

*Mycoplasma pneumoniae's perspectives in
systems and synthetic biology.*

Jaime Cano Martínez-Botas



19th JUNE 2015

POMPEU FABRA UNIVERSITY

Degree in Human Biology

Faculty in life and health sciences

Index

Supervisor authorisation	2
Abstract	2
Problem and objective	2
Introduction	3
1. Discovery and classification	3
2. Morphology and metabolism	4
3. Systems biology and Minimal Cell.....	5
4. Systems biology approaches to minimal cell	5
5. Chassis cell	8
6. Models and Simulations	10
7. Pathogenesis	11
8. Epidemiology	13
9. Host cells and immunity	14
10. Current advances in M. pneumoniae.....	15
Methods	17
Results	17
Conclusions	18
References	19

Supervisor authorisation

Maria Lluch Senar, the supervisor of this project, authorizes to Jaime Cano Martínez to present this work as a final project for his degree in Human Biology.

Abstract

Mycoplasma pneumoniae was initially described in the 1940s as a highly pathogenic organism. Since then it has become recognized as a human lung pathogen which produces primary atypical pneumonia, among other extrapulmonary complications. Moreover, it stands as one of the smallest and best characterized bacteria, known by its reduced genome and lack of cell wall. There is a classification with two main subgroups or clinical isolates of this bacteria, types I and II, which appear to switch predominance in specific geographical areas through time. Nowadays, *M. pneumoniae*, thanks to its reduced genome, stands as an ideal candidate for achieving both goals of minimal cell and chassis cell within systems and synthetic biology. For these purposes it is necessary the identification of those genetic features associated with pathogenicity within the bacteria. Here we study 22 sequenced clinical isolates of *M. pneumoniae* through the analysis of SNPs, missense, indels and genome rearrangements. They revealed a new classification of strains and higher levels of CARDs toxin in Type II strains which may indicate increased virulence and pathogenicity in this group.

Problem and objective

The aims of this study is to overview the current status of *Mycoplasma pneumoniae* in the fields of systems and synthetic biology and the reasons why this bacterium stands as an ideal candidate for achieving the goals of both minimal cell and chassis cell. Furthermore it aims to explain the most recent advances in assessing *M. pneumoniae*'s mechanisms of pathogenicity, which will help to better understand this small bacterium.

This last study on *M. pneumoniae*'s pathogenicity is based in the idea that the classification of the different clinical isolates is up to now carried out by using nucleic acid amplification tests (NAATs) that compare only a few genes and intergenetic regions. This methodology makes unclear of what other differences could exist in their genomes and therefore its effect on pathogenesis, the immune system or their response to antibiotics. For this reason, this study aims for the identification of virulence factors and antigenic proteins by doing a comparative genome study with 27 clinical isolates. Moreover, a full transcriptome and proteome analysis of four strains, representative of both main subgroups, was carried out as well.

Introduction

1. Discovery and classification

Mycoplasma pneumoniae was firstly isolated from the sputum of a patient in the 1940s as a highly pathogenic organism, originally thought to be a virus, named the “Eaton Agent” (1). Several tests with volunteers in the 1950s and early 1960s proved that the pathogen caused lower respiratory tract infections and that was treatable with antibiotics, therefore it was proven that it was not a virus, yet until 1963 it wasn’t cultured in free-cell medium and taxonomically classified (2).

The term “mycoplasma” (Greek; “mykes” = fungus and “plasma” = formed) appeared in the 1950s and replaced the older genera terminology, though the allusion to the fungus-like growth pattern is only applicable to *M. mycoides* (3). This genera belongs to the class Mollicutes (figure 1), which derives from Latin words meaning soft (“mollis”) and skin (“cutis”). They are distinguished by the absence of cell wall, for its small size (between 0.2 and 0.3 μm) and for being parasites, living on or in their host’s cells.

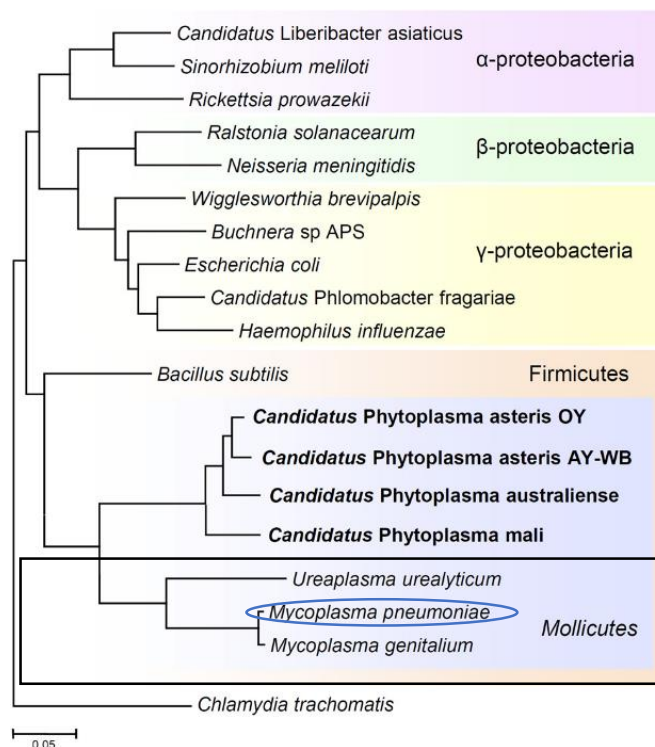


Figure 1. Phylogenetic distance tree constructed by the neighbor-joining method, comparing the 16S rRNA gene sequences. (42)

Mycoplasmas are up to now the smallest self-replicating organisms capable of cell-free existence, regarding to both cellular and nuclear dimensions. *M. pneumoniae* possesses

a circular genome with about 800 kb that encodes 687 genes. Its small genome seems to be the result of a gradual reduction in genome size from a gram-positive ancestor, shared for example with streptococci or bacilli gram-positive bacteria, by the process of degenerative evolution (4). This process is responsible for the genome size reduction of other bacteria such as *Buchnera aphidicola* (5).

Mycoplasmas appear in nature widespread as parasites of humans, mammals, reptiles, fish, arthropods, and plants. There have been identified more than 200 *Mycoplasmas* and there are continuously increasing, yet among all of them the best known and most studied is the human parasite *Mycoplasma pneumoniae* (6)(7).

2. Morphology and metabolism

M. pneumoniae has a spindle-shaped morphology with 1-2 μm long and 0.1-0.2 μm wide, compared with a standard bacillus of 1-4 μm in length and 0.5-1.0 μm width. Its cell volume is less than a 5% of the one of a typical bacillus. This small size volume allows it to pass through the 0.45 μm pore filters that are commonly used to filter sterilized media. Furthermore, due to its small cellular mass it cannot be detected by light microscopy nor it can produce turbidity in liquid growth medium.

It has a very limited biosynthetic activity for the production of proteins, carbohydrates and lipids. It needs and scavenges for nucleotide precursors as it is not able to synthesize purine and pyrimidines *de novo*. The ATP is obtained by fermentation of glucose to lactic and acetic acid, by substrate-level phosphorylation, mediated by several kinases such as pyruvate kinase and acetate kinase; other products, for example glycerol and small carbohydrates can also be used for ATP generation. As a remark, its capacity to reduce tetrazolium has used historically for distinguishing the bacteria from other commensal oropharyngeal mycoplasmas (4).

This limited biosynthetic capabilities along with its small genome are responsible for the complex media requirements needed for this organism to be cultivated *in vitro*. This requirements were obtained in a study in which a manually curated metabolic network catalysed by 129 enzymes was produced, allowing the design of minimal medium with 19 essential nutrients for both *M. pneumoniae* and *M. genitalium* (8). As an example, they require sterols in its growth media, as they are an essential component of the triple-layered cell membrane which provides structural support to this osmotically fragile organism.

Mycoplasmas are not found freely living in nature, they are parasites, and for this reason they depend on a host to be supplied with the essential nutrients. The reductive

evolutionary process that led to the small genome of *M. pneumoniae* suggests that they might exist as a facultative intracellular pathogen. There is evidence that supports this idea of facultative intracellular pathogen. Some recent studies have proven *in vitro* that the bacteria can take up residence and replicate within cultured cells for long periods. However it remains to be proven during natural or *in vivo* infections. Lately, in a differentiated normal human bronchial epithelial cell culture this intracellular invasion was not observed after infection, therefore more studies should be carried out to test the hypothesis. (7)

3. Systems biology and Minimal Cell

The cell is the most complex micrometre-sized known structure for humans. Despite all the advances on the recent years, for example with the characterisation of new omics, we are still far from understanding the interactions between its components that, in the end, are responsible for its proper functioning. To face this issue, we should consider an idea gathered from the beginning of the era of molecular systems biology, that a cell could be as simple as we could define life in its simplest form (9).

This is the concept of minimal cell, which at present can only be defined, on a semiabstract level, as a living cell with a minimal and sufficient number of components (10). It should have this three main features: (I) the metabolism responsible for the acquisition of the molecular building blocks and energy required for synthesizing its cellular components, (II) the genetic replication machinery or an equivalent information processing and transfer machinery mechanism, and (III) a boundary or membrane that separates the cell from its environment. (9)

There are two common but different ideas regarding the concept of minimal cell. The first one is relating minimal cell with the cell that has the smallest number of components, meaning genes and proteins, while the second one is relating minimal cell with the so-called simpler cell or lowest complexity cell, which in theory, once achieved, should be easier to predict and manipulate. However, up to now the first concept is far easier to assess and measure whereas the second remains an issue to be tackled in the future (9).

4. Systems biology approaches to minimal cell

There are four categories in which systems biology approaches to minimal cell can be divided. The first two ones are the classical top-down, based on deconstruction of systems in an analytical point of view, and bottom-up, pursuing synthetically construction of systems. The third one is the middle-out approach which consists in a large-scale data integration, modelling and simulations for the study of simpler and minimal cells. Finally,

the fourth one are the system-level comparative studies, based on whole “-omics” comparison between organisms, both for systems biology understanding and minimal cell construction (9) (figure 2).

1- Top-Down approach

The Top-Down approach has traditionally referred to as reductionism and implies the removal of the nonessential components of a system until it is no longer functional. It allows to understand the individual function of each of these components within the system as a whole. In the minimal genome search it has been carried out by knocking out genes to determine which ones are not essential. An example is the reduction of *B. subtilis* genome by sequentially combining smaller deletions (11).

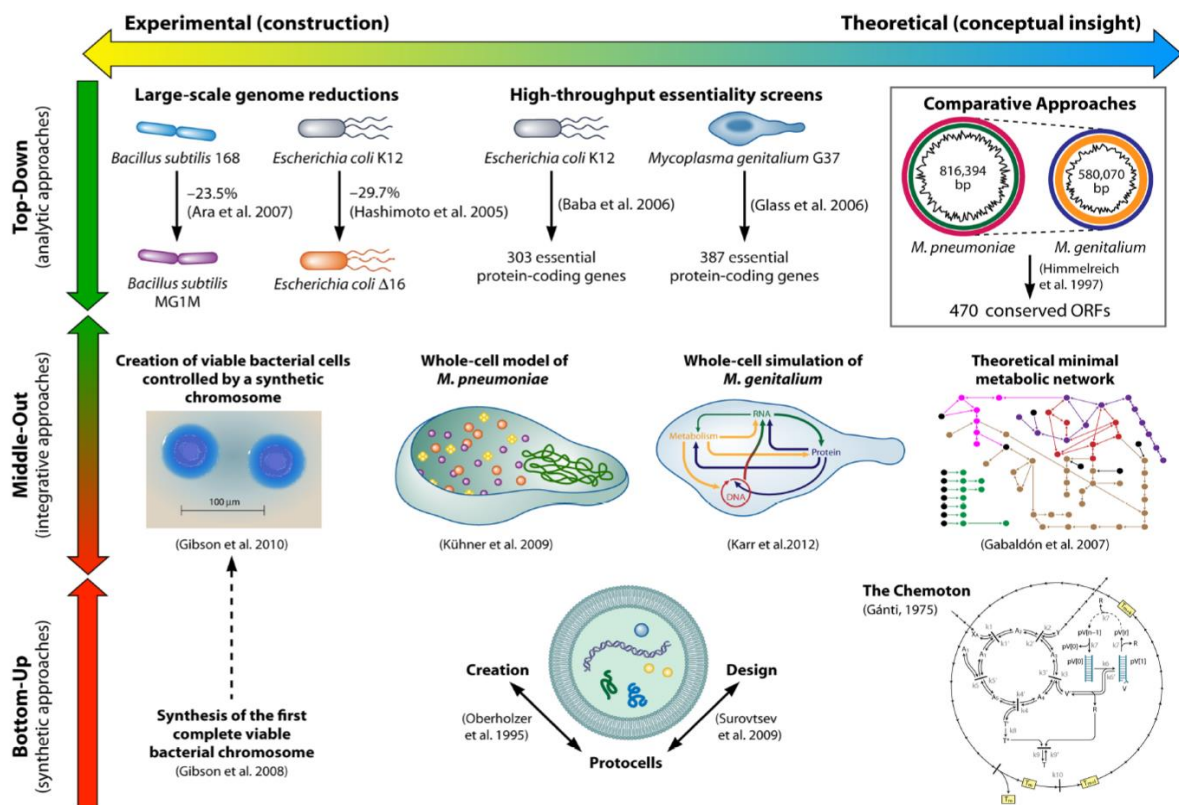


Figure 2: Systems approaches and relevant results toward understanding and designing minimal or simpler cells (9).

2- Middle-Out Approach

The Middle-Out approach was originally attributed to Sydney Brenner by Kohl and Noble, as he coined it during a Novartis Foundation symposium. There is an adapted definition for prokaryotic cells from Noble’s original definition, that is: “the approach which starts at any level (gene, RNA, protein, metabolic or regulatory pathways) at which there are sufficient data and reaches (up, down and across) toward other levels and components”

(9). The middle-out approach is sometimes difficult to distinguish from the above mentioned classical approaches, yet it is comprised of those studies which integrate different layers of information in a holistic model or construct following the concepts of minimal cell (figure 3).

An example of an integrative approach which results in an original construct is the whole-cell tomogram of *M. pneumoniae*, obtained by electron tomography of 26 entire cells (12). A whole-cell illustration of the spatial organisation of the proteome was accomplished by a combination of a pattern recognition and classification of algorithms for protein complexes positioning.

The major achievement from the middle-out approach which represents the climax of experimental projects towards the creation of artificial cells is the fully transplantation of an artificial chromosome from *M. mycoides* to the *M. capricolum* recipient cell, creating new cells that were controlled by the synthetic chromosome (13).

TABLE 1 Concepts related to minimal or simpler cells

Concept or construct	Short definition
Minimal genome	Simplified genome without nonessential genes (under specific environmental conditions)
LUCA	Life form commonly accepted to have existed before the divergence of the domains <i>Bacteria</i> , <i>Archaea</i> , and <i>Eukarya</i> ; hypothesized to have been inorganically hosted ^a
Chassis cell	Cell designed for use in industrial production processes with a high degree of controllability and efficiency
Artificial/semiartificial cell	Cell built in the laboratory (at least partially) with resources to extant genetic and other biological material
Minimal cell models	
Protocells	<i>In vitro</i> models of a minimal cell usually containing some kind of biological material encapsulated in liposomes or other lipidic vesicles
<i>In silico</i> minimal cells	Virtual model/reconstruction of any of the possible constructs described above or any other model of a minimal “ome” relevant to the study of the minimal cell

Figure 3. Concepts related to minimal or simpler cells (9).

3- Comparative approach

The comparative approach is based on comparative genomics, involving whole genomes and inferred proteomes. It is based on the idea that those genes that are conserved during evolution have much more probability to be essential. The best known is the 16S rRNA, used for phylogeny.

Now there is a new wave of comparatives studies that integrate proteogenomics to validate genetic conservation, using high-throughput tandem mass spectrometry to verify

the expression of the predicted and conserved coding regions. (9) With this new exponential increase of studies that compare proteomes, transcriptomes, or fluxomes, there is a need for an additional layer of information, the experimental conditions. The comparison of this omics has huge promising expectations, yet it can be really challenging due to the fact that most of the studies were not done under the same conditions. Even small variations in a complex media can impair comparisons and for this reason it is imperative the generation of a controlled experimental data.

Ultimately, the recognition that the Top-Down and comparative approaches are insufficient to reduce the complexity to the level of cell comprehension has led to the idea that the only way left is to build or synthesize a minimal cell from its parts, exactly what the Bottom-Up approach intends to achieve.

4- Bottom-Up Approach

The Bottom-Up approach, also called synthetic approach, is aimed at assembling the minimal cell in the laboratory. Most of this studies have focused in physical and chemical properties of the cell or in the dynamics of the building blocks of life.

Most of the focus has been placed on the creation of protocells, by inserting genetic material (RNA or DNA) or enzymes inside lipidic vesicles. They are useful for studying properties such as stability, selfreproduction, permeability as well as the dynamics of biochemical reactions.

The synthesis of the first artificial bacterial chromosome is probably the major landmark of this kind of study (14). Though a complete cell was not created per se, the technology for the creation of the code for an artificial cell was established. However, even if we are getting deeper understanding of biosystems for now it is still a nascent technology (9), as there are important barriers such as information on gene networks and their interactions that remain to be overcome (15).

However, we are in an era in which the advances are speeding up. Between these new milestones, it has been estimated that the synthetic genome of *M. mycoides* harbouring 828 ORFs, assembled and “booted-up” by Gibson et al.(13), will be reduced about 450 genes by removing non-essential genes (16).

5. Chassis cell

A concept that is related to the minimal cell is the platform or chassis cell. It can be defined as a cell with reduced complexity that is designed for one or more

biotechnological applications and can be modified and controlled with precision in a predictive manner (9,17).

It seems without doubt that for a chassis cell design a pragmatic, integrative and knowledge based approach is needed. The biotechnological applications of the minimal cell or minimal data set achievements are probably the reason of the huge interest that is catching recently the topic. In fact, as stated above, the terms chassis cell or factory cell are preferred when we refer to the minimal cell that is intentionally simplified for its use in industry. It implies a much more specialized and comprehensive cell for biological production of chemicals or pharmaceutical molecules, among other applications. There are some characteristics that have been described as essential requisites for an optimal minimal cell (figure 4) from two recent reviews (17,18).

TABLE 2 Requirements for an industrially relevant chassis cell	
Requirement for a chassis cell	
Overall simplicity	
Minimal no. of carbon sinks and other nonoptimal flux paths	
Predictable metabolic and regulatory networks (more control over growth and production)	
Simplified translation code	
Reduced genetic drift and limited evolvability	
Robust mechanisms for genome replication, cytokinesis, and coordination in between	
Robust cell membrane and cell wall that confers resistance to shear stress in bioreactors	
Efficient transcription, translation, and regulation for optimization of cellular fluxes to desired goals	
Availability of predictive mathematical models that save expensive trial resources	
Process-specific modules for implementation of different industrial solutions (particular for each process)	
Other stress tolerance mechanisms, such as:	
Product tolerance	
High substrate tolerance	
Tolerance to low O ₂ levels	

Figure 4. Requirements for industrially relevant chassis cell (9).

In the industrial bioprocesses, in an opposite way to the scientific discovery, no surprises are desired. For this reason, the chassis cell which ensures a total control over the cell processes, meaning not only over the minimal processes of the cell but also of any novel biological pathway introduced into it (19), is ideal.

However, this is an area of great debate. Some people argue that evolution is inevitable and it is intrinsically linked to DNA replication, so the stability of those chassis cells remains to be studied. Others argue that natural, robust cells constitute a better chassis, as large genomes of robust cells encode multiple metabolic pathways and are therefore

self-sustainable under diverse growth conditions (20), which minimal cells are not yet able to do. Only the construction of the true minimal cells and comparison with robust cells will provide a definitive answer on its suitability for industrial applications, yet tighter control and the possibility of easier manipulation in the desired direction might turn out to be the deciding arguments in favour of the minimal cell (15).

6. Models and Simulations

Up to now the simpler or minimal cells are just conceptual ideas and for this reason theoretical and mathematical models are needed for the advancement of the field. The whole cell model simulation is a field which is very scattered right now. In general, it requires modelling of different biological networks in an appropriate scale. There are three categories in which these models could be classified: interaction models or network representations, constraint-based models and mechanistic models (9). Yet all of them must achieve a series of milestones for being optimal.

The first of them would be a complete, verifiable mechanistic simulation of all relevant molecular interactions involved in a bottom-up implementation of dynamic processes such as reproduction or membrane formation, as well as of each proto-(metabolic, transcriptional, signalling) networks (21,22). Secondly, the other milestone would be the integration of this models into a comprehensive simulation of the entire organism, in order to get its cell cycle.

Meeting goals offers the unique possibility of developing a co-design strategy (23), which would be the simultaneous step-by-step development of experimental techniques and computational models in order to obtain complete knowledge of what goes into the experiments. This strategy could totally change the way in which bottom-up research is performed. Ultimately, harnessing and engineering a collective multi-cellular behaviour would require a detailed understanding of the sources of noise in (proto)cellular systems as well as the practical strategies needed for programming both “patterned” noise and cell-to-cell communication. A key goal will be the integrated simulation of large hybrid protocell-cell systems (22).

There is a recent work which reported a whole-cell model of *M. genitalium* that allows accurate phenotypic predictions (24). The model divides the total functionality of the cell into 28 essential cellular processes represented in different submodels and fall into five categories: DNA, RNA, protein, metabolism, and other (such as cytokinesis and host interaction). Each model has been independently build and has its own mathematical representation, due to the fact that some process are better studied or understood than others. The key to this study is the integration of these sub-models, where the used an

integrative strategy that is based in the assumption that on short time scales (less than one second) the submodels are approximately independent. However, at each step the submodels depend on the values of variables determined by the other submodels at the previous step. These whole-cell model allows the most complete simulation of *M. genitalium* so far, it explains a variety of emergent behaviors in terms of molecular interactions, it provides insight into biological processes for which experimental assessment is not readily feasible yet and enables the rapid identification of novel gene functions as well as specific cellular parameters (24).

Even though this formulation of independent and decoupled modules represents the most complete simulation of *M. genitalium* so far it is only a “first draft”, therefore extensive effort are required before the model can be considered complete. Between the drawbacks of the project there is the lack of quantitative data from *M. genitalium*, which should be updated with the actual knowledge, along with some technical and modelling aspects that need refinement and expansion such as the integration time of only 1 second which is not really reproducing the biology of the cell in real time. In future attempts for creating a whole cell model a series of improvements derived from this initial idea must be pursued for a complete success *in silico* cell modelling (9,24).

Taking all into account, this simulations, along with their experimental counterpart, will be crucial goals towards the future engineering of robust, reliable and efficient synthetic organisms for producing smart drug delivery systems, tissue/organ enhancement techniques or more general applications of artificial living matter (22).

7. Pathogenesis

P1 Adhesin protein

There is evidence to suggest that the attachment of *M. pneumoniae* to the respiratory epithelium is a prerequisite for the events that eventually produce disease. The interaction itself protects the bacteria from its removal by the host's mucociliary clearance mechanism, allowing the production of local cytotoxic effects. There it lies the reason why this attachment organelle has evolved as a complex and specialized tip that has many adhesin proteins such as P1 adhesin protein.

The P1 adhesin is a 170 kDa protein that is known to be the major structure responsible for the interaction of *M. pneumoniae* with its host. Its loss through spontaneous mutation or by trypsin treatment produces avirulence due to the reduction in adherence to the eukaryotic cells. The restoration of the functional phenotype is accompanied by reappearance of the structural and functional tips and full infectivity. Moreover,

monoclonal antibodies against P1 block adherence in hamster models while on other proteins they have no effect on the attachment (3).

Peroxide

Besides cytoadherence, *Mycoplasma pneumoniae* is believed to cause disease in part through generation of peroxide. In a *M. mycoides* study, it has been observed a correlation between pathogenicity and peroxide formation in the form of a by-product of glycerol metabolism, L- α -glycerophosphate oxidase (25). *M. pneumoniae* has the same enzyme and metabolism of glycerol which result in peroxide production, it is reasonable to propose that the same metabolic pathway contributes to disease in *M. pneumoniae*. The loss of glutathione, peroxidation of lipids, denaturalization or haemoglobin and lysis of the cell are among the ultrastructural effect that peroxide has on host cells such as erythrocytes.

The superoxide anion produced by *M. pneumoniae* inhibits the catalase in host cells, making them much more susceptible to oxidative damage, as it is inhibiting the defensive reduction of peroxides induced endogenously by the catalase, rendering the host cell much more susceptible to oxidative damage. *M. pneumoniae*'s hemadsorption and lysis of erythrocytes that are low in catalase is property used for diagnostic testing and for distinguishing between other commensal bacteria found in the respiratory tract, as they lack the hydrogen peroxide production mechanism and do not hemadsorb.

CARDS toxin

ADP-ribosyltransferase protein was identified in 2005 in *M. pneumoniae*, encoded in the gene *mpn 372*, which has a significant sequence homology with S1 subunit of pertussis toxin and it has been suggested as a potential exotoxine during mycoplasma infection.

This toxin is now called community-acquired respiratory distress syndrome toxin (CARDS TX).

The gene encoding the toxin was found to have mRNA levels increased substantially during infection of mammalian cells. Has been proven that this protein produces a robust inflammatory response in mice consistent with allergic disease (26), that it binds with high affinity to human surfactant protein-A, along with the production of specific biological activities such as mono-ADP ribosylation and vacuolization (27). In mice, differences of CARDS toxin levels have been reported to differ between strains, suggesting that its levels could be related to the ability of the specific strain to replicate, colonize and persist (28).

However, *M. pneumoniae* lacks the homologous proteins that translocate it from the pathogen to the host cell cytoplasm found in the pertussis toxin, meaning it remains unclear how it reaches the host. It could be that the C-terminal moiety of the CARDS toxin could have this ability, as the protein itself is immunodominant and has a different coding sequence than its homologue. Moreover, it is still unclear how this toxin relates with surfactant protein A by means of function. A deeper understanding of this protein could lead to a proper comprehension of *M. pneumoniae* infection mechanism as it is clearly related to the bacteria pathogenicity.

8. Epidemiology

M. pneumoniae infections can involve both the upper and the lower respiratory tracts and occur both endemically and epidemically worldwide in children and adults. Although nor climate or geography are thought to be of major significance, the proportion of patients with pneumonia due to *M. pneumoniae* is greatest during the summer in temperate climates due to the lower incidence of other respiratory pathogens on that season. Despite most of the data comes from studies performed in the United States, Europe and Japan, seroprevalence investigations in arctic and tropical zones have proved the presence of *M. pneumoniae* antibody, which suggests that populations in these regions have had infections induced by this organism (29).

In 2000 Dorigo-Zetsma et al. genotyped *M. pneumoniae* clinical isolates and grouped them into eight subtypes within two genomic groups based on P1 adhesin subtypes, type I and II subgroups (30). Through the characterization of over 200 *M. pneumoniae* isolates from Europe and the United States, *M. pneumoniae* has proven to be a rather uniform microorganism, as most of the isolates could be classified into the two initial subtypes. This classification is based now not only on the sequences of the P1 adhesin gene but also in the ORF6 gene, and the P65 gene and by a typical DNA restriction fragment pattern maintains the initial subgroups (I and II) (3). These studies found either one or the other subgroup tended to predominate in a specific geographical area, and over time there were changes with respect to which subgroup the majority of the isolates belonged to. The pathogen seem to act in a series of cycle epidemics pattern, as observed in a study carried out in Denmark over the 50 year period between 1946 and 1995 the peaks occur every 3 to 5 years (3,4).

Furthermore, *M. pneumoniae* was responsible for 15-20% of all cases of community-acquired pneumonia (CAP) between 1962 and 1975 in Seattle (7). Extrapolating data from different studies provides an estimated number of CAP cases in the US that exceeds the 100.000 on an annual bases. The incidence of pneumonias caused by *M.*

pneumoniae increases by age in hospitalized adults, only second to the one produced by *Streptococcus pneumoniae* (7). Moreover, many of CAP cases are treated as outpatient and this numbers could be a much greater.

The persistence in the respiratory tract, low transmission rate and the long incubation period may explain this prolonged duration of epidemics. Hopefully, morbidity is uncommon, yet the acute illnesses are quite often disruptive and consume large amount of resources.

Finally, *M. pneumoniae* has been implicated along with proinflammatory cytokines release as a possible mechanism to chronic pulmonary diseases such as bronchial asthma and adult respiratory distress syndrome (ARDS) (29). The important question for the researchers is if it is the primary cause or just a cofactor in its development.

9. Host cells and immunity

M. pneumoniae possesses both protein and glycolipid antigens that elicit antibody responses in infected patients. The P1 protein is the target of many of the antibodies that are produced by the host in response to the infection and has been used as well as a target for development of serological assays. In addition to *M. pneumoniae* specific antibodies, a variety of cross-reactive antibodies may develop in association with *M. pneumoniae* infection.

The extensive sequence homology of the *M. pneumoniae* adhesin proteins and other proteins, such as glycolipids from the cell membrane, with mammalian tissues is today a well-known example of molecular mimicry that might trigger autoimmune disorders that involve different organ systems. This process is produced through the formation of antibodies against endogen substances such as myosin, keratin, fibrinogen, brain, liver, kidney, smooth muscle, and lung tissues (29). This reaction was observed back in the early 1918s, when pneumonias were often associated with cold agglutinins, which are characterized for recognizing the antigen I of erythrocytes, a carbohydrate of surface glycolipids and proteins and therefore produce autoimmune disease (7). There is strong data that links peripheral neurologic syndromes to pathologic antibodies against carbohydrate moieties expressed on variety of gangliosides, such as GM1. Moreover, about 5-15% of Guillain-Barré syndrome cases have been associated with a preceding *M. pneumoniae* infection, this were much more likely to have antibodies against galactocerebrosides. Moreover, there is some data linking *M. pneumoniae* with encephalitis, optic neuritis, acute demyelinating encephalomyelitis (ADEM) (7).

Besides this antigenic mimicry as autoimmunity mechanism, there are studies that showed directly activation of cells from the immune system for cytokine production. There is an induction of IL-4 by mast cells dependent of sialic acid residues on the target cell membrane and the P1 adhesin protein. A model on cellular activation has been developed and links adherence by the bacterium to surface sialoglycoproteins to cellular activation by receptor-signalling mechanisms, the process is enhanced in mast cells by H₂O₂ from the organism (7,31).

The inflammation induced by cytokine production in leukocytes, lymphocytes, respiratory epithelial cells and monocytes does not require cell contact and can be directly produce from culture supernatants. This activation is carried out through the activation of toll-like receptors 1 and 2 (TLR1 and TLR2). The role of all this cytokines and other substances in pathogenesis has been a topic of interest during the past years. Current evidence suggests that they may either minimize disease through the enhancement of host defences or exacerbate the infection through immunological lesion development. This immune-mediated lung disease provides a basis for consideration of immunomodulatory therapeutics along with conventional therapies, based on antimicrobial drugs, for the management of the *M. pneumoniae* infection (3).

10. Current advances in *M. pneumoniae*

The genome, thanks to sequencing advances, was the first of the “-omics” to be available for systems biology research, this implies that it is the most studied and characterized one. For this reason, the search for the minimal genome represents the state of the art of the minimal cell field.

A first definition of this minimal genome could be: “the smallest possible group of genes sufficient to sustain a functional cellular life form under the most favourable conditions imaginable, which is the presence of a full complement of essential nutrients and the absence of environmental stress” (9,32). The phrase “most favourable conditions” must be emphasized as it indicates that in practice the minimal cell could have extremely complex nutritional requirements. The smallest prokaryotic genomes sequence to date belong to species that are not autonomously alive, as by missing essential genes became dependent from the much more complex host.

Besides genomes, minimal protein sets have been recently begun to be inferred thanks to the integration of experimental data from different sources. This effort is a step forward from the inference given by minimal genomes towards a new one of minimal proteomes. There is a study on *M. pneumoniae* which reveals the interactome of this bacteria (12). The authors concluded that most of Clusters of Orthologous Groups of proteins (COGs)

in the Mollicutes core proteome (140 COGs) are expected to be associated in protein complexes and that 54 COGs are also participating in more than one complex. These complexes are involved in secondary functions such as the maintenance of the overall cellular and genomic stability, which could explain the maintenance of incomplete metabolic pathways in reduced genomes. For the authors the concept of minimal genome should be more than a set of essential functions a set of essential structures (33).

In this context, *M. pneumoniae* has several favourable properties for organism-wide analysis. It has a massive genome reduction into only 800kb, encoding 689 proteins. Its expressed proteome is low complexity, with only three orders of magnitude in abundance. It lacks metabolic pathways forcing it to acquire building blocks from the environment, and it is able to be grown autonomously in a laboratory culture.

Recent studies have reported a large variety of analysis on *M. pneumoniae*. In these studies it has been unveiled the proteome (12), transcriptome (34), a metabolic network that has allowed the identification of the medium that supports *M. pneumoniae* and *M. genitalium* growth (8), protein modifications (35), proteins half-lives (36) and a completely re-annotated genome of the M129 strain (37).

Furthermore, it was recently unveiled the *M. pneumoniae*'s metabolome (38). In this work a comprehensive metabolic model was developed, based on a previous manually curated metabolic map (8). Through the establishment of the biomass composition and analysis of energy metabolism it was verified that the major energy sinks are protein and RNA production as well as biomass production, in accordance with previous models from *M. genitalium* (24). Interestingly, the ATPase uses about 57-80% of total energy generated for maintain a favourable proton gradient across the membrane, which could be explained by the small size of the bacteria that makes it much more susceptible to membrane leaking.

One of the latest studies has unveiled an essentiality study in *M. pneumoniae* (39). The essential genome of an organism comprises protein-coding regions (ORF), regulatory and structural elements, yet most of the studies on this topic have used conventional genome annotations which are biased against small proteins and regulatory elements. An accurate essentiality study is limited to the completeness of the genome annotation, this is the reason why *M. pneumoniae* stands an ideal organism. In this key study carried out by Lluch-Senar et al., they used two mini-transposon mutant libraries and high-throughput insertion tracking by deep sequencing (HITS) to determine the essentiality map of the bacteria. Among their discoveries they observed that multi-domain proteins

involved in protein complexes are frequently more essential than proteins with single domain, as they are involved in important cellular processes such as DNA replication. This sets a new criterion for future studies, in which essentiality should be considered at a protein resolution and that smORFs along with regulatory elements, frequently essential genomic elements, should be considered for the minimal genome.

All these studies are a tremendous step forward in the integration of “-omics” in the minimal cell panorama and for the usage of the holistic system perspective for the study of single species.

Methods

I have contributed in the study by analysing and refining the genomes of those above mentioned strains. For analysing the genomes I have used the Integrative Genomics Viewer (IGV) which is a high-performance viewer that efficiently handles large heterogeneous data sets while providing a smooth and intuitive user experience at all levels of genome resolution. It focuses on the integrative nature of genomic studies, with support for both array-based and next-generation sequencing data, and the integration of clinical and phenotypic data (40,41). With the use of this program I was able to localize insertions and deletions in the genome which were used to cluster the strains in a new fashion.

For the verification of those overlapping regions of the genome I have use PCR and sequencing techniques, in order to clarify the data belonging to the four representative strains.

Results

Genome sequencing allowed the identification of all SNPs, missense, Indels and rearrangements in the genome of different *M. pneumoniae* strains refining the typing and revealing sub-classes in the two main groups. It allowed for the creation of a new classification of strains based on genome differences (figure 5).

Moreover, integrative analysis of *in vitro* gene essentiality and mutation rates allowed the identification of some virulence factors and antigenic proteins; revealing that glycerol metabolism and peroxide production are important factors in the physiology of these pathogenic strains. Additionally, transcriptome and proteome data allowed characterizing the impact of mutations in the levels of RNA and proteins. The study revealed that a single mutation in the *mpn372* gene increases the levels of CARDs toxin

in Type II strains. Since this protein was shown to induce a strong immune response, seems that Type II strains could be more virulent than Type I strains.

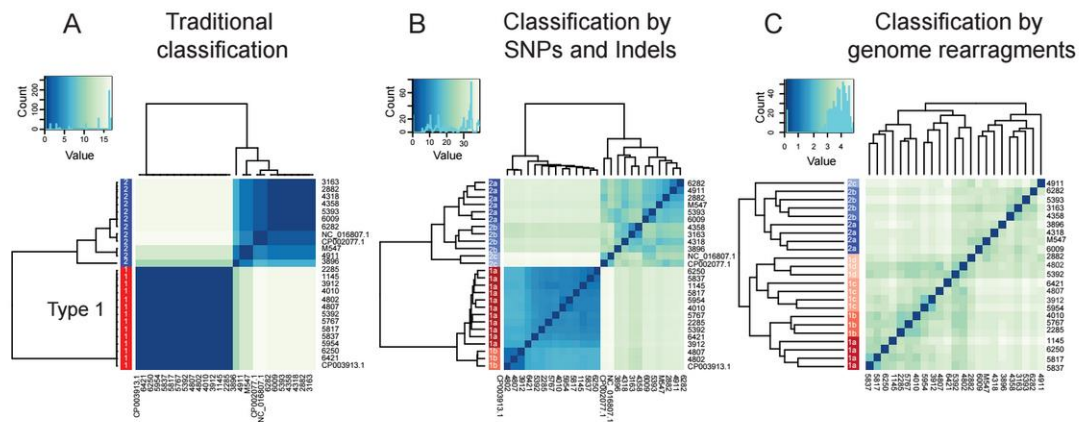


Figure 5. Clustering of different *M. pneumoniae* strains (data still unpublished).

Conclusions

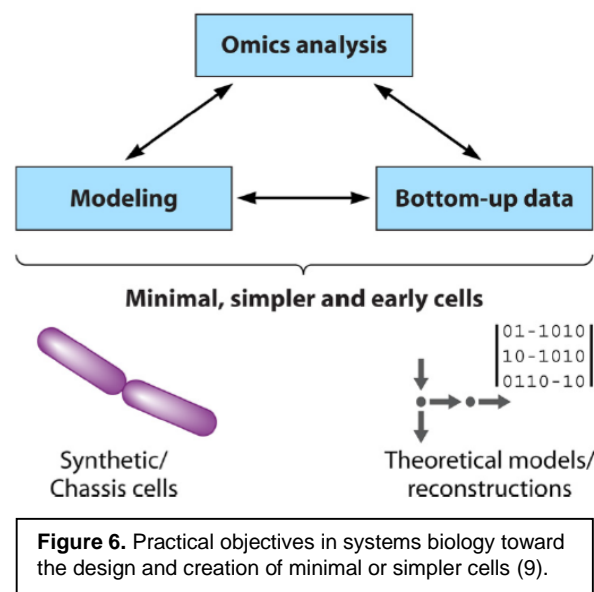
The famous American physicist and Nobel Prize Richard Philips Feynman wrote this on his blackboard in 1988: ‘what I cannot create I do not understand’. By reversing this, we get the more obvious ‘what I do not understand I cannot create’, which might in fact be the perfect metaphor of the paradoxical nature of synthetic biology (22).

This complexity of a cell is shown the above mentioned reports by the group of Luis Serrano and co-workers on the naturally reduced bacterium *M. pneumoniae*. They concluded that this minimal organism exhibits a proteome complexity that ‘could not be directly inferred from its genome composition and organization or from extensive transcriptional analysis’ (12). In other words, the closer to life understanding get the most intricate life appears to be, which makes the machine-like orthogonalization an utopical, oversimplified metaphor for synthetic biology. Then, if we apply Feynman’s philosophy, there’s no hope for truly synthetic life in a close future.

However the fact is that we will be able to create what we do not fully understand, by using already functional parts and by integrating them in rational processes under the constant guiding aide of selection and evolution as well as modelling bioinformatics. This ‘assisted biological design’, combined with the high throughput DNA synthesis and transplantation techniques, might prove a revolutionary tool for the implementation of large variety of biotechnological applications such as sustainable energy production, bioremediation strategies or biomedicine (22).

This work in *M. pneumoniae* mechanism of pathogenesis has allowed the identification of all SNPs, missense, Indels and rearrangements in the genome of different *M. pneumoniae* strains refining the typing and revealing sub-classes in the two main groups. This new data will allow for further studies to search specifically for those differences between strains and their effect in pathogenesis and response to antibiotics and immune system. The study as a whole revealed one of this differences, that a single mutation in the *mpn372* gene increases the levels of CARDS toxin in Type II strains, a protein that induces a strong immune response as it has been stated above. This led to the conclusion that Type II strains should be more virulent than Type I strains (data still unpublished).

This project has pictured an overview of the most recent advances of the systems and synthetic biology fields in relation to the small bacterium *M. pneumoniae* and the final goal of achieving the minimal cell and the chassis cell. In this context the unpublished pathogenesis study stands as another step forward in the direction of a better understanding *M. pneumoniae*. Further work with this genome reduced bacterium is required for addressing these final objectives (figure 6).



Acknowledgments

I would like to express my very great appreciation to Dr. Maria Lluch Senar and Dr. Eulàlia de Nadal Clanchet for their willingness to give their time so as to provide many valuable and constructive suggestions during the planning and development of this final degree project.

References

1. Eaton MD, Meiklejohn G, van Herick W. STUDIES ON THE ETIOLOGY OF PRIMARY ATYPICAL PNEUMONIA : A FILTERABLE AGENT TRANSMISSIBLE TO COTTON RATS, HAMSTERS, AND CHICK EMBRYOS. J Exp Med [Internet]. 1944 Jun 1 [cited 2015 May 19];79(6):649–68. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2135382&tool=pmcentrez&rendertype=abstract>
2. CHANOCK RM. Mycoplasma pneumoniae: proposed nomenclature for atypical pneumonia organism (Eaton agent). Science [Internet]. 1963 May 10 [cited 2015 May 30];140(3567):662. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14020096>
3. Waites KB, Talkington DF. Mycoplasma pneumoniae and its role as a human pathogen. Clin Microbiol Rev [Internet]. 2004 Oct [cited 2015 Feb 9];17(4):697–728, table of contents. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=523564&tool=pmcentrez&rendertype=abstract>
4. Maniloff J, McElhaney RN, Finch LR, Baseman JB (Editors). Mycoplasmas: molecular biology and pathogenesis. American Society for Microbiology; 1992 [cited 2015 May 19]; Available from: <http://www.cabdirect.org/abstracts/19932279979.html>
5. Van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, et al. Reductive genome evolution in Buchnera aphidicola. Proc Natl Acad Sci U S A [Internet]. 2003 Jan 21 [cited 2015 May 30];100(2):581–6. Available from: <http://www.pnas.org/content/100/2/581.full>
6. Razin S, Yogev D, Naot Y. Molecular biology and pathogenicity of mycoplasmas. Microbiol Mol Biol Rev [Internet]. 1998 Dec [cited 2015 Apr 6];62(4):1094–156. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=98941&tool=pmcentrez&rendertype=abstract>
7. Atkinson TP, Balish MF, Waites KB. Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections. FEMS Microbiol Rev [Internet]. The Oxford University Press; 2008 Nov 1 [cited 2015 Apr 12];32(6):956–73. Available from: <http://femsre.oxfordjournals.org/content/32/6/956.abstract>
8. Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen W-H, et al. Impact of genome reduction on bacterial metabolism and its regulation. Science [Internet]. 2009 Nov 27 [cited 2015 Apr 21];326(5957):1263–8. Available from: <http://www.sciencemag.org.sare.upf.edu/content/326/5957/1263.full>
9. Xavier JC, Patil KR, Rocha I. Systems biology perspectives on minimal and simpler cells. Microbiol Mol Biol Rev [Internet]. 2014 Sep 1 [cited 2015 May 21];78(3):487–509. Available from: <http://mmbr.asm.org.sare.upf.edu/content/78/3/487.full>

10. Henry C, Overbeek R, Stevens RL. Building the blueprint of life. *Biotechnol J* [Internet]. 2010 Jul [cited 2015 Jun 2];5(7):695–704. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20665643>
11. Tanaka K, Henry CS, Zinner JF, Jolivet E, Cohoon MP, Xia F, et al. Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res* [Internet]. 2013 Jan 7 [cited 2015 Jun 3];41(1):687–99. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3592452&tool=pmcentrez&rendertype=abstract>
12. Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, et al. Proteome organization in a genome-reduced bacterium. *Science* [Internet]. 2009 Nov 27 [cited 2015 Apr 8];326(5957):1235–40. Available from: <http://www.sciencemag.org/content/326/5957/1235>
13. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang R-Y, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* [Internet]. 2010 Jul 2 [cited 2014 Jul 11];329(5987):52–6. Available from: <http://www.sciencemag.org/content/329/5987/52.abstract>
14. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* [Internet]. 2008 Feb 29 [cited 2014 Jul 15];319(5867):1215–20. Available from: <http://www.sciencemag.org/content/319/5867/1215.abstract>
15. Juhas M. On the road to synthetic life: the minimal cell and genome-scale engineering. *Crit Rev Biotechnol* [Internet]. Informa Healthcare USA, Inc. New York; 2015 Jan 12 [cited 2015 Apr 24];1–8. Available from: <http://informahealthcare.com/doi/abs/10.3109/07388551.2014.989423>
16. Glass JI. Synthetic genomics and the construction of a synthetic bacterial cell. *Perspect Biol Med* [Internet]. 2012 Jan [cited 2015 Jun 3];55(4):473–89. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23502559>
17. Vickers CE, Blank LM, Krömer JO. Grand challenge commentary: Chassis cells for industrial biochemical production. *Nat Chem Biol* [Internet]. Nature Publishing Group; 2010 Dec 15 [cited 2015 May 25];6(12):875–7. Available from: <http://www.nature.com/sare.upf.edu/nchembio/journal/v6/n12/full/nchembio.484.html>
18. Foley PL, Shuler ML. Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnol Bioeng* [Internet]. 2010 Jan 1 [cited 2015 May 28];105(1):26–36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19816966>
19. Juhas M, Eberl L, Glass JI. Essence of life: essential genes of minimal genomes. *Trends Cell Biol* [Internet]. 2011 Oct [cited 2015 Jun 3];21(10):562–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21889892>
20. Juhas M, Eberl L, Church GM. Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol* [Internet]. 2012 Nov [cited

- 2015 May 12];30(11):601–7. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/22951051>
21. Fellermann H, Rasmussen S, Ziöck H-J, Solé R V. Life cycle of a minimal protocell--a dissipative particle dynamics study. *Artif Life* [Internet]. 2007 Jan [cited 2015 Jun 3];13(4):319–45. Available from:
<http://www.ncbi.nlm.nih.gov/pubmed/17716015>
22. Porcar M, Danchin A, de Lorenzo V, Dos Santos VA, Krasnogor N, Rasmussen S, et al. The ten grand challenges of synthetic life. *Syst Synth Biol* [Internet]. 2011 Jun [cited 2015 May 25];5(1-2):1–9. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3159694&tool=pmcentrez&rendertype=abstract>
23. Staunstrup J, Wolf W, editors. *Hardware/Software Co-Design: Principles and Practice* [Internet]. Boston, MA: Springer US; 1997 [cited 2015 Jun 3]. Available from: <http://link.springer.com/10.1007/978-1-4757-2649-7>
24. Karr JR, Sanghvi JC, Macklin DN, Gutschow M V, Jacobs JM, Bolival B, et al. A whole-cell computational model predicts phenotype from genotype. *Cell* [Internet]. Elsevier; 2012 Jul 20 [cited 2014 Jul 9];150(2):389–401. Available from: <http://www.cell.com/article/S0092867412007763/fulltext>
25. Pilo P, Vilei EM, Peterhans E, Bonvin-Klotz L, Stoffel MH, Dobbelaere D, et al. A metabolic enzyme as a primary virulence factor of *Mycoplasma mycoides* subsp. *mycoides* small colony. *J Bacteriol* [Internet]. 2005 Oct [cited 2015 May 31];187(19):6824–31. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1251598&tool=pmcentrez&rendertype=abstract>
26. Medina JL, Coalson JJ, Brooks EG, Winter VT, Chaparro A, Principe MFR, et al. *Mycoplasma pneumoniae* CARDS toxin induces pulmonary eosinophilic and lymphocytic inflammation. *Am J Respir Cell Mol Biol* [Internet]. 2012 Jul [cited 2015 May 31];46(6):815–22. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3380286&tool=pmcentrez&rendertype=abstract>
27. Krishnan M, Kannan TR, Baseman JB. *Mycoplasma pneumoniae* CARDS toxin is internalized via clathrin-mediated endocytosis. *PLoS One* [Internet]. 2013 Jan 7 [cited 2015 May 31];8(5):e62706. Available from:
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0062706>
28. Kannan TR, Musatovova O, Balasubramanian S, Cagle M, Jordan JL, Krunkosky TM, et al. *Mycoplasma pneumoniae* Community Acquired Respiratory Distress Syndrome toxin expression reveals growth phase and infection-dependent regulation. *Mol Microbiol* [Internet]. 2010 Jun 1 [cited 2015 Jun 7];76(5):1127–41. Available from:
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2883071&tool=pmcentrez&rendertype=abstract>
29. Waites KB, Talkington DF. *Mycoplasma pneumoniae* and its role as a human pathogen. *Clin Microbiol Rev* [Internet]. 2004 Oct 1 [cited 2015 Feb 9];17(4):697–728, table of contents. Available from:
<http://cmr.asm.org/content/17/4/697.short>

30. Kenri T, Okazaki N, Yamazaki T, Narita M, Izumikawa K, Matsuoka M, et al. Genotyping analysis of *Mycoplasma pneumoniae* clinical strains in Japan between 1995 and 2005: type shift phenomenon of *M. pneumoniae* clinical strains. *J Med Microbiol* [Internet]. 2008 Apr [cited 2015 May 21];57(Pt 4):469–75. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18349367>
31. Atkinson TP, Dai Y, Duffy LB. Role of Hydrogen Peroxide in *Mycoplasma pneumoniae*-induced Mast Cell IL-4 Production. *J Allergy Clin Immunol* [Internet]. Elsevier; 2007 Jan 1 [cited 2015 May 21];119(1):S54. Available from: <http://www.jacionline.org/article/S0091674906025802/fulltext>
32. Koonin E V. How many genes can make a cell: the minimal-gene-set concept. *Annu Rev Genomics Hum Genet* [Internet]. 2000 Jan [cited 2015 Apr 27];1:99–116. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11701626>
33. Fisunov GY, Alexeev DG, Bazaleev NA, Ladygina VG, Galyamina MA, Kondratov IG, et al. Core proteome of the minimal cell: comparative proteomics of three mollicute species. *PLoS One* [Internet]. 2011 Jan [cited 2015 May 26];6(7):e21964. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3139596&tool=pmcentrez&rendertype=abstract>
34. Güell M, van Noort V, Yus E, Chen W-H, Leigh-Bell J, Michalodimitrakakis K, et al. Transcriptome complexity in a genome-reduced bacterium. *Science* [Internet]. 2009 Nov 27 [cited 2015 May 16];326(5957):1268–71. Available from: <http://www.sciencemag.org.sare.upf.edu/content/326/5957/1268.long>
35. Van Noort V, Seebacher J, Bader S, Mohammed S, Vonkova I, Betts MJ, et al. Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol* [Internet]. 2012 Jan [cited 2015 Jun 9];8:571. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3293634&tool=pmcentrez&rendertype=abstract>
36. Maier T, Schmidt A, Güell M, Kühner S, Gavin A-C, Aebersold R, et al. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* [Internet]. 2011 Jan [cited 2015 Jun 9];7:511. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3159969&tool=pmcentrez&rendertype=abstract>
37. Wodke JAH, Alibés A, Cozzuto L, Hermoso A, Yus E, Lluch-Senar M, et al. MyMpn: a database for the systems biology model organism *Mycoplasma pneumoniae*. *Nucleic Acids Res* [Internet]. 2015 Jan 28 [cited 2015 Jun 9];43(Database issue):D618–23. Available from: <http://nar.oxfordjournals.org/content/43/D1/D618>
38. Wodke J a H, Puchałka J, Lluch-Senar M, Marcos J, Yus E, Godinho M, et al. Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol Syst Biol* [Internet]. 2013 Jan [cited 2014 Jun 12];9(653):653. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3658275&tool=pmcentrez&rendertype=abstract>

39. Lluch-senar M, Delgado J, Chen W, Lloréns-rico V, Reilly FJO, Wodke JAH, et al. Defining a minimal cell : essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. 2015;1–7.
40. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform [Internet]. 2013 Mar 1 [cited 2014 Jul 9];14(2):178–92. Available from: <http://bib.oxfordjournals.org/content/14/2/178.full?keytype=ref&%2520ijkey=qTgjFwbRBAzRZWC>
41. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Jan [cited 2014 Nov 19];29(1):24–6. Available from: <http://dx.doi.org/10.1038/nbt.1754>
42. Oshima K, Maejima K, Namba S. Genomic and evolutionary aspects of phytoplasmas. Front Microbiol [Internet]. Frontiers; 2013 Jan 14 [cited 2015 May 19];4:230. Available from: <http://journal.frontiersin.org/article/10.3389/fmicb.2013.00230/abstract>