

## LE GENOME DE *MYCOBACTERIUM LEPRAE* : DE L'ANALYSE DE LA SEQUENCE AUX ENJEUX THERAPEUTIQUES

N. HONORE

• Travail de l'Unité de Génétique Moléculaire Bactérienne (N.H., Ingénieur de recherche), Institut Pasteur, 28, rue du Docteur Roux, 75724 Paris Cedex 15 •  
Tel. : +33 (0) 1 40 61 37 19 • Fax : +33 (0) 1 40 61 35 83 • E-mail: nhonore@pasteur.fr •

*Med. Trop.* 2002 ; **62** : 473-479

En 1873, le Docteur G.H.A. Hansen découvrit que la lèpre est une maladie contagieuse transmise par *Mycobacterium leprae*. Il établit ainsi, pour la première fois dans l'histoire, la relation entre une bactérie et une maladie infectieuse humaine. Il n'imaginait pas à l'époque que, plus d'un siècle plus tard, cette maladie très invalidante poserait toujours, dans de nombreux pays, un important problème de santé publique. En effet, l'incidence de cette maladie reste encore aujourd'hui très inquiétante, avec plus de 690 000 nouveaux malades enregistrés chaque année (1). Et ce, malgré la mise en place par l'OMS, à partir de 1982, de la polychimiothérapie (PCT) (2) et de la vaccination par le BCG (3, 4), qui ont considérablement réduit la prévalence de la lèpre. Celle-ci est à présent estimée à environ 1 à 2 millions de cas.

La lèpre est curable et elle peut passer inaperçue, mais le plus souvent elle se caractérise chez les malades par des lésions cutanées qui mènent parfois à des mutilations très invalidantes. *M. leprae* se multiplie préférentiellement aux extrémités du corps, dans les macrophages et les cellules de Schwann du système nerveux périphérique. Les cellules de Schwann infectées perdent alors la gaine de myéline qui les entoure et sont détruites par les défenses immunitaires de l'hôte. Ceci provoque la destruction des nerfs, et la perte de la sensibilité qui ont pour conséquence les stigmates irréversibles bien connus. Actuellement, 4 % des nouveaux patients présentent déjà des séquelles neurologiques qui dans la plupart des cas sont irréversibles, alors qu'un diagnostic et un traitement précoces de la maladie auraient suffi à éviter ou au moins à limiter ce phénomène. Statistiquement, pour 70 % des patients, le diagnostic de lèpre est établi

grâce à l'observation clinique qui révèle la présence de taches cutanées accompagnées d'une perte de la sensibilité ou de l'augmentation du volume des nerfs périphériques. Cependant, 30 % des malades, et parmi eux beaucoup de multibacillaires, ne présentent pas ce symptôme. Bien que l'examen de frottis, en laboratoire, complète l'examen clinique et classe les différents cas de lèpre suivant le nombre de bacilles présents en lèpre multibacillaire (MB) ou paucibacillaire (PB), il est important d'établir de nouveaux tests sérologiques ou cutanés, afin d'être certain du diagnostic et d'identifier tous les cas infectieux. Ainsi, la mise au point de tests de diagnostic immunologique, la recherche de traitements de plus en plus performants mais aussi la découverte d'un nouveau vaccin devraient mener à l'éradication de la lèpre, qui est un des principaux objectifs de l'OMS.

### Etude du génome de *M. leprae*

*M. leprae* est un pathogène intra-cellulaire obligatoire qui ne se multiplie pas en milieu axénique. Cette propriété a, à elle seule, limité considérablement les recherches, qui auraient sans doute permis d'élucider de nombreux mystères entourant ce bacille, à commencer par son tropisme particulier: les cellules de Schwann.

Néanmoins, la découverte de différents hôtes animaux a permis l'étude de *M. leprae* en laboratoire. En effet, Shepard a montré en 1960 que *M. leprae* pouvait se multiplier dans les coussinets plantaires de la souris, mais avec un temps de génération très lent estimé à 14 jours (5). La souris est encore utilisée de nos jours pour tester la sensibilité des bacilles aux antibiotiques et contrôler les traitements (6). Cependant,

c'est chez le tatou (7) que l'on a réussi à produire une quantité de bacilles suffisante pour permettre la réalisation des premières études biochimiques et la purification de l'ADN. Et c'est grâce à cette avancée que les études génétiques chez *M. leprae* ont pu être entreprises et que l'on a pu estimer la taille de son génome et son contenu en G+C (8). Les valeurs obtenues à l'époque (G+C 56 % et taille 3,38 mega bases (Mb)) se révélèrent d'ailleurs très proches de celles obtenues à partir du séquençage du génome effectué quelques années plus tard (57,8 % et 3,268 Mb) (9). Parallèlement à ces découvertes, les énormes progrès réalisés au niveau des techniques de biologie moléculaire, notamment les techniques de donage et de séquençage à haut débit de fragments d'ADN, couplés à l'analyse par bio-informatique, ont donné un nouvel essor aux recherches sur la lèpre.

### Séquençage du génome

Le séquençage du génome de *M. leprae* a été réalisé par étapes successives: d'abord la purification de l'ADN à partir des bacilles, puis le sous-clonage de fragments d'ADN de grande taille dans des cosmides ou de taille aléatoire dans un plasmide («Shotgun» du génome), et enfin le séquençage à haut débit des clones.

La première étape a donc été la purification d'ADN à partir de bacilles isolés chez un patient de Tamil Nadu (TN), et multipliés chez le tatou. Cette préparation a été difficile et l'ADN purifié n'était pas de très bonne qualité, car déjà légèrement dégradé. Ceci résultait certainement des traitements violents utilisés pour l'extraction, traitements dus à la présence chez *M. leprae* d'une enveloppe très résistante, mais pouvait aussi résulter de l'existence d'une

# Avancée

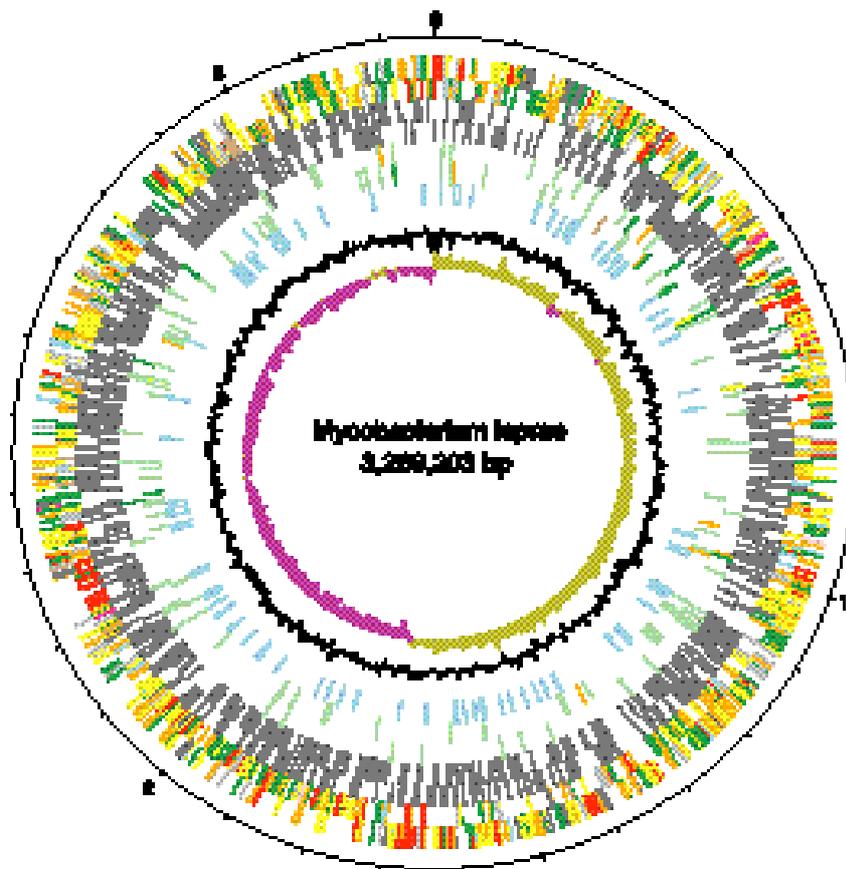


Figure 1 - Carte circulaire du génome de *M. leprae* montrant la position et l'orientation des gènes connus, des pseudogènes et des séquences répétées. Depuis l'extérieur, les cercles 1 et 2 correspondent respectivement aux gènes présents sur les brins moins et plus du chromosome; les cercles 3 et 4 correspondent aux pseudogènes; 5 et 6 aux gènes spécifiques à *M. leprae*; 7 aux séquences répétées; 8 au contenu en G+C, 9 au rapport (G+C)/(G-C). Les gènes ont une couleur différente suivant leur fonction: métabolisme des lipides en noir; respiration et métabolisme intermédiaire en jaune; machinerie cellulaire en rouge; protéines régulatrices en bleu clair; protéines hypothétiques conservées en orange; protéines ayant une fonction inconnue en vert clair; séquences d'insertion et fonctions relatives aux phages en rose; ARN en bleu foncé; paroi cellulaire en vert; les familles de protéines PE et PPE en magenta; adaptation, virulence et détoxification en maron. Les nombres à l'extérieur indiquent l'échelle en Mégabases.

population de bacilles hétérogènes chez le tatou dont l'ADN était déjà plus ou moins endommagé. Néanmoins cet ADN fut digéré par des enzymes de restriction, et des fragments de grande taille (~40 kilo bases (kb)) furent clonés dans différents cosmides. Les profils de restriction d'environ 1000 clones indépendants, issus du clonage dans le cosmide *Loist6*, furent comparés et analysés par la technique du «fingerprint» (10, 11). Une carte physique du chromosome, représentée par les cosmides chevauchants, fut donc établie et on put réaliser l'ébauche d'une carte génétique en positionnant sur cette carte physique différents gènes et loci de *M. leprae* préalablement identifiés. Cependant, cette carte intégrée ne représentait pas la totalité du chromosome, elle était en effet constituée de 4 contigs non reliés entre eux, et les tentatives pour la compléter à l'aide d'une autotransformation cosmique, construite dans le vecteur navette pYUB18 (12), échouèrent. Vraisemblablement, quelques régions chromosomiques, peut-être sous-représentées

dans les préparations d'ADN, étaient impossibles à cloner dans les cosmides. Toutefois, ces régions étaient présentes dans les plasmides recombinants générés lors du «shotgun» du génome. Ce clonage fut réalisé globalement de la façon suivante: l'ADN de *M. leprae* (TN) fut fractionné de façon aléatoire en fragments de petites tailles (~2 à 3 kb) et ces fragments furent donés dans le plasmide pUC18.

La séquence totale du génome de la souche TN de *M. leprae*, quant à elle, fut obtenue en combinant la séquence de cosmides sélectionnés et la séquence de la totalité des clones obtenus par «shotgun» (9). Celle-ci fut ensuite soumise à l'analyse bio-informatique.

## Analyse de la séquence

Le chromosome de *M. leprae* est circulaire. Il contient 3 268 203 paires de bases avec, en moyenne, un pourcentage de G+C de 57,8 et l'analyse bio-informatique a

montré que 1605 gènes codent pour des protéines. On constate que ces valeurs sont beaucoup plus basses que celles trouvées dans le génome de *M. tuberculosis*, où environ 4000 gènes ont été identifiés pour une taille de 4 411 532 avec un G+C de 65,6 % (13).

Déjà, lors de l'analyse de la séquence du premier cosmide, le B1790 (14), il avait été montré que seule la moitié de la séquence était potentiellement codante. Cela semblait très étonnant, mais à l'époque aucun génome mycobactérien n'avait encore été séquencé et les comparaisons se limitaient aux séquences des gènes d'autres espèces. En revanche, lorsque la séquence totale de *M. leprae* fut terminée, la séquence de *M. tuberculosis* était disponible et les résultats des comparaisons entre les génomes et protéomes de ces deux pathogènes furent très informatifs (15). On remarqua alors que 49,5 % du génome de *M. leprae* contiennent des phases de lectures codantes et que dans les 50,5 % restants, 27 % contiennent des pseudogènes

# Avancée en Avancée

Tableau I - Comparaison des caractéristiques des deux génomes.

	<i>Mycobacterium leprae</i>	<i>Mycobacterium tuberculosis</i>
Taille du génome (pb)	3 268 203	4 411 532
G+C (%)	57,79	65,61
Séquences codantes (%)	49,5	90,8
Nombre de gènes	1605	3959
Nombre de pseudogènes	1116	6
Taille moyenne des gènes (pb)	1011	1012
Taille moyenne des gènes inconnus (pb)	338	653

vraisemblablement dépourvus de fonction mais identifiables grâce à la comparaison avec *M. tuberculosis*, où ils sont encore parfaitement fonctionnels. Les 23,5 % de séquences restantes ne paraissent pas être codantes. Le chromosome comporte donc 1116 pseudogènes, qui semblent être répartis au hasard, et 1605 gènes potentiellement actifs organisés en clusters flanqués de longues régions non-codantes (Fig. 1). Parmi ces gènes, 1439 se trouvent chez les deux pathogènes, et 165 sont propres à *M. leprae*. L'analyse informatique montre que dans ce dernier groupe, 29 gènes ont une fonction probable alors que les 136 autres ne présentent aucune similarité avec des gènes connus. Certains de ces derniers sont peut-être des pseudogènes dont on ne connaît pas l'ancêtre, et ce dans la mesure où ils sont plus petits que la moyenne des gènes (9). La comparaison des principales caractéristiques des deux génomes est indiquée dans le tableau I.

## Evolution réductive

Si on considère que toutes les mycobactéries descendent d'un ancêtre commun dont la taille du génome devait être environ de 4,4 Mb, on constate que la réduction de celui de *M. leprae*, qui a dû avoir lieu au cours de l'évolution, est d'une importance significative. Si tout le génome de *M. leprae* et celui de *M. tuberculosis* étaient codants, on attendrait 3000 protéines pour le premier et 4000 pour le second. Les premières analyses du protéome ont mis en évidence la présence de 391 protéines solubles chez *M. leprae* (16) et presque 1800 chez *M. tuberculosis* (17). Ceci semble confirmer que les pseudogènes sont inertes. *M. leprae* aurait perdu quelques 2000 gènes et l'évolution réductive aurait défini l'ensemble minimum de gènes nécessaires à la survie de ce pathogène intracellulaire. La perte d'information génétique à grande échelle a été précédemment décrite chez des parasites intracellulaires obligatoires tel que *Rickettsia* (18) ou des bactéries symbiotiques telle que *Buchnera spp* (19). Chez celle-ci la taille du génome a été réduite de

~4,5 à 0,64 Mb par de grandes délétions. Cependant, contrairement à *M. leprae*, on note la présence d'un petit nombre de pseudogènes. Les auteurs suggèrent que les gènes sont inactivés à partir du moment où leur fonction n'est plus utilisée, car ces bactéries colonisent des niches très spécialisées. De plus, ces organismes n'ayant pas de cycles sexués, les événements de recombinaison avec des ADN étrangers sont peu probables, hormis peut-être avec l'ADN de leur hôte. Ce transfert horizontal de gène semble avoir eu lieu chez *M. leprae* au niveau de *proS*, qui code pour une prolyl-tRNA synthétase. En effet les comparaisons de séquences montrent une similarité plus grande avec les enzymes eucaryotes telles que celle de l'homme et de la drosophile, qu'avec l'enzyme de *M. tuberculosis*. Au contraire de *M. tuberculosis* H37Rv, qui contient au moins deux prophages et 56 séquences d'insertion plus ou moins intactes (IS) (13, 20), *M. leprae* ne contient que trois gènes ayant très certainement une origine phagique, et 26 vestiges de transposons. La perte d'ADN répété, tel que les IS, est également un signe de l'évolution réductive.

Bien que l'on ne connaisse pas les mécanismes mis en œuvre pour la formation des pseudogènes, on suppose que l'inactivation de certains gènes qui sont impliqués dans la recombinaison, la réplication de l'ADN et le maintien de l'intégrité du génome, joue un rôle important. Chez *M. leprae* le gène *dnaQ*, qui code pour la sous-unité  $\beta$  de l'ADN polymérase III dont l'activité de «relecture» vérifie la fidélité de la réplication (21), est délété. Il en est de même pour les gènes *mutL* et *mutS* qui, eux, participent à la réparation des mésappariements. Aucun système connu ne corrige les erreurs de réplication chez *M. leprae* et la voie de recombinaison générale, catalysée par l'enzyme RecBCD, a aussi été éliminée. De plus, il semble tout à fait évident que des événements de recombinaison homologue au niveau des séquences répétées ont été responsables à la fois des délétions et du réaménagement du génome de *M. leprae* par rapport à celui de *M. tuberculosis*. Si l'on compare les deux génomes, on peut délimiter 65 segments qui sont presque super-

posables mais arrangés différemment. Chez *M. leprae*, l'ordre des gènes est interrompu par des séquences répétées, par des gènes d'ARN de transfert, ou par des régions dépourvues de gènes. Globalement, 2 % du génome sont composés de séquences répétées. Ces dernières ont été classées en 4 familles: RLEP (37 copies), REPLEP (15 copies), LEPREP (8 copies), LEPRPT (5 copies). Bien qu'aucune homologie de séquence n'ait été trouvée avec des éléments transposables, il semble que les séquences RLEP soient capables de dissémination car on les trouve de façon prédominante à l'extrémité 3' des gènes mais aussi à l'intérieur de certains pseudogènes (22).

## Familles multigéniques

Il existe chez *M. tuberculosis* une véritable redondance fonctionnelle : en effet la moitié des gènes sont issus d'événements de duplication génique et de remaniement (15). On compte ainsi 213 familles avec 2 membres et 72 avec 3 membres, alors que chez *M. leprae* seulement 87 et 37 familles ont respectivement 2 et 3 membres. *M. tuberculosis*, quant à lui, contient un nombre très important de gènes qui sont impliqués dans le métabolisme des lipides ou qui codent pour deux nouvelles familles de protéines, les PE et PPE (23). Les protéines de la famille PE, qui compte environ 100 membres, ont un domaine N-terminal commun de 110 acides aminés avec en particulier les motifs Pro-Glu (PE) aux positions 8 et 9. Cette famille se divise en trois groupes dont les membres sont classés suivant la nature de leur domaine C-terminal. Le groupe le plus important est celui des PE-PGRS (polymorphic GC-rich sequence) qui se composent de motifs répétés (Gly-Gly-Ala). Les membres de la famille PPE ont, eux, un domaine N-terminal commun de 180 acides aminés avec le motif Pro-Pro-Glu (PPE) aux positions de 7 à 9, et sont au nombre de 68. De même que les protéines PE, les PPE sont composées, pour la plupart, de séquences répétées en tandem. Ces 167 gènes exceptionnellement riches en G+C et dont on ne connaît pas la fonction, occupent 8 % du génome de *M. tuberculosis*. Il est surprenant de constater que chez *M. leprae* la plupart des gènes qui codent pour les PE et PPE sont des pseudogènes, ils sont au nombre de 30, et que seulement neuf gènes PE et PPE sont intacts, dont aucun PE-PGRS, alors que tous les génomes mycobactériens séquencés en contiennent un nombre beaucoup plus important. Ainsi, la disparition de la plupart des membres de ces grandes familles de

# Avancée en Avancée

gènes riches en G+C peut expliquer la petite taille et le faible contenu en G+C du génome de *M. leprae* par rapport à celui de *M. tuberculosis*.

On peut enfin constater que les familles de protéines les plus importantes sont toujours celles impliquées dans le métabolisme des lipides et des polykétides, la synthèse de la paroi (transférases, glycosyltransférases), le transport des molécules (ABC transporteurs, les protéines MmpL) et la régulation des gènes (TetR, WhiB, systèmes à 2 composants). La liste de tous les gènes de *M. leprae* est accessible sur le site: [http://www.pasteur.fr/recherche/unites/Lgmb/NATURE\\_DATA\\_gene\\_list](http://www.pasteur.fr/recherche/unites/Lgmb/NATURE_DATA_gene_list). Pour obtenir le maximum d'informations concernant la séquence, il suffit de consulter la page suivante: <http://genolist.pasteur.fr/Leproma>. Sur Leproma, plusieurs critères de recherche sont utilisables: le nom ou la fonction du gène, ou encore, la région du génome (24), et il est aussi possible de comparer des séquences nucléotidiques ou protéiques en utilisant les programmes BLAST (25, 26) ou FASTA (27).

Comme tentent de le montrer les exemples suivants, l'analyse du génome a, dès à présent, apporté de nombreuses réponses aux questions posées concernant la physiologie de *M. leprae*. Toutefois, il ne s'agit ici que d'un bref résumé et non pas d'une liste exhaustive.

## Biosynthèse de l'enveloppe

L'enveloppe des mycobactéries est constituée d'une membrane cytoplasmique et d'une paroi qui se compose essentiellement d'un réseau de peptidoglycane, d'arabinogalactane et d'acides mycoliques. Chez *M. tuberculosis*, le peptidoglycane contient de l'alanine et il existe trois classes d'acides mycoliques, les  $\alpha$ , kéto et méthoxymycolates. En revanche, chez *M. leprae*, le peptidoglycane contient de la glycine et les acides mycoliques sont de deux types, les  $\beta$  et les kéto mycolates (28). Il est vrai que l'on ne connaît pas les effets de ces modifications sur les propriétés physiques de la paroi, mais l'analyse du génome a montré que les méthoxymycolates étaient absents parce que les gènes qui codent pour les enzymes MmaA2 et MmaA3, dont le rôle est d'attacher le groupe méthoxy chez *M. tuberculosis* (29), étaient déléétés.

La paroi joue un rôle très important dans la survie des mycobactéries pathogènes, surtout lors de leur multiplication dans les macrophages. Ainsi, il a été montré que la cyclopropanation des acides

mycoliques contenus dans la paroi par la CMAS-1 (cyclopropane mycolique synthase-1), protège le bacille tuberculeux du stress oxydatif (30). La séquence du génome de *M. leprae* indique que les gènes responsables des fonctions de propanation des acides mycoliques ont été déléétés. En effet, seuls les gènes *mmaA2* et *umaA2*, décrits chez *M. tuberculosis* (31) ont échappé à l'évolution réductive. Malgré cela, *M. leprae* peut se multiplier dans les macrophages, il doit donc disposer de l'arsenal nécessaire pour se protéger des oxydants létaux présents dans ces cellules. Le phénol-glycolipide PGLI devrait faire partie de cet arsenal, dans la mesure où il a été montré qu'il piègeait les radicaux hydroxyl et les anions superoxyde. Le PGLI est constitué d'un lipide phénolique, commun à toutes les mycobactéries, lié à un trisaccharide spécifique à *M. leprae*. Ce trisaccharide semble très important, d'une part il serait très immunogène, d'autre part il interviendrait dans la liaison de *M. leprae* avec les cellules de Schwann.

## Métabolisme

La production d'énergie ou catabolisme chez *M. leprae* est très réduite: un tiers des gènes fonctionnels chez *M. tuberculosis* a disparu. Que ce soit au niveau du cycle de Krebs ou de la glycolyse, les copies des gènes ont été déléétés. D'une part, et ce bien que *M. leprae* possède le cycle des glyoxylates intact, les voies anaplerotiques ont été perdues ou modifiées. L'enzyme malique, qui transforme le pyruvate en malate, est déléété et la pyruvate carboxylase, qui transforme le pyruvate en oxaloacétate, est remplacée par une phospho-enol-pyruvate carboxylase. D'autre part, les sources de carbone utilisables par *M. leprae* sont très réduites. Par exemple, ce bacille ne peut utiliser ni l'acétate (les gènes *ackA*, *pta* et *acs* sont déléétés), ni le galactose. Ceci implique une spécificité de milieu de culture très importante. Enfin, la chaîne respiratoire aérobie étant tronquée et la chaîne anaérobie étant inactive, la production d'ATP à partir de l'oxydation du NADH est impossible et la régénération du NAD doit être très limitée.

En revanche, *M. leprae* a gardé les principales voies anaboliques. Fait marquant pour un pathogène intracellulaire obligatoire les voies de biosynthèse des purines, et des pyrimidines sont intactes. Néanmoins, *M. leprae* pourrait être auxotrophe pour la méthionine puisque le gène *metC*, qui code pour la cystathionine

$\beta$ -lyase, est un pseudogène et devrait avoir une source organique de soufre, puisque les gènes qui codent pour le système de transport de cette molécule, *cysT*, *W* et *A* sont inactivés. Cette bactérie est aussi probablement auxotrophe pour la cobinamide, car il manque les gènes *cobD* et *U*, alors que les gènes nécessaires à la synthèse de la vitamine B12, à partir de la cobinamide, sont présents.

L'ensemble de ces observations montre à quel point *M. leprae* a besoin d'un milieu très particulier pour se multiplier. C'est pourquoi jusqu'à maintenant aucun milieu synthétique, permettant la culture de ce bacille, n'a pu être élaboré.

## Régulation

Les protéines régulatrices, de même que les fonctions cataboliques, sont très réduites chez *M. leprae*. En effet, elles sont au nombre de 46 alors que l'on en dénombre plus de 110 chez *M. tuberculosis*. Ceci suggère que *M. leprae*, avec des facultés d'adaptation restreintes, se développe très certainement dans une niche plus stable.

Il existe des différences majeures entre les bacilles lépreux et les bacilles tuberculeux au niveau de la reconnaissance des promoteurs et de l'initiation de la transcription. Ces différences peuvent être expliquées par leurs caractéristiques physiologiques. Les mécanismes concernés mettent en jeu les facteurs sigma de l'ARN polymérase. On compte 13 facteurs sigma appartenant à la famille sigma 70, dont 10 font partie de la sous-famille ECF (extra-cytoplasmic function), chez *M. tuberculosis*. Il a été montré récemment que la capacité de ce bacille à se multiplier à 37°C est contrôlée par SigB, à travers les facteurs sigma ECF E et H. De plus, la perte du gène *sigE* affecte la résistance au SDS et au stress oxydatif des bactéries, ainsi que leur survie dans les macrophages (32, 33). On constate chez *M. leprae* l'existence de pseudogènes pour la plupart des facteurs sigma, exceptés SigA, SigB et les 2 sigma ECF, SigC et SigE. On peut donc imaginer que la diminution de l'expression de SigB due à la perte de SigH fait que *M. leprae* se multiplie à des températures plus basses que *M. tuberculosis* et colonise en conséquence préférentiellement les parties froides du corps humain telle que la peau. Enfin, on remarque à nouveau que cette bactérie a gardé le minimum de gènes qui lui permettent de s'adapter au milieu dans lequel elle doit survivre, notamment les macrophages.

## Pathogénicité

Pour certaines bactéries, la combinaison de la génomique et de l'analyse informatique ont permis l'identification de facteurs de virulence, le meilleur exemple étant le cas du bacille de la peste *Yersinia pestis* (34). Dans ce cas, l'examen de la séquence à l'aide de logiciels informatiques d'analyse utilisant le biais des dinucléotides par exemple (35, 36), a montré l'existence des îlots de pathogénicité. Cependant, la même analyse chez *M. leprae* n'a donné aucun résultat, et de tels îlots n'ont pas été identifiés. De plus, les gènes qui, chez *M. tuberculosis*, pourraient coder des facteurs de virulence tel que l'opéron Mce (15, 37), ont été réduits en nombre ou délétés chez *M. leprae*. Il semble que les capacités de ce bacille à survivre dans les macrophages et les cellules de Schwann représentent à elles seules des déterminants de pathogénicité majeurs. Pourtant, la survie dans les macrophages semble délicate, car non seulement sa paroi protéomique est moins bien, mais en plus *M. leprae* a perdu un de ses moyens de défense: la catalase-peroxydase KatG (38). Cependant les deux superoxydes dismutases, SodA et C, persistent. En revanche, la survie de *M. leprae* dans les cellules de Schwann doit être beaucoup plus facile. En effet ces cellules, qui ne sont pas des phagocytes professionnels, sont incapables de détruire les pathogènes. De plus, elles protègent *M. leprae* du système immunitaire de l'hôte et des antibiotiques. Récemment un modèle d'interaction entre ces cellules et *M. leprae* a été proposé. Une protéine de la paroi mycobactérienne ML-LBP21 (21 kilo Dalton (kDa)), qualifiée d'adhésine, se fixerait à un récepteur de la membrane basale des cellules, via l'extrémité carboxy-terminale de la laminin-2 (LN[2]G) (39-41). La découverte de ces récepteurs bactériens et cellulaires offre de nouvelles possibilités pour traiter et prévenir la destruction des nerfs de façon précoce et plus ciblée.

## Traitements

Les informations déduites de la séquence sont très importantes pour le traitement de la lèpre. La connaissance de l'ensemble des gènes, enzymes, protéines et voies métaboliques permet de comprendre pourquoi tel antibiotique est actif ou non et d'envisager l'utilisation de nouveaux médicaments.

Le traitement recommandé par l'OMS (PCT) comprend la rifampicine, la dapsonne et la clofazimine. Puisque la rifampicine est l'antibiotique clé de la PCT, le

projet de séquençage a commencé par le cosmide B1790 (14), qui contient la région du chromosome entourant le gène *rhoB*. En effet, les études précédentes, réalisées chez *E. coli* et d'autres micro-organismes, ont montré que la sous-unité  $\beta$  de l'ARN polymérase, codée par *rhoB*, est la cible de la rifampicine. Les bases moléculaires de la résistance à la rifampicine ont été établies à partir des résultats de la séquence et de l'étude d'isolats cliniques résistants à cet antibiotique (42). De plus, ces informations ont été utilisées pour mettre au point des tests diagnostiques moléculaires qui permettent de détecter rapidement les bacilles résistants à la rifampicine à partir de biopsies (43, 44). Il faut préciser que, chez *M. leprae*, la réalisation d'un antibiogramme classique est impossible puisque ce bacille ne se multiplie pas en dehors d'un hôte animal, et qu'il est donc nécessaire, pour évaluer la sensibilité des bacilles, de les injecter dans le coussinet plantaire de la souris ce qui implique une procédure longue, contraignante et coûteuse. Les techniques moléculaires qui sont basées sur une amplification génique et une hybridation présentent en revanche de nombreux avantages. D'abord, elles permettent d'éviter la culture bactérienne, l'amplification ayant lieu directement à partir des biopsies, ensuite, elles sont très rapides dans la mesure où les résultats sont obtenus en 48 heures.

La PCT a été développée en raison d'une large émergence de la résistance à la dapsonne, qui avait été administrée aux malades en monothérapie dans les années 60. De nouveau, la génomique a été d'une grande importance pour comprendre les bases moléculaires de la résistance à la dapsonne. En effet, les auteurs ont montré que c'est le gène *folP1*, qui code pour une folate-synthase, qui est muté chez les bactéries résistantes à la dapsonne (45). De plus, l'étude des mécanismes enzymatiques et du mode d'action de la dapsonne, réalisés grâce à la cristallisation de cette enzyme chez *M. tuberculosis*, devraient permettre de synthétiser de nouvelles molécules mieux adaptées pour lutter contre la lèpre.

Le troisième composé de la PCT, la clofazimine, est l'antibiotique le moins bien connu quant à son mode d'action, et les mécanismes de résistance n'ont pas encore été étudiés.

Le fait que *M. leprae* soit relativement sensible à un plus grand nombre d'antibiotiques que ne l'est le bacille tuberculeux peut s'expliquer par sa composition génique. D'abord, il possède moins de gènes codant pour des protéines capables de

modifier ou d'inhiber les antibiotiques. Par exemple, *M. leprae* est plus sensible aux  $\beta$ -lactames que le bacille tuberculeux car il ne possède que deux gènes codant potentiellement pour des  $\beta$ -lactamases (ML0270 et ML1923), alors que *M. tuberculosis* en comporte sept. Ensuite, la capacité de *M. leprae* à transporter les molécules vers l'extérieur de la cellule est très limitée, en effet les 4 grandes familles de protéines transmembranaires (ABC, MFS, SMR, RND), qui assurent les mécanismes d'efflux et contribuent à la résistance naturelle des bactéries, comportent peu de membres comparé à *M. tuberculosis* (46).

Néanmoins, la perte de certains gènes à cause de l'évolution réductive ne représente pas toujours un handicap pour *M. leprae*, vis-à-vis des antibiotiques. Par exemple, le bacille lépreux est résistant à l'isoniazide car l'enzyme qui active cette molécule, la catalase-peroxydase KatG, est absente.

## Discussion

L'accès au génome de *M. leprae* a été possible grâce à la construction des banques cosmidiqes. Ces banques sont encore aujourd'hui un outil d'une valeur inestimable. Elles fournissent en effet une source inépuisable d'ADN, dans la mesure où l'ADN cloné est amplifié chez *E. coli*, et ce sans avoir à manipuler cette bactérie pathogène impossible à cultiver *in vitro*. De plus, chaque banque est utilisée suivant sa spécificité. D'une part, les clones navettes (pYUB18), grâce à leur origine de répllication mycobactérienne, permettent l'étude de l'expression des gènes de *M. leprae* chez d'autres mycobactéries à croissance rapide et non pathogènes, telles que *Mycobacterium smegmatis* ou *M. bovis* (BCG). D'autre part, les clones issus du «shotgun» sont utilisés pour la réalisation de micro-puces à ADN. Celles-ci sont destinées, dans un premier temps, à la recherche de régions variables dans le génome de *M. leprae*, afin d'entreprendre des études épidémiologiques. En effet, la différenciation des souches du bacille lépreux permettrait d'identifier chaque agent infectieux, afin de suivre sa dissémination, et de faire la distinction entre les cas de rechute ou de réinfection. Jusqu'à maintenant, les techniques disponibles et l'accès limité au génome de ce bacille n'avaient pas permis la détection de différences dans la séquence d'ADN de bacilles d'origines géographiques diverses (47) et le génome de *M. leprae*, qui a énormément changé au cours de l'évolution, semblait très stable.

# Avancée en Avancée

Or, on a récemment pu mettre en évidence un certain polymorphisme, au niveau du gène *rpoT* (48) et au niveau du nombre de répétitions du tri nucléotide TTC (49). Ces résultats sont très prometteurs, et la mise au point de tests discriminants est en cours d'étude. Dans un deuxième temps, les micro-puces seront utilisées pour l'étude du transcriptome de *M. leprae* afin de déterminer si les gènes et pseudogènes, repérés grâce à l'analyse informatique, sont transcrits.

Les banques génomiques ont en premier lieu permis la réalisation de la séquence du génome de *M. leprae*. L'analyse de cette séquence a montré que le génome a subi une évolution réductive très importante, qui n'a laissé à ce bacille qu'un set minimum de gènes fonctionnels pour vivre ou survivre dans des niches très sélectives. Néanmoins, on a pu identifier dans ce set minimum, des gènes potentiellement codants, mais pour des protéines de fonction inconnue, qui n'existent que chez *M. leprae*. Ceux-ci semblent particulièrement intéressants dans la mesure où, parmi eux, pourraient se trouver des facteurs de virulence ou des antigènes spécifiques au bacille lépreux. Quelques protéines produites après le clonage de ces gènes sont d'ores et déjà à l'étude. Si elles s'avèrent être des antigènes reconnus par des sérums de malades, elles pourront être intégrées dans des tests immunodiagnostiques ou être utilisées pour la mise au point d'un nouveau vaccin, ce qui constituerait une avancée déterminante dans la lutte contre la lèpre.

**Remerciements** • Cette étude a été réalisée en collaboration avec le Sanger Centre, et le support financier de l'Institut Pasteur, du New York Community Trust, de l'Association Française Raoul Follereau et du Wellcome Trust. Je remercie particulièrement le Professeur Stewart Cole pour ses conseils avisés.

## Résumé •

L'analyse génomique de *Mycobacterium leprae*, l'agent pathogène de la lèpre, a été réalisée grâce au séquençage à haut débit des cosmides et plasmides préparés à partir de l'ADN d'une souche isolée d'un patient. Elle a montré que ce bacille possède un seul chromosome circulaire, dont la taille (3,2 Mb) et la composition en G+C (57,8 %) sont réduits par rapport aux autres chromosomes mycobactériens connus. L'analyse informatique de la séquence a mis en évidence que seule la moitié de la séquence est codante. L'autre moitié contient des pseudogènes et des séquences non codantes. Le génome de *M. leprae* a donc subi une évolution réductive très importante qui ne lui a laissé qu'un set minimum de gènes fonctionnels pour vivre. Les informations déduites de la partie codante de la séquence permettent d'appréhender les propriétés biologiques particulières de cet agent infectieux à développement intra-cellulaire obligatoire. De plus, la disparition de nombreuses voies enzymatiques par rapport à *M. tuberculosis*, autre pathogène intra-cellulaire en certains points semblable à *M. leprae*, explique les différences observées entre les deux organismes. Enfin, l'analyse génomique du bacille lépreux permet de comprendre les bases moléculaires de sa résistance à différents antibiotiques et d'identifier des cibles potentielles pour la mise au point de nouveaux traitements.

## Mots-clés •

*Mycobacterium leprae* - Banques génomiques - Séquence - Evolution réductive - Antibiotiques.

## Abstract •

### THE MYCOBACTERIUM LEPRAE GENOME : FROM SEQUENCE ANALYSIS TO THERAPEUTIC IMPLICATIONS

The genome of *Mycobacterium leprae*, the causative agent of leprosy, was analyzed by rapid sequencing of cosmids and plasmids prepared from DNA isolated from one patient's strain. Results showed that the bacillus possesses a single circular chromosome that differs from other known mycobacterial chromosomes with regard to size (3.2 Mb) and G+C content (57.8%). Computer analysis demonstrated that only half of the sequence contains protein-coding genes. The other half contains pseudogenes and non-coding sequences. These findings indicate that *M. leprae* has undergone a major reductive evolution leaving a minimal set of functional genes for survival. Study of the coding region of the sequence provides evidence accounting for the particular pathogenic properties of *M. leprae* which is an obligate intracellular parasite. Disappearance of numerous enzymatic pathways in comparison with *M. tuberculosis*, an intracellular pathogen comparable to *M. leprae*, could explain the differences observed between the two organisms. Genomic analysis of the leprosy bacillus also provided insight into the molecular basis for resistance to various antibiotics and allowed identification of several potential targets for new drug treatments.

## Key words •

*Mycobacterium leprae* - Genomic libraries - Sequence - Reductive evolution - Antibiotics.

## REFERENCES

- 1 - WHO - *Weekly Epidemiological Record* 1998 ; **73** : 40.
- 2 - WHO SG - Chemotherapy of leprosy for control programmes. *World Health Organ Tech Report Series* 1982, n° 675.
- 3 - FINE PE - BCG vaccination against tuberculosis and leprosy. *Br Med Bull* 1988 ; **44** : 691-703.
- 4 - FINE PE, SMITH PG - Vaccination against leprosy the view from 1996. *Lepr Rev* 1996 ; **67** : 249-252.
- 5 - SHEPARD CC - The experimental disease that follows the injection of human leprosy bacilli into foot pads of mice. *J Exp Med* 1960 ; **112** : 445-454.
- 6 - SHEPARD CC, CHANG YT - Effect of several anti-leprosy drugs on multiplication of human leprosy bacilli in footpads of mice. *Proc Soc Exp Biol Med* 1962 ; **109** : 636-638.
- 7 - KIRCHHEIMER WF, STORRS EE - Attempts to establish the armadillo as a model for the study of leprosy. *Int J Lep* 1971 ; **39** : 693.
- 8 - CLARK-CURTISS JE, JACOBS WR, DOCHERTY MA *et Coll* - Molecular analysis of DNA and construction of genomic libraries of *Mycobacterium leprae*. *J. Bacteriol.* 1985 ; **161** : 1093-1102.
- 9 - COLE ST, EIGLMEIER K, PARKHILL J *et Coll* - Massive gene decay in the leprosy bacillus. *Nature* 2001 ; **409** : 1007-1011.
- 10 - COULSON A, SULSTON J, BRENNER S, KARN J - Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 1986 ; **83** : 7821-7825.

# Avancée

- 11 - SULSTON J, MALLET F, STADEN R *et Coll* - Software for genome mapping by fingerprinting techniques. *CABIOS* 1988; **4** : 125-132.
- 12 - JACOBS WR, KALPANA GV, CIRILLO JD *et Coll* - Genetic systems for mycobacteria. *Methods in Enzymology* 1991; **204** : 537-555.
- 13 - COLE ST, BROSCHE R, PARKHILL J *et Coll* - Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; **393** : 537-544.
- 14 - HONORE N, CHANTEAU S, DOUCET-POPULAIRE F *et Coll* - Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project : structure and function of the Rif-Str regions. *Mol Microbiol* 1993; **7** : 207-214.
- 15 - TEKAIA F, GORDON S V, GARNIER T *et Coll* - Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis* 1999; **79** : 329-342.
- 16 - MARQUES MA, CHITALE S, BRENNAN PJ, PESSOLANI MC - Mapping and identification of the major cell wall-associated components of *Mycobacterium leprae*. *Infect Immun* 1998; **66** : 2625-2631.
- 17 - JUNGBLUT PR, SCHAIBLE UE, MOLLENKOPF HJ *et Coll* - Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains : towards functional genomics of microbial pathogens. *Mol Microbiol* 1999; **33** : 1103-1117.
- 18 - ANDERSSON JO, ANDERSSON SG - Genomic rearrangements during evolution of the obligate intracellular parasite *Rickettsia prowazekii* as inferred from an analysis of 52015 bp nucleotide sequence. *Microbiology* 1997; **143** : 2783-2795.
- 19 - SHIGENOBUS, WATANABE H, HATTORI M *et Coll* - Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 2000; **407** : 81-86.
- 20 - GORDON SV, HEYM B, PARKHILL J *et Coll* - New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 1999; **145** : 881-892.
- 21 - MIZRAHI V, DAWES S, RUBIN H - DNA replication. - In «HATFULL GF, JACOBS WRJ - Molecular Genetics of Mycobacteria». ASM Press ed, Washington DC, 2000, pp 159-172.
- 22 - COLE ST, SUPPLY P, HONORE N - Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. *Lepr Rev* 2001; **72** : 449-461.
- 23 - Cole ST, Barrell BG - Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Novartis Found Symp* 1998; **217** : 160-172.
- 24 - JONES L, MOSZER I, COLE ST - Leproma: a *Mycobacterium leprae* genome browser. *Lepr Rev* 2001; **72** : 470-477.
- 25 - ALTSCHUL S, GISH W, MILLER W *et Coll* - A basic local alignment search tool. *J Mol Biol* 1990; **215** : 403-410.
- 26 - GISH W, STATES DJ - Identification of protein coding regions by database similarity search. *Nat Genet* 1993; **3** : 266-272.
- 27 - PEARSON W, LIPMAN D - Improved tools for biological sequence comparisons. *Proc Natl Acad USA* 1988; **85** : 2444-2448.
- 28 - DAFRE M, DRAPER P - The envelope layers of mycobacteria with reference to their pathogenicity. *Adv Microb Physiol* 1998; **39** : 131-203.
- 29 - YUAN Y, BARRY CE - A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 1996; **93** : 12828-12833.
- 30 - YUAN Y, ZHU Y, CRANE DD, BARRY CE - The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in *Mycobacterium tuberculosis*. *Mol Microbiol* 1998; **29** : 1449-1458.
- 31 - GLICKMAN MS, COX JS, JACOBS WR - A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol Cell* 2000; **5** : 717-727.
- 32 - MANGANELLI R, DUBNAU E, TYAGI S *et Coll* - Differential expression of 10 sigma factor genes in *Mycobacterium tuberculosis*. *Mol Microbiol* 1999; **31** : 715-724.
- 33 - MANGANELLI R, VOSKUIL M I, SCHOOLNIK G K, SMITH I - The *Mycobacterium tuberculosis* ECF sigma factor SigmaE : role in global gene expression and survival in macrophages. *Mol Microbiol* 2001; **41** : 423-437.
- 34 - PARKHILL J, WREN BW, THOMSON NR *et Coll* - Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 2001; **413** : 523-527.
- 35 - LOBRY JR - Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* 1996; **13** : 660-665.
- 36 - KARLIN S - Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr Opin Microbiol* 1998; **1** : 598-610.
- 37 - ARRUDA S, BOMFIM G, KNIGHTS R *et Coll* - Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* 1993; **261** : 1454-1457.
- 38 - EIGLMEIER K, FSIHI H, HEYM B, COLE ST - On the catalase-peroxidase gene, katG, of *Mycobacterium leprae* and the implications for treatment of leprosy with isoniazid. *FEMS Microbiol Lett* 1997; **149** : 273-278.
- 39 - SHIMOJI Y, NG V, MATSUMURA K, FISCHETTI V A *et Coll* - A 21-kDa surface protein of *Mycobacterium leprae* binds peripheral nerve laminin-2 and mediates Schwann cell invasion. *Proc Natl Acad Sci USA* 1999; **96** : 9857-9862.
- 40 - RAMBUKKANA A - How does *Mycobacterium leprae* target the peripheral nervous system? *Trends Microbiol* 2000; **8** : 23-28.
- 41 - RAMBUKKANA A, ZANAZZI G, TAPINOS N, SALZER J L - Contact-dependent demyelination by *Mycobacterium leprae* in the absence of immune cells. *Science* 2002; **296** : 927-931.
- 42 - HONORE N, COLE ST - The molecular basis of rifampin-resistance in *Mycobacterium leprae*. *Antimicrob Agents Chemother* 1993; **37** : 414-418.
- 43 - HONORE N, PERRANI E, TELENTI A *et Coll* - A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. *Int J Lepr Other Mycobact Dis* 1993; **61** : 600-604.
- 44 - HONORE N, ROCHE P W, GROSSET J H, COLE ST - A method for rapid detection of rifampicin-resistant isolates of *Mycobacterium leprae*. *Lepr Rev* 2001; **72** : 441-448.
- 45 - KAI M, MATSUOKA M, NAKATA N, MAEDA S *et Coll* - Diaminodiphenylsulfone resistance of *Mycobacterium leprae* due to mutations in the dihydropteroate synthase gene. *FEMS Microbiol Lett* 1999; **177** : 231-235.
- 46 - GROSSET JH, COLE ST - Genomics and the chemotherapy of leprosy. *Lepr Rev* 2001; **72** : 429-440.
- 47 - WILLIAMS DL, GILLIS TP, PORTAELS F - Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol* 1990; **4** : 1653-1659.
- 48 - MATSUOKA M, MAEDA S, KAI M *et Coll* - *Mycobacterium leprae* typing by genomic diversity and global distribution of genotypes. *Int J Lepr Other Mycobact Dis* 2000; **68** : 121-128.
- 49 - SHIN YC, LEE H, WALSH GP *et Coll* - Variable numbers of TTC repeats in *Mycobacterium leprae* DNA from leprosy patients and use in strain differentiation. *J Clin Microbiol* 2000; **38** : 4535-4538.