

Le génome humain entre médecine et science

Les premiers succès de la génétique humaine moléculaire, qui datent du début des années 1980, ont été obtenus en dépit de l'absence de cartes du génome. Ces dernières, qui commencent à être disponibles, provoquent une accélération considérable de l'acquisition des connaissances. De nouveaux outils ont facilité cette recherche, au premier rang desquels il faut citer le clonage en chromosomes artificiels de levure permettant de manipuler de grands fragments d'ADN, et la PCR. Avec l'analyse de grands génomes, la biologie est entrée dans une phase d'acquisition massive de données posant de difficiles problèmes de logistique. Les structures actuelles sont encore inadéquates pour le recueil massif de données et les moyens requis bien supérieurs au budget des organismes de recherche. Les retombées majeures de ces grands programmes, aussi bien dans le domaine fondamental que dans celui des applications, ne se feront, pour la plupart, qu'au cours du siècle prochain.

Jean Weissenbach

Le virage moléculaire de la génétique humaine remonte au début des années 1980 et coïncide presque avec la création de *médecine/sciences* qui a consacré une part importante de ses colonnes à cette révolution. Cet article se propose de revenir sur quelques faits saillants de cette nouvelle page de l'aventure humaine, dont l'écriture s'accélère à mesure que progresse la cartographie systématique du génome humain.

1980 : la génétique humaine se fait moléculaire

La découverte du polymorphisme de restriction de l'ADN humain [1]

marque le point de départ de toute une série de premières cartographiques, d'abord du *linkage* de maladies génétiques liées au sexe avec des marqueurs polymorphes du chromosome X [2] puis de la localisation de maladies autosomiques [3]. Il est donc possible, en ce début de décennie, de cartographier les maladies génétiques, mais les succès demeurent souvent le fait du hasard (proximité d'un marqueur particulièrement informatif) et de nombreuses maladies fréquentes, comme la mucoviscidose, ne sont pas encore localisées. Tout le monde déplore l'absence d'une carte génétique [4]. Mais son établissement pose aussi de sérieux problèmes que la création du Centre d'étude du polymorphisme humain

ADRESSE

J. Weissenbach : directeur de recherche au Cnrs. Cnrs URA 1922, Gépéthon, 1, rue de l'Internationale, 91002 Évry Cedex, France.

RÉFÉRENCES

1. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 1978 ; 75 : 5631-5.
2. Murray JM, Davies KE, Harper PS, Meredith L, Mueller CR, Williamson R. Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature* 1982 ; 300 : 69-71.
3. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 1983 ; 306 : 234-8.
4. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet* 1980 ; 32 : 314-31.
5. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. Program description: Centre d'étude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 1990 ; 6 : 575-8.
6. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984 ; 37 : 67-75.
7. Collins FS, Weissman SM. Directional cloning of DNA fragments at large distance from an initial probe: a circularization method. *Proc Natl Acad Sci USA* 1984 ; 81 : 6812-6.
8. Kunkel LM, Monaco AP, Middlesworth W, Ochs HD, Latt SA. Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc Natl Acad Sci USA* 1985 ; 82 : 4778-82.
9. Davies KE, Young BD, Elles RG, Hill ME, Williamson R. Cloning of a representative genomic library of the human X chromosome after sorting by flow cytometry. *Nature* 1981 ; 293 : 374-6.
10. Cavenee WK, Dryja TP, Phillips RA, Benedict WF, Godbout R, Gallie BL, Murphree AL, Strong LC, White RL. Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 1983 ; 305 : 779-84.
11. Marshall CJ. Tumor suppressor genes. *Cell* 1991 ; 64 : 313-26.
12. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971 ; 68 : 820-3.

(CEPH) va, en partie, aider à résoudre. Jean Dausset, Daniel Cohen et leurs collègues proposent aux principaux cartographes du monde entier d'établir la carte génétique à l'aide d'un matériel de référence commun [5]. Établies sur les mêmes produits méiotiques, les données peuvent être mises en commun et chaque collaborateur pourra contribuer à l'œuvre commune. Cela permettra de localiser tous les marqueurs génétiques les uns par rapport aux autres et donc de gagner du temps et d'éviter des situations de monopole.

Dans la foulée des premiers *linkage*, on essaye de se rapprocher des gènes déjà localisés sur la carte. Toutefois, cette recherche de gènes est dépendante de techniques particulièrement lourdes ou qui souvent tiennent plus de l'acrobatie que de protocoles aux conditions bien contrôlables [6-8]. Le génome humain se laisse aborder, mais c'est une aventure incertaine, de très longue haleine, où il faut plus compter sur la chance que sur l'efficacité des outils. Ceux-ci restent les outils de la première vague de la génétique moléculaire : la cartographie physique et l'isolement de nouvelles sondes font surtout appel aux banques d'ADN génomiques en bactériophages lambda ou cosmides et la cartographie génétique recourt pour l'essentiel au polymorphisme de longueur des fragments de restriction (RFLP, *restriction fragment length polymorphism*). Les banques d'ADN spécifiques d'un chromosome trié par cytométrie de flux commencent à se répandre [9]. Avec de tels moyens, le travail de cartographie des maladies et d'identification des gènes reste fastidieux, répétitif et peu efficace. Mais les enjeux sont alors tels que de nombreuses équipes se lancent avec une certaine inconscience dans la chasse aux gènes de maladies. Il est *a posteriori* très clair que le manque de cartes et des collections de réactifs génomiques bien caractérisés et cartographiés a été un très lourd handicap qui s'est prolongé jusqu'à fin 1992. À l'inverse, le passage aux applications a été immédiat, mais le pouvoir prédictif du diagnostic anténatal a de quoi effrayer. Ce diagnostic est néanmoins une révolution de la médecine et l'on observe de nouvelles nais-

sances dans des familles où le risque génétique avait pratiquement fait abandonner ce projet.

Le printemps de la génétique humaine

En dépit de leur qualité rudimentaire, les outils du début des années 1980 permettent de faire des observations majeures dans le domaine de la génétique humaine (*Tableau I*). Leur amélioration et l'introduction de nouvelles techniques ne feront qu'accélérer le rythme de ces acquisitions (*Tableau I*). Petit à petit, les gènes des maladies génétiques les plus fréquentes sont localisés. L'utilisation de marqueurs polymorphes permet aussi de mettre en évidence la perte d'hétérozygotie, d'abord dans le rétinoblastome [10], puis dans d'autres prédispositions génétiques aux tumeurs [11]. La mise en évidence de ce phénomène est une confirmation éclatante du modèle de tumorigenèse en deux temps de Knudson [12]. Grâce à un heureux concours de circonstances, le gène du rétinoblastome est cloné en peu de temps. L'analyse des délétions incluant le gène permet de définir un intervalle de quelques kilobases [13]. C'est, en 1986, le premier succès du clonage positionnel, suivi de près par l'isolement d'exons du gène de la myopathie de Duchenne [14]. Ces exons ont été trouvés à la suite d'un travail délicat d'isolement de sondes ciblé sur une région de quelques mégabases [8]. Mais ces premières restent isolées en raison du manque de réactifs de cartographie. C'est ainsi qu'il faudra dix ans pour identifier le gène de la chorée de Huntington (*m/s n° 4, vol. 9, p. 488*) [15] alors qu'il fut le premier gène autosomique cartographié à l'aide de marqueurs d'ADN polymorphes [3]. Cependant, petit à petit, les difficultés sont surmontées, l'utilisation de nouvelles techniques (*voir ci-dessous*) permet d'identifier d'autres gènes par clonage positionnel vers la fin des années 1980 et le début des années 1990. C'est notamment l'utilisation de l'électrophorèse en champ pulsé qui permet, après sept ans d'efforts, de déceler le *locus* de l'X fragile [16]. L'analyse moléculaire de ce dernier aboutit à la mise en évidence d'une expansion de répétitions d'un trin-

Tableau I

SÉLECTION DE FAITS MARQUANTS DES 15 PREMIÈRES ANNÉES D'EXPLORATION DU GÉNOME

1979	Premier RFLP (<i>restriction fragment length polymorphism</i>) [1]
1980	Proposition de cartographier le génome à l'aide de RFLP [4]
1981	Première banque d'ADN spécifique d'un chromosome humain [9]
1982	Localisation du gène de la myopathie de Duchenne <i>DMD</i> [2]
1983	Localisation du gène de la maladie de Huntington <i>HD</i> [3]
1984	Perte d'hétérozygotie dans le rétinoblastome (gène <i>RB</i>) [10]
1984	Électrophorèse en champ pulsé [6]
1984	Création du Centre d'études du polymorphisme humain (CEPH) [5]
1985	Localisation du gène de la mucoviscidose <i>CF</i> (<i>m/s n° 8, vol. 5, p. 589</i>)
1985	Amplification <i>in vitro</i> de l'ADN par PCR (<i>m/s n° 8, vol. 4, p. 515</i>) [22]
1986	Clonage du gène <i>RB</i> (<i>m/s n° 6, vol. 3, p. 363</i>) [13]
1986	Premiers exons du gène <i>DMD</i> (<i>m/s n° 3, vol. 4, p. 141</i>) [14]
1986	Lecture automatisée de séquences d'ADN marqué par fluorescence [36]
1987	Carte génétique humaine de RFLP [25]
1987	ADN humain cloné en chromosome artificiel de levure (YAC) (<i>m/s n° 8, vol. 3, p. 496</i>) [21]
1989	Clonage du gène <i>CF</i> (<i>m/s n° 8, vol. 5, p. 589</i>)
1990	Lancement du programme génome américain (<i>m/s n° 9, vol. 7, p. 957</i>)
1991	Extension de triplets répétés dans <i>FRX1</i> (<i>m/s n° 6, vol. 7, p. 637</i>)
1992	Séquençage du premier chromosome de levure [31]
1992	Hybridation <i>in situ</i> sur fibres de chromatine étirée [29]
1992	Premières cartes physiques de chromosomes humains (<i>m/s n° 9, vol. 8, p. 966</i>)
1992	Carte génétique humaine de microsatellites (<i>m/s n° 1, vol. 9, p. 84</i>)
1993	Première ébauche de carte physique du génome humain (<i>m/s n° 1, vol. 9, p. 84</i>) [27]

cléotide chez les sujets malades. Ce nouveau type de mutation due à l'instabilité d'une séquence répétée permet de rendre compte de toute une série d'observations cliniques qui ne cadraient pas strictement avec un mode de transmission lié à l'X [17]. Depuis, ces expansions de triplets répétés ont été observées dans tout un ensemble de maladies neurologiques héréditaires où se produit une anticipation se traduisant par une apparition plus précoce ou une aggravation des signes cliniques de génération en génération [18].

À ce jour, près d'une cinquantaine de gènes de maladies monogéniques ont été identifiés en une dizaine d'années et ce rythme est appelé à croître. Tous ces gènes n'ont cependant pas été identifiés par clonage purement positionnel, mais souvent par une combinaison d'approches utilisant, en particulier, la localisation de gènes potentiellement candidats, au vu de leurs propriétés biologiques ou structurales, dans la région contenant le gène de la maladie.

De nombreuses autres observations hautement significatives qu'il est impossible de passer en revue ici ont émaillé ce premier âge d'or de la gé-

nétiq ue humaine. Parmi les avancées les plus récentes, il convient de mentionner la mise en évidence, par des études de liaison, de régions génomiques impliquées dans le diabète insulino-dépendant [19, 20], ouvrant ainsi la voie de l'étude de maladies multifactorielles.

Les nouveaux outils

La mise au point de nouvelles techniques et l'accroissement des collections de réactifs ont facilité les découvertes de gènes, surtout ces dernières années. Trois techniques majeures d'analyse du génome ont émergé pendant cette décennie.

Les deux premières, l'électrophorèse en champ pulsé [6] et le clonage en chromosomes artificiels de levure (YAC, *yeast artificial chromosome*) [21], sont étroitement liées : elles requièrent la capacité de manipuler des grands fragments d'ADN, un savoir-faire à présent maîtrisé. De plus, le clonage en YAC est directement dépendant du champ pulsé. Ces deux techniques comblent le fossé méthodologique entre les techniques de faible résolution (cytogénétique et analyse de liaison) dont la limite in-

férieure est de l'ordre du mégabase et les techniques moléculaires classiques (électrophorèse à champ uniforme, clonage en cosmides et phages) dont la limite supérieure se situe autour de cinquante kilobases. Alors que de nombreuses variantes de séparation électrophorétique en champ pulsé ont été mises au point et sont pratiquées par de nombreux laboratoires, le clonage en YAC reste l'apanage de quelques laboratoires qui ont acquis un savoir-faire et entretiennent leurs banques où des dizaines de milliers de clones sont conservés individuellement. Ces laboratoires ont mis en place un service qui assure la distribution des clones pour répondre aux nombreuses demandes du monde entier.

La troisième technique, la PCR, est d'utilisation quasi universelle en biologie moléculaire et génétique [22]. L'importance qu'elle a acquise dans le domaine du génome se mesure à la part qu'elle représente dans les budgets de fonctionnement des laboratoires engagés dans la cartographie. Parmi les nombreuses applications, certaines, comme la PCR inter-Alu [23] et la PCR à grande échelle (*long range PCR*) [24], ont été déve-

RÉFÉRENCES

13. Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. A human DNA segment with properties of the gene that predispose to retinoblastoma and osteosarcoma. *Nature* 1986 ; 323 : 643-6.
14. Monaco AP, Neve RL, Coletti-Feener C, Bertelson CJ, Kurnitt DM, Kunkel LM. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* 1986 ; 323 : 646-50.
15. Group HDJR. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993 ; 72 : 971-83.
16. Vincent A, Heitz D, Petit C, Kretz C, Oberle I, Mandel JL. Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. *Nature* 1991 ; 349 : 624-6.
17. Mandel JL, Heitz D. Molecular genetics of the fragile-X syndrome: a novel type of unstable mutation. *Curr Op Genet Dev* 1992 ; 2 : 422-30.
18. Willems PJ. Dynamic mutations hit double figures. *Nature Genet* 1994 ; 8 : 213-5.
19. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, et al. (15 auteurs). A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994 ; 371 : 130-6.
20. Hashimoto L, Habita C, Beressi JP, Delphine M, Besse C, et al. (14 auteurs). Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. *Nature* 1994 ; 371 : 161-4.
21. Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987 ; 236 : 806-12.
22. Saiki RK, Scharf SJ, Faloona F, Mullis KB, Horn CT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin sequences and restriction site analysis for diagnosis of cycle cell anemia. *Science* 1985 ; 230 : 1350-4.
23. Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH, Caskey CT. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci USA* 1989 ; 86 : 6686-90.
24. Cheng S, Chang SY, Gravitt P, Respass R. Long PCR. *Nature* 1994 ; 369 : 684-5.

loppées pour des études cartographiques. Le recours à la PCR inter-Alu est devenu quasi systématique dans un travail de cartographie physique d'une région du génome humain.

Il est vraisemblable que sans ces trois piliers, les objectifs du premier plan quinquennal du programme génome américain n'auraient pas été atteints. Les grands programmes de cartographie sont des utilisateurs effrénés de la PCR et il est maintenant courant de trouver dans les «centres génomes» ou des entreprises privées des thermocycleurs réalisant 1 600 ou 3 200 réactions simultanément.

Des rails sur la prairie

Les premiers succès laissent penser que tous les gènes de maladies génétiques se laisseraient découvrir à l'identique. Or, cet isolement fut particulièrement laborieux et il était impératif que de grands travaux d'infrastructure prennent le relais des efforts anarchiques et quelque peu dérisoires qui trouvaient surtout leur légitimité dans la gravité des maladies abordées.

Une première carte génétique financée par le secteur privé avait été établie dès 1987 [25]. Malheureusement, cette carte, surtout constituée de marqueurs RFLP bi-alléliques peu informatifs, n'était pas vraiment adaptée à la cartographie de la grande majorité des maladies monofactorielles pour lesquelles la rareté des familles constitue un facteur limitant.

Une carte génétique de deuxième génération [26] et une carte physique [27], financées pour l'essentiel par l'Association française contre les myopathies, sont en cours d'achèvement. La carte génétique de Généthon a déjà provoqué une accélération impressionnante et mesurable du nombre et de la précision des localisations génétiques. Avec la carte physique CEPH-Généthon, même incomplète, la phase d'identification de nombreux gènes de maladies a changé de visage. Les contributeurs américains peuvent être reconnaissants envers les donateurs français du Téléthon qui ont permis d'atteindre les objectifs du plan quinquennal mentionné plus haut. Peut-être plus qu'en termes d'accélération, est-ce

en termes d'économie de moyens qu'il faudra évaluer l'impact de ces réalisations cartographiques.

D'autres efforts plus ciblés, essentiellement américains (*m/s n° 8-9, vol. 10, p. 898*), viennent utilement compléter et préciser les cartes existantes. La masse de données cartographiques accumulées dans le monde entier est sans précédent. Leur gestion et leur distribution posent de sérieux problèmes d'intendance auxquels les bases de données internationales, comme GDB (*genome data base*), essaient de faire face. Cette tâche est d'autant plus ardue que la nature des besoins évolue d'année en année. Par ailleurs, l'accès à ces bases de données demeure souvent laborieux. En France notamment, de nombreux laboratoires et centres hospitaliers ne sont pas, ou mal, connectés aux réseaux informatiques internationaux.

Le futur immédiat

L'infrastructure cartographique qui a fait défaut si longtemps et dont nous disposons maintenant est encore largement perfectible. Plusieurs dizaines de milliers d'ADNc partiellement séquencés et disponibles sous forme d'EST (*expressed sequence tags*) peuvent dès à présent être intégrés dans les cartes existantes. Cela devrait beaucoup faciliter l'identification de gènes de maladies. En matière de cartographie, il a toujours été préférable de déterminer des positions et des ordres par plusieurs méthodes indépendantes. C'est pourquoi, deux voies complémentaires de cartographie d'EST se dessinent à l'heure actuelle. Une première consiste à assigner ces EST à des YAC d'une collection ordonnée couvrant la carte physique CEPH-Généthon. La deuxième s'appuie sur l'analyse d'hybrides d'irradiation [28] représentant l'ensemble du génome. Ces hybrides renferment des fragments de chromosomes humains obtenus par irradiation de cellules humaines par les rayons X puis fusionnées avec des cellules de rongeur. La fréquence de cassures entre deux marqueurs (mesurée par la fréquence de leur ségrégation indépendante dans différents hybrides) reflète la distance physique les séparant. Cette mesure permet aussi de proposer des ordres entre marqueurs fondés sur une éva-

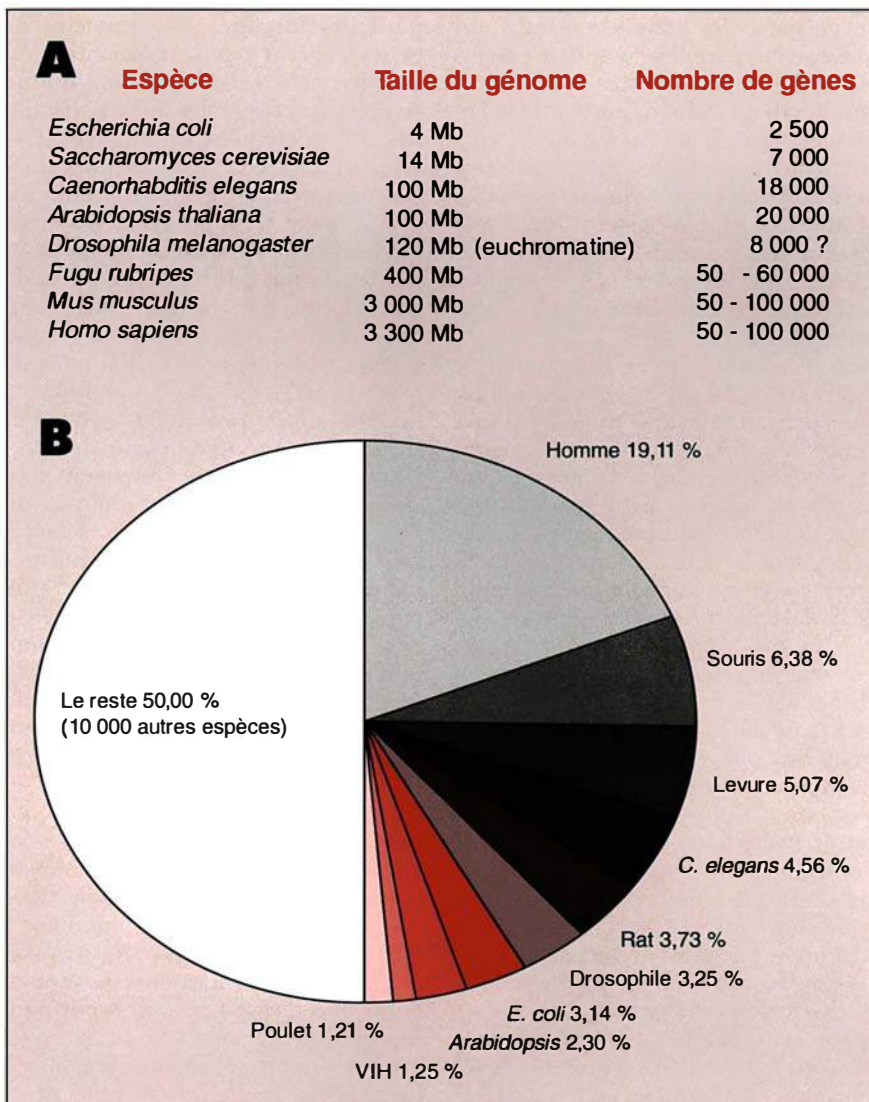


Figure 1. **A** : Les génomes d'organismes modèles. **B** : Quantité de données de séquences dans les bases de données EMBL GENBANK.

luation statistique. Des collections de ces hybrides sont actuellement caractérisées par des marqueurs régulièrement espacés le long de la carte génétique. Il devient donc possible de positionner n'importe quel EST par rapport à ces marqueurs polymorphes par analyse des hybrides. D'une précision de l'ordre de la mégabase, cette méthode, fondée sur une analyse statistique des résultats, présente l'avantage d'une plus grande robustesse et permet de couvrir l'ensemble du génome.

Des progrès remarquables dans la manipulation des molécules d'ADN et des fibres de chromatine permet-

tent de les observer au microscope avec des résolutions de l'ordre de la kilobase [29]. Il est ainsi devenu possible, à l'aide de l'hybridation *in situ* fluorescente, d'ordonner des fragments d'ADN les uns par rapport aux autres et de mesurer des distances entre fragments. L'observation directe des fibres est encore laborieuse mais la possibilité d'aligner des molécules d'ADN côte à côte a été rapportée très récemment [30]. Si la mise en œuvre de ces méthodes se banalise, elles ont de grandes chances de se substituer aux méthodes actuelles de cartographie. Moyennant l'introduction d'une relative automatisation

couplée à l'analyse d'images, elles pourraient, à terme, permettre un examen immédiat de l'ensemble d'un caryotype avec une résolution de l'ordre de la kilobase et de déceler ainsi une foule de micro-remaniements chromosomiques qui restent à mettre en évidence dans de nombreuses situations pathologiques.

Et la séquence...

Dans la mesure où seule la séquence génomique donnera un inventaire complet des gènes, il faudra d'une manière ou d'une autre passer à cette forme ultime de la carte. Les programmes de séquençage d'organismes modèles (figure 1), actuellement en cours, progressent à pas de géants. Pour de petits génomes, comme celui de la levure ou des bactéries, le travail est réparti entre des laboratoires académiques de dimension classique qui consacrent une partie de leurs moyens à une œuvre collective. Le séquençage du génome de levure, commencé par le chromosome III [31], est à mi-chemin, et devrait se terminer fin 1996. Les extrapolations au génome complet, fondées sur les chromosomes de levure séquencés, laissent prévoir l'identification d'environ 3 000 gènes aux fonctions non prédictibles sur un total de 7 000. Sur ce même total, près de 3 000 gènes pourraient être redondants. La connaissance du génome de la levure va susciter une quantité de nouvelles expériences et d'analyses informatiques qui se poursuivront pendant de nombreuses années.

Une telle répartition du travail se prête au séquençage des micro-organismes, d'autant plus que l'identification d'un nouveau gène peut parfois susciter, à chaud, une série d'expériences biologiques rapides menées dans le même laboratoire. Le séquençage de génomes plus importants nécessite, en revanche, des structures *ad hoc* constituées d'unités de séquençage avec une division du travail poussée. C'est ainsi que la séquence du génome de *Caenorhabditis elegans* est déterminée dans deux grandes structures (à Saint-Louis, Missouri, USA et au Centre Sanger, Hinxton, Grande-Bretagne) qui devraient produire chacune, annuellement, de l'ordre d'une dizaine de

RÉFÉRENCES

25. Donis-Keller H, Green P, Helms C, Carlinhour S, Weiffenbach B, *et al.* (33 auteurs). A genetic linkage map of the human genome. *Cell* 1987 ; 51 : 319-37.
26. Gyapay G, Morissette J, Vignal A, *et al.* The 1993-1994 Génethon human genetic linkage map. *Nature Genet* 1994 ; 7 : 246-339.
27. Cohen D, Chumakov I, Weissenbach J. A 1st-generation physical map of the human genome. *Nature* 1993 ; 366 : 698-701.
28. Cox DR, Burmeister M, Price ER, Kim S, Myers RM. Radiation hybrid mapping: a somatic cell genetics method for constructing high resolution maps of mammalian chromosomes. *Science* 1990 ; 250 : 245-50.
29. Houseal TW, Klinger KW. What's in a spot. *Hum Mol Genet* 1994 ; 3 : 1215-6.
30. Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D. Alignment and sensitive detection of DNA by a moving interface. *Science* 1994 ; 365 : 2096-8.
31. Oliver SG, Van der Aart QJM, Agostoni-Carbone ML, *et al.* (136 auteurs de 38 laboratoires). The complete DNA sequence of yeast chromosome III. *Nature* 1992 ; 357 : 38-46.
32. Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet* 1994 ; 6 : 84-9.
33. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachibiri Y, Simon M. Cloning and stable maintenance for 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* 1992 ; 89 : 8794-7.
34. Uberbacher EC, Mural RJ. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci USA* 1991 ; 88 : 11261-5.
35. Auffray C, Behar G, Bois F, *et al.* Analyse intégrée des transcrits du génome humain. *CR Acad Sci Paris* 1995 (sous presse).
36. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE. Fluorescence detection in automated DNA sequence analysis. *Nature* 1986 ; 321 : 674-9.0
- mégabases. Ces grands centres poursuivent en parallèle d'autres projets de séquençage. D'autres laboratoires publics de grand séquençage sont envisagés aux États-Unis où ont également fleuri plusieurs entreprises privées consacrées au séquençage massif d'ADNc. Les biologistes vont avoir besoin de ces grands outils, de même que la physique des hautes énergies a eu besoin d'accélérateurs de particules.
- Avec l'amélioration de la technologie courante de séquençage, se concrétisant par la mise sur le marché d'une nouvelle génération de machines à haut débit capables d'analyser plusieurs centaines de réactions par jour, les États-Unis envisagent sérieusement de se lancer dans le séquençage en grand du génome humain. Mais la collection de fragments clonés et ordonnés à séquencer n'est pas encore réalisée. Les YAC sont satisfaisants pour le clonage positionnel mais leur taux élevé de réarrangements oblige à recourir à une autre source pour le séquençage. On s'oriente de plus en plus vers des vecteurs bactériens à réplication sous contrôle strict, qui se maintiennent à raison d'une à deux copies par cellule bactérienne. En première analyse, il semble que de tels vecteurs soient moins sujets à réarrangements et il y a des raisons de penser que les séquences réputées inclonables dans les vecteurs à nombre élevé de copies puissent être propagées à l'aide de ces nouveaux vecteurs. Des banques dont les insertions ont une taille moyenne supérieure à 100 kb ont été construites dans des vecteurs dérivés, soit du phage P1 (PAC, P1 *artificial chromosome*), soit du facteur F (BAC, *bacterial artificial chromosome*) [32, 33], mais l'ordonnement de ces clones reste à réaliser.
- L'inventaire complet des gènes dépend aussi d'une analyse par des programmes informatiques capables de détecter la totalité des exons avec une fiabilité supérieure à celle atteinte par des programmes actuels de recherche de séquences codantes comme GRAIL [34]. Dans l'attente de cet inventaire, d'autres approches vont s'avérer utiles. Les entreprises de séquençage massif d'ADNc ont constitué des bases de données privées contenant plusieurs centaines de milliers de fragments correspondant à plusieurs dizaines de milliers de gènes. L'accès à cette information est possible mais à des conditions qui ne semblent acceptables que pour des besoins strictement scientifiques. Des efforts publics de moindre importance, notamment la contribution du programme Genexpress avec le séquençage de plus de 8 000 clones d'une banque normalisée d'ADNc de cerveau de nouveau-né, constituent néanmoins à ce jour la plus grande source librement accessible de séquences d'ADNc [35]. Il importe maintenant de positionner ces gènes sur les cartes génomiques.
- D'autres alternatives pourraient aussi faciliter l'inventaire génique de petites régions. Le génome du Fugu, huit fois plus compact que celui des mammifères, semble montrer une conservation de la synténie entre gènes trouvés proches chez les mammifères (*m/s n° 11, vol. 10, p. 1154*). Le séquençage de petits segments de Fugu pourrait contribuer à l'inventaire de régions environ huit fois plus étendues d'un génome de mammifère. Il a aussi été proposé de procéder à un séquençage partiel (*shallow sequencing* ou *skimming sequencing*) d'un ou deux équivalents de régions génomiques humaines. L'analyse informatique de ces séquences devrait permettre le repérage rapide d'une fraction importante des exons.
- Le séquençage complet de fragments de génome humain de l'ordre de la mégabase est actuellement en cours dans plusieurs centres. Toutefois, à ce jour, l'émergence espérée de nouvelles techniques de séquençage n'a pas véritablement eu lieu. Le séquençage par hybridation, très séduisant, sera plutôt utilisé pour le séquençage massif de fragments de séquence connue, dans un but de diagnostic, de recherche de variants ou pour des analyses de populations de séquences. Il est donc vraisemblable que dans les prochaines années le séquençage *de novo* à grande échelle se fasse à l'aide des techniques existantes améliorées [36].

L'après-génome

L'exploration systématique du génome va produire une énorme quantité de données dont l'utilité dépendra avant tout de la qualité. La cartogra-

phie s'appuie sur des manipulations d'une reproductibilité imparfaite, qui ne sont pas toujours dupliquées en raison du nombre trop élevé d'expériences à réaliser. La qualité des réactifs biologiques est aussi sujette à caution. Les utilisateurs connaissent ces difficultés et sont en général capables de tirer le meilleur de l'information existante et de faire les corrections qui s'imposent. De plus, la multiplication des données débouche de plus en plus rapidement sur la détection de résultats discordants qui apparaissent donc d'emblée comme suspects. Les exigences de qualité sont plus déterminantes en matière de séquençage, mais le niveau de reproductibilité atteint assure une fiabilité satisfaisante des résultats.

Avec l'analyse de grands génomes, la biologie est entrée dans une phase d'acquisition massive de données. Les besoins d'observations systématiques, bien plus complexes que les analyses structurales actuelles, telles que l'observation de phénotypes mutés et, d'une manière plus générale, la compréhension de milliers de fonctions, l'analyse des systèmes multifactoriels, etc., vont se multiplier. La levure servira à nouveau de précurseur. Les expériences à 50 ou 100 chercheurs vont se banaliser. Il sera intéressant de voir comment sera acceptée la nécessaire division du travail.

Les criblages systématiques sont devenus inévitables. Ils permettront d'enregistrer des paramètres bien établis, mais il importera plus que jamais de ne se lancer dans une expérience qu'après de solides mises au point. Le recueil d'observations *a priori* bien définies requiert une attention toute particulière : on a vite fait de laisser échapper un détail crucial ne faisant pas partie du schéma expérimental prévu. Rappelons-nous que l'établissement des grands inventaires en zoologie et botanique aux XVIII^e et XIX^e siècles a été fait par les plus grands experts de l'époque et qu'il est le fruit d'un travail minutieux et d'une remise en question fréquente des critères de classification. Il se pose en outre un problème logistique : les structures existantes sont inadéquates pour le recueil massif de données et les moyens requis bien au-delà du budget des organismes de recherche.

Détracteurs et bonimenteurs

Le domaine du génome se caractérise par la dilution de quelques idées motrices parfois banales dans une masse énorme de travail extrêmement répétitif, soit exactement l'inverse du rêve de tout scientifique. On comprend donc la réticence, voire le mépris, de nombreux biologistes pour le projet génome : les progrès de la technologie s'accompagnent d'une mise en veilleuse de la pensée. Mais nous ne ferons pas l'économie du génome. La controverse née lors du débat sur le séquençage du génome humain, largement fondée sur une série de malentendus entre biologistes, tend aujourd'hui à s'estomper. L'histoire la classera sans doute parmi les avatars de la querelle des anciens et des modernes. Un consensus plus ou moins tacite semble s'être établi autour de quelques vérités premières qui admettent que le génome n'est pas une fin en soi, mais que l'acquisition systématique de cette information représentera en fin de compte un gain de temps et de moyens. Le génome est devenu une condition nécessaire, mais non suffisante, de la biologie et de la médecine modernes.

Des chiffres d'affaires colossaux dans le domaine des applications de la génétique humaine ont été avancés par les faiseurs de prospectives. Ils paraissent singulièrement surévalués et les contraintes de temps globalement ignorées. Le progrès scientifique est certes imprévisible, mais les développements sont beaucoup plus lents que les prévisions ne l'imaginent. Les applications, dans un domaine aussi complexe que le vivant, nécessitent à chaque instant la résolution de questions fondamentales pour lesquelles les connaissances sont simplement absentes. De surcroît, la nécessaire sécurité pour les applications humaines ne fera que multiplier le nombre nécessaire d'expériences de contrôle préalable. Il est donc illusoire de croire que les thérapies nouvelles capables de soigner le diabète (pour lequel la plupart des gènes impliqués restent à identifier), l'obésité, etc., sont pour l'aube du prochain millénaire. N'oublions pas qu'il s'est écoulé quatre-vingts longues années entre la réfutation de la génération

spontanée et la découverte des antibiotiques ■

Summary

The human genome between medicine and science

The molecular move of human genetics stands back to the early eighties and the first results were obtained despite the lack of any genetic map. Genetic maps being now available, the pace of knowledge acquisition has considerably speeded up. New tools, such as cloning in yeast artificial chromosomes allowing the manipulation of large DNA fragments, or *in vitro* amplification of DNA (PCR), greatly facilitated this research. With the analysis of large genomes, biology is now acquiring massive lots of data, raising new logistic problems. Actual research structures are not yet appropriate for dealing with such a mass of data and are not funded adequately. The major fallout of these programs concerning basic research as well as medical applications will not be available before next century.

Remerciements

Je remercie Sabine Fauré, Sophie Marc et Monique Meugnier pour leur aide dans la préparation de ce manuscrit.

TIRÉS À PART

J. Weissenbach.