



HAL
open science

Analyse génétique exhaustive du mécanisme de transformation naturelle chez *Legionella pneumophila* et caractérisation d'un nouvel acteur conservé chez les bactéries Gram-négative

Léo Hardy

► **To cite this version:**

Léo Hardy. Analyse génétique exhaustive du mécanisme de transformation naturelle chez *Legionella pneumophila* et caractérisation d'un nouvel acteur conservé chez les bactéries Gram-négative. Microbiologie et Parasitologie. Université de Lyon, 2021. Français. NNT : 2021LYSE1322 . tel-04348091

HAL Id: tel-04348091

<https://theses.hal.science/tel-04348091>

Submitted on 16 Dec 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.



THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 341
« Evolution, Ecosystèmes, Microbiologie, Modélisation »

Spécialité de doctorat : Aspect moléculaire et cellulaire de la biologie
Discipline : Microbiologie moléculaire

Soutenue publiquement le 15/12/2021, par :
Léo Hardy

Analyse génétique exhaustive du mécanisme de transformation naturelle chez *Legionella pneumophila* et caractérisation d'un nouvel acteur conservé chez les bactéries Gram-négative

Devant le jury composé de :

Madame LEBLOND-BOURGET Nathalie, rapportrice, Professeure des Universités, Université de Lorraine, UMR 1128 - DynAMic

Monsieur RADICELLA Pablo, rapporteur, Directeur de Recherche, CEA Fontenay-aux-roses, UMR SGCSR - Laboratoire de Recherche sur l'Instabilité Génétique

Madame ANSALDI Mireille, examinatrice, Directrice de Recherche CNRS, CNRS Marseille, UMR 7283 – Laboratoire de Chimie Bactérienne

Madame JARRAUD Sophie, Professeure des Universités, Praticien Hospitalier, Faculté de Médecine Lyon, U1111 - CIRI

Monsieur CHARPENTIER Xavier, directeur de thèse, directeur de recherche, INSERM Lyon, Inserm U1111 - CIRI

Université Claude Bernard – LYON 1

| | |
|---|-------------------------|
| Administrateur provisoire de l'Université | M. Frédéric FLEURY |
| Président du Conseil Académique | M. Hamda BEN HADID |
| Vice-Président du Conseil d'Administration | M. Didier REVEL |
| Vice-Président du Conseil des Etudes et de la Vie Universitaire | M. Philippe CHEVALLIER |
| Vice-Président de la Commission de Recherche | M. Jean-François MORNEX |
| Directeur Général des Services | M. Pierre ROLLAND |

COMPOSANTES SANTE

| | |
|---|--|
| Département de Formation et Centre de Recherche en Biologie Humaine | Directrice : Mme Anne-Marie SCHOTT |
| Faculté d'Odontologie | Doyenne : Mme Dominique SEUX |
| Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux | Doyenne : Mme Carole BURILLON |
| Faculté de Médecine Lyon-Est | Doyen : M. Gilles RODE |
| Institut des Sciences et Techniques de la Réadaptation (ISTR) | Directeur : M. Xavier PERROT |
| Institut des Sciences Pharmaceutiques et Biologiques (ISBP) | Directrice : Mme Christine VINCIGUERRA |

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

| | |
|---|---|
| Département Génie Electrique et des Procédés (GEP) | Directrice : Mme Rosaria FERRIGNO |
| Département Informatique | Directeur : M. Behzad SHARIAT |
| Département Mécanique | Directeur M. Marc BUFFAT |
| Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon) | Directeur : Gérard PIGNAULT |
| Institut de Science Financière et d'Assurances (ISFA) | Directeur : M. Nicolas LEBOISNE |
| Institut National du Professorat et de l'Education | Administrateur Provisoire : M. Pierre CHAREYRON |
| Institut Universitaire de Technologie de Lyon 1 | Directeur : M. Christophe VITON |
| Observatoire de Lyon | Directrice : Mme Isabelle DANIEL |
| Polytechnique Lyon | Directeur : Emmanuel PERRIN |
| UFR Biosciences | Administratrice provisoire : Mme Kathrin GIESELER |
| UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS) | Directeur : M. Yannick VANPOULLE |
| UFR Faculté des Sciences | Directeur : M. Bruno ANDRIOLETTI |

Résumé

La transformation naturelle est un mécanisme de transfert horizontal de gène très répandu chez les bactéries. C'est l'un des moteurs de la diversité bactérienne qui peut conduire à l'émergence de nouvelles bactéries pathogènes et de nouvelles résistances aux antibiotiques. La transformation naturelle désigne la capacité de certaines bactéries dites « transformables » à capturer de l'ADN nu exogène et de l'intégrer dans leur génome par recombinaison homologue. Elle met en jeu de nombreux acteurs protéiques regroupés sous le nom de « machinerie d'import de l'ADN ». Cette machinerie est largement conservée chez les différentes espèces transformables. Elle n'est exprimée que lorsque la bactérie est dans un état génétiquement programmé et le plus souvent transitoire qu'on appelle « la compétence ». Les conditions d'induction ainsi que les voies de régulation qui aboutissent à cet état physiologique particulier varient largement en fonction des espèces transformables. Malgré des avancées considérables ces dernières années, les situations qui stimulent la transformation naturelle restent encore mal comprises. De même, le fonctionnement de la machinerie d'import n'est que partiellement élucidé. Tous les composants ne sont pas encore identifiés et il est possible que des processus fondamentaux de la cellule participent au fonctionnement du système.

Mon premier objectif était d'identifier de façon exhaustive tous les gènes requis pour la transformation naturelle de *Legionella pneumophila*, une bactérie gram négative à multiplication intracellulaire. À cette fin, j'ai mis au point une méthode de Tn-seq chez ce pathogène. Le Tn-seq est une méthode puissante qui permet, à partir d'une banque de mutants insertionnelles, d'identifier de façon exhaustive les mutants des gènes ayant perdu une fonction donnée. Les étapes de construction de la banque de *L. pneumophila* jusqu'à l'analyse bioinformatique des données de Tn-seq, jamais réalisés jusqu'alors pour cette espèce, ont fait l'objet d'une publication technique. L'analyse des données de Tn-seq issue de cribles de transformation m'a permis de confirmer le rôle de la plupart des gènes déjà connus pour être impliqués dans la transformation naturelle chez cette espèce et a permis de préciser le rôle de pilines mineures et majeures. Cette analyse m'a surtout aussi permis d'identifier un nouvel acteur : YraN, conservé chez la majorité des espèces gram négative et jusqu'alors de fonction inconnue. La caractérisation de ce nouvel acteur par des approches génétiques et moléculaires a permis de montrer qu'YraN était une nucléase impliquée dans les voies moléculaires dédiées à la recombinaison de l'ADN transformant.

Abstract

Natural transformation is a widespread mechanism of horizontal gene transfer in bacteria. It is one of the drivers of bacterial diversity that can lead to the emergence of new pathogenic bacteria and antibiotic resistance. Natural transformation refers to the ability of certain so-called "transformable" bacteria to capture exogenous naked DNA and integrate it into their genome by homologous recombination. It involves numerous protein players grouped under the name of "DNA uptake machinery". This machinery is largely conserved in the various transformable species. It is only expressed when the bacterium is in a genetically programmed and most often transient state called "competence". The conditions of induction as well as the regulatory pathways that lead to this particular physiological state vary widely between transformable species. Despite considerable progress in recent years, the situations that stimulate natural transformation are still poorly understood. Similarly, the functioning of the import machinery is only partially elucidated. Not all components have been identified yet and it is possible that fundamental processes in the cell are involved in the functioning of the system.

My first objective was to comprehensively identify all the genes required for the natural transformation of *Legionella pneumophila*, a gram-negative bacterium which replicates intracellularly. To this end, I developed a Tn-seq method for this pathogen. Tn-seq is a powerful method that allows, from a library of insertional mutants, to exhaustively identify mutants of genes that have lost a given function. The steps from the construction of the *L. pneumophila* library to the bioinformatics analysis of the Tn-seq data, which has never been done before for this species, have been the subject of a technical publication. The analysis of Tn-seq data from transformation screens allowed me to confirm the role of most of the genes already known to be involved in natural transformation in this species and to clarify the role of minor and major pilins. This analysis also allowed me to identify a new actor: YraN, conserved in the majority of gram-negative species and previously of unknown function. The characterisation of this new actor by genetic and molecular approaches showed that YraN was a nuclease involved in the molecular pathways dedicated to the recombination of transforming DNA.

Remerciements

Premièrement je tiens à remercier, les membres du jury Nathalie Leblond-Bourget et Pablo Radicella pour leur rôle d'examineurs, et Mireille Ansaldi et Sophie Jarraud pour leur rôle de rapportrices, pour le temps que vous consacrez à l'évaluation de ce travail de thèse.

Je tiens à remercier particulièrement Xavier pour m'avoir fait confiance dès mon stage de master 2 et pendant ces 4 années de thèse [et encore un petit peu de rab en 2022 :-)]. Alors que je n'étais qu'un étudiant « lambda » venu du fin fond de ma Normandie natale, assez peu confiant en moi, tu as su croire en mon potentiel sur un simple entretien téléphonique pour le stage de master 2. A ce moment-là je crois même que tu croyais plus en moi que moi-même. D'un point de vue professionnel, vu mon niveau technique en début de stage de master 2, je peux dire que tu m'as tout appris. D'ailleurs, j'espère, à ce moment-là, ne pas t'avoir fait regretter ton choix quand tu as compris, qu'à part les PCRs, je ne savais pas faire grand-chose :-P. Mais, je crois que davantage encore que ces aspects techniques, j'ai acquis à travers nos discussions, toujours très enrichissantes, une manière particulière de « penser » la recherche, d'imaginer des manip', de concevoir des présentations ou même de rédiger. J'ai adoré me creuser les méninges pour trouver la petite astuce, parfois saugrenue, pour répondre à une question précise : on l'aura ce foutu D-loop, on l'aura ! (oui, même si la situation est désespérée, je ne désespère pas :-D). D'ailleurs, je tiens à te remercier de m'avoir permis de réaliser tout ce que j'avais envie de réaliser, de m'avoir permis d'avoir accès aux outils de dernier cri (je parle par exemple du Nanopore), moi qui suis « fan » de nouvelles technologies, je ne pouvais pas rêver mieux ! Et tout ceci va avec ton incroyable habilité à obtenir des financements : je pense que, sans toi, je n'aurais jamais obtenu ce financement FRM qui m'a permis d'obtenir cette quatrième année de thèse et donc pour ça aussi je te remercie.

D'un point de vue évolution personnelle, même s'il me reste certainement encore du chemin à parcourir, tu as su me redonner confiance et me montrer que cette timidité qui me caractérise n'est pas un frein. Je sais que je n'ai pas été le doctorant le plus jovial et que mon introversion, dans ce monde de la recherche qui demande souvent communication et sociabilité, a pu parfois être pesante, mais je te remercie vivement pour ta patience, pour m'avoir accepté tel que je suis. A tes côtés, j'ai découvert un métier, une passion et un exemple inspirant du chercheur que je souhaiterai être. Maintenant, j'appréhende beaucoup la suite, car trouver un PI et une équipe aussi bienveillante que celle-ci, ça va être compliqué :{.

Je tiens à remercier l'ensemble des membres de mon équipe à commencer par Isabelle sans qui cette thèse aurait sûrement été beaucoup moins « fun ». Pourtant, ça avait plutôt mal commencé : désolé encore d'avoir piqué ta blouse dans les premiers jours où tu es arrivée :-p. Mais passée cette carapace qui nous caractérise, nous nous sommes trouvés de nombreux points communs : j'ai encore du mal à m'en remettre quand je pense que tes neveux sont jumeaux et qu'ils portent les mêmes prénoms que mon jumeau et moi :O. C'est encore plus fou de savoir que tu étais collégienne dans le collège que mon père dirige actuellement alors même que nous ne sommes pas du tout Haut-Savoyards d'origine. Au fil de ces années, nous avons développé une belle complicité. Je te remercie pour tous ces moments de fous rires : rien que de penser à ton piège odieux « des boulettes », j'en ai les larmes (de rire et aussi de honte) aux yeux :D. Et que dire des chansons qu'on chantait à tour de rôle ? L'hymne du labo restant pour moi Emilie Jolie ou l'Aigle Noir :-D. Je te remercie aussi pour ton écoute et ces longues discussions sérieuses qu'on a pu avoir sur notre fameux banc (dont seul toi connais le surnom) qui m'ont véritablement aidées dans ma construction professionnelle et personnelle. En revanche, je ne te remercie pas de m'avoir déculpabilisé de manger autant de MnM's :D : cette thèse m'aura fait prendre quelques kg (bon j'avoue je crois que le confinement y est aussi

pour beaucoup). Je te remercie aussi pour ta confiance à toute épreuve surtout quand il s'agissait de te servir de mes pipettes ou de mes solutions :-P et aussi quand il s'agissait de me confier les milieux communs :P. Plus sérieusement, merci aussi pour toutes ces nombreuses fois où je t'ai demandé des conseils techniques auquel tu as toujours répondu de façon précise et avec pédagogie, et aussi pour toutes ces fois où tu m'as fait des petites manip. Bref, merci beaucoup pour tous ces moments passés en ta compagnie, mon éternelle voisine de bureau qui fait des étincelles ;-). N'oublie pas que le fantôme du labo te surveille.

Je tiens aussi à remercier tout particulièrement Alexandre, qui, au même titre qu'Isabelle est devenu un véritable ami. Je me souviendrais toujours de tes expressions légendaires parmi lesquelles « par tous les dieux de l'enfer et du ciel réunis » si je te cite bien, que je réutilise maintenant à mon compte. Cette expression reste dans mon esprit parce qu'avec tout ce qu'il s'est passé au labo : entre les travaux, le COVID, l'obus retrouvé, l'explosion de la bombonne de gaz, les manips qui ne marchent pas... et j'en passe, elle aura beaucoup servi à témoigner notre stupéfaction et notre colère parfois. Tu es la personne avec qui, hors labo, j'ai pu partager le plus de choses entre nos restos, nos cinémas ou encore la piscine (et le mythique Aquaboulevard), tu auras contribué à rendre cette vie lyonnaise plus agréable. Tu m'auras aussi beaucoup fait rire avec tes anecdotes de prépa, d'école véto ou encore en tant qu'enseignant que tu étais et que tu es redevenu : tu devrais écrire un livre je te dis :D !! Je te remercie aussi et surtout pour tous les services que tu as pu me rendre au labo ou hors labo : je ne savais pas à quel point on pouvait entasser autant de choses dans une petite Nissan Micra :D. Sous tes airs de « Voldemort... » pardon, de « celui dont on ne prononce le nom » comme tu te plais à le dire, tu es quelqu'un sur qui j'ai pu toujours compter et je pense que c'est ça la définition d'un vrai ami.

Je tiens aussi à remercier particulièrement Laetitia pour ses conseils et ses longues discussions très enrichissantes qu'on a pu avoir sur à peu près tout. Tu es ce genre de personne avec qui on apprend toujours quelque chose, de sciences bien sûr, mais je n'ai jamais autant parlé de géopolitiques ou de religions qu'avec quelqu'un d'autre que toi. Malgré nos caractères très différents, tu es l'une des personnes qui, au labo, m'a le plus poussé à sortir de ma timidité et d'un point de vue évolution personnelle tu ne le sais peut-être pas, mais tu m'as beaucoup aidé. Tu es aussi quelqu'un sur qui on peut compter et à toute heure, en témoigne la relecture de ma thèse pour laquelle j'ai reçu une super correction à 4H du matin :O. Même si mes westerns ne seront jamais assez beaux à tes yeux :(, merci pour tous les conseils que tu as pu me donner, même s'ils ne sont pas toujours donnés avec l'art et la manière :P, ils ont toujours eu pour but de m'améliorer. Grâce à toi, je sais maintenant que les ARNm ça s'écrit comme ça : *rocR* et les gènes comme ça : *RocR* :P. Eh oui, vu que je ne suis pas suicidaire, je jure de toujours appeler *RocC* -> *ProQ* lors du mécanisme traductionnel de la compétence constitutive de *L. pneumophila*. Allez j'arrête, car sinon tu vas casser encore une fois ton ordi :D.

Je tiens aussi à remercier Maria, pour ses précieux conseils et sa bienveillance. Même si je n'ai pas été souvent très bavard, j'ai beaucoup aimé partager ce bureau avec toi, Isabelle et Xavier que tu as animé de ta bonne humeur pendant ces 4 années de thèse. Tu as toujours su me venir en aide de façon très réactive, même à distance. Je suis ravi d'avoir fait partie du « camp » Acineto (la meilleure !) de cette équipe et d'avoir pu apprendre de tes connaissances sur cette espèce. Je te souhaite des recherches très prolifiques sur ce pathogène et en espérant qu'il soit toujours d'actualité : que la force soit avec toi avec ton projet « Ab-one ».

Je tiens aussi à remercier Kévin pour sa bonne humeur. Grâce (ou à cause de) toi, maintenant j'ai des paroles de chanson traduite en français dans ma tête que je ne connaissais même pas en anglais :D. Avec ces chansons et quelques pas de danse, tu auras vivement contribué à la bonne ambiance du labo et agrémenté mes longues séances de PSM. Merci pour tes anecdotes de soirée à l'ENS qui auront contribué à me faire vivre une vie étudiante que je n'ai pas vraiment vécu. Par contre je ne suis pas sûr que je me servais de ton crumble si toutefois tu en ramenaient un :D. J'ai l'impression qu'en termes de manip, on s'est un peu suivi et je te remercie pour cette coopération : moi j'ai récupéré quelques-unes de tes souches de double hybride, de tes primers et toi tu as récupéré quelques-unes de mes

souches flag (même si elles ne sont pas toujours bien rangées :D). Bon courage pour cette quatrième année de thèse qui, je suis sûr, t'apportera une foule de résultats !

Merci à Félix, le globe-trotter et le musicien du labo. T'inquiète c'est pas parce qu'on chante du Barbara avec Isabelle qu'on aime ça (...quoique). Merci, pour ta jovialité, ça a toujours été un plaisir de discuter et de travailler en ta compagnie. Tu m'as fait me sentir important en me demandant des conseils de manip. Désolé encore pour le DAP qui n'en était pas :D. Je te souhaite le meilleur pour la suite avec tes amibes !

Je tiens à remercier aussi Chloé, ma première (et seule) stagiaire. Je te remercie vivement pour ton implication et ton dynamisme qui ont apporté beaucoup à mes manips (et aussi à ma paillasse :D). Ce fut un plaisir d'avoir travaillé avec toi et j'espère t'avoir donné le goût pour la recherche.

Je tiens aussi à remercier Anne-Sophie, Corentin, Juliette et Margaux avec qui j'ai beaucoup apprécié travailler durant ces 4 années. Il est vrai que je n'ai pas été le doctorant le sociable, ni certainement le plus avenant de l'étage et, à regret, nous n'avons pas eu l'occasion d'avoir beaucoup de grandes discussions ensemble, mais celle qu'on a pu avoir ont toujours été bienveillantes et enrichissantes. Merci à Anne-Sophie avec qui j'ai partagé quelques manips et pour les nombreuses souches que tu as construites qui ont pu notamment me servir par la suite. Merci aussi beaucoup à toi Corentin, pour tous tes conseils sur les manips : je confirme les prot' c'est compliqué :-D. Merci à Margaux et Juliette pour votre dynamisme et votre bonne humeur, j'ai appris à vous découvrir finalement un peu tard au séminaire de la SFM, mais j'ai vraiment aimé partager ce moment à vos côtés.

Dans les autres personnes de l'étage,

Particulièrement, je tiens à remercier Claire pour sa bienveillance, mais aussi pour son implication sans égale dans le bon fonctionnement du laboratoire. Merci, d'avoir été avec Christophe les portes-paroles du labo face à ces maudits travaux. Merci aussi à toi d'avoir pris régulièrement de mes nouvelles surtout dans la période COVID qui a été très difficile pour moi avec la maladie de ma mère.

Pour la même raison, merci vivement à Anne. Tu es aussi la personne qui m'a motivé à faire ACE qui m'ont beaucoup apporté en termes de confiance en moi. J'ai d'ailleurs beaucoup aimé encadrer tes TD qui resteront mes tout premiers enseignements. Je te remercie pour ta bienveillance tout au long de ma thèse.

Je tiens aussi à remercier les normands... euh les bretons de l'équipe pardon, Annelise et Johann pour m'avoir guidé au début de ma thèse dans ce monde totalement inconnu pour moi de la biologie cellulaire. Merci pour vos conseils prodigués et merci aussi pour toutes ces discussions non scientifiques qu'on a eu : vous êtes sur quelles séries en ce moment ? :D.

Merci aussi à Nathalie, Elisabeth, Patricia et Christophe pour tout leur investissement dans la vie de l'étage. Nous n'avons pas eu souvent l'occasion de discuter ensemble, mais j'ai toujours apprécié évoluer à vos côtés.

Enfin merci à toutes les personnes que j'ai pu croiser qui ont pu m'aider de près ou de loin à certaines manips notamment Patrice Polard, Violette Morales, Christophe Ginevra, Antoine Corbin et Marie Cariou. Merci à Nathalie Poussereau pour m'avoir fait confiance pour l'enseignement des TP. Merci à tous les membres de mon comité de pilotage, Christian Lesterlin pour son rôle de tuteur et Erwan Gueguen, Zeynep Baharoglu, Rémi Fronzes et Karen Moreau pour leur rôle de jury.

Hors labo,

Je tiens vivement à remercier Olivier qui a été mon pilier pendant ces années de thèse et même bien avant. Tapis dans l'ombre, c'est probablement grâce à toi que j'ai vécu sereinement ces années qui ont pourtant été intenses et ponctuées de super moments comme de grosses galères (4 appart' en

4 ans ça fait beaucoup...). Je sais qu'il n'est pas toujours facile de vivre avec un thésard, mais tu as su faire preuve de patience et de compréhension pour ce domaine qui ne t'est pas familier. Alors, je ne vais pas épiloguer, car j'ai et j'aurais d'autres occasions de te le signifier, mais merci pour tout ! Mention spéciale à Vénice qui est notre petit rayon de soleil quand il ne s'amuse pas à fermer la page word de mon manuscrit de thèse en jouant avec le curseur de mon écran tactile :(.

Je tiens aussi à remercier ta famille et particulièrement tes parents, ta sœur Camille et son copain Thomas qui ont fait de ces weekends parisiens (et autre) de superbes weekends qui m'ont permis de me libérer régulièrement l'esprit le temps de quelques jours (et de quelques jeux :P) et qui, je suis sûr, ont fait partie intégrante de mon équilibre psychologique durant cette thèse ! Notamment, sans vous, je crois que la période du premier confinement aurait été très dure à vivre : On a pu s'insulter à distance sur Mario Kart et c'était cool :P. Je fais une petite dédicace au petit Maxou ici, qui est né cette année et qui aura donc l'âge de cette thèse s'il la regarde un jour (mais ça voudra probablement dire que je suis vieux ou mort :/).

Je tiens aussi à remercier mon frère Axel et sa copine Justine pour ces weekends « haut savoyard » qui m'ont permis de m'aérer (en plus des bonnes pizzas :D) surtout après cette ère post-Covid. Particulièrement, je tiens à remercier Justine qui m'a accompagné jusqu'au master et avec qui j'ai formé un binôme de choc lors des TP. C'est probablement parce que tu as permis de rendre plus agréables ces années d'études que j'en suis ici aujourd'hui et je t'en remercie.

Je tiens aussi à remercier mon frère jumeau Tom avec qui j'ai passé toute ma scolarité jusqu'en prépa et dont la gentille guéguerre des notes que nous avons menée a probablement abouti à là où nous en sommes tous les deux aujourd'hui. Oui, tu as réussi la prépa, oui tu as un diplôme de vétérinaire, et alors ? Moi j'ai un bac+9 :P.

Je tiens à remercier aussi mon frère Vincent et son mari Ludo qui malgré la distance, ont toujours été bienveillants à mon égard, de ces fameuses vacances en Alsace et toujours aujourd'hui. Vous avez été, pour moi, des exemples d'acceptation de soi et ça a beaucoup compté dans mon évolution personnelle.

Je tiens à remercier tous mes amis et particulièrement Audrey et Adrien qui sont probablement les meilleurs témoins d'une descente aux enfers (qu'a été la prépa) et de mon évolution aujourd'hui : qui aurait pu imaginer qu'un jour je puisse parler dans un grand amphi devant plein de monde alors que je mourrai de stress en khôlle devant une seule personne ? (bon, j'avoue je stresse toujours autant pour les séminaires, mais je gère beaucoup plus). C'est vrai qu'on ne se voit pas souvent, mais c'est toujours un plaisir de se retrouver comme si rien n'avait changé. On se posait plein de questions à l'époque, mais nous voilà tous dans une voie qui nous épanouit.

Finalement, ceux à qui je dois tout, ce sont mes parents. Je tiens à vous remercier pour m'avoir soutenu et poussé tout au long de ma scolarité malgré mes indécisions. Après une année de baigne...euh... de prépa ratée qui m'a pas mal bousculée, la « honte suprême » selon les dires de certains était de revenir à la fac. C'est pourquoi j'avais postulé et été accepté dans une école d'ingénieur dont les frais de scolarité vous auront couté un SMIC... sans que jamais je n'y mette les pieds. Après moult réflexions, j'ai décidé « oh malheur » de revenir à la fac. Je tiens aujourd'hui à vous dire que c'est certainement la décision la plus difficile, mais aussi la plus bénéfique que j'ai prise de toute ma vie. Outre le fait que financièrement, les frais de scolarité d'une école d'ingénieur privée, auraient fait de moi un étudiant endetté, ce revirement m'a permis de découvrir une passion. Sans ce revirement de dernière minute, jamais je n'aurais fait la connaissance la microbiologie et jamais probablement, je n'aurais pu découvrir ce formidable métier de chercheur.

Table des matières

| | |
|---|-----------|
| CHAPITRE I : INTRODUCTION BIBLIOGRAPHIQUE..... | 20 |
| Préambule historique | 21 |
| I. Variabilité des mécanismes de régulation de la compétence..... | 25 |
| A. Des « fenêtres » de compétence variables | 25 |
| B. Une variété des signaux inducteurs de la compétence..... | 27 |
| I.B.1. Signaux liés à l'habitat naturel | 27 |
| I.B.2. Stress environnementaux..... | 29 |
| I.B.3. Molécules signal | 31 |
| C. Diversité des mécanismes de régulation de la compétence | 31 |
| I.C.1. Activation de la compétence chez <i>S. pneumoniae</i> | 32 |
| I.C.2. Activation de la compétence chez <i>B. subtilis</i> | 34 |
| I.C.3. Activation de la compétence chez <i>V. cholerae</i> | 36 |
| II. Une machinerie de transformation conservée | 39 |
| A. Capture de l'ADN exogène | 39 |
| II.A.1. Eléments conservés des filaments impliqués dans la transformation naturelle | 40 |
| II.A.2. Spécificité des différents systèmes et leur rôle dans la capture de l'ADN exogène . | 44 |
| II.A.2.1. Le pilus de type 4..... | 44 |
| II.A.2.1.1. Généralités | 44 |
| II.A.2.1.2. Rôle du T4P dans la transformation naturelle..... | 46 |
| II.A.2.2. Le pilus de compétence..... | 49 |
| II.A.2.2.1. Généralités | 49 |
| II.A.2.2.2. Rôle dans la transformation naturelle..... | 50 |
| II.A.3. Le système de sécrétion de type 4 : le cas particulier d' <i>H. pylori</i> | 52 |
| B. Les étapes périplasmiques du processus de transformation | 54 |
| II.B.1. Transfert de l'ADNt dans le périplasme | 54 |
| II.B.2. Fragmentation de l'ADNt dans le périplasme | 57 |
| II.B.3. Transport de l'ADNt dans le cytoplasme..... | 58 |
| II.B.3.1. Rôle de ComEC | 58 |
| II.B.3.2. Les acteurs pivots du système de transformation | 61 |
| C. Les étapes cytoplasmiques du processus de transformation naturelle | 63 |
| II.C.1. Protection de l'ADNsb | 63 |
| II.C.2. Médiation de la recombinaison de l'ADNt | 66 |
| II.C.3. Recombinaison de l'ADNt..... | 72 |
| II.C.3.1. Généralités | 72 |
| II.C.3.2. Rôle de RecA dans la recombinaison homologue | 73 |
| II.C.3.3. Processing du D-loop de recombinaison de l'ADNt..... | 79 |

| | | |
|------------|--|-----------|
| III. | <i>Legionella pneumophila</i> : un modèle original de bactérie naturellement transformable | 83 |
| A. | Une bactérie d'origine environnementale et pathogène accidentel de l'homme..... | 83 |
| III.A.1. | Découverte | 83 |
| III.A.2. | Une espèce environnementale..... | 83 |
| III.A.3. | ... Et un pathogène accidentel de l'Homme | 84 |
| III.A.3.1. | Données épidémiologiques | 84 |
| III.A.3.2. | Le SST4 : principal facteur de virulence de <i>L. pneumophila</i> | 85 |
| III.A.3.3. | LetA/S : un système de régulation pleiotropique conservé de <i>L. pneumophila</i> .. | 86 |
| B. | Une bactérie transformable | 88 |
| III.B.1. | Des caractéristiques génomiques conservées..... | 88 |
| III.B.2. | ... Dissimulant un génome hautement dynamique | 89 |
| III.B.3. | Un mécanisme atypique de régulation de la compétence..... | 91 |
| | Projet de thèse..... | 94 |
| | CHAPITRE II : RESULTATS | 96 |
| I. | Mise au point d'une méthode de Transposition-sequencing chez <i>L. pneumophila</i> | 97 |
| II. | Analyse génétique exhaustive des gènes impliqués dans la viabilité et la transformation naturelle de <i>L. pneumophila</i> | 135 |
| A. | Publication..... | 135 |
| B. | Résultat complémentaire : Recherche de mutants hypercompétents dans la banque de mutant de <i>L. pneumophila</i> | 158 |
| III. | Caractérisation d'un nouvel acteur impliqué dans la transformation naturelle des bactéries gram négative..... | 160 |
| A. | Publication..... | 160 |
| B. | Expérience complémentaire 1 : Recherche d'interactants potentiels <i>in vivo</i> | 187 |
| III.B.1. | Par co-immunoprécipitation | 187 |
| III.B.2. | Par une expérience de double hybride <i>in vivo</i> | 190 |
| C. | Expérience complémentaire 2 : Détection du brin déplacé du D-loop <i>in vivo</i> | 193 |
| | CONCLUSION GENERALE ET PERSPECTIVES | 195 |
| | RÉFÉRENCES BIBLIOGRAPHIQUES | 198 |

TABLE DES FIGURES

| | |
|---|----|
| FIGURE 1 : LES 3 PRINCIPAUX MECANISMES DE THG CHEZ LES BACTERIES..... | 23 |
| FIGURE 2 : TRANSFORMABILITE DE <i>L. PNEUMOPHILA</i> PARIS AU COURS DE LA CROISSANCE..... | 26 |
| FIGURE 3 : L'EXPRESSION DU PILUS DE <i>S. ELONGATUS</i> EST CONTROLEE PAR LE CYCLE CIRCADIEN. | 27 |
| FIGURE 4 : INDUCTION DE LA COMPETENCE SOUS L'EFFET DE STRESS GENOTOXIQUES CHEZ <i>L. PNEUMOPHILA</i> | 30 |
| FIGURE 5 : ACTIVATION DE LA COMPETENCE CHEZ <i>S. PNEUMONIAE</i> | 33 |
| FIGURE 6 : SCHEMA SIMPLIFIE DES CASCADES MOLECULAIRES PERMETTANT L'ACTIVATION DE LA COMPETENCE CHEZ <i>B. SUBTILIS</i> | 35 |
| FIGURE 7 : INDUCTION DE LA COMPETENCE CHEZ <i>V. CHOLERA</i> E PAR LA CHITINE (PANEL DE GAUCHE), LA REPRESSION CATABOLIQUE (PANEL DU MILIEU) ET LE QUORUM SENSING (PANEL DE DROITE). | 39 |
| FIGURE 8 : REPRESENTATION SCHEMATIQUE DES DIFFERENTS FILAMENTS DE TYPE 4 IMPLIQUES DANS LA CAPTURE DE L'ADNT AU COURS DE LA TRANSFORMATION NATURELLE. | 40 |
| FIGURE 9 : REPRESENTATION MOLECULAIRE DE LA PILINE MAJEURE PILA DU T4P DE <i>PSEUDOMONAS AERUGINOSA</i> . (A) ET DU PILUS ASSEMBLE (B). | 41 |
| FIGURE 10 : ALIGNEMENT DE LA SEQUENCE N-TERMINAL DES PILINES. EXTRAIT DE (GILTNER ET AL., 2012). | 42 |
| FIGURE 11 : MODELE D'ASSEMBLAGE DU T4P PAR ROTATION DE L'ATPASE D'ASSEMBLAGE..... | 43 |
| FIGURE 12 : VISUALISATION PAR MICROSCOPIE ELECTRONIQUE DE FAISCEAUX DE T4P DE <i>N. MENINGITIDIS</i> PAR COLORATION NEGATIVE A L'ACIDE PHOSPHOTUNGSTIQUE. | 45 |
| FIGURE 13 : MODELE DE CAPTURE DE L'ADNT PAR <i>NEISSERIA SPP</i> | 47 |
| FIGURE 14 : LE T4P DE <i>V. CHOLERA</i> E LIE L'ADNT A DISTANCE DE LA CELLULE ET LE RAMENE A LA SURFACE. | 48 |
| FIGURE 15 : VISUALISATION DIRECT DU PILUS COM DE <i>S. PNEUMONIAE</i> PAR MICROSCOPIE ELECTRONIQUE. | 49 |
| FIGURE 16 : LE PILUS COM DE <i>S. PNEUMONIAE</i> LIE L'ADN ET SE RETRACTE. | 50 |
| FIGURE 17 : MODELE DE CAPTURE DE L'ADNT CHEZ <i>S. PNEUMONIAE</i> | 52 |
| FIGURE 18 : IMPORT DE L'ADN PAR LE SST4 COMB DE <i>H. PYLORI</i> | 54 |
| FIGURE 19 : LOCALISATION DE COMEA DANS LES CELLULES COMPETENTES..... | 56 |
| FIGURE 20 : MODELE D'ACTION DE COMEA AU COURS DE LA TRANSFORMATION NATURELLE..... | 57 |
| FIGURE 21 : TOPOLOGIE DE COMEC DE <i>B. SUBTILIS</i> (A) ET MODELE DE COMEC COMME CANAL DE L'ADNT (B). | 59 |
| FIGURE 22 : ROLE DE COMFA ET COMFC DANS LA TRANSFORMATION NATURELLE. | 62 |
| FIGURE 23 : BILAN DU ROLE DE SsbB DANS LA TRANSFORMATION DE <i>S. PNEUMONIAE</i> | 66 |
| FIGURE 24 : STRUCTURE DE DPRA..... | 68 |
| FIGURE 25 : ROLE DE DPRA DANS LES ETAPES CYTOPLASMIQUES DE LA TRANSFORMATION NATURELLE CHEZ <i>S. PNEUMONIAE</i> | 69 |
| FIGURE 26 : INTERACTION ENTRE RECA ET DPRA DURANT LE PROCESSING DE L'ADNSB CHEZ <i>S. PNEUMONIAE</i> | 71 |
| FIGURE 27 : STRUCTURE LINEAIRE (A) ET 3D DE RECA ET IMPLICATION DES DIFFERENTS DOMAINES DANS LA RECOMBINAISON HOMOLOGUE (B)..... | 74 |
| FIGURE 28 : MECANISME DE RECHERCHE D'HOMOLOGIE MEDIEE PAR RECA..... | 76 |
| FIGURE 29 : MECANISME MOLECULAIRE DE LA RECOMBINAISON HOMOLOGUE MEDIEE PAR RECA..... | 78 |
| FIGURE 30 : SCHEMATISATION DE L'EXPERIENCE DE CO-MIGRATION DE BRANCHE..... | 80 |
| FIGURE 31 : MECANISME UNIFIE DU PROCESSING DE D-LOOP AU COURS DE LA RECOMBINAISON DE L'ADNT. | 82 |

| | |
|--|----|
| FIGURE 32: ROLES DES EFFECTEURS SECRETES PAR LE SST4B DANS LE CYCLE INTRACELLULAIRE DE <i>L. PNEUMOPHILA</i> | 86 |
| FIGURE 33 : REGULATION DE LA PHASE TRANSMISSIVE CHEZ <i>L. PNEUMOPHILA</i> | 87 |
| FIGURE 34: REGULATION DE LA COMPETENCE CHEZ <i>L. PNEUMOPHILA</i> | 92 |

Abbreviations

ACES *N-(2-acetamido)-2-aminoethanesulfonic acid*
ADN Acide désoxyribonucléique
ADNdb ADN double-brin
ADNsb ADN simple-brin
ADNnc : ADN non codant
AMPc Adénosine MonoPhosphate cyclique
ARN Acide ribonucléique
ARNm ARN messenger
AYE *ACES buffered yeast extract*
CRP *cAMP receptor protein*
CSP *Competence Stimulation Protein*
CSF *Competence Stimulation Factor*
CYE *ACES-buffered charcoal yeast extract*
Dot *defect in organelle trafficking*
DU-pilus *DNA-uptake pilus*
GFP *Green fluorescent protein*
h heure
Icm *intracellular multiplication*
IPTG Isopropyl β -D-1-thiogalactopyranoside
kb kilobase
LCV *Legionella Containing Vacuole*
min minute
mL millilitre
MOI *Multiplicity of infection*
pb paire de bases
PCR *Polymerisase Chain Reaction*
RBS : *Ribosome Binding Site*
RFU Unités de Fluorescence Relative
RMP Protéine médiatrice de la recombinaison
SST2 Système de sécrétion de type II
SST4 Système de sécrétion de type IV
T4P : Type 4 pilus
THG Transfert horizontal de gène
TCS : *Two Component System*
T4P : *Type 4 Pilus*
UFC Unités formant colonies
UV Ultra-violet
 μ g microgramme

CHAPITRE I : INTRODUCTION BIBLIOGRAPHIQUE

Préambule historique

La transformation naturelle est, avec la conjugaison et la transduction l'un des trois principaux mécanismes de transfert horizontal de gène (THG) chez les bactéries. Ces mécanismes de THG permettent la transmission du matériel génétique chez des bactéries qui n'ont pas de liens de parenté avec leur descendant, ce qui implique que ces transferts peuvent se faire entre espèces phylogénétiquement éloignées (Soucy et al., 2015). Si le « sexe » est défini comme un processus au cours duquel les matériels génétiques de deux individus se retrouvent dans une seule et même cellule au sein de laquelle ils peuvent recombinaison pour former de nouveaux allèles, alors les THG peuvent être considérés comme des modes parasexuels chez les bactéries (Narra and Ochman, 2006). Seulement, contrairement aux autres organismes, le « sexe » chez les bactéries est unidirectionnel et implique le plus souvent une petite partie du génome du donneur et tout ou partie du génome du receveur.

Avant même qu'on parle de THG ou de « sexe » chez les bactéries, la toute première évidence d'acquisition de traits phénotypiques entre bactéries non apparentées a été révélée par Frédéric Griffith en 1928 au cours d'une célèbre expérience réalisée chez la bactérie pathogène *Streptococcus pneumoniae* (Griffith, 1928). Peu de temps après avoir injecté à une souris, une souche non virulente dite « R » (pour « Rough », d'aspect rugueux sur boîte) en même temps qu'une souche virulente et encapsulée dite « S », (pour « Smooth », d'aspect lisse sur boîte), mais inactivée par la chaleur, Griffith constata la mort de la souris et la présence, dans son sang, de bactéries dites « S ». Il en conclut que la bactérie « R » avait acquis les caractéristiques phénotypiques de la bactérie « S », il désigna ce processus sous le terme de « transformation » et suggéra la présence « d'un principe transformant » présent dans le sang infecté à l'origine de ces nouveaux traits acquis par la souche « R ».

Quelques années plus tard, en 1944, les travaux de Griffiths furent repris par Oswald Avery (Avery et al., 1944). Au lieu d'injecter à une souris un mélange de la souche « S » de *S. pneumoniae* inactivée et de la souche « R », il injecta ce qu'il appelait à l'époque « la substance transformante » de cette souche « S » et obtint les mêmes résultats que Griffiths. Par analyse physicochimique, il s'aperçut que cette « substance transformante » était en réalité l'ADN de la bactérie et conclut son expérience en proposant que « l'acide nucléique de type désoxyribose est l'unité fondamentale du principe transformant du Pneumococcus Type III » (Avery et al., 1944). Il venait là de montrer que l'ADN dont l'existence était connue depuis 1869 (Dahm, 2005), est le support de l'information génétique et non les protéines comme ce qui était suggéré à l'époque notamment par Griffith lui-même. Ces expériences marquent en quelque sorte le point de départ de la génétique moderne.

On a commencé à véritablement parler de l'existence de « sexe » chez les bactéries à la fin des années 1940s grâce à une expérience menée par Joshua Lederberg et Edward Tatum (Tatum and Lederberg, 1947). Ils cultivèrent deux souches de *E. coli* : une souche A [*met*⁻ *bio*⁻ *thr*⁺ *leu*⁺ *thi*⁺] nécessitant obligatoirement la méthionine et la biotine pour croître et une autre souche B [*met*⁺ *bio*⁺ *thr*⁻ *leu*⁻ *thi*⁻] nécessitant obligatoirement la thréonine, la leucine, et la thiamine pour croître, dans un même milieu liquide contenant l'ensemble des nutriments requis pour les deux espèces. Après étalement de la souche A, ou de la souche B seule ou du mélange souche A/souche B sur un milieu minimum, ils s'aperçurent que seules les bactéries issues du mélange souche A/souche B étaient capables de croître sur un tel milieu à une fréquence de 1 bactérie sur 10⁷ cellules étalées (Griffiths et al., 2000). Puisqu'il faut les caractéristiques des deux souches pour croître sur un tel milieu, ils suggérèrent que ces bactéries étaient issues de formes de recombinaisons génétiques du génome de la souche A et de la souche B qui permettaient de reconstituer une souche WT [*thr*⁺ *leu*⁺ *thi*⁺ *met*⁺ *bio*⁺]. Ils comparèrent ce processus de recombinaison génétique à « une sorte de stade sexuel » au cours duquel, tout comme les organismes sexués des recombinaisons génétiques pouvaient se produire. Quelques années plus tard, en prouvant que ce processus impliquait un contact direct entre cellules (Davis, 1950) et qu'il était unidirectionnel (Wollman et al., 1956), on s'aperçut que le mécanisme « sexuel » décrit par Lederberg et Tatum était un nouveau mécanisme de ce qu'on n'appelait pas encore THG à l'époque et qu'on dénomma « conjugaison ». Aujourd'hui on définit la conjugaison comme un mécanisme de THG au cours duquel, l'ADN est transféré par contact direct entre les cellules et l'établissement de ce qu'on appelle un pilus sexuel (von Wintersdorff et al., 2016) (Figure 1c).

Voulant reproduire ces expériences chez *Salmonella typhimurium* pour montrer sa capacité à conjuguer, Joshua Lederberg et Norton Zinder découvrirent par hasard le troisième principal mécanisme de THG : la transduction (Zinder and Lederberg, 1952). Ils cultivèrent en milieu minimum séparément une souche *phe*⁻ *trp*⁻ *tyr*⁻, et une autre souche *met*⁻ *his*⁻ mais, cette fois-ci, dans les deux compartiments respectifs d'un tube de culture particulier où deux compartiments sont séparés par un filtre possédant une ouverture trop petite pour laisser passer les bactéries d'un compartiment à un autre. Dans ces conditions, après étalement de la culture sur milieu minimum, de façon surprenante, ils observèrent des colonies à des fréquences similaires à ce qui avait été observé chez *E. coli*. En faisant varier la taille de l'ouverture présente dans le filtre, ils s'aperçurent que la taille minimale pour laquelle ils observèrent des colonies était de l'ordre de la taille d'un phage (P22) préalablement découvert chez *S. typhimurium*, ils en conclurent que c'était ce phage qui était responsable de ce transfert génétique et ils nommèrent ce mécanisme : « la transduction ». La transduction est aujourd'hui définie comme un mécanisme de THG au cours duquel l'ADN est transféré à la cellule receveuse par un phage ayant préalablement infecté une souche donneuse. Il existe deux types de transductions : la transduction

généralisée au cours de laquelle un phage lytique encapsule accidentellement l'ADN de la souche donneuse ou la transduction spécialisée au cours de laquelle un prophage, en s'excisant du génome de façon impropre, encapsule un génome hybride (phage-ADN bactérien) (von Wintersdorff et al., 2016) (Figure 1b).

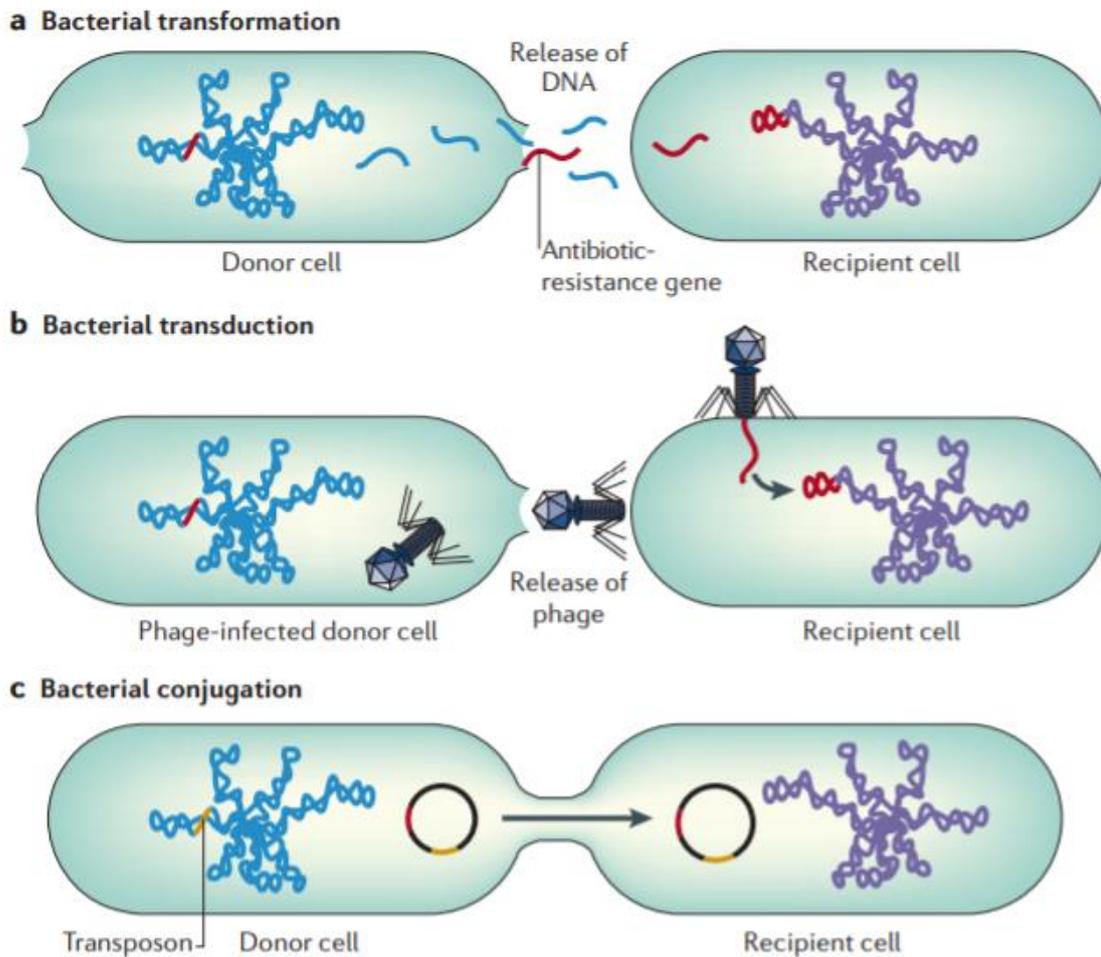


Figure 1 : Les 3 principaux mécanismes de THG chez les bactéries.

Le concept de « Transfert Horizontal de gène » ou de « transfert latéral de gène » n'est apparu que bien plus tard, dans les années 1990 avec l'essor de la phylogénie moléculaire (Hilario and Gogarten, 1993). Des analyses phylogénétiques ont révélé une complexité sous-estimée de l'évolution des génomes microbiens (Koonin, 2016; Soucy et al., 2015). On s'est aperçu que les arbres phylogénétiques générés pour des protéines appartenant à la même espèce microbienne étaient incongruents et très vite certains scientifiques ont évoqué les transferts de gènes, bien caractérisés à cette époque, comme étant à l'origine de ces arbres conflictuels (Hilario and Gogarten, 1993). Ce constat a été renforcé par la suite, avec l'évolution des techniques de séquençage et de la génomique qui ont permis de comparer plus largement les génomes. De grands débats ont eu lieu entre les

partisans en faveur d'un rôle majeur et ceux d'un rôle mineur des THG dans l'évolution des génomes (Koonin, 2016). Le point culminant de ces débats aurait eu lieu en 2006 d'après Koonin (2016), lorsque le premier « Arbre de la vie » a été généré automatiquement (Ciccarelli et al., 2006). Il fut considéré comme une avancée considérable en matière de phylogénie par certains pour être directement qualifié « d'arbre des 1% » par d'autres considérant qu'il ne reflétait qu'une petite partie de l'histoire évolutive des gènes (Koonin, 2016). Aujourd'hui, on représente plutôt « l'arbre de la vie » comme un arbre dont les branches définissent les transferts verticaux de gènes, et les ramifications, les THG (Doolittle, 1999).

Aujourd'hui, même si les arbres phylogénétiques sont toujours autant utilisés notamment, car ils définissent « une tendance de l'évolution verticale des gènes » (Koonin, 2016), on sait que les THG jouent un rôle majeur dans l'évolution des génomes.

C'est notamment l'une des hypothèses majeures quant au rôle de la transformation naturelle, le mécanisme de THG que j'étudie durant ma thèse (Johnston et al., 2014). Ce mécanisme permet à une bactérie dite transformable de capturer de l'ADN nu exogène et de l'intégrer dans son génome par recombinaison homologue (**Figure 1a**). C'est un mécanisme très répandu chez les bactéries puisqu'on dénombre à ce jour plus de 80 espèces expérimentalement transformables bien réparties dans les différents clades bactériens (Johnston et al., 2014). Elle a principalement été étudiée chez les bactéries gram positives *S. pneumoniae* et *Bacillus subtilis* et chez les bactéries gram négatives *Neisseria gonorrhoeae*, *Vibrio cholerae* et *Helicobacter pylori*. C'est le seul des trois principaux mécanismes de THG qui est entièrement régulé par la bactérie receveuse lors d'un processus génétiquement programmé qu'on appelle la compétence. Lorsque les bactéries sont compétentes, elles expriment ce qu'on appelle « la machinerie d'import de l'ADN » (i.e. *DNA uptake system*, en anglais) qui permet la capture de l'ADN exogène, son transport dans le périplasma, puis son internalisation dans le cytoplasme et sa recombinaison au sein de chromosome.

Comme nous allons le voir de façon plus détaillée dans cette partie introductive, les voies de régulation de la compétence sont très divergentes au sein des différentes espèces transformables (I) tandis qu'à l'inverse, les voies moléculaires qui mènent à la recombinaison de l'ADNt sont très conservées au sein de ces différentes espèces (II).

I. Variabilité des mécanismes de régulation de la compétence

Les bactéries dites « naturellement transformables » sont des bactéries, qui dans des conditions naturelles ou dans des conditions reproduites en laboratoire expriment l'ensemble des protéines leur permettant de capturer l'ADN exogène et l'intégrer dans leur génome par recombinaison homologue (Blokesch, 2016). Cette capacité de transformation est codée génétiquement lorsque ces bactéries sont dans un état physiologique programmé et le plus souvent transitoire qu'on appelle la compétence. En cela, la transformation naturelle se distingue de la transformation artificielle qui est une technique très utilisée en biologie moléculaire permettant par des procédés physiques, chimiques ou électriques de perméabiliser les membranes bactériennes pour permettre l'internalisation de plasmides par exemple (Ren et al., 2019).

Les conditions et mécanismes par lesquels les bactéries deviennent compétentes sont dépendant de la phase de croissance, de facteurs environnementaux et de voies de régulation qui divergent largement d'une espèce à une autre (Johnston et al., 2014). Dans cette partie, je vais développer ces différents aspects à travers l'étude de divers organismes modèles transformables.

A. Des « fenêtres » de compétence variables

La grande majorité des espèces bactériennes développent leur compétence de façon transitoire durant, ce que Johnston et al. dans leur revue de 2014, ont appelé « une fenêtre d'opportunité ». En fonction des espèces, cette fenêtre durant laquelle elles expriment leur machinerie de transformation est très variable au cours de leur phase de croissance. Généralement, pour connaître l'influence de la phase de croissance sur la compétence, la bactérie est placée dans des conditions de transformation appropriées et les fréquences de transformation (ratio entre le nombre de transformants et le nombre total de cellules) sont calculées au cours du temps. Pour la plupart des bactéries, la « courbe de compétence » ainsi obtenue a la forme d'un pic plus ou moins élargi. Cela signifie qu'au cours du temps, il y a une augmentation plus ou moins progressive de la proportion de cellules compétentes au sein de la population jusqu'à atteindre un maximum (« pic de compétence »), suivi d'une diminution, là aussi plus ou moins progressive de cette proportion. Le début et la fin de ce pic correspondent respectivement à l'entrée et à la sortie de la phase de compétence. Alternativement, une fusion rapportrice d'un gène de transformation peut être utilisée pour tracer cette courbe. Cette courbe de compétence peut ensuite être superposée à la courbe de croissance des bactéries pour identifier la période de la croissance durant laquelle les bactéries sont compétentes (voir **Figure 2**, un exemple de courbe de compétence obtenue pour *L. pneumophila*).

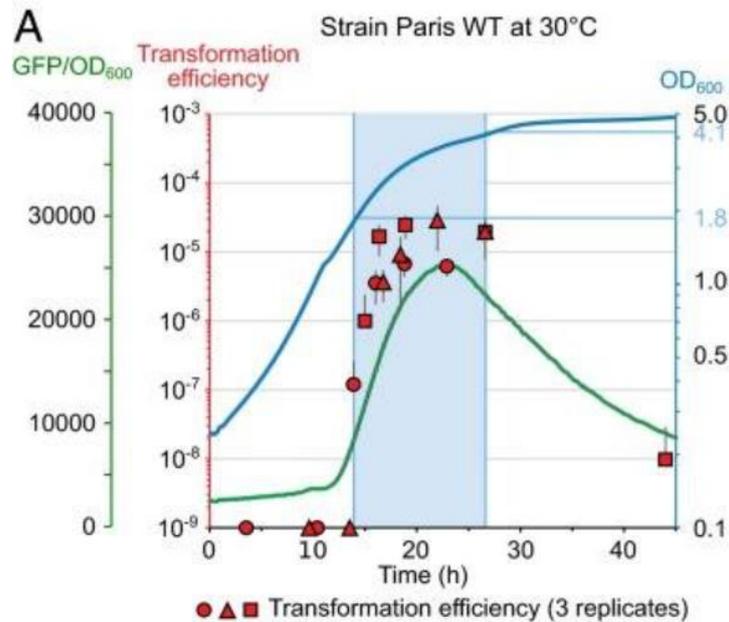


Figure 2 : Transformabilité de *L. pneumophila* Paris au cours de la croissance.

D'après (Attaiech et al., 2016). Courbe bleue : croissance en fonction du temps (lecture de la DO à 600nm). Courbe verte : fluorescence par cellule en fonction du temps (les cellules portent la fusion rapportrice comEA-gfp). Points rouge : transformabilité de *L. pneumophila* à différents temps. Bande bleue : fenêtre de compétence de *L. pneumophila*

Certaines bactéries sont dites constitutivement transformables ou constitutivement compétentes lorsqu'elles expriment à toutes les phases de la croissance et dans n'importe quelles conditions, la machinerie de transformation. C'est le cas par exemple des *Neisseria spp* (Sparling, 1966; Treangen et al., 2008), *Acinetobacter baylyi* (Cruze et al., 1979; Ellison et al., 2021) ou encore d'*Helicobacter pylori* (Dorer et al., 2010; Seitz and Blokesch, 2013a), cependant chez ces deux dernières, l'existence de variations dans la courbe de compétence suggère une régulation sous-jacente de sa compétence (Baltrus and Guillemin, 2006).

La plupart des bactéries deviennent compétentes durant leur phase exponentielle. C'est le cas par exemple de *S. pneumoniae* [durant une courte période au début de la phase exponentielle (Tomasz and Hotchkiss, 1964)], *Acinetobacter baumannii* [(Vesel and Blokesch, 2021) après 90 min de croissance en milieu riche], *Xylella fastidiosa* [après l'entrée en phase exponentielle (Kung and Almeida, 2011)], ou encore *L. pneumophila* [cf **Figure 2** (Attaiech et al., 2016; Sexton and Vogel, 2004)]. Au contraire, certaines bactéries développent leur compétence plus tardivement à l'exemple de *Bacillus subtilis* ou *Haemophilus influenzae* qui deviennent compétentes à la transition entre la phase exponentielle et la phase stationnaire (Dubnau, 1991; Rj, 1991).

Par ailleurs il existe des bactéries qui développent leur compétence indépendamment de la phase de croissance, mais sous l'effet de certains inducteurs, c'est le cas de *Vibrio cholerae* avec la chitine

(voir ci-après) (Meibom et al., 2005) ou de la cyanobactérie *Synechococcus elongatus* dont la compétence ne dépend pas de la croissance, mais du cycle circadien (Taton et al., 2020).

Bien qu'il y ait une corrélation notable entre la phase de croissance et la phase de compétence, ce n'est pas pour autant la croissance qui induit la compétence chez ces espèces. En effet de nombreux autres processus dépendent de la croissance et de la densité cellulaire tels que le quorum sensing ou la limitation en nutriments, et ce sont potentiellement ces processus qui sont les véritables inducteurs de la compétence.

B. Une variété des signaux inducteurs de la compétence

I.B.1. Signaux liés à l'habitat naturel

Les bactéries sont retrouvées dans tous les écosystèmes de la planète du plus simple au plus extrême. Pendant des millions d'années, elles ont été soumises aux signaux environnementaux propres à leur habitat et ont développé des stratégies d'adaptation dédiées pour y répondre. L'une de ces stratégies est la compétence naturelle.

Le rythme circadien. Un exemple original de bactéries induisant la compétence sous l'effet de conditions naturelles particulières sont les cyanobactéries. Les cyanobactéries sont considérées comme étant les premiers organismes présents sur terre. Ce sont les seuls organismes procaryotes capables d'effectuer la photosynthèse oxygénique et sont à ce titre tenus pour responsables de l'oxygénation de l'atmosphère et des océans durant la Grande Oxygénation, il y a 2,4 milliards d'années, aboutissant à l'émergence de vies plus complexes sur Terre (Demoulin et al., 2019). De ce fait, la source principale d'énergie des cyanobactéries est la lumière et leur métabolisme cellulaire est largement contrôlé par un cycle circadien endogène (Johnson et al., 2008). La capacité des cyanobactéries à transformer est connue depuis les années 1970 (Shestakov and Khyen, 1970) et l'espèce la plus étudiée pour sa transformabilité est *S. elongatus* (Porter, 1986). Récemment, une

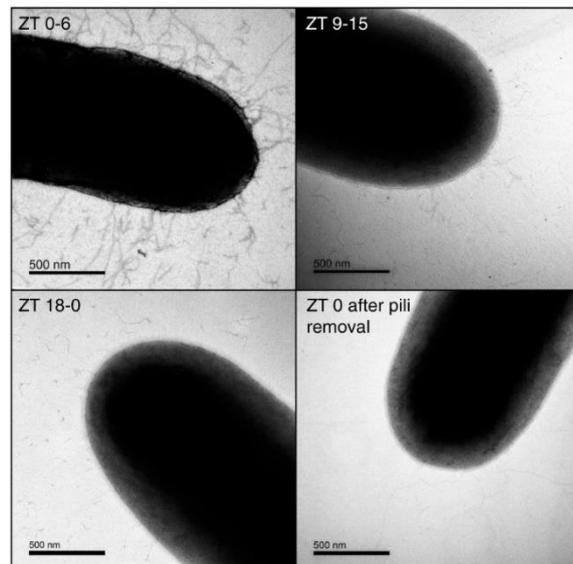


Figure 3 : L'expression du pilus de *S. elongatus* est contrôlée par le cycle circadien.

ZT (Zeitgeber Time) : Temps en heure écoulé depuis l'allumage de la lumière. Entre 0 et 6H (ZT 0-6, considéré comme étant le matin) les bactéries présentent de nomb

méthode de *Random Barcoding Transposon-sequencing* (RB-Tnseq) associée à une analyse transcriptomique a permis d'identifier les gènes essentiels à la transformation de *S. elongatus* et de montrer que sa compétence était régulée par le cycle circadien (Taton et al., 2020). En effet, il a été montré que la plupart des gènes codant pour le système de pilus essentiel à la capture de l'ADN exogène (voir II.A) étaient exprimés quotidiennement le matin (**Figure 3**), alors que les gènes codant par exemple pour ComEA, ComEC, ComF ou DprA essentiels à l'internalisation et/ou la recombinaison de l'ADNt étaient induits par l'obscurité et exprimés au crépuscule. En entraînant les bactéries à des cycles de jours plus ou moins longs, il a été aussi montré que cette régulation de la compétence pouvait s'adapter aux changements saisonniers de la durée des jours (Taton et al., 2020).

La chitine. L'interaction de *V. cholerae* avec son environnement est un bon exemple de l'incidence que peut avoir le milieu naturel sur l'induction de la compétence. *V. cholerae* est une bactérie aquatique, que l'on retrouve le plus souvent dans les régions côtières ou les rivières, associée ou non à des crustacés (Matthey and Blokesch, 2016). Alors qu'elle est étudiée depuis des décennies pour être l'agent pathogène responsable du choléra, une maladie diarrhéique qui tue 100,000 personnes dans le monde chaque année (Jahan, 2016), ce n'est que depuis quelques années seulement que l'on connaît sa capacité de transformation grâce à la découverte du principal inducteur de sa compétence : la chitine (Meibom et al., 2005). La chitine est le polymère aminopolysaccharide le plus abondant dans la nature, car il fait partie, entre autres, de l'exosquelette des crustacés ou des insectes (Elieh-Ali-Komi and Hamblin, 2016). En utilisant une méthode de microarray pour identifier les gènes activés lors de l'interaction de *V. cholerae* avec une carapace chitineuse de crabe, Meibom et al. (2005) se sont aperçus que la plupart des gènes codant pour le pilus de type 4 (T4P) de cette bactérie étaient induits dans ces conditions. En plus de déterminer les conditions inductrices de la compétence chez cette espèce, ils venaient là de montrer pour la première fois que *V. cholerae* était transformable. Depuis, expérimentalement, la chitine est systématiquement ajoutée au milieu de transformation pour induire la compétence de *V. cholerae* qui sert même d'organisme modèle pour l'étude de la transformation naturelle chez les bactéries à gram négatif (Blokesch, 2012).

Les facteurs d'hôte. *X. fastidiosa* est une bactérie phytopathogène responsable notamment de la maladie de Pierce des vignes ou de la chlorose panachée des agrumes. Depuis 1880, cette bactérie a détruit plus de 35,000 hectares de vignes en Californie du Sud, mais aussi dans d'autres régions du monde et a donc un impact économique important (Chatterjee et al., 2008). Pour cette raison elle est classée comme organisme nuisible. Son mode d'action est encore peu connu, mais elle formerait des biofilms à l'intérieur des vaisseaux de xylèmes des plantes qu'elle infecte, provoquant une obstruction des vaisseaux et un assèchement des parties aériennes de la plante aboutissant à la mort de celle-ci (Chatterjee et al., 2008). Elle est transmise d'une plante à une autre par des insectes se nourrissant de

la sève du xylème telle que les cicadelles (Chatterjee et al., 2008). Récemment, *X. fastidiosa* a été décrite comme naturellement compétente et il a été suggéré que cette habilité à échanger de l'ADN lui permettrait d'augmenter sa diversité génétique et son spectre d'hôte (Kung and Almeida, 2011). Récemment, il a été observé que *X. fastidiosa* transformait à de hautes fréquences en chambre microfluidique alimentée avec de la sève de vigne. Ce qui suggère que cet élément issu de son environnement naturel s'il n'induit pas forcément sa compétence a une incidence dessus (Kandel et al., 2016). Il aussi été montré qu'un autre phytopathogène, *Ralstonia solanacearum*, devenait compétent au cours de l'infection de son hôte (Bertolla et al., 1999).

I.B.2. Stress environnementaux

La plupart des bactéries de par leur mode de vie sont directement soumises aux fluctuations environnementales qui peuvent menacer leur survie. Ces stress environnementaux peuvent être des limitations en nutriments ou peuvent provoquer des dommages au niveau de leur ADN, ce qui peut être synonyme de mort. Pour faire face à ces conditions adverses, les bactéries sont capables de sentir leur environnement et de développer des stratégies d'adaptation en modulant un grand nombre de voies moléculaires dont fait partie la compétence.

Stress génotoxique. Deux études menées chez *S. pneumoniae* et *L. pneumophila* (plus longuement étudiée en III) ont montré que les antibiotiques tels que les fluoroquinolones, les ultraviolets ou encore la mitomycine C induisaient la compétence chez ces bactéries (Charpentier et al., 2011; Prudhomme et al., 2006). Par exemple chez *L. pneumophila*, il a été montré que l'expression d'une fusion ComEA-GFP était induite par la norfloxacine, la mitomycine C ou la bicyclomycin (et par les UV) (**Figure 4**). Ces facteurs de stress génotoxiques provoquent des dommages à l'ADN sous la forme de cassures double brin, en agissant sur la gyrase répllicative pour les fluoroquinolones ou par action directe sur l'ADN pour les UV ou la mitomycine C. En réponse à ces lésions de l'ADN la plupart des bactéries induisent la réponse SOS. La réponse SOS est un réseau de régulation qui permet aux bactéries de survivre face à ces stress génotoxiques en réarrangeant et en mutant leur génome. En plus de permettre la survie des cellules, la réponse SOS augmente la diversité des génomes par divers mécanismes (McKenzie et al., 2000). La transformation naturelle contribue aussi à ces fonctions de recombinaison et de diversification. *L. pneumophila* et *S. pneumoniae*, sont dépourvus de systèmes SOS. Au vu du fait que les stress génotoxiques induisent la compétence chez ces bactéries, il a été supposé que chez ces bactéries privées de réponse SOS, la transformation naturelle aussi montrée comme participant à la diversification et/ou à la recombinaison des génomes pouvait remplacer la réponse SOS, en utilisant les ADN donneurs pour réparer les génomes (Charpentier et al., 2012).

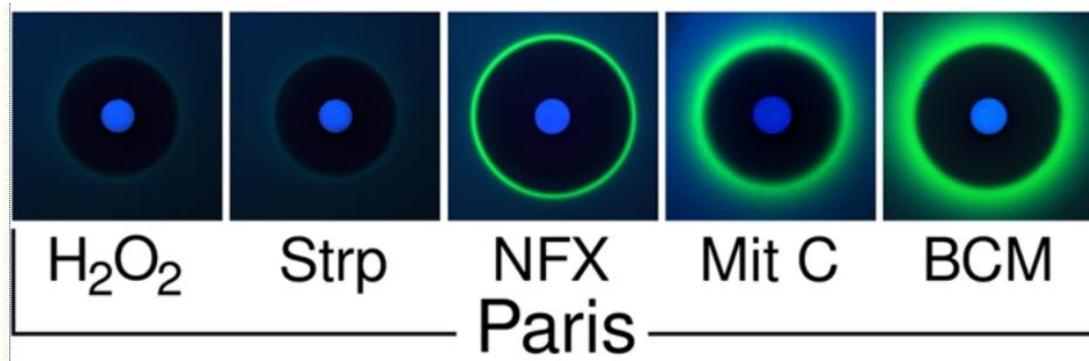


Figure 4 : Induction de la compétence sous l'effet de stress génotoxiques chez *L. pneumophila*. Extrait de (Charpentier et al., 2011). Différentes molécules ont été testées pour leur capacité à induire l'expression de ComEA-GFP chez la souche Paris de *L. pneumophila* rapporteur de l'induction de la compétence. La norfloxacine, la mitomycine C et la bicyclomycin induisent la compétence chez *L. pneumophila* (visible par un halo fluorescent vert caractéristique d'une expression de la GFP) alors que H_2O_2 et la streptomycine ne l'induisent pas.

Carence en nutriments. La disponibilité en nutriments est un facteur récurrent de l'induction de la compétence (Johnston et al., 2014). Par exemple, chez *B. subtilis*, *H. influenzae* ou *Actinobacillus pleuropneumoniae* la compétence est induite à l'entrée en phase stationnaire (ou à la transition entre la phase exponentielle et la phase stationnaire) (Bossé et al., 2009; Dubnau, 1991; Rj, 1991). Chez les bactéries l'entrée en phase stationnaire étant généralement considéré comme un ralentissement de la croissance dû à un appauvrissement des nutriments dans le milieu, il est suggéré que chez ces espèces cette entrée tardive en compétence est due à cet appauvrissement des ressources. Chez *H. influenzae*, ce qui vient corroborer cette hypothèse est que sa compétence est aussi induite lorsqu'on la bascule d'un milieu riche à un milieu pauvre en nutriments (Herriott et al., 1970). Cette induction de la compétence lorsque les bactéries sont placées dans un milieu minimal a aussi été observée chez *X. fastidiosa* (Kung and Almeida, 2011) *R. solanacearum* (Bertolla et al., 1997) ou encore *Micrococcus luteus* (Lichev et al., 2019). Chez *V. cholerae*, même si la situation est plus complexe, il a été aussi récemment montré que sa compétence était induite en l'absence de glucose dans un mécanisme faisant intervenir la répression catabolique (Blokesch, 2012). En accord avec cette capacité des bactéries à induire la compétence dans des milieux appauvris en nutriments, certains chercheurs ont suggéré que l'ADN acquis par transformation pouvait servir de nourriture pour la bactérie et ont ainsi établi une nouvelle hypothèse quant au rôle de la transformation : l'hypothèse de « *DNA for food* » (i.e, l'ADN comme nourriture) (Redfield, 1993). Un argument en faveur de cette hypothèse est la découverte récente que chez *H. influenzae* et des espèces proches telles que *A. pleuropneumoniae* et *Actinobacillus suis*, la compétence est induite suite à un appauvrissement en purines intra et extracellulaires (Sinha et al., 2013). Les auteurs suggèrent que l'internalisation de l'ADN via transformation pourrait permettre, après dégradation de cet ADN, de combler ce manque en purine. Seulement, l'un des arguments en défaveur de cette théorie est que chez certaines bactéries

la limitation en nutriment ne semble pas être corrélée à une induction de la compétence comme c'est le cas de *S. pneumoniae* qui est compétente uniquement au début de la phase exponentielle, en milieu riche, là où les nutriments sont les plus abondants (Johnston et al., 2014).

I.B.3. Molécules signal

Les bactéries sont capables de communiquer entre elles grâce à un langage bien particulier appelé « quorum sensing ». Le quorum sensing est défini comme « un processus de communication entre cellules dans lequel les bactéries utilisent la production et la détection de substances chimiques extracellulaires appelées autoinducteurs pour contrôler la densité de la population cellulaire » (Ng and Bassler, 2009). Comme chez les organismes plus complexes, ces autoinducteurs sont considérés comme une sorte d'hormone permettant de synchroniser l'expression des gènes de l'ensemble de la population suite à une altération de celle-ci (Waters and Bassler, 2005). Le quorum sensing induit des cascades de régulation qui contrôlent un large éventail de processus physiologiques telles que la symbiose, la mobilité, la sporulation, la conjugaison, la formation de biofilm et ce qui nous intéresse ici : la compétence (Miller and Bassler, 2001). Les autoinducteurs sont des petites molécules diffusibles sécrétées par les bactéries ; les gram négatives utilisent en général de l'homosérine lactone acétylée comme autoinducteur alors que les bactéries gram positives utilisent de petits peptides préalablement processés pour communiquer (Miller and Bassler, 2001). Ces autoinducteurs sont fréquemment utilisés par les bactéries pour induire la compétence (Johnston et al., 2014).

C. Diversité des mécanismes de régulation de la compétence

La plupart des bactéries dont les voies de régulation de la compétence sont connues ont recours à ce qu'on appelle un « régulateur central de la compétence » qui régule directement le régulon de compétence (i.e, les gènes requis pour la transformation). Il peut s'agir de facteurs sigmas alternatifs, d'activateurs transcriptionnels ou des co-régulateurs transcriptionnels (Johnston et al., 2014).

Chez *B. subtilis* et *S. pneumoniae*, les cascades de régulation qui aboutissent à l'expression des gènes de compétences sont assez similaires (Claverys et al., 2006). Chez ces deux espèces, la compétence est induite par un peptide (*Competence Stimulated Peptide* (CSP) pour *S. pneumoniae* et ComX pour *B. subtilis*) qui est détecté par le senseur du système à deux composantes (ComD pour *S. pneumoniae* et ComP pour *B. subtilis*) qui transmet le signal à un régulateur de réponse (ComE pour *S. pneumoniae* et ComA pour *B. subtilis*). Ces régulateurs de réponse sont notamment impliqués dans l'expression du gène codant pour un régulateur central de la compétence spécifique à chaque espèce (σ_x pour *S. pneumoniae* et ComK pour *B. subtilis*). Ces régulateurs centraux de la compétence induisent

à leur tour l'expression de gènes qui codent notamment pour la machinerie d'import de l'ADNt. Il existe néanmoins des spécificités inhérentes à chacun des modèles bactériens.

I.C.1. Activation de la compétence chez *S. pneumoniae*

S. pneumoniae est un résidant commun des voies respiratoires supérieures de la plupart des enfants de moins de 5 ans et de certains adultes où il est impliqué dans une relation commensale (Weiser et al., 2018). Mais c'est aussi un pathogène opportuniste responsable d'un large éventail de maladies. Notamment, c'est la principale cause de pneumonie sévère dans le monde entraînant la mort de près de 1 000 000 enfants chaque année (O'Brien et al., 2009). Malgré l'existence de vaccins et de thérapies antibiotiques dirigées contre cette bactérie, *S. pneumoniae* continue d'être un problème majeur de santé publique. Ceci s'explique notamment par l'incroyable plasticité de son génome médiée par de hautes fréquences de recombinaison via THG qui lui permettent de développer des résistances aux antibiotiques et des mécanismes d'échappement aux vaccins (Straume et al., 2015). La transformation naturelle est le principal HGT utilisé par *S. pneumoniae* pour cela. Bien que *S. pneumoniae* ne soit pas constitutivement transformable contrairement à *Neisseria spp* ou *H. pylori* qui sont deux bactéries très recombinogènes, elle transforme à de très hautes fréquences de transformation, ce qui s'explique notamment par le fait que dans une culture à croissance exponentielle près de 100% des bactéries deviennent compétentes simultanément (Claverys et al., 2006).

Chez *S. pneumoniae*, la compétence est principalement régulée par un mécanisme apparenté au « quorum sensing ». Seulement, on ne peut pas à proprement parler de « quorum sensing », mais plutôt de « communication cellulaire », car l'expression des gènes à l'échelle de la population bien qu'induite par des autoinducteurs n'est pas reliée à la densité cellulaire, mais plutôt à des stimuli environnementaux (Moreno-Gámez et al., 2017). Ces autoinducteurs nommés CSPs sont de nature peptidique et peuvent être qualifiés de phéromones (Johnston et al., 2014). *S. pneumoniae* produit deux types de CSP : CSP-1 et CSP-2 (1 type de CSP par souche) qui diffèrent par leurs séquences primaires alors que les autres espèces de ce genre tel que *Streptococcus mitis* en produisent une multitude qui sont classés en phérotypes (Steinmoen et al., 2002).

La **Figure 5** présente les voies moléculaires permettant l'induction de la compétence par ces CSPs chez *S. pneumoniae*. Ces CSPs sont d'abord produits sous la forme de précurseurs pré-CSP (ComC) codés par le gène *comC* qui appartiennent à l'opéron *comCDE*. Cet opéron code en plus pour un récepteur histidine kinase ancré dans la membrane plasmique ComD et son régulateur de réponse ComE. ComC est sécrétée puis processée par le transporteur de la membrane plasmique ComAB codé

par l'opéron *comAB* pour former le peptide mature. Cette CSP induit la compétence lorsqu'il atteint une concentration critique dans le milieu extracellulaire. À cette concentration, ComD perçoit le signal et phosphoryle le régulateur de réponse ComE grâce à son activité histidine kinase. ComE agit alors comme un activateur transcriptionnel en se fixant au niveau des promoteurs des gènes de compétence notamment, *comCDE* ou *comAB* eux-mêmes, pour un rétrocontrôle positif ou au niveau du promoteur du gène codant le régulateur central de compétence ComX (Lee and Morrison, 1999).

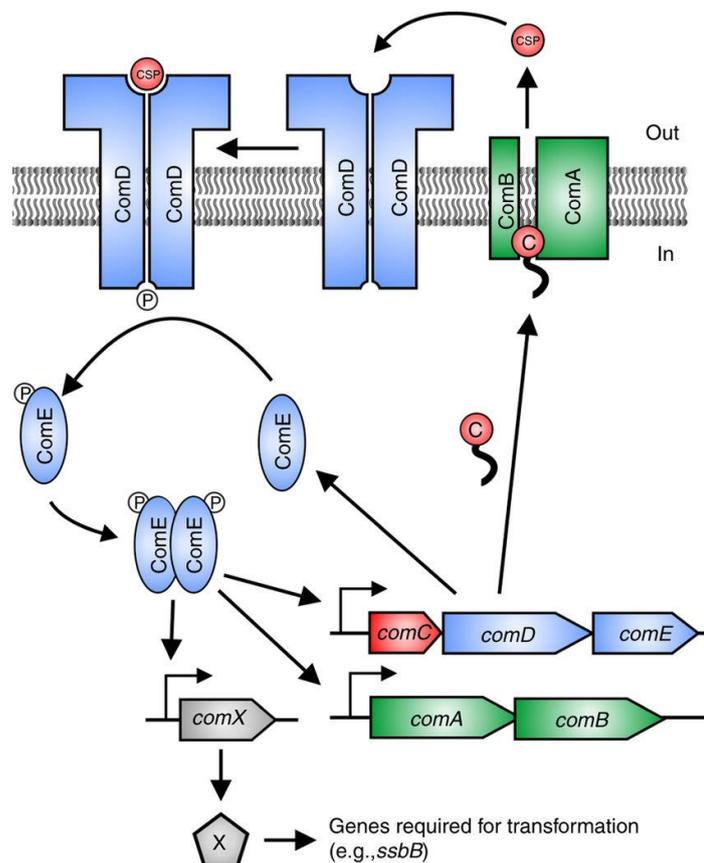


Figure 5 : Activation de la compétence chez *S. pneumoniae*.

Extrait de (Moreno-Gómez et al., 2017). Voir texte pour une description détaillée.

Contrairement à *B. subtilis*, le régulateur de compétence ComX de *S. pneumoniae* est un facteur sigma (σ_x). Les facteurs sigma permettent d'orienter l'ARN polymérase responsable de la transcription vers des gènes ayant des caractéristiques communes au niveau de leur promoteur pour coordonner une réponse génétique à des stimuli environnementaux. Ces facteurs sigmas ont 3 rôles principaux : ils reconnaissent une séquence spécifique sur le promoteur, permettent de positionner l'ARN polymérase sur ce dernier et facilitent l'ouverture de la double hélice pour permettre la transcription (Burgess, 2017). Pour répondre à tel ou tel besoin de la cellule, les bactéries produisent un large éventail de facteurs sigma. σ_x fait partie des facteurs sigmas dits « alternatifs » qui sont les facteurs sigma dédiés à des réponses spécifiques. Lorsqu'il est induit, il reconnaît une séquence de 8 pb

commune à tous les gènes de compétence (appelée *combox* (Peterson et al., 2000)) et active leur transcription coordonnée.

Des analyses initiales de profils d'expression de gènes en microarray puis en RNAseq a permis de montrer que l'induction des gènes de compétence se faisait en deux temps après induction avec le CSP : d'abord les gènes dits « précoces » qui sont induits par ComE puis les gènes « tardifs » qui sont induits par ComX (Peterson et al., 2004; Slager et al., 2019). D'après les données récentes de RNAseq (Slager et al., 2019), en tout c'est un peu plus d'une centaine de gènes qui sont directement induits par le CSP dont une quarantaine directement régulés par ComE (ce qui correspond à environ 15 ARNm polycistroniques) dont les opérons *comCDE* et *comAB* et une soixantaine régulés par ComX (20 ARNm polycistroniques) dont notamment tous les gènes requis pour la capture, l'internalisation et la recombinaison de l'ADNt. Cela représente 4% du génome de *S. pneumoniae* directement induit durant la compétence par ComE ou ComX. Ces gènes induits induisent eux-mêmes l'expression d'autres gènes, et au total c'est l'expression de plus de 360 gènes qui est affectée (17% du génome). Des expériences de mutagenèse ont permis de montrer que parmi tous ces gènes, seuls 23 sont requis pour la transformation de *S. pneumoniae* suggérant que d'autres processus biologiques que la transformation naturelle sont induits durant la compétence (Peterson et al., 2004). Ainsi, cet état a été qualifié de « X-state » pour enlever toute connotation fonctionnelle à cet état de compétence chez *S. pneumoniae* qui fait souvent référence à la transformation naturelle (Claverys et al., 2006).

I.C.2. Activation de la compétence chez *B. subtilis*

B. subtilis est une espèce de bactérie à gram positive non pathogène que l'on retrouve fréquemment dans les sols. Elle est caractérisée par sa grande capacité à survivre dans des environnements hostiles. Outre sa capacité à acquérir de l'ADN par transformation, *B. subtilis* est l'une des rares bactéries, notamment avec *Clostridium difficile* et *Bacillus anthracis*, à être capable de réaliser ce qu'on appelle « la sporulation » (Tan and Ramamurthi, 2014). La sporulation est un mécanisme au cours duquel la cellule bactérienne se divise de façon asymétrique pour former une cellule dormante métaboliquement inactive appelée spore dans laquelle une copie de l'ADNg est séquestrée. Ces spores sont extrêmement résistantes à tous types de stress et lorsque les conditions environnementales redeviennent propices à la croissance, elles sont capables de former de nouvelles cellules métaboliquement actives au cours d'un processus appelé germination (Tan and Ramamurthi, 2014).

Chez *B. subtilis*, la régulation de la compétence et de la sporulation est interconnectée et dépend du quorum sensing. Cette régulation utilise deux circuits distincts qui s'excluent mutuellement : l'un aboutissant à l'induction de la compétence et l'autre à l'activation de la sporulation (Waters and

Bassler, 2005). *B. subtilis* exprime deux peptides autoinducteurs spécifiques : ComX et CSF (Competence and Sporulation factor) et c'est la concentration intracellulaire de l'un par rapport à l'autre qui aboutit à l'une ou l'autre des deux voies.

Les cascades de régulation de la compétence et de la sporulation de *B. subtilis* et la façon dont les deux mécanismes s'excluent mutuellement est complexe (Maier, 2020). De façon simplifiée (**Figure 6**), ComX est synthétisé sous la forme d'un pré-ComX qui est processé et sécrété par le transporteur de la membrane plasmique ComQ. ComX processé est détecté par le senseur histidine kinase ancrée dans la membrane : ComP qui s'autophosphoryle. ComP~P transmet son phosphate au régulateur de réponse ComA. ComA~P active alors la transcription de divers gènes qui aboutissent à l'expression du régulateur central de la compétence ComK. ComK active lui-même à son tour, les gènes requis pour la transformation. L'autre autoinducteur CSF est d'abord codé par *phrC* sous sa forme non mature PhrC et sécrété dans le milieu extracellulaire par le système de sécrétion général Sec. PhrC est processé pour devenir le CSF mature qui est alors réinternalisé par le transporteur de peptides Opp. À faible concentration, CSF active indirectement la compétence en désactivant la protéine RapC qui en temps normal bloque l'action de ComA. Cependant, à forte concentration, CSF inhibe la cascade ComP->ComA et donc la compétence et favorise la sporulation en inhibant RapB qui en temps normal inactive par déphosphorylation un autre régulateur de réponse Spo0F. Ainsi à forte concentration Spo0F est dans sa forme active phosphorylé et active les gènes de sporulation.

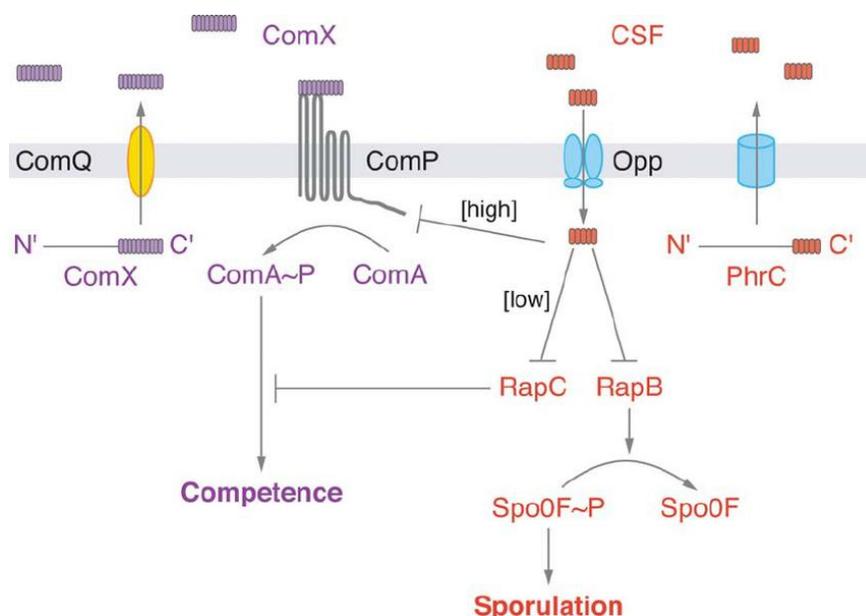


Figure 6 : Schéma simplifié des cascades moléculaires permettant l'activation de la compétence chez *B. subtilis*.

Extrait de (Waters and Bassler, 2005). Voir texte pour une description détaillée.

Contrairement au régulateur central de la compétence de *S. pneumoniae*, ComK n'est pas un facteur sigma, mais un activateur transcriptionnel qui peut se fixer sous la forme d'un dimère de dimère dans la région promotrice des gènes qu'il régule au niveau de ce qu'on appelle la K-box (Hamoen et al., 2002). Des études préliminaires de profils d'expression de gènes induits durant la compétence ont montré qu'environ une centaine de gènes étaient surexprimés durant la compétence induite par ComK (Berka et al., 2002; Claverys et al., 2006). Parmi ces gènes, 25 seraient impliqués dans la transformation et 6 dans la sporulation, les autres gènes seraient impliqués dans d'autres processus cellulaires tels que la division cellulaire ou le métabolisme de l'ADN (Berka et al., 2002; Claverys et al., 2006). C'est pourquoi tout comme chez *S. pneumoniae*, pour définir l'ensemble des processus induits au cours de l'état de compétence de *B. subtilis* Berka et al. ont qualifié cet état de K-state. Contrairement à *S. pneumoniae*, il n'existe pas de rétrocontrôle positif permettant d'augmenter la production de l'autoinducteur ComX. C'est peut-être pour cette raison que contrairement à *S. pneumoniae*, seuls 10% de la population bactérienne devient réellement compétentes (ie, exprime ComK) (Smits et al., 2005).

I.C.3. Activation de la compétence chez *V. cholerae*

On a vu précédemment que la compétence de *V. cholerae* pouvait être induite de plusieurs façons différentes : via la chitine, via le quorum sensing ou encore via la répression catabolique. Ces différents inducteurs entraînent des cascades de régulation différentes parfois interconnectées aboutissant à l'expression des gènes de compétence.

Induction par la chitine (Figure 7, panel de gauche). Les voies moléculaires qui aboutissent à l'activation de la compétence par la chitine font intervenir un senseur histidine kinase qui détecte la présence de chitine dans le milieu extracellulaire et, suite à des cascades moléculaires, active TfoX le régulateur central de la compétence de *V. cholerae* (Seitz and Blokesch, 2013a). Même si initialement il était suggéré que ChiS fonctionnait comme un système à deux composantes en tandem avec un potentiel régulateur de réponse, cette hypothèse est maintenant exclue. Il a été montré qu'il était plutôt couplé à un régulateur ancré dans la membrane interne TfoS (Dalia et al., 2014; Yamamoto et al., 2014). En absence de chitine, ChiS serait inhibé par la protéine de liaison à la chitine CPB, ce qui laisserait TfoS inactif. En présence de chitine, au contraire, la chitine lierait le CBP, ce qui lèverait l'inhibition de ChiS imposé par ce dernier. Sous l'induction de ChiS ou de la chitine seule, TfoS dimériserait pour passer à une forme active capable d'interagir avec l'ADN au niveau du promoteur du gène *tfoR* codant pour le petit ARN TfoR. TfoR chaperonné par Hfq interagirait avec la partie 5' non traduite de l'ARNm de *tfoX* pour favoriser l'accès au *Ribosome Binding Site* (RBS) par le ribosome. Une fois traduit, TfoR activerait les gènes requis pour la transformation de *V. cholerae*.

Induction par le quorum sensing (Figure 7, panel de droite). En plus de la chitine qui permet d'induire des voies spécifiques aboutissant à la compétence de *V. cholerae*, le quorum sensing joue aussi un rôle majeur dans l'induction de la compétence chez ce pathogène (Antonova and Hammer, 2015). *V. cholerae* produit deux autoinducteurs impliqués dans l'induction de la compétence : CAI-1 (pour *Cholerae AutoInducer 1*) et AI-2 (pour *AutoInducer 2*) qui sont produits et sécrétés dans le milieu extracellulaire respectivement par la synthase CqsA et l'enzyme LuxS. CAI-1 est reconnu par le récepteur membranaire histidine kinase CqsS tandis que AI-2 est reconnu par un récepteur périplasmique LuxP qui est associé à l'histidine kinase LuxQ. À de faibles densités cellulaires, LuxQ et CqsS agissent comme des kinases qui entraînent des phosphorylations en cascade, d'abord de LuxU, qui transmet ensuite son phosphate à LuxO qui, en tandem avec le facteur sigma RpoN, active alors la régulation d'au moins 4 petits ARN Qrrs (*quorum sensing RNAs*) (Antonova and Hammer, 2015). Par un mécanisme de silencing médié par la chaperonne à ARN Hfq, ces Qrrs vont s'hybrider au niveau du RBS de l'ARNm codant pour le facteur de transcription HapR. HapR étant l'activateur transcriptionnel de QstR qui est lui-même l'activateur transcriptionnel des gènes *comEA* et *comEC*, on en déduit qu'en l'absence d'autoinducteur le mécanisme qui aboutit à l'expression des gènes de compétence est réprimé en continu. Lorsqu'une concentration critique de CAI-1 ou AI-2 est atteinte, il y a, au contraire, une cascade de déphosphorylation qui aboutit à la levée de répression du système et l'activation de la transcription des gènes de compétence (Antonova and Hammer, 2015).

Induction par la répression catabolique (Figure 7, panel du milieu). Lorsqu'on a découvert que la chitine induisait la compétence de *V. cholerae*, on a aussi montré que le glucose inhibait la transformation naturelle chez cette espèce (Meibom et al., 2005). Cette inhibition par le glucose d'une voie métabolique rappelle le phénomène de répression catabolique retrouvé chez un grand nombre de bactéries (Deutscher et al., 2006). Il a donc été proposé que ce phénomène jouait un rôle dans la régulation de la compétence de *V. cholerae* (Blokesh, 2012). La répression catabolique est un mécanisme de régulation générale du catabolisme bactérien qui permet à une bactérie de métaboliser prioritairement sa source de carbone préférée (généralement le glucose) en inhibant la synthèse des enzymes impliquées dans l'utilisation d'autres sources de carbone (Deutscher et al., 2006). Les acteurs les plus importants de ce mécanisme sont le système phosphoenolpyruvate (PEP): carbohydate phosphotransferase (PTS) qui permet l'absorption et la phosphorylation de nombreuses sources de carbone, et l'enzyme productrice d'AMPc (l'adénylate cyclase CyaA) avec son récepteur CRP (*cAMP Receptor Protein*). Dans le système PTS, les glucides traversent la membrane par l'enzyme EIIBC et entrent dans une cascade de phosphorylation dont le donneur primaire est le phosphoenolpyruvate et dont le transfert du groupe phosphoryl est réalisé successivement par l'enzyme EI (EI), la protéine Hpr et l'enzyme EIIA. En l'absence de la source de carbone préférentielle (par exemple, du glucose),

l'enzyme EIIA reste phosphorylée et agit directement sur l'adénylate cyclase pour augmenter le niveau d'AMPc dans la cellule. Cet AMPc se fixe sur son récepteur CRP. Le complexe AMPc-CRP est connu pour contrôler l'expression de très nombreux gènes impliqués dans diverses voies métaboliques en se fixant au niveau de séquences appelées *CRP-sites* présentes dans le promoteur des gènes cibles (Zheng et al., 2004). La délétion d'AMPc ou de CRP inhibe la voie de régulation de la compétence induite par la chitine (Blokesch, 2012). Par ailleurs, l'absence de l'enzyme dégradant l'AMPc : l'AMPc phosphodiesterase (CdpA) entraîne une augmentation des fréquences de transformation (Scrudato and Blokesch, 2012). Il a aussi été montré que CRP-AMPc jouait un rôle dans le quorum sensing notamment en activant la transcription de *hapR* (Silva and Benitez, 2004). Chez *H. influenzae*, il a aussi été montré que l'adénylate cyclase et le CRP étaient essentiels pour le développement de sa compétence (Chandler, 1992; Dorocicz et al., 1993). Des analyses du profil d'expression de gènes en microarray ont montré que l'ensemble des gènes faisant partie du régulon de compétence de cette espèce possédait une séquence de fixation pour le régulateur central de la compétence Sxy étroitement liée aux séquences de fixation de l'AMPc-CRP et il a ainsi été suggéré que Sxy permettait de faciliter l'accès et/ou la stabilisation de CRP à sa séquence de fixation (Redfield et al., 2005). Ces séquences particulières de fixation à la fois au CRP et à Sxy ont été nommées CRP-S chez *H. influenzae* (Cameron and Redfield, 2006). Par des analyses *in silico*, il a été montré que de telles séquences étaient aussi présentes chez *V. cholerae* (Cameron and Redfield, 2006). Or, Sxy est un homologue de TfoX de *V. cholerae*. Il a donc été suggéré que TfoX, tout comme Sxy joue aussi un rôle dans la fixation de CRP-AMPc à ses gènes cibles (Scrudato and Blokesch, 2012).

Suivant ces différentes observations, un modèle a été proposé dans lequel la répression catabolique joue un rôle dans l'induction de la compétence en agissant possiblement sur le régulateur central de la compétence TfoX et/ou sur le quorum sensing (Blokesch, 2012; Scrudato and Blokesch, 2012; Seitz and Blokesch, 2013a).

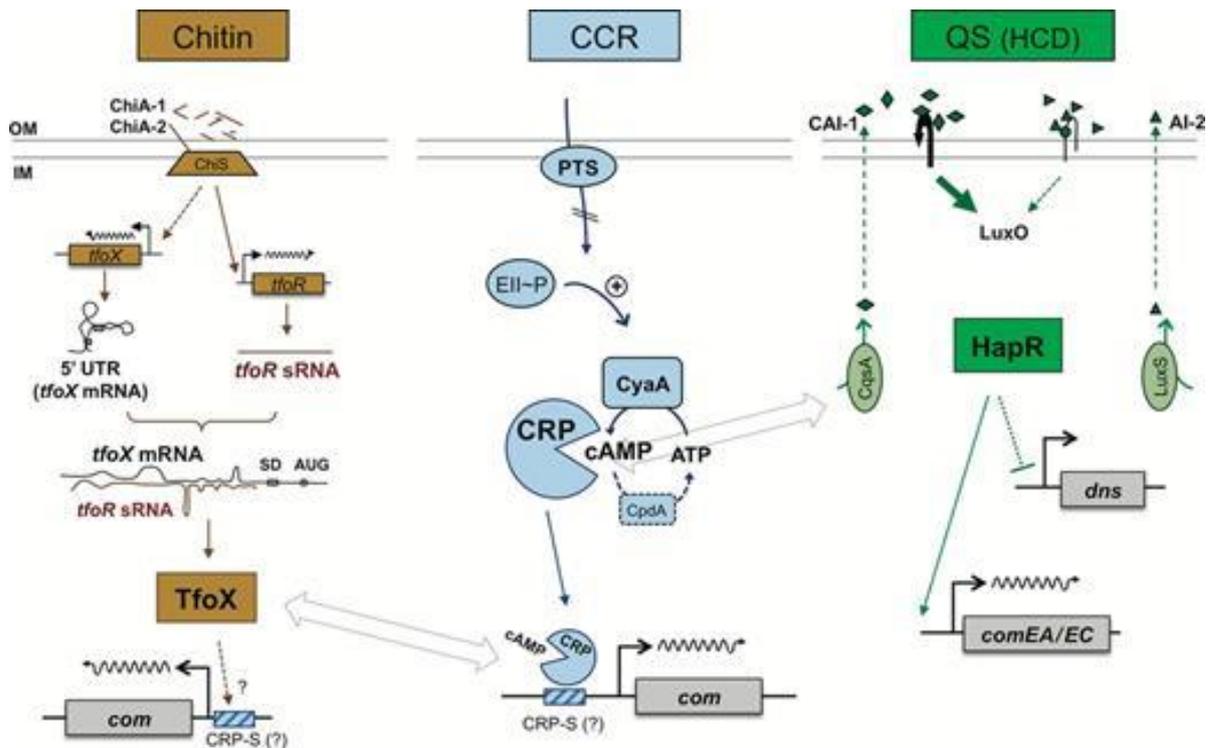


Figure 7 : Induction de la compétence chez *V. cholerae* par la chitine (panel de gauche), la répression catabolique (panel du milieu) et le quorum sensing (panel de droite).

Extrait de (Moreno-Gómez et al., 2017). Voir texte pour une description détaillée.

L'ensemble de ces mécanismes de régulation de la compétence décrit ici, sont basés sur des activateurs transcriptionnels. Nous allons voir en partie III.B que la régulation de la compétence de *L. pneumophila* est, en revanche, basée sur un mécanisme de répression post-transcriptionnelle et c'est à ce jour le seul mécanisme de ce type jamais décrit chez une bactérie transformable (Attaiech and Charpentier, 2017).

Nous avons vu que ces mécanismes et les cascades moléculaires qui en découlent sont très divergents, mais permettent dans tous les cas l'expression des acteurs nécessaires à la mise en place de ce qu'on appelle « la machinerie d'import de l'ADN ». Paradoxalement, les acteurs moléculaires engagés dans cette machinerie qui permettent notamment la capture, l'internalisation et la recombinaison de l'ADNt sont relativement conservés au sein des différentes espèces transformables comme nous allons le voir dans une seconde partie.

II. Une machinerie de transformation conservée

A. Capture de l'ADN exogène

À l'exception notable de *H. pylori*, toutes les bactéries transformables utilisent des appendices extracellulaires apparentés au pilus de type IV et/ou au système de sécrétion de type II pour capturer

l'ADN exogène au cours du processus de transformation naturelle (Piepenbrink, 2019). Ces structures incluent le pilus de type IV (T4P), le pilus de compétence des bactéries gram positives (pilus Com, historiquement appelés pseudopilus, et le pilus *Flp/tad-like* de *Micrococcus luteus* récemment découvert (Angelov et al., 2015) non décrit ici. Ils sont regroupés, selon la littérature, sous le nom de *DNA-uptake pilus* (DU-pilus) (Dubnau and Blokesch, 2019) ou filaments de type IV (T4F) (Piepenbrink, 2019) (Figure 8).

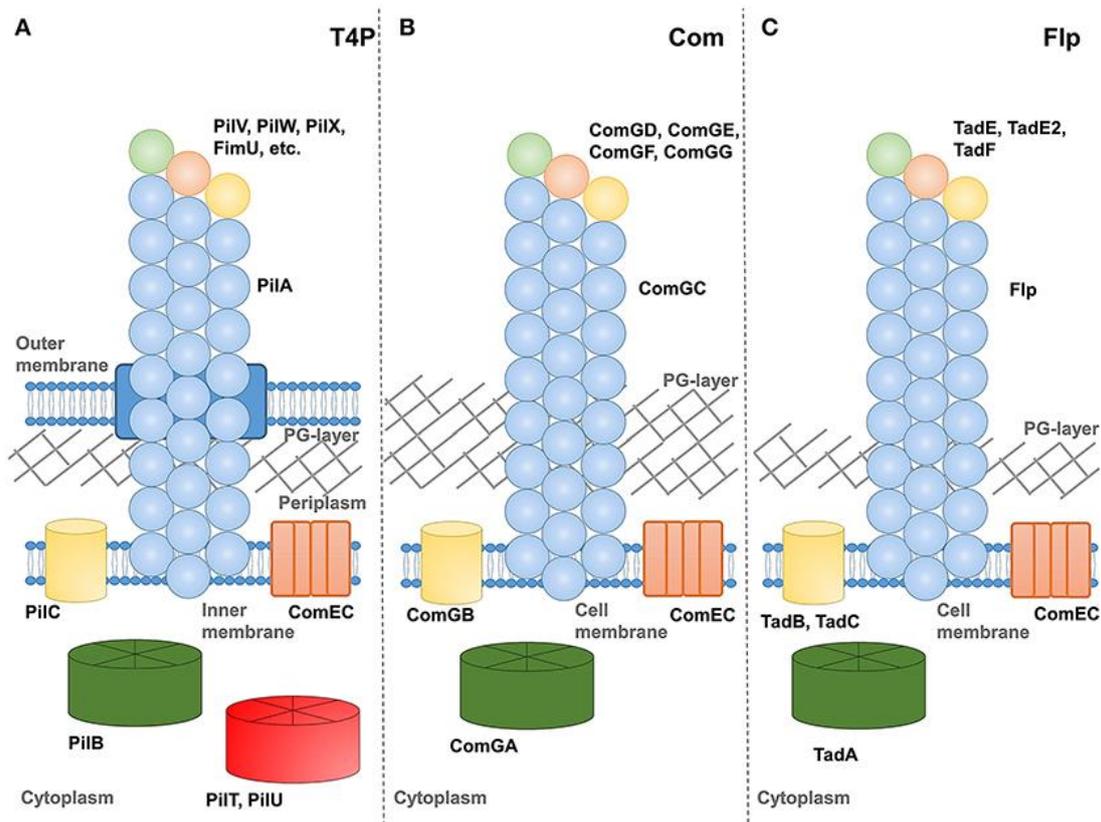


Figure 8 : Représentation schématique des différents filaments de type 4 impliqués dans la capture de l'ADN au cours de la transformation naturelle.

De gauche à droite : le pilus de type 4 ou T4P, le pilus de compétence ou pilus Com et le pilus *Flp/tad-like* de *Micrococcus luteus*. Extrait de (Piepenbrink, 2019)

II.A.1. Éléments conservés des filaments impliqués dans la transformation naturelle

Indépendamment du type de filament, l'architecture générale du pilus est conservée. En effet, ils résultent tous de l'assemblage de petites sous-unités protéiques appelées « pilines ». Ils sont, dans tous les cas, constitués de plusieurs milliers de copies d'une piline dite « majeure » (PiliA, ComGC et Flp respectivement pour le T4P, le pilus Com et le pilus *Flp/tad-like*) et de quelques sous-unités de pilines dites « mineures ». Bien que la séquence primaire des pilines soit divergente, elles partagent toute une structure en « sucette » caractéristique (Giltner et al., 2012) (**Figure 9a**). En effet, elles possèdent

toutes un domaine N-terminal hydrophobe en α -hélice relié à une tête globulaire variable, amphiphile en C-terminal constituée de plusieurs feuillets β antiparallèles. Le domaine N-terminal est lui-même divisé en 2 sous-domaines : α -1N et α -1C qui s'étendent jusque dans la tête globulaire. Le domaine α -1N joue un rôle très important avant et pendant l'assemblage du pilus. En effet, il sert, d'une part, de segment transmembranaire pour permettre l'ancrage de la piline dans la membrane cytoplasmique, mais aussi d'autre part, de domaine d'interaction pour l'incorporation de nouvelles pilines au sein du pilus en formation (Giltner et al., 2012). Au sein du pilus, les α -hélices sont liées entre elles par des interactions hydrophobes qui constituent le cœur du pilus, ce qui permet l'exposition des têtes globulaires à l'extérieur, qui, ensemble, forment l'enveloppe du pilus (Craig et al., 2006) (**Figure 9b**).

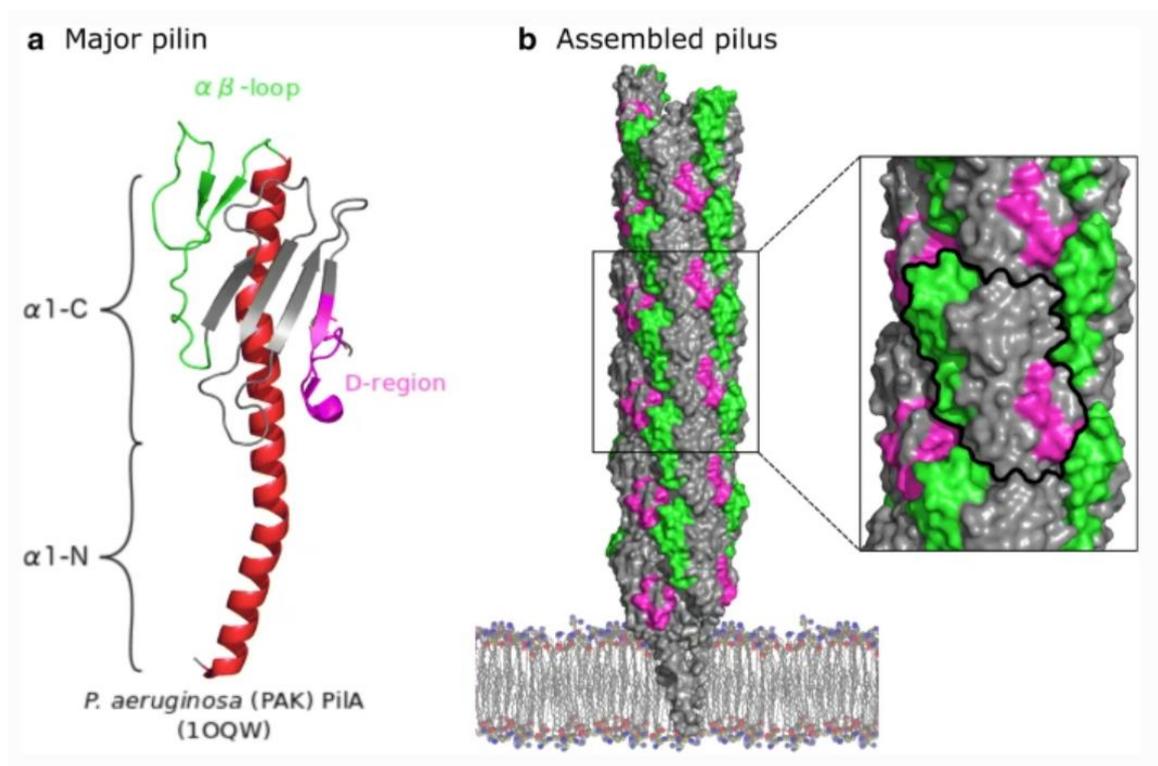


Figure 9 : Représentation moléculaire de la piline majeure pilA du T4P de *Pseudomonas aeruginosa*. (a) et du pilus assemblé (b).

Les pilines adoptent une structure en forme de « sucette » constituée d'un domaine N-terminal en α -hélice (α 1-N + α 1-C) (en rouge) relié par l' $\alpha\beta$ -loop (vert) à un domaine C-terminal globulaire constitué de plusieurs feuillets β (gris) et d'une région D hypervariable (rose). Une fois assemblées, les pilines forment un filament avec les têtes globulaires présentées à l'extérieur et constituant l'enveloppe du pilus (gris + rose + vert) et les domaines N-terminaux compactés ensemble à l'intérieur du pilus pour former son squelette. Extrait de (Jacobsen et al., 2019)

Peu importe le type de piline, toutes sont synthétisées sous la forme d'une prépiline qui se distingue par la présence d'un peptide leader en N-terminal. Ce peptide leader est chargé positivement et permettrait l'adressage des pilines à la membrane cytoplasmique dans la bonne orientation (Giltner

et al., 2012) à savoir l' α -hélice ancrée dans la membrane cytoplasmique et le domaine globulaire dans le compartiment périplasmique. La maturation de ces prépilines en pilines est réalisée par une prépiline peptidase qui existe dans chacun des systèmes (PilD pour le T4P (Nunn and Lory, 1991), ComC pour le pilus Com (Chung and Dubnau, 1995) et TadV pour le pilus *Flp/tad-like* (Angelov et al., 2015)) et qui clive le peptide leader après un résidu d'acide aminé qui dépend de chaque système au niveau du côté cytoplasmique de la membrane (Berry and Pelicic, 2015) (Figure 10). Cette étape de maturation est indispensable pour permettre leur assemblage (Strom and Lory, 1991). Une fois processées, l'ensemble des pilines ancrées dans la membrane interne forment un réservoir directement utilisable pour l'assemblage du pilus.

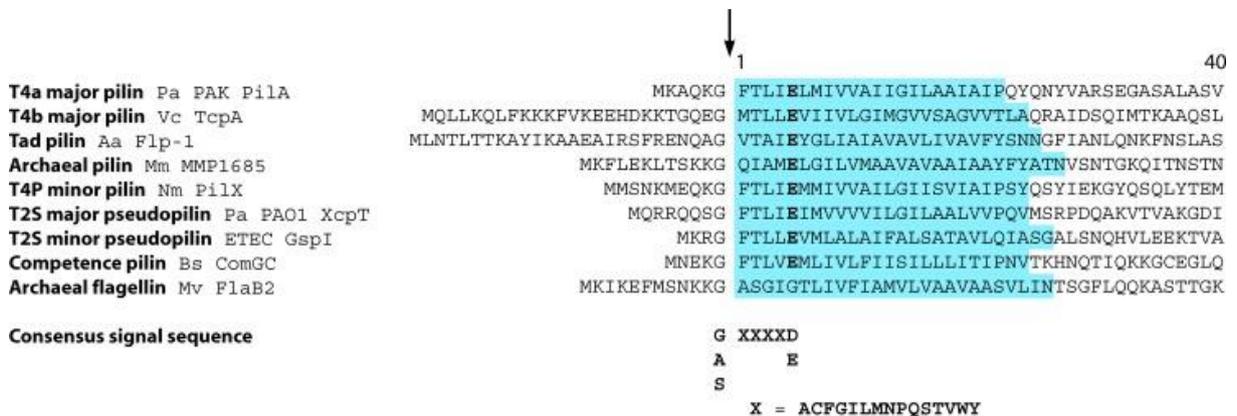


Figure 10 : Alignement de la séquence N-terminal des pilines. Extrait de (Giltner et al., 2012).

La flèche représente le site de coupure par la prépiline péptidase. En gras est représenté le résidu glutamate invariable. En bleu est surlignées les parties ancrées dans la membrane plasmique.

En plus d'avoir un motif en N-terminal conservé, la majorité de ces pilines possèdent un résidu glutamate (E) invariable situé, quel que soit le type de pilus ou de piline, à la position 5 de la piline mature (Figure 10). Ce résidu est le seul résidu hydrophile du tronçon N-terminal et permettrait selon les modèles de pilus actuels de former un pont salin entre les parties N-terminales de la piline N et de la piline N+1 nouvellement assemblée (Craig et al., 2003). Certaines pilines mineures (une par espèce) ne possèdent pas de glutamate en position 5. C'est le cas pour PilX de *A. baumannii* et ComGF de *S. pneumoniae*, et il est suggéré que ces pilines soient situées à la pointe du pilus et n'ont donc pas besoin de faire un pont salin avec la piline N+1 (Piepenbrink, 2019).

Pour l'ensemble des filaments de type IV, en plus des pilines, l'assemblage requière au moins deux composants indispensables : une ATPase hexamérique AAA+ d'assemblage cytoplasmique [PilB pour le T4P (Sukmana and Yang, 2018), ComGA pour le pseudopilus et TadA pour le pilus *Flp/tad-like* (Bhattacharjee et al., 2001)] et une protéine membranaire intégrale qui forme une plateforme dans la membrane interne capable de transduire le signal (PilC/G pour le T4P (Craig et al., 2019), ComGB pour

le pseudopilus (Muschiol et al., 2015) et TadB/C pour le pilus *Flp/tad-like* (Tomich et al., 2007) (Figure 8).

Par ailleurs, le rôle des pilines mineures est encore largement méconnu, mais plutôt que de participer à la structure du pilus il est suggéré que certaines sont, par exemple, responsables de la liaison à l'ADNsb comme c'est le cas de ComP chez *N. meningitidis* (Cehovin et al., 2013; Piepenbrink, 2019).

Le mécanisme d'assemblage des filaments de type 4 est encore peu connu, mais plusieurs modèles d'assemblage ont été suggérés basés sur les structures obtenues en cryo-microscopie électronique et tomographique du T4P (et notamment de PilB) chez différentes espèces (Chang et al., 2016; Gold et al., 2015; Mancl et al., 2016). L'un de ces modèles suggère un assemblage des pilines en commençant par la pointe du pilus, par rotation de la plateforme de la membrane interne (Craig et al., 2006, 2019) (**Figure 11**). L'hydrolyse de l'ATP par l'ATPase d'assemblage induirait un changement conformationnel de cette dernière qui entrainerait une rotation de la plateforme de la membrane interne situé à proximité. Cette rotation surélèverait cette plateforme placée à la base du pilus en formation. Par un effet levier, cela entrainerait l'extrusion du pilus de 8 à 10 Å à l'extérieur de la membrane, laissant un espace entre la base du pilus et la membrane plasmique. Cet interstice serait directement comblé par une des pilines du réservoir membranaire diffusant à travers la membrane interne, attirée par complémentarité des charges du Glu5 chargé négativement et du premier résidu de l' α 1-N de la piline chargé positivement. Une fois le phosphate consommé, le système reviendrait à son état initial pour un nouveau cycle d'incorporation de piline. Pour le T4P, des études récentes ont montré que cet assemblage se déroulait dans le sens horaire et était couplé à la rétraction par désassemblage des pilines dans le sens antihoraire (McCallum et al., 2017; Solanki et al., 2018).

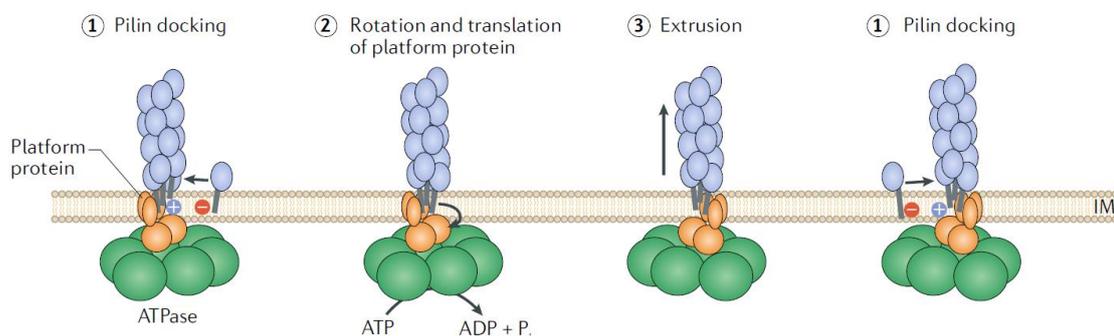


Figure 11 : Modèle d'Assemblage du T4P par rotation de l'ATPase d'assemblage.

L'ATPase d'assemblage, la plateforme de la membrane interne et les pilines majeures formées d'une tête globulaire et de leur α 1-N sont respectivement représentée en vert, orange et bleu. (1) Une piline majeure provenant du réservoir membranaire diffuse dans l'ADN pour s'assembler au pilus par complémentarité des charges de son Glu5 et de l' α 1-N de la piline N+1 préalablement

assemblée. (2) La liaison de l'ATP à l'ATPase d'assemblage provoque un changement conformationnel de cette dernière qui est transmis à la plateforme de la membrane interne, provoquant sa rotation. (3) En tournant, par effet levier, cette plateforme provoque l'extrusion du pilus de quelques nm, laissant un espace entre la base du pilus et la membrane interne. (1) Cet espace est directement comblé par une nouvelle piline, recommençant le cycle d'assemblage. Extrait de (Craig et al., 2019)

Malgré une architecture conservée de ces systèmes, il existe des spécificités inhérentes à chacun d'entre eux qui dépendent de la physiologie de la bactérie et qui ont une influence sur les stratégies mises en place par ces systèmes pour capturer l'ADN transformant (ADNt).

II.A.2. Spécificité de différents systèmes et leur rôle dans la capture de l'ADN exogène

II.A.2.1. Le pilus de type 4

II.A.2.1.1. Généralités

Le T4P a été décrit pour la première fois chez *P. aeruginosa* en 1972 (Bradley, 1972) et a ensuite été identifié chez un très grand nombre d'espèces gram négatives (Denise et al., 2019; Imam et al., 2011). Alors qu'on le pensait spécifique des espèces gram négative, la disponibilité croissante des séquences complètes de génomes bactériens a montré que les gènes codant les constituants de ce système – qu'on appellera dorénavant locus *pil* – étaient aussi très répandus chez les espèces gram positives (Denise et al., 2019; Imam et al., 2011; Melville and Craig, 2013). Des trois filaments, le T4P est de loin le plus étudié notamment, car, outre la transformation naturelle, il est impliqué dans de nombreux processus cellulaires tels que l'adhésion aux surfaces biotiques ou abiotiques, la mobilité twitching ou encore la formation de biofilms ou de microcolonies ; ce qui lui a valu d'être qualifié de « couteau Suisse procaryote » (Berry and Pelicic, 2015). Néanmoins, son implication dans la capture de l'ADN durant le processus de transformation naturelle n'a pour l'heure été établie que chez les espèces gram négatives.

Les T4P sont généralement divisés en deux sous-types T4aP et T4bP (Craig et al., 2004) bien qu'un troisième sous-type T4cP dont ferait partie les *Flp/tad pili* ait été introduit récemment (Ellison et al., 2017). Cette classification est basée sur la séquence du peptide leader et l'assemblage des pilines. Alors que les gènes codant pour les T4b et T4c pili sont regroupés en opéron, les gènes codant le T4a sont dispersés en différents clusters dans tout le génome à des positions conservées (Denise et al., 2019; Pelicic, 2008). Les T4aP sont les plus répandus puisqu'on les retrouve chez diverses espèces alors que les T4bP sont restreints aux bactéries entériques. Les T4aP sont les seuls T4P pour lesquels

l'implication dans la capture de l'ADN transformant (ADNt) au cours de la transformation naturelle a été décrite à ce jour (Piepenbrink, 2019).

Les pili de type IV sont de longs (plusieurs μM) et fins (entre 60 et 80 \AA) filaments (Berry and Pelicic, 2015) capables d'interagir latéralement pour former des faisceaux (**Figure 12**). Ils sont extrêmement flexibles et dynamiques capables de se polymériser et se dépolymériser à une vitesse de ~ 1000 sous-unités par seconde (Merz et al., 2000). Qu'ils soient seuls ou en faisceau ils sont très puissants générant des forces de plus de 100 pN (Maier et al., 2002) allant jusqu'à plusieurs nM lorsqu'ils sont en faisceaux (Biais et al., 2008).

Bien qu'on retrouve dans les trois systèmes au moins une ATPase AAA+ conservée, seul le T4P est pourvu, en plus de son ATPase d'assemblage PilB, d'une ou de plusieurs ATPases de rétraction PilT, PilU et/ou PilT2. PilT semble être dans la totalité des espèces gram négatives l'ATPase essentielle à la rétraction du T4P puisqu'il a été montré que la délétion du gène *pilT* induit une perte de mobilité twitching et/ou de transformabilité dans bon nombre d'espèces gram négatives (par exemple, chez *V.*

cholerae (Adams et al., 2019), *Pseudomonas stutzeri* (Graupner et al., 2001), *A. baylyi* (Leong et al., 2017), *Acinetobacter baumannii* (Harding et al., 2013), *Dichelobacter nosodus* (Han et al., 2008), *N. meningitidis* (Pujol et al., 1999), *Neisseria gonorrhoeae* (Wolfgang et al., 1998) ou encore la cyanobactérie *Synechocystis sp. PCC6803* (Bhaya et al., 2000)). Dans certaines de ces études, la délétion de *pilT* a même été associée à un phénotype d'hyperpiliation des bactéries suggérant le rôle prépondérant de PilT dans la rétraction du pilus. Dans ces espèces, PilU est la plupart du temps codé en opéron en aval de PilT et semble jouer un rôle subsidiaire notamment dans la transformation naturelle. Certaines espèces, par exemple de *Neisseria spp.* ou *Geobacter sulfurreducens*, sont dotées d'autres ATPase tels que PilT2 et de jusqu'à 4 ATPases paralogues de PilT (Kurre et al., 2012; Speers et al., 2016).

Une autre différence notable entre le T4P des gram négatives et les autres systèmes est qu'il comporte une sécrétine PilQ qui permet au pilus de traverser la membrane externe pour pouvoir émerger à l'extérieur de la bactérie. PilQ est essentielle à la transformation des bactéries gram négative car une délétion du gène *pilQ* induit une absence de transformation et aucun pili et/ou import d'ADN

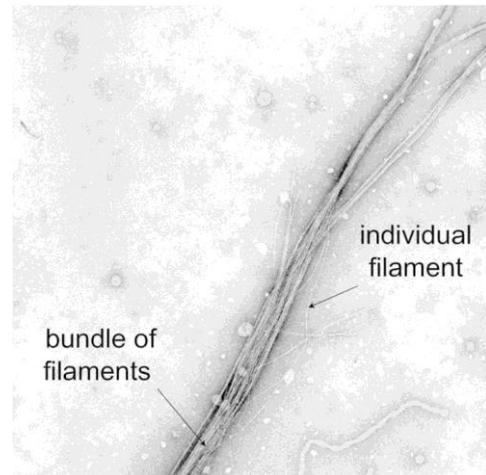


Figure 12 : Visualisation par microscopie électronique de faisceaux de T4P de *N. meningitidis* par coloration négative à l'acide phosphotungstique.

D'après (Berry and Pelicic, 2015)

transformant n'est détecté pour ce mutant chez diverses espèces transformables (Drake and Koomey, 1995; Friedrich et al., 2002, 2002). Au moins chez *N. meningitidis* il a été montré que cette sécrétine PilQ liait l'ADNt de façon non spécifique (Assalkhou et al., 2007) suggérant qu'elle pourrait aussi être nécessaire au transport de l'ADNt.

II.A.2.1.2. Rôle du T4P dans la transformation naturelle

La capture de l'ADN exogène par le T4P au cours de la transformation naturelle a été principalement étudiée chez *N. gonorrhoeae* et *V. cholerae* (Ellison et al., 2018; Hamilton and Dillard, 2006; Matthey and Blokesch, 2016).

On sait depuis plusieurs décennies maintenant que lors de ce processus de transformation naturelle *N. gonorrhoeae* est capable de reconnaître une séquence spécifique appelées DNA-uptake sequence (DUS) (Goodman and Scocca, 1988) de 12 pb (ATGCCGTCTGAA). Cette habilité à reconnaître une séquence spécifique est inhabituelle chez les bactéries transformables puisqu'à ce jour, seule *H. influenzae* est l'unique autre espèce capable d'un tel mécanisme (Mell et al., 2012). La DUS reconnue par *N. gonorrhoeae* représente 1% de son génome (Smith et al., 1999) et cette bactérie étant constitutivement transformable, certains suggèrent que cette spécificité est importante pour reconnaître l'ADN de ses semblables et limiter l'intégration d'ADN parasite (Dubnau and Blokesch, 2019). En revanche, ce n'est que récemment que le récepteur de ces DUS a été identifié chez *Neisseria* (Cehovin et al., 2013). Il s'agit de ComP une piline mineure qui avait pourtant été identifiée plusieurs années auparavant comme indispensable à la transformation naturelle sans pour autant être essentielle à la mise en place du pilus (Wolfgang et al., 1999). ComP est disposée à différents endroits le long du pilus (Berry et al., 2013), ce qui suggérerait que le pilus puisse lier l'ADNt en plusieurs points contrairement à ce qui a été observé chez *V. cholerae* où l'ADNt semble être lié uniquement par la pointe du T4P (Ellison et al., 2018).

Ces observations ont mené les chercheurs à établir un modèle de capture de l'ADN chez *N. gonorrhoeae* qui est à ce jour le plus abouti (Berry et al., 2016; Dubnau and Blokesch, 2019) (**Figure 13**). Par l'intermédiaire de son pilus et de ComP, *N. gonorrhoeae* lierait l'ADNt. La rotation du pilus exercée par PilT/C lors de sa rétraction entrainerait l'enroulement de l'ADNt sur le filament au niveau de ses sillons. Cet enroulement de l'ADN autour du pilus permettrait de réduire le diamètre du complexe ADN-pilus, ce qui permettrait à l'ensemble de passer à travers la sécrétine PilQ pour pénétrer dans le périplasme.

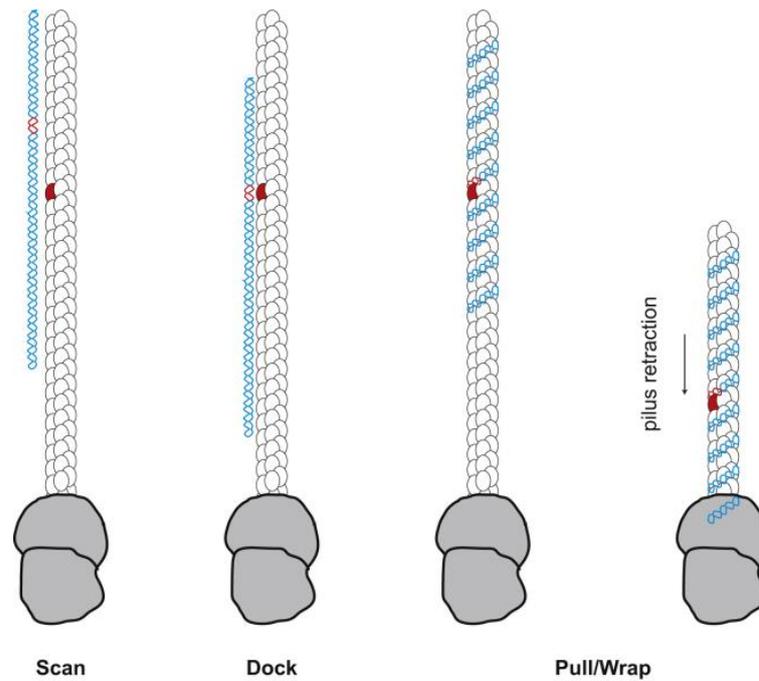


Figure 13 : Modèle de capture de l'ADNt par *Neisseria spp.*

Extrait de (Piepenbrink, 2019). La piline mineure ComP (en rouge) est localisée sporadiquement le long du T4P (en blanc). Le pilus fixe l'ADNt (en cyan) qui est scanné pour la présence de DUS (région rouge de l'ADN). Une fois la DUS reconnue l'ADNt se fixe spécifiquement au pilus et s'enroule autour. La rétraction du pilus permet au complexe ADNt-pilus de passer à travers la membrane externe par l'intermédiaire du canal PilQ (en gris).

Ce modèle a l'avantage de concilier toutes les hypothèses formulées à ce jour quant au mécanisme de capture de l'ADNt : rôle prépondérant des pilines mineures dans la liaison de l'ADNt, rotation et rétraction du pilus, passage de l'ADNt par le canal PilQ. Cependant, il manque encore un certain nombre d'éléments pour qu'il puisse être généralisé à toutes les bactéries transformables.

En effet, même si l'utilisation de pilines mineures semble être la norme pour lier l'ADNt comme cela a été montré chez un certain nombre d'espèces outre *Neisseria spp* : VC0858 et VC0859 pour *V. cholerae* (Ellison et al., 2018), ComZ pour *Thermus thermophilus* (Salleh et al., 2019) ou encore FimT pour *L. pneumophila* (Braus et al., 2021), on ne peut pas exclure un rôle direct ou indirect des pilines majeures dans la liaison de l'ADNt. De même, au vu du manque d'homologie qui existe entre ces différentes pilines réceptrices, les mécanismes par lesquels elles fixent l'ADNt sont probablement très différents. Également, les sites de fixation au pilus semblent variés d'une espèce à une autre en fonction de si les pilines réceptrices sont situées à la pointe du pilus comme cela a été démontré chez *V. cholerae* (Ellison et al., 2018) ou réparties sur toute sa longueur comme c'est le cas chez *N. gonorrhoeae*. Par ailleurs, la rétraction semble être le mécanisme privilégié pour permettre l'entrée de l'ADNt dans le périplasme comme l'indiquent les différentes études précitées sur PilT et comme

cela a été montré pour la première fois en temps réel chez *V. cholerae* dans l'étude de Ellison et al., publiée en 2018 (Figure 14).

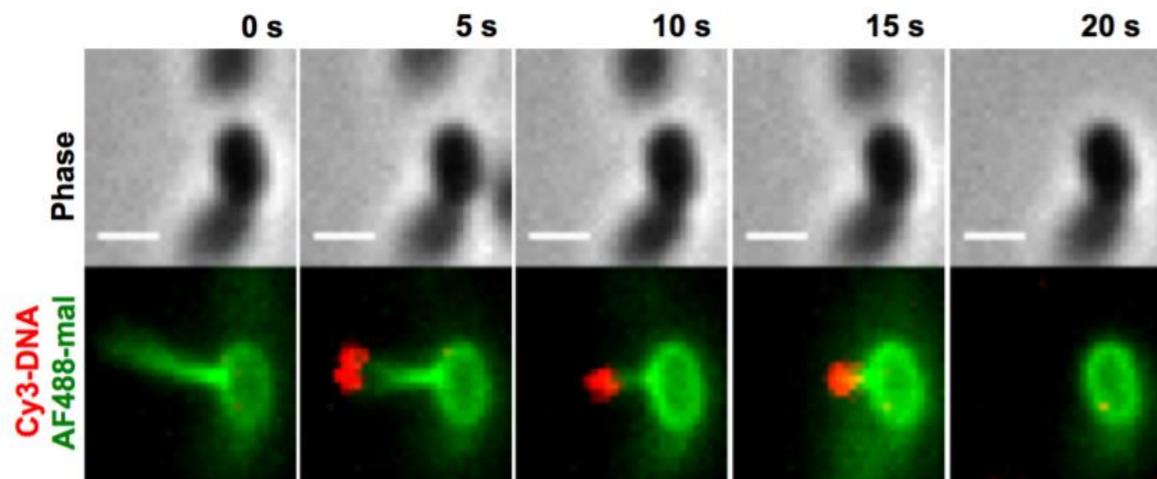


Figure 14 : Le T4P de *V. cholerae* lie l'ADNt à distance de la cellule et le ramène à la surface. Montage d'images issues de timelapse de microscopie à fluorescence d'une souche pilA-cys marquée avec du AF488-mal (en vert) et incubée avec de l'ADN Cy3 en rouge. Echelle (barre blanche) = 1 μ M. Extrait de (Ellison et al., 2018)

Cependant, dans cette même étude, les auteurs suggèrent que la rétraction seule n'est probablement pas suffisante et que ComEA pourrait avoir un rôle dans cette internalisation (voir II.B.1). Quant à la rotation du pilus, même s'il existe des preuves biochimiques, sans preuve visuelle, elle n'est pour l'instant qu'une hypothèse. Enfin, le canal PilQ étant relativement étroit (7-8 nm), certains suggèrent qu'il serait peu enclin à laisser passer d'aussi gros complexes tel que le complexe pilus-ADNt (Piepenbrink, 2019) même si des études suggèrent que l'ouverture de PilQ est dynamique et qu'il serait capable de s'accommoder au passage de différents substrats (Gold et al., 2015).

Ainsi, même si le mécanisme général de capture de l'ADN exogène par l'intermédiaire du T4P est connu, il reste un certain nombre d'interrogations qui ne permettent pas de généraliser le modèle actuel prévalant chez *Neisseria*, à toutes les bactéries transformables. Notamment, il faudrait étendre les recherches à d'autres espèces pour identifier les possibles similitudes et spécificités de ces systèmes.

A l'évidence, ce modèle n'est pas applicable aux espèces gram positives pour lesquelles ni d'homologues d'ATPase de rétraction, ni de sécrétine tel que PilQ n'ont été retrouvées.

II.A.2.2. Le pilus de compétence

II.A.2.2.1. Généralités

Contrairement aux loci *pil* qui sont retrouvés à la fois chez les gram positives et négatives, les gènes qui codent pour le pilus de compétence – qu'on appellera « loci *com* » pour parler des gènes ou « pilus Com » pour parler du filament – sont restreints à l'unique phylum des Firmicutes, représentants majeurs des bactéries gram positives (Sheppard et al., 2020). Plus précisément, d'après cette même étude, les loci *com* dans leur intégralité ont été retrouvés dans plus de 2300 génomes bactériens, ce qui représente 350 espèces différentes dont la quasi-totalité appartient aux Bacillales et aux Lactobacillales, deux sous-classes de Firmicutes. Justement, les pili Com ont été principalement étudiés chez *B. subtilis* et *S. pneumoniae*, deux espèces appartenant respectivement à chacune de ces sous-classes.

Contrairement au T4P, le pilus Com intervient exclusivement dans la transformation naturelle. Il est codé par l'opéron *comG* qui comprend 7 gènes : *comGA*, *B*, *C*, *D*, *E*, *F* et *G* dont l'organisation est conservée quelle que soit l'espèce. Aussi bien chez *B. subtilis* que chez *S. pneumoniae*, ces gènes sont induits uniquement durant la phase de compétence des bactéries (Dubnau, 1991; Peterson et al., 2004) et sont absolument requis pour la capture de l'ADN exogène au cours du processus de transformation naturelle (Chung and Dubnau, 1995, 1998; Muschiol et al., 2015). Le gène *comC* de *B. subtilis* (*pppA* chez *S. pneumoniae*) homologue de *pilD* et qui code pour une prépiline peptidase est placé à un autre endroit du génome et est aussi requis dans le processus (Chung and Dubnau, 1995).

Le pilus Com de *B. subtilis* n'a pour l'heure jamais pu être observé et c'était aussi le cas pour le pilus Com de *S. pneumoniae* jusqu'à récemment. En effet, deux études ont mis en évidence deux structures très différentes de présumés pilus Com (**Figure 15**). Laurenceau et al. en 2013, sont les premiers à avoir pu observer par microscopie électronique des filaments similaires aux T4P codés par le locus *pil*. Ils ont en effet décrit de longs et fins filaments de 2-3 μ M de longueur et de 5 à 6 nm de diamètre (50-60 Å) principalement constitués de la piline ComGC (**Figure 15A**). Les auteurs ont considéré que malgré le fait qu'il soit codé par les loci *com*, ce long pilus est un

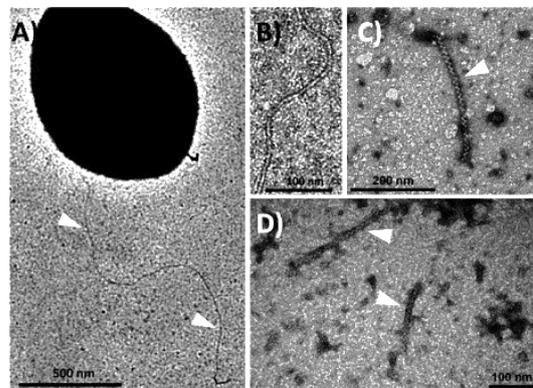


Figure 15 : Visualisation directe du pilus Com de *S. pneumoniae* par microscopie électronique.

(A) et (B) issue de (Laurenceau et al., 2013) : Observation d'une cellule compétente et de son long pilus (flèche blanche). (C) et (D) issue de (Balaban et al., 2014) Observation de pili courts et d'aspect tressé.

authentique T4P (« *bona fide* T4P ») (Laurenceau et al., 2013). A contrario, Balaban et al. en 2014, ont observé chez cette même espèce des filaments présumés être des pili Com très courts (~100 nm) et d'aspect tressé qu'ils n'avaient pas observé dans un mutant du gène *comGC* (Figure 15B,C,D) suggérant que ComGC était le constituant majeur de ce pilus. En observant ces filaments détachés des bactéries, ils ont suggéré que ces pili étaient auto-sécrétables et qu'ils avaient pu être manqués précédemment du fait de leur trop petite taille. Ils pensaient avoir observé pour la première fois des « pseudopili », qui, comparés aux authentiques T4P et bien qu'ils n'aient jamais été observés auparavant, seraient de petites structures ne s'étendant pas au-delà de la paroi bactérienne. Cependant, Laurenceau et al. dans leur publication de 2015, ont très vite réfuté cette observation en montrant clairement que ces courtes structures tressées étaient aussi relarguées par des bactéries non compétentes et/ou mutées pour leur système de pilus et qu'il s'agissait en réalité de spiroosomes (Laurenceau et al., 2015).

II.A.2.2.2. Rôle dans la transformation naturelle

Dans leur étude de 2013, Laurenceau et al. montrent que le *com pilus* est capable de lier l'ADNt à plusieurs endroits et suggèrent que ce dernier, malgré l'absence d'homologue à PilT est capable de se rétracter comme ce qui a été décrit chez *N. gonorrhoeae*. Cette hypothèse a été confirmée par une très récente étude de Lam et al. parue en 2021. Par une méthode similaire à ce qui avait permis de visualiser en temps réel le T4P de *V. cholerae* (Ellison et al., 2018, 2019a), Lam et al. montrent clairement que le pilus Com de *S. pneumoniae* lie l'ADNt et se rétracte (Figure 16). D'après cette étude, l'assemblage du pilus semble se produire au milieu de la cellule et s'étend jusqu'à 500 nm à 1 µM, ce qui est légèrement inférieur à ce qu'avait observé Laurenceau et al. en 2013 chez cette même espèce, mais toutefois, bien plus long que les structures décrites par Balaban et al. en 2014.

À la différence du T4P impliqué dans la capture de l'ADN exogène des bactéries gram négatives, les systèmes de pilus codés par les loci *com* des gram positives ne sont pourvus ni d'ATPase de rétraction, ni de sécrétine pour permettre à l'ADNt de traverser l'épaisse couche de peptidoglycane caractéristiques des espèces gram positives (Matias and Beveridge, 2005, 2006). Ainsi, comment

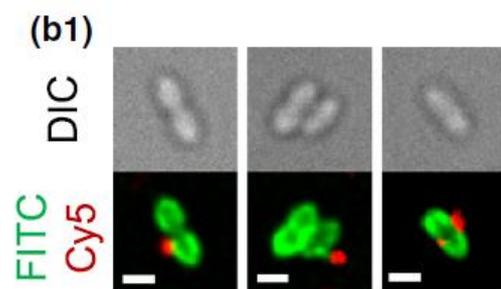


Figure 16 : Le pilus Com de *S. pneumoniae* lie l'ADN et se rétracte.

Montage d'images issues de timelapse (de gauche à droite à $t=24s$, $t=28s$ et $t=32s$) de microscopie à fluorescence d'une souche *comGC-cys* marquée avec du AF488-mal (en vert) et incubée avec de l'ADN Cy5-labeled en rouge. Echelle (barre blanche) = 1 µM. (Lam et al., 2021).

expliquer la rétraction pourtant observée du pilus ? Comment expliquer le passage du pilus et de l'ADNt à travers l'épaisse couche de peptidoglycane ?

Il existe plusieurs hypothèses pour répondre à la première question. (1) Même s'il n'existe pas d'homologue à PilT chez les gram positives, on ne peut exclure qu'une ATPase originale puisse catalyser la rétraction du pilus. Dubnau et Blokesch dans leur revue de 2019 émettent l'hypothèse que ComFA, une petite ATPase de liaison à l'ADN simple brin (ADNsb) associée à la membrane interne et impliquée dans l'internalisation de l'ADNt dans le cytoplasme, puisse jouer ce rôle (Dubnau and Blokesch, 2019). Cependant, Lam et al. doute qu'une si petite protéine puisse, en plus de sa fonction initiale, permettre la rétraction du pilus (Lam et al., 2021). (2) Lam et al. proposent comme alternative que l'ATPase d'assemblage ComGA, puisse être une ATPase réversible capable de réaliser à la fois le rôle d'ATPase d'assemblage et de rétraction, comme c'est le cas de CpaF, une ATPase nouvellement découverte chez *Caulobacter crescentus* (Ellison et al., 2019b; Lam et al., 2021). Cependant, à l'heure actuelle aucune donnée n'est disponible pour confirmer ou non cette hypothèse. (3) Une troisième hypothèse serait que la rétraction du pilus s'effectue spontanément sans avoir recourt à une ATPase. Cette hypothèse est étayée par le fait que chez *V. cholerae* ou *N. gonorrhoeae*, même en l'absence de PilT, un désassemblage est observé (Ellison et al., 2018; Zöllner et al., 2019). Cependant, cela générerait de trop faibles forces qui semblent incompatibles avec celles nécessaires pour tirer l'ADN du milieu extérieur jusque dans le compartiment apparenté au périplasme.

Pour répondre à la deuxième question, plusieurs hypothèses ont été formulées (**Figure 17**). Sur la base du long pilus rétractile observé par Laurenceau et al. en 2013 et Lam et al. en 2021, une possibilité serait que lors de son assemblage le pilus crée un accès dans la paroi bactérienne, qui est ensuite utilisé lors de sa rétraction pour amener l'ADNt jusqu'au pore d'entrée dans la membrane plasmique. Cette hypothèse est étayée par le fait que chez *S. pneumoniae*, deux enzymes lytiques impliquées dans le remodelage du peptidoglycane ont été retrouvées surexprimées durant la compétence et pourraient aider le pilus à traverser la couche de peptidoglycane (Peterson et al., 2004; Piepenbrink, 2019). Une autre hypothèse concernant ces longs pili est la possibilité qu'il existe un canal ou un transporteur non homologue à PilQ permettant le passage de l'ADNt à travers la paroi (Piepenbrink, 2019). Cependant, à ce jour, aucun élément ne permet de confirmer l'une ou l'autre de ces hypothèses. Balaban et al. dans leur publication de 2014, même s'il s'avère que leur observation est erronée, ont émis une hypothèse intéressante qui pourrait s'appliquer aux pili Com. Ils ont suggéré que ce qu'ils pensaient être des pseudopili aurait pour rôle de perforer la paroi de la bactérie, créant un trou, permettant à l'ADNt de traverser la couche de peptidoglycane jusque dans le périplasme sans l'intervention du pilus (hypothèse « *hole in the wall* »). Même si la mise en place d'un complexe macromoléculaire si sophistiqué pour uniquement induire une perforation de la paroi semble être disproportionnée, ce

modèle a l'avantage d'expliquer l'entrée de l'ADN dans le périplasma indépendamment d'un canal et/ou d'une ATPase de rétraction.

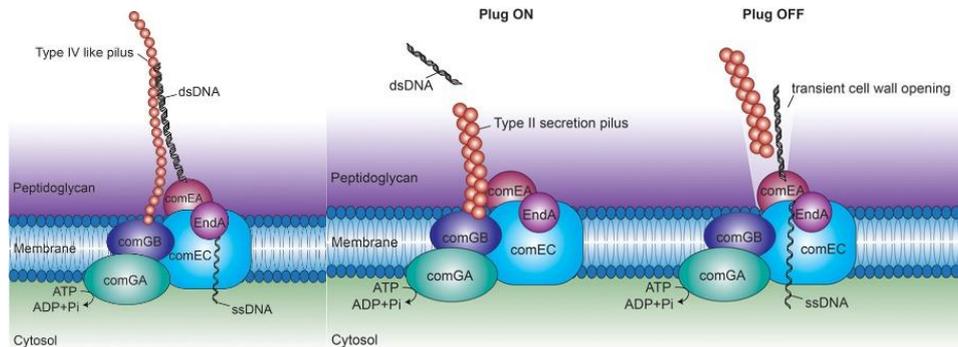


Figure 17 : Modèle de capture de l'ADNt chez *S. pneumoniae*.

(A) Hypothèse de Laurenceau et al. et Lam et al. : le pilus lie l'ADNt et se rétracte pour l'amener dans le périplasma. Hypothèse « hole in the wall » de Balaban et al. : le pilus s'autosécrète et perce la paroi pour permettre à l'ADNt de rentrer dans le périplasma. (Muschiol et al., 2015)

II.A.3. Le système de sécrétion de type 4 : le cas particulier d'*H. pylori*

H. pylori est une bactérie gram négative très étudiée pour sa capacité à persister dans l'estomac humain pendant des années. Colonisant la muqueuse gastrique de plus de la moitié de la population mondiale, c'est un pathogène majeur dont l'infection est notamment la première cause de cancer gastrique (Atherton, 2006). C'est aussi l'une des bactéries dont le génome est le plus variable s'expliquant à la fois par un taux de mutations spontanées élevé, mais aussi et surtout par des recombinaisons intra-espèces très fréquentes (Go et al., 1996; Suerbaum and Josenhans, 2007). C'est sans doute cette étonnante plasticité génomique qui lui permet de s'adapter et de survivre à son hôte et qui explique son succès en tant que pathogène.

Le principal mécanisme à l'origine de ces recombinaisons est la transformation naturelle. La transformabilité de *H. pylori* est connue depuis plusieurs décennies quand elle s'appelait encore *Campylobacter pylori* (Nedenskov-Sørensen et al., 1990), mais au fur et à mesure des recherches elle est très vite apparue comme « le mouton noir » des bactéries transformables (Johnston et al., 2014). En effet, le système de transformation d'*H. pylori* n'a nul autre équivalent dans le monde procaryote. Alors que toutes les autres bactéries transformables utilisent un système apparenté au T4P ou au SST2 pour capturer l'ADN exogène, *H. pylori* utilise un système de sécrétion de type IV (SST4) pour assurer cette fonction. En effet, aucun homologue aux systèmes de pilus décrits précédemment n'a été retrouvé dans le génome d'*H. pylori*. Initialement identifié par un crible de transformation sur une banque de mutants de gènes sécrétoires (Hofreuter et al., 1998) et caractérisé plus en détail par la suite (Hofreuter et al., 2001; Karnholz et al., 2006), on sait aujourd'hui que ce SST4 particulier nommé

ComB pour son implication dans la compétence pour la transformation, est codé par 8 gènes regroupés en deux opérons *comB2-4* et *comB6-10* numérotés sur la base de la nomenclature du SST4 Vir de *Agrobacterium tumefaciens*. Excepté la protéine codée par *comB7*, tous les composants de ce système sont indispensables à la transformation puisque la délétion de chacun de leur gène induit une absence de transformation (Fernandez-Gonzalez and Backert, 2014).

L'architecture du système est prédite sur la base des caractéristiques structurales et d'homologie de séquences avec le système Vir, et reste donc très hypothétique (Hofreuter et al., 2001, 2003; Karnholz et al., 2006). ComB2 par son homologie à VirB2, serait une sous-unité du pseudopilus associée au SST4. Elle serait présentée à la face externe de la bactérie et aurait pour rôle de faciliter la capture de l'ADN exogène lors de la transformation. Concernant ComB3 à part qu'elle est essentielle au processus de transformation, aucun rôle ne lui a pour l'heure été attribué. ComB4, en raison de sa forte homologie de séquence avec VirB4 et de la présence d'un motif de liaison à l'ATP serait la seule ATPase impliquée dans ce système. Les protéines ComB6 à 10 formeraient un large complexe qui s'étendrait de la membrane interne jusqu'à la membrane externe avec probablement ComB6, 8 et 10 ancrées dans la membrane interne et ComB7 et 9 présentes dans le périplasme ou formant un canal au niveau de la membrane externe (**Figure 18**).

On sait très peu de choses sur le rôle de ce système de SST4 ComB dans la transformation naturelle, mais des études préliminaires suivant le devenir d'un ADNt fluorescent lors du processus de transformation d'*H. pylori* a permis révéler certains aspects du processus (**Figure 18**). Une première étude a permis de montrer que le transport de l'ADNt jusqu'au périplasme se faisait en deux étapes (Stingl et al., 2010). L'ADNt est d'abord importé du milieu extérieur dans le périplasme par un mécanisme actif consommant de l'ATP et dépendant du SST4 ComB (dont possiblement son ATPase ComB4) et est ensuite transporté dans le cytoplasme par un processus dépendant du canal de la membrane interne ComEC comme c'est le cas chez les autres espèces transformables. Les auteurs suggèrent que cette translocation de l'ADNt du périplasme au cytoplasme est couplée à un système de dégradation de l'ADN double brin en ADN simple brin comme c'est le cas pour les autres systèmes de transformation (Stingl et al., 2010). Cette même étude a permis de montrer que ce processus se faisait indépendamment de la structure de l'ADNt et qu'il ne nécessitait pas de reconnaissance préalable de DUS comme c'est le cas par exemple chez *Neisseria* (Stingl et al., 2010). La découverte récente du récepteur à ADN double brin périplasmique ComH essentiel à la transformation d'*H. pylori* a permis de compléter ce modèle (Damke et al., 2019). ComH bien que ne présentant pas de similarité de séquences avec ComEA (décrit en II.B.1) jouerait un rôle similaire. Par l'intermédiaire de son domaine N-terminal, elle lierait l'ADNt arrivant dans le périplasme, possiblement en agissant comme une force supplémentaire pour l'amener dans ce compartiment. Ensuite, grâce à l'interaction de son

domaine C-terminal et de l'*oligonucleotide binding domaine* de ComEC, ComH délivrerait l'ADNt à ce dernier remplissant une fonction homologue à ComEA présente dans les autres systèmes. Les étapes menant à l'internalisation et à la recombinaison de l'ADNt sont communes aux autres espèces et décrites ci-après.

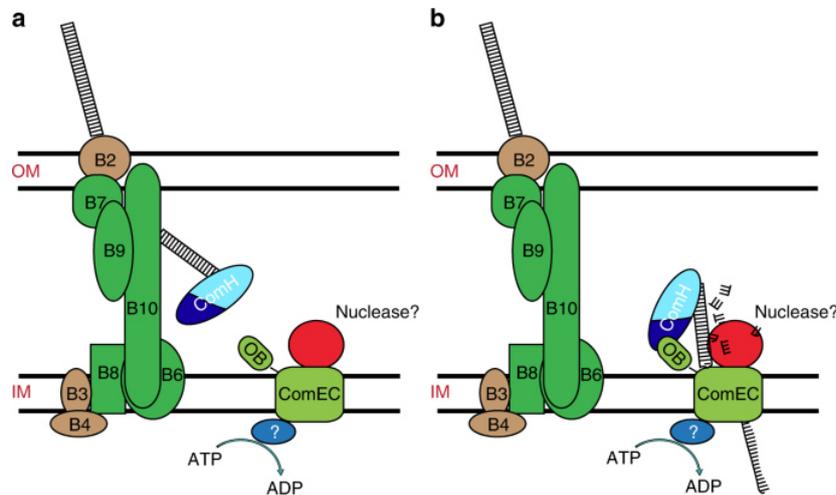


Figure 18 : Import de l'ADN par le SST4 ComB de *H. pylori*.

D'après (Damke et al., 2019). L'ADNt est lié par le pseudopilus du SST4 est amené dans le périplasma par un processus actif nécessitant de l'ATP. L'ADNt entrant serait alors lié ComH (par le domaine N-terminal, en bleu clair) qui induirait une force supplémentaire pour le tirer dans le périplasma. ComH présenterait cet ADNt à ComEC (via son domaine C-terminal, en bleu foncé)

B. Les étapes périplasmiques du processus de transformation

II.B.1. Transfert de l'ADNt dans le périplasma

Chez l'ensemble des bactéries transformables, après le passage du peptidoglycan (et de la membrane externe chez les gram négatives), l'ADNt est universellement pris en charge par un récepteur à ADN homologue à ComEA, initialement identifié chez *B. subtilis* (Inamine and Dubnau, 1995; Provvedi and Dubnau, 1999) excepté chez *H. pylori* (voir paragraphe précédent). ComEA et ses homologues possèdent en N-terminal une séquence signal d'adressage au périplasma et deux domaines hélice-tour-hélice (HTH) conservés qui sont bien connus pour permettre la liaison non spécifique à de l'ADN double brin (Doherty et al., 1996). Chez les bactéries gram négative ComEA est soluble dans le périplasma (Dubnau and Blokesch, 2019) tandis qu'au moins chez les gram positives *B. subtilis* et *S. pneumoniae* la protéine mature est pourvue en N-terminal d'un domaine transmembranaire lui permettant de s'ancrer dans la membrane plasmique (Provvedi and Dubnau, 1999). La délétion de *comEA* chez *S. pneumoniae* ou *B. subtilis* entraîne une absence totale de transformants suggérant que ComEA est absolument requise pour la transformation. Étonnamment,

la délétion de *comEA* entraîne différents phénotypes chez les bactéries gram négatives. Tandis qu'un mutant *comEA* de *V. cholerae* est totalement déficient pour la transformation (Meibom et al., 2005), un mutant *comE* de *N. gonorrhoeae* (Chen and Gotschlich, 2001), de *comE1* de *H. influenzae* (Sinha et al., 2012) ou *comEA* de *L. pneumophila* (Charpentier et al., 2008) n'est que sévèrement ou partiellement déficient pour la transformation respectivement suggérant que chez ces espèces il existe une autre protéine capable de suppléer sa fonction. De façon intéressante, des expériences de complémentation d'un mutant *comEA* chez *V. cholerae* par ses homologues d'autres espèces (et notamment de *B. subtilis*) suffit à restaurer la transformabilité de *V. cholerae* suggérant un mécanisme d'action conservé de ComEA chez les bactéries transformables (Seitz et al., 2014).

Le rôle de ComEA dans la transformation naturelle a principalement été étudié chez *N. gonorrhoeae*, *V. cholerae* et *B. subtilis* qui servent respectivement de modèles pour les bactéries gram négatives et positives. Chez *V. cholerae*, des études de localisation de fusions de ComEA à des protéines fluorescentes ont permis de montrer qu'en l'absence d'ADNt, ComEA-mCherry était distribuée de façon uniforme dans le périplasme et qu'une fois en présence d'ADNt, elle formait des *foci* colocalisés à l'ADNt marqué (Seitz et al., 2014) (**Figure 19A**). Des résultats similaires ont été observés pour ComE chez *N. gonorrhoeae* (Gangel et al., 2014). De plus, une méthode de détection de l'ADNt par PCR sur cellules après exposition successive des bactéries à l'ADNt et à la DNase I, montrent que l'ADNt n'est pas détecté dans les cellules alors qu'il est détecté dans une souche sauvage ou dans un mutant *comEC* (Seitz et al., 2014). Ces données suggèrent que ComEA est recrutée au niveau du point d'entrée de l'ADNt (à la base du pilus) et qu'il participe à la capture de l'ADNt. Chez *B. subtilis*, des expériences similaires ont permis de montrer que contrairement à *V. cholerae*, ComEA était plutôt localisé sous la forme de *foci* distribués dans l'intégralité du périplasme et que cette distribution ne changeait pas avec la présence ou non d'ADNt (Hahn et al., 2021; Kaufenstein et al., 2011)(**Figure 19B**). Des expériences de cinétiques en microscopie 3D à fluorescence ont révélé que chez un mutant *comEA*, l'ADNt marqué pouvait se fixer à la surface de la bactérie, mais pas à entrer dans l'espace apparenté au périplasme suggérant que tout comme chez *V. cholerae*, ComEA est absolument requis, en plus du pilus, pour permettre l'entrée de l'ADNt dans ce compartiment (Hahn et al., 2021).

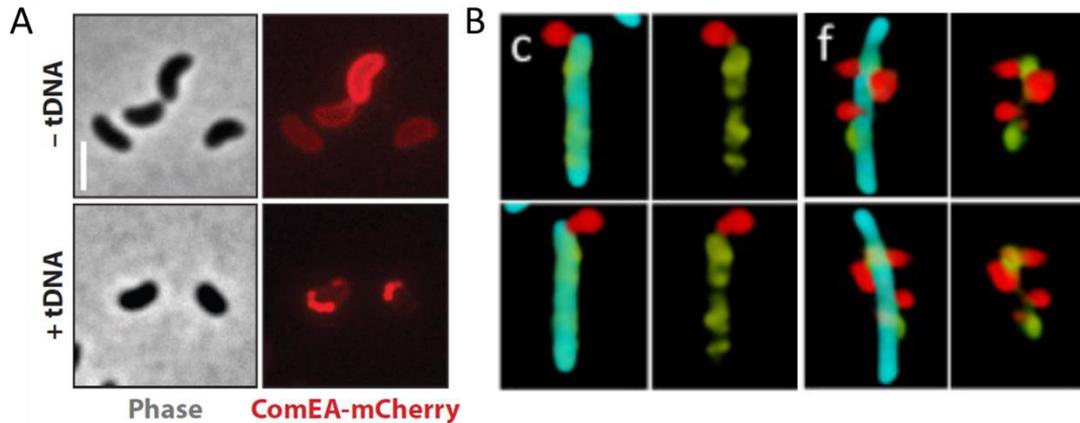


Figure 19 : Localisation de ComEA dans les cellules compétentes.

Figure A, issue de Dubnau et Blokesch, 2019 : localisation de ComEA-mcherry chez *V. cholerae* en présence ou en absence d'ADNt. ComEA-mcherry est réparti uniformément dans le périplasme en absence d'ADNt (photos du haut) et s'accumule au site d'entrée de l'ADNt en sa présence formant des *foci*. Figure B, issue de Hahn et al., 2021 : images tridimensionnelles (3D) déconvoluées et reconstruites en volume de cellules compétentes de *B. subtilis* (marqué à la CFP (cyan) qui délimite le cytoplasme) exprimant la YFP-ComEA (en jaune) en présence d'ADNt marqué à la rhodamine (en rouge). C : Exemple d'une cellule compétente dans différentes orientations (images du dessus ou du dessous) après 1 min d'incubation en présence d'ADNt, celui-ci est localisé au niveau des pôles de la cellule (le pilus étant situé aux pôles chez *B. subtilis*) et semble faiblement associé à YFP-ComEA qui est localisée sous forme de *foci* répartis dans la majeure partie du périplasme. F : Exemple d'une autre cellule compétente après 10 min d'incubation avec l'ADNt. Pas de modifications de l'organisation des foci de ComEA, l'ADNt semble lui, par contre, plus intimement associé aux *foci* de YFP-ComEA et parfois situé loin du pôle (observations confirmées sur différentes autres cellules compétentes).

Des études biophysiques menées chez *N. gonorrhoeae* utilisant des techniques à l'échelle d'une molécule unique pour caractériser la vitesse et la force d'absorption de l'ADN ont montré qu'un ADN de 10 kB (3,4 μ M) était importé en quelques secondes sans interruption dans le périplasme (Hepp and Maier, 2016, 2017). Cette forte processivité et l'absence de pause est incohérente avec un modèle dans lequel le pilus serait le seul responsable de l'import de l'ADNt dans le périplasme. En effet, le pilus subit des cycles d'assemblages et de rétraction qui impliquent nécessairement des pauses lors de l'importation de l'ADN exogène. Par ailleurs, en montrant que la vitesse d'absorption de l'ADNt est proportionnelle à la concentration périplasmique de ComE, les auteurs suggèrent plutôt que c'est ComE qui remplirait ce rôle via un mécanisme à cliquet brownien (ie, *brownian ratchet mechanism*) (Hepp and Maier, 2016, 2017). C'est aussi l'hypothèse envisagée chez les autres organismes modèles (Hahn et al., 2021; Seitz et al., 2014).

Sur la base de ces observations et hypothèses, un modèle a été proposé chez *V. cholerae* et peut être appliqué aux autres bactéries transformables gram négatives (Figure 20) (Matthey and Blokesch, 2016; Seitz et al., 2014). Une fois l'ADNt lié, le pilus, en se rétractant, aurait uniquement pour rôle de permettre l'ouverture du canal PilQ et le passage d'une portion limitée d'ADNt à travers ce pore ; assez

pour permettre sa liaison par ComEA via l'un de ses domaines HhH. ComEA permettrait la stabilisation de l'ADNt entrant et rendrait irréversible sa capture par le pilus en tirant l'ADNt dans le périplasma via un mécanisme à cliquet brownien. La capacité d'accumulation de ComEA permettrait de compacter l'ADNt au sein du périplasma qui constituerait alors un réservoir à ADNt directement utilisable par ComEC (Gangel et al., 2014; Kaufenstein et al., 2011).

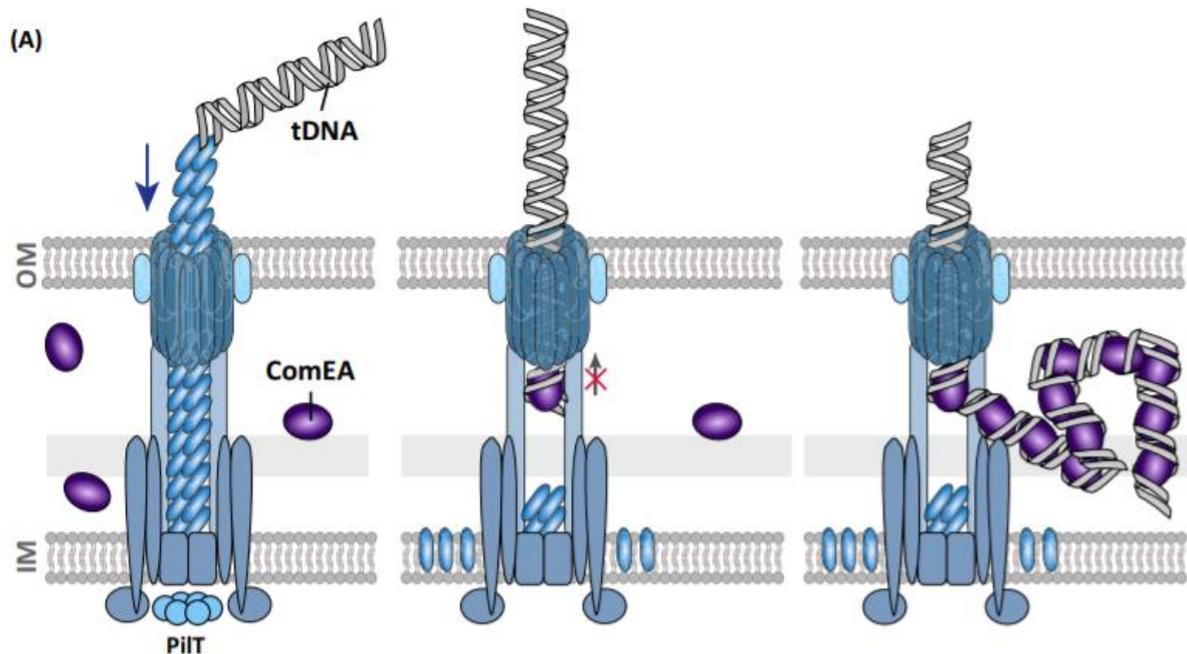


Figure 20 : Modèle d'action de ComEA au cours de la transformation naturelle.

Le T4P lie l'ADNt exogène et se rétracte (flèche bleue, schéma de gauche). La rétraction du pilus permet d'ouvrir le canal pilQ à travers lequel passe une petite portion de l'ADNt qui devient accessible pour le récepteur à ADN ComEA. ComEA lie l'ADNt via son domaine HhH et le tire via un mécanisme à cliquet brownien empêchant une rétrodiffusion de l'ADNt dans le compartiment extracellulaire (flèche noire, schéma du centre). ComEA est capable de s'accumuler au point d'entrée de l'ADNt favorisant son compactage au sein du périplasma (schéma de droite). D'après (Matthey and Blokesch, 2016).

II.B.2. Fragmentation de l'ADNt dans le périplasma

ComEA délivre probablement l'ADNt à une nucléase qui dégrade l'ADN double brin (ADNdb) en ADN simple brin (ADNsb). En effet, on sait depuis bien longtemps que pour que l'ADNt soit transporté dans le cytoplasme et intégré par recombinaison homologe, il doit être converti en ADNsb (Chaussee and Hill, 1998; Lacks, 1962; Piechowska and Fox, 1971). Chez *S. pneumoniae*, on a très tôt découvert que la nucléase responsable de cette conversion est EndA (Lacks et al., 1974). Il a été montré que EndA était une enzyme attachée à la membrane plasmique (côté périplasmique) et que sa délétion induisait un défaut de transformation de l'ordre de 3 à 4-log par rapport à la souche parente (Bergé et al., 2002; Puyet et al., 1990). Le suivi d'un ADNt radiomarqué au cours du processus de transformation chez *S. pneumoniae*, a permis de montrer que (1) le brin non transporté n'était pas dégradé en l'absence de

ComGA, ComGC et ComEA, ce qui suggère que le pilus est indispensable pour amener l'ADNt au sein du périplasma et que ComEA est indispensable pour amener l'ADNt auprès de EndA, mais que (2) ce brin était bien dégradé en l'absence de ComEC ou ComF, ce qui suggère que le récepteur à ADNt ComEA et le transport de l'ADNt dans le cytoplasme fonctionnent de manière indépendante (Bergé et al., 2002). Contrairement aux autres gènes impliqués dans la transformation naturelle de *S. pneumoniae*, *endA* est exprimé de façon constitutive. Récemment Bergé et al. (2013) ont observé que chez les bactéries non compétentes, EndA était localisée de façon uniforme au niveau de la membrane plasmique et qu'une fois devenues compétentes, EndA était recrutée au milieu de la cellule dans un processus dépendant de ComEA. Les auteurs suggèrent que ComEA permettrait de moduler l'activité de EndA pour qu'un seul brin soit dégradé. Toujours, chez *S. pneumoniae*, en utilisant des ADNt radiomarqués en 5' ou en 3' d'un seul des deux brins, il a été montré que le marquage en 3' n'était pas internalisé, ce qui suggère que l'ADNt simple brin rentre dans le cytoplasme avec une polarité de 3' → 5' et que l'autre brin est dégradé avec une polarité inversée (Méjean and Claverys, 1988).

Bien qu'étant un acteur clé du système de transformation, aucun homologue à EndA, n'a été identifié chez les autres espèces transformables. Seul chez *B. subtilis*, une autre nucléase NuCA induisant la cassure double brin de l'ADNt, a été identifiée, mais elle n'est pas responsable de la conversion de l'ADNdb en ADNsb (Provvedi et al., 2001).

Très récemment, il a été montré que chez certaines de ces espèces, la large protéine polytopique transmembranaire ComEC pouvait exercer ce rôle de nucléase en plus de sa fonction initiale (Baker et al., 2016; Silale et al., 2021).

II.B.3. Transport de l'ADNt dans le cytoplasme

II.B.3.1. Rôle de ComEC

Chez l'ensemble des bactéries transformables, le transport de l'ADNt à travers la membrane plasmique, sous sa forme simple brin, requiert obligatoirement la protéine ComEC. En effet, la délétion de *comEC* abolit systématiquement la transformation naturelle comme c'est le cas par exemple chez *A. baylyi* (Friedrich et al., 2001), *H. pylori* (Damke et al., 2019), *N. gonorrhoeae* (Facijs and Meyer, 1993), *V. cholerae* (Matthey and Blokesch, 2016), *S. pneumoniae* (Bergé et al., 2002), *B. subtilis* (Inamine and Dubnau, 1995), ou encore chez *L. pneumophila* (Voir Chapitre II partie II.A, Hardy et al., 2021).

Une analyse topologique *in silico* des séquences protéiques de ComEC chez *B. subtilis*, *V. cholerae*, *S. pneumoniae*, *Acinetobacter spp* ou encore *N. gonorrhoeae* suggère que c'est une large protéine

transmembranaire polytopique constituée de 9 à 13 segments transmembranaires et deux domaines extramembranaires N et C terminal (Dubnau and Blokesch, 2019; Tsirigos et al., 2015).

Des analyses évolutives (Pimentel and Zhang, 2018) ont montré que l'ensemble des protéines ComEC possédait un domaine central très conservé appelé « domaine de compétence ». Chez *B. subtilis* des études de topologie expérimentale avec des fusions *lacZ* ou *phoA* ont montré que ce domaine était constitué de trois segments transmembranaires et d'une hélice amphipathique cytoplasmique (Draskovic and Dubnau, 2005) (**Figure 21A**). À la vue de la forte conservation de ce domaine au sein des différentes espèces, ce domaine de compétence est présumé former le canal par lequel l'ADNt est transporté (Draskovic and Dubnau, 2005). Dans cette même étude, d'après leur analyse topologique, Draskovic and Dubnau propose un modèle dans lequel ComEC dimériserait pour former un canal (**Figure 21B**).

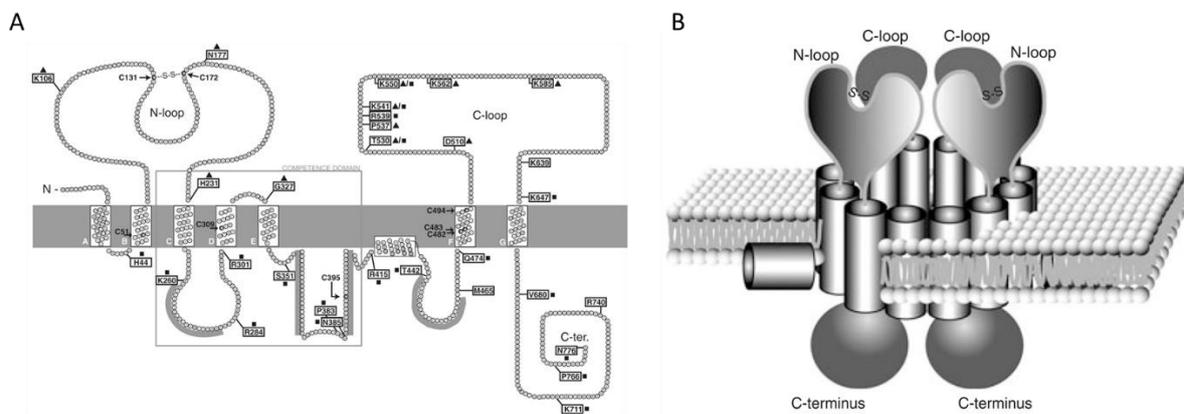


Figure 21 : Topologie de ComEC de *B. subtilis* (A) et modèle de ComEC comme canal de l'ADNt (B). Extrait de (Draskovic and Dubnau, 2005). D'après des études topologiques, ComEC chez *B. subtilis* est une protéine constituée de 7 fragments transmembranaires (actuellement le modèle envisagé est plutôt 11 domaines transmembranaires (Pimentel and Zhang, 2018)). Elle est constituée de trois domaines : un domaine central nommé « domaine de compétence » (encadré) bordé par deux domaines extramembranaires N-terminal (N-loop) et C-terminal. Le domaine N-terminal possède un pont disulfide. Le domaine C-terminal peut-être divisé en deux parties : une partie périplasmique (C-loop) et une partie cytoplasmique (extrémité C-terminal). (B) d'après leur étude topologique Draskovic and Dubnau, 2005 propose un modèle dans lequel ComEC dimériserait pour former un canal.

Récemment, des analyses *in silico* ont montré que son domaine N-terminal soluble contenait un domaine « *oligonucleotide binding fold* » (OBD) de liaison à l'ADN simple brin (ADNsb) (Baker et al., 2016). Draskovic et Dubnau (2005) ont établi que ce domaine était extracellulaire (périplasmique) et que sa mutation abolissait la transformation sans affecter la stabilité du canal. D'après ces résultats, Baker et al. (2016) suggèrent que ce domaine OBD permettrait la manipulation de l'ADNss par ComEC possiblement pour le réceptionner auprès de ComEA et permettre son passage au travers du canal, sans toutefois que de preuves expérimentales ne soit encore disponibles.

In silico, il a aussi été montré que le domaine soluble C-terminal était homologue aux protéines de la superfamille des métallo- β -lactamases (Baker et al., 2016). Ce sont des protéines capables d'hydrolyser par clivage un certain nombre de substrats et notamment des acides nucléiques (Hall and Barlow, 2004). D'après les données topologiques, ce domaine C-terminal aurait une partie extracellulaire (périplasmique) et une partie intracellulaire cytoplasmique (Draskovic and Dubnau, 2005). Baker et al. (2016) suggèrent que c'est la partie extracellulaire qui porte l'activité catalytique métallo- β -lactamase et proposent que ce domaine soit la nucléase responsable de la dégradation de l'ADNdb en ADNsb pour passer dans le cytoplasme. Dans une très récente étude (Silale et al., 2021), cette hypothèse a été validée expérimentalement chez *B. subtilis*. Silale et al. (2021) ont purifié le domaine C-terminal de la bactérie gram positive *Moorella thermoacetica* (MthCTD) et ont montré qu'elle avait une activité nucléase *in vitro* capable de dégrader des plasmides surenroulés, des ADNsb et des ADNdb linéaires. De façon importante, le mutant catalytique de deux sites putatifs aspartate de liaison au Mn^{2+} de MthCTD (MthCTD-DADA) ont été aussi produits et présente une activité nucléase réduite. Ces données suggèrent que le domaine C-terminal de ComEC de *M. thermoacetica* est une nucléase dépendante du Mn^{2+} . Le fait qu'elle puisse dégrader un plasmide suggère que c'est une endonucléase. En substituant génétiquement les sites putatifs aspartates de liaison au Mn^{2+} chez *B. subtilis*, ils ont montré que le mutant présentait un défaut de transformation (diminution d'efficacité de 10 à 100x) par rapport à la WT suggérant que l'activité nucléase est requise pour la transformation naturelle. Cela rappelle les données obtenues pour un mutant *endA* chez *S. pneumoniae* dont la transformabilité est réduite de 1000 à 10 000 fois par rapport à la *wt* (Puyet et al., 1990)

Les auteurs expliquent que le fait que la transformation ne soit pas totalement abolie (contrairement à un mutant ComEC) chez ce mutant catalytique pourrait être dû au fait que (1) le CTD muté sur la base des données de *M. thermoacetica* exerce malgré tout une activité de nucléase résiduelle chez *B. subtilis* ou que (2) l'ADNt présente une extrémité simple brin qui peut être manipulée par le domaine OBD de ComEC et que la traction exercée par ComFA (voir ci-après) serait suffisante pour forcer le passage de l'ADNt et extruder le brin non transformant lors du passage par le canal (Silale et al., 2021).

L'hypothèse de ComEC et son domaine CTD comme nucléase responsable de la conversion de l'ADNsb en ADNdb ne serait cependant pas généralisable à l'ensemble des bactéries transformables, car près de la moitié de ces espèces ne possèdent pas de CTD équivalent à celui de *B. subtilis* (Pimentel and Zhang, 2018). Ainsi pour ces organismes, l'identification d'une telle nucléase reste toujours à déterminer.

II.B.3.2. Les acteurs pivots du système de transformation

D'autres acteurs participent probablement au transport de l'ADNt dans le cytoplasme et sont susceptibles de jouer le rôle d'intermédiaires entre les machineries d'internalisation et de recombinaison de l'ADNt. Chez *B. subtilis*, ces acteurs potentiels sont codés par un seul et même opéron : l'opéron *comF*. Cet opéron code pour trois protéines : ComFA, ComFB et ComFC, toutes jouant un rôle dans la transformation naturelle avec des implications plus ou moins fortes (Londoño-Vallejo and Dubnau, 1993). Bien que leur implication dans ce mécanisme de THG ait été démontrée il y a plusieurs décennies maintenant, ce n'est que très récemment que le rôle de ComFA et ComFC dans ce mécanisme a été établi (Damke et al., 2021, preprint; Diallo et al., 2017). ComFB a été démontrée comme non impliquée dans cette machinerie avec probablement un rôle de régulateur (Sysoeva et al., 2015).

ComFA est présente uniquement chez les bactéries gram positives. Elle est induite durant la compétence de *S. pneumoniae* et *B. subtilis* au moins. Son absence chez *S. pneumoniae* induit une perte totale de transformabilité alors que celle-ci est affectée moins drastiquement chez *B. subtilis* (de l'ordre de 10 000 fois) (Bergé et al., 2002; Londoño-Vallejo and Dubnau, 1993). Pour autant, il a été montré qu'un mutant *comFA* de *S. pneumoniae* était toujours capable de lier et de dégrader l'ADNt (Bergé et al., 2002). Avant que son rôle dans la transformation naturelle ne soit proposé, des analyses bio-informatiques ont permis de montrer qu'elle appartenait à la famille des hélicases/translocases à motif DEAD. De plus, il a été montré qu'elle co-localisait et/ou interagissait dans le cytoplasme avec des protéines aussi bien de la machinerie d'import telles que ComGA, ComFC ou ComEC que des protéines de la machinerie de recombinaison telles que DprA ou SsbB (Hahn et al., 2005; Kaufenstein et al., 2011; Kramer et al., 2007). Récemment ces données de colocalisation et d'interaction ont été partiellement confirmées par double hybride bactérien (Diallo et al., 2017). Il a été montré que ComFA interagissait avec elle-même, DprA et ComFC. Collectivement, ces données suggèrent que ComFA est située à l'interface entre la machinerie d'import et de recombinaison. Dans cette même étude de Diallo et al. (2017), des tests d'activité ATPase et de liaison à des substrats ADN ont permis de montrer que ComFA liait l'ADNsb et que son activité ATPase était stimulée en présence d'ADNsb. Ces résultats ont permis aux auteurs d'établir un modèle dans lequel ComFA permettrait de fournir l'énergie nécessaire pour tracter l'ADNsb à travers le canal ComEC dans la membrane plasmique (Diallo et al., 2017) (**Figure 22**). Chez les gram négatives, pour l'heure, aucune protéine de ce type n'a été identifiée, pourtant il est fortement suggéré que l'internalisation de l'ADN dans le cytoplasme est aussi un processus actif (Johnston et al., 2014).

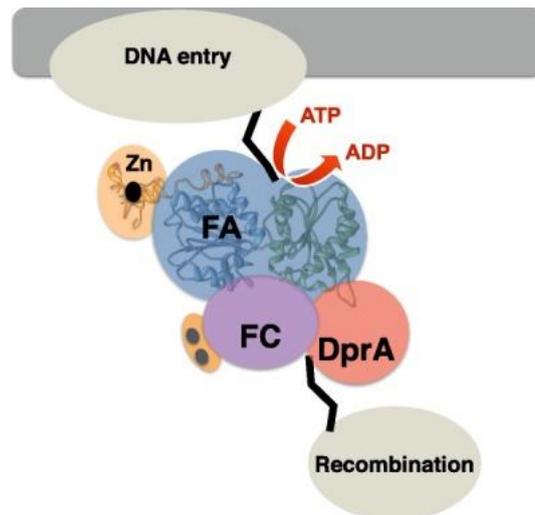


Figure 22 : Role de ComFA et ComFC dans la transformation naturelle.

ComFA et ComFC sont situées à l'interface entre machinerie d'internalisation et de recombinaison de l'ADNt chez les gram positives

ComFC chez les bactéries gram positives et son orthologue ComF chez les bactéries gram négatives, est présente chez l'ensemble des bactéries transformables. Sa délétion chez *S. pneumoniae* (Diallo et al., 2017), *V. cholerae* (Seitz and Blokesch, 2013b) ou encore *L. pneumophila* (Voir Chapitre II partie II.A, Hardy et al., 2021) abolit totalement leur transformabilité alors qu'elle n'est que modérément affectée chez *B. subtilis* (Londoño-Vallejo and Dubnau, 1994). Tout comme ComFA, ce n'est que très récemment qu'un rôle dans la transformation naturelle a été proposé pour ComF (Damke et al., 2021). En étudiant le devenir d'un l'ADNt fluorescent au cours de la transformation naturelle de *H. pylori*, Damke et al., (2021) ont montré par microscopie à fluorescence que l'ADNt était bien capturé et transporté dans le périplasma chez un mutant *comF* mais que de la même manière qu'un mutant *comEC*, l'ADNt restait bloqué dans ce compartiment (*foci* persistant), suggérant tout comme ComEC que ComF est nécessaire au transport de l'ADNt à travers la membrane interne. Des résultats similaires ont été décrits chez *V. cholerae* par une autre méthode de suivi de l'ADNt par détection par PCR (Seitz and Blokesch, 2013b). Damke et al. (2021, préprint) ont montré que ComF était associé au côté cytoplasmique de la membrane interne chez *H. pylori* et pour pouvoir étudier le devenir de l'ADNt au sein du cytoplasme, ils ont testé la transformabilité d'un mutant *comF* en électroporant un ADNt simple brin afin d'outrepasser la barrière que constitue la membrane interne chez un mutant *comF*. Dans ces conditions, un mutant *comF* présente une transformabilité similaire à un mutant *dprA* bien connu pour être impliqué dans les étapes cytoplasmiques de la transformation. Il a aussi été montré que ComF était capable de lier l'ADNss avec une forte affinité et que son équivalent ComFC chez *S. pneumoniae* au moins, interagissait avec DprA (Diallo et al., 2017). Ainsi, d'après ces résultats Damke et al. (2021) suggèrent que ComF, en se liant à l'ADNss une fois entré dans le cytoplasme, permettrait de faire l'intermédiaire entre la machinerie d'import de l'ADNt et le

système cytoplasmique de recombinaison dédiée à son intégration dans le chromosome grâce à son interaction avec DprA.

C. Les étapes cytoplasmiques du processus de transformation naturelle

Les étapes cytoplasmiques du processus de transformation naturelle font intervenir de nombreux acteurs qui sont, pour la plupart, conservés au sein des bactéries transformables. L'élément principal de ces voies menant à la recombinaison et l'intégration de l'ADNt dans le chromosome est la recombinase RecA. Elle est assistée par un certain nombre de protéines qui ont pour la plupart été caractérisées chez les bactéries gram positives *B. subtilis* et *S. pneumoniae* et qui peuvent être regroupées en 3 classes : les protecteurs, les médiateurs et les effecteurs (Kidane et al., 2012).

II.C.1. Protection de l'ADNsb

Une fois entré dans le périplasme, l'ADNsb est pris en charge par des protéines qui vont permettre sa protection. Chez *B. subtilis*, il est fortement présumé que les premières protéines à venir se fixer sur l'ADNsb sont les *single-stranded binding proteins* (SSBs) SsbA et/ou SsbB (aussi appelé YwpH)(Lindner et al., 2004; Yadav et al., 2012). SsbA est une protéine essentielle à la viabilité cellulaire qui possède une grande similarité de séquence avec le domaine N-terminal de liaison à l'ADN et C-terminal de liaison de la protéine SSB de *E. coli*, largement étudiée pour son implication dans la maintenance du chromosome (Lohman and Ferrari, 1994). SsbB, quant à elle, est une plus petite protéine qui possède 63% d'identité de séquence avec le domaine N-terminal de liaison à l'ADN de SsbA, mais ne possède pas de domaine C-terminal de liaison aux protéines chez *B. subtilis*. Contrairement à SsbA, SsbB est uniquement impliqué dans la transformation naturelle et ne possède pas d'homologue chez *E. coli*. SsbA et SsbB sont toutes deux induites durant la compétence de *B. subtilis* contrairement à *S. pneumoniae*. En effet, bien que les deux protéines soient présentes, chez *S. pneumoniae* seule SsbB est induite spécifiquement durant la compétence, SsbA étant exprimé de façon constitutive (Claverys et al., 2009; Grove et al., 2005). Chez cette espèce, le complexe nucléoprotéique dans lequel est retenu l'ADNsb dans les premières étapes cytoplasmiques est appelé complexe d'éclipse (Morrison, 1977). SsbB en est le constituant majeur (Attaiech et al., 2011; Morrison et al., 2007).

En fonction des études, il existe une grande hétérogénéité dans l'ampleur des défauts de transformation associés à une délétion de *ssbB* que ce soit chez *B. subtilis* ou *S. pneumoniae*. Dans tous les cas, l'absence de SsbB réduit modérément l'efficacité de transformation [en fonction des études de 3 à 30 fois pour *S. pneumoniae* (Attaiech et al., 2011; Bergé et al., 2003; Campbell et al., 1998;

Morrison et al., 2007) et de 5 à 50 fois chez *B. subtilis* (Berka et al., 2002; Lindner et al., 2004)] suggérant qu'il existe d'autre(s) acteur(s) impliqué(s) dans la protection de l'ADNsb. L'étude de SsbA par des méthodes de génétique conventionnelles est rendu difficile par son essentialité pour la viabilité bactérienne ; il a cependant été montré qu'elle faisait partie intégrante du complexe d'éclipse chez *S. pneumoniae* (Morrison et al., 2007). Les bactéries gram négatives transformables ne codent que pour une seule *single-stranded binding protein* SSB dont les similarités de séquences suggèrent qu'il s'agit d'un homologue à SsbA. Cependant, tout comme SsbA, SSB est essentielle à la viabilité cellulaire et pour cette raison, son implication dans la transformation naturelle de ces espèces reste à clarifier.

SsbB et SsbA forment des homotétramères en solution et lient l'ADNsb avec une forte affinité (Yadav et al., 2012). Les structures 3D de ces protéines ont été résolues chez diverses espèces et montrent une architecture globale conservée et une surface de liaison à l'ADNsb similaire (Chan et al., 2009; Raghunathan et al., 2000; Yadav et al., 2012). Chez *S. pneumoniae*, il a été établi que SsbB était 20 fois plus abondante que SsbA dans les cellules compétentes avec environ 70 000 molécules par cellule, ce qui représente une capacité de couverture de 1,15 Mb d'ADNsb (Attaiech et al., 2011).

Ces deux protéines accomplissent diverses fonctions au cours du processus de transformation naturelle. Tout d'abord, l'implication de SsbB dans la protection de l'ADNsb entrant a clairement été démontrée par des cinétiques de dégradation d'ADNsb radiomarqué chez *S. pneumoniae* (Attaiech et al., 2011). Il a été montré que dans un mutant Δ ssbB compétent, la quantité d'ADNsb radiomarqué décroissait beaucoup plus vite au cours du temps que chez une souche *WT* et que cette protection n'était pas dépendante de la partie C-terminale de la protéine. Au contraire, chez un mutant *ssbB* Δ 7 (déléte de cette partie C-terminal), il a été montré que les quantités d'ADNsb récupérées au cours du temps sont toujours plus importantes que pour la *WT*. Cependant, lorsqu'un mutant *ssbB* Δ 7 ou Δ ssbB est transformé avec une faible quantité d'ADNt, les fréquences de transformation pour ces deux mutants sont similaires et sont réduites par rapport à la souche sauvage. Puisque la délétion entière du gène *ssbB* et la délétion de la région codant la partie C-terminale de SsbB (mutant *ssbB* Δ 7), ont des effets opposés dans la stabilisation de l'ADNsb, mais entraînent des phénotypes de transformation similaires, les auteurs suggèrent qu'en plus d'être impliquée dans la stabilité et la protection de l'ADNsb, SsbB via son domaine C-terminal serait aussi impliquée dans le processing de l'ADNsb en recombinants en interagissant avec des protéines importantes pour la recombinaison de l'ADNsb (Attaiech et al., 2011).

De façon intéressante, dans cette même étude, les auteurs ont montré que plus la quantité d'ADNt à cible chromosomique donné à un mutant Δ ssbB de *S. pneumoniae* était importante, plus le défaut de transformation (comparé à la *WT*) augmentait. A contrario, ils ont montré que chez un mutant *sbbB*,

lorsque cet ADNt était un plasmide, la fréquence de transformation augmentait avec la concentration d'ADNt dans les mêmes proportions que pour une souche WT. Les auteurs en ont donc déduit que SsbB avait un rôle crucial dans la transformation chromosomique notamment à de fortes concentrations d'ADNsb, tout en ayant un rôle antagoniste sur la transformation plasmidique (Attaiech et al., 2011) (Figure 23). Par ailleurs, cette variation des fréquences de transformation en fonction de la concentration et de la nature de l'ADNt explique pourquoi il existe une telle hétérogénéité des phénotypes de transformation associés à une délétion de *ssbB* en fonction des différentes études.

Au vu de ces données et de l'abondance de SsbB dans les cellules compétentes, les auteurs suggèrent que la « raison d'être » de SsbB chez *S. pneumoniae* est de constituer un réservoir d'ADNsb directement utilisable pour permettre des recombinaisons successives d'ADNsb dans le chromosome (Attaiech et al., 2011) (**Figure 23**). Récemment, ces multiples événements de recombinaison ont été directement observés chez *S. pneumoniae* par microscopie en temps réel sur cellules vivantes (Kurushima et al., 2020). Puisque le nombre estimé de molécules de SsbB aurait la capacité de stocker l'équivalent de presque la moitié de son génome en ADNsb (1,15 Mb sur une taille totale du génome de 2,35 Mb) et sous réserve qu'une cellule bactérienne puisse emmagasiner tout cet ADNsb, la présence et l'abondance de SsbB pourraient expliquer à elles seules l'importante plasticité du génome de cette espèce. Il est fort probable qu'une grande partie de l'ADNsb présent dans ce réservoir soit issu de cellules sœurs ; le fratricide, phénomène où les cellules compétentes provoquent la lyse des non compétentes, étant courant chez *S. pneumoniae* (Claverys et al., 2007). Ainsi, Attaiech et al. (2011) suggèrent que cet ADNsb pourrait aussi servir de matrice pour la réparation des cassures chromosomiques. Au vu de l'hétérogénéité des phénotypes de transformation aussi constatés chez un mutant *ssbB* de *B. subtilis*, il est probable que le rôle de SsbB chez cette espèce soit similaire, d'autant que SsbB colocalise avec le pore d'entrée de l'ADNsb chez cette espèce (Kramer et al., 2007).

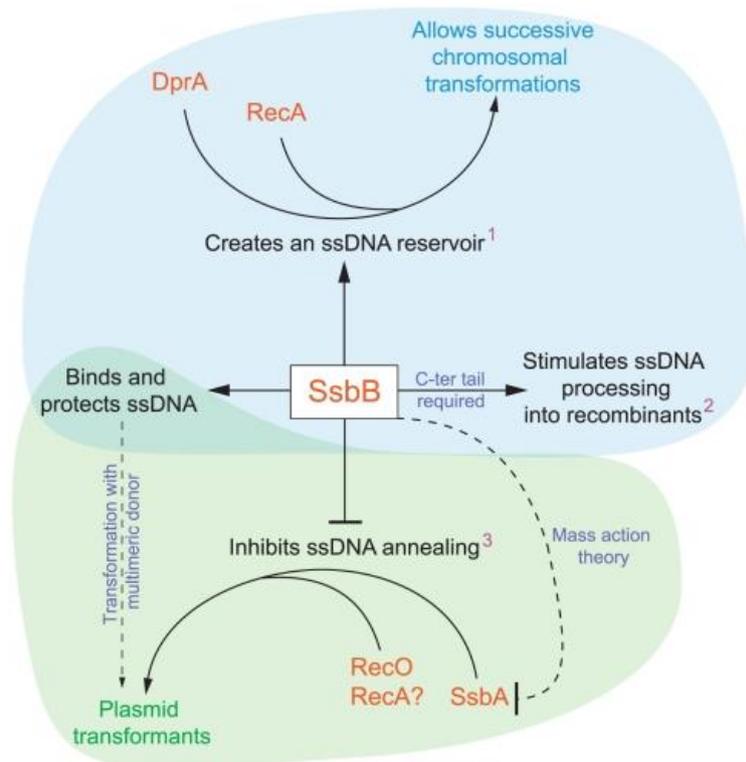


Figure 23 : Bilan du rôle de SsbB dans la transformation de *S. pneumoniae*.
 Extrait de (Attaiech et al., 2011). Voir texte.

Chez *B. subtilis* et *S. pneumoniae*, un mutant *ssbB* reste déficient pour la transformation alors même que *SsbA* est présente, ce qui suggère qu’au lieu de suppléer sa fonction, *SsbA* aurait d’autres rôles, seule ou en association avec *SsbB*. Chez *S. pneumoniae*, *SsbA* est retrouvée, de façon minoritaire, dans le complexe d’éclipse suggérant qu’elle pourrait aussi être impliquée dans la protection de l’ADNsb (Claverys et al., 2009; Morrison et al., 2007). En association, chez *B. subtilis*, il a été montré que *SsbA* et *SsbB* permettaient d’enlever les structures secondaires présentes sur l’ADNsb et limiteraient l’initiation de l’assemblage de *RecA* (appelée « nucléation ») sur ces nucléofilaments (Yadav et al., 2012, 2013). En ce sens, ces SSBs agiraient comme une barrière pour limiter les liaisons non productives de *RecA* sur les nucléofilaments. Cette barrière serait levée par l’interaction de ces SSBs avec des protéines qui, au contraire, faciliterait la liaison de *RecA* aux nucléofilaments de SSBs et qu’on appelle médiateurs de la recombinaison.

II.C.2. Médiation de la recombinaison de l’ADNt

L’un des médiateurs les plus importants intervenant dans la recombinaison de l’ADNt est la *DNA processing protein* *DprA* aussi anciennement appelée *Smf* chez *B. subtilis*. *DprA* a d’abord été identifiée chez *H. influenzae* par un crible de transformation ayant permis d’identifier un mutant fortement déficient pour la transformation, mais capable d’internaliser l’ADNt (Karudapuram et al., 1995; Tomb et al., 1989). Depuis, l’implication de *DprA* dans la transformation naturelle a été déterminée chez un

très grand nombre d'espèces gram positives et négatives : tandis que l'absence de DprA abolit totalement la transformation d'un mutant de *S. pneumoniae* (Bergé et al., 2003), de *N. meningitidis* (Hovland et al., 2017), de *N. gonorrhoeae* (Duffin and Barber, 2016), de *V. cholerae* (Scrudato and Blokesch, 2012) ou encore de *Riemerella anatipestifer* (Huang et al., 2019), la capacité à réaliser la transformation naturelle n'est que partiellement (bien que sévèrement) altérée chez des mutants de *B. subtilis* (Berka et al., 2002; Tadesse and Graumann, 2007), *H. pylori* (Ando et al., 1999), *Deinococcus radiodurans* (Ithurbide et al., 2020), *Campylobacter jejuni* (Takata et al., 2005) ou encore *L. pneumophila* (source personnelle). Outre son rôle dans la transformation, il a été montré que DprA chez *N. meningitidis* affectait la variation antigénique des pilines (Duffin and Barber, 2016). Chez *S. pneumoniae*, il a aussi été montré que DprA était impliqué dans l'extinction de la compétence via son interaction avec le régulateur central de la compétence ComE (Mirouze et al., 2013).

À part ces deux exemples, l'absence de DprA n'entraîne aucun autre phénotype connu : les mutants sont efficaces pour la recombinaison générale et présentent une sensibilité normale aux UV (Bergé et al., 2003; Claverys et al., 2007; Duffin and Barber, 2016; Mortier-Barrière et al., 2007). Excepté les espèces pour lesquelles la transformation est constitutive, DprA est la plupart du temps induite durant la compétence (Claverys et al., 2007). Paradoxalement, DprA n'est pas présente uniquement chez ces espèces transformables. D'après l'outil en ligne AnnoTree (<http://annotree.uwaterloo.ca/>) qui permet de déterminer la distribution phylogénétique d'une annotation au sein de différents génomes bactériens ou archéens, sur plus de 30000 génomes bactériens que compte cette base de données, plus de 27000 possèdent le principal domaine pfam04281 de DprA (données datées de Septembre 2021). Ce qui fait de DprA une protéine quasi-ubiquitaire. Ces données suggèrent que DprA est un acteur majeur de la transformation naturelle et que ce mécanisme de THG est probablement bien plus répandu que ce qu'on pense actuellement. Récemment, il a été montré que la protéine DprA de *Fusobacterium naviforme*, exprimée chez un mutant *dprA* de *R. anatipestifer* restaurait partiellement sa capacité à transformer. *F. naviforme* n'ayant pour l'heure jamais été rapportée pour être naturellement transformable et sa protéine DprA ne possédant que 30% d'identité de séquence avec DprA de *R. anatipestifer*, cette complémentation partielle suggère que la fonction de DprA est conservée au sein de l'ensemble des espèces la possédant.

Selon les espèces, des analyses bio-informatiques et structurales ont permis de montrer que les protéines appartenant à la famille DprA (ou SMF) étaient définies sur la base d'alignements multiples par le domaine Pfam02481 (aussi appelé DNA_processg_A <http://pfam.xfam.org/family/PF02481>) et constituées de jusqu'à trois domaines (Hovland et al., 2017; Huang et al., 2019; Quevillon-Cheruel et al., 2012; Wang et al., 2014) (**Figure 24**). Chez *N. meningitidis*, *R. anatipestifer* ou encore *H. influenzae*

DprA est constituée d'un domaine N-terminal en alpha motif SAM (sterile-alpha motif), d'un domaine central Rossman Fold (RF) qui coïncide presque entièrement avec le domaine pfam04281 et d'un domaine C-terminal DML-1-like aussi appelé $Z\alpha$. En revanche, *S. pneumoniae* ne possède que les domaines SAM et RF tandis que *H. pylori* ne possède que les domaines RF et $Z\alpha$. Présent chez toutes les espèces avec un pourcentage d'identité de l'ordre de 30 à 37%, le domaine RF (pfam04281) apparait comme le domaine cœur de la protéine et constitue sa signature phylogénétique.



Figure 24 : Structure de DprA.

(A) Structure linéaire des protéines DprA de *N. meningitidis* (Nm), *R. anatipestifer* (Rp), *H. influenzae* (Hi), *S. pneumoniae* (Sp) et *H. pylori* (Hp). D'après (Huang et al., 2019). (B) Structure cristallisée de multimères de DprA chez *S. pneumoniae*. D'après (Quevillon-Cheruel et al., 2012) (A) DprA est constitué de plusieurs domaines selon les espèces (SAM, RF ou $Z\alpha$). Chez *S. pneumoniae*, les domaines RF (domaine RF étendu (eRF) chez *S. pneumoniae*) de DprA interagissent entre eux pour former un dimère et les domaines SAM (S) de ces dimères interagissent entre eux pour former des multimères (via une interaction SAM/SAM ou N/N).

Initialement, le rôle de DprA dans la transformation naturelle a été étudié chez *S. pneumoniae*. Des expériences de suivi d'ADNt radiomarqué dans des cellules compétentes de *S. pneumoniae* ont permis de montrer que DprA, tout comme SsbB, permettait de protéger l'ADNsb tout de suite après son entrée dans le cytoplasme (Bergé et al., 2003).

Ce n'est que quelques années plus tard que son rôle dans la transformation a véritablement été révélé dans l'étude de Mortier-Barrière et al. parue en 2007. Par des expériences de liaison de différents substrats d'ADN *in vitro* par des protéines DprA purifiées (de *B. subtilis* ou *S. pneumoniae*), ils ont montré que DprA liait l'ADNsb de façon non spécifique et qu'il s'assemblait pour former avec l'ADNsb de larges complexes nucléoprotéiques (Mortier-Barrière et al., 2007). Par microscopie électronique à transmission, il a même été montré que DprA était capable de juxtaposer plusieurs molécules d'ADNsb et de favoriser l'appariement de molécules complémentaires (Mortier-Barrière et al., 2007). Par des méthodes de co-immunoprécipitation et de double hybride en levure, les auteurs ont montré que DprA était capable d'interagir avec RecA et de former avec elle des nucléofilaments. Aussi, par des méthodes de microscopie électronique à transmission, ils ont montré que DprA était capable d'interagir avec des ADNsb recouverts de SSB (SsbB prioritairement et peut-être aussi SsbA)

probablement via la partie C-terminal de cette dernière (Attaiech et al., 2011; Hovland et al., 2017). Finalement, en montrant qu'*in vitro*, la simple présence de DprA permettait de (1) supprimer l'effet inhibiteur de SSB sur l'activité ATPase dépendante de l'ADNsb de RecA et (2) favoriser la nucléation de RecA sur l'ADNsb, ils ont démontré que le principal rôle de DprA, était de lever la barrière SSB pour RecA. Par toutes ces observations, Mortier-Barrière et al. ont montré que DprA possédait toutes les caractéristiques d'une protéine médiatrice de la recombinaison : RMP (*Recombination Mediator Protein*) dédiée à la recombinaison de l'ADNt.

Les auteurs ont aussi montré que DprA, outre sa capacité à lier de l'ADNsb, était aussi capable de se lier à de l'ADN superenroulé négativement. Comme l'ADNt entrant dans le cytoplasme et l'ADN chromosomique présentent respectivement ces caractéristiques et au vu de la capacité de DprA à favoriser l'appariement de molécules complémentaires, les auteurs ont suggéré que DprA avait aussi un rôle dans la recherche d'homologie médiée par RecA. Toutes ces données, leur ont permis de proposer un modèle d'action de DprA au cours du processus de transformation naturelle de *S. pneumoniae* qui reste compatible avec l'hypothèse de SsbB comme réservoir de l'ADNsb (**Figure 25**).

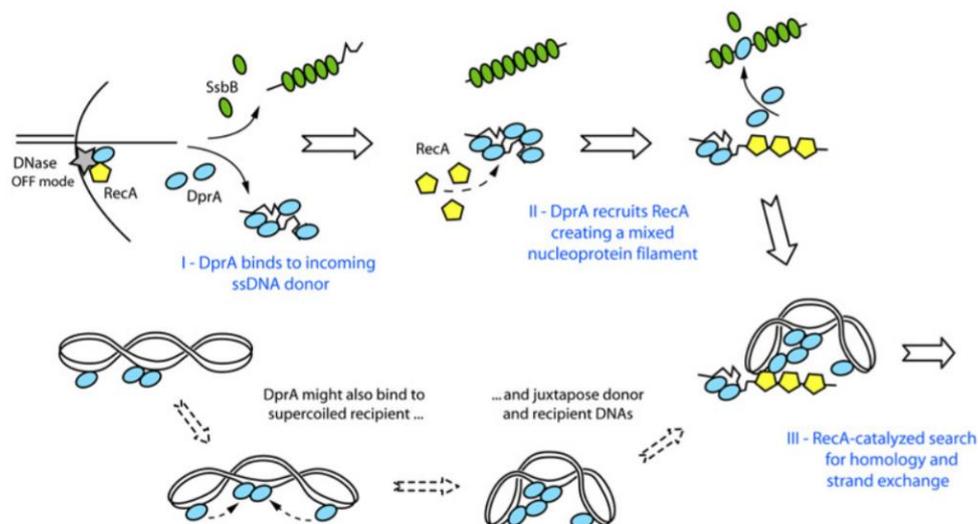


Figure 25 : Rôle de DprA dans les étapes cytoplasmiques de la transformation naturelle chez *S. pneumoniae*.

Extrait de (Mortier-Barrière et al., 2007). (I) Une fois entré, l'ADNt est soit lié par DprA pour directement être intégré dans la voie de la recombinaison homologue, soit lié par SsbB pour former un réservoir d'ADNsb. En parallèle DprA est capable de se lier à l'ADN chromosomique grâce à sa capacité à lier l'ADN superenroulé. (II) Par son rôle de RMP, DprA lié à un ADNsb, recrute RecA et/ou déloge SsbB des ADNsb en réserve, levant ainsi la barrière SSB et permettant la nucléation de RecA sur l'ADNsb. (III) En permettant la juxtaposition de l'ADNsb et de l'ADN chromosomique superenroulé et par sa capacité à favoriser l'appariement de molécules complémentaires DprA favorise la recombinaison homologue médiée par RecA.

Plus tard, l'analyse de filaments cristallisés de DprA a montré qu'elle était retrouvée associée deux à deux en dimère via son domaine C-terminal RF (interaction RF/RF ou C/C) et que chaque dimère était relié l'un avec l'autre via les domaines N-terminaux SAM (interaction SAM/SAM ou N/N) pour former des multimères (Quevillon-Cheruel et al., 2012) (**Figure 24, Figure 26**). Il a été montré que ce domaine SAM de DprA de *S. pneumoniae* interagissait aussi avec le régulateur central de la compétence ComE chez cette espèce pour éteindre la compétence (Mirouze et al., 2013).

Chez *H. pylori* qui ne possède pas de domaine SAM, il a été montré que l'interaction entre 2 dimères de DprA impliquait le domaine C-terminal DML-1-like (interaction DML-1/DML-1 ou C/C) pour former des multimères sur l'ADN (ie, des nucléofilaments) et que ces interactions DML-1/DML-1 pouvaient aussi permettre la formation de complexe de nucléofilaments réticulés (Dwivedi et al., 2015). De façon intéressante, chez DprA de cette espèce, il a été montré que le domaine RF en plus d'être impliqué dans la dimérisation de DprA (via une interaction RF/RF), possédait aussi une poche constituée de résidus chargés positivement responsables à la fois de la liaison à l'ADN^{sb}, mais aussi de façon surprenante à l'ADN^{db} (Dwivedi et al., 2015).

L'étude de mutants catalytiques de DprA de *S. pneumoniae* incapable d'interagir avec elle-même ou avec RecA, a montré que la dimérisation de DprA ainsi que son interaction avec RecA étaient déterminantes pour la formation de nucléofilaments et l'obtention de transformants (Quevillon-Cheruel et al., 2012). Dans cette même étude, des comparaisons des structures de dimères cristallisés de DprA de *S. pneumoniae* et *Rhodopseudomonas palustris*, une espèce gram négative phylogénétiquement distante, a permis de montrer qu'elle était aussi associée en dimère chez cette espèce toujours via le domaine RF et ce en dépit du fait que DprA de *R. palustris* possède un domaine C-terminal additionnel Z α (Quevillon-Cheruel et al., 2012). De même, il a été montré que les résidus à l'interface de ces dimères étaient pour la plupart conservés entre les DprA de ces deux espèces et que 3 d'entre eux étaient identiques chez plus de 60 domaines pfam04281 analysés. Cette forte conservation ainsi que des évidences expérimentales de dimérisation de DprA chez d'autres espèces phylogénétiquement distantes *H. pylori* ou encore *N. meningitidis* (Dwivedi et al., 2015; Hovland et al., 2017; Wang et al., 2014) suggèrent que cette dimérisation est essentielle à la fonction de DprA chez l'ensemble des bactéries transformables.

En revanche, les résidus impliqués dans son interaction avec RecA ne semblent pas conservés entre les DprA des différentes espèces alors que, pourtant, cette interaction a été observée chez différentes espèces et semble donc essentielle à la fonction de DprA au cours de la transformation naturelle (Lisboa et al., 2014)(voir chapitre II, partie III). En étudiant des mutants de DprA mutés au niveau de résidus conservés impliqués dans sa dimérisation, Quevillon-Cheruel et al. (2012) se sont aperçus que

certaines d'entre eux étaient aussi incapables d'interagir avec RecA. Les auteurs ont donc suggéré que les surfaces d'interaction de DprA avec elle-même et d'interaction avec RecA se chevauchaient, certains résidus étant à la fois impliqués dans l'une ou l'autre des interactions. Les auteurs ont même établi un modèle dans lequel ces deux interactions s'excluent mutuellement pour permettre la nucléation de RecA chez *S. pneumoniae* (Figure 26).

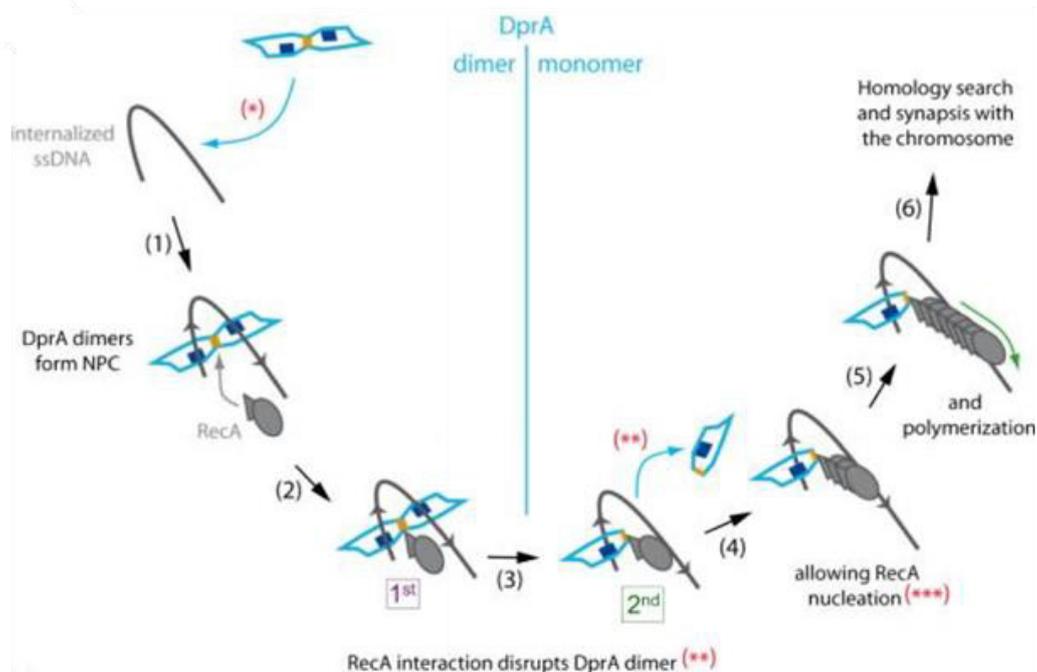


Figure 26 : Interaction entre RecA et DprA durant le processing de l'ADNsb chez *S. pneumoniae*. Extrait de (Quevillon-Cheruel et al., 2012). (*) Seul DprA dimérisé (en forme de papillon bleu clair sur la figure) est capable de former des nucléofilaments (1) DprA dimérisé lie l'ADNsb probablement via des poches constituées de résidus chargés positivement (en bleu)(Wang et al., 2014). (2) Par son rôle de RMP DprA recrute RecA au niveau du nucléofilament. (3) Les résidus impliqués dans la dimérisation de DprA étant aussi impliqués dans son interaction avec RecA (en jaune), la fixation de RecA implique un changement conformationnel du dimère de DprA, voire une rupture de ce dernier. (***) La nucléation est l'étape initiale d'assemblage de RecA sur le nucléofilament. (4, 5 et 6) Le maintien de l'interaction de DprA-ADNsb avec RecA entraîne de façon successive sa nucléation (4), sa polymérisation (5) et la recherche d'homologie (6).

Au moins chez *B. subtilis*, un autre médiateur de la recombinaison jouant un rôle similaire à DprA a été montré comme impliqué dans la recombinaison de l'ADNt : il s'agit de RecO (Yadav et al., 2012, 2013). Il a été montré que RecO collaborait avec DprA pour permettre la transformation chromosomique en favorisant l'accès de RecA sur un ADNsb recouvert de SsbA et/ou SsbB, un double mutant *recO dprA* étant très fortement impacté pour ce type de transformation (Yadav et al., 2012). Par ailleurs, il a été montré que RecO via son interaction spécifique avec SsbA était recruté par cette dernière pour permettre la transformation plasmidique en favorisant l'appariement d'ADNsb complémentaires, prérequis indispensable à la reconstitution d'un ADNdb plasmidique intact (Yadav et al., 2012). Chez *D. radiodurans*, il a été montré qu'en plus de DprA, c'est tout le complexe RecFOR

qui permettrait de jouer ce rôle de médiateur de la recombinaison en compensant le rôle de DprA pour permettre le chargement de RecA sur l'ADNsb, des doubles mutants *recF dpra* ou *recO dpra* étant totalement déficient pour la transformation chromosomique (Ithurbide et al., 2020). L'existence d'autres médiateurs de la recombinaison impliqués dans le processus de transformation naturelle pourrait expliquer pourquoi en l'absence de DprA certaines espèces sont toujours capables de réaliser la transformation naturelle.

II.C.3. Recombinaison de l'ADNt

II.C.3.1. Généralités

La recombinaison homologue est un mécanisme de recombinaison génétique durant lequel des séquences nucléiques sont échangées entre deux ADN identiques (recombinaison homologue) ou quasi identiques (recombinaison homéologue). Ce mécanisme de recombinaison homologue est très conservé dans les trois domaines du vivant : Bacteria, Eucarya et Archaea et se retrouve aussi chez les virus. La conservation structurelle et fonctionnelle de la principale protéine impliquée dans ce mécanisme, RecA chez les bactéries (RAD51 chez les eucaryotes ou RadA chez les archées), prouve l'importance et l'universalité des mécanismes d'échange d'ADN dans le vivant. À titre d'exemple, au niveau génétique, *recA* est le gène participant au métabolisme de l'ADN évoluant le plus lentement avec une séquence conservée de 60 à 70% chez l'ensemble des bactéries (Bell and Kowalczykowski, 2016).

Alors qu'un mutant *recA* chez les bactéries est souvent viable, une protéine RAD51 non fonctionnelle chez les humains entraîne une augmentation des mutations dans l'ADN. Cette augmentation des mutations s'explique par la perte de la fonction de réparation de l'ADN dépendante de la recombinaison homologue. Cette observation permet de mettre en évidence le rôle principal de la recombinaison homologue qui est de réparer les cassures double brin (Wright et al., 2018). Ces cassures double-brin sont les lésions les plus sévères qu'un ADN puisse subir puisqu'ils peuvent aboutir à la perte de fragments chromosomiques. Elles peuvent être provoquées par des stress endogènes tels que des espèces réactives de l'oxygène, exogènes comme des radiations ionisantes ou simplement lors du métabolisme cellulaire normal. Si ces cassures ne sont pas réparées de façon appropriée, cela peut provoquer des instabilités chromosomiques aboutissant à la mort de la cellule ou à l'apparition de maladies neuro-dégénératives et de certains cancers chez l'homme.

Outre son implication dans la réparation de l'ADN, la recombinaison homologue permet la production de nouvelles combinaisons de séquences durant la méiose chez les eucaryotes aboutissant à l'accroissement de la diversité génétique de la descendance ; de fait, c'est un mécanisme majeur

impliqué dans l'adaptation des espèces au cours de l'évolution. Chez les bactéries, les THG servent cette fonction de diversification génétique et impliquent souvent la recombinaison homologue. Le mécanisme de transformation naturelle qui est décrit ici est entièrement dépendant de la recombinaison homologue et donc de RecA.

II.C.3.2. Rôle de RecA dans la recombinaison homologue

RecA est définie comme étant une « *ATP-mediated DNA strand exchange protein* » qui est une famille de protéines capables après hydrolyse de l'ATP de (1) s'assembler coopérativement sur un ADNsb pour former un filament nucléoprotéique (phase présynaptique), (2) rechercher des séquences complémentaires entre ce nucléofilament et un ADNdb et (3) catalyser l'appariement de l'ADNsb du nucléofilament avec un des brins de l'ADNdb cible et l'échange de brins (phase synaptique) (Bell and Kowalczykowski, 2016; Reitz et al., 2021). Au cours de ce processus, il y a formation d'un intermédiaire branché et de ce que l'on appelle une « D-loop » (ie, *Displacement loop*) qui sont résolus pour former un ADN recombinant (phase postsynaptique).

RecA a été identifiée pour la première fois en 1965 chez *E. coli* dans un crible de mutants déficients pour leur capacité à générer des recombinants lors du processus de conjugaison bactérienne (Clark and Margulies, 1965). Depuis, de nombreuses études ont été menées, mais RecA de *E. coli* reste la plus étudiée (Kowalczykowski et al., 1994). Bien qu'un mutant RecA reste viable, l'absence de RecA affecte de nombreuses voies moléculaires dans la cellule bactérienne, notamment celles impliquées dans la réparation de l'ADN, la régulation du système SOS, la division cellulaire ou encore la ségrégation chromosomique (Bell and Kowalczykowski, 2016; Kowalczykowski et al., 1994). Par ailleurs, chez *V. cholerae*, *N. gonorrhoeae*, *D. radiodurans*, *B. subtilis*, *H. pylori* ou encore *S. pneumoniae* au moins et probablement chez l'ensemble des espèces transformables, l'absence de RecA abolit la transformation chromosomique (Ithurbide et al., 2020; Koomey and Falkow, 1987; Martin et al., 1995; Schmitt et al., 1995; Seitz and Blokesch, 2013b; Serrano et al., 2018) ce qui suggère que RecA est essentielle à la transformation chromosomique et qu'aucun autre acteur ne permet de suppléer sa fonction. De façon intéressante, l'absence de RecA chez *D. radiodurans* ou *B. subtilis*, n'est pas délétère pour la transformation de plasmide ce qui suggère que l'établissement de plasmide par transformation implique d'autres acteurs chez ces espèces (Ithurbide et al., 2020; Serrano et al., 2018).

La première structure cristallisée de RecA a été résolue en 1992 (Story et al., 1992), mais ce n'est qu'en 2008 qu'elle a été pour la première fois obtenue sous la forme d'un nucléofilament de RecA cristallisé en complexe avec l'ADNsb (complexe présynaptique) ou double brin (complexe synaptique) permettant de révéler les mécanismes moléculaires mis en jeu durant la recombinaison homologue (Chen et al., 2008). Sur la base de ces données structurales chez *E. coli*, RecA est une protéine de 38

kDa et de 352 acides aminés (**Figure 27**). Elle possède un large domaine central ATPase (Core ATPase Domain, CAD) bordé par les domaines N-terminal et C-terminal (N-terminal Domain, NTD et C-terminal domaine, CTD respectivement). Le CAD est constitué de plusieurs sites : le site de liaison à l'ATP (site I) qui est aussi le site de liaison à l'ADNsb, le site de liaison à l'ADNdb (site II) et des résidus responsables de l'hydrolyse de l'ATP (motif [KR]x[KR] et E96). Le CTD peut-être lui-même divisé en deux sous-domaines : une passerelle permettant le bon positionnement de l'ADNdb et son extrémité capable de moduler l'activité enzymatique. Au sein de tous ces différents domaines, il existe des régions flexibles : à la jonction entre NTD et CAD, au niveau du domaine cœur (loop L1 et loop L2) et au niveau du CTD. Chacun de ces sites/domaine a une fonction particulière au cours du processus de recombinaison homologue médié par RecA qui s'effectue en plusieurs étapes.

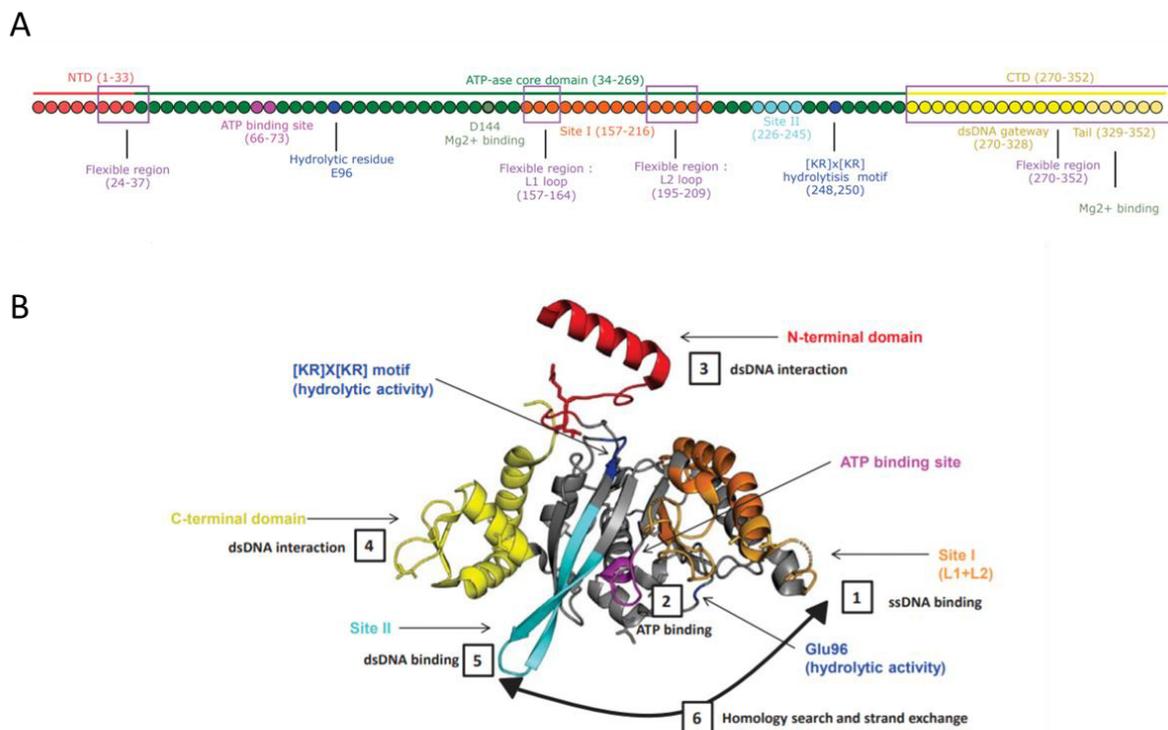


Figure 27 : Structure linéaire (A) et 3D de RecA et implication des différents domaines dans la recombinaison homologue (B).

Extrait de (Del Val et al., 2019), voir texte pour plus de détails.

Nucléation de RecA et formation du filament présynaptique. RecA interagit avec le squelette phosphaté de l'ADNsb qui est chargé négativement via le site I de son CTD qui est constitué de résidus chargés positivement. Dans cette conformation, en l'absence de ses cofacteurs, RecA est inactive. L'étape de nucléation est lente et se termine lorsqu'un dimère de RecA est attaché sur l'ADNsb. Il s'en suit une étape d'extension du filament de RecA par l'ajout de monomère de RecA de façon contiguë. Cette extension se fait dans les deux directions bien que l'extension dans le sens 5'→3' soit plus rapide.

RecA s'assemble de façon hélicoïdale pour former un filament hélicoïde nucléoprotéique appelé aussi filament présynaptique.

Cette nucléation de RecA et l'extension ultérieure du filament de RecA se font probablement en compétition directe avec les SSBs. On a vu précédemment que les SSBs étaient probablement les premières protéines à venir se fixer sur l'ADNsb une fois entrée dans le cytoplasme pour le protéger. On a aussi vu que ces SSBs constituaient une barrière pour limiter cette nucléation lorsque celle-ci n'est pas nécessaire. Paradoxalement, ces SSBs sont aussi bien connues pour favoriser l'extension du filament de RecA en participant à l'élimination des structures secondaires souvent générées spontanément sur les ADNsb (Kowalczykowski and Krupp, 1987). Cet autre rôle est aussi envisagé pour SsbA et/ou SsbB au cours de la transformation naturelle d'après des données obtenues chez *B. subtilis* (Yadav et al., 2012, 2013). Chez *E. coli*, l'hypothèse a été formulée que ce double rôle antagoniste de SSB sur RecA permettait de réguler sa nucléation et/ou son extension (Wu et al., 2017). Supposant que RecA (et/ou ses médiateurs) déloge SSB au cours de ces processus, en présence de quantités limitées de RecA lors de dommages importants à l'ADN par exemple, SSB pourrait privilégier un mode de liaison plus « difficile » à déplacer pour RecA permettant ainsi de l'économiser pour que celle-ci se focalise sur les événements de recombinaison les plus importants.

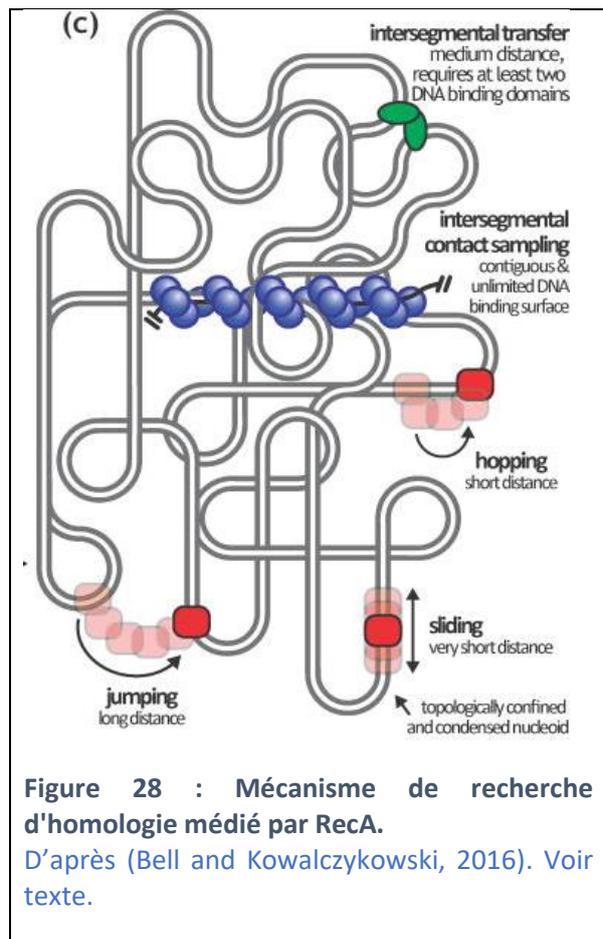
Activation du filament présynaptique. Initialement, le nucléofilament de RecA est dans une conformation inactive appelé « état O ». En présence de ses cofacteurs Mg^{2+} et de l'ATP qui se fixent sur leurs sites respectifs, il adopte une conformation active appelée « état A » (Del Val et al., 2019). Dans cette conformation, le site I de RecA possède une forte affinité pour l'ADNsb. À l'intérieur du filament de RecA, l'ADNsb est étiré de l'ordre de 1,5 fois sa taille d'origine. Le nucléofilament est alors caractérisé par un pas d'hélice plus étendu passant de ~ 6 nm à ~ 9 nm et est plus aplati passant de 12 nm à 9 nm (Egelman and Stasiak, 1986). Cette étape est appelée l'étape de « *stretching* ». D'un point de vue stœchiométrique, dans cette conformation active, RecA interagit avec 3 nucléotides (contre 5 dans sa forme inactive). Tout ceci permet d'obtenir un ADN « sousenroulé » permettant une conformation plus ouverte indispensable pour la recherche et la capture de séquence d'ADN homologue au cours de la phase synaptique (Bell and Kowalczykowski, 2016).

Recherche d'homologie (« homology probing ») (Figure 29). Le nucléofilament de RecA effectue une recherche dans l'ensemble du génome pour trouver une séquence d'ADNdb homologue à son ADNsb. Ce filament présynaptique se fixe transitoirement et non spécifiquement sur un des brins de l'ADNdb via le site II de RecA à la suite d'interactions successives entre les différents domaines de RecA notamment le NTD et le CTD et l'ADNdb (étape de « *docking* ») rendue possible par les différentes régions flexibles. Dans le nucléofilament, l'ADNdb est aussi maintenu dans une conformation

« *stretch* », ce qui perturbe l'association entre les bases de la double hélice causant un mésappariement et une ouverture locale de l'ADN. Les nucléotides, détachés du duplex d'ADN, tournent sur eux-mêmes pour faire face à l'ADNsb situé dans le site I avec qui ils établissent des liaisons hydrogène (Del Val et al., 2019). Tant qu'une région homologe n'est pas trouvée, la synapse n'est pas stable.

La recherche d'homologie se fait très rapidement et différents mécanismes ont été proposés pour expliquer comment elle est réalisée à l'échelle de temps métabolique (Bell and Kowalczykowski, 2016; Nandakumar and Patel, 2015; Renkawitz et al., 2014) (**Figure 28**). (1) Le nucléofilament de RecA diffuse

dans toutes les directions ce qui augmente la probabilité de rencontrer un ADN avec qui il va initier une recombinaison non homologue (« *collision* » *mechanism*). (2) Puisque l'ADNdb *in vivo* est très compacté, des régions qui linéairement sont éloignées de plusieurs dizaines de kilobases à plusieurs mégabases, se retrouvent tridimensionnellement proches l'une de l'autre permettant à un même nucléofilament de réaliser de multiples contacts simultanés à des régions éloignées du chromosome (« *Intersegmental contact sampling* » ou « *intergmental transfer* » *mechanism*). (3) La capacité du nucléofilament de RecA à se fixer transitoirement et non spécifiquement à l'ADNdb lui permet de glisser le long de l'ADNdb et le scanner très rapidement (« *sliding* » *mechanism*). (4 et 5) Cette liaison nucléofilament-ADNdb étant relativement faible,



ce nucléofilament de RecA peut se dissocier et se lier rapidement à une région linéairement proche (« *hopping* » *mechanism*) ou éloignée (« *jumping* » *mechanism*) de la zone initiale de liaison. (6) Enfin, en glissant sur l'ADNdb le nucléofilament de RecA scanne et rejette toutes les régions homologues inférieures à 8 pb jusqu'à former un complexe stable (« *sliding* » ou « *kinetic discrimination* » *mechanism*).

Formation du filament post-synaptique et échange de brin. Lorsqu'une région d'au moins 8 pb est trouvée, le complexe synaptique adopte une conformation métastable dont la durée de vie augmente exponentiellement toutes les 3 pb (Del Val et al., 2019; Lee et al., 2016). Lorsqu'une région de 15-20 nucléotides d'homologie est trouvée, l'appariement homologue de l'ADNsb situé dans le site I et le brin complémentaire de l'ADNdb situé dans le site II de RecA peut avoir lieu dans un processus qu'on appelle « échange de brin » ou « invasion de brin ». Il se forme ce qu'on appelle une D-loop (displacement loop) qui est un intermédiaire à trois brins (**Figure 29**). Cet intermédiaire de recombinaison est constitué, d'une part, du brin déplacé de l'ADNdb non complémentaire qui reste dans le site II et d'un hétéroduplex d'ADN néoformé, d'autre part, qui reste attaché au site I et qui est formé du brin donneur (brin échangé) appareillé à la chromatide sœur du brin déplacé (Renkawitz et al., 2014). Au cours de la recombinaison homologue cette D-loop est processée au niveau de la jonction des 3 brins (« 3 way junction » ou « 3J ») pour permettre son extension au cours d'un processus qu'on appelle la migration de branche (voir II.C.3.3). L'établissement de ce complexe est favorisé, car le site II de RecA a une plus forte affinité pour l'ADNsb déplacé (Del Val et al., 2019).

Bien que l'échange de courtes régions d'ADN au cours de ce processus ait lieu indépendamment de l'hydrolyse de l'ATP (Menetski et al., 1990; Rosselli and Stasiak, 1990), l'acquisition de larges régions d'ADN est, elle, couplée à l'hydrolyse de l'ATP. Cette réaction permet de dissocier les protomères de RecA situés à l'extrémité opposée à celle où l'assemblage se produit. Les RecA ainsi dissociés peuvent être alors recyclés à l'avant du nucléofilament pour être de nouveau assemblés et permettre une nouvelle recherche d'homologie. En ce sens, le nucléofilament agit comme un moteur rotatif qui propage l'échange de brin sur de grandes distance permettant l'échange de larges régions d'ADN (Cox, 2003, 2007).

Dissociation : Lorsque l'échange de brins est achevé, l'hydrolyse de l'ATP réalisée par coopération des motifs [KR]X[KR] et E96 converti la forme RecA-ATP de haute affinité pour l'ADN en la forme RecA-ADP qui possède une faible affinité pour l'ADN, ce qui entraîne la dissociation de RecA. Puisque ces motifs et le site de fixation à l'ATP sont situés à l'interface des différents monomères de RecA dans le nucléofilament, le changement de conformation induit par l'hydrolyse de l'ATP se transmet aux monomères voisins qui se dissocient aussi (Del Val et al., 2019). Cela entraîne le relargage de l'hétéroduplex nouvellement formé et dans le cadre de la transformation naturelle, probablement la dégradation du brin déplacé. Il est suggéré que ce relargage est favorisé par la liaison de SSB à l'ADNsb en facilitant son émancipation du nucléofilament (Lavery and Kowalczykowski, 1992).

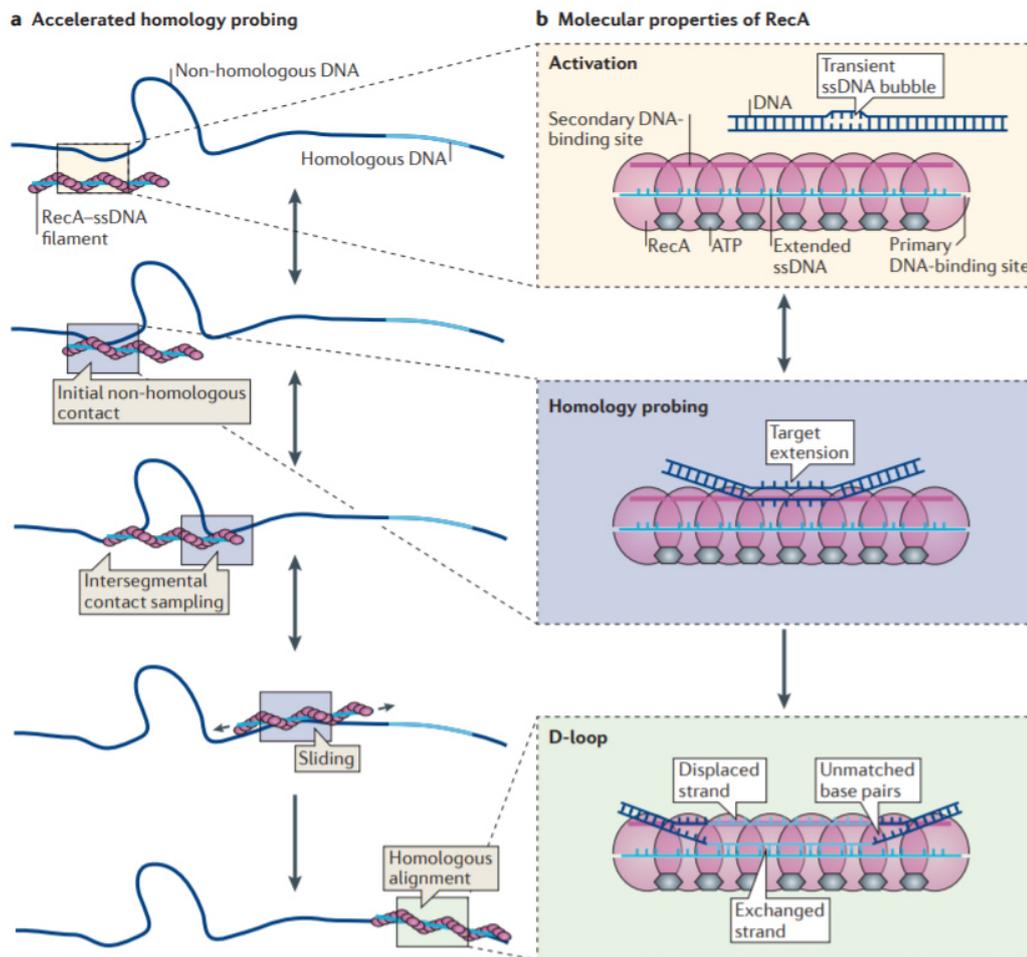


Figure 29 : Mécanisme moléculaire de la recombinaison homologue médiée par RecA.

a. Stratégie de recherche rapide d'homologie par RecA. b. Etape principales du processus de recombinaison médiée par RecA. D'après (Renkawitz et al., 2014)

Activation : La fixation de l'ATP sur RecA augmente son affinité pour l'ADNsb. RecA fixe alors l'ADNsb au niveau de son site I et s'assemble de façon hélicoïdale sur l'ADNsb pour former un nucléofilament. Au sein du filament de RecA l'ADNsb est étendu et maintenu dans une conformation ouverte.

Homology probing : Le nucléofilament de RecA effectue une recherche dans l'ensemble du génome pour trouver une séquence d'ADNdb homologue. Pour cela, il se fixe transitoirement et non spécifiquement sur un des brins de l'ADNsb via le site II de RecA. Différentes stratégies sont envisagées pour diminuer le temps de recherche. RecA effectue la recherche d'homologie selon les mécanismes décrits dans la partie « recherche d'homologie » du texte ci-dessus et illustré en (a.).

D-loop : dans ce nucléofilament, l'ADNdb est aussi étiré, ce qui perturbe l'association entre les bases de la double hélice causant un mésappariement et une ouverture locale de l'ADN. Les nucléotides, détachés du duplex d'ADN, tournent sur eux-mêmes pour faire face à l'ADNsb situé dans le site I avec qui ils établissent des liaisons hydrogènes. Lorsqu'une région de 15-20 nucléotides d'homologie est trouvée, l'appariement homologue de l'ADNsb situé dans le site I et le brin complémentaire de l'ADNdb situé dans le site II de RecA peut avoir lieu dans un processus qu'on appelle « échange de brin » ou « invasion de brin ». Il se forme ce qu'on appelle une D-loop qui est un intermédiaire à trois brins. Cet intermédiaire de recombinaison est constitué, d'une part, du brin déplacé de l'ADNdb non complémentaire qui reste dans le site II et d'un hétéroduplex d'ADN néoformé, d'autre part, qui reste attaché au site I et qui est formé du brin donneur (brin échangé) apparié à la chromatide sœur du brin déplacé.

II.C.3.3. Processing du D-loop de recombinaison de l'ADNt

Aujourd'hui, nous n'avons qu'un petit aperçu des mécanismes moléculaires mis en jeu pour permettre l'intégration de l'ADNt par recombinaison homologe. Néanmoins, récemment, des équipes indépendantes ont décrit le rôle de deux hélicases conservées remplissant des rôles similaires au cours de la recombinaison homologe de l'ADNt : RadA pour les Firmicutes (Marie et al., 2017; Torres et al., 2019) et ComM pour les bactéries gram négative (Nero et al., 2018).

RadA (aussi appelé Sms) est une protéine très conservée présente à la fois chez les gram positives et négatives (présente dans près de 25000 des 30000 génomes analysés par AnnoTree). RadA a été initialement caractérisé chez *E. coli* (Beam et al., 2002) et montrée comme étant impliquée dans la réparation de l'ADN en agissant de concert avec RecA au niveau de la jonction 3J de la D-loop de recombinaison pour favoriser la migration de branche de manière ATP dépendante (Cooper and Lovett, 2016). Par la suite, elle a aussi été montrée comme étant impliquée dans la transformation naturelle des bactéries gram positives *S. pneumoniae* et *B. subtilis*, par des cribles identifiant des mutants déficients pour la transformation (Burghout et al., 2007; Carrasco et al., 2002). Chez ces deux espèces, l'efficacité de transformation chromosomique d'un mutant *radA* est modérément réduite de l'ordre de 100 fois par rapport à la *WT* et de l'ordre de 10 fois pour la transformation de plasmide selon les données obtenues chez *B. subtilis* (Torres et al., 2019).

Contrairement à RadA, ComM n'est retrouvé que chez les gram négative et n'a pour l'heure pas été montré comme impliquée dans d'autres mécanismes que la transformation naturelle suggérant que son rôle est dédié à ce processus. Néanmoins, elle est aussi retrouvée chez les bactéries gram négatives non transformables, ce qui suggère qu'elle joue d'autres rôles chez ces espèces ou que ces organismes sont aussi naturellement transformables, mais que les conditions d'induction de la compétence n'ont pas encore été découvertes. De façon intéressante, certaines espèces d'*A. baumannii* possèdent un îlot génomique de résistance de type AbaR intégré dans ComM (Bonnin et al., 2012). Ces souches présentent des fréquences de transformation relativement faibles et il a récemment été montré que la réparation du gène *comM* par l'excision de cet îlot, rétablissait de hautes fréquences de transformation (Godeux et al., 2020). Cette inactivation d'un gène de compétence par un élément génétique mobile n'est pas inédite (Durieux et al., 2019).

L'implication de ComM dans la transformation naturelle a été initialement montrée chez *H. influenzae* (Gwinn et al., 1998), mais n'a été caractérisée que récemment chez *V. cholerae* et *A. baylyi* (Nero et al., 2018). ComM fait partie du régulon de compétence chez ces espèces et tout comme RadA, son absence réduit modérément l'efficacité de transformation chez un mutant *comM* de l'ordre de 100 fois. Contrairement à RadA chez *B. subtilis*, ComM n'est pas impliquée dans la transformation de

plasmide au moins chez *V. cholerae* (Nero et al., 2018). Par ailleurs, bien que RadA soit présente chez les gram négatives, elle n'est pas requise pour la transformation de ces espèces (Nero et al., 2018). ComM est prédite pour être une AAA+ ATPase possédant des similarités de structure avec l'hélicase répllicative MCM2–7 des eucaryotes.

Sur la base de ces données prédites pour ComM et des données existantes chez *E. coli* pour RadA, les auteurs des études Nero et al. (2018) et Marie et al. (2017) ont réalisé des expériences similaires *in vivo* pour montrer l'implication de ces hélicases putatives dans la migration de branche associée à la recombinaison de l'ADNt chez *V. cholerae* (ComM) et *S. pneumoniae* (RadA) respectivement. Ici, je vais décrire l'expérience réalisée pour ComM chez *V. cholerae* car une méthode adaptée a été utilisée durant mes travaux de thèse pour caractériser le rôle d'un nouvel acteur impliqué dans la transformation naturelle (Chapitre II, III.A). Des souches de *V. cholerae* (ie, le mutant *comM* et sa souche parente WT) porteuses du gène *lacZ* fonctionnel ont été transformées avec un ADNt possédant un premier marqueur de sélection conférant la résistance à la spectinomycine (Spec^R), qui est génétiquement lié à un second marqueur situé en aval qui introduit après transformation une mutation qui inactive le gène *lacZ* et qui rend la bactérie blanche après recombinaison/réplication (Figure 30).

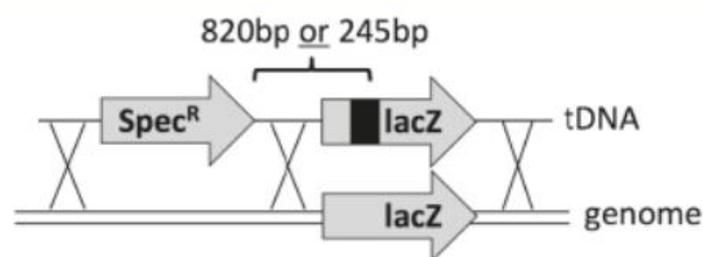


Figure 30 : Schématisation de l'expérience de co-migration de branche.
D'après (Nero et al., 2018). Voir texte.

Deux ADNt ont été utilisés présentant respectivement la mutation de *lacZ* placée à 245 pb ou à 820 pb de la fin de la cassette Spec^R. Après détermination des efficacités de co-migration de marqueur (% du nombre de transformant Spec^R ayant aussi acquis la mutation de *lacZ*), il en résulte une efficacité de co-migration d'autant plus réduite que la mutation de *lacZ* est éloignée de la cassette de résistance pour un mutant *comM* par rapport à sa souche parente. Ce qui se caractérise phénotypiquement par une diminution du nombre de colonies blanches sur milieu supplémenté en spectinomycine et en X-gal. Cette mesure de l'efficacité de co-migration est en fait une mesure indirecte de l'efficacité de migration de branche (ie, d'extension de la D-loop). En effet, cette diminution de l'efficacité de co-migration à mesure que la délétion de *lacZ* est éloignée du marqueur de sélection Spec^R s'explique par le fait qu'en l'absence de ComM, la D-loop qui a permis d'acquérir la cassette Spec^R n'a pas réussi à

s'étendre efficacement jusqu'à la mutation inactivant *lacZ* et d'autant plus lorsque celle-ci est éloignée. Ces résultats suggèrent que ComM joue un rôle dans la migration de branche en permettant d'augmenter la taille de l'événement de recombinaison (Nero et al., 2018). Une expérience similaire réalisée chez *S. pneumoniae in vivo* a conclu que RadA, tout comme ComM, était impliquée dans la migration de branche lors de la recombinaison durant la transformation naturelle (Marie et al., 2017).

Des structures cristallisées ont été obtenues pour ComM chez *V. cholerae* et RadA chez *S. pneumoniae* à des résolutions respectives de 13,8 Å et 2,5 Å (Marie et al., 2017; Nero et al., 2018). Les reconstructions 3D révèlent que ces protéines adoptent une structure en anneau hexamérique similaire à l'hélicase répliquative bactérienne DnaB. Tandis qu'en solution la liaison de l'ATP est requise pour l'hexamérisation de ComM (bien qu'induite possiblement pas d'autres types de nucléotide), elle n'est pas requise pour l'oligomérisation de RadA. De façon importante, ComM possède deux domaines chélatase Mg^{2+} qui catalysent habituellement l'insertion de Mg^{2+} dans la protoporphyrin IX des organismes photosynthétiques. ComM est la seule protéine connue à ce jour avec un tel domaine chez un organisme non photosynthétique. Chez *S. pneumoniae*, RadA possède 2 autres domaines conservés en plus de son domaine ATPase central (domaine H) : un domaine en doigt de zinc de type C4 (domaine C) en N-terminal et un domaine protéase (domaine P) en C-terminal.

In vitro, ces deux protéines présentent une activité hélicase capable d'ouvrir des substrats d'ADN synthétiques en forme de fourche. Il a été montré que RadA était incapable d'ouvrir un ADNdb s'il ne possédait pas une extrémité simple brin débordante, suggérant que l'activité hélicase de RadA est initié au niveau d'un ADNsb. Avec d'autres types de substrats, il a été montré que RadA seul ouvrait la double hélice en se déplaçant sur l'ADNsb dans le sens 5'->3' de la même façon que DnaB et la plupart des hélicases connues (Marie et al., 2017). ComM, au contraire, est un peu plus singulière, car il a été montré qu'elle était capable d'ouvrir seule l'ADNdb dans les deux directions, ce qui suggère que ComM est une hélicase bidirectionnelle (Nero et al., 2018). De façon importante, ces activités hélicases sont dépendantes de l'hydrolyse de l'ATP, car la mutation de résidus conservés de leur domaine ATPase respectif abolit toute activité hélicase *in vitro*.

Puisque toutes les hélicases ne sont pas capables de migration de branche, des tests *in vitro* de migration de branche ont été réalisés sur des substrats ADN synthétiques imitant les jonctions 3J délimitant la D-loop de recombinaison formée par RecA. Il a été montré que les deux hélicases sont capables d'ouvrir ce type de substrat de façon ATP-dépendante. De façon intéressante, des expériences de double hybride bactérien ont montré que RadA de *S. pneumoniae* interagissait avec RecA et que son domaine C était essentiel à cette interaction (Marie et al., 2017). Lorsque les expériences d'ouverture de molécules 3J sont réalisées en présence de RecA purifiée, il a été montré

que RadA était capable d'ouvrir ce type de substrat dans le sens 3'->5' (Marie et al., 2017). Cela suggère que RadA seule est capable d'ouvrir de l'ADN dans le sens 5'->3', mais qu'elle nécessite l'interaction avec RecA pour permettre l'ouverture dans le sens 3'->5', ce qui fait que, tout comme ComM, RadA serait capable d'ouvrir la D-loop dans les deux directions et donc de réaliser la migration de branche *in vivo*. Chez *B. subtilis*, où l'interaction RecA/RadA au cours de la recombinaison de l'ADNt a aussi été démontrée récemment, il est suggéré que RecA permettrait le recrutement de RadA au niveau des jonction 3J de la D-loop (Torres et al., 2019).

Bien que structurellement très différentes, et au vu de leur grande conservation et de leur mode d'action similaire, un modèle unifié de processing de la D-loop peut être élaboré et potentiellement être applicable à l'ensemble des bactéries transformables (**Figure 31**).

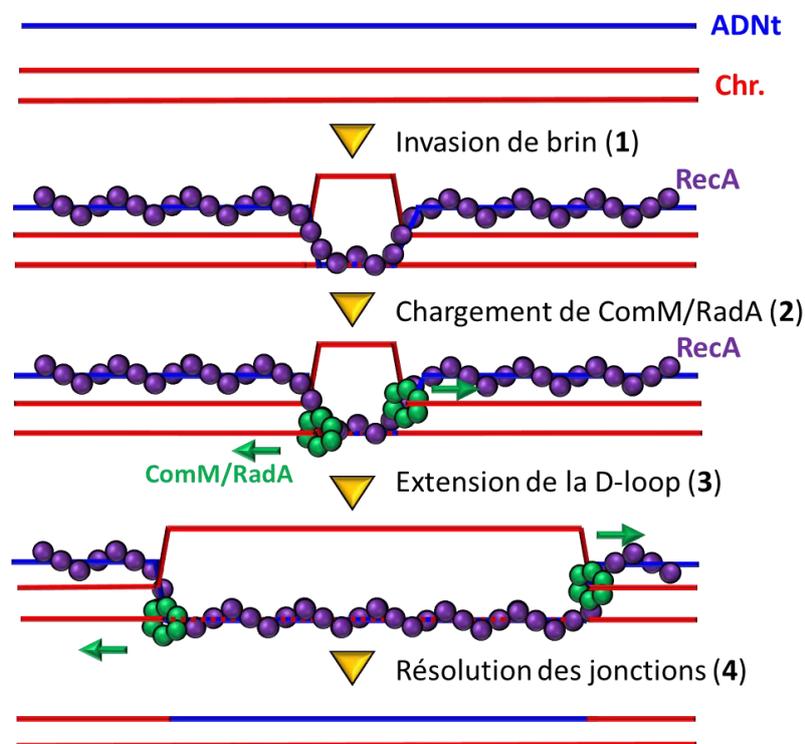


Figure 31 : Mécanisme unifié du processing de D-loop au cours de la recombinaison de l'ADNt.

Réalisé avec BioRender. Dans le cytoplasme l'ADNt simple brin est chargé de DprA et/ou SSB(s) (non représentées ici). (1) Une fois l'homologie trouvée RecA catalyse l'invasion de l'ADNsb dans l'ADNdb, ce qui forme une D-loop flanquée de jonctions 3J. (2) ComM/RadA se chargent au niveau de la jonction 3J et ouvre la double hélice via leur activité hélicase. (3) ComM/RadA catalyse l'extension de la D-loop aussi appelé migration de branche pour augmenter la portion d'ADN intégrée. (4) Les jonctions sont résolues par un mécanisme encore inconnu formant un hétéroduplex d'ADN.

III. *Legionella pneumophila* : un modèle original de bactérie naturellement transformable

A. Une bactérie d'origine environnementale et pathogène accidentel de l'homme

III.A.1. Découverte

En juillet 1976, une épidémie de maladie respiratoire atypique causée par un agent pathogène jusqu'alors inconnu toucha les participants du congrès de la Légion Américaine qui se tenait à Philadelphie. Sur 182 cas recensés, 29 succombèrent d'une pneumonie sévère aujourd'hui appelée maladie du légionnaire ou legionellose (Legionnaires' disease) (Fraser et al., 1977). Après d'intenses investigations, la bactérie responsable de cette épidémie fut isolée en 1977 et baptisée *Legionella pneumophila* en référence aux circonstances de la survenue de cette épidémie (McDade et al., 1977).

III.A.2. Une espèce environnementale...

L. pneumophila est un bacille de type gram négatif qui fait partie d'un genre bien plus vaste comprenant plus de 65 espèces (Mondino et al., 2020). *L. pneumophila* est retrouvée de façon ubiquitaire dans les environnements d'eau douce naturels ou anthropisés tels que les rivières, les lacs, les fontaines, les tours aéroréfrigérantes ou les spas (Newton et al., 2010). Dans son environnement naturel, contrairement à d'autres espèces environnementales, *L. pneumophila* n'est pas retrouvée libre, mais associée à des hôtes protozoaires tels que les amibes. C'est d'ailleurs la première espèce bactérienne pour laquelle a été décrite une multiplication au sein d'hôtes protozoaires (Rowbotham, 1980). Aujourd'hui, c'est au moins 13 espèces d'amibes et 2 espèces de protozoaires ciliés qui ont été reportées comme étant des potentiels hôtes naturels de *Legionella* (Fields, 1996). Dans les environnements anthropisés, *L. pneumophila* est retrouvée la plupart du temps sous la forme de biofilms complexes associée à d'autres microorganismes eucaryotes ou procaryotes (Taylor et al., 2009). Qu'elle soit retrouvée dans ses hôtes protozoaires ou sous forme de biofilm, ces niches fournissent à *L. pneumophila* une source continue de nutriments nécessaire à sa croissance (Declerck, 2010) et lui permettent de se protéger des conditions environnementales difficiles et donc de persister dans l'environnement (Berjeaud et al., 2016).

III.A.3. ... Et un pathogène accidentel de l'Homme

III.A.3.1. Données épidémiologiques

Parmi les 65 espèces connues de *Legionella*, *L. pneumophila* est responsable de plus de 90% des cas confirmés de légionellose en Europe et aux Etats-Unis (Yu et al., 2002) et bien qu'il existe 16 sérogroupes connus au sein même de cette espèce, plus de 85% des cas de légionellose sont imputés au séro groupe 1 (sg1) alors même que celui-ci n'est pas majoritaire dans l'environnement (Doleans et al., 2004).

Les infections humaines à *L. pneumophila* sont contractées quasi exclusivement par l'inhalation d'aérosols contaminés par cette bactérie qui sont générés dans la majorité des cas par des infrastructures humaines vétustes (douches, bains à remous, systèmes d'air conditionnés) dans lesquelles l'eau est maintenue à des températures élevées (30-40°C) favorisant notamment le développement de biofilms (Rogers et al., 1994). Seules les sources naturelles d'eau chaude sont considérées comme une potentielle origine naturelle de contamination (Mashiba et al., 1993). Bien qu'au moins un cas de transmission interhumaine ait été recensé, ce mode transmission est considéré comme extrêmement rare (Correia et al., 2016). Les individus les plus susceptibles de développer des infections à *L. pneumophila* sont des hommes âgés de plus de 50 ans possédant des comorbidités (tabagisme, obésité, diabète, traitement immunosuppresseur)(Mondino et al., 2020). Le développement des infrastructures aquatiques qui a favorisé l'exposition humaines aux aérosols potentiellement contaminés, le vieillissement de la population et possiblement les changements climatiques (Walker, 2018) ont entraîné une recrudescence des cas de légionellose ces dernières années. À titre d'exemple en France, le nombre de cas de légionellose est passé d'environ 600 cas en 2000 à plus de 2000 en 2018, année record du nombre de cas (<https://www.santepubliquefrance.fr/>). Néanmoins, les moyens de surveillance et de détection ayant considérablement progressé ces 20 dernières années, on ne peut exclure une sous-évaluation du nombre de cas dans les années 2000 et auparavant.

Une fois inhalée *L. pneumophila* peut directement atteindre les alvéoles pulmonaires où elle est internalisée par les macrophages alvéolaires. Ces cellules phagocytaires de l'immunité innée constituent en temps normal la première ligne de défense contre les pathogènes et notamment les bactéries en permettant leur élimination par fusion du phagosome et du lysosome. Mais, contrairement à la plupart des bactéries, *L. pneumophila* utilise tout un arsenal pour empêcher la fusion de ces vésicules et permettre sa multiplication par un mécanisme proche de celui utilisé pour parasiter ses hôtes protozoaires naturels (Al-Quadani et al., 2012). Ces derniers sont d'ailleurs considérés comme un « terrain d'entraînement » dans lequel *L. pneumophila* aurait acquis au cours de

l'évolution des mécanismes lui permettant de survivre dans les macrophages (Molmeret et al., 2005). En effet, les macrophages et les hôtes protozoaires de *Legionella*, toutes deux des cellules eucaryotes, possèdent des voies moléculaires conservées qui sont ciblées par *L. pneumophila* pour permettre sa multiplication (Escoll et al., 2013). À ce titre, *L. pneumophila* est considéré comme un pathogène accidentel de l'homme ; la transmission interhumaine étant (quasi) inexistante, l'infection des macrophages constitue une impasse évolutive pour la bactérie.

III.A.3.2. Le système de sécrétion de type 4 : principal facteur de virulence de *L. pneumophila*

L'un des facteurs de virulence clé permettant à *L. pneumophila* de survivre et de se multiplier dans ses hôtes est le système de sécrétion de type IVB Dot/Icm (SST4B) (Kubori and Nagai, 2016; Segal et al., 1998; Vogel et al., 1998). Localisé aux pôles de la cellule bactérienne (Jeong et al., 2017), il induit la sécrétion de plus de 300 effecteurs qui permettent le détournement des voies moléculaires de la cellule hôte au profit de la bactérie (Mondino et al., 2020). En outre, ce SST4B est crucial pour permettre aux bactéries phagocytées d'échapper à la voie endocytique en empêchant la fusion phagosome – lysosome (Horwitz, 1983) par la formation d'un compartiment intracellulaire particulier appelé « Legionella-containing vacuole » (LCV) (Isberg et al., 2009) qui sert de niche répliquative pour la bactérie au sein de sa cellule hôte. À ce jour, parmi les ~330 effecteurs qu'injecte *L. pneumophila* dans le cytoplasme de la cellule hôte, seule une cinquantaine ont été étudiés et se sont vus attribuer une fonction (Finsel and Hilbi, 2015; Mondino et al., 2020). Ils interviennent et gouvernent l'ensemble des étapes du cycle intracellulaire de *L. pneumophila* comme présenté en **Figure 32**.

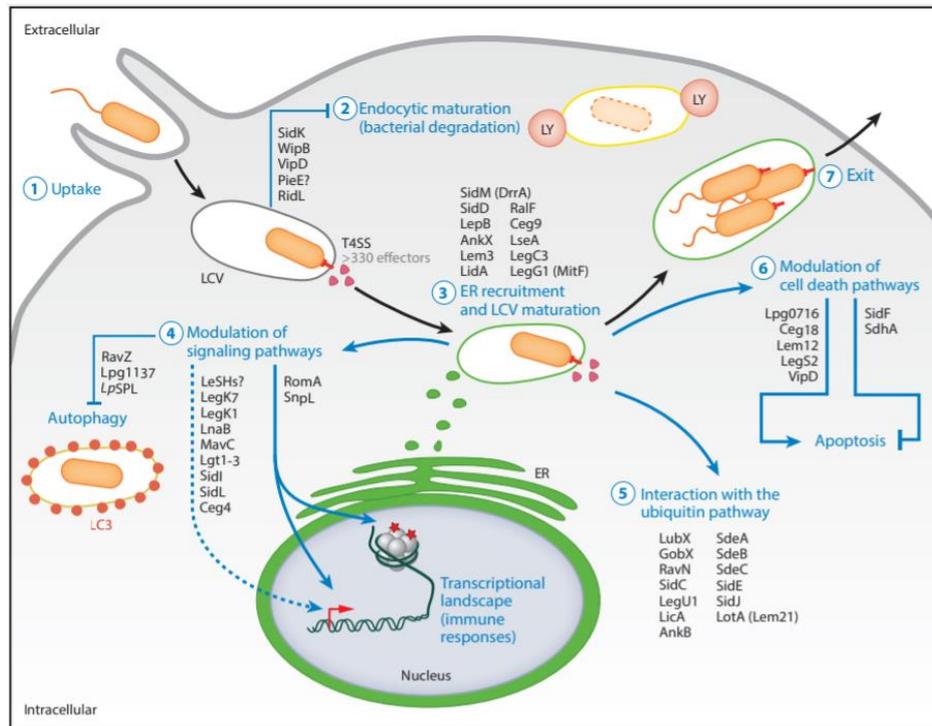


Figure 32: Rôles des effecteurs sécrétés par le SST4B dans le cycle intracellulaire de *L. pneumophila*
D'après (Mondino et al., 2020)

Face aux fluctuations environnementales imposées par ce cycle de réplication, *L. pneumophila* adopte un comportement biphasique passant d'une forme avirulente dite « réplivative » lorsqu'elle se multiplie au sein de la LCV à une forme virulente dite « transmissive » lorsqu'elle sort de la cellule infectée pour en infecter une nouvelle. Ce processus d'adaptations aux conditions environnementales est appelé « différenciation » (Molofsky and Swanson, 2004).

III.A.3.3. LetA/S : exemple d'un système de régulation pléiotropique conservé de *L. pneumophila*

Ce mécanisme de différenciation est favorisé par le système à deux composantes LetA/LetS de *L. pneumophila* qui est un bon exemple de système de régulation pléiotropique conservé. En plus d'être parfois impliqués dans la compétence comme vu précédemment, les TCS sont largement utilisés par les bactéries pour s'adapter aux fluctuations environnementales (Bijlsma and Groisman, 2003). LetA/S est un TCS particulier appartenant à une famille de protéines transductrices de signaux qui comprend les systèmes BvgA/BvgS de *Bordetella*, BarA/UvrY de *E. coli* (Pernestig et al., 2001) ou GacS/GacA de *Pseudomonas spp* (Marutani et al., 2008) caractérisés par un phosphorelay à quatre étapes His-Asp-His-Asp. Ces TCS particuliers induisent l'activation de la transcription de petits ARN non codants (ARNnc) de type Csr/Rsm (Kay et al., 2005; Lenz et al., 2005; Miller et al., 2016; Zere et al., 2015). Grâce à des séquences spécifiques, ces ARNnc se lient aux protéines de liaison à l'ARN CsrA (pour *L.*

pneumophila, *V. cholerae* ou *E. coli*) ou RsmA (pour *P. aeruginosa*) qui empêchent, en temps normal, la traduction des ARNm cibles en bloquant l'accès aux ribosomes. En présence des ARNnc, ces protéines régulatrices sont séquestrées et ne peuvent plus se lier au RBS, ce qui induit la traduction des ARNm et la réponse aux stimuli environnementaux (exemple de fonctionnement du système LetA/LetS chez *L. pneumophila* en **Figure 33**). De manière non exhaustive, ces TCS et les cascades régulatrices associées sont impliquées dans un large éventail de traits biologiques chez différentes espèces allant de la régulation des gènes de virulence (Herren et al., 2006; Martínez et al., 2011; Marutani et al., 2008), la formation de biofilms (Mitra et al., 2013), l'acquisition du fer (Frangipani et al., 2014), la production de toxines (Reimann et al., 1997), ou encore la régulation du quorum sensing (Lenz et al., 2005). Chez *Legionella*, le système LetA/LetS a été identifié pour la première fois dans un crible de mutants déficients dans l'expression de la flagelline (Hammer et al., 2002) et a depuis été montré comme étant impliqué dans l'activation de divers traits de virulence ainsi que dans la croissance intracellulaire chez les amibes (Bachman and Swanson, 2004; Gal-Mor and Segal, 2003; Shi et al., 2006). Comme on le verra en dans le chapitre II, nous avons montré que le système LetA/S était impliqué dans la régulation de la compétence de *L. pneumophila*.

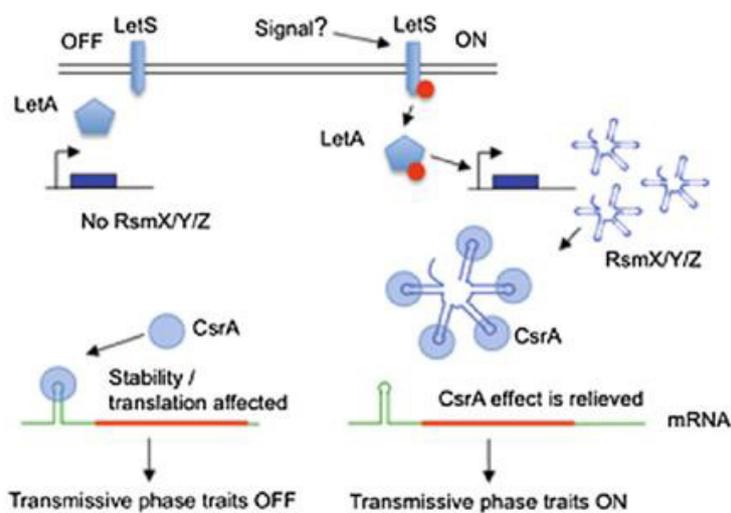


Figure 33 : Régulation de la phase transmissive chez *L. pneumophila*.

D'après (Hoffman et al., 2007). En l'absence de signaux inducteurs, CsrA bloque l'accès au RBS par le ribosome des gènes impliqués dans le passage à la phase transmissive. Sous l'induction de stimuli environnementaux, le senseur LetS s'autophosphoryle au niveau d'un résidu histine conservé, puis, le phosphate (en rouge) est transféré séquentiellement à l'acide aspartique et aux résidus histidines de LetS et enfin à un acide aspartique de LetA. LetA phosphorylé va alors activer la transcription de ses gènes cibles et notamment ceux codant pour les petits ARNnc RsmX/Y/Z. Ces ARNnc permettent la séquestration de CsrA qui ne peut plus aller bloquer l'accès des RBS des gènes cibles par le ribosome, ce qui déclenche la phase transmissive.

Le mode de vie intracellulaire de *L. pneumophila* qui a coévolué pendant des millions d'années au sein de ses hôtes eucaryotes a façonné son génome dans des dimensions que l'on commence seulement aujourd'hui à entrevoir comme on va le voir maintenant.

B. Une bactérie transformable

III.B.1. Des caractéristiques génomiques conservées...

Les principales caractéristiques génomiques de *L. pneumophila* ont été déterminées dans différentes études (D'Auria et al., 2010; Gomez-Valero and Buchrieser, 2013; Gomez-Valero et al., 2011) et sont résumées dans le **Tableau 1** dérivé de l'étude de Gomez-Valero et Buchrieser parue en 2013 portant sur l'analyse des génomes séquencés de 8 souches endémiques de *L. pneumophila* appartenant au Sg1.

| | (A) <i>L. pneumophila</i> | | | | | | | |
|--------------------------------|---------------------------|------------|--------------|------------|------------|------------|------------|------------|
| | Paris | Lens | Philadelphia | Corby | Alcoy | 130b | Lorraine | HL06041035 |
| Chromosome size (Mb) | 3.5 | 3.3 | 3.4 | 3.6 | 3.5 | 3.5 | 3.5 | 3.5 |
| G + C content (%) | 38 | 38 | 38 | 38 | 38 | 38 | 38 | 38 |
| Number of genes | 3178 | 3034 | 3083 | 3290 | 3197 | 3293 | 3170 | 3184 |
| Number of protein-coding genes | 3079 | 2921 | 2999 | 3193 | 3191 | 3288 | 3080 | 3079 |
| Pseudogenes | 45 | 59 | 32 | 44 | ? | ? | 37 | 53 |
| tRNA | 43 | 43 | 43 | 44 | ? | 42 | 44 | 43 |
| 16S/23S/5S | 03/03/2003 | 03/03/2003 | 03/03/2003 | 03/03/2003 | 03/03/2003 | 03/03/2003 | 03/03/2003 | 03/03/2003 |
| Coding density (%) | 87 | 87 | 88 | 87 | 86 | | 87 | 87 |
| Plasmids | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |

Tableau 1 : Principales caractéristiques des génomes de *L. pneumophila*.
D'après (Gomez-Valero and Buchrieser, 2013)

D'après cette étude, les principales caractéristiques génomiques sont conservées au sein des différentes souches de cette espèce. En effet, toutes possèdent un unique chromosome circulaire riche en adénine et en thymine (G+C content ~38%) dont la taille varie de 3,3Mb à 3,5Mb selon les souches et est constituée d'environ 3000 gènes. De plus, l'ordre des gènes est conservé à quelques exceptions près. L'analyse comparative des génomes montre qu'elles partagent environ ~2500 gènes ce qui représente un core génome d'environ 80% des gènes prédits et qu'elles ont entre 154 et 271 gènes spécifiques de souche ce qui représente un total de plus de 1000 gènes accessoires. Ce nombre est probablement largement sous-estimé, car des études récentes réalisées sur un large éventail de génomes de *Legionella* spp (et non uniquement de *L. pneumophila*) ont montré que le pangénome de *Legionella* spp était constitué d'au moins 30 000 gènes (Joseph et al., 2016). Ainsi, bien que la structure générale des génomes soit conservée, les souches de *L. pneumophila* et plus largement les souches du genre *Legionella* possèdent un génome accessoire très divergent. La plupart de ces gènes spécifiques de souche sont situés dans des îlots génomiques de résistance ou de pathogénicité (D'Auria et al.,

2010) ou sur des éléments génétiques mobiles probablement acquis par THG (voir paragraphe suivant).

Il est généralement admis que les parasites intracellulaires obligatoires subissent une réduction de leur génome au cours de leur co-évolution avec leur hôte, car ils perdent certaines fonctions redondantes exercées par ce dernier (Andersson and Kurland, 1998). Ce n'est pas le cas de *L. pneumophila* et en général des parasites d'amibes qui, au contraire, ont subi une extension de leur génome au cours de leur co-évolution avec leurs hôtes. En effet, ils présentent des génomes plus grands que les espèces non intracellulaires les plus proches (Moliner et al., 2010). Par exemple, le génome de *L. pneumophila* (~3,5Mb) est presque 3 fois plus large que le génome de *Coxiella burnetii* ou *Francisella tularensis* ses parents les plus proches. Ceci est encore plus frappant chez une autre espèce de Legionella : *L. drancourtii* qui possède un génome de plus de 4Mb soit le plus long génome du genre Legionella découvert à ce jour. Ceci suggère qu'au cours de son évolution avec son hôte, Legionella a acquis de nombreux gènes probablement par THG.

III.B.2. ... Dissimulant un génome hautement dynamique

On a longtemps pensé que *L. pneumophila* évoluait principalement par mutations spontanées transmises verticalement et possédait donc une structure de population clonale (Edwards et al., 2008; Selander et al., 1985). Aujourd'hui, alors que deux des trois principaux mécanismes de THG ont été décrits chez *L. pneumophila* (Dreyfus and Iglewski, 1985; Stone and Kwaik, 1999) [l'existence de phage de ce pathogène étant encore incertaine] et qu'un nombre croissant de génomes complets sont disponibles, des études récentes tendent à démontrer qu'à l'instar d'espèces hautement « recombino-gènes » tel que *S. pneumoniae* ou *N. meningitidis*, les événements de recombinaison ont un rôle important, si ce n'est prépondérant, dans l'évolution de *L. pneumophila*. En effet, une première étude à grande échelle basée sur l'analyse de 364 génomes d'isolats cliniques de *L. pneumophila* appartenant aux 5 principaux ST responsables de la majorité des cas européens de legionellose, a montré qu'au sein d'un même ST, plus de 95% des SNPs présents dans les génomes des isolats cliniques étaient hérités d'événements de recombinaison et que plus de 90% de ces SNPs proviennent de THG (David et al., 2016). Ces événements de recombinaison peuvent soit donner lieu au remplacement d'un gène par un variant allélique, soit donner lieu à une acquisition de nouveaux gènes.

Une autre récente étude portant sur l'analyse de plus de 291 génomes d'isolats cliniques appartenant aux 6 principaux ST responsable de la majorité des cas mondiaux de legionellose, a montré que parmi les plus de 95% de SNPs hérités d'événements de recombinaison, 33 à 80% dérivent potentiellement d'événement de recombinaison (David et al., 2017). D'après cette même étude, la

taille des tracts de recombinaison homologue serait relativement petite avec une majorité de tracts de taille inférieure à 10 kB même si de plus larges événements sont susceptibles de se produire, le plus grand tract observé atteignant presque 100kB chez le ST37. Les auteurs de cette étude suggèrent que ces événements de recombinaison homologue ont joué un rôle prédominant dans l'émergence de nouveaux pathogènes de *L. pneumophila* en permettant l'échange de gènes entre souches d'importance clinique. Même s'il n'existe pas de preuves formelles, on peut raisonnablement suggérer que ces événements ont été acquis par transformation naturelle puisque *L. pneumophila* est naturellement transformable et puisque c'est le seul mécanisme de THG existant chez *L. pneumophila* capable de réaliser ce type de recombinaison.

Outres ces transferts de gènes par transformation, il peut y avoir des transferts de gène intersouches voire interespèces médiées par conjugaison. En effet, en plus de disposer d'un SST4B essentiel pour sa réplication intracellulaire et aussi capable de transfert d'ADN (Guzmán-Herrador et al., 2017), il a été montré qu'un grand nombre de souches de Legionella, disposait de SST4A conjuguatifs (de type-F ou P ou des 2) suggérant que les transferts de gènes par conjugaison sont fréquents chez les espèces de ce genre bactérien (Gomez-Valero et al., 2011, 2019). Notamment, il a été montré que la région codant pour le SST4A Lvh, homologue au système vir de conjugaison d'*A. tumefaciens*, était présente dans 147 des 217 génomes analysés dans cette étude (Cazalet et al., 2008). Cette région est retrouvée totalement ou partiellement, intégrée dans les chromosomes ou excisée sous la forme de réplicons autonomes et a subi de nombreux réarrangements entre les souches, ce qui témoigne indubitablement d'une haute fréquence de transfert par conjugaison. Par ailleurs, les plasmides conjuguatifs pLPP (~130kB) et pLPL (~60kB) initialement découverts chez les souches de *L. pneumophila* Paris et Lens respectivement (Cazalet et al., 2004) sont en fait très répandus chez les différentes souches de *L. pneumophila* suggérant une large dissémination par conjugaison. Plus récemment, il a même été montré que différentes souches de *L. pneumophila* partageaient de larges portions continues de leurs chromosomes jusqu'à 200 kB avec des modifications mineures suggérant qu'ils ont été acquis récemment, la conjugaison étant le mécanisme envisagé par les auteurs (Gomez-Valero et al., 2011).

Enfin, une des caractéristiques majeures des génomes de *L. pneumophila* et de Legionella en général est qu'ils possèdent un très grand nombre de gènes de type eucaryote (*eukaryotic-like genes*). Ces gènes particuliers codent pour des protéines ou des motifs qui ont davantage de similitudes avec des protéines eucaryotes qu'avec des protéines procaryotes. D'après une étude récente réalisée sur 58 espèces de Legionella plus de 137 motifs eucaryotes et près de 300 protéines de types eucaryotes ont été retrouvés dans leur génome, la plupart codant pour des effecteurs du SST4 icm/dot (Gomez-Valero et al., 2019). Ils ont probablement été acquis par transfert de gène interrègne au cours de leur

évolution avec leur hôte eucaryote (Cazalet et al., 2004; Lurie-Weinberger et al., 2010). Bien qu'il n'existe pas encore de certitudes quant aux mécanismes d'acquisition de ces gènes particuliers, plusieurs hypothèses existent à ce jour (Gomez-Valero and Buchrieser, 2013). L'une d'entre elles stipule qu'ils pourraient avoir été acquis par un mécanisme d'évolution convergente durant laquelle les conditions imposées par la vie au sein de son hôte eucaryote auraient exercé une pression de sélection qui aurait naturellement fait converger la structure des protéines de *L. pneumophila* vers des structures de types eucaryotes. Cependant, ce type de mécanisme agit davantage sur la structure tridimensionnelle de la protéine et non sur les séquences des gènes qui pourtant, eux aussi, sont de types eucaryotes comme l'ont révélé les analyses génomiques. Une hypothèse originale serait que ces gènes de types eucaryotes puissent avoir été acquis par « transfert horizontal d'ARN » suivi d'une intégration par reverse transcription. Cette hypothèse est étayée par le fait que les gènes de type eucaryote présents chez *Legionella* sont dépourvus d'introns et qu'il existe chez *Legionella* une reverse transcriptase (Gomez-Valero and Buchrieser, 2013). Néanmoins, cette hypothèse n'a jamais encore été explorée et reste donc à vérifier. L'hypothèse la plus probable serait que ces gènes puissent avoir été acquis par transformation naturelle, car c'est le seul mécanisme de THG qui permettrait de réaliser un transfert interrègne : conjugaison et transduction faisant appel à des vecteurs qui ont une spécificité d'hôte et transférés uniquement entre bactéries (phage, ICE).

Ces échanges de gènes sont favorisés en contexte intracellulaire au sein des amibes qui sont considérées aujourd'hui comme les *melting-pots* de l'évolution (Moliner et al., 2010). En effet, elles hébergent de nombreuses espèces bactériennes distinctes du genre *Legionella*, mais aussi des virus géants avec lesquels *L. pneumophila* pourrait recombinaison, complexifiant encore les échanges. C'est probablement aussi parce que son contexte intracellulaire la protège des phages que les acquisitions de gènes par transduction n'ont pour l'heure pas encore été montré chez *L. pneumophila*.

III.B.3. Un mécanisme atypique de régulation de la compétence

La transformabilité de *L. pneumophila* a été décrite pour la première fois en 1998 soit plus de 20 ans après sa découverte (Stone and Abu Kwaik, 1998). Bien qu'initialement, *L. pneumophila* ait été décrite comme transformable à 37°C en condition microaérophile, on sait aujourd'hui que la majorité des souches de *L. pneumophila* deviennent compétentes dans des conditions suboptimales de température (Durieux et al., 2019) ou après exposition à des agents endommageant l'ADN (Charpentier et al., 2011). *L. pneumophila* devient compétente uniquement transitoirement dans une courte fenêtre qui commence à la moitié de sa phase exponentielle de croissance ($OD_{600} > 1,8$) et se termine juste avant son entrée en phase stationnaire ($OD_{600} < 4$) (Attaiech et al., 2016)(**Figure 1**). La

compétence de *L. pneumophila* est régulée par un mécanisme de répression post-transcriptionnelle et c'est à ce jour l'unique mécanisme de ce genre qui a été décrit chez une espèce transformable (Attaiech and Charpentier, 2017; Attaiech et al., 2016). Le processus fait intervenir deux partenaires qui sont conservés au sein du genre *Legionella* : une chaperonne à ARN RocC (initialement, faussement annotée « ProQ » par homologie partielle avec une protéine d'*E. coli*) et un petit ARN non codant (ARNnc) RocR qui interfère avec les ARNm des gènes du régulon de compétence selon un mécanisme de silencing classique (blocage de la traduction par masquage probable du site de fixation du ribosome) (**Figure 34**).

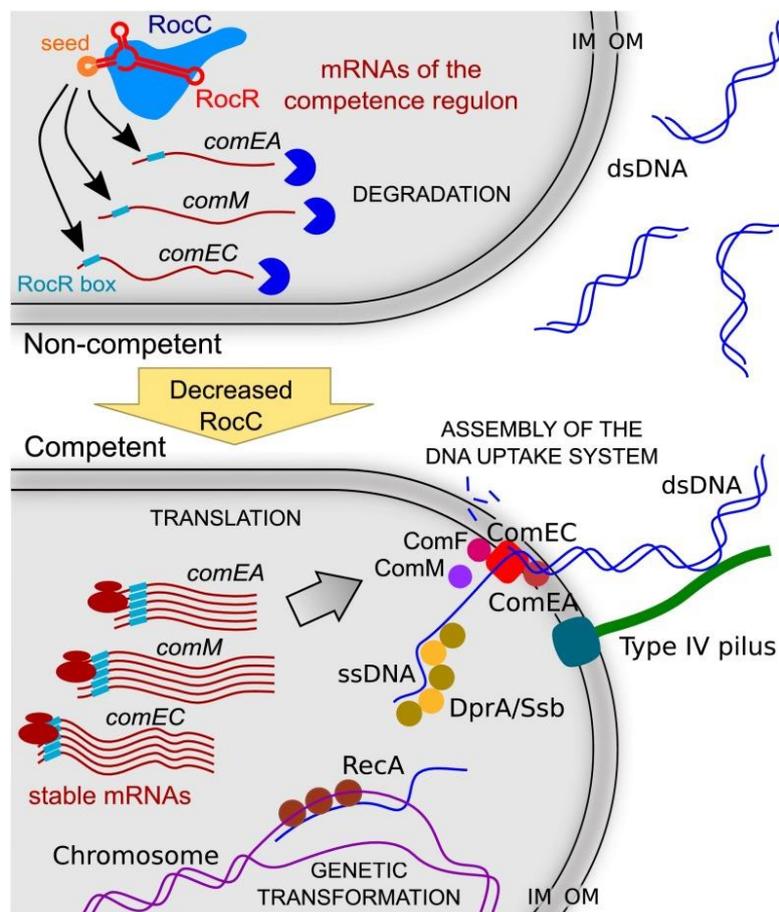


Figure 34: Régulation de la compétence chez *L. pneumophila*.

Extrait de (Attaiech et al., 2016). Lorsque les cellules ne sont pas compétentes, RocC stabilise RocR et ce complexe ribonucléoprotéique se fixe sur certains ARNm codant pour les constituants de la machinerie d'import, au niveau d'une petite séquence nucléotidique conservée de 6pb appelée RocR-box. Cela bloque l'accès au RBS par le ribosome et entraîne la dégradation de ces ARNm. Lorsque la compétence est induite, l'expression de RocC diminue par un mécanisme encore inconnu provoquant la déstabilisation de RocR, le RBS des ARNm de transformation devient alors accessible et ces ARNm sont traduits en protéines qui s'assemblent pour former la machinerie d'import de l'ADN exogène.

La mutation de l'un ou l'autre des partenaires empêche le silencing des gènes de compétence et les mutants deviennent par conséquent hypercompétents, ce qui montre que RocC et RocR agissent

directement comme répresseur de ce système. Par des cribles génétiques, deux autres mutants hypercompétents ont été identifiés : le mutant du gène codant pour la protéine ComR (Sexton and Vogel, 2004) et le mutant du gène codant pour la protéine RNaseR (Charpentier et al., 2008) sans pour autant qu'à ce jour des liens n'aient été établis entre ces protéines et RocC ou RocR. L'existence d'une part, de ces autres répresseurs, mais aussi d'autre part, de phénomènes inexplicables telle que la diminution de l'expression de RocC/RocR précédant la compétence, ou le fait que des stress réplcatifs puissent induire la compétence montrent que de nombreux aspects de ce mécanisme de régulation sont encore méconnus.

Ce mécanisme de régulation est d'autant plus complexe qu'il existe une grande variabilité des phénotypes de transformation au sein des différentes souches de *L. pneumophila* suggérant que ce système est modulable. Tandis que des souches telles que la souche Paris et plus généralement les isolats cliniques de ST1 présentent des fréquences de transformation élevées (10^{-4} - 10^{-6}), d'autres souches de *L. pneumophila* sont partiellement ou totalement réfractaires à la transformation comme c'est le cas de la souche Lens et de la majorité des isolats cliniques de ST47.

Récemment, il a été montré grâce à une méthode de *Genome Wide Association Study*, que, pour une partie de ces souches non transformables (dont fait partie Lens, mais pas les ST47), cette inhibition de la transformation était due à la présence d'un plasmide conjugatif qui interfère avec le mécanisme de régulation décrit précédemment (Durieux et al., 2019). Ce plasmide appelé pLPL, identifié initialement dans la souche Lens, code pour un petit ARNnc homologue à RocR nommé RocRp (pour RocR plasmidique). Il a été montré que RocRp substituait le rôle de RocR uniquement en son absence (c'est-à-dire à partir de la moitié de la phase exponentielle de croissance) en étant stabilisé par RocC. Ainsi, chez les souches porteuses du pLPL, il est en résulte un silencing continu des ARNm encodant les protéines nécessaires à la transformation tout au long de la croissance et donc une inhibition totale de la compétence. Cependant, le pLPL n'est présent que dans 1/3 des souches peu ou non transformables suggérant qu'il existe d'autres systèmes capables de moduler la compétence de *L. pneumophila*.

Projet de thèse

Bien que de nombreuses études ont permis d'élucider bien des aspects de la régulation de la compétence et de la machinerie d'import et de recombinaison de l'ADNt, ce mécanisme de transformation reste encore incomplet.

Chez les bactéries gram négatives, bien que les mécanismes par lesquels le pilus s'assemble et se rétracte pour capturer et internaliser l'ADNt semblent partiellement élucidés, à quelques exceptions près on ne connaît pas l'identité et le rôle des pilines mineures dans ce complexe macromoléculaire. Ensuite, les processus périplasmiques qui permettent de dégrader les ADNdb qui se retrouve en ADNsb une fois internalisé dans le cytoplasme sont encore largement méconnus. Existe-t-il une nucléase dédiée ? Ou est-ce ComEC qui accomplit ce rôle ? Chez les bactéries gram négatives l'internalisation de l'ADNt dans le cytoplasme est-il un processus actif nécessitant de l'énergie ? Existe-t-il comme c'est le cas chez les gram positives avec ComFA, une ATPase dédiée ? Quel est le rôle précis de ComF chez les et de son équivalent ComFC chez les gram positive dans cette internalisation ?

Les étapes cytoplasmiques qui aboutissent à la recombinaison de l'ADNt sont aussi très lacunaires et particulièrement chez les espèces gram négative. Pour ces dernières, les mécanismes envisagés durant ces étapes cytoplasmiques proviennent largement de données obtenues chez les bactéries gram positives *S. pneumoniae* et *B. subtilis* et des connaissances calquées sur la recombinaison générale. Seulement, la récente découverte de l'hélicase ComM qui n'est présente que chez les bactéries gram négatives et spécifiquement dédiée à la recombinaison de l'ADNt, tend à démontrer que la voie de recombinaison de l'ADNt est différente de celle de la recombinaison générale. Existe-t-il d'autres acteurs dédiés ou non à cette voie de recombinaison de l'ADNt ?

Par ailleurs, les mécanismes d'induction et de régulation de la compétence sont encore méconnus chez la plupart des espèces. En effet, il y a tout d'abord, les espèces telles que les entérobactéries pour lesquelles il existe des traces de THG dans leur génome et qui possèdent tous les constituants génétiques essentiels à la machinerie d'import de l'ADNt, mais pour lesquelles aucune évidence expérimentale de transformation naturelle n'existe. Ensuite, il y a les bactéries qui transforment à de hautes fréquences, mais pour lesquelles les mécanismes de régulation de la compétence sont encore totalement méconnus, c'est le cas par exemple des espèces du genre *Acinetobacter* : chez ces organismes la compétence est-elle régulée par un mécanisme similaire à ce qui existe déjà (activation par le quorum sensing, répression catabolique, induction par des signaux liés à l'habitat naturel) avec un régulateur central de la compétence ? Ou, comme c'est le cas de *L. pneumophila*, est-ce un mécanisme de régulation totalement inédit ? Enfin, il existe des espèces comme c'est le cas de *L.*

pneumophila, pour lesquelles certains aspects de la régulation de la compétence sont connus, mais encore très parcellaires. Outre le système central RocR/RocR, quels sont les autres éléments de régulation et qui pourraient agir en amont ?

Pour tenter de répondre à ces questions, dans cette thèse j'ai envisagé de rechercher l'ensemble des gènes impliqués dans la compétence et la transformation naturelle dans une même espèce. Une approche possible était la méthode Tn-seq, qui permet, à partir d'une banque de mutant saturante, d'identifier les gènes essentiels à une fonction biologique d'intérêt exerçant une pression de sélection. J'ai donc (1) mis au point une méthode de Tn-seq chez *L. pneumophila*. J'ai (2) utilisé cette méthode pour trouver les acteurs manquants du système de transformation de cette espèce. Cela m'a permis de trouver des acteurs spécifiques (impliqué dans la régulation de la compétence chez *L. pneumophila*) et des acteurs conservés (impliqué dans la machinerie). Finalement, j'ai (3) caractérisé plus en détail un tout nouvel acteur conservé chez les bactéries gram négatives, impliqué dans la recombinaison de l'ADNt.

L'ensemble de ces travaux ont permis d'obtenir une vision plus complète de la régulation de la compétence chez *L. pneumophila*, mais également de combler un vide de connaissance dans la voie de recombinaison de l'ADNt, et ce bien au-delà du modèle d'étude.

CHAPITRE II : RESULTATS

I. Mise au point d'une méthode de Transposition-sequencing chez *L. pneumophila*

Le premier objectif de ce travail de thèse était de mettre au point une méthode de Tn-seq (*Transposition-sequencing*) chez *L. pneumophila* afin de l'utiliser par la suite pour identifier les gènes impliqués dans le mécanisme de transformation naturelle chez cette espèce.

Le prérequis initial à la mise au point d'une méthode de Tn-seq est de disposer d'une banque de mutants de l'organisme d'intérêt qui soit la plus dense possible. En effet, la force d'une méthode de tn-seq repose sur le nombre de mutants uniques constituant la banque. Notre objectif fixé arbitrairement était de disposer de 10 à 100 fois plus de mutants que de gènes présents dans le génome de notre organisme d'intérêt de façon à ce que dans cette banque il y ait statistiquement au moins 10 à 100 mutants de chaque gène du génome. Cela nécessite de choisir (1) la souche, (2) le système de mutagène et (3) la technique de mutagenèse permettant de générer le plus de mutants possible.

Une fois la banque générée, il a fallu mettre au point une méthode permettant de préparer la banque en vue de cartographier les sites d'insertion des transposons. Pour cela, il existe un certain nombre de techniques (reviewées récemment ici : Cain et al., 2020) qui permettent d'amplifier et de séquencer les régions d'ADN chromosomique flanquant le transposon. Notre objectif était ici d'utiliser une technique adaptée au transposon utilisé pour la mutagenèse et qui soit facilement réalisable avec les outils à disposition dans notre laboratoire.

Ensuite, il a fallu déterminer les outils bio-informatiques les plus adaptés à notre méthode de cartographie des sites d'insertion permettant de (1) traiter les données de séquençage, (2) cartographier ces sites d'insertion et (3) d'identifier les gènes essentiels à la fonction testée. Là aussi, il existe un certain nombre d'outils bio-informatiques dédiés au Tn-seq dont les principaux sont listés dans le tableau 1 de la revue de Peng et al. publiée en 2017.

Enfin, il nous a fallu valider cette méthode en l'utilisant dans un crible permettant d'identifier les gènes requis pour un mécanisme déjà connu.

Dans l'étude qui suit, je détaille sous la forme d'un article informel, les étapes d'optimisation de cette méthode de Tn-seq chez *L. pneumophila* qui ont permis l'élaboration d'un protocole qui a été publié en 2019 dans le chapitre 7 du volume 1921 de la série « *Methods in Molecular Biology* » intitulée « *Legionella: Methods and Protocols* » qui recense l'ensemble des protocoles utilisés pour étudier diverses caractéristiques de ce pathogène intracellulaire (Buchrieser et Hilbi, 2019).

A. Article

INTRODUCTION

Notre compréhension de la plupart des grandes fonctions biologiques chez les bactéries est le plus souvent basée sur des méthodes de génétique conventionnelle : il s'agit de muter volontairement un gène par exemple, par l'insertion d'un transposon et d'étudier le phénotype du mutant afin d'identifier la fonction perdue ; on appelle ceci la mutagenèse insertionnelle par transposition.

Cette technique peut être appliquée à grande échelle par la génération d'une banque de mutants recouvrant l'intégralité des gènes d'un génome. Dans ce cas, c'est la propriété d'insertion aléatoire des transposons qui est utilisée et on parle de mutagenèse globale par transposition. Traditionnellement, chaque mutant peut être testé dans un crible phénotypique pour déterminer l'implication du gène muté dans le processus biologique considéré. De façon alternative, l'ensemble de la banque peut être soumise à un test phénotypique, à condition de pouvoir repérer chaque mutant et son devenir lors du test. Cela a été initialement développé avec les méthodes telles que la *Signature-Tagged Mutagenesis* (Mazurkiewicz et al., 2006) ou la *Transposon-Site Hybridization* (Murry et al., 2008). Cette dernière technique utilise l'hybridation sur des *whole genome microarrays* pour identifier le gène inactivé par l'insertion d'un transposon. Ces méthodes autorisent des « sélections négatives » *in vivo* qui permettent d'identifier des mutants qui ont perdu une propriété ayant pour conséquence leur élimination de la banque de mutants : par exemple, des mutants incapables de coloniser un hôte. Ces techniques ont permis l'identification de nombreux facteurs de virulence chez de nombreuses bactéries pathogènes. Néanmoins, elles pâtissent d'un certain nombre de problèmes liés aux microarrays elles-mêmes, comme leur coût de développement qui les rend inaccessibles pour des organismes peu étudiés.

Pour résoudre ces problèmes, grâce à l'essor des nouvelles technologies de séquençage, un ensemble de techniques regroupées sous le nom de « Transposition-sequencing » a vu le jour à la fin des années 2000 (Gawronski et al., 2009; Goodman et al., 2009; Langridge et al., 2009; van Opijnen et al., 2009). Ces approches ne dépendent pas d'une microarray préexistante, mais sont basées sur l'utilisation combinée de la mutagenèse globale par transposition et des méthodes de séquençage haut débit en parallèle (Massively Parallele Sequencing, MPS). Ces techniques suivent un workflow à peu près identique comprenant la génération d'une banque de mutants par mutagenèse globale par transposition, le passage de cette banque à travers un crible impliquant une forte sélection, l'extraction et la fragmentation de l'ADN des mutants à l'entrée et à la sortie du crible, la ligation d'un adaptateur aux fragments d'ADN pour faciliter l'amplification par PCR des sites d'insertion et enfin le séquençage des fragments amplifiés afin de déterminer la localisation des sites d'insertion des transposons sur le génome et d'identifier les gènes essentiels à la fonction testée (**Figure 1B**). Néanmoins, ces techniques possèdent quelques spécificités expliquées dans la revue de Chao et al. publiée en 2016 notamment en ce qui

concerne le choix du transposon et les étapes de digestion et d'analyse des données de séquençage. Aujourd'hui, après plus de dix ans d'existence, le Tn-seq a été combiné à d'autres techniques qui ont permis d'améliorer encore sa puissance pour permettre de répondre à des questions biologiques encore plus complexes (listés dans la revue de Cain et al. parue en 2020). Ainsi le Tn-seq est devenu une technique incontournable de la génétique bactérienne pour l'identification des gènes essentiels à l'échelle d'un génome entier si bien qu'on dispose aujourd'hui, pour la quasi-totalité des espèces bactériennes communément étudiées, des banques de mutants de transposon analysables par Tn-seq.

À mon arrivée au laboratoire en 2017, aucune banque de mutant saturante de *L. pneumophila* utilisable pour le Tn-seq n'existait encore. Dans cette étude nous nous proposons d'élaborer un protocole de Tn-seq chez ce pathogène. Sont décrites les étapes d'optimisation de cette méthode de Tn-seq de la construction de la banque de mutant de transposon jusqu'à sa validation par criblage de cette banque dans un contexte d'infection de macrophage et/ou d'amibes.

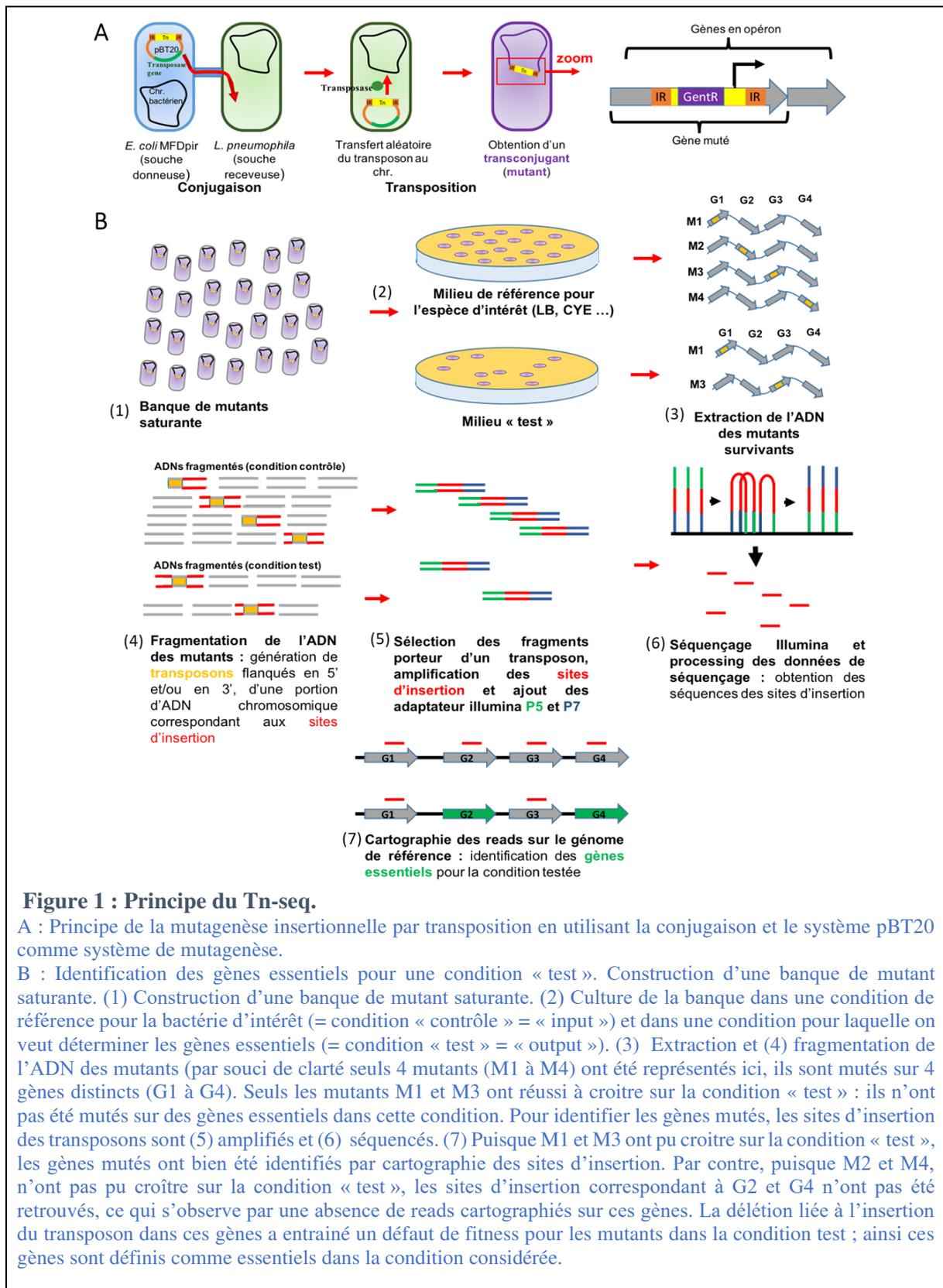


Figure 1 : Principe du Tn-seq.

A : Principe de la mutagenèse insertionnelle par transposition en utilisant la conjugaison et le système pBT20 comme système de mutagenèse.

B : Identification des gènes essentiels pour une condition « test ». Construction d'une banque de mutant saturante. (1) Construction d'une banque de mutant saturante. (2) Culture de la banque dans une condition de référence pour la bactérie d'intérêt (= condition « contrôle » = « input ») et dans une condition pour laquelle on veut déterminer les gènes essentiels (= condition « test » = « output »). (3) Extraction et (4) fragmentation de l'ADN des mutants (par souci de clarté seuls 4 mutants (M1 à M4) ont été représentés ici, ils sont mutés sur 4 gènes distincts (G1 à G4). Seuls les mutants M1 et M3 ont réussi à croître sur la condition « test » : ils n'ont pas été mutés sur des gènes essentiels dans cette condition. Pour identifier les gènes mutés, les sites d'insertion des transposons sont (5) amplifiés et (6) séquencés. (7) Puisque M1 et M3 ont pu croître sur la condition « test », les gènes mutés ont bien été identifiés par cartographie des sites d'insertion. Par contre, puisque M2 et M4, n'ont pas pu croître sur la condition « test », les sites d'insertion correspondant à G2 et G4 n'ont pas été retrouvés, ce qui s'observe par une absence de reads cartographiés sur ces gènes. La délétion liée à l'insertion du transposon dans ces gènes a entraîné un défaut de fitness pour les mutants dans la condition test ; ainsi ces gènes sont définis comme essentiels dans la condition considérée.

RÉSULTATS

1. Construction de la banque de mutant

Pour construire la banque de mutant de *L. pneumophila* par mutagenèse insertionnelle par transposition, une technique de conjugaison a été employée pour transférer le transposon à la souche à mutagénéiser (Figure 1A). Cette conjugaison a été réalisée en utilisant la souche d'*E. coli* MFDpir comme souche donneuse ; elle possède le système conjugatif RP4 (Ferrières et al., 2010) capable de délivrer chez *L. pneumophila* un plasmide non répliatif. Le plasmide est porteur du transposon utilisé pour la mutagenèse et d'un gène codant la transposase. Pour réaliser cette mutagenèse, deux plasmides ont été testés : le plasmide pBT20 et le plasmide pSC189, tous deux portent des variantes du transposon Mariner qui a la spécificité de s'insérer uniquement dans les sites dinucléotidiques « TA » présents dans le génome. Contrairement à celui porté par le pSC189, comme montré en Figure 1, le transposon porté par le pBT20 possède un promoteur sortant qui permet de limiter l'effet polaire imputé à l'insertion du transposon dans un opéron. Afin de tester la capacité de ces deux systèmes de mutagenèse à générer des mutants, un criblage a été réalisé sur différentes souches de *L. pneumophila*.

Ce criblage a été initialement réalisé sur des souches phylogénétiquement proches de la souche Paris couramment utilisée en laboratoire et toutes de *Sequence Type* 1 (ST1) (**Tableau 1, partie supérieure**). La mutagenèse réalisée avec le transposon du pSC189 n'a pas permis d'obtenir de mutants, pour aucune des souches testées. La mutagenèse réalisée avec le transposon du pBT20 a permis d'obtenir au maximum 60 UFC/100 μ L et seulement pour la souche LG07162016. Ainsi, pour ces souches ST1, le nombre de mutants obtenus par mutagenèse insertionnelle, aussi bien avec le pBT20 qu'avec le pSC189 est trop faible pour obtenir une banque de mutants saturante : ces souches ST1 ne semblent donc pas être de bonnes candidates pour la génération d'une banque de mutants saturante de *L. pneumophila*.

Dans un deuxième temps, le criblage a donc été réalisé sur d'autres isolats cliniques phylogénétiquement divers et différents du ST1 (**Tableau 1, partie inférieure**). La mutagenèse réalisée avec le transposon du pSC189 n'a permis d'obtenir un grand nombre de mutants que pour la souche HL07093014. Cependant, la mutagenèse réalisée avec le pBT20 a permis d'obtenir un grand nombre de mutants (>1000 UFC/100 μ L) pour cinq des huit souches testées à savoir : LG07242038, HL7102013, LG7211006, LG7245003, HL7093014. Ainsi, c'est avec le système de mutagenèse porté par le pBT20 que, par la suite, les expériences de mutagenèse insertionnelle par transposition pour la génération d'une banque de Tn-seq saturante chez *L. pneumophila* ont été menées et c'est ce système de mutagenèse que nous avons préconisé d'utiliser dans notre protocole de Tn-seq publié.

Parmi les cinq souches candidates pour la réalisation de cette banque de Tn-seq saturante, une a été sélectionnée pour poursuivre le protocole de Tn-seq ; il s'agit de la souche HL07093014 (renommé en HL77) dont le profil des boîtes obtenues après sélection des mutants est observable en **Figure 2**. Elle a été sélectionnée, car c'est celle qui nous a permis d'obtenir le plus de mutants. Après dilution en série

et dénombrement des tests de conjugaison réalisées avec cette souche, 28000 mutants (en UFC/mL) ont été obtenus en moyenne pour la mutagenèse/conjugaison réalisée avec HL77 comme souche réceptrice. De plus, cette souche a été choisie, car son efficacité de transformation est bonne et ceci est indispensable dans la perspective d'identifier par la méthode de Tn-seq, les gènes essentiels pour la transformation (voir partie « objectif 2 » de la thèse).

Pour constituer définitivement la banque de mutants, pour des raisons pratiques, cette souche a été rendue streptomycine résistante (HL77S) et une expérience finale a été effectuée en réalisant dix fois le protocole de conjugaison/sélection pour cette souche. Il en résulte que la banque de mutants de HL77S générée compte ~280 000 mutants. Il reste à déterminer par cartographie des sites d'insertion des transposons de combien de mutants uniques (site d'insertion unique) est constituée cette banque.

| Condition de sélection des mutants > | | Nombre de mutants sélectionnés en CFU/100µL | | | |
|--------------------------------------|----------------------|---|--------|---------------------------------------|------|
| | | pBT20 (TransposonMariner) | | pSC189 (TransposonMariner TnSC189) | |
| | | Gentamicine 10µg/mL | | Kanamycine 30µg/mL | |
| | | Contrôle | test | Contrôle | test |
| Souches « Paris » (ST1) | HL0631 2017 | 0 | 0 | n.c | n.c |
| | HL0636 5016 | 0 | 0 | n.c | n.c |
| | HL0638 5028 | 0 | 1 | n.c | n.c |
| | HL0639 2009 | 0 | 12 | 0 | 0 |
| | HL0639 2021 | 0 | 1 | n.c | n.c |
| | Paris | 0 | 4 | 0 | 0 |
| | HL0644 2011 | 0 | 60 | n.c | n.c |
| | HL0702 5017 | 0 | 3 | n.c | n.c |
| | LG0712 2011 | 0 | 0 | n.c | n.c |
| | LG0713 5012 | 0 | 0 | 0 | 0 |
| | LG0716 2016 | 0 | 3 | n.c | n.c |
| | LG0725 2030 | 0 | 1 | n.c | n.c |
| LG0725 4014 | 0 | 0 | n.c | n.c | |
| Isolats cliniques (non Paris) | LG0724 2038 | 0 | > 1000 | 0 | 0 |
| | HL 0710 2013 (ST75) | 0 | > 1000 | 0 | 3 |
| | HL 0641 3006 (ST701) | 0 | 1 | 0 | 0 |
| | HL 0714 2022 (ST59) | 0 | 0 | 0 | 0 |
| | LG 0721 1006 (ST702) | 0 | > 1000 | 0 | 0 |
| | LG 0724 5003 (ST22) | 0 | > 1000 | 0 | 0 |
| | LG 0727 5020 (ST23) | 0 | 0 | n.c | n.c |
| HL 0709 3014 (ST18) | 0 | > 1000 | 0 | > 1000 | |

Tableau 1 : Nombre de mutants obtenus par conjugaison en CFU/100µL pour chacune des souches de *L. pneumophila* testée avec les deux systèmes de mutagenèse de *E. coli* MFDpir : le pBT20 et le pSC189.

La sélection des mutants est réalisée respectivement sur milieu CYE supplémenté en gentamicine (10 µg/mL) ou sur milieu CYE supplémenté en kanamycine (30 µg/mL). Pour chacune des souches, une condition contrôle a été réalisée en remplaçant dans la conjugaison MFDpir par du PBS. Dans la partie supérieure du tableau est indiqué le nombre de mutants obtenus pour les souches « Paris » (de ST1) et dans la partie inférieure du tableau est indiqué le nombre de mutants obtenus pour des isolats cliniques de *L. pneumophila*, qui ne sont pas des ST1. Les deux souches sélectionnées pour constituer la banque de Tn-seq sont surlignées en vert.

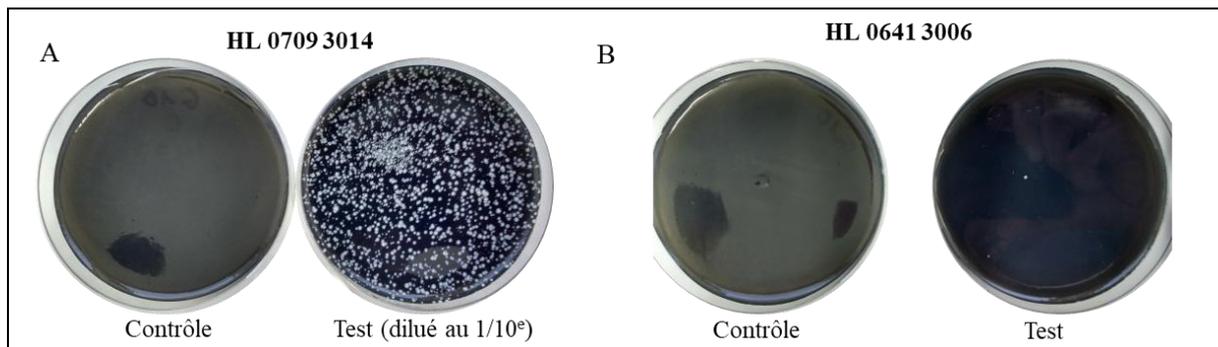


Figure 2 : Profil des boîtes obtenues après mutagenèse des souches HL 0709 3014 et HL 0641 3006 avec le système de mutagenèse pBT20. De nombreuses colonies correspondant à des transconjugants (mutants) de HL07093014 (HL77) ayant acquis le transposon GentR sont visibles sur les boîtes « test » (supplémenté en gentamicine) après mutagenèse avec le système pBT20 transféré par conjugaison. En revanche, lorsque la mutagenèse est réalisée sur une souche ST1 telle que HL06413006 presque aucun transconjugant n'est obtenu sur boîte « test » avec un profil de boîte ressemblant à la boîte « contrôle » (conjugaison réalisée en remplaçant la souche donneuse par du PBS).

2. Mise au point de la méthode de cartographie des sites insertions des transposons

La méthode de cartographie des sites d'insertion des transposons que nous avons mis au point pour *L. pneumophila* est la méthode dite « TdT ». Elle est basée sur l'HTML-PCR pour *homopolymer tail-mediated ligation PCR* qui a été développée par (Lazinski and Camilli, 2013). Elle s'effectue en plusieurs étapes (voir **Figure 3**).

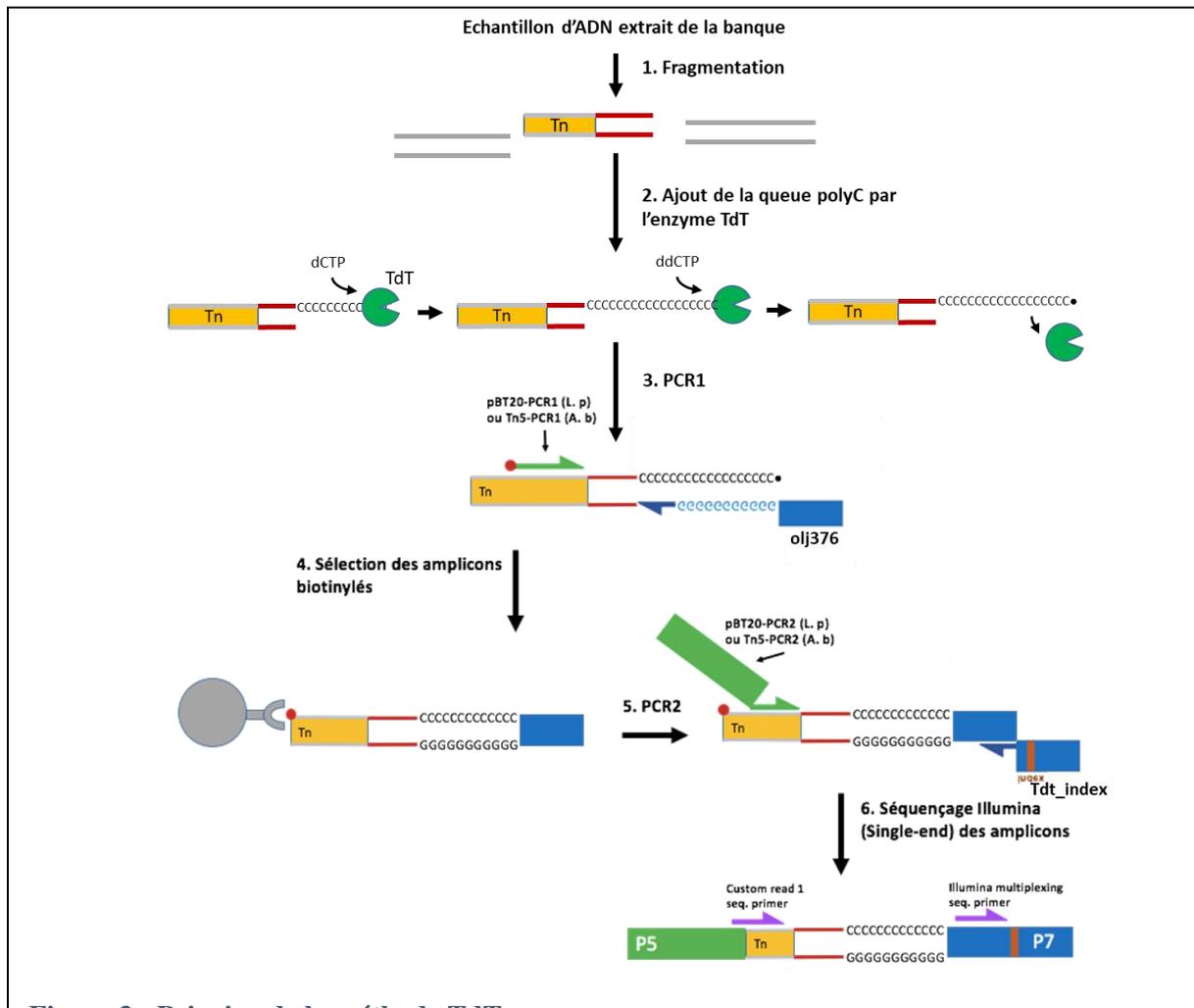
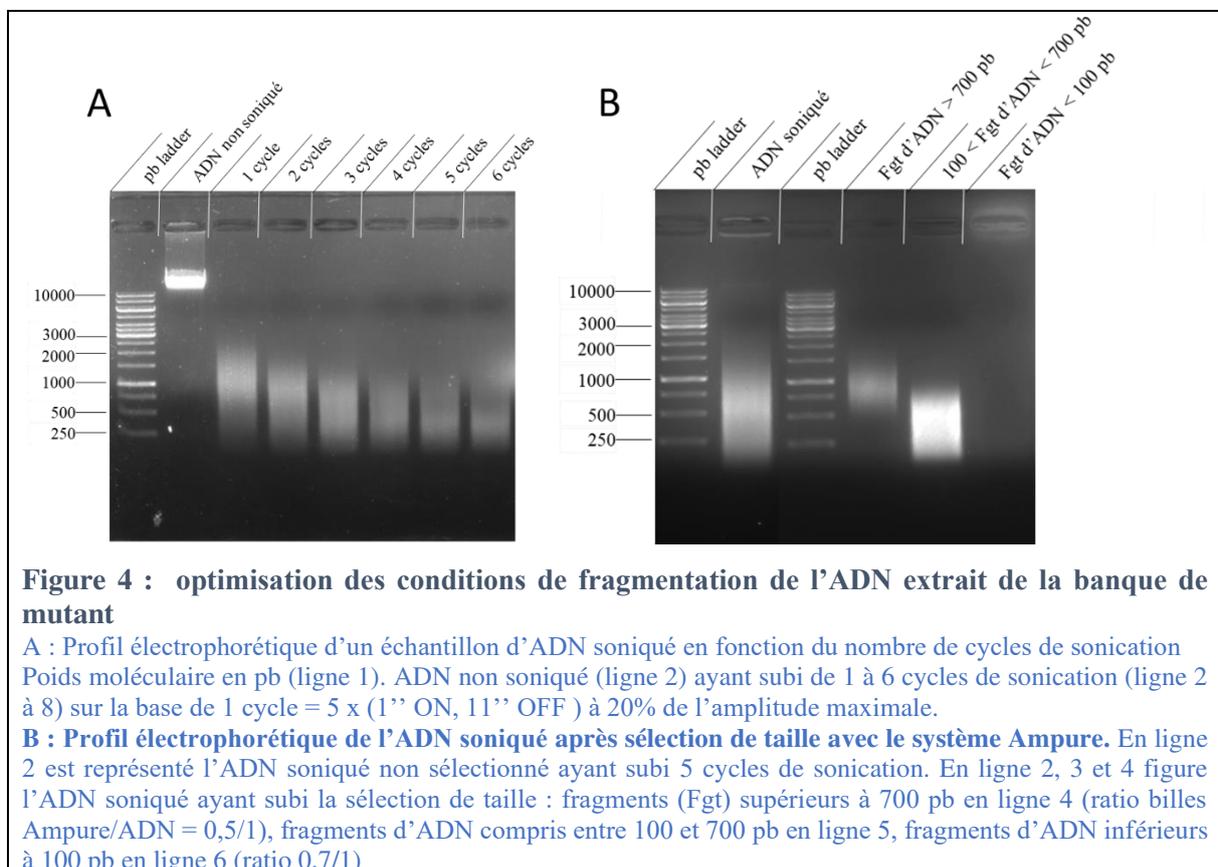


Figure 3 : Principe de la méthode TdT.

(1) Fragmentation mécanique de l'ADN extrait des banques de mutants et obtention de fragments porteurs ou non d'un transposon (en jaune), flanqué de portions d'ADN chromosomique correspondant aux sites d'insertion (en rouge) (2) Polymérisation de dCTP par l'enzyme TdT en 3' des fragments d'ADN (par souci de clarté seul un fragment porteur d'un transposon a été représenté, mais l'ajout de la queue polyC s'effectue pour l'ensemble des fragments d'ADN), la réaction de polymérisation se termine lorsque l'enzyme TdT incorpore un terminateur de chaîne ddCTP. (3) PCR1 réalisée sur les fragments d'ADN traités avec l'enzyme TdT grâce à un primer biotinylé (cercle rouge) spécifique du transposon utilisé (pBT20-PCR1 pour *L. pneumophila*) et un primer spécifique de la queue polyC porteur d'une partie de l'adaptateur Illumina P7 (olj376 en bleu). (4) Sélection des fragments biotinylés avec des billes recouvertes de streptavidine (en gris). (5) PCR2 réalisée sur les amplicons de la PCR1 (liés aux billes de streptavidine) grâce à un primer spécifique de la partie du transposon amplifiée lors de la PCR1 et un autre primer spécifique de la partie de l'adaptateur illumina ajoutée lors de la PCR1. (6) Séquençage Illumina single-end du site d'insertion grâce à un primer bordant le site d'insertion du transposon (Custom read 1 sequencing primer) et un primer servant à séquencer l'index pour identifier l'origine des séquences (en rose).

2.1. Fragmentation des banques d'ADNs et sélection de taille des fragments

La fragmentation de l'ADN extrait des banques de mutants est réalisée par sonication grâce à un sonicateur qui permet de générer des ultrasons de durée et d'intensité variable à partir d'une sonde plongée directement dans l'échantillon d'ADN à fragmenter. Ces paramètres d'intensité et de durée de l'onde transmise ont été optimisés dans l'objectif d'obtenir des fragments d'ADN compris entre 100 et 700 pb. Pour cela, le sonicateur a été paramétré de la façon suivante : 1 cycle = 1 minute = 5 x (1 s ON et 11 s OFF, duty cycle) à 20% de la puissance maximale. Pour un échantillon d'ADN, ce cycle a été réalisé 6 fois et après chacun de ces cycles, un aliquot d'ADN soniqué a été prélevé et utilisé pour réaliser une électrophorèse sur gel d'agarose de façon à déterminer la « gamme » de tailles des fragments obtenus après chaque cycle (**Figure 4A**). On observe que l'ADN, initialement présent sous la forme d'une bande de haut poids moléculaire (**Figure 4A, ligne 2**), est retrouvé sous la forme d'une trainée (smear) dès le premier cycle de sonication (**Figure 4A, ligne 3**) ; ce qui signifie que la sonication a bien entraîné la fragmentation de l'ADN. Plus le nombre de cycles de sonication est important, plus les fragments générés sont de petites tailles (**Figure 4A, ligne 3 à 6**). En effet, on passe d'une gamme de taille allant de ~100 à ~3000 pb après un premier cycle de sonication à une gamme de taille allant de ~0 à ~900 pb après six cycles de sonication. Après cette optimisation, les paramètres de fragmentation retenus dans notre protocole final sont les suivants : 4 cycles de sonication à 20% de la puissance maximale du sonicateur Branson (ce qui équivaut à 4 min (1 s ON et 11 s OFF, duty cycle)), ce qui génère des fragments allant de ~0 à ~1000 pb (**ligne 5**).



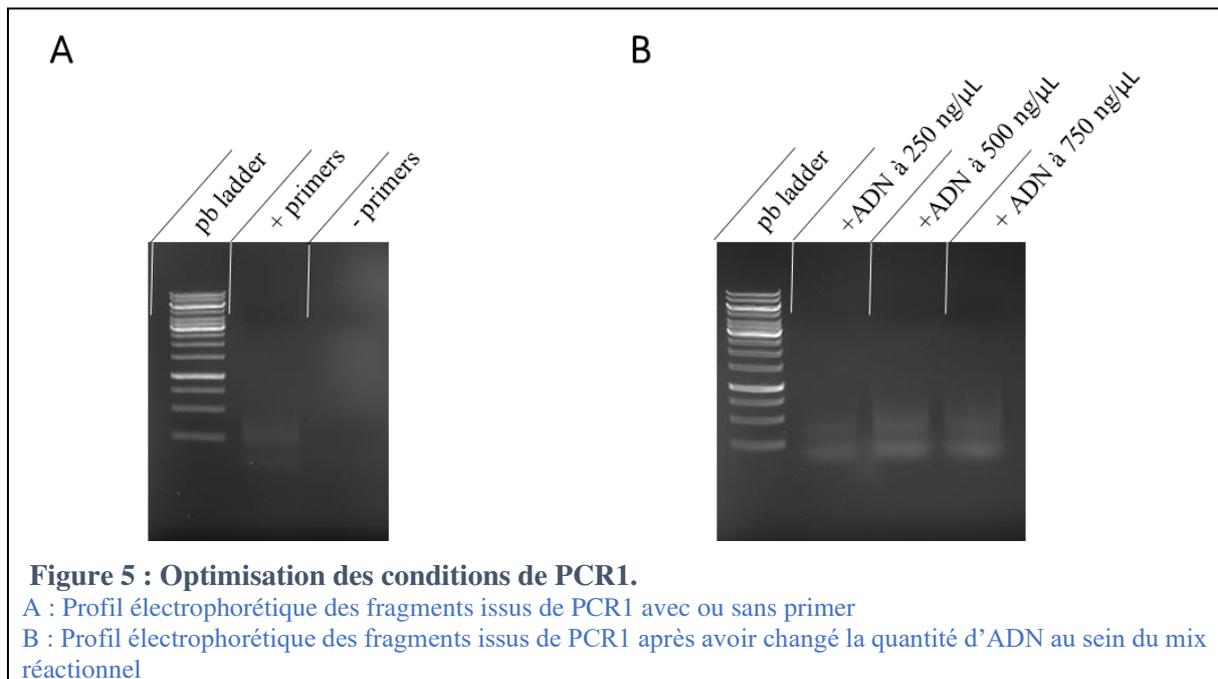
Pour éliminer les fragments indésirables (> 700 ou < 100), une sélection de taille a été réalisée avec la technologie Ampure qui permet de sélectionner ou d'éliminer des fragments de certaines tailles en fonction du ratio « volume de billes Ampure/volume de l'échantillon d'ADN ». D'après l'analyse électrophorétique d'aliqots prélevés après chaque étape de la sélection de taille, on observe qu'en plus d'être précis, ce système Ampure permet d'obtenir un bon rendement d'ADN, le signal obtenu sur gel pour les fragments de tailles d'intérêt étant de forte intensité (**Figure 4B**).

Finalement, d'après des expériences additionnelles (non montré ici) il s'avère que l'étape de sélection de taille permettant l'élimination des grands fragments n'est pas nécessaire. Ainsi, pour le protocole final de Tn-seq que nous avons publié, une seule sélection de taille dans le but d'enlever les petits fragments uniquement a été retenue en utilisant comme ratio 0,6/1 de billes Ampure par rapport au volume d'ADN à cette étape.

2.2. Amplification des sites d'insertion par PCR

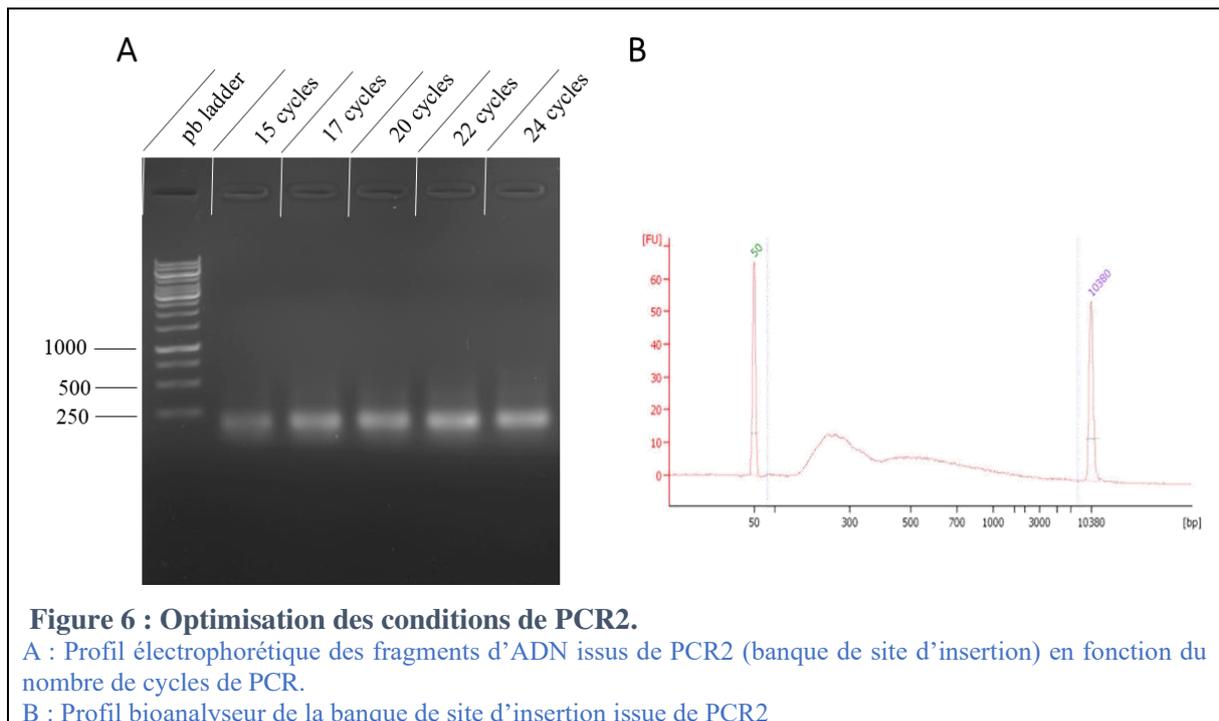
Les fragments d'ADN sélectionnés en taille ont été traités avec la TdT pour ajouter une queue polyC en 3' de chaque fragment. Cette queue polyC génère une matrice universelle pour l'amplification, lors de la PCR1, des sites d'insertion des transposons grâce à l'utilisation d'un primer polyG (olj376) complémentaire de cette queue polyC et un autre primer spécifique du transposon (pBT20-PCR1). Afin de contrôler l'efficacité d'amplification de la PCR1, les échantillons d'ADN amplifiés ont été analysés par électrophorèse sur gel d'agarose (**Figure 5**). Le signal observé sur gel s'apparente à un smear d'intensité très faible centré sur de faibles poids moléculaires. Pour tenter d'observer un signal d'intensité plus élevée, tenant compte d'une amplification plus efficace, différents paramètres ont été augmentés comme la quantité d'ADN introduite dans le mix PCR (**Figure 5**), la température d'hybridation des primers ou le nombre de cycles de PCR (données non illustrées). Dans tous les cas, le signal observé sur gel a toujours été de faible intensité. En réalité, cette PCR n'a pas vocation à générer une amplification exponentielle des sites d'insertion des transposons, mais uniquement à biotinyler les amplicons porteurs d'un transposon, de façon à les sélectionner par capture sur les billes de streptavidine.

Pour notre protocole final publié, nous avons retenu des conditions de PCR1 dans lesquelles 500 ng d'ADN-polyC ont été introduits dans le mix réactionnel avec 25 cycles de PCR1.



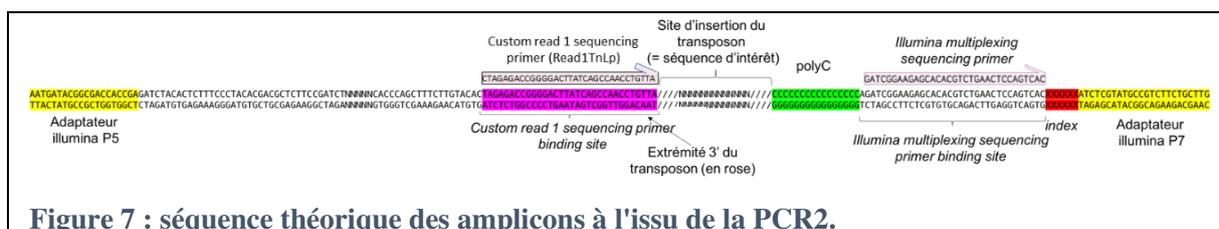
Contrairement à ce qui a été observé pour la PCR1, lorsque les échantillons d'ADN issus de la PCR2 sont analysés par électrophorèse sur gel d'agarose, un smear plus intense est observé (**Figure 6A**). D'après le profil des échantillons issu de PCR2 qui correspond à la banque de site d'insertion directement utilisable pour séquençage, les fragments d'ADN ont une distribution en taille bimodale avec une majorité des fragments compris entre 100 et 400 pb (**Figure 6B**). Pour cette PCR, les échantillons ont été soumis à différents nombres de cycles de PCR afin d'identifier les échantillons pour lesquels un niveau de saturation des amorces a été atteint. Ceci est une alternative à la réalisation d'une qPCR. Le niveau de saturation des amorces est atteint lorsque, sur gel, la bande correspondant aux amorces n'est plus observée ; cela correspond à la phase de plateau de la courbe caractéristique d'une qPCR.

La bande caractéristique des amorces n'étant déjà plus visible à 15 ou 17 cycles et voulant maximiser la quantité d'ADN utilisable pour le séquençage, pour notre protocole final, nous avons préconisé de réaliser la PCR2 avec 25 cycles de PCR.



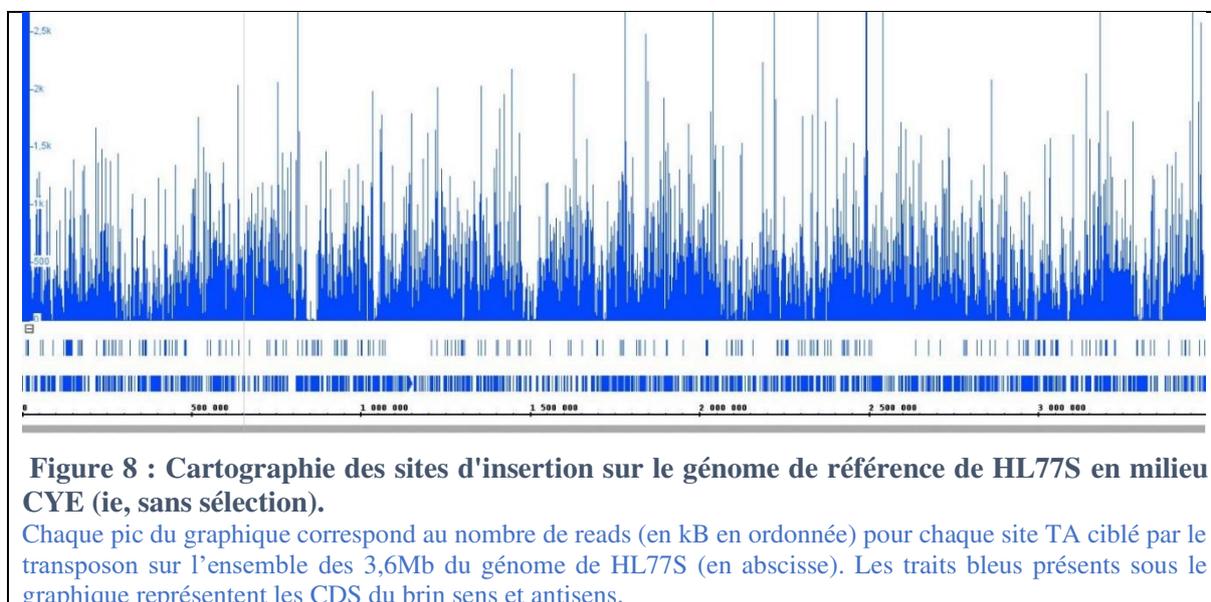
2.3. Séquençage des sites d'insertion des transposons

La structure théorique des amplicons obtenus à l'issue des deux PCR est donnée en **Figure 7**. Le primer de séquençage (*Custom read 1 sequencing primer* (Read1TnLp)) fourni au prestataire de séquençage est conçu de sorte que la première base « lue » lors du séquençage corresponde à la première base du site d'insertion du transposon (ie, la première base de la région d'ADNg flanquant le transposon). Expérimentalement, les séquences obtenues après séquençage Illumina des banques de site d'insertion sont très similaires à la structure théorique de la **Figure 7** excepté la partie comprenant l'adaptateur illumina P5 et le *custom read 1 sequencing primer* qui n'est pas séquençable. Comme attendu, la première base des reads obtenus correspond à la première base du site d'insertion du transposon qui correspond à la séquence d'intérêt à cartographier pour identifier la coordonnée génomique exacte de l'insertion du transposon. Ces séquences sont de tailles variables (dues à une fragmentation aléatoire de l'ADN en début de protocole). Elles sont suivies par une séquence polyC, elle aussi de taille variable en fonction des reads. En effet, même si le ratio de ddCTP/dCTP est égal à 1/20 dans le mix réactionnel TdT, nous nous attendions à obtenir une distribution de taille centrée autour de 20 pb, ce qui est le cas dans les séquences obtenues de nos différents runs.



2.4. Cartographie des sites d'insertion des transposons

Après séquençage et traitement des reads, en fonction des runs de séquençage 2,5 à 4 millions de reads ont été cartographiés sur le génome de références de HL77S (exemple en **Figure 8** de la cartographie obtenus après culture et analyse Tn-seq de la banque dans sa condition initiale d'isolement sur milieu CYE). Ce qui correspond à environ 60,000 insertions uniques. Le génome de HL77S étant constitué d'environ 250000 sites TA, ciblés spécifiquement par le transposon *Mariner* utilisé pour réaliser la mutagenèse, cela signifie qu'environ 1 site TA sur 4 est occupé par un transposon dans la banque de mutant. On observe des régions sans insertion qui correspondent aux régions génomiques (gènes ou opérons) essentiels à la viabilité de la bactérie. Cette analyse des gènes essentielles à la viabilité de *L. pneumophila* a été plus largement étudiée dans la partie « objectif 2 » de ce chapitre de résultat de cette thèse.



3. Validation de la méthode par culture de la banque de mutant dans un contexte d'infection

L'objectif ici n'est pas d'identifier de nouveaux gènes impliqués dans un mécanisme particulier, mais plutôt d'obtenir une preuve de concept que la banque de mutants et le protocole de Tn-seq que nous avons mis au point sont efficaces pour identifier des gènes requis pour une fonction d'intérêt. Une preuve d'efficacité serait d'identifier des gènes déjà connus pour être impliqués dans les processus d'infection et de multiplication intracellulaire très étudiés chez *L. pneumophila*. Pour cela, la banque de mutant d'HL77S a été cultivée en milieu liquide AYE qui sert ici de condition « contrôle » (ou d'« input »). En parallèle cette banque a été cultivée en présence des macrophages humain (U937) ou murin (RAW) et d'amibes (*A. castellanii*) qui servent ici de condition « test » ou « d'output ». Après application de la méthode de Tn-seq décrite ci-avant sur les mutants issus de la condition contrôle ou

ayant été soumis aux cribles de multiplication intracellulaire mentionnés ci-avant, une analyse d'essentialité conditionnelle a été réalisée avec la méthode statistique « resampling » du logiciel Transit. Cette méthode permet d'identifier les gènes pour lesquels il y a significativement plus ou moins d'insertion dans la condition « test » par rapport à la condition « contrôle ». Avec cette méthode, nous avons identifié 27, 19 et 5 gènes pour lesquels les insertions provoquent une diminution ($\log_2FC < -2$; $p\text{-value} < 0,05$) des mutants dans la population récoltée respectivement après infection des macrophages murins, humains et des amibes (**Tableau 2**). La plupart de ces gènes codent pour le SST4 Dot/Icm ou le système de flagelle de *L. pneumophila*, ce qui est cohérent avec une analyse transcriptomique réalisée précédemment en contexte d'infection des macrophages qui montrent que ces gènes sont parmi les plus surexprimés (Faucher et al., 2011). Ce sont même des opérons entiers qui semblent essentiels dans les conditions de multiplication intracellulaire comme nous pouvons l'observer au niveau de la région *icm* en **Figure 9** où visuellement on s'aperçoit qu'il y a moins d'insertion dans ces conditions de multiplication intracellulaire par rapport aux contrôles. Particulièrement, trois gènes *icmE*, *B* et *Q* sont retrouvés comme significativement essentiels en contexte de multiplication intracellulaire, quel que soit l'hôte.

| RAW | | | U937 | | | <i>Acanthamoeba castellanii</i> | | |
|---------------|--------|---------|---------------|--------|---------|---------------------------------|--------|---------|
| #Orf | log2FC | p-value | #Orf | log2FC | p-value | #Orf | log2FC | p-value |
| lcmE (DotG) | -3,89 | 0 | barA | -2,2 | 0,0007 | lcmE (DotG) | -3,07 | 0,0007 |
| lcmB (DotO) | -2,95 | 0,0002 | flgE | -2,26 | 0,0276 | lcmB (DotO) | -2,06 | 0,0025 |
| DotA protein | -3,12 | 0,0003 | flgF | -4,44 | 0,0207 | DotA protein | -2,37 | 0,004 |
| magA | -4,98 | 0,0004 | lcmB (DotO) | -2,81 | 0,0002 | yrbG | -2,59 | 0,0191 |
| lcmK (DotH) | -2,96 | 0,0008 | lcmC (DotE) | -2,38 | 0,0188 | lcmQ proteir | -2,58 | 0,045 |
| lcmF protein | -2,88 | 0,001 | lcmE (DotG) | -3,69 | 0 | | | |
| zraS_1 | -2,63 | 0,0048 | lcmF protein | -2,52 | 0,0041 | | | |
| ptsP | -2,64 | 0,0109 | lcmK (DotH) | -2,7 | 0,001 | | | |
| lcmL (DotI) p | -4,2 | 0,0119 | lcmL (DotI) p | -3,97 | 0,0131 | | | |
| flgF | -4,7 | 0,0121 | lcmO (DotL) | -3,03 | 0,0206 | | | |
| flgI | -3,04 | 0,0144 | lcmQ proteir | -3,22 | 0,0263 | | | |
| fliC | -2,06 | 0,0167 | lcmV protein | -4,36 | 0,0467 | | | |
| tamB | -3,15 | 0,0169 | magA | -3,08 | 0,0336 | | | |
| lcmO (DotL) | -3,17 | 0,0187 | nhaS3 | -2,2 | 0,0348 | | | |
| General secr | -3,4 | 0,0222 | OmpA/MotB | -3,07 | 0,0422 | | | |
| flgE | -2,46 | 0,0227 | tamB | -3,03 | 0,0223 | | | |
| lcmC (DotE) | -2,2 | 0,0254 | uvrY | -2,61 | 0,0109 | | | |
| nhaS3 | -2,33 | 0,0268 | yrrB | -2,01 | 0,0394 | | | |
| flgL | -3,24 | 0,0269 | zraS_1 | -2,07 | 0,0173 | | | |
| lcmQ proteir | -3,09 | 0,0303 | | | | | | |
| greA | -2,81 | 0,0326 | | | | | | |
| OmpA/MotB | -3,27 | 0,0328 | | | | | | |
| lcmV protein | -4,61 | 0,0421 | | | | | | |
| rsxB | -4,01 | 0,0473 | | | | | | |
| dtpA_4 | -3,06 | 0,0476 | | | | | | |
| aroB | -3,09 | 0,0484 | | | | | | |
| ptsH | -2,98 | 0,0489 | | | | | | |

Tableau 2 : Gènes conditionnellement essentiels identifiés par Tn-seq requis pour l'infection et/ou la multiplication intracellulaire de *L. pneumophila* HL77S dans les macrophages murins, humains et dans les amibes.

Ne sont représentés ici que les gènes ayant significativement moins d'insertion dans les conditions de multiplication intracellulaire par rapport à la condition contrôle (culture en AYE) ayant un log2FC < -2 et une p-value < 0,05. En rouge sont entourés les gènes retrouvés par Tn-seq dans deux des trois conditions de multiplication intracellulaire (RAW, U937 ou *A. castellanii*) et en vert les gènes retrouvés dans les 3 conditions.

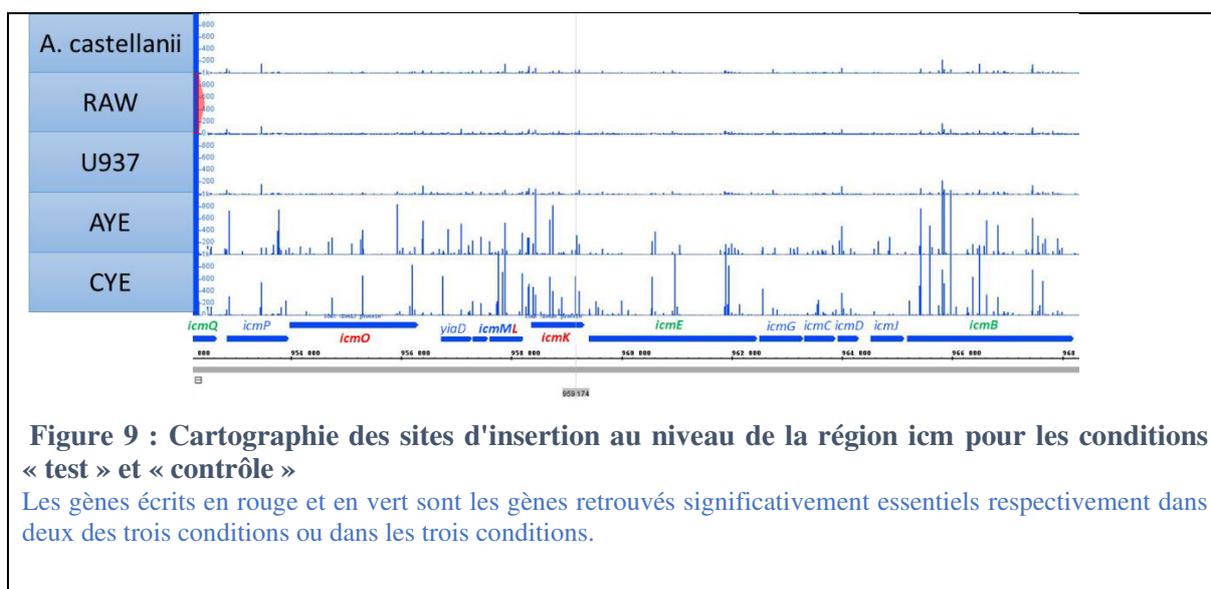


Figure 9 : Cartographie des sites d'insertion au niveau de la région icm pour les conditions « test » et « contrôle »

Les gènes écrits en rouge et en vert sont les gènes retrouvés significativement essentiels respectivement dans deux des trois conditions ou dans les trois conditions.

DISCUSSION ET CONCLUSION

Dans cette étude, nous avons mis au point une méthode de Tn-seq pour la bactérie pathogène *L. pneumophila*. La méthode utilisée ici est différente de la technique originelle de Tn-seq, mais elle devient de plus en plus répandue. En effet, la méthode originelle de Tn-seq développée par Van Opijnen (van Opijnen et al., 2009) n'est réalisable qu'avec un transposon *Mariner* particulier : il s'agit d'un transposon *HimarI* modifié au niveau de ses séquences inversées répétées pour constituer un site de reconnaissance pour l'enzyme de restriction MmeI. Lorsque cette enzyme reconnaît ce site, elle clive l'ADN chromosomique à une distance fixe de 16 nucléotides de son site de reconnaissance, ce qui entraîne l'excision du transposon qui se retrouve flanqué en 5' et en 3' de séquences de taille unique d'ADN chromosomique, correspondant aux sites d'insertion du transposon. Par rapport à cette technique « MmeI », la technique « TdT » que nous avons mise en œuvre, possède un certain nombre d'avantages. Premièrement, elle est applicable à tout type de transposon et c'est un avantage considérable compte tenu du fait que tous les transposons n'ont pas la même efficacité de mutagenèse comme on l'a vu avec le transposon porté par le système de mutagenèse pSC189. D'autre part, avec la technique TdT la taille des séquences correspondant aux sites d'insertion des transposons est plus importante que celles générées avec la technique MmeI. En effet, la taille des séquences correspondant aux sites d'insertion peut s'étendre à plusieurs dizaines de pb (dans la limite de 50 pb imposée par le mode de séquençage) avec la technique TdT, ce qui est bien plus important qu'avec la technique MmeI qui ne génère que des sites d'insertion de taille fixe de 16 pb. L'obtention de sites d'insertion « longs » permet d'effectuer de manière plus précise leur cartographie sur le génome de référence de la bactérie d'intérêt.

Le fait que nous n'ayons pas réussi à mutagénéiser les souches de *L. pneumophila* de ST1 est intrigant. Ni cette technique de conjugaison/transposition ni d'autres techniques utilisées au laboratoire n'ont permis d'obtenir un grand nombre de transconjugants pour ces souches de ST1 (Brunel, 2016). À l'heure actuelle, nous ne savons pas si ces souches sont imperméables à la conjugaison ou à la transposition. Néanmoins, ce comportement réfractaire vis-à-vis de l'un et/ou de l'autre de ces mécanismes de transfert de gène rappelle ce qui est observé chez un certain nombre de souches de *L. pneumophila* réfractaires à la transformation naturelle comme la souche *Lens* (Durieux et al., 2019). Comme discuté en introduction de cette thèse, une méthode de *Genome Wide Association Study* a permis de montrer que ce phénomène était dû à un plasmide conjugatif interférant avec le mécanisme de régulation RocC/RocR dans ces souches. Une telle méthode pourrait être appliquée à un certain nombre d'isolats cliniques de *L. pneumophila* réfractaire à la conjugaison/transposition pour tenter d'élucider les mécanismes sous-jacents de barrière à ces transferts de gènes.

Le nombre de reads obtenus (entre 5 et 6 millions par condition) après séquençage des banques cultivées dans les différentes conditions appliquées dans cette étude (multiplication intracellulaire et CYE/AYE) est très disparate en fonction des conditions et assez faible compte tenu du fait que notre prestataire prévoit plutôt entre 10 à 15 millions de reads par condition (150 millions de reads par run).

Pour résoudre ce problème, pour les études suivantes nous avons réalisé un pré-run de séquençage qui permet de déterminer les abondances relatives en termes de nombre de reads de chaque banque (identifiées par un barcode spécifique) et de relancer après coup le run en rééquilibrant les quantités d'ADN introduite. Cela a considérablement augmenté et rééquilibré le nombre de reads obtenus en fin de run, plus en adéquation avec les prévisions du prestataire de séquençage. De plus, seuls ~45% des reads obtenus sont réellement cartographiés sur le génome de référence, une grande partie des reads étant perdue au cours des étapes bio-informatiques de traitement, de filtration et même de cartographie des reads. Cela peut être dû au fait que les fragments générés au cours des étapes de fragmentation et de sélection de taille sont trop courts et ne passent pas le filtre de sélection de taille appliqué bio-informatiquement (voir matériel et méthode ou protocole en Annexe). Pour cela, comme mentionné plus haut et comme nous l'avons indiqué dans le protocole publié, nous avons essayé de favoriser les grands fragments en n'appliquant pas de sélection de taille « vers le haut ».

Après cartographie des sites d'insertion issus des mutants cultivés en condition de référence CYE, environ 60000 sites uniques ont été détectés ce qui signifie que la banque est constituée de 60000 mutants uniques. Le nombre de gènes du génome de *L. pneumophila* étant approximativement composé de 3000 gènes, il y a statistiquement 20 mutants par gène du génome dans cette banque ce qui est en adéquation avec l'objectif fixé de densité de la banque qui est souvent considérée comme saturante lorsque qu'il y a 10 à 100 fois plus de mutants dans la banque que de gènes dans le génome de la bactérie d'intérêt. Dans l'étude suivante, l'optimisation des paramètres de séquençage qui a permis d'obtenir un nombre plus important de reads a permis de réévaluer ce nombre à la hausse. On considère maintenant que la banque de HL77S que nous avons générée ici est constituée de 110000 insertions uniques, soit une insertion toutes les 31 pb ou autrement dit 36 mutants par gène du génome.

Bien qu'encore préliminaires, les données de Tn-seq obtenues en contexte d'infection étant cohérentes avec la littérature permettent d'apporter une preuve suffisante que (1) la banque de mutant d'HL77S générée est assez dense pour identifier de façon significative des gènes impliqués dans un mécanisme d'intérêt et (2) que les méthodes de préparation, de séquençage et de cartographie des sites d'insertion sont fonctionnelles.

La banque de mutant et le protocole mis en œuvre et publié (Annexe) sont donc utilisables pour identifier les gènes requis pour d'autres fonctions biologiques. Notamment, cette banque et ce protocole ont été utilisés pour identifier de façon exhaustive les gènes impliqués dans le processus de transformation naturelle chez *L. pneumophila*.

MATERIEL ET METHODES

Pour plus de détails, se référer à l'article méthodologique présent en Annexe.

Souches, amorces et plasmides.

Toutes les manipulations des souches de *L. pneumophila* ont été réalisées en laboratoire de biosécurité de niveau 2. Les souches de *L. pneumophila* utilisées sont des isolats cliniques issus du Centre National de Reference des Legionnelles. Les génomes des souches de cette étude sont séquencés et annotés. Les systèmes de mutagenèse portés par les plasmides pBT20 (Powell et al., 2016) et pSC189 (Chiang and Rubin, 2002) ont été utilisés indépendamment pour réaliser la mutagenèse insertionnelle par transposition chez les souches de *L. pneumophila*. Ces plasmides sont portés par la souche de *Escherichia coli* MFDpir qui est auxotrophe au 2,3-DiAminoPropionic acid (DAP) et qui est utilisé comme souche donneuse pour la conjugaison. Ces plasmides portent tous deux un transposon de type *Mariner* toutefois, dotés d'un gène de sélection différent : un gène de résistance à la gentamicine pour le transposon porté par le pBT20 et un gène de résistance à la kanamycine pour le transposon porté par le pSC189. Les amorces utilisées pour réaliser les PCR de cette étude sont publiées dans la partie 2.2 de l'article méthodologique présent en annexe.

Milieux de culture.

La souche d'*E. coli* MFDpir auxotrophe au DAP a été cultivée en milieu solide LB agar (Lysogeny Broth agar) ou en milieu LB liquide supplémenté en ampicilline (100µg/mL) et en DAP (0,01%) à 37°C pendant 24 heures. Les souches de *L. pneumophila* ont été cultivées, sauf mention contraire, en milieu liquide AYE (ACES-buffered Yeast Extract) à 37°C ou en milieu solide CYE (ACES – buffered Charcoal, Yeast Extract) à 37°C. Le temps d'incubation pour obtenir des bactéries en phase stationnaire, en milieu liquide, est d'une nuit à partir d'une colonie isolée sur boîte. En milieu solide, le temps d'incubation pour obtenir des colonies isolées est de 3 à 4 jours. Pour réaliser des sélections, le milieu CYE a pu être supplémenté en antibiotique comme de la kanamycine (30µg/mL) ou de la gentamicine (10µg/mL).

Construction de la banque de mutants de *Legionella pneumophila* par conjugaison.

La conjugaison est réalisée en co-cultivant la souche donneuse MFDpir (portant le pBT20 ou le pSC189) avec une souche de *L. pneumophila* à mutagéniser. Cette co-culture est réalisée à partir de 1 mL de cultures liquides en phase stationnaire de ces deux souches. Les cultures de *L. pneumophila* et *E. coli* sont centrifugées (10,000g ; 1 min) et resuspendues dans 50 µL d'eau et 25 µL de PBS, respectivement. Les deux concentrats sont mélangés et le mélange (75 µL) est ensuite déposé sur un milieu CYE dépourvu de fer et cystéine, mais supplémenté en DAP (0,01%). Ces conditions sont indispensables à la survie de MFDpir tout en permettant la croissance de *L. pneumophila*, car la présence habituelle de fer et de cystéine d'un milieu CYE est toxique pour cette souche d'*E. coli*. Un contrôle négatif est réalisé en remplaçant MFDpir par du PBS. Après six heures d'incubation à 37°C, les spots « test » et « contrôle », sont resuspendus dans 1 mL d'eau stérile. Les mutants de *L. pneumophila* ayant

intégré le transposon sont dénombrés sur des boîtes CYE supplémentées en gentamicine (10 µg/mL) ou en kanamycine (30µg/mL) selon le transposon utilisé. Pour déterminer le nombre de mutants générés, ceux-ci sont directement dénombrés sur boîte en UFC/100µL sauf mention contraire. Dans certains cas, le dénombrement a nécessité la dilution en série du mélange de conjugaison jusqu'à une dilution au 10⁴. Le nombre de mutants obtenus pour une souche de *L. pneumophila* est déterminé en réalisant le différentiel du nombre d'UFC/100µL obtenus pour la condition « test » par rapport à la condition « contrôle ».

Extraction et purification de l'ADN génomique.

L'extraction de l'ADN des banques s'effectue après culture des banques *L. pneumophila* dans les milieux « test » respectifs. L'objectif ici est de passer d'une banque de mutants à une banque d'ADN mutés. L'extraction et la purification de l'ADN ont été réalisées à partir de 1 mL de culture liquide avec le *Wizard® Genomic DNA purification kit* (Promega) selon les recommandations du fabricant et aboutissent à l'élution de l'ADN dans 50µL de Tris HCl (10 mM, pH8). Sauf indication contraire, la concentration et la pureté des préparations d'ADNs ont été déterminées par mesure de l'absorbance à 260 et 280 nm.

Amplification des sites d'insertion des transposons et préparation des banques pour le séquençage.

Les ADN extraits des banques sont fragmentés par sonication en utilisant un sonicateur Branson réglé sur le mode « pulse » et paramétré de la façon suivante 4 min (1 s « ON » et 11 s « OFF », *duty cycle*) à 20% de l'amplitude maximale. Une sélection de taille des fragments générés est menée grâce à l'utilisation de billes Ampure (Beckman Coulter). Selon les recommandations du fabricant, un ratio de 0,6/1 pour éliminer les fragments inférieurs à 100 pb.

Une queue poly-Cytosine (polyC) est ajoutée en 3' de l'ensemble des fragments d'ADN générés à l'issue de l'étape de fragmentation par polymérisation de dCTP catalysée par l'enzyme Terminal deoxynucleotidyl Transferase (TdT). La réaction est réalisée dans un volume final de 50µL en introduisant : 3µg d'ADN fragmenté, 1,25µL de TdT (Promega), 10µL de 5x *rTdT reaction buffer* et 2,5µL d'un mélange de dCTP (9,5 mM) et de ddCTP (0,5 mM). Ce mix réactionnel est incubé à 37°C pendant 1 heure puis soumis à une étape d'inactivation enzymatique pendant 20 min à 75°C. Le terminateur de chaîne ddCTP, ajouté au mix réactionnel, permet de limiter la taille de la queue polyC en bloquant l'incorporation de nouveaux dCTPs par l'enzyme TdT. A l'issue de cette réaction TdT, une purification avec les billes Ampure est réalisée avec un ratio de 1/1 pour éliminer les petits fragments (<100 bp) ainsi que l'ensemble des réactifs qui ont servi à réaliser la réaction TdT.

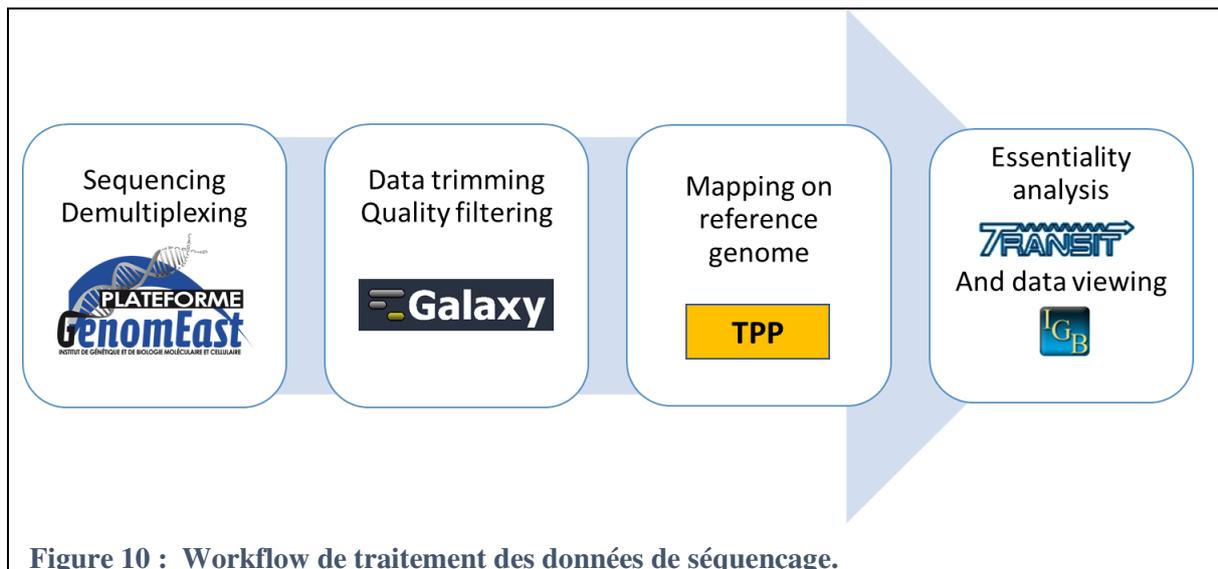
Pour amplifier les sites d'insertion des transposons, les banques d'ADN fragmentés, traités avec l'enzyme TdT sont soumises à deux PCR successives. La première PCR (PCR1) est menée dans un volume final de 50µL, en introduisant : 500 ng d'ADN fragmenté – polyC, 2,5µL de dNTPs (10 mM), 1µL de primer pBT20_PCR1 biotinylé (30 µM), 3µL du primer olj376 (30 µM), 0,75µL d'ADN

polymérase Q5 (New England Biolabs) et 10µL de 5x *reaction buffer*. Le thermocycleur est programmé de la façon suivante : 98°C, 5 min ; 25 x (98°C, 30 sec.; 65°C, 30 sec.; 72°C, 30 sec.) ; 72°C, 5 min. Une purification des produits de PCR est ensuite menée avec les billes Ampure en appliquant un ratio de 1/1. Une seconde purification est réalisée avec des billes magnétiques recouvertes de streptavidine (Dynabeads, ThermoFisher Scientific). Ces billes lient spécifiquement les produits de PCR1 qui sont biotinylés par la présence du primer pBT20-PCR1 ou Tn5-PCR1. Le protocole de liaison billes–amplicon est réalisé selon les indications du fabricant. Les billes liées aux amplicons de la PCR1 sont directement introduites dans la PCR2. La seconde PCR (PCR2) permet l’ajout des adaptateurs Illumina portés par les primers de PCR. Elle est menée dans un volume final de 50µL, en introduisant les billes liées aux amplicons de la PCR1, 2,5µL de dNTPs (10 mM), 1µL de primer pBT20_PCR2 (30 µM) en fonction de la souche utilisée, 1µL de primer TdT_index_X (30 µM), 0,75µL d’ADN polymérase Q5 et 10µL de 5x *reaction buffer*. 98°C, 5min ; 25 x (98°C, 30 sec.; 55°C, 30 sec.; 72°C, 30 sec.) ; 72°C, 5 min. La purification des produits de PCR est réalisée avec le système Ampure en appliquant un ratio de 1/1. Afin de vérifier la qualité des banques de Tn-seq devenues des « banques de site d’insertion des transposons » (en anglais, *transposon insertion libraries*) à l’issue de ces deux PCR, celles-ci ont été analysées par électrophorèse sur puce et quantification fluorimétrique (Bioanalyseur, Agilent).

Les banques obtenues ont été séquencées sur un séquenceur Illumina HiSeq 4000 en single-end 50 bp en utilisant l'amorce de séquençage customisée Read1TnLp fourni au prestataire de séquençage.

Traitement et Analyse des données de Tn-seq.

Pour chaque condition, entre 5 et 6,5 millions de reads ont été obtenus et filtrés avec les outils du serveur public Galaxy (Goecks et al., 2010). Fastx_clipper a été utilisé pour trimer les queues poly(C) et supprimer les reads trop courts [<15 pb après la coupure poly(C)]. Ensuite, les reads ont été filtrés sur la base de leur qualité en utilisant Trimmomatic et leur qualité a été vérifiée avec FastQC. Ces reads traités ont été cartographiés sur le génome de référence à l'aide du logiciel dédié « Tn-seq TPP » (Tn-seq preprocessor) (DeJesus et al., 2015). Les fichiers wig de sortie de TPP ont été utilisés pour effectuer une analyse d'essentialité à l'aide de Transit (DeJesus et al., 2015). L'analyse d'essentialité conditionnelle a été réalisée avec la méthode de "resampling" selon la documentation du logiciel Transit. La séquence complète du génome de HL07093014 a été obtenue (via le Centre National de Référence des Légionelles) et annotée avec Prokka (Seemann, 2014).



Crible de multiplication intracellulaire

La banque de mutant de HL77S a été cultivée en contexte d'infection en présence de son hôte naturel ambien *Acanthamoeba castellanii* ou d'hôtes accidentels : les macrophages humains U937 ou murin RAW. La capacité des mutants de la banque à infecter ces cellules hôtes a été déterminée comme suit. À J-2 avant l'infection, la banque de mutant de HL77S a été cultivé à 37°C une nuit sur boîte CYE supplémentée avec 50 µg/mL de streptomycine et 10 µg/mL de gentamicine pour obtenir une culture fraîche de la banque de HL77S et limiter les contaminations. À J-1, à partir de ces cultures fraîches l'équivalent de 1.10^7 bactéries par millilitre ont été inoculées dans 3 mL d'AYE aussi supplémentés avec 50 µg/mL de streptomycine et 10 µg/mL de gentamicine pour limiter les contaminations et incubé à 37°C pour la nuit. En parallèle les macrophages (cellules U937 et RAW 264.7 différenciées) ont étéensemencés dans des plaques 6 puits, dans leur milieu de culture respectif (DMEM pour les cellules RAW 264.7 et RPMI 1640 pour les cellules U937) à raison de $1,5.10^6$ cellules/puits (1 mL/puits). Les amibes, elles, ont été remises en suspension dans du milieu PYG à une concentration de 5.10^5 amibes/ml. La suspension a été répartie dans une plaque 6 puits à fond plat (2 ml par puits ; 1.10^6 amibes par puits) et incubée pendant 3 heures à 30°C pour permettre aux amibes d'adhérer aux puits.

Une partie de la culture de nuit en AYE (densité optique [DO] de ~ 5) a été culotée pour former la condition « contrôle » ou « input » du Tn-seq. Cette culture a aussi été diluée dans le milieu de culture cellulaire approprié (DMEM pour les cellules RAW 264.7, RPMI 1640 pour les cellules U937 et milieu PY pour les amibes) et incubée pendant 1 h à 37°C. Les cellules hôte ont été lavées et infectées par ces bactéries à une MOI de 1. Les plaques ont été centrifugées (à $500 \times g$ pendant 5 minutes pour les macrophages et à $650 \times g$ pendant 10 minutes pour les amibes) pour favoriser le contact bactérie-cellule et incubées à 37°C et 5% de CO₂ pendant 48 h pour l'infection des macrophages U937, 72 h pour l'infection des macrophages RAW et à 30°C, 72H pour les amibes. Après incubation, le contenu des puits pour chaque condition a été transféré dans un tube de 1,5 ml et centrifugé à $16\ 000 \times g$ pendant 5

minutes. Les culots contenant les cellules hôtes infectés ont été remis en suspension dans de l'eau distillée stérile pour lyser les cellules et libérer les bactéries.

Pour amplifier la banque après son passage dans les macrophages, ces suspensions de bactéries ont été spotées sur milieu CYE et incubé 24H à 37°C. Après incubation, ce spot a été récolté et resuspendu dans de l'AYE puis culoté pour former la condition « test » ou « output » du Tn-seq. À partir des culots des conditions « input » et « output », les banques de sites d'insertion ont été préparées et séquencées selon les modalités précisées dans ce matériel et méthode. Une analyse d'essentialité conditionnelle a été réalisée en comparant les abondances relatives des insertions dans chacun des gènes entre les conditions « input » et « output ».

B. Annexe



Querying *Legionella* Genomes Using Transposition-Sequencing

Léo Hardy and Xavier Charpentier

Abstract

Transposition-sequencing (Tn-seq) has recently emerged as a powerful technique to query bacterial genomes. Tn-seq can be used to query the bacterial genome with unprecedented resolution, allowing the identification of small genes (e.g., noncoding RNA) that may be missed in conventional screening approaches. Tn-seq can be used to predict genes essential for in vitro growth and to directly identify genetic requirements for survival under multiple conditions. For instance, Tn-seq can be applied to determine the genes, and cellular processes, required to resist an antibacterial treatment or to acquire new resistance genes, to adapt to intracellular life or to compete with other bacteria. Virtually any assay that involves a selection pressure can be used to identify the associated genetic determinants. So far, genome-wide Tn-seq has not been applied to *Legionella* species. Here, we provide a protocol covering all the different steps to conduct a Tn-seq analysis in *L. pneumophila*. This includes generating a high-density library of insertional mutants, setting up a selection screen, sequencing the libraries, mapping the insertion sites, and analyzing the data to obtain the list of genes involved in surviving the applied selection.

Key words Transposition, Sequencing, Gene, Genome, Gene function, *Legionella pneumophila*

1 Introduction

Pathogens can adopt specific strategies to colonize and infect a host or to adapt and resist therapeutic and prophylactic interventions. Deciphering these strategies requires determining their genetic basis. For a given pathogen, this is typically done by analyzing the phenotype of a given gene mutant in a model of infection/adaptation/resistance. Alternatively, random mutagenesis can generate libraries of thousands of mutants that can be individually tested in a phenotyping model. This can lead to the identification of the genes required for a specific process (i.e., virulence, resistance). Yet, these strategies are tedious, and a limited number of infection models and/or conditions can be tested for a given pathogen. Transposition-sequencing (Tn-seq, HITS, Tradis, or Inseq) has recently emerged as a way to drastically increase the throughput

of such approaches [1–4]. Transposition-sequencing (Tn-seq) combines transposition mutagenesis and deep-sequencing mapping and allows the identification of mutants from an insertional library that have lost a given function (negative selection) without the need to test individual mutants. By monitoring a large library of single transposon insertion mutants with high-throughput sequencing, this method can rapidly identify, with high resolution, genomic regions that contribute to organismal fitness under any condition that can be assayed in the laboratory (Fig. 1) [5]. For instance, a recent study defined a “fine scale phenotype-genotype virulence map” of the human pathogen *Streptococcus pneumoniae* by screening 17 in vitro and 2 in vivo (carriage and infection) conditions [6]. Moreover, Tn-seq can be used to query the bacterial genome with unprecedented resolution, allowing the identification of small genes (e.g., noncoding RNA genes) that may be missed in conventional screening approaches [7]. Once a Tn-seq library has been established for a pathogen, it can be screened in multiple assays with near endless possibilities, from predicting genes essential for in vitro growth to directly assaying

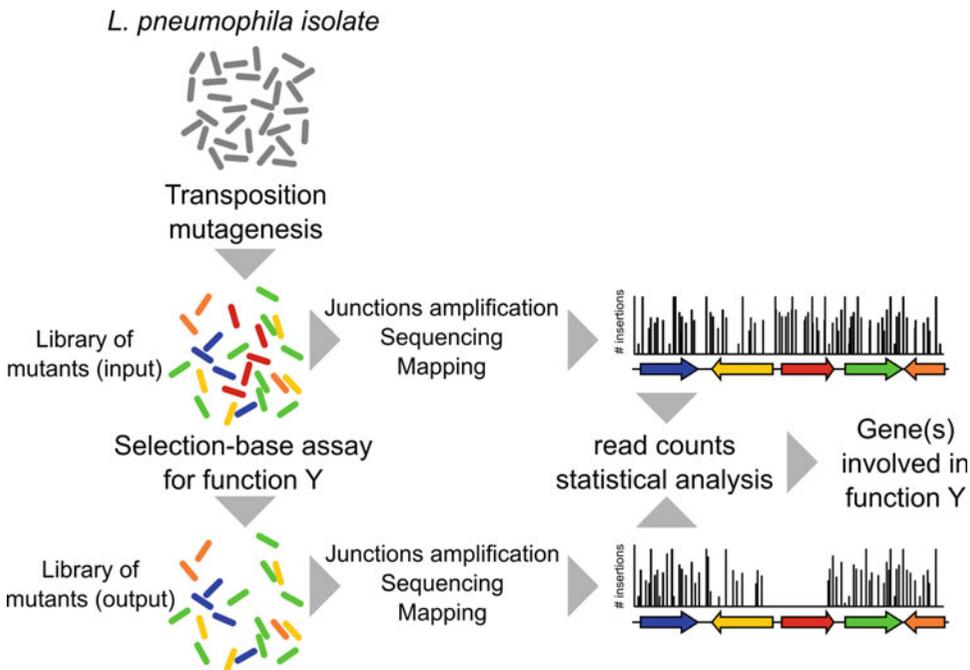


Fig. 1 Diagram of a Tn-seq analysis to identify genes required for the function Y. The *L. pneumophila* strain must first be subjected to random insertional mutagenesis in order to obtain tens or hundreds of thousands of insertional mutants. The library is then subjected to an assay that will counterselect mutants that have lost the ability to perform function Y. Both the input and output libraries will be treated to amplify the junctions of the transposon used to generate the library of mutants. Sequencing and mapping these regions on a reference genome will show how many insertions in a given gene are present/absent in the input and output libraries. Genes in which insertions are underrepresented in the output are involved in function Y

requirements for survival under infective conditions in vivo. Tn-seq can be applied to determine the genes, and cellular processes, required to resist an antibacterial treatment or to acquire new resistance genes, to adapt to intracellular life or to compete with other bacteria. Virtually any assay that applies a selection pressure can be used to identify the genetic determinants involved in a selection process.

In *L. pneumophila*, before the advent of massively paralleled sequencing, saturating mutagenesis and transposon site hybridization (TraSH) provided the first identification of the genes required for growth in nutrient-rich medium [8]. More recently, a transposition-sequencing approach, but limited to a sub-library of insertion mutants in Type IV-secretion effector genes, revealed how multiple effectors contribute to the virulence of *L. pneumophila* [9]. This study demonstrated the power of Tn-seq to identify virulence determinants, but the approach was targeted to only ~10% of the *L. pneumophila* genes. Hence, the full power of the Tn-seq technique has not yet been exploited to query the *Legionella* genomes.

We here provide a protocol to conduct a genome-wide Tn-seq analysis in a *Legionella pneumophila* isolate. First, we describe the procedure to generate a collection of insertion mutants that will be suitable for a Tn-seq analysis. Several hundred thousand mutants can be obtained with this protocol. The transposition mutant library can then be used in any selection-based screen. Second, we provide a detailed protocol to identify the transposition insertion sites with the terminal deoxynucleotidyl transferase (TdT) method. Finally, we present an example of data analysis to identify the genes required for the function tested in the selection-based screen.

2 Materials

2.1 Random Insertion by Mutagenesis

1. The 2,3-Diaminopropionic acid (DAP) auxotrophic *Escherichia coli* MFD_{pir} donor strain with chromosome-integrated RP4 conjugative system [10].
2. Pir-dependent, mobilizable plasmid pBT20 carrying the mariner himar1 transposon with gentamicin-resistant gene [11].
3. *L. pneumophila* strain of interest (*see Note 1*).
4. AYE medium. For 1 liter dissolve 0.30 g iron (III) polyphosphate hydrate, 0.50 g L-cysteine, 12 g yeast extract (BD Bacto Yeast Extract, #212750) (*see Note 2*), and 10 g ACES; adjust pH to 6.9 with 1 M KOH. Fill volume to 1 L with distilled water and filter sterilize.
5. CYE plates. For 1 liter dissolve 10 g yeast extract granulated (Merck Millipore, #1.03753) (*see Note 2*) and 10 g ACES, adjust pH to 6.9 with 1 M KOH, add 15 g of agar and 2 g of

activated charcoal, and autoclave. Add 10 mL of filter-sterilized cysteine 40 g/L and 10 mL of filter-sterilized ferric nitrate 25 g/L. When appropriate, supplement medium with 10 µg/mL gentamicin.

6. CYE + DAP plates (CYED) for conjugation. Same as CYE plates but with addition of DAP and without addition of cysteine and ferric nitrate (*see Note 3*). For 1 liter, add 10 mL of DAP at 10 g/L (1% stock solution). A 1% (10 g/L) stock solution of DAP is prepared by mixing 1 g of DAP in 100 mL of water. Add a few drops of 1 M HCl until DAP fully dissolves.
7. Lysogeny broth (LB). When appropriate, supplement medium with 100 µg/mL ampicillin.
8. Phosphate-buffered saline solution (PBS).
9. Glycerol.
10. Screw-cap cryotubes.

2.2 *Tn-seq TdT* Method

1. Wizard Genomic DNA Purification Kit (Promega).
2. 10 mM Tris-HCl pH 8.0.
3. Sonicator.
4. Agarose gel electrophoresis equipment.
5. Agencourt Ampure XL Beads (Beckman Coulter).
6. Recombinant terminal deoxynucleotidyl transferase (rTdT) 30 U/µL (Promega).
7. 5× rTdT buffer (Promega).
8. dCTP/ddCTP mix with 9.5 mM deoxy-CTP (dCTP) and 0.5 mM di-deoxy-dCTP (ddCTP).
9. Thermocycler.
10. Q5 High-Fidelity DNA Polymerase (New England BioLabs).
11. 10 mM dNTP.
12. Primer olj376 (30 µM): GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGGGGGGGGGGGG GG.
13. Primer pBT20-PCR1 (30 µM): Biotin-TCGTATAATGTGT GGAATTGTGAGCGG.
14. Primer pBT20-PCR2 (30 µM): AATGATACGGCGACCACC GAGATCTACACTCTTTGGACTCTAGAGGATCACCCAG CTTTCTTG.
15. HPLC-purified primer TdT_index_X (*see Note 4* for index list): CAAGCAGAAGACGGCAT ACGAGATXXXXXXGT-GACTGGAGTTCA GACGTGTGCTCTTCCGATCT.
16. Dynabeads M-280 Streptavidin (Invitrogen, 11205D).
17. Magnet for separation.

18. 2× binding and wash (B&W) buffer (2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5).
19. Low TE (LoTE) buffer (3 mM Tris-HCl, 0.2 mM EDTA, pH 7.5).
20. Primer Read1TnLp (Read 1 custom sequencing primer, 100 μM, HPLC-purified): CTAGAGACCGGGGACTTATCAGCCAACCTGTTA.

2.3 Bioinformatic Analysis

1. Genome file (FASTA format) and corresponding annotation file (GFF3 format) of the tested strain.
2. Galaxy server. Either the public server (<https://usegalaxy.org/>) or one of the mirror servers (<https://galaxyproject.org/public-galaxy-servers/>).
3. TRANSIT—a software tool for Himar1 TnSeq analysis [12]. TRANSIT is available at <http://saclab.tamu.edu/essentiality/transit/> and has extensive documentation <http://transit.readthedocs.io/en/latest/>.

3 Methods

3.1 Generation of a Library of Insertion Mutants Using Transposition-Based Random Mutagenesis

Transposon mutagenesis is a powerful method to perform random gene disruptions in bacterial genomes. For that purpose, we use a Himar1-based transposon suicide vector pBT20 [11] carried by the auxotrophic strain *E. coli* MFD*pir* [10]. This bacterium has a RP4 conjugative system that allows the transfer of the mutagenesis system to the *L. pneumophila* recipient strain to be mutagenized. There are other techniques that allow such transfers by transformation or electroporation, but we found these techniques to be less effective than conjugation for *L. pneumophila* mutagenesis. Once in the recipient strain, the Himar1 transposon inserts randomly and specifically at TA dinucleotide sites by nonconservative transposition catalyzed by the Himar1 C9 transposase encoded by pBT20. In addition to the gentamicin resistance gene, this transposon carries an outward-facing promoter that limits polar effect when it is inserted in genes organized in operon.

1. Streak *L. pneumophila* recipient strain of interest (*see Note 1*) from a frozen stock on a CYE plate, and incubate it for 72 h at 37 °C.
2. With a sterile loop, scrape off enough bacteria from the plate to inoculate 15 mL of AYE medium at an OD₆₀₀ ~0.1.
3. In parallel, grow *E. coli* MFD*pir* donor strain in 7.5 mL LB broth containing 100 μg/mL ampicillin and 1% DAP.
4. Incubate both cultures overnight at 37 °C with agitation to stationary phase (OD₆₀₀ = 4–5 for *L. pneumophila*).

5. Centrifuge the cultures at $5000 \times g$ for 10 min. Discard supernatant. Resuspend the *L. pneumophila* cell pellet in 1.5 mL sterile water and the *E. coli* cell pellet in 750 μ L sterile PBS.
6. Mix both concentrated pellets together by pipetting.
7. Spot 100 μ L of this mixture on nonselective CYED plates (i.e., CYE plates without iron and cysteine but with DAP) without spreading. Up to 20 spots can be arranged on 1 plate. When the liquid is completely absorbed, incubate plates 5–6 h at 37 °C. During this incubation time, both the conjugation mediated by the RP4 conjugative plasmid and the following mutagenesis by transposition will take place.
8. Resuspend each spot in 1 mL sterile water and then combine all suspensions. Spread aliquots of the suspension on CYE plates (500 μ L per plate) containing 10 μ g/mL gentamicin to obtain the transconjugants (i.e., the mutants). Make a serial dilution using 100 μ L of the remaining suspension to approximate the number of mutants obtained. Incubate plates 72 h at 37 °C.
9. Sum the total number of mutants obtained on all plates. This number varies greatly depending on the tested strain (*see Note 1*). Permissive strains may yield from 250,000 to 500,000 mutants. A library containing fewer than 200,000 mutants will not be suitable for a Tn-seq analysis.
10. Scrape off all the mutants from all the plates, and resuspend them in 25 mL AYE medium. Add 25 mL glycerol 30% and mix gently. Aliquot 500 μ L of the suspension in cryotubes and store at -80 °C.
11. The following day, thaw a frozen aliquot, and plate serial dilutions to determine the titer of the library in CFU/mL. Typically, 250,000–500,000 mutants resuspended in 50 mL of AYE+glycerol yield a titer of 10^{10} CFU/mL.
12. The mutant library is now ready to be used in any Tn-seq screen.

3.2 Performing a Selection-Based Tn-seq Screen

1. Design the Tn-seq screen and the conditions under which the library will be grown. (*see Note 5* for important considerations when designing a Tn-seq screen.)
2. Determine the starting inoculum to obtain at least nine generations between the starting and end points of the assay. For instance, if the selection involves growth in AYE, keep in mind that a culture in stationary phase reaches $5 \cdot 10^9$ CFU/mL. So, a starting inoculum of a 10 mL culture would thus be $5 \cdot 10^9 \times 10/2^9 = 1 \cdot 10^8$ CFU.

3. Thaw completely one cryotube containing the mutant library (input). Given the above calculation and the titer of the library, only a few microliters of the library will be used.
4. Inoculate the culture with the determined CFU, and incubate in the appropriate conditions until you have reached the end point of the assay.
5. Pellet the cells that are remaining in the cryotube (should be over 10^9 cells). This is the “input pellet.” Store the pellet at $-20\text{ }^{\circ}\text{C}$ to extract genomic DNA at a later time.
6. Once the culture has reached the end point, recover the surviving mutants (i.e., the output), and centrifuge at maximum speed to pellet them. This is the “output pellet.” Store it at $-20\text{ }^{\circ}\text{C}$ to extract genomic DNA at a later time.

3.3 Transposon Junction Amplification and Sequencing by the TdT Method

Different methods have been developed to amplify and sequence the transposon junctions. A common method involves the introduction of a MmeI restriction site in the inverted repeats of the transposon. This is only possible for Himar1 transposon, and most Himar1-based transposons natively lack this feature. Here we use an alternate method called “TdT method” (Fig. 2) [1]. This method involves fragmentation of the genomic DNA and the subsequent addition of a homopolymeric cytosine-tail (C-tail) to the 3' end of all fragments, a reaction catalyzed by the terminal deoxynucleotidyl transferase (TdT). This 3' C-tail is then used as a universal template for a poly-guanine (poly-G) reverse primer. Together with a forward biotinylated transposon-specific primer, a first round PCR is performed resulting in 3' transposon-genomic DNA junction amplification. Affinity of streptavidin to biotin is then used to capture the biotinylated PCR products. A second round of PCR is then performed to enrich the products and add Illumina sequencing adapters and indexes for multiplexing. Finally, sequencing is performed on an Illumina platform using a custom sequencing primer. The following protocol is adapted from [1, 13]:

1. Carry out DNA genomic extraction on the “input pellet” and “output pellet” using Wizard Genomic DNA Purification Kit or any other kit of your choice. Make sure to process enough material to get 50–100 μg of genomic DNA.
2. Mechanically shear 150 μL genomic DNA at 200 $\text{ng}/\mu\text{L}$ (30 μg) by sonication. We use a Branson sonifier with the following parameters: 4 min (1 s ON, 11 sec OFF duty cycle, 20% intensity).
3. Perform agarose gel electrophoresis (2%) using 5 μL of the sample to observe the sheared DNA smear. For whatever shearing method used, a smear of 100–1000 bp size range with maximum intensity between 300 and 600 bp should be obtained. If another shearing method or equipment is used, a

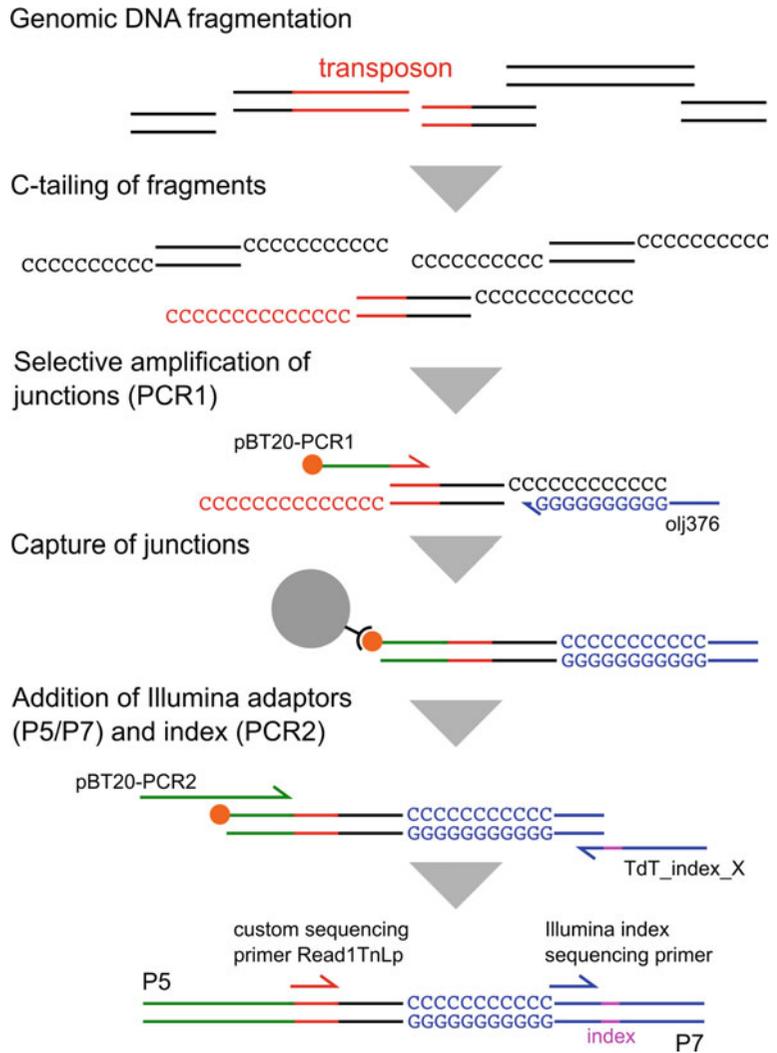


Fig. 2 Schematic representation of the TdT method to amplify the transposon junctions. Following fragmentation, the DNA extracted from the input and output libraries of mutants is treated with the TdT enzyme to add a poly-C-tail to DNA fragments. The transposon junctions are amplified by PCR using a 5'-biotinylated transposon-specific primer (pBT20-PCR1) and the poly-G primer olj376. These fragments are captured and purified by affinity with streptavidin-coupled beads. A final PCR (PCR2) is used to add the Illumina adaptors and the index for multiplexing. Sequencing uses the custom sequencing primer Read1TnLp

preliminary study should be conducted to determine the required parameters.

4. Perform a size selection using Agencourt Ampure XL magnetic beads to remove short fragments. This is done by following steps 5–11.

5. To a 130 μL sample of sheared DNA, add 0.6 \times volume (78 μL) of Ampure XL bead suspension (vortex and mix well before use). Incubate 5 min at room temperature.
6. Place tube on a magnetic stand to collect the beads. Remove and discard the supernatant.
7. With the tube still on the magnetic stand, wash the beads by addition of 200 μL of 80% ethanol. Incubate 30 s and remove the supernatant.
8. Repeat **step 9**.
9. Let the pellet of beads dry for 10 min at room temperature.
10. Remove tube from the magnetic stand, and elute by resuspending the beads in 50 μL of 10 mM Tris-HCl pH 8.0. Incubate for 5 min at room temperature.
11. Place the tube back on the magnetic stand and collect the supernatant. Quantify DNA by absorbance. A total of about 10 μg of DNA should remain in the tube.
12. Prepare the TdT C-tailing reaction in a final volume of 50 μL by mixing 3 μg of size-selected and sheared DNA, 2.5 μL dCTP/ddCTP mix, 10 μL TdT buffer 5 \times , and 1.25 μL TdT, and complete with ultrapure water.
13. Incubate the mixture 1 h at 37 $^{\circ}\text{C}$ followed by an inactivation step of 20 min at 75 $^{\circ}\text{C}$.
14. Purify the C-tailed DNA mixture by adding 1 \times (50 μL) of Ampure XL beads. Incubate for 5 min at room temperature.
15. Place tube on a magnetic stand to collect the beads. Remove and discard the supernatant.
16. Wash by addition of 200 μL of 80% ethanol. Incubate for 30 s and remove the supernatant.
17. Repeat **step 18**.
18. Remove tube from the magnetic stand, and elute by resuspending the beads in 50 μL of 10 mM Tris-HCl pH 8.0. Incubate for 5 min at room temperature.
19. Place the tube back on the magnetic stand and collect the supernatant. Quantify DNA by absorbance. At least 1 μg of DNA should remain in the tube.
20. Prepare the first PCR reaction (PCR1) in a final volume of 50 μL by mixing 500 ng C-tailed DNA, 1 μL biotinylated pBT20-PCR1 primer (30 μM), 3 μL olj376 primer (30 μM), 2.5 μL dNTPs (10 mM), 10 μL 5 \times Q5 reaction buffer, and 0.75 μL Q5 High-Fidelity DNA Polymerase, and complete with ultrapure water.
21. Perform PCR amplification on a thermocycler following those parameters: 98 $^{\circ}\text{C}$, 5 min.; 25 \times (98 $^{\circ}\text{C}$, 30 s; 65 $^{\circ}\text{C}$, 30 s; 72 $^{\circ}\text{C}$, 30 s); 72 $^{\circ}\text{C}$, 5 min. Analyzing the PCR by agarose gel

electrophoresis is not necessary as no detectable PCR products are usually observed at this stage.

22. Purify the PCR reaction with $1 \times$ (50 μL) of Ampure XL beads (*see steps 14–19*).
23. Prepare Dynabeads for the capture of biotinylated PCR1 products: resuspend Dynabeads by vortexing. Add 32 μL Dynabeads to a 1.5 mL Eppendorf tube. Place the Eppendorf tube on a magnetic stand, and wash the beads three times with 1 mL $1 \times$ B&W buffer. Remove the final wash, and resuspend the beads in 52 μL $2 \times$ B&W.
24. Add 50 μL of purified PCR1 product to the resuspended beads, and put the tube on a rotator at room temperature for 30 min to promote streptavidin-biotin interactions.
25. Place the tube on a magnetic stand, and wash the beads once with 100 μL $1 \times$ B&W and twice with 1 mL of LoTE. Remove all supernatant.
26. Prepare the second PCR reaction (PCR2) in a final volume of 50 μL by mixing 1 μL pBT20-PCR2 primer (30 μM), 1 μL TdT_index_X primer (30 μM), 2.5 μL dNTPs (10 mM), 10 μL $5 \times$ Q5 reaction buffer, and 0.75 μL Q5 High-Fidelity DNA Polymerase, and complete with ultrapure water.
27. Resuspend the beads (which have PCR1 products bound to them) in the PCR2 reaction mix.
28. Perform the PCR amplification on a thermocycler following those parameters: 98 $^{\circ}\text{C}$, 5 min.; $25 \times$ (98 $^{\circ}\text{C}$, 30 s; 55 $^{\circ}\text{C}$, 30 s; 72 $^{\circ}\text{C}$, 30 s); 72 $^{\circ}\text{C}$, 5 min.
29. Perform agarose gel electrophoresis (2%) with 5 μL of the PCR2 reaction to observe the amplified PCR2 products. You must get a smear of 150–250 bp size range.
30. Transfer the PCR reaction to a 1.5 mL Eppendorf tube, and place it on a magnetic stand to collect the beads. Transfer the supernatant to a fresh tube. Discard the beads.
31. Purify the PCR reaction with $1 \times$ (50 μL) of Ampure XL beads (**steps 14–19**).
32. Perform agarose gel electrophoresis (2%) with 5 μL purified PCR2 to make sure that **step 31** effectively removed the primers and the smear remains. Alternatively, the sample may be analyzed by capillary electrophoresis (Agilent Bioanalyzer, Thermo Fragment Analyzer).
33. Your sample is now ready to be sequenced. We sequence our samples on an Illumina HiSeq 4000 using the custom sequencing primer ReadITnLp which must be provided to the sequencing facility. *See Note 6* on specific instructions that must be given to the sequencing facility in order to set up the sequencing run.

3.4 Data Analysis

The sequencing facility should return several million reads per sample. These reads will be used to determine the insertion sites of the transposon and the abundance of each mutant with a given insertion. The genes for which insertions are underrepresented under the tested conditions (output) relative to the control (input) correspond to the genes required for the function tested by the screen. In essence, the analysis may be similar to the analysis of RNA-seq datasets. Yet, several methods have been developed to perform such analysis, and a recent review provides a nice overview of most of them [14]. Invariably the different methods begin with preparing the reads and mapping them against the genome of the tested strain. Then, the reads corresponding to defined features of the genome (CDS, sRNA) will be counted. A statistical analysis is then performed to determine nonrandom difference in read counts. We here describe a data analysis that uses the Galaxy Project to check the quality of the reads and preprocess them for analysis. Reads will then be processed using the TPP tool [15] and used as input for the TRANSIT software that was designed to perform analysis of HimarI-based Tn-seq data [12]. TRANSIT is a cross-platform software with a convenient graphical interface. In addition to performing comparative analysis between conditions (output vs. input), TRANSIT also uses statistical methods to identify essential genes based on the frequency of insertion in a given gene. This latter part will not be covered here. As an example, we describe the analysis for a screen with a single input and the corresponding output. Obviously, similarly to an RNA-seq analysis, a typical Tn-seq analysis should include replicates. In order to make Tn-seq available to most users, we provide here a step-by-step analysis conducted through graphic user interfaces. Experienced users may access the same tools using command-line programs and scripts.

1. Make sure your Tn-seq data is demultiplexed and in .fastq format. You should have one fastq file for each of the tested condition (for this example, one input.fastq, one output.fastq).
2. Create a galaxy account on the Galaxy Project's public server usegalaxy.org (User tab > Register), and login with your ID.
3. Go to "Analyse Data" tab, and load your Tn-seq datasets with the "Get Data" tool present in the "Tools" menu (right drop-down list). With this tool you can load data from a local file, from UCSC, or from other servers. Your data will appear in your History (left drop-down list).
4. Trim the reads with the `fastx_clipper` following those steps: In the "Tools" menu, search for "clip adapter sequences." Click on "clip adapter sequences," and the tool interface will appear (id: `fastx_clipper`). In the field "Library to clip," choose from your History the library you want to clip. If your .fastq file does

not appear, edit the attributes of your file to change the datatypes into `.fastqsanger` (in the History >, click on your `.fastq` file >, click on the pencil >, click on “datatypes” tab, and choose “.fastqsanger”). Set the “Minimum sequence length” to 15. This will cause the `fastx_clipper` tool to output only reads (clipped or unclipped) that are longer than 15 bp (i.e., reads that can be mapped specifically of the reference genome). Under “Source,” choose “Enter custom sequence.” Then enter `CCCCCCCCCCC` in the field “Enter custom clipping sequence.” This will indicate the program to cut out from the reads the poly-C and every base that follow it. In the two next fields, leave default value “0” (to discard the adapter sequence and x bases that follow it) and “yes” (to discard sequences with unknown (N) bases). In the “Output options,” choose “Output both clipped and non-clipped sequences.” Click on “execute” to run the tool. In your History, the job will be added to the queue. Once completed, you will be able to access the trimmed library and use this new file for the next steps.

5. Filter and trim reads by quality with Trimmomatic. In the left “Tools” menu, search for “Trimmomatic” tool. Configure the tool by selecting “Single-end,” and select the clipped data file obtained in **step 4**. For the other options, leave the default settings. Click on “execute” to run the tool and to add it to the queue in your history.
6. Control the quality of your trimmed library using FastQC Read Quality reports. In the left “Tools” menu, search for “FastQC” tool. Click on and the tool interface will appear. Simply choose the output from Trimmomatic (**step 5**), and run the tool with the default parameters.
7. Check the following results to ensure the good quality of your sequencing library after trimming. In “Basic Statistics,” check for the number of total sequences left after trimming. It is expected to lose about 30% of reads after trimming. In the “per base sequence quality” graph, make sure the quality score across all bases is over 30. In “per base sequence content” graph, make sure cytosine is not overrepresented.
8. If the quality of your trimmed library is good, download the output file from Trimmomatic.
9. Mapping the reads to the reference genome. Install the python-based Tn-seq software TPP (Tn-seq preprocessor) and TRANSIT according to the developer’s instructions (<http://transit.readthedocs.io/en/latest/>). Download and install the mapping software BWA (Burrows-Wheeler Aligner) from <http://bio-bwa.sourceforge.net/>. Configure TPP with the following parameters: specify the path to where the BWA executable is installed; specify the path to where the reference genome

in .fasta format is located; in “choose the fastq file for read1,” specify the path to your trimmed library (i.e., the galaxy output), and leave blank for Read 2; choose a prefix to use for the output file name; specify the Tn-seq protocol used to prepare sequencing library. Here, choose “Tn5” (i.e., the TdT method); in “transposon used,” choose “[custom],” and leave blank for the primer because in our protocol, the custom sequencing primer will directly read the first base of the insertion and not the end of the transposon. For other parameters, leave blank if you want neither to allow mismatch for the reads mapping nor to map only a defined number of reads. Run TPP. During the run some output files are created. Two files are particularly interesting: [1] the .tn_stats file which contains all mapping statistics such as the number of reads mapped and the number of unique hits and [2] the .wig file which will be used for the TRANSIT analysis.

10. Start TRANSIT, and set up the analysis by providing the path to the .gff3 annotation file (*see Note 7*).
11. Add the control and experimental dataset (and/or different replicates) in .wig format (i.e., the output from TPP software).
12. Run TRANSIT. The output of TRANSIT includes a .dat file that will contain all necessary informations (\log_2 FC, P-value) to identify the genes required for the tested function (*see Note 8*).
13. The insertions and their abundance can also be visually inspected either using the genome viewer embedded in TRANSIT or using other genome browsers, such as IGV and IGB.

4 Notes

1. Not all *L. pneumophila* strains may be suitable for a Tn-seq analysis. This is mostly due to the ability to obtain large number of transposition mutants. For instance, in our hands, delivery of the transposon by electroporation of a plasmid-encoded Himar1 transposon in the lab strain JR32 resulted in 1000–10,000 mutants. Delivery of a Himar1 transposon by conjugation proved much more effective, but we observed large differences in the number of transformants depending on the tested *L. pneumophila* isolate. Lab strains (Paris, JR32) performed poorly, but some clinical isolates were much susceptible to transposition mutagenesis, and with several isolates, we obtained over 100,000 mutants in a single mating experiment. There was no connection between the origin of the isolates, their phylogenetic relationships, and the number of mutants obtained.

2. Different sources of yeast extract were found to differentially support *L. pneumophila* growth in liquid (AYE) or solid (CYE) media. For instance, using BD Bacto Yeast Extract (#212750) in AYE, the culture can reach an optical density (600 nm) of 5–6, while it may only reach 3–4 with granulated yeast extract (Merck Millipore, #1.03753). Conversely, granulated yeast extract in CYE yields larger colonies than BD Bacto Yeast Extract.
3. Regular CYE should not be used for mating with the DAP auxotroph MFD pir strain. Regular CYE contains a large amount of L-cysteine that oxidizes to L-cystine. DAP and L-cystine are structurally similar and imported by the same transporter system. However, because L-cystine has a stronger affinity to the transporter than DAP [16], it will prevent import of DAP and affect viability and growth of MFD pir . Thus, L-cysteine is omitted from the plates. Iron is also omitted.
4. We recommend the TruSeq index system. Indexing primers 1–12 are obtained by replacing the XXXXX sequence of the primer by each of the following sequences: CGTGAT, ACATCG, TTAGGC, GCCTAA, TGGTCA, CACTGT, ATTGGC, GATCTG, TCAAGT, CTGATC, AAGCTA, GTAGCC, and TACAAG.
5. A good screen typically involves a strong selection. If the screen involves growth, a minimum of nine generations should be obtained between the starting and end points. Fewer generations will result in a small fold change in the analysis. Keep in mind that your assay may represent a bottleneck. If a strong selection is applied to an inoculum that has fewer individual mutants than the complexity of the library, some mutants may be lost by chance rather than by selection. We recommend that at any stage of the screen, the population should contain at least ten individuals of each mutant (e.g., for a library of 300,000 insertions, the starting culture should contain at least $3 \cdot 10^6$ cells).
6. Request the sequencing facility to perform a SR50 run (single end, 50 bp). Sequencing more than 50 bp is not necessary and may even be detrimental. Indeed, because of C-tailing reactions, many reads will show a 15–25 long stretch of C. This will introduce an imbalance in the relative abundance of each of the four nucleotides in the reads. This is known to affect the general quality of the reads. To limit this effect, request the facility to add at least 10% PhiX DNA to the samples. Provide the facility with the Read 1 custom sequencing primer, Read1TnLp, at 100 μ M. This primer should be added to the Illumina Read 1 sequencing primer (which will read the PhiX sequence). If multiplexing multiple samples, make sure to

obtain at least ten million reads per sample. The first base read with the Read1TnLp primer corresponds to the insertion site.

7. If the tested strain is not referenced, we recommend annotating the genome with a tool such as Rast (available at <http://rast.nmpdr.org/>) or Prokka (available at <http://www.vicbioinformatics.com/software/prokka.shtml> or in <https://usegalaxy.org/>). Both tools create a .gff3 file that is properly interpreted by TRANSIT.
8. Please refer to the TRANSIT documentation <http://transit.readthedocs.io/en/latest/>.

Acknowledgments

We thank Jean-Marc Ghigo for providing the strain MFD_{pir}. Clinical isolates tested for the construction of *L. pneumophila* mutant libraries have been collected from the French Legionella Reference Center (Lyon, France). This work was supported by the LABEX ECOFECT (ANR-11-LABX-0048) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

References

1. van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6:767–772
2. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 19:2308–2316
3. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6:279–289
4. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci* 106:16422–16427
5. van Opijnen T, Camilli A (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* 11:435–442
6. van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* 22:2541–2551
7. Barquist L, Boinett CJ, Cain AK (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol* 10:1161–1169
8. O’Connor TJ, Adepoju Y, Boyd D, Isberg RR (2011) Minimization of the Legionella pneumophila genome reveals chromosomal regions involved in host range expansion. *Proc Natl Acad Sci U S A* 108:14733–14740
9. Shames SR, Liu L, Havey JC, Schofield WB, Goodman AL, Roy CR (2017) Multiple Legionella pneumophila effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc Natl Acad Sci U S A* 114:E10446–E10454
10. Ferrières L, Hémery G, Nham T, Guéroult A-M, Mazel D, Beloin C, Ghigo J-M (2010) Silent mischief: Bacteriophage Mu insertions contaminate products of escherichia coli random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J Bacteriol* 192:6418–6427
11. Kulasekara HD, Ventre I, Kulasekara BR, Lazdunski A, Filloux A, Lory S (2005) A

- novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol Microbiol* 55:368–380
12. DeJesus MA, Ambadipudi C, Baker R, Sasseti C, Ioerger TR (2015) TRANSIT-a software tool for Himar1 TnSeq analysis. *PLoS Comput Biol* 11:e1004401
 13. Stacy A, Fleming D, Lamont RJ, Rumbaugh KP, Whiteley M (2016) A commensal bacterium promotes virulence of an opportunistic pathogen via cross-respiration. *MBio* 7: e00782-16
 14. Burby PE, Nye TM, Schroeder JW, Simmons LA (2017) Implementation and data analysis of Tn-seq, whole-genome resequencing, and single-molecule real-time sequencing for bacterial genetics. *J Bacteriol* 199: e00560–e00516
 15. Long JE, DeJesus M, Ward D, Baker RE, Ioerger T, Sasseti CM (2015) Identifying essential genes in *Mycobacterium tuberculosis* by global phenotypic profiling. *Methods Mol Biol* 1279:79–95
 16. Berger EA, Heppel LA (1972) A binding protein involved in the transport of cystine and diaminopimelic acid in *Escherichia coli*. *J Biol Chem* 247:7684–7694

II. Analyse génétique exhaustive des gènes impliqués dans la viabilité et la transformation naturelle de *L. pneumophila*

Dans l'étude suivante qui a fait l'objet d'une publication dans Journal of Bacteriology en 2021 dont je suis le premier auteur, nous avons utilisé la banque de mutant et la méthode de Tn-seq décrite dans le paragraphe précédent pour réaliser une analyse génétique exhaustive des gènes essentiels à la viabilité bactérienne et requis pour le processus de transformation de *L. pneumophila*. Cela nous a permis d'identifier (1) le set de gènes essentiels à la viabilité de la bactérie et requis pour sa fitness, (2) les constituants majeurs de la machinerie d'import de l'ADN et (3) des gènes spécifiques impliqués dans la régulation de la compétence. Des études plus approfondies des gènes codant pour le T4P ont permis de définir le set de piline majeure et mineures impliquées dans la capture de l'ADN exogène.

A. Publication



Transposon Insertion Sequencing in a Clinical Isolate of *Legionella pneumophila* Identifies Essential Genes and Determinants of Natural Transformation

Léo Hardy,^a Pierre-Alexandre Juan,^a Bénédicte Coupat-Goutaland,^a Xavier Charpentier^a

^aCentre International de Recherche en Infectiologie, Team Horizontal gene transfer in bacterial pathogens, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Normale Supérieure de Lyon, Université Lyon, Villeurbanne, France

ABSTRACT *Legionella pneumophila* is a Gram-negative bacterium ubiquitous in freshwater environments which, if inhaled, can cause a severe pneumonia in humans. The emergence of *L. pneumophila* is linked to several traits selected in the environment, the acquisition of some of which involved intra- and interkingdom horizontal gene transfer events. Transposon insertion sequencing (TIS) is a powerful method to identify the genetic basis of selectable traits as well as to identify fitness determinants and essential genes, which are possible antibiotic targets. TIS has not yet been used to its full power in *L. pneumophila*, possibly because of the difficulty of obtaining a high-saturation transposon insertion library. Indeed, we found that isolates of sequence type 1 (ST1), which includes the commonly used laboratory strains, are poorly permissive to saturating mutagenesis by conjugation-mediated transposon delivery. In contrast, we obtained high-saturation libraries in non-ST1 clinical isolates, offering the prospect of using TIS on unaltered *L. pneumophila* strains. Focusing on one of them, we then used TIS to identify essential genes in *L. pneumophila*. We also revealed that TIS could be used to identify genes controlling vertical transmission of mobile genetic elements. We then applied TIS to identify all the genes required for *L. pneumophila* to develop competence and undergo natural transformation, defining the set of major and minor type IV pilins that are engaged in DNA uptake. This work paves the way for the functional exploration of the *L. pneumophila* genome by TIS and the identification of the genetic basis of other life traits of this species.

IMPORTANCE *Legionella pneumophila* is the etiologic agent of a severe form of nosocomial and community-acquired pneumonia in humans. The environmental life traits of *L. pneumophila* are essential to its ability to accidentally infect humans. A comprehensive identification of their genetic basis could be obtained through the use of transposon insertion sequencing. However, this powerful approach had not been fully implemented in *L. pneumophila*. Here, we describe the successful implementation of the transposon-sequencing approach in a clinical isolate of *L. pneumophila*. We identify essential genes, potential drug targets, and genes required for horizontal gene transfer by natural transformation. This work represents an important step toward identifying the genetic basis of the many life traits of this environmental and pathogenic species.

KEYWORDS *Legionella pneumophila*, cell viability, genetic competence, genome analysis, natural transformation systems

Legionella pneumophila is a Gram-negative bacterium, ubiquitous in freshwater environments, where it can be found in planktonic form, in biofilm communities, or associated with amoebic protozoa which constitute its natural hosts (1). *L. pneumophila* can resist predation by amoebae and even establish an intracellular vacuole in which it can multiply while being protected from external environment (2). Man-made water

Citation Hardy L, Juan P-A, Coupat-Goutaland B, Charpentier X. 2021. Transposon insertion sequencing in a clinical isolate of *Legionella pneumophila* identifies essential genes and determinants of natural transformation. J Bacteriol 203:e00548-20. <https://doi.org/10.1128/JB.00548-20>.

Editor Laurie E. Comstock, Brigham and Women's Hospital/Harvard Medical School

Copyright © 2021 Hardy et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Xavier Charpentier, xavier.charpentier@univ-lyon1.fr.

Received 2 October 2020

Accepted 4 November 2020

Accepted manuscript posted online 9 November 2020

Published 11 January 2021

systems have offered a new breeding ground for the development of *L. pneumophila*. Inhalation by humans of aerosols produced by these systems and contaminated by *L. pneumophila* can cause Legionnaires' disease (3). This community-acquired disease, which is most often characterized by a severe pneumonia, occurs when *L. pneumophila* infects alveolar macrophages (4). In both macrophages and its natural amoebic hosts, *L. pneumophila* replicates intracellularly by hijacking the host cellular machinery (5). This requires the *lcm/Dot* type IV system (6, 7), a conjugative system that can secrete more than 300 different effector proteins (8, 9). The genome of the endemic strain Paris provided early evidence of genes encoding proteins like those of eukaryotes (10), and eukaryotic-like proteins encoded by the Philadelphia-1 genome (11) were found to be the effector protein substrates of the *lcm/Dot* system (12). Phylogenetic analyses suggest that these genes would have been acquired by interkingdom horizontal gene transfer (HGT) during coevolution of *Legionella* and its natural host for millions of years (13). Hundreds of genome sequences of *L. pneumophila* clinical isolates have now revealed that recombination events are common in this species (14–16). Thus, intraspecies and interkingdom HGT events are playing a major role in the evolution and adaptation of this species.

The high degree of plasticity of the genomes of *L. pneumophila* is consistent with the fact that it is competent for natural transformation (17). Natural transformation refers to the ability of certain bacteria to capture exogenous DNA and integrate it into their genomes by homologous recombination (18). It is one of the driving forces for bacterial evolution that can lead to the emergence of new pathogenic bacteria and new antibiotic-resistant recombinants. It is a widespread mechanism of HGT in bacteria, with more than 80 experimentally confirmed transformable species (19). The DNA uptake mechanisms and associated proteins constituting the so-called DNA uptake machinery are highly conserved (18), suggesting that most species are potentially transformable. DNA uptake first involves a type IV pilus (T4P) (20) whose direct observation supports a model in which it binds DNA via its tip and its retraction allows the internalization of DNA into the periplasm (21). The periplasmic DNA-binding protein ComEA serves as a ratchet (22, 23), and large amounts of DNA can accumulate in the periplasm before being converted into single-stranded DNA (ssDNA) and translocated across the cytoplasmic membrane through the ComEC inner membrane channel (24). In the cytoplasm, the ssDNA is protected by the transformation-dedicated protein DprA (25) and the single-stranded binding protein SsbB (26). If the internalized ssDNA possesses regions homologous to those in the bacterial chromosome, it is integrated by homologous recombination mediated by the recombinase RecA, which interacts with DprA (27). In Gram-negative bacteria, the newly discovered ComM helicase is also involved in this recombination process (28).

In most transformable species, these proteins are not expressed constitutively but only when the bacterium is in a genetically programmed and transient state called competence (29). *L. pneumophila* was first reported to be competent when grown at 37°C under some form of stress, such as under microaerophilic conditions (17) or with exposure to DNA-damaging agents (30). In the absence of any stress, *L. pneumophila* becomes transiently competent when grown at 30°C at the transition between the exponential and stationary growth phases (31, 32). *L. pneumophila* is unique in that the regulation of competence does not involve transcriptional activation of the competence regulon. Rather, the core genes encoding the DNA uptake system (*comEC*, *comEA*, *comFC*, and *comM*) are subjected to posttranscriptional repression by a ribonucleoprotein complex consisting of a small RNA, RocR, and an RNA chaperone, RocC (32). At the onset of the stationary phase, the expression of RocR decreases and the translation of the mRNAs encoding the DNA uptake system allows *L. pneumophila* to take up and recombine extracellular DNA. Most *L. pneumophila* clinical isolates do transform under these conditions, yet some isolates fail to develop competence, and in some instances, this is due to the presence of a mobile genetic element (MGE) that encodes a RocR homolog that acts as a substitute of the chromosome-encoded RocR (33). Competence is further repressed in

stationary phase by the quorum-sensing system (34, 35). The regulation of competence in *L. pneumophila* remains poorly understood (36).

Regulation of competence is best understood in the Gram-positive *Streptococcus pneumoniae*, in which the comprehensive genetic approach of transposon insertion sequencing (TIS) has recapitulated decades of findings (37). Beyond the identification of additional regulatory or functional elements of natural transformation, this approach gave rise to a better understanding of the biology of this bacterium by identifying genes involved in virulence and in resistance against stresses. TIS approaches encompass a number of similar methods (transposon sequencing [Tn-seq], transposon-directed insertion site [TraDIS] sequencing, insertion sequencing [INseq], and high-throughput insertion tracking by deep sequencing [HITS]) (38–41) that have been used for the identification of essential genes on a genome-wide scale in a number of species (42). TIS relies on the mapping and quantification of transposon insertion mutants by high-throughput DNA sequencing, and a critical factor is to obtain high-saturation libraries of transposition mutants (43). TIS was recently applied to *L. pneumophila* with a focus on effector-encoding genes and their conditional involvement in intracellular replication (44, 45). However, the libraries of mutants were either targeted for effectors (44) or of low coverage (45). Thus, the full power of TIS has not yet been harnessed to understand fundamental or specific aspects of the biology of *L. pneumophila*, possibly because of the difficulty of obtaining high-saturation mutant libraries. In addition, the current libraries were constructed in strain Ip02, which has lost competence regulation during its laboratory domestication (46). Here, we sought to obtain a high-coverage library for Tn-seq in *L. pneumophila* that could be used to apprehend the genetic basis of the many life traits of this species. We found that some clinical isolates of *L. pneumophila* are more permissive to transposon mutagenesis than the commonly used laboratory strains. We obtained a high-coverage Tn-seq library in an unaltered clinical isolate and identified genes essential for fitness and growth in axenic medium. We then applied Tn-seq to identify the genes involved in competence and natural transformation.

RESULTS AND DISCUSSION

High-saturation Tn-seq library of *L. pneumophila*. With the objective of obtaining a Tn-seq library of *L. pneumophila*, we tested the conjugative delivery of the Himar1-based transposon carried by the *pir*-dependent mobilizable plasmid pBT20 to the commonly used strain Paris. Conjugation assays with the MFD_{pir} donor strain produced only a few insertional mutants. We hypothesized that the Paris strain was particularly resistant and tested 12 other clinical isolates belonging to sequence type 1 (ST1). Similarly to the Paris strain, none of the ST1 isolates generated a meaningful number of mutants. We concluded that for an unknown reason, the ST1 isolates (which would include the Philadelphia-1 derived laboratory strains Ip02 and JR32) were poorly permissive to conjugative transfer and/or to transposition by Himar1. We thus tested 8 other non-ST1 clinical isolates. We obtained several thousand mutants for 5 of these. We decided to continue with isolate HL-0709-3014, for which we obtained a complete genome composed of a circular chromosome of 3,405 kb and a plasmid of 106 kb (see Materials and Methods). A total of 3,183 open reading frames were detected, 2,791 and 2,741 of which have orthologs in the Paris and Philadelphia-1 strains, respectively (see Data Set S1 in the supplemental material). HL-0709-3014 belongs to the ST18 lineage, which is closely related to the ST1 lineage. Hence, it is phenotypically similar to the Paris strain, it is naturally transformable, and it shows similar intracellular replication rates in amoebae (Fig. S1). It also effectively replicates in human and murine macrophages (>2-log growth in 72 h) (Fig. S1). We isolated HL77S, a spontaneous streptomycin-resistant mutant of HL-0709-3014, and subjected it to mutagenesis with the transposon of pBT20. This mariner-based transposon inserts at TA sites and includes an outward facing *P_{tac}* promoter that can minimize possible polar effects on operon and downstream genes. About 250,000 colonies of mutants were isolated on charcoal-yeast extract (CYE) plates and collected (initial isolation). The library was then cultured

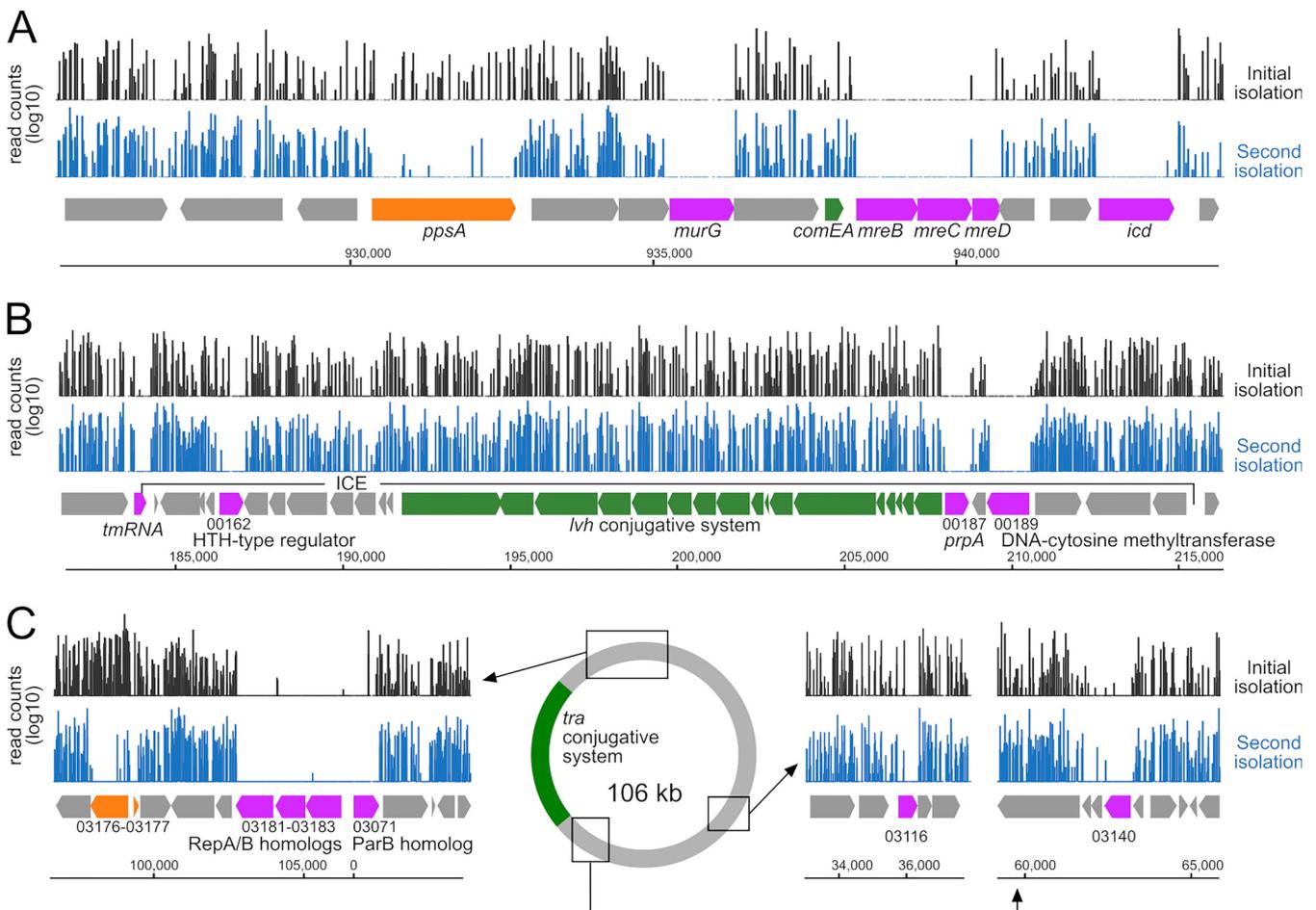


FIG 1 Tn-seq analysis of *L. pneumophila* strain HL77S. (A) Log₁₀ read counts of transposon insertions after initial library isolation (black) and second isolation (blue). Genes identified as essential are in magenta, and fitness determinants are in orange. Other genes of interest are in green. (B) Transposon insertion coverage in a region encompassing an integrative conjugative element (ICE) harboring essential genes (magenta) and genes encoding a conjugative system (green). The duplicated sequence GCGGGTTCGATCCCGCGCTCCACCA of the tmRNA and located 66 kb away delineates the boundaries of the ICE. (C) Essential genes and fitness determinants in the conjugative plasmid of HL77S.

in rich medium at 30°C and reisolated on CYE (second isolation). Sequencing of the transposon insertion sites revealed a maximum of 110,679 unique insertion out of 255,021 possible TA sites (43% saturation) and an average of one insertion site every 31 bp. This represents a significant improvement over the previously published library in the ST1 strain Ip02, which contained 17,781 unique insertion sites (7% saturation) (45). Thus, we obtained a high-saturation Tn-seq library in an *L. pneumophila* clinical isolate that can be used as a surrogate for the commonly used laboratory strains (Paris, JR32, Ip02, and AA100).

Analysis of gene essentiality. The high saturation allowed the identification of genes essential for growth. To do so, we used two statistical methods: the Gumbel method (47), a Bayesian model based on longest consecutive sequence of TA sites without insertion in the genes, and the hidden Markov model (HMM) method based on the detection of genes with unusually low read counts (48). Both methods gave similar results with 401 (Gumbel) and 500 (HMM) genes identified as essential, 382 of which were identified as strictly essential by both methods (Data Set S1), consistent with the average number (391) of essential genes identified by TIS in other bacterial species (49). This is lower than the number of essential genes (588) identified by the TraSH method in the Ip02 strain of *L. pneumophila*, but these also included fitness determinants (50). The data confirmed our previous observation of the essentiality of the actin-like protein MreB (51) but also that of MreC and MreD, while intergenic insertions between *mreC* and *mreD* are tolerated (Fig. 1A).

Comparative analysis of the second and initial isolation identified 181 genes which were nonessential at the initial isolation but whose inactivation impaired fitness (\log_2 fold change [\log_2 FC] < -2; $P < 0.05$). These include the gene encoding exoribonuclease R, whose growth defect was previously reported (52), and the RNA chaperone Hfq but also, more surprisingly, the substrates of the Lcm/Dot type IV secretion AnkQ and SdbB (Data Set S1). Presumably because their inactivation lowered the fitness so dramatically, 61 genes not essential after the initial isolation were deemed essential on the second isolation. These include the phosphoenolpyruvate synthase-encoding gene *ppsA* (Fig. 1A) and the genes encoding the sigma factor RpoS, the recombinase RecA, the tyrosine recombinase XerC (which is involved in chromosome dimer resolution), and the transfer-messenger RNA (tmRNA)-binding protein SmpB, which is involved in *trans* translation. Indeed, this is consistent with our previous demonstration that *trans* translation is essential in *L. pneumophila* (53) and that no insertions are observed in the tmRNA-encoding gene (Fig. 1B).

The vast majority of the essential genes that encode proteins have orthologs in the Paris and Philadelphia-1 genome, as expected for genes that encode proteins involved in the fundamental processes of the cell. However, of all the genes found to be essential on either the initial or second isolation, 14 have no orthologs in the Paris and Philadelphia-1 strains. How could strain-specific genes be essential? Three of these genes (HL77S_01135, HL77S_01141, and HL77S_01146) encode the antitoxin component of toxin-antitoxin (TA) modules clustered within a 4-kb segment. HL77S_01068 is located next to a gene encoding a toxin of the type II TA system, suggesting that it also encodes an antitoxin. Others have no known function, such as HL77S_02141, which encodes a protein with a predicted helix-turn-helix (HTH) motif, or HL77S_00079, encoding a protein with a conserved domain of unknown function (DUF3800).

Consistent with their being part of the accessory genome, genomic comparison with other complete genomes of *L. pneumophila* indicates that all of these genes reside in highly variable regions, often in proximity to putative transposase and prophage integrase. However, no genetic structure corresponding to a complete mobile genetic element (MGE) could be detected. In contrast, two essential genes (HL77S_00162 and HL77S_00189) are within a recognizable MGE corresponding to an integrative conjugative element (ICE) inserted at the 3' end of the tmRNA-encoding gene and carrying a conjugative system homologous to the Lvh system (Fig. 1B). HL77S_00162 is an HTH-type regulator and HL77S_00189 is predicted to encode a DNA-cytosine methyltransferase. The ICE shows a third essential gene (HL77S_00187, *prpA*), conserved in the Lvh ICE of the Paris strain and encoding a LexA/Ci-like repressor homolog (Fig. 1B). Another strain-specific essential gene (HL77S_00197) is located just downstream of the ICE, in a unique region that may represent a remnant of another MGE. Another three of the 14 strain-specific essential genes (HL77S_03181, HL77S_03182, HL77S_03183) are part of the 106-kb conjugative plasmid and clustered with another essential gene (HL77S_03071) which has a homolog on the pLPP plasmid of the Paris strain (Fig. 1C). Encoding Rep or Par homologs, these genes are involved in replication/partition of the plasmid, and their inactivation likely resulted in plasmid loss.

Also on this plasmid, two other genes of unknown function also appear essential, one that is unique and has no conserved domain (HL77S_03140) and one with a homolog on pLPP (HL77S_03116, *pipp0094*) containing an N-terminal HTH motif and a C-terminal nucleotidyltransferase (NT) domain also found in DNA polymerase beta (Fig. 1C). Insertions in two divergently oriented genes (HL77S_03176 and HL77S_03177) are also associated with strong fitness defects and are of unknown function (Fig. 1C).

Overall, we found that many essential genes can be found within MGEs. Insertion in genes controlling vertical transmission can result in the loss of the MGE (and thus of transposon insertions, making the corresponding gene seemingly essential). This might be the case for repressor of excision of ICE or genes required for replication/partition of plasmids. Other genes might be required to limit the cost of the MGE to the fitness of their host. This might be the case for the LexA/Ci-like repressor of ICE, as exemplified

by the *Vibrio cholerae* SXT ICE, for which inactivation of the LexA/Ci-like repressor SetR is deleterious to its host (54). Whatever the mechanism, the genes characterized as essential in MGEs ensure their vertical transmission. Thus, our result indicates that in addition to identifying genes required for the fundamental functions of the cell, Tn-seq analyses can also reveal novel genes that contribute to vertical transmission of MGEs, representing an untapped resource to study bacterium-MGE coevolution.

Tn-seq analysis of natural transformation in *L. pneumophila*. We sought to use the Tn-seq library to identify all genes required for competence and subsequent natural transformation. Mutants defective for expression of competence, DNA uptake, protection, or recombination would not be able to undergo transformation and would thus be missing in the transformed population. We subjected the Tn-seq library to natural transformation with two distinct transforming DNAs carrying a kanamycin resistance cassette inserted in the *legK2* gene (encoding an Icm/Dot substrate) or in the *ihfB* gene (encoding the B subunit of the integration host factor [IHF]). This strategy should limit false-positive results arising from epistatic interactions between the gene in which the selected resistance cassette is inserted (*legK2* or *ihfB*) and the transposon-disrupted genes.

Transformation frequencies of the HL77S Tn-seq library were in the range of 1×10^{-5} to 6×10^{-4} , and to avoid a bottleneck effect (43), we collected over 5×10^6 transformants. The control, nontransformed populations were subsampled to obtain a similar number of isolated colonies. We observed 28 genes in which insertions cause a decrease ($\log_2FC < 2$; $P < 0.05$) of the mutants in the population transformed with either the *legK2::Kan* or the *ihfB::Kan* DNA (Fig. 2A). As expected, these include the gene encoding the periplasmic DNA receptor ComEA and the genes required for DNA transport across the inner membrane (*comEC* and *comFC*), for ssDNA protection in the cytoplasm (*dprA*), and for recombination (*comM*). Many of the other genes encode factors known to be involved in type IV pilus assembly, confirming the role of this system in natural transformation of *L. pneumophila* (17). These include the retraction ATPase PilT (lpp1995), the extension ATPase PilB, the PilQ secretin, the PilC platform protein and the proteins of the PilMNOP complex. Interestingly, we observed no transformation defect for insertions in the gene encoding another putative pilus retraction ATPase (lpp2271). Two putative pilins, PilE (lpp0681) and PilA2 (lpp1890), were also identified as essential for transformation.

Other genes potentially involved in natural transformation or regulation of competence were also identified (*lpp1976*, *lpp1977*, *lpp1978*, *lpp3030*, *lpp2632*, *djlA*, *letA*, and *letS*). In order to confirm their role and disentangle their involvement in DNA uptake or in regulation, we constructed gene deletion mutants in a Paris strain with a premature stop codon of the RocC chaperone (Paris *rocC_{TAA}*) which is defective for repression of competence and constitutively transformable (32). Deletion mutants corresponding to genes known to be involved in natural transformation were defective for transformation as expected (Fig. 2B). The *comEC* and *comFC* mutants were totally defective for transformation, and the *comM* and *comEA* mutants showed an ~ 100 -fold decrease in transformation frequencies, as observed for other species (28, 55). Similar partial transformation defects were observed for mutants of *djlA*, encoding a DnaJ-like protein required for intracellular replication in *Legionella dumoffii* (56), and *lpp3030*, a *Legionellaceae*-specific gene encoding an uncharacterized protein with a putative signal peptide. However, in this constitutively competent background, we could not confirm the involvement of *lpp2632*, which encodes a glutaryl coenzyme A (glutaryl-CoA) dehydrogenase, indicating that this gene is dispensable for the transformation process (Fig. 2B). Mutants of this gene show a reduced fitness ($\log_2FC = -1.99$; $P < 0.01$) (Data Set S1), suggesting that the transformation defect observed in the Tn-seq analysis is an indirect consequence of the mutants limited growth that could prevent entry into the competence state at the onset of the stationary phase.

Intriguingly, in the constitutively competent strain, a deletion mutant of *letS* also showed no transformation defect (Fig. 2B). LetS is the sensor of the LetA/LetS two-

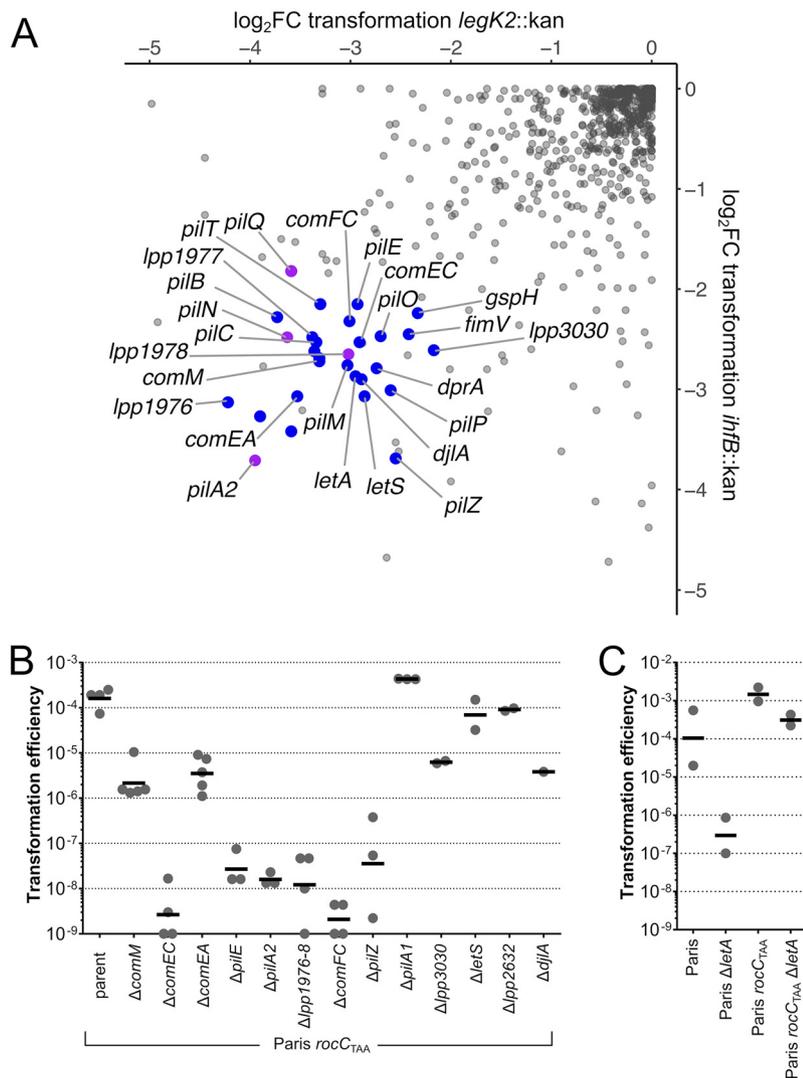


FIG 2 Identification of genes required for natural transformation by Tn-seq. (A) Scatterplot of fold change (\log_2) of insertions in the corresponding genes in two tested transformation conditions. HL77S was subjected to natural transformation with a 4-kb PCR fragment of the *legK2* or *ihfB* genes interrupted by a kanamycin resistance gene. Reads counts per gene were determined and expressed as fold change between the nontransformed population and the *legK2::kan*- or *ihfB::kan*-transformed populations. Individual genes (gray dots) were considered to be required for natural transformation if \log_2 FC was >2 or <-2 and if P was <0.05 under one (magenta dots) or both (blue dots) conditions. (B) Natural transformation efficiency of reconstructed mutants in the Paris *rocC_{TAA}* strain, which is constitutively competent for natural transformation. Transformation experiments were performed at least three times independently, and transformation frequencies were plotted (gray dots) along with the geometric means (black lines). (C) Natural transformation efficiency of the reconstructed $\Delta letA$ mutant in the original Paris strain and the constitutively competent Paris *rocC_{TAA}* strain. Transformation experiments were performed twice independently, and transformation frequencies were plotted (gray dots) along with the geometric means (black lines).

component system (TCS) homologous to the BarA/UvrY system in *Escherichia coli* (57) and GacS/GacA in *Pseudomonas* spp (58). In *L. pneumophila*, the LetA/LetS system has been identified for the first time in a screen of mutants deficient in the expression of flagellin (59) and has since been shown to be involved in the activation of various virulence traits as well as intracellular growth in amoebae (60–63). One of the major roles of the LetA/S TCS is to enable the transition from the transmissive to the replicative phase (64). The fact that both LetA and LetS are output together in the transformation screens while the *rocC_{TAA}* $\Delta letS$ mutant is not defective for transformation suggests

that this TCS is involved in the regulation of competence in *L. pneumophila*. To test this, we reconstructed insertion mutants of the *letA* gene, encoding the response regulator/activator of this TCS, in the Paris strain and *roc*_{TAA} genetic backgrounds and tested them for their ability to undergo transformation. Consistent with the Tn-seq data, inactivation of LetA in the Paris strain reduced transformability by over 500-fold (Fig. 2C). In contrast, like the Δ *letS* mutant, the Δ *letA* mutant in the constitutively competent strain *roc*_{TAA} background is only marginally affected for natural transformation (Fig. 2C). These data suggest that the LetA/S TCS is involved in the regulation of competence upstream of the regulation controlled by the RocC/RocR system. Further work will be needed to determine the precise role of this TCS and the associated regulatory cascades in the regulation of *L. pneumophila* competence.

Major and minor pilins required for natural transformation. With the remarkable exception of *Helicobacter pylori* (65), in all Gram-negative bacteria DNA uptake requires type IV pili (20). Type IV pili are extracellular filaments resulting from the assembly of thousands copies of an abundant major pilin but also of less abundant minor pilins that could be embedded in the filaments (core minor pilins) or at its tip (noncore minor pilins) (66, 67). The nomenclature of pilins is relatively confusing, but the major pilin is generally called PilA, although in *Neisseria* spp., that protein is called PilE (67). In addition, some species carry multiple copies of pilins, and at least in *Thermus thermophilus*, two major pilins (PilA4 and PilA5) are assembled into distinct filaments, required for natural transformation and twitching motility, respectively (68). The *L. pneumophila* genomes show two putative PilA homologs encoded by two consecutive genes (*pilA1* [*lpp1889*] and *pilA2* [*lpp1890*]) in a locus removed from any other genes encoding type IV pilus components. Both the Tn-seq data and reconstructed mutants show that PilA2 is required for natural transformation, while PilA1 is dispensable (Fig. 2A and B). The two copies of PilA-encoding genes may have resulted from a gene duplication event, followed by the loss of function of one of the two copies.

The Tn-seq data show that a putative pilin PilE (Lpp0681) appears to be required for transformation, while five genes upstream of *pilE* (*lpp0686* to *lpp0682*) and respectively annotated as encoding PilC/PilY1 and minor pilins PilX, PilW, PilV, and GspH/FimT appear to be dispensable. Targeted gene deletion also confirmed the Tn-seq result that PilE is required for natural transformation, corroborating an initial observation that a mutant of the *pilE* gene (then designated *pilE_L*) is not competent for transformation (17). Based on sequence comparison with PilA from *Pseudomonas aeruginosa*, *pilE_L* was then proposed to encode a type IV pilin structural gene (69). We thus investigated which of PilE and PilA2 constitute the major pilin in *L. pneumophila*. We tested the complementation of the Δ *pilE* and Δ *pilA2* deletion mutants obtained in the constitutively competent *roc*_{TAA} strain. Both *pilE* (*lpp0681*) and *pilA2* (*lpp1890*) were ectopically expressed from an IPTG (isopropyl- β -D-thiogalactopyranoside)-inducible promoter to produce fusion proteins with a C-terminal FLAG epitope. Western blot analysis showed that both PilE-FLAG and PilA2-FLAG could be expressed in an IPTG-dependent manner, with PilE-FLAG always being expressed at a higher level than PilA2-FLAG, likely reflecting the efficiency of their ribosome-binding sites (Fig. 3A). Data show that a low expression of PilE-FLAG is sufficient to restore natural transformation in the *roc*_{TAA} Δ *pilE* mutant, as full complementation of the transformation phenotype is obtained even in the absence of IPTG (Fig. 3B). In contrast, a higher concentration of IPTG and thus a higher expression of PilA2-FLAG are required to obtain a functional complementation of the *roc*_{TAA} Δ *pilA2* mutant (Fig. 3B).

The results are consistent with a model in which PilA2 is the major pilin while PilE is a low-abundance minor pilin. In addition, when expressed ectopically in the *roc*_{TAA} strain, PilA2-FLAG assembles in long extracellular filaments (Fig. 4A and B). In the *roc*_{TAA} Δ *pilE* strain, fewer PilA2-FLAG filaments are observed by microscopy, and Western blotting confirmed a lower abundance of extracellular PilA2-FLAG (Fig. 4A and B). This indicates that PilE, while not strictly essential, still plays a role in pilus formation. Minor pilins have been proposed to localize at the tip of the pilus and stabilize it

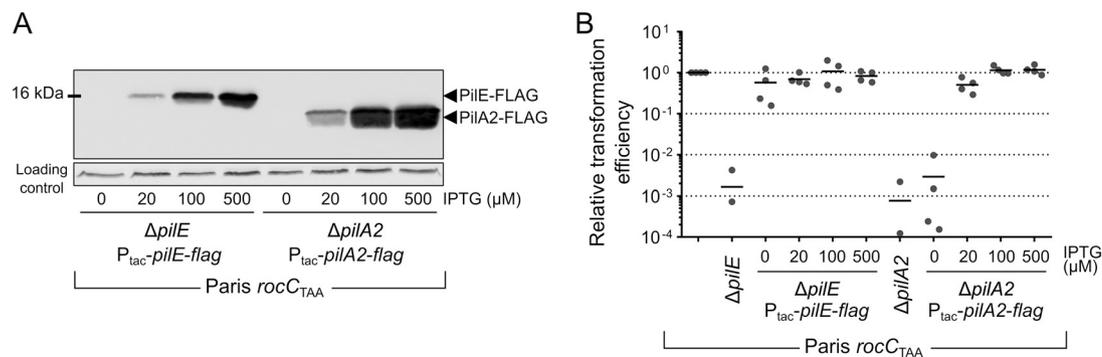


FIG 3 PilA2 is the major pilin of the *L. pneumophila* transformation pilus. (A) Western blot analysis of ectopically expressed PilA2-FLAG (encoded by p1890F) and PilE (encoded by p0681F) as a function of the IPTG inducer. (B) Complementation of the $\Delta pilE$ and $\Delta pilA2$ mutants in the Paris rocC_{TAA} strain by the ectopic expression of PilA2-FLAG (encoded by p1890F) and PilE (encoded by p0681F). Transformation frequencies were determined four times independently as a function of the IPTG inducer and normalized to 1 for the parental strain (Paris rocC_{TAA}).

(66). In *Vibrio cholerae*, DNA binding has been observed to occur at the tip of the pilus (21). Because PilE is not strictly essential for pilus assembly but required for transformation and DNA internalization (Fig. 3B and 4C), we propose that PilE is the DNA receptor at the tip of a pilus composed of PilA2 subunits.

Genes of unknown function and pilZ. In addition to the pilin mutants that were strongly defective for natural transformation, we investigated the underlying reason for the strong transformation defect of the mutant with a deletion of the operon *lpp1976-lpp1977-lpp1978* (Fig. 2B). Consistent with being important for natural transformation, this operon was found to be upregulated in the constitutively transformable rocC_{TAA} mutant (32). Automated annotation did not assign a predicted function for the

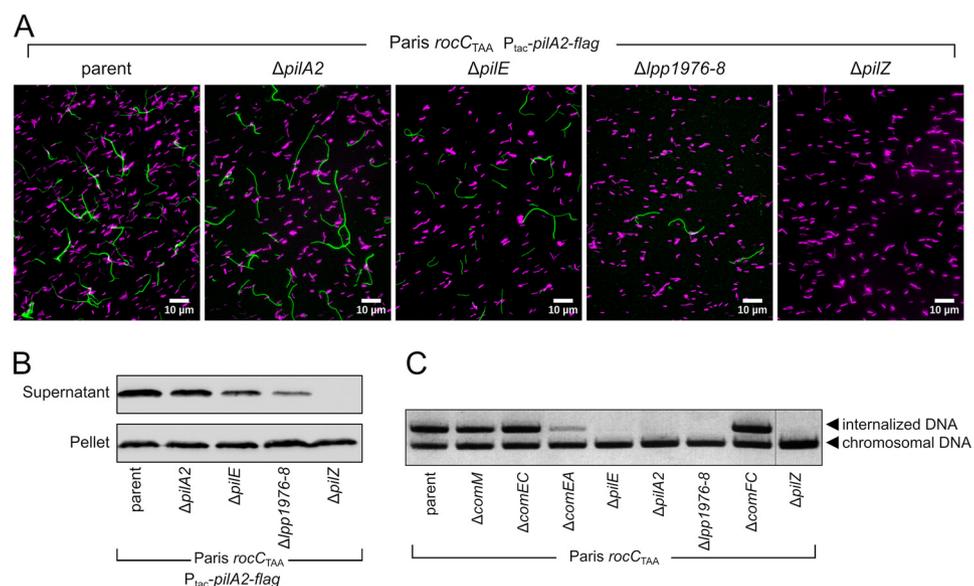


FIG 4 PilA2 assembly into extracellular filaments depends on *pilE*, the operon *lpp1976-lpp1978*, and *pilZ*. (A) Visualization of PilA2-FLAG filaments (green) by immunofluorescence microscopy using fluorescein-conjugated anti-FLAG antibody. Bacteria were visualized by labeling DNA with Hoechst 33288 (magenta). (B) Western blot detection of extracellular PilA2. Bacteria were vortexed to release pili which were precipitated from supernatants. PilA2-FLAG was detected in supernatant and whole-cell lysates (pellet) using anti-FLAG antibodies. (C) DNA uptake assay of the reconstructed mutants defective for natural transformation. The transformation-deficient mutants were tested for the ability to internalize pGEM-HYG1, a nonreplicative plasmid. Following incubation with the DNA and subsequent DNase I treatment, the internalized DNA was detected by PCR for pGEM-HYG. As a control, chromosomal DNA was also detected by PCR. This multiplex PCR was analyzed by agarose gel electrophoresis and labeling of DNA with ethidium bromide. Images are representative of experiments performed three times independently.

three genes, and a BLAST search failed to identify homologs outside the genus *Legionella*. The mutant with the deletion of the entire operon ($\Delta lpp1976-lpp1978$) was unable to take up DNA (Fig. 4C), indicative of a defect in type IV pilus-mediated DNA import. Indeed, in this mutant, the levels of extracellular PilA2-FLAG were strongly reduced (Fig. 4B). The mutant produced few, and short, PilA2-FLAG filaments (Fig. 4A), revealing a major defect in type IV pilus assembly or stability. A search for conserved domains in the three predicted proteins identified, in the 268-amino-acid Lpp1977, only a partial homology with the N-terminal part of the Tfp pilus assembly protein PilW. This suggested that *lpp1976-lpp1977-lpp1978* would encode a set of minor pilins. Indeed, PilFind (70) identified an N-terminal transmembrane segment in all three predicted proteins and a type II signal in Lpp1977 and Lpp1978. The operonic organization of these three genes is reminiscent of the operon encoding four minor pilins of the type IV pilus of *Neisseria meningitidis* (*pilHIJK*) and *P. aeruginosa* (*fimU-pilVWX*) and of the type II secretion system (T2SS) of enterotoxigenic *E. coli* (*gspHIJK*). In the latter system, the last three genes (*gspIJK*) encode minor pseudopilins which assemble into a stable complex (71). This complex of minor pilins would form in the inner membrane to establish a platform for the assembly of the major pilin (72) and remain at the tip of the pilus, stabilizing it (66). Such heterotrimeric complex may be formed by minor pilins of limited homology but displaying structural similarity (73). Altogether, this supports the hypothesis that in *Legionella* species, the initiation complex of the transformation pilus is formed by Lpp1976, Lpp1977, and Lpp1978, which serve as a scaffold for assembly of the major pilin PilA2.

Another gene whose deletion resulted in strong deficiency in natural transformation is *pilZ*. The $\Delta pilZ$ mutant is defective for DNA uptake and is totally unable to produce extracellular PilA2 or assemble PilA2 filaments (Fig. 4). PilZ was originally identified in *P. aeruginosa* as required for the secretion of PilA polymers, pilus genesis, and type IV pilus-dependent motility (74). Although the *P. aeruginosa* PilZ served as the founding member of a diverse family of proteins with PilZ domains (75), some of which bind the cyclic di-GMP (c-di-GMP) second messenger, it itself does not bind c-di-GMP (76). A *pilZ* mutant in *Xanthomonas campestris* pv. *campestris* displays a minor defect in type IV pilus-dependent motility (77), and this PilZ ortholog directly binds to the PilB ATPase and the c-di-GMP-interacting FimX protein (78). However, no homolog of FimX could be identified in *L. pneumophila*, and Tn-seq did not reveal any c-di-GMP synthesis enzyme required for natural transformation. However, Tn-seq did show that PilB was important for transformation (Fig. 2A). We thus speculate that, in *L. pneumophila*, PilZ controls type IV pilus assembly independently of c-di-GMP signaling and through a direct interaction with PilB.

Conclusion. We report on a clinical isolate of *L. pneumophila*, which displays phenotypes (intracellular replication and competence for natural transformation) similar to those of commonly used laboratory strains. In contrast to laboratory strains, a high-saturation Tn-seq library could be obtained and allowed to define essential genes, including strain-specific genes in MGEs. Tn-seq analyses of transformation, with follow-up work performed in the Paris strain, defined the set of major and minor type IV pilins that are engaged in DNA uptake. While we focused here on mutants that were strongly deficient for natural transformation, Tn-seq also identified potential regulators of competence as well as genes of unknown function that also participate in natural transformation (for instance, *djIA* and *lpp3030*). We demonstrate here that strain HL775 could represent a surrogate for the commonly used lab strains to perform Tn-seq analysis. Unleashing the full power of Tn-seq is a major step toward the identification of the genetic basis of traits that turned *L. pneumophila* into a successful pathogen, such as its ability to form biofilms and to resist biocides and unicellular predators.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Legionella pneumophila* strains were grown in liquid medium containing ACES [N-(2-acetamido)-2-aminoethanesulfonic acid] and buffered yeast extract (AYE) or on solid medium plates containing ACES and buffered charcoal yeast extract (CYE). When appropriate,

kanamycin, gentamicin, and streptomycin were added at 15 $\mu\text{g/ml}$, 10 $\mu\text{g/ml}$, and 50 $\mu\text{g/ml}$, respectively. Clinical isolates of *L. pneumophila*, including HL-0709-3014 (referred to as HL77), were provided by the Centre National de Référence des Légionelles, Lyon, France. A streptomycin-resistant of HL-0709-3014 mutant was obtained by plating 1 ml of culture on a streptomycin-containing CYE plate and was named HL77S. *Escherichia coli* MFDpir (79) with a chromosome-integrated RP4 conjugative system was used as a donor strain for conjugative transfer of the mutagenesis system pBT20 (80), which carries a Himar1 transposon bearing a gentamicin resistance gene and an outward-facing promoter. MFDpir is auxotrophic for diaminopimelic acid (DAP) and thus was always cultivated with 1% DAP. Axenic *Acanthamoeba castellanii* cells were grown in proteose-yeast extract-glucose (PYG) medium at 30°C and split once or twice a week. Human U937 cells were maintained in RPMI 1640 with 10% heat-inactivated fetal calf serum and 1% penicillin-streptomycin at 37°C and 5% CO₂. Differentiation into macrophages was induced by the addition of PMA (phorbol 12-myristate 13-acetate) at a final concentration of 100 ng/ml. Murine macrophages, RAW 264.7, were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C and 5% CO₂.

Intracellular growth experiments. The ability of HL77S strain to infect host cells was compared to that of the Paris strain. Paris ΔdotA was used as a negative control, as this mutant is unable to multiply in host cells. The ability of HL77S to replicate in the amoeba *Acanthamoeba castellanii* was determined as follows. Amoebas were resuspended in PYG medium at a concentration of 1×10^6 amoebas/ml. The suspension was distributed in a flat-bottom 6-well plate (2 ml per well; 2×10^6 amoebas per well) and incubated for 3 h at 30°C to allow amoebas to settle and adhere to the plate. One milliliter of a PY medium suspension containing 2×10^6 bacteria (from a culture in stationary phase, optical density [OD] of ~ 5) were added in each well to obtain a multiplicity of infection (MOI) of 1. The plate was centrifuged for 10 min at $650 \times g$ and incubated at 30°C for 72 h. At 0 h, 48 h, and 72 h, 250 μl of supernatant of each well was serially diluted, spotted onto CYE plates, and incubated at 37°C for 72 h to determine the number of CFU per milliliter.

The ability of *L. pneumophila* strains to infect macrophages was determined as follows. Overnight cultures of bacterial strains (optical density [OD] of ~ 5 in AYE medium) were diluted (1:10) in the appropriate cell culture medium (DMEM for RAW 264.7 cells and RPMI 1640 for U937 cells) and incubated for 1 h at 37°C. Host cells (differentiated U937 and RAW 264.7 cells) were seeded in 24-well plates, 3 wells per condition. Cells were washed and were infected with bacteria at an MOI of 1 or 10. Plates were centrifuged at $500 \times g$ for 5 min to promote bacterium-cell contact and incubated at 37°C and 5% CO₂ for 24 h to 72 h. Every 24 h, the contents of one well per condition was transferred to a 1.5-ml tube and centrifuged at $16,000 \times g$ for 5 min. The pellet containing the infected macrophages was resuspended in sterile distilled water to lyse the macrophages and release the bacteria. The suspension was serially diluted, spotted onto CYE plates, and incubated at 37°C for 72 h to determine the titer, in CFU per milliliter.

Plasmid and strain constructions. Plasmid pJET1.2-*legk2::kan*, used for natural transformation experiments, was constructed by cloning a 6-kb fragment consisting of the *legK2::kan* gene (81) and 2 kb of its flanking regions in the pJET1.2/blunt cloning vector (Thermo Fisher) according to the manufacturer's instructions. All the mutants generated in this study are derived from *L. pneumophila* Paris or *L. pneumophila* Paris *rocC-TAA*. All the genes suspected to be involved in natural transformation were deleted by replacement with a kanamycin resistance gene. To do so, the upstream (PCR product A [PCRA]; 2 kb) and downstream (PCRC; 2 kb) regions of each suspected gene were amplified, respectively, with the primer pairs X_{P1}/X_{P2}-tail-pKD4 and X_{P3}-tail-pKD4/X_{P4} (where X designates the gene to be deleted). X_{P2}-tail-pKD4 and X_{P3}-tail-pKD4 carry 30-nucleotide sequences complementary to the ends of the kanamycin cassette. This complementarity was used to assemble PCRA and PCRC to the kanamycin resistance cassette (PCRB; 1,490 kb amplified from plasmid pGEMPKD4 (32) with primer pair pKD4_P1/pKD4_P2) by overlap extension PCR. Overlapping PCR products were naturally transformed in the desired strain. Transformants were selected on CYE supplemented with kanamycin (15 $\mu\text{g/ml}$). Integration of the *kan* cassette at the correct locus was finally verified by colony PCR. Plasmids p1890F and p0681F, encoding the FLAG-tagged PiiA2 and PiiE, were constructed by amplifying *lpp1890* (*pilA2*) and *lpp0681* (*pilE*) with primers *lpp1890-F/lpp1890F-R* and *lpp0681-F/lpp0681F-R*, respectively. The PCR products and the recipient plasmid pMMB207C were digested with HindIII/BamHI and ligated to place the genes under the control of the *Ptac* promoter. All strains, plasmids, and oligonucleotides are listed in Table S1 in the supplemental material.

Generation of a transposon insertion mutant library of *Legionella pneumophila*. Transposon-based random mutagenesis was performed as previously described (82) by conjugative delivery of the Himar1-based transposon suicide vector pBT20 from the donor strain *E. coli* MFDpir to the recipient strain of *L. pneumophila* to be mutagenized. To do so, both bacteria were cultivated overnight at 37°C with shaking in their corresponding liquid media: 7.5 ml LB broth containing 100 $\mu\text{g/ml}$ ampicillin and 1% DAP for *E. coli* and 15 ml standard AYE medium for *L. pneumophila*. Once in stationary phase (OD of ~ 5), the *L. pneumophila* and *E. coli* cultures were concentrated by centrifugation ($5,000 \times g$, 10 min), and cell pellets were resuspended, respectively, in 1.5 ml sterile water and 0.750 ml sterile phosphate-buffered saline (PBS). To promote cell-to-cell contacts and the subsequent conjugation, both concentrated cultures were mixed together by pipetting and spotted on CYE plates without iron or cysteine but supplemented with DAP (CYED) (82) until the sample was exhausted. Plates were incubated at 37°C for 5 to 6 h. All the spots were resuspended in sterile water and used to inoculate transconjugant-selective plates (i.e., CYE plates supplemented with 10 $\mu\text{g/ml}$ of gentamicin). In parallel, the suspension was 10-fold diluted and spotted onto transconjugant-selective plates to evaluate the number of mutants in the library. After 72 h of incubation at 37°C, a mutant library was obtained by collecting all colonies from

the plates and resuspending them in AYE–5% glycerol. The suspension was aliquoted and stored at -80°C until its use for a Tn-seq screen. This library, called initial isolation, is named sample XRCR13.

Natural transformation Tn-seq screen. Transposon mutants of *L. pneumophila* HL775 were screened for their ability to undergo transformation. To avoid a bottleneck effect, a volume of the -80°C frozen library containing 10 times the number of mutants in the library was spotted on a CYE plate supplemented with gentamicin ($10\ \mu\text{g}/\text{ml}$) and streptomycin ($50\ \mu\text{g}/\text{ml}$) to obtain exponentially growing cells. After 24 h of incubation at 37°C , fresh bacteria from the spot were resuspended in AYE to an OD of ~ 0.2 . This suspension was used to perform transformation assays using $2\ \mu\text{g}/\text{ml}$ of either pGEM-ihfB-kan or pGET1.2-legK2-kan as transforming DNA, both conferring resistance to kanamycin. For both transforming DNAs, the transformation screen was conducted in duplicate. The suspensions were cultivated at 30°C with shaking for 40 h to ensure that bacteria underwent transformation and achieved an OD of ~ 5 . These conditions were expected to give about $\sim 10^5$ transformants/ml for the DNA conditions. Regarding the no-DNA condition, cultures were diluted in AYE to obtain the same number of CFU on nonselective plates as the output condition on selective plates. Each sample (with DNA and without DNA) was used to inoculate, respectively, nonselective (i.e., CYE) and transformant-selective (i.e., CYE supplemented with $20\ \mu\text{g}/\text{ml}$ of kanamycin) plates. In parallel, no-DNA samples were 10-fold diluted and spotted on nonselective plates to determine transformation frequencies as mentioned above. After 72 h of incubation at 37°C , colonies were collected from the plates and resuspended in AYE–15% glycerol until the preparation of DNA libraries. The no-DNA condition is also referred to as the second isolation used in the fitness analysis.

DNA library preparation and sequencing. Libraries were prepared as previously described (82). Mutant libraries from the -80°C frozen stock were thawed and centrifuged at maximum speed to pellet them. Genomic DNA (gDNA) extraction was carried out directly on the pellet cells with the Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions. Approximately $30\ \mu\text{g}$ of DNA was mechanically sheared by sonication using a Branson sonifier for 4 min (1 s on and 11 s off; 20% intensity) in 0.5-ml PCR tubes kept on ice. Small gDNA molecules were removed by mixing sonicated gDNA with $0.6\times$ Agencourt Ampure XL magnetic beads (Beckman Coulter) according to the manufacturer's instructions. These treatments led to gDNA fragments being between 300 and 1,000 bp. Homopolymeric cytosine tails (C-tails) were then added to the 3' ends of all fragments by incubation of $3\ \mu\text{g}$ of size-selected DNA fragments with the recombinant terminal deoxynucleotidyl transferase (rTdT; $30\ \text{U}/\mu\text{l}$; Promega) at 37°C for 1 h, followed by heat inactivation at 75°C for 20 min. TdT reagents were then removed by purifying the TdT reaction mixture with $1\times$ of Ampure XL beads.

To amplify transposon junctions, a first-round of PCR (PCR1) was performed in a final volume of $50\ \mu\text{l}$ by mixing $500\ \text{ng}$ C-tailed DNA, $1\ \mu\text{l}$ biotinylated pBT20-PCR1 primer ($30\ \mu\text{M}$), $3\ \mu\text{l}$ olj376 primer ($30\ \mu\text{M}$), $2.5\ \mu\text{l}$ deoxynucleoside triphosphates (dNTPs) ($10\ \text{mM}$), $10\ \mu\text{l}$ Q5 reaction buffer, and $0.75\ \mu\text{l}$ Q5 high-fidelity DNA polymerase (New England Biolabs). PCR1 products were purified using $1\times$ Ampure beads. Biotinylated and purified PCR1 products were then selectively captured using streptavidin-containing Dynabeads (M-280; Invitrogen) according to the manufacturer's instructions. A second round of PCR was carried out in a final volume of $50\ \mu\text{l}$ by resuspending Dynabeads (which had the PCR1 products bound to them) in the prepared PCR2 reaction mix consisting of $1\ \mu\text{l}$ pBT20-PCR2 primer ($30\ \mu\text{M}$), $1\ \mu\text{l}$ TdT_index_X primer ($30\ \mu\text{M}$), $2.5\ \mu\text{l}$ dNTPs ($10\ \text{mM}$), $10\ \mu\text{l}$ Q5 reaction buffer, and $0.75\ \mu\text{l}$ Q5 high-fidelity DNA polymerase. PCR2 products were purified with $1\times$ Ampure XL beads. The obtained libraries were sequenced on an Illumina HiSeq 4000 in single-end 50 bp using the custom sequencing primer Read1TnLp. Samples and conditions are listed in Data Set S1. Essentiality analysis was performed using reads from sample XRCR13. Fitness analysis was performed by comparing reads from samples XRCR24, -26, -36, and -38 (no-DNA conditions from the transformation screen) versus XRCR13. Analysis of transformation was performed by comparing samples XRCR27 and -39 (*legK2::kan* transforming DNA) to samples XRCR26 and -38 (no-DNA control) and by comparing samples XRCR25 and -37 (*ihfB::kan* transforming DNA) to samples XRCR24 and -36 (no-DNA control).

Tn-seq data analysis. For each condition, 10 to 50 million reads were obtained and trimmed with tools from the Galaxy Project's public server. Fastx_clipper was used to cut poly(C) tails and remove short reads [<15 bp after poly(C) clipping]. Then, reads were filtered by quality using Trimmomatic and quality checked with FastQC. Trimmed reads in the fastq output file were mapped to the reference genome using the Tn-seq software TPP (Tn-seq preprocessor) (83). Output wig files from TPP were used to perform essentiality analysis using Transit (84). Single-condition essentiality analysis was performed with the HMM (48) or Gumbel (47) method. Conditional essentiality analysis was performed with the "resampling" method according to the Transit software documentation. The complete genome sequence of HL-0709-3014 was obtained (see "Genome sequencing" below) and annotated with Prokka (85). An orthology search was carried out between the strains of *L. pneumophila* HL775, Paris, and Philadelphia-1 using the orthology detection eggNOG mapper (86), and COG and KEGG numbers were assigned to each gene.

Transformation assays. Natural transformation assays were conducted differently depending on the genetic background of the *L. pneumophila* strain used.

For the constitutively transformable *rocC_{TAA}* strains, natural transformation was conducted on solid medium at 37°C as follows. The strains were streaked on CYE solid medium from a frozen stock culture and incubated for 72 h at 37°C . The strains were then restreaked on a new CYE plate and incubated overnight at 37°C to obtain freshly growing cells. Bacteria were resuspended in sterile water to an OD_{600} of 1 to obtain a suspension of 1.10^9 CFU/ml. A $10\text{-}\mu\text{l}$ portion of the suspensions ($\sim 1 \times 10^7$ CFU) was spotted on CYE with $1.5\ \mu\text{g}$ of transforming DNA. Once the spots were absorbed by the agar, plates were incubated at 37°C for 24 h. Each spot was resuspended in $200\ \mu\text{l}$ sterile water and used to perform 10-fold serial dilutions, which were then plated on nonselective medium and selective medium. Plates

were incubated at 37°C for 72 h. Finally, transformation frequencies were calculated as the ratio of the number of CFU counted on selective medium divided by the number of CFU counted on nonselective medium. For all the *rocC_{TAA}* strains, the *rpsL* PCR product was used as transforming DNA. This transforming DNA is obtained by amplifying the 2-kb regions upstream and downstream of the *rpsL* single point mutation conferring resistance to streptomycin (PCR primer pair *rpsL_F/rpsL_R*). Transformation experiments on strains bearing plasmids p0681F and p1890F were performed the same way, using CYE plates containing different concentrations of IPTG.

For the non-constitutively transformable strains of *L. pneumophila*, transformation was carried out in liquid medium at 30°C as follows. Strains were streaked on CYE solid medium from a frozen stock culture, incubated for 72 h at 37°C, restreaked on a new CYE plate, and incubated overnight at 37°C. Fresh bacteria were resuspended in 3 ml of AYE in 13-ml tubes to an OD of ~0.2 with 2 µg of transforming DNA and cultivated at 30°C with shaking for 24 h. Tenfold serial dilutions of each culture were then performed, plated on nonselective medium and selective medium, and incubated at 37°C for 72 h. Finally, transformation frequencies were determined as described above. (3) For *letA* mutants of constitutively and non-constitutively transformable strains of *L. pneumophila*: strains were streaked on CYE solid medium from a frozen stock culture, incubated for 72 h at 37°C, restreaked on a new CYE plate, and incubated overnight at 37°C. Fresh bacteria were resuspended in 3 ml of AYE in 13-ml tubes to an OD of ~0.2 and cultivated at 30°C with shaking until an OD of ~2 to 4 (corresponding to the competence phase of *L. pneumophila*) was reached. A volume corresponding to 1×10^8 bacteria was spotted on CYE plates with 1.5 µg of the *rpsL* PCR product. The following steps were the same as for the transformation of constitutively transformable *rocC_{TAA}* strains as mentioned in item 1 above.

Detection of extracellular pilin by Western blotting. Strains bearing plasmid p1890F were grown overnight at 37°C on CYE containing 500 µM IPTG and were then resuspended in 2 ml AYE at an optical density at 600 nm (OD₆₀₀) of ~1.5. One milliliter of the suspension was then subjected to maximum-speed vortex agitation for 1 min and centrifuged for 15 min at $21,000 \times g$ and 4°C. Supernatants were recovered in a new tube and centrifuged again, while pellets were saved on ice. After centrifugation, 900 µl of supernatants were recovered and proteins were precipitated by adding 100 µl of trichloroacetic acid (TCA; final concentration of 10%). After 30 min of incubation on ice, a 15-min centrifugation at $21,000 \times g$ and 4°C was performed. Pellets were washed three times with acetone, dried at room temperature, and resuspended with 100 µl of 1× Laemmli sample buffer. Pellets previously saved on ice were resuspended with 150 µl of 1× sample buffer. Samples were then analyzed by Western blotting. Aliquots were boiled for 5 min and subjected to denaturing polyacrylamide gel electrophoresis. Proteins from SDS-polyacrylamide gels were electrophoretically transferred to nitrocellulose membranes (Schleicher and Schuell) and subsequently stained with Ponceau S (Sigma) to check the loading of the lanes. Membranes were incubated with monoclonal anti-FLAG antibody (1:1,000 dilution; no. F1804; Sigma) as a primary antibody and an anti-mouse immunoglobulin-peroxidase conjugate (1:50,000 dilution; no. A0168; Sigma) as a secondary antibody. Nitrocellulose membranes were revealed with the SuperSignal West Dura detection system (Pierce) and an imaging workstation equipped with a charge-coupled device camera (Thermo).

Determination of the DNA uptake ability. The ability of the transformation-deficient mutants to take up DNA was determined as follows: strains were inoculated in AYE medium at an OD₆₀₀ of 0.05, and tubes were incubated overnight under constant shaking at 30°C. When an OD₆₀₀ of 0.9 was reached, 1 ml of each culture was centrifuged for 3 min at $5,000 \times g$, and pellets were resuspended in 200 µl ultrapure water containing 2 µg of pGEM-HYG1. This plasmid is a nonreplicative plasmid in *L. pneumophila* and, as it contains no homology with the *L. pneumophila* genome, it cannot integrate by recombination. After 20 min of incubation at 37°C, tubes were centrifuged for 3 min at $5,000 \times g$, and pellets were resuspended in 200 µl AYE liquid medium containing 10 U of DNase I (Sigma). After 20 min of incubation at 37°C, DNase I was removed, and bacteria were washed by two successive centrifugations for 3 min at $5,000 \times g$ and resuspension in 1 ml of water. Pellets were finally resuspended in 100 µl ultrapure water and incubated for 30 min at 65°C to complete DNase I inactivation and kill bacteria. The DNA uptake ability of each mutant was then determined by PCR, using two pairs of primers amplifying on the one hand the chromosomal *mreB* gene (*lpp0873*) and on the other hand a part of pGEM-HYG1, giving, respectively, PCR products of 1,194 bp (*mreB*seqF/*mreB*seqR) and 1,657 bp [M13F(-47)/M13R(-48)].

Microscopy. Bacteria expressing the FLAG-tagged pilins were grown as spots on CYE plates with 0.5 mM IPTG for 24 h at 37°C. Bacteria were gently collected with a pipette tip. In order to limit shearing and breaking of the pilus, the pipette tip was left standing an Eppendorf tube with 1 ml of water for a few minutes. Once the collected bacterial culture was starting to dissociate and falling off the tip, the bacterial pellet was resuspended gently by slowly pipetting up and down. The collected 1-ml suspensions were centrifuged for 3 min at $5000 \times g$, and pellets were gently resuspended in 300 µl PBS-3.7% formaldehyde and incubated at room temperature for 30 min. Acid-washed (ethanol-HCl, 1 M) glass coverslips were coated with poly-L-lysine by immersion in a 0.01% poly-L-lysine solution in distilled water (Sigma-Aldrich) for 5 min. Fixed bacteria in PBS-3.7% formaldehyde were pipetted (250 µl) on the air-dried coverslips and allowed to settle and stick to the coverslips for about 30 min. Coverslips were then washed twice with PBS and incubated with monoclonal anti-FLAG M2-fluorescein conjugates at 1/200 in PBS for 1 h. Coverslips were then washed twice with PBS, and DNA was labeled using Hoechst 33288 (12 µg/ml in PBS) for 1 h. Coverslips were washed twice in PBS and mounted using 8 µl of mounting solution (1,4-diazabicyclo[2.2.2]-octane [DAPCO]). After an overnight incubation at 4°C, slides were observed and imaged with an epifluorescence microscope (Zeiss Axioplan 2).

Genome sequencing. The complete genome of isolate HL-0709-3014 was obtained using Illumina MiSeq paired-end reads from previously available SRA sample [ERS1305867](#) and long reads from Oxford

Nanopore sequencing on a MinION sequencer according to the manufacturer's instructions (Oxford Nanopore). Illumina and Nanopore reads were then used for short-read/long-read hybrid assembly using Unicycler v0.4.6 (87).

Data availability. The complete genome of isolate HL-0709-3014 is available under accession numbers [CP048618.1](https://doi.org/10.1093/bioinformatics/btad018) (chromosome) and [CP048619.1](https://doi.org/10.1093/bioinformatics/btad018) (plasmid). The strain is listed under the name *Legionella pneumophila* strain ERS1305867 (BioProject no. [PRJEB15241](https://doi.org/10.1093/bioinformatics/btad018); BioSample no. SAMEA4394418). Raw sequencing reads of the Tn-seq samples are available at the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/>) under the study accession number [PRJEB40244](https://doi.org/10.1093/bioinformatics/btad018).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

SUPPLEMENTAL FILE 2, XLSX file, 1.7 MB.

ACKNOWLEDGMENTS

We warmly thank Christophe Ginevra (Centre National de Référence des Légionelles) for kindly providing Nanopore sequencing reads and full genome assembly of HL-0709-3014. We also thank Chloé Vallantin for technical assistance with clinical isolates and Annelise Chapalain and Johann Guillemot for providing mouse and human monocyte cultures and helpful advice on infection with *L. pneumophila*. We thank Vladimir Shevchik for his insight into type IV secretion and minor pili. We acknowledge Laetitia Attaiech and Maria-Halima Laaberki for their critical assessment of the manuscript.

This work was supported in part by the LABEX ECOFECT (ANR-11-LABX-0048) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). This work was supported by an intramural grant from the Centre International de Recherche en Infectiologie.

L.H., P.-A.J., and B.C.-G. designed and performed experiments and analyzed data. L.H. and X.C. analyzed Tn-seq data. L.H. and X.C. wrote the manuscript. X.C. conceptualized and supervised the project.

REFERENCES

- Molmeret M, Horn M, Wagner M, Santic M, Abu Kwaik Y. 2005. Amoebae as training grounds for intracellular bacterial pathogens. *Appl Environ Microbiol* 71:20–28. <https://doi.org/10.1128/AEM.71.1.20-28.2005>.
- Rowbotham TJ. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Pathol* 33:1179–1183. <https://doi.org/10.1136/jcp.33.12.1179>.
- Fields BS, Benson RF, Besser RE. 2002. Legionella and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* 15:506–526. <https://doi.org/10.1128/CMR.15.3.506-526.2002>.
- Nash TW, Libby DM, Horwitz MA. 1984. Interaction between the Legionnaires' disease bacterium (*Legionella pneumophila*) and human alveolar macrophages. Influence of antibody, lymphokines, and hydrocortisone. *J Clin Invest* 74:771–782. <https://doi.org/10.1172/JCI111493>.
- Gao LY, Harb OS, Abu Kwaik Y. 1997. Utilization of similar mechanisms by *Legionella pneumophila* to parasitize two evolutionarily distant host cells, mammalian macrophages and protozoa. *Infect Immun* 65:4738–4746. <https://doi.org/10.1128/IAI.65.11.4738-4746.1997>.
- Vogel JP, Andrews HL, Wong SK, Isberg RR. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279:873–876. <https://doi.org/10.1126/science.279.5352.873>.
- Segal G, Purcell M, Shuman HA. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc Natl Acad Sci U S A* 95:1669–1674. <https://doi.org/10.1073/pnas.95.4.1669>.
- Zhu W, Banga S, Tan Y, Zheng C, Stephenson R, Gately J, Luo Z-Q. 2011. Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of *Legionella pneumophila*. *PLoS One* 6:e17638. <https://doi.org/10.1371/journal.pone.0017638>.
- Burstein D, Zusman T, Degtyar E, Viner R, Segal G, Pupko T. 2009. Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathog* 5:e1000508. <https://doi.org/10.1371/journal.ppat.1000508>.
- Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, Tichit M, Jarraud S, Bouchier C, Vandenesch F, Kunst F, Etienne J, Glaser P, Buchrieser C. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* 36:1165–1173. <https://doi.org/10.1038/ng1447>.
- Chien M, Morozova I, Shi S, Sheng H, Chen J, Gomez SM, Asamani G, Hill K, Nuara J, Feder M, Rineer J, Greenberg JJ, Steshenko V, Park SH, Zhao B, Teplitskaya E, Edwards JR, Pampou S, Georghiou A, Chou I-C, Iannuccilli W, Ulz ME, Kim DH, Geringer-Sameth A, Goldsberry C, Morozov P, Fischer SG, Segal G, Qu X, Rzhetsky A, Zhang P, Cayanis E, De Jong PJ, Ju J, Kalachikov S, Shuman HA, Russo JJ. 2004. The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* 305:1966–1968. <https://doi.org/10.1126/science.1099776>.
- de Felipe KS, Pampou S, Jovanovic OS, Pericone CD, Ye SF, Kalachikov S, Shuman HA. 2005. Evidence for acquisition of *Legionella* type IV secretion substrates via interdomain horizontal gene transfer. *J Bacteriol* 187:7716–7726. <https://doi.org/10.1128/JB.187.22.7716-7726.2005>.
- Lurie-Weinberger MN, Gomez-Valero L, Merault N, Glöckner G, Buchrieser C, Gophna U. 2010. The origins of eukaryotic-like proteins in *Legionella pneumophila*. *Int J Med Microbiol* 300:470–481. <https://doi.org/10.1016/j.ijmm.2010.04.016>.
- David S, Rusniok C, Mentasti M, Gomez-Valero L, Harris SR, Lechat P, Lees J, Ginevra C, Glaser P, Ma L, Bouchier C, Underwood A, Jarraud S, Harrison TG, Parkhill J, Buchrieser C. 2016. Multiple major disease-associated clones of *Legionella pneumophila* have emerged recently and independently. *Genome Res* 26:1555–1564. <https://doi.org/10.1101/gr.209536.116>.
- Sánchez-Busó L, Comas I, Jorques G, González-Candelas F. 2014. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet* 46:1205–1211. <https://doi.org/10.1038/ng.3114>.
- David S, Sánchez-Busó L, Harris SR, Marttinen P, Rusniok C, Buchrieser C, Harrison TG, Parkhill J. 2017. Dynamics and impact of homologous recombination on the evolution of *Legionella pneumophila*. *PLoS Genet* 13:e1006855. <https://doi.org/10.1371/journal.pgen.1006855>.

17. Stone BJ, Kwai YA. 1999. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. *J Bacteriol* 181:1395–1402. <https://doi.org/10.1128/JB.181.5.1395-1402.1999>.
18. Dubnau D, Blokesch M. 2019. Mechanisms of DNA uptake by naturally competent bacteria. *Annu Rev Genet* 53:217–237. <https://doi.org/10.1146/annurev-genet-112618-043641>.
19. Johnsborg O, Eldholm V, Håvarstein LS. 2007. Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* 158:767–778. <https://doi.org/10.1016/j.resmic.2007.09.004>.
20. Piepenbrink KH. 2019. DNA uptake by type IV filaments. *Front Mol Biosci* 6:1. <https://doi.org/10.3389/fmolb.2019.00001>.
21. Ellison CK, Dalia TN, Vidal Ceballos A, Wang JC-Y, Biais N, Brun YV, Dalia AB. 2018. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat Microbiol* 3:773–780. <https://doi.org/10.1038/s41564-018-0174-y>.
22. Seitz P, Pezeshgi Modarres H, Borgeaud S, Bulushev RD, Steinbock LJ, Radenovic A, Dal Peraro M, Blokesch M. 2014. ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. *PLoS Genet* 10:e1004066. <https://doi.org/10.1371/journal.pgen.1004066>.
23. Hepp C, Maier B. 2016. Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. *Proc Natl Acad Sci U S A* 113:12467–12472. <https://doi.org/10.1073/pnas.1608110113>.
24. Draskovic I, Dubnau D. 2005. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol Microbiol* 55:881–896. <https://doi.org/10.1111/j.1365-2958.2004.04430.x>.
25. Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol Microbiol* 50:527–536. <https://doi.org/10.1046/j.1365-2958.2003.03702.x>.
26. Attaiech L, Olivier A, Mortier-Barrière I, Soulet A-L, Granadel C, Martin B, Polard P, Claverys J-P. 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLoS Genet* 7:e1002156. <https://doi.org/10.1371/journal.pgen.1002156>.
27. Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, Fichant G, Martin B, Noirot P, Le Cam E, Polard P, Claverys J-P. 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* 130:824–836. <https://doi.org/10.1016/j.cell.2007.07.038>.
28. Nero TM, Dalia TN, Wang JC-Y, Kysela DT, Bochman ML, Dalia AB. 2018. ComM is a hexameric helicase that promotes branch migration during natural transformation in diverse Gram-negative species. *Nucleic Acids Res* 46:6099–6111. <https://doi.org/10.1093/nar/gky343>.
29. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* 12:181–196. <https://doi.org/10.1038/nrmicro3199>.
30. Charpentier X, Kay E, Schneider D, Shuman HA. 2011. Antibiotics and UV radiation induce competence for natural transformation in *Legionella pneumophila*. *J Bacteriol* 193:1114–1121. <https://doi.org/10.1128/JB.01146-10>.
31. Buchrieser C, Charpentier X. 2013. Induction of competence for natural transformation in *Legionella pneumophila* and exploitation for mutant construction. *Methods Mol Biol* 954:183–195. https://doi.org/10.1007/978-1-62703-161-5_9.
32. Attaiech L, Boughammoura A, Brochier-Armanet C, Allatif O, Peillard-Fiorente F, Edwards RA, Omar AR, MacMillan AM, Glover M, Charpentier X. 2016. Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc Natl Acad Sci U S A* 113:8813–8818. <https://doi.org/10.1073/pnas.1601626113>.
33. Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, Charpentier X. 2019. Diverse conjugative elements silence natural transformation in *Legionella* species. *Proc Natl Acad Sci U S A* 116:18613–18618. <https://doi.org/10.1073/pnas.1909374116>.
34. Kessler A, Schell U, Sahr T, Tiaden A, Harrison C, Buchrieser C, Hilbi H. 2013. The *Legionella pneumophila* orphan sensor kinase LqsT regulates competence and pathogen-host interactions as a component of the LAL-1 circuit. *Environ Microbiol* 15:646–662. <https://doi.org/10.1111/j.1462-2920.2012.02889.x>.
35. Hochstrasser R, Hilbi H. 2017. Intra-species and inter-kingdom signaling of *Legionella pneumophila*. *Front Microbiol* 8:79. <https://doi.org/10.3389/fmicb.2017.00079>.
36. Attaiech L, Charpentier X. 2017. Silently transformable: the many ways bacteria conceal their built-in capacity of genetic exchange. *Curr Genet* 63:451–455. <https://doi.org/10.1007/s00294-016-0663-6>.
37. van Opijnen T, Camilli A. 2012. A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res* 22:2541–2551. <https://doi.org/10.1101/gr.137430.112>.
38. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* 19:2308–2316. <https://doi.org/10.1101/gr.097097.109>.
39. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci U S A* 106:16422–16427. <https://doi.org/10.1073/pnas.0906627106>.
40. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JL. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6:279–289. <https://doi.org/10.1016/j.chom.2009.08.003>.
41. van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6:767–772. <https://doi.org/10.1038/nmeth.1377>.
42. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. 2020. A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* 21:526–540. <https://doi.org/10.1038/s41576-020-0244-x>.
43. Chao MC, Abel S, Davis BM, Waldor MK. 2016. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol* 14:119–128. <https://doi.org/10.1038/nrmicro.2015.7>.
44. Shames SR, Liu L, Havey JC, Schofield WB, Goodman AL, Roy CR. 2017. Multiple *Legionella pneumophila* effector virulence phenotypes revealed through high-throughput analysis of targeted mutant libraries. *Proc Natl Acad Sci U S A* 114:E10446–E10454. <https://doi.org/10.1073/pnas.1708553114>.
45. Park JM, Ghosh S, O'Connor TJ. 2020. Combinatorial selection in amoebal hosts drives the evolution of the human pathogen *Legionella pneumophila*. *Nat Microbiol* 5:599–511. <https://doi.org/10.1038/s41564-019-0663-7>.
46. Sexton JA, Vogel JP. 2004. Regulation of hypercompetence in *Legionella pneumophila*. *J Bacteriol* 186:3814–3825. <https://doi.org/10.1128/JB.186.12.3814-3825.2004>.
47. DeJesus MA, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, Ioerger TR. 2013. Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. *Bioinformatics* 29:695–703. <https://doi.org/10.1093/bioinformatics/btt043>.
48. DeJesus MA, Ioerger TR. 2013. A hidden Markov model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* 14:303. <https://doi.org/10.1186/1471-2105-14-303>.
49. Luo H, Lin Y, Gao F, Zhang C-T, Zhang R. 2014. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res* 42:D574–D580. <https://doi.org/10.1093/nar/gkt1131>.
50. O'Connor TJ, Adepoju Y, Boyd D, Isberg RR. 2011. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. *Proc Natl Acad Sci U S A* 108:14733–14740. <https://doi.org/10.1073/pnas.1111678108>.
51. Juan P-A, Attaiech L, Charpentier X. 2015. Natural transformation occurs independently of the essential actin-like MreB cytoskeleton in *Legionella pneumophila*. *Sci Rep* 5:16033. <https://doi.org/10.1038/srep16033>.
52. Charpentier X, Faucher SP, Kalachikov S, Shuman HA. 2008. Loss of RNase R induces competence development in *Legionella pneumophila*. *J Bacteriol* 190:8126–8136. <https://doi.org/10.1128/JB.01035-08>.
53. Brunel R, Charpentier X. 2016. Trans-translation is essential in the human pathogen *Legionella pneumophila*. *Sci Rep* 6:37935. <https://doi.org/10.1038/srep37935>.
54. Beaber JW, Hochhut B, Waldor MK. 2002. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* 184:4259–4269. <https://doi.org/10.1128/JB.184.15.4259-4269.2002>.
55. Sinha S, Mell JC, Redfield RJ. 2012. Seventeen Sxy-dependent cyclic AMP receptor protein site-regulated genes are needed for natural transformation in *Haemophilus influenzae*. *J Bacteriol* 194:5245–5254. <https://doi.org/10.1128/JB.00671-12>.
56. Ohnishi H, Mizunoe Y, Takade A, Tanaka Y, Miyamoto H, Harada M,

- Yoshida S. 2004. Legionella dumoffii DjlA, a member of the DnaJ family, is required for intracellular growth. *Infect Immun* 72:3592–3603. <https://doi.org/10.1128/IAI.72.6.3592-3603.2004>.
57. Pernestig AK, Meleforts O, Georgellis D. 2001. Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*. *J Biol Chem* 276:225–231. <https://doi.org/10.1074/jbc.M001550200>.
 58. Marutani M, Taguchi F, Ogawa Y, Hossain MM, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y. 2008. Gac two-component system in *Pseudomonas syringae* pv. tabaci is required for virulence but not for hypersensitive reaction. *Mol Genet Genomics* 279:313–322. <https://doi.org/10.1007/s00438-007-0309-y>.
 59. Hammer BK, Tateda ES, Swanson MS. 2002. A two-component regulator induces the transmission phenotype of stationary-phase *Legionella pneumophila*. *Mol Microbiol* 44:107–118. <https://doi.org/10.1046/j.1365-2958.2002.02884.x>.
 60. Gal-Mor O, Segal G. 2003. The *Legionella pneumophila* GacA homolog (LetA) is involved in the regulation of *icm* virulence genes and is required for intracellular multiplication in *Acanthamoeba castellanii*. *Microb Pathog* 34:187–194. [https://doi.org/10.1016/S0882-4010\(03\)00027-5](https://doi.org/10.1016/S0882-4010(03)00027-5).
 61. Lynch D, Fieser N, Glöggler K, Forsbach-Birk V, Marre R. 2003. The response regulator LetA regulates the stationary-phase stress response in *Legionella pneumophila* and is required for efficient infection of *Acanthamoeba castellanii*. *FEMS Microbiol Lett* 219:241–248. [https://doi.org/10.1016/S0378-1097\(03\)00050-8](https://doi.org/10.1016/S0378-1097(03)00050-8).
 62. Shi C, Forsbach-Birk V, Marre R, McNealy TL. 2006. The *Legionella pneumophila* global regulatory protein LetA affects DotA and Mip. *Int J Med Microbiol* 296:15–24. <https://doi.org/10.1016/j.ijmm.2005.09.003>.
 63. Bachman MA, Swanson MS. 2004. The LetE protein enhances expression of multiple LetA/LetS-dependent transmission traits by *Legionella pneumophila*. *Infect Immun* 72:3284–3293. <https://doi.org/10.1128/IAI.72.6.3284-3293.2004>.
 64. Molofsky AB, Swanson MS. 2003. *Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication. *Mol Microbiol* 50:445–461. <https://doi.org/10.1046/j.1365-2958.2003.03706.x>.
 65. Hofreuter D, Odenbreit S, Haas R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Mol Microbiol* 41:379–391. <https://doi.org/10.1046/j.1365-2958.2001.02502.x>.
 66. Jacobsen T, Bardiaux B, Francetic O, Izadi-Pruneyre N, Nilges M. 2020. Structure and function of minor pilins of type IV pili. *Med Microbiol Immunol* 209:301–308. <https://doi.org/10.1007/s00430-019-00642-5>.
 67. Giltner CL, Nguyen Y, Burrows LL. 2012. Type IV pilin proteins: versatile molecular modules. *Microbiol Mol Biol Rev* 76:740–772. <https://doi.org/10.1128/MMBR.00035-12>.
 68. Neuhaus A, Selvaraj M, Salzer R, Langer JD, Kruse K, Kirchner L, Sanders K, Daum B, Averhoff B, Gold VAM. 2020. Cryo-electron microscopy reveals two distinct type IV pili assembled by the same bacterium. 1. *Nat Commun* 11:2231. <https://doi.org/10.1038/s41467-020-15650-w>.
 69. Stone BJ, Abu Kwaik Y. 1998. Expression of multiple pili by *Legionella pneumophila*: identification and characterization of a type IV pilin gene and its role in adherence to mammalian and protozoan cells. *Infect Immun* 66:1768–1775. <https://doi.org/10.1128/IAI.66.4.1768-1775.1998>.
 70. Imam S, Chen Z, Roos DS, Pohlschröder M. 2011. Identification of surprisingly diverse type IV pili, across a broad range of Gram-positive bacteria. *PLoS One* 6:e28919. <https://doi.org/10.1371/journal.pone.0028919>.
 71. Korotkov KV, Hol WGJ. 2008. Structure of the GspK-GspL-GspJ complex from the enterotoxigenic *Escherichia coli* type 2 secretion system. *Nat Struct Mol Biol* 15:462–468. <https://doi.org/10.1038/nsmb.1426>.
 72. Cisneros DA, Bond PJ, Pugsley AP, Campos M, Francetic O. 2012. Minor pseudopilin self-assembly primes type II secretion pseudopilus elongation. *EMBO J* 31:1041–1053. <https://doi.org/10.1038/emboj.2011.454>.
 73. Karuppiiah V, Thistlethwaite A, Derrick JP. 2016. Structures of type IV pilins from *Thermus thermophilus* demonstrate similarities with type II secretion system pseudopilins. *J Struct Biol* 196:375–384. <https://doi.org/10.1016/j.jsb.2016.08.006>.
 74. Alm RA, Boderer AJ, Free PD, Mattick JS. 1996. Identification of a novel gene, pilZ, essential for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. *J Bacteriol* 178:46–53. <https://doi.org/10.1128/JB.178.1.46-53.1996>.
 75. Amikam D, Galperin MY. 2006. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* 22:3–6. <https://doi.org/10.1093/bioinformatics/bti739>.
 76. Merighi M, Lee VT, Hyodo M, Hayakawa Y, Lory S. 2007. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Mol Microbiol* 65:876–895. <https://doi.org/10.1111/j.1365-2958.2007.05817.x>.
 77. McCarthy Y, Ryan RP, O'Donovan K, He Y-Q, Jiang B-L, Feng J-X, Tang J-L, Dow JM. 2008. The role of PilZ domain proteins in the virulence of *Xanthomonas campestris* pv. *campestris*. *Mol Plant Pathol* 9:819–824. <https://doi.org/10.1111/j.1364-3703.2008.00495.x>.
 78. Guzzo CR, Salinas RK, Andrade MO, Farah CS. 2009. PILZ protein structure and interactions with PILB and the FIMX EAL domain: implications for control of type IV pilus biogenesis. *J Mol Biol* 393:848–866. <https://doi.org/10.1016/j.jmb.2009.07.065>.
 79. Ferrières L, Hémerly G, Nham T, Guérou A-M, Mazel D, Beloin C, Ghigo J-M. 2010. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J Bacteriol* 192:6418–6427. <https://doi.org/10.1128/JB.00621-10>.
 80. Kulasekara HD, Ventre I, Kulasekara BR, Lazdunski A, Filloux A, Lory S. 2005. A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol Microbiol* 55:368–380. <https://doi.org/10.1111/j.1365-2958.2004.04402.x>.
 81. Hervet E, Charpentier X, Vianney A, Lazzaroni J-C, Gilbert C, Atlan D, Doublet P. 2011. Protein kinase LegK2 is a type IV secretion system effector involved in endoplasmic reticulum recruitment and intracellular replication of *Legionella pneumophila*. *Infect Immun* 79:1936–1950. <https://doi.org/10.1128/IAI.00805-10>.
 82. Hardy L, Charpentier X. 2019. Querying *Legionella* genomes using transposition-sequencing. *Methods Mol Biol* 1921:107–122. https://doi.org/10.1007/978-1-4939-9048-1_7.
 83. Long JE, DeJesus M, Ward D, Baker RE, Ioerger T, Sassetti CM. 2015. Identifying essential genes in *Mycobacterium tuberculosis* by global phenotypic profiling. *Methods Mol Biol* 1279:79–95. https://doi.org/10.1007/978-1-4939-2398-4_6.
 84. DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR. 2015. TRANSIT—a software tool for Himar1 TnSeq analysis. *PLoS Comput Biol* 11:e1004401. <https://doi.org/10.1371/journal.pcbi.1004401>.
 85. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 86. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.
 87. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.

Transposon-insertion sequencing in a clinical isolate of *Legionella pneumophila* identifies essential genes and determinants of natural transformation

Léo Hardy, Pierre-Alexandre Juan, Bénédicte Coupat-Goutaland and Xavier Charpentier

Supplemental material

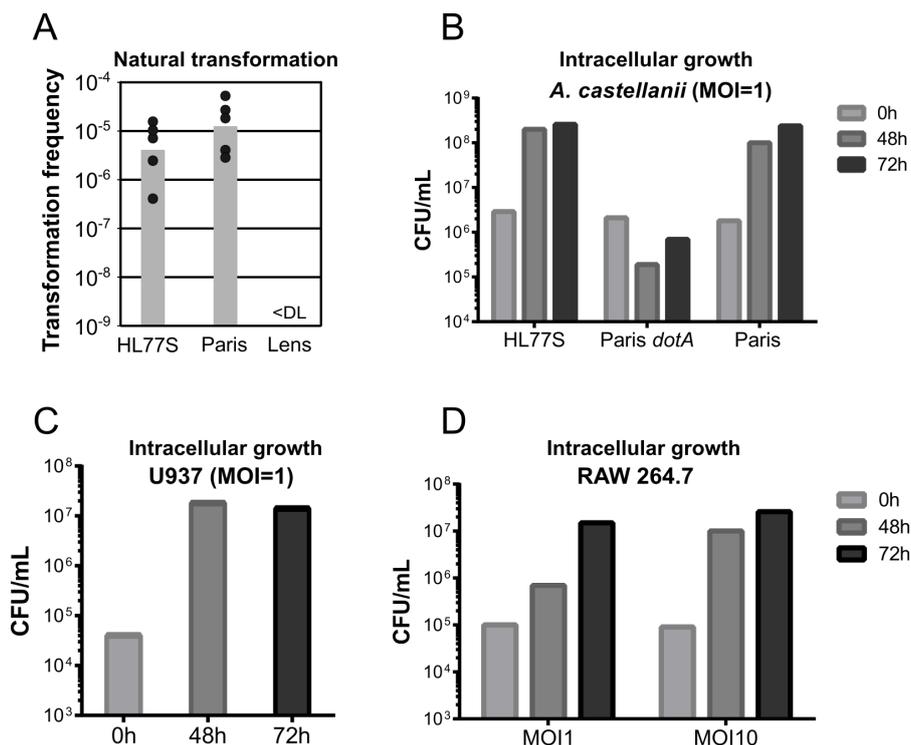


Figure S1. Phenotypic characterization of strain HL77S, a streptomycin-resistant mutant of the clinical isolate HL-0709-3014.

(A) Natural transformability of HL77S compared to the Paris and Lens strain. Transformation was tested by growing the strains in AYE at 30°C for 24h in the presence of 2 µg of transformation DNA consisting of a kanamycin resistance gene interrupting the *ihfB* gene. (B) Intracellular replication of HL77S in the amoeba *A. castellanii*. Cells were infected with a suspension of HL77S at a multiplicity of infection (MOI) of 1. At the initial time point, after 48h and 72h of culture at 30°C, colony-forming units are determined by plating on CYE medium. (C) Intracellular replication of HL77S in differentiated human monocytes of the U937 cell line. Cells were infected with a suspension of HL77S at a multiplicity of infection (MOI) of 1. At the initial time point, after 48h and 72h of culture at 30°C, colony-forming units are determined by plating on CYE. (D) Intracellular replication of HL77S in the murine macrophage-like cell line RAW 264.7. Cells were infected with a suspension of HL77S at a multiplicity of infection (MOI) of 1 or 10. At the initial time point, after 48h and 72h of culture at 30°C, colony-forming units are determined by plating on CYE. Data from panel A, B and C are representative of experiment performed at least three times.

Dataset S1. Tn-seq data for strain HL77S in Excel spreadsheet. Spreadsheet “Analysis – Summary” lists gene annotations and gene essentiality on 1st and 2nd isolation (NE, non-essential; E, essential; U, uncertain; GA, growth advantage; GD, growth disadvantage; S, too short), gene fitness value (log₂ fold-change 2nd vs 1st isolation) and gene role in natural transformation by the *ihfB::kan* and *legK2::kan* markers (log₂ fold-change between transformed and control populations). Following spreadsheets includes all raw data for the analysis displayed in the summary. Spreadsheet “Sample description” provides description of the samples and read mapping statistics.

Table S1. Strains, plasmids and oligonucleotides used in this study

| Strains | | |
|--|--|--|
| Name | Genotype | Source |
| Paris WT | Wild-type <i>Legionella pneumophila</i> | Paris Outbreak isolate CIP107629, CNR Lyon |
| Paris_S | Spontaneous streptomycin-resistant mutant of Paris; StrepR | This study |
| HL-0709-3014 | Wild-type <i>Legionella pneumophila</i> | Clinical isolate, CNR Lyon |
| HL77S | Spontaneous streptomycin-resistant mutant of HL-0709-3014; StrepR | This study |
| Paris <i>rocC_{TAA}</i> | Paris <i>rocC_{TAA}</i> (=lpp0148 _{TAA}); <i>rocC</i> allele with a premature stop codon TAA | (1) |
| Paris <i>rocC_{TAA} letS::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp1887::kan; KanR | This study |
| Paris <i>rocC_{TAA} letA::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp2699::kan; KanR | This study |
| Paris <i>rocC_{TAA} comM::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp2632::kan; KanR | This study |
| Paris <i>rocC_{TAA} comEC::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp0680::kan; KanR | This study |
| Paris <i>rocC_{TAA} pilE::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp0681::kan; KanR | This study |
| Paris <i>rocC_{TAA} comEA::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp0872::kan; KanR | This study |
| Paris <i>rocC_{TAA} pilA1::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp1890::kan; KanR | This study |
| Paris <i>rocC_{TAA} lpp1976-1977-1978::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp1976-1977-1978::kan; KanR | This study |
| Paris <i>rocC_{TAA} comF::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp2280::kan; KanR | This study |
| Paris <i>rocC_{TAA} pilA2::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp1889::kan; KanR | This study |
| Paris <i>rocC_{TAA} pilZ::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp1356::kan; KanR | This study |
| Paris <i>rocC_{TAA} lpp3030::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp3030::kan; KanR | This study |
| Paris <i>rocC_{TAA} djlA::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp2289::kan; KanR | This study |
| Paris <i>rocC_{TAA} lpp2632::kan</i> | Paris <i>rocC_{TAA}</i> ; lpp2632::kan; KanR | This study |
| Paris <i>letA::kan</i> | Paris lpp2699::kan; KanR | This study |
| Paris <i>dotA::kan</i> | Paris <i>dotA::kan</i> ; KanR | This study |
| MFDpir | <i>Escherichia coli</i> ; MG1655 RP4-2-Tc::[ΔMu1::aac(3)IV-ΔaphA-Δnic35-ΔMu2::zeo] ΔdapA::(erm-pir) ΔrecA | (2) |

| Plasmids | | | |
|-----------------|-----------------|---|---------------|
| Name | Genotype | Description | source |
| pGEM- | pGEM-T Easy:: | plasmid bearing the <i>ihfB</i> gene of <i>L. pneumophila</i> | (1) |

| | | | |
|------------------------|---|--|------------|
| ihfB::kan | <i>ihfB::nptII</i> ; AmpR, KanR | Paris interrupted by a kanamycin cassette ; Kan R | |
| pGEM-HYG1 | pGEM-T Easy::hygR | plasmid bearing an hygromycine cassette ; HygR | This study |
| pJET1.2- legk2::kan | pJET1.2 ; <i>legk2::kan</i> from HL77 ; AmpR, KanR | plasmid bearing the <i>legK2</i> gene of <i>L. pneumophila</i> HL77 interrupted by a kanamycin cassette ; Kan R | This study |
| pBT20 | Gmr; Himar1C9 transposon | Pir-dependent, mobilizable plasmid carrying the Mariner Himar1 transposon with gentamycin resistant gene | (3) |
| pMMB207C | RSF1010:: <i>lacI^q- tacp mobA</i> | cloning vector for expression genes under the P _{tac} promoter, derived from RSF1010 ; CmR | (4) |
| p0681F | pMMB207C ; lpp0681::FLAG | complementation plasmid for the lpp0681::kan mutant, expressing the lpp0681-FLAG (P _{ilE} - FLAG) fusion protein ; CmR | This study |
| p1890F | pMMB207C ; lpp1890::FLAG | complementation plasmid for the lpp1890::kan mutant, expressing the lpp1890-FLAG (P _{ilA2} - FLAG) fusion protein ; CmR | This study |

Oligonucleotides

| Name | sequence | use |
|---------------------------|--|---|
| LH1_lpp3030_P1 | ctgcatgggatattggtatcactgc | Forward primer to amplify a 2kB fragment upstream of lpp3030 |
| LH2_lpp3030_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCtggcgatgataacaagcgagagcc | Reverse primer to amplify a 2kB fragment upstream of lpp3030 |
| LH3_lpp3030_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCgaaaatctgccgatactgtttctgc | Forward primer to amplify a 2kB fragment downstream of lpp3030 |
| LH4_lpp3030_P4 | ggtccatctcattctccttaatcc | Reverse primer to amplify a 2kB fragment downstream of lpp3030 |
| LH5_lpp1887_P1 | gggtaatttctctgagacagtggagg | Forward primer to amplify a 2kB fragment upstream of letS |
| LH6_lpp1887_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCggcaggaataagagttagtcttagc | Reverse primer to amplify a 2kB fragment upstream of letS |
| LH7_lpp1887_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCctaagtctgataatttaaccggagcc | Forward primer to amplify a 2kB fragment downstream of letS |
| LH8_lpp1887_P4 | ggtggaggtaccatagttatgaccc | Reverse primer to amplify a 2kB fragment downstream of letS |
| LH9_lpp2699_P1 | ggtatctctgtctgacagtctaaacc | Forward primer to amplify a 2kB fragment upstream of letA |
| LH10_lpp2699_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCccccattctaaccaatgcatggtc | Reverse primer to amplify a 2kB fragment upstream of letA |
| LH11_lpp2699_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCcgattattgaacacccaatg | Forward primer to amplify a 2kB fragment downstream of letA |
| LH12_lpp2699_P4 | caggggaaatcaaagacattggcc | Reverse primer to amplify a 2kB fragment downstream of letA |
| LH21_lpp2632_P1 | cggcttctgtatattattgagagac | Forward primer to amplify a 2kB fragment upstream of lpp2632 |
| LH22_lpp2632_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCcgagacacactatctcgtatcatgc | Reverse primer to amplify a 2kB fragment upstream of lpp2632 |
| LH23_lpp2632_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCgtgcatactttgtcttgggaagac | Forward primer to amplify a 2kB fragment downstream of lpp2632 |
| LH24_lpp2632_P4 | gatggagaattttccgctcatcc | Reverse primer to amplify a 2kB fragment downstream of lpp2632 |
| LH29_lpp2289_P1 | caggaatgtcacactgaattttcc | Forward primer to amplify a 2kB fragment upstream of djIA |
| LH30_lpp2289_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCccccaccaggtgttattacaaag | Reverse primer to amplify a 2kB fragment upstream of djIA |
| LH31_lpp2289_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCgctcaaggtttaccgaagaatg | Forward primer to amplify a 2kB fragment downstream of djIA |
| LH32_lpp2289_P4 | cgagtcaggaagtgtttacagg | Reverse primer to amplify a 2kB fragment downstream of djIA |

| | | |
|---------------------------|---|---|
| rpsL_Fw | GCAGCTCCAGATGGCTCAATC | Forward primer to amplify a 2kB fragment downstream of rpsL |
| rpsL_Rv | CAACCATACATGTCCATATTGACCAC | Reverse primer to amplify a 2kB fragment upstream of rpsL |
| lpp0640_P1 | TCATCCAACTCATCTCGCAATCG | Forward primer to amplify a 2kB fragment upstream of comM |
| lpp0640_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCACTGCTGTTTCAGCAAGTCCT ACC | Reverse primer to amplify a 2kB fragment upstream of comM |
| lpp0640_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCGTCTGCTCGTGGCTACCATCGC | Forward primer to amplify a 2kB fragment downstream of comM |
| lpp0640_P4 | GCTGTAGGACAGCGGCTAACTTG | Reverse primer to amplify a 2kB fragment downstream of comM |
| lpp0681_P1 | ATGGGAGCTGGCGTAGATCCTG | Reverse primer to amplify a 2kB fragment downstream of pilE |
| lpp0681_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCGGCAATTGAAACCAGAATGC CC | Reverse primer to amplify a 2kB fragment downstream of pilE |
| lpp0681_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCAAACAACGCCGAACGGGTAC | Reverse primer to amplify a 2kB fragment downstream of pilE |
| lpp0681_P4 | CATTGCCATTGCGGGTATGAATG | Reverse primer to amplify a 2kB fragment downstream of pilE |
| lpp1889_P1 | GTGAACTGCAGCAAGCTCCATCC | Reverse primer to amplify a 2kB fragment downstream of pilA1 |
| lpp1889_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCGATGGCAACCAGAATCCCAA G | Reverse primer to amplify a 2kB fragment downstream of pilA1 |
| lpp1889_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCTGCAAACGCAGGTAATGGCAC | Reverse primer to amplify a 2kB fragment downstream of pilA1 |
| lpp1889_P4 | GCGAAATGGCCGTTACTGCTTG | Reverse primer to amplify a 2kB fragment downstream of pilA1 |
| lpp1890_P1 | GTCAGGTAATAACCGGGTTTGCC | Reverse primer to amplify a 2kB fragment downstream of pilA2 |
| lpp1890_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCGCGATTGCTGCCAAAATACCG | Reverse primer to amplify a 2kB fragment downstream of pilA2 |
| lpp1890_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCAATTGCTATTGGTGCGAACGG | Reverse primer to amplify a 2kB fragment downstream of pilA2 |
| lpp1890_P4 | CAGTCTGGTGGTGTGACCGCTG | Reverse primer to amplify a 2kB fragment downstream of pilA2 |
| lpp2280_P1 | CCGGCAACTGGAAAAGGGAG | Reverse primer to amplify a 2kB fragment downstream of comF |
| lpp2280_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCTCCATGCAATTAGAGCAAAC TGC | Reverse primer to amplify a 2kB fragment downstream of comF |
| lpp2280_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCAAGCCAACCTTGATGGAGAGC | Reverse primer to amplify a 2kB fragment downstream of comF |
| lpp2280_P4 | GCCCCAAGCTACAAATACCATAG | Reverse primer to amplify a 2kB fragment downstream of comF |
| lpp1976-1978_P1 | CAATCAATCAAACCTCTCTCAAGAACG | Reverse primer to amplify a 2kB fragment downstream of the operon of unknown function |
| lpp1976-1978_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC ACAATCGCCAGATAAGATGATTGATTA GATCTC | Reverse primer to amplify a 2kB fragment downstream of the operon of unknown function |
| lpp1976-1978_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCGGTGAAGGAAGGGGTAAGCA GC | Reverse primer to amplify a 2kB fragment downstream of the operon of unknown function |
| lpp1976-1978_P4 | GAAAAATGAATGGGAGCTTCTGG | Reverse primer to amplify a 2kB fragment downstream of the operon of unknown function |
| lpp0872_P1 | GTAGTGGGAGGAAGCTTAGG | Reverse primer to amplify a 2kB fragment downstream of comEA |
| lpp0872_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTAC | Reverse primer to amplify a 2kB |

| | | |
|----------------------|---|---|
| | ACAATCGAGGCTCGTTCTCAGCTTGA GAAGAG | fragment downstream of comEA |
| lpp0872_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACC ATGGCGCGGAAGTCAAGGGTATAGG | Reverse primer to amplify a 2kB fragment downstream of comEA |
| lpp0872_P4 | GCGACAGGGCTAACTGTAAC | Reverse primer to amplify a 2kB fragment downstream of comEA |
| olj376 | GTGACTGGAGTTCAGACGTGTGCTCT TCCGATCTGGGGGGGGGGGGGGGG | Primer for Tn-seq to amplify transposon junction and add polyC-tail during PCR1 |
| pBT20-PCR1 | Biotin- TCGTATAATGTGTGGAATTGTGAGCGG | Biotinylated primer for Tn-seq to amplify transposon junction during PCR1 |
| pBT20-PCR2 | AATGATACGGCGACCACCGAGATCTAC ACTCTTTGGACTCTAGAGGATCACCCA GCTTTCTTG | Primer for Tn-seq to amplify transposon junction during PCR2 |
| TdT_Index_X | CAAGCAGAAGACGGCATAACGAGATXX XXXXGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCT | HPLC-purified primer for Tn-seq to amplify transposon junction and add index for multiplexing during PCR2 |
| Read1TnLp | CTAGAGACCGGGACTTATCAGCCAA CCTGTTA | HPLC-purified custom sequencing primer for Tn-seq library sequencing |
| lpp0681-F | CCGGGGATCCGCAAATTCAATAGAGG ATACCCAAATG | cloning of lpp0681 in pMM207C |
| lpp0681F-R | AGCCAAGCTTTTACTTGTGTCATCGTCGT CCTTGTAATCAGCGCTACCGGGATTCC AGCATTCTGGTTG | cloning of lpp0681 in pMM207C |
| lpp1890-F | CCGGGGATCCGTTAACTATGGAGATGG TCATGAGAC | cloning of lpp1890 in pMM207C |
| lpp1890F-R | AGCCAAGCTTTTACTTGTGTCATCGTCGT CCTTGTAATCAGCGCTTGGTCTGCAGC TGGCAGGTCG | cloning of lpp1890 in pMM207C |
| mreBseqF | TCCGGTTAAAAGCAGTTGTCTGG | Forward primer to amplify <i>mreB</i> , chromosomal control for uptake test |
| mreBseqR | CCCAGAGAACTGTGTCCGCCC | Reverse primer to amplify <i>mreB</i> , chromosomal control for uptake test |
| M13F(-47) | CGCCAGGGTTTTCCAGTCACGAC | Forward primer to amplify a 1657 pb fragment of the pGEM-HYG1 |
| M13R(-48) | AGCGGATAACAATTTACACAGGA | Reverse primer to amplify a 1657 pb fragment of the pGEM-HYG1 |

References

1. Juan P-A, Attaiech L, Charpentier X. 2015. Natural transformation occurs independently of the essential actin-like MreB cytoskeleton in *Legionella pneumophila*. *Sci Rep* 5:16033.
2. Ferrières L, Hémerly G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo J-M. 2010. Silent Mischief: Bacteriophage Mu Insertions Contaminate Products of *Escherichia coli* Random Mutagenesis Performed Using Suicidal Transposon Delivery Plasmids Mobilized by Broad-Host-Range RP4 Conjugative Machinery. *J Bacteriol* 192:6418–6427.
3. Kulasekara HD, Ventre I, Kulasekara BR, Lazdunski A, Filloux A, Lory S. 2005. A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol Microbiol* 55:368–380.
4. Segal G, Shuman HA. 1998. Intracellular multiplication and human macrophage killing by *Legionella pneumophila* are inhibited by conjugal components of IncQ plasmid RSF1010. *Mol Microbiol* 30:197–208.

B. Résultat complémentaire : Recherche de mutants hypercompétents dans la banque de mutant de *L. pneumophila*

Un mutant hypercompétent est un mutant transformable expérimentalement dans n'importe quelles conditions de culture et le plus souvent à n'importe quelle phase de la croissance. Une souche *WT* de *L. pneumophila* Paris est compétente et transformable entre $DO_{600} \sim 2$ et ~ 4 en culture liquide AYE à 30°C (Attaiech et al., 2016). Les mutants $\Delta rocC$ (anciennement $\Delta proQ$), $\Delta rocR$ et $\Delta comR$, eux sont hypercompétents c'est-à-dire qu'ils transforment aussi bien en culture liquide AYE qu'en culture solide CYE à 30°C comme à 37°C. Le quatrième et dernier mutant hypercompétent connu de *L. pneumophila*, le mutant Δrnr du gène codant pour la RNase R est un peu particulier, car il transforme, certes en culture solide et liquide, mais uniquement à 30°C.

En s'inspirant des travaux de Sexton and Vogel en 2004 qui avaient permis d'identifier les mutants $\Delta rocC$ et $\Delta comR$ par crible génétique chez la souche de *L. pneumophila* AA100 grâce à l'utilisation d'une banque non saturante de mutants, nous avons entrepris d'utiliser notre banque de mutant de Tn-seq dans le but d'identifier de nouveaux possibles répresseurs de la compétence développant ce phénotype d'hypercompétence. Pour cela nous avons réalisé la transformation de la banque de mutant dans des conditions où normalement la souche HL77S utilisée pour construire la banque n'est pas transformable. À savoir, nous avons cultivé la banque de mutants en présence d'ADNt en milieu solide CYE à 37°C. Après 24H d'incubation, les tests de transformation ont été récoltés et étalés sur milieu sélectif pour sélectionner les potentiels transformants qui ne peuvent qu'être nécessairement des transformants de mutants hypercompétents dans ces conditions de transformation.

Par ce principe des clones ont été isolés et le gène muté a été identifié par la méthode ayant permis d'identifier les sites d'insertion des transposons par Tn-seq, mais cette fois-ci sur une cellule unique.

Sur ces principes, il en résulte qu'aucun autre mutant hypercompétent n'a été identifié que ceux déjà découverts. Plus précisément, nous n'avons identifié que des mutants délétés par l'insertion du transposon au niveau du gène *comR* ou *rocC*. Dans ces conditions de transformation, il est normal de ne pas trouver de mutants *rnr* (qui ne seraient révélés qu'à 30°C). Par contre, dans ces conditions, nous aurions dû sélectionner des mutants de *rocR*. Après analyse de la cartographie des sites d'insertion de la banque en milieu CYE, il se trouve qu'aucun site d'insertion de transposon n'a été cartographié dans le gène codant pour cet ARNnc. Cela s'explique par le fait que *rocR* est un tout petit gène de 66 pb et il n'a donc possiblement pas été ciblé lors de la mutagenèse initiale.

En conclusion, en identifiant des répresseurs de la compétence déjà connus, ces données montrent que la banque de mutants de HL77S que nous avons générée est utilisable pour trouver de possibles

répresseurs d'un mécanisme particulier. Lorsque ce principe est utilisé pour trouver de potentiels nouveaux répresseurs de la compétence caractérisés par un phénotype d'hypercompétence, aucun nouveau répresseur n'a été retrouvé. La banque de mutant que nous avons générée étant très dense, ceci suggère potentiellement que les seuls répresseurs du système de compétence de *L. pneumophila* sont ceux déjà découverts. Néanmoins, on ne peut exclure qu'il puisse exister d'autres petits gènes codant pour des répresseurs non ciblés par la mutagenèse.

III. Caractérisation d'un nouvel acteur impliqué dans la transformation naturelle des bactéries gram négatives

L'analyse génétique exhaustive de la transformation naturelle par Tn-seq chez *L. pneumophila* m'a permis d'identifier un nouvel acteur : YraN, conservé chez la majorité des espèces gram négatives et jusqu'alors de fonction inconnue. Dans cette partie, écrite sous la forme d'un article en préparation, sont décrites les étapes ayant permis la caractérisation de ce nouvel acteur au cours de la transformation naturelle des bactéries gram négatives.

A. Publication

1
2
3
4
5
6
7
8
9
10
11
12
13

YraN is a highly conserved nuclease involved in the processing of the recombination D-loop during natural transformation

Léo Hardy^a, Violette Morales^{b,c}, Patrice Polard^{b,c} and Xavier Charpentier^{a,*}

^a CIRI, Centre International de Recherche en Infectiologie, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Normale Supérieure de Lyon, Univ Lyon, 69100, Villeurbanne, France. Team Horigene.

^b Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Integrative (CBI), Centre National de la Recherche Scientifique (CNRS), 31062 Toulouse, France.

^c Université de Toulouse, Université Paul Sabatier, 31062 Toulouse, France.

* Correspondence to :

xavier.charpentier@univ-lyon1.fr | @labXC

Running title: YraN, a nuclease involved in natural transformation

Abstract

Natural transformation is a major mechanism of horizontal gene transfer, resulting from the uptake of DNA and its stable integration in the bacterial chromosome. Exogenous DNA follows a pathway involving a series of genes required for its the capture, internalisation, protection and recombination with the chromosome. Most of these genes were identified through the isolation of transformation-defective mutants and/or on the basis of their expression preceding natural transformation. Yet, genes required for key steps of the pathway remain elusive. We sought to identify any missing component by re-analysing Tn-seq data obtained in two distantly-related transformable Gram-negative species, the human pathogen *Legionella pneumophila* and the cyanobacterium *Synechococcus elongatus*. We identified *yraN*, a highly conserved gene of unknown function which belongs to the core genetic pathway of natural transformation. We provide evidence that YraN is a nuclease associated to the ComM helicase, which cooperate to promote the recombination of transforming DNA. We propose a model in which YraN cleaves the displaced strand of the D-loop formed by the invasion of the transforming DNA in the parental DNA strands. The identification of this broadly conserved nuclease supports the hypothesis that most bacteria possess a conserved pathway for the transport and recombination of exogenous DNA.

32 **Introduction**

33

34 One of the remarkable features of bacteria is their ability to naturally undergo genetic transformation.
35 This property is linked to the unique ability of bacteria to import exogenous DNA and integrate it in
36 their chromosome by homologous recombination (1). Despite major differences in their cell wall
37 architecture, the process is generally conserved between monoderms (bacteria with a single cell
38 membrane, like most Gram-positive) and diderms (bacteria with an outer membrane like most Gram-
39 negative) bacteria (2). Exogenous DNA is captured by an extracellular type IV filament (3, 4). While
40 in diderms, the Type IV filament also serves other functions such as motility and adhesion (5), the
41 competence Type IV filament of diderms is so far dedicated to the transformation pathway. In
42 diderms, retraction of the filament conveys DNA through the outer membrane by passing through the
43 PilQ secretin (4). Retraction of the competence pilus also occurs in the monoderm *Streptococcus*
44 *pneumoniae*, and is essential for natural transformation (6). In the periplasmic space, the ComEA
45 DNA receptor would function as a ratchet for double-strand DNA (dsDNA) (7, 8). In a process not
46 understood, the dsDNA is converted into single strand DNA (ssDNA) and transported across the
47 cytoplasmic membrane, presumably through a channel formed by the ComEC protein (9). In diderms,
48 the ComFC protein is also required for this transport stage, yet the function of this protein remains
49 elusive (10–12) Upon entry in the cytoplasm, the ssDNA is covered by the ssDNA-binding protein
50 DprA to form a nucleofilament which both protects the incoming DNA from the action of nuclease
51 but also promotes RecA-dependent recombination with the chromosome (13, 14). RecA-dependent
52 strand invasion displaces the homologous parental strand, forming a displacement (D-loop) onto
53 which the ComM helicase is recruited (15). The helicase activity of ComM is proposed to favor
54 integration of transforming DNA by promoting extension of the D-loop formed by the invasion of the
55 transforming ssDNA in the chromosomal dsDNA (16). In many bacteria, the expression of the genes
56 involved in this DNA uptake and recombination pathway is concerted, occurring in a physiological
57 state known as competence, and controlled by divergent regulatory pathway (1).

58 All proteins specific to the pathway have been identified by mutants defective for transformation or
59 by their observed co-regulation in the state of competence-induced (11, 17–20). Yet as, several key
60 steps of the process remain unresolved (ssDNA conversion, energy, mechanism of D-loop resolution)
61 it remain opens whether some players could not be identified through these approaches, either because
62 the encoding genes are small (and thus evade mutagenesis) or not part of the competence regulon.
63 Here, we sought to identify any other highly conserved genes involved in this pathway, through the
64 reanalysis of comprehensive genetic screen for competence development and which were based on
65 the ability of bacteria to capture exogenous DNA by natural transformation.

66 Results

67 YraN, a protein of unknown function is required for natural transformation in diverse 68 Gram-negative bacteria

69 In order to identify conserved genes involved in natural transformation, we compared the results of
70 transposon-insertion sequencing (Tn-seq) screen conducted in two Gram-negative bacteria, the
71 human pathogen *Legionella pneumophila* (21) and the cyanobacterium *Synechococcus elongatus*
72 (22). Both screens were initially designed to identify all genes required for competence (gene
73 regulation) and natural transformation (21, 22). Only 13 genes were found to be required for natural
74 transformation in both species (Table S1). As expected, no genes correspond to regulatory elements,
75 as the regulation of competence largely differs between the two organisms. Consistent with the role
76 of Type IV pilus in natural transformation, 7 genes encode the core components of Type IV pilus,
77 including the components of the basal body (PilM, PilN, PilO, PilP and PilC), the PilB ATPase and
78 the PilQ secretin. 5 other genes encode the highly conserved known players of the DNA transport and
79 processing pathway (ComEC, ComEA, ComFC, DprA and ComM). The last gene, annotated as *yraN*,
80 encodes a protein of unknown function, belonging to the uncharacterized protein family UPF0102.
81 Although mutants in the above genes appear equally defective in the Tn-seq screen, inactivation of
82 some of them are known to lead to transformation defects whose large amplitude further reflects their
83 specific role. For instance, the loss of the putative DNA import channel ComEC is known to
84 completely abrogates transformability in all tested species (2). In contrast, the loss of the ComM
85 helicase reduces the transformation frequency of single nucleotide polymorphism (SNP) by ~7 to ~15
86 fold in *Acinetobacter baylyi* and *Vibrio cholerae*, respectively (16).

| Gene name | COG | Annotation | Ratio transformed vs control populations (log ₂ fold change) | |
|--------------|---------|---|---|---------------------|
| | | | <i>L. pneumophila</i> | <i>S. elongatus</i> |
| <i>comM</i> | COG0606 | ComM DNA helicase | -3.3 | -1.4 |
| <i>comEC</i> | COG0658 | DNA transport transmembrane protein ComEC | -2.9 | -2.4 |
| <i>comEA</i> | COG1555 | Competence protein ComEA | -3.5 | -2.5 |
| <i>pilM</i> | COG4972 | Tfp pilus assembly protein PilM | -3.0 | -2.9 |
| <i>pilN</i> | COG3166 | Tfp pilus assembly protein PilN | -3.6 | -2.8 |
| <i>pilO</i> | COG3167 | Tfp pilus assembly protein PilO | -2.7 | -2.6 |
| <i>pilP</i> | COG3168 | Tfp pilus assembly protein PilP | -2.6 | -2.4 |
| <i>pilQ</i> | COG4796 | Tfp pilus assembly protein PilQ | -3.6 | -2.7 |
| <i>pilB</i> | COG2804 | Tfp pilus assembly protein PilB | -3.7 | -3.1 |
| <i>pilC</i> | COG1459 | Tfp pilus assembly protein PilC | -3.3 | -3.1 |
| <i>comFC</i> | COG1040 | Competence protein ComFC | -3.0 | -2.0 |
| <i>dprA</i> | COG0758 | DNA protection and recombination DprA | -2.7 | -2.3 |
| <i>yraN</i> | COG0792 | UPF0102 protein, unknown function | -2.8 | -1.0 |

Table1. Genes required for competence and natural transformation in both *L. pneumophila* and *S. elongatus*.

Data were retrieved from previous studies (21, 22). CDS were consistently annotated using eggNOG-mapper (23) and orthologous genes in *L. pneumophila* and *S. elongatus* were matched based on gene name and/or COG when available. Values correspond to the log₂ fold change (log₂FC) in abundance of insertions in the corresponding genes between the population obtained following transformation and the control, non-transformed population. Data were sorted and filtered for log₂FC < -1 and with P-value < 0.05.

87
88 We thus sought to test the contribution of YraN in natural transformation in three distinct organisms,
89 *L. pneumophila*, *Acinetobacter nosocomialis* and *A. baylyi*. In all species, inactivation of *yraN* leads
90 to a 7- to 40-fold reduction in the transformation efficiency using DNA with homology to the
91 chromosome and carrying a selectable SNP (Figure 1A). Similar results were obtained using
92 transforming DNA carrying a heterologous sequence conferring antibiotic resistance and flanked by
93 regions with homology to the chromosome (Figure S1). In contrast, inactivation of *yraN* had no
94 impact when using a replicative plasmid DNA whose natural transformation does not involve
95 recombination with the chromosome (Figure S2A). Similarly, conjugative transfer of the p45
96 integrative conjugative element and its site-specific recombination (24) in the chromosome of *L.*
97 *pneumophila* is insensitive to inactivation of *yraN* (Figure S2B). This suggested that *yraN* is involved
98 in homologous recombination. Yet, in contrast to the *recA* mutant, *yraN* mutants of *L. pneumophila*
99 and *A. baylyi* are not defective for the repair of DNA damage caused by UV irradiation (Figure S3).
100 Submission of the *L. pneumophila* YraN protein-coding gene to the STRING database returned a
101 network of co-occurrence with genes encoding ComEC, DprA, ComFC and with the highest
102 confidence score with the ComM helicase (Figure S4). Indeed, in all three tested species, inactivation
103 of *comM* leads to a transformation defect of SNP that is similar to that of the *yraN* mutants (Figure
104 1A). Moreover, these transformation defects are not additive, as a *yraN/comM* mutants do not show
105 stronger transformation defects than the single mutants (Figure 1A). This clearly suggested that YraN
106 and ComM are involved in the same recombination pathway.

107 The functional association of ComM and YraN is further supported by coevolution analyses which
108 predicted the direct interaction of these proteins (25, 26) as well as the interaction between ComM
109 and DprA (25). We thus tested these direct interactions using the bacterial two-hybrid assay based on
110 the functional reconstitution of the *Bordetella pertusis* adenylate cyclase from its T18 and T25
111 domains fused to the bait and prey proteins (27). We found that, as expected, both DprA and RecA
112 could self-interact and interact with one another (Figure 1B). While we detected no interaction
113 between RecA and ComM, we did observe the predicted interaction of ComM and DprA, but only
114 when ComM is fused to T18 (Figure 1B). Fusing the larger T25 part to the C-terminus of ComM may
115 make it not functional, as previously reported with fusion to sfGFP (16). We found that YraN could
116 also interact with the ComM fusion which interacts with DprA but also with both T18 and T25 fusions
117 of DprA (Figure 1B). Taken together with genetic analyses, protein association networks and co-
118 evolution analyses, this experimental interaction assay indicates that YraN is a component of the
119 pathway dedicated to the recombination of single-strand transforming DNA.

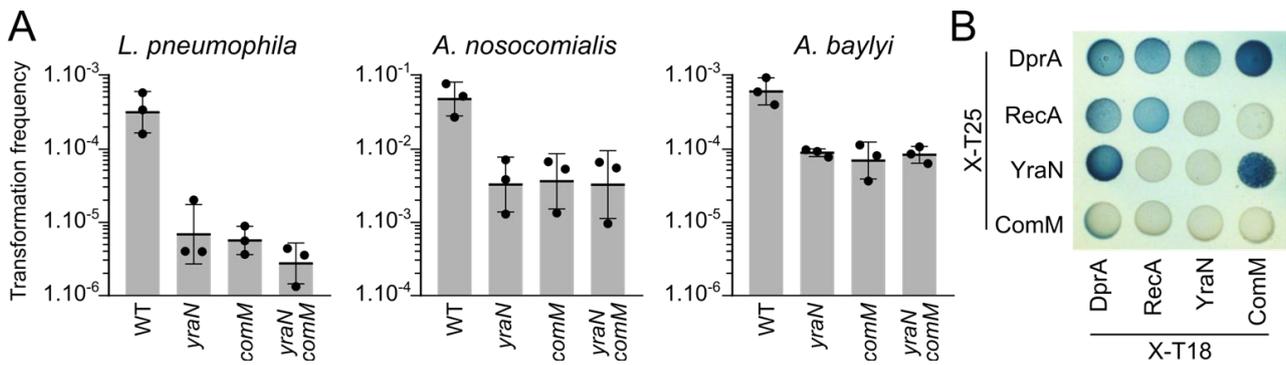


Figure 1. YraN, a protein of unknown function is required for natural transformation.

A. Efficiency of natural transformation of a SNP conferring resistance to streptomycin (*L. pneumophila*, Lp) or rifampicin (*A. nosocomialis*, An and *A. baylyi*, Ab). Wild type (WT) and mutants of the *yraN* and *comM* genes, were exposed to transforming DNA consisting of a PCR product of the *rpsL* allele for Lp or *rpoB* allele for An and Ab. Transformation frequencies represent the ratio of colony-forming units on selective (containing streptomycin or rifampicin) over the total CFU count on non-selective plates (CYE for Lp, LB for An and Ab). Data represent the geometric average of three independent experiments. B. Bacterial two-hybrid assay to detect interaction between the *L. pneumophila* YraN, DprA, RecA and ComM proteins. Each of these protein was expressed in *E. coli* as a fusion with their C-terminus fused to the T18 of T25 domains (X-T18 or X-T25) of the adenylate cyclase of *B. pertusis*. Protein interactions reconstitute of the adenylate cyclase, to produce cAMP, a cofactor required for the expression of the *lacZ* gene. Expression of LacZ is detected using the chromogenic beta-galactosidase substrate X-Gal. Strains expressing combinations of the fusions were spotted on LB plates containing X-Gal and IPTG. Blue coloration indicate interaction between the tested proteins.

120
121

122 **YraN belongs to the PD-(D/E)XK superfamily of phosphodiesterase, with a nuclease**
123 **activity required for its biological function**

124 YraN is a small protein of 118 amino acids (in *L. pneumophila*) which consists of the single domain
125 UPF0102. This domain adopts a common fold consisting of four β -sheets sandwiched between two
126 α -helices which form the active site of the large superfamily of PD-(D/E)XK phosphodiesterases (28,
127 29). Enzymes belonging to this family are very diverse, and while retaining the same structural
128 scaffold for the phosphodiesterase activity, show little sequence conservation outside of the residues
129 involved in the active site. They are found in prokaryotes, eukaryotes and archaea, performing
130 important functions related to nucleic acid maintenance (restriction enzymes, tRNA splicing
131 endonucleases, holliday junction resolvase, transposase) (28, 29). Many display variations in the PD-
132 (D/E)XK structural domain, and/or carry additional domains (link to UPF0102). In contrast, the
133 structure of YraN, predicted by alphafold2 (30), shows that it solely consist of a minimal PD-
134 (D/E)XK structural domain with the canonical signature of the phosphodiesterase active site (Figure
135 2A). On the *L. pneumophila* YraN, the active site is thus composed by D₃₇, E₅₀ and R₅₂ (Figure 2A).
136 In order to test its enzymatic activity, the *L. pneumophila* YraN protein and the E50A mutant YraN^{E50A}
137 were purified to homogeneity (Figure 2B). We found that YraN displays a nicking activity on double-
138 strand DNA, relaxing supercoiled circular DNA. As anticipated for this family of phosphodiesterase,
139 this single strand endonuclease activity requires the residue E50 (Figure 2C). Importantly, YraN also
140 acts on single-strand DNA, with no or highly relaxed site specificity (Figure 2D). In all, we found

141 that YraN is an endonuclease capable of cleaving single-strand DNA. We thus tested whether the
 142 enzymatic activity of YraN is required for its function in natural transformation. The wild-type YraN,
 143 expressed ectopically on the *L. pneumophila* native large plasmid pLPP, could restore the loss of
 144 transformability caused by the deletion of the chromosomal *yraN* gene (Figure 2E). In contrast,
 145 mutants of the active site YraN^{E50A} and YraN^{E37A} could not restore transformability (Figure 2E). This
 146 shows that the endonuclease activity of YraN is central to its contribution to the recombination of the
 147 transforming DNA with the chromosome.

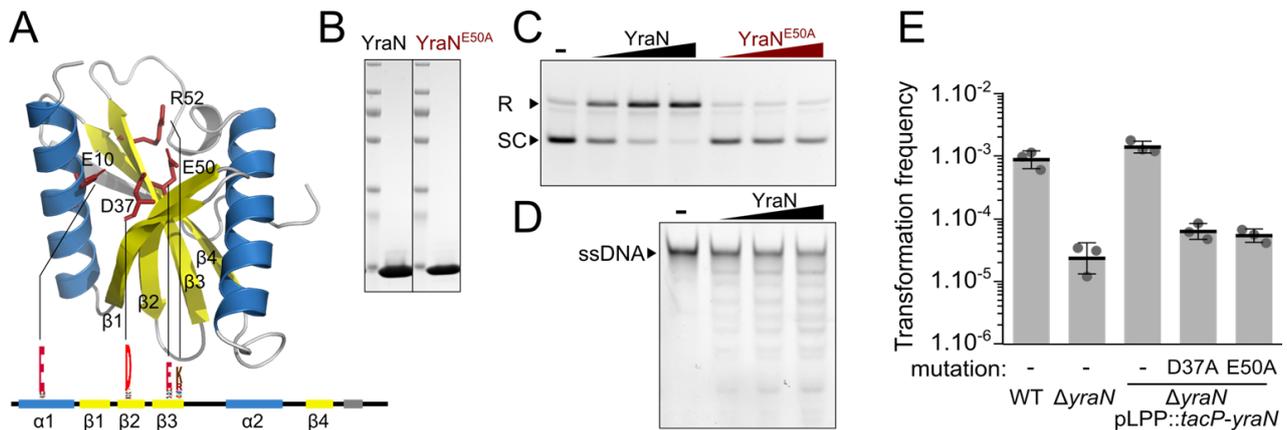


Figure 2. YraN displays endonuclease activity required for its in vivo biological function.

A. Predicted structure of the *L. pneumophila* YraN using alphafold2. YraN displays the typical fold of PD-(D/E)XK phosphodiesterase consisting of four beta strands (yellow) sandwiched by who alpha helices (blue). The linear representation of YraN shows the highly conserved residues (red) indicates as weblogo of the UPF0102 domain. The same residues are highlighted in the 3D structure. B. The *L. pneumophila* YraN and E50A mutant protein were expressed in *E. coli* and purified to homogeneity. C. Agarose gel electrophoresis of plasmid DNA incubated with increasing amount of YraN and the E50A mutant. Incubation with YraN, but not with the mutant E50A, produce a shift of the supercoiled (SC) fraction to the relaxed (R) form, indicative of ssDNA phosphodiesterase activity. D. Acrylamide gel electrophoresis of ssDNA incubated with increasing amount of YraN. The ladder below the main ssDNA band reveals endonuclease activity. E. The endonuclease activity of YraN is essential to promote recombination of transforming DNA. Efficiency of natural transformation of a SNP conferring resistance to streptomycin. Wild type (WT), mutant of the *yraN* gene, or expressing *yraN* using a *tac* promoter from endogenous plasmid pLPP were exposed to transforming DNA consisting of a PCR product of the *rpsL* allele. Transformation frequencies represent the ratio of colony-forming units on selective plates over the total CFU count on non-selective plates.

148

149 YraN promotes D-loop extension in association with the ComM helicase

150 The functional association of YraN and ComM suggests that YraN participates in the efficient
 151 integration of transforming DNA promoted by ComM. ComM was previously proposed to function
 152 as a DNA helicase, recruited at the D-loop formed by the invasion of the single-strand transforming
 153 DNA in the chromosome (15, 16). The helicase activity of ComM is thought to allows extension of
 154 the D-loop, which *in vivo* translates into the efficient comigration of linked genetic markers (16). To
 155 test the hypothesis that YraN might assist ComM in D-loop extension, we determined the efficiency
 156 of integration of genetic markers of separated by increasing distances. An *A. nosocomialis* strain
 157 carrying an integrated tetracycline resistance gene followed by the *lacZ* gene was exposed to
 158 transforming DNA encompassing the tetracycline resistance, but replaced by a kanamycin resistance

159 gene, and with a 7-bp deletion in the *lacZ* gene at increasing distance from the kanamycin resistance
160 marker (Figure 3A). We then determined comigration efficiency (*kan^R lacZ* CFU vs all *kan^R* CFU)
161 as a function of the distance separating the markers. In the wild type background, the comigration
162 efficiency decreases with the distance separating the markers, from nearly 1 at 200 bp distance down
163 to 0.7 at 3,000 bp (Figure 3B). While comigration at 200 bp distance is unchanged in a ComM-
164 deficient strain, it rapidly decreases with increasing distance, down to 0.4 at 3,000 bp (Figure 3B).
165 This confirms previous results obtained in *V. cholerae* and *A. baylyi* (16). The *yraN* mutant is also
166 severely affected for comigration at 3,000 bp, similarly to the *comM* mutant, suggesting that *yraN* is
167 required for the ComM-dependent D-loop extension. Yet, at 1,500 bp, the *yraN* mutant is less
168 defective than the *comM* mutant, suggesting that YraN facilitates ComM-dependent D-loop extension
169 over long distance. Consistent with results from transformation assays, the deficiency in comigration
170 of the *comM* and *yraN* mutant does not seem additive. In order to ascertain the role of ComM in
171 facilitating D-loop extension and thus integration of larger homologous sequence, we sought to
172 directly detect the recombination tracts resulting from natural transformation. An *A. baylyi* strain was
173 provided with a linear 11 kb-long DNA fragment encompassing the *rpoB* gene and obtained from a
174 rifampicin-resistant isolate of *A. nosocomialis*. Gene synteny is maintained in this region which
175 contains essential genes involved in transcription and translation. However, the sequence differs by
176 1081 SNPs which can be used to identify the recombination tracts of the *A. baylyi* transformants and
177 resulting from the acquisition of the sequence of *A. nosocomialis* (Figure 3C and 3D). We defined as
178 recombination tracts, segments of DNA with at least 2 SNPs converted to the donor variant and
179 separated by more than 3 non-converted SNPs. We found that transformants of the wild-type strain
180 of *A. baylyi*, present largely continuous recombination tracts and ranging from 1.1 kb to up to 9.8 kb,
181 with a mean length of 6.17 kb (Figure 3E). Mutants of *yraN* and *comM* show tracts ranging from 8
182 kb and down to 9 bp, with a mean length significantly shorter than for the wild-type (Mann–Whitney
183 U test, $p < 0.01$) and of 4.04 and 4.05 kb respectively. This confirms that both *yraN* or *comM* are
184 needed to maximize integration of transforming DNA. Interestingly, the double mutant *yraN/comM*
185 shows even shorter tracts than single mutants (Mann–Whitney U test, $p < 0.01$), with a mean size of
186 less than 2.9 kb, less than half the size as in the wild-type (Figure 3E). This analysis, of higher
187 resolution than the comigration or transformation assays, thus reveals that ComM and YraN act
188 synergistically to promote D-loop extension and the generation of larger recombination tracts.

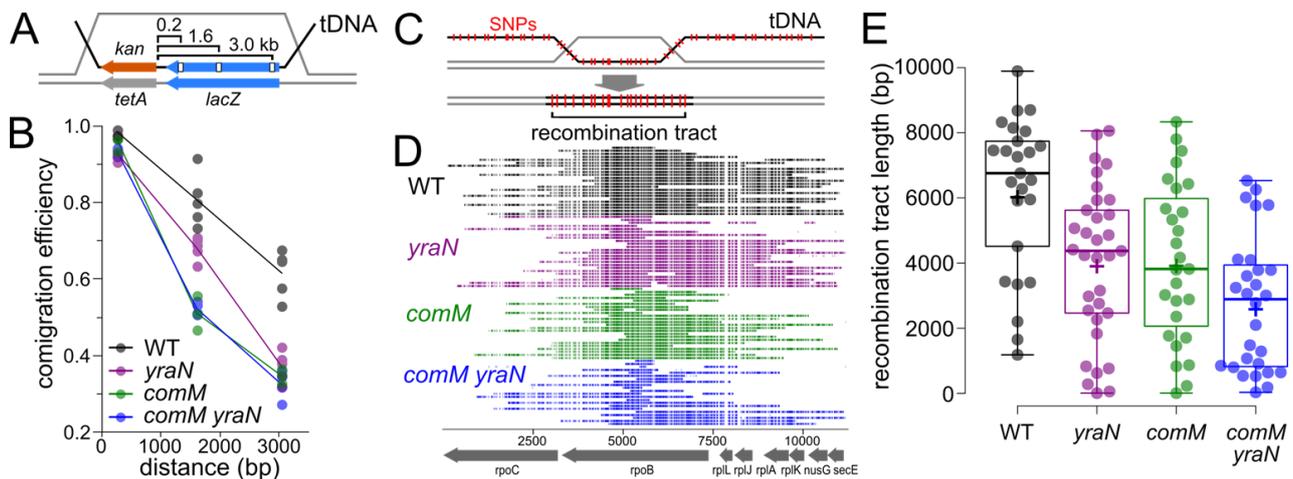


Figure 3. YraN promotes D-loop extension in association with the ComM helicase.

A. Principle of the *in vivo* branch migration assay based on marker comigration. An *A. nosocomialis* strain carrying an integrated tetracycline resistance gene followed by the *lacZ* gene is exposed to transforming DNA (tDNA) encompassing the tetracycline resistance, but replaced by a kanamycin resistance gene, and with a 7-bp deletion in the *lacZ* gene at a distance of 0.2, 1.6 or 3 kb from the end of the kanamycin marker. B. Comigration efficiency ($\text{kan}^R \text{ lacZ}$ CFUs vs all kan^R CFUs) as a function of the distance separating the markers in WT and mutants of *A. nosocomialis*. C. Schematic representation of the assay to detect recombination tracts. *A. baylyi* is exposed to an 11-kb PCR fragment from the *rpoB* region and amplified from *A. nosocomialis* resistant to rifampicin. Over a 1,000 SNPs differentiate the sequence from *A. baylyi* and *A. nosocomialis*. Sequencing of the transformant is thus used to determine the acquired SNPs, defining recombination tracts. D. Graphical representation of the acquired DNA fragments by transformants of *A. baylyi* when exposed to an 11-kb PCR fragment amplified from *A. nosocomialis*. Acquired SNPs resulting from variant calling and identification of converted markers (SNPs of *A. nosocomialis* acquired by *A. baylyi*) delineate the acquired regions (see Material and Methods). E. Quantitative analyses of the recombination tracts observed in panel D. Recombination tracts are defined as segments of DNA with at least 2 acquired SNPs and separated by more than 3 non-converted SNPs. Whiskers delineate the range, box delineate the 1st and 3rd quartile, horizontal line shows the median, and cross correspond to the mean.

189

190 Discussion

191

192 We here report on the functional characterization of YraN, a protein largely conserved in bacteria
 193 with 23,388 out of 30,238 genomes (77 %) showing a YraN (COG0792) homolog (Figure 4A).
 194 Interestingly, YraN is absent from the genomes of Firmicutes. The function of YraN was, until now,
 195 unknown. First, based on Tn-seq results, we found that YraN plays an important role in promoting
 196 natural transformation in two Gram-negative transformable bacterial species, *L. pneumophila* and *S.*
 197 *elongatus*. By comparing Tn-seq analyses of genetic determinants of natural transformation in these
 198 two species, we confirmed that the transformation pathway requires genes encoding the components
 199 of the Type IV pilus, and all genes previously identified as being specifically involved in natural
 200 transformation. Surprisingly, only one other gene was required for transformation, *yraN*. The role of
 201 *yraN*, was further confirmed using deletion mutants in *L. pneumophila*, *A. nosocomialis*, *A. baylyi*
 202 (Figure 1) but also *A. baumannii* (data not shown) and *Vibrio cholerae* (Ankur Dalia, University of
 203 Indiana; personal communication). Being a member of a large family of phosphodiesterase, one

204 possible role for YraN could have been to process the periplasmic dsDNA to ssDNA transported
 205 across the inner membrane. However, the *yraN* mutant is not defective for transformation of plasmid
 206 which can replicate in the cytoplasm (Figure S2A). This readily excluded a generic role in DNA
 207 uptake. This, in turn, and given the list of genes required for transformation in *L. pneumophila* and *S.*
 208 *elongatus*, also indicates that the nuclease activity involved in the conversion of dsDNA to ssDNA is
 209 either carried out by one of the known components of the pathway (ComEC, ComFC) or by an
 210 enzyme that is either distinct between species or essential for viability. The differential requirement of
 211 YraN for transformation of DNA that depends or not on recombination pointed to a specific role in
 212 the recombination pathway. This was further supported by computational evidence of co-occurrence
 213 of YraN and ComM as well as co-evolution analyses. We indeed confirmed that YraN could interact
 214 with ComM (Figure 1). The lack of YraN produces a transformation deficiency similar to the absence
 215 of ComM, and consistent with a functional association of ComM and YraN, the mutants of the
 216 corresponding genes display similar deficiency in the comigration assay (Figure 3). Analyses of
 217 recombination tracts produced by mutants of *yraN* and *comM* indicate that both are needed to produce
 218 extended recombination tracts (Figure 3) which may explain the transformation deficiency when
 219 tested by acquisition of single SNP conferring resistance (Figure 1). While the helicase activity of
 220 ComM in extending the D-loop can explain the production of larger recombination tracts, the
 221 contribution of a nuclease was unexpected.

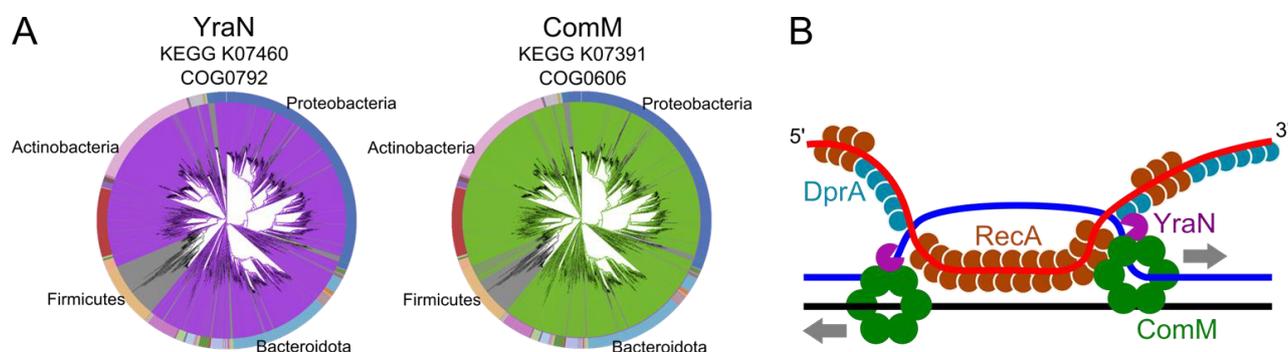


Figure 4. YraN is a widely conserved nuclease promoting extension of the recombination D-loop during natural transformation.

A. Phylogenetic distribution of YraN (COG0792) and ComM (COG0606) according to the AnnoTree database. AnnoTree allows the visualization of annotations across large phylogenetic trees, consisting here in 30,238 bacterial genomes (31). Absence (grey) or presence of an annotated YraN (purple) or ComM (green) homolog is reported in genomes (one genome per species). B. Proposed model of the transformation D-loop with extension promoted by ComM and YraN. YraN would associate with ComM and exert its endonuclease activity on the parental strand (blue), displaced by the RecA-dependent invasion of the DprA-protected transforming ssDNA (red).

222

223 Given that the endonuclease activity of YraN is essential for its biological function (Figure 2), how
 224 could cleavage of ssDNA promote integration of transforming DNA? On the basis of our results and
 225 the model in Figure 4B, we propose that the interaction of YraN with ComM and DprA can promote
 226 its recruitment at the RecA-mediated ssDNA-invasion event. Being loaded with DprA and RecA, this

227 ssDNA strand (red strand) is protected from the action of nuclease (32). As the chromosomal DNA is
228 invaded by the tDNA, its strand complementary to tDNA (black strand) is engaged in heteroduplex
229 formation, while the opposite strand is displaced (blue strand), resulting in a displacement loop (D-
230 loop). This may allow for the recruitment of the ComM helicase at the recombination site, which
231 would unwind the DNA, promoting extension of the D-loop (15). However, only about 25% of
232 invasion events followed by the recruitment of ComM successfully produced recombinants (15),
233 suggesting that D-loop extension and strand invasion are reversible events. Recruitment of the YraN
234 nuclease at the D-loop may introduce a cleavage of the unprotected displaced strand (blue in the
235 model of Figure 4B). This cleavage may prevent backtracking of the ComM helicase, resulting in
236 higher success rate of recombination events. In addition, ComM-dependent unwinding of the helix
237 during D-loop extension is anticipated to cause supercoiling of the DNA ahead of ComM.
238 Supercoiling will increase with the progression of ComM and may generate a topological roadblock.
239 Cleavage of the displaced strand by YraN, acting as a relaxase would relieve the topological stress,
240 ensuring that ComM extend the D-loop over long distances (several kb). This role of YraN may be
241 consistent with the weaker comigration defect of the *yraN* mutant than the *comM* mutant at short
242 distance (1.6 kb), while the two mutants display stronger and identical defect at longer distance (3
243 kb) (Figure 3B). YraN could be acting at the front of ComM. In that case, whichever strand is cleaved
244 by YraN, it must be repaired to either allow the progression of ComM (travelling on the blue strand,
245 Figure 4B) and to maintain the integrity of the maintained strand (black strand) which will duplex
246 with the tDNA. Alternatively, YraN may be positioned behind ComM, where the interaction of YraN
247 with DprA may direct it toward cleavage of the displaced strand (blue strand, Figure 4B). Further
248 work is needed to establish the structural configuration of the cooperation of YraN and ComM.
249 However, the YraN-ComM association represents a new example of direct interactions between
250 helicase and nuclease involved in DNA requiring the unwinding of the parental DNA strands (33).

251 Why would bacteria have evolved such system? Genome-wide analyses of recombination tracts in
252 transformable Gram-negative species have highlighted that natural transformation produces large
253 recombination tracts, with a mean length of 14 kb in *Haemophilus influenzae* (34) or 10 kb in
254 *Neisseria meningitidis* (35). Indeed, under mixed-culture conditions, where DNA is naturally
255 released, the mean length of the total acquired DNA in *V. cholerae* is even higher and was found to
256 be of ~70 kbp, with recombination events of up to 168 kbp (36). Similar observations were made in
257 mixed cultures of *A. nosocomialis* and *A. baumannii* (37). Thus, Gram-negative bacteria might have
258 evolved the YraN/ComM nuclease/helicase system to favor the recombination of sequence over long
259 distances. Interestingly, this system seems restricted to Gram-negative bacteria (Figure 4B). In light
260 of work in *S. pneumoniae*, the RadA helicase might have been coopted by Gram-positive bacteria to

261 also serve in the recombination of transforming DNA (38). It would thus be expected that a functional
262 equivalent of YraN also exists in Gram-positive bacteria, yet its identity is not known.

263 In conclusion, the results of this study reveals the involvement a dedicated nuclease in promoting large
264 recombination events resulting from the uptake of DNA by natural transformation. This broadly
265 conserved nuclease is yet another indication that most bacteria possess a complete and functional
266 pathway for the transport and recombination of exogenous DNA.

267 **Material and methods**

268 **Bacteria and growth conditions**

269 All the *Legionella pneumophila* strains are derived from the Paris or Philadelphia-1 strains. They
270 were cultured in solid medium containing ACES [*N*-(2-acetamido)-2-aminoethanesulfonic acid] and
271 buffered charcoal yeast extract (CYE) or in liquid medium containing ACES and buffered yeast
272 extract (AYE). When appropriate, media were supplemented with kanamycin (15 µg/mL for Paris
273 strain and 50 µg/mL for Philadelphia strains), streptomycin (50 µg/mL), gentamycin (10 µg/mL) or
274 apramycin (15 µg/mL). All the *Acinetobacter* strains are derived from *A. baumannii* AB5075-UW, *A.*
275 *nosocomialis* M2 or *A. baylyi* ADP1 (also know as BD413). When appropriate, they were grown in
276 LB broth or LB agar supplemented with tetramycin (5 µg/mL), kanamycin (50 µg/mL), apramycin
277 (30 µg/mL) or rifampicin (100 µg/mL). The methods for constructing mutants derived from these
278 different strains were previously reported (39–41) and described in the supplementary material.

279 **Transformation assays**

280 In transformation assay of *L. pneumophila*, the parent (WT) strain is a constitutively transformable
281 strain obtained by introducing a premature stop codon in the post-transcriptional repressor of
282 competence *rocC* (39, 40). Overnight streak of the desired *L. pneumophila* Paris strain was
283 resuspended in 1 mL sterile water to and OD600 of 1 to obtain $\sim 1.10^9$ CFU/mL. A 10-µL portion of
284 this suspension ($\sim 1.10^7$ CFUs) was spotted on CYE plate supplemented with 0.5 mM IPTG when
285 appropriate. Then, 10 µL of transforming DNA (a 4 kb-long PCR product encompassing the *rpsL*
286 mutation conferring resistance to streptomycin (42)) at the indicated concentration was spotted
287 directly on the bacteria spot and left absorbed onto the agar. Plates were incubated at 37°C for 24
288 hours. Each spot was then resuspended in 200 µL sterile water and ten-fold serial dilution were plated
289 on nonselective and selective medium. Plates were incubated at 37°C for 72 hours and CFU counting
290 was performed.

291 Transformation assays of *A. baumannii* and *A. nosocomialis* were performed as previously described
292 (41, 43). 2 mL LB were inoculated with overnight streak of the desired strain and incubated at 37°C
293 until reaching an OD600 of 1 to obtain $\sim 1.10^9$ CFU/mL. In sterile water, culture was 100-fold diluted

294 to obtain 10^7 CFU/mL. Then, 10 μ L of this suspension was mixed with 10 μ L of transforming DNA
295 at the indicated concentration and 2.5 μ L of this mixture was stabbed into freshly prepared motility
296 medium (5 g/liter agarose, 5 g/liter tryptone, and 2.5 g/liter NaCl) in 2-mL Eppendorf tubes. Tubes
297 were incubated at 37°C overnight and after incubation time, bacteria were resuspended in 200 μ L
298 sterile water by vigorous vortexing. Ten-fold serial dilutions of this suspension were then plated on
299 nonselective or selective medium. Plates were incubated at 37°C overnight and CFU counting was
300 performed.

301 Transformation assays of *A. baylyi* were conducted as follows: 70 μ L of overnight culture in LB liquid
302 medium at 30°C of the desired strain was used to inoculate 1 mL LB. After 1 hour incubation at 30°C
303 with shaking, tDNA was added at the indicated concentration and the cultures were re-incubated at
304 30°C with shaking for 1 hour to overnight and following DNase I treatment to stop DNA uptake (see
305 figure legend). Ten-fold serial dilution of this culture were then plated on nonselective or selective
306 medium. Plates were incubated at 30°C overnight and CFU counting was performed.

307 For all the strains, transformation frequencies were calculated as the ratio of the number of CFU
308 counted on selective medium (ie, the transformants) to the number of CFU counted on nonselective
309 medium (ie, total population). All transformation assays were performed at least three times
310 independently (on different days).

311 **Bacterial two-hybrid system experiment**

312 Two hybrid experiment was carried out to test the interaction between proteins involved in
313 transformation-associated homologous recombination. To that end, *recA*, *dprA*, *yraN* and *comM*
314 genes from *L. pneumophila* Paris were cloned into pKT25, pKNT25, pUT18 or pUT18C plasmids in
315 frame with T25 or T18 catalytic domains of *B. pertussis* adenylate cyclase (27) by Sequence and
316 Ligation Independent Cloning (SLIC) (44) (see supplementary material for more details). With this
317 method, for each protein of interest, 2 recombinant plasmids encoding 2 hybrid proteins were
318 propagated in their respective *Escherichia coli* DH5a competent hosts : pKNT25-X and pUT18-X,
319 encoding respectively X-T25 and T18-X. Each plasmid encoding a X-T25 fusion protein was then
320 co-transformed in competent BTH101 *cya*- strains with a plasmid encoding X-T18 fusion protein.
321 The co-transformed bacteria obtained were pre-cultured in LB liquid medium supplemented with 25
322 μ g/mL kanamycin and 100 μ g/mL ampicillin at 37°C overnight with shaking. After incubation, these
323 cultures were 1:10 diluted and spotted onto LB agar supplemented with 25 μ g/mL kanamycin, 100
324 μ g/mL ampicillin, 0.5 mM IPTG and 100 μ g/mL X- β -Gal (Roth) and incubated at 20°C for 48 to 72
325 hours. Test of interaction between DprA and RecA are used as positive controls in this experiment as
326 their interaction was already described (13).

327 **Expression and purification of YraN**

328 In order to produced C-terminal His tagged YraN in *E. coli*, *yraN* from *L. pneumophila* was amplified
329 by PCR and introduced in pKHS vector (a pET28 derivative) by InFusion methodology. The E50A
330 variant was further obtained by site-directed mutagenesis using an adapted protocol from the
331 Stratagene "QuickChange" with only one mutator primer carrying the modified sequence and with
332 both the Pfu turbo polymerase (Stratagene) and the Taq ligase (New England Biolabs) in the PCR
333 reaction. pKHS-*yraN* and its derivative pKHS-*yraN*-E50A were introduced in *E. coli* BL21-
334 Rosetta(DE3)-pLysS (CmR) cells (Novagen). The expression clones were grown at 30°C in 400 ml
335 LB medium with Cm (10 µg.mL⁻¹), Kan (50 µg.mL⁻¹), 0.2% glucose until the OD600 reached 0.7.
336 Protein expression was then induced by adding 0.5 mM IPTG followed by growth at 30°C for 2 hours.
337 Cell pellets were collected by centrifugation (6000g; 20min; 4°C) and further resuspended in 30 ml
338 solution containing 50 mM Na-Phosphate pH 7.5, 300 mM NaCl, 10 mM imidazol before breaking
339 cells and DNA with sonifier (Vibracell). The soluble extracts obtained after clarification
340 centrifugation (15000 g; 30min; 4°C in JA20) were injected first onto an His-Trap crude 1 ml column
341 and further purified by FPLC (Äkta purifier-10, GE Healthcare) using a solution of 50 mM Na-
342 Phosphate pH 7.5, 300 mM NaCl, and an increasing gradient of imidazol. The eluted fractions were
343 diluted 2.5 fold in 50 mM Na-Phosphate pH 7.5, to reach 120 mM of NaCl and were further passed
344 through a Hi-Trap Q HP 1 mL column to remove residual DNA before binding on a Hi-Trap heparin
345 HP 1 mL column (GE Healthcare) and elution with a NaCl gradient. Purified YraN proteins or variant
346 were finally injected and separated on a Superdex 75 16/600 sizing column (GE healthcare)
347 equilibrated with 50 mM Na-Phosphate pH 7.5, 200 mM NaCl. Part of the purified protein were
348 stored at -80°C after addition of 10% glycerol, 1 mM TCEP for longer storage or at -20°C after
349 addition of 50% of glycerol. Another part was dialyzed against 50 mM Hepes pH 7.6, 200 mM NaCl,
350 0,5 mM TCEP, 10 % glycerol using PD miniTraP 25 buffer exchange column for buffer comparative
351 assays on YraN activities. The purity and the concentrations of the purified proteins were analyzed
352 by SDS-PAGE and coomassie staining analysis.

353 **Nuclease assay**

354 Increasing amounts of YraN and YraN-E50A (150 to 1000 nM) were incubated 30 min at 37°C with
355 either 100 ng of supercoiled pUC18 plasmid or 10 nM of Cy3 labelled DNA substrates, in 10 µL
356 solutions containing 50 mM Hepes pH 7.6; 50 mM NaCl; 0,1 mg/ml BSA; 0,5 mM TCEP, 6%
357 Glycerol, 0,3 to 3 mM MnCl₂ or/and 10 mM MgO(acetate)₂. In order to quench the reaction and to
358 deproteinize the DNA, 0.1% SDS and 1 mM EDTA were added to the reaction for 5 min at 37°C
359 before loading half of the reaction on 1% agarose gel / TAE 1X for long DNA molecules or on 4%
360 acrylamide (29:1) / 0.3X TBE gels for short DNA structures (100 nt). Cy3 labeled oligonucleotides

361 were detected with the Typhoon Trio imager and the cold plasmid DNA was detected after Ethyidium
362 bromide staining.

363 **Comigration marker experiment**

364 Comigration marker experiment was adapted from the comigration assay conducted in *V. cholerae*
365 (16). This experiment was conducted in *A. nosocomialis* M2 and *A. baylyi* ADP1 mutants in a similar
366 manner as follows. Strains that carries genetically linked markers : *lacZ* gene placed adjacent to a
367 tetracyclin cassette were constructed as described in the supplementary material. These strains called
368 “*lacZ*⁺-tet” were naturally transformed with the following PCR products: “2kB-*lacZ*^{dlt1}-kan-2kB” or
369 “2kB-*lacZ*^{dlt2}-kan-2kB” or “2kB-*lacZ*^{dlt3}-kan-2kB” as tDNA. These tDNA have 2kB upstream and
370 downstream homologous regions with “*lacZ*⁺-tet” recipient strains that allow the replacement of the
371 tetracyclin cassette by kanamycin cassette and the deletion of a 7pb-part of *lacZ* (*lacZ*^{dltX}) by
372 homologous recombination. The 7 pb deletion was designed so that (i) it is not a substrate for the
373 mismatch repair and (ii) it shifts the reading frame of *lacZ* gene to produce a TGA stop codon resulting
374 in *lacZ*- phenotype. Depending on the tDNA used, these 7 bp deletions were designed at increasing
375 distances from the original TAA stop codon of the kanamycin cassette (distance between *kan*_{TAA} and
376 *lacZ*_{TGA} for *dlt1*=3043bp, for *dlt2*=253bp and for *dlt3*=1606bp). Ten-fold serial dilution of the
377 transformation cultures were plated on LB agar supplemented with 50 µg/mL kanamycin and 100
378 µg/mL X-gal or on nonselective LB plates and incubated in appropriate conditions described before.
379 Co-migration efficiency, which refers to the number of kanamycin-resistant transformants that have
380 acquired the *lacZ* deletion, was calculated, as the ratio of the number of white transformants
381 (*kanR/lacZ*-) divided by the number of total transformants (*KanR/lacZ*⁺ & *KanR/lacZ*-) for each
382 tDNA and each mutant.

383 **Characterization of recombination tract length**

384 Direct detection of transformation recombination tracts was carried out as follows. *A. baylyi* strain
385 was transformed with a 11 kb-long PCR product as tDNA that confer resistance to rifampicin upon
386 recombination. This tDNA was obtained by amplifying a 11-kb portion of *A. nosocomialis* M2 *rif*^R
387 genome (37) that carries the *rpoB* gene, using forward and reverse primers at equal distance from the
388 mutation conferring resistance to rifampicin (see primer list in the supplementary material). *A. baylyi*
389 mutants were pré-cultivated for 2H at 30°C (from -80°C frozen stock) and transformed with 200 ng
390 of DNA for 1 hour, DNase I was added to the culture, further incubated for 1 hour, and plated
391 rifampicin-containing plates. 24 colonies were picked for each mutant and the region of interest was
392 amplified as two overlapping PCR products, which were used to construct a sequencing library (Rapid
393 Barcoding Kit 96 SQK-RBK110.96, Oxford Nanopore) and sequenced on a MinIon instrument
394 (Oxford Nanopore). Reads were basecalled using the high-accuracy model (hac, guppy 5). Variant

395 calling was performed using Medaka (v1.3.4) using the donor and recipient sequence as reference.
396 Converted SNP correspond to haplotype which differs between donor and recipient, and match donor
397 haplotype. Recombination tracts are defined as series of at least 2 converted SNPs that are not
398 interrupted by more than 3 unconverted SNPs.

399

400 **Acknowledgments**

401 We thank Marie Cariou (CIRI) for her assistance in writing scripts for analyses of recombination
402 tracts. This work was supported in part by grant ANR-20-CE12-0004 (TransfoConflict) from Agence
403 Nationale de la Recherche (awarded to XC). LH is supported by a fellowship "Fin de thèse" from the
404 Fondation pour la Recherche Médicale (FDT202001010890).

405

406 **References**

407

1. Johnston C, Martin B, Fichant G, Polard P, Claverys J-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* 12:181–196.
2. Dubnau D, Blokesch M. 2019. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu Rev Genet* 53:217–237.
3. Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, Dujeancourt A, Campo N, Chamot-Rooke J, Le Cam E, Claverys J-P, Fronzes R. 2013. A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*. *PLoS Pathog* 9:e1003473.
4. Ellison CK, Dalia TN, Vidal Ceballos A, Wang JC-Y, Biais N, Brun YV, Dalia AB. 2018. Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat Microbiol* 3:773–780.
5. Berry J-L, Pelicic V. 2015. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol Rev* 39:134–154.
6. Lam T, Ellison CK, Eddington DT, Brun YV, Dalia AB, Morrison DA. 2021. Competence pili in *Streptococcus pneumoniae* are highly dynamic structures that retract to promote DNA uptake. *Mol Microbiol* 116:381–396.
7. Provvedi R, Dubnau D. 1999. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Molecular Microbiology* 31:271–280.
8. Seitz P, Pezeshgi Modarres H, Borgeaud S, Bulushev RD, Steinbock LJ, Radenovic A, Dal Peraro M, Blokesch M. 2014. ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. *PLoS Genet* 10:e1004066.
9. Draskovic I, Dubnau D. 2005. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol Microbiol* 55:881–896.
10. Diallo A, Foster HR, Gromek KA, Perry TN, Dujeancourt A, Krasteva PV, Gubellini F, Falbel TG, Burton BM, Fronzes R. 2017. Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. *Molecular Microbiology* 105:741–754.
11. Sinha S, Mell JC, Redfield RJ. 2012. Seventeen Sxy-Dependent Cyclic AMP Receptor Protein Site-Regulated Genes Are Needed for Natural Transformation in *Haemophilus influenzae*. *J Bacteriol* 194:5245–5254.
12. Seitz P, Blokesch M. 2013. DNA-uptake machinery of naturally competent *Vibrio cholerae*. *PNAS* 110:17987–17992.

13. Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, Fichant G, Martin B, Noiro P, Le Cam E, Polard P, Claverys J-P. 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* 130:824–36.
14. Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol Microbiol* 50:527–536.
15. Dalia AB, Dalia TN. 2019. Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. *Cell* 179:1499–1511.e10.
16. Nero TM, Dalia TN, Wang JC-Y, Kysela DT, Bochman ML, Dalia AB. 2018. ComM is a hexameric helicase that promotes branch migration during natural transformation in diverse Gram-negative species. *Nucleic Acids Res* 46:6099–6111.
17. Albano M, Hahn J, Dubnau D. 1987. Expression of competence genes in *Bacillus subtilis*. *J Bacteriol* 169:3110–3117.
18. Fani R, Mastromei G, Polsinelli M, Venema G. 1984. Isolation and characterization of *Bacillus subtilis* mutants altered in competence. *J Bacteriol* 157:152–157.
19. Gwinn ML, Ramanathan R, Smith HO, Tomb JF. 1998. A new transformation-deficient mutant of *Haemophilus influenzae* Rd with normal DNA uptake. *J Bacteriol* 180:746–748.
20. Karudapuram S, Zhao X, Barcak GJ. 1995. DNA sequence and characterization of *Haemophilus influenzae* dprA+, a gene required for chromosomal but not plasmid DNA transformation. *J Bacteriol* 177:3235–3240.
21. Hardy L, Juan P-A, Coupat-Goutaland B, Charpentier X. 2021. Transposon Insertion Sequencing in a Clinical Isolate of *Legionella pneumophila* Identifies Essential Genes and Determinants of Natural Transformation. *J Bacteriol* 203:e00548-20.
22. Taton A, Erikson C, Yang Y, Rubin BE, Rifkin SA, Golden JW, Golden SS. 2020. The circadian clock and darkness control natural competence in cyanobacteria. *Nat Commun* 11:1688.
23. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* 34:2115–2122.
24. Christensen LM, Sule P, Strain M, Cirillo JD. 2019. *Legionella pneumophila* p45 element influences host cell entry and sensitivity to sodium. *PLoS One* 14:e0218941.
25. Cong Q, Anishchenko I, Ovchinnikov S, Baker D. 2019. Protein interaction networks revealed by proteome coevolution. *Science*.
26. A multi-scale coevolutionary approach to predict interactions between protein domains.
27. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *PNAS* 95:5752–5756.
28. Laganeckas M, Margelevicius M, Venclovas C. 2011. Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. *Nucleic Acids Res* 39:1187–1196.
29. Steczkiewicz K, Muszewska A, Knizewski L, Rychlewski L, Ginalski K. 2012. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Research* 40:7016–7045.
30. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.
31. Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* <https://doi.org/10.1093/nar/gkz246>.
32. Bergé M, Mortier-Barrière I, Martin B, Claverys J-P. 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol Microbiol* 50:527–536.
33. Duguet M. 1997. When helicase and topoisomerase meet! *J Cell Sci* 110 (Pt 12):1345–1350.
34. Mell JC, Shumilina S, Hall IM, Redfield RJ. 2011. Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog* 7:e1002151.
35. Alfsnes K, Frye SA, Eriksson J, Eldholm V, Brynildsrud OB, Bohlin J, Harrison OB, Hood DW, Maiden MCJ, Tønnum T, Ambur OH. 2018. A genomic view of experimental intraspecies and interspecies transformation of a rifampicin-resistance allele into *Neisseria meningitidis*. *Microb Genom* 4.

36. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. 2019. Neighbor predation linked to natural competence fosters the transfer of large genomic regions in *Vibrio cholerae*. *eLife* 8:e48212.
37. Godeux A-S, Sveldhom E, Barreto S, Potron A, Venner S, Charpentier X, Laaberki M-H. 2021. Interbacterial transfer of carbapenem resistance and large antibiotic resistance islands by natural transformation in pathogenic *Acinetobacter*.
38. Marie L, Rapisarda C, Morales V, Bergé M, Perry T, Soulet A-L, Gruget C, Remaut H, Fronzes R, Polard P. 2017. Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. *Nat Commun* 8:15638.
39. Attaiech L, Boughammoura A, Brochier-Armanet C, Allatif O, Peillard-Fiorente F, Edwards RA, Omar AR, MacMillan AM, Glover M, Charpentier X. 2016. Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc Natl Acad Sci USA* 113:8813–8818.
40. Juan P-A, Attaiech L, Charpentier X. 2015. Natural transformation occurs independently of the essential actin-like MreB cytoskeleton in *Legionella pneumophila*. *Scientific Reports* 5:16033.
41. Godeux A-S, Svedholm E, Lupo A, Haenni M, Venner S, Laaberki M-H, Charpentier X. 2020. Scarless Removal of Large Resistance Island AbaR Results in Antibiotic Susceptibility and Increased Natural Transformability in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* 64.
42. Durieux I, Ginevra C, Attaiech L, Picq K, Juan P-A, Jarraud S, Charpentier X. 2019. Diverse conjugative elements silence natural transformation in *Legionella* species. *PNAS* 116:18613–18618.
43. Godeux A-S, Lupo A, Haenni M, Guette-Marquet S, Wilharm G, Laaberki M-H, Charpentier X. 2018. Fluorescence-Based Detection of Natural Transformation in Drug-Resistant *Acinetobacter baumannii*. *Journal of Bacteriology* 200:e00181-18.
44. Li MZ, Elledge SJ. 2012. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol Biol* 852:51–59.

Supplementary information

YraN is a highly conserved nuclease involved in the processing the recombination

D-loop during natural transformation

Léo Hardy^a, Violette Morales^{b,c}, Patrice Polard^{b,c} and Xavier Charpentier^{a,*}

^a CIRI, Centre International de Recherche en Infectiologie, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Normale Supérieure de Lyon, Univ Lyon, 69100, Villeurbanne, France. Team Horigene, <https://horigene.fr>

^b Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Centre National de la Recherche Scientifique (CNRS), 31062 Toulouse, France.

^c Université de Toulouse, Université Paul Sabatier, 31062 Toulouse, France.

* Correspondence to :

xavier.charpentier@univ-lyon1.fr | @labXC

Running title: YraN, a nuclease involved in natural transformation

Figure S1. YraN is required for optimal natural transformation of heterologous DNA sequence.

Figure S2. YraN is dispensable for natural transformation of plasmid and conjugation of ICE.

Figure S3. YraN is dispensable for repair of DNA damage caused by UV irradiation.

Figure S4. Output of the lpp3065 (*L. pneumophila* yraN) query to the string-db with co-occurrence as sole active interaction sources.

Table S1. List of strains

Table S2. List of primers used in this study

Supplementary material and methods

28

29

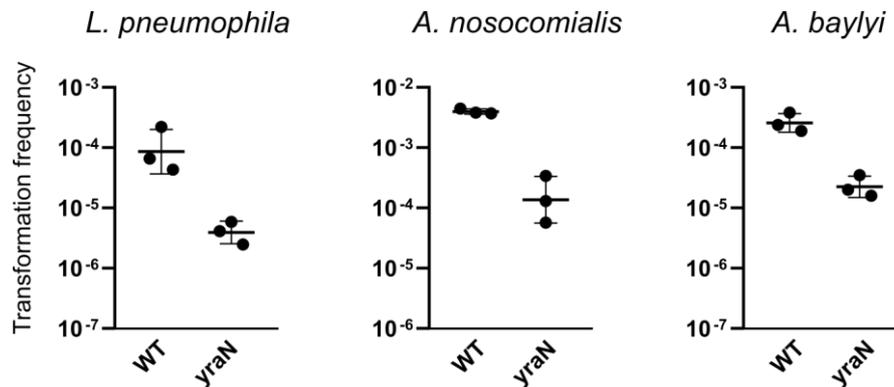


Figure S1. YraN is required for optimal natural transformation of heterologous DNA sequence.

L. pneumophila (Lp), *A. nosocomialis* (An) or *A. baylyi* (Ab), parental genotype or deleted for the *yraN* gene, were exposed to transforming DNA consisting of an apramycin (Lp) or kanamycin (An and Ab) resistance gene flanked by 2 kb-long regions with homology to the chromosome. Transformation frequencies represent the ratio of colony-forming units on selective (containing apramycin or kanamycin) over the total CFU count on non-selective plates (CYE for Lp, LB for An and Ab).

30

31

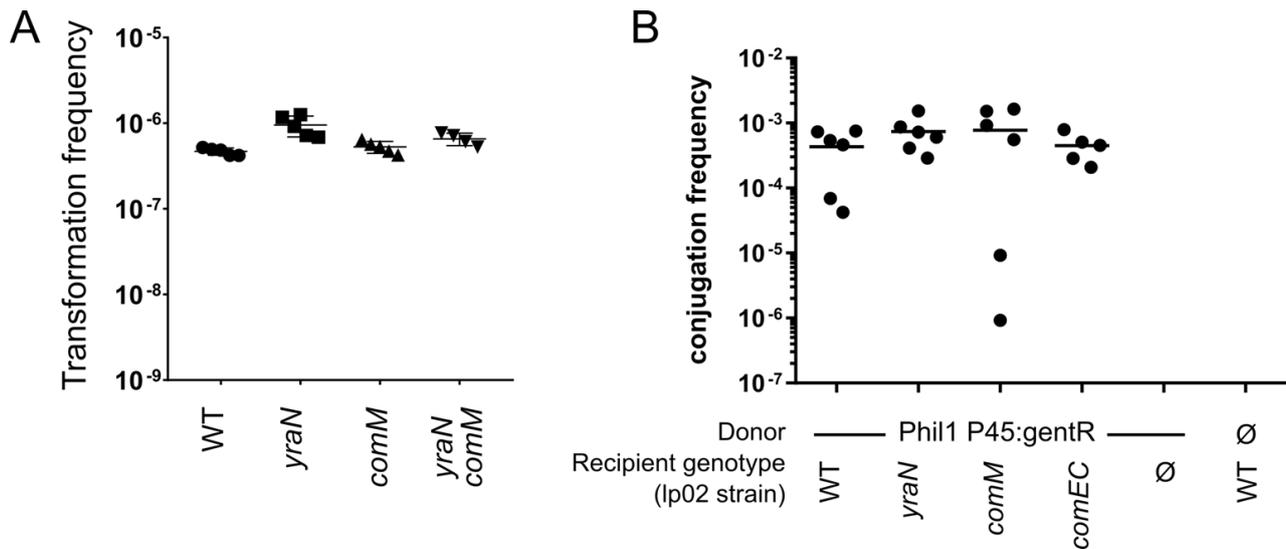


Figure S2. YraN is dispensable for natural transformation of plasmid and conjugation of ICE

A. Transformation frequencies of *A. baylyi* (Ab) and mutants by the plasmid pX5K, an RSF1010-derived plasmid (*incQ*) capable of episomal replication in a wide range of hosts and carrying the kanamycin resistance gene. Ab was exposed to the purified plasmid for 24h and cultures were plated on selective plates containing kanamycin. Transformation frequencies represent the ratio of colony-forming units on selective over the total CFU count on non-selective plates.

B. Conjugation efficiency of the P45 integrative conjugative element (ICE) carrying a gentamicin resistance gene (*gentR*) from the Philadelphia-1 strain of *L. pneumophila* to the lp02 strain of *L. pneumophila* of parental genotype or deleted for *yraN*, *comM* or *comEC*. The latter demonstrates that transfer of P45 is indeed independent of natural transformation. Following co-culture in AYE broth for 48h, cultures were plated on CYE plates containing streptomycin (which selects for the lp02 donor) and gentamicin (which selects for the acquisition of P45::*gentR*). Conjugation frequencies represent the ratio of colony-forming units on selective (containing streptomycin and gentamicin) over the total CFU count on non-selective plates.

33

34

35

36

37

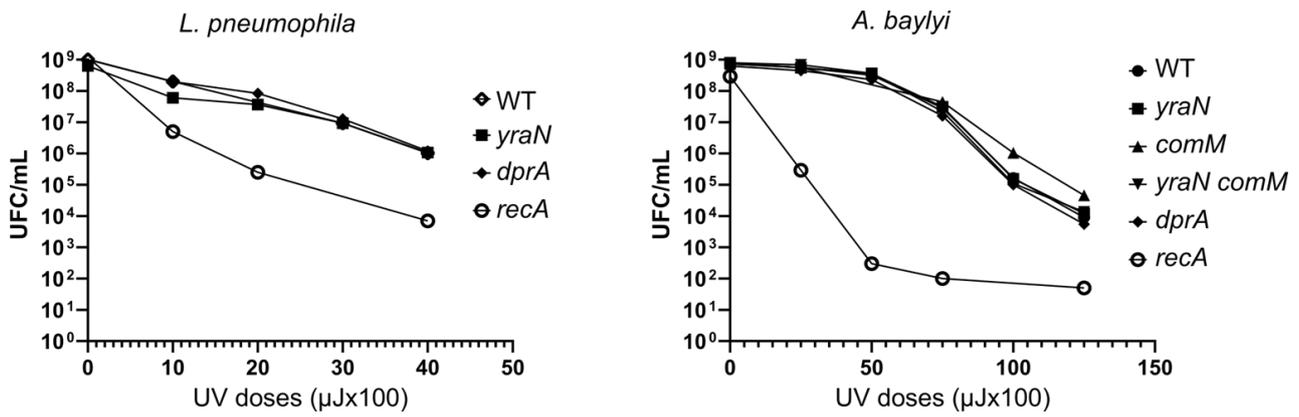


Figure S3. YraN is dispensable for repair of DNA damage caused by UV irradiation. From overnight streak, bacteria were cultivated in their respective media and at their respective incubation temperatures overnight. Overnight cultures were diluted in water to reach an OD₆₀₀ of 1 and deletions were ten-fold serial diluted and 20 μL of each dilution was spotted on the respective solid media of each species. The plates with the spots were incubated directly at different UV doses in a UV oven. Finally, the plates were incubated in the respective conditions of each species to obtain visible colonies and CFU counting was performed.

38

39

40

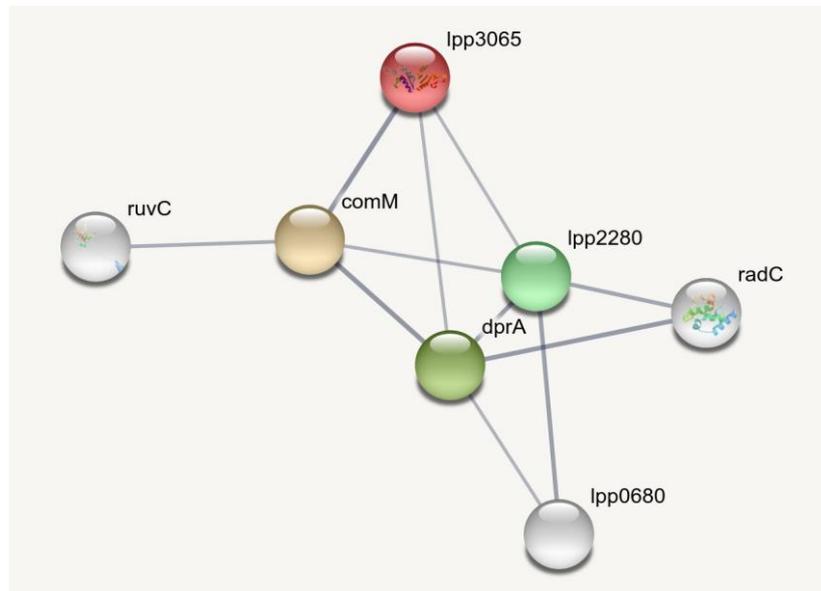


Figure S4. Output of the *lpp3065* (*L. pneumophila yraN*) query to the string-db with co-occurrence as sole active interaction sources. Thickness of the connecting line indicates the strength of data support. *Lpp2280* is ComFC; *lpp0680* is ComEC.

41

Strains

| Organism | Strain | Genotype | Source |
|---|---|---|--|
| <i>L. pneumophila</i> | Paris WT | Wild-type <i>Legionella pneumophila</i> | Paris Outbreak isolate CIP107629, CNR Lyon |
| | Paris_S | Spontaneous streptomycin-resistant mutant of Paris; <i>strepR</i> | This study |
| | Paris <i>rocC_{TAA} lpp2553::apra</i> | Paris <i>rocC_{TAA} ; lpp2553::apra ; ApraR</i> | This study |
| | Paris <i>yraN::kan</i> | Paris ; <i>yraN::kan ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan</i> | Paris <i>rocC_{TAA} ; yraN::kan ; KanR</i> | This study |
| | Paris <i>recA::kan</i> | Paris ; <i>recA::kan ; KanR</i> | This study |
| | Paris <i>dpra::kan</i> | Paris ; <i>dpra::kan ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} recA::kan</i> | Paris <i>rocC_{TAA} ; recA::kan ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} dpra::kan</i> | Paris <i>rocC_{TAA} ; dpra::kan ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} comM::gent</i> | Paris <i>rocC_{TAA} ; comM::gent ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan comM::gent</i> | Paris <i>rocC_{TAA} ; yraN::kan ; comM::gent ; GentR ; KanR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan pLPP::lacIQ-ptac-yraN-gent</i> | Paris <i>rocC_{TAA} ; yraN::kan ; pLPP::lacIQ-ptac-yraN-gent ; KanR, GentR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan pLPP::lacIQ-ptac-yraN-flag-gent</i> | Paris <i>rocC_{TAA} ; yraN::kan ; pLPP::lacIQ-ptac-yraN-flag-gent ; KanR, GentR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan pLPP::lacIQ-ptac-yraN^{D37A}-gent</i> | Paris <i>rocC_{TAA} ; yraN::kan ; pLPP::lacIQ-ptac-yraN^{D37A}-gent ; KanR, GentR</i> | This study |
| | Paris <i>rocC_{TAA} yraN::kan pLPP::lacIQ-ptac-yraN^{E50A}-gent</i> | Paris <i>rocC_{TAA} ; yraN::kan ; pLPP::lacIQ-ptac-yraN^{E50A}-gent ; KanR, GentR</i> | This study |
| | Lp02 <i>yraN::kan</i> | Lp02 ; <i>yraN::kan ; KanR</i> | This study |
| | Lp02 <i>comM::kan</i> | Lp02 ; <i>comM::kan ; KanR</i> | This study |
| | Lp02 <i>comEC::kan</i> | Lp02 ; <i>comEC::kan ; KanR</i> | This study |
| | Lp02 <i>comM::kan</i> | Lp02 ; <i>comM::kan ; KanR</i> | This study |
| | <i>A. nosocomialis</i> | M2 WT | Wild-type <i>Acinetobacter nosocomialis</i> |
| M2_rifR | | Spontaneous rifampicin-resistant mutant of M2 ; <i>RifR</i> | This study |
| M2 <i>yraN::apra</i> | | M2 ; <i>yraN::apra ; ApraR</i> | This study |
| M2 Δ <i>comM yraN::apra</i> | | M2 ; Δ <i>comM ; yraN::apra ; ApraR ; lacZ+</i> | This study |
| M2 <i>atn7::lacZ-tet</i> | | M2 ; <i>atn7::lacZ-tet ; ApraR ; TetR ; lacZ+</i> | This study |
| M2 Δ <i>comM atn7::lacZ-tet</i> | | M2 ; Δ <i>comM ; atn7::lacZ-tet ; ApraR ; TetR ; lacZ+</i> | This study |
| M2 <i>yraN::apra atn7::lacZ-tet</i> | | M2 ; <i>yraN::apra ; atn7::lacZ-tet ; ApraR ; TetR ; lacZ+</i> | This study |
| M2 Δ <i>comM yraN::apra atn7::lacZ-tet</i> | | M2 ; <i>yraN::apra ; ΔcomM ; atn7::lacZ-tet ; ApraR ; TetR ; lacZ+</i> | This study |
| M2 <i>atn7::lacZ_{dt1}-kan</i> | | M2 ; <i>yraN::apra ; atn7::lacZ_{TAA}-kan ; KanR ; lacZ-</i> | This study |
| M2 <i>atn7::lacZ_{dt2}-kan</i> | | M2 ; <i>yraN::apra ; atn7::lacZ_{TAA}-kan ; KanR ; lacZ-</i> | This study |
| M2 <i>atn7::lacZ_{dt3}-kan</i> | | M2 ; <i>yraN::apra ; atn7::lacZ_{TAA}-kan ; KanR ; lacZ-</i> | This study |
| <i>A. baylyi</i> | | ADP1 WT | Wild-type <i>Acinetobacter baylyi</i> |
| | ADP1_rifR | Spontaneous rifampicin-resistant mutant of ADP1 ; <i>RifR</i> | This study |
| | BD413 WT | WT | This study |
| | ADP1 <i>yraN::apra</i> | BD413 ; <i>yraN::apra ; ApraR</i> | This study |
| | ADP1 <i>comM::gent</i> | BD413 ; <i>comM::gent ; gentR</i> | This study |
| | ADP1 <i>comM::gent yraN::apra</i> | BD413 ; <i>comM::gent yraN::apra ; gentR</i> | This study |
| <i>E. coli</i> | <i>BTH101</i> | <i>BTH101 ; cya-</i> | |

42

43 **Table S1 (1). Strains**

Plasmids

| Name | Genotype | Aim | source |
|----------------|--|--|-----------------------------|
| pUT18 | <i>AmpR</i> | vector encoding the T18 fragment that is fused downstream of the MCS. Used for BACTH system experiment; | Karimova et al, (PNAS 1998) |
| pKNT25 | <i>KanR</i> | number vector encoding the T25 fragment that is fused downstream of the MCS. Used for BACTH system experiment; | Karimova et al, (PNAS 1998) |
| pKNT25-dprA | <i>dprA-T25 ; kanR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T25 fused in frame to C-terminal of <i>dprA</i> gene from <i>L. pneumophila</i> | This study |
| pKNT25-comM | <i>comM-T25 ; kanR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T25 fused in frame to C-terminal of <i>comM</i> gene from <i>L. pneumophila</i> | This study |
| pKNT25-yraN | <i>yraN-T25 ; kanR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T25 fused in frame to C-terminal of <i>yraN</i> gene from <i>L. pneumophila</i> | This study |
| pKNT25-recA | <i>recA-T25 ; kanR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T25 fused in frame to C-terminal of <i>recA</i> gene from <i>L. pneumophila</i> | This study |
| pUT18-dprA | <i>dprA-T18 ; ampR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T18 fused in frame to C-terminal of <i>dprA</i> gene from <i>L. pneumophila</i> | This study |
| pUT18-recA | <i>recA-T18 ; ampR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T18 fused in frame to C-terminal of <i>comM</i> gene from <i>L. pneumophila</i> | This study |
| pUT18-yraN | <i>yraN-T18 ; ampR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T18 fused in frame to C-terminal of <i>yraN</i> gene from <i>L. pneumophila</i> | This study |
| pUT18-comM | <i>comM-T18 ; ampR</i> | plasmid used for two-hybrid experiment expressing IPTG-inducible T18 fused in frame to C-terminal of <i>recA</i> gene from <i>L. pneumophila</i> | This study |
| pX5K | <i>KanR</i> | derivative of the RSF1010-derived plasmid | This study |
| pMMB207C | <i>RSF1010::lacI^q tacp mobA</i> | cloning vector for expression genes under the Ptac promoter, derived from RSF1010 ; CmR | This study |
| pKHS | | Derivative of the commercial pET28 vector | |
| pKHS-yraN | <i>pKHS ; yraN</i> | <i>L. pneumophila</i> | This study |
| pKHS-yraN-E50A | <i>pKHS ; yraN^{E50A}</i> | plasmid for the production of YraN ^{E50A} catalytic mutant of YraN from <i>L. pneumophila</i> | This study |

44

Table S1 (2). Plasmids

| | | |
|---------------------------|--|--|
| LH13_lpp3065_P1 | CAACGCTTTCAAGGTATTAAGAGC | Forward primer to amplify a 2kB fragment upstream of <i>yraN</i> from Lp |
| LH14_lpp3065_P2-tail-pKD4 | GGAACTTCGAAGCAGCTCCAGCCTACACAATCG CCAATTGTTTCAGCAAACCTTCC | Reverse primer to amplify a 2kB fragment upstream of <i>yraN</i> from Lp |
| LH15_lpp3065_P3-tail-pKD4 | GAACTAAGGAGGATATTCATATGGACCATGGCG GTAAAGAACGCATTTGATGCAGG | Forward primer to amplify a 2kB fragment downstream of <i>yraN</i> from Lp |
| LH16_lpp3065_P4 | GACAACATTGCAGATCTCATCCATG | Reverse primer to amplify a 2kB fragment downstream of <i>yraN</i> from Lp |
| LH45_recA_P1 | GCAATTAGTGACTGCGGAATC | Forward primer to amplify a 2kB fragment upstream of <i>recA</i> from Lp |
| LH46_recA_P2 | GGAACTTCGAAGCAGCTCCAGCCTACACAATCC AATATCCCGTGATACCGTGC | Reverse primer to amplify a 2kB fragment upstream of <i>recA</i> from Lp |
| LH47_recA_P3 | GAACTAAGGAGGATATTCATATGGACCATGGCG TTGTCTGTGTTGGCTAGTTC | Forward primer to amplify a 2kB fragment downstream of <i>recA</i> from Lp |
| LH48_recA_P4 | CCTTTATCTCCTACCTGACCACC | Reverse primer to amplify a 2kB fragment downstream of <i>recA</i> from Lp |
| LH55_dprA_P1 | CTCCGATTTCAGCATCACC | Forward primer to amplify a 2kB fragment upstream of <i>dprA</i> from Lp |
| LH56_dprA_P2 | GGAACTTCGAAGCAGCTCCAGCCTACACAATCG GGCGAGGAGGGGTTGCAG | Reverse primer to amplify a 2kB fragment upstream of <i>dprA</i> from Lp |

| | | |
|------------------------------|---|--|
| LH57_dprA_P3 | GAACTAAGGAGGATATTCATATGGACCATGGCC TACTTATTGCAGCAGGGTGC | Forward primer to amplify a 2kB fragment downstream of dprA from Lp |
| LH58_dprA_P4 | CTACCACGTATTTCGCGCTG | Reverse primer to amplify a 2kB fragment downstream of dprA from Lp |
| LH66_pUT18_F | AGCTCGAATTCAGCCGCCAGCG | Forward primer to amplify T18 from pUT18 plasmid by SLIC |
| LH67_pUT18_R | CATAGCTGTTTCCTGTGTGAAATTGTTATC | Reverse primer to amplify T18 from pUT18 plasmid by SLIC |
| LH74_pKNT25_R | CATAGCTGTTTCCTGTGTG | Forward primer to amplify T25 from pUT25 plasmid by SLIC |
| LH75_pKNT25_F | AGCTCGAATTCATGACCATGC | Reverse primer to amplify T25 from pUT25 plasmid by SLIC |
| LH76_yraN-pKNT25_F | CACACAGGAAACAGCTATGACCCAAGAAAAAG GAAAGTTTGTCTG | Forward primer to amplify YraN for cloning in pKNT25 by SLIC |
| LH77_yraN-pKNT25_R | GCATGGTCATTGAATTCGAGCTGCATCCTGCATC AAATGCGTTC | Reverse primer to amplify YraN for cloning in pKNT25 by SLIC |
| LH82_comM-pUT18_F | TCACACAGGAAACAGCTATGAGTCTCGCATTTA CCAAAACGC | Forward primer to amplify comM for cloning in pUT18 by SLIC |
| LH83_comM-pUT18_R | CTGGCGGCTGAATTCGAGCTTTTCGGTAAAGTGT AAATTTTGTGTTG | Reverse primer to amplify comM for cloning in pUT18 by SLIC |
| LH86_comM-pKNT25_F | CACACAGGAAACAGCTATGAGTCTCGCATTTAC CAAAAACGC | Forward primer to amplify comM for cloning in pKNT25 by SLIC |
| LH87_comM-pKNT25_R | GCATGGTCATTGAATTCGAGCTCTATTTTCGGTAA GTGTAATTTTGTGTTG | Reverse primer to amplify comM for cloning in pKNT25 by SLIC |
| LH90_dprA-pUT18_F | TCACACAGGAAACAGCTATGAATAATCTGCAAC CCCTCCTCG | Forward primer to amplify dprA for cloning in pUT18 by SLIC |
| LH91_dprA-pUT18_R | CTGGCGGCTGAATTCGAGCTTCTTTCATATTCAC ACCTTATATAGCC | Reverse primer to amplify dprA for cloning in pUT18 by SLIC |
| LH94_dprA-pKNT25_F | CACACAGGAAACAGCTATGAATAATCTGCAACC CCTCCTCG | Forward primer to amplify dprA for cloning in pKNT25 by SLIC |
| LH95_dprA-pKNT25_R | GCATGGTCATTGAATTCGAGCTTCTTTCATATTC ACACCTTATATAGCC | Reverse primer to amplify dprA for cloning in pKNT25 by SLIC |
| LH98_recA-pUT18_F | TCACACAGGAAACAGCTATGGAAGAGAATAAAA CAAAAAGCAC | Forward primer to amplify recA for cloning in pUT18 by SLIC |
| LH99_recA-pUT18_R | CTGGCGGCTGAATTCGAGCTGTCTATAGTCT CAAAAAGATCTTC | Reverse primer to amplify recA for cloning in pUT18 by SLIC |
| LH102_recA-pKNT25_F | CACACAGGAAACAGCTATGGAAGAGAATAAAC AAAAAGCAC | Forward primer to amplify recA for cloning in pKNT25 by SLIC |
| LH103_recA-pKNT25_R | GCATGGTCATTGAATTCGAGCTGTCTATAGT CTCAAAAAGATCTTC | Reverse primer to amplify recA for cloning in pKNT25 by SLIC |
| LH118_YraN_D37A_F | ATCATTGCCGTTTAGGAGAAATTGCCTTAATTAT GCGTGAAGGATCTTATC | Forward primer to replace codon coding for D37 to A37 in yraN |
| LH119_YraN_D37A_R | GATAAGATCCTTACGCATAATTAAGGCAATTC TCCTAAACGGCAATGAT | Reverse primer to replace codon coding for D37 to A37 in yraN |
| LH120_yraN_E50A_F | GTGAAGGATCTTATCTGGTTTTTATTGCCGTTTCG TTCGCGTTCAAATATG | Forward primer to replace codon coding for E50 to A50 in yraN |
| LH121_yraN_E50A_R | CATATTTGAACGCGAACGAACGGCAATAAAAAAC CAGATAAGATCCTTCAC | Reverse primer to replace codon coding for E50 to A50 in yraN |
| LH140_plpp0110_P1 | GCCAAACGCTATCAGGAGTTACCTC | P1 primer to clone yraN in pLPP |
| LH141_plpp0110_P2-tail-LacIq | CTTGCTGGCGTTTCGGGAGCAGAAGAGCATAGAT TGTCCCATCTCCTTGAACC | P2 primer to clone yraN in pLPP |
| LH142_plpp0110_P3-tail-CmR | GTTCCGCGTCTTGCATAACTGTGTTACGCGA AGAATCAAGAATAACGCCAC | P3 primer to clone yraN in pLPP |
| LH143_plpp0110_P4 | CTGATGAGGCGCTGCGAAGAG | P4 primer to clone yraN in pLPP |
| LH144_plpp3065-LacIq_F | TATGCTCTTCTGCTCCCGAACGCCAGCAAG | F primer to clone yraN in pLPP |
| LH145_plpp3065-CmR_R | GTAACACAGTATTGCAAGGACGCGGAAC | R primer to clone yraN in pLPP |
| LH146_2kB-AttTn7_P1 | GTTCTTTCTGCTGTGCGAAAGGAC | P1 primer to insert lacZ-attB cassette in attn7 from M2 |
| LH147_AttTn7-tail-lacZ_P2 | ATTATACCTAGGACTGAGCTAGCTGTCAACGTG AAAGAAAATCTAGTTTGATAGTAGAG | P2 primer to insert lacZ-attB cassette in attn7 from M2 |
| LH148_attTn7-tail-kan_P3-v4 | CATCAACCATATCAGCAAAAAGTATACGGTAAA AGCGCTTGTAGCGATATGAATG | P3 primer to insert lacZ-attB cassette in attn7 from M2 |
| LH149_attTn7-2kB_P4 | CGTGCTCAACAAATTTCTAAGGC | P4 primer to insert lacZ-attB cassette in attn7 from M2 |
| LH150_TcR-tail-lacZ_F | CATTCATATCGTACACAAGCGCTTTTAACATCG CAAAAAGGTGTAATATTATAATGC | forward primer to insert lacZ-attB cassette in attn7 from M3 |

| | | |
|--------------------------------|---|--|
| LH151_TcR-tail-attTn7_R | GCCAGTGTACAACCAATTAACCAATTCTGATAC TTCCATTTCAGGTCGAGGTGGCCCGGC | reverse primer to insert lacZ-atbR cassette in attn7 from M4 |
| LH152_2k-TcR_P1 | TTCGCATTATCCGAACCATCCGC | P1 primer to amplify tet cassette |
| LH153_2kb-TcR_P2 | CAGAATTGGTTAATTGGTTGTAACACTGGC | P2 primer to amplify tet cassette |
| LH154_TcR-2kB_P3 | TAAAAGCGCTTGTGTAGCGATATGAATG | P3 primer to amplify tet cassette |
| LH155_strgPRM_lacZ_F | TTGACAGCTAGCTCAGTCCTAGGTATAATTCAC ACAGGAAACAGAATTCTTTAAGAAGGAGATATA CATATGGTTCGTTTTACAA | F primer to add a strong promoteur to lacZ |
| LH156_kan_R-V4 | CCGTATCACTTTTGGCTGATATGGTTGATG | Reverse primer to amplify kan cassette |
| LH159_lacZ-dlt1_P2 | AACTGTTGGGAAGGGCGATC | reverse primer to delete 7 pb at the beginning of of lacZ gene |
| LH160_lacZ-dlt1-tailed_P3 | CCCGCACCGATCGCCCTTCCCAACAGTTGTGAA TGGCGAATGGCGCTTTG | forward primer to delete 7 pb at the beginning of of lacZ gene |
| LH161_lacZ-dlt2_P2 | GCCTTCTCCGCGTGCAGCAG | reverse primer to delete 7 pb at the end of of lacZ gene |
| LH162_lacZ-dlt2-tailed_P3 | CATCGCCATCTGCTGCACGCGGAAGAAGGCTGA ATATCGACGGTTTCCATATGGG | forward primer to delete 7 pb at the end of of lacZ gene |
| LH163_lacZ-dlt3_P2 | CGTCTCTCCAGGTAGCGAAAAG | reverse primer to delete 7 pb at the middle of of lacZ gene |
| LH164_lacZ-dlt3-tailed_P3 | AAAAAATGGCTTTCGCTACTGGAGAGACGTG ATCCTTTGCGAATACGCC | forward primer to delete 7 pb at the middle of of lacZ gene |
| LH222_yraN-baylyi_P1 | CAGCCTTAACCGAGAGGTC | Forward primer to amplify a 1kB fragment uptstream of yraN from BD413 |
| LH223_yraN-baylyi_P2-tail-apra | TGCGGGAGGGCAAGGGCTCCAAGGATCGGCAT AGCTTCTAGAGTTCTGTGC | Reverse primer to amplify a 1kB fragment uptstream of yraN from BD413 |
| LH224_yraN-baylyi_P3-tail-apra | CCGCTCGCCAGTCGATTGGCTGAGCTCATGACC GCTGCCAATGAGGTAC | Forward primer to amplify a 1kB fragment uptstream of yraN from BD413 |
| LH225_yraN-baylyi_P4 | GGAGGAAGAAATTAGCGACGC | Reverse primer to amplify a 1kB fragment uptstream of yraN from BD413 |
| LH228_comM-baylyi_P1 | CCCAATTGGCACAGCATTTG | Forward primer to amplify a 1kB fragment uptstream of comM from BD413 |
| LH229_comM-baylyi_P2-tl-gent | CGGTAAATTGTCACAACGCCCGCAGGTGGCCGC TTACCCCTCAAGAGAATC | Reverse primer to amplify a 1kB fragment uptstream of comM from BD413 |
| LH230_comM-baylyi_P3-tl-gent | TCCGTTTCCACGGTGTGCGTCCATGGGCAAGTG GCAAAATACTCGTAAGCG | Forward primer to amplify a 1kB fragment uptstream of comM from BD413 |
| LH231_comM-baylyi_P4 | GCAGCAGCGACAAAGCTATG | Reverse primer to amplify a 1kB fragment uptstream of comM from BD413 |
| LH232_gentR_F | TTGCCCATGAGCGCACACCGTG | Forward primer to amplify gent cassette |
| LH233_GentR_R | GCCACCTGGCGGCGTTGTG | Reverse primer to amplify gent cassette |
| LH259_ADP1_perM_P1 | GAGGCACCCCACTAATCGC | P1 primer to insert lacZ-atbR cassette in perM from BD413 |
| LH260_ADP1_perM_P2-tl-lacZ | CCTAGGACTGAGCTAGCTGTCAACGTGCATGGC GCTACGTTTGGC | P2 primer to insert lacZ-atbR cassette in perM from BD413 |
| LH261_ADP1_perM_P3-tl-kan | CATCAACCATATCAGCAAAAGTGATACGGGGTA TGCGACACCCATGGC | P3 primer to insert lacZ-atbR cassette in perM from BD413 |
| LH262_ADP1_perM_P4 | GCAGCATCTGCCAAAGAACG | P4 primer to insert lacZ-atbR cassette in perM from BD413 |
| LH280_rpoB_M2/ADP1_12kB_R | GCTGTTACAGGCAAGATTCAACAG | Reverse to amplify 11kB region from M2 for tract experiment |
| LH281_rpoB_M2/ADP1_10kB_F | GTTGGGGGTTTCGAGTCCCTC | Forward to amplify 11kB region from M2 for tract experiment |

Table S2 : Primers.

Supplementary material and methods.

Construction of mutants of genes involved in natural transformation:

45 Genes were disrupted by inserting a resistance cassette into them. To that end, the ~2kB upstream
46 region (PCR product A [PCRA]) and the ~2kB downstream region (PCRC) were amplified by PCR
47 respectively using primer pairs S_X_P1/S_X_P2-tail-Y and S_X_P3-tail-Y/S_X_P4 (where S
48 designates the strain, X the gene to be disrupted and Y the type of resistance cassette). X_P2-tail-Y
49 and X_P3-tail-Y carry ~30-nucleotide sequences complementary to the ends of the Y cassette. This
50 complementarity was used to assemble the PCRA and PCRC with the Y cassette (PCRB, amplified
51 with Y_F/Y_R primer pair) by overlap extension PCR. Overlapping PCR products were naturally
52 transformed in appropriate conditions for each organism and plated on corresponding selective
53 medium.

54

Construction of mutants for comigration marker experiment.

56 Mutants for the co-migration marker experiment were constructed as follows. A first construct was
57 generated by similar manner as described in the previous paragraph. First, a cassette (PCRB)
58 comprising the lacZ gene upstream of the tetracycline resistance cassette which are antisense to each
59 other ("lacZ-tet" cassette) was generated by assembly PCR of the "tet cassette" and "the lacZ gene"
60 using strgPRM_lacZ_F forward primer - which -allowed the addition of a strong promoter upstream
61 of the lacZ gene - and tcR-tail-attn7_F reverse primer. Then, 2kB upstream (PCRA : 2kB-
62 attTn7_P1/attTn7-P2) and downstream (PCRC: TcR-2kB_P3/2kB-attTn7_P4) regions homologous to
63 the integration region of the attn7 site were added to the PCRB "lacZ-tet" cassette by assembly PCR.
64 This "2kB-lacZ-tet-2kB" construct was directly used to transform *A. nosocomialis* strains M2 WT,
65 M2 *yraN::apra*, M2 $\Delta comM$ and M2 $\Delta comM$ *yraN::apra* to form the different recipient
66 "attTn7::lacZ-tet" strains of this experiment. In the same way "attTn7::lacZ-kan" strain was
67 generated only in the genetic background of the M2 WT. DNA from this strain was used as a template
68 to generate the lacZ deletion constructs ("attTn7::lacZ^{DltX}-kan"). A first PCR (PCRA) was performed
69 with the attTn7-2kB_P4 primer and the lacZ-dltX-tailed_P3 primer which was designed to generate
70 a 7 bp deletion in lacZ. This primer also has a complementary tail to the end of the PCRB fragment.
71 The PCRB fragment is generated with the primers 2kB-attTn7_P1 and lacZ-dltX_P2 which is
72 complementary to the tail of the lacZ-dltX-tailed_P3 primer. PCRA and B were assembled by overlap
73 extension PCR using 2kB-attTn7_P1 and attn7-2kB_P4 to finally generate the following
74 construction: 2kB-lacZ^{DltX}-kan-2kB which is used to transform "attn7::lacZ-tet" strains. Finally, the
75 DNAs from these strains were used as a template to generate the tDNA used for the co-migration
76 marker experiment (2kB-lacZ^{Dlt1,2,3}-kan-2kB) by PCR with 2kB-attn7_P1 and attn7-2kB_P4 primers.

77

Strain construction for bacterial two-hybrid experiment.

79 Plasmids pUT18 and pKNT25 were first linearized using the primers pUT18_F/R or pKNT25_F/R
80 respectively. The genes to be cloned into these vectors are amplified by PCR with X-pUT18_F/R
81 (for cloning into pUT18 and X-pKNT25_F/R (for cloning into pKNT25) which allowed to add to
82 their ends, complementary regions to the ends of the linearized vector. All fragments: linearized
83 vectors and inserts are purified on gel and treated with DpnI. The SLIC reaction is then performed in
84 20 μ L of final volume completed with water by adding a 3:1 ratio of insert:vector with 0.5 μ L of T4
85 DNA ligase and 2 μ L of NEB1.2 buffer. The whole is incubated at room temperature for 2 minutes
86 and 30 seconds. The product of the SLIC reaction which theoretically corresponds to the recirculated
87 vector with the insert in it, is directly used for cloning into *E. coli DH5a* competent strains to form
88 strain for the two-hybrid experiment.

89

90

91

B. Expérience complémentaire 1 : Recherche d'interactants potentiels *in vivo*

Les expériences de double hybride chez *E. coli* nous ont permis de déterminer des interactions directes inconnues jusqu'alors entre les protéines de *L. pneumophila* : YraN et DprA, YraN et ComM et entre ComM et DprA. Toutefois, ce système de double hybride chez *E. coli* est limité. C'est un système hétérologue dans lequel on regarde l'interaction de protéines d'une autre espèce dont l'expression est induite chez *E. coli* qui n'est pas naturellement transformable. Nous avons développé d'autres techniques pour confirmer dans des conditions plus natives ces interactions observées en double hybride chez *E. coli*.

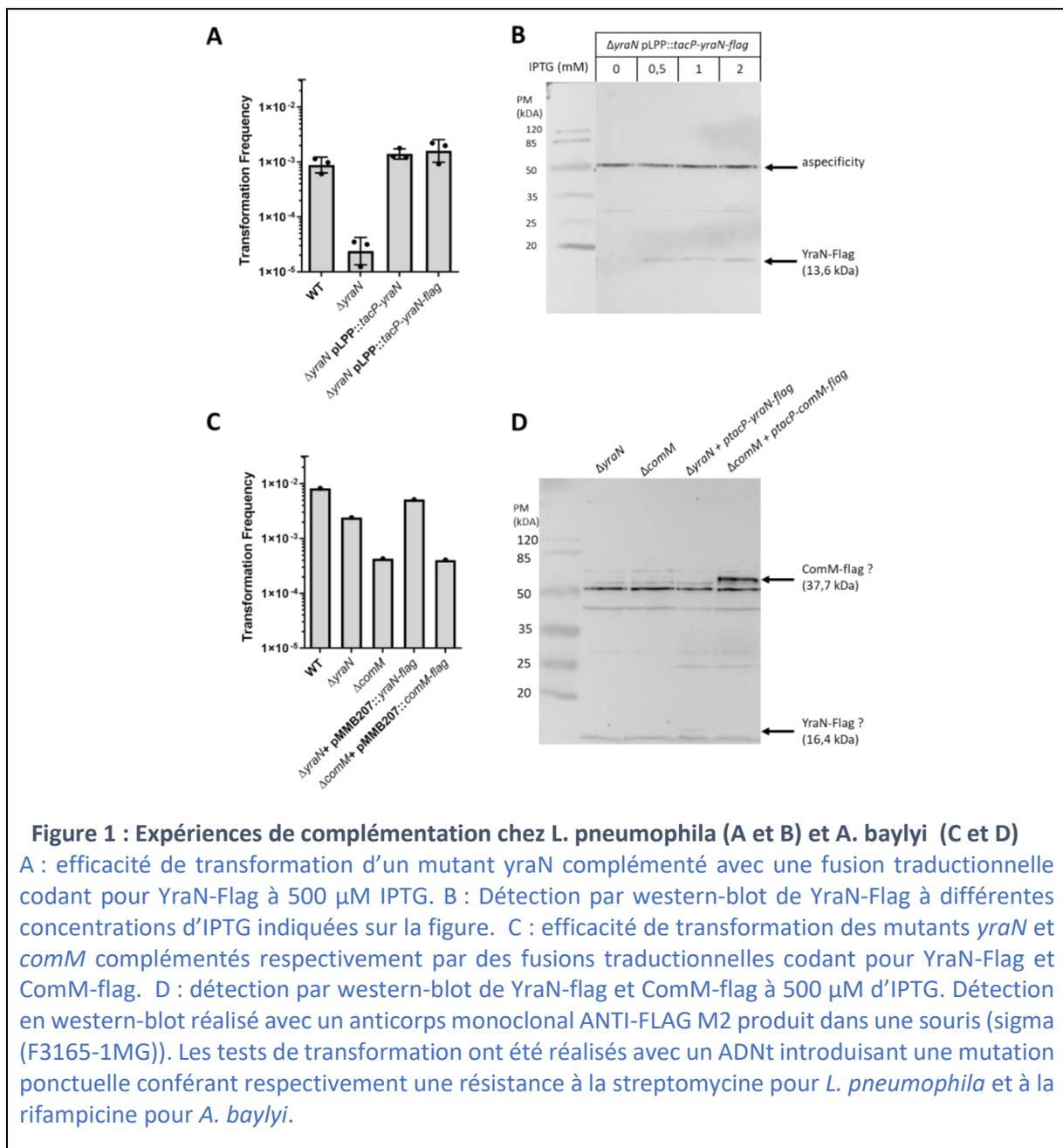
III.B.1. Par co-immunoprécipitation

L'objectif ici est de déterminer par co-immunoprécipitation de YraN fusionné au peptide Flag, les possibles interactants de cette protéine.

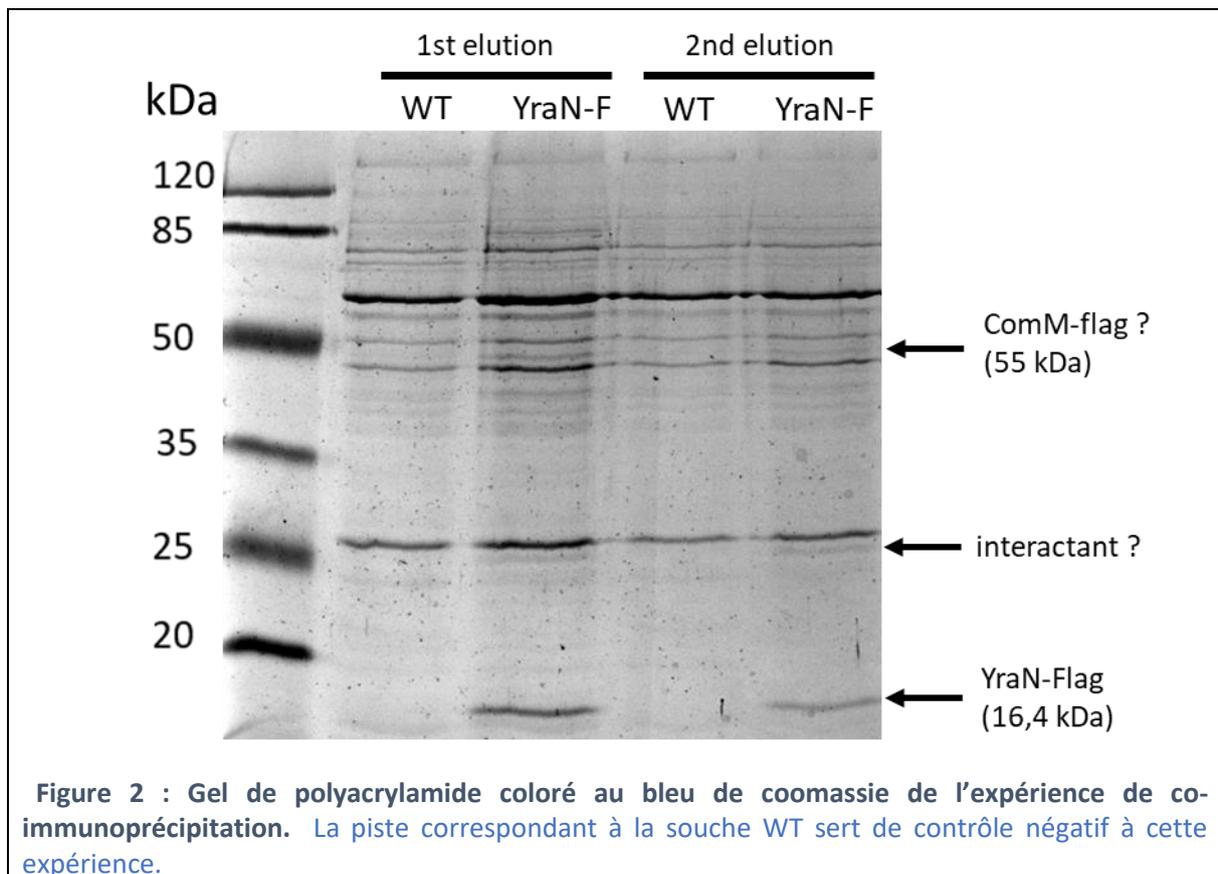
Tout d'abord, j'ai généré chez un mutant *yraN* de *L. pneumophila*, des fusions traductionnelles YraN-flag. En plus d'avoir construit cette fusion YraN-flag chez un mutant *yraN* de *A. baylyi*, j'ai aussi construit une fusion ComM-Flag chez un mutant *comM* de cette espèce. Ces différentes constructions ont été placées respectivement sur le pLPP (plasmide natif) de *L. pneumophila* et un plasmide (pMMB207) introduit chez *A. baylyi*. Ces constructions sont sous le contrôle d'un promoteur pTAC et donc inductibles à l'IPTG. Les expériences de complémentation montrent que les fusions YraN-flag sont fonctionnelles chez les deux espèces (**Figure 1AC**) et détectables par western-blot en utilisant un anticorps anti-flag après induction à l'IPTG (**Figure 1BC**). Comme on l'observe sur les profils de western-blot, pour les deux espèces, YraN semble très faiblement exprimée même à de fortes concentrations d'IPTG (jusqu' 2 mM chez *L. pneumophila*).

La fusion ComM-flag générée chez *A. baylyi* n'est pas fonctionnelle (**Figure 1C**) bien que détectable par western-blot (**Figure 1D**). Dans des données non montrées ici, une fusion ComM-mCherry avait été générée chez *L. pneumophila* et elle aussi n'était pas fonctionnelle. Ces données sont cohérentes avec celles déjà publiées chez *V. cholerae* où une fusion ComM-gfp avait été générée et elle non plus n'était pas fonctionnelle contrairement à une fusion gfp-ComM (Nero et al., 2018, supplementary). Ceci suggère que la partie C-terminal de la protéine est importante pour sa fonction chez les différentes espèces gram négatives. Même si des analyses plus approfondies devront être réalisées afin de comprendre son rôle dans la fonction de ComM, on peut formuler l'hypothèse que cette partie C-terminale puisse jouer un rôle dans l'interaction fortement présumée avec YraN (ou

DprA) ou tout simplement dans l'hexamérisation de ComM. Le signal détecté par western-blot (~50-60kDa) ne correspond pas à la taille attendue de la fusion ComM-Flag (~37,7 kDa). Cette fusion étant non fonctionnelle, on peut se demander si ce n'est pas un artefact dû à sa non-fonctionnalité. Plus audacieusement, on peut formuler l'hypothèse que ce signal puisse dissimuler un complexe de ComM avec une autre protéine, par exemple YraN, puisque la somme des tailles de ComM-flag et d'YraN (54,1 kDa) correspond environ à la taille du signal observé. Cette dernière hypothèse est assez peu probable au vu du traitement subi par les protéines au cours de cette expérience de western-blot (chauffage, agents dénaturants) et contredirait l'hypothèse formulée juste avant selon laquelle la partie C-terminale (non fonctionnelle) servirait à interagir avec d'autres protéines.



Pour réaliser l'expérience de co-immunoprécipitation, des cellules en phase exponentielle d'une souche WT de *L. pneumophila* ou du mutant *yraN* complémentés avec la fusion *yraN-flag* induite avec 500 μ M d'IPTG ont été soumises à une lyse chimique (avec du B-per) puis physique (par des cycles de congélation/décongélation suivie d'une sonication). Les lysats ont été incubés avec des billes magnétiques portant un anticorps anti-flag de façon à capturer les protéines YraN-flag et les possibles interactants selon le principe de la co-immunoprécipitation. Après plusieurs lavages des billes et deux éluations avec le peptide flag, les éluats correspondants ont été analysés sur gel de polyacrylamide révélé au bleu de Coomassie (**Figure 2**). On observe un très grand nombre de bandes sur le gel qui sont présentes aussi bien dans le contrôle négatif (WT) que dans la condition test (*yraN-flag*) témoignant de mauvais lavage ou d'une trop grande aspécificité de l'anticorps anti-flag lors du pull-down. Néanmoins, un signal correspondant à la taille de la fusion YraN-Flag se distingue dans la condition test par rapport à la condition contrôle. Deux autres bandes semblent être détectées dans les conditions test par rapport aux conditions contrôles à environ 55 kDa et 25 kDa. On peut émettre l'hypothèse que la bande observée à 55kDa puisse être ComM. Ces données nécessiteraient une vérification par spectrométrie de masse, mais surtout de refaire l'expérience en créant des fusions avec d'autres tags qui ne généreraient pas autant d'aspécificité.



Finalement, ces données de co-immunoprécipitation n'ont pas été très fructueuses. La faible quantité d'YraN présente dans les cellules, combinée à un tag peu spécifique dans les conditions réalisées ici, sont probablement les raisons principales de cet échec. Dans les expériences futures, il faudrait réitérer ces expériences, peut-être avec d'autres conditions de lyse, à une autre phase de la croissance, en présence d'ADNt et/ou avec un autre tag pour capturer spécifiquement YraN et ses possibles interactants. Une analyse des éluats en spectrométrie de masse augmenterait drastiquement la précision de cette expérience.

III.B.2. Par une expérience de double hybride *in vivo*

Avec l'aide de Chloé Cruzillac, stagiaire de Master 1 que j'ai encadré au cours de ma thèse, nous avons mis au point une méthode de double hybride innovante chez *A. baylyi* récemment développée chez *Clostridium difficile* (Oliveira Paiva et al., 2019). Le principe de cette méthode est similaire à celui du système classique de double hybride chez *E. coli*, sauf que cette méthode n'est pas basée sur la reconstitution de l'adénylate cyclase mais sur la reconstitution d'une luciférase synthétique particulière appelée NanoBit. Cette enzyme peut être divisée en deux sous-unités : une large appelée LgBit (~18 kDa) et une petite appelée SmBit (~1,3kDa) qui ont été optimisées pour réduire leur autoassociation et augmenter leur stabilité (Dixon et al., 2016). Lorsque deux protéines d'intérêt sont taguées avec l'une ou l'autre de ces sous-unités et qu'elles sont proches l'une de l'autre dans la cellule, la NanoBit peut être régénérée par l'interaction des deux sous-unités LgBit et SmBit, ce qui produit de la lumière en présence de son substrat. L'avantage de ce système par rapport au système de double hybride classique chez *E. coli* est qu'il peut être développé directement chez la bactérie d'intérêt pour détecter les interactions *in vivo* et en temps réel.

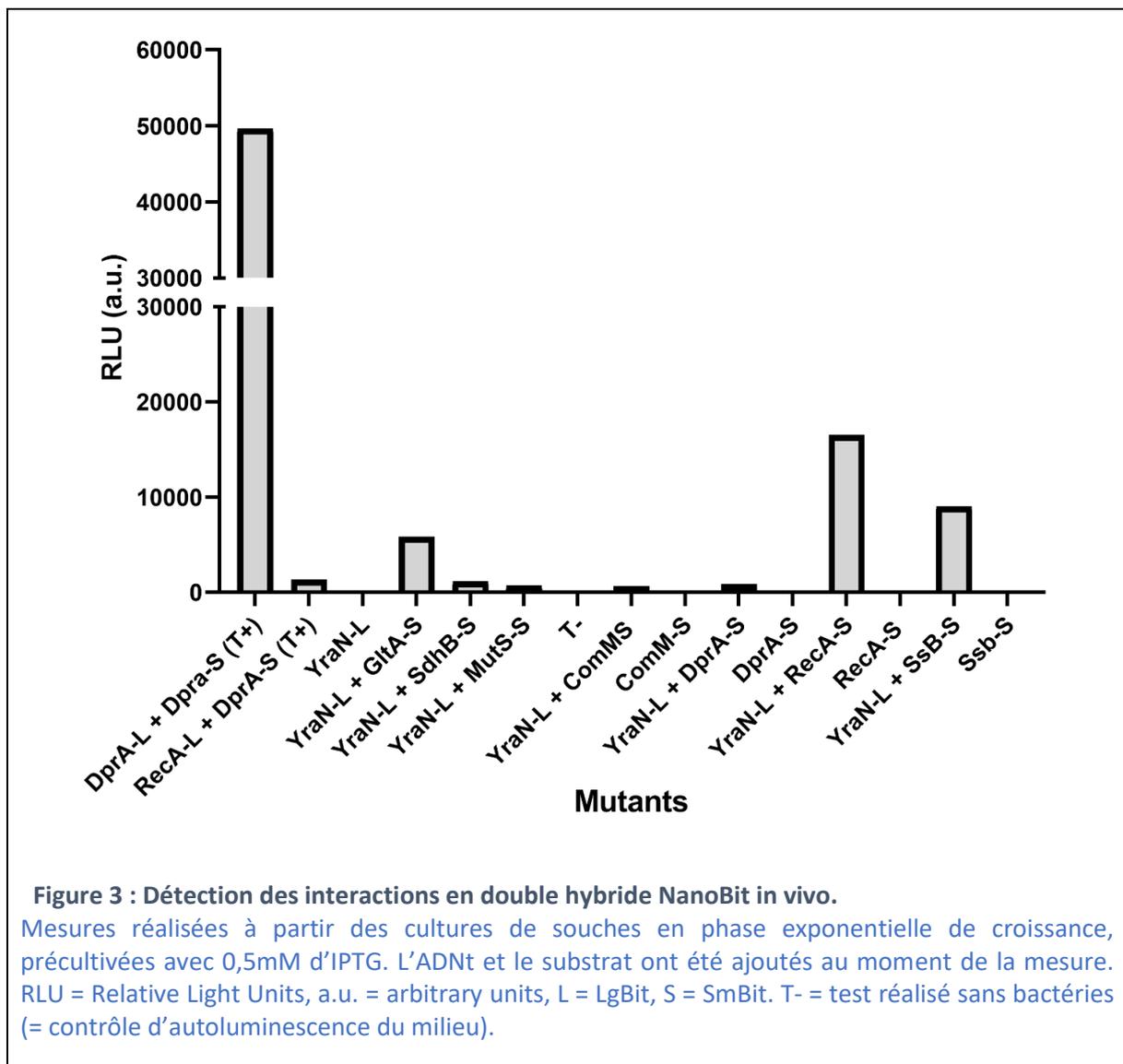
Nous avons développé ce système chez la bactérie transformable *A. baylyi* afin de confirmer les interactions détectées en système hétérologue et particulièrement celles impliquant YraN. Pour cela nous avons construit des fusions traductionnelles des interactants potentiels de YraN avec le SmBit placé en C-terminal à leurs *loci* respectifs et sous induction de leur promoteur natif (souches nommées I-SmBit). Plusieurs interactants (I) potentiels ont été fusionnés et testés en double hybride : ComM, DprA et RecA, pour confirmer les résultats du double hybride classique, mais aussi SsB et LigA pour tenter d'identifier de nouvelles interactions. Une fois que les fusions ont été validées génétiquement, nous avons testé les phénotypes de ces mutants I-SmBit pour vérifier la fonctionnalité des protéines taguées et elles sont bien fonctionnelles (données non montrées). Des fusions contrôles (pour test négatif d'interaction) ont été générées de la même manière en fusionnant le SmBit à des protéines qui, a priori, ne sont pas impliquées dans les voies moléculaires de la transformation à savoir GlbA et

SdhB qui sont impliquées dans le métabolisme énergétique et MutS qui est impliqué dans le métabolisme de l'ADN.

En parallèle nous avons construit des plasmides exprimant, sous induction à l'IPTG, des fusions traductionnelles de YraN taguées en C-terminal avec le LgBit (souches nommées LgBit-Y ou Y-LgBit). Ces fusions ont été introduites dans un mutant *yraN* de *A. baylyi* et des tests de transformation sous induction à l'IPTG (0,5 mM) ont été réalisés pour tester la fonction de ces protéines hybrides et elles sont bien fonctionnelles (données non montrées). Enfin, ces plasmides exprimant les fusions LgBit-Y ou Y-LgBit ont été introduits dans les mutants exprimant les fusions I-SmBit. En parallèle des plasmides exprimant les fusions DprA-LgBit et RecA-LgBit inductibles à l'IPTG ont été générées de la même manière pour être introduits dans les souches exprimant les fusions DprA-SmBit pour servir de contrôle positif d'interaction, les interactions DprA-DprA ou RecA-DprA étant connues.

Des tests d'interactions ont été réalisés avec ces souches hybrides en phase exponentielle de croissance préalablement cultivées en présence d'IPTG (0,5 mM) pour induire les fusions YraN, en présence d'ADNt pour favoriser les interactions avec les protéines potentiellement impliquées dans la recombinaison de l'ADNt et en présence du substrat de la NanoBit (ajouté au moment de la mesure). La luminescence a été mesurée grâce à un luminomètre dédié. Ces tests ont aussi été réalisés avec les souches exprimant uniquement YraN-LgBit ou I-SmBit pour vérifier l'autoluminescence des constructions. Les résultats sont présentés dans la **Figure 3**. Un signal élevé de luminescence relative (RLU) pour le contrôle positif d'interaction DprA-LgBit/DprA-SmBit est détecté de la même façon qu'on avait observé en double hybride classique un signal bleu caractéristique, ce qui confirme *in vivo* la capacité de cette protéine à interagir avec elle-même. En revanche, pour l'interaction RecA-LgBit/DprA-SmBit un faible signal de luminescence est détecté ce qui suggère que dans ces conditions, les protéines interagissent peu ou pas entre elles contrairement à ce qu'on avait pu observer en double hybride classique ou ce qui est connu de la littérature. De façon tout aussi surprenante, un signal de luminescence relativement élevé est observé pour le test d'interaction YraN-LgBit/GltA-SmBit généré à la base pour servir de contrôle négatif d'interaction, GltA étant la citrate synthase impliquée dans le cycle de Krebs et aucun élément ne laisse penser que ces deux protéines puissent interagir. Seuls deux tests d'interaction présentent des niveaux de luminescences supérieures à ce contrôle : les tests d'interaction YraN-LgBit/RecA-SmBit et YraN-LgBit/SsB-SmBit ce qui suggère que ces protéines interagissent potentiellement entre elles *in vivo*. Ces différents tests ont été reproduits plusieurs fois (à des facteurs de RLU près), avec les mêmes résultats avec ou sans ADN, à différentes concentrations d'IPTG ou à différents timing de la croissance. D'autres souches hybrides ont été générées pour tester les interactions de LgBit-YraN/ I-SmBit ou SmBit-YraN/I-LgBit ou encore YraN-SmBit/I-LgBit pour des résultats encore moins convaincants.

Finalement même si par cette méthode, l'interaction de DprA avec elle-même semble confirmée *in vivo* et que deux nouvelles interactions (YraN/RecA et YraN/SsB non testées en double hybride classique) semblent présentes, la non-validité du contrôle négatif nous incite à interpréter ces résultats avec prudence. Même si non observé ici, dans d'autres expériences réalisées au laboratoire, une autoluminescence de mutants exprimant la fusion YraN-LgBit seule a été observée pouvant expliquer ces résultats incohérents. L'induction à l'IPTG, même à faibles concentrations, induit probablement une surproduction de YraN dans les cellules pouvant suffire à générer cette autoluminescence. Dans des expériences futures, il faudrait pouvoir réitérer cette expérience en créant une fusion de *yraN* à son locus sous influence de son propre promoteur et optimiser les conditions de détection de ces interactions en faisant varier, par exemple, plus précisément le timing de croissance, ou la nature et la concentration d'ADNt, car rien n'indique que les protéines impliquées dans la recombinaison de l'ADNt interagissent de façon continue entre elles.



C. Expérience complémentaire 2 : Détection du brin déplacé du D-loop de transformation *in vivo*

Comme mentionné dans l'article en préparation, nous proposons que YraN, par son activité endonucléase, clive le brin déplacé de la D-loop formé par l'invasion de l'ADNt dans l'ADNdb chromosomique (**Figure 4a**). Pour valider cette hypothèse, nous avons imaginé une expérience permettant de détecter indirectement cette D-loop, par reconnaissance du brin déplacé *in vivo*. Ici, je vais décrire uniquement le principe de cette expérience qui est toujours en cours d'optimisation au laboratoire.

Cette expérience est basée sur l'hypothèse préalable que ce brin déplacé ne soit pas totalement dégradé par YraN dont on suggère plutôt qu'elle réalise un ou des nicks au niveau de ce brin déplacé générant une extrémité libre d'ADNsb qui reste attachée au duplex d'ADN chromosomique (comme montré sur le schéma) ou qui est libérée dans le cytoplasme (**Figure 4b**).

Cette expérience est réalisée chez *A. baylyi* qui semble constitutivement transformable et qui est caractérisée par des fréquences élevées de transformation. Après obtention d'une culture en phase exponentielle de croissance, un ADNt qui est homologue à l'ADN du receveur [à ceci près qu'il possède centralement, une délétion (« dlt3 » sur le schéma) de 7pb par rapport à l'ADN receveur] est introduit dans le milieu de culture pour permettre la transformation des cellules bactériennes supposées être compétentes. Directement après introduction de l'ADNt ou à différents temps qu'il reste à déterminer après cette introduction, une fraction de la culture est récupérée et directement placée sur glace pour arrêter le métabolisme cellulaire et espérer bloquer le processus de recombinaison de l'ADNt de façon à ce que le D-loop de recombinaison reste figée dans une configuration similaire à la **Figure 4b**. Sur ces fractions cellulaires récupérées à différents temps après l'introduction d'ADNt une extraction d'ADN au phénol-chloroforme est réalisée dans le but de récupérer tout l'ADN de ces cellules (ADNg/ADNt/ADNsb).

Pour détecter spécifiquement par PCR l'ADN déplacé et le distinguer de l'ADNt ou de l'ADNg nous avons imaginé une méthode basée sur le principe de la méthode TdT décrite en partie I de ce chapitre de Résultats. Dans cette méthode, l'enzyme TdT par son activité 3' déoxynucléotidyl-transférase est utilisée pour ajouter une queue polyC en 3' de tous les ADN extraits des cellules récupérées de l'expérience de transformation (**Figure 4c**). Une amorce biotinylée spécifique du brin déplacé (en noir et mauve sur le schéma) qui est désignée dans la région génomique délétée sur l'ADNt, mais présente sur l'ADN déplacé est utilisée en association avec une amorce polyG (en turquoise sur le schéma) qui s'hybride au niveau du polyC pour amplifier par PCR, spécifiquement, le brin déplacé (**Figure 4d**). Les

amplicons biotinylés sont capturés sélectivement par des billes de streptavidine et sont soumis à un deuxième round de PCR pour (1) enrichir ces fragments correspondant au brin déplacé et (2) amplifier le signal lors de sa visualisation sur gel d'agarose après électrophorèse (**Figure 4e**).

Pour montrer spécifiquement l'implication de YraN dans le clivage de cet ADN déplacé, cette expérience pourra être réalisée chez une souche *WT* en comparaison à un mutant *yraN* pour lequel aucun signal ne devrait être détecté. Si tel est le cas, les amplicons de PCR2 obtenus pour la souche *WT* pourront être séquencés pour identifier, s'il existe, le site de clivage de YraN qui devrait être situé en amont de la queue polyG ainsi générée (flèche « YraN nick » sur le schéma).

L'application de cette méthode fait face à un certain nombre de verrous techniques qui ne sont, pour le moment, pas encore tous levés. Premièrement, même si d'après des expériences préliminaires menées au laboratoire, *A. baylyi* semble transformable à toutes les phases de la croissance, nous devons définir le bon timing durant lequel un maximum de bactéries incorpore l'ADN en même temps, de façon à maximiser nos chances de « figer » l'intermédiaire du D-loop de recombinaison. De plus, il faut optimiser les conditions d'extraction de l'ADN de façon à ne pas perdre cet intermédiaire. Enfin, il faudrait trouver un outil moléculaire permettant d'éliminer spécifiquement les ADNsb issus du métabolisme normal de l'ADN qui sont susceptibles d'interférer avec la détection de l'ADNsb déplacé.

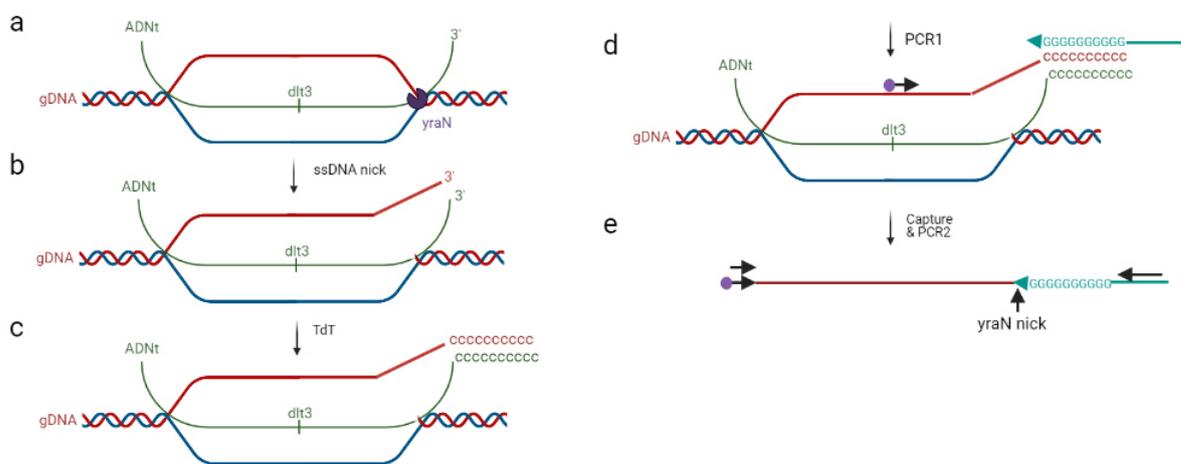


Figure 4 : Détection du brin déplacé de la D-loop de transformation.

Schéma réalisé avec BioRender. Voir description de la méthode dans le texte. En vert : l'ADNt simple brin, en rouge : l'ADN déplacé et en bleu : l'ADN parental.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

À mon arrivée en thèse en 2017, le système de transformation était encore largement incompris, particulièrement chez les bactéries gram négatives. Ces quatre dernières années, des avancées considérables ont été réalisées permettant de mieux comprendre ce mécanisme conservé de THG. Par exemple, bien que présumée auparavant, la rétraction du pilus capturant l'ADNt a été observée en temps réel pour la première fois chez *V. cholerae* en 2018 (Ellison et al., 2018). Cette même année l'hélicase ComM a aussi été décrite comme spécifiquement requise pour le processing du D-loop de recombinaison (Nero et al., 2018). Pour la première fois depuis la description détaillée du rôle de DprA en 2007 chez *S. pneumoniae* (Mortier-Barrière et al., 2007), une nouvelle protéine conservée dédiée aux voies moléculaires de la recombinaison de l'ADNt était ainsi découverte.

Concernant les conditions d'induction et les mécanismes de régulation de la compétence qui peuvent être spécifiques à chaque espèce, beaucoup restent à découvrir puisqu'ils ne sont que partiellement connus chez une petite poignée d'espèces transformables. Si on s'intéresse spécifiquement à *L. pneumophila*, notre organisme d'étude, une importante avancée a été obtenue peu avant ma thèse en 2016 avec la description du système RocC/RocR (Attaiech et al., 2016). Il s'agissait de la première fois qu'un processus de répression post-transcriptionnelle était suggéré comme mécanisme principal d'induction de la compétence, toutes les voies décrites auparavant étant basées sur des activateurs. Les données de Tn-seq issues de cribles de transformation m'ont aussi permis d'identifier de nouveaux potentiels régulateurs spécifiques de la compétence sous la forme du système à deux composantes LetA/LetS qui agirait possiblement en amont de la répression imposée par le système RocC/RocR. Des investigations futures devront être menées pour préciser le rôle de ce système pléiotropique dans la régulation de la compétence. Notamment, il faudra déterminer s'il agit directement sur le système RocC/RocR par l'intermédiaire des petits ARNnc RsmZ/Y ou s'il existe des voies intermédiaires et d'autres régulateurs impliqués dans ce système.

Outre ces découvertes qui sont restreintes à *L. pneumophila*, les données de Tn-seq m'ont permis d'identifier un nouvel acteur majeur du système de transformation conservé chez l'ensemble des bactéries gram négatives : YraN. La caractérisation fonctionnelle de ce nouvel acteur a permis de montrer qu'YraN était une nucléase agissant de concert avec ComM pour permettre l'extension du D-loop de recombinaison au cours de la recombinaison de l'ADNt. Nous suggérons qu'YraN puisse être la pièce manquante de ce système permettant la résolution des jonctions 3J entraînant l'intégration irréversible de l'ADNt dans le chromosome. Outre ce rôle de résolvasse, nous suggérons qu'YraN grâce à son activité nucléase puisse aussi jouer le rôle de relaxase pour faciliter la progression de ComM sur la double hélice. Des recherches futures devront être menées pour préciser ces rôles. Il pourra, par

exemple, être appliquée la méthode de détection de brin déplacé décrite en III.C pour montrer l'activité nucléase de YraN *in vivo*. Plus généralement, au vu de l'homologie fonctionnelle de ComM avec son équivalent RadA, chez les Gram positives, l'existence d'une telle nucléase pourra être explorée chez ces espèces. On peut aussi se poser la question des raisons pour lesquelles ce système nucléase/hélicase a été sélectionné. En effet, son absence ne se traduit pas par une perte totale de transformabilité, mais seulement une réduction d'un facteur 10 à 100. Si des fréquences de transformation suffisamment élevées sont nécessaires pour faire face à la stochasticité environnementale, comme proposé par Carvalho et al, (Carvalho et al., 2020) la réponse pourrait se trouver dans la nature des événements de recombinaison qu'autorise ce système. Produire des tracts de recombinaison étendus, permettrait d'augmenter les chances d'intégrer des polymorphismes avantageux. Il n'est pas exclu également que les conditions *in vitro* ne révèlent pas une nécessité plus importante de YraN/ComM dans d'autres conditions. Compte tenu d'un rôle de relaxase, YraN peut favoriser la transformation dans des conditions où le chromosome est maintenu dans une conformation surenroulée. Cette hypothèse pourrait être testée en augmentant le surenroulement du chromosome (par l'utilisation d'inhibiteur des gyrases par exemple). Une autre possibilité serait que l'hélicase ComM, aidée de YraN, permette l'intégration d'ADNt dans des régions à forte contrainte topologique, causée par les G-quadruplex (G4) ou encore par la présence de complexes nucléoprotéiques (répresseurs transcriptionnels). Ces hypothèses devront être explorées. Enfin, seules des approches intégrées, combinant génétique, mais surtout structurales permettront d'y voir plus clair sur le mécanisme moléculaire de ce système. Un des résultats surprenants de notre analyse Tn-seq des gènes impliqués dans la transformation naturelle de *L. pneumophila* comparé à celle réalisée chez *S. elongatus* est qu'elle n'a pas permis d'identifier d'autres acteurs conservés de ce processus de THG. Cela n'implique pas nécessairement qu'il n'en existe pas d'autres. L'une des limites inhérentes à ces méthodes basée sur l'étude de banque de mutant est qu'elles permettent d'identifier uniquement les gènes non essentiels, les gènes essentiels ne tolérant pas, par définition, les insertions. Ainsi, il est possible, comme c'est le cas de RecA ou SsB, qu'un acteur essentiel à la viabilité cellulaire puisse être impliqué dans la transformation naturelle. Pour les identifier, il faudrait développer des techniques alternatives, comme par exemple l'inactivation conditionnelle par CRISPRi (Liu et al., 2017).

Finalement mes travaux de thèse ont permis d'identifier des pièces manquantes du système de transformation naturelle chez *L. pneumophila* et plus largement des gram négatives. Cependant, il reste encore de nombreuses zones d'ombre aussi bien dans les mécanismes spécifiques de régulation de la compétence que dans le fonctionnement de la machinerie avec un exemple phare de toutes les questions que les chercheurs travaillant sur la transformation se posent : quid de la nucléase qui permet de convertir l'ADNdb en ADNsb lors de son internalisation dans le cytoplasme ?

Cet exemple précis montre qu'on est encore loin d'avoir élucidé entièrement ce mécanisme conservé de transfert horizontal de gène.

RÉFÉRENCES BIBLIOGRAPHIQUES

Adams, David.W., Stutzmann, S., Stoudmann, C., and Blokesch, M. (2019). DNA-uptake pili of *Vibrio cholerae* are required for chitin colonisation and capable of kin recognition via sequence-specific self-interaction. *Nat Microbiol* 4, 1545–1557.

Al-Quadani, T., Price, C.T., and Kwaik, Y.A. (2012). Exploitation of evolutionarily conserved amoeba and mammalian processes by *Legionella*. *Trends Microbiol* 20, 299–306.

Andersson, S.G., and Kurland, C.G. (1998). Reductive evolution of resident genomes. *Trends Microbiol* 6, 263–268.

Ando, T., Israel, D.A., Kusugami, K., and Blaser, M.J. (1999). HP0333, a Member of the *dprA* Family, Is Involved in Natural Transformation in *Helicobacter pylori*. *J Bacteriol* 181, 5572–5580.

Angelov, A., Bergen, P., Nadler, F., Hornburg, P., Lichev, A., Übelacker, M., Pachi, F., Kuster, B., and Liebl, W. (2015). Novel Flp pilus biogenesis-dependent natural transformation. *Frontiers in Microbiology* 6, 84.

Antonova, E.S., and Hammer, B.K. (2015). Genetics of Natural Competence in *Vibrio cholerae* and other *Vibrios*. *Microbiol Spectr* 3.

Assalkhou, R., Balasingham, S., Collins, R.F., Frye, S.A., Davidsen, T., Benam, A.V., Bjørås, M., Derrick, J.P., and Tønjum, T. (2007). The outer membrane secretin PilQ from *Neisseria meningitidis* binds DNA. *Microbiology (Reading)* 153, 1593–1603.

Atherton, J.C. (2006). The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annu Rev Pathol* 1, 63–96.

Attaiech, L., and Charpentier, X. (2017). Silently transformable: the many ways bacteria conceal their built-in capacity of genetic exchange. *Curr Genet* 63, 451–455.

Attaiech, L., Olivier, A., Mortier-Barrière, I., Soulet, A.-L., Granadel, C., Martin, B., Polard, P., and Claverys, J.-P. (2011). Role of the Single-Stranded DNA-Binding Protein SsbB in Pneumococcal Transformation: Maintenance of a Reservoir for Genetic Plasticity. *PLoS Genet* 7.

Attaiech, L., Boughammoura, A., Brochier-Armanet, C., Allatif, O., Peillard-Fiorente, F., Edwards, R.A., Omar, A.R., MacMillan, A.M., Glover, M., and Charpentier, X. (2016). Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8813–8818.

Avery, O.T., Macleod, C.M., and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES: INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J Exp Med* 79, 137–158.

Bachman, M.A., and Swanson, M.S. (2004). The LetE protein enhances expression of multiple LetA/LetS-dependent transmission traits by *Legionella pneumophila*. *Infect Immun* 72, 3284–3293.

Baker, J.A., Simkovic, F., Taylor, H.M.C., and Rigden, D.J. (2016). Potential DNA binding and nuclease functions of ComEC domains characterized in silico. *Proteins* 84, 1431–1442.

Balaban, M., Bättig, P., Muschiol, S., Tirier, S.M., Wartha, F., Normark, S., and Henriques-Normark, B. (2014). Secretion of a pneumococcal type II secretion system pilus correlates with DNA uptake during transformation. *Proc Natl Acad Sci U S A* *111*, E758–E765.

Baltrus, D.A., and Guillemin, K. (2006). Multiple phases of competence occur during the *Helicobacter pylori* growth cycle. *FEMS Microbiol Lett* *255*, 148–155.

Beam, C.E., Saveson, C.J., and Lovett, S.T. (2002). Role for radA/sms in Recombination Intermediate Processing in *Escherichia coli*. *Journal of Bacteriology* *184*, 6836–6844.

Bell, J.C., and Kowalczykowski, S.C. (2016). RecA: Regulation and Mechanism of a Molecular Search Engine. *Trends Biochem. Sci.* *41*, 491–507.

Bergé, M., Moscoso, M., Prudhomme, M., Martin, B., and Claverys, J.-P. (2002). Uptake of transforming DNA in Gram-positive bacteria: a view from *Streptococcus pneumoniae*. *Molecular Microbiology* *45*, 411–421.

Bergé, M., Mortier-Barrière, I., Martin, B., and Claverys, J.-P. (2003). Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Molecular Microbiology* *50*, 527–536.

Bergé, M.J., Kamgoué, A., Martin, B., Polard, P., Campo, N., and Claverys, J.-P. (2013). Midcell Recruitment of the DNA Uptake and Virulence Nuclease, EndA, for Pneumococcal Transformation. *PLoS Pathog* *9*, e1003596.

Berjeaud, J.-M., Chevalier, S., Schlusshuber, M., Portier, E., Loiseau, C., Aucher, W., Lesouhaitier, O., and Verdon, J. (2016). *Legionella pneumophila*: The Paradox of a Highly Sensitive Opportunistic Waterborne Pathogen Able to Persist in the Environment. *Frontiers in Microbiology* *7*, 486.

Berka, R.M., Hahn, J., Albano, M., Draskovic, I., Persuh, M., Cui, X., Sloma, A., Widner, W., and Dubnau, D. (2002). Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Molecular Microbiology* *43*, 1331–1345.

Berry, J.-L., and Pelicic, V. (2015). Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiol Rev* *39*, 1–21.

Berry, J.-L., Cehovin, A., McDowell, M.A., Lea, S.M., and Pelicic, V. (2013). Functional Analysis of the Interdependence between DNA Uptake Sequence and Its Cognate ComP Receptor during Natural Transformation in *Neisseria* Species. *PLOS Genetics* *9*, e1004014.

Berry, J.-L., Xu, Y., Ward, P.N., Lea, S.M., Matthews, S.J., and Pelicic, V. (2016). A Comparative Structure/Function Analysis of Two Type IV Pilin DNA Receptors Defines a Novel Mode of DNA Binding. *Structure* *24*, 926–934.

Bertolla, F., Van Gijsegem, F., Nesme, X., and Simonet, P. (1997). Conditions for natural transformation of *Ralstonia solanacearum*. *Appl Environ Microbiol* *63*, 4965–4968.

Bertolla, F., Frostegård, Å., Brito, B., Nesme, X., and Simonet, P. (1999). During Infection of Its Host, the Plant Pathogen *Ralstonia solanacearum* Naturally Develops a State of Competence and Exchanges Genetic Material. *MPMI* *12*, 467–472.

Bhattacharjee, M.K., Kachlany, S.C., Fine, D.H., and Figurski, D.H. (2001). Nonspecific Adherence and Fibril Biogenesis by *Actinobacillus actinomycetemcomitans*: TadA Protein Is an ATPase. *Journal of Bacteriology* *183*, 5927–5936.

Bhaya, D., Bianco, N.R., Bryant, D., and Grossman, A. (2000). Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. *Molecular Microbiology* *37*, 941–951.

Biais, N., Ladoux, B., Higashi, D., So, M., and Sheetz, M. (2008). Cooperative Retraction of Bundled Type IV Pili Enables Nanonewton Force Generation. *PLoS Biol* *6*, e87.

Bijlsma, J.J.E., and Groisman, E.A. (2003). Making informed decisions: regulatory interactions between two-component systems. *Trends in Microbiology* *11*, 359–366.

Blokesch, M. (2012). Chitin colonization, chitin degradation and chitin-induced natural competence of *Vibrio cholerae* are subject to catabolite repression. *Environmental Microbiology* *14*, 1898–1912.

Blokesch, M. (2016). Natural competence for transformation. *Current Biology* *26*, R1126–R1130.

Bonnin, R.A., Poirel, L., and Nordmann, P. (2012). AbaR-type transposon structures in *Acinetobacter baumannii*. *Journal of Antimicrobial Chemotherapy* *67*, 234–236.

Bossé, J.T., Sinha, S., Schippers, T., Kroll, J.S., Redfield, R.J., and Langford, P.R. (2009). Natural competence in strains of *Actinobacillus pleuropneumoniae*. *FEMS Microbiology Letters* *298*, 124–130.

Bradley, D.E. (1972). Evidence for the retraction of *Pseudomonas aeruginosa* RNA phage pili. *Biochem Biophys Res Commun* *47*, 142–149.

Braus, S.A.G., Short, F.L., Holz, S., Stedman, M.J.M., Gossert, A.D., and Hospenthal, M.K. (2021). The Molecular Basis of FimT-mediated DNA Uptake during Bacterial Natural Transformation.

Burgess, R.R. (2017). Sigma Factors☆. In *Reference Module in Life Sciences*, (Elsevier), p.

Burghout, P., Bootsma, H.J., Kloosterman, T.G., Bijlsma, J.J.E., de Jongh, C.E., Kuipers, O.P., and Hermans, P.W.M. (2007). Search for Genes Essential for Pneumococcal Transformation: the RadA DNA Repair Protein Plays a Role in Genomic Recombination of Donor DNA. *J Bacteriol* *189*, 6540–6550.

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* *21*, 526–540.

Cameron, A.D.S., and Redfield, R.J. (2006). Non-canonical CRP sites control competence regulons in *Escherichia coli* and many other γ -proteobacteria. *Nucleic Acids Res* *34*, 6001–6014.

Campbell, E.A., Choi, S.Y., and Masure, H.R. (1998). A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* *27*, 929–939.

Carrasco, B., Fernández, S., Asai, K., Ogasawara, N., and Alonso, J. (2002). Effect of the recU suppressors *sms* and *subA* on DNA repair and homologous recombination in *Bacillus subtilis*. *Mol Gen Genomics* *266*, 899–906.

Cazalet, C., Rusniok, C., Brüggemann, H., Zidane, N., Magnier, A., Ma, L., Tichit, M., Jarraud, S., Bouchier, C., Vandenesch, F., et al. (2004). Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* *36*, 1165–1173.

Cazalet, C., Jarraud, S., Ghavi-Helm, Y., Kunst, F., Glaser, P., Etienne, and Buchrieser, C. (2008). Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Res* 18, 431–441.

Cehovin, A., Simpson, P.J., McDowell, M.A., Brown, D.R., Noschese, R., Pallett, M., Brady, J., Baldwin, G.S., Lea, S.M., Matthews, S.J., et al. (2013). Specific DNA recognition mediated by a type IV pilin. *PNAS* 110, 3065–3070.

Chan, K.-W., Lee, Y.-J., Wang, C.-H., Huang, H., and Sun, Y.-J. (2009). Single-stranded DNA-binding protein complex from *Helicobacter pylori* suggests an ssDNA-binding surface. *J Mol Biol* 388, 508–519.

Chandler, M.S. (1992). The gene encoding cAMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. *Proc Natl Acad Sci U S A* 89, 1626–1630.

Chang, Y.-W., Rettberg, L.A., Treuner-Lange, A., Iwasa, J., Sjøgaard-Andersen, L., and Jensen, G.J. (2016). Architecture of the type IVa pilus machine. *Science* 351, aad2001.

Charpentier, X., Faucher, S.P., Kalachikov, S., and Shuman, H.A. (2008). Loss of RNase R Induces Competence Development in *Legionella pneumophila*. *J Bacteriol* 190, 8126–8136.

Charpentier, X., Kay, E., Schneider, D., and Shuman, H.A. (2011). Antibiotics and UV radiation induce competence for natural transformation in *Legionella pneumophila*. *J Bacteriol* 193, 1114–1121.

Charpentier, X., Polard, P., and Claverys, J.-P. (2012). Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS? *Curr Opin Microbiol* 15, 570–576.

Chatterjee, S., Almeida, R.P.P., and Lindow, S. (2008). Living in two worlds: the plant and insect lifestyles of *Xylella fastidiosa*. *Annu Rev Phytopathol* 46, 243–271.

Chaussee, M.S., and Hill, S.A. (1998). Formation of Single-Stranded DNA during DNA Transformation of *Neisseria gonorrhoeae*. *J Bacteriol* 180, 5117–5122.

Chen, I., and Gotschlich, E.C. (2001). ComE, a Competence Protein from *Neisseria gonorrhoeae* with DNA-Binding Activity. *J Bacteriol* 183, 3160–3168.

Chen, Z., Yang, H., and Pavletich, N.P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453, 489–484.

Chung, Y.S., and Dubnau, D. (1995). ComC is required for the processing and translocation of comGC, a pilin-like competence protein of *Bacillus subtilis*. *Mol Microbiol* 15, 543–551.

Chung, Y.S., and Dubnau, D. (1998). All Seven comG Open Reading Frames Are Required for DNA Binding during Transformation of Competent *Bacillus subtilis*. *J Bacteriol* 180, 41–45.

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* 311, 1283–1287.

Clark, A.J., and Margulies, A.D. (1965). ISOLATION AND CHARACTERIZATION OF RECOMBINATION-DEFICIENT MUTANTS OF *ESCHERICHIA COLI* K12*. *Proc Natl Acad Sci U S A* 53, 451–459.

Claverys, J.-P., Prudhomme, M., and Martin, B. (2006). Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* 60, 451–475.

Claverys, J.-P., Martin, B., and Håvarstein, L.S. (2007). Competence-induced fratricide in streptococci. *Mol Microbiol* *64*, 1423–1433.

Claverys, J.-P., Martin, B., and Polard, P. (2009). The genetic transformation machinery: composition, localization, and mechanism. *FEMS Microbiol Rev* *33*, 643–656.

Cooper, D.L., and Lovett, S.T. (2016). Recombinational branch migration by the RadA/Sms paralog of RecA in *Escherichia coli*. *Elife* *5*, e10807.

Correia, A.M., Ferreira, J.S., Borges, V., Nunes, A., Gomes, B., Capucho, R., Gonçalves, J., Antunes, D.M., Almeida, S., Mendes, A., et al. (2016). Probable Person-to-Person Transmission of Legionnaires' Disease. *N Engl J Med* *374*, 497–498.

Cox, M.M. (2003). The bacterial RecA protein as a motor protein. *Annu Rev Microbiol* *57*, 551–577.

Cox, M.M. (2007). Motoring along with the bacterial RecA protein. *Nat Rev Mol Cell Biol* *8*, 127–138.

Craig, L., Taylor, R.K., Pique, M.E., Adair, B.D., Arvai, A.S., Singh, M., Lloyd, S.J., Shin, D.S., Getzoff, E.D., Yeager, M., et al. (2003). Type IV Pilin Structure and Assembly: X-Ray and EM Analyses of *Vibrio cholerae* Toxin-Coregulated Pilus and *Pseudomonas aeruginosa* PAK Pilin. *Molecular Cell* *11*, 1139–1150.

Craig, L., Pique, M.E., and Tainer, J.A. (2004). Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* *2*, 363–378.

Craig, L., Volkmann, N., Arvai, A.S., Pique, M.E., Yeager, M., Egelman, E.H., and Tainer, J.A. (2006). Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions. *Molecular Cell* *23*, 651–662.

Craig, L., Forest, K.T., and Maier, B. (2019). Type IV pili: dynamics, biophysics and functional consequences. *Nature Reviews Microbiology* *17*, 429–440.

Cruze, J.A., Singer, J.T., and Finnerty, W.R. (1979). Conditions for quantitative transformation in *Acinetobacter calcoaceticus*. *Current Microbiology* *3*, 129–132.

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Dev Biol* *278*, 274–288.

Dalia, A.B., Lazinski, D.W., and Camilli, A. (2014). Identification of a Membrane-Bound Transcriptional Regulator That Links Chitin and Natural Competence in *Vibrio cholerae*. *MBio* *5*, e01028-13.

Damke, P.P., Di Guilmi, A.M., Varela, P.F., Velours, C., Marsin, S., Veaute, X., Machouri, M., Gunjal, G.V., Rao, D.N., Charbonnier, J.-B., et al. (2019). Identification of the periplasmic DNA receptor for natural transformation of *Helicobacter pylori*. *Nat Commun* *10*, 5357.

Damke, P.P., Celma, L., Kondekar, S., Guilmi, A.M.D., Marsin, S., Dépagne, J., Veaute, X., Legrand, P., Walbott, H., Vercruyssen, J., et al. (2021). ComF is a key mediator in single-stranded DNA transport and handling during natural transformation.

D'Auria, G., Jiménez-Hernández, N., Peris-Bondia, F., Moya, A., and Latorre, A. (2010). *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics* *11*, 181.

David, S., Rusniok, C., Mentasti, M., Gomez-Valero, L., Harris, S.R., Lechat, P., Lees, J., Ginevra, C., Glaser, P., Ma, L., et al. (2016). Multiple major disease-associated clones of *Legionella pneumophila* have emerged recently and independently. *Genome Res* 26, 1555–1564.

David, S., Sánchez-Busó, L., Harris, S.R., Marttinen, P., Rusniok, C., Buchrieser, C., Harrison, T.G., and Parkhill, J. (2017). Dynamics and impact of homologous recombination on the evolution of *Legionella pneumophila*. *PLoS Genet* 13.

Davis, B.D. (1950). NONFILTRABILITY OF THE AGENTS OF GENETIC RECOMBINATION IN *ESCHERICHIA COLI*. *J Bacteriol* 60, 507–508.

Declerck, P. (2010). Biofilms: the environmental playground of *Legionella pneumophila*. *Environ Microbiol* 12, 557–566.

Del Val, E., Nasser, W., Abaibou, H., and Reverchon, S. (2019). RecA and DNA recombination: a review of molecular mechanisms. *Biochem Soc Trans* 47, 1511–1531.

Demoulin, C.F., Lara, Y.J., Cornet, L., François, C., Baurain, D., Wilmotte, A., and Javaux, E.J. (2019). Cyanobacteria evolution: Insight from the fossil record. *Free Radical Biology and Medicine* 140, 206–223.

Denise, R., Abby, S.S., and Rocha, E.P.C. (2019). Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLOS Biology* 17, e3000390.

Deutscher, J., Francke, C., and Postma, P.W. (2006). How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiology and Molecular Biology Reviews* 70, 939–1031.

Diallo, A., Foster, H.R., Gromek, K.A., Perry, T.N., Dujeancourt, A., Krasteva, P.V., Gubellini, F., Falbel, T.G., Burton, B.M., and Fronzes, R. (2017). Bacterial transformation: ComFA is a DNA-dependent ATPase that forms complexes with ComFC and DprA. *Molecular Microbiology* 105, 741–754.

Dixon, A.S., Schwinn, M.K., Hall, M.P., Zimmerman, K., Otto, P., Lubben, T.H., Butler, B.L., Binkowski, B.F., Machleidt, T., Kirkland, T.A., et al. (2016). NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* 11, 400–408.

Doherty, A.J., Serpell, L.C., and Ponting, C.P. (1996). The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res* 24, 2488–2497.

Doleans, A., Aurell, H., Reyrolle, M., Lina, G., Freney, J., Vandenesch, F., Etienne, J., and Jarraud, S. (2004). Clinical and Environmental Distributions of *Legionella* Strains in France Are Different. *J Clin Microbiol* 42, 458–460.

Doolittle, W.F. (1999). Phylogenetic classification and the universal tree. *Science* 284, 2124–2129.

Dorer, M.S., Fero, J., and Salama, N.R. (2010). DNA damage triggers genetic exchange in *Helicobacter pylori*. *PLoS Pathog* 6, e1001026.

Dorocicz, I.R., Williams, P.M., and Redfield, R.J. (1993). The *Haemophilus influenzae* adenylate cyclase gene: cloning, sequence, and essential role in competence. *J Bacteriol* 175, 7142–7149.

Drake, S.L., and Koomey, M. (1995). The product of the *pilQ* gene is essential for the biogenesis of type IV pili in *Neisseria gonorrhoeae*. *Molecular Microbiology* 18, 975–986.

Draskovic, I., and Dubnau, D. (2005). Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol Microbiol* 55.

Dreyfus, L.A., and Iglewski, B.H. (1985). Conjugation-mediated genetic exchange in *Legionella pneumophila*. *J Bacteriol* 161, 80–84.

Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiol Rev* 55, 395–424.

Dubnau, D., and Blokesch, M. (2019). Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* 53, 217–237.

Duffin, P.M., and Barber, D.A. (2016). DprA is required for natural transformation and affects pilin variation in *Neisseria gonorrhoeae*. *Microbiology (Reading)* 162, 1620–1628.

Durieux, I., Ginevra, C., Attaiech, L., Picq, K., Juan, P.-A., Jarraud, S., and Charpentier, X. (2019). Diverse conjugative elements silence natural transformation in *Legionella* species. *PNAS* 116, 18613–18618.

Dwivedi, G.R., Srikanth, K.D., Anand, P., Naikoo, J., Srilatha, N.S., and Rao, D.N. (2015). Insights into the Functional Roles of N-Terminal and C-Terminal Domains of *Helicobacter pylori* DprA. *PLOS ONE* 10, e0131116.

Edwards, M.T., Fry, N.K., and Harrison, T.G. (2008). Clonal population structure of *Legionella pneumophila* inferred from allelic profiling. *Microbiology (Reading)* 154, 852–864.

Egelman, E.H., and Stasiak, A. (1986). Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP. *J Mol Biol* 191, 677–697.

Elieh-Ali-Komi, D., and Hamblin, M.R. (2016). Chitin and Chitosan: Production and Application of Versatile Biomedical Nanomaterials. *Int J Adv Res (Indore)* 4, 411–427.

Ellison, C.K., Kan, J., Dillard, R.S., Kysela, D.T., Ducret, A., Berne, C., Hampton, C.M., Ke, Z., Wright, E.R., Biais, N., et al. (2017). Obstruction of pilus retraction stimulates bacterial surface sensing*. *Science* 358, 535–538.

Ellison, C.K., Dalia, T.N., Vidal Ceballos, A., Wang, J.C.-Y., Biais, N., Brun, Y.V., and Dalia, A.B. (2018). Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nature Microbiology*.

Ellison, C.K., Dalia, T.N., Dalia, A.B., and Brun, Y.V. (2019a). Real-time microscopy and physical perturbation of bacterial pili using maleimide-conjugated molecules. *Nat Protoc* 14, 1803–1819.

Ellison, C.K., Kan, J., Chlebek, J.L., Hummels, K.R., Panis, G., Viollier, P.H., Biais, N., Dalia, A.B., and Brun, Y.V. (2019b). A bifunctional ATPase drives tad pilus extension and retraction. *Sci Adv* 5, eaay2591.

Ellison, C.K., Dalia, T.N., Klancher, C.A., Shaevitz, J.W., Gitai, Z., and Dalia, A.B. (2021). *Acinetobacter baylyi* regulates type IV pilus synthesis by employing two extension motors and a motor protein inhibitor. *Nat Commun* 12, 3744.

Escoll, P., Rolando, M., Gomez-Valero, L., and Buchrieser, C. (2013). From amoeba to macrophages: exploring the molecular mechanisms of *Legionella pneumophila* infection in both hosts. *Curr Top Microbiol Immunol* 376, 1–34.

Facius, D., and Meyer, T.F. (1993). A novel determinant (comA) essential for natural transformation competence in *Neisseria gonorrhoeae* and the effect of a comA defect on pilin variation. *Mol Microbiol* 10, 699–712.

Fernandez-Gonzalez, E., and Backert, S. (2014). DNA transfer in the gastric pathogen *Helicobacter pylori*. *J Gastroenterol* 49, 594–604.

Fields, B.S. (1996). The molecular ecology of legionellae. *Trends Microbiol* 4, 286–290.

Finsel, I., and Hilbi, H. (2015). Formation of a pathogen vacuole according to *Legionella pneumophila*: how to kill one bird with many stones. *Cellular Microbiology* 17, 935–950.

Frangipani, E., Visaggio, D., Heeb, S., Kaever, V., Cámara, M., Visca, P., and Imperi, F. (2014). The Gac/Rsm and cyclic-di-GMP signalling networks coordinately regulate iron uptake in *Pseudomonas aeruginosa*. *Environ Microbiol* 16, 676–688.

Fraser, D.W., Tsai, T.R., Orenstein, W., Parkin, W.E., Beecham, H.J., Sharrar, R.G., Harris, J., Mallison, G.F., Martin, S.M., McDade, J.E., et al. (1977). Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med* 297, 1189–1197.

Friedrich, A., Hartsch, T., and Averhoff, B. (2001). Natural Transformation in Mesophilic and Thermophilic Bacteria: Identification and Characterization of Novel, Closely Related Competence Genes in *Acinetobacter* sp. Strain BD413 and *Thermus thermophilus* HB27. *Appl Environ Microbiol* 67, 3140–3148.

Friedrich, A., Prust, C., Hartsch, T., Henne, A., and Averhoff, B. (2002). Molecular Analyses of the Natural Transformation Machinery and Identification of Pilus Structures in the Extremely Thermophilic Bacterium *Thermus thermophilus* Strain HB27. *Applied and Environmental Microbiology* 68, 745–755.

Gal-Mor, O., and Segal, G. (2003). The *Legionella pneumophila* GacA homolog (LetA) is involved in the regulation of *icm* virulence genes and is required for intracellular multiplication in *Acanthamoeba castellanii*. *Microb. Pathog.* 34, 187–194.

Gangel, H., Hepp, C., Müller, S., Oldewurtel, E.R., Aas, F.E., Koomey, M., and Maier, B. (2014). Concerted Spatio-Temporal Dynamics of Imported DNA and ComE DNA Uptake Protein during Gonococcal Transformation. *PLOS Pathogens* 10, e1004043.

Giltner, C.L., Nguyen, Y., and Burrows, L.L. (2012). Type IV Pilin Proteins: Versatile Molecular Modules. *Microbiol Mol Biol Rev* 76, 740–772.

Go, M.F., Kapur, V., Graham, D.Y., and Musser, J.M. (1996). Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. *J Bacteriol* 178, 3934–3938.

Godeux, A.-S., Svedholm, E., Lupo, A., Haenni, M., Venner, S., Laaberki, M.-H., and Charpentier, X. (2020). Scarless Removal of Large Resistance Island *AbaR* Results in Antibiotic Susceptibility and Increased Natural Transformability in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* 64, e00951-20.

Gold, V.A., Salzer, R., Averhoff, B., and Kühlbrandt, W. (2015). Structure of a type IV pilus machinery in the open and closed state. *ELife* 4, e07380.

Gomez-Valero, L., and Buchrieser, C. (2013). Genome Dynamics in Legionella: The Basis of Versatility and Adaptation to Intracellular Replication. *Cold Spring Harb Perspect Med* 3, a009993.

Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., Medigue, C., Etienne, J., and Buchrieser, C. (2011). Extensive recombination events and horizontal gene transfer shaped the Legionella pneumophila genomes. *BMC Genomics* 12, 536.

Gomez-Valero, L., Rusniok, C., Carson, D., Mondino, S., Pérez-Cobas, A.E., Rolando, M., Pasricha, S., Reuter, S., Demirtas, J., Crumbach, J., et al. (2019). More than 18,000 effectors in the Legionella genus genome provide multiple, independent combinations for replication in human cells. *PNAS* 116, 2265–2273.

Goodman, S.D., and Scocca, J.J. (1988). Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae. *PNAS* 85, 6982–6986.

Graupner, S., Weger, N., Sohni, M., and Wackernagel, W. (2001). Requirement of Novel Competence Genes pilT and pilU of Pseudomonas stutzeri for Natural Transformation and Suppression of pilT Deficiency by a Hexahistidine Tag on the Type IV Pilus Protein PilAI. *J Bacteriol* 183, 4694–4701.

Griffith, F. (1928). The Significance of Pneumococcal Types. *J Hyg (Lond)* 27, 113–159.

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. (2000). Bacterial conjugation. An Introduction to Genetic Analysis. 7th Edition.

Grove, D.E., Willcox, S., Griffith, J.D., and Bryant, F.R. (2005). Differential single-stranded DNA binding properties of the paralogous SsbA and SsbB proteins from Streptococcus pneumoniae. *J Biol Chem* 280, 11067–11073.

Guzmán-Herrador, D.L., Steiner, S., Alperi, A., González-Prieto, C., Roy, C.R., and Llosa, M. (2017). DNA Delivery and Genomic Integration into Mammalian Target Cells through Type IV A and B Secretion Systems of Human Pathogens. *Front Microbiol* 8, 1503.

Gwinn, M.L., Ramanathan, R., Smith, H.O., and Tomb, J.-F. (1998). A New Transformation-Deficient Mutant of Haemophilus influenzae Rd with Normal DNA Uptake. *J Bacteriol* 180, 746–748.

Hahn, J., Maier, B., Haijema, B.J., Sheetz, M., and Dubnau, D. (2005). Transformation proteins and DNA uptake localize to the cell poles in Bacillus subtilis. *Cell* 122, 59–71.

Hahn, J., DeSantis, M., and Dubnau, D. (2021). Mechanisms of Transforming DNA Uptake to the Periplasm of Bacillus subtilis. *MBio* 12, e0106121.

Hall, B.G., and Barlow, M. (2004). Evolution of the serine beta-lactamases: past, present and future. *Drug Resist Updat* 7, 111–123.

Hamilton, H.L., and Dillard, J.P. (2006). Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Mol Microbiol* 59, 376–385.

Hammer, B.K., Tateda, E.S., and Swanson, M.S. (2002). A two-component regulator induces the transmission phenotype of stationary-phase Legionella pneumophila. *Mol. Microbiol.* 44, 107–118.

Hamoen, L.W., Smits, W.K., de Jong, A., Holsappel, S., and Kuipers, O.P. (2002). Improving the predictive value of the competence transcription factor (ComK) binding site in Bacillus subtilis using a genomic approach. *Nucleic Acids Res* 30, 5517–5528.

Han, X., Kennan, R.M., Davies, J.K., Reddacliff, L.A., Dhungyel, O.P., Whittington, R.J., Turnbull, L., Whitchurch, C.B., and Rood, J.I. (2008). Twitching Motility Is Essential for Virulence in *Dichelobacter nodosus*. *J Bacteriol* *190*, 3323–3335.

Harding, C.M., Tracy, E.N., Carruthers, M.D., Rather, P.N., Actis, L.A., and Munson, R.S. (2013). *Acinetobacter baumannii* Strain M2 Produces Type IV Pili Which Play a Role in Natural Transformation and Twitching Motility but Not Surface-Associated Motility. *MBio* *4*, e00360-13.

Hardy, L., Juan, P.-A., Coupat-Goutaland, B., and Charpentier, X. (2021). Transposon Insertion Sequencing in a Clinical Isolate of *Legionella pneumophila* Identifies Essential Genes and Determinants of Natural Transformation. *J Bacteriol* *203*.

Hepp, C., and Maier, B. (2016). Kinetics of DNA uptake during transformation provide evidence for a translocation ratchet mechanism. *Proc Natl Acad Sci U S A* *113*, 12467–12472.

Hepp, C., and Maier, B. (2017). Bacterial Translocation Ratchets: Shared Physical Principles with Different Molecular Implementations. *BioEssays* *39*, 1700099.

Herren, C.D., Mitra, A., Palaniyandi, S.K., Coleman, A., Elankumaran, S., and Mukhopadhyay, S. (2006). The BarA-UvrY two-component system regulates virulence in avian pathogenic *Escherichia coli* O78:K80:H9. *Infect Immun* *74*, 4900–4909.

Herriott, R.M., Meyer, E.M., and Vogt, M. (1970). Defined Nongrowth Media for Stage II Development of Competence in *Haemophilus influenzae*. *J Bacteriol* *101*, 517–524.

Hilario, E., and Gogarten, J.P. (1993). Horizontal transfer of ATPase genes--the tree of life becomes a net of life. *Biosystems* *31*, 111–119.

Hoffman, P., Friedman, H., and Bendinelli, M. (2007). *Legionella Pneumophila: Pathogenesis and Immunity* (Springer Science & Business Media).

Hofreuter, D., Odenbreit, S., Henke, G., and Haas, R. (1998). Natural competence for DNA transformation in *Helicobacter pylori*: identification and genetic characterization of the *comB* locus. *Molecular Microbiology* *28*, 1027–1038.

Hofreuter, D., Odenbreit, S., and Haas, R. (2001). Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Molecular Microbiology* *41*, 379–391.

Hofreuter, D., Karnholz, A., and Haas, R. (2003). Topology and membrane interaction of *Helicobacter pylori* ComB proteins involved in natural transformation competence. *Int J Med Microbiol* *293*, 153–165.

Horwitz, M.A. (1983). The Legionnaires' disease bacterium (*Legionella pneumophila*) inhibits phagosome-lysosome fusion in human monocytes. *J Exp Med* *158*, 2108–2126.

Hovland, E., Beyene, G.T., Frye, S.A., Homberset, H., Balasingham, S.V., Gómez-Muñoz, M., Derrick, J.P., Tønjum, T., and Ambur, O.H. (2017). DprA from *Neisseria meningitidis*: properties and role in natural competence for transformation. *Microbiology (Reading)* *163*, 1016–1029.

Huang, L., Tian, X., Liu, M., Wang, M., Biville, F., Cheng, A., Zhu, D., Jia, R., Chen, S., Zhao, X., et al. (2019). DprA Is Essential for Natural Competence in *Riemerella anatipestifer* and Has a Conserved Evolutionary Mechanism. *Frontiers in Genetics* *10*, 429.

Imam, S., Chen, Z., Roos, D.S., and Pohlschröder, M. (2011). Identification of Surprisingly Diverse Type IV Pili, across a Broad Range of Gram-Positive Bacteria. *PLOS ONE* 6, e28919.

Inamine, G.S., and Dubnau, D. (1995). ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J Bacteriol* 177, 3045–3051.

Isberg, R.R., O'Connor, T.J., and Heidtman, M. (2009). The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* 7, 13–24.

Ithurbide, S., Coste, G., Lisboa, J., Eugénie, N., Bentchikou, E., Bouthier de la Tour, C., Liger, D., Confalonieri, F., Sommer, S., Quevillon-Cheruel, S., et al. (2020). Natural Transformation in *Deinococcus radiodurans*: A Genetic Analysis Reveals the Major Roles of DprA, DdrB, RecA, RecF, and RecO Proteins. *Frontiers in Microbiology* 11, 1253.

Jahan, S. (2016). *Cholera – Epidemiology, Prevention and Control* (IntechOpen).

Jeong, K.C., Ghosal, D., Chang, Y.-W., Jensen, G.J., and Vogel, J.P. (2017). Polar delivery of *Legionella* type IV secretion system substrates is essential for virulence. *Proc Natl Acad Sci U S A* 114, 8077–8082.

Johnson, C.H., Mori, T., and Xu, Y. (2008). A Cyanobacterial Circadian Clockwork. *Curr Biol* 18, R816–R825.

Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* 12, 181–196.

Joseph, S.J., Cox, D., Wolff, B., Morrison, S.S., Kozak-Muiznieks, N.A., Frace, M., Didelot, X., Castillo-Ramirez, S., Winchell, J., Read, T.D., et al. (2016). Dynamics of genome change among *Legionella* species. *Sci Rep* 6, 33442.

Kandel, P.P., Lopez, S.M., Almeida, R.P.P., and De La Fuente, L. (2016). Natural Competence of *Xylella fastidiosa* Occurs at a High Frequency Inside Microfluidic Chambers Mimicking the Bacterium's Natural Habitats. *Appl Environ Microbiol* 82, 5269–5277.

Karnholz, A., Hoefler, C., Odenbreit, S., Fischer, W., Hofreuter, D., and Haas, R. (2006). Functional and Topological Characterization of Novel Components of the comB DNA Transformation Competence System in *Helicobacter pylori*. *J Bacteriol* 188, 882–893.

Karudapuram, S., Zhao, X., and Barcak, G.J. (1995). DNA sequence and characterization of *Haemophilus influenzae* dprA+, a gene required for chromosomal but not plasmid DNA transformation. *J Bacteriol* 177, 3235–3240.

Kaufenstein, M., van der Laan, M., and Graumann, P.L. (2011). The Three-Layered DNA Uptake Machinery at the Cell Pole in Competent *Bacillus subtilis* Cells Is a Stable Complex. *Journal of Bacteriology* 193, 1633–1642.

Kay, E., Dubuis, C., and Haas, D. (2005). Three small RNAs jointly ensure secondary metabolism and biocontrol in *Pseudomonas fluorescens* CHA0. *Proc Natl Acad Sci U S A* 102, 17136–17141.

Kidane, D., Ayora, S., Sweasy, J.B., Graumann, P.L., and Alonso, J.C. (2012). The cell pole: the site of cross talk between the DNA uptake and genetic recombination machinery. *Crit Rev Biochem Mol Biol* 47, 531–555.

Koomey, J.M., and Falkow, S. (1987). Cloning of the *recA* gene of *Neisseria gonorrhoeae* and construction of gonococcal *recA* mutants. *J Bacteriol* *169*, 790–795.

Koonin, E.V. (2016). Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Res* *5*, F1000 Faculty Rev-1805.

Kowalczykowski, S.C., and Krupp, R.A. (1987). Effects of *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA protein: Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. *Journal of Molecular Biology* *193*, 97–113.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W.M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* *58*, 401–465.

Kramer, N., Hahn, J., and Dubnau, D. (2007). Multiple interactions among the competence proteins of *Bacillus subtilis*. *Molecular Microbiology* *65*, 454–464.

Kubori, T., and Nagai, H. (2016). The Type IVB secretion system: an enigmatic chimera. *Curr Opin Microbiol* *29*, 22–29.

Kung, S.H., and Almeida, R.P.P. (2011). Natural Competence and Recombination in the Plant Pathogen *Xylella fastidiosa*. *Appl Environ Microbiol* *77*, 5278–5284.

Kurre, R., Höne, A., Clausen, M., Meel, C., and Maier, B. (2012). PilT2 enhances the speed of gonococcal type IV pilus retraction and of twitching motility. *Mol Microbiol* *86*, 857–865.

Kurushima, J., Campo, N., van Raaphorst, R., Cerckel, G., Polard, P., and Veening, J.-W. (2020). Unbiased homeologous recombination during pneumococcal transformation allows for multiple chromosomal integration events. *ELife* *9*, e58771.

Lacks, S. (1962). Molecular fate of DNA in genetic transformation of *Pneumococcus*. *J Mol Biol* *5*, 119–131.

Lacks, S., Greenberg, B., and Neuberger, M. (1974). Role of a Deoxyribonuclease in the Genetic Transformation of *Diplococcus pneumoniae*. *Proc Natl Acad Sci U S A* *71*, 2305–2309.

Lam, T., Ellison, C.K., Eddington, D.T., Brun, Y.V., Dalia, A.B., and Morrison, D.A. (2021). Competence pili in *Streptococcus pneumoniae* are highly dynamic structures that retract to promote DNA uptake. *Molecular Microbiology* *116*, 381–396.

Laurenceau, R., Péhau-Arnaudet, G., Baconnais, S., Gault, J., Malosse, C., Dujeancourt, A., Campo, N., Chamot-Rooke, J., Cam, E.L., Claverys, J.-P., et al. (2013). A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*. *PLoS Pathogens* *9*.

Laurenceau, R., Krasteva, P.V., Diallo, A., Ouarti, S., Duchateau, M., Malosse, C., Chamot-Rooke, J., and Fronzes, R. (2015). Conserved *Streptococcus pneumoniae* spiroosomes suggest a single type of transformation pilus in competence. *PLoS Pathog* *11*, e1004835.

Lavery, P.E., and Kowalczykowski, S.C. (1992). A postsynaptic role for single-stranded DNA-binding protein in *recA* protein-promoted DNA strand exchange. *J Biol Chem* *267*, 9315–9320.

Lee, M.S., and Morrison, D.A. (1999). Identification of a New Regulator in *Streptococcus pneumoniae* Linking Quorum Sensing to Competence for Genetic Transformation. *J Bacteriol* *181*, 5004–5016.

Lee, J.Y., Qi, Z., and Greene, E.C. (2016). ATP hydrolysis Promotes Duplex DNA Release by the RecA Presynaptic Complex. *J Biol Chem* *291*, 22218–22230.

Lenz, D.H., Miller, M.B., Zhu, J., Kulkarni, R.V., and Bassler, B.L. (2005). CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Mol Microbiol* *58*, 1186–1202.

Leong, C.G., Bloomfield, R.A., Boyd, C.A., Dornbusch, A.J., Lieber, L., Liu, F., Owen, A., Slay, E., Lang, K.M., and Lostroh, C.P. (2017). The role of core and accessory type IV pilus genes in natural transformation and twitching motility in the bacterium *Acinetobacter baylyi*. *PLoS One* *12*.

Lichev, A., Angelov, A., Cucurull, I., and Liebl, W. (2019). Amino acids as nutritional factors and (p)ppGpp as an alarmone of the stringent response regulate natural transformation in *Micrococcus luteus*. *Sci Rep* *9*, 11030.

Lindner, C., Nijland, R., van Hartskamp, M., Bron, S., Hamoen, L.W., and Kuipers, O.P. (2004). Differential Expression of Two Paralogous Genes of *Bacillus subtilis* Encoding Single-Stranded DNA Binding Protein. *J Bacteriol* *186*, 1097–1105.

Lisboa, J., Andreani, J., Sanchez, D., Boudes, M., Collinet, B., Liger, D., Tilbeurgh, H. van, Guérois, R., and Quevillon-Cheruel, S. (2014). Molecular determinants of the DprA–RecA interaction for nucleation on ssDNA. *Nucleic Acids Research* *42*, 7395–7408.

Lohman, T.M., and Ferrari, M.E. (1994). *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu Rev Biochem* *63*, 527–570.

Londoño-Vallejo, J.A., and Dubnau, D. (1993). comF, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol Microbiol* *9*, 119–131.

Londoño-Vallejo, J.A., and Dubnau, D. (1994). Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. *J Bacteriol.* *176*, 4642–4645.

Lurie-Weinberger, M.N., Gomez-Valero, L., Merault, N., Glöckner, G., Buchrieser, C., and Gophna, U. (2010). The origins of eukaryotic-like proteins in *Legionella pneumophila*. *Int J Med Microbiol* *300*, 470–481.

Maier, B. (2020). Competence and Transformation in *Bacillus subtilis*. *Curr Issues Mol Biol* *37*, 57–76.

Maier, B., Potter, L., So, M., Seifert, H.S., and Sheetz, M.P. (2002). Single pilus motor forces exceed 100 pN. *Proc Natl Acad Sci U S A* *99*, 16012–16017.

Mancl, J.M., Black, W.P., Robinson, H., Yang, Z., and Schubot, F.D. (2016). Crystal Structure of a Type IV Pilus Assembly ATPase: Insights into the Molecular Mechanism of PilB from *Thermus thermophilus*. *Structure* *24*, 1886–1897.

Marie, L., Rapisarda, C., Morales, V., Bergé, M., Perry, T., Soulet, A.-L., Gruget, C., Remaut, H., Fronzes, R., and Polard, P. (2017). Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. *Nat Commun* *8*, 15638.

Martin, B., García, P., Castanié, M.P., and Claverys, J.P. (1995). The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. *Mol Microbiol* *15*, 367–379.

Martínez, L.C., Yakhnin, H., Camacho, M.I., Georgellis, D., Babitzke, P., Puente, J.L., and Bustamante, V.H. (2011). Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HliD. *Mol Microbiol* *80*, 1637–1656.

Marutani, M., Taguchi, F., Ogawa, Y., Hossain, M.M., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2008). Gac two-component system in *Pseudomonas syringae* pv. *tabaci* is required for virulence but not for hypersensitive reaction. *Mol Genet Genomics* *279*, 313–322.

Mashiba, K., Hamamoto, T., and Torikai, K. (1993). [A case of Legionnaires' disease due to aspiration of hot spring water and isolation of *Legionella pneumophila* from hot spring water]. *Kansenshogaku Zasshi* *67*, 163–166.

Matias, V.R.F., and Beveridge, T.J. (2005). Cryo-electron microscopy reveals native polymeric cell wall structure in *Bacillus subtilis* 168 and the existence of a periplasmic space. *Mol Microbiol* *56*, 240–251.

Matias, V.R.F., and Beveridge, T.J. (2006). Native Cell Wall Organization Shown by Cryo-Electron Microscopy Confirms the Existence of a Periplasmic Space in *Staphylococcus aureus*. *J Bacteriol* *188*, 1011–1021.

Matthey, N., and Blokesch, M. (2016). The DNA-Uptake Process of Naturally Competent *Vibrio cholerae*. *Trends in Microbiology* *24*, 98–110.

McCallum, M., Tammam, S., Khan, A., Burrows, L.L., and Howell, P.L. (2017). The molecular mechanism of the type IVa pilus motors. *Nat Commun* *8*, 15091.

McDade, J.E., Shepard, C.C., Fraser, D.W., Tsai, T.R., Redus, M.A., and Dowdle, W.R. (1977). Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. *N Engl J Med* *297*, 1197–1203.

McKenzie, G.J., Harris, R.S., Lee, P.L., and Rosenberg, S.M. (2000). The SOS response regulates adaptive mutation. *PNAS* *97*, 6646–6651.

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.-Y., and Schoolnik, G.K. (2005). Chitin induces natural competence in *Vibrio cholerae*. *Science* *310*, 1824–1827.

Méjean, V., and Claverys, J.-P. (1988). Polarity of DNA entry in transformation of *Streptococcus pneumoniae*. *Mol Gen Genet* *213*, 444–448.

Mell, J.C., Hall, I.M., and Redfield, R.J. (2012). Defining the DNA uptake specificity of naturally competent *Haemophilus influenzae* cells. *Nucleic Acids Res* *40*, 8536–8549.

Melville, S., and Craig, L. (2013). Type IV pili in Gram-positive bacteria. *Microbiol Mol Biol Rev* *77*, 323–341.

Menetski, J.P., Bear, D.G., and Kowalczykowski, S.C. (1990). Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *PNAS* *87*, 21–25.

Merz, A.J., So, M., and Sheetz, M.P. (2000). Pilus retraction powers bacterial twitching motility. *Nature* *407*, 98–102.

Miller, M.B., and Bassler, B.L. (2001). Quorum sensing in bacteria. *Annu Rev Microbiol* *55*, 165–199.

Miller, C.L., Romero, M., Karna, S.L.R., Chen, T., Heeb, S., and Leung, K.P. (2016). RsmW, *Pseudomonas aeruginosa* small non-coding RsmA-binding RNA upregulated in biofilm versus planktonic growth conditions. *BMC Microbiol* *16*, 155.

Mintz, C.S., and Shuman, H.A. (1987). Transposition of bacteriophage Mu in the Legionnaires disease bacterium. *Proc Natl Acad Sci U S A* *84*, 4645–4649.

Mirouze, N., Bergé, M.A., Soulet, A.-L., Mortier-Barrière, I., Quentin, Y., Fichant, G., Granadel, C., Noirot-Gros, M.-F., Noirot, P., Polard, P., et al. (2013). Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. *PNAS* *110*, E1035–E1044.

Mitra, A., Palaniyandi, S., Herren, C.D., Zhu, X., and Mukhopadhyay, S. (2013). Pleiotropic roles of *uvrY* on biofilm formation, motility and virulence in uropathogenic *Escherichia coli* CFT073. *PLoS One* *8*, e55492.

Moliner, C., Fournier, P.-E., and Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiology Reviews* *34*, 281–294.

Molmeret, M., Horn, M., Wagner, M., Santic, M., and Abu Kwaik, Y. (2005). Amoebae as Training Grounds for Intracellular Bacterial Pathogens. *Appl Environ Microbiol* *71*, 20–28.

Molofsky, A.B., and Swanson, M.S. (2004). Differentiate to thrive: lessons from the *Legionella pneumophila* life cycle. *Molecular Microbiology* *53*, 29–40.

Mondino, S., Schmidt, S., Rolando, M., Escoll, P., Gomez-Valero, L., and Buchrieser, C. (2020). Legionnaires' Disease: State of the Art Knowledge of Pathogenesis Mechanisms of *Legionella*. *Annu Rev Pathol* *15*, 439–466.

Moreno-Gámez, S., Sorg, R.A., Domenech, A., Kjos, M., Weissing, F.J., van Doorn, G.S., and Veening, J.-W. (2017). Quorum sensing integrates environmental cues, cell density and cell history to control bacterial competence. *Nat Commun* *8*, 854.

Morrison, D.A. (1977). Transformation in pneumococcus: existence and properties of a complex involving donor deoxyribonucleate single strands in eclipse. *J Bacteriol* *132*, 576–583.

Morrison, D.A., Mortier-Barrière, I., Attaiech, L., and Claverys, J.-P. (2007). Identification of the Major Protein Component of the Pneumococcal Eclipse Complex. *J Bacteriol* *189*, 6497–6500.

Mortier-Barrière, I., Velten, M., Dupaigne, P., Mirouze, N., Piétrement, O., McGovern, S., Fichant, G., Martin, B., Noirot, P., Le Cam, E., et al. (2007). A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* *130*, 824–836.

Muschiol, S., Balaban, M., Normark, S., and Henriques-Normark, B. (2015). Uptake of extracellular DNA: Competence induced pili in natural transformation of *Streptococcus pneumoniae*. *Bioessays* *37*, 426–435.

Nandakumar, D., and Patel, S.S. (2015). Finding the Right Match Fast. *Cell* *160*, 809–811.

Narra, H.P., and Ochman, H. (2006). Of What Use Is Sex to Bacteria? *Current Biology* 16, R705–R710.

Nedenskov-Sørensen, P., Bukholm, G., and Bøvre, K. (1990). Natural competence for genetic transformation in *Campylobacter pylori*. *J Infect Dis* 161, 365–366.

Nero, T.M., Dalia, T.N., Wang, J.C.-Y., Kysela, D.T., Bochman, M.L., and Dalia, A.B. (2018). ComM is a hexameric helicase that promotes branch migration during natural transformation in diverse Gram-negative species. *Nucleic Acids Res.* 46, 6099–6111.

Newton, H.J., Ang, D.K.Y., van Driel, I.R., and Hartland, E.L. (2010). Molecular pathogenesis of infections caused by *Legionella pneumophila*. *Clin Microbiol Rev* 23, 274–298.

Ng, W.-L., and Bassler, B.L. (2009). Bacterial Quorum-Sensing Network Architectures. *Annual Review of Genetics* 43, 197–222.

Nunn, D.N., and Lory, S. (1991). Product of the *Pseudomonas aeruginosa* gene *pilD* is a prepilin leader peptidase. *Proc Natl Acad Sci U S A* 88, 3281–3285.

Oliveira Paiva, A.M., Friggen, A.H., Qin, L., Douwes, R., Dame, R.T., and Smits, W.K. (2019). The Bacterial Chromatin Protein HupA Can Remodel DNA and Associates with the Nucleoid in *Clostridium difficile*. *Journal of Molecular Biology* 431, 653–672.

Pelicic, V. (2008). Type IV pili: e pluribus unum? *Mol Microbiol* 68, 827–837.

Peng, C., Lin, Y., Luo, H., and Gao, F. (2017). A Comprehensive Overview of Online Resources to Identify and Predict Bacterial Essential Genes. *Frontiers in Microbiology* 8, 2331.

Pernestig, A.K., Melefors, O., and Georgellis, D. (2001). Identification of UvrY as the cognate response regulator for the BarA sensor kinase in *Escherichia coli*. *J. Biol. Chem.* 276, 225–231.

Peterson, S., Cline, R.T., Tettelin, H., Sharov, V., and Morrison, D.A. (2000). Gene Expression Analysis of the *Streptococcus pneumoniae* Competence Regulons by Use of DNA Microarrays. *Journal of Bacteriology* 182, 6192–6202.

Peterson, S.N., Sung, C.K., Cline, R., Desai, B.V., Snesrud, E.C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., et al. (2004). Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol Microbiol* 51, 1051–1070.

Piechowska, M., and Fox, M.S. (1971). Fate of Transforming Deoxyribonucleate in *Bacillus subtilis*. *J Bacteriol* 108, 680–689.

Piepenbrink, K.H. (2019). DNA Uptake by Type IV Filaments. *Front Mol Biosci* 6.

Pimentel, Z.T., and Zhang, Y. (2018). Evolution of the Natural Transformation Protein, ComEC, in Bacteria. *Front Microbiol* 9, 2980.

Porter, R.D. (1986). Transformation in Cyanobacteria. *CRC Critical Reviews in Microbiology* 13, 111–132.

Provvedi, R., and Dubnau, D. (1999). ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Molecular Microbiology* 31, 271–280.

Provvedi, R., Chen, I., and Dubnau, D. (2001). NucA is required for DNA cleavage during transformation of *Bacillus subtilis*. *Molecular Microbiology* 40, 634–644.

Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B., and Claverys, J.-P. (2006). Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. *Science* 313, 89–92.

Pujol, C., Eugène, E., Marceau, M., and Nassif, X. (1999). The meningococcal PilT protein is required for induction of intimate attachment to epithelial cells following pilus-mediated adhesion. *Proc Natl Acad Sci U S A* 96, 4017–4022.

Puyet, A., Greenberg, B., and Lacks, S.A. (1990). Genetic and structural characterization of endA. A membrane-bound nuclease required for transformation of *Streptococcus pneumoniae*. *J Mol Biol* 213, 727–738.

Quevillon-Cheruel, S., Campo, N., Mirouze, N., Mortier-Barrière, I., Brooks, M.A., Boudes, M., Durand, D., Soulet, A.-L., Lisboa, J., Noirot, P., et al. (2012). Structure–function analysis of pneumococcal DprA protein reveals that dimerization is crucial for loading RecA recombinase onto DNA during transformation. *Proc Natl Acad Sci U S A* 109, E2466–E2475.

Raghunathan, S., Kozlov, A.G., Lohman, T.M., and Waksman, G. (2000). Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nat Struct Biol* 7, 648–652.

Redfield, R.J. (1993). Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* 84, 400–404.

Redfield, R.J., Cameron, A.D.S., Qian, Q., Hinds, J., Ali, T.R., Kroll, J.S., and Langford, P.R. (2005). A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. *J Mol Biol* 347, 735–747.

Reimann, C., Beyeler, M., Latifi, A., Winteler, H., Foglino, M., Lazdunski, A., and Haas, D. (1997). The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Mol Microbiol* 24, 309–319.

Reitz, D., Chan, Y.-L., and Bishop, D.K. (2021). How strand exchange protein function benefits from ATP hydrolysis. *Curr Opin Genet Dev* 71, 120–128.

Ren, J., Karna, S., Lee, H.-M., Yoo, S.M., and Na, D. (2019). Artificial transformation methodologies for improving the efficiency of plasmid DNA transformation and simplifying its use. *Appl Microbiol Biotechnol* 103, 9205–9215.

Renkawitz, J., Lademann, C.A., and Jentsch, S. (2014). Mechanisms and principles of homology search during recombination. *Nat Rev Mol Cell Biol* 15, 369–383.

Rj, R. (1991). *sxy-1*, a *Haemophilus influenzae* mutation causing greatly enhanced spontaneous competence. *Journal of Bacteriology* 173.

Rogers, J., Dowsett, A.B., Dennis, P.J., Lee, J.V., and Keevil, C.W. (1994). Influence of temperature and plumbing material selection on biofilm formation and growth of *Legionella pneumophila* in a model potable water system containing complex microbial flora. *Appl Environ Microbiol* 60, 1585–1592.

Rosselli, W., and Stasiak, A. (1990). Energetics of RecA-mediated recombination reactions. Without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle. *J Mol Biol* 216, 335–352.

Rowbotham, T.J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *J Clin Pathol* 33, 1179–1183.

Salleh, M.Z., Karuppiah, V., Snee, M., Thistlethwaite, A., Levy, C.W., Knight, D., and Derrick, J.P. (2019). Structure and Properties of a Natural Competence-Associated Pilin Suggest a Unique Pilus Tip-Associated DNA Receptor. *MBio* 10, e00614-19.

Schmitt, W., Odenbreit, S., Heuermann, D., and Haas, R. (1995). Cloning of the *Helicobacter pylori* recA gene and functional characterization of its product. *Mol Gen Genet* 248, 563–572.

Scrudato, M.L., and Blokesch, M. (2012). The Regulatory Network of Natural Competence and Transformation of *Vibrio cholerae*. *PLOS Genetics* 8, e1002778.

Segal, G., Purcell, M., and Shuman, H.A. (1998). Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc Natl Acad Sci U S A* 95, 1669–1674.

Seitz, P., and Blokesch, M. (2013a). Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiol. Rev.* 37, 336–363.

Seitz, P., and Blokesch, M. (2013b). DNA-uptake machinery of naturally competent *Vibrio cholerae*. *PNAS* 110, 17987–17992.

Seitz, P., Pezeshgi Modarres, H., Borgeaud, S., Bulushev, R.D., Steinbock, L.J., Radenovic, A., Dal Peraro, M., and Blokesch, M. (2014). ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. *PLoS Genet.* 10, e1004066.

Selander, R.K., McKinney, R.M., Whittam, T.S., Bibb, W.F., Brenner, D.J., Nolte, F.S., and Pattison, P.E. (1985). Genetic structure of populations of *Legionella pneumophila*. *J Bacteriol* 163, 1021–1037.

Serrano, E., Carrasco, B., Gilmore, J.L., Takeyasu, K., and Alonso, J.C. (2018). RecA Regulation by RecU and DprA During *Bacillus subtilis* Natural Plasmid Transformation. *Frontiers in Microbiology* 9, 1514.

Sexton, J.A., and Vogel, J.P. (2004). Regulation of Hypercompetence in *Legionella pneumophila*. *J Bacteriol* 186, 3814–3825.

Sheppard, D., Berry, J.-L., Denise, R., Rocha, E.P.C., Matthews, S., and Pelicic, V. (2020). The major subunit of widespread competence pili exhibits a novel and conserved type IV pilin fold. *Journal of Biological Chemistry* 295, 6594–6604.

Shestakov, S.V., and Khyen, N.T. (1970). Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. *Mol Gen Genet* 107, 372–375.

Shi, C., Forsbach-Birk, V., Marre, R., and McNealy, T.L. (2006). The *Legionella pneumophila* global regulatory protein LetA affects DotA and Mip. *Int. J. Med. Microbiol.* 296, 15–24.

Silale, A., Lea, S.M., and Berks, B.C. (2021). The DNA transporter ComEC has metal-dependent nuclease activity that is important for natural transformation. *Molecular Microbiology* *116*, 416–426.

Silva, A.J., and Benitez, J.A. (2004). Transcriptional Regulation of *Vibrio cholerae* Hemagglutinin/Protease by the Cyclic AMP Receptor Protein and RpoS. *Journal of Bacteriology* *186*, 6374–6382.

Sinha, S., Mell, J.C., and Redfield, R.J. (2012). Seventeen Sxy-Dependent Cyclic AMP Receptor Protein Site-Regulated Genes Are Needed for Natural Transformation in *Haemophilus influenzae*. *J Bacteriol* *194*, 5245–5254.

Sinha, S., Mell, J., and Redfield, R. (2013). The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. *Mol Microbiol* *88*, 1106–1119.

Slager, J., Aprianto, R., and Veening, J.-W. (2019). Refining the Pneumococcal Competence Regulon by RNA Sequencing. *Journal of Bacteriology* *201*, 25.

Smith, H.O., Gwinn, M.L., and Salzberg, S.L. (1999). DNA uptake signal sequences in naturally transformable bacteria. *Research in Microbiology* *150*, 603–616.

Smits, W.K., Eschevins, C.C., Susanna, K.A., Bron, S., Kuipers, O.P., and Hamoen, L.W. (2005). Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. *Mol Microbiol* *56*, 604–614.

Solanki, V., Kapoor, S., and Thakur, K.G. (2018). Structural insights into the mechanism of Type IVa pilus extension and retraction ATPase motors. *The FEBS Journal* *285*, 3402–3421.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: building the web of life. *Nat Rev Genet* *16*, 472–482.

Sparling, P.F. (1966). Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. *J Bacteriol* *92*, 1364–1371.

Speers, A.M., Schindler, B.D., Hwang, J., Genc, A., and Reguera, G. (2016). Genetic Identification of a PilT Motor in *Geobacter sulfurreducens* Reveals a Role for Pilus Retraction in Extracellular Electron Transfer. *Front Microbiol* *7*, 1578.

Steinmoen, H., Knutsen, E., and Håvarstein, L.S. (2002). Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proc Natl Acad Sci U S A* *99*, 7681–7686.

Stingl, K., Müller, S., Scheidgen-Kleyboldt, G., Clausen, M., and Maier, B. (2010). Composite system mediates two-step DNA uptake into *Helicobacter pylori*. *PNAS* *107*, 1184–1189.

Stone, B.J., and Abu Kwaik, Y. (1998). Expression of multiple pili by *Legionella pneumophila*: identification and characterization of a type IV pilin gene and its role in adherence to mammalian and protozoan cells. *Infect Immun* *66*, 1768–1775.

Stone, B.J., and Kwaik, Y.A. (1999). Natural Competence for DNA Transformation by *Legionella pneumophila* and Its Association with Expression of Type IV Pili. *J Bacteriol* *181*, 1395–1402.

Story, R.M., Weber, I.T., and Steitz, T.A. (1992). The structure of the *E. coli* recA protein monomer and polymer. *Nature* *355*, 318–325.

Strom, M.S., and Lory, S. (1991). Amino acid substitutions in pilin of *Pseudomonas aeruginosa*. Effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. *J Biol Chem* 266, 1656–1664.

Suerbaum, S., and Josenhans, C. (2007). *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol* 5, 441–452.

Sukmana, A., and Yang, Z. (2018). The type IV pilus assembly motor PilB is a robust hexameric ATPase with complex kinetics. *Biochem J* 475, 1979–1993.

Sysoeva, T.A., Bane, L.B., Xiao, D.Y., Bose, B., Chilton, S.S., Gaudet, R., and Burton, B.M. (2015). Structural characterization of the late competence protein ComFB from *Bacillus subtilis*. *Biosci Rep* 35, e00183.

Tadesse, S., and Graumann, P.L. (2007). DprA/Smf protein localizes at the DNA uptake machinery in competent *Bacillus subtilis* cells. *BMC Microbiol* 7, 105.

Takata, T., Ando, T., Israel, D.A., Wassenaar, T.M., and Blaser, M.J. (2005). Role of dprA in transformation of *Campylobacter jejuni*. *FEMS Microbiology Letters* 252, 161–168.

Tan, I.S., and Ramamurthi, K.S. (2014). Spore formation in *Bacillus subtilis*. *Environ Microbiol Rep* 6, 212–225.

Taton, A., Erikson, C., Yang, Y., Rubin, B.E., Rifkin, S.A., Golden, J.W., and Golden, S.S. (2020). The circadian clock and darkness control natural competence in cyanobacteria. *Nat Commun* 11, 1688.

Tatum, E.L., and Lederberg, J. (1947). Gene Recombination in the Bacterium *Escherichia coli*. *J Bacteriol* 53, 673–684.

Taylor, M., Ross, K., and Bentham, R. (2009). *Legionella*, protozoa, and biofilms: interactions within complex microbial systems. *Microb Ecol* 58, 538–547.

Tomasz, A., and Hotchkiss, R.D. (1964). REGULATION OF THE TRANSFORMABILITY OF PNEUMOCOCCAL CULTURES BY MACROMOLECULAR CELL PRODUCTS*. *Proc Natl Acad Sci U S A* 51, 480–487.

Tomb, J.F., Barcak, G.J., Chandler, M.S., Redfield, R.J., and Smith, H.O. (1989). Transposon mutagenesis, characterization, and cloning of transformation genes of *Haemophilus influenzae* Rd. *J Bacteriol* 171, 3796–3802.

Tomich, M., Planet, P.J., and Figurski, D.H. (2007). The tad locus: postcards from the widespread colonization island. *Nat Rev Microbiol* 5, 363–375.

Torres, R., Serrano, E., and Alonso, J.C. (2019). *Bacillus subtilis* RecA interacts with and loads RadA/Sms to unwind recombination intermediates during natural chromosomal transformation. *Nucleic Acids Research* 47, 9198–9215.

Treangen, T.J., Ambur, O.H., Tonjum, T., and Rocha, E.P. (2008). The impact of the neisserial DNA uptake sequences on genome evolution and stability. *Genome Biology* 9, R60.

Tsirigos, K.D., Peters, C., Shu, N., Käll, L., and Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 43, W401–W407.

Vesel, N., and Blokesch, M. (2021). Pilus Production in *Acinetobacter baumannii* Is Growth Phase Dependent and Essential for Natural Transformation. *J Bacteriol* 203, e00034-21.

Vogel, J.P., Andrews, H.L., Wong, S.K., and Isberg, R.R. (1998). Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* 279, 873–876.

Walker, J.T. (2018). The influence of climate change on waterborne disease and *Legionella*: a review. *Perspect Public Health* 138, 282–286.

Wang, W., Ding, J., Zhang, Y., Hu, Y., and Wang, D.-C. (2014). Structural insights into the unique single-stranded DNA-binding mode of *Helicobacter pylori* DprA. *Nucleic Acids Res* 42, 3478–3491.

Waters, C.M., and Bassler, B.L. (2005). QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology* 21, 319–346.

von Wintersdorff, C.J.H., Penders, J., van Niekerk, J.M., Mills, N.D., Majumder, S., van Alphen, L.B., Savelkoul, P.H.M., and Wolffs, P.F.G. (2016). Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front Microbiol* 7, 173.

Wolfgang, M., Lauer, P., Park, H.S., Brossay, L., Hébert, J., and Koomey, M. (1998). PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in pilated *Neisseria gonorrhoeae*. *Mol Microbiol* 29, 321–330.

Wolfgang, M., van Putten, J.P., Hayes, S.F., and Koomey, M. (1999). The *comP* locus of *Neisseria gonorrhoeae* encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation. *Mol Microbiol* 31, 1345–1357.

Wollman, E.L., Jacob, F., and Hayes, W. (1956). Conjugation and genetic recombination in *Escherichia coli* K-12. *Cold Spring Harb Symp Quant Biol* 21, 141–162.

Wright, W.D., Shah, S.S., and Heyer, W.-D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* 293, 10524–10535.

Wu, H.-Y., Lu, C.-H., and Li, H.-W. (2017). RecA-SSB Interaction Modulates RecA Nucleoprotein Filament Formation on SSB-Wrapped DNA. *Sci Rep* 7, 11876.

Yadav, T., Carrasco, B., Myers, A.R., George, N.P., Keck, J.L., and Alonso, J.C. (2012). Genetic recombination in *Bacillus subtilis*: a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Res* 40, 5546–5559.

Yadav, T., Carrasco, B., Hejna, J., Suzuki, Y., Takeyasu, K., and Alonso, J.C. (2013). *Bacillus subtilis* DprA Recruits RecA onto Single-stranded DNA and Mediates Annealing of Complementary Strands Coated by SsbB and SsbA. *J Biol Chem* 288, 22437–22450.

Yamamoto, S., Mitobe, J., Ishikawa, T., Wai, S.N., Ohnishi, M., Watanabe, H., and Izumiya, H. (2014). Regulation of natural competence by the orphan two-component system sensor kinase ChiS involves a non-canonical transmembrane regulator in *Vibrio cholerae*. *Molecular Microbiology* 91, 326–347.

Yu, V.L., Plouffe, J.F., Pastoris, M.C., Stout, J.E., Schousboe, M., Widmer, A., Summersgill, J., File, T., Heath, C.M., Paterson, D.L., et al. (2002). Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *J Infect Dis* 186, 127–128.

Zere, T.R., Vakulskas, C.A., Leng, Y., Pannuri, A., Potts, A.H., Dias, R., Tang, D., Kolaczowski, B., Georgellis, D., Ahmer, B.M.M., et al. (2015). Genomic Targets and Features of BarA-UvrY (-SirA) Signal Transduction Systems. *PLoS One* 10.

Zheng, D., Constantinidou, C., Hobman, J.L., and Minchin, S.D. (2004). Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res* 32, 5874–5893.

Zinder, N.D., and Lederberg, J. (1952). Genetic exchange in Salmonella. *J Bacteriol* 64, 679–699.

Zöllner, R., Cronenberg, T., and Maier, B. (2019). Motor Properties of PilT-Independent Type 4 Pilus Retraction in Gonococci. *J Bacteriol* 201, e00778-18.

(2019). *Legionella: Methods and Protocols* (Humana Press).

Bilan des cas de légionellose survenus en France en 2019.