

**THÈSE PRÉSENTÉE À L'UNIVERSITÉ D'ORLÉANS  
POUR OBTENIR LE GRADE DE  
DOCTEUR DE L'UNIVERSITÉ D'ORLÉANS**

**Discipline : Environnement et Santé**

PAR

**GUERNIER Vanina**

**Combiner analyse spatiale et épidémiologie  
pour l'aide à la décision dans la lutte contre  
la tuberculose en Guyane française**

Soutenue publiquement le 19 janvier 2006

**MEMBRES DU JURY :**

DEPARIS Xavier	Médecin de Santé Publique, DeSP, Paris	Examinateur
FOTSING Jean-Marie	Professeur, Université d'Orléans	Directeur de thèse
GUEGAN Jean-François	Directeur de Recherche, IRD, Montpellier	Co-directeur de thèse
GUISSO Nicole	Docteur es-sciences, Institut Pasteur, Paris	Rapporteur
PONCET Yveline	Directeur de Recherche, IRD, Orléans	Présidente, examinateur
VIEL Jean-François	Professeur, Université de Besançon	Rapporteur

*« Pour approfondir la médecine, il faut considérer d'abord les saisons, connaître la qualité des eaux, des vents, étudier les divers états du sol et le genre de vie des habitants »*

*Airs, eaux, lieux*  
Hippocrate (460 - 377 av. J.-C.)

# Remerciements

Si l'on tient compte, avant la thèse, du travail de recherche bibliographique et de la rédaction de mon projet de thèse (ou devrais-je dire mes...), de la soutenance orale pour l'obtention d'une bourse ministérielle, de l'attente pour l'attribution des bourses... Si l'on tient compte des trois mois passés, après la soutenance, à faire les corrections demandées par les différents membres du jury afin de proposer le manuscrit définitif (qui doit bien cacher encore quelques coquilles...), je crois que l'on peut estimer à quatre années l'investissement total de cette thèse. Et maintenant que je suis au bout, que c'est vraiment fini, maintenant que j'ai envoyé trois copies de mon manuscrit au service recherche de l'université, et que le soulagement d'en avoir fini s'installe, le temps m'est enfin donné de penser aux nombreuses personnes qui ont été près de moi pendant ces années de thèse et qui ont contribué, concrètement ou par leur soutien moral, à la réussite de ce travail. Il me tient aujourd'hui à cœur de les remercier...

Pour la construction initiale de mon projet de recherche, son insertion dans un cadre thématique porteur, et son acceptation par l'école doctorale, je tiens à remercier Françoise, Jean-Marie et Yveline du DEA Aden, qui ont porté le projet et soutenu ma candidature pour l'obtention de la bourse MNRT. Mon seul mérite n'aurait pas suffi sans leur soutien.

Je remercie Jean-François, Jean-Marie et Frédéric qui ont permis à cette thèse d'exister, malgré les difficultés du début, et les diverses réorientations du sujet. Merci aux membres de l'US ESPACE et du GEMI qui m'ont apporté le soutien logistique et financier dont j'avais besoin pour travailler, mais aussi pour les encouragements et l'intérêt qu'ils ont porté à mon sujet.

Parmi ceux qui ont permis à cette thèse de voir le jour, je dois beaucoup au Dr. Pierre Huguet, médecin inspecteur général à la DSDS, avec qui j'ai eu divers entretiens lors de mon arrivée en affectation en Guyane en août 2003, et qui m'a le premier intéressée au problème de la tuberculose en Guyane. Il m'a informé du travail réalisé par les différents médecins de Cayenne, et par le centre de référence des mycobactéries de l'Institut Pasteur de Guadeloupe. C'est de là que tout est parti...

Je remercie grandement Monique et Isabelle qui m'ont aidées très souvent, en particulier lors de la rédaction des conventions de recherche entre l'IRD et les différents instituts partenaires, et lorsque j'ai eu des problèmes de voyages, de logement, ou de sous, tout cela étant loin d'être mineur dans la gestion du quotidien, surtout lorsque l'on mène une vie de nomade... Merci à Frédéric qui a bien voulu supporter les frais supplémentaires de voyages et autres, du fait de la collaboration avec l'équipe de Guadeloupe.

Je remercie vivement les nombreuses personnes ayant contribué au recensement des cas de tuberculose de l'île de Cayenne, données qui étaient la clé de voûte de l'ensemble de ce travail : le Dr Deparis, médecin épidémiologiste à l'Institut Pasteur de la Guyane ; le Dr Guillot et le Dr Bourbigot, pneumologues au centre hospitalier de Cayenne ; le Dr Quintard, médecin au centre de Lutte Anti-Tuberculeuse de Cayenne ; le Dr Duval, médecin biologiste à l'Institut Pasteur de la Guyane ; le Dr Sobesky, directeur du Département d'Informatique Médicale de l'hôpital de Cayenne ; le Dr Cottrelle, médecin inspecteur général à la DSDS ; E. Miatti, secrétaire médicale en Médecine B ; le personnel du bureau des entrées de l'hôpital de Cayenne. Ces nombreuses personnes m'ont consacré un temps précieux alors que j'avais bien peu à leur apporter en échange (outre mes promesses de formidables résultats) et je les en remercie infiniment. Merci également au Dr Morvan, directeur de l'Institut Pasteur de la Guyane, qui a facilité nombre de mes démarches au sein de l'Institut, à Mr Le Behec, directeur du centre hospitalier de Cayenne, qui m'a autorisé l'accès aux données tuberculose, ainsi qu'aux médecins de l'hôpital qui m'ont, à un moment ou à un autre, consacré du temps : le Dr Louvel, le Dr Couppié, le Dr Nacher et le Dr Carme.

Je garde également de formidables souvenirs de ma mission tuberculose en pays amérindien, ce que je dois au Dr Michel Joubert coordinateur des centres de santé de Guyane, et à Dominique Polidori qui me l'a présenté. Merci aux membres de l'équipée sauvage, Danielle, Véra et Jérôme, alias « Docteur rasta » pour leur accueil, leur engagement, et leur cœur gros comme ça.

Merci également à Annelise et Jérôme du LRT de Cayenne pour leur aide concernant le recueil des 366 points GPS relevés au domicile des patients tuberculeux recensés sur l'île de Cayenne, recueil qui nécessitait un travail en binôme pilote / copilote. Un merci tout spécial à Annelise qui m'a servi de coach, d'encadrante, de prof, de guide touristique, et d'amie, en Guyane, puis à Montpellier. Une grande part de la réussite de ce travail, je le dois à son accueil au LRT de Cayenne à mon arrivée, et à l'amitié qu'elle m'a offerte, en plus des nombreuses heures passées à m'initier à la télédétection et aux Systèmes d'Information Géographique, et je n'ai pas toujours été une élève facile.

## REMERCIEMENTS

---

Merci également à Laurent, à Antoine et aux autres membres du LRT pour leurs différentes contributions.

En Guadeloupe, je remercie très sincèrement les membres du centre de référence des mycobactéries de l'Institut Pasteur de Pointe-à-Pitre. N'étant pas directement impliqués dans ma thèse au niveau administratif, ils l'ont été bien plus concrètement que sur le papier. Merci pour le temps qu'ils m'ont consacré, l'accueil qu'ils m'ont réservé, et les moments « hors labo » qui ont contribué au bon déroulement de cette partie du travail. Toute mon amitié à Nalin, Christophe, Séverine, Frédéric, Karine et Thierry, ainsi qu'aux autres membres du centre de référence des mycobactéries.

Merci aux (très) nombreuses personnes qui pendant mes nombreux déplacements en Guyane, en Guadeloupe et à Montpellier, ont bien voulu m'héberger, chercheurs en vacances ou amis de tous horizons. Mentions spéciales à Patrice, oscar du nombre d'appartements partagés, à mes cinq colocataires de Guyane, oscars de la bonne humeur, à Zac, oscar du « bienvenu, fais comme chez toi », et à Rachel, oscar de la difficulté, trois derniers mois de la thèse oblige.

Je tiens également remercier mes rapporteurs de thèse, en particulier le Pr Viel, pour leurs commentaires avisés, et l'ensemble des membres de mon jury pour avoir accepté de participer à la soutenance de ma thèse, ainsi que mes parents et amis qui se sont déplacés pour y assister.

Pour finir, difficile de citer tous les amis qui m'ont rendu, pendant toutes ces années, le travail plus léger, que ce soit les amis de longue date qui se demandaient perpétuellement sous quelle latitude je pouvais bien être, ou les amis rencontrés pendant la thèse, qui m'ont aidé à supporter les constants voyages et décalages horaires. Je ne veux pas ici tous les citer, au risque d'en oublier... Une pensée à tous mes collègues de la MTD, de l'IRD et du LRT, à toute la « Gwada team », en particulier l'équipe des Manouchs, à mes amis de Guyane, à ceux de Dijon et de Montpellier, et ceux de la Réunion... Ils ont enrichi ces années de moments inoubliables.

Je n'oublie pas mes parents, à qui cette thèse a causé presque autant de soucis qu'à moi. Et je n'oublie pas non plus mon frère... même si cette thèse ne lui a causé à lui aucun souci.

# Avant-propos

Ce travail de recherche doctorale s'inscrit dans le prolongement d'une formation universitaire de cinq ans en médecine, biologie générale, écologie et environnement, et d'un premier travail de recherche de D.E.A. concernant l'occurrence, la distribution spatiale, et la persistance d'espèces pathogènes de l'homme dans le monde, stage réalisé en 2002 à la Maison de la Télédétection (unité S140 ESPACE -Expertise et SPAtialisation des Connaissances en Environnement-) sous la direction de Jean-François Guégan (UMR 2724, GEMI -Génétique et Évolution des Maladies Infectieuses- IRD de Montpellier).

Le sujet de thèse initialement envisagé concernait l'identification des processus de risques majeurs dans l'émergence de la fièvre Q en Guyane française, et l'intérêt de la spatialisation pour la compréhension des dynamiques spatio-temporelles de la maladie dans un environnement modifié. Ce sujet devait prolonger les premiers résultats des recherches engagées en 2001 par Annelise Tran au Laboratoire Régional de Télédétection (unité ESPACE) de l'IRD de Cayenne sur l'intérêt de l'utilisation de la Télédétection pour l'étude de l'épidémiologie de la maladie. Mais des contraintes se sont présentées : fin du programme de trois ans financé par l'Institut Pasteur et départ du responsable de programme ; aucun résultat positif lors des tentatives d'isolement de la bactérie responsable de la maladie, que ce soit chez les individus malades ou chez des vecteurs animaux potentiels ; difficultés à récolter les résultats d'analyses des patients suspectés de fièvre Q auprès de l'hôpital de Cayenne... Le début de la thèse prenant du retard, il a été demandé une affectation de 6 mois à l'IRD de Cayenne, au sein de l'unité ESPACE, afin d'évaluer la possibilité d'un nouveau sujet directement en concertation avec les différents acteurs de la santé en Guyane. Cette affectation a pris effet le 5 août 2003. Après discussion avec le responsable du service épidémiologie du Département de la Santé et du Développement Social (DSDS), ainsi que des médecins de l'Institut Pasteur de Guyane et de l'hôpital de Cayenne, un nouveau sujet a pu être défini concernant la tuberculose. La relation à l'environnement était moins évidente que pour la fièvre Q, mais les données génétiques disponibles à l'Institut Pasteur de Guadeloupe concernant les bactéries isolées chez les patients guyanais semblaient pouvoir apporter une information intéressante du point de vue spatial. Le travail a donc été engagé en septembre 2003 en relation avec l'Institut Pasteur de Guyane, des médecins du centre hospitalier de Cayenne (en particulier les deux pneumologues), les

responsables successifs du département épidémiologie de la DSDS, et le service de Lutte Anti-Tuberculeuse de Cayenne. Une collaboration a également été mise en place entre l'IRD de Cayenne et l'équipe de recherche de l'Unité de la Tuberculose et des Mycobactéries de l'Institut Pasteur de Guadeloupe, afin de travailler sur les données génétiques. Au total, quatre missions de terrain ont été réalisées en Guyane, et deux en Guadeloupe, pour une durée totale effective d'un peu plus d'un an. Le présent mémoire est donc le fruit d'un travail de recherche de deux ans, qui fait suite à l'évolution progressive de la réflexion autour du sujet initial.

Cette recherche a été financée par le Ministère de la Recherche et de la Technologie. Le travail a été réalisé à l'Institut de Recherche pour le Développement (IRD), sous la co-direction de Jean-Marie Fotsing (unité IRD S140 ESPACE) et de Jean-François Guégan (UMR 2724 IRD-CNRS). Les deux équipes ont assuré le soutien logistique et financier pour les nombreuses missions de terrain réalisées en Guyane et en Guadeloupe. Le travail a été réalisé en collaboration étroite avec des équipes de l'Institut Pasteur de la Guyane et de l'Institut Pasteur de la Guadeloupe, après signature de conventions de recherche, et l'Unité de la Tuberculose et des Mycobactéries de l'Institut Pasteur de Guadeloupe a assuré un accueil de plusieurs mois au cours des missions réalisées en Guadeloupe. Un certain nombre de personnels médicaux ont par ailleurs pris part à la thèse, et ont contribué au bon déroulement du travail de terrain.

Ce travail de recherche a permis de concrétiser mon propre intérêt, et celui des partenaires scientifiques associés au travail, pour l'étude des relations entre l'homme, sa santé et son environnement, en particulier au travers des modifications anthropiques et des changements environnementaux. Son intérêt est de mettre en évidence l'importance de la prise en compte de l'environnement et de l'aspect spatial pour l'étude des maladies, y compris une maladie à transmission non-vectorielle telle que la tuberculose.

De part son caractère interdisciplinaire, ce mémoire s'adresse à des lecteurs de disciplines diverses. Pour faciliter la lecture transversale du document, certains aspects médicaux, méthodologiques et techniques ont été détaillés. Le chapitre 3, qui présente les différents outils et méthodes utilisés, a été créé dans le but d'alléger la lecture des chapitres ultérieurs. Des notes de bas de page, des encadrés méthodologiques insérés dans le document, et un glossaire final doivent également favoriser la compréhension du document pour tous les lecteurs néophytes intéressés de près ou de loin par ce travail.

### **Questions de déontologie sur l'utilisation de données nominatives**

Concernant l'aspect déontologique de l'étude, il est également important de souligner dès à présent que ce travail, qui fait appel à des données de santé directement et indirectement nominatives, garantit la protection des données

confidentielles, conformément aux règles d'éthique médicale et à la législation sur la recherche médicale et scientifique.

Les données patients utilisées dans le cadre de notre étude proviennent de trois établissements adhérents au réseau de soins pour la prise en charge de la tuberculose en Guyane : l'Institut Pasteur (de Guyane et de Guadeloupe), le Département d'Information Médicale de l'hôpital de Cayenne, et le service de Lutte Anti-Tuberculeuse. L'objectif d'un réseau de soins est d'améliorer la prise en charge d'une pathologie grâce à la coordination des professionnels de santé, et ainsi de favoriser la continuité des soins et l'utilisation des moyens existants de manière plus efficace. Le réseau de soins pour la prise en charge des patients tuberculeux sur l'île de Cayenne fonctionne de la façon suivante :

Durant la phase contagieuse tout d'abord, qui peut durer de deux à trois semaines, les malades sont pris en charge par le Centre Hospitalier de Cayenne (CHC) qui dispose de chambres d'isolement. Étant donné le délai de diagnostic bactériologique (4 semaines en moyenne), la mise en route du traitement anti-tuberculeux repose généralement sur un diagnostic anatomopathologique, c'est-à-dire la mise en évidence de lésions histologiques, et ce, dans un contexte clinique évocateur (affaiblissement, amaigrissement, etc.). Le Département d'Information Médicale (DIM) de l'hôpital de Cayenne collecte et archive informatiquement l'ensemble des informations médicales concernant les patients admis au CHC : nom, prénom, date de naissance, adresse, problèmes de santé et résultats des différentes analyses médicales pratiquées.

Certaines de ces analyses médicales ne peuvent pas être pratiquées à l'hôpital. Sur l'île de Cayenne, le laboratoire des mycobactéries de l'Institut Pasteur de la Guyane est en charge des examens directs (au microscope) des produits d'expectoration des patients tuberculeux, et de la mise en culture des prélèvements effectués à l'hôpital. Ces cultures sont ensuite envoyées à l'Institut Pasteur de Guadeloupe, au laboratoire régional de référence pour la tuberculose et les mycobactéries. Ce laboratoire est en charge de l'identification des souches de mycobactéries présentes sur la culture, de l'analyse génomique de ces souches, ainsi que de la réalisation des antibiogrammes (contrôle de la résistance des souches). Au final, les analyses réalisées au sein de l'Institut Pasteur permettent de confirmer l'existence d'une tuberculose-maladie et de vérifier la contagiosité de patient.

Lorsqu'il n'est plus contagieux et s'il n'a pas besoin d'autres soins, le patient est envoyé au centre de Lutte Anti-Tuberculeuse (LAT) pour poursuivre son traitement. Ce dispensaire est coordonné par le Conseil Général en lien avec le Département de la Santé et du Développement Social (DSDS) de Guyane. Il a pour mission le suivi et le traitement des malades tuberculeux à la sortie de leur phase d'hospitalisation. Le patient doit se rendre à la LAT une fois par mois, généralement pendant 6 à 9 mois, pour recevoir son traitement qui lui est délivré gratuitement. La LAT est également en charge des enquêtes de

proximité conduites autour des cas (voisins, personnes du même foyer, etc.) permettant de détecter les personnes ayant été en contact avec un individu malade, et d'engager si nécessaire un traitement préventif.

Au sein des différents établissements du réseau de soins s'échangent donc des informations personnelles et médicales concernant les patients. Et étant donné le circuit de prise en charge des patients tuberculeux dans le réseau de soins, détaillé ci-dessus, les cas recensés auprès de chacun des établissements du réseau devraient être les mêmes, et également identiques aux patients ayant fait l'objet d'une déclaration obligatoire auprès de la DSDS. Sur l'île de Cayenne, le partage électronique de données n'existe pas, et les différents membres du réseau de soins font appel à des dossiers « papier », ce qui peut augmenter le risque d'erreurs lors de la transmission d'information. Or, des incohérences ont été notées au début de l'étude entre les chiffres relevés dans les bilans d'activité annuels de l'Institut Pasteur et les déclarations obligatoires publiées par l'Institut de Veille Sanitaire. Partant de ce constat, l'un des objectifs de notre étude était de vérifier si les patients étaient correctement recensés au sein des trois organismes du réseau de soins, et donc d'évaluer l'efficacité de la prise en charge de la tuberculose sur l'île de Cayenne.

Afin de recouper les cas recensés auprès des différentes sources (Institut Pasteur, hôpital de Cayenne et LAT), les enquêtes que nous avons menées incluent nécessairement la collecte de données directement et indirectement nominatives telles que le nom, le prénom, le sexe et la date de naissance. Conformément aux règles d'éthique médicale et à la législation sur la recherche médicale et scientifique, deux conventions de recherche ont été mises en place spécifiquement dans le cadre du travail de thèse. La convention de recherche établit un lien juridique et pédagogique étroit entre la structure d'accueil de la thèse, ici l'Institut de Recherche pour le Développement (IRD), et un établissement public ou privé avec lequel l'étudiant collabore, en l'occurrence la DSDS et l'Institut Pasteur. Un engagement de confidentialité a par ailleurs été signé, conjointement avec les directeurs des différents organismes ayant donné accès à leurs données, cet accord stipulant que je m'engage à ne porter aucune atteinte au secret professionnel qui s'attache aux informations nominatives auxquelles j'ai eu accès dans le cadre de mon travail de thèse, et qu'en retour les parties sont convenues de m'accorder un accès aux données concernant la tuberculose en Guyane (*cf.* document en Annexe 1). Après croisement des données des différentes sources, et recoupement des cas communs, un fichier secondaire réunissant l'ensemble des informations a été créé et anonymé, avant d'être utilisé pour la cartographie des cas grâce à un Système d'Information Géographique. Les adresses des patients ont également été éliminées de la base de données, qui de ce fait ne constitue pas un fichier indirectement nominatif. Seul ce fichier a été conservé. Ces aspects déontologiques sont repris au chapitre 2 du présent mémoire.

# Table des matières

<b>Avant propos</b>	<b>i</b>
<b>Table des figures</b>	<b>xiii</b>
<b>Liste des tableaux</b>	<b>xxi</b>
<b>Introduction</b>	<b>1</b>
<b>I. Contexte général : de la tuberculose en général et de la Guyane en particulier</b>	<b>7</b>
<b>1. Généralités sur la tuberculose humaine</b>	<b>7</b>
1.1. La tuberculose dans le monde	7
1.1.1. Les estimations officielles	8
1.1.2. Évolution de l'incidence	8
1.2. Données médicales sur la tuberculose humaine	9
1.2.1. Les données cliniques	10
1.2.1.1. L'agent responsable de la maladie	10
1.2.1.2. Les différentes formes cliniques	10
1.2.1.3. La transmission de la maladie	10
1.2.1.4. Les phases évolutives de la maladie	11
1.2.2. Le diagnostic	12
1.2.2.1. L'intradermoréaction à la tuberculine	12
1.2.2.2. La radiographie pulmonaire standard	13
1.2.2.3. La recherche bactériologique	13
1.2.2.4. Typage et antibiogramme	13
1.2.2.5. Cas des tuberculoses extra-pulmonaires	14
1.2.3. Le traitement	15
1.2.3.1. Les différents types de traitement	15
1.2.3.2. Délai de mise en route du traitement	15
1.2.3.3. Délai de guérison	16
1.3. La tuberculose à bacilles multirésistants	16
1.3.1. Causes et implications	16
1.3.2. Traitement	16
1.4. Contraintes liées aux caractéristiques intrinsèques de la	16

tuberculose et implications en santé publique	
1.4.1. Contraintes diagnostiques	16
1.4.2. Contraintes liées au pathogène	17
1.4.3. Contraintes associées au traitement	17
1.5. Bilan	17
<b>2. Le contexte guyanais</b>	<b>18</b>
2.1. Caractéristiques géographiques générales	18
2.2. Quelques caractéristiques socio-démographiques	19
2.2.1. Une multi-ethnicité	19
2.2.2. Une population très jeune et en forte croissance	20
2.2.3. Des difficultés sociales et économiques	20
2.3. Le contexte sanitaire	21
2.3.1. Un territoire mal maîtrisé	21
2.3.2. Un réseau sanitaire insuffisant	21
2.3.3. Situation épidémiologique	22
2.3.4. Des infrastructures urbaines très inégales	22
2.3.5. Un taux élevé de Sida	23
2.4. Bilan de la situation en Guyane	23
<b>3. Conclusion</b>	<b>24</b>
<b>II. Données chiffrées de la tuberculose en Guyane : analyse critique</b>	<b>25</b>
<b>1. La tuberculose en Guyane en chiffres</b>	<b>25</b>
1.1. Les données officielles	25
1.1.1. Les sources de données	26
1.1.2. Définition d'un cas	26
1.1.3. Les données de l'Institut de Veille Sanitaire	26
1.1.4. Les données du Département de la Santé et du Développement Social (DSDS)	27
1.1.5. Au final, les données officielles sont-elles fiables ?	29
1.2. Les données issues d'enquêtes	29
1.2.1. Éthique et déontologie de l'étude	29
1.2.2. Les sources de données	30
1.2.2.1. Le centre hospitalier de Cayenne	30
1.2.2.2. Les laboratoires de référence	31
1.2.2.3. Les centres de soins et de prévention	31
1.2.3. Choix de la période d'étude	32
1.2.4. Méthode : croisement des trois sources de données	32
1.2.5. Résultats	33
1.2.6. Exhaustivité des sources d'information	34
1.2.7. Au final, des données officielles non exhaustives	35

<b>2. Épidémiologie de tuberculose en Guyane</b>	<b>36</b>
2.1. Cas introduits dans l'étude	36
2.2. Résultats	37
2.2.1. Répartition des cas par ville (1996-2003)	37
2.2.2. Répartition par âge et par sexe des cas de 1996-2003	38
2.2.2.1. Répartition des cas par tranches d'âge	38
2.2.2.2. Répartition des cas en fonction du sexe	41
2.2.2.3. Répartition des cas par âge et par sexe	41
2.2.3. Répartition des cas en fonction de la nationalité	43
2.2.3.1. Prise en compte de la nationalité seule	44
2.2.3.2. Répartition des cas par nationalité et par tranches d'âge	45
2.2.4. Le taux de co-infection tuberculose / VIH	48
2.3. Discussion	48
2.3.1. Répartition par âge et par sexe	48
2.3.2. Pic d'incidence chez les personnes âgées	48
2.3.3. Pic d'incidence chez les jeunes adultes	48
2.3.4. Patients de nationalité étrangère	49
2.3.5. Co-infections tuberculose / VIH	49
2.4. Bilan des facteurs de risque humains	49
<b>3. Évaluation statistique de l'incidence actuelle de la tuberculose en Guyane</b>	<b>50</b>
3.1. La méthode de capture-recapture	50
3.1.1. Historique de la méthode	50
3.1.2. Application en épidémiologie	51
3.1.3. Principe et conditions d'application	52
3.2. Matériel et méthodes	52
3.2.1. Les sources de données	52
3.2.2. Respect des conditions de validité	53
3.2.3. Les modèles utilisés	54
3.2.3.1. Les estimateurs de Petersen et Chapman	54
3.2.3.2. Les modèles log-linéaires	55
3.2.3.3. L'approche « <i>sample coverage</i> »	56
3.3. Résultats	57
3.3.1. Simple croisement des trois sources de données	57
3.3.2. Résultats de capture-recapture	57
3.3.2.1. Méthode de Petersen et Chapman	57
3.3.2.2. Modèles log-linéaires	58
3.3.2.3. L'approche « <i>sample coverage</i> »	59
3.3.3. Récapitulatif des résultats en termes de nombre de cas	60
3.3.3.1. Sur l'île de Cayenne	61
3.3.3.2. Sur l'ensemble de la Guyane	62

3.3.4. Résultats en termes d'incidence de la tuberculose	62
3.3.4.1. Incidence théorique espérée de l'Ile de Cayenne	62
3.3.4.2. Incidence estimée sur l'ensemble de la Guyane	63
3.4. Discussion	65
3.4.1. Détection et traitement des cas	65
3.4.1.1. Taux de détection des nouveaux cas	65
3.4.1.2. Taux de réussite des traitements	66
3.4.1.3. Les cas de tuberculose multi-résistante	66
3.4.2. Améliorer le système de prise en charge	67
3.4.2.1. La stratégie DOTS	67
3.4.2.2. Meilleure prise en charge de la tuberculose en Guyane	69
<b>4. Conclusion</b>	<b>70</b>
<b>III. Relever le défi de la lutte contre la tuberculose en Guyane : quels outils pour quels besoins ?</b>	<b>71</b>
<b>1. La tuberculose humaine en Guyane : quels sont les besoins, et comment y répondre</b>	<b>71</b>
1.1. Comment améliorer le contrôle de la maladie ?	71
1.2. Quelles sont les questions à résoudre ?	72
1.2.1. Objectifs généraux de l'étude et problématique	72
1.2.2. Quand ? La dynamique temporelle de la maladie	73
1.2.3. Où ? La dynamique spatiale de la maladie	74
1.2.4. À qui ? Les individus à risque de tuberculose	75
<b>2. L'analyse spatiale appliquée aux maladies</b>	<b>75</b>
2.1. État de l'art	75
2.1.1. Analyse spatiale en écologie	75
2.1.2. Analyse spatiale en épidémiologie	76
2.2. Les outils de la spatialisation et de la géomatique	77
2.2.1. Les outils de télédétection	77
2.2.2. Les Systèmes d'Information Géographique	78
2.3. Application de ces outils à notre étude	80
2.3.1. Application à la tuberculose	80
2.3.2. Application au cas de la Guyane	80
<b>3. Les apports de l'épidémiologie moléculaire</b>	<b>81</b>
3.1. Épidémiologie et épidémiologie moléculaire	81
3.2. Épidémiologie moléculaire et tuberculose	82
3.3. Limites de la méthode	83
<b>4. Intérêt de la combinaison des deux approches</b>	<b>84</b>

<b>5. Conclusion</b>	<b>85</b>
<b>IV. Analyse des relations entre environnement spatial et tuberculose sur l'Île de Cayenne</b>	<b>87</b>
<b>1. Distribution spatiale des cas de tuberculose sur l'Île de Cayenne</b>	<b>87</b>
1.1. Distribution spatiale des cas de tuberculose	88
1.2. Distribution de l'incidence des cas	91
1.2.1. L'incidence de tuberculose par IRIS	91
1.2.1.1. Matériel et méthodes	91
1.2.1.2. Résultats	93
1.2.1.3. Discussion	93
1.2.2. L'incidence de tuberculose par zone habitée	94
1.2.2.1. Matériel et méthodes	94
1.2.2.2. Résultats	95
1.2.2.3. Discussion	97
1.3. Distribution des ratios standardisés d'incidence	97
1.3.1. Matériel et méthodes	98
1.3.2. Résultats	99
1.3.3. Discussion	102
1.4. Améliorer ces résultats	102
1.4.1. Des méthodes fiables validées : revue de détails	102
1.4.1.1. Estimation de l'incidence par télédétection	102
1.4.1.2. Classification du bâti par télédétection	103
1.4.2. Typologie d'habitat et incidence : nos résultats	103
1.4.2.1. Typologie d'habitat	103
1.4.2.2. Calcul de la population par pixel	105
1.5. Bilan préliminaire	107
<b>2. Identification des agrégats spatiaux</b>	<b>108</b>
2.1. Définition d'un agrégat	108
2.2. Statistiques appliquées à la recherche d'agrégats	108
2.2.1. État de l'art	108
2.2.2. Comparaison de 3 méthodes	108
2.3. La méthode « <i>spatial scan statistic</i> »	109
2.3.1. Le logiciel SaTScan	109
2.3.2. Application de la méthode à nos données	111
2.3.2.1. Les données utilisées	111
2.3.2.2. Les paramètres d'analyse choisis	112
2.4. Résultats	112
2.4.1. Dans les 35 IRIS, par âge et par sexe	112
2.4.1.1. Agrégats significatifs	112

2.4.1.2. Cartographie des ratios standardisés d'incidence ( <i>RIS</i> )	114
2.4.2. Dans les 35 IRIS, par nationalité et par sexe	115
2.4.2.1. Agrégats significatifs	115
2.4.2.2. Cartographie des ratios standardisés d'incidence ( <i>RIS</i> )	116
2.4.3. À l'échelle des quartiers, sans inclure de covariable	117
2.5. Discussion	118
2.5.1. Comparaison entre les trois approches	118
2.5.2. Agrégats de tuberculose et habitat insalubre	119
2.5.2.1. Définitions	119
2.5.2.2. Résultats	120
2.5.2.3. Recherche d'agrégats avec la covariable insalubrité	121
<b>3. Conclusion</b>	<b>122</b>
<b>V. Épidémiologie moléculaire de la tuberculose en Guyane</b>	<b>123</b>
<b>1. Identification des grappes génétiques de <i>Mycobacterium tuberculosis</i></b>	<b>123</b>
1.1. Introduction	123
1.1.1. Typage moléculaire des souches	123
1.1.2. Grappes génétiques	124
1.2. Méthodes de typage moléculaire	124
1.2.1. Critères de choix d'une méthode de typage	124
1.2.2. Méthodes de typage de <i>Mycobacterium tuberculosis</i>	125
1.2.2.1. Les méthodes de référence	125
1.2.2.2. Quelles méthodes pour quelles données ?	126
1.2.2.3. Application à nos propres données	127
1.2.3. Principe général du spoligotypage	127
1.2.4. Principe général du typage par VNTR	129
1.2.5. Mise en œuvre des analyses moléculaires	131
1.3. Résultats	132
1.3.1. Spoligotypes identifiés (ou shared-type, ou ST)	132
1.3.2. Les grappes génétiques	136
1.4. Discussion	139
1.4.1. Cas en grappes et contamination	139
1.4.2. Spécificité géographique	141
1.5. Conclusion	141
<b>2. Diversité des souches de <i>Mycobacterium tuberculosis</i> et phylogénie</b>	<b>142</b>

2.1. Introduction	142
2.1.1. La méthode n et la méthode n-1	142
2.1.2. Taux de mutation des marqueurs moléculaires	143
2.2. Méthodes	143
2.2.1. L'indice de mutation-transmission	143
2.2.2. Visualiser les liens génétiques entre souches	144
2.2.3. Phylogénie de <i>Mycobacterium tuberculosis</i>	146
2.2.4. Construction des « <i>cluster-graphs</i> »	149
2.3. Résultats	149
2.3.1. Les familles de spoligotypes	149
2.3.2. Les « <i>cluster-graphs</i> »	152
2.3.3. Méthode n, méthode n-1 et TMI	156
2.3.4. Phylogénie spoligotype + VNTR	156
2.4. Discussion	159
2.4.1. Arbre couvrant minimal et phylogénie manuelle	159
2.4.2. <i>Cluster-graphs</i>	159
2.4.3. Transmission récente de la tuberculose en Guyane	160
2.4.4. Susceptibilité génétique et tuberculose	161
<b>3. Conclusion</b>	<b>162</b>
<b>VI. La dynamique de transmission de la tuberculose en Guyane</b>	<b>165</b>
<b>1. Mise en relation des agrégats spatiaux de cas et des grappes génétiques</b>	<b>165</b>
1.1. Introduction	165
1.1.1. Combiner SIG et épidémiologie moléculaire	165
1.1.2. Hypothèses de travail	166
1.2. Méthode	166
1.2.1. Les données	166
1.2.2. Les analyses	168
1.2.2.1. Transmission active dans les IRIS	168
1.2.2.2. Transmission dans les agrégats géographiques	169
1.3. Résultats	169
1.3.1. Pourcentage de cas en grappes par IRIS	169
1.3.2. Pourcentage de cas en grappes par agrégat	170
1.4. Discussion	171
1.4.1. Cas en grappes dans les agrégats géographiques	171
1.4.2. Cas en grappes dans les IRIS	171
1.4.3. Discussion des résultats	171
1.5. Bilan	171
<b>2. Analyse cartographique du risque tuberculeux en Guyane et sur l'île de Cayenne</b>	<b>172</b>

2.1. Introduction	172
2.1.1. Définition du risque	172
2.1.2. Application à l'épidémiologie	173
2.1.3. Cas de la tuberculose humaine	174
2.2. Cartographies de l'aléa, de la vulnérabilité et du risque tuberculeux sur l'île de Cayenne	175
2.2.1. Le Système d'Information Géographique	176
2.2.2. Cartographie de l'aléa	176
2.2.2.1. Méthode	176
2.2.2.2. Résultats	178
2.2.3. Cartographie de la vulnérabilité	179
2.2.3.1. Méthode	179
2.2.3.2. Résultats	181
2.2.4. Cartographie du risque tuberculeux	182
2.2.4.1. Méthode	182
2.2.4.2. Résultats	183
2.3. Discussion	184
2.3.1. Un outil d'aide à la décision	184
2.3.2. Des indicateurs simples et accessibles	184
2.3.3. Surveillance et retour d'information	185
2.3.4. Validation du modèle	185
<b>3. Conclusion</b>	<b>185</b>
<b>Conclusion générale</b>	<b>187</b>
<b>Bibliographie</b>	<b>193</b>
<b>Glossaire</b>	<b>221</b>

# Table des figures

1.1	Incidences estimées de la tuberculose par pays en 2003 [WHO, 2005]. Remarque : la Guyane est dans la classe « No estimate », pas d'estimation.	8
1.2	Organigramme de confirmation d'un cas de tuberculose. IDR, intradermoréaction ; ED, examen direct ; TC, tableau clinique.	14
1.3	Situation géographique de la Guyane française.	19
1.4	Les structures sanitaires de Guyane. (Source : DSDS de Guyane. Photo : Vanina Guernier, 2004).	22
2.1	Croisement des données de l'IPG, de la LAT et du DIM, 1996-2003. Le nombre de cas présents dans une, deux ou trois sources est noté en gras. Le point d'interrogation illustre le nombre de cas absents de ces sources. N est le nombre de cas recensés dans chacune des 3 sources entre 1996 et 2003.	34
2.2	Répartition géographique des cas de tuberculose confirmés bactériologiquement, de 1996 à 2003. Les données sont issues de l'enquête menée à l'Institut Pasteur de Guyane.	37
2.3	Courbes de distribution des âges (a) de la population totale de Guyane au dernier recensement 1999 et (b) des cas de tuberculose confirmés recensés sur l'ensemble de la Guyane entre 1996 et 2003, par tranches de cinq ans.	38
2.4	Taux d'incidence de la tuberculose par tranche d'âge, calculé en rapportant, pour chaque tranche d'âge, le nombre de cas de tuberculose confirmés de 1996-2003 (soit 8 ans) à 100000 habitants. Ce rapport est ensuite divisé par huit pour obtenir une incidence annuelle.	40
2.5	Pourcentage homme/femme (A) de la population totale de Guyane au dernier recensement 1999 et (B) des cas confirmés de tuberculose 1996-2003 recensés en Guyane.	41
2.6	Pyramide des âges (A) des cas confirmés de tuberculose recensés sur l'ensemble de la Guyane entre 1996 et 2003, et (B) de la population totale de Guyane d'après le dernier recensement de l'INSEE réalisé en 1999.	42
2.7	Répartition des cas confirmés de tuberculose recensés en Guyane entre	45

TABLE DES FIGURES

---

	1996 et 2003 en fonction de leur nationalité (en pourcentage du total de cas). Les patients d'origine « France » incluent les créoles, métropolitains, amérindiens, etc. ayant la nationalité française.	
2.8	Taux d'incidence moyen de la tuberculose par tranche d'âge (1) chez les guyanais d'origine française et (2) d'origine étrangère (exprimés respectivement en nombre de cas de tuberculose / 100 000 français / an ou / 100 000 étrangers /an).	46
2.9	Pourcentage de cas par tranche d'âge sur l'ensemble des cas de tuberculose (1) de nationalité française et (2) de nationalité étrangère, recensés entre 1996 et 2003.	47
2.10	Principe de la méthode de capture-recapture : (a) capture d'un premier échantillon N1, (b) marquage (symbolisé par des croix) des individus capturés, qui sont ensuite relâchés, (c) capture d'un second échantillon N2, et (d) comptage des individus marqués n2.	51
2.11	Évolution annuelle du nombre de tuberculeux, Ile de Cayenne, 1996-2003. Sources : Département de la Santé et du Développement Social (DSDS), Institut Pasteur de Guyane (IPG), service de lutte anti-tuberculeuse (LAT), département d'information médicale de l'hôpital de Cayenne (DIM), après croisement de trois sources (IPG*DIM*LAT) et après correction par capture-recapture selon l'approche <i>sample coverage</i> (+11,6%).	61
3.1	Illustration de la superposition des couches thématiques dans un Système d'Information Géographique. Ces informations géoréférencées concernent ici les paysages, le sol, l'hydrographie, et la topographie. (Source : <a href="http://www.cev.washington.edu/lc/CEVIMAGES/dhsvm-static-layers.jpg">http://www.cev.washington.edu/lc/CEVIMAGES/dhsvm-static-layers.jpg</a> ).	79
4.1	Répartition géographique des cas de tuberculose résidant sur l'Ile de Cayenne à la date de déclaration de la maladie (cas du 1 <sup>er</sup> janvier 1996 au 31 décembre 2003). Le découpage en IRIS (Îlots Regroupés pour l'Information Statistique) correspond aux unités spatiales de référence utilisées pour les recensements de l'INSEE.	90
4.2	(A) Zones urbanisées de l'Ile de Cayenne et (B) incidence annuelle de tuberculose par IRIS sur l'Ile de Cayenne (nombre de nouveaux cas par an pour 100 000 habitants). (cf. encadré 4.2 pour le choix du nombre de classes et des valeurs extrêmes des classes).	93
4.3	Le quartier de la Madeleine sur l'Ile de Cayenne. (A) Image SPOT 5-panchromatique du 29 août 2003 (programme ISIS, © CNES, distribution Spotimage), (B) photographie aérienne (© IGN) et (C) résultat de notre photo-interprétation, c'est-à-dire polygones de bâti habité et non habité en 2003 obtenus sous GeoConcept (chaque polygone indépendant correspond à ce que nous avons appelé un « quartier »).	95
4.4	(A) Densité de population estimée par quartier (en habitants par km <sup>2</sup> ),	96

et (B) incidence annuelle de la tuberculose par quartier. Ces deux calculs reposent sur les estimations de populations par quartier obtenues en redistribuant la population totale de l'IRIS sur les surfaces réellement urbanisées (§1.2.2.1.). Les encadrés (a) et (b) représentent des zooms respectivement de (A) et (B) sur une même zone correspondant au centre ville de Cayenne (marquée par un rectangle coloré sur les vues générales A et B).

4.5	Cartographies sur l'île de Cayenne des ratios standardisés d'incidence ( <i>RIS</i> ) à l'échelle des IRIS, en tenant compte (A) de la répartition par âge et par sexe, et (B) de la répartition par sexe et par nationalité de la population (cartes obtenues avec le logiciel GeoDa). Les IRIS numérotés de 1 à 7 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette ; 5 Mango ; 6 Mont Baduel ; 7 Balata. En fonction des covariables prises en compte, seuls sont indiqués les IRIS où l'on note une sur-incidence de tuberculose (cf. tableau 4.1).	99
4.6	Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés avec les covariables sexe et nationalité à l'échelle des IRIS, (A) sans notion de contiguïté spatiale, et (B) en tenant compte de la contiguïté spatiale. Résultats obtenus avec le logiciel GeoDa. Les IRIS numérotés de 1 à 4 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette.	100
4.7	Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés avec les covariables âge et sexe à l'échelle des IRIS, (A) sans notion de contiguïté spatiale, et (B) en tenant compte de la contiguïté spatiale. Résultats obtenus avec le logiciel GeoDa. Les IRIS numérotés de 1 à 4 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette.	101
4.8	Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés à l'échelle des quartiers. Seul le résultat sans notion de contiguïté spatiale (lissage global) est présenté. En effet, dans la mesure où tous les quartiers ne sont pas contigus et où il existe des espaces non habités, il n'est pas possible d'appliquer la seconde méthode.	101
4.9	La commune de Cayenne. (A) Image binaire du bâti (valeur 1 pour le bâti, et 0 pour le non bâti, respectivement en blanc et en noir) [Tran <i>et al.</i> , 2002], (B) typologie d'habitat 2003 par quartier (réalisée au cours de cette étude), et (C) typologie d'habitat 2003 par pixel, obtenue en appliquant le masque (A) sur l'image (B).	104
4.10	Schéma illustrant les difficultés inhérentes à la manipulation simultanée des formats d'image vecteur (le quartier) et raster (les pixels d'habitats). Le but est de compter le nombre de pixels de chaque couleur inclus dans le quartier. Visuellement, la solution est simple, mais pas de façon automatisée car les SIG manipulent rarement les deux formats.	107
4.11	Unique agrégat significatif de tuberculose identifié à l'échelle de l'IRIS en tenant compte des covariables âge et sexe. À droite, détail des IRIS	114

- inclus dans cet agrégat. RIS, ratio standardisé d'incidence.
- 4.12 Unique agrégat significatif de tuberculose identifié à l'échelle de l'IRIS en tenant compte des deux covariables nationalité et sexe. À droite, détail des IRIS inclus dans cet agrégat de tuberculose. RIS, ratio standardisé d'incidence. 116
- 4.13 (A) Agrégats de tuberculose identifiés par l'analyse SaTScan à l'échelle des quartiers ; (B) Agrégat principal (n°1) et agrégat secondaire n°2 ; (C) Agrégats secondaires n°4, 6 et 8 ; (D) Agrégat secondaire n°5. À cette échelle, les agrégats n°3 et 7 ne sont pas visibles. 118
- 5.1 (A) Structure du locus DR du génome des mycobactéries. (B) Amplification *in vitro* de la région DR par PCR grâce aux amorces DRa et DRb. Chaque DR du locus peut servir de départ à l'amplification, donc le résultat d'amplification donne un mélange de fragments d'ADN de taille multiples. (C) Détection des sondes inter-DR (IDR) par hybridation sur un système matriciel, à gauche (membrane sur laquelle sont liées de façon covalente les 43 séquences oligonucléotidiques correspondant aux 43 espaceurs). À droite, photographie après révélation par chimioluminescence (ECL, Amersham). (D) Motifs d'hybridation (spoligotypes) des souches de *M. tuberculosis* testées (une souche par ligne). La coloration noire indique la présence de l'espaceur dans la souche considérée alors que le blanc indique l'absence de cet espaceur. (Schéma d'après Groenen *et al.*, 1993 et Kamerbeek *et al.*, 1997). 128
- 5.2 Exemple d'un locus contenant un nombre variable de régions répétées en tandem (VNTR). La figure montre l'ADN génomique au niveau du locus ETR-B de *M. tuberculosis* H37Rv. Le locus est amplifié par PCR grâce à des amorces complémentaires des régions d'ADN qui encadrent le locus (flèches). Le produit de PCR obtenu est de 292 paires de bases (pb) contenant 3 copies de régions répétées en tandem de 57 paires de bases, et 8 bases supplémentaires qui correspondent au début d'une autre région répétée en tandem. 129
- 5.3 Polymorphisme de taille de différentes souches de *M. tuberculosis* isolées en Guyane pour l'ETR-E. La photo montre le profil de migration des échantillons d'ADN sur gel après électrophorèse. Chaque colonne contient le produit de PCR d'une souche différente (ici, 12 souches) ; le dépôt n°12 n'a pas migré. M correspond aux marqueurs de poids moléculaire "100bp ladder" (100, 200, 300... à 1000 pb). La vitesse de migration des fragments dépend de leur poids moléculaire ; le fragment de 100 pb migre le plus vite, c'est donc la bande la plus éloignée du puit). Le polymorphisme observé correspond à l'addition ou à la délétion d'une unité répétée en tandem de 53 pb. Le poids moléculaire (en pb) est noté au-dessus des profils de migration ; juste en-dessous entre parenthèses, est indiqué le nombre de copies correspondantes. 130
- 5.4 Spécificité géographique des souches (après attribution de l'indice *CI*) et le nombre (1) de patients et (2) de spoligotypes concernés dans 136

TABLE DES FIGURES

	chaque classe.	
5.5	Histogrammes récapitulatifs. (A) Nombre de cas associés aux spoligotypes recensés en Guyane, dérivé du tableau 5.1 ; le spoligotype qui recense le plus grand nombre de cas, situé le plus à gauche, est ST 50. (B) Tailles des grappes, c'est-à-dire nombre de grappes incluant 2, 3, 4... patients, en fonction de la méthode (dérivées du tableau 5.3).	138
5.6	<i>Cluster-graphs</i> représentant six génotypes ( $g_1$ à $g_6$ ). Un nœud représente une grappe de cas ayant le même génotype, et la taille des grappes est indiquée à l'intérieur du cercle. (A) Représentation d'après [Tanaka et Francis, 2005]. (B) Représentation personnelle, plus informative. Les cercles en pointillés indiquent les sauts mutationnels : entre $g_1$ et $g_3$ il y a une mutation, alors qu'entre $g_1$ et $g_4$ il y en a deux.	145
5.7	Identification des liens phylogéniques entre spoligotypes ("Type"), définis par des événements de délétion de certains espaceurs (IDR). Type, identifiant numérique des spoligotypes. "Description binaire", présence / absence des 43 IDR définies par analyse moléculaire ; les carrés noirs et blancs signalent respectivement la présence ou l'absence d'IDR. Les chiffres près des flèches donnent le nombre de sauts mutationnels. Les quatre spoligotypes sont de la famille T, définie par l'absence des espaceurs 33 à 36. ST 53 est à l'origine de tous les spoligotypes de la famille T (ancêtre commun). Le lien phylogénique entre ST 53 et ST 86 est défini par la délétion de l'espaceur 22.	146
5.8	<i>Cluster-graph</i> de la famille Haarlem. Sur 17 spoligotypes (ST) représentés, 13 ont été recensés en Guyane (1996-2003), dont ST 46 de la famille « likely Haarlem ». La taille des cercles définit la taille des grappes. Quand il est $> 1$ cas, le nombre de cas est inscrit dans le cercle. Les désignations des spoligotypes sont inscrites près des nœuds.	152
5.9	<i>Cluster-graph</i> de la famille EAI. Sur 36 spoligotypes représentés, 24 ont été recensés en Guyane (1996-2003). ST 236 serait l'ancêtre commun des autres spoligotypes de cette famille mais il n'a pas été recensé en Guyane (du moins sur la période étudiée).	153
5.10	<i>Cluster-graph</i> de la famille X. Sur les 11 spoligotypes représentés, 8 ont été recensés en Guyane (1996-2003). L'ancêtre commun probable, ST 119, n'a pas été recensé.	153
5.11	<i>Cluster-graph</i> de la famille LAM. Sur 38 ST représentés, 33 ont été recensés en Guyane (1996-2003). ST 42 est l'ancêtre commun de cette famille.	154
5.12	<i>Cluster-graph</i> des familles T et S. Sur 43 ST de la famille T représentés, 38 ont été recensés en Guyane (1996-2003). Les 2 spoligotypes de la famille S ont été recensés.	154
5.13	<i>Cluster-graph</i> complet avec les 120 spoligotypes recensés en Guyane entre 1996 et 2003, auxquels s'ajoutent 29 spoligotypes absents de Guyane, qui sont schématisés par un petit point.	155

5.14	Arbre couvrant minimal des souches de <i>Mycobacterium tuberculosis</i> isolées en Guyane entre 1996 et 2003, construit sur la base de deux marqueurs génétiques indépendants : spoligotypes et VNTR. À chaque petit cercle correspond un génotype. Les paramètres choisis dans BioNumerics pour la construction de cet arbre incluent la création de génotypes hypothétiques, matérialisés par les petits cercles blancs.	157
5.15	Zoom sur les branches A à D de l'arbre phylogénétique de la figure 5.14. Chaque petit cadre décrit les données moléculaires correspondant au génotype. Le premier chiffre indique le spoligotype, le deuxième indique le codage VNTR-A à E.	158
5.16	Description binaire des spoligotypes 4, 2 et 1 (présence/absence des 43 espaceurs) et familles auxquelles ils appartiennent.	159
6.1	Taille des populations étudiées (en nombre de cas N), en fonction des analyses moléculaires (spoligotypage et VNTR) et du lieu de résidence des cas. Sur les 345 cas confirmés de départ, trois cultures contaminées ont été éliminées de l'étude. Les analyses ont par ailleurs mis en évidence deux cas contaminés successivement par deux souches différentes, ce qui donne un total 344 souches au lieu de 342. Les "incertains" correspondent aux cas dont le typage par VNTR était incomplet.	167
6.2	Pourcentage de cas en grappes dans chaque IRIS de l'île de Cayenne (A) en tenant compte uniquement des spoligotypes et (B) de la combinaison spoligotype + VNTR.	170
6.3	Illustration de la notion de risque et de ses composantes, traditionnellement utilisée en gestion des risques naturels. La flèche en pointillés indique que la survenue d'une catastrophe n'est pas systématique lorsqu'il y a risque, mais dépend de l'intervention d'un élément extérieur potentiel, qu'illustre le point d'interrogation. (Source des images : <a href="http://www.irma-grenoble.com/04risques/index.htm">http://www.irma-grenoble.com/04risques/index.htm</a> ).	173
6.4	Grille de décision pour l'identification des zones à risque épidémiologique pour la trypanosomose animale au Burkina Faso (d'après [Michel <i>et al.</i> , 2001]). À l'horizontale se trouve l'espace des bovins, défini par la « charge surfacique » en bétail ; à la verticale se trouve l'espace des glossines, lié à certaines classes de paysages de bas-fonds, auxquels se réfèrent les chiffres 8, 1 et 2, correspondant respectivement à un paysage de forêt, un paysage de galerie large bordée de savanes herbeuses, et un paysage de galerie étroite bordée de savanes arbustives ou arborées lâches. Ces trois paysages sont ceux pour lesquels les densités de glossines observées sont significatives.	174
6.5	Arbre de décision permettant de définir les trois classes d'aléa de tuberculose, en fonction de la présence de cas contagieux BAAR+, et de la présence de cas de tuberculose BAAR- mais potentiellement contagieux. L'aléa est fort si, dans le quartier considéré, il y a au moins un cas contagieux ; l'aléa est faible si il n'y a aucun cas de tuberculose	177

## TABLE DES FIGURES

---

- détecté comme contagieux (BAAR+) mais qu'il y a au moins un cas de tuberculose recensé dans le quartier ; l'aléa est nul si aucun cas de tuberculose n'a été recensé entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003 (probabilité de l'évènement "présence d'un cas de tuberculose" = 0).
- 6.6 Cartographie de l'aléa par quartier sur l'île de Cayenne (associé à la présence de cas de tuberculose contagieux, et à la présence de cas non contagieux dans le quartier). La classification est basée sur l'arbre de décision de la figure 6.5. appliqué aux cas de tuberculose recensés sur l'île de Cayenne entre 1996 et 2003 dont nous avons recensé l'adresse (366 cas spatialisés). 178
- 6.7 Tableau récapitulatif de la vulnérabilité face à la tuberculose, associée à un quartier en fonction (1) du type d'habitat (spontané, mixte, résidentiel dense ou diffus, dispersé), et (2) de l'insalubrité de l'habitat, définie par une précédente étude de la DDE de Guyane [DDE-AU&S, 1997]. Les cinq classes de vulnérabilité de la population sont : très faible, faible, moyenne, forte, très forte. 180
- 6.8 Cartographie de la vulnérabilité de la population de l'île de Cayenne face à la tuberculose, par quartier. La classification est basée sur la grille de décision de la figure 6.7 tenant compte des deux paramètres typologie d'habitat et insalubrité de l'habitat. La cartographie a été réalisée grâce au système de requêtes du SIG GéoConcept. 181
- 6.9 Grille de décision pour l'identification des zones à risque de tuberculose sur l'île de Cayenne. En lignes, sont notées les cinq classes de vulnérabilité de la population (de très forte à très faible) définissant l'espace des hôtes. En colonnes, sont notées les trois classes d'aléa, correspondant à l'occurrence ou non de cas tuberculeux potentiellement contagieux, définissant l'espace des pathogènes. Le croisement de l'aléa et de la vulnérabilité aboutit à une échelle de risque tuberculeux comprenant sept classes de risque (de fort à nul). 182
- 6.10 Cartographie du risque de tuberculose par quartier sur l'île de Cayenne (associé à la vulnérabilité de la population et à l'aléa), la classification étant basée sur la grille de décision présentée sur la figure 6.9. 183



# Liste des tableaux

2.1	Taux d'incidence de la tuberculose en France métropolitaine, en Ile de France et en Guyane, et déclarations obligatoires (DO) en Guyane, issues de la DSDS, 1996-2003.	27
2.2	Nombre de cas de tuberculose déclarés en Guyane sur la période 1996-2003 en fonction du lieu de résidence des patients (Surinam et quelques villes de Guyane).	27
2.3	Nombre de cas de tuberculose déclarés sur la période 1996-2003, en fonction de la ville ou du pays de résidence des patients, et après corrections.	28
2.4	Nombre de cas de tuberculose recensés sur l'Ile de Cayenne entre 1996 et 2003, auprès de la DSDS, de l'IPG, du DIM, et de la LAT, et dans ces 3 dernières sources à la fois après élimination des cas communs à plusieurs sources (doublons et triplets).	33
2.5	Cas de tuberculose confirmés bactériologiquement ou non confirmés ayant déclaré la maladie entre le 1 <sup>er</sup> janvier 1996 et le 31 décembre 2003 et résidants en Guyane.	36
2.6	Ratios standardisés d'incidence (RIS) de tuberculose calculés pour classes d'âge de 1996 à 2003 de la population de l'Ile de Cayenne.	40
2.7	Estimations du nombre de cas de tuberculose totaux basées sur les estimateurs de Petersen et Chapman, les échantillons étant pris deux à deux.	58
2.8	Estimations des cas de tuberculose totaux dans les modèles log-linéaires à trois sources. De haut en bas : modèle d'indépendance ; modèles à 1 ou 2 interactions avec indépendance relative (modèles 13/2 à 23/13) ; modèles tenant compte de l'hétérogénéité entre individus (quasi-symétrique et partiels quasi-symétriques) ; modèle saturé.	58
2.9	Résultats obtenus avec l'approche <i>sample coverage</i> . Les trois estimateurs calculés tiennent compte de l'hétérogénéité de redondance de l'information entre les trois échantillons (ou sources) considéré(e)s.	60
2.10	Estimations annuelles de l'incidence de la tuberculose sur l'Ile de Cayenne durant la période 1996-2003 (incidence en nombre de cas / an / 100 000 habitants).	63

---

2.11	Données officielles de l'InVS, estimations brutes et estimations corrigées par capture-recapture avec la méthode <i>sample coverage</i> du nombre de cas totaux de tuberculose présents en Guyane entre 1996 et 2003, et incidences calculées sur la base des données corrigées par capture-recapture avec l'approche <i>sample coverage</i> (incidences exprimées en nombre de cas par an pour 100 000 habitants).	64
4.1	Ratios standardisés d'incidence ( <i>RIS</i> ) de tuberculose calculés par IRIS, en fonction (a) de la répartition par âge et par sexe, et (b) de la répartition par sexe et par nationalité de la population de l'île de Cayenne. Seuls sont détaillés les résultats des IRIS où l'on note une sur-incidence significative de tuberculose ( <i>RIS</i> significativement supérieur à 1).	100
4.2	Unique agrégat significatif de tuberculose, identifié par l'analyse SaTScan. L'IRIS « Eau Lisette », sur lequel est centré l'agrégat, lui donne son nom.	113
4.3	Détail des IRIS inclus dans l'unique agrégat significatif, dit « Eau Lisette », avec le nombre de cas de tuberculose attendu ( <i>loc_att</i> ), le nombre de cas réellement observé ( <i>loc_obs</i> ) et le rapport entre nombre de cas observé sur nombre de cas attendu ( <i>loc_o/a</i> ). La probabilité associée à l'agrégat (et donc aux 12 IRIS qu'il comprend) est $p = 0,0001$ .	113
4.4	Agrégats de tuberculose identifiés par l'analyse SaTScan. Seul l'agrégat nommé « Palmistes », comprenant 9 IRIS, est significatif ( $p = 0,0001$ ).	115
4.5	Détail des IRIS inclus dans le seul agrégat de tuberculose significatif, dit « Palmistes ». Le nombre de cas de tuberculose attendu dans l'IRIS ( <i>loc_att</i> ), le nombre de cas réellement observé ( <i>loc_obs</i> ) et le rapport entre cas observés sur cas attendus ( <i>loc_o/a</i> ) sont précisés. La probabilité associée à l'agrégat (et aux 9 IRIS qu'il comprend) est $p = 0,0001$ .	115
4.6	Agrégats de tuberculose identifiés par l'analyse SaTScan. Sur ces huit agrégats, seuls les agrégats notés 1, 2 et 3 sont significatifs ( $p < 0,05$ ).	117
4.7	Zones d'habitat insalubre sur l'île de Cayenne identifiées par la DDE de Guyane, et typologie d'habitat associée [DDE-AU&S, 1997].	120
5.1	Tableau récapitulatif des spoligotypes isolés en Guyane sur la période 1996-2003, et des spécificités géographiques et abondances associées. Les 25 spoligotypes orphelins se sont pas intégrés dans ce tableau puisqu'ils sont uniques en Guyane et dans le monde (en l'état actuel des connaissances, c'est-à-dire au 1 <sup>er</sup> janvier 2005).	134
5.2	Grappes obtenues d'une part par la méthode de spoligotypage seule, et d'autre part en couplant le spoligotypage et le typage par VNTR (nombre d'isolats et fréquences).	137

## LISTE DES TABLEAUX

---

5.3	Taille des grappes (en nombre de cas), et nombre de grappes correspondantes.	137
5.4	Nombre de cas par VNTR pour 5 spoligotypes regroupant plus de 15 cas.	137
5.5	Déductions concernant le taux de transmission de tuberculose en fonction de la diversité des génotypes observés (d'après [Tanaka et Francis, 2005]).	143
5.6	Spoligotypes recensés en Guyane entre 1996 et 2003, et leurs familles.	150
5.7	Nombre de cas et de spoligotypes (ST) de Guyane (1996-2003) par famille.	152

# Introduction

Jusqu'à présent et depuis l'avènement, au début du XIX<sup>e</sup> siècle de l'approche pastorienne des maladies (à un germe correspond une maladie), la médecine et les disciplines connexes se sont essentiellement préoccupées de comprendre les pathologies humaines dans un contexte socio-économique où les agents étiologiques pouvaient être éradiqués grâce aux progrès de la médecine. Cependant, l'évolution actuelle des problèmes de santé dans le monde montre certaines faiblesses de la médecine moderne face aux risques d'émergence de maladies nouvelles ou de ré-émergence de maladies que l'on pensait avoir éradiquées, et face à l'apparition de résistances thérapeutiques et d'allergies. Et si les pays en développement restent les plus touchés en termes de morbidité et de mortalité, les pays industrialisés sont de plus en plus affectés par ces nouveaux risques. En France, les départements français d'Outre-Mer, et spécialement la Guyane française, sont particulièrement à risque. Cette évolution incite aujourd'hui certains scientifiques à parler d'une quatrième transition épidémiologique, celle de la « ré-émergence » [McMichael, 2001].

Au XX<sup>e</sup> siècle, la pensée scientifique s'est progressivement imprégnée de la notion de complexité, et de la nécessité d'une approche systémique [Froment, 1997], y compris dans le domaine médical. Au sein du système d'interactions hôte-pathogène s'est donc ajouté, au fil du temps, la prise en compte d'un troisième élément, l'environnement. Replacer le système pathogène dans son environnement suppose l'intégration, dans la démarche analytique, des différents acteurs pouvant influencer le fonctionnement de ce système. Depuis quelques dizaines d'années, tend donc à se développer une approche pluridisciplinaire visant à prendre en compte les causes « externes » dans le développement de pathologies, telles que les facteurs physiques, chimiques climatiques, écologiques, mais aussi les comportements personnels ou culturels pouvant favoriser le maintien, l'extension, l'émergence ou la ré-émergence de maladies infectieuses et parasitaires.

Dans ce contexte, si l'on veut aujourd'hui comprendre les problèmes complexes liés à l'émergence, la dispersion, et la distribution spatiale des agents étiologiques responsables des maladies humaines, il faut replacer les agents pathogènes dans leur contexte écologique, biologique et social, c'est-à-dire considérer de façon intégrée les systèmes « éco-sociopathogènes » [Amat

Rose et Gentillini, 1995], pour pouvoir décrire, analyser, modéliser, et prédire la dynamique spatio-temporelle de maladies. C'est autour de ce thème que se sont développées les thématiques de recherche sur les relations entre environnement et santé (animale ou humaine) [Avruskin *et al.*, 2004 ; Cadot *et al.*, 1998 ; Craig *et al.*, 2000 ; de La Rocque *et al.*, 2004 ; de Savigny et Wijeyaratne, 1995 ; Dossou-Yovo *et al.*, 1998 ; Epstein, 1999 ; Hay, 2000 ; Linthicum *et al.*, 1999 ; Reeves *et al.*, 1994 ; Washino et Wood, 1994].

Un certain nombre de facteurs ou de déterminants influant sur la santé ont des caractéristiques spatiales et temporelles distinctes. C'est par exemple le cas du climat, d'où l'importance d'étudier les effets du changement climatique à l'échelle globale. Si l'on en vient aujourd'hui à étudier la spatialisation des données épidémiologiques, c'est tout d'abord parce que l'on se rend compte de l'importance de l'espace dans la structuration des phénomènes de contagion-diffusion de maladies infectieuses ou parasitaires, mais également parce qu'il existe maintenant de nouveaux outils permettant d'intégrer cette information. Depuis plusieurs années, les outils de télédétection et les Systèmes d'Information Géographique (SIG) sont utilisés pour de telles recherches. Ces outils permettent d'intégrer une composante spatiale à l'observation des dynamiques de maladies et d'identifier certains paramètres environnementaux corrélés à la présence de la maladie, ces paramètres étant extraits des images. Cependant, il existe à l'heure actuelle encore peu de recherches concernant les pathogènes et les maladies pour lesquels la relation à l'environnement est moins directe et intuitive que pour les maladies transmises par un insecte vecteur. C'est par exemple le cas de la tuberculose, où l'analyse spatiale doit intégrer les disparités socio-spatiales. Parallèlement, les progrès réalisés en génétique humaine, qui découlent en grande partie du Projet du génome humain, sont en train de bouleverser les pratiques cliniques et médicales, en particulier dans les domaines de la prévention de la maladie et de la promotion de la santé. L'évolution rapide des méthodes moléculaires et de la génétique donne aujourd'hui accès à de nouveaux outils permettant de suivre les dynamiques spatio-temporelles des agents infectieux à l'intérieur même de la population hôte, et ce, à l'échelle du gène.

Le présent travail de recherche se propose de comprendre la dynamique spatiale de propagation et de persistance de la tuberculose en Guyane, et ce, par l'identification des facteurs de risque liés à l'apparition ou à la persistance de cas de la maladie. La compréhension de tels processus passe par une approche à la fois épidémiologique, écologique, statistique, génétique et géographique de la maladie. Cette étude pluridisciplinaire prend en compte des niveaux complémentaires et hiérarchisés.

La Guyane, seule région de France et d'Europe basée en Amérique du Sud, connaît une situation ambivalente qui a un certain nombre d'impacts sur sa situation socio-économique et sanitaire. Bénéficiant des avantages sociaux et techniques de la France métropolitaine, la Guyane connaît pourtant des difficultés sanitaires et sociales marquées, avec entre autres une fréquence encore importante de décès par maladies infectieuses et parasitaires, un taux élevé de Sida, une forte surmortalité masculine, et l'analphabétisme d'une partie de la population. De nombreuses équipes de recherche en santé se tournent aujourd'hui vers cette région en pleine transition épidémiologique, c'est-à-dire que la Guyane évolue d'une situation où les grandes pandémies sont la principale cause de mortalité, en particulier le paludisme dans les régions de l'intérieur, vers une situation où les maladies dégénératives et les sociopathies prennent le pas sur les autres causes de mortalité, ce qui concerne plutôt les communes du littoral.

Cette recherche a été développée pour le cas de la tuberculose en Guyane, avec cependant la volonté affirmée de proposer une méthode transposable à d'autres pays pour lesquels la situation épidémiologique et géographique est similaire. En nous appuyant d'une part sur l'analyse des limites actuelles de la surveillance épidémiologique de la tuberculose en Guyane, et d'autre part sur la détermination des besoins actuels au vu de la situation internationale, cette recherche vise à répondre aux principales questions scientifiques suivantes.

1. Comment évaluer le poids réel de la maladie en terme de santé publique, dans les pays où l'incidence de la maladie est sous-évaluée ?
2. Quelles sont les dynamiques spatiales de transmission de la tuberculose pouvant expliquer la propagation et/ou la persistance de la maladie ?
3. Quelle est l'influence de l'environnement social et culturel sur l'évolution de maladies telles que la tuberculose ?
4. Quels sont les besoins et les actions à mettre en œuvre pour l'aide à la décision dans la surveillance et le contrôle de la tuberculose ?

C'est sur la base de ces grandes questions qu'a été construit le programme de recherche doctorale énoncé ci-après. Ce travail s'insère dans le cadre conceptuel d'une prise de conscience de l'association étroite entre la santé des sociétés humaines et leur environnement au sens large, devenue paradigme. Elle s'inscrit également dans la dynamique actuelle d'un développement actif d'activités de recherche transversales et intégratives en matière de santé et d'environnement.

Les différentes étapes de la recherche engagée concernant la dynamique de transmission de la tuberculose en Guyane sont expliquées ci-dessous au travers d'un résumé des six chapitres que compte ce mémoire.

Le premier chapitre a pour objectif de donner une vue d'ensemble du contexte général de l'étude. Dans un premier temps, les caractéristiques intrinsèques de la tuberculose humaine seront détaillées, ainsi que les contraintes de diagnostic, de suivi, et de traitement qui y sont liées, et qui contribuent à freiner l'éradication de la maladie dans le monde, en particulier dans les pays en développement. Dans un deuxième temps, seront développées certaines caractéristiques géographiques, sociales, et démographiques de la Guyane française, qui font de cette région un terrain favorable à la diffusion et/ou à la persistance de la tuberculose, qualifiée de « maladie de la pauvreté ». Cette analyse préliminaire permettra d'intégrer aux analyses ultérieures certaines spécificités de la maladie et de son contexte géographique qui peuvent avoir une influence sur la dynamique de la tuberculose.

Le deuxième chapitre sera consacré à l'analyse critique des données chiffrées de la tuberculose en Guyane. Il recense l'ensemble des données chiffrées existantes sur la tuberculose dans cette région, que ce soit les données officielles publiées par les organismes de santé publique en charge de la surveillance de la maladie ou les données obtenues après enquête auprès des acteurs de la lutte anti-tuberculeuse en Guyane sur la période 1996-2003 :

- l'Institut Pasteur de Guyane (IPG),
- le Département d'Information Médicale (DIM) de l'hôpital de Cayenne, et
- le service de Lutte Anti-Tuberculeuse (LAT) géré par le conseil général.

Les données chiffrées seront analysées afin de mettre en avant une sous-évaluation de l'importance de la tuberculose en Guyane, et la nécessité d'évaluer l'incidence théorique espérée de cette maladie. Un certain nombre d'analyses descriptives concernant l'épidémiologie des cas seront réalisées dans le but d'identifier les facteurs de risques humains tels que l'âge, le sexe, etc. spécifiques à la Guyane. Puis, la méthode dite de « capture-recapture » sera utilisée pour estimer la proportion de cas non recensés par les systèmes de prise en charge de la tuberculose sur l'île de Cayenne, en Guyane, ces cas contribuant à maintenir un pool d'agents pathogènes au sein de la population hôte. Ces résultats permettront d'évaluer la nécessité d'améliorer ou non la détection des cas, et/ou le bon suivi du traitement.

Le troisième chapitre a pour but d'expliquer le cheminement scientifique qui a orienté nos différents choix d'étude, et de faciliter la lecture du reste du mémoire. Cette partie posera donc les bases des différentes disciplines et approches méthodologiques abordées au cours du travail de recherche. En prenant appui sur le chapitre précédent, les questions à résoudre pour améliorer les stratégies de lutte contre la tuberculose seront présentées. La problématique générale de l'étude sera posée, avec comme question centrale : « Quels sont les processus en cause dans la propagation et le maintien de la tuberculose en Guyane ? » La réponse à cette question s'appuiera sur deux

volets fondamentaux : (1) l'analyse spatiale appliquée aux maladies, avec l'utilisation des outils de la spatialisation et de la géomatique, et d'un Système d'Information Géographique (SIG), et (2) l'apport de l'épidémiologie moléculaire à cette étude. Un bref historique et état de l'art de chacun de ces deux volets sera décrit, avant de développer les aspects spécifiques applicables au cas de la tuberculose. Enfin, nous présenterons l'organisation de la suite du travail de recherche qui consistera, dans un premier temps, à déterminer les facteurs intrinsèques (génétiques) et extrinsèques (sociaux et environnementaux) pouvant expliquer la persistance de la tuberculose en Guyane, puis, dans un second temps, à identifier les populations à risque de tuberculose, et ce par la combinaison d'un SIG et de l'analyse moléculaire des souches de tuberculose recensées en Guyane entre 1996 et 2003.

Dans le chapitre quatre, l'analyse statistique des relations entre occurrence de cas de tuberculose et environnement spatial sera présentée. La distribution de l'incidence de cas par zone géographique sera examinée afin de vérifier l'hypothèse nulle ( $H_0$ ) selon laquelle la répartition des cas de tuberculose sur l'île de Cayenne est aléatoire. Différentes covariables telles que l'âge, le sexe et la nationalité, seront par la suite intégrées à l'analyse afin de vérifier si la distribution de la tuberculose peut être corrélée à la distribution spatiale de la population générale en correspondance avec certaines de ces covariables. Différentes échelles spatiales d'étude seront par ailleurs testées et discutées. Un outil de statistique spatiale permettra ensuite d'apporter une validation statistique à cette première approche, au travers de la caractérisation des agrégats spatiaux de cas de tuberculose, c'est-à-dire les zones où se concentrent principalement les cas. Cette agrégation peut être liée à deux hypothèses que nous allons chercher à valider ou infirmer : (1) une transmission localisée de la maladie dans des foyers épidémiques ou (2) une agrégation de cas corrélée à l'existence ponctuelle de facteurs de risques, se traduisant par exemple par une ségrégation spatiale des individus vulnérables.

Le chapitre cinq se préoccupe essentiellement de caractériser les patients porteurs de la même souche bactérienne, ces individus étant potentiellement issus d'une même source de contamination. La comparaison des souches de la bactérie identifiées en Guyane avec une base de données internationale permettra de dégager les spécificités géographiques de certaines souches, en particulier les souches recensées uniquement en Guyane, et de proposer des hypothèses concernant la dynamique de transmission de la maladie, basée sur l'importation de cas, la réactivation de tuberculoses anciennes, et/ou la transmission active récente de la maladie. Afin de comprendre les liens évolutifs entre souches, un arbre phylogénétique sera proposé sur la base de l'analyse moléculaire des souches. Enfin, une méthode mathématique récemment développée permettra, sur la base de cette phylogénie, d'évaluer

quantitativement le taux de transmission récente de la tuberculose en Guyane, calcul qui permettra de juger de la sévérité de l'endémie / épidémie.

Le dernier chapitre conclura ce mémoire en récapitulant les informations recensées sur la dynamique de la tuberculose en Guyane, et en combinant les informations spatiales et génétiques. Plus précisément, les agrégats spatiaux de tuberculose sur l'île de Cayenne seront mis en parallèle avec les grappes génétiques incluant les cas porteurs d'une même souche, afin de vérifier s'il existe bien une contamination de proximité (dans la même rue ou le même quartier) ou si la transmission de la maladie se fait à une échelle plus importante. Sur la base du résultat obtenu, des cartographies de l'aléa et de la vulnérabilité de la population seront construites, le croisement de ces deux couches d'information aboutissant finalement à une cartographie du risque. Le but de cette cartographie est d'apporter une aide à la décision pour la lutte contre la tuberculose en Guyane française

# Chapitre 1

## Contexte général : de la tuberculose en général et de la Guyane en particulier

*Ce premier chapitre a pour but de présenter le contexte général de l'étude, défini à la fois par la maladie étudiée et le cadre géographique dans lequel elle se place. Dans un premier temps, nous présentons quelques caractéristiques de la tuberculose humaine, en particulier celles pouvant expliquer les difficultés de lutte contre cette maladie, tant en termes de suivi épidémiologique que de traitement. Puis nous détaillons la région d'étude, la Guyane française, avec sa situation géographique et quelques données socio-économiques, culturelles et sanitaires, le but étant d'évaluer dans quelle mesure ce contexte peut avoir une incidence sur la dynamique de la tuberculose dans cette région.*

### 1. Généralités sur la tuberculose humaine

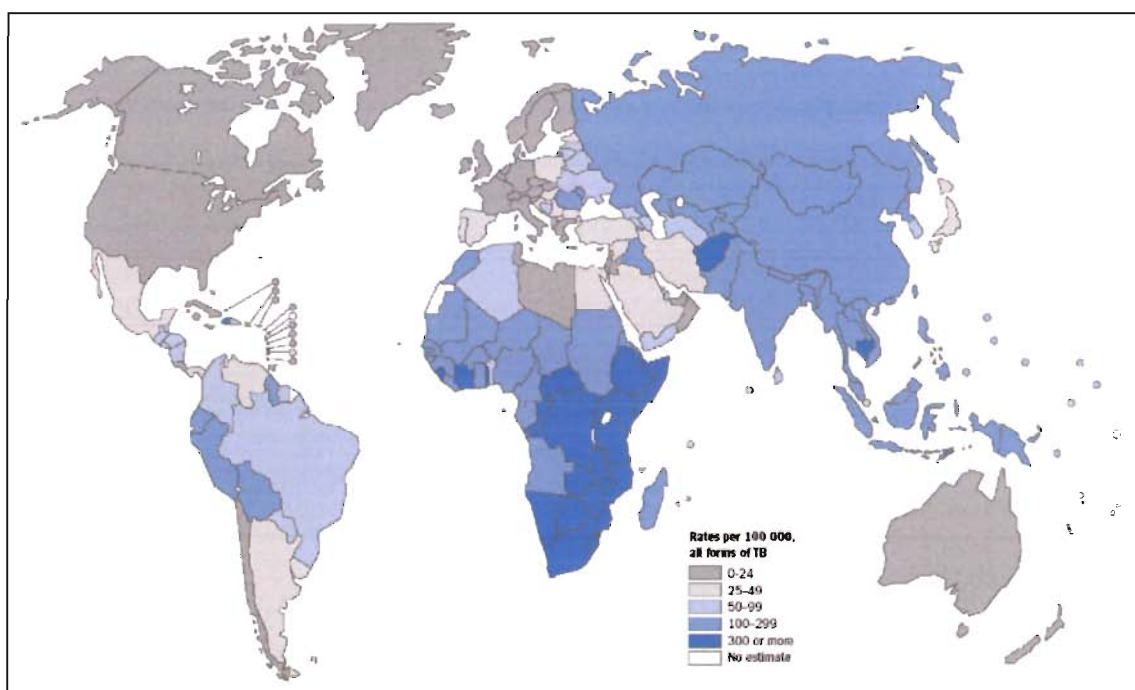
L'objet de cette partie n'est pas de faire une analyse descriptive de la tuberculose humaine, mais de mettre en évidence quelles sont les contraintes particulières à la maladie pouvant freiner la lutte contre la tuberculose, qui fait partie des cinq maladies infectieuses les plus meurtrières au monde. Pour cela, nous avons dû détailler certaines caractéristiques épidémiologiques, cliniques, et diagnostiques de la maladie (§1.2 à 1.6), pour finalement aboutir à leurs implications en terme de santé publique (§1.7).

#### 1.1. La tuberculose dans le monde

Après un bref récapitulatif des chiffres officiels concernant les taux de morbidité et de mortalité dus à la tuberculose, nous détaillons l'évolution du nombre de nouveaux cas infectés par la maladie par rapport à la population totale, dans les pays industrialisés, où la maladie est en ré-émergence, et dans les pays en développement, où elle est en forte recrudescence.

### 1.1.1. Les estimations officielles

L'Organisation Mondiale de la Santé (OMS) estime à 8,8 millions le nombre de nouveaux cas de tuberculose en 2002, et à 2 millions le nombre de cas qui meurent de la maladie chaque année [Dye *et al.*, 1999]. Au total, un tiers de la population mondiale est infecté, et 95% des cas sont localisés dans le Tiers-Monde (figure 1.1).



### 1.1.2. Évolution de l'incidence

L'un des moyens les plus courants d'évaluer l'importance quantitative d'une maladie dans un pays ou une région est de calculer l'incidence de cette maladie, c'est-à-dire le nombre de nouveaux cas infectés par rapport à la population totale sur une période de temps donnée [Hellenbrand *et al.*, 2001].

#### *Dans les pays industrialisés*

Après une forte diminution de l'incidence de la tuberculose vers 1960-70, on a assisté à la fin des années 1980 / début des années 1990 à une résurgence de la tuberculose dans de nombreux pays industrialisés [Hopewell, 2002]. Les facteurs responsables de cette évolution sont nombreux : relâchement de la lutte anti-tuberculeuse ; immigration de pays à forte incidence de tuberculose ; paupérisation de certains quartiers urbains ; transmission de la maladie dans

des lieux de vie collectifs tels que les hôpitaux et les prisons ; croissance mondiale de l'épidémie de VIH<sup>1</sup> qui, en cas de co-infection, augmente le risque d'activation des tuberculoses latentes; émergence de souches de tuberculose multirésistantes aux antibiotiques [CDC, 1999].

En Guyane, région qui nous intéresse, certains de ces facteurs jouent certainement un rôle prépondérant. Des données récentes suggèrent par ailleurs que le facteur « émergence de souches multirésistantes » est en passe de devenir la principale cause de morbidité et de mortalité aux États-Unis, en Europe et en Amérique Latine (références par pays dans [Farmer et Kim, 1998]). Dans la majorité des cas, les mesures prises pour réorganiser la lutte ont conduit à une baisse de morbidité. Mais dans certains pays, en particulier en Europe de l'Est et dans les pays de l'ex-URSS, le nombre de cas et la morbidité continuent d'augmenter [Raviglione *et al.*, 1994].

### *Dans les pays en développement*

Dans ces pays, les taux de notification de la tuberculose par l'OMS n'ont cessé d'augmenter entre 1990 et 1999 [Raviglione *et al.*, 1995]. En 1997 par exemple, 22 pays recensent à eux seuls 80% des cas incidents, avec plus de 50% des cas situés dans 5 pays d'Asie du Sud-Est. D'autre part, 9 des 10 pays les plus touchés en terme d'incidence sont des pays africains [Dye *et al.*, 1999] (figure 1.1).

Au final, les régions les plus touchées sont l'Asie du Sud-Est, l'Afrique et l'Europe de l'Est (80% des cas dans le monde). Dans ces pays, l'importance de l'endémie tuberculeuse n'est pas tant due aux problèmes de multirésistance des souches qu'à un manque d'accès aux soins, ainsi qu'aux taux très importants de co-infection tuberculose/VIH observés dans certains pays africains [Farmer et Kim, 1998 ; Dye *et al.*, 1999].

## **1.2. Données médicales sur la tuberculose humaine**

Préalablement à la description d'un travail de recherche qui concerne la tuberculose en Guyane, il est nécessaire de préciser les particularités cliniques de la maladie qui vont avoir de l'importance pour la suite de l'étude, qui inclut entre autres le suivi de la contamination. Les informations concernant les formes cliniques de la tuberculose, son mode de transmission et sa contagiosité sont d'abord détaillées (§1.2.1.).

Puis nous développons les aspects diagnostiques (1.2.2.) et thérapeutiques (1.2.3.) de la maladie. Ces caractéristiques ont de fortes implications en terme de lutte contre la maladie, en vue de son éradication.

---

<sup>1</sup> Virus d'Immunodéficience Humaine

## 1.2.1. Les données cliniques

### 1.2.1.1. L'agent responsable de la maladie

La tuberculose est une maladie infectieuse, contagieuse et inoculable. L'agent responsable de la maladie est le bacille de Koch, bactérie appartenant au complexe *Mycobacterium tuberculosis* constitué de trois sous-espèces : *M. tuberculosis*, *M. bovis* et *M. africanum*. Ces deux dernières sous-espèces sont beaucoup plus rarement responsables de la tuberculose humaine [Dautzenberg et Bouvet, 1996]. Contrairement à l'hypothèse couramment véhiculée selon laquelle *M. tuberculosis* aurait évolué à partir de souches animales telles que *M. bovis*, une étude récente suggère que *M. tuberculosis* existerait depuis 15 à 20 000 ans et serait donc antérieur à l'apparition de *M. bovis*. La souche bactérienne ancestrale du bacille tuberculeux ressemblerait plutôt à *M. tuberculosis* ou *M. canettii*, et pourrait bien avoir été dès l'origine une bactérie pathogène de l'homme [Brosch *et al.*, 2002].

### 1.2.1.2. Les différentes formes cliniques

L'infection par la tuberculose peut prendre différentes formes cliniques : soit elle envahit rapidement l'ensemble de l'organisme (on parle alors de granulie), soit elle reste localisée dans un tissu ou un organe où elle se développe. La localisation pulmonaire est la plus commune (80% des cas) [Dautzenberg et Bouvet, 1996].

### 1.2.1.3. La transmission de la maladie

#### *Le mode de contamination*

La contamination est essentiellement inter-humaine, et se fait, dans le cas d'une tuberculose pulmonaire, par voie aérienne, par l'intermédiaire des gouttelettes de salive qui contiennent le bacille pathogène et qui sont propulsées lorsque le malade éternue ou tousse, par exemple [Dautzenberg et Bouvet, 1996]. La contamination peut également se faire, dans une moindre mesure, soit par l'ingestion de lait contaminé dans les régions où l'on recense des cas de tuberculose bovine, soit lors d'un contact cutané avec des lésions infectées.

#### *La contagiosité*

La forme pulmonaire de la maladie est la plus contagieuse [Dautzenberg et Bouvet, 1996]. En effet, cette forme clinique offre à la bactérie, par le biais des voies respiratoires, une voie de sortie vers l'environnement extérieur, et donc une possibilité de contact avec d'autres hôtes.

Le risque de contagion augmente avec le taux de concentration des bacilles à l'expectoration, ainsi qu'avec la proximité et la durée d'exposition au patient-source [Rieder, 1999]. Le bacille se caractérise par son caractère acido-alcoolo résistant, c'est-à-dire que sa paroi résiste aux acides et aux alcools. Cette propriété est à l'origine de deux colorations permettant la mise en évidence du bacille. La présence de bacilles acido-alcoolo-résistants (BAAR) à l'examen microscopique direct (ED) des produits d'expectoration du patient est un marqueur de la contagiosité [CSHPF, 2003].

#### **1.2.1.4. Les phases évolutives de la maladie**

Lors du premier contact avec le bacille, l'infection tuberculeuse connaît deux phases évolutives, la primo-infection tuberculeuse et la tuberculose maladie, avec parfois une phase intermédiaire de latence.

##### *La primo-infection tuberculeuse (PIT)*

C'est la phase d'infection initiale pendant laquelle le bacille de Koch va disséminer par l'intermédiaire de la voie sanguine et s'installer dans l'organisme [CSHPF, 2003]. Cette phase de bacillémie est le plus souvent asymptomatique.

##### *La tuberculose-infection latente*

Cette phase quiescente, dite aussi tuberculose-infection latente, peut à tout moment évoluer en une phase de maladie, symptomatique. La phase de latence précédant cette évolution peut durer de quelques semaines à plusieurs années [CSHPF, 2003]. Cette phase de latence a de nombreuses implications en épidémiologie. Elle a, en particulier, pour conséquence d'induire une incertitude sur la date exacte du contact patient / bactérie, qui peut précéder de plusieurs années la déclaration de la maladie. Dans une étude comme celle qui relève de notre mémoire, et qui concerne la dynamique de transmission de la tuberculose, ceci implique l'impossibilité d'un suivi temporel non biaisé des cas sur la base de la date de déclaration de la maladie. Cette particularité explique certains de nos choix ultérieurs.

##### *La tuberculose maladie*

L'évolution vers cette phase, qui s'exprime par des symptômes cliniques et/ou radiologiques, peut avoir lieu par exemple à l'occasion d'un déficit de défenses immunitaires de l'individu [CSHPF, 2003]. Dans environ 90% des cas, la primo-infection guérit spontanément et définitivement. Dans environ 10% des cas, les patients développent une tuberculose maladie, dont 5% dans l'année qui suit leur PIT, et 5 % au-delà de ce délai [Dautzenberg et Bouvet, 1996].

## 1.2.2. Le diagnostic

Le diagnostic de tuberculose pulmonaire se fait en plusieurs étapes. Devant toute suspicion de tuberculose, une intradermoréaction (IDR) à la tuberculine est pratiquée, ainsi qu'une radiographie pulmonaire. En cas de résultat anormal, on réalise un examen direct (ED) du produit d'expectoration au microscope et une recherche bactériologique approfondie. En dernier lieu, lorsque la présence de bactéries est confirmée, on réalise un antibiogramme pour analyser la sensibilité des souches bactériennes aux antibiotiques.

### 1.2.2.1. L'intradermoréaction à la tuberculine

#### *Méthode*

Les tests cutanés à la tuberculine, ou intradermoréaction, sont utilisés pour le diagnostic de primo-infection tuberculeuse. Ils consistent en une injection intra-dermique de 0,1 ml de solution liquide de tuberculine, mélange standardisé de fragments de mycobactéries. Un "virage" positif se traduit par une réaction cutanée tuberculinique (RCT) caractérisée par un érythème ainsi qu'une induration mettant en évidence une réaction d'hypersensibilité retardée. Cette réaction est induite par la présence d'antigènes de *M. tuberculosis*, ou de certaines mycobactéries atypiques, ainsi que par les antigènes du BCG<sup>2</sup>, raison pour laquelle l'IDR doit être interprétée en fonction des antécédents de vaccination [CSHPF, 2003].

#### *Interprétation de l'intradermoréaction*

L'interprétation du test repose sur la taille de l'induration. En règle générale, le critère de positivité correspond à un diamètre d'induration  $\geq$  à 5 mm, et une IDR dont le diamètre d'induration est  $\geq$  à 10 mm est en faveur d'une infection tuberculeuse suite à un contact infectant direct avec *M. tuberculosis* [CSHPF, 2003]. Pour les diamètres faiblement positifs, entre 5 et 10 mm, il est préférable de répéter le test 2 mois plus tard. L'interprétation doit tenir compte (1) de la prévalence de la tuberculose dans la zone considérée (zone de faible ou forte endémie), (2) des antécédents de vaccination (chez les sujets vaccinés, le résultat est à interpréter en fonction de l'ancienneté du BCG, l'immunité s'atténuant en 10 ou 15 ans), et (3) du statut VIH (chez les sujets infectés par le VIH, il est possible d'envisager la possibilité d'une tuberculose infection dès 5 mm) [CSHPF, 2003]. En cas de doute, on engage généralement un traitement prophylactique préventif (ou chimioprophylaxie) (figure 1.2).

---

<sup>2</sup> Le bacille de Calmette-Guérin (BCG) est une souche vivante atténuée de *Mycobacterium bovis*, utilisée pour la vaccination de la tuberculose depuis les années 1940 sur le plan mondial. Le vaccin par le BCG a un caractère obligatoire en France depuis 1952.

### **1.2.2.2. La radiographie pulmonaire standard**

Elle permet de faire un bilan des lésions thoraciques. Mais les anomalies radiologiques constatées peuvent être des séquelles d'une tuberculose ancienne.

### **1.2.2.3. La recherche bactériologique**

En première intention, on fait une recherche de bacilles dans un crachat, ou après tubage gastrique si le malade ne crache pas. Si cet examen ne révèle aucun BAAR, on réalise en deuxième intention une fibroscopie bronchique ou une mise en culture des prélèvements (figure 1.2). Le bacille de Koch est une bactérie à croissance lente, et il faut compter en moyenne quatre semaines de culture. L'identification des bactéries obtenues en culture se fait ensuite grâce à des épreuves biochimiques permettant de différencier les bactéries du complexe *Mycobacterium tuberculosis* des mycobactéries atypiques (*M. avium*, *M. gordonae*, *M. marinum*, *M. kansasii*...). Dans ce mémoire, il sera systématiquement précisé à quel type de cas nous nous référons (BAAR+, culture (+), etc.) Ces différentes analyses diagnostiques (figure 1.2) révèlent des stades de la maladie qui n'ont pas les mêmes implications en santé publique.

### **1.2.2.4. Typage et antibiogramme**

Le typage et l'antibiogramme ne font pas, à proprement parler, partie des étapes de diagnostic de la maladie, mais ils apportent des informations complémentaires sur les souches bactériennes isolées chez les patients tuberculeux en fournissant des données génétiques pour ce qui est du typage, ou des données de résistance pour ce qui est de l'antibiogramme.

#### *L'antibiogramme*

Il consiste à mesurer la sensibilité des bactéries aux antibiotiques par une nouvelle mise en culture en présence d'antibiotiques. Le résultat est obtenu en 4 à 6 semaines, ce qui explique le délai de 2 mois préconisé dans le traitement en quadrithérapie.

#### *Le typage moléculaire*

La reproduction du bacille tuberculeux est de type clonal, ce qui correspond à une reproduction asexuée, monoparentale, qui produit des individus strictement identiques sur le plan génétique entre eux et par rapport à l'individu dont ils sont issus [Sola *et al.*, 2000]. Le typage moléculaire, c'est-à-dire l'analyse génétique des souches bactériennes isolées chez les patients, permet de mettre en évidence les patients porteurs de la même souche clonale. Cette information est importante puisque la mise en évidence de nombreux

clones identiques (donc potentiellement issus d'une même bactérie "mère") chez des patients, peut être le reflet d'une épidémie [Hunter, 1990].

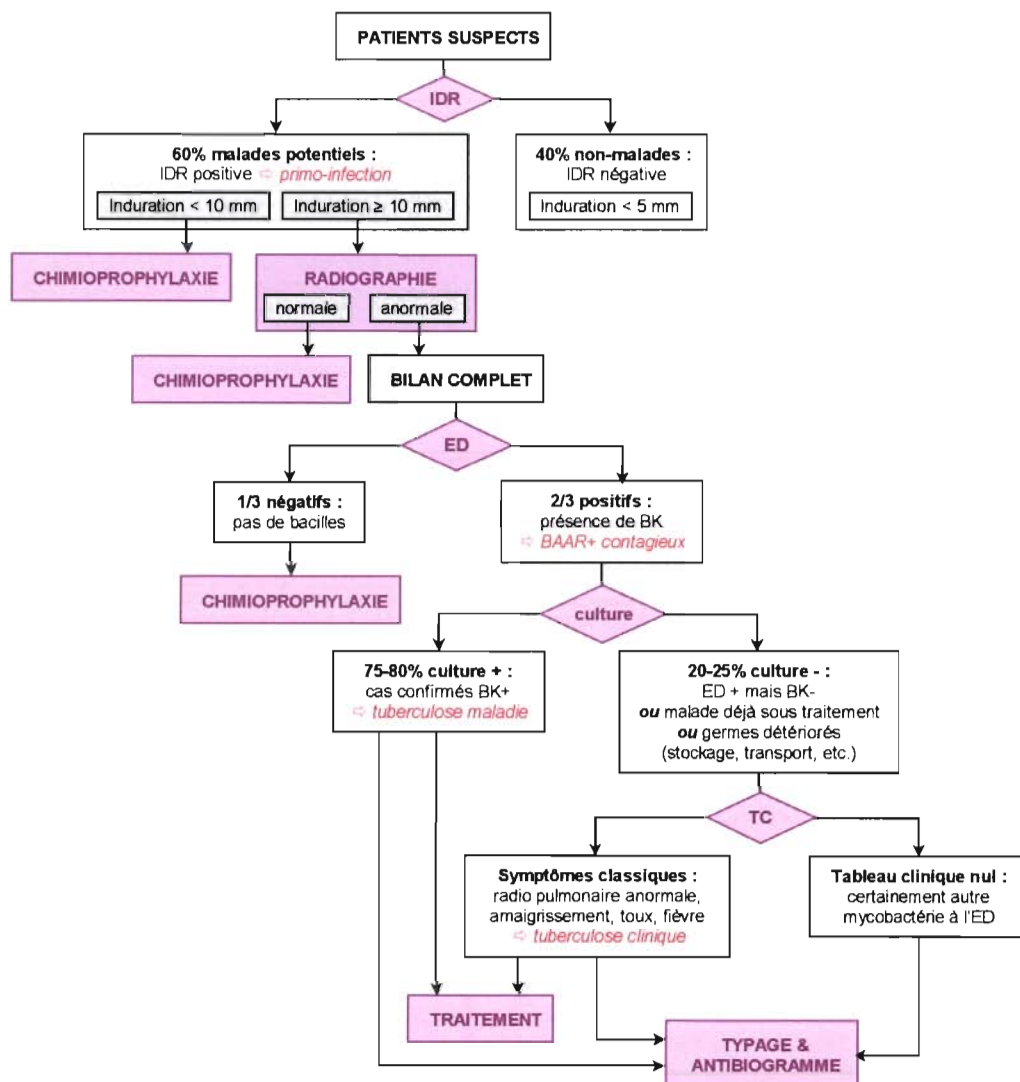


Figure 1.2 – Organigramme de confirmation d'un cas de tuberculose. IDR, intradermoréaction ; ED, examen direct ; TC, tableau clinique.

### 1.2.2.5. Cas des tuberculoses extra-pulmonaires

Les tuberculoses extra-pulmonaires peuvent être ganglionnaires, miliaires<sup>3</sup>, méningées, digestives, articulaires, etc. Pour ces formes cliniques, qui sont beaucoup plus difficiles à diagnostiquer, l'examen anatomopathologique, c'est-à-dire basé sur les lésions organiques observées, est décisif. Il peut ensuite s'accompagner d'un diagnostic bactériologique (prélèvements d'urine, de liquide céphalorachidien,...) et/ou histologique (exérèse).

<sup>3</sup> Tuberculose à dissémination sanguine.

### **1.2.3. Le traitement**

Le traitement antibiotique de la tuberculose est codifié par l'OMS, ainsi que les examens cliniques et paracliniques de surveillance devant être effectués.

#### **1.2.3.1. Les différents types de traitements**

##### *Le traitement préventif*

La chimioprophylaxie (traitement préventif combinant deux antibiotiques différents) est appliquée, soit aux tuberculoses-infections latentes (ou PIT) pour éviter que le patient ne développe plus tard une tuberculose active, soit aux sujets à risque exposés à un cas de tuberculose contagieuse.

##### *Le traitement curatif*

Le schéma thérapeutique de la tuberculose maladie recommandé en France chez l'adulte est la prise quotidienne d'un traitement en deux phases, avec dans un premier temps l'association de quatre antibiotiques anti-tuberculeux pendant deux mois (isoniazide, rifampicine, pyrazinamide et éthambutol), puis, dans un deuxième temps, la prise de deux antibiotiques (isoniazide et rifampicine) pendant quatre mois [SPLF, 2004]. Cette combinaison de différents antibiotiques (ou polychimiothérapie) permet, en cas de mauvaise observance<sup>4</sup> du traitement, d'éviter la sélection de mutants résistants, risque auquel expose la conduite d'une monothérapie. Pendant le premier mois, la tolérance au traitement est explorée à l'aide d'un bilan hépatique hebdomadaire. La primo-infection patente avec signes radiologiques et/ou généraux doit être traitée comme une tuberculose maladie [CSHPF, 2003].

#### **1.2.3.2. Délai de mise en route du traitement**

Étant donné le délai de diagnostic bactériologique (4 semaines en moyenne), la mise en route du traitement anti-tuberculeux repose généralement sur un diagnostic anatomopathologique, c'est-à-dire sur la mise en évidence de lésions histologiques (par exemple des « cavernes pulmonaires » visibles sur les radiographies dans le cas d'une tuberculose pulmonaire) et ce, dans un contexte clinique évocateur (affaiblissement, amaigrissement, etc.). On parle alors de tuberculose clinique (que l'on appellera tuberculose maladie si elle est confirmée ultérieurement par culture).

---

<sup>4</sup> L'observance se définit comme la capacité, pour un patient, de respecter la prescription du médecin. Le terme de "compliance" est également utilisé.

### **1.2.3.3. Délai de guérison**

Sous traitement anti-tuberculeux bien conduit, on considère qu'un patient atteint de tuberculose à bacilles sensibles n'est plus contagieux après 15 jours de traitement. La guérison complète est acquise au bout de six mois si le traitement a été bien suivi [CSHPPF, 2003].

## **1.3. La tuberculose à bacilles multirésistants**

### **1.3.1. Causes et implications**

Une mauvaise observance du traitement ou un traitement écourté peuvent conduire à une rechute du patient et/ou à l'émergence de souches bactériennes résistantes aux antibiotiques [Neville *et al.*, 1994 ; Lambregts-van Weezenbeek et Veen, 1995]. Cette résistance, dite résistance acquise, par opposition à une résistance primaire [Yew et Chau, 1995], limite sévèrement l'efficacité du traitement chez le patient concerné, et peut conduire au maintien d'un pool de bactéries potentiellement résistantes dans la population, constituant ainsi un frein à l'éradication de la maladie [Farmer et Kim, 1998].

### **1.3.2. Traitement**

Le traitement antibiotique de ces cas rares doit reposer sur l'association de 4 ou 5 anti-tuberculeux, et ce pendant deux ans (car les antibiotiques utilisables sont beaucoup moins actifs que l'isoniazide et la rifampicine).

## **1.4. Contraintes liées aux caractéristiques intrinsèques de la tuberculose et implications en santé publique**

### **1.4.1. Contraintes diagnostiques**

Les étapes de la confirmation d'un cas de tuberculose sont nombreuses (figure 1.2), ce qui rend le diagnostic de cette maladie fastidieux. De plus, le diagnostic implique un certain nombre d'outils (en particulier pour le volet génétique) et des analyses parfois onéreuses, surtout pour des pays en voie de développement. Ceci implique également un bon retour de l'information entre les organismes ou les laboratoires en charge des différentes étapes de diagnostic, s'ils sont plusieurs. Mais en Guyane par exemple, cet échange

d'informations fonctionne mal, les médecins n'ayant pas systématiquement de retour des analyses réalisées, le traitement étant de toute façon déjà engagé.

#### **1.4.2. Contraintes liées au pathogène**

Il existe aussi des contraintes liées aux caractéristiques épidémiologiques particulières de la maladie telles que (1) la période de latence asymptomatique pendant laquelle il est particulièrement difficile de détecter les nouveaux cas, (2) une croissance lente de la bactérie en culture, (3) des formes extra-pulmonaires difficiles à diagnostiquer, et (4) la possibilité de réactivation endogène de tuberculoses anciennes, en particulier chez les sujets âgés.

#### **1.4.3. Contraintes associées au traitement**

Il existe à l'heure actuelle au moins deux contraintes majeures associées au traitement de la tuberculose.

##### *Possibilité de faux positifs*

Fait inhérent à la croissance lente de la bactérie, le délai de diagnostic est long et de ce fait, le traitement est généralement initié avant confirmation bactériologique. Ceci peut conduire au traitement de faux positifs, par exemple chez les patients ayant des lésions séquellaires de tuberculoses anciennes.

##### *Un traitement contraignant*

D'autre part, le traitement anti-tuberculeux est très contraignant pour le patient : certains des antibiotiques peuvent être mal tolérés ou toxiques ; le traitement est long (tous les jours pendant 6 mois, parfois plus selon le cas) ; la posologie dépend du poids du patient et elle varie pour les quatre antibiotiques utilisés ce qui peut amener le patient à prendre 10 cachets par jour ; en cas de co-infection par le VIH, le traitement est encore plus lourd (plus de 20 comprimés par jour dont la tolérance digestive est souvent médiocre) [CSHPF, 2003]. Les formes galéniques combinant deux ou quatre antibiotiques dans un même comprimé sont donc recommandées pour favoriser l'observance des patients [SPLF, 2004].

### **1.5. Bilan**

Ces informations mettent en évidence les problèmes inhérents à la prise en charge de la tuberculose, tant du point de vue du diagnostic que du point de vue du traitement : une phase asymptomatique de la maladie qui peut durer

plusieurs années ; une croissance lente de la bactérie qui ne permet pas un diagnostic rapide ; un traitement long et contraignant pour le patient ; des problèmes de résistance aux antibiotiques ; le coût associé au traitement par polychimiothérapie [Farmer et Kim, 1998]. Ces différents points contribuent à freiner l'éradication de la maladie. D'autre part, s'ils ne sont pas identifiés et traités de façon appropriée, les cas de tuberculose non recensés par les services de surveillance contribuent au maintien de l'agent pathogène dans l'environnement [WHO, 1997]. À ces contraintes s'ajoutent celles inhérentes au pays dans lequel circule la maladie, développées ci-après pour la Guyane.

## **2. Le contexte Guyanais**

La Guyane a été choisie comme site d'étude pour notre travail de recherche car son taux d'incidence de tuberculose est le plus important de France (*cf.* chapitre 2). Il s'agit ici de présenter le cadre géographique général de l'étude, mais surtout de développer certains aspects démographiques, sociaux, économiques ou sanitaires de cette région qui contribuent à freiner l'éradication de la tuberculose.

### **2.1. Caractéristiques géographiques générales**

Avec une surface de 84 000 km<sup>2</sup>, la Guyane est le plus grand département français d'Outre-Mer (DOM). Elle est située au Nord-Est de l'Amérique du Sud, encadrée par l'océan Atlantique au Nord et les deux fleuves frontaliers qui sont le Maroni, frontière avec le Surinam à l'Ouest, et l'Oyapock qui borde le Brésil à l'Est (figure 1.3).

Située entre 2 et 6° de latitude nord, et 52 et 54° de longitude ouest, la Guyane a un climat de type équatorial, chaud et humide, et 90% du territoire est occupé par la forêt amazonienne. Le reste de l'espace, une plaine côtière de 40 km de large, concentre l'essentiel de la population totale, évaluée à 157 213 habitants au dernier recensement de l'INSEE (Institut National de la Statistique et des Études Économiques ; <http://www.insee.fr>) en 1999.

Les principales agglomérations sont la région de Cayenne, appelée localement "Ile de Cayenne", et les villes de Kourou et St Laurent du Maroni (figure 1.3). L'Ile de Cayenne est composée de trois communes : Cayenne, Rémire-Montjoly et Matoury. Elle constitue le principal pôle d'attraction économique et humain, et regroupe plus de 50% de la population du département avec 84 181 habitants en 1999 (dernier recensement INSEE).

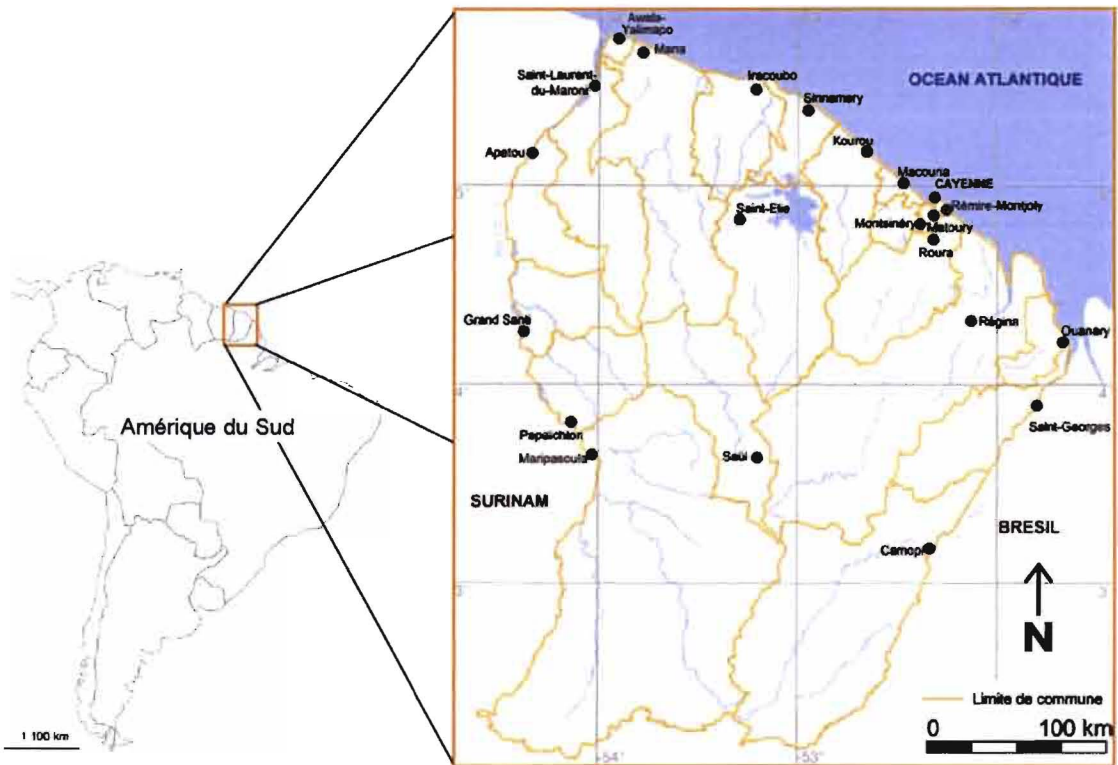


Figure 1.3 – Situation géographique de la Guyane française.

## 2.2. Quelques caractéristiques socio-démographiques

Le passé historique de la Guyane lui confère une double identité de département français d'Outre-Mer (DOM) intégré dans l'espace européen, et de département français d'Amérique (DFA), inscrit dans l'espace insulaire caribéen et l'espace Sud-Américain. Cette situation hybride a des impacts sur la situation socio-économique de ce pays.

### 2.2.1. Une multiethnicité

#### *Un territoire multiethnique*

Sur le plan démographique proprement dit, la Guyane présente la particularité d'abriter sur son sol des groupes nationaux et ethniques très divers qui contribuent au pluriethnisme et au multiculturalisme de cette région : amérindiens, noirs marrons<sup>5</sup>, créoles, métropolitains, et divers groupes migrants (chinois, haïtiens, brésiliens, surinamais, guyaniens, hmongs, libanais, etc.) [Ureña Rib, 2002]. La Guyane compte actuellement 30% de

<sup>5</sup> Descendants d'esclaves fugitifs de l'ex-Guyane néerlandaise, actuel Surinam, majoritairement implantés sur le cours inférieur du fleuve Maroni.

population de nationalité étrangère (dont 35% de surinamais, 30% de haïtiens et 15% de brésiliens), et moins de 50% de population née sur le territoire [Léglise, 2004].

### *Histoire des migrations en Guyane*

L'importante tradition d'immigration en Guyane débute au XVII<sup>ème</sup> siècle avec la conquête du plateau des Guyanes [Barret, 2001]. À cette période, les amérindiens autochtones voient arriver européens et esclaves, dont les descendants (créoles et noirs marrons) résident aujourd'hui en Guyane [Ureña Rib, 2002]. Au XIX<sup>ème</sup>, d'autres populations s'implantent à l'occasion de la construction du bagne ou de la ruée vers l'or (1890). L'immigration en provenance du Brésil débute dès 1965. Échappant aux difficultés économiques de leur pays, des brésiliens viennent participer aux grands travaux initiés en Guyane, en particulier la construction du Centre Spatial Guyanais à Kourou (1965), puis du barrage de Petit-Saut (1989) [Barret, 2001]. L'immigration en provenance d'Haïti débute dans les années 1970, puis se tarit à partir de 1986, suite aux changements politiques en Haïti et à la nouvelle politique d'immigration en Guyane. Suit une deuxième vague dans les années 1990, avec la politique de regroupement familial [Charrier, 2002]. L'immigration en provenance du Surinam commence au milieu des années 1980, avec la guerre civile dans ce pays [Léglise, 2004]. À l'initiative du gouvernement français, on voit également s'implanter en Guyane en 1977 une communauté d'agriculteurs Hmongs venus du Laos [Barret, 2001].

### **2.2.2. Une population très jeune et en forte croissance**

Le solde naturel et le solde migratoire (qui représentent respectivement 75% et 25% du gain de population) se combinent pour assurer à la Guyane une très forte croissance démographique. La population étrangère, constituée essentiellement de jeunes en âge de travailler, contribue au dynamisme démographique du département, avec un taux d'accroissement de la population de 3,6% par an. L'excédent de la natalité sur la mortalité est le principal facteur de cette augmentation. Le taux de natalité est de 31,2 naissances pour 100 000 habitants en 1999, contre 12,7 en métropole [INSEE, 1999]. Cette dynamique démographique explique l'extrême jeunesse de la population guyanaise (50 % des habitants ont moins de 25 ans).

### **2.2.3. Des difficultés sociales et économiques**

Parallèlement à cette explosion démographique, on observe un taux de chômage de 23% de la population active, un Produit Intérieur Brut (PIB) par habitant moitié moindre de celui de la Métropole (mais qui demeure le plus fort

d'Amérique du Sud), et une population clandestine croissante évaluée à 10 000 personnes, constatations qui font craindre des difficultés sanitaires futures.

## **2.3. Le contexte sanitaire**

Le contexte guyanais est très particulier, au regard de la métropole et des autres départements d'Outre-Mer. Il existe en effet sur le territoire une forte hétérogénéité des équipements urbains (réseau routier, structures sanitaires), qui se traduit par un contraste très marqué entre les régions du littoral, qui bénéficient d'un certain nombre d'aménagements, et les régions de l'intérieur, pour la plupart difficilement accessibles et fortement isolées. À cela s'ajoute une forte hétérogénéité des populations, et de leurs conditions de vie.

Ces hétérogénéités et ce qu'elles impliquent en terme de santé pour la Guyane sont détaillées ci-dessous.

### **2.3.1. Un territoire mal maîtrisé**

Dans les années 50, la départementalisation de la Guyane s'accompagne d'une volonté d'amélioration des structures de soins, ainsi que de campagnes sanitaires avec, par exemple, en 1949 la première campagne de lutte contre le paludisme par pulvérisation de DDT [Léopold Aït-Ouada, 2002].

Mais il persiste des difficultés importantes à maîtriser cet immense territoire où vivent par endroits des populations isolées et difficilement accessibles. Le réseau routier, de densité extrêmement faible, ne dessert que les communes du littoral ; les communes de l'intérieur ne sont accessibles que par voie fluviale ou, pour certaines, uniquement par voie aérienne, ce qui est par exemple le cas de Saül (figure 1.4).

### **2.3.2. Un réseau sanitaire insuffisant**

Le département dispose de trois hôpitaux, situés à Cayenne, Kourou, et Saint Laurent du Maroni, et de trois cliniques sur Cayenne, auxquels s'ajoutent 21 structures de santé (centres de santé, avec au moins un médecin à temps plein, et postes de santé) qui assurent la continuité des soins sur tout le territoire (figure 1.4).

Mais, malgré une distribution relativement homogène de centres médicaux sur l'ensemble du territoire, ces structures souffrent d'un manque cruel de personnel médical.

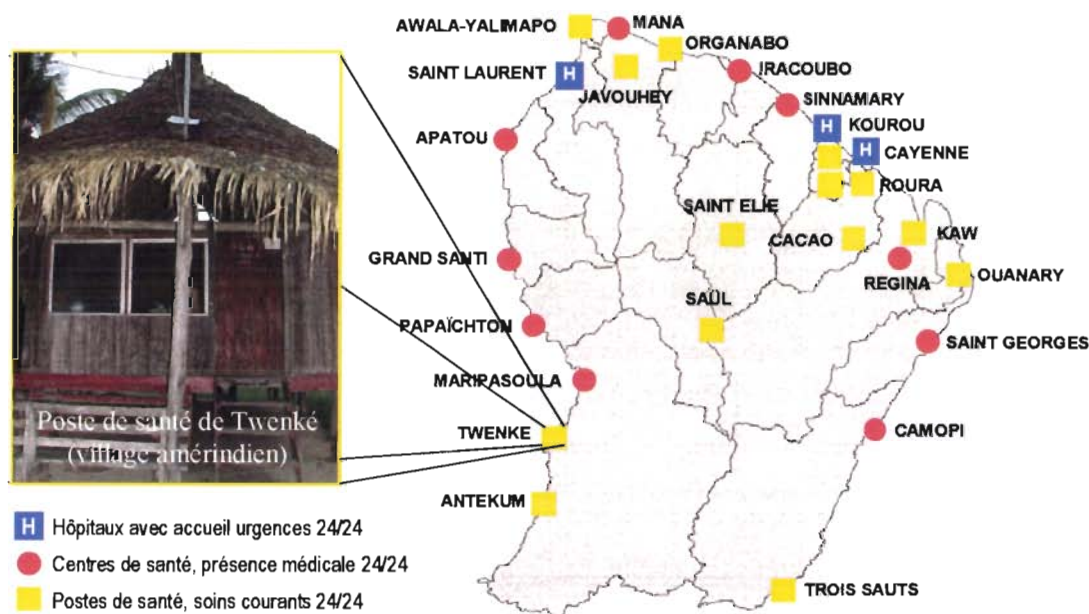


Figure 1.4 – Les structures sanitaires de Guyane.  
(Source : DSDS de Guyane. Photo : Vanina Guernier, 2004)

### 2.3.3. Situation épidémiologique

Les modifications du mode de vie ont contribué à l'augmentation de la proportion de maladies cardio-vasculaires et de sociopathies<sup>6</sup>, tandis que le taux de maladies infectieuses et parasitaires (MIP) reste élevé. La Guyane est en phase de transition épidémiologique [Bourgarel, 1994], qui se caractérise par la diminution graduelle de la part des maladies infectieuses au profit des maladies chroniques, et annonce une stabilisation future de la mortalité à un faible niveau [Omran, 1977]. Selon une étude de l'INSERM (Institut National de la Santé et de la Recherche Médicale) les maladies de l'appareil circulatoire occupent la première place parmi les causes de décès en Guyane (24,9% des décès, contre 8,3% pour les MIP, dont 47% de Sida et infections par le VIH) [INSEE, 1999].

La persistance en Guyane d'une fréquence élevée de décès dus aux MIP est liée à l'immigration, à la diffusion récente du Sida (en forte croissance depuis 1979), et à une surmortalité masculine, les femmes ayant fortement bénéficié des progrès sanitaires [Bourgarel, 1994].

### 2.3.4. Des infrastructures urbaines très inégales

En zone urbaine (principalement les communes du littoral), les structures sanitaires sont plus importantes (figure 1.4), de même que les ressources

<sup>6</sup> Affection qui engendre des phénomènes d'interaction morbide entre un individu et son milieu social.

économiques, et le taux de scolarisation, mais il existe de fortes hétérogénéités spatiales et sociales. Ainsi, en 1990, 33% des résidences ne disposent ni de l'eau ni de l'électricité à l'intérieur du logement [INSEE, 1991]. L'Ile de Cayenne recense à elle seule au moins 10 petits bidonvilles dans lesquels les conditions sociales et d'hygiène sont difficiles, et des quartiers d'habitat informel [DDE<sup>7</sup>-AU&S, 1997]. Trois facteurs définissent l'habitat informel : le non respect des règles d'urbanisme, le non respect des règles de construction et l'implantation sur des terrains non constructibles [DDE-AU&S, 1997].

Au total, on a recensé en 1997 sur l'Ile de Cayenne 24 zones d'habitat insalubre, c'est-à-dire qui ne répondent pas aux normes de maintien de la santé des habitants [DDE-AU&S, 1997]. Les populations migrantes, qui sont bien souvent des groupes à faible pouvoir d'achat, non couverts par la sécurité sociale et qui ne parlent pas toujours le français, sont particulièrement touchées par cette précarité [Bourgarel, 1994].

### **2.3.5. Un taux élevé de Sida**

La Guyane est le département français le plus touché, avec un taux d'incidence estimé à environ 43,8 cas pour 100 000 habitants entre janvier 2001 et mars 2002. D'après les estimations de la DSDS<sup>8</sup> (Département de la Santé et du Développement Social), on en serait au total à 3 000 séropositifs fin 2001, soit près d'un habitant sur cinquante.

Au 31 décembre 1995, 79% des cas de Guyane relèvent d'une transmission hétérosexuelle, 10% d'une transmission materno-foetale, et 5% d'une transmission homosexuelle ou bi-sexuelle. La situation est très différente de la France métropolitaine où les groupes les plus touchés sont les homo/bisexuels (49%) et les toxicomanes (25%) [CNS, 1996].

Ces taux de sida sont très importants à prendre en compte dans la mesure où l'infection VIH a un effet amplificateur sur l'épidémie tuberculeuse (*cf.* co-infection VIH/tuberculose chapitre 2 §2.2.4.).

## **2.4. Bilan de la situation en Guyane**

Il existe en Guyane des populations défavorisées, vivant dans des conditions de vie précaires. Cette situation est aggravée par le fait que ces populations n'ont pas toujours recours aux structures sanitaires existantes, de part leur clandestinité sur le territoire guyanais, ou simplement parce qu'elles sont immigrées et ne parlent pas français. On relève également un problème d'accès

---

<sup>7</sup> Direction Départementale de l'Équipement, ici celle de Guyane.

<sup>8</sup> La dénomination DSDS est spécifique à la région Antilles-Guyane. Elle correspond au regroupement d'une Direction régionale (Drass) et d'une Direction départementale des affaires sanitaires et sociales (Ddass).

aux soins, avec un nombre de structures sanitaires insuffisant, et des difficultés d'accès dans les régions de l'intérieur

Si l'on ajoute à ces facteurs socio-économiques le fait que la Guyane se situe en zone tropicale, contexte favorable à l'occurrence de nombreux pathogènes infectieux et parasitaires [Guernier *et al.*, 2004], il est clair que la Guyane cumule de nombreux facteurs de risque en termes sanitaires.

### 3. Conclusion

La première partie de ce chapitre, consacrée aux généralités sur la tuberculose, a permis de mettre en évidence les difficultés propres à la lutte contre cette maladie, et la nécessité de mettre en place des stratégies de contrôle qui passent par l'identification précoce et exhaustive des nouveaux cas. Cette identification doit être précoce, non seulement pour traiter les nouveaux malades potentiellement contagieux, mais aussi pour engager les enquêtes de proximité permettant d'identifier les personnes ayant été au contact du malade, et devant être traitées préventivement [Pillay et Clarke, 2003]. L'identification des cas doit également être exhaustive pour garantir une mesure correcte de l'incidence de la maladie, contrôler l'évolution des cas, détecter les épidémies, et planifier des stratégies de lutte et de prévention. Pour toutes ces raisons, il est particulièrement important de pouvoir (1) identifier rapidement l'ensemble des cas de tuberculose à traiter, et (2) comprendre les facteurs génétiques, écologiques et humains qui permettent la persistance de la maladie dans un environnement donné, en l'occurrence celui de la Guyane. Ces deux points constituent les objectifs principaux du présent travail de thèse. Ce travail se trouve être particulièrement justifié dans le contexte de la Guyane où l'incidence de la tuberculose est importante et largement sous-estimée par les organismes de notification, comme nous le montrons dans le chapitre suivant. Nous avons par ailleurs montré dans la seconde partie de ce chapitre que cette région réunit plusieurs facteurs favorables à la persistance de la tuberculose : existence de populations défavorisées, d'un taux élevé de sida, de populations immigrantes en provenance de pays à forte incidence de tuberculose, etc. L'objectif final de la thèse étant de contribuer à l'aide à la décision dans la lutte contre la tuberculose en Guyane, le travail de recherche a systématiquement été replacé dans le contexte particulier de ce département. Mais il ouvre également des perspectives intéressantes pour les pays dont le contexte global apparaît être similaire ou proche. Le chapitre suivant propose une analyse critique des données épidémiologiques de la tuberculose en Guyane.

## Chapitre 2

# Données chiffrées de la tuberculose en Guyane : analyse critique

*L'objet de ce chapitre est de proposer une analyse épidémiologique critique de toutes les données concernant la tuberculose en Guyane. Nous passons tout d'abord en revue les données chiffrées existantes, que ce soit les données officielles publiées par l'État, ou celles obtenues après enquête auprès des différents acteurs de la lutte anti-tuberculeuse en Guyane sur la période 1996-2003, et ce, en accord avec les règles de confidentialité. Ensuite, nous procédons sur la base de ces données à un certain nombre d'analyses descriptives concernant l'épidémiologie des cas afin de mettre en évidence les facteurs de risques humains (âge, sexe, etc.) spécifiques de Guyane. Enfin, une évaluation statistique par la méthode de capture-recapture est conduite dans le but d'estimer l'incidence théorique espérée de la tuberculose en Guyane, et donc d'évaluer dans quelle mesure les résultats obtenus sur la base des données d'enquêtes sont représentatifs de l'ensemble des cas de tuberculose de Guyane sur la période 1996-2003.*

## 1. La tuberculose en Guyane en chiffres

Les informations détaillées ci-après concernent la période 1996-2003 pour laquelle les données sont disponibles dans l'ensemble des sources étudiées, c'est-à-dire d'une part les organismes officiels de surveillance de la tuberculose, et d'autre part les informations obtenues après des enquêtes de terrain que nous avons réalisées auprès de trois organismes de prise en charge de la tuberculose en Guyane (cf. §1.2.4.).

### 1.1. Les données officielles

En France, la tuberculose fait l'objet d'une déclaration obligatoire (DO) depuis 1964 auprès de l'Institut de veille sanitaire (InVS), établissement public

placé sous la tutelle du ministère des Solidarités, de la Santé et de la Famille, qui a pour mission de surveiller l'état de santé de l'ensemble de la population, et d'alerter les pouvoirs publics en cas de menace pour la santé publique.

C'est le Département de la Santé et du Développement Social (DSDS) de Guyane qui fournit chaque semaine à l'InVS le recueil des déclarations obligatoires faites par les médecins pour ce département d'Outre-Mer.

### **1.1.1. Les sources de données**

Les données de l'InVS sont issues des différents rapports sur la tuberculose publiés en ligne sur le site <http://www.invs.sante.fr/>. Les données de la DSDS, ont été récupérées directement auprès de cet organisme, basé à Cayenne.

### **1.1.2. Définition d'un cas**

La déclaration obligatoire est basée sur la présence d'une tuberculose, confirmée ou non, mais ayant conduit à la mise en route d'un traitement avec au moins trois antituberculeux [Rieder *et al.*, 1996 ; Schwoebel *et al.*, 1996], ce qui exclut les patients sous traitement prophylactique. Les mycobactéries atypiques (*M. avium*, *M. gordonae*, *M. marinum*, *etc.*) doivent également être exclues des DO, dans la mesure où les résultats d'identification de la bactérie ont bien été transmis au médecin traitant, en charge de la DO.

### **1.1.3. Les données de l'Institut de Veille Sanitaire**

Selon un récent rapport publié par l'InVS [Che et Bitar, 2005], la région la plus touchée de France est l'Ile de France, avec un taux d'incidence trois fois et demi supérieur à la moyenne nationale hors Ile de France et hors DOM. À titre comparatif, le tableau 2.1 récapitule les taux d'incidence calculés en France métropolitaine, en Ile de France et en Guyane sur la base des DO.

On constate que le taux d'incidence en métropole diminue régulièrement depuis 1996. Globalement, le taux d'incidence en Ile de France a également diminué depuis 1996, même si cette décroissance est beaucoup moins régulière qu'en métropole (petite ré-augmentation en 1998 et 2000), et que le taux annuel reste en moyenne deux fois plus important que celui de métropole. En Guyane en revanche, l'incidence de la tuberculose est très irrégulière, avec une forte diminution de 1996 à 1998, une augmentation avec un taux maximal en 2000 dépassant le taux de 1996, et enfin une re-diminution jusqu'en 2003 où le taux notifié est le même que celui de métropole. Ceci traduit en fait une forte irrégularité de la déclaration obligatoire par les médecins de Guyane, c'est-à-dire que le formulaire de DO devant être rempli, signé et envoyé à la DSDS pour chaque malade, ne l'est pas systématiquement [Léopold Aït-Ouada, 2002].

Tableau 2.1 – Taux d'incidence de la tuberculose en France métropolitaine, en Ile de France et en Guyane, et déclarations obligatoires (DO) en Guyane, issues de la DSDS, 1996-2003<sup>1</sup>.

Année	Métropole	Ile de France	Guyane	
	Incidence (cas / 100 000 hab.)		Incidence	DO
1996	12,8	30,6	36,6	43
1997	11,5	26,7	14,7	24
1998	11,1	27,5	3,7	6
1999	11,2	26,4	11,6	19
2000	11,2	28,8	39,4	62
2001	10,8	27,2	38,2	61
2002	10,5	27,1	24,9	43
2003	10,2	24,8	10,2	37
<b>Moyenne</b>	<b>11,2 (±)</b>	<b>27,4 (±)</b>	<b>22,4 (±)</b>	<b>36,9 (±)</b>

#### 1.1.4. Les données du Département de la Santé et du Développement Social (DSDS)

##### *Données détaillées par commune*

Les données de l'InVS représentent le cumul de l'ensemble des données de Guyane. Cependant, le recueil exhaustif des données que nous avons réalisé par enquête auprès des différents acteurs de la lutte anti-tuberculeuse (§1.2.) ne concerne que l'Ile de Cayenne. Pour pouvoir comparer les données issues des différentes sources, nous avons donc récupéré les informations par commune auprès de la DSDS. Le tableau 2.2 présente les données brutes recensées.

Tableau 2.2 – Nombre de cas de tuberculose déclarés en Guyane sur la période 1996-2003 en fonction du lieu de résidence des patients (Surinam et quelques villes de Guyane).

Année	Total <sup>a</sup>	Surinam	Ile de Cayenne	Kourou	St Laurent	Autres <sup>b</sup>
1996	43	0	25	5	5	8
1997	24	0	12	2	7	3
1998	6	0	4	0	1	1
1999	19	1	8	1	6	3
2000	66	2	35	8	9	12
2001	61	1	35	8	4	13
2002	43	1	23	4	6	9
2003	37	2	18	4	9	4

<sup>a</sup> total des cas, que le lieu de résidence soit en Guyane ou au Surinam

<sup>b</sup> villes de Guyane autres que Cayenne, Matoury, Rémire-Montjoly, Kourou et St Laurent

<sup>1</sup> Des données antérieures à cette période sont disponibles mais, afin de permettre la comparaison avec les autres sources de données (§1.2.), l'analyse des données InVS et DSDS se limite à la période pour laquelle des données sont disponibles pour l'ensemble des sources étudiées (cf. §1.2.4. Période de l'étude).

*Données corrigées*

Nous avons dans un deuxième temps apporté quelques corrections aux données fournies par la DSDS afin d'éliminer certaines erreurs de déclaration. Ainsi, la consultation des résultats d'analyses bactériologiques concernant l'identification des prélèvements bactériens a révélé que 12 patients déclarés sur la période 1996-2003 étaient en fait porteurs de mycobactéries atypiques. Ces cas sont notifiés par année dans le tableau 2.3, et ont été par la suite éliminés des données.

D'autre part, afin de suivre la chronologie d'apparition des cas, nous avons corrigé certaines dates. À l'InVS et à la DSDS, la date retenue pour le décompte des cas est la date à laquelle le médecin a fait la déclaration, qui peut suivre de plusieurs mois le début de la maladie. Dans les autres sources d'information, la date retenue est celle du début du traitement des malades, qui est plus pertinente pour le suivi temporel des cas. Afin d'homogénéiser les données des différentes sources, nous avons donc choisi comme référence temporelle la date de début du traitement. Après recoupement des patients dans les différentes bases, les cas déclarés à la DSDS ont été reclassés par année en fonction de la date de début du traitement du patient (tableau 2.3). Sept patients dont le traitement avait en fait commencé en 1995 ont, par exemple, été éliminés.

Les données du tableau 2.3 offrent peu de différences avec les données du tableau 2.2, à quelques cas près qui ont été ajoutés ou enlevés. Le maximum de cas est observé pour 2000-2001, et le minimum pour 1998.

Tableau 2.3 – Nombre de cas de tuberculose déclarés sur la période 1996-2003, en fonction de la ville ou du pays de résidence des patients, et après corrections.

Année	Total	Surinam	Ile de Cayenne	Kourou	St Laurent	Autres
1996	37	0	22	3	6	6
1997	23	0	12 <sup>a</sup>	2	6	3
1998	8	0	4 <sup>d</sup>	0	1	3
1999	26	1	13	3	6	3 <sup>a</sup>
2000	62	2	33 <sup>a</sup>	7	9	11
2001	60	1	34 <sup>c</sup>	8	4	13 <sup>a</sup>
2002	42	1	22 <sup>b</sup>	4 <sup>a</sup>	6	9 <sup>a</sup>
2003	34	2	17	3	9 <sup>a</sup>	3

<sup>a</sup> dont un cas infecté par une mycobactérie atypique

<sup>b</sup> dont deux cas infectés par une mycobactérie atypique

<sup>c</sup> dont trois cas infectés par une mycobactérie atypique

<sup>d</sup> dont 1 cas inconnu (non retrouvé dans les enquêtes ultérieures)

*Remarque*

Cette correction n'a pas pu être appliquée aux données de l'InVS pour l'ensemble de la Guyane, ces données étant anonymées et cumulées par année.

Pour cette raison, les comparaisons ultérieures de nombre de cas ou d'incidences entre les différentes sources de données font référence aux données corrigées de la DSDS concernant l'île de Cayenne. Les enquêtes que nous avons par la suite réalisées au sein de différents organismes de prise en charge de la tuberculose se sont donc limitées à cette zone d'étude (cf. §1.2.4. ci-après).

### **1.1.5. Au final, les données officielles sont-elles fiables ?**

Notre examen des registres de la DSDS a montré que sept cas enregistrés à la DSDS résident au Surinam, alors que seuls les patients résidants en France sont soumis à déclaration. Il y a fort à penser que ces patients vivent sur les berges du fleuve Maroni, et ont été soignés à l'hôpital de Saint Laurent. La nationalité des patients n'étant pas précisée, il est possible qu'ils soient français, ce qui expliquerait que les médecins aient déclaré ces cas en Guyane.

D'autre part, des cas porteurs de mycobactéries atypiques qui n'auraient pas dû être déclarés l'ont été. Ceci s'explique par le fait que les traitements antituberculeux démarrent bien souvent avant l'obtention des résultats d'analyses bactériologiques, du fait du délai d'attente d'environ quatre semaines (cf. chapitre 1). Il existe donc un biais dans les données officielles de DO, et les moyens de l'éliminer sont discutés §1.2.6.

Afin d'évaluer l'importance de ce biais, ces données ont été croisées avec les données issues de trois autres sources d'information, auprès desquelles nous avons mené une enquête rétrospective exhaustive.

## **1.2. Les données issues d'enquêtes**

### **1.2.1. Éthique et déontologie de l'étude**

En tant que département français, la Guyane dispose, tout comme la Métropole, d'un réseau de soins pour la prise en charge de la tuberculose. Ce réseau intègre différentes structures de prévention, de traitement, et de lutte anti-tuberculeuse, auprès desquelles nous avons mené nos enquêtes.

Étant donné le circuit de prise en charge des patients tuberculeux, qui passe successivement par chacun des différents organismes du réseau de soins (cf. §1.2.2. ci-après), les cas recensés auprès de chacun de ces organismes devraient être les mêmes, et également identiques aux patients ayant fait l'objet d'une déclaration obligatoire auprès de la DSDS. Le but de cette partie de l'étude est justement de vérifier si c'est bien le cas.

Afin de recouper les informations recensées auprès des différentes sources, les enquêtes que nous avons menées incluent nécessairement la collecte de données directement et indirectement nominatives telles que le nom, le

prénom, le sexe et la date de naissance. Il est donc important de souligner que nos enquêtes garantissent la protection des données confidentielles, conformément aux règles d'éthique médicale et à la législation sur la recherche médicale et scientifique. Deux conventions de recherche ont été mises en place spécifiquement dans le cadre du travail de thèse, respectivement entre la structure d'accueil, c'est-à-dire l'Institut de Recherche pour le Développement (IRD) et la DSDS, et entre l'IRD et l'Institut Pasteur. Un accord de confidentialité a également été signé, conjointement avec les directeurs des différents organismes ayant donné accès à leurs données (cf. Annexe 1). Après croisement des données des différentes sources, le fichier commun réunissant l'ensemble des informations a été anonymé. Il ne constitue pas par ailleurs un fichier indirectement nominatif.

### **1.2.2. Les sources de données**

Nos enquêtes ont été réalisées auprès de trois systèmes de prise en charge de la tuberculose en Guyane : le centre hospitalier de Cayenne, les laboratoires de références (l'Institut Pasteur de Guyane et, dans une moindre mesure, l'Institut Pasteur de Guadeloupe), et le centre de lutte anti-tuberculeuse.

#### **1.2.2.1. Le centre hospitalier de Cayenne**

La prise en charge des malades tuberculeux se fait soit par les centres hospitaliers, soit par des médecins libéraux, à condition que le patient ne soit plus contagieux. Pendant la phase contagieuse qui peut durer de deux à trois semaines, les malades sont obligatoirement hospitalisés et maintenus seuls dans des chambres d'isolement, jusqu'à disparition de la contagiosité. Sur l'île de Cayenne, les cas contagieux sont traités au Centre Hospitalier de Cayenne (CHC). Le Département d'Information Médicale (DIM) de l'hôpital de Cayenne collecte et archive informatiquement l'ensemble des informations médicales concernant les patients traités ou admis à l'hôpital : nom, prénom, date de naissance, adresse, problèmes de santé et résultats des différentes analyses médicales pratiquées. Les diagnostics sont enregistrés en utilisant le système de codage correspondant à la Classification Internationale des Maladies (10<sup>ème</sup> révision, CIM-10). Pour les besoins de l'étude, un listing des cas de tuberculose a été extrait par le Dr. Sobeski de la base du DIM sur la base des diagnostics principal et secondaire. En cas de doute, une confirmation a été demandée au service de l'hôpital ayant pris en charge le patient.

En l'absence de risque contagieux, les malades peuvent être pris en charge par des médecins libéraux. Ces cas ne sont donc pas recensés au niveau du DIM mais ils apparaissent généralement dans les bases de données des laboratoires d'analyse ou du dispensaire de soins (cf. § suivants).

### **1.2.2.2. Les laboratoires de référence**

Le laboratoire des mycobactéries de l'Institut Pasteur de la Guyane<sup>2</sup> (que nous notons par la suite IPG) basé à Cayenne est en charge des examens directs et de la mise en culture des prélèvements effectués à l'hôpital ou par des médecins exerçant en cabinet libéral. Ces cultures sont ensuite envoyées à l'Institut Pasteur de Guadeloupe (que nous notons par la suite IPA, A pour Antilles), au laboratoire régional de référence pour la tuberculose et les mycobactéries. Ce laboratoire est en charge de l'identification des souches de mycobactérie isolées chez les patients des trois départements de la région Antilles-Guyane (Martinique, Guadeloupe, Guyane), de la surveillance des souches résistantes (antibiogrammes), et il assure depuis 1994 l'analyse génomique de toutes les souches reçues.

#### *Non exhaustivité des laboratoires*

Étant donné qu'une analyse génomique nécessite l'existence d'une culture viable de *M. tuberculosis*, les données recensées à l'IPG ou à l'IPA excluent les cas de tuberculose clinique ou certaines tuberculoses extra-pulmonaires pour lesquelles il est difficile, voire impossible, d'obtenir un isolat bactérien. Pour cette raison, les données de l'IPG ne peuvent être entièrement exhaustives.

#### *Non centralisation des analyses*

Jusqu'en 1999, l'ensemble des examens bactériologiques de Guyane était centralisé à l'IPG. Mais depuis cette date, le laboratoire de l'hôpital de Saint Laurent effectue l'examen direct des prélèvements, puis les envoie au laboratoire du CHU d'Angers pour la mise en culture [Léopold Aït-Ouada, 2002]. Cette non centralisation des analyses est préjudiciable à la surveillance de la tuberculose en Guyane, et c'est une autre des raisons pour lesquelles nous avons limité notre étude à l'île de Cayenne.

### **1.2.2.3. Les centres de soins et de prévention**

#### *Cas des patients ayant déclaré la maladie en Guyane*

Le centre de Lutte Anti-Tuberculeuse (LAT) est coordonné par le Conseil Général en lien avec la DSDS de Guyane, et il est le seul dispensaire de ce type existant en Guyane. Il a pour mission le suivi et le traitement des malades tuberculeux pendant 6 à 9 mois, à la sortie de leur phase d'hospitalisation. Le traitement est délivré gratuitement (en accord avec l'article L.3112-5 du code de la Santé Publique). La LAT est également en charge des enquêtes de proximité conduites autour des cas (voisins, personnes du même foyer, etc.)

---

<sup>2</sup> Site web : <http://www.pasteur-cayenne.fr/>

permettant de détecter les personnes ayant été en contact avec un individu malade, et d'engager un traitement prophylactique post-exposition si nécessaire. En théorie, l'hôpital doit informer la LAT à chaque hospitalisation d'un cas tuberculeux, afin que celle-ci puisse conduire immédiatement l'enquête de proximité. En pratique, la LAT n'est généralement informée qu'à la sortie du malade de l'hôpital, ce qui retarde cette enquête de plusieurs semaines.

#### *Cas des patients immigrés déjà sous traitement*

Les patients immigrés depuis peu ayant débuté un traitement dans leur pays d'origine sont également suivis par la LAT à leur arrivée en Guyane. Il n'est cependant pas possible de dissocier à la lecture des registres les patients nouvellement immigrés des autres cas étrangers. Ces cas ont donc été inclus dans notre étude, malgré le fait qu'ils aient été contaminés à l'étranger.

### **1.2.3. Choix de la période d'étude**

Pour permettre la comparaison et le croisement des données issues des différentes sources, il était indispensable de choisir une période d'étude pour laquelle les informations concernant la tuberculose étaient disponibles et fiables dans l'ensemble des organismes ciblés. Les données du DIM, de la LAT et de l'IPG sont disponibles de façon plus ou moins exhaustive depuis 1990-92, avec des problèmes d'archivages des données « papier ». En revanche, le laboratoire de référence régional pour la tuberculose aux Antilles-Guyane n'existe que depuis 1993, et les analyses génétiques sur les souches de *M. tuberculosis* réalisées par ce laboratoire ne sont exhaustives que depuis 1996. Pour ces raisons, l'ensemble des données analysées dans ce travail de recherche se restreint à la période allant du 1<sup>er</sup> janvier 1996 au 31 décembre 2003.

### **1.2.4. Méthode : croisement des trois sources de données**

#### *Critères d'inclusion des cas*

Outre la résidence sur l'île de Cayenne au moment de la déclaration de la maladie (cf. remarque §1.2.4.), l'inclusion des cas dans chacune des sources considérées a été faite sur la base des mêmes critères que précédemment : cas ayant débuté un traitement avec au moins trois anti-tuberculeux, excluant les traitements prophylactiques et les infections par mycobactéries atypiques. À noter également que, sur la période d'étude, aucun cas de *Mycobacterium bovis* n'a été recensé, donc les cas sont tous infectés par *M. tuberculosis*. Lorsqu'un cas apparaissait dans plus d'une source, la date retenue pour les analyses ultérieures était celle du premier enregistrement, correspondant, en général, à la date d'hospitalisation ou de mise en route du traitement.

*Identification des doublons et des triplets*

Des identifiants communs aux différentes sources de données étaient nécessaires pour identifier les cas présents dans plus d'une source. À cette fin, des informations personnelles telles que nom, sexe, date de naissance et date d'enregistrement du malade ont été collectées pour chaque cas. Les doublons (présents dans deux sources) et les triplets (présents dans trois sources) ont été identifiés sur la base du nom et de la date de naissance, et occasionnellement sur la base d'un troisième identifiant en cas de doute.

**1.2.5. Résultats**

*Croisement des trois sources*

Les résultats obtenus après recensement et croisement des informations par année et dans chacune des sources considérées sont résumés dans le tableau 2.4, auquel sont également incluses, à titre comparatif, les données corrigées de la DSDDS précédemment obtenues. Après simple croisement des données de l'IPG, du DIM et de la LAT, 381 cas de tuberculose ont été comptabilisés sur l'île de Cayenne entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003, soit 2,5 fois plus de cas que notifiés par la DSDDS sur la même période.

Tableau 2.4 – Nombre de cas de tuberculose recensés sur l'île de Cayenne entre 1996 et 2003, auprès de la DSDDS, de l'IPG, du DIM, et de la LAT, et dans ces 3 dernières sources à la fois après élimination des cas communs à plusieurs sources (doublons et triplets).

Année	DSDDS	IPG	DIM	LAT	IPG*DIM*LAT
1996	22	22	20	29	38
1997	11	33	31	32	41
1998	4	20	15	16	28
1999	13	25	25	23	35
2000	32	29	35	33	46
2001	31	33	54	47	62
2002	20	31	46	41	61
2003	17	42	57	50	70
Total	150	235	283	271	381

DSDDS, Département de la Santé et du Développement Social ; IPG, Institut Pasteur de Guyane ; DIM, Département d'Information Médicale de l'hôpital de Cayenne ; LAT, service de Lutte Anti-Tuberculeuse ; IPG\*DIM\*LAT, données combinées de l'IPG, du DIM et de la LAT.

*Cas recensés dans une ou plusieurs sources*

Sur les 381 cas recensés au total, 111 (29,1%) ont été identifiés dans une seule source, 132 (34,6%) dans seulement deux sources, et 138 cas (36,2%) dans les trois sources d'information.

Si individuellement, chacune des sources recense davantage de cas que ce que révèlent les déclarations obligatoires, aucune source n'est cependant exhaustive pour la notification des cas. En effet, l'IPG, le DIM et la LAT recensent respectivement 61,7%, 74,3% et 71,1% des 381 cas recensés au total. Le nombre cumulé de cas de tuberculose identifiés dans une, deux, ou trois sources de données est illustré par un diagramme de Venn [Nardone et Decludt, 2000] (figure 2.1).

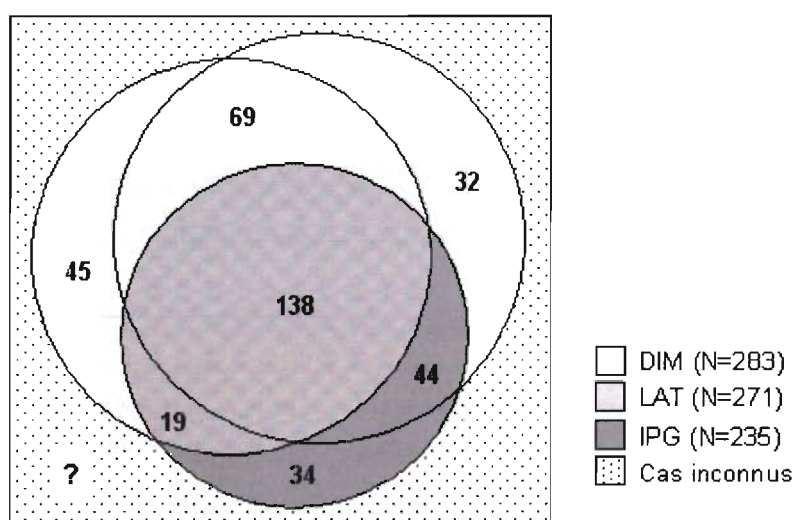


Figure 2.1 – Croisement des données de l'IPG, de la LAT et du DIM, 1996-2003. Le nombre de cas présents dans une, deux ou trois sources est noté en gras. Le point d'interrogation illustre le nombre de cas absents de ces sources. N est le nombre de cas recensés dans chacune des 3 sources entre 1996 et 2003.

### 1.2.6. Exhaustivité des sources d'information

#### *Exhaustivité de l'IPG*

Les données recueillies auprès de l'IPG ne peuvent être exhaustives dans la mesure où elles ne concernent que les confirmations bactériologiques de cas. En revanche, certains patients ne sont recensés qu'à l'IPG (9%), ce qui s'explique aisément dans le contexte local. L'île de Cayenne est un « grand village » où beaucoup de gens se connaissent ; de ce fait, certains malades choisissent d'être suivis uniquement par des médecins de ville et n'apparaissent donc ni à l'hôpital, ni à la LAT, mais apparaissent dans les listings de l'IPG si leurs examens bactériologiques sont positifs.

#### *Exhaustivité du DIM*

Outre les malades suivis en cabinet de ville, qui échappent au DIM, ce dernier ne recense pas non plus les malades qui arrivent en Guyane en ayant déjà démarré un traitement anti-tuberculeux ailleurs, et qui seront suivis

uniquement par la LAT. À côté de cela, une petite minorité de cas (8%) n'est recensée qu'au DIM. Ce sont généralement des patients n'ayant pas été jusqu'au terme de leur hospitalisation (pour cause de fugue, de décès prématuré ou de retour dans le pays d'origine, par exemple) et qui, de fait, n'apparaissent ni à l'IPG, ni à la LAT.

### *Exhaustivité de la LAT*

Certaines causes de la non-exhaustivité de la LAT ont déjà été notées ci-dessus. Les patients suivis durant tout leur traitement par un médecin libéral ou à l'hôpital (certains cas de co-infection HIV par exemple) ne sont pas recensés à la LAT, de même que les « perdus de vue » (fugues de l'hôpital ou malades ne se rendant pas à la LAT). Pour éviter ce dernier cas, l'hôpital a récemment mis en place un service d'accompagnement des patients à la LAT par le Service d'Aide Médicale d'Urgence à leur sortie d'hôpital. Cette première visite permet aux services de la LAT de prendre les renseignements nécessaires à l'enquête de voisinage et de remettre au malade son premier mois de traitement. Les médecins espèrent ainsi favoriser la compliance<sup>3</sup> des patients.

### **1.2.7. Au final, des données officielles non exhaustives**

Ces résultats confirment les résultats d'une précédente étude [Rastogi *et al.*, 1998] ayant mis en évidence une sous-déclaration des cas de tuberculose en Guyane sur la période 1994-1996. Après enquête auprès de trois organismes de surveillance de la tuberculose et croisement des données obtenues, 381 cas de tuberculose ont été recensés entre 1996 et 2003. Par comparaison avec les données corrigées de la DSDS, la sous-déclaration représente donc 61% des cas de tuberculose sur l'île de Cayenne, et on peut s'attendre au même ordre de grandeur concernant l'ensemble de la Guyane. De plus, les données de la DSDS (donc celles de l'InVS) sont soumises à caution puisque certains cas infectés par des mycobactéries atypiques ont été déclarés à tort. Enfin, la déclaration des cas, lorsqu'elle a lieu, peut suivre la date de début du traitement du patient de plusieurs semaines à plusieurs mois. Ce manque de réactivité exclut la possibilité d'un suivi en temps réel de l'épidémie, qui permettrait une intervention plus rapide des organismes de santé publique.

Suite à ce constat, il apparaît important d'évaluer quels sont les facteurs de risque humains (âge, sexe, nationalité, etc.) spécifiques de la tuberculose en Guyane. Cette analyse épidémiologique de la tuberculose en Guyane fait l'objet du paragraphe suivant.

---

<sup>3</sup> La compliance (ou observance, cf. chapitre 1) se définit comme la capacité, pour un patient, de respecter la prescription du médecin.

## 2. Épidémiologie de la tuberculose en Guyane

Dans ce sous-chapitre sont détaillées des analyses descriptives concernant l'épidémiologie des cas de tuberculose en Guyane. Il s'agit en fait d'analyser les distributions statistiques observées par âge, sexe, nationalité, etc. des cas de tuberculose sur la période 1996-2003. Le but est de mettre en évidence les facteurs de risques humains associés à la tuberculose en Guyane.

### 2.1. Cas introduits dans l'étude

Nous avons, dans la partie précédente, comptabilisé 381 cas de tuberculose, recensés sur l'île de Cayenne entre 1996 à 2003. Si l'on inclut les cas recensés hors île de Cayenne auprès de l'IPG (donc confirmés bactériologiquement), on comptabilise un total de 491 cas sur cette même période (tableau 2.5).

Tableau 2.5 – Cas de tuberculose confirmés bactériologiquement ou non confirmés ayant déclaré la maladie entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003 et résidents en Guyane.

	GUYANE		
	Ile de Cayenne	Autre	Total
Confirmés	235	110	345
Non confirmés	146	0	146
<b>Total</b>	<b>381</b>	<b>110</b>	<b>491</b>

Ce résultat n'est pas exhaustif dans la mesure où, pour les patients résidant hors île de Cayenne, on ne dispose que des cas confirmés bactériologiquement, disponibles à l'IPG. Il n'était cependant pas possible, par manque de temps, d'envisager un recueil de données supplémentaires auprès des hôpitaux de Kourou et Saint Laurent, et de l'ensemble des services de santé du département. Ce travail reste toutefois envisageable dans le futur.

N'ayant pas pu établir un recueil exhaustif des cas de tuberculose de Guyane, nous disposons au final, de deux sous-populations : (1) les cas de tuberculose de l'île de Cayenne, qui constituent un échantillon de l'ensemble des cas de Guyane, et (2) les cas de tuberculose confirmés de Guyane, qui constituent un échantillon de l'ensemble des cas de tuberculose (incluant les tuberculoses cliniques). Les analyses qui suivent ont été réalisées sur l'un ou l'autre de ces deux échantillons en fonction des données auxquelles on les compare. Par exemple, dans le cas de la répartition géographique des cas de tuberculose entre villes de Guyane, seuls les cas confirmés sont pris en compte. Si l'on incluait les cas non confirmés, qui sont disponibles uniquement pour l'île de Cayenne, la répartition serait biaisée en faveur de cette région.

*Remarque*

Il faut garder à l'esprit qu'une partie des données de l'IPG concernant les confirmations bactériologiques de cas de Saint Laurent du Maroni est susceptible d'être lacunaire depuis 1999 (cf. §1.2.1.1.), et ce, sans qu'il soit possible d'estimer dans quelle proportion.

## 2.2. Résultats

Les résultats énoncés ci-après concernent la répartition par ville, par âge et par sexe des cas de tuberculose recensés en Guyane entre 1996 et 2003, que ce soit les cas confirmés et/ou non confirmés.

### 2.2.1. Répartition des cas par ville (1996-2003)

L'île de Cayenne regroupe à elle seule 48,1% de l'ensemble des cas de tuberculose confirmés bactériologiquement entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003 (figure 2.2) ; suivent les villes de Saint Laurent (11,9%) et Kourou (7,0%), ce qui correspond globalement à la répartition de la population de Guyane (cf. chapitre 1).

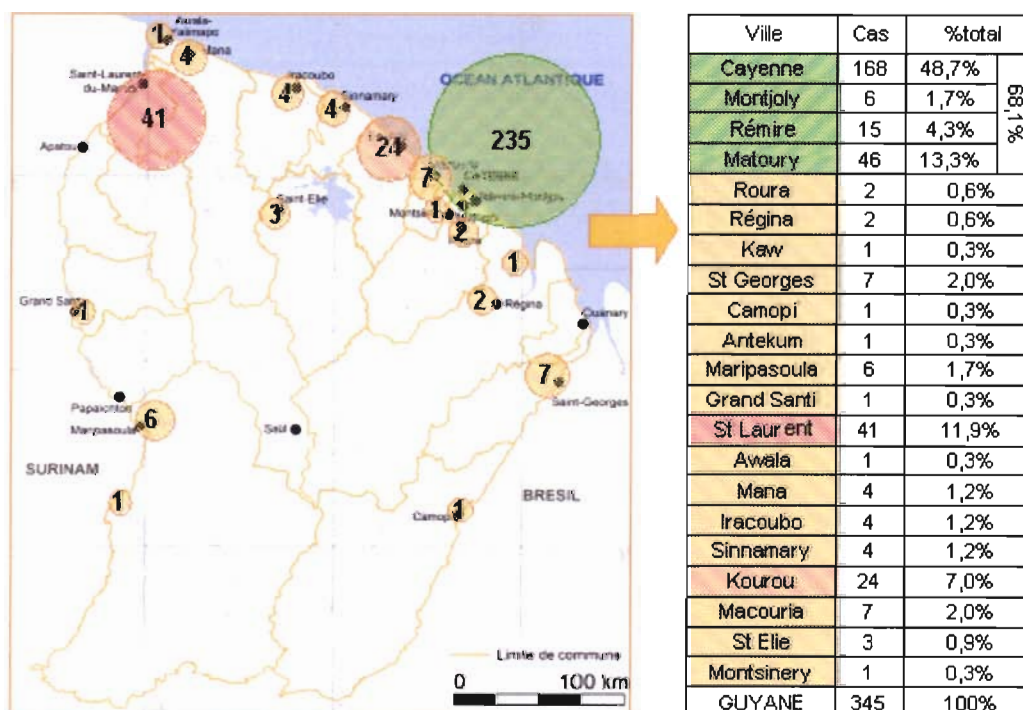


Figure 2.2 – Répartition géographique des cas de tuberculose confirmés bactériologiquement, de 1996 à 2003. Les données sont issues de l'enquête menée à l'Institut Pasteur de Guyane.

Parmi les 235 cas que regroupe l'île de Cayenne, 168 sont recensés à Cayenne (71,5%), contre 21 à Rémire-Montjoly (8,9%), et 46 à Matoury (19,6%). Ceci ne correspond pas tout à fait à la répartition de la population totale, dont 60% est basée à Cayenne, 18,5% à Rémire-Montjoly, et 21,5% à Matoury. Ce pourrait être lié au fait que la population de Rémire-Montjoly est globalement plus aisée que celle de Cayenne [Diman Antenor, 1992].

## 2.2.2. Répartition par âge et par sexe des cas de 1996-2003

Les données de population issues du recensement 1999 de l'INSEE sont détaillées par sexe et par tranches d'âge à l'échelle de la Guyane, ainsi qu'à l'échelle de l'île de Cayenne. Cependant, l'île de Cayenne n'étant pas représentative de l'ensemble de la population, nous avons choisi d'analyser les distributions par âge et sexe à l'échelle de la Guyane, ce qui implique que seuls les cas confirmés de tuberculose ont été pris en compte dans cette partie.

### 2.2.2.1. Répartition des cas par tranches d'âge

#### *Cas de tuberculose et population*

Pour cinq des 345 cas de tuberculose étudiés, l'âge du patient à la date du début du traitement est inconnu. La figure 2.3 montre la distribution de la population totale et des patients tuberculeux par tranches d'âge de cinq ans.

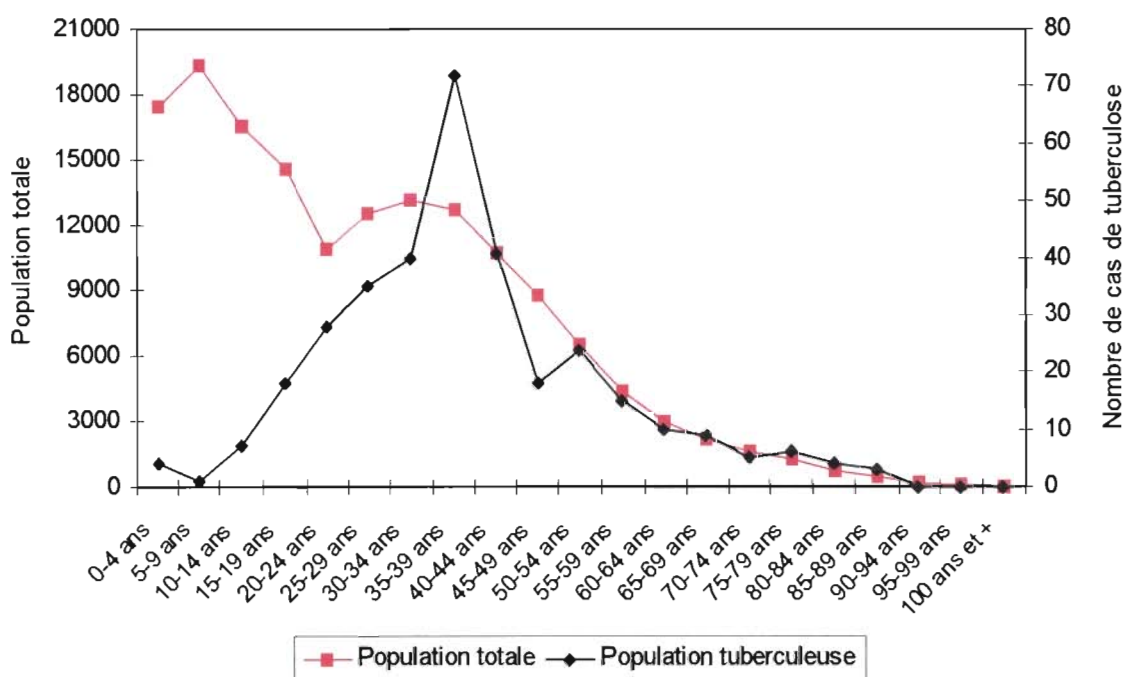


Figure 2.3 – Courbes de distribution des âges (a) de la population totale de Guyane au dernier recensement 1999 et (b) des cas de tuberculose confirmés recensés sur l'ensemble de la Guyane entre 1996 et 2003, par tranches de cinq ans.

La population de Guyane est une population jeune, avec 50 % de moins de 25 ans (cf. chapitre 1). En revanche, l'histogramme de distribution des âges des patients tuberculeux montre une répartition différente : le nombre de cas de tuberculose augmente avec l'âge, avec un maximum de cas pour la tranche d'âge 35-39 ans, puis le nombre de tuberculeux rediminue (figure 2.3). Ce résultat est cohérent avec la vaccination obligatoire des enfants avant 6 ans, et la possibilité de réactivation de la maladie après plusieurs années.

La moyenne d'âge des cas de tuberculose confirmés en Guyane sur la période 1996-2003 est de 38,9 ans. La répartition par tranches d'âge des 381 cas de tuberculose (confirmés et non confirmés) recensés sur l'île de Cayenne sur cette même période n'est pas présentée ici, mais la moyenne d'âge des cas, calculée à titre comparatif, est exactement la même.

L'âge moyen de la population de Guyane n'étant pas disponible, il n'est pas possible de faire une comparaison de moyennes. Nous avons donc choisi de comparer les distributions observées grâce au test de Kolmogorov-Smirnov. Le résultat montre que les deux distributions (figure 2.3) sont significativement différentes ( $D = 0,9048$  ;  $p < 0,0001$ ).

#### **Encadré 2.1 – Le test de Kolmogorov-Smirnov**

Le test de Kolmogorov-Smirnov est utilisé pour comparer deux fonctions de répartition empiriques [Dagnelie, 1975]. Il permet de vérifier que deux échantillons suivent la même loi, ou qu'un échantillon suit la même distribution que la population dont il est issu.

D'une façon plus précise, il s'agit de déterminer l'écart maximum, en valeur absolue, entre les deux fonctions que l'on cherche à comparer, soit  $|F'(x) - F(x)|$  aux points  $X_i$ , et l'on teste si la plus grande distance entre les fonctions de répartition est au-delà d'une valeur critique ou non [Dagnelie, 1975]. On rejette l'hypothèse nulle d'égalité des fonctions lorsque l'écart maximum observé est supérieur ou égal aux valeurs critiques qui sont données par des tables.

#### *Incidence par tranche d'âge*

Si la population la plus touchée en terme de nombre de cas de tuberculose est la population des 25-44 ans (54,5% des cas), en terme d'incidence le résultat obtenu est différent. En effet, la figure 2.4 ci-après montre que l'incidence de la tuberculose par classe d'âge suit une distribution bimodale avec un fort excès de cas par rapport à la population considérée parmi les jeunes adultes (avec une incidence maximale pour le groupe des 35-39 ans) et chez les plus de 80 ans. Ces résultats sont analysés dans la discussion (§2.6.2. et 2.6.3.)

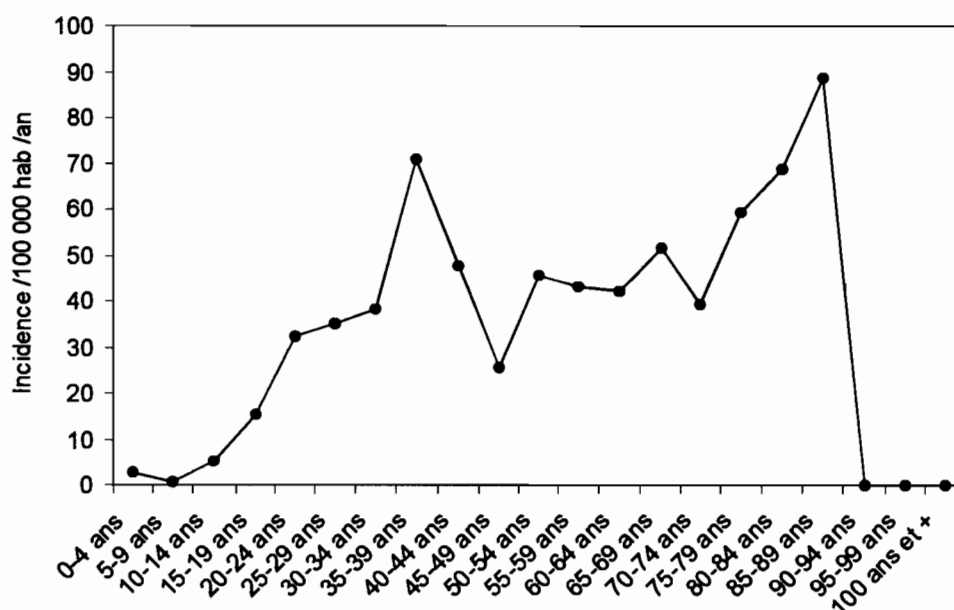


Figure 2.4 – Taux d'incidence de la tuberculose par tranche d'âge, calculé en rapportant, pour chaque tranche d'âge, le nombre de cas de tuberculose confirmés de 1996-2003 (soit 8 ans) à 100000 habitants. Ce rapport est ensuite divisé par huit pour obtenir une incidence annuelle.

### *Ratios standardisés d'incidence*

Les ratios d'incidence standardisés (*RIS*), correspondant au ratio du nombre de cas observés rapporté au nombre de cas attendus, ont été déterminés pour différentes tranches d'âges. Les effectifs attendus de tuberculose par classe d'âge ont été calculés grâce aux tailles de population par classe d'âge (issues du recensement INSEE de 1999) et au nombre total de cas de tuberculose, toutes classes confondues. Le tableau 2.6 récapitule les résultats de *RIS*.

Tableau 2.6 – Ratios standardisés d'incidence (*RIS*) de tuberculose calculés pour classes d'âge de 1996 à 2003 de la population de l'île de Cayenne.

	Attendu (E)	Observé (O)	RIS (O/E)	$p^a$	IC 95% <sup>b</sup>
0-14 ans	115,5	12	0,10	< 0,001	[0,05 - 0,18]
15-29 ans	82,2	82	1,00	0,981	[0,79 - 1,24]
30-34 ans	28,5	40	1,40	0,031	[1,00 - 1,91]
35-39 ans	27,6	71	2,58	< 0,001	[2,01 - 3,25]
40-44 ans	23,3	41	1,76	< 0,001	[1,26 - 2,39]
45-49 ans	18,9	18	0,95	0,831	[0,56 - 1,50]
50-54 ans	14,2	24	1,69	0,009	[1,08 - 2,51]
55-59 ans	9,4	15	1,59	0,069	[0,89 - 2,63]
60-74 ans	14,7	24	1,64	0,015	[1,05 - 2,44]
75 ans et +	5,7	13	2,27	< 0,001	[1,21 - 3,88]

<sup>a</sup>  $p$ , degré de significativité du test unilatéral du test de comparaison  $RIS > 1$ .

<sup>b</sup> IC 95%, limites inférieure et supérieure de l'intervalle de confiance à 95%.

Le calcul des ratios standardisés d'incidence montre qu'il existe une sur-incidence significative de tuberculose parmi trois classes d'âge, 35-39 ans, 40-44 ans, et 75 ans et plus ( $RIS > 1$  ;  $p < 0,001$ ), ainsi qu'une sous-incidence significative de tuberculose parmi les 0-14 ans ( $RIS < 1$  ;  $p < 0,001$ ).

### 2.2.2.2. Répartition des cas en fonction du sexe

Avec 78 963 hommes et 77 827 femmes recensés par l'INSEE en 1999, le ratio homme/femme est égal à 1,01 homme(s)/femme. Le ratio homme/femme des cas de tuberculose est en revanche déséquilibré vers les hommes, avec environ 2/3 des cas confirmés de sexe masculin (figure 2.5).

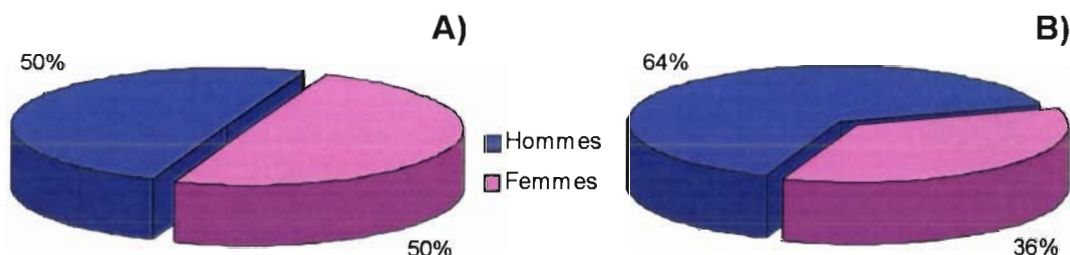


Figure 2.5 – Pourcentage homme/femme (A) de la population totale de Guyane au dernier recensement 1999 et (B) des cas confirmés de tuberculose 1996-2003 recensés en Guyane.

### 2.2.2.3. Répartition des cas par âge et par sexe

#### *Pyramide des âges*

Afin de tenir compte à la fois de l'âge et du sexe dans la comparaison de la distribution des cas (confirmés) de tuberculose par rapport à la population totale de Guyane, une pyramide des âges par sexe a été établie dans les deux populations (figure 2.6).

Le test de Kolmogorov-Smirnov montre que les distributions de cas de tuberculose par tranches d'âge en fonction du sexe (figure 2.6 - A) ne diffèrent pas significativement entre les hommes et les femmes au seuil alpha de 5% ( $D = 0,2178$  ;  $p = 0,0121$ ). Cependant, la probabilité associée au test étant à la limite du seuil de rejet de l'hypothèse nulle ( $H_0 =$  égalité des fonctions), nous avons également comparé l'âge moyen chez les cas de tuberculose homme et femme, qui est de 40,2 ans chez les hommes ( $\pm 15,9$ ) et 36,7 ans chez les femmes ( $\pm 15,8$ ). Cette différence est non significative ( $t = 1,9353$  ;  $df = 338$  ;  $p = 0,0538$ ).

Si l'on compare de nouveau la distribution des cas de tuberculose avec celle de la population grâce au test de Kolmogorov et Smirnov, mais en tenant cette fois-ci compte du sexe, le résultat est le même que précédemment, c'est-à-dire que les distributions par âge de la population totale d'une part, et des cas de tuberculose d'autre part, sont significativement différentes pour les deux sexes (hommes :  $D = 0,9208$  ;  $p < 0,0001$  et femmes :  $D = 0,9604$  ;  $p < 0,0001$ ).

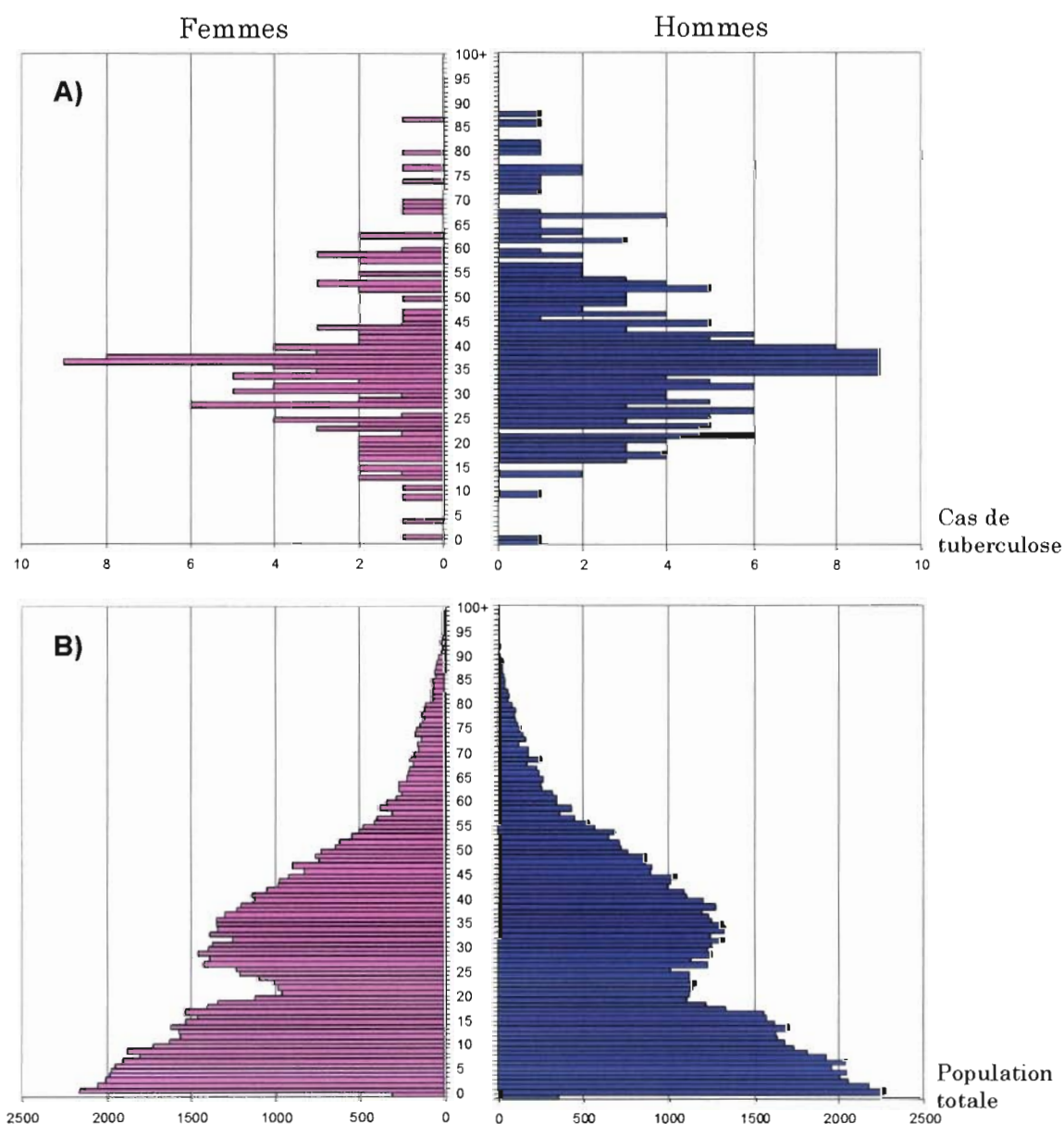


Figure 2.6 – Pyramide des âges (A) des cas confirmés de tuberculose recensés sur l'ensemble de la Guyane entre 1996 et 2003, et (B) de la population totale de Guyane d'après le dernier recensement de l'INSEE réalisé en 1999.

### *Ratios standardisés d'incidence*

Les ratios d'incidence standardisés ont été calculés par âge et par sexe, en prenant comme population de référence la population masculine de Guyane. Le *RIS* calculé est de 54,8% c'est-à-dire que le taux d'incidence de tuberculose chez les femmes est inférieur de 45,2% au taux attendu si l'on appliquait les taux d'incidence de tuberculose par âge de la population masculine à la structure par âge des femmes. On peut donc conclure que, après standardisation sur l'âge, le risque relatif de tuberculose dans la population féminine guyanaise, est presque deux fois inférieur par rapport à la population masculine de Guyane.

### Encadré 2.2 – Les ratios standardisés d'incidence

La comparaison directe de taux bruts entre différentes régions, et donc différentes populations, peut entraîner des conclusions erronées si les populations ont des structures d'âge différentes par exemple.

Le but de la standardisation, qu'elle soit directe ou indirecte, est de faire abstraction des modifications liées à la différence de structure d'âge (et/ou sexe, nationalité, etc.) des populations considérées, dans le but de les rendre comparables [Rumeau-Rouquette *et al.*, 1993]. Les taux d'incidence standardisés ne doivent pas être appréhendés comme des taux bruts puisqu'ils varient avec la structure d'âge choisie comme référence.

Le principe de la standardisation indirecte repose sur la comparaison du nombre total de cas observés dans la population étudiée avec le nombre de cas auquel on aurait pu s'attendre si l'incidence de la pathologie considérée dans cette population était la même que l'incidence de référence, soit la moyenne d'incidence par tranche d'âge dans la population de référence. On obtient ainsi un ratio d'incidence standardisé :

$$RIS = \frac{K}{\sum_{x=1}^g \left( E_x \times \frac{C_x}{R_x} \right)} \times 100$$

avec  $K$  le nombre de cas dans la population considérée,  $g$  le nombre de tranches d'âge considéré,  $E_x$  l'effectif dans la population étudiée,  $C_x$  le nombre de cas observé dans la population de référence, et  $R_x$  l'effectif dans la population de référence. Le  $RIS$  s'exprime en pourcentage. C'est une mesure du risque relatif de la population étudiée par rapport à la population que l'on a choisie pour référence.

### 2.2.3. Répartition des cas en fonction de la nationalité

Il existe probablement une hétérogénéité spatiale de la répartition des populations de différentes nationalités sur le territoire guyanais. Par exemple, de part la situation géographique de la ville de Saint Laurent du Maroni, située à la frontière du Surinam, il est probable que le nombre de cas de tuberculose d'origine surinamaïse recensé à Saint Laurent sera supérieur à celui de l'île de Cayenne.

Pour cette raison, la proportion de patients tuberculeux par nationalité est analysée sur l'ensemble des cas de Guyane, et pas uniquement de l'île de Cayenne. Ce sont donc les cas confirmés de l'ensemble de la Guyane qui sont ici pris en compte.

### 2.2.3.1. Prise en compte de la nationalité seule

#### *Critères d'inclusion des cas*

Dans les différents registres consultés au cours de nos enquêtes, la nationalité des patients n'était pas toujours indiquée, cette information n'étant pas demandée de façon systématique par les personnels soignant. Dans les registres de traitement de la LAT cependant, la nationalité des patients était très souvent précisée. Cette information est en effet utile dans la mesure où seuls les patients de nationalité française font l'objet d'une vaccination obligatoire au BCG, et où les antécédents de vaccination sont utiles au diagnostic, (cf. chapitre 1).

Il faut cependant noter que les enfants nés en Guyane de parents étrangers sont inclus dans les cas français, ce qui peut induire un biais lorsque l'on cherche à évaluer si le risque de tuberculose concerne davantage certaines nationalités que d'autres.

#### *Nationalités "autres"*

La figure 2.7 illustre la proportion de chaque nationalité parmi les cas de tuberculose confirmés entre 1996 et 2003. Pour plus de lisibilité, certaines nationalités en très faible effectif ont été regroupées sous la catégorie « autres ». Cette classe exclut les pays frontaliers ou proches de la Guyane, ainsi que les pays de la Caraïbe. On y trouve des cas originaires de Chine (4 cas), d'Europe (1 cas de Pologne) et d'Afrique (2 cas), ainsi que 2 cas marqués « autre », sans aucune précision, dans les registres de la LAT.

Dans cette catégorie, nous avons également inclus trois cas amérindiens. Ceci ne correspond en fait pas à une nationalité, mais cette population étant particulièrement surveillée, nous avons tout de même relevé cette précision, ce qui ne veut pas dire par ailleurs que d'autres cas amérindiens n'ont pas été inclus dans des cas de nationalité française.

#### *Résultats*

Si l'on exclut les 75 cas dont la nationalité est inconnue (cf. figure 2.7), les populations les plus touchées de Guyane en terme de tuberculose sont les haïtiens (34,0%), les français (31,7%), et les brésiliens (21,1%), puis, dans une moindre mesure, les guyaniens (originaires du Guyana) (6,0%) et les surinamais (3,4%). Pour comparaison, la proportion de ces nationalités dans la population totale est de 9% d'haïtiens, 70,3% de français, 4,6% de brésiliens, 1,5% de guyaniens, et 11,3% de surinamais [INSEE, 1999].

Au total, 69,8% des cas de tuberculose confirmés recensés entre 1996 et 2003 en Guyane sont de nationalité étrangère.

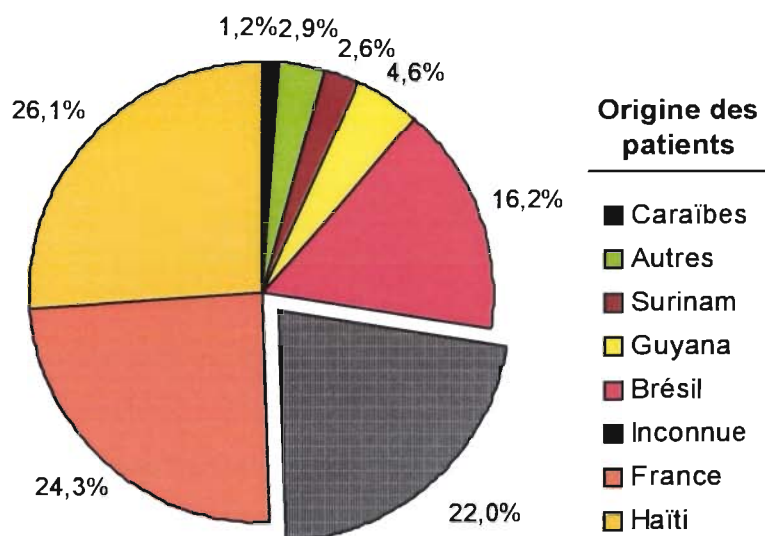


Figure 2.7 – Répartition des cas confirmés de tuberculose recensés en Guyane entre 1996 et 2003 en fonction de leur nationalité (en pourcentage du total de cas). Les patients d'origine « France » incluent les créoles, métropolitains, amérindiens, etc. ayant la nationalité française.

### *Taux d'incidence par nationalité*

Au recensement 1999, l'INSEE fournit un chiffre de 46 576 étrangers vivant en Guyane, contre 110 637 guyanais de nationalité française. Si on élimine les 75 cas pour lesquels on n'a pas la nationalité, le taux d'incidence annuel est de 9,5 cas pour 100 000 personnes de nationalité française, et de 49,7 cas pour 100 000 personnes de nationalité étrangère, sachant cependant que la population étrangère estimée est très certainement sous-évaluée en Guyane [Gardel, 2000].

### **2.2.3.2. Répartition des cas par nationalité et tranches d'âges**

#### *Taux d'incidence par tranches d'âge*

Les tranches d'âge les plus à risque de tuberculose sont susceptibles de varier en fonction de la nationalité des patients. Il se pourrait par exemple que, parmi les guyanais de nationalité française, les individus à risque soient les personnes âgées (réactivation de tuberculoses anciennes), et plutôt de jeunes adultes chez les guyanais d'origine étrangère.

Afin de vérifier cette hypothèse, les taux d'incidence de tuberculose par tranches âge ont été recalculés en tenant compte de l'origine des patients tuberculeux (figure 2.8). Les deux distributions observées (incidence par tranche d'âge chez les patients de nationalité française ou étrangère) sont significativement différentes ( $D = 0,7647$  ;  $p < 0,0001$ ).

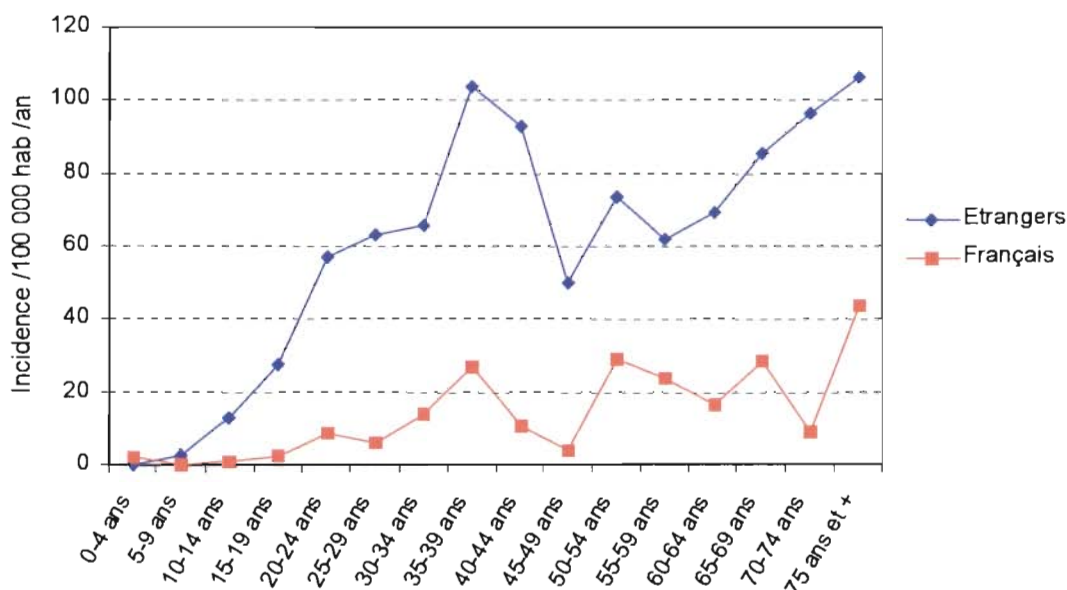


Figure 2.8 – Taux d'incidence moyen de la tuberculose par tranche d'âge (1) chez les gyanais d'origine française et (2) d'origine étrangère (exprimés respectivement en nombre de cas de tuberculose / 100 000 français / an ou / 100 000 étrangers / an).

### *Populations d'origine étrangère*

Dans les populations d'origine étrangère, la distribution bimodale relevée précédemment (§2.3.1.) est bien visible, avec les valeurs d'incidence les plus élevées pour le groupe des 35-39 ans (103,7 cas / 100 000 personnes d'origine étrangère / an) et les 75 ans et plus (106,5 cas / 100 000 / an).

### *Populations d'origine française*

Pour les populations de nationalité française, une structuration particulière de la courbe d'incidence est plus difficile à mettre en évidence. L'incidence la plus forte relevée concerne les 75 ans et plus (43,7 cas / 100 000 personnes d'origine française / an). Dans l'ordre décroissant d'incidence, on trouve ensuite le groupe des 50-54 ans (29,3 cas / 100 000 / an), le groupe des 65-69 ans (28,6 cas / 100 000 / an) et le groupe des 35-39 ans (27,3 cas / 100 000 / an).

### *Ratios standardisés d'incidence*

Les ratios d'incidence standardisés (*cf.* encadré 2.2) ont été calculés pour chaque tranche d'âge et par nationalité (nationalité française versus autres nationalités), en prenant comme population de référence la population de nationalité française de Guyane. Le *RIS* calculé est de 514% c'est-à-dire que le taux d'incidence de tuberculose dans la population de nationalité étrangère est 5 fois supérieur au taux attendu si l'on appliquait les taux d'incidence de

tuberculose par âge de la population française à la structure par âge des étrangers.

*Pourcentage des cas par tranches d'âge*

Le pourcentage de cas tuberculeux dans chaque tranche d'âge a été calculé en considérant indépendamment les cas de nationalité française et étrangère. Ceci permet de mettre en évidence la répartition des cas de tuberculose, non pas par rapport à la population totale considérée (quelle que soit la nationalité), mais par rapport à l'ensemble des cas d'une nationalité, puis de l'autre (figure 2.9).

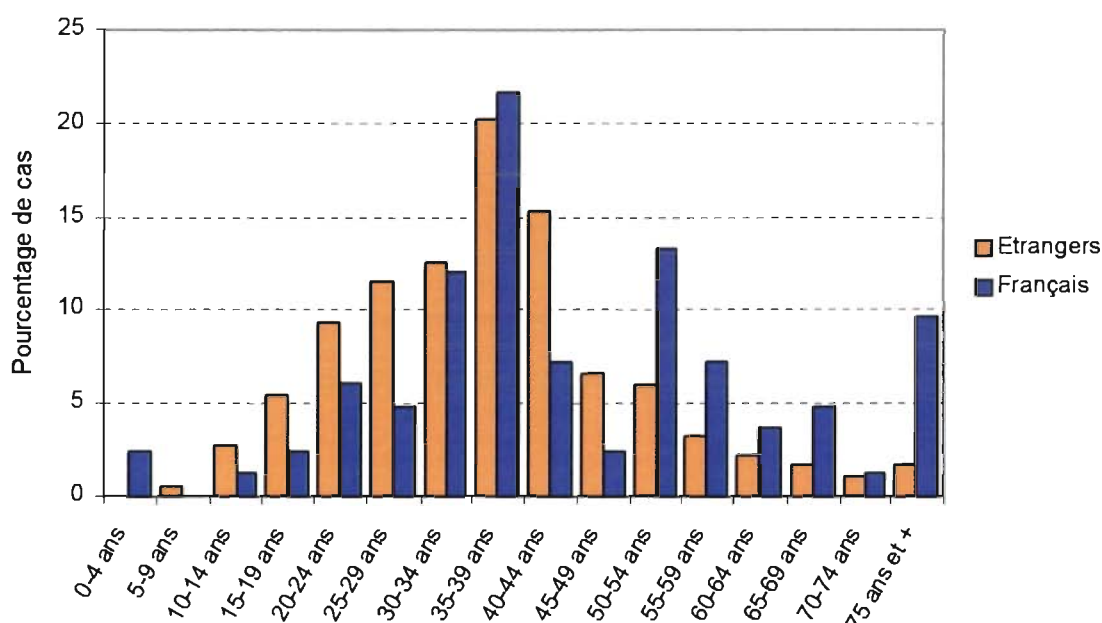


Figure 2.9 – Pourcentage de cas par tranche d'âge sur l'ensemble des cas de tuberculose (1) de nationalité française et (2) de nationalité étrangère, recensés entre 1996 et 2003.

Les deux distributions observées ne sont pas significativement différentes (test de Kolmogorov-Smirnov :  $D = 0,1875$  ;  $p = 0,9523$ ). Cependant, l'analyse plus poussée de la figure 2.9 montre qu'avant 50 ans, le pourcentage de tuberculeux de nationalité étrangère dans les tranches d'âge concernées est presque systématiquement supérieur au pourcentage de malades tuberculeux français, alors qu'après 50 ans la tendance s'inverse.

Chez les patients de nationalité étrangère, 84% des cas ont moins de 50 ans, contre seulement 60% de moins de 50 ans chez les patients de nationalité française. Les deux histogrammes mettent en évidence que chez les jeunes (enfants et adolescents) et les jeunes adultes, les individus touchés par la tuberculose sont plutôt de nationalité étrangère, alors que chez les plus de 50 ans, les individus touchés sont plus souvent des français.

#### **2.2.4. Le taux de co-infection tuberculose / VIH**

Le statut sérologique VIH n'était disponible que pour les cas de tuberculose recensés sur l'île de Cayenne, qu'ils soient confirmés bactériologiquement ou non. Sur les 381 cas de tuberculose concernés, 23,9% sont co-infectés par le VIH. Les 91 patients séropositifs sont plutôt des hommes (71,4%), de nationalité étrangère (74,6%), dont l'âge est compris entre 18 et 95 ans, 53% des cas ayant entre 35 et 55 ans.

Il n'est pas possible d'évaluer dans quelle mesure ces pourcentages sont représentatifs de l'ensemble de la Guyane. À titre comparatif, une précédente étude réalisée sur les cas de Guyane de 1990 à 2000 donnait un taux de co-infection VIH de 33% des cas [Léopold Aït-Ouada, 2002].

### **2.3. Discussion**

#### **2.3.1. Répartition par âge et par sexe**

En Guyane, l'incidence de tuberculose par classe d'âge suit une répartition bimodale avec un fort excès de cas parmi les jeunes adultes de sexe masculin, ce qui correspond à la situation générale observée en France [Decludt et Vaillant, 1997 ; Hubert *et al.*, 1995].

#### **2.3.2. Pic d'incidence chez les personnes âgées**

Le pic d'incidence chez les personnes âgées, observé quelle que soit la nationalité, pourrait refléter une exposition ancienne au bacille tuberculeux, et une réactivation du bacille à la faveur d'une altération de l'état général due au vieillissement.

#### **2.3.3. Pic d'incidence chez les jeunes adultes**

Le pic d'incidence chez les jeunes adultes, plutôt marqué chez les patients de nationalité étrangère, s'expliquerait par l'association étroite des infections par le VIH et le bacille tuberculeux. En France, cette cause est la seule identifiée mais n'expliquerait que le tiers de l'excès de cas observé chez les jeunes adultes [Hubert *et al.*, 1995].

En Guyane, il est probable qu'en plus de la co-infection VIH, s'ajoutent des comportements à risque tels que la drogue ou la prostitution, même si nous avons noté précédemment qu'en Guyane la toxicomanie ne constitue pas un facteur de risque majeur pour l'infection VIH [CNS, 1996] (chapitre 1 §2.3.5.).

En effet, sur les 381 cas de tuberculose de l'île de Cayenne, au moins 15 cas sont des sans domicile fixe (SDF), dont 4 sont notés dans les registres de la LAT comme étant suivis par le Relais Drogue Solidarité, et 8 cas sur 381 sont incarcérés à la prison de Rémire.

#### **2.3.4. Patients de nationalité étrangère**

Il serait important de pouvoir identifier dans quelle mesure les malades tuberculeux de nationalité étrangère sont des cas d'importation ou des cas contaminés localement. Ce point sera étudié dans le chapitre 5 du présent mémoire, au travers des analyses moléculaires des souches isolées en Guyane entre 1996 et 2003.

#### **2.3.5. Co-infections tuberculose / VIH**

L'étude menée par Léopold Ait-Ouada (2002) révèle que les formes extra-pulmonaires de tuberculose sont rencontrées presque deux fois plus souvent chez les patients séropositifs pour le VIH. Les formes extra-pulmonaires étant les plus difficiles à diagnostiquer, il pourrait être intéressant d'envisager un dépistage systématique de tuberculose extra-pulmonaire, avec des analyses bactériologiques et histologiques plus poussées, chez l'ensemble des patients séropositifs au VIH.

### **2.4. Bilan des facteurs de risque humains**

En conclusion, cette analyse fait apparaître qu'en Guyane, les individus les plus exposés au risque tuberculeux sont les personnes âgées de plus de 75 ans, les adultes jeunes de sexe masculin et les personnes de nationalité étrangère. Ce risque augmente avec la séropositivité au VIH. L'incidence de la tuberculose est très hétérogène selon les caractéristiques socio-démographiques des populations et ces données sont prises en compte dans la prise en charge de la maladie en Guyane. Elles sont cependant insuffisantes pour affiner les stratégies de lutte, qui doivent également prendre en compte la dynamique locale de transmission de la maladie.

Dans la perspective d'une évaluation de nouvelles stratégies de lutte à mettre en place, nous avons d'abord voulu estimer l'incidence théorique espérée de la tuberculose sur l'île de Cayenne et en Guyane. Les résultats obtenus sont présentés ci-après.

### 3. Évaluation statistique de l'incidence actuelle de la tuberculose en Guyane

Après la mise en évidence d'une sous-déclaration de la tuberculose sur l'île de Cayenne, et l'analyse descriptive de l'épidémiologie de la maladie, il apparaît important d'évaluer dans quelle mesure le recensement réalisé au sein de trois systèmes de prise en charge de la tuberculose (IPG, DIM et LAT) est représentatif de l'ensemble des cas de tuberculose. Plus précisément, l'objet de cette partie est d'évaluer la proportion de cas non recensés après enquête auprès de l'IPG, du DIM et de la LAT, et qui sont donc à priori non traités et susceptibles de transmettre la maladie. Cette analyse a été réalisée grâce à la méthode de capture-recapture<sup>4</sup>.

#### 3.1. La méthode de capture-recapture

##### 3.1.1. Historique de la méthode

À l'origine, cette méthode a été introduite par les écologues comme moyen d'estimer la taille de populations animales lorsqu'un comptage exhaustif n'est pas réalisable [Chapman, 1954 ; Cormack, 1968].

Cette méthode d'estimation est basée sur le protocole suivant : dans la population dont on cherche à estimer la taille, un certain nombre d'animaux sont capturés une première fois, marqués, puis libérés sur leur lieu de capture. Puis, une seconde capture est effectuée dans la même zone. La proportion d'animaux marqués présents parmi les individus capturés lors de ce second piégeage (donc re-capturés) est comptabilisée (figure 2.10).

La population totale (N) estimée est calculée sur la base d'un produit en croix qui prend en compte le nombre d'animaux capturés dans le 1<sup>er</sup> échantillon (N<sub>1</sub>), le nombre d'animaux capturés dans le 2<sup>nd</sup> échantillon (N<sub>2</sub>) et le nombre d'animaux re-capturés (donc marqués) présents dans le 2<sup>nd</sup> échantillon (n<sub>2</sub>) :

$$(n_2 / N_2) \approx (N_1 / N)$$

d'où l'estimation de

$$N \approx (N_1 * N_2) / n_2$$

Cette méthode de capture-recapture a, par la suite, été appliquée aux analyses démographiques afin de compenser le sous-dénombrement dans les recensements, et d'estimer les taux de natalité et de mortalité dans les pays en développement [Sekar et Deming, 1949]. Elle est actuellement utilisée dans les études épidémiologiques.

---

<sup>4</sup> Cette partie a fait l'objet d'une publication : Guernier V., Guégan, J-F., Deparis X. An evaluation of the actual incidence of tuberculosis in French Guiana using a capture-recapture model (accepté le 8 septembre 2005 dans *Microbes and Infection*). Le manuscrit tel qu'il a été resoumis est lisible en Annexe 2.

### 3.1.2. Application en épidémiologie

La méthode de capture-recapture est utilisée depuis plus de 50 ans en épidémiologie pour améliorer les estimations d'incidence et de mortalité des maladies [Wittes et Sidel, 1968 ; Wittes *et al.*, 1974 ; Chao, 1998]. Son utilisation s'est, par ailleurs, particulièrement accrue ces 20 dernières années [Hook et Regal, 1995a ; Bernillon *et al.*, 1997 ; Deparis *et al.*, 1997 ; Nanan et White, 1997 ; Tocque *et al.*, 2001].

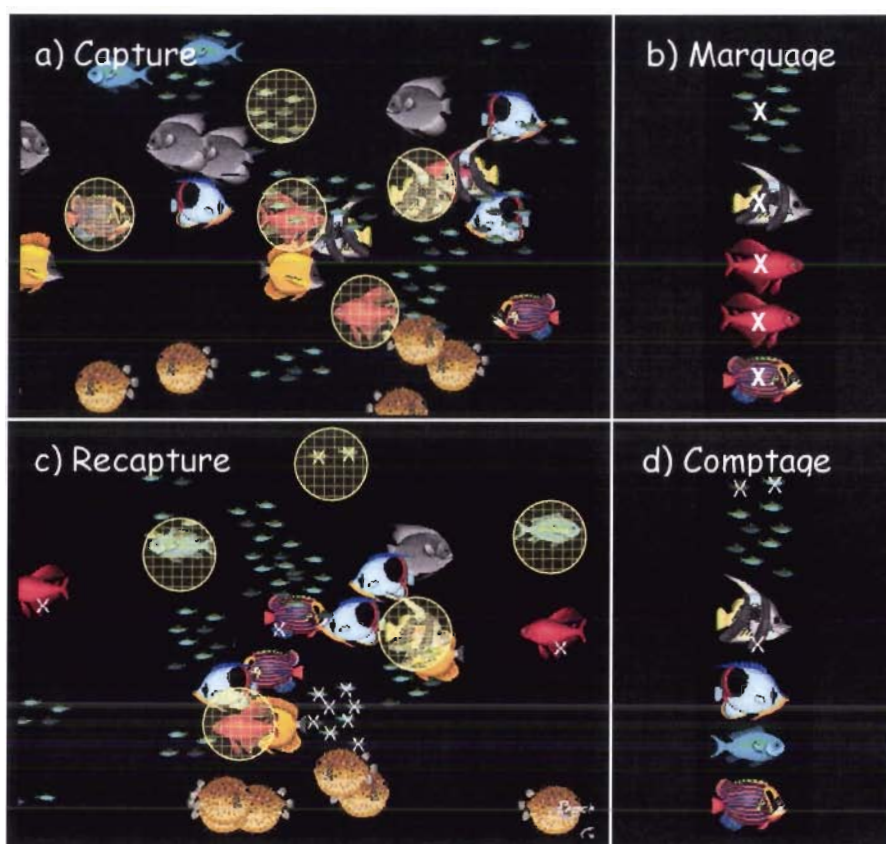


Figure 2.10 – Principe de la méthode de capture-recapture : (a) capture d'un premier échantillon N1, (b) marquage (symbolisé par des croix) des individus capturés, qui sont ensuite relâchés, (c) capture d'un second échantillon N2, et (d) comptage des individus marqués n2.

En épidémiologie, le fait d'être « capturé dans un échantillon » revient à « figurer sur une liste ». Ces listes renvoient aux organismes permettant de recenser des malades. Les bases de données systématiquement recueillies, comme les registres de maladies, les données sur les sorties des hôpitaux, les certificats de décès et les ordonnances, peuvent être utilisées comme sources. Les sources d'information sont choisies en fonction de la population dont on cherche à évaluer la taille (cas d'une certaine maladie, usagers de drogues, etc.), l'important étant de pouvoir recouper les individus entre les sources.

### 3.1.3. Principe et conditions d'application

Les estimations de populations par capture-recapture consistent à croiser les informations d'au moins deux bases de données afin d'identifier le nombre d'individus communs aux listes prises deux par deux (cas appariés), et d'utiliser ces recouvrements dans des analyses statistiques pour évaluer le nombre d'individus manquants et la taille de la population totale [Darroch, 1958 ; Seber, 1982 ; Nichols et Pollock, 1983 ; Magurran, 1988 ; Mathematical Sciences Education Board, 1993].

Quatre conditions sous-tendent l'analyse par capture-recapture [IWGDMF, 1995 ; Tocque *et al.*, 2001 ; Ding et Feinberg, 1996] :

- (1) L'homogénéité de capture des cas : tous les cas étudiés doivent avoir une probabilité identique d'être identifiés par une source. La notification des cas dans une source ne doit donc pas être liée à des variables caractérisant les cas comme l'âge, le sexe, le lieu de résidence ou la gravité de maladie. Cependant, cette probabilité peut être différente entre les sources.
- (2) Les cas ne doivent apparaître qu'une seule fois dans chaque source donc tous les vrais cas communs doivent être identifiés. Le critère d'identification des doublons doit être suffisamment discriminant.
- (3) L'indépendance des sources entre elles : la probabilité qu'un cas soit recensé dans une source ne doit pas dépendre de la probabilité qu'il soit recensé dans une autre source.
- (4) La population doit être close, c'est-à-dire qu'il ne doit pas y avoir de natalité, mortalité, ou migration importante pendant la durée de l'étude.

Ces conditions, si elles ne sont pas respectées, peuvent biaiser la validité et la précision des estimations. Le respect de ces conditions dans notre propre étude est donc discuté au §3.2.2. ci-après.

## 3.2. Matériel et méthodes

### 3.2.1. Les sources de données

Dans le cas présent, les trois systèmes de prise en charge de la tuberculose détaillés au §1.2.1., c'est-à-dire le Département d'Information Médicale (DIM) de l'hôpital de Cayenne, le service de Lutte Anti-Tuberculeuse (LAT) et l'Institut Pasteur de Guyane (IPG) constituent des sources de données idéales pour notre analyse de capture-recapture, à condition que les clauses de validité de la méthode pré-citées soient respectées. Cette analyse, dont les résultats suivent (§3.3.), concerne donc uniquement les cas de tuberculose recensés entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003 sur l'île de Cayenne.

### **3.2.2. Respect des conditions de validité**

#### *Homogénéité de capture des cas*

Dans la mesure où notre travail d'enquête s'est limité à l'île de Cayenne, le lieu de résidence des patients n'a pas d'influence sur leur recensement dans les trois sources considérées puisqu'il n'existe sur l'île de Cayenne qu'un hôpital, un laboratoire de référence, et un dispensaire. En revanche, il est difficile d'affirmer avec certitude que la notification des cas au DIM, à la LAT et à l'IPG, ne dépend pas de caractéristiques du malade, même si, sur la base de nos connaissances, cela semble être le cas.

#### *Identification des vrais cas communs*

Les cas inclus dans nos données ayant été soumis à nos propres critères de jugement lors des enquêtes, nous pouvons être sûrs que la même définition de ce qu'est un cas de tuberculose a bien été utilisée dans les trois sources.

Il est conseillé dans la littérature d'utiliser un unique identifiant (le nom par exemple) pour trouver les cas communs aux trois sources [Chao *et al.*, 2001]. Mais étant donné les erreurs possibles (archives papier, erreurs d'orthographe des noms, contradictions sur les dates de naissance), nous avons préféré la combinaison de plusieurs critères. Cependant, même s'il ne s'agit pas d'un identifiant unique, les critères que nous avons choisis sont suffisamment discriminants pour limiter les faux positifs ou faux négatifs (*cf.* §1.2.3. pour les détails concernant la méthode d'appariement des cas).

#### *Population close*

Avec une croissance annuelle de 3,6% par an, la population de Guyane ne peut pas être considérée comme close sur la période 1996-2003, ce qui constitue une limite majeure à notre analyse.

#### *Indépendance des sources*

L'hypothèse d'indépendance des sources est la plus difficile à respecter en épidémiologie, et en particulier dans notre étude. La majorité des cas recensés à l'IPG et à la LAT viennent de l'hôpital de Cayenne, donc on peut supposer une dépendance entre DIM et IPG et entre DIM et LAT, et une indépendance entre IPG et LAT.

Au final, seules les deux premières conditions semblent bien respectées. Le problème de non indépendance entre sources est cependant résolu par le fait que l'une des méthodes de capture-recapture choisies, qui nécessite au moins trois sources de données, permet d'intégrer dans les modèles statistiques les relations de dépendance entre sources prises deux à deux (les différents modèles de cette méthode sont détaillés ci-après §3.2.3.).

La condition d'une population close et stable n'est jamais satisfaite. Nous avons cependant relevé plusieurs études épidémiologiques appliquant la méthode de capture-recapture à des données pluriannuelles, avec une probable augmentation de la population, sans que cela semble remettre en cause les résultats [Gjini *et al.*, 2004 ; Iñigo *et al.*, 2003]. D'autre part, aucune étude n'a jusqu'à présent évalué dans quelle mesure le non-respect de cette condition a un impact sur les résultats d'analyse de capture-recapture.

### 3.2.3. Les modèles utilisés

Un logiciel spécifique des modèles de capture-recapture a été utilisé afin d'estimer la taille de la population guyanaise infectée par la tuberculose : CARE-1 (pour CAPture-REcapture) [Chao *et al.*, 2001] développé pour un environnement S-Plus [MathSoft, 1997]. Le logiciel est téléchargeable gratuitement à l'adresse <http://chao.stat.nthu.edu.tw/softwareCE.html>. Les trois méthodes statistiques intégrées dans ce logiciel, ont été testées : (1) les estimateurs de Petersen et Chapman, (2) les modèles log-linéaires et (3) l'approche « *sample coverage* » (défini ici comme le « taux de couverture », aucun terme français correspondant n'ayant été trouvé dans la littérature).

#### 3.2.3.1. Les estimateurs de Petersen et Chapman

L'estimateur de Petersen a été corrigé par Chapman pour les petits échantillons. Ces estimateurs, basés sur des échantillons appariés, sont utilisés sous la condition d'indépendance des sources. Ce n'est probablement pas le cas ici, mais ces estimateurs sont utiles pour une analyse préliminaire.

#### Encadré 2.3 – Estimateurs de Petersen et Chapman

Les estimateurs de Petersen et Chapman sont les modèles les plus simples d'estimation d'une population par capture-recapture. Ils sont définis par les formules suivantes [Chapman et Junge, 1956] :

$$\text{Estimateur de Petersen : } \hat{N}_p = \frac{n_1 \times n_2}{m_2}$$

$$\text{Estimateur de Chapman : } \hat{N}_c = \frac{(n_1 + 1) \times (n_2 + 1)}{(m_2 + 1)} - 1$$

où  $\hat{N}_p$  et  $\hat{N}_c$  correspondent à la population totale estimée,  
 $n_1$  est le nombre d'individus capturés dans le 1<sup>er</sup> échantillon et marqués,  
 $n_2$  est le nombre d'individus capturés dans le 2<sup>nd</sup> échantillon,  
 $m_2$  est le nombre d'individus marqués présents dans le 2<sup>nd</sup> échantillon.

### 3.2.3.2. Les modèles log-linéaires

Les modèles log-linéaires sont basés sur une table de contingence de  $2^k$  cellules. Ils permettent d'analyser des interactions entre variables en modélisant les relations deux à deux (et donc la dépendance) entre les sources. Chaque interaction correspond à un modèle. Le nombre de cas qui n'ont été identifiés par aucune des sources (les « cas manquants ») est estimé à partir du modèle le plus satisfaisant par rapport aux données, ou en calculant une combinaison de différents modèles, pondérés les uns par rapport aux autres [Fienberg, 1972 ; Hook et Regal, 1995b].

#### Encadré 2.4 – Les modèles log-linéaires

Les modèles log-linéaires permettent d'analyser l'association entre variables catégorielles (ou qualitatives) au travers d'une table de contingence [Christensen, 1997]. La table de contingence entre trois variables possédant chacune deux modalités (oui ou non) peut être définie comme suit :

		Variable 1			
		oui		non	
		Variable 2			
		oui	non	oui	non
Variable 3	oui	$Z_{111}$	$Z_{101}$	$Z_{011}$	$Z_{001}$
	non	$Z_{110}$	$Z_{100}$	$Z_{010}$	$Z_{000}$

L'effectif estimé d'une cellule du tableau de contingence correspond au produit  $(n_i \times n_j) / n_{ij}$  avec  $n_i$  l'effectif total de la ligne  $i$ ,  $n_j$  l'effectif total de la colonne  $j$ , et  $n_{ij}$  l'effectif total observé du tableau de contingence. Le passage aux logarithmes transforme ce modèle multiplicatif en modèle additif. Le modèle général correspondant à 3 variables  $i, j$  et  $k$  s'écrit :

$$\log E(Z_{ijk}) = u + u_1 I(i=1) + u_2 I(j=1) + u_3 I(k=1) + u_{12} I(i=j=1) + u_{13} I(i=k=1) + u_{23} I(j=k=1) + u_{123} I(i=j=k=1)$$

Dans le détail, ceci nous donne les 7 modèles suivants :

Interaction d'ordre 1	$\log E(Z_{001}) = u + u_3$
	$\log E(Z_{010}) = u + u_2$
	$\log E(Z_{100}) = u + u_1$
Interaction d'ordre 2	$\log E(Z_{110}) = u + u_1 + u_2 + u_{12}$
	$\log E(Z_{101}) = u + u_1 + u_3 + u_{13}$
	$\log E(Z_{011}) = u + u_2 + u_3 + u_{23}$
Interaction d'ordre 3	$\log E(Z_{111}) = u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23} + u_{123}$

Les 13 modèles inclus dans le logiciel CARE-1 [Chao *et al.*, 2001] sont :

Modèle d'indépendance :  $u_{12} = u_{13} = u_{23} = 0$

Modèle 12/3 (interaction entre 1 et 2) :  $u_{13} = u_{23} = 0$

Modèle 13/2 (interaction entre 1 et 3) :  $u_{12} = u_{23} = 0$

Modèle 23/1 (interaction entre 2 et 3) :  $u_{12} = u_{13} = 0$

Modèle 12/13 (2 interactions, entre 1 et 2, et entre 1 et 3) :  $u_{23} = 0$

Modèle 12/23 (2 interactions, entre 1 et 2, et entre 2 et 3) :  $u_{13} = 0$

Modèle 13/23 (2 interactions, entre 1 et 3, et entre 2 et 3) :  $u_{12} = 0$

Modèle symétrique :  $u_1 = u_2 = u_3$  et  $u_{12} = u_{13} = u_{23}$

Modèle quasi-symétrique :  $u_{12} = u_{13} = u_{23}$

Modèle partiel quasi-symétrique (part-qs1) :  $u_{12} = u_{23}$

Modèle partiel quasi-symétrique (part-qs2) :  $u_{12} = u_{13}$

Modèle partiel quasi-symétrique (part-qs3) :  $u_{13} = u_{23}$

Modèle saturé : intègre toutes les relations de dépendance

### 3.2.3.3. L'approche « *sample coverage* »

L'approche *sample coverage* intègre le biais de corrélation d'un modèle à trois échantillons en tenant compte du taux de capture des sources, et des dépendances entre ces sources. Cette approche est utilisée pour mesurer le taux de recouvrement (overlap fraction) entre échantillons.

Trois types d'estimateurs de la taille de population totale sont calculés avec CARE-1 :  $\hat{N}_0$  pour des échantillons indépendants ;  $\hat{N}$  pour une couverture de l'échantillon suffisante (c'est-à-dire lorsque l'erreur standard (*se*) se rapportant au calcul est inférieure au tiers de l'estimation (*est*) de la population) et  $\hat{N}_1$  pour un taux de capture faible (beaucoup de singletons, c'est-à-dire de cas présents dans une unique source) [Chao *et al.*, 2001].

#### Encadré 2.5 – L'approche *sample coverage*

Soit une table de contingence concernant des patients inclus dans trois sources. L'approche *sample coverage* est basée sur l'estimation de la population totale via la relation entre la taille de population et le « taux de couverture » de l'échantillon, qui tient compte de la proportion de cas recensés dans une ou plusieurs sources.

Les relations de dépendance entre sources sont modélisées par les coefficients de covariation (CCV). Le calcul des CCV intègre la probabilité de rencontrer l'individu *i* dans la source *j*, et la probabilité de rencontrer ce même individu dans une autre source, *k*. Si la covariance entre ces deux probabilités est nulle, les deux sources sont indépendantes.

Le taux de couverture  $\hat{C}$  d'un échantillon correspond à la moyenne de la proportion de cas trouvés dans plus d'une source [Chao et Tsay, 1998] :

$$\hat{C} = 1 - \frac{1}{3} \times \left( \frac{Z_{100}}{n_1} + \frac{Z_{010}}{n_2} + \frac{Z_{001}}{n_3} \right)$$

Le taux de recouvrement  $D$  est défini par le nombre moyen d'individus présents dans deux sources mais pas la troisième :

$$D = M - \frac{1}{3} \times (Z_{100} + Z_{010} + Z_{001})$$

avec  $M$  le nombre total de cas identifiés par combinaison des 3 sources. Les estimations de la population totale sont calculées en fonction des paramètres  $\hat{C}$  et  $D$  ( $\hat{N}_0 = D/\hat{C}$  ; cf. [Chao *et al.*, 2001] pour le détail de  $\hat{N}$  et  $\hat{N}_1$ ). Pour chaque estimateur ( $\hat{N}_0$ ,  $\hat{N}$  et  $\hat{N}_1$ ), l'intervalle de confiance à 95% correspondant (IC<sub>95</sub>) est calculé grâce à une transformation logarithmique, ce qui permet d'obtenir une limite inférieure de l'intervalle de confiance qui soit toujours supérieure à la valeur observée du nombre de cas [Chao, 1987]. L'erreur standard associée (*se*) est calculée par une méthode de bootstrap, basée sur 1000 réplifications.

### 3.3. Résultats

#### 3.3.1. Simple croisement des trois sources de données

Le croisement des cas de tuberculose de trois sources (IPG, DIM et LAT) après identification des doublons et triplets a été fait précédemment (§1.2.4.), et donne un total de 381 cas de tuberculose recensés sur l'île de Cayenne entre 1996 et 2003. Rappelons également que l'IPG, le DIM et la LAT recensent respectivement 61,7%, 74,3% et 71,1% des cas. Le nombre de cas de tuberculose identifiés dans chacune des sources et le nombre cumulé de cas identifiés dans une, deux ou trois sources de données a été détaillé dans la figure 2.1.

#### 3.3.2. Résultats de capture-recapture

##### 3.3.2.1. Méthode de Petersen et Chapman

Les estimateurs de Petersen et Chapman donnent une fourchette de 365 à 406 cas (tableau 2.7). L'estimateur de Petersen sous-estime la taille réelle de la population lorsque deux sources sont positivement dépendantes, et la sur-estime en cas de dépendance négative. Le nombre de cas obtenu par simple

croisement étant de 381 cas, on en déduit une dépendance positive entre les sources 1 et 2, et 2 et 3 ( $\hat{N}_p(1,2)$  et  $\hat{N}_p(2,3) < 381$ ) ; ces relations de dépendance entre échantillons sont confirmées dans la méthode *sample coverage* ci-après.

Tableau 2.7 – Estimations du nombre de cas de tuberculose totaux basées sur les estimateurs de Petersen et Chapman, les échantillons étant pris deux à deux.

	Petersen	Chapman	<i>se</i> <sup>a</sup>	<i>cil</i> <sup>b</sup>	<i>ciu</i> <sup>b</sup>
pair (1,2) <sup>d</sup>	365	365	8	354	384
pair (1,3)	406	405	12	386	434
pair (2,3)	370	370	6	361	387

<sup>a</sup> *se*, erreur standard.

<sup>b</sup> *cil*, limite inférieure de l'intervalle de confiance à 95% (IC<sub>95</sub>) ; *ciu*, limite supérieure.

<sup>d</sup> Les chiffres se réfèrent aux sources appariées (1, IPG ; 2, DIM ; 3, LAT).

### 3.3.2.2. Modèles log-linéaires

Le résultat de la modélisation log-linéaire sous CARE-1 est résumé sur le tableau 2.8. Le choix du meilleur modèle, qui dépend de la qualité d'ajustement (*goodness-of-fit*) des modèles aux données observées, est expliqué ci-après.

Tableau 2.8 – Estimations des cas de tuberculose totaux dans les modèles log-linéaires à trois sources. De haut en bas : modèle d'indépendance ; modèles à 1 ou 2 interactions avec indépendance relative (modèles 13/2 à 23/13) ; modèles tenant compte de l'hétérogénéité entre individus (quasi-symétrique et partiels quasi-symétriques) ; modèle saturé.

	<i>dev</i> <sup>a</sup>	<i>ddl</i>	<i>est</i>	<i>se</i> (±)	<i>cil</i>	<i>ciu</i>
indépendant	39,75	3	395	4	389	407
13/2 <sup>b</sup>	38,77	2	393	4	387	406
23/1	27,47	2	406	7	395	424
12/3	28,27	2	403	6	394	419
12/23	1,84	1	462	28	423	536
12/13	28,20	1	402	7	392	421
23/13	27,47	1	406	9	394	430
symétrique	32,75	4	463	34	419	560
quasi-sym	17,28	2	464	34	419	561
part-qs1	0,32	1	495	49	431	638
part-qs2	11,37	1	483	43	427	606
part-qs3	15,03	1	460	33	417	555
saturé	0,00	0	498	51	433	647

<sup>a</sup> *dev*, déviance ; *ddl*, degré de liberté ; *est*, estimation de la population totale ; *se* (±), erreur standard ; *cil*, limite inférieure de IC<sub>95</sub> ; *ciu*, limite supérieure de IC<sub>95</sub>.

<sup>b</sup> 1, IPG ; 2, DIM ; 3, LAT ; en cas de dépendance entre les sources 1 et 2, le modèle est noté 12/3. Si les sources 1 et 3 sont également dépendantes, le modèle est noté 12/13, et ainsi de suite pour les autres interactions (pour les autres modèles, cf. encadré 2.3).

### *Choix du meilleur modèle*

La qualité d'ajustement des modèles aux données observées est mesurée par la déviance (*dev*), qui doit être la plus faible possible. Mais l'erreur standard doit également être faible et le modèle le moins complexe possible. Le modèle saturé qui intègre toutes les relations de dépendance, est le plus complexe et sa déviance est toujours la plus faible. Entre deux modèles, on choisit toujours le modèle le plus simple, dans la mesure où sa qualité d'ajustement est la même que le modèle plus complexe.

### *Modèle log-linéaire retenu*

Les deux modèles les plus adéquats sont 12/23 et part-qs1, mais celui retenu comme étant le meilleur est le modèle 12/23 (erreur standard inférieure et modèle plus simple), modèle qui tient compte d'une dépendance entre les sources 1 et 2, et entre 2 et 3. L'estimation obtenue avec ce modèle est 462 cas, avec un intervalle de confiance à 95% (IC<sub>95</sub>) de [423-536]. Les relations de dépendance entre 1 et 2 (IPG et DIM) et entre 2 et 3 (DIM et LAT) étaient fortement intuitives (*cf.* §3.2.2.). La « sensibilité » d'un système de prise en charge est évaluée par le rapport du nombre de cas observé après enquête, sur le nombre de cas théorique calculé par capture-recapture. Avec ce modèle, les « sensibilités » de capture du DIM, de la LAT et de l'IPG sont respectivement de 61,3%, 58,7% et 50,9% (le nombre de cas théorique étant 462).

### **3.3.2.3. L'approche « *sample coverage* »**

#### *Choix de l'estimateur*

Pour choisir entre les trois estimateurs  $\hat{N}_0$ ,  $\hat{N}$  et  $\hat{N}_1$ , il faut prendre en compte l'erreur standard (*se*), qui doit être inférieure au tiers de l'estimation [Chao *et al.*, 2001], ce qui est le cas ici. Le taux de couverture estimé, de 85,9% ( $\hat{C} = 0,859$ ), peut être considéré comme suffisamment élevé pour corriger le sous-recensement des cas, et l'estimateur  $\hat{N}$  est utilisé.

#### *Résultats*

Le calcul de *M* correspond au simple croisement des trois sources (§3.3.1.). Le taux moyen de recouvrement des cas, c'est-à-dire le nombre de cas présents dans deux listes à l'exclusion de la troisième, est égal à  $D = 344$  (tableau 2.9). L'estimateur  $\hat{N}$  donne une population totale de cas de tuberculose sur l'île de Cayenne de 425 cas, avec un IC<sub>95</sub> de [407-453] (tableau 2.9). Les données des trois sources considérées sont donc suffisantes pour fournir une estimation fiable de la taille de la population tuberculeuse qui échappe aux trois systèmes de surveillance épidémiologique et de soins évalués. Avec l'estimateur  $\hat{N}$ , les « sensibilités » des trois systèmes de prise en charge de la tuberculose (sont

respectivement de 66,6% pour le DIM, de 63,8% pour la LAT, et de 55,3% pour l'IPG. Ces valeurs sont discutées plus loin (§3.4.)

Tableau 2.9 – Résultats obtenus avec l'approche *sample coverage*. Les trois estimateurs calculés tiennent compte de l'hétérogénéité de redondance de l'information entre les trois échantillons (ou sources) considéré(e)s.

	$M^a$	$D$	$\hat{C}$	$est$	$se (\pm)$	$cil$	$ciu$
$\hat{N}_0^d$	381	344	0,859	401	6	392	415
$\hat{N}$	381	344	0,859	425	11	407	453
$\hat{N}_1$	381	344	0,859	419	9	405	441

	$u_1^b$	$u_2$	$u_3$	$r_{12}^c$	$r_{13}$	$r_{23}$	$r_{123}$
$\hat{N}_0^d$	0,59	0,71	0,68	0,10	-0,01	0,08	0,06
$\hat{N}$	0,55	0,67	0,64	0,16	0,05	0,15	0,03
$\hat{N}_1$	0,56	0,68	0,65	0,15	0,03	0,13	0,03

<sup>a</sup>  $M$ , nombre de cas présents dans au moins une liste ;  $D$ , nombre moyen de cas présents dans 2 listes et pas la 3<sup>ème</sup> ;  $\hat{C}$ , estimation du taux de couverture de l'échantillon ;  $est$ , estimation de la taille de population ;  $se (\pm)$ , erreur standard ;  $cil$ , limite inférieure de IC<sub>95</sub> ;  $ciu$ , limite supérieure de IC<sub>95</sub>.

<sup>b</sup>  $u$ , probabilités moyennes estimées, dépendantes des estimations de  $N$  pour les échantillons 1, 2 et 3 respectivement (1, IPG ; 2, DIM ; 3, LAT).

<sup>c</sup>  $r$ , estimations du coefficient de covariation (CCV) dépendant des estimations de  $N$ , pour différentes interactions entre échantillons 1, 2 et 3.

<sup>d</sup> Se réfère à la population estimée pour des sources indépendantes ( $\hat{N}_0$ ), un taux de capture des cas suffisamment élevé ( $\hat{N}$ ), ou un taux de capture faible ( $\hat{N}_1$ ).

### 3.3.3. Récapitulatif des résultats en termes de nombre de cas

Les méthodes de Petersen et Chapman ne fournissent pas d'estimateurs fiables puisque nos sources de données sont non indépendantes entre elles. En revanche, les modèles log-linéaires et l'approche *sample coverage* donnent des estimateurs fiables.

Chacune de ces deux méthodes a un certain nombre d'avantages et d'inconvénients (cf. [Chao *et al.*, 2001] pour plus de détail). Un désavantage majeur des modèles log-linéaires est de proposer plusieurs modèles ayant la même qualité d'ajustement, mais qui vont donner des estimations pouvant être très différentes [Coull et Agresti, 1999], alors qu'avec l'approche *sample coverage*, aucune sélection ou comparaison de modèles n'est nécessaire [Chao *et al.*, 2001]. Pour plus de cohérence dans la lecture et la comparaison des résultats de l'Ile de Cayenne et de la Guyane, nous avons choisi pour les analyses qui vont suivre de nous baser sur une unique estimation, soit celle obtenue avec l'approche *sample coverage* sur l'Ile de Cayenne.

### 3.3.3.1. Sur l'île de Cayenne

#### *Nombre de cas cumulés sur la période 1996-2003*

Cette étude est à notre connaissance la première à estimer le nombre de cas de tuberculose sur l'île de Cayenne entre 1996 et 2003. L'estimation de 381 cas obtenue par simple croisement de trois sources a été corrigée par la méthode de capture-recapture. L'approche *sample coverage* donne une estimation de 425 cas (IC<sub>95</sub> : 407-453), ce qui représente une augmentation de 11,6% de cas. Les estimations obtenues avec les modèles log-linéaires sont similaires.

#### *Nombre de cas annuel*

Étant donné le nombre de cas de tuberculose recensés annuellement sur l'île de Cayenne (entre 40 et 70 cas par an après croisement des données de l'IPG, du DIM et de la LAT), et afin d'éviter la multiplication des biais liés aux faibles effectifs, les estimations de capture-recapture précédentes ont été estimées une unique fois sur la base du total des cas de tuberculose recensés entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003. Pour obtenir une estimation de l'évolution annuelle du nombre de cas de tuberculose sur l'île de Cayenne depuis 1996, nous avons donc appliqué une augmentation de 11,6% aux données annuelles obtenues par simple croisement des données de l'IPG, du DIM et de la LAT, ceci simulant la correction obtenue par la méthode de capture-recapture (approche *sample coverage*). La figure 2.11 résume pour chaque source les données annuelles de tuberculose sur la période 1996-2003.

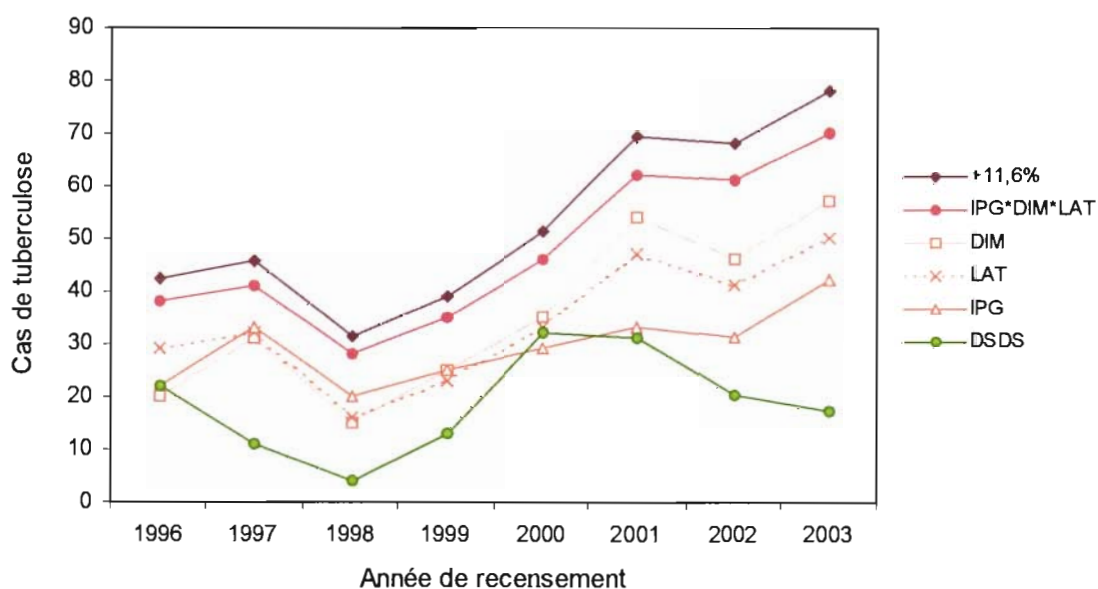


Figure 2.11 – Évolution annuelle du nombre de tuberculeux, île de Cayenne, 1996-2003. Sources : Département de la Santé et du Développement Social (DSDS), Institut Pasteur de Guyane (IPG), service de lutte anti-tuberculeuse (LAT), département d'information médicale de l'hôpital de Cayenne (DIM), après croisement de trois sources (IPG\*DIM\*LAT) et après correction par capture-recapture selon l'approche *sample coverage* (+11,6%).

Cette figure fournit un élément simple de comparaison visuelle du degré de sensibilité des différents systèmes de prise en charge de la tuberculose en Guyane, par rapport au nombre de cas déclarés officiellement à la DSDS.

### **3.3.3.2. Sur l'ensemble de la Guyane**

#### *Résultats cumulés sur 1996-2003*

Pour des raisons déjà explicitées, il était difficile d'effectuer un recensement de l'ensemble des cas confirmés et non confirmés pour l'ensemble de la Guyane depuis 1996 (§2.1). Il est cependant possible d'estimer le nombre de cas cumulés de tuberculose entre 1996 et 2003 sur l'ensemble de la Guyane en utilisant le résultat selon lequel 61,7% des cas de l'Île de Cayenne sont des cas confirmés bactériologiquement. Si ce même ratio est appliqué aux 345 cas confirmés sur l'ensemble de la Guyane entre 1996 et 2003, on peut évaluer à 559 cas le nombre de cas totaux que l'on aurait recensé en Guyane en tenant compte des cas non confirmés. Ce résultat est cependant grossier, eu égard à l'hétérogénéité environnementale, sociale, économique et démographique du territoire guyanais, l'Île de Cayenne n'étant qu'un cas particulier.

#### *Résultats par année*

Disposant de la date de début du traitement des cas confirmés, on peut de la même façon que ci-dessus estimer le nombre de cas totaux par année. Le tableau 2.11 récapitule les estimations annuelles obtenues (notées "est"). Une correction a par ailleurs été appliquée, afin de prendre en compte les cas ayant échappé au recensement (on garde la même correction que précédemment, soit une augmentation de 11,6% des cas). Les résultats de ce calcul sont également détaillés dans le tableau 2.11 (§3.3.4.2.).

### **3.3.4. Résultats en termes d'incidence de la tuberculose**

#### **3.3.4.1. Incidence théorique espérée de l'Île de Cayenne**

##### *Incidence moyenne sur la période 1996-2003*

Nous avons estimé l'incidence moyenne de tuberculose sur l'Île de Cayenne sur la base des 425 cas totaux estimés par la méthode *sample coverage* divisé par huit années, soit 53,1 cas par an pour les 84 181 habitants de l'Île de Cayenne au dernier recensement [INSEE, 1999], soit une incidence évaluée à 63,1 cas par an pour 100 000 habitants. À titre comparatif, pour cette même période (moyenne des chiffres 1996-2003), les chiffres reportés par l'InVS sont de 22,4 cas par an pour 100 000 habitants pour l'ensemble de la Guyane, 27,4 pour l'Île de France, et 11,2 pour la France métropolitaine (*cf.* tableau 2.1).

*Incidence théorique espérée actuelle*

Le tableau 2.10 récapitule les données d'incidence annuelle obtenues par simple croisement des données de l'IPG, du DIM et de la LAT sans correction par l'approche *sample coverage*. Les recensements de population totale de l'INSEE étant disponibles uniquement pour 1990 et 1999, les populations intermédiaires ont été calculées sous l'hypothèse que le taux d'accroissement est constant entre deux recensements.

Tableau 2.10 – Estimations annuelles de l'incidence de la tuberculose sur l'île de Cayenne durant la période 1996-2003 (incidence en nombre de cas / an / 100 000 habitants).

Année	Ile de Cayenne		
	IPG*DIM*LAT <sup>a</sup>	Population <sup>b</sup>	Incidence
1996	38	77 094	49,3
1997	41	79 456	51,6
1998	28	81 819	34,2
1999	35	84 181	41,6
2000	46	87 237	52,7
2001	62	90 294	68,7
2002	61	93 350	65,3
2003	70	96 406	72,6

<sup>a</sup> Se réfère au nombre annuel de cas de tuberculose estimé par simple croisement des données de l'IPG, du DIM et de la LAT.

<sup>b</sup> Population totale de l'île de Cayenne. Le chiffre de 1999 est celui du recensement INSEE ; les autres ont été estimées sous l'hypothèse d'une croissance régulière de la population entre 1990 et 1999.

Les résultats du tableau 2.10 montrent une nette croissance de l'incidence de la tuberculose depuis 1998 : on passe de 34,2 cas par an pour 100 000 habitants en 1998 à 72,6 en 2003, cette croissance étant quasiment linéaire dans le temps (petite baisse de l'incidence en 2002, avant une ré-augmentation en 2003). La moyenne d'incidence de 63,1 cas par an pour 100 000 habitants estimée ci-dessus est donc certainement sous-évaluée par rapport à la situation réelle actuelle sur l'île de Cayenne.

**3.3.4.2. Incidence estimée sur l'ensemble de la Guyane**

*Incidence moyenne 1996-2003*

En utilisant l'estimation de 559 cas totaux de tuberculose en Guyane (§3.3.3.2.), l'incidence moyenne pour 1996-2003 est de 44,4 cas par an pour 100 000 habitants. Cette incidence est certainement sous-estimée par rapport à la situation réelle actuelle, vu l'augmentation de l'incidence de la tuberculose depuis 1998, mais elle est malgré tout supérieure aux chiffres officiels : la

moyenne des incidences de 1996 à 2003 publiées par l'InVS pour la Guyane [<http://www.invs.sante.fr>] donne une incidence moyenne de 22,4 cas par an pour 100 000 habitants, avec un maximum de 39,4 en 2000.

### *Incidence annuelle réelle*

L'incidence annuelle réelle de la tuberculose en Guyane a été calculée sur la base des chiffres annuels corrigés par capture-recapture avec l'approche *sample coverage* (augmentation de 11,6% par rapport aux cas observés). Les données annuelles sont résumées sur le tableau 2.11. On note une incidence croissante de la tuberculose depuis 1999 en Guyane, avec un maximum pour l'année 2003 estimé à 67,8 cas / an / 100 000 habitants.

Tableau 2.11 – Données officielles de l'InVS, estimations brutes et estimations corrigées par capture-recapture avec la méthode *sample coverage* du nombre de cas totaux de tuberculose présents en Guyane entre 1996 et 2003, et incidences calculées sur la base des données corrigées par capture-recapture avec l'approche *sample coverage* (incidences exprimées en nombre de cas par an pour 100 000 habitants).

Année	Guyane			
	InVS <sup>a</sup>	est <sup>b</sup>	+11,6% <sup>c</sup>	Incidence <sup>d</sup>
1996	43	50	56	39,2
1997	24	70	78	52,6
1998	6	55	61	40,3
1999	19	55	61	39,1
2000	66	60	67	41,3
2001	61	81	90	54,3
2002	43	81	90	52,8
2003	37	107	119	67,8
Total	299	559	624	

<sup>a</sup> InVS, nombre de cas de tuberculose notifiés par l'InVS.

<sup>b</sup> est, estimation du total des cas de tuberculose sous l'hypothèse que les cas confirmés représentent 61,7% du total des cas.

<sup>c</sup> total des cas de tuberculose corrigé par l'approche *sample coverage*, soit une augmentation de 11,6% par rapport à "est".

<sup>d</sup> incidences annuelles de tuberculose en Guyane ; les calculs sont basés sur les données de populations estimées sous l'hypothèse d'une croissance régulière de la population entre 1990 et 1999

### *Fiabilité de ces estimations*

Il faut souligner que ces données restent des approximations puisque deux corrections successives ont été appliquées : (1) une correction sous l'hypothèse que les cas confirmés représentent 61,7% du total des cas, et (2) une correction sous l'hypothèse que la différence de 11,6% entre cas estimés (par croisement de trois sources de données) et cas corrigés (par l'approche *sample coverage*)

pour l'île de Cayenne est la même pour la Guyane. Mais l'intérêt de ces estimations est de donner des ordres de grandeur des données réelles, ce qui est en soi une forte plus-value par rapport aux données officielles qui sont fortement biaisées.

Rappelons d'autre part que sur la carte d'incidence mondiale publiée par l'OMS (figure 1.1, chapitre 1), la Guyane figure comme une zone où ces données sont inconnues.

### **3.4. Discussion**

Nous avons pu, au travers de notre étude, évaluer le taux de détection de nouveaux cas de tuberculose, et nous en discutons ici. Les analyses réalisées ne permettent cependant pas d'évaluer la proportion de cas qui ont été jusqu'à une guérison complète. De même le cas particulier des cas de tuberculose multi-résistante n'est pas pris en compte. Après avoir soulevé ces deux points, nous proposons différentes solutions susceptibles d'améliorer le système de prise en charge de la tuberculose sur l'île de Cayenne et en Guyane.

#### **3.4.1. Détection et traitement des cas**

##### **3.4.1.1. Taux de détection des nouveaux cas**

###### *Sensibilité des 3 sources combinées (IPG, DIM, LAT)*

Le taux de détection des cas est défini ici comme le nombre de cas recensés dans une ou plusieurs sources sur le nombre de cas réel estimé par capture-recapture. Les résultats obtenus montrent que la sensibilité de capture des trois systèmes de surveillance IPG, LAT et DIM combinés est bonne (381 cas recensés sur 425 réels estimés, soit 89,6% des cas).

Cependant, aucun organisme de santé publique sur l'île de Cayenne ne centralise l'ensemble des données de l'IPG, du DIM et de la LAT ; les données combinées de ces différents organismes ne sont donc pas directement disponibles et exploitables, certaines ayant même pu échapper à notre propre recensement. D'autre part, la proportion d'individus que nous avons recensés, c'est-à-dire suivis par au moins une des trois sources, n'est pas forcément représentative de la proportion d'individus traités jusqu'à leur guérison totale, proportion qu'il est pourtant nécessaire d'estimer si l'on veut évaluer l'efficacité réelle du système de soins (cf. §3.4.1.2. ci-après).

###### *Sensibilité propre de chacune des 3 sources (IPG, DIM, LAT)*

Sur la base de l'estimation du nombre de cas totaux de tuberculose par le modèle *sample coverage*, la sensibilité de capture des trois systèmes de

surveillance considérés varie entre 55,3 et 66,6% des cas, contre 35,3% pour le système de notification des cas (DSDS).

Afin de vérifier la cohérence de nos résultats, nous les avons comparés à ceux obtenus dans des études antérieures similaires : par exemple, une étude concernant le paludisme dans les armées française donne des sensibilités des systèmes de surveillance de 49,6 à 56,2% [Deparis *et al.*, 1997], et une autre concernant la maladie du Légionnaire en France donne des sensibilités de 58 à 66% [Infuso *et al.*, 1998]. Ces valeurs sont similaires aux nôtres, ce qui étaye nos résultats, mais montre aussi que la sensibilité de capture de nos trois sources (IPG, LAT et DIM) n'est pas moins satisfaisante que celle d'autres organismes similaires basés en France métropolitaine.

#### 3.4.1.2. Taux de réussite des traitements

Malgré la sous-déclaration de la tuberculose en Guyane, la combinaison des trois sources d'information choisies donne un taux de couverture important (« *sample coverage estimate* » de 85,9%). Mais cette estimation ne nous donne pas la proportion de cas réellement traités et guéris.

On peut supposer que l'ensemble des patients recensés au service de Lutte Anti-Tuberculeuse (LAT) a au moins débuté un traitement, ce qui représente 71,1% des cas (par rapport aux 381 cas recensés dans au moins une des trois sources, ou 63,8% par rapport aux 425 cas estimés par capture-recapture, tenant compte des cas manquants). Ces chiffres sont à comparer avec les taux de traitement moyens, qui vont de 70 à 95% des cas dans les pays développés, et de 50 à 75% dans les pays en développement [Blower *et al.*, 1996].

Mais la proportion de patients traités est très loin du taux de guérison des patients. Une étude réalisée en Guyane sur la période 1990-2000 [Léopold Aït-Ouada, 2002] estime à 40,5% le taux de réussite des traitements anti-tuberculeux en 2000, à 11,9% le taux de rechutes sur l'ensemble des cas diagnostiqués entre 1990 et 2000, et enfin à 22,6% la proportion de patients perdus de vue en 2000. Cette faible efficacité à mener à bien le traitement jusqu'à guérison complète des malades tuberculeux a de fortes implications en santé publique, puisqu'elle conduit à la propagation de souches résistantes et multirésistantes en Guyane, cette résistance aux antibiotiques étant mise en évidence par les antibiogrammes réalisés par l'Institut Pasteur de Guadeloupe (cf. §3.4.1.3. ci-après).

#### 3.4.1.3. Les cas de tuberculose multi-résistante

##### *Tuberculose multirésistante et « point chaud »*

Une tuberculose multi-résistante (notée MDRTB dans la littérature anglophone, pour *multidrug-resistant tuberculosis*) est définie par une

tuberculose résistante aux deux antibiotiques anti-tuberculeux majeurs, c'est-à-dire l'isoniazide et la rifampicine [Blower et Chou, 2004].

Un point chaud (ou « *hot zone* ») est défini comme une région où plus de 5% des cas de tuberculose sont des cas de tuberculose multirésistants [Blower et Chou, 2004]. Les données concernant la résistance aux antibiotiques ont été collectées à l'Institut Pasteur de Guadeloupe, au laboratoire régional de référence pour la tuberculose et les mycobactéries qui est en charge des antibiogrammes. Les antibiogrammes ne sont disponibles que pour les patients dont on dispose d'une culture viable de la bactérie, c'est-à-dire les cas confirmés. Sur les 345 cas confirmés de tuberculose recensés sur l'ensemble de la Guyane entre 1996 et 2003, on dispose de 340 antibiogrammes dont 9 révèlent des cas de MDRTB soit 2,6% ( $9 / 340 = 0,026$ ). La Guyane n'est donc pas un point chaud.

#### *Inclusion des cas non confirmés (sans antibiogramme)*

En prenant en compte l'estimation de 559 cas incluant les tuberculoses non confirmées recensées sur l'île de Cayenne, la Guyane ne peut être considérée comme un point chaud que si le taux de MDRTB excède 5% des cas, soit un rapport de 28 MDRTB sur 559 cas. Cette condition est respectée si, parmi les 214 cas supplémentaires ( $345 + 214 = 559$ ), il y a au moins 19 cas de MDRTB. Nous avons donc calculé la probabilité d'obtenir 19 cas de MDRTB parmi 214.

Soit  $x$  le nombre de MDRTB parmi les cas totaux, avec  $p$  la probabilité qu'un cas pris au hasard soit un cas MDRTB. On estime cette probabilité par rapport au nombre de MDRTB observés parmi les 340 antibiogrammes, soit 9.

$$p(x=1) = 9/340 = 0,026 \text{ et } q = 1 - p = 0,974$$

La probabilité d'observer au moins 19 cas de MDRTB est :

$$p(x \geq 19) = 1 - \left( \sum_{i=0}^{18} \binom{214}{i} p^i \times (1-p)^{214-i} \right) \approx 5.332 \times 10^{-6}$$

La Guyane a donc une probabilité nulle de devenir un point chaud en terme de tuberculose multi-résistante, simplement en intégrant les cas non confirmés, sur lesquels il n'est pas possible d'obtenir un antibiogramme.

### **3.4.2. Améliorer le système de prise en charge**

#### **3.4.2.1. La stratégie DOTS**

##### *Définition et éléments essentiels de la stratégie*

Dans les années 1990, l'OMS et l'UICTRM (Union Internationale Contre la Tuberculose et les Maladies Respiratoires) ont commencé à parler de la

tuberculose comme d'une urgence à l'échelon mondial, et à développer une approche globale de la maladie. C'est ainsi qu'un nouveau programme de lutte contre la tuberculose a été lancé en 1993-1994 : la stratégie DOTS (pour *directly observed treatment, short course*, ou traitement de courte durée sous supervision directe).

Cette stratégie DOTS repose sur cinq points essentiels [WHO, 2001] :

- engagement des pouvoirs publics à lutter contre la tuberculose,
- dépistage par examen microscopique de l'expectoration de tous les patients suspects de tuberculose (avec en plus une culture si possible, mais ce n'est pas indispensable),
- traitement standardisé court, de 6 à 8 mois, des cas à frottis positif (BAAR+) sous surveillance régulière (avec une surveillance directe des malades pendant au moins les deux premiers mois pour s'assurer qu'ils prennent bien leurs médicaments),
- aide à l'établissement et au maintien d'un système d'approvisionnement régulier et ininterrompu en médicaments anti-tuberculeux essentiels,
- établissement et maintien d'un système standardisé d'enregistrement et de déclaration des cas pour suivre le progrès des traitements.

### *Objectifs de l'OMS*

Les objectifs définis par l'OMS sont inscrits dans le cadre de la stratégie DOTS, basée sur la détection des frottis positifs (BAAR+), qui sont la principale source de contagion et ont un taux de mortalité supérieur par rapport aux BAAR-. Le taux de détection des cas est défini par l'OMS comme le nombre de notifications annuelles de BAAR+ sur l'incidence annuelle estimée de BAAR+ [WHO, 2005]. La stratégie DOTS visait à détecter, à l'horizon 2005, 70% des nouveaux cas de tuberculose contagieux (BAAR+), et à guérir 85 % des personnes diagnostiquées.

### *Défaillances de la stratégie*

L'outil de diagnostic de la tuberculose active (examen direct au microscope) a été développé en 1882 et n'identifie que la moitié des personnes souffrant de tuberculose [MSF, 2004]. En effet, cet examen ne détecte pas certaines formes pulmonaires de la tuberculose, ni les formes extra-pulmonaires, il détecte très difficilement les malades co-infectés VIH/tuberculose, et il exclut les enfants, qui ont beaucoup de mal à produire des crachats [MSF, 2004]. De plus, cette stratégie, qui a pour but de limiter la contagion, ne prend en compte que les cas positifs à l'examen direct, considérés comme les plus contagieux ; or, une précédente étude concernant la ville de San Francisco a montré que les patients BAAR- étaient responsables de plus de 21% des cas recensés dans cette ville [Behr *et al.*, 1999].

D'autre part, il a déjà été souligné que, même s'il est atteint, l'objectif de l'OMS de traiter 85% des cas à frottis positifs ne devrait pas conduire à l'élimination globale de la tuberculose, l'éradication d'une maladie étant définie par une incidence inférieure à 1 cas pour 1 million d'habitants [Blower *et al.*, 1996 ; Kochi, 1991].

### **3.4.2.2. Meilleure prise en charge de la tuberculose en Guyane**

D'après Blower *et al.* (1996), il existerait au moins deux moyens d'améliorer les stratégies de contrôle de la tuberculose : (1) en augmentant la détection des cas, et donc la proportion de cas traités, et (2) en améliorant la prise en charge des patients, et donc leur compliance.

#### *Meilleure détection des nouveaux cas*

Actuellement, la détection des cas de tuberculose en Guyane est une détection passive, qui nécessite une démarche du malade vers une structure de soin. Un moyen d'améliorer la détection des cas serait d'identifier les populations à risque pour la tuberculose, afin de cibler des interventions de détection active de cas par les services de la LAT.

Cependant, une stratégie de contrôle permettant une meilleure détection des cas, sans une augmentation parallèle du suivi des cas risque de mener à un échec des traitements, et au développement de tuberculoses résistantes. Un effort simultané doit donc être mené sur les deux fronts.

#### *Meilleure suivi des cas au cours du traitement*

Une précédente étude avait déjà proposé des solutions pour améliorer le suivi des patients en Guyane [Léopold Aït-Ouada, 2002] : mise en place d'un carnet de suivi individuel, écrit en français et dans la langue maternelle du patient, permettant d'assurer une continuité des soins en cas de déménagement, ou de retour dans le pays d'origine pour les patients étrangers ; instauration d'un suivi hebdomadaire plutôt que mensuel pendant les 2 premiers mois de traitement ; notification obligatoire des cas diagnostiqués par les laboratoires d'analyse.

Il serait également important d'améliorer le retour d'information entre les différents organismes, certains médecins se plaignant par exemple de n'avoir pas toujours les résultats d'analyses. L'extension de la stratégie DOTS publiée en 2001 demandait la création de comités nationaux de coordination inter-institutions [WHO, 2005] ; il serait intéressant de mettre en place un système de ce type à l'échelle régionale de la Guyane.

Enfin, dans le contexte de la Guyane, région qui connaît d'importants flux migratoires en provenance de pays à forte incidence de tuberculose, avec des populations qui sont souvent dans des situations précaires, il paraît important

de mettre en place un accompagnement social du malade tuberculeux, prenant en compte ses besoins, sa situation sociale et familiale, son emploi, son logement, etc. L'intégration d'assistantes sociales au sein des acteurs de la lutte anti-tuberculeuse en Guyane permettrait une prise en charge plus adaptée de certains groupes réfractaires aux méthodes de la médecine traditionnelle, ou ayant à gérer des situations d'exclusion.

## 4. Conclusion

Cette analyse souligne à quel point l'incidence de la tuberculose, et donc le poids de cette maladie en terme de santé publique, sont sous-évalués en Guyane. En 2003 par exemple, nos estimations donnent une incidence de la tuberculose de 67,8 cas par an pour 100 000 habitants (après correction par capture-recapture), alors que les chiffres officiels publiés par l'InVS sont de 10,2 cas par an pour 100 000 habitants pour la France métropolitaine comme pour la Guyane [Che et Bitar, 2005]. Il faut cependant préciser que les personnels médicaux exerçant en Guyane sont eux très conscients de la situation réelle, mais aucun programme spécifique à ce département n'a encore été mis en œuvre afin d'améliorer la situation.

Une des solutions proposées dans ce chapitre pour augmenter la détection des cas était d'identifier les populations à risque de tuberculose. L'objectif de la suite de notre travail de recherche est justement de comprendre la dynamique de la tuberculose en Guyane, afin de créer un index de vulnérabilité des populations vivant sur l'Île de Cayenne, et de produire, au final, une cartographie du risque de tuberculose. Nous développons dans le chapitre suivant les différentes approches et méthodes utilisées dans la suite de l'étude, et qui, du fait de la transversalité de ce travail de recherche, relèvent de disciplines variées.

## Chapitre 3

# Relever le défi de la lutte contre la tuberculose en Guyane : quels outils pour quels besoins ?

*Ce chapitre a pour objectif d'identifier les besoins spécifiques à la lutte contre la tuberculose en Guyane, en tenant compte des caractéristiques propres à cette région qui ont été précédemment décrites, et de présenter les différents outils et méthodes susceptibles de répondre aux besoins définis dans ce contexte. Les outils et méthodes choisis ne sont pas nouveaux, mais leur application est innovante et passe par une approche pluridisciplinaire. Le but de ce chapitre est également d'expliquer aux lecteurs pourquoi des approches aussi variées que la géomatique, l'épidémiologie et la génétique ont été utilisées, et de retracer la logique globale du travail de recherche réalisé. Pour cela nous allons présenter tour à tour un bref état de l'art de l'analyse spatiale et des outils associés, et de l'épidémiologie moléculaire, puis l'application de ces différentes méthodes à l'étude de la tuberculose.*

## 1. La tuberculose humaine en Guyane : quels sont les besoins, et comment y répondre

### 1.1. Comment améliorer le contrôle de la maladie ?

#### *Les besoins actuels*

Au vu de la situation globale actuelle concernant l'épidémie tuberculeuse, les besoins prioritaires pour améliorer la situation ont été définis : (1) trouver des solutions thérapeutiques plus efficaces (création de nouveaux traitements ou vaccins) [O'Brien et Nunn, 2001], (2) améliorer la compliance des patients (suivi social et thérapeutique renforcés) [Blower *et al.*, 1996 ; Hill et Ramachandran, 1992], ou (3) identifier plus efficacement les nouveaux cas de tuberculose [Blower *et al.*, 1996]. Les deux premiers volets cités relèvent

davantage d'une décision politique que d'une question scientifique. En revanche, l'identification des nouveaux cas de tuberculose peut faire l'objet d'une recherche approfondie, permettant de cibler les individus prioritaires.

### *Détection active ou détection passive ?*

Depuis de nombreuses années, l'OMS recommande la détection passive de cas en tant que stratégie prioritaire de lutte contre la tuberculose, au travers de la stratégie DOTS [Murray et Salomon, 1998]. Les arguments contre la détection active de cas sont basés sur : (1) des études en Europe et au Japon qui ont montré que seul 15 à 25% des nouveaux cas BAAR+ étaient identifiés par cette méthode [Krivinka *et al.*, 1974], (2) la supposition que, à un certain stade de la maladie, la majorité des patients se présentent aux services de santé [Aluoch *et al.*, 1987], et (3) le coût élevé des programmes de détection active de cas [Murray et Salomon, 1998 ; Taylor et O'Brien, 2001].

Il a cependant été démontré par la suite que la détection active des cas de tuberculose peut permettre de réduire fortement la mortalité liée à la maladie [Murray et Salomon, 1998]. Pour améliorer le rapport coût-efficacité de cette détection active, il est nécessaire de cibler les populations à risque qui, outre les individus ayant été en contact direct avec des malades et qui sont dépistés systématiquement, peuvent être les populations carcérales [Jones et Schaffner, 2001] ou immigrantes [Khan, 2002], par exemple.

## **1.2. Quelles sont les questions à résoudre ?**

Le présent travail de thèse a été imaginé pour trouver un moyen de détecter de façon précoce et exhaustive les nouveaux cas de tuberculose en Guyane. Le but est au final de proposer un outil d'aide à la décision efficace permettant de cibler les actions de lutte contre la tuberculose en Guyane, et devant être potentiellement applicable à d'autres pays où le contexte est similaire ou proche. Après une présentation des objectifs généraux du travail de recherche et de la problématique générale qui y est associée, nous allons successivement passer en revue les trois grandes questions qui nous intéressent : quand, où, et à qui se transmet la tuberculose en Guyane.

### **1.2.1. Objectifs généraux de l'étude et problématique**

#### *Objectif général de l'étude*

La présente étude se propose, par la compréhension de la dynamique spatiale et temporelle de la tuberculose en Guyane, d'identifier les facteurs de risque liés à la persistance de la maladie dans cette région, et plus

particulièrement sur l'île de Cayenne. La compréhension de cette dynamique passe par une approche à la fois épidémiologique, écologique et géographique.

L'utilisation couplée d'une analyse épidémiologique et moléculaire, et d'une analyse géographique, au travers d'outils de la spatialisation, doit permettre d'identifier les zones favorables au développement de la maladie, où se trouvent donc les populations les plus vulnérables. D'autre part, l'identification des zones où se concentrent depuis plusieurs années les cas de tuberculose permettra de valider ces résultats, dans le but d'établir une cartographie fine du risque tuberculeux sur l'île de Cayenne.

#### *Intérêt de l'étude*

L'intérêt primordial de cette recherche est de fournir des résultats permettant, à un niveau préventif, l'intervention efficace et rapide des organismes décisionnels de santé, en particulier lorsque les sites d'intervention prioritaires sont mal ciblés, faute de moyens d'identification.

Le but n'est pas de remettre en cause l'efficacité de techniques telles que les enquêtes épidémiologiques, qui permettent d'identifier les facteurs de risque associés aux maladies, mais plutôt de tester ce que de nouvelles méthodes peuvent apporter à la connaissance de la maladie et à sa prévention. Les cartes de risque pourraient ainsi représenter un outil efficace pour améliorer la prise de décisions dans le domaine de la santé, et le suivi des épidémies par l'élaboration d'un système d'alerte.

#### *Problématique de l'étude*

« Quels sont les processus en cause dans la propagation et le maintien de la tuberculose en Guyane ? » C'est autour de cette question centrale que s'organise notre problématique de recherche. Mais elle soulève d'autres interrogations plus ciblées : comment la tuberculose se transmet-elle en Guyane, à quelle vitesse, à quelle échelle spatiale, et comment persiste-elle sur le long terme, malgré les efforts d'éradication mis en œuvre ? De façon plus simple, où, quand, et à qui se transmet la maladie ?

La réponse à ces questions s'appuie sur l'analyse de la dynamique spatiale et temporelle de la maladie sur plusieurs années. Comprendre les processus et la dynamique spatio-temporelle de transmission de la tuberculose en Guyane, c'est pouvoir définir les facteurs favorables à la maladie, et donc pouvoir potentiellement prédire les facteurs de persistance de la tuberculose dans d'autres zones du globe réunissant les mêmes conditions et/ou contraintes.

### **1.2.2. Quand ? La dynamique temporelle de la maladie**

Dans notre étude, la date de survenue de la maladie chez un patient est la seule information temporelle dont nous disposons. Or, cette information est

insuffisante pour identifier précisément la date à laquelle le malade a été mis en contact avec la bactérie, qui peut précéder de plusieurs années la date de déclaration de la maladie, du fait de la latence potentielle de l'infection [Blower *et al.*, 1995]. Seule une enquête épidémiologique auprès des patients est à même de fournir des informations concernant les contacts étroits qui ont eu lieu entre individus malades, et à quelle date. Cette enquête étant inenvisageable dans le cadre de notre étude, il ne nous a pas été possible d'identifier les cas sources, c'est-à-dire les patients qui ont été contaminés les premiers et qui ont ensuite contaminé d'autres individus (cas secondaires), informations qui auraient permis de retracer le circuit de la contamination entre les patients au fil des mois ou des années.

#### *Analyses de séries temporelles*

Il est cependant possible d'exploiter l'information temporelle dont nous disposons. Le suivi temporel de l'apparition de cas malades est un outil prédictif important [Reis et Mandl, 2003], car il permet de déterminer la dynamique de la maladie qui peut être de type épidémique (forte augmentation ponctuelle de l'incidence) ou endémique (la transmission est stable dans le temps), ces deux dynamiques supposant des stratégies de lutte différentes.

Le suivi temporel des cas de tuberculose sur l'île de Cayenne, où nous avons recensé l'ensemble des cas sur la période 1996-2003, offre donc la possibilité de suivre l'apparition au cours du temps d'épidémies ou micro-épidémies qui peuvent être périodiques ou liées à un événement ponctuel (immigration en masse en provenance d'un pays à forte incidence de tuberculose par exemple).

### **1.2.3. Où ? La dynamique spatiale de la maladie**

#### *Suivi spatial des cas de maladie*

La géographie de la santé, avec l'utilisation grandissante d'outils de statistique spatiale, est une discipline qui se démocratise pour identifier les zones à forte incidence de maladie, et vérifier s'il existe une signature spatiale particulière d'une maladie. Cette méthode a déjà été utilisée très récemment pour étudier la distribution spatiale de cas de tuberculose humaine [Kistemann *et al.*, 2002 ; Moonan *et al.*, 2004 ; Munch *et al.*, 2003].

#### *Suivi spatial des souches pathogènes*

Le suivi spatial des cas ayant la même souche génétique (potentiellement des clones issus d'une même source de contamination) fournit une information sur l'échelle à laquelle se fait la transmission (contamination dans la même maison, la même rue, la même ville...), et permet d'émettre des hypothèses concernant le mode de contamination préférentiel de la maladie (au sein du

même contexte social, familial, du même lieu de résidence, etc.). Par suite, ceci permet d'identifier les comportements ou les personnes à risque de tuberculose.

#### **1.2.4. À qui ? Les individus à risque de tuberculose**

L'analyse spatio-temporelle de la tuberculose a pour but de comprendre les conditions favorables à la transmission de la maladie, et à identifier quels sont les individus les plus vulnérables que doivent cibler les actions de détection active (IDR, radiographies), en plus des enquêtes de proximité ciblant les personnes ayant été en contact avec des malades, et de la détection passive des malades qui se rendent de leur propre initiative dans un centre de soins.

## **2. L'analyse spatiale appliquée aux maladies**

Après un bref état de l'art concernant l'approche spatiale dans les domaines de l'écologie et de la santé, quelques outils d'analyse spatiale sont décrits afin de fournir au lecteur quelques informations concernant le type de données obtenues avec ces outils, et leur utilisation en épidémiologie. La tuberculose étant une maladie où l'analyse spatiale n'a encore été que très peu utilisée, la dernière partie dresse un bref bilan des études déjà réalisées, et de ce qu'il est possible de transposer à notre propre étude.

### **2.1. État de l'art**

Nous évoquons ici l'importance de l'intégration de la composante spatiale à l'écologie et les concepts qui en ont découlé, puis nous discutons de l'application progressive de ces concepts aux problématiques de santé. Enfin, nous donnons quelques exemples d'études récentes liant santé et environnement spatial.

#### **2.1.1. Analyse spatiale en écologie**

La prise en considération des caractéristiques spatiales en écologie est aussi ancienne que la discipline elle-même, les espèces vivantes étant intimement liées au milieu dans lequel elles vivent. On peut, par exemple, évoquer la théorie insulaire biogéographique [Mac Arthur et Wilson, 1967] qui spécifie que les processus d'immigration / extinction des espèces d'une île dépendent de la taille des îles, et de la distance entre îles ou avec le continent ; la théorie des métapopulations [Hanski et Ovaskainen, 2000 ; Levin, 1970] qui décrit une

population comme un ensemble de groupes d'individus en interactions dans un espace commun ; l'écologie du paysage, qui stipule que l'hétérogénéité d'un paysage contraint les flux de matières ainsi que les déplacements d'organismes vivants en son sein [Forman et Godron, 1986] ; la biologie des populations [Legay et Debouzie, 1985] qui établit les liens qui existent entre la démographie d'une population, la répartition spatiale des individus qui la composent, et le fonctionnement global de l'écosystème.

D'un point de vue général, l'analyse spatiale appliquée à l'écologie permet d'identifier la combinaison de facteurs et les mécanismes écologiques qui sous-tendent l'occurrence (c'est-à-dire la présence/absence) d'individus dans leur environnement (mécanismes de dispersion, existence de variables climatiques régionales, etc.), mais aussi leur abondance et diversité locales, ces études ayant mené au développement de la macroécologie [Ricklefs et Schuller, 1993].

### 2.1.2. Analyse spatiale en épidémiologie

L'état de santé d'une population humaine (et, dans une certaine mesure, animale) est l'expression, en un lieu, à un moment donné, d'une combinaison de facteurs qui relève du contexte individuel et collectif (biologique, social, culturel, politique, économique...) [OMS, 2002]. L'analyse spatiale des schémas (ou patrons) de distribution des maladies (considérant soit les patients eux-mêmes, soit les souches du pathogène responsable de la maladie) est devenue un outil d'identification de l'étiologie des maladies [Mayer, 1983], c'est-à-dire des causes des maladies. De nombreuses maladies sont liées, de façon plus ou moins forte, à leur environnement. Ce lien est particulièrement évident pour les maladies vectorielles (transmises par un insecte vecteur du type moustique) pour lesquelles il existe une corrélation entre la distribution des vecteurs et certains paramètres environnementaux qui déterminent leur habitat favorable : pluviométrie, température, type de végétation etc. [Curran *et al.*, 2000 ; Hay *et al.*, 1996].

De nombreuses recherches visant à démontrer des liens entre maladies et environnement se sont développées ces dernières décennies. Elles ont, entre autre, mis en évidence l'existence de relations entre déforestation et échinococcose alvéolaire [Craig *et al.*, 2000], culture du cacao et transmission de l'onchocercose [Cadot *et al.*, 1998], perturbations bio-climatiques et choléra [Reeves *et al.*, 1994], riziculture et paludisme [Dossou-Yovo *et al.*, 1998], etc. Dans le même temps, l'utilisation de la spatialisation et de la géomatique<sup>1</sup> dans les problèmes épidémiologiques a suscité un intérêt croissant [Washino et Wood, 1994 ; De Savigny et Wijeyaratne, 1995 ; Epstein, 1999 ; Linthicum *et al.*, 1999 ; Hay, 2000 ; Avruskin *et al.*, 2004 ; de La Rocque *et al.*, 2004], et ce

---

<sup>1</sup> Discipline ayant pour objet la gestion des données à référence spatiale et qui fait appel aux sciences et aux technologies reliées à leur acquisition, leur stockage, leur traitement et leur diffusion (par exemple, la topométrie, la cartographie, la géodésie, la photogrammétrie, la télédétection et l'informatique).

pour des maladies très différentes d'un point de vue épidémiologique : choléra, paludisme, schistosomose, trypanosomose humaine ou animale, fièvre de la vallée du Rift, maladie de Lyme, leishmaniose, dengue, etc. Un des premiers articles en la matière, datant de 1970, qualifie ces outils de « nouveaux yeux pour les épidémiologistes » [Cline, 1970].

## **2.2. Les outils de la spatialisation et de la géomatique**

Il s'agit ici de fournir quelques éléments de compréhension sur ce que sont les outils de télédétection et les Systèmes d'Information Géographique, qui sont de plus en plus appliqués à l'étude des maladies.

### **2.2.1. Les outils de télédétection**

#### *Définition de la télédétection*

La télédétection est l'ensemble des connaissances et des techniques utilisées pour acquérir des informations sur l'environnement par des mesures effectuées à distance, sans contact matériel avec celui-ci [Cracknell et Hayes, 1991]. La télédétection moderne est née de la photographie aérienne, et a ensuite intégré les développements plus récents de la recherche spatiale, avec l'acquisition d'images satellitaires [Bonn et Rochon, 1992].

#### *Paramètres dérivés des images*

Les techniques d'observation de la terre sont à même de fournir des informations sur l'environnement physique, économique ou social, que l'on cherche à mettre en relation avec des maladies, et ce à différentes échelles spatiales, temporelles et spectrales.

Les images de télédétection délivrent une information sous la forme d'une matrice de pixels, un pixel étant la plus petite surface homogène constitutive d'une image numérisée, à laquelle correspond une valeur radiométrique [Bonn et Rochon, 1992]. L'extraction de paramètres environnementaux à partir d'images de télédétection se fait généralement grâce à des logiciels de traitement d'images.

#### *Exemples de paramètres liés aux maladies*

Des articles de revue [Goetz *et al.*, 2000 ; Hay, 2000] ont inventorié les paramètres environnementaux couramment extraits des images satellitaires (généralement des indices) et qui peuvent potentiellement être liés à des indices d'occurrence de maladie. Un indice est calculé par combinaison de

plusieurs bandes spectrales<sup>2</sup> destinées à mettre en évidence les particularités radiométriques d'une surface donnée [Robin, 2002]. L'un des indices les plus utilisés est le NDVI, un indice de végétation étroitement corrélé à l'activité chlorophyllienne [Asrar *et al.*, 1989]. Il existe aussi des indices de pluviométrie [Hay *et al.*, 1998], de turbidité de l'eau [Yagamata *et al.*, 1988], etc.

L'occupation du sol est également une des couches d'information les plus souvent utilisées en épidémiologie [Curran *et al.*, 2000]. Elle dérive de la photo-interprétation d'une photographie aérienne ou de la classification d'une image satellitale. La classification est un processus visant à caractériser des classes correspondant à différents milieux (sol nu, eau, savane, forêt primaire, forêt secondaire, etc.) sur la base de données visuelles, ou d'informations radiométriques contenues dans l'image, lorsqu'il s'agit d'une image satellitale. On obtient au final une cartographie d'occupation du sol.

### 2.2.2. Les Systèmes d'Information Géographique

#### *Définition*

D'un point de vue général, un Système d'Information Géographique (SIG) est un outil informatique (logiciel) permettant d'effectuer des traitements divers sur toutes sortes d'informations, du moment qu'elles peuvent être référencées dans l'espace [Bordin, 2002]. Ce peut être des informations graphiques (objets géographiques, cartes) ou des informations attributaires, qui décrivent les caractéristiques des objets géographiques. Par exemple, si les objets géographiques considérés sont des maisons, préalablement géoréférencées, les données attributaires associées peuvent être la surface habitable, le nombre de pièces et le nombre de personnes qui y vivent. Les informations dérivées du traitement des images de télédétection décrites précédemment (*cf.* §2.2.1.) peuvent donc être intégrées dans un Système d'Information Géographique.

Fonctionnellement, un SIG est un outil informatique qui permet de saisir, représenter, transformer, interroger, mettre à jour les informations référencées dans l'espace dont nous avons parlé plus haut [Bordin, 2002]. L'analyse spatiale proprement dite se fait grâce à des requêtes du type inclusion, exclusion, contiguïté, proximité, ces requêtes permettant d'établir les liens existant entre les différents paramètres inclus dans le système (les couches thématiques). Le SIG permet également l'édition des résultats de l'analyse sous forme cartographique.

Classiquement, le principe du SIG est illustré par la superposition de différentes cartes thématiques (figure 3.1).

---

<sup>2</sup> Une bande spectrale correspond au domaine de sensibilité optique d'un capteur de télédétection dans une certaine largeur de bande, correspondant à un intervalle de longueur d'onde.



Figure 3.1 – Illustration de la superposition des couches thématiques dans un Système d'Information Géographique. Ces informations sont géoréférencées et concernent ici les paysages, le sol, l'hydrographie, et la topographie. (Source : <http://www.cev.washington.edu/lc/CEVIMAGES/dhsvm-static-layers.jpg>)

### *Application à l'épidémiologie*

L'utilisation des systèmes d'information géographique dans l'analyse du risque en santé tend à se démocratiser. Les SIG peuvent en effet être utilisés pour identifier les zones favorables à une maladie (par exemple les zones de présence du vecteur, les zones de contact hôte/vecteur, les cours d'eau contaminés, etc.), donc potentiellement à risque en cas de contact avec un hôte susceptible [Dale *et al.*, 1998 ; Michel *et al.*, 2002].

L'approche méthodologique consiste à établir des corrélations entre certains paramètres environnementaux et des données épidémiologiques, l'ensemble de ces informations étant au préalable inclut dans un SIG. L'une des méthodes possibles repose sur le calcul de la proportion de chaque type d'occupation du sol dans un certain périmètre autour du cas de la maladie, qui est géoréférencé [Beck *et al.*, 1994]. Le risque est évalué en fonction de la fréquence d'apparition d'un ou plusieurs paramètres autour des cas. Ainsi, si la présence d'eau dans un rayon de 200 mètres autour des cas est détectée dans 95% des points observés, il est possible de créer un indice de risque en fonction de la quantité d'eau présente dans un environnement considéré. Cette corrélation ne définit cependant en rien une relation de cause à effet. Ce type d'analyse aboutit généralement à l'édition de cartes de la densité de population de vecteurs, de cartes d'incidence d'une maladie, de cartes de risque, etc.

## **2.3. Application de ces outils à notre étude**

### **2.3.1. Application à la tuberculose**

#### *État de l'art*

Jusqu'à présent, très peu de recherches utilisant les SIG ont été menées sur la tuberculose, certainement parce que le lien avec l'environnement est moins évident que pour les maladies vectorielles. Seules trois études sur la tuberculose utilisant réellement le SIG à des fins d'analyse spatiale et pas seulement de représentation cartographique de la distribution des cas, ont été recensées : Kistemann *et al.*, 2002 ; Munch *et al.*, 2003 ; Moonan *et al.*, 2004. Mais dans ces trois études il s'agit plutôt d'une analyse statistique de la distribution des cas, que d'une analyse spatiale du risque tuberculeux telle que nous souhaitons la mettre en œuvre dans notre travail.

#### *Paramètres liés à la tuberculose*

La tuberculose est une maladie « sociale », liée à des facteurs de type socio-économique. Les indices ou les paramètres environnementaux (au sens large) à intégrer dans un SIG sur la tuberculose doivent donc être corrélés avec des variables socio-économiques. Dans leur étude, Kistemann *et al.* (2002) ont cherché à mettre au point un indice de pauvreté (correspondant à l' « *index of deprivation* » existant en Angleterre) qui tenait à la fois compte des conditions d'habitat (densité de population, taille des ménages, etc.), du statut ethnique (total d'immigrants, de non-immigrants et d'étrangers), et des conditions économiques (taux de chômage, total d'ouvriers, d'allocataires sociaux, etc.). Des corrélations spatiales ont pu être mises en évidence entre quatre de ces variables, et le taux d'incidence de tuberculose à Cologne, en Allemagne : taux de chômage, total de personnes percevant une allocation, taux d'immigrants non allemands, et total d'immeubles recensés, tout ceci par unité spatiale. Mais l'utilisation de ces mêmes variables est-elle transposable en Guyane ?

### **2.3.2. Application au cas de la Guyane**

Les études citées précédemment utilisent des données démographiques et socio-économiques qui nécessitent, en général, des enquêtes exhaustives, coûteuses en temps et en argent. Or ces informations ne sont pas toujours disponibles. Dans le cas de la Guyane, des erreurs dans les recensements de population de l'INSEE (Institut National de la Statistique et des Études Économiques) ont, par exemple, été relevées [Gardel, 2000].

Une solution alternative est d'accéder à ces informations de manière indirecte. Des méthodes ont été développées pour relier les informations issues

des images de télédétection et certains paramètres démographiques auquel il est difficile d'avoir accès sur le terrain. Par exemple, Faure *et al.* (2003) ont créé un indice de densité de population corrélé avec la densité du bâti, ce bâti étant extrait d'images satellitales après classification. Cet exemple souligne une potentialité supplémentaire des outils de télédétection, qui sera développée pour la production d'une cartographie du risque tuberculeux sur l'île de Cayenne (chapitre 6).

### **3. Les apports de l'épidémiologie moléculaire**

Dans cette partie, nous expliquons ce qu'est l'épidémiologie moléculaire, quel est le but de ce type d'analyse, et l'intérêt de la méthode dans le cadre de notre étude. Certains détails méthodologiques sont également développés, ainsi que les avantages et les limites de la méthode, ce qui nous permet de poser dès à présent les bases permettant au lecteur de comprendre le travail de recherche développé par la suite au chapitre 5.

#### **3.1. Épidémiologie et épidémiologie moléculaire**

##### *L'épidémiologie*

Les définitions sont nombreuses, la plus générale étant celle-ci : discipline qui a pour objet l'influence de divers facteurs (constitution individuelle, environnement, mode de vie, milieu social) sur les maladies, notamment sur leur fréquence, distribution et étiologie, ainsi que sur tout autre phénomène biologique ou social déterminé [TLFI]. L'épidémiologie peut également être définie comme une science quantitative qui étudie le comportement des maladies à l'intérieur d'une population d'hôtes. Elle vise à l'interprétation de la distribution et des déterminants de l'infection au travers de mesures des paramètres de la maladie, et à la recherche des méthodes d'intervention les plus efficaces [Anderson, 1998].

Ces deux définitions sont complémentaires. Dans le cadre de notre étude nous avons intégré les composantes qualitative et quantitative des facteurs qui déterminent la maladie.

##### *L'épidémiologie moléculaire*

Cette discipline est née du développement de nouveaux outils moléculaires et génétiques, et permet d'identifier et de suivre les souches pathogènes (de *M. tuberculosis* par exemple) isolées chez les malades [Burgos et Pym, 2002]. Le but de la méthode est le même qu'en épidémiologie classique, mais l'analyse est

centrée sur la population de pathogènes, et non plus la population hôte, et à un niveau d'analyse bien inférieur, celui du gène.

### 3.2. Épidémiologie moléculaire et tuberculose

Les mycobactéries font partie des bactéries les mieux caractérisées sur le plan génétique, en particulier *Mycobacterium tuberculosis* qui a fait l'objet de très nombreux travaux génétiques et génomiques. Actuellement, on dispose de la séquence complète de *M. tuberculosis* H37Rv [Cole *et al.*, 1998]. De ce fait, l'épidémiologie moléculaire a été particulièrement appliquée à cette bactérie.

#### *Identification des clusters*

Dans le cas de la tuberculose, l'épidémiologie moléculaire est basée sur l'hypothèse que les patients infectés par une même souche de *M. tuberculosis* sont le résultat d'une transmission récente, alors que ceux infectés par des souches uniques (dans l'échantillon de souches considéré) sont des cas d'activation ou de réactivation d'une tuberculose plus ancienne [Burgos et Pym, 2002 ; McConkey *et al.*, 2002]. Les cas s'étant révélés porteurs de la même souche bactérienne, donc potentiellement issus d'une même source de contamination, sont appelés « cas en grappe », et l'ensemble des souches ayant un profil génétique identique définit une grappe, ou « *cluster* » [Small et Behr, 1997].

#### *Applications pratiques*

L'analyse des différentes souches tuberculeuses isolées dans une population a de nombreuses applications pratiques. Elle permet entre autres de suivre l'évolution statistique du nombre de cas associés à un génotype donné, et donc la diversité des souches ; d'évaluer le taux de transmission récente (et par conséquent, l'efficacité des systèmes de contrôle de la tuberculose) ; de suivre la circulation globale des souches ; de déterminer les spécificités géographiques de certaines souches ; de définir des relations phylogénétiques entre souches et de proposer un schéma de l'évolution des bacilles tuberculeux [Burgos et Pym, 2002]. L'objectif final de l'application de l'épidémiologie moléculaire à l'étude de la tuberculose humaine est de comprendre la dynamique de transmission de cette maladie dans une communauté donnée, ce que nous cherchons à faire pour le cas de la Guyane, au travers de l'analyse de l'ensemble des souches isolées en Guyane entre 1996 et 2003.

L'analyse peut se faire à différents niveaux dans une même région. Les sous-populations où le taux de transmission récente est élevé constituent des groupes à haut risque, qui contribuent (quantitativement) de façon majeure à

la persistance de l'infection ; l'identification de ces groupes peut permettre de cibler les actions d'intervention.

### 3.3. Limites de la méthode

#### *Limite due à la technique*

Le lien épidémiologique entre patients de mêmes souches n'est que potentiel [Pfyffer *et al.*, 1998]. Ce n'est pas l'ensemble du génome du pathogène qui est étudié, mais seulement certaines séquences de son ADN, les marqueurs génétiques. De ce fait, deux souches reconnues « identiques » par une même méthode d'identification, c'est-à-dire ayant le même profil génétique, peuvent se révéler être des souches différentes dans la réalité. À l'inverse, deux patients porteurs de deux souches différentes avec une même analyse moléculaire peuvent s'avérer liés épidémiologiquement, la différence observée entre souches étant due à un unique événement de mutation [Benjamin *et al.*, 2001]. Il faut garder ces points à l'esprit lors de l'analyse des grappes. Ce biais peut cependant être atténué par le choix de marqueurs génétiques fortement discriminants (en général des séquences d'ADN fortement variables), et par l'utilisation couplée de deux techniques d'analyse moléculaire ciblant des séquences du génome différentes [Filliol *et al.*, 2000].

#### *Limite due à la diversité des souches*

Un autre biais potentiel est introduit par l'existence de souches plus abondantes et plus dispersées que les autres. Ainsi, une étude concernant l'Afrique du Sud, la Tanzanie, le Zimbabwe, le Kenya et le Malawi, a révélé que des souches prélevées dans des régions géographiques très dispersées, avaient des empreintes génomiques identiques, sans qu'aucun lien épidémiologique puisse être identifié [Picard, 2000].

Des méthodes d'analyse standardisées ont été développées et ont permis la mise en place récente de bases de données regroupant les données de plusieurs pays et continents, par exemple Spolddb3 et Spolddb4<sup>3</sup>. Ces bases de données internationales offrent la possibilité de comparer les souches entre elles à l'échelle de la planète. Des indicateurs de l'abondance et de la dispersion géographique des souches au niveau mondial ont par exemple pu être développés [Filliol *et al.*, 2003]. Certaines souches de *Mycobacterium tuberculosis* ont ainsi été identifiées comme étant des souches ubiquistes, c'est-à-dire qu'elles sont présentes partout dans le monde. L'interprétation des clusters de souches ubiquistes s'avère de ce fait délicate, car il est difficile de

---

<sup>3</sup> Les informations constituant cette base de données internationale sont centralisées au centre de référence des mycobactéries de l'Institut Pasteur de Guadeloupe.

prédire si un lien épidémiologique réel existe entre les cas porteurs de cette souche, ou s'ils ont été contaminés par cette même souche mais de façon totalement indépendante.

## 4. Intérêt de la combinaison des deux approches

### *L'étude de Moonan et al. (2004)*

À l'heure actuelle, l'utilisation couplée de l'analyse moléculaire des souches de tuberculose et d'un Système d'Information Géographique n'a fait l'objet que d'un unique travail de recherche très récent [Moonan *et al.*, 2004] dont le but était d'identifier les zones de transmission active de la maladie.

Les conclusions de cet article sont très prometteuses, puisque la méthode a permis d'identifier certains facteurs de risque de transmission (forte densité de population, pauvreté et surpeuplement, par exemple), ainsi qu'un certain nombre d'épidémies que les méthodes traditionnelles n'avaient pas détecté. Le risque de transmission active a été circonscrit à trois unités géographiques, sur 59 étudiées. D'autre part, des zones à forte incidence de tuberculose, mais où la transmission active est faible, ont également été déterminées, ce qui peut permettre de cibler la prévention dans certaines zones fortement touchées par la maladie, avant qu'une épidémie ne se déclenche et ne soit repérée par les systèmes de surveillance.

### *Choix de la méthodologie de l'étude*

Dans notre travail de recherche sur la tuberculose en Guyane, nous avons également combiné épidémiologie moléculaire et SIG dans le but d'identifier, les zones de transmission active sur l'île de Cayenne. Mais à la différence de l'étude de Moonan *et al.* (2004), nous avons également identifié les zones à risque de tuberculose, le but étant de vérifier si ces zones à forte transmission et les zones à risques sont superposables, ce qui n'est pas forcément le cas, par exemple dans les régions où l'incidence de tuberculose est forte, mais où la transmission active est faible.

Les zones de transmission récente sont étudiées par analyse génétique des souches et spatialisation des clusters génétiques de cas, que l'on suppose être les indicateurs d'une transmission récente, alors que les zones à risque sont identifiées en fonction d'un indice de vulnérabilité sociale de la population que nous avons créé, et de la distribution spatiale des cas de tuberculose. L'objectif final est de fournir une cartographie du risque tuberculeux, mais aussi de comprendre la dynamique de transmission de la maladie qui explique sa persistance sur le territoire.

## **5. Conclusion**

Ces différentes analyses montrent que combiner des outils géomatiques tels que les Systèmes d'Information Géographiques, et l'épidémiologie moléculaire est une méthode qui semble prometteuse pour comprendre la dynamique de la tuberculose en Guyane, en particulier sur l'île de Cayenne. Cette méthode ne peut cependant être appliquée que dans les pays où les analyses moléculaires des isolats bactériens sont faites en routine, ce qui est par chance le cas en Guyane. Au final, ce mémoire espère proposer des modes de prévention et d'intervention adaptés au contexte local guyanais pour l'aide à la décision dans la lutte contre la maladie. Les chapitres suivants développent les résultats obtenus concernant l'analyse spatiale de la distribution des cas sur l'île de Cayenne (chapitre 4), l'épidémiologie moléculaire de la tuberculose en Guyane (chapitre 5) et l'analyse de la dynamique de transmission de la tuberculose en Guyane, aboutissant à la production d'une cartographie du risque tuberculeux sur l'île de Cayenne (chapitre 6).



## Chapitre 4

# Analyse des relations entre environnement spatial et tuberculose sur l'Ile de Cayenne

*La première étape dans l'étude de la dynamique de la tuberculose en Guyane consiste à analyser la distribution spatiale des malades tuberculeux résidant sur l'Ile de Cayenne, recensés sur la période 1996-2003. Dans un premier temps, nous analysons la distribution de l'incidence de cas par zone géographique, afin de vérifier si la répartition des cas de tuberculose est aléatoire ou si elle dépend de la distribution spatiale de la population en fonction de l'âge, du sexe et de la nationalité. Différentes échelles spatiales d'étude (les IRIS et les quartiers) sont testées et discutées. Une deuxième étape, faisant cette fois appel à une validation statistique de l'analyse spatiale, consiste à mettre en évidence des agrégats spatiaux de cas, à différentes échelles et en tenant compte de la répartition de la population par sexe, par âge, et par nationalité. Le but est d'identifier les zones où existe une forte concentration de cas, pouvant potentiellement constituer un foyer de départ d'une épidémie de tuberculose sur l'Ile de Cayenne. La corrélation de ces agrégats avec les zones d'habitat insalubre de l'Ile de Cayenne est finalement discutée.*

## 1. Distribution spatiale des cas de tuberculose sur l'Ile de Cayenne

La cartographie de l'incidence (nombre de nouveaux cas infectés par rapport à la population totale) des cas d'une maladie est une méthode basique pour étudier la distribution d'une maladie dans une zone donnée, mais elle reste descriptive. Outre une brève analyse de la cartographie de l'incidence sur l'Ile de Cayenne, le but de cette partie est plus précisément de discuter de problèmes méthodologiques tels que (1) les choix d'échelles dans les analyses spatiales et leur implication pour l'interprétation de nos résultats, ainsi que (2) la fiabilité des données démographiques utilisées dans nos calculs d'incidence, et les outils de télédétection pouvant offrir des mesures indirectes de ces

données. L'analyse spatiale des cas étant ici principalement descriptive, une analyse plus poussée basée sur des méthodes de statistiques spatiales est développée dans la deuxième partie de ce chapitre.

## **1.1. Distribution spatiale des cas de tuberculose**

### *Cas intégrés dans l'étude*

Afin de conserver la cohésion de l'ensemble de la recherche et de pouvoir comparer les différentes analyses entre elles, nos données se limitent aux cas recensés sur l'île de Cayenne au cours de la période du 1<sup>er</sup> janvier 1996 au 31 décembre 2003, région et période pour lesquelles nous avons effectué un recueil exhaustif de multiples informations nécessaires à notre analyse (*cf.* chapitres précédents).

### *Référentiel pour la localisation des cas dans l'espace*

Pour référencer les cas de tuberculose dans l'espace, il nous a fallu choisir un site de référence : lieu d'habitation, de travail, lieu de contamination... Ce qui nous intéresse ici n'est pas tant le lieu où se fait la mise en contact avec la bactérie, mais plutôt les conditions de vie du patient au moment de la déclaration de l'infection. En effet, l'activation de l'infection à un instant  $t$  est le reflet de conditions favorables au développement de la maladie, en général une baisse des défenses immunitaires, qui peut être liée au lieu de résidence de l'individu (habitat insalubre, isolement, absence de structures de soins à proximité, etc.). Ce sont ces conditions potentiellement à risque que nous cherchons à mettre en évidence au niveau spatial.

Nous avons donc choisi de localiser les patients sur leur lieu de résidence à la date de déclaration de la maladie. Ce choix est renforcé par le taux élevé de chômage que l'on observe en Guyane, les personnes sans emploi passant davantage de temps à leur domicile que les personnes ayant une activité professionnelle (sachant par ailleurs qu'il existe une forte association entre tuberculose et chômage [Darbyshire, 1995]).

### *Recensement des données*

Les adresses des 381 cas de tuberculose recensés sur l'île de Cayenne entre 1996 et 2003, ont été recueillies auprès du bureau des entrées de l'hôpital de Cayenne et dans les registres du service de lutte anti-tuberculeuse (LAT). Sur l'ensemble des cas, seuls 366 ont été spatialisés, dont 220 cas confirmés, et 146 non confirmés par culture. Parmi les cas non géoréférencés, six patients étaient notifiés comme étant sans domicile fixe. Pour les neuf autres patients, aucune adresse précise n'a pu être trouvée dans les registres examinés, outre la ville de

résidence. Ces quinze cas ont cependant été intégrés dans les analyses des chapitres précédents car ils résident bien sur l'Île de Cayenne.

### *Cas particuliers*

Dans le chapitre 2, nous avons mentionné 15 patients « sans domicile fixe ». Outre les six patients n'ayant absolument aucune adresse de notifiée (donc non géoréférencés), trois patients ont été localisés à l'adresse du squat qu'ils occupent, quatre à l'adresse du Relais Drogue Solidarité, et deux sont localisés dans la rue où ils passent la majeure partie de leur temps. D'autre part, les huit cas de tuberculose en détention au moment de la déclaration de la maladie ont été géoréférencés à l'adresse du centre pénitentiaire de Rémire.

### *Géoréférencement*

Les coordonnées géographiques du domicile des patients ont été relevées sur le terrain grâce à un GPS<sup>1</sup> (WGS 84 - UTM 22 Nord). Ces coordonnées ont ensuite été intégrées dans un système d'information géographique avec les informations suivantes : âge, sexe, nationalité (les informations génétiques sur les souches de *M. tuberculosis* également intégrées dans le SIG sont détaillées ultérieurement). Après traitement statistique, les informations personnelles ont été supprimées afin d'obtenir un fichier non indirectement nominatif. Le logiciel de SIG utilisé est GeoConcept 5.5 [GeoConcept, 2000]. La distribution spatiale des cas de tuberculose telle qu'observée est illustrée sur la figure 4.1.

#### **Encadré 4.1 – Système géodésique et coordonnées**

Toute mesure de position sur la Terre se fait par rapport à des lignes imaginaires : la longitude et la latitude. Un système géodésique définit l'endroit précis où ces lignes passent sur la Terre. Du fait de la rotation de la Terre, sa surface est un ellipsoïde aplati aux deux pôles. En faisant abstraction des irrégularités de surface, on définit mathématiquement un ellipsoïde de référence, caractérisé par la valeur de son demi grand axe et le rapport grand axe sur petit axe.

Avec l'avènement du GPS, un système géodésique international a été mis au point : l'ellipsoïde WGS 84 (*World Geodesic System of 1984*). Les cartes IGN utilisent les systèmes ED 50, NTF (nouvelle triangulation de la France) ou WGS 84. Il est donc important de vérifier l'adéquation entre le système géodésique du GPS et celui de la carte que l'on utilise.

Dans un système géodésique, les coordonnées géographiques sont exprimées en latitude et longitude, communément exprimées en degrés. Mais l'inconvénient de ce système est que la distance d'un degré de longitude varie avec la latitude. Afin d'unifier les représentations, l'armée

---

<sup>1</sup> Global Positioning System. Système de localisation par satellite, existant à l'échelle du globe.

américaine a mis au point le système de coordonnées UTM (*Universal Transverse Mercator*).

La projection se fait à partir du centre de la Terre sur un cylindre tangent à l'équateur, les parallèles et les méridiens se coupant à angle droit. La Terre est divisée en 60 bandes de largeur constante du nord au sud, elles mêmes divisées en 20 bandes depuis le parallèle 80° sud jusqu'au parallèle 84° nord. Il existe donc 600 zones UTM.

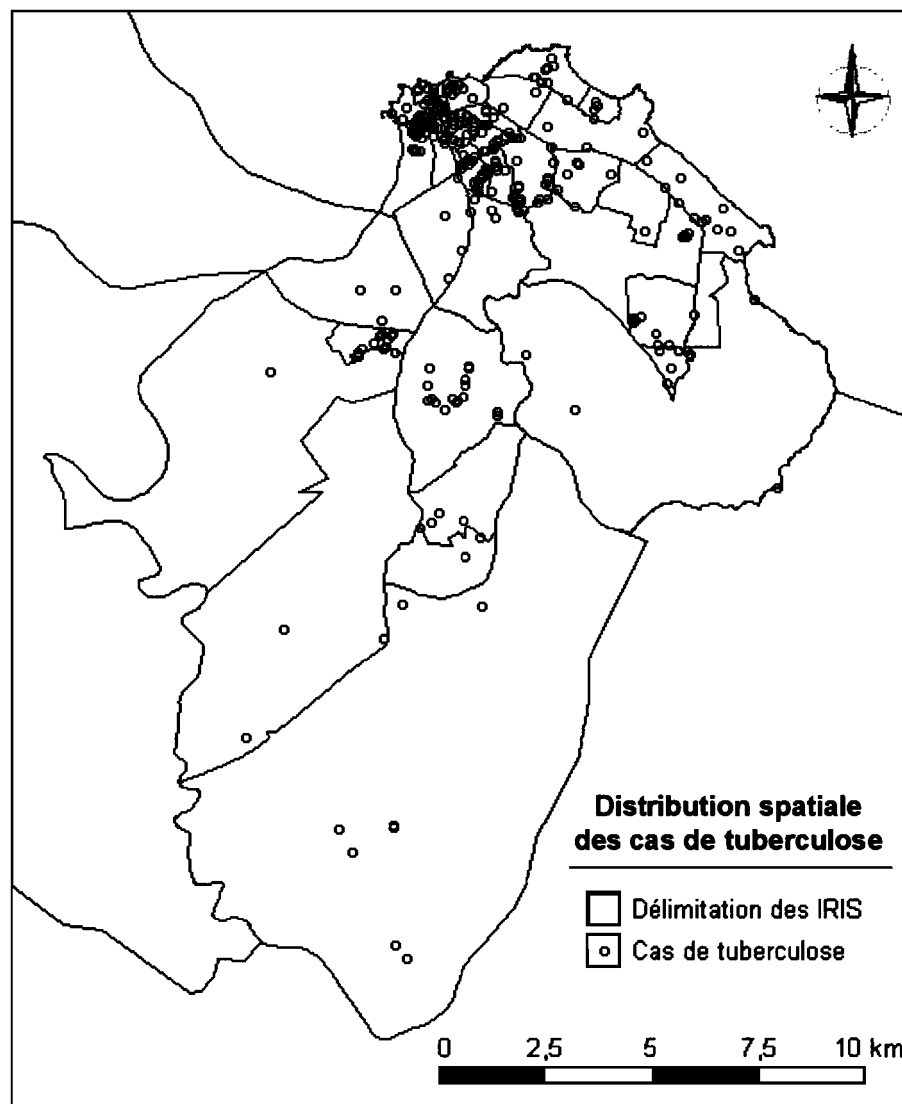


Figure 4.1 – Répartition géographique des cas de tuberculose résidant sur l'île de Cayenne à la date de déclaration de la maladie (cas du 1<sup>er</sup> janvier 1996 au 31 décembre 2003). Le découpage en IRIS (Îlots Regroupés pour l'Information Statistique) correspond aux unités spatiales de référence utilisées pour les recensements de l'INSEE.

#### *Répartition de la tuberculose et de la population*

La cartographie des cas de tuberculose sur l'île de Cayenne (figure 4.1) semble mettre en évidence plusieurs zones géographiques où le nombre de cas

est concentré à très concentré, au nord en particulier. Cependant, il se peut que ceci soit uniquement dû à une hétérogénéité de la répartition de la population. Ainsi, aucun cas de tuberculose ne sera recensé dans une zone non habitée, alors que dans une zone résidentielle, la probabilité de recrutement d'un cas de tuberculose augmente, par simple corrélation avec la présence d'individus, ce qui est ici nettement le cas (*cf.* §1.2.2.2.).

Il existe au moins deux moyens de tenir compte de l'hétérogénéité spatiale de la population : (1) calculer l'incidence des cas par zones géographiques, qui permet de rapporter le nombre de cas à la population totale, ou (2) calculer les ratios standardisés d'incidence par zones géographiques, ce qui permet en plus d'intégrer l'hétérogénéité démographique de la population (âge, sexe, etc.)

## **1.2. Distribution de l'incidence des cas**

L'incidence, qui correspond au ratio du nombre de cas de tuberculose sur la population totale dans une zone définie et s'exprime généralement en nombre de cas malades par an pour 100 000 habitants, permet de tenir compte d'une hétérogénéité spatiale potentielle de la population. Nous avons donc calculé les incidences de tuberculose pour plusieurs unités spatiales : les îlots regroupés pour l'information statistique (§1.2.1) et les zones habitées (§1.2.2.).

### **1.2.1. L'incidence de tuberculose par IRIS**

#### **1.2.1.1. Matériel et méthodes**

##### *Unité spatiale de référence*

Un îlot définit l'unité géographique de base pour la statistique et la diffusion du recensement par l'INSEE. En zone bâtie dense, l'îlot représente le plus souvent un « pâté de maison » alors qu'en zone « périphérique », l'îlot est un ensemble limité par des voies (ou autres limites visibles) découpant cette zone en plusieurs morceaux [INSEE, 1999]. Les îlots peuvent être vides d'habitants (par exemple une zone portuaire).

Les IRIS (îlots regroupés pour l'information statistique) correspondent, comme leur nom l'indique, à des regroupements de plusieurs îlots. Dans les zones résidentielles, les IRIS sont homogènes quant au type d'habitat et regroupent une population de 1 800 à 5 000 habitants [INSEE, 1999].

##### *L'incidence d'une maladie*

Elle est définie par le nombre de nouveaux cas infectés par rapport à la population totale sur une période de temps donnée [Hellenbrand *et al.*, 2001].

L'incidence annuelle moyenne de la tuberculose sur l'île de Cayenne, a été calculée sur la base du recensement de population 1999 de l'INSEE par IRIS. Elle est mise en parallèle avec la cartographie des zones bâties (habitées et non habitées) de l'île de Cayenne (figure 4.2).

### *Discrétisation des données*

Pour une meilleure représentation, le nombre de classes a été choisi d'après l'équation de Huntsberger [EDATER, 2001] :

$$K = 1 + 3,3 \log_{10} n$$

avec  $K$  le nombre de classes, et  $n$  le nombre de valeurs de la donnée.

Dans notre étude,  $n = 35$  IRIS, donc on obtient  $K = 6$  classes. La distribution de l'incidence de tuberculose étant dissymétrique à gauche, nous avons effectué une discrétisation par progression géométrique<sup>2</sup>, c'est-à-dire que l'amplitude des classes augmente au fur et à mesure de la croissance des valeurs de densité et d'incidence [Rican, 1998].

#### **Encadré 4.2 – La discrétisation de données continues**

La représentation cartographique d'une variable statistique continue (situation la plus souvent rencontrée pour la répartition spatiale d'indicateurs de santé tels que les taux de mortalité, de prévalence, etc.) nécessite un découpage statistique préalable des valeurs de la série. Cette procédure de mise en classes des données est appelée discrétisation, et a pour but de rendre l'information géographique contenue dans la série statistique étudiée plus lisible [Béguin et Pumain, 1994].

Différentes méthodes de discrétisation ont été mises au point afin de minimiser (sans les annuler) les pertes d'information liées aux coupures de la série statistique. Le choix d'une méthode s'effectue par rapport à trois critères : (1) la nature de l'information statistique, (2) la forme de la distribution statistique, et (3) l'objectif visé (si l'on cherche à mettre en évidence des valeurs extrêmes par exemple) [Rican, 1998].

À titre d'exemple, voici quelques méthodes de discrétisation, et la distribution statistique pour laquelle elles sont préconisées :

- méthode des seuils naturels : distribution statistique plurimodale
- classes d'amplitudes égales : distribution uniforme ou normale
- utilisation de la moyenne et de l'écart-type : distribution normale
- discrétisation par progression géométrique : distribution fortement dissymétrique à gauche

<sup>2</sup> Une progression géométrique  $n$  de raison  $r$  est une série de valeurs dont le  $i^{\circ}$  terme est égal à :  $n_i = r * n_{i-1}$  avec  $r = (max/min)^{1/K}$ ,  $K$  le nombre de classes et  $min$  et  $max$  les valeurs minimale et maximale de la donnée.

### 1.2.1.2. Résultats

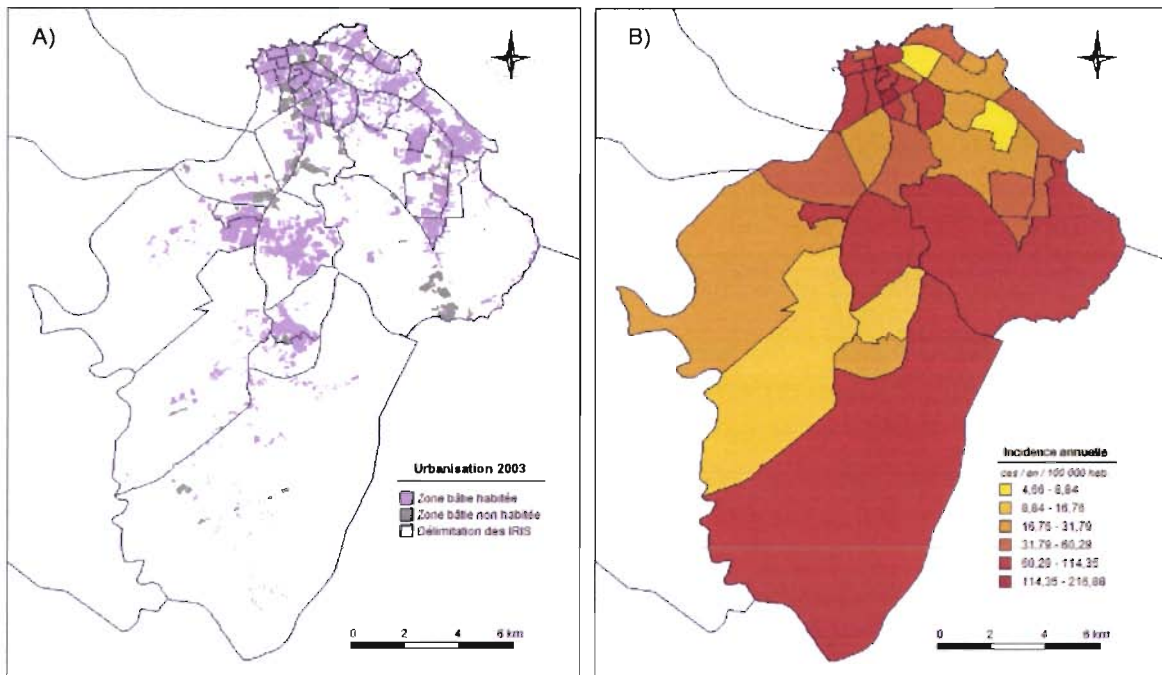


Figure 4.2 – (A) Zones urbanisées de l'île de Cayenne et (B) incidence annuelle de tuberculose par IRIS sur l'île de Cayenne (nombre de nouveaux cas par an pour 100 000 habitants). (cf. encadré 4.2 pour le choix du nombre de classes et des valeurs extrêmes des classes).

L'incidence de tuberculose est très variable suivant les IRIS : de 4,7 à 216,9 cas par an pour 100 000 habitants (figure 4.2). La corrélation de Spearman montre qu'il n'existe aucune corrélation entre la population totale d'un IRIS et l'incidence de tuberculose dans cet IRIS ( $r = -0,126$  ;  $p = 0,47$ ), c'est-à-dire que les zones où l'on recense le plus d'habitants ne correspondent pas forcément aux zones où le nombre de cas de tuberculose rapporté à 100 000 habitants est le plus fort.

### 1.2.1.3. Discussion

L'analyse de la cartographie de l'incidence de la tuberculose sur l'île de Cayenne est insuffisante pour conclure sur l'homogénéité, ou non, de la répartition des cas sur ce territoire. En effet, si les populations de chaque IRIS sont relativement homogènes, les tailles de ces IRIS ne le sont pas (de 0,14 à 70 km<sup>2</sup>), et il se pourrait que certaines zones soient totalement inhabitées. Pour avoir une idée de la répartition plus précise de l'incidence de tuberculose sur l'île de Cayenne, il faut travailler à une échelle plus fine. D'autre part, une analyse spatiale de l'incidence qui exclurait la prise en compte de l'hétérogénéité spatiale potentielle de la population en fonction de l'âge, du sexe, etc. peut également conduire à une interprétation biaisée des résultats. Par exemple, nous avons noté dans le chapitre 2 que les individus les plus

touchés par la tuberculose sur l'île de Cayenne sont de sexe masculin et ont entre 35 et 39 ans. Si dans un IRIS cette population représente 60% du total, une forte incidence de tuberculose dans ce même IRIS pourrait être corrélée à cette concentration d'individus statistiquement plus touchés. L'analyse statistique de la distribution spatiale des cas de tuberculose sur l'île de Cayenne réalisée dans la deuxième partie de ce chapitre tient compte de l'hétérogénéité spatiale potentielle de certains paramètres démographiques (âge, sexe, nationalité).

### 1.2.2. L'incidence de tuberculose par zone habitée

Afin d'intégrer la première remarque énoncée ci-dessus, c'est à dire la prise en compte de l'hétérogénéité de répartition de la population sur l'île de Cayenne, nous avons recalculé les incidences de tuberculose uniquement sur les surfaces urbanisées.

#### 1.2.2.1. Matériel et méthodes

##### *Cartographie de l'habitat*

Comme nous l'avons souligné au chapitre 3, les données de télédétection peuvent être utilisées pour cartographier des éléments de l'environnement, par exemple les zones bâties. Utilisant le logiciel de système d'information géographique GeoConcept 5.0, nous avons délimité manuellement l'ensemble des zones bâties de l'île de Cayenne, et ce par photo-interprétation d'une photographie aérienne de l'IGN (Institut Géographique National) prise en septembre 2001. Ces données ont été complétées par l'analyse d'une image SPOT 5 panchromatique<sup>3</sup> (résolution à 2,5 mètres) acquise le 29 août 2003, ce qui nous donne au final une cartographie de l'urbanisation 2003 de l'île de Cayenne. Le principe de la photo-interprétation est illustré par la figure 4.3.

Dans un deuxième temps, les zones bâties habitées ont été différenciées des zones bâties non habitées (zones industrielles et d'activités, équipements et zones agricoles), et ce en nous appuyant sur une précédente étude concernant l'identification des paysages urbains de l'île de Cayenne pour l'année 1998 [Gardel, 2001] (cf. figure 4.2-A présentée précédemment).

*NB* : notons que certains polygones apparaissant comme une seule plage bâtie correspondent en fait à différents « quartiers ». Le découpage des quartiers est arbitraire, même s'il se base principalement sur la voirie, et ne correspond donc pas obligatoirement à des délimitations administratives réelles.

---

<sup>3</sup> Une image multispectrale inclut l'observation de 5 bandes des gammes de fréquence visible et infrarouge du spectre électromagnétique. L'image panchromatique résulte de la fusion des images multispectrales.

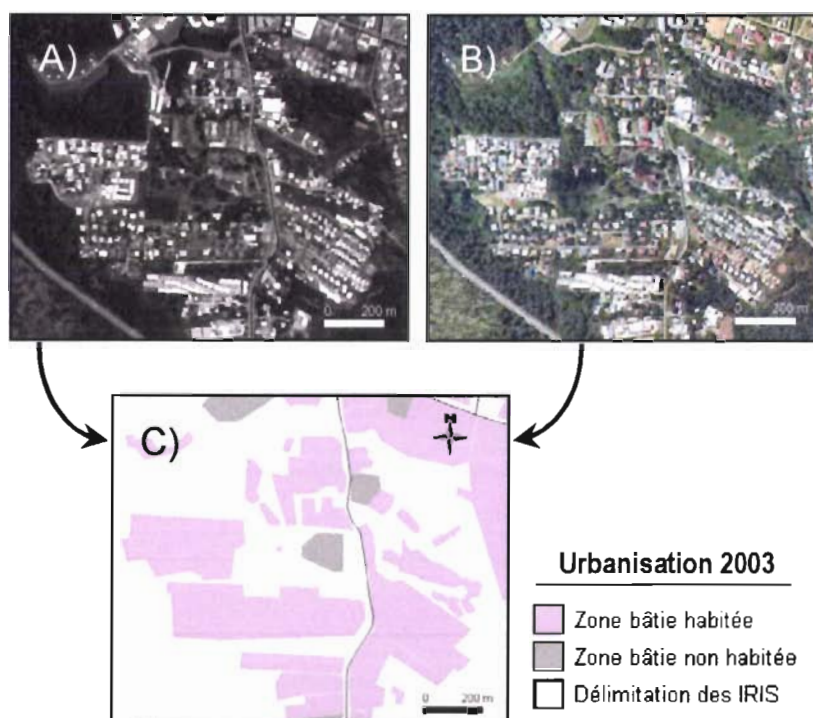


Figure 4.3 – Le quartier de la Madeleine sur l’île de Cayenne. (A) Image SPOT 5-panchromatique du 29 août 2003 (programme ISIS, © CNES, distribution Spotimage), (B) photographie aérienne (© IGN) et (C) résultat de notre photo-interprétation, c’est-à-dire polygones de bâti habité et non habité en 2003 obtenus sous GeoConcept (chaque polygone indépendant correspond à ce que nous avons appelé un « quartier »).

### *Calcul de l’incidence*

Les systèmes d’information géographique permettent de calculer la surface des polygones définissant nos « quartiers ». Pour estimer la taille de population par quartier, nous avons réaffecté la population totale de l’IRIS à chacun des quartiers en fonction de leur surface par rapport à la surface urbanisée totale de l’IRIS. L’incidence de tuberculose par quartier (figure 4.4) a été recalculée sur la base de cette population estimée.

#### **1.2.2.2. Résultats**

Lorsque l’on travaille à l’échelle des quartiers, on met en évidence une forte variabilité de l’incidence, avec des quartiers où l’incidence de tuberculose est nulle (en étant cette fois certain qu’il s’agit bien de zones habitées) (figure 4.4). Cette forte hétérogénéité d’incidence entre quartiers pourrait signifier que la répartition spatiale de la tuberculose sur l’île de Cayenne n’est pas aléatoire et ne dépend pas (ou pas uniquement) de la répartition de la population. Cette première hypothèse formulée doit être vérifiée statistiquement (*cf.* partie 2.). Il faut cependant noter que les effectifs par quartiers sont plus faibles que par IRIS, donc les taux obtenus sont plus instables, et les résultats moins robustes.

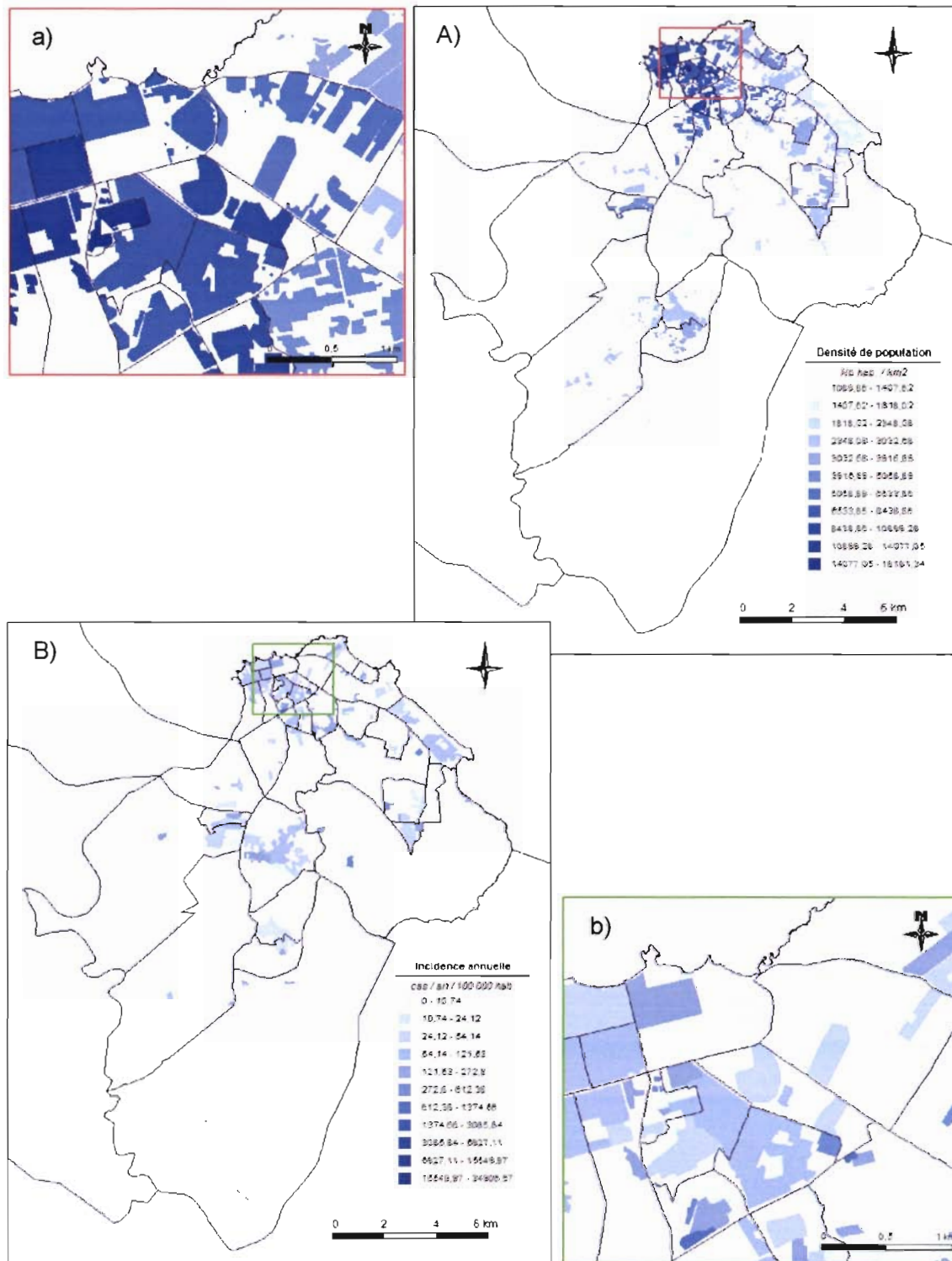


Figure 4.4 – (A) Densité de population estimée par quartier (en habitants par km<sup>2</sup>), et (B) incidence annuelle de la tuberculose par quartier. Ces deux calculs reposent sur les estimations de populations par quartier obtenues en redistribuant la population totale de l'IRIS sur les surfaces réellement urbanisées (§1.2.2.1.). Les encadrés (a) et (b) représentent des zooms respectivement de (A) et (B) sur une même zone correspondant au centre ville de Cayenne (marquée par un rectangle coloré sur les vues générales A et B).

### 1.2.2.3. Discussion

#### *Limite inférieure de précision*

Dans certaines zones de l'île de Cayenne, les habitations sont dispersées et isolées (figure 4.4). Dans ce cas, les polygones construits sur le SIG n'englobent qu'une unique maison, voire deux ou trois. S'il n'y a aucun cas de tuberculose, l'incidence est nulle ; en revanche, s'il y a un cas de tuberculose pour un « quartier » qui regroupe moins d'une dizaine d'habitants, l'incidence calculée rapportée à 100 000 habitants est énorme, et n'a plus de signification réelle.

Il existe donc en quelque sorte un « seuil de résolution » en dessous duquel on ne gagne pas en précision, mais on perd en pertinence de l'information.

#### *Limites méthodologiques*

De même que précédemment, le problème de la répartition spatiale de la population par âge et par sexe n'a pas été résolu. D'autre part, ces analyses de l'incidence sont relativement descriptives et ne mettent pas en évidence une agrégation statistique des cas. Il existe cependant des méthodes de statistiques spatiales permettant de résoudre ce type de problèmes, ce que nous allons étudier dans la suite du chapitre (partie 2).

#### *Densité réelle de population*

Dans le calcul de densité de population effectué pour pouvoir cartographier la répartition de la population de l'île de Cayenne (à comparer avec l'incidence), nous avons posé l'hypothèse que cette densité est la même pour une même surface de quartier considérée. Or, ce n'est pas le cas. En effet, pour une même surface de bâti, la densité de population sera supérieure pour un immeuble par rapport à une zone résidentielle, ce qui introduit un nouveau biais dans l'interprétation des résultats. Pour obtenir un résultat plus juste, il nous faut tenir compte de la typologie d'habitat dont dépend la densité de population.

## 1.3. Distribution des ratios standardisés d'incidence

Lorsque l'on étudie la distribution spatiale de l'incidence d'une maladie, il est important de pouvoir la comparer à la cartographie de la distribution attendue, calculée à l'aide d'un modèle théorique (loi normale ou de Poisson), ou sous l'hypothèse que l'incidence est uniforme sur l'ensemble de la zone étudiée, et égale à l'incidence moyenne de la maladie.

Cependant, les degrés de signification les plus extrêmes surviennent dans les zones les plus peuplées, d'où une perte de robustesse en fonction des tailles de population, et il faut donc utiliser ce que l'on appelle des taux lissés.

### 1.3.1. Matériel et méthodes

Nous avons procédé à deux types d'analyses : (1) *RIS* par IRIS sans lissage, et (2) *RIS* par IRIS et par quartier, avec l'application d'un lissage bayésien aux taux de tuberculose. Les analyses et cartographies ont été réalisées grâce au logiciel GeoDa 9.5 (CSISS, Santa Barbara) [Anselin *et al.*, 2006].

#### *RIS sans lissage bayésien*

Le *RIS* a été calculé classiquement en calculant le ratio du nombre de cas observés rapporté au nombre de cas attendus, et ce pour chacune des unités géographiques considérées (cf. chapitre 2 encadré 2.2).

Cette analyse a été faite uniquement à l'échelle des IRIS, dans la mesure où à l'échelle des quartiers, d'une part il n'existe aucune information démographique (donc pas de covariables), et d'autre part, plus de 50% des unités géographiques ont une incidence nulle, ce qui cause un problème de représentation sous GeoDa, qui utilise des percentiles. Les *RIS* ont été calculés en fonction (1) de l'âge et du sexe, et (2) du sexe et de la nationalité.

L'analyse des *RIS* calculés permet de mettre en évidence au moins trois classes de taux d'incidence : (1) taux significativement inférieur, (2) taux significativement supérieur, et (3) taux sans différence significative, ce par rapport au taux de référence. Les IRIS pour lesquels le taux d'incidence de tuberculose est significativement supérieur au taux de référence sont ceux où il existe un excès de risque de tuberculose, et ce, relativement aux covariables considérées. Ceci permet donc d'établir une cartographie de l'excès de risque tuberculeux sur l'île de Cayenne.

#### *RIS avec lissage bayésien*

Le but des modèles bayésiens est de trouver un compromis entre les taux standardisés classiques (ici *SIR*) et la moyenne globale de la zone géographique considérée, pour démêler une authentique structure spatiale sous-jacente de la variabilité aléatoire des taux [Colonna, 1997].

Nous avons dans un premier temps procédé à un lissage bayésien empirique sans notion de contiguïté spatiale, ou « lissage global ». Dans ce cas, toutes les observations, mêmes celles éloignées géographiquement, vont influencer avec le même poids la prévision en n'importe quel point du champ. L'estimateur lissé de chaque unité géographique est alors égal à :

$$\theta_i = \frac{O_i + \nu}{E_j + \alpha} \quad \text{avec } O_i \text{ les taux observés, et } E_j \text{ les taux attendus}$$

Cependant, l'examen visuel des variations spatiales de variables montre souvent une contiguïté spatiale de ces variables, avec des valeurs semblables observées dans des zones géographiquement proches, et non une distribution

aléatoire des valeurs sur la carte. Nous avons donc dans un deuxième temps appliqué un lissage bayésien empirique avec une notion de contiguïté spatiale, ou « lissage local », qui permet de diminuer l'impact sur l'interpolation pour une région donnée des observations éloignées de cette région. Les fichiers géoréférencés correspondant aux IRIS et aux quartiers ont été exportés de GeoConcept pour être importés sous GeoDa.

### 1.3.2. Résultats

#### *RIS sans lissage bayésien*

Les ratios standardisés d'incidence ont été calculés à l'échelle des IRIS, avec différents groupes de covariables (âge et sexe d'une part, et sexe et nationalité d'autre part). Ces résultats sont résumés sur les cartographies des *RIS* sur l'île de Cayenne (figure 4.5), et ils permettent d'identifier les zones d'excès de risque tuberculeux.

Le tableau 4.1 récapitule les données relatives aux IRIS pour lesquels il existe un excès de risque pour la tuberculose (IRIS pour lesquels  $RIS > 1$  et  $p < 0,005$ ). Les IRIS « Eau Lisette » et « Leblond » apparaissent quelles que soient les covariables prises en compte.

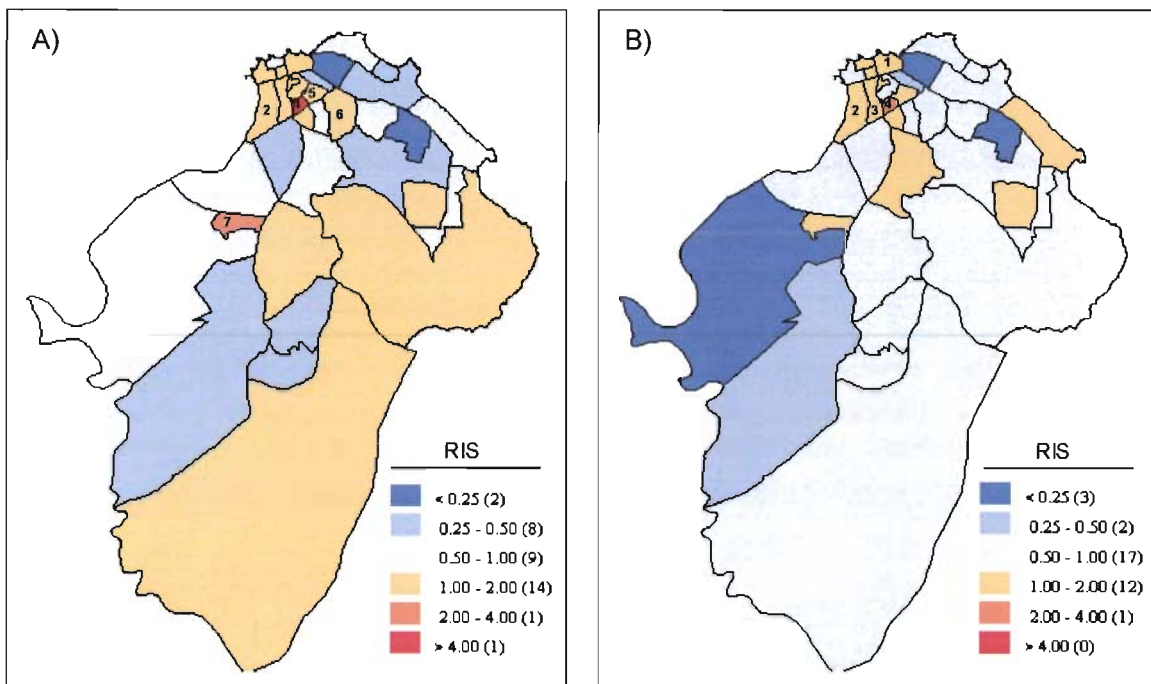


Figure 4.5 – Cartographies sur l'île de Cayenne des ratios standardisés d'incidence (*RIS*) à l'échelle des IRIS, en tenant compte (A) de la répartition par âge et par sexe, et (B) de la répartition par sexe et par nationalité de la population (cartes obtenues avec le logiciel GeoDa). Les IRIS numérotés de 1 à 7 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette ; 5 Mango ; 6 Mont Baduel ; 7 Balata. En fonction des covariables prises en compte, seuls sont indiqués les IRIS où l'on note une sur-incidence de tuberculose (cf. tableau 4.1).

Tableau 4.1 – Ratios standardisés d'incidence (*RIS*) de tuberculose calculés par IRIS, en fonction (a) de la répartition par âge et par sexe, et (b) de la répartition par sexe et par nationalité de la population de l'île de Cayenne. Seuls sont détaillés les résultats des IRIS où l'on note une sur-incidence significative de tuberculose (*RIS* significativement supérieur à 1).

a)	Attendu (E)	Observé (O)	<i>RIS</i> (O/E)	$p^a$	IC 95% <sup>b</sup>
Balata	8,47	17	2,01	0,003	[1,17 - 3,21]
Eau Lisette	5,26	22	4,18	< 0,001	[2,62 - 6,33]
Leblond	14,44	28	1,94	< 0,001	[1,29 - 2,80]
Mango	10,10	19	1,88	0,005	[1,13 - 2,94]
Mont Baduel	17,03	28	1,64	0,008	[1,09 - 2,38]

b)	Attendu (E)	Observé (O)	<i>RIS</i> (O/E)	$p^a$	IC 95% <sup>b</sup>
Buzaret	6,09	12	1,97	0,017	[1,02 - 3,44]
Eau Lisette	8,67	19	2,19	< 0,001	[1,32 - 3,42]
Galmot	8,57	16	1,87	0,011	[1,07 - 3,03]
Leblond	15,91	28	1,76	0,002	[1,17 - 2,54]

<sup>a</sup>  $p$ , degré de significativité du test unilatéral du test de comparaison  $RIS > 1$ .

<sup>b</sup> IC 95%, limites inférieure et supérieure de l'intervalle de confiance à 95%.

### *RIS avec lissage bayésien*

Les figures 4.6 et 4.7 montrent les résultats lissés obtenus à l'échelle des IRIS avec différentes covariables, lorsque l'on tient compte de l'autocorrélation entre unités géographiques proches (figures B) ou non (figures A).

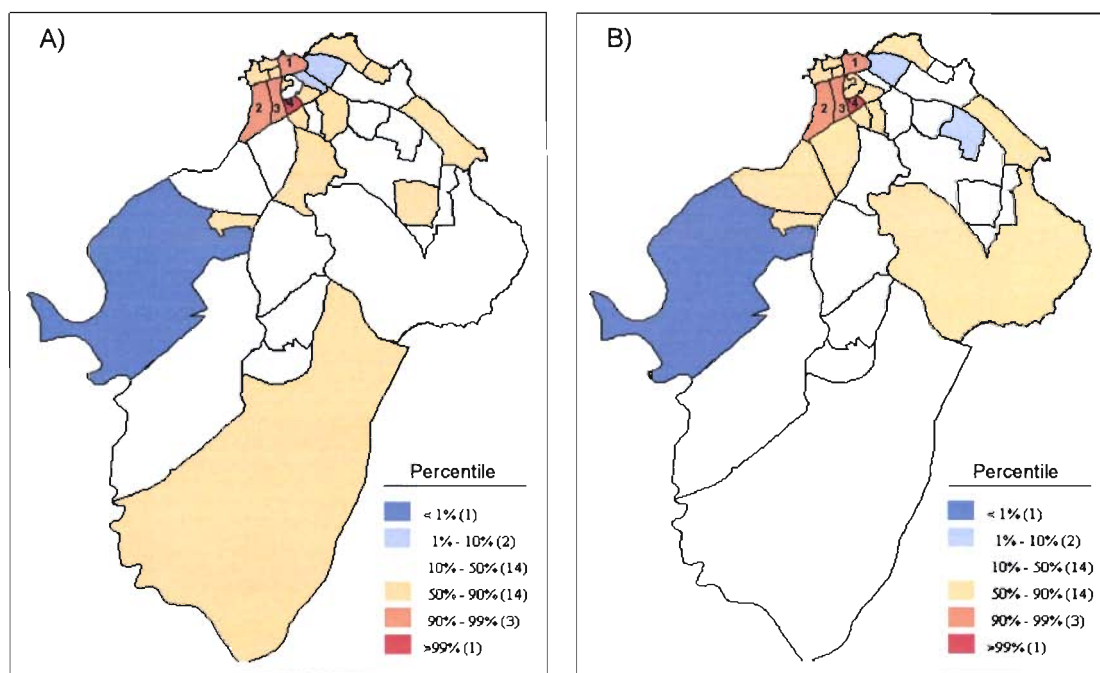


Figure 4.6 – Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés avec les covariables sexe et nationalité à l'échelle des IRIS, (A) sans notion de contiguïté spatiale, et (B) en tenant compte de la contiguïté spatiale. Résultats obtenus avec le logiciel GeoDa. Les IRIS numérotés de 1 à 4 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette.

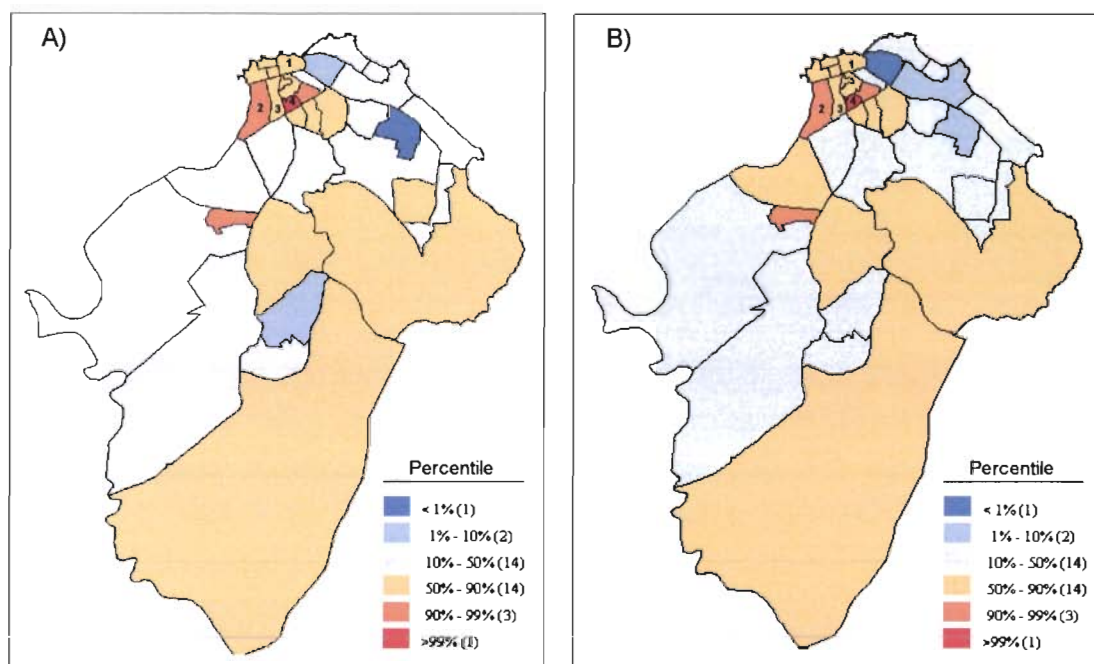


Figure 4.7 – Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés avec les covariables âge et sexe à l'échelle des IRIS, (A) sans notion de contiguïté spatiale, et (B) en tenant compte de la contiguïté spatiale. Résultats obtenus avec le logiciel GeoDa. Les IRIS numérotés de 1 à 4 sont : 1 Buzaret ; 2 Leblond ; 3 Galmot ; 4 Eau Lisette.

La figure 4.8 montre les résultats lissés obtenus à l'échelle des quartiers.

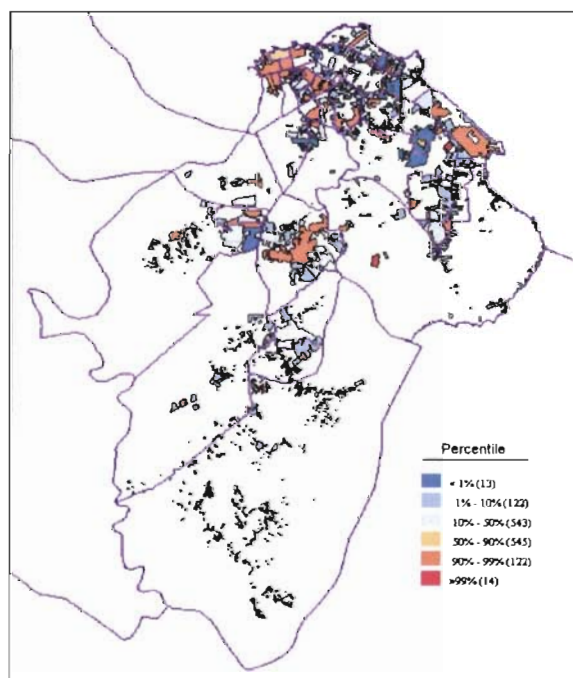


Figure 4.8 – Cartographie sur l'île de Cayenne des taux d'incidence de tuberculose lissés à l'échelle des quartiers. Seul le résultat sans notion de contiguïté spatiale (lissage global) est présenté. En effet, dans la mesure où tous les quartiers ne sont pas contigus et où il existe des espaces non habités, il n'est pas possible d'appliquer la seconde méthode.

### **1.3.3. Discussion**

Les estimations lissées spatialement apportent une solution au problème de la dispersion et elles conviennent donc très bien dans les évaluations de la variation entre secteurs géographiques. Toutes les méthodes de lissage, y compris celle utilisée dans cette étude, comportent une part d'arbitraire [Lawson *et al.*, 1999]. Les méthodes que nous avons choisies tiennent compte ou non du voisinage des secteurs étudiés, et intègrent différents groupes de covariables. Ces choix influent sur la valeur lissée de l'incidence et doivent être pris en considération dans l'examen des variations selon le secteur.

Les estimations brutes et les estimations lissées diffèrent considérablement. Les résultats varient également en fonction des covariables prises en compte. Il est cependant intéressant de noter que certains IRIS apparaissent de façon systématique comme étant en sur-incidence de tuberculose, en particulier les IRIS « Eau Lisette » et « Leblond ».

L'échelle des quartiers pose différents problèmes susceptibles de limiter la pertinence de cette échelle : (1) aucune information n'a pu être introduite en tant que covariable, (2) le lissage intégrant la notion de voisinage n'a pas pu être fait, et (3) plus l'unité spatiale est petite, avec peu d'habitants, plus l'erreur théorique est grande. Malgré tout, l'information lissée obtenue est plus pertinente que la cartographie de l'incidence brute, et permet de mettre en évidence les quartiers de l'île de Cayenne en sur-incidence de tuberculose.

## **1.4. Améliorer ces résultats**

Sont présentées ici de façon succincte certaines recherches déjà menées sur l'évaluation de la densité de population et de la typologie d'habitats sur l'île de Cayenne, recherches susceptibles de nous fournir des données intéressantes pour corriger nos précédentes analyses. Certaines pistes de travail visant à améliorer nos résultats sont explorées. Malgré les difficultés méthodologiques auxquelles nous nous sommes heurtées, nous développons ici la démarche, ainsi que l'intérêt des analyses engagées pour répondre à ce type de problématique.

### **1.4.1. Des méthodes fiables validées : revue de détails**

#### **1.4.1.1. Estimation de l'incidence par télédétection**

L'utilisation de la télédétection dans le cadre d'études démographiques se base sur l'hypothèse d'un lien existant entre la population et la morphologie urbaine [Lo, 1995 ; Chen *et al.*, 2000], c'est-à-dire entre la densité du bâti et la

densité de population. Les outils de télédétection permettent également de discriminer précisément les zones de paysages naturels des zones urbanisées.

À titre d'exemple, nous présentons un travail de recherche ayant intégré l'utilisation d'outils de télédétection pour cartographier une estimation de l'incidence de fièvre Q, sur l'île de Cayenne [Faure *et al.*, 2003 ; Tran *et al.*, 2002] : dans un premier temps, les chercheurs ont procédé à une classification supervisée à l'issue de laquelle les résultats ont été réorganisés sous forme binaire, les pixels de l'image classés urbains prenant la valeur 1 et les autres la valeur 0. Ensuite, un indice de densité de population a été calculé à partir de la densité urbaine en chaque point de l'image, évaluée en analysant les pixels voisins dans un rayon de 200 mètres (ainsi un pixel en périphérie de la « tache urbaine » aura un indice de densité de population inférieur à un pixel en plein centre de la tache). Un indice de la densité de cas incidents a ensuite été calculé par pixel en appliquant la même procédure. L'incidence est égale au rapport entre la densité de cas incidents et la densité de population au même pixel. Une interpolation<sup>4</sup> a ensuite permis d'obtenir un gradient de l'incidence que l'on peut cartographier [Tran *et al.*, 2002].

#### **1.4.1.2. Classification du bâti par télédétection**

L'extraction du thème bâti peut également se faire par morphologie mathématique. Cette méthode tient compte de la forme, du contraste, de la couleur et de la disposition des objets dans l'image [Bonn et Rochon, 1992], et permet, entre autres, de classer les bâtiments en fonction de leur taille. La structure géométrique et la nature du revêtement de la voirie sont également des bons indicateurs du type d'habitat [Gardel, 2000]. Par exemple, les habitats spontanés, c'est-à-dire « réalisés en dehors des règles d'urbanisme et en auto-construction par leurs habitants » [DDE-AU&S, 1997], sont souvent associés à des voiries en latérite, dont le tracé est sinueux.

Associé à un indice de densité de bâti (défini ci-dessus, §1.3.1.1.), cela permet d'obtenir une typologie précise de l'habitat tenant compte de la densité (habitat dense, diffus) et des conditions socio-économiques (habitat spontané, résidentiel ou mixte) [Gardel, 2000].

### **1.4.2. Typologie d'habitat et incidence : nos résultats**

#### **1.4.2.1. Typologie d'habitat**

##### *Typologie par quartier*

Il existe déjà des travaux de caractérisation de l'habitat de l'île de Cayenne en 1987 et 1998 [Gardel, 2001]. La typologie d'habitat de cette précédente

---

<sup>4</sup> Action consistant à ajouter des valeurs ou des termes intermédiaires dans une série de valeurs.

étude, établie en accord avec la DDE de Guyane (Direction Départementale de l'Équipement) a été reprise, à l'exclusion des zones non habitées (zones industrielles et d'activités, et grands équipements). Les classes retenues sont donc les suivantes : (1) habitat spontané, (2) habitat collectif, (3) habitat mixte, (4) habitat résidentiel dense, (5) habitat résidentiel diffus, et (6) habitat dispersé. Notre typologie d'habitat a été réalisée sur la base des contours de bâti obtenus par photo-interprétation (§1.2.2.1).

Partant de la typologie de 1998 sous GeoConcept obtenue auprès de A. Gardel, nous avons réactualisé la typologie de l'Île de Cayenne par photo-interprétation de l'image SPOT-5 du 29 août 2003, ainsi que par une confirmation éventuelle sur le terrain, lors de la présence de nuages sur l'image par exemple (figure 4.9 - B).

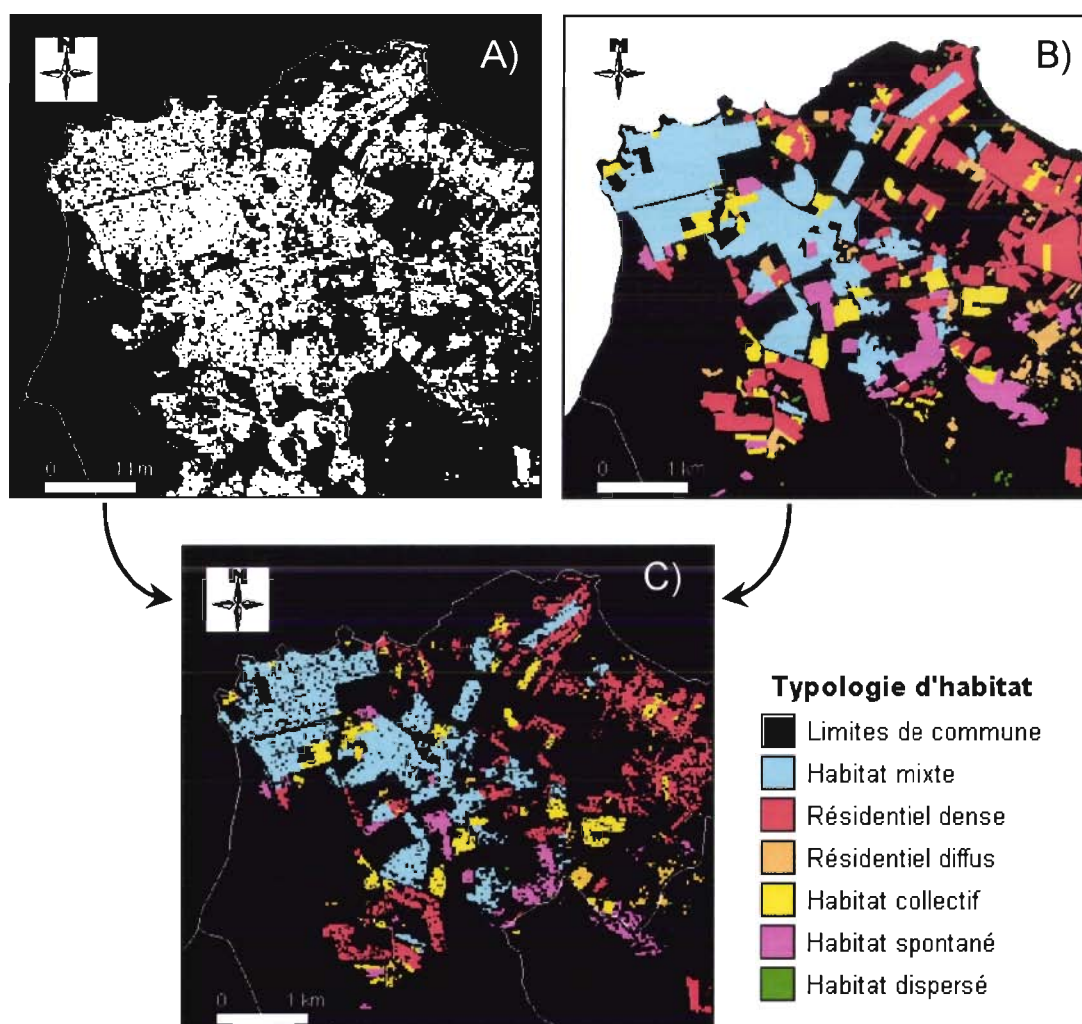


Figure 4.9 – La commune de Cayenne. (A) Image binaire du bâti (valeur 1 pour le bâti, et 0 pour le non bâti, respectivement en blanc et en noir) [Tran *et al.*, 2002], (B) typologie d'habitat 2003 par quartier (réalisée au cours de cette étude), et (C) typologie d'habitat 2003 par pixel, obtenue en appliquant le masque (A) sur l'image (B).

### *Typologie par pixel*

Les polygones de bâti qui sont la base de notre analyse typologique ne correspondent pas aux surfaces réellement bâties, dans la mesure où ils incluent jardins, cours intérieures, etc. Afin d'obtenir un calcul plus juste de la densité de bâti, dont dépendra la densité de population, nous avons donc récupéré la classification précise, sous forme binaire, du bâti de l'île de Cayenne réalisée au cours de l'étude de Tran *et al.* (2002) (figure 4.9 - A), sachant que la taille des pixels est de 20m x 20m. Cette typologie binaire est utilisée pour appliquer un masque à la typologie d'habitat que nous avons réalisé, ce qui permet d'avoir une bonne précision des zones bâties, auxquelles nous ajoutons la précision de typologie d'habitat (figure 4.9).

Dans la mesure où la fragmentation de l'habitat est ici prise en compte par l'exclusion des pixels non bâtis, les classes habitat résidentiel dense, diffus et dispersé ont ensuite été regroupées en une seule classe « habitat planifié ».

#### **1.4.2.2. Calcul de la population par pixel**

Comme nous l'avons dit plus haut, il faut corriger nos données pour que, par exemple, la densité de population attribuée à un pixel d'habitat collectif soit supérieure à celle d'un pixel d'habitat résidentiel.

Les données dont nous disposons sont la population totale de l'IRIS, et le nombre de pixels de chaque type d'habitats dans cet IRIS. Il s'agit donc pour nous de redistribuer la population totale de l'IRIS à chacun des pixels habités, en tenant compte du type d'habitat. Il faut pour cela calculer ce que nous avons appelé un « coefficient de densité » pour chaque type d'habitat. Par exemple, il y a 10 fois plus d'habitants dans un pixel collectif que dans un pixel résidentiel, les « coefficients de densité » correspondant à ces deux types d'habitat sont respectivement de 10 et 1. Pour trouver ces coefficients, il faut résoudre une matrice d'équations, que l'on peut écrire sous la forme :

$$P_i = p_{i1} \cdot C_1 + p_{i2} \cdot C_2 + p_{i3} \cdot C_3 + p_{i4} \cdot C_4$$

avec  $P_i$  la population de l'IRIS  $i$ ,

$C_1$  à  $C_4$  les « coefficients de densité » des types d'habitat 1 à 4, et,

$p_{i1}$  à  $p_{i4}$  le nombre de pixels de chaque type d'habitat 1 à 4 dans l'IRIS  $i$ .

Au final, le système d'équations comprend donc 35 équations, correspondant aux 35 IRIS.  $P$  et  $p$  sont connues pour chaque IRIS et les inconnues du système sont  $C_1$  à  $C_4$ . Le système a été résolu grâce à la méthode des moindres carrés, et les résultats obtenus sont les suivants :

Coefficient habitat collectif	$C_1 = 7,13$
Coefficient habitat mixte	$C_2 = 4,14$
Coefficient habitat spontané	$C_3 = 4$
Coefficient habitat planifié	$C_4 = 1,63$

### Encadré 4.3 – La méthode des moindres carrés

La méthode des moindres carrés permet de comparer des données expérimentales, souvent entachées d'erreurs de mesure, à un modèle mathématique censé décrire ces données. Dans le cas le plus courant, ce modèle est une famille de fonctions  $f(x, \alpha)$  d'une ou plusieurs variables muettes  $x$ , indexées par un ou plusieurs paramètres  $\alpha$  inconnus. La méthode des moindres carrés permet de sélectionner parmi ces fonctions, celle qui reproduit le mieux les données expérimentales. On parle dans ce cas d'ajustement par la méthode des moindres carrés, et donne une estimation indirecte de la valeur de ces paramètres.

L'hypothèse posée est que la fonction  $f(x, \alpha)$  qui décrit "le mieux" les données est celle qui minimise la somme quadratique des déviations des mesures aux prédictions de  $f(x, \alpha)$ . Si par exemple nous disposons de  $N$  mesures  $(y_i)$  ( $i = 1 \dots N$ ), les paramètres  $\alpha$  "optimaux" au sens de la méthode des moindres carrés sont ceux qui minimisent la quantité :

$$K = \sum_{i=1}^N y_i - f(x, \alpha))^2$$

#### *Population réelle par quartier*

Il s'agit maintenant de repasser à l'échelle des quartiers. La population par quartier peut être obtenue en multipliant le coefficient d'un type d'habitat au nombre de pixels du type d'habitats correspondant, puis en faisant la somme des produits obtenus pour les 4 types d'habitats. C'est ici que l'on se heurte à la manipulation des formats d'image vecteur et raster (*cf.* définitions ci-dessous), et aux difficultés liées à leur interaction dans un SIG.

Les données raster sont analogues au mode image, avec une structure composée de pixels individuels auxquels sont associées une valeur représentant un attribut (dans le cas qui nous préoccupe, le type d'habitat). Les données vecteur (ou vectorielles) sont représentées sous la forme de points, de lignes et de polygones. Lorsque nous avons travaillé sur les pixels, nous avons perdu l'information sur les polygones, définissant nos quartiers, et nous n'avons pas réussi par la suite à croiser ces deux informations, ce qui aurait permis d'aller jusqu'au bout de cette démarche. Nous illustrons ce problème par le schéma de la figure 4.10.

Il existe certainement des solutions alternatives, comme appliquer une grille reproduisant les pixels de 20 mètres sur 20 mètres, et calculer dans chaque carré de la grille la proportion de surface occupée de chacun des types d'habitat. Mais nous nous sommes finalement limité à cette étape car d'autres problèmes ont été soulevés au cours de notre étude : l'habitat dispersé, trop petit, a tendance à disparaître lors de la transformation des polygones en pixels, la classification du bâti de Tran *et al.* (2002) a été faite sur une image

SPOT de 1998, et la matrice de 1203 x 1199 pixels devient difficile à manipuler lorsque l'on veut faire des calculs statistiques.



Figure 4.10 – Schéma illustrant les difficultés inhérentes à la manipulation simultanée des formats d'image vecteur (le quartier) et raster (les pixels d'habitats). Le but est de compter le nombre de pixels de chaque couleur inclus dans le quartier. Visuellement, la solution est simple, mais pas de façon automatisée car les SIG manipulent rarement les deux formats.

### *Perspectives*

Il y a certainement malgré tout des pistes à explorer pour améliorer les estimations des recensements. Ceci est d'autant plus important que les données de population de l'INSEE sont bien souvent sous-estimées, d'une part parce que les populations sont très mobiles, et d'autre part parce que certains quartiers sont difficiles d'accès [Gardel, 2000]. Les résultats des coefficients de densité que nous avons obtenus confirment cette idée, dans la mesure où il est probable que la densité de population en zone d'habitat spontané est supérieure à la densité en zone d'habitat mixte, qui regroupe à la fois de l'habitat spontané et de l'habitat résidentiel.

Cependant, la pertinence du choix d'une échelle aussi fine que celle des quartiers ayant été discutée précédemment (*cf.* §1.3.3.), et les estimations de population par quartier ayant donné lieu à des difficultés méthodologiques, il serait préférable à l'avenir de travailler à l'échelle des îlots (*cf.* définition page 91), intermédiaires entre les IRIS et les quartiers, et pour lesquels l'INSEE dispose d'informations démographiques. Malheureusement, ces données n'ont pas été accessibles au moment de l'étude.

## **1.5. Bilan préliminaire**

Les résultats obtenus dans cette première partie semblent prédire une distribution spatiale de la tuberculose sur l'île de Cayenne qui ne dépendrait pas uniquement de celle de la population. Afin d'identifier les sources de contamination potentielles de tuberculose, nous avons donc cherché à vérifier statistiquement, s'il existe des agrégats significatifs de cas, et à identifier où se trouvent ces zones agrégées, ce qui est l'objet de la suite de ce chapitre.

## 2. Identification des agrégats spatiaux

### 2.1. Définition d'un agrégat

Lorsque l'on s'intéresse à la surveillance des maladies, un agrégat (en anglais « *cluster* ») peut se définir comme un groupement de cas malades géographiquement proches, de taille et de concentration suffisante pour qu'il y ait peu de chance qu'il soit uniquement dû au hasard. Cet agrégat peut également être référencé dans le temps, pour détecter les pics épidémiques [Kulldorff et Nagarwalla, 1995].

### 2.2. Statistiques appliquées à la recherche d'agrégats

#### 2.2.1. État de l'art

De très nombreuses méthodes statistiques ont été développées pour tester si le patron de distribution des cas associés à une maladie est uniquement dû au hasard [Besag et Newell, 1991 ; Bithell, 1999 ; Cliff et Ord, 1973 ; Cuzick et Edwards, 1990 ; Diggle et Chetwynd, 1991 ; Kulldorff, 1999 ; Martuzzi et Hills, 1995 ; Tango, 1995 ; Tango, 2000 ; Turnbull *et al.*, 1990 ; Walter, 1994]. Afin d'évaluer la robustesse statistique de ces multiples méthodes, Song et Kulldorff (2003) ont procédé à une analyse comparative de huit d'entre elles. Leurs résultats montrent que la majorité de ces tests (6 sur 8) ont une bonne capacité de détection des agrégats de cas, à condition d'inclure les paramètres appropriés, et de bien caractériser à priori la démarche épistémologique que l'on souhaite adopter.

#### 2.2.2. Comparaison de 3 méthodes

À titre d'exemple, citons les résultats de l'étude de Perez *et al.* (2002) comparant trois méthodes pour la détection des agrégats de tuberculose bovine en Argentine : l'autocorrélation spatiale, le test de Cuzick et Edwards, et la méthode « *spatial scan statistic* ».

(1) L'inconvénient de l'autocorrélation spatiale [Moran, 1950 ; Cliff et Ord, 1973] est qu'elle ne tient pas compte de la distribution hétérogène naturelle des populations auxquelles on s'intéresse, ici les cas de tuberculose bovine, qui correspond pourtant à une réalité biologique.

(2) En revanche, le test de Cuzick et Edwards pour les populations non-homogènes [Cuzick et Edwards, 1990] tient compte de cette hétérogénéité des

cas en comparant la distribution des cas malades à celle de cas contrôle. Ce test donne des informations sur l'existence probable d'agrégats spatiaux.

(3) La méthode « *spatial scan statistic* » [Kulldorff et Nagarwalla, 1995] tient également compte de l'hétérogénéité de la population et détecte la présence d'agrégats spatiaux. Cette méthode présente en outre l'avantage de localiser précisément ces agrégats [Odoi *et al.*, 2004 ; Viel *et al.*, 2000]. C'est cette dernière méthode que nous avons choisie d'appliquer à notre étude sur la tuberculose humaine.

## 2.3. La méthode « *spatial scan statistic* »

### 2.3.1. Le logiciel SaTScan

#### *Généralités*

Le logiciel SaTScan [Kulldorff and Information Management Services, Inc., 2005] a été développé pour l'application de la méthode « *spatial scan statistic* » spécifiquement aux études épidémiologiques. Il permet d'identifier les agrégats de cas d'une maladie sous l'hypothèse nulle ( $H_0$ ) que l'excès de cas observé dans une certaine zone géographique est uniquement dû au hasard. Un balayage horizontal de toute la zone d'étude est réalisé grâce une fenêtre circulaire dont le diamètre varie de façon continue entre zéro et une valeur limite fixée par l'utilisateur, d'où le nom de « balayage spatial » également donné à cette méthode. La fenêtre étant placée successivement au centroïde de chaque unité spatiale, la variation de la taille de fenêtre permet d'inclure les unités spatiales voisines à plus ou moins longue distance, les agrégats pouvant donc apparaître à différentes échelles.

#### *Paramétrer le modèle statistique*

Le logiciel permet de choisir le modèle de distribution théorique auquel l'on va comparer la distribution observée des cas, soit la loi de Poisson, soit la loi de Bernoulli. Lorsque l'on choisit la loi de Poisson, les points théoriques sont générés selon un « processus de Poisson inhomogène », c'est-à-dire qui varie en fonction des populations à risque, et donc de l'espace. La loi de Bernoulli est basée sur la comparaison cas/contrôles et nécessite la localisation d'un groupe de cas et d'un groupe d'individus contrôle. Dans l'étude de Perez *et al.* (2002) sur la tuberculose bovine par exemple, on compare les cheptels infectés et non infectés. La loi de Poisson suppose une distribution continue des cas selon la loi de Poisson [Kulldorff, 1997]. Sous l'hypothèse nulle que les cas sont répartis de façon aléatoire dans l'espace, le risque est le même pour toutes les fenêtres.

*Paramétrer les covariables*

Lorsque aucune covariable n'est introduite dans le modèle, le nombre de cas attendu (ici, de tuberculose) dans la fenêtre d'observation est proportionnel à la population contenue dans cette fenêtre : cette distribution « idéale » constitue l'hypothèse nulle  $H_0$ . L'utilisateur peut également spécifier certaines covariables à prendre en compte dans le calcul du nombre de cas attendus [Kulldorff, 1997], ce qui permet donc d'intégrer l'hétérogénéité de distribution de la population, comme nous l'avions signalé plus haut. Dans le cas d'une maladie gériatrique par exemple, la distribution spatiale des individus de plus de 60 ans peut être importante à prendre en compte. En effet, si leur répartition sur le territoire n'est pas homogène, c'est-à-dire si cette population est sur-représentée dans certaines régions données, on pourrait voir apparaître dans ces zones des agrégats de cas simplement corrélés à la présence de cette population âgée.

Une bonne connaissance "à priori" des facteurs pouvant influencer l'analyse spatiale des cas est donc très importante, afin de pouvoir les inclure dans l'analyse et ainsi éviter les biais. Dans notre étude sur la tuberculose en Guyane, l'analyse préliminaire d'épidémiologie développée au chapitre 2 nous a permis de mettre en évidence certains de ces facteurs. Leur choix est développé dans la partie concernant l'application de la méthode à nos données (§2.2.2.1.)

**Encadré 4.4 – Le rapport du maximum de vraisemblance**

Pour chaque localisation et taille de la fenêtre glissante, l'hypothèse nulle ( $H_0$ ) suppose que pour chaque point la probabilité  $p$  d'être dans la fenêtre est égale à la probabilité  $q$  d'être à l'extérieur de cette fenêtre, alors que l'hypothèse alternative ( $H_1$ ) suppose que le risque à l'intérieur de la fenêtre est plus élevé qu'à l'extérieur, tous les cas inclus dans cette fenêtre constituant alors un agrégat statistique.

La fonction de vraisemblance étudiée est différente en fonction du modèle de distribution choisi, c'est-à-dire la loi de Poisson ou celle de Bernoulli. Notre étude propre ne concernant pas une étude cas-contrôle, nous avons choisi de ne développer ici que les équations associées au modèle de Poisson. Sous cette hypothèse, la fonction du maximum de vraisemblance pour une fenêtre donnée est proportionnelle à :

$$\left( \frac{n_z}{\mu(Z)} \right)^{n_z} \left( \frac{N - n_z}{N - \mu(Z)} \right)^{N - n_z} I(.)$$

où  $N$  est le nombre total de cas dans l'unité spatiale considérée,  
 $n_z$  est le nombre de cas dans le cercle  $Z$  (la fenêtre de glissement),  
 $\mu(Z)$  est le nombre de cas attendus à l'intérieur de la fenêtre sous l'hypothèse nulle, ajusté en fonction des covariables introduites.

$I(.)$  est une fonction indicatrice<sup>5</sup>, égale à 1 lorsque la fenêtre recense plus de cas observés que de cas attendus sous  $H_0$ , ou à 0 dans les autres cas.

L'agrégat le plus probable correspond au cercle pour lequel on obtient le maximum des rapports de vraisemblance. Soit  $L(Z)$  le maximum de vraisemblance du cercle  $Z$ ,  $\mathcal{E}$  l'ensemble des fenêtres possibles,  $L_0$  la vraisemblance sous l'hypothèse  $H_0$ . Le maximum de vraisemblance de chaque fenêtre est obtenu par la fonction :

$$S_{\mathcal{E}} = \max_{Z \in \mathcal{E}} \frac{L(Z)}{L_0}$$

$$\text{avec } \frac{L(Z)}{L_0} = \left( \frac{n_z}{\mu(Z)} \right)^{n_z} \left( \frac{N - n_z}{N - \mu(Z)} \right)^{N - n_z} \quad \text{si } n_z > \mu(Z) ; \text{ sinon } \frac{L(Z)}{L_0} = 1$$

La valeur du degré de significativité est obtenue par itération du rapport du maximum de vraisemblance sous  $H_0$ , à l'aide d'une simulation de Monte Carlo. Le nombre d'itérations est choisi par l'utilisateur, le minimum conseillé étant 9999 [Kulldorff *et al.*, 1998]. L'ensemble de la méthode est détaillé dans [Kulldorff, 1997].

## 2.3.2. Application de la méthode à nos données

### 2.3.2.1. Les données utilisées

Le logiciel SaTScan fonctionne avec trois fichiers : un fichier « cas » avec le nombre de cas par unité spatiale, tenant compte des covariables ; un fichier « population » indiquant la population totale par unité spatiale, en fonction des covariables ; un fichier « coordonnées » avec les coordonnées géographiques des unités spatiales (coordonnées au centroïde de l'unité, par exemple un IRIS).

Nous avons choisi de mettre de côté l'aspect temporel de l'analyse et avons cherché à mettre en évidence des agrégats spatiaux, plutôt que des agrégats spatio-temporels. Dans le fichier « cas », nous avons inclus l'ensemble des cas recensés sur l'île de Cayenne entre 1996 et 2003, sans tenir compte de l'année de déclaration afin d'éviter une dilution de l'information (même si les articles explicatifs du logiciel SaTScan ne précisent aucun nombre minimum de cas à entrer). Les populations par nationalité, par sexe et par tranches d'âge de 5 ans (de 0 à 100 ans et plus) correspondent au recensement 1999 de l'INSEE.

Nous avons d'autre part testé différentes échelles spatiales (IRIS versus quartiers) et différentes covariables (âge, sexe, nationalité) dans la recherche d'agrégats de tuberculose sur l'île de Cayenne : (1) agrégats sur les IRIS, par sexe et par âge, (2) agrégats sur les IRIS, par sexe et par nationalité, (3) agrégats sur les quartiers, sans covariable.

<sup>5</sup> Si A et B sont deux ensembles tels que A est inclus dans B, on appelle fonction indicatrice de A la fonction  $I_A$  définie de B dans  $\{0,1\}$ , et telle que :  $I_A(x) = 1$ , si x est dans A ou  $I_A(x) = 0$ , si x n'est pas dans A.

### *Remarques*

Dans le test (2) concernant le test d'agrégation sur les IRIS, par sexe et par nationalité, les patients pour lesquels nous avons une incertitude sur la nationalité ont été éliminés, soit 51 patients sur les 366 référencés (14%). Les données de nationalité de l'INSEE étant détaillées par sexe, mais pas par âge, nous n'avons pas pu tester l'agrégation en tenant compte simultanément de l'âge, du sexe et de la nationalité, d'où les deux tests effectués.

D'autre part, la covariable « nationalité » n'inclut en fait que deux groupes, la population de nationalité étrangère et la population de nationalité française, car les données de population INSEE ne sont détaillées que pour ces deux groupes, et pas pour chaque nationalité représentée en Guyane. La covariable « nationalité » mérite malgré tout d'être testée étant donné l'existence d'une forte ségrégation spatiale entre les différents groupes nationaux sur l'île de Cayenne [Gardel, 2000]. On parle, par exemple, de « Little Haïti » ou du « Village brésilien » pour désigner certains quartiers [Gardel, 2001].

#### **2.3.2.2. Les paramètres d'analyse choisis**

L'analyse est purement spatiale ; le modèle choisi est la loi de Poisson ; la recherche est basée sur la recherche d'agrégats ayant de forts taux de maladie (on peut choisir de mettre en évidence les faibles taux).

Puisque l'on cherche à mettre en évidence les aires (les cercles) où le taux de maladie est exceptionnellement élevé (plutôt que faible), la taille maximale choisie pour la fenêtre de glissement correspond à la taille pour laquelle 50% de la population incluse dans la fenêtre est « à risque » ; la taille minimale de fenêtre est zéro.

Pour chacune des analyses sous SaTScan, nous avons lancé 9999 itérations du test de Monte Carlo correspondant au minimum d'itérations conseillé par [Kulldorff *et al.*, 1998] pour assurer l'indépendance des valeurs échantillonnées (voir aussi [Manly, 1991] pour le choix du nombre d'itérations).

## **2.4. Résultats**

### **2.4.1. Dans les 35 IRIS, par âge et par sexe**

#### **2.4.1.1. Agrégats significatifs**

Pour cette analyse, un agrégat est reconnu comme significatif lorsque le  $\log_{10}$  du rapport de vraisemblance calculé (noté LRV, correspondant à  $L(Z)/L_0$ ) est supérieur ou égal à 7,406 au seuil de 0,01, ou à 5,742 au seuil de 0,05 (ces seuils sont donnés par SaTScan avec l'ensemble des résultats d'analyse).

La méthode met en évidence 5 agrégats de tuberculose, dont un seul est significatif (LRV = 27,99 ;  $p = 0,0001$ ) (tableau 4.2).

Tableau 4.2 – Unique agrégat significatif de tuberculose, identifié par l'analyse SaTScan. L'IRIS « Eau Lisette », sur lequel est centré l'agrégat, lui donne son nom.

Agrégat	Rayon <sup>a</sup>	N	obs	att	o/a	LRV	$p$
Eau Lisette	1662,75	12	195	125,06	1,56	27,99	0,0001

<sup>a</sup> Rayon, rayon de la fenêtre de glissement de l'agrégat ; N, nombre d'IRIS inclus dans l'agrégat ; obs, nombre de cas de tuberculose observés dans l'agrégat ; att, nombre de cas de tuberculose attendus dans l'agrégat ; o/a, ratio du nombre de cas observés sur le nombre de cas attendus ; LRV, log du rapport de vraisemblance ;  $p$ , probabilité associée à l'agrégat.

Cet agrégat, nommé « Eau Lisette », regroupe 12 IRIS au total (tableau 4.3). La probabilité associée à cet agrégat, et donc à l'ensemble des 12 IRIS qu'il comprend est  $p = 0,0001$ . Le logiciel SaTScan donne également trois autres variables que nous n'avons pas ajoutées au tableau car elles sont identiques pour les douze IRIS considérés, ces variables étant toujours identiques pour les unités spatiales (IRIS) incluses dans un même agrégat : clu\_obs, nombre de cas de tuberculose observés dans l'agrégat, qui est égal à 195, clu\_att, nombre de cas de tuberculose attendus dans l'agrégat, égale à 125,06 et clu\_o/a, ratio du nombre de cas observés sur le nombre de cas attendus, égal à 1,559.

Tableau 4.3 – Détail des IRIS inclus dans l'unique agrégat significatif, dit « Eau Lisette », avec le nombre de cas de tuberculose attendu (loc\_att), le nombre de cas réellement observé (loc\_obs) et le rapport entre nombre de cas observé sur nombre de cas attendu (loc\_o/a). La probabilité associée à l'agrégat (et donc aux 12 IRIS qu'il comprend) est  $p = 0,0001$ .

IRIS	loc_obs	loc_att	loc_o/a
Eau Lisette	22	5,26	4,182
Bonhomme	13	8,41	1,546
Thémire	16	12,46	1,284
Galmot	17	10,69	1,590
Mango	19	10,10	1,881
Mirza	8	7,01	1,142
Cabassou	9	10,81	0,832
Anatole	4	8,07	0,496
Leblond	28	14,44	1,939
De Gaulle	15	9,46	1,586
Buzaret	16	11,32	1,413
Mont Baduel	28	17,03	1,644

*Remarque*

Un troisième tableau fourni par SaTScan récapitule le nombre de cas observés, le nombre de cas attendus, le ratio de ces deux nombres, et le risque relatif (RR) pour chacun des 35 IRIS. L'information contenue dans ce troisième tableau de SaTScan étant redondante avec celle des deux précédents tableaux, nous ne présentons pas ces résultats.

**2.4.1.2. Cartographie des ratios standardisés d'incidence (RIS)**

Pour rappel (*cf.* chapitre 2), le risque relatif est défini comme le risque respectif des différentes unités spatiales d'un agrégat, par rapport au risque de la population totale [Klassen *et al.*, 2005]. Numériquement, le risque relatif correspond en fait au rapport du nombre de cas observés sur le nombre de cas attendus (*o/a*) sans approximation [comm. pers. J.-F. Viel]. En français, on lui préfère le terme de « ratio standardisé d'incidence » [Viel *et al.*, 2000], car le terme de « risque relatif » recouvre une toute autre définition en épidémiologie (risque relatif d'une population exposée par rapport à une population non exposée à un agent cancérogène, par exemple).

La figure 4.11 illustre le ratio standardisé d'incidence (noté RIS) associé à chaque IRIS de l'agrégat dit « Eau Lisette », qui varie de 0,50 pour l'IRIS Anatole, à 4,18 pour l'IRIS Eau Lisette.

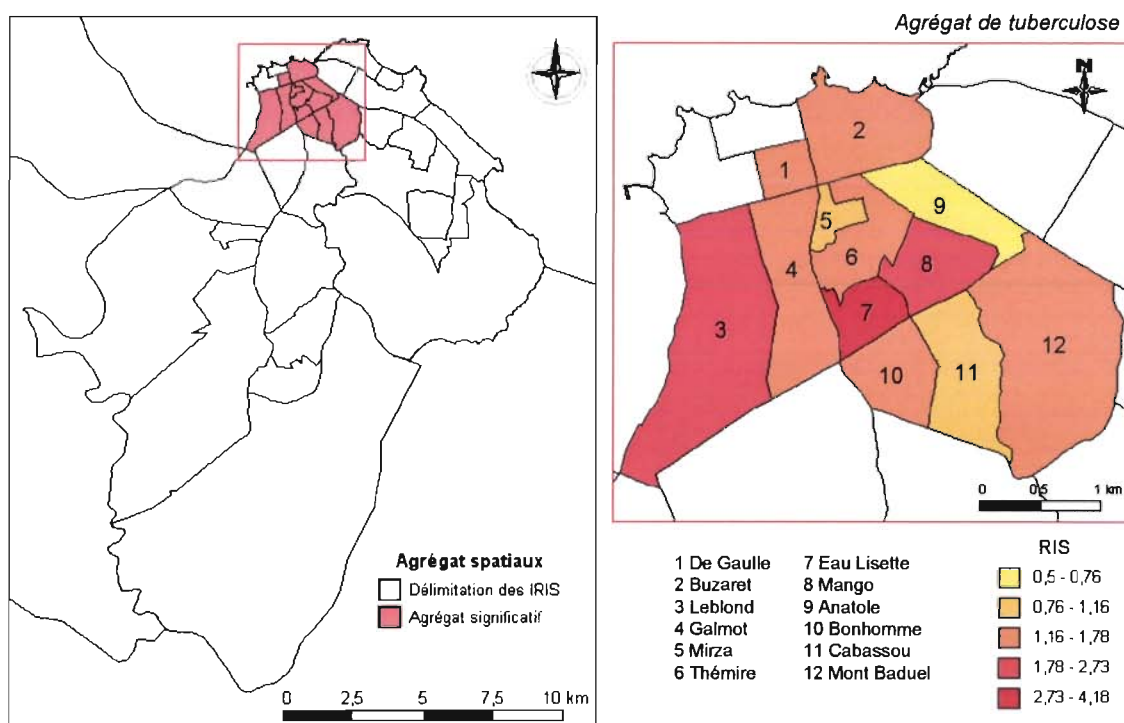


Figure 4.11 – Unique agrégat significatif de tuberculose identifié à l'échelle de l'IRIS en tenant compte des covariables âge et sexe. À droite, détail des IRIS inclus dans cet agrégat. RIS, ratio standardisé d'incidence.

Comme explicité dans le chapitre précédent, le nombre de classes pour la discrétisation des RIS suit l'équation de Huntsberger (formule §1.2.1.1.). La forme de la distribution statistique est asymétrique à gauche, donc on applique une discrétisation par progression géométrique [Rican, 1998] (cf. encadré 4.2 pour le choix de la méthode de discrétisation).

## 2.4.2. Dans les 35 IRIS, par nationalité et par sexe

### 2.4.2.1. Agrégats significatifs

Pour cette analyse, l'agrégat est reconnu significatif lorsque le log du rapport de vraisemblance ( $LRV = L(Z/L_0)$ ) calculé est supérieur ou égal à 7,207 au seuil de 0,01, ou à 5,518 au seuil de 0,05.

La méthode met en évidence trois agrégats de tuberculose (tableau 4.4), dont un seul est significatif ( $LRV = 14,65$  ;  $p = 0,0001$ ) (figure 4.12). Cet agrégat, nommé « Palmistes », regroupe 9 IRIS au total (tableau 4.5).

Tableau 4.4 – Agrégats de tuberculose identifiés par l'analyse SaTScan. Seul l'agrégat nommé « Palmistes », comprenant 9 IRIS, est significatif ( $p = 0,0001$ ).

Agrégats	Rayon <sup>a</sup>	N	obs	att	o/a	LRV	p
Palmistes	1786,20	9	127	82,77	1,53	14,65	0,0001
Montjoly Plage	0	1	12	7,07	1,70	1,46	0,96
Mome Coco-BP134	0	1	10	9,07	1,10	0,05	1,00

<sup>a</sup> Rayon, rayon de la fenêtre de glissement de l'agrégat de cas ; N, nombre d'IRIS inclus dans l'agrégat ; obs, nombre de cas de tuberculose observés dans l'agrégat ; att, nombre de cas de tuberculose attendus dans l'agrégat ; o/a, ratio du nombre de cas observés sur le nombre de cas attendus ; LRV, log du rapport de vraisemblance ; p, probabilité associée à l'agrégat.

Tableau 4.5 – Détail des IRIS inclus dans le seul agrégat de tuberculose significatif, dit « Palmistes ». Le nombre de cas de tuberculose attendu dans l'IRIS (loc\_att), le nombre de cas réellement observé (loc\_obs) et le rapport entre cas observés sur cas attendus (loc\_o/a) sont précisés. La probabilité associée à l'agrégat (et aux 9 IRIS qu'il comprend) est  $p = 0,0001$ .

IRIS	loc_obs	loc_att	loc_o/a
Palmistes	12	12,28	0,977
Les Amandiers	8	6,66	1,201
De Gaulle	14	8,65	1,619
Leblond	28	15,91	1,760
Mirza	7	3,80	1,844
Galmot	16	8,57	1,867
Buzaret	12	6,09	1,970
Thémire	11	12,14	0,906
Eau Lisette	19	8,67	2,191

La probabilité associée à l'agrégat « Palmistes » incluant 9 IRIS (tableau 4.5) est  $p = 0,0001$ . De même que précédemment les variables identiques pour l'ensemble des IRIS inclus dans l'agrégat ne sont pas précisées dans le tableau. Leurs valeurs sont les suivantes : *clu\_obs*, nombre de cas de tuberculose observés dans l'agrégat, qui est égal à 127, *clu\_att*, nombre de cas de tuberculose attendus dans l'agrégat, égale à 82,77, et *clu\_o/a*, ratio du nombre de cas observés sur le nombre de cas attendus, égal à 1,534.

#### 2.4.2.2. Cartographie des ratios standardisés d'incidence (RIS)

Les ratios standardisés d'incidence associés à chaque IRIS de l'agrégat « Palmistes » (figure 4.12), varient de 0,91 pour l'IRIS Thémire, à 2,19 pour l'IRIS Eau Lisette. L'agrégat identifié en tenant compte de la nationalité n'inclut pas les mêmes IRIS que lorsque l'on tient compte de l'âge et du sexe, mais l'IRIS Eau Lisette reste celui ayant le RIS le plus élevé.

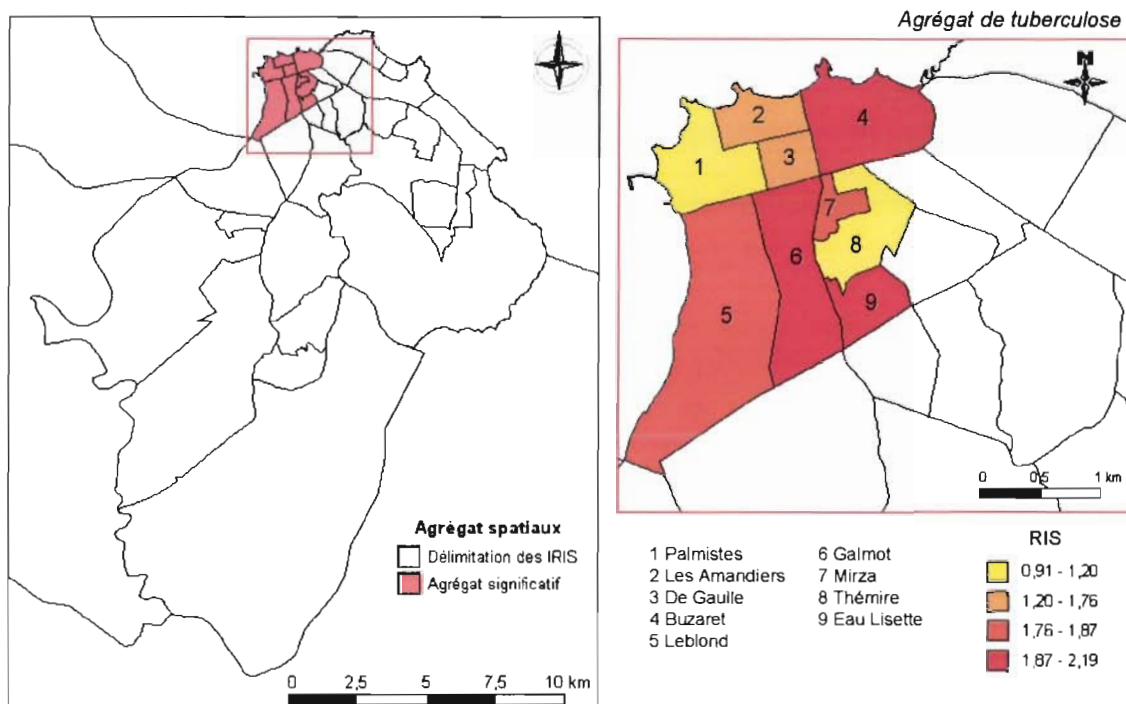


Figure 4.12 – Unique agrégat significatif de tuberculose identifié à l'échelle de l'IRIS en tenant compte des deux covariables nationalité et sexe. À droite, détail des IRIS inclus dans cet agrégat de tuberculose. RIS, ratio standardisé d'incidence.

Le logiciel GeoConcept possède un module de discrétisation des données continues qui permet de déterminer la forme de la distribution statistique des données, ici les RIS. La forme de la distribution statistique suit cette fois une loi normale, et on applique une discrétisation en classes d'amplitudes égales [Rican, 1998] (cf. encadré 4.2 pour le choix de la méthode de discrétisation). L'amplitude des classes est obtenue en divisant l'étendue de la distribution

(différence entre valeurs maximale et minimale) par le nombre de classes (défini grâce à l'équation de Huntsberger). La valeur d'amplitude est ajoutée à la limite inférieure de chaque classe pour obtenir les limites de classes.

### 2.4.3. À l'échelle des quartiers, sans inclure de covariable

L'agrégat est reconnu comme étant significatif lorsque le log du rapport de vraisemblance ( $LRV = L(Z)/L_0$ ) est supérieur ou égal à 10,822 au seuil de 0,01, ou à 8,952 au seuil de 0,05 (seuils fournis par SaTScan). La méthode met en évidence un agrégat principal significatif (n°1) ( $LRV = 28,29$  ;  $p = 0,0001$ ) regroupant 89 « quartiers », et sept agrégats secondaires (n°2 à 8), dont seuls deux sont significatifs (n° 2 et 3) au seuil alpha de 5%. L'ensemble des résultats d'agrégation a été reporté dans le tableau 4.6, sauf ceux concernant 8 agrégats supplémentaires non significatifs notifiés par SaTScan ( $p = 1$ ).

Les agrégats de tuberculose n°1 à 8 sont localisés figure 4.13. L'agrégat n°1 englobe le centre ville historique de Cayenne et les IRIS périphériques. L'agrégat n°2 regroupe les quartiers Ploërmel, Louise Elmiré (dit aussi Source de Baduel) et Troubiran (figure 4.13-B). L'agrégat n°4 (figure 4.13-C) correspond à la prison de Rémire, où l'on a recensé 7 cas de tuberculose sur la période 1996-2003. L'agrégat n°6 correspond au village brésilien dit « BP 134 », incluant 7 cas. L'agrégat n°8 correspond à un agrégat de 4 cas de tuberculose dans une même rue du quartier Vieux Chemin. L'agrégat n°5, constitué de 16 quartiers, s'étend de Balata, à l'ouest, à Cogneau Lamirande, à l'est (figure 4.13-D). Les agrégats n°3 et 7 sont tous deux constitués d'une unique maison dans les environs de Stoupan, chacune regroupant deux cas de tuberculose.

Tableau 4.6 – Agrégats de tuberculose identifiés par l'analyse SaTScan. Sur ces huit agrégats, seuls les agrégats notés 1, 2 et 3 sont significatifs ( $p < 0,05$ ).

Quartier	Agrégat	Rayon <sup>a</sup>	N	obs	att	o/a	LRV	p
1116731	1	1476,75	89	190	120,16	1,58	28,29	0,0001
1117205	2	559,54	26	24	7,73	3,10	11,29	0,0069
1124699	3	0	1	2	<0,001	461,02	10,28	0,0192
1117141	4	0	1	7	0,82	8,54	8,88	0,0641
1116806	5	1485,63	16	29	14,16	2,05	6,26	0,4174
1117223	6	0	1	7	1,28	5,47	6,22	0,4249
1124882	7	0	1	2	0,03	57,63	6,15	0,4541
1116603	8	0	1	4	0,60	6,68	4,21	0,9495

<sup>a</sup> Rayon, rayon en mètres de l'agrégat (supposé circulaire) ; N, nombre d'IRIS inclus dans l'agrégat ; obs, nombre de cas de tuberculose observés dans l'agrégat ; att, nombre de cas de tuberculose attendus dans l'agrégat ; o/a, ratio du nombre de cas observés sur le nombre de cas attendus ; LRV, log du rapport de vraisemblance ; p, probabilité associée à l'agrégat.

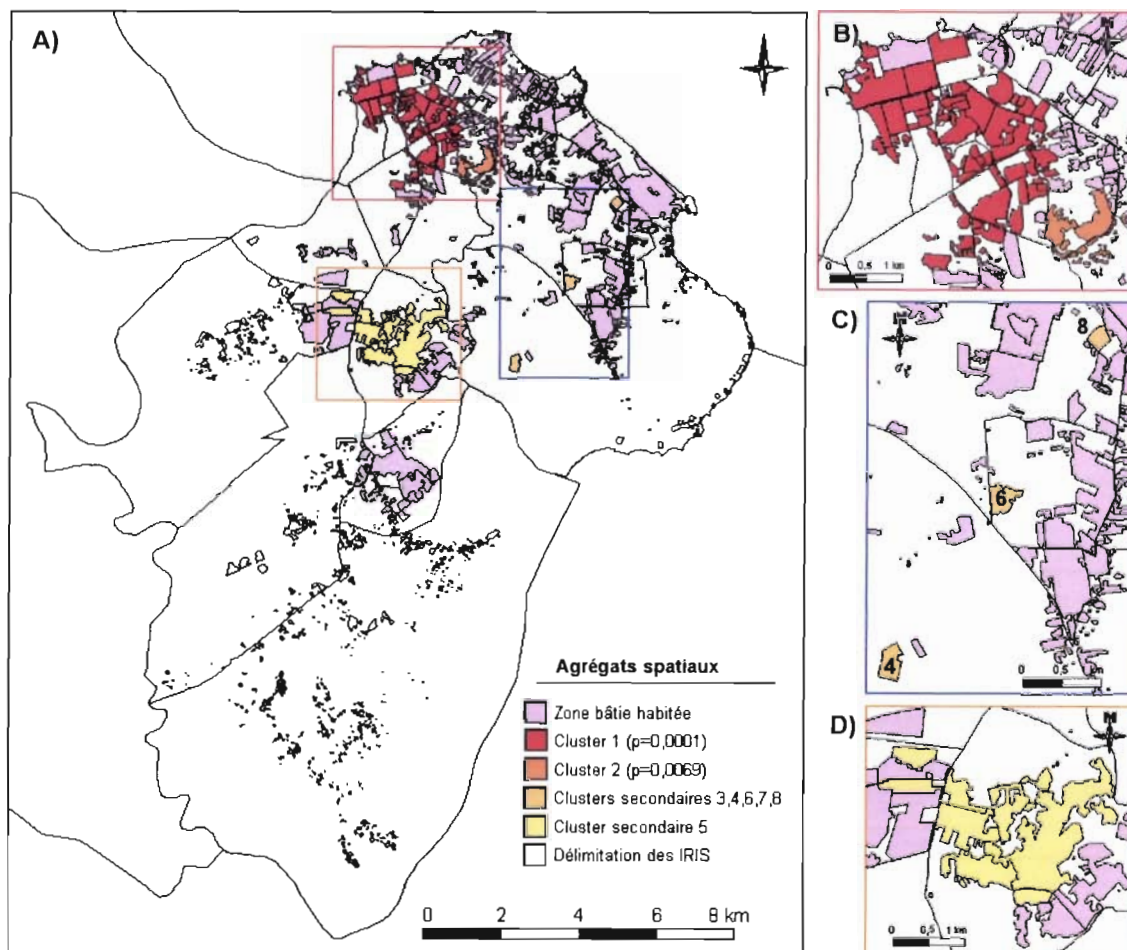


Figure 4.13 – (A) Agrégats de tuberculose identifiés par l'analyse SaTScan à l'échelle des quartiers ; (B) Agrégat principal (n°1) et agrégat secondaire n°2 ; (C) Agrégats secondaires n°4, 6 et 8 ; (D) Agrégat secondaire n°5. À cette échelle, les agrégats n°3 et 7 ne sont pas visibles.

## 2.5. Discussion

### 2.5.1. Comparaison entre les trois approches

Il est intéressant de voir qu'en fonction des covariables introduites dans les analyses, et de l'échelle d'étude, certaines zones géographiques apparaissent de manière systématique, alors que d'autres sont occasionnelles.

Avec la première méthode, qui tient compte de la répartition par âge et par sexe de la population, on obtient une forte agrégation de cas de tuberculose autour du quartier Eau Lisette, et globalement à la périphérie du centre ville de Cayenne (sauf l'IRIS De Gaulle qui fait, lui, partie du centre ville).

Lorsque l'on tient compte de la répartition de la population en fonction de la nationalité, on obtient également une agrégation marquée autour de Eau Lisette mais les quartiers du centre ville historique (IRIS Palmistes et Les

Amandiers) apparaissent inclus dans l'agrégat, alors que les IRIS Mango, Anatole, Bonhomme, Cabassou et Mont Baduel ont disparu. Ceci pourrait indiquer que l'agrégation de cas de tuberculose observée dans ces 5 derniers IRIS avec la première approche (âge et sexe) est simplement corrélée à une sur-représentation de la population étrangère dans ces IRIS. Cependant, cette hypothèse ne correspond pas aux données de population de l'INSEE ; ainsi, l'IRIS Anatole regroupe 28% d'étrangers, alors que l'IRIS Eau Lisette en regroupe 52%. En réalité, la délimitation des agrégats ne doit pas être vue comme quelque chose de strictement défini, l'inclusion d'un IRIS de plus ou de moins étant sujette à caution [comm. pers. J.-F. Viel].

La dernière méthode donne une délimitation plus précise des agrégats. Il est, par exemple, intéressant de noter qu'à cette échelle, l'agrégat de cas au niveau de la prison de Rémire apparaît, de même que celui du village brésilien. Si ces agrégats ne sont pas significatifs, c'est peut-être que les unités spatiales sur lesquelles nous travaillons sont trop petites. Mais à une échelle de travail plus large, ces agrégats très ponctuels n'apparaissent pas. En revanche, puisque à l'échelle des quartiers nous n'avons intégré aucune covariable, il est plus difficile d'identifier les paramètres de la population qui peuvent interférer sur ce patron de distribution.

## 2.5.2. Agrégats de tuberculose et habitat insalubre

### 2.5.2.1. Définitions

Afin de vérifier si la répartition de nos agrégats de tuberculose peut correspondre à l'insalubrité de l'habitat, nous avons repris dans le tableau 4.7 les données d'une étude sur l'identification des zones d'habitat informel et d'habitat insalubre sur l'île de Cayenne [DDE-AU&S, 1997]. La lecture de ce tableau nécessite quelques définitions préalables [DDE-AU&S, 1997].

- *Habitat informel* : trois facteurs définissent cet habitat : le non-respect des règles d'urbanisme, le non-respect des normes de construction, et une implantation sur des terrains non constructibles.
- *Bidonville* : agglomération d'abris de fortune, de constructions sommaires réalisées à partir de matériaux de récupération (bois, tôles, fûts) ou de matériaux détournés de leur fonction originelle, et dans laquelle les habitants vivent dans des conditions sociales et d'hygiène difficile. Type d'habitat généralement péri-urbain.
- *Habitat insalubre* : habitat qui ne répond pas aux normes de maintien de la santé des habitants.
- *Habitat groupé* : ensemble bâti formé de petits groupes de constructions denses, implantées de façon dispersée sur un territoire.

Tableau 4.7 – Zones d’habitat insalubre sur l’île de Cayenne identifiées par la DDE de Guyane, et typologie d’habitat associée [DDE-AU&S, 1997].

Toponymie IGN	Toponymie vernaculaire	Typologie d’habitat
Cité Nzilla	Cité Nzilla	Formel groupé
Cité Eau Lisette	Cité Eau Lisette	Bidonville
sans	Cité Jean-François	Bidonville
Cité Pascaline	Cité Pascaline	Formel groupé
Suzini-Ratier	Cité Suzini	Bidonville
sans	Mont Baduel-route de Raban	Informel groupé
Ploërmel	Vendome	Formel groupé
Ploërmel	Cité Ploërmel	Informel groupé
Louise Elmire (sud)	Sources de Baduel	Informel groupé et diffus
Trou Biran	Trou Biran	Bidonville
Palika	Palika	Informel diffus et formel isolé
Collery	ZA Collery	Informel groupé et diffus
Les Abricots	Quartier Vieux Chemin	Informel et formel groupé
Pont Lacroix	Chemin Léo	Informel groupé
D1	Route des plages	Informel groupé et diffus
Rémire	Centre du village	Bidonville
Village Brésilien	BP 134	Formel isolé
D1	Ancienne route de Dégrad des cannes	Informel diffus
Carrefour RN1/D191	Village amérindien	Informel groupé
Cité Balata Ouest	Cité Balata	Informel groupé
Ancienne distillerie	La mirande	Informel diffus
Cogneau Est	Cogneau	Informel groupé et bidonville
La Désirée	La Désirée	Informel diffus
La Levée	Piste de la Levée	Informel diffus et formel isolé

### 2.5.2.2. Résultats

L’agrégat n°1 comprend cinq zones d’habitat insalubre ; l’agrégat n°2 en regroupe trois, ainsi que l’agrégat n°5 ; les agrégats n°6 et 8 sont des zones d’habitat insalubre. Seuls trois agrégats (n°3, 4 et 7) ne font pas partie des zones d’habitat insalubre : la prison de Rémire, et les deux habitats isolés à Stoupan. La prison est un cas particulier, et il est logique qu’elle ne soit pas notée dans les habitats insalubres. Les deux autres agrégats apparaissent car il y a deux cas dans la même maison alors que la population attendue est de 1, ce qui donne un rapport o/a (nombre de cas observés sur nombre de cas attendus) très important. Pour les cas habitant la même maison, nous avons vérifié s’il s’agit d’une contamination au sein d’un même foyer : dans un cas cette vérification n’est pas possible car la souche est inconnue pour l’un des deux cas de la maison, et pour l’autre il s’agit de deux souches différentes. Il

pourrait donc s'agir d'un habitat insalubre isolé, même s'il n'est pas inclus dans les zones d'habitat insalubre de la DDE.

Si l'on regarde l'ensemble des habitats insalubres, qui sont au nombre de 24, on en compte 23 où des cas de tuberculose ont été recensés entre 1996 et 2003 (l'exception étant le quartier de La Désirée, à Matoury) ; en revanche, seuls 13 de ces zones insalubres sont incluses dans nos agrégats de tuberculose. Mais ces zones pourraient constituer des zones plus favorables que les autres à l'apparition d'agrégat, la présence de tuberculose étant malgré tout notée.

L'ensemble des agrégats de tuberculose que nous avons identifié constituent des sources potentielles de contamination, d'où peuvent démarrer des épidémies. Ceci est d'autant plus vrai que ces agrégats se trouvent dans des zones insalubres, où les populations sont certainement plus vulnérables qu'ailleurs d'un point de vue sanitaire.

### 2.5.2.3. Recherche d'agrégat avec la covariable insalubrité

Afin de tester si le facteur insalubrité peut avoir un effet sur la détection des agrégats de tuberculose, nous avons lancé une analyse sous SaTScan avec le facteur insalubrité du quartier en tant que covariable.

#### *Agrégats significatifs*

Un agrégat est reconnu comme étant significatif lorsque le log du rapport de vraisemblance ( $LRV = L(Z)/L_0$ ) est supérieur ou égal à 10,634 au seuil de 0,01, ou à 9,028 au seuil de 0,05 (seuils fournis par SaTScan).

La méthode met en évidence un agrégat principal, qui correspond au cluster principal déterminé précédemment (figure 4.13) et trois agrégats secondaires significatifs : la prison de Rémire, la cité Abriba, dans le quartier Balata, et une maison avec deux cas à Stoupan (celle identifiée précédemment comme étant un agrégat significatif de tuberculose : agrégat n°3, tableau 4.6).

#### *Discussion*

L'agrégat principal, qui comprend une partie du centre ville de Cayenne et ses quartiers périphériques, continue d'apparaître même lorsque l'on introduit la covariable « insalubrité ». Il semble donc exister un lien entre la présence de cas de tuberculose et la présence d'habitats insalubres, mais en revanche pas obligatoirement entre la caractérisation d'un agrégat de cas et les conditions d'insalubrité.

Ceci étant dit, nous tenons à souligner que l'identification des zones d'habitat insalubre sur laquelle nous nous basons [DDE-AU&S, 1997] date de 1997. De plus, notre connaissance du terrain nous permet de dire que certaines zones de l'île de Cayenne, si elles ne sont pas incluses dans les zones d'habitat insalubre de l'étude de la DDE, présentent néanmoins une forte précarité.

Dans le centre ville de Cayenne par exemple, les maisons délabrées côtoient les cases créoles rénovées. Il est donc possible que, parmi ces habitats vétustes, existent des habitats insalubres isolés qui n'ont pas été identifiés précédemment.

### **3. Conclusion**

Les résultats obtenus ici mettent clairement en évidence l'existence de zones géographiques où l'accumulation locale de cas de tuberculose n'est pas due simplement au hasard. De plus, outre la corrélation de certains agrégats avec l'âge, le sexe ou la nationalité des populations, un lien semble exister entre la présence de cas de tuberculose et l'insalubrité de l'habitat, ce qui est intuitif par rapport à ce que nous connaissons de la tuberculose, qui est une maladie de la pauvreté et de la précarité.

Il peut exister deux hypothèses pour expliquer l'agrégation spatiale des cas de tuberculose : soit il existe une forte transmission à l'échelle locale, avec une diffusion de proche en proche autour du foyer source qui donne lieu à un agrégat de cas de tuberculose correspondant à une micro-épidémie à l'échelle locale ; soit la contamination de proximité est très faible, et dans ce cas c'est un facteur de risque lié à la localisation géographique, par exemple le fait d'habiter dans un habitat insalubre, qui explique l'agrégation de cas. Pour vérifier ces deux hypothèses, il nous faut identifier les agrégats cette fois génétiques, c'est-à-dire les groupes de patients issus d'une même source de contamination. C'est ce que nous développons dans le chapitre suivant.

## Chapitre 5

# Épidémiologie moléculaire de la tuberculose en Guyane

*Il existe différentes méthodes de typage moléculaire pour la caractérisation infra-spécifique des souches de Mycobacterium tuberculosis. Nous en avons utilisé deux, l'une seule tout d'abord, puis les deux associées, afin de caractériser les clones identiques, ou « souches en grappes », et de suivre la transmission grâce à ces grappes génétiques. La comparaison des souches identifiées en Guyane avec les bases de données internationales nous a également permis de déterminer les spécificités géographiques de certaines souches. L'identification des grappes génétiques nous a par ailleurs permis de construire des arbres phylogénétiques basés sur l'analyse des spoligotypes. La construction de ce que l'on appelle des « cluster-graph », qui a pour l'instant fait l'objet d'un unique article en janvier 2005, permet ensuite le calcul d'indices de diversité / regroupement et du taux de transmission récente, qui permet de juger de la sévérité de l'endémie / épidémie.*

## 1. Identification des grappes génétiques de *Mycobacterium tuberculosis*

### 1.1. Introduction

#### 1.1.1. Typage moléculaire des souches

En plus d'une identification précise et d'un antibiogramme détaillé, il est intéressant pour le clinicien et l'épidémiologiste de savoir si une souche bactérienne isolée chez un patient provient de la même source de contamination que d'autres souches apparemment identiques et isolées chez d'autres patients ou dans l'environnement. Cela permet de déceler la présence d'une chaîne infectieuse responsable d'une épidémie [Small et Behr, 1997]. La mise en évidence d'une telle chaîne infectieuse est réalisable en utilisant une

méthode de typage qui permet une différenciation précise entre des souches de la même espèce.

L'hypothèse sous-jacente à l'utilisation du typage moléculaire pour le suivi de la transmission est que les changements génétiques (mutations) sont moins rapides que la transmission [Kulaga *et al.*, 1990], ce qui n'est pas vrai pour tous les marqueurs (*cf.* §1.2.1. pour le choix du marqueur).

### 1.1.2. Grappes génétiques

L'identification des souches bactériennes de *Mycobacterium tuberculosis* permet d'identifier les patients ayant un profil génétique bactérien identique.

Pour rappel (*cf.* chapitre 4), les cas s'étant révélés porteurs de la même souche bactérienne, donc potentiellement issus d'une même source de contamination, sont appelés « cas en grappe », et l'ensemble des souches ayant un profil génétique identique définissent un « *cluster* » ou un grappe génétique [Crawford *et al.*, 2002 ; Möstrom *et al.*, 2002]. Ces grappes définissent une chaîne infectieuse potentielle.

## 1.2. Méthodes de typage moléculaire

### 1.2.1. Critères de choix d'une méthode de typage

Lorsque l'on évalue des méthodes de typage moléculaire, les critères les plus importants à prendre en compte sont la typabilité, la reproductibilité, et le pouvoir discriminant [Hunter, 1990].

- *la typabilité* est la capacité à obtenir un résultat positif, non-ambigu pour chaque souche [Hunter, 1990].
- *la reproductibilité* est la capacité de la méthode à donner le même résultat lorsque la même souche est testée plusieurs fois avec la même méthode [Hunter, 1990].
- *le pouvoir discriminant* est la capacité à différencier les souches non apparentées. Idéalement, une méthode de typage très discriminante devrait identifier chaque souche comme unique. Pratiquement, la méthode peut être considérée comme utile lorsque la probabilité que deux souches non apparentées appartiennent au même type est inférieure à 5 %. Plus le pouvoir discriminant augmente, plus la méthode est capable de détecter des variations minimales ou moins fréquentes [Hunter, 1990].

La reproduction de *Mycobacterium tuberculosis* est clonale (*cf.* chapitre 1). La conclusion que deux souches ayant un même profil génétique viennent d'un

même clone, dépend du pouvoir discriminant de la méthode utilisée et de la diversité génétique de la population des souches examinées. Si la méthode a un pouvoir discriminant très élevé, elle pourra détecter des différences qui n'ont aucun intérêt épidémiologique, telles que des mutations spontanées par exemple. La méthode de typage choisie ne doit donc pas forcément être la plus discriminante possible.

### Encadré 5.1 – L'index de discrimination ( $D$ )

Gaston et Hunter (1988) ont suggéré que le pouvoir discriminant d'une méthode peut être défini mathématiquement comme la probabilité que deux souches choisies au hasard dans une population de souches non liées entre elles soient différenciées par la méthode choisie. Ce principe a permis de définir un index de discrimination ( $D$ ). La probabilité associée est donnée par la formule :

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^s x_j(x_j - 1)$$

où  $s$  est le nombre total de souches,  $x_j$  le nombre d'individus appartenant au type  $j$ , et  $N$  la taille de la population totale. Un  $D$  égal à 1 signifie donc que la méthode est très discriminante, car elle permet de distinguer chacun des membres de la population de souches par rapport aux autres, alors qu'à l'inverse, un  $D$  égal à 0 indique que toutes les souches de la population étudiée sont reconnues comme étant une souche identique par la méthode utilisée, qui est donc non discriminante.

## 1.2.2. Méthodes de typage de *Mycobacterium tuberculosis*

### 1.2.2.1. Les méthodes de référence

Il existe actuellement un large panel de techniques moléculaires utilisables pour typer les souches de *Mycobacterium tuberculosis*.

Les séquences cibles de ces techniques (ou marqueurs) peuvent être soit des séquences d'insertion, dont les plus utilisées sont IS6110 et IS1081, soit des éléments répétitifs du génome tels que le locus DR (*direct repeat*), les loci PGRS (*polymorphic GC-rich sequence* ou séquences polymorphes riches en GC), ETR (*exact tandem repeat*) et MIRU (*Mycobacterial interspersed repetitive units*) [Kanduma *et al.*, 2003 ; Möstrom *et al.*, 2002].

Les différentes méthodes d'investigation de ces marqueurs sont la technique RFLP (*restriction fragment length polymorphism*), qui utilise des enzymes de restriction pour découper l'ADN chromosomal en un certain nombre de fragments de restriction, la technique d'amplification PCR de certaines régions

cibles de l'ADN ou de l'ARN de *M. tuberculosis*, la LM-PCR (*ligation mediated-PCR*) et le spoligotypage. La technique de F-ALFP (*fluorescent amplified-fragment length polymorphism*) est moins courante [Kanduma *et al.*, 2003 ; Möstrom *et al.*, 2002].

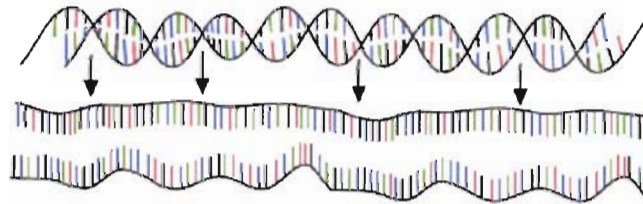
### Encadré 5.2 – Le principe de la PCR

Imaginée par Mullis *et al.* (1986), la « *Polymerase Chain Reaction* » ou PCR est une technique de répliation *in vitro* permettant d'obtenir, à partir d'un échantillon complexe et peu abondant, d'importantes quantités d'un fragment d'ADN spécifique et de longueur définie.

Le principe est basé sur une succession de réactions de répliation d'une matrice double brin d'ADN. Chaque réaction met en oeuvre deux amorces oligonucléotidiques dont les extrémités 3' pointent l'une vers l'autre, ces amorces définissant alors, en la bornant, la séquence à amplifier. Au fil des cycles de synthèse (*cf.* 3 étapes ci-dessous), la quantité d'ADN augmente de façon exponentielle. En théorie, après 20 cycles, il y a environ un million de copies de la séquence d'ADN cible.

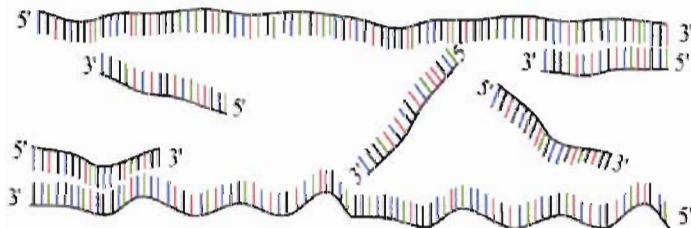
**Étape 1 :** dénaturation de l'ADN double brin en ADN simple brin

1 minute à 94°C



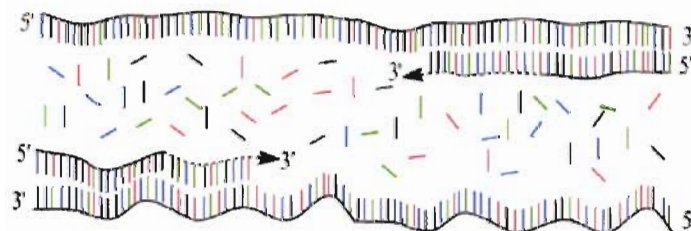
**Étape 2 :** hybridation avec les sondes complémentaires en 3' et en 5' servant d'amorces

45 secondes à 54°C



**Étape 3 :** ajout des nucléotides dATP, dCTP, dGTP, dTTP (désoxynucléotides tri-phosphates) et élongation des brins d'ADN grâce à l'enzyme Taq polymérase

2 minutes à 72°C



#### 1.2.2.2. Quelles méthodes pour quelles données ?

L'ensemble de ces méthodes ont été testées sur *M. tuberculosis*, mais toutes n'ont pas la même validité et la même application [Kremer *et al.*, 1999]. Par exemple, la méthode IS6110-RFLP est actuellement le standard de référence

car il existe un protocole consensuel international [van Embden *et al.*, 1993]. Kanduma *et al.* (2003) ont établi un diagramme permettant de choisir la méthode adéquate en fonction de l'application souhaitée. Le dernier consensus suggère de typer les souches par spoligotype et MIRU 12 (*Mycobacterial interspersed repetitive units*), puis celles qui sont toujours en grappes par IS6110-RFLP [CDC, 2005].

### 1.2.2.3. Application à nos propres données

#### *Méthodes choisies*

Dans notre propre étude, nous avons choisi de coupler deux méthodes : spoligotypage et VNTR (*Variable number of tandem DNA repeats*), détaillées respectivement §1.3. et 1.4. Cette combinaison de méthodes permet d'affiner l'identification des grappes génétiques, et d'obtenir un pouvoir discriminant presque équivalent à la méthode IS6110-RFLP [Filliol *et al.*, 2000]. Les avantages et inconvénients de ces deux méthodes sont décrites ci-après.

#### *Remarque*

La combinaison spoligotypage et MIRU est encore plus efficace [CDC, 2005] mais cette méthode n'était que peu utilisée à l'Institut Pasteur de Guadeloupe lors de la mission réalisée sur place, pour des raisons de temps (12 VNTR à tester au lieu de 5 pour les ETR-A à E) et de coût.

### 1.2.3. Principe général du spoligotypage

#### *Principe*

Le spoligotypage, ou typage oligonucléotidique des espaceurs du locus DR (*cf.* détails ci-après), est une méthode pré-conceptualisée par Groenen *et al.* (1993) et développé par Kamerbeek *et al.* (1997). Cette méthode repose sur la détection du polymorphisme de 43 séquences oligonucléotides du génome de *M. tuberculosis*, intégrées dans le locus<sup>1</sup> DR (*direct repeat*). Le locus DR est constitué de courtes séquences d'ADN répétées (DR) de 36 paires de base (pb), séparées par des séquences variables de 35 à 41 pb appelées espaceurs ou séquences inter-DR [Devallois *et al.*, 1998 ; Sola *et al.*, 2000]. Ces séquences varient d'une souche à l'autre par leur longueur, leur séquence et leur nombre. La technique consiste à amplifier le locus DR par PCR, puis à l'hybrider sur une membrane où sont fixées de façon covalente 43 espaceurs (94 ont été identifiés, mais seulement 43 sont utilisés en routine). Les fragments hybridés sur le gel sont ensuite révélés par chimioluminescence. Les résultats montrent la présence ou l'absence des 43 espaceurs dans la souche testée (figure 5.1).

---

<sup>1</sup> Emplacement précis d'un gène sur un chromosome

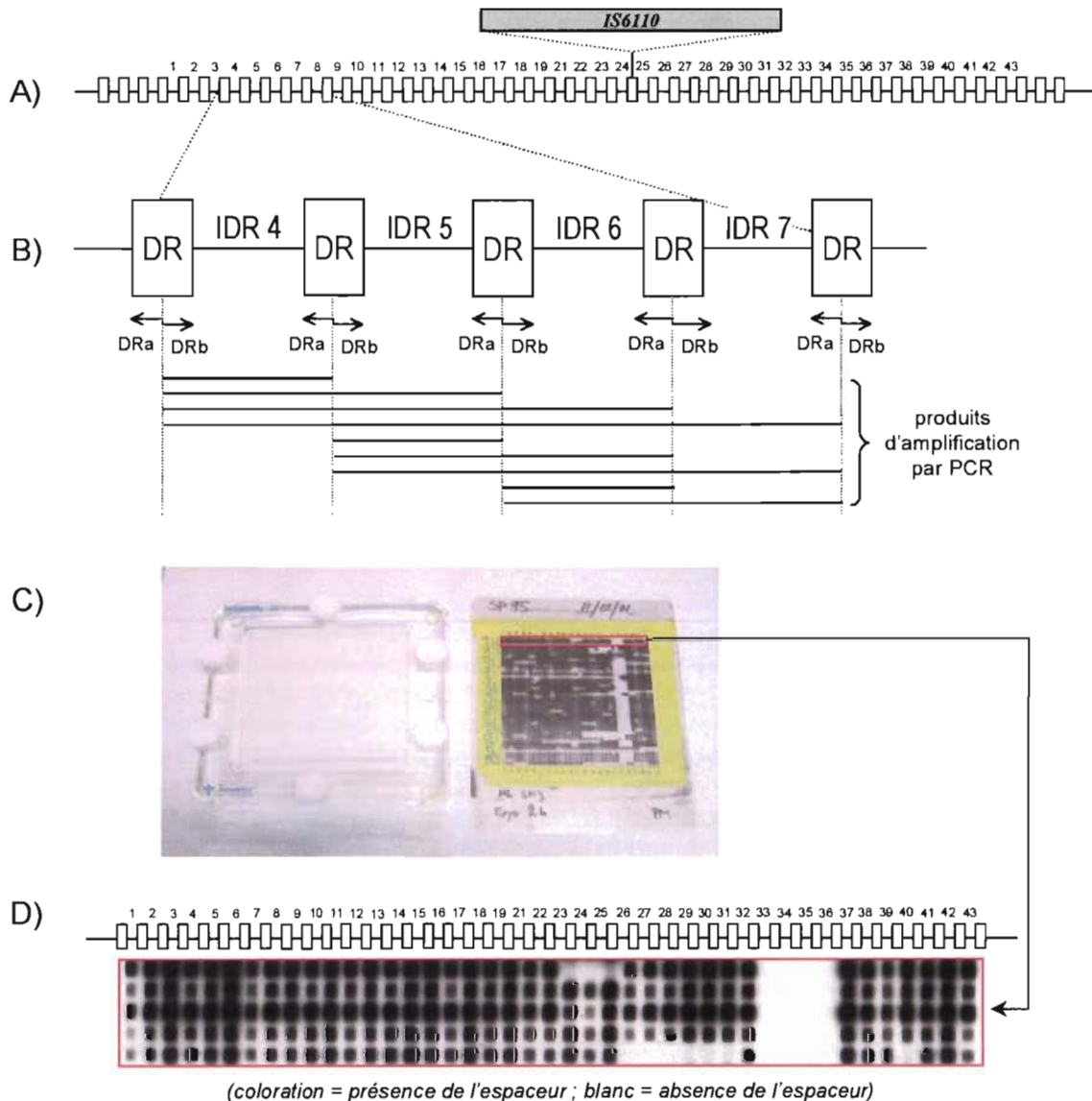


Figure 5.1 – (A) Structure du locus DR du génome des mycobactéries. (B) Amplification in vitro de la région DR par PCR grâce aux amorces DRa et DRb. Chaque DR du locus peut servir de départ à l'amplification, donc le résultat d'amplification donne un mélange de fragments d'ADN de taille multiples. (C) Détection des sondes inter-DR (IDR) par hybridation sur un système matriciel, à gauche (membrane sur laquelle sont liées de façon covalente les 43 séquences oligonucléotidiques correspondant aux 43 espaceurs). À droite, photographie après révélation par chimioluminescence (ECL, Amersham). (D) Motifs d'hybridation (spoligotypes) des souches de *M. tuberculosis* testées (une souche par ligne). La coloration noire indique la présence de l'espaceur dans la souche considérée alors que le blanc indique l'absence de cet espaceur. (Schéma d'après Groenen *et al.*, 1993 et Kamerbeek *et al.*, 1997).

### Avantages

Cette méthode étant spécifique de *Mycobacterium tuberculosis*, elle permet d'éliminer les autres sous-espèces de *Mycobacterium* et les mycobactéries atypiques [Burgos et Pym, 2002]. Cette méthode présente également l'avantage de pouvoir être utilisée directement sur des isolats cliniques car, grâce à la

PCR, elle nécessite très peu d'ADN [Kamerbeek *et al.*, 1997]. Enfin, cette méthode est simple, rapide, robuste, et peu coûteuse [Kanduma *et al.*, 2003].

### *Inconvénients*

Le pouvoir discriminant de cette méthode est assez faible par rapport à IS6110-RFLP [Burgos et Pym, 2002 ; Kanduma *et al.*, 2003]. Cependant, cette méthode peut être appliquée en première intention avant d'appliquer une seconde technique ayant un pouvoir discriminant plus important, ce qui est de plus en plus conseillé [Burgos et Pym, 2002 ; Möstrom *et al.*, 2002].

## 1.2.4. Principe général du typage par VNTR

### *Principe*

Le séquençage complet du génome de *M. tuberculosis* H37Rv a permis de mettre en évidence certaines régions d'ADN (ou loci) contenant un nombre variable de régions répétées en tandem (VNTR). Frothingham et Meeker-O'Connell (1998) ont étudié onze loci hypervariables du génome de *M. tuberculosis*. Cinq loci sont des MPTR (*major polymorphic tandem repeat*), contenant des séquences d'ADN très polymorphes, six sont des ETR (*exact tandem repeat*) contenant des séquences identiques d'ADN répétées en tandem. Le nombre de répétitions en tandem à chaque locus (ou copies) est variable d'une souche à l'autre. Ainsi, la souche H37Rv possède 3 copies de l'allèle A et 3 copies de l'allèle B (figure 5.2).

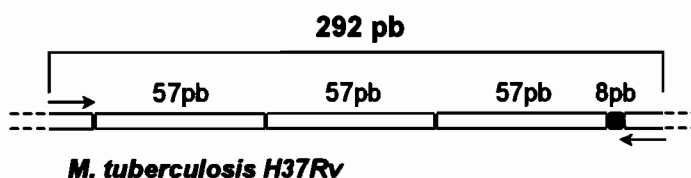


Figure 5.2 – Exemple d'un locus contenant un nombre variable de régions répétées en tandem (VNTR). La figure montre l'ADN génomique au niveau du locus ETR-B de *M. tuberculosis* H37Rv. Le locus est amplifié par PCR grâce à des amorces complémentaires des régions d'ADN qui encadrent le locus (flèches). Le produit de PCR obtenu est de 292 paires de bases (pb) contenant 3 copies de régions répétées en tandem de 57 paires de bases, et 8 bases supplémentaires qui correspondent au début d'une autre région répétée en tandem.

Chaque locus est amplifié par PCR, et les produits de PCR sont soumis à une électrophorèse<sup>2</sup> sur gel d'agarose. Les fragments d'ADN sont révélés par coloration au bromure d'ethidium (BET), fluorescent sous UV. L'électrophorèse permet de déterminer précisément le poids moléculaire des fragments, dont on déduit ensuite le nombre de copies de chaque VNTR.

<sup>2</sup> Technique de laboratoire permettant la séparation de fragments d'ADN grâce à leur différence de masse, en présence d'un champ électrique. La vitesse de migration des fragments dépend de leur poids moléculaire.

Le calcul des poids moléculaires se fait par comparaison avec des marqueurs de poids moléculaire que l'on fait migrer en parallèle avec les échantillons à analyser, et qui sont des mélanges de fragments d'ADN de poids connus (figure 5.3).

Après calcul du nombre de copies de chaque VNTR, on aboutit à un code chiffré représentatif du polymorphisme de la souche considérée sur les régions répétées en tandem. Par exemple, si l'on tient compte des ETR-A à E, le profil VNTR de H37Rv est 33433, ce qui signifie qu'il y a 3 copies d'ETR-A, 3 copies d'ETR-B, etc. [Kamerbeek *et al.*, 1997 ; Frothingham et Meeker-O'Connell, 1998].

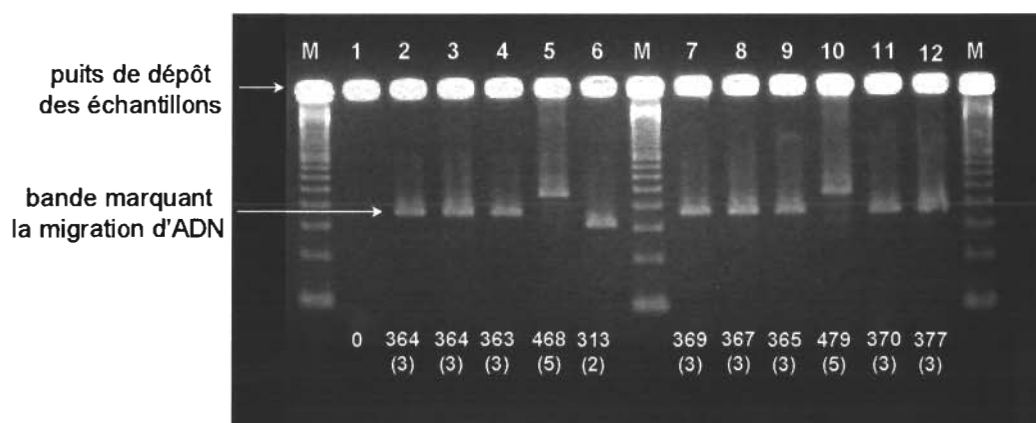


Figure 5.3 – Polymorphisme de taille de différentes souches de *M. tuberculosis* isolées en Guyane pour l'ETR-E. La photo montre le profil de migration des échantillons d'ADN sur gel après électrophorèse. Chaque colonne contient le produit de PCR d'une souche différente (ici, 12 souches) ; le dépôt n°12 n'a pas migré. M correspond aux marqueurs de poids moléculaire "100bp ladder" (100, 200, 300... à 1000 pb). La vitesse de migration des fragments dépend de leur poids moléculaire ; le fragment de 100 pb migre le plus vite, c'est donc la bande la plus éloignée du puit). Le polymorphisme observé correspond à l'addition ou à la délétion d'une unité répétée en tandem de 53 pb. Le poids moléculaire (en pb) est noté au-dessus des profils de migration ; juste en-dessous entre parenthèses, est indiqué le nombre de copies correspondantes.

#### Remarque

À l'Institut Pasteur de Guadeloupe, le logiciel Taxotron (Institut Pasteur, Paris, France) est utilisé pour convertir la distance de migration mesurée sur la photo numérisée, en poids moléculaires, dont on déduit ensuite le nombre de copies pour chaque locus (figure 5.3).

#### Avantages

Cette méthode, également basée sur une méthode PCR, peut se faire directement sur des isolats [Kanduma *et al.*, 2003]. De plus, la méthode est simple et rapide.

### *Inconvénients*

Utilisée seule, cette méthode est faiblement discriminante, mais utilisée en seconde intention, complémentaire à une première analyse, elle permet de subdiviser les grappes de souches ubiquistes, ces grappes étant obtenus par une méthode préliminaire [Filliol *et al.*, 2000].

### **1.2.5. Mise en œuvre des analyses moléculaires**

Sur les 345 cas confirmés de tuberculose recensés sur l'ensemble de la Guyane entre 1994 et 2003, toutes les souches bactériennes correspondantes ont été typées par le centre de référence des Mycobactéries de l'Institut Pasteur de Guadeloupe sauf trois (ADN non disponible ou culture contaminée), soit 342 cas. En revanche, deux doublons qui avaient été retirés de l'étude précédemment ont été réintroduits dans cette partie de l'étude, car les analyses moléculaires ont mis en évidence deux infections tuberculeuses par deux souches différentes. Avec ces deux patients porteurs de deux souches différentes, on arrive à un total 344 souches étudiées.

### *Le spoligotyping*

Il a été réalisé selon le protocole décrit dans Kamerbeek *et al.* (1997) et Sola *et al.* (1998). Trois étudiantes en thèse ont eu la charge de pratiquer ces analyses sur les souches de Guyane : Ingrid Filliol, Sandrine Ferdinand et Karine Brudey. Tous les spoligotypes étaient disponibles sur la période du 1<sup>er</sup> janvier 1996 au 31 décembre 2003, et nous avons donc simplement récupéré cette information dans les bases de données. À chaque spoligotype est attribué un numéro d'identification dès que deux souches identiques sont identifiées ; les autres sont des souches orphelines. Il existe au Centre de Référence des Mycobactéries de l'Institut Pasteur de Guadeloupe une base de données qui recense l'ensemble de ces spoligotypes (au 1<sup>er</sup> août 2005, le nombre total de spoligotypes labellisées est 2393, le nombre total d'entrées (nombre de patterns) est de 31 642, correspondant à environ 45 000 souches, provenant de 114 pays). Cette base de données (Spolddb4) est en cours de publication [Brudey *et al.*, en préparation] mais les données sont par ailleurs disponibles sur demande à Nalin Rastogi ou Christophe Sola<sup>3</sup>.

### *L'analyse VNTR*

Elle a été faite en suivant le protocole décrit par Frothingham et Meeker-O'Connell (1999), avec quelques légères modifications décrites dans Filliol *et al.* (2000). Sandrine Ferdinand et Karine Brudey ont réalisé une partie des analyses ; mais au moment de l'étude, une partie des analyses VNTR n'avaient

---

<sup>3</sup> [nrastogi@pasteur-guadeloupe.fr](mailto:nrastogi@pasteur-guadeloupe.fr) ; [csola@pasteur-guadeloupe.fr](mailto:csola@pasteur-guadeloupe.fr) ; téléphone : 05 90 89 76 61.

pas pu être réalisées. Nous avons donc procédé au reste des analyses de typage par VNTR au cours d'une mission de deux mois à l'Institut Pasteur de Guadeloupe, en janvier et février 2005. L'étude de Kremer *et al.* (1999) ayant mis en évidence que seuls cinq loci sur onze (les ETR-A à E) étaient suffisamment discriminants, ce sont ces cinq ETR dont le polymorphisme entre souches a été analysé à l'Institut Pasteur de Guadeloupe.

### 1.3. Résultats

Remarque préliminaire : il est à noter que les comparaisons du pool de souches isolées en Guyane avec l'ensemble des souches isolées dans le monde s'appuient sur la base internationale disponible à l'Institut Pasteur de Guadeloupe (Spolddb 4), base qui par nature ne peut pas être exhaustive, mais dont l'objectif principal est de décrire la diversité génétique mondiale des populations de BK afin d'en étudier la structure.

D'autre part, seuls les cas confirmés, pour lesquels on dispose d'une culture, sont intégrés à ces analyses moléculaires et aux grappes génétiques.

#### 1.3.1. Spoligotypes identifiés (ou shared-type, ou ST)

Sur les 344 isolats bactériens analysés, on a mis en évidence l'existence de 120 spoligotypes différents, dont 25 orphelins (21%), qui n'ont donc jusqu'à maintenant été identifiés nulle part ailleurs dans le monde. Le tableau 5.1 récapitule l'ensemble des spoligotypes identifiés en Guyane, ainsi que le nombre de souches concernées, et fournit des informations sur la fréquence du spoligotype en Guyane et dans le monde (en nombre de souches et en pourcentage du total des spoligotypes), et le nombre de pays où la souche a été recensée. Des indices ont par ailleurs été développés pour décrire la spécificité géographique des différentes souches, et leur prévalence [Filliol *et al.*, 2003].

*Matching code (MC)* : ce code résume la (les) spécificité(s) géographique(s) d'un spoligotype donné ; à un continent correspond un chiffre, qui est inclus dans le MC lorsque le spoligotype a été recensé sur ce continent (1, Afrique ; 2, Amérique du Nord ; 3, Amérique centrale ; 4, Amérique du Sud ; 5, Europe ; 6, Asie centrale et Moyen-orient ; 7, Asie du Sud-Est ; 8, Océanie).

*Spreading index (SI)* : cet indice de dispersion est obtenu en divisant le nombre total d'isolats de ce spoligotype par le nombre de pays où il a été recensé. Contrairement à l'indice MC, SI est un indicateur quantitatif.

*C1* est un indicateur de la dispersion géographique du spoligotype. Il correspond aux qualificatifs suivants : endémique<sup>4</sup>, localisé ou ubiquiste.

*C2* est un indicateur de la prévalence (en terme de nombres d'isolats ayant le même spoligotype). Il correspond aux qualificatifs suivants : rare, récurrent, commun ou très abondant. Un spoligotype peut être à la fois endémique mais abondant, ou ubiquiste mais rare.

Les règles de définition de *C1* et *C2* sont les suivantes :

- si  $MC = 1$  continent, *C1* = endémique
- si  $MC = 2$  continents, *C1* = localisé
- si  $MC \geq 3$  continents et *ST* présent dans  $\leq 5$  pays, *C1* = localisé
- si  $MC \geq 3$  continents et *ST* présent dans  $\geq 6$  pays, *C1* = ubiquiste
  
- si  $SI \geq 30$ , *C2* = épidémique
- si  $10 \leq SI \leq 29$ , *C2* = commun
- si  $3 \leq SI \leq 9$ , *C2* = récurrent
- si  $SI \leq 2$ , *C2* = rare

Le tableau 5.1 récapitule l'ensemble des résultats de spoligotypage sur les 344 souches isolées en Guyane entre 1996 et 2003, ainsi que les spécificités géographiques et prévalences associées, correspondant aux indices que nous avons définis ci-dessus. Sur les 344 souches, seuls les 25 spoligotypes orphelins ne sont pas précisés dans le tableau, puisqu'ils sont, en l'état actuel des connaissances, endémiques et rares.

#### *Analyse des résultats de spoligotypage*

Ce tableau indique 31 spoligotypes (*ST*) endémiques à la Guyane (25,8%), 40 « localisés »<sup>5</sup> (33,3%), et 49 ubiquistes (40,8%). Mais la proportion de cas attribués aux spoligotypes endémiques, localisés et ubiquistes est différente de la proportion en termes de nombre de *ST*. Afin de mettre en évidence cette nuance, la figure 5.4 récapitule ces proportions. Ainsi les souches ubiquistes représentent 40,8% des 120 *ST* de Guyane, mais représentent 70,6% des patients totaux.

Outre les spoligotypes orphelins, les spoligotypes 66, 72, 385, 958, 1337, 1339, 1340, 1486, 1762, 1763, et 1935 représentent plus de 50% du total des souches de chacun de ces *ST* au niveau mondial. Au total, 40 spoligotypes sur les 120 identifiés en Guyane (33,3%) sont donc fortement sur-représentés dans ce département.

---

<sup>4</sup> Attention, endémique est ici à prendre au sens écologique du terme (espèce endémique, propre à une zone géographique donnée), et non pas au sens épidémiologique (maladie endémique, qui persiste dans le temps).

<sup>5</sup> Traduction de l'anglais « localized ». En fait, ce sont plutôt des souches intercontinentales, mais dont la dispersion est limitée à 5 pays maximum.

CHAPITRE 5 : Épidémiologie moléculaire de la tuberculose en Guyane

Tableau 5.1 – Tableau récapitulatif des spoligotypes isolés en Guyane sur la période 1996-2003, et des spécificités géographiques et abondances associées. Les 25 spoligotypes orphelins se sont pas intégrés dans ce tableau puisqu'ils sont uniques en Guyane et dans le monde (en l'état actuel des connaissances, c'est-à-dire au 1<sup>er</sup> janvier 2005).

ST	NGUF	Nmonde	% GUF	% Monde	Npays	MC	(MC)	SI	C1	C2
001	5	3127	1,50	10,76	58	12345678	8	53,91	Ubiquiste	Épidémique
002	16	216	4,79	0,74	26	123456	6	8,31	Ubiquiste	Récurrent
004	2	95	0,60	0,33	28	123457	6	3,39	Ubiquiste	Récurrent
005	2	21	0,60	0,07	5	234	3	4,20	Localisé	Récurrent
007	3	28	0,90	0,10	14	123457	6	2,00	Ubiquiste	Rare
011	3	198	0,90	0,68	20	2345678	7	9,90	Ubiquiste	Récurrent
014	1	33	0,30	0,11	5	234	3	6,60	Localisé	Récurrent
017	6	287	1,80	0,99	19	123457	6	15,11	Ubiquiste	Commun
020	9	359	2,69	1,23	35	12345678	8	10,26	Ubiquiste	Commun
031	2	30	0,60	0,10	13	234568	6	2,31	Ubiquiste	Rare
032	2	63	0,60	0,22	12	24578	5	5,25	Ubiquiste	Récurrent
033	1	238	0,30	0,82	27	123456	6	8,81	Ubiquiste	Récurrent
034	4	322	1,20	1,11	37	12345678	8	8,70	Ubiquiste	Récurrent
037	1	138	0,30	0,47	36	12345678	8	3,83	Ubiquiste	Récurrent
041	1	88	0,30	0,30	18	12458	5	4,89	Ubiquiste	Récurrent
042	10	999	2,99	3,44	59	12345678	8	16,93	Ubiquiste	Commun
044	3	95	0,90	0,33	25	1234578	7	3,80	Ubiquiste	Récurrent
046	1	64	0,30	0,22	19	12345678	8	3,37	Ubiquiste	Récurrent
047	3	640	0,90	2,20	40	1234568	7	16,00	Ubiquiste	Commun
048	2	215	0,60	0,74	22	1245678	7	9,77	Ubiquiste	Récurrent
050	48	1194	14,37	4,11	57	12345678	8	20,95	Ubiquiste	Commun
051	6	137	1,80	0,47	21	2345678	7	6,52	Ubiquiste	Récurrent
052	1	348	0,30	1,20	48	12345678	8	7,25	Ubiquiste	Récurrent
053	39	2079	11,68	7,15	74	12345678	8	28,09	Ubiquiste	Commun
058	1	80	0,30	0,28	14	12345	5	5,71	Ubiquiste	Récurrent
060	2	74	0,60	0,25	25	123456	6	2,96	Ubiquiste	Rare
062	2	141	0,60	0,49	26	1234568	7	5,42	Ubiquiste	Récurrent
063	1	5	0,30	0,02	4	345	3	1,25	Localisé	Rare
064	13	115	3,89	0,40	21	12345678	8	5,48	Ubiquiste	Récurrent
066	1	2	0,30	0,01	1	4	1	2,00	Endémique	Rare
067	2	16	0,60	0,06	5	234	3	3,20	Localisé	Récurrent
070	5	67	1,50	0,23	7	23457	5	9,57	Ubiquiste	Récurrent
072	6	11	1,80	0,04	3	234	3	3,67	Localisé	Récurrent
073	5	103	1,50	0,35	24	12345678	8	4,29	Ubiquiste	Récurrent
076	2	2	0,60	0,01	1	4	1	2,00	Endémique	Rare
077	3	10	0,90	0,03	4	234	3	2,50	Localisé	Rare
078	2	26	0,60	0,09	9	2456	4	2,89	Ubiquiste	Rare
086	1	42	0,30	0,14	9	2456	4	4,67	Ubiquiste	Récurrent
088	1	3	0,30	0,01	3	134	3	1,00	Localisé	Rare
089	1	62	0,30	0,21	11	234578	6	5,64	Ubiquiste	Récurrent
091	3	119	0,90	0,41	12	2345	4	9,92	Ubiquiste	Récurrent
092	5	109	1,50	0,37	16	123458	6	6,81	Ubiquiste	Récurrent
093	2	153	0,60	0,53	13	23458	5	11,77	Ubiquiste	Commun
094	5	5	1,50	0,02	1	4	1	5,00	Endémique	Récurrent
095	2	9	0,60	0,03	4	245	3	2,25	Localisé	Rare
096	4	35	1,20	0,12	3	246	3	11,67	Localisé	Commun
102	1	37	0,30	0,13	15	12345678	8	2,47	Ubiquiste	Rare
118	1	64	0,30	0,22	24	1234567	7	2,67	Ubiquiste	Rare
129	2	13	0,60	0,04	7	13456	5	1,86	Ubiquiste	Rare
131	9	22	2,69	0,08	11	123456	6	2,00	Ubiquiste	Rare
134	1	9	0,30	0,03	5	345	3	1,80	Localisé	Rare

CHAPITRE 5 : Épidémiologie moléculaire de la tuberculose en Guyane

ST	NGUF	Nmonde	% GUF	% Monde	Npays	MC	(MC)	SI	C1	C2
136	2	30	0,60	0,10	10	12457	5	3,00	Ubiquiste	Récurrent
137	2	648	0,60	2,23	20	1234578	7	32,40	Ubiquiste	Épidémique
150	3	58	0,90	0,20	14	12345	5	4,14	Ubiquiste	Récurrent
152	1	26	0,30	0,09	7	2457	4	3,71	Ubiquiste	Récurrent
176	2	11	0,60	0,04	3	124	3	3,67	Localisé	Récurrent
177	2	18	0,60	0,06	8	12457	5	2,25	Ubiquiste	Rare
179	1	8	0,30	0,03	3	45	2	2,67	Localisé	Rare
181	1	117	0,30	0,40	12	1245	4	9,75	Ubiquiste	Récurrent
193	1	14	0,30	0,05	3	234	3	4,67	Localisé	Récurrent
240	1	7	0,30	0,02	5	2345	4	1,40	Localisé	Rare
245	1	10	0,30	0,03	9	1245	4	1,11	Ubiquiste	Rare
274	1	6	0,30	0,02	5	245	3	1,20	Localisé	Rare
306	1	14	0,30	0,05	6	1245	4	2,33	Ubiquiste	Rare
333	1	10	0,30	0,03	3	34	2	3,33	Localisé	Récurrent
334	3	35	0,90	0,12	17	123456	6	2,06	Ubiquiste	Rare
344	1	7	0,30	0,02	3	45	2	2,33	Localisé	Rare
347	1	13	0,30	0,04	5	2458	4	2,60	Localisé	Rare
385	2	3	0,60	0,01	2	34	2	1,50	Localisé	Rare
414	1	11	0,30	0,04	5	24567	5	2,20	Localisé	Rare
453	1	7	0,30	0,02	4	245	3	1,75	Localisé	Rare
483	1	7	0,30	0,02	5	2457	4	1,40	Localisé	Rare
492	1	6	0,30	0,02	4	2345	4	1,50	Localisé	Rare
702	1	12	0,30	0,04	3	134	3	4,00	Localisé	Récurrent
753	1	11	0,30	0,04	6	1245	4	1,83	Ubiquiste	Rare
924	3	9	0,90	0,03	6	456	3	1,50	Ubiquiste	Rare
926	1	7	0,30	0,02	5	2456	4	1,40	Localisé	Rare
958	2	4	0,60	0,01	3	245	3	1,33	Localisé	Rare
1059	1	3	0,30	0,01	2	47	2	1,50	Localisé	Rare
1062	1	3	0,30	0,01	3	146	3	1,00	Localisé	Rare
1066	1	5	0,30	0,02	5	1245	4	1,00	Localisé	Rare
1084	4	4	1,20	0,01	1	4	1	4,00	Endémique	Récurrent
1214	1	11	0,30	0,04	4	456	3	2,75	Localisé	Rare
1284	1	4	0,30	0,01	3	245	3	1,33	Localisé	Rare
1318	1	6	0,30	0,02	3	146	3	2,00	Localisé	Rare
1321	1	7	0,30	0,02	4	2456	4	1,75	Localisé	Rare
1337	4	6	1,20	0,02	2	45	2	3,00	Localisé	Récurrent
1339	1	2	0,30	0,01	2	24	2	1,00	Localisé	Rare
1340	3	3	0,90	0,01	1	4	1	2,00	Endémique	Rare
1486	1	2	0,30	0,01	2	24	2	1,00	Localisé	Rare
1526	2	2	0,60	0,01	1	4	1	2,00	Endémique	Rare
1761	1	3	0,30	0,01	3	45	2	1,00	Localisé	Rare
1762	1	2	0,30	0,01	2	34	2	1,00	Localisé	Rare
1763	1	2	0,30	0,01	2	45	2	1,00	Localisé	Rare
1935	1	2	0,30	0,01	2	45	2	1,00	Localisé	Rare

ST, pour share-type, numéro correspondant au spoligotype ; NGUF, nombre de souches du ST recensées en Guyane ; Nmonde, nombre de souches du ST recensées dans le monde (au 1<sup>er</sup> janvier 2005) ; % GUF, pourcentage du ST dans la base Guyane ; % Monde, pourcentage du ST dans la base internationale soit 29 069 souches (au 1<sup>er</sup> janvier 2005) ; Npays, nombre de pays dans lesquels le ST a été identifié ; MC, matching code ; (MC), nombre de continents dans lesquels le ST a été identifié ; SI, spreading index ; C1 et C2, cf. détail dans le texte ci-avant.

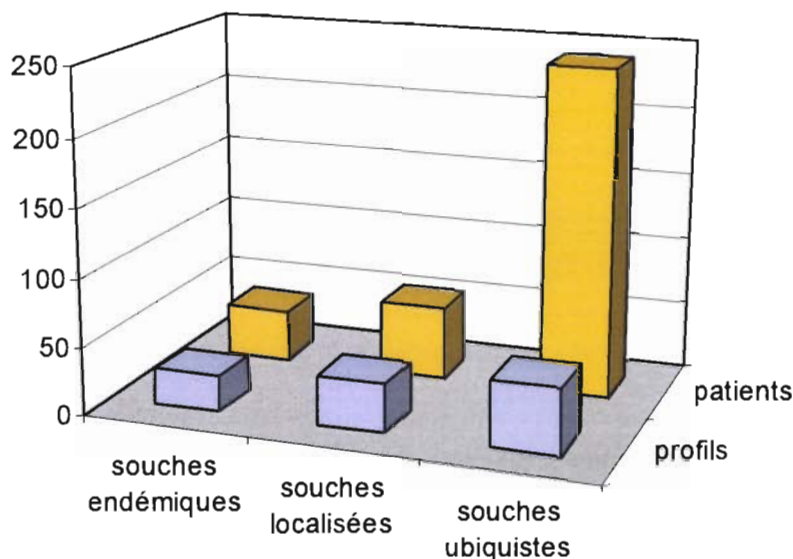


Figure 5.4 – Spécificité géographique des souches (après attribution de l'indice C1) et le nombre (1) de patients et (2) de spoligotypes concernés dans chaque classe.

### 1.3.2. Les grappes génétiques

Du point de vue de la prévalence, 71 ST (59,2% des ST) sont uniques en Guyane (une seule souche isolée par ST). Ces cas sont dits pseudo-orphelins, ou « non en grappes », par opposition aux cas dont on a recensé au moins 2 souches identiques, qui sont « en grappes » ou en *cluster*, selon le terme anglais.

Le tableau 5.2 contenant les informations sur les grappes obtenues avec les 2 méthodes moléculaires choisies, est en fait un simple résumé des données du tableau 5.1 et du typage par VNTR, dont les résultats ne sont pas détaillés ici. Avec la méthode spoligotypage + VNTR, il manque 5 cas correspondant aux cas dont la VNTR a totalement échoué. On trouve également 9 cas dont la VNTR a échoué pour au moins un des 5 ETR et pour lesquels il n'était pas possible de conclure (203 cas en grappe + 127 cas non en grappe + 9 NA = 339)

Le tableau 5.3 précise par ailleurs le nombre de cas inclus dans chaque grappe. Avec la méthode de spoligotypage seul par exemple, il existe une unique grappe incluant 48 cas, correspondant à la souche ubiquiste ST 50 ; en revanche, il existe 20 grappes, correspondant donc à 20 souches différentes, n'incluant que 2 patients. Cette information est intéressante pour mesurer la transmission entre patients, et cette analyse sera développée ultérieurement (§2.2.1.). Lorsque l'on ajoute les résultats de l'analyse par VNTR, les grappes obtenues par spoligotypage seul sont re-segmentées, ce qui aboutit à un nombre supérieur de grappes (52 au lieu de 49 grappes). De plus, la taille moyenne des grappes diminue (elle passe de 5,57 à 3,90 cas en moyenne par grappe), et la proportion des patients totaux qui sont inclus dans une grappe diminue (37,5% de souches uniques avec le couplage du spoligotypage et du typage par VNTR).

Tableau 5.2 – Grappes obtenues d'une part par la méthode de spoligotypage seule, et d'autre part en couplant le spoligotypage et le typage par VNTR (nombre d'isolats et fréquences).

<b>Spoligotypage seul</b>	
total des cas de TB étudiés	344
nombre de grappes identifiées	49
nombre moyen de cas par grappe	5,57
nombre de patients en grappes (et %)	273 (79,4%)
nombre de patients non en grappes (et %)	71 (20,6%)
<b>Spoligotypage + VNTR</b>	
total des cas de TB étudiés	339
nombre de grappes identifiées	52
nombre moyen de cas par grappe	3,90
nombre de patients en grappes (et %)	203 (59,9%)
nombre de patients non en grappes (et %)	127 (37,5%)

Tableau 5.3 – Taille des grappes (en nombre de cas), et nombre de grappes correspondantes.

<b>Spoligotypage seul</b>	<b>Spoligotypage + VNTR</b>
1 grappe de 48 cas	
1 grappe de 39 cas	1 grappe de 24 cas
1 grappe de 16 cas	1 grappe de 10 cas
1 grappe de 13 cas	2 grappes de 9 cas
1 grappe de 10 cas	1 grappe de 7 cas
2 grappes de 9 cas	2 grappes de 6 cas
3 grappes de 6 cas	7 grappes de 5 cas
5 grappes de 5 cas	8 grappes de 4 cas
4 grappes de 4 cas	5 grappes de 3 cas
10 grappes de 3 cas	25 grappes de 2 cas
20 grappes de 2 cas	

Si l'on s'intéresse aux 5 spoligotypes qui regroupent le nombre de cas le plus important (ST 2, 42, 50, 53 et 64), on trouve respectivement pour chacune de ces grappes 2, 7, 8, 14 et 4 codes VNTR différents (tableau 5.4). Pour les autres grappes, le nombre de VNTR différents oscille entre 1 et 2, plus rarement 3.

Tableau 5.4 – Nombre de cas par VNTR pour 5 spoligotypes regroupant plus de 15 cas.

<b>Spoligotype</b>	<b>N t</b>	<b>N g</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>	<b>f</b>	<b>g</b>	<b>h</b>	<b>i</b>	<b>j</b>	<b>k</b>	<b>l</b>	<b>M</b>	<b>n</b>
ST 02	16	15	10	5												
ST 42	10	5	3	2	1	1	1	1	1							
ST 50	42	39	24	7	4	2	2	1	1	1						
ST 53	39	35	6	5	5	4	4	3	2	2	2	2	1	1	1	1
ST 64	13	11	9	2	1	1										

N t, nombre de cas totaux ; N g, nombre de cas inclus dans une grappe ; a à n, codes VNTR factices. Le code du VNTR « a » n'est pas le même pour les différents ST. Par exemple, pour ST 02, a = 32333 ; pour ST 42, a = 21432. Les chiffres donnent le nombre de souches par ST+VNTR.

La figure 5.5 récapitule les informations contenues dans les tableaux 5.1 à 5.3 sur la distribution des différents ST et la taille des grappes.

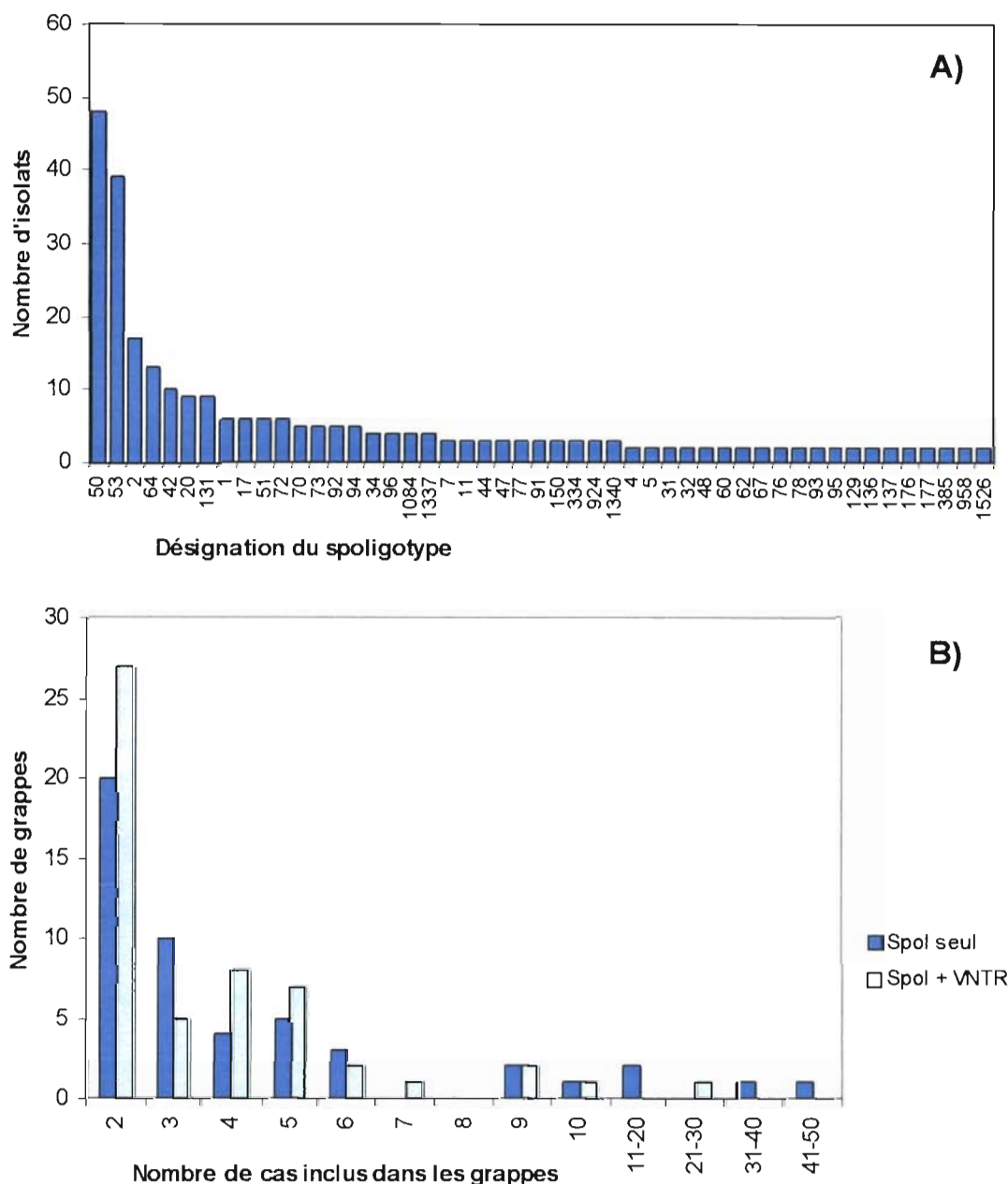


Figure 5.5 – Histogrammes récapitulatifs. (A) Nombre de cas associés aux spoligotypes recensés en Guyane, dérivé du tableau 5.1 ; le spoligotype qui recense le plus grand nombre de cas, situé le plus à gauche, est ST 50. (B) Tailles des grappes, c'est-à-dire nombre de grappes incluant 2, 3, 4... patients, en fonction de la méthode (dérivées du tableau 5.3).

Si l'on tient compte de la nationalité des cas inclus dans les grappes obtenues par spoligotypage, le calcul de l' « odds ratio » entre les cas français et étrangers nous permet de conclure qu'un cas de nationalité française a significativement deux fois moins de chance d'être inclus dans une grappe génétique qu'un cas de nationalité étrangère (OR = 1,97, IC<sub>95%</sub> = 0,98 ; 3,96). Si

l'on tient compte des grappes par spoligotypage + VNTR, (outre les 14 cas pour lesquels on n'a pas le VNTR, et 73 autres cas dont on n'a pas la nationalité), on obtient la même tendance (OR = 1,76, IC<sub>95%</sub> = 1 ; 3.09).

### Encadré 5.3 – Le calcul de l' « odds ratio »

L' « odds ratio » (ou rapport des cotes) est un outil statistique très utilisé en épidémiologie, dans les enquêtes de type cas/témoins marquées par l'impossibilité de mesurer les risques de la maladie chez les sujets exposés et les non-exposés ; dans ces situations, l' « odds ratio » va permettre une estimation du risque relatif.

	Malades Cas dans les enquêtes cas-témoins	Non malades Témoins dans les enquêtes cas-témoins
Exposés	A	B
Non exposés	C	D

$$OR \text{ (ou RC)} = (A \cdot D / B \cdot C)$$

Interprétation : les « odds » de l'exposition parmi les cas sont « OR fois » ceux des témoins. Autrement dit, les « odds » de la maladie parmi les exposés sont « OR fois » celles parmi les non exposés.

Il est possible de remplacer les malades et non malades par deux caractères (ex : en grappes ou non) dont on cherche à évaluer le risque dans deux groupes d'individus (ex : cas français et cas étrangers).

## 1.4. Discussion

### 1.4.1. Cas en grappes et contamination

Nos résultats mettent en évidence une forte diversité de spoligotypes en Guyane, avec 120 ST pour 344 patients (déclarés entre 1996 et 2003, sans pouvoir intégrer les cas non confirmés). Si 80% des 344 cas sont inclus dans des grappes (62% si l'on intègre les résultats VNTR), cela n'indique pas pour autant une contamination tuberculeuse active en Guyane. En effet, les tailles de grappe sont en moyenne plutôt faibles, ce qu'illustre bien la figure 5.5 – B : on a 20 grappes de 2 cas sur 49, soit 40,8% des grappes (ou 27 grappes de 2 cas sur 59, soit 45,8% avec le typage par VNTR). Et si l'on considère les profils plutôt que les patients, 71 ST sur 120 (59,2%) sont uniques, c'est-à-dire que ce sont des cas qui, à priori, n'ont pas été source de contamination en Guyane, du moins concernant les cas enregistrés sur la période 1996-2003.

D'autre part, les grappes importantes, incluant 48 et 39 cas pour les plus grosses, ne sont pas forcément le reflet d'une épidémie tuberculeuse. La comparaison avec la base de donnée internationale SpolDb 4 a mis en évidence l'ubiquité de ces spoligotypes, présents sur les 5 continents et dans de nombreux pays (57 pays pour ST 50 et 74 pays pour ST 53, avec des indices de dispersion SI<sup>6</sup> de 20,95 et 28,09 respectivement). La combinaison de l'analyse par typage VNTR à la première méthode a permis de re-segmenter ces deux grappes principales en 22 grappes. Malgré tout, il reste une grappe de 24 cas (ST 50, VNTR 32333) dont on ne peut affirmer qu'il corresponde effectivement à un lien épidémiologique entre les cas (5 nationalités différentes et 5 villes de résidence différentes, outre les 3 villes de l'Ile de Cayenne). Ainsi, la méthode a ses limites, en particulier pour les spoligotypes ubiquistes, et la signification épidémiologique de la clonalité des isolats ne peut alors être démontrée que par des enquêtes de terrain [Filliol *et al.*, 1999].

#### *Persistance de la tuberculose en Guyane : des hypothèses à tester*

Tout ceci prône en faveur d'une dynamique de transmission de type plutôt endémique de la tuberculose en Guyane. Mais l'incidence de tuberculose est malgré tout très forte dans cette région. Si la transmission est faible et la diversité de souches importante, deux hypothèses peuvent expliquer la persistance de l'endémie : (1) la réactivation de tuberculoses anciennes, et/ou (2) l'importation de cas de par l'immigration en provenance de pays à forte incidence de tuberculose.

Un moyen de vérifier l'hypothèse de réactivation est de comparer la moyenne d'âge des individus inclus ou non dans les grappes [Moonan *et al.*, 2004] : si les individus porteurs de souches uniques (non inclus dans une grappe) ont une moyenne d'âge significativement supérieure aux individus inclus dans des grappes, le fait d'être non inclus dans une grappe pourrait être lié à une réactivation de tuberculose ancienne chez les individus plus âgés. Dans notre étude, l'âge moyen des individus inclus dans une grappe génétique est de  $39,0 \pm 16,1$  (écart-type) et de  $38,3 \pm 15,2$  pour les individus non inclus dans une grappe. L'âge moyen ne diffère pas significativement entre ces deux groupes (test de Student ;  $p = 0,720$ ).

L'hypothèse de réactivation de tuberculoses anciennes ne semble donc pas jouer un rôle majeur dans la dynamique de la tuberculose en Guyane. L'importance de ces deux phénomènes de réactivation/importation dans la dynamique de la transmission de la tuberculose en Guyane est analysée plus avant dans la partie suivante.

---

<sup>6</sup> Pour mémoire, l'indice SI (Spreading Index) est obtenu en divisant le nombre total d'isolats d'un spoligotype donné dans la base internationale, par le nombre de pays où il a été recensé.

## 1.4.2. Spécificité géographique

### *Cas des souches endémiques*

Si l'on se limite aux données obtenues, les spoligotypes spécifiques de Guyane représentent 25,8% des ST (31/120) et 12,2% des cas (42/344). L'hypothèse d'importation de nouveaux cas par immigration évoquée ci-dessus ne peut donc expliquer à elle seule la persistance de la tuberculose en Guyane, et ce pour au moins 12,2% des cas. Une des hypothèses pour expliquer l'existence de ces souches endémiques est qu'il pourrait s'agir de souches ancestralement présentes en Guyane. D'autre part, ces souches ont pu évoluer localement par mutation au cours de l'endémie tuberculeuse, et donner des souches que l'on ne retrouve aujourd'hui nulle part ailleurs. Une dernière hypothèse serait que cette endémicité n'est qu'apparente ; ces souches existeraient ailleurs qu'en Guyane mais n'ont jusqu'à présent pas été recensées dans la base internationale sur laquelle repose notre analyse géographique.

### *Dynamique de la tuberculose en Guyane : hypothèse finale*

Au final, on peut poser l'hypothèse d'un système de micro-épidémies à double entrée, avec des souches endémiques dues à la transmission locale de souches historiquement présentes en Guyane, et des souches non-endémiques issues de l'importation de cas de pays à forte incidence de tuberculose. Il est également à noter que le taux de « *clustering* » est significativement plus important parmi les cas de nationalité étrangère que parmi les cas de nationalité française, ce qui indique que les immigrants sont davantage source de transmission.

## 1.5. Conclusion

L'analyse des grappes génétiques des souches isolées en Guyane entre 1996 et 2003 a permis d'apporter un certain nombre de réponses, ou tout du moins des hypothèses, concernant la dynamique de transmission de la tuberculose en Guyane, avec la circulation interne de souches ancestrales, et l'importation de nouvelles souches grâce aux échanges inter-frontaliers.

Mais ceci ne définit pas de façon quantitative la transmission récente de la tuberculose en Guyane, même si elle semble assez faible au vu de la taille moyenne des grappes qui est plutôt faible. Seule l'analyse phylogénétique et phylogéographique des souches peut nous permettre de calculer le taux de transmission récente, ce à quoi nous nous appliquons dans la seconde partie de ce chapitre. À noter également que ces résultats n'expliquent pas l'existence des agrégats spatiaux de cas de tuberculose observés sur l'Île de Cayenne.

## 2. Diversité des souches de *Mycobacterium tuberculosis* et phylogénie<sup>7</sup>

### 2.1. Introduction

#### 2.1.1. La méthode n et la méthode n-1

Comme nous l'avons expliqué dans le chapitre 6, l'épidémiologie moléculaire est basée sur le concept que les cas de maladie ayant un lien épidémiologique sont porteurs des mêmes souches de pathogènes, alors que les cas n'ayant aucun lien entre eux diffèrent sur certains loci du génome du pathogène [Murray, 2002]. Les isolats de *M. tuberculosis* qui sont inclus dans des grappes génétiques -car portant les mêmes empreintes génétiques- sont censés représenter d'une chaîne de transmission récente ; ceux ayant une empreinte génétique unique sont susceptibles d'être plutôt dus à une infection acquise moins récemment [Burgos et Pym, 2002 ; McConkey *et al.*, 2002].

Deux méthodes différentes sont couramment utilisées pour estimer la proportion de cas en grappes ; ce sont les méthodes dites n et n-1 [Glynn *et al.*, 1999 ; Murray, 2002]. Elles consistent toutes deux à calculer des indices de transmission récente (ou RTI pour *recent transmission index*) [Tanaka et Francis, 2005], basés sur l'idée intuitive que plus les données de l'échantillon sont homogènes génétiquement, plus la transmission est sévère.

#### *La méthode n*

Le premier indice est  $RTI_n = n_c / n$  [Alland *et al.*, 1994], où n est le nombre total de cas dans l'échantillon, et  $n_c$  le nombre total de cas inclus dans des grappes de deux cas ou plus. Cet indice est uniquement dépendant de la taille de l'échantillon et du nombre des cas uniques génétiquement.

#### *La méthode n-1*

Le second indice est  $RTI_{n-1} = (n_c - c)/n$  [Small *et al.*, 1994], où n est le nombre total de cas de l'échantillon, c le nombre de génotypes représentés par au moins deux cas, et  $n_c$  le nombre total de cas inclus dans des grappes de deux cas ou plus. Cet indice dépend de la taille de l'échantillon et du nombre de génotypes.

Contrairement à la méthode n, qui utilise le nombre total de cas inclus dans des grappes, la méthode n-1 pose l'hypothèse que un cas par grappe est un cas de réactivation d'une tuberculose ancienne, et donc élimine un cas par grappe pour évaluer le nombre réel de cas en grappes.

---

<sup>7</sup> Une publication est actuellement en préparation sur cette partie : Guernier *et al.* (2006) Using cluster-graphs from spoligotype data to quantify the extent of recent transmission of tuberculosis in French Guiana.

### 2.1.2. Taux de mutation des marqueurs moléculaires

Les deux méthodes décrites ci-dessus ne tiennent pas compte des évènements de mutation qui peuvent augmenter artificiellement la diversité des souches. Un taux de mutation élevé du marqueur choisi aura tendance à réduire la taille des grappes dans les isolats issus d'une épidémie, ce qui peut conduire à une sous-estimation de la transmission si l'hétérogénéité des données est due à la vitesse de mutation plutôt qu'à une faible transmission. Par exemple, il a été démontré que le marqueur IS6110 évolue à une vitesse suffisante pour produire des évènements de mutation sur de courtes périodes de temps [Yeh *et al.*, 1998 ; Niemann *et al.*, 1999 ; Rosenberg *et al.*, 2003]. À l'inverse, l'observation d'une forte homogénéité des génotypes porte à conclure à une forte transmission, alors que ce résultat peut être dû à un processus de mutation très faible.

Le tableau 5.5 résume les problèmes d'interprétation liés au taux de mutation des marqueurs génétiques. Il est donc important d'intégrer le taux de mutation des marqueurs dans les études d'épidémiologie moléculaire qui tentent d'évaluer la sévérité des épidémies [Tanaka et Francis, 2005].

Tableau 5.5 – Déductions concernant le taux de transmission de tuberculose en fonction de la diversité des génotypes observés (d'après [Tanaka et Francis, 2005]).

Taux de mutation	Données génétiques homogènes	Données génétiques hétérogènes
Élevé	Épidémie sévère	Ambigu
Faible	Ambigu	Épidémie moyenne

## 2.2. Méthodes

### 2.2.1. L'indice de mutation-transmission

Un modèle de mutation basique couplé à l'organisation des données sous forme de « *cluster-graph* » (description ci-après) peuvent être utilisés pour développer un indice alternatif aux  $RTI_n$  et  $n-1$  : le TMI (*transmission mutation index*) [Tanaka et Francis, 2005].

$$TMI = \tilde{\mu} \frac{n-g+\nu_1}{\nu_1}$$

avec  $\tilde{\mu}$  défini comme une estimation du taux de mutation du marqueur génétique choisi,  $n$  le nombre total de cas considérés,  $g$  le nombre de génotypes différents, et  $\nu_1$  le nombre d'évènements de mutations en un pas attribuables

aux données. L'hypothèse qui sous-tend ce modèle est la théorie des allèles infinis, qui suppose que chaque mutation est unique et aboutit à un génotype différent ; dans ce cas, un locus génétique particulier est le même chez deux individus si et seulement si aucune mutation n'a eu lieu (pas de mutations convergentes). En réalité, cette hypothèse n'est pas tout à fait exacte dans le cas des spoligotypes pour lesquels de rares cas de convergence génétique ont été démontrés [Warren *et al.*, 2002].

Le TMI, toujours compris entre 0 et 1, est interprété sous la forme d'un pourcentage. Par exemple, si on trouve un TMI de 0,4, on en conclut que 40% des cas en grappes résultent d'un événement de transmission récente.

### 2.2.2. Visualiser les liens génétiques entre souches

La classification des souches bactériennes en grappes de clones génétiques ne fournit aucune information sur les liens évolutifs génétiques entre souches, qui sont obtenus par phylogénie, domaine d'étude qui consiste à retracer l'arbre évolutif entre souches. Un compromis entre ces deux types d'information (phylogénie et grappes) peut être obtenu en construisant ce que Tanaka et Francis (2005) ont appelé des « *cluster-graph* » (nous avons conservé ce terme).

#### *Définition*

Le *cluster-graph* est un mode de représentation graphique des liens phylogéniques entre grappes. Les génotypes sont illustrés par des cercles, et les traits entre les cercles (ou nœuds) traduisent des liens évolutifs directs, c'est-à-dire des événements mutationnels (mutation en un pas<sup>8</sup> ou en plusieurs pas). La taille des grappes associée à chaque génotype est illustrée par la taille du cercle, et/ou un chiffre à l'intérieur du cercle correspondant au nombre de cas [Tanaka et Francis, 2005] (figure 5.6). Le spoligotypage produit en général des données qui sont plus riches en signaux phylogénétiques que des données issues par exemple de IS6110 [Tanaka et Francis, 2005]. De ce fait, les *cluster-graphs* obtenus par spoligotypage sont plus informatifs.

#### *Notre représentation graphique*

La formule de l'indice de mutation-transmission ne tient compte que des mutations en un pas ( $v_1$ ). Lorsqu'il faut au moins deux événements de mutation pour passer d'un génotype à un autre, Tanaka et Francis (2005) n'ont établi aucun lien (ou trait) entre les deux génotypes, aboutissant donc à deux graphes indépendants. Afin d'améliorer cette représentation, nous avons choisi d'ajouter aux *cluster-graphs* de grands cercles concentriques permettant de

---

<sup>8</sup> Une mutation en un pas (en anglais « *one-step mutation* » ou « *single-step mutation* ») traduit simplement le fait qu'un seul événement de mutation est à l'origine du génotype observé.

visualiser les sauts évolutifs entre génotypes. Le spoligotype le plus ancestral est placé au centre des cercles, et les spoligotypes issus d'évènements mutationnels successifs sont situés sur les différents cercles. Les spoligotypes les moins mutés (donc à priori plus anciens) sont situés sur les cercles les plus proches du centre ; à l'inverse, les spoligotypes les plus récents (qui ont connu plusieurs mutations successives dans le temps) sont sur les cercles les plus extérieurs (figure 5.6 – B).

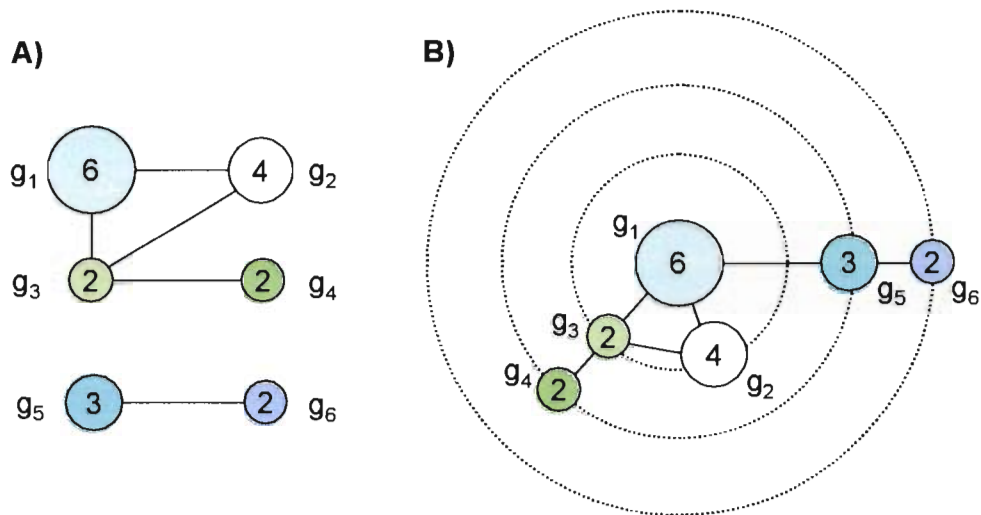


Figure 5.6 – *Cluster-graphs* représentant six génotypes ( $g_1$  à  $g_6$ ). Un nœud représente une grappe de cas ayant le même génotype, et la taille des grappes est indiquée à l'intérieur du cercle. (A) Représentation d'après [Tanaka et Francis, 2005]. (B) Représentation personnelle, plus informative. Les cercles en pointillés indiquent les sauts mutationnels : entre  $g_1$  et  $g_3$  il y a une mutation, alors qu'entre  $g_1$  et  $g_4$  il y en a deux.

### Paramètres de l'équation

Si l'on reprend l'exemple de la figure 5.6 – A, les paramètres du calcul de l'indice de mutation-transmission sont les suivants :  $g = 6$  génotypes ;  $n = 6+4+2+2+3+2 = 19$  isolats ou cas. Pour calculer  $v_1$ , il faut prendre en compte indépendamment les différentes unités du *cluster-graph* qui ne sont pas connectées entre elles ; dans l'exemple considéré, il y a une unité incluant  $g_1$  à  $g_4$ , et une seconde incluant  $g_5$  et  $g_6$ . Pour chaque unité, le nombre de mutations en un pas est obtenu en soustrayant 1 au nombre de génotypes, donc au nombre de nœuds ; puis on additionne le résultat de chaque unité pour obtenir le nombre de mutations total :  $v_1 = (4-1) + (2-1) = 4$ .

Avec la figure 5.6 – B, toutes les « unités » sont connectées entre elles sur le graphe, puisque l'on ne tient pas compte uniquement des sauts mutationnels en un pas. Mais les calculs restent exactement les mêmes :  $g$  et  $n$  sont identiques, et pour le calcul de  $v_1$  il suffit de considérer comme des unités indépendantes les groupes de génotypes séparés par plus d'un évènement mutationnel, comme c'est le cas entre les génotypes  $g_1$  et  $g_5$ .





### *Phylogénie incluant les VNTR*

Différentes études phylogénétiques ont été réalisées afin d'établir les liens entre les différentes familles de *M. tuberculosis*, et avec les autres espèces du complexe *M. tuberculosis* [Sola *et al.*, 1999 ; Sola *et al.*, 2000 ; Sola *et al.*, 2001a ; Duchêne *et al.*, 2004]. Ces arbres tendent à se préciser au fil de l'approvisionnement des bases de données internationales en spoligotypes, et de l'inclusion des données MIRU [Gutacker *et al.*, 2002]. Afin de permettre la comparaison de nos données avec celles basées sur des méthodes phylogéniques standardisées, nous avons procédé à l'analyse phylogénique de nos isolats en couplant spoligotypage et VNTR, et ce grâce à la méthode « *minimum spanning trees* » (arbres couvrants minimaux), et au logiciel BioNumerics (Applied Maths, Kortrijk, Belgium).

#### **Encadré 5.5 – Phylogénie et arbres couvrants minimaux**

La phylogénie (ou phylogenèse) est une reconstruction de l'histoire évolutive des êtres vivants. Le terme phylogenèse, introduit par Haeckel en 1860, désigne "l'histoire du développement paléontologique des organismes". Un arbre phylogénétique est une représentation graphique de la phylogenèse d'un groupe de plusieurs taxa, les nœuds représentant les unités taxonomiques et les branches définissant les relations entre les taxa en terme de descendance [Ridley, 1997].

Lorsque l'on considère une matrice de distances entre  $n$  taxa, l'arbre couvrant minimal (ACM) est l'arbre qui connecte les taxa entre eux de façon à minimiser la distance additionnelle de toutes les branches de l'arbre (la somme des arêtes du graphe). La technique est basée sur le principe de parcimonie qui postule que, pour un groupe d'espèces, la phylogénie la plus vraisemblable est celle qui nécessite le plus petit nombre de changements évolutifs. L'arbre phylogénétique des espèces est conçu de manière à impliquer le minimum d'événements évolutifs [Darlu et Tassy, 1993]. Le taxon ayant le maximum de connexions est désigné comme nœud central, ce qui donne des arbres en étoile, avec une classification correcte pour les systèmes où le taux de mutation ou de recombinaison est suffisamment important pour permettre la production d'un grand nombre de variants à un locus (ou SLV pour single locus variant) à partir d'un type commun [Smith *et al.*, 1993].

*Avantages de BioNumerics.* Pour calculer un ACM, deux conditions importantes doivent être respectées : (1) l'étude doit porter sur une courte période de temps, afin de garantir que tous les types existent toujours, et (2) les données doivent être suffisamment complètes pour permettre la construction d'un arbre cohérent. BioNumerics utilise un algorithme qui présente l'avantage de créer des types hypothétiques lorsque des données sont manquantes (types intermédiaires) [Applied Maths BVBA, 2000].

#### **2.2.4. Construction des « *cluster-graphs* »**

##### *Méthode*

La construction d'un *cluster-graph* passe par au moins trois étapes : (1) l'identification des paires de spoligotypes (génotypes) reliées entre elles par un ou plusieurs événement de mutation, (2) l'attribution à chaque cercle illustrant un spoligotype d'un nom et d'une taille (taille de la grappe), et (3) la création de traits entre cercles illustrant les liens phylogénétiques entre spoligotypes.

La construction des *cluster-graphs* procède donc d'une analyse purement visuelle des événements de délétions d'un ou plusieurs espaceurs, ces événements permettant de passer d'un spoligotype à l'autre. À l'heure actuelle, un outil informatique baptisé Spolnet permettant d'établir ces liens de façon automatisée est développé à l'Institut Pasteur de Guadeloupe.

##### *Représentation graphique*

Le logiciel « BioLayout » est un logiciel de visualisation de graphes, développé entre autres pour visualiser les similarités entre protéines sur la base de leurs enchaînements d'acides aminés [Enright et Ouzounis, 2001 ; Goldovsky *et al.*, 2005]. Nous avons utilisé cet outil pour la représentation graphique de nos *cluster-graphs*.

##### *Spoligotypes absents de Guyane*

Afin de proposer une représentation graphique visuellement correcte de la phylogénie des souches (spoligotypes), nous avons par ailleurs intégré aux *cluster-graphs* que nous avons construit l'ensemble des spoligotypes intermédiaires, même lorsque ceux-ci n'étaient pas présents en Guyane sur la période 1996-2003. Lorsque ces spoligotypes intermédiaires n'ont pas été indiqués sur les graphes, c'est qu'ils n'étaient pas présents dans la base de donnée internationale SpolDb4 consultée à l'Unité de la Tuberculose et des Mycobactéries de l'Institut Pasteur de Guadeloupe au 1<sup>er</sup> janvier 2005.

### **2.3. Résultats**

#### **2.3.1. Les familles de spoligotypes**

Le tableau 5.6 résume la classification en familles et sous-familles de l'ensemble des spoligotypes isolés en Guyane entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003. Lorsqu'il existe une incertitude (pour certains spoligotypes qui respectent deux règles de classification à la fois), les deux sous-familles ont été notées (voir par exemple ST 95, tableau 5.6).



CHAPITRE 5 : Épidémiologie moléculaire de la tuberculose en Guyane

Type	Famille	Description binaire
041	LAM7 TUR	██████████
004	LAM3 ou S	██████████
042	LAM	██████████
150	LAM	██████████
177	LAM	██████████
274	LAM	██████████
492	LAM	██████████
orph 11	LAM	██████████
orph 14	LAM	██████████
orph 15	LAM	██████████
046	likely Haarlem	██████████
088	S	██████████
034	S	██████████
005	T	██████████
031	T	██████████
077	T	██████████
453	T	██████████
926	T	██████████
1084	T	██████████
1284	T	██████████
1318	T	██████████
1526	T	██████████
1761	T	██████████
orph 18	T	██████████
orph 19	T	██████████
orph 24	T	██████████
orph25	T	██████████
007	T1	██████████
053	T1	██████████
063	T1	██████████
086	T1	██████████
118	T1	██████████
240	T1	██████████
245	T1	██████████
334	T1	██████████
344	T1	██████████
orph 17	T1	██████████
orph 20	T1	██████████
051	T2	██████████
052	T2	██████████
078	T2	██████████
073	T2 ou 3	██████████
1339	T2 ou 3	██████████
037	T3	██████████
102	T3	██████████
131	T3	██████████
1214	T3	██████████
orph 16	T3	██████████
094	T4	██████████
044	T5	██████████
058	T5 MAD2	██████████
385	X1	██████████
137	X2	██████████
347	X2	██████████
014	X3	██████████
070	X3	██████████
091	X3	██████████
092	X3	██████████
1486	X3	██████████

orph, spoligotype orphelin. Les numéros sont arbitraires et servent simplement à les identifier dans les *cluster-graphs*. Description binaire, présence/absence des 43 IDR (carrés noirs/blancs). Les spoligotypes notés en gras correspondent aux prototypes de chaque famille.

### 2.3.2. Les « clusters-graphs »

Le tableau 5.7 récapitule le nombre de cas et de spoligotypes différents recensés en Guyane pour chaque famille. La majorité des cas de notre étude (32,6%) sont de la famille T<sup>9</sup>, suivi de LAM (23,3%), Haarlem (22,7%), EAI (12,2%) et X (5,8%). Les autres familles sont minoritaires (moins de 2%).

Tableau 5.7 – Nombre de cas et de spoligotypes (ST) de Guyane (1996-2003) par famille.

Famille	Nb de cas	Nb de ST
Africanum	1	1
Beijing	5	1
EAI	42	24
Haarlem	78	12
LAM	80	33
likely H	1	1
S	5	2
T	112	38
X	20	8
<b>Total</b>	<b>344</b>	<b>120</b>

Pour plus de lisibilité des *cluster-graphs* les familles majoritaires en nombre de cas sont représentées sur des figures indépendantes (figures 5.8 à 5.12). Une figure finale présente la phylogénie de l'ensemble des spoligotypes (figure 5.13).

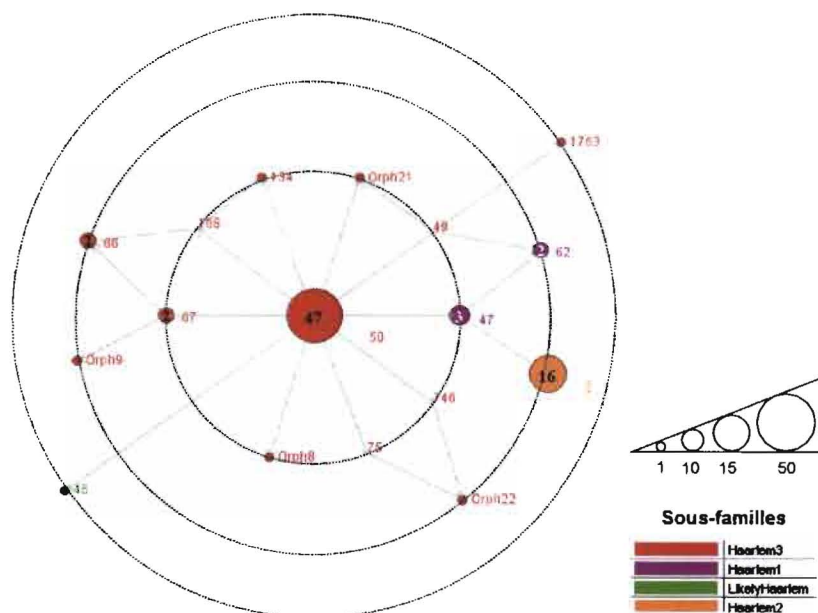


Figure 5.8 – *Cluster-graph* de la famille Haarlem. Sur 17 spoligotypes (ST) représentés, 13 ont été recensés en Guyane (1996-2003), dont ST 46 de la famille « *likely Haarlem* ». La taille des cercles définit la taille des grappes. Quand il est > 1cas, le nombre de cas est inscrit dans le cercle. Les désignations des spoligotypes sont inscrites près des nœuds.

<sup>9</sup> La famille T n'est pas une vraie famille génétique homogène. Elle a été définie par défaut et inclut des génotypes encore mal connus, caractérisés uniquement par l'absence des espaceurs 33 à 36.

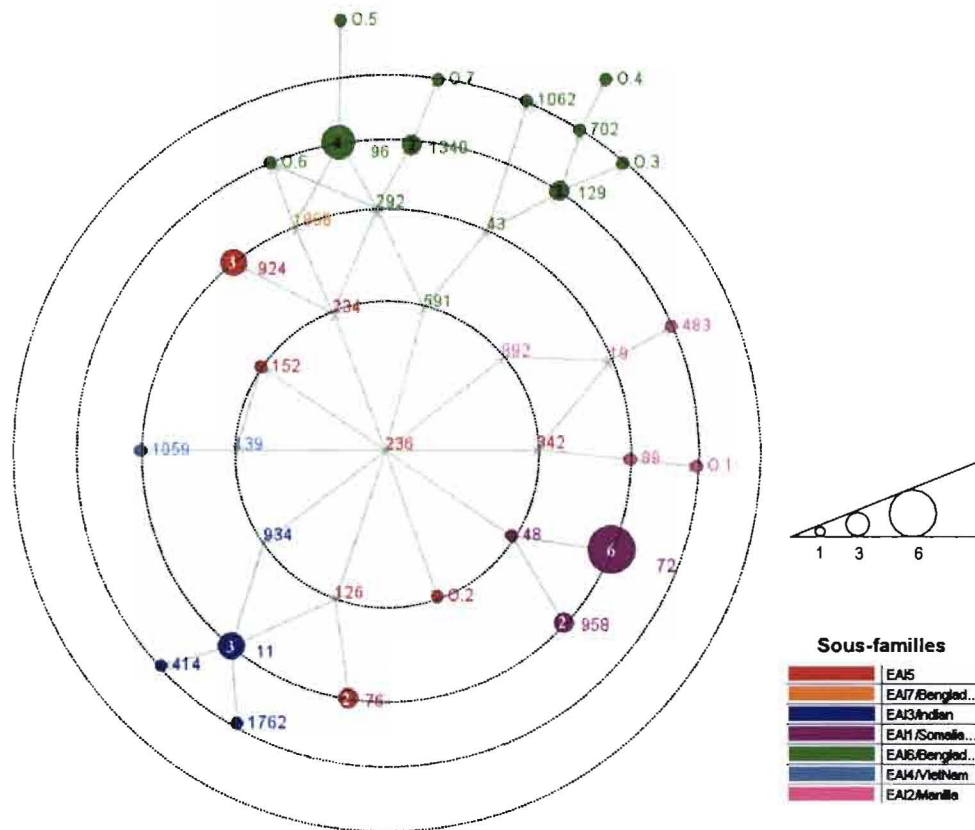


Figure 5.9 – *Cluster-graph* de la famille EAI. Sur 36 spoligotypes représentés, 24 ont été recensés en Guyane (1996-2003). ST 236 serait l'ancêtre commun des autres spoligotypes de cette famille mais il n'a pas été recensé en Guyane (du moins sur la période étudiée).

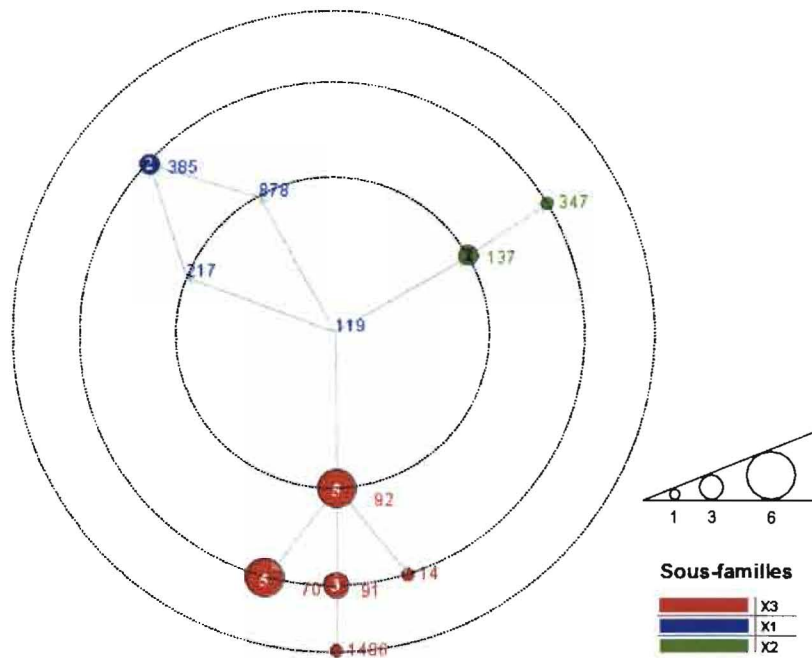


Figure 5.10 – *Cluster-graph* de la famille X. Sur les 11 spoligotypes représentés, 8 ont été recensés en Guyane (1996-2003). L'ancêtre commun probable, ST 119, n'a pas été recensé.

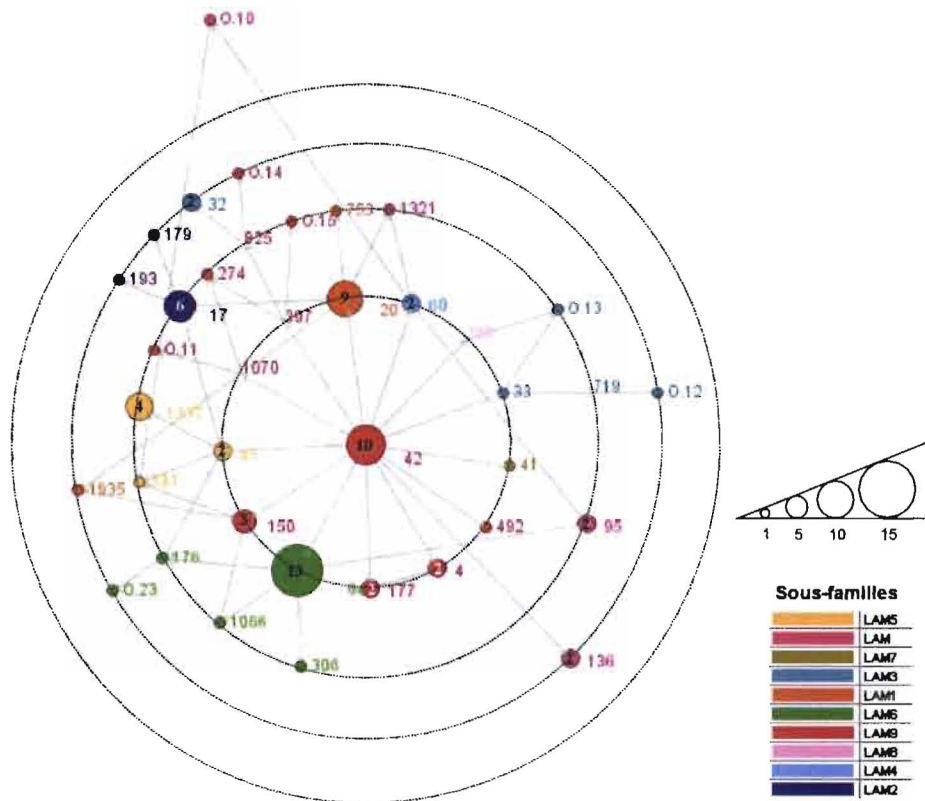


Figure 5.11 – *Cluster-graph* de la famille LAM. Sur 38 ST représentés, 33 ont été recensés en Guyane (1996-2003). ST 42 est l'ancêtre commun de cette famille.

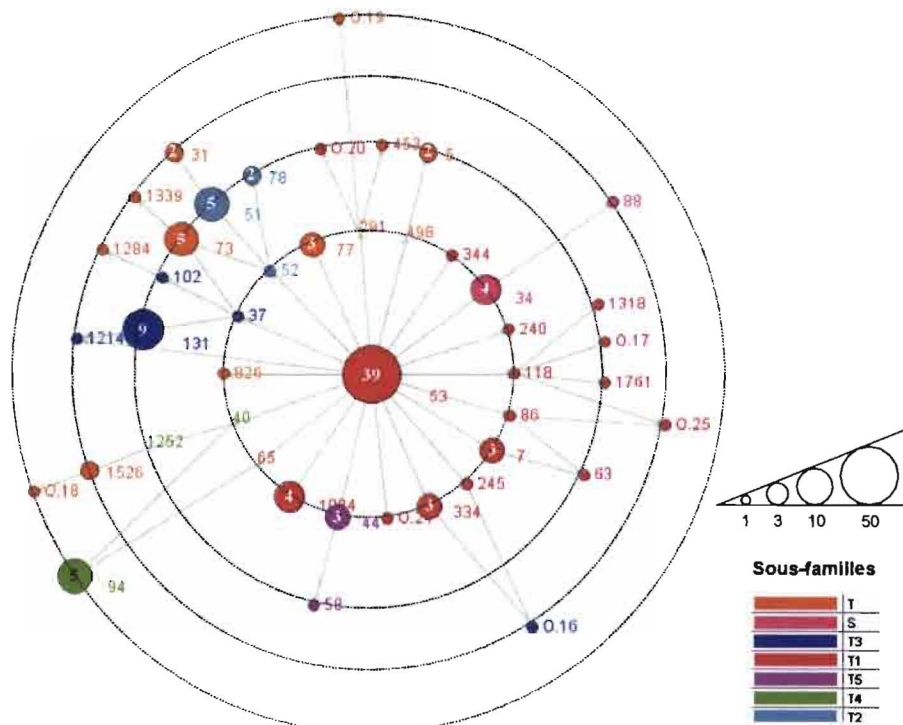


Figure 5.12 – *Cluster-graph* des familles T et S. Sur 43 ST de la famille T représentés, 38 ont été recensés en Guyane (1996-2003). Les 2 spoligotypes de la famille S ont été recensés.

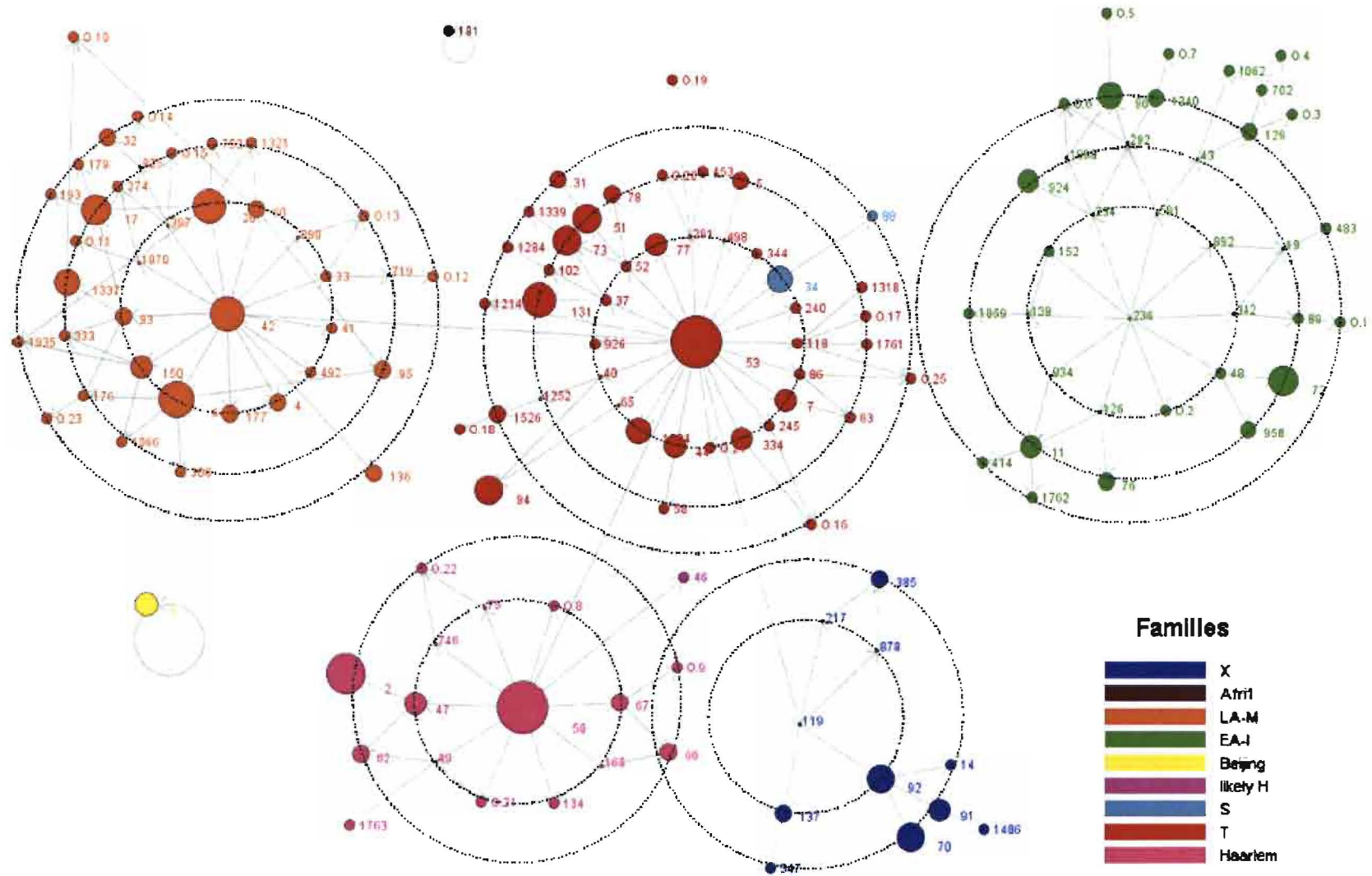


Figure 5.13 – Cluster-graph complet avec les 120 spoligotypes recensés en Guyane entre 1996 et 2003, auxquels s'ajoutent 29 spoligotypes absents de Guyane, qui sont schématisés pas un petit point (ST 119 par exemple, en bleu ci-dessus).

### 2.3.3. Méthode n, méthode n-1 et TMI

#### *Indice de mutation-transmission*

Les *cluster-graphs* que nous avons construits permettent au final de calculer l'indice de mutation-transmission (TMI).

Le nombre de cas totaux est  $n = 344$  ; le nombre de spoligotypes différents est  $g = 120$ . Le taux de mutation associé aux spoligotypes a été précédemment estimé [Tanaka et Francis, 2005] et est égal à 0,039 évènements de mutation par an. Il existe 65 spoligotypes connectés à ST 53, soit une « unité » de 66 ST. À cette unité s'ajoutent 9 unités indépendantes de 2 à 5 spoligotypes. Il existe par ailleurs 28 spoligotypes qui, pour être reliés au reste des spoligotypes, nécessitent au moins deux sauts mutationnels et ne sont donc pas pris en compte dans le calcul de  $\nu_1$ . On obtient au final :  $\nu_1 = (66-1) + (2-1) + (2-1) + (5-1) + (3-1) + (3-1) + (2-1) + (5-1) + (2-1) + (2-1) = 82$ .

$$TMI = \tilde{\mu} \frac{n - g + \nu_1}{\nu_1} = 0,039 \frac{344 - 120 + 82}{82} = 0,146$$

#### *Méthodes n et n-1*

À titre comparatif, les méthodes n et n-1 ont également été testées. Rappelons pour les calculs des indices de transmission récente (RTI) qu'il y a 344 cas confirmés de tuberculose, répartis en 120 spoligotypes dont 25 sont des spoligotypes orphelins ; 71 cas ont des spoligotypes uniques, et les 273 autres cas sont inclus dans 49 grappes.

$$RTI_n = n_c / n = 273 / 344 = 0,794$$

$$RTI_{n-1} = (n_c - c) / n = (273-49) / 344 = 0,651$$

Les valeurs de RTI avec les méthodes n et n-1 sont particulièrement élevées par rapport à la valeur de l'indice TMI qui indique 14,6% de transmission récente, contre 79,4 et 65,1% avec  $RTI_n$  et  $RTI_{n-1}$  (résultats discutés §2.4.3.).

### 2.3.4. Phylogénie spoligotype + VNTR

La méthode de phylogénie ACM (arbres couvrants minimaux) combinant deux marqueurs génétiques indépendants (spoligotypes et VNTR-A à E) donne l'arbre phylogénétique présenté sur la figure 5.14. Le résultat montre une grappe centrale dont partent trois branches principales. L'examen détaillé des souches incluses dans chacune des branches (figure 5.15), ne permet pas de mettre en évidence un regroupement par familles. Outre la branche D qui regroupe uniquement des spoligotypes de la famille LAM, la branche A présente des spoligotypes des familles T, Haarlem, Beijing, LAM et X, la branche B des spoligotypes des familles Haarlem, T, S et X, et la branche C des spoligotypes des familles Haarlem, EAI, LAM, likely Haarlem, et Africanum.

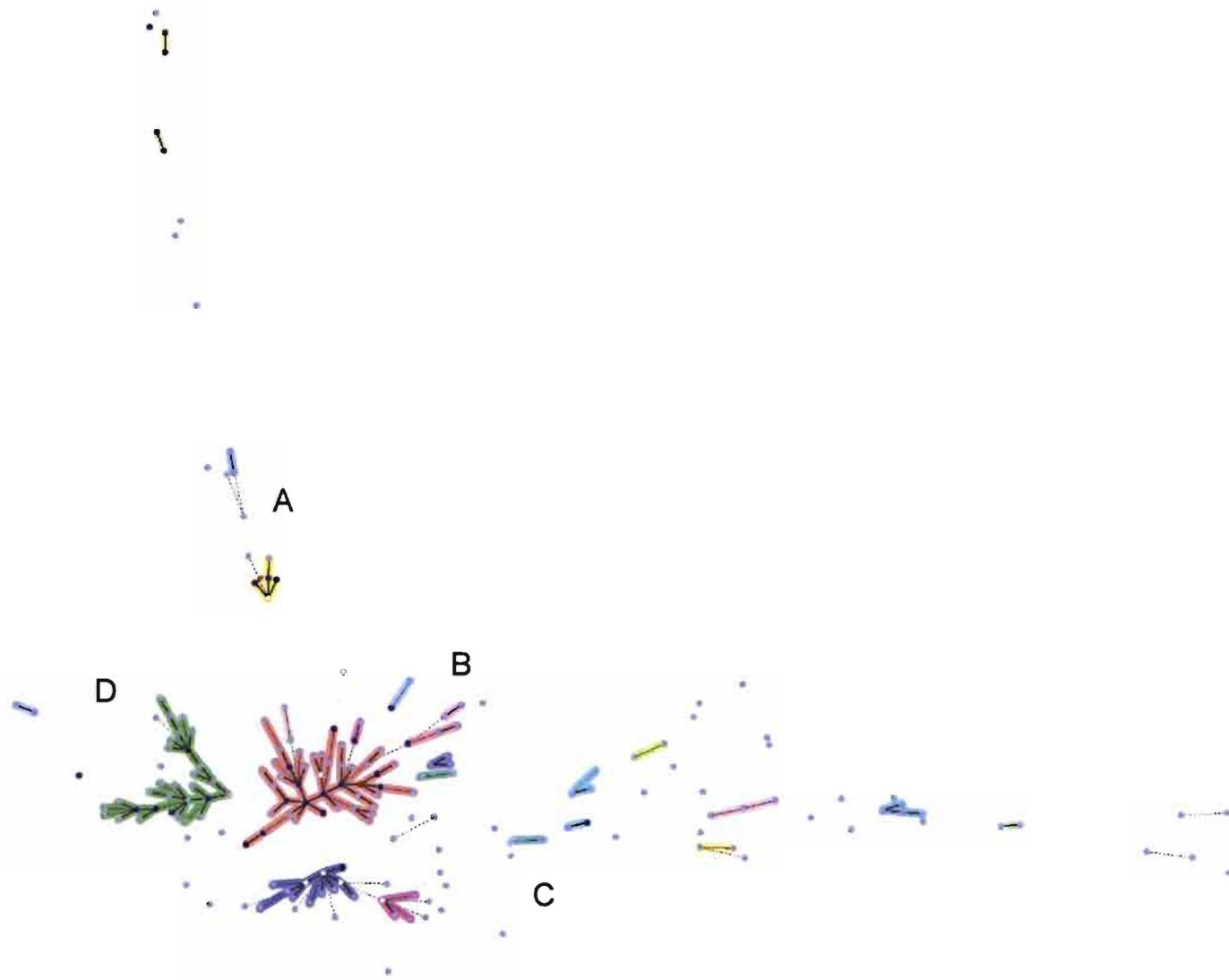


Figure 5.14 – Arbre couvrant minimal des souches de *Mycobacterium tuberculosis* isolées en Guyane entre 1996 et 2003, construit sur la base de deux marqueurs génétiques indépendants : spoligotypes et VNTR. À chaque petit cercle correspond un génotype. Les paramètres choisis dans BioNumerics pour la construction de cet arbre incluent la création de génotypes hypothétiques, matérialisés par les petits cercles blancs .

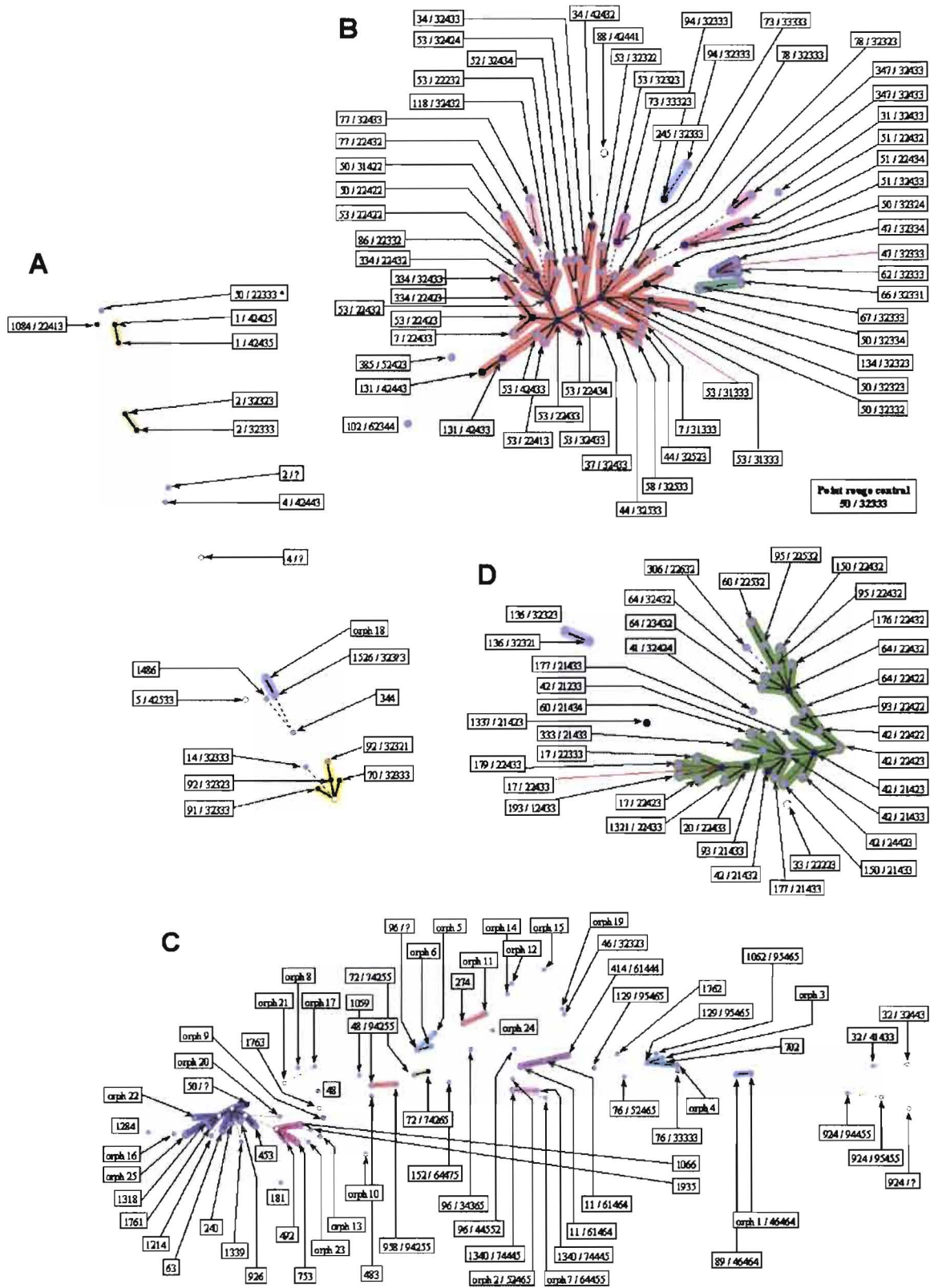


Figure 5.15 – Zoom sur les branches A à D de l'arbre phylogénétique de la figure 5.14. Chaque petit cadre décrit les données moléculaires correspondant au génotype. Le premier chiffre indique le spoligotype, le deuxième indique le codage VNTR-A à E.

## 2.4. Discussion

### 2.4.1. Arbre couvrant minimal et phylogénie manuelle

La comparaison des résultats obtenus avec chacune des deux méthodes d'analyse phylogénétique permet de mettre en évidence que la méthode manuelle que nous avons appliquée donne des résultats plus cohérents que l'ACM. En effet, l'algorithme utilisé par BioNumerics tient compte du maximum de similarité entre spoligotypes, ces similarités ne correspondant pas pour autant aux mécanismes d'évolution des spoligotypes, c'est-à-dire à des événements de délétion des espaceurs. La figure 5.16 met parfaitement en évidence ce résultat. Les spoligotypes 4, 2 et 1 sont reliés phylogénétiquement sur la branche A de l'arbre (figure 5.15) du fait de leur similarité du point de vue du nombre d'espaceurs qui diffèrent, mais ils appartiennent pourtant à trois familles différentes : Haarlem, LAM et Beijing.

Type	Famille	Description binaire
004	LAM9	
002	Haarlem2	
001	Beijing	

Figure 5.16 – Description binaire des spoligotypes 4, 2 et 1 (présence/absence des 43 espaceurs) et familles auxquelles ils appartiennent.

En fait, la méthode des arbres couvrants minimaux (ACM) est surtout productive lors d'un recrutement exhaustif, sans allèles manquants, et sur un territoire bien circonscrit, et une population close (pas de flux de population) [comm. pers. C. Sola], ce qui n'est pas le cas ici.

Lors de l'interprétation ultérieure des résultats de phylogénie, nous ne parlerons donc que des *cluster-graphs* que nous avons construits manuellement sur la base des mutations.

### 2.4.2. Cluster-graphs

#### *Familles présentes et histoire de la Guyane*

La majorité des souches (33%) sont de la super-famille T, qui est largement répandue (ubiquiste), et qui nous apporte de ce fait peu d'informations. En revanche, les autres résultats obtenus sont cohérents avec l'histoire migratoire de la Guyane française. Ainsi, l'existence de 23% de souches Haarlem, majoritairement européennes, peut correspondre à l'introduction historique de souches en provenance d'Europe ; la présence de 23% de souches LAM est également cohérente avec la situation géographique de la Guyane, et les

échanges avec les pays d'Amérique du Sud dont elle est limitrophe et dont nous avons parlé dans la première partie au sujet du contexte guyanais.

#### *Cas particulier des souches EAI*

Concernant les souches EAI (12% des souches), il est intéressant de noter que les souches susceptibles d'être les plus ancestrales sont absentes de notre jeu de données : absence de la souche centrale ST 236 ; absence de 7 souches sur 10 sur le premier cercle (marquant une unique mutation de ST 236) ; puis encore des souches intermédiaires manquantes sur le deuxième cercle. En revanche, certaines souches observées nécessitent au moins cinq sauts mutationnels pour évoluer depuis ST 236.

Ce résultat pourrait être lié à l'introduction tardive en 1977 d'une population d'agriculteurs Hmongs du Laos, même si la première introduction (ou tentative d'introduction) d'une communauté asiatique en Guyane est beaucoup plus vieille, puisqu'elle date de 1820. Il est possible que (1) les premiers asiatiques n'aient pas emmené avec eux de tuberculose, et/ou que (2) les souches introduites au début du XX<sup>ème</sup> siècle aient évolué en Guyane pour aboutir aux souches EAI observées actuellement, dont 7 sur 24 (29%) correspondent à des spoligotypes orphelins.

### **2.4.3. Transmission récente de la tuberculose en Guyane**

#### *Interprétation du TMI*

Le calcul du TMI donne un taux de transmission récente de la tuberculose en Guyane de 0,146 (soit 14,6%) sur la période 1996-2003. Ce résultat est cohérent avec les informations obtenues dans la première partie de ce chapitre, à savoir que la transmission de la tuberculose en Guyane se fait plutôt sur un mode endémique, avec peu de contamination locale, mais une persistance de la maladie due à l'importation de cas de pays à forte incidence et/ou à des réactivations de tuberculose ancienne.

#### *Discussion de la méthode*

Concernant la méthode utilisée, il existe deux biais pouvant conduire à des conclusions opposées, l'un conduisant à la mise en évidence de plusieurs épidémies qui n'en sont en fait qu'une, l'autre conduisant à la mise en évidence d'une seule épidémie<sup>10</sup> alors qu'il y en a en fait plusieurs.

---

<sup>10</sup> L'usage du terme « épidémie » peut porter à caution dans la mesure où tous les cas de tuberculose peuvent être considérés comme faisant partie de la même -très vieille- épidémie. Une épidémie désigne, en général, un seuil d'un certain nombre de cas recensés sur une période de temps donnée ; elle désigne ici un groupe de cas de tuberculose issus d'une unique source de contamination, sans seuil temporel ou quantitatif défini.

1. Si des cas importés ou issus d'une réactivation sont porteurs de souches qui diffèrent par une unique mutation d'une souche à l'origine d'une épidémie, ces cas seront associés au *cluster-graph*, et donc à l'épidémie, alors qu'ils ont abouti au même génotype de façon indépendante. L'hypothèse des allèles infinis qui sous-tend la méthode exclut cette possibilité qui peut pourtant exister.
2. Lorsque les cas inclus dans l'étude des grappes représentent un échantillon de la population étudiée, il est possible que certaines souches nous échappent, par simple biais d'échantillonnage. Les problèmes liés à ces biais ont déjà été soulevés dans la littérature [Glynn *et al.*, 1999 ; Murray, 2002]. Un échantillon étant par nature incomplet, l'échantillonnage des données peut conduire à une sous-estimation des grappes. Si les données étudiées représentent un échantillon, il est possible que l'on observe une unique épidémie, alors qu'il y en a en fait plusieurs.
3. Enfin, on peut discuter un troisième biais potentiel lié à cette méthode. L'hypothèse faite ici est que, lorsqu'un individu A infecte un individu B, les génotypes des bactéries que portent A et B diffèrent au maximum d'une seule mutation. Cette hypothèse est vraisemblable dans la mesure où le taux de mutation des spoligotypes est faible. Il semble que les possibilités d'erreur liées aux points 1 et 2 soient plus importantes que celles liées à ce dernier point.

Dans notre jeu de données, nous avons recensé les spoligotypes 53 et 137 reliés entre eux par deux sauts mutationnels, mais pas ST 119 qui est intermédiaire. Soit ST 119 existe bien en Guyane mais n'a pas été recensé (biais de type 2), soit ST 137 n'est pas relié à ST 53 et fait partie d'une épidémie indépendante. C'est cette seconde possibilité qui est prise comme hypothèse dans le modèle de calcul de TMI présenté ici. Une troisième hypothèse est la possibilité d'une homoplasie, c'est-à-dire que ST 137 pourrait être le produit d'une histoire évolutive indépendante à partir d'un autre génotype ancestral, mais ayant conduit à un génotype proche.

#### **2.4.4. Susceptibilité génétique et tuberculose**

L'expression de la tuberculose résulte d'interactions complexes entre le bacille tuberculeux, les facteurs du milieu, et des facteurs propres à l'hôte. Un aspect qui n'a pas été développé dans notre étude, mais qu'il est important d'évoquer, est celui de la susceptibilité génétique de l'homme face à cette maladie. Il est important de noter que la grande majorité (90%) des individus infectés ne présentera jamais de symptomatologie clinique [Bloom et Small, 1998]. La grande variabilité de la réponse à l'infection par *Mycobacterium tuberculosis*, et en particulier le fait que la majorité des individus peut être

considérée comme résistant au développement d'une tuberculose clinique, est un argument très fort en faveur du rôle de facteurs génétiques de l'hôte dans cette maladie infectieuse [Casanova et Abel, 2002]. Le rôle des facteurs génétiques dans la prédisposition à la tuberculose a été suggéré par plusieurs études épidémiologiques, comme par exemple :

- la mise en évidence de fortes différences inter-ethniques, montrant en particulier une plus grande susceptibilité des populations d'origine africaine à la tuberculose comparé aux populations d'origine caucasienne [Bloom et Small, 1998],
- un taux de concordance pour la maladie plus grand chez les jumeaux monozygotes (~60%) que chez les dizygotes (~35%) [Comstock, 1978],
- un rôle majeur des facteurs génétiques de l'hôte dans la réponse aux infections mycobactériennes sur des modèles animaux [Lam-Yuk-Tseung et Gros, 2003].

La grande majorité des recherches génétiques réalisées jusqu'à présent sur la tuberculose humaine sont des études d'association avec certains gènes candidats : antigène HLA-DR2, antigène DQ1, gène NRAMP1, gène récepteur de la vitamine D..., les gènes impliquant pouvant varier en fonction de l'origine ethnique des patients [Remus *et al.*, 2005]. D'autre part, le récent article de Gutierrez *et al.* (2005) montre que le bacille tuberculeux est capable d'échanger des parties de son génome avec d'autres souches, un processus important dans les systèmes de co-évolution hôte-pathogène. Ces différents points pourraient expliquer la plus grande susceptibilité de certaines populations.

#### *Perspectives de recherche*

Étant donné l'implication des facteurs génétiques dans la susceptibilité à la tuberculose, il serait donc important d'intégrer la composante « ethnique » dans l'étude de la tuberculose en Guyane. Par exemple, il serait pertinent de réaliser une étude d'association gènes/maladie de type cas-témoins, avec des témoins familiaux, sur les différentes populations de Guyane.

### **3. Conclusion**

L'utilisation de *cluster-graphs* a permis d'organiser les données et de visualiser très concrètement les spoligotypes de *M. tuberculosis* effectivement en circulation en Guyane sur la période 1996-2003, ainsi que les liens génétiques effectifs entre souches, ce que ne permet pas la méthode ACM. La reconstruction de « réseaux d'allèles » est ici plus appropriée que les arbres construits sur les calculs de distance.

L'utilisation de l'indice TMI a également permis d'évaluer quantitativement le poids respectif de (1) la transmission récente et (2) des réactivations /

importations dans la persistance de la tuberculose en Guyane ces dernières années, conduisant à la forte incidence observée. Cette quantification est particulièrement importante pour savoir dans quelle direction renforcer les systèmes de détection de cas et/ou de prise en charge active de la tuberculose en Guyane.

Enfin, ce travail offre un éclairage nouveau sur la méthode récemment développée par Tanaka et Francis (2005) et qui, selon les recommandations des auteurs, nécessitent un approfondissement. L'utilisation des *cluster-graphs* semble particulièrement pertinente, et les résultats que nous avons obtenus par le calcul de TMI sont beaucoup plus cohérents avec l'analyse préliminaire des données génétiques que les résultats obtenus avec les RTI anciennement utilisés.



## Chapitre 6

# La dynamique de transmission de la tuberculose en Guyane

*Dans ce chapitre, il s'agit dans un premier temps de mettre en relation les agrégats spatiaux de tuberculose de l'Île de Cayenne et les grappes génétiques, afin de vérifier si les cas inclus dans les agrégats se superposent aux grappes génétiques, c'est-à-dire si les grappes génétiques correspondent ou non à une contamination de proximité (dans la même rue ou le même quartier). Ce type d'analyse liant données génétiques et analyse spatiale est très nouveau, et, concernant la tuberculose, n'a fait l'objet que d'une seule publication en 2004. L'objectif de ce chapitre est, finalement, de proposer une cartographie du risque tuberculeux sur l'Île de Cayenne, basée sur la construction d'un indice de vulnérabilité de la population et sur la répartition observée des cas de tuberculose recensés entre 1996 et 2003. La construction de l'indice de vulnérabilité se réfère à tous les précédents résultats énoncés dans ce mémoire concernant les facteurs explicatifs de la distribution observée des cas. La cartographie du risque finale devrait être un outil d'aide à la décision efficace pour la lutte contre la tuberculose en Guyane française.*

## **1. Mise en relation des agrégats spatiaux de cas et des grappes génétiques**

### **1.1. Introduction**

#### **1.1.1. Combiner SIG et épidémiologie moléculaire**

Jusqu'à présent, les études d'épidémiologie moléculaire se préoccupaient essentiellement d'identifier les facteurs de risque liés à l'existence de grappes génétiques de cas, même si quelques études ont cherché à évaluer si les grappes mises en évidence apparaissaient dans des zones géographiques définies [Tanser et Le Sueur, 2002 ; Yang *et al.*, 2001 ; Kistemann *et al.*, 2002 ;

Bishai *et al.*, 1998, Verver *et al.*, 2004], ou, à une échelle internationale, ont cherché à analyser la répartition géographique globale des souches [Duchêne *et al.*, 2004]. Pourtant, si l'on considère que les grappes génétiques sont le résultat d'une transmission active, l'identification des lieux précis où se fait cette transmission permettrait de cibler la recherche active des nouveaux cas de tuberculose et leur traitement.

Alors que ce travail de recherche combinant systèmes d'information géographique (SIG) et épidémiologie moléculaire était déjà engagé depuis deux ans, une équipe américaine a publié en octobre 2004 un article utilisant la même approche [Moonan *et al.*, 2004], réalisé sur les cas de tuberculose de Tarrant County, Texas, du 1<sup>er</sup> janvier 1993 au 31 décembre 2000 (N = 991 cas), et dont les résultats sont probants. Nous avons donc poursuivi notre travail de recherche sur la tuberculose dans la région de l'Île de Cayenne en Guyane en confortant nos hypothèses de travail sur la base de ce premier article.

### 1.1.2. Hypothèses de travail

Dans le chapitre 4, nous avons énoncé deux hypothèses ( $H_1$  et  $H_2$ ) pour expliquer l'agrégation spatiale de cas de tuberculose dans certaines zones :

- ( $H_0$ ) l'hypothèse nulle de distribution aléatoire des cas a déjà été testée au chapitre 4 et nous avons montré que cette hypothèse est ici rejetée
- ( $H_1$ ) une contamination à l'échelle locale, de proche en proche, qui aboutit à une grappe génétique de cas qui est aussi un agrégat spatial de cas, ou
- ( $H_2$ ) une contamination de proximité très faible (ou micro-épidémie) ; dans ce cas c'est un facteur de risque lié au lieu qui pourrait expliquer l'agrégation spatiale de cas.

Les deux hypothèses alternatives ( $H_1$  et  $H_2$ ) étant complémentaires, il suffit de vérifier l'une d'elles pour conclure. Ainsi, si la démonstration d'une contamination locale dans une même rue ou un même quartier ( $H_1$ ) s'avère impossible, cela voudra dire que la contamination de proximité est faible et que l'agrégation de cas est liée à l'existence locale d'un risque géographique ( $H_2$ ).

## 1.2. Méthode

### 1.2.1. Les données

Sur les 491 cas de tuberculose recensés en Guyane sur la période 1996-2003, seuls 342 cas confirmés ont pu être analysés par biologie moléculaire. La figure 6.1 récapitule le nombre de cas inclus dans les différents groupes étudiés.

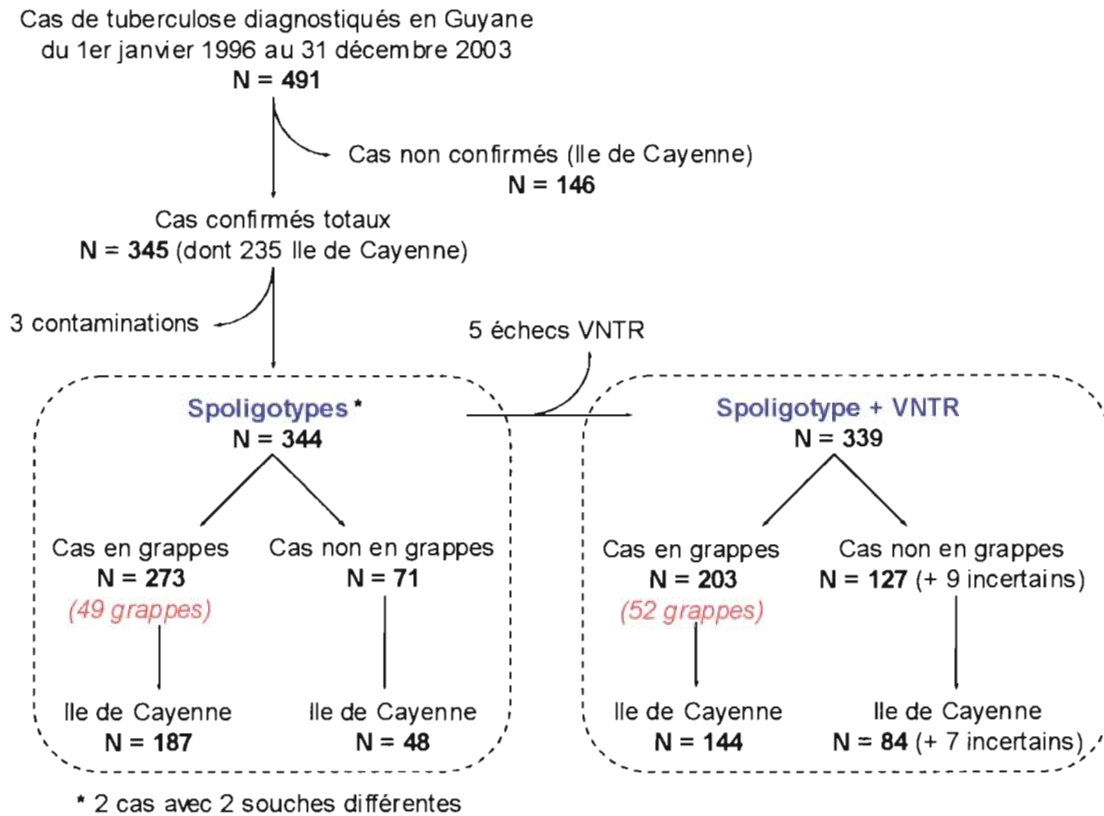


Figure 6.1 – Taille des populations étudiées (en nombre de cas N), en fonction des analyses moléculaires (spoligotypage et VNTR) et du lieu de résidence des cas. Sur les 345 cas confirmés de départ, trois cultures contaminées ont été éliminées de l'étude. Les analyses ont par ailleurs mis en évidence deux cas contaminés successivement par deux souches différentes, ce qui donne un total 344 souches au lieu de 342. Les "incertains" correspondent aux cas dont le typage par VNTR était incomplet.

### Grappes génétiques

Lorsque l'on s'intéresse aux grappes génétiques de cas, la population d'étude regroupe l'ensemble des cas confirmés de Guyane, soit 342 patients mais 344 spoligotypes (deux patients ont en effet déclaré une première puis une seconde tuberculose durant la période de l'étude, avec deux souches différentes pour chacune des infections).

### Agrégats géographiques

Les agrégats géographiques déterminés au chapitre 4 se limitent à l'Ile de Cayenne. L'étude de ces agrégats inclut donc uniquement les cas résidant sur l'Ile de Cayenne (au moment de la déclaration de la maladie), soit 381 cas. Quinze de ces cas n'ont pas pu être spatialisés (voir chapitre 4 §1.1.) donc le système d'information géographique mis en place sur l'Ile de Cayenne compte uniquement 366 cas de tuberculose (220 cas confirmés et 146 non confirmés).

## 1.2.2. Les analyses

Vérifier l'hypothèse 1, c'est-à-dire la possibilité d'une contamination de proximité, globalement dans une même rue ou un même quartier, voire la même commune, revient à démontrer que les cas inclus dans les agrégats géographiques identifiés sont tous porteurs de la même souche bactérienne (avec la méthode couplant spoligotype + VNTR), c'est-à-dire qu'il existe une superposition des grappes génétiques et des agrégats géographiques. C'est ce qui se passe dans le cas d'une épidémie causée par la même souche clonale [Maynard Smith *et al.*, 1993]. Pour démontrer cette corrélation, nous avons procédé à différentes analyses.

### 1.2.2.1. Transmission active dans les IRIS

#### *Pourcentage de cas en grappe par IRIS*

Utilisant le SIG mis en place sur l'île de Cayenne, nous avons dans un premier temps calculé le pourcentage de cas en grappes pour chacun des 35 IRIS. Pour cela nous avons, indépendamment pour chaque IRIS, divisé le nombre de cas inclus dans une grappe par le nombre de cas totaux confirmés de l'IRIS. Suivant la procédure utilisée par Moonan *et al.* (2004), les cas non confirmés présents dans les IRIS sont exclus de ces calculs.

Il est important de préciser que nous avons considéré comme « cas en grappe » les cas porteurs de souches que l'on a recensées au moins deux fois en Guyane sur la période d'étude, que ces cas aient été recensés ou non dans le même IRIS.

#### *Corrélation incidence et pourcentage de cas en grappe*

Le pourcentage de cas en grappes par IRIS ne fournit pas à lui seul une information pertinente sur le taux de transmission active de tuberculose. En effet, si tous les cas de l'IRIS sont inclus dans une grappe (c'est-à-dire qu'il existe au moins un autre patient porteur de cette même souche qui a été identifié en Guyane durant la période de l'étude), mais que tous ces cas sont porteurs de souches différentes, cela ne traduit en aucun cas un évènement de transmission active à l'intérieur même de l'IRIS ; la transmission a pu avoir lieu entre IRIS, ou sur une longue distance à travers le territoire, par déplacement d'individus infectés.

Dans leur étude, Moonan *et al.* (2004) ont testé, dans chaque unité géographique, la corrélation entre incidence de tuberculose et pourcentage de cas en grappes grâce au test de corrélation de Pearson. C'est ce que nous avons fait sur les 35 IRIS de l'île de Cayenne, en ne considérant donc que les cas en grappes présents sur l'île de Cayenne.

### **1.2.2.2. Transmission active dans les agrégats géographiques**

Les agrégats géographiques de cas que nous avons identifiés au chapitre 4 nous apportent une information supplémentaire par rapport à l'étude de Moonan *et al.* (2004). Dans cette étude, un agrégat géographique est défini par au moins 2 patients ayant la même souche qui vivent dans la zone d'étude. Dans notre cas, les agrégats géographiques ont été identifiés statistiquement à l'aide de SatScan. En effet, dans le cas de la Guyane où les IRIS sont inégaux en taille, et parfois très étendus, il est plus pertinent de travailler sur ces agrégats de cas reconnus significatifs, que sur une segmentation arbitraire de l'espace qui ne correspond pas forcément à la réalité du terrain.

Nous avons donc, dans un deuxième temps, calculé le pourcentage de cas en grappes dans les agrégats géographiques identifiés (1 agrégat principal, 2 agrégats secondaires significatifs, et 5 agrégats secondaires non significatifs). Mais, à la différence du calcul par IRIS, nous n'avons tenu compte que des cas inclus dans l'agrégat géographique lui-même. Par exemple, si l'on trouve un unique cas porteur du spoligotype 64 dans l'agrégat de cas, cette souche sera considérée dans notre analyse comme étant une souche « unique » même si en réalité, il en existe d'autres en Guyane, le but étant de définir si l'agrégation observée dans une zone particulière correspond à une transmission active et localisée dans cette zone.

## **1.3. Résultats**

### **1.3.1. Pourcentage de cas en grappe par IRIS**

Les résultats obtenus diffèrent en fonction de l'analyse moléculaire prise en compte (spoligotypage seul, ou spoligotypage + VNTR).

Avec le spoligotypage, on trouve 49 grappes incluant 273 cas, et 71 cas uniques, c'est-à-dire non inclus dans une grappe génétique (chapitre 5).

Avec l'analyse couplée des deux méthodes, les grappes sont plus nombreuses mais plus petites, c'est-à-dire qu'elles regroupent moins de cas, et 52 cas supplémentaires se retrouvent exclus des grappes (soit 123 cas uniques au total), ce qui se traduit sur la figure 6.2 – B par une diminution du pourcentage de cas en grappes par IRIS. Sur les 35 IRIS de l'Ile de Cayenne, 6 ont alors plus de 83,3% de cas en grappes, 20 ont plus de 66,7% de cas en grappes, et 25 en ont plus de 50%.

Le test de corrélation entre incidence de tuberculose par IRIS et pourcentage de cas en grappes par IRIS montre qu'il n'existe donc aucune corrélation entre ces deux variables ( $cor = 0,04$  ;  $p = 0,81$ ).

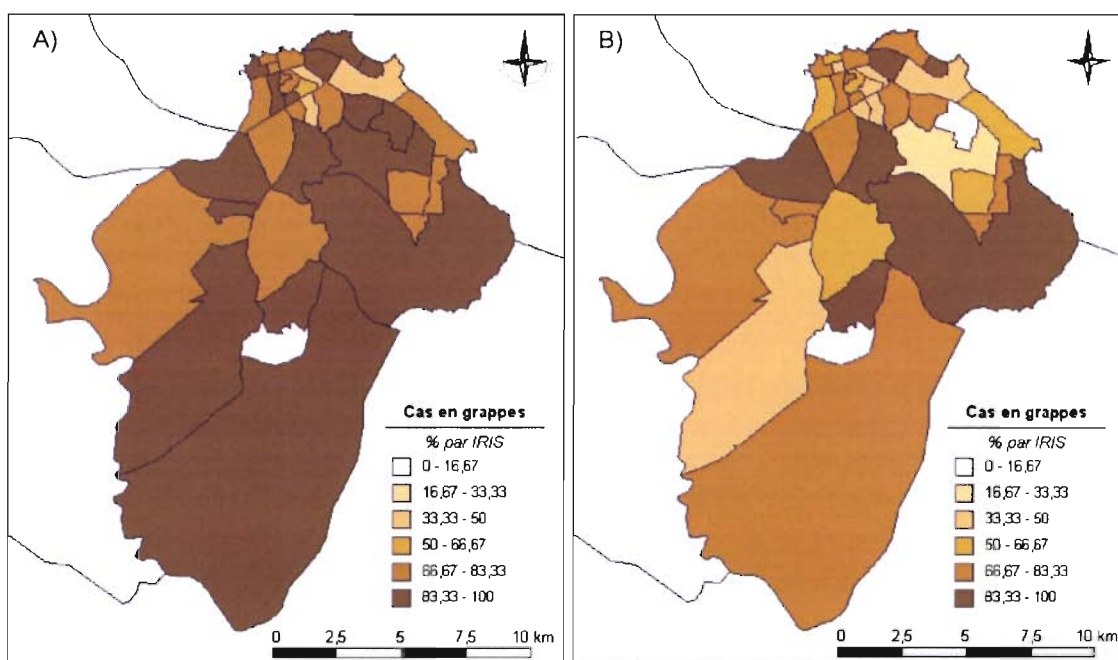


Figure 6.2 – Pourcentage de cas en grappe dans chaque IRIS de l'île de Cayenne (A) en tenant compte uniquement des spoligotypes et (B) de la combinaison spoligotype + VNTR.

### 1.3.2. Pourcentage de cas en grappe par agrégat

*L'agrégat n°1* (centre ville de Cayenne et quartiers proches) regroupe 190 cas de tuberculose, dont 78 sont des cas non confirmés. Sur les 112 autres cas, on recense 64 cas uniques, une grappe de 7 cas, une de 5 cas, une de 4 cas, trois grappes de 3 cas, et dix grappes de 2 cas (3 VNTR étant incomplets, on ne peut pas conclure pour 3 cas). Le pourcentage de cas en grappes est de 40,2%.

*L'agrégat n°2* (quartiers sud-est de Cayenne) regroupe 23 cas, dont 10 cas non confirmés. On trouve trois grappes de 2 cas, et 7 souches uniques, soit un pourcentage de cas en grappes de 46,2%.

*L'agrégat n°3* (vers Stoupan) ne regroupe que deux cas dans la même maison. Seul l'un des deux étant un cas confirmé, il n'est pas possible de conclure quant à la parenté des deux souches. À titre d'information, nous avons également analysé les souches dans les cinq agrégats secondaires non significatifs obtenus avec SaTScan (par exemple la prison de Rémire ; cf. chapitre 4) pour voir si l'on relevait des résultats intéressants. En fait, ces agrégats ne regroupent que des cas porteurs de souches uniques. Dans le cas de la prison, cela tendrait à montrer que la contagion est d'origine extérieure ou qu'elle est antérieure à l'emprisonnement.

## **1.4. Discussion**

### **1.4.1. Cas en grappe dans les agrégats géographiques**

L'analyse des souches inclues dans les trois agrégats significatifs de cas identifiés sur l'île de Cayenne n'a mis en évidence que très peu de grappes génétiques, ce qui est cohérent par rapport aux résultats obtenus sur l'échelle de transmission de la maladie. Dans l'agrégat principal, qui recense 190 cas de tuberculose, on ne recense que 40,2% de cas en grappe, la moitié étant des grappes de deux cas. La transmission concerne 2 à 3 cas, rarement plus.

### **1.4.2. Cas en grappe dans les IRIS**

Les résultats concernant le pourcentage de cas en grappes par IRIS vont également dans ce sens. Les cartes montrent bien l'existence d'IRIS pour lesquels le pourcentage de cas en grappe est élevé. Mais dans la mesure où il n'existe aucune corrélation avec l'incidence de tuberculose, ces cas en grappes ne sont pas représentatifs d'une forte transmission active localement.

### **1.4.3. Discussion des résultats**

Il existe différentes hypothèses pour expliquer ces résultats, notamment le fait que la transmission de la tuberculose nécessite généralement une forte proximité entre individus, et un contact prolongé [Rieder, 1999]. Il se pourrait également que le niveau de vie en Guyane soit suffisamment élevé pour limiter la transmission active de la tuberculose. Mais s'il existe, malgré le peu de contamination locale, une agrégation marquée de la tuberculose sur l'île de Cayenne, il doit exister un facteur de risque local pouvant expliquer cette agrégation, ce qui validerait donc notre hypothèse n°2 énoncée §1.1.2. Cette hypothèse étant validée, la construction d'une carte du risque tuberculeux sur l'île de Cayenne se trouve justifiée. Cette dernière étape est présentée ci-après.

## **1.5. Bilan**

Même s'il est difficile de définir de façon précise le mode de transmission de la tuberculose qui semble se faire de manière prépondérante à une échelle inter-communale, il semble que la contamination de proximité, à l'échelle de la maison ou du quartier, soit relativement faible. D'autre part, l'analyse des cas inclus dans les agrégats géographiques n'a pas permis de mettre en évidence beaucoup de cas ayant la même souche génétique de bactérie. Il semblerait

donc qu'un facteur géographique, lié au lieu de résidence des patients soit responsable d'une concentration de cas de tuberculose dans certaines zones de l'île de Cayenne, par exemple l'insalubrité de l'habitat (*cf.* chapitre 4). La confrontation de l'ensemble des résultats obtenus dans ce mémoire a été utilisée pour établir une cartographie du risque de tuberculose sur l'île de Cayenne, présentée ci-après.

## **2. Analyse et cartographie du risque tuberculeux en Guyane et sur l'île de Cayenne**

Ce travail s'appuie principalement sur de précédentes études concernant des maladies transmises par des insectes vecteurs, car il n'existe à l'heure actuelle aucune étude de cartographie du risque concernant une maladie à transmission directe, et liée à des facteurs socio-économiques, comme c'est le cas de la tuberculose.

Nous n'avons recensé que deux études cherchant à identifier les facteurs socio-économiques liés à la présence de tuberculose. Les résultats de notre travail de recherche, présentés tout au long de ce mémoire, nous ont permis d'identifier des facteurs de risque propres à la Guyane, et de construire une cartographie du risque intégrant ces spécificités.

### **2.1. Introduction**

#### **2.1.1. Définition du risque**

La littérature concernant la gestion des risques naturels définit le risque comme une fonction de l'aléa, ou danger, (le terme anglais correspondant est *hazard*), de la vulnérabilité, et de l'exposition [Crichton, 1999] (figure 6.3). Ces termes ont diverses significations en fonction des contextes et des disciplines.

#### *Vulnérabilité*

En sciences sociales, la vulnérabilité est définie comme l'ensemble des facteurs socio-économiques qui déterminent la capacité des individus à lutter contre un stress ou un changement [Allen, 2003]. En climatologie, par exemple, la définition de la vulnérabilité est très différente puisqu'elle est vue en termes de probabilité d'occurrence et d'impacts des événements liés au temps et au climat [Nicholls *et al.*, 1999].

### *Aléa (danger)*

Il définit à la fois l'évènement possible, qui peut être un processus naturel, technologique, social, ou économique, et la probabilité d'occurrence de ce phénomène en un point donné [Veyret, 2003].

### *Exposition*

Elle traduit simplement la présence simultanée de l'aléa et de la vulnérabilité. L'interaction aléa / vulnérabilité induit un risque, mais il n'y a pas forcément « catastrophe » (figure 6.3). Pour cela, il faut un évènement déclencheur supplémentaire. Dans le domaine de la santé, la catastrophe se mesure en termes de morbidité ou de mortalité.



Figure 6.3 – Illustration de la notion de risque et de ses composantes, traditionnellement utilisée en gestion des risques naturels. La flèche en pointillés indique que la survenue d'une catastrophe n'est pas systématique lorsqu'il y a risque, mais dépend de l'intervention d'un élément extérieur potentiel, qu'illustre le point d'interrogation. (Source des images : <http://www.irma-grenoble.com/04risques/index.htm>).

### **2.1.2. Application à l'épidémiologie**

Les très nombreux articles publiés concernant les risques en santé ne cartographient bien souvent pas le risque à proprement parler, mais plutôt les individus malades [Michael et Bundy, 1997 ; Wint *et al.*, 2002], ou les zones de présence potentielle d'un vecteur de la maladie étudiée, ce qui constitue donc une cartographie de l'aléa [ Buskirk et Ostfeld, 1998 ; Dale *et al.*, 1998 ; Michel *et al.*, 2001].

#### *Exemple de la trypanosomose*

L'étude de Michel *et al.* (2001) à propos de l'évaluation du risque trypanosomien pour le bétail au Burkina Faso illustre très bien les deux composantes du risque, aléa et vulnérabilité. Les chercheurs ont d'abord modélisé (1) la distribution de la densité de vecteurs, les glossines<sup>1</sup>, sur la base des paysages qui leur sont favorables, puis (2) la fréquentation de l'espace par les bovins. Les zones épidémiologiquement dangereuses, correspondant à la

<sup>1</sup> Pour rappel, en épidémiologie, un vecteur est un organisme qui transmet un agent infectieux. Les glossines, plus connues sous le nom de « mouches tsé-tsé » sont des insectes diptères hématophages qui transmettent les trypanosomoses animales et la maladie du sommeil.

rencontre hôte-vecteur, se situent donc à l'intersection de l'espace des glossines et de l'espace des bovins.

Une grille de risque a été établie pour rechercher les combinaisons les plus favorables à la transmission parasitaire, en intégrant un indicateur spatial des densités de glossines, et un indicateur spatial des densités de bovins (figure 6.4). Les deux indicateurs ont été croisés pour faire apparaître des tronçons de réseau hydrographique au niveau desquels les interfaces entre les glossines et les bovins sont les plus intenses. Ces zones sont supposées être épidémiologiquement les plus dangereuses.

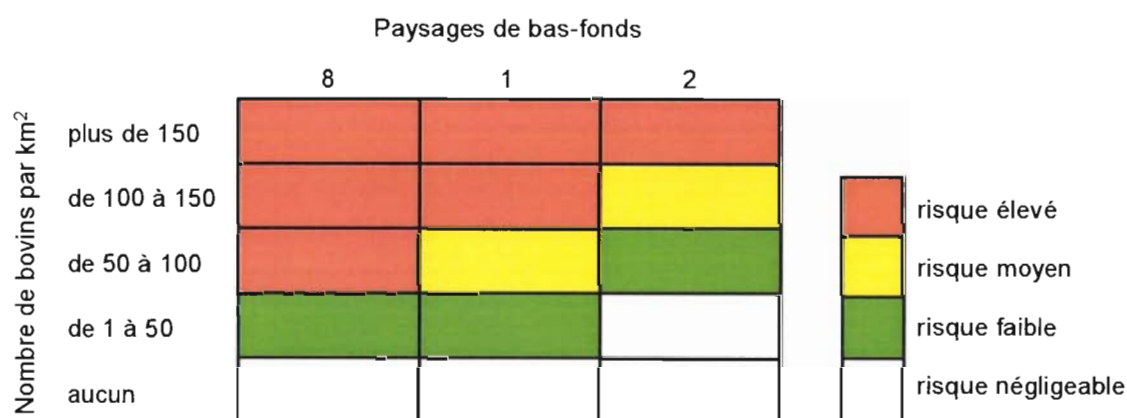


Figure 6.4 – Grille de décision pour l'identification des zones à risque épidémiologique pour la trypanosomose animale au Burkina Faso (d'après [Michel *et al.*, 2001]). À l'horizontale se trouve l'espace des bovins, défini par la « charge surfacique » en bétail ; à la verticale se trouve l'espace des glossines, lié à certaines classes de paysages de bas-fonds, auxquels se réfèrent les chiffres 8, 1 et 2, correspondant respectivement à un paysage de forêt, un paysage de galerie large bordée de savanes herbeuses, et un paysage de galerie étroite bordée de savanes arbustives ou arborées lâches. Ces trois paysages sont ceux pour lesquels les densités de glossines observées sont significatives.

### 2.1.3. Cas de la tuberculose humaine

#### *La vulnérabilité*

La tuberculose humaine est une « maladie sociale », donc lorsque l'on parle de vulnérabilité, il s'agit d'une vulnérabilité sociale, déterminée par des facteurs tels que la pauvreté, la surpopulation, la malnutrition, ou la qualité de l'habitat [Blaikie *et al.*, 1994]. Ces différentes conditions sont favorables à une baisse de l'immunité, et peuvent conduire au développement d'une tuberculose maladie en cas de contact avec le pathogène. La vulnérabilité sociale associée au risque tuberculeux est difficile à mesurer quantitativement, donc il est plus pertinent de chercher à définir un indice de vulnérabilité.

Un taux correspond à une valeur réelle, alors qu'un indice a une signification plus conceptuelle (échelle de valeur de  $x_1$  à  $x_i$ ). Un indice est construit par combinaison de plusieurs variables, représentatives du paramètre

que l'on cherche à quantifier. On utilise un indice lorsque la variable à étudier est trop complexe pour être mesurée directement ou lorsque l'on n'a pas accès à cette mesure. Par exemple, l'indice de végétation normalisé (IVN ou NDVI pour « *Normalized Difference Vegetation Index* ») est basé sur la combinaison de deux variables : les réflectances dans le spectre rouge et le proche infra-rouge mesurées par les satellites [Asrar *et al.*, 1989].

### *L'aléa*

Il est défini par la probabilité d'occurrence d'un cas de tuberculose dans l'environnement, en un point donné. Dans le cas présent, cette probabilité peut être évaluée par rapport à la distribution observée des cas recensés sur l'île de Cayenne entre janvier 1996 et décembre 2003.

### *Le risque tuberculeux*

Seule la présence simultanée de ces deux conditions crée le risque. S'il existe certaines zones géographiques où résident des populations pauvres et fragiles (vulnérabilité), mais que la probabilité de rencontrer un cas de tuberculose pouvant transmettre la maladie dans cette région (aléa) est nulle ou très faible, le risque tuberculeux est également nul ou très faible.

À contrario, si la probabilité d'occurrence d'un cas contagieux de tuberculose dans une population donnée est forte, mais que cette population n'est pas vulnérable, il y aura contact hôte-pathogène sans qu'il y ait, *à priori*, risque de tuberculose. Cependant, la bactérie responsable de la tuberculose humaine est capable de survivre à l'état latent dans l'organisme hôte et peut être activée plusieurs années plus tard si l'individu devient vulnérable. Contrairement au premier exemple cité, le risque associé à cette deuxième situation sera donc considéré comme non nul (voir §2.2.).

## **2.2. Cartographies de l'aléa, de la vulnérabilité et du risque tuberculeux sur l'île de Cayenne**

Dans un premier temps, nous décrivons ici brièvement comment le SIG a été mis en place, et les couches que nous y avons intégré. Nous présentons ensuite successivement les procédures que nous avons suivies pour établir les cartographies de l'aléa (§2.2.), de la vulnérabilité (§2.3.), et enfin du risque tuberculeux (§2.4.). Pour chacune des composantes du risque et le risque lui-même, nous expliquons les choix des différents facteurs pris en compte pour la cartographie, ainsi que les grilles de décision permettant de définir les classes d'aléa, de vulnérabilité et de risque. Les cartographies, toutes réalisées à l'échelle fine des quartiers, sont finalement présentées.

### 2.2.1. Le Système d'Information Géographique

En 2002, il existait déjà à l'IRD<sup>2</sup> de Cayenne un Système d'Information Géographique élaboré sous GeoConcept, intégrant un certain nombre de couches d'information sur l'île de Cayenne, telles que le réseau routier, le réseau hydrographique, les limites administratives, les limites des IRIS, ainsi que différents fonds de carte : la carte IGN au 1/25000<sup>e</sup> de l'île de Cayenne (n°4713) et la photographie aérienne de l'IGN prise en septembre 2001, auxquelles nous avons ajouté l'image SPOT 5 acquise le 29 août 2003, après l'avoir recalée. Il existait également une première typologie de l'habitat établie en 1998 par A. Gardel [Gardel, 2001]. À ces informations, nous avons ajouté :

- les cas de tuberculose recensés entre 1996 et 2003 (366 cas spatialisés sur les 381 cas totaux ; cf. chapitre 4 §1.1.). Des attributs sont associés à ces cas : âge, sexe, nationalité, année de déclaration de la maladie, inclusion ou non dans un cluster géographique de cas, etc.
- la typologie d'habitat de l'île de Cayenne réactualisée pour 2003
- les zones d'habitat insalubre de l'île de Cayenne (cf. chapitre 4 §2.5.2.)

### 2.2.2. Cartographie de l'aléa

#### 2.2.2.1. Méthode

##### *Paramètres pris en compte*

L'aléa est associé à la probabilité de présence de cas de tuberculose dans une zone donnée. Nous avons défini 3 classes d'aléa : fort, faible et nul. Cette classification tient compte (1) de la contagiosité des cas, et (2) de la présence de cas de tuberculose dans le quartier considéré.

Le marqueur de la contagiosité utilisé est la présence de bacilles acido-alcoolo-résistants (BAAR) à l'examen microscopique direct des produits d'expectoration du patient (cf. chapitre 1). L'aléa est donc proportionnel au nombre de cas contagieux (BAAR+) par rapport aux cas totaux de tuberculose dans le quartier, et au nombre de cas totaux de tuberculose.

La prise en compte des cas non contagieux dans la mesure de l'aléa est justifiée par le fait que, s'ils sont moins contagieux, les patients BAAR- seraient néanmoins capables, au moins dans une certaine mesure, de transmettre *Mycobacterium tuberculosis* [Behr *et al.*, 1999]. Une étude concernant la ville de San Francisco a ainsi montré que les patients BAAR- étaient responsables de plus de 21% des cas recensés dans cette ville sur la période 1991-1996 [Behr *et al.*, 1999].

---

<sup>2</sup> Institut de Recherche pour le Développement. Le Laboratoire Régional de Télédétection (LRT) de l'IRD de Cayenne a été le laboratoire d'accueil lors de toutes les missions de terrain effectuées en Guyane.

*Paramètres non pris en compte*

La différenciation entre cas confirmés et non-confirmés n'a pas été prise en compte, de même que la forme clinique de la maladie, car nous avons considéré que le danger (ou aléa) associé à la présence d'un cas non contagieux de tuberculose est le même, qu'il s'agisse d'une tuberculose pulmonaire ou extra-pulmonaire.

L'arbre décisionnel correspondant aux trois classes d'aléa que nous avons choisies est illustré par la figure 6.5.

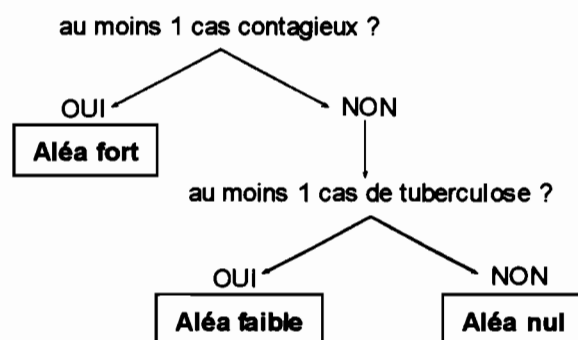


Figure 6.5 – Arbre de décision permettant de définir les trois classes d'aléa de tuberculose, en fonction de la présence de cas contagieux BAAR+, et de la présence de cas de tuberculose BAAR- mais potentiellement contagieux. L'aléa est fort si, dans le quartier considéré, il y a au moins un cas contagieux ; l'aléa est faible si il n'y a aucun cas de tuberculose détecté comme contagieux (BAAR+) mais qu'il y a au moins un cas de tuberculose recensé dans le quartier ; l'aléa est nul si aucun cas de tuberculose n'a été recensé entre le 1<sup>er</sup> janvier 1996 et le 31 décembre 2003 (probabilité de l'évènement "présence d'un cas de tuberculose" = 0).

La fonction de requête spatiale offerte par les SIG, dont le logiciel GeoConcept utilisé ici, permet de réaliser des calculs, des transformations ou des combinaisons diverses sur des groupements d'objets. Elle repose sur un certain nombre d'opérateurs dits relationnels comme l'intersection, l'union, la jointure [Bordin, 2002]. Par exemple, pour la classe « aléa faible », la requête est la suivante : colorier les objets : *urbanisé habité* vérifiant la condition : « *casTB* >= 1 ET au moins 1 cas contagieux = oui ».

Le type « urbanisé habité » de notre SIG, correspondant à la délimitation des quartiers, inclut un champ « cas TB quartier » auquel est associé le nombre de cas de tuberculose du quartier, et le champ « au moins 1 cas contagieux », champ rempli « oui » si au moins un des cas de tuberculose du quartier était BAAR+ à l'examen direct. Une couleur différente est associée à chacune des trois requêtes correspondant aux trois classes d'aléa définies (nul, faible, fort), et la cartographie de l'aléa est automatiquement générée en trois couleurs après exécution des requêtes.

### 2.2.2.2. Résultats

La figure 6.6 montre de nombreuses zones d'aléa nul. Elles correspondent aux quartiers où aucun cas de tuberculose n'a été recensé entre 1996 et 2003.

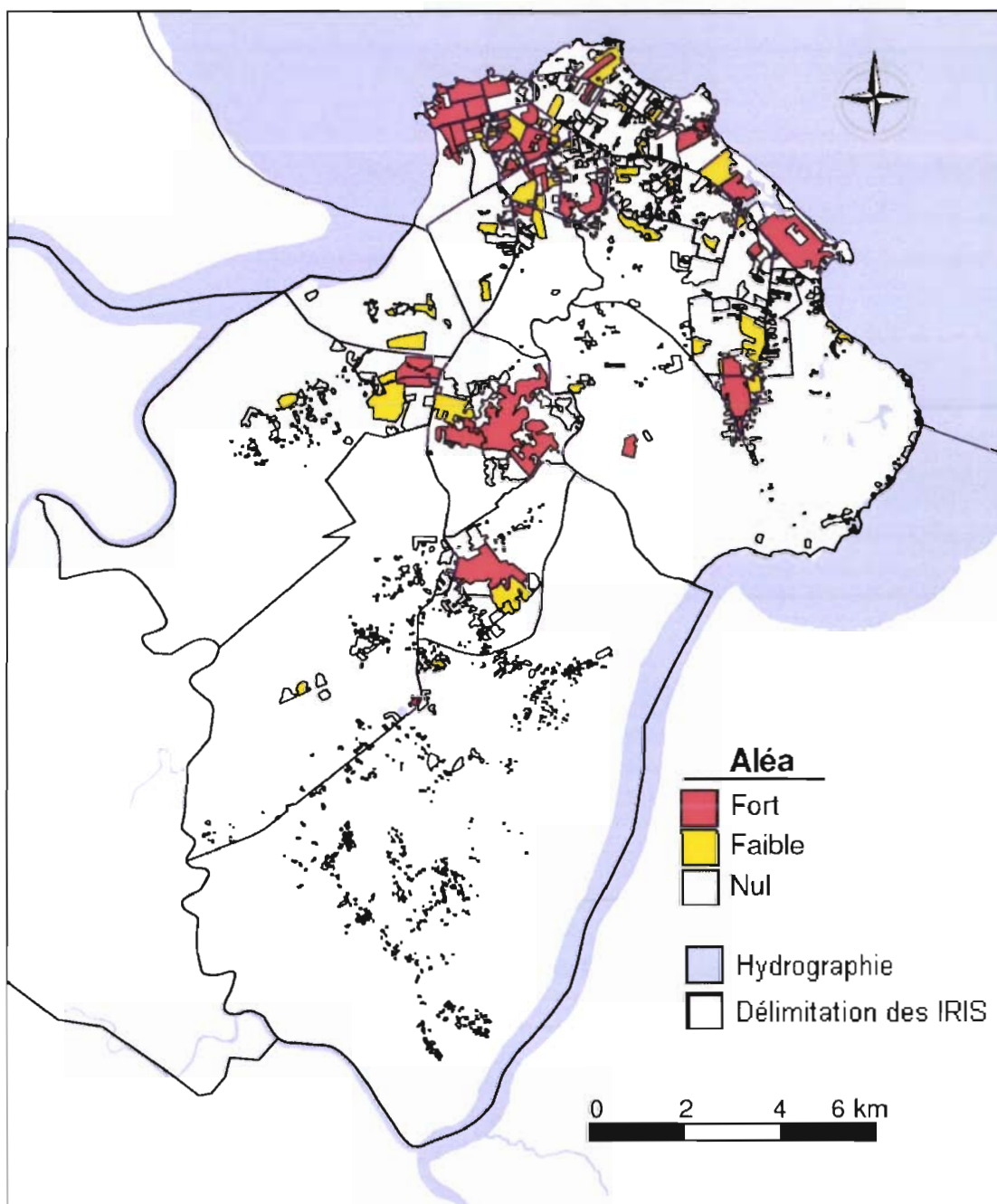


Figure 6.6 – Cartographie de l'aléa par quartier sur l'île de Cayenne (associé à la présence de cas de tuberculose contagieux, et à la présence de cas non contagieux dans le quartier). La classification est basée sur l'arbre de décision de la figure 6.5. appliqué aux cas de tuberculose recensés sur l'île de Cayenne entre 1996 et 2003 dont nous avons recensé l'adresse (366 cas spatialisés).

### 2.2.3. Cartographie de la vulnérabilité

#### 2.2.3.1. Méthode

##### *État de l'art*

Il existe dans la littérature peu d'informations concernant la « vulnérabilité sociale » et les données socio-économiques qui peuvent y être associées. Nous avons cependant relevé deux études dont le but était de rechercher les facteurs socio-économiques liés à la présence de tuberculose.

L'étude de Kistemann *et al.* (2002) a mis en évidence une corrélation entre quatre variables socio-économiques (sur douze étudiées), et le taux d'incidence de tuberculose à Cologne, en Allemagne : taux de chômage, total de personnes percevant une allocation, taux d'immigrants non allemands, et total d'immeubles recensés, ceci par unité spatiale.

L'article de Krieger *et al.* (2003) a analysé la relation potentielle entre la présence de cas de tuberculose et neuf variables socio-économiques. Parmi elles, seules deux sont significativement liées à des taux élevés de tuberculose : la surpopulation et une forte pauvreté.

##### *Application à la Guyane*

Nous avons noté deux points importants concernant ces deux études. Tout d'abord, ces variables sont pour la plupart des variables qui relèvent d'enquêtes socio-économiques du type de celles réalisées en France par l'INSEE. En Guyane, les statistiques sociales, économiques et sanitaires de l'Île de Cayenne fournies par l'INSEE ne sont disponibles qu'à l'échelle des IRIS<sup>3</sup>. Pour obtenir ces informations à l'échelle du quartier, il faut donc passer par des mesures indirectes. Dans une étude ultérieure, des données satellites SPOT ont par exemple permis de construire des indices de qualité de vie sur la base de la taille des maisons, de la densité de l'habitat, de la présence de voies de communications et de la proximité d'espaces verts [Weber et Hirsch, 1992].

D'autre part, les facteurs corrélés à la présence de tuberculose ne sont pas forcément les mêmes d'un pays à l'autre. Nous avons donc choisi des variables socio-économiques tenant compte de l'ensemble des résultats décrits dans la présente étude pour construire un indice de vulnérabilité pertinent à l'échelle de l'Île de Cayenne.

Nous avons précédemment montré que les principaux facteurs associés à la persistance de la tuberculose sur l'Île de Cayenne sont l'importation de cas et les réactivations. Il faut donc tenir compte des quartiers à forte concentration d'immigrants, et ceux où les conditions de vie précaire pourraient favoriser des réactivations. Nous avons également montré que la transmission locale de la

---

<sup>3</sup> En réalité, ces informations sont également disponibles à une échelle inférieure, celles des filots INSEE (cf. chapitre 4 §1.4.2.2.). Malheureusement, ces données n'ont pas été accessibles au moment de l'étude.

maladie est faible, et l'agrégation de cas de tuberculose dans certaines zones est liée à un facteur géographique local. Ce facteur « lieu » est lié pour une part au niveau de vie et à la qualité de vie des habitants de l'Île de Cayenne. C'est l'environnement d'un point de vue humain que nous cherchons ici à caractériser et à cartographier. Il faut donc trouver des paramètres environnementaux qui peuvent être des indicateurs du niveau de vie et de la qualité de vie des habitants.

### *Variables liées à la vulnérabilité de la population*

Dans le présent travail, nous avons choisi de caractériser le niveau de vie des habitants de l'Île de Cayenne sur des critères simples, à la fois qualitatifs et quantitatifs reflétant la qualité de vie des habitants :

- la typologie d'habitat élaborée précédemment (*cf.* chapitre 4 §1.3.2.), qui tient compte d'une part d'un critère quantitatif, la densité d'habitat (dense, diffus, ou dispersé) et d'autre part d'un critère qualitatif, les conditions socio-économiques (habitat spontané, résidentiel ou mixte), et

- l'insalubrité de l'habitat. Les zones d'habitat insalubre de l'Île de Cayenne (*cf.* définition des 24 quartiers insalubres, chapitre 4 tableau 4.7) ont été recensées par une étude de la DDE de Guyane [DDE-AU&S, 1997] qui précise les coordonnées géographiques de ces zones. L'insalubrité a été mesurée sur la base de la densité de population dans des zones impropres à l'urbanisation (habitat informel) et surtout, non équipées en VRD (voiries et réseaux divers) [Gardel, 2000].

Notre indice de vulnérabilité de la population pour la tuberculose a donc été construit sur la base de ces deux paramètres. La figure 6.7 récapitule le croisement des deux variables, aboutissant à cinq classes de vulnérabilité (de très forte à très faible). Le système de requêtes du SIG a permis d'établir la cartographie finale, en croisant les deux paramètres (figure 6.8).

Typologie d'habitat	Habitat insalubre	Habitat salubre	Échelle de vulnérabilité
spontané	[orange]	[orange]	
collectif		[orange]	
mixte	[orange]	[jaune]	
résidentiel dense	[jaune]	[vert clair]	
dispersé	[jaune]	[vert clair]	
résidentiel diffus	[jaune]	[vert foncé]	très faible

Figure 6.7 – Tableau récapitulatif de la vulnérabilité face à la tuberculose, associée à un quartier en fonction (1) du type d'habitat (spontané, mixte, résidentiel dense ou diffus, dispersé), et (2) de l'insalubrité de l'habitat, définie par une précédente étude de la DDE de Guyane [DDE-AU&S, 1997]. Les cinq classes de vulnérabilité de la population sont : très faible, faible, moyenne, forte, très forte.

### 2.2.3.2. Résultats

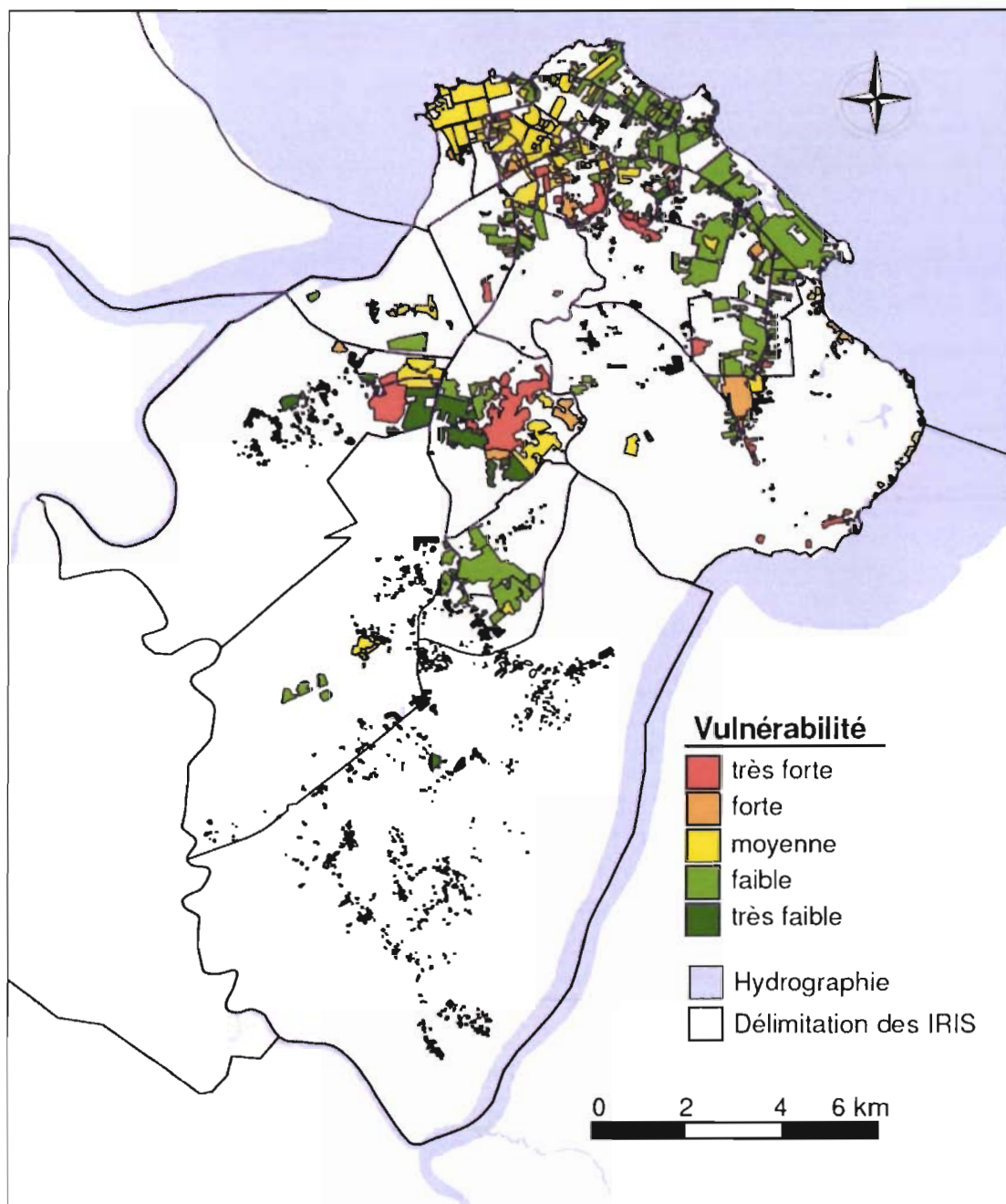


Figure 6.8 – Cartographie de la vulnérabilité de la population de l'île de Cayenne face à la tuberculose, par quartier. La classification est basée sur la grille de décision de la figure 6.7 tenant compte des deux paramètres typologie d'habitat et insalubrité de l'habitat. La cartographie a été réalisée grâce au système de requêtes du SIG GéoConcept.

Les quartiers regroupant une population vulnérable face à la tuberculose (classes moyenne à très forte) sont très localisés au niveau du centre ville de Cayenne et sa périphérie, de Balata et de Cogneau-Lamirande, ainsi que les quartiers sud de Rémire et la zone de Dégrad des Cannes.

## 2.2.4. Cartographie du risque tuberculeux

### 2.2.4.1. Méthode

Les couches d'information intervenant dans notre modèle du risque tuberculeux, c'est-à-dire l'aléa et la vulnérabilité, ont été croisées dans le SIG, faisant apparaître les zones au niveau desquelles les interactions entre une population vulnérable face à la tuberculose et des patients atteints de tuberculose qui sont susceptibles de transmettre la maladie (BAAR+ ou non) sont les plus intenses. Ces zones sont supposées être les plus à risque, ou encore les plus « dangereuses épidémiologiquement » pour la tuberculose humaine [Michel *et al.*, 2001].

La grille de décision que nous avons construite pour l'identification des zones à risque de tuberculose sur l'île de Cayenne (figure 6.9) ressemble à celle de la figure 6.4 de Michel *et al.* (2001). Les cinq classes de vulnérabilité de la population (de très faible à très forte) sont indiquées en lignes, et les trois classes d'aléa (de nul à fort) sont notées en colonnes. Au final, le croisement des deux couches aboutit à 7 classes de risque tuberculeux.

On utilise encore une fois le système de requêtes du SIG (sous GéoConcept) pour construire la carte de risque tuberculeux de l'île de Cayenne. Il est nécessaire de créer, au préalable, de nouveaux champs que l'on nomme « aléa » et « vulnérabilité » dans lesquels on inscrit pour chaque quartier la classe correspondant aux résultats obtenus par les deux analyses spatiales précédentes, afin de croiser les deux champs.

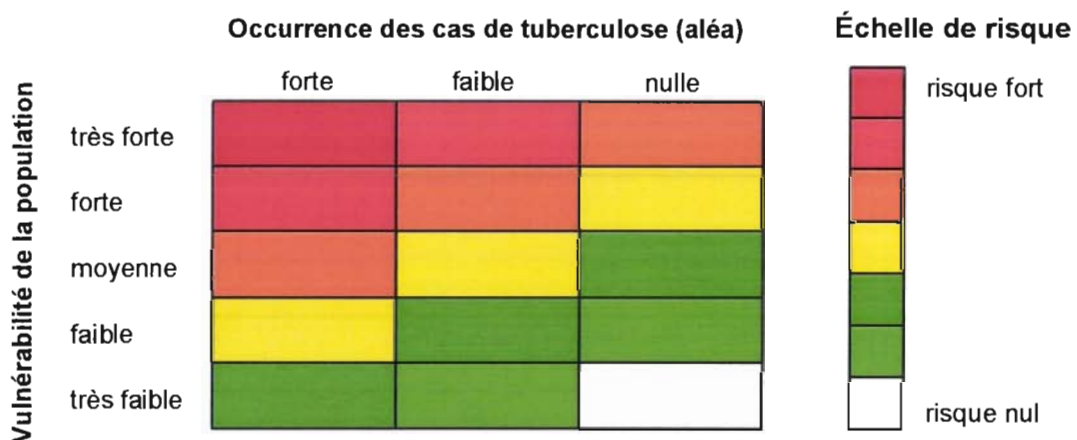


Figure 6.9 – Grille de décision pour l'identification des zones à risque de tuberculose sur l'île de Cayenne. En lignes, sont notées les cinq classes de vulnérabilité de la population (de très forte à très faible) définissant l'espace des hôtes. En colonnes, sont notées les trois classes d'aléa, correspondant à l'occurrence ou non de cas tuberculeux potentiellement contagieux, définissant l'espace des pathogènes. Le croisement de l'aléa et de la vulnérabilité aboutit à une échelle de risque tuberculeux comprenant sept classes de risque (de fort à nul).

### 2.2.4.2. Résultats

La figure 6.10 représente la cartographie du risque de tuberculose sur l'île de Cayenne réalisée sur la base de données épidémiologiques recensées avant 2004. La combinaison de l'aléa et de la vulnérabilité aboutit à un gradient de risque, avec la mise en évidence des quartiers les plus à risque de tuberculose.

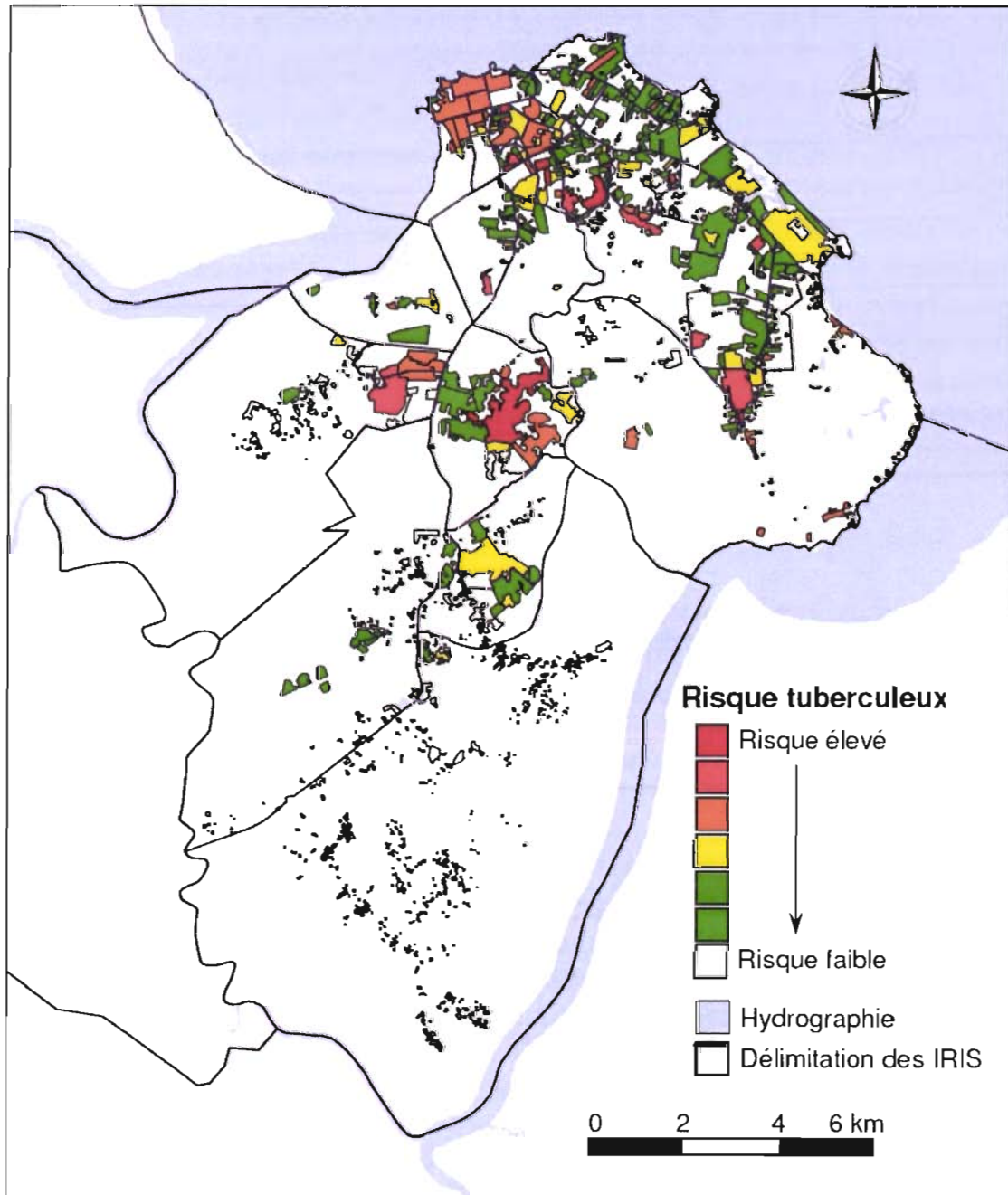


Figure 6.10 – Cartographie du risque de tuberculose par quartier sur l'île de Cayenne (associé à la vulnérabilité de la population et à l'aléa), la classification étant basée sur la grille de décision présentée sur la figure 6.9.

## **2.3. Discussion**

### **2.3.1. Un outil d'aide à la décision**

La cartographie du risque obtenue montre de façon précise les quartiers à risque de tuberculose, avec par exemple 15 quartiers inclus dans la classe de risque le plus important, et correspondant aux toponymies IGN suivantes : cité Eau Lisette, Thémire, cité Pascaline, Troubiran, Ploërmel, cité Suzini (village haïtien), prison de Rémire, Balata Ouest, Cogneau Est, et Dégrad des Cannes.

La mise en place de notre SIG pour le suivi géographique des cas de tuberculose, et l'identification des zones à risque de tuberculose doit permettre de cibler les actions de prévention et de lutte contre la tuberculose sur l'île de Cayenne, où se concentrent entre 65 et 75% de l'ensemble des cas de Guyane.

À noter également que, bien souvent, les médecins qui travaillent localement sont tout à fait conscients des mesures à engager pour améliorer la lutte contre la tuberculose, et des zones où mettre en œuvre les actions prioritaires, mais ils manquent généralement de moyens. La production d'une carte de risque pourrait avoir un impact important au niveau des organismes décisionnels et politiques chargés de prendre en charge la situation sanitaire, cette carte constituant un outil d'aide à la décision très concret.

### **2.3.2. Des indicateurs simples et accessibles**

Dans la mesure où nous avons précédemment montré que les informations de l'INSEE obtenues en Guyane sont biaisées par la difficulté d'accès à certains quartiers et à certaines populations, la cartographie du risque tuberculeux que nous avons créée est intéressante dans la mesure où elle intègre l'utilisation d'informations indirectes, dérivées de photographies aériennes ou d'images de télédétection. De plus, cela nous a permis de travailler à l'échelle beaucoup plus fine des quartiers, plutôt que celle des IRIS.

L'utilisation de ce type de données est d'autant plus pertinent que la Guyane dispose de stations de réception satellites permettant de mettre à jour très régulièrement les informations qui en sont dérivée ; il serait par exemple possible de mettre à jour chaque année la typologie de l'habitat que nous avons réalisée. Le SIG que nous avons mis en place peut donc constituer la base d'une surveillance épidémiologique future de la tuberculose sur l'île de Cayenne, à condition que les informations soient mises à jour régulièrement.

D'autre part, ce type d'outil n'est opérationnel qu'à la condition d'être basé sur des identificateurs fiables et simples. Nous avons essayé de limiter au maximum les couches d'information à combiner afin d'éviter l'obtention d'un résultat trop compliqué et difficile à interpréter, et toutes les informations que nous avons utilisées sont facilement accessibles.

### **2.3.3. Surveillance et retour d'information**

Un avantage supplémentaire est que l'utilisation de ce type d'outil facilite la mise à jour des données et peut permettre le suivi en temps réel de l'évolution de la tuberculose, d'où une potentielle amélioration de la surveillance épidémiologique, et des systèmes d'alerte. Il suffit pour cela d'intégrer au SIG les nouveaux cas malades, au fur et à mesure de leurs déclarations.

D'autre part, l'utilisation d'un système commun à l'ensemble des services de prise en charge de la tuberculose en Guyane permettrait une meilleure circulation de l'information, et un retour de l'information à chaque étape de l'analyse des cas (diagnostic clinique, diagnostic bactériologique, mise sous traitement, antibiogramme...), ce qui n'est actuellement pas le cas, du moins sur l'île de Cayenne.

### **2.3.4. Validation du modèle**

À ce niveau de l'étude, et même si la démarche proposée ici semble efficace, il nous manque une dernière étape qu'il aurait fallu ajouter : la validation de la cartographie du risque. Les perspectives de recherche relatives à ce point sont développées dans la conclusion.

## **3. Conclusion**

L'utilisation d'un système d'information géographique nous a permis de mettre en relation des informations spatiales, aboutissant à une cartographie précise du risque tuberculeux à l'échelle du quartier sur l'île de Cayenne. De cette analyse émerge des propositions d'intervention plus ciblées dans et autour des zones présentant les plus forts risques. L'originalité de ce travail est par ailleurs d'intégrer la notion de latence inhérente à la dynamique intrinsèque de la tuberculose : au-delà du risque lié au contact individu sain / individu malade qu'appréhendent les médecins, notre approche met en évidence l'importance de la dualité individu vulnérable / individu non vulnérable. Si améliorer les stratégies de contrôle de la tuberculose passe par l'augmentation de la détection des nouveaux cas [Blower *et al.*, 1996], l'identification des populations cibles est un premier pas vers ces nouvelles stratégies.



## Conclusion générale

Au terme de ce travail, il est important, dans un premier temps, de revenir sur les enjeux actuels du contrôle de la tuberculose dans le monde, puis, après avoir rappelé les objectifs généraux du travail de thèse, de présenter et de discuter les résultats obtenus, l'apport de notre approche méthodologique par rapport à des méthodes plus traditionnelles telles que les enquêtes épidémiologiques, et les limites du travail réalisé. Nous abordons ensuite une réflexion sur l'intégration potentielle de nos résultats dans un système de lutte contre la tuberculose en Guyane, et développons quelques propositions susceptibles de compléter les stratégies actuelles de contrôle de la maladie. Enfin, les perspectives de recherche de certaines approches développées au cours de la thèse sont présentées.

Avec une incidence qui augmente annuellement de 2%, la tuberculose est redevenue depuis une vingtaine d'années une priorité de santé publique au niveau de l'OMS. Bien qu'existante, la thérapie antituberculeuse s'avère lourde (traitement de 6 à 12 mois), et elle se limite à un petit nombre d'antibiotiques dont l'utilisation est remise en question par l'apparition de souches multi-résistantes à ces composés. De plus, l'efficacité de la vaccination contre la tuberculose, par l'utilisation du bacille Calmette et Guérin (BCG) depuis les années 50, est très variable, avec un taux de protection de 0 à 80% selon les études et selon les individus. De ce fait, le développement de nouvelles stratégies thérapeutiques et vaccinales contre ces bacilles reste une priorité de santé publique [O'Brien et Nunn, 2001]. Outre ce volet thérapeutique, des études ont également montré que l'on pouvait améliorer les stratégies de contrôle de la tuberculose en augmentant la détection des cas, et donc la proportion de cas traités, et en améliorant la prise en charge et le suivi individuel des patients, et donc leur compliance [Blower *et al.*, 1996].

Mais les stratégies actuelles de surveillance et de contrôle de la tuberculose, en particulier la stratégie DOTS préconisée par l'OMS, ne ciblent que les patients contagieux, c'est-à-dire ayant un frottis positif à l'examen direct [WHO, 2005], et ne prennent pas en compte la multiplicité des facteurs associés à la propagation de la maladie, ni ne tiennent compte des spécificités locales. Tant que l'on ne dispose pas de nouvelles solutions thérapeutiques, le moyen d'améliorer le contrôle de la maladie réside dans la prévention des populations à risque, ce qui passe par la caractérisation des facteurs de risque liés à la

tuberculose, ainsi que par la mise en place d'un système de détection précoce de ces populations à risque.

Le travail de recherche que nous avons réalisé s'inscrit donc dans l'optique de proposer des solutions nouvelles, de préférence simples à mettre en œuvre et peu coûteuses, afin d'améliorer le contrôle de la tuberculose en Guyane. Les résultats obtenus sont un premier pas vers une compréhension plus fine des mécanismes qui déterminent la répartition spatiale de la tuberculose sur l'île de Cayenne, et favorisent la persistance de la maladie, malgré les systèmes de prise en charge de la maladie existant. Les analyses mises en œuvre permettent de répondre aux grandes questions énoncées dans l'introduction, en particulier quelles sont les dynamiques de transmission de la tuberculose d'un point de vue spatial, et quelle est l'influence de l'environnement social et culturel sur l'évolution de cette maladie. Les résultats du travail de recherche sont également intéressants dans la mesure où ils apportent une plus-value par rapport aux études classiques d'épidémiologie clinique et médicale, qu'ils ne prétendent par ailleurs pas remplacer.

Nous avons d'abord démontré que la dynamique de la tuberculose sur l'île de Cayenne suit un mode endémique plutôt qu'épidémique. La transmission active de la maladie est faible, puisque seuls 15% des cas environ relèveraient d'une transmission récente de la maladie, ce résultat étant en adéquation avec le faible nombre de grappes génétiques identifiées. La persistance de la maladie serait en fait due à un système de micro-épidémies à double entrée, avec la présence de souches endémiques dues à la réactivation ou à la transmission locale de souches historiquement présentes en Guyane, et l'introduction de souches non-endémiques du fait de l'importation de cas par l'immigration en provenance de pays à forte incidence de tuberculose. Ces deux systèmes expliquent la forte diversité de souches de *Mycobacterium tuberculosis* observée en Guyane.

L'existence d'une agrégation spatiale de cas non aléatoire dans certaines zones de l'île de Cayenne a également été démontrée. Nous avons énoncé deux hypothèses pour expliquer cette agrégation : soit l'existence d'une forte contamination de proximité, avec l'existence de petits foyers épidémiques locaux, soit l'existence d'un facteur de risque lié à la localisation géographique, avec par exemple une superposition de la distribution des agrégats de tuberculose avec celle des habitats défavorisés de l'île de Cayenne. Notre analyse des souches en circulation à l'intérieur de ces agrégats spatiaux a montré qu'il existe en fait peu de patients porteurs de la même souche génétique dans ces agrégats. L'insalubrité de l'habitat, si elle n'est pas corrélée à la présence d'agrégats de cas, semble en revanche bien être un facteur lié à la présence de tuberculose.

Enfin, nous avons montré qu'il était possible, en croisant des informations spatiales et environnementales (typologie d'habitat et insalubrité), d'obtenir une cartographie des populations vulnérables de l'Île de Cayenne face à la tuberculose. En liant cette cartographie de la vulnérabilité avec des données épidémiologiques sur l'occurrence précise des cas depuis 8 ans, nous avons pu établir une cartographie du risque tuberculeux sur l'Île de Cayenne, directement exploitable pour l'aide à la décision dans la lutte contre la tuberculose en Guyane.

Il existe cependant des limites à notre étude. En premier lieu, il aurait fallu disposer d'un recensement exhaustif de toutes les données de tuberculose, confirmée et non confirmée, pour l'ensemble de la Guyane. Ceci aurait permis de mieux appréhender les processus en oeuvre au niveau régional.

D'autre part, dans la partie épidémiologie moléculaire, nous n'avons pas utilisé la combinaison d'outils moléculaires la plus pertinente pour l'analyse phylogénique des souches, les analyses MIRUs étant incomplètes au moment de la thèse. Cependant, dans la mesure où les analyses MIRUs ne sont pas encore utilisées de façon universelle, le choix de cette technique aurait limité les possibilités de comparaison dans la base de données internationale.

L'analyse des facteurs pouvant être liés à la vulnérabilité de la population a également été limitée par le manque de données cartographiées sur l'Île de Cayenne. Ainsi, il n'existe aucun SIG indiquant les réseaux aériens et souterrains d'alimentation électrique, les réseaux d'eau potable et d'eaux usées, ou les logements disposant de sanitaires par exemple. L'inclusion de ce type de données aurait peut-être permis d'affiner la cartographie du risque tuberculeux sur l'Île de Cayenne.

Enfin, il faut noter que, pour valider la cartographie du risque tuberculeux obtenue, il aurait fallu vérifier si les nouveaux cas de tuberculose recensés depuis le 1<sup>er</sup> janvier 2004 résident effectivement dans des zones que nous avons identifiées comme étant à haut risque de tuberculose. Malheureusement, nous avons manqué de temps pour acquérir ces nouvelles données. De plus, le géoréférencement des cas par GPS aurait nécessité une mission de terrain supplémentaire en Guyane, ce que la durée de l'étude ne permettait pas. Cette vérification mériterait d'être faite dans le futur.

Outre ces limites, les résultats obtenus nous permettent malgré tout d'émettre des propositions pour une action plus ciblée de la lutte contre la tuberculose en Guyane. La propagation et la persistance de la tuberculose en Guyane sont contrôlées par différents facteurs, qui relèvent d'interactions complexes impliquant la génétique, l'écologie, la démographie, le contexte socio-politique, ainsi que les comportements socio-culturels de la population. Les spécificités locales de l'Île de Cayenne que nous avons pu mettre en évidence au cours de notre recherche permettent de proposer des solutions visant à

renforcer le contrôle de la maladie dans cette région, et dans une certaine mesure, sur l'ensemble de la Guyane.

1. Au vu de la forte mobilité de la population guyanaise, la mise en place d'un carnet de suivi individuel, écrit en français et dans la langue maternelle du patient, permettrait d'assurer une continuité des soins en cas de déménagement, ou de retour dans le pays d'origine pour les patients étrangers.
2. L'analyse par la méthode de capture-recapture a montré que la proportion d'individus malades recensés par les organismes de lutte contre la tuberculose est en fait beaucoup plus forte que ce que laisse croire la proportion d'individus déclarés. Cependant, la notification obligatoire des cas de tuberculose diagnostiqués constitue un outil important dans le choix des priorités de santé au niveau politique et doit donc être améliorée.
3. Il serait intéressant de mettre en place un comité de coordination inter-institutions, comme préconisé par l'OMS, à l'échelle régionale de la Guyane, afin de centraliser l'information de l'ensemble des services de lutte et de prise en charge de la tuberculose en Guyane, et de favoriser le retour d'informations, et la réactivité des organismes de prévention.
4. La cartographie du risque nous semble un outil important sur lequel appuyer notre démarche dans le futur. Le caractère visuel, et donc concret d'une carte, peut s'avérer utile pour appuyer les demandes de crédits nécessaires à la mise en place de programmes de dépistage actifs de cas par exemple. De plus, la cartographie que nous avons élaborée peut être améliorée au cours du temps, car les données qui y sont intégrées telles que la typologie de l'habitat et les zones d'habitat insalubre peuvent être actualisées, de même que la cartographie au lieu de résidence des nouveaux patients depuis le 1<sup>er</sup> janvier 2004.
5. Ayant montré que l'importation de cas en provenance de pays à forte incidence de tuberculose est un acteur majeur dans la persistance de la tuberculose, il serait peut-être intéressant de mettre en place un dépistage systématique par radiographie pulmonaire de l'ensemble des étrangers entrant sur le territoire. L'évaluation préalable du coût d'une telle opération doit cependant être faite.
6. Enfin, l'intégration d'assistantes sociales au sein des acteurs de la lutte anti-tuberculeuse en Guyane semble également indispensable afin de prendre en charge certains groupes particulièrement vulnérables, et vivant dans des contextes sociaux difficiles.

Cette étude a montré que l'intégration des facteurs environnementaux dans l'étude de la dynamique des maladies ne doit pas se limiter aux maladies vectorielles. Elle a également montré l'intérêt de la combinaison de l'analyse spatiale et de l'épidémiologie moléculaire pour la compréhension des dynamiques de transmission de maladies infectieuses, approche qui a jusqu'à présent été peu envisagée [Moonan *et al.*, 2004]. Des travaux similaires pourraient donc être entrepris sur la tuberculose dans d'autres pays, ou sur d'autres maladies infectieuses, afin d'intégrer l'aspect spatial lorsqu'il a été peu ou pas pris en compte.

D'autre part, sachant qu'une analyse épidémiologique de trois ans a été engagée sur la région Antilles-Guyane en 2003, les résultats de cette étude devraient permettre d'affiner l'analyse du risque en y intégrant d'autres facteurs dont nous n'avons pas tenu compte, et d'étendre nos résultats concernant l'île de Cayenne à l'ensemble de la Guyane.

Enfin, la mise en place d'équipes pluridisciplinaires réunissant médecins, écologues, modélisateurs, télédéTECTEURS et géographes sur des problématiques de recherche en santé devrait permettre d'ouvrir de nouvelles perspectives de recherche pour des maladies présentes en Guyane mais dont on explique encore très mal l'existence et la persistance. C'est par exemple le cas de la fièvre Q ou de l'ulcère de Buruli.

Ce travail de recherche, original de part son approche spatiale et les outils qu'il combine (SIG et épidémiologie moléculaire), a ainsi permis de mettre en évidence certains aspects de la dynamique spatiale de la tuberculose en Guyane qui n'avaient pu jusqu'à présent être mis en évidence, ou du moins formalisés. Et même si des limites demeurent, nous avons montré l'intérêt de développer de telles méthodes appliquées à la surveillance des maladies et à l'implémentation de systèmes d'alerte opérationnels.



# Bibliographie

- Alland, D., Kalkut, G.E., Moss, A.R., McAdam, R.A., Hahn, J.A., Bosworth, W., *et al.* (1994). Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *New England Journal of Medicine* 330 : 1710-1716.
- Allen, K. (2003). Vulnerability reduction and the community-based approach. Dans Pelling, M. (Ed.), *Natural disasters and development in a globalising world*. Routledge, London, pp. 170-184.
- Aluoch, J.A., Oyoo, D., Swai, O.B., Kwamanga, D., Agwanda, R., Edwards, E.A. (1987). A study of the use of maternity and child welfare clinics in case-finding for pulmonary tuberculosis in Kenya. *Tubercle* 68 : 93-103.
- Amat-Rose, J.M. et Gentillini, M. (1995). La santé en carte. Dans Morlin, E. (Dir.), *Penser la terre, stratèges et citoyens, le réveil des géographes*. Autrement, série Mutations, Paris, pp. 119-137.
- Anderson, R.M. (1998). Epidemiology. Dans Cox, F.E.G. (Ed.), *Modern parasitology - A textbook of parasitology. 2<sup>nd</sup> Edition*. Blackwell Scientific Publications, Oxford, Boston, pp. 75-116.
- Anselin, L., Syabri, I., Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis* 38(1) : 5-22.
- Applied Maths BVBA (2000). Bionumerics. The integral study of biological relationships. Manual version 2.5. Kortrijk, Belgium, 262 pages.

- Asrar, G., Myneni, R.B., Li, Y., Kanemasu, E.T. (1989). Measuring and modeling spectral characteristics of a tallgrass prairie. *Remote Sensing of Environment* 27 : 143-155.
- Avruskin, G.A., Jacquez, G.M., Meliker, J.R., Slotnick, M.J., Kaufmann A., Nriagu, J.O. (2004). Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *International Journal of Health Geographics* 3 : 26.
- Barret, J. (Dir.) (2001). *Atlas illustré de la Guyane, 2<sup>de</sup> édition*. IRD éditions, France, Limoges, 219 pages.
- Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., *et al.* (1994). Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *American Journal of Tropical Medicine and Hygiene* 51(3) : 271-280.
- Béguin, M. et Pumain, D. (1994). La représentation des données géographiques : statistique et cartographie. Armand Colin, Paris, 192 pages.
- Behr, M.A., Warren, S.A., Salamon, H., Hopewell, P.C., Ponce de Leon, A., Daley C.L., Small, P.M. (1999). Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 353 : 444-449.
- Benjamin, W.H. Jr., Lok, K.H., Harris, R., Brook, N., Bond, L., Mulcahy, D., *et al.* (2001). Identification of a contaminating *Mycobacterium tuberculosis* strain with a transposition of an IS6110 insertion element resulting in an altered spoligotype. *Journal of Clinical Microbiology* 39(3) : 1092-1096.
- Bernillon, P., Lievre, L., Pillonel, J., Laporte, A., Costagliola, D., Groupe d'épidémiologie clinique des CISIH. (1997). Estimation de la sous-déclaration des cas de Sida en France par la méthode de capture-recapture. *Bulletin Épidémiologique Hebdomadaire* 97 : 19-21.

- Besag, J.E. et Newell, J.N. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A* 154 : 143-155.
- Bishai, W.R., Graham, N.M., Harrington, S., Pope, D.S., Hooper, N., Astemborski, J., et al. (1998). Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *Journal of the American Medical Association* 280 : 1679-1684.
- Bithell, J.F. (1999). Disease mapping using the relative risk function estimated from areal data. Dans Lawson, A.B. et al. (Eds.), *Disease Mapping and Risk Assessment for Public Health*. Wiley, London, 482 pages.
- Blaikie, P., Cannon, T., Davis, I., Wisner, B. (1994). *At risk: Natural hazards, people vulnerability, and disasters*. Routledge, London, 464 pages.
- Bloom, B.R. et Small, P.M. (1998). The evolving relation between humans and *Mycobacterium tuberculosis*. *New England Journal of Medicine* 338 : 677-678.
- Blower, S.M., McLean, A.R., Porco, T.C., Small, P.M., Hopewell, P.C., Sanchez, M.A., Moss, A.R. (1995). The intrinsic transmission dynamics of tuberculosis epidemics. *Nature Medicine* 1(8) : 815-821.
- Blower, S.M., Small, P.M., Hopewell, P.C. (1996). Control strategies for tuberculosis epidemics: new models for old problems. *Science* 273 : 497-500.
- Blower, S.M. et Chou, T. (2004). Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance. *Nature Medicine* 10(10) : 1111-1116.
- Bonn, F. et Rochon, G. (1992). *Précis de télédétection. Volume 1 : principes et méthodes*. Presses de l'Université du Québec, Sillery, 485 pages.
- Bordin, P. (2002). *SIG: concepts, outils et données*. Hermes Science Publications, Paris, 259 pages.

- Bourgarel, S. (1994). *Santé et géographie en Guyane*. L'Harmattan, Paris, 239 pages.
- Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., *et al.* (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences of the United States of America* 99(6) : 3684-3689.
- Brudey, K., Gordon, M., Moström, P., Svensson, L., Jonsson, B., Sola, C., *et al.* (2004). Molecular Epidemiology of *Mycobacterium tuberculosis* in Western Sweden. *Journal of Clinical Microbiology* 42 : 3046-3051.
- Burgos, M.V. et Pym, A.S. (2002). Molecular epidemiology of tuberculosis. *European Respiratory Journal* 20 (Suppl. 36) : 54s–65s.
- Buskirk, J. V. et Ostfeld, R.S. (1998). Habitat heterogeneity, dispersal, and local risk of exposure to Lyme disease. *Ecological Applications* 8(2) : 365-378.
- Cadot, E., Barbazan, P., Boussinesq, M. (1998). Les déterminants géographiques de la transmission de l'onchocercose en zone de transition forêt-savane : l'exemple de deux villages du foyer de Mbam (région du Centre, Cameroun). *Cahiers Santé* 8(6) : 429-435.
- Casanova, J.L. et Abel, L. (2002). Genetic dissection of immunity to mycobacteria : the human model. *Annual Review of Immunology* 20 : 581-620.
- Centers for Disease Control and Prevention (CDC) (1999). *Tuberculosis elimination revisited: obstacles, opportunities, and a renewed commitment*. Advisory Council for the Elimination of Tuberculosis (ACET). MMWR. 48 (RR-9):[inclusive page numbers].
- Centers for Disease Control and Prevention (CDC) (2005). New CDC Program for rapid genotyping of *Mycobacterium tuberculosis* isolates. *Journal of the American Medical Association* 293 : 2086.

- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43 : 783-791.
- Chao, A. (1998). Dans Armitage, P. et Colton, T. (Eds.), *Encyclopedia of Biostatistics, Volume 1*. Wiley, New York, pp. 482-486.
- Chao, A., et Tsay, P.K. (1998). A sample coverage approach to multiple-system estimation with application to census undercount. *Journal of the American Statistical Association* 93 : 283-293.
- Chao, A., Tsay, P.K., Lin, S-H., Shau, W-Y., Chao, D-Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine* 20 : 3123-3157.
- Chapman, D.G. (1954). The estimation of wildlife populations. *Annals of Mathematical Statistics* 25 : 1-15.
- Chapman, D.H. et Junge, C.O. (1956). The estimation of the size of a stratified animal population. *Annals of Mathematical Statistics* 27 : 375-389.
- Charrier, R. (2002). Guyane, des peuples et des histoires. *Antiane-Éco* 54 : 14-17.
- Che, D. et Bitar, D. (2005). Les cas de tuberculose déclarés en France en 2003. *Bulletin Épidémiologique Hebdomadaire* n°17-18 : 66-69.
- Chen, S., Zeng, S., Xie, C. (2000). Remote sensing and GIS for urban growth analysis in China. *Photogrammetric Engineering and Remote Sensing* 66 : 593-598.
- Christensen, R. (1997). *Log-linear models and logistic regression*, 2<sup>nd</sup> edition. Springer, New York.
- Cliff, A.D. et Ord, J.K. (1973). *Spatial autocorrelation*. Pion, London, 178 pages.

- Cline, B. (1970). New eyes for the epidemiologists: aerial photography and other remote sensing techniques. *Journal of Epidemiology* 92(2) : 85-89.
- Colonna, M. (1997). Estimations bayésiennes empiriques de risques relatifs : principes et exemples d'application. *Revue d'Épidémiologie et de Santé Publique* 45(2) : 142-149.
- Comstock, G.W. (1978). Tuberculosis in twins: a re-analysis of the Proffit survey. *American Review of Respiratory Disease* 117 : 621-624.
- Conseil National du Sida (CNS) (1996). Spécificités et inégalités. Le sida dans les départements français d'Amérique. Rapport suivi de recommandations. [http://www.cns.sante.fr/htm/avis/rapports\\_pdf/96\\_06\\_17\\_ethique\\_recherche.pdf](http://www.cns.sante.fr/htm/avis/rapports_pdf/96_06_17_ethique_recherche.pdf)
- Conseil Supérieur d'Hygiène Publique de France (CSHPF) (2003). Prévention et prise en charge de la tuberculose en France. Synthèse et recommandations du groupe de travail du Conseil Supérieur d'Hygiène Publique de France (2002-2003). *Revue des Maladies Respiratoires* 20(6) : 7S13-7S106.
- Cormack, R.M. (1968). The statistics of capture-recapture methods. *Oceanography and Marine Biology Annual Review* 6 : 455-506.
- Coull, B.A. et Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics* 55(1) : 294-301.
- Cracknell, A.P. et Hayes, L.W.B. (1991). *Introduction to Remote Sensing*. Taylor & Francis, London, 278 pages.
- Craig, P.S., Giraudoux, P., Shi, D., Bartholomot, B., Barnish, G., Delattre, P., et al. (2000). An epidemiological and ecological study of human alveolar echinococcosis transmission in south Gansu, China. *Acta tropica* 77 : 167-177.

- Crawford, J.T., Braden, C.R., Schable, B.A., Onorato, I.M. (2002). National Tuberculosis Genotyping and Surveillance Network: design and methods. *Emerging Infectious Diseases* 8(11) : 1192-1196.
- Crichton D. (1999) *Natural Disaster Management*. Tudor Rose, London, pp 102-103.
- Curran, P.J., Atkinson, P.M., Foody, G.M. et Milton, E.J. (2000). Linking remote sensing, land cover and disease. *Advances in Parasitology* 47 : 37-80.
- Cuzick, J. et Edwards, R. (1990). Spatial clustering for inhomogenous populations. *Journal of the Royal Statistical Society* 52 : 73-104.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393 : 537-544.
- Dale, P.E.R., Ritchie, S.A., Territo, B.M., Morris, C.D., Muhar, A., Kay, B.H. (1998). An overview of remote sensing and GIS for surveillance of mosquito vector habitats and risk assessment. *Journal of Vector Ecology* 23(1) : 54-61.
- Darbyshire, J.H. (1995). Tuberculosis: old reasons for a new increase? *British Medical Journal* 310(6985) : 954-955 .
- Darlu, P. et Tassy, P. (1993). *Reconstruction phylogénétique. Concepts et méthodes*. Coll. Biologie théorique n°7. Masson, Paris, 245 pages.
- Darroch, J.N. (1958). The multiple-recapture census I: estimation of a closed population. *Biometrika* 64 : 435-447.
- Dautzenberg, B. et Bouvet, E. (1996). Tuberculose et infections à mycobactéries atypiques (SIDA exclu). Dans Aubier, M., Fournier, M., Pariente, R., *Pneumologie*. Médecine-Sciences, Flammarion, 38 : 349-363.

- DDE-AU&S (1997). *Les territoires de l'exclusion. Descriptif des zones d'habitat insalubre de l'île de Cayenne*. DDE Guyane.
- Decludt, B. et Vaillant, V. (1997). Les cas de tuberculose déclarés en France en 1995. *Bulletin Épidémiologique Hebdomadaire* n° spécial : 16-19.
- Dagnelie, P. (1975). *Théorie et méthodes statistiques : applications agronomiques. Volume 2 : les méthodes de l'inférence statistique*. Les Presses Agronomiques de Gembloux, Belgique, Gembloux, 463 pages.
- De La Rocque, S., Michel, V., Plazanet, D., Pin, R. (2004). Remote sensing and epidemiology: examples of applications for two vector-borne diseases. *Comparative Immunology, Microbiology and Infectious Diseases* 27(5) : 331-341.
- Deparis, X., Pascal, B., Baudon, D. (1997). Evaluation of the completeness of the epidemiological surveillance system for malaria by the capture-recapture system in the French armies in 1994. *Tropical Medicine and International Health* 2 : 433-439.
- De Savigny, D. et Wijeyaratne, P. (1995). GIS for health and the environment. International Development Research Centre, Ottawa, 172 pages.
- Devallois, A., Horgen, L., Sola, C., Rastogi, N. (1998). Le typage moléculaire des mycobactéries. *Pathologie Biologie* 46(8) : 625-636.
- Diggle, P.J. et Chetwynd, A.G. (1991). Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* 47 : 1155-1163.
- Diman Antenor, D. (1992). Approche des milieux Sociaux. *Antiane-Éco* 17 : 6-8.
- Ding, Y. et Feinberg, S.E. (1996). Multiple sample estimation of population and census undercount in the presence of matching errors. *Survey Methodology* 22 : 55-64.

- Dossou-Yovo, J., Doannio, J.M., Diarrassouba, S., Chauvancy, G. (1998). Impact d'aménagements de rizières sur la transmission du paludisme dans la ville de Bouaké, Côte d'Ivoire. *Bulletin de la Société de Pathologie Exotique* 91 : 327-33.
- Duchêne, V., Ferdinand, S., Filliol, I., Guegan, J-F., Rastogi, N., Sola, C. (2004). Phylogenetic reconstruction of *Mycobacterium tuberculosis* within four settings of the Carribean region: tree comparative analysis and first appraisal on their phylogeography. *Infection, Genetics and Evolution* 4 : 5-14.
- Dye, C., Scheele, S., Dolin, P., Pathania, V., Raviglione, M.C. for the WHO Global Surveillance and Monitoring Project (1999). Global burden of tuberculosis: Estimated incidence, prevalence, and mortality by country. *Journal of the American Medical Association* 282(7) : 677-686.
- EDATER (2001). *La représentation cartographique. Guide méthodologique*. CERTU - GIP ATEN, Lyon, Montpellier, 87 pages.
- Enright, A.J. et Ouzounis, C.A. (2001). BioLayout – an automatic graph layout algorithm for similarity visualization. *Bioinformatics* 17(9) : 853-854.
- Epstein, P.R. (1999) Climate and Health. *Science* 285(5426) : 347-348.
- Farmer, P. et Kim, J.Y. (1998). Community based approaches to the control of multidrug resistant tuberculosis: introducing "DOTS-plus". *British Medical Journal* 317 : 671-674.
- Faure, J.F., Tran, A., Gardel, A., Polidori, L. (2003). Élaboration d'un indice de densité de population et analyse de sa distribution spatiale à Belém (Brésil) et Cayenne (Guyane française). *Bulletin de la Société Française de Photogrammétrie et Télédétection* 173/174 : 135-144.
- Ferdinand, S. (2004). *Contribution du génotypage à l'étude de l'épidémiologie et de la génétique évolutive de Mycobacterium tuberculosis en milieux insulaires*. Thèse de doctorat, Université Paris VII, spécialité microbiologie.

- Fienberg, S.E. (1972). The multiple-recapture census for closed populations and incomplete 2k contingency tables. *Biometrika* 59 : 591-603.
- Filliol, I., Sola, C., Legrand, E., Rastogi, N. (1999). Diversité génotypique de *Mycobacterium tuberculosis* dans la région Antilles-Guyane. *Bulletin de la Société de Pathologie Exotique* 92(5) : 292-298.
- Filliol, I., Ferdinand, S., Negroni, L., Sola, C., Rastogi, N. (2000). Molecular typing of *Mycobacterium tuberculosis* based on variable number of tandem DNA repeats used alone and in association with spoligotyping. *Journal of Clinical Microbiology* 38(7) : 2520-2524.
- Filliol, I., Driscoll, J.R., Van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valetudie, G., *et al.* (2002). Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerging Infectious Diseases* 8(11) : 1347-1349.
- Filliol, I., Driscoll, J.R., van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valétudie, G., *et al.* (2003). Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *Journal of Clinical Microbiology* 41(5) : 1963-1970.
- Forman, R.T.T. et Godron, M. (1986). Landscape ecology. John Wiley & sons, New York, 619 pages.
- Froment, A. (1997). Une approche écoanthropologique de la santé publique. *Nature Sciences Sociétés* 5 : 5-11.
- Froment, A. (2002). Maladies et évolution humaine : une approche anthropobiologique. Dans Susanne, Ch., Chiarelli, B. et Rebato, E. (Dir.), *Sur les Traces de la Biologie Humaine. Les Fondements de l'Anthropologie Biologique*. De Boeck, Bruxelles, pp. 579-584.

- Frothingham, R. et Meeker-O'Connell, W.A. (1998). Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* 144 : 1189-1196.
- Gardel, A. (2000). *Téledétection et enjeux d'aménagement en contexte urbain tropical : application à l'île de Cayenne, Guyane*. Thèse de doctorat, Université du Littoral, Côte d'Opale.
- Gardel, A. (2001). Les paysages urbains de l'île de Cayenne, Guyane Française. *Mappemonde* 63 : 16-21.
- Geiter, L. (Ed.) (2000). *Ending neglect: the elimination of tuberculosis in the United States*. National Academy Press, Washington DC, 292 pages.
- GeoConcept (2000). *Manuel d'utilisation, version 5.0*. GeoConcept SA.
- Gjini, A., Stuart, J.M., George, R.C., Nichols, T. Heyderman, R.S. (2004). Capture-recapture analysis and pneumococcal meningitis estimates in England. *Emerging Infectious Diseases* 10(1) : 87-93.
- Glynn, J.R., Vynnycky, E., Fine, P.E. (1999). Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. *American Journal of Epidemiology* 149(4) : 366-371.
- Goetz, S.J., Prince, S.D., Small, J. (2000). Advances in satellite remote sensing of environmental variables for epidemiological applications. Dans Hay, S., Rogers, D. et Randolph, S. (Eds.), *Remote Sensing and GIS in Public Health. Advances in Parasitology* 47, Academic Press, New York, pp. 289-307.
- Goldovsky, L., Cases, I., Enright, A.J., Ouzounis, C.A. (2005). BioLayoutJava: versatile network visualisation of structural and functional relationships. *Applied Bioinformatics* 4(1) : 71-74.

- Green, C., Hoppa, R.D., Young, T.K., Blanchard, J.F. (2003). Geographic analysis of diabetes prevalence in an urban area. *Social Science & Medicine* 57(3) : 551-60.
- Groenen, P.M., Bunschoten, A.E., van Soolingen, D., van Embden, J.D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*: application for strain differentiation by a novel typing method. *Molecular Microbiology* 10 : 1057-1065.
- Guernier, V., Hochberg, M.E., Guegan, J.-F. (2004) Ecology drives the worldwide distribution of human diseases. *PLoS Biology* 2(6) : e141. DOI : 10.1371/journal.pbio.0020141.
- Gutacker, M.M., Smoot, J.C., Migliaccio, C.A., Ricklefs, S.M., Hua, S., Cousins, D.V., *et al.* (2002). Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 162 : 1533-1543.
- Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., *et al.* (2005). Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. *Plos Pathogens* 1(1) : e5. DOI : 10.1371/journal.ppat.0010005.
- Hanski, I. et Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature* 404 : 755-758.
- Hay, S.I. (2000). An overview of remote sensing and geodesy for epidemiology and public health application. Dans Hay, S., Rogers, D. et Randolph, S. (Eds.), *Remote Sensing and GIS in Public Health. Advances in Parasitology* 47, Academic Press, New York, pp. 1-35.
- Hay, S.I., Tucker, C.J., Rogers, D.J., Packer, M.J. (1996). Remotely sensed surrogates of meteorological data for the study of the distribution and

- abundance of arthropod vectors of disease. *Annals of Tropical Medicine and Parasitology* 90 : 1-19.
- Hay, S.I., Snow, R.W., Rogers, D.J. (1998). Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 92(1) : 12-20.
- Hellenbrand, W., Breuer, T., Petersen, L. (2001). Changing epidemiology of Q fever in Germany, 1947-1999. *Emerging Infectious Diseases* 7(5) : 789-796.
- Hill, J.P. et Ramachandran, G. (1992). A simple scheme to improve compliance in patients taking tuberculosis medication. *Tropical Doctor* 22(4) : 161-163.
- Hook, E.B. et Regal, R.R. (1995a). Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews* 17 : 243-264.
- Hook, E.B. et Regal, R.R. (1995b). Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. *American Journal of Epidemiology* 142 : 48-52.
- Hopewell, P.C. (2002). Tuberculosis control: how the world has changed since 1990. *Bulletin of the World Health Organization* 80(6) : 427-427.
- Hubert, B., Decludt, B., Vaillant, V. (1995) Analyse critique de la situation épidémiologique de la tuberculose en France et propositions. *Médecine et Maladies Infectieuses* 25 : 291-298.
- Hunter, P. (1990). Reproducibility and indices of discriminatory power of microbial typing methods. *Journal of Clinical Microbiology* 28 : 1903-1905.
- Hunter, P.R. et Gaston, M.A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology* 26 : 2465-2466.

- Infuso, A., Hubert, B., Etienne, J. (1998). Underreporting of legionnaires' disease in France: the case for more active surveillance. *Euro Surveill* 3 : 48-50.
- Iñigo, J., Arce, A., Martin-Moreno, J.M., Herruzo, R., Palenque, E., Chaves, F. (2003). Recent transmission of tuberculosis in Madrid: application of capture–recapture analysis to conventional and molecular epidemiology. *International Journal of Epidemiology* 32 : 763-769.
- Institut National de la Statistique et des Études Économiques (INSEE) (1991). La part de l'habitat précaire. *Antiane-Éco* 14 : 30-31.
- Institut National de la Statistique et des Études Économiques (INSEE) (1999). [http://www.recensement.insee.fr/RP99/rp99/page\\_accueil.paccueil](http://www.recensement.insee.fr/RP99/rp99/page_accueil.paccueil)  
Recensement 1999 de la population.
- International Working Group for Disease Monitoring and Forecasting (IWGDMF) (1995). Capture-recapture and multiple-record systems estimation II: applications in human diseases. *American Journal of Epidemiology* 142 : 1059-1068.
- Jones, T. et Schaffner, W. (2001). Miniature chest radiograph screening for tuberculosis in jails: a cost-effectiveness analysis. *American journal of respiratory and critical care medicine* 164 : 77-81.
- Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S. et al. (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology* 35 : 907-914.
- Kanduma, E., McHugh, T.D., Gillespie, S.H. (2003). Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *Journal of Applied Microbiology* 94 : 781-791.

- Khan, K.M., Muennig, P., Behta, M., Zivin, J.G. (2002). Global drug-resistance patterns and the management of latent tuberculosis infection in immigrants to the United States. *New England Journal of Medicine* 347 : 1850-1859.
- Kistemann, T., Munzinger, A., Dangendorf, F. (2002). Spatial patterns of tuberculosis incidence in Cologne (Germany). *Social Science & Medicine* 55(1) : 7-19.
- Klassen, A.C., Kulldorff, M., Curriero, F. (2005). Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. *International Journal of Health Geographics* 4 : 1.
- Kochi, A. (1991). The global tuberculosis situation and the new control strategy of the World Health Organization (editorial). *Tubercle* 72 : 1-6.
- Kremer, K., van Soolingen, D., Frothingham, R., Haas, W.H., Hermans, P.W., Martin, C., *et al.* (1999). Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: Interlaboratory study of discriminatory power and reproducibility. *Journal of Clinical Microbiology* 37 : 2607-2618.
- Krivinka, R., Drapela, J., Kubik, A., Dankova, D. (1974). Epidemiological and clinical study of tuberculosis in the district of Kolin, Czechoslovakia. *Bulletin of the World Health Organization* 51: 59-69.
- Kulaga, S., Behr, M.A., Schwartzman, K. (1990). Genetic fingerprinting in the study of tuberculosis transmission. *Canadian Medical Association Journal* 161(9) : 1165-1169.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26 :1481-1496.
- Kulldorff, M. (1999). An isotonic spatial scan statistic for geographical disease surveillance. *Journal of the National Institute of Public Health* 48(2) : 94-101.

- Kulldorff, M. et Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in Medicine* 14 :799-810.
- Kulldorff, M., and Information Management Services, Inc. (2005). SaTScan™ version 6.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>
- Kulldorff, M., Athas, W.F., Feuer, E.J., Miller, B.A., Key, C.R. (1998). Evaluating cluster alarms: a space-time statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health* 88(9) :1377-1380.
- Lam-Yuk-Tseung, S. et Gros, P. (2003). Genetic control of susceptibility to bacterial infections in mouse models. *Cellular Microbiology* 5(5) : 299-313.
- Lambregts-van Weezenbeek, C.S. et Veen, J. (1995). Control of drug-resistant tuberculosis. *Tubercle and Lung Disease* 76 : 455-459.
- Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J-F., Bertollini, R. (1999). *Disease Mapping and Risk Assessment for Public Health*, John Wiley & Sons, Toronto, 482 pages.
- Legay, J-M. et Debouzie, D. (1985) *Introduction à une Biologie des Populations*. Masson, Paris, 149 pages.
- Léglise, I. (2004). Langues frontalières et langues d'immigration en Guyane Française. *Glottopol* 4 : 108-124 (revue accessible en ligne : <http://www.univ-rouen.fr/dyalang/glottopol>).
- Léopold Aït-Ouada, V. (2002). *La tuberculose en Guyane*. Thèse de médecine, Montpellier I, 159 pages.
- Levin, R. (1970). Extinctions. Dans Gerstenhaber, M. (Ed.), *Some mathematical questions in biology*. American Mathematical Society, Providence, Rhode Island, pp. 77-107.

- Lhuillier, N. (2001). En Guyane, ça déménage. *Antiane-Éco* 48 : 23-24.
- Linthicum, K.J., Anyamba, A., Tucker, C.J., Kelley, P.W., Myers, M.F., Peters, C.J. (1999). Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science* 285 (5426) : 397-400.
- Lo, C.P. (1995). Automated population dwelling unit estimation from high-resolution satellite images: a GIS approach. *International Journal of Remote Sensing* 16(1) :17-34.
- Mac Arthur, R.H. et Wilson, E.O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton, 203 pages.
- Magurran, A.E. (1988). *Ecological diversity and its measurement*. Chapman & Hall, London, 179 pages.
- Manly, B.F.J. (1991). Randomization and Monte-Carlo methods in biology. Second edition. Chapman et Hall, London, 281 pages.
- Martuzzi, M. et Hills, M. (1995). Estimating the degree of heterogeneity between event rates using likelihood. *American Journal of Epidemiology* 1995; 141: 369-74.
- Mathematical Sciences Education Board, National Research Council (1993). *Measuring what counts: a policy brief*. National Science Teachers Association, Washington DC.
- MathSoft (1997). *S-PLUS 4 User's Guide*. Data Analysis Products Division, MathSoft Inc., Seattle, WA.
- Mayer, J.D. (1983). The role of spatial analysis and geographic data in the detection of disease causation. *Social Science & Medicine* 17(16) : 1213-1221.

- Maynard Smith, J., Smith, N.H., O'Rourke, M., Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* 90 : 4384-4388.
- McConkey, S.J., Williams, M., Weiss, D., Adams, H., Cave, M.D., Yang, Z., *et al.* (2002). Prospective Molecular Typing for TB. *Clinical Infectious Diseases* 34 : 612-619.
- McMichael, A.J. (2001). Human culture, ecological change, and infectious disease: are we experiencing history's fourth great transition? *Ecosystem Health* 7(2) : 107-115.
- Médecins Sans Frontières (MSF) (2004). *Running Out of Breath: TB Care in the 21st Century. Campaign for access to essential medicines*. Geneva. Disponible sur <http://www.msf.org/source/annual/tbday/2004/TBDay2004report.doc>
- Michael, E. et Bundy, D.A.P. (1997). Global mapping of lymphatic filariasis. *Parasitology Today* 13(12) : 472-476.
- Michel, J.-F., Dray, S., de La Rocque, S., Desquesnes, M., Solano, P., De Wispelaere, G., Cuisance, D. (2002). Modelling bovine trypanosomosis spatial distribution by GIS in an agro-pastoral zone of Burkina Faso. *Preventive Veterinary Medicine* 56: 5-18.
- Moonan, P.K., Bayona, M., Quitugua, T.N., Oppong, J., Dunbar, D., Jost, K.C. Jr. *et al.* (2004). Using GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographics* 3(1) : 23.
- Moran, P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika* 37 : 17-23.
- Moström, P., Gordon, M., Sola, C., Ridell, M., Rastogi, N. (2002). Methods used in the molecular epidemiology of tuberculosis. *Clinical Microbiology and Infection* 8 : 694-704.

- Mullis, K.B., Faloona, F., Scharf, S., Saiki, R.K., Horn, G.T., Erlich, H.A. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51 : 263-273.
- Munch, Z., Van Lill, S.W., Booyesen, C.N., Zietsman, H.L., Enarson, D.A., Beyers, N. (2003). Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *International Journal of Tuberculosis and Lung Disease* 7(3) : 271-277.
- Murray, M. (2002). Sampling bias in the molecular epidemiology of tuberculosis. *Emerging Infectious Diseases* 8(4) : 363-369.
- Murray, C.J.L. et Salomon, J.A. (1998). Expanding the WHO tuberculosis control strategy: rethinking the role of active case-finding. *The International Journal of Tuberculosis and Lung Disease* 2 (Supp. 1) : S9-S15(1).
- Nanan, D.J. et White, F. (1997). Capture-recapture: reconnaissance of a demographic technique in Epidemiology. *Chronic Diseases in Canada* 18 : 144-148.
- Nardone, A. et Decludt, B. (2000). Évaluation épidémiologique du système de surveillance de la légionellose en France en 1998. Rapport de l'Institut de Veille Sanitaire, 45 pages.
- Neville, K., Bromberg, A., Bromberg, R., Bank, S., Hanna, B.A., Ross, W.N. (1994). The third epidemic: multi-drug resistant tuberculosis. *Chest* 105 : 45-48.
- Nicholls, R.J., Hoozemans, F.M.J., Marchand, M. (1999). Increasing flood risk and wetland losses due to global sea-level rise: regional and global analyses. *Global Environmental Change* 9 : S69-S87.

- Nichols, J.D. et Pollock, K.H. (1983). Estimation methodology in contemporary small mammal capture-recapture studies. *Journal of Mammalogy* 64 : 253-260.
- Niemann, S., Richter, E., Rusch-Gerdes, S. (1999). Stability of *Mycobacterium tuberculosis* IS6110 restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. *Journal of Clinical Microbiology* 37 : 409-12.
- O'Brien, R.J. et Nunn, P.P. (2001). The need for new drugs against tuberculosis: obstacles, opportunities, and next steps. *American Journal of Respiratory and Critical Care Medicine* 163 :1055-1058.
- Odoi, A., Martin, S.W., Michel, P., Middleton, D., Holt, J., Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International Journal of Health Geographics* 3 : 11.
- Omran, A.R. (1977). Epidemiologic transition in the US: the health factor in population change. *Population Bulletin* 32 : 2-41.
- Organisation Mondiale de la Santé (OMS) (2002). *Rapport sur la santé en Europe 2002*. OMS, Publications régionales. Série européenne n° 97.
- Perez, A.M., Ward, M.P., Torres, P., Ritacco, V. (2002). Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. *Preventive Veterinary Medicine* 56(1) : 63-74.
- Pfyffer, G.E., Strassle, A., Rose, N., Wirth, R., Brandli, O., Shang, H. (1998). Transmission of tuberculosis in the metropolitan area of Zurich: a 3 year survey based on DNA fingerprinting. *European Respiratory Journal* 11(4) : 804-808.
- Picard, B. (2000). Molecular epidemiology of large bacterial endemics in sub-Saharan Africa. *Bulletin de la Société de Pathologie Exotique* 93 : 219-223.

- Pillaye, J. et Clarke, A. (2003). An evaluation of completeness of tuberculosis notification in the United Kingdom. *BMC Public Health* 3 : 31.
- Rastogi, N., Schlegel, L., Pfaff, F., Jeanne, I., Magnien, C., Lajoinie, G., et al. (1998). La tuberculose en région Antilles-Guyane. Situation épidémiologique de 1994-1996. *Bulletin Épidémiologique Hebdomadaire* 11 : 45-47.
- Raviglione, M.C., Rieder, H.L., Styblo, K., Khomenko, A.G., Esteves, K., Kochi, A. (1994). Tuberculosis trends in Eastern-Europe and the former Ussr. *Tubercle and Lung Disease* 75(6) : 400-416.
- Raviglione, M.C., Snider, D.E. Jr., Kochi, A. (1995). Global epidemiology of tuberculosis - Morbidity and mortality of a worldwide epidemic. *Journal of the American Medical Association* 273(3) : 220-226.
- Reeves, W.C., Hardy, J.L., Reisen, W.K., Milby, M.M. (1994). The potential effect of global warming on mosquito-borne arboviruses. *Journal of Medical Entomology* 31(3) : 323-332.
- Reis, B.Y. et Mandl, K.D. (2003). Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making* 3(1) : 2.
- Remus, N., El Baghdadi, J., Abel, L., Casanova, J.L. (2005). Génétique et Immunité de la tuberculose. *Archives de pédiatrie* 12 : S74-S79.
- Rican, S. (1998). La cartographie des données épidémiologiques. Les principales méthodes de discrétisation et leur importance dans la représentation cartographique. *Cahiers Santé* 8 : 461-470.
- Ricklefs, R.E. et Schuller, D. (Eds.) (1993). Species diversity in ecological communities. Historical and geographical perspectives. University of Chicago Press, Chicago, 414 pages.
- Ridley, M. (1997). *Évolution biologique* (2<sup>nd</sup> Ed.). DeBoeck Université, Bruxelles, 719 pages.

- Rieder, H.L. (1999). Infection with tubercle bacilli. Dans Rieder, H.L. (Ed.), *Epidemiologic basis of tuberculosis control*. International Union Against Tuberculosis and Lung Disease (IUATLD), Paris, pp.17-63.
- Rieder, H.L., Watson, J.M., Raviglione, M.C., Forssbohm, M., Migliori, G.B., Schwoebel, V., *et al.* (1996). Surveillance of tuberculosis in Europe. Recommendations of a Working Group of the World Health Organization (WHO) and the European Region of the International Union Against Tuberculosis and Lung Disease (IUATLD) for uniform reporting on tuberculosis cases. *European Respiratory Journal* 9 : 1097-1104.
- Robin, M. (2002). *Téledétection, des satellites aux SIG*. 2<sup>nd</sup> Ed. Nathan Université, Paris, 318 pages.
- Rosenberg, N.A., Tsolaki, A.G., Tanaka, M.M. (2003). Estimating change rates of genetic markers using serial samples: applications to the transposon IS6110 in *Mycobacterium tuberculosis*. *Theoretical Population Biology* 63(4) : 347-363.
- Rumeau-Rouquette, C., Blondel, B., Kaminski, M., Bréart, G. (1993). *Épidémiologie : méthodes et pratiques*. Flammarion Médecine-Sciences, Paris, 312 pages.
- Schwoebel, V., Rieder, H.L., Watson, J., Raviglione, M., for the working group for uniform reporting on TB cases in Europe (1996). Surveillance of TB in Europe, *Euro Surveillance* 1 : 5-8.
- Seber, G.A.F. (1982). *The estimation of animal abundance and related parameters*, 2nd ed., Griffin, London, 654 pages.
- Sekar, C. et Deming, E.W. (1949). On a method of estimating birth and death rates and extent of registration. *Journal of the American Statistical Association* 44 : 1059-68.

- Small, P.M., Hopewell, P.C., Singh, S.P., Paz, A., Parsonnet, J., Ruston, D.C., *et al.* (1994). The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *New England Journal of Medicine* 330 : 1703-1709.
- Small, P. et Behr, M. (1997). Molecular fingerprinting of *Mycobacterium tuberculosis*: how can it help the clinician? *Clinical Infectious Diseases* 25 : 806-10.
- Smith, J.M., Smith, N.H., Orourke, M., Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America* 90 : 4384-4388.
- Société de Pneumologie de Langue Française (SPLF) (2004). Recommandations de la Société de Pneumologie de Langue Française sur la prise en charge de la tuberculose en France. *Revue des Maladies Respiratoires* 21 : 414-420.
- Sola, C., Devallois, A., Horgen, L., Maisetti, J., Filliol, I., Legrand, E., Rastogi, N. (1999). Tuberculosis in the Carribean: using spacer oligonucleotide typing to understand strain origin and transmission. *Emerging Infectious Diseases* 5 : 404-414.
- Sola, C., Filliol, I., Legrand, E., Rastogi, N. (2000). Recent developments of spoligotyping as applied to the study of epidemiology, biodiversity and molecular phylogeny of the *Mycobacterium tuberculosis* complex. *Pathologie Biologie* 48(10) : 921-932.
- Sola, C., Filliol, I., Gutierrez, M.C., Mokrousov, I., Vincent, V., Rastogi, N. (2001a). Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerging Infectious Diseases* 7(3) : 390-396.
- Sola, C., Filliol, I., Legrand, E., Mokrousov, I., Rastogi, N. (2001b). *Mycobacterium tuberculosis* phylogeny reconstruction based on combined

- analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. *Journal of Molecular Evolution* 53 : 680-689.
- Song, C. et Kulldorff, M. (2003). Power evaluation of disease clustering tests. *International Journal of Health Geographics* 2 : 9.
- Tanaka, M.M. et Francis, A.R. (2005). Methods of quantifying and visualising outbreaks of tuberculosis using genotypic information. *Infection, Genetics and Evolution* 5(1) : 35-43.
- Tango, T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine* 14 : 2323-2334.
- Tango, T. (2000). A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine* 19 : 191-204.
- Tanser, F.C. et Le Sueur, D. (2002). The application of geographical information systems to important public health problems in Africa. *International Journal of Health Geographics* 1(1) : 4.
- Taylor, Z. et O'Brien, R.J. (2001). Tuberculosis elimination - Are we willing to pay the price? *American Journal of Respiratory and Critical Care Medicine* 163(1) : 1-2.
- TLFI (Le trésor de la langue française informatisé) <http://atilf.atilf.fr/>
- Tocque, K., Bellis, M.A., Beeching, N.J., Davies, P.D. (2001). Capture-recapture as a method of determining the completeness of tuberculosis notifications. *Communicable Disease and Public Health* 4 : 141-143.
- Tran, A., Gardon, J., Weber, S., Polidori, L. (2002). Mapping disease incidence in suburban areas using remotely sensed data. *American Journal of Epidemiology* 156(7) : 662-668.

- Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L., Clark, L.C. (1990). Monitoring for clusters of disease: Application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 132 : S136-S143.
- Ureña Rib, P. (2002). Les jeunes de la Guyane et la pluri-ethnicité. Dans Doumenge, J-P. (Ed.), *La France et les Outre-Mers. L'enjeu multiculturel. Hermes* 32-33 : 69-81.
- van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., *et al.* (1993). Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *Journal of Clinical Microbiology* 31 : 406-409.
- Verver, S., Warren, R.M., Munch, Z., Vynnycky, E., van Helden, P.D., Richardson, M. *et al.* (2004). Transmission of tuberculosis in a high incidence urban community in South Africa. *International Journal of Epidemiology* 33(2) : 351-357.
- Veyret, Y. (Ed.) (2003). *Les risques*. Dossiers des Images Économiques du Monde. SEDES, Paris, 255 pages.
- Viel, J.F., Arveux, P., Baverel, J., Cahn, J.Y. (2000). Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 152(1) : 13-19.
- Walter, S.D. (1994). A simple test for spatial pattern in regional health data. *Statistics in Medicine* 13 : 1943-1950.
- Warren, R.M., Streicher, E.M., Sampson, S.L., van der Spuy, G.D., Richardson, M., Nguyen, D., *et al.* (2002). Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *Journal of Clinical Microbiology* 40(12) : 4457-4465.

- Washino, R. et Wood, B. (1994). Application of remote sensing to arthropod vector surveillance and control. *American Journal of Tropical Medicine and Hygiene* 50(6) : 134-144.
- Weber, C. et Hirsch, J. (1992). Some urban measurements from SPOT data: urban life quality indices. *International Journal of Remote Sensing* 13(17) : 3251-3261.
- Wint, G.R., Robinson, T.P., Bourn, D.M., Durr, P.A., Hay, S.I., Randolph, S.E., Rogers, D.J. (2002). Mapping bovine tuberculosis in Great Britain using environmental data. *Trends in Microbiology* 10(10) : 441-444.
- Wittes, T.J. et Sidel, W.V. (1968). A generalisation of the simple capture-recapture model with application to epidemiological research. *Journal of Chronic Diseases* 21 : 287-301.
- Wittes, J.T., Colton, T., Sidel, V.W. (1974). Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources *Journal of Chronic Diseases* 27 : 25-36.
- World Health Organisation (WHO) (1997). *Anti-tuberculosis drug resistance in the world: the WHO/IUTLD global project on anti-tuberculosis drug resistance surveillance 1994-7*. WHO, Geneva.
- World Health Organization (WHO) (2001). *Global tuberculosis control. WHO report 2005. Communicable diseases*. World Health Organization, Geneva.
- World Health Organization (WHO) (2005). *Global tuberculosis control: surveillance, planning, financing*. WHO report 2005. World Health Organization, Geneva, 247 pages.
- Yamagata, Y., Wiegand, C., Akiyama, T., Shibayama, M. (1988). Water turbidity and perpendicular vegetation Indices for paddy rice flood damage analyses. *Remote Sensing of Environment* 26(3) : 241-251.

- Yang, Z.H., Rendon, A., Flores, A., Medina, R., Ijaz, K., Llaca, J., *et al.* (2001). A clinic-based molecular epidemiologic study of tuberculosis in Monterrey, Mexico. *International Journal of Tuberculosis and Lung Disease* 5(4) : 313-20.
- Yeh, R.W., Ponce de Leon, A., Agasino, C.B., Hahn, J.A., Daley, C.L., Hopewell, P.C., Small, P.M. (1998). Stability of *Mycobacterium tuberculosis* DNA genotypes. *Journal of Infectious Diseases* 198, 177 : 1107-1111.
- Yew, W.W. et Chau, C.H. (1995). Drug-resistant tuberculosis in the 1990s, *European Respiratory Journal* 8 : 1184-1192.
- Yuen, K.Y., Chan, C.M., Chan, K.S., Yam, W.C., Ho, P.L., Chau, P.Y. (1995). IS6110 based amplotyping assay and RFLP fingerprinting of clinical isolates of *Mycobacterium tuberculosis*. *Journal of Clinical Pathology* 48 : 924-928.



# Glossaire

**Antibiogramme** : Étude complète d'un germe à l'égard de tous les antibiotiques et de leurs associations, aussi bien *in vitro* (antibiogramme quantitatif en milieu liquide) qu'*in vivo* (évaluation de la virulence du germe sur un animal et recherche sur cet animal des doses actives des antibiotiques et de leurs associations).

**Bacilles acido-alcoolo-résistants (BAAR)** : Les bacilles sont de bactéries en forme de bâtonnets qui sont à l'origine de nombreuses maladies. Le bacille tuberculeux se caractérise par son caractère acido-alcoolo résistant, c'est-à-dire que sa paroi résiste aux acides et aux alcools. Cette propriété est à l'origine de deux colorations permettant la mise en évidence du bacille par examen direct au microscope.

**BCG (bacille de Calmette-Guérin)** : C'est une souche vivante atténuée de *Mycobacterium bovis*, utilisée pour la vaccination de la tuberculose depuis les années 1940 sur le plan mondial. Le vaccin par le BCG a un caractère obligatoire en France depuis 1952.

**Contagiosité** : La tuberculose n'est contagieuse que dans sa forme pulmonaire ou laryngée. Le risque de transmission de personne à personne dépend de la densité de bacilles dans l'expectoration, de la proximité du contact avec un individu sain, et de la durée de ce contact. Les formes extra-pulmonaires ne sont qu'exceptionnellement contagieuses.

**DOTS** : Directly Observed Treatment, Short course, ou traitement de courte durée sous supervision directe.

**DSDS** : Département de la Santé et du Développement Social. La dénomination DSDS est spécifique à la région Antilles-Guyane. Elle correspond au regroupement d'une Direction régionale (Drass) et d'une Direction départementale des affaires sanitaires et sociales (Ddass).

**Endémie** : Maladie qui se manifeste en permanence dans une région déterminée, ou y existe à l'état latent.

**Épidémie** : Augmentation inhabituelle et subite du nombre d'individus atteints d'une maladie transmissible existant à l'état endémique dans une région ou une population donnée; apparition d'un nombre plus ou moins élevé de cas d'une maladie transmissible n'existant pas normalement à l'état endémique dans une région donnée.

**Épidémiologie** : Discipline qui a pour objet l'influence de divers facteurs (constitution individuelle, environnement, mode de vie, milieu social) sur les maladies, notamment sur leur fréquence, distribution et étiologie, ainsi que sur tout autre phénomène biologique ou social déterminé.

**Géomatique** : Discipline ayant pour objet la gestion des données à référence spatiale et qui fait appel aux sciences et aux technologies reliées à leur acquisition, leur stockage, leur traitement et leur diffusion (par exemple, la topométrie, la cartographie, la géodésie, la photogrammétrie, la télédétection et l'informatique).

**GPS (Global Positioning System)** : Système de localisation à l'échelle du globe : constellation de satellites dont le système de fonctionnement permet à un utilisateur muni d'un récepteur GPS de déterminer sa localisation, en tout point du globe et en temps réel, avec une précision de quelques mètres.

**Incidence** : Le taux d'incidence d'une maladie est égal au nombre de nouveaux cas pendant une période donnée divisée par la population totale pendant cette période.

**Intradermo-réaction (IDR)** : Réaction inflammatoire locale provoquée par l'injection intradermique d'une petite quantité d'antigène ou de toxine, et qui permet d'étudier le degré d'immunité ou de réceptivité du sujet à l'égard des maladies correspondantes. Elle est utilisée dans le diagnostic de certaines infections, en particulier la primo-infection tuberculeuse.

**Multirésistance** : Résistance à au moins deux antibiotiques anti-tuberculeux, en général l'isoniazide et la rifampicine.

**Observance** : Elle se définit comme la capacité, pour un patient, de respecter la prescription du médecin. Le terme de "compliance" est également utilisé.

**PCR** : Polymerase Chain Reaction (réaction de polymérisation en chaîne). Technique de répllication in vitro permettant d'obtenir, à partir d'un échantillon complexe et peu abondant d'ADN, et avec l'intervention d'une ADN-polymérase, d'importantes quantités d'un fragment d'ADN spécifique et de longueur définie.

***Prophylaxie*** : Ensemble des moyens destinés à prévenir l'apparition, la propagation ou l'aggravation des maladies. **Chimioprophyllaxie** : action préventive obtenue grâce à l'utilisation de certains produits chimiques permettant d'éviter l'infection.

***Résistance primaire*** : Résistance à un ou plusieurs antibiotiques observée chez un patient n'ayant jamais reçu auparavant de traitement anti-tuberculeux.

***Résistance secondaire*** : Résistance à un ou plusieurs antibiotiques observée chez un patient ayant déjà reçu un traitement anti-tuberculeux (soit plus de deux mois de traitement spécifique).

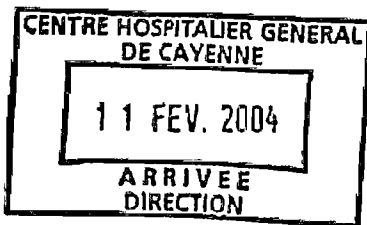
***SIG (Système d'Information Géographique)*** : Outil informatique qui permet de saisir, représenter, transformer, interroger, mettre à jour toutes sortes d'informations graphiques ou attributaires référencées dans l'espace. Classiquement, son principe est illustré par la superposition de cartes de thèmes différents.

***Téledétection*** : Ensemble des connaissances et des techniques utilisées pour acquérir des informations sur l'environnement par des mesures effectuées à distance, sans contact matériel avec celui-ci.



## **ANNEXE 1**

### **ENGAGEMENT DE CONFIDENTIALITÉ**



## ENGAGEMENT DE CONFIDENTIALITÉ

*Je soussignée Vanina GUERNIER, certifie avoir été informée de mes devoirs en matière de secret professionnel et notamment du contenu de l'article 72 du code de Deontologie médicale ainsi que du contenu de l'article 226-13 du code penal qui précise les sanctions encourues en cas de « révélation d'une information a caractere secret par une personne qui en est dépositaire, soit par état, soit en raison d'une fonction ou d'une mission temporaire »*

*Je m'engage à ne porter aucune atteinte au secret professionnel qui s'attache aux informations nominatives auxquelles j'aurai accès dans le cadre de l'étude que je mène et qui fait le sujet de mon travail de thèse « Épidémiologie de la tuberculose et de la fièvre Q sur l'Île de Cayenne. Analyse des facteurs de risque et étude de la dynamique spatio-temporelle des maladies »*

Accord

à Cayenne, le 26 janvier 2004

Vanina GUERNIER

- Vu la compétence et l'expertise particulières de la Direction de la Santé et du Développement Social (DSDS) de Guyane dans le domaine de la santé publique, et la mission qui lui a été conférée concernant le recueil prospectif des données de santé de la population guyanaise,
- Vu la mission conférée à l'Institut Pasteur de Guyane concernant le recueil et la transmission au médecin investigateur du signalement des cas cliniques et des données bactériologiques (examens directs, cultures et antibiogramme),
- Au vu des conventions signées dans le cadre de la thèse de Mlle Vanina Guernier entre l'Institut de Recherche pour le Développement (IRD) et la DSDS d'une part, et entre l'IRD et l'Institut Pasteur de Guyane d'autre part,

Les parties sont convenues d'accorder à Mlle Guernier l'accès à l'ensemble des données nominatives et informatisées concernant les cas de tuberculose et de fièvre Q survenus en Guyane sur la période considérée (de 1996 à 2003)

Fait à Cayenne, le 26 janvier 2004 en 3 exemplaires originaux, chacune des versions faisant pareillement foi

Pour l'Institut Pasteur

Docteur Xavier DEPARIS

Pour la DSDS

Docteur Benoît COTTRELLE

## **ANNEXE 2**

**GUERNIER V., GUÉGAN, J-F., DEPARIS X. An evaluation of the actual incidence of tuberculosis in French Guiana using a capture-recapture model (accepté le 8 septembre 2005 dans *Microbes and Infection*).**



Original article

## An evaluation of the actual incidence of tuberculosis in French Guiana using a capture-recapture model

Vanina Guernier<sup>a,b,\*</sup>, Jean-François Guégan<sup>a</sup>, Xavier DeParis<sup>c</sup>

<sup>a</sup> UMR 2724 IRD-CNRS, Génétique et Évolution des Maladies Infectieuses, Equipe Évolution des Systèmes Symbiotiques, 911 avenue Agropolis, BP 64501, 34394 Montpellier cedex 05, France

<sup>b</sup> Unité S140 Expertise et Spatialisation des Connaissances en Environnement, IRD, Maison de la Télé-détection, 500 rue J-F. Breton, 34093 Montpellier cedex 05, France

<sup>c</sup> Institut Pasteur de la Guyane, BP 6010, 97306 Cayenne, French Guiana

Received 28 April 2005; accepted 7 September 2005

Available online 13 January 2006

### Abstract

In order to estimate the level of under-reporting and to improve estimates of the incidence of tuberculosis (TB) in the vicinity of Cayenne, French Guiana, we performed capture-recapture analysis from 1996 through 2003. We cross-linked data from the Institut Pasteur, the Département d'Information Médicale of Cayenne Hospital, and the Service de Lutte Anti-Tuberculeuse. The estimate of 381 TB cases obtained after matching those three sources was revised to 425 (95% confidence interval: 407, 453) using the capture-recapture model based on sample coverage. The corresponding average annual incidence was 63.1 TB cases per 100,000 population. The evaluated sensitivity of the compulsory notification system was 35.3%, indicating wide under-notification of TB in the vicinity of Cayenne. The estimated coverage reported by the three sources was fairly accurate (i.e. 85.9%), but not sufficient to evaluate the risk of transmission of TB in the Ile-de-Cayenne (Cayenne and its suburbs).

© 2006 Elsevier SAS. All rights reserved.

**Keywords:** Tuberculosis; French Guiana; Disease notification; Transmission; Surveillance; Incidence; Statistical models

### 1. Introduction

Human tuberculosis (TB) is caused by the bacterium *Mycobacterium tuberculosis* (and occasionally by *Mycobacterium bovis*, which is generally more pathogenic for animals). This

major human disease, responsible for 1.9 million deaths per year worldwide, has re-emerged as a serious public health threat. Almost two billion people are latently infected with the tubercle bacillus and more than eight million new cases are diagnosed each year [1]. *M. tuberculosis* is transmitted from person to person via the aerosol route through small droplets of saliva and sputum expelled when an infectious patient sneezes or coughs. Active TB infection requires treatment with a combination of three or four antibiotics for at least six months. Failure to complete the full course of drug therapy can lead to the emergence of drug-resistant strains of mycobacteria [2,3]. This so-called acquired resistance, as opposed to primary resistance [4], severely limits effective treatment options. Moreover, it can lead to the maintenance of a pool of potentially resistant bacteria in the environment, thus hindering arrest of the chain of transmission. Likewise, if not

**Abbreviations:** CI, confidence interval; DIM, Département d'Information Médicale; INSEE, National Institute for Statistics and Economic Studies; InVS, Institut de Veille Sanitaire; IPG, Institut Pasteur de la Guyane; LAT, Lutte Anti-tuberculeuse; MDRTB, multidrug-resistant tuberculosis cases; TB, tuberculosis.

\* Corresponding author. UMR 2724 IRD-CNRS, Génétique et Évolution des Maladies Infectieuses, Equipe Évolution des Systèmes Symbiotiques, 911 avenue Agropolis, BP 64501, 34394 Montpellier cedex 05, France. Tel.: +33 4 67 54 87 12; fax: +33 4 67 54 87 00.

E-mail address: [guernier@mpl.ird.fr](mailto:guernier@mpl.ird.fr) (V. Guernier).

screened and treated appropriately, persons with unnotified cases are likely to transmit the disease and/or to develop drug-resistant TB [5,6].

These points underscore the need for effective control strategies, specifically for TB. Early identification of cases ensures that treatment is completed and enables close-contact investigations and screening [7]. It also provides data to measure disease incidence, monitor epidemiological trends, detect outbreaks and plan new strategies for prevention and treatment. For these reasons, obligatory notification of all TB cases is particularly useful. Several studies have evaluated the completeness of TB notification over a defined period and within particular geographical areas [8–11], while recent studies often use capture-recapture methodology [12]. In France, TB has been statutorily considered a notifiable disease since 1964. The Health Monitoring Institute (InVS, Institut de Veille Sanitaire), in charge of epidemiological surveillance, reported the highest TB rates in French Guiana and in the suburbs of Paris. But an evaluation of the completeness of notification in French Guiana, comparing culture-positive cases with notified cases from 1994 to 1996, has shown strong evidence of under-notification of TB in that department [13]. For example, in 1998, only six cases were notified to the InVS for the all French Guiana, whereas the Institut Pasteur reported 20 biologically confirmed cases in the Ile-de-Cayenne only (i.e. Cayenne and its suburbs). At present, dissemination of health services in the territory (3 hospitals on the coast and 21 health centers dispersed throughout the territory) and under-reporting of TB by medical practitioners preclude data completeness and analysis of a larger data set for the whole of French Guiana. This is the reason why we considered only TB cases in the population of the Ile-de-Cayenne, where most new cases occur and for which most health information is available. In the present work, we performed capture-recapture analysis to evaluate the actual TB incidence in the Ile-de-Cayenne and to provide estimates of TB incidence for all of French Guiana.

## 2. Materials and methods

### 2.1. Study period

The analysis was restricted to the period from January 1, 1996 to December 31, 2003, which is the most reliable period in terms of available data sources. TB cases were notified using the date of registry in each monitoring system, thus providing daily data.

### 2.2. Study site and population

French Guiana is a French overseas department comprising 32,432 square miles. It is located in the northeast of South America, between Surinam and Brazil. The study sector, locally referred to as the Ile-de-Cayenne, is located along the Atlantic coast in the north of French Guiana. This area is made up of three communities, Cayenne (the capital), Rémire-Montjoly and Matoury, and constitutes the main economic pole of attraction with the highest population density

in French Guiana. More than 90% of the 157,213 inhabitants of French Guiana (census 1999 from INSEE – National Institute for Statistics and Economic Studies; <http://www.insee.fr>) are concentrated in the coastal region, a tropical rainforest spreading over 90% of the area, and the Ile-de-Cayenne represents more than 50% of the total population of French Guiana (84,181 inhabitants in 1999).

### 2.3. Data sources

In order to select patients to be included in our analysis, we used the European consensus definition of a notifiable case of TB [14,15]. A case is defined as any person administered a full course of antituberculosis treatment (at least three antibiotics). This definition excludes prophylaxis TB and atypical mycobacteria, but includes definite cases in which *M. tuberculosis* was confirmed by direct examination and/or culture, as well as clinical cases, non-confirmed in the laboratory but treated based on a clinician's decision. Cases which appeared more than once in a source, and were generally relapsed cases, were included once in our data set, even when the cases were reported several times upon obligatory notification. The earliest record was retained from each duplicate set. The same rule applied to two cases notified as chronic. Cases already treated for TB before January 1, 1996 and again treated after that date were ignored. Since no proper databases yet exist, we censused all cases via surveys conducted in three different health monitoring systems.

The Mycobacterium Laboratory of the Institut Pasteur de la Guyane (IPG) is in charge of the culture of isolates from patients suspicious for TB and forwarded by hospitals or private general practitioners. Cultures are then sent to the Institut Pasteur de la Guadeloupe (French West Indies), where identification of the mycobacterium and molecular analysis are performed. Because genotyping requires the availability of a viable isolate of *M. tuberculosis*, the population for which the culture is positive and for which genotyping is performed constitutes a subset of all cases, excluding clinical cases or extrapulmonary TB for which biopsy specimens are not always available for bacteriology. For this reason, data censused at the IPG are not exhaustive for TB notification.

The Département d'Information Médicale (DIM) of Cayenne Hospital collects and computes medical information on all patients treated in or admitted to the Cayenne Hospital, such as the name, surname, date of birth, address, health problems and results of key medical tests. Diagnoses are recorded using the 10th International Classification of Diseases (ICD-10) coding system. For the needs of the study, a listing of cases was automatically extracted from the DIM general database based on principal as well as secondary diagnosis of TB. When there was doubt concerning the diagnosis, the medical record was inspected for confirmation of a true case of TB.

The Service de Lutte Anti-Tuberculeuse (LAT), a branch of the Conseil Général of Cayenne, i.e. a Department Council, is in charge of follow-up and treatment of TB cases when patients leave the hospital in which they were originally treated in individual isolation rooms during the infectious stage of the

disease. The LAT is also in charge of screening surveys conducted among close relatives or neighbors of the patients in order to detect persons at risk of exposure, and to provide post-exposure prophylaxis if necessary. Patients immigrating from surrounding countries where they started a treatment are followed up by the LAT upon their arrival in French Guiana, and they were also included in our data set.

#### 2.4. Linkage among registries

Identifiers common to the three sources were needed to match patients present in more than one source. Personal information (name, sex, date of birth, address and date of registry in the system) was collected for each case and used to detect duplicates. It should be noted that this research proposal, including the collection of data from clinical records, ensures protection of confidentiality according to the code of medical ethics, legislation and research findings. Duplicate cases were identified by examining the name and date of birth, and occasionally a third item when in doubt. The earliest record was retained from each duplicate set, and the registered date was used for all further analyses.

#### 2.5. Capture-recapture method

Population estimates using capture-recapture analysis consist of cross-matching the information from two or more databases in order to identify the number of individuals common to paired lists (matched cases) and then using overlapping information and statistical techniques to evaluate the number of missing individuals and, consecutively, the total population size [16–20]. This method has been used in the epidemiology for over 50 years to improve estimates of disease incidence and deaths [21–23], with increased use during the past 20 years [24–28]. Four assumptions underlie this method [12,28,29]: cases must have equal probabilities of being observed (captured) in any source; cases must be uniquely matched among the various data sources; case ascertainment by sources must be independent; the population under study must be 'closed'. Two sources are considered independent when the probability of notification of an event in one source is not dependent upon the probability of notification of that event in the other source, and when the probability of notification of an event is not influenced by its characteristics (e.g. age, gender, symptoms, etc.) [30].

When only two data sources are available, the assumption of independence is difficult to justify in epidemiological surveys. For example, in our study, the majority of cases reported from IPG and LAT comes from the Cayenne Hospital where suspicious TB cases are first noted; therefore, dependence should occur between DIM and IPG, and DIM and LAT, and independence between IPG and LAT. But with a multiple source approach (more than two sources), which is the case here, the assumption of independence can be dropped by using specified statistical techniques which take into account interdependence among data sets, the most commonly

used being log-linear models and the sample coverage approach [31].

#### 2.6. Estimate of the number of TB cases

Specific software for capture-recapture models was used to estimate the size of the population of French Guiana infected with TB: CARE-1 [31], developed for an S-Plus environment [32] (available at <http://chao.stat.nthu.edu.tw/softwareCE.html>). Three different statistical methods were tested: Petersen and Chapman estimators, log-linear models and the sample coverage approach. The standard Petersen estimator and the nearly unbiased Chapman estimator, based on any pair of samples, must be used under the assumption of independence, which is unlikely to be met here, but those estimators are useful in preliminary analysis. With the log-linear modeling approach, data are considered as a  $2^t$  contingency table (with  $t$  the number of sources) with an empty cell, corresponding to the number of cases not identified by any of the lists. Those missing cases are estimated while modeling two-way interactions (and therefore dependence) between sources in various log-linear models. The estimate of missing cases can be derived from the model that best fits the data [33,34]. The sample coverage approach incorporates the correlation bias contained in a three-sample model, taking into account recapture information and sample dependencies. It is used as a measure of the overlap fractions among sources. Three types of estimators are calculated: the population size estimate for independent samples ( $\hat{N}_0$ ), a sufficiently high number of sample coverage cases ( $\hat{N}$ ) and low sample coverage cases ( $\hat{N}_1$ ) (see [31] for details).

### 3. Results

#### 3.1. Simple matching of sources

After cross-linking of records from the IPG, DIM and LAT, 381 TB cases were reported in the Ile-de-Cayenne from January 1, 1996 to December 31, 2003. Among these 381 identified cases, 235 (61.7%) were confirmed by isolation and cell culture of *M. tuberculosis*, 111 (29.1%) were found in only one registry, 132 (34.7%) in only two registries and 138 cases (36.2%) in all three sources of information. The cumulative number of TB cases found in one, two or three registries is illustrated in Fig. 1. During the census, no cases of *M. bovis* infection were identified.

#### 3.2. Capture-recapture estimates

The Petersen and Chapman estimates are in the range of 365–406 (see Table 1 for details), given that the Petersen estimator underestimates the true size when both samples are positively dependent and overestimates the true size in case of negative dependence (see the sample coverage approach for identification of dependent sources).

Results for all log-linear models considered are summarized in Table 2, with corresponding deviances and estimates

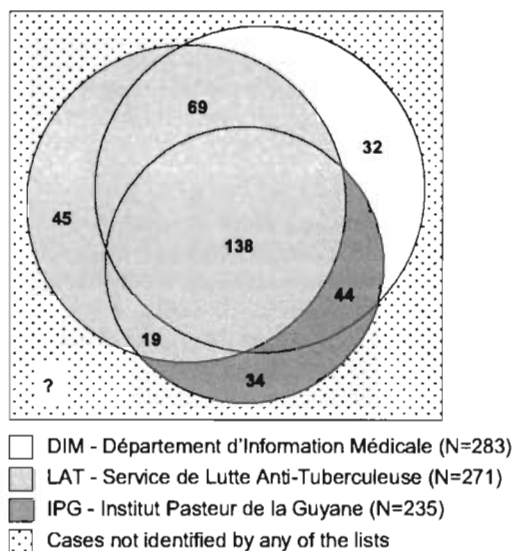


Fig. 1. Three-source combinations of epidemiological reports for the Ile-de-Cayenne for 1996 through 2003 (N = 381). The number of cases identified in one, two or three sources is noted.

of the total number of TB cases. The best-fitting model is 12/23 (i.e. dependence arises between sources one and two and sources two and three), which produces an estimate of 462 cases (95% confidence interval (CI): 423, 536). The dependence between sources one and two (IPG and DIM) and between sources two and three (DIM and LAT) was highly intuitive (see Section 2). The degree of ascertainment of monitoring systems (or “sensitivity”) was 61.3%, 50.9%, and 58.7%, respectively, for DIM, IPG and LAT, with this model.

Using the third approach, the sample coverage estimate for our data set, i.e. the proportion of cases that were not unique, was 85.9%, and the average of the overlapped cases, i.e. the number of cases found in two lists, excluding the third, was equal to  $D = 344$  (Table 3). The sample coverage estimate of 85.9% was sufficiently high to correct for undercount, and estimator  $\hat{N}$  was used. It gives an estimated population size of 425 (95% CI: 407, 453) using 1000 bootstrap replications. Sensitivities were 66.6% for DIM, 55.3% for IPG and 63.8% for LAT. These data were sufficient to produce a reliable

Table 1  
Population size estimates based on any pair of samples using Petersen and Chapman estimators

	Petersen	Chapman	se <sup>a</sup>	cil <sup>b</sup>	ciu <sup>c</sup>
Pair (1,2) <sup>d</sup>	365	365	8	354	384
Pair (1,3)	406	405	12	386	434
Pair (2,3)	370	370	6	361	387

<sup>a</sup> se, Asymptotic standard error.  
<sup>b</sup> cil, 95% CI lower limit. A log-transformation is used so that the lower limit is always greater than the number ascertained. See [17] for details.  
<sup>c</sup> ciu, 95% CI upper limit.  
<sup>d</sup> Refers to paired samples (1, Institut Pasteur de la Guyane; 2, Département d'Information Médicale; 3, Service de Lutte Anti-Tuberculeuse).

Table 2  
Estimates based on log-linear models

	dev <sup>a</sup>	df	est	se	cil	ciu
Independent	39.75	3	395	4	389	407
13/2 <sup>b</sup>	38.77	2	393	4	387	406
23/1	27.47	2	406	7	395	424
12/3	28.27	2	403	6	394	419
12/23	1.84	1	462	28	423	536
12/13	28.20	1	402	7	392	421
23/13	27.47	1	406	9	394	430
Symmetry <sup>c</sup>	32.75	4	463	34	419	560
Quasi-sy	17.28	2	464	34	419	561
Part-qs1	0.32	1	495	49	431	638
Part-qs2	11.37	1	483	43	427	606
Part-qs3	15.03	1	460	33	417	555
Saturated	0.00	0	498	51	433	647

All possible models derived from a triple sample are considered: independent model, models with one or two interactions which consider local independence only (i.e. 13/1, 23/1, 12/3, 12/23, 12/13, 23/13), models taking heterogeneity into account (quasi-symmetric and partial quasi-symmetric models), and saturated models.

<sup>a</sup> dev, Deviance statistics; df, degree of freedom; est, population size estimate; se, asymptotic standard error; cil, 95% confidence interval lower limit; ciu, 95% confidence interval upper limit; 1, IPG; 2, DIM; 3, LAT.

<sup>b</sup> If dependence arises in samples 1 and 2, the model is denoted as model 12/3. If dependence also appears in samples 1 and 3, the model is noted 12/13, and so on for other interactions.

<sup>c</sup> See [31] for details about the models.

population size estimate of TB cases escaping epidemiological surveillance and health care systems.

### 3.3. Incidence estimates

Considering the low number of TB cases counted in our survey in the Ile-de-Cayenne (38–70 cases per year), and in order to avoid multiplication of biases, capture-recapture

Table 3  
Outputs obtained with the sample coverage approach

	M <sup>a</sup>	D	$\hat{C}$	est	se	cil	ciu
$\hat{N}_{\text{hat-0}}^d$	381	344	0.859	401	6	392	415
$\hat{N}_{\text{hat}}$	381	344	0.859	425	11	407	453
$\hat{N}_{\text{hat-1}}$	381	344	0.859	419	9	405	441
	$u1^b$	$u2$	$u3$	$r12^c$	$r13$	$r23$	$r123$
$\hat{N}_{\text{hat-0}}^d$	0.59	0.71	0.68	0.10	-0.01	0.08	0.06
$\hat{N}_{\text{hat}}$	0.55	0.67	0.64	0.16	0.05	0.15	0.03
$\hat{N}_{\text{hat-1}}$	0.56	0.68	0.65	0.15	0.03	0.13	0.03

Three estimators were calculated, taking into account the heterogeneity of redundant information between the three samples (or sources) considered.

<sup>a</sup> M, number of cases listed in at least one list; D, average number of cases found in two lists excluding the third;  $\hat{C}$ , sample coverage estimate; est, population size estimate; se, asymptotic standard error; cil, 95% CI lower limit; ciu, 95% CI upper limit.

<sup>b</sup> u, Estimated mean probabilities depending on the estimate of N for samples 1, 2 and 3.

<sup>c</sup> r, Estimated coefficient of covariation (CCV) depending on the estimate N for different interactions between samples 1, 2 and 3.

<sup>d</sup> Refers to the population size estimate for independent samples ( $\hat{N}_{\text{hat-0}}$ ), sufficiently high sample coverage cases ( $\hat{N}_{\text{hat}}$ ) or low sample coverage cases ( $\hat{N}_{\text{hat-1}}$ ). See [31] for details.

estimates were calculated once with the sum of cases reported from January 1, 1996 to December 31, 2003. The estimated mean incidence was therefore calculated using the estimated number of 425 cases divided by eight years, i.e. 53.1 cases per year for 84,181 inhabitants. Thus, the average annual incidence rate determined by capture-recapture estimates was 63.1 TB cases per 100,000 population in the Ile-de-Cayenne. However, this might represent an underestimation of the actual incidence, since the number of TB cases increased from 1998 in the Ile-de-Cayenne. During the same period, annual incidences per 100,000 population officially reported by the InVS were 24.2 for all of French Guiana, 27.8 for the Ile-de-France (Paris and its suburbs) and 11.3 for metropolitan France.

#### 4. Discussion

This is the first study to estimate the true number of TB cases in the Ile-de-Cayenne from 1996 through 2003. Indeed, the disease is recognized to be widespread but is also highly underestimated in French Guiana. Using the capture-recapture method, the estimate of 381 TB cases obtained by simple linking of information from three sources was revised to 425 (95% CI: 407, 453) with the sample coverage approach, which represents an increase of 11.6%. Similar estimates were found using log-linear models. The difference between notified cases and cross-linking results is probably due to the fact that there are only two pulmonary disease specialists in the Ile-de-Cayenne and a general shortage of physicians who do not have enough time for notifications, whereas differences between results from cross-linking and sample coverage might be related to cases that are less likely to be treated, among homeless and non-sedentary populations, for example.

Despite the low declaration of TB cases in French Guiana, the sensitivity of the three combined systems in the Ile-de-Cayenne was quite good, but such cross-linked data are not available, as there is no centralization of data. For this reason, the degree of ascertainment of each monitoring system was evaluated independently. Estimated sensitivity was 35.3% for the notification system compared to sample coverage approach results. Sensitivities of the three monitoring systems varied between 55.3% and 66.6% as a function of the model used. These values are similar to those evaluated in previous studies, for example, on malaria in French armies, with sensitivities between 49.6% and 56.2% [26], and Legionnaire's disease in France, with sensitivities between 58% and 66% [35]. This implies that the sensitivity of health monitoring systems in French Guiana is no worse than that of similar French institutions concerning different diseases.

It is possible to obtain an overall picture of the TB incidence in the whole of French Guiana based on the fact that 61.7% of cases in the Ile-de-Cayenne were culture-confirmed. If the same ratio is applied to the 345 confirmed cases censused by the IPG throughout the region during the same time period, the total number of TB would be about 559, i.e. an average annual incidence rate of 44.4 TB cases per

100,000 population. This value is undoubtedly underestimated, given that the TB incidence has increased in French Guiana over the past five years. Nevertheless, such an estimate is higher than the incidence previously reported by the InVS: 23.8 per 100,000 population per annum on an average for the period 1996–2003, with a maximum of 39.4 in 2000 (<http://www.invs.sante.fr>). The number of TB cases actually occurring annually since 1996 in French Guiana has been estimated so as to enable comparison between monitoring system estimates and reality (Table 4 and Fig. 2). In 2003, for example, the number of biologically confirmed cases (from IPG) was 66, providing an estimate of  $66 \times 100/61.7 = 107$  cases in French Guiana. The population size in French Guiana was evaluated at 176,117 inhabitants in 2003, hypothesizing a linear monotonous increase in the population between the 1990 and 1999 censuses; therefore, the 2003 incidence may be almost 60.8 TB cases per 100,000 population, showing a strong increase in the TB threat over the last few years. Moreover, if the actual number of cases (including missing cases) is presumed to be 11.6% higher, as evaluated by the sample coverage approach, the estimated 2003 incidence would be 67.8 per 100,000 population. Fig. 2 provides a comparison of the degree of ascertainment of the different TB monitoring systems, taking into account the number of cases reported.

When interpreting our findings, certain limitations should be noted. First, the final calculated TB incidence in French Guiana is an approximate result based on different assumptions. However, from a public health point of view, it is important to estimate the number of people who actually have TB and are thus potentially contagious. The aim of our evaluation

Table 4  
Comparisons between obligatory notifications of (TB) from governmental institutions and the estimated number of cases that should have actually occurred (1) in the Ile-de-Cayenne and (2) in all of French Guiana from 1996 through 2003

Year <sup>a</sup>	Ile-de-Cayenne		French Guiana	
	DSDS <sup>b</sup>	IPG × DIM × LAT <sup>c</sup>	InVS <sup>d</sup>	est <sup>e</sup>
1996	22	38	43	50
1997	11	41	24	70
1998	4	28	6	55
1999	13	35	19	55
2000	33	46	66	60
2001	34	62	61	81
2002	21	61	43	81
2003	17	70	37	107
Total	160	381	299	559

<sup>a</sup> To homogenize the results between sources and enable comparisons, notifications from DSDS were corrected, excluding two declared atypical cases and revising the year of notification by the year notified at the beginning of treatment.

<sup>b</sup> DSDS, number of TB cases notified by the Département de la Santé et du Développement Social in the Ile-de-Cayenne.

<sup>c</sup> IPG × DIM × LAT, number of matching cases from IPG, DIM and LAT in the Ile-de-Cayenne.

<sup>d</sup> InVS, number of TB notifications from Health Monitoring Institute for all of French Guiana.

<sup>e</sup> est, Estimated number of total cases, considering that confirmed cases represent 61.7% of the total report (without any capture-recapture correction).

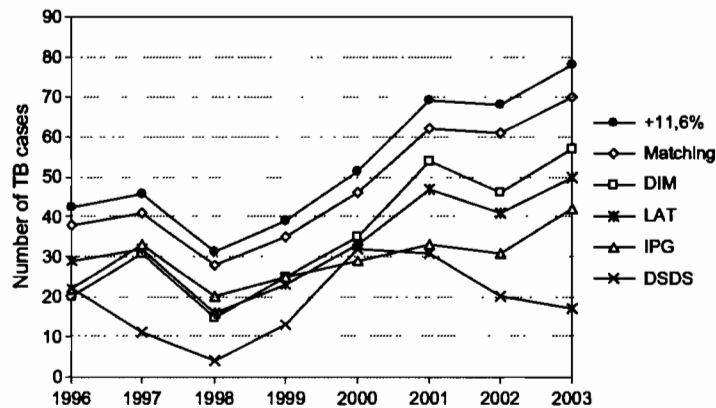


Fig. 2. Evolution of the number of TB cases in the Ile-de-Cayenne between 1996 and 2003 according to the different monitoring systems, i.e. IPG (Institut Pasteur de la Guyane), DIM (Département d'Information Médicale), LAT (Service de Lutte Anti-Tuberculeuse), and the DSDS (Département de la Santé et du Développement Social) which rebroadcast obligatory notifications to the InVS. So-called "cross-linking" data corresponding to a simple match between data from the IPG, DIM and LAT are also given. "+11.6%" is the estimated number of TB cases per year, taking into account the sample coverage approach, which gives 11.6% more TB cases than after simple matching of the three sources.

was not to provide an exact incidence rate, but to give an order of magnitude, which is already a significant advance compared to biased notified data.

A second limitation is related to assumptions inherent in capture-recapture methods. In this study, the closure assumption, i.e. the constant size of the population during the entire study period, is difficult to evaluate, as is the assumption of equal probabilities of being observed (captured) in any sample. Further research is required to evaluate the bias induced by not taking into account these assumptions and to identify the minimum data sources required using capture-recapture. Nevertheless, in spite of this remark, the combined use of cross-linkage over three sources, along with capture-recapture methods, could help to adjust for underestimates in order to achieve a more accurate view of risks and risk factors.

Finally, we should underline that our estimates do not provide an idea of the proportion of patients reported in the three monitoring systems who are presumed to have at least begun treatment and who have been cured. This rate could also be estimated, since treatment failure strongly affects TB dynamics, with the development of drug-resistant cases, which may remain infectious for long periods of time, thus maintaining the chain of transmission.

#### Acknowledgements

We would like to thank the physicians and health care professionals who provided generous help in reporting cases in the Ile-de-Cayenne: Dr. Guillot and Dr. Quintard from the Service de Lutte Anti-Tuberculeuse of Cayenne; Dr. Sobeski from the Département d'Information Médicale of the Cayenne Hospital; Dr. Duval from the Laboratory of the Institut Pasteur de la Guyane; Dr. Rastogi, Dr. Sola and K. Brudey from the Mycobacteria Reference Center of the Institut Pasteur de la Guadeloupe; and Dr. Huguet, Dr. Cottrelle and N. Joly from the Département de la Santé et du Développement Social (DSDS) of French Guiana. We would also like to thank the

personnel of the "Main Entry-Reception" of Cayenne Hospital, and E. Miatti, of the Médecine B Service of the hospital, who gave of their time. We are grateful to M. Choisy and O. Gimenez for useful discussions. V.G. is supported by a PhD fellowship from the French Ministry of Higher Education and Research, J.-F.G. by both the IRD and the CNRS, and X.D. by the Institut Pasteur.

#### References

- [1] C. Dye, S. Scheele, P. Dolin, V. Pathania, M.C. Raviglione, Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. *WHO Global Surveillance and Monitoring Project, JAMA* 282 (1999) 677–686.
- [2] K. Neville, A. Bromberg, R. Bromberg, S. Bank, B.A. Hanna, W.N. Ross, The third epidemic: multi-drug resistant tuberculosis, *Chest* 105 (1994) 45–48.
- [3] C.S. Lambregts-van Weezenbeek, J. Veen, Control of drug-resistant tuberculosis, *Tuber. Lung Dis.* 76 (1995) 455–459.
- [4] W.W. Yew, C.H. Chau, Drug-resistant tuberculosis in the 1990s, *Eur. Respir. J.* 8 (1995) 1184–1192.
- [5] World Health Organisation, Anti-Tuberculosis Drug Resistance in the World: The WHO/IUTLD Global Project on Anti-Tuberculosis Drug Resistance Surveillance 1994–7, WHO, Geneva, 1997.
- [6] P. Farmer, J.Y. Kim, Community based approaches to the control of multidrug resistant tuberculosis: introducing "DOTS-plus", *Br. Med. J.* 317 (1998) 671–674.
- [7] J. Pillay, A. Clarke, An evaluation of completeness of tuberculosis notification in the United Kingdom, *BMC Public Health* 3 (2003) 31.
- [8] C.D. Sheldon, K. King, H. Cock, P. Wilkinson, N.C. Barnes, Notification of tuberculosis: how many cases are never reported? *Thorax* 47 (1992) 1015–1018.
- [9] G.B. Migliori, A. Spanevello, L. Ballardini, M. Neri, C. Gambarini, M.L. Moro, et al., Validation of the surveillance system for new cases of tuberculosis in a province of northern Italy. Varese Tuberculosis Study Group, *Eur. Respir. J.* 8 (1995) 1252–1258.
- [10] R.M. Washko, T.R. Frieden, Tuberculosis surveillance using death certificate data, New York City, 1992, *Public Health Rep.* 11 (1996) 251–255.
- [11] A.K. Mukerjee, Ascertainment of non-respiratory tuberculosis in five boroughs by comparison of multiple data source, *Commun. Dis. Public Health* 2 (1999) 143–144.

- [12] International Working Group for Disease Monitoring and Forecasting (IWGDMF), Capture-recapture and multiple-record systems estimation II: applications in human diseases, *Am. J. Epidemiol.* 142 (1995) 1059–1068.
- [13] N. Rastogi, L. Schlegel, F. Pfaff, I. Jeanne, C. Magnien, G. Lajoie, et al., La tuberculose en région Antilles-Guyane. Situation épidémiologique de 1994–1996, *BEH* 11 (1998) 45–47.
- [14] H.L. Rieder, J.M. Watson, M.C. Raviglione, M. Forssbohm, G.B. Migliori, V. Schwoebel, et al., Surveillance of tuberculosis in Europe. Recommendations of a Working Group of the World Health Organization (WHO) and the European Region of the International Union Against Tuberculosis and Lung Disease (IUATLD) for uniform reporting on tuberculosis cases, *Eur. Respir. J.* 9 (1996) 1097–1104.
- [15] V. Schwoebel, H.L. Rieder, J. Watson, M. Raviglione, for the working group for uniform reporting on TB cases in Europe, *Surveillance of TB in Europe, Euro. Surveill.* 1 (1996) 5–8.
- [16] J.N. Darroch, The multiple-recapture census I: estimation of a closed population, *Biometrika* 64 (1958) 435–447.
- [17] G.A.F. Seber, The estimation of animal abundance and related parameters, second ed., Griffin, London, 1982.
- [18] J.D. Nichols, K.H. Pollock, Estimation methodology in contemporary small mammal capture-recapture studies, *J. Mammal.* 64 (1983) 253–260.
- [19] A.E. Magurran, *Ecological Diversity and Its Measurement*, Chapman & Hall, London, 1988.
- [20] Mathematical Sciences Education Board, National Research Council, *Measuring What Counts: A Policy Brief*, National Science Teachers Association, Washington DC, 1993.
- [21] T.J. Wittes, W.V. Sidel, A generalisation of the simple capture-recapture model with application to epidemiological research, *J. Chronic Dis.* 21 (1968) 287–301.
- [22] J.T. Wittes, T. Colton, V.W. Sidel, Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources, *J. Chronic Dis.* 27 (1974) 25–36.
- [23] A. Chao, in: P. Armitage, T. Colton (Eds.), *Encyclopedia of Biostatistics*, vol. 1, Wiley, New York, 1998, pp. 482–486.
- [24] E.B. Hook, R.R. Regal, Capture-recapture methods in epidemiology: methods and limitations, *Epidemiol. Rev.* 17 (1995) 243–264.
- [25] P. Bernillon, L. Lievre, J. Pillonel, A. Laporte, D. Costagliola, Groupe d'épidémiologie clinique des CISH, Estimation de la sous-déclaration des cas de Sida en France par la méthode de capture-recapture, *BEH* 97 (1997) 19–21.
- [26] X. Deparis, B. Pascal, D. Baudon, Evaluation of the completeness of the epidemiological surveillance system for malaria by the capture-recapture system in the French armies in 1994, *Trop. Med. Int. Health* 2 (1997) 433–439.
- [27] D.J. Nanan, F. White, Capture-recapture: reconnaissance of a demographic technique in epidemiology, *Chronic Dis. Can.* 18 (1997) 144–148.
- [28] K. Tocque, M.A. Bellis, N.J. Beeching, P.D. Davies, Capture-recapture as a method of determining the completeness of tuberculosis notifications, *Commun. Dis. Public Health* 4 (2001) 141–143.
- [29] Y. Ding, S.E. Feinberg, Multiple sample estimation of population and census undercount in the presence of matching errors, *Surv. Methodol.* 22 (1996) 55–64.
- [30] J. Iñigo, A. Arce, J.M. Martin-Moreno, R. Herruzo, E. Palenque, F. Chaves, Recent transmission of tuberculosis in Madrid: application of capture-recapture analysis to conventional and molecular epidemiology, *Int. J. Epidemiol.* 32 (2003) 763–769.
- [31] A. Chao, P.K. Tsay, S.H. Lin, W.Y. Shau, D.Y. Chao, Tutorial in biostatistics: the application of capture-recapture models to epidemiological data, *Stat. Med.* 20 (2001) 3123–3157.
- [32] Mathsoft, *S-Plus User's Manual*, Version 4.0, Mathsoft Inc, Seattle, 1997.
- [33] S.E. Fienberg, The multiple-recapture census for closed populations and incomplete 2k contingency tables, *Biometrika* 59 (1972) 591–603.
- [34] E.B. Hook, R.R. Regal, Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence, *Am. J. Epidemiol.* 142 (1995) 48–52.
- [35] A. Infuso, B. Hubert, J. Etienne, Under-reporting of legionnaires' disease in France: the case for more active surveillance, *Euro. Surveill.* 3 (1998) 48–50.

**« Combiner analyse spatiale et épidémiologie pour l'aide à la décision dans la lutte contre la tuberculose en Guyane française »**

Le travail de thèse vise à comprendre les processus et les dynamiques spatiales de la tuberculose en Guyane, en particulier sur l'île de Cayenne, et ce par l'identification des facteurs de risque liés à l'émergence et/ou à la persistance de la maladie sur plusieurs années (1996-2003).

La caractérisation des agrégats spatiaux de cas de tuberculose est combinée à une analyse phylogéographique des souches de *Mycobacterium tuberculosis* isolées en Guyane afin de caractériser les grappes génétiques de cas, et de déterminer les circuits de transmission de la maladie.

Un Système d'Information Géographique permet ensuite de cartographier les populations à risque de tuberculose sur l'île de Cayenne. Cette carte est produite en couplant la cartographie des cas de tuberculose à un indice de vulnérabilité de la population, basé sur l'analyse qualitative de l'environnement urbain.

Cette étude révèle certains patrons de transmission de la tuberculose sur l'île de Cayenne, apportant à la fois des réponses pour les politiques de santé en Guyane et de nouvelles perspectives de recherche.

**Mots-clés :** épidémiologie moléculaire, analyse spatiale, cartographie du risque, dynamique de transmission, tuberculose, Guyane française

---

**« Combining spatial analysis and epidemiology to improve decision-making in the fight against tuberculosis in French Guiana »**

This thesis work aims to understand the processes and spatial dynamics of human tuberculosis in the Ile-de-Cayenne, French Guiana. This is done by the identification of factors related to the diffusion and the persistence of the disease over several years (1996-2003).

The spatial cluster analysis of tuberculosis cases is combined to a phylogeographic analysis of *Mycobacterium tuberculosis* strains isolated in French Guiana. This analysis dealt with the description of genetic and spatial clusters, as well as the routes of transmission responsible for the maintenance of the disease.

A Geographic Information System is then used to map tuberculosis risk among populations of the Ile-de-Cayenne. This map is established by coupling the distribution of tuberculosis cases with a vulnerability index of populations, derived from a qualitative analysis of urban landscape.

This study reveals some transmission patterns of tuberculosis in the Ile-de-Cayenne, providing support for the development of health policies in French Guiana, and offers perspectives on issues for future research.

**Keywords :** molecular epidemiology , spatial analysis, risk maps, transmission dynamics, tuberculosis, French Guiana

---

**Ecole Doctorale :** Sciences de l'Homme et de la Société

**Discipline :** « Environnement et Santé »

**Laboratoires d'accueil :** (1) Unité IRD S140, Expertise et SPAtialisation des Connaissances en Environnement, Montpellier, (2) Unité 2724 IRD-CNRS, Génétique et Evolution des Maladies Infectieuses, Montpellier.