

Bertrand Jordan

Chroniques d'une séquence annoncée

1992-2002 :
DIX ANS
DE PROGRAMMES GÉNOME



**CHRONIQUES
D'UNE
SÉQUENCE ANNONCÉE**

**1992-2002 : dix ans
de programmes Génome**

Ce recueil est dédié à Axel Kahn, qui m'a « lancé » dans l'écriture, à François Flori qui, non content de suivre ces chroniques depuis des années, a veillé sur la réalisation de ce livre, et surtout à Anne, dont le patient travail de relecture critique a bénéficié à beaucoup de chroniques et à tous les commentaires qui les accompagnent.

CHRONIQUES D'UNE SÉQUENCE ANNONCÉE

**1992-2002 : dix ans
de programmes Génome**

Bertrand Jordan



Outre les *Chroniques Génomiques* et de nombreux articles de vulgarisation, Bertrand Jordan a publié « Voyage autour du Génome » (Éditions John Libbey Eurotext, 1993), fondé sur les chroniques décrivant un tour du monde qui faisait le point sur le projet Génome à ses débuts. D'autres ouvrages ont suivi, les plus récents sont « Les Imposteurs de la Génétique » (Éditions du Seuil, 2000), qui traite des déformations et déviations de la Génétique, « Le chant d'amour des concombres de mer » (Éditions du Seuil, 2002), méditation marine sur la nature, l'évolution et la condition humaine, et « Les marchands de clones » (Éditions du Seuil, 2003), mise au point sur les aspects scientifiques, économiques et sociétaux du clonage animal et humain.

Le dessin de couverture provient du livre de résumés du *Genome mapping and sequencing meeting* tenu en 1989 à Cold Spring Harbor Laboratory et est dû à James Duffy. Il est reproduit avec l'autorisation de l'auteur et de « Cold Spring Harbor Laboratory Meetings and Courses Programs ».

Éditions E.D.K.
10, villa d'Orléans
75014 PARIS
Tél. : 01 53 91 06 06

© Éditions E.D.K., Paris, 2003
ISBN : 2-84254-089-1

Il est interdit de reproduire intégralement ou partiellement le présent ouvrage – loi du 11 mars 1957 – sans autorisation de l'éditeur ou du Centre Français d'Exploitation du Droit de Copie (CFC) 20, rue des Grands-Augustins, 75006 Paris.

SOMMAIRE

PRÉFACE

Des gènes et des hommes <i>Axel Kahn</i>	8
---	---

1. ENTRÉE EN MATIÈRE	11
<i>Avant-propos</i>	12
Années Génome : les dix glorieuses.....	13

2. L'ÈRE DES CARTOGRAPHES	19
Des cartes en voie d'intégration ?.....	20
Généthon : la réussite d'un pari.....	23
Carte physique du génome humain : l'état des lieux.....	29
Fugu Story.....	36

3. LE DÉTOUR PAR LES ADN COMPLÉMENTAIRES	41
Le festival des ADNc.....	42
La valse des étiquettes.....	51
Génome humain : l'annuaire nouveau est arrivé.....	57
ADNc : les incontournables.....	62
<i>ADNc ou ADN génomique ?</i>	67

4. COORDINATION OU CONCURRENCE ?	69
Les contradictions du génome.....	70
Génome : la caverne d'Ali Baba... ou le supplice de Tantale ?.....	79
Les HGM se suivent et ne se ressemblent pas.....	85
Du Programme Génome à la « Pharmacogénomique ».....	91
Génome : les méandres de la technologie.....	96

5. AUX QUATRE COINS DU GLOBE	103
Génome au Japon : au-delà des mythes.....	104
USA : un programme génome solidement installé.....	109
Grande-Bretagne : un programme Génome à dimension humaine.....	113
Où en est le Programme Génome russe ?.....	119
Allemagne : enfin un programme Génome humain.....	127
Génome : quand la Chine s'éveillera... ..	131

6. HÉSITATIONS HEXAGONALES	135
Flash	136
Génome français : de grandes espérances	137
Très Grand Séquençage : trompe-l'œil politique ou nécessité scientifique ?.....	143
<i>La belle et triste histoire du « Programme Génome Français »</i>	146
7. L'IRRUPTION DES PUCES	149
Voyage au pays des puces.....	150
Jusqu'où iront les puces ?.....	159
Puces à ADN : les brevets contre le progrès ?	165
Puces-actualités.....	171
<i>Puces « recherche », puces « cliniques » et « labopuces »</i>	177
8. UNE SÉQUENCE TANT ATTENDUE...	179
Les heurs et malheurs du séquençage d'ADN à grande échelle.....	180
Séquençage génomique : le deuxième souffle	184
Un compte d'apothicaire	190
9. EN GUISE DE CONCLUSION...	199

PRÉFACE

DES GÈNES ET DES HOMMES

Il est rare dans l'histoire qu'un même événement eut été annoncé par trois fois, à plusieurs années d'intervalle, simultanément par les Chefs d'États et de Gouvernements des pays dominants sur le plan scientifique et technique. Tel est le cas du séquençage du génome humain dont les résultats très avancés, puis presque complets, puis enfin pratiquement complets furent présentés en grande pompe à la presse en 2000, le 14 et 15 février 2001 et, enfin, en avril 2003 afin de coïncider avec le cinquantième anniversaire de la découverte de la structure en double hélice de l'ADN. Chaque fois, le caractère exceptionnel de cette médiatisation institutionnelle internationale fut justifié par le caractère symbolique, presque métaphysique, du déchiffrement par l'homme de son programme génétique, le « *blueprint of human life* » et par les extraordinaires bienfaits que ne saurait manquer d'entraîner cette performance technique et scientifique.

Bien avant cette débauche de dithyrambes et de spéculations, Bertrand Jordan fut un observateur avisé, lucide et distancié de l'entreprise, aussi bien dans ses dimensions techniques que politiques, économiques et éthiques. À la fin des années 1980, il avait entrepris une reconversion thématique de l'équipe qu'il dirigeait au Centre d'Immunologie de Marseille-Luminy. Spécialiste des gènes codant les protéines du complexe majeur d'histocompatibilité chez l'homme, thème en parfaite cohérence avec les priorités d'un Institut d'Immunologie, Bertrand Jordan avait décidé de se consacrer plus à la génétique en elle-même. Il s'intéressait, en particulier, à la recherche par clonage positionnel de gènes de maladies génétiques à une époque où le premier grand succès en ce domaine, l'identification du gène modifié dans la myopathie de Duchenne, venait d'être publié. Un autre témoin de ce bouleversement scientifique était la revue *médecine/sciences*, créée à la suite d'une initiative conjointe franco-québécoise, en 1985. C'est au titre de membre fondateur et de Rédacteur en Chef de *médecine/sciences* que Bertrand Jordan me sollicita afin de me présenter ses projets. Il envisageait en particulier de prendre une année sabbatique afin de pouvoir visiter, tout autour du monde, les principaux laboratoires engagés dans la « nouvelle recherche en génétique », ce que l'on allait bientôt appeler la génomique. C'est alors qu'émergea l'idée d'une collaboration régulière entre Bertrand Jordan et *médecine/sciences*, sous la forme de « chroniques génomiques » pour lesquelles j'inventais un logo évocateur, celui d'un individu stylisé arpentant à grands pas la double hélice d'ADN. Ces chroniques débutèrent avant que le tour du monde des grands laboratoires du domaine ne fût entrepris, mais servirent à accueillir, à chacune des étapes de celui-ci, les billets du Docteur Jordan. Ce voyage d'études se situait juste entre deux épisodes très significatifs qui firent entrer les études du génome dans le champ de la « *big science* », c'est-à-dire celui d'une étude à grande échelle systématisée, optimisée et robotisée. Il s'agit du séquençage systématique des ADN complémentaires par Craig Venter, alors chercheur des NIH américains, puis de l'exploit, en 1992, de l'équipe de Daniel Cohen et Jean Weissenbach au Généthon d'Évry. Grâce à Bertrand Jordan, *médecine/sciences* devint donc un outil privilégié de l'éveil du monde scientifique francophone à ce nouveau cours des choses. Il s'agissait en fait pour *médecine/sciences* d'une fonction logique puisque la revue avait été créée en 1985, l'année même de la publication dans *Nature* de l'identification formelle par l'équipe de Louis Kunkel, aux États-Unis, du gène muté ou délété dans la myopathie de Duchenne. C'est d'ailleurs grâce à la prise de

conscience par le directeur de l'Association Française de Lutte contre les Myopathies de l'époque, Bernard Barataud, du pouvoir de la génétique à la suite de cet exploit des américains que l'AFM s'engageait dans l'aventure du Téléthon et contribuait alors à écrire l'histoire de la génomique. L'actuel ouvrage de Bertrand Jordan commence au terme de ces prémisses, alors que se pose la question des étapes suivantes. C'est donc à un compte rendu détaillé des événements et péripéties ayant conduit au séquençage du génome humain qu'est convié le lecteur. Ce dernier pourra se familiariser avec l'évolution des stratégies expérimentales mises en œuvre, se plonger dans les arcanes de la compétition scientifique et industrielle, s'initier aux subtilités des relations complexes entre le monde scientifique et politique, prendre conscience, enfin, des incertitudes qui persistent quant à la signification réelle de ce long enchaînement d'un peu plus de trois milliards de paires de bases. Celles-ci sont en particulier illustrées par les débats actuels sur la diversité de la séquence du génome dans l'espèce humaine, et sur le nombre de gènes. Craig Venter, dirigeant l'effort de séquençage de la Société *Celera Genomics* publié le 15 février 2001 dans la revue *Science*, commenta les résultats obtenus sur un petit échantillon d'ADN appartenant à des personnes d'ethnies différentes. L'échantillon « caucasien », c'est-à-dire le sujet blanc d'origine européenne, s'avéra être le sien. On ne peut cependant en déduire que Monsieur Venter « se connaisse » pour autant mieux que le commun des mortels, marquant les limites « programmatiques » de la séquence... Quoiqu'il en soit, Craig Venter insista longuement sur le fait que la très grande homogénéité des séquences d'ADN de sujets asiatiques, africains ou européens disqualifiait les thèses racistes. En d'autres termes, la réfutation philosophique des discours inégalitaires fondant les thèses racistes pouvait maintenant laisser place à l'évidence scientifique offerte par le séquençage du génome qui devenait ainsi, vraiment, une œuvre d'une profonde dimension morale. Malheureusement, on pourrait aussi bien indiquer que l'étude du génome démontrait aussi qu'existaient en moyenne entre des individus jusqu'à 3 millions de différences... Lorsque l'on connaît les effets phénotypiques possibles d'une mutation unique, il y a bien évidemment là de quoi largement laisser place à tous les délires. D'ailleurs, nul ne conteste que les quelque 1,4 % de différences entre les génomes d'*Homo sapiens* et de *Pan troglodytes* aient des conséquences phénotypiques notables.

L'autre observation dérivée de l'analyse du génome humain et présentée comme une bouleversante surprise fut que l'on n'y pouvait détecter qu'un petit nombre de « gènes », de l'ordre de 25 000 à 30 000. Ce décompte, non encore définitif, repose avant tout sur l'identification des unités de transcription conservées entre espèces et ayant la capacité de coder des protéines. La stupeur déclenchée par cette nouvelle aussi bien que sa signification biologique méritent d'être commentées. Sur le premier plan, la métaphore d'un programme génétique vu comme un langage dont le gène constituerait les mots, et dont la signification serait ainsi largement combinatoire, conduit à n'être nullement surpris de ce qu'une régulation différente de gènes pratiquement semblables en nombre et très voisins en séquence puisse aboutir à des phénotypes extrêmement différents. Quant à l'aspect quantitatif lui-même, tout dépend à l'évidence de ce que l'on entend par « gènes ». Dans le sens des « facteurs héréditaires » de Gregor Mendel, un gène doit correspondre à tout élément transmissible du génome susceptible d'influer le phénotype. À ce titre, faut-il parler d'un ou de plusieurs gènes lorsqu'une même unité de transcription peut, par des phénomènes d'épissages alternatifs multiples ou d'utilisation de promoteurs optionnels, engendrer des transcrits et des protéines de structure et de fonction fort différentes ? Et comment considérer tous les éléments du génome qui sont transcrits, relativement conservés au cours de l'évolution mais non codants ? Si seul 1,6 % du génome des mammifères semblent avoir le potentiel de coder pour des protéines, les régions transcrites et conservées pourraient être dix fois plus étendues. Parmi elles, nul doute que figurent, en particulier, des ARN interférants, des micro-ARN ou d'autres espèces encore dont on sait qu'elles peuvent finement moduler l'expression, voire la structure du génome à différents niveaux : réarrangement génique, méthylation, trans-

cription, stabilité des transcrits, traduction... Or, chacune des unités transcriptionnelles codant pour de telles espèces d'ARN régulateurs peut avoir un effet phénotypique caractérisé et, par conséquent, mérite alors d'être appelé « gène » au sens initial du terme. Ainsi, suivant la convention à laquelle on se réfère, le génome humain, comme d'ailleurs celui des autres espèces, doit avoir de 25 000 à plusieurs centaines de milliers de « gènes ».

Ainsi apparaît-il que le séquençage du génome humain dont rend compte l'ouvrage de Bertrand Jordan n'est qu'une étape, certes importante, des progrès de la biologie. Pour reprendre la métaphore linguistique précédemment convoquée, l'on pourrait assimiler le travail ainsi réalisé à une vaste entreprise lexicographique : il va de soi que la possession d'un bon dictionnaire, constitué ici des gènes « annotés » est un outil essentiel de toute étude linguistique et littéraire, c'est-à-dire, ici, biologique. Au-delà de cette réalité, fleurissent des déclarations manifestant une incompréhension profonde de ce qu'est le génome ou reflétant des desseins divers, d'essence idéologique, commerciale ou politique. La lecture de l'ouvrage de Bertrand Jordan donnera à ses lecteurs les clés nécessaires pour, dans l'avenir, différencier, dans le discours autour des relations entre l'homme et ses gènes, ce qui revient probablement à la prudence scientifique ou bien ressort d'un discours d'une toute autre nature.

Axel Kahn

1. ENTRÉE EN MATIÈRE

AVANT-PROPOS

J'ai commencé à écrire pour médecine/sciences en 1989, sur l'amicale sollicitation d'Axel Kahn, alors Rédacteur en chef. Il s'agissait d'un éditorial sur la « Génétique inverse », terme employé à l'époque pour la démarche qui allait, en quelques années, identifier plus d'un millier de gènes impliqués dans différentes maladies génétiques humaines. Cette collaboration est devenue régulière et intensive lorsque, en 1991, j'ai consacré une année entière à un « Tour du monde du Génome ». Il s'agissait en fait d'une enquête menée dans près d'une centaine de laboratoires, des États-Unis à l'Europe en passant par le Japon ou l'Australie, afin de voir où en était réellement ce « Programme Génome » alors à ses débuts. Les « Chroniques génomiques » rédigées au fil des visites et aussitôt publiées dans médecine/sciences ont donné naissance à deux livres relatant ce voyage, l'un du point de vue scientifique et organisationnel¹, l'autre d'une manière plus personnelle, moins axée sur la science mais insistant sur la découverte des pays, des cultures et des individus rencontrés au cours de ce périple². Le présent ouvrage se rapporte à une période ultérieure : il reproduit les chroniques parues dans médecine/sciences de 1992 à 2002. Regroupées par thèmes, elles constituent un aperçu – personnel, partiel et sans doute parfois partial – du déroulement de ce Programme Génome qui, en une décennie, est passé des premières cartes grossières et truffées d'erreurs à la séquence quasiment complète des trois milliards de lettres qui constituent notre patrimoine héréditaire. Elles illustrent les aléas de cette aventure, les illusions et parfois les déconvenues qui l'ont marquée, montrent la formidable évolution de certaines techniques et la relative stagnation d'autres. Elles font aussi une large part aux problèmes d'organisation inévitablement rencontrés au cours de cette entreprise très nouvelle pour la biologie par son ampleur et son caractère systématique. Ces textes sont reproduits tels qu'ils ont paru³, accompagnés de quelques commentaires qui les resituent dans leur contexte ou examinent si les pronostics formulés se sont ou non avérés corrects.

J'espère que les lecteurs qui ont apprécié ces chroniques lors de leur parution les retrouveront avec plaisir, et que d'autres découvriront à travers elles les péripéties parfois oubliées de ce projet sans précédent, dont l'achèvement nous ouvre aujourd'hui des perspectives aussi passionnantes qu'à certains égards inquiétantes...

1. *Voyage autour du Génome, le tour du monde en 80 labs*. Paris : John Libbey Eurotext, 1993.

2. *Voyage au pays des gènes*. Paris : Les Belles Lettres, 1996.

3. Nous avons même conservé les conventions typographiques de l'époque de parution, y compris la référence à la revue, passée de *médecine/sciences* à *Med Sci*...

ANNÉES GÉNOME : LES DIX GLORIEUSES

Ce premier texte n'est pas une « vraie » chronique, mais un éditorial écrit pour le numéro de médecine/sciences de janvier 2000 consacré à « La révolution du génome ». Rédigé peu avant l'obtention du premier « brouillon » de la séquence de l'ADN humain, il situe globalement les différents éléments du Programme génome et constitue à ce titre, me semble-t-il, une bonne introduction à ce recueil.

1988 : les Programmes Génome se mettent en place, après plusieurs années de débats lancés par le colloque de Santa Cruz en mai 1985. La communauté scientifique, dans sa grande majorité, est sceptique. Des biologistes de renom (David Botstein, Robert Weinberg, entre autres) expriment même une opposition farouche : certains se convertiront bientôt. L'image publique du projet, véhiculée par les lobbyistes qui ont réussi à convaincre le Congrès américain d'investir des sommes qui paraissent colossales (deux cent millions de dollars, « qui seraient tellement mieux employés à faire de la vraie biologie »), apparaît très décalée par rapport à la réalité. Pour le journaliste ou l'homme de la rue (si tant est qu'il en ait entendu parler), c'est le « Grand programme du séquençage du Génome », projet international centré essentiellement sur la séquence. Et l'on croit que cette entreprise est strictement coordonnée par une sorte de Comité central du Génome dont on attribuera bientôt le rôle à HUGO, la *Human Genome Organisation*, fondée cette année-là.

Les faits sont tout autres : plusieurs programmes nationaux démarrent effectivement, mais dans une grande dispersion accompagnée d'une vive concurrence : deux programmes aux États-Unis, ceux du *National Institutes of Health* et du *Department of Energy*, quatre ou cinq au Japon, un en Grande-Bretagne... Et surtout, le contenu effectif de ces travaux porte pour l'essentiel sur la construction des cartes (carte génétique, carte physique), préalable incontournable au séquençage global, qui semble une entreprise chimérique ou en tout cas fort lointaine. C'est ce recentrage sur la construction des cartes, objectif plus réaliste et, par ailleurs, directement utile à la Génétique médicale, qui entraînera le ralliement d'une partie de la communauté ou du moins une baisse sensible de son hostilité.

Un colloque fondateur

Le premier colloque *Genome Mapping and Sequencing* tenu à *Cold Spring Harbor* (New York, USA) au printemps 1988 reflète ces tendances. J'en rapportais quelques éléments dans une *Chronique Génomique* avant la lettre (le début officiel de cette série se situant en 1989) intitulée *Chronique de Cold Spring Harbor* [1]. Un peu plus de dix ans après, il est instructif de la relire et de consulter le livre des *abstracts* correspondant. Plusieurs éléments sont frappants. Le séquençage n'occupe qu'une demie-journée, la matinée du dimanche (lorsque beaucoup de participants se préoccupent surtout de rejoindre l'aéroport Kennedy à l'aide d'incertaines limousines aux horaires aléatoires), et est vu sous l'angle des premières machines (aujourd'hui si courantes) et de quelques tentatives aux ambitions à la fois modestes dans l'absolu et démesurées par rapport aux

réalisations de l'époque. Déchiffrer une ou deux mégabases dans une région du génome particulièrement intéressante : ces tentatives (comme celle, à l'époque, du CEPH – Centre d'Étude du Polymorphisme Humain – qui se proposait de séquencer les deux ou trois mégabases du complexe majeur d'histocompatibilité, ou celle du laboratoire de Fred Blattner qui, avec le groupe Japonais de Katsumi Isono à Kobé, se lançait très logiquement dans le déchiffrement des quatre mégabases du génome d'*Escherichia coli*), vont échouer lamentablement. Les problèmes d'organisation, de gestion des flux, de contrôle de qualité, les efforts nécessaires à l'identification et à l'élimination des goulets d'étranglement dans une entreprise à l'échelle industrielle avaient été largement sous-estimés. Ils entraînaient l'arrêt de ces projets après une ou deux années au cours desquelles les séquences obtenues représentaient un cinquième ou un dixième des prévisions, tandis que les coûts, eux, dépassaient allègrement le mythique « un dollar par base » que l'on avait pourtant espéré réduire rapidement à un chiffre dix fois plus faible [2].

Le séquençage intégral du génome humain semblait reporté aux calendes grecques, et beaucoup (moi-même compris) doutaient fort qu'il soit effectué un jour – compte tenu de l'énormité de la tâche, des difficultés induites par la multiplicité des séquences répétées, et du maigre intérêt des 95 % de séquences non codantes qui encombrant notre ADN. A moins qu'une technique révolutionnaire ne multiplie subitement par dix ou cent la vitesse d'obtention des données tout en réduisant considérablement les coûts ? Aucune des approches sur lesquelles reposaient à l'époque nos espoirs (lecture directe par microscopie en champs proches, analyse à très haute sensibilité de bases excisées d'une molécule d'ADN, ou même séquençage par hybridation) ne s'est aujourd'hui encore révélée réellement opérationnelle, elles étaient d'ailleurs pratiquement absentes de cette première édition du colloque. Si aujourd'hui le séquençage est réellement engagé, avec déjà près de 20 % de nos trois mille mégabases déchiffrées au cours des deux dernières années, c'est à l'aide de la méthode publiée dès 1977 par Fred Sanger. Méthode rendue plus rapide, plus puissante, grâce à une multitude de petites améliorations, à une automatisation encore partielle, et surtout à une organisation de type industriel, quasiment militaire : les grandes « usines à séquencer » regroupent des dizaines ou même des centaines de machines, un personnel technique spécialisé très encadré et une informatique performante, le tout dans le cadre de projets très définis et structurés de manière rigoureuse.

Une autre absente de taille en 1988 : l'approche des EST (*expressed sequence tags*) (séquences partielles de milliers de clones d'ADNc), initiée deux plus tard par le précurseur Craig Venter à contre-courant de la doctrine officielle, révélée au grand jour en 1961 par le « scandale » des brevets sur ces entités, et qui aujourd'hui, avec trois millions d'EST dans le domaine public et bien plus dans les bases de données industrielles privées, occupe une place centrale dans la plupart des programmes de Génomique. Bel exemple des bouleversements que peut introduire une idée originale, et de l'impossibilité de planifier strictement une entreprise pourtant aussi routinière et stupide (d'après ses détracteurs) que l'établissement de l'anatomie du Génome humain...

Les cartes avant la séquence

Le colloque de 1988 était donc majoritairement consacré aux cartes. La première carte génétique de l'ensemble du génome humain, coordonnée par Helen Donis-Keller et qui fit la couverture de *Cell* en 1987, n'avait pas apparemment suscité beaucoup d'émules : seuls deux ou trois *abstracts* font état d'efforts dans ce domaine : enrichissement de la carte générale, ou cartographie plus précise d'un ou deux chromosomes individuels. Son deuxième souffle, fondé sur la découverte et l'utilisation massive des microsatellites, plus fréquents, plus polymorphes et se prêtant mieux à une exploitation semi-automatisée que les incommodes RFLP (*restriction fragment length polymor-*

phisms), n'était pas encore manifeste. Il faudrait attendre l'aventure du Généthon, l'investissement massif de l'AFM grâce à l'impulsion de Bernard Barataud, l'audace de Daniel Cohen et la rigueur de Jean Weissenbach pour qu'apparaisse, en 1992, la première carte de deuxième génération qui devait, entre autres, multiplier instantanément par dix le rythme de la localisation de maladies génétiques. La carte génétique de troisième génération, édifiée sur l'exploitation intégrale du polymorphisme humain, est aujourd'hui bien engagée : la découverte et la détection des SNP (*single nucleotide polymorphisms*), pour laquelle plusieurs techniques d'analyse performantes s'affrontent encore, a déjà permis de positionner plusieurs dizaines de milliers de balises et promet l'obtention de la « carte à cent mille marqueurs » indispensable pour l'élucidation du déterminisme des maladies multigéniques et pour de nombreuses approches de « pharmacogénétique » [3]. Comme pour les EST, secteur public, entreprises industrielles, et aussi consortiums public-privé contribuent à l'accumulation de ces données non sans concurrence, duplication d'efforts et coups médiatiques dirigés plus vers les investisseurs ou la Bourse que vers la communauté scientifique.

Les cartes physiques, en revanche, étaient au centre du débat en 1988 et constituaient l'objet principal du débat. Deux approches s'opposaient : les cartes « par le haut » (*top down*) et « par le bas » (*bottom up*). Par le haut : l'invention en 1984 par David Schwartz et Charles Cantor de la séparation de très grands fragments d'ADN par électrophorèse en champs pulsés et la découverte des enzymes de restriction à site rare ouvraient, en principe, la possibilité de construire des cartes de restriction de chromosomes entiers. En fait, ces cartes, peu fiables en raison de leur sensibilité aux polymorphismes et à la méthylation de l'ADN, devaient s'avérer très délicates à construire et encore plus à exploiter, et leur contribution au balisage de notre génome a été modeste, sans commune mesure avec les efforts déployés. L'approche « par le bas » était la bonne : la constitution de collections de clones contigus, des *contigs*, couvrant de grandes régions et si possible des chromosomes entiers, pouvait rendre accessible à une expérimentation détaillée tout point du génome (élément crucial pour la Génétique médicale) et préparer la voie au séquençage. Mais caractériser et ordonner cinq ou dix mille cosmides pour « couvrir » le chromosome 16 ou 19 (tâche à laquelle s'étaient attelés les laboratoires de Robert Moysis et Anthony Carrano sous l'égide du *Department of Energy*) n'était pas une mince affaire, et beaucoup doutaient que ce soit possible. De nouveaux systèmes de clonage permettant de propager de grands fragments d'ADN allaient faciliter la tâche.

L'arrivée des chromosomes artificiels de levure, les YAC (*yeast artificial chromosomes*) (publiés en 1987 par David Burke, Georges Carle et Maynard Olson, mais dont la mise en place effective dans les laboratoires concernés devait prendre beaucoup de temps), et plus tard celle des P1 et des BAC (*bacterial artificial chromosomes*), vecteurs de clonage bactériens plus maniables et plus fidèles, allait sauver les cartes physiques, aujourd'hui établies de manière à peu près sûre sur l'ensemble du génome après quelques tentatives très médiatisées mais peu opérationnelles. Le séquençage impose encore l'affinement et la fiabilisation de ces cartes (obtention de cartes « prêtes à séquencer », *sequence-ready*), généralement effectué au sein des grands centres de séquençage qui ont pris en charge le déchiffrement de tel ou tel chromosome – à moins de faire l'impasse sur cet aspect et de tenter, comme le fait Craig Venter, le séquençage en *shotgun* intégral (on séquence au hasard, on assemble ensuite). Approche qui a montré sa puissance pour les petits génomes, et qui, pour les plus grands (drosophile, homme, souris), s'appuiera sans doute (sans trop le dire) sur des éléments de cartographie et de séquence déjà présents dans les bases de données publiques...

Aujourd'hui plus personne (ou presque) ne doute de l'importance, de l'intérêt, du caractère opérationnel des données sur les génomes. Les génomes : les organismes-modèle, bactéries, levure, nématode, *Arabidopsis thaliana*, drosophile et j'en passe, ont pris une place qui n'était pas évidente il y a dix ans.

Place psychologique : le succès du Projet Levure, pourtant éparpillé entre une cinquantaine de laboratoires européens et objet à son début des sarcasmes de nos collègues d'outre-Atlantique, a redonné confiance aux séquenceurs en montrant qu'il était effectivement possible de déchiffrer intégralement plus d'une dizaine de mégabases [4].

Place scientifique surtout, dans la mesure où la parenté des gènes d'organismes très éloignés dans l'échelle de l'évolution s'est révélée forte (grâce aussi à des techniques bioinformatiques de comparaison de plus en plus performantes), allant parfois jusqu'à une interchangeabilité fonctionnelle dont on n'osait pas rêver.

Il n'est pratiquement plus de projet de recherche biologique qui ne fasse appel à un moment ou à un autre aux données génomiques accessibles grâce aux bases de données disponibles sur Internet ; beaucoup sont même largement fondés sur leur exploitation à l'aide d'une bioinformatique très diversifiée, dont l'épanouissement a été accompagné par une fantastique accélération des performances des ordinateurs et un extraordinaire développement du réseau Internet. Bioinformatique presque absente en 1988 (aucune communication sur ce sujet, et à peine trois *posters* sur une quarantaine), et qui joue aujourd'hui un rôle central dans toute la génomique.

Les industriels s'intéressent de près au génome

L'irruption du secteur privé est sans doute le changement le plus fondamental apparu au cours de cette décennie du Génome. Certes, l'industrie n'était pas totalement absente à *Cold Spring Harbor* en avril 1988 : n'oublions pas que le travail d'Helen Donis-Keller était effectué dans le cadre de la firme *Collaborative Genetics*. Mais cela semblait presque une bizarrerie à l'époque, et n'a d'ailleurs pas porté bonheur à la firme en question. L'ambiance du colloque était encore très universitaire. Aujourd'hui l'industrie détient sans doute plus de données génomiques (séquences de bactéries et de pathogènes végétaux, EST d'homme, de souris, de rat, collections de SNP...) que les bases publiques. Certaines de ces informations sont rendues disponibles au bout d'un certain délai, après que les brevets aient été déposés ; d'autres restent confidentielles et sont parfois réétablies par la recherche académique.

L'industrie pharmaceutique (mais n'oublions pas non plus les firmes agro-alimentaires) s'est lourdement investie dans l'exploitation (et parfois l'acquisition) de données sur les génomes : effort compréhensible en regard de l'accélération attendue dans la découverte de nouveaux médicaments pour lesquels les dépenses de recherche et développement sont de l'ordre de cinq cents millions de dollars par molécule. Que représentent quelques millions investis ou payés à un sous-traitant s'ils permettent de hâter ce processus, de mieux choisir les cibles et donc de réduire le coût total ? L'investissement peut être direct, en développant les programmes de recherche génomique *in house* ; il est le plus souvent indirect, prenant la forme de contrats avec des *start-ups* spécialisées dans tel ou tel aspect de la génomique : Incyte et beaucoup d'autres aux États-Unis, Genset chez nous. Ces importantes mises de fonds démontrent (s'il en était encore besoin) l'intérêt des données obtenues sur les génomes et la perspicacité de ceux qui ont lancé ces projets dans les années quatre-vingts ; elles posent aussi de redoutables problèmes aux équipes académiques qui se retrouvent parfois en concurrence directe avec un département de recherche industriel sans disposer des mêmes ressources, notamment en bioinformatique ou pour la mesure en grand de l'expression des gènes, deux approches centrales en Génomique.

Les années « biologie »

Les années qui viennent, on peut l'affirmer sans risque, seront celles de la biologie. Le seul phénomène scientifico-industriel d'importance comparable est celui de l'informatique et du réseau Internet. Encore s'agit-il là avant tout d'impact sur la société, sur son mode de fonctionnement, sur le positionnement des individus. La Génomique, tout en ayant des conséquences sociétales du même ordre de grandeur, constitue de plus un enjeu de première dimension au niveau cognitif : la compréhension du vivant, appuyée maintenant sur une connaissance précise et détaillée de sa base génétique, va se développer dans des directions dont nous pouvons imaginer les grandes lignes (développement, neurobiologie, réseaux d'interactions) mais dont les révélations sont par nature imprévisibles. Les années 1980 et 1990 ont été une époque fabuleuse pour la biologie : attendons-nous, dans les années 2000, à d'autres progrès, d'autres surprises... et bien sûr aussi à d'autres problèmes !

Références

1. Jordan B. Cartographie et séquence du génome humain. Chronique de *Cold Spring Harbor Laboratory. Med Sci* 1988 ; 4 : 448-50.
2. Wada A. Automated high-speed DNA sequencing. *Nature* 1987 ; 325 : 771-2.
3. Jordan B. Du Programme Génome à la « pharmacogénomique ». *Med Sci* 1997 ; 13 : 1176-8.
4. Jordan B. Séquençage génomique : le deuxième souffle. *Med Sci* 1992 ; 8 : 854-7.

Depuis cet éditorial (paru au tout début de l'année 2000), les séquences « brouillon » puis « finies » du génome humain et de plusieurs autres génomes ont été publiées, et les surprises n'ont pas manqué. Le nombre (probable) de gènes humains a été revu à la baisse et est passé de plus de 100 000 à 30 ou 40 000 (voir Chapitre 8). Les phénomènes d'épissage alternatif se sont avérés bien plus fréquents que prévu, et touchent sans doute la majorité des gènes. Enfin, la découverte des ARNi, ces petits ARN transcrits mais non traduits, ouvre une nouvelle fenêtre sur les processus de régulation tout en donnant un rôle à une partie de l'ADN non codant. Le panorama effectué reste néanmoins d'actualité, et j'ai regroupé les chroniques dans un ordre qui correspond approximativement au survol effectué ci-dessus. Nous allons donc commencer par les cartes...

Vj ku' r ci g'kpvgpvkqpcmf 'igh'dre pm

2. L'ÈRE DES CARTOGRAPHES

Les premiers temps du programme ont été dominés par la réalisation des cartes du génome humain. On oublie un peu à quel point ces études étaient imparfaites et parcellaires dans les années 1980 : la première carte génétique générale date de 1987, et ses RFLP espacés de vingt ou quarante centimorgans la rendent peu utilisable. Quant aux cartes physiques, beaucoup pensaient lors des débuts du programme Génome qu'elles seraient impossibles à établir à l'échelle de chromosomes entiers. Ce balisage apparaissait pourtant indispensable au séquençage (objectif encore lointain compte tenu des techniques alors disponibles), mais aussi, et surtout, il était directement utile à la génétique médicale pour sa recherche de gènes impliqués dans des maladies héréditaires. De 1988 (commencement réel des études à grande échelle) jusque vers la fin des années 1990, la construction de cartes génétiques et physiques a donc constitué l'essentiel du travail de nombreux laboratoires et Genome centers. Cette aventure, dans laquelle la France a joué un rôle de premier plan grâce au CEPH, à l'AFM et à Généthon, a connu de nombreux rebondissements dont l'écho transparaît dans les chroniques publiées à l'époque.

Références

1. Pennisi E. Human Genome Project. And the gene number is... ? *Science* 2000 ; 288 : 1146-7.
2. Dunham I, Shimizu N, Roe BA, *et al.* The DNA sequence of human chromosome 22. *Nature* 1999 ; 402 (6761) : 489-95.
3. Boguski MA, Schuler GD. ESTablishing a human transcript map. *Nat Genet* 1995 ; 10 : 369-71.
4. Adams MD, Kerlavage AR, Fleischmann RD, *et al.* Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 1995 ; 377 : 3-174.
5. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. Gene index analysis of the human genome estimates approximately 120 000 genes. *Nat Genet* 2000 ; 25 : 239-40.
6. Hattori M, Fujiyama A, Taylor TD, *et al.* The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature* 2000 ; 405 : 311-9.
7. Jordan B R. Fugu Story. *Med Sci* 1994 ; 10 : 1154-6.
8. Elgar G, Sandford R, Aparicio S, Macrae A, Venkatesh B, Brenner S. Small is beautiful : comparative genomics with the pufferfish (*Fugu rubripes*). *Trends Genet* 1996 ; 12 : 145-50.
9. Roest Crolius H, Jaillon O, Bernot A, *et al.* Estimate of human gene number provided by genome-wide analysis using tetraodon *nigroviridis* DNA sequence. *Nat Genet* 2000 ; 25 : 235-8.
10. Ewing B, Green P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet* 2000 ; 25 : 232-4.
11. Mironov AA, Fickett JW, Gelfand MS. Frequent alternative splicing of human genes. *Genome Res* 1999 ; 12 : 1288-93.
12. Bretta D, Hanke J, Lehmann G, *et al.* EST comparison indicates 38 % of human mRNAs contain possible alternative splice forms. *FEBS Lett* 2000 ; 474 : 83-6.

Le pari de Cold Spring Harbor a trouvé son épilogue au printemps 2003. En fait, on ne connaît toujours pas le nombre exact des gènes humains, quoique la plupart des scientifiques s'accordent sur une valeur de l'ordre de 30 000. Mais les termes du pari lancé en 2000 stipulaient que le vainqueur devait être annoncé en 2003 ; ceci fut donc fait, en se fondant sur les données du site Ensembl (www.ensembl.org), projet commun à l'EMBL (Laboratoire européen de biologie moléculaire) et à l'EBI (European Bioinformatics Institute) et qui constitue sans doute la meilleure analyse actuelle du génome humain. Ensembl répertorie 24 847 gènes humains, et c'est donc Lee Rowen (Institute for Systems Biology, Seattle, États-Unis) qui a emporté le « pot » constitué par les 461 parieurs : il avait prévu 25 947 gènes.

9. EN GUISE DE CONCLUSION...

Ces chroniques émanent d'un observateur jouissant d'une relative indépendance. Durant la période où elles ont été rédigées, je n'étais ni directeur d'institut, ni responsable de département au CNRS ou à l'Inserm, mais simple chef d'équipe dans un bon laboratoire (le Centre d'Immunologie Inserm/CNRS de Marseille-Luminy) bénéficiant d'un financement très correct. Ayant effectué en 1991 une sabbatique-enquête d'une année qui me donnait une vision d'ensemble du monde du Génome, participant à plusieurs comités nationaux et étrangers intervenant dans ce domaine, je n'avais pourtant pas de responsabilité institutionnelle lourde. Je pouvais donc m'exprimer librement, développer une vision personnelle (d'ailleurs parfois erronée) sans non-dit ni langue de bois, d'autant que la revue médecine/sciences n'a jamais exercé aucune censure sur la forme ou sur le fond de mes chroniques. L'histoire ainsi racontée à travers ces textes est avant tout sincère, malgré les antagonismes très vifs qui régnaient dans le milieu, les oppositions entre AFM, GRÉG, ministère, organismes de recherche et les personnalités parfois explosives de certains protagonistes. Je ne suis pas peu fier, à cet égard, du fait que mon texte sur le Très Grand Séquençage, paru dans médecine/sciences, ait été repris tel quel et in extenso tant dans le bulletin des Sciences de la Vie du CNRS que dans le Journal de l'AFM.

Pour autant, l'histoire ainsi racontée est incomplète : par exemple, elle ne décrit pas la compétition entre le programme public et l'entreprise Celera pour l'obtention du premier brouillon de la séquence humaine. D'une manière générale, elle témoigne d'une prise de conscience un peu tardive du rôle des entreprises en recherche génomique. Je pense pourtant que ce recueil, malgré ses défauts et ses manques, aura permis au lecteur de revivre une période passionnante, et peut-être même d'en tirer quelques leçons... Ces dix années ont vu la recherche biologique passer (au moins en partie) de l'artisanat à la Big science, tandis que faisaient irruption la technologie et l'informatique, et que les questions posées changeaient de nature. Ce basculement déjà très avancé va certainement se poursuivre ; il n'est pourtant pas dit qu'il doive s'étendre à l'ensemble de cette science. Au contraire même, certaines disciplines abusivement délaissées au profit de la toute-puissante biologie moléculaire, comme la