

Thématique de recherche

Tristan Lefébure

Mes recherches incluent une grande variété de modèles biologiques (des crustacés aux bactéries), d'habitats (des eaux souterraines aux tissus des mammifères colonisés par des pathogènes), et de champs d'applications (de la gestion de la biodiversité à la médecine). Cette apparente hétérogénéité recouvre en fait un seul et unique but : comprendre les mécanismes évolutifs et adaptatifs à l'origine de la biodiversité en utilisant le matériel porteur de l'information génétique, l'ADN et le génome. L'essor des méthodes moléculaires, hier la PCR et aujourd'hui les nouvelles techniques de séquençage [1], mais aussi l'essor des méthodes de modélisation probabiliste ainsi que l'accroissement de la puissance de calcul, ouvrent de nouvelles voies quant à notre compréhension de l'évolution biologique. J'ai développé ces approches dans des cadres universitaires où l'évolution n'est généralement pas le principal sujet d'étude (un laboratoire d'écologie et une école vétérinaire). En cela je pense que mes recherches ont ouvert de nouvelles perspectives et ont une nouvelle fois vérifié l'universalité des sciences de l'évolution et leur rôle fondamental en biologie. Concernant le futur, je m'intéresse maintenant à un modèle souterrain en particulier, *Proasellus*, pour comprendre les phénomènes adaptatifs à l'échelle génomique. Ce modèle très riche permettra de s'attaquer à des questions tels que l'évolution du taux de substitution, de la taille des génomes, de la composition des protéines, du devenir des pseudo-gènes, de la régulation du vieillissement, ..., le tout dans un cadre comparatif extrêmement riche.

1 - Biodiversité, évolution et biogéographie souterraine

La très grande majorité des organismes souterrains présentent les mêmes caractéristiques [2-6]: anophtalmie, dépigmentation, cycle de vie allongé, ralentissement du métabolisme, résistance à des périodes prolongées de jeûne et d'hypoxie ...A cela on peut rajouter un fort endémisme associé à un fort rapport # [7, 8], représentatif des fortes radiations spécifiques. La plupart de ces caractéristiques ont été acquises par convergence au cours de l'adaptation aux conditions extrêmes des milieux souterrains. Dans ce cadre, mon travail de recherche a tout d'abord été de vérifier à quel point ce phénomène massif de convergence pouvait avoir une influence sur notre estimation de la biodiversité souterraine. J'ai pour cela étudié la phylogéographie de deux espèces contrastées de *Niphargus* à l'aide d'un fort échantillonnage et du séquençage de plusieurs gènes [9, 10]. Ces deux travaux ont démontré l'importance de la biodiversité cachée en milieu souterrain, à savoir la sous-estimation par des techniques de taxonomie morpho-anatomique classique du nombre d'espèces. La synthèse de ces travaux et d'autres [11], ont également montrés combien la distribution des organismes souterrains est difficile à appréhender *a priori* sans approche moléculaire : dans bien des cas l'endémisme est sous-estimé, mais dans d'autre cas, moins nombreux, il peut être sur-estimé.

En parallèle, j'ai étudié l'impact des phénomènes historiques, et particulièrement les glaciations du quaternaire, sur l'évolution et la distribution des organismes souterrains. Classiquement, le milieu souterrain est considéré comme un milieu stable à l'échelle des temps géologiques, ce qui aurait permis à certaines lignées dites " fossiles " ou " reliques " d'y subsister [12]. Le corollaire à cette vision statique du milieu souterrain, est de le considérer comme un milieu refuge où s'accumule au cours des temps les espèces " reliques ". Néanmoins il est également supposé que des évènements comme les glaciations ou des périodes d'arridification ont pu détruire la faune souterraine. Mes

recherches phylogéographiques ont tout d'abord montré que certaines populations de *Niphargus rhenorhodanensis* ont probablement survécu aux glaciations au cœur même des Alpes [10]. Enfin, une analyse de la structure génétique couplée à une analyse de distribution de *Niphargus virei* à l'échelle du Jura, ont montré combien la dernière glaciation a repoussé la distribution de cette espèce aux limites de l'extension des glaciers, laissant, plusieurs milliers d'années après le retrait des glaciers, une forte empreinte sur la distribution et la structure génétique de *Niphargus virei* [13].

Le dernier aspect que j'ai étudié est la structuration des populations souterraines et leur capacité de dispersion. Il est communément établi dans la littérature que les organismes souterrains ont de très faibles capacités de dispersion, générant ainsi un fort endémisme. Toutes mes recherches confortent cette idée [9, 10, 13]. Cependant nous ne savons toujours pas quels sont les obstacles à la dispersion en milieu souterrain : est-ce la biologie de ces organismes ou bien la structure de leur habitat ? Dans ce dernier cas, quels sont les obstacles physiques à la dispersion ? Répondre à ces questions aurait de fortes implications en biologie de la conservation, car dans bien des cas les espèces souterraines endémiques sont potentiellement menacées par la pollution ou la sur-exploitation des aquifères. Hors, nous n'avons presque aucune idée de la taille et de la distribution des populations, voire de la continuité de leur habitat. Il est alors très difficile de faire des pronostics quant à la survie d'une espèce et quant aux stratégies qui devraient être conduites. Pour répondre à ces questions, j'ai procédé à un important échantillonnage régional, et j'ai analysé la structure génétique de *Niphargus virei* à l'aide de marqueurs AFLP. Un Système d'Information Géographique de la région étudiée recoupant la distribution des aquifères et leurs caractéristiques, a été développé. Le couplage de ces informations génétiques et géographiques est le sujet d'un article en préparation, déjà partiellement publié dans ma thèse [14].

2 - Le concept d'espèce, la taxonomie moléculaire et le *DNA barcoding*

D'un côté, le développement des techniques de séquençage haut-débit, et de l'autre, la quantité gigantesque d'espèces à décrire alors que le nombre de spécialistes ne fait que diminuer, ont conduit certains auteurs à proposer un système taxonomique où l'ADN aurait un rôle central [15-19]. Cette proposition a soulevé de vives réactions où, pêle-mêle, sont confrontés des arguments théoriques, techniques et de politiques scientifiques [20-27]. Une partie importante du débat est issue d'une confusion sur la définition même de la taxonomie moléculaire. Sous les termes de "DNA taxonomy" et de "DNA barcoding" se cachent en fait deux objectifs: (1) le développement de techniques d'identification moléculaire (auxquelles le terme "DNA barcoding" serait le plus approprié), et (2) le développement de méthodes de définition et de classification de nouveaux taxons. Alors que le premier objectif est finalement la recherche de synapomorphies moléculaires, le second implique que l'hypothétique classification moléculaire et la classification actuelle soient cohérentes. Cela sous-entend, entre autre, qu'une gamme de divergence moléculaire puisse être attribuée à un rang taxonomique, nommé alors MOTU ("Molecular Operational Taxonomic Unit", [19]). Les détracteurs comme les avocats de la taxonomie moléculaire s'entendent sur le fait que ce dernier point est une limitation majeur au développement de cette démarche [15, 19, 23, 24, 28]. J'ai mis au point un test de la corrélation entre divergence moléculaire et rang taxonomique chez les crustacés à partir des séquences de COI et de 16S disponibles dans les bases de données génomiques [29]. Nous avons observé une corrélation globale entre la divergence moléculaire et les rangs taxonomiques. Cette corrélation est particulièrement forte pour le COI au niveau des faibles rangs taxonomiques. Pour les niveaux taxonomiques plus élevés, la corrélation diminue fortement jusqu'à devenir très

faible entre familles. En supprimant de notre jeu de données les cas reconnus de diversité cachée, ainsi qu'un genre de décapode possédant des divergences inter-espèces très faibles, nous avons montré que les divergences intra-espèces et inter-espèces peuvent être discriminées avec un très fort succès (99%) en utilisant un seuil moléculaire de 0,16 substitutions par site. Il semble qu'il existe une entité morphologique et moléculaire qui correspond à ce qui est couramment appelé une espèce chez les crustacés. Dans ce cadre, le terme d'espèce désigne la plus petite unité taxonomique et est ainsi indépendant d'une définition particulière de l'espèce. On peut alors le remplacer par celui d'OTU (i.e. "Operational taxonomic unit"). Cette adéquation nous a permis de proposer un outil pour aider la délimitation des espèces de crustacés. Elle fournit également un marqueur pragmatique de la biodiversité des eaux souterraines.

Afin de tester la capacité universelle du COI à délimiter les espèces, je suis entrain de généraliser le même type d'analyse à tous les métazoaires. Les résultats préliminaires montrent que les divergences moléculaires du COI permettent de délimiter la plupart des espèces avec succès. Ils montrent néanmoins qu'il n'existe pas un seuil moléculaire universel, mais plutôt autant de seuil que de genre. Certains groupes montrent d'importantes exceptions à cette règle, comme les crustacés, certains ordres d'insectes et les chiroptères, où la délimitation des espèces reste possible avec un unique seuil moléculaire. Ces succès sont à la fois à mettre au compte d'une homogénéité dans les taux de mutations du COI et dans les techniques de délimitation employées par les taxonomistes au sein de ces groupes, mais ils sont surtout à mettre au compte d'importants "barcoding gaps", à savoir la présence de grande variation entre les divergences intra et inter-spécifiques.

3 - L'évolution des pathogènes bactériens

Beaucoup de groupes de pathogènes montrent d'étonnantes capacités d'adaptations. Par exemple, certaines espèces ou souches sont strictement adaptées à un hôte alors que d'autres parasitent un large éventail d'hôtes, certaines espèces sont sujettes à des sauts d'hôtes et de rapide pandémie alors que d'autres pas, beaucoup de pathogènes acquièrent de nouveaux facteurs de virulence ou deviennent résistants aux antibiotiques dans des périodes de temps records ... Bien que beaucoup de génomes de pathogènes bactériens est été séquencés, on connaît encore assez peu l'évolution des pathogènes et leurs mécanismes adaptatifs [30-33]. Les transferts latéraux de gènes sont supposés être un mécanisme fondamental [34-37], mais l'on ne connaît pas leur étendue à l'échelle génomique, leur fréquence, ainsi que leur part dans l'adaptation des pathogènes. Mon travail de recherche est alors de décrire les processus adaptatifs de certains des principaux groupes de pathogènes, comme les genres *Escherichia*, *Streptococcus*, *Campylobacter* et *Staphylococcus*, à partir de données génomiques, de méthodes phylogénétiques et de modélisations de l'évolution moléculaire, autrement dit, par une approche phylogénomique.

J'ai tout d'abord étudié l'évolution du genre *Streptococcus*, et l'importance relative des transferts de gènes et de la sélection naturelle positive sur l'évolution des protéines, et ce à différentes échelles évolutives et dans différents compartiments génomiques [38]. Cette étude a montré combien la sélection naturelle avait modelé et adapté une grande partie du génome-noyau¹ des différentes espèces de *Streptococcus*. Cette étude a également souligné l'importance des transferts de gènes à toutes les échelles évolutives, générant un large pan-génome², mais participant également à l'évolution du génome-

1 Le génome-noyau, "the core-genome", est la portion du génome partagée par tous les membres de la même espèce (généralement 75% du génome)

2 Le pan-génome, "the pan-genome", est l'ensemble des gènes trouvés chez une espèce

noyau. Une analyse similaire chez le genre *Campylobacter* a montrée que plus de 90% du génome-noyau est sous sélection positive. Bien que les différentes espèces de *Campylobacter* partagent le même habitat (les intestins de mammifères), très peu de cas de convergence moléculaire ont été détecté, laissant au contraire supposer que la sélection positive est liée à un processus de divergence fonctionnelle. Ces découvertes supportent l'idée selon laquelle l'intestin est un habitat très hétérogène où la compétition entre taxon est très forte, entraînant une très forte dynamique adaptative.

Par la suite, nous avons séquencé le génome de *Streptococcus canis*, le groupe frère de *Streptococcus pyogenes*, l'un des principaux pathogènes humains. La particularité de *S. pyogenes* est d'être un pathogène strictement humain, alors que les autres espèces du groupe "pyogenic" sont des généralistes colonisant plusieurs hôtes. L'évolution de *S. pyogenes* a donc été principalement marquée par la stricte adaptation à l'hôte humain. Une analyse phylogénomique a montré que cette adaptation est corrélée à des gains de facteurs de virulence, mais aussi au développement de nouveaux systèmes de régulation de la virulence, ainsi que l'adaptation à de nouvelles sources de carbone et d'énergie (publication en révision majeur). Une étude similaire a été entreprise pour déterminer comment *Staphylococcus aureus* est devenu un pathogène humain, cette fois en séquencant le génome de *S. simiae*, le groupe frère de *S. aureus* et pathogène des singes Saïmiri.

J'ai également participé à plusieurs travaux sur l'acquisition de la résistance à l'amoxicilline chez *Streptococcus pneumoniae*. Un premier article a permis de déterminer la fréquence relative de propagation de la résistance par transmission horizontale (transfert de gène) et verticale (dispersion clonale) [39]. Un deuxième article [40] propose de nouveaux sites au sein de plusieurs des "Penicillin Binding Proteins" (*pbp1a*, *pbp2b*, *pbp2x*) qui pourraient avoir une implication dans la résistance à l'amoxicilline. Ces sites présentent des excès de substitutions non-synonymes, indiquant de la sélection naturelle positive, et sont corrélés à des phénotypes résistants.

Depuis un an, je participe à un programme du NIH qui vise à reconstruire l'évolution du genre *Streptococcus*, en particulier le groupe "pyogenic". Je coordonne le séquençage et l'analyse des génomes d'une dizaine d'espèces du groupe, mais aussi le séquençage d'une centaine de souches pour trois espèces de *Streptococcus*. Ce fort échantillonnage permettra, à l'aide de méthodes phylogénomiques et de génomique des populations, d'obtenir une vision globale de l'évolution du groupe et de déterminer les événements adaptatifs qui ont conduit à la diversification d'un groupe particulièrement virulent et diversifié.

4 - Le concept d'espèce bactérienne

La définition du concept d'espèce est l'une des plus compliquées et controversées de la biologie [41]. Elle devient d'autant plus un défi pour les microbiologistes [42-44], certains considérant que le concept d'espèce n'est pas applicable aux bactéries [45]. Tout d'abord, les prokaryotes n'ont pas recours à la reproduction sexuée, réfutant ainsi toute notion biologique de l'espèce bactérienne. Enfin, l'afflux de données génomiques a démontré l'importance à toutes les échelles évolutives des transferts latéraux de matériel génétique, remettant en cause une vision phylogénétique uniquement verticale, et par la même, mettant à mal les notions phylogénétiques de l'espèce bactérienne. Récemment, le séquençage de plusieurs souches au sein de la même espèce bactérienne, a dévoilé l'existence d'un génome-noyau et d'un pan-génome [46, 47]. Il a alors été proposé que ce qui caractérise l'espèce est son genome-noyau. Néanmoins, comme démontré par mes recherches [38], il est maintenant clair que le génome-noyau est lui aussi soumis à la

(pouvant être 2 à 3 fois plus grand que le génome)

recombinaison. Une autre approche, “the species core-genome hypothesis”, permet d’accomoder l’existence de transferts de gènes au sein du génome-noyau [48]. Selon cette hypothèse, les individus de la même espèce bactérienne partageraient un génome-noyau au sein duquel ils pourraient échanger du matériel génétique, alors que ces échanges seraient stoppées entre espèces différentes, permettant ainsi la formation de groupes cohésifs, à l’instar des espèces à reproduction sexuée. Ce comportement a été modélisé et fonctionne *in silico* [49], néanmoins il n’existe pas de test *in vivo*. A partir de séquençage Illumina pour 80 souches appartenant à deux espèces soeurs de *Campylobacter*, j’ai pu démontrer à la fois (i) l’existence d’un pan-génome fini, (ii) l’existence d’un set de gène unique à chaque espèce, (iii) mais surtout la quasi absence de transfert de gènes au sein du génome coeur entre les deux espèces, alors que ceux-ci sont courant au sein de la même espèce [50]. Cette analyse conforte donc à l’échelle génomique la théorie du “core-genome hypothesis”. Une seconde publication en cours de préparation tirera cette fois partie du grand nombre de génomes disponibles pour plusieurs espèces des genres *Streptococcus*, *Campylobacter*, *Salmonella*, *Escherichia*, *Yersinia*, *Bacillus*, *Staphylococcus* et *Clostridium*, et tentera de valider l’universalité des résultats obtenus lors de l’analyse précédente.

References

- [1] Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–80.
- [2] Poulson T, White W (1969) The cave environment. *Science* 165:971–981.
- [3] Culver D (1982) *Cave life, evolution and ecology* (Harvard University Press, Cambridge).
- [4] Hervant F, Mathieu J, Barr H (1999) Comparative study on the metabolic responses of subterranean and surface-dwelling amphipods to long-term starvation and subsequent refeeding. *J. Exp. Biol.* 202:3587–3595.
- [5] Malard F, Hervant F (1999) Oxygen supply and the adaptations of animals in groundwater. *Freshwat. Biol.* 41:1–30.
- [6] Hüppop K (2000) in *Subterranean Ecosystems*, Ecosystems of the world, eds Wilkens H, Culver D, Humphreys W (Elsevier) Vol. 30, pp 159–188.
- [7] Stoch F (1995) The ecological and historical determinants of crustacean diversity in groundwaters, or : why are there so many species ? *Mémoires de Biospéologie* 23:139–160.
- [8] Gibert J, Deharveng L (2002) Subterranean ecosystems: a truncated functional biodiversity. *Bioscience* 52:473–481.
- [9] Lefébure T, et al. (2006) Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Mol. Ecol.* 15:1797–806.
- [10] Lefébure T, Douady CJ, Malard F, Gibert J (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Mol. Phylogenet. Evol.* 42:676–86.
- [11] Trontelj P, et al. (2009) A molecular test for cryptic diversity in groundwater: how large are the ranges of macro-stygobionts? *Freshwat. Biol.* 54:727–744.
- [12] Humphreys W (2000) in *Subterranean ecosystems*, Ecosystems of the world, eds

Wilkens H, Culver D, Humphreys W (Elsevier), pp 417–431.

- [13] Foulquier A, Malard, F and Lefébure T, Douady CJ, Gibert J (2008) The imprint of quaternary glaciers on the present-day distribution of the obligate groundwater amphipod *Niphargus virei* (Niphargidae). *Journal of Biogeography* 35:552–564.
- [14] Lefébure T (2005) Ph.D. thesis (Université Lyon 1, Villeurbanne (Fr)).
- [15] Hebert P, Cywinska A, Ball S, deWaard J (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* 270:313–21.
- [16] Hebert P, Ratnasingham S, deWaard J (2003) Barcoding animal life : cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B (Suppl.) Biology Letters* 270:S96–S99.
- [17] Stoeckle M (2003) Taxonomy, DNA, and the Bar Code of Life. *Bioscience* 53:796–797.
- [18] Tautz D, Arctander P, Minelli A, Thomas R, Vogler A (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.* 18:70–74.
- [19] Blaxter ML (2004) The promise of DNA taxonomy. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences* 359:669–679.
- [20] Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: A comment on DNA taxonomy. *Trends Ecol. Evol.* 18:65–66.
- [21] Mallet J, Willmott K (2003) Taxonomy: renaissance or tower of babel ? *Trends Ecol. Evol.* 18:57–59.
- [22] Seberg O, et al. (2003) Shortcuts in systematics? a commentary on DNA-based taxonomy. *Trends Ecol. Evol.* 18:63–65.
- [23] Will K, Rubinoff D (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20:47–55.
- [24] Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biol.* 2:e354.
- [25] Schindel D, Miller S (2005) DNA barcoding a useful tool for taxonomists. *Nature* 435:17.
- [26] Gregory T (2005) DNA barcoding does not compete with taxonomy. *Nature* 434:1067.
- [27] Ebach M, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. *Nature* 434:697.
- [28] Ferguson J (2002) On the use of genetic divergence for indentifying species. *Biol. J. Linn. Soc. Lond.* 75:509–516.
- [29] Lefébure T, Douady CJ, Gouy M, Gibert J (2006) Relationship between morphological taxonomy and molecular divergence within crustacea: proposal of a molecular threshold to help species delimitation. *Mol. Phylogenet. Evol.* 40:435–47.
- [30] Pallen MJ, Wren BW (2007) Bacterial pathogenomics. *Nature* 449:835–42.
- [31] Maurelli AT (2007) Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. *FEMS Microbiology Letter* 267:1–8.
- [32] Wirth T, et al. (2006) Sex and virulence in escherichia coli: an evolutionary perspective. *Mol. Microbiol.* 60:1136–51.
- [33] Wright GD (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* 5:175–86.
- [34] Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.
- [35] Awadalla P (2003) The evolutionary genomics of pathogen recombination. *Nat. Rev. Genet.* 4:50–60.
- [36] Hao W, Golding GB (2006) The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome Res.* 16:636–643.
- [37] Marri P, Hao W, Golding G (2006) Gene Gain and Gene Loss in *Streptococcus*: Is it

Driven by Habitat? *Mol. Biol. Evol.* 23:2379–2391.

- [38] Lefébure T, Stanhope MJ (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8:R71.
- [39] Stanhope MJ, et al. (2007) The relative frequency of intraspecific lateral gene transfer of penicillin binding proteins 1a, 2b, and 2x, in amoxicillin resistant *Streptococcus pneumoniae*. *Infect. Genet. Evol.* 7:520–34.
- [40] Stanhope MJ, et al. (2008) Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. *Infect. Genet. Evol.* 8:331–339.
- [41] Claridge M, Dawah H, Wilson M (1997) *Species: The Units of Biodiversity* (Chapman & Hall, London).
- [42] Cohan FM (2002) What are bacterial species? *Annu. Rev. Microbiol.* 56:457–87.
- [43] Hanage WP, Fraser C, Spratt BG (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol.* 3:6.
- [44] Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17:R373–86.
- [45] Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. *Genome Biol.* 7:116.
- [46] Tettelin H, et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. U.S.A.* 102:13950–13955.
- [47] Obert C, et al. (2006) Identification of a candidate streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. *Infect. Immun.* 74:4766–77.
- [48] Lan R, Reeves PR (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol.* 8:396–401.
- [49] Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. *Science* 315:476–80.
- [50] Lefébure T, Pavinski Bitar P, Suzuki H, Stanhope M (2010) Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biol. Evol.* 2:646–655.