

Université PARIS EST-CRETEIL

Université PARIS SUD



Ecole nationale Vétérinaire d'Alfort

MASTER 2<sup>ème</sup> ANNEE

Santé publique Paris Sud-Saclay et Santé UPEC

Dominante

**SURVEILLANCE EPIDEMIOLOGIQUE DES MALADIES  
HUMAINES ET ANIMALES**

---

# RAPPORT DE STAGE

L'épidémiologie moléculaire de la tuberculose bovine – Rôle de la faune sauvage  
et des bovins dans la transmission dans les Pyrénées-Atlantiques et les Landes

Présenté par

Hélène DUAULT

Réalisé sous la direction de : Laetitia CANINI et Lorraine MICHELET

Organisme et pays : Unité Epidémiologie, Laboratoire de Santé Animale, Anses - France

Période du stage : 6 janvier 2020 au 3 juillet 2020

Date de soutenance : 24 juin 2020

Année universitaire 2019-2020

## Remerciements

*A Laetitia Canini et Lorraine Michelet, pour leurs conseils, enseignements et leur patience,*  
*A Maud Marsot, pour ses encouragements, son soutien et sa gestion des accidents improbables,*  
*A Benoît Durand, pour ses conseils et sa disponibilité,*  
*A toute l'équipe d'EPI et d'EPI-MAI, pour leur accueil, leur ambiance et leur soutien inoubliables,*  
*Et à toute l'équipe d'UZB, pour leur accueil chaleureux et leur travail de manipulation indispensable.*

## Résumé court

La tuberculose bovine est une maladie chronique bactérienne réglementée à l'échelle européenne. La France est reconnue officiellement indemne depuis 2001. Cependant, des foyers persistent notamment dans le Sud-Ouest de la France où des cas sont détectés à la fois chez les bovins et la faune sauvage. L'objectif de notre étude était d'identifier les rôles joués par les blaireaux et les sangliers dans la transmission de *Mycobacterium bovis* dans une zone d'étude de 3754 km<sup>2</sup>. Pour ce faire, nous avons analysé des données de séquençage complets de 220 souches (174 de spoligotype F7 et 46 de spoligotype F15) issues de la surveillance de la tuberculose bovine dans les Pyrénées-Atlantiques et les Landes et isolées chez des bovins, des blaireaux et des sangliers. Les souches isolées chez des blaireaux et des bovins ont été analysées par inférence bayésienne. Aucun modèle de transition inter-espèce n'a été privilégié par un facteur comparatif appelé facteur de Bayes (BF). Le modèle de transition asymétrique prédisait un taux de transition de blaireau à bovin 6.7 fois supérieur au taux de transition de bovin à blaireau et le modèle symétrique prédisait un taux de transition très faible entre ces deux populations. Le rôle de la faune sauvage dans la transmission n'a donc pas pu être élucidé. Cependant les phénomènes de transition du blaireau au bovin restent relativement rares. Cela pourrait évoluer à terme si la prévalence de la maladie augmente dans la faune sauvage.

## Table des matières

Remerciements .....	2
Résumé court .....	3
Table des matières .....	4
Liste des tableaux .....	6
Liste des figures .....	7
Résumé long .....	9
Synthèse bibliographique .....	13
I) La tuberculose bovine en France.....	13
a. Modes de transmission et facteurs de risque .....	14
b. Systèmes de surveillance et contrôle de la tuberculose bovine en France .....	14
i. Chez les bovins .....	14
ii. Dans la faune sauvage .....	15
II) Séquençage du génome de <i>Mycobacterium bovis</i> .....	16
a. Méthodes utilisées .....	16
b. Caractéristiques génétiques des souches en Pyrénées Atlantiques/Landes .....	16
III) Phylogénétique moléculaire : les principes.....	17
a. Définition d'un arbre phylogénétique.....	17
b. Modèles évolutifs : reconstruction d'un arbre phylogénétique .....	18
i. Modèles de substitution.....	18
ii. Horloge moléculaire .....	18
iii. Modèles de population .....	19
c. Covariables ou traits.....	19
IV) Inférence phylogénétique .....	20
a. Méthode de parcimonie.....	20
b. Inférence bayésienne .....	21
c. Sélection du modèle.....	22
Travail personnel.....	23
I) Contexte .....	23
II) Matériel et Méthode .....	23
a. Sélection des souches.....	23
b. Séquençage et sélection des SNP.....	25
c. Données géographiques.....	25
d. Analyse phylogénétique .....	26

i.	Maximum de parcimonie .....	26
ii.	Inférence bayésienne .....	26
III)	Résultats .....	27
a.	Souches sélectionnées dans la zone d'étude .....	27
b.	Données de séquençage.....	29
c.	Données géographiques.....	29
i.	Spoligotype F15 .....	29
ii.	Spoligotype F7 .....	30
d.	Analyse par méthode de parcimonie .....	32
i.	Spoligotype F15 .....	32
ii.	Spoligotype F7 .....	34
e.	Inférence bayésienne .....	35
V)	Discussion .....	41
VI)	Conclusion .....	44
	Bibliographie.....	45
	Liste des annexes.....	49

## Liste des tableaux

Tableau 1 : Interprétation du facteur de Bayes selon sa valeur .....	23
Tableau 2 : Modèles testés avec BEAST. Le modèle de diffusion spatial discret est détaillé en Annexe 10.....	27
Tableau 3 : Répartition des souches échantillonnées dans la zone d'étude en fonction de leur spoligotype (F15 ou F7) et de l'espèce hôte. ....	28
Tableau 4 : Caractéristiques des clades pour le spoligotype F15 : espèce hôte, profondeur temporelle et nombre de communes concernées.....	32
Tableau 5 : Caractéristiques des clades pour le spoligotype F7 : espèce hôte, profondeur temporelle et nombre de communes concernées.....	34
Tableau 6 : Probabilité que l'espèce hôte soit un bovin et probabilité postérieure du MRCA de chaque clade et de la racine de l'arbre consensus. ASYM correspond au modèle asymétrique et SYM au modèle symétrique .....	36
Tableau 7 : Paramètres de sorties pour chaque modèle (ASYM : asymétrique et SYM : symétrique) avec leur moyenne, intervalle de crédibilité à 95% et leur ESS ainsi que les paramètres calculés à partir de ces sorties. ....	38

## Liste des figures

Figure 1 : Evolution de la prévalence et de l'incidence de la tuberculose bovine de 1995 à 2017 (Delavenne et al., 2019). .....	13
Figure 2 : Répartition géographique des foyers incidents de tuberculose bovine de 2015 à 2017 en France (Delavenne et al., 2019).....	14
Figure 3 : Comparaison de la topologie entre un arbre enraciné (à droite) et l'arbre non enraciné dont il est issu (à gauche). .....	17
Figure 4 : Reconstruction par processus coalescent de l'arbre associé à un échantillon de sept individus issus d'une population de taille constante. A gauche, la représentation schématique de l'évolution de la population dans le cas d'une population de taille constante. Chaque ligne représente une génération de la population et chaque rond, un individu. Les individus échantillonnés sont représentés en vert, leurs ancêtres en bleu, les ancêtres communs à au moins deux individus en jaune et le MRCA en orange. A droite, l'arbre enraciné généré à partir de ces sept échantillons.....	19
Figure 5 : Deux arbres phylogénétiques A et B réalisés à partir des mêmes taxa définis par une séquence de trois nucléotides. Le nombre de substitutions nécessaires ( $N_s$ ) entre deux nœuds est indiqué à chaque branche. Le score de parcimonie de chaque arbre peut être calculé en sommant les $N_s$ . Le score de parcimonie de l'arbre A ( $\sum(N_{sA})=1+0+0+1+1+0=3$ ) est de 3 et celui de B ( $\sum(N_{sB})=1+0+1+0+2+1=5$ ) est de 5. La méthode de parcimonie privilégiera donc l'arbre A. ....	21
Figure 6 : Représentation schématique de l'approximation de la distribution a posteriori d'un vecteur de paramètres $\theta$ par une chaîne de Markov Monte Carlo de longueur 12. A gauche, la distribution a priori d'un vecteur $\theta$ contenant un seul paramètre qui suit une loi de distribution uniforme. Au centre, la représentation du processus d'acceptation ou de rejet d'un état (i.e. d'une valeur de $\theta$ ). A droite, la distribution a posteriori constituée par la chaîne en considérant que l'on échantillonnait à chaque étape de la chaîne (d'après Lee et al., 2015).....	22
Figure 7 : Représentation de la zone d'étude en bleu (à gauche) au sein des Pyrénées-Atlantiques dans le Sud-Ouest de la France (à droite) .....	24
Figure 8 : Résumé de l'obtention des données géographiques par espèce.....	26
Figure 9 : Nombre de souches isolées de spoligotype F7 (en A) et F15 (en B) au cours du temps en fonction des espèces hôtes. ....	28
Figure 10 : Arbre phylogénétique des souches de spoligotype F15 construit par maximum de parcimonie sur le logiciel Bionumerics mettant en évidence une souche très éloignée des autres et dont le profil VNTR diffère (6 5 5 3 11 2 4s 7 vs. 6 5 5 3 11 2 4s 8).....	29
Figure 11 : Géolocalisation des cas de <i>M. bovis</i> de spoligotype F15 (A) et de spoligotype F7 (B) dans la zone d'étude. (A) En vert, on retrouve les 2 souches isolées chez des sangliers, en violet chez des blaireaux (9 souches) et en orange, chez des bovins (37 souches). Les souches éliminées car les critères de qualité n'étaient pas remplis, sont marquées par une étoile noire ; celle éliminée car trop éloignée des autres, par un astérisque rose. (B) En vert, on retrouve les 6 souches isolées chez des sangliers, en violet chez des blaireaux (21 souches) et en orange, chez des bovins (147 souches). L'astérisque rose marque l'emplacement de la souche éliminée lors de la sélection des souches. ....	31

Figure 12 : Arbre phylogénétique des souches de spoligotype F15 construit par méthode de parcimonie. Les lettres A, B, C et D correspondent au MRCA de chaque clade. .... 32

Figure 13 : Répartition géographique des clades A, B, C et D, des souches de spoligotype F15, identifiés à partir de l'arbre de maximum de parcimonie. Les communes où se trouve au moins une souche appartenant au clade identifié par la lettre à gauche de la figure sont colorées en gris. .... 33

Figure 14 : Arbre phylogénétique des souches de spoligotype F7 construit par méthode de parcimonie. Les lettres A, B, C et D correspondent au MRCA de chaque clade. .... 34

Figure 15 : Arbres enracinés consensus (en (A) du modèle asymétrique et en (B), du modèle symétrique) des souches de spoligotype F7 (inférence bayésienne) selon le maximum de crédibilité de clade. Les nœuds internes dont la probabilité de l'espèce était inférieure à 0,70 sont indiqués comme incertains. En (A), le clade A1 (en vert foncé) comporte 88 souches, le clade B1 (en rose) 28 souches, le clade C1 (en vert pâle) 22. Il reste 29 souches n'appartenant à aucun clade. En (B), le clade A2 (en jaune) comporte 14 souches que l'on retrouve dans le cercle jaune en (A), dans le clade B2 (en vert foncé) 15 souches sont retrouvées dans le clade A1 (en A, entourées en vert), le clade C2 (en vert pâle), 22 souches est le même que le clade C1 (A) et de même pour le clade D2 et B1 (en rose), 28 souches. Il reste 88 souches n'appartenant à aucun clade..... 37

Figure 16 : Répartition géographique des clades A, B et C et des souches sans clade, identifiés à partir de l'arbre du modèle de transition asymétrique. Les communes où se trouve au moins une souche appartenant au clade, identifié par la lettre en haut à gauche de la figure, sont colorées en gris. .... 39

Figure 17 : Répartition géographique des clades A, B, C et D et des souches sans clade, identifiés à partir de l'arbre du modèle de transition symétrique. Les communes où se trouve au moins une souche appartenant au clade, identifié par la lettre en haut à gauche de la figure, sont colorées en gris. .... 40

## Résumé long

*Mycobacterium bovis* est l'agent le plus fréquent de la tuberculose bovine en France et peut à la fois infecter la faune domestique (bovins, porcins, caprins, ovins) et sauvage (blaireaux, sangliers, cervidés et renards). C'est un agent zoonotique mais de nos jours, les cas chez l'homme sont rares en France. La tuberculose bovine est règlementée à l'échelle européenne et le statut indemne, que la France possède depuis 2001, facilite le commerce d'animaux vivants au sein de l'Union Européenne. Cependant, des foyers persistent dans certaines régions comme le Sud-Ouest de la France et menacent le statut indemne.

Dans ces foyers, des cas sont retrouvés à la fois chez les bovins et dans la faune sauvage (blaireaux, sangliers, cervidés). Dans un système multi-hôte, chaque espèce peut avoir un rôle épidémiologique différent dans la transmission voire la persistance de cet agent pathogène. Pour mettre en place des mesures de lutte efficaces, il est donc indispensable de bien comprendre le rôle joué par chaque espèce.

Depuis 2011, le programme de surveillance Sylvatub suit l'évolution de la tuberculose bovine chez le cerf élaphe (*Cervus elaphus*), le sanglier (*Sus scrofa*), le chevreuil (*Capreolus capreolus*) et le blaireau (*Meles meles*). Les modalités de mise en œuvre de ce programme varient (du ramassage de carcasses à la mise en place d'opérations de piégeages) selon la situation épidémiologique du département. Dans les élevages de bovins, un dépistage régulier dont la périodicité dépend du statut à risque ou non de l'élevage a lieu d'octobre à avril. De plus, les carcasses de bovins sont systématiquement inspectées en abattoir.

L'épidémiologie moléculaire est un outil qui permet d'investiguer des maladies infectieuses à partir de marqueurs moléculaires du pathogène (ADN par exemple). Ces marqueurs moléculaires sont notamment utilisés pour reconstruire des arbres phylogénétiques qui présentent les liens de parenté entre les souches isolées chez des animaux infectés.

Dans les zones de forte incidence, les méthodes de typage classiques comme le spoligotypage et le typage VNTR, ne suffisent pas à faire la différence entre les souches. Par exemple, dans les Pyrénées-Atlantiques pour 224 souches isolées, 2 spoligotypes (F15 et F7) et 10 profils VNTR ont été identifiés. L'identification de mutations ponctuelles ou « single nucleotide polymorphism » (SNP) obtenue grâce au séquençage complet du génome des souches et permettant d'avoir une résolution plus fine, s'avère alors nécessaire.

Il existe de nombreuses méthodes de reconstruction d'arbres phylogénétiques. Une des plus simples et rapides est celle du maximum de parcimonie, elle permet de regrouper les souches selon le nombre de substitutions qui les séparent. Elle ne permet cependant pas de prendre en compte les données épidémiologiques telles que l'espèce hôte ou encore la localisation géographique des souches.

La reconstruction d'arbres phylogénétiques par inférence bayésienne est plus complexe et nécessite de renseigner les modèles de substitution, d'horloge (*i.e.* strict ou relâchée) et de population dans la distribution *a priori*. De plus, cette méthode permet de prendre en compte des caractéristiques épidémiologiques comme l'espèce-hôte ou encore des données géographiques et ainsi d'estimer des paramètres comme le taux de transition inter-espèce ou la diffusion spatiale. Il est possible de sélectionner le modèle qui correspond le mieux aux données en comparant deux à deux les modèles grâce au facteur de Bayes (BF).

L'approche bayésienne a déjà été utilisée en épidémiologie moléculaire pour étudier la transmission de la tuberculose bovine dans des systèmes multi-hôtes en Dordogne, en Nouvelle-Zélande, au Royaume-Uni et aux Etats-Unis.

Dans notre étude, l'objectif était d'élucider, à partir des données génomiques, les rôles joués par les blaireaux et les sangliers dans la transmission de la tuberculose bovine au sein d'une zone d'étude de 3754 km<sup>2</sup> située dans les Pyrénées Atlantiques et les Landes.

Après vérification de la qualité du séquençage, nous avons inclus 174 souches de spoligotype F7 (isolées chez 146 bovins, 21 blaireaux, 6 sangliers et un porc) et 46 de spoligotype F15 (isolées chez 34 bovins, 9 blaireaux, 2 sangliers et un humain).

Les données géographiques ont mis en évidence que la distance entre un blaireau et le bovin le plus proche était inférieure à 5 km (maximum de 4,4 km pour le spoligotype F15 et 2,6 km pour F7). Ceci est compatible avec un phénomène de transmission entre les deux espèces sachant que les données de mouvements de blaireaux en Europe estime qu'une majorité de déplacements est inférieure à 5 km.

Dans un premier temps, toutes les espèces hôtes ont été considérées. La reconstruction des arbres phylogénétiques par maximum de parcimonie a permis de mettre en évidence la présence de plusieurs clades comportant à la fois des bovins, des blaireaux et des sangliers. Ces résultats suggèrent que plusieurs phénomènes de transmission pourraient avoir eu lieu entre ces espèces mais nous ne pouvons conclure sur le sens des échanges ni quelle(s) espèce(s) ont été infectée(s) par ou ont infecté des bovins.

Dans l'approche bayésienne, seules les souches de spoligotype F7 isolées chez des bovins et des blaireaux ont été étudiées. Les données bibliographiques ont permis de faire le choix d'un modèle de substitution HKY. D'autre part, le BF a permis la sélection d'un modèle d'horloge moléculaire stricte et d'un modèle de population constant mais pas celle du modèle de transition inter-espèce.

Un modèle de diffusion spatiale discret a été testé pour essayer de mettre en évidence une diffusion spatiale mais celle-ci n'a pas pu être démontrée. Ceci pourrait s'expliquer par le fait que la maladie est d'évolution lente et que notre zone d'étude est relativement restreinte. L'échelle de temps voire la taille de la zone pourrait ne pas être suffisant pour mettre en évidence cette diffusion, si elle existe. L'absence de prise en compte de mouvements de bovins et des animaux de la faune sauvage dans le modèle pourrait également contribuer à ce résultat.

Le taux de substitution de 0,35 substitutions/génome/an (HPD : Highest posterior density à 95% : [22,40 ; 41,38] vs. [22,17 ; 40,61]) estimé par le modèle est cohérent avec les estimations d'autres études concernant *M. bovis*.

La méthode de migration nous a permis d'estimer des taux de transition inter-espèce qui représente la probabilité qu'une souche passe d'une espèce hôte à une autre au cours d'une année. Avec le modèle symétrique, on prédit un taux de transition de 0,03 transition par souche et par an. A l'inverse, avec le modèle asymétrique on prédit un taux de transition de blaireau à bovin de 0,40 transition par souche et par an 6.7 fois supérieur au taux de transition de bovin à blaireau de 0,06 transition par souche et par an. En Dordogne, le modèle avec une transition asymétrique entre trois espèces a prédit des taux de transition plus élevés variant entre 0,27 transitions/souche/an (du bovin au sanglier) et 1,5 transitions/souche/an (du sanglier au blaireau, résultats non publiés) avec un rôle prédominant du sanglier.

Dans notre approche, l'arbre phylogénétique qui résume les résultats est l'arbre de maximum de crédibilité de clade. Dans notre étude, cet arbre prédit quatre événements de transition du blaireau au bovin pour le modèle asymétrique et un seul événement pour le modèle symétrique.

Dans le modèle asymétrique (vs. symétrique), l'ancêtre commun le plus récent aurait circulé en 1985 (vs. 1986) (HPD à 95% : [1975 ; 1995] vs. [1976 ; 1995]) et l'espèce-hôte de cet ancêtre est incertaine (vs. un bovin). Dans l'arbre consensus, on peut définir 3 (vs. 4) clades de tailles variables (22 à 88 souches vs. 14 à 28 souches). Cependant, 29 souches isolées uniquement chez des bovins (vs. 88 souches isolées à la fois chez des bovins et des blaireaux) n'ont pu être classées dans un clade. Nous avons pu identifier deux clades identiques entre les deux arbres.

Il serait intéressant de pouvoir inclure des données plus récentes dans la reconstruction des arbres phylogénétiques de la région. De plus, contrairement aux arbres phylogénétiques qui représentent la proximité des souches dans l'histoire évolutive, les arbres de transmission permettent de reconstituer l'histoire de transmission entre les différentes populations (bovins, blaireaux et sangliers). Ces arbres de transmission prendraient en compte les nouvelles données phylogénétiques ainsi que l'inclusion de données épidémiologiques (estimation de la date d'infection probable par exemple). Ils permettraient d'évaluer le rôle des deux espèces de la faune sauvage et de modéliser la diffusion spatiale de la maladie dans la zone.

En conclusion, bien qu'on ne puisse pas déterminer le rôle exact du blaireau et qu'on ne connaisse pas du tout celui du sanglier, quel que soit le modèle de transition inter-espèce, les phénomènes de transition du blaireau au bovin restent relativement rares du fait du faible nombre de blaireaux infectés. Cela pourrait évoluer à terme si la prévalence de la maladie augmente dans la faune sauvage.



## Synthèse bibliographique

### I) La tuberculose bovine en France

La tuberculose bovine (bTB) touche principalement les bovins mais peut aussi infecter la faune sauvage (blaireaux, sangliers, cervidés et renards), d'autres animaux domestiques (porcins, caprins, ovins...) et dans de rares cas, l'homme. La bTB est une maladie bactérienne chronique à déclaration obligatoire en France et règlementée à l'échelle européenne. Le statut indemne permet de déroger au dépistage lors des échanges commerciaux de bovins vivants au sein de l'Union Européenne (Directive 64/432/CEE).

En 2001, la France a obtenu le statut officiellement indemne de tuberculose bovine (Directive 2003/467/CE). Les critères d'obtention du statut indemne sont une prévalence annuelle des troupeaux infectés inférieure à 0,1% pendant six ans et un taux de troupeaux officiellement indemnes au 31 décembre de chaque année, supérieur à 99,9% pendant six ans. Malgré le fait que l'incidence de la bTB reste stable (0,05% des troupeaux bovins depuis 2012), sa prévalence quant à elle continue d'augmenter faiblement pour atteindre 0,10% en 2017 (Figure 1). Ceci est une situation nationale préoccupante, le statut indemne étant menacé.

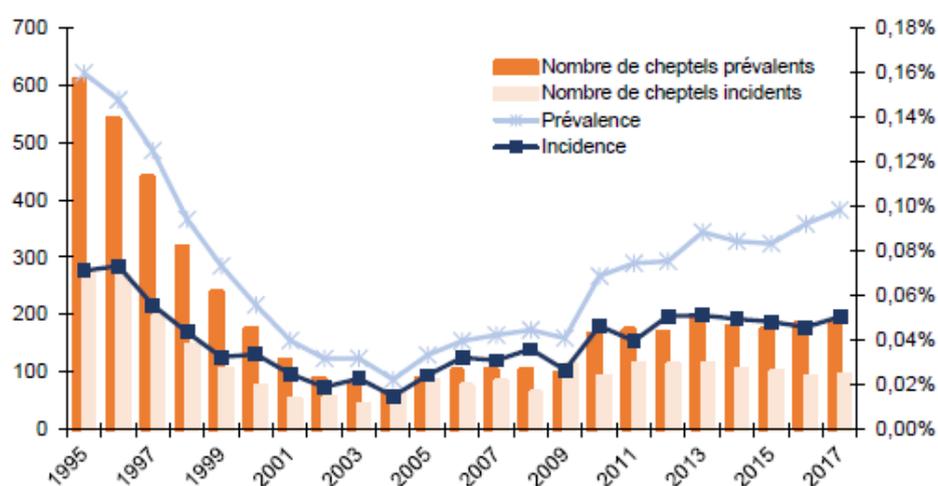
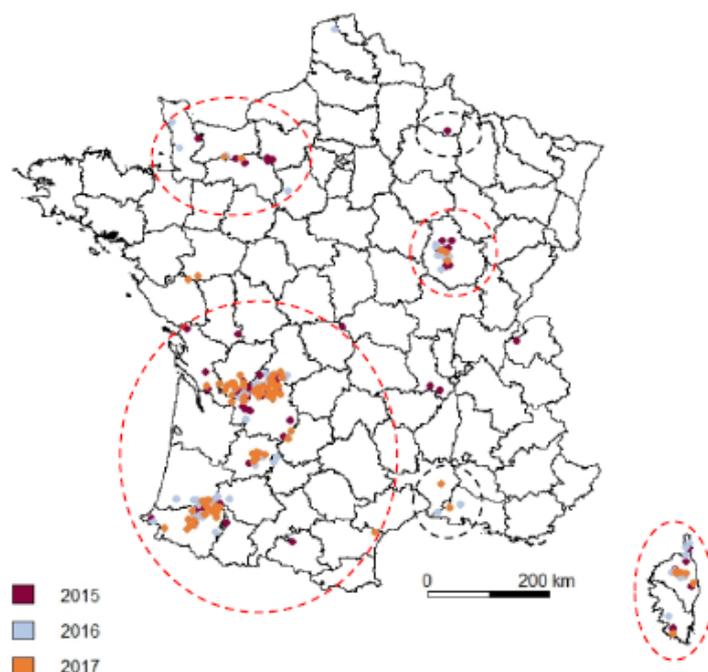


Figure 1 : Evolution de la prévalence et de l'incidence de la tuberculose bovine de 1995 à 2017 (Delavenne et al., 2019).

Quatre zones concentrent la majorité des foyers : le Sud-Ouest (avec d'un côté une zone en Dordogne, Charente et Haute-Vienne, et de l'autre, une zone dans les Pyrénées-Atlantiques et les Landes), la Corse, la Côte-d'Or et la Normandie (Figure 2). En Normandie, le nombre de foyers incidents détectés est en diminution depuis 2000 ; cette tendance est également observée en Côte-d'Or depuis 2010. Dans le Sud-Ouest, on observe au contraire une expansion régulière des zones d'infection. En Corse, l'absence de système de contention efficace chez la plupart des éleveurs rend l'évaluation de la situation réelle difficile (Delavenne et al., 2019).



*Figure 2 : Répartition géographique des foyers incidents de tuberculose bovine de 2015 à 2017 en France (Delavenne et al., 2019).*

#### a. Modes de transmission et facteurs de risque

La répartition des lésions tuberculeuses majoritairement dans l'appareil respiratoire (poumons et nœuds lymphatiques associés) observée en autopsie et des études expérimentales réalisées chez des veaux ont permis d'identifier l'inhalation d'aérosols comme principale voie d'infection chez les bovins (O'Reilly et Daborn, 1995).

Les facteurs de risque de la tuberculose bovine à l'échelle d'un troupeau varient selon les études, les régions et le contexte épidémiologique : mouvement de bovins, présence de bTB dans les élevages à proximité, taille du cheptel voire la proximité de la faune sauvage (blaireaux), pour en citer quelques-uns (Skuce *et al.*, 2012). Une étude cas-témoin dans les Ardennes, la Côte-d'Or et la Dordogne a permis d'identifier plusieurs facteurs de risque en France : la proximité avec un foyer récent et la présence d'un bâtiment (stockant de la nourriture ou hébergeant des bovins) éloigné des zones d'habitation. L'étude met en relation ce dernier facteur avec la possibilité de contacts avec, et donc de contamination par, la faune sauvage (Marsot *et al.*, 2016).

#### b. Systèmes de surveillance et contrôle de la tuberculose bovine en France

##### i. Chez les bovins

Les bovins à l'abattoir subissent systématiquement une inspection *post mortem* (arrêté du 15 septembre 2003). Si des lésions sont détectées, des prélèvements d'organes et de leurs nœuds lymphatiques associés ont lieu en vue d'une recherche de mycobactéries par détection d'une séquence d'ADN spécifique (« Polymerase Chain Reaction » ou PCR) et par culture. La sensibilité de cette surveillance a été estimée à 28,5% et sa spécificité à 100% (Keck *et al.*, 2014).

En élevage, un dépistage par intradermo-tuberculation (IDT), dont la périodicité est fonction de la situation sanitaire du département et du statut à risque ou non de l'élevage, a lieu d'octobre à avril (arrêté du 15 septembre 2003, Annexe 1) et lors de l'introduction d'un bovin à risque (Annexe 2). Ce dépistage est effectué à l'aide d'un dérivé protéique purifié, la tuberculine, inoculé en région cervicale. Après l'inoculation de la tuberculine, une réaction d'hypersensibilité retardée (lecture après

72 heures) est mise en évidence chez un sujet ayant été en contact avec la bactérie. L'IDT peut être simple (IDS), utilisation de la tuberculine bovine seule, ou comparée (IDC), utilisation conjointe de la tuberculine aviaire. Elle peut être accompagnée selon le contexte, par un dosage sanguin de la libération in vitro d'interféron gamma (IFN- $\gamma$ ) après stimulation des lymphocytes T à la tuberculine (Keck et al., 2014).

La sensibilité (Se) et la spécificité (Sp) varient selon les techniques (Keck *et al.*, 2014):

- IDS : Se [min : 58,6%- max : 91%] et Sp [min : 75%- max : 99,7%],
- IDC : Se d'environ 50% et Sp supérieure à 99,9%,
- IFN : Se [min : 82%- max : 100%] et Sp [min : 88% - max : 99%].

La gestion d'un élevage suspect dépend du niveau de suspicion, lui-même déterminé par une analyse de risque qui tient compte de critères épidémiologiques. Deux cas de figure existent, soit un abattage diagnostique des animaux réagissant est réalisé, soit un contrôle par IFN est possible. Une recherche de mycobactéries par PCR et par culture est réalisée après chaque abattage diagnostique. La culture de mycobactéries étant très lente (1 à 3 mois), le délai pour infirmer une suspicion peut s'avérer long et avoir des conséquences néfastes sur la pérennité de l'élevage bloqué.

Lorsqu'une suspicion est confirmée, l'abattage total est recommandé en vue d'assainir le troupeau ; même s'il existe une dérogation possible pour un abattage sélectif, où les animaux sont testés par IDT et IFN et seuls les animaux réagissant sont abattus. Le temps de récupération de la qualification indemne est long et nécessite trois contrôles négatifs espacés d'un minimum de deux mois et un nettoyage-désinfection (Delavenne *et al.*, 2019).

## ii. Dans la faune sauvage

Depuis 2011, un programme de surveillance appelé Sylvatub créé dans le cadre de la Plateforme d'Epidémiosurveillance en Santé Animale, permet de suivre l'évolution de la bTB dans la faune sauvage. Les espèces sauvages concernées par ce programme sont le cerf élaphe (*Cervus elaphus*), le sanglier (*Sus scrofa*), le chevreuil (*Capreolus capreolus*) et le blaireau (*Meles meles*) (Rivière *et al.*, 2013). De plus, une étude est actuellement en cours en Dordogne pour déterminer le rôle des renards (*Vulpes vulpes*), suite à la découverte de 4 renards infectés en 2015 (Michelet *et al.*, 2018).

Ce dispositif repose sur la surveillance événementielle (ex : sangliers collectés lors de la chasse), la surveillance événementielle renforcée (ex : blaireaux retrouvés morts en bord de route) et la surveillance programmée (ex : piégeages de blaireaux) (Annexe 3). La mise en œuvre de ces modalités dépend du niveau de surveillance défini à l'échelle d'un département (Annexe 4 et Annexe 5).

Les animaux sauvages abattus ou retrouvés morts sont autopsiés et des prélèvements sont réalisés en vue d'une recherche de mycobactéries par PCR et culture (Réveillaud, 2013). En cas de culture ou de PCR positive, un typage moléculaire est réalisé (cf. Partie II).

Du fait de la dégradation des carcasses dans l'environnement et des possibles contaminations bactériennes, les prélèvements réalisés dans le cadre de Sylvatub sont généralement de moins bonne qualité que ceux effectués chez les bovins en abattoir. D'après des avis d'experts, la culture bactérienne réalisée à partir de ces prélèvements aurait une sensibilité diminuée de 35% (Rivière *et al.*, 2015).

La surveillance de la faune sauvage est complémentaire de celle effectuée dans les élevages, la majorité des cas détectés ayant pu être mise en relation avec la présence de l'infection chez des bovins (similitude des types moléculaires et zones géographiques) (Desvaux *et al.*, 2019).

## II) Séquençage du génome de *Mycobacterium bovis*

Trois espèces de *Mycobacterium* peuvent être responsables de la tuberculose chez les bovins : *Mycobacterium tuberculosis*, *Mycobacterium caprae* et enfin la plus fréquente en France, *Mycobacterium bovis* (Michelet *et al.*, 2019). Le génome de *M. bovis* comprend environ 4,3 million de paires de bases (pb) parmi lesquelles on compte un fort pourcentage (environ 65%, (Garnier *et al.*, 2003)) de bases guanines (G) ou cytosines (C).

Du fait de la reproduction asexuée de *M. bovis* et de l'absence d'échanges de matériel génétique entre mycobactéries, elles sont clonales et peu variables dans le temps (Michelet *et al.*, 2019). Il est donc intéressant de procéder à une combinaison de plusieurs méthodes de génotypage pour augmenter la discrimination entre les souches de *M. bovis*.

### a. Méthodes utilisées

Le spoligotypage est une technique de typage basée sur le polymorphisme d'une région conservée de l'ADN, le locus Direct Repeat (DR). Les DR sont de courtes séquences de nucléotides fréquemment répétées dans le génome (par exemple : **TTAAC** dans la séquence 5' **TTAAC**ggttacatctat**TTAAC**tg 3'). Le génome de *M. bovis* possède 43 espaceurs (séquences de nucléotides situées entre 2 DR) dont la position est fortement conservée. Leur présence ou absence est codée de manière binaire et caractérise la souche. *M. bovis* ne possède jamais les espaceurs 3, 9, 16 et 39 à 43 (Kamerbeek *et al.*, 1997).

Le typage VNTR (« Variable Number of Tandem Repeat ») identifie le nombre de répétitions en tandem (Tandem Repeat) présentes dans des loci dispersés dans le génome bactérien. Un Tandem Repeat est une suite de motifs nucléotidiques, adjacents, qui se répètent à l'identique (exemple : TTCGA dans 5' TTCGATTCGATTCGA 3'). Les loci sont dans des régions polymorphiques et hypervariables du génome. Le nombre de répétitions associées à chaque locus est combiné pour générer un code numérique appelé profil VNTR (Annexe 6) (Roring *et al.*, 2002). Huit loci sont préconisés par le consortium européen et utilisés en France (2165, 2461, 577, 580, 2163a, 2163b, 4052 et 3232) (Michelet *et al.*, 2019).

Entre 1978 et 2013, parmi 270 souches de spoligotype SB0134 isolées en France, 53 profils VNTR différents ont été mis en évidence. Cependant, 56% de ces souches présentaient le même profil VNTR (6 4 5 3 6 4 3 6) isolé majoritairement en Côte-d'Or (Hauer, 2015). Lors d'une étude épidémiologique dans une zone de forte incidence comme la Côte-d'Or, si la majorité des souches présentent le même spoligotype et le même profil VNTR, il est difficile de faire la différence entre elles. Une méthode plus précise est nécessaire : l'identification de mutations ponctuelles ou SNP (« Single Nucleotide Polymorphism »).

Avec les avancées de la science en matière de séquençage, il est possible à l'heure actuelle de faire appel au WGS ou « whole genome sequencing ». Les étapes du WGS (fragmentation, séquençage puis alignement) permet d'obtenir la séquence nucléotidique des souches (Annexe 7) par homologie avec une séquence de référence (Hauer, 2015).

Pour *M. bovis*, selon la méthode présentée en Annexe 7, le génome complet n'est en réalité jamais obtenu du fait de nombreuses séquences répétées qui compliquent l'alignement mais le résultat du WGS permet toutefois l'identification des SNPs.

### b. Caractéristiques génétiques des souches en Pyrénées Atlantiques/Landes

En 2015, une étude portant sur plus de 4000 souches de *M. bovis* isolées dans toute la France entre 1978 et 2013, a permis l'identification d'une famille de souches, appelée famille F4, provenant principalement du Sud de la France. Dans cette famille, le spoligotype était caractérisé par l'absence

de l'espaceur 33 et le profil VNTR présentait au locus 4052 une répétition tronquée (indiquée par un « s » dans le code numérique) (Hauer, 2015). L'existence de cette famille F4 a pu être confirmée par analyse des SNP de souches circulantes en France (Hauer *et al.*, 2019).

Parmi les membres de cette famille, on retrouve un génotype présentant le spoligotype SB0821 (ou F7) avec un profil VNTR 6 5 5 3 11 2 5s 8. Ce génotype a été retrouvé à la fois chez les bovins, les blaireaux et les sangliers dans les Pyrénées-Atlantiques et les Landes (Hauer, 2015). Le fait que les mêmes types moléculaires isolés dans la faune sauvage sont également isolés chez les bovins suggère fortement un système multi-hôte faune sauvage-faune domestique (Hars et Richomme, 2010). Un système multi-hôte est un ensemble d'espèces pouvant héberger un agent pathogène et dont chaque espèce peut avoir un rôle épidémiologique différent selon son importance dans la transmission et la persistance de cet agent pathogène (Bouchez-Zacria *et al.*, 2017).

### III) Phylogénétique moléculaire : les principes.

#### a. Définition d'un arbre phylogénétique

Un arbre phylogénétique est une représentation graphique de la parenté supposée entre différents échantillons étudiés, appelés « taxa ». Ces taxa constituent les nœuds terminaux ou feuilles de l'arbre. Les nœuds internes représentent quant à eux des ancêtres communs hypothétiques (car non échantillonnés).

Les feuilles et les nœuds sont reliés entre eux par des branches et la topologie de l'arbre correspond à la forme générale de l'arbre. Un clade est un groupe constitué par un ancêtre commun et l'ensemble de ses descendants (Baum et Offner, 2008). Une lignée désigne le chemin évolutif reliant un taxa et son ancêtre hypothétique.

Les arbres non enracinés décrivent les relations évolutives entre les taxa, la longueur des branches traduit la distance génétique qui sépare les nœuds (pour son calcul, voir partie III) b.). Dans un arbre enraciné, un nœud interne correspond à l'ancêtre le plus récent commun à tous les taxa (MRCA : Most Recent Common Ancestor) et est aussi appelé racine (Figure 3). La hauteur d'un arbre enraciné permet de calculer quand circulait ce MRCA. Les arbres enracinés intègrent donc une dimension temporelle contrairement aux arbres non enracinés. Pour chaque arbre non enraciné, plusieurs arbres enracinés peuvent correspondre selon l'emplacement de la racine (Drummond et Bouckaert, 2015).

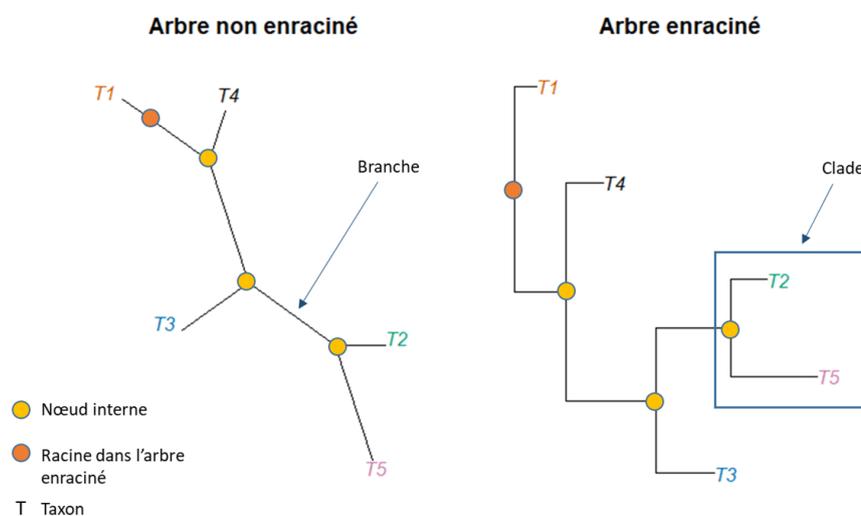


Figure 3 : Comparaison de la topologie entre un arbre enraciné (à droite) et l'arbre non enraciné dont il est issu (à gauche).

## b. Modèles évolutifs : reconstruction d'un arbre phylogénétique

### i. Modèles de substitution

La manière la plus simple de mesurer la distance entre deux séquences alignées est le nombre de sites sur lesquels elles diffèrent. Ce nombre est appelé distance de Hamming ( $h$ ) et rapporté à la taille de la séquence ( $l$ ), il donne la proportion de sites qui diffèrent ( $p=h/l$ ) (Hamming, 1950).

Cependant, l'alphabet des séquences nucléotidiques ne comporte que 4 lettres {A, C, G, T} et pour chaque site connu à un instant  $t$  seules 3 substitutions sont possibles. Certains changements peuvent ne pas être prédits (par exemple si le nucléotide est A chez un ancêtre, a muté en G puis a muté à nouveau en A chez le *taxon*, le passage par G n'est pas identifié dans le modèle) et les distances génétiques sont alors sous-estimées. Pour compenser cette tendance, il est possible de corriger l'estimation de la distance en utilisant un modèle de substitution. Ce modèle peut prendre en compte la fréquence des bases dans le génome ou encore la nature des bases impliquées dans la substitution (Drummond et Bouckaert, 2015), *i.e.* les pyrimidines (cytosine, C et thymine, T) et les purines (guanine, G et adénine, A).

Le modèle de substitution le plus simple est celui de Jukes et Cantor développé en 1969 (JC69) selon lequel toutes les bases sont présentes à la même fréquence et toutes les substitutions sont équiprobables. Dans d'autres modèles, les taux de substitution peuvent varier selon la nature de bases impliquées (Kimura, 1980), les fréquences des bases peuvent être inégales (Felsenstein, 1981) ou les deux variantes peuvent être prises en compte (Hasegawa Kishino Yano ou HKY et Generalised Time-Reversible ou GTR dans lequel les taux de substitution varient selon les bases en elles-mêmes et pas seulement leur nature) (Hasegawa *et al.*, 1985) (Annexe 8).

Le taux de substitution peut varier selon les sites du génome : par exemple, les gènes hautement exprimés ont tendance à évoluer plus lentement (Zhang et Yang, 2015). La distribution gamma discrète proposée par Yang (Yang, 1994), dont le paramètre  $\alpha$  permet de faire varier l'allure de la courbe d'une courbe en L ( $\alpha < \text{ou} = 1$ ) à une courbe en cloche ( $\alpha \gg 1$ ) et donc de modéliser des situations très variées, est souvent utilisée pour modéliser cette hétérogénéité. De plus, le fait que ces sites conservés car une mutation du gène concerné y serait délétère peut être modélisé en ajoutant une proportion de sites invariants (Gu *et al.*, 1995).

### ii. Horloge moléculaire

L'horloge moléculaire est l'hypothèse selon laquelle, les taux de substitution ne varient pas au cours du temps ni entre les lignées. On suppose un taux d'évolution constant et égal entre chaque branche. Prenons deux souches S1 et S2 correspondant à des séquences nucléotidiques de souches isolées respectivement en 2010 et 2020 et un MRCA circulant en 2000. La distance génétique entre S1 et la racine ( $d_{S1-R}$ ) sera 2 fois moins grande que la distance génétique entre S2 et la racine ( $d_{S2-R}$ ), selon l'hypothèse d'horloge moléculaire.

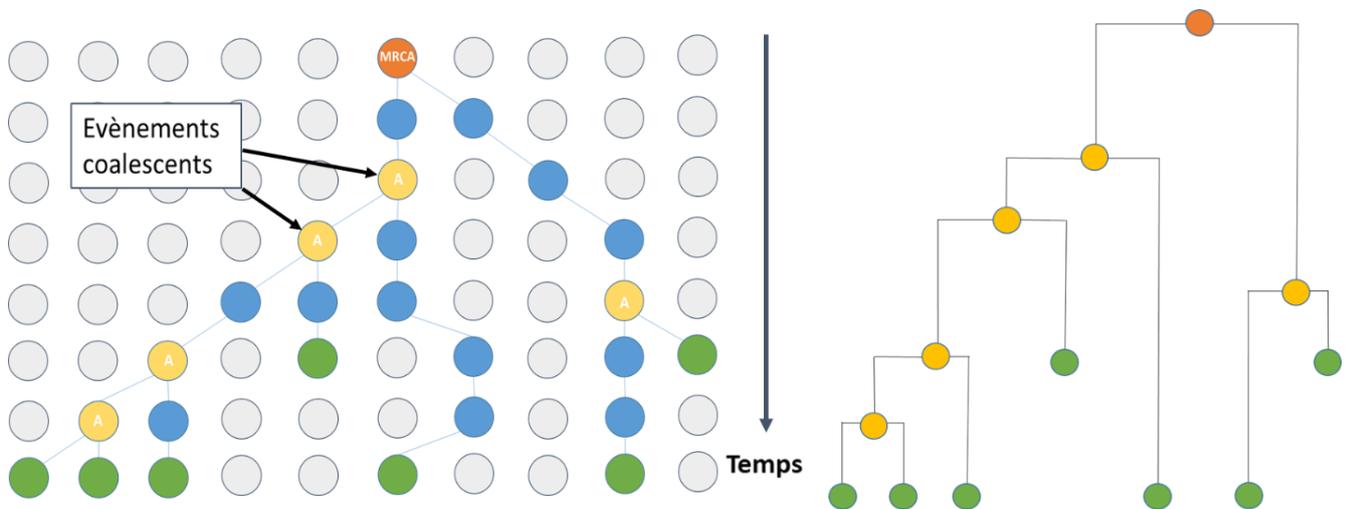
Dès 1990, des études génomiques réfutent cette hypothèse en mettant en évidence pour un même gène, des taux de substitutions différents entre différentes espèces (Li, 1993). Il est possible de relâcher ce modèle d'horloge dit strict, en introduisant des variations des taux de substitution en fonction du temps ou en fonction des lignées (dos Reis *et al.*, 2016). Les modèles d'horloge relâchée peuvent être non corrélés (le taux d'évolution le long d'une branche est indépendant de celui des autres branches) ou auto-corrélés (le taux d'évolution évolue graduellement d'une branche à l'autre, deux branches proches ont un taux d'évolution semblable) (Drummond *et al.*, 2006). Par exemple, dans les modèles non corrélés, le taux d'évolution de chaque branche est tiré au sort à partir d'une même distribution. Si la distribution choisie est exponentielle, les valeurs élevées ont une plus grande probabilité d'être tirées et inversement pour une distribution log-normale.

Les modèles d'horloges moléculaires aussi bien strictes que relâchées permettent d'intégrer une dimension temporelle aux modèles de substitution, on aboutit à la construction d'un arbre phylogénétique enraciné.

### iii. Modèles de population

Pour représenter les relations entre les différentes séquences échantillonnées (estimation de la topologie de l'arbre), des modèles de population sont utilisés. Ils déterminent la façon dont l'arbre est construit, le plus fréquemment utilisé est le modèle de coalescence.

Selon la théorie de la coalescence de Kingman, on peut reconstruire l'arbre associé à un échantillon de  $n$  séquences nucléotidiques ( $n=7$  en vert dans la [Figure 4](#)) en remontant le temps avec  $n - 1$  événements coalescents. Un événement coalescent correspond à la convergence de deux lignées, manifestée par la présence d'un ancêtre commun. Chaque événement réduit de 1 le nombre de lignées et à la fin de ce processus de coalescence, on aboutit au MRCA.



*Figure 4 : Reconstruction par processus coalescent de l'arbre associé à un échantillon de sept individus issus d'une population de taille constante. A gauche, la représentation schématique de l'évolution de la population dans le cas d'une population de taille constante. Chaque ligne représente une génération de la population et chaque rond, un individu. Les individus échantillonnés sont représentés en vert, leurs ancêtres en bleu, les ancêtres communs à au moins deux individus en jaune et le MRCA en orange. A droite, l'arbre enraciné généré à partir de ces sept échantillons.*

Dans les modèles de population, on peut intégrer le fait que la taille de la population évolue au cours du temps. Prenons le cas d'une population composée de souches de *M. bovis* isolées chez des bovins. Si l'incidence est stable, la population de *M. bovis* peut être considérée constante. Si l'incidence augmente, il peut être judicieux de choisir un modèle exponentiel pour représenter une population de *M. bovis* en expansion. Le modèle « skyline » est adapté à un contexte où l'évolution de la population est plus complexe et une tendance globale ne peut être identifiée. Dans un modèle « skyline », le temps est découpé en intervalles dans lesquels un taux de croissance différent est estimé (Pybus *et al.*, 2000).

### c. Covariables ou traits

Les modèles utilisés peuvent prendre en compte des traits ou covariables représentant une caractéristique associée aux séquences échantillonnées ; par exemple, leur localisation géographique ou encore l'espèce chez qui elle a été isolée. Lors de la construction de l'arbre phylogénétique, la valeur prise par cette covariable (ou état) dans les nœuds internes doit alors être estimée à partir de celle dans les nœuds terminaux et du choix du modèle de transition. Le modèle de transition d'un état à un

autre (par exemple : la transition d'une espèce-hôte A à une autre B) peut être considéré symétrique (taux de transition de A vers B égal au taux de B vers A) ou asymétrique (taux différents).

Il existe plusieurs méthodes dites « phylogéographiques » qui permettent l'inférence de la structure de la population à partir des phylogénies. La méthode de migration considère la migration d'un état à un autre comme un processus indépendant de la génération de l'arbre phylogénétique et elle peut conduire à des biais dans l'estimation des vitesses de migration quand l'échantillonnage est biaisé (Lemey *et al.*, 2009).

Le modèle de coalescence structurée est une extension de la théorie de la coalescence et décrit un processus de coalescence dans des sous-populations (définies par leur état) entre lesquelles les individus peuvent migrer. Une analyse phylogénétique se basant sur ce modèle permet l'estimation de paramètres tels que les vitesses de migration et la taille de population effective de chaque sous-population (correspondant à l'inverse de la vitesse de coalescence). Cette méthode requiert l'inférence de l'état de tous les nœuds internes de la phylogénie et demande donc de longs temps de calcul. L'utilisation de chaîne de Markov Monte Carlo (MCMC, voir partie IV) b)) pour réaliser cette inférence réduit le nombre d'états pouvant être étudié à deux ou trois (Müller *et al.*, 2017).

#### IV) Inférence phylogénétique

Plusieurs méthodes permettent d'inférer l'arbre phylogénétique et se regroupent en deux écoles distinctes suivant le critère utilisé : la distance ou les caractères. Dans les méthodes basées sur la distance, on calcule la distance génétique entre chaque paire de séquences et on génère une matrice de distances à laquelle est appliqué un algorithme permettant la construction de l'arbre. Ces méthodes ont l'avantage de ne pas demander une puissance de calcul très importante ; cependant elles sont déconseillées quand les séquences sont très divergentes et sont sensibles aux lacunes dans la séquence d'alignement (Yang et Rannala, 2012).

Les méthodes basées sur les caractères comparent simultanément toutes les séquences sur un caractère à la fois et calculent un score pour chaque arbre généré. Selon la méthode, le score obtenu peut correspondre à la somme du minimum de changements nécessaires à chaque site (score de parcimonie), à la vraisemblance (maximum de vraisemblance) ou encore à la probabilité de la distribution postérieure (inférence bayésienne) (Yang et Rannala, 2012).

##### a. Méthode de parcimonie

La méthode de parcimonie n'utilise pas explicitement un modèle du taux de substitution (Yang et Rannala, 2012). Elle se base sur le principe de parcimonie : l'explication la plus simple est la plus plausible. L'algorithme de Fitch permet de calculer le score de parcimonie de l'arbre, l'arbre avec le score le plus faible est celui qui requiert le moins de substitutions dans le processus évolutif. Dans l'exemple qui suit (Figure 5), on choisirait l'arbre A dont le score de parcimonie est de 3 (comparé à 5 pour l'arbre B).

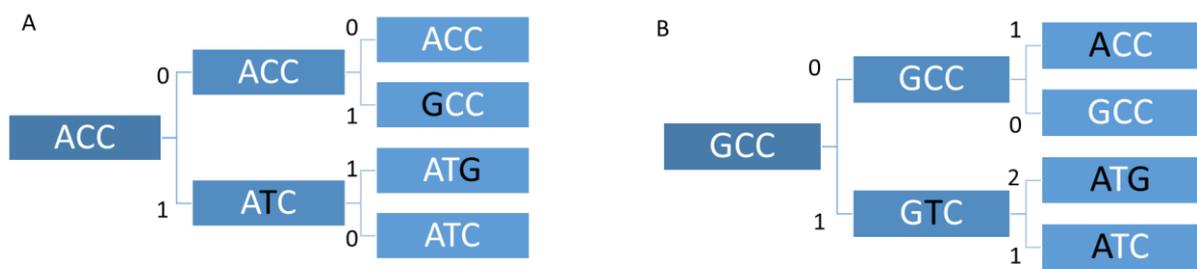


Figure 5 : Deux arbres phylogénétiques A et B reconstruits à partir des mêmes taxa définis par une séquence de trois nucléotides. Le nombre de substitutions nécessaires ( $N_s$ ) entre deux nœuds est indiqué à chaque branche. Le score de parcimonie de chaque arbre peut être calculé en sommant les  $N_s$  sur l'ensemble de l'arbre. Le score de parcimonie de l'arbre A ( $\sum(N_{sA})=1+0+0+1+1+0=3$ ) est de 3 et celui de B ( $\sum(N_{sB})=1+0+1+0+2+1=5$ ) est de 5. La méthode de parcimonie privilégiera donc l'arbre A.

**b. Inférence bayésienne**

L'inférence bayésienne se base sur la formule de Bayes :

$$f(\theta|D) = \frac{\text{Pr}(D|\theta) \cdot f(\theta)}{\text{Pr}(D)}$$

où D représente les données (ici l'ensemble des séquences nucléotidiques) et  $\theta$  le vecteur des paramètres dans le modèle.  $f(\theta)$  est la distribution *a priori* des paramètres,  $\text{Pr}(D|\theta)$  est la vraisemblance,  $\text{Pr}(D)$  est la vraisemblance marginale (ou VM) et  $f(\theta|D)$  est appelé la distribution *a posteriori*, que l'on souhaite estimer.

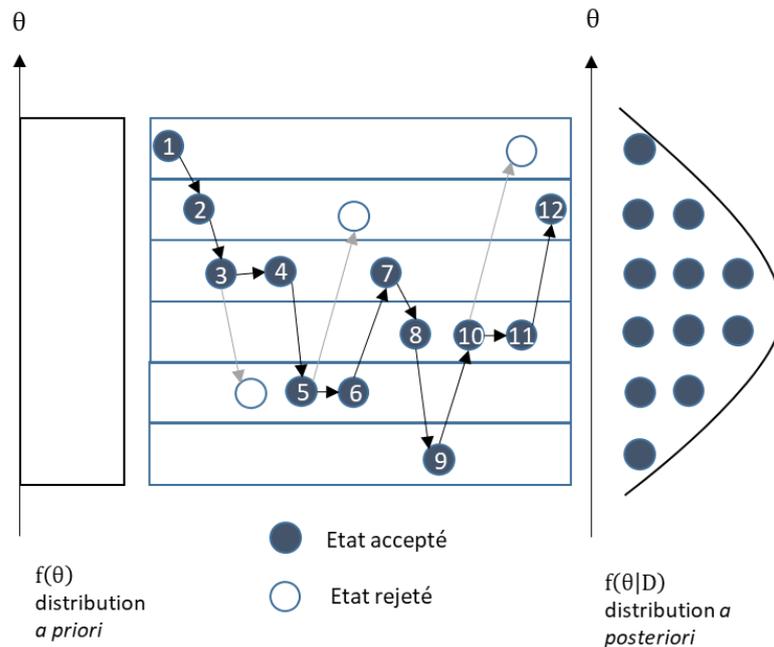
La distribution *a priori*, ( $f(\theta)$  ou prior), est choisie selon des résultats d'études déjà menées, des avis d'experts ou la littérature. L'inférence bayésienne est une actualisation de nos connaissances sur un modèle à partir de connaissances préexistantes (les priors) et de nos données D. Le poids relatif de chaque terme dépend du caractère très, peu voire non-informatif (qui représente l'absence de connaissances *a priori* sur le paramètre) de la distribution *a priori* choisie.

Il n'est pas possible de calculer directement la distribution *a posteriori* mais nous pouvons faire son approximation à partir d'échantillons qui en sont issus. L'algorithme de Markov Chain Monte Carlo (MCMC) permet d'explorer l'espace de la distribution *a posteriori*. Lors de l'exploration de cet espace, la probabilité de passer d'un état  $\theta$  à un autre état  $\theta'$  est de  $f(\theta'|\theta)$ . L'algorithme de Metropolis-Hastings (Hastings, 1970) nous donne la probabilité d'accepter cet état :

$$\alpha = \min \left( 1, \frac{f(\theta'|D) f(\theta|\theta')}{f(\theta|D) f(\theta'|\theta)} \right)$$

Un nombre est tiré au sort à partir d'une distribution uniforme comprise entre 0 et 1, si ce nombre est inférieur à  $\alpha$ , on accepte ce nouvel état et  $\theta$  devient  $\theta'$ , sinon on le rejette. Cette étape est répétée et l'ensemble des états visités constitue la chaîne de Markov. Etant donné que deux états  $\theta$  et  $\theta'$  sont généralement proches, ils dépendent l'un de l'autre. Cependant, quand un nombre suffisant d'étapes les sépare, les états deviennent des échantillons indépendants de la distribution *a posteriori* (Huelsenbeck *et al.*, 2001). Ce nombre d'étapes détermine la fréquence d'échantillonnage des états à partir de la chaîne (Figure 6). Les échantillons obtenus sont utilisés pour caractériser la distribution *a posteriori*, par exemple en permettant le calcul de la moyenne et l'écart-type des paramètres (Drummond et Bouckaert, 2015). L'Effective sample size (ESS) correspond au nombre de tirages

indépendants réalisés par MCMC, plus ce nombre est élevé, meilleure est l'approximation de la véritable distribution *a posteriori*.



*Figure 6* : Représentation schématique de l'approximation de la distribution a posteriori d'un vecteur de paramètres  $\theta$  par une chaîne de Markov Monte Carlo de longueur 12. A gauche, la distribution a priori d'un vecteur  $\theta$  contenant un seul paramètre qui suit une loi de distribution uniforme. Au centre, la représentation du processus d'acceptation ou de rejet d'un état (i.e. d'une valeur de  $\theta$ ). A droite, la distribution a posteriori constituée par la chaîne en considérant que l'on échantillonnait à chaque étape de la chaîne (d'après Lee et al., 2015).

Dans le cas de l'inférence bayésienne appliquée à des données de phylogénétique, les échantillons correspondent à des arbres caractérisés par leur topologie ainsi que l'ensemble des paramètres du modèle. L'arbre consensus permet de résumer l'information contenu dans ces arbres échantillonnés et plusieurs méthodes existent pour l'obtenir. Dans la méthode du maximum de crédibilité de clade, l'arbre consensus est celui dont le produit des probabilités postérieures des nœuds est le plus élevé, parmi les centaines voire les milliers échantillonnés. (Heled et Bouckaert, 2013).

### c. Sélection du modèle

Après avoir vérifié la validité et la plausibilité des modèles, vient l'étape du choix du modèle le mieux ajusté. Dans l'approche bayésienne, un indicateur appelé facteur de Bayes (BF) permet la comparaison de deux modèles.

Soit  $M_1$  et  $M_2$  deux modèles à comparer ( $M_1$  étant le modèle le plus simple) et  $VM_1$  et  $VM_2$  leurs vraisemblances marginales respectives, alors on peut définir le BF comme le rapport des VM :

$$BF = VM_1/VM_2 \text{ ou encore } \log(BF) = \log(VM_1) - \log(VM_2)$$

Si  $\log(BF) > 0$ , alors on favorise le modèle  $M_1$  sinon c'est  $M_2$  qui est choisi.

Pour chaque modèle, un algorithme (par exemple l'algorithme de « Nested Sampling » (Maturana et al., 2019) présenté en Annexe 9) permet d'obtenir une approximation de la vraisemblance marginale (VM) représentant la qualité moyenne de l'ajustement du modèle aux données.

*Tableau 1 : Interprétation du facteur de Bayes selon sa valeur*

<b>log(BF)</b>	<b>Interprétation</b>
<b>0-0,5</b>	Niveau de preuve faible
<b>0,5-1,3</b>	Niveau de preuve substantiel
<b>1,3-2,2</b>	Niveau de preuve fort
<b>&gt; 2,2</b>	Niveau de preuve décisif

## Travail personnel

### I) Contexte

Dans les Pyrénées-Atlantiques et les Landes, une augmentation du nombre de cas d'infection à *M. bovis* est observée depuis une dizaine d'années. Cette augmentation contribue à menacer le statut indemne de tuberculose bovine en France et une bonne compréhension de la situation épidémiologique (notamment des relations entre les espèces infectées) est nécessaire pour orienter les efforts de lutte.

Les espèces intervenant dans le système multi-hôte faune sauvage-faune domestique de *M. bovis* varient selon les régions et au cours du temps. Dans le rôle de réservoir, on peut citer le blaireau (*Meles meles*) en Irlande du Nord (Biek *et al.*, 2012 ; Trewby *et al.*, 2016) et au Royaume-Uni (Crispell *et al.*, 2019), le sanglier (*Sus scrofa*) en Espagne (Naranjo *et al.*, 2008), le phalanger-renard (*Trichosurus vulpecula*) en Nouvelle-Zélande (Kean *et al.*, 1999) ou encore le cerf de Virginie (*Odocoileus virginianus*) dans le Michigan aux Etats-Unis (Salvador *et al.*, 2019). Dans les Pyrénées-Atlantiques et les Landes, une thèse universitaire a notamment porté sur la transmission de *M. bovis* entre bovins et blaireaux dans une zone d'étude restreinte (Figure 7) (Bouchez-Zacria *et al.*, 2017)

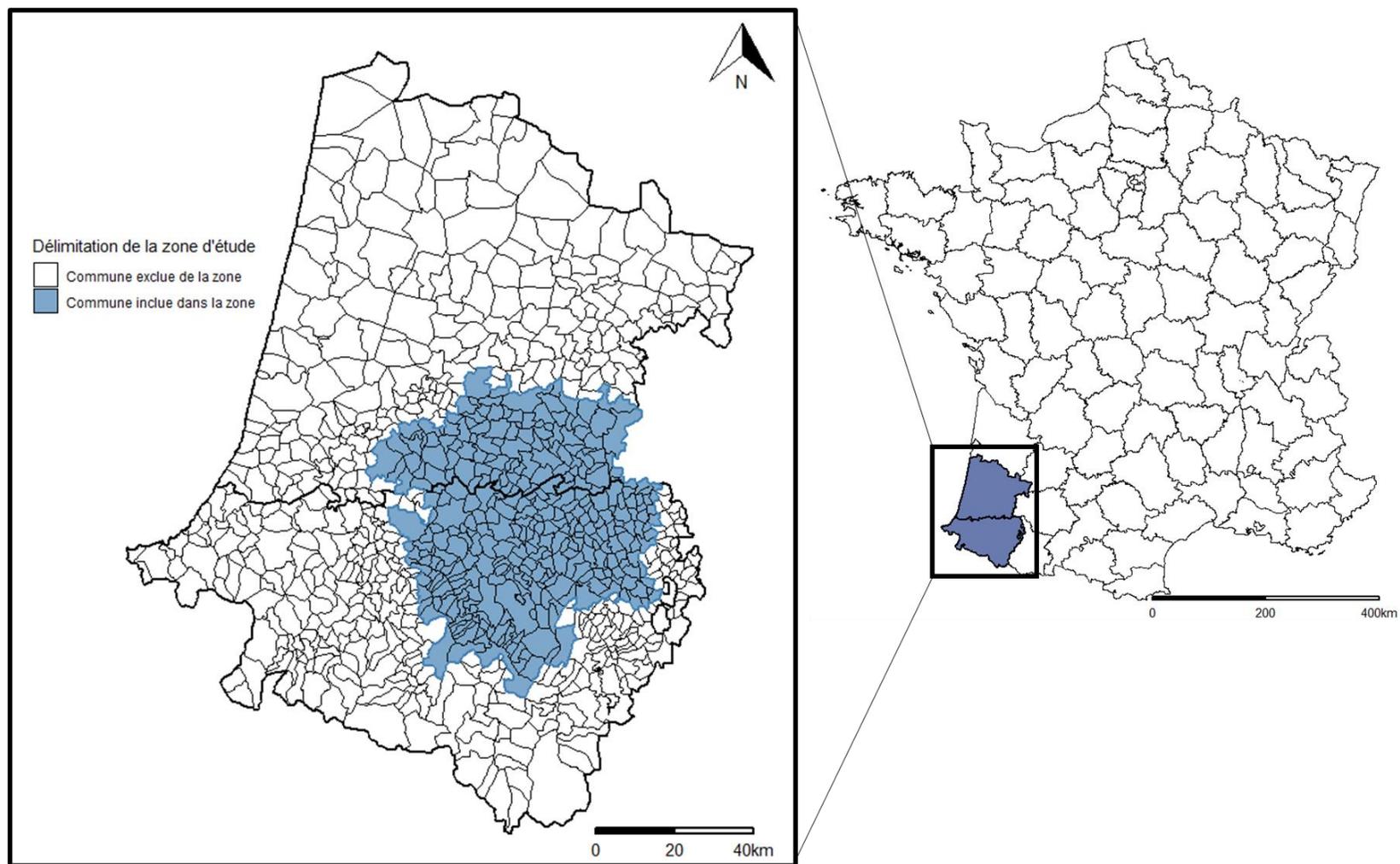
L'épidémiologie moléculaire est un outil permettant d'investiguer plus précisément le rôle joué par chacune des espèces hôtes dans la transmission de *M. bovis*. Elle est notamment utilisée pour estimer le taux de transition entre espèces, la direction dans laquelle cette transition s'effectue (Crispell *et al.*, 2019) ou encore la vitesse de diffusion spatiale de l'agent pathogène (Trewby *et al.*, 2016).

Notre étude a pour objectif l'investigation des rôles des blaireaux et des sangliers dans les Pyrénées-Atlantiques et les Landes, à partir des données de séquençage de *M. bovis* isolées chez ces espèces et chez des bovins.

### II) Matériel et Méthode

#### a. Sélection des souches

L'étude porte sur les souches de spoligotypes F7 et F15 échantillonnées entre 2002 et 2017 dans une zone de 3754 km<sup>2</sup>, située au Nord-Ouest de la ville de Pau, à cheval entre les Pyrénées-Atlantiques (64) et les Landes (40) (Figure 7). Cette zone composée de 335 communes, a été premièrement définie selon les limites d'un recensement des terriers de blaireaux réalisé en 2015, au cours de la thèse universitaire portant sur la transmission de *M. bovis* dans le système bovins-blaireaux en Pyrénées-Atlantiques et les Landes (Bouchez-Zacria *et al.*, 2017).



*Figure 7 : Représentation de la zone d'étude en bleu (à gauche) au sein des Pyrénées-Atlantiques dans le Sud-Ouest de la France (à droite)*

Toutes les souches isolées chez des bovins provenaient du système de surveillance de bTB de la région dans la limite de 3 souches par APDI et par élevage. Dans cette zone d'étude, des campagnes de dépistage ont eu lieu chaque année entre 2005 et 2017 dans les élevages des communes où au moins un élevage a été détecté comme foyer. Dans les autres communes de la zone d'étude, le rythme est biennal dans les Landes et triennal dans les Pyrénées-Atlantiques (Annexe 3). Jusqu'en 2016, le dépistage est effectué par IDC dans les élevages laitiers et IDS dans les autres types d'élevages, majoritaires dans la zone, puis l'utilisation de l'IDC a été privilégiée. Selon le contexte, un test positif ou non négatif est suivi d'un contrôle ou d'un abattage diagnostique avec dépistage sur tout ou une partie du troupeau.

Toutes les souches isolées chez des animaux infectés de la faune sauvage, obtenues par Sylvatub à partir de 2013 et pour lesquelles le typage moléculaire était possible, ont été incluses dans l'étude. Dans ces départements de niveau de surveillance 3, une surveillance programmée est réalisée sur les blaireaux avec des opérations de piégeages (Annexe 3 et 4). Le protocole de cette surveillance est régulièrement mis à jour selon les résultats épidémiologiques des années précédentes. Pour l'année 2017, il était demandé de prélever tous les blaireaux situés dans la zone d'infection (aires d'un rayon de 1,5 km aux abords des exploitations atteintes depuis moins de 6 ans et 2 km autour des terriers infectés depuis 2013) et un blaireau par terrier dans les autres zones ainsi que tous les blaireaux retrouvés morts en bord de route (Annexe 5).

#### b. Séquençage et sélection des SNP

Après réception, les souches sélectionnées ont été mises en culture liquide au Laboratoire National de Référence (LNR) et après chauffage et lavage, un lysat de souche est obtenu (ADN bactérien avec des débris cellulaires). Ce lysat a été envoyé à l'ICM (Institut du Cerveau et de la Moelle Epinière) pour séquençage de l'ADN par méthode ILLUMINA (Annexe 7) après une étape de purification. La taille des fragments de lecture obtenus était en moyenne de 150 pb. La qualité du séquençage a été contrôlée avec FASTQC en prenant pour critère d'acceptation d'une souche, un score de qualité de Phred (Q) supérieur ou égal à 30. L'équation permettant de définir ce score est la suivante :

$$Q = -10 \cdot \log(p)$$

Le paramètre p est la probabilité moyenne d'erreur de positionnement d'une base sur un fragment de lecture. Avec un score Q égal à 30, on admet un risque d'erreur de positionnement pour un site donné d'un fragment de 1 pour 1000.

Le logiciel Bionumerics version 7.6 a été utilisé pour réaliser l'alignement et l'assemblage des fragments par homologie à une souche de référence britannique AF2122/97 isolée d'un bovin en 1997. Après obtention de la séquence consensus, les souches dont le nombre de fragments alignés était inférieur à 300 000 ont été retirées. Les critères de sélection des SNP étaient : une distance inter-SNP minimale de 12 pb et une couverture minimale de 5 fragments (la mutation doit être présente sur au moins 5 fragments de lecture). Les SNP ambigus (plusieurs bases possibles) ou indéterminés ont été exclus ainsi que ceux qui étaient présents sur une région répétée du génome.

#### c. Données géographiques

Les données du Registre Parcellaire Graphique (RPG) déclarées chaque année par les éleveurs et utilisées dans le cadre de la Politique Agricole Commune (PAC) nous ont renseignés sur les pâtures dans la zone d'étude appartenant à chaque élevage (numéro EDE) en 2013. Nous avons fait l'hypothèse que ces parcelles restaient sensiblement les mêmes de 2007 à 2018. Pour chaque élevage, les coordonnées du centroïde de la commune contenant le plus de parcelles lui appartenant ont été choisies comme localisation. Si plusieurs communes avaient le même nombre maximal de parcelles, l'une d'entre elles était tirée au sort. Pour les élevages dont les données parcellaires n'étaient pas

disponibles et pour la seule souche isolée chez un porc, nous avons pris le centroïde de la commune correspondant au numéro EDE (*i.e.* les 5 premiers chiffres du numéro EDE).

Les données recueillies dans le cadre de la thèse universitaire déjà évoquée (Bouchez-Zacria *et al.*, 2017) comprenaient l'attribution d'un terrier avec ses coordonnées (longitude et latitude) aux blaireaux échantillonnés dans cette zone. Si les données ne permettaient pas d'obtenir la localisation précise d'un blaireau, les coordonnées du centroïde de la commune où l'animal a été trouvé ont été utilisées.

Pour les souches isolées de sangliers, les coordonnées du centroïde de la commune où l'animal a été trouvé ont été utilisées.

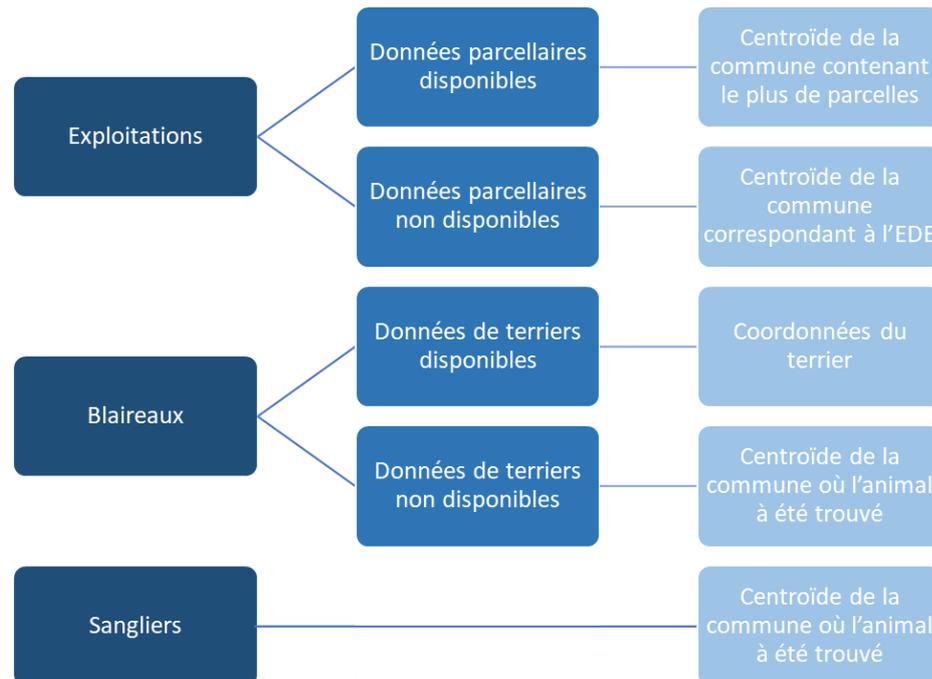


Figure 8 : Résumé de l'obtention des données géographiques par espèce.

#### d. Analyse phylogénétique

##### i. Maximum de parcimonie

Les données de SNP obtenues ont été importées sur le logiciel R 3.5.2 pour construire un arbre phylogénétique pour chaque spoligotype en utilisant la méthode de parcimonie. La fonction *pratchet* disponible dans le package *phangorn 2.5.5* (Schliep, 2011) permet d'obtenir le(s) arbre(s) avec le meilleur score de parcimonie (calculé par l'algorithme de Fitch). Parmi les arbres obtenus, nous avons choisi celui qui ne présentait pas de polytomies (arbre strictement binaire).

##### ii. Inférence bayésienne

Après exclusion des souches isolées chez des sangliers dont le faible nombre aurait entraîné des problèmes de convergence et d'estimations des paramètres, nous avons ajusté un modèle évolutif bayésien et construit un arbre consensus.

Pour ce faire, les modèles évolutifs ont été testés grâce au logiciel BEAST2, après annotation de chaque modèle avec l'interface BEAUti (Bayesian Evolutionary Analysis Utility) (Bouckaert *et al.*, 2019). Nous avons d'abord testé trois modèles de substitution (JC, HKY et GTR), associés à un modèle d'horloge moléculaire stricte et un modèle de population constante. Pour le premier modèle testé (JC), nous avons fait varier la longueur de la chaîne MCMC pour fixer celle qui permettait de répondre aux

critères de qualité (convergence et ESS ou « Effective Sample Size » suffisant). Par la suite, tous les modèles ont été estimés avec une chaîne de longueur égale à 500 millions avec une fréquence d'échantillonnage de 1/20 000 après une période de burn-in de 50 millions (10% de la chaîne).

Les graphiques de sortie de chaque paramètre ont été inspectés sur TRACER (Rambaut *et al.*, 2018) pour s'assurer de la convergence de la chaîne MCMC : oscillation autour d'une distribution stationnaire représentée par un aspect de « Hairy Caterpillar » (HC) ou « chenille poilue ». Les ESS ont également été inspectés, nous avons pris le seuil d'acceptabilité généralement fixé qui est de 200 (Robert *et al.*, 1997).

Le facteur de Bayes (BF) comparant les modèles deux à deux a permis de sélectionner le modèle de substitution le mieux ajusté aux données. L'approximation numérique de la VM a été calculée avec l'algorithme de « Nested Sampling » (Maturana *et al.*, 2019), implémenté dans le package NS (Annexe 9).

Nous avons ensuite testé les différents modèles présentés dans [Tableau 2](#). Pour le modèle de trait, nous avons choisi de tester le modèle de migration (que ce soit pour les espèces hôtes et pour la localisation géographique). Dans le modèle de diffusion spatiale discret (détaillé en Annexe 10), la zone où se trouvaient la majorité des souches (zone restreinte de 1610 km<sup>2</sup>) a été divisée en quatre cases de taille égale et le numéro de la case a été identifié comme trait. Une fois le modèle choisi, l'arbre consensus avec le maximum de crédibilité de clade (*i.e.* dont le produit des probabilités postérieures des nœuds est le plus élevé) a été obtenu grâce à Tree Annotator à partir des arbres estimés et avec une période de burn-in de 10%.

*Tableau 2 : Modèles testés avec BEAST. Le modèle de diffusion spatiale discret est détaillé en Annexe 10.*

Nom du modèle	Substitution	Horloge	Population	Transition Inter-espèce	Diffusion spatiale discret
JC1	JC	Stricte	Constante	Symétrique	non
HKY1	HKY	Stricte	Constante	Symétrique	non
GTR1	GTR	Stricte	Constante	Symétrique	non
HKY2	HKY	Exponentielle	Constante	Symétrique	non
HKY3	HKY	Log Normale	Constante	Symétrique	non
HKY4	HKY	Stricte	Exponentielle	Symétrique	non
HKY5	HKY	Stricte	« Skyline »	Symétrique	non
HKY6	HKY	Stricte	Constante	Asymétrique	non
HKY7	HKY	Stricte	Constante	Asymétrique	oui

### III) Résultats

#### a. Souches sélectionnées dans la zone d'étude

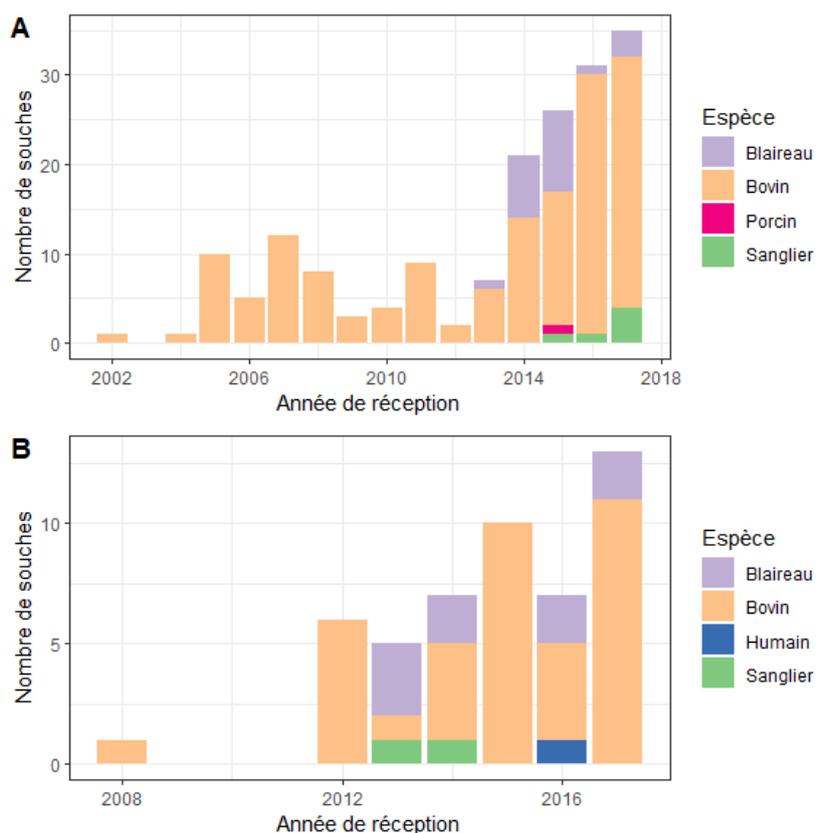
Les 224 souches de *M. bovis* sélectionnées ont été collectées entre 2002 et 2017 et provenaient de 5 espèces différentes (30 blaireaux, 184 bovins, 8 sangliers, 1 porc et 1 humain) dont les caractéristiques sont résumées dans le Tableau 3. Les souches isolées appartenait à la famille F4 et présentaient deux spoligotypes différents F7 (ou SB0821) et F15 (ou SB0832) et dix profils VNTR ont pu être déterminés sur 134 souches dont un profil incomplet (Annexe 11 et 12).

*Tableau 3 : Répartition des souches échantillonnées dans la zone d'étude en fonction de leur spoligotype (F15 ou F7) et de l'espèce hôte.*

Espèce	Spoligotype F7	Spoligotype F15	Total
Blaireau	21	9	30
Bovin	147	37	184
Humain	0	1	1
Porcin	1	0	1
Sanglier	6	2	8
Total	175	49	224

Les bovins proviennent de 107 élevages différents. Parmi ces 107 élevages, 50 comptaient plus d'un animal infecté (de 2 à 5). La répartition de la collecte des souches en fonction de l'espèce hôte et du temps (Figure 9) montre une augmentation du nombre de souches isolées chez des bovins depuis 2014 ainsi qu'une augmentation globale du nombre de souches durant la période d'étude.

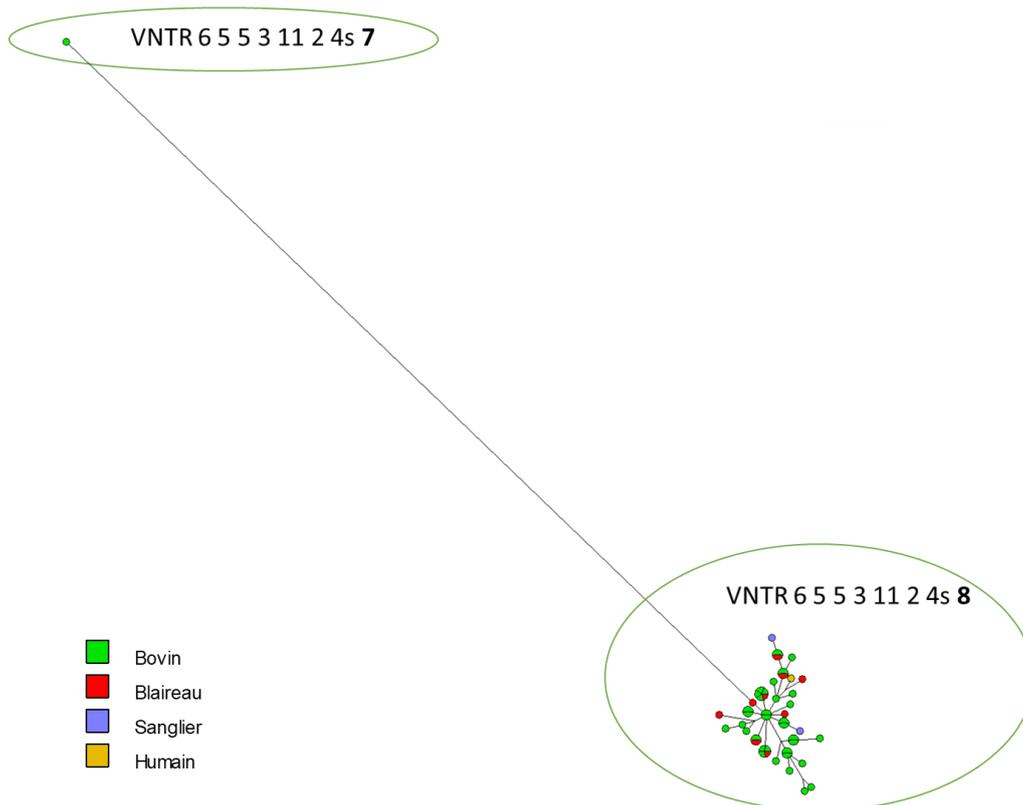
Un seul spoligotype a été isolé par élevage (le spoligotype F7 concernait 86 élevages et le spoligotype F15, 21 élevages). Les souches de spoligotype F15 ont été détectées chez des bovins plus tardivement (1<sup>er</sup> isolat en 2008) que les souches de spoligotype F7 (1<sup>er</sup> isolat en 2002). Les années d'isolement des souches chez des blaireaux, (2013 à 2017) sont les mêmes pour les deux spoligotypes contrairement aux données d'isolement chez les sangliers (2013 et 2014 pour F15, puis 2015 à 2017 pour F7).



*Figure 9 : Nombre de souches isolées de spoligotype F7 (en A) et F15 (en B) au cours du temps en fonction des espèces hôtes.*

## b. Données de séquençage

Après exclusion des souches dont le nombre de fragments alignés était inférieur à 300 000 (2 souches F15 et une souche F7, toutes provenant de bovins), la génération d'arbres non enracinés par maximum de parcimonie a permis d'identifier dans la famille F15, une souche trop éloignée des autres (et vérifiant pourtant les critères de qualités mais dont le profil VNTR était différent) qui a été exclue ([Figure 10](#)). Cette souche est la seule souche de spoligotype F15 qui a été isolée en 2008 et son élimination restreint la profondeur temporelle pour ce spoligotype à la période comprise entre 2012 et 2017.



*Figure 10 : Arbre phylogénétique des souches de spoligotype F15 construit par maximum de parcimonie sur le logiciel Bionumerics mettant en évidence une souche très éloignée des autres et dont le profil VNTR diffère (6 5 5 3 11 2 4s 7 vs. 6 5 5 3 11 2 4s 8).*

La position des SNP a permis d'identifier les SNP peu fiables (car présents dans des zones répétitives de l'ADN) et d'en restreindre le nombre final. Finalement, 47 SNP ont été obtenus pour F15 et 172 (pour la totalité des souches) ou 171 (en prenant en compte seulement les 167 souches isolées chez des bovins et des blaireaux) pour F7.

## c. Données géographiques

### i. Spoligotype F15

Pour les souches de spoligotype F15, les données parcellaires étaient disponibles pour 12 numéros EDE sur 20 (soit 23/37 bovins) et les données de terrier étaient disponibles pour 6/9 blaireaux. Sur la carte [Figure 11,A](#), nous observons une majorité (44/48) des souches de spoligotype F15 dans le Sud-Ouest de

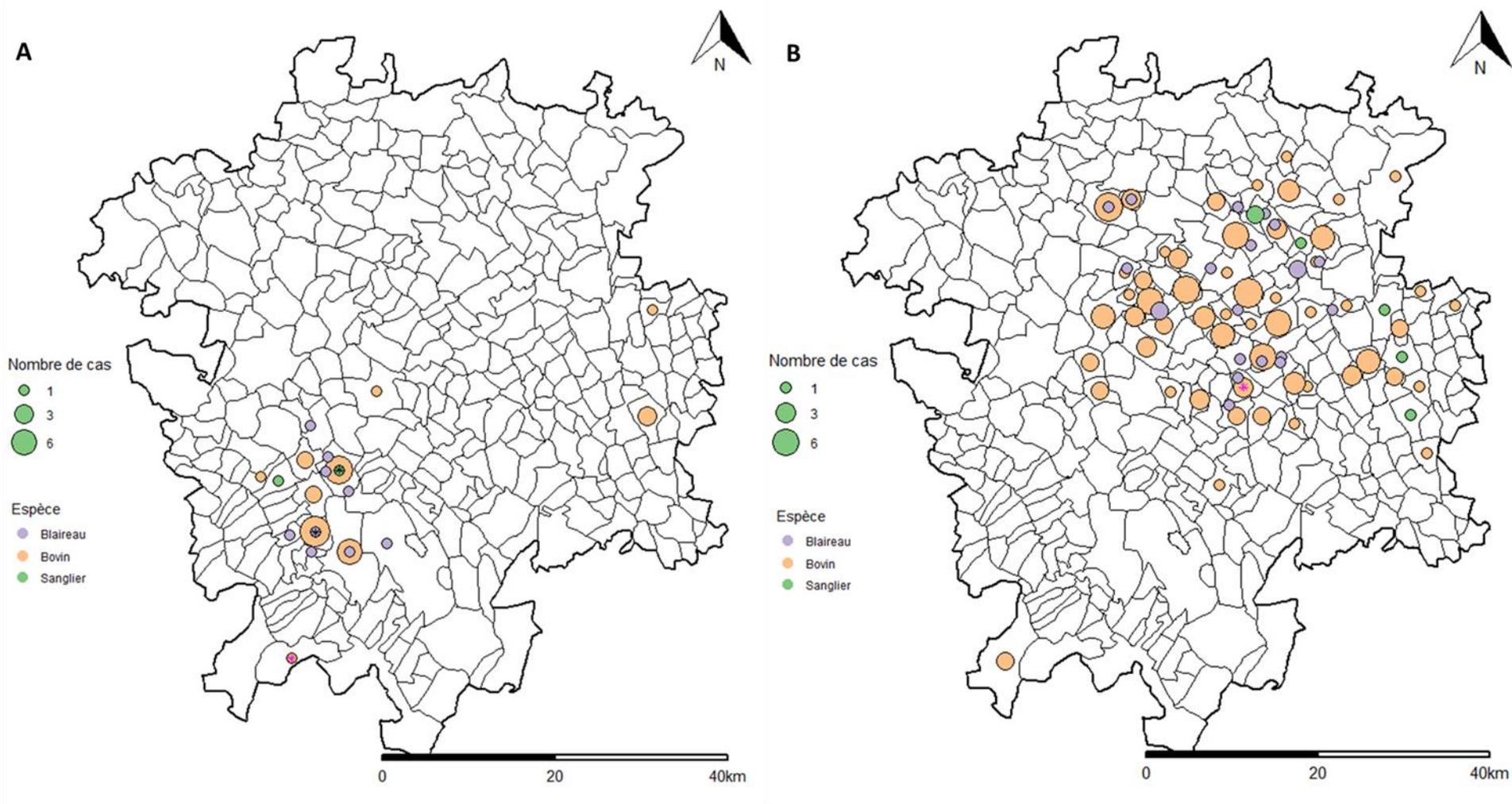
la zone d'étude. Il est à noter que les coordonnées de localisation de la souche de spoligotype F15 isolée chez l'homme ne sont pas disponibles pour des raisons d'anonymat.

Les souches provenant de blaireaux se situaient en moyenne à 2,2 km du bovin contaminé le plus proche (min-max : 0 - 4,4 km) (Figure 11,A). De même, les souches isolées de sangliers se situaient à 0 et 2,1 km du bovin contaminé le plus proche et à 1,6 et 5,5 km de la souche la plus proche isolée chez un blaireau.

#### *ii. Spoligotype F7*

Pour les souches de spoligotype F7, nous avons obtenu les données parcellaires pour 39 numéros EDE sur 85 (soit 82/147 bovins) et les données de géolocalisation de terriers pour 19/21 blaireaux. Sur la carte Figure 11,B, on observe que les souches de spoligotype F7 se retrouvent principalement (172/174) dans le centre, le Nord et l'Est de la zone d'étude. Il est à noter que la localisation de la souche de spoligotype F7 isolée chez un porc n'est pas représentée car elle se trouve en-dehors de la zone d'étude (Annexe 13).

Nous pouvons observer que les souches provenant de blaireaux se situent en moyenne à 1,5 km du bovin contaminé le plus proche (min-max : 0 - 2,6 km) (Figure 11,B). De même, les souches isolées de sangliers se situent en moyenne à 2,5 km du bovin contaminé le plus proche (min : 0 km, max : 3,4 km) et à 6 km de la souche la plus proche isolée chez un blaireau (min : 1,2 km, max : 15,2 km).

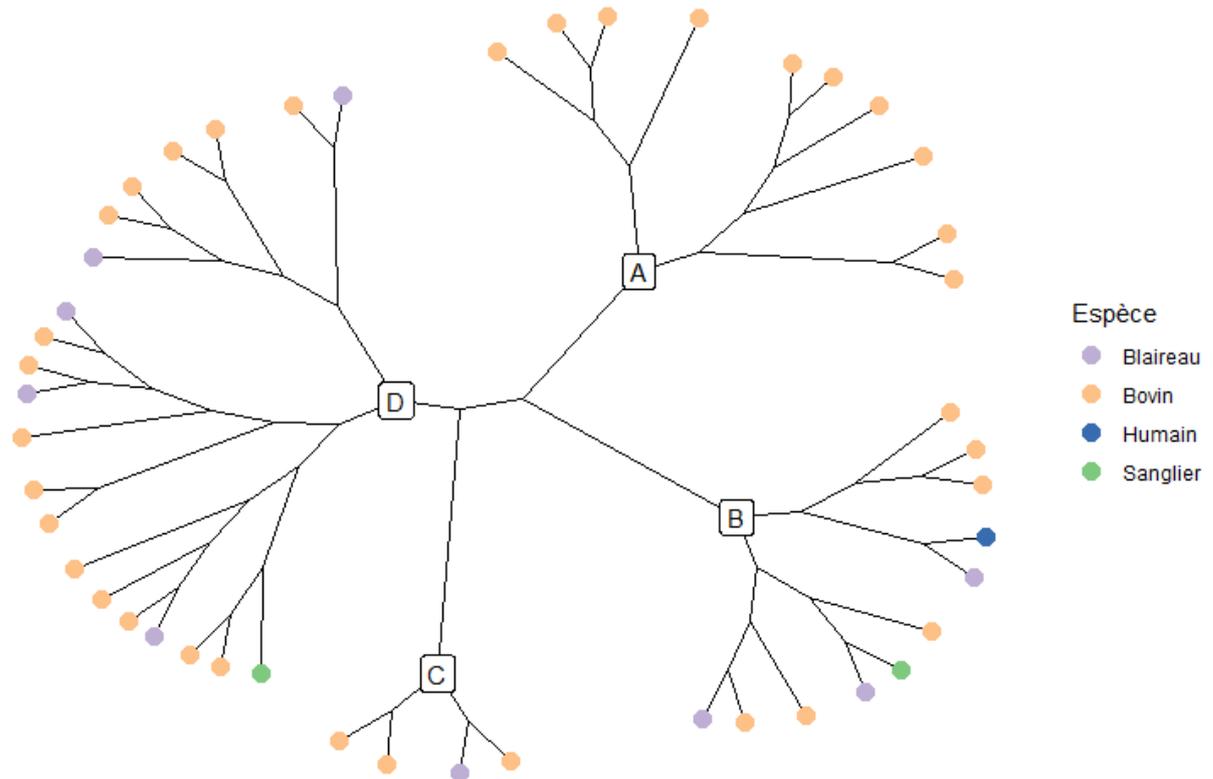


*Figure 11 : Géolocalisation des souches de M. bovis de spoligotype F15 (A) et de spoligotype F7 (B) dans la zone d'étude. (A) En vert, on retrouve les 2 souches isolées chez des sangliers, en violet chez des blaireaux (9 souches) et en orange, chez des bovins (37 souches). Les souches éliminées car les critères de qualité n'étaient pas remplis, sont marquées par une étoile noire ; celle éliminée car trop éloignée des autres, par un astérisque rose. (B) En vert, on retrouve les 6 souches isolées chez des sangliers, en violet chez des blaireaux (21 souches) et en orange, chez des bovins (147 souches). L'astérisque rose marque l'emplacement de la souche éliminée lors de la sélection des souches.*

#### d. Analyse par méthode de parcimonie

##### i. Spoligotype F15

L'arbre phylogénétique pour les 46 souches de spoligotype F15 avec 47 SNP, par méthode de parcimonie est présentée sur la [Figure 12](#). L'inspection visuelle de la topologie de l'arbre a permis de déterminer quatre clades.



*Figure 12 : Arbre phylogénétique des souches de spoligotype F15 construit par méthode de parcimonie. Les lettres A, B, C et D correspondent au MRCA de chaque clade.*

Le clade A ([Tableau 4](#)) possède exclusivement des souches isolées de bovins (10/10). Dans le clade C, le plus petit, nous retrouvons des souches isolées à la fois de bovins (3/4) et de blaireaux (1/4). Le plus grand clade (D) comporte 21 souches dont 1 isolée d'un sanglier, 5 de blaireaux et 15 de bovins. Le dernier clade (B) est le plus diversifié avec 1 souche isolée d'un sanglier, 3 de blaireaux, 6 de bovins et enfin la seule souche isolée chez un humain.

La [Figure 13](#) montre que seul le clade A possède des souches éloignées les unes des autres à plus de 20 km avec 2 clusters.

*Tableau 4 : Caractéristiques des clades pour le spoligotype F15 : espèce hôte, profondeur temporelle et nombre de communes concernées.*

Clade	Nombre de souches isolées				Profondeur temporelle	Nombre de communes
	Bovin	Blaireau	Sanglier	Humain		
A	10	0	0	0	3 ans	4
B	6	3	1	1	5 ans	3
C	3	1	0	0	6 ans	2
D	15	5	1	0	6 ans	10

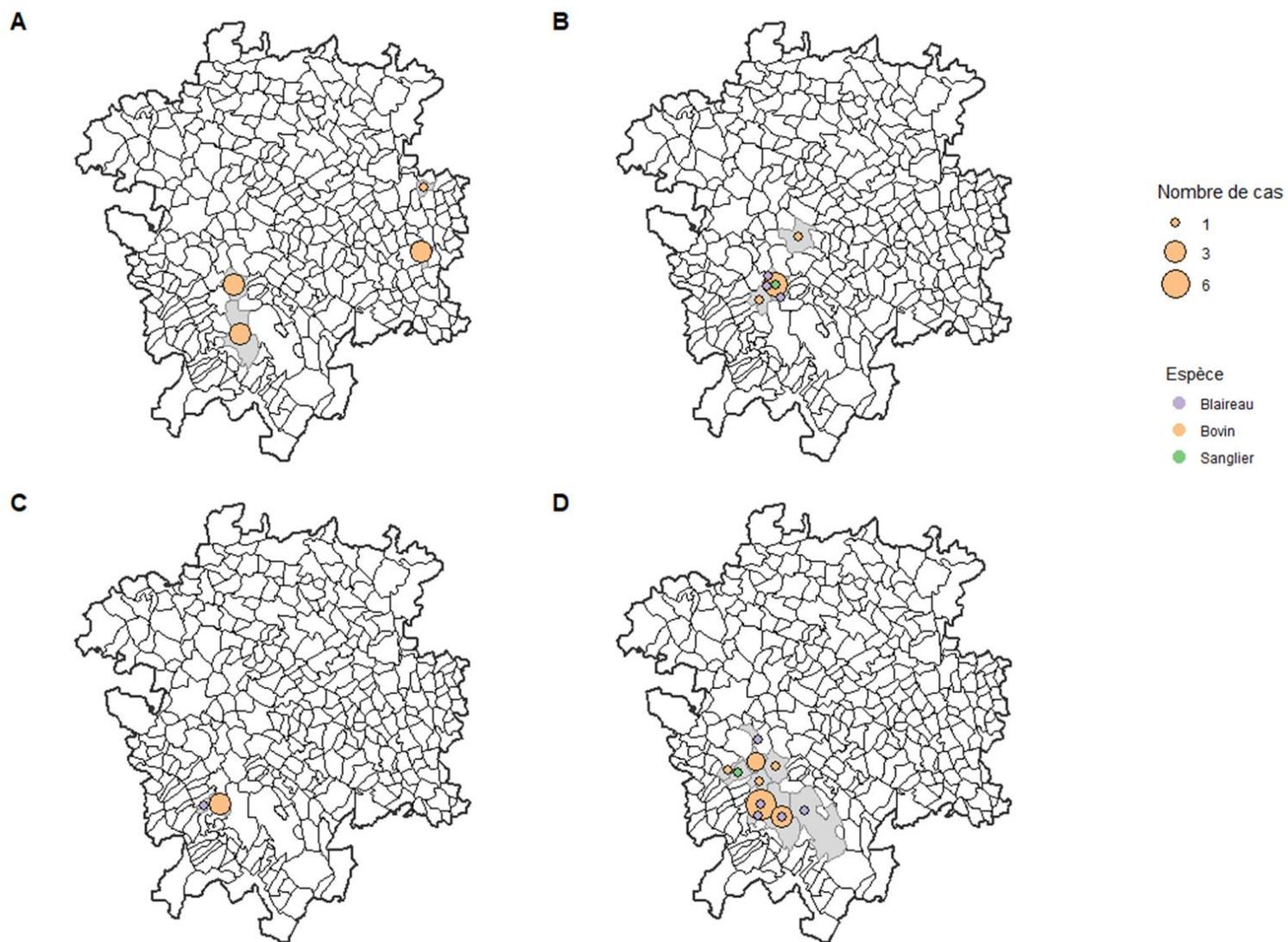
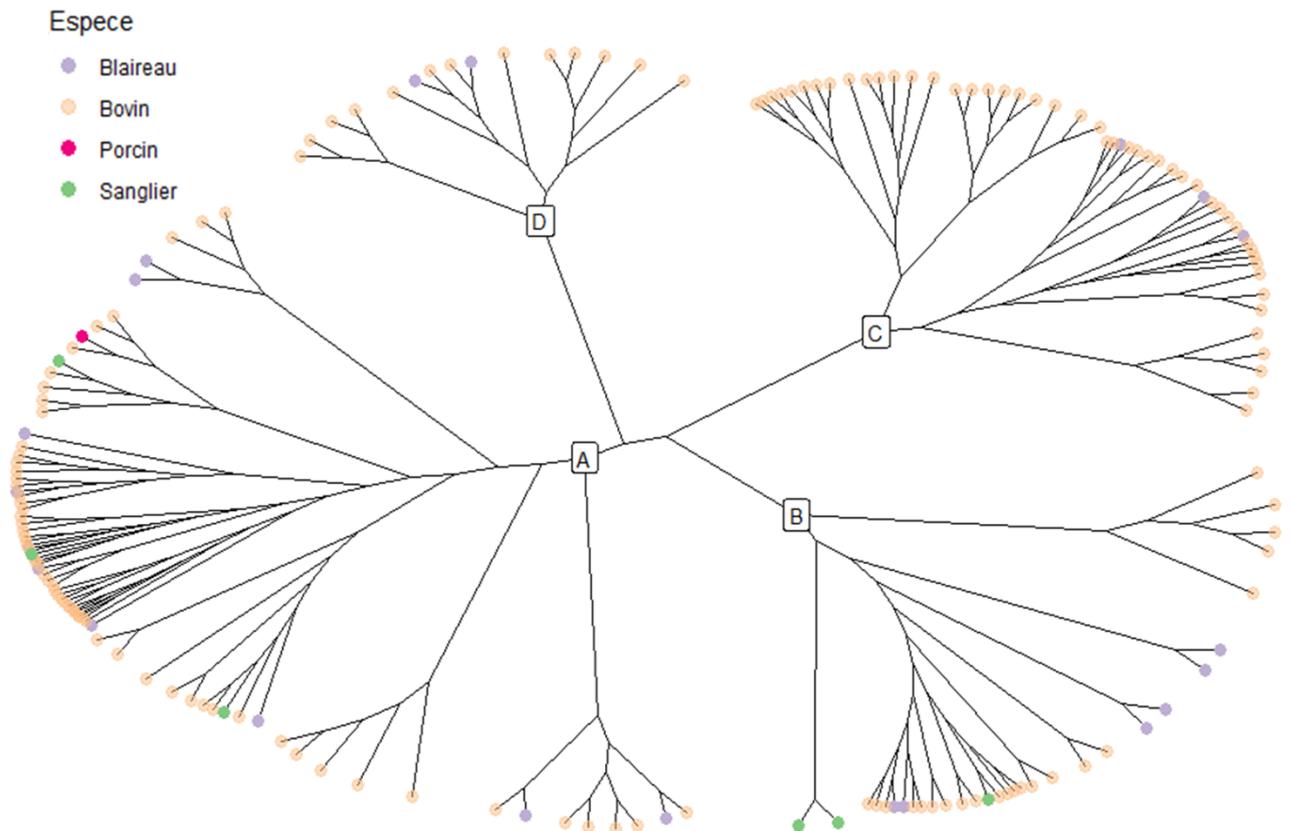


Figure 13 : Répartition géographique des clades A, B, C et D, des souches de spoligotype F15, identifiés à partir de l'arbre de maximum de parcimonie. Les communes où se trouve au moins une souche appartenant au clade identifié par la lettre à gauche de la figure sont colorées en gris.

## ii. Spoligotype F7

De manière analogue à F15, pour les 174 souches et 172 SNP, la répartition en clades a été réalisée selon la topologie de l'arbre obtenu par méthode de parcimonie ([Figure 14](#)).

La construction de cet arbre permet de visualiser les souches non prises en compte dans le modèle évolutif ; à savoir, les souches isolées chez des sangliers et chez le porcine que l'on retrouve dans les clades A et B. On ne retrouve aucun clade mono-espèce, *a minima* les souches sont isolées à la fois de bovins et de blaireaux (clade C et D). La souche provenant du seul porcine se situe dans le clade de plus grande taille, A ([Tableau 5](#)).



[Figure 14](#) : Arbre phylogénétique des souches de spoligotype F7 construit par méthode de parcimonie. Les lettres A, B, C et D correspondent au MRCA de chaque clade.

[Tableau 5](#) : Caractéristiques des clades pour le spoligotype F7 : espèce hôte, profondeur temporelle et nombre de communes concernées.

Clade	Nombre de souches isolées				Profondeur temporelle	Nombre de communes
	Bovin	Blaireau	Sanglier	Porcin		
A	13	2	0	0	11 ans	7
B	48	3	0	0	16 ans	22
C	22	6	3	0	14 ans	21
D	63	10	3	1	13 ans	33

### e. Inférence bayésienne

L'inférence bayésienne n'a concerné que les souches de spoligotype F7 isolées chez des bovins (146 souches) et des blaireaux (21 souches). L'Annexe 14 présente les modèles testés. Tous les modèles ont convergé (aspect graphique en « chenille poilue » et les ESS de chaque paramètre était supérieur au seuil fixé à 200).

Le résultat de la comparaison de ces 9 modèles (Annexe 14) montre que nous avons fait le choix du modèle de substitution HKY par rapport au modèle GTR (en cohérence avec de précédents modèles d'évolution bayésien (Crispell et al., 2017 ; Salvador et al., 2019 ; Modenesi, 2019)). Le modèle sélectionné inclut une horloge moléculaire stricte, un modèle de population constant et pas de diffusion spatiale. Le BF n'a pas permis de favoriser un modèle de transition entre espèce hôte, les deux modèles sont donc présentés.

La [Figure 15](#) représente les arbres consensus obtenus pour chaque modèle de transition inter-espèce. On peut remarquer que l'espèce hôte prédite pour la racine est incertaine pour le modèle asymétrique contrairement au modèle symétrique qui prédit un bovin comme espèce hôte ([Tableau 6](#)). La hauteur estimée moyenne de l'arbre est de 31,28 ans vs. 30,5 ans (HPD : Highest posterior density à 95% : [22,40 ; 41,38] vs. [22,17 ; 40,61]). Le MRCA aurait donc probablement circulé entre 1975 et 1995 (vs. 1976 et 1995) ([Tableau 7](#)). Le taux de substitution estimé (nombre de SNP × taux d'horloge) est de 0,35 substitutions/génome/an pour les deux modèles (HPD à 95% : [0,26 ; 0,45] vs. [0,25 ; 0,44]).

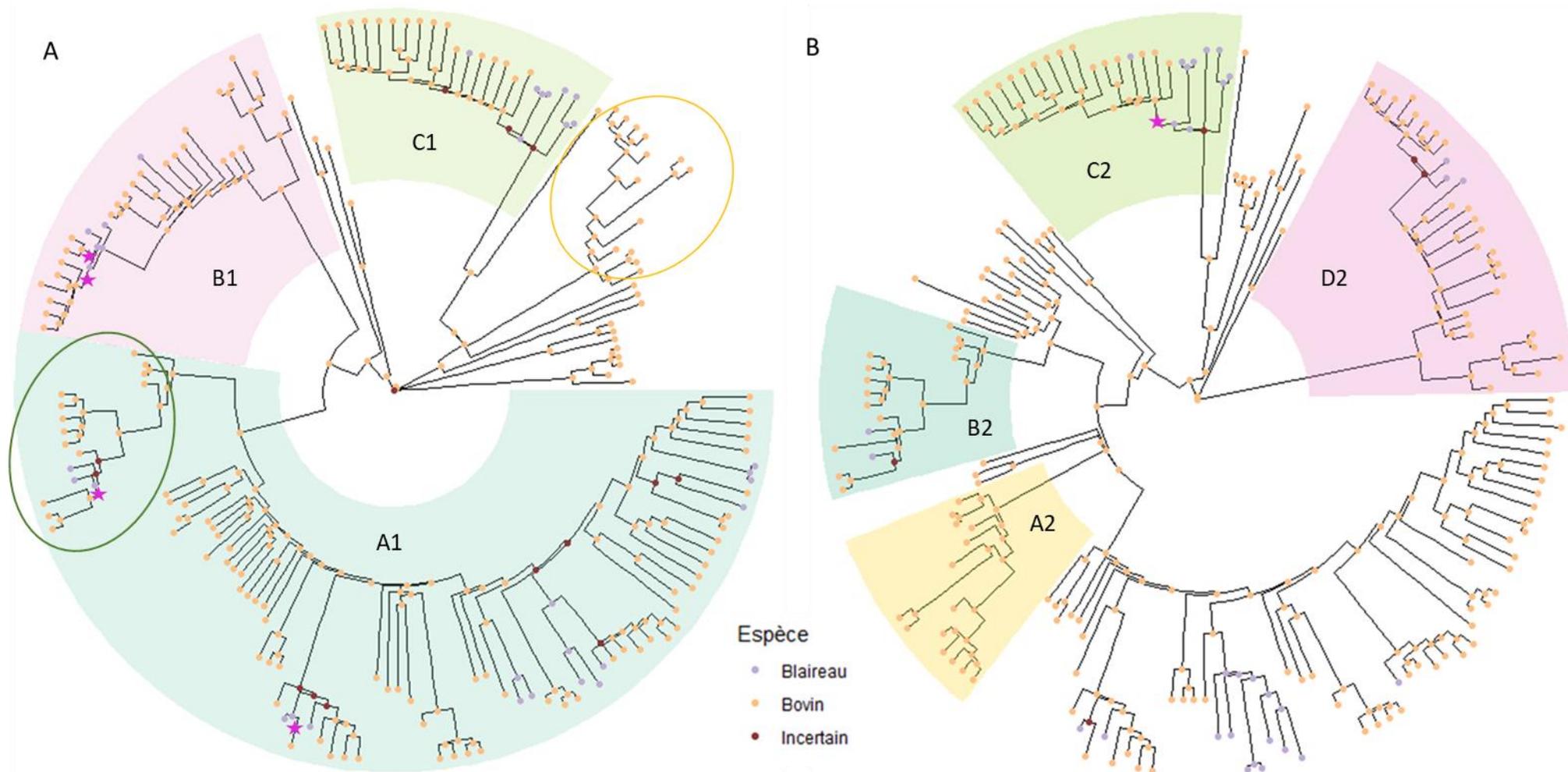
Les taux de transition inter-espèce ont été calculés comme le produit (taux de transition relatif × taux d'horloge de l'espèce hôte). Le modèle asymétrique prédit un taux de transition de blaireau à bovin (0,40 transition par souche et par an) 6.7 fois supérieur au taux de transition de bovin à blaireau (0,06 transition par souche et par an). Le modèle symétrique prédit quant à lui, 0,03 transition par souche et par an. Sur l'arbre phylogénétique reconstruit à partir de 146 souches isolées chez des bovins et 21 souches isolées chez des blaireaux, quatre événements de transition de blaireaux à bovins sont prédits par le modèle asymétrique ([Figure 15,A](#)) et un seul, par le modèle symétrique ([Figure 15,B](#)). On peut noter toutefois que les trois événements de transition supplémentaires dans le modèle asymétrique correspondent à des ancêtres dont l'espèce hôte n'a pu être prédite.

Un clade est défini par dix isolats minimum (feuilles) dont le MRCA a une probabilité postérieure supérieure à 0,65. La répartition en clades ([Figure 15](#)) montre des similitudes entre les 3 clades de tailles variables (22 à 88 souches) définis dans le modèle asymétrique et les 4 clades (de 14 à 28 souches). En effet, le clade B2 dans le modèle symétrique possède 15 souches retrouvées dans le clade de même couleur (clade A1) dans le modèle asymétrique. Les clades C1 et C2 correspondent aux mêmes 22 souches dans les deux modèles, de même pour les 28 souches du clade B1 (modèle asymétrique) et D2 (modèle symétrique). On remarque la présence de 29 vs. 88 souches que nous n'avons pas pu attribuer à un clade.

Dans la [Figure 16](#) et la [Figure 17](#), seules les souches appartenant à des clades composés uniquement de bovins présentent des distances supérieures à 20 km. Les clades C1 et C2 sont identiques dans les deux cartes et présentent un regroupement des souches isolées chez des blaireaux d'un côté (à l'Ouest) et de celles isolées chez des bovins de l'autre (à l'Est) ce qui est cohérent le scénario décrit dans les deux arbres ([Figure 15](#)). Dans ce scénario, un seul phénomène de transition aurait lieu entre la population de blaireaux et celle de bovins puis l'agent pathogène aurait circulé séparément dans les deux populations.

*Tableau 6 : Probabilité que l'espèce hôte soit un bovin et probabilité postérieure du MRCA de chaque clade et de la racine de l'arbre consensus. ASYM correspond au modèle asymétrique et SYM au modèle symétrique*

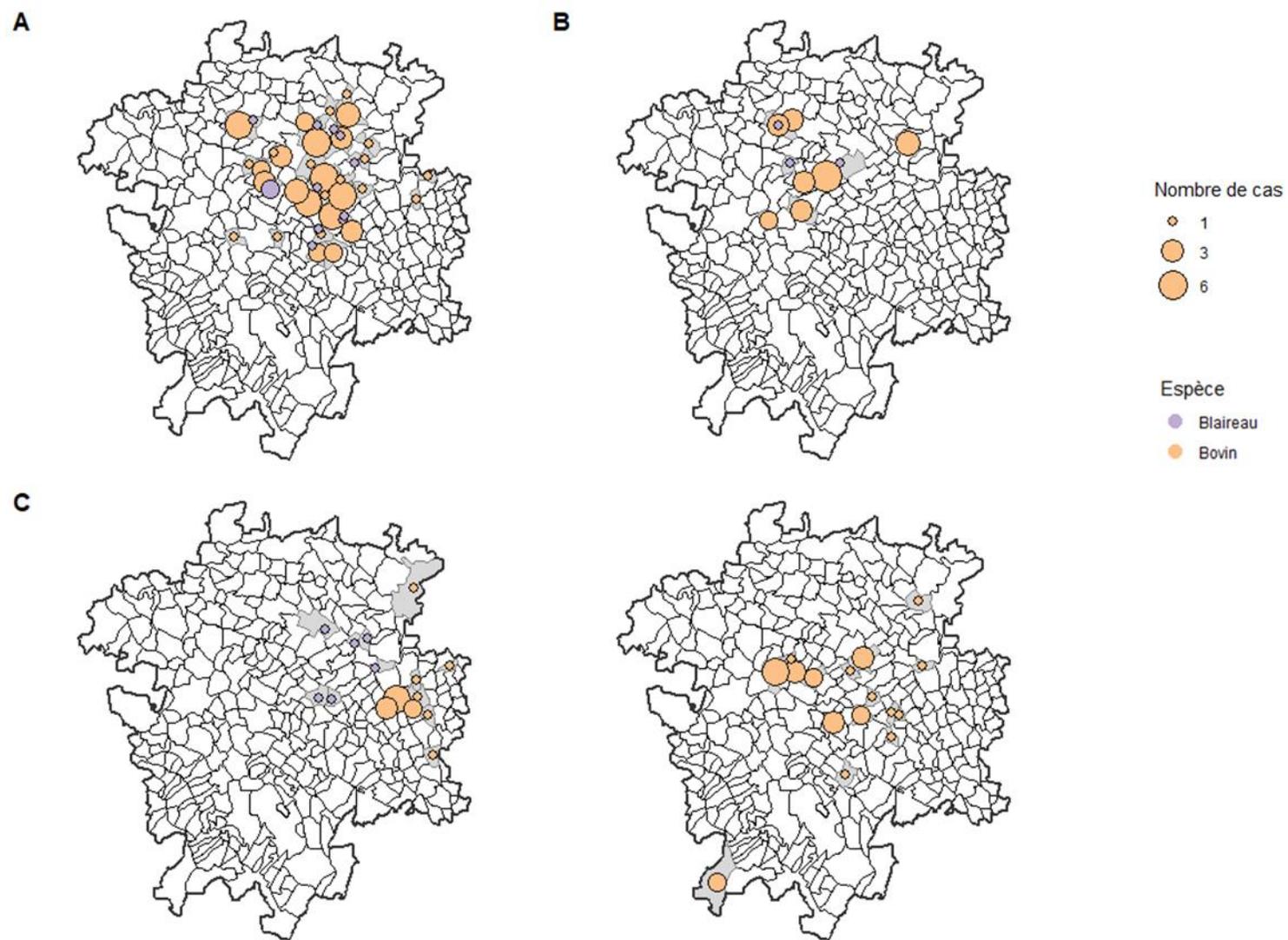
Nœud	Probabilité (Bovin)		Probabilité postérieure	
	ASYM	SYM	ASYM	SYM
<b>Racine</b>	0,51	0,92	1	1
<b>MRCA du clade A</b>	0,93	0,99	0,70	0,97
<b>MRCA du clade B</b>	0,92	1	1	1
<b>MRCA du clade C</b>	0,62	0,65	0,90	0,89
<b>MRCA du clade D</b>	-	0,99	-	1



*Figure 15* : Arbres enracinés consensus (en (A) du modèle asymétrique et en (B), du modèle symétrique) des souches de spoligotype F7 (inférence bayésienne) selon le maximum de crédibilité de clade. Les nœuds internes dont la probabilité de l'espèce était inférieure à 0,70 sont indiqués comme incertains. En (A), le clade A1 (en vert foncé) comporte 88 souches, le clade B1 (en rose) 28 souches, le clade C1 (en vert pâle) 22. Il reste 29 souches n'appartenant à aucun clade. En (B), le clade A2 (en jaune) comporte 14 souches que l'on retrouve dans le cercle jaune en (A), dans le clade B2 (en vert foncé) 15 souches sont retrouvées dans le clade A1 (en A, entourées en vert), le clade C2 (en vert pâle), 22 souches est le même que le clade C1 (A) et de même pour le clade D2 et B1 (en rose), 28 souches. Il reste 88 souches n'appartenant à aucun clade.

*Tableau Z : Paramètres de sorties pour chaque modèle (ASYM : asymétrique et SYM : symétrique) avec leur moyenne, intervalle de crédibilité à 95% et leur ESS ainsi que les paramètres calculés à partir de ces sorties.*

Paramètre	Unité	Moyenne et HPD 95%		ESS	
		ASYM	SYM	ASYM	SYM
<b>Modèle de transition</b>					
<b>Hauteur de l'arbre</b>	ans	31,28 [22,40 ; 41,38]	30,50 [22,17 ; 40,61]	5812	5835
<b>Taux d'horloge</b>	substitutions/site/an	$2,04 \cdot 10^{-3}$ [ $1,50 \cdot 10^{-3}$ ; $2,61 \cdot 10^{-3}$ ]	$2,03 \cdot 10^{-3}$ [ $1,49 \cdot 10^{-3}$ ; $2,58 \cdot 10^{-3}$ ]	3147	3185
<b>Taux de substitution</b> (=Taux d'horloge*nombre de SNP)	substitutions/génome/an	0,35 [0,26 ; 0,45]	0,35 [0,25 ; 0,44]	-	-
<b>Taux d'horloge de l'espèce hôte</b>		0,23 [0,07 ; 0,41]	0,03 [0,01 ; 0,06]	10372	6988
<b>Taux de transition relatif (Blaireau &lt;&gt; Bovin)</b>		-	$0,999$ [ $3,61 \cdot 10^{-6}$ ; 2,96]	-	22501
<b>Taux de transition Blaireau &lt;&gt; Bovin</b> (=Taux d'horloge de trait * taux de transition relatif)		-	<b>0,03</b>	-	-
<b>Taux de transition relatif (Blaireau &gt; Bovin)</b>		1,75 [0,03 ; 4,15]	-	22501	-
<b>Taux de transition Blaireau &gt; Bovin</b> (=Taux d'horloge de trait * taux de transition relatif)	transitions/souche/an	<b>0,40</b>	-	-	-
<b>Taux de transition relatif (Bovin &gt; Blaireau)</b>		$0,25$ [ $4,42 \cdot 10^{-3}$ ; 0,66]	-	21968	-
<b>Taux de transition Bovin &gt; Blaireau</b> (=Taux d'horloge de trait * taux de transition relatif)	transitions/souche/an	<b>0,06</b>	-	-	-



*Figure 16 : Répartition géographique des clades A, B et C et des souches sans clade, identifiés à partir de l'arbre du modèle de transition asymétrique. Les communes où se trouve au moins une souche appartenant au clade, identifié par la lettre en haut à gauche de la figure, sont colorées en gris.*



*Figure 17* : Répartition géographique des clades A, B, C et D et des souches sans clade, identifiés à partir de l'arbre du modèle de transition symétrique. Les communes où se trouve au moins une souche appartenant au clade, identifié par la lettre en haut à gauche de la figure, sont colorées en gris.

## V) Discussion

Dans notre étude, l'objectif était d'utiliser des données de séquençage de *M. bovis* pour répondre à la question des rôles joués par les blaireaux et les sangliers dans la transmission de la bTB dans les Pyrénées Atlantiques et les Landes. Pour ce faire nous disposons de 174 souches de spoligotype F7 et 46 de spoligotype F15, isolées chez de multiples espèces (bovins, blaireaux, sangliers, un humain et un porc).

Les données concernant les bovins ont été obtenues principalement par deux niveaux différents du système de surveillance de la tuberculose bovine en France : le dépistage et la surveillance en abattoir. En ce qui concerne les souches isolées dans la faune sauvage, elles proviennent des données de surveillance événementielle ainsi que des opérations de piégeages réalisées dans le cadre de Sylvatub à partir de 2013.

Nous avons construit des arbres phylogénétiques par maximum de parcimonie pour les deux spoligotypes. La présence de plusieurs clades comportant à la fois des bovins, des blaireaux et des sangliers suggèrent que plusieurs phénomènes de transmission pourraient avoir eu lieu pour les deux spoligotypes. Cependant, à partir de ces arbres, nous n'avons pas d'information sur la direction des échanges : la transmission s'est-elle effectuée des bovins vers la faune sauvage ou inversement ? De plus, on ne peut savoir si cette transmission bovins-faune sauvage concerne les blaireaux et les sangliers ou si une espèce joue un rôle d'intermédiaire par exemple.

Par la suite, nous avons adopté une approche bayésienne pour analyser les données de spoligotype F7 concernant les bovins et les blaireaux uniquement. Cette approche permet, à l'aide d'un modèle d'évolution, de reconstruire l'état ancestral des nœuds internes, ici l'espèce hôte, et d'estimer les paramètres d'intérêt tels que le taux de transition inter-espèce.

Le modèle d'évolution qui a été choisi a permis de déterminer que le modèle d'horloge stricte et de population constant s'ajustait le mieux aux données. Le modèle de substitution HKY, plus simple que GTR, a été privilégié pour être cohérent avec de précédents modèles d'évolution bayésien ajustés sur des données provenant de Dordogne (Modenesi, 2019), de Nouvelle-Zélande et aux Etats-Unis (Crispell *et al.*, 2017 ; Salvador *et al.*, 2019).

Le taux de substitution estimé est du même ordre de grandeur que le taux de 0,2 95% HPD : [0,1 - 0,3] substitutions/génome/an estimé dans une étude concernant des souches isolées au Royaume-Uni (Trewby *et al.*, 2016) ou encore le taux de 0,37 95% HPD : [0,24 – 0,51] substitutions/génome/an estimé dans le Michigan (Etats-Unis) par Salvador *et al.*, 2019. De même, en France, une étude sur des souches isolées en Dordogne a estimé un taux de 0,28 95% HPD : [0,24 – 0,34] substitutions/génome/an (Modenesi, 2019). Cependant, Biek *et al.*, 2012 ont estimé 0,15 95% HPD : [0,04 - 0,26] substitutions/génome/an pour *M. bovis* au Royaume-Uni et Crispell *et al.*, 2017, 0,53 95% HPD : [0,22 – 0,94] substitutions/génome/an en Nouvelle-Zélande.

Les différences entre ces résultats pourraient avoir plusieurs causes : le taux de substitution pourrait dépendre de la population échantillonnée, notamment de l'espèce hôte (Kao *et al.*, 2016). Ces études portent toutes sur des souches isolées à la fois chez des bovins et dans la faune sauvage. Cependant, le nombre de souches isolées dans la faune sauvage et chez des bovins varient selon les études : 66 vs. 234 (22%) pour Modenesi, 2019 et 6 vs. 139 (0,04%) pour Trewby *et al.*, 2016. Dans

notre étude on se situe dans la fourchette haute et on a ainsi une assez bonne représentation de la faune sauvage avec 21 vs. 146 (12,6%) des échantillons collectés chez des blaireaux. De plus, les espèces de la faune sauvage étudiées diffèrent, on peut citer le cerf de Virginie et l'élan (Salvador *et al.*, 2019) ou encore le blaireau comme dans notre étude, celle de Biek *et al.*, 2012, de Trewby *et al.*, 2016 et de Modenesi, 2019 qui a également pris en compte les données provenant de sangliers.

Les différences entre ces études pourraient également dépendre de la lignée étudiée de *M. bovis* (Crispell *et al.*, 2017), qui elle-même peut dépendre de la région étudiée (Hauer, 2015) ; par exemple, en Dordogne la totalité des souches était de spoligotype BCG appartenant au complexe clonal Eu3 et notre estimation du taux de substitution était basé sur des souches de spoligotype F7 faisant partie de la famille F4 (Hauer *et al.*, 2019).

Notre approche nous a permis d'estimer des taux de transition inter-espèce qui correspond à la probabilité qu'une souche passe d'une espèce hôte à une autre au cours d'une année. En ce qui concerne cette transition entre espèces hôtes aucun modèle n'a pu être sélectionné, le modèle asymétrique prédit un taux de transition de blaireau à bovin 6.7 fois supérieur au taux de transition de bovin à blaireau. Le modèle de transition symétrique quant à lui prédit un taux de transition très faible entre les deux populations.

Ces résultats diffèrent de ceux trouvés pour la Dordogne dont le modèle avait inclus les données des sangliers. Dans cette étude, le modèle avec une transition asymétrique entre les bovins et la faune sauvage prédisait des taux de transition plus élevés variant entre 0,27 transitions/souche/an (du bovin au sanglier) et 1,5 transitions/souche/an (du sanglier au blaireau, résultats non publiés) et un rôle prédominant du sanglier.

Le fait que les souches de spoligotype F7 ont commencé à être mises en évidence plus tôt chez les bovins que dans la faune sauvage ne signifie pas forcément que ces souches circulaient chez les bovins en premier. Le suivi plus tardif de la faune sauvage pourrait avoir un impact sur la reconstruction de l'arbre. Un biais temporel semblable avait été identifié dans l'étude de Crispell *et al.*, 2017, les souches provenant de la faune sauvage étaient plus nombreuses par rapport à celles provenant de bovins au début de la période d'étude ; puis au cours du temps, la tendance s'était inversée. Les auteurs ont donc conclu au fait que la direction de transmission de la maladie observée (de la faune sauvage vers les bovins) pouvait être le reflet de la réalité ou causée par un artefact dû à la stratégie d'échantillonnage.

Dans notre cas, en suivant ce raisonnement on s'attendrait à ce que le MRCA ait infecté un bovin du simple fait de ce biais. L'espèce-hôte du MRCA de l'arbre phylogénétique prédite est un bovin dans le modèle symétrique mais reste incertaine dans le modèle asymétrique. Cela peut s'expliquer par le fait que dans le modèle symétrique, le taux de transition/souche/an étant le même d'une espèce à une autre, la reconstruction de l'état des nœuds ancestraux dépend davantage du nombre de souches échantillonnées dans chaque population et dans notre cas, ce nombre est plus important pour la population de bovins. De même, dans le modèle asymétrique, le taux de transition des blaireaux aux bovins étant nettement supérieur, le modèle prédit donc plus de sauts inter-espèces du blaireau au bovin que le modèle précédent mais le faible nombre de souches isolées entre 2013 et 2017 chez des blaireaux conduit à plus d'incertitude dans cette reconstruction.

Une étude des données de mouvements de blaireaux en Europe a estimé qu'une majorité de déplacements est inférieure à 5 km avec de rares déplacements sur de plus longues distances (jusqu'à 22 km) (Byrne *et al.*, 2014). Nos données de localisation de souches montrant la présence de souches de blaireaux à moins de 5 km des souches bovines concordent avec la possibilité de phénomènes de transmission entre ces deux espèces. De plus, une étude portant sur la distribution de la tuberculose

bovine dans les élevages de cette région conclut à un rôle majeur du blaireau dans la diffusion de la maladie (avec une part non négligeable du rôle du commerce entre élevages) (Bouchez-Zacria *et al.*, 2018).

Un clade identifié à partir de l'inférence bayésienne et constitué de souches de spoligotype F7 isolées à la fois chez des bovins et des blaireaux, présente un regroupement des souches isolées chez des blaireaux d'un côté (à l'Ouest) et de celles isolées chez des bovins de l'autre (à l'Est). Cette répartition géographique et l'arbre phylogénétique est cohérent avec un scénario où un seul phénomène de transition aurait eu lieu entre la population de blaireaux et celle de bovins puis l'agent pathogène aurait circulé séparément dans les deux populations. Le taux de cette transmission intra-espèce n'a pas été estimé dans notre modèle.

Le modèle de diffusion spatiale discret n'a pas été sélectionné, cela pourrait être dû au fait que la maladie est d'évolution lente et que notre zone d'étude (de 1610 km<sup>2</sup> après mise en place de la grille) est relativement restreinte. L'échelle de temps sur laquelle la maladie a été étudiée (16 ans) voire la taille de la zone pourrait ne pas être suffisant pour mettre en évidence cette diffusion, si elle existe.

Pour les données géographiques implémentées dans ce modèle de diffusion, le lieu de contamination étant inconnu, la commune avec le maximum de pâtures n'était qu'une approximation du lieu du contact infectieux. Cependant, dans les clades ne possédant que des souches isolées chez des bovins, certaines souches géographiquement éloignées des autres provenaient d'élevages ne possédant pas de parcelles plus proches. Il serait intéressant de vérifier sur la base de données nationale d'identification (BDNI) si des déplacements de bovins appartenant aux exploitations contaminées ont eu lieu entre élevages de la zone. Or, notre modèle ne prenait pas en compte les mouvements de bovins et restreignait le domaine de vie des animaux de la faune sauvage à un seul point. Les limites de ce modèle pourraient contribuer à l'absence de mise en évidence d'une diffusion spatiale.

Cette étude a donc permis d'intégrer des données génomiques dans l'étude du rôle du blaireau dans la transmission de *M. bovis* aux bovins mais ces données ont leurs limites. En effet, il existe des disparités entre le recueil des données chez les bovins et celui dans la faune sauvage. Le rythme de dépistage des bovins varie selon la commune et le département mais ce dépistage concerne toutes les exploitations bovines. A ce dépistage, s'ajoute le fait que les carcasses subissent systématiquement une recherche de lésions tuberculeuses même si cette surveillance possède une moins bonne sensibilité. En théorie, la totalité de la population bovine âgée de plus de 24 mois est donc surveillée et testée de manière régulière.

De plus, le test de dépistage utilisé chez les animaux vivants est principalement l'IDS (les élevages laitiers étant minoritaires dans cette région) jusqu'en 2016, quand la décision de privilégier l'IDC a été prise. De cette dichotomie, on pourrait donc s'attendre à une sous-détection des cas dans les élevages utilisant l'IDC par rapport à ceux utilisant l'IDS. Cependant, en pratique l'IDC serait mieux réalisée et donc plus sensible que l'IDS. En effet, la note de service du 5 décembre 2017 (DGAL/SDSPA/2017-979) met l'accent sur le fait que dans les zones où l'IDC a remplacé l'IDS, le taux de déclaration de résultats non négatifs a été doublé. Globalement, ces tests ont une sensibilité médiocre (IDS : Se [min : 58,6%- max : 91%] et IDC : Se d'environ 50%) et des performances variables selon les conditions de réalisation et peuvent donc conduire à une sous-détection des cas chez les bovins.

Les protocoles de piégeage des blaireaux varient selon les années et les zones étudiées, la totalité des blaireaux ne peut être surveillée pour des raisons pratiques et financières. De plus, la

dégradation et la contamination possibles des carcasses dans l'environnement (contrairement aux conditions de conservation des prélèvements effectués en abattoir) diminuent la sensibilité des tests (PCR et culture bactérienne) (Rivière *et al.*, 2015) et rendent ainsi une sous-détection de cas provenant de la faune sauvage inévitable. Dans notre étude, la sous-détection des cas chez les blaireaux pose un problème lors de la reconstruction de l'arbre, une partie de l'histoire évolutive nous manque et cela pourrait conduire à une mauvaise estimation du nombre de phénomènes de transition entre bovins et blaireaux.

Le nombre de souches de spoligotype F7 isolées chez des sangliers étant trop restreint, leur inclusion dans l'analyse phylogénétique par inférence bayésienne n'a pas été possible contrairement à l'étude en Dordogne (Modenesi, 2019). La possibilité de phénomènes de transmission entre sangliers et bovins et/ou blaireaux a été mise en évidence dans l'arbre construit par maximum de parcimonie. Cette espèce joue donc un rôle dans le système multi-hôte de la tuberculose bovine dans les Pyrénées-Atlantiques et les Landes, même si l'importance et la nature de ce rôle restent inconnues. Il serait intéressant de suivre l'évolution du nombre de cas mis en évidence chez les sangliers et de voir s'il est possible d'inclure ces souches pour mieux caractériser le rôle de cette espèce dans ce système multi-hôte.

## VI) Conclusion

Nous avons étudié un système multi-hôte de la tuberculose bovine à partir des données de séquençage obtenues dans une zone d'étude située à cheval entre les Pyrénées-Atlantiques et les Landes. Ce système multi-hôte met en jeu au moins trois espèces : les blaireaux, les sangliers et les bovins dont nous avons pu isoler pour analyse 174 souches de spoligotype F7 et 46 de spoligotype F15.

Nous avons réalisé par inférence bayésienne la reconstruction d'arbres phylogénétiques. Dans ces arbres, des phénomènes de transition entre la faune sauvage et les bovins ont pu être mis en évidence mais le modèle de transition (symétrique vs. asymétrique) n'a pu être sélectionné. Le rôle du blaireau n'a donc pas été élucidé et les données provenant de sangliers n'ont pas été incluses. Cependant, que ce soit dans le modèle symétrique ou asymétrique, les événements prédits de transition du blaireau au bovin restent relativement rares. Cela pourrait évoluer à terme si la prévalence de la maladie augmente dans la faune sauvage.

Pour aller plus loin dans la compréhension de ce système multi-hôte, il serait nécessaire de pouvoir intégrer plus de données notamment les données de séquençage depuis 2017, en particulier ceux provenant de la faune sauvage dans une approche bayésienne. De plus, l'intégration des données de mouvements et la reconstitution de l'histoire de transmission permettraient d'élucider la dynamique de la transmission dans la zone d'étude.

## Bibliographie

- BAUM D.A., OFFNER S. (2008) Phylogenies & Tree-Thinking. *Am. Biol. Teach.* 70(4), 222-229
- BIEK R., O'HARE A., WRIGHT D., *et al.* (2012) Whole Genome Sequencing Reveals Local Transmission Patterns of *Mycobacterium bovis* in Sympatric Cattle and Badger Populations. *PLOS Pathog.* 8(11), e1003008
- BONNIN R.A., NAAS T., DORTET L. (2017) Impact du séquençage d'ADN à haut débit sur la surveillance des épidémies de bactéries multi-résistantes aux antibiotiques. *Feuill. Biol.* , 14
- BOUCHEZ-ZACRIA M., COURCOUL A., JABERT P., RICHOMME C., DURAND B. (2017) Environmental determinants of the *Mycobacterium bovis* concomitant infection in cattle and badgers in France. *Eur. J. Wildl. Res.* 63
- BOUCHEZ-ZACRIA M., COURCOUL A., DURAND B. (2018) The Distribution of Bovine Tuberculosis in Cattle Farms Is Linked to Cattle Trade and Badger-Mediated Contact Networks in South-Western France, 2007–2015. *Front. Vet. Sci.* 5, 173
- BOUCKAERT R., VAUGHAN T.G., BARIDO-SOTTANI J., *et al.* (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* 15(4), e1006650
- BYRNE A.W., QUINN J.L., O'KEEFE J.J., *et al.* (2014) Large-scale movements in European badgers: has the tail of the movement kernel been underestimated? *J. Anim. Ecol.* 83(4), 991-1001
- CRISPELL J., ZADOKS R.N., HARRIS S.R., *et al.* (2017) Using whole genome sequencing to investigate transmission in a multi-host system: bovine tuberculosis in New Zealand. *BMC Genomics* 18(1), 180
- CRISPELL J., BENTON C.H., BALAZ D., *et al.* (2019) Combining genomics and epidemiology to analyse bi-directional transmission of *Mycobacterium bovis* in a multi-host system. *eLife* 8
- DELAVENTE C., PANDOLFI F., GIRARD S., *et al.* (2019) Tuberculose bovine: Bilan et évolution de la situation épidémiologique entre 2015 et 2017 en France Métropolitaine. *Bull. Épidémiologique* , 22
- DESVAUX S., RÉVEILLAUD É., RICHOMME C., *et al.* (2019) Sylvatub: Bilan 2015-2017 de la surveillance de la tuberculose dans la faune sauvage. *Bull. Épidémiologique* , 10
- DOS REIS M., DONOGHUE P.C.J., YANG Z. (2016) Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* 17(2), 71-80
- DRUMMOND A.J., HO S.Y.W., PHILLIPS M.J., RAMBAUT A. (2006) Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* 4(5)
- DRUMMOND A.J., BOUCKAERT R.R. (2015) Bayesian Evolutionary Analysis with BEAST. Cambridge University Press
- FELSENSTEIN J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17(6), 368-376

- GARNIER T., EIGLMEIER K., CAMUS J.-C., *et al.* (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc. Natl. Acad. Sci. U. S. A.* 100(13), 7877-7882
- GU X., FU Y.X., LI W.H. (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 12(4), 546-557
- HAMMING R. (1950) Error-detecting and error-correcting codes. *Bell Syst. Tech. J.* 29(2), 147-160
- HARS J., RICHOMME C. (2010) La tuberculose bovine dans la faune sauvage en France. *Bull. Épidémiologique*, 5
- HASEGAWA M., KISHINO H., YANO T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22(2), 160-174
- HASTINGS W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), 97-109
- HAUER A. (2015) Etude des souches de *Mycobacterium bovis* à l'origine de foyers de tuberculose bovine en France de 1978 à aujourd'hui : une approche moléculaire et génomique. thesis. Tours
- HAUER A., MICHELET L., COCHARD T., *et al.* (2019) Accurate Phylogenetic Relationships Among *Mycobacterium bovis* Strains Circulating in France Based on Whole Genome Sequencing and Single Nucleotide Polymorphism Analysis. *Front. Microbiol.* 10
- HELED J., BOUCKAERT R.R. (2013) Looking for trees in the forest: summary tree from posterior samples. *BMC Evol. Biol.* 13, 221
- HUELSENBECK J.P., RANNALA B., MASLY J.P. (2001) An Introduction to Bayesian Inference of Phylogeny. *Science*, 7
- KAMERBEEK J., SCHOOLS L., KOLK A., *et al.* (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* 35(4), 907-914
- KAO R.R., PRICE-CARTER M., ROBBE-AUSTERMAN S. (2016) Use of genomics to track bovine tuberculosis transmission. *Rev. Sci. Tech. Int. Off. Epizoot.* 35(1), 241-258
- KEAN J.M., BARLOW N.D., HICKLING G.J. (1999) Evaluating potential sources of bovine tuberculosis infection in a New Zealand cattle herd. *N. Z. J. Agric. Res.* 42(1), 101-106
- KECK N., MOYEN J.-L., GUENEAU E., BOSCHIROLI M.-L. (2014) Particular features of screening and diagnosis of bovine tuberculosis. *Epidémiologie Santé Anim.* 65, 5-19
- KIMURA M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16(2), 111-120
- LEE J., SUNG W., CHOI J.-H. (2015) Metamodel for Efficient Estimation of Capacity-Fade Uncertainty in Li-Ion Batteries for Electric Vehicles. *Energies* 8, 5538-5554
- LEMEY P., RAMBAUT A., DRUMMOND A.J., SUCHARD M.A. (2009) Bayesian Phylogeography Finds Its Roots. *PLoS Comput. Biol.* 5(9)

- LI W.H. (1993) So, what about the molecular clock hypothesis? *Curr. Opin. Genet. Dev.* 3(6), 896-901
- MARSOT M., BÉRAL M., SCOIZEC A., *et al.* (2016) Herd-level risk factors for bovine tuberculosis in French cattle herds. *Prev. Vet. Med.* 131, 31-40
- MATURANA P., BREWER B.J., KLAERE S., BOUCKAERT R. (2019) Model selection and parameter inference in phylogenetics using Nested Sampling. *Syst. Biol.* 68(2), 219-233
- MICHELET L., DE CRUZ K., HÉNAULT S., *et al.* (2018) Mycobacterium bovis Infection of Red Fox, France. *Emerg. Infect. Dis.* 24(6), 1151-1153
- MICHELET L., DURAND B., BOSCHIROLI M.L. (2019) Tuberculose bovine: Bilan génotypique de M. bovis à l'origine des foyers bovins entre 2015 et 2017 en France Métropolitaine. *Bull. Épidémiologique* , 9
- MODENESI G. (2019) Epidémiologie moléculaire de la tuberculose bovine –Rôle de la faune sauvage et de bovins dans la transmission en Dordogne et Haute-Vienne. Surveillance épidémiologique des maladies humaines et animales. Maisons-Alfort, ENVA
- MÜLLER N.F., RASMUSSEN D.A., STADLER T. (2017) The Structured Coalescent and Its Approximations. *Mol. Biol. Evol.* 34(11), 2970-2981
- NARANJO V., GORTAZAR C., VICENTE J., DE LA FUENTE J. (2008) Evidence of the role of European wild boar as a reservoir of Mycobacterium tuberculosis complex. *Vet. Microbiol.* 127(1-2), 1-9
- O'REILLY L.M., DABORN C.J. (1995) The epidemiology of Mycobacterium bovis infections in animals and man: a review. *Tuber. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis.* 76 Suppl 1, 1-46
- PYBUS O.G., RAMBAUT A., HARVEY P.H. (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155(3), 1429-1437
- RAMBAUT A., DRUMMOND A.J., XIE D., BAELE G., SUCHARD M.A. (2018) Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67(5), 901-904
- RÉVEILLAUD É. (2013) Sylvatub un dispositif de surveillance de la tuberculose bovine dans la faune sauvage. *Nouv. Prat. Vét.* 6(25), 8
- RIVIÈRE J., FAURE E., HENDRIKX P., FEDIAEVSKY A. (2013) Sylvatub : bilan d'une première année de surveillance de la tuberculose bovine dans la faune sauvage en France. *Bull. Épidémiologique Santé Anim. Aliment.* , 6
- RIVIÈRE J., LE STRAT Y., DUFOUR B., HENDRIKX P. (2015) Sensitivity of Bovine Tuberculosis Surveillance in Wildlife in France: A Scenario Tree Approach. *PLoS ONE* 10(10)
- ROBERT M., KAAS R.E., BRADLEY P., *et al.* (1997) Markov Chain Monte Carlo in Practice: A Roundtable Discussion
- RORING S., SCOTT A., BRITAIN D., *et al.* (2002) Development of Variable-Number Tandem Repeat Typing of Mycobacterium bovis: Comparison of Results with Those Obtained by Using Existing Exact Tandem Repeats and Spoligotyping. *J. Clin. Microbiol.* 40(6), 2126-2133

- SALVADOR L.C.M., O'BRIEN D.J., COSGROVE M.K., *et al.* (2019) Disease management at the wildlife-livestock interface: Using whole-genome sequencing to study the role of elk in *Mycobacterium bovis* transmission in Michigan, USA. *Mol. Ecol.* 28(9), 2192-2205
- SCHLIEP K.P. (2011) phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4), 592-593
- SKUCE R.A., ALLEN A.R., MCDOWELL S.W.J. (2012) Herd-Level Risk Factors for Bovine Tuberculosis: A Literature Review. *Vet. Med. Int.* 2012
- TREWBY H., WRIGHT D., BREADON E.L., *et al.* (2016) Use of bacterial whole-genome sequencing to investigate local persistence and spread in bovine tuberculosis. *Epidemics* 14, 26-35
- YANG Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* 39(3), 306-314
- YANG Z., RANNALA B. (2012) Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* 13(5), 303-314
- ZHANG J., YANG J.-R. (2015) Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* 16(7), 409-420

## Liste des annexes

Annexe 1 : Extrait de l'arrêté du 15 septembre 2003 (fixant les mesures techniques et administratives relatives à la prophylaxie collective et à la police sanitaire de la tuberculose des bovinés et des caprins).....	50
Annexe 2 : Extrait de la note de service (DGAL/SDSPA/N2006-8051) .....	51
Annexe 3 : Répartition en fonction des départements des niveaux de surveillance de la tuberculose bovine dans la faune domestique et sauvage en France .....	52
Annexe 4 : Le dispositif Sylvatub, Modalités de surveillance en fonction des niveaux de surveillance (note de service DGAL/SDSPA/2018-708).....	53
Annexe 5 : Extrait de l'arrêté préfectoral n°645/2016 portant sur la surveillance des blaireaux autour des foyers de tuberculose bovine dans les Landes. ....	54
Annexe 6 : Un exemple de typage VNTR.....	55
Annexe 7 : Méthode de séquençage Illumina (d'après Illumina.com).....	56
Annexe 8 : Modèles de substitution et matrices associées .....	57
Annexe 9 : Principe de l'algorithme de « Nested Sampling » permettant l'estimation de la vraisemblance marginale. ....	58
Annexe 10 : Le modèle de diffusion spatiale discret pour les souches de spoligotype F7 représentée sur la zone d'étude restreinte à une grille de 1610 km <sup>2</sup> (les deux souches les plus éloignées n'ont pas été prise en compte). ....	60
Annexe 11 : Correspondance entre les spoligotypes présents dans la zone d'étude et leur code binaire (d'après Mbovis.org). ....	61
Annexe 12 : Profils VNTR identifiés dans la zone d'étude.....	61
Annexe 13 : Carte permettant la localisation des pâtures des élevages dont les souches étaient éloignées ainsi que la localisation de la souche porcine (point rouge).....	62
Annexe 14 : Sélection du modèle évolutif.....	63
Annexe 15 : Tableaux récapitulatif et comparatif des clades selon les méthodes de reconstruction utilisées : à savoir l'inférence bayésienne avec un modèle asymétrique (ASYM) et symétrique (SYM) ainsi que le maximum de parcimonie (MP). Les couleurs des cases correspondent aux couleurs des clades dans les arbres reconstruits par inférence bayésienne. Les couleurs des cases pour les clades MP sont arbitraires. Les cases SANS dans le modèle asymétrique correspondent aux souches qui sont entourées en jaune dans l'arbre. ....	64

Chapitre I : Dispositions générales

**Article 6**

Le préfet, sur proposition du directeur départemental en charge de la protection des populations et après accord du ministre chargé de l'agriculture (sous-direction de la santé et de la protection animales), peut prendre toutes **dispositions complémentaires** aux mesures définies dans le présent arrêté afin de rendre plus efficiente la protection des élevages et de la santé publique à l'égard de la tuberculose. Il prescrit des **mesures renforcées de surveillance** notamment **vis-à-vis des troupeaux présentant un risque sanitaire particulier à l'égard de la tuberculose** soit en raison d'un risque d'exposition accru, soit en raison d'un risque particulier pour la santé publique ou la santé animale.

Dans ces troupeaux, il peut prescrire un **rythme de dépistage supérieur** à celui des autres troupeaux du département et des obligations de dépistage lors du mouvement des animaux.

Il peut également demander un **dépistage des animaux d'autres espèces sensibles** à la tuberculose détenus de façon non distincte du troupeau de bovinés.

Sont susceptibles de présenter un **risque sanitaire particulier** à l'égard de la tuberculose :

a) Les troupeaux **ayant retrouvé leur qualification** après avoir été reconnus atteints de tuberculose pendant une **durée maximale de dix ans**;

b) Les troupeaux pour lesquels un **lien épidémiologique à risque** a été constaté avec un animal ou un troupeau atteint de tuberculose;

c) Les troupeaux pour lesquels un **lien épidémiologique à risque** est constaté avec un foyer confirmé de tuberculose dans la faune sauvage;

d) Les troupeaux pour lesquels il est établi que des **dispositions réglementaires** relatives à l'identification ou à la circulation des animaux ou aux conditions de maintien de la qualification "officiellement indemne" de tuberculose **n'ont pas été respectées**;

e) Les troupeaux **livrant directement au consommateur du lait cru ou des produits au lait cru**;

f) Les troupeaux fournissant des animaux participant à la **monte publique naturelle ou artificielle**;

g) Les **troupeaux présentés au public**.

l) Dérogation au dépistage brucellose et tuberculose lors de mouvements de bovins

La dérogation aux tests de dépistage de la brucellose et de la tuberculose bovines lors de mouvements entre exploitations est applicable pour les bovins provenant de **cheptels officiellement indemnes** de brucellose et de tuberculose bovines, à condition que la durée du transfert entre la sortie de l'exploitation d'origine et l'introduction dans l'exploitation de destination **n'excède pas 6 jours**. Ces bovins ne sont donc pas soumis à des contrôles de dépistage lors de mouvements.

Cependant les bovins **provenant d'exploitations** présentant un **risque sanitaire particulier** ou **introduits dans un cheptel dont le taux de rotation annuel supérieur à 40 %** et en **provenance d'un département dont la prévalence de la tuberculose cumulée sur 5 ans est supérieure à la moyenne nationale** sont exclus du bénéfice de la dérogation, quel que soit le délai de livraison.

[...]

Annexe 1 (Extrait) : Liste des exploitations de bovins à risque, exclues des dérogations aux contrôles lors de mouvement entre exploitations.

Risques sanitaires spécifiques
<b>1- Risque de résurgence</b>
- 5 ans après abattage total du cheptel infecté - 10 ans après abattage partiel du cheptel infecté
<b>2- Lien épidémiologique par voisinage avec un cheptel infecté</b> <i>Sont concernés les exploitations identifiées par l'enquête épidémiologique conduite par la DDPP après confirmation de l'infection</i>
- maintien des contrôles lors de mouvements entre exploitations pour les bovins issus du cheptel tant que celui-ci est soumis à des mesures de dépistage annuel (maximum de trois ans) sauf pour les animaux destinés à l'engraissement ou à l'abattage direct.
<b>3- Risque lié à la faune sauvage</b>
- existence de cas confirmés de tuberculose sur des animaux de la faune sauvage dans le département (ou à proximité dans un département limitrophe) et - évaluation locale du risque par la DDPP, notamment en fonction de la localisation des élevages et de la probabilité des contacts entre faune sauvage et bovins.

Annexe 3 : Répartition en fonction des départements des niveaux de surveillance de la tuberculose bovine dans la faune domestique et sauvage en France

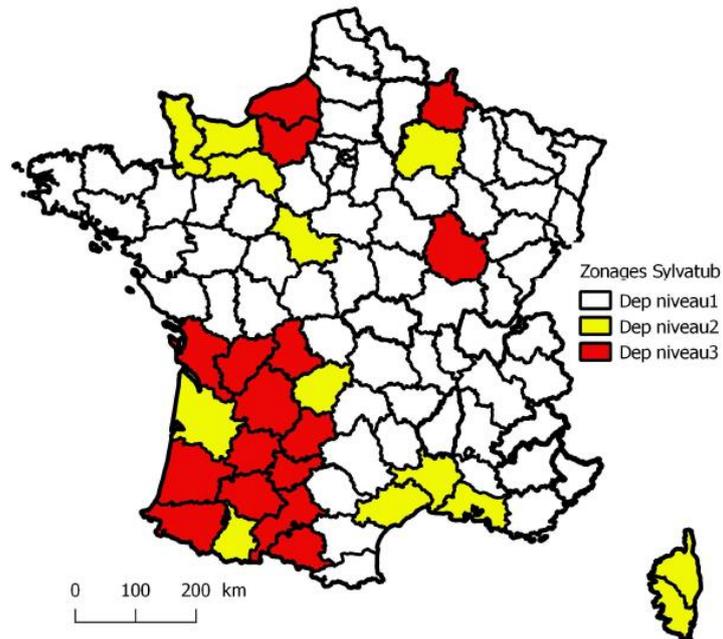


Figure 1 : Zonages des niveaux de surveillance en France du dispositif Sylvatub (note de service DGAL/SDSPA/2018-699)

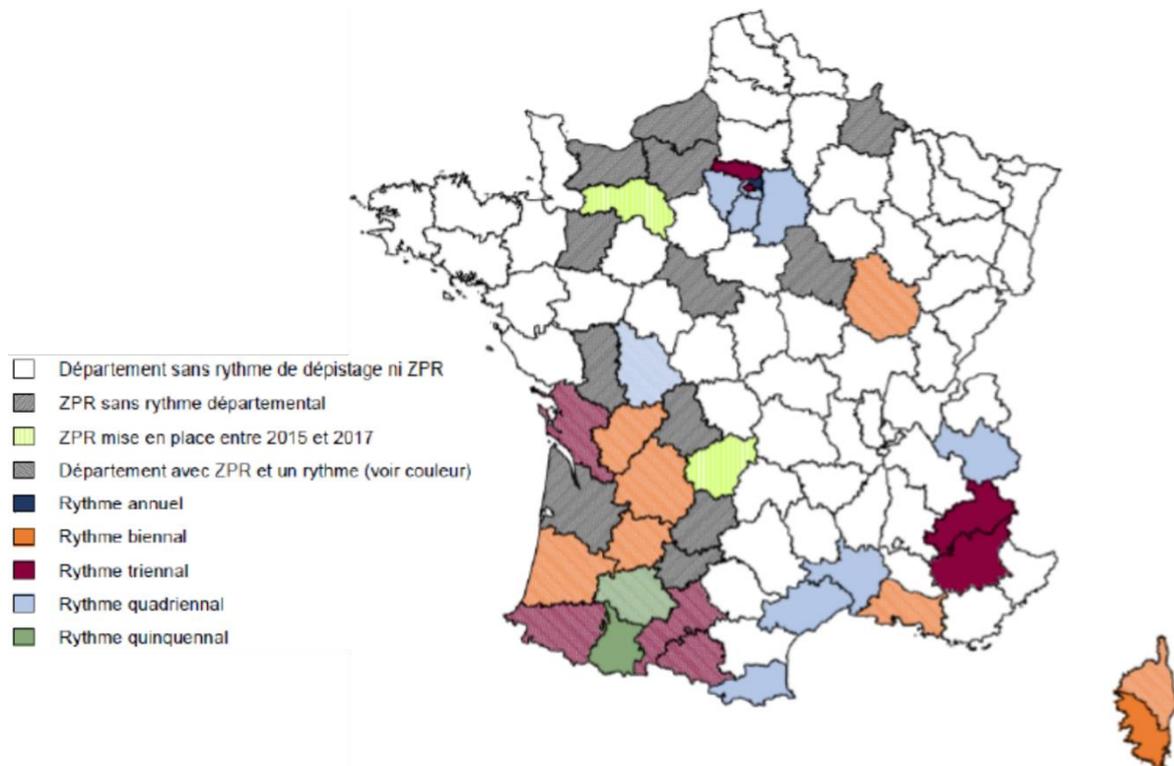


Figure 2 : Rythmes de dépistage de la tuberculose bovine en élevage selon les départements de 2015 à 2017 (Delavenne et al., 2019)

Annexe 4 : Le dispositif Sylvatub, Modalités de surveillance en fonction des niveaux de surveillance (note de service DGAL/SDSPA/2018-708)

<b>Type de surveillance</b>	<b>Modalité de surveillance</b>	<b>Niveau 1</b>	<b>Niveau 2</b>	<b>Niveau 3</b>
<b>Evènementielle</b>	<i>Recherche de lésions suspectes chez les cervidés et sangliers lors de l'examen de carcasse dans le cadre d'une pratique de chasse habituelle</i>	X	X	X
	<i>Recherche de lésions évocatrices de tuberculose chez les sangliers, cervidés et blaireaux collectés dans le cadre du réseau Sagir (animaux morts ou mourants) dans son fonctionnement normal</i>	X	X	X
<b>Evènementielle renforcée</b>	<i>Recherche analytique systématique de tuberculose chez les sangliers, cerfs et blaireaux collectés dans le cadre du renforcement du réseau Sagir</i>		X	X
	<i>Recherche analytique systématique de tuberculose chez les cadavres de blaireaux collectés sur les routes dans le cadre du renforcement réseau Sagir. Ce renforcement des analyses doit s'accompagner d'un renfort de collecte sur l'ensemble des zones de prospection et des zones tampon.</i>		X	X
<b>Programmée</b>	<i>Recherche systématique de tuberculose sur un échantillon de blaireaux prélevés dans les zones infectées de la zone à risque ou en zone de prospection</i>		X	X
	<i>Recherche systématique de tuberculose sur un échantillon de sangliers prélevés sur l'ensemble de la zone à risque.</i>			X

Annexe 5 : Extrait de l'arrêté préfectoral n°645/2016 portant sur la surveillance des blaireaux autour des foyers de tuberculose bovine dans les Landes.

**ARTICLE 1er :** Zones de prélèvements

Des opérations de prélèvement de blaireaux sont ordonnées afin de dépister, sur les individus prélevés, la présence de *Mycobacterium bovis*, agent responsable de la tuberculose bovine.

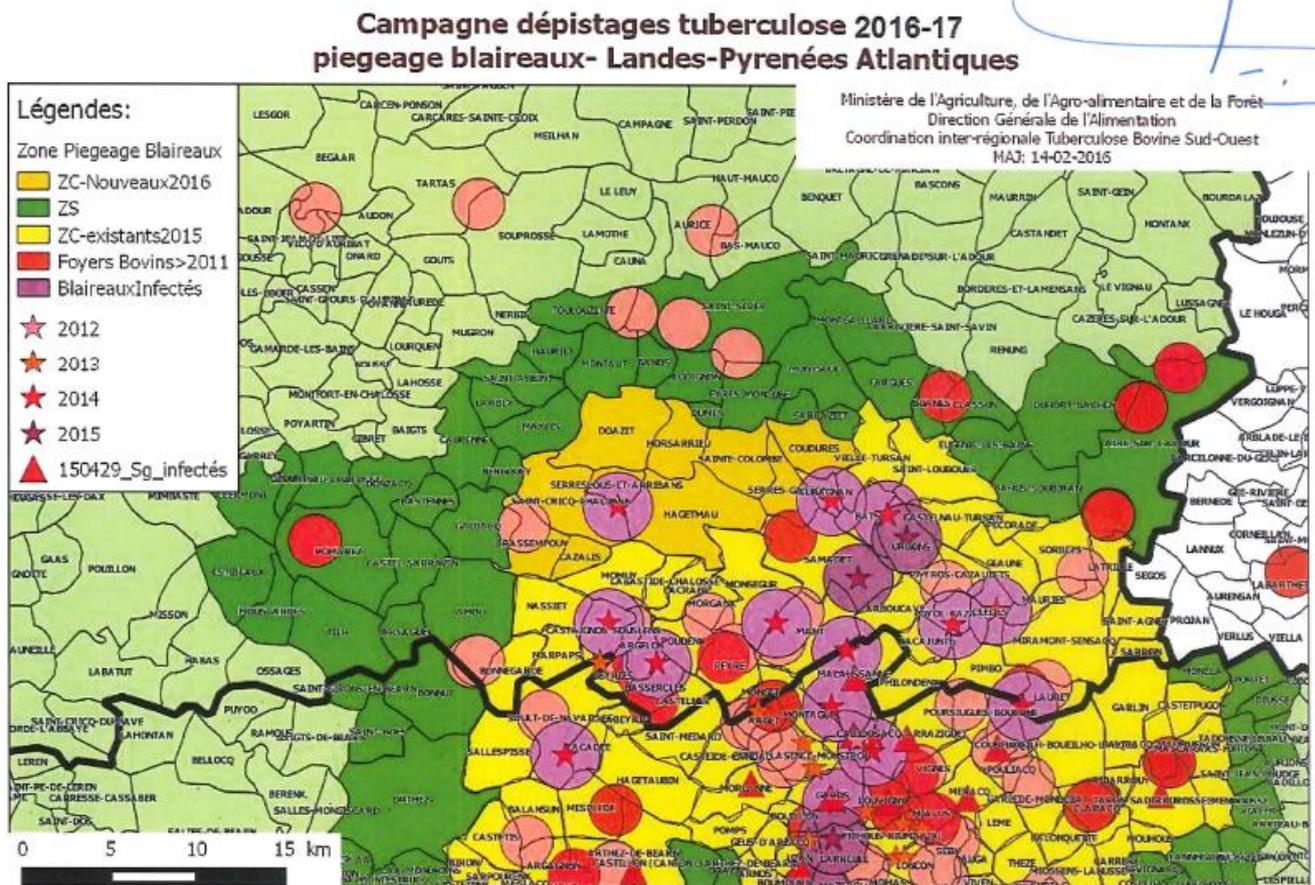
À cette fin, trois types de zones concernées par ces opérations sont définies :

→Zones d'infection : aires d'un rayon de 1,5 Km de rayon aux abords immédiats des exploitations bovines atteintes depuis moins de 6 ans et de 2 Km de rayon autour des terriers trouvés infectés de tuberculose depuis 2013, sur la base des éléments épidémiologiques recueillis par la direction départementale de la protection des populations ;

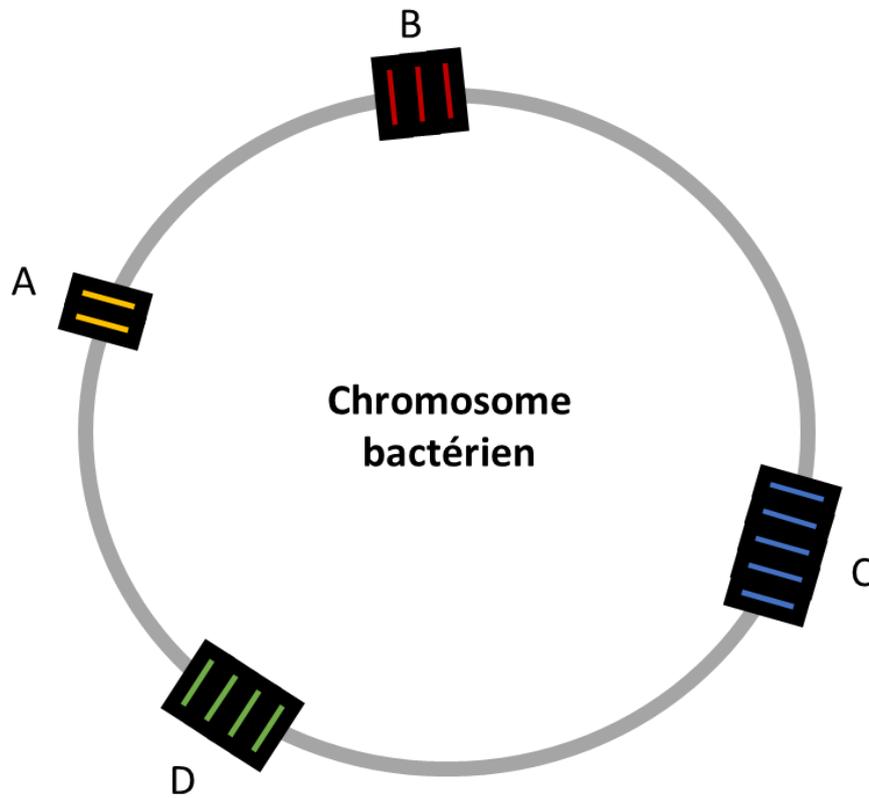
→Zones de contrôle : l'ensemble des communes dont une partie du territoire se situe dans un rayon approximatif de 3 Km autour des zones d'infection sus-définies, ainsi que des zones d'infection des années précédentes ; Une commune située en zone de contrôle dont plus de 75% de son territoire est constitué de zone d'infection acquière le statut de zone d'infection pour tout son territoire.

→Zones de surveillance : l'ensemble des communes dont une partie du territoire se situe dans un rayon approximatif de 2 à 5 Km autour des zones de contrôle sus-définies, en fonction des éléments épidémiologiques recueillis par la direction départementale de la protection des populations, non comprises les zones de contrôle elles-mêmes. Les différents périmètres sus-cités sont précisés par cartographie jointe en annexe 1 du présent arrêté.

Les différents périmètres sus-cités sont précisés par cartographie jointe en annexe 1.



Annexe 6 : Un exemple de typage VNTR



*Figure 1: Représentation schématique d'un chromosome bactérien. Quatre loci situés dans des régions polymorphiques et hypervariables du génome sont représentés par les encadrés A, B, C et D. Le nombre de répétitions associées à chaque locus (2 pour A, 3 pour B, 5 pour C et 4 pour D) est combiné pour générer un code numérique (2 3 5 4). Ce code est le profil VNTR.*

### I- Préparation de l'échantillon

L'échantillon à analyser est fragmenté en morceaux d'ADN double brin (db) aux extrémités desquels s'ajoutent des adaptateurs et des amorces. Les morceaux d'ADN double brin sont alors dénaturés en simple brin (sb), puis une de leurs extrémités se fixe (grâce aux adaptateurs) à un des deux types d'oligonucléotides présents à la surface de la « flow cell ».

### II- Génération de clusters

Une fois le fragment sb fixé à la « flow cell », une polymérase crée un brin complémentaire. Le fragment db nouvellement créé est dénaturé et la séquence d'origine est éliminée. Puis l'extrémité libre du fragment sb restant se fixe à l'autre type d'oligonucléotide à la surface de la « flow cell » (formation d'une figure « en pont »). Une polymérase crée alors un nouveau brin complémentaire (dit brin antisens), le fragment db formé est alors dénaturé et on obtient deux fragments sb. Ce processus (aussi appelé « bridge amplification ») est répété en boucle et a lieu simultanément sur toute la surface de la « flow cell ».

### III- Séquençage et alignement

A chaque amorce, quatre nucléotides marqués entrent en compétition pour la première base ; un seul est incorporé à la séquence complémentaire. Après incorporation du nucléotide, le cluster est soumis à une source lumineuse et le nucléotide incorporé émet une fluorescence dont la longueur d'onde est caractéristique de sa nature. Ce processus est répété jusqu'à obtenir le fragment de lecture qui est alors éliminé et on recommence à partir d'une autre amorce. Des brins sens et antisens étant séquencés, on obtient donc des fragments de lecture dans les deux sens.

La comparaison des fragments de lecture par des programmes bio-informatiques permet d'aligner les séquences qui se chevauchent. On obtient alors des séquences plus longues appelées contigs que l'on peut ensuite relier entre eux. Ce processus d'assemblage peut se faire à l'aide d'un génom de référence s'il en existe un, sinon cela s'appelle un séquençage de novo.

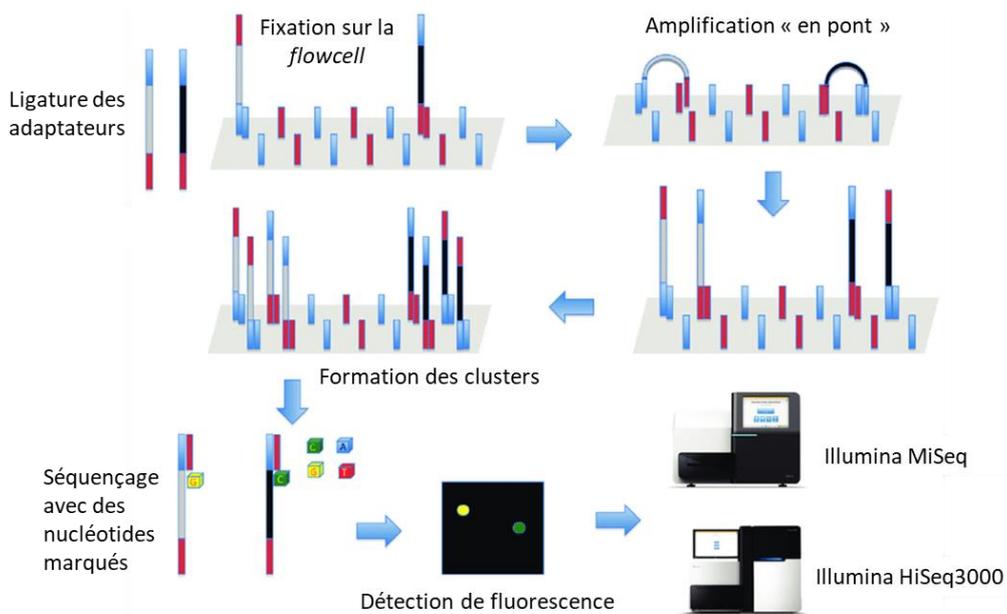


Figure 1 : Représentation schématique de la méthode Illumina (Bonnin et al., 2017)

Annexe 8 : Modèles de substitution et matrices associées

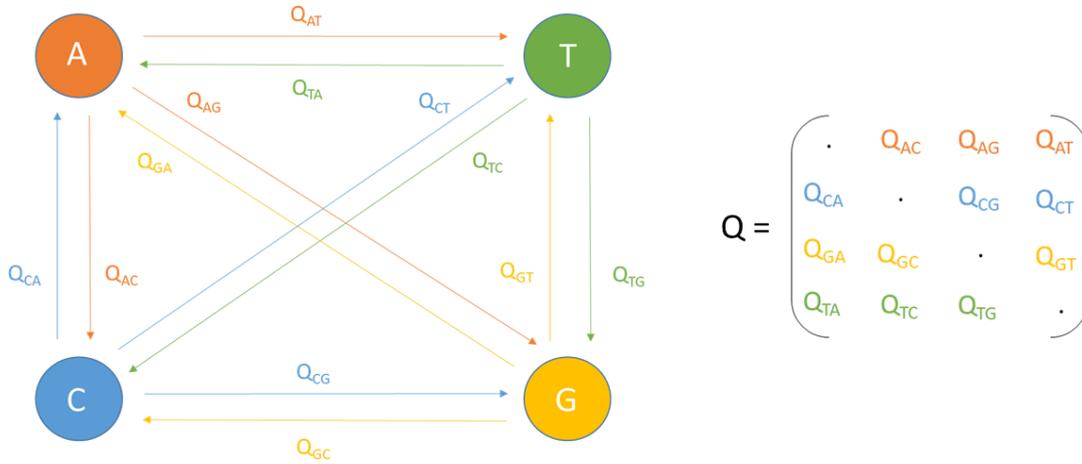


Figure 1 : Représentation schématique des substitutions possibles (à gauche) et la matrice générale associée (à droite). Les quatre cercles de couleurs représentent les quatre nucléotides (les purines, A et G, ainsi que les pyrimidines, C et T). La substitution d'un nucléotide vers l'autre est désignée par une flèche de la couleur du nucléotide d'origine. A droite, se trouve la matrice des taux de substitutions associée. Par exemple,  $Q_{GA}$  est le taux de substitution de G en A.

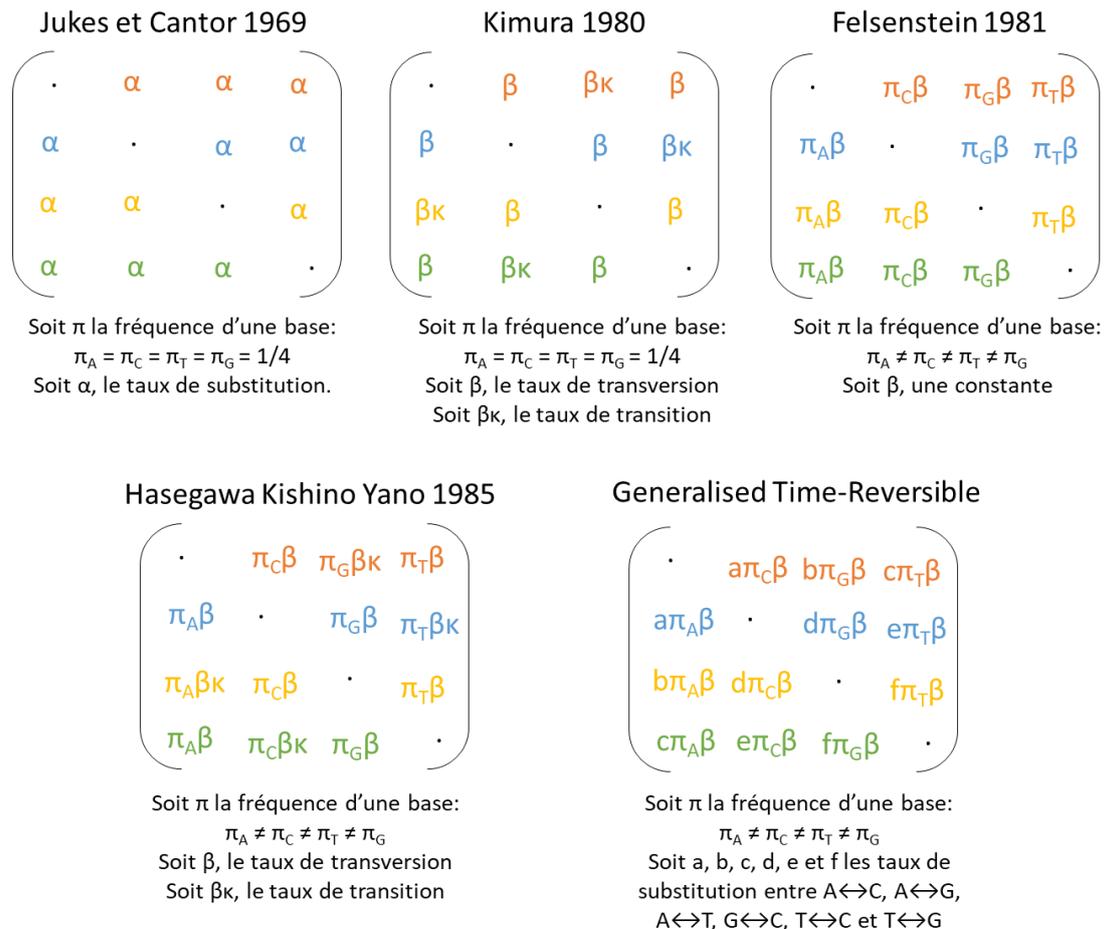


Figure 2 : Matrice de substitution  $Q$  associée à chaque modèle présenté : JC69, K80, F81, HKY et GTR. Ces modèles peuvent prendre en compte la fréquence d'une base désignée par  $\pi$  avec le nom de la base en indice. Une transition est une substitution entre deux pyrimidines ( $T \leftrightarrow C$ ) ou entre deux purines ( $A \leftrightarrow G$ ). Une transversion est une substitution entre deux bases de nature différente.

Annexe 9 : Principe de l'algorithme de « Nested Sampling » permettant l'estimation de la vraisemblance marginale.

La vraisemblance marginale (VM) représente la qualité moyenne de l'ajustement d'un modèle aux données et permet de comparer les modèles entre eux. La VM est la moyenne des vraisemblances sur la distribution *a priori*, on note :

$$VM = \int \pi(\theta).L(\theta)d\theta$$

Avec  $\theta$  un vecteur de paramètres,  $L(\theta)$  la vraisemblance et  $\pi(\theta)$  la distribution *a priori*.

Son calcul étant complexe, il est nécessaire d'approximer cette valeur avec des algorithmes comme celui du « Nested Sampling ».

Etape d'initialisation : Echantillonnage à partir de la distribution *a priori* de  $N$  points (aussi appelés particules)  $\theta_1, \dots, \theta_N$  (dont les vraisemblances respectives sont notées  $L(\theta_1), \dots, L(\theta_N)$ ).

Soit  $i \in \{1, \dots, N\}$  tel que  $L(\theta_i)$  est le minimum de  $\{L(\theta_1), \dots, L(\theta_N)\}$ . Alors le point  $\theta_i$  est mis de côté et remplacé par un nouveau point  $\theta$  échantillonné à partir de la distribution *a priori* sous la condition :

$$L(\theta) > L(\theta_i).$$

Répétition de ce processus jusqu'à satisfaire un critère d'arrêt. On obtient alors une séquence de vraisemblances croissantes ( $L_1, \dots, L_k$ ) et les points mis de côté ( $\theta_1, \dots, \theta_k$ ) correspondants qui vont contribuer à l'estimation de la VM.

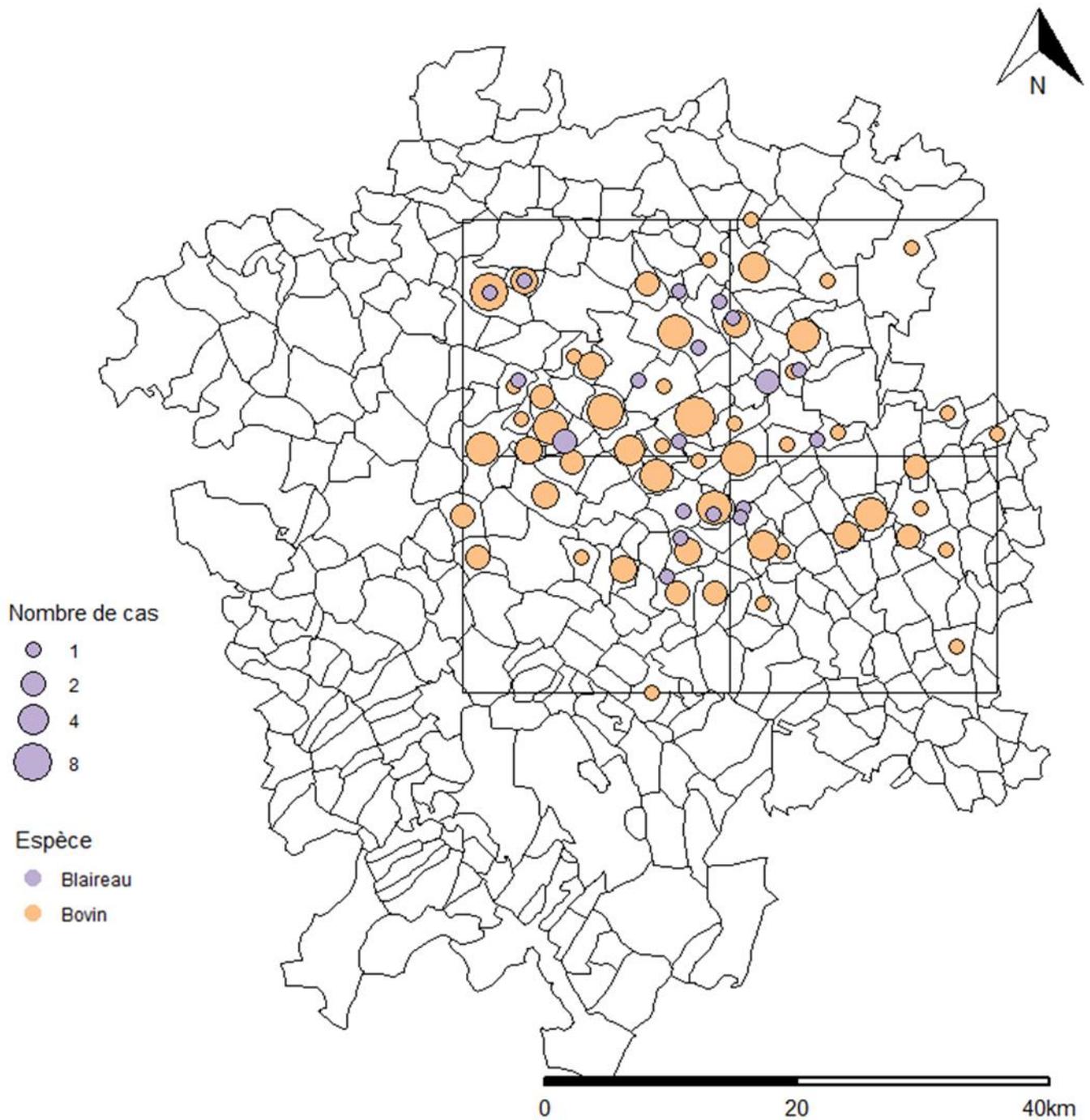
Deux paramètres sont à spécifier : d'une part,  $N$  qui détermine le nombre de particules permettant d'explorer l'espace de la distribution *a priori*. D'autre part, la longueur de la sous-chaîne définit le nombre d'étapes MCMC permettant de générer le nouveau point  $\theta$  à partir la distribution *a priori*.

Dans notre étude, nous avons choisi de fixer la longueur de la sous-chaîne à 100 000 et pour commencer,  $N$  était égal à 1 Si  $\log(\text{BF}) < 2\sqrt{(\text{SD}1^2 + \text{SD}2^2)}$ , alors la différence observée pouvait être due à la randomisation et il fallait augmenter le nombre de particules, sinon le résultat était interprétable selon le Tableau 1. Nous avons fixé arbitrairement un maximum pour  $N$  égal à 10, si aucune différence ne permettait d'être observée pour cette valeur, nous avons choisi le modèle cohérent avec la littérature ou le contexte épidémiologique.

1. Sample  $N$  points  $\theta_1, \dots, \theta_N$  from the prior  $\pi(\theta)$ ;
2. Initialise  $\mathcal{Z} = 0$  and  $\xi_0 = 1$ ;
3. Repeat for  $i = 1, \dots, k$ ;
  - i) out of the  $N$  live points, take the one with the lowest likelihood which we call  $\theta_l$  with corresponding likelihood  $L_i = L(\theta_l)$ , where  $l \in \{1, \dots, N\}$ ;
  - ii) set  $\xi_i = \exp(-i/N)$ ;
  - iii) set  $w_i = \xi_{i-1} - \xi_i$  (or  $w_i = (\xi_{i-1} - \xi_{i+1})/2$  for the trapezoidal rule);
  - iv) update  $\mathcal{Z} = w_i L_i + \mathcal{Z}$ ; and
  - v) update the set of active points  $\theta_1, \dots, \theta_N$  replacing  $\theta_l$  by drawing a new point  $\theta$  from the prior distribution restricted to  $L(\theta) > L(\theta_l)$ .

Figure 1 : L'algorithme tel qu'il est décrit dans Maturana et al, 2019 (avec  $Z = VM$ ).

Annexe 10 : Le modèle de diffusion spatiale discret pour les souches de spoligotype F7 représentée sur la zone d'étude restreinte à une grille de 1610 km<sup>2</sup> (les deux souches les plus éloignées n'ont pas été prise en compte).



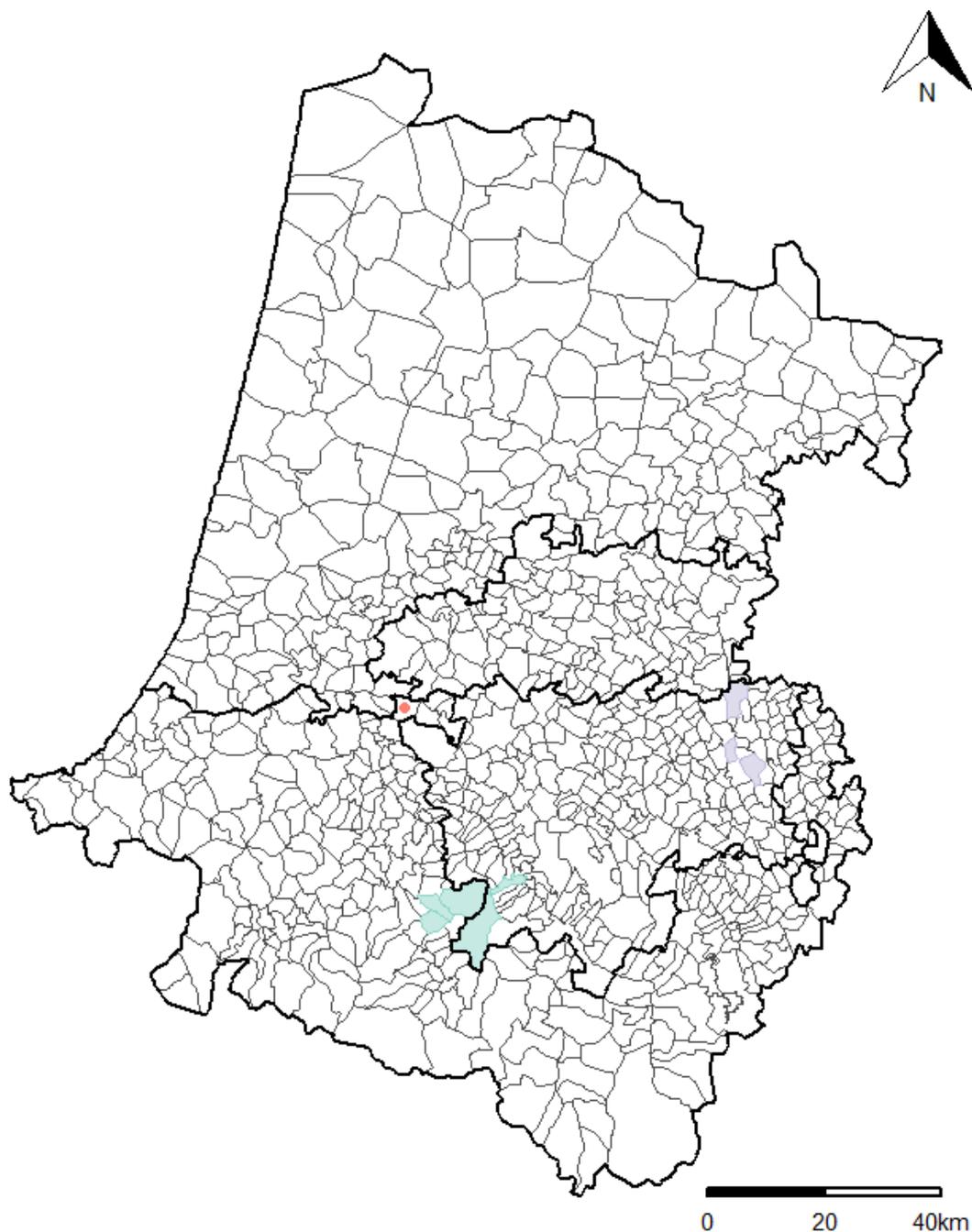
Annexe 11 : Correspondance entre les spoligotypes présents dans la zone d'étude et leur code binaire (d'après Mbovis.org).

Ancienne Nomenclature	Nomenclature Homogénéisée	Code binaire correspondant
<b>F7</b>	SB0821	11010011011111110000001111111111101111100000
<b>F15</b>	SB0832	11010011011111110111111111111111101111100000

Annexe 12 : Profils VNTR identifiés dans la zone d'étude.

Profils identifiés								Nombre de souches
6	3+5	5	3	11	2	2+5 s	8	1
6	4	5	3	11	2	5 s	8	1
6	5	5	3	11	2	4 s	7	1
6	5	5	3	11	2	4 s	8	27
6	5	5	3	11	2	5 s	8+9	1
6	5	5	3	11	2	5 s	6+8	1
6	5	5	3	11	2	5 s	8	99
6	5	5	3	9+11	2	4 s	8	1
6	5	5	3	-	2	5 s	8	1
6	8	5	3	11	2	5 s	8	1

Annexe 13 : Carte permettant la localisation des pâtures des élevages dont les souches étaient éloignées ainsi que la localisation de la souche porcine (point rouge).



Les quatre souches de spoligotype F15 les plus à l'Est ont été isolées en 2017 et trois d'entre elles concernaient le même élevage. Cet élevage possédait des parcelles dans d'autres communes comprises dans la zone mais celles-ci se trouvaient plus au Nord (présenté sur la carte en mauve). Nous n'avions pas de données parcellaires de l'élevage où la quatrième souche avait été isolée.

Les deux souches de spoligotype F7 les plus au Sud ont été isolées en 2017 et chez des bovins appartenant au même élevage. Cet élevage possédait des parcelles dans d'autres communes mais celles-ci se trouvaient plus à l'Ouest, à l'extérieur de la zone d'étude (présenté sur la carte en vert).

Annexe 14 : Sélection du modèle évolutif

Tableau 1 : Modèles testés avec BEAST. La colonne Aspect graphique renseigne sur l'observation (« ok ») ou non d'une convergence i.e. un aspect en « chenille poilue » de la courbe de sortie de chaque paramètre.

Nom du modèle	Modèles					convergence	
	Substitution	Horloge	Population	Transition Inter-espèce	Grille	Aspect graphique	ESS > 200
JC1	JC	Stricte	Constante	Symétrique	non	ok	oui
HKY1	HKY	Stricte	Constante	Symétrique	non	ok	oui
GTR1	GTR	Stricte	Constante	Symétrique	non	ok	oui
HKY2	HKY	Exponentielle	Constante	Symétrique	non	ok	oui
HKY3	HKY	Log Normale	Constante	Symétrique	non	Ok	oui
HKY4	HKY	Stricte	Exponentielle	Symétrique	non	Ok	oui
HKY5	HKY	Stricte	« Skyline »	Symétrique	non	Ok	oui
HKY6	HKY	Stricte	Exponentielle	Asymétrique	non	Ok	oui
HKY7	HKY	Stricte	Exponentielle	Asymétrique	oui	Ok	oui

Tableau 2 : Sélection des modèles avec le facteur de Bayes. Modèle 1 et Modèle 2 correspondent aux codes des modèles dans le Tableau 4. N correspond au nombre de particules, ML1 (2) au log du maximum de vraisemblance du Modèle 1 (2), et SD, à l'écart-type. Log(BF) est la différence entre ML1 et ML2. Le signe du log(BF) donne le modèle qui est favorisé (si >0, le Modèle 1 est favorisé et inversement). Les paramètres ne sont donnés que pour les modèles dont on peut interpréter les résultats.

Modèle 1	Modèle 2	N	ML1	ML2	SD1	SD2	Log(BF)	Modèle favorisé (choisi)	Niveau de preuve
JC1	HKY1	10	-1995,24	-1933,18	6,84	6,96	-62,06	HKY1	Décisif
HKY1	GTR1	10	-	-	-	-	-	(HKY1)	-
HKY1	HKY2	1	-1894,05	-39504,85	21,57	6,29	37610,8	HKY1	Décisif
HKY1	HKY3	1	-1894,05	-2203,36	21,57	22,97	309,31	HKY1	Décisif
HKY1	HKY4	10	-1933,18	-2004,58	6,96	7,4	71,4	HKY1	Décisif
HKY1	HKY5	10	-1933,18	-1958,8	6,96	7,03	25,62	HKY1	Décisif
HKY4	HKY6	10	-	-	-	-	-	Aucun	-
HKY6	HKY7	1	-1917,91	-2082,78	20,99	21,78	164,87	HKY6	Décisif

Annexe 15 : Tableaux récapitulatif et comparatif des clades selon les méthodes de reconstruction utilisées : à savoir l'inférence bayésienne avec un modèle asymétrique (ASYM) et symétrique (SYM) ainsi que le maximum de parcimonie (MP). Les couleurs des cases correspondent aux couleurs des clades dans les arbres reconstruits par inférence bayésienne. Les couleurs des cases pour les clades MP sont arbitraires. Les cases SANS dans le modèle asymétrique correspondent aux souches qui sont entourées en jaune dans l'arbre.

Tableau 1 : Tableau récapitulatif des équivalences entre les clades des différentes méthodes.

Méthode ASYM		Méthode SYM		Méthode MP	
A	NA	B	NA	D	A
B	NA	D	NA	Une partie de B	
C		C		C	
SANS	NA	A	NA	Une partie de B	

Tableau 2 : Tableau comparatif des clades entre les différentes méthodes.

Echantillon	Clade ASYM	Clade SYM	Clade MP
50369100	A	NA	D
50392400	A	NA	D
50428100	A	NA	D
50428200	A	NA	D
60082400	A	NA	D
60127200	A	NA	D
60195600	A	NA	D
60197100	A	NA	D
70094400	A	NA	D
70094500	A	NA	D
70094600	A	NA	D
70100900	A	B	A
70111900	A	NA	D
70122400	A	NA	D
70122700	A	NA	D
70124700	A	NA	D
70131200	A	NA	D
70246800	A	B	A
70273700	A	B	A
80016700	A	NA	D
80087400	A	B	A
80087500	A	NA	D
80244400	SANS	A	B
90036800	SANS	A	B
90043200	SANS	A	B
90062400	SANS	A	B
10Z004552	A	NA	D
10Z007578	B	D	B
10Z015065	B	D	B

10Z015066	B	D	B
11Z003023	B	D	B
11Z003260	A	NA	D
11Z003748	SANS	A	B
11Z004019	A	NA	D
11Z004023	A	NA	D
11Z004190	A	NA	D
11Z004191	A	NA	D
11Z006232	A	NA	D
11Z006414	A	NA	D
12Z004357	A	NA	D
13Z000481	B	D	B
13Z001986	A	NA	D
13Z002332	A	B	A
13Z002809	SANS	A	B
13Z003041	C	C	C
13Z004716	B	D	B
13Z008676	B	D	B
14Z000030	C	C	C
14Z003370	A	B	A
14Z003763	A	NA	D
14Z004112	A	NA	D
14Z004114	A	B	A
14Z004231	B	D	B
14Z004232	B	D	B
14Z004649	B	D	B
14Z005257	A	B	A
14Z006229	A	B	A
14Z006289	B	D	B
14Z006290	B	D	B
14Z006291	B	D	B
14Z006602	C	C	C
14Z006603	A	NA	D
14Z006803	A	B	A
14Z006806	A	B	A
14Z006807	A	B	A
14Z006808	C	C	C
14Z008502	B	D	B
14Z009303	A	NA	D
15Z000246	A	NA	D
15Z000571	A	NA	D
15Z002045	SANS	A	B
15Z002136	B	D	B
15Z002137	C	C	C
15Z002138	C	C	C
15Z002139	A	NA	D

15Z002347	B	D	B
15Z002489	C	C	C
15Z004056	B	D	B
15Z004057	C	C	C
15Z004058	A	NA	D
15Z006343	B	D	B
15Z006344	B	D	B
15Z006498	C	C	C
15Z006817	C	C	C
15Z008799	A	NA	D
15Z009913	A	NA	D
15Z010972	C	C	C
15Z011421	C	C	C
15Z011426	A	NA	D
15Z011428	A	NA	D
16Z000024	A	NA	D
16Z000028	A	NA	D
16Z000935	SANS	A	B
16Z003009	A	NA	D
16Z003010	A	NA	D
16Z003283	SANS	A	B
16Z004257	B	D	B
16Z004258	A	NA	D
16Z004259	A	NA	D
16Z004260	A	NA	D
16Z004261	A	NA	D
16Z004702	SANS	A	B
16Z004717	A	NA	D
16Z004719	A	NA	D
16Z004721	A	NA	D
16Z005211	A	NA	D
16Z005651	A	NA	D
16Z005652	B	D	B
16Z005653	B	D	B
16Z006301	SANS	A	B
16Z006469	B	D	B
16Z006486	B	D	B
16Z006888	A	NA	D
16Z006896	B	D	B
16Z007493	A	NA	D
16Z007496	A	NA	D
16Z007497	A	NA	D
16Z007684	A	NA	D
16Z008629	A	NA	D
16Z011478	B	D	B
17Z000235	A	NA	D

17Z000752	B	D	B
17Z001188	A	NA	D
17Z001189	A	NA	D
17Z002139	A	B	A
17Z002140	B	D	B
17Z002142	C	C	C
17Z002143	C	C	C
17Z002743	C	C	C
17Z003123	A	NA	D
17Z003241	C	C	C
17Z003413	A	B	A
17Z003414	A	B	A
17Z003416	A	NA	D
17Z003447	C	C	C
17Z004070	A	NA	D
17Z004366	A	NA	D
17Z004367	A	NA	D
17Z004368	A	NA	D
17Z005433	A	NA	D
17Z005492	C	C	C
17Z005738	C	C	C
17Z005739	C	C	C
17Z005740	A	NA	D
17Z006003	C	C	C
17Z006005	A	NA	D
17Z006288	SANS	A	B
17Z006453	C	C	C
17Z006499	A	NA	D
17Z006676	A	NA	D
17Z007511	SANS	A	B
80259500	NA	A	B
20478601	NA	NA	B
40440200	NA	NA	C
50367000	NA	NA	B
60096900	NA	NA	C
50367100	NA	NA	B
50433800	NA	NA	C
80100800	NA	NA	B
80297800	NA	NA	B
17Z001187	NA	NA	C
50391100	NA	NA	B
50433900	NA	NA	C
50438200	NA	NA	C
70124800	NA	NA	B
80118200	NA	NA	B
12Z005361	NA	NA	B

<i>15Z000042</i>	NA	NA	C
<i>15Z002348</i>	NA	NA	D
<i>15Z002488</i>	NA	NA	C
<i>16Z004722</i>	NA	NA	D
<i>17Z000236</i>	NA	NA	D
<i>17Z000237</i>	NA	NA	D
<i>17Z003415</i>	NA	NA	C