UPPSALA
UNIVERSITET

Evolutionary and functional analysis of RavC, a *Legionellales*-wide conserved effector

Emma Brodin

Master Degree Project in Infection Biology, 45 credits. November 2021- June 2022
Department: Department of Medical Biochemistry and Microbiology (IMBIM)
Supervisor: Lionel Guy
Co-Supervisor: Celestine Chi

Table of contents

Abstract 4
Popular science summary 5
1. Introduction 6
1.1 Epidemiology 6
1.2 Infection of host cells and the importance of effectors 6
1.3 Evolutionary history of RavC 7
2. Aim 8
3. Materials and methods 8
3.1 Buffers 8
3.2 Bacterial strains and cultures 8
3.3 Cell cultures used for infections 9
3.4 Extraction of genomic DNA from *L. pneumophila* and pGEM-Kanamycin-mazF plasmid from *E. coli* 9
3.5 Double-joint PCR to get the first step construct for transformation 9
3.6 Joint PCR to get the second step construct for transformation 10
3.7 Natural transformation to achieve ΔRavC mutant 10
3.8 PCR analysis of mutant 11
3.9 Growth curve experiments 11
3.10 Infection of *L. pneumophila* into host cells 12
3.11 Competition assay of ΔRavC and WT in U937 cells 12
3.12 Extraction of RNA from different time points during growth 12
3.13 RT-qPCR 12
3.14 Expression of RavC for protein analysis 13
3.15 Harvest and extraction of RavC 13
3.16 Immobilized Metal Affinity Chromatography (IMAC) 14
3.17 SDS-PAGE 14
3.18 Phylogenetic tree of RavC 14
4. Results 14
4.1 Growth curves of SYFP2 14
4.2 Growth of ΔRavC and WT 15
4.3 Competition assay 16
4.4 RT-qPCR 18
4.5 Phylogenetics 19
4.6 Purification of RavC 19
5. Discussion 20
6. Acknowledgements 24

7. References 25
8. Appendix A 30
9. Appendix B 32
10. Appendix C 35

Abstract

Legionella have adapted to an intracellular life cycle in protozoans. To infect and replicate within their host cells, bacteria within the *Legionella* genus have an arsenal of effector proteins which they release into the host via their type IV B Secretion System. There are over 18 000 effectors among the 50 *Legionella* species, but only nine of them are conserved. One of them, RavC, has presumably been acquired in the last *Legionellales* common ancestor (LLCA), circa two billion years ago. Little is known about the function and origin of RavC and why it is so conserved within the genus. The aim of this study is therefore to study the evolution and function of RavC in *Legionella pneumophila*. A phylogenetic tree of RavC homologs suggests that RavC was horizontally transferred from eukaryotes to all host-adapted bacteria one single time, prior to LLCA. A mutant lacking *ravC* (ΔRavC) was created and used for infections in *Acanthamoeba castellanii* and U937 cells. Additionally, a competition assay was performed in U937 cells comparing the growth of ΔRavC against its wild type (WT). No differences in growth rate are noticeable during infection or competition in U937 cells. Surprisingly, ΔRavC displays an increased growth in *A. castellanii*, a natural *Legionella* host.

Keywords: *Legionella pneumophila*; effector protein; phylogenetics; knock-out mutant; infection; competition assay

Popular science summary

Legionellae are bacteria that can infect and replicate within other cells, such as amoebae and human macrophages. When infecting humans, *Legionella pneumophila* will most commonly cause a severe form of pneumonia. To infect their hosts, *Legionella* use an arsenal of effector proteins that they release into the host cell to hijack important cellular mechanisms. Among the 50 *Legionella* species there are 18 000 effectors but only nine of them are conserved. One of them – RavC, is believed to have been acquired already in the last *Legionellales* common ancestor (LLCA), more than two billion years ago. Since RavC is one of few conserved effectors, it must have played an important role for the first host-adapted bacteria to infect their host cells. However, very little is known about RavC and its function. Therefore, the aim for this project was to learn more about the function and evolutionary history of RavC and why it still is conserved among all *Legionella* species.

To investigate where RavC might originate from, a phylogenetic tree of proteins with the same evolutionary origin as RavC was made to get an understanding of the evolutionary history of the effector. When analysing the tree, all host-adapted bacteria (*Legionella, Chlamydia, Coxiella, Rickettsia*) where RavC was present are found in the same clade. As hypothesised, this suggests that RavC was horizontally transferred only once from eukaryotes to all host-adapted bacteria, more than two billion years ago.

A mutant of *L. pneumophila* lacking the gene for RavC was created to see what role RavC plays in growth and infection of host cells. The mutant and the wild type of *L. pneumophila* was then used to infect *Acanthamoeba castellanii* and macrophages to compare the growth between the two bacterial strains. Contrary to the hypothesis, the mutant showed a higher rate of growth compared to the wild type when infecting *A. castellanii*. However, no clear difference was seen in growth when infecting macrophages. One possible explanation for this may be that *L. pneumophila* is more adapted to a life cycle within *A. castellanii* compared to macrophages or that RavC only is of importance during infection of certain host species.

In conclusion, RavC seems to have been gained more than two billion years ago and must have played an important role for the host-adapted bacteria, including *Legionella*, at least during some point of evolution. No difference in growth could be seen between the mutant and the wild type during infection in macrophages in this study. However, there was an increase of growth in the mutant compared to the wild type in *A.* castellanii. This may result from the fact that amoeba more commonly act as hosts for *L. pneumophila* compared to macrophages.

1. Introduction

Bacteria within the *Legionella* genus are gram-negative bacteria that have adapted to a facultatively intracellular lifestyle (1,2). The bacteria can survive and possibly reproduce outside of host cells, but most commonly protozoa act as hosts for *Legionella* to replicate. Some of the first confirmed protist hosts of *Legionella pneumophila* were amoeba within the genera *Acanthamoeba* and *Naegleria* (3–5). To infect their hosts, *L. pneumophila* will adhere to the host cell which triggers phagocytosis or micropinocytosis (6). Once inside the host cell, the bacterium is found within a vacuole where it can multiply and replicate by hi-jacking cellular pathways by releasing effector proteins. After a while, the level of nutrients will decrease inside the host, and this will cause the bacteria to emerge from the vacuole and out from the host cell where they are prepared to infect a new host (6).

L. pneumophila can also be found in abundant numbers in water, including drinking water (7). There are also examples of the bacterium being found in soil, sediment and biofilms (4). Protists within the phyla *Choanozoa* and *Euglenozoa* have been shown to be present when studying *Legionella*-containing biofilms in engineered water systems, thus indicating that these protists may interact with *Legionella* (8). However, the interaction between *Legionella* and other organisms found in biofilms is not well understood today, but it is hypothesized that the other organisms play a role in supplying the bacteria with nutrients needed for them to survive (9).

1.1 Epidemiology

The main route of infection with *Legionella* is through inhalation of the bacteria into the airways (10). Risk factors are, for example, immunosuppression, male gender, smoking, increasing age, and diabetes (11). When infecting humans, *Legionella* can cause Legionnaires' disease, which is a severe form of pneumonia (12). *L. pneumophila,* followed by *L. longbeachae,* are the most common human pathogens among the *Legionella* species (13,14). The bacteria are able to infect and replicate within macrophages found in the alveoli and, eventually, the replication results in the death of macrophages, which then leads to Legionnaires' disease (14). The infected patients will show common pneumonia-symptoms, such as shortness of breath, fever and coughing (10). However, unlike community-acquired pneumonia, patients infected with *Legionella* often also have gastro-intestinal and neurological symptoms. Pontiac fever is another form of disease caused by *Legionella* that show similar symptoms to Legionnaires' disease, but without the pneumonia (10).

A seasonal pattern is seen for Legionnaire's disease, where most cases occur during late summer and beginning of autumn (15,16). One suggested explanation to this is the warmer and more humid weather seen during these months (15). Contaminated cooling towers seem to be the main source of *Legionella,* but other water systems, such as fountains and whirlpools, where aerosol is created can also spread the disease (15,17). Moreover, there are studies which have shown cases linked to potting soil as well, but these cases are mainly caused by *L. longbeachae* (17,18). As of today, only one suggested case of human-to-human transmission has been reported, this one being in Portugal where a person is suspected of having transmitted *L. pneumophila* to his mother (19). Apart from that, all reported cases are mainly related to aerosol coming from water systems or potting soil, as previously mentioned.

1.2 Infection of host cells and the importance of effectors

When *Legionella* infects host cells, they secrete effectors via their type IV B Secretion System (T4BSS), also referred to as Dot/Icm (defective organelle trafficking/intracellular multiplication), to be able to survive and replicate within the host (20,21). The Dot/Icm complex found among the different *Legionella* species is, due to its mode of action, under a strong

evolutionary pressure (22). However, the T4SS must still be able to exert its function of secreting effectors into the host cell but also adapt to new niches and avoid host defence. This means that T4SS is under both positive and negative selection (22). For example, the proteins IcmQ and IcmR undergo fast and positive evolution and are essential for *Legionella* during infection where they insert pores into the host cells membrane (22,23). The proteins that make up the base of the cytoplasmic part of Dot/Icm, such as DotO and DotB, will instead be under a negative selection since they most likely are essential for the function of the entire complex (22).

Once *Legionella* is inside the host cell, it is able to manipulate different pathways by secreting its effectors via T4BSS into the host to allow for growth of the bacteria (14). The secretion of effectors is triggered when the bacterium is phagocytosed by the host and the effectors then allow for *Legionella* to avoid cell degradation by instead forming a vacuole, also known as a *Legionella*-containing vacuole (LCV) (24). The LCV is needed for the survival and replication of the bacteria within the host and for *Legionella* to be able to produce the proteins needed, amino acids from the host will be transferred into the LCV where the bacteria can use them for protein synthesis (14,25). It is also likely that *Legionella* can recruit mitochondria and endoplasmic reticulum (ER) vesicles to the LCV, which promotes the growth of the bacteria (24). Moreover, the effectors will also prevent fusion of the LCV with the lysosome and can also hinder acidification inside the vacuole by blocking the host's vacuolar ATPase (24). There are also effectors secreted which contain eukaryotic F-box and U-box domains, which indicates that *Legionella* can alter the host cell's ubiquitin system (26).

In 2004, the entire genome of *L. pneumophila* was sequenced and it was found that the bacterium contains a high number of genes that encode proteins that are either highly similar to eukaryotic proteins or may have eukaryotic domains on them (14,27,28). This indicates that there are protein-protein interactions between the bacterial and host proteins, most likely as a result from *L. pneumophila* co-evolving with amoebae. Proteins consisting of the eukaryotic domain ankyrin repeats are the most common among different *Legionella* species and have been shown important during replication of *L. pneumophila* within the host (26,29).

The genomes of *Legionella* species have around 10% strain-specific genes, which is high compared to other intracellular species such as *Chlamydia trachomatis* (30,31). Additionally, the genomes have plenty of mobile genetic elements and a high plasticity. It is likely that *Legionella* can transfer genes from many different organisms, such as other bacterial species, but also from their eukaryotic hosts since many proteins in *Legionella* genomes have eukaryotic domains (31). In total, there are around 300 predicted effector proteins in *L. pneumophila* that carry any of these features (32). From phylogenetic trees, these effector proteins seem to have been acquired by *Legionella* species from protozoa, plants, fungi, and animals (26).

1.3 Evolutionary history of RavC

T4BSS can be found in all *Legionella* species, but out of the 18 000 effectors secreted through T4BSS within the genus, only nine core common effectors have been found when studying the genomes from multiple *Legionella* species (26,33–35). The core effectors show high similarity at the protein level and phylogenetically, which indicates that they have evolved together with the *Legionella* genus (26,34). One of the core effectors, RavC, was first discovered in *L. pneumophila* by identifying proteins that were translocated via the T4BSS and contained carboxyl terminal glutamates, since this was shown to be necessary to allow translocation of a reporter protein (36). It was found that *ravC*, alongside with other *rav*-genes, are co-regulated

with T4BSS-translocated substrates and this group therefore got its name; region allowing vacuole colocalization (36).

RavC is the most conserved of the core effectors and was early found present in the genome of five different *Legionella* species (33,35). Homologues of this protein can be found in various other species, for example in other intracellular living bacteria such as *Chlamydia*, but also in eukaryotes like fungi and mammals (34,37,38). RavC seems to belong to the RMD1 protein family. In yeasts, such as *Saccharomyces cerevisiae*, it functions as a sporulation protein and is required for meiotic nuclear division, while the protein seems to be important for mitochondrial translation in humans (37,38) Interestingly, most of the other effectors found within the *Legionella* genus do not show homology to proteins found in other organisms, but are exclusively found in this specific genus (39).

In the last *Legionellales* common ancestor (LLCA), the T4BSS was gained together with RavC and legA3 (lpg 2300), which is one of the other core effectors (1). The gain of these two proteins in LLCA indicates that it was able to infect other cells, most likely eukaryotic ancestors. It is probable that the host-adaptation only occurred once in the *Legionellales* order since it is unlikely that different host-adaptation events would have involved the same genes every time. The order *Legionellales* includes *Legionella* and *Coxiella*, which both are intracellular bacteria growing inside eukaryotic cells (1). However, there is a great variation in lifestyle among the different species in the order, where some live as obligate intracellular and others as facultative intracellular (40). LLCA existed approximately 1.89 billion years ago and T4BSS, RavC and legA3 have been conserved since then which indicates that those proteins are, or were at some point, critical for the bacteria (1). Together, the core effectors and T4BSS may have been the key to success of *Legionellales*, and how they managed to infect the first eukaryotes (1). The T4BSS has remained conserved in *Legionellales*, but there is a great versatility of effectors within the order (26,41). RavC is one of the very few conserved effectors within *Legionellales*, however, very little is known about RavC and what function it may have for *Legionella* during infection of host cells and why it is conserved within all species of *Legionellales*.

2. Aim

This project aimed to study the evolution and function of the conserved effector RavC in *Legionella*. This was addressed by doing both evolutionary research on the protein and by using different experimental approaches.

3. Materials and methods

3.1 Buffers

The buffers used for protein purification were GST buffer which contained 50 mM Tris (pH 8.0), 150 mM NaCl, and 0.1% Triton X-100; a Protein refolding buffer that contained 500 mM L-Arginine, 20 mM, DHPC, 50 mM, Tris (pH 8.0), and 100 mM NaCl; and an Elution buffer which contained 250 mM Imidazole, and 0.1% Triton X-100. A Running buffer that contained 1% SDS, 192 mM Glycine, and 25 mM g Tris was used when performing SDS-PAGE.

3.2 Bacterial strains and cultures

Three different strains of *Legionella pneumophila*; Paris (DA57510), dTomato (DA57509), and SYFP2 (DA57508) were stored at -80 °C in 43% glycerol stocks. Strains dTomato and SYFP2 are *L. pneumophila* strain Paris containing the fluorescent marker dTomato and SYFP2 respectively. The strains were cultured at 37 °C on charcoal yeast extract (CYE) agar or ACES buffered yeast extract (AYE) broth, both supplemented with 0.4 mg/ml L-cysteine and 0.25 mg/ml ferric pyrophosphate. When needed, 15 µg/ml of kanamycin or 5 µl/ml of isopropyl β-

D-1-thiogalactopyranoside (IPTG) was added to the agar or broth. In addition, *Escherichia coli* DH5α containing a pGEM-kanamycin-mazF plasmid was grown at 37 °C on LB agar or LB broth containing 50 µg/ml of kanamycin. For protein purification, BL21 Gold Competent Cells *E. coli* were used. They were grown at 37 °C on either LB agar or LB broth containing 1 µl/ml ampicillin.

3.3 Cell cultures used for infections

For infection of *L. pneumophila, Acanthamoeba castellanii* and U937 cells differentiated into macrophages were used. To prepare the cells for infection, *A. castellanii* was grown in Peptone yeast glucose broth (PYG) containing 1 µl/ml of Penicillin-Streptomycin (Pen-Strep) from Sigma-Aldrich and was incubated at 30 °C. The U937 cells were differentiated into macrophages by growing them in RPMI Medium 1640 (1X) with phenol red and GlutaMAX from Gibco supplemented with 10% heat inactivated Fetal Bovine Serum (FBS), 1% Pen-Strep and the differentiation medium phorbol 12-myristate 13-acetate (PMA) with a concentration of 100 ng/ml. The cells were then incubated at 37 °C. The next day, the media was removed from the macrophages and was replaced with RPMI Medium 1640 (1X) from Gibco without phenol red, complemented with 10% FBS and 1% GlutaMAX.

3.4 Extraction of genomic DNA from *L. pneumophila* and pGEM-Kanamycin-mazF plasmid from *E. coli*

To extract DNA, strains Paris, dTomato, and SYFP2 were streaked on CYE agar and put in an incubator at 37 °C to grow for three days. Colonies of the three strains were scraped off from the agar using cotton swabs and resuspending them in 500 µl of ddH_2O. Using MasterPure Complete DNA & RNA Purification Kit from Lucigen, the DNA was extracted by following the protocol provided by manufacturer. The extracted DNA was then run on a 0.7% agarose gel to ensure DNA had been correctly extracted.

E. coli strain DH5α containing the pGEM-Kanamycin-mazF plasmid was streaked from frozen culture onto LB-agar. The next day, an overnight culture was prepared by adding 5 ml of LB broth into two Falcon tubes. Thereafter, kanamycin with a concentration of 50 mg/ml was added to each of the tubes. Colonies of *E. coli* from the agar plate were added to the tubes and they were then incubated overnight in a shaker at 37 °C shaking at 190 rpm. The next day, a plasmid DNA extraction and purification was performed by using the E.Z.N.A Plasmid DNA Mini Kit I from Omega Biotech, following protocol provided by manufacturer. To ensure that the mazF plasmid was obtained, 2 µl of the DNA was linearised by using the restriction enzymes EcoRI and XbaI and following the FastDigest protocol from Thermo Scientific. The linearised plasmid was then run on a 0.7% agarose gel for confirmation.

3.5 Double-joint PCR to get the first step construct for transformation

In order to make the mutant, two steps of natural transformation were performed (42). The construct used for the first step transformation contained an upstream fragment and a downstream fragment from *ravC* and the mazF cassette between the two fragments. Transforming the first step construct into *L. pneumophila* would remove *ravC* and the mazF would allow for selection. For the second step transformation, a smaller construct only containing the upstream and the downstream fragment was used to remove the mazF cassette.

Two different PCR reactions were performed by using Thermo Scientific Phusion High-Fidelity PCR Kit to amplify the upstream and downstream regions of the *ravC* gene in *L. pneumophila* Paris (accession number NC_006368). See Appendix A for PCR protocol. For the upstream region, the reversed primer was designed to contain an overhang of mazF. The primers used to

amplify the upstream were Upstream Forward (P1) and Upstream Reverse (P2). See Appendix A for primer sequences. The forward primer for the downstream region was also designed with an overhang of mazF. The primers used for the downstream amplifications were Downstream Forward (P3) and Downstream Reverse (P4). To amplify the mazF cassette, a PCR reaction was performed using the Thermo Scientific Phusion High-Fidelity PCR Kit with the primers mazF Forward and mazF Reverse.

The PCR products were run on a 0.7% agarose gel for purification using GeneJET Gel Extraction Kit from Thermo Scientific, following the protocol provided by manufacturer. For the double-joint PCR, the purified upstream, mazF, and downstream fragments were mixed in a 1:4:1 volume ratio. A Thermo Scientific Phusion High-Fidelity PCR master mix was prepared with primers P1 and P4. Thereafter, the double-joint PCR was performed, and the product was gel purified by using GeneJET Gel Extraction Kit from Thermo Scientific to contain only the desired band at around 5700 bp.

3.6 Joint PCR to get the second step construct for transformation

To make the small construct, two separate PCRs using Thermo Scientific Phusion High-Fidelity PCR Kit were performed to get the upstream fragment with a 3' overhang corresponding to the downstream fragment and to get the downstream fragment with a 5' overhang corresponding to the upstream fragment. The primers used for the upstream PCR were Forward (P1) and Reversed (P5). For the downstream PCR Forward (P6) and Reverse (P4) were used. See Appendix A for primer sequence and PCR protocol.

For the joint PCR, the purified upstream and downstream fragments were mixed in a 1:1 ratio and a Thermo Scientific Phusion High-Fidelity PCR master mix was prepared using primers P1 and P4. The PCR was then performed, and the PCR product was gel purified using GeneJET Gel Extraction Kit from Thermo Scientific and following the protocol provided by the manufacturer to achieve the correct DNA-fragment of around 2500 bp.

3.7 Natural transformation to achieve ΔRavC mutant

For the first step of transformation, the 5700 bp construct was transformed into *L. pneumophila*. Firstly, a loop-full of cultures of the bacteria was added to 3 ml of AYE broth in a 10 ml Falcon tube to make a concentrated stock solution. OD 600 was measured of the stock to calculate the correct volume to transfer from the stock to a new tube to get an OD of 0.05 and a liquid culture with a final volume of 3 ml. 1 µg of the first step construct was then added to the culture and they were then vortexed carefully at low speed. The culture was then grown for 24 hours at 30 °C, shaking at 190 rpm to allow for the bacteria to take up the construct.

After incubating for 24 hours, the liquid culture, containing the construct, was plated on five CYE plates containing kanamycin by adding 400 µl of the culture to each of the plates. All of the CYE plates were then incubated for five days at 37 °C. Colonies appearing on the CYE+kanamycin plates were patched on both CYE+kanamycin and CYE+IPTG and were incubated at 37 °C for 24-48 hours to select for mutants that were kanamycin resistant but IPTG sensitive. Thereafter, the correct mutant clones were streaked for isolated colonies on CYE+kanamycin and incubated at 37 °C for 24 hours.

The colonies confirmed to contain the 5700 bp construct were used to make the second step of the transformation by following the same protocol that was used for the first step of mutant isolation. However, this time, 1 µg of the second step 2500 bp construct was added to the liquid cultures instead of the first step 5700 bp construct.

After growing for 24 hours at 30 °C, the liquid cultures were diluted to 10^{-1}, 10^{-2}, and 10^{-3}. 100 µl of each of the dilutions from every culture was then plated on CYE+IPTG and were incubated for three to five days. After this, 30-50 colonies were picked and patched on both CYE+kanamycin and CYE+IPTG to select for mutant clones that were kanamycin sensitive but IPTG resistant. The plates were incubated at 37 °C for 24-48 hours. To verify the second step mutant, the exact same procedure was performed here as for the PCR analysis of the first step mutant.

3.8 PCR analysis of mutant

Following the first and second step of transformation, up to 10 mutant colonies were picked to analyse if they contained the correct construct. This was performed by resuspending each colony in 50 µl of nuclease free H_2O and incubating it for 15 minutes at 95 °C. Following this, the resuspended colonies were centrifuged at 13 000 *x g* for five minutes at room temperature. A PCR was then performed by using 5 µl of the supernatant mixed with Thermo Scientific DreamTaq PCR MasterMix (2x) and primers P1 and P4. Standard parameters were used, and the protocol ran for 24 cycles. The PCR product was then confirmed by running it on a 0.7% agarose gel. See Appendix A for primer sequence and PCR protocol.

3.9 Growth curve experiments

To see when *L. pneumophila* SYFP2 reaches the early exponential, mid-exponential, post-exponential and stationary phase, growth curve experiments were performed. Firstly, *L. pneumophila* SYFP2 was streaked for single colonies onto a CYE+IPTG plate and incubated at 37 °C for approximately 72 hours. To make the growth curves, bacteria were taken from three different parts of the plate (see **Fig. 1**), corresponding to one, two and three days of growth respectively, and these were used to make stock solutions with OD=1. Using the stock solutions, dilutions were made to get cultures with an OD of 0.1 and 0.05 for each zone, so that there were three tubes for each OD. 200 µl was thereafter aliquoted from each tube into six different wells on a 96-well plate. The 96-well plate was then put in a humidity cassette in a Tecan Spark Multimode Microplate Reader at 37 °C shaking orbitally at 180 rpm for 35 hours, with measurements taken every 30 minutes.

Figure 1. SYFP2 streaked for isolated colonies on CYE+IPTG. The plate was then divided into three zones; 1, 2, and 3. Zone 1 corresponds to approximately three days of growth while zone 2 corresponds to two days of growth and zone 3 for one day of growth. To make growth curves for SYFP2, bacteria was taken from the three different zones to compare the growth between the zones. Created in BioRender.com

3.10 Infection of *L. pneumophila* into host cells

To gain knowledge about the potential function of RavC, infections in host cells were performed. First, *A. castellanii* were grown in 96-well plates with approximately $1x10^5$ cells per well. 100 µl of LoFlo medium from Formedium supplemented with 5 mM of IPTG was added to the wells. 100 µl of bacteria with an MOI of 25, 50 and 100 of the ΔRavC mutant or SYFP2 was then added to the wells. The 96-well plate was thereafter put in a humidity cassette in Tecan Spark Multimode Microplate Reader at 30 °C for 70 hours with measurements taken every 30 minutes. The same procedure was then repeated for U937 cells differentiated into macrophages but using RPMI medium without phenol red from Gibco supplemented with 5 mM of IPTG instead of LoFlo. The humidity cassette was used for the U937 cells as well and the 96-well plate was put in the Tecan Spark Multimode Microplate Reader at 37 °C for 70 hours with measurements taken every 30 minutes. To measure the fluorescence for both the *A. castellanii* and the U937 cells infection, the excitation wavelength was 508 nm, and the emission wavelength was set to 555 nm. Growthcurver was used to measure the growth rate of the two strains by using all measurements as input (43).

3.11 Competition assay of ΔRavC and WT in U937 cells

A competition assay between ΔRavC and dTomato (WT) was performed to see if one had advantage over the other during infection. To do this, U937 cells were differentiated into macrophages and grown in a 96-well plate containing $2x10^5$ cells per well. The cells were then infected by adding 100 µl of a mixed culture of ΔRavC and WT in a 1:1 ratio with MOI's of 25, 50 and 100. The 96-well plate was then put in a humidity cassette in Tecan Spark Multimode Microplate Reader at 37 °C for 72 hours with measurements taken every 30 minutes. For the fluorescence reading, the excitation wavelength was set to 508 nm and the emission wavelength to 555 nm. The growth rates of ΔRavC and WT were calculated using Growthcurver and all measurements as input (43)

3.12 Extraction of RNA from different time points during growth

In order to extract RNA at different time points, the optimal OD and zones on the plates were chosen by utilizing the growth curved made for *L. pneumophila* SYFP2. It was decided to proceed with OD=0.1 using bacteria taken from zone 2 on the plate. Six 5 ml liquid AYE cultures with OD=0.1 were incubated at 37 °C shaking at 190 rpm. Three of the cultures were then taken after 15 hours of incubation, while the remaining three cultures were taken out after 23 hours. To extract the RNA, 500 µl from each culture was immediately added to 1000 µl of RNAprotect Bacteria reagent (Qiagen) in a 2 ml microfuge tube. The tubes were then mixed by inversion and incubated on ice for 5 minutes. After the incubation, the tubes were centrifuged at 17 000 *x g* for 2 minutes at 4 °C. Thereafter the supernatant was removed, and the pellet was saved for the proceeding step. The RNeasy mini kit from Qiagen was used to extract the RNA from the pellet by following the protocol provided by the manufacturer. After that, the extracted RNA was DNAse treated by using the TURBO DNA-*free* Kit from Invitrogen, following the protocol provided with the kit.

3.13 RT-qPCR

To make cDNA from the extracted and DNAse treated RNA, the High-capacity Reverse Transcription Kit from Applied Biosystems was used following the protocol provided by manufacturer. Firstly, each of the RNA-samples were diluted to get two tubes of 25 µl containing 500 ng of RNA in total per sample. A master mix was then prepared to have the following per sample: 5 µl 10x RT buffer, 2 µl 25x dNTP mix, 5 µl random primer mix and 10.5 µl nuclease free H_2O. For each sample, two reactions were performed where one contained 2.5 µl MultiScribe Reverse Transcriptase and in the other reaction the reverse transcriptase was

substituted with 2.5 µl of nuclease free H_2O as a negative control. In each PCR tube, 22.5 µl of the master mix along with either the Reverse Transcriptase or H_2O was added together with 25 µl of one of the RNA-dilutions, so that there were 12 PCR tubes in total containing 50 µl each.

The gene gyrA (lpp 1372) was chosen to be used as a norm since it is a housekeeping gene that is being expressed at steady levels (44). To ensure that the primers amplify the desired regions, a PCR was performed by using three different types of templates: colony of SYFP2, extracted DNA from WT *L. pneumophila* and cDNA. DreamTaq PCR MasterMix (2X) was used and 1 µl of template was added to each of the PCR tube.

The qPCR was performed using SYBR green from Sigma-Aldrich together with primers amplifying RavC and gyrA. Dilutions of 10^{-1}, 10^{-2}, and 10^{-3} of the cDNA and the negative controls were used as template. See Appendix A for qPCR protocol and primer sequence.

3.14 Expression of RavC for protein analysis

For expression of RavC, two pET21 plasmids were ordered from GenScript, one containing the gene of *ravC* (101) and the other containing *ravC* with an additional lipoyl domain (*Escherichia Coli*) tag for aiding expression and solubility (102). Both plasmids were then transformed into BL21 Gold cells by adding 1 µl of each plasmid in separate tubes, each containing 20 µl of BL21 Gold cells. The cells were put on ice for 30 minutes and afterwards they were heat shocked at 42 °C for 45 seconds and then put back on ice for approximately 20 seconds. 100 µl of SOC broth was then added to each of the tubes before they were put in an incubator for one hour at 37 °C shaking at 220 rpm. The cells were then plated on LB agar containing 1 µl ampicillin per ml and incubated at 37 °C over-night.

The next day, the cultures from each plate were transferred into 250 ml LB broth with an ampicillin concentration of 1µl/ml. Two flasks containing liquid cultures were made for each plasmid, so that there in total were four liquid cultures. The liquid cultures were then placed on a shaker at 220 rpm at 37 °C until an OD of approximately 0.8 was reached. Thereafter, 250 µl of IPTG with a concentration of 1 mM was added to each of the liquid cultures to induce expression. One liquid culture of each plasmid was incubated at 37 °C shaking at 220 rpm for three hours. The remaining two flasks were instead incubated over night at 30 °C.

3.15 Harvest and extraction of RavC

In order to harvest and extract RavC to be able to purify the protein, the bacterial cultures were transferred to centrifuge bottles after incubation and were then centrifuged (Thermo Scientific SL 40R) at 4500 RPM for 15 minutes at 4 °C. Thereafter, the supernatant was removed, and the pellets were harvested using GST buffer, containing 50 mM Tris with a pH of 8.0 and 150 mM NaCl. The harvested bacterial cells were then sonicated (Bandelin Sonoplus HD2200 homogeniser) during four minutes for six pulse cycles at 30% of its amplitude, which is at 20 kW. Subsequently, the sonicated cells were centrifuged (Sorvall RC 26 Plus) at 20 000 RPM for one hour at 4 °C. To extract the protein from the pellet, 20 ml of 6M Guanadinium chloride (GdHCl) was prepared. The pellet was placed in the solution together with a magnetic stirrer and was left overnight to dissolve. The next day the solution with the dissolved pellet was centrifuged again (Sorvall RC 26 Plus) at 20 000 RPM for one hour at 4 °C. To refold RavC after it had been dissolved in GdHCl, 7,5 ml of the supernatant from the protein was added to the protein refolding buffer while stirring and was then left overnight to stir so that the protein would be refolded. The buffer containing the protein was then concentrated using Sartorius Vivaspin 20 down to 5 ml.

3.16 Immobilized Metal Affinity Chromatography (IMAC)

To purify the protein from the sample, it ran through a 3 ml Ni Sepharso column (Cytiva) that was charged with 0.5 M NiCl and equilibrated using 15 ml of GST buffer. The sample was run through the column by gravity and when the sample had passed through the column, GST buffer was added to remove unbound material inside the column. Thereafter, the sample was eluted from the column by adding 9 ml of Elution buffer. To concentrate the samples, Sartorius Vivaspin 2 concentrators (centrifuging at 4500 rpm, 4°C), were used to get a volume of 200 µl of each of the eluted protein samples.

3.17 SDS-PAGE

Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) was performed by using a 4-12% SDS-PAGE gel in order to verify the purity of the protein of interest. The samples were diluted by adding 25 µl of DTT and SDS containing loading dye to 50 µl of protein sample. When loading the gel, 6 µl of PageRuler prestained protein ladder (Fisher BioReagents) was added to the first well. Thereafter, 20 µl of each of the samples was loaded into separate wells. A 1 % SDS running buffer was used and the SDS-PAGE was set to 150 V and ran for approximately one hour. To stain the gel, Comassie brilliant blue stain was used. The gel, together with the stain were firstly heated close to boiling and was then allowed to gently shake for 10 minutes in room temperature. After this, the stain was replaced with destain containing 10% acetic acid and 10 % isopropanol which was heated close to boiling point and was then left to shake at room temperature until the stain had been removed.

3.18 Phylogenetic tree of RavC

To construct a phylogenetic tree of homologues to RavC, PSI-BLAST command line was used (45). The query used was the protein sequence for RavC in *L. pneumophila* strain Philadelphia 1 (accession number AAU26214.1). To conduct the search, the following parameters were used: 5000 alignments, minimum e-value of 10^{-6} and a maximum of 10 iterations. From here, 21 representative sequences from the *Legionellales* order were manually selected to perform a multiple sequence alignment by using version v7.310 of MAFFT with the L-INS-I algorithm, default parameters were used (46). This alignment was then used to perform a new PSI-BLAST using the same parameters as before, but with the multiple sequence alignment as a query instead. From the 51 505 hits in PSI-BLAST, the sequences were clustered using the online version of H-CD-HIT (47,48) by doing three CD-HIT runs with a cut-off at 90% for the first run, 75% for the second run and 60% for the third run, default parameters were used. Additionally, 51 sequences were manually selected from the PSI-BLAST results separately. The results from the clustering and the 51 manually selected sequences were subsequently used to make a MAFFT L-INS-i alignment, using default parameters. The two alignments were trimmed of positions having a majority of gaps by using trimAl (49) with a gap threshold of 0.5. Phylogenetic trees were then inferred with IQ-TREE (50) version 2.0.3 by using an integrated model finder together with 1000 ultrafast bootstrap replicates (51) and an optimizer for the ultrafast bootstrap by using NNI on bootstrap alignment. For the H-CD-HIT clustered sequences, IQ-TREE model finder chose LG+R10, while it chose LG+R5 for the selected sequences. To visualize the tree, FigTree was used.

4. Results

4.1 Growth curves of SYFP2

The difference in growth rate between *L. pneumophila* taken from different parts of the plate corresponding to one, two and three days of growth was measured along with the difference of when the bacteria reached exponential and stationary phase depending on the OD. To measure this, liquid cultures of *L. pneumophila* were measured on a 96-well plate by Tecan Spark

Multimode Microplate Reader every 30 minutes until stationary phase was reached (it differed between the different cultures). As seen on **Fig 2**, the growth curves varied depending on where on the plate the bacteria were taken from, but also on the initial OD of the culture. The cultures with a starting OD of 0.1 reached exponential growth faster compared to the cultures with a starting OD of 0.05. For all OD's, bacteria taken from zone 3 (corresponding to one day of growth) reached exponential growth and stationary phase faster than bacteria taken from zone 2 (corresponding to two days of growth) and 1 (corresponding to three days of growth).

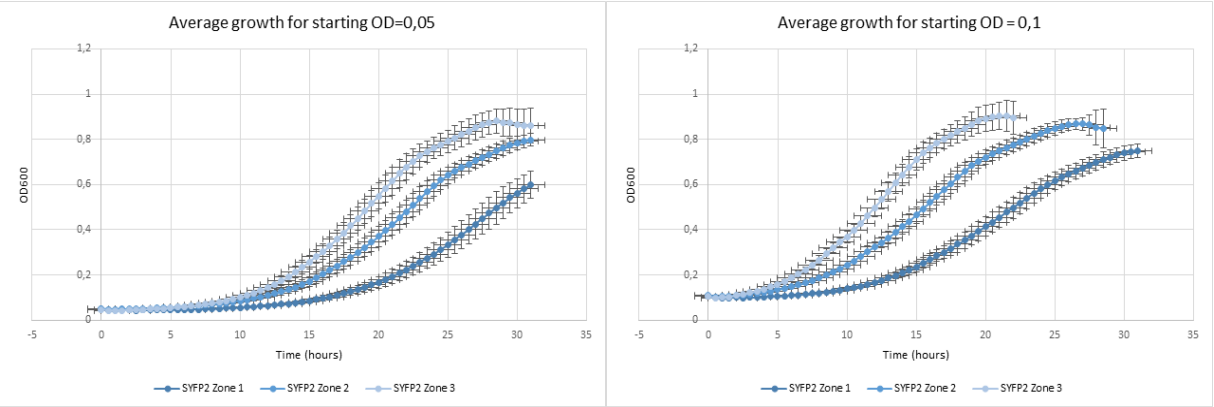


***Figure 2**:* The graphs show the growth for *L. pneumophila* at either OD 0.1 or 0.05. For each OD, bacteria were taken from different parts of an agar plate corresponding to one (zone 3), two (zone 2) and three days of growth (zone 1). Depending on how many days the bacteria had been growing and what the starting OD was, the growth curves looked different. For both OD's, bacteria corresponding to one day of growth entered exponential growth the fastest. Overall, the cultures with a starting OD of 0.1 reached exponential and stationary phase faster compared to the cultures with a starting OD of 0.05.

4.2 Growth of ΔRavC and WT

ΔRavC and WT (SYFP2) were grown in MOI's of 25, 50 and 100 together with *A. castellanii* in a 96-well plate to see what effect the lack of RavC would have on the intracellular growth of the mutant. To measure the growth of the bacteria while infecting *A. castellanii*, the relative fluorescence was measured every 30 minutes for 70 hours in total using Tecan Spark Multimode Microplate Reader. The relative fluorescence showed that ΔRavC had a higher growth rate in *A. castellanii* compared to the WT (**Fig 3**). ΔRavC and SYFP2 started at the same MOI and approximately 18 hours post-infection ΔRavC went into exponential growth and reached an RFU of around 1000. WT went into exponential phase 28 hours after infection and was not able to reach the same levels of fluorescence as ΔRavC. After 50 hours post-infection, ΔRavC starting at MOI 50 and 100 began to grow more again, showing that they went into a second infection where they infected the remaining live *A. castellanii* cells. WT with a starting MOI of 50 and 100 also went into a second round of infection, but after about 55 hours post-infection. There was no second infection for MOI 25 of ΔRavC and WT. In **Table 1,** the average growth rate can be seen for the different MOI's.

Using the same method as for *A. castellanii*, U937 cells were infected with ΔRavC and WT (SYFP2) in MOI's of 25, 50 and 100. There was no clear difference in growth between ΔRavC and the WT when looking at the relative fluorescence. There was no big difference in average growth rate, as can be seen in **Table 1**. In Appendix B, **Appendix B Figure 1** and **Appendix B Figure 2** display the error bars with the standard deviation for the growth curves of ΔRavC and SYFP2 in *A. castellanii* and U937 cells. During lag phase and exponential phase, the standard deviation is small. However, as the bacteria move into stationary phase, the standard deviation increases dramatically. This is expected as there may be many dead cells during that

time and bacteria emerging from the host cells and potentially also begin a second round of infection in the host cells that are still alive.

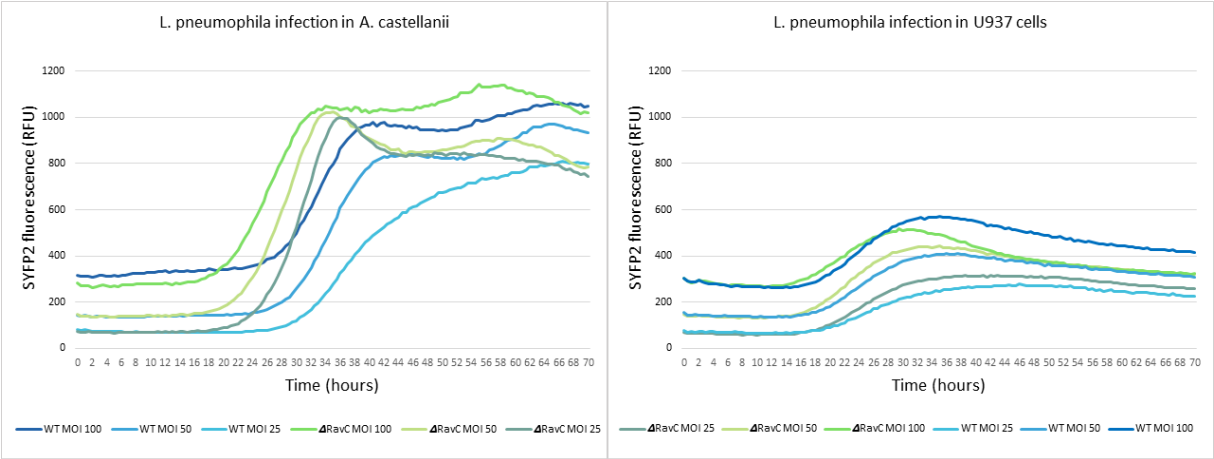


***Figure 3**:* Left: Growth curves showing the relative fluorescence of ΔRavC and WT (SYFP2) when infecting *A. castellanii* at MOI's of 25, 50 and 100. ΔRavC reached exponential growth 18 hours post-infection, while WT reached if after 28 hours. ΔRavC also reached a higher relative fluorescence compared to WT. An increase of growth can be seen for both ΔRavC and WT after about 50- and 55-hours post-infection respectively. This indicates that there was a second round of infection for those bacteria. Right: The growth curves are showing the relative fluorescence of ΔRavC and WT (SYFP2) during infection in U937 cells differentiated into macrophages. MOI's of 25, 50 and 100 were used for the infections. No major difference in growth rate can be seen between ΔRavC and WT.

Growth rates during infection

	MOI 25	*MOI 50*	*MOI 100*
A. castellanii ΔRavC	0.586 (±0.02)	0.550 (±0.02)	0.360 (±0.01)
A. castellanii WT	0.207 (±0.03)	0.325 (±0.04)	0.341 (±0.003)
U937 cells ΔRavC	0.384 (±0.02)	0.539 (±0.02)	0.795 (±0.01)
U937 cells WT	0.336 (±0.03)	0.504 (±0.04)	0.596 (±0.003)
Competition assay: ΔRavC	0.467 (±0.02)	0.639 (±0.02)	0.189 (±0.009)
Competition assay: WT	0.426 (±0.01)	0.445 (±0.02)	0.249 (0.008)

***Table 1**:* The table lists the average growth rate per hour for strains ΔRavC and WT at MOI's 25, 50 and 100 during infection in *A. castellanii*, U937 cells and during a competition assay in U937 cells. Growth rate is measured by the increase in fluorescence per hour. Standard deviation is shown in parenthesis. During infection in *A. castellanii*, ΔRavC shows a higher growth rate compared to WT, especially in MOI's 25 and 50. However, during infection in U937 cells and during the competition assay, no clear difference can be observed between ΔRavC and WT.

4.3 Competition assay

A competition assay was performed in U937 cells by using ΔRavC and strain WT (dTomato) at MOI's of 25, 50 and 100 to see if one strain has an advantage over the other. WT, encoding the fluorescent protein dTomato is relatively brighter than SYFP2 and therefore also show a higher relative fluorescence when measuring the fluorescence during the competition assay. Both ΔRavC and WT go into exponential phase at the same time and there are no major differences between growth rates of the two strains (**Fig 4A-C, Fig 5**). For MOI 25 and 50 ΔRavC show a slightly higher growth rate than WT, but the opposite can be seen for MOI 100. When performing fluorescent microscopy, it seemed as if the two strains did not co-infect host cells that often. Instead, ΔRavC and WT infected cells separately in most cases (**Fig 4D**).

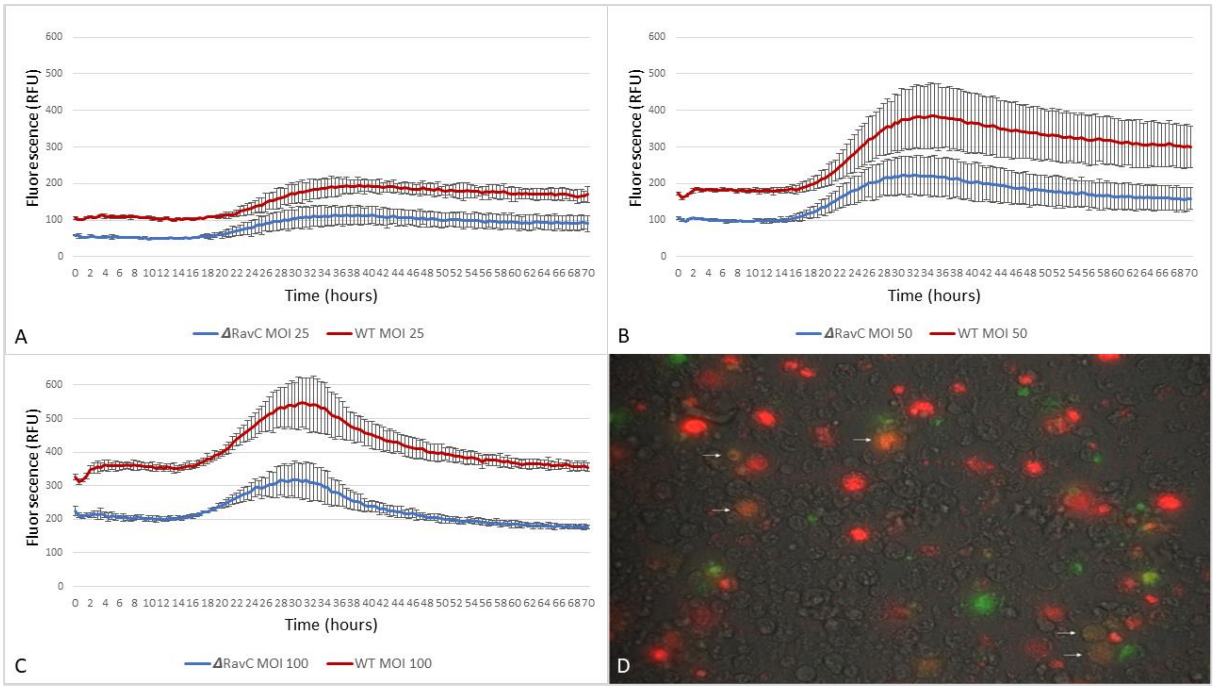


***Figure 4**:* A-C: Growth curves showing the relative fluorescence during competition assay of ΔRavC and WT in U937 cells at MOI's 25, 50 and 100. The slope of the growth curves shows that there is no clear difference in growth between ΔRavC and WT. D: Fluorescent microscopy picture of competition assay in U937 cells, four days post-infection. Green is ΔRavC and red is WT. Arrows indicate where ΔRavC and WT are co-infecting the same host-cell, those cells emit a more orange fluorescence. Though, after four days of infection, most cells are infected with only one of the strains (either ΔRavC or WT).

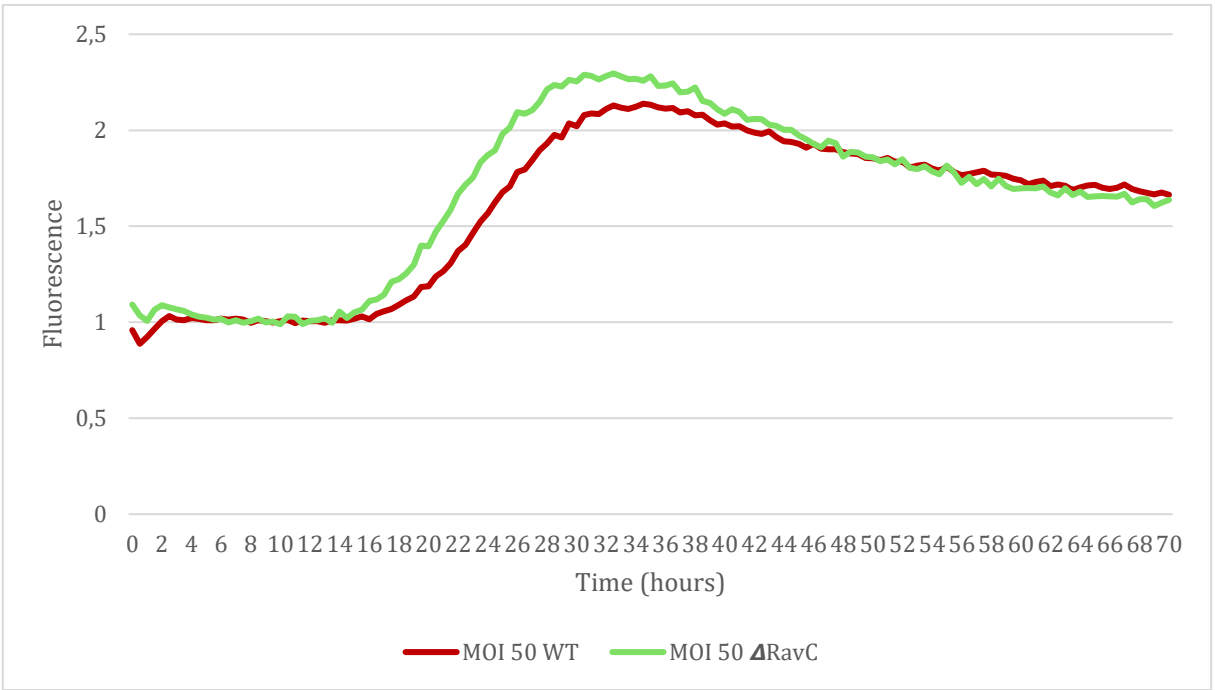


***Figure 5**:* This graph shows the normalized growth curves of WT and ΔRavC during competition assay in U937 cells with an MOI of 50. In the beginning of the infection, the two strains have a 1:1 ratio. When the bacteria reach exponential phase, there is an increase in intracellular growth of ΔRavC compared to WT. However, as the bacteria progress into stationary phase, the ratio becomes 1:1 again.

4.4 RT-qPCR

RNA was extracted from three samples at exponential growth (after 15 hours of incubation) and three samples at stationary growth (after 23 hours of incubation) from *L. pneumophila* growing in AYE as a pure culture (see **Fig 2**, OD=0.1, zone 2 for reference). To see at which of the time points *L. pneumophila* potentially was expressing RavC, RT-qPCR was performed. However, no conclusions can be drawn from the results of the RT-qPCR as the negative controls (RNA without reverse transcriptase) showed false positive results. As seen from **Fig 6** and **Table 2**, the negative controls have a Cq-value that is 2,5 cycles higher on average compared to the positive samples and they reach similar levels of fluorescence after 40 cycles. All potential contamination sources were checked, and it turned out that the primers were amplifying things even without RNA or DNA present in the master mix.

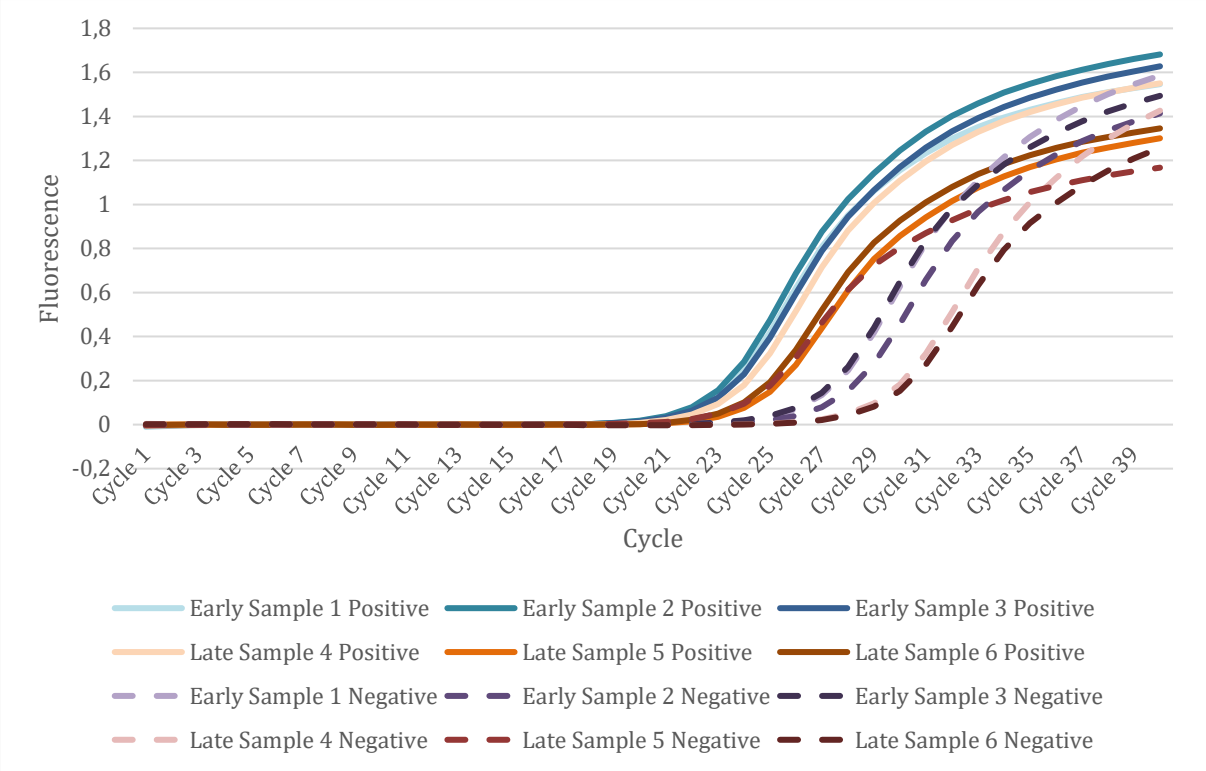


***Figure 6**:* A graph showing the amplification of the different samples during a RT-qPCR to see if/when RavC is expressed when *L. pneumophila* is growing in a pure culture. However, primers were amplifying even without template present and the results can therefore not be trusted. Both samples and negative controls are notably increasing in fluorescence before 30 cycles has passed and all reach a similar fluorescence after 40 cycles.

Cq-values

	Early Sample 1	*Early Sample 2*	*Early Sample 3*	*Late Sample 4*	*Late Sample 5*	*Late Sample 6*
Positive	23,8	24,8	25,9	25,4	26,0	24,9
Negative	26,7	27,5	28,4	26,4	27,5	28,9

***Table 2**:* The table lists the average Cq-value of the six different samples (early 1-3 and late 4-6) and their negative controls. The samples have a similar Cq-value with an average of 25,1 for the positive samples, while the negative controls have an average Cq-value of 27,6. The average difference is therefore 2,5 cycles between the positive samples and the negative controls.

4.5 Phylogenetics

The maximum-likelihood phylogenetic tree inferred by IQ-TREE using the manually selected sequences (**Fig 7**) show that the protein sequence of RavC results in *Legionellales* as a monophyletic group (*Legionellaceae* shown in red and *Coxiellaceae* shown in purple). Moreover, all the host-adapted bacteria, including *Rickettsiae* (shown in turquoise) and *Chlamydiae* (shown in blue), belong to the same clade. In that clade, *Acanthamoeba castellanii* can also be found. This could be a fluke in reconstruction or a retro-transfer from bacteria and back into *A. castellanii* as it acts as host for many species of host-adapted bacteria. The RavC homologue found in archaea (shown in orange) is very distant to the RavC protein found in *Legionella* and was used to root the tree. In Appendix B, a supplementary figure showing the tree inferred using the H-CD-HIT clustered sequences shows a similar pattern (**Appendix B Figure 3**).

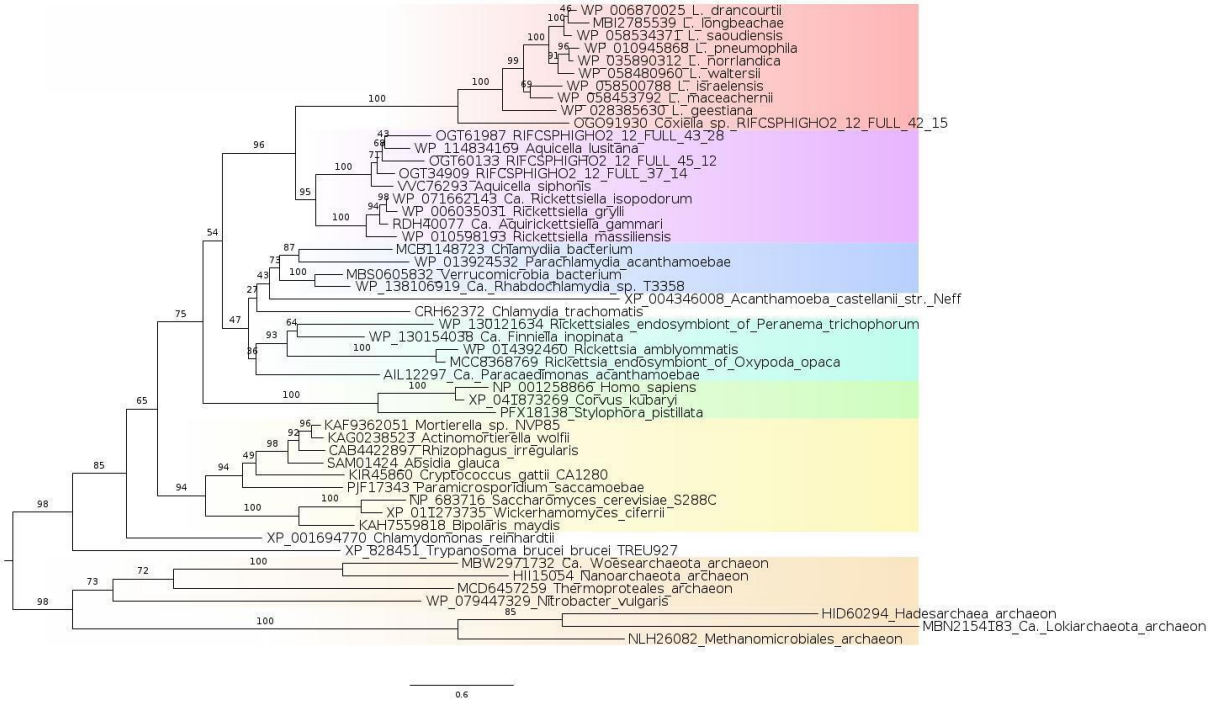


***Figure 7**:* A phylogenetic tree showing the relationship between RavC in *L. pneumophila* and homologues in other species. Numbers on branches represent bootstrap values. The host-adapted bacteria are found in a monophyletic group which points to one early single transfer of RavC.

4.6 Purification of RavC

In the process of trying to purify RavC, SDS-PAGE was performed to see if we had managed to isolate and purify only the protein of interest. When the harvested cells had been sonicated and centrifuged, a band could clearly be seen at the appropriate size of the protein (26 kDa) when running an SDS-PAGE. Since this step was done prior to purification, plenty of other bands can be seen as well. However, after purifying the supernatant through an IMAC column and concentrating it, the protein could no longer be seen when running the gel (**Fig 8**). The same thing happened when the protein was refolded in buffer (figure not shown). A weak band could be seen for the supernatant of sample 102 after the sample had been dissolved in GdHCl and centrifuged. Unfortunately, this band was not of the expected size.

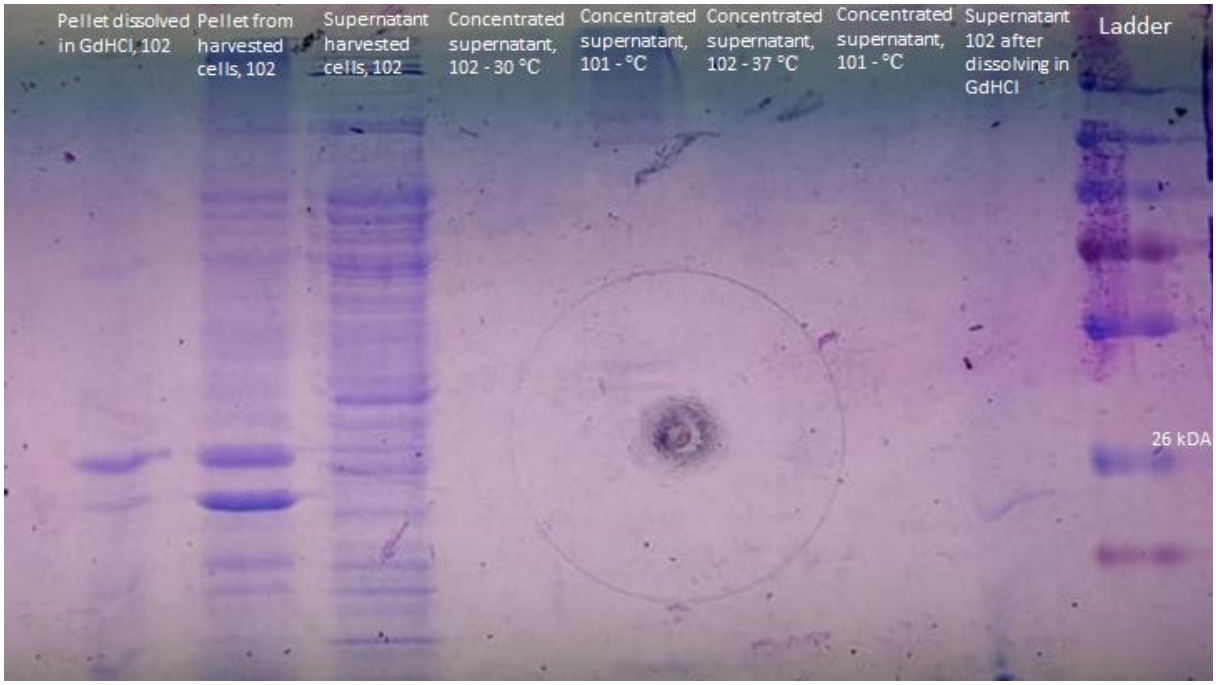


***Figure 8**:* SDS-page showing the bands after the different steps in trying to purify the RavC protein. A band of the expected size (~26 kDA) can be seen in in the pellet after the cells have been harvested, but also after the pellet had been dissolved in GdHCl. A faint band of the wrong size can be seen in the supernatant of sample 102 after it had been dissolved in GdHCl and centrifuged.

5. Discussion

The purpose of this project was to study the evolution and function of the *Legionella pneumophila* effector RavC. To investigate its function, we created a knock-out mutant lacking the gene encoding RavC (ΔRavC). However, there were issues with the natural transformation steps to remove *ravC* and create the mutant. For the first step transformation there were no colonies at first. In the protocol used at first, the bacteria were grown for 12-18 hours while shaking at 37 °C before adding DNA. After adding DNA, the culture was incubated again but growing at 30 °C without shaking (42). Since we got no colonies by using this protocol, this protocol was modified by adding higher concentrations of DNA and by adding the DNA directly to the cultures. Adding a higher concentration of DNA (1 µg) directly to the cultures and incubating them at 30 °C while shaking for 24 hours showed to be the most efficient method in this case. This agrees with the results from another study which showed that the transformation was as most efficient when DNA was added during post-exponential phase, right before the bacteria reached stationary phase (52). Bacteria that have grown for one day are still in exponential phase and are not as competent as during post-exponential phase. However, in our study the bacteria were growing for 24 hours together with the DNA and will progressively become more competent and can therefore take up the DNA in a later growth phase.

When we managed to achieve our mutant, we wanted to test whether ΔRavC showed a difference in growth compared to the WT. Therefore, *A. castellanii* was infected with ΔRavC and SYFP2 (WT) at different MOI's. The growth curves show that ΔRavC has a higher growth rate compared to the WT in *A. castellanii*. This is contrary to the expectations since it is

hypothesized that RavC was important for LLCA to be able to infect their first host, and therefore most likely still is important during infection since it is conserved within all *Legionella* species (1,26). Additionally, previous studies have shown that the deletion of *ravC* resulted in no significant decrease in intracellular growth during infection (34). We instead show that the deletion of *ravC* causes an increase in intracellular growth during exponential phase of the infection in *A. castellanii*, which correlates with the findings in a study by Park et al from 2020 (53). They did not find a significant increase compared to the WT, but still saw a slight fitness advantage for the mutant in both *A. castellanii* and U937 cells. As we only had time to perform one infection of ΔRavC in *A. castellanii* and U937 cells, we cannot say for sure whether our results are statistically significant or not. However, the difference in growth rate is less prominent between ΔRavC and WT during infection in U937 cells compared to infection in *A. castellanii*. One potential hypothesis for this is that *L. pneumophila* is more adapted to having an intracellular life cycle in its environmental host *A. castellanii* compared to macrophages. Another interesting hypothesis is that RavC is important during certain stages of infection, such as during stationary phase. In the graphs representing the growth curves during infection, we can see that ΔRavC no longer seem to have the same fitness advantage when the bacteria have reached stationary growth. It could also be that RavC only is of importance during infection of certain host species or in combination with other effectors. This has been shown to be the case for other *rav*-genes, such as *ravO* (53). Moreover, studies done on one of the other *rav*-genes, r*avY*, have shown that deletion affects the growth of *L. pneumophila*. Mutants of *L. pneumophila* strain Paris lacking the gene for RavY have an impaired replication in amoebae and show no growth in macrophages (53). When instead deleting RavY from *L. pneumophila* wild-type (WT) strain Lp01, the mutant bacteria could grow, but showed a 2-log decrease compared to the WT (54). Because of these studies, along with the high conservation of *ravC*, similar results were expected from our ΔRavC knock-out since they both are effector proteins released during infection and belong to the same group of genes that encode region allowing vacuole colocalization (*rav*) proteins.

After the single infections, a competition assay was performed in U937 cells to see if ΔRavC or WT has an advantage over the other during infection. There was no clear difference in growth between ΔRavC and WT, which is in line with the result when ΔRavC and WT were infecting U937 cells separately. To my knowledge, there have been no studies showing the difference in growth during competition assays in U937 cells. However, competition assays between ΔRavC and WT in *A. castellanii* have been performed in another study where they saw no significant difference in growth between the two strains (35). Compared to our mutant, the ΔRavC created in that study still had fragments of *ravC* left in the genome which may affect the results. Additionally, no data was shown on the growth of ΔRavC separately during infection of *A. castellanii* in that study. This would be of interest since we saw a difference in growth between ΔRavC and WT during single infections in our study.

Thereafter, we wanted to see during which growth phase RavC might be expressed. To do this, we performed a RT-qPCR. However, only RNA extracted from *L. pneumophila* in pure culture was analysed. We tried to extract RNA from *L. pneumophila* infecting both *A. castellanii* and U937 cells, but we did not manage to isolate high enough concentrations of RNA to be able to make cDNA and therefore we could not analyse the potential expression of RavC during infection of host cells. The same protocol was used to extract RNA from the pure culture as from the cultures with *L. pneumophila* infecting *A. castellanii* and U937 cells. Another protocol may have to be used to be able to get high enough concentrations of RNA in order to analyse the expression levels. When performing the RT-qPCR for the extracted RNA from pure culture, false positive results were seen for the negative controls. Therefore, we ordered new primers

and used a new tube of qPCR mix (SYBR green). However, there were still false positive results when using the qPCR primers for RavC. A qPCR was therefore set up with only SYBR green and primers (no cDNA or RNA) but there was still a strong fluorescence for RavC. The primers for gyrA worked well and showed no false positive results which is why we believe that it is the primers for *ravC* that are amplifying without any DNA present. To rule out that the primers were contaminated, we used new, nuclease free water from Thermo Scientific as well as filter tips when diluting the new primer stocks. Filter tips and nuclease free water was also used when making the primer dilutions and setting up the qPCR, but the negative controls still came out as false positive. If there was more time for troubleshooting to avoid the false positive results, the RT-qPCR would be an interesting method to use to get a better idea of when RavC might be expressed as it has been successfully used for other proteins expressed by *L. pneumophila,* both during infection and as free-living bacteria (55).

Finally, we wanted to see the evolutionary history of RavC by inferring phylogenetic trees. In an evolutionary perspective, RavC seems to have been important since the phylogenetic tree indicates that RavC only has been transferred once to an ancestor of *Legionellales,* possibly in the LLCA, and has remained conserved within the species since (1). Another study tested different evolutionary scenarios to see if RavC has been horizontally transferred between genera or if it was gained in a common ancestor and has adapted from there (56). When comparing these two evolutionary models, there was a higher likelihood for the gene to have been horizontally transferred (56). However, since all of the host-adapted bacteria are found in the same clade in our phylogenetic trees, together with *A. castellanii*, it points to one early transfer of RavC to an ancestor of this clade and that the protein then has been either been retro-transferred from bacteria to *A. castellanii* since it acts as a host for several species found in that clade, or that it is a mistake in the tree reconstruction (3–5). However, it is not possible discriminate between those hypothesises in this tree.

There are several limitations in this study. Firstly, due to a lack of time, the infection experiments were not repeated to make sure that the results are reproducible. We can also see that there is quite a big variation between the replicates (illustrated by the error bars) during the competition assay (**Fig 4A-C**). It would therefore be interesting to set up yet another competition assay in U937 cells to see if the variation remains that big or if other results would be obtained. The lack of time we experienced for the infections was a result of the struggles with the natural transformation to achieve ΔRavC. The natural transformation was yet another limitation since it requires several PCRs and double-joint PCRs where mutations may occur in the fragments that are being amplified. These PCRs would also need some further optimization since there were troubles with getting the right construct during the double-joint PCR. The selection for mutants is yet another limitation since there were cases when colonies grew on both Kanamycin and IPTG after the second step transformation. Additionally, there could also be mixed populations in the colonies that seemingly only grew on IPTG after the second step which then could affect the results when doing the infection of host cells.

In the future it would be of interest to perform a competition assay in *A. castellanii* as well, and not only in U937 cells since there was a difference in growth rate between ΔRavC and WT in *A. castellanii*. Moreover, a plasmid encoding the gene could be transformed into the bacteria to see if the mutant would revert to normal growth rate when gaining the gene again. A plasmid containing the *ravC* gene and a fluorescent gene could also be used to transfect *A. castellanii* and U937 cells to see where the protein is co-localized in the host and therefore get a better understanding of the actual function of the protein. To get a better understanding of RavC, it would also be useful to know the expression levels of it during infection but also in pure culture.

Since it is an effector protein, it should mainly be expressed during infection and not while growing in a pure culture, but it would still be of interest to know if that actually is the case. However, as our primers for the RT-qPCR were not working as desired, this is something to investigate further in the future.

In conclusion, according to our phylogenetic tree, RavC is likely to have been acquired more than two billion years ago. The effector must have played an important role for the host-adapted bacteria, including *Legionella*, at least during some point of evolution and most likely still does since it remains conserved within all species in the genus. No difference in growth could be seen between the mutant and the wild type during infection in U937 cells. Interestingly, there was an increase of growth in the mutant compared to the wild type in *A. castellanii*. This difference in intracellular growth may be a result of *Legionella's* co-evolution with *A. castellanii* since the amoebae is a more common environmental host for the bacteria compared to macrophages. However, more research needs to be done on the subject to get a better understanding of what role RavC may have played in the first host-adapted bacteria and what potential role it still may play in the host-adapted bacteria we have today.

6. Acknowledgements

I would like to thank my supervisor **Lionel Guy** for letting me do my master thesis in his group and supporting me during my time here. I would also like to thank **Anaisa Moreno** and **Kiran Paranjape** for helping me and supporting me with all my lab work in such a great way. Lastly, I would like to thank my co-supervisor **Celestine Chi** for working with me on the biochemistry part of the thesis.

7. References

1. Hugoson E, Guliaev A, Ammunét T, Guy L. Host-adaptation in Legionellales is 1.9 Ga, coincident with eukaryogenesis. Mol Biol Evol. 2022 Feb 15;msac037.

2. 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. 2006 Feb;9(1):86–94.

3. Rowbotham TJ. Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. J Clin Pathol. 1980 Dec;33(12):1179–83.

4. Graells T, Ishak H, Larsson M, Guy L. The all-intracellular order Legionellales is unexpectedly diverse, globally distributed and lowly abundant. FEMS Microbiol Ecol [Internet]. 2018 Dec 1 [cited 2021 Sep 26];94(12). Available from: https://academic.oup.com/femsec/article/doi/10.1093/femsec/fiy185/5110392

5. Khodr A, Kay E, Gomez-Valero L, Ginevra C, Doublet P, Buchrieser C, et al. Molecular epidemiology, phylogeny and evolution of Legionella. Infect Genet Evol. 2016 Sep 1;43:108–22.

6. Eisenreich W, Heuner K. The life stage-specific pathometabolism of Legionella pneumophila. FEBS Lett. 2016;590(21):3868–86.

7. Borella P, Montagna MT, Romano-Spica V, Stampi S, Stancanelli G, Triassi M, et al. Legionella Infection Risk from Domestic Hot Water. Emerg Infect Dis. 2004 Mar;10(3):457–64.

8. Valster RM, Wullings BA, van der Kooij D. Detection of Protozoan Hosts for Legionella pneumophila in Engineered Water Systems by Using a Biofilm Batch Test. Appl Environ Microbiol. 2010 Nov;76(21):7144–53.

9. Best A, Abu Kwaik Y. Evolution of the Arsenal of Legionella pneumophila Effectors To Modulate Protist Hosts. mBio. 2018 Oct 9;9(5):e01313-18.

10. National Academies of Sciences, Engineering, and Medicine. Management of Legionella in Water Systems [Internet]. Washington, DC: The National Academies Press; 2020 [cited 2022 Jan 10]. 290 p. Available from: https://www.nap.edu/catalog/25474/management-of-legionella-in-water-systems

11. Marston BJ, Lipman HB, Breiman RF. Surveillance for Legionnaires' Disease: Risk Factors for Morbidity and Mortality. Arch Intern Med. 1994 Nov 14;154(21):2417–22.

12. Fields BS, Benson RF, Besser RE. Legionella and Legionnaires' Disease: 25 Years of Investigation. Clin Microbiol Rev. 2002 Jul 1;15(3):506–26.

13. Chambers ST, Slow S, Scott-Thomas A, Murdoch DR. Legionellosis Caused by Non-Legionella pneumophila Species, with a Focus on Legionella longbeachae. Microorganisms. 2021 Feb;9(2):291.

14. Cazalet C, Rusniok C, Brüggemann H, Zidane N, Magnier A, Ma L, et al. Evidence in the Legionella pneumophila genome for exploitation of host cell functions and high genome plasticity. Nat Genet. 2004 Nov;36(11):1165–73.

15. Phin N, Parry-Ford F, Harrison T, Stagg HR, Zhang N, Kumar K, et al. Epidemiology and clinical management of Legionnaires' disease. Lancet Infect Dis. 2014 Oct 1;14(10):1011–21.

16. Beauté J, Plachouras D, Sandin S, Giesecke J, Sparén P. Healthcare-Associated Legionnaires' Disease, Europe, 2008−2017 - Volume 26, Number 10—October 2020 - Emerging Infectious Diseases journal - CDC. [cited 2022 Apr 3]; Available from: https://wwwnc.cdc.gov/eid/article/26/10/18-1889_article

17. van Heijnsbergen E, Schalk JAC, Euser SM, Brandsema PS, den Boer JW, de Roda Husman AM. Confirmed and Potential Sources of Legionella Reviewed. Environ Sci Technol. 2015 Apr 21;49(8):4797–815.

18. Whiley H, Bentham R. Legionella longbeachae and Legionellosis. Emerg Infect Dis. 2011 Apr;17(4):579–83.

19. Correia AM, Ferreira JS, Borges V, Nunes A, Gomes B, Capucho R, et al. Probable Person-to-Person Transmission of Legionnaires' Disease [Internet]. https://doi-org.ezproxy.its.uu.se/10.1056/NEJMc1505356. Massachusetts Medical Society; 2016 [cited 2022 Jan 10]. Available from: http://www.nejm.org/doi/10.1056/NEJMc1505356

20. Segal G, Feldman M, Zusman T. The Icm/Dot type-IV secretion systems of Legionella pneumophila and Coxiella burnetii. FEMS Microbiol Rev. 2005 Jan 1;29(1):65–81.

21. Isberg RR, O'Connor TJ, Heidtman M. The Legionella pneumophila replication vacuole: making a cosy niche inside host cells. Nat Rev Microbiol. 2009 Jan;7(1):13–24.

22. 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. 2019 Sep 1;11(9):2619–32.

23. Coers J, Kagan JC, Matthews M, Nagai H, Zuckman DM, Roy CR. Identification of Icm protein complexes that play distinct roles in the biogenesis of an organelle permissive for Legionella pneumophila intracellular growth. Mol Microbiol. 2000;38(4):719–36.

24. Mondino S, Schmidt S, Rolando M, Escoll P, Gomez-Valero L, Buchrieser C. Legionnaires' Disease: State of the Art Knowledge of Pathogenesis Mechanisms of *Legionella*. Annu Rev Pathol Mech Dis. 2020 Jan 24;15(1):439–66.

25. Schunder E, Gillmaier N, Kutzner E, Herrmann V, Lautner M, Heuner K, et al. Amino Acid Uptake and Metabolism of Legionella pneumophila Hosted by Acanthamoeba castellanii. J Biol Chem. 2014 Jul 25;289(30):21040–54.

26. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE, Rolando M, 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. 2019 Feb 5;116(6):2265–73.

27. Gomez-Valero L, Rusniok C, Jarraud S, Vacherie B, Rouy Z, Barbe V, et al. Extensive recombination events and horizontal gene transfer shaped the Legionella pneumophila genomes. BMC Genomics. 2011 Nov 1;12(1):536.

28. Albert-Weissenberger C, Cazalet C, Buchrieser C. Legionella pneumophila — a human pathogen that co-evolved with fresh water protozoa. Cell Mol Life Sci. 2007 Feb;64(4):432–48.

29. Habyarimana F, Al-khodor S, Kalia A, Graham JE, Price CT, Garcia MT, et al. Role for the Ankyrin eukaryotic-like genes of Legionella pneumophila in parasitism of protozoan hosts and human macrophages. Environ Microbiol. 2008;10(6):1460–74.

30. Gomez-Valero L, Rusniok C, Buchrieser C. Legionella pneumophila: Population genetics, phylogeny and genomics. Infect Genet Evol. 2009 Sep 1;9(5):727–39.

31. Swart AL, Gomez-Valero L, Buchrieser C, Hilbi H. Evolution and function of bacterial RCC1 repeat effectors. Cell Microbiol. 2020;22(10):e13246.

32. Burstein D, Zusman T, Degtyar E, Viner R, Segal G, Pupko T. Genome-Scale Identification of Legionella pneumophila Effectors Using a Machine Learning Approach. PLOS Pathog. 2009 Jul 10;5(7):e1000508.

33. Gomez-Valero L, Rusniok C, Rolando M, Neou M, Dervins-Ravault D, Demirtas J, et al. Comparative analyses of Legionella species identifies genetic features of strains causing Legionnaires' disease. Genome Biol. 2014 Nov 3;15(11):505.

34. Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, et al. Genomic analysis of 38 Legionella species identifies large and diverse effector repertoires. Nat Genet. 2016 Feb;48(2):167–75.

35. Wexler M, Zusman T, Linsky M, Lifshitz Z, Segal G. The Legionella genus core effectors display functional conservation among orthologs by themselves or combined with an accessory protein. Curr Res Microb Sci. 2022 Jan 1;3:100105.

36. Huang L, Boyd D, Amyot WM, Hempstead AD, Luo ZQ, O'Connor TJ, et al. The E Block motif is associated with Legionella pneumophila translocated substrates. Cell Microbiol. 2011;13(2):227–45.

37. Janer A, Antonicka H, Lalonde E, Nishimura T, Sasarman F, Brown GK, et al. An RMND1 Mutation Causes Encephalopathy Associated with Multiple Oxidative Phosphorylation Complex Deficiencies and a Mitochondrial Translation Defect. Am J Hum Genet. 2012 Oct 5;91(4):737–43.

38. Enyenihi AH, Saunders WS. Large-scale functional genomic analysis of sporulation and meiosis in Saccharomyces cerevisiae. Genetics. 2003 Jan;163(1):47–54.

39. Qiu J, Luo ZQ. Legionella and Coxiella effectors: strength in diversity and activity. Nat Rev Microbiol. 2017 Oct;15(10):591–605.

40. Duron O, Doublet P, Vavre F, Bouchon D. The Importance of Revisiting Legionellales Diversity. Trends Parasitol. 2018 Dec;34(12):1027–37.

41. Chen C, Banga S, Mertens K, Weber MM, Gorbaslieva I, Tan Y, et al. Large-scale identification and translocation of type IV secretion substrates by Coxiella burnetii. Proc Natl Acad Sci. 2010 Dec 14;107(50):21755–60.

42. Bailo N, Kanaan H, Kay E, Charpentier X, Doublet P, Gilbert C. Scar-Free Genome Editing in Legionella pneumophila. Methods Mol Biol Clifton NJ. 2019;1921:93–105.

43. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics. 2016 Apr 19;17(1):172.

44. Brüggemann H, Hagman A, Jules M, Sismeiro O, Dillies MA, Gouyette C, et al. Virulence strategies for infecting phagocytes deduced from the in vivo transcriptional program of Legionella pneumophila. Cell Microbiol. 2006;8(8):1228–40.

45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389–402.

46. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013 Apr 1;30(4):772–80.

47. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150–2.

48. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics. 2010 Mar 1;26(5):680–2.

49. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug 1;25(15):1972–3.

50. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020 May 1;37(5):1530–4.

51. UFBoot2: Improving the Ultrafast Bootstrap Approximation | Molecular Biology and Evolution | Oxford Academic [Internet]. [cited 2022 May 2]. Available from: https://academic.oup.com/mbe/article/35/2/518/4565479

52. Buchrieser C, Charpentier X. Induction of Competence for Natural Transformation in Legionella pneumophila and Exploitation for Mutant Construction. In: Buchrieser C, Hilbi H, editors. Legionella: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2013 [cited 2022 May 5]. p. 183–95. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-62703-161-5_9

53. Park JM, Ghosh S, O'Connor TJ. Combinatorial selection in amoebal hosts drives the evolution of the human pathogen Legionella pneumophila. Nat Microbiol. 2020 Apr;5(4):599–609.

54. Liu L, Roy CR. The Legionella pneumophila Effector RavY Contributes to a Replication-Permissive Vacuolar Environment during Infection. Infect Immun. 89(12):e00261-21.

55. Faucher S, Mueller C, Shuman H. Legionella Pneumophila Transcriptome during Intracellular Multiplication in Human Macrophages. Front Microbiol [Internet]. 2011 [cited 2022 Jun 5];2. Available from: https://www.frontiersin.org/article/10.3389/fmicb.2011.00060

56. Burstein D, Amaro F, Zusman T, Lifshitz Z, Cohen O, Gilbert JA, et al. Uncovering the Legionella genus effector repertoire - strength in diversity and numbers. Nat Genet. 2016 Feb;48(2):167–75.

57. The use of sarkosyl in generating soluble protein after bacterial expression. | PNAS [Internet]. [cited 2022 May 13]. Available from: https://www-pnas-org.ezproxy.its.uu.se/doi/abs/10.1073/pnas.88.4.1192

58. Hahn AWA, Regenass S, Kern F, Buhler FR, Resink TJ. Expression of Soluble and Insoluble Fibronectin in Rat Aorta: Effects of Angiotensin II and Endothelin-1. Biochem Biophys Res Commun. 1993 Apr 15;192(1):189–97.

59. Przybycien TM, Dunn JP, Valax P, Georglou G. Secondary structure characterization of ß-lactamase inclusion bodies. Protein Eng Des Sel. 1994 Jan 1;7(1):131–6.

60. Frangioni JV, Neel BG. Solubilization and Purification of Enzymatically Active Glutathione S-Transferase (pGEX) Fusion Proteins. Anal Biochem. 1993 Apr 1;210(1):179–87.

61. Schnaitman CA. Solubilization of the Cytoplasmic Membrane of Escherichia coli by Triton X-100. J Bacteriol. 1971 Oct;108(1):545–52.

62. Tsumoto K, Ejima D, Kumagai I, Arakawa T. Practical considerations in refolding proteins from inclusion bodies. Protein Expr Purif. 2003 Mar 1;28(1):1–8.

63. Lin SH, Guidotti G. Chapter 35 Purification of Membrane Proteins. In: Burgess RR, Deutscher MP, editors. Methods in Enzymology [Internet]. Academic Press; 2009 [cited 2022 Mar 30]. p. 619–29. (Guide to Protein Purification, 2nd Edition; vol. 463). Available from: https://www.sciencedirect.com/science/article/pii/S0076687909630354

8. Appendix A

Appendix A contains the primers used for the different PCRs mentioned in this report, together with the PCR parameters used.

Upstream Forward (P1)	5’-GTTGAGGAGATGGCTAAGCATC-3’
Upstream Reverse (P2)	5’-GGCCCAATTCGCCCTATAGTGAGTCGGGTTTTTCCCTTAATTTATGCCCA-3’
Downstream Forward (P3)	5’-GGGTTTGCTCGGGTCGGTGGCATATGTGATTTTAATAACCCAACAATGAA-3’
Downstream Reverse (P4)	5’- GTAACATTTATATTAAAAATAGCAGACTCAG-3’
Upstream Reverse (P5)	5'-TTCATTGTTGGGTTATTAAAATCAGGTTTTTCCCTTAATTTATGCCCA-3'
Downstream Forward (P6)	5'-GTGGGCATAAATTAAGGGAAAAACTGATTTTAATAACCCAACAATGAA-3'
mazF Forward	5’-CGACTCACTATAGGGCGAATTGGGCCGCTTTCCAGTCGGGAAACCTG-3’
mazF Reverse	5’-CATATGCCACCGACCCGAGCAAACCCGAAGAAGTTGTCCATATTGGCCAC-3’

Appendix A Table 1. The primers listed in this table are the primers used to amplify the fragments that were used to transform *L. pneumophila* so that the bacteria would lose its gene encoding RavC. Overhangs of the primers are shown in underline.

RavC Forward	5’-ATGGCCATGGCGACTTTC-3’
RavC Reversed	5’-CCATGCTTGAGCGCTATTTAC-3’
gyrA Forward	5’-ATGCTATGCTCCCTGTCCGT-3’
gyrA Reversed	5’-GACTCCGACGAGGCTGTCAT-3’

Appendix A Table 2. The primers listed here are the primers used for the qPCR.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	60 °C	30 seconds
Extension	72 °C	1 minute 30 seconds
Final extension	72 °C	10 minutes

Appendix A Table 3. This table shows the parameters used to amplify the Upstream fragment using primers P1 and P2. The PCR was set to 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	55 °C	30 seconds
Extension	72 °C	1 minute
Final extension	72 °C	10 minutes

Appendix A Table 4. This tables shows the parameters used to amplify the Downstream fragment using primers P3 and P4. The PCR was set to 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	72 °C	30 seconds
Extension	72 °C	2 minutes
Final extension	72 °C	10 minutes

Appendix A Table 5. This tables shows the parameters used to amplify the mazF cassete using primers mazF Forward and mazF Reversed. The PCR was set to 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	60 °C	30 seconds
Extension	72 °C	3 minutes
Final extension	72 °C	10 minutes

Appendix A Table 6. This tables shows the parameters used for the first step double-joint PCR using primers P1 and P4. The PCR was set for 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	60 °C	30 seconds
Extension	72 °C	1 minute 30 seconds
Final extension	72 °C	10 minutes

Appendix A Table 7. This tables shows the parameters used to amplify the Upstream fragment using primers P1 and P5. The PCR was set for 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	55 °C	30 seconds
Extension	72 °C	1 minute
Final extension	72 °C	10 minutes

Appendix A Table 8. This tables shows the parameters used to amplify the Downstream fragment using primers P6 and P4. The PCR was set for 35 cycles.

Initial denaturation	98 °C	1 minute
Denaturation	98 °C	10 seconds
Annealing	72 °C	30 seconds
Extension	72 °C	3 minutes
Final extension	72 °C	10 minutes

Appendix A Table 9. This tables shows the parameters used for the second step joint PCR using primers P1 and P4. The PCR was set to 35 cycles.

25 °C	10 minutes
37 °C	2 hours
85 °C	5 minutes (inactivate enzyme)
4 °C	Hold

Appendix A Table 10. This table shows the parameters used for the cDNA synthesis.

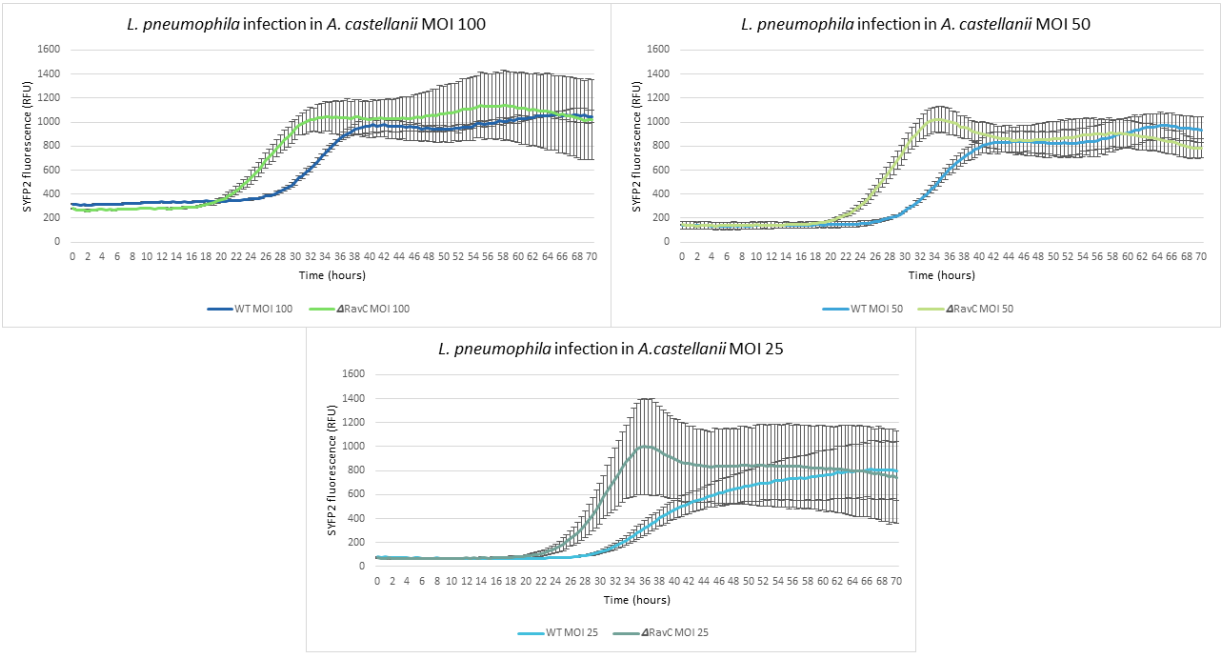
Polymerase activation	95 °C	10 minutes
PCR cycling,	95 °C	10 seconds

40 cycles	60 °C	30 seconds
Melt curve	95 °C	15 seconds
	55 °C	15 seconds
	95 °C	15 seconds

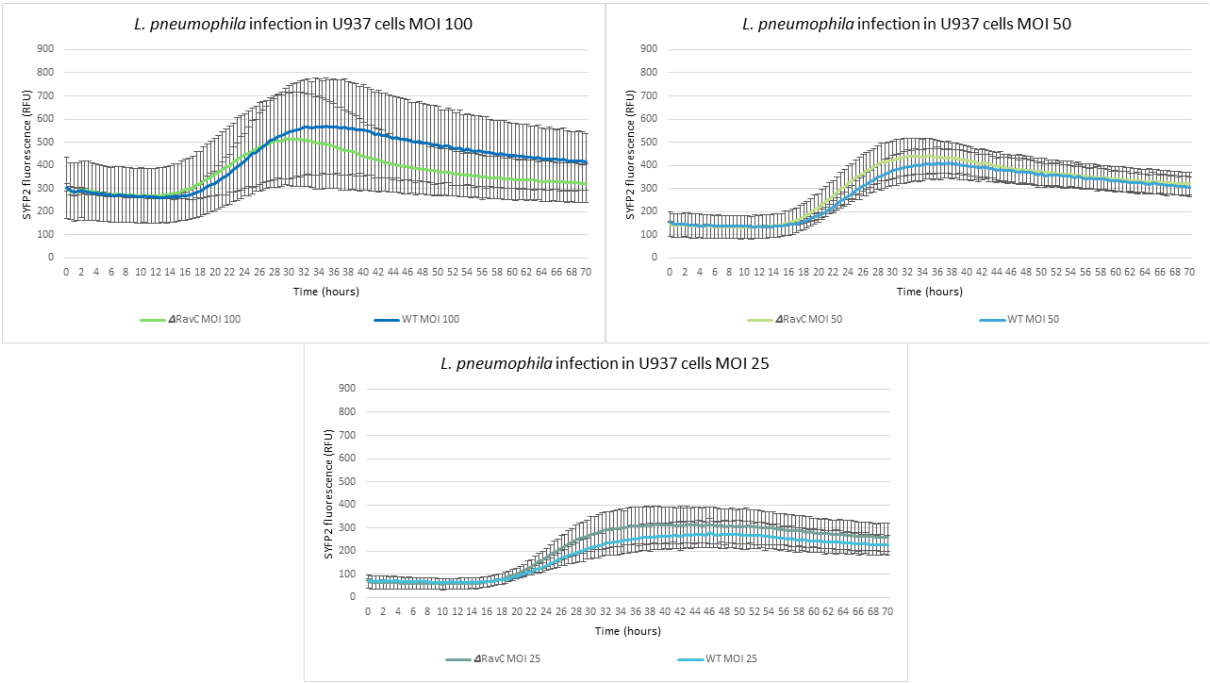
Appendix A Table 11. This table shows the parameters used for the qPCR.

9. Appendix B

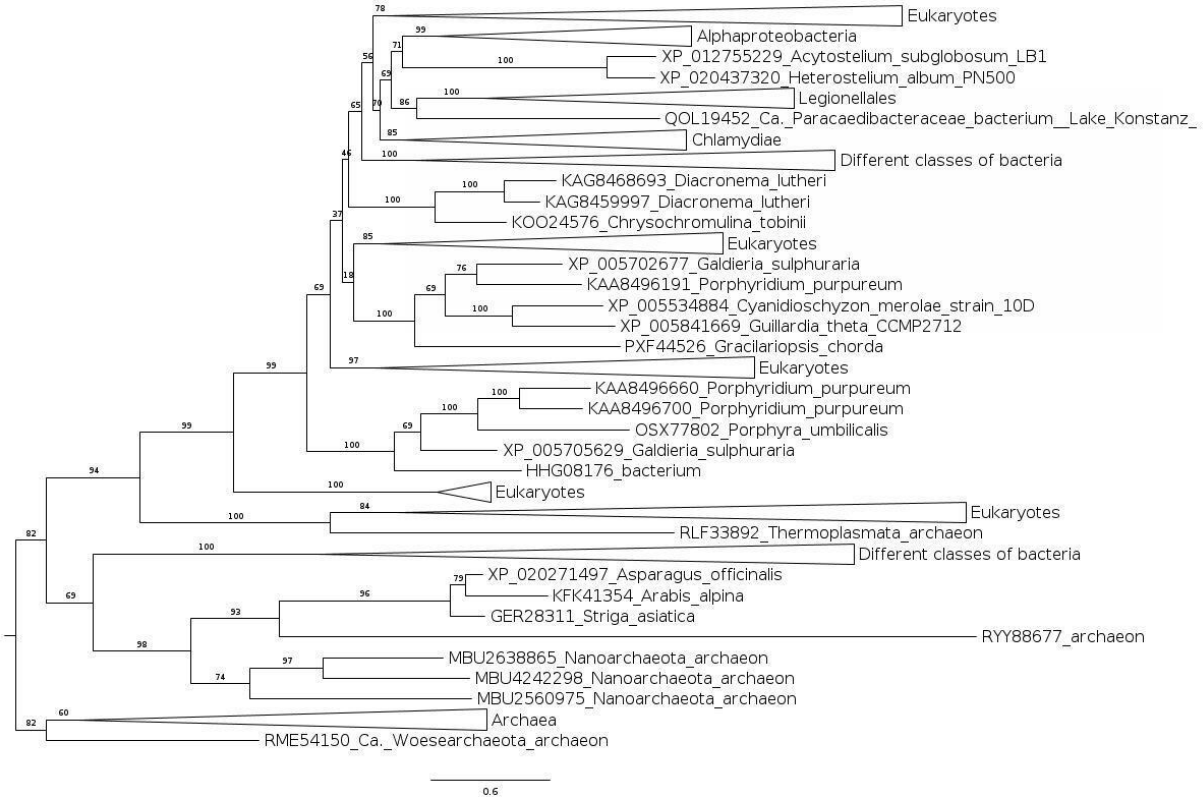
Appendix B contains supplementary figures that complement the figures shown in the results section.



Appendix B Figure 1: Growth curves showing the standard deviation (represented by error bars) during infection of ΔRavC and WT in *A. castellanii* at MOI's of 25, 50 and 100.



Appendix B Figure 2: Growth curves showing the standard deviation (represented by error bars) during infection of ΔRavC and WT in U937 cells at MOI's of 25, 50 and 100.



Appendix B Figure 3*:* An inferred phylogenetic tree based on H-CD-HIT clustering of sequences. Here showing the relationship between RavC in *L. pneumophila* and homologues in other species. Numbers on branches represent bootstrap values. The host-adapted bacteria are found in a monophyletic group.

10. Appendix C

Appendix C contains a supplementary discussion about protein purification of RavC.

We wanted to get an idea of the protein structure of RavC and to do this, we tried to purify RavC to later analyse the properties of the protein. Firstly, *E. coli* was transformed to express RavC and during harvesting of the cells, inclusion bodies were seen, which is especially common for proteins of eukaryotic origin and recombinant proteins (57–60). Even when using Triton X-100, which is a detergent that should allow for membrane proteins to solubilize, the protein was still found in the pellet (61). Because of this, we tried dissolving the pellet in GdHCl but the protein was still found in the pellet after centrifugation, and not in the supernatant as expected even though 6 M of GdHCl was used (62). Because of this, it is most likely that RavC is a transmembrane protein. This explains the difficulties of trying to extract and purify RavC from the transformed *E. coli.* It would require another protocol than what is being used for soluble proteins, since proteins must become soluble in an aqueous solution to purify (63). Due to lack of time, it was not possible to do more trouble-shooting by trying different detergents or to order the *ravC* gene again containing another tag, which may have solved the issues of not being able to purify the protein.