodes ou de bases de données peuvent éviter de se soucier de porter une application informatique sur un système différent du leur. Ce point est d'importance car il augmente les possibilités de collaborations à distance. Il présente néanmoins des inconvénients : les sources de programmes ne sont plus relues par d'autres que leurs auteurs et il est donc moins facile de savoir quelles opérations sont effectivement réalisées par ces programmes ; puisque les ressources utilisées ne sont pas celles des utilisateurs, elles sont parfois mises à disposition trop parcimonieusement ; enfin lorsqu'il s'agit de bases de données, leur accès ne peut être réalisé qu'avec les méthodes prévues par les auteurs, ce qui interdit de lancer des requêtes originales.

### 4.4 POSITION DE LA RECHERCHE BIOINFORMATIQUE FRANÇAISE

Les revues *Bioinformatics* et *Journal of Computational Biology* (*J. Comp. Bol.*) sont les supports de publication majeurs en bioinformatique. On peut se donner une vision instantanée de la position de la recherche nationale en bioinformatique en examinant quelle est la part de publications provenant de laboratoires français dans ces revues. 17 numéros de *Bioinformatics* publiés en 2003 contiennent 226 articles originaux et 153 notes applicatives. Les notes applicatives sont de très courts articles décrivant un outil bioinformatique original (programme, base de données ou service web). 25 articles ont été publiés dans *J. Comp. Bol.* Sur l'ensemble des 251 (226 + 25) articles, 13 ont une origine française, 121 proviennent d'institutions américaines, 35 du Royaume-Uni, 19 d'Allemagne et 13 de Chine. En ce qui concerne les notes applicatives, 64 proviennent des USA, 16 du Royaume-Uni, 13 d'Allemagne, 3 ont une origine française.

Par ailleurs, la revue *Nucleic Acids Research* a publié deux numéros, l'un entièrement consacré aux bases de données de biologie moléculaire (*Nucleic Acids Research* 2003, 31 n°1), l'autre aux serveurs web pour la biologie (*Nucleic Acids Research* 2003, 31 n°13). Dans le premier, 130 articles décrivent des bases de données, dans le second 121 décrivent des serveurs web. En ce qui concerne les bases de données, 63 articles proviennent des USA, 20 du Royaume-Uni, 19 d'Allemagne et 13 de France. En ce qui concerne les serveurs web, 64 articles sont originaires des USA, 17 d'Allemagne, 14 de France et 13 du Royaume-Uni.

La production américaine est de l'ordre du double de la production européenne. Au sein de celle-ci, la production française arrive en troisième position derrière le Royaume-Uni et l'Allemagne, sauf en ce qui concerne les serveurs Web.

## 4.5 CONCLUSION

La bioinformatique française s'est activement développée dans le courant des années 90, avec un léger retard par rapport aux USA ou au Royaume-Uni. Il importe que cet effort soit maintenu : la biologie produira de plus en plus de données, de types de plus en plus variés et sur un mode de plus en plus systématique et rapide. L'analyse de cette information sera une des clés des prochaines découvertes, et si les techniques de biologie moléculaire ont marqué la seconde moitié du siècle qui vient de s'écouler, il est vraisemblable que celles de la bioinformatique marqueront ce siècle en biologie.

Pour qu'une discipline se développe, il est important de promouvoir les formations supérieures qui l'enseignent. Au niveau national, un effort sensible a été réalisé pour former des ingénieurs bio-informaticiens (niveau DESS). Un effort pour atteindre le niveau du Doctorat reste nécessaire.

Le financement continu de la maintenance de bases de données est une nécessité à laquelle nous avons échappé partiellement en nous concentrant sur des bases de données spécialisées et en évitant de développer une base de données majeure : il en résulte que les bases de données sur les séquences nucléotidiques sont pour l'essentiel produites aux USA, et celles sur les protéines le sont en Suisse et au Royaume-Uni. Notons à nouveau sur ce sujet que les USA, parmi bien d'autres projets de bioinformatique, consacrent 15 millions de dollars au financement de l'intégration des trois bases de données de protéines pour les trois prochaines années. Ceci est à comparer aux 2 millions d'euros qui sont consacrés en France au financement des programmes de recherche bioinformatique à trois ans par l'action concertée incitative IMPBIO. Cette action est la seule qui a financé la recherche bioinformatique française en 2003.

L'avance du Royaume-Uni est essentiellement due à trois causes : un développement

précoce de la bioinformatique dans ce pays, le fait que le centre européen de bioinformatique (EBI) soit localisé près de Cambridge à proximité d'un centre de génomique de haut niveau, le Sanger Centre qui a contribué pour plus de 25 % au projet de séquençage du génome humain, et l'implication financière très conséquente de Wellcome Trust.

## 5 GÉNÉTIQUE, GÉNOMIQUE, ÉVOLUTION ET ENVIRONNEMENT

La totale continuité entre génétique, écologie et évolution est un fait récent. Il paraissait impensable, il y a quelques années encore, que puisse être comblé le fossé historique qui les séparait. L'éclairage apporté par les approches génétiques joue un rôle central non seulement dans les études sur l'évolution, dont la cible évidente est le génome lui-même, mais aussi dans la plupart des études sur la biodiversité, les relations des organismes à leur environnement, leurs réponses et adaptations aux stress, la dynamique et la stabilité des populations, les interactions parasites ou symbiotiques, l'organisation des communautés plurispécifiques.

Dès les années 80, les axes génétique des populations, biologie et écologie évolutives, ont émergé en France sous l'impulsion d'un petit nombre de centres, principalement à Paris, Lyon, Montpellier, avant d'essaimer pour pénétrer à des degrés divers la plupart des autres pôles scientifiques. Cet essor s'est appuyé sur le développement parallèle des concepts, intégrant notamment la culture mathématique et probabiliste de la génétique des populations et de la génétique quantitative, et des outils, en particulier ceux de la biologie moléculaire, de la modélisation et de l'informatique.

## 5.1 ANALYSE DESCRIPTIVE ET FONCTIONNELLE DE LA DIVERSITÉ DU VIVANT

Les outils de la biologie moléculaire autorisent un marquage d'une puissance et d'une fiabilité inégalées, et sont souvent les seuls disponibles. La description des taxons, leurs relations phylogéniques, la reconstitution de leur généalogie, de l'histoire du peuplement de la planète à différentes échelles spatiales et temporelles ont pris une nouvelle dimension, renforcée par l'étude difficile des ADN anciens qui a ouvert une discipline nouvelle : la paléogénétique, dont la lisibilité commence à s'affirmer en France.

On retrouve les mêmes développements à l'échelle infra-spécifique où l'analyse moléculaire des populations (présentes ou passées) permet de reconstituer leur histoire, de comprendre leur structure, leur fonctionnement et leur dynamique actuels, de comprendre les forces structurant la distribution de la variation génétique (dérive, migration, sélection, démographie historique etc.). Cette révolution de l'étude des structures et de la dynamique spatio-temporelle des populations a permis d'amorcer la synthèse entre les approches concrètes et les modèles théoriques en génétique, dynamique et écologie des populations. On peut commencer à prévoir leur évolution et à organiser la gestion de leurs réponses aux changements anthropiques. La position française est plutôt bonne pour ce qui est des outils et de la compréhension théorique, mais des efforts sont à faire pour développer des modèles sophistiqués incluant ces différentes forces.

Une mention particulière doit être faite pour la microbiologie, où détection et typages à grande échelle sont devenus courants, et surtout pour l'écologie microbienne. Les milieux complexes comme les sols, les eaux continentales, marines ou lagunaires, et le milieu vivant lui-même sont peuplés d'une foule de bactéries, pour la plupart non cultivables et non détectables par les méthodes traditionnelles, et de virus. Il est devenu possible

de détecter et d'identifier ces microorganismes, et même d'aborder l'étude de leur rôle fonctionnel. C'est un champ nouveau de la biologie et de l'écologie qui vient de s'ouvrir, dans lequel prévaudra une approche plus intégrative des génomes, et où l'étude des interactions entre procaryotes et eucaryotes va prendre une place grandissante.

Les champs d'application sont immenses : l'environnement (notamment gestion et conservation de la biodiversité), l'agronomie (biologie des sols, gestion des résistances, protection des cultures, développement durable, traçabilité sanitaire et commerciale), et la santé (pathogènes et parasites, populations vectorielles).

## 5.2 RELATIONS DU GÉNOME À L'ENVIRONNEMENT, MÉCANISMES DE L'ADAPTATION ET DE L'ÉVOLUTION

À côté de ces marqueurs désormais incontournables, la génétique fournit des outils d'investigation fonctionnelle, aussi bien à travers les mécanismes de sélection, qu'à travers l'expression différentielle des gènes en fonction des conditions externes.

Très peu d'études concernent les effets de l'environnement sur la structure et la composition des génomes, ou sur les interactions entre ses différents composants. Ces études sont pourtant indispensables pour comprendre l'adaptation et l'évolution des espèces (génétique des espèces invasives, réponses aux stress, aspects génétiques de la biologie de la conservation). Ainsi connaît-on mal les mécanismes responsables des modifications de taille et de composition des génomes, et l'impact des grands bouleversements historiques du génome sur la variabilité génétique des populations. De même, on ne sait pas prévoir les invasions : quelles espèces ont la capacité à envahir ? Quels changements génétiques associés à ces invasions ? Quels profils d'expression de leurs gènes dans les nouveaux environnements ?

Réciproquement, les effets des êtres vivants sur la physico-chimie de leur milieu, très étudiés à l'échelle des bilans globaux (flux carbonés, cycles biogéochimiques, etc.), sont presque inconnus à l'échelle fonctionnelle locale des organismes et des populations qui en sont les acteurs. L'approche spatio-temporelle fine de ces processus devra intégrer une dimension génétique, aussi bien en ce qui concerne le fonctionnement et l'expression des systèmes génétiques impliqués, que l'organisation et les différenciations génétiques des populations concernées.

Le rôle des associations parasitaires et symbiotiques (les « interactions durables ») dans les processus d'adaptation et d'évolution est encore mal compris, et probablement sous-estimé. Chez les plantes comme chez les animaux, c'est une collection de génomes (y compris procaryotes) qui sont associés à un niveau de chaque « individu », qui interagissent de différentes manières (de la synergie mutualiste à l'antagonisme parasitaire), pour conduire au phénotype opérationnel (le « phénotype étendu »). L'étude de cette confrontation entre génomes distincts demande à être approfondie aussi bien dans une perspective évolutive (co-adaptations, co-évolution, édification des génomes) que dans une perspective fonctionnelle (expression des génomes, réponses à l'environnement).

La réalisation des phénotypes, interface entre génomes et environnement, soulève d'immenses questions, et une mention particulière doit être faite du contrôle génétique et épigénétique des caractères quantitatifs à variation continue. Si certaines adaptations à des perturbations environnementales spécifiques, naturelles ou anthropiques, relèvent de fonctions précises contrôlées par un nombre très réduit de loci, c'est le plus souvent à travers des caractères quantitatifs que sont réalisées les adaptations des organismes aux facteurs environnementaux majeurs (climatiques, trophiques, biotiques, etc.). La génétique quantitative utilise actuellement des outils mathématiques dérivés des méthodes d'amélioration, et son application aux populations naturelles demande de nouveaux développements théoriques et

méthodologiques pour étudier des questions centrales comme l'héritabilité de la valeur sélective, la plasticité phénotypique, les normes de réaction et leur contrôle génétique.

L'étude des transcriptomes à haut débit va donner un nouvel essor à ces recherches en permettant de repérer les gènes impliqués dans ces fonctions, d'étudier leur expression dans des conditions ou sous des environnements naturels ou contrôlés, d'enregistrer les variations entre individus.

La France est globalement peu présente dans ce domaine. Il est urgent que les écologistes, écophysiologistes, évolutionnistes de notre pays prennent conscience de l'apport considérable de la génétique, qui dépasse très largement la seule mise à disposition de marqueurs performants.

### 5.3 L'ÉVOLUTION DES GÉNOMES

L'évolution biologique repose sur la fixation dans le génome d'erreurs produites par différents mécanismes, dont les multiples conséquences vont du simple changement ponctuel d'une base jusqu'à des remaniements profonds du génome ou la génération de nouveaux gènes à partir de gènes préexistants. Le fonctionnement du génome, sa plasticité, ses mécanismes de réparation et de recombinaison portent ainsi la responsabilité directe de la nature des « mutations » soumises aux processus stochastiques ou adaptatifs, et indirecte de la réponse des organismes aux contraintes. L'évolution moléculaire est donc le lieu de rencontre privilégié des approches évolutives et fonctionnelles.

Les aspects bioinformatiques et théoriques de la génomique évolutive (phylogénie, génétique des populations, modélisation dans le domaine de l'évolution) sont bien développés en France, notamment la phylogénie, et il faut souligner le très bon fonctionnement pluridisciplinaire de cette communauté de biologistes et informaticiens autour de la phylogénie et de la génomique comparatives.



L'analyse fine de l'évolution de certaines familles de gènes, l'évolution du code, de la structure introns-exons, des ADN répétés, etc., permet de faire le lien entre le niveau moléculaire et l'environnement auquel doivent s'adapter les espèces.

L'analyse comparative de génomes très distants permet de décrire de manière exhaustive les « patterns » de variations à long terme (les « conséquences » de l'évolution moléculaire). Elle constitue un outil très puissant pour mettre en parallèle structure et fonction des gènes, et comprendre les évolutions différentielles qu'ils ont pu connaître.

L'approche de génomique des populations développe les comparaisons entre espèces proches ou à l'intérieur des espèces, et prend en compte la variabilité entre individus pour un ensemble de gènes. Elle s'appuie sur des modèles de génétique des populations intégrant effets sélectifs, conversion, recombinaison et variabilité des systèmes génétiques. Elle implique des collaborations entre bioinformaticiens et généticiens des populations, et requiert beaucoup de données sur la variabilité fine du génome, notamment l'accès à de nombreux fragments d'ADN séquencés chez de nombreux individus. Elle débouche sur une « génomique microévolutive » dans laquelle les données sont encore peu nombreuses, et où la France, avec sa bonne culture dans le domaine de la génétique des populations, pourrait jouer un rôle.

## 5.4 CONCLUSION

Les connexions entre génétique et biologie moléculaire d'une part, et écologie et évolution d'autre part, se situent dans deux domaines : l'histoire des êtres vivants, et les mécanismes de l'évolution et de l'adaptation. La génétique apporte des outils et des concepts qui permettent de dépasser les approches descriptives et consolident les approches théoriciennes plus récentes en offrant de véritables possibilités d'investigation fonctionnelle et expérimentale. Réciproquement, les

recherches sur l'organisation des génomes, leur évolution et les fonctions des gènes ne peuvent plus se concevoir sans référence aux relations qui s'établissent entre les organismes et leur milieu, soit à court terme au niveau de la réalisation des phénotypes, soit à plus long terme au travers des pressions de sélection et processus d'adaptation.

La communauté concernée en France, au CNRS en particulier, mais aussi à l'INRA, est bien structurée autour de plusieurs pôles d'excellence de renommée internationale et de plusieurs GDR. Paradoxalement, sa situation à l'interface de différents champs disciplinaires la met en grande difficulté car jusqu'ici les sciences de l'environnement et de la biodiversité s'intéressent beaucoup aux faits, résultats et évolutions socialement perceptibles mais très peu aux mécanismes sous-jacents, et réciproquement, la génétique conserve encore une approche fortement réductionniste. Alors que les clivages scientifiques et intellectuels entre ces disciplines commencent à s'atténuer en France, il en va tout autrement au niveau institutionnel. Il est extrêmement urgent de promouvoir l'interdisciplinarité au niveau de ces sous-disciplines biologiques, aussi bien en incitant à la programmation d'opérations de recherche transversales ou d'interface, qu'en veillant à la cohérence et à la modernité des programmes de formation dès les premiers cycles universitaires.

## 6 – GÉNOMES ET MICROBIOLOGIE

### 6.1 LES ENJEUX DE LA RECHERCHE EN MICROBIOLOGIE

Outre le rôle clé joué par l'étude des microbes comme systèmes modèles dans l'essor de la biologie moléculaire, ces modèles

ont également fourni un apport essentiel au développement de l'ensemble des outils et approches constituant la génomique. Les premiers génomes complètement séquencés étaient ceux de deux microorganismes, le procaryote bactérien *Haemophilus influenzae* et l'eucaryote *Saccharomyces cerevisiae*. Début 2004, les séquences de 1 487 génomes viraux et 159 génomes bactériens (eubactéries et archea confondus) étaient présents dans les bases de données publiques, et presque 400 génomes additionnels de procaryotes sont en cours d'analyse. À ces chiffres, il faut ajouter les génomes de microorganismes eucaryotes. Ces données ne sont pas si surprenantes si nous considérons la taille réduite des génomes microbiens par rapport aux eucaryotes supérieurs, l'énorme diversité de la vie microbienne (tant du point de vue des espèces que de la variété des niches écologiques) et l'impact des microbes dans tous les domaines de la vie.

L'émergence ou la re-émergence des maladies infectieuses dans le monde, les problèmes posés par les infections nosocomiales et les phénomènes de résistance aux antibiotiques, ont remis la microbiologie sur le devant de la scène, associant exclusivement, dans l'esprit du grand public, les microorganismes aux maladies infectieuses. Mais cet aspect ne représente qu'une partie infime de l'impact des microbes sur notre vie. Les microbes constituent une partie significative de la biomasse, pouvant aller jusqu'à 90 % selon certaines estimations. Les virus bactériens à eux seuls sont vraisemblablement les entités les plus nombreuses dans la biosphère et représentent un pool génique énorme. Les microorganismes occupent l'ensemble d'une biosphère large de 30 kilomètres, depuis les failles océaniques les plus profondes jusqu'à plusieurs kilomètres dans l'atmosphère, et peuvent résister aux pressions, aux températures, aux conditions de salinité et de pH les plus extrêmes. Ils jouent un rôle essentiel dans les grands cycles écologiques tels ceux du carbone et de l'azote, dans la croissance des plantes ou des animaux via des symbioses délicates,

dans la corrosion, dans la régénération des sols. Ils sont depuis longtemps exploités par l'homme pour la production d'aliments, ou plus récemment pour la production d'antibiotiques. Leur potentialité d'utilisation pour le recyclage de la biomasse, la dépollution, la production de molécules à haute valeur ajoutée, est immense.

Malgré les enjeux de connaissances fondamentales et les enjeux sociaux et économiques liés au développement de la microbiologie, force est de constater que l'effort français de recherche dans ce domaine n'est pas à la hauteur. Au niveau de la recherche fondamentale, on a assisté, dès le milieu des années 80, à une désaffection des chercheurs pour la microbiologie. Ce mouvement s'est accompagné d'une baisse des financements et d'une certaine désaffection des étudiants, qui voient en elle, à travers l'enseignement qu'ils reçoivent, une discipline « historique ». Bien que des laboratoires français aient joué un rôle important dans le séquençage des génomes de *Bacillus subtilis* et de la levure, qu'ils continuent de jouer un rôle important pour le séquençage de plusieurs microorganismes, et qu'ils soient impliqués dans plusieurs programmes de génomique fonctionnelle de microorganismes, globalement la contribution française à la mine d'information génomique a été bien moindre que celle des États-Unis et de la Grande Bretagne. De plus, au niveau de l'industrie, on assiste au niveau mondial, et principalement en Europe, à un désengagement des grands groupes pharmaceutiques de la recherche sur les anti-infectieux, qui par ailleurs n'était plus depuis longtemps une priorité majeure.

Pourtant, le développement de la génomique, en particulier de la génomique fonctionnelle, relance l'intérêt des modèles microbiens, ce qui s'est déjà traduit par un certain renversement de tendance aux USA, alors que le déclin semble se poursuivre non seulement en France mais en Europe. Il est temps de renverser la tendance à notre tour.

## 6.2 GÉNOMIQUE ET RENAISSANCE DE LA MICROBIOLOGIE

### L'apport de la génomique en microbiologie

Comme pour tout organisme, la connaissance et l'analyse de la séquence génomique, l'analyse globale de l'expression génique, du protéome, des interactions entre protéines, des voies métaboliques, facilite une vision intégrée du fonctionnement cellulaire et peut parfois déboucher sur l'identification de fonctions cryptiques. Pour ce qui concerne les microorganismes, les analyses du transcriptome et du protéome dans différentes conditions de culture commencent à porter leurs fruits en fournissant une carte globale de la dynamique de l'expression d'un génome et de ses réponses aux différents signaux externes. Cela constitue un atout essentiel pour la compréhension des circuits métaboliques et des systèmes de perception de l'environnement et de transduction de signal. La génomique comparative entre espèces proches a déjà fourni des perspectives inattendues sur l'évolution des processus comme la pathogénicité et la fixation de l'azote. La comparaison d'organismes voisins pathogènes et non pathogènes ou ayant des hôtes différents (comparaison de génomes entiers ou soustraction génomique) permet d'identifier des candidats pour les gènes de pathogénèse ou les gènes déterminant la spécificité d'hôte, et donc de cibler plus efficacement les études ultérieures.

### L'intérêt des modèles microbiens en génomique

**1. Fonctionnement et évolution des génomes :** La détermination des génomes de microorganismes reste beaucoup plus aisée que celle d'organismes multicellulaires macroscopiques, au moins pour ce qui est des microorganismes cultivables. L'accumulation des données concernant ces microorganismes représente une base d'analyse largement

majoritaire en ce qui concerne les études évolutives. De plus, les microorganismes sont présents dans les trois grands domaines du vivant, représentant à eux seuls l'intégralité de deux de ces règnes (eubactéries et archées), et une large partie de la biodiversité des eucaryotes. Pour comprendre la diversification des différentes branches du vivant, il est important de pouvoir analyser et comparer de nombreux exemples de génomes dans ces trois grands domaines. Il ne faut pas oublier non plus les virus et bactériophages, dont on apprécie de plus en plus la diversité biologique ainsi que le rôle important qu'ils ont pu jouer dans l'évolution, par le biais de transferts de gènes. Indépendamment des virus, les transferts horizontaux entre microorganismes partageant des mêmes niches écologiques sont de plus en plus documentés et contribuent fortement à la plasticité génomique. De tels transferts peuvent conduire à la transmission de gènes de pathogénicité ou de résistance aux antibiotiques.

**2. Les microorganismes comme organismes modèles de fonctionnement cellulaire intégré :** Un des objectifs essentiels de la génomique est sans doute de comprendre le fonctionnement intégré d'un génome au niveau d'une cellule entière. Or les organismes microbiens modèles tels la levure, *E. coli* ou *B. subtilis*, restent les meilleurs systèmes d'étude pour réaliser cet objectif. L'analyse du fonctionnement intégré du génome dans la physiologie cellulaire passe non seulement par le développement de la génomique et des disciplines connexes (bioinformatique par exemple), mais par un nouveau développement de la génétique. En effet, même chez des organismes possédant un nombre relativement limité de gènes (quelques milliers pour les bactéries ou la levure) une proportion importante de ces gènes sont soit totalement « orphelins », soit de fonction précise inconnue, même s'ils peuvent être rattachés à une classe fonctionnelle connue ou à une famille de gènes. Découvrir la fonction de ces gènes est un des véritables défis de la génomique fonctionnelle. Si les approches dites globales ou à haut débit fournissent à la communauté scientifique des ressources indispensables, l'exploitation complète des données

ainsi accumulées ne pourra se faire que grâce à des efforts concertés associant les compétences de nombreux laboratoires, à travers un renouveau de l'intérêt pour la physiologie cellulaire, s'appuyant nécessairement, à côté des approches génomiques et structurales plus soutenues, sur des approches génétiques.

### **6.3 LA RECHERCHE PUBLIQUE FRANÇAISE EN MICROBIOLOGIE ET LES APPROCHES GÉNOMIQUES**

Le CNRS représente une part importante de la microbiologie française. De très nombreuses équipes, en particulier au CNRS, mais aussi des équipes universitaires, s'intéressent aux divers aspects de la physiologie des enveloppes microbiennes (transports, énergétique, composants), ainsi qu'aux mécanismes d'adaptation des microorganismes à leur environnement et aux interactions cellulaires (en particulier procaryotes/eucaryotes). Les études hôtes-pathogènes sont très présentes à l'Institut Pasteur, en association avec le CNRS, et dans une moindre mesure avec l'INSERM, mais également dans diverses unités du CNRS en association avec des universités, que ce soit pour des pathogènes bactériens ou des parasites. Par contre, les études sur les anti-infectieux, les cibles potentielles, les mécanismes de résistance, au-delà de quelques laboratoires phares, paraissent insuffisamment développées, au regard des enjeux, que ce soit au CNRS ou à l'INSERM. La microbiologie semble moins bien représentée à l'INSERM où elle concerne surtout la virologie et l'étude de la réponse immunitaire. La plupart des équipes travaillant sur des maladies infectieuses (bactériennes ou dues à des parasites) sont associées au CNRS et/ou à l'Institut Pasteur. L'épidémiologie des maladies infectieuses semble relativement limitée.

Les microorganismes restent largement utilisés comme modèles d'étude des mécanismes de base de la cellule, particulièrement en ce qui concerne le métabolisme de l'ADN

et l'ARN et la photosynthèse. Le CEA est très présent dans ce dernier domaine, essentiellement à travers des études biophysiques et structurales. Au CEA on peut aussi citer l'utilisation de protéines de la levure comme modèle pour l'étude du prion et de la transcription. La France possède également une tradition d'utilisation des champignons filamenteux comme organismes modèles.

Plusieurs laboratoires de l'INRA, particulièrement à Jouy en Josas et à Grignon, ont joué un rôle important dans le développement de la génomique de *Bacillus subtilis* et de différentes levures. Des projets importants portent sur les bactéries lactiques. D'autres projets s'intéressent aux microorganismes d'intérêt agroalimentaire et industriel tels que levures et bactéries du vin, pathogènes de plantes ou régulateurs d'insectes ravageurs. La génomique se développe dans le domaine de l'étude des flores microbiologiques (microorganismes de la rhizosphère, particulièrement avec la contribution de l'INRA et du CNRS à Toulouse, colon humain, rumen, aliments, eau, fruits, etc.), au niveau des écosystèmes ou des propriétés physiologiques, métaboliques, et génétiques des souches utilisées, avec plusieurs projets de génomique fonctionnelle.

De nombreux laboratoires s'intéressent à la biodiversité et à l'écologie microbienne, dans le cadre du département des sciences de la vie du CNRS, mais aussi du département des sciences de l'univers. Nombre de ces laboratoires seront amenés à intégrer les approches génomiques, bioinformatiques et de génétique moléculaire.

Un des points forts de la recherche française est de bien couvrir l'ensemble de la biodiversité microbienne (virus, bactéries des différentes branches, archées, protistes, champignons et algues microscopiques) ce qui est un atout certain à préserver. L'étude des aspects moléculaires de cette diversité, en particulier grâce aux apports de la génomique, est très certainement à renforcer.

Il semble que dans tous les organismes de recherche, les laboratoires s'intéressant aux microorganismes affirment leur volonté de se tourner vers des approches génomiques. De

nombreux laboratoires ont fait du génome un objet d'études : mécanismes de fonctionnement des génomes (stabilité, plasticité, transfert de gènes), comparaison génomique d'espèces proches, évolution, biodiversité.

## 6.4 LES MOYENS POUR LA GÉNOMIQUE À LA DISPOSITION DES MICROBIOLOGISTES

### Séquençage

L'immense majorité des génomes séquencés à ce jour, ou en cours de séquençage, sont des génomes de microorganismes.

Loin de s'atténuer, les besoins en matière de séquençage de génomes microbiens vont probablement rester importants. La connaissance de la biodiversité et le développement de l'écologie microbienne passe par le séquençage de nouveaux génomes. Le développement de nouveaux systèmes modèles hôtes-pathogènes particuliers, l'émergence de nouvelles pathologies infectieuses (voir l'exemple récent du SARS), nécessiteront le séquençage de nouveaux pathogènes. De plus, la comparaison de génomes très proches ou de sérotypes pathogènes et non pathogènes est riche d'enseignements tant au niveau des mécanismes de l'évolution que de l'identification de facteurs de pathogénicité ou de spécificité d'hôte.

L'effort français de séquençage, pour ce qui est des programmes à grande échelle, se fait essentiellement au Genoscope. Plusieurs projets en cours ou qui viennent d'être achevés concernent des microorganismes (environ 16 projets sur 45 répertoriés). Il est probable que la capacité de séquençage du Genoscope ne saurait répondre aux besoins, alors même qu'il n'existe aucun programme, en particulier au ministère de la recherche, pour financer des projets de séquençage qui pourraient éventuellement être réalisés sur d'autres plates-formes.

Les seuls génopoles ayant des projets de génomique fonctionnelle de microorganismes sont, en région parisienne, le site de Pasteur/Necker et, en province, Toulouse, Marseille et Montpellier. La génomique des microorganismes marins pourrait peut-être trouver une place dans le programme mer du génopole ouest.

### Les programmes nationaux

Il y a deux types de programmes qui pourraient soutenir la génomique des microorganismes : les programmes de génomique proprement dits, ou les programmes « microbiologie ». En ce qui concerne le séquençage du génome de microorganismes d'intérêt pour les chercheurs français, il n'existe aucun programme, en dehors des appels d'offres annuels du Genoscope, qui ne sauraient répondre à l'ensemble des besoins. Au niveau du CNRS, un programme « puces à ADN » est maintenant « achevé » (mais pas les recherches en la matière !). Il est vraiment dommageable pour la recherche que les quelques efforts entrepris ne soient que ponctuels et rapidement classés « sans suite ». Le programme ministériel qui subventionne les puces affymetrix ne répond pas nécessairement aux besoins des utilisateurs. Le seul autre programme de génomique est le programme « protéomique et génie des protéines » du CNRS qui a retenu en 2003 26 projets, dont 5 concernent la protéomique de microorganismes.

En ce qui concerne les programmes axés sur la microbiologie, les difficultés du nouveau programme microbiologie, à l'avenir très incertain, illustrent bien les difficultés de la communauté des microbiologistes à trouver des financements. En ce qui concerne l'écologie et la biodiversité microbienne, le seul programme existant est Geomex, très ciblé et qui est très loin de couvrir l'ensemble des besoins dans ce domaine. Enfin la génomique fonctionnelle implique nécessairement de s'allier à des approches génétiques. Or, comme nous l'avons noté plus haut, il n'existe à l'heure actuelle plus aucun financement spécifique pour la génétique en France, qu'il s'agisse de microorganismes ou d'autres organismes.

## 6.5 CONCLUSION

La microbiologie peut offrir de nombreux modèles pour la génomique fonctionnelle, et dans le même temps, plusieurs secteurs de la microbiologie peuvent se renouveler de manière considérable grâce à la génomique (biodiversité microbienne, réponses aux stress et adaptation aux milieux, interactions hôtes-pathogènes, etc.). Comme pour d'autres organismes, tirer les bénéfices de la révolution génomique ne se fera pas sans un nouveau développement de la génétique. Une réflexion sur l'articulation entre approches globales et approches plus classiques, entre plates-formes et travail au niveau des équipes, est également souhaitable. Les moyens financiers et d'infrastructure nécessaires sont importants pour la microbiologie comme pour d'autres disciplines. La faiblesse des moyens attribués à ce secteur, comme le manque d'infrastructures et plates-formes réellement orientées vers la microbiologie ne fait que refléter les difficultés de la microbiologie française, qui n'est plus une priorité depuis des dizaines d'années. Beaucoup d'organismes de recherche et de génopoles orientent plutôt les moyens dont ils disposent vers la génomique fonctionnelle des plantes et/ou de modèles animaux. On peut gager qu'on ne fera pas l'économie de développer en parallèle la génomique des microorganismes, a priori beaucoup plus abordable, et présentant souvent l'avantage de pouvoir s'appuyer sur des études génétiques approfondies.

# 7 – LES GÉNOMES ET L'AGRONOMIE

## 7.1 LES GÉNOMES ET LA GÉNÉTIQUE VÉGÉTALE

Les plantes sont l'une des composantes majeures de notre environnement et jouent un rôle considérable dans la production d'oxygène,

l'absorption du gaz carbonique et dans le cycle de nombreux éléments minéraux. Enfin, en matière de santé, un très grand nombre de médicaments utilisent des principes actifs d'origine végétale (aspirine, antitumoraux tels que taxol et vinblastine, morphine, etc.) et la biodiversité végétale est loin d'avoir été complètement exploitée par l'homme pour assurer sa nutrition. Toute avancée sur la connaissance des voies de biosynthèse de ces produits a, ou aura, un impact sur la nutrition ou la production de médicaments et donc sur la santé. Au plan pratique, le développement de la génétique végétale a permis l'essor de l'amélioration des plantes et de la sélection variétale sur laquelle repose toute notre agriculture et une partie des biotechnologies végétales.

La génétique et la génomique végétales revêtent une importance considérable, tant sur le plan théorique que pratique. Au plan de la contribution à l'acquisition des connaissances, les travaux de génétique végétale ont d'abord contribué à établir les lois fondamentales de la génétique avec les études réalisées sur le pois par Gregor Mendel. Ceux sont les travaux de Barbara McClintock sur le maïs qui ont conduit à la mise en évidence des éléments transposables et aux concepts de fluidité des génomes. Plus récemment, les travaux de mise au point de plantes transgéniques ont révélé l'existence de mécanismes d'inactivation de l'expression des gènes. Les travaux pionniers allemands et anglais sur la génétique du développement floral du mufler ont conduit à identifier les gènes homéotiques de développement floral. Ces divers travaux ont illustré les particularités des caractéristiques du développement des végétaux supérieurs. De façon générale, la génétique végétale a permis au cours des dix dernières années d'analyser les processus de développement, et les voies de transduction des signaux hormonaux ou environnementaux. Ces découvertes orientent dès maintenant les stratégies d'amélioration des plantes et d'utilisation des intrants.

Les efforts de la génomique végétale ont initialement porté sur des espèces ayant à la fois « une bonne génétique » et une valeur de modèle ; ils s'étendent maintenant aux principales espèces d'intérêt agronomique dont les

génomomes sont le plus souvent beaucoup plus gros. Il était réaliste d'envisager le séquençage des petits génomes comme celui d'*Arabidopsis* (130 Mégabases) ou du riz (440 Mégabases), beaucoup moins d'envisager de séquencer celui du maïs (2 600 Mégabases) ou du blé (16 000 Mégabases). Pour la plupart des espèces agronomiques l'étape initiale a consisté à réaliser des cartes saturées par des marqueurs RFLP puis à dresser des catalogues de gènes à partir d'étiquettes ADNc (EST).

### **Les travaux de recherche sur l'espèce modèle *Arabidopsis* et les autres espèces modèles**

Ce n'est qu'au cours des années 80 que le choix d'*Arabidopsis thaliana* comme espèce modèle s'est imposé au niveau mondial, et a finalement « détrôné » aux USA les espèces phares maïs, soja et tomate. En effet *Arabidopsis thaliana* est la première espèce végétale pour laquelle diverses ressources génomiques ont simplifié son utilisation pour des travaux de génétique moléculaire. En premier lieu une ressource EST a été constituée dès 1992 (consortium français), puis une ressource YAC de bonne qualité couvrant l'ensemble du génome (consortium français) qui a permis le clonage positionnel de nombreux gènes. Une collaboration Allemagne-USA a par la suite réalisé une couverture BAC du génome permettant le lancement d'un programme international de séquençage du génome qui s'est terminé en 2000, programme auquel le Genoscope, et ESSA (« *European Scientists Sequencing Arabidopsis* ») impliquant quelques laboratoires français, ont apporté une contribution de qualité. Ce programme est complété par trois initiatives (USA, Japon, France) de production de séquences d'ADNc pleine longueur qui permettent une annotation solide du génome et qui couvrent de l'ordre de 24 000 gènes. Les divers efforts visant à maîtriser l'usage de transposons pour l'étiquetage de gènes ont finalement été distancés par la mise au point à Versailles d'une technique de mutagenèse d'insertion de l'ADN-T d'*Arabidopsis* par transformation *in planta*. Cette

technique a permis de générer des collections importantes de lignées étiquetées dont certaines sont exploitables pour des approches de génétique réverse par séquençage systématique des sites d'insertion. Parmi celles-ci les collections Génoplante, GABI, Signal et Syngenta sont les plus importantes (avec des contraintes d'accès fortes pour cette dernière). Pour compléter cet outil de génétique réverse, un programme européen, coordonné par la France, a été créé pour produire, à terme, une collection de lignées « RNAi » affectée sur chacun des 20 000 gènes annotés du génome d'*Arabidopsis*. De puissants outils d'analyse du transcriptome sont maintenant également disponibles ou en cours d'élaboration. La firme Affymetrix a été la première à commercialiser une puce oligonucléotides de 8000 gènes d'*Arabidopsis*. Affymetrix produit maintenant une puce de 22000 séquences en même temps qu'une batterie de puces « génomiques » couvrant sous forme d'oligonucléotides l'ensemble du génome. Parallèlement le consortium européen CATMA construit une puce de 25000 éléments basée sur l'emploi de sondes gène-spécifiques qui est maintenant opérationnelle. Enfin une collaboration CNRS-CSHL USA vise à créer une puce chromosomique permettant d'étudier les mécanismes de contrôle épigénétique de l'expression des gènes sur le Chromosome 4. Cet éventail d'outils est complété par la constitution de ressources visant à analyser la diversité biologique de l'espèce *Arabidopsis*. Les Pays-Bas ont été pionniers dans ce domaine, suivis par le MPI de Jena (Allemagne) puis le Salk Institute. En France, une « core collection » et un ensemble de RIL ont été constitués à Versailles, ce qui représente une ressource significative pour la communauté scientifique. Les outils SNP permettant d'analyser le polymorphisme dans ces populations sont développés dans le cadre d'une collaboration INRA-CNG. Les travaux d'analyse du protéome menés en France, principalement au CEA de Grenoble et à l'INRA de Montpellier ont une bonne tenue scientifique, par contre un important retard a été pris dans le domaine de l'analyse du métabolome d'*Arabidopsis*, domaine dans lequel d'autres pays d'Europe, l'Allemagne en particulier, ont pris

une longueur d'avance. Globalement l'INRA et le CNRS tiennent une place significative dans la recherche en génétique moléculaire menée sur *Arabidopsis* au niveau international. La France se situe au cinquième rang mondial des citations par pays, derrière les USA, le Royaume Uni, l'Allemagne et le Japon, ce qui reste honorable.

Le riz est devenu, à partir de 1997 la seconde espèce végétale dont le génome serait séquencé. Les équipes françaises, principalement à Montpellier et Perpignan, se sont mobilisées pour assurer les contacts et réaliser le travail préliminaire de cartographie du chromosome 12 qui a été finalement séquencé en totalité par le Genoscope dans le cadre du consortium international public. La même stratégie de création de ressources et de mise en place de collections de mutants d'insertion a été réalisée, en particulier à Montpellier.

Plus récemment, une luzerne sauvage, *Medicago truncatula*, a été poussée sur le devant de la scène grâce aux efforts des équipes INRA et CNRS de Toulouse, Gif sur Yvette et Dijon. Son séquençage génomique est en cours, le Genoscope prenant en charge l'un des huit chromosomes.

## Le programme Génoplante

*Arabidopsis* est un excellent modèle pour l'étude de nombreuses fonctions, mais s'agissant d'une plante sauvage, elle est dépourvue de nombreux caractères connus sur les espèces cultivées, ce qui contribue à limiter son intérêt pour certaines approches (par exemple gènes support de domestication, formation des tubercules, des fruits, fixation racinaire de l'azote, formation du bois, etc.). Ces divers éléments montrent l'importance des travaux de génétique sur les plantes cultivées. Qu'en est-il des recherches menées aujourd'hui ? Aux USA une forte tradition de travaux de génétique sur les plantes cultivées a fourni la justification de travaux de génomique. Un important programme de génomique portant sur un nombre conséquent d'espèces cultivées, parallèle au programme

« *Arabidopsis* for year 2010 » a été financé par la NSF avec le support déterminé du congrès américain. Malheureusement, il n'en a pas été de même au niveau européen. L'Union Européenne a financé une partie du séquençage du génome d'*Arabidopsis* dans le cadre du « framework 4 », mais n'a plus individualisé les thématiques végétales dans ses appels d'offre à partir du « framework 5 », et a réorienté par la suite ses programmes sur les questions de sécurité alimentaire. La contribution de l'Union Européenne à la génomique végétale a donc été très limitée dans le cadre des « frameworks » 5 et 6, malgré la qualité des projets et des équipes proposant. Seul un projet de génomique des légumineuses est retenu dans le cadre du « framework 6 ».

Notre pays a un savoir faire reconnu dans le domaine de la génétique quantitative, mais a tardé à introduire les outils de la génétique moléculaire pour étudier les génomes de plantes cultivées. Deux programmes européens (« framework 4 »), EGRAM et EuDicotMap, auxquels ont participé des équipes du CIRAD, du CNRS et de l'INRA, ont apporté une contribution internationalement reconnue en effectuant les travaux de cartographie RFLP comparée établissant l'existence d'importantes conservations de synténie entre génomes de céréales et plus limitées entre espèces dicotylédones et *Arabidopsis*. Le nombre réduit de travaux de génétique mendélienne sur les plantes cultivées en France a limité les initiatives d'identification des supports moléculaires de caractères agronomiques et de développement des outils de génomiques nécessaires à ces approches (banques BAC, puces ADN). Le constat a été fait par la direction de l'INRA et celle de firmes telles que Aventis et Limagrain que le retard accumulé pouvait avoir des conséquences graves pour la compétitivité de la recherche publique, mais aussi une incidence très négative sur la compétitivité du secteur semencier. Par exemple, en 1999, aucun gène « d'importance agronomique » n'avait encore été identifié en France contrairement à de nombreux autres pays (USA, Royaume-Uni, Allemagne, Pays-Bas, et Australie) qui mènent sur la base de ce type de découvertes une politique offensive de



protection par brevet. Il a donc été décidé de mettre en place un programme mobilisateur, Génoplante, financé en principe pour un tiers par les instituts publics, un tiers par le secteur privé et un tiers par le ministère de la recherche (en réalité 40/40/20 !). Après quatre années, les résultats obtenus sont impressionnants et notre pays a pu créer des outils de génomique opérationnels pour cinq espèces cultivées (maïs, blé, colza, riz, pois) en plus du soutien qu'il a apporté aux travaux de génomique sur *Arabidopsis*. Des ressources génomiques ont été créées en premier lieu. Parmi celles-ci notons la production de marqueurs microsatellites ou SNP, banques BAC (par exemple blé hexaploïde), collections importantes d'EST (par exemple 150 000 EST de Maïs, 100 000 EST de blé, etc.) et puces ADN dérivées, bases de données et outils logiciels bioinformatiques pour l'annotation des génomes. Chez le maïs et le riz des ressources de génétique réverse ont été aussi constituées. Des travaux d'analyse QTL ont été menés sur toute une série de caractères agronomiques sur lesquels des travaux de physiologie ayant un intérêt cognitif peuvent s'appuyer (remplissage de la graine, adaptation au stress hydrique, précocité de floraison, lignification, etc.). Enfin trois gènes d'importance agronomique, un gène de résistance aux pucerons, VAT, un restaurateur de fertilité permettant la production de semences hybrides de colza, Rfo, et un gène de résistance récessive aux virus, Pvr2, ont été récemment clonés. Dans ce même cadre, trois plates-formes de génomique ont été constituées qui jouent un rôle déterminant dans la conduite de ces nouveaux programmes à Evry (URGV INRA-CNRS, laboratoire RHOBIO privé) et au CIRAD de Montpellier. Au moins aussi important, une dynamique nouvelle a été initiée qui a permis à un certain nombre d'équipes travaillant sur plantes cultivées de renouveler leurs approches et de former une réelle communauté scientifique en génétique moléculaire des plantes.

Il faut aussi souligner que Génoplante est loin de couvrir l'ensemble des projets de génétique moléculaire des plantes : la plupart des espèces cultivées tropicales, cibles de l'IRD et du CIRAD, en sont exclues ; la vigne, la tomate,

les arbres fruitiers ou forestiers n'ont que peu bénéficié du programme car ces espèces n'étaient pas identifiées comme prioritaires par les partenaires privés de Génoplante, surtout céréaliers. Cependant, les équipes travaillant sur ces espèces ont largement bénéficié de la dynamique créée.

### **Quel avenir pour la recherche en génomique végétale ?**

De nombreuses ressources moléculaires et génétiques ont été créées au cours des dernières années et il est maintenant réaliste d'identifier le support moléculaire d'un caractère mendélien par clonage positionnel. De nouveaux outils renouvellent complètement nos démarches expérimentales (par exemple TILLING, puces à ADN, etc.) et ouvrent de nouvelles perspectives de recherche et d'applications. Il est donc indispensable de poursuivre l'effort de génomique entrepris, en particulier sur les plantes cultivées. C'est ici la mission des organismes publics de recherche, de l'INRA, du CIRAD et de l'IRD en particulier qui ont la capacité de gérer les ressources biologiques nécessaires. Ce domaine de recherche présente de vastes perspectives d'application dans le domaine de l'amélioration des plantes cultivées, grâce à la sélection assistée par marqueurs qui peut dorénavant s'envisager, dans les domaines de la traçabilité ou la découverte des mécanismes qui sous-tendent la qualité des produits. Tout en poursuivant les travaux sur *Arabidopsis*, il semble important de développer des travaux de génétique sur d'autres espèces, en particulier sur le riz, modèle d'étude des céréales et sur *Medicago truncatula*, légumineuse modèle qui est, avec le maïs et la tomate, la prochaine priorité internationale en matière de séquençage. Le riz présente l'intérêt particulier d'être un bon modèle pour l'analyse de la variabilité génétique, ce qui, pour être exploité, nécessite un investissement important dans le séquençage d'allèles. Le nombre d'espèces cultivées sur lesquelles il est judicieux de faire des travaux de génétique et de génomique de grande ampleur est une question dont la réponse dépend des

moyens que l'on veut investir. Il est essentiel de participer activement aux grands projets internationaux et de partager les charges les plus lourdes. Dans une vision minimale, il semble cependant important d'étudier au moins une céréale cultivée en France (le blé), une légumineuse (le pois, légumineuse à grosses graines source de protéines végétales), un oléagineux (le colza, candidat à devenir une ressource majeure en biocarburants), enfin une espèce ligneuse et productrice de fruits, la vigne, espèce en péril dans notre pays du fait de la pression exercée par de nombreux pathogènes qui en font la première culture consommatrice de phytosanitaires. Le caractère stratégique des travaux menés sur la vigne plaide en faveur du séquençage rapide du génome de cette espèce afin d'y localiser en particulier les gènes impliqués dans la protection contre l'attaque des pathogènes. Une espèce fruitière tropicale telle que la banane est aussi une possibilité intéressante envisagée par le CIRAD.

Quels thèmes majeurs aborder en France grâce à la génomique ? Nombreuses sont les équipes de recherche qui ont intégré la génétique moléculaire à leurs approches. A de rares exceptions près, il n'y a plus guère de laboratoire végétaliste, INRA ou CNRS, qui n'utilise pas *Arabidopsis* comme outil de génétique moléculaire pour étayer ses travaux. Cette situation est considérablement différente de celle à laquelle nous étions confrontés il y a 15 ans. La France est compétitive au niveau international dans différentes thématiques : formation de la paroi et lignification, formation du fruit, remplissage et qualité de la graine, adaptation aux stress hydriques et nutritionnels, membranes et systèmes de transport, nodulation, biogenèse des plastes, système reproducteur, recombinaison et méiose, méristèmes, processus épigénétiques de contrôle de l'expression des gènes, stabilité du génome et rétrotransposons, mécanismes de réparation de l'ADN. La liste n'est pas exhaustive, mais il reste par ailleurs des domaines importants insuffisamment couverts dans notre pays, parmi lesquels il faut encourager en priorité l'étude des fonctions qui sont originales aux végétaux. Citons la mycorhization, la synthèse des

métabolites secondaires (les plantes fournissent 50 % des molécules d'intérêt pharmaceutique), l'induction et le développement floral, le mode d'action des phytohormones, les mécanismes de défense contre les pathogènes et prédateurs et de co-évolution associés, etc. Grâce à l'accumulation des données génomiques, un important effort de recherches est à mener dans le domaine émergent de la biologie intégrative des systèmes complexes. Il y aurait lieu de créer de nouvelles équipes sur ces thèmes de recherche, de préférence sur des sites où existe une masse critique de compétences.

### **Un sujet sensible : les OGM végétaux et la biotechnologie végétale**

Même si la création de plantes transgéniques (OGM) n'est que l'une des multiples retombées de la génétique tant dans un but de recherche que d'applications, elle est souvent assimilée à la génétique végétale. La mauvaise perception des OGM a donc un impact sur l'essor futur de la génomique végétale. Les travaux pionniers dans ce domaine ont été conduits principalement en Europe au cours des années 80, en particulier sous l'impulsion de J. Schell et M. Van Montagu à Gand, et de Jacques Tempé à Versailles. La machinerie cellulaire de la bactérie phytopathogène *Agrobacterium tumefaciens*, qui lui permet d'insérer un fragment oncogène d'ADN bactérien dans le génome d'une cellule végétale, a été utilisée pour développer des méthodes efficaces de transformation génétique des plantes, ouvrant la voie à de multiples travaux de recherche fondamentale, en particulier à tous ceux qui permettent d'établir la fonction d'un gène par son inactivation ou par complémentation d'une mutation. Parmi ces travaux notons la découverte majeure des mécanismes d'inactivation épigénétique des gènes (« silencing »). Mais par ailleurs ces techniques ont aussi permis le développement de biotechnologies présentant un intérêt agronomique (résistance aux herbicides, création de nouvelles résistances aux maladies ou

aux insectes, modification des réserves des graines, production de molécules d'intérêt thérapeutique, dépollution, etc.). L'essor de la génomique permettra aussi de réaliser des OGM mieux définis, plus performants et mieux traçables. Pourtant ces travaux ont été, et sont encore, considérablement freinés en Europe par l'existence du moratoire sur les OGM, alors qu'ils sont soutenus massivement aux USA. Aujourd'hui 67,7 millions d'hectares d'OGM, soit plus de quatre fois la surface cultivable française, sont cultivés à travers le monde dont une part infime (moins du millième) en Europe, sans que cela aboutisse aux catastrophes écologiques annoncées. Il en résulte malheureusement un désengagement des industries de ce secteur et une désaffection marquée en matière d'orientation professionnelle des jeunes vers les métiers de la recherche dans le domaine végétal, où les candidatures à des DEA de biologie moléculaire végétale ont été divisées par cinq en cinq ans. De la réponse apportée à ce problème dépend largement la compétitivité de notre industrie agroalimentaire et l'avenir de ce secteur de recherche, mais aussi notre PNB et notre bien être collectif.

## 7.2 LES GÉNOMES ET LA GÉNÉTIQUE ANIMALE

Les travaux de recherche en génétique animale menés dans le domaine public en France, essentiellement à l'INRA, concernent l'analyse des génomes, le déterminisme des caractères phénotypiques, et les méthodes de gestion des populations animales domestiques.

### Génomique et analyse des génomes

L'étude des génomes d'animaux domestiques a débuté au début des années 90 par la construction des cartes génétiques des espèces agronomiques majeures dans le cadre de consortiums européens. L'arrêt des financements européens en 1997 a d'abord

sensiblement ralenti ces programmes et le niveau de coopération internationale, puis a provoqué l'émergence d'un consortium national, AGENAE.

L'effort a porté initialement sur la production de marqueurs microsatellites et la construction de cartes génétiques, cytogénétiques et, plus récemment, de cartes d'irradiation, domaine dans lequel la France occupe une place de premier plan. L'INRA a ainsi eu une contribution importante, parfois majeure, dans la construction des cartes du bovin, de l'ovin, du caprin, du porc, de la poule, du cheval, de la caille et du lapin.

Depuis plusieurs années, la stratégie prioritaire est le développement de la cartographie comparée entre espèces, en particulier avec les génomes de l'homme et de la souris, permettant ainsi le transfert de l'information des espèces les mieux connues vers les espèces agronomiques auxquelles des moyens beaucoup plus réduits sont consacrés. D'abord à l'échelle chromosomique ou sub-chromosomique par peinture hétérologue au milieu des années 90, elle se base maintenant sur les cartes physiques et les comparaisons de séquences. Des développements bioinformatiques ont débouché sur le logiciel ICCARE, un outil puissant de comparaison des séquences entre espèces. Vers la fin des années 90, des banques de BAC ont été produites pour la plupart des espèces étudiées et, pour le bovin et le porc, elles ont été assemblées et ancrées sur le génome humain à la fois par des marqueurs et par séquençage d'extrémités des BAC, en collaboration avec le Genoscope. Dans ces deux espèces, ce travail réalisé dans le cadre d'AGENAE permet à l'INRA de participer aux consortiums de séquençage des génomes complets, en cours pour le bovin, en projet pour le porc.

À la fin des années 90, un effort de génomique fonctionnelle a été réalisé en parallèle avec la construction de banques d'ADNc normalisées de différents tissus, le séquençage d'EST, en vue de produire des puces génériques ou dédiées (glycogénome par exemple). Ces travaux, réalisés pour l'essentiel dans le cadre d'AGENAE, concernent quatre espèces,

le bovin, le porc, la poule et la truite. Ils sont complétés par des accords d'échange internationaux avec l'USDA et différentes universités américaines. Alors que le bilan de nos contributions au travail international est très significatif chez la truite (50 %), et non négligeable chez la poule et le porc, l'arrivée récente et massive de séquences d'EST pour plusieurs espèces, incite à réfléchir sur les options futures. Au minimum, il convient de soustraire les clones déjà décrits avant de séquencer davantage. Dans certains cas, des banques spécifiques de tissus pourront être construites pour combler des fonctions éventuellement absentes ou mal représentées. Enfin, l'opportunité de se tourner vers les offres commerciales de production de puces à oligonucléotides est l'objet de réflexions.

Depuis trois ans, ces actions ont été réalisées dans le cadre du programme AGENAE qui est structuré autour des principes suivants :

- concentrer les moyens pour acquérir des données génériques sur les principaux génomes d'intérêt en élevage (cartographie génétique et physique, transcriptome), dans le contexte de consortiums internationaux visant le séquençage complet ;

- fédérer l'ensemble des équipes des différentes disciplines (génétique, physiologie, pathologie, nutrition) autour de l'usage de matériels (puces) et d'outils (transcriptomique, protéomique) pour répondre à des questions scientifiques et finalisées.

Initié au sein de l'INRA, le programme AGENAE a ensuite été ouvert à une collaboration organisée entre recherche publique et professions de l'élevage, se traduisant par la création du GIS AGENAE et le soutien du ministère chargé de la recherche (Direction de la Recherche et Direction de la Technologie), à des projets de recherche de type générique (soutien du FNS) et de type finalisé (soutien conjoint du FRT et les partenaires professionnels). Un premier appel à projets « GENANIMAL » a été lancé en avril 2003. Globalement, un budget de 50 millions d'euros sur cinq ans a été retenu, dont environ 50 % apporté par les établissements de recherche et 50 % de contributions spécifiques attribuées

sur appels à projets (5 millions d'euros par le FNS, 10 millions d'euros par les partenaires et 10 millions d'euros par le FRT). Les travaux s'appuient sur les infrastructures collectives (centres nationaux, génopoles), sur le Centre National de Ressources Génomiques Animales mis en place à Jouy en Josas, et sur des plateaux techniques installés dans les principaux centres INRA. Un dispositif système informatique unique, dérivé de celui de Génoplante, est mis en place pour l'ensemble des génomes animaux, avec un groupe de projets réunissant actuellement sept personnes.

## Variabilité génétique des phénotypes

Cette thématique regroupe une part substantielle de la capacité de recherche ainsi que l'essentiel de la capacité expérimentale. Elle implique la définition des caractères analysés et de leurs composantes, leur méthode de mesure, la modélisation statistique de leurs variations, l'estimation de leurs paramètres génétiques, la recherche de QTL ou de gènes majeurs (par des dispositifs familiaux), leur cartographie fine (généralement par déséquilibre de liaison) et l'identification des gènes impliqués et du polymorphisme causal.

Les caractères agronomiques étudiés touchent la production, la reproduction, la composition corporelle, l'ingestion et l'efficacité alimentaire, la qualité du lait et de la viande, la résistance à diverses maladies (salmonelles, coccidies, mammites, tremblante, parasites gastro-intestinaux, etc.), le comportement, les aptitudes sportives (chez le cheval) ou la coloration. Les dispositifs les plus courants sont la sélection de lignées divergentes, les dispositifs de détection de QTL (F2, BC), les dispositifs de cartographie fine de gènes. En complément du dispositif expérimental, on mobilise aussi le dispositif national de contrôle de performances, centralisé au CTIG (INRA de Jouy en Josas) en collaboration avec les professionnels de la sélection animale. Un observatoire national des anomalies génétiques a été mis en place pour compléter le dispositif. Ces différentes sources d'information sont très largement complémentaires : le dispositif

national couvrant des effectifs considérables, il permet l'estimation de paramètres à l'échelle de populations ou le criblage d'événements rares ; par ailleurs, l'expérimentation donne accès à des génotypes ou des structures de populations spécifiques, et à des phénotypes définis de façon beaucoup plus précise, et à des échantillons biologiques.

Dans certains cas, les espèces agronomiques sont des modèles biomédicaux, comme les porcs présentant une forte susceptibilité au mélanome. Ces modèles sont sources de collaborations et de liens entre l'INRA et les autres organismes de recherche.

Quelques exemples de résultats : identification du gène RN responsable d'un défaut de qualité de viande chez le porc, de plusieurs gènes d'hyperprolifération chez l'ovine, du gène PIS d'intersexualité et d'absence de cornes chez la chèvre, des différentes formes du gène PrP et de leur effet sur la résistance à la tremblante. Par ailleurs, plusieurs centaines de gènes majeurs ou QTL ont été localisés, en particulier chez le bovin, le porc, la poule et le mouton, dont les plus importants sont en phase de cartographie fine et de caractérisation.

## Gestion des populations

Enfin, au-delà des études de génomique et de génétique des caractères agronomiques, l'INRA et son département de génétique animale ont une mission dans la gestion de certaines populations d'élevage, conformément à la loi sur l'élevage de 1966. Cette mission se traduit par :

- (a) le maintien des bases de données nationales de contrôle de performances au CTIG (identification, pedigree, situation physique et performances des animaux) ;

- (b) l'évaluation génétique des reproducteurs, c'est-à-dire l'estimation de leur valeur génétique, à partir de leurs performances et celles de leurs apparentés et, de plus en plus, à partir de l'information sur les polymorphismes de l'ADN (mutations causales ou marqueurs liés) ;

- (c) le développement de méthodes de sélection et de gestion des populations (définition des objectifs de sélection, optimisation des programmes de sélection, caractérisation, de la diversité génétique, maintien de la variabilité génétique à long terme, méthodes de conservation).

## 7.3 RECOMMANDATIONS COMMUNES AUX TRAVAUX DE GÉNÉTIQUE ET DE GÉNOMIQUE DANS LE DOMAINE AGRONOMIQUE

- Poursuivre le soutien aux programmes Génoplante et AGENAE portant sur la génomique des espèces modèles et de quelques animaux domestiques et plantes cultivées d'importance majeure, à la fois pour promouvoir les travaux de recherche fondamentale et la création d'outils performants en génétique animale et en génétique et amélioration des plantes ;

- lancer des appels d'offre nouveaux pour le soutien de programmes de biologie animale et végétale utilisant les divers outils de la génomique et de la génétique, mais aussi d'autres types d'outils (analyse d'image, biochimie, etc.) et d'autres approches (en particulier les démarches de biologie des systèmes intégrés) ;

- créer de nouvelles équipes analysant des fonctions ayant une importance zootechnique ou agronomique, au sens large du terme (par exemple physiologie de la reproduction, adaptation aux stress) selon la procédure des ACI ;

- informer les enseignants de biologie, les scolaires et le grand public de l'importance de la recherche en génétique et amélioration des animaux domestiques et des plantes, et en biotechnologie ;

- soutenir les travaux de recherche fondamentaux sur les OGM animaux et végétaux, pour éviter que la France accumule du retard dans ce domaine, et définir un cadre socialement acceptable pour leur exploitation économique éventuelle.

## **8 – GÉNÉTIQUE, GÉNOMIQUE ET BIOTECHNOLOGIE**

La société finance la recherche, quel retour en attend-elle ? Répondre à cette question est une des clés du bon fonctionnement du contrat social entre le chercheur et le citoyen. La première réponse est certainement de faire des découvertes inattendues grâce à l'acquisition de connaissances nouvelles. Cette acquisition est essentielle au développement de l'humanité et elle répond – en partie – à des questions existentielles. Ces questions sont légitimes, tout un chacun les exprime plus ou moins consciemment ; les généticiens doivent revendiquer cette part de rêve (et pas seulement de cauchemar) que d'autres scientifiques en recherche fondamentale assument sans hésitation. La deuxième réponse est, sans nul doute, que la recherche doit aussi apporter des réponses concrètes : ne plus avoir faim, vivre en meilleure santé, avoir un environnement préservé pour les générations futures.

Pour ces demandes de progrès économique, il est possible d'identifier une contribution de la recherche en génétique. Pourtant, ici comme dans d'autres domaines, la France parvient difficilement à suivre le rythme imposé par les énormes investissements publics et privés des États-Unis ou d'autres pays européens. Il n'est pas question dans ce chapitre de rediscuter de la pertinence à financer de la recherche fondamentale, ou appliquée, dans des institutions de recherche académique ou des entreprises privées, mais de souligner les points spécifiques à la génétique ou à la génomique. En effet, il y a une réelle tendance des pouvoirs publics à sous-estimer les enjeux de la génétique dans le domaine industriel (industrie pharmaceutique, agroalimentaire, énergétique et environnementale), en raison du souci de garder un profil bas sur des questions sensibles (par exemple O GM, clonage animal et humain) et aussi tout simplement par méconnaissance du potentiel technique de la génétique. Les points essentiels, développés ci-dessous, sont que :

- (i) les applications de la génétique existent déjà ;
- (ii) elles voient le jour dans des entreprises de toute taille et profil ;
- (iii) elles découlent d'une articulation réussie entre recherche académique et recherche privée.

Enfin, il faut noter que la génétique, avec l'apparition de la génomique (dans le sens étude à grande échelle d'un grand nombre d'objets biologiques, gènes, protéines ou autres), est le premier secteur de la recherche en biologie où l'on voit apparaître une industrialisation de la recherche fondamentale.

### **8.1 APPORTS DE LA GÉNÉTIQUE DANS LE DOMAINE DES APPLICATIONS**

La génétique et la génomique ont déjà des retombées potentielles dans de nombreux secteurs industriels :

- pharmaceutique, par exemple souches mutantes ou transgéniques pour la production de vitamines ou d'antibiotiques, variants viraux pour des vaccins, modèles animaux pour des maladies, gènes de virulence dans les bactéries pathogènes comme cibles thérapeutiques, validation de cibles pharmacologiques sur récepteurs clonés ;
- agro-alimentaire, avec la production de variants génétiques sélectionnés pour de nombreuses espèces végétales (ici, nous parlons de sélection de variants naturels). On peut aussi évoquer la recherche de combinaisons génétiques optimales chez les animaux de boucherie ;
- procédés industriels proprement dits avec la fabrication par des microorganismes génétiquement modifiés de produits biochimiques entrant dans la composition de produits industriels (lessives, produits aromatiques), ou pour le traitement de déchets (sélection de souches bactériennes pour détoxifier des

boues industrielles). L'usage d'organismes génétiquement modifiés dans une filière alimentaire devrait aussi être débattu dans un cadre plus sérieux et moins polémique qu'aujourd'hui.

Ces secteurs sont à l'affût d'innovations. À l'heure actuelle, ces marchés sont toujours en forte croissance. Les innovations sont recherchées, même sur du long terme (avec développements prévus des produits à 10-15 ans) et sur des paris scientifiques risqués (haute valeur ajoutée).

## 8.2 APPLICATIONS GÉNÉTIQUES DANS TOUS LES TYPES D'ENTREPRISES

Les innovations par la génétique arrivent dans tous les types d'entreprises :

- les grands groupes internationaux (par exemple Procter & Gamble, laboratoires pharmaceutiques divers, Monsanto ou DuPont) ayant des départements de recherche qui font de la génétique. Les politiques affichées à court terme par ces groupes, parfois pour des raisons liées au marché boursier, ne doivent pas faire perdre de vue les grandes options stratégiques à long terme qui intègrent toujours de la recherche amont en génétique ;

- les sociétés établies dans des créneaux niches qui ont émergé, surtout aux États-Unis, au cours des 20 dernières années. À partir de la révolution de la biologie moléculaire, des sociétés se sont créées et ont très vite développé en interne et/ou en partenariat avec des chercheurs du secteur académique de la recherche en génétique de qualité. Chiron ou Amgen sont des exemples de telles sociétés aux États-Unis. Serono, leader en production de protéines recombinantes, cherche à s'inscrire dans ce schéma avec l'acquisition de Genset ;

- les nouvelles sociétés de type start-up, créées à partir de nouveaux concepts biologiques ou de méthodologies expérimentales

qui font appel à de sérieuses connaissances en génétique. Dans un milieu de très grande compétitivité, les seules sociétés qui survivront seront celles qui, au-delà d'une innovation technologique, sauront intégrer une recherche amont interne ou avec des partenariats académiques de très haut niveau. La génomique réduite à de la cuisine de biologie moléculaire mâtinée de bioinformatique de salon ne permettra pas à ces sociétés de développer en interne ou de proposer à des partenaires industriels des produits innovants. On assistera à des feux de paille associés à des coups boursiers.

## 8.3 ARTICULATION ENTRE RECHERCHE ACADÉMIQUE ET INDUSTRIELLE

Les gènes (et produits dérivés) peuvent être des outils et les données de la génétique peuvent donner des informations pour des décisions sur des développements industriels. Cet enjeu est bien perçu dans de nombreux pays, l'articulation entre recherche académique et sociétés industrielles existe déjà, la biologie/génétique se trouvant très souvent en amont des phases de développement que les entreprises maîtrisent en interne (citons, à l'échelle européenne, le Royaume-Uni, les Pays-Bas et la Suisse). Dans les pays où cette articulation fait défaut, et la France fait malheureusement partie de ceux-là, la capacité des entreprises à créer des produits innovants est faible et le milieu de la recherche académique ne forme pas (ou mal) des chercheurs qui auraient comme vocation à transformer en produit une découverte scientifique. C'est clairement le problème de la frontière (pris dans le sens *lieu de passage* et non *mur de séparation*) entre découverte et invention qui est posé.

## 8.4 INDUSTRIALISATION DE LA RECHERCHE FONDAMENTALE EN GÉNÉTIQUE

Enfin, la génétique, avec l'arrivée de la génomique, est le premier domaine de la recherche en biologie où la dimension industrielle a été prise en compte (robotique, automatisation, grand nombre d'échantillons, etc.) y compris dans les structures de recherche fondamentale. Cette dimension existait déjà dans le monde médical (analyse médicale à grande échelle avec robots de biochimie et/ou sérologie). Mais les chercheurs en biologie ignoraient souvent cette culture du grand nombre d'échantillons ou des robots. Dans de nombreux centres de recherche académique, génétique et génomique sont maintenant clairement associées et complémentaires pour des projets de recherche fondamentale. En France, les centres de recherche fondamentale qui bénéficient de plates-formes à haut débit d'un niveau national ou international sont rares. Citons le Centre National de Séquençage et le Centre national de Génotypage à Évry, la Clinique de la Souris à Strasbourg, et une plate-forme émergente de protéomique au CEA à Grenoble. L'analyse de ces diverses structures, en France comme à l'étranger, démontre que le succès de ces plates-formes passe par l'excellence scientifique des équipes qui en sont les animatrices. Le mythe de plates-formes technologiques performantes qui seraient dirigées indépendamment de projets scientifiques évalués et reconnus comme excellents doit être dénoncé : le coût de ces structures, en général très élevé, et l'obligation d'une action dans la durée (5 à 10 ans, au minimum) justifient la mise en place de mécanismes d'évaluation précis et performants.

## 8.5 RECOMMANDATIONS

Au-delà du soutien direct que les pouvoirs publics peuvent ou non apporter aux développements des entreprises industrielles en France

et qui relève d'un choix politique à assumer (maintien d'emplois industriels, création d'entreprises, etc.), il faut permettre à la recherche française en génétique d'être compétitive par rapport à ses voisines en Europe, sur le continent américain et en Asie. Les actions principales sont à mener sur le long terme et ne peuvent donc pas s'inscrire dans des effets d'annonce budgétaire. Il faut partir du constat que, dans le domaine de la génétique et de la génomique, les acteurs mondiaux se trouvent à la fois parmi les institutions de recherche académiques et dans les entreprises, françaises ou non, avec des liens variés entre ces structures, créés sur la base d'intérêts respectifs bien compris :

- cofinancement de plates-formes technologiques et de développements d'outils méthodologiques ;

- création de projets scientifiques mixtes où les cultures scientifiques se nourrissent respectivement ;

- identification soigneuse des retombées possibles des projets, tant sur le plan des publications (notoriété, visibilité scientifique) que sur la création de propriété ou produits industriels avec la reconnaissance de la contribution de chaque partenaire.

En France, pour réaliser ces objectifs il faudrait soutenir ou imaginer des initiatives telles que :

- laboratoires associés publics-privés. Ici nous nous référons à des structures, observées à l'étranger (Royaume-Uni, Japon par exemple) où cohabitent des chercheurs travaillant sur des sujets de recherche fondamentale et finalisée. Les chercheurs peuvent passer d'un type de projet à l'autre. Il en résulte un échange de pratiques professionnelles qui sont très fructueux. Mais cela suppose une grande rigueur dans la définition des objectifs et des moyens mis en œuvre ;

- structures de valorisation associées aux laboratoires de recherche fondamentale et aux entreprises. Cette association doit se faire au contact des équipes de recherche, avec une compétence mixte de nature industrielle et scientifique au sein de la structure de valorisation. La



traduction en valeur économique d'un résultat de recherche requiert un réel savoir-faire qui ne s'invente pas ;

- passerelles entre structures industrielles et de recherche académique pour des chercheurs et techniciens sur des missions à durée déterminée (1, 2, ou même 5 ans) ;

- aide à la création d'entreprises basées sur une expertise scientifique et technique permettant d'établir un métier dont on mesure le succès par la création d'emplois dans la durée et non par la seule hausse des actions en bourse sur la base d'une communication sans support concret.

Dans le domaine de la génétique et génomique, de telles initiatives pourraient très vite se mettre en place ou se développer.

En France, pour améliorer la mise en œuvre de tels projets, il faut certainement évaluer les expériences menées, soit sous la forme de partenariats bilatéraux (financés par le réseau génome par exemple), soit dans le cadre de consortia (Génoplante), soit sur la base de projets scientifiques retenus par les comités scientifiques mis en place auprès des plates-formes (CNS, par exemple).

Un dernier point mérite d'être souligné : en France la culture scientifique est disjointe de la culture technique, et l'opposition des deux conduit trop souvent à créer des structures non performantes. Ainsi il est souvent difficile de présenter un thème de recherche méthodologique dans un laboratoire de recherche fondamentale comme un sujet de thèse légitime. On reconnaît moins souvent qu'ailleurs la contribution déterminante de la levée d'un verrou technologique pour répondre à une question scientifique.

La responsabilité de cet état de fait ne revient pas uniquement aux décideurs de l'attribution des deniers publics, mais révèle une méconnaissance en profondeur des mécanismes de création scientifique et d'innovation. Cet état de fait s'enracine dès le système éducatif. La culture scientifique, particulièrement en biologie, s'acquiert encore trop peu au niveau des collèges et des lycées et est

trop souvent absente chez un grand nombre de décideurs à tous les niveaux. Par ailleurs, l'introduction de l'enseignement de la technologie au collège, à côté de la physique, de la chimie et des sciences de la vie, n'a pas répondu à la demande : on a créé de nouvelles frontières entre savoirs, au lieu d'insister sur la synergie entre eux et la fertilisation croisée possible entre ces champs de connaissance. Dans le domaine de la génétique et de la génomique, cet aspect culturel est particulièrement criant : c'est le mélange subtil entre cultures scientifique et technique qui a créé de la synergie entre découverte scientifique et invention d'un produit, par exemple dans le cas du séquençage de l'ADN ou des kits de diagnostic pour des agents viraux émergents. C'est l'absence de liens entre science et technologie qui freine la création et interdit le fonctionnement pérenne de structures industrielles innovantes pratiquant réellement la recherche et le développement dans notre pays. Génétique et génomique ont des atouts, ensemble, pour servir de chantier pilote pour améliorer cette situation.

## 9 – GÉNÉTIQUE ET SOCIÉTÉ

La politique scientifique nationale est clairement devenue un enjeu majeur dans les pays développés et les gouvernements se chargent de façon de plus en plus évidente d'en définir les objectifs par rapport à la société. Les citoyens financent directement la recherche scientifique publique, et indirectement la recherche privée par les crédits d'impôts, et l'élévation du niveau culturel moyen des électeurs leur permet de demander aux scientifiques un retour sur leur investissement. Or, les coûts sont devenus considérables, parce qu'en plus de la recherche universitaire, liée à l'appréciation positive de la formation des jeunes par la recherche, dans tous les pays développés existent des agences dédiées exclusivement à la recherche, surtout avec des visées applicatives. D'autre part, les méthodes d'études sont devenues très complexes, sophistiquées et chères.

Cette formulation des rapports entre chercheurs et société peut apparaître brutale, mais c'est souvent dans ces termes que les questions posées par le financement de la recherche universitaire ou des EPST (Entreprises Publiques Scientifiques et Technologiques) sont discutées actuellement en France. Par exemple, les tutelles ministérielles demandent aux chercheurs et aux enseignants-chercheurs du secteur public d'accepter l'évaluation rigoureuse de leurs travaux de recherche, pour justifier leurs salaires et leurs financements. Il est intéressant alors d'examiner les critères utilisés pour l'évaluation, puisqu'ils n'indiquent pas seulement les standards que se donne la science, mais aussi les obligations que la société pose aux scientifiques. D'autre part, de plus en plus les recherches sont financées sur des contrats établis entre le bailleur de fonds et le scientifique, qui voit donc son activité orientée par le besoin d'obtenir l'argent nécessaire pour ses expériences.

Finalement, les parlements nationaux et le parlement européen sont sollicités de plus en plus souvent pour statuer sur des problèmes provoqués par l'avancée des résultats scientifiques. Après les applications de la physique moderne dans le domaine militaire, ce sont bien les avancées de la biologie, et de la génétique en particulier, qui inquiètent aujourd'hui de nombreux électeurs dans les pays développés.

Notre analyse des interactions entre la recherche en génétique et notre société doit alors s'inscrire dans le cadre global de la vision sociale actuelle de la science et des volontés de pilotage de ses orientations qui se manifestent dans les pays développés.

Il convient donc de s'interroger de façon générale sur les questions que la société contemporaine pose aux scientifiques, puisque leur identification aidera à comprendre les normes actuelles qui règlent le contrat social entre les chercheurs et leurs concitoyens. Dans nos sociétés, il apparaît que la demande principale actuelle soit celle de vivre mieux et plus longtemps.

Ce concept d'apparence simple est en réalité très complexe, puisque la définition d'une *bonne* vie est différente selon les

époques, les civilisations et la conception qu'on a de la condition humaine. De plus, prolonger la vie n'a d'intérêt dans nos sociétés que si l'état général, et l'état mental en particulier, sont maintenus à un niveau satisfaisant. Malgré ces complications, c'est bien l'exigence d'une vie globalement meilleure qui est à l'origine du développement des fondements et des applications de la médecine, de la biologie, de la chimie et de la physique modernes, mais aussi des sciences humaines et sociales.

Si l'on se limite à ce constat, on pourrait avoir l'impression que la seule raison d'intérêt des citoyens des pays occidentaux pour la science se trouve dans ses applications pratiques. Or, la situation est plus complexe. En effet, d'une part les applications de la science ont parfois généré des interrogations, voire des révoltes parmi les citoyens et, d'autre part, la demande scientifique sociale s'adresse aussi avec grande force à des problèmes dont les solutions n'apparaissent pas immédiatement utiles dans la vie de tous les jours.

Pour commencer par ce dernier point, il faut souligner que l'astrophysique, les études interdisciplinaires sur l'origine de l'espèce humaine, l'archéologie, l'histoire des religions, l'histoire de l'art, la philosophie, la psychologie, etc., suscitent un grand intérêt dans le public, sans que leur soit associée la possibilité d'une application immédiate. Il apparaît donc que la société se pose des questions très fondamentales sur notre origine, notre nature, notre place et notre destin dans un univers complexe et immense. Or, pour répondre à ces interrogations, le public a besoin de l'aide des chercheurs et il est donc prêt à subventionner leurs recherches. À partir de cette observation, on peut aussi souligner qu'il semble accepté par l'opinion que la capacité de notre espèce à s'interroger sur les questions scientifiques fondamentales constitue un élément central des moyens d'analyse et de représentation qui ont contribué à l'évolution de notre espèce.

Depuis que la science rationnelle, que nous pratiquons aujourd'hui, s'est développée, quelques exemples majeurs qui ont marqué notre histoire par leur ampleur et leur non

utilitarisme immédiat, peuvent être cités. L'un est celui des observations de Copernic et Galilée qui sortirent la terre du centre de l'univers. Les montres n'ont pas été changées pour autant, et le soleil se lève toujours à l'est. Un autre est l'hypothèse évolutive de Darwin qui sortit l'homme du centre de la création. L'homme n'en fut ni meilleur ni plus mauvais. Ces deux événements ébranlèrent la société, et si l'univers galiléen nous est aujourd'hui familier, l'idée darwinienne, qui postule que l'homme, malgré la culture qui le distingue du reste des êtres vivants, est de même nature que l'animal, le végétal ou le microbe, n'est toujours pas assimilé réellement dans notre société.

Le bilan d'un siècle de génétique et ses perspectives d'avenir doivent donc être intégrés dans ce contexte d'intérêt et parfois de résistance profonds du public pour les réponses apportées par une discipline scientifique. Il serait simpliste de nous concentrer exclusivement sur les possibilités nouvelles offertes par les applications de notre discipline à la solution de problèmes d'intérêt immédiat. L'histoire de la science montre bien que les retombées les plus extraordinaires et inattendues sur le plan pratique sont venues de recherches fondamentales, motivées par la seule curiosité du scientifique et sa capacité à se poser des questions originales et à en chercher les réponses de façon rigoureuse et sans préjugés ni contraintes sociales.

D'autre part, le renvoi obsédant de certains courants de pensée à l'idée que la science doit s'attaquer à satisfaire en priorité les besoins immédiats de l'humanité a induit certains groupes d'opinion à définir la recherche scientifique comme le moteur de l'exploitation irréflective et égoïste des ressources. Or, des exemples récents tragiques de l'histoire humaine confirment la nécessité d'une réflexion très large sur les rapports entre la science, ses applications et leur utilisation par les sociétés humaines.

Nous vivons une période d'une telle fécondité scientifique en biologie qu'il est de notre devoir d'apporter à notre société des éléments nouveaux de réflexion sur notre

présent et notre futur, faute de quoi des réactions hostiles à la science peuvent s'amplifier. Nos concitoyens sont conscients que des découvertes très importantes ont déjà été réalisées en génétique et que d'autres vont sans doute suivre. Cependant, la génétique est surtout assimilée à certaines de ses applications, comme la lutte contre le cancer, les organismes génétiquement modifiés (OGM), les techniques de procréation médicalement assistée, le diagnostic prénatal de maladies génétiques, les manipulations génétiques de toutes sortes et en fin de compte le clonage humain. Toutes ces applications sont loin de faire l'unanimité dans le public, bien au contraire certaines suscitent même des réactions de franche hostilité. Laisser identifier l'étude des gènes et des génomes avec ces applications serait donc non seulement simpliste et erroné, mais aussi dangereux parce que cela pourrait contribuer à susciter une vague d'opposition à la recherche scientifique en général.

Nous avons donc la responsabilité de convaincre nos concitoyens que nos recherches contribuent à répondre aux questions fondamentales qui sont posées par la société depuis toujours, et qu'elles ne peuvent pas être résumées seulement par ses applications. Cependant, il faut aussi courir le risque d'expliquer que certaines applications ne sont pas forcément mauvaises et que ce n'est pas parce qu'elles suscitent des interrogations nouvelles qu'il faut en avoir peur. Une société démocratique avancée devrait être capable, en y mettant l'énergie et le temps nécessaires, d'évaluer avec ses membres l'impact social des avancées des sciences, sans ouvrir la porte à des réactions irrationnelles.

Dans les domaines de recherche couverts par les groupes qui étudient les structures, les fonctions et les régulations des gènes et des génomes, plusieurs axes complémentaires peuvent être définis pour les prochaines années, qui vont de la physiopathologie de l'ADN à la reconstitution de l'évolution moléculaire qui a mené à la vie telle qu'on la connaît aujourd'hui, à la naissance de l'homme, à la complexité cellulaire et à la variété des formes de vie dans notre planète (et ailleurs ?).

L'étude de l'organisation et du fonctionnement des gènes et des génomes est loin d'être accomplie. L'identification de la séquence du génome humain a sans doute marqué les esprits : cependant même le public commence à se rendre compte que ce n'est qu'une étape dans la compréhension du fonctionnement de la molécule qui organise le fonctionnement et l'hérédité de la plupart des organismes, c'est-à-dire l'ADN. La connaissance du rôle de l'ADN a fait son chemin dans la culture de base grâce au bon système d'études secondaires français en même temps que la notion que des aberrations de sa structure ou de son fonctionnement provoquent des maladies. La constitution de puissantes associations de familles de patients atteints de maladies génétiques, associations qui sont capables de mobiliser les médias, de récolter de grandes sommes d'argent, d'influencer les politiciens et d'orienter la recherche biomédicale, a contribué à propager la notion du rôle central joué par l'ADN et celle des dangers de ses aberrations. Malheureusement, ces élans ont contribué à alimenter des faux espoirs de guérisons rapides liées à l'utilisation de thérapies géniques. Les déceptions de ces espoirs trop optimistes, mais aussi les progrès réels qui ont été réalisés par la mise en jeu de grandes sommes d'argent, devraient convaincre l'opinion publique de la nécessité de développer les recherches fondamentales en génétique, en particulier sur le métabolisme de l'ADN. D'autre part, la lutte contre le cancer a été déclarée grande cause nationale en France. Elle passe nécessairement par la prévention, le dépistage et les soins médicaux mais aussi par la compréhension du rôle central joué dans son déclenchement par les dysfonctionnements de l'ADN. L'effort des généticiens devrait se concentrer sur l'explication que des applications utiles contre les maladies génétiques et le cancer ne peuvent venir que d'une recherche fondamentale bien soutenue financièrement et d'excellente qualité.

L'étude des causes génétiques de la complexité cellulaire est présente dans tous les projets destinés à comprendre le fonctionnement intégré d'une cellule. Jusqu'ici nous avons démonté ce microcosme en

sous-système (mitochondrie, noyau, chaînes de biosynthèse, etc.). L'articulation entre ces sous-ensembles doit être abordée plus finement. Il s'agit à la fois des aspects de flux de l'information présente dans le génome (signalisation, codage, rétroactivité, etc.), de la dynamique de l'architecture fonctionnelle cellulaire (assemblage et désassemblage) et de l'organisation spatio-temporelle des éléments (mouvements des protéines, des organites). Un nouveau concept commence aussi à émerger : comme toute organisation vivante, de la molécule biologique à la biosphère, une cellule n'a pas de déterminisme absolu, elle est le fruit d'une histoire. Chaque cellule naît, vit et meurt. À chaque instant où nous la voyons, elle est à un moment de son histoire unique, héritière de son propre passé individuel et devant un avenir qui est inconnu de par le fait même des événements stochastiques qui surviendront. L'approche de la cellule unique sera sans doute un chemin plein de surprise et les traces de son histoire laissées dans son patrimoine génétique sont à la fois causes de dysfonctionnement mais aussi matériels utilisés par l'évolution.

L'étude de la variété de la vie tient de beaucoup plus près des questions d'identité de l'homme et de sa place sur la terre. En plus de l'étude de son organisation, qu'on a déjà discutée, un aspect plus étrange de la cellule est sa « robustesse » dans le sens scientifique, c'est-à-dire son pouvoir de continuer son chemin en dépit des heurts et accidents, des variations faibles mais incessantes de son environnement. Un autre niveau d'analyse est celui des différences entre individus à l'intérieur d'une espèce donnée. Ces variations peuvent être d'amplitude variable, chaque espèce étant marquée fortement par son histoire : expansion, heurts avec des pressions de sélections fortes ou pas, constantes ou variables, étranglement des effectifs de population, etc. Enfin, il nous reste à découvrir la partie cachée de l'iceberg de la vie. Nous ne connaissons qu'environ 1 à 10 % des bactéries et la mer, refuge de la grande majorité de la variété biologique, du fait de sa taille et de sa difficile accessibilité est une réserve immense. Les résultats des analyses

génomiques et fonctionnelles qui viendront de l'ensemble de ces thèmes de recherche peuvent être très spectaculaires autant sur le plan fondamental que sociologique et appliqué.

Un autre élément de grande fascination pour le public des pays développés est celui de la mise en place des organes, en particulier le cerveau et son fonctionnement, et jusqu'aux individus complets. Dans ce même contexte s'inscrivent les recherches, plus techniques mais tout aussi fascinantes, sur les cellules souches présentes chez l'adulte. En effet, tout en posant la question fondamentale des divisions asymétriques qui permettent leur maintien, les cellules souches de l'adulte ouvrent des pistes applicatives de thérapie cellulaire alternatives à la thérapie génique et au clonage thérapeutique. On est là bien évidemment dans des domaines où seules des études interdisciplinaires pourront trouver des réponses satisfaisantes à des problèmes si complexes. Cependant, l'identification du rôle joué par le patrimoine génétique et son plan global d'expression et de régulation reste un point central de ces analyses. Sans tomber dans les excès qui ont nui à notre discipline, comme les annonces erronées de la découverte des gènes de la criminalité ou de l'homosexualité, des sujets passionnants de recherche concernent les déterminants génétiques qui ont permis à notre espèce de développer un langage complexe et la capacité d'abstraction et de représentation symbolique. Le public semble très intéressé par les études sur les composants matériels des activités cérébrales complexes, autant pour analyser les comportements rationnels que sentimentaux.

L'histoire évolutive des caractères qui nous ont rendus différents des autres animaux constitue donc un défi intellectuel qui dépasse les murs de nos laboratoires. Les études génétiques et génomiques d'espèces modèles autres que l'homme ont permis de commencer à tracer le chemin suivi par les gènes au cours de l'évolution qui, à partir de lointains ancêtres communs, a conduit à l'apparition de nouvelles formes biologiques. Le débat récurrent sur l'origine de l'espèce humaine, sur nos relations avec les grands singes, avec lesquels nous partageons tant d'information génétique, et finalement les questions politiquement très délicates sur l'existence de races humaines

et sur le sens à donner aux différences évidentes de couleur de peau ou de forme des yeux, indiquent bien la portée profonde dans l'imaginaire du public de recherches de génétique qui n'ont pas pourtant d'applications immédiates.

Finalement, la connaissance de l'évolution des espèces sur la terre (naissance, apogée, disparition), la capacité de notre espèce de modifier les conditions de sélection grâce aux avancées technologiques, et l'énorme accélération des connaissances en biologie posent des questions extraordinaires qui touchent au fondement même de la notion d'humanité. Par exemple, le débat sur le clonage thérapeutique conduit à des réponses différentes dans différents pays : cela indique bien que les problèmes posés sortent du strict cadre scientifique, qui est le même partout, pour assumer des dimensions philosophiques et idéologiques qui sont propres aux diverses cultures qui peuplent notre planète. Dans une période de respect des différences, il est difficile d'imposer une opinion ou une ligne de conduite à l'ensemble des pays capables de réaliser le clonage humain. D'autre part, des énormes intérêts économiques sont en jeu et on ne peut pas exclure que même dans des pays culturellement hostiles au clonage thérapeutique la pression des forces financières ne conduise pas à un changement de l'opinion publique. L'impact profond de ces thèmes sur l'imaginaire collectif est prouvé par la déclaration d'un mouvement religieux sectaire d'avoir réalisé le clonage reproductif humain. Sans donner trop d'importance à de telles déclarations, dont les motivations ne sont pas scientifiques, on ne peut pas passer sous silence l'obtention, publiée en mars 2004 par un groupe sud-coréen (Hwang et al. « Évidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst », *Science*, vol.303, 1669-1674, 12 mars 2004), d'un embryon humain obtenu par insertion d'un noyau dérivé d'une cellule différenciée dans un œuf humain énucléé. Le but de cette expérience semble limité à l'étude des conditions expérimentales nécessaires à sa réalisation et à la production de cellules souches embryonnaires utilisables en thérapie, c'est-à-dire au clonage thérapeutique. Cependant, le pas a été franchi, et pour la première fois la construction d'un

clone humain capable de dépasser le stade de 8 à 10 cellules a été documentée au sein de la communauté scientifique.

Même si la perspective de la modification programmée du patrimoine génétique de notre espèce est une perspective qui semble encore lointaine, il serait imprudent de sous-estimer la rapidité de l'accumulation des données biologiques et celle des évolutions technologiques. Déjà une partie de nos concitoyens s'interroge sur la possibilité d'un nouvel eugénisme scientifique généré par les résultats de la génétique. Ce débat ne doit pas être escamoté. Le diagnostic pré-implantatoire est déjà utilisé pour sélectionner l'embryon, et les tests génétiques pour des embryons conçus naturellement seront de plus en plus « faciles ». Seulement la mise en place rapide d'une confrontation ouverte et approfondie des différents points de vue scientifiques, la discussion de tous les aspects éthiques, et l'implication de toutes les composantes qui régissent l'existence d'une société démocratique, permettront aux chercheurs et aux citoyens d'avoir un rôle actif dans ce débat fondamental.

## **10 – QUELQUES RECOMMANDATIONS POUR UN PROGRAMME INTERDISCIPLINAIRE POUR LE DÉVELOPPEMENT DE LA RECHERCHE SUR LES GÉNOMES**

Le Ministère de la Recherche, les organismes de recherche, l'ensemble des tutelles qui ont en charge la recherche, devraient se préoccuper de manière urgente de l'état du financement de la génétique et de la recherche sur les génomes en France. Un retard s'accumule depuis des années par rapport aux autres pays. Il est surprenant et inexplicable de constater qu'en ce début du XXI<sup>e</sup> siècle, alors que la génétique

est citée partout dans les médias, qu'elle suscite tant d'espoirs mais aussi beaucoup d'interrogations, alors que les grandes nations scientifiques engagent des moyens considérables dans la recherche sur les génomes, aucun programme dédié au soutien à la génétique n'existe dans notre pays. À terme, c'est l'une des activités essentielles au développement des sciences de la vie qui risque de s'étioler d'une façon qui pourrait être irréversible.

Un programme interdisciplinaire pour le développement de la génétique et de la recherche sur les génomes devrait permettre de contribuer à la structuration de la communauté scientifique travaillant en France sur ces problèmes, et soutenir efficacement les équipes travaillant sur des sujets novateurs dans des thèmes concernant :

- l'analyse globale de l'expression des génomes et de sa régulation ;
- les régulations génétiques et épigénétiques de l'expression des génomes ;
- les approches génétique et structurale de l'étude des ARN à fonction régulatrice ;
- le développement intégré des approches génétique et cellulaire pour l'analyse des phénotypes ;
- la machinerie cellulaire responsable de l'intégrité, de la variabilité, et de l'évolution des génomes ;
- la création et le maintien de la diversité génétique ;
- l'évolution des molécules, des formes et des fonctions du vivant ;
- la bioinformatique pour l'analyse structurale, fonctionnelle et évolutive des génomes ;
- les aspects fondamentaux de la recherche en génétique humaine ;
- les aspects fondamentaux de la recherche en génétique animale, végétale et des micro-organismes ;
- le débat social et éthique induit par la recherche en génétique.

## ANNEXE

### ANNEXE 1 : LISTE DES AUTEURS

#### **Giuseppe Baldacci**

Directeur de Recherche au CNRS  
Génotoxicologie et cycle cellulaire (UMR2027),  
Institut Curie, Bât. 110,  
15 rue Georges Clémenceau 91405 Orsay cedex

#### **Didier Boichard**

Directeur de Recherche à l'INRA  
INRA – Département de Génétique Animale,  
Domaine de Vilvert 78352 Jouy-en-Josas cedex

#### **Michel Boulétreau**

Directeur de Recherche au CNRS  
Biométrie et biologie évolutive (UMR5558),  
Université Claude Bernard (Lyon I), Bât. G.  
Mendel et 711, 43 bd. du 11 Novembre 1918  
69622 Villeurbanne cedex

#### **Alain Bucheton**

Directeur de Recherche au CNRS  
Institut de génétique humaine – IGH  
(UPR1142), 141 rue de la Cardonille,  
34396 Montpellier cedex 5

#### **Michel Caboche**

Directeur de Recherche à l'INRA  
Laboratoire de biologie des semences, INRA,  
route de Saint-Cyr 78026 Versailles cedex

#### **Michael Chandler**

Directeur de Recherche au CNRS  
Laboratoire de microbiologie et génétique  
moléculaires (UMR5100), Centre national de  
la recherche scientifique, LMGM, 118 route  
de Narbonne 31062 Toulouse cedex 4

#### **Claude Chevalet**

Directeur de Recherche à l'INRA  
INRA, Génopole Toulouse Midi-Pyrénées,  
Chemin de Borde-Rouge, Auzeville BP 27  
31326 Castanet Tolosan cedex

#### **Michel Delseny**

Directeur de Recherche au CNRS  
Laboratoire Génome et développement des  
plantes (UMR5096), Université de Perpignan  
Bat. C, 52 av. de Villeneuve,  
66860 Perpignan

#### **Bernard Dujon**

Professeur  
Génétique des génomes (URA2171),

Institut Pasteur, Site Fernbach,  
25 rue du Docteur Roux, 75724 Paris cedex 15

#### **Annie Jacq**

Chargée de Recherche au CNRS  
Institut de génétique et microbiologie – IGM  
(UMR8621), Université Paris Sud (Paris XI),  
Bât. 400 -409 -360,  
15 rue Georges Clémenceau  
91405 Orsay cedex

#### **Pierre Legrain**

Directeur de Recherche au CNRS  
Protéines membranaires transductrices  
d'énergie (PMTE) (URA2096), Commissariat  
à l'énergie atomique, DSV/DBJC/SBFM,  
Bât. 532 – Pièce 205 91191 Gif sur Yvette cedex

#### **Arnold Munnich**

Professeur d'Université, Praticien Hospitalier  
Hôpital des enfants malades,  
149 rue de Sèvres 75743 Paris cedex 19

#### **Alain Nicolas**

Directeur de Recherche au CNRS  
Dynamique de l'information génétique :  
Bases fondamentales et cancer (UMR7147),  
Institut Curie, 26 rue d'Ulm 75248 Paris cedex 05

#### **Véronique Paquis**

Professeur d'Université, Praticien Hospitalier  
Laboratoire de Génétique, Hôpital de l'Archet 2,  
151 route St Antoine de Ginestrière BP 3079  
06202 Nice cedex 3

#### **William Saurin**

Directeur de Recherche au CNRS  
Structure et évolution des génomes  
(UMR8030), GENOSCOPE – CNS,  
2 rue Gaston Crémieux BP 5706  
91057 Évry cedex

#### **Hélène Richard-Foy**

Directeur de Recherche au CNRS  
Laboratoire de biologie moléculaire  
des eucaryotes – LBME (UMR5099), Bât. IBCG,  
118 route de Narbonne  
31062 Toulouse cedex 4

#### **Jean Weissenbach**

Directeur de Recherche au CNRS  
Structure et évolution des génomes  
(UMR8030), Genoscope – CNS  
2 rue Gaston Crémieux BP 5706  
91057 Évry cedex

