



HAL
open science

Evolution de la virulence et du spectre d'hôtes chez la bactérie pathogène émergente *Legionella pneumophila*

Guillaume Carrillo

► **To cite this version:**

Guillaume Carrillo. Evolution de la virulence et du spectre d'hôtes chez la bactérie pathogène émergente *Legionella pneumophila*. Microbiologie et Parasitologie. Université Lyon 1 - Claude Bernard, 2023. Français. NNT: . tel-04766146

HAL Id: tel-04766146

<https://theses.hal.science/tel-04766146v1>

Submitted on 4 Nov 2024

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 CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 341
Evolution, Ecosystèmes, Microbiologie, Modélisation**

Discipline : Microbiologie

Soutenue publiquement le 29 Septembre 2023, par :
Guillaume Carrillo

**Evolution de la virulence et du spectre
d'hôtes chez la bactérie pathogène
émergente *Legionella pneumophila***

Devant le jury composé de :

ALLEGRA, Séverine, Professeure des Universités, Université Jean Monnet St-Etienne. Rapporteuse
GUY, Lionel, Professeur associé, Université d'Uppsala. Rapporteur
SAMBA-LOUAKA Ascel, Maître de Conférences, Université de Poitiers. Rapporteur
MOREAU Karen, Professeure des Universités, Université Claude Bernard Lyon 1. Présidente
DAUBIN Vincent, Directeur de Recherche, Université Claude Bernard Lyon 1. Directeur de thèse
KAY Elisabeth, Chargée de Recherche, Université Claude Bernard Lyon 1. Co-directrice de thèse

Université Claude Bernard – LYON 1

Président de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN ADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Comité de Coordination des Etudes Médicales	Mme Carole BURILLON
Faculté d'Odontologie	M. Jean Christophe MAURIN
Faculté de Médecine et Maïeutique Lyon Sud – Charles Mérieux	M. Philippe PAPAREL
Faculté de Médecine Lyon-Est	M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	M. Jacques LUAUTE
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	M. Claude DUSSART

COMPOSANTES SECTEUR SCIENCES & TECHNOLOGIE

Département génie Electrique et des Procédés (GEP)	Mme. Rosaria FERRIGNO
Département Informatique	Mme. Saida BOUAKAZ BRONDEL
Département Mécanique	M. Marc BUFFAT
Institut de Science Financière et d'Assurances (ISFA)	M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education (INSPé)	M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1 (IUT)	M. Michel MASSENZIO
Observatoire de Lyon	M. Bruno GUIDERDONI
Polytech Lyon	M. Emmanuel PERRIN
UFR Biosciences	Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	M. Guillaume BODET
URF Faculté des Sciences	M. Bruno ANDRIOLETTI

Acknowledgements

While this manuscript is written in English, I'll use my mother tongue to write the acknowledgements part, as the jury of this PhD is French-speaking and I hope I'll be able to better convey in French the gratitude I want to share with the people acknowledged here.

Je voudrais avant tout remercier Mme Séverine Allegra, M. Lionel Guy et M. Ascel Samba-Louaka pour avoir accepté de juger mon travail de thèse, et ce malgré les différentes embuches rencontrées en fin de parcours. Vous avez toute ma reconnaissance pour votre disponibilité et votre abnégation en acceptant de relire ce travail en si peu de temps (et sur la période estivale). Pour ça, merci. J'espère que mon travail sera à la hauteur.

Merci également à Mme Karen Moreau d'avoir accepté de faire partie de ce jury. Karen, merci de m'avoir suivi tout au long de ces années. C'est avec toi que j'ai commencé mon parcours dans la recherche, il apparaît donc naturel que tu puisses participer à l'évaluation de ce travail. Merci de m'avoir fait confiance pour ce stage de master et merci de m'avoir fait prendre goût à l'étude des bactos. Je n'oublierai également jamais ta disponibilité et ta compréhension lorsque, en plein milieu de mon stage, j'ai dû m'absenter pendant quelques temps afin de régler certains problèmes de santé.

Merci à Vincent, mon directeur de thèse. Merci de m'avoir fait confiance avec ce projet et merci de m'avoir initié à l'étude de la bioinfo. Même si le projet initial n'a pas abouti comme espéré, j'ai beaucoup appris à ton contact. Ton professionnalisme, ta connaissance scientifique, ta capacité exceptionnelle à ne jamais voir de problèmes et ta coolitude ont été autant de petits plus que je garde en mémoire et espère avoir fait miens et implémentés dans ma vie professionnelle et personnelle.

Et surtout merci à Elisabeth. Babeth (j'ai pas osé en 7 ans, je me permets cette fantaisie maintenant), au-delà d'avoir été une fantastique co-directrice de thèse tu as surtout été une rencontre formidable. Nous voilà arrivés au bout de ce voyage qui a duré 7 ans et pas une fois tu ne m'as fait sentir autre que quelqu'un de valable et capable. Tu as toujours su m'épauler au niveau scientifique comme personnel, dans les mauvais moments (et Dieu sait qu'il y en a eu !) comme dans les bons (heureusement là aussi on peut dresser une longue liste). J'ai énormément appris à tes côtés. Ta disponibilité et ton soutien sans failles sont certainement la raison pour laquelle je suis en mesure d'écrire ce paragraphe. Je sais que tu penseras que c'est faux car tu es trop humble pour le reconnaître, mais c'est grâce à toi qu'on est arrivés au bout. Si tu m'as bien appris quelque chose, c'est qu'il ne faut jamais douter de soi et avec ce travail, on a encore une autre preuve que tes hypothèses sont visionnaires, bravo ! Pour terminer, je me suis amusé à me poser la question de quelle définition du dictionnaire t'irait le mieux, j'avoue, j'hésites entre « bienveillance » et « résilience ». Dans tous les cas, merci.

Je voudrais maintenant remercier toutes les personnes avec qui j'ai collaboré tout au long de ma thèse. J'écris ces pages bien longtemps après avoir posé ma dernière pipette, aussi j'espère ne pas en oublier.

Merci à Patricia de m'avoir accueilli dans son équipe et d'avoir toujours gardé un esprit ouvert sur la suite à donner à mon projet après la fin de mon contrat doctoral. Merci également à Sophie pour la confiance que tu m'as apporté et pour avoir fait en sorte d'être mis en lien avec Vincent et Elisabeth au début de la thèse.

Merci à Christophe et Anne, les mémoires du labo. Quel plaisir d'avoir pu travailler à vos côtés et d'avoir pu partager vos expériences. Mention particulière à Anne avec qui j'ai partagé, comme beaucoup, de longues discussions sur à peu près tous les sujets possibles et des cours de sport endiablés. Tu m'as dit une fois que tu étais impressionnée par le fait de combiner un travail à plein temps et l'écriture d'une thèse. C'est fait ! J'espère que tu ne seras pas déçue en lisant les lignes qui vont suivre.

Merci à Nat et Claire sans qui le labo se serait écroulé depuis longtemps. Encore une fois, quel plaisir d'avoir pu travailler entouré de personnes si compétentes.

Merci à Annelise et Johann, je n'ai que de bons souvenirs qui vous sont reliés. Merci de votre présence presque certaines lors des weekends au labo et merci pour vos discussions, scientifiques ou non.

Merci à l'équipe de la Croix Rousse et surtout à Christophe, sans ses nombreuses contributions, mon travail n'aurait pas la même saveur !

Merci également à mes stagiaires, Pierre-Yves et Juliette. C'est étrange de te présenter comme stagiaire alors que toi aussi tu écris ta thèse Juliette. Merci de ton aide précieuse et j'espère t'avoir laissé un bon sujet sur lequel travailler. Bon courage pour l'écriture !

Il est temps maintenant de changer d'équipe et de remercier également l'équipe HORIZONE (qui pour moi resteront « les compétents »). Merci à tous de m'avoir enrichi à un moment ou à un autre de savoirs scientifiques ou de papotages plus personnels.

Merci surtout aux doctorants avec qui j'ai partagé cette galère, Kévin, Léo, Félix, Margaux. C'est drôle de penser que vous avez tous commencés votre thèse après moi et que je dois maintenant vous remercier en tant que docteurs. Merci Corentin pour tout, vraiment.

Remercier les doctorants ne serait rien sans la meilleure d'entre nous, merci Anne-So. Je ne vais pas m'étaler sur tout ce que te rencontrer m'a apporté, ce serait trop long et tu sais bien déjà tout.

Merci à ceux que j'ai oublié, ne m'en voulez pas et si jamais, contactez-moi, ça fera l'occasion de prendre une bière !

Je voudrais aussi remercier ma famille qui m'a suivi et soutenu pendant tout ce long parcours. Ça y est Steph, je vais bientôt te rejoindre dans la team docteur. Merci à mon frère, merci à ma sœur, merci à mes belles-sœurs. Merci à mes neveux et nièces aussi, ils n'ont pas vraiment activement participé à ce travail (et pour deux d'entre eux ne pourront pas lire ce message avant quelques années), mais c'est toujours important de se sentir entourés. Merci à ma maman.

Enfin, le plus gros des mercis à Cédrick. Je n'en serais pas là sans toi et « là » veut dire plus que la finalisation de ce travail de thèse. Merci de m'avoir boosté pour terminer tout ça, merci pour ta patience, ta compréhension et ton soutien pendant les derniers mois. Merci de t'être dévoué pour patauger au bord de la piscine en Grèce pendant que j'écrivais ce manuscrit. Merci d'être là, merci merci merci d'être TOI. Je t'aime.

Table of Contents

Acknowledgements	5
Summary	13
<i>In English</i>	13
<i>In French</i>	15
General objectives	17
Disclaimer	21
1. Literature review	23
1.1. <i>Experimental evolution methods and their contribution to the study of bacterial processes</i>	23
1.1.1. Introduction	23
1.1.2. The Long Term Evolution Experiment (LTEE) by Richard Lenski	30
1.1.3. Experimental evolution and pathogenesis	33
1.1.4. Experimental evolution and symbiosis	35
1.1.5. Experimental evolution as a way to study emergence of resistance mechanisms in <i>L. pneumophila</i> .	39
1.2. <i>Legionella pneumophila, a fascinating emerging bacterial pathogen</i>	42
1.2.1. <i>Legionella pneumophila</i>	42
1.2.2. Legionellosis	43
1.2.3. Ecology and genetics	47
1.2.4. The biphasic lifestyle of <i>Legionella</i>	59
1.2.5. The Type IVb Dot/Icm secretion system of <i>Legionella</i>	61
1.2.6. Effectors of the Type IVb secretion system	65
2. Experimental evolution leads to the discovery of a novel mechanism of virulence control in <i>Legionella pneumophila</i>	75
2.1. <i>Introduction</i>	75
2.2. <i>Results</i>	78
2.2.1. Mutation accumulation experiment in <i>L. pneumophila</i>	78
2.2.2. Growth rate and competitive fitness of <i>L. pneumophila</i> lineages evolved under axenic conditions.	79
2.2.3. Altered intracellular phenotypes of <i>Legionella pneumophila</i> evolved lineages.	81
2.2.4. The genetic basis of altered virulence phenotypes of evolved populations.	83
2.2.5. Spontaneous chromosomal integration of the pLPP plasmid occurs in all lineages	85
2.2.6. pLPP integration into the chromosome is correlated with intracellular phenotypic change.	88
2.2.7. Major transcriptomic shifts in pLPP-integrated populations	89
2.2.8. Reacquisition of virulence traits depends on the environmental conditions in which bacteria replicate.	92
2.2.9. Reversible integration of pLPP in re-evolved populations.	95
2.2.10. Accumulation of multiple mutations in re-evolved lines results in loss-of-function of RelA, LetS/LetA regulators	97
2.3. <i>Discussion</i>	101
2.4. <i>Conclusion</i>	106
3. <i>Legionella norrlandica</i>, an environmental species of <i>Legionella</i> to study transfer of virulence genes	107

3.1.	<i>Introduction on HGT and natural transformation in Legionella pneumophila</i>	107
3.2.	<i>Results</i>	110
3.2.1.	Monitoring intracellular replication ability of <i>L. norrlandica</i>	110
3.2.2.	Study of natural transformation in <i>L. norrlandica</i>	112
3.3.	<i>Discussion</i>	117
4.	Global Discussion and future prospects	121
5.	Materials and Methods	127
5.1.	<i>Bacterial strains, macrophages and amoebae strains and growth conditions</i>	127
5.2.	<i>Experimental evolution strategies</i>	128
5.2.1.	Mutation accumulation strategy:	128
5.2.2.	Adaptative mutation strategy:	128
5.3.	<i>Bacterial competition assays</i>	129
5.4.	<i>Intracellular replication assays</i>	129
5.5.	<i>Endoplasmic reticulum recruitment measurement</i>	129
5.6.	<i>TEM translocation assay</i>	130
5.7.	<i>Genomic DNA preparation, whole-genome sequencing and comparative genomic analysis.</i>	130
5.8.	<i>Detection of the episomal or integrated form of the pLPP plasmid by PCR and Digital droplet PCR</i>	131
5.9.	<i>RNA isolation, depletion of rRNA and RNA sequencing</i>	132
6.	References	133

List of figures

Figure 1: Charles Darwin's 'I think' tree.	24
Figure 2: Reconciling a gene tree with a species tree.	25
Figure 3: Schematic representation of experimental evolution strategies and associated population size progression.	28
Figure 4: Microscopic images of <i>Legionella pneumophila</i> .	43
Figure 5: Evolution of the number of cases and the incidence rate of Legionellosis in France	44
Figure 6: number of cases and the associated case-fatality of legionellosis in France between 1996 and 2018.	46
Figure 7: Phylogenetic tree of the experimentally defined hosts of <i>L. pneumophila</i>	50
Figure 8: <i>L. pneumophila</i> hosts span across the Eukaryota superkingdom	51
Figure 9: Hypothetical reconstruction of major features of amoebozoan evolution.	52
Figure 10: Phylogenetic reconstructions of the genus <i>Legionella</i> with <i>Legionella</i> -like amoebal pathogens (LLAPs) based on 16S RNA sequences.	55
Figure 11: Bacterial biofilm development.	56
Figure 12: Control of the biphasic lifestyle of <i>Legionella pneumophila</i>	60
Figure 13: Type IVb Dot/Icm secretion system of <i>Legionella pneumophila</i> .	64
Figure 14: Type IVa secretion system.	65
Figure 15: Distribution of 50 selected effectors from <i>Legionella pneumophila</i> in 58 different <i>Legionella</i> species and 80 <i>Legionella</i> strains.	66
Figure 16: Composition of the carboxy terminal end of effectors.	68
Figure 17: Summary of genes coding for eukaryotic-like proteins in the genomes of 3 <i>L. pneumophila</i> strains.	70
Figure 18: Intracellular infection cycle of <i>L. pneumophila</i>	73
Figure 19: Summary of <i>L. pneumophila</i> effectors implicated in subversion of host-cell functions	74
Figure 20: Mutation Accumulation experiment (MA) with <i>L. pneumophila</i>	79
Figure 21: Growth rate and competition of <i>L. pneumophila</i> Ancestor, Paris 4/5/6 strains	80
Figure 22: Phenotypes of the MA – generated Paris 4/5/6 lineages	82
Figure 23: Integration of the 131-kb pLPP plasmid into the chromosome of MA-evolved clones	87
Figure 24: Transcriptomic landscape of MA-evolved clones	91
Figure 25: Adaptive evolution experiment (AE) strategies	93
Figure 26: Growth kinetics of Evolvir and AX and MX populations within amoebae	94
Figure 27: Percentage of pLPP integration into the chromosome of the amoeba or liquid medium-evolved populations at several stage of evolution	96
Figure 28: Identification of the mutations in the LetA/LetS two component response system and in the (p)pGpp synthase RelA	100
Figure 29: Model of regulation of natural transformability in <i>L. pneumophila</i>	108
Figure 30: Growth kinetics of <i>L. norrlandica</i> in amoebae and human lung epithelial cells.	111
Figure 31: Tests of natural transformation in <i>L. norrlandica</i> .	113
Figure 32: Construction of the Kan ^R cassette	114
Figure 33: Co-occurrence tree of competence genes within the <i>Legionella</i> species	116
Figure 34: Analysis of the expression of sRNA RocR and mRNA of comEA by Northern blot.	117

List of Tables

Table 1: Table of hypothesis addressed with experimental evolution and model organisms.	27
Table 2: Examples of bacterial-eukaryotic-host interactions.....	36
Table 3: Experimentally confirmed <i>Legionella pneumophila</i> protozoan hosts.....	48
Table 4: List of <i>L. pneumophila</i> proteins preferentially found within eukaryotic proteins and their distribution in three sequenced species	71
Table 5: List of mutations observed in 3 lineages after 700 generations	84
Table 6: Occurrence of the IS21-like family Insertion sequence on the chromosomes and plasmids of <i>Legionella</i> strains.....	86
Table 7: Mutations observed in the AX and BX lineages.....	98

Summary

In English

Bacteria from the genus *Legionella* are responsible in Human for a disease called Legionnaire's disease (LD) or legionellosis. This acute form of pneumonia can be fatal in 7 to 30% of cases, depending on the immunological status of the patient. Among more than sixty species described to this day, the serogroup 1 of *Legionella pneumophila* is alone implicated in 90% of LD cases worldwide. This clinical overrepresentation cannot be linked with an increased prevalence in the environment. What is more, recent comparative genomics studies were not successful in identifying genes or functions specifically present in the 5 more frequently disease-associated subtypes. In the environment, *Legionella pneumophila* colonizes aquatic ecosystems where it is present either in mixt biofilms or associated with protozoan hosts (mostly amoeba) that the bacteria will penetrate and utilize at its advantage to extensively replicate and ultimately kill. In humans, the bacteria infect alveolar macrophages in a fashion similar to that of amoeba infection. The key element of the virulence is the Dot/Icm type IV secretion system (T4SS) that is used to manipulate the host, by translocating more than 300 bacterial proteins called effectors into the infected cell. Several of these effectors are thought to have been acquired by the bacteria through horizontal gene transfer during its lasting co-evolution with its protozoan hosts. The vast effector repertoire as well as their functional redundancy and the high variability within species are all criteria making extremely difficult to find a set of genes and functionality explaining the over prevalence of *L. pneumophila* in clinical cases. For this thesis work, we used two different setting of experimental evolution to allow a clinical strain of *L. pneumophila* to greatly lessen its infectious capacity towards amoeba and to reacquire it in a second time. Genomic sequencing was performed in parallel on specimens at different stages of evolution/re-evolution that allowed us to identify the role of a naturally occurring plasmid (pLPP) in the modification of the virulence in *L. pneumophila*. Indeed, this plasmid has the ability to insert into the genome and excise from it using an insertion sequence present both on the plasmid and in the genome. These insertion/excision events are directly linked to the diminution and reacquisition of the virulence, respectively. We also performed infection experiment on different type of host cells to show that the pattern of diminution and reacquisition of virulence towards amoeba can be applied to other hosts as well, including human cells. We also began the characterisation of an effector protein

called SidH, whose two transcribed genes are located at each side of the insertion sequence in the genome of the bacteria.

In the second part of this work, we tested the hypothesis that events of horizontal gene transfer may have been the main drive for the emergence of *L. pneumophila* as the main pathogenic species of the *Legionella* genus. To do so, we set up an experiment to test the natural competence for transformation in *L. norrlandica*, the closest relative to *L. pneumophila* in the evolutionary tree, but unknown to cause legionellosis in human. Although *L. norrlandica* possessed all required genes for natural transformation to occur, we weren't successful in triggering it in the tested conditions. We also showed that *L. norrlandica* can replicate inside amoebae, although to a lesser extent than *L. pneumophila*, but not in the human epithelial lung cells A549.

Taken together, these results show a new mechanism to regulate virulence in *L. pneumophila* and pave the way for future research using experimental evolution to study the emergence of pathogenic bacteria.

In French

Les bactéries du genre *Legionella* sont responsables chez l'homme d'une maladie appelée légionellose. Cette forme aiguë de pneumonie peut être mortelle dans 7 à 30% des cas, selon le statut immunologique du patient. Parmi plus de soixante espèces décrites à ce jour, le sérotype 1 de *Legionella pneumophila* est impliqué à lui seul dans 90% des cas de légionellose dans le monde. Cette surreprésentation clinique n'est pas liée à une prévalence accrue dans l'environnement. De plus, des études de génomique comparative n'ont pas permis d'identifier des gènes ou des fonctions spécifiquement présents dans les cinq sous-types les plus fréquemment associés à la maladie. Dans l'environnement, *L. pneumophila* colonise les écosystèmes aquatiques où elle est présente soit dans des biofilms mixtes, soit associée à des protozoaires hôtes (principalement des amibes) que la bactérie pénètre et utilise à son avantage pour se répliquer largement et finalement les tuer. Chez l'homme, la bactérie infecte les macrophages alvéolaires de la même manière que les amibes. L'élément clé de la virulence est le système de sécrétion Dot/Icm de type IV qui est utilisé pour manipuler l'hôte, en transloquant plus de 300 protéines bactériennes appelées effecteurs dans la cellule infectée. Plusieurs de ces effecteurs auraient été acquis par la bactérie par transfert horizontal de gènes au cours de sa coévolution avec ses hôtes protozoaires. Le vaste répertoire d'effecteurs ainsi que leur redondance fonctionnelle et la grande variabilité au sein des espèces sont autant de critères qui rendent extrêmement difficile la recherche d'un ensemble de gènes et de fonctionnalités expliquant la prévalence excessive de *L. pneumophila* dans les cas cliniques. Pour ce travail de thèse, nous avons utilisé deux expériences d'évolution expérimentale pour permettre à une souche clinique de *L. pneumophila* de réduire considérablement sa capacité infectieuse envers les amibes et de l'acquérir à nouveau dans un second temps. Le séquençage génomique réalisé en parallèle sur des spécimens à différents stades d'évolution nous a permis d'identifier le rôle d'un plasmide naturel (pLPP) dans la modification de la virulence de *L. pneumophila*. En effet, ce plasmide a la capacité de s'insérer dans le génome et de s'en exciser à l'aide d'une séquence d'insertion présente à la fois sur le plasmide et dans le génome. Ces événements d'insertion/excision sont directement liés à la diminution et à la ré-acquisition de la virulence, respectivement. Nous avons également réalisé des expériences d'infection sur différents types de cellules hôtes pour montrer que le schéma de diminution et de ré-acquisition de la virulence envers les amibes s'applique à d'autres hôtes, y compris les cellules humaines. Nous avons également

commencé à caractériser un effecteur appelé SidH, dont les deux gènes transcrits sont situés de part et d'autre de la séquence d'insertion dans le génome de la bactérie.

Dans la deuxième partie de ce travail, nous avons testé l'hypothèse selon laquelle des événements de transfert horizontal de gènes seraient le principal moteur de l'émergence de *L. pneumophila* comme principale espèce pathogène du genre *Legionella*. Pour ce faire, nous avons mis en place une expérience visant à tester la capacité naturelle de transformation de *L. norrlandica*, la plus proche parente de *L. pneumophila* dans l'arbre évolutif, mais qui n'a jamais été retrouvée en clinique. Bien que *L. norrlandica* possède tous les gènes nécessaires à la transformation naturelle, nous n'avons pas réussi à la déclencher dans les conditions testées. Nous avons également montré que *L. norrlandica* peut se répliquer à l'intérieur des amibes, mais dans une moindre mesure que *L. pneumophila*, mais pas dans les cellules épithéliales pulmonaires humaines A549.

Dans l'ensemble, ces résultats montrent un nouveau mécanisme de régulation de la virulence chez *L. pneumophila* et ouvrent la voie à de futures recherches utilisant l'évolution expérimentale pour étudier l'émergence de bactéries pathogènes.

General objectives

Bacterial pathogens often emerge or re-emerge owing to evolutionary changes that allow them to colonize new host species, transmit between individual hosts by new means, and resist therapeutic antibiotics. To identify the cellular and molecular mechanisms of emergence of new pathogens is crucial to understanding the origin of infectious diseases and to develop new therapeutic strategies.

Legionella is ubiquitous of aquatic environments, including both freshwater sources and human-made water distribution systems, but in some rare circumstances, it is also a human pathogen that can infect alveolar macrophages and cause a severe and often fatal pneumonia called Legionnaire's Disease (LD) ¹⁻⁵. Of the 60 species that comprise the genus of *Legionella*, only a small number of species have been reported to cause disease in humans, including the species *L. pneumophila* ⁶⁻⁸. It has been speculated that the interaction of *L. pneumophila* with protozoa generated a pool of virulence traits that enable the pathogenic species to successfully parasitize mammalian macrophages⁹. Thus, the emergence of *L. pneumophila* as a pathogen reflects complex evolutionary processes of general importance, involving historical factors, (*i.e.* traits selected in the history of the species) and environmental factors (host spectrum, aquatic environments, development of aerosol-generating devices...).

In this context, the objectives of my thesis were to understand the evolutionary history of *L. pneumophila*, but also to decipher the generic and dynamic evolutionary mechanisms that led an environmental bacterium to become a human pathogen. This project also aimed to combine the study of genome dynamics with that of the integrated and complex phenotype of virulence.

For several years, my host laboratory had addressed the question of the evolution of virulence and host spectrum *L. pneumophila*. In particular, they wondered whether bacteria evolving in the absence of hosts for hundreds of generations will lose or retain their infection ability and how?

To answer this question, they had initiated evolving lineages founded from a common ancestor, *L. pneumophila* Paris strain, a worldwide-distributed epidemic clone implicated in

outbreaks and sporadic cases of LD and frequently isolated from environment and for which the complete genome was sequenced¹⁰⁻¹². They had propagated six lineages in parallel for 700 generations through multiple passages of individual colonies on standard *Legionella* agar medium. Such evolution experiments, called “mutation accumulation”, reduce selection against slightly deleterious mutations and therefore allow the substitution of beneficial, neutral and slightly deleterious mutations. When I arrived at the laboratory, phenotypic comparisons between the ancestor and the evolved lineages had just begun and revealed a strong attenuation of the virulence after 700 generations of genetic drift. **The first objective of my thesis** was therefore to pursue the phenotypic, genetic and transcriptomic comparisons of the evolved clones. I characterized the phenotypic changes as well as the underlying mutations and transcriptional changes in 3 evolved lineages. Then, **in a second time**, I aimed to re-evolve one of those attenuated clones under conditions where the virulence and/or host spectrum were expected to rapidly evolve. The objectives were to assess how the selective pressure of a specific host regimen might improve the pathogenic capacities of attenuated virulent strain, and to determine what are the genetic and molecular mechanisms underlying these adaptations. As the evolved individuals from each lineage were isolated and frozen at different times, I was therefore able to “rewind the tape” of the evolution and to compare the evolutionary processes between lineages. The global idea of this axis was to identify genes or functions selected during evolution strategies as markers of increased virulence. The results of these axis of my thesis are described in chapter 2 of this manuscript (**Experimental evolution leads to the discovery of a novel mechanism of virulence control in *Legionella pneumophila***) and should give rise to a publication that we would like to submit to ISME Journal this autumn.

In the first part of my thesis, I focused on the clinical isolate Paris that was, in a way, already pre-adapted to the human host. Thus, in the second part, I wanted to focus on the evolutionary and adaptive capacities of a strictly environmental isolate. Indeed, while most, if not all, *Legionella* species are capable of infecting a broad spectrum of amoebae, only a small number are able to cause LD. Thus, another major challenge of my thesis was to understand which factors in *L. pneumophila* constitute a pre-adaptation to human cell infection.

The second axis of my thesis was then to develop a new strategy of experimental evolution from an environmental strain, *L. norrlandica*, genetically closest to *L. pneumophila* and never

implicated in clinical settings. Our initial objective was to propagate *L. norrlandica* within various hosts for hundreds of generations, in the presence or not of genomic DNA from *L. pneumophila* Paris. The underlying idea of this axis was to determine whether *L. norrlandica* was capable of acquiring DNA from Paris strain during the infection process, and to determine the consequences this might have on the evolution of its virulence and host spectrum.

The prerequisites for such an experiment were first, to determine whether *L. norrlandica* replication within host cells was sufficient to initiate this evolutionary strategy, and in a second time to assess its ability to acquire foreign DNA by natural transformation. We thus aimed at characterizing the replicative capacity of *L. norrlandica* within *Acanthamoeba* spp., U937 macrophages and also A549 human pulmonary epithelial cells. In parallel, we and colleagues from the laboratory, had confirmed that *L. norrlandica* has all the main genes necessary to undergo natural transformation so we sought to test its capacity to develop this competence state by acquiring antibiotic resistance through HGT and homologous recombination.

Disclaimer

This manuscript is the fruit of 3.5 years of research in the laboratory. The COVID-19 pandemic crisis and lockdown started at the end of this 3.5-year period, with all the consequences we all sadly experienced. At that time, I was trying to write this manuscript.

This stressful period precipitated my need to start new projects and to look for a new career path following this PhD.

I was fortunate enough to find a job position in the industry in the end of the summer of 2020 as a clinical research associate within a contract research organization.

But this meant having to learn a completely new and demanding job while pursuing the redaction of this manuscript whenever possible. In the end, it took me an extra three years... Few thought that this manuscript could ever be completed, and this could never have been done without the help of a handful of people, that I thanked in the acknowledgment section of this manuscript and thank again here, deeply.

1. Literature review

1.1. Experimental evolution methods and their contribution to the study of bacterial processes

1.1.1. Introduction

In the early 1800's, Jean Baptiste Lamarck (1744-1829) became one of the first naturalist to propose a theory on the evolution of species. He founded his hypothesis on the idea that Speciation occurs through the adaptation of the organs and behaviour of organisms to a given environment. In his ground-breaking book *On the origin of species*, Charles Darwin capitalized on Lamarck's work and succeeded – where Lamarck had stalled – in finding the driving force of evolution¹³. In 1859, he introduced a concept that later became a cornerstone of evolutionary biology: Natural Selection. Crucial to the development of this theory was Darwin's journey onboard the 'Beagle', a survey ship that completed a trip around the globe between 1831 and 1836. Darwin published the memoirs of this voyage in 1839 in a book called *Journal and Remarks* or more commonly *The voyage of the Beagle*¹⁴. Almost half of the trip consisted in the exploration of the South American continent. Traveling south, he encountered many different animal and vegetal species. One of particular interest to him was a family of birds called rheas. He witnessed slight changes in the birds' appearance as he went southward, with closely allied individuals replacing one another. Consequent to these observations, he ventured the hypothesis that the two types of birds descended from a common parent. Hence, he made a first step towards invalidating the theory of the transmutation of species – in which one argues that species change by blows – he previously supported¹⁵. As a side note, this goes against Lamarck's theory of the grading of species. Lamarck thought that species form a continuum from the simplest to the most complex. Interestingly, this theory doesn't allow for species extinction and according to Lamarck, species either go on living in parts of the world undiscovered at the time, or more importantly they simply change into new life forms¹⁶. The branching event from one common ancestor to new species was initially called 'inosculation' by Darwin, as a parallel to the medical term designating the joining of one blood vessel to another. To illustrate his theory, he drew one of the first phylogenetic tree on page 36 of his red Notebook B on the transmutation of species (**figure 1**).

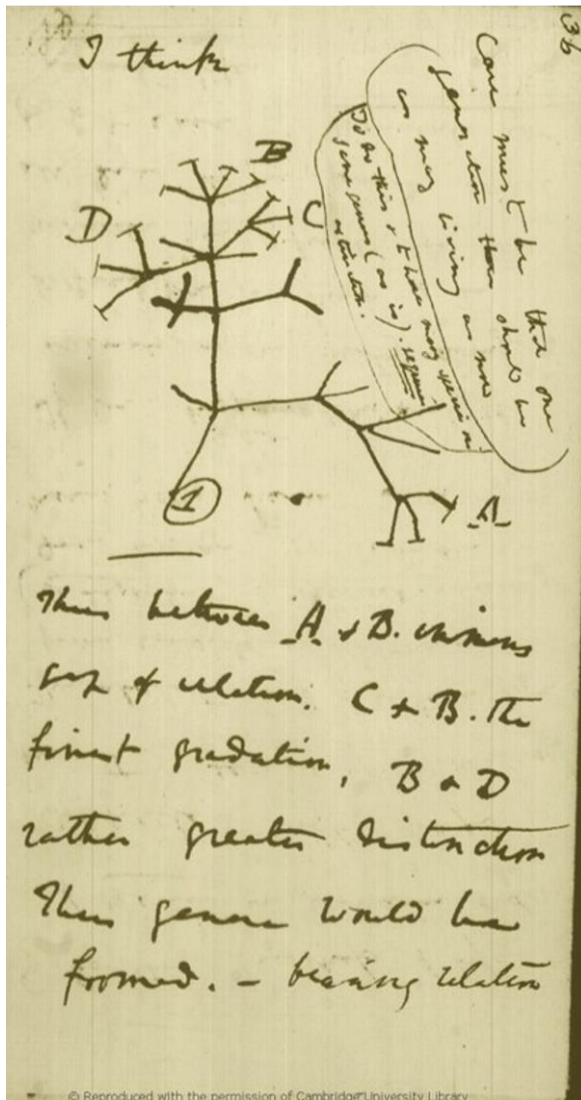


Figure 1: Charles Darwin's 'I think' tree.

Darwin made this sketch of a tree representing the potential link between different species, evolved from a common ancestor, in page 36 of his red Notebook B in 1837. The legend reads 'I think [sketch] Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction. Thus, between A & B immense gap of relation. C & B the finest graduation, B & D rather greater distinction. Thus, genera would be formed. - bearing relation'

A lot has changed since this first depiction of the relationship between species, living and extinct. Thanks to the rise of modern technologies, we now know that the evolution of species is quite more complicated when the evolution of the species' genomes is taken into account. Indeed, genes are not necessarily transmitted directly from one ancestor to its descendants but undergo events of horizontal transfer (see part 3 of this manuscript), duplication and loss. This leads to rather complicated evolutionary histories, when species trees and gene trees need to be reconciled ¹⁷ (figure 2).

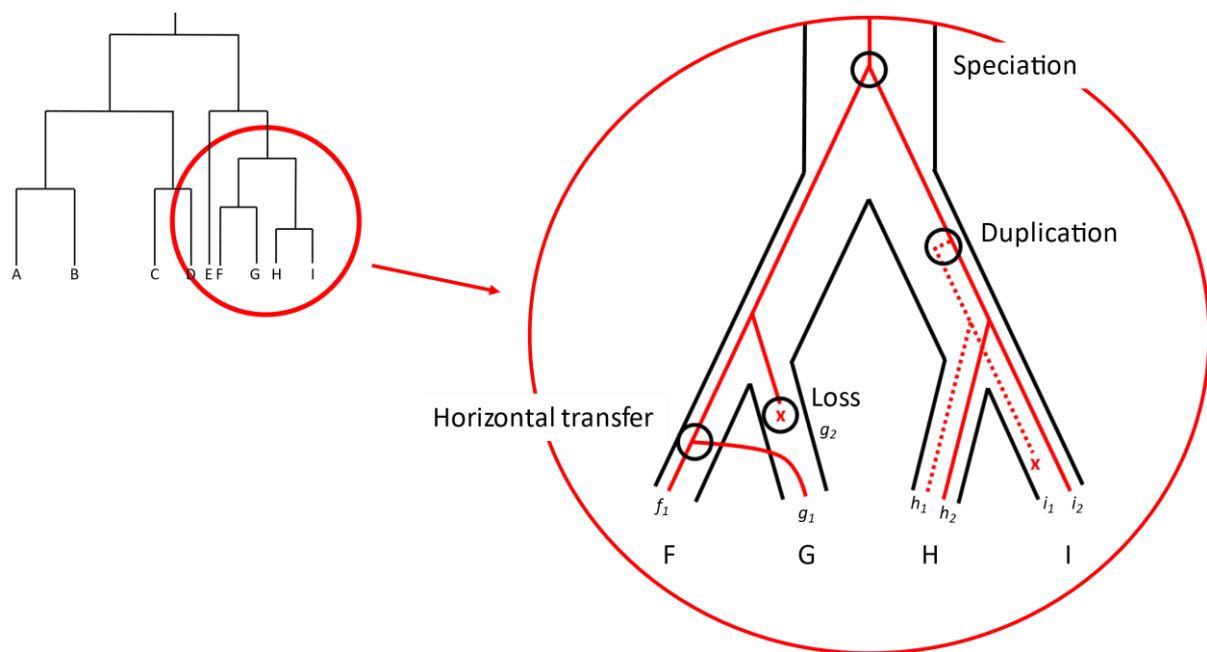


Figure 2: Reconciling a gene tree with a species tree.

A species tree (black lines) along with a gene history (red lines) involving events of speciation, duplication, horizontal transfer and loss. The leaves of the tree are labelled in capital letters for species and in lowercase, italic letters for gene copies.

Comparative genomics is indeed a tremendous asset if one wants to study the evolution of genomes and species, but it cannot alone answer all questions as a lot of it is based on the occurrence of probabilistic events. Scientists need other tools that, together with comparative analyses, could depict a more accurate vision of the evolution of organisms. Experimental evolution is one of those tools that allow the investigator to look at evolution in ‘real time’. Noteworthy, in his book *On the Origin of Species*, Darwin discussed the instance of artificial selection of varieties (pigeons, as an example), a setting that resembles that of modern experimental evolution experiments, i.e., the adaptation to a given environment. Powerful, yet having some drawbacks mainly linked to experimental settings, experimental evolution allows scientists to monitor the adaptation of model organisms to a specific environment. It allows them to understand the evolutionary processes leading to the rise of new metabolic abilities, new infectious strategies or the establishment of symbiotic relations.

But first, as now customary in this manuscript, a bit of history on the development of experimental strategies as a means to study evolutionary processes.

The reverend William Henry Dallinger (1839-1909) is considered one of the first scientist to set up a documented experimental evolution experiment. In addition to his clerical duties, he was an accomplished scientist who closely work with Darwin and became a pioneer in the field of protozoology. In 1880, he started a 7-year experiment on the adaptation of monads to increasing temperature. These quick-generating protozoans were first cultivated at 60°F (15°C). As Dallinger raised the temperature of the system in small increments, portions of the population died, forcing him to wait a month for the organisms to recover their former activity before raising the temperature again. In 1886, an accident caused the cultivating apparatus to be destroyed, thus ending the experiment with monads comfortably thriving at an astonishing temperature of 158°F (70°C). Very interestingly, the high temperature resisting monads weren't able to grow back at 65°F (18°C) in their natural cultivating medium ¹⁸.

This experiment proved the adaptability of organisms to slow changes and even if the cause of this adaptation wasn't known at the time, it paved the way for future research on the contribution of adaptive mutations to the evolution of organisms.

Kawecki and collaborators define experimental evolution as 'the study of evolutionary changes occurring in experimental populations as a consequence of conditions (environmental, demographic, genetic, social, and so forth) imposed by the experimenter' ¹⁹. With this definition, they exclude artificial selection due to breeding and insist on the random aspect of evolution. Indeed, they allow the action of experimental evolution on 'any and all traits and genes relevant to fitness under the environmental regimes of interest'.

One intrinsically inherent downside of using experimental evolution to study biological processes is the generation time of studied organisms. Even though it is basically possible to use it on every mutating (thus living) being, microbes are preferentially utilized compared to greater mammals or slow growing plants for instance. With that in mind, there is still a wide range of species used to answer a wide range of questions, a non-exhaustive list is being provided in **table 1**.

Hypothesis	Organisms
Mutation and adaptation	Bacteria, bacteriophages
Genetic drift and inbreeding	Insects
Environmental variability	Bacteria, plants, viruses, algae, digital organisms
Sexual selection and conflict	Insects, fungi, acarids, fishes
Life history and sex allocation	Insects, plants, fishes, acarids
Sexual reproduction and mating systems	Bacteriophages, bacteria, fungi, algae, rotifers, plants, digital organisms
Kin selection and cooperation	Insects, bacteriophages, bacteria, fungi, protozoa
Behaviour and cognition	Insects
Host-parasite interactions	Bacteria-bacteriophages, insects-fungi, algae-rotifers, gastropods-trematode, algae-bacteria
Speciation	Insects, fungi
Repeatability of evolution	Bacteria, bacteriophages, insects, digital organisms

Table 1: Table of hypothesis addressed with experimental evolution and model organisms.

This non-exhaustive list aims at providing examples of biological questions – with associated organisms – that can be answered using experimental evolution. Digital organisms are self-replicating computer programs that mutate, adapt and evolve²⁰. Adapted from Kawecki *et al.* 2012.

For a few decades, there has been an increase in the use of experimental evolution in the lab, that is why in this part we will particularly focus on bacteria as model organisms and we will see with a few examples how beneficial experimental evolution can be to study pathogenesis, metabolic adaptation, symbiosis, drug resistance and social behaviours in microorganisms.

Experimentally evolving bacteria leads to two main outcomes in terms of population genetics, depending on the initial hypothesis and the experimental settings (**figure 3**).

Mutation accumulation is a form of random mutagenesis experiment in which single bacterial colonies are picked and propagated on agar plates. This technique applies a strong bottleneck, drastically reducing the population size at each propagation or ‘passage’ as seen on the population size panel in **figure 3a**. These bottlenecks allow the bacteria to fix mutations randomly – regardless of their impact on fitness – and purge genetic diversity. Mutation accumulation experiments have concluded that the mean rate of single base mutations in bacteria is of the order of 10^{-10} – 10^{-9} per base per generation²¹, if hypermutator lineages are not taken into account. These particular lineages arise in some experiments when genes involved in key functions are mutated, such as the mismatch repair systems^{22,23}, leading to mutational rates as high as 138 fold what is naturally observed²⁴.

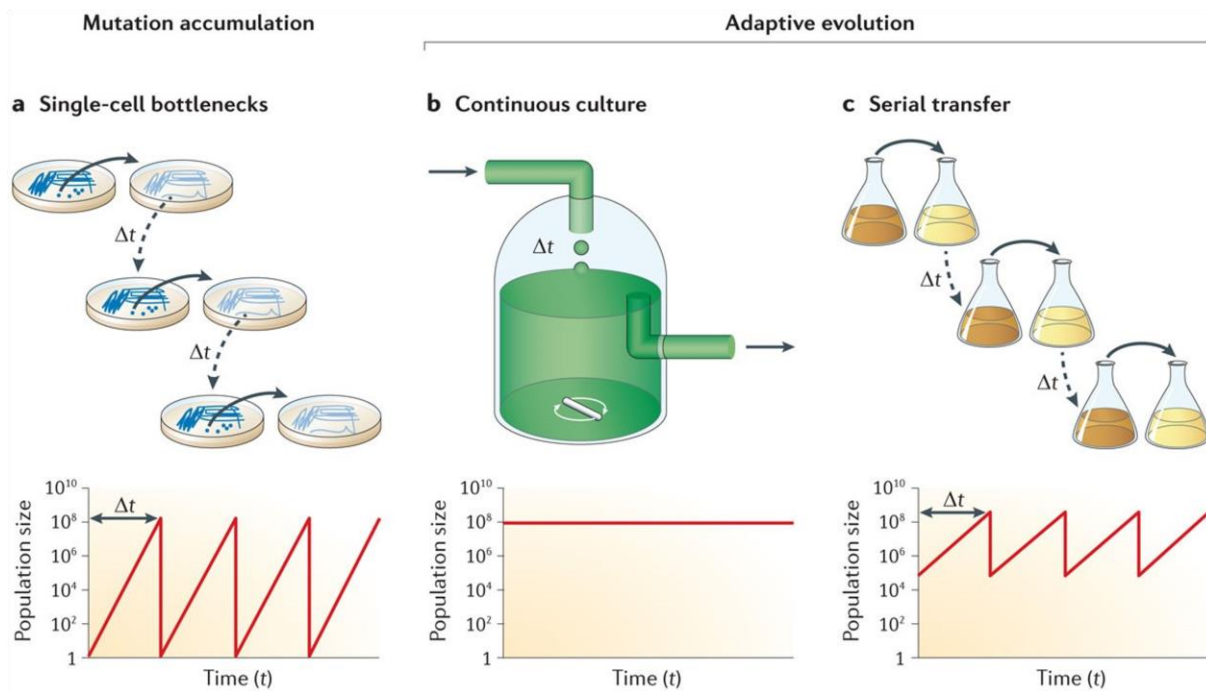


Figure 3: Schematic representation of experimental evolution strategies and associated population size progression.

Three types of population propagation experiments are depicted in the above panel with the associated population size evolution on the below panel. Mutation accumulation experiments use single-cell bottlenecks (a) and adaptive evolution can be achieved through continuous culture (b) or serial transfer (c). Barrick and Lenski 2013.

It is interesting to note that there is usually a bias in single base substitutions, depending on the mutated pathways. For instance, in *E. coli*, a loss of function in the *mutM* or *mutY* gene leads to increased G:C → T:A transversions, whereas a genome with an impaired *mutT* gene results in increased A:T → G:C transitions²⁵. As long as they are not too deleterious, mutations in this type of experiment can virtually impact all of the bacterial processes, making it a powerful without *a priori* technique of mutagenesis. Classically, mutation accumulation has been associated with whole genome sequencing to identify mutations underlying a changed phenotype. In those studies, a lot of the sequencing is done using the Illumina platform (see refs^{26–28} for examples), which is fine as it is efficient and usually answers desired questions at a ridiculously low price²⁹. However, we will show in the work described in this manuscript that short reads produced by Illumina sequencers are ideal for the identification of substitution mutation, but they lack precision when it comes to larger chromosomal rearrangements. Numerous examples in the literature tackle the problem of larger scale genome modification because they can alter a lot of bacterial processes. They can for instance be decisive to the rise of new pathogenic bacteria like the Shiga toxin-producing *E. coli* O157:H7 for instance,

whose toxin is believed to have been acquired through horizontal gene transfer mediated by bacteriophages^{30,31}. We argue that to have a broader vision of all modifications in the bacterial genome after experimental evolution, we should systematically associate Illumina-based sequencing to have shorter reads but a high coverage, with technologies like PacBio to have lesser coverage but longer reads.

The second main objective of experimental evolution is setting up assays to drive changes in organisms in response to a controlled environment. Called adaptive evolution, this technique allows the accumulation of bacterial mutants better adapted to the given environment than their ancestors. The simplest way to achieve this is to propagate a clonal population in a liquid medium, either in continuous culture where the input of fresh nutrients and the output of wastes are tightly controlled (figure 3b) or by serial dilution of part of the population in fresh medium (figure 3c). In continuous culture, the size of the population hardly varies across time, whereas a closed microcosm like the one depicted in figure 3c requires the dilution to allow the population to keep growing. The difference between serial transfer and mutation accumulation lies within the size of the population transferred from one passage to another. In serial transfer, there is no bottlenecks as a rather substantial part of the population is transferred to the next passage, allowing different variants to be propagated while a single clonal colony is peaked in mutation accumulation. This is pivotal as in serial dilution, the part of the population that accumulated the fitter changes would naturally dominate within the inoculum used to initiate the next passage.

As hypothesis become more complex, adjustments need to be applied to experimental setups, adding environmental parameters to the basic serial dilution and continuous culture setups. For instance, the team of Eric Alm looked at the adaptation of *Bacteroides fragilis* to the gut of healthy patients³², Jahn and collaborators studied the adaptation of *E. coli* to antibiotics under four different regimes³³ and Martin and collaborators studied the evolution of microbial interactions with *P. fluorescens* or *Burkholderia cenocepacia* in biofilms³⁴. As in mutation accumulation experiments, adaptive evolution is usually paired with whole genome sequencing.

Later on in this manuscript, experiments of adaptive evolution of *L. pneumophila* in liquid medium and *in vivo* in amoebae will be largely discussed (Carrillo *et al.* in preparation).

1.1.2. The Long-Term Evolution Experiment (LTEE) by Richard Lenski

No introduction on experimental evolution would be complete without mention of one of the most inspiring and prolific experiment in evolutionary biology, the long-term evolution experiment or LTEE. Inspiring, by the sheer simplicity of the initial setup, and prolific by the number of scientific articles that it fathered.

The LTEE, initiated in 1988 by Richard Lenski, represents a ground-breaking study that has provided invaluable insights into the dynamics of evolutionary processes. The experiment consists in the culture of twelve independent lineages of *E. Coli*, all initially founded by a single clone, propagated under identical conditions. Samples were regularly taken from each population and frozen for future analysis, allowing for the examination of genetic changes over time, and in response to a resource-limited environment. The number of results derived from this experiment is astronomical and the bacterial populations, reaching 75 000 generations as of June 2022, have been utilized to describe plenty of evolutionary aspects. Rather than making a complete review of all of them – which would be one (or more) PhD dissertation by itself – we will make a far from exhaustive summary of the main discoveries of this fascinating work.

1.1.2.1. Adaptation and fitness improvement

One of the key features to quantify adaptability of a population to a given environment is the measure of fitness, which classically results of the competition between evolved and ancestral populations. Being able to identify patterns of adaptation through measurement of fitness change or even better predict future adaptation of bacterial populations is a central part of most of studies interested in population evolution. In one of the (many) articles derived from this experiment, Lenski and his team tackled the presupposed idea that fitness must ultimately reach an upper limit beyond which it cannot further increase³⁵. They based themselves onto two models described by Wisner and collaborators: a hyperbolic model, in which there is an upper limit to fitness increased, and a power-law model, in which fitness virtually tends towards infinity in dependence on the logarithm of time. Both models however predict a decreasing rate of fitness improvement, meaning that changes to the fitness will have less and less positive impact on fitness, but without ever becoming negative³⁶. By performing competition assays using population clones from 40 000, 50 000, and 60 000 generations, they presented strong evidence that evolution preferably responds to the power-

law model, showing that all the lineages tested continued to accumulate mutations leading to increase of fitness of 5.1% between 40 000 and 60 000 generations. The power-law model predicted a fitness increase of 3.9% when the hyperbolic model predicted a fitness increase of 1.33%. These results are quite astonishing as they prove two points: first, that mathematical models created to explain evolution can be empirically validated and second, that evolution in the sense of fitness increase does not seem to have a limit in a fixed environment.

Interestingly, Wiser *et al.*'s power-law model also predicts that hypermutator populations that arose early in the LTEE will show higher rates of fitness improvement (an hypothesis verified in Lenski's study).

1.1.2.2. Mutational dynamics

Linked with the study of adaption, the LTEE has also served to study mutational dynamics in bacterial populations faced with long-term evolution in a constant environment since it remains to this day poorly documented empirically. By using high throughput genetic sequencing over 60 000 generations of bacteria derived from the LTEE, researchers from Harvard University were able to show that the dynamics of population genetic changes is vast and varied, contrasting to what was thought to be an evolutionary quasi-halt when approaching fitness optimum³⁷. Authors identified many point mutations but also a lot of events mediated by insertion sequence elements, suggesting that the study of evolution sometimes need to zoom out from the classical one nucleotide change → one mutation paradigm and use a combination of techniques allowing to account for larger chromosomal rearrangements. This reflexion will prove to show useful in later parts of this manuscript when discussing the thesis work. Interestingly, they close their argument by suggesting that evolution targets vary over time, by creating new 'niches' of evolution within the genome that is caused by everchanging ecological interactions and genetic background, thus allowing for mutation rate to stay high even after 60 000 generations. As mentioned above, models for genetic evolution, can predict for hypermutator phenotypes to arise in the population. It was observed in the LTEE where six of the twelve lineages evolved this phenotype due to mutations in the DNA repair system or the removal of oxidized nucleotides³⁸. However, hypermutator phenotypes not only increase the chance for the bacteria to evolve beneficial mutations, but they also equally contribute to the apparition of deleterious mutations. This happened in one of the lineages of the LTEE, for which the population exhibited a *mutT*

mutation, increasing the mutation rate by approximately 150-fold, but later fixating a *mutY* mutation, reducing the rate by approximately 50%³⁹. Study by Tenaillon *et al.* gracefully summarizes the global processes of genetic change in long-term evolution³⁸. Focusing on the non-hyper-mutator lineages, they showed that population genetic models need to account for beneficial mutation but also for neutral hitchhikers. They also produced the evidence that the majority of mutations in early stages of evolution are beneficial and that the proportion of those mutations declined over time while still remaining to account for – which fits the power-law model described above³⁶. They finally show that there is strong parallel evolution across the 12 lineages of the LTEE³⁸.

1.1.2.3. Parallel evolution and convergent phenotypic changes

One of the significant discoveries resulting from the LTEE is the phenomenon of parallel evolution. Parallel evolution refers to the independent emergence of similar traits or adaptations in different populations or lineages. Lenski's experiment revealed that despite the random nature of mutations and the complexity of evolutionary processes, certain genetic changes and phenotypic adaptations can occur repeatedly and predictably.

Through the course of the LTEE, occurrence of parallel changes in the evolving *E. coli* populations have been observed at the phenotypic level. We previously discussed the fitness increase, but we can also cite increase in cell size and modification of cell shape⁴⁰.

However, underlying genetic mechanisms for this parallel phenotypic evolution remains poorly understood. Study by Woods *et al.* on the LTEE bacteria showed that by sequencing 4 loci without *a priori* in the 12 lineages, it is possible to retrieve parallel changes at the gene level but that these changes were not common at the nucleotide level⁴¹. To better understand how genetic changes can lead to parallel phenotypes, Favate *et al.* took the instance of parallel fitness increase in the LTEE bacteria⁴². They very interestingly argue that genetic changes leading to parallel increase in fitness are indeed not linked to base pair mutation as what we discussed before, but they might rather come from modification at a larger scale, through the transcriptional landscape of the evolved bacteria. This result is fascinating as it may explain the link between disparate genetic mutations and similar phenotypic outcomes. This result is also pivotal to explain results of the thesis work presented later in the manuscript.

1.1.2.4. Historical contingency and chance events

We showed that the population of bacteria in the LTEE have evolved similarly in many ways, but they also diverge in some ⁴³, suggesting that although principally evolving in a parallel fashion, lineages take slightly different paths towards adaptation . These different paths can partly be explained by an evolutionary process called historical contingency, that can be roughly defined as the fact that ‘evolutionary outcomes are contingent on the prior history of the evolving population’ ⁴⁴. Once again, the LTEE lets us gain knowledge on another evolutionary process thanks to one of the lineage that evolved the capacity to grow on citrate which had been added to the culture medium merely as a chelating agent ⁴⁵. The favoured hypothesis to explain this extremely rare occurrence in *E. coli* (only second reported case of Citrate utilization in the bacteria mediated by mutation ⁴⁶) is as follows. The initial mutation allowing for seldom citrate utilization was present in early generations of evolution (around 20 000 generations ⁴⁷) , arising from a point mutation switching on the CitT transporter, that was only slightly beneficial. This can potentially happen at any time of the evolution. However, mutations kept on accumulating on this genetic background, cumulating at 33 000 generations where the population was able to efficiently grow on Citrate as a sole source of carbon ⁴⁸. These results suggest that Citrate utilization is constrained by previous evolutionary events that first allowed a slightly beneficial mutation to happen, before other refining mutations could arise.

Through observation, genomic analysis and documentation of key evolutionary events, the LTEE has made significant contributions to our understanding of evolutionary dynamics in bacteria, providing ground-breaking insights into various bacterial traits linked with evolutionary patterns. This experiment paved the way for future investigations and research on the fundamental principles underlying processes of evolution.

1.1.3. Experimental evolution and pathogenesis

Experimental evolution has been used quite early on, not so much in the scope of fundamental research but rather as a tool for the development of vaccines. The US army sponsored a study in 1945 on the development of a vaccine against Dengue virus. The authors reported that the cycling of the virus in mice resulted in an attenuated version that did not cause a severe reaction in healthy volunteer inmates of the State Prison of New Jersey. It

appeared clear to the authors that as soon as the 2nd passage in mice, the virus had already mutated but it wasn't before the 7th passage that they were convinced that the attenuated virus would be suitable to generate immunity in Humans ⁴⁹. Now, scientists are using more and more experimental evolution in the lab to study the evolution of pathogenesis.

One question that has been puzzling scientists for a long time are the reasons underlying host specificity or generalisation in pathogens, most of them infecting more than one host with the exceptions of few species ⁵⁰. *Salmonella* and *Yersinia* genus are particularly interesting when studying this. The different serovars of the *S. enterica* species do not exhibit the same regimen of host infection. While some subspecies of the *S. enterica* species are generalists and have reservoirs in mammals, birds and reptiles like *S. enterica* serovar Typhimurium, some are much more restricted like typhoidal *Salmonella* serovars whose only known reservoirs are Humans, or *S. enterica* s. Gallinarum which causes fowl typhoid in poultry but is avirulent for other species. Another interesting example are the three *Yersinia* species of bacteria. Two of them, *Y. enterocolitica* and *Y. pseudotuberculosis* are responsible diarrheal diseases in Humans and are associated with wild animals and can be found in the environment ^{51,52}. The third species is the infamous *Y. pestis* responsible for the plague. Although closely related to *Y. pseudotuberculosis* from which it recently diverged ⁵³, its mode of transmission from rodent to Human *via* fleas differ drastically from the faecal-oral transmission of its ancestors ⁵⁴. Adaptive evolution could then be a way to answer why some bacteria 'choose' either one of the described lifestyles and what impact it could have on the fitness of the bacteria to change from one regimen to the other. One way to verify this is to setup experiments in which bacteria are propagated in one host, are cycling between hosts or between the host and the environment.

L. pneumophila is a well-suited test subject for hypothesis of host cycling. As previously said, these environmental bacteria infect a wide variety of eukaryotic hosts and can, when inhaled by Humans, cause a potentially fatal pneumonia. Scientists have asked themselves over the years if the complex and variable environments in which they live could be the reason for their extraordinary host range (see part I of the manuscript). Ensminger and collaborators were the first to directly test host adaptation of *L. pneumophila* using experimental evolution ⁵⁵. In their study they described a protocol in which they propagated *L. pneumophila* inside mouse macrophages for hundreds of generations. They observed several mutations that improved the replication of the bacteria inside macrophages. However, some of these mutations,

though being beneficial to macrophage infection, happened at the expense of the fitness of the bacteria when it came to replication inside amoebae. This illustrates a concept of microbial ecology well-known to evolutionary scientists. The reason behind this modification of the pathogenic power is supposed to be a trade-off between the ability to infect various hosts and the cost of maintaining such functions. Trade-offs have been identified in almost every key functions (motility or biofilm, DNA repair or evolvability, metabolic rate or metabolic yield, etc) and they help structure bacterial diversity⁵⁶. As habitual in this type of experiment, they observed similar evolution between the parallel lineages they initiated from the same ancestor strain, suggesting that bacteria have 'preferred pathways' of evolution when constrained within the same type of environment. Reasons underlying this phenomenon are still not entirely clear although natural selection is believed to play an important part⁵⁷. Interestingly, research is conducted on influenza virus⁵⁸ and bacteriophage T7⁵⁹ to see if the repeatability of mutations in parallel evolution can be used to predict adaptability of organisms.

A second important issue in microbial ecology is to unveil the reasons that drive the evolution of environmental bacteria into pathogens. Answering this question can also be in the scope of experimental evolution. In the part 2 of this manuscript, we discuss a strategy in which a clinical strain of *L. pneumophila* with an experimentally attenuated virulence is propagated within amoebae during hundreds of generations. By 'forcing' the bacteria to infect its natural host, they reacquired their initial virulence, allowing us to identify novel virulence regulatory pathways (Carrillo *et al.* in preparation).

1.1.4. Experimental evolution and symbiosis

Symbiosis is justifiably a critical phenomenon in events of speciation, of such importance that a general consensus attributes the rise of the modern eukaryotic cell to endosymbiosis between an ancestral host (likely archaeal) and an α -proteobacteria around 1.5 billion years ago^{60,61}. This first association created the mitochondria inside the heterotrophic eukaryotic cell and in some cases can be followed by a second event of endosymbiosis, this time by the acquisition of a cyanobacteria to engender the ancestral autotrophic eukaryotic cell⁶¹.

More modern occurrences of symbiosis are documented between a large spectrum of species giving birth to complex, optimal relationships amongst somewhat distant organisms. Examples of bacterial symbiosis are given in **table 2**.

Phylum of Eukaryotic host	Host	Prokaryotic partner	Benefit
Protozoa	<i>Amoeba proteus</i>	<i>Legionella jeonii</i>	Unknown
Mollusca	<i>Lyrodus pedicellatus</i> (shipworm)	<i>Teredinibacter turnerae</i>	Cellulose degradation
	<i>Euprymna scolopes</i>	<i>Vibrio fischeri</i>	Bioluminescence
Arthropoda	Aphidae	<i>Buchnera</i> spp.	Provision of essential amino acids
Vertebrates	Ruminantia	<i>Bacteroides</i> spp. <i>Ruminococcus</i> spp.	Cellulose digestion
Plants	Leguminous	<i>Rhizobium</i> spp. <i>Bradyrhizobium</i> spp.	Nitrogen fixation

Table 2: Examples of bacterial-eukaryotic-host interactions.

Adapted from Steinert *et al.* 2000⁶² and McFall-Ngai *et al.* 2012⁶³

Symbiosis usually helps one or both of partners. The bacteria can help its host by digesting elements of its alimentation, helping it grow by providing amino acids or environmental nutrient that the host would not be able to acquire or produce by itself. The bacteria in turn receive protection by their host or are also provided by nutrients otherwise inaccessible for them. These types of symbiosis are beneficial and are often referred as ‘mutualistic’⁶⁴. Other instances of symbiosis are not so gentle for the host and often lead to adverse outcome. This is for example the case for the intracellular bacteria *Wolbachia* spp. which colonise several arthropods and are well known for the manipulation of the sexual reproduction of its hosts, ensuring its own vertical transmission⁶⁵.

One case of symbiosis highlighted in table 2 is puzzling enough to be worth a few lines in this manuscript. We refer to the association of *Amoeba proteus* with *Legionella jeonii* (ex X-bacteria). As previously mentioned, following internalisation by amoebae, *Legionella* face two distinct outcomes: either degradation by the endocytic pathway or hijack of the host cell machinery and establishment of a replicative vacuole. With *Legionella jeonii*, scientists have observed a third fate of the bacteria-amoeba relationship, unlike everything discovered yet. The xD strain of *Amoeba proteus*, derived from the D strain, arose spontaneously after infection of a gram-negative bacteria⁶⁶ later identified as belonging to the *Legionella* species and given the name *L. jeonii*⁶⁷. Once infected by the bacteria, the amoeba becomes

dependant of this relationship and dies if the bacteria are removed from their cytoplasm ⁶⁸. It has been experimentally tested that naïve amoebae require around 200 divisions in presence of the bacteria to develop this dependant bond ⁶⁹. Once established in their host, the bacteria divide in vesicles to reach a total of up to 42,000 individuals and induce genetic changes in the host nucleus ⁷⁰. *L. jeonii* is the only member of the *Legionella* species to form this type of relationship with amoebae. It would be very revealing to pursue on this work and try to explain how the *L. jeonii*-*A. proteus* association is possible. Scientists may employ experimental evolution here again since it has already been used to decipher the establishment of symbiosis for other pathogens.

In a 2010 article, Marchetti and collaborators experimentally evolved a plant pathogen into a legume symbiont. They argue that to achieve this symbiosis, the bacteria underwent metabolic changes following an event of horizontal gene transfer, a hypothesis we too defend in this manuscript to explain the emergence of *Legionella* bacteria as broad range pathogens. Marchetti and collaborators worked with a system composed of the devastating pathogen ⁷¹ *Ralstonia solanacearum*, the rhizobium bacteria *Cupriavidus taiwanensis* (rhizobia are legume-associated nitrogen-fixing bacteria ⁷²) and the *C. taiwanensis* host *Mimosa* spp. They showed that although necessary, transferring the symbiosis-promoting plasmid of *C. taiwanensis* inside *R. solanacearum* was not sufficient. To allow the pathogen to establish nodules on the roots of the host and intracellularly infect its cells, modified *R. solanacearum* cells underwent global genetic modification mainly resulting from the inactivation of the global virulence pathway regulated by *hrpG*. This inactivation was consequential to the *Mimosa* environment pressure exerted on modified *R. solanacearum* through cycling challenge of the host plant by the bacteria ⁷³. Hence, although the symbiotic relationship obtained with this experiment is not complete – the bacteria could not fixate nitrogen inside nodules – experimental evolution help prove that horizontal gene transfer alone is sometimes not enough to change a bacterium's regimen and rewiring of the virulence pathway is also necessary.

Symbiosis, as any mutualistic way of life, requires a fine-tuning of both the symbiont and the host response to one another in order for a partner not to overthrow the other. In eukaryote-microbe systems, the non-pathologic host innate immune system recognizes a microbe as beneficial or deleterious to properly react towards it. This recognition is mainly achieved by

pattern-recognition receptors that sense microbial molecules. These molecules can thus originate from either pathogens or residents of the host microbiota ⁷⁴. Although it is unclear how the host discriminates between the two types of microbes, its immune system is key in determining whether to engage a fight against or establish a symbiotic relation with the microbe.

A lot of work still needs to be engaged on this matter. However, some answers can be found using experimental evolution (we've seen it come a mile away, haven't we?). But as mentioned, it is a vastly uncharted territory so we will slightly shift our point of view and report an experiment of host-eukaryote interaction.

The fungus *Candida albicans* can be found in the gastro-intestinal tract of mammals where it remains under the control of the immune system of immunocompetent individuals but can become pathogenic in incapacitated patients. The hallmark of the switch from mutualistic to pathogenic is the formation of hyphae by the normally unicellular cells ⁷⁵. Tso and collaborators were interested in better understanding the establishment of mutualistic relations between mammals and microbes. In 2018, they published an article on the experimental evolution of *C. albicans* in the gut of antibiotic-treated mice. Through serial faecal transplants from colonised to naïve mice, they experimentally adapted the yeast to the gut. By doing that, they observed two interesting outcomes. The first one is the rise of clones of yeast with an attenuated virulence phenotype that were unable to form hyphae. Conversely, the evolved clones were less fit in competition with the ancestral, filamentous forms. Their second finding was that gut-adapted yeast conferred immunity to various pathogens, from fungal and bacterial origins (*Aspergillus fumigatus*, *Staphylococcus aureus* and *Pseudomonas aeruginosa*), through the stimulation of pro-inflammatory cytokines ⁷⁶. Manipulating host-immunity can then be an evolutionary strategy allowing the microbe to form a stable association with its host by protecting it and preventing other microbes to oust the symbiont from its environment.

The previous experiment has been conducted in guts of mice treated with antibiotics, thus without considering the resident microbiota. When the authors performed the experiment in normal guts, they were only able to observe the persistence of pathogenic forms of *C. albicans*⁵⁰. In Humans, the gut is home to more than 10^{14} microorganisms ⁷⁷, it is then impossible to assess the role of each member on the others. To simplify this intricate environment, King and collaborators developed a tripartite experimental evolution within the

gut of the worm *C. elegans* using two species of bacteria that normally colonize the gut of both Humans and worms: *Enterococcus faecalis* and *S. aureus* that are responsible for mild and severe infections, respectively. They either allow *E. faecalis* to evolve alone in the gut of *C. elegans* or they used a non-evolving *S. aureus* to serially co-infect the worm's gut with *E. faecalis*. As a side note, for the first co-inoculation, the authors allowed *E. faecalis* to colonize the gut 24h before infecting it with *S. aureus*. They also performed simultaneous infections of the worm's gut with *S. aureus* and *S. aureus*-naïve, gut-evolved *E. faecalis*. They elegantly showed that colonization of the gastrointestinal tract by *E. faecalis* protected the worm against *S. aureus* infection when *E. faecalis* is allowed to evolve conjointly with *S. aureus*. This protection was due to the increased production of antibacterial superoxide by *E. faecalis*. They did not however see any form of protection provided by *E. faecalis* when it had alone evolved within the host gut ⁷⁸. This exemplifies even more the extremely complex processes leading to the establishment of within-host mutualism and the importance of considering the host pre-existing microbiota.

1.1.5. Experimental evolution as a way to study emergence of resistance mechanisms in *L. pneumophila*.

The growing number of Antibiotic treatment failures in clinical settings is one of the most alarming public health issue in recent years with an estimate from the ECDC of 33000 deaths in Europe due to antibiotic resistant bacteria ⁷⁹.

In this part we will focus on studies regarding the antibiotic resistance schemes in our pathogen of choice, *L. pneumophila* and how experimental evolution can help unravel resistance mechanisms.

Historically, antibiotic resistance has not been a major concern when it comes to *Legionella* infections, even when despite appropriate use of antibiotics (macrolides and fluoroquinolones) a steady 10% death rate is observed in human infections. It has however been increasingly reported in recent years a capacity of *L. pneumophila* to develop resistance mechanisms giving rise to macrolide resistance such as efflux pumps-coding *lpeAB* genes ^{80,81} and single point mutation in 23S rRNA ⁸².

To help discover the mechanisms underlying the rise of macrolides and fluoroquinolone resistance in *L. pneumophila*, we will focus on two elegant experimental evolution studies by the teams of Sophie Jarraud⁸³ and Max Maurin⁸⁴, respectively.

In these two studies, several clones founded from the ancestral strain *L. pneumophila* Paris, responsible for clinical outbreaks in France and Europe, were propagated in parallel either on BCYE plates or ACES buffer, with increasing concentration of antibiotics. To reconstruct the evolutionary pathways of the bacteria under such conditions, clones of each lineage were sampled at several passages and frozen at -80°C to allow for later whole genome or targeted sequencing.

All clones under erythromycin (macrolide) pressure reached a MIC of 4096-fold the minimal starting concentration as soon as 9 to 11 passages and 7 to 14 passages for 3 clones under azithromycin (macrolide) pressure, the other ones reaching a MIC of 256-fold the initial concentration. Interestingly, the authors observed two distinct resistance profiles, irrespectively of the starting antibiotic pressure (erythromycin or azithromycin). Half of the final clones from the 12 parallel lineages developed high-level of resistance to both erythromycin and azithromycin in addition to other macrolides, while the other half developed high-level resistance to erythromycin but low-level resistance to other macrolides. The authors were able to associate the different resistance profiles to the number of mutated copies of the *rml* gene (23S rRNA), the mutated nucleotide, and/or the nature of the mutation. However, as *rml* mutations were combined with *rplD* (50S ribosomal protein L4) and/or *rplV* (50S ribosomal protein L22) mutations in all but three lineages, absolute direct correlation of macrolides resistance and *rml* profile is not attainable.

Study by the team of M. Maurin focused on the effect of selective pressure on the Quinolone Resistance Determining Regions (QRDRs) of the DNA gyrase *gyrA* according to a previous study⁸⁵ but also *gyrB* and topoisomerase IV (*parC* and *parE*) genes.

They showed that some clones under moxifloxacin (fluoroquinolone) pressure reached a maximum MIC of 512-fold the minimal starting concentration as soon as 8 passages, with the lineages reaching a MIC of 16 or 32mg/L after 12 passages when the starting sub-inhibitory concentration was 0.5mg/L.

Authors were able to reconstruct the mutational pathways of the different clones, as previously said, by sequencing DNA gyrase and topoisomerase IV coding genes at different steps of evolution. They observed first a nucleotide substitution in the *gyrA* gene in all clones

of all lineages as early as the fourth passage. The next detected mutation was in the *parC* gene with 6 of the 8 lineages bearing the same codon change (codon 78) at passage 6 or 7 depending on the lineage. The two remaining lineages experienced transient mutation by displaying a different mutation (codon 80) at passage 6 or 8, but eventually also harboured mutation in codon 78 at the end of the experiment. Finally, clones from lineages 6 to 8 had another mutation in the *gyrA* gene, with the other 5 lineages experiencing either of two mutations in *gyrB* gene before eventually developing the same second *gyrA* mutation.

Excitingly, authors were able to link the mutational steps with the incremental increase in moxifloxacin resistance of the bacteria, leading to the final combination of the two mutations in *gyrA* + the mutation in *parC* + one of the mutation in *gyrB* responsible to the final MIC of 32mg/L. Noteworthy, authors mentioned that some mutations on other genes than *gyrAB* and *parC* were probably missed, especially for lineage 4 where MIC increased from 16mg/L to 32mg/L with no apparition of the second *gyrA* mutation.

Although clones from the two studies underwent different evolutionary pathways with genome modification at different locations giving rise to either fluoroquinolones or macrolides resistance, it is extremely interesting to insist on the fact that clones from each study evolved in a parallel fashion, accordingly to their own selective pressure. This suggests that evolution works in 'preferred' pathways where fittest mutations tend to not only accumulate in the population, which is a well-known concept, but also same mutations or combination of mutations arise to give selective advantage under a given environmental pressure. This further proves the extraordinary adaptation capacity of living organisms and more so of bacteria, emphasising the importance of research to study mechanisms responsible for the rise of antibiotic resistance.

1.2. *Legionella pneumophila*, a fascinating emerging bacterial pathogen

1.2.1. *Legionella pneumophila*

Legionella pneumophila is an aerobic, facultative intracellular, Gram-negative bacillus belonging to the gammaproteobacteria class. Depending on its growth phase, it can be mobile with a lopho- or polytrichious flagellar arrangement. Bacteria measure 0.3 to 0.9 μm in width and 2 to $\geq 20\mu\text{m}$ (when in a filamentous form) in length ⁸⁶ (**figure 4**). They are present in both natural and man-made freshwater ecosystems. In the later, they thrive in temperatures ranging from 25°C to 45°C ⁸⁷. However, they have been documented to resist to a higher array of temperatures in the environment (from 6 to 63°C) ⁸⁸. *L. pneumophila* cells can be present either in a planktonic state, associated with mixt biofilms ⁸⁹ or parasiting a large array of Protozoan hosts ranging from *Amoebozoa* to *Percolozoa*, and *Chromista* ⁹⁰. However, amoebae is the predominant phylum which *L. pneumophila* is isolated from with more than 20 amoebae species described, such as *Acanthamoeba* spp. or *Vermamoeba vermiformis* ⁹⁰. Soil has also been identified as a reservoir for *L. pneumophila* ⁹¹, where they can be associated with amoeba species from genera such as *Acanthamoeba* or *Naegleria* ⁹².

The mechanisms responsible for the survival of the bacteria inside this broad range of hosts also enable the bacteria to infect Humans. This accidental contamination occurs through the inhalation of little droplets of water containing the bacteria. Once the bacteria reaches the lung, it can replicate inside alveolar macrophages and pneumocytes ^{93 94}. The multiplication and destruction of the pulmonary cells eventually causes a severe pneumonia called Legionnaire's disease or legionellosis.

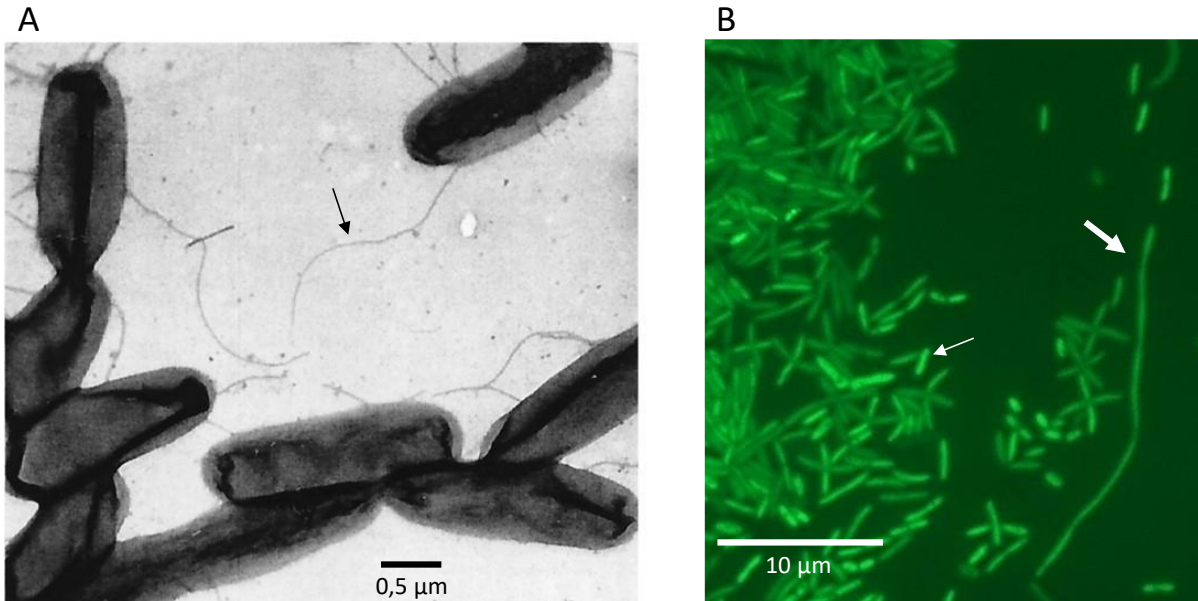


Figure 4: Microscopic images of *Legionella pneumophila*.

A: Transmission electron microscopy of stationary phase *L. pneumophila* with the black arrow pointing at the flagellum. (adapted from Dietrich *et al.* 2001)⁹⁵ **B:** Epifluorescence microscopy of CFP-stained *L. pneumophila*. The arrows point at the classical form of the bacterium (thin arrow) and the filamentous form (thick arrow).

1.2.2. Legionellosis

Legionnaire's disease's or legionellosis' first outbreak followed a reunion of members from the American Legion in July 1976 in the Bellevue-Stratford Hotel in Philadelphia, USA. More than 2,000 legionnaires, with a majority of elderly males, attended the event. Three days after the convention ended, the first fatality was recorded with the death of Ray Brennan, a 61-year-old Air Force and American Legion veteran. Within a week following Brennan's death, more than 130 people, mostly men had been hospitalized. They all suffered from the same symptoms, namely fatigue, chest pain, cough, lung congestion and fever. The total number of cases, of both congress attendees and from the surrounding area of the hotel rose to 182, from which 29 persons died⁹⁶ 4. The role of Joseph McDade – whose name was later given to the species *Legionella micdadei* – was pivotal in the discovery of the etiological agent of what was then referred in media as "Legion fever" or "Philadelphia fever" (B. D. Colen (January 31, 1977). "'Legion Fever' Germ Killed 16 Here in 1965". Washington Post). After overlooking at what resembled rod-shape bacteria in slides of guinea pig tissues injected with lung tissue from deceased patients, McDade decided to resume its experiment of cultivating material from the patients in egg yolks, only this time he did not add antibiotics. The eggs eventually died, driving McDade to look at the content under a microscope. He grazed at

samples full of bacteria that would later be identified and subsequently named *Legionella pneumophila* in 1977⁹⁷ 5. The bacteria had been replicating in the cooling tower of the air conditioning system of the Bellevue-Stratford hotel and then spread to the entire building ⁹⁸ (Ronald Kotulak (August 31, 1986), "Legionnaires' Disease Less Mysterious, Still Deadly", Chicago Tribune, p. 3, retrieved January 18, 2014). Scientists discovered that the transmission was airborne, through water droplets, explaining why people both inside and around the hotel got sick.

Now, *Legionella* are responsible for 10 – 15 cases of LD per million inhabitants in Europe, the United States and Australia⁹⁹. Legionellosis is a reportable disease, allowing authorities to keep track of the incidence of the disease through appropriate surveillance networks. We are facing an increased incidence with +31% cases between 2018 and 2017 in France, with 2133 cases reported (1630 cases in 2017)¹⁰⁰. In 2018, the number of cases reached its highest value since the initiation of the legionellosis surveillance network by Santé Publique France in 1988 (figure 5).

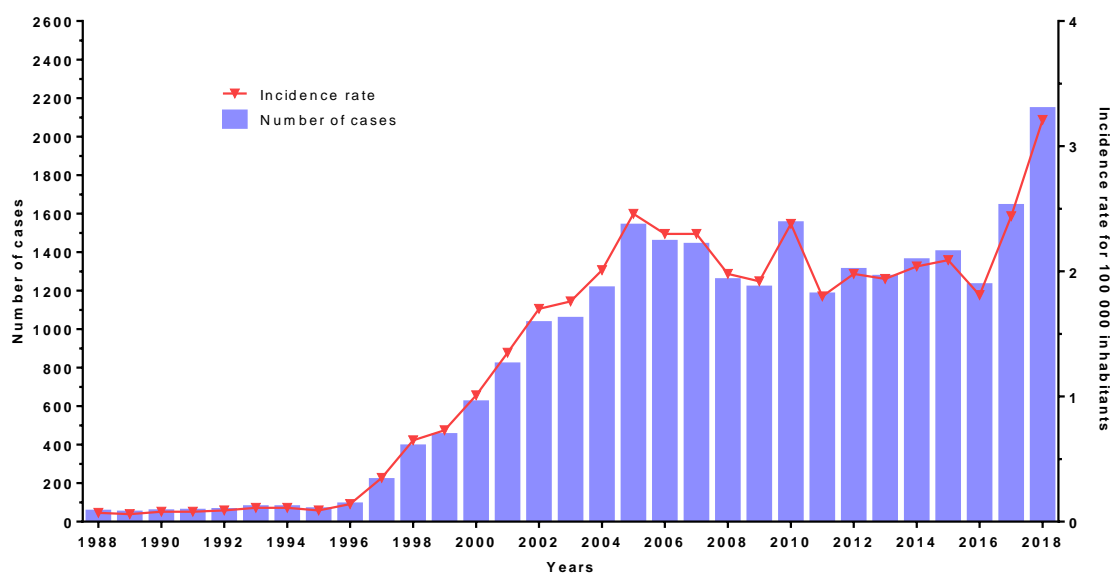


Figure 5: Evolution of the number of cases and the incidence rate of Legionellosis in France
Data from Santé Publique France

The French national reference centre for legionella infection defines as Legionnaire’s disease a radiologically confirmed pneumonia with clinical manifestations progressively appearing over 2 to 3 days. Those symptoms include an asthenia, a mild fever at the beginning of the illness raising to 39-40°C at day 3, myalgia and headache. A non-productive cough is initially observed, with a progression towards the production of mucoid and/or bloody sputum. This

clinical presentation can be associated with gastrointestinal disorders such as diarrhoea, nausea and vomiting and neurological disorders such as confusion and delirium. The chest X-ray shows an often-bilateral pneumonia most often systematized with an alveolar or alveolo-interstitial syndrome. The infection can lead to respiratory failure, acute renal failure, as well as extra-pulmonary manifestations (endocarditis, etc.) and rhabdomyolysis. A number of factors can explain the prevalence in certain part of the population such as a compromised immune system, a possible co-morbidity, smoking, being a male, an advanced age, a delay in starting a suitable treatment. Legionellosis is exceptionally encountered in children (only in very immunocompromised children).

To confirm a case, the clinical and radiological signs of pneumonia must be associated with one of the following biological criteria: isolation of *Legionella* spp. in a clinical sample (mainly sputum), a 4-fold increase in antibody titres with a minimum second titre of 128 and the presence of soluble urinary antigens. Only the cultivation of the bacteria from lung sampling or haemocultures allows the identification of the species and the strain, subsequently leading to the diagnosis and the epidemiological survey. PCR amplification can also be used to establish a diagnosis of legionellosis, particularly when non-*pneumophila* *Legionella* are implicated¹⁰¹.

Antibiotic treatment depends on the severity of legionellosis and the general health of the patient:

- in non-severe cases (patient hospitalised in ambulatory, classic or emergency wings): monotherapy with macrolides, preferably azithromycin or clarithromycin.
- in severe cases (patient hospitalised in the ICU, in reanimation and/or immunocompromised):
 - o either monotherapy with fluoroquinolones (levofloxacin, ofloxacin or ciprofloxacin, by order of preference)
 - o or through the association of two of the three following families: macrolides of class IV (spiramycin rather than erythromycin), fluoroquinolones and rifampicin (association with rifampicin are not to be encouraged).

Treatment usually lasts from 8 to 14 days for the non-severe cases and up to 21 days for the severe cases or in immunocompromised patients.

The case-fatality of legionellosis ranges from 8 to 25%, mostly depending on the immunological status of the patients, even with the use of an appropriate treatment.

Figure 6 depicts the number of cases and associated case-fatality (available for cases for which the evolution of the disease is known, between 77 and 97% of cases) in France between 1996 and 2018. The case-fatality dropped starting from 2000 and stabilised at around 11% in 2002 and onward. It has been suggested by experts that this can be explained by a better and an earlier detection of the disease through the titration of urinary antigens. There is no explanation however about why the lethality rate is constant at 11%. In 2018, there has been a particularly high number of cases with a relatively low associated case-fatality. Patients were globally younger than usual, thus more likely to clear the infection.

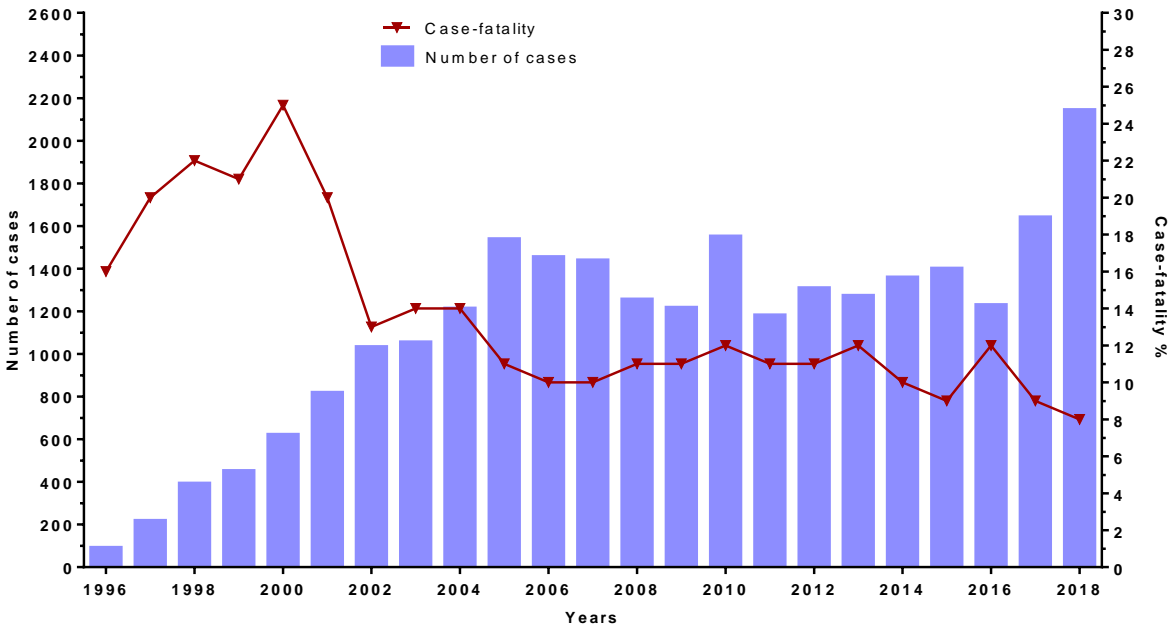


Figure 6: number of cases and the associated case-fatality of legionellosis in France between 1996 and 2018.

Data from Santé Publique France

Bacteria from the *Legionella* genus are also responsible for a milder disease called Pontiac fever¹⁰². It has been described as an influenza-like syndrome without pneumopathy and a spontaneous clearing in 2 to 5 days.

1.2.3. Ecology and genetics

Strikingly, only one case has been documented of probable man-to-man transmission of LD¹⁰³, putting the emphasis on the importance of environmental reservoirs. Furthermore, several studies showed that cycling to amoebae leads to *L. pneumophila* with an increased virulence, compared to bacteria grown in broth^{104 105 106}. Garduño and collaborators formed the hypothesis that highly infectious forms of *Legionella*, called MIFs for Mature Intracellular Form, accumulate after the growth, replicative phase. These MIFs are highly environmentally resistant, highly infectious and harbour the post-stationary phase characteristics typical of the transmissible environmental form causing LD¹⁰⁷. As a result, it is crucial to understand how *Legionella* is able to accidentally cross the species barrier and establish as an accidental human pathogen. Hence, we must further investigate the ecology of *Legionella* and the intricate relation between the bacteria and their environmental niche.

1.2.3.1. *Legionella pneumophila* and protozoa

As previously mentioned, *Legionella* cells in the environment are mainly associated with protozoa in liquid ecosystems. The diversity of unicellular eukaryotes in which *Legionella* can replicate is astounding and only few of the protozoa species and even genera that could facilitate *Legionella* replication are thought to be described.

Two recent reviews extensively screened the literature to try and have a broad, quite nearly exhaustive view of experimentally validated protozoan hosts of *Legionella pneumophila*^{108,90}(**Table 3**).

Using several techniques of microscopy (epifluorescence, phase-contrast, electron, confocal and light) as well as co-culture and CFU counting, the authors from the studies mentioned in these reviews were able to identify the protozoan hosts at the species level (sometimes even stains), the strains of *L. pneumophila* and the fate of these bacterial strains after engulfment by the protozoa.

Grazing of *L. pneumophila* by protozoa have different outcomes depending on which *L. pneumophila* strain is phagocytized by which protozoan strain. *L. pneumophila* can undergo three different main events once inside their hosts: intracellular multiplication (e.g. *Acanthamoeba castellanii* + *L. pneumophila* strain Paris)¹⁰⁹, intracellular survival (e.g. *Naegleria fowleri* + *L. pneumophila* strain Dallas 1E)¹¹⁰ and packaging and expulsion of live bacteria in pellets (e.g. *Tetrahymena tropicalis* + *L. pneumophila* strain Lens)² (**Table 3**).

Protozoan genus	Protozoan species	Phylum	Fate of <i>L. pneumophila</i>
<i>Acanthamoeba</i>	spp., sp., <i>astronyxis</i> , <i>castellanii</i> , <i>lenticulata</i> , <i>palestinensis</i> , <i>polyphaga</i> , <i>royreba</i>	<i>Amoebozoa</i>	Intracellular multiplication, live cells packaged in expelled pellets, intracellular survival
<i>Balamuthia</i>	<i>mandrillaris</i>	<i>Amoebozoa</i>	Intracellular multiplication
<i>Ciliophrya</i>	sp.	<i>Ciliophora</i>	Intracellular survival
<i>Dictyostelium</i>	<i>discoideum</i>	<i>Amoebozoa</i>	Intracellular multiplication
<i>Echinamoeba</i>	<i>exudans</i>	<i>Amoebozoa</i>	Intracellular multiplication
<i>Hartmannella</i>	<i>cantabrigiensis</i>	<i>Amoebozoa</i>	Intracellular multiplication
<i>Naegleria</i>	spp., <i>fowleri</i> , <i>gruberi</i> , <i>jadini</i> , <i>lovaniensis</i>	<i>Percolozoa</i>	Intracellular multiplication, intracellular survival
<i>Oxytricha</i>	<i>bifaria</i>	<i>Ciliophora</i>	Intracellular survival
<i>Paramecium</i>	<i>caudatum</i>	<i>Ciliophora</i>	Intracellular multiplication
<i>Stylonychia</i>	<i>mytilus</i>	<i>Ciliophora</i>	Intracellular survival
<i>Tetrahymena</i>	sp., <i>pyriformis</i> , <i>thermophila</i> , <i>tropicalis</i> , <i>vorax</i>	<i>Ciliophora</i>	Intracellular multiplication, live cells packaged in expelled pellets, intracellular survival
<i>Tetramitus (Vahlkampfia)</i>	<i>jugosus (jugosa)</i>	<i>Percolozoa</i>	Intracellular multiplication
<i>Vermamoeba (Hartmannella)</i>	<i>vermiformis</i>	<i>Amoebozoa</i>	Intracellular multiplication, intracellular survival
<i>Willaertia</i>	<i>magna</i>	<i>Percolozoa</i>	Intracellular multiplication

Table 3: Experimentally confirmed *Legionella pneumophila* protozoan hosts.

Adapted from Boamah et al. 2017 and Best and Kwaik 2018. All bacterial species were identified by culture and/or various microscopical techniques.

The overwhelming majority of cases of Legionnaire’s disease are due to *L. pneumophila*¹¹¹, explaining why this study mainly focuses on this species (aspects of the virulence of this bacteria will be discussed further on). However, we feel one particular case of amoeba ‘food-poisoning’ by the species *Legionella steelei* is surprising enough to deserve a bit of explanation. *Solomitrus palustris* is a member of the *Percolozoa* phylum that prey on *L. pneumophila*. Following internalisation by the protozoa, the bacteria is consumed as a source of nutrients, in a fashion similar to the degradation of non-protozoa-degrading bacteria like *Escherichia coli*^{112 113}. *L. steelei* has been documented to cause Legionnaire’s disease and to actively replicate

inside *A. castellanii*¹¹⁴ but Amaro and collaborators reported the first instance of *L. steelei* – and more generally *Legionella* – killing of its host through a mechanism akin to food-poisoning¹¹³. Co-culture experiments between the two organisms showed a decreased in viable trophozoites, with only about 0.5% of the amoebae being able to grow back on *E. coli*, after 72h of incubation. Furthermore, this killing is dependent on the type 4 secretion system of the bacteria, which is known to deliver various proteins inside the host cytosol¹¹⁵. The bacteria are not able to replicate inside the amoebae but rather inflict cellular damage to their hosts such as less granular cytoplasm, damaged mitochondria, and granular masses of degraded cytoplasm. Finally, the toxic effect of the bacteria on the amoebae is dependent on the bacteria to amoebae ratio, with a more lethal effect following an increase in the foresaid ratio. All in all, these results are consistent with a food-poisoning-like mechanism of resistance deployed by *L. steelei* to resist degradation by *S. palustris*.

Although it appears that *L. pneumophila* can replicate inside a wide variety of hosts, three main genera overshadow the others. These being *Acanthamoeba* (which species include *A. castellanii*, *A. astronyxis*, *A. lenticulata*, *A. palestinensis*, *A. polyphaga* and *A. royreba*), *Naegleria* (*N. fowleri*, *N. gruberi*, *N. jadini* and *N. lovaniensis*) and *Tetrahymena* (*T. pyriformis*, *T. thermophila*, *T. tropicalis* and *T. vorax*). Interestingly, these genera are from three different phylum, spanning across the Eukaryote superkingdom (**figure 7 and 8**): *Acanthamoeba* belongs to the *Amoebozoa* phylum, *Naegleria* to the *Percolozoa* phylum and *Tetrahymena* to the *Ciliophora* phylum.

The *Amoebozoa* phylum is perhaps one of the most interesting group of Protista. Its evolutionary interest has been assessed by Cavalier-Smith and collaborators^{116 117}. They argue that the last common ancestor to all amoebozoan probably was a unikont, suggesting a close evolutionary link between *Amoebozoa* and *Opisthokonta* (which contains *Fungi* and *Metazoa*). Since the 18S rRNA gene trees of *Amoebozoa* have failed to clearly resolve its internal topology, external characteristics remain the gold standard to assign a particular species to this phylum. However, the difficulty of this task lies in the extremely diverse forms and shapes *Amoebozoa* members can take, as they include amoeba slime moulds (i.e. *Dictyostelium discoideum*), naked (lacking any hard coverings like *Acanthamoeba* sp.) and testate (enclosed within a hard shell, like members of the *Arcellinida* order) amoebae and some flagellates (plasmodial slime moulds from the *Myxogastria* class). Even with so many differences

amongst its members, the monophyly of the *Amoebozoa* phylum is quasi unrefuted, with the pleiotropy of phenotypes being explained by multiple losses of characters from the ancestral unikont (such as branched tubular mitochondrial cristae and non-filipodial amoeboid motion) across the evolutionary tree, as well as the independent apparition of others ¹¹⁷ (**figure 9**). Later work from the same author showed significant changes in the phylogeny of *Amoebozoa*, with the dichotomy of the phylum in two sub-phyla: *Lobosa* and *Conosa*, which in turn separate into *Tubulinea* and *Discosea*, and *Archaemoeba* and *Semiconosia*, respectively. However, the monophyly of the phylum is still undisputed ¹¹⁸.

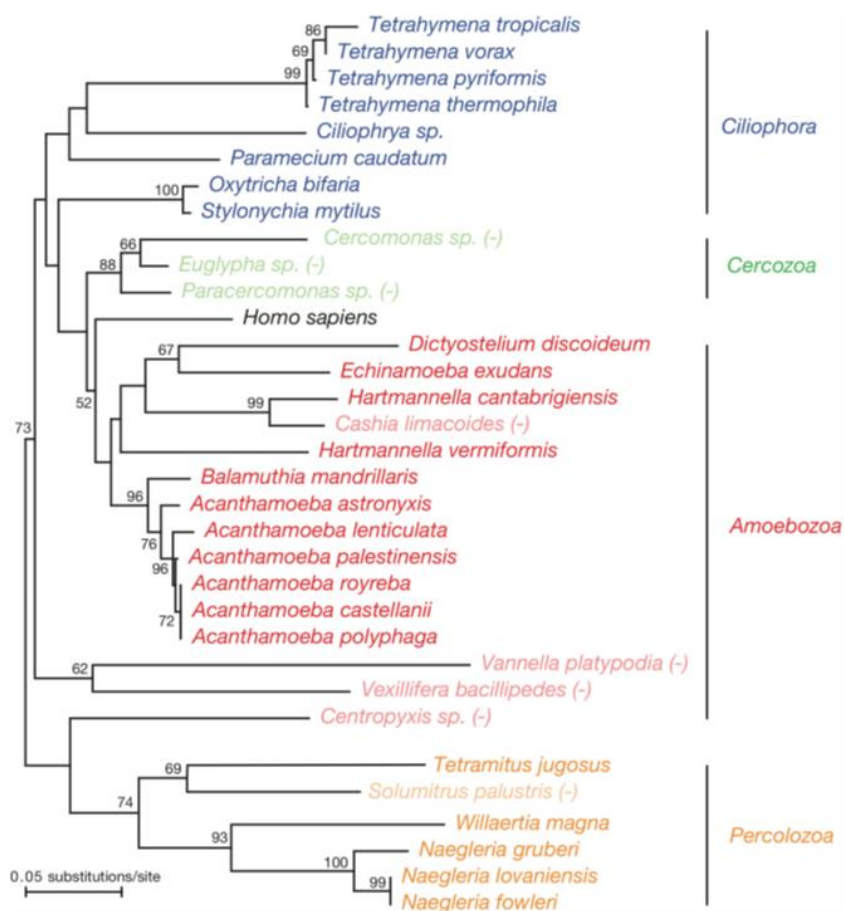


Figure 7: Phylogenetic tree of the experimentally defined hosts of *L. pneumophila*

Adapted from Boamah *et al.* 2017. Tree was built on 18S sequences alignments using the Neighbor-Joining method. Species that do not support *L. pneumophila* replication or survival are indicated by a lighter shading and the annotation “(-)”. Taxonomic phylum designation is based on the classification system outlined by Ruggiero *et al.* (2015) ¹¹⁹.

This proves that, not unlike other phyla, the phylogeny of *Amoebozoa* is constantly changing and adapting to the use of new phylogenetic tools and the growing contribution of whole genome sequencing.

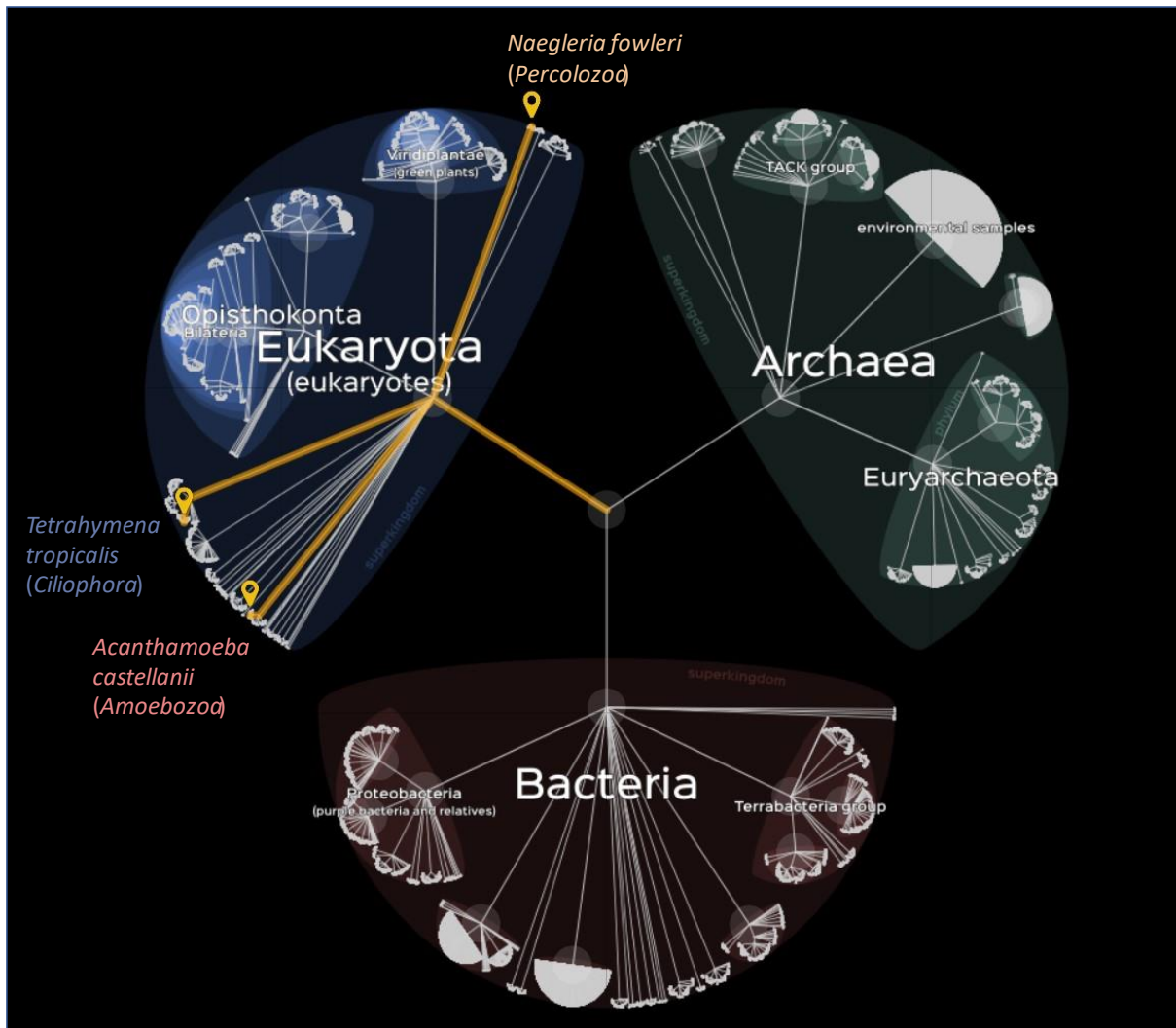


Figure 8: *L. pneumophila* hosts span across the Eukaryota superkingdom

Representation of the Tree of Life with three *L. pneumophila* eukaryotic hosts highlighted. These hosts span across the entire Eukaryota superkingdom, from Amoebozoa to Percolozoa. From Lifemap: Exploring the Entire Tree of Life, de Vienne DM (2016)

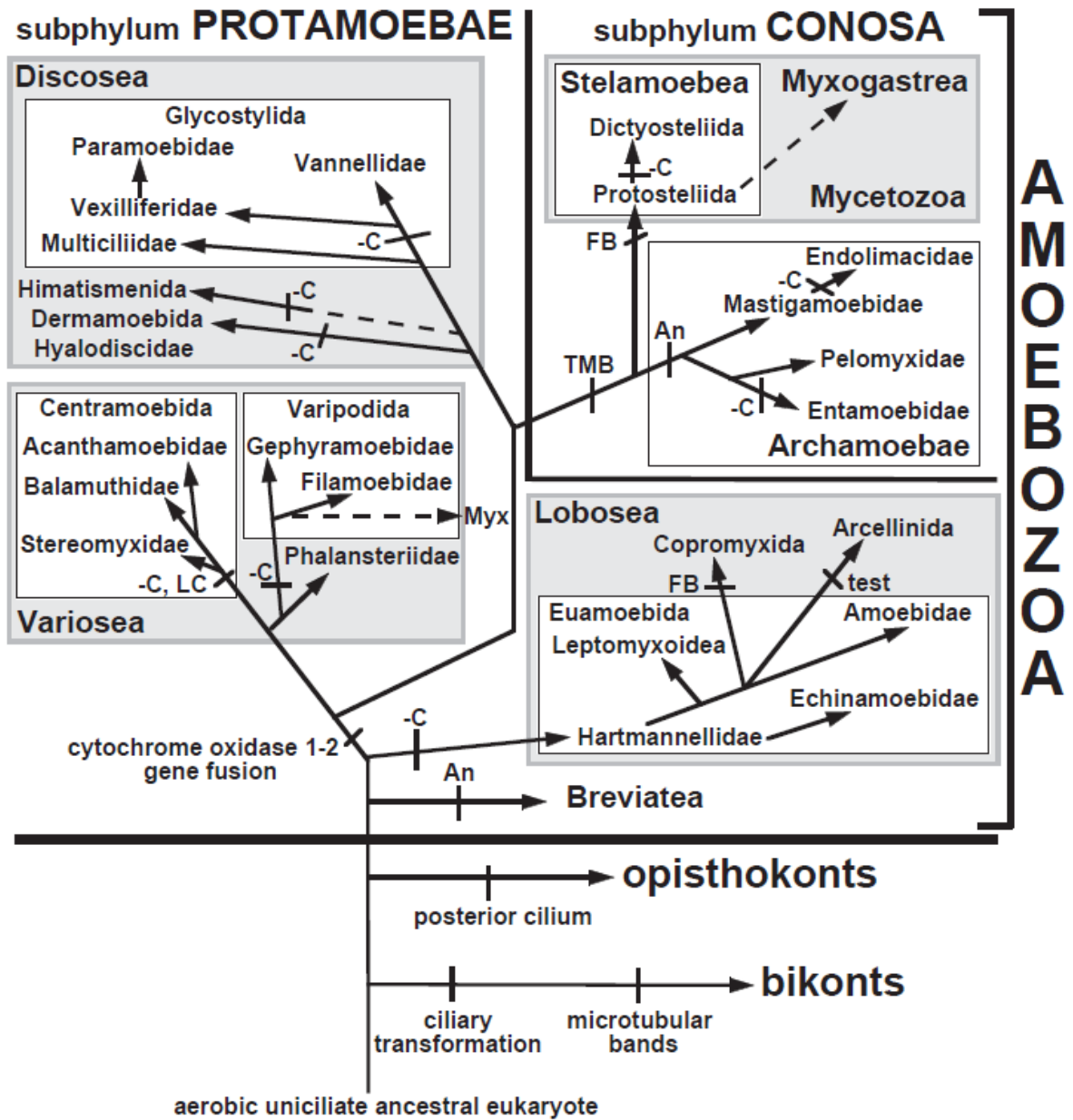


Figure 9: Hypothetical reconstruction of major features of amoebozoan evolution.

The topology shown involves 9 ciliary losses (-C); a different ranking order within Discosea could allow fewer losses. Mitochondria were converted into mitosomes or lost in two separate secondarily anaerobic lineages (An). LC = lamellate centrosome. FB = fruiting body. TMB = transverse microtubular band. Because of the uncertainty in the position of myxogastriids (Myx) alternatives are shown with dashed lines. (from Cavalier-Smith *et al.* 2004) More recent studies modified the branching and phylogeny of this reconstruction (explained in the text)

1.2.3.2. *Legionella*-like amoebal pathogens

Amoebae are also vital for the maintenance in the environment of *Legionella*-related bacteria called LLAPs for *Legionella*-like amoebal pathogens. These gram-negative bacilli are non-culturable bacteria, but they infect and replicate inside amoebae in the same way that *Legionella* do¹²⁰. They have historically been called LLAP because of their replicative strategy and their close relatedness to *Legionella*. However, they do not replicate on the standard laboratory medium for the cultivation of *Legionella* (BCYE)¹²¹. 14 LLAPs have been identified to date by 16S rRNA-sequencing. Isolation and culturing have been successful for 9 LLAPs strains and were thus given a species name (i.e., *L. drozanskii* str. LLAP-1, *L. rowbothamii* str. LLAP-6, *L. lytica* str. LLAP-3,7,9, *L. fallonii* str. LLAP-10 and *L. drancourtii* str. LLAP-11 and 12)¹²²¹²³. The phylogenetical analysis of the LLAP using 16S RNA sequences showed that they do not form a distinctive clade in the *Legionella* genus, but are rather branching all across the tree. However, the internal resolution of the tree is uncertain, with different topologies depending on the technique used to produce the tree (**figure 10**)⁶⁷¹²⁴.

Using 129 16S rRNA sequences of approximately 440bp long (71 sequences already published and 58 from *de novo* sequencing) and maximum likelihood to reconstruct the phylogeny, Goma and collaborators created a tree in which LLAPs are grouped within 4 different clusters (figure 6A). *L. drozanskii* is grouped closest to *L. nautarum*; *L. fallonii*, *L. lytica*, *L. rowbothamii* and LLAP-2 are within the same cluster with *L. fallonii* being closest to *L. moravica* and the other three being closest to *L. waltersii*; *L. drancourtii* and LLAP-4 are more closely related to *L. pneumophila* and finally LLAP-8 is closest to *L. tucsonensis*. In an earlier study, Park and collaborators used 53 16sRNA sequences to align unambiguously 1303 nucleotides and created a tree with the neighbour-joining method. In their tree, LLAPs formed a strong cluster of 9 sequences (LLAPs 9, 7, 3, 2, 6, 11, 12, 4 and *L. lytica*) with very few nucleotide substitutions between the leaves and a relatively low bootstrap support (41%) between branches of this cluster and the closest sequence (LLAP-10), and 19% with in turn the nearest node to the cluster at the top of the tree. LLAP-14 is closest to *L. shakespearei*, LLAP-8 is branched outside of the cluster containing *L. tucsonensis* and finally LLAP-1 is closest to *L. feeleii*. The differences between these two trees can come from a variety of factors. Among those, the choice of building intragenic trees using only the 16S RNA sequences can be questionable. Indeed, even if 16S RNA sequence-based phylogeny is the gold standard for the analysis of bacterial communities, it encounters some limitations at the genus level due to the lack of variability.

PCR amplicons used for sequencing classically target just one of the hypervariable region of the 16S RNA ^{125 126}. The resulting tree reconstruction can lead to a poorly resolved internal topology. Although multigene alignments can seem more appropriate in cases when a high resolution is needed, using 16S RNA sequences to build a phylogeny has become fast and affordable for almost every laboratory. It is also a good starting point when working with bacteria with no available genome.

The choice of the method of tree reconstruction can also probably explain the differences between the two trees ^{127 128}. Indeed, the Maximum likelihood (ML) method has been proven to be more powerful in reconstructing an accurate tree topology than the Neighbour-joining (NJ) method ^{129 130}. Ogden and Rosenberg ¹³⁰ even found ML to be more accurate than NJ in every simulations they performed on a defined dataset.

Interestingly, the study of Gomaa and collaborators also described for the first time *Legionella*-like bacteria associated with testate amoeba. They tested the presence of what they called LLAB (*Legionella*-like *Arcella*-associated bacteria) in members of the *Arcella* genus. It is not yet clear which type of relation these bacteria establish with their hosts but based on the authors' 16S RNA analysis, the LLABs are deeply branched in the *Legionella* tree (**figure 10A**).

Since their first description in 1991 ¹³¹, LLAPs remain largely unstudied and their implication in pneumonia cases as the main cause or acting as a co-pathogen seem to be underestimated ¹³².

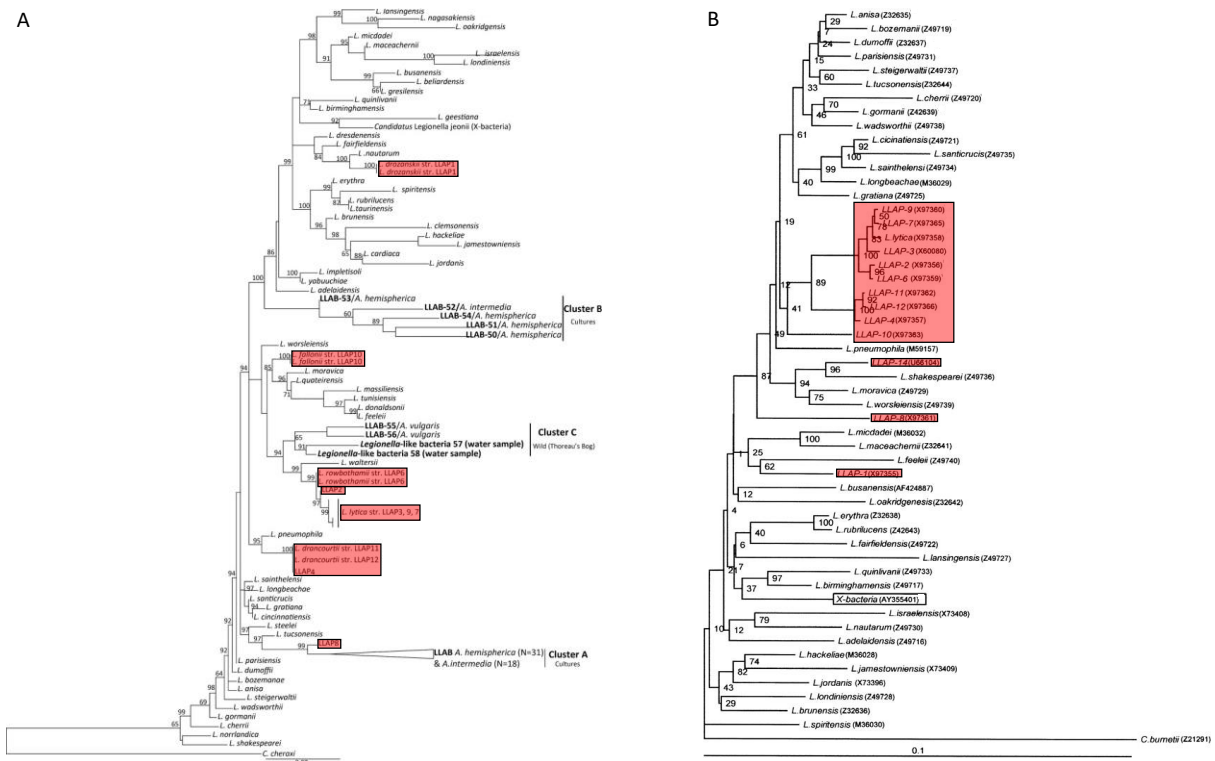


Figure 10: Phylogenetic reconstructions of the genus *Legionella* with *Legionella*-like amoebal pathogens (LLAPs) based on 16S RNA sequences.

LLAPs are outlined within red boxes. Numbers at the nodes represent bootstrap support. A: Maximum likelihood phylogeny illustrating the relationships among *Legionella*-like *Arcella*-associated bacteria (LLAB – bold taxa), *Legionella*-like amoebal pathogens and other members of the *Legionella* genus. *Coxiella cheraxi* was used as an outgroup. scale bar = 0.07 substitutions/site. B: Phylogenetic relationships between *Legionella* species. *Coxiella burnetii* was used as an outgroup. Treeing was achieved using the Neighbour-joining method. The scale bar represents 10 substitutions per 100 nucleotides. (Adapted from Gomaa *et al.* 2018 and Park *et al.* 2004)

1.2.3.3. *Legionella* within biofilms

L. pneumophila are fastidious to grow in laboratory settings and require the addition of L-cysteine, iron pyrophosphate, activated charcoal and ACES buffer in addition to an extremely pure water. However, we previously said that *Legionella pneumophila* in the environment can be in a planktonic form or associated with biofilms. The sessile way of life presents a lot of advantages for the bacteria in aquatic settings, which can be extremely complex and changing. *L. pneumophila* have been shown to produce monospecies biofilm *in vitro*¹³³ but the more (and maybe only) naturally-occurring type of biofilm in which *L. pneumophila* take part are multispecies biofilms¹³⁴.

Biofilms are complex microbial communities that are irreversibly attached to a surface and enclosed in an extracellular matrix composed mainly of polysaccharides associated with DNA, proteins, lipids and ions¹³⁵. This is an ecological niche where bacteria are protected from

outside aggressions thanks to the extracellular matrix that impedes the penetration of antibiotics and protects against desiccation thanks to its very high hydration¹³⁶. Interestingly, multispecies biofilms can act as evolutionary ‘boosts’ as the close proximity of bacteria from different species or genera and the presence of extracellular DNA seem to promote horizontal gene transfer¹³⁷.

On **figure 11** is depicted the canonical development pathway of bacterial biofilms. First, bacteria (usually motile) initially attach to a substratum (step 1). This first transient attachment is then followed by the irreversible adhesion of the bacteria to the surface and the beginning of the production of the extracellular matrix (step 2). Then the biofilm grows into a three-dimensional, usually mushroom-like structure (step 3) and matures (step 4). Finally, when environmental condition become unfavourable, bacteria detach from the biofilm and swim away to find a better settlement (step 5)¹³⁸.

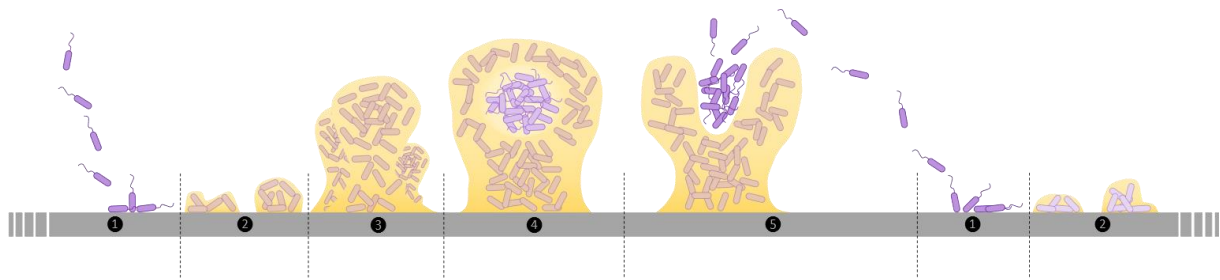


Figure 11: Bacterial biofilm development.

All steps are discussed in the text. 1: Initial attachment; 2: Irreversible adhesion; 3: Growth; 4: Maturation; 5: Dispersion. The yellow cap represents the extracellular matrix.

L. pneumophila is able to colonize existing biofilms¹³⁹ with other bacteria promoting its persistence¹⁴⁰ while others tend to be detrimental for its survival in this environment¹⁴¹. However, little is known of the bacteria-encoded factors promoting its attachment to existing biofilms¹⁴². As for other numerous organisms, the bacterial second messenger cyclic di-GMP seems to play a role in *L. pneumophila* in promoting the biofilm formation by increasing its global or local levels in the bacteria^{143–147}.

What is more interesting in the context of this study is the possible association of *L. pneumophila* with amoebae in biofilms.

Interestingly, it has been shown that the number of *L. pneumophila* cells in a faucet biofilm was directly linked to the number of amoebae cells ¹⁴⁸. Another study showed that the association of *L. pneumophila* with *A. castellanii* in biofilms can have a great beneficial impact on the survival of the bacteria to heat and chlorine based treatments ¹⁴⁹, two of the most used procedures to treat drinking water systems ¹⁵⁰.

However, the nature and characteristics of the relationship between *Legionella* and amoebae in water biofilm are still largely unknown.

A study by Declerck and collaborators aimed at a more comprehensive characterisation of the relation between *A. castellanii* and *L. pneumophila* in water distribution biofilms. As such investigations are challenging *in situ*, they reproduced a flow of water similar to that in a section of a water distribution system with a rotating reactor and tested the fate of both the amoebae and *L. pneumophila* in biofilms made by 4 other bacterial species (*Aeromonas hydrophila*, *Escherichia coli*, *Flavobacterium breve* and *Pseudomonas aeruginosa*) which are relevant to this study as they can be found in natural biofilms where they are known to influence the presence in *L. pneumophila* ¹⁵¹. They showed that the number of *Legionella* cells highly increased nearly 3 log units as soon as 48h after the addition of *A. castellanii* to the biofilm. Furthermore, their experimental setup supported the results obtained by Murga and collaborators stating that *L. pneumophila* might only be surviving in biofilms in the absence of protozoan hosts and not actively replicating ¹⁵². As can be expected from the current knowledge of the relationship between free-living protozoa and planktonic *L. pneumophila*, the association of *A. castellanii* and *L. pneumophila* in biofilm leads to an intense replication of the bacteria. An interesting hypothesis was formulated by Shaheen and collaborators concerning the trophic habits of amoebae in biofilms ¹⁵³. Grazing preferences of amoebae for non-*legionella* species in mixt biofilms could lead to the increase in amoebae population and subsequent relative enrichment in *L. pneumophila* of the biofilm. Then, a large number of amoebae are forced to feed on *L. pneumophila*, leading eventually to the release of significant quantities of bacteria in the environment. ¹⁵³ Their last founding concerns the persistence of *L. pneumophila* in water systems. They showed that 2 years after initiating their experiment, the total count of viable cells was reduced by 2 log units at room temperature, but it increased when the temperature reached 40°C. Again, these findings are coherent with results obtained by Declerck and collaborators ¹⁵¹ and Murga and collaborators ¹⁵². Upon starvation, chlorinated or heat treatment, *Legionella* cells in the environment can enter the transient

viable but non culturable (VBNC) state, in which they are not actively replicating and are not able to grow on standard laboratory media ¹⁵⁴. Although it has been shown that serial passages in amoebae were sufficient to render the bacteria culturable again ¹⁵⁵, VBNC *Legionella* cells' importance in Legionnaire's disease aetiology has long been overlooked. Two recent studies from members of the university of Vienna (Austria) took interest in the infectivity of VBNC *Legionella*. Their results are fascinating, they first showed that *L. pneumophila* in a VBNC state induced by starvation in ultrapure water remain virulent towards *A. castellanii*, THP-1 human macrophages and monocyte-derived primary human macrophages for up to a year ¹⁵⁶. Then, they induced the VBNC state with a heat treatment, to have results more relevant to what happens in water distribution networks and found that after a 60°C treatment, *L. pneumophila* was able to infect *A. castellanii* and THP-1 cells for up to 85 days ¹⁵⁷. Although we still don't know if VBNC *L. pneumophila* can cause disease *in vivo*, these results raise the question of the adequacy between the actual population of *L. pneumophila* in water pipes and the detection and treatment methods.

The presence of *L. pneumophila* within biofilms in these environments poses a real threat by ensuring the dissemination of the bacteria even with the use of classical treatment procedures.

1.2.3.4. *Legionella* and animals

The last example of environmental setting for *L. pneumophila* is rather peculiar, as it involves the relationship of the bacteria with an animal host. Nematodes have been estimated to represent a total community of approximately 4.4×10^{20} individuals in surface soils across the world ¹⁵⁸. They are the dominant metazoan taxon ¹⁵⁹ in soils and play a crucial role in regulating its structure and functions ¹⁶⁰. Among them is the *Caenorhabditis elegans* roundworm. It is best known for its contribution to the fields of cell biology, genetics and physiology as a study model for all sorts of biological processes (e.g. cellular signalling ¹⁶¹, obesity and aging ¹⁶², cell death and stress response ¹⁶³, autism ¹⁶⁴, Parkinson's disease ¹⁶⁵, etc.). There even is a growing use of *C. elegans* as model to study host-virus interaction ¹⁶⁶. As a bacteriovorous soil worm, *C. elegans* is permanently in contact with bacteria. Their interaction mainly leads to two distinct outcomes: predation of the bacteria by the nematode or infection of the nematode by the bacteria (with either host resistance or host death) ¹⁶⁷. Already used for decades as a host model for intracellular pathogens such as *Shigella flexneri*

¹⁶⁸, *Salmonella enterica* ¹⁶⁹, *Listeria monocytogenes* ¹⁷⁰ and *Cryptococcus neoformans* ¹⁷¹, *C. elegans* has also been cited as an environmental reservoir for foodborne pathogens ¹⁷². Brassinga and collaborators documented the first association of *Legionella* (sp. *pneumophila* and sp. *longbeachae*) and *C. elegans* in laboratory assays and artificial soil environments ¹⁷³. The bacteria are able to colonize and persist inside the nematode digestive tract without shortening much its lifespan and without triggering a pathogen avoidance behaviour from the worm. Furthermore, they show that the tolerance of the nematode to *L. pneumophila* is dependent on the host immune response and the bacteria Dot/Icm type IV secretion system. Finally, their most important finding is that after colonization of the nematode, *Legionella* differentiate into highly infectious and resistant MIFs, that are then excreted by the worm in the environment. This suggests that *C. elegans* can act as a natural reservoir for *Legionella* in soils and as a potentiator for *Legionella* infections. As *C. elegans* belongs to the Metazoa kingdom, they can act as a missing link to explain the adaptation of *L. pneumophila* from Protozoa to Human infections.

1.2.4. The biphasic lifestyle of *Legionella*

1.2.4.1. *Legionella* as a facultative intracellular pathogen

The biphasic lifestyle of *L. pneumophila* consists in the alternance between a replicative (inside its host) and a transmissive phase (in its environment, either the outside environment or the human lung). To initiate its replication cycle, the bacteria will first be engulfed by endocytosis (phagocytosis or macropinocytosis) by its host-cell. Then, the bacteria will secrete numerous effectors in a fine-tuned fashion through the type IVb Dot/Icm secretion system (T4BSS) in the infected cell cytosol ¹⁷⁴ to hijack the host-cell machinery, avoid degradation and allow replication. First, the bacteria will escape the endosomal degradation pathway by initially preventing the fusion of the bacteria-containing endosome with lysosomes. Then, thanks again to the secreted effectors, they rewire the cell trafficking and allow the establishment of a replicative vacuole, called the LCV for *Legionella*-containing vacuole, by recruiting various organelles to the membrane of the endosome. The bacteria will next extensively replicate before switching to its transmissive phase and being released of the LCV and eventually into the environment.

1.2.4.2. Control of the biphasic switch

The transition between the two lifestyles of *L. pneumophila*, replicative and transmissive (virulent), is finely regulated by various environmental and metabolic signals ¹⁷⁵ (figure 12). One of the metabolic trigger of the switch from replicative to transmissive phase is the nutrient depletion of the host cell once bacteria has extensively replicated inside the LCV ¹⁷⁶. The low concentration of both amino acids and fatty acids inside the host cell sends a signal to two ppGpp (guanosine 3'-diphosphate 5'-diphosphate) synthases, RelA and SpoT (respectively) ¹⁷⁷. The accumulation in the cell of the alarmone (p)ppGpp in turns induces the expression of the *rpoS* gene coding for the alternative sigma factor RpoS and the activation of the two-component system LetA/S.

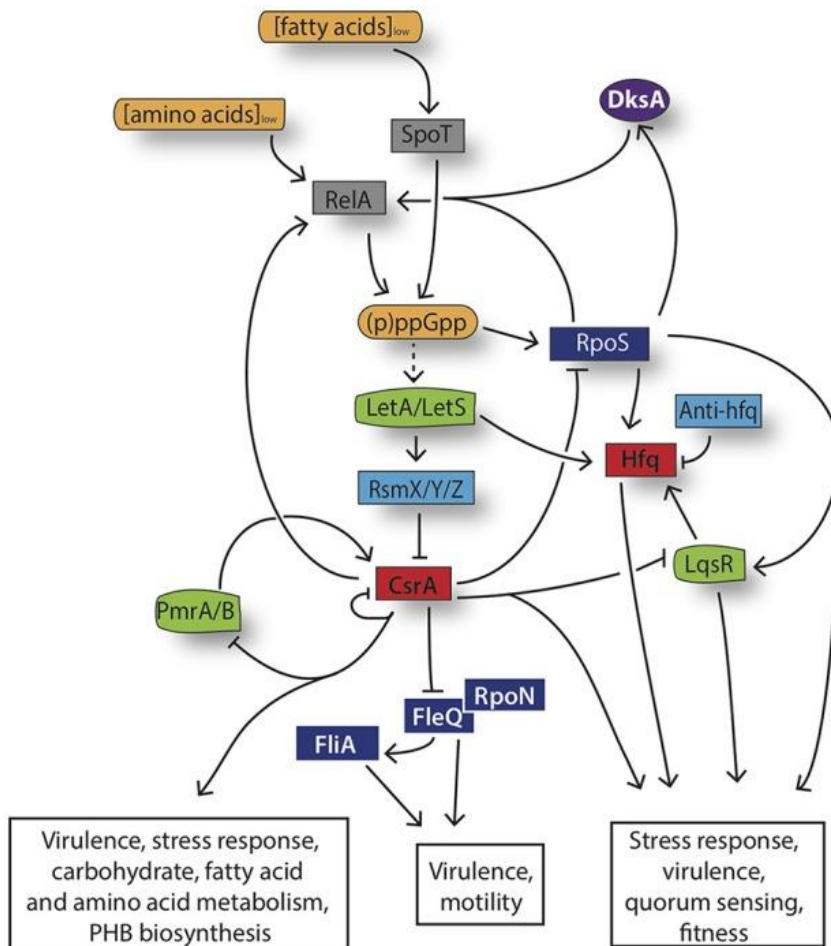


Figure 12: Control of the biphasic lifestyle of *Legionella pneumophila*

The LetA/S two-component system is formed by the LetS sensor histidine kinase that phosphorylates the LetA response regulator at the end of a four-step phosphorelay. The cues activating the LetS kinase are largely unknown¹⁷⁸. However, it has been suggested that (p)ppGpp accumulation is likely responsible for this phenomenon¹⁷⁹. Once phosphorylated, the LetA protein is able to regulate the expression of the small noncoding RNAs RsmX/Y/Z, responsible for the transcriptional derepression of the global virulence regulator protein CsrA. By sequestering CsrA, RsmX/Y/Z prevent the interaction of the protein with its target mRNAs, ultimately relieving the expression of transmissible-trait-coding genes^{180–182}.

It has been empirically proven that mutants of *letA* and *rsmYZ* are less cytotoxic, sodium resistant, non-motile (at least for the *letA* mutants) and less capable to efficiently infect both macrophages and amoebae^{183,184}.

L. pneumophila's capacity to replicate inside such evolutionary distant hosts as amoebae and human macrophages, demonstrates its amazing adaptability and is key to understand the evolutionary emergence of pathogenesis in bacteria. As discussed, this extraordinary feature of the bacteria is largely facilitated by the T4bSS.

1.2.5. The Type IVb Dot/Icm secretion system of *Legionella*

To understand the extraordinary capacity of *L. pneumophila* to subvert the entire cellular process of pathogen degradation by specialised (amoebae and macrophages) and non-specialised (pneumocytes, HeLa) cells, we must first further characterise the type IVb Dot/Icm secretion system (T4bSS).

Present in all described *Legionella* species¹⁸⁵, the T4bSS is the key element of the virulence in *L. pneumophila* and is used by the bacteria to inject more than 330 effectors in the infected cell¹⁸⁶ (**figure 13**). It has been shown that its presence (especially at the pole of the bacteria¹⁸⁷) is paramount for the proper biogenesis of the LCV as well as all steps critical for the intracellular replication of the bacteria^{188,189}. Interestingly, the studies by Marra and collaborators in 1992 and Berger and Isberg in 1993 both described the same genetic locus that restored the pathogenic capacity of *L. pneumophila*. Both groups named this locus according to the defect their absence induced to cellular processes. Marra and collaborators named this locus 'icm' for intracellular multiplication and Berger and Isberg, 'dot' for defect in organelle trafficking.

The T4bSS is a complex assembly of proteins coded by 27 genes on the *dot/icm* locus ¹⁹⁰.

The function of the proteins coded by these genes is to date not completely understood, but they can however be placed in 6 categories:

- Coupling protein complex LvgA/IcmS/IcmW and DotL/DotM/DotN

IcmS/LvgA and IcmS/IcmW form heterodimers described as adapters for the T4bSS and are thought to help present substrates to the Dot/Icm machinery through interactions with the ATPase activity complex DotL/DotM/DotN ¹⁹¹⁻¹⁹³. Together, they form what is called a 'coupling protein complex' (figure 8).

- Essential cytosolic components IcmR and IcmQ

The IcmQ protein inserts into the inner membrane bilayer, forming a pore potentially allowing substrates of the T4bSS to be translocated outside the bacterial cell. This protein can be associated with the IcmR chaperone that inhibits both the membrane insertion and the pore-forming activity of IcmQ ¹⁹⁴.

- Core transmembrane complex DotC/DotD/DotF/DotG/DotH

These 5 proteins assemble to shape the central node of the secretion complex. Through a strong interaction, they act as a link between the outer and inner membranes. The complex is formed of two pairs of protein dimers anchored to the inner membrane: DotF and DotG, associated with DotC, DotD and DotH, with their own anchoring to the outer membrane ¹⁹⁵.

- Inner membrane accessory factors DotU and IcmF

As other members of the T4bSS, the presence of DotU and IcmF are crucial to the proper translocation of effectors ¹⁹⁶. Recently, their role has been clarified and it seems they act as some sort of beacon to recruit the core transmembrane complex proteins to the pole of the bacteria. A *dotU/icmF* mutant still produces the DotC,D,F,G and H proteins, but rather than assembling to form the classical complex, they remain in a cytosolic form ¹⁹⁷.

- Cytoplasmic ATPase complex DotO/DotB

A recent study by Chetrit and collaborators focused on resolving the structure of the Dot/Icm secretion system by cryo-ET experiment. They also characterised the association of the DotO/DotB energy complex to the T4bSS. The DotB protein is described as a dynamic ATPase present in the cytosol of the bacteria and capable of being recruited to its pole by the DotO ATPase. Once docked at the base of the inner membrane complex of the T4bSS, the DotO protein recruits the ATP-bound DotB to form two hexamer rings stacked one on top of the other. This energy complex creates a cytoplasmic channel through which the secreted effector proteins would transit ¹⁹⁸.

- Components of unknown functions DotK, DotI, DotJ, DotE, DotP, DotV, DotA, IcmX and IcmT

The functions of some of the proteins associated with the T4bSS still remain unknown to this date. However, their absence is generally linked with impaired intra-host activities for *L. pneumophila*. DotK is a lipoprotein associated with the outer membrane. It may help the anchoring of the Dot/Icm machine to the peptidoglycan layer thanks to its OmpA family domain (which is known to be conserved in peptidoglycan-associated bacterial proteins) ¹⁹⁹. DotI and its partial paralog DotJ form an heterocomplex in the inner membrane and seem to poorly associate with the core complex ²⁰⁰. DotI is the homolog of the T4aSS VirB8 protein ²⁰⁰. In *A. tumefaciens*, VirB8 might have a crucial role in the assembly of the secretion machinery ²⁰¹. The DotE/IcmC and DotP/IcmD inner membrane proteins are required for macrophage killing ²⁰². DotE/IcmC also has a role in DNA conjugation ²⁰³. DotV is closely related to DotE ¹⁹⁹. The exact role of the DotA protein remains unknown, despite $\Delta dotA$ mutant bacteria being largely used as a non-virulent control in infection studies. Strikingly, DotA is secreted into in culture supernatant in a T4bSS-dependant fashion ²⁰⁴. IcmV is an inner membrane protein whose expression can be induced early following exposure to antimicrobial copper oxide nanoparticles ²⁰⁵. IcmX is a periplasmic protein, essential for the biogenesis of the LCV ²⁰⁶. Finally, the role of the C-terminus part of the IcmT protein is pivotal in creating membrane pores allowing bacteria to egress from macrophages ²⁰⁷. In addition to its implication in pore-mediated egress, IcmT is essential in preventing phagosome-lysosome fusion in *A. polyphaga*²⁰⁸.

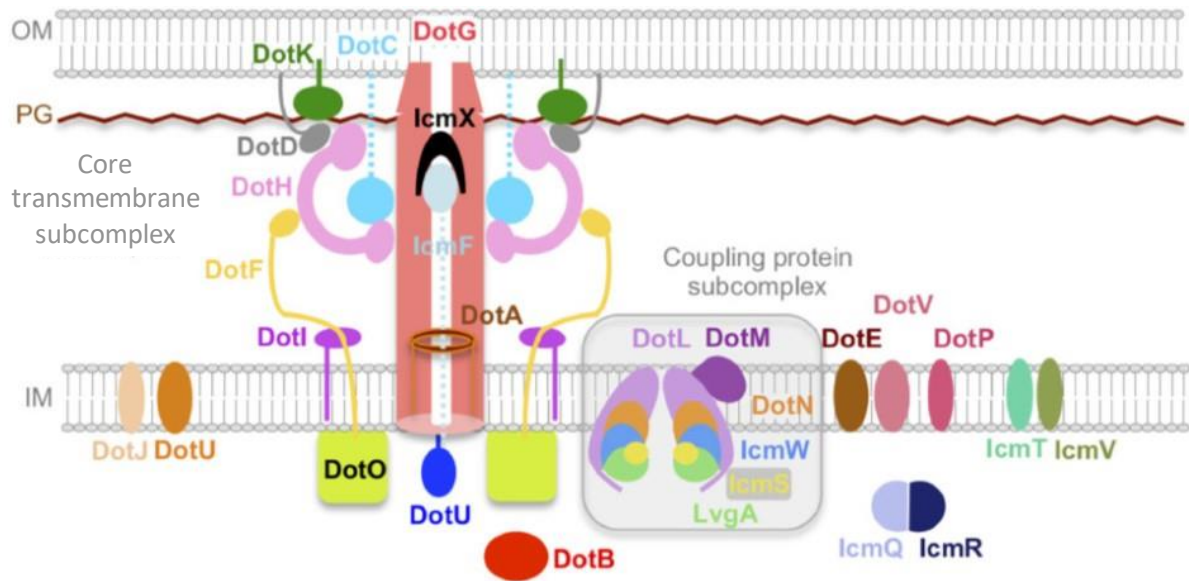


Figure 13: Type IVb Dot/Icm secretion system of *Legionella pneumophila*.

Adapted from Gomez-Valero *et al.* 2019

Legionella also possess a T4aSS called the Lvh secretion system, but contrary to the T4bSS, its absence does not impact the intracellular replication of the bacteria. Furthermore, the T4aSS seems to have been acquired through horizontal gene transfer as the genes encoding this system are clustered on a genomic island with an overall GC content higher than the chromosome's²⁰⁹. To further support the hypothesis of an acquisition after speciation of the *Legionella* bacteria, it has been shown that the Lvh secretion system is randomly distributed across the *Legionella* species tree¹⁸⁵. The archetypical T4aSS present in *Agrobacterium tumefaciens*²¹⁰, which is formed of 'only' 12 proteins (**figure 14**), is substantially less complex than the T4bSS Dot/Icm.

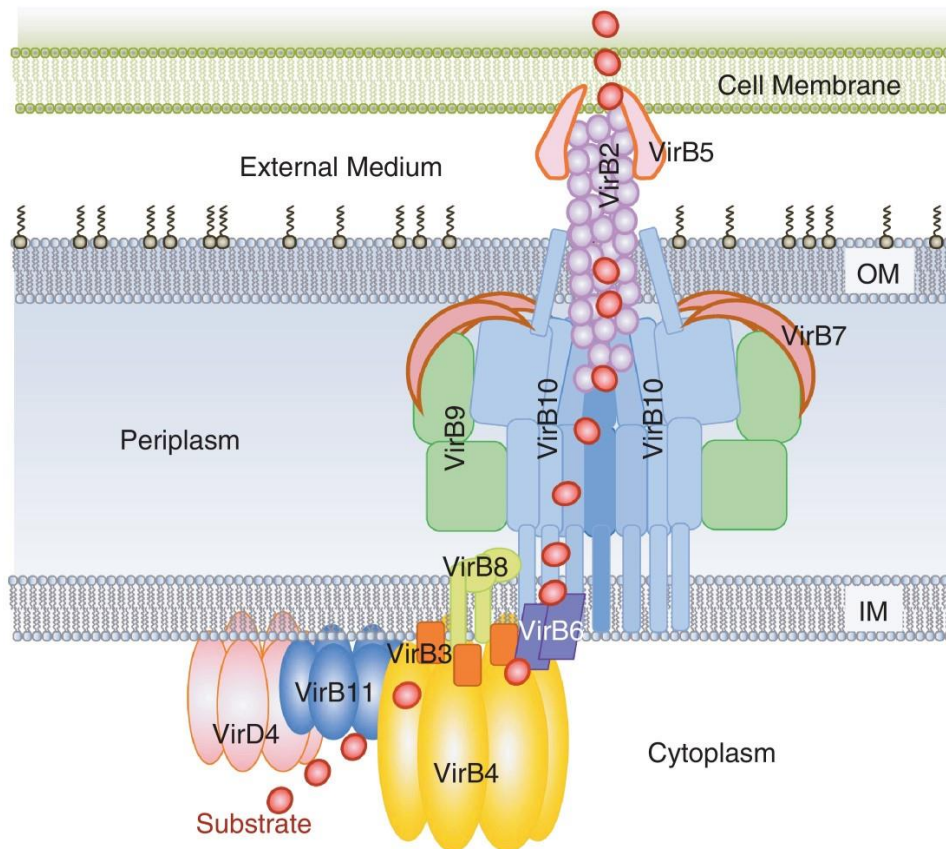


Figure 14: Type IVa secretion system.

From Waksman and Orlova, 2014

1.2.6. Effectors of the Type IVb secretion system

As previously mentioned, the type IVb Dot/Icm secretion system is essential for the virulence of *L. pneumophila*. To hijack the host-cell machinery and use it to its advantage, *L. pneumophila* uses the T4bSS to translocate more than 330 effector proteins. More than 18,000 effectors in total have been identified in 58 different species of *Legionella*²¹¹. It has been suggested that a specific subset of these effectors could be responsible for the propensity of *L. pneumophila* over other species to be isolated in patients presenting a case of LD. But shockingly, in 47 experimentally tested species for intra macrophage replication, those able to replicate (60%) do not share a common combination of effectors²¹¹. This supports the findings that although quite conserved amongst *L. pneumophila* strains, the effector repertoire is extremely variable across the species tree¹⁸⁶ and that no specific genes or functions can be directly linked to the heightened prevalence of *L. pneumophila* (**figure 15**).

A striking feature of **figure 15** concerning the effector repertoire concerns the species *Legionella longbeachae*. It is the second most frequent cause of LD worldwide and the most frequent in Australia and New Zealand ²¹². Contrary to *L. pneumophila*, *L. longbeachae* is mainly associated with moist and potting soils ²¹³ although one study found it may have been responsible for an outbreak linked with emanations from a cooling tower ²¹⁴. While being only 30% similar ¹⁸⁶, the effector repertoire of *L. longbeachae* seems to be closest to *L. pneumophila*'s (at least concerning the 50 analysed effectors in the study from Mondino and collaborator). Furthermore, when applying a cutoff of 30% identity, some effectors appear to be present practically only in those two species, such as RavZ (also present in *L. santicrucis* and *L. sainthelensi*) or SdeB (present also in *L. gratiana*). Interestingly, although closely related to *L. longbeachae* (within the same cluster), these species don't behave the same way towards human cells. *L. gratiana* and *L. santicrucis* are not known to cause infection in human. *L. gratiana* appears also not to be able to effectively infect amoebae or macrophages. However, this isn't clear concerning *L. santicrucis* with one study stating that the bacterial load decreases after infection of U937 cells – even dropping below the detection threshold after 48h ²¹⁵ – and an earlier one finding that $10^{3.47}$ legionella cells were necessary to reach an ID₅₀ towards U937 cells at 72h, meaning that there are able to infect those cells, albeit poorly ²¹⁶. *L. sainthelensi* however, effectively infects U937 cells but not THP-1 cells. This exemplifies yet more the difficulty of pointing out a set of effectors directly responsible for the increased prevalence of *L. pneumophila*-related cases of Legionnaires disease.

1.2.6.1. Generic features of the effectors

To be addressed to the T4bSS, the vast majority of proteins harbour a signal sequence on the carboxy terminal end of their sequence. It consists at the minimum in a hydrophobic residue at the -3rd position and a region of 6-8 residues enriched in glutamate called the E-Block region (**figure 16**) ²¹⁷.

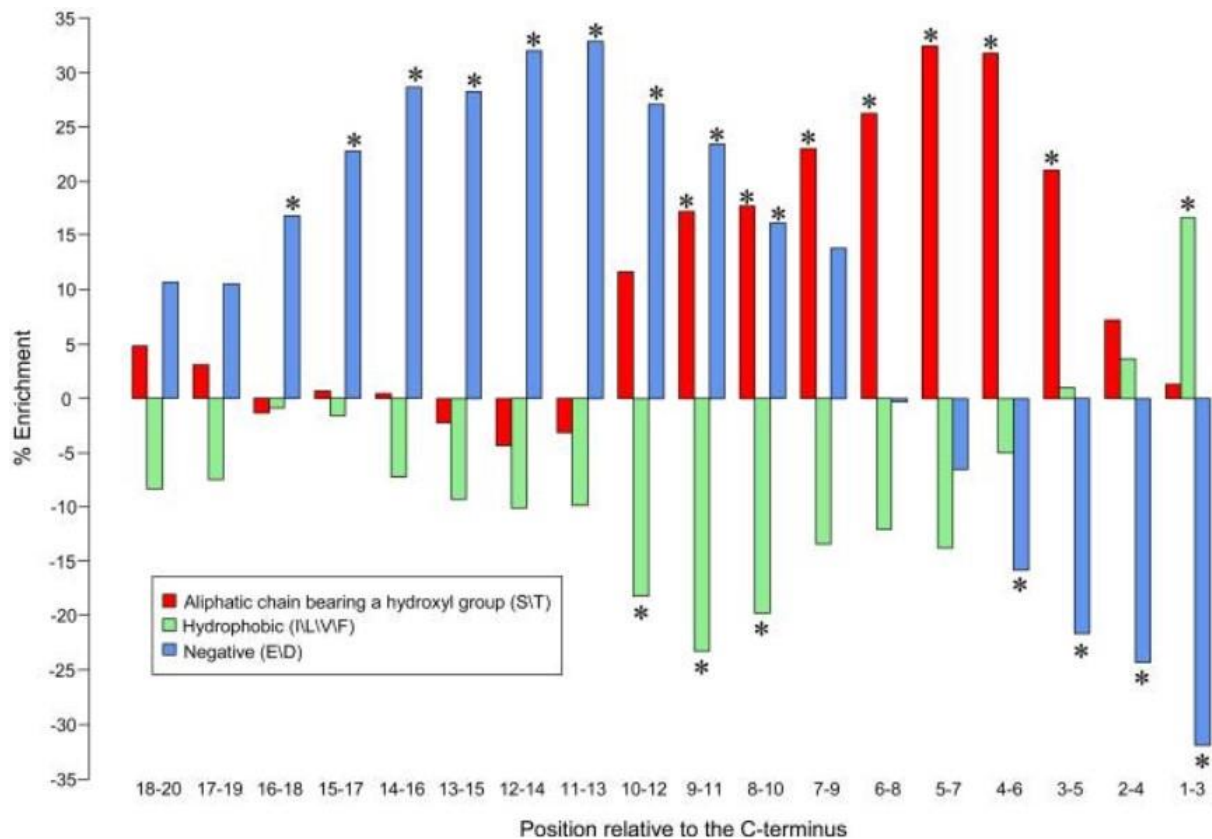


Figure 16: Composition of the carboxy terminal end of effectors.

The enrichment and depletion pattern of groups of amino acids within the 20 c-terminal residues of 145 validated effectors is shown. Amino acids with aliphatic side-chains bearing a hydroxyl group (Ser, Thr) are in red, hydrophobic amino acids (Ile, Leu, Val, Phe) in green; and negatively charged amino acids (Glu, Asp) in blue. Statistically significant enrichments or depletions (G -test; p -value < 0.01 after Bonferroni correction) are marked with asterisks. From Burstein *et al.* 2009²¹⁸.

In addition to the Carboxy terminal signal which molecular recognition by the T4bSS remains to be clarified, some effectors harbour an internal sequence designed to facilitate the association to the protein coupling complex of the Dot/Icm secretion system (**figure 13**) *via* the IcmS/IcmW proteins²¹⁰.

1.2.6.2. The eukaryotic-like proteins and eukaryotic domain-containing proteins

L. pneumophila's effector repertoire stands out by its record size across all bacteria. But the number of these proteins is not the only peculiar feature to be noted. First on this list is the relatively high number of eukaryotic-like proteins and eukaryotic-like domains. It is suggested that up to 3% of the proteome is made of these proteins²¹⁹. Throughout their history of co-evolution with their eukaryotic hosts, bacteria from the *Legionella* genus and more importantly *L. pneumophila* underwent extensive events of horizontal gene transfers

(HGT) ²²⁰. These events are thought to have occurred not only between *L. pneumophila* and their hosts but also in a tripartite relation with the amoeba giant virus *A. polyphaga* Mimivirus (APMV) which contains a patchwork of both amoebal and bacterial genes ²²¹. This suggests that amoebae can participate in enriching the *Legionella* genus gene pool both by transferring its own as well genes obtained from the genome of other organisms. Interestingly, the APMV has recently been reported as the first virus to encode a Rab GTPase protein ²²². Together with Ras, Rho, Ran and Arf proteins, Rab GTPase proteins form one of the five major branches of a family of small guanosine triphosphatases, called the Ras superfamily. Mostly found in humans with more than 150 members, they are conserved throughout the eukaryotic tree with orthologs identified in *Drosophila*, *C. elegans*, *S. cerevisiae*, *S. pombe*, *Dictyostelium* and plants ²²³. These proteins act as molecular switches for the fine regulation of a plethora of cellular processes. They are present in two forms, either active i.e., GTP-bound or inactive, i.e., GDP-bound. The cycling between the two forms is facilitated by two types of regulators: guanine-nucleotide-exchange factors and GTPase-activating proteins that promote the active and inactive forms of the GTPase proteins, respectively ²²⁴. Strikingly, the presence of Ras superfamily proteins in the genome of bacteria seems to be restricted to the *Legionella* genus. In 2010, Cazalet and collaborators identified three Ras-related GTPases in the genome of *L. longbeachae*, making it the first occurrences of such proteins in prokaryotes ²²⁵. Later studies discovered that the *Legionella* species actually encode 184 small GTPases, 104 of which being characterised as belonging to the Ras superfamily ²¹¹. This exemplifies the importance of HGT between *Legionella* and eukaryotes in allowing the bacteria to subvert many cellular pathways when infecting its various hosts. Proteins of eukaryotic origin exhibit a high rate of conservation at the species level and may participate in the hijacking of host cellular processes, suggesting that they play a pivotal role in the virulence of *L. pneumophila*¹². Figure 12 displays a summary of genomic characteristics of 102 identified eukaryotic-like proteins – out of more than 250 ²¹¹ – in 3 strains of *L. pneumophila*. Interestingly, although extremely conserved at the species level (**figure 17 A**) – 73 out of 91, 90 and 83 proteins in strain Paris, Philadelphia and Lens, respectively, are common to all three strains – they are far more scarcely found in other species of the *Legionella* genus, suggesting a strong evolutionary pressure on the corresponding genes. Of the 102 genes, 1 is from archaeal origin, 36 are from other bacteria, 31 from eukaryotes, 23 are only found in *Legionella*. Authors of the study calculated the d_N/d_S ratio of the eukaryotic-like genes – depending on their putative origins –

to identify the selection forces governing their evolution in the genome of *L. pneumophila*, thus estimating the balance between purifying (selection against non-synonymous substitutions) and positive selection (diversifying selection). The calculation was based on the ratio of the number of non-synonymous (d_N) to synonymous substitutions (d_S) at any given site, in genes encoding eukaryotic-like proteins that had at least three homologs. They used a model that did not assume a constant rate of substitutions. A d_N/d_S ratio < 1 would describe purifying selection and a d_N/d_S ratio > 1 , a positive selection. In **figure 17 B** are displayed the average ratio, modified to exclude the positions under diversifying selection.

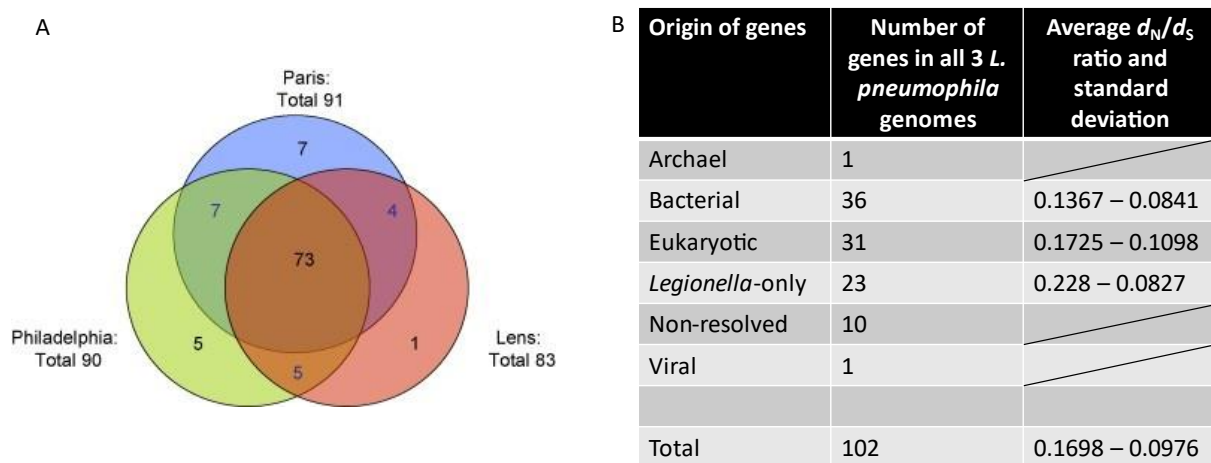


Figure 17: Summary of genes coding for eukaryotic-like proteins in the genomes of 3 *L. pneumophila* strains.

A: Venn diagram of the distribution of the genes in *L. pneumophila* strains Paris, Philadelphia and Paris. B: Origin, number and d_N/d_S ratio of genes coding for the eukaryotic-like proteins in the genome of the 3 *L. pneumophila* strains. Adapted from Lurie-Weinberger *et al.* 2010

All genes meeting the analysis criteria (80 out of the 102) were found to be under purifying selection, with values ranging from 0.0299 to 0.7897. In average, genes from bacterial origin appear to be under a stronger selection against non-synonymous substitution than eukaryotic and ultimately *Legionella*-only origins. The fact that genes only found in a particular genus or species seem to be under less strong purifying selection, has already been documented for γ -proteobacteria²²⁶. Lurie-Weinberger and collaborators formulate a rather interesting hypothesis to explain the higher ratios observed for genes from eukaryotic and *Legionella*-only origins. They analysed their dataset and found that 29% of proteins from eukaryotic and *Legionella*-only origins are at least partly composed of positions under diversifying selection

(whereas 19% for proteins from bacterial origins). Genes encoding proteins exposed to the extracellular milieu are likely to be under diversifying selection ²²⁷.

Thus, having surface proteins evolving at a fast pace would allow the bacteria to stay competitive in the evolutive ‘arms race’ between pathogens and hosts and ultimately efficiently evade immune recognition. This eventually supports the hypothesis that eukaryotic-like proteins would more likely be implicated in host-bacteria interaction ²²⁸.

The sequencing of protein-encoding eukaryotic-like genes identified various domains that are classically associated with eukaryotic proteins (**table 4**).

<i>L. pneumophila</i> strains				Motif identified
Paris	Lens	Philadelphia	Product	
Lpp2692	Lpl2564	Lpg2639	EnhC	21 sel-1 domains
Lpp1174	Lpl1180	Lpg1172	Lidl	Six sel-1 domains
Lpp1310	Lpl1307	Lpg1356		Four sel-1 domains
LpnE (Lpp2174)	LpnE (Lpl1303)	Lpg2222	LpnE	Three sel-1 domains
–	Lpl1059	Lpg1062		Seven sel-1 domains
Lpp1932	Lpl1919	Lpg1950	RalF	Sec7 domain
Lpp0267	Lpl0262	Lpg0208	LegK4	Ser/thr proteine kinase domain
Lpp2626	Lpl2481	Lpg2556	LegK3	Ser/thr proteine kinase domain
Lpp1439	Lpl1545	Lpg1483	LegK1	Ser/thr proteine kinase domain
Lpp2065	Lp2055	Lpg2214		Ankyrin repeat
Lpp0037	Lpl0038	Lpg0038	AnkQ	Ankyrin repeat
Plpp0098	–	–		Ankyrin repeat
Lpp2058	Lpl2048	–		Ankyrin repeat
Lpp0750	Lpl0732	Lpg0695	AnkX	Ankyrin repeat
Lpp2061	Lpl2051	–		Ankyrin repeat
Lpp2270	Lpl2242	Lpg2322	AnkK	Ankyrin repeat
Lpp0503	Lpl0479	Lpg0436	AnkJ	Ankyrin repeat
Lpp1905	–	–		Ankyrin repeat
Lpp1683	Lpl1682	Lpg1718	AnkI	Ankyrin repeat + SET domain
Lpp2248	Lpl2219	Lpg2300	AnkH	Ankyrin repeat
Lpp0202	–	–		Ankyrin repeat
Lpp0469	Lpl0445	Lpg0403	AnkG	Ankyrin repeat
Lpp2517	Lpl2370	Lpg2452	AnkF	Ankyrin repeat
Lpp1100	–	–		Ankyrin repeat
Lpp0126	Lpl0111	Lpg0112		Ankyrin repeat
Lpp0356	–	–		Ankyrin repeat
Lpp2522	Lpl2375	Lpg2456	AnkD	Ankyrin repeat
Lpp0547	Lpl0523	Lpg0483	AnkC	Ankyrin repeat
–	Lpl1681	–		Ankyrin repeat
–	Lpl2344	–		Ankyrin repeat
–	Lpl2058	Lpg2128		Ankyrin repeat
–	–	Lpg0402	AnkY	Ankyrin repeat
–	–	Lpg2131	LegA6	Ankyrin repeat
Lpp2082	Lpl2072	Lpg2144	AnkB	F-Box domain + ankyrin repeat
Lpp2486	–	–		F-Box domain + coiled-coil
–	–	Lpg2224	PpgA	F-Box domain
Lpp0233	Lpl0234	Lpg0171	LegU1	F-Box domain
Lpp2887	–	Lpg2830	LubX	Two U-Box domains

Table 4: List of *L. pneumophila* proteins preferentially found within eukaryotic proteins and their distribution in three sequenced species

Lpp indicate CDSs of strain Paris, Plpp indicates predicted CDS of the pLpp plasmid of the strain Paris, Lpl indicate predicted CDSs of strain Lens and Lpg indicate predicted CDSs of strain Philadelphia. Adapted from Brüggemann *et al.* 2006 ²²⁹ and Gomez Valero *et al.* 2011 ²²⁸

Those domains are ankyrin repeats, SEL1 (tetratricopeptide), Set, Sec7, serine-threonine kinase, U-box and F-box. Ankyrin repeats are some of the most widespread protein-protein interaction motifs in eukaryotes and are very scarcely found in prokaryote and virus genomes, in which occurrences are probably a result of horizontal gene transfer ²³⁰, as mentioned before. These motifs' presence in proteins is as widespread as their functions. They can act as transcriptional initiators, cell-cycle regulators, cytoskeleton remodulators, ion transporters and signal transducers ²³¹. The SEL1 domain belongs to the tetratricopeptide-like helical domain superfamily. Like ankyrin repeats, they are present in a large array of proteins with various functions such as cell cycle regulation, transcriptional regulators, mitochondrial and peroxisomal protein transport and protein folding. They play a role in protein-protein interaction and multiprotein complex assembly ²³². Set domains are found again in functionally diverse proteins, spanning the eukaryote tree. They play a role at the nucleus level with the modulation of gene activities, chromatin structure and histone methylation ²³³. Sec7 domain-containing proteins are guanine nucleotide exchange factors specific for the ADP-ribosylation factors. As such, they are important for vesicular protein trafficking and membrane remodelling processes ²³⁴. Kinases are a superfamily of enzymes whose main function is to post-translationally and reversibly add a phosphate group to certain amino acid residues of target proteins thanks to their conserved catalytic domain. The family is subdivided into two main groups with respect to the substrate specificity: the tyrosine and the serine-threonine protein kinases. The kinase activity targets very broad cellular processes, including division, proliferation, apoptosis and differentiation. Phosphorylated proteins usually undergo changes in activity, location or association with other proteins ²³⁵. U-box and F-box domains are present in proteins whose activity is linked with protein degradation through ubiquitination. U-box and F-box-containing proteins form E3 ligases that bind both adenylated (activated) ubiquitin – conjugated to a E2 enzyme – and the target protein to address it to the proteasome for degradation. F-box containing proteins form an atypical class of ubiquitin ligases that need to associate with SCFs (Skp1-cullin-F-box protein ligase) complexes before catalysing the ubiquitination of their target protein ^{236,237}.

As mentioned before, the roles of the eukaryotic-like proteins (especially through the action of their eukaryotic domains) are mainly associated with the interaction of the bacteria with its hosts, more precisely to subvert and modulate host cell functions to allow the replication of the bacteria ¹¹ (**figure 18 and 19**).

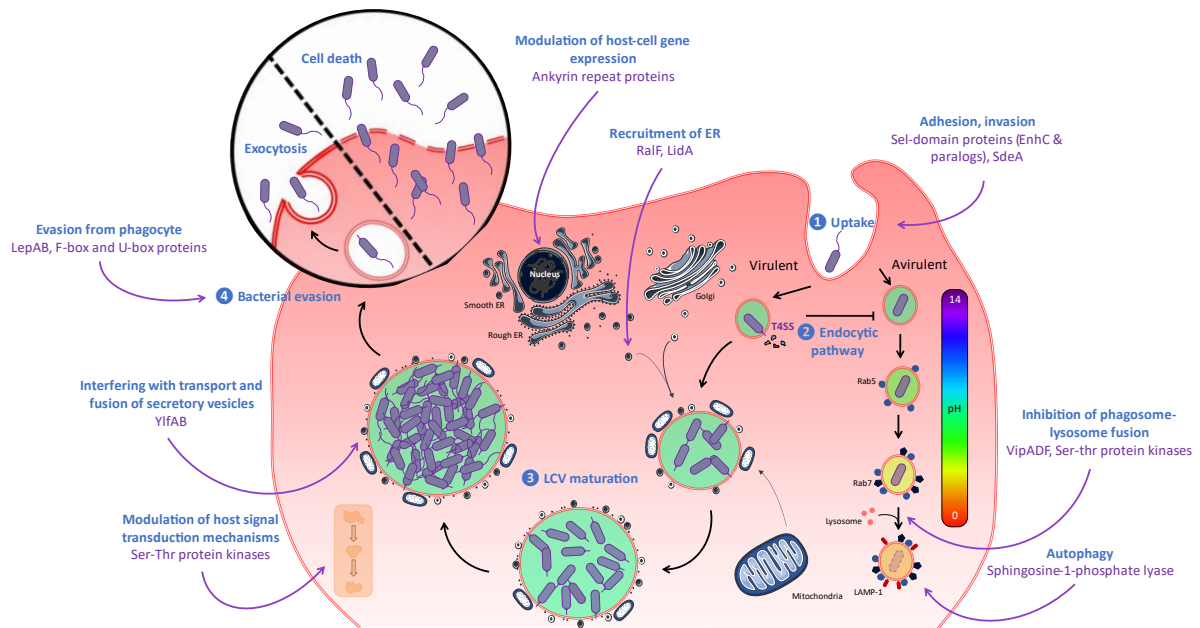


Figure 18: Intracellular infection cycle of *L. pneumophila*

On this figure are schematically represented the different steps of the infectious cycle of *L. pneumophila*, from uptake by phagocytic cells to bacterial evasion. 1: uptake of the bacteria by the host cell, 2: inhibition of the endocytic degradation pathway, 3: recruitment of organelles from the host cell to allow the maturation of the Legionella-containing vacuole (LCV), 4: bacterial evasion through host cell death or exocytosis. Purple arrows indicate steps in which Eukaryotic-like proteins are involved, and in purple are given some examples of these proteins. At the right-hand side is represented the canonical endocytic degradation pathway underwent by avirulent bacteria. All steps are further explained in the text.

1. Uptake

Initial adhesion of *L. pneumophila* is both facilitated by the host (amoebae and mammalian cells) and the bacteria. Indeed, bacterial proteins help with the engulfment of the bacteria by its future host cell. Those factors include, without being limited to, a type I secretion system coded by the *rtxA* locus and a type IV pili coded by the *pilEL* locus²³⁸. However, even if other bacterial proteins can be involved in the initial attachment to the host cell, the internalization of the bacteria itself is mediated by the host phagocytic and macropinocytic pathways¹⁸⁶.

2/3. LCV pathway

Soon after the internalization inside the host cell, the bacteria continue to deploy its effector proteins, whole or in part, to avoid being degraded by the canonical endocytic pathway. Bacteria-containing phagosomes classically undergo maturation by fusing with early, then late endosomes, leading to a strong acidification of the vacuole²³⁹. On the other hand, minutes after being internalized, *L. pneumophila* inside the LCV (for *Legionella*-containing vacuole), will use its protein arsenal to recruit vesicles derived from the host's endoplasmic reticulum. These

vesicles bring membranes and various nutrients to the lumen of the LCV, allowing the bacteria to actively replicate²⁴⁰. Recent studies (reference du labo) also have proven the capacity of *L. pneumophila* to interfere with the autophagocytic pathway, once again at its advantage, avoiding destruction by the host cell machinery.

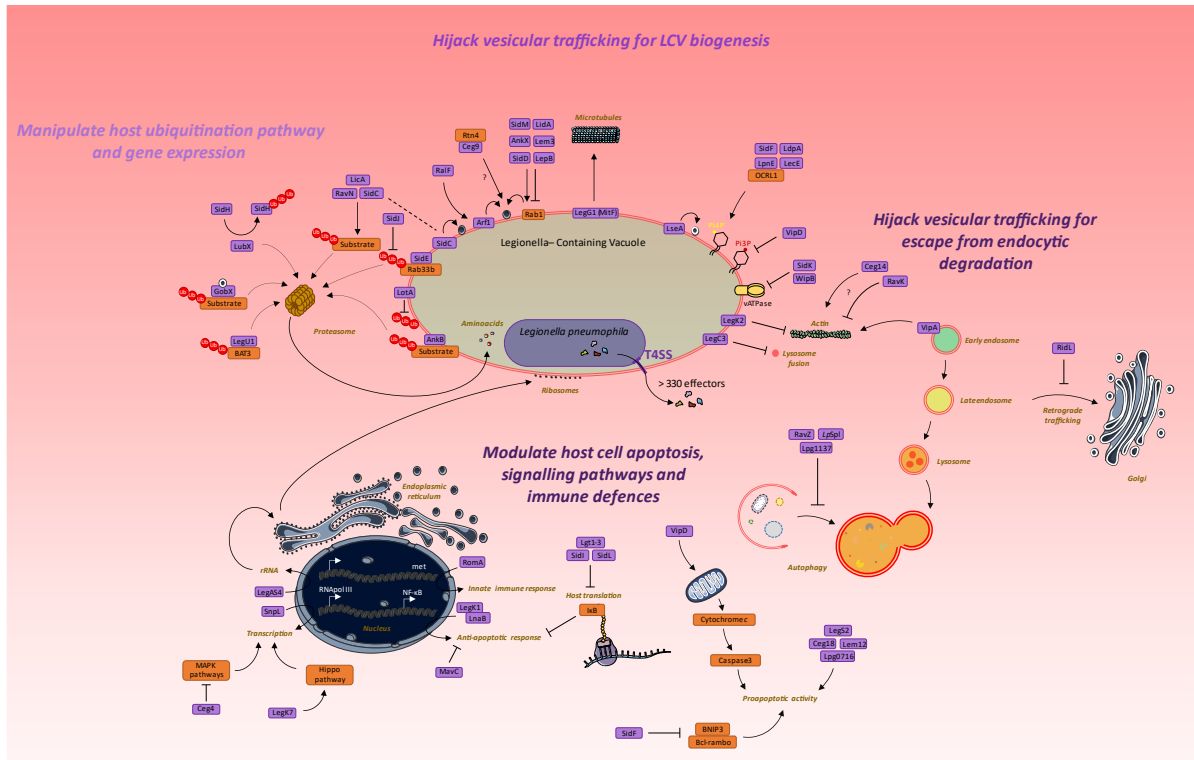


Figure 19: Summary of *L. pneumophila* effectors implicated in subversion of host-cell functions

Summary of effector proteins of *L. pneumophila* translocated in the host cell by the type 4 b secretion system Dot/Icm (purple boxes) subverting different cellular processes: host ubiquitination, gene expression, vesicular trafficking for Legionella-containing vacuole and escape from endocytic degradation, apoptosis, signalling and immune defences. Orange boxes represent host proteins or pathways.

4. Evasion

Following the intense replication of the bacteria, nutrients depletion inside the host cell triggers a metabolic switch in *L. pneumophila*, allowing it to change from replicative to transmissive phase. One of the most striking feature change for the bacteria is the reappearance of the flagella associated with the transcription of genes necessary for the expression of the virulence traits necessary for evasion from the LCV, form the host cell and for the infection of a new, fit host or survival in the extracellular environment¹⁷⁹.

2. Experimental evolution leads to the discovery of a novel mechanism of virulence control in *Legionella pneumophila*

2.1. Introduction

The Gram-negative bacteria *Legionella pneumophila* are protozoan parasites and accidental pathogens of humans. In the environment, these bacteria are ubiquitous of natural and man-made fresh-water environments where they parasitize and replicate within a diverse range of protozoan hosts¹⁻³. In humans, *L. pneumophila* can infect alveolar macrophages and cause a severe and often fatal pneumonia called Legionnaire's Disease (LD)^{4,5}. Human infections occur through inhalation of contaminated water droplets generated from various aerosol-generating devices thus making these bacteria a major public health concern and LD an emerging disease. Since its first description in 1976, over 60 *Legionella* species have been identified. The species *L. pneumophila* and especially *L. pneumophila* serogroup 1 (Lp1) is implicated in 83% of cases of LD worldwide^{6,8,87}. The remaining species are rarely isolated from humans, and some species are classified as pathogenic based on a single reported human infection⁸. Comparative genomic studies highlighted significant syntenic differences between strains that are usually associated with mobile genetic elements (plasmids, prophages, or integrative conjugative elements or pathogen islands), signifying that numerous genomic rearrangements continue to occur also after speciation²⁴¹. This strong genomic plasticity of *Legionella* genomes probably leads to the genetic and pathogenic heterogeneity within species⁹.

The key virulence factor indispensable for *L. pneumophila*'s intracellular replication within both protozoan and macrophages is the type 4 secretion system (T4SS) Dot/Icm, which translocates over 300 bacterial proteins, called effectors, into the host cell cytoplasm^{203,242}. These effectors target host proteins and subvert cellular pathways processes, in particular some of them redirect trafficking of the *L. pneumophila* phagosome to escape from the endosomal pathway and mediate its conversion into an ER-derived compartment supportive for intracellular bacterial replication^{217,243-247}. In the last decade, substantial efforts and progress have been made in understanding the cell biological processes and effector activities²⁴⁸⁻²⁵¹, however, the exact function of the vast majority remains unknown.

Remarkably, recent whole genome sequencing of about 40 *Legionella* species (*ie.* about two third of known *Legionella* species) revealed that they have a large non-overlapping effector repertoire, with only seven core effectors shared by all species studied^{9,252}. Species-specific effectors had atypically low GC content and harbor eukaryotic-like domains or resemble eukaryotic proteins, suggesting exogenous acquisition through horizontal gene transfers (HGT), possibly from their natural protozoan hosts and/or co-infecting pathogens^{111,252,253}. Another striking feature of these effectors is their high genetic and/or functional redundancies. Indeed, individual deletions of effector genes are almost always dispensable for growth in macrophages or amoeba^{254–256}. The fact that *Legionella* encounters and replicates within a remarkably broad range of amoebae and ciliates in its natural environments¹⁰⁸ presumably led to the selection for distinct and “better-suited” effectors, thus causing redundancy in gene repertoire but also making *Legionella* a generalist pathogen. Furthermore, the conservation of targeted-proteins or pathways by *Legionella* between protozoan and human, is presumably the cause of the accidental but successful infection of human macrophages^{256–258}. Consistent with this model, Park et al. (2020) have elegantly demonstrated that, among the large repertoire of effectors, only 18 were important to promote growth in all amoebal hosts and macrophages while the vast majority of other “auxiliary” effectors are only required for replication in one or more amoebae but not in macrophages²⁵⁹.

As variability in the effector repertoire is not the hallmark of successful replication within macrophages, this raises the question of which bacterial determinants are actually important for infection of human cells. To answer this question, Ensminger and colleagues⁵⁵ have developed an experimental approach in which *L. pneumophila* was restricted to grow within mouse macrophages for hundreds of generations. They showed that the evolved strains had accumulated several adaptive mutations as they replicated better in macrophages than the ancestor strain. Indeed, several mutations, such as those found in the genes encoding the flagellar regulator FlhN or the lysine biosynthesis pathway, correlated with the fitness increase and conferred advantage for the replication in macrophages compared to the ancestral strain. Furthermore, identical mutations were present in different lineages that had undergone the same growth restriction, demonstrating parallel evolution due to the selective pressures exerted by the macrophages. These results revealed how the environment may modify a

genome and support the hypothesis that cycling generalist pathogen through only one cell type restrict host-range spectrum.

In this work, we developed experimental evolution strategies with *Legionella pneumophila* with the objective of evolving its virulence properties. In particular, we addressed the question of the intracellular replication robustness and host spectrum. We wondered whether bacteria evolving in the absence of hosts for hundreds of generations lose or retain their ability to infect protists and mammalian cells. We first conducted a mutation accumulation evolution experiment in which several replicate populations, founded from a *L. pneumophila* clinical isolate, were serially propagated on standard agar medium. After several hundred generations of genetic drift, evolved clones exhibited a decreased ability to infect both amoebae and human macrophages compared to the ancestor. This intracellular growth defect was correlated with a reduced capacity to establish the replication-permissive vacuole although the type 4 secretion system, the key element of virulence in *Legionella*, was still functional. Using whole genome sequencing, we have identified all the genomic modifications accumulated during evolution. We especially characterized one mutation, the integration of a 131-kb plasmid element into the chromosome, that was shared by all the independently evolved lineages and conferred attenuated virulent phenotype. In a second time, we aimed at re-evolving one of these attenuated virulent clones under different conditions where we expected the fitness and/or the virulence to rapidly increase. Genetic and phenotypic analyses of re-evolved individuals isolated at different time revealed that the plasmid integration and the associated virulent phenotypes can be reversed when bacteria re-evolved within amoeba host.

Overall, these combined evolution strategies allowed us to “rewind the tape” of the evolution and to decipher the genetic, molecular, and cellular mechanisms underlying the re-emergence of virulence and the adaptation to the eukaryotic hosts.

2.2. Results

2.2.1. Mutation accumulation experiment in *L. pneumophila*

Evolving lineages founded from a common ancestor, *L. pneumophila* Paris strain, were propagated by single-cell bottlenecks on axenic agar medium for several hundred generations (**Figure 20**). *L. pneumophila* Paris is a worldwide-distributed epidemic clone implicated in outbreaks and sporadic cases of LD, frequently isolated from environment and for which the complete genome was sequenced¹⁰⁻¹². Six isolated colonies from *L. pneumophila* Paris clone (the ancestor clone of this experiment) were streaked on new plates of standard *Legionella* agar medium (CYE), thereby generating six independent lineages. We then propagated these six lineages in parallel through several cycles of “single colony isolation and re-streaking on CYE plates”. Each cycle implies bottlenecks corresponding to one single cell, which, after 3 days of growth, gives rise to a single colony. Repeated population bottlenecks have significant implications for bacterial genome evolution due to their potential to lead to genetic drift. Indeed, they interfere with selective processes, including purifying selection of deleterious mutations as well as positive selection for beneficial ones, which results in a reduction in the genetic diversity. These types of evolution experiments, called “mutation accumulation” (MA), lead to the accumulation of all type of mutations in the genome, except those that are highly deleterious or lethal, at the rate of DNA copying errors or unrepaired DNA. The mutations present in the individual (here, a colony) chosen to propagate the lineage are unavoidably fixed in subsequent descendants.

We propagated the six lineages (named Paris 1 to Paris 6) for thirty-five cycles and every five cycles evolved populations isolated from each lineage were frozen at -80°C for subsequent comparison with the ancestor clone Paris (**Figure 20**). Knowing the average number of cells within a single colony after 3 days of growth (approx. 10^6 cells, which correspond to 20 generations) and the number of passages (35 passages), we estimated that we have propagated these six lineages for 700 generations. At this step, we compared these evolved lineages to the ancestral strain at phenotypic and genetic levels as well as the global transcriptomic profiles, which allow for a simultaneous comparison of thousands of expression phenotypes.

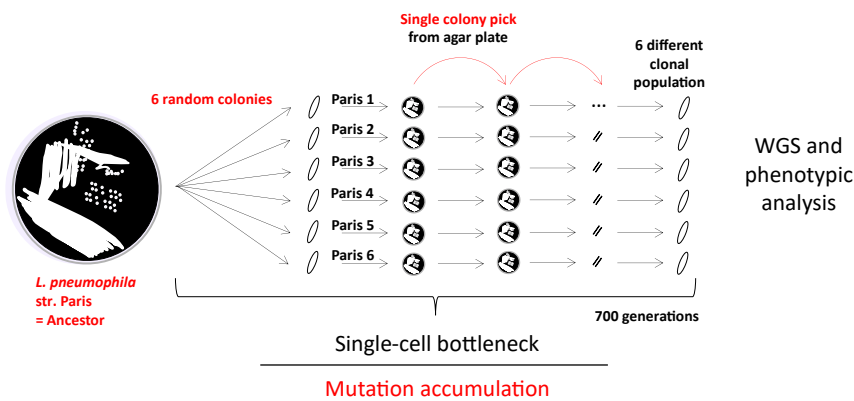


Figure 20: Mutation Accumulation experiment (MA) with *L. pneumophila*

6 lineages founded from a common ancestor – the *Legionella pneumophila* strain Paris – were propagated on standard agar medium for 700 generations (35 passages). Each passage corresponds to the streaking of a single colony from the previous passage on a new plate after approximately 3 days of growth. This strategy leads to the accumulation of non-lethal mutations and changes in fitness and associated phenotypes such as virulence and host-spectrum through the use of a strong population bottleneck.

2.2.2. Growth rate and competitive fitness of *L. pneumophila* lineages evolved under axenic conditions.

After 700 generations of evolution on agar plates, evolved clones were compared to the clinical Paris ancestor for various phenotypes. First, their ability to grow in rich broth was assessed by performing independent growth kinetics. Three isolated colonies from three evolved lineages (Paris 4, Paris 5 and Paris 6) as well as the ancestral clone were used to inoculate 3 mL of AYE rich medium. After 24h of growth at 37°C with continuous shaking, cultures were diluted in fresh medium adjusted to an optical density at 600 nm (OD_{600}) of 0.2 and allowed to grow for a further 24 hours during which OD_{600} was measured every 30 minutes (**Figure 21A**). Cultures from the lineages (Paris 4, Paris 5 and Paris 6) and the ancestor exhibited comparable growth curves, meaning that the 700 generations of evolution under axenic conditions did not affect the replication ability in rich medium of the evolved populations. To check for more subtle changes in the replication capacity, competition experiments between the evolved clones and the ancestor were performed. Indeed, competition assays have the advantage of exacerbating phenotypic differences between two distinct genotypes. Thus, when competing two populations for the same resources – such as nutrients – the population with a slight disadvantage in fitness (ability to duplicate) suffers from a drop in its population count, and accordingly, the proportion of the other population

(presenting a better fitness) increases in the mix. To evaluate the competitive fitness of evolved clones, a competition experiment in broth was performed by using isogenic strain expressing two different fluorescent proteins (Cfp or Yfp) as described in Materials & Methods.

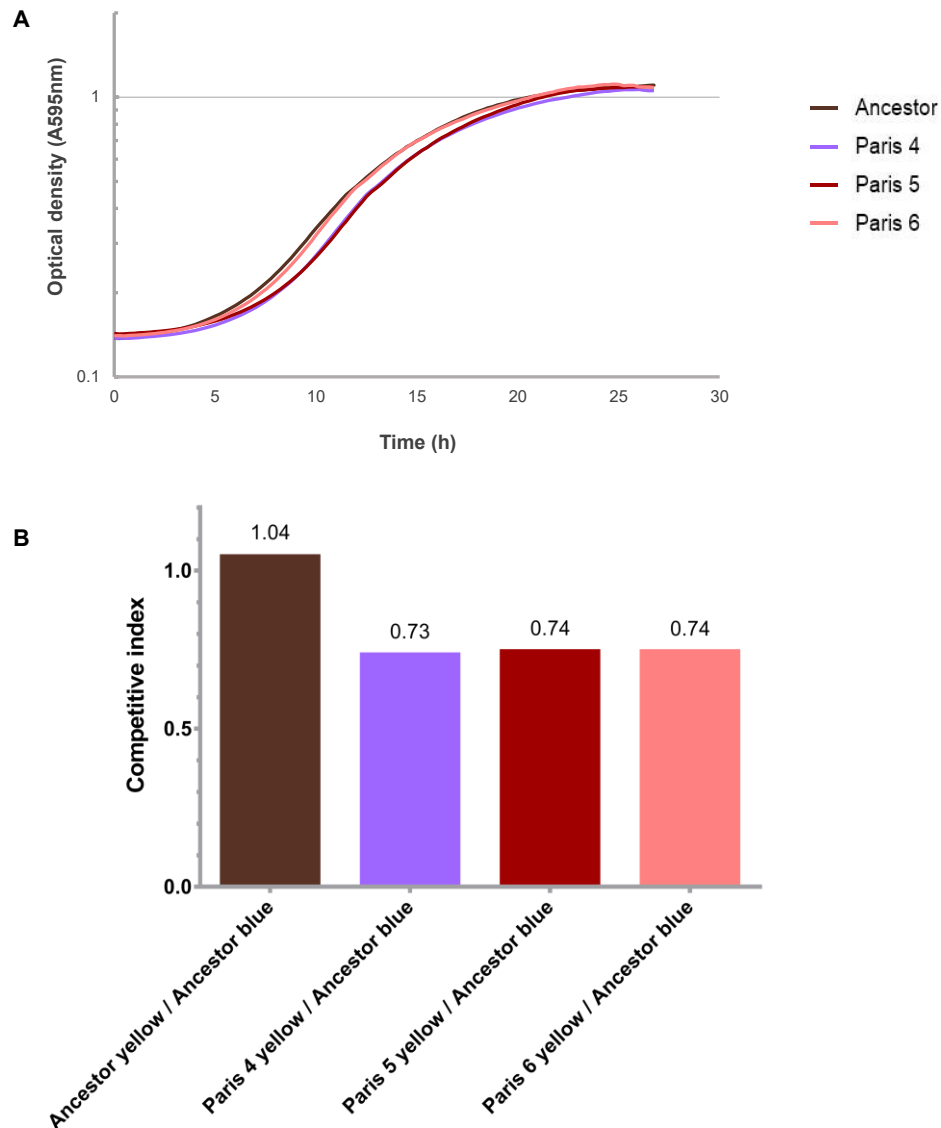


Figure 21: Growth rate and competition of *L. pneumophila* Ancestor, Paris 4/5/6 strains

A: Ancestor strain and Paris 4/5/6 lineages were grown independently in liquid rich medium and OD₆₀₀ was measured to follow growth. **B:** Ancestor was grown in competition with either Paris 4, Paris 5 or Paris 6 lineages in liquid rich medium. Reporter genes of different colours (Cfp for blue or Yfp for yellow) were used to follow growth of the competing populations. Ratio of yellow to blue fluorescence was used to calculate the competitive index

We then calculated the competitive index (CI) after 24h of incubation of both competitors in the same cultures (**Figure 21B**). As expected, the CI of the ancestor was ~1 (1.04) while those of evolved Paris 4, Paris 5 and Paris 6 were 0.73, 0.74 and 0.74 respectively, which corresponds to only 1.4-fold decreased fitness compared to the ancestor.

2.2.3. Altered intracellular phenotypes of *Legionella pneumophila* evolved lineages.

To determine if this slight decreased fitness in rich medium is accompanied by other phenotypic changes, the evolved clones were assessed for their intracellular growth capacities. Paris ancestral clone and its derivatives after 700 generations (Paris 4, Paris 5 and Paris 6) expressing the mCherry-fluorescent protein were used to infect either *Acanthamoeba castellanii* amoebae or U937 macrophages, and the mCherry fluorescence, which reflects bacterial intracellular growth, was monitored every hour over a period of 60h. As expected, fluorescence increases for *L. pneumophila* Paris ancestor (**Figure 22A**) but not for the avirulent $\Delta dotB$ mutant mutant which lacks a functional Dot/Icm T4SS (**Figure S1**). Paris 4, 5 and 6 clones appeared to have a defect in their replicative capacity inside both amoebae and human macrophages (**Figure 22A and 22B**). For *Acanthamoeba* infections, evolved clones exhibited a delay in replication of approximately 24 hours compared to the ancestor while within macrophages, this delay was less pronounced, probably due to the higher permissiveness of these human cell line to the infection by *L. pneumophila*. Furthermore, the evolved clones did not reach the same fluorescence plateau as the ancestor, suggesting not only a delay in infection but also a less pronounced ability to replicate inside human macrophages.

A key step in the infectious process of *Legionella* is the biogenesis of an endoplasmic reticulum (ER)-derived vacuole called LCV, which allows subsequent replication of bacteria. This step is regulated by effector proteins secreted into the host cytoplasm by the Dot/Icm type 4 secretion system (T4SS), including RalF, SidM, LidA, LepB, SdcA and SidC²⁶⁰⁻²⁶². Bacteria lacking the T4SS or one of its critical components (for instance the *dotB* gene) are unable to efficiently deliver the effectors to the cytoplasm of the infected cells, leading to an impaired or delayed intracellular growth. LCV biogenesis was assessed by infecting *Dyctiostelium discoideum* amoeba producing GFP-fused calnexin, an ER-protein, with m-Cherry producing *L. pneumophila* Paris ancestor, its derivatives Paris 4, 5 and 6 or the $\Delta dotB$ mutant.

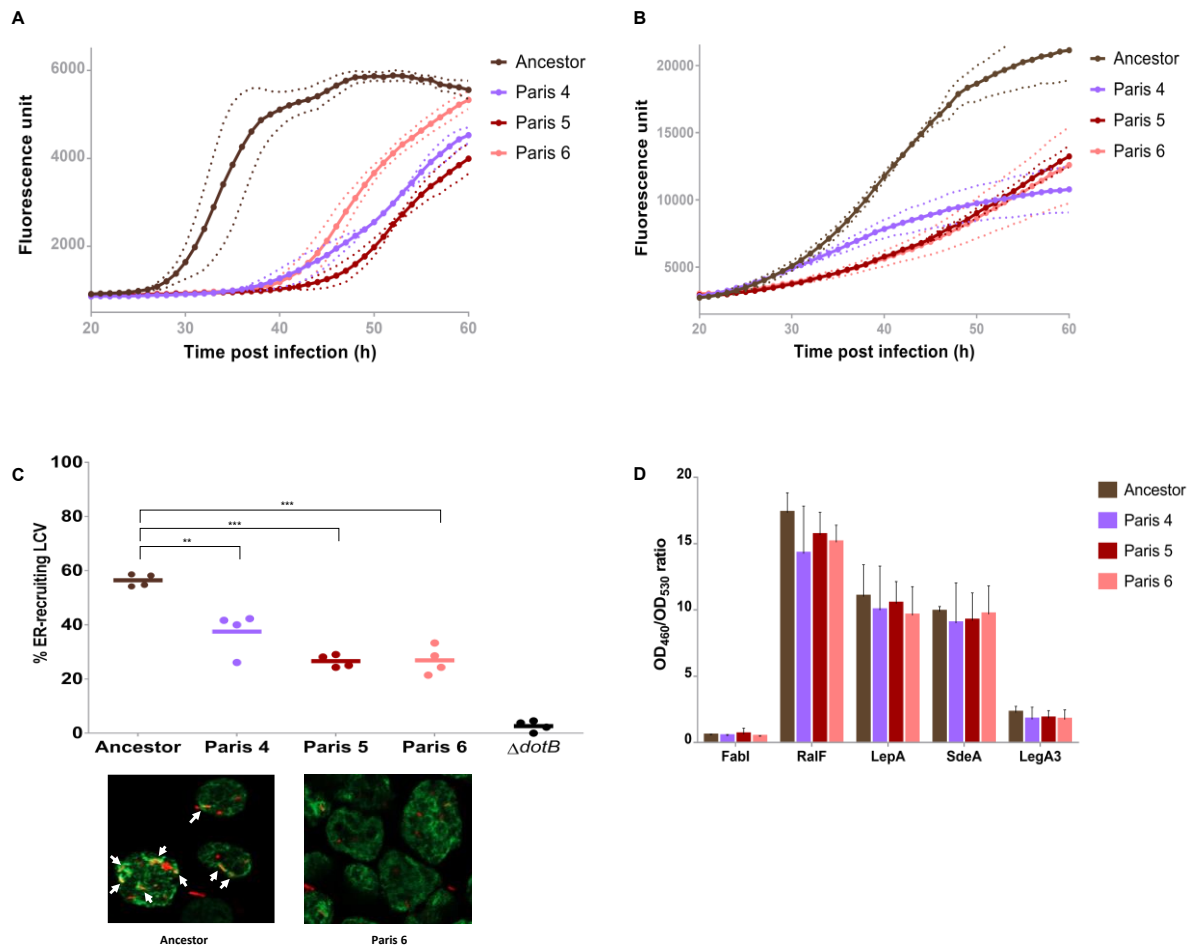


Figure 22: Phenotypes of the MA – generated Paris 4/5/6 lineages

Ancestor and three evolved clones isolated from independent lineages after 700 generations of mutation accumulation were independently grown within either amoeba *Acanthamoeba* spp. (A) or human monocyte-derived macrophages U937 cell line (B) at an MOI of 0,5. C: *Dictyostelium discoideum* amoeba expressing a calnexin-GFP fusion were infected with mCherry-expressing bacteria. ER-recruitment was observed 1h post infection (arrows point green calnexin-positive vacuoles surrounding red bacteria). **: $p \leq 0,05$; ***: $p \leq 0,001$ student test. D: U937 cells are infected with bacteria carrying the various TEM- β lactamase/effector fusion. Infected cells are loaded with the β lactamase substrate CCF4/AM. Then, translocation efficiency is determined by measuring the ratio between cleaved and uncleaved substrate which results in blue (460nm) and green (530nm) fluorescence, respectively.

As observed by confocal microscopy (Figure 22C), 56% of the vacuoles containing ancestor Paris were ER-calnexin positive at 90 min post-infection while the $\Delta dotB$ mutant failed to recruit ER since than 2% of the LCV are decorated with calnexin. The three evolved lineage displayed a reduced percentage of ER-derived vesicles recruitment compared to the ancestor with 37, 26 and 27% for Paris 4, 5, 6, respectively. The most significant drop was with the Paris 5 lineage with 30% less recruitment than the ancestor ($p=0.05$, student t-test).

Since evolved clones exhibited a defect in ER recruitment, we wondered whether this reflected a defect in effector translocation via the T4SS. To assess the functionality of the Dot/Icm machinery in the evolved clones, we thus performed a translocation assay using the

TEM/ β -lactamase reporter system, previously validated to monitor Dot/Icm-dependent effectors translocation by *L. pneumophila*^{263,264}(De Felipe et al. 2008; Allombert et al. 2013). We tested the translocation of 4 effectors arbitrary chosen (RalF, LepA, SdeA and LegA3) and a control protein, FabI, which is cytoplasmic and not translocated by the T4SS (**Figure 22D**). Briefly, plasmids carrying β -lactamase translational fusions TEM-RalF, TEM-LepA, TEM-SdeA, TEM-LegA3 and TEM-FabI under the control of the IPTG-inducible (Ptac) promoter were transformed in Paris ancestor, evolved Paris 4, 5 and 6 and $\Delta dotB$ mutant. These strains were then used to infect U937-derived human macrophages and translocation level of the different effector-TEM fusions was detected 90 minutes post-infection by adding CCF4/AM substrate and measuring FRET reaction. CCF4/AM is a fluorescent substrate composed of a blue fluorochrome (coumarin) and a green one (fluorescein) linked by a B-lactam ring, that rapidly accumulates in eucaryotic cells cytoplasm. Once cleaved, the emitted fluorescence changes from green (emission wavelength 460 nm) to blue (emission wavelength 530 nm) and measuring the fluorescence ratio blue/green provides a quantitative measure of effector translocation efficiency. Of the 4 effectors tested, none displayed significant decrease in translocation level compared to the ancestor (**Figure 22D**). This result suggests that the intracellular growth delay observed in the evolved clones was not directly linked to effector translocation default and that the Dot/Icm T4SS is still fully functional in evolved clones.

Together, these results demonstrated that the populations evolved on agar medium for 700 generations exhibited a strong attenuated virulent phenotype towards both amoebae and macrophages and an altered capacity to create a suitable replicative vacuole within hosts whereas their capacity to replicate in axenic media was not strongly impacted.

2.2.4. The genetic basis of altered virulence phenotypes of evolved populations.

To identify the underlying genetic changes that lead to the altered intracellular phenotypes, we sequenced the genome of Paris 4, Paris 5 and Paris 6 at 700 generations as well as the ancestor clone. Considering that the bacterial cells within the colony are clonal and that a single colony was isolated and re-streaked at each passage, we thus assumed that the population at each passage is quite homogeneous and therefore, it was not necessary to sequence several clones per line and per generation. Nevertheless, in order to identify all the mutations fixed in the population, it seemed appropriate to sequence the same genomes

using two different technologies, Illumina and PacBio. Indeed, PacBio technology provides longer read length than Illumina and thus offer better opportunity for genome assembly and detection of repeated sequences (such as insertion sequences, transposons, CRISPR), gene duplication events or large chromosomal rearrangements.

The whole genome sequencing by Illumina of Paris 4, 5 and 6 identified a few mutations in clones Paris 5 and Paris 6 but none in Paris 4 (**Table 5**). The two mutations detected in Paris 5 are SNPs (single nucleotide polymorphism): one is a synonymous mutation at position 1071 in the *icmS* gene, encoding a component of the Dot/Icm T4SS machinery; the other one is a substitution (G>C) resulting in amino acid change in the receiver domain of LerC (lpp0112). Interestingly, LerC is a single-domain response regulator connecting the PmrAB and the LetAS two-component systems, both of which known to directly regulate the expression of numerous Dot/Icm effector-encoding genes. Thus, LerC appears as a modulator in the sophisticated regulatory network that govern virulence genes expression in *L. pneumophila*. Whether the mutation found in Paris 5 affect the structure/function of LerC, and thus its activity, is still unknown but it would be worth investigating it further. The evolved clone Paris 6 contains the highest number of mutations, resulting in either amino acid changes in the coding sequences of *mreC*, *aph* and *lpp2849* genes, or in a base pair change at one hundred base-pair upstream the promoter region of *lpp2137* gene. As far as we know, none of these genes encode known virulence factors or regulators of gene expression.

Gene name	Description	Position of the mutation		Paris 4	Paris 5	Paris 6	Impact
		Absolute	Relative to ATG				
lpp0112	Putative response regulator receiver	119813	48	-	G/C SNP	-	Leu → Phe
lpp0524	icmF – Component of the Dot/Icm secretion system	570498	1071	-	T/A SNP	-	Synonymous
lpp0874	mreC – rod shape determining protein	984355	793	-	-	C/T SNP	Pro → Ser
lpp1448	aph – spectinomycin phosphotransferase	1614669	251	-	-	A/C SNP	Met → Leu
• lpp2137 • lpp2138	• Unknown function • alk – inner membrane protein	2449963	Intergenic	-	-	A/G SNP	Upstream promoter
lpp2849	Unknown function	3245650	654	-	-	C/G SNP	Asn → Lys
lpp2884 lpp2885	Insertion sequence inserted in the sidH effector coding gene	3291076- 3292913	-	Insertion pLPP	Insertion pLPP	Insertion pLPP	-

Table 5: List of mutations observed in 3 lineages after 700 generations

After 700 generations of evolution, the genomes of the Paris 4, 5 and 6 lineages were sequenced with Illumina and PacBio technologies. Each mutation is recorded and their impact on the proteins identified

The PacBio sequencing allowed the detection of one additional mutation which is shared by the three evolved clones, namely the integration of the 131kb pLPP plasmid into the chromosome (**Table 5**). The plasmid has been integrated into the chromosome by homologous recombination between two copies of the same insertion sequence (IS), one copy presents on the plasmid and another one located on the chromosome. This IS, named ISLpn12, belongs to the IS21 family and encodes two genes organized as an operon, a DDE-type transposase (*istA*, *lpp2885*) and a transposition helper protein containing an AAA-family ATPase domain (*istB*, *lpp2884*)^{265,266}. Both proteins, IstA and IstB were found to be necessary for *in vitro* and *in vivo* transposition^{265,267}.

Thus, the sequencing data from Illumina and PacBio allowed for the identification of all mutations that had accumulated over the course of 700 generations of evolution in the various lineages Paris. This also enabled us to calculate a mutation rate of the bacteria in such conditions, using the following formula: $\frac{\text{Pooled number of mutations in all lineages}}{\text{genome size} \times \text{number of generations} \times \text{clones}}$ ²³

We obtained $\frac{9}{3,5 \cdot 10^6 \times 700 \times 3} = 1,1 \times 10^{-9}$ mutation per base-pair per generation, which is slightly higher but in the same range of mutation rate as those reported for other MA with other pathogens^{23,24,268}.

2.2.5. Spontaneous chromosomal integration of the pLPP plasmid occurs in all lineages

As indicated above, the 131-kb pLPP plasmid was integrated into the chromosome of evolved lines during the evolution experiment. This plasmid was originally found in the clinical Paris strain¹¹ but a more recent study reports that it is commonly found in both, clinical and environmental isolates of the cluster of *L. pneumophila* ST1²⁶⁹. Like most of transmissible plasmids, it encodes a certain number of “accessory” genes that may be involved in the survival of *Legionella* in different environmental conditions (biocides resistance genes, detoxification systems) and also functions required for its own replication, maintenance, and transfer functions (*repABC*, *parA/B*, toxin-antitoxin system, conjugative T4SS...). It also contains a certain number of genes (14) coding for GNAT-acetyltransferase family proteins having an effect on a large scale of substrates (this study) and potential regulators of the virulence (a paralogue of the global repressor of transmissible traits CsrA as well as the TCS

IrpR-IskS shown to be implicated in virulence of other species) ^{11,111,270}. Very recently, it has been shown the role of a deacetylase in epigenetic modifications of the histones of the host cell by *L. pneumophila*, allowing the bacteria to repress genes of the immune response of the host ²⁷¹. We can only hypothesize that proteins from the GNAT-acetyltransferase family could also act as regulators of the host response. Of particular importance in our work, is also the presence of one copy of IS-Lpn12 on the pLPP. Interestingly, this IS is also present twice in the chromosome of *L. pneumophila* strain Paris. One copy is located at position 3 291 075 to 3 292 913 and interrupt the open reading frame (ORF) of the effector-encoding gene *sidH*. The other copy is integrated into another IS (IS-Lpn11 belonging to the ISL3 family), at position 1 742 601 to 1 744 414. The pLPP integration occurred into the IS copy that interrupt the *sidH* gene. We then looked at the occurrence of this IS in other genomes and found it in 21 *L. pneumophila* subsp. *pneumophila* and 9 *L. pneumophila* subsp. *fraseri* strains (**Table 6**). The number of IS per chromosome varies from 1 to 4 with higher copy number in the subspecies *fraseri*. The IS was also present on the plasmids of 3 strains other than Paris. Our analysis indicated the presence of the IS on the plasmid but not on the chromosome of one strain (C9_S) (**Table 6**).

Species/Strains	N° IS on chromosome	Identity on chromosome	N° IS on plasmid	Identity on plasmid
<i>L. pneumophila</i> subsp. <i>pneumophila</i>				
Albuquerque	2	100%		
Flint	2	100%		
HL06041035	1	99%		
OLDA	1	100%		
F4468	2	99%		
F4469	2	99%		
D-7630	2	99%		
D-7631	2	99%		
D-7632	2	99%		
NCTC12179	3	99%		
E5_N	2	99%	1	99%
C9_S			1	99%
D5945	3	97%		
D6026	3	97%		
D-4040	1	97%		
D-5265	1	97%		
D-7787	1	97%		
Thunder Bay	1	100%		
LPE509	1	100%		
Paris	2	100%	1	100%
Corby	1	99%		
<i>L. pneumophila</i> subsp. <i>fraseri</i>				
D-4058	3	97%	1	99%
F-4198	3	97%		
Los Angeles	2	97%		
D-5387	3	97%		
D-3137	4	97%		
D-5744	3	97%		
Lansing	2	97%		
Dallas	3	97%		
Detroit	3	97%		

Table 6: Occurrence of the IS21-like family Insertion sequence on the chromosomes and plasmids of *Legionella* strains.

The presence of the IS21-like insertion sequence has been investigated by blasting the sequence of the IS found in the Paris strain, against all available genomes in the NCBI. For each occurrence, the percentage of identity compared to the IS in the Paris strain has been noted.

As the pLPP integration was common to all the three evolved lineages, we focused on this genome modification. In particular, we sought to determine the dynamic of pLPP plasmid integration over the 700 generations. To do so, we performed PCR amplifications with primer pairs flanking the IS on the chromosome and on the pLPP in order to detect the episomal or integrated form of the plasmid within the bacterial population (**Figure 23A**).

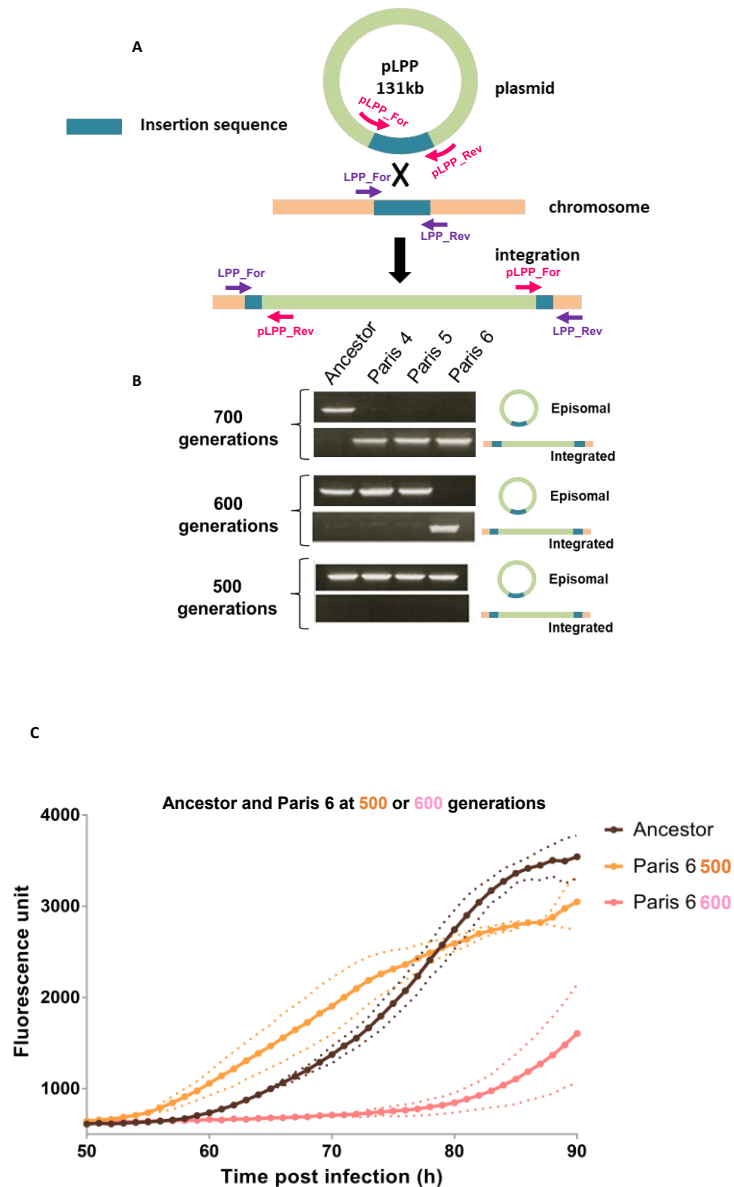


Figure 23: Integration of the 131-kb pLPP plasmid into the chromosome of MA-evolved clones

A: An insertion sequence (IS) present on both the plasmid and the chromosome enables plasmid integration by homologous recombination. **B:** PCR amplifications of either the episomal or integrated form of the pLPP plasmid in Paris 4 to Paris 6 evolved lineages at 700, 600 and 500 generations. **C:** *Acanthamoeba* spp. amoeba infected with evolved clones at 500 or 600 generations

As a confirmation of Pacbio sequencing, we indeed observed the integrated form of the plasmid into the chromosomes of our evolved clones after 700 generations of evolution (**Figure 23B**). We also confirmed the integration occurred in the three other lines Paris 1, 2 and 3 (**Figure S2**). To establish the dynamic of pLPP integration over the time course of the experimental evolution, we tested clones isolated at earlier passages. At 500 generations, the plasmid is in an episomal form in all lines, as in the ancestor strain (**Figure 23B**). At 600 generations, all but one evolved population (Paris 6) still maintained the circular form of pLPP, while between 600 and 700 generations, the plasmid was integrated into the chromosome in all lines. The evolutionary history of Paris 6 is interesting since it demonstrates that the plasmid remains integrated into the chromosome even after an additional 5 passages (approx. 100 generations) after its integration.

Additionally, even though the IS was present twice on the chromosome of *L. pneumophila* Paris, we were able to demonstrate using qPCR amplification that roughly 99% of the time, the homologous recombination occurs between the same two copies of the IS (data not shown, collaboration with C. Ginevra). Hence, the pLPP plasmid integration occurred spontaneously at the same location and almost at the same time in 6 out of 6 evolved lineages. To our knowledge, our results showed for the first time “in live” large molecular rearrangements leading to *Legionella* pLPP integration into the chromosome and highlight the importance of IS elements in this process.

2.2.6. pLPP integration into the chromosome is correlated with intracellular phenotypic change.

Then, to examine the impact of such a mutation we tested the intracellular replicative capacity of Paris 6 evolved individuals before and after plasmid integration, i.e., at 500 and 600 generations. As described above, the clones were transformed with a plasmid expressing the mCherry-fluorescent protein to monitor bacterial growth in amoebae (**Figure 23C**). Interestingly, while the evolved clone at 500 generations had no significant phenotype in comparison to the ancestor, we observed that the intracellular replication of the evolved clone at 600 generations was importantly delayed (**Figure 23C**).

Thus, our results suggest a strong impact of plasmid integration on the intracellular fitness of bacteria, however, the epistatic effect of other mutations on this phenotype cannot be ruled out. Remarkably, our findings also revealed a parallel evolution between the six lineages which is commonly regarded as evidence of natural selection. This suggests that bacteria have been exposed to common selection pressures, such as common environmental factors. *In fine*, this specific genetic modification likely confers an adaptive advantage to bacteria under these specific evolutionary conditions, but not within host cells.

2.2.7. Major transcriptomic shifts in pLPP-integrated populations

To identify the molecular mechanisms underlying the intracellular growth defect in evolved populations, we examined the global gene expression profiles of Paris 4, 5 and 6 (at 700 generations) and compared them with each other and with the ancestor. Total RNAs were extracted from bacteria grown in liquid media at 37°C until they reach the post-exponential growth phase, mimicking the entry into the transmissive phase ²²⁹. The threshold for differential expression was set at $|\log_2FC| \geq 2$ with an adjusted p-value $p_{adj} \geq 0.05$.

A large number of genes are differentially regulated in each of the evolved clones compared with the ancestor, 16.5%, 21.6% and 14.7% of genes were downregulated and 15.7%, 21.8% and 15.1% were upregulated in Paris 4, Paris 5 and Paris 6, respectively ([Suppl. results](#)). Venn diagrams show the number of up- and down-regulated genes in each clone (**Figure 24A**). Considering the large number of expression changes in all three clones, we first decided to focus on common genes in order to identify common evolutionary pathways between lineages. Genes shared between evolved clones are, at least in part, those whose expression is modified following plasmid integration since this mutation is common to all three clones (**Figure 24A**). Many functional categories are impacted: The genes involved in metabolism, transport, translational process and typeIV pili as well as 46 Dot/Icm effectors genes are up-regulated, while those linked to flagella biogenesis (26 out of 45 flagellar genes) and virulence-regulatory pathway (RpoS, LetE, LetS, Hfq, HU, Fis2) are globally down-regulated (**Figure 24B and [suppl. results](#)**). Our results also showed that the vast majority of plasmid genes (138/142 genes) are repressed in the evolved clones (\log_2FC varies between -1 and -2,5). We hypothesized that the slight down-regulation of plasmid genes expression did not reflect a real repression of genes but was rather due to a gene dosage effect caused by a decrease in plasmid copy number following its chromosomal integration.

A heatmap representing the 50 genes whose expression has been most significantly altered in Paris 4, 5 and 6 is shown in **Figure 24C**. Strikingly, Paris 4 and Paris 6 displayed similar transcriptional profiles in comparison to Paris 5. Paris 4 and 6 have no common mutation except for the integration of the plasmid (**Table 5**), suggesting that the gene expression changes are linked to this mutation. In contrast, Paris 5 harbours an additional mutation in *LerC* (*lpp0112*), a connector between *PmrA/B* and *LetA/S* regulatory pathways, which in turn regulate the expression of several effector genes. It's plausible that *LerC* mutation impacts a large number of genes and likely masks the global transcriptomic effects due to pLPP integration. Most of the genes (39/50) are upregulated. The functional categories to which the proteins coded by these genes belong to are Dot/Icm substrates (24.5%), pili biogenesis (8.2%) surfactant biosynthesis (8.2%) and stress response (8.2%).

On the opposite, of the 11 downregulated genes, 6 are coding for the biogenesis of the flagellum (12.2% of the 50 genes). According to these data, the evolved clones appeared less motile than the ancestor strain on live microscopic observations. It has been shown that flagellum is expressed in post-exponential phase and participate in triggering host-cell immune response²⁷². *Legionella* motility is also required for proper host-cell invasion as it promotes contacts with host cells but has no impact on intracellular replication²⁷².

Together, these transcriptomic data reflect the huge transcriptomic shift linked to the chromosomal integration of pLPP. It is plausible that plasmid genes directly regulate the expression of genes carried by the chromosome, but this would have only a limited impact on gene expression. Faced with so many transcriptomic changes, it is likely that global regulatory mechanisms of gene expression are involved. Changes in the expression of global regulators such as *RpoS*, *LetE*, *LetS*, *Hfq*, *HU* or *Fis2* would certainly lead to strong pleiotropic effects. we can also hypothesize that the integration of a 131-kb plasmid leads to topological modifications of the chromosome that would render certain regions of the chromosome more or less accessible to RNA polymerase and thus impact global gene expression.

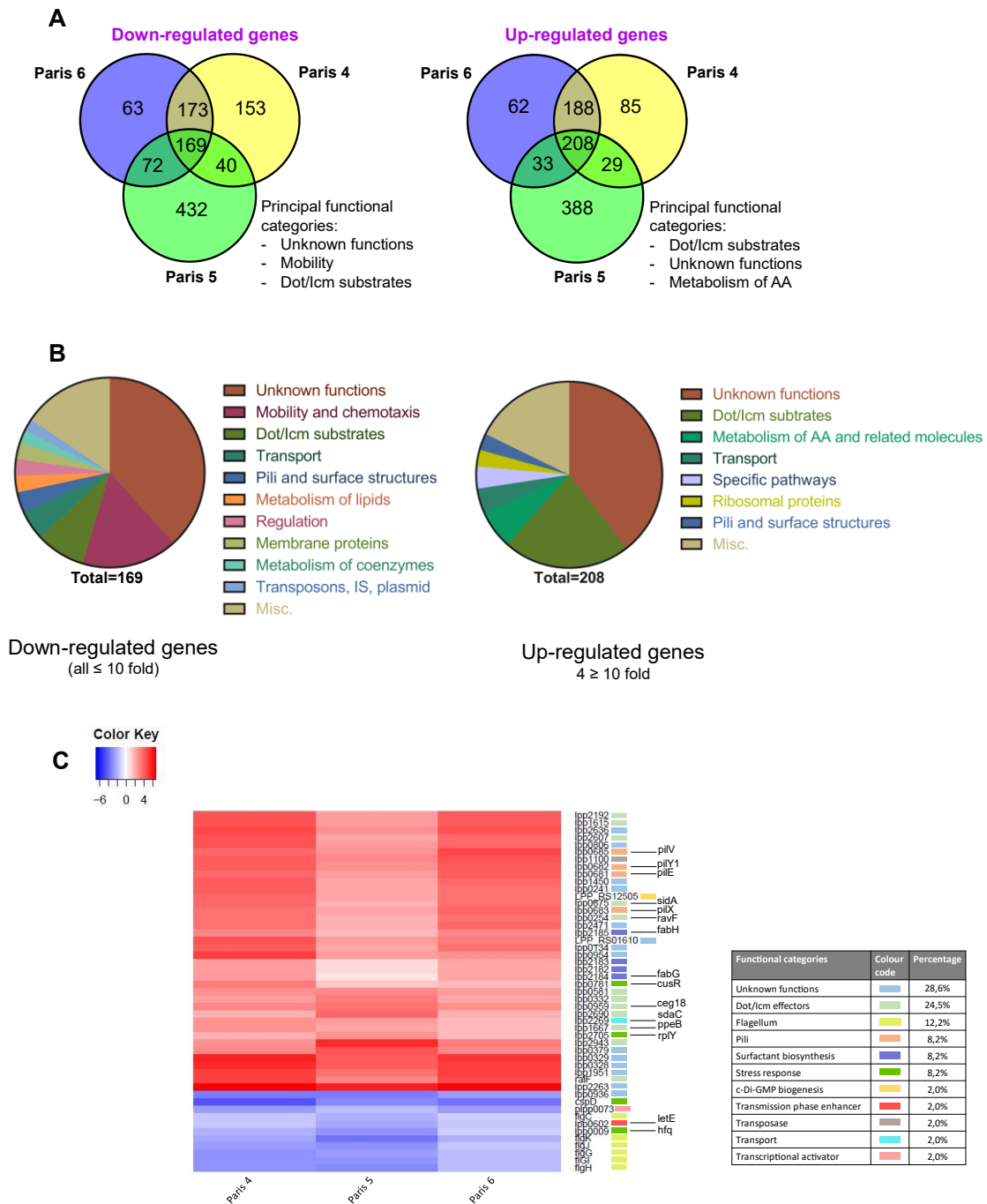


Figure 24: Transcriptomic landscape of MA-evolved clones

A: Venn diagrams with all down- or up-regulated genes in Paris 4/5/6 compared with the ancestor strain and associated principal functions of coded proteins. **B:** Pie charts of functional categories represented by all down- or up-regulated genes, common to all three lineages. **C:** Coloured heatmaps were generated with results from RNA-sequencing by comparing the expression of the 50 most differentially expressed genes between clones founded from the mutation accumulation experiment (Paris 4, 5 and 6) and the clinical ancestor strain Paris (left). The Color Key panel indicates the relative expression of each gene compared to the ancestor's, from dark blue (highly under expressed) to dark red (highly overexpressed). The right table displays the functional categories which the indicated genes belong to with the proportion in percentage of each category.

2.2.8. Reacquisition of virulence traits depends on the environmental conditions in which bacteria replicate.

In order to determine if the partial loss of intracellular replication capacities is reversible and to identify the molecular mechanisms of the virulence reacquisition, we sought to re-evolve one of the virulence-attenuated clones under two distinct experimental conditions. In one part of the experiment, the bacteria were propagated in presence of eukaryotic hosts so that they depend on the infection and replication in host cells to survive while in the other part, the bacteria were replicated in liquid broth. In both conditions, bacteria were propagated through serial passages with large-population bottlenecks at each passage (**Figure 25**). These conditions are known to lead to the predominance of the most adapted individuals in each environment (medium adaptation) and to select for mutations that increase the fitness under the conditions tested.

Ten independent lineages were founded from the clone Paris 6 describe above, which we have renamed in this context the "Evolvir ancestor". Five lineages were propagated within amoebae (named AX) and five in broth (named BX) (**Figure 25A**) according to the following conditions: amoebae were infected with Evolvir ancestor at a MOI of 0.5 and after 3 days of infection and the complete lysis of amoeba cells, a fraction of the bacterial population was used to re-infect a fresh batch of amoebae (at the same MOI). In parallel, five other cultures of Evolvir were propagated in AYE medium, by serial passage three times a week of a fraction of the population (100 to 1000-fold dilutions).

Both, the "Amoeba" lines (AX) and "Broth" lines (BX) were propagated for 35 passages (corresponding to 300 and 400 generations respectively) (**Figure 25A**). As the bacteria quickly adapted to their environment during the evolution experiment, we had to reduce the inoculum size and/or the time between two passages in order to maintain the bacteria in similar physiological conditions at each passage. Indeed, our main concern was to avoid bacteria from AX lineages remaining too long outside the amoebae after their lysis or in a prolonged stationary phase for the BX lines. These variations in frequency and inoculum size were schematized in **Figure 25B**.

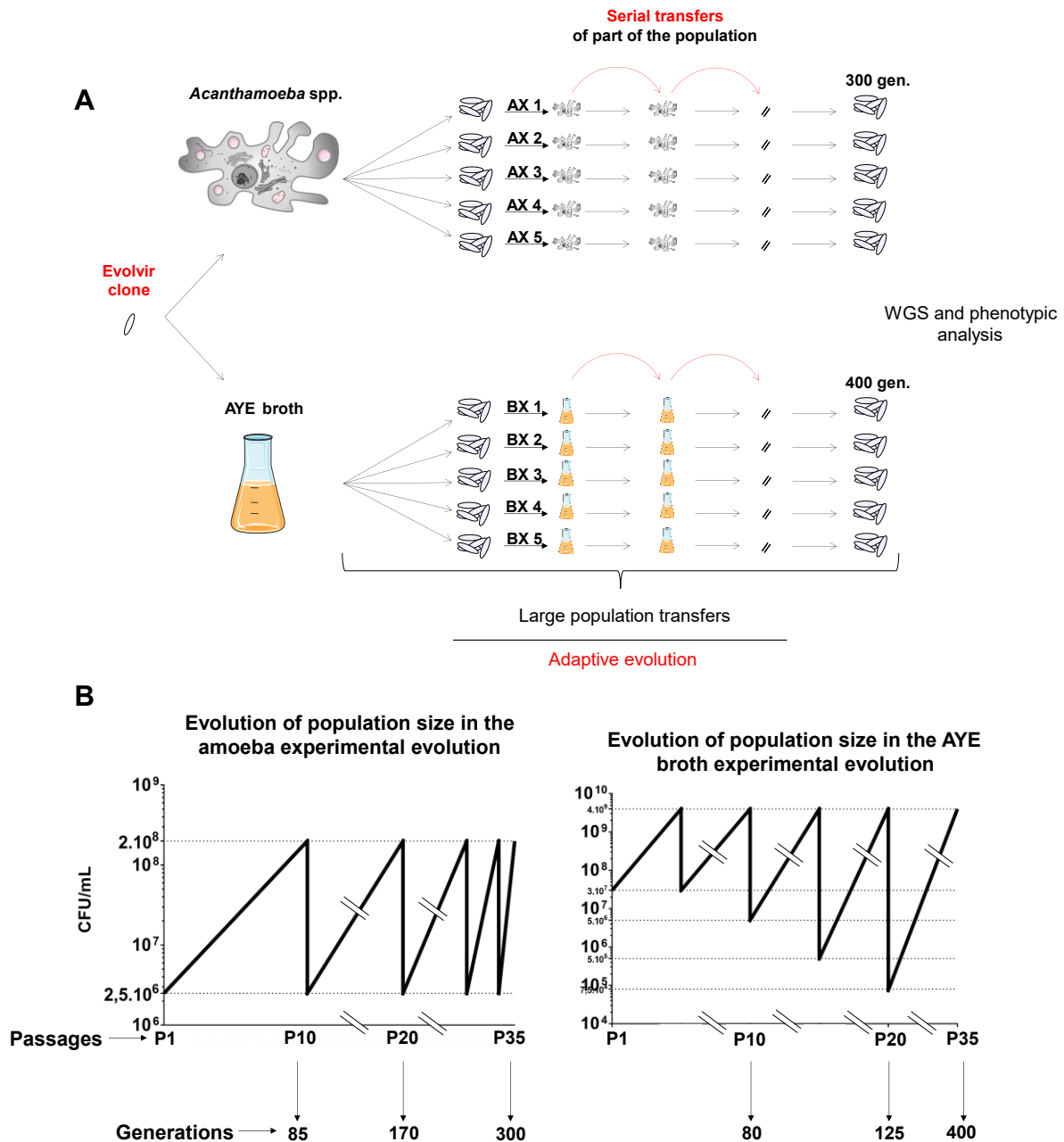


Figure 25: Adaptive evolution experiment (AE) strategies

A: Five lineages, derived from the attenuated virulent clone Evolvir, were propagated inside *Acanthamoeba* spp. (AX lineages) or in AYE broth (BX lineages). For AX lines, at each passage, a fraction of the bacterial population was used to re-infect fresh amoebae until complete host-cell lysis. For BX lines, at each passage, a fraction of the bacterial population was used to re-inoculate fresh medium. This experimental design leads to adaptive evolution because ample genetic diversity is maintained through each transfer. B: Schematic representation of the population size and evolution across both AE strategies. At passage 10, 20 and 35, number of generations was calculated and reported below.

After 35 passages, we analysed the phenotypic changes in evolved populations of both strategies. As before, amoebae and macrophages were challenged with populations from each lineage and mCherry-fluorescence was monitored over 60-80 hours post-infection to determine the growth dynamics in both host cell types (**Figure 26**). The lineages passaged within amoebae for 300 generations replicated more efficiently in *A. polyphaga* and U937 macrophages than did their Evolvir ancestor. Their replication rate is even as high as that of the Paris clinical strain (**Figure 26A and 26B**). In contrast, the lineages propagated in broth displayed similar growth in *A. polyphaga* than Evolvir clone (**Figure 26C**), and more interestingly, they seemed to have completely lost the capacity to replicate within macrophages (**Figure 26D**).

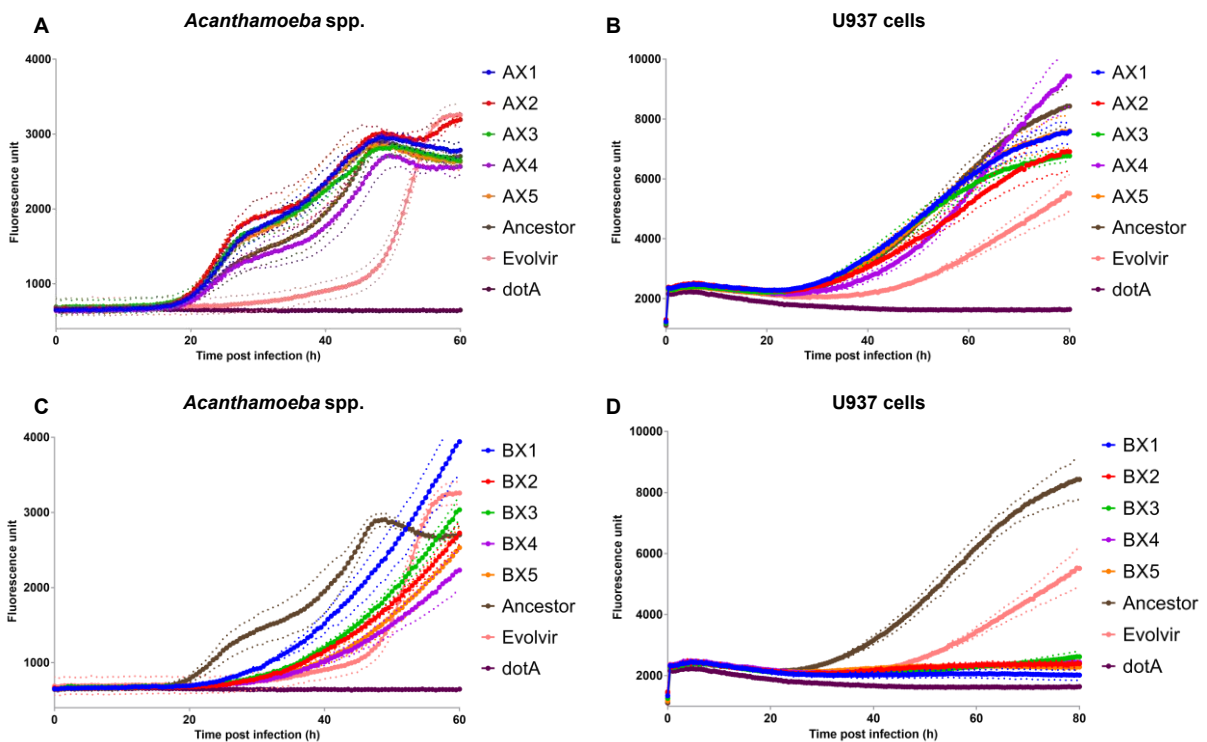


Figure 26: Growth kinetics of Evolvir and AX and MX populations within amoebae *Acanthamoeba* spp. or U937 macrophages were infected by Ancestor, Evolvir and clones from adaptive evolution populations after 300 (AX lineages, **A and B**) or 400 (BX lineages, **C and D**) generations at an MOI of 0.5.

On the one hand, prolonged passage of an attenuated strain in a single host cell type (amoebae) for hundreds of generations restores the virulent phenotype without narrowing the host spectrum, since the evolved strain replicates very well in macrophages. On the other hand, lines re-evolved in broth without the selection pressure of the host cell do not reacquire virulence. On the contrary, it appears that the changes occurred during adaptative evolution are highly deleterious for the replication in macrophages.

2.2.9. Reversible integration of pLPP in re-evolved populations.

So far, our findings suggest that the chromosomal integration of the pLPP negatively impacts the intracellular growth capacities of *L. pneumophila* Paris. Then, to determine whether the virulence reacquisition, is accompanied by pLPP excision from the chromosome, we tested the status of the pLPP plasmid in the lineages generated from the Evolvir ancestor. Instead of performing conventional PCR on one single colony as described above, we performed digital droplet PCR experiments directly on the bacterial population. We first had to improve the protocol with the objective to statistically detect only one single individual per droplet. This allowed us to have an overview of the plasmid status at the single cell level in the entire population, and for each lineage. Populations from three different passages were tested (passages 10, 20 and 35). For each population, approximately 10,000 droplets were analyzed by ddPCR quantification system to determine the proportion of episomal and integrated form of the plasmid. The results of the experiment are presented in **Figure 27**.

In the Evolvir ancestor population, the percentage of plasmid-integrated bacteria was 99,1%. After 10 passages within amoebae (85 generations), this percentage decreased but the PLPP was still integrated in the vast majority of bacteria (between 85% and 65% depending on the lineage) (**Figure 27A**). At passage 20, less than 1% of bacteria harbored the integrated form of pLPP, except in lineage AX3 in which it was still predominant (65% of bacteria). At passage 35, the pLPP is episomal in the whole population of every lineage. Thus, the excision of the plasmid from the chromosome is correlated with the reacquisition of intracellular replication capacities in these re-evolved populations.

The scenario was much more contrasted in the bacteria that have evolved in broth environment (BX lineages) (**Figure 27B**). For instance, in lineage BX1, the pLPP was excised from the chromosome in most of the bacteria earlier than passage 10 (80 generations) while in BX5 line the plasmid remained mostly integrated into the chromosome throughout the

experiment evolution (passages 10, 20 and 35). In BX2, 3 and 4, the integrated form was detected in a fraction of the population at passages 10 and 20 and was still present, albeit at low level, after 35 passages. In all lines, except BX5, it seems that the rate of excision never reaches 100%, meaning that a subpopulation retained an integrated form of the plasmid.

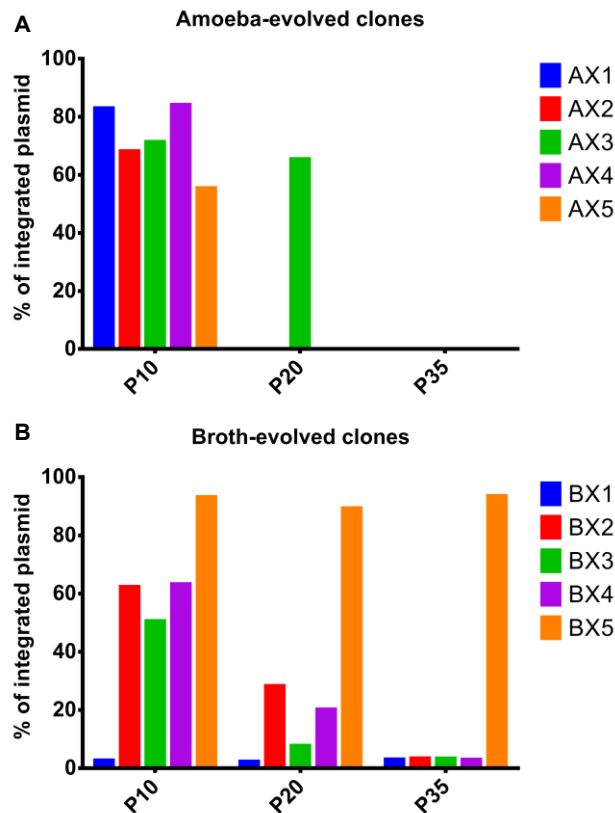


Figure 27: Percentage of pLPP integration into the chromosome of the amoeba or liquid medium-evolved populations at several stage of evolution

Digital droplet PCR was performed directly on bacteria from AX (A) or MX (B) populations to verify the form of the pLPP plasmid, either episomal, or integrated into the chromosome

These observations suggest either that pLPP must be excised from the entire population to restore the virulent phenotype, or, more likely, that other epistatic mutations have accumulated in the genome of these populations and thus mask the phenotypic effects related to pLPP excision.

In summary, all these results demonstrate that pLPP not only spontaneously integrate into the chromosome of *L. pneumophila* but can also subsequently excise, restoring the phenotypes unless other epistatic mutations appeared. Hence, the integration and excision of this large

plasmid element might represent one mechanism that *L. pneumophila* exploits to better adapt to different environments, and specifically to its natural host, the amoeba.

2.2.10. Accumulation of multiple mutations in re-evolved lines results in loss-of-function of RelA, LetS/LetA regulators

To investigate the diversity of the mutations that accumulated during AX or BX evolution experiments, 3 independent clones per lineage (at passage 35) were sequenced using Illumina technology. All the mutations are reported in **Table 7**. Surprisingly, we found only 6 mutations in the AX lines after 300 generations of evolution (**Table 7A**): one mutation in AX1 and AX2 and two independent mutations in AX3 and in AX4. In each case, the mutations are present in all 3 sequenced clones, except for AX1 line in which the mutation identified is present in only 1 clone out of 3. Even though an impact of these mutations on the bacterial physiology cannot be ruled out, our findings strongly imply that the excision of the pLPP plasmid in AX lines is most likely responsible for the reacquisition of virulence in those lines. The situation is a bit different in the BX lines, with a total of 25 mutations identified in BX1, BX3 and BX5 lineages (**Table 7B**). The genomic sequences of BX2 and BX4 clones have not yet been analysed. Strikingly, several independent mutations affecting *relA*, *letA* and *letS* genes were identified, suggesting that modifications in these loci confer a growth advantage in broth. RelA is an enzyme involved in the synthesis of (p)ppGpp (guanosine tetraphosphate), an alarmone produced when bacteria lack nutrients, particularly an amino acids starvation^{176,273}. The *letA* and *letS* genes encode two-component system (TCS) LetA/LetS (*Legionella* transmission activator and sensor, respectively), an important regulator system inducing the expression of the non-coding (nc)-RNAs RsmX, Y and Z which ultimately affect the expression of a large set of virulence genes^{178,274,275}. Both, RelA and LetS/LetA TCS are part of the regulatory network that governs the transition from non-virulent replicative to virulent transmissive phase in *Legionella*, including morphological differentiation (including flagellum) and virulence traits expression¹⁷⁹.

A

Gene name	Description	Position of the mutation		AX1	AX2	AX3	AX4	AX5	# of clones	Impact
		Absolute	Relative to ATG							
lpp1610	conserved membrane protein of unknown function	1795109	283	A/T SNP	-	-	-	-	1/3	premature STOP codon
lpp0597	sucA, oxoglutarate decarboxylase involved in TCA	642534	2115	-	C/G SNP	-	-	-	3/3	synonymous
lpp2360	mRNA Cleavage and polyadenylation specificity factor	2704169	357	-	-	C/T SNP	-	-	3/3	synonymous
lpp2884	putative transposase	3410374	1613	-	-	A/T SNP	-	-	3/3	Leu → Gln
lpp2296	dienoyl-CoA reductase (Fatty acid and phosphatidic acid)	2637591	1807	-	-	-	C/G SNP	-	3/3	Asp → His
lpp0104	hypothetical protein	108003	488	-	-	-	C/A SNP	-	3/3	Thr → Asn

B

Gene name	Description	Absolute position of the mutation	BX1	BX3	BX5	# of clones
Upstream lpp0087 and lpp0088	Intergenic	90653	-	ATC Insertion	-	1/3
lpp0525	lcmH/DotU – component of the Dot/ICM system	572053	-	C/A SNP	-	1/3
lpp0873	Mbl – rod shape determining	983274	G/A SNP	-	-	1/3
Upstream lpp0930	Toluene ABC transporter	1012203	-	G/A SNP	-	1/3
lpp1328	rodA – cell wall determining protein	1482169	-	T/C SNP	-	1/3
lpp1328	rodA – cell wall determining protein	1482440	G/T SNP	-	-	2/3
lpp1328	rodA – cell wall determining protein	1482912	-	A/T SNP	-	1/3
lpp1328	rodA – cell wall determining protein	1482919	-	-	C/T SNP	3/3
lpp1329	mrDA	1484210	-	C/T SNP	-	1/3
lpp1413	RelA – stringent stress response	1571513	-	T insertion	-	1/3
lpp1413	RelA – stringent stress response	1571585	-	A/T SNP	-	1/3
lpp1413	RelA – stringent stress response	1571747	-	-	T insertion	3/3
lpp1413	RelA – stringent stress response	1571889	T/A SNP	-	-	1/3
lpp1413	RelA – stringent stress response	1571972	-	G/T SNP	-	1/3
lpp1413	RelA – stringent stress response	1572312	T/A SNP	-	-	2/3
lpp1744	motD – flagellar motor protein	1966435	-	G/C SNP	-	1/3
lpp1880	multidrug DMT transporter permease	2124942	C/T SNP	-	-	1/3
lpp1887	gacS/letS – <i>Legionella</i> transmission sensor LetS (His kinase)	2133283	T deletion	-	-	3/3
lpp1887	gacS/letS – <i>Legionella</i> transmission sensor LetS (His kinase)	2134035	-	-	29 bp deletion	3/3
lpp1887	gacS/letS – <i>Legionella</i> transmission sensor LetS	2134067	-	A deletion	-	1/3
lpp2323	PHA synthase	2664860	-	C/T SNP	-	2/3
lpp2699	gacA/letA – <i>Legionella</i> transmission activator LetA	3081466	-	A/T SNP	-	2/3
lpp2987	conserved protein of unknown function	3537306	T/C SNP	-	-	1/3
lpp3049	M4 peptidase	3605409	-	-	A deletion	1/3

Table 7: Mutations observed in the AX and BX lineages

Sequencing of various evolved clones from the Amoeba (AX; panel A)- or Broth (BX; panel B)-evolved lineages yielded the mutations presented in the table. Characteristics of the mutations are reported as follows: name of the gene mutated, its function, absolute position of mutation in the genome and relative to the ATG of the gene (panel A only), nature of the mutation, number of clones concerned and impact on the protein (panel A only). Lineages BX2 and BX4 are pending analysis

To investigate further the diversity and distribution of the *relA*, *letA* and *letS* mutations in BX lines, we performed targeted-NGS (next-generation sequencing), which allows to focus on specific regions of interest. Finally, a total of 18 different mutations were identified 4 in *letA*, 7 in *letS*, and 7 in *relA* genes (**Figure 28**).

LetS is a 103-kDa sensor kinase predicted to have three cytoplasmic signaling domains, namely, a transmitter, a receiver, and a histidine phosphotransfer domain (**Figure 28**). By analogy to its orthologs BarA/UvrY in *Escherichia coli* or GacS/GacA in *Pseudomonas*, it is predicted that, upon receiving a signal, LetS autophosphorylates on a conserved histidine residue in the transmitter domain. Then, the phosphate is sequentially transferred to an aspartic acid residue in the receiver domain, on a histidine in the phosphotransfer domain and finally to the cognate response regulator, LetA (**Figure 28**). Most of the mutations identified in LetS introduce a premature stop codon resulting in a truncated protein. In these circumstances, it is very likely that phosphate transfer to the cognate response regulator LetA is interrupted, as well as the downstream regulatory cascade.

LetA is a 43-kDa activator kinase that contains a receiver domain and a helix-turn-helix (HTH) motif required for DNA binding (**Figure 28**). The receiver domain exhibits a conserved aspartic acid residue which is indeed the *phosphate* acceptor site. Phosphorylation of this aspartate induces conformational changes that mediates dimerization of LetA and its activation as a transcriptional regulator^{184,276,277}. The mutations identified in BX lines are located in the receiver domain of LetA or between the receiver and HTH domains. It is possible that these mutations induce structural changes that affect phosphorylation, dimerization or DNA binding of LetA but, to date, we have no data to confirm these hypotheses.

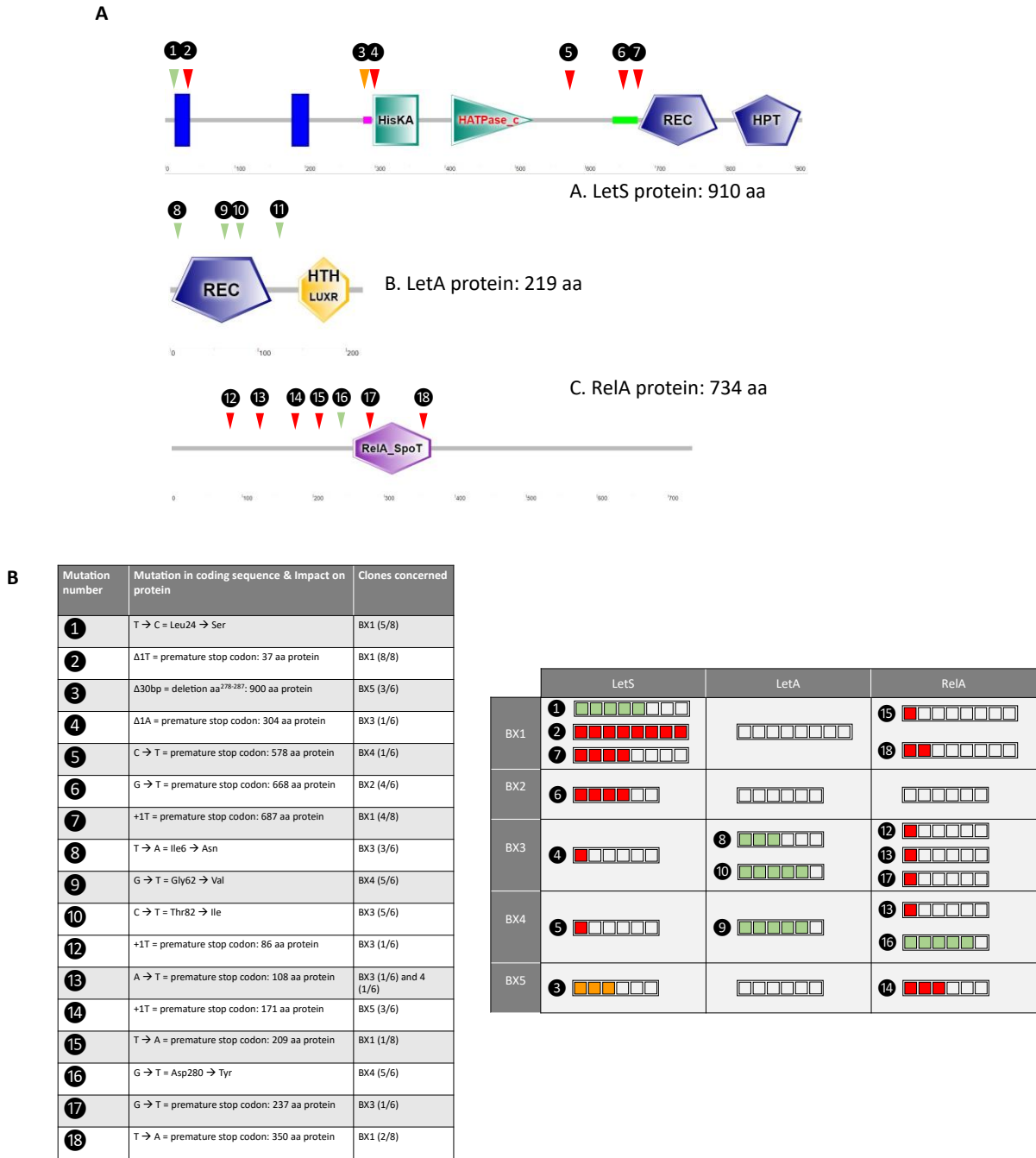


Figure 28: Identification of the mutations in the LetA/LetS two component response system and in the (p)pGpp synthase RelA

Whole genome sequencing was performed on 8 (MX1), 6 (MX2, MX3, MX4 and MX5), or 5 (AncAX) clones from the evolved lineages after 400 generations of adaptive evolution in broth (MX lineages) or amoebae (AX lineage). **A:** SMART analysis of LetA, LetS and RelA with arrows pointing at the mutated zones. Colours indicate the consequence of the mutations on the protein (red: stop codon, orange: deletion, green: change in amino acid sequence). Numbers above arrows are used to identify the mutation in table **B**. **B:** Tables with depiction of the mutations found in the LetA/LetS/RelA coding genes and their impact on protein. On the right-hand side, coloured squares represent each a clone of a given lineage

RelA is a 734 amino acids protein containing two distinct domains, one in the N-terminal domain (from aa 256 to 366) which is responsible for (p)ppGpp synthesis, and the other in the C-terminal domain which is thought to be involved in the regulation of RelA activity²⁷⁸ (**Figure 28**). All but one of the mutations found in RelA introduced a premature stop codon the catalytic domain, or upstream, suggesting that this enzyme can no longer synthesize (p)ppGpp.

Taken together, these results demonstrate that BX lines have accumulated multiple mutations in RelA and LetA/S TCS, leading to the loss of function in one or both regulators. This probably causes the shutdown of the downstream regulatory cascade that governs the entry into the transmissive phase and the expression of the most important virulence genes, which ultimately leads to loss of virulence in these lines.

2.3. Discussion

In this study, we developed experimental evolution strategies in order to have a better understanding of the evolutionary processes leading to increased virulence in some *L. pneumophila* strains. The *L. pneumophila* ST1, also known as the “Paris-like strain”, have been detected worldwide¹² and accounts for a disproportionately high number of clinical cases²⁷⁹. These observations suggest that the expansion of this particular clone may be related to specific fitness and/or virulence traits. The experimental approaches we have developed have enabled phenotypic and genetic comparisons to be made between an ancestor and all its descendants, allowing us to examine how natural selection is able to (re-)shape and improve genomes and which genes or functions are crucial for the emergence *Legionella* virulence.

We found that the large conjugative pLPP plasmid, commonly found in *L. pneumophila* ST1^{11,269}, has a global influence on the physiology of Paris clinical isolate. Integrated pLPP attenuates its fitness and virulence while the episomal form of the plasmid improved its survival in both amoebae and macrophages, highlighting the importance of this plasmid. Large conjugative plasmids are often critical for the virulence and/or the spread of antimicrobial resistance in many bacterial pathogens. They can confer beneficial characteristics to bacteria in certain circumstances but can also become highly deleterious

when these environmental conditions change^{280,281}. To ensure their replication and maintenance within a bacterial population, plasmids have evolved dedicated mechanisms such as replication and partitioning systems as well as, frequently, toxin and antitoxin (TA) systems. TA systems encode a toxin and its antidote; the former is stable while the latter is more labile. Following cell division, the plasmid, and its cognate TA module, is transmitted to daughter cells; if one daughter cell loses the plasmid, it is unable to produce any more antitoxin to counteract the residual toxin, which ultimately inhibit the growth of the plasmid-free cell²⁸². Little is known about the functions encoded by the pLPP, however we do know that it contains at least one TA system (plpp0089-0090) to ensure its maintenance within the population. Here, we described a novel mechanism that promotes pLPP maintenance *via* its spontaneous integration into the chromosome. This inter-molecular event occurs through homologous recombination between two copies of IS-Lpn12, one located on the plasmid and the other on the chromosome. Interestingly, we also demonstrated for the first time that this recombination event is completely reversible underlining the extreme plasticity of *Legionella* genome and how this genetic plasticity contribute to plasmid adaptation and maintenance within a population.

Even more unexpectedly, our results also illustrate how plasmids manipulate the expression of a wide range of bacterial phenotypes, enabling bacteria to adapt to new ecological niches. Indeed, upon plasmid integration, which happens systematically under our evolution conditions, *Legionella* cells operate a major shift in global gene expression. This event not only affects the plasmid genes, but also triggers changes in the expression of a large number of chromosomal factors. For instance, our transcriptomic data revealed that metabolic pathways involved in bioenergetics or metabolism of amino acid, nucleotides and lipids are up-regulated while most flagellar genes are down-regulated, which may account for the reduced motility observed in bacteria with the integrated form of the plasmid. Other virulence-related traits are affected by the pLPP integration such as the short and long pili clusters as well as 60 Dot/Icm effector-encoding genes. It is highly probable that all these changes in chromosomal genes expression affect the fitness and the virulence of *Legionella*.

The molecular mechanisms underlying the phenotypic and genetic changes following pLPP integration are not yet understood. However, several hypotheses can be put forward to explain these fluctuations. First, the diminution in plasmid copy number following the

chromosomal integration of pLPP may lead to a gene-dosage effect and an overall decrease in plasmid gene expression that might result in phenotypic changes, notably in virulence. As an example, in *Yersinia pseudotuberculosis*, the pIBX copy number is critical for virulence²⁸³. Indeed, only one copy of the virulence plasmid encoding the type III secretion system (T3SS) is not sufficient to drive a systemic infection in mice whereas three copies are sufficient to kill them. Thus, *Yersinia* has evolved mechanisms to increase the plasmid copy number during the course of infection despite the increased metabolic burden caused by the three copies²⁸⁴.

A second hypothesis lies on the presence of several transcriptional factors (TF) on the pLPP plasmid that could directly alter the expression of chromosomal genes. In particular, we noticed the presence of three TFs belonging to the MerR-family regulators, encoded by the genes *plpp0129*, *plpp0088* and *plpp0073*, the latter being strongly repressed in the 3 evolved clones in parallel. The MerR-family TFs usually respond to environmental stimuli, such as oxidative stress, cellular metal ions or antibiotics, that are specifically recognized by their "effector-binding domain" located in the C-terminal of the protein, while the N-terminal helix-turn-helix domain is required for DNA binding^{285,286}. Their mode of action is original as they function as transcription repressors under "normal" circumstances, but quickly switch into activators upon stimuli detection²⁸⁶. To date, there is no evidence that these three plasmid-encoded regulators recognize *Legionella* chromosomal genes and modulate their expression. It would therefore be interesting to investigate their target genes and regulatory mechanisms in future works.

A third possibility is that plasmid-encoded factors might interact with chromosomal global regulator that in turn modify the expression of a large set of chromosomal genes. This hypothesis is reinforced by our global transcriptomic data showing that the expression of several regulators such as RpoS, LetE, LetS, Hfq, HU, Fis2 are all repressed in plasmid-integrated strains. RpoS, LetE, LetS and Hfq all interact in the same regulatory cascade that governs the differentiation and biphasic *Legionella* life cycle and thus are known to regulate hundreds of genes²⁸⁷. Similarly, both HU and Fis2 are nucleoid-associated proteins known to bind DNA and modulate the expression of a large set of genes in bacteria²⁸⁸

Bacterial plasmids that control chromosomally-encoded gene have already been reported in several bacterial pathogens ^{289–294}. For instance, in *Rhodococcus equi*, a facultative intracellular pathogen it has been suggested that following the acquisition of the virulence plasmid, the two plasmid-transcriptional regulators VirR and VirS integrated themselves into the chromosomal regulatory network and thus modified the expression of a large number of chromosomal genes that ultimately improved the intracellular growth capacities of the pathogen ²⁹⁵. In *Acinetobacter baumannii*, it was shown that large conjugative plasmids like pAB5, affect pili and exopolysaccharides biosynthesis implicated in adherence and biofilm formation and this modulation of chromosomally-encoded factors account for differential outcomes in *Acinetobacter* isolates, such as an improved survival in urinary tract but attenuated-virulence in pulmonary model ²⁹⁶. In another study, it was shown that pAB5 also encodes a regulator that inhibits the activity of the chromosomal type-IV secretion system (T6SS), a bacterial weapon used to kill non-kin competitors, including the putative plasmid donor or recipient cells ²⁹⁷. Therefore, in *A. baumannii*, T6SS activity is detrimental to the conjugative transfer of multidrug-resistant (MDR) plasmids. By silencing chromosomal T6SS-encoding genes, pAB5-like plasmids thus ensure their dissemination within bacterial population ²⁹⁸. In each case, these studies highlight the major role played by plasmids in the enhanced growth and adaptation of host bacteria to specific niches, and how plasmids, by enhancing bacterial survival, ultimately increase their own fitness and maintenance.

Other striking results from the experimental evolution strategies we developed, is the diversity of molecular mechanisms that *Legionella* has evolved to adapt to the particular selection regime, and also the rapidity with which these mechanisms have occurred. Indeed, in both cases AX and BX lines, the excised form of pLPP was detected as early as 10 passages (around 80 generations) suggesting that this mutation has a strong beneficial impact on fitness. Indeed, in the Long-Term Evolution Experiment (LTEE) conducted by Lenski and co-workers, it has been shown that the first fixed mutations in the *E. coli* lines conferred strong fitness gain, whereas those appearing later have a lesser impact in fitness, although both adaptation and divergence can continue indefinitely ^{21,35}.

Another common feature between our findings and those of LTEE or others EE strategies, is the high level of genetic *parallelism* and *convergent evolution* between *populations* evolving in the same environments. The convergence of mutations is particularly striking notably in

BX lines. Indeed, these populations have accumulated multiple mutations in the same global regulators, namely LetA/LetS and RelA, and this appeared to be correlated to a severe parallel decrease in virulence in all lineages. Consistent with this hypothesis, it has been shown that a *letA* mutant does not replicate efficiently in amoebae^{183,273,299}. However, it is worth to notice that a *L. pneumophila relA* mutant shows no intracellular growth defect. Nevertheless, as these regulators belong to the same virulence regulatory network, it is possible that the accumulation of several mutations simultaneously in both regulators could lead to severe dysfunction and ultimately a decreased virulence. The impact of these mutations on virulence traits, alone or in combination, is currently under investigations. According to these results, it is reasonable to think that *Legionella* has developed convergent mechanisms for deactivating its virulence traits under environmental conditions where it is no longer useful to express these costly functions. The reduced expression of these “non-essential” functions would thus alleviate the metabolic burden on the bacteria, thereby resulting in better growth.

Interestingly, spontaneous mutants in LetA/LetS TCS homologue have already been found in *Pseudomonas* species and are heavily documented³⁰⁰⁻³⁰³. In all these cases, spontaneous mutants have appeared under nutrients rich conditions, as in our experimental design with *Legionella* BX lines. The ecological reasons for the appearance of these spontaneous mutants are not yet clear, however, it has been argued that the functions controlled by this TCS become of little use when bacteria grow in nutrient-rich axenic media. This hypothesis may be true but probably too simple. For instance, experimental evolution conducted with *Pseudomonas fluorescens* have also shown that bacteria accumulated mutations almost exclusively in the GacA/GacS TCS to compensate the reduced fitness due to the cost of megaplasmid carriage^{304,305}. Therefore, GacA/GacS compensatory mutations would not only increased fitness but also would allow for plasmid domestication. It is tempting to speculate that similar mechanisms occurred in our BX lines that have evolved under rich nutrient conditions and in which the status of the plasmid is both integrated and episomal at the same time.

Altogether, our results and those of other groups, highlight the underestimated importance of the plasmids in the evolutionary mechanisms of bacteria. They also clearly suggest that it exists an interdependent relationship between plasmids and their hosts, with on one hand, an adaptive manipulation of chromosomal genes by the plasmid and on the other hand, a generic bacterial response to facilitate plasmid acquisition and maintenance.

2.4. Conclusion

At the microbiology level, evolutionary mechanisms leading to both the emergence of pathogenic strains and the associated evolution of virulence are still poorly understood. The strong restructuration of the environment of microorganisms by human activities needs to be integrated into a model that will allow predictions of how these activities can influence microbial evolution. Such objectives make *L. pneumophila* an especially interesting model organism: the development of man-made hot water systems has seemingly played a crucial role in the emergence of its pathogenicity. Elucidating the mechanisms of host spectrum evolution and the molecular targets of selection in *L. pneumophila*, will shed light on the interplay between environmental changes and pathogen evolution.

3. *Legionella norrlandica*, an environmental species of *Legionella* to study transfer of virulence genes

3.1. Introduction on HGT and natural transformation in *Legionella pneumophila*

Bacterial genomes are shaped by gene exchanges, chromosomal rearrangements and the activity of mobile genetic elements³⁰⁶. As discussed in the part I of this manuscript, one of the preferred ways of *Legionella* to acquire new and potentially virulence-linked genes, is through horizontal gene transfer. Horizontal gene transfer (HGT) being a process by which bacteria can acquire DNA from the environment or other cells through three main distinct events: transduction (mediated by a bacteriophage), conjugation (transfer between bacteria of a plasmid mediated by a conjugative pilus) or natural transformation (acquisition of exogen DNA mediated by transformation pili)³⁰⁷. Transfer of genetic material can also occur to a lesser extent through vesicles or special structures called nanotubes³⁰⁸.

Natural transformation and associated DNA uptake mechanisms are conserved features of HGT in bacteria with more than 80 experimentally confirmed transformable species^{309,310}. Internalization of exogenous DNA starts by the binding of the DNA with the type IV pilus (T4P)³¹¹. The literature suggests that once DNA is attached to the tip of the T4P, its internalization into the periplasm is allowed by the retraction of the T4P³¹². Once DNA reaches the periplasm, it accumulates and binds to the ComEA protein. Studies have shown that ComEA not only passively binds to the DNA but rather actively promotes its passage through the outer membrane by exerting a ratcheting force³¹³. DNA is then converted into single-stranded DNA (ssDNA) and is translocated into the bacterial cytoplasm through the ComEC inner membrane channel³¹⁴. In the cell, ssDNA is protected from degradation by binding to two proteins, SsbB and DprA^{315,316}. The latter is able to interact with the DNA recombinase RecA who will facilitate the integration of the ssDNA into the bacterial chromosome, providing that the two molecules of DNA possess homologous regions³¹⁷. In Gram-negative bacteria, another competence protein named ComM is also implicated in the process of homologous recombination in naturally competent Gram-negative bacteria³¹⁸. In *L. pneumophila* like in most naturally competent bacterial species, competence genes are not constitutively expressed but in response to environmental cues³¹⁹. To reach its competence state, *L. pneumophila* needs to be cultivated at 30°C and be at the transition

between its exponential and stationary growth phases (**figure 29**) (which can be roughly compared to the replicative and transmissive phases of infection, respectively)³²⁰. The role of transcriptional activators was thought to be crucial and pivotal to start the transcription of genes implicated in the activation of the competence state in a vast majority of bacteria³²¹. However, a recent study by the team of Xavier Charpentier, unmasked a new and revolutionary mechanism of post-transcriptional regulation of the competence genes in *L. pneumophila*³²². In this study, scientists describe a model in which a small RNA (sRNA) called RocR associates with an RNA chaperone, RocC, to negatively control the stability of the competence genes mRNA (**figure 29**).

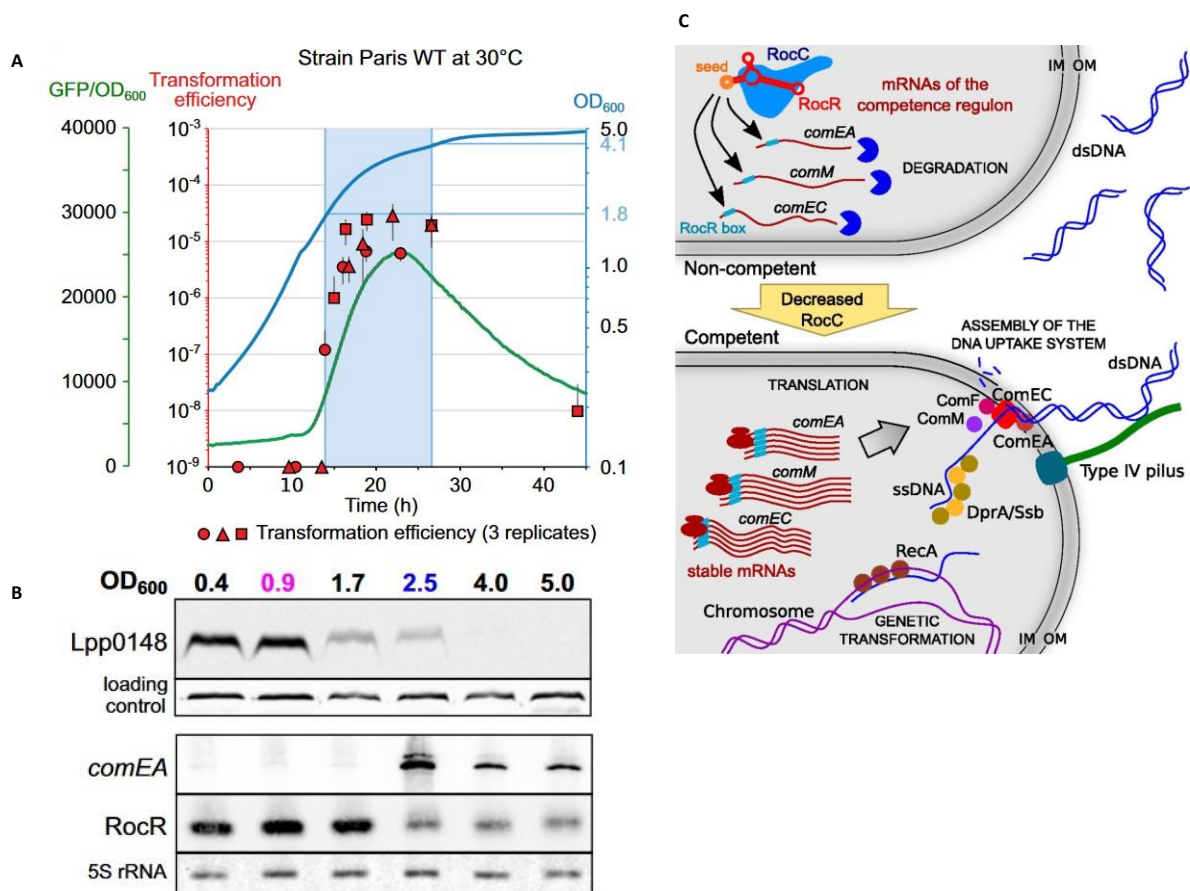


Figure 29: Model of regulation of natural transformability in *L. pneumophila*

The *L. pneumophila* Paris wild strain transiently develops competence for natural transformation during growth at 30°C. **A:** Expression of *comEA* was followed using a *comEA*-GFP transcriptional fusion carried by pXDC91 (GFP/OD₆₀₀, green line) and natural transformability (red circle, triangle and square) determined at different time points during growth (OD₆₀₀, blue line). Error bars on natural efficiency data represent SE. **B:** Expression of Lpp0148 (RocC) and RocR decrease at the onset of the transformability phase at 30 °C. Expression of Lpp0148 was analyzed by Western blot, and expression of *comEA* and RocR was determined by Northern blot analysis. A cross-reacting band and the 5S rRNA were used as loading controls for the Western blot and Northern blot,

respectively. **C:** When cells are noncompetent, RocC stabilizes RocR and this ribonucleoprotein complex can bind the mRNAs of genes of the competence regulon via a conserved 6-nt sequence called RocR box and promote their degradation. Under competence-inducing conditions (for example, at the end of the exponential phase at 30 °C), RocC expression decreases. This triggers the destabilization of RocR and thus the stabilization of the mRNAs of the competence genes. These are then translated, the DNA uptake system is assembled, and horizontal gene transfer by natural genetic transformation can occur. IM, inner membrane; OM, outer membrane. Adapted from Attaïech *et al.* 2016

In bacteria in a non-competent state, RocC is constitutively expressed and its association with RocR helps stabilize the sRNA. The ribonucleoprotein complex targets a region in the 5'UTR of mRNAs of competence genes called the RocR box, promoting their degradation. As the bacteria enters its competence state, RocC expression decreases, destabilizing RocR and eventually stabilizing the mRNAs of competence genes. mRNAs can be translated and the DNA competence complex can be assembled (**figure 29**). RocR is conserved among the *Legionella* genus with a substantial sequence identity percentage ³²², which suggests a shared mechanism of competence regulation in all *Legionella* species.

L. pneumophila can acquire genes by using all of the above-described systems, facilitating its adaptation to new environments or new hosts. Indeed, comparative genomics studies have demonstrated that *Legionella* species have acquired many eukaryotic-like proteins that likely modulate specific host functions to allow intracellular survival and replication in eukaryotic host cells ¹¹¹. This repertoire is highly variable among the different *Legionella* species suggesting several gain and loss events during the evolution of the genus ²¹¹. Thus, the ability of *Legionella* to replicate in macrophages may have been achieved by the fortuitous acquisition of different virulence properties initially required to colonize a specific natural host, such as the amoeba. Then, due to the high conservation of key cellular and signaling pathways in professional phagocytes, such as amoebae and human macrophages, different combinations of virulence traits may have allowed some *Legionella* species to infect higher eukaryotic cells by chance. This raises the question of the evolutionary capacities of species isolated from the environmental reservoir. Can environmental *Legionella* species evolve and become a new human pathogen like *L. pneumophila*? And what are the molecular mechanisms underlying the emergence of virulence?

In this context, we aim to set up a novel experimental evolution strategy consisting to evolve the closest relative of *L. pneumophila*, *Legionella norrlandica*, an environmental species, in

different eukaryotic and in presence (or not) of genomic DNA from *L. pneumophila*. *L. norrlandica* was found in a wood processing plant in Sweden³²³, capable of infecting different species of amoebas³²⁴ but unknown to cause legionellosis to humans. The objective of this strategy is to determine how *L. norrlandica* evolve and adapt to eukaryotic hosts, in particular the human macrophages, and whether under this selective pressure it is capable of acquiring exogenous DNA that could potentially improve its fitness. The prerequisites for this type of long-term evolution experiment are first to evaluate the intracellular replication capacities of *L. norrlandica* within different eukaryotic host, which must be sufficient to allow the transfer of a fraction of the population at each passage; and secondly, to characterize the competence state of *L. norrlandica* and test its natural transformation capacity. According to whole genome sequencing (this study and Kevin Picq PhD these, unpublished), *L. norrlandica* in particular possesses the genes encoding the Dot/Icm Type IV secretion system as well as those encoding for type IV pili and the competence genes needed for natural transformation.

3.2. Results

3.2.1. Monitoring intracellular replication ability of *L. norrlandica*

As extensively discussed in this manuscript, *Legionella* species are known to actively replicate inside amoeba cells in the environment. We naturally wanted to confirm results already published in the sole study relative to *L. norrlandica* and test the replicating capacity of this bacteria in A549 human lung epithelial cells, a permissive cell type for *L. pneumophila* infection³²⁵.

We carried out an experiment in which we infected *Acanthamoeba castellanii* and A549 human lung epithelial cells by either *L. norrlandica* or *L. pneumophila*, to be able to visually compare the replication capacity of the two species in the two cell types. Different strains of both bacteria are used, harboring either the pXDC50 or p3041 plasmids. Both plasmids are coding for mCherry fluorescent protein under Ptac control and are culture under chloramphenicol or gentamicin selection, respectively. Both plasmids are independently used to transform *L. pneumophila* and *L. norrlandica* to rule out fitness cost of the plasmid on the bacteria replication capacity. mCherry under Ptac is used to track fluorescence increase, which is directly linked to the increase in number of bacteria.

Results depicted in **figure 30** show a strong increase in fluorescence for *L. pneumophila* in amoebae starting at 17h post-infection (**figure 30A**) or 25h post-infection (**figure 30B**) depending on the plasmid used. For *L. norrlandica*, the fluorescence increase is milder and starts later than for *L. pneumophila* (45h and 70h post infection for pXDC50 and p3041 plasmids, respectively). There seems to be a higher cost in fitness brought by the p3041 plasmid compared to pXDC50, which is why we chose to continue using pXDC50 in all further experiments. Figure 30C confirms replication capacity of *L. pneumophila* in A549 cells and especially shows that *L. norrlandica* is incapable of infecting this cell type. Overall, these results show a decreased capacity of replication for *L. norrlandica* compared to *L. pneumophila* in *A. castellanii*, both in terms of start and rapidity of infection and amount of replication cycles and seem to suggest a narrower host range for *L. norrlandica* compared to *L. pneumophila*.

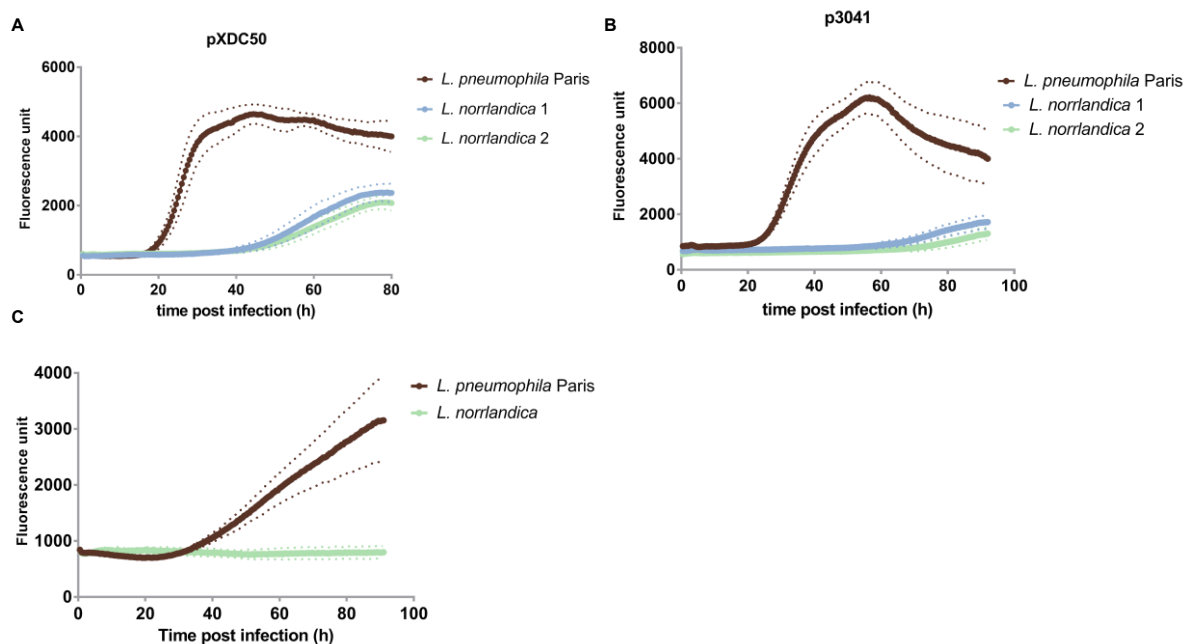


Figure 30: Growth kinetics of *L. norrlandica* in amoebae and human lung epithelial cells.

Intracellular replication is monitored by measuring the fluorescence emitted by the mCherry protein. **A:** 2 clones of *L. norrlandica* and 1 clone of *L. pneumophila*^R containing the plasmid pXDC50 (mCherry⁺, Chl^R) are cultured in the presence of amoebae in the depleted medium PYs+Chl+IPTG (Chl at 10µg/mL, IPTG at 0.25mM). **B:** 2 clones of *L. norrlandica* and 1 clone of *L. pneumophila* containing the plasmid p3041 (mCherry⁺, Genta^R) are cultured in the presence of amoebae in the depleted medium PYs+Genta+IPTG (Genta at 10µg/mL, IPTG at 0.25mM). **C:** One clone of *L. norrlandica* and 1 clone of *L. pneumophila* with pXDC50 plasmid are used to infect human lung epithelial cells A549

3.2.2. Study of natural transformation in *L. norrlandica*

3.2.2.1. Natural transformation assay in *L. norrlandica*

To test the transformation capacity of *L. norrlandica*, we chose a transforming DNA which, when integrated, confers a selective advantage to the transformers, here the resistance to streptomycin. Resistance to streptomycin is provided by the mutated *rpsL* gene. The wild type *rpsL* gene encodes the ribosomal protein S12 of subunit 30S of ribosome³²⁶. The mutation of the S12 protein makes it possible to reduce the interaction of the 30S subunit with streptomycin and thus to confer resistance. To place ourselves in the best conditions of transformation and facilitate homologous recombination, we have chosen a DNA sequence that is 4 kb long and homologous to the genome of the recipient bacterium except at the level of this mutation which confers resistance. Different DNA types (all conferring resistance to streptomycin) are tested: genomic DNA (gDNA), PCR product and plasmid DNA. We first extracted gDNA from a streptomycin-resistant *L. norrlandica* strain (DNA was concentrated to 682 ng/ μ L DNA). Then we performed a PCR of the mutated *rpsL* region (primers, see Table 1) which gives an amplicon of 4257 base pairs. For plasmid DNA, it was necessary to clone the 4kb region amplified by PCR in a vector. We tried 2 different vectors (pJet1.2 or pBluescript II SK), but we failed to have an *E. coli* transformant regardless of the vector used. Natural transformation tests with *L. norrlandica* are therefore performed only with the PCR product and purified genomic DNA.

L. pneumophila acquires its competent state at the mid-exponential phase until early stationary phase ($1.5 > DO_{600} > 2.5$), when cultivated at 30°C (**figure 29**). As the species genetically the closest to *L. pneumophila*, we postulated that *L. norrlandica* would require the same culture conditions to reach the competence state for natural competence (i.e., culture in a rich medium with agitation at 30°C). We recovered bacterial cells at different stages of their growth ($DO=0.7, 0.9, 1.1, 1.6, 3.7, \text{ and } 5$). The bacteria were then incubated either with gDNA from a Streptomycin^R *L. pneumophila* clone or PCR product – both conferring resistance to streptomycin – or without DNA (negative control). Bacteria were then spread over BCYE plates, with or without adjunction of streptomycin to select for resistant clones.

Transformation rate was then be calculated as follows:

$$\frac{\text{Number of streptomycin-resistant clones}_{+gDNA}}{\text{Total number of bacteria}} \text{ or } \frac{\text{Number of streptomycin-resistant clones}_{+PCR \text{ product}}}{\text{Total number of bacteria}}$$

and then compared to $\frac{\text{Number of streptomycin-resistant clones}_{\text{Negative control}}}{\text{Total number of bacteria}}$

Transformation rates are very low ($\approx 10^{-7}/10^{-8}$) and correspond to the frequency of mutant appearance by spontaneous mutation (**figure 31**). When transformation rates are higher ($\approx 10^{-6}$), they nevertheless appear to be non-statistically significant compared to rates of spontaneous mutation in negatives controls who did not receive transforming products.

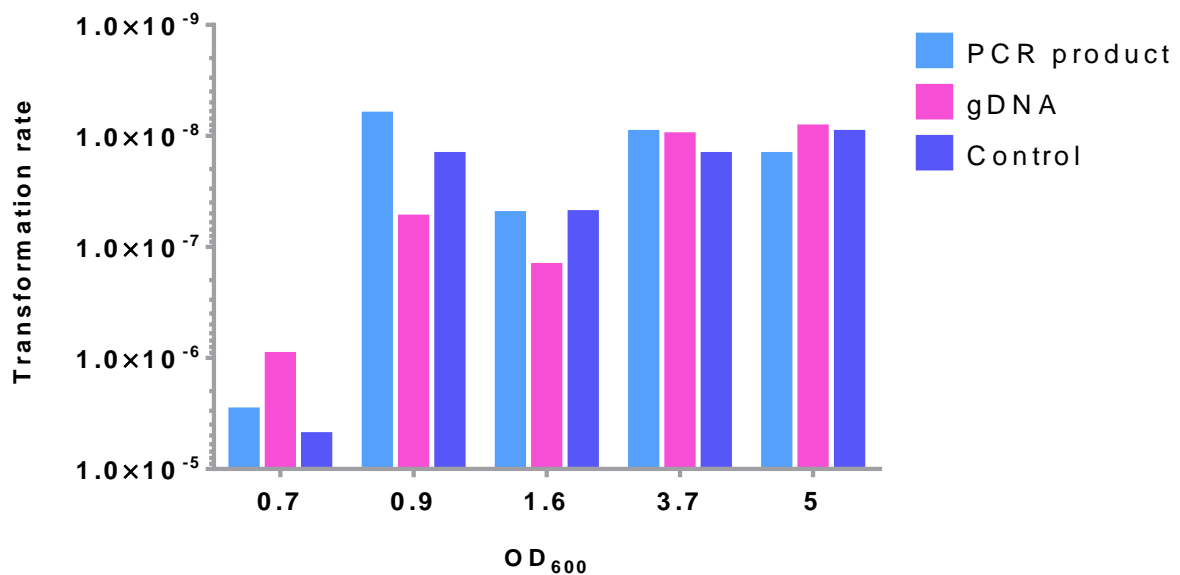


Figure 31: Tests of natural transformation in *L. norrlandica*.

L. norrlandica was cultivated at 30°C and a subpopulation harvested at different OD₆₀₀, corresponding to different growth stage. Harvested bacteria were cultivated in contact with either gDNA from a Streptomycin^R *L. pneumophila* clone or *rpsL* PCR product. Cultures were then spread on CYE or CYE+Strep plates and UFC numbered. Transformation rate was calculated by dividing the number of bacteria that grew on CYE+Strep plates by the number of bacteria on CYE plates. Control are *L. norrlandica* cells cultivated without foreign DNA

The *rpsL* gene mutation conferring streptomycin resistance is not a good marker for measuring the natural transformation frequency of *L. norrlandica*, which appears to be low. Therefore, we decided to use another marker gene conferring resistance to kanamycin (*aphA3*), an antibiotic for which the spontaneous resistance rate is lower ($<10^{-9}$).

3.2.2.2. Improvement of transformation assays using kanamycin resistance conferred by *aphA3* gene

In order to test the natural transformation capacity in more stringent conditions (i.e., with less spontaneous mutation arising), we proceeded to assemble a Kan^R cassette, containing the *aphA3* gene flanked by regions homologous to the chromosome of *L. norrlandica*.

In the laboratory, we had identified a region called “neutral” in the *L. pneumophila* chromosome where insertion can be done without much impact on fitness or global gene expression. This ≈4kb region is located between position 2 300 387 and 2 304 530, encompassing tRNA coding sequences.

By homology with *L. pneumophila*, a similar region was detected in the genome of *L. norrlandica* between positions 1 006 683 and 1 010 836. We chose this region to insert the *aphA3* gene.

An overview of the protocol for amplification of the Kan^R cassette flanked with regions homologous to the neutral site in *L. norrlandica* is presented in **figure 32**.

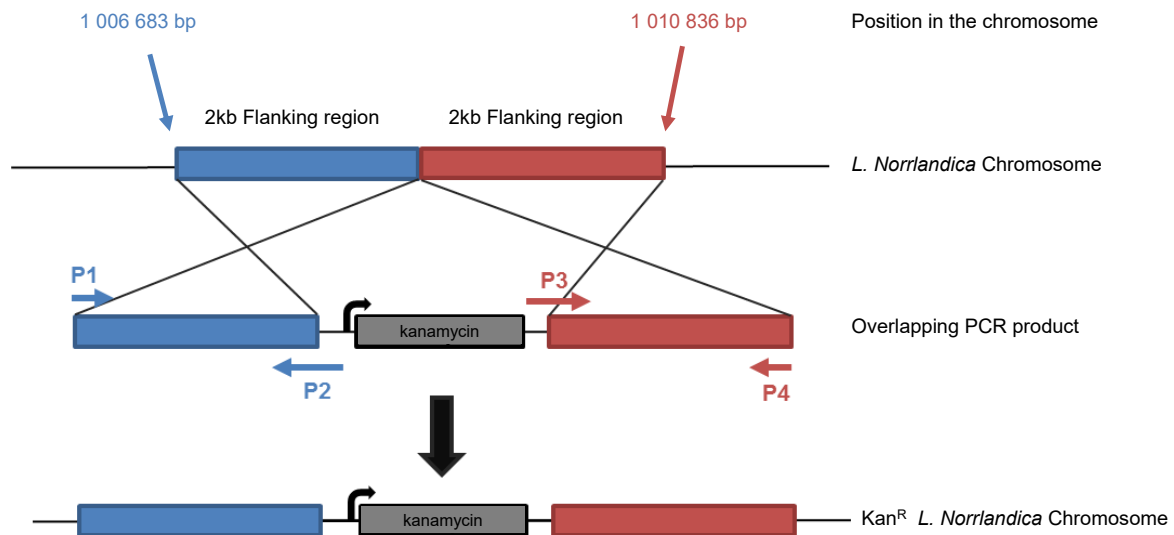


Figure 32: Construction of the Kan^R cassette

First, an overlapping PCR was performed using on the one end, P1 and P2 primers to amplify upstream region of the neutral site with 20 nucleotides complementary of the 5' region of the Kan^R cassette, and on the other hand, P3 and P4 primers to amplify the downstream region of the neutral site with 20 nucleotides complementary of the 3' region of the Kan^R cassette. Then, PCR products are mixed and amplified using the P1 and P4 primers to create the Kan^R used as transformation product

The Kan^R cassette was used as transformation product to test natural transformation in *L. norrlandica* with conditions similar to the one previously described with the *rpsL* gene. Unfortunately, results were not usable and will need to be repeated, maybe using different culture conditions.

3.2.2.3. Regulation of competence development in *L. norrlandica*

Our previous experience let us think that *L. norrlandica* is not competent for natural transformation under the different conditions tested (30 ° C, rich medium, different stages of growth). However, a co-occurrence analysis on the competence genes showed that *comEA* (essential to natural transformation), *comR*, *rocC* and the small RNA regulator RocR are all present in most *Legionella* species, including *L. norrlandica* (**figure 32**). We then sought to understand the underlying regulatory mechanisms of competence state acquisition of *L. norrlandica*. To do so, we performed a Northern Blot analysis directed towards the mRNA of *comEA* and the competence genes repressor (i.e., the sRNA RocR). We grew *L. norrlandica* at 30°C in rich liquid medium and sampled bacteria at various growth phases (corresponding to OD₆₀₀= 0.6, 1.1, 1.4, 2, 2.7 and 4) on which we tested the presence of the two RNAs (**figure 33**).

The Northern Blot analysis does not permit to reveal any bands related to *comEA* mRNA for any of the OD₆₀₀ tested. Conversely, RocR has a strong and constant expression throughout all phases of growth. This would therefore support the model described in *L. pneumophila* where competence genes, including *comEA* are subject to a strong repression by RocR. Our experiment also confirms the previous results suggesting that the regulation mechanisms of the competence state in *L. norrlandica* are different from the ones described for *L. pneumophila*.

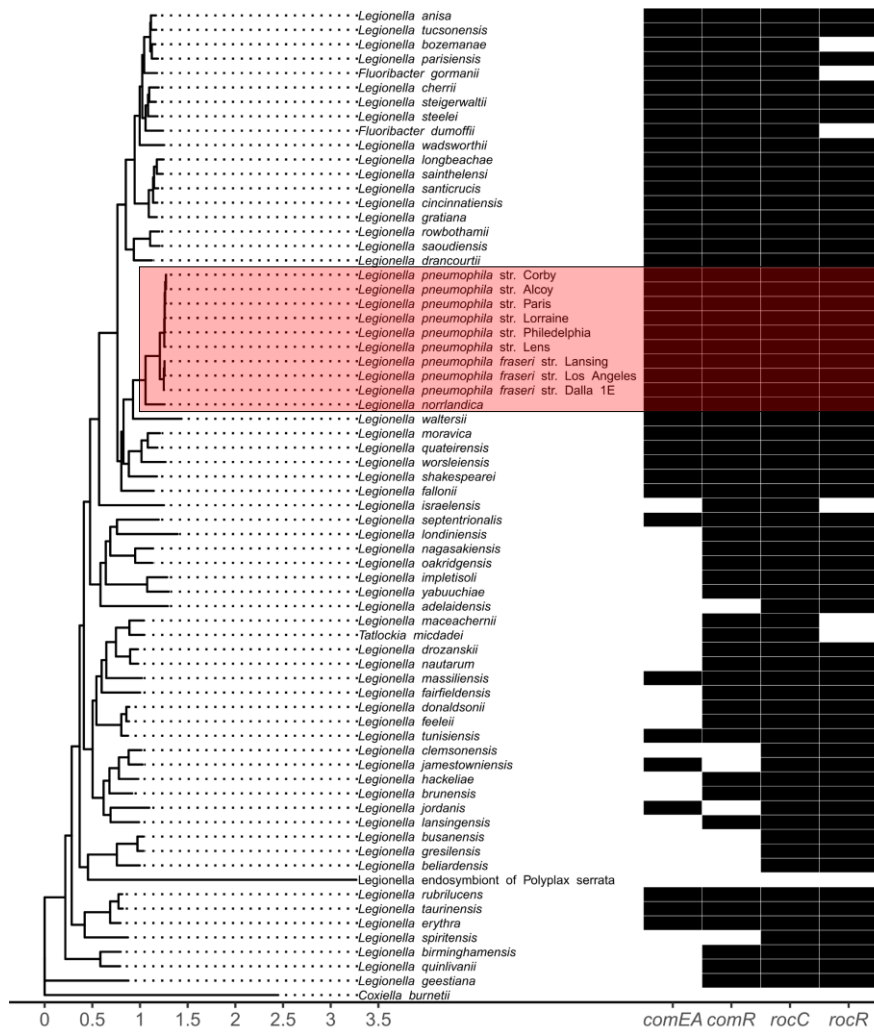


Figure 33: Co-occurrence tree of competence genes within the *Legionella* species

A species tree of *Legionella* species was constructed (with *Coxiella* as an external clade) to show co-occurrence of four of the genes implicated in the competence regulation. The red box is used to highlight the clade grouping *L. pneumophila* strains and *L. norrlandica*.

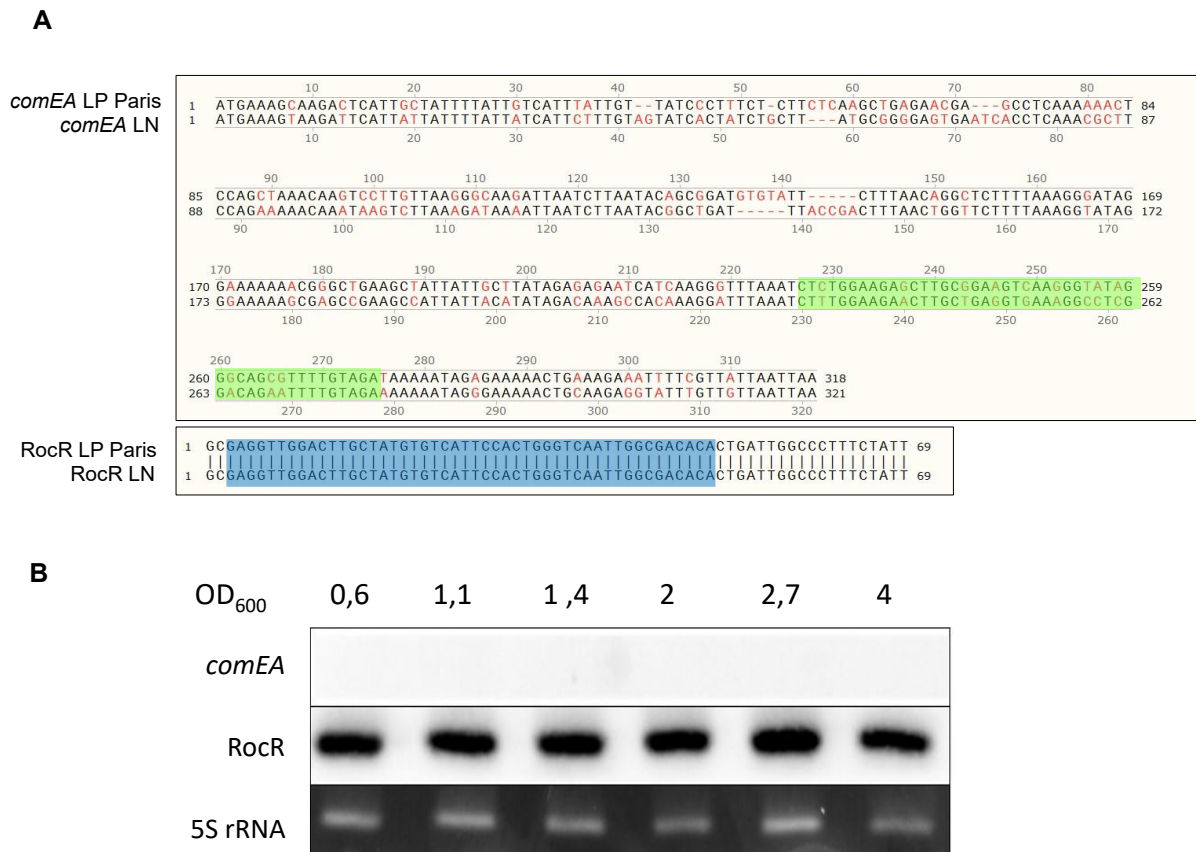


Figure 34: Analysis of the expression of sRNA RocR and mRNA of *comEA* by Northern blot.
A: Alignments of *comEA* genes (upper panel) and RocR sRNA (lower panel) from *L. pneumophila* Paris and *L. norrlandica*. Green and blue boxes represent genome equivalent of RNA location for hybridization of Northern blot probes for *comEA* and RocR, respectively. **B:** Northern blot analysis of the expression of RocR sRNA and *comEA* mRNA at various growth time points (OD₆₀₀).

3.3. Discussion

Several recent studies tend to show that the numerous interdomain and interbacterial HGT as well as the combined selective pressures of its diverse natural hosts shape the genomes of *Legionella* species and drive for the emergence of human pathogen from environmental bacteria ¹¹¹. *Legionella pneumophila* are such bacteria that infect various species of protozoan hosts in the environment. They also evolved to be able to infect human macrophages accidentally, causing legionellosis, a sometimes-fatal form of pneumonia. Mechanisms that led to the ability of these bacteria to invade a wide range of hosts, as well as the predominance of some strains of *L. pneumophila* in clinical settings remain still largely unknown.

L. norrlandica is a species of *Legionella* recently discovered in a wood processing plant in

Sweden. They happen to be the closest relative to *L. pneumophila* but have not been identified in any epidemic outbreak. Yet genome sequencing revealed that *L. norrlandica* possesses all major virulence mechanisms known to be present in *L. pneumophila*, including the type 4 secretion system. Only one study confirmed *L. norrlandica* is able to infect *A. castellanii*. Our objectives were to test the bacteria competence for natural transformation and further characterize the replication capacity of *L. norrlandica* towards *A. castellanii*.

We confirmed results obtained by Dobrowsky and collaborators³²⁴ and replicated *L. norrlandica* in the amoeba *A. castellanii*. For this, we used two strains of *L. norrlandica* each with a distinct plasmid, pXDC50 or p3041, both harboring the gene coding for the same fluorescent protein but selected by two different antibiotics, Chloramphenicol or Gentamycin, respectively. Both strains of *L. norrlandica* were able to replicate in amoebae but it seems p3041 has a more important cost on the bacteria's fitness. This is probably due to the cost of the gentamycin coding gene vs. the chloramphenicol coding gene as the pXDC50 and p3041 plasmids backbones are otherwise the same. Plasmid pXDC50 would be our preferred choice to test *L. norrlandica* intracellular replication capacity. Comparing the replication kinetics of *L. norrlandica* with those of *L. pneumophila*, there is a clear difference between the two species. Even if we did show that *L. norrlandica* replicates in, they replicate less and at a much slower rate than *L. pneumophila*. We were also able to show that *L. norrlandica* does not replicate inside human lung epithelial cells, contrary to *L. pneumophila*. The next natural step would then be to test the replication capacity of *L. norrlandica* in other cells, known to be permissive to *L. pneumophila*, such as human lung macrophages U937 or THP-1 human monocytic cell line. A recent study²¹¹ showed that among 47 different species of *Legionella*, 23 (60%) were able to replicate in THP-1. However, the vast majority of these strains have never been described as causing cases of Legionellosis. These strains are also vastly dispersed in the phylogenetic tree of the *Legionella* genus. Authors of the study make the suggestion from these results that the ability to infect human cells was acquired several times in the evolution of the genus. By comparing the genomes of these different species, no group of effectors or virulence factors common to all these strains was identified. Thus, different random combinations of effectors or virulence factors could allow certain environmental strains to infect human cells.

As part of our second objective, we were able to show through transformation assays that *L. norrlandica* is not naturally competent to acquire foreign DNA when grown at 30°C, contrary to *L. pneumophila*. Indeed, there is no significant difference in the rate of appearance of streptomycin resistance compared to negative control, at each of the tested growth phases and independently to the DNA source used (gDNA or PCR product). This result can be correlated to the Northern Blot analysis showing constitutive, strong expression of RocR and no mRNA for *ComEA* throughout all tested growth phases. This could imply two hypotheses. On the one hand, it would appear that RocR acts the same way in *L. norrlandica* as it does in *L. pneumophila*, *i.e.*, by repressing the transcription of competence genes. To test this, it would be interesting to delete RocR in an otherwise wild type *L. norrlandica* strain and perform the same transformation experiments as described above. If RocR acts the same way in *L. pneumophila* as in *L. norrlandica*, the repression of competence genes would be lifted, allowing for a much higher transformation rate compared to wild type *L. norrlandica*. On the other hand, although phylogenetically close, *L. pneumophila* and *L. norrlandica* do not necessarily share the same competence regulation pathways. For instance, bacteria from the *Streptococcus* genus have developed strong phenotypic heterogeneity in regards to transformation frequencies due partly to fluctuations in the activation of the σ^X alternative sigma factor (partly due to environmental cues), considered as the master regulator of the competence state in *Streptococcus* bacteria³²⁷. It would be interesting to test other culture conditions (temperature, culture media, presence of amoeba...) to see if competence state is differently triggered in *L. norrlandica* compared to *L. pneumophila*. Another explanation of the different transformability between *L. pneumophila* and *L. norrlandica* could come from the recent work of Xavier Charpentier's team²⁶⁹. In their study, they elegantly demonstrated the impact of a plasmid-borne small RNA (RocRp) present in non-competent clades of *L. pneumophila*. Much like RocR, that we previously characterized in this study, RocRp is acting as a strong repressor of transcription of competence genes, with the difference that it is expressed in the mid/late growth phases of *L. pneumophila*, following the decrease of expression of RocR. Even if RocR's expression does not seem to diminish in *L. norrlandica*, thus not requiring RocRp to take over the competence-switching-off role, it could be interesting to check for the presence and explore the role of conjugative elements in *L. norrlandica*.

Another idea could be to test alternate source of transforming DNA. Indeed, the mutated *rpsL* gene used in this study may not be the adequate tool to test transformation frequency due to the high prevalence of spontaneous mutations apparition, giving resistance to streptomycin. We actually counted almost the same number of streptomycin-resistant colonies with or without transforming DNA. Some clones were able to grow on streptomycin-containing medium in our negative control condition (bacteria without DNA), with a frequency of around 1.10^{-7} - 1.10^{-8} , which is what is observed in *Mycobacterium smegmatis*³²⁸ for instance and which is significantly lower than the natural transformation rate of *L. pneumophila* (around 1.10^{-5} at OD₆₀₀=3 and 30°C)³²⁹. To obtain better transformation efficiency and free us of spontaneous mutations clouding results, we constructed a Kanamycin resistance cassette containing the *aphA3* gene encoding an aminoglycoside 3'-phosphotransferase conferring resistance to kanamycin. Indeed, rate of onset of spontaneous mutation for kanamycin resistance is lower than that of streptomycin. This Kan^R cassette will need to be used to test homologous recombination *via* natural transformation in *L. norrlandica*.

Noteworthy, of all bacteria belonging to the *L. pneumophila* clade (*L. fallonii*, *L. shakespearei*, *L. worsleiensis*, *L. quateirensis*, *L. moravica*, *L. waltersi*, *L. pneumophila*, *L. norrlandica*), *L. norrlandica*, though branching closest to *L. pneumophila*, seems to have less in common with *L. pneumophila* than the other species, at least on the tested features. For instance, *L. norrlandica* has the smallest genome, with the smallest GC% and the smallest number of singletons (except for *L. pneumophila Philadelphia*, *Lens*, *Lorraine*, *Paris* and *HL06041035*). The acquisition of Eukaryotic domains seems also to have been reduced in *L. norrlandica* compared with other species of the clade, with the smallest number of Ankyrin motif, the only species with no F-box domain, no SET domain, or no Sec-7 domain. On the other hand, it is the only species of the clade with mitochondrial substrate/solute carrier and T-complex 10/11. Altogether, this could point to a modified ability of *L. norrlandica* to acquire foreign DNA compared to *L. pneumophila*, which highlights the importance of continuing experiment on this species.

4. Global Discussion and future prospects

Legionella pneumophila is the main species implicated in clinical settings of Legionnaires disease, of all the 60+ species described to that day. This prevalence cannot be fully explained with current results of comparative genomics, owing to the fact that species share a rather small core genome and a plethora of variable effector proteins. The emergence of *L. pneumophila* as human pathogen from the environment is also one of the big questions of *Legionella*-based research. How do the bacteria evolved to infect humans? What are the predeterminants to this evolution?

Finding the answers to these questions is crucial not only for *L. pneumophila* but also for the fields of evolutionary microbiology and public health in general. Indeed, pathogens often emerge or re-emerge owing to evolutionary changes enabling them to colonize new host species or to better spread from one host to another. These questions have been lately exposed to the general public with the Covid-19 pandemic. This disease was due to a type of coronavirus (SARS-CoV-2) that emerged from an unknown environmental reservoir, probably bats or pangolins. The reason why it crossed the species barrier to infect humans remains unknown, but it is likely due to mutations in the genome allowing it to either replicate more intensively within humans or to better spread between hosts. We witnessed the formidable mutational capacity of the virus during the pandemic with the emergence of multiple variants, some of them deemed 'of concern' by the WHO (Alpha, Beta, Gamma, Delta and Omicron). These variants either exhibited higher transmissibility, higher virulence, or high resistance to vaccines.

With the growing world population and increase in intercontinental exchanges, it is pivotal to better apprehend these mechanisms underlying the rise of new pathogens.

My PhD project is intimately embedded into this global context and may seem innovative as the project had been initiated long before the questions of pathogen emergence were known by the general public.

My objectives were to gain knowledge on how host spectrum of *L. pneumophila* evolves, what are the impacts of environmental factors such as host pressure, what are the costs and pace

of this evolution and how can bacteria in general, and *L. pneumophila* in particular, evolve from an avirulent bacteria to be the cause of a potentially lethal form of pneumonia.

To answer these questions, I developed experimental evolution (EE) strategies capitalizing on previous work conducted in the laboratory. This almost magical experiment seemed to be insanely adapted to the questions posed by my PhD subject. EE can allow us to evolve organisms in control environment, but also rewinding the tape by experimenting on frozen populations at different steps of evolution, and unveiling genomic changes underlying phenotypic modification observed, by performing massive sequencing and transcriptomic analysis of the populations.

Before I started my PhD, six independent lineages had been generated from the clinical strain *L. pneumophila* Paris by a mutation accumulation (MA) experiment. Briefly, clones were independently and in parallel propagated on standard agar medium for 700 generations. This type of experiment allows for the accumulation of neutral to beneficial mutations without (much) *a priori*. I started by characterizing mutants resulting from this experiment at the phenotypic level. I was amazed to see that all clones from the six lineages had greatly reduced their capacity to infect both their natural hosts (*Acanthamoeba* spp.) and their accidental ones (human pulmonary macrophages U937 cells). We then proceeded to characterize these clones at the genetic level by sequencing 3 of the 6 lineages with Illumina technology. I was a bit disappointed when results showed that Paris 4 had no mutations, Paris 5 had only 2 SNPs (one of which was synonymous) and 4 SNPs in Paris 6. No genetic mutations could explain the observed phenotype at least in Paris 4 lineage. This also contrasts with what has been described in other EE experiments, where parallel evolution can be observed between evolved lineages¹. We thus tried a different approach by re-sequencing the clones using the PacBio technology. This would allow to account for larger chromosomal rearrangements, and indeed, we detected the integration of the pLPP plasmid (a natural, widely distributed plasmid in the *L. pneumophila* ST1 cluster²) into the chromosome of the three lineages! This integration was facilitated by an Insertion Sequence (IS) present both on the plasmid and on the genome of the bacteria. To assess the dynamics of this event, we designed a PCR experiment to detect either the integrated or episomal form of the pLPP in evolved clones, at different generations of their evolution and saw that integration occurs between 500 and 700 generations of evolution.

Amazingly, by testing the replicative capacity of the clones at different generations, we were able to directly link the loss of virulence to the integration of the plasmid into the genome. This finding raised a lot of questions as to how the plasmid integration could impact so strongly the virulence for these clones. RNAseq analysis was performed to compare the transcriptomic landscape of the evolved clones compared to the ancestor. We saw that a large number of genes were impacted in all clones, so by focusing on only the genes commonly deregulated in all lineages, we confirmed that many functional categories were impacted such as metabolism, transport, translational process and type IV pili as well as 46 Dot/Icm effectors genes being up-regulated, while those linked to flagella and virulence-regulatory pathway are globally down-regulated. These results will later be discussed as part of the perspectives of this project.

The second part of my work was designed to test if the clones could reacquire their virulence, giving us high-quality settings to study emergence of the virulence in real-time. I designed another EE experiment in which these attenuated clones would be propagated in either axenic rich broth, or by repeatedly infecting *Acanthamoeba* spp., giving rise to the BX and AX lineages, respectively.

Restoration of the virulent phenotype was observed in AX lineages where avirulent clones were constrained to infecting their host in order to replicate and survive. Interestingly, AX lineages reacquired also their virulence towards U937 cells, adding evidence to the common assumption that similar processes allow *L. pneumophila* to infect their different host cells, natural or accidental. The main discovery of this experiment was that reacquisition of virulence could be directly linked to the excision of the plasmid from the genome, confirming the results I had hoped for.

On the other hand, results for the BX lines were not so clear as plasmid re-excision was not as definite as in AX lineages. Results seem to imply that a subset of the population retains the integrated form of the plasmid within each BX lineage throughout the re-evolution. What's interesting to note here is that BX lineages did not fully regain their virulence capacity towards amoebae and more strikingly, they seemed to have completely lost their virulence capacity towards U937 cells. I wasn't surprised that bacteria evolving without host selective pressure

did not fully regain the virulence capacity since they would trivially do not need to burden themselves with maintaining this costly feature. However, I was sure that other underlying mechanisms were involved to explain the complete loss of virulence towards U937 cells. Indeed, sequencing of 3 of the 6 BX lineages showed a large number of additional mutations, with in particular, 18 mutations in either/or the LetA/LetS two component system and in the (p)pGpp synthase RelA. These proteins are at the center of the switch from replicative to transmissive virulent phenotype (**figure 12**), suggesting that mutation accumulation in the BX lineages led to the loss of function of one or both regulators, ultimately leading to the loss of virulence.

To further support this assumption, none of the AX lineage had mutation in these systems. Surprisingly, all clones from all lineages retained the plasmid in an episomal form which strongly suggest a crucial role for this plasmid.

After validating the hypothesis that experimental evolution is a powerful tool to study evolution in bacterial pathogen using a modified *L. pneumophila*, I tackled the third part of my work.

In this part, I designed yet another EE experiment to study virulence emergence in a species of *Legionella* unknown to cause Legionnaire's disease. The main goal was to evolve *L. norrlandica* -a strictly environmental species, closest phylogenetically to *L. pneumophila* on the species tree – in presence or absence of amoebae, with or without DNA from the *L. pneumophila* Paris strain. This project might seem a little risky, but we planned to perform this evolution in fully controlled conditions.

The two postulates of this experiment were that *L. norrlandica* could replicate inside amoebae and that it was competent for natural transformation. During my PhD, I first validated the results available in one study and confirmed that *L. norrlandica* can replicate inside *Acanthamoeba* spp. Although in a lesser extent than *L. pneumophila*. I also showed that *L. norrlandica* however can't replicate inside A459 human pulmonary epithelial cells, contrary to *L. pneumophila*. These results support what is seen in "real life" and comforted the idea that *L. norrlandica* is the perfect organism for our experiment. However, it would be interesting to test the replicative capacity of the bacteria in other cell types such as U937 cells.

Recent results have shown that *L. norrlandica* possess all the genes required to enter into a competent state for natural transformation. Then, I created two genetic tools (two cassettes conferring resistance to either streptomycin or kanamycin) and to follow gene acquisition by the bacteria *via* natural transformation. The results obtained with the streptomycin cassette were not conclusive due to clouding of the transformation efficacy by the high rate of spontaneous mutations arising. The kanamycin cassette would appear more efficient in this setting as spontaneous mutation towards kanamycin is lower than that of streptomycin. This tool still needs to be empirically validated with transformation assays.

In parallel, I tested the natural competence of *L. norrlandica* by following the transcription of *comEA* and RocR by Northern blot analysis, in the same conditions known to trigger competence for natural transformation in *L. pneumophila*. My results showed that sRNA RocR, the repressor of competence was constitutively express across all growth phases of *L. norrlandica*. Conversely, *comEA* transcription was not detected. This could imply either a different control of competence by *L. norrlandica*, or that conditions triggering this state are not the same as *L. pneumophila*.

Together, my results proved that Experimental Evolution combined with high throughput sequencing are formidable tools to study emergence of pathogens.

Our main finding suggests that plasmid seem to have major effects on the bacterial genome and on how bacteria interact with their environment. This field of research is emerging, and more and more studies are taking interests in mobile genetic elements and how they regulate major pathways in bacteria. My PhD fully contributed to illustrate this field.

The perspectives on my work are plenty and it could be interesting to look at the broader distribution of not only pLPP but also other plasmids in the *Legionella* genome and characterize further their impact. It could also be interesting to look at the co-distribution of the IS and plasmid in the ST1 subtype, as preliminary results seem to indicate that they are largely distributed. One of my interrogations was also to find how plasmid integration can impact global chromosomal gene expression. Does it have an impact on chromosome architecture, or transcription regulators? Could the integration directly impact transcription of the genes in the neighborhood of the IS? A new project has been initiated in the laboratory to answer this question. Indeed, we showed that IS in the chromosome is flanked by two genes

coding for an effector called SidH. It was previously thought that SidH was inactive as the gene coding for this effector were in two parts. However, preliminary results demonstrated that the protein is actually composed of two modules that interact with each other with LubX, a meta-effector that directs it to the proteasome.

Finally, in lights of recent work on mobile genetic elements, it could be interesting to look at the pLPP genes' impact on bacterial and host functions. Our work showed an over-representation of proteins from the GNAT acetyltransferase family in the chromosome (14 out of 277 genes of the plasmid). A recent study has exhibited the role of a *L. pneumophila* deacetylase on epigenetic modification of the host cell histones. We can only hypothesize similar roles for the GNAT family proteins.

Another perspective on this work could be brought by bioinformatics. It would be interesting to reconstruct the evolutionary story of *L. pneumophila* to account for events of horizontal gene transfers, deletion or duplications and reconcile the gene tree with the species tree. As part of my PhD, I had started to recover and clean genomes of *Legionella* but due to time constraints, I did not move further with this project.

Further work is needed also on *L. norrlandica* to better characterize the bacteria and initiate the experimental evolution lineages. If successful, this experiment could lead to a breakthrough discovery on the emergence of *Legionella pneumophila* as a public health issue.

5. Materials and Methods

All experiments were performed several times and the data shown are from one representative experiment with 3 or more replicates each experiment.

5.1. Bacterial strains, macrophages and amoebae strains and growth conditions

Bacterial strains, cells lines used in this study are summarized in [Supplementary Data S1](#). *Legionella pneumophila* strain Paris and subsequent evolved clones were grown at 37°C on BCYE (ACES-buffered charcoal yeast extract) agar plates supplemented with 0.4 mg.mL⁻¹ L-cysteine and 0.25 mg.mL⁻¹ iron pyrophosphate. AYE broth (CYE without agar and charcoal) was used as liquid medium. When mentioned, media were supplemented with 10µg.mL⁻¹ and/or 0.25 mM IPTG. Human, monocyte-derived U937 macrophages cell line was maintained in RPMI 1640 (ThermoFisher Scientific) medium, without glutamic acid, supplemented with 10% heat-inactivated FBS (HyClone) at 37°C in a 5% CO₂ atmosphere. For infections, U937 cells were induced to differentiate 3 days by exposure to 100 ng.mL⁻¹ of phorbol 12-myristate 13-acetate (PMA). *Acanthamoeba polyphaga* or *Acanthamoeba* spp. cells were maintained in PY (20 mg.mL⁻¹ peptone ; 1 mg.mL⁻¹ yeast extract ; 0.81M MgSO₄ ; 1M CaCl₂ ; 1M sodium citrate ; 5mM Fe(NH₄)₂ (SO₄)₂ ; 0.25M Na₂HPO₄ ; 0.25M KH₂PO₄) medium, supplemented with 2% glucose and grown at 30°C.

Dictyostelium discoideum Dd04 expressing calnexin-green fluorescent protein (GFP) (DBS0236184) was obtained from the Dicty Stock Center (<http://dictybase.org/StockCenter/StockCenter.html> ; depositor, A. Muller-Taubenberger). *D. discoideum* cells were grown at 22°C in HL5 medium supplemented with 20 µg.mL⁻¹ neomycin when necessary.

5.2. Experimental evolution strategies

These experiments were conducted for 35 passages. Each 5 passage, bacteria were harvested and kept frozen at -80°C.

5.2.1. Mutation accumulation strategy:

L. pneumophila Paris strain was streaked onto supplemented BCYE plates and left to grow at 37°C for 3 days. Then 6 randomly chosen colonies were picked and used to generate six independent lineages by re-streaking it on new supplemented BCYE plates and left to grow at 37°C. Then, every 3 days, one colony per lineage was randomly picked and re-streaked on new supplemented BCYE plates. This is equivalent to one (1) passage.

5.2.2. Adaptative mutation strategy:

5.2.2.1. Propagation in *Acanthamoeba* spp.:

Acanthamoeba spp. cells were plated at a concentration of $5 \cdot 10^6$ cells/mL in T25 flasks (Sarstedt) in PY special (PY medium without peptone and yeast extract) and left to adhere during 1h30-2h at 30°C. After that time, flasks were visually inspected to confirm adherence to the plate. For the first passage, $2.5 \cdot 10^6$ bacteria suspended in PY special were used to infect the amoebae. Flasks were kept at 30°C until complete amoebae lysis. For subsequent passages, totality of the flask was harvested, and centrifuged 1min at 1500rpm. 100µL of the supernatant was collected and diluted in 900µL of PY special. Then 10µL of the bacterial suspension was used to seed new flasks with plated amoebae. This is equivalent to one (1) passage.

5.2.2.2. Propagation in AYE medium:

AYE broth was seeded with $3 \cdot 10^7$ bacteria and kept at 37°C under agitation for approx. 2days or until bacteria grew to a concentration of $4 \cdot 10^9$ bacteria/mL. This is equivalent to one (1) passage. Until Passage 13, 8µL of the previous passage was used to seed fresh AYE broth. Until passage P17, bacterial suspension from previous passage was diluted by half and 2.5µL of suspension was used to seed fresh AYE broth. Until passage P20, bacterial suspension from previous passage was diluted by 20 and 2.5µL of suspension was used to seed fresh AYE broth. Until passage P35, bacterial suspension from previous passage was diluted by 100 and 2.5µL of suspension was used to seed fresh AYE broth.

5.3. Bacterial competition assays

Isolated colonies from three evolved lineages (Paris 4, Paris 5 and Paris 6) as well as the ancestral clone were naturally transformed using homologous recombination with the insertion of a fluorescent cassette for either Yellow or Blue fluorescent protein (Yfp or Cfp, respectively) under control of an IPTG-inducible promoter, using the Kan/MazF protocol published by Bailo *et al* in *Methods Mol Biol.* 2019. These isolates were used to independently inoculate 3 mL of AYE rich medium with addition of IPTG. After 24h of growth at 37°C with continuous shaking, cultures were diluted in fresh medium adjusted to an optical density at 600 nm (OD600) of 0.2 and allowed to grow together at a 1:1 ratio for a further 24 hours during which blue or yellow fluorescence was measured every 30 minutes. Finally, ratio of yellow to blue fluorescence was used to calculate the competitive index.

5.4. Intracellular replication assays

For bacterial replication analysis, 1.10^5 macrophages per well were plated in 200 μ L of CO₂-independent medium supplemented with 5% FBS, 10 μ g.mL⁻¹ chloramphenicol and 0.25 mM IPTG in black, clear flat bottom 96-well plates (Greiner). *Acanthamoeba* spp. were plated at a density of 1.10^5 cells per well in 200 μ L PY supplemented with 10 μ g.mL⁻¹ chloramphenicol and 0.25 mM IPTG in black, clear flat bottom 96-well plates (Greiner). Eukaryotic cells were then challenged by *Legionella pneumophila* harbouring an IPTG inducible, mCherry-coding plasmid (pXDC50, obtained from Xavier Charpentier) at a MOI of 0.5. Plates were centrifuged at 1500 rpm for 5min and incubated at 30°C for amoebae and 37°C for macrophages. Intracellular growth was automatically monitored by measuring mCherry fluorescence at λ_{ex} = 587 nm 147 and λ_{em} = 610 nm every 30 min on an Infinite M200 microplate reader (Tecan).

5.5. Endoplasmic reticulum recruitment measurement

D. discoideum Dd04 cells producing a calnexin-GFP fusion protein were left to adhere onto sterile glass coverslips on 24 well plates at 1.10^6 cells/well in MB medium at 22°C overnight. MB medium was then carefully removed and replaced with 1mL of bacterial suspension containing 5.10^7 *L. pneumophila* cells transformed with the pXDC50 plasmid (MOI of 50) in MB medium supplemented with 0.25 mM IPTG. The plate was centrifuged at 1500rpm for 10min and incubated for 1h at 25°C. The medium was carefully removed, and cultures were fixed 30 min in the dark at room temperature by adding 2 mL of *paraformaldehyde* (PAF) 4% in PBS.

Coverslips were carefully rinsed in PBS, mounted in a Dabco+Mowiol mix, sealed and observed with an inverted confocal microscope.

5.6. TEM translocation assay

Translocation assays were performed as previously described (Allombert et al. 2014). Briefly, *L. pneumophila* cells carrying various TEM- β lactamase-effector fusion proteins (TEM-RalF, TEM-LepA, TEM-SdeA, TEM-LegA3 and TEM-FabI as negative control) were grown on BCYE plates containing chloramphenicol and IPTG for 72h to induce the production of the fusion proteins and replica-plated on fresh plate for an additional 24h. 10 μ L of bacteria re-suspended in RPMI at 1.10^8 cells.mL⁻¹ were used to infect differentiated U937 cells at a MOI of 10. After centrifugation (880 x *g*, 10 min), the plate was incubated at 37°C with CO₂. At 1h post-infection, 10 μ M of CCCP was added to block the secretion through the Dot/Icm T4SS. Cells monolayers were then loaded with the fluorescent substrate by adding 20 μ L of 6X CCF4/AM solution (LiveBLAzer-FRET B/G Loading Kit, Invitrogen) containing 15mM Probenecid (Sigma). The cells were then incubated at room temperature.

5.7. Genomic DNA preparation, whole-genome sequencing and comparative genomic analysis.

Genomic DNA was extracted using the Wizard™ Genomic DNA Purification kit (Promega) according to the manufacturer's instructions. Whole-genome sequencing (WGS) was performed using both Pacific Biosciences (PacBio) and Illumina technology. PacBio was used for getting long-read sequencing in order to detect duplications, deletions and large genomic rearrangements.

Long-read sequencing: The Pacific Biosciences libraries DNA libraries were produced using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences) according to the manufacturer's instructions. The libraries were sequenced on the PacBio RS instrument using C4 chemistry and P6 polymerase and 2 SMRT cells by Lausanne Genomic Technologies Facility GTF (Lausanne, Switzerland)

Short-read sequencing: Illumina DNA prep kit was used for library preparation followed by paired-end 2x150 bp sequencing using a Nextseq 550 sequencer (Illumina®, San Diego, USA). Genome assembly and polishing were performed with flye 2.9.1 and medaka 1.7.3,

respectively (Kolmogorov et al., 2019) (<https://github.com/nanoporetech/medaka>). The obtained genomes were annotated with bakta 1.7.0 (Schwengers et al., 2021).

Variant calling: To determine SNP frequency within populations, amplicon libraries of the target regions were constructed using a dual barcoded protocol. Briefly, target regions in *letA*, *letS* and *relA* genes were amplified using sequence-specific primers: LetA_for (ccaaatattcatataagggc) and LetA_rev (cggttcacttgggagtttcg); LetS_for (cccaccaacgtgtaaaaaacc) and LetS_rev (ggttgatacagatTTTTactttg; RelA_for (ggtaacgcaggcaattcagt) and RelA_rev (cacagagcagttgttcagg). The resulting PCR products were assessed on a 1% agarose gel. Equimolar ratios of the three amplicons were prepared and then used as template to produce DNA libraries using the Nextera XT DNA Library Prep kit (Illumina). Briefly, the DNA is fragmented and bound to the appropriate barcodes. After amplification, fragments of the desired size are selected using magnetic beads and the quality and yield of the final amplicon libraries were assessed on an Agilent Bioanalyzer (Agilent Technologies). The resulting libraries were sequenced using on an Illumina HiSeq sequencing platform. Sequencing adapters and low-quality bases were trimmed, the read numbers normalized using custom scripts.

5.8. Detection of the episomal or integrated form of the pLPP plasmid by PCR and Digital droplet PCR

PCRs were performed on chromosomal and plasmid DNA to detect either the episomal form of the plasmid (LPP-For: 5'-ACAAGAGATAGGGCTGATAAACG-3'; LPP-Rev: 5'-TGGCTGCCTTGACCAAAC-3') or the integrated form (pLPP-For: 5'-CCAACCTCAAAGAGTATTGCACA-3'; pLPP-Rev: 5'-AGCAAGGGCAGAGACCACC-3'). Additional PCRs were performed to cross-verify the results: LPP-For X pLPP-Rev and pLPP-For X LPP-Rev. To assess the status of the plasmid in the entire bacterial population, digital droplet PCR (ddPCR) was performed. A mix of the pLPP-For, pLPP-Rev and LPP-For LPP-Rev primers as well as two fluorescent probes pb2: [FAM] AGGCAGATTACTAACTCCATTACA [BHQ1] and pb4: [HEX] AGTCCTGAAGTAACACTCGGTAA [BHQ1], coupled with 6-carboxyfluorescein (FAM) and hexachloro-fluorescein (HEX), respectively, was used with the ddPCR super Mix to perform the reaction directly on bacterial populations at a concentration of $8 \cdot 10^5$ cells.mL⁻¹. The ddPCR protocol was improved in order to detect only one single bacterium per droplet. After partitioning the ddPCR reaction mix into thousands droplets with the QX200 Droplet Generator, the PCR was then performed as follows: initial denaturation at 98°C for 10min,

then 80 cycles of denaturation at 95°C for 20 sec, annealing at 55°C for 30 sec and elongation at 72°C for 2min and a final denaturation at 98°C for 10min. After PCRs, droplets from each sample are analyzed individually on the QX200 Droplet Reader. The PCR-positive and PCR-negative droplets are counted providing an absolute quantification of bacteria.

5.9. RNA isolation, depletion of rRNA and RNA sequencing

L. pneumophila Paris and the three evolved clones Paris 4, 5 and 6 were grown in AYE medium until post-exponential phase (OD_{600nm} 4 and visual check of motility acquisition), harvested by centrifugation (5 min, 179 7000 rpm, 4°C) and immediately frozen at -80°C for subsequent RNAs extraction. Total RNA from bacterial cultures was extracted according to a previously described procedure (Attaiech et al. 2016). Briefly, pellets of 10⁹ bacterial cells were lysed in 50 µl of RNAsnap buffer (18 mM EDTA, 0.025% SDS, 95% formamide) and total RNAs were extracted using a tri-reagent solution (acid guanidinium thiocyanate–phenol–chloroform) and isopropanol-precipitated. After precipitation, we performed an additional step of RNA purification on silica-based columns (DirectZol kit, ZymoResearch) by following the manufacturer's recommendations. RNA sample purity and concentration were determined by spectrophotometric analysis on a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo). RNA sequencing was performed following ribosomal RNA depletion and cDNA libraries preparation on an NovaSeq platform (Illumina) with paired-end 150bp (Genewiz-Azenta, Leipzig, Germany). After mapping sequence reads to the reference genome and extraction of gene hit counts, the comparison of gene expression between the defined groups of samples was performed using DESeq2. The BAM files were imported into IGV software (V2.15.2) and reads were aligned with the genome sequence of *L. pneumophila* Paris strain (NCBI accession number: NC_006368).

6. References

1. Fields, B. S. The molecular ecology of legionellae. *Trends Microbiol.* **4**, 286–290 (1996).
2. Faulkner, G., Berk, S. G., Garduño, E., Ortiz-Jiménez, M. A. & Garduño, R. A. Passage through *Tetrahymena tropicalis* Triggers a Rapid Morphological Differentiation in *Legionella pneumophila*. *J. Bacteriol.* **190**, 7728–7738 (2008).
3. Molmeret, M., Bitar, D. M., Han, L. & Kwaik, Y. A. Cell biology of the intracellular infection by *Legionella pneumophila*. *Microbes Infect.* **6**, 129–139 (2004).
4. Fraser, D. W. *et al.* Legionnaires' Disease. *N. Engl. J. Med.* **297**, 1189–1197 (1977).
5. McDade, J. E. *et al.* Legionnaires' Disease. *N. Engl. J. Med.* **297**, 1197–1203 (1977).
6. Beauté, J. Legionnaires' disease in Europe, 2011 to 2015. *Eurosurveillance* **22**, 30566 (2017).
7. Fields, B. S., Benson, R. F. & Besser, R. E. Legionella and Legionnaires' Disease: 25 Years of Investigation. *Clin. Microbiol. Rev.* **15**, 506–526 (2002).
8. Yu, V. L. *et al.* 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 (2002).

9. Gomez-Valero, L. *et al.* Comparative analyses of Legionella species identifies genetic features of strains causing Legionnaires' disease. *Genome Biol.* **15**, 505 (2014).
10. Lawrence, C. *et al.* Single clonal origin of a high proportion of Legionella pneumophila serogroup 1 isolates from patients and the environment in the area of Paris, France, over a 10-year period. *J. Clin. Microbiol.* **37**, 2652–2655 (1999).
11. Cazalet, C. *et al.* Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36**, 1165–1173 (2004).
12. Cazalet, C. *et al.* Multigenome analysis identifies a worldwide distributed epidemic Legionella pneumophila clone that emerged within a highly diverse species. *Genome Res.* **18**, 431–441 (2008).
13. Natural selection. *Darwin Correspondence Project*
<https://www.darwinproject.ac.uk/commentary/evolution/natural-selection> (2016).
14. A Naturalist's Voyage Round the World: Title. <http://www.gutenberg.org/files/3704/3704-h/3704-h.htm>.
15. Herbert, S. ed. 1980. The red notebook of Charles Darwin. Bulletin of the British Museum (Natural History) Historical Series 7 (24 April): 1-164. <http://darwin-online.org.uk/content/frameset?viewtype=text&itemID=F1583e&pageseq=9>.
16. Behie, A. M. & Oxenham, M. F. *Taxonomic Tapestries: The Threads of Evolutionary, Behavioural and Conservation Research*. (ANU Press, 2015).
17. Boussau, B. & Scornavacca, C. Reconciling Gene trees with Species Trees. 24.
18. Haas, J. W. The Reverend Dr William Henry Dallinger, F.R.S. (1839-1909). 13.
19. Kawecki, T. J. *et al.* Experimental evolution. *Trends Ecol. Evol.* **27**, 547–560 (2012).
20. Ostrowski, E. A., Ofria, C. & Lenski, R. E. Ecological Specialization and Adaptive Decay in Digital Organisms. *Am. Nat.* **169**, E1–E20 (2007).
21. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).

22. Maharjan, R. P. *et al.* Mutation accumulation and fitness in mutator subpopulations of *Escherichia coli*. *Biol. Lett.* **9**, 20120961 (2013).
23. Long, H., Miller, S. F., Williams, E. & Lynch, M. Specificity of the DNA Mismatch Repair System (MMR) and Mutagenesis Bias in Bacteria. *Mol. Biol. Evol.* **35**, 2414–2421 (2018).
24. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci.* **109**, E2774–E2783 (2012).
25. Couce, A. & Tenaillon, O. Mutation bias and GC content shape antimutator invasions. *Nat. Commun.* **10**, 3114 (2019).
26. Bosshard, L., Peischl, S., Ackermann, M. & Excoffier, L. Dissection of the mutation accumulation process during bacterial range expansions. *BMC Genomics* **21**, 253 (2020).
27. Castañeda-García, A. *et al.* Specificity and mutagenesis bias of the mycobacterial alternative mismatch repair analyzed by mutation accumulation studies. *Sci. Adv.* **6**, eaay4453 (2020).
28. Niccum, B. A. *et al.* New complexities of SOS-induced “untargeted” mutagenesis in *Escherichia coli* as revealed by mutation accumulation and whole-genome sequencing. *DNA Repair* **90**, 102852 (2020).
29. Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* **15**, 141–161 (2015).
30. Raeside, C. *et al.* Large Chromosomal Rearrangements during a Long-Term Evolution Experiment with *Escherichia coli*. *mBio* **5**, e01377-14 (2014).
31. Yue, W.-F., Du, M. & Zhu, M.-J. High Temperature in Combination with UV Irradiation Enhances Horizontal Transfer of *stx2* Gene from *E. coli* O157:H7 to Non-Pathogenic *E. coli*. *PLOS ONE* **7**, e31308 (2012).
32. Zhao, S. *et al.* Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host Microbe* **25**, 656-667.e8 (2019).

33. Jahn, L. J., Munck, C., Ellabaan, M. M. H. & Sommer, M. O. A. Adaptive Laboratory Evolution of Antibiotic Resistance Using Different Selection Regimes Lead to Similar Phenotypes and Genotypes. *Front. Microbiol.* **8**, (2017).
34. Martin, M., Hölscher, T., Dragoš, A., Cooper, V. S. & Kovács, Á. T. Laboratory Evolution of Microbial Interactions in Bacterial Biofilms. *J. Bacteriol.* **198**, 2564–2571 (2016).
35. Lenski, R. E. *et al.* Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proc. R. Soc. B Biol. Sci.* **282**, 20152292 (2015).
36. Wisner, M. J., Ribbeck, N. & Lenski, R. E. Long-Term Dynamics of Adaptation in Asexual Populations. *Science* **342**, 1364–1367 (2013).
37. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The Dynamics of Molecular Evolution Over 60,000 Generations. *Nature* **551**, 45–50 (2017).
38. Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* **536**, 165–170 (2016).
39. Wielgoss, S. *et al.* Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 222–227 (2013).
40. Grant, N. A., Abdel Magid, A., Franklin, J., Dufour, Y. & Lenski, R. E. Changes in Cell Size and Shape during 50,000 Generations of Experimental Evolution with *Escherichia coli*. *J. Bacteriol.* **203**, e00469-20 (2021).
41. Woods, R., Schneider, D., Winkworth, C. L., Riley, M. A. & Lenski, R. E. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9107–9112 (2006).
42. Favate, J. S., Liang, S., Cope, A. L., Yadavalli, S. S. & Shah, P. The landscape of transcriptional and translational changes over 22 years of bacterial adaptation. *eLife* **11**, e81979 (2022).
43. Lenski, R. E. Convergence and Divergence in a Long-Term Experiment with Bacteria. *Am. Nat.* **190**, S57–S68 (2017).

44. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **105**, 7899–7906 (2008).
45. Blount, Z. D., Lenski, R. E. & Losos, J. B. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, eaam5979 (2018).
46. Hall, B. G. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *J. Bacteriol.* **151**, 269–273 (1982).
47. Turner, C. B., Blount, Z. D., Mitchell, D. H. & Lenski, R. E. Evolution and coexistence in response to a key innovation in a long-term evolution experiment with *Escherichia coli*. 020958 Preprint at <https://doi.org/10.1101/020958> (2015).
48. Leon, D., D'Alton, S., Quandt, E. M. & Barrick, J. E. Innovation in an *E. coli* evolution experiment is contingent on maintaining adaptive potential until competition subsides. *PLoS Genet.* **14**, e1007348 (2018).
49. Sabin, A. B. & Schlesinger, R. W. Production of Immunity to Dengue with Virus Modified by Propagation in Mice. *Science* **101**, 640–642 (1945).
50. Woolhouse, M. E. J., Taylor, L. H. & Haydon, D. T. Population Biology of Multihost Pathogens. *Science* **292**, 1109–1112 (2001).
51. Kapperud, G. *Yersinia enterocolitica* in food hygiene. *Int. J. Food Microbiol.* **12**, 53–65 (1991).
52. Fukushima, H. & Gomyoda, M. Intestinal carriage of *Yersinia pseudotuberculosis* by wild birds and mammals in Japan. *Appl. Environ. Microbiol.* **57**, 1152–1155 (1991).
53. Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14043–14048 (1999).
54. Bäumlner, A. & Fang, F. C. Host Specificity of Bacterial Pathogens. *Cold Spring Harb. Perspect. Med.* **3**, (2013).

55. Ensminger, A. W., Yassin, Y., Miron, A. & Isberg, R. R. Experimental Evolution of *Legionella pneumophila* in Mouse Macrophages Leads to Strains with Altered Determinants of Environmental Survival. *PLoS Pathog.* **8**, e1002731 (2012).
56. Ferenci, T. Trade-off Mechanisms Shaping the Diversity of Bacteria. *Trends Microbiol.* **24**, 209–223 (2016).
57. McDonald, M. J. Microbial Experimental Evolution – a proving ground for evolutionary theory and a tool for discovery. *EMBO Rep.* **20**, e46992 (2019).
58. Łuksza, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507**, 57–61 (2014).
59. Bull, J. J. & Molineux, I. J. Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity* **100**, 453–463 (2008).
60. Martin, W. F., Garg, S. & Zimorski, V. Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140330 (2015).
61. Symbiose et évolution : à l'origine de la cellule eucaryote. *Encyclopédie de l'environnement* <https://www.encyclopedie-environnement.org/vivant/symbiose-evolution-lorigine-de-cellule-eucaryote/> (2016).
62. Steinert, M., Hentschel, U. & Hacker, J. Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* **87**, 1–11 (2000).
63. McFall-Ngai, M., Heath-Heckman, E. A. C., Gillette, A. A., Peyer, S. M. & Harvie, E. A. The secret languages of coevolved symbioses: Insights from the *Euprymna scolopes*–*Vibrio fischeri* symbiosis. *Semin. Immunol.* **24**, 3–8 (2012).
64. Leung, T. & Poulin, R. Parasitism, commensalism, and mutualism: exploring the many shades of symbioses. (2008).
65. Zimmer, C. Wolbachia: A Tale of Sex and Survival. *Science* **292**, 1093–1095 (2001).
66. Unusual intra-cellular bacterial infection in large, free-living amoebae. *Exp. Cell Res.* **48**, 236–240 (1967).

67. Park, M., Yun, S. T., Kim, M. S., Chun, J. & Ahn, T. I. Phylogenetic characterization of Legionella-like endosymbiotic X-bacteria in *Amoeba proteus*: a proposal for 'Candidatus Legionella jeonii' sp. nov. *Environ. Microbiol.* **6**, 1252–1263 (2004).
68. Choi, J. Y., Lee, T. W., Jeon, K. W. & Ahn, T. I. Evidence for Symbiont-induced Alteration of a Host's Gene Expression: Irreversible Loss of SAM Synthetase from *Amoeba proteus*. *J. Eukaryot. Microbiol.* **44**, 412–419 (1997).
69. Jeon, K. W. & Ahn, T. I. Temperature Sensitivity: A Cell Character Determined by Obligate Endosymbionts in Amoebas. *Science* **202**, 635–637 (1978).
70. Jeon, K. W. The Large, Free-living Amoebae: Wonderful Cells for Biological Studies. *J. Eukaryot. Microbiol.* **42**, 1–7 (1995).
71. Genin, S. Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum*. *New Phytol.* **187**, 920–928 (2010).
72. Hirsch, A. M., Lum, M. R. & Downie, J. A. What Makes the Rhizobia-Legume Symbiosis So Special? *Plant Physiol.* **127**, 1484–1492 (2001).
73. Marchetti, M. *et al.* Experimental Evolution of a Plant Pathogen into a Legume Symbiont. *PLOS Biol.* **8**, e1000280 (2010).
74. Chu, H. & Mazmanian, S. K. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nat. Immunol.* **14**, 668–675 (2013).
75. Calderone, R. A. & Fonzi, W. A. Virulence factors of *Candida albicans*. *Trends Microbiol.* **9**, 327–335 (2001).
76. Tso, G. H. W. *et al.* Experimental evolution of a fungal pathogen into a gut symbiont. *Science* **362**, 589–595 (2018).
77. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* **474**, 1823–1836 (2017).
78. King, K. C. *et al.* Rapid evolution of microbe-mediated protection against pathogens in a worm host. *ISME J.* **10**, 1915–1924 (2016).

79. Résistance aux antibiotiques. *Institut Pasteur* <https://www.pasteur.fr/fr/centre-medical/fiches-maladies/resistance-aux-antibiotiques> (2017).
80. Massip, C. *et al.* Macrolide resistance in *Legionella pneumophila*: the role of LpeAB efflux pump. *J. Antimicrob. Chemother.* **72**, 1327–1333 (2017).
81. Vandewalle-Capo, M. *et al.* Minimum inhibitory concentration (MIC) distribution among wild-type strains of *Legionella pneumophila* identifies a subpopulation with reduced susceptibility to macrolides owing to efflux pump genes. *Int. J. Antimicrob. Agents* **50**, 684–689 (2017).
82. Shadoud, L. *et al.* Hidden Selection of Bacterial Resistance to Fluoroquinolones In Vivo: The Case of *Legionella pneumophila* and Humans. *EBioMedicine* **2**, 1179–1185 (2015).
83. Descours, G. *et al.* Ribosomal Mutations Conferring Macrolide Resistance in *Legionella pneumophila*. *Antimicrob. Agents Chemother.* **61**, e02188-16, e02188-16 (2017).
84. Almahmoud, I., Kay, E., Schneider, D. & Maurin, M. Mutational paths towards increased fluoroquinolone resistance in *Legionella pneumophila*. *J. Antimicrob. Chemother.* **64**, 284–293 (2009).
85. Jonas, D. *et al.* Development and mechanism of fluoroquinolone resistance in *Legionella pneumophila*. *J. Antimicrob. Chemother.* **51**, 275–280 (2003).
86. Winn, W. C. Legionnaires disease: historical perspective. *Clin. Microbiol. Rev.* **1**, 60–81 (1988).
87. Fields, B. S., Benson, R. F. & Besser, R. E. *Legionella* and Legionnaires' Disease: 25 Years of Investigation. *Clin. Microbiol. Rev.* **15**, 506–526 (2002).
88. Fliermans, C. B. *et al.* Ecological distribution of *Legionella pneumophila*. *Appl. Environ. Microbiol.* **41**, 9–16 (1981).
89. Steinert, M., Hentschel, U. & Hacker, J. *Legionella pneumophila*: an aquatic microbe goes astray. *FEMS Microbiol. Rev.* **26**, 149–162 (2002).
90. Best, A. & Kwaik, Y. A. Evolution of the Arsenal of *Legionella pneumophila* Effectors To Modulate Protist Hosts. *mBio* **9**, e01313-18 (2018).

91. Schalk, J. A. C. *et al.* Soil as a source of Legionella pneumophila sequence type 47. *Int. J. Infect. Dis.* **27**, 18–19 (2014).
92. Rowbotham, T. J. Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. *J. Clin. Pathol.* **33**, 1179–1183 (1980).
93. Kwaik, Y. A. Fatal attraction of mammalian cells to Legionella pneumophila. *Mol. Microbiol.* **30**, 689–695 (1998).
94. Cianciotto, N. P., Stamos, J. K. & Kamp, D. W. Infectivity of Legionella pneumophila mip mutant for alveolar epithelial cells. *Curr. Microbiol.* **30**, 247–250 (1995).
95. Dietrich, C., Heuner, K., Brand, B. C., Hacker, J. & Steinert, M. Flagellum of Legionella pneumophila Positively Affects the Early Phase of Infection of Eukaryotic Host Cells. *Infect. Immun.* **69**, 2116–2122 (2001).
96. THE PHILADELPHIA KILLER. *Time* (1976).
97. 40 years later, scientist who first discovered Legionnaires' disease is still learning lessons. *WHYY* <https://whyy.org/articles/40-years-later-scientist-who-first-discovered-legionnaires-disease-is-still-learning-lessons/>.
98. Star-News - Google News Archive Search. <https://news.google.com/newspapers?nid=1454&dat=19780903&id=bLssAAAAIBAJs&sjid=LRMEAAAAIBAJs&pg=2736,405917>.
99. Van Kenhove, E., Dinne, K., Janssens, A. & Laverge, J. Overview and comparison of Legionella regulations worldwide. *Am. J. Infect. Control* **47**, 968–978 (2019).
100. Légionellose. /maladies-et-traumatismes/maladies-et-infections-respiratoires/legionellose.
101. Traitement antibiotique de la legionnellose chez l'adulte - Mise au point.
102. Blaser, M. HOT-BATH SYNDROME, PONTIAC FEVER, AND LEGIONNAIRES' DISEASE. *The Lancet* **310**, 1226 (1977).
103. Correia, A. M. *et al.* Probable Person-to-Person Transmission of Legionnaires' Disease. (2016) doi:10.1056/NEJMc1505356.

104. Cirillo, J. D., Falkow, S. & Tompkins, L. S. Growth of *Legionella pneumophila* in *Acanthamoeba castellanii* enhances invasion. *Infect. Immun.* **62**, 3254–3261 (1994).
105. Cirillo, J. D. *et al.* Intracellular Growth in *Acanthamoeba castellanii* Affects Monocyte Entry Mechanisms and Enhances Virulence of *Legionella pneumophila*. *Infect. Immun.* **67**, 4427–4434 (1999).
106. Brieland, J. *et al.* Coinoculation with *Hartmannella vermiformis* enhances replicative *Legionella pneumophila* lung infection in a murine model of Legionnaires' disease. *Infect. Immun.* **64**, 2449–2456 (1996).
107. Garduño, R. A., Garduño, E., Hiltz, M. & Hoffman, P. S. Intracellular Growth of *Legionella pneumophila* Gives Rise to a Differentiated Form Dissimilar to Stationary-Phase Forms. *Infect. Immun.* **70**, 6273–6283 (2002).
108. Boamah, D. K., Zhou, G., Ensminger, A. W. & O'Connor, T. J. From Many Hosts, One Accidental Pathogen: The Diverse Protozoan Hosts of *Legionella*. *Front. Cell. Infect. Microbiol.* **7**, (2017).
109. Mengue, L. *et al.* *Legionella pneumophila* prevents proliferation of its natural host *Acanthamoeba castellanii*. *Sci. Rep.* **6**, 1–12 (2016).
110. Buse, H. Y. & Ashbolt, N. J. Differential growth of *Legionella pneumophila* strains within a range of amoebae at various temperatures associated with in-premise plumbing. *Lett. Appl. Microbiol.* **53**, 217–224 (2011).
111. Khodr, A. *et al.* Molecular epidemiology, phylogeny and evolution of *Legionella*. *Infect. Genet. Evol.* **43**, 108–122 (2016).
112. Anderson, O. R., Wang, W., Faucher, S. P., Bi, K. & Shuman, H. A. A New Heterolobosean Amoeba *Solumitrus palustris* n. g., n. sp. Isolated from Freshwater Marsh Soil. *J. Eukaryot. Microbiol.* **58**, 60–67 (2011).
113. Amaro, F., Wang, W., Gilbert, J. A., Anderson, O. R. & Shuman, H. A. Diverse protist grazers select for virulence-related traits in *Legionella*. *ISME J.* **9**, 1607–1618 (2015).

114. Edelstein, P. H., Edelstein, M. A., Shephard, L. J., Ward, K. W. & Ratcliff, R. M. *Legionella steelei* sp. nov., isolated from human respiratory specimens in California, USA, and South Australia. *Int. J. Syst. Evol. Microbiol.* **62**, 1766–1771 (2012).
115. Segal, G., Purcell, M. & Shuman, H. A. 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.* **95**, 1669–1674 (1998).
116. Cavalier-Smith, T. Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *Eur. J. Protistol.* **49**, 115–178 (2013).
117. Cavalier-Smith, T., Chao, E. E.-Y. & Oates, B. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont Phalansterium. *Eur. J. Protistol.* **40**, 21–48 (2004).
118. Cavalier-Smith, T. *et al.* Multigene phylogeny resolves deep branching of Amoebozoa. *Mol. Phylogenet. Evol.* **83**, 293–304 (2015).
119. Ruggiero, M. A. *et al.* A Higher Level Classification of All Living Organisms. *PLOS ONE* **10**, e0119248 (2015).
120. Marrie, T. J. *et al.* *Legionella*-like and other amoebal pathogens as agents of community-acquired pneumonia. *Emerg. Infect. Dis.* **7**, 1026–1029 (2001).
121. Adeleke, A. *et al.* *Legionella*-like amoebal pathogens--phylogenetic status and possible role in respiratory disease. *Emerg. Infect. Dis.* **2**, 225–230 (1996).
122. Adeleke, A. A. *et al.* *Legionella drozanskii* sp. nov., *Legionella rowbothamii* sp. nov. and *Legionella fallonii* sp. nov.: three unusual new *Legionella* species. *Int. J. Syst. Evol. Microbiol.* **51**, 1151–1160 (2001).
123. Scola, B. L. *et al.* *Legionella drancourtii* sp. nov., a strictly intracellular amoebal pathogen. *Int. J. Syst. Evol. Microbiol.* **54**, 699–703 (2004).

124. Gomaa, F., Gersh, M. & Cavanaugh, C. M. Diverse Legionella-Like Bacteria Associated with Testate Amoebae of the Genus Arcella (Arcellinida: Amoebozoa). *J. Eukaryot. Microbiol.* **65**, 661–668 (2018).
125. Yang, B., Wang, Y. & Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**, 135 (2016).
126. Schriefer, A. E. *et al.* A multi-amplicon 16S rRNA sequencing and analysis method for improved taxonomic profiling of bacterial communities. *J. Microbiol. Methods* **154**, 6–13 (2018).
127. Nei, M. Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* **30**, 371–403 (1996).
128. Rosenberg, M. S. & Kumar, S. Heterogeneity of Nucleotide Frequencies Among Evolutionary Lineages and Phylogenetic Inference. *Mol. Biol. Evol.* **20**, 610–621 (2003).
129. Lees, J. A. *et al.* Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study. *Wellcome Open Res.* **3**, (2018).
130. Ogden, T. H. & Rosenberg, M. S. Multiple Sequence Alignment Accuracy and Phylogenetic Inference. *Syst. Biol.* **55**, 314–328 (2006).
131. DROZANSKI, W. J. *Sarcobium lyticum* gen. nov., sp. nov., an Obligate Intracellular Bacterial Parasite of Small Free-Living Amoebae†. *Int. J. Syst. Evol. Microbiol.* **41**, 82–87 (1991).
132. Lamoth, F. & Greub, G. Amoebal pathogens as emerging causal agents of pneumonia. *FEMS Microbiol. Rev.* **34**, 260–280 (2010).
133. Pécastaings, S., Bergé, M., Dubourg, K. M. & Roques, C. Sessile Legionella pneumophila is able to grow on surfaces and generate structured monospecies biofilms. *Biofouling* **26**, 809–819 (2010).
134. Declerck, P. Biofilms: the environmental playground of Legionella pneumophila. *Environ. Microbiol.* **12**, 557–566 (2010).
135. Donlan, R. M. Biofilms: Microbial Life on Surfaces. *Emerg. Infect. Dis.* **8**, 881–890 (2002).
136. Donlan, R. M. Role of Biofilms in Antimicrobial Resistance: *ASAIO J.* **46**, S47–S52 (2000).

137. Schroeder, M., Brooks, B. D. & Brooks, A. E. The Complex Relationship between Virulence and Antibiotic Resistance. *Genes* **8**, 39 (2017).
138. Watnick, P. & Kolter, R. Biofilm, City of Microbes. *J. Bacteriol.* **182**, 2675–2679 (2000).
139. Vervaeren, H., Temmerman, R., Devos, L., Boon, N. & Verstraete, W. Introduction of a boost of *Legionella pneumophila* into a stagnant-water model by heat treatment. *FEMS Microbiol. Ecol.* **58**, 583–592 (2006).
140. Stewart, C. R., Muthye, V. & Cianciotto, N. P. *Legionella pneumophila* persists within biofilms formed by *Klebsiella pneumoniae*, *Flavobacterium* sp., and *Pseudomonas fluorescens* under dynamic flow conditions. *PLoS One* **7**, e50560 (2012).
141. Guerrieri, E. *et al.* Effect of Bacterial Interference on Biofilm Development by *Legionella pneumophila*. *Curr. Microbiol.* **57**, 532–536 (2008).
142. Abu Khweek, A. & Amer, A. O. Factors Mediating Environmental Biofilm Formation by *Legionella pneumophila*. *Front. Cell. Infect. Microbiol.* **8**, (2018).
143. Valentini, M. & Filloux, A. Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from *Pseudomonas aeruginosa* and Other Bacteria. *J. Biol. Chem.* **291**, 12547–12555 (2016).
144. Fazli, M. *et al.* Regulation of *Burkholderia cenocepacia* biofilm formation by RpoN and the c-di-GMP effector BerB. *MicrobiologyOpen* **6**, (2017).
145. Liu, W. *et al.* BsmR degrades c-di-GMP to modulate biofilm formation of nosocomial pathogen *Stenotrophomonas maltophilia*. *Sci. Rep.* **7**, 4665 (2017).
146. Hsieh, M.-L., Hinton, D. M. & Waters, C. M. VpsR and cyclic di-GMP together drive transcription initiation to activate biofilm formation in *Vibrio cholerae*. *Nucleic Acids Res.* **46**, 8876–8887 (2018).
147. Pécastaings, S. *et al.* New insights into *Legionella pneumophila* biofilm regulation by c-di-GMP signaling. *Biofouling* **32**, 935–948 (2016).
148. Liu, R. *et al.* Pyrosequencing analysis of eukaryotic and bacterial communities in faucet biofilms. *Sci. Total Environ.* **435–436**, 124–131 (2012).

149. Storey, M. V., Winiecka-krusnell, J., Ashbolt, N. J. & Stenström, T. The Efficacy of Heat and Chlorine Treatment against Thermotolerant Acanthamoebae and Legionellae. *Scand. J. Infect. Dis.* **36**, 656–662 (2004).
150. Cervero-Aragó, S., Rodríguez-Martínez, S., Puertas-Bennasar, A. & Araujo, R. M. Effect of Common Drinking Water Disinfectants, Chlorine and Heat, on Free Legionella and Amoebae-Associated Legionella. *PLOS ONE* **10**, e0134726 (2015).
151. Declerck, P. *et al.* Replication of Legionella pneumophila in biofilms of water distribution pipes. *Microbiol. Res.* **164**, 593–603 (2009).
152. Murga, R. *et al.* Role of biofilms in the survival of Legionella pneumophila in a model potable-water system. *Microbiology*, **147**, 3121–3126 (2001).
153. Shaheen, M., Scott, C. & Ashbolt, N. J. Long-term persistence of infectious Legionella with free-living amoebae in drinking water biofilms. *Int. J. Hyg. Environ. Health* **222**, 678–686 (2019).
154. Oliver, J. D. Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEMS Microbiol. Rev.* **34**, 415–425 (2010).
155. Steinert, M., Emödy, L., Amann, R. & Hacker, J. Resuscitation of viable but nonculturable Legionella pneumophila Philadelphia JR32 by Acanthamoeba castellanii. *Appl. Environ. Microbiol.* **63**, 2047–2053 (1997).
156. Dietersdorfer, E. *et al.* Starved viable but non-culturable (VBNC) Legionella strains can infect and replicate in amoebae and human macrophages. *Water Res.* **141**, 428–438 (2018).
157. Cervero-Aragó, S. *et al.* Viability and infectivity of viable but nonculturable Legionella pneumophila strains induced at high temperatures. *Water Res.* **158**, 268–279 (2019).
158. Hoogen, J. van den *et al.* Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
159. Bardgett, R. D. & Putten, W. H. van der. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).

160. Ferris, H. Contribution of Nematodes to the Structure and Function of the Soil Food Web. *J. Nematol.* **42**, 63–67 (2010).
161. Deng, X. & Kolesnick, R. *Caenorhabditis elegans* as a model to study sphingolipid signaling. *Biol. Chem.* **396**, 767–773 (2015).
162. Shen, P., Yue, Y. & Park, Y. A living model for obesity and aging research: *Caenorhabditis elegans*. *Crit. Rev. Food Sci. Nutr.* **58**, 741–754 (2018).
163. Hsu, T.-Y., Butler, V. J. & Kao, A. W. The Use of *Caenorhabditis elegans* to Study Progranulin in the Regulation of Programmed Cell Death and Stress Response. *Methods Mol. Biol. Clifton NJ* **1806**, 193–206 (2018).
164. Schmeisser, K. & Parker, J. A. Worms on the spectrum - *C. elegans* models in autism research. *Exp. Neurol.* **299**, 199–206 (2018).
165. Cooper, J. F. & Van Raamsdonk, J. M. Modeling Parkinson's Disease in *C. elegans*. *J. Park. Dis.* **8**, 17–32 (2018).
166. Gammon, D. B. *Caenorhabditis elegans* as an Emerging Model for Virus-Host Interactions. *J. Virol.* **91**, (2017).
167. Khan, F., Jain, S. & Oloketuyi, S. F. Bacteria and bacterial products: Foe and friends to *Caenorhabditis elegans*. *Microbiol. Res.* **215**, 102–113 (2018).
168. Burton, E. A., Pendergast, A. M. & Aballay, A. The *Caenorhabditis elegans* ABL-1 Tyrosine Kinase Is Required for *Shigella flexneri* Pathogenesis. *Appl. Environ. Microbiol.* **72**, 5043–5051 (2006).
169. Aballay, A., Yorgey, P. & Ausubel, F. M. *Salmonella typhimurium* proliferates and establishes a persistent infection in the intestine of *Caenorhabditis elegans*. *Curr. Biol.* **10**, 1539–1542 (2000).
170. Forrester, S., Milillo, S. R., Hoose, W. A., Wiedmann, M. & Schwab, U. Evaluation of The Pathogenicity of *Listeria* spp. in *Caenorhabditis elegans*. *Foodborne Pathog. Dis.* **4**, 67–73 (2007).

171. Mylonakis, E., Ausubel, F. M., Perfect, J. R., Heitman, J. & Calderwood, S. B. Killing of *Caenorhabditis elegans* by *Cryptococcus neoformans* as a model of yeast pathogenesis. *Proc. Natl. Acad. Sci.* **99**, 15675–15680 (2002).
172. Anderson, G. L., Kenney, S. J., Millner, P. D., Beuchat, L. R. & Williams, P. L. Shedding of foodborne pathogens by *Caenorhabditis elegans* in compost-amended and unamended soil. *Food Microbiol.* **23**, 146–153 (2006).
173. Brassinga, A. K. C. *et al.* *Caenorhabditis* is a metazoan host for *Legionella*. *Cell. Microbiol.* **12**, 343–361 (2010).
174. Allombert, J. *et al.* Orchestrated delivery of *Legionella* effectors by the Icm/Dot secretion system. *bioRxiv* 754762 (2019) doi:10.1101/754762.
175. Molofsky, A. B. & Swanson, M. S. Differentiate to thrive: lessons from the *Legionella pneumophila* life cycle. *Mol. Microbiol.* **53**, 29–40 (2004).
176. Hammer, B. K. & Swanson, M. S. Co-ordination of *Legionella pneumophila* virulence with entry into stationary phase by ppGpp. *Mol. Microbiol.* **33**, 721–731 (1999).
177. Richards, A. M., Von Dwingelo, J. E., Price, C. T. & Abu Kwaik, Y. Cellular microbiology and molecular ecology of *Legionella*–amoeba interaction. *Virulence* **4**, 307–314 (2013).
178. Edwards, R. L., Jules, M., Sahr, T., Buchrieser, C. & Swanson, M. S. The *Legionella pneumophila* LetA/LetS Two-Component System Exhibits Rheostat-Like Behavior. *Infect. Immun.* **78**, 2571–2583 (2010).
179. Oliva, G., Sahr, T. & Buchrieser, C. The Life Cycle of *L. pneumophila*: Cellular Differentiation Is Linked to Virulence and Metabolism. *Front. Cell. Infect. Microbiol.* **8**, 3 (2018).
180. Sahr, T. *et al.* Deep sequencing defines the transcriptional map of *L. pneumophila* and identifies growth phase-dependent regulated ncRNAs implicated in virulence. *RNA Biol.* **9**, 503–519 (2012).
181. Faucher, S. P. & Shuman, H. A. Small Regulatory RNA and *Legionella pneumophila*. *Front. Microbiol.* **2**, (2011).

182. Sahr, T. *et al.* The Legionella pneumophila genome evolved to accommodate multiple regulatory mechanisms controlled by the CsrA-system. *PLoS Genet.* **13**, (2017).
183. Gal-Mor, O. & Segal, G. 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 (2003).
184. Hammer, B. K., Tateda, E. S. & Swanson, M. S. A two-component regulator induces the transmission phenotype of stationary-phase Legionella pneumophila. *Mol. Microbiol.* **44**, 107–118 (2002).
185. Qin, T., Zhou, H., Ren, H. & Liu, W. Distribution of Secretion Systems in the Genus Legionella and Its Correlation with Pathogenicity. *Front. Microbiol.* **8**, (2017).
186. Mondino, S. *et al.* Legionnaires' Disease: State of the Art Knowledge of Pathogenesis Mechanisms of Legionella. *Annu. Rev. Pathol. Mech. Dis.* **15**, 439–466 (2020).
187. Jeong, K. C., Ghosal, D., Chang, Y.-W., Jensen, G. J. & Vogel, J. P. Polar delivery of Legionella type IV secretion system substrates is essential for virulence. *Proc. Natl. Acad. Sci.* **114**, 8077–8082 (2017).
188. Berger, K. H. & Isberg, R. R. Two distinct defects in intracellular growth complemented by a single genetic locus in Legionella pneumophila. *Mol. Microbiol.* **7**, 7–19 (1993).
189. Marra, A., Blander, S. J., Horwitz, M. A. & Shuman, H. A. Identification of a Legionella pneumophila locus required for intracellular multiplication in human macrophages. *Proc. Natl. Acad. Sci.* **89**, 9607–9611 (1992).
190. Gomez-Valero, L., Chiner-Oms, A., Comas, I. & Buchrieser, C. Evolutionary Dissection of the Dot/Icm System Based on Comparative Genomics of 58 Legionella Species. *Genome Biol. Evol.* **11**, 2619–2632 (2019).
191. Vincent, C. D. & Vogel, J. P. The Legionella pneumophila IcmS–LvgA protein complex is important for Dot/Icm-dependent intracellular growth. *Mol. Microbiol.* **61**, 596–613 (2006).

192. Xu, J. *et al.* Structural insights into the roles of the IcmS–IcmW complex in the type IVb secretion system of *Legionella pneumophila*. *Proc. Natl. Acad. Sci.* **114**, 13543–13548 (2017).
193. Meir, A., Chetrit, D., Liu, L., Roy, C. R. & Waksman, G. *Legionella* DotM structure reveals a role in effector recruiting to the Type 4B secretion system. *Nat. Commun.* **9**, 507 (2018).
194. Duménil, G., Montminy, T. P., Tang, M. & Isberg, R. R. IcmR-regulated Membrane Insertion and Efflux by the *Legionella pneumophila* IcmQ Protein. *J. Biol. Chem.* **279**, 4686–4695 (2004).
195. Vincent, C. D. *et al.* Identification of the core transmembrane complex of the *Legionella* Dot/Icm type IV secretion system. *Mol. Microbiol.* **62**, 1278–1291 (2006).
196. VanRheenen, S. M., Duménil, G. & Isberg, R. R. IcmF and DotU Are Required for Optimal Effector Translocation and Trafficking of the *Legionella pneumophila* Vacuole. *Infect. Immun.* **72**, 5972–5982 (2004).
197. Ghosal, D. *et al.* Molecular architecture, polar targeting and biogenesis of the *Legionella* Dot/Icm T4SS. *Nat. Microbiol.* **4**, 1173–1182 (2019).
198. Chetrit, D., Hu, B., Christie, P. J., Roy, C. R. & Liu, J. A Unique Cytoplasmic ATPase Complex Defines the *Legionella pneumophila* Type IV Secretion Channel. *Nat. Microbiol.* **3**, 678–686 (2018).
199. Nagai, H. & Kubori, T. Type IVB Secretion Systems of *Legionella* and Other Gram-Negative Bacteria. *Front. Microbiol.* **2**, (2011).
200. Kuroda, T. *et al.* Molecular and structural analysis of *Legionella* DotI gives insights into an inner membrane complex essential for type IV secretion. *Sci. Rep.* **5**, (2015).
201. Kumar, N. *et al.* Analyzing the role of CagV, a VirB8 homolog of the type IV secretion system of *Helicobacter pylori*. *FEBS Open Bio* **7**, 915–933 (2017).
202. Purcell, M. & Shuman, H. A. The *Legionella pneumophila* icmGCDJBFGenes Are Required for Killing of Human Macrophages. *Infect. Immun.* **66**, 2245–2255 (1998).
203. Vogel, J. P., Andrews, H. L., Wong, S. K. & Isberg, R. R. Conjugative Transfer by the Virulence System of *Legionella pneumophila*. *Science* **279**, 873–876 (1998).

204. Nagai, H. & Roy, C. R. The DotA protein from *Legionella pneumophila* is secreted by a novel process that requires the Dot/Icm transporter. *EMBO J.* **20**, 5962–5970 (2001).
205. Lu, J. *et al.* *Legionella pneumophila* Transcriptional Response following Exposure to CuO Nanoparticles. *Appl. Environ. Microbiol.* **79**, 2713–2720 (2013).
206. Matthews, M. & Roy, C. R. Identification and Subcellular Localization of the *Legionella pneumophila* IcmX Protein: a Factor Essential for Establishment of a Replicative Organelle in Eukaryotic Host Cells. *Infect. Immun.* **68**, 3971–3982 (2000).
207. Bitar, D. M., Molmeret, M. & Kwaik, Y. A. Structure–function analysis of the C-terminus of IcmT of *Legionella pneumophila* in pore formation-mediated egress from macrophages. *FEMS Microbiol. Lett.* **242**, 177–184 (2005).
208. Molmeret, M. *et al.* The C-terminus of IcmT is essential for pore formation and for intracellular trafficking of *Legionella pneumophila* within *Acanthamoeba polyphaga*. *Mol. Microbiol.* **43**, 1139–1150 (2002).
209. Segal, G., Russo, J. J. & Shuman, H. A. Relationships between a new type IV secretion system and the icm/dot virulence system of *Legionella pneumophila*. *Mol. Microbiol.* **34**, 799–809 (1999).
210. Kubori, T. & Nagai, H. The Type IVB secretion system: an enigmatic chimera. *Curr. Opin. Microbiol.* **29**, 22–29 (2016).
211. Gomez-Valero, L. *et al.* More than 18,000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. *Proc. Natl. Acad. Sci.* **116**, 2265–2273 (2019).
212. Currie, S. L. & Beattie, T. K. Compost and *Legionella longbeachae*: an emerging infection?: *Perspect. Public Health* (2015) doi:10.1177/1757913915611162.
213. Kenagy, E. *et al.* Risk Factors for *Legionella longbeachae* Legionnaires' Disease, New Zealand. *Emerg. Infect. Dis.* **23**, 1148–1154 (2017).

214. Thornley, C. N. *et al.* Legionella longbeachae detected in an industrial cooling tower linked to a legionellosis outbreak, New Zealand, 2015; possible waterborne transmission? *Epidemiol. Infect.* **145**, 2382–2389 (2017).
215. Alli, O. A. T., Zink, S., von Lackum, N. K. & Abu-Kwaik, Y. Comparative assessment of virulence traits in Legionella spp. *Microbiology*, **149**, 631–641 (2003).
216. O’Connell, W. A., Dhand, L. & Cianciotto, N. P. Infection of macrophage-like cells by Legionella species that have not been associated with disease. *Infect. Immun.* **64**, 4381–4384 (1996).
217. Zhu, W. *et al.* Comprehensive Identification of Protein Substrates of the Dot/Icm Type IV Transporter of Legionella pneumophila. *PLoS ONE* **6**, e17638 (2011).
218. Burstein, D. *et al.* Genome-Scale Identification of Legionella pneumophila Effectors Using a Machine Learning Approach. *PLoS Pathog.* **5**, (2009).
219. Lurie-Weinberger, M. N. *et al.* The origins of eukaryotic-like proteins in Legionella pneumophila. *Int. J. Med. Microbiol.* **300**, 470–481 (2010).
220. Franco, I. S., Shuman, H. A. & Charpentier, X. The perplexing functions and surprising origins of Legionella pneumophila type IV secretion effectors. *Cell. Microbiol.* **11**, 1435–1443 (2009).
221. Moliner, C., Raoult, D. & Fournier, P.-E. Evidence of horizontal gene transfer between amoeba and bacteria. *Clin. Microbiol. Infect.* **15**, 178–180 (2009).
222. Zade, A., Sengupta, M. & Kondabagil, K. Extensive in silico analysis of Mimivirus coded Rab GTPase homolog suggests a possible role in virion membrane biogenesis. *Front. Microbiol.* **6**, (2015).
223. Colicelli, J. Human RAS Superfamily Proteins and Related GTPases. *Sci. STKE Signal Transduct. Knowl. Environ.* **2004**, RE13 (2004).
224. Wennerberg, K., Rossman, K. L. & Der, C. J. The Ras superfamily at a glance. *J. Cell Sci.* **118**, 843–846 (2005).

225. Cazalet, C. *et al.* Analysis of the *Legionella longbeachae* Genome and Transcriptome Uncovers Unique Strategies to Cause Legionnaires' Disease. *PLoS Genet.* **6**, e1000851 (2010).
226. Daubin, V. & Ochman, H. Bacterial Genomes as New Gene Homes: The Genealogy of ORFans in *E. coli*. *Genome Res.* **14**, 1036–1042 (2004).
227. Aguilera, G., Refrégier, G., Yockteng, R., Fournier, E. & Giraud, T. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect. Genet. Evol.* **9**, 656–670 (2009).
228. Gomez-Valero, L., Rusniok, C., Cazalet, C. & Buchrieser, C. Comparative and Functional Genomics of *Legionella* Identified Eukaryotic Like Proteins as Key Players in Host–Pathogen Interactions. *Front. Microbiol.* **2**, (2011).
229. Brüggemann, H., Cazalet, C. & Buchrieser, C. Adaptation of *Legionella pneumophila* to the host environment: role of protein secretion, effectors and eukaryotic-like proteins. *Curr. Opin. Microbiol.* **9**, 86–94 (2006).
230. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct. Funct. Bioinforma.* **17**, 363–374 (1993).
231. Ankyrin repeat (IPR002110) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR002110/>.
232. Tetratricopeptide-like helical domain superfamily (IPR011990) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR011990/>.
233. SET domain (IPR001214) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001214/>.
234. Sec7 domain (IPR000904) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000904/>.
235. Serine-threonine/tyrosine-protein kinase, catalytic domain (IPR001245) - InterPro entry - InterPro. <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001245/>.

236. U box domain (IPR003613) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR003613/>.
237. F-box domain (IPR001810) - InterPro entry - InterPro.
<https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001810/>.
238. Chauhan, D. & Shames, S. R. Pathogenicity and Virulence of Legionella: Intracellular replication and host response. *Virulence* **12**, 1122–1144.
239. Fountain, A., Inpanathan, S., Alves, P., Verdawala, M. B. & Botelho, R. J. Phagosome maturation in macrophages: Eat, digest, adapt, and repeat. *Adv. Biol. Regul.* **82**, 100832 (2021).
240. Robinson, C. G. & Roy, C. R. Attachment and fusion of endoplasmic reticulum with vacuoles containing Legionella pneumophila. *Cell. Microbiol.* **8**, 793–805 (2006).
241. Gomez-Valero, L. & Buchrieser, C. Genome dynamics in Legionella: the basis of versatility and adaptation to intracellular replication. *Cold Spring Harb. Perspect. Med.* **3**, a009993 (2013).
242. Segal, G. & Shuman, H. A. Legionella pneumophila utilizes the same genes to multiply within Acanthamoeba castellanii and human macrophages. *Infect. Immun.* **67**, 2117–2124 (1999).
243. Isberg, R. R., O'Connor, T. & Heidtman, M. The Legionella pneumophila replication vacuole: making a cozy niche inside host cells. *Nat. Rev. Microbiol.* **7**, 13–24 (2009).
244. Newton, H. J., Ang, D. K. Y., van Driel, I. R. & Hartland, E. L. Molecular Pathogenesis of Infections Caused by Legionella pneumophila. *Clin. Microbiol. Rev.* **23**, 274–298 (2010).
245. Huang, L. *et al.* The E Block motif is associated with Legionella pneumophila translocated substrates. *Cell. Microbiol.* **13**, 227–245 (2011).
246. Qiu, J. & Luo, Z.-Q. Effector translocation by the Legionella Dot/Icm type IV secretion system. *Curr. Top. Microbiol. Immunol.* **376**, 103–115 (2013).
247. Ensminger, A. W. & Isberg, R. R. Legionella pneumophila Dot/Icm translocated substrates: a sum of parts. *Curr. Opin. Microbiol.* **12**, 67–73 (2009).
248. Finsel, I. & Hilbi, H. Formation of a pathogen vacuole according to Legionella pneumophila: how to kill one bird with many stones. *Cell. Microbiol.* **17**, 935–950 (2015).

249. Isaac, D. T. & Isberg, R. Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol.* **9**, 343–359 (2014).
250. Pearson, J. S., Zhang, Y., Newton, H. J. & Hartland, E. L. Post-modern pathogens: surprising activities of translocated effectors from *E. coli* and *Legionella*. *Curr. Opin. Microbiol.* **23**, 73–79 (2015).
251. Rolando, M. & Buchrieser, C. *Legionella pneumophila* type IV effectors hijack the transcription and translation machinery of the host cell. *Trends Cell Biol.* **24**, 771–778 (2014).
252. Burstein, D. *et al.* Genomic analysis of 38 *Legionella* species identifies large and diverse effector repertoires. *Nat. Genet.* **48**, 167–175 (2016).
253. Gomez-Valero, L. *et al.* Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics* **12**, 536 (2011).
254. Bardill, J. P., Miller, J. L. & Vogel, J. P. IcmS-dependent translocation of SdeA into macrophages by the *Legionella pneumophila* type IV secretion system. *Mol. Microbiol.* **56**, 90–103 (2005).
255. Jeong, K. C., Sexton, J. A. & Vogel, J. P. Spatiotemporal regulation of a *Legionella pneumophila* T4SS substrate by the metaeffector SidJ. *PLoS Pathog.* **11**, e1004695 (2015).
256. O'Connor, T. J., Adepoju, Y., Boyd, D. & Isberg, R. R. Minimization of the *Legionella pneumophila* genome reveals chromosomal regions involved in host range expansion. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14733–14740 (2011).
257. Chien, M. *et al.* The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305**, 1966–1968 (2004).
258. Ensminger, A. W. *Legionella pneumophila*, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world. *Curr. Opin. Microbiol.* **29**, 74–80 (2016).
259. Park, J. M., Ghosh, S. & O'Connor, T. J. Combinatorial selection in amoebal hosts drives the evolution of the human pathogen *Legionella pneumophila*. *Nat. Microbiol.* **5**, 599–609 (2020).

260. Joshi, A. D. & Swanson, M. S. Secrets of a successful pathogen: legionella resistance to progression along the autophagic pathway. *Front. Microbiol.* **2**, 138 (2011).
261. Mishra, A. K., Campo, C. M. D., Collins, R. E., Roy, C. R. & Lambright, D. G. The Legionella pneumophila GTPase Activating Protein LepB Accelerates Rab1 Deactivation by a Non-canonical Hydrolytic Mechanism *. *J. Biol. Chem.* **288**, 24000–24011 (2013).
262. Wasilko, D. J., Huang, Q. & Mao, Y. Insights into the ubiquitin transfer cascade catalyzed by the Legionella effector SidC. *eLife* **7**, e36154 (2018).
263. de Felipe, K. S. *et al.* Legionella eukaryotic-like type IV substrates interfere with organelle trafficking. *PLoS Pathog.* **4**, e1000117 (2008).
264. Allombert, J., Vianney, A. & Charpentier, X. Monitoring Effector Translocation using the TEM-1 Beta-Lactamase Reporter System. *Methods Mol. Biol. Clifton NJ* **1615**, 489–499 (2017).
265. Reimann, C. & Haas, D. The istA gene of insertion sequence IS21 is essential for cleavage at the inner 3' ends of tandemly repeated IS21 elements in vitro. *EMBO J.* **9**, 4055–4063 (1990).
266. Arias-Palomo, E. & Berger, J. M. An Atypical AAA+ ATPase Assembly Controls Efficient Transposition through DNA Remodeling and Transposase Recruitment. *Cell* **162**, 860–871 (2015).
267. Spínola-Amilibia, M., Araújo-Bazán, L., de la Gándara, Á., Berger, J. M. & Arias-Palomo, E. IS21 family transposase cleaved donor complex traps two right-handed superhelical crossings. *Nat. Commun.* **14**, 2335 (2023).
268. Dillon, M. M., Sung, W., Sebra, R., Lynch, M. & Cooper, V. S. Genome-Wide Biases in the Rate and Molecular Spectrum of Spontaneous Mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Mol. Biol. Evol.* **34**, 93–109 (2017).
269. Durieux, I. *et al.* Diverse conjugative elements silence natural transformation in Legionella species. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18613–18618 (2019).
270. Doyle, R. M. & Heuzenroeder, M. W. A mutation in an ompR-like gene on a Legionella longbeachae serogroup 1 plasmid attenuates virulence. *Int. J. Med. Microbiol. IJMM* **292**, 227–239 (2002).

271. Schator, D. *et al.* Legionella para-effectors target chromatin and promote bacterial replication. *Nat. Commun.* **14**, 2154 (2023).
272. Appelt, S. & Heuner, K. The Flagellar Regulon of Legionella—A Review. *Front. Cell. Infect. Microbiol.* **7**, (2017).
273. Zusman, T., Gal-Mor, O. & Segal, G. Characterization of a Legionella pneumophila relA Insertion Mutant and Roles of RelA and RpoS in Virulence Gene Expression. *J. Bacteriol.* **184**, 67–75 (2002).
274. Rasis, M. & Segal, G. The LetA-RsmYZ-CsrA regulatory cascade, together with RpoS and PmrA, post-transcriptionally regulates stationary phase activation of Legionella pneumophila Icm/Dot effectors. *Mol. Microbiol.* **72**, 995–1010.
275. Sahr, T. *et al.* Two small ncRNAs jointly govern virulence and transmission in Legionella pneumophila. *Mol. Microbiol.* **72**, 741–762 (2009).
276. Heeb, S. & Haas, D. Regulatory roles of the GacS/GacA two-component system in plant-associated and other gram-negative bacteria. *Mol. Plant-Microbe Interact. MPMI* **14**, 1351–1363 (2001).
277. Humair, B., Wackwitz, B. & Haas, D. GacA-controlled activation of promoters for small RNA genes in Pseudomonas fluorescens. *Appl. Environ. Microbiol.* **76**, 1497–1506 (2010).
278. Gropp, M., Strausz, Y., Gross, M. & Glaser, G. Regulation of Escherichia coli RelA requires oligomerization of the C-terminal domain. *J. Bacteriol.* **183**, 570–579 (2001).
279. David, S. *et al.* Seeding and Establishment of Legionella pneumophila in Hospitals: Implications for Genomic Investigations of Nosocomial Legionnaires' Disease. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **64**, 1251–1259 (2017).
280. Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C. & San Millán, Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* **19**, 347–359 (2021).

281. Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat. Rev. Microbiol.* **8**, 260–271 (2010).
282. Jurénas, D., Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol.* **20**, 335–350 (2022).
283. Wang, H. *et al.* Increased plasmid copy number is essential for *Yersinia* T3SS function and virulence. *Science* **353**, 492–495 (2016).
284. Schneiders, S. *et al.* Spatiotemporal Variations in Growth Rate and Virulence Plasmid Copy Number during *Yersinia pseudotuberculosis* Infection. *Infect. Immun.* **89**, e00710-20 (2021).
285. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27**, 145–163 (2003).
286. Fang, C. & Zhang, Y. Bacterial MerR family transcription regulators: activation by distortion. *Acta Biochim. Biophys. Sin.* **54**, 25–36 (2022).
287. Oliva, G., Sahr, T., Rolando, M., Knoth, M. & Buchrieser, C. A Unique cis-Encoded Small Noncoding RNA Is Regulating *Legionella pneumophila* Hfq Expression in a Life Cycle-Dependent Manner. *mBio* **8**, (2017).
288. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8**, 185–195 (2010).
289. Pilla, G., McVicker, G. & Tang, C. M. Genetic plasticity of the *Shigella* virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. *PLOS Genet.* **13**, e1007014 (2017).
290. Yang, C. *et al.* Chlamydial Lytic Exit from Host Cells Is Plasmid Regulated. *mBio* **6**, e01648-01615 (2015).
291. Patton, M. J. *et al.* Plasmid Negative Regulation of CPAF Expression Is Pgp4 Independent and Restricted to Invasive *Chlamydia trachomatis* Biovars. *mBio* **9**, e02164-17 (2018).
292. Ronin, I., Katsowich, N., Rosenshine, I. & Balaban, N. Q. A long-term epigenetic memory switch controls bacterial virulence bimodality. *eLife* **6**, e19599 (2017).

293. Vial, L. & Hommais, F. Plasmid-chromosome cross-talks. *Environ. Microbiol.* **22**, 540–556 (2020).
294. Billane, K., Harrison, E., Cameron, D. & Brockhurst, M. A. Why do plasmids manipulate the expression of bacterial phenotypes? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **377**, 20200461 (2022).
295. Coulson, G. B. *et al.* Transcriptome reprogramming by plasmid-encoded transcriptional regulators is required for host niche adaption of a macrophage pathogen. *Infect. Immun.* **83**, 3137–3145 (2015).
296. Di Venanzio, G. *et al.* Urinary tract colonization is enhanced by a plasmid that regulates uropathogenic *Acinetobacter baumannii* chromosomal genes. *Nat. Commun.* **10**, 2763 (2019).
297. Weber, B. S., Ly, P. M., Irwin, J. N., Pukatzki, S. & Feldman, M. F. A multidrug resistance plasmid contains the molecular switch for type VI secretion in *Acinetobacter baumannii*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9442–9447 (2015).
298. Di Venanzio, G. *et al.* Multidrug-resistant plasmids repress chromosomally encoded T6SS to enable their dissemination. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1378–1383 (2019).
299. Sahr, T. *et al.* Two small ncRNAs jointly govern virulence and transmission in *Legionella pneumophila*. *Mol. Microbiol.* **72**, 741–762 (2009).
300. Laville, J. *et al.* Global control in *Pseudomonas fluorescens* mediating antibiotic synthesis and suppression of black root rot of tobacco. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 1562–1566 (1992).
301. Rich, J. J., Kinscherf, T. G., Kitten, T. & Willis, D. K. Genetic evidence that the *gacA* gene encodes the cognate response regulator for the *lemA* sensor in *Pseudomonas syringae*. *J. Bacteriol.* **176**, 7468–7475 (1994).
302. Grewal, S. I., Han, B. & Johnstone, K. Identification and characterization of a locus which regulates multiple functions in *Pseudomonas tolaasii*, the cause of brown blotch disease of *Agaricus bisporus*. *J. Bacteriol.* **177**, 4658–4668 (1995).

303. Bull, C. T. *et al.* Characterization of spontaneous *gacS* and *gacA* regulatory mutants of *Pseudomonas fluorescens* biocontrol strain CHAO. *Antonie Van Leeuwenhoek* **79**, 327–336 (2001).
304. Harrison, E., Guymer, D., Spiers, A. J., Paterson, S. & Brockhurst, M. A. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol. CB* **25**, 2034–2039 (2015).
305. Bird, S. M. *et al.* Compensatory mutations reducing the fitness cost of plasmid carriage occur in plant rhizosphere communities. *FEMS Microbiol. Ecol.* **99**, fiad027 (2023).
306. Darmon, E. & Leach, D. R. F. Bacterial genome instability. *Microbiol. Mol. Biol. Rev. MMBR* **78**, 1–39 (2014).
307. Godeux, A.-S. Un agent nosocomial humain et animal : étude de la dynamique du transfert de gènes de résistance aux antibiotiques chez *Acinetobacter*. (Université de Lyon, 2020).
308. Hall, J. P. J., Brockhurst, M. A. & Harrison, E. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, 20160424 (2017).
309. Johnsborg, O., Eldholm, V. & Håvarstein, L. S. Natural genetic transformation: prevalence, mechanisms and function. *Res. Microbiol.* **158**, 767–778 (2007).
310. Dubnau, D. & Blokesch, M. Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annu. Rev. Genet.* **53**, 217–237 (2019).
311. Piepenbrink, K. H. DNA Uptake by Type IV Filaments. *Front. Mol. Biosci.* **6**, (2019).
312. Ellison, C. K. *et al.* Retraction of DNA-bound type IV competence pili initiates DNA uptake during natural transformation in *Vibrio cholerae*. *Nat. Microbiol.* **3**, 773–780 (2018).
313. Seitz, P. *et al.* ComEA Is Essential for the Transfer of External DNA into the Periplasm in Naturally Transformable *Vibrio cholerae* Cells. *PLOS Genet.* **10**, e1004066 (2014).
314. Draskovic, I. & Dubnau, D. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds. *Mol. Microbiol.* **55**, 881–896 (2005).

315. Attaiech, L. *et al.* Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. *PLoS Genet.* **7**, e1002156 (2011).
316. Bergé, M., Mortier-Barrière, I., Martin, B. & Claverys, J.-P. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol. Microbiol.* **50**, 527–536 (2003).
317. Mortier-Barrière, I. *et al.* A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell* **130**, 824–836 (2007).
318. Nero, T. M. *et al.* ComM is a hexameric helicase that promotes branch migration during natural transformation in diverse Gram-negative species. *Nucleic Acids Res.* **46**, 6099–6111 (2018).
319. Huang, M. *et al.* The activation and limitation of the bacterial natural transformation system: The function in genome evolution and stability. *Microbiol. Res.* **252**, 126856 (2021).
320. Buchrieser, C. & Charpentier, X. Induction of competence for natural transformation in *Legionella pneumophila* and exploitation for mutant construction. *Methods Mol. Biol. Clifton NJ* **954**, 183–195 (2013).
321. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J.-P. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**, 181–196 (2014).
322. Attaiech, L. *et al.* Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8813–8818 (2016).
323. Rizzardi, K. *et al.* *Legionella norrlandica* sp. nov., isolated from the biopurification systems of wood processing plants. *Int. J. Syst. Evol. Microbiol.* **65**, 598–603 (2015).
324. Dobrowsky, P. H., Khan, S., Cloete, T. E. & Khan, W. Molecular detection of *Acanthamoeba* spp., *Naegleria fowleri* and *Vermamoeba (Hartmannella) vermiformis* as vectors for *Legionella* spp. in untreated and solar pasteurized harvested rainwater. *Parasit. Vectors* **9**, 539 (2016).

325. Furugen, M. *et al.* Legionella pneumophila infection induces programmed cell death, caspase activation, and release of high-mobility group box 1 protein in A549 alveolar epithelial cells: inhibition by methyl prednisolone. *Respir. Res.* **9**, 39 (2008).
326. Nair, J., Rouse, D. A., Bai, G. H. & Morris, S. L. The rpsL gene and streptomycin resistance in single and multiple drug-resistant strains of Mycobacterium tuberculosis. *Mol. Microbiol.* **10**, 521–527 (1993).
327. Fontaine, L., Wahl, A., Fléchar, M., Mignolet, J. & Hols, P. Regulation of competence for natural transformation in streptococci. *Infect. Genet. Evol.* **33**, 343–360 (2015).
328. Springer, B. *et al.* Mechanisms of Streptomycin Resistance: Selection of Mutations in the 16S rRNA Gene Conferring Resistance. *Antimicrob. Agents Chemother.* **45**, 2877–2884 (2001).
329. Juan, P.-A., Attaiech, L. & Charpentier, X. Natural transformation occurs independently of the essential actin-like MreB cytoskeleton in Legionella pneumophila. *Sci. Rep.* **5**, 16033 (2015).