

## Révéler la lèpre pré-contact européen dans les Amériques et sa persistance durable

*Texte traduit en français à partir de l'original en anglais*

### Abstract:

La lèpre, principalement provoquée par *Mycobacterium leprae*, est généralement considérée comme une maladie introduite dans les Amériques durant la colonisation européenne. Toutefois, la découverte récente d'un second agent de la lèpre, *M. lepromatosis*, principalement présent dans les Amériques, remet cette hypothèse en question. Nous montrons ici que *M. lepromatosis* infectait déjà les populations humaines des Amériques avant le contact européen. En analysant 389 échantillons anciens et 408 échantillons contemporains, nous avons presque triplé le volume de données génétiques disponibles pour cette espèce. Les analyses phylogénétiques révèlent des clades distincts infectant l'humain, dont l'un dominant l'Amérique du Nord depuis l'époque coloniale. La présence de souches vieilles de plusieurs millénaires en Amérique du Nord et du Sud indique une distribution étendue de *M. lepromatosis* durant l'Holocène tardif, démontrant ainsi une histoire de la lèpre due à *M. lepromatosis* bien antérieure à l'arrivée des Européens.

**Résumé en une phrase :** La lèpre causée par *M. lepromatosis* a infecté les humains dans les Amériques pendant des millénaires.

### Corps de Texte

Les maladies infectieuses ont profondément influencé et façonné l'histoire des Amériques durant la conquête européenne (1–5). Au cours de ce processus critique, toute une gamme d'agents pathogènes, attestés dans les sources testimoniales ou actuellement prévalents dans la région, a été introduite depuis d'autres parties du monde (5–9). Cependant, les maladies infectieuses qui existaient sur le continent avant l'invasion européenne, ainsi que leur influence sur les paysages épidémiologiques contemporains, sont beaucoup moins bien comprises et ont été déduites à partir de quatre sources principales : i) l'endémisme et la prévalence accrue au sein du continent par rapport à d'autres régions (par ex. la maladie de Chagas (10)), ii) les données paléopathologiques issues des restes squelettiques (11–13), iii) les analyses phylogénomiques de souches pathogènes actuelles indiquant une origine putative dans les Amériques (par ex. les tréponématoses telles que la syphilis endémique) (14), et iv) les approches basées sur l'ADN ancien (aDNA) (par ex. l'hépatite B (15), les tréponématoses (*Treponema pallidum*) (16) et la tuberculose causée par *Mycobacterium pinnipedii* (2, 17)).

La lèpre, l'une des plus anciennes (18) et plus stigmatisantes maladies de l'histoire humaine (19–21), est considérée comme une maladie infectieuse introduite dans les Amériques lors de la colonisation européenne (22, 23). Cette introduction était associée à *Mycobacterium leprae*, la bactérie responsable de la lèpre la plus répandue dans le monde (24–27), qui demeure également la principale cause de lèpre dans les Amériques. Plus récemment, en 2008, une

## Révéler la lèpre pré-contact européen dans les Amériques et sa persistance durable

*Texte simplifié pour un public non-spécialiste*

### Abstract:

La lèpre est une maladie depuis longtemps associée à une bactérie appelée *Mycobacterium leprae*, arrivée dans les Amériques avec les colonisateurs européens. Mais des scientifiques ont récemment découvert une seconde bactérie, *Mycobacterium lepromatosis*, qui provoque également la lèpre et se trouve principalement dans les Amériques. Cela suggère que la lèpre pourrait avoir existé dans les Amériques avant même l'arrivée des Européens.

Dans notre étude, nous avons analysé près de 800 échantillons, anciens (provenant d'Ancêtres Autochtones) et modernes (provenant de personnes atteintes de lèpre aujourd'hui), et avons trouvé des preuves manifestes de la présence de *M. lepromatosis* dans les Amériques il y a plus de mille ans. Notre analyse génétique montre que cette bactérie se propage et évolue dans la région depuis très longtemps.

**Résumé en une phrase :** La lèpre causée par *M. lepromatosis* a infecté les humains dans les Amériques pendant des millénaires.

### Corps de Texte

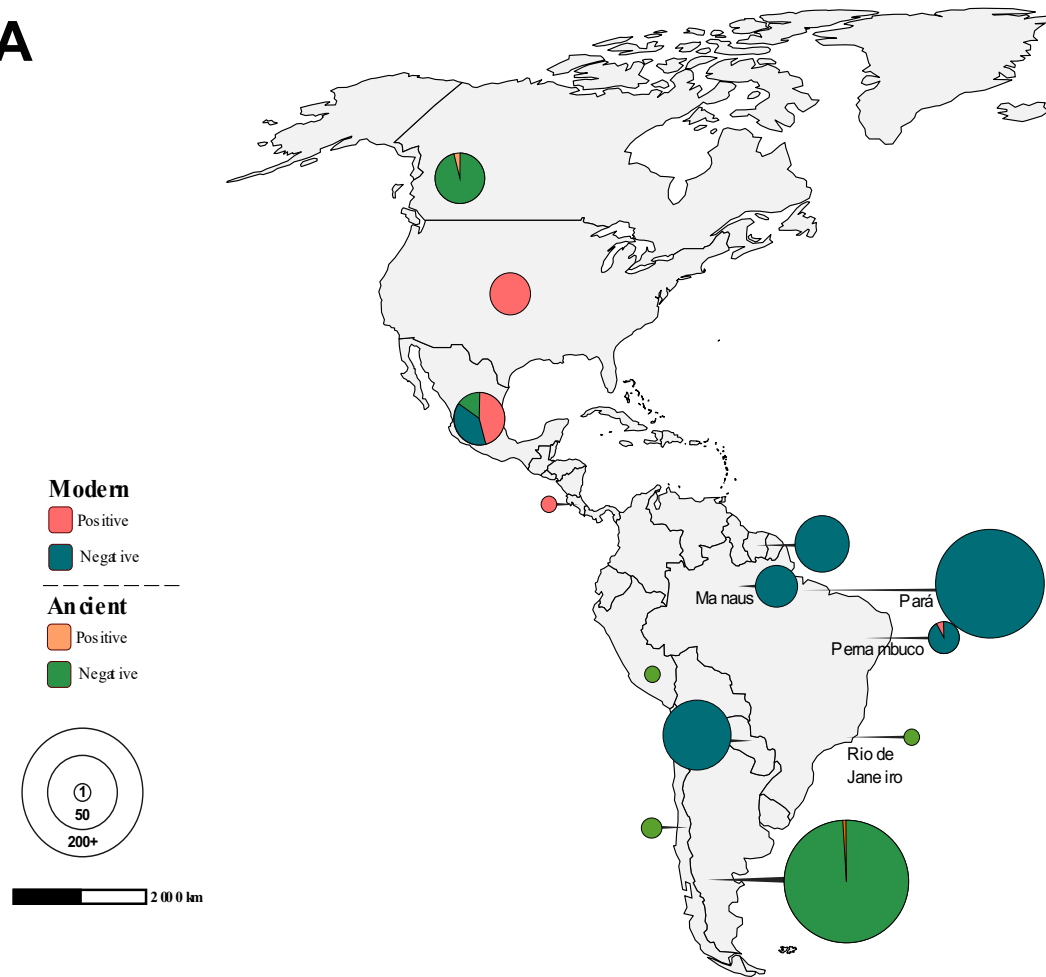
Les maladies apportées par les Européens ont eu un impact majeur sur les communautés Autochtones des Amériques, et nombre d'entre elles sont bien documentées. Mais les maladies qui étaient déjà présentes dans les Amériques avant cette époque restent encore mal comprises. Cette étude contribue à combler cette lacune en recherchant des agents pathogènes anciens - des micro-organismes responsables de maladies - dans des restes humains très anciens ainsi que dans des échantillons provenant de populations contemporaines du continent.

La lèpre est l'une des maladies les plus anciennes et les plus socialement stigmatisées qui soient. Historiquement, les scientifiques pensaient qu'elle était arrivée dans les Amériques avec les colonisateurs européens, lorsqu'une bactérie appelée *Mycobacterium leprae* y avait été introduite. Cette espèce demeure aujourd'hui la principale cause de lèpre dans les Amériques. Cependant, en 2008, des chercheurs ont découvert une deuxième espèce responsable de la lèpre, *Mycobacterium lepromatosis*. Contrairement à *M. leprae*, cette espèce n'a été trouvée quasiment

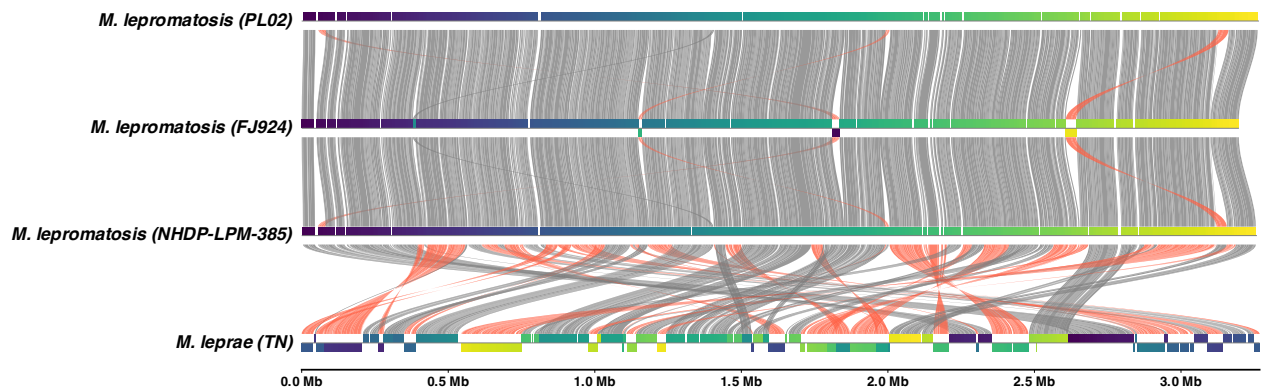
seconde espèce responsable de la lèpre, *Mycobacterium lepromatosis*, a été identifiée (28) ; cette espèce a été signalée presque exclusivement dans les Amériques, principalement au Mexique, avec des cas supplémentaires aux États-Unis, au Canada, au Brésil et dans les Caraïbes (29–32). Très peu de cas ont été signalés sur d'autres continents, avec seulement deux cas à Singapour (33), deux cas au Myanmar (31) et quelques-uns chez des écureuils roux de Grande-Bretagne et d'Irlande (34) (tableau S1). *M. leprae* et *M. lepromatosis* ne peuvent pas être cultivés in vitro, et leur diagnostic différentiel ne peut être effectué que par des analyses moléculaires (35, 36). À ce jour, les seules données génomiques disponibles pour *M. lepromatosis* proviennent de trois échantillons humains de patients d'origine mexicaine (24, 34, 37, 38) et de sept écureuils roux de Grande-Bretagne/Irlande (39), limitant ainsi la compréhension actuelle de la diversité génétique et de l'histoire de cette espèce. Bien que la prévalence plus élevée de *M. lepromatosis* dans les Amériques suggère une émergence plausible de l'espèce sur ce continent (24), les données phylogénomiques actuelles n'excluent pas une introduction possible depuis d'autres continents (31, 33).

qu'exclusivement dans les Amériques - surtout au Mexique - avec quelques cas signalés aux États-Unis, au Canada, au Brésil et dans les Caraïbes. Seule une poignée de cas a été rapportée ailleurs dans le monde, notamment chez des écureuils roux au Royaume-Uni et en Irlande. Les scientifiques ne peuvent cultiver aucune de ces deux bactéries en laboratoire ; ils doivent donc recourir à des techniques avancées basées sur l'ADN pour les détecter et les étudier. Jusqu'à présent, les chercheurs n'avaient séquencé l'ADN de *M. lepromatosis* qu'à partir de trois échantillons humains (tous provenant de personnes d'origine mexicaine) et de sept écureuils roux. Ces informations très limitées rendent difficile la compréhension complète de l'histoire et de la variabilité de cette espèce. Bien qu'il semble probable que *M. lepromatosis* soit apparue pour la première fois dans les Amériques, les données disponibles ne permettent pas d'exclure la possibilité qu'elle provienne d'ailleurs.

**A**



**B**



**Fig. 1. Sites d'origine des échantillons contemporains et anciens dans les Amériques et carte de synténie entre les génomes de bactéries responsables de la lèpre (A)** Les diagrammes circulaires indiquent la localisation des sites ou les pays d'où proviennent les échantillons analysés. Les graphiques sont codés par couleur selon la présence ou l'absence de *M. lepromatosis* et selon qu'ils proviennent de biopsies actuelles ou de contextes archéologiques anciens.

**(B)** Illustration d'alignements progressifs réalisés avec Mauve montrant les réarrangements possibles et les blocs de synténie entre quatre génomes bactériens responsables de la lèpre, visualisés avec pyGenomeViz (48). Les espèces (et les souches entre parenthèses) sont indiquées en gras à gauche. La taille des génomes (en mégabases, Mb) est indiquée sous chaque entrée. La palette Viridis (Matplotlib) représente la position le long

**Figure 1. Lieux d'origine des échantillons anciens et modernes et comparaison des génomes bactériens**

**(A)** Cette carte montre où ont été sourcés les échantillons anciens et modernes analysés dans cette étude. Les diagrammes circulaires précisent si le site contient des échantillons positifs pour *M. lepromatosis* et si ceux-ci proviennent de patients actuels ou de restes archéologiques.

**(B)** Cette image compare l'ADN de quatre bactéries causant la lèpre. Les bandes colorées montrent à quel point leurs structures génétiques sont similaires. Les souches d'une même espèce (*M. lepromatosis*) présentent des motifs de couleur très proches, indiquant que leur ADN est organisé de la même façon. Cependant, une souche de *M. leprae* affiche un motif très différent, indiquant un génome organisé de manière distincte. Les marques orange soulignent les segments d'ADN orientés à l'inverse. Ces

du génome de référence *M. lepromatosis* (NHDP-LPM-385), avec un gradient continu du début à la fin. Les génomes entièrement synténiques avec la référence - comme les autres souches de *M. lepromatosis* - présentent un ordre de couleurs presque identique, reflet d’une structure génomique conservée. À l’inverse, le génome de *M. leprae* TN montre une séquence de couleurs perturbée en raison de différences d’organisation par rapport aux coordonnées de la référence *M. lepromatosis*. Des lignes grises relient les régions colinéaires, tandis que les surlignages orange signalent les blocs inversés (sur le brin « antisens ») par rapport à la référence NHDP-LPM-385.

Informations aident les chercheurs à comprendre comment ces bactéries ont évolué et se sont modifiées au fil du temps.

Tree branch / souche	Source	Pays / Région	Datation	Profondeur Moyenne de couverture (X)	Étendue de couverture (%)	SNPs (nombre)
NHDP-LPM-9(†) NHDP-LPM-6	Cette étude	USA	Contemporain	59	97	4111
Ecureuils roux (7 souches)	Avanzi et al. 2016	British Isles	Contemporain	m: 16 – Av: 59 – M: 126	m: 97 – M: 99	m: 423 – Av: 471 – M: 514
XVII-B-357	This study	Canada	1310 Cal AP	32	98	422
AR0353	This study	Argentina	942 Cal AP	0.67	27	635
AR0018	This study	Argentina	861 Cal AP	92	99	388
PDDC (24 souches) (†)	Mostly this study	North/Central America	Contemporain	m: 0.53 – Av: 44 – M: 171	m: 15 – Av: 85 – M: 99	m: 57 – Av: 80 – M: 171

**Tableau 1.** Statistiques récapitulatives des souches infectant l’humain et les animaux sauvages incluses dans cette étude. Les branches de l’arbre comportant plusieurs souches (deuxième et dernière ligne) ainsi que les valeurs minimale (m), moyenne (Av) et maximale (M) sont indiquées. (†) Seuls les génomes de haute qualité ont été pris en compte pour les trois statistiques présentées des clades NHDP-LPM-9/6 et PDDC (p). Voir le tableau S2 pour des statistiques récapitulatives étendues.

Ce tableau résume les informations génétiques clés des souches de *M. lepromatosis* trouvées à la fois chez l’homme et chez les animaux sauvages analysées dans cette étude. Pour les groupes de souches formant des branches spécifiques dans l’arbre phylogénétique, il indique les valeurs minimale, moyenne et maximale de certaines mesures génétiques. Ces valeurs permettent de comparer la similarité ou la différence des souches au sein de chaque groupe. Seules les données génomiques de haute qualité ont été utilisées pour calculer ces chiffres pour deux groupes importants : le clade NHDP-LPM-9/6 et le clade PDDC. Une version plus détaillée de ce tableau est disponible dans les matériaux complémentaires.

Pour étudier la distribution et la diversité actuelle de *M. lepromatosis*, nous avons recherché l’ADN de cette bactérie en criblant une vaste collection de 408 échantillons - prélevés pour l’essentiel chez des patients atteints de la lèpre récemment - provenant de cinq pays des Amériques (États-Unis, Mexique, Guyane française, Brésil et Paraguay) choisis pour leur pertinence géographique et l’accessibilité des échantillons. Notre test de PCR quantitative (qPCR) (voir *Matériel et méthodes*, section 2.5 ; texte supplémentaire, note 3 ; tableau de données S1) a révélé un total de 34 cas positifs (fig. 1A, tableau de données S1). Fait notable, un seul des 360 échantillons testés en Amérique du Sud s’est avéré positif (tableau de données S1), malgré des signalements antérieurs de *M. lepromatosis* en Amérique du Sud (29, 31, 40). À partir de ces échantillons positifs, nous avons obtenu des données génomiques pour 23 spécimens en combinant séquençage haut débit à courtes lectures et stratégies d’enrichissement ciblé (tableau 1, tableau S2), augmentant ainsi de manière significative le jeu de données disponible sur les souches infectant l’humain.

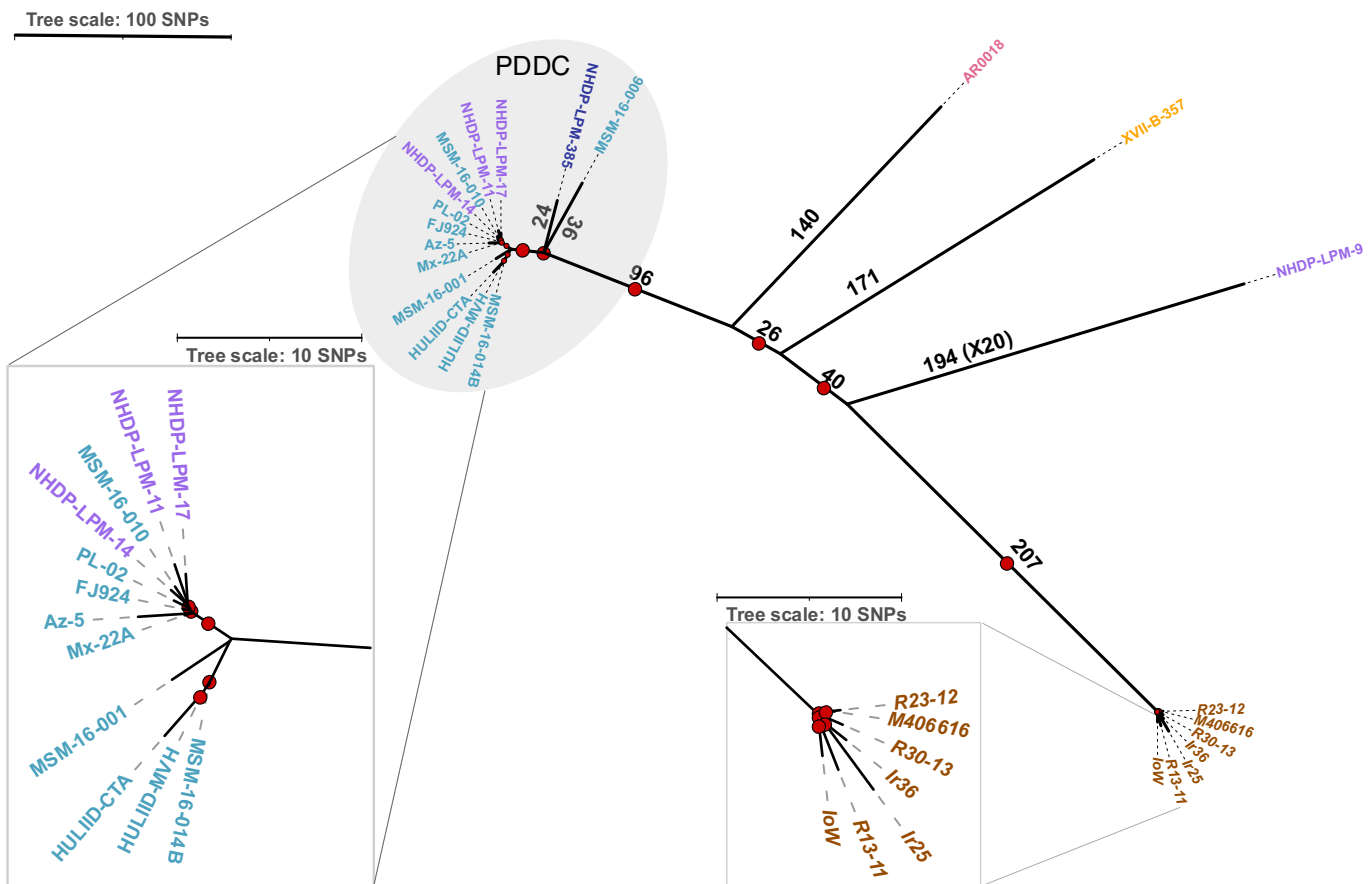
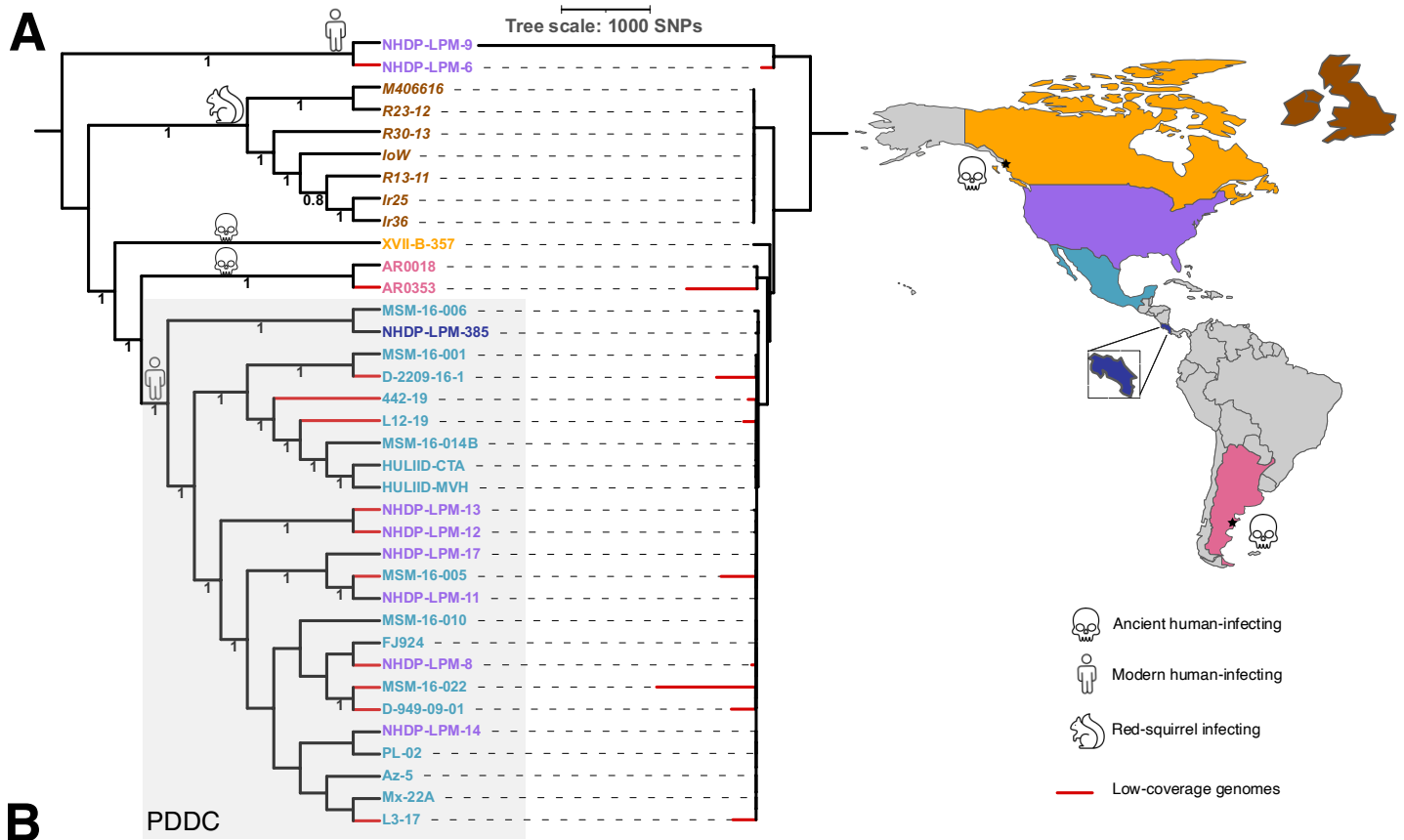
Pour comprendre où est trouvée *M. lepromatosis* aujourd’hui et sa diversité génétique, les chercheurs ont testé 408 échantillons médicaux provenant de personnes de cinq pays des Amériques : les États-Unis, le Mexique, la Guyane française, le Brésil et le Paraguay. Ces échantillons provenaient surtout de personnes atteintes de lèpre ou présentant des symptômes suggérant la lèpre. L’équipe a utilisé une technique de laboratoire appelée PCR quantitative (ou qPCR), qui peut détecter de très petites quantités d’ADN d’un organisme spécifique. Elle a trouvé 34 échantillons positifs à *M. lepromatosis*. Fait surprenant, un seul de ces échantillons positifs provenait d’Amérique du Sud, alors que des études antérieures avaient signalé la présence de la bactérie dans cette région. Sur les 34 cas positifs, les chercheurs ont pu récupérer des données ADN utilisables pour 23 d’entre eux. Ils ont utilisé des techniques avancées combinant un séquençage d’ADN rapide et une « pêche » ciblée du matériel génétique de la bactérie, ce qui les a aidés à augmenter considérablement le nombre de génomes connus de souches de *M. lepromatosis* infectant l’humain.

En raison du caractère incultivable de l'espèce, il est difficile d'obtenir un ADN pur et de haute qualité directement à partir de biopsies. Pour cette raison, le génome de référence actuel de *M. lepromatosis* (FJ924) a été obtenu grâce à plusieurs cycles de ré-assemblage de courtes lectures initialement appariées au génome de *M. leprae*, puis comblées par PCR et séquençage Sanger (37, 38). Nous avons surmonté ces limitations en générant un ADN de grande taille et de haute qualité de *M. lepromatosis* à partir d'échantillons de coussinets plantaires de souris inoculées avec la souche NHDP-LPM-385, isolée d'un échantillon humain du Costa Rica (29). En combinant des lectures longues Nanopore et des lectures courtes Illumina, le génome NHDP-LPM-385 a été assemblé de novo en un seul contig de 3 259 657 paires de bases (bp) avec une profondeur médiane de couverture de 133× (table S3) et une identité moyenne nucléotidique (ANI) de 99,80 % par rapport au génome FJ924 (38) (fig. S3). Fait intéressant, la **carte de synténie** des génomes *M. lepromatosis* NHDP-LPM-385 et *M. leprae* TN (41) a révélé un grand nombre de réarrangements génomiques, incluant 19 inversions (Fig. 1B, texte supplémentaire note 4). Cette observation concorde globalement avec l'analyse récente du génome FJ924 (38), qui a identifié 24 cassures de synténie et 18 événements d'inversion par rapport au génome TN. Toutefois, nos analyses ont également mis en évidence des modifications de synténie précédemment ignorées (voir figs. S1 et S2, tables S3, S4, S5 et S6, et texte supplémentaire note 4).

Comme le nouveau génome NHDP-LPM-385 a été assemblé en un contig unique grâce aux lectures longues, puis repolié par cartographie de courtes lectures à haute couverture, nous jugeons sa résolution plus précise que celle du génome de référence actuel FJ924, nous l'avons donc utilisé comme référence pour le génotypage et la reconstruction phylogénétique (Fig. 2). La reconstruction phylogénétique par maximum de vraisemblance montre que, à l'exception de deux, toutes les souches infectant l'humain aujourd'hui (24 sur 26) appartiennent à un clade phylogénétique dominant actuel quasi clonale (Present-Day Dominant phylogenetic Clade, PDDC) (Fig. 2A, fig. S4, table S2). La topologie de l'arbre demeure robuste, quel que soit le génome de référence ou le sous-ensemble de gènes utilisé (figs. S5 et S6), avec seulement de légères incohérences au sein du PDDC en raison du très faible nombre de variants nucléotidiques simples (SNV) différenciateurs. Fait remarquable, une souche humaine (NHDP-LPM-9) se révèle basale par rapport à l'ensemble des autres souches animales et humaines (Fig. 2, fig. S4), présentant 4 111 SNV par rapport à NHDP-LPM-385 (Data table S2), ce qui suggère l'existence d'un clade *M. lepromatosis* jusqu'ici non reconnu infectant l'humain à l'heure actuelle.

Parce que *M. lepromatosis* ne peut pas être cultivé en laboratoire, il est difficile d'obtenir un ADN propre et de haute qualité directement à partir de tissus humains infectés. Le premier génome de référence (nommé FJ924) a été assemblé à partir de fragments d'ADN initialement appariés à *M. leprae*, l'autre bactérie responsable de la lèpre. Pour améliorer ce résultat, les chercheurs ont fait proliférer la bactérie dans les coussinets plantaires de souris (une technique particulièrement délicate), en utilisant une souche appelée NHDP-LPM-385 prélevée à l'origine chez un patient costaricien. Ils ont ensuite combiné deux technologies de séquençage de l'ADN - Nanopore pour les longues lectures et Illumina pour des lectures courtes et précises - afin de construire un génome complet en un seul fragment (un « contig ») avec une couverture très élevée, c'est-à-dire de nombreuses lectures qui se chevauchent pour chaque région du génome, assurant ainsi l'exactitude. Le nouveau génome compte 3,26 millions de lettres d'ADN et est identique à 99,8 % à l'ancienne référence. En comparant sa structure à celle de *M. leprae*, les chercheurs ont découvert de nombreux réarrangements - segments d'ADN inversés ou déplacés -, appelés « modifications de synténie ». Certaines étaient déjà connues, mais ce nouvel assemblage de haute qualité a permis d'en détecter plusieurs qui n'avaient jamais été observées auparavant.

Le nouveau génome de *M. lepromatosis* (NHDP-LPM-385) a été assemblé en un fragment continu unique grâce à des lectures longues, puis affiné au moyen de lectures courtes de haute qualité. Il est donc plus précis que l'ancien génome de référence (FJ924) et a servi de base pour la comparaison de toutes les autres souches. Lorsque les chercheurs ont construit un arbre phylogénétique pour comprendre les relations entre ces souches, ils ont découvert que 24 des 26 souches modernes isolées chez des patients appartenaient à un même groupe très apparenté, baptisé Present-Day Dominant Clade (PDDC). L'arbre est resté très stable, quel que soit le génome de référence ou le sous-ensemble de gènes utilisé, ce qui confirme la fiabilité de ce résultat. Cependant, une souche particulière, NHDP-LPM-9, se distingue : elle n'appartient pas à ce clade dominant et présente plus de 4 000 variants nucléotidiques simples (SNV), ce qui suggère qu'elle représente une lignée totalement nouvelle et jusqu'alors inconnue de *M. lepromatosis* infectant l'humain encore aujourd'hui.



**Fig. 2. Phylogénie par maximum de vraisemblance (ML) des souches de *M. lepromatosis*.** (A) Arbre par maximum de vraisemblance comprenant les 36 souches de *M. lepromatosis* disposant de données

**Figure 2. Arbre de parenté des souches de *M. lepromatosis* à partir de leur ADN génomique.** (A) Cette image présente les relations évolutives entre tous les génomes connus de *M. lepromatosis* (nouveaux et déjà

génomiques (26 nouvellement séquencées, 10 déjà publiées), présenté à gauche en cladogramme (longueurs de branche ignorées) et à droite en arbre échelonné selon les longueurs de branche. Les étiquettes des extrémités (tip labels) sont codées par couleur selon le pays d'origine (carte insérée) et le type d'hôte est indiqué par des icônes (humains anciens, humains actuels, écureuils roux). Les branches en rouge correspondent à des génomes à faible couverture et peuvent afficher des branches artificiellement longues en raison d'appels de SNV erronés dus à un soutien de lecture insuffisant. Le cladogramme facilite la visualisation des relations phylogénétiques et du support nodal (proportion de répliquats bootstrap indiquée lorsque  $> 0,7$ ).

(B) Arbre par maximum de vraisemblance ne comprenant que les souches à haute couverture ( $n = 24$ ), présenté non enraciné afin d'améliorer la résolution de la structure des clades. Cinq clades bien soutenus sont visibles (support bootstrap  $> 0,7$ , indiqué par des points rouges), chacun séparé de son nœud ancestral par plus de 100 SNP. Au sein des clades PDDC et « écureuil roux », les courtes branches terminales reflètent une diversification récente (les encadrés agrandissent leurs topologies internes). La longue branche menant à NHDP-LPM-9 a été réduite d'un facteur 20 pour la lisibilité. Les longueurs de branche représentent des comptes de SNP, obtenus en multipliant les longueurs de branche RAXML par la longueur de l'alignement, les échelles sont indiquées au-dessus de chaque panneau.

Pour confirmer que la souche était bien *M. lepromatosis*, nous avons réalisé une analyse phylogénétique incluant les souches anciennes et contemporaines de *M. leprae* disponibles publiquement, reposant sur 1 311 gènes orthologues partagés par les deux espèces (fig. S7 ; voir Matériels et méthodes, section 3.19 ; note 6 du texte supplémentaire ; Data table S3). Les résultats positionnent NHDP-LPM-9 comme une souche basale de *M. lepromatosis*, très éloignée de *M. leprae* (fig. S7). Le positionnement basal de NHDP-LPM-9 est également étayé par les analyses de vraisemblance marginale et de facteur de Bayes logarithmique, qui identifient cette branche comme le groupe externe le plus basal par rapport aux autres lignées de l'arbre (table S7, Matériel supplémentaire, section 3.18). La branche basale est aussi confortée par une seconde souche, NHDP-LPM-6 : malgré sa couverture limitée (profondeur de couverture =  $0,2\times$  ; étendue de couverture = 11 %), elle se place de façon fiable dans ce clade (Fig. 2A, Table 1, table S2, fig. S4), avec de nombreuses substitutions partagées exclusivement par les deux souches (figs. S8 et S9). Le génome NHDP-LPM-9 a toutefois accumulé un grand nombre de SNV singletons, générant une branche très longue dans l'arbre (Fig. 2B). Des tests statistiques et une inférence bayésienne indiquent un taux évolutif particulièrement élevé pour NHDP-LPM-9, suggérant qu'il pourrait s'agir d'une souche **hypermutatrice** (texte supplémentaire, figs. S10 et S11). Ces résultats demeurent inchangés après l'application de filtres stricts visant à éliminer les variants artefactuels dus à des séquences contaminantes ou au traitement des échantillons (Matériel supplémentaire, section 3.9 ; table S2 ; fig. S12). Des souches hypermutatrices de *M. leprae* ont déjà été décrites chez des patients lépreux et attribuées à des mutations délétères dans des gènes de réparation de l'ADN (*nth*, endonucléase III), favorisant l'émergence de résistances aux médicaments (22). De même, plusieurs mutations faux-sens ont été détectées dans NHDP-LPM-9 au niveau de gènes codant des protéines de réparation de l'ADN, notamment le facteur de couplage transcription-réparation Mfd, ce qui pourrait expliquer le génotype hypermutateur (table S8, texte supplémentaire). Toutefois, aucun des quatre gènes cibles associés à la résistance médicamenteuse connus chez *M. leprae* (22) ne présente de mutation dans NHDP-LPM-9, ce qui exclut que ce génotype hypermutateur potentiel ait favorisé l'apparition de

publiés). À gauche, la version simplifiée de l'arbre montre toutes les branches de même longueur pour mettre en évidence la structure. À droite, les branches sont mises à l'échelle de la différence génétique entre les souches (leur longueur est proportionnelle au nombre de mutations accumulées). Les couleurs et les icônes placées à côté des noms indiquent le pays d'origine de chaque souche et l'hôte infecté (humain ancien ou actuel, ou écureuil roux). Certaines branches sont en rouge parce qu'elles reposent sur des données de faible qualité et peuvent ainsi paraître plus longues qu'elles ne le sont réellement.

(B) Il s'agit d'un arbre plus détaillé qui ne comporte que les 24 génomes de meilleure qualité. Il met en évidence cinq groupes principaux (appelés clades), très distincts les uns des autres. Certains groupes, comme ceux provenant des écureuils roux et de nombreux cas humains actuels (PDDC), sont étroitement apparentés et n'ont divergé que récemment, ce que révèlent leurs courtes branches terminales. L'arbre montre également une souche atypique (NHDP-LPM-9) qui a accumulé un nombre de mutations nettement plus élevé que les autres, elle est si différente que sa branche a dû être raccourcie pour tenir dans l'image.

Pour s'assurer que la souche inhabituelle NHDP-LPM-9 était bien *M. lepromatosis*, les chercheurs l'ont comparée à de nombreux génomes anciens et modernes de *M. leprae* et de *M. lepromatosis*, en se concentrant sur plus de 1 300 gènes partagés par les deux espèces. Cela a confirmé que NHDP-LPM-9 appartient clairement à *M. lepromatosis*, mais qu'elle a accumulé davantage de différences par rapport aux autres souches connues. Dans l'arbre phylogénétique, elle apparaît à la base de la branche *M. lepromatosis* — ce qui signifie qu'elle s'en est détachée très tôt. Une autre souche, NHDP-LPM-6, s'est également groupée avec elle, soutenant l'idée qu'il s'agit d'une lignée distincte et ancienne. Fait intéressant, NHDP-LPM-9 présentait un nombre inhabituellement élevé de changements génétiques uniques (appelés SNV singleton), ce qui a allongé sa branche dans l'arbre bien plus que les autres. Cela suggère qu'il pourrait s'agir d'un « hypermutateur », c'est-à-dire qu'il accumule les mutations plus rapidement que d'habitude, possiblement à cause de défauts dans son système de réparation de l'ADN. Les chercheurs ont effectivement trouvé des mutations dans certains de ses gènes de réparation de l'ADN, mais pas dans les gènes liés à la résistance aux antibiotiques ; cette hypermutation ne semble donc pas rendre la souche résistante aux médicaments. Dans l'ensemble, ces résultats montrent qu'au moins deux lignées très différentes de *M. lepromatosis* infectent encore des personnes en Amérique du Nord aujourd'hui.



résistances (table S9, note 7 du texte supplémentaire). Il est important de noter que la découverte des souches basales NHDP-LPM-9 et NHDP-LPM-6 montre qu'il existe au moins deux lignées indépendantes de *M. lepromatosis* infectant actuellement l'humain en Amérique du Nord.

Pour clarifier l'histoire et l'ancienne prévalence de *M. lepromatosis* dans les Amériques, nous avons **criblé** une collection de 389 jeux de données d'ADN ancien (aDNA) provenant d'Ancêtres pré-contact européen des Amériques, nouvellement générés ou accessibles publiquement (voir Matériels et méthodes). Ce travail sur l'aDNA a compris diverses actions d'engagement communautaire, décrites en détail dans la section 1.1 du Matériel et méthodes. Fait intéressant, un Ancêtre du territoire actuel du Canada en Amérique du Nord (XVII-B-357) (42) et deux Ancêtres (AR0018 et AR0353) du Cône Sud du continent se sont révélés positifs pour *M. lepromatosis* et négatifs pour *M. leprae* (Fig. 1A, Fig. 2A, tableau S10, voir Matériels et méthodes section 1.3). Bien que la maxillaire XVII-B-357 présente de légères altérations pathologiques compatibles, mais non spécifiques, avec les premiers stades de la lèpre (texte supplémentaire, note 1), aucun signe pathologique macroscopique n'était apparent sur les squelettes des deux autres individus. Ce résultat est attendu, la pathologie osseuse ne se manifestant que chez une faible fraction de la population (43). Les restes des trois Ancêtres ont été datés au radiocarbone de périodes antérieures au contact européen, dans un intervalle relativement court de l'Holocène tardif, il y a environ mille ans (XVII-B-357 : 1 310 calBP ; AR0018 : 860 calBP ; AR0353 : 942 calBP, fig. S13). Conformément aux datations, les génomes des trois Ancêtres montrent uniquement une ascendance génétique Autochtone des Amériques (texte supplémentaire, note 11 ; fig. S14 ; tableau S11).

En utilisant une combinaison de séquençage à haute profondeur et de stratégies d'enrichissement ciblé (ces dernières n'ont pas été nécessaires pour XVII-B-357), nous avons obtenu des génomes de *M. lepromatosis* à très forte couverture pour XVII-B-357 (31×) et AR0018 (92×). Les profils de dommages de l'ADN étaient conformes aux rapports précédents sur les génomes mycobactériens anciens (2, 23) (figs. S15, S16 et S17, table S14, note 8 du texte supplémentaire). Fait remarquable, malgré leurs localisations extrêmement éloignées, ces deux souches anciennes se regroupent l'une à côté de l'autre et à la base du PDDC d'Amérique du Nord et centrale, XVII-B-357 divergeant en premier (Fig. 2B). Cependant, les deux branches sont bien séparées et relativement longues, formant apparemment des clades distincts, ce qui suggère que chacune a suivi une longue histoire indépendante. Bien que le génome d'AR0353 présente une faible couverture (profondeur = 0,67× ; étendue = 27 %), il se regroupe systématiquement avec AR0018 dans l'ensemble de nos tests phylogénétiques (notamment ceux de la Fig. 2A et de la fig. S4), ce qui renforce l'idée d'un clade distinct en Amérique du Sud.

Our results combining ancient and present-day DNA data revealed five distinct clades of *M. lepromatosis*, identified by the following strains: (i) the most basal, represented by the NHDP-LPM-9 and NHDP-LPM-6 strains, (ii) the red squirrel-infecting

Pour déterminer à quelle point *M. lepromatosis* était commune, les chercheurs ont analysé 389 échantillons d'ADN ancien (aDNA) provenant de restes humains antérieurs à l'arrivée des Européens dans les Amériques. Ces échantillons provenaient d'études précédentes ou étaient nouvellement collectés, et la recherche a été menée avec le consentement des communautés Autochtones. Ils ont identifié trois individus clairement infectés par *M. lepromatosis* : un provenant de l'actuel Canada (XVII-B-357) et deux d'Argentine (AR0018 et AR0353). Aucun d'entre eux ne présentait d'ADN de *M. leprae*, l'autre bactérie de la lèpre. Un des squelettes (XVII-B-357) montrait quelques signes légers dans la mâchoire pouvant être liés aux premiers stades de la lèpre, mais ces indices n'étaient pas concluants. Les deux autres squelettes ne présentaient pas de modifications osseuses visibles, ce qui n'est pas surprenant car la plupart des personnes atteintes de lèpre ne développent pas de lésions touchant le squelette. La datation au radiocarbone a montré que les trois individus ont vécu il y a environ 1 000 ans, bien avant la colonisation européenne. L'analyse génétique a également confirmé que ces individus avaient une ascendance propre aux peuples Autochtones des Amériques. Cela signifie que *M. lepromatosis* circulait déjà parmi les populations Natives d'Amérique du Nord et du Sud bien avant l'arrivée des Européens.

Pour mieux comprendre la génétique des souches anciennes, les chercheurs ont utilisé un séquençage profond de l'ADN, c'est-à-dire la lecture d'un très grand nombre de molécules d'ADN (plus de 100 millions de fragments), afin de garantir la précision et une couverture génomique élevée. Pour deux des individus (XVII-B-357, Canada, et AR0018, Argentine), ils ont réussi à récupérer des génomes de *M. lepromatosis* de haute qualité. Les profils de dommages de l'ADN dans ces échantillons correspondaient à ceux observés habituellement dans d'autres ADN bactériens anciens, confirmant leur authenticité (c'est-à-dire qu'ils sont anciens). Lorsque les chercheurs ont comparé ces deux génomes anciens, ils ont été surpris de constater que — malgré la grande distance entre le Canada et l'Argentine — les souches étaient étroitement apparentées et se détachaient près de la base de l'arbre phylogénétique, avant le Present-Day Dominant Clade (PDDC), qui regroupe la plupart des souches modernes du Mexique et des États-Unis. La souche canadienne (XVII-B-357) diverge en premier, suivie de la souche argentine (AR0018). Cela suggère que ces deux souches anciennes provenaient de lignées différentes mais apparentées, qui évoluaient déjà indépendamment depuis longtemps. Bien que le troisième génome ancien (individu AR0353, Argentine) soit de qualité et de couverture inférieures, il se regroupe néanmoins avec AR0018 dans tous les tests génétiques. Cela renforce l'idée qu'il existait, en Amérique du Sud à l'époque pré-coloniale, une lignée distincte désormais attestée de *M. lepromatosis*.



strains, (iii) the ancient XVII-B-357 strain, (iv) the AR0018/AR0353 strains, and (v) the 24 PDDC strains (Fig. 2B). The phylogenetic tree shows that each clade is clearly differentiated by multiple SNVs, with the nodes defining them being well-supported. Additionally, the branches extending from these nodes to the sampled strains are relatively long, indicating a substantial evolutionary history of unsampled diversity since the initial divergence of each clade. However, within the red squirrel and PDDC clades, the sampled strains exhibit very short terminal branches from their most recent common node, suggesting that while these clades experienced prolonged periods of divergence from their respective ancestors, the extant strains within them have only recently diverged from each other. To assess the timing of divergence between strains and clades, we conducted molecular clock analyses, including the contemporary and ancient strains of highest quality (breadth-of-coverage >95% and average depth-of-coverage >10X) (fig. S18), using a relaxed clock model best-fitting our data after thorough testing (fig. S19, see materials and methods, fig. S21, supplementary text note 9). The estimated most recent common ancestor (MRCA) of all clades, including NHDP-LPM-9, dates to a median of ~9400 BP (Fig. 3A, Internal Node 1 in fig. S20A), followed by an MRCA of median ~3200 BP for the node of the red squirrels branch, ~2200 BP for the XVII-B-357 and ~1500 BP to the node separating the AR0018/AR0353 and PDDC. In our view, the most parsimonious interpretation of these results is that the divergence of these clades occurred within the Americas, as the estimated dates fall within a period of limited contact between the Americas and other continents. They also reveal a very recent expansion of the red squirrels and PDDC clades, with strains diverging from a common node dated at a median of ~105 BP and ~280 BP, respectively, both falling within a post-European contact period.

En exploitant nos nouvelles données, nous avons recalculé le temps de divergence entre *M. leprae* et *M. lepromatosis* en analysant des souches anciennes et contemporaines des deux espèces. Le dernier ancêtre commun a été estimé à 720 ka à l'aide d'un modèle d'horloge détendue avec une distribution coalescente exponentielle (valeur médiane ; intervalle crédible à 95 % : 660–792 ka, Fig. 3B). Nous avons également testé un second modèle, utilisant comparativement une distribution plus naïve de la taille de population (1/X), qui donne des dates plus anciennes : médiane 2,015 Ma (intervalle crédible à 95 % : 1,507–2,583 Ma, Fig. 3B). Ces estimations sont nettement plus récentes que les 14 Ma rapportés auparavant (24, 38), ce qui souligne l'importance d'une taille d'échantillon accrue - obtenue grâce aux nouveaux génomes modernes - et d'une plus grande profondeur temporelle -grâce à l'inclusion de génomes anciens - pour affiner les horloges moléculaires chez les pathogènes à évolution lente.

En effet, les taux évolutifs calculés pour ces bactéries (et d'autres du même genre) suggèrent environ une substitution tous les cinq ans à l'échelle du génome (0,19 subst./an pour un génome de 3,2 Mbp) ; un temps de divergence de plusieurs millions d'années saturerait donc la plupart des sites du génome. Par conséquent,

En combinant des données d'ADN ancien et moderne, les chercheurs ont identifié cinq groupes génétiques clairement distincts — ou clades — de *Mycobacterium lepromatosis* :

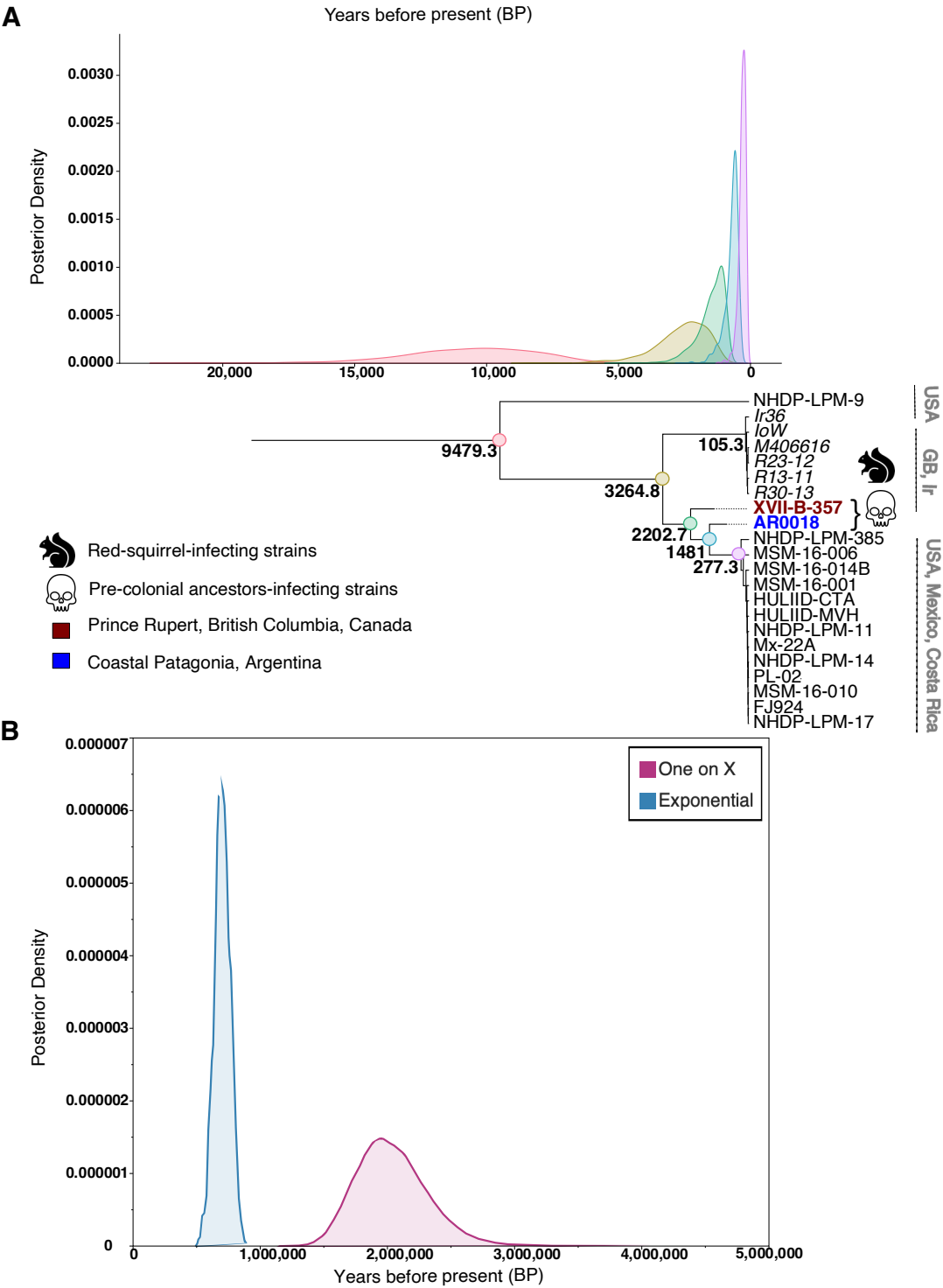
1. Le clade le plus ancien (ou « basal »), constitué de deux souches modernes des États-Unis (NHDP-LPM-9 et NHDP-LPM-6).
2. Un clade formé par sept souches isolées chez des écureuils roux des îles Britanniques.
3. La souche ancienne du Canada (XVII-B-357).
4. Les deux souches anciennes d'Argentine (AR0018 et AR0353).
5. Un grand groupe de souches modernes très proches, principalement trouvées au Mexique et aux États-Unis, appelé Present-Day Dominant Clade (PDDC).

Ces groupes ont été définis à l'aide d'un arbre phylogénétique - comparable à un arbre généalogique pour les bactéries. Chaque clade est caractérisé par de nombreuses différences génétiques uniques (SNV), et plus la « branche » menant à une souche est longue, plus elle évolue de façon indépendante depuis longtemps. Certains clades, comme celui des écureuils roux et le PDDC moderne, présentent des branches terminales courtes, signe d'une diversification relativement récente. En revanche, les longues branches qui relient ces clades au reste de l'arbre indiquent que de nombreuses souches intermédiaires manquent ou n'ont pas été échantillonnées, révélant des lacunes dans nos connaissances sur l'histoire et la diversification du pathogène. Pour estimer le moment où ces groupes se sont séparés, les chercheurs ont appliqué une analyse d'horloge moléculaire. Cette méthode calcule les dates de divergence entre souches et clades en comptant le nombre de mutations accumulées au fil du temps et en supposant qu'elles s'accumulent à un rythme mesurable, comme le tic-tac d'une horloge. Les résultats suggèrent que les cinq clades ont commencé à diverger d'un ancêtre commun il y a environ 9 400 ans. Le groupe des écureuils roux s'est séparé il y a environ 3 200 ans, la souche canadienne il y a 2 200 ans, et le groupe sud-américain/moderne il y a 1 500 ans. Les souches modernes des écureuils et des humains semblent s'être étendues très récemment, après le contact européen, il y a respectivement environ 100 ans et 280 ans. Ces périodes d'émergence estimées pour chacune des cinq lignées de *M. lepromatosis* confirment qu'elles ont toutes évolué au sein des Amériques.

Grâce aux nouvelles données génétiques disponibles, les chercheurs ont ré-estimé le moment où *Mycobacterium leprae* et *Mycobacterium lepromatosis* - les deux bactéries qui provoquent la lèpre - ont partagé pour la dernière fois un ancêtre commun. Cela est important pour comprendre depuis combien de temps ces deux espèces se sont séparées l'une de l'autre. En utilisant un type de modélisation génétique appelé « horloge moléculaire détendue », qui estime le temps en fonction du nombre de mutations accumulées, ils ont trouvé que la date de divergence la plus probable était d'environ 720 000 ans. Un second modèle, moins informatif, a donné une estimation plus ancienne, d'environ 2 millions d'années. Mais l'information essentielle est que les deux estimations sont bien plus récentes que les 14 millions d'années avancés par des études antérieures. Pourquoi est-ce important ?

cette estimation plus récente apparaît comme le scénario le plus plausible, compte tenu des similarités génomiques entre les deux espèces (92 % d'ANI, fig. S22) et de la symptomatologie très proche qu'elles induisent.

Parce que, même si les bactéries de ce groupe ont tendance à muter lentement - environ une modification génétique tous les cinq ans - , si les deux espèces s'étaient réellement séparées il y a 14 millions d'années, leurs génomes seraient devenus beaucoup plus différents qu'ils ne le sont. Or, elles partagent encore environ 92 % de leur ADN et provoquent des symptômes très similaires. Ces nouvelles estimations de divergence, plus jeunes, sont donc plus raisonnables et mieux étayées par les données.



**Fig. 3. Analyses d'horloge moléculaire de *M. lepromatosis* et estimation du temps de divergence par rapport à *M. leprae*.** (A) Arbre phylogénétique daté (calibré dans le temps) inféré de manière bayésienne à partir de 21 génomes de *M. lepromatosis* obtenus chez des écureuils roux (en italique, n = 6) de Grande-Bretagne et d'Irlande, ainsi que chez des

**Figure 3. Estimation de l'apparition de *M. lepromatosis* et de la date de sa séparation avec *M. leprae*.** (A) Cet arbre montre à quel point les souches de *M. lepromatosis* diffèrent les unes des autres et estime le moment où elles se sont séparées. Il inclut l'ADN de 21 génomes : 6 provenant d'écureuils roux du Royaume-Uni et d'Irlande, 13 issus de cas

humains contemporains ( $n = 13$ ) et anciens ( $n = 2$ ). L'arbre a été reconstruit avec BEAST v2.6.6 (49) en utilisant un modèle d'horloge moléculaire détendue. Les souches humaines contemporaines proviennent principalement des États-Unis et du Mexique, tandis que le génome de référence (NHDP-LPM-385) vient d'un patient du Costa Rica (29). Les étiquettes des nœuds indiquent les dates médianes de divergence (en années avant le présent) et les distributions colorées au-dessus de certains nœuds montrent les densités de probabilité a posteriori des âges des nœuds, estimées à partir de 1 000 échantillons postérieurs.

(B) Distributions a posteriori des temps de divergence estimés entre *M. leprae* et *M. lepromatosis* en utilisant des modèles d'horloge détendue avec deux a priori différents sur la distribution de la taille de population :  $1/X$  (maximalement non informatif) et exponentiel. Malgré les différences notables entre ces distributions a posteriori, les deux a priori déduisent un temps de divergence beaucoup plus récent que les 14 millions d'années précédemment estimés (24, 38).

Les études archéologiques, paléopathologiques et phylogénomiques antérieurs avaient écarté la présence de la lèpre dans les Amériques à l'époque pré-coloniale, faute de preuves ostéologiques claires dans les restes humains anciens du continent (44, 45), alors que de telles preuves sont couramment observées dans les vestiges archéologiques européens. Cette conclusion était également appuyée par des études phylogénomiques et d'ADN ancien montrant que la lèpre due à *M. leprae* avait été introduite dans les Amériques depuis l'Europe et l'Afrique, avec un impact certain sur les populations Autochtones (22, 23, 46). Nos résultats ne remettent pas en cause l'introduction de la lèpre à *M. leprae* par les populations européennes et africaines, mais ils révèlent que *M. lepromatosis* circulait déjà parmi les populations humaines du continent plusieurs siècles avant l'arrivée des Européens. Ils suggèrent donc fortement qu'une lèpre provoquée par *M. lepromatosis* constituait une maladie endémique de longue durée, à l'échelle du continent, avant l'époque européenne.

Malgré nos efforts de criblage approfondis sur des échantillons anciens et modernes, la diversité des souches de *M. lepromatosis* récupérées dans cette étude montre combien l'histoire évolutive et épidémiologique de ce pathogène reste méconnue. Les longues branches observées entre les nœuds et les extrémités de chaque clade, sans souches intermédiaires, suggèrent l'existence — ou l'ancienne existence — d'une diversité non échantillonnée, laissant d'importantes lacunes dans la chronologie de son évolution et de sa diffusion. Cela contraste avec la structure phylogénétique de *M. leprae* (fig. S7B), où un échantillonnage plus vaste couvrant davantage de périodes et de régions révèle des branches plus profondes et une diversification intra-lignée plus complexe. Ces observations soulignent l'urgence de poursuivre l'échantillonnage pour élucider la dynamique de transmission et la dispersion historique de *M. lepromatosis*, non seulement dans les Amériques, mais aussi dans d'autres régions comme l'Asie, où le pathogène est signalé mais dont le génome n'a pas encore été caractérisé.

Néanmoins, nos données fournissent des informations précieuses et suscitent de nouvelles hypothèses quant à l'histoire du pathogène. La découverte de souches anciennes en Amérique du Nord et en Amérique du Sud confirme que *M. lepromatosis* infectait les humains avant l'arrivée des Européens et qu'il était largement répandu sur le continent. La reconstruction phylogénétique révèle au moins cinq clades distincts, dont quatre présents dans les Amériques. La présence d'un cinquième clade infectant les écureuils roux des îles Britanniques soulève l'hypothèse que cette lignée est également originaire des Amériques, compte tenu de son temps de divergence estimé à 3 200

humains modernes (principalement aux États-Unis et au Mexique) et 2 provenant d'humains anciens. L'arbre a été construit à l'aide d'une horloge moléculaire détendue, qui permet aux chercheurs d'estimer la date de chaque divergence en se basant sur les taux de mutation. Les barres colorées indiquent le niveau de confiance associé à chaque date estimée.

(B) Ce graphique compare deux façons d'estimer l'époque à laquelle *M. lepromatosis* et *M. leprae* partageaient un ancêtre commun. Les deux méthodes suggèrent que ces bactéries se sont séparées beaucoup plus récemment qu'on ne le pensait — entre 700 000 et 2 millions d'années, plutôt qu'il y a 14 millions d'années, comme l'indiquaient deux publications précédentes.

Jusqu'à présent, les scientifiques pensaient que la lèpre n'existait pas dans les Amériques avant l'arrivée des Européens. Cette idée reposait sur deux observations : 1) les archéologues n'avaient pas trouvé de signes clairs de lèpre dans les squelettes anciens du continent, et 2) les études génétiques montraient que *M. leprae*, l'agent de la lèpre le plus connu, avait été introduit dans les Amériques par les Européens et les Africains après la colonisation. Cette étude confirme que *M. leprae* est bien arrivé avec les colonisateurs, mais elle apporte un élément nouveau : *M. lepromatosis* — l'autre bactérie responsable de la lèpre — était déjà présente dans les Amériques bien avant cela. La découverte de souches anciennes en Amérique du Nord comme en Amérique du Sud prouve que *M. lepromatosis* circulait parmi les peuples Autochtones depuis des siècles. Cela suggère qu'une forme de lèpre était déjà répandue sur l'ensemble du continent, faisant de la maladie une endémie américaine avant le contact européen.

Même si cette étude a examiné de nombreux échantillons anciens et modernes, l'histoire et la propagation de *M. lepromatosis* restent très mal connues. Dans les arbres phylogénétiques montrant les relations entre souches, les branches reliant les différentes souches de *M. lepromatosis* sont longues et comptent peu d'étapes intermédiaires ; cela signifie que de nombreuses souches anciennes ou stades évolutifs intermédiaires manquent encore dans nos données, laissant de grandes lacunes dans la chronologie de la diversification et de la diffusion de cette bactérie. C'est très différent pour *M. leprae*, mieux étudiée dans le temps et l'espace ; son arbre phylogénétique révèle une évolution beaucoup plus détaillée. Ces résultats soulignent l'importance de continuer à rechercher de nouveaux échantillons de *M. lepromatosis*, tant dans les Amériques que dans d'autres régions — en particulier en Asie, où la bactérie a été signalée mais demeure sans étude génétique approfondie.

Même si de nombreuses questions restent en suspens, les résultats de cette étude apportent des informations importantes et ouvrent de nouvelles pistes de recherche. La découverte de souches anciennes de *M. lepromatosis* en Amérique du Nord et en Amérique du Sud confirme que la bactérie infectait déjà les humains sur tout le continent bien avant l'arrivée des Européens. En analysant les relations génétiques entre les souches, cinq grands groupes (appelés clades) ont été identifiés, dont quatre proviennent d'échantillons prélevés dans les Amériques. Fait intéressant, le cinquième clade a été découvert chez des écureuils roux des îles Britanniques. Comme ce groupe semble s'être séparé des autres il y a environ 3 200 ans - bien avant tout contact connu entre les Amériques et l'Europe - il est très probable que ce clade d'écureuils

ans avant le présent, bien antérieur à toute connexion historique ou zoonotique probable entre les Amériques et d'autres continents qui aurait pu permettre une propagation transcontinentale du pathogène. De plus, la plupart des souches obtenues dans cette étude appartiennent à un seul clade de souches très similaires (PDDC), qui semble prédominer au Mexique et dans d'autres régions d'Amérique du Nord. Bien que cela suggère une prévalence plus élevée du pathogène en Amérique du Nord aujourd'hui, l'absence de *M. lepromatosis* dans la majorité des échantillons sud-américains testés peut aussi refléter des lacunes dans les efforts d'échantillonnage actuels plutôt qu'une véritable absence du pathogène dans la région.

La structure phylogénétique de *M. lepromatosis* soulève également des questions intéressantes quant aux événements de transmission récents. Le clade des écureuils roux présente une longue branche depuis le nœud racine jusqu'aux souches actuelles, suivie de branches terminales extrêmement courtes, ce qui suggère une longue période de diversité non échantillonnée avant une expansion récente. Le temps estimé jusqu'au plus récent ancêtre commun (MRCA) des souches d'écureuils roux, d'environ 105 AP, correspond aux introductions d'animaux médiées par l'homme dans les îles Britanniques et au Royaume-Uni (comme celle des écureuils gris au XIX<sup>e</sup> siècle (47)), ce qui étaye l'hypothèse d'un débordement récent depuis un réservoir américain. Toutefois, en l'absence de branches intermédiaires, on ne peut exclure qu'une introduction plus ancienne ait eu lieu à n'importe quel point de cette longue branche et depuis des régions encore largement non échantillonnées, l'expansion au Royaume-Uni et dans les îles Britanniques n'étant survenue que récemment. De même, le MRCA du clade PDDC, estimé à environ 280 AP, laisse penser que son expansion a pu être influencée par les bouleversements démographiques et sociaux consécutifs à la colonisation européenne. La période coloniale initiale a connu d'importants déplacements de population, des perturbations écologiques et des modifications des interactions homme-animal, autant de facteurs susceptibles d'avoir facilité la propagation du pathogène. Néanmoins, étant donné le nombre limité de souches analysées et l'absence d'échantillons couvrant diverses régions et époques, cela demeure une hypothèse plutôt qu'une conclusion définitive. Enfin, les estimations des temps de divergence entre les clades nord- et sud-américains indiquent que *M. lepromatosis* a connu une dispersion à grande échelle sur le continent durant l'Holocène tardif : le MRCA entre le PDDC et AR0018 est évalué à environ 1 500 BP, et le nœud séparant la souche ancienne XVII-B-357 en Amérique du Nord remonte à environ 2 200 AP. Ces résultats suggèrent des mouvements intercontinentaux du pathogène au cours des derniers millénaires, mais, en raison des vastes intervalles qui séparent ces clades, la direction de cette dispersion reste incertaine.

En intégrant la génomique des agents pathogènes anciens et modernes, cette étude constitue une première étape essentielle dans la reconstruction de l'histoire évolutive de *M. lepromatosis*. Toutefois, les larges lacunes évolutives non échantillonnées mises en évidence par nos reconstructions phylogénétiques soulignent l'ampleur des inconnues qui subsistent. Étendre les efforts d'échantillonnage à d'autres régions géographiques et périodes historiques, rechercher d'éventuels réservoirs non humains et renforcer la surveillance épidémiologique seront indispensables pour affiner notre compréhension de la dynamique de transmission,

soit en réalité originaire des Amériques. La plupart des souches modernes mises en évidence appartiennent à un groupe quasi génétiquement identique et très répandu au Mexique et dans certaines régions des États-Unis, ce qui suggère que la bactérie est plus commune aujourd'hui en Amérique du Nord. Cependant, il est aussi possible que des données manquent pour d'autres régions — comme l'Amérique du Sud — parce que nous n'y avons pas encore suffisamment échantillonné ou pas aux bons endroits.

La structure génétique de *M. lepromatosis* nous donne aussi des indices sur la façon dont il a pu se propager à des époques plus récentes. Le groupe de souches trouvées chez les écureuils roux des îles Britanniques présente un long vide dans son histoire génétique, suivi de la différenciation de souches aujourd'hui très proches. Ce schéma suggère que ce clade a existé et s'est diversifié pendant longtemps dans un lieu inconnu, sans être échantillonné, et que nous n'avons collecté que des souches issues d'une expansion récente. L'ancêtre commun estimé de ces souches d'écureuils remonte à environ 105 ans, ce qui correspond à l'introduction historique d'animaux comme les écureuils gris dans les îles Britanniques au XIX<sup>e</sup> siècle. Cela appuie l'idée que la maladie a pu être importée des Amériques avec des animaux, et n'a commencé à se répandre chez les écureuils du Royaume-Uni que plus récemment. Cependant, faute d'échantillons intermédiaires dans le temps ou l'espace, on ne peut exclure qu'une introduction plus ancienne ait eu lieu mais soit restée inaperçue. Un schéma similaire apparaît dans le clade PDDC — le groupe de souches courant en Amérique du Nord aujourd'hui — dont l'ancêtre commun récent date d'environ 280 ans. Cette période recoupe les débuts de la colonisation européenne, marqués par de grands bouleversements démographiques, environnementaux et dans les interactions homme-animal, qui ont pu faciliter la propagation du pathogène. Mais, étant donné le nombre limité de souches et le faible échantillonnage à travers les époques et les régions, cela reste une hypothèse plutôt qu'une preuve confirmée. Enfin, les temps de divergence estimées entre les souches nord- et sud-américaines indiquent que *M. lepromatosis* s'est largement dispersé sur le continent au cours des derniers millénaires. Par exemple, les clades PDDC et AR0018 partagent un ancêtre commun vieux d'environ 1 500 ans, et la souche ancienne d'Amérique du Nord (XVII-B-357) s'est séparée des autres il y a environ 2 200 ans. Cela signifie que la maladie se déplaçait probablement sur de longues distances dans les Amériques durant l'Holocène tardif, même si la direction exacte de sa propagation et les moyens impliqués demeurent inconnus.

En étudiant l'ADN d'échantillons anciens et modernes, ce travail réalise une première étape importante pour retracer l'histoire de *M. lepromatosis*. Cependant, de nombreuses parties de cette histoire manquent encore parce que nous ne disposons pas encore d'ADN provenant de suffisamment de périodes et de lieux. Pour comprendre pleinement comment cette bactérie s'est répandue et a survécu au fil du temps, il faudra obtenir davantage d'échantillons — en particulier provenant de différentes régions, époques et éventuels vecteurs animaux. Combler ces lacunes permettra non seulement de déterminer l'origine de *M. lepromatosis*, mais aussi la façon dont elle s'est propagée entre les humains et peut-être entre humains et animaux. Cela aidera également les scientifiques à mieux comprendre comment les maladies, en général, sont influencées par les déplacements

de la persistance et de la diffusion historique de ce pathogène. Comblant ces lacunes permettra non seulement de préciser l'origine et la trajectoire évolutive de *M. lepromatosis*, mais aussi d'offrir une perspective plus large sur l'interaction complexe entre migrations humaines, facteurs environnementaux et réservoirs zoonotiques structurant l'épidémiologie des maladies infectieuses.

humains, les changements environnementaux et les animaux susceptibles de porter des infections.

## References and Notes

### References

1. M. Woodman, I. L. Haeusler, L. Grandjean, Tuberculosis Genetic Epidemiology: A Latin American Perspective. *Genes (Basel)* **10**, 53 (2019).
2. Å. J. Vågene, *et al.*, Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. *Nat Commun* **13**, 1195 (2022).
3. V. Rougeron, J. Daron, M. C. Fontaine, F. Prugnolle, Evolutionary history of *Plasmodium vivax* and *Plasmodium simium* in the Americas. *Malar J* **21**, 141 (2022).
4. A. Koch, C. Brierley, M. M. Maslin, S. L. Lewis, Earth system impacts of the European arrival and Great Dying in the Americas after 1492. *Quaternary Science Reviews* **207**, 13–36 (2019).
5. C. F. Merbs, A new world of infectious disease. *Am. J. Phys. Anthropol.* **35**, 3–42 (1992).
6. D. A. Sack, R. B. Sack, G. B. Nair, A. Siddique, Cholera. *The Lancet* **363**, 223–233 (2004).
7. J. C. Riley, Smallpox and American Indians Revisited. *Journal of the History of Medicine and Allied Sciences* **65**, 445–477 (2010).
8. J. E. Bryant, E. C. Holmes, A. D. T. Barrett, Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS Pathog* **3**, e75 (2007).
9. A. A. Guzmán-Solis, *et al.*, Ancient viral genomes reveal introduction of human pathogenic viruses into Mexico during the transatlantic slave trade. *eLife* **10**, e68612 (2021).
10. J. R. Coura, P. A. Viñas, A. C. Junqueira, Ecoepidemiology, short history and control of Chagas disease in the endemic countries and the new challenge for non-endemic countries. *Mem Inst Oswaldo Cruz* **109**, 856–862 (2014).
11. T. Gann, Recent discoveries in central america proving the pre-columbian existence of syphilis in the new world. *The Lancet* **158**, 968–970 (1901).
12. M. M. M. Okumura, S. Eggers, The people of Jabuticabeira II: reconstruction of the way of life in a Brazilian shellmound. *HOMO* **55**, 263–281 (2005).
13. S. Eggers, C. C. Petronilho, K. Brandt, C. Jericó-Daminello, J. Filippini, K. J. Reinhard, How does a riverine setting affect the lifestyle of shellmound builders in Brazil? *HOMO* **59**, 405–427 (2008).
14. K. N. Harper, *et al.*, On the Origin of the Treponematoses: A Phylogenetic Approach. *PLoS Negl Trop Dis* **2**, e148 (2008).
15. A. Kocher, *et al.*, *Science* **374**, 182–188 (2021).
16. R. Barquera, *et al.*, Ancient genomes reveal a deep history of *Treponema pallidum* in the Americas. *Nature*, doi: 10.1038/s41586-024-08515-5 (2024).
17. K. I. Bos, *et al.*, Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
18. K. Köhler, *et al.*, Possible cases of leprosy from the Late Copper Age (3780–3650 cal BC) in Hungary. *PLoS ONE* **12**, e0185966 (2017).
19. A. S. Law, *Kalaupapa: A Collective Memory* (University of Hawai'i Press, Honolulu, 2012).

20. S. B. Marahatta, *et al.*, Perceived stigma of leprosy among community members and health care providers in Lalitpur district of Nepal: A qualitative study. *PLoS ONE* **13**, e0209676 (2018).
21. L. Santacroce, R. Del Prete, I. A. Charitos, L. Bottalico, *Mycobacterium leprae*: A historical study on the origins of leprosy and its social stigma. *Infez Med* **29**, 623–632 (2021).
22. A. Benjak, *et al.*, Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*. *Nat Commun* **9**, 352 (2018).
23. V. J. Schuenemann, *et al.*, Genome-Wide Comparison of Medieval and Modern *Mycobacterium leprae*. *Science* **341**, 179–183 (2013).
24. P. Singh, *et al.*, Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc Natl Acad Sci USA* **112**, 4459–4464 (2015).
25. M. Tió-Coma, *et al.*, Genomic Characterization of *Mycobacterium leprae* to Explore Transmission Patterns Identifies New Subtype in Bangladesh. *Front. Microbiol.* **11**, 1220 (2020).
26. C. Avanzi, *et al.*, Population Genomics of *Mycobacterium leprae* Reveals a New Genotype in Madagascar and the Comoros. *Front. Microbiol.* **11**, 711 (2020).
27. P. Dwivedi, M. Sharma, A. Ansari, P. Singh, Genetic diversity of *Mycobacterium leprae*: Need to move towards genome-wide approaches. *Indian J Med Res* **159**, 121 (2024).
28. L. Vera-Cabrera, W. G. Escalante-Fuentes, M. Gomez-Flores, J. Ocampo-Candiani, P. Busso, P. Singh, S. T. Cole, Case of Diffuse Lepromatous Leprosy Associated with “*Mycobacterium lepromatosis*.” *Journal of Clinical Microbiology* **49**, 4366–4368 (2011).
29. R. Sharma, *et al.*, Isolation of *Mycobacterium lepromatosis* and Development of Molecular Diagnostic Assays to Distinguish *Mycobacterium leprae* and *M. lepromatosis*. *Clin Infect Dis* **71**, e262–e269 (2020).
30. A. Virk, B. Pritt, R. Patel, J. R. Uhl, S. A. Bezalel, L. E. Gibson, B. M. Stryjewska, M. S. Peters, *Mycobacterium lepromatosis* Lepromatous Leprosy in US Citizen Who Traveled to Disease-Endemic Areas. *Emerg. Infect. Dis.* **23**, 1864–1866 (2017).
31. X. Y. Han, F. M. Aung, S. E. Choon, B. Werner, Analysis of the Leprosy Agents *Mycobacterium leprae* and *Mycobacterium lepromatosis* in Four Countries. *American Journal of Clinical Pathology* **142**, 524–532 (2014).
32. P. Dets, S. M. Collin, *Mycobacterium lepromatosis* as a Second Agent of Hansen’s Disease. *Front. Microbiol.* **12**, 698588 (2021).
33. X. Y. Han, K. C. Sizer, H.-H. Tan, Identification of the leprosy agent *Mycobacterium lepromatosis* in Singapore. *Journal of drugs in dermatology: JDD* **11**, 168–172 (2012).
34. C. Avanzi, *et al.*, Red squirrels in the British Isles are infected with leprosy bacilli. *Science* **354**, 744–747 (2016).
35. P. Singh, S. T. Cole, *Mycobacterium leprae* : genes, pseudogenes and genetic diversity. *Future Microbiology* **6**, 57–71 (2011).
36. X. Y. Han, N. A. Mistry, E. J. Thompson, H.-L. Tang, K. Khanna, L. Zhang, Draft Genome Sequence of New Leprosy Agent *Mycobacterium lepromatosis*. *Genome Announc.* **3**, e00513-15, /ga/3/3/e00513-15.atom (2015).
37. F. J. Silva, D. Santos-Garcia, X. Zheng, L. Zhang, X. Y. Han, Construction and Analysis of the Complete Genome Sequence of Leprosy Agent *Mycobacterium lepromatosis*. *Microbiol Spectr*, e01692-21 (2022).
38. N. Cardona-Castro, M. V. Escobar-Builes, H. Serrano-Coll, L. B. Adams, R. Lahiri, *Mycobacterium lepromatosis* as Cause of Leprosy, Colombia. *Emerg Infect Dis* **28**, 1067–1068 (2022).
39. S. T. Cole, *et al.*, Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007–1011 (2001).
40. J. Lindo, *et al.*, A time transect of exomes from a Native American population before and after European contact. *Nat Commun* **7**, 13175 (2016).
41. J. E. Buikstra, Ed., *Ortner’s Identification of Pathological Conditions in Human Skeletal Remains* (Academic Press, London, Third edition., 2019).

42. S. Mark, Early Human Migrations (ca. 13,000 Years Ago) or Postcontact Europeans for the Earliest Spread of *Mycobacterium leprae* and *Mycobacterium lepromatosis* to the Americas. *Interdisciplinary Perspectives on Infectious Diseases* **2017**, 1–8 (2017).
43. D. J. Ortner, W. Putschar, 1985 Identification of Pathological Conditions in Human Skeletal Remains. *Smithsonian Contributions to Anthropology* **28**.
44. S. Pfrengle, *et al.*, *BMC Biol* **19**, 220 (2021).
45. A. L. Signorile, P. W. W. Lurz, J. Wang, D. C. Reuman, C. Carbone, Mixture or mosaic? Genetic patterns in UK grey squirrels support a human-mediated ‘long-jump’ invasion mechanism. *Diversity and Distributions* **22**, 566–577 (2016).
46. Shimoyama, Y. (2024). pyGenomeViz: A genome visualization python package for comparative genomics (Version 1.2.0) [Computer software]. <https://github.com/moshi4/pyGenomeViz>.
47. R. Bouckaert, *et al.*, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**, e1006650 (2019).
48. Institut Pasteur. (2025). Owey platform – Repository used for ancient DNA data [Data repository]. Owey. <https://doi.org/10.48802/owey.9YsfbYxw>
49. M. Lopopolo, N. Rascovan, J. Fumey, G. Y. Ponce Soto, S. Duchene (2025). Code used for uncovering pre-European contact leprosy in the Americas and its enduring persistence [Code]. Zenodo. <https://doi.org/10.5281/zenodo.15396306>

## Acknowledgments

We acknowledge Susana Salceda and Gustavo Barrientos, chiefs of the Anthropological Division at Museum La Plata and Analía A. Lanteri, director of Museum La Plata for their support and excellent willingness to provide access to the museum collections for this research. A special recognition to Florencia Borella and the Community of Traun Kutral in Las Grutas for their help carrying out community engagement actions. A special mention to Malena Vazquez and Julio Avalos from RENYCOA for their assistance in carrying out the necessary procedures for the exportation of the archaeological material. We would like to thank Jaymes H. Collins and Roena Stevenson for excellent technical support. We would also thank the HPC Core Facility for their support for computational analyses and Marc Monot and Laurence Motreff from the Biomix Platform (supported by France Génomique (ANR-10-INBS-09-09 and IBISA) for their assistance in sample sequencing. The team in French Guiana would like to thank the patients and all the people of the teams that took part at Cayenne hospital and in particular Aline Masson for her meticulous cutting of the biopsies. We thank Julien Fumey (Microbial Paleogenomics Unit) for assisting with data upload and helping ML reorganize GitLab repositories and scripts for improved clarity, security, and reproducibility. We also thank Damien Mornico (Bioinformatics Hub, Institut Pasteur) for his help on the genomic annotation of genetic variants. Sequencing data from present-day samples is available at NCBI under BioProject PRJNA1188782 and ancient DNA data at the Owey platform (48). Code used for bioinformatic analyses is available via Zenodo (49). The skeletal remains of Ancestor XVII-B-357 are curated at the Canadian Museum of History, Canada. Requests to access these remains should be directed to the Lax Kw’alaams and Metlakatla First Nations, as guardians of this Ancestor. For Lax Kw’alaams, please contact Valeen Knott ([valeen\\_knott@laxband.com](mailto:valeen_knott@laxband.com)) or Mark Bowler ([lands\\_director@laxband.com](mailto:lands_director@laxband.com)). For the Metlakatla First Nation, please contact Barbara Petzelt ([metlch@metlakatla.ca](mailto:metlch@metlakatla.ca)). All remaining materials related to this individual, including DNA and raw sequencing data, have been returned to the communities. The skeletal remains of individuals AR0353 and AR0018 are curated at the Museum of La Plata, Argentina. No formal restrictions have been expressed by associated Indigenous communities regarding data reuse. However, future requests to access these data must be done through a formal access request process hosted on the Institut Pasteur’s Owey platform (49).

## Funding:

ERC-2020-STG - PaleoMetAmerica – 948800 (NR)

Institut Pasteur and CNRS UMR funding (NR)

INCEPTION program (Investissement d’Avenir grant ANR-16-CONV-0005) (NR, ML,SD)

Fondation pour la Recherche Médicale (FRM) “bourse de fin de thèse” PhD fellowship (ML)



The Heiser Program of the New York Community Trust for Research in Leprosy grants P18-000250 and P21-000127 (CA, JSS)

The Fondation Raoul Follereau (CA, MJ)

The Association de Chimiothérapie Anti-Infectieuse of the Société Française de Microbiologie (CA)

ML is a student from the FIRE PhD program funded by the Bettencourt Schueller foundation and the EURIP graduate program (ANR-17-EURE-0012).

Interagency agreement AAI23014-001\_24\_004 between HRSA and NIH/NIAID for supporting part of the study (LBA, RL).

The EPI-LEPR project is co-financed by the Centre Hospitalier de Cayenne, the Institut Pasteur de la Guyane, the European Union PO FEDER-FSE 2014-2020 Guyane (Synergie GY0012083) and the Labex CEBA (Centre d'Etude de la Biodiversité Amazonienne).

ERC-2023-SyG- Horse Power – 101071707 (LO)

NSF-BCS-2018200 (RSM)

SR was supported by the Novo Nordisk Foundation (NNF14CC0001 and NNF23SA0084103)

### **Author contributions:**

Conceptualization and supervision: NR

Experimental design: ML, CA, GT

Sampling: MdP, ADF, JSC, JM, VTM, LBA, NO, CW, AB, GMO, VPS, CFP, LVC, JGP, JSS, CGS, XYH, PSR, AJF, AFB, ANBF, COF, PNS, RS, PC, RHC, MSM, IEG, JSC, RL

Modern DNA extraction, qPCR screening and sequencing: CA, AKW, AJF, AFB, COF, SEGV, JH, KDA, VTM, KDA

Modern genome sequencing laboratory work and pathogen DNA enrichment: CA

Ancient DNA laboratory work: ML, GT, ADF, JM, MIO, LO, HS

Data analysis: ML, NR, SD, PL, GYPS, SN, FL, CA, CMP, RL

Archaeological contextualization: MdP, JSC, EAN

Funding acquisition: NR, CA, JSS, MJ, LBA, RL, RS, RSM

Writing – original draft: NR, ML, CA

Writing – review & editing: LQM, SR, LO, RSM

Community engagement: JSC, RSM, JM, NR, PL

**Competing interests:** Authors declare that they have no competing interests.

### **Data and materials availability:**

Modern sequencing data (from *M. lepromatosis* strains and the assembled NHDP-LPM-385 genome) are available at BioProject PRJNA1188782. The *de novo* assembled genome of NHDP-LPM-385 is available at NCBI under accession number CP155806.

Ancient DNA data are accessible through the Owey platform hosted by Institut Pasteur (37). For data derived from Ancestor XVII-B-357 (Canada), requests must be approved by the Lax Kw'alaams and Metlakatla First Nations (contact information in Acknowledgements). For individuals from Argentina (AR0353 and AR0018), access is granted via the Owey request process; community consultation may be added at their discretion.

The code used for bioinformatic analyses and supplementary tables are available through Zenodo (38).

### **Disclaimer**

The views expressed in this article are solely the opinions of the authors and do not necessarily reflect the official policies of the U. S. Department of Health and Human Services or the Health Resources and Services Administration, nor does mention of the department or agency names imply endorsement by the U.S. Government.

### **Supplementary Material**

Table of contents	
Materials and Methods	
Supplementary text	
Figs. S1 to S22	
Tables S1-S11	
Data tables S1-S4	
References 39 to 100	