

REPUBLIQUE DU SENEGAL

Un peuple - Un but - Une foi

MINISTERE DE L'AGRICULTURE ET DE L'EQUIPEMENT RURAL



Institut Sénégalais de Recherches Agricoles (ISRA)

FICHE TECHNIQUE n° 2 : ETUDE DU PAN - GENOME DE *BACILLUS anthracis*.

M. Mbengue¹*, F. T. Lo¹, A. A. Diallo¹, Y. S. Ndiaye¹, M. Diouf¹ and M. Ndiaye²

1 Laboratoire National de l'Elevage et des Recherches Vétérinaires (LNERV – ISRA) - Hann – Dakar – BP 2057, Sénégal.

2 Laboratoire de Biologie Cellulaire; de Microbiologie et de Rickettsiologie de la Faculté des Sciences et Techniques de l'UCAD – Dakar – Sénégal

Email : mbenguem@yahoo.fr

ISSN : 0850 – 9980

**ISRA
VISA**
Commission de validité des documents
scientifiques et techniques
ISSN n°
Date 25 JAN 2017

Le Président

Vol 9, n°4

[Signature]
Directeur Scientifique
de l'ISRA
Dr Et Hadji TRAORÉ

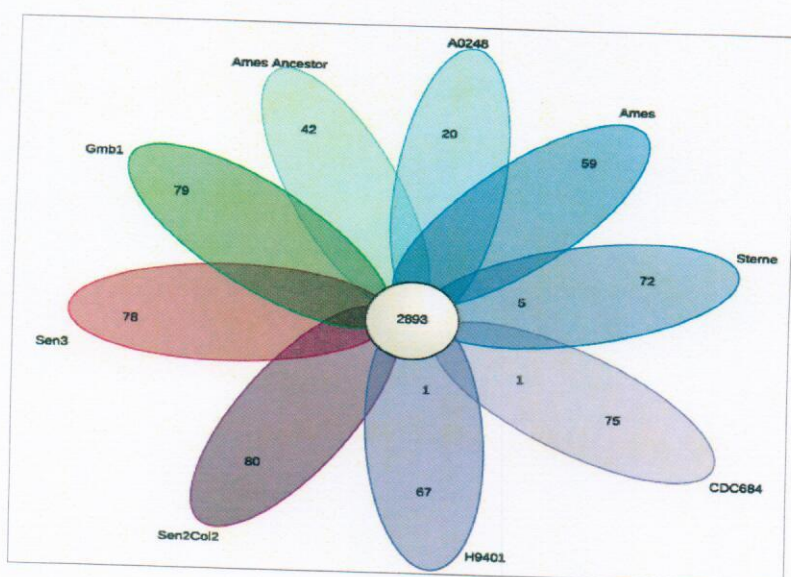


Figure 1: pot de fleurs montrant le pan - génome de *Bacillus anthracis*. Le nombre dans le cercle est le génome de base. Les chiffres dans la partie supérieure des pétales constituent le nombre de gènes accessoires présents dans chaque souche (sur les 85 au total). Les chiffres à la partie inférieure des pétales est le nombre de gènes uniques dans chaque souche.

ETUDE D'UN PANGENOME:

L'étude pangénomique première du genre a été publiée en 2005 (Tettelin H *et al.*, 2005) et, depuis, ce genre d'études a rapidement pris de l'ampleur. Ce travail sur le pan- génome a permis une comparaison entre les différentes espèces ou souches.

Objectif général

L'objectif général de cette activité est de contribuer à une meilleure connaissance de l'espèce *Bacillus anthracis* dans des localités différentes, dans le but de prendre de bonnes dispositions pour la prévention de cette zoonose au Sénégal et dans la sous régions Ouest Africaine

Objectifs spécifiques:

Mettre au point un vaccin recombinant protégeant contre la maladie charbonneuse des ruminants une fois l'agent identifié dans le souci de contribuer à l'amélioration de la santé animale voire humaine. Augmentation de la productivité en matière de protéines animales et du produit intérieur brut..

Matières premières: Souches de *Bacillus anthracis* originaires de la Gambie, du Sénégal et d'autres régions du monde

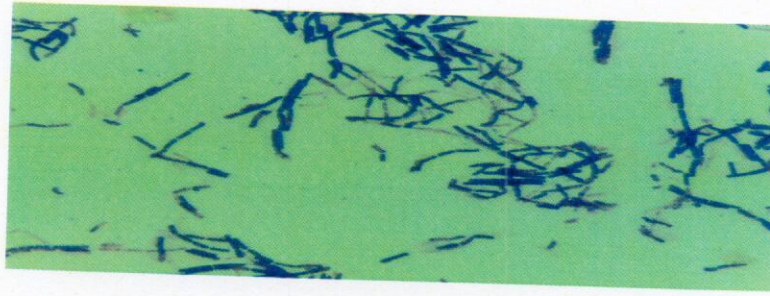


Figure 2 : Frottis de sang coloré par la méthode de GRAM, examiné au Microscope à l'immersion

Caractéristiques : L'examen du frottis coloré par la méthode de Gram a montré des bacilles en longues chaînettes à Gram + (figure 2) : Nous avons aussi noté la présence de spores non déformantes en position subterminale et même terminale sur la culture de 24 heures. La capsule a été bien visualisée après coloration à l'encre de chine à partir du frottis par apposition réalisé sur la rate de la souris morte ayant subi l'inoculation de la bactérie.

MATERIEL D'ETUDES :

Les différentes souches à tester sont les suivantes :

- Sénégal : 3 souches

- Souche n°1 : *Bacillus anthracis* isolée sur vache de race Holstein de la ferme de Wayembam (Niayes du Sénégal) le 09 Février 2010 à partir d'organes (Ganglion, rate, poumon, foie) (**SEN1 = n° 124 21 899**) :
- Souche n°2 : *Bacillus anthracis* isolée à partir de poumons sur Autruche male âgée, du parc zoologique de Hann morte en Juillet 2010 (**SEN 2 = n° 124 21 900**) :
- Souche n°3 : *Bacillus anthracis*, isolée sur mouton male Touabire de la ferme laitière de Wayembam (Niayes du Sénégal) en Avril 2010 à partir de prélèvements d'organes (foie, poumon, rate et sang total). (**SEN 3 = n° 124 21 901**) :

- Gambie : 1 souche

- Souche n°1 : *Bacillus anthracis*, isolée sur bovin Ndama en provenance de la localité d'Abuko (laboratoire Central Vétérinaire) en Gambie le 28 Septembre 2011 à partir de sang entier (couleur noire) récolté sur EDTA. (**Gamb 1 = n° 124 21 902**) :
- **Autres souches référenciées : 6 souches**
- D'autres souches ont été utilisées dans l'étude génomique à titre comparatif, ce sont les suivantes avec leurs caractéristiques particulières Ames (Read, TD., 2003), Ames

Ancêtre (Ravel, J *et al.*, 2009), A0248, CDC 684 (Richard, T *et al.*, 2008)), H9401 (Read, TD *et al.*, 2002) et Sterne. voir <http://www.biomedcentral.com/1471-2164/12/477>

- **Animaux** : Pour la mise en œuvre de cette étude, différents animaux ont constitués les cibles de *Bacillus anthracis* à savoir :
 - Un bovin de race Holstein dont la production moyenne observée est de 20 litres par jour.. Elle dispose d'une robe pie noire
 - Une Autruche âgée du parc zoologique de Hann Dakar (6 années d'âge) trouvée morte à la suite de maladie. Cette autruche est originaire de la ferme de Bandia (Localité de Mbour – Banlieue du Sénégal - 50 Km de Dakar)
 - Un mouton de race Touabire, disposant d'une fertilité et d'une fécondité très améliorée par rapport aux autres races telles que la race Djallonké. a présenté un pelage blanc
 - Un bovin de race Ndama trypanotolérante. Ce bovin male de race Ndama est à 4 années d'âge porte des cornes courtes.
- **Procédure d'identification** :
 - Analyse bactériologique et moléculaire suivie d'une interprétation des séquences génomiques obtenues par la méthode bioinformatique : les résultats obtenus après ces analyses nous permettent d'envisager la connaissance de l'identité de chacune de ces souches et d'apporter une réponse sur leur éventuelle variabilité génomique, une recherche qui n'a jamais été menée dans ces zones Sénégalaises et Gambiennes et qui a déjà fait l'objet de recherches dans certaines régions du monde (Keim,P *et al.*, 2000).
 - Réalisation d'une importante étude sur le pan - génome de *Bacillus anthracis*. basée sur le génome de trois souches (Sen2, Sen3 et Gmb1) et sur ceux de six autres souches de références.
 - **Etudes bactériologiques** : La procédure utilisée pour la mise en œuvre de cette étude bactériologique, fait référence à celle menée par Leon Leminor *et al.* (Leminor *et al.*, 1998) : Après ensemencement le milieu est incubé à 37°C pendant 24 heures.
 - Prélèvements d'organes réalisés dans les différents sites selon des procédures bio sécuritaires, mis en culture dans du bouillon ordinaire.

- Isolement s'en suit sur gélose au sang de mouton (hémoculture) qui est incubé comme précédemment.
- Coloration de Gram est effectuée à partir des isolats, de même que l'ensemencement de milieux pour l'étude des caractères biochimiques.
- Etude sur la production de l'Acetyl Méthyl Carbinol (AMC) de même que la fermentation de certains sucres a été réalisée.(2)
- Réaction de Voges Proskauer est un test microbiologique pour mettre en évidence la production d'acétoïne.
- Epreuve sur l'étude du pouvoir pathogène expérimental (PPE) chez la souris : Elle consiste à procéder à l'inoculation de 500 µl de la souche isolée en culture par la voie intra péritonéale et de suivre l'animal pendant 24 Heures : Après la mort une autopsie est faite de même qu'un frottis par apposition sur la rate hypertrophiée : le frottis est par la suite coloré par la méthode de Giemsa et observé au Microscope avec le grossissement 100 à l'immersion.
- Autopsie et la mise en évidence de *Bacillus anthracis* par confection de frottis par apposition au niveau de la rate suivie d'une technique de coloration par la méthode Giemsa
- Mise en évidence de la capsule : Elle est réalisée en déposant sur une lame une goutte de suspension bactérienne à laquelle une goutte de colorant d'encre de chine a été ajoutée. La lamelle est recouverte de façon à obtenir un gradient de teinte noire. L'existence de la capsule est caractérisée par la présence d'un halo clair autour de la bactérie : la capsule est souvent rencontrée dans les produits pathologiques.

Etudes moléculaires :

- Protocole ayant pour but de décrire les modalités d'extraction des agents pathogènes hautement contagieux en assurant une bonne application des Bonnes Pratiques d'Exécution des Analyses (GBEA) respectant ainsi les mesures techniques de prévention, notamment de confinement, à mettre en œuvre dans les industries et les laboratoires de recherche et d'enseignement où les travailleurs sont susceptibles d'être exposés à des agents biologiques pathogènes suivant le protocole PO5/BIO/CFX-96 : Protocole d'utilisation du CFX 96 Biorad. (Voir Annexe3)
- **Extraction de l'ADN :**
- Prélèvement d'une colonie est mis en suspension dans un tube d'Eppendorf dans 500µl d'eau (conserver 200 µl à - 20°C pour vérifications ultérieures, et utiliser 100 µl pour faire une coloration de Gram).

- 200 µl de cette solution est mélangée à 500 µl de Chelex (tubes prêts au réfrigérateur). Ce tube muni d'un capuchon anti-ouverture est chauffé à 100°C pendant trente minutes sur le bloc chauffant dans le laboratoire L1 puis centrifugé à 10 000 tours /minutes pendant 20 minutes (centrifugeuse à froid ; localisée dans la zone Stock).
- Après centrifugation, le surnageant est récolté (en prenant garde de ne pas aspirer de bille) et placé dans un tube d'Eppendorf neuf libellé du numéro Humox ou Biotox correspondant.
- A partir d'une poudre ou d'un prélèvement clinique : une suspension est faite avec 1ml d'eau distillée de la manière suivante : 200 µl d'échantillon prélevé auquel il faut ajouter 500 µl de Chelex (tubes prêts au réfrigérateur). Le reste de l'extraction a suivi le protocole ci – dessus.

- **Amplification**

Pour chaque colonie isolée par plaque, nous avons ensuite procédé à une centrifugation des tubes contenant 200 µl de Tampon d'élution (TE) constitué par du tampon Tris Hcl à pH 8,0 et de l'EDTA à 1,0mM.

Les colonies ont été ensuite remises en suspension par vortex ou par pipetage répété afin de bien homogénéiser.

La suspension cellulaire a été chauffée à 95°C pendant 20 minutes et refroidie à la température ambiante.

Les débris cellulaires ont été obtenus après centrifugation à 15000 g pendant une minute.

La centrifugation a été menée dans un endroit sécurisé absent d'aérosols.

Le surnageant a été transféré à un nouveau tube pour stockage.

1 µl de lysat renfermant suffisamment d'ADN nécessaire pour une PCR simple pour 200 réactions.

Nous avons optimisé une réaction par titrage des concentrations en lysats utilisant des séries de dilutions.

Cependant le procédé de la lyse thermique est plus rapide et plus facile et est la méthode la plus utilisée pour la majorité de nos échantillons.

Le mélange réactionnel est constitué de la façon suivante :

4 µl H₂O + 0,5 µl UDG + 10 µl de Mix + 0,5 µl d'amorce **S** + 0,5 µl d'amorce **R** + 0,5 µl de Soude
+ 4 µl d'extrait ADN = 20 µl de volume de mélange réactionnel par puits au total.

La réaction au total comprend les éléments suivants :

Réaction :

1 X PCR tampon : 0,2 mM de dNTPS : 1 µl de Platinum à 0,04 U : Taq DNA Polymerase 0,04 U : 0,2 µg d'extrait d'ADN.

PCR

Amplification proprement dite :

Thermocycleur est programmé pour les réactions à mener : d'abord à 94°C pendant 5 minutes pour activer l'ADN polymérase, ensuite à 94°C pendant 20 secondes : 60°C pendant 20 secondes et 65°C pour 20 secondes : ces trois étapes sont ainsi répétées 34 fois. L'étape finale étant de 65°C pendant 5 minutes.

Après amplification, on procède à l'analyse automatique des génotypes par la technique de l'électrophorèse sur gel d'agarose par le système Perkin Elmer Applied Biosystèmes.

Les gels obtenus après électrophorèse ont par la suite été analysés et la taille apparente des fragments d'ADN a été déterminée par le microséquençage par le système SOLID (Annexe 2) qui nous permis d'obtenir les éléments de comparaison par rapport à la souche de référence par GENBANK

Microséquençage :

Cette méthode de microséquençage peut être réalisée par la méthode de Sanger (Annexe 3) mais avec l'évolution des biotechnologies, d'autres méthodes plus sensibles sont mises en œuvre c'est le cas du système SOLID 4 Life technologies.

La bibliothèque a été construite jumelée à partir de 1 µg de l'ADN génomique extrait de chaque souche.

Pour la course (run), un montant total de 708 millions P2 perles positifs ont été chargées sur la cellule. La sortie de la longueur de lecture a été comme prévue 85bp (50x35bp).

3 des génomes *B anthracis*. (Sen2col-Sen3-Gmb1) ont été séquencés par les codes à barres E 3,2 E +6,3.1 +6,3.9 E +6 et lus alors que le génome d'une souche à savoir la Sen1 ayant généré moins de lectures (2.3E +6) n'a pu être lue : ce qui a conduit respectivement aux valeurs suivantes à 273Mb, 262Mb, 382Mb 188Mb et le projet de séquençage global de ces 4 génomes a procédé à l'utilisation de données à 1.1GB.

Le projet de séquençage global de ces 3 génomes a été réalisé à 917Mb de données.

Analyse génomique :

Au cours de notre étude, nous avons eu à effectuer la caractérisation générale des souches utilisées, à mener une étude à partir de différents logiciels composés d'outils et de méthodes informatiques.

Et surtout une étude pangénomique a été réalisée sur l'ensemble des souches de *Bacillus anthracis* :

Pour la caractérisation des souches, la souche de référence utilisée fut Ames. Le logiciel CLC Genomic a été utilisé.

Nous avons opté pour la solution du mapping car l'assemblage de novo de reads SOLiD ne fonctionne pas (nombres de contigs énormes ...).

Trois outils informatiques ont été utilisés :

Pour les outils informatiques nous avons eu à utiliser certains dont trois principalement :

- **Logiciel COG** (Cluster of Orthologues Groups) nécessaire pour l'analyse fonctionnelle des génomes (<http://www.ncbi.nlm.nih.gov>)

- **Logiciel KEGG** (Kyoto Encyclopedia of Genes and Genomes) utile pour mieux comprendre l'analyse métabolique des génomes. Ce logiciel est surtout utilisé après l'identification des gènes impliqués dans des phénotypes d'intérêt. Il devrait se poser le problème de l'étude de ces gènes.

Pour faciliter la tâche nous étions dans l'obligation de créer un fichier HTML qui puisse nous donner toutes les informations sur ces gènes à travers Kyoto Encyclopedia of Genes and Genomes.

- **Logiciel PCA** (Analyse de la composante principale ou Principle Component Analysis) utilisé pour l'analyse statistique des génomes

Deux méthodes informatiques ont aussi été utilisées :

- **Méthode d'alignement - MAUVE** (Alignement multiple du génome en présence d'événements évolutifs ou Multiple Genome Alignment ou Molecular Evolutionary Genetics Analysis) qui est une méthode d'alignement.

L'alignement multiple a été fait grâce à ClustalW. A partir de celui-ci, cet arbre fut généré via la méthode des distances, aussi connue sous le nom de Neighbor Joining.

Afin de générer tout ceci, nous avons utilisé le logiciel MEGA5.

- **MEGA5** (Analyse moléculaire génétique évolutive ou Molecular Evolutionary Genetics Analysis) qui est une méthode intégrée pour la réalisation d'alignement de séquences et la possibilité d'inférer des arbres phylogénétiques.

La construction de l'arbre phylogénétique est utilisée par les programmes d'alignement multiples de séquence afin d'éliminer une grande partie des alignements et de limiter ainsi le temps de calcul. Il guide l'alignement total : la méthode la plus employée est celle du Neighbour joining (NJ) ou méthode des distances.

MEGA est utilisé par le biologiste dans un grand nombre de laboratoires pour la reconstruction de l'histoire évolutive des espèces et de déduire l'ampleur de la nature des forces sélectives et de façonner l'évolution des gènes et des espèces.

L'analyse moléculaire génétique évolutive (MEGA) est un logiciel qui a été développé dans le but de former une approche centrée biologiste, avec une suite intégrée d'outils d'analyse statistique de l'ADN à partir du point de vue évolutif.

Le Clustering hiérarchique (groupage hiérarchique ou Hierarchic Clustering) qui peut être considéré en temps qu'arbre phylogénétique d'autant plus que dans les options, nous avons demandé un Clustering optimal. L'orthologie est un lien évolutif entre deux gènes présents chez deux espèces différentes si elles descendent d'une séquence unique présente dans le dernier ancêtre commun aux deux espèces : deux gènes sont dits orthologues s'ils se retrouvent dans des espèces différentes (spéciation sans duplication)

Pour que deux gènes soient considérés comme orthologues, il faut donc établir qu'ils sont homologues c'est-à-dire qu'ils descendent d'un même gène ancestral. Ensuite il faut montrer que c'est après l'événement d'une spéciation qu'il y'a eu une évolution séparée des deux gènes.

D'une manière générale ; la procédure suivie pour cette analyse génomique est la suivante :

L'alignement du pan-génome et du mobilome

Tout d'abord, nous avons effectué deux types d'alignements: un alignement global du génome de mauve (Darling *et al.*, 2010) et avec un alignement multiple MEGA5 (Tamura *et al.*, 2011) (en utilisant l'algorithme ClustalW) à la suite d'une reconstruction de l'arbre phylogénétique (méthode des distances, encore appelée méthode de jonction (Neighbour Joining ou procédé NJ) . Pour la mise en place de l'arbre phylogénétique, nous nous sommes basés sur le MLST (typage génomique multilocus), qui est composé de 7 gènes de ménage (*glpF*, *gmk*, *ilvD*, *pta*, *pur*, *pycA* and *tpi*) (Kim *et al.*, 2005).

Merci à MAUVE et son fichier de sortie épine dorsale (Sheppard *et al.*, 2013), à partir duquel nous avons calculé la proportion du génome de base en fonction de la taille du pan-génome, une épreuve qui nous a permis d'évaluer la nature fermée ou ouverte du pan-génome.

Etude d'un pan - génome proprement dit:

L'étude pangénomique première du genre a été publiée en 2005 (Tettelin H *et al.*, 2005) et, depuis, ce genre d'études a rapidement pris de l'ampleur. Ce travail sur le pangénome a permis une comparaison entre les différentes espèces ou souches.

CAMERA (Sun S, *et al.*, 2011) est un portail bioinformatique où nous pouvons faire plusieurs types d'analyse. Nous l'avons utilisé pour générer les données COG (Tatusov RL, Natale DA, Garkavtsev 2001). Il est un outil commun, utilisé pour attribuer les annotations fonctionnelles des protéines. Les protéines ont été classées en catégories (la liste est disponible à l'<http://www.ncbi.nlm.nih.gov/COG/old/palox.cgi?fun=all>). Pour obtenir les données KEGG (Ogata H, Goto S, Sato K, Fujibuchi 1999), il faut donc étudier les voies métaboliques : Ainsi pour y parvenir nous avons utilisé les KAAS (KEGG serveur d'annotation automatique) qui est un outil en ligne.

En KEGG, les protéines ont été réparties entre les classes et sous-classes : Alignements (orthologie/PCA). Tout d'abord, nous avons réalisé deux types d'alignements: un alignement du génome global avec MAUVE (Darling AE *et al.*, 2010) et un alignement multiple avec MEGA5 (Tamura K *et al.*, 2011) en utilisant l'algorithme ClustalW à la suite de la réalisation d'un arbre pour tous les génomes (méthode à distance, ou méthode de jonction).

Ensuite, nous avons utilisé le logiciel OrthoMCL (Chen F *et al.*, 2006) pour obtenir une liste des orthologues et déterminer la composition du pan - génome (noyau, accessoire et les gènes uniques).

Pour les résultats obtenus avec COG et KEGG, nous avons effectué une analyse en composante principale (ACP) en utilisant R, pour avoir une meilleure vue de la répartition des différentes classes.

Nous avons utilisé MeV (Saeed AI *et al.*, 2003) ; Saeed AI *et al.*, 2006) (Viewer expérience multi) pour mieux visualiser la distribution d'accessoires et de gènes afin d'effectuer une classification hiérarchique.

RESULTATS : Tout d'abord, nous avons étudié les gènes uniques. Nous avons trouvé 5 pour la souche Sterne (2 pas trouvés, une protéine conservée hypothétique, la déshydrogénase, EmrB / QACA transporteur résistance famille de médicaments et de liaison au zinc), 1 dans la souche CDC 684 (pas de fonds sur le NCBI) et 1 pour la souche H9401 (yfeT liaison à l'ADN de la transcription régulateur). Ensuite, nous avons examiné en détail les 85 gènes accessoires à partir de la figure 24 suivante montrant bien le clustering hiérarchique : (Fig 3)

Le mobilome examiné a montré la même tendance dans toutes les souches: 2 transposases, 1 ou 2 CRISPRs confirmés et beaucoup de phages ou des protéines / protéines associées aux prophages (123 en moyenne). Nous avons également examiné cette question au niveau des plasmides. Nous avons constaté que tous les plasmides pX01 appartiennent aux transposases 5, 1 et 0 des phages CRISPRs. Tous les plasmides pX02 possèdent 2 transposases, 8 phages et 0 CRISPRs. Le pan-génome est composé de 2893 gènes de base, 7 gènes uniques et 85 gènes accessoires (Figure 1).

Discussion et Conclusion

Dans ce travail, nous avons donné une validation importante telle que montrée précédemment (Tettlin *et al.*, 2005), à savoir le pan-génome de *B.anthraxis* qui est de type étroit. Pour confirmer notre étude, nous avons utilisé différents outils afin de comparer et valider nos résultats.

Tout va dans le même sens après l'addition des trois souches Sénégalaises et Gambiennes qui n'influe pratiquement pas sur la nature fermée de ce pan-génome (2893 gènes essentiels et 85 gènes accessoires), avec un ratio core / pan-génome de 99%. Ce ratio core / pan-génome est très proche des autres souches pathogènes pour l'homme (tableau V) comme *Rickettsia rickettsii*.

Tableau I : Pan - génome de souches pathogènes humaines avec la colonne pourcentage (%) qui correspond au ratio core/pan - génome

Espèces	Génome utilisé	Style de vie	Intracellulaire	Niche	Taille pan génome	Taille core génome	%
<i>B .anthracis</i>	9	allopatric	Non	Animal	47041887	46513801	99
<i>Rickettsia rickettsii</i>	8	allopatric	Non	Tiques	10129221	100112432	99
<i>Chlamydia trachomatis</i>	20	allopatric	Non	Humain	20960000	20689197	99
<i>Rickettsia prowazeki</i>	8	allopatric	Non	Humain	8888959	8869530	100

Cependant, il y'a une discordance entre la présence d'un mobilome (beaucoup de phages et, au moins, un CRISPRs confirmés), et la forte présence de protéines impliquées dans la transcription et la traduction, caractères propres pour un pan - génome fermé.

Néanmoins, *Bacillus anthracis* dérivé du groupe *Bacillus cereus*, une espèce sympatrique qui n'est pas intracellulaire. Donc, *Bacillus anthracis* peut devenir allopatrique. Ceci peut être expliqué par le fait que *B.anthraxis* est une bactérie ancienne (au moins 150 ans) qui a cessé d'évoluer. Nous avons testé cette hypothèse en étudiant le SNP basé sur le contenu génomique de base. Nous avons constaté que les SNPs sont au nombre total de 2786, dont 896 sont logés dans le génome de base. En outre, le rapport de transition / transversion est très faible (0,32).

Ce manque de SNPs peuvent valider notre hypothèse sur l'arrêt de l'évolution de cette espèce. *B.anthraxis* qui est une souche ancienne qui s'est stabilisée avec le temps et qui présente un pan - génome de type étroit.

Références bibliographiques :

1. Kim, K., E. Cheon, K. E. Wheeler, Y. Youn, T. J. Leighton, C. and S. I. Chung. 2005. Determination of the most closely related bacillus isolates to *Bacillus anthracis* by multilocus sequence typing. *Yale J.Biol.Med.* **78**:1-14.
2. MBENGUE M, KONTE M 2001 – Mise au point d'un vaccin recombinant , bivalent contre les deux maladies charbonneuses bovines : Thèse Doc 3^{ème} cycle N° 57
- 2.Sheppard, S. K., X. Didelot, K. A. Jolley, A. E. Darling, B. Pascoe and D. Falush. 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Mol.Ecol.* **22**:1051-1064. doi:10.1111/mec.12162 [doi].
- 3.Sun S, Chen J, Li W, Altintas I, Lin A, Peltier S and Wooley J (2011) Community cyberinfrastructure for advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res* **39**: D546-D551. gkq1102 [pii];10.1093/nar/gkq1102 [do]
4. Tettelin H, Massignani V, Cieslewicz MJ, Donati C and Fraser CM (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* **102**: 13950-13955. 0506758102 [pii];10.1073/pnas.0506758102 [doi].