Des études de diversité des bactéries d'intérêt agronomique, vétérinaire et médical aux perspectives de gestion

Xavier Bailly

▶ **To cite this version:**

Xavier Bailly. Des études de diversité des bactéries d'intérêt agronomique, vétérinaire et médical aux perspectives de gestion. Sciences du Vivant [q-bio]. Université Clermont Auvergne (UCA), Clermont-Ferrand, FRA., 2020. tel-04127661

HAL Id: tel-04127661
https://hal.inrae.fr/tel-04127661

Submitted on 14 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Université Clermont Auvergne
Ecole Doctorale des Science de la Vie, Santé, Agronomie & Environnement

Dossier de candidature
Habilitation à Diriger des Recherches

Discipline : Biologie des populations et écologie
Spécialité : Génomique évolutive et des populations

Des études de diversité des bactéries d'intérêt agronomique, vétérinaire et médical aux perspectives de gestion

Xavier Bailly

Soutenue le 30 mars 2020 devant le jury composé de :
Samuel Alizon, Directeur de Recherche CNRS, Montpellier
Gilles Béna, Directeur de Recherche IRD, Montpellier
Christiane Forestier, Professeur UCA, Clermont-Ferrand
Emmanuelle Gilot-Fromont, Professeur VetAgro Sup, Lyon
Philippe Glaser, Responsable de Groupe, Institut Pasteur, Paris
Frédérique Le Roux, Directrice de Recherche IFRMER, Roscoff
Olivier Tenaillon, Directeur de Recherche INSERM, Paris

Remerciements

Je tenais à remercier les rapporteurs et membres du jury qui m'ont fait le plaisir d'accepter d'évaluer mon dossier et qui m'ont encouragé à aller au bout de cette démarche, malgré différents contre-temps.

Parmi les personnes qu'il me faut remercier, je pense particulièrement à ceux qui m'ont mis le pied à l'étrier, encadré, conseillé. J'espère avoir un impact aussi positif auprès de ceux que j'accompagne.

Un grand merci aux collègues de l'unité EPIA. Cette HDR est le témoignage de nos interactions chaleureuses et enrichissantes, de votre soutien et votre confiance.

Merci a mon équipe de recherche de bonheurs quotidiens, Céline, Léonie et Raphaël.

Contents

1 Curriculum vitae **3**

2 Des bancs de l'Université à l'HDR **10**
2.1 Mon cursus 10
2.2 Développement et financement de projets 12
2.3 Contribution au fonctionnement de réseaux de recherche 12
2.4 Formation par la recherche 12
2.5 Valorisation des activités de recherche 14
2.6 Contribution au dialogue entre recherche et porteurs d'enjeux en épidémiologie 15
2.7 Reconnaissance de l'expertise scientifique et travaux d'intérêt collectif 15
2.8 Dynamique collective au sein d'EPIA 16
2.9 Direction de l'unité EPIA 17

3 Illustration des activités de recherche **19**
3.1 Questions posées par les bactéries d'intérêt agronomique, vétérinaire ou médical 19
3.2 Questions d'écologie et d'évolution bactérienne traitées 20

4 Impact de la recombinaison sur les patrons de diversité génomique **21**
4.1 Recombinaison entre taxons apparentés: des mesures de déséquilibre de liaison aux taux de recombinaison intra et inter-taxons 22
4.1.1 Recombinaison chez les rhizobia associés aux luzernes, *S. meliloti* et *S. medicae* 22
4.1.2 Recombinaison au sein des espèces cryptiques incluses dans *Rhizobium leguminosarum* 24
4.1.3 Diminution de la recombinaison homologue entre espèces du complexe *Borrelia burgdroferi* 25
4.2 Distribution des événements de recombinaison au sein des génomes bactériens : des hypothèses liées à l'observation de marqueur au point de vue génomique 27
4.2.1 Recombinaison homologue et transferts au sein des différentes unités de réplication des génomes de *S. meliloti* et *S. medicae* 27
4.2.2 Polymorphisme des plasmides cp26 du complexe d'espèce *Borrelia burgdorferi*: quand la sélection exacerbe les traces de la recombinaison 29

5 Effet de la sélection sur les patrons de diversité & identification des déterminants génétiques de phénotypes **31**
5.1 Spécificité d'hôte chez les rhizobia, de l'étude de régions candidates aux associations entre diversité génomique et phénotype d'intérêt 32
5.2 Distribution de gènes accessoires déterminant d'autres caractères adaptatifs dans les populations de rhizobia 35
5.3 Sélection et dynamique épidémiologique chez *Borrelia burgdorferi* 37
5.4 Variation du potentiel zoonotique chez *Borrelia garinii* 38

6 Épidémiologie moléculaire de pathogènes zoonotiques : des patrons de diversité aux cycles de transmission des pathogènes 42
6.1 Épidémiologie des pathogènes et diversité observable à l'aide de données de surveillance . . . 42
6.1.1 Diversité de souches de *Coxiella burnetii* issues de cas d'avortement chez différentes espèces d'élevage . . . 43
6.1.2 Diversité des populations d'*Anaplasma phagocytophilum* chez les bovins et d'autres espèces de vertébrés . . . 44
6.2 Des données de co-infection aux hypothèses épidemiologiques . . . 45
6.2.1 Mesure de l'impact des chevreuils sur l'infection des tiques par *Anaplasma phagocytophilum* . . . 45
6.2.2 Identification de cycles de transmission de *Borrelia burgdorferi* en Forêt de Sénart . . . 47
6.3 Identification d'associations dans les microbiotes, une perspective en santé animale . . . 50
6.3.1 Étude des associations entre microbes au sein de données métagénomiques 50
6.3.2 Étude du microbiote d'*Ixodes ricinus* en fonction des stades de développement . . . 52

7 Projet de recherche 55
7.1 Développer l'utilisation des approches phylodynamiques pour les pathogènes animaux et zoonotiques . . . 55
7.2 Comprendre les processus adaptatifs des agents pathogènes dans leur environnement . . . 56
7.3 Étudier la dynamique épidémiologique des agents pathogènes au sein et entre leurs hôtes . . . 58
7.4 Happy end! . . . 59

8 Bibliographie 60

1 Curriculum vitae

Formation

-1996 Bac S option Sciences Naturelles

-1999 DEUG Sciences Biologiques et Naturelles Université Montpellier 2

-2000 Licence Biologie des Organismes et des Populations Université Montpellier 2

-2001 Maîtrise Biologie des Populations et des Ecosystèmes Université Montpellier 2

-2002 DEA Ressources Phytogénétiques et Interactions Biologiques Université Montpellier 2

-2006 Doctorat, École Doctorale Biologie Intégrative, spécialité Ressources Phytogénétiques et Interactions Biologiques Université Montpellier 2. Direction de thèse : Isabelle Olivieri (UMR ISEM) et Gilles Béna (UMR LSTM)

Expériences professionnelles

-2006-2008 Post-doctorat à l'Université de York (UK) : Génomique des populations de rhizobia

-2008 Ingénieur de Recherche au Laboratoire d'Épidémiologie des maladies Animales et zoonotiques (UR INRA 0346) : Épidémiologie moléculaire des pathogènes animaux et zoonotiques

Responsabilités collectives

-2003-2006 Représentant doctorant au conseil de l'école doctorale « Biologie des Systèmes Intégrés, Agronomie – Environnement » - Université Montpellier II

-2011-2015 Représentant adjoint élu au conseil scientifique du centre INRA Auvergne Rhône Alpes

-2012-2016 Représentant nommé du département de Santé Animale auprès de la Cellule Bioinformatique de l'INRA

-2016-2020 Représentant élu au conseil scientifique du département de Santé Animale de l'INRA

-2018 Directeur de l'Unité Mixte de Recherche Épidémiologie des maladies animales et zoonotiques

-2019 Représentant adjoint d'INRAE au Conseil National d'Épidémiosurveillance en Santé animale

Financements en tant que porteur de projet

-2003 Bourse « Jean et Marie-Louise Dufrenoy » de l'Académie Française d'Agriculture

-2009 AIP Bioressources INRA

-2009-2011 FEDER, Région Auvergne

-2012 AIP Bioressources INRA

-2015-2016 Projet de Métaprogramme INRA

-2015-2018 Bourse de thèse INRA

Participation à des programmes de recherche nationaux et internationaux

-2011-2015 Projet OSCAR – Outil de Simulation Cartographique à l'Échelle du Paysage Agricole du Risque Acarologique. Financement : Agence Nationale de la Recherche

-2012-2016 Projet BIODIS – Décrypter les liens entre Biodiversité et maladies infectieuses émergentes. Financement : Centre de synthèse et d'analyse sur la biodiversité

-2020-2024 Projet MOOD – Monitoring outbreak events for disease surveillance in a data science context. Financement : Programme cadre Horizon 2020.

Encadrement

Doctorants

-Maude Jacquot - 2011-2014 - Diversité génomique des bactéries pathogènes du complexe d'espèces *Borrelia burgdorferi* : évolution et épidémiologie moléculaire.
Directeur de thèse Christian Ducrot, encadrement à 90%

-Amélie Chastagner - 2011-2014 - Étude des cycles épidémiologiques d'*Anaplasma phagocytophilum* en France : apport des approches de caractérisation génétique.
Directrice de thèse Agnès Leblond, co-encadrement avec Gwenaël Vourc'h à 40%

-Arnaud Cougoul - 2015-2019 - Analyses statistiques de réseaux d'associations entre espèces microbiennes à partir de données métagénomiques.
Directrice de thèse Gwenaël Vourc'h, co-encadrement avec Patrick Gasqui à 40%

Master 2

-Maude Jacquot

-Amélie Chastagner
-Audrey Ganteil
-Oshma Chakoory

Master 1/Maîtrise
-Stéphane de Mita
-Gaëlle Léonard
-Marie-Mathilde Perrinaud
-Sébastien Guizard
-Maxime Bisseux

Magistère/Licence
-Santiago Wadsworth
-Emilie Bévilacqua
-Benjamin Faure
-Mathilde Carrias

Publications

Publications avec un doctorant co-encadré (en italique)

1- ***Chastagner, A.***, **Bailly, X.**, Leblond, A., Pradier, S., Vourc'h, G., 2013. Single Genotype of Anaplasma phagocytophilum Identified from Ticks, Camargue, France. Emerg. Infect. Dis. 19, 825–827.
doi:10.3201/eid1905.121003

2- ***Chastagner, A.***, Dugat, T., Vourc'h, G., Verheyden, H., Legrand, L., Bachy, V., Chabanne, L., Joncour, G., Maillard, R., Boulouis, H.-J., Haddad, N., **Bailly, X.**, Leblond, A., 2014. Multilocus sequence analysis of Anaplasma phagocytophilum reveals three distinct lineages with different host ranges in clinically ill French cattle. Vet. Res. 45.
doi:10.1186/s13567-014-0114-7

3- ***Jacquot, M.***, Bisseux, M., Abrial, D., Marsot, M., Ferquel, E., Chapuis, J.-L., Vourc'h, G., **Bailly, X.**, 2014a. High-Throughput Sequence Typing Reveals Genetic Differentiation and Host Specialization among Populations of the Borrelia burgdorferi Species Complex that Infect Rodents. PLoS One 9.
doi:10.1371/journal.pone.0088581

4- ***Jacquot, M.***, Gonnet, M., Ferquel, E., Abrial, D., Claude, A., Gasqui, P., Choumet, V., Charras-Garrido, M., Garnier, M., Faure, B., Sertour, N., Dorr, N., De Goër, J., Vourc'h, G., **Bailly, X.**, 2014b. Comparative population genomics of the Borrelia burgdorferi species complex reveals high degree of genetic isolation among species and underscores benefits and constraints to studying intra-specific epidemiological processes. PLoS One 9, e94384.
doi:10.1371/journal.pone.0094384

5- Dugat, T., ***Chastagner, A.***, Lagree, A.-C., Petit, E., Durand, B., Thierry, S., Corbiere, F., Verheyden, H., Chabanne, L., **Bailly, X.**, Leblond, A., Vourc'h, G., Boulouis, H.-J., Maillard, R., Haddad, N., 2014. A new multiple-locus variable-number tandem repeat analysis reveals different clusters for Anaplasma phagocytophilum circulating in domestic and wild ruminants. Parasit. Vectors 7.
doi:10.1186/1756-3305-7-439

6- Vourc'h, G., Abrial, D., Bord, S., ***Jacquot, M.***, Masseglia, S., Poux, V., Pisanu, B., **Bailly, X.**, Chapuis, J.-L., 2016. Mapping human risk of infection with Borrelia burgdorferi sensu lato, the agent of Lyme borreliosis, in a periurban forest in France. Ticks Tick. Borne. Dis. 7, 644–652.
doi:10.1016/j.tttbdis.2016.02.008

7- ***Jacquot, M.***, Abrial, D., Gasqui, P., Bord, S., Marsot, M., Masseglia, S., Pion, A., Poux, V., Zilliox, L., Chapuis, J.-L., Vourc'h, G., **Bailly, X.**, 2016. Multiple independent transmission cycles of a tick-borne pathogen within a local host community. Sci. Rep. 6.
doi:10.1038/srep31273

8- ***Chastagner, A.***, Pion, A., Verheyden, H., Lourtet, B., Cargnelutti, B., Picot, D., Poux, V., Bard, E., Plantard, O., McCoy, K.D., Leblond, A., Vourc'h, G., **Bailly, X.**, 2017. Host specificity, pathogen exposure, and superinfections impact the distribution of Anaplasma phagocytophilum genotypes in ticks, roe deer, and livestock in a fragmented agricultural landscape. Infect. Genet. Evol. 55, 31–44.
doi:10.1016/j.meegid.2017.08.010

9- ***Cougoul A.***, **Bailly X.**, Vourc'h G., Gasqui P., 2019. Rarity of microbial species: In search of reliable associations. PLoS One. 2019;14(3):e0200458.
doi: 10.1371/journal.pone.0200458

Autres publications

10- **Bailly, X.**, Migeon, A., Navajas, M., 2004. Analysis of microsatellite variation in the spider mite pest Tetranychus turkestani (Acari: Tetranychidae) reveals population genetic structure and raises questions about related ecological factors. Biol. J. Linn. Soc. 82, 69–78.
doi:10.1111/j.1095-8312.2004.00316.x

11- Rasolomampianina, R., **Bailly, X.**, Fetiarison, R., Rabevohitra, R., Béna, G., Ramaroson, L., Raherimandimby, M., Moulin, L., De Lajudie, P., Dreyfus, B., Avarre, J.-C., 2005. Nitrogen-fixing nodules from rose wood legume trees (Dalbergia spp.) endemic to Madagascar host seven different genera belonging to alpha- and beta-Proteobacteria. Mol. Ecol. 14, 4135–46.
doi:10.1111/j.1365-294X.2005.02730.x

12- Villegas, M.D.C., Rome, S., Mauré, L., Domergue, O., Gardan, L., **Bailly, X.**, Cleyet-Marel, J.-C., Brunel, B., 2006. Nitrogen-fixing sinorhizobia with Medicago laciniata constitute

a novel biovar (bv. medicaginis) of S. meliloti. Syst. Appl. Microbiol. 29, 526–38.
doi:10.1016/j.syapm.2005.12.008

13- Merabet, C., Bekki, A., Benrabah, N., Bey, M.B.H., Bouchentouf, L., Ameziane, H., Rezki, M.A., Domergue, O., Cleyet-Marel, J.C., Avarre, J.C., Bena, G., **Bailly, X.**, de Lajudie, P., 2006. Distribution of Medicago species and their microsymbionts in a saline region of Algeria. Arid L. Res. Manag. 20, 219–231.
doi:10.1080/15324980600705685

14- **Bailly, X.**, Béna, G., Lenief, V., de Lajudie, P., Avarre, J.-C., 2006. Development of a lab-made microarray for analyzing the genetic diversity of nitrogen fixing symbionts Sinorhizobium meliloti and Sinorhizobium medicae. J. Microbiol. Methods 67, 114–24.
doi:10.1016/j.mimet.2006.03.006

15- **Bailly, X.**, Olivieri, I., De Mita, S., Cleyet-Marel, J.-C., Béna, G., 2006. Recombination and selection shape the molecular diversity pattern of nitrogen-fixing Sinorhizobium sp. associated to Medicago. Mol. Ecol. 15, 2719–2734.
doi:10.1111/j.1365-294X.2006.02969.x

16- **Bailly, X.**, Olivieri, I., Brunel, B., Cleyet-Marel, J.-C., Béna, G., 2007. Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of Medicago species. J. Bacteriol. 189, 5223–36.
doi:10.1128/JB.00105-07

17- Chaintreuil, C., Rigault, F., Moulin, L., Jaffre, T., Fardoux, J., Giraud, E., Dreyfus, B., **Bailly, X.**, 2007. Nickel Resistance Determinants in Bradyrhizobium Strains from Nodules of the Endemic New Caledonia Legume Serianthes calycina. Appl. Environ. Microbiol. 73, 8018–8022.
doi:10.1128/AEM.01431-07

18- Pavoine, S., **Bailly, X.**, 2007. New analysis for consistency among markers in the study of genetic diversity: development and application to the description of bacterial diversity. Bmc Evol. Biol. 7.
doi:156 10.1186/1471-2148-7-156

19- **Bailly, X.**, Giuntini, E., Sexton, M.C., Lower, R.P.J., Harrison, P.W., Kumar, N., Young, J.P.W., 2011. Population genomics of Sinorhizobium medicae based on low-coverage sequencing of sympatric isolates. Isme J. 5, 1722–1734.
doi:10.1038/ismej.2011.55

20- Suzan, G., Garcia-Pena, G.E., Castro-Arellano, I., Rico, O., Rubio, A. V, Tolsa, M.J., Roche, B., Hosseini, P.R., Rizzoli, A., Murray, K.A., Zambrana-Torrelio, C., Vittecoq, M., **Bailly, X.**, Aguirre, A.A., Daszak, P., Prieur-Richard, A.-H., Mills, J.N., Guegan, J.-F., 2015. Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. Ecol. Evol. 5, 865–873.

doi:10.1002/ece3.1404

21- Joulie, A., Laroucau, K., **Bailly, X.**, Prigent, M., Gasqui, P., Lepetitcolin, E., Blanchard, B., Rousset, E., Sidi-Boumedine, K., Jourdain, E., 2015. Circulation of Coxiella burnetii in a Naturally Infected Flock of Dairy Sheep: Shedding Dynamics, Environmental Contamination, and Genotype Diversity. Appl. Environ. Microbiol. 81, 7253–7260.
doi:10.1128/AEM.02180-15

22- Ezenwa, V.O., Prieur-Richard, A.-H., Roche, B., **Bailly, X.**, Becquart, P., Garcia-Pena, G.E., Hosseini, P.R., Keesing, F., Rizzoli, A., Suzan, G., Vignuzzi, M., Vittecoq, M., Mills, J.N., Guegan, J.-F., 2015. Interdisciplinarity and Infectious Diseases: An Ebola Case Study. Plos Pathog. 11.
doi:10.1371/journal.ppat.1004992

23- Kumar, N., Lad, G., Giuntini, E., Kaye, M.E., Udomwong, P., Shamsani, N.J., Young, J.P.W., **Bailly, X.**, 2015. Bacterial genospecies that are not ecologically coherent: population genomics of Rhizobium leguminosarum. Open Biol. 5.
doi:10.1098/rsob.140133

24- Joulie, A., Sidi-Boumedine, K., **Bailly, X.**, Gasqui, P., Barry, S., Jaffrelo, L., Poncet, C., Abrial, D., Yang, E., Leblond, A., Rousset, E., Jourdain, E., Consortium, A.D.L., 2017. Molecular epidemiology of Coxiella burnetii in French livestock reveals the existence of three main genotype clusters and suggests species-specific associations as well as regional stability. Infect. Genet. Evol. 48, 142–149.
doi:10.1016/j.meegid.2016.12.015

25- Ras, G., **Bailly, X.**, Chacornac, J.-P., Zuliani, V., Derkx, P., Seibert, T.M., Talon, R., Leroy, S., 2017. Contribution of nitric oxide synthase from coagulase-negative staphylococci to the development of red myoglobin derivatives. Int. J. Food Microbiol.
doi:10.1016/j.ijfoodmicro.2017.11.005

26- Hosseini, P.R., Mills, J.N., Prieur-Richard, A.-H., Ezenwa, V.O., **Bailly, X.**, Rizzoli, A., Suzan, G., Vittecoq, M., Garcia-Pena, G.E., Daszak, P., Guegan, J.-F., Roche, B., 2017. Does the impact of biodiversity differ between emerging and endemic pathogens? The need to separate the concepts of hazard and risk. Philos. Trans. R. Soc. B-Biological Sci. 372.
doi:10.1098/rstb.2016.0129

27- **Bailly, X.**, 2017. Hidden Markov phylogenetic models offer an interesting perspective to identify "high risk lineages" of environmental pathogens. Infect. Genet. Evol. 55, 45–47.
doi:10.1016/j.meegid.2017.08.007

28- Ageorges, V., Schiavone, M., Jubelin, G., Caccia, N., Ruiz, P., Chafsey, I., **Bailly, X.**, Dague, E., Leroy, S., Paxman, J., Heras, B., Chaucheyras-Durand, F., Rossiter, A.E., Henderson, I.R., Desvaux, M. 2019 Differential homotypic and heterotypic interactions of antigen 43 (Ag43) variants in autotransporter-mediated bacterial autoaggregation. Sci Rep. 2019;9.

doi:10.1038/s41598-019-47608-4

29- Binetruy, F., **Bailly, X.**, Chevillon, C., Martin, O.Y., Bernasconi, V. M., Duron, O. Phylogenetics of the Spiroplasma ixodetis endosymbiont reveals past transfers between ticks and other arthropods. Ticks Tick Borne Dis. 2019;10(3):575–84.
doi: 10.1016/j.ttbdis.2019.02.001

30- Giangaspero, M., Steinbach, F., Strong, R., Decaro, N., Buonavoglia, C., Domenis, L., Gargano, P., **Bailly, X.**, Apicella, C., Turno, P. 2020. Characterization of internal ribosome entry sites according to secondary structure analysis to classify border disease virus strains. J Virol Methods. 2020 Jan;275.
doi: 10.1016/j.jviromet.2019.113704

2 Des bancs de l'Université à l'HDR

2.1 Mon cursus

Une démarche scientifique se construit au fil des expériences et des échanges. Mon parcours a débuté à l'Université de Montpellier dans le cadre d'une formation en biologie des organismes, des populations et des écosystèmes. Intrigué par le concept de biologie intégrative mis en avant dans les formations de l'Université, j'ai eu l'opportunité d'explorer en parallèle de cette formation le potentiel que pouvaient offrir les approches génétiques et moléculaires, plus centrales dans d'autres cursus de formation, aux travaux d'écologie et d'évolution. J'ai ainsi effectué des stages dans les unités de recherche "Biochimie et Physiologie Moléculaire des Plantes" dans l'équipe de Pierre Berthomieu et "Différenciation Musculaire et Métabolisme" sous la direction de Chantal Wrutniak-Cabello en 1999 et 2000.

Ce cursus m'a naturellement conduit à m'orienter vers le domaine de l'écologie évolutive et moléculaire. Je souhaitais confronter des observations de terrain et des hypothèses biologiques. J'ai commencé à appliquer cette démarche dans le cadre d'une étude visant à tester la spécialisation d'un acarien phytophage vis-à-vis de différentes plantes hôtes en étudiant la structure génétique de plusieurs populations sous la direction de Maria Navajas. J'ai réalisé cette étude au sein du "Centre de Biologie pour la Gestion des Populations" en 2001, une troisième unité empreinte des problématiques d'INRAE (l'INRA... à l'époque). Cette expérience a fini d'ancrer mon intérêt pour le développement de travaux sur des modèles biologiques appliqués, mais le manque de données permettant de faire le lien entre structure génétique et adaptation chez les eucaryotes m'a paru limitant.

Je me suis alors orienté vers un Diplôme d'Études Approfondies formant dans les domaines des "Ressources Phytogénétiques et Interactions Biologiques". Cette formation offrait des bases sur différentes approches de modélisation. Même si je ne suis pas modélisateur, ces bases m'ont donné le goût de collaborer avec des partenaires plus exercés dans le domaine pour enrichir mon travail empirique. De plus cette formation m'a conduit à m'intéresser à l'écologie évolutive et moléculaire chez les bactéries. Leur biologie évolutive me semblait offrir des perspectives intéressantes, avec un lien plus direct avec les approches de génétique fonctionnelle.

De 2002 à 2008, durant mon DEA, ma thèse et mon post-doctorat, je me suis intéressé à l'évolution de bactéries appelées rhizobia, qui fixent l'azote atmosphérique en interaction avec les légumineuses. Les laboratoires que j'ai fréquentés, le "Laboratoire des Symbioses Tropicales et Méditerranéennes" et l'"Institut des Sciences de l'Évolution de Montpellier" encadré par Gilles Béna et Isabelle Olivieri, de 2002 à 2006, puis le groupe de Peter Young à l'Université de York, de 2006 à 2008, ont constitué des environnements favorables à plusieurs points de vue. Ils m'ont permis de développer des recherches en écologie et évolution moléculaire prenant en compte des connaissances fonctionnelles et génomiques. Mon expérience anglaise m'a également permis d'intégrer assez rapidement les outils de séquençages haut-débit à mes travaux.

A mon arrivée en post-doctorat, les approches de séquençage haut-débit ont modifié la forme et le fond des recherches en microbiologie, impactant les questions de recherche abordables, les protocoles expérimentaux à envisager, et l'analyse des données. Il m'a rapidement semblé important de développer des recherches tirant parti de cette évolution. Ne disposant

pas d'une formation en bio-informatique pour prendre en main les données de séquençage haut-débit, je me suis efforcé de développer mes compétences dans ce domaine au fil des expériences. Ainsi, une large partie de mon travail s'est tournée vers l'exploration des opportunités offertes par les approches de séquençage haut-débit pour comprendre l'évolution et l'écologie des bactéries en interaction avec des hôtes eucaryotes.

Cette orientation technique et la volonté de travailler sur des sujet susceptibles de déboucher sur des applications concrètes placent mon travail à l'interface entre ingénierie et recherche. Ce profil m'a permis de rejoindre en 2008 l'unité d'"Épidémiologie des Maladies Animales et Zoonotiques" (EPIA) en tant qu'ingénieur de recherche. L'objectif était d'intégrer des approches d'épidémiologie moléculaire, prenant en compte des notions d'évolution et d'écologie microbienne, dans les projets de l'unité. Ces projets était alors centrés sur l'analyse de données de cas dans le cadre d'approches d'épidémiologie quantitative. Cette mission m'a donné l'opportunité de m'approprier de nouvelles problématiques appliquées, de confronter mes connaissances à de nouveaux modèles biologiques, de développer de manière autonome des problématiques de recherche en développant des interactions avec mes collègues vétérinaires, écologues, notamment Gwenaël Vourc'h qui avait suivi mon recrutement, informaticiens et statisticiens, dont David Abrial et Patrick Gasqui que j'inonde de questions depuis mon arrivée. Dans ce contexte, je me suis intéressé à l'évolution, l'écologie et l'épidémiologie de bactéries susceptibles d'affecter la santé animale et humaine, en utilisant comme modèle d'étude principal des bactéries pathogènes transmises par les tiques.

J'ai pu consacrer une dizaine d'années à la conception de projets d'ingénierie et de recherche, à la réalisation de projets financés, et au développement de méthodes permettant d'aborder les questions que je me posais (Figure 1). Au delà des interactions avec les collègues de mon unité et nos partenaires, j'ai pu participer à la formation par la recherche et accompagner de jeunes chercheurs. Au fil des ans, j'ai contribué à des tâches d'intérêt collectif, notamment au sein de mon unité et d'INRAE.

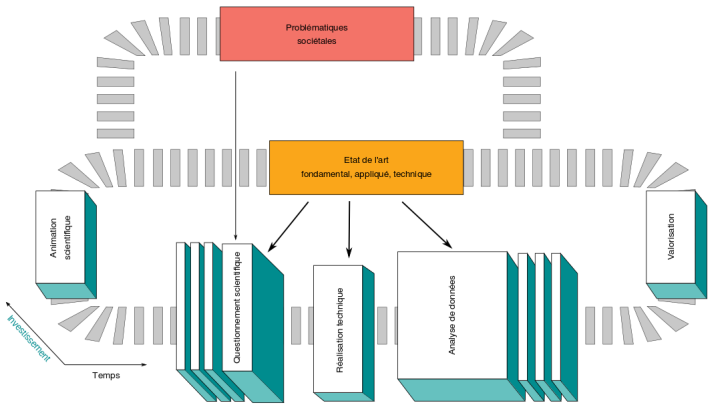

Figure 1: Schéma décrivant ma démarche de travail au sein de l'unité d'épidémiologie des maladies animales et zoonotiques entre 2008 et 2018, figurées en bleu, dans le cadre d'un processus de recherche.

2.2 Développement et financement de projets

Depuis ma thèse, j'ai eu une démarche proactive dans la conception de projets et la recherche de financements pour les réaliser. J'ai été porteur d'initiatives financées par l'INRA, la région Auvergne, ou le Fond FEDER. J'ai participé activement, en tant que collaborateur, à la préparation de projets financés par l'ANR (Outil de Simulation Cartographique à l'échelle du paysage Agricole du Risque acarologique - OSCAR) ou le programme cadre Horizon 2020 (MOnitoring Outbreak events for Disease surveillance in a data science context - MOOD).

La dynamique créée au sein de réseaux de recherche est centrale pour développer de nouveaux projets et obtenir leur financement.

2.3 Contribution au fonctionnement de réseaux de recherche

J'ai participé aux activités de différents réseaux en lien avec mes thèmes de recherche et mon environnement de travail. Je me suis investi dans l'organisation d'événements : les journées des microbiologistes de l'INRA 2012 (comité scientifique, animateur de session), les journées du département de Santé Animale de l'INRA 2013 (comité d'organisation et comité scientifique), les Journées Ouvertes en Biologie, Informatique, et Mathématiques 2015 (comité scientifique) et le congrès Pathobiome 2018 (comité scientifique, animateur de session).

J'ai été sollicité pour illustrer les problématiques de reconstruction de réseaux d'association entre microbes à partir de données métagénomiques dans le cadre d'une école chercheur en écologie microbienne INRA en 2016 et, en mars 2019, les problématiques de phylodynamique dans le cadre d'une introduction à l'écologie et l'évolution des maladies infectieuses et parasitaires organisée par INRAE.

Je me suis impliqué dans des groupes de travail internationaux, notamment via le projet Biodis financé par la Fondation pour la Recherche sur la Biodiversité via le CEntre de Synthèse et d'Analyse sur la Biodiversité dont l'objectif était de mieux comprendre les liens entre biodiversité et santé.

Compte tenu du profil sur lequel j'ai été recruté, j'ai centré mes contributions au fonctionnement de réseaux de recherche sur des initiatives de l'INRA et de son département de Santé Animale. Aujourd'hui, je pense qu'il serait judicieux de consacrer plus d'énergie à des groupes de travail disciplinaires plus larges, pour enrichir ma démarche et participer à de nouveaux projets collaboratifs.

2.4 Formation par la recherche

Les projets de recherche engagés m'ont apporté de nombreuses expériences, notamment dans le domaine de la formation par la recherche.

J'ai eu l'opportunité d'accompagner un post-doctorant, Mathieu Gonnet. La construction de nouvelles opportunités passera notamment par le développement de projets permettant d'intégrer le financement de post-doctorants et par une augmentation de la visibilité et de l'attractivité des recherches conduites.

Je me suis investi dans l'encadrement de trois doctorants :

- Maude Jacquot, avec un encadrement à 90% partagé avec son directeur de thèse Christian Ducrot, soutenance en octobre 2014
- Amélie Chastagner, avec un encadrement à 40% partagé avec ses directrices de thèse Agnès Leblond et Gwenaël Vourc'h, soutenance en octobre 2014
- Arnaud Cougoul, avec un encadrement à 40% partagé avec Patrick Gasqui et sa directrice de thèse Gwenaël Vourc'h, soutenance en novembre 2019

Amélie et Maude ont soutenu leur thèse dans les temps impartis. Je partage avec Amélie et Maude huit articles, qui ont parfois été publiés bien après leur soutenance. Les derniers articles publiés avec Amélie et Maude n'ont été valorisés qu'en 2017. Amélie et Maude ont néanmoins connus plusieurs expériences post-doctorales en lien avec leur thèse dans des contextes compétitifs. La thèse d'Arnaud à pris sept mois de plus que prévu. Arnaud a pour l'instant valorisé un article et produit un pré-print. Trois autres projets de publication sont associés à sa thèse. Arnaud a effectué un premier CDD suite à son contrat de thèse et devrait commencer un post-doctorat en mars 2020.

En lien avec la formation par la recherche, j'ai pu compléter mon expérience en participant aux comités de thèse de Cécile Rangin (UMR LSTM Montpellier), de Rémy Melkonian (UMR LSTM Montpellier), de Caroline Bournaud (UMR LSTM Montpellier), d'Enrique Ortega-Abboud (UMR BGPI Montpellier), de Chervin Hassel (UMR LMGE Clermont-Ferrand), de Florian Binetruy (UMR MIVEGEC Montpellier), de Romain Daveu (UMR BIOEPAR Nantes), et d'Hélène Gardon (UMR LMGE).

J'ai également participé au jurys de thèse de Jean-Philippe Buffet (UMR BIPAR, Maisons-Alfort, 2012), Elsa Quillery (UMR BioEpAR, Nantes, 2013), Alice Leduc (UMR PVBMT, Saint Pierre, 2015) et Jonas Durand (Université de Neuchâtel, Neuchâtel, 2016).

L'obtention de l'habilitation à diriger les recherches me permettra de poursuivre des activités de formation par la recherche au niveau doctoral. Elles sont sources d'échanges qui enrichissent mes réflexions et me permettent de prendre du recul sur mes activités. Il me faudra être attentif à la durée de thèse et aux délais de valorisation des travaux des doctorants que j'encadre. Ce sont des points importants pour qu'ils puissent faire fructifier leur investissement à court terme. Plus généralement, j'espère que les doctorants que je contribue à former, capables de prendre en compte l'écologie et l'évolution des micro-organismes dans des domaines appliqués et maîtrisant des outils de bio-informatique, pourront obtenir un emploi en phase avec leur cursus. Il devraient pour cela bénéficier du développement des questions de recherche et des problématiques sociétales liées à la gestion des problématiques de santé simultanément à l'échelle des populations et de leurs individus.

Depuis mon arrivée à Clermont-Ferrand, je me suis aussi impliqué dans la formation par la recherche d'étudiants aux niveaux licence et master. Les travaux initiés à York m'ont donné l'opportunité de co-encadrer le stage de Master 1 Bioinformatique de Sébastien Guizard et celui de Master 2 Bioinformatique de Maude Jacquot (Université Clermont Auvergne). Les analyses de diversité génétique du complexe d'expèce *Borrelia burgdorferi* m'ont amené à encadrer les stages de DUT Génie Biologique de Mathilde Carrias (IUT de Clermont-Ferrand), de Licence Bioinformatique de Benjamin Faure (IUT d'Aurillac), Master 1 de Maxime Bisseux

(ENS Lyon), de Master 2 bioinformatique d'Audrey Ganteil (Université de Rennes I). J'ai participé au co-encadrement du Master 2 Ecologie, Evolution, Biométrie d'Amélie Chastagner (Université Claude-Bernard Lyon I) sur *Anaplasma phagocytophilum*. J'encadre cette année le stage de Master 2 Bioinformatique d'Oshma Chakoory (UCA) sur la sensibilité du modèle CorHMM.

Au cours de mon parcours, j'ai aussi pris part à quelques modules de formation initiale. Dans le cadre de ma thèse, j'avais effectué des enseignements dans différents contextes : des interventions en première et seconde année d'école d'ingénieur à l'Agro-Montpellier (16 et 48h), les sessions de TP de Licence de Physiologie Végétale à l'Université de Nîmes (100h), une intervention en DEA Ressources phytogénétiques et Interactions Biologiques à l'Université de Montpellier (3h). Depuis mon arrivée à Clermont-Ferrand, j'ai également construit un cours sur les approches de phylogénie à l'IUT d'Aurillac dans le cadre d'une licence de Bioinformatique (24h), et j'interviens ponctuellement depuis deux ans dans le module optionnel de Master 2 de Microbiologie "Émergence et diffusion des pathogènes" à l'UCA (6h). J'ai participé en 2018 au jury de stage de ce Master 2.

Je devrais autant que faire ce peut développer ma participation à l'encadrement de stage et aux enseignements au niveau Master. La perte de l'accès au concours pour les bourses ministérielles de notre école doctorale de rattachement ne m'a probablement pas incité à être plus actif. Il n'en reste pas moins vrai que commencer à travailler durant le Master 2 sur un potentiel sujet de thèse met le doctorant dans une situation plus confortable lorsqu'on arrive à décrocher un financement alternatif. De plus, même si une continuité entre Master et thèse ne peut pas être construite, il est critique d'être identifié par les étudiants potentiellement intéressés par les sujets de thèse proposés. Happé par d'autres activités, j'ai souvent saisi l'opportunité d'encadrer des stagiaires de Master suite à des sollicitations d'étudiants, plus que par une démarche systématique de proposition de sujets de stage. De plus, ma contribution à des enseignements au niveau Master est restée marginale au regard des faibles volumes horaires effectués. Ainsi, je souhaiterais être plus proactif pour proposer des interventions et des sujets de stage au niveau Master, notamment dans les cursus dans lesquels je peux être amené à encadrer, par exemple dans le cadre formations de microbiologie, de bio-informatique, ou d'écologie et d'évolution. La création d'un master international en santé publique vétérinaire soutenu par l'Université de Lyon constitue également une opportunité de m'insérer plus dans ces activités.

2.5 Valorisation des activités de recherche

Les projets de recherche auxquels j'ai participé, dans un cadre souvent collaboratif et avec le support d'étudiants, ont conduit à des résultats reconnus par la communauté scientifique. Je suis associé à un nombre de publications en phase avec les critères quantitatifs de qualification aux fonctions de Professeur des Universités de la section biologie des populations et écologie du conseil national des Universités, avec environ 2.5 publications par an depuis ma soutenance en 2006. Ces travaux ont eu un écho favorable dans les communautés intéressées. Ils sont cités régulièrement, 2.5 citations par an par publication.

Au delà de l'aspect quantitatif, j'ai essayé de construire ou de contribuer dès que possible à

des projets étudiant l'écologie et l'évolution de mes modèles biologiques sous différents angles. J'ai essayé de progresser, si possible d'innover, dans les approches techniques et les analyses utilisées pour aborder mes sujets. J'espère par cette démarche contribuer aux questions que se pose la communauté scientifique. J'ai essayé de valoriser ces travaux de manière intégrée.

En accord avec les problématiques de mon unité, j'essaye de faire évoluer mes sujets pour que mes travaux soient plus pertinents d'un point de vue épidémiologique et répondent plus directement aux problèmes des porteurs d'enjeux du domaine et plus généralement sociétaux. Je veux autant que faire ce peut inscrire mes activités dans ce cadre, en profitant des synergies qui peuvent exister entre le besoin de recueillir des informations pour la gestion de problèmes sanitaires et les questions de recherche posées par l'analyse de ces informations.

2.6 Contribution au dialogue entre recherche et porteurs d'enjeux en épidémiologie

Dans cette optique, j'ai contribué à un groupe d'expert de l'ANSES sur la gestion d'une pestivirose causée par le *Border-Disease Virus* à l'interface entre la faune sauvage et les animaux d'élevage, ici l'isard et les ovins dans les Pyrénées.

Je me suis investi dans le processus de renforcement du groupe d'appui transversal de la plateforme nationale d'épidémiosurveillance en santé animale ESA suite à l'inclusion de l'INRA dans le dispositif en 2018. La plateforme ESA implique des acteurs majeurs, publiques et privés, du domaine. Dans ce cadre, un groupe de sept personnes est en cours de constitution au sein de notre unité pour contribuer à l'amélioration de la surveillance sanitaire des maladies animales et zoonotiques en France. Cette plateforme apporte un appui méthodologique et opérationnel pour la conception, le déploiement, l'animation, la valorisation et l'évaluation des dispositifs de surveillance sanitaire. La plateforme ESA apporte également un appui pour l'investigation épidémiologique de problématiques sanitaires. Enfin, elle a pour mission de développer une interface entre surveillance opérationnelle et recherche.

J'ai contribué aux démarches de conventionnement nécessaires à la mise en place du groupe INRA, aujourd'hui INRAE, au sein de la plateforme, au recrutement et au suivi des cinq personnes ayant pour l'instant contribué au dispositif, et à la promotion de cette démarche. Je suis représentant adjoint d'INRAE au Comité National d'Épidémiosurveillance en Santé Animale qui pilote ce dispositif.

J'espère que notre participation à la plateforme ESA constituera une des voies facilitant le développement et le transfert des recherches développées pour améliorer la surveillance et la gestion des maladies animales et zoonotiques. De part son caractère collaboratif, c'est une source potentielle de nouveaux partenariats.

2.7 Reconnaissance de l'expertise scientifique et travaux d'intérêt collectif

Mes travaux m'ont permis d'être invité dans différents évènements nationaux ou internationaux : une édition des Rencontres plantes-bactéries, un Forgarty's Research and Policy for

Infectious Disease Dynamics Meeting à Glasgow, un Symposium sur *Borrelia* à Munich.

Je suis sollicité pour des activités de référé pour les journaux Molecular Ecology (3), the ISME Journal (2), FEMS microbial-ecology (1), Infection Genetics and Evolution (1), Journal of Evolutionary Biology (1), Journal of Theoretical Biology (1), PLOS One (1).

Le département de Santé Animale et le métaprogramme Adapation au Changement Climatique de l'Agriculture et de la Forêt de l'INRA, le DIM1Health d'Ile de France et le National Environmental Research Council de Grande-Bretagne ont fait appel à mes services pour l'évaluation de projets dans le cadre d'appels d'offre. Enfin, j'ai participé au jury HCERES de l'UMR VIRO en 2018.

Au sein de l'INRA, j'ai participé au titre du département de Santé Animale à une cellule de travail concernant la situation et l'évolution de la bioinformatique au sein de l'Institut. J'ai effectué un mandat en tant que membre élu adjoint au comité scientifique du centre INRA Auvergne Rhône-Alpes. J'ai participé à la cellule d'animation sur le thème du "Pathobiome" dans le cadre du métaprogramme Meta-omiques des systèmes Microbiens. J'ai participé à l'élaboration de la prospective biologie prédictive de l'Institut produite par l'institut en 2019. Je suis depuis 2016 membre élu du conseil scientifique du département de Santé Animale.

J'ai été sollicité pour participer aux jurys de différents concours. J'ai participé à quatre concours externes de recrutement INRA, deux au niveau ingénieur d'étude et deux au niveau ingénieur de recherche, pour des profils de bioinformaticiens et d'épidémiologistes. J'ai participé à un jury d'évaluation conseil pour des unités CIRAD en 2015. J'ai participé à un jury de recrutement de cadre de recherche pour le CIRAD. J'ai présidé en 2019 le concours interne de recrutement de l'INRA au niveau ingénieur d'étude.

2.8 Dynamique collective au sein d'EPIA

Au sein de l'unité EPIA, j'ai le sentiment que les objectifs initiaux associés à mon profil de recrutement ont été remplis. Les activités d'épidémiologie moléculaire, au sens large, sont intégrées dans le travail de notre collectif. Je suis de ce fait impliqué dans une part significative des projets de l'unité avec mes collègues. J'ai participé à un tiers des publications mises en avant pour illustrer nos productions dans le cadre de l'évaluation HCERES de l'unité en décembre 2019.

Au delà de ces valorisations, j'ai collaboré avec les chercheurs de l'unité, Gwenaël Vourc'h et Elsa Jourdain notamment. Une technicienne en biologie moléculaire, Séverine Barry, et deux ingénieures en bioinformatique, Émilie Bard et Aminah Keliet, ont rejoint notre unité. Trois ingénieurs en statistique et en informatique, David Abrial, Jocelyn de Goër, et Patrick Gasqui, se sont fortement impliqués sur les problématiques sur des projets d'intérêt commun. Nous avons recruté une chargée de recherche, Anaïs Bompard, qui s'intéresse notamment à l'optimisation d'interactions microbiennes dans une perspective de gestion des problématiques sanitaires en élevage en 2018 et un chargé de recherche en phylodynamique des maladies animales et zoonotiques, Julien Thézé, en 2019.

Je pense que la dynamique enclenchée peut perdurer tant il existe encore un besoin de développement de recherches sur l'épidémiologie des pathogènes animaux et zoonotiques à l'interface entre : i) écologie & évolution ; ii) infectiologie, sciences vétérinaires & zootechnie ; iii) informatique & mathématique.

2.9 Direction de l'unité EPIA

Mon travail scientifique et mon investissement pour le collectif m'ont permis d'appréhender le fonctionnement d'une unité de recherche en interne et dans son environnement, de mieux connaître le fonctionnement et les ressources d'INRAE et de ses partenaires. Notre directrice d'unité ayant pris une disponibilité d'un an, j'ai accepté de prendre la direction de l'unité d'Épidémiologie des maladies animales et zoonotiques en avril 2018.

Notre unité comprend actuellement une petite quarantaine d'agents distribués à parts égales sur deux sites, sur le site de Theix du centre INRAE Clermont Auvergne-Rhône-Alpes et le campus vétérinaire de VetAgro Sup de Marcy L'Étoile, proche de Lyon. Les dernières années ont été caractérisées par des changements, encore en cours, dans la structure, la composition et l'administration de notre unité. L'inclusion d'enseignants-chercheurs VetAgro Sup et, en conséquence, le passage de notre unité du statut d'unité propre INRA à unité mixte de recherche en 2017 en sont des marqueurs importants. L'accueil sur le site de Marcy l'Étoile des personnels INRAE dédiés à la plateforme nationale d'épidémio-surveillance en santé animale ESA et de deux chargés de recherche sur le site de Theix depuis 2018 constituent également des changements importants.

Notre collectif a été évalué par le Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur dans le cadre la vague A 2019-2020. Nous avons commencé à préparer notre bilan et notre nouveau projet scientifique en 2018. L'implication de l'INRA dans l'accueil de nouveaux collaborateurs de la plateforme ESA dans notre unité est devenue une position institutionnelle claire au début de l'année 2018. Ainsi, lors de d'une assemblée générale d'unité au printemps 2018, j'ai organisé une rencontre avec la coordination de la plateforme et pour que nous puissions échanger sur les conséquences de l'accueil d'agents dédiés aux points de vue scientifique et organisationnels. Lors de l'assemblée de fin d'année 2018, j'ai organisé un travail collectif pour que nous partagions les enjeux sous-tendant notre travail et définissions les thèmes prioritaires pour notre prochain projet, car nous ne sommes pas structurés en équipes. Les résultats ont été validés par les membres de l'unité et par nos institutions de tutelle. L'assemblée générale du printemps 2019 a été consacrée à la définition des objectifs que nous poursuivrons pour chacun des thèmes.

Pour le prochain contrat, notre cœur d'activité reste d'améliorer la compréhension de la dynamique des maladies animales et zoonotiques. D'un point de vue appliqué, nous souhaitons renseigner les porteurs d'enjeux sur les risques infectieux, et les conseiller sur les possibilités d'optimisation de leurs stratégies de surveillance et de gestion. Nous avons identifié trois thèmes principaux pour le prochain contrat qui agrégeront les forces de l'unité : (i) la «Gestion des enjeux de santé dans un contexte de changement global »; (ii) l'étude de l'«Épidémiologie, l'Écologie et l'Évolution des agents pathogènes animaux et zoonotiques»; iii) l'«Epidémiosurveillance des maladies animales et zoonotiques». Compte tenu des défis scientifiques et sociétaux actuels, ces thèmes développeront leurs activités sur les agents pathogènes animaux résistants aux antibiotiques, les agents pathogènes vectorisés et les agents pathogènes animaux ou zoonotiques critiques. La pertinence de ces thèmes a été relevée par le jury d'évaluation HCERES. La nécessité de définir un plan d'action plus opérationnel visant à atteindre nos objectifs scientifiques à toutefois été souligné.

Dans le cadre d'un tel plan, il sera important d'investir dans la dynamique scientifique de nos sites d'implantation, à Clermont-Ferrand et Lyon. A L'échelle du site Clermontois, nos

activités s'inscriront dans le challenge 1 de I-Site CAP 20-25 (Agro-écosystèmes durables dans un contexte de changements globaux) et la Fédération de Recherche sur les Systèmes Microbiens. Sur le site Lyonnais nous souhaitons notamment développer nos interactions avec les unités participant au réseau Santé et bien être Animal en Auvergne Rhône Alpes (initiative VetAgro Sup, INRAE, ANSES), la communauté du Labex Ecofect auquel nous sommes rattachés, et un hub en Santé Publique Vétérinaire impliquant des partenaires publics et privés financés par l'Idex Université de Lyon. Dans le cadre de ce dernier projet, VetAgro Sup développe avec ses unités de recherche, principalement l'UMR LBBE, et un partenaire privé, Boehringer Ingelheim, une chaire de Santé Publique Vétérinaire. Les thématiques développées au sein de cette chaire couvrirait la recherche et la formation à la modélisation des processus épidémiques à des fins d'inférence, la prédiction des conséquences éco-évolutives potentielles des stratégies de gestion, et le développement d'outils décisionnels associés, ce qui est en parfaite adéquation avec notre projet. Compte tenu de l'éclatement des forces de recherche en santé animale dans la région, des initiatives permettant une meilleure structuration de nos activités et de plus amples collaborations devront être développées. En outre, nous avons besoin d'établir des collaborations plus intenses avec les partenaires nationaux et internationaux intéressés par nos thèmes de recherche. La question du développement international est une question récurrente pour notre unité. Le développement de collaborations à ces échelles, permettant d'envisager des projets plus ambitieux, devra également constituer une priorité. Notre engagement dans des projets européens comme MOOD doit nous servir de tremplin pour participer à de nouveaux projets.

J'ai accepté de poursuivre mes activités de Directeur d'Unité dans le cadre de notre prochain contrat, sur la période 2021-26. J'espère pouvoir contribuer à l'accomplissement de nos objectifs. J'estime que la défense de cette habilitation à diriger les recherches est une étape nécessaire dans cette perspective.

3 Illustration des activités de recherche

Dans les paragraphes suivants, je vais m'attacher à illustrer ma démarche scientifique. Elle s'articule autour de trois questions de génomique des populations appliquées aux bactéries d'intérêt agronomique, vétérinaire ou médical : i) Quel est l'impact de la recombinaison sur la diversité des populations étudiées ? ii) Comment la sélection influence la distribution du polymorphisme dans les populations étudiées ? iii) Quelle est la relation entre la dynamique des populations ou la dynamique épidémiologie et la diversité observée ? Enfin, la dernière partie de ce document sera consacrée aux perspectives de recherche que je souhaite développer, dans la continuité de activités passées et dans le contexte de mon unité.

3.1 Questions posées par les bactéries d'intérêt agronomique, vétérinaire ou médical

Les populations de bactéries que j'étudie établissent des interactions qui se situent le long du continuum entre relations mutualistes et antagonistes. Par leurs actions positives ou négatives, elles jouent un rôle important dans des systèmes biologiques anthropisés. Compte tenu de ce rôle, mon objectif est d'identifier des leviers permettant de mieux maîtriser les interactions qu'elles établissent avec leurs hôtes. Pour cela, je documente comment se crée, se distribue et se maintient la diversité génétique et fonctionnelle au sein de populations de microbes afin de comprendre les étapes clef de leurs cycles de vie.

A une extrémité de ce continuum, les rhizobia sont responsables de la moitié de la fixation de diazote assimilable par les eucaryotes. Ils expriment généralement cette fonction dans le cadre d'une interaction symbiotique avec les plantes légumineuses. Ils représentent ainsi des ressources génétiques clefs dans une perspective agro-écologique. Savoir quels sont les rhizobia capables d'interagir avec une plante d'intérêt agronomique, quels sont les plus efficaces et quel peut être leur devenir en population naturelle sont des questions essentielles pour gérer au mieux ces ressources génétiques. Cette gestion passe par la caractérisation de l'écologie et de l'évolution de ces bactéries symbiotiques. Dans ce contexte, je me suis intéressé à différents modèles : i) *Sinorhizobium meliloti* et *Sinorhizobium medicae*, les symbiotes des luzernes du genre *Medicago*, dont la luzerne modèle *Medicago truncatula*, ii) *Rhizobium leguminosarum* le symbiote des vesces et des trèfles, des genres *Vicia* et *Trifolium*, mais aussi de plantes cultivées comme la lentille *Lens culinaris*, ou iii) des souches de *Bradyrhizobium* sp. associées à des arbustes Néo-Calédoniens, utilisés pour réhabiliter des remblais riches en nickel.

Les bactéries pathogènes constituent l'autre extrémité du continuum. Les populations des modèles que j'étudie persistent généralement via l'infection d'animaux, mais peuvent également infecter l'homme. Les maladies causées par ces pathogènes sont appelées zoonoses. Une grande partie de mon travail porte sur des pathogènes bactériens transmis par des tiques. Comprendre quels sont les acteurs clefs des cycles de transmission de ces pathogènes, définir leur potentiel de diffusion et leur potentiel zoonotique sont des questions qui peuvent s'appuyer sur des travaux d'écologie moléculaire et d'évolution. J'étudie ces différentes questions à travers différents agents pathogènes transmis par la tique *Ixodes ricinus* : *Borrelia bugdorferi*, l'agent de la maladie de Lyme, et *Anaplasma phagocytophilum*, l'agent de l'anaplasmose granulocytique. Je participe également à des projets visant à étudier des pathogènes à transmission directe comme *Coxiella burnetii*, l'agent de la fièvre Q.

3.2 Questions d'écologie et d'évolution bactérienne traitées

Mon cursus, via les questions posées par différentes bactéries modèles et le travail dans des laboratoires ayant des approches diverses -*i.e.* plus mécanistes puis plus analytiques-, m'a amené à travailler au cours du temps sur trois problématiques majeures ancrées dans le domaine de la génomique des populations bactériennes :

- L'étude des mécanismes créant de la diversité génomique chez les bactéries étudiées. Cette problématique s'inscrit dans la continuité de ma formation initiale en biologie évolutive. Il me semble en effet essentiel de comprendre comment mutations et recombinaisons peuvent créer la diversité qui va être façonnée par différents processus écologiques et évolutifs.
- L'identification de loci impliqués dans l'interaction avec l'hôte d'intérêt et l'identification des pressions de sélection qui influencent ces régions génomiques. Au delà des connaissances sur les contraintes écologiques et évolutives auxquelles sont confrontées les bactéries étudiées, cette activité peut ouvrir la voie à différentes applications.
- Documenter l'impact des processus de dynamique des populations ou de transmission sur les patrons de diversité observés. Cette problématique est centrale dans le cadre d'une unité d'épidémiologie. N'ayant que peu d'expérience dans ce domaine lors de mon recrutement, j'ai fait évoluer mes travaux vers cette problématique en prenant en compte : i) les fronts de sciences en écologie de la santé et en épidémiologie moléculaire ; ii) les implications potentielles de travaux de recherche pour les parties prenantes confrontées aux problématiques sanitaires.

Si cette présentation en trois problématiques est pratique, elle reste théorique tant les processus présentés sont généralement imbriqués Figure 2. Ils influencent, chacun à leur manière, les patrons de diversité génomique observables et leurs effets ne sont pas simplement additifs (Rice, 2013).

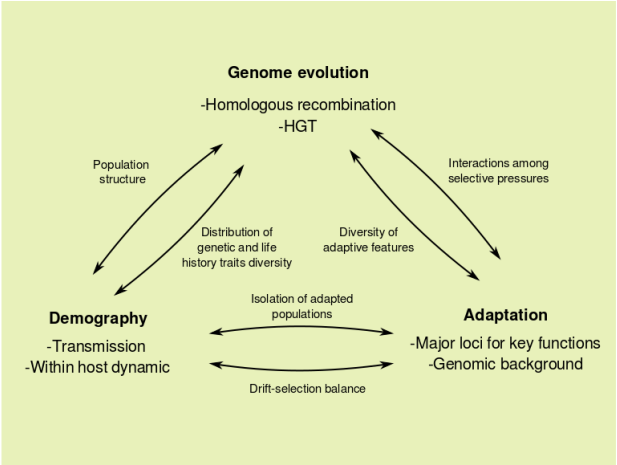

Figure 2: Illustration d'interactions entre processus écologiques et évolutifs étudiés dans mes travaux et identification de sources d'informations exploitées pour les documenter.

4 Impact de la recombinaison sur les patrons de diversité génomique

Dans l'optique d'étudier des aspects évolutifs liés aux interactions durables qui pouvaient déboucher sur des applications concrètes, j'ai intégré à la rentrée 2000 un DEA dont la formation comprenait des modules consacrés à l'amélioration des plantes et aux interactions qu'elles établissaient avec les microbes. Un des aspects marquant de cette formation était l'importance des processus de recombinaison sur l'émergence de polymorphismes d'intérêt chez les plantes et sur leur valorisation au travers des approches de sélection variétale. Toutefois, j'avais pu observer que l'étude des processus adaptatifs chez les eucaryotes était compliquée, *via* le screening d'une banque de mutants aléatoires sur *Arabidopsis thaliana*, ou l'étude du niveau de spécialisation d'un acarien phytophage à l'aide de marqueurs microsatellites neutres (Bailly *et al.*, 2004). Comme indiqué plus haut, j'ai donc voulu transposer ces questions à l'interface entre écologie évolutive et fonctionnelle chez les bactéries, la caractérisation fonctionnelle étant plus facilement abordable sur ces modèles biologiques.

Ce positionnement s'est fait à une période d'acquisition de connaissances sur les modalités évolutives de nombreux groupes bactériens avec la standardisation des approches de caractérisation par séquençage multilocus à la fin des années 90 (Maiden *et al.*, 1998) puis par des approches génomiques (Maiden *et al.*, 2013). Ces approches ont notamment permis de mieux décrire l'impact des processus de recombinaison sur la diversité génétique des populations bactriennes. Deux types de recombinaison, sont différenciés en fonction de leurs conséquences sur les génomes bactériens : la recombinaison homologue, qui aboutit au remplacement d'un fragment d'ADN par une séquence similaire, et la recombinaison hétérologue, qui aboutit à l'insertion et/ou la délétion de matériel génétique.

L'impact de la recombinaison homologue sur les patrons de diversité bactériens a notamment été formalisé au début des années 1990 par un article publié par John Maynard Smith *et al.* (Maynard Smith *et al.*, 1993), qui résume une grande partie des problématiques explorées durant les vingt années suivantes dans ce domaine : i) la diversité des taux de recombinaison entre espèces bactériennes ; ii) L'impact du taux de recombinaison sur la diversité des populations bactériennes ; iii) l'impact de la diversité générée par recombinaison sur les processus adaptatifs et démographiques dans les populations bactériennes. Il existe une forte hétérogénéité du ratio du taux de recombinaison homologue sur le taux de mutation, ρ/μ, entre espèces bactérienne (Hanage *et al.*, 2006). Le taux de recombinaison diminue avec la divergence (Vulić *et al.*, 1997). Dans un modèle neutre, ces variations jouent un rôle important dans la diversité et la structure génétique des populations bactériennes (Fraser *et al.*, 2007). Pour des faibles ratios ρ/μ, différents groupes génétiques émergent naturellement. Toutefois, la recombinaison permetre d'homogénéiser le patrimoine génétique de différentes lignées génétiques. Quand la probabilité de recombinaison r devient plus importante que la probabilité de mutation m, les mutations en ségrégation peuvent se retrouver en équilibre de liaison. Ceci contribue à la cohésion des fonds génétiques à l'échelle génomique, et permettre en théorie aux pressions de sélections d'agir indépendamment sur différents loci (Hill & Robertson, 1966). En fonction de la distribution de la valeur sélective des allèles sous sélection et du niveau de

recombinaison, les pressions de sélection les plus intenses vont avoir une forte influence sur la distribution de la diversité observée à l'échelle génomique (Kim & Stephan, 2000; Majewski & Cohan, 1999).

L'intensité de la recombinaison hétérologue, mise en évidence par un fort polymorphisme de contenu en gène dans les populations bactériennes, a été illustrée dès les premiers projets de génomique comparative au sein d'espèces bactériennes, notamment chez *Escherichia coli* (Lawrence & Ochman, 1998). Les importantes différences de composition en gènes entre souches de la même espèce ont mené à l'émergence de la notion de pan-génome, l'ensemble des gènes portés par ces individus. Le pan génome se décompose en un *core-génome*, partagé par l'ensemble des bactéries étudiées, et d'un génome accessoire qui n'est pas présent chez l'ensemble des individus (Tettelin *et al.*, 2008). La présence d'une fraction accessoire dans les génomes bactériens a posé différentes questions, notamment sur la contribution de ces gènes à la valeur sélective des bactéries (Gogarten & Townsend, 2005). Bien que différents modèles ont permis de définir un certains nombre d'hypothèses sur l'évolution du pan-génome, de nombreuses zones d'ombre persistent encore concernant sa dynamique et son impact sur la valeur sélective (Brockhurst *et al.*, 2019).

Dans ce contexte, je me suis intéressé à la distribution des événements de recombinaison, homologue et hétérologue, à différentes échelles évolutive et génomique. Je me suis intéressé à la distribution des événements de recombinaison entre groupes génétiques d'apparentement variable, afin de mieux cerner le rôle de la recombinaison en tant que force de cohésion génétique au sein des populations étudiés. J'ai également étudié la distribution des événements de recombinaison le long des génomes étudiés, afin de mieux déterminer si ceci pouvait expliquer la capacité de certains caractères d'intérêt à évoluer de façon indépendante des fonds génétiques.

4.1 Recombinaison entre taxons apparentés: des mesures de déséquilibre de liaison aux taux de recombinaison intra et inter-taxons

Ainsi, depuis ma thèse, j'ai développé des travaux visant à illustrer l'impact de la recombinaison par des mesures indirectes, basées sur les niveaux de diversité observés en populations naturelles et tirant parti du développement des méthodes de séquençage.

4.1.1 Recombinaison chez les rhizobia associés aux luzernes, *S. meliloti* et *S. medicae*

L'article de Maynard Smith *et al.* cité plus haut mettait en évidence un déséquilibre de liaison important lié à la présence de deux groupes de souches génétiquement isolées à partir d'un jeu de données d'allozymes caractérisant les symbiotes de luzernes du genre *Medicago*. Ces deux groupes correspondent aux espèces *S. meliloti* et *S. medicae* décrites plus tard sur la base de données de séquence 16S et de tests phénotypiques. Au sein de *S. meliloti* une caractérisation des séquences de l'ARNr 16S et des tests phénotypiques basés sur la spécificité d'hôte de différents isolats ont mené à la description de biovars, des groupes de souches associés à des plantes "hôtes" différentes au sein de la même espèce bactérienne. Juste avant le démarrage

de mes travaux, le génome de la souche 1021 de *S. meliloti* a été publié. A l'image de ce génome, *S. meliloti* et *S. medicae* ont -au moins- un chromosome et deux mégaplasmides circulaires (Capela *et al.*, 2001). Il existe une différence marquée de distribution des gènes impliqués dans différentes fonctions : i) les gènes liés au métabolisme de base sont localisés sur le chromosome, ii) le premier mégaplasmide inclut un grand nombre de gènes impliqués dans des métabolisme secondaires et dans la synthèse de polysaccharides de surface, iii) Le dernier mégaplasmide porte la quasi totalité des gènes impliqués dans les fonctions symbiotiques.

Ces données offraient des conditions particulièrement favorables pour caractériser par séquençage multi-locus des populations de *S. meliloti* et *S. medicae* afin d'approfondir, entre autres, les connaissances sur les patrons de recombinaison entre ces deux espèces, entre les biovars de *S. meliloti*, et au sein de chaque biovar. Mes premiers travaux ont été menés sur des panels de souches isolées à partir de sols provenant de trois localités mis en contact ex-situ avec différents génotypes de plantes hôtes (Bailly *et al.*, 2006, 2007). Ils ont notamment impliqué Stéphane de Mita dans le cadre d'un stage de M1. Le choix de travailler à partir de collections de souches isolées à partir des mêmes sols se basait sur l'hypothèse qu'une partie significative des échanges de gènes entre les lignées étudiées puisse se faire localement. Du fait de leur localisation sur les trois unités de réplication majeures, le choix des marqueurs s'écartait de la définition classique d'un schéma de caractérisation multilocus. L'idée sous-jacente à ce choix était qu'une différence significative pouvait exister entre les taux de transfert horizontaux observés sur les différentes unités de réplication. Le désavantage de ce choix était que, outre l'existence de mécanismes de recombinaison spécifiques, les pressions de sélection agissant sur chaque marqueur puissent différer et participer à l'établissement des patrons de polymorphisme observés. En outre, les marqueurs utilisés étaient situés dans des régions inter-géniques. Ce positionnement visait à maximiser le niveau de polymorphisme observé, en espérant que la décroissance du déséquilibre de liaison le long des génomes étudiés soit forte et que la sélection joue par conséquent un rôle moins important sur le niveau de polymorphisme des régions non-codantes. Il faisait toutefois l'hypothèse que la faible divergence entre les deux espèces étudiées, considérées comme "sœurs" à l'époque, se traduise par une forte synténie.

A l'aide d'une approche phylogénétique nous avons pu démontrer la monophylie de *S. medicae*, sur les différents marqueurs utilisés. Ce résultat confortait les précédent résultats pointant l'isolement reproducteur de *S. meliloti* et de *S. medicae*.

Au contraire, au sein de *S. meliloti* et de *S. medicae* des approches populationnelles, différenciation et déséquilibre de liaison, illustraient un impact significatif de la recombinaison homologue au niveau sub-spécifique, même entre biovars.

Les données accumulées ne laissaient guère de doute sur un déficit de recombinaison entre les espèces étudiées. Il m'a alors semblé intéressant de documenter de manière plus quantitative les patrons de recombinaison intra-spécifiques et d'isolement inter-spécifique pour documenter les processus mis en jeux lors de l'émergence de nouvelles lignées/espèces bactériennes. L'acquisition de données génomiques par séquençage haut-débit, encore balbutiante, offrait des perspectives intéressantes pour identifier assez de polymorphisme pour poser cette question.

4.1.2 Recombinaison au sein des espèces cryptiques incluses dans *Rhizobium leguminosarum*

Mon post-doctorat m'a donné l'opportunité de travailler dans cette voie à partir d'un jeu de données comportant 72 génomes partiels issus de souches de *Rhizobium leguminosarum* isolées en sympatrie: 36 génomes de souches du biovar viciae associées aux vesces et 36 génomes de souches du biovar trifolii associé aux trèfles (Kumar *et al.*, 2015).

La diversité génomique observé au sein du groupe *Rhizobium leguminosarum* c'est avérée bien plus importante que chez *S. meliloti* et surtout *S. medicae* (Figure 3). De fait, le calcul de l'identité nucléotidique moyenne sur un panel de "*core genes*" a mis en évidence l'existence de plusieurs groupes, représentant potentiellement plusieurs espèces génomiques sur la base des critères ad hoc (Jain *et al.*, 2018). De même, le patron de divergence entre les espèces génomiques de notre jeu de données en terme de présence/absence des gènes identifiés dans le génome de la souche *R. leguminosarum* biovar viciae 3841 est relativement marqué (Figure 4). Le développement d'outils permettant d'intégrer dans un cadre phylogénétique les processus de substitution et de recombinaison à partir d'alignements à l'échelle génomique offrait la possibilité d'intégrer le long d'un arbre, des mesures des taux de recombinaison et des taux de mutation (Didelot & Falush, 2007). A l'aide de ce jeu de données, des analyses ont été conduites de manière itérative à différentes échelles évolutives, mettant en évidence des différences substantielles des ratios des taux de recombinaison sur le taux de mutation.

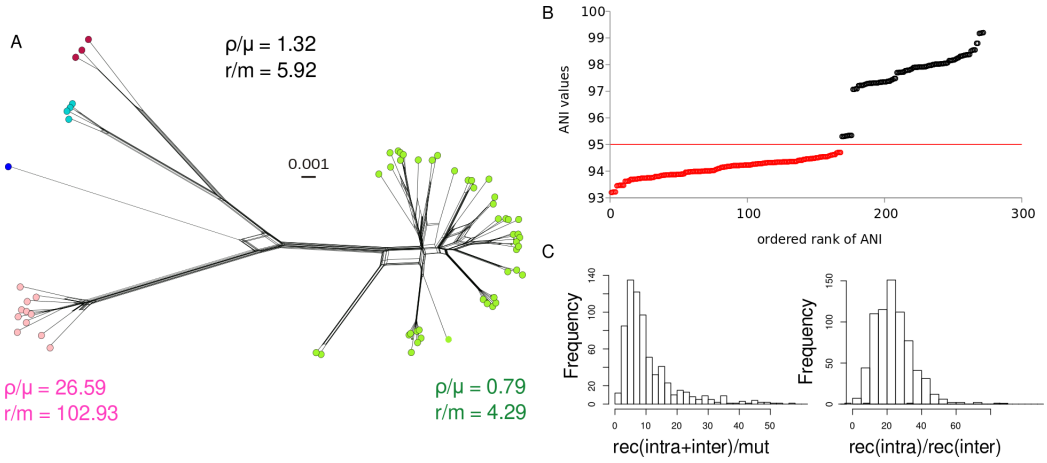

Figure 3: Diversité génomique de 72 souches de *Rhizobium leguminosarum*. A) Réseau phylogénétique basé sur l'information de 305 *core-genes*, les couleurs des feuilles représentent les espèces génomiques identifiées par analyse de l'identité nucléotidique moyenne. Les inférences des ratios i) des taux de recombinaison/mutation (ρ/μ, et ii) des probabilités de recombinaison et de mutation par site (r/m) obtenus par Clonal Frame pour l'ensemble du jeu de données et les deux espèces les plus échantillonnées sont figurés respectivement en noir, rouge et bleu. B) Distribution de l'identité nucléotidique moyenne entre paires de génomes étudiés. C) Distributions a posteriori du ratio des probabilités qu'un site soit affecté par une recombinaison plutôt que par une recombinaison et du ratio des taux de recombinaison intra-espèce génomique et interespèces génomiques obtenues par calcul bayésien approximé à l'aide d'un simulateur de coalescent avec recombinaison structurée décrit dans le chapitre suivant.

Les analyses effectuées sur *Rhizobium leguminosarum* permettent d'obtenir une idée plus précise de l'impact de la recombinaison sur l'évolution de ce groupe bactérien. Ces mesures

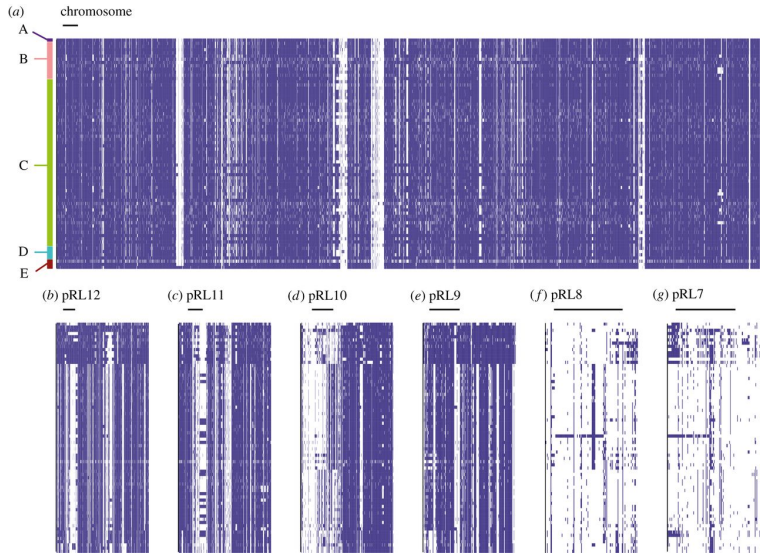

Figure 4: Diversité de présence/absence des gènes identifiés dans le génome de la souche *R. leguminosarum* biovar viciae 3841 parmi nos 72 isolats de *Rhizobium leguminosarum*.

suggèrent de forts taux de recombinaison à l'échelle intra-spécifique. Ces taux semblent susceptibles d'empêcher la divergence de lignées au moins dans un contexte de neutralité sélective. Des approches expérimentales ont montré chez le genre frère *Agrobacterium* (*Agrobacterium* et *Rhizobium* sont des genres paraphylétiques) que le différentiel de recombinaison hétérologue de part et d'autre de la frontière d'espèces était très limité (Costechareyre *et al.*, 2009). Dans la mesure où nous avions identifié un isolement reproducteur, il m'a semblé intéressant d'essayer d'obtenir des mesures indépendantes pour les taux de recombinaison intra et inter-spécifiques.

4.1.3 Diminution de la recombinaison homologue entre espèces du complexe *Borrelia burgdroferi*

Pour aller plus loin, je me suis intéressé aux approches de calcul bayésien approximé s'appuyant sur des simulations de coalescent avec des taux de recombinaison paramétrables alors que je quittais l'université de York pour Clermont-Ferrand (Beaumont, 2010). Malgré différents problèmes de mise à l'échelle des éléments génomiques simulés, cette solution plastique et méthodologiquement abordable me semblait séduisante.

Ces approches ont été abordées dans le cadre du modèle *Rhizobium leguminosarum* lors du stage de Master 2 de Maude Jacquot. Elle ont ensuite été étudiées sur des génomes issues de souches du complexe d'espèce *Borrelia burgdorferi*, qui inclut les agents de la maladie de Lyme, dans le cadre d'un projet qui a constitué une partie du sujet de thèse de Maude Jacquot. Ce projet a également impliqué Mathieu Gonnet (post-doctorant). Il a été réalisé en collaboration avec les membres de l'Institut Pasteur alors responsables du le Centre National de Recherche sur *Borrelia*, notamment Muriel Cornet et Elisabeth Ferquel. Nous avons séquencé les génomes de 63 souches de *Borrelia garinii*, *Borrelia burgdorferi* et *Borrelia afzelii* .

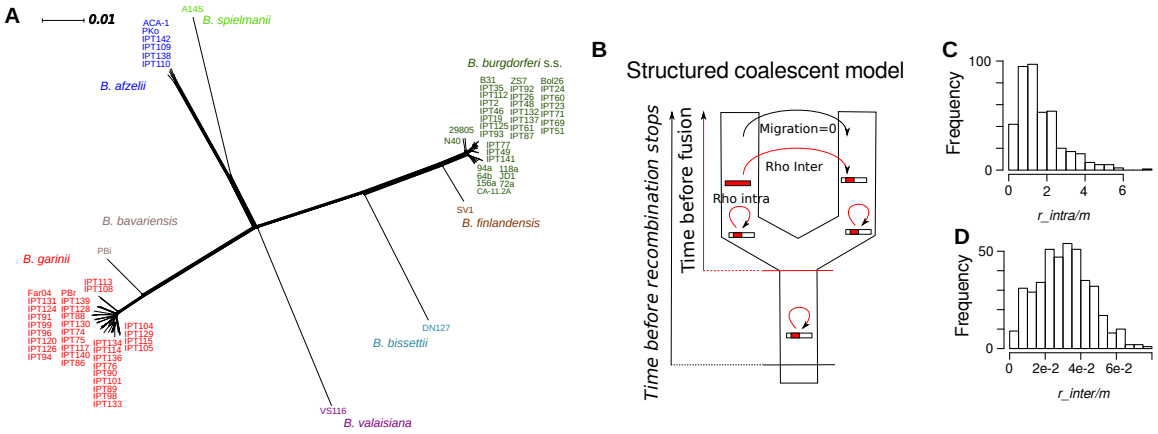

Figure 5: Structure génétique des chromosomes d'un panel de souche du complexe d'espèces *Borrelia burgdorferi* isolées en Alsace de tiques *Ixodes ricinus* à l'affût. A) Phylogénie chromosomique des souches étudiées et des références disponibles au moment de la publication. B) Schéma illustrant le protocole de simulations utilsé pour étudier les taux de recombinaison et de mutation dans les souches de *B. garinii* et *B. burgdorferi* étudiées. C et D) Distribution a posteriori des ratios des probabilités de mutation intra et inter-population par site sur la probabilité de mutation.

Les espèces échantillonnées peuvent facilement être différenciées sur la base de phylogénies chromosomiques (Figure 5). Nous avons donc initialement vérifié par comparaison à un modèle nul que des traces de recombinaison étaient identifiables dans ce jeu de données. En calculant les statistiques nécessaires pour conduire une approche ABC, nous avons vérifié que la recombinaison induit chez les espèces de ce complexe, comme attendu, une décroissance du déséquilibre de liaison en fonction de la distance physique, contrairement à ce qui avait été indiqué dans la littérature, (Haven *et al.*, 2011). Enfin, des simulations de coalescent ont été réalisées en différenciant des taux de recombinaison intra et inter-spécifiques et les résultats ont été utilisés pour approcher ces taux sur la base de l'écart à différentes statistiques observées pour *B. burgdorferi* sensu stricto et *Borrelia garinii* (Jacquot *et al.*, 2014b).

Le ratio du taux de recombinaison sur le taux de mutation semble plus faible que celui observé pour *Rhizobium leguminosarum*. Les résultats du modèle suggèrent également une probablité de recombinaison intra-spécifique 50 fois plus importante que la probabilité inter-spécifique. Sur la base de l'étude des niveaux attendus de différenciation à partir du modèle de coalescent, les taux de recombinaison observés semblent toutefois assez important pour empêcher l'apparition de génotypes indépendants sur la base d'un modèle neutre et sans contraintes en terme de contact entre les différents génotypes. Ce patron de diversité, comme celui de *R. leguminosarum*, pose de fait la question des rôles joués par la sélection sur la divergence des espèces étudié.

Il est envisageable d'étudier avec le modèle développé le niveau de différenciation attendu entre groupes génétiques en fonction des taux de mutation, de recombinaison et d'une diminution de la recombinaison entre populations d'un point de vue plus général. Un nombre important de modèles se sont intéressés à la divergence des lignées bactériennes. Deux types de modèles principaux existent à ce sujet : i) des modèles neutres pointant le rôle de la recombinaison comme un mécanisme créant une force de cohésion au sein des lignées. Au delà d'un seuil dans le ratio du taux de recombinaison sur le taux de mutation, la recombinaison ho-

mogénéise la population simulée; ii) des modèles où la sélection influence le degré de structure de sous-population via des balayages sélectifs globaux ou limités à chaque sous population. Dans les deux cas, une décroissance du taux de recombinaison avec le niveau de divergence a été assumée. Étudier l'isolement reproducteur en terme de différence entre recombinaison inra et inter-lignées apporte ainsi un point de vue singulier, permettant de déterminer à partir de quel niveau de différence la divergence devient inéluctable (Figure 6). Ces travaux menés dans le cadre de la thèse de Maude Jacquot n'ont pas été publiés. Une publication qui aborde de manière relativement similaire cette question, présente des conclusions semblables (Marttinen & Hanage, 2017).

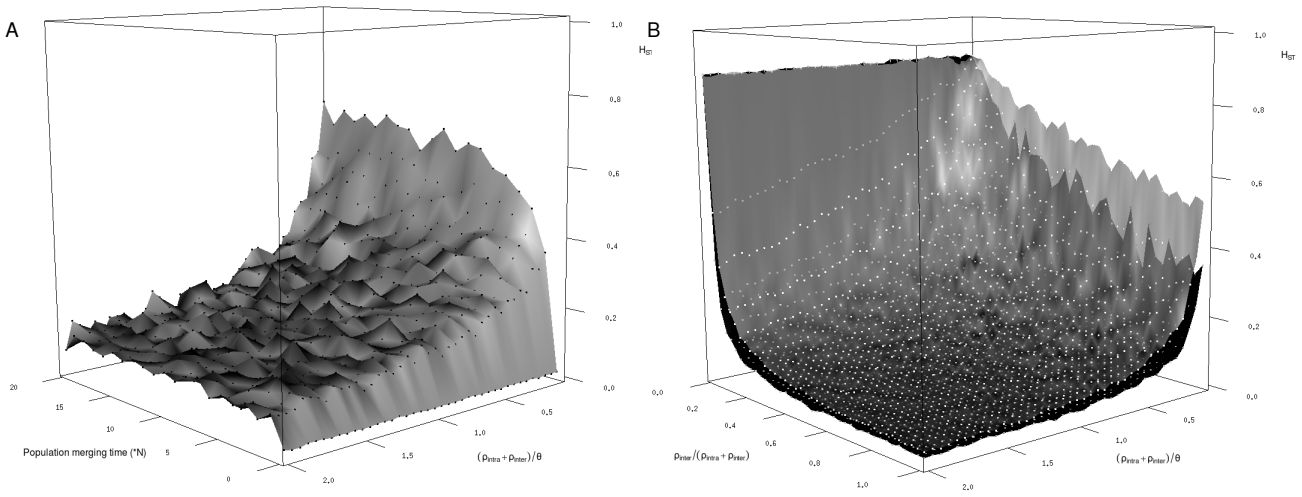

Figure 6: Étude des propriété du modèle de coalescent décrit plus haut. A) On peut déterminer le temps minimum nécessaire pour atteindre un équilibre de différenciation dans les populations simulées en fonction des taux de recombinaison intra et inter-populations. B) Le niveau de différenciation entre populations à l'équilibre dépend à la fois du taux de recombinaison global de chaque population et de l'influence relative des taux de recombinaison intra et inter-populations.

4.2 Distribution des événements de recombinaison au sein des génomes bactériens : des hypothèses liées à l'observation de marqueur au point de vue génomique

Les génomes bactériens ont évolué vers des structures complexes, à plusieurs reprises (Ochman, 2002). Ces structures complexes incluent plusieurs éléments de grande taille qui ont été appelés chromides (Harrison *et al.*, 2010). La ségrégation des fonctions généralement encodées par ces différentes unités de réplications posent la question d'une différence potentielle de dynamique évolutive entre les différents réplicons, tant en terme d'expression, de sélection, ou de recombinaison (Cooper *et al.*, 2010).

4.2.1 Recombinaison homologue et transferts au sein des différentes unités de réplication des génomes de *S. meliloti* et *S. medicae*

Cette problématique peut s'appliquer à différents rhizobia dont *S. meliloti* et *S. medicae*. Je me suis attaché à étudier des différences potentielles de taux de recombinaison homologue

entre unités de réplication chez ces deux espèces bactériennes. J'ai initié des approches basées sur l'association statistique entre les patrons de diversité à différents loci: le déséquilibre de liaison. Comme indiqué précédemment, ces travaux se sont basés sur les séquences de plusieurs marqueurs répartis sur les réplicons majeurs des bactéries étudiées. L'analyse des données de séquençage multi-locus obtenues durant ma thèse suggéraient une différence de taux de recombinaison entre le chromosome, pour lequel les marqueurs présentaient un déséquilibre de liaison fort, et les mégaplasmides, pour lesquels de faible valeurs de déséquilibre de liaison intra et inter-réplicon ont été observées. Ce dernier patron permettait d'émettre l'hypothèse que les mégaplasmides sont caractérisés par des taux de recombinaison homologue forts, chaque événement impactant seulement une partie du réplicon. Malgré ces résultats encourageants, cette d'étude se heurtait : i) d'une part au faible niveau de polymorphisme des populations de *S. medicae*, pas toujours suffisant pour utiliser des statistiques de déséquilibres de liaison; ii) d'autre part à la proximité de marqueurs sélectionnés dans le voisinage des marqueurs utilisés. La validation de ces hypothèses sur un plus large jeu de données était donc nécessaire.

Durant mon post-doctorat, j'ai eu l'occasion de prolonger cette étude par une approche de génomique des populations menée sur un ensemble de souches de *S. medicae* isolées en sympatrie en Angleterre (Bailly *et al.*, 2011). Au moment où les méthodes de séquençage haut débit prenaient leur essor, une majorité des projets de séquençage génomique visaient un séquençage complet à forte couverture avec une finission à façon. Nous avons fait le choix de travailler sur des séquences de génomes partielles sur une douzaine de souches. Si cette stratégie est bien sûr limitante pour différents aspects, elle offrait un compromis pertinent pour s'intéresser à différents aspects de la biologie des populations de *S. medicae*, dont l'impact de la recombinaison sur les patrons de diversité observés. Les capacités de séquençage et d'assemblage augmentant, les couvertures se sont sensiblement améliorées mais les études sur des génomes non complétés en génomique des populations bactériennes est devenu une approche usuelle.

En s'appuyant sur la comparaison des informations phylogénétiques contenues dans des fenêtres contiguës définies le long des génomes reconstitués, les hypothèses émises à partir des résultat de typage par séquençage multilocus ont été confirmées (Figure 7). L'information phylogénétique portée par les mégaplasmides, notamment le mégaplasmide portant les fonctions symbiotique, est plus hétérogène que celle portée par le chromosome. La recombinaison au sein des mégaplasmide prend la même forme que sur un chromosome, dominée par l'intégration de fragments de taille limitée.

Ces données sont à mettre en relation avec une dynamique de perte et de gain de gènes également structurée par unité de réplication (Figure 8). La composition du chromosome parait en effet plus stable que celle des mégaplasmides, le mégaplasmide portant les fonctions symbiotiques présentant un polymorphisme particulièrement important. De fait, des liens mecanistes peuvent existent entre recombinaison homologue et hétérologue (Vetsigian & Goldenfeld, 2005). Ce résultat confortait les hypothèses émises par une équipe italienne concernant la dynamique de perte et d'acquisition de gènes issues de l'analyse de données d'hybridation génomique comparatives réalisées sur un panel de souches de *S. meliloti* (Giuntini *et al.*, 2005).

Cette différence de dynamique évolutive entre unités de réplication se retrouve dans le cadre de comparaisons génomiques entre *S. medicae* et *S. meliloti*. On observe en effet une différenciation variable en fonction des unités de réplications des contenus en gènes de *S.*

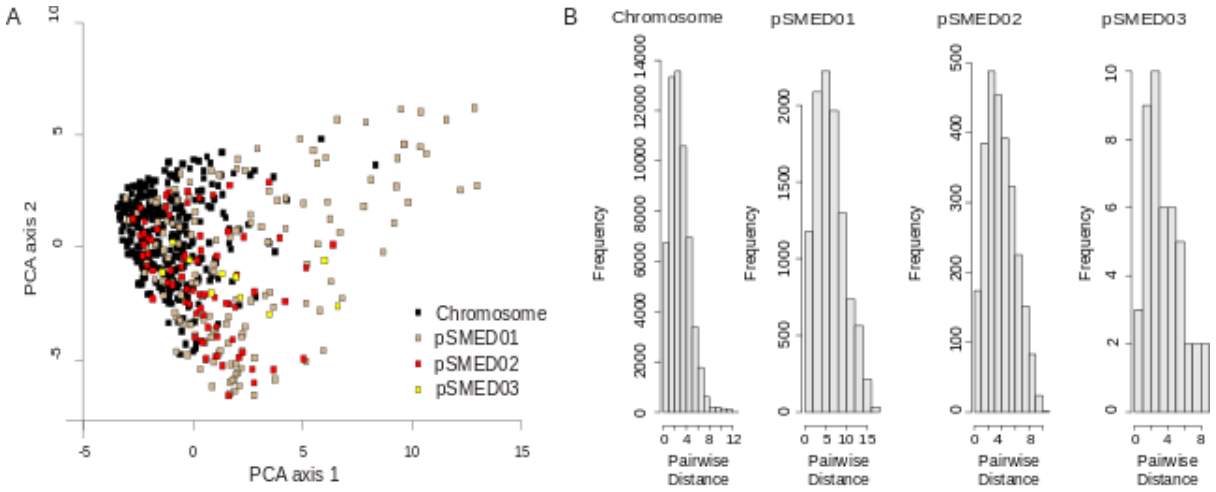

Figure 7: Analyse en composante principale illustrant la diversité des patrons de divergence entre les 12 souches de S. medicae étudiées entre des fenêtres de 10 kb définies sur la base du génome de S. medicae WSM 419. Distribution des distances euclidiennes entre les projections de chaque fenêtre et le barycentre pour les différentes unités de réplication.

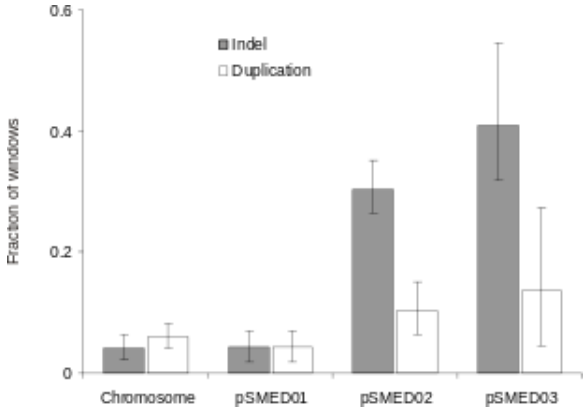

Figure 8: Fraction de fenêtres de 10 kb pour lesquelles une duplication ou une délection a été détectée parmi les 12 souches de S. medicae étudiées en comparant la couverture observée à une distribution de couverture attendue.

meliloti et *S. medicae* (Figure 9). Toutefois, les mécanismes mis en jeu restent à déterminer. Au delà de mécanismes de recombinaisons spécifiques aux mégaplasmides, on peut envisager l'effet de pressions de sélection de premier ordre, via la sélection de caractères influençant la valeur sélective des individus, tout comme des pressions de sélection de second ordre, via l'augmentation de la capacité à générer de nouveaux variants ou une augmentation de la robustesse par modularité, n'ont pas été réfutés.

4.2.2 Polymorphisme des plasmides cp26 du complexe d'espèce *Borrelia burgdorferi*: quand la sélection exacerbe les traces de la recombinaison

Dans d'autres cas, le rôle de la sélection dans l'émergence de traces de recombinaison locus-spécifiques est moins ambigu. Dans le cadre des études génomiques menées au sein du complexe d'espèce *B. burgdorferi*, nous avons porté une attention particulière au patron de poly-

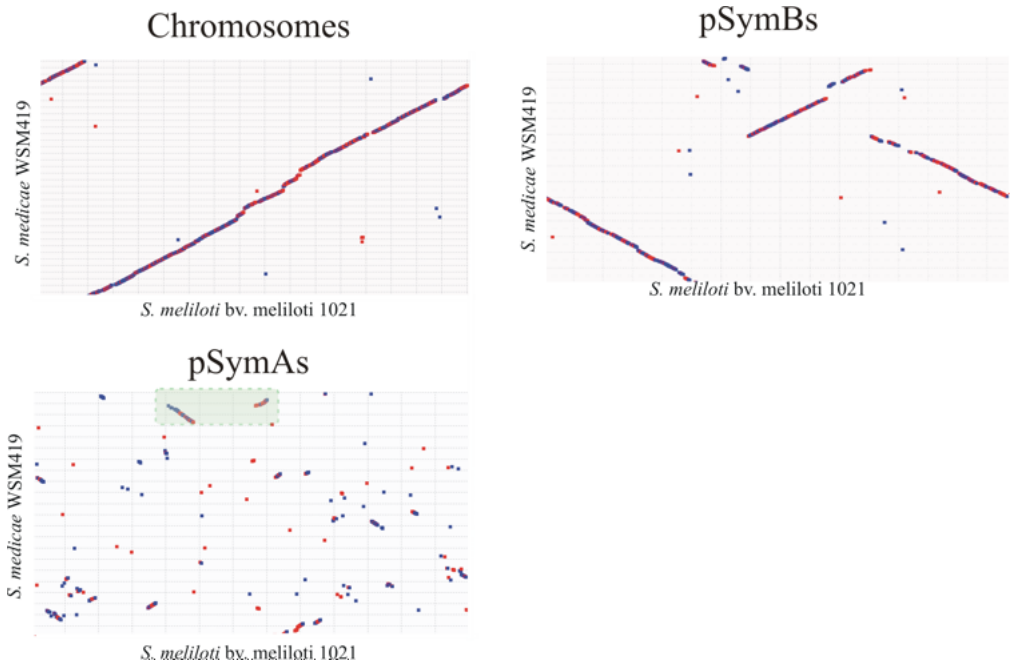

Figure 9: Illustration de la conservation de k-mers nucléotidiques entre deux souches de *S. meliloti* et *S. medicae* en fonction de l'unité de réplication considérée. Les k-mers conservés correspondant aux gènes de nodulation et de fixation d'azote sont encadrés en vert.

morphisme observé au sein du plasmide cp26. Ce plasmide contient en effet un gène encodant un antigène majeur reconnu par les systèmes immunitaires de vertébrés, *ospC*. Même si ce gène possède des fonctions susceptibles d'entraîner des pressions de sélection variées, son potentiel immunogène est responsable d'une sélection fréquence dépendance négative promouvant un fort niveau de polymorphisme. La recombinaison joue un rôle majeur dans l'évolution de ce locus, en créant de nouveaux allèles présentant de forts niveaux de divergence globaux par rapport à la diversité existante. Ces nouveaux allèles, s'ils restent fonctionnels, sont alors susceptibles d'augmenter en fréquence dans la population.

De fait, on peut en constater une forte baisse du déséquilibre de liaison au niveau du gène *ospC* (Figure 10). Une étude phylogénétique des régions flanquant le gène *ospC* suggère que les recombinaisons inter-spécifiques, malgré leur relativement faible fréquence à l'échelle du génome, ont eu un rôle significatif sur l'évolution de ce locus.

Cette observation souligne l'importance des interactions, à l'échelle des génomes, entre recombinaison et sélection. Elle illustre le potentiel des approches basées sur des études de diversité pour identifier des régions clefs dans l'adaptation des taxons bactériens.

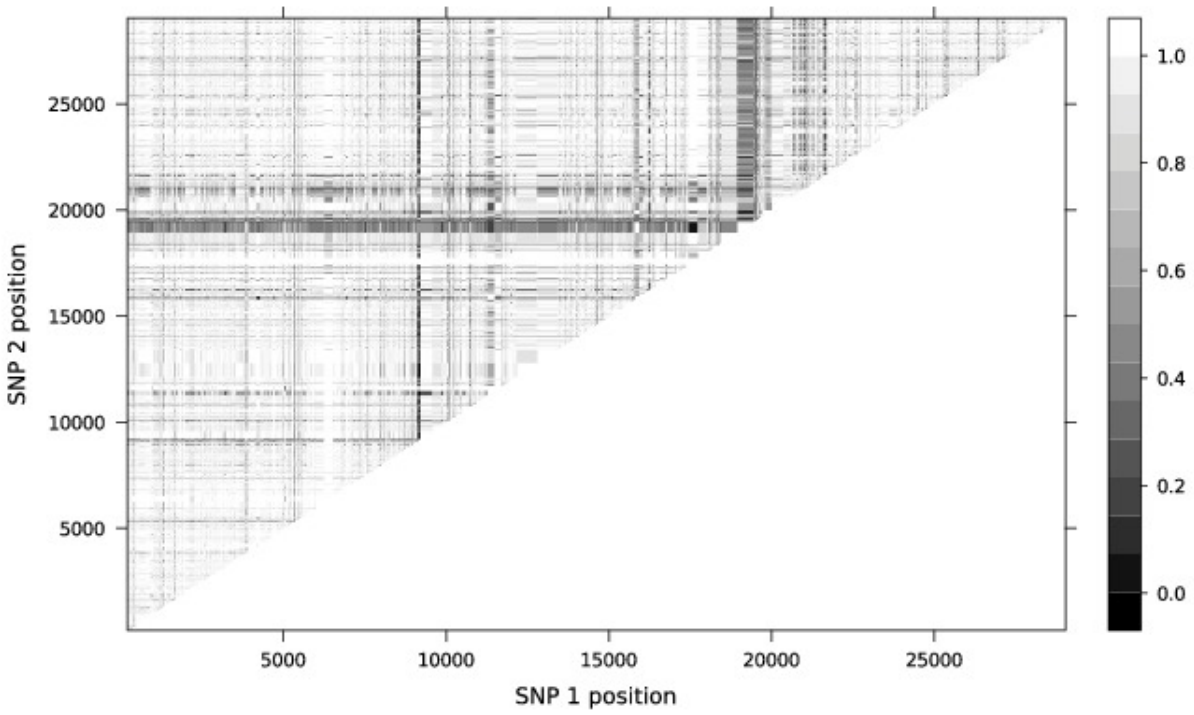

Figure 10: Déséquilibre de liaison à l'échelle interspécifique sur le plasmide cp26

5 Effet de la sélection sur les patrons de diversité & identification des déterminants génétiques de phénotypes

La connaissance des fonctions impliquées dans la valeur sélective et la spécialisation écologique des bactéries repose encore largement sur les résultats des approches de génétique, notamment de génétique inverse. Si l'approche génétique est indispensable pour obtenir une validation fonctionnelle des gènes, elle présente aussi certaines limites. Les outils permettant la mise en place d'une approche de mutagenèse, de complémentation fonctionnelle, et a fortiori les approches culturales, doivent être disponibles. De plus, la redondance fonctionnelle peut poser certains problèmes pour identifier la fonction d'un gène d'intérêt. Enfin, ces expérimentations sont souvent limitées à un nombre limité de souches modèles, ce qui pose des problèmes compte tenu de la notion de pan-génome : certains gènes codant pour des fonctions importantes peuvent être absents du génome étudié, et les gènes étudiés lors d'approches mécanistes ne sont pas forcément ceux qui sont les plus intéressants d'un point de vue évolutif. Ce dernier point est d'autant plus important que le débit des approches génétiques reste dans un grand nombre de cas limité. Le temps nécessaire pour obtenir une caractérisation peut être limitant par rapport au nombre de gènes qui pourraient être caractérisés fonctionnellement. Dans ce contexte, les approches moléculaires, indirectes, peuvent complémenter les approches génétiques. Ainsi, les approches transcriptomiques sont par exemple devenues courantes pour identifier des gènes candidats, dont la fonction est étudiée a-posteriori.

Des approches se basant sur des notions de biologie évolutive peuvent également compléménter cet ensemble d'outils. La mesure des associations entre patrons de polymorphisme et phénotypes, dans l'esprit des approches de génétique quantitative communes en amélioration des plantes, ou plus généralement les approches de "scan génomique" peuvent conduire à l'identification de régions génomiques d'intérêt au sein des génomes bactériens (Falush & Bowden, 2006). Ces approches avaient été peu utilisées chez les bactéries avant la révolution génomique. Compte tenu du mode de recombinaison des bactéries, l'efficacité de ces approches restait à prouver.

Différentes études illustrent des résultats convaincants (Shapiro *et al.*, 2012), même si les gènes identifiés ont parfois des fonctions surprenantes au premier abord (Sheppard *et al.*, 2013). L'utilisation d'approches basées non pas sur la fréquence de différents polymorphismes, mais sur leur nature (*i.e.* mutations synonymes/non synonymes) peuvent également apporter des informations critiques pour l'identification de gènes candidats (*e.g.* Almeida *et al.*, 2017; Mongodin *et al.*, 2013; Vayssier-Taussat *et al.*, 2010). Au delà de l'identification des gènes d'intérêt, ces patrons de diversité renseignent aussi sur les pressions de sélection auxquels ils sont soumis. Ceci peut être utile pour repenser le lien entre observation in-situ et résultats expérimentaux et pour envisager la réussite de stratégies de maîtrise basées sur ces gènes, par exemple en estimant le risque de contournement d'une stratégie de lutte contre un pathogène (Vera Cruz *et al.*, 2000), ou en envisageant le maintien dans une population bactérienne d'un gène ou d'un allèle permettant une symbiose efficace.

Après avoir caractérisé les patrons de polymorphisme de régions candidates, je me suis intéressé à l'identification et à l'évolution des gènes impliqués dans la mise en place d'interactions symbiotiques spécifiques chez les rhizobia à une échelle génomique. Appliquer cette approche à un modèle biologique qui avait fait l'objet d'études génétiques me semblait particulièrement pertinent pour valider les résultats obtenus. J'ai recherché des gènes accessoires importants pour la valeur sélective de rhizobia. Au sein du complexe *B. burgdorferi*, j'ai étudié la distribution du polymorphisme génomique au sein de leur tiques vectrices à la recherche de régions susceptibles d'influencer le spectre d'hôte. Je m'intéresse aujourd'hui à la relation entre la diversité génomique d'une espèce du complexe et la capacité à infecter l'homme.

5.1 Spécificité d'hôte chez les rhizobia, de l'étude de régions candidates aux associations entre diversité génomique et phénotype d'intérêt

Les déterminants génétiques liés à la mise en place de l'interaction entre bactéries fixatrices d'azote et plantes ont fait l'objet d'études approfondies car la maîtrise des mécanismes mis en jeux est cruciale pour optimiser les applications agronomiques (Masson-Boivin *et al.*, 2009). Dans le cas de la majorité des rhizobia, un ensemble de gènes appelés *nod*, souvent co-localisés au sein d'îlots génomiques symbiotiques, sont apparus comme des facteurs clefs dans le dialogue établi avec les légumineuses pour mettre en place une symbiose fonctionnelle. Par ailleurs, un ensemble d'autres mécanismes semblent influencer soit la mise place de l'interaction, soit son efficacité. Par exemple, dans de nombreux couples rhizobium-légumineuse, les gènes impliqués dans la synthèse de polysaccharides de surface jouent un rôle

important dans l'établissement de l'infection de la plante par la bactérie. La théorie voudrait que pour les différents systèmes de reconnaissance, une pression de sélection purificatrice soit à l'œuvre. En effet, un mutant affecté dans sa capacité à reconnaître son partenaire symbiotique devrait voir sa valeur sélective sensiblement amoindrie.

J'ai participé durant ma thèse à la description de *S. meliloti* biovar meliloti et *S. meliloti* biovar medicaginis, deux groupes de souches conspécifiques présentant des spectres d'hôte différents (Villegas *et al.*, 2006). J'ai ensuite pu étudier le niveau de différenciation entre des souches isolées de deux sols et appartenant soit aux biovars meliloti et biovar medicaginis de *S. meliloti*, soit à *S. medicae* (Bailly *et al.*, 2007). Dans ce but, les souches isolées ont été caractérisées à l'aide de quatre marqueurs dont un localisé au sein de l'îlot symbiotique.

Les biovars de *S. meliloti* forment des groupes paraphylétiques, sauf dans le cas des marqueurs liés aux gènes symbiotiques. Ce patron est en accord avec une étude génétique qui a identifié le gène *nodC* comme le déterminant de la différence de spécificité entre ces deux biovars de *S. meliloti* (Barran *et al.*, 2002). Au niveau du marqueur situé au sein de l'îlot symbiotique, l'étude différentes statistiques indique des traces de forte sélection disruptive entre biovars. Plus conformément aux attendus, de traces de sélections purificatrices sont identifiées au sein des biovars. Ces résultats suggèrent que la recombinaison au sein de *S. meliloti*, au delà de son rôle cohésif au niveau des fonds génétiques, permet une évolution indépendante du système de spécificité : les différents variants fonctionnels de l'îlot symbiotique sont distribués dans une large part des fonds génétiques existants.

Ces observations interrogent sur la proportion du génome des rhizobia sensiblement structurée par la spécificité d'hôte au delà des gènes de nodulation *nod*. Le développement des méthodes de séquençage haut-débit offrait la perspective de prolonger ces études sur régions candidates en identifiant les déterminants de l'adaptation à la plante hôte à l'échelle génomique.

Dans cette perspective, la masse de données à étudier et la difficulté à comparer les structures génétiques inférées à partir de différents marqueurs représentait un challenge. J'ai contribué avec Sandrine Pavoine, qui travaille sur des problèmes statistiques appliqués à l'écologie, au développement d'une analyse multivariée visant à comparer les patrons de structuration génétique observés pour un grand nombre de locus. Cette méthode a été validée sur données simulées et sur données réelles (Pavoine & Bailly, 2007). La réalisation de cette analyse repose toutefois sur l'identification a priori de groupes de souches associées à des conditions environnementales différentes et sur l'obtention de jeux de données sans données manquantes. Cette collaboration a été importante pour moi, afin de mieux envisager les analyses qui pouvaient être effectuées sur les jeux de données génomiques. C'est toutefois à l'aide d'outils plus classiques que je me suis penché avec Sébastien Guizard et Maude Jacquot à Clermont, ainsi que Nitin Kumar à York, sur l'étude de la spécificité d'hôte au sein des génomes de souches de *R. leguminosarum* issues de vesces ou de trèfles durant mon post-doctorat. *Rhizobium leguminosarum* inclus les biovars viciae et trifolii associées respectivement aux trèfles et aux vesces (notamment), qui sont également largement paraphylétiques (Figure 11).

La divergence entre les espèces cryptiques identifiées plus haut au sein de *R. legumi-*

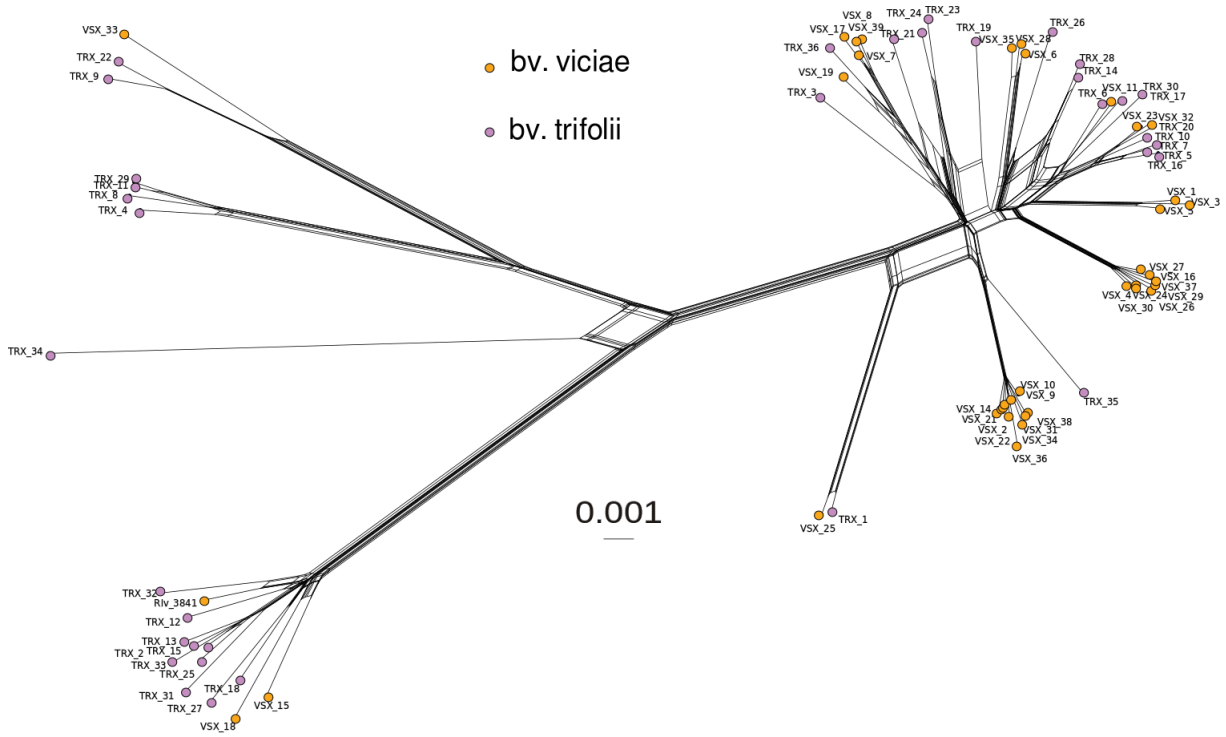

Figure 11: Diversité phylogénétique et symbiotique des isolats de *R. leguminosarum* séquencés. Les souches du biovars trifolii sont indiquées en violet et celles du biovar viciae sont figurées en orange.

nosarum explique une forte partie de la variance génétique observée. Ce signal phylogénétique est indépendant de la spécificité symbiotique (Kumar *et al.*, 2015). Dans ce contexte favorable, la distribution des patrons de différenciation nucléotidique entre biovars a été étudiée le long des génomes pour identifier des déterminants de la spécificité symbiotique. Une seule région génomique porte une marque de différenciation forte entre les deux biovars (Figure 12). Cette région est comme attendu centrée sur les gènes *nod* qui, comme dans le cas des biovars meliloti et medicaginis de *S. meliloti*, déterminent le spectre d'hôte de la bactérie.

D'un point de vue appliqué, le résultat obtenu suggère que l'identification des polymorphismes bactériens liés à la capacité d'établir une interaction symbiotique est largement dominée par la présence des allèles *nod* appropriés. Toutefois, les tris de collections basés uniquement sur les polymorphismes de gènes *nod* sont probablement peu efficaces pour sélectionner des souches à inoculer au delà de leur capacité à former une symbiose fonctionnelle. L'identification de génotypes bactériens associés à une meilleure efficacité symbiotique -du point de vue de l'agronome- doivent être recherchés par une association directe entre polymorphisme génomique bactérien et qualité agronomique du couple plante / rhizobium.

L'étude de ce type de phénotype, plus quantitatif, constituera un champs d'investigations à l'interface entre génomique des populations et biologie systémique. Elle a des implications dans le domaine de l'épidémiologie, par exemple dans les domaines de l'évolution de la virulence ou de résistance aux traitements. La faible réussite des approches d'association conduites sur la capacité de souches de *R. leguminosarum* à utiliser différentes sources de carbone illustre le travail à accomplir dans le domaine. Elle implique d'identifier les modèles

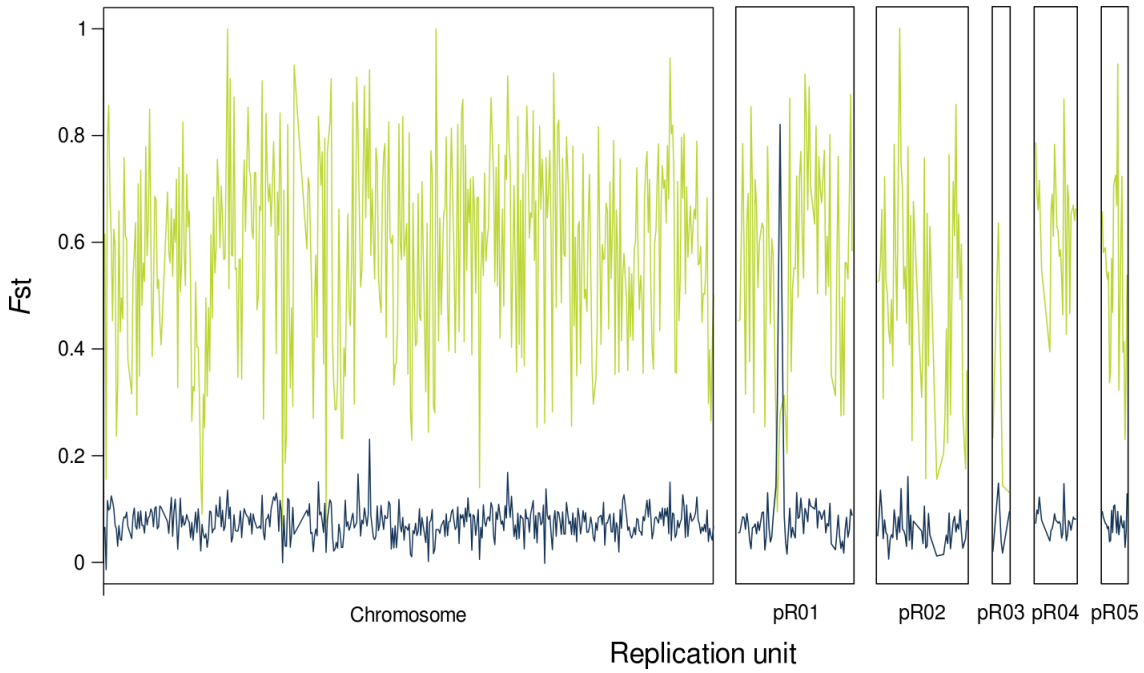

Figure 12: Distribution de mesures de différenciation génétique le long du génome de *Rhizobium leguminosarum* considérant les espèces génomiques (en vert) ou les plantes hôtes d'isolement (en bleu) comme facteur structurant.

permettant de retracer l'évolution de caractères quantitatifs comme l'expression de gènes le long d'une phylogénie en prenant en compte les contraintes biologiques associés à ces modèles (Bastide *et al.*, 2018), et d'envisager plus en détail les implications évolutives de ces résultats (Sabarly *et al.*, 2011).

5.2 Distribution de gènes accessoires déterminant d'autres caractères adaptatifs dans les populations de rhizobia

Différentes études montrent que le génotype des rhizobia influence, soit directement, soit par une interaction génotype bactérien / génotype hôte la productivité de la plante. Dans le cadre de l'analyse du contenu des génomes de souches de *S. medicae* présentés plus haut, j'ai recherché au sein de notre échantillon des gènes à haute fréquence dans la population étudiée, pourtant absents des génomes de référence déjà séquencés. Ce type de distribution, pour un gène potentiellement accessoire, représente une trace potentielle de rôle adaptatif, dans la mesure où s'approcher de la fixation dans une population de taille si importante semble difficile par le biais de la dérive (Gogarten & Townsend, 2005).

Le cluster de gène spécifique de notre population et le plus fréquent -présent dans l'ensemble de nos isolats- est composé de gènes *rtx*. Les homologues de ces gènes sont retrouvés chez de nombreuses bactéries associées aux plantes. Ils sont responsables de la synthèse d'une molécule appelée rhizobitoxine. Cette dernière entraîne chez les plantes hôtes une diminution de la synthèse d'éthylène. Or, ce composé est une hormone végétale qui régule de nombreux processus de développements chez les plantes. Notamment, l'éthylène inhibe la formation des nodosités chez les légumineuses. En manipulant la quantité d'éthylène produite, les bac-

téries pourraient donc accroître le nombre de nodosités, donc la capacité d'accueil de leur milieu. Bien que la validation fonctionnelle de ce cluster de gène n'ai pas été effectuée, il est probable qu'il soit un candidat intéressant pour expliquer un potentiel différentiel formation de nodules et potentiellement de productivité végétale lié au génotype bactérien. Un autre gène s'attaquant à la même voie métabolique, mais via une activité ACC-désaminase, est également présent à haute fréquence dans d'autres populations géographiques de symbiotes de luzernes. Ces gènes constituent une des bases génétiques d'un polymorphisme d'efficacité symbiotique.

Une étude récente illustre l'importance des gènes accessoires dans la compétitivité de différents isolats de Rhizobium leguminosarum du même biovars vis à vis du pois (*Pisum sativum*) et de la fève (*Viscia faba*) (Boivin *et al.*, 2019).

Les derniers projets que j'ai envisagés concernant les bactéries fixatrices d'azote concernaient l'adaptation des couples symbiotiques à l'environnement abiotique. On peut penser que dans différents cas, le résultat de l'interaction entre un microbe et son hôte puisse être dépendant non seulement du potentiel de croissance intrinsèque des individus engagés dans l'association mais aussi, de la réponse de l'interaction face au contraintes de l'environnement. Dans ce contexte, déterminer dans quelle mesure la divergence fonctionnelle liée à l'adaptation des organismes en interaction à des conditions abiotiques contrastées pouvait influencer l'évolution de leur spécificité d'interaction m'a semblé intéressant.

Ce point de vue m'a amené à participer à l'étude des couples symbiotiques impliquant la légumineuse *Serianthes calycina* et les bactéries du genre *Bradyrhizobium* associées. Cette association présente l'intérêt de se développer, en Nouvelle-Calédonie, sur des sols présentant des teneurs en nickel fortement variables. En effet, une partie de ces sols sont issus de la dégradation de roches ultramaphiques, très riches en nickel. L'étude préliminaire réalisée visait notamment à déterminer si les symbiotes des sols nickélifères et ceux des sols sans nickel se différenciaient en terme d'adaptation aux conditions édaphiques et d'identifier les déterminants de cette adaptation (Chaintreuil *et al.*, 2007).

En utilisant les données disponibles sur les déterminants de résistances aux métaux chez les bactéries, deux gènes candidats identifiés à partir de la bibliographie, *nreB* et *cnrA*, ont été recherchés parmi les isolats disponibles. Ces gènes ont été recherchés à l'aide de tests PCR basés sur des amorces établies à partir de l'alignement de séquences de référence. Alors que les deux gènes présentent un polymorphisme de présence / absence, seule la présence de *cnrA* est corrélée à la croissance des souches bactériennes sur milieu fortement enrichi en nickel. Un mutant du gène *cnrA*, obtenu par le Clémence Chaintreuil, se caractérise par une perte du phénotype de résistance au nickel. Ceci suggère un rôle important de *cnrA* dans le phénotype d'intérêt.

Un très bel article, que j'ai eu le plaisir de "reviewer", étudie cette problématique à l'échelle génomique, identifiant les déterminants potentiels de tolérance au nickel au sein des génomes de souches de *Mesorhizobium* par une approche d'association (Porter *et al.*, 2017).

5.3 Sélection et dynamique épidémiologique chez *Borrelia burgdorferi*

L'analyse de génomique des populations conduite sur des isolats du complexe d'espèce *Borrelia burgdorferi* issus de tiques nous a donné l'occasion de rechercher des gènes qui pourraient être impliqués dans l'adaptation aux différents hôtes vertébrés de la bactérie (Jacquot *et al.*, 2014b). La capture des hôtes potentiels de *Borrelia burgdorferi sensu lato* demande des moyens significatifs. Il est de ce fait important d'obtenir un maximum d'informations sur les différentes populations bactériennes retrouvées dans les tiques à l'affût. Toutefois, la détection de gènes potentiellement impliqués dans l'adaptation aux hôtes dans ce contexte ne peut pas se faire de manière directe, sur la base d'un screening phénotypique.

Dans chaque espèce, nous avons recherché des gènes présentant des patrons de polymorphisme typique d'une sélection diversifiante, qui pourrait être induite par l'adaptation à différents hôtes en fonction des génotypes, dans une collection de génomes correspondant à des isolats obtenus chez la tique vectrice. Malgré les résultats précédents qui illustraient l'importance potentielle des gènes accessoires dans la valeur sélective des génomes bactériens, nous avons dû nous contenter d'étudier les spectres de fréquence alléliques observés dans des gènes conservés dans les populations étudiées à l'aide du D de Tajima. Utilisant le fait que les souches avaient été prélevées sur deux sites distincts, nous avons également étudié l'évolution du niveau de différenciation le long des génomes étudiés, car ces données peuvent être impactées par les traces d'adaptation aux multiples hôtes infectés dans chaque site.

Ces analyses ont mis en évidence que les gènes impliqués dans la synthèse de lypo-protéines sont des cibles majeures de pressions de sélection diversifiante dans les trois unités de réplication majeures des espèces du complexe *B. burgdorferi*. Ces résultats corroborent ceux obtenus par une équipe concurrente à l'aide des ratios de mutation synonymes et non-synonymes (Mongodin *et al.*, 2013). Seul le gène *opsC*, évoqué plus haut, présente des traces de sélection significatives au sein des trois espèces, soulignant la diversité potentielle des pressions de sélection s'appliquant sur les différentes espèces du complexe *B. burgdorferi* et qui traduit peut-être différentes stratégies d'infection.

Au delà de ces résultats potentiellement intéressants, les analyses effectuées ont illustré trois problèmes conséquents :

- Le premier est lié à la structure des données de polymorphisme bactériens et aux limites associées. Le fort niveau de déséquilibre de liaison et structuration de la diversité chromosomique peut limiter l'efficacité des approches visant à étudier les relations entre génotype et phénotype sur la base de spectres de fréquences alléliques.
- Le second tient à la complexité du matériel plasmidique chez *Borrelia*. Les duplications multiples au sein des plasmides cryptiques mènent à des problèmes d'assemblage qui nous a conduit à ne pas utiliser une partie du matériel génétique bactérien dans le cadre de nos analyses. Définir une stratégie d'analyse permettant d'utiliser cette information semblait importante, d'autant plus que les plasmides cryptiques de ces bactéries semblent porter des gènes importants pour sa capacité d'infection.

- Le troisième tient au design expérimental utilisé : nous avons essayer de mettre en évidence des loci à caractère adaptatif dans les populations naturelles du complexe d'espèce *Borrelia burgdorferi*. Il me semble en effet important de comprendre quels sont les facteurs qui vont conduire à une forte exposition des populations humaines au pathogènes via une prévalence forte dans les tiques. Pour apporter des informations complémentaires sur le risque d'infection, il m'a semblé important d'étudier si certaines souches du complexe d'éspèce étaient plus capables d'infecter l'homme.

5.4 Variation du potentiel zoonotique chez *Borrelia garinii*

La maladie de Lyme représente un enjeu sanitaire et économique certain en Europe. En Allemagne et aux Pays-Bas, des études ont été conduites pour estimer les coûts associés aux cas humains. Les impacts ont ainsi été estimés des coûts sociétaux de 7 et 19 millions d'euros respectivement (Lohr *et al.*, 2015; Van Den Wijngaard *et al.*, 2017). Sur la bases des chiffres avancés dans ces études, et en assumant 26584 cas par an dont 954 hospitalisations reportés en France, on peut penser que les coûts de prise en charge des patients atteints de maladie de Lyme à l'échelle nationale sont au minimum de 4.5 millions d'euros par an. Les prises en charge hospitalières représentent la majorité de ces coûts, à raison de 3600 et 5700 euros alloués aux cas complexes en Allemagne et aux Pays-Bas. Dans ces conditions, il semble raisonnable d'investir pour améliorer la prévention des cas complexes de maladie de Lyme et leur prise en charge.

Seules certaines espèces du complexe d'espèce *Borrelia burgdorferi* sont préférentiellement associées à des cas humains (Stanek *et al.*, 2012). L'identification des lignées à risque et le suivi de leur distribution dans les tiques à l'affût constitue donc une information pertinente pour définir une politique de prévention. D'autre part, l'identification des mécanismes utilisés par les lignées à risque pour infecter l'homme peut ouvrir de nouvelles perspectives cliniques (Marcnkiewicz *et al.*, 2018). Dans ce contexte, *B. garinii* représente un modèle particulièrement intéressant. Cette espèce est largement impliquée dans les cas de neuroborréliose humaine de la majorité des cas de neuro-borréliose (Coipan *et al.*, 2016). Elle est fréquente dans les *Ixodes ricinus* à l'affût en Europe, une méta-analyse en Europe reportant sa présence dans 23% des tiques porteuses de bactéries du complexe. Cette espèce utilisant des oiseaux comme réservoir, les génotypes à risque sont susceptibles de diffuser rapidement sur l'aire de répartition (Vollmer *et al.*, 2011). Pourtant la mise en évidence de lignées pathogènes dans l'espèce s'avère difficile (Coipan *et al.*, 2016). De plus, on connaît mal les déterminants génétiques permettant aux lignées à risque d'échapper aux défenses immunitaires humaines chez *B. garinii* (Kraiczy, 2016). Dans ce contexte, j'ai initié différentes études visant à identifier et caractériser les lignées de *B. garinii* capables d'infecter l'homme. Elles sont réalisées au sein de l'unité avec l'aide de mes collègues David Abrial, Emilie Bard et Aminah Kelliet.

Malgré différents problèmes potentiels liés à des modalités d'échantillonnages hétérogènes entre les cas humains et environnementaux, le ratio des fréquences d'un génotype dans des cas humains et dans les tiques semble être un indicateur intéressant pour évaluer le risque zoonotique. Toutefois, une étude visant à mettre en évidence des différences significatives de ce ratio à l'échelle de génotypes multilocus n'avait pas donné de résultat concluant, malgré des

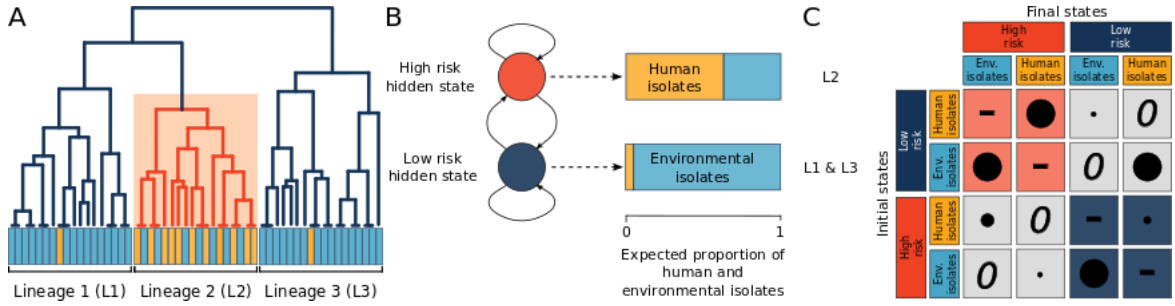

Figure 13: Principe d'utilisation des approches d'arbre de Markov caché pour identifier des lignées de pathogène zoonotique à risque. A) On cherche à identifier des lignées présentant une plus forte fréquence d'isolats issus de cas humain. B) On considère que les lignées vont être caractérisées par un caractère caché associés à différentes probabilité de découverte d'isolats humains et environnementaux. C) Implémentation possible d'une matrice de transition entre états du caractère caché et du caractère observé.

tendances marquées. Au delà d'un potentiel déficit de puissance statistique, des améliorations méthodologiques pourraient être utilisées pour approfondir la question. Dans cette optique, je m'intéresse aux approches phylogénétiques utilisant des arbres de Markov cachés pour aborder ce problème (Bailly, 2017, Figure 13). J'ai pu illustrer le potentiel de ces approches pour discriminer des lignées de *B. garinii* présentant un risque zoonotique plus important. Une étude de la sensibilité de ces approches est nécessaire. Ce travail est actuellement réalisé dans le cadre d'un stage de Master 2 Bioinformatique par Oshma Chakoory.

Les progrès dans la mise en évidence des lignées à risque ouvre la voie à l'étude des mécanismes moléculaires mis en jeux dans l'infection de l'homme par *B. garinii*. Une étude sur cette question a été initiée dans le cadre du stage de Master 2 Bioinformatique d'Audrey Ganteil, en profitant des progrès des approches de génétique d'association appliquées aux bactéries qui visent à prendre en compte la structure génétique des populations étudiées (Collins & Didelot, 2018). Alors que les approches les plus récentes ont donné des résultats convaincants, particulièrement pour la recherche de déterminants d'antibio-résistance, les applications sur des traits de spécificité, de virulence ou d'antigénicité sont plus rares. Différents partenaires européens ont été contactés afin d'aboutir à un échantillonnage permettant d'aborder cette question. Les génomes d'une centaine de souches issues de tiques à l'affût et de cas humains en France, en Allemagne, ou en Slovaquie ont été séquencés. Une phylogénie calibrée à l'aide d'une horloge moléculaire a été produite à partir des chromosomes des différentes souches.

L'étude a été conduite en utilisant comme caractère d'intérêt soit l'origine des isolats, soit le caractère caché inféré à partir des origines des isolats à l'aide d'une approche d'arbre de Markov caché. Afin de contourner les problèmes liés à la complexité de structure des génomes de *Borrelia*, nous avons étudié l'association entre le caractère d'intérêt et les génotypes bactérien en utilisant le contenu en gène et la présence de 10-mers d'acides aminés déduits à partir de l'analyse in-silico des protéomes des souches étudiées.

Les résultats obtenus dans le cadre de cette étude illustrent la possibilité de mettre en évidence par des approches de génétique d'association des gènes candidats potentiellement impliqués dans la capacité à infecter l'homme par des souches de *Borrelia* (Figure 14). Ils pointent un certain nombre de gènes dont l'annotation suggère une influence potentielle sur

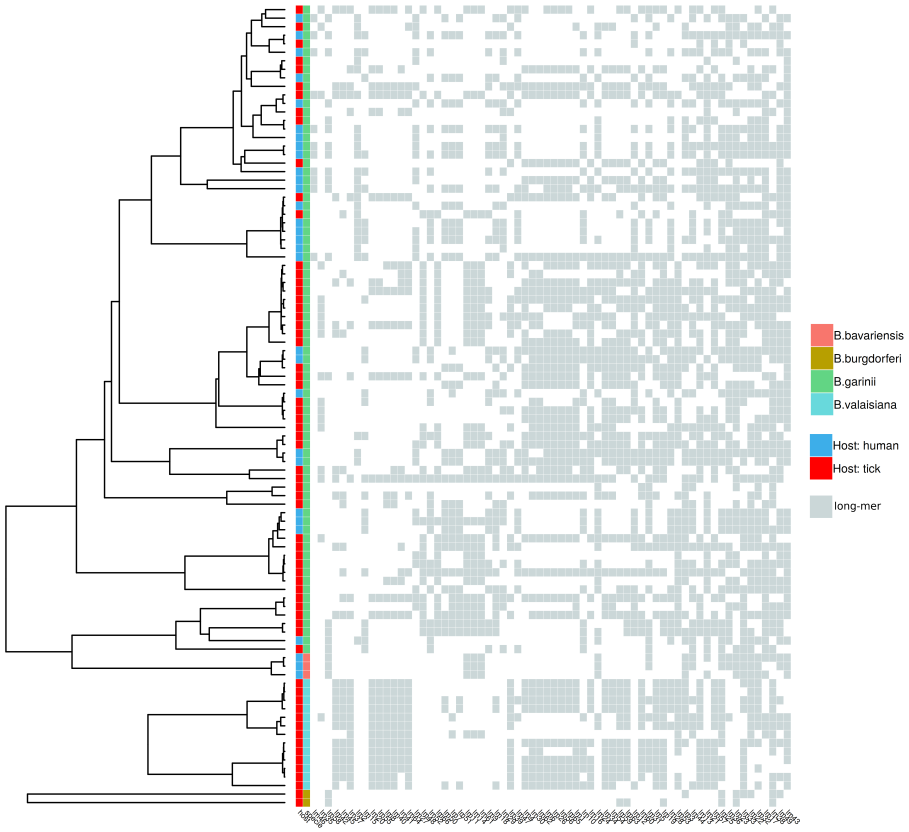

Figure 14: Mise en relation de la distribution des kmers d'acide-aminés associés à un isolement issu de patient humain avec la phylogénie chromosomique des génomes de *Borrelia* séquencés.

la capacité à interagir avec des hôtes vertébrés. L'étude du contexte génétique de ces gènes suggère qu'une large majorité sont localisés sur des plasmides.

Notamment, nous avons mis en évidence un gène de la famille *crasp*, localisé sur un plasmide accessoire de la bactérie. Le représentant identifié n'a pas été testé pour sa capacité à contourner le système du complément (Cordes *et al.*, 2006). Son implication dans la capacité de *B. burgdorferi* d'infecter la souris a toutefois été démontrée (Singh *et al.*, 2017).

D'un point de vue méthodologique, les différents scores utilisés pour tester l'association (Treewas) n'ont pas forcément donné des résultats congruents. Ces scores se différencient par la prise en compte du signal observé sur les branches terminales ou le long de la phylogénie. Différentes explications peuvent être envisagées pour ce résultat. Notamment, les différents scores doivent avoir des sensibilités différentes à la vitesse et aux modalités d'évolution des déterminants du caractère d'intérêt (e.g. recombinaison vs mutation).

En effet, la profondeur au sein de la phylogénie le long de laquelle l'association phenotype / génotype est pertinente doit dépendre de la capacité des gènes responsables du phénotype d'intérêt à être transmis horizontalement. Pour des gènes portés par des plasmides fréquemment gagnés ou perdus, l'association ne pourrait être pertinente qu'au niveau des branches terminales. Les gènes identifiés par les scores qui prennent en compte la profondeur de la phylogénie pourrait constituer des gènes importants mais non essentiel à la capacité à infecter

l'homme.

Ainsi, les résultats des approches d'arbre de Markov caché et d'identification de marqueurs génétiques suggèrent que certaines lignées de *B. garinii* seraient plus susceptibles d'infecter l'homme. Il semble improbable que la capacité à infecter l'homme constitue une source de valeur sélective pour les bactéries étudiées. Les gènes identifiés pourraient notamment offrir aux bactéries qui les portent la capacité d'exploiter d'autres hôtes dans la nature. Certaines souches de *B. garinii* peuvent de fait infecter des mammifères, alors que l'espèce est décrite comme majoritairement associée aux oiseaux.

Il serait important de mieux comprendre quelle est l'importance des déterminants de la capacité à affecter l'homme dans la valeur sélective des souches de *Borrelia garinii*. Si l'impact de ces gènes est important, ils sont susceptibles d'augmenter en fréquence dans l'espèce et conduire à un accroissement du risque zoonotique. Différents outils récemment développés visent à étudier quelles lignées / mutations ont eu un fort succès évolutif (Rasigade *et al.*, 2017; Merker *et al.*, 2018). Ils pourraient être utilisés à terme pour étudier notre question.

Si ces dernières approches offrent des perspectives excitantes, leur interprétation pourrait toutefois s'avérer difficile. Comme mentionné plus haut, le fait de ne pas connaître à coup sur le phénotype associé aux polymorphismes identifiés peut être frustrant. C'est pourtant un des challenges majeurs pour comprendre l'évolution des bactéries, notamment celles d'intérêt agronomique, vétérinaire ou médical de l'échelle des gènes à celle des écosystèmes. L'accumulation de données épidémiologiques, sur les caractéristiques des hôtes infectés, des population susceptibles sous-jacentes et de l'environnement peuvent fournir des métadonnées qui seront à même d'enrichir ce type d'approches.

6 Épidémiologie moléculaire de pathogènes zoonotiques : des patrons de diversité aux cycles de transmission des pathogènes

Les données moléculaires constituent une source d'information importante pour mieux identifier les cycles de transmissions et documenter les dynamiques épidémiologiques des pathogènes. Elles doivent permettre de mieux estimer la distribution ou l'évolution des risques infectieux et le cas échéant, de planifier des mesures permettant de limiter ce risque. Les challenges liée à l'intégration et l'inférence des processus épidémiologiques à partir des données moléculaire constitue un front de science appelé phylodynamique (Grenfell *et al.*, 2004). Ce champ de recherche s'est principalement intéressé à l'évolution de virus ARN compte tenu de la taille de leur génome et de leur dynamique évolutive (Biek *et al.*, 2015). Dans le domaine de l'épidémiologie des maladies infectieuses affectant l'homme, l'approche phylodynamique la plus courante est d'acquérir des données à l'échelle génomique sur des isolats issus de différents individus afin d'affiner les inférences concernant les forces évolutives majeures qui influencent la structure génétique des populations de pathogènes, leur démographie, leur interaction avec les populations d'hôtes, la distribution spatio-temporelle des hôtes sensibles.

Ces approches peuvent être appliquées aux pathogènes circulant dans les populations d'animaux sauvages ou domestiques. Ceci ouvre des perspectives dans le domaine de recherche de notre unité. Toutefois, la dynamique des pathogènes dans les populations animales diffèrent sensiblement de ce qui est maintenant généralement rencontré chez l'homme dans les pays du Nord. Les animaux peuvent-être exposés de manière plus fréquentes aux pathogènes. Un certain nombre d'informations permettant de comprendre la circulation des pathogènes pourraient être obtenues en caractérisant la diversité intra et inter-spécifique des microbes au sein des hôtes infectés, notamment pour les maladies transmises par les tiques (Diuk-Wasser *et al.*, 2016).

Au cours des dernières années, je me suis investi dans des projets visant à caractériser la diversité intra-spécifique de pathogènes à l'échelle nationale en prenant en compte la diversité des espèces hôtes afin de mieux comprendre leur cycle de transmission. Compte tenu de différents résultats, je me suis engagé dans des travaux visant à prendre en compte la diversité des pathogènes au sein d'hôtes et/ou de vecteurs infectés à des échelles spatiales plus fines. Enfin, je me suis intéressé au potentiel des études de microbiotes pour caractériser des risques sanitaires.

6.1 Épidémiologie des pathogènes et diversité observable à l'aide de données de surveillance

Un nombre important de pathogènes animaux font l'objet d'analyses de détection dans le cadre de suivis vétérinaires conventionnels et / ou de protocoles de surveillance exploratoires ou réglementaires. L'analyse de ces données se limite souvent à l'identification des cas d'infection avérés et au calcul des prévalences observées pour les données réglementaires.

Ces analyses constituent toutefois une source d'information extrêmement riche pour aller plus loin et affiner nos connaissances sur la circulation des dangers sanitaires étudiés. Notamment, la caractérisation de génotypes différents au sein de différentes espèces hôtes peut

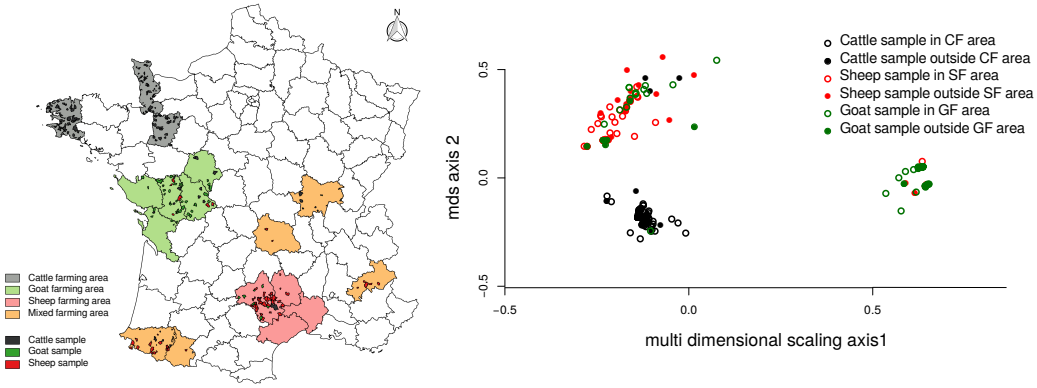

Figure 15: Échantillonnage de génotypes de *Coxiella burnetii* issus de cas cliniques et projection de la diversité phylogénétique observée.

amener à reconsidérer le risque d'infection déduit de données de cas. Dans cette optique, j'ai participé à des études sur *Coxiella burnetii* et *Anaplasma phagocytophylum*.

6.1.1 Diversité de souches de *Coxiella burnetii* issues de cas d'avortement chez différentes espèces d'élevage

La fièvre Q est une zoonose mondiale causée par la bactérie *Coxiella burnetii*. Chez les ruminants domestiques, les principales manifestations cliniques de la fièvre Q sont les avortements. Compte tenu de sa prévalence importante et du risque sanitaire qu'elle représente, la fièvre Q fait l'objet d'une surveillance animée dans le cadre de la plateforme nationale d'épidémiosurveillance en santé animale (ESA).

Alors que la fièvre Q est enzootique en France et qu'elle touche de nombreux hôtes d'intérêt, la diversité génétique de *C. burnetii* n'avait pas été étudiée de manière systématique. Dans la cadre de travaux initiés par Elsa Jourdain et réalisés dans le cadre de la thèse d'Aurélien Joulié dirigée par Agnès Leblond, j'ai participé à une étude visant à identifier les associations potentielles entre les génotypes de *C. burnetii* et différentes espèces de ruminants (Joulié *et al.*, 2017). Pour cela, la répartition des génotypes de *C. burnetii* a été étudiée dans les principales régions d'élevage de France, qui sont caractérisées par une forte spécialisation dans les espèces élevées.

Nous avons utilisé des échantillons d'ADN prélevés entre 2006 et 2015 sur 301 femelles de ruminants (160 vaches, 76 brebis, 65 chèvres) ayant subi un avortement assigné à un épisode de Fièvre Q dans sept régions agricoles françaises. La diversité de *C. burnetii* au sein de ces échantillons a été caractérisée à l'aide de 17 marqueurs mini-satellites (MLVA).

Une approche phylogénétique a permis d'identifier trois principaux groupes de génotypes au sein de l'échantillon. Ces groupes étaient significativement associés aux espèces de ruminants : presque tous les génotypes de bovins étaient regroupés au sein d'un clade spécifique. Les génotypes bactériens issus d'ovins et de caprins étaient regroupés au sein de deux autres groupes (Figure 15).

Au delà du fort effet hôte observé, nous avons mis en évidence que la diversité des souches

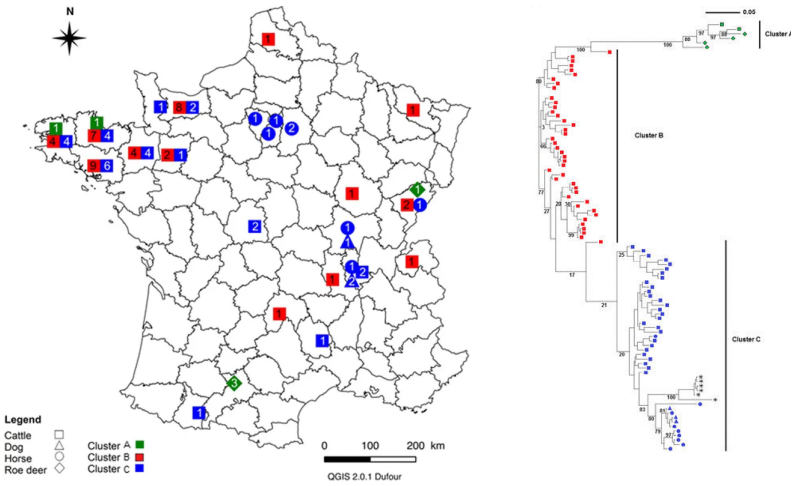

Figure 16: Échantillonnage de génotypes d'*Anaplasma phagocytophilum* issus de données de cas clinique et analyse phylogénétique des génotypes multilocus observés.

de *C. burnetii* retrouvé sur les prélèvement dépendaient également de l'interaction entre l'espèce hôte et la région d'origine. Ces observations pourraient être liées à des "spill-overs", les infections à l'origine de ce patron étant liées à la forte circulation de génotypes bactériens aspécifiques générant des infections atypiques.

Les résultats obtenus montrent que les modèles décrivant l'épidémiologie, le risque de transmission de *C. burnetii* pourraient être enrichis par les données de typage moléculaire. La caractérisation fine des processus aboutissant à des spill-overs revêt à mon sens une importance tant pour comprendre l'apparition de pathologie chez les animaux d'élevage que pour envisager plus efficacement comment se construit le risque zoonotique.

6.1.2 Diversité des populations d'*Anaplasma phagocytophilum* chez les bovins et d'autres espèces de vertébrés

Une étude similaire a été conduite sur *A. phagocytophilum*, une bactérie transmise par les tiques qui affecte un large éventail d'animaux sauvages et domestiques. Cet agent pathogène est potentiellement zoonotique et la déclaration des infections humaines est obligatoire aux États-Unis. Ces travaux ont été initiés avec Gwenaël Vourc'h et Agnès Leblond dans le cadre de la thèse d'Amélie Chastagner (Chastagner *et al.*, 2014). Ils ont été conduits en partenariat avec le laboratoire d'analyse vétérinaire Labeo et l'unité Biologie Moléculaire et Immunologie Parasitaire qui s'intéressent à ce pathogène.

Nous avons caractérisé la diversité génétique d'*A. phagocytophilum* dans plus de cent bovins français identifiés comme infectés par le pathogène suite à une suspicion d'anaplasmose. Nous avons ensuite comparé les génotypes observés avec ceux trouvés chez treize chevaux, trois chiens et quarante chevreuils infectés et des données de référence pour déterminer si les génotypes d'*A. phagocytophilum* qu'ils portaient étaient partagés. Dans ce but, nous avons utilisé une analyse de séquences multilocus.

L'analyse phylogénique a révélé trois groupes génétiques bactériens chez les animaux étudiés (Figure 16). Les deux groupes principaux comprenaient 98% des génotypes bactériens trouvés chez les bovins. Le premier comprenait uniquement des génotypes issus de nos échantillons de bovins, tandis que le second contenait des génotypes de nos échantillons de bovins, de chevaux et de chiens. Ces deux groupes semblaient largement différenciés par rapport aux génotypes portés par les chevreuils. En effet, le troisième groupe observé contenait tous les génotypes de chevreuils et seulement trois génotypes de bovins.

Cette dernière observation est importante dans la mesure où le chevreuil est souvent considéré comme une source potentielle du risque infectieux pour les bovins. En effet, le cycle de transmission des génotypes associés aux deux espèces hôtes semblent dissociés. Les données observées interrogent également sur les conditions permettant le maintien de l'importante diversité d'*A. phagocytophilum* et les conséquences sanitaires associées. Ces questions et l'identification de traces potentielles de co-infection au sein de notre jeu de données appelaient le développement d'études spécifiques.

6.2 Des données de co-infection aux hypothèses épidemiologiques

La co-infection par différents génotypes d'un agent pathogène peut, dans certains systèmes biologiques, être commune et avoir des origines multiples (Sofonea *et al.*, 2017). Les dynamiques des différents génotypes au sein des hôtes infectés sont susceptibles d'avoir de multiples conséquences : i) d'un point de vue proximal, elles peuvent avoir des conséquences directes sur l'épidémiologie du pathogène à l'échelle des populations hôtes, dans la mesure ou la dynamique de diffusion globale du pathogène dépend de la dynamique de l'agent pathogène au sein des individus infectés (Mideo *et al.*, 2008) ; ii) Les interactions entre les genotypes qui se retrouvent co-infectant les mêmes individus peuvent déboucher sur l'évolution de traits critiques en terme de gestion sanitaire comme la virulence (Alizon *et al.*, 2013). L'analyse des patrons de séquençage haut-débit peut fournir un point de vue nouveau sur sur les patrons de co-infection et offrir de nouvelles perspectives d'analyses (Alizon *et al.*, 2011).

Compte tenu des niveaux d'expositions rencontrés, les résultats de modèles simplifiés visant à objectiver les dynamiques épidémiologiques, de traces moléculaires identifiées, et des perspectives potentielles, je me suis intéressé à la diversité génétique d'*A. phagocytophilum* et du complexe d'espèce *B. burgdorferi* au sein d'hôtes et de vecteurs infectés. Ces problématiques ont été développées en lien avec le projet ANR OSCAR porté par Olivier Plantard.

6.2.1 Mesure de l'impact des chevreuils sur l'infection des tiques par *Anaplasma phagocytophilum*

Après avoir cherché à documenter la diversité d'*A. phagocytophilum* à grande échelle chez les bovins et plus sporadiquement chez d'autres espèces comme les chevreuil, la thèse d'Amélie Chastagner s'est orientée vers la caractérisation de patrons de co-infection au sein d'hôtes et de vecteurs. On peut constater dans la littérature que la bactérie peut atteindre des prévalences importantes dans ces réservoirs naturels, e.g. souvent plus de 80% chez le chevreuil, même si elle est souvent rare chez les nymphes de la tique considérée comme son vecteur principal, <5% chez les nymphes d'*Ixodes ricinus* (Stuen *et al.*, 2013). Cette situation suggère

que les co-infections sont nombreuses au moins chez les hôtes.

Dans ce cadre, une étude de caractérisation par séquençage haut-débit des génotypes d'*Anaplasma phagocytophilum* infectant des chevreuils et nymphes d'*Ixodes ricinus* échantillonnées à l'affût a été menée à l'échelle d'un paysage agricole fragmenté localisé en Haute-Garonne, en collaboration avec différents partenaires dont Hélène Verheyden de l'unité Comportement et Écologie de la Faune Sauvage.

L'objectif de ce travail était double : i) Confronter la prévalence d'*A. phagocytophilum* observée dans les tiques et dans les chevreuil à l'aide d'un modèle exploratoire afin de comprendre comment elles peuvent être compatibles avec la circulation du pathogène tout en étant si différentes ; ii) définir si les chevreuils représentent une source importante de tiques infectées par *A. phagocytophilum*.

Les données acquises indiquent des prévalences d'*Anaplasma phagocytophilum* chez les chevreuils et les nymphes d'*Ixodes ricinus* en phase avec la littérature, 75% et 3% respectivement (Chastagner *et al.*, 2017). Les données de séquence permettent de différencier nettement les population bactériennes associées au chevreuil et celles généralement présentes dans les nymphes (Figure 17). En effet, la majorité des génotypes d'*Anaplasma phagocytophilum* retouvés dans les tiques sont apparentés à ceux précédemment identifiés chez les bovins. Seuls 23% des nymphes positives analysées portent des génotypes de chevreuil. Le risque d'infection pour le bétail et l'homme serait plus important que ce qui est actuellement suspecté.

Un modèle assumant la prévalence observée des génotypes d'*A. phagocytophilum* associés aux chevreuil dans les tiques et la charge en tiques moyenne observée pour le chevreuil indiquent que moyennant une efficacité suffisante de la transmission d'*A. phagocytophilum* de la tique vers le chevreuil, les prévalences observées chez le chevreuil peuvent être atteintes si des super-infections sont possibles quelque soit le statut infectieux de l'hôte. De fait, des traces de co-infections sont identifiées sur les trois loci bactériens caractérisés par séquençage haut-débit dans le cadre de cet étude.

Ces résultats illustrent qu'il serait intéressant de développer un modèle permettant de prédire la diversité attendue aux sein des hôtes infectés. Ce type de modèle devrait offrir des informations complémentaires sur la dynamique épidémiologique d'*A. phagocytophilum* à différentes échelles et sur le niveau de compétition au sein des hôtes entre lignées bactériennes.

L'étude conduite illustre de nombreux intérêts potentiels de l'étude de la dynamique intra-hôte des génotypes pathogènes à des fins épidémiologiques. J'imagine que ces dynamiques intra-hôte sont susceptibles de mener à des spécialisations fonctionnelles spécifiques aux différentes espèces hôtes, qui sont caractérisées par différentes durées de vie, dynamiques de cellules hôtes, niveaux d'exposition aux génotypes de pathogènes. Ces spécialisation pourraient avoir un impact sur la circulation des génotypes pathogènes au sein des communautés d'hôtes.

Plus pragmatiquement, cette étude illustre l'intérêt de ces approches pour mieux décrire les patrons de circulation des génotypes pathogènes entre les espèces hôtes disponibles. Toutefois, les effectifs de vecteurs disponible et l'utilisation d'un hôte unique limitait la portée de ce travail.

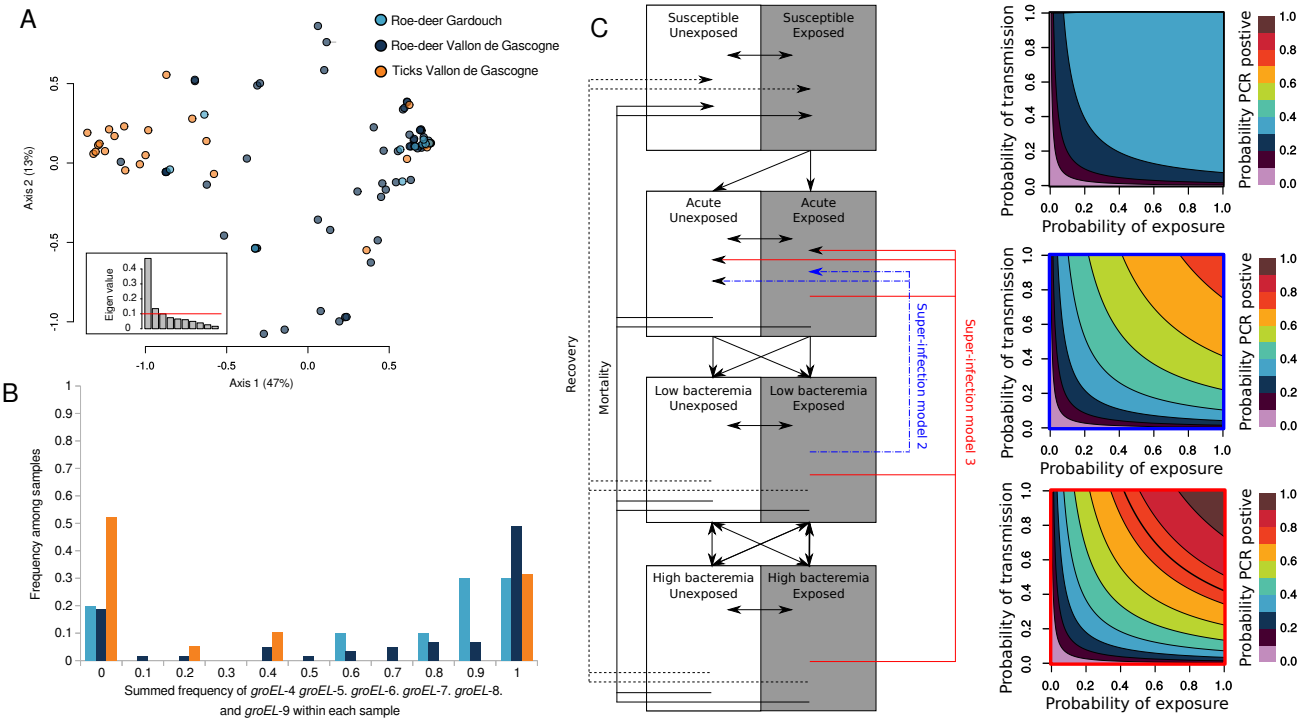

Figure 17: Analyse de patrons de co-infection par *Anaplasma phagocytophilum* chez des tiques vectrices et des chevreuil. A) projection de la diversité génotypes d'*A. phagocytophilum* portés par les échantillons étudiés. B) Illustration des patrons de co-infections observés au locus *groEL*. C) Structure et propriétés de modèles visant à détermeiner la prévalence d'A. phagocytophilum dans les populations de cherveuil en fonction des données observées et différentes hypothèses sur les modalités de super-infection.

6.2.2 Identification de cycles de transmission de *Borrelia burgdorferi* en Forêt de Sénart

Ces limites étaient contournables grâce à des échantillons collectés dans le cadre de projets initiés par Gwenaël Vourc'h pour des études concernant la circulation de *B. burgdorferi*. Dans le cadre de collaborations impliquant Jean-Louis Chapuis du Museum National d'Histoire Naturelle et la composante du CNR Borrelia de l'Institut Pasteur, qui avaient conduit à la thèse de Maude Marsot et de Séverine Bord, nous disposions dans l'unité d'un échantillonnage conséquent de micro-mammifères et de tiques échantillonnées à l'affût en sympatrie dans la forêt de Sénart.

De manière analogue avec le travail réalisé sur *A. phagocytophylum*, une modélisation du niveau de diversité circulant dans les tiques de deux localités par calcul bayésien approximé nous avait suggéré que de forts niveaux de co-infection par différents génotypes de *B. burgdorferi* étant anticipés sous hypothèse neutre (Jacquot *et al.*, 2014b).

De fait, de nombreuses données supportent l'hypothèse de nombreuses co-infections chez les réservoirs de pathogènes transmis par les tiques. Or, la méthode la plus utilisée pour effectuer leur suivi moléculaire est basée sur un séquençage Sanger de produits PCR obtenus à partir d'ADN issus d'hôtes ou de tiques prélevées à l'affût. Cette méthode ne permet pas d'aborder le problème des co-infections de manière précise.

Dans une perspective d'ingénierie écologique visant à réduire l'impact des réservoirs majeurs des agents de la maladie de Lyme, il m'a semblé pertinent d'étudier s'il était possible de différencier finement les cycles de transmissions associés à différentes espèces hôtes (Brisson *et al.*, 2012). La notion de spectre d'hôte au sein du complexe d'espèce *B. burgdorferi* est en effet souvent présentée comme un élément central pour comprendre la distribution de la diversité au sein du complexe d'espèce *Borrelia burgdorferi* (Margos *et al.*, 2011). Toutefois ce spectre d'hôte est souvent décrit de manière assez générale, via une association des espèces du complexe à des grandes classes d'hôtes : mammifères, oiseaux, reptiles.

Nous avons utilisé une approche haut-débit, permettant le séquençage indépendant d'un grande nombre de produits PCR, afin de caractériser les populations du complexe d'espèce *B. burgdorferi* infectant 125 tamias (*Tamias sibiricus*), 93 campagnols (*Myodes glareolus*), and 10 mulots (*Apodemus sylvaticus*), ainsi que 478 nymphes d'*Ixodes ricinus* (Jacquot *et al.*, 2014a, 2016).

Compte tenu des patrons de diversité génomique observés au préalable, nous avons souhaité caractériser les souches sur un marqueur localisé dans un gène de ménage ,*rlpB*, et un marqueur localisé dans le gène *ospC*. Ces deux marqueurs se différencient par leur niveau de polymorphisme, leur contexte génomique et leurs structures génétiques. Il était donc intéressant de voir si les informations apportées peuvent être complémentaires dans le cadre d'une étude visant à définir la contribution de différentes espèces hôtes à la circulation de *B. burgdorferi*.

Les résultats obtenus montrent une forte différence de composition entre les populations infectants les tamias et les campagnols quelque soit le marqueur envisagé (Figure 18). La diversité observée permet notamment de différencier les souches infectant les deux hôtes même pour le marqueur chromosomique *rlpB*, au sein de *Borrelia afzelii*. Les données issues de tiques à l'affût illustrent une diversité bien plus importante. Un modèle de mélange résolu par ABC permet de déterminer que presque la moitié des nymphes infectées analysées se sont gorgées sur d'autres hôtes que ceux envisagés. Il est notamment probable que les oiseaux représentent une source importante de tique infectées compte tenu de la fréquence de génotypes de *Borrelia garinii* et *Borrelia valaisianna*.

D'autre part, ce modèle indique que les réservoirs potentiellement pluri-spécifiques qui incluent respectivement les tamia et les campagnols contribuent de manière extrêmement déséquilibrée à la densité de nymphes infectées, et donc au risque d'exposition pour l'homme. De fait, les deux génotypes du complexe que nous avons pu caractériser à partir de souches isolées de cas humains issus de la forêt de Sénart correspondent à des génotypes rencontrés chez des tamia et potentiellement associés aux oiseaux.

Depuis, différentes études ont également étudié la diversité des souches de *Borrelia burgdorferi* s.l. par des approches permettant de prendre en compte la diversité intra échantillon (Durand *et al.*, 2017a,b; Strandh & Råberg, 2015; Råberg *et al.*, 2017; Walter *et al.*, 2016). Les objectifs poursuivis étaient souvent différents, mais toutes ces études ont pointé une forte diversité intra-échantillon et une faible évolution temporelle de la fréquence de génotypes rencontrés localement. Au sein de l'université de Neuchâtel, les paramètres de transmission de ces différents variants de *Borrelia afzelii* ont été étudiés chez la souris, illustrant une re-

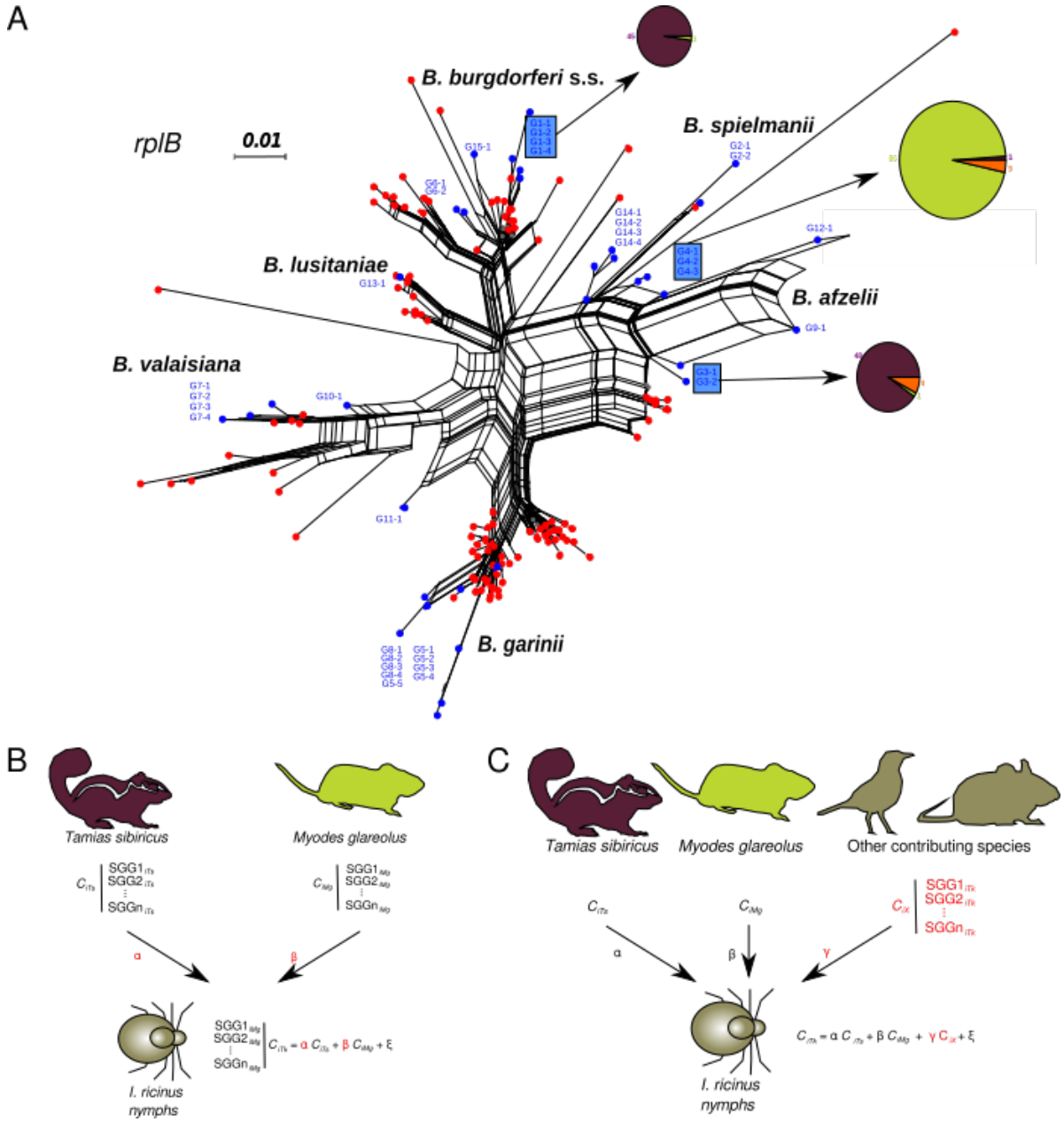

Figure 18: Contribution de différentes espèces hôtes à la diversité des génotypes du complexe *Borrelia burgdorferii* circulant dans les nymphes d'*Ixodes ricinus* en forêt de Sénart. A) Réseau phylogénétique représentant les relations de parenté entre les génotypes observés pour le marqueur *rlpB*. B) Première phase du modèle de mélange visant à estimer la contribution des hôtes échantillonnés à la diversité des génotypes circulant au sein des tiques vectrices. C) Seconde phase du modèle cherchant à identifier la contribution d'hôtes non-échantillonnés.

lation significative, malgré une certaine variance résiduelle, entre fréquence observée et R0 calculé à partir des paramètres mesurés (Durand *et al.*, 2017b). Ceci suggérerait, au contraire des résultats que nous avons observés une influence relativement faible de l'effet hôte sur la structure des populations (Estrada-Peña & de la Fuente, 2017). Des études mécanistes (Hart *et al.*, 2018) et la comparaison des patrons de diversité rencontrés entre différentes localités

(Mechai *et al.*, 2016) devrait permettre d'avoir une meilleure vision de la spécificité et de la stabilité des associations entre hôtes et *Borrelia burgdorferi*.

Les études de co-infection conduites sur *A. phagocytophilum* et *B. burgdorferi* m'ont amené à m'intéresser plus activement à la diversité du microbiote d'*Ixodes ricinus* : i) dans une perspective de systématisation des études de diversité pour différents pathogènes utilisant les mêmes vecteurs et/ou hôtes ; ii) pour affiner les analyses permettant de détecter les associations entre génotypes au sein de jeux de données métagénomiques.

6.3 Identification d'associations dans les microbiotes, une perspective en santé animale

Chaque eucaryote constitue un ensemble de niches écologiques colonisables par des bactéries plus où moins favorables. Les approches métagénomiques ouvrent des perspectives intéressantes concernant la caractérisation des communautés bactériennes qui colonisent ces niches écologiques.

Alors que la majorité des études en épidémiologie des maladies infectieuses se focalisent sur la circulation d'un pathogène donné dans son système biologique, la métagénomique offre la possibilité technique d'obtenir des réponses concernant un panel plus large de génotypes ou de taxons qui sont susceptibles d'interagir. Elle ouvre la perspective de faire le lien entre la structure générale des microbiotes en fonction de facteurs structurants et la dynamique de taxons d'intérêt, via l'identification des associations microbiennes les plus pertinentes. Elles permettent ainsi d'envisager d'une manière plus générale les problématiques épidémiologiques.

Ainsi, je me suis investi sur des problématiques de métagénomique. D'un point de vue méthodologique, les questions posées par les travaux de diversité intra-hôte et l'ambition associée à ces études m'ont amené à m'investir sur l'étude des approches visant à détecter des associations entre microbes dans les jeux de données métagénomiques. Au niveau biologique, je me suis focalisé sur la diversité du microbiotes de la tique *Ixodes ricinus*.

6.3.1 Étude des associations entre microbes au sein de données métagénomiques

Les réseaux d'associations constituent une approche de première intention pour identifier grâce aux patrons d'association statistique de potentielles interactions entre les unités taxonomiques opérationnelles (OTUs) identifiées par des approches métagénomiques. Si ces approches connaissent un fort succès, elles reposent sur une diversité de pratiques bio-informatiques et d'approches statistiques dont les limites restent floues. J'ai eu l'opportunité de participer à l'encadrement, avec Gwenaël Vourc'h et Patrick Gasqui, de la thèse d'Arnaud Cougoul sur ce sujet. Arnaud Cougoul s'est intéressé au potentiel et limites de ces approches avant de proposer de nouvelles solutions pour conduire ce type d'analyse et d'appliquer ces solutions.

Dans un premier temps, Arnaud Cougoul s'est intéressé aux difficultés statistiques associées au traitement des données métagénomiques dans le cadre de recherche d'associations (Cougoul *et al.*, 2019a). En effet, les microbiotes contiennent généralement des centaines voir des milliers d'OTUs qui sont pour la plupart rares. Cette structure de communauté peut entraîner des difficultés méthodologiques: des simulations ont montré que les méthodes de

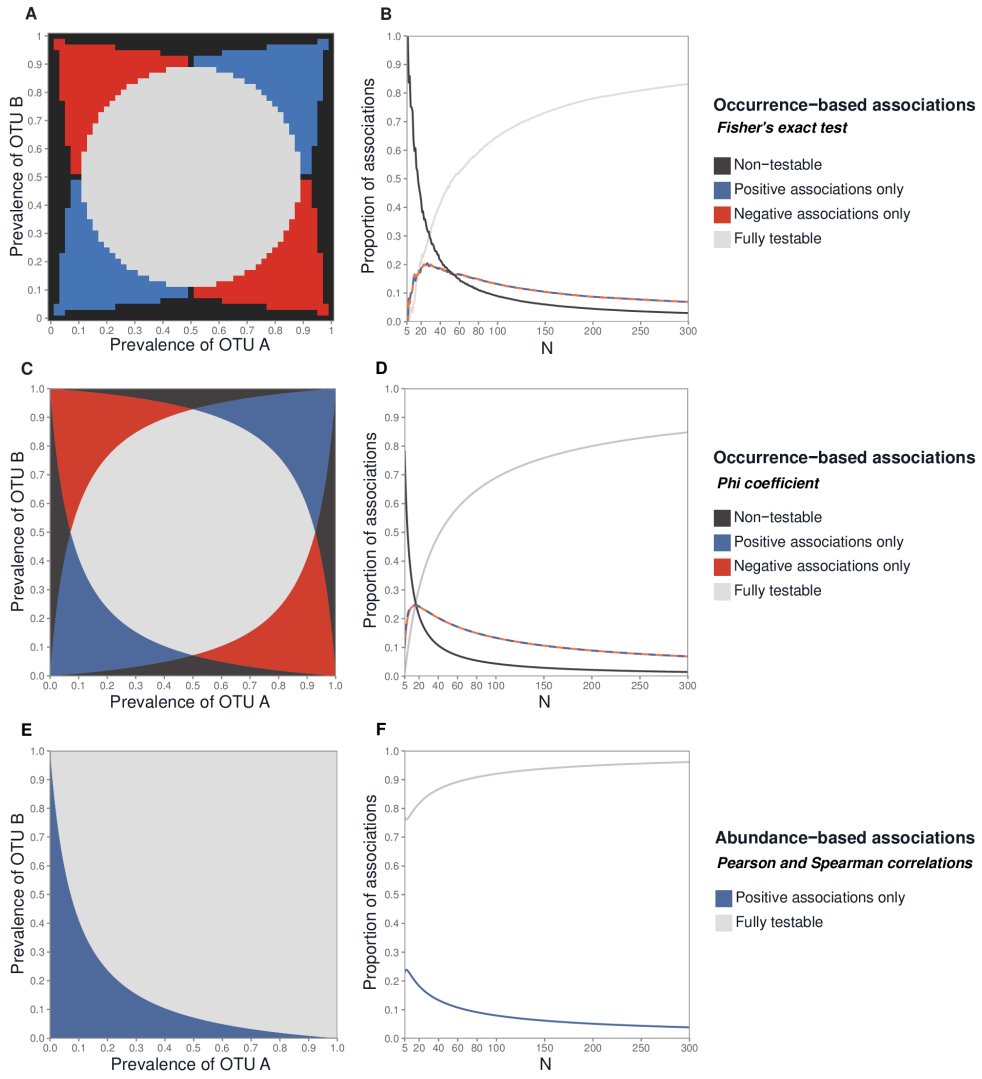

Figure 19: Impact de la prévalence de deux OTUs pour tester la corrélation de leur abondance à partir de données métagénomiques. A, C, E) Identifications des couples de prévalence impliquant des tests problématiques. B, D, F) Évolution de la proportion de tests affectés en fonction de l'effectif étudié.

détection des associations par paires entre OTU, qui reflètent vraisemblablement les interactions, donnent des résultats problématiques (Weiss *et al.*, 2016). Les performances des outils de détection d'association sont altérées en cas de forte proportion de zéros dans les tables OTU. Cette même étude a mis en évidence un différentiel de sensibilité pour la détection d'association négatives comparativement aux associations positives.

Notre objectif était de comprendre l'impact de la rareté des OTUs et des caractéristiques des associations sur leur détection. Nous avons exploré par des approches mathématiques et des approches de simulations la performance de statistiques communes pour tester les associations, la sensibilité des mesures alternatives d'association, et les performances des outils d'inférence de réseau couramment utilisés dans la littérature. Ces résultats ont été replacés dans la perspective de la recherche d'associations microbiennes dans les jeux de données métagénomiques.

Arnaud Cougoul a pu définir des fonctions décrivant à partir de quelles limites les pré-

valences de deux OTUs pour lesquels on cherche a étudier une association ne permettaient plus de conduire un test significatif (Figure 19). Nous avons défini à partir de ce résultat la proportion d'associations par paire qui peuvent être testées efficacement. Ce résultat permet de définir plus clairement les associations dont on ne peut pas attendre de résultat pertinent, par rapport à celle où l'absence d'association ne peut être rejetée, ou une association positive ou négative peut-être validée.

D'un point de vue quantitatif, nos résultats indiquent qu'une grande proportion d'associations par paires, en particulier les associations négatives, ne peuvent pas être testées de manière fiable au sein d'un ensemble de microbiotes. Cette contrainte pourrait par exemple entraver l'identification d'agents de lutte biologique candidats qui pourraient être utilisés pour contrôler des agents pathogènes rares.

Nous avons enfin vérifié que l'identification des associations testables pourrait servir de méthode objective pour filtrer les ensembles de données, en remplacement des approches empiriques actuelles. Nous avons vérifié que cette stratégie réduisait le temps de calcul nécessaire pour déduire les réseaux d'association à l'aide des méthodes couramment utilisées et donne des résultats plus robustes comparativement aux filtres classiques.

Ces résultats et leurs implications ont permis de définir différentes pistes d'amélioration concernant l'analyse des associations entre OTUs au sein de microbiotes. En prenant en compte ces pistes, Arnaud a produit un paquet R appelé R-MAGMA dédié à ces analyses (Cougoul *et al.*, 2019b).

La base de cette méthode est d'utiliser un modèle graphique gaussien pour approcher des corrélations partielles parmi un large nombre de variables. Les corrélations sont évaluées pour chaque paires de variables en faisant abstraction des effets indirects liés aux autres variables. L'inférence des corrélations se fait par une approche lasso qui influence de manière conséquentes l'ajustement du modèle aux données observées.

Le modèle graphique gaussien faisant l'hypothèse que les variables utilisées ont une distribution gaussienne, les données de comptage de séquences par OTU doivent être normalisées. Pour chaque OTU, les données de comptage sont d'abord modélisées pour rendre en compte certaines spécificités des données métagénomiques dont : i) un excès de dénombrements nuls liés à l'absence de l'OTU et une sur-dispersion des dénombrements en cas de présence grâce à une loi ZI-BN ; ii) l'effet d'une potentielle covariable ; iii) la compositionnalité des données métagénomique en utilisant un offset basé sur la moyenne génométrique de ratios de comptage par paire.

La méthode implémentée utilise ensuite un modèle de copules gaussiennes afin d'obtenir une distribution des données compatibles avec modèle graphique gaussien. Arnaud a utilisé son modèle pour analyser différents jeux de données métagénomiques.

6.3.2 Étude du microbiote d'*Ixodes ricinus* en fonction des stades de développement

Arnaud a notamment participé à l'analyse d'un jeu de données visant à étudier l'évolution du microbiote de la tique *Ixodes ricinus* en fonction du stade de développement. Les tiques sont les vecteurs de nombreuses bactéries pathogènes et de nombreux taxons apparentés potentiellement mutualistes ou dont la niche écologique reste méconnue. Étudier la diversité

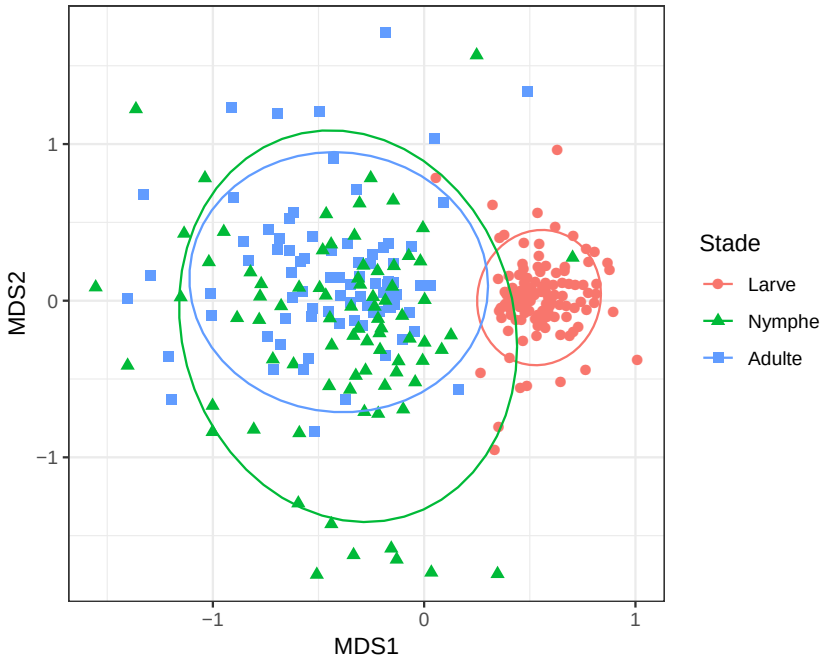

Figure 20: Projection de la diversité des microbiotes portés par les individus étudiés à différents stades de développement.

des communautés bactériennes dans les tiques est donc susceptible de fournir des informations intéressantes quant à la dynamique des populations de pathogènes potentiels et à leur niche écologique. Elle donne aussi l'occasion d'explorer de potentielles interactions entre microbes aboutissant à une modification de l'épidémiologie des pathogènes. Elle permet enfin d'explorer les interactions biotiques qui influencent la dynamique et les traits d'histoire de vie des tiques. Dans le cadre de ces études, je me suis intéressé à la structure des communautés bactériennes retrouvées dans différents stades de développement de la tique *Ixodes ricinus*. Ainsi les microbiotes de 150 larves, 150 nymphes et 150 adultes d'*Ixodes ricinus* échantillonnés dans la forêt de Sénart ont été caractérisés par séquençage haut-débit d'une région hyper variable de l'ARNr 16S.

Bien que d'une taille relativement modeste et basé sur une technologie de séquençage maintenant ancienne (454), le projet engagé illustre différentes questions d'actualité concernant l'analyse des données de microbiote.

En terme de système biologique, les données obtenues permettent élargir le point de vue sur un pathosystème. Il existe une forte différence entre les microbiotes de tiques au stade larvaire et celles aux stades nymphal ou adulte, qui ont déjà pris au moins un repas sanguin (Figure 20). Ces données sont en phase avec les résultats obtenus pour d'autres espèces du genre *Ixodes* (Kwan *et al.*, 2017; Ross *et al.*, 2018).

Deux modules ont retenu notre attention dans le réseau inféré avec R-MAGMA (Figure 21). Un module de grande taille caractérisé par des associations positives pourrait constituer un "core-microbiote" d'*Ixodes ricinus* dans la forêt de Sénart. Ce module est lié via une association négative à un second module. Ce second module implique un symbiote de la tique à transmission verticale, *Spiroplasma ixodetis*, et des taxons des genres *Arsenophonus* et *Wol-*

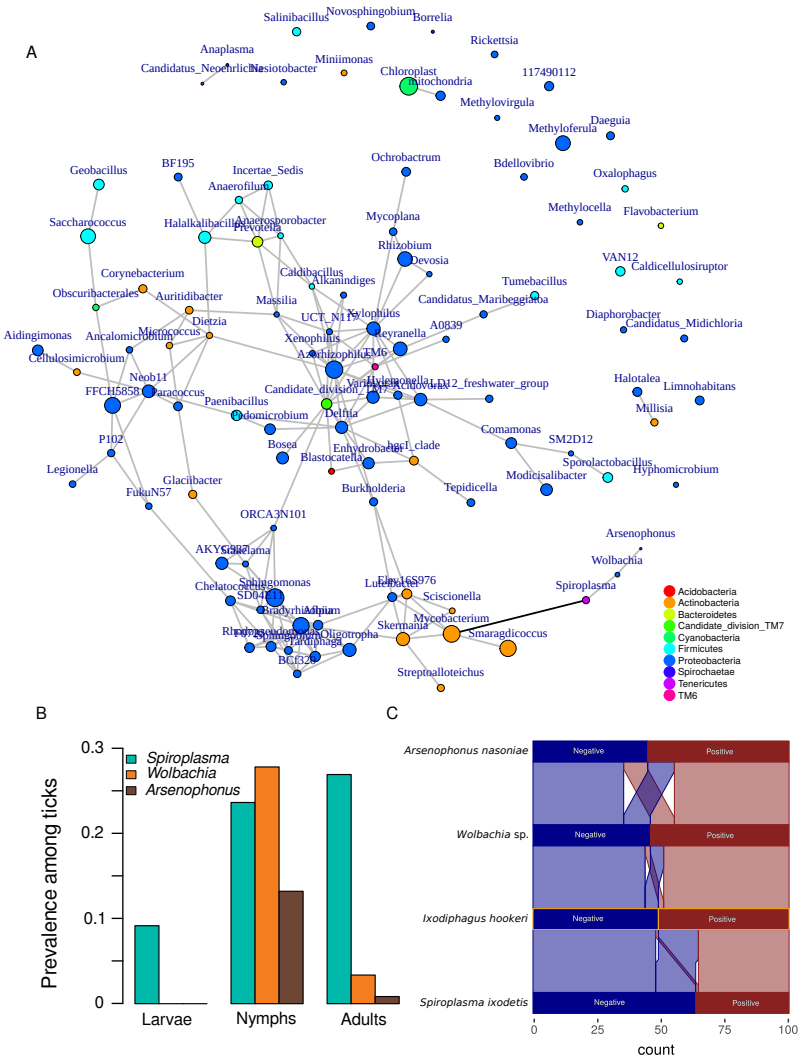

Figure 21: Analyse des associations entre taxons microbiens inclus dans les microbiotes de la tique *Ixodes ricinus*. A) Réseau d'association inféré à l'aide de l'approche R-MAGMA prennant en compte un effet stade. Les associations positives sont figurées par des liens gris, les associations négatives sont figurées par des liens noirs. B) Prévalence de *Spiroplasma*, *Arsenophonus* et *Wolbachia* obtenues à partir de données métagénomiques pour différents stades de développement pour les tiques étudiées. C) Confirmation par qPCR de la forte association entre les présences / absences du parasitoïde *Ixodiphagus hookeri*, de ses symbiotes des genres *Arsenophonus* et *Wolbachia* et du symbiote de la tique Spiroplasma ixodetis.

bachia, portés à priori par le parasitoïde *Ixodiphagus hookeri* qui infeste les nymphes d'*Ixodes ricinus*. Des approches de PCR quantitiative nous ont permis de confirmer la forte association entre le parasitoïde et ces trois taxons bactériens au sein des nymphes d'*Ixodes ricunus*.

Cette étude illustre l'intérêt de prendre en compte les associations microbiennes pour identifier des facteurs susceptibles d'influencer l'épidémiologie de pathogènes animaux et zoonotiques.

7 Projet de recherche

Je souhaite inscrire mes futurs travaux de recherche dans la continuité des travaux déjà engagées. Je souhaite notamment continuer faire évoluer mon activité vers des projets plus pertinents d'un point de vue épidémiologique, potentiellement plus utiles aux porteurs d'enjeux. Pour cela, j'espère orienter mes recherches dans trois directions : i) Développer l'utilisation des approches phylodynamiques pour les pathogènes animaux et zoonotiques ; ii) Comprendre les processus adaptatifs des agents pathogènes à leur environnement ; iii) Étudier les liens entre la dynamique épidémiologique des agents pathogènes au sein et entre leurs hôtes. Ci-dessous, je décris succinctement mes motivations pour m'engager dans chacun de ces axes et les projets qui peuvent être envisagés à court terme. A plus long terme, j'espère concilier le développement de recherche sur ces trois problématiques via des projets plus intégrés.

Ces travaux seront réalisés dans le cadre du thème Épidémiologie, Écologie et Évolution des agents pathogènes animaux et zoonotiques de l'unité EPIA, en interaction avec les autres problématiques de travail que nous développons. En effet, nous souhaitons étudier au sein de ce thème comment l'utilisation de données d'épidémiologie moléculaire à grande échelle peuvent contribuer à l'étude et à la gestion des maladies animales et zoonotiques à fort impact.

Compte-tenu de mes fonctions de directeur d'unité, je devrai faire des choix sur les projets scientifiques à réaliser. J'espère que l'organisation de notre unité et le recrutement de jeunes scientifiques me permettra de continuer à participer au développement de projets scientifiques dans l'unité, en tant que meneur ou en tant que collaborateur.

7.1 Développer l'utilisation des approches phylodynamiques pour les pathogènes animaux et zoonotiques

Les attentes vis-à-vis des recherches épidémiologiques sur ces pathogènes sont de plus en plus fortes alors que les situations à évaluer sont de plus en plus complexes. Ces attentes sont notamment liées à un impact socio-économique important dans les filières d'élevage, à un besoin de rationalisation des approches de lutte, et à une demande d'information de la part du public impacté par les pathogènes. Les modifications du climat, de la gestion des populations animales, des modalités de gestion sanitaire ou de la structure des contacts entre les populations humaines et les animaux modifient la dynamique des agent pathogènes. De nombreuses hypothèses alternatives peuvent être envisagées pour mieux comprendre et limiter leur impact. Afin de répondre au mieux à ces attentes, il est important de pouvoir fournir une vision quantitative des processus épidémiologiques à l'œuvre, nourrie par l'accroissement des capacités d'acquisition de données.

Les modèles classiques de dynamique épidémiologique peuvent servir à définir comment les agents pathogènes se maintiennent ou diffusent dans les populations susceptibles. Ils sont généralement appliqués à l'étude de distributions de cas. Les données moléculaires permettent de lever certaines ambiguïtés associées aux données de cas. Comme illustré plus haut, elles permettent par exemple d'identifier les infections épidémiologiquement liées ou de clarifier l'échelle de temps des processus épidémiologiques. Elles constituent donc une source d'information importante pour évaluer les facteurs qui influencent la distribution des pathogènes. La prise en compte de ces informations repose sur des modèles de phylody-

namique, à l'interface entre dynamique épidémiologique et biologie évolutive. Je souhaite positionner l'essentiel de mon travail dans ce champ de recherche en développement.

La réussite de ce type de projet passe par une association avec les acteurs de la surveillance, du diagnostique ou de la référence sur les pathogènes ciblés, afin de recueillir les informations pertinentes.

Je dois travailler au développement de différentes opportunités sur ce sujet. Dans le cadre du projet H2020 MOOD, j'espère travailler avec les partenaires intéressés par les approches phylodynamiques : i) Au outils facilitant la valorisation de telles approches dans une optique d'appui aux politiques publiques ; ii) à leur son application à différents modèles dont l'encéphalite virale à tique. Dans le cadre de la plateforme ESA, il existe des besoins similaires au point de vue méthodologique et applicatif sur des modèles biologiques comme la tuberculose ou la grippe. A l'heure actuelle, je participe à la construction d'un premier projet en lien avec cette orientation. Dans le cadre des travaux développés par Julien Thézé et en interaction avec Maude Jacquot, nous souhaitons développer un projet de recherche sur l'épidémiologie du virus de la diarrhée virale bovine (BVDV). Ce virus, encore très répandu en France, a fait l'objet de campagnes d'éradication dans différents pays de la Communauté européenne. Ces pays ont réussi à limiter grandement l'incidence du BVDV. Dans ces conditions, le gouvernement français a adopté le 31 juillet 2019 un décret visant à préciser les modalités de surveillance et de gestion du BVDV dans les élevages bovins Français. Le dispositif vise à identifier les animaux Immunotolérants Infectés Permanents (IPI) contrôler l'enzootie. En Auvergne-Rhône-Alpes, tous les veaux nouveaux-nés IPI devraient être identifiés et éliminés. Nous souhaitons nous greffer à cette campagne de détection pour construire un projet d'épidémiologie moléculaire à grande échelle. Dans la phase initiale du projet, des milliers d'animaux seront testés chaque année par notre partenaire du laboratoire Terana sur six départements en interaction avec les Groupement de Défense Sanitaire concernés. Les premières données suggèrent une prévalence de l'ordre de 0.1%. Nous voulons réaliser des séquençages ciblés pour l'essentiel des échantillons infectés et un séquençage génomique pour la caractérisation des échantillons clefs pour améliorer l'inférence de processus épidémiologiques. Une réponse à l'appel d'offre de la région Auvergne-Rhône-Alpes a été déposé pour soutenir ce projet et nous sommes en contact avec Boehringer Ingelheim pour envisager de futurs développements.

7.2 Comprendre les processus adaptatifs des agents pathogènes dans leur environnement

Les études permettant de comprendre la diffusion des pathogènes à partir de données de surveillance doivent intégrer l'effet des facteurs à l'origine des dynamiques observées. De ce fait, je continuerai à travailler sur des problématiques d'adaptation des micro-organismes à leur environnement. Les données de surveillance offrent l'opportunité d'obtenir une vision plus représentative de la diversité des pathogènes dans le temps, l'espace, et différentes conditions de circulation. Cela permet de mieux envisager l'impact de la sélection sur la diversité observée dans ces populations. Tout particulièrement, cela permet d'envisager la réponse des agents pathogènes à une modification de leur conditions de circulation, par exemple suite à

l'implémentation de nouveaux protocoles de gestion zootechnique ou sanitaire.

D'un point de vue méthodologique, l'étude de la sensibilité des méthodes phylodynamiques pour identifier de l'effet de facteurs affectant la diffusion de pathogènes est une question clef. Maude Jacquot souhaite s'investir sur ce sujet dans la continuité de ces projets actuels. Nous essayerons de mettre en place les conditions d'une nouvelle collaboration, en développant si possible ces travaux au sein de l'unité EPIA.

Plus généralement, la distribution du polymorphisme en ségrégation et son évolution au cours du temps doivent nous permettre de mieux comprendre quelle part de la diversité observée a un rôle adaptatif et éventuellement à quelle vitesse les pressions de sélection actuellement à l'œuvre modifient la dynamique des mutations au sein des populations de pathogène. La mise en relation des distributions de polymorphisme et des connaissances fonctionnelles sont une première étape pour l'identification des phénotypes importants pour la valeur sélective des pathogènes et pour déterminer quel est l'impact de ces phénotypes sur l'épidémiologie du pathogène. Les populations de *Mycoplasma bovis* pourraient constituer un modèle intéressant pour aborder ces questions. Cette bactérie est susceptible d'évoluer rapidement. Elle a connu un balayage sélectif dans les années 1990, probablement lié à la fixation de résistances aux antibiotiques. Elle présente depuis une diversité limitée ce qui facilite le suivi de l'apparition et de l'évolution de fréquence de potentielles mutations. Le dispositif de surveillance VigiMyc vise a documenter les souches de mycoplasme circulant en France depuis 10 ans. Des centaines de souches de cet agent pathogène sont centralisées chaque année par l'UMR Mycoplasmoses des ruminants dans le cadre de ce dispositif. Il serait intéressant de sonder l'évolution de la diversité des souches de *M. bovis* collectées depuis le départ de ce dispositif jusqu'à aujourd'hui pour observer le potentiel d'évolution de fréquence allélique sur cette période. Cette étude constituerait un préalable pour essayer d'estimer la distribution des effets sur la valeur sélective des mutations en ségrégation. Je souhaiterais participer à un projet exploratoire sur ce sujet, par exemple financé par un appel d'offre du réseau de recherche en Santé et bien-être Animal en Auvergne-Rhône-Alpes.

Dans un premier temps et dans la continuité des projets développés par Elsa Jourdain, nous souhaitons étudier la diffusion de *Coxiella burnetii* à une échelle régionale. Ce projet s'inscrit dans le cadre d'un protocole de surveillance mis en place à la suite d'infections zoonotiques répétées dans le département des Deux-Sèvres. Il implique des partenaires de l'ANSES, de l'agence régionale de santé de la région Grande Aquitaine, des GDS des Deux-Sèvres et des départements limitrophes ainsi que du laboratoire Qualyse. Des chiffonnettes ont servi à prélever des poussières au sein de bâtiments d'élevage de la zone d'étude. Une détection de *C. burnetii* a été effectuée sur ces échantillons, montrant une prévalence importante. Les traces d'ADN du pathogène détecté sont probablement issues d'épisodes de sécrétion suite à une infection au sein de l'élevage. Les génotypes présents dans les chiffonnettes présentant des traces d'infection sont à l'heure actuelle caractérisés par un séquençage haut-débit de deux produits PCR. Nous avons ainsi étudié la distribution de génotypes de *C. burnetii* observés au sein de plus d'une centaine de sites spécialisés dans l'élevage d'espèces différentes. Cet échantillonnage sera réalisé durant deux années consécutives au moins. Des données préliminaires indiquent que deux grandes lignées de *C. burnetii* sont observées, une lignée spécifique des élevages caprins et une lignée partagée par des élevages ovins et caprins. Les

élevages bovins ne semblent pas présenter les génotypes couramment rencontrés au sein de l'espèce. Les processus aboutissant au maintien de ces deux génotypes de *C. burnetii* seront étudiés en prenant en compte la distribution spatiale des échantillons présentant des traces de *C. burnetii* et la distribution des espèces hôtes au sein des élevages. Le financement de cette étude sera présenté au second tour de l'appel à projet PRN-EST de l'ANSES.

7.3 Étudier la dynamique épidémiologique des agents pathogènes au sein et entre leurs hôtes

Au delà de patrons globaux, je souhaiterais continuer à m'intéresser aux interactions entre dynamique des populations de bactéries intra-hôtes et à l'échelle de populations. Les modèles biologiques où les co-infections d'hôtes par différents génotypes d'agents pathogènes me semblent particulièrement pertinents pour aborder ces questions. Ces questions s'inscrivent notamment dans les thématiques de la Fédération de Recherche Systèmes Microbiens de Clermont-Ferrand.

D'un point de vue méthodologique, je travaille avec David Abrial à l'amélioration des méthodes utilisées pour délimiter des génotypes conspécifiques dans le cas d'infections multiples à partir de données de séquençage à haut débit. Ces méthodes nous aideront à étudier différentes questions biologiques portant les liens entre de la dynamique d'exposition des hôtes aux agents pathogènes, et la dynamique des interactions en agents pathogènes au sein des hôtes.

En premier lieu, il y aurait matière à continuer les études engagées sur les maladies transmises par les tiques. Je souhaiterais particulièrement continuer à travailler sur les patrons de co-infection par différents génotypes d'*A. phagocytophilum* au sein de ses hôtes. Nous disposons du suivi longitudinal de l'infection par *A. phagocytophilum* chez une population de chevreuils élevée en captivité sur près de dix ans. Des tiques à l'affût ont été régulièrement récoltées sur les parcelles utilisées par cet élevage. Une diminution progressive de la densité de chevreuil est planifiée dans le cadre de cette installation expérimentale.

Cette configuration pourrait nous permettre d'évaluer les processus permettant le maintien de la diversité d'*A. phagocytophilum* infectant cette population ainsi que la relation entre le niveau de co-infection de la population hôte et la densité en animaux. Un projet supervisé par Hélène Verheyden a été déposé sur ce sujet auprès d'une fondation.

D'autre part, des sujets seraient envisageables en lien avec les activités de recherche développées par Anaïs Bompard. Anaïs s'intéresse notamment aux processus écologiques permettant la diminution de l'utilisation des intrants médicamenteux et le contrôle du niveau d'antibio-résistance dans les populations d'animaux d'élevage. Ses premiers projets portent notamment sur la dynamique d'*Escherichia coli* et de *Staphylococcus aureus*. Malgré l'étendue des travaux menés sur ces modèle biologiques, j'ai le sentiment que des études sur la structure génétique des populations de ces espèces cibles prenant mieux en compte les co-infections au sein d'hôtes animaux aurait du sens. En effet, une grande partie des études conduites in-vivo se base sur l'isolement d'isolats en nombre relativement restreint par rapport aux populations d'intérêt. De ce fait, une partie des processus évolutifs à l'œuvre pourraient être mieux

documentée pour les comprendre et adapter les stratégies de gestion sanitaire.

Dans ce contexte, les premières études que j'envisagent donneraient l'occasion d'étudier i) comment évolue la diversité de ces bactéries au sein et entre les individus hôtes d'un élevage, au grès des traitements et autres perturbations, ii) le lien potentiel entre composition des populations des espèces bactériennes cibles, voir des communautés microbiennes associées, et la nécessité de traitements antibiotiques, et iii) La distribution de l'antibio-résistance dans les populations bactériennes cibles en fonction de la gestion de l'utilisation des traitements disponibles. J'ai déjà proposé, sans succès, quelques projets sur ces sujets par le passé. La mise en place de collaborations permettra, s'il y a lieu de persévérer, de renforcer la pertinence de mes propositions passées.

7.4 Happy end!

A travers les directions de recherche et les projets que j'envisage, je souhaite continuer à contribuer à ma communauté scientifique et apporter des informations utiles pour les porteurs d'enjeux dans le domaine des maladies animales et zoonotiques. Ces orientations visent à continuer à développer des collaborations : i) au sein de mon unité pour participer à une évolution cohérente de nos activités, et ii) en dehors pour m'ouvrir à de nouvelles questions et à de nouvelles perspectives. Je compte également m'investir dans la formation par la recherche, en espérant que ces interactions seront profitables pour les personnes encadrées. J'espère enfin arriver à concilier ces activités de recherche avec la direction de l'unité d'Épidémiologie des maladies animales et zoonotiques.

Cette habilitation à diriger les recherches est une première étape dans cette voie. C'est plein d'enthousiasme que m'engage sur ce chemin.

8 Bibliographie

References

Alizon S, Luciani F, Regoes RR (2011) Epidemiological and clinical consequences of within-host evolution. *Trends in Microbiology*, **19**, 24–32.

Alizon S, de Roode JC, Michalakis Y (2013) Multiple infections and the evolution of virulence. *Ecology Letters*, **16**, 556–567.

Almeida A, Rosinski-Chupin I, Plainvert C, *et al.* (2017) Parallel Evolution of Group B Streptococcus Hypervirulent Clonal Complex 17 Unveils New Pathoadaptive Mutations. *mSystems*, **2**, 1–15.

Bailly X (2017) Hidden Markov phylogenetic models offer an interesting perspective to identify "high risk lineages" of environmental pathogens. *Infection, Genetics and Evolution*, **55**.

Bailly X, Giuntini E, Sexton MC, *et al.* (2011) Population genomics of Sinorhizobium medicae based on low-coverage sequencing of sympatric isolates. *ISME Journal*, **5**, 1722–1734.

Bailly X, Migeon A, Navajas M (2004) Analysis of microsatellite variation in the spider mite pest Tetranychus turkestani (Acari: Tetranychidae) reveals population genetic structure and raises questions about related ecological factors. *Biological Journal of the Linnean Society*, **82**, 69–78.

Bailly X, Olivieri I, Brunel B, Cleyet-Marel JC, Béna G (2007) Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of Medicago species. *Journal of Bacteriology*, **189**, 5223–5236.

Bailly X, Olivieri I, De Mita S, Cleyet-Marel JC, Béna G (2006) Recombination and selection shape the molecular diversity pattern of nitrogen-fixing Sinorhizobium sp. associated to Medicago. *Molecular Ecology*, **15**, 2719–2734.

Barran LR, Bromfield ES, Brown DC (2002) Identification and cloning of the bacterial nodulation specificity gene in the Sinorhizobium meliloti - Medicago laciniata symbiosis. *Canadian Journal of Microbiology*, **48**, 765–771.

Bastide P, Ané C, Robin S, Mariadassou M (2018) Inference of Adaptive Shifts for Multivariate Correlated Traits. *Systematic Biology*, **67**, 662–680.

Beaumont MA (2010) Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Biek R, Pybus OG, Lloyd-Smith JO, Didelot X (2015) Measurably evolving pathogens in the genomic era. *Trends in Ecology and Evolution*, **30**, 306–313.

Boivin S, Ait Lahmidi N, Sherlock D, *et al.* (2019) Host-specific competitiveness to form nodules in Rhizobium leguminosarum symbiovar viciae . *New Phytologist*.

Brisson D, Drecktrah D, Eggers CH, Samuels DS (2012) Genetics of Borrelia burgdorferi. *Annual Review of Genetics*, **46**, 515–536.

Brockhurst MA, Harrison E, Hall JP, Richards T, McNally A, MacLean C (2019) The Ecology and Evolution of Pangenomes. *Current Biology*, **29**, R1094–R1103.

Capela D, Barloy-Hubler F, Gouzy J, *et al.* (2001) Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9877–9882.

Chaintreuil C, Rigault F, Moulin L, *et al.* (2007) Nickel resistance determinants in Bradyrhizobium strains from nodules of the endemic new caledonia legume Serianthes calycina. *Applied and Environmental Microbiology*, **73**.

Chastagner A, Dugat T, Vourc'h G, *et al.* (2014) Multilocus sequence analysis of Anaplasma phagocytophilum reveals three distinct lineages with different host ranges in clinically ill French cattle. *Veterinary Research*, **45**.

Chastagner A, Pion A, Verheyden H, *et al.* (2017) Host specificity, pathogen exposure, and superinfections impact the distribution of Anaplasma phagocytophilum genotypes in ticks, roe deer, and livestock in a fragmented agricultural landscape. *Infection, Genetics and Evolution*, **55**, 31–44.

Coipan EC, Jahfari S, Fonville M, *et al.* (2016) Imbalanced presence of Borrelia burgdorferi s.l. multilocus sequence types in clinical manifestations of Lyme borreliosis. *Infection, Genetics and Evolution*, **42**, 66–76.

Collins C, Didelot X (2018) A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Computational Biology*, **14**, 1–21.

Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ (2010) Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Computational Biology*, **6**.

Cordes FS, Kraiczy P, Roversi P, *et al.* (2006) Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of Borrelia burgdorferi. *International Journal of Medical Microbiology*, **296**, 177–184.

Costechareyre D, Bertolla F, Nesme X (2009) Homologous recombination in Agrobacterium: Potential implications for the genomic species concept in bacteria. *Molecular Biology and Evolution*, **26**, 167–176.

Cougoul A, Bailly X, Vourc'h G, Gasqui P (2019a) Rarity of microbial species: In search of reliable associations. *PloS one*, **14**, e0200458.

Cougoul AP, Bailly X, Wit EC (2019b) MAGMA: inference of sparse microbial association networks. *bioRxiv*, S. 538579.

Didelot X, Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics*, **175**, 1251–1266.

Diuk-Wasser MA, Vannier E, Krause PJ (2016) Coinfection by Ixodes Tick-Borne Pathogens: Ecological, Epidemiological, and Clinical Consequences. *Trends in Parasitology*, **32**, 30–42.

Durand J, Herrmann C, Genné D, Sarr A, Gern L, Voordouw MJ (2017a) Multistrain infections with lyme borreliosis pathogens in the tick vector. *Applied and Environmental Microbiology*, **83**, 1–14.

Durand J, Jacquet M, Rais O, Gern L, Voordouw MJ (2017b) Fitness estimates from experimental infections predict the long-term strain structure of a vector-borne pathogen in the field. *Scientific Reports*, **7**, 1–9.

Estrada-Peña A, de la Fuente J (2017) Host Distribution Does Not Limit the Range of the Tick Ixodes ricinus but Impacts the Circulation of Transmitted Pathogens. *Frontiers in Cellular and Infection Microbiology*, **7**, 1–10.

Falush D, Bowden R (2006) Genome-wide association mapping in bacteria? *Trends in Microbiology*, **14**, 353–355.

Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. *Science*, **315**, 476–480.

Giuntini E, Mengoni A, De Filippo C, *et al.* (2005) Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of Sinorhizobium meliloti natural strains. *BMC Genomics*, **6**, 1–11.

Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology*, **3**, 679–687.

Grenfell BT, Pybus OG, Gog JR, *et al.* (2004) Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*, **303**, 327–332.

Hanage WP, Fraser C, Spratt BG (2006) The impact of homologous recombination on the generation of diversity in bacteria. *Journal of Theoretical Biology*, **239**, 210–219.

Harrison PW, Lower RP, Kim NK, Young JPW (2010) Introducing the bacterial 'chromid': Not a chromosome, not a plasmid. *Trends in Microbiology*, **18**, 141–148.

Hart T, Nguyen NTT, Nowak NA, *et al.* (2018) *Polymorphic factor H-binding activity of CspA protects Lyme borreliae from the host complement in feeding ticks to facilitate tick-to-host transmission*, Bd. 14.

Haven J, Vargas LC, Mongodin EF, *et al.* (2011) Pervasive recombination and sympatric genome diversification driven by frequency-dependent selection in borrelia burgdorferi, the lyme disease bacterium. *Genetics*, **189**, 951–966.

Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genetical research*, **8**, 269–294.

Jacquot M, Abrial D, Gasqui P, *et al.* (2016) Multiple independent transmission cycles of a tick-borne pathogen within a local host community. *Scientific Reports*, **6**.

Jacquot M, Bisseux M, Abrial D, *et al.* (2014a) High-throughput sequence typing reveals genetic differentiation and host specialization among populations of the Borrelia burgdorferi species complex that infect rodents. *PLoS ONE*, **9**.

Jacquot M, Gonnet M, Ferquel E, *et al.* (2014b) Comparative population genomics of the Borrelia burgdorferi species complex reveals high degree of genetic isolation among species and underscores benefits and constraints to studying intra-specific epidemiological processes. *PLoS ONE*, **9**.

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, **9**, 1–8.

Joulié A, Sidi-Boumedine K, Bailly X, *et al.* (2017) Molecular epidemiology of Coxiella burnetii in French livestock reveals the existence of three main genotype clusters and suggests species-specific associations as well as regional stability. *Infection, Genetics and Evolution*, **48**, 142–149.

Kim Y, Stephan W (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, **155**, 1415–1427.

Kraiczy P (2016) Hide and seek: How lyme disease spirochetes overcome complement attack. *Frontiers in Immunology*, **7**.

Kumar N, Lad G, Giuntini E, *et al.* (2015) Bacterial genospecies that are not ecologically coherent: Population genomics of rhizobium leguminosarum. *Open Biology*, **5**.

Kwan JY, Griggs R, Chicana B, Miller C, Swei A (2017) Vertical vs. horizontal transmission of the microbiome in a key disease vector, Ixodes pacificus. *Molecular Ecology*, **26**, 6578–6589.

Lawrence JG, Ochman H (1998) Molecular archaeology of the Escherichia coli genome. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 9413–9417.

Lohr B, Müller I, Mai M, Norris DE, Schöffski O, Hunfel KP (2015) Epidemiology and cost of hospital care for Lyme borreliosis in Germany: Lessons from a health care utilization database analysis. *Ticks and Tick-borne Diseases*, **6**, 56–62.

Maiden MC, Van Rensburg MJ, Bray JE, *et al.* (2013) MLST revisited: The gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, **11**, 728–736.

Maiden MCJ, Bygraves JA, Feil E, *et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, **95**, 3140–3145.

Majewski J, Cohan FM (1999) Adapt globally, act locally: The effect of selective sweeps on bacterial sequence diversity. *Genetics*, **152**, 1459–1474.

Marcnkiewicz AL, Lieknina I, Kotelovica S, *et al.* (2018) Eliminating factor H-Binding activity of Borrelia burgdorferi CspZ combined with virus-like particle conjugation enhances its efficacy as a Lyme disease vaccine. *Frontiers in Immunology*, **9**, 1–11.

Margos G, Vollmer SA, Ogden NH, Fish D (2011) Population genetics, taxonomy, phylogeny and evolution of Borrelia burgdorferi sensu lato. *Infection, Genetics and Evolution*, **11**, 1545–1563.

Marttinen P, Hanage WP (2017) Speciation trajectories in recombining bacterial species. *PLoS Computational Biology*, **13**, 1–15.

Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends in Microbiology*, **17**, 458–466.

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences*, **90**, 4384–4388.

Mechai S, Margos G, Feil EJ, *et al.* (2016) Evidence for host-genotype associations of Borrelia burgdorferi sensu stricto. *PLoS ONE*, **11**, 1–25.

Merker M, Barbier M, Cox H, *et al.* (2018) Compensatory evolution drives multidrug-resistant tuberculosis in central Asia. *eLife*, **7**, 1–31.

Mideo N, Alizon S, Day T (2008) Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in Ecology and Evolution*, **23**, 511–517.

Mongodin EF, Casjens SR, Bruno JF, *et al.* (2013) Inter- and intra-specific pan-genomes of Borrelia burgdorferi sensu lato: Genome stability and adaptive radiation. *BMC Genomics*, **14**, 1.

Ochman H (2002) Bacterial evolution: Chromosome arithmetic and geometry. *Current Biology*, **12**, 427–428.

Pavoine S, Bailly X (2007) New analysis for consistency among markers in the study of genetic diversity: Development and application to the description of bacterial diversity. *BMC Evolutionary Biology*, **7**, 1–16.

Porter SS, Chang PL, Conow CA, Dunham JP, Friesen ML (2017) Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic Mesorhizobium. *ISME Journal*, **11**, 248–262.

Råberg L, Hagström, Andersson M, *et al.* (2017) Evolution of antigenic diversity in the tick-transmitted bacterium Borrelia afzelii: a role for host specialization? *Journal of Evolutionary Biology*, **30**, 1034–1041.

Rasigade JP, Barbier M, Dumitrescu O, *et al.* (2017) Strain-specific estimation of epidemic success provides insights into the transmission dynamics of tuberculosis. *Scientific Reports*, **7**, 1–12.

Rice WR (2013) Nothing in Genetics Makes Sense Except in Light of Genomic Conflict. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 217–237.

Ross BD, Hayes B, Radey MC, *et al.* (2018) Ixodes scapularis does not harbor a stable midgut microbiome. *The ISME journal*, S. 198267.

Sabarly V, Bouvet O, Glodt J, *et al.* (2011) The decoupling between genetic structure and metabolic phenotypes in Escherichia coli leads to continuous phenotypic diversity. *Journal of Evolutionary Biology*, **24**, 1559–1571.

Shapiro BJ, Friedman J, Cordero OX, *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science*, **335**, 48–51.

Sheppard SK, Didelot X, Meric G, *et al.* (2013) Genome-wide association study identi fi es vitamin B 5 biosynthesis as a host speci fi city factor in Campylobacter. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1–5.

Singh P, Verma D, Backstedt BT, *et al.* (2017) Borrelia burgdorferi BBI39 paralogs, targets of protective immunity, reduce pathogen persistence either in hosts or in the vector. *Journal of Infectious Diseases*, **215**, 1000–1009.

Sofonea MT, Alizon S, Michalakis Y (2017) Exposing the diversity of multiple infection patterns. *Journal of Theoretical Biology*, **419**, 278–289.

Stanek G, Wormser GP, Gray J, Strle F (2012) Lyme borreliosis. *The Lancet*, **379**, 461–473.

Strandh M, Råberg L (2015) Within-host competition between B. afzelii ospC strains as revealed by massively parallell amplicon sequencing. *Philosophical transactions of the Royal Society of London. B*, **370**.

Stuen S, Granquist EG, Silaghi C (2013) Anaplasma phagocytophilum–a widespread multi-host pathogen with highly adaptive strategies. *Frontiers in cellular and infection microbiology*, **3**, 31.

Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, **11**, 472–477.

Van Den Wijngaard CC, Hofhuis A, Wong A, *et al.* (2017) The cost of Lyme borreliosis. *European Journal of Public Health*, **27**, 538–547.

Vayssier-Taussat M, Rhun DL, Deng HK, *et al.* (2010) The Trw type IV secretion system of Bartonella mediates host-specific adhesion to erythrocytes. *PLoS Pathogens*, **6**.

Vera Cruz CM, Bai J, Oña I, *et al.* (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 13500–13505.

Vetsigian K, Goldenfeld N (2005) Global divergence of microbial genome sequences mediated by propagating fronts. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7332–7337.

Villegas M, Rome S, Mauré L, *et al.* (2006) Nitrogen-fixing sinorhizobia with Medicago laciniata constitute a novel biovar (bv. medicaginis) of S. meliloti. *Systematic and Applied Microbiology*, **29**.

Vollmer SA, Bormane A, Dinnis RE, *et al.* (2011) Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Environmental Microbiology*, **13**, 184–192.

Vulić M, Dionisio F, Taddei F, Radman M (1997) Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9763–9767.

Walter KS, Carpi G, Evans BR, Caccone A, Diuk-Wasser MA (2016) Vectors as Epidemiological Sentinels: Patterns of Within-Tick Borrelia burgdorferi Diversity. *PLoS Pathogens*, **12**, 1–18.

Weiss S, Van Treuren W, Lozupone C, *et al.* (2016) Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME Journal*, **10**, 1669–1681.