Engineering the genome of minimal bacteria using CRISPR/Cas9 tools

Iason Tsarmpopoulos

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE

Science de la vie et de la Santé

SPÉCIALITÉ

Microbiologie and Immunologie

Par Iason TSARMPOPOULOS

Ingénierie de génome de bactéries minimales par des outils CRISPR/Cas9

Sous la direction de : Monsieur Pascal SIRAND-PUGNET

Soutenue le ***jeudi 07 décembre 2017*** à 14h00
Lieu : INRA, 71 avenue Edouard Bourlaux 33882 Villenave d'Ornon
salle Amphithéâtre Josy et Colette Bové

Membres du jury :

Mme Cécile BEBEAR	Université de Bordeaux et CHU de Bordeaux	Président
Mme Florence TARDY	Anses-Laboratoire de Lyon	Rapporteur
M. Matthieu JULES	Institut Micalis, INRA and AgroParisTech	Rapporteur
M. David BIKARD	Institut Pasteur	Examinateur
M. Fabien DARFEUILLE	INSERM U1212 - CNRS UMR 5320	Invité
Mme Carole LARTIGUE-PRAT	INRA - Université de Bordeaux	Invité
M. Pascal SIRAND-PUGNET	Université de Bordeaux	Directeur de thèse

Ingénierie de génome de bactéries minimales par des outils CRISPR/Cas9

Les mycoplasmes sont des bactéries pathogènes, dotées de petits génomes d'environ 1Mbp, avec une faible teneur en G+C. L'intérêt de la communauté scientifique pour ces bactéries a été récemment renouvelé par des avancées dans les domaines de la synthèse et de la transplantation de génomes. Ces nouvelles approches ont ouvert la voie à l'ingénierie génomique à grande échelle des mycoplasmes. Les systèmes CRISPR/Cas sont des systèmes de défense adaptatifs procaryotes contre les acides nucléiques invasifs. Le système CRISPR de *Streptococcus pyogenes* est composé d'une endonucléase (SpCas9) et de deux CRISPR ARNs (crRNA et tracrRNA) qui dirigent Cas9 vers sa séquence d'ADN cible. La reconnaissance de l'ADN cible se fait par appariement du crRNA et de la présence en aval d'une séquence nommée protospacer adjacent motif (PAM). Apres cette reconnaissance, Cas9 coupe l'ADN cible. A partir de ce système, un outil génétique simplifié composé de Cas9 et d'un ARN guide (gRNA) a été développé pour de nombreux organismes. Le premier objectif de ma thèse était de combiner les méthodes de biologie synthétique de clonage et de la transplantation de génomes avec les outils CRISPR/Cas9 pour l'ingénierie des génomes de mycoplasmes clonés dans la levure. Nous avons réussi à utiliser cette approche pour enlever des gènes et des régions génomiques dans trois espèces: *Mycoplasma mycoides* subsp*. capri* (Mmc), *M. capricolum* subsp. *capricolum* et *M. pneumoniae*. Afin de développer un système plus adapté aux mycoplasmes, nous avons ensuite caractérisé le système CRISPR/Cas9 de *Mycoplasma gallisepticum* (Mg). En utilisant une combinaison d'approches in silico et in vivo, la séquence PAM de MgCas9 a été caractérisée comme NNNAAAA. Nous avons alors entrepris de développer un système CRISPR/Cas minimal de *M. gallisepticum* pour une utilisation directe dans les cellules de mollicutes: le gène codant MgCas9 a été introduit dans le génome de Mmc, mais son activation avec un gRNA chimère entre le crRNA et le tracrRNA de *M. gallisepticum* n'a pas été obtenue pour le moment.

Mots clés : Biologie de synthèse, CRISPR/Cas9, Mycoplasma

Engineering the genome of minimal bacteria using CRISPR/Cas9 tools

Mycoplasmas are small pathogenic bacteria that are characterized by reduced genomes of about 1 Mbp with a low G+C content. The interest of the scientific community towards these species has been recently renewed by successful synthesis of their genome and transplantation experiments. These new genetic tools opened the way to further applications and developments for large-scale genome engineering programmes. CRISPR/Cas systems are natural systems that provide bacteria and archaea with an adaptive defense mechanism against invading nucleic acids. The CRISPR system from *Streptococcus pyogenes* includes an endonuclease (SpCas9) and two CRISPR RNAs (crRNA et tracrRNA) which role are to drive Cas9 to a target sequence. Target recognition depends on a specific pairing of the crRNA and the presence of a motif named protospacer adjacent motif (PAM). After recognition, Cas9 cleaves the targeted DNA. From the natural *S. pyogenes* system, a simplified genetic tool including Cas9 and a guide RNA (gRNA) was developed for many organisms . The first goal of my thesis was to combine the synthetic biology methods of genome cloning in yeast and back transplantation into recipient cells with a CRISPR/Cas9 tool for efficient engineering of mycoplasma genomes cloned in yeast. We succeeded in removing genes and genomic regions in three different species, *Mycoplasma mycoides* subsp*. capri* (Mmc)*, M. capricolum* subsp*. capricolum* and *M. pneumoniae*. Then, in order to develop a system optimized for mycoplasma genome editing, we characterized a natural CRISPR/Cas9 system derived from *Mycoplasma gallisepticum* (Mg). Using a combination of in silico and in vivo approaches, MgCas9 PAM sequence was characterized as NNNAAAA. We then started to develop a minimal CRISPR/Cas system from *M. gallisepticum* for direct genome editing in mollicutes. Thus we introduced MgCas9 encoding gene in Mmc and tried to activate it with a newly designed gRNA, a chimeric molecule between the crRNA and the tracrRNA of *M. gallisepticum*, without success yet.

Keywords : Synthetic Biology, CRISPR/Cas9, Mycoplasma

UMR 1332 Biologie du Fruit et Pathologie
INRA, IBVM, 71 Av. E. Bourlaux, CS 20032
33882 Villenave d'Ornon Cedex, France

Table of contents

Results

Abbreviation list

A, C, G, T, U = Adenine, Cytosine, Guanine, Thymine, Uracil

Cas proteins = CRISPR associated proteins

CDS = Coding DNA Sequence

CRISPR = Clustered Regularly Interspaced Short Palindromic Repeats

crRNA = Crispr RNA

DNA = Deoxyribo-Nucleic Acid

DR = Direct Repeat

DSB(s) = Double strand Break(s)

glpO = glycerol-3-phosphate oxidase

gRNA = guide RNA

HCO = Hexaammine-CObalt

HDR = Homologous Directed Repair

HGT = Horizontal Gene Transfer

ICE = Integrative and Conjugative Element

Mcap = *Mycoplasma capricolum* subsp. *capricolum*

MgCas9 = *Mycoplasma gallisepticum* Cas9 protein

Mmc = *Mycoplasma mycoides* subsp. *capri*

Mmm = *Mycoplasma mycoides* subsp. *mycoides*

PAM = Protospacer adjacent motif

RNA = Ribo-Nucleic Acid

tracrRNA = trans-activating crRNA

TREC = Tandem Repeat coupled with Endonuclease Cleavage

TREC-IN = TREC-assisted gene knock-IN

Introduction

Introduction

I. Mollicutes

Mollicutes is a class of bacteria without a cell wall that are relatively small (between 0.3 and 0.8 µm in diameter, while *E. coli* are 2 µm long), and characterized by a reduced genome with sizes ranging between 580 kbp for *Mycoplasma genitalium* and 2,200 kbp for *Spiroplasma ixodetis*. The percentage of G+C (Guanine and Cytosine) is also lower than in other bacteria, with an average for the class of 27.3% and a range of 23.7% for *Mycoplasma capricolum* subsp. *capricolum* to 40% for *Mycoplasma pneumoniae*. This general low G+C content is more pronounced in the non-coding regions of the genome, sometimes reaching 10-20%. In the coding regions, a codon bias can be also observed, with a preference for adenine or thymine at the third base of many codons (Razin 1998). Finally, mollicutes use UGA as a tryptophan codon and not as a stop codon, except for the related genera *Acholeplasma* and *Candidatus* phytoplasma (Blanchard, 1990).

Genome analysis of mollicutes has revealed an absence of many genes and metabolic pathways that are found in most bacteria. One main difference is the lack of genes involved in the biosynthesis of the peptidoglycan cell wall, thus their name: molli= soft, cutis= skin (in Latin). This lack of cell wall makes these bacteria impossible to characterize by gram-based staining and explains the various cell shapes observed for many of these species. For the mycoplasma cells, the most common shape is a sphere but others can be observed, such as pear-like shapes (Razin 1978 and 1998). Comparative studies of mollicute metabolism indicate that they have not developed extensive biosynthetic pathways, but instead are dependent on the acquisition of all the substances necessary for energy generation from the extracellular environment (Razin 1998). This means that mollicutes need an outside source of sterols, fatty acids, amino acids, vitamins and the precursors of nucleic acids, all of which they are incapable of synthesizing. This is consistent with these bacteria having a parasitic way of life and their culture requires rich and complex media. There is no simple relationship between genome size and cultivability of mollicutes and some mollicutes, like phytoplasmas and hemoplasmas, are still un-cultivable.

Despite their small genome and their relatively simple metabolism, mollicutes are characterized as minimal self-replicating organisms. One representative species is *M. genitalium* with a genome of 580 kbp that encodes only 482 proteins. This is the main reason why mycoplasmas have attracted the interest of the scientific community for novel applications such as the development of synthetic genomes at the J. Craig. Venter Institute (*M. genitalium*, *M. mycoides* subsp. *capri* and reduced versions Syn1.0 to Syn3.0) by Gibson and colleagues (Hutchison et al., 1999, Gibson et al., 2010, Hutchison et al., 2016) and the transplantation of Mycoplasma genome from one species to another (Lartigue 2007). Systems biology approaches including transcriptomics, proteomics, metabolomics and *in silico* whole-cell modeling approaches have also been achieved in some species like *Mycoplasma pneumoniae* (Kühner et al., 2009, Yus et al., 2009, Güell et al., 2009, Maier et al., 2013) in order to identify the way these minimal organisms function, with the perspective to decipher some of the secrets of life. The combined approach of systems biology and synthetic biology now opens possibilities to engineer and redesign the genome of these “minimal cells”. All these results and all the ongoing work will allow a variety of applications, from

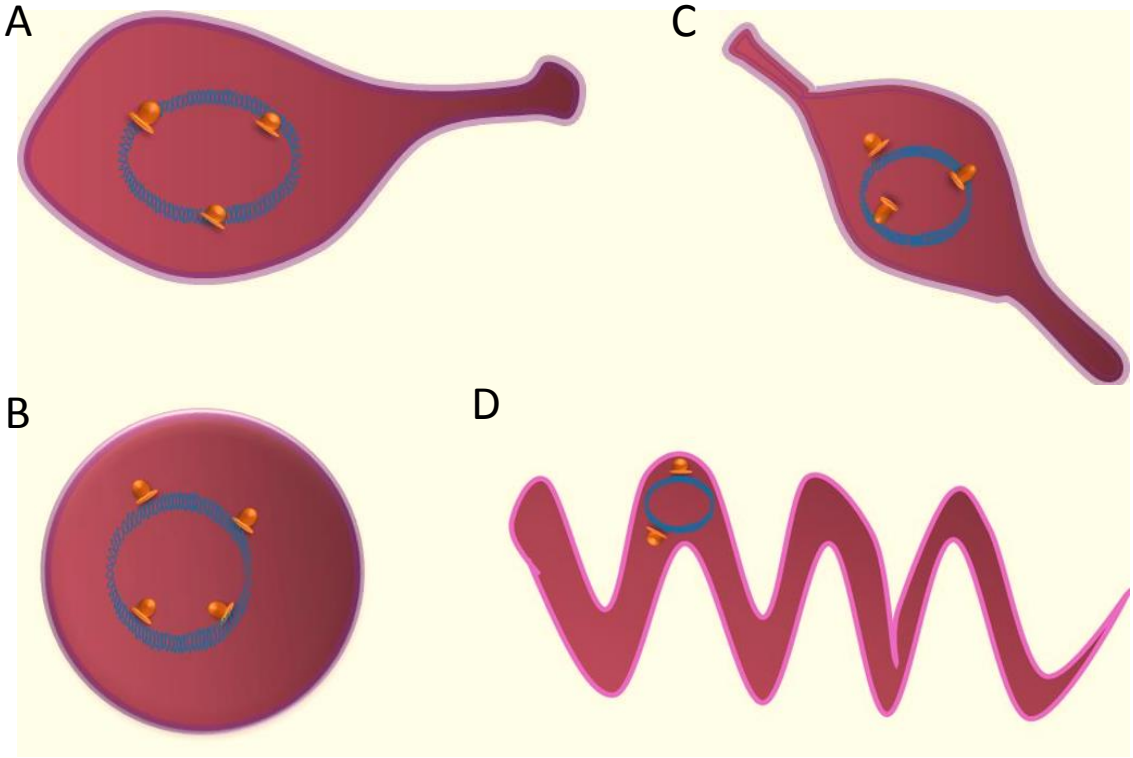


Figure 1. Morphology of Mollicutes. Mollicutes cells have various shapes including pear (A), sphere (B), elongated shape (C) of *M. pneumoniae* and helicoid of *S. citri* (D). In orange are the ribosomes. In purple the lipoprotein membrane

production of molecules of interest in an organism with entirely controlled capabilities but also developing attenuated strains of pathogenic bacteria to be used as vaccines.

II. Taxonomy and Phylogeny

Mollicutes are part of the phylum Tenericutes that contains bacteria without cell wall. They are divided in two main branches: The Spiroplasmataceae-Entomoplasmataceae-Mycoplasmatacea (SEM) branch that contains the genera *Mycoplasma, Ureaplasma, Entomoplasma, Mesoplasma* and *Spiroplasma* and the Acholeplasmataceae-Aneroplasma-Phytoplasma (AAP) branch that contains the genera *Acholeplasma, Candidatus phytoplasma, Anaeroplasma* and *Haloplasmatales.*

a. Taxonomy

The *Mollicutes* class includes 4 distinct orders, divided in 5 families and 8 genera (Razin et al., 2002). The order of *Mycoplasmatales* consists in one family, the *Mycoplasmataceae* which contain two genera: *Mycoplasma* and *Ureaplasma*. The organisms of this family are mainly aerobic and their growth requires cholesterol. The species of the *Ureaplasma* genus have the capacity to hydrolyze urea as a carbon and energy source, in contrast with mycoplasmas that use sugars or, for some species, the amino acid arginine for energy metabolism. Mycoplasmas and ureaplasmas can infect a large variety of animal hosts including reptiles, birds, fishes and many mammals, including humans. Recently some hemotropic mycoplasmas have been characterized, the Hemoplasma and the Hepatoplasma. Hemoplasmas are uncultivable, which makes it difficult to properly classify this species. Classification using the *gapA* and *dnaK* as markers for phylogenetic analysis, instead of the 16S rRNA, provided clear evidence of their classification within the *Mycoplasma* genus (Hicks et al., 2014). The Hepatoplasma is also a member of the *Mycoplasma* genus as shown by Leclercq and his colleagues (Leclercq et al., 2014). In this case, an analysis was conducted on 127 orthologous genes conserved among mollicutes to allow a proper classification.

The *Entomoplasmatales* are mollicutes that have been isolated from arthropods and from the surface of plants. This order includes two families: the *Entomoplasmataceae* family that contains two genera: *Mesoplasma* and *Entomoplasma* and the *Spiroplasmataceae,* containing the single genus *Spiroplasma.* Spiroplasmas typically infect invertebrates. Three species are considered phytopathogenic (*Spiroplasma citri, Spiroplasma kunkelii, Spiroplasma phoeniceum*) and have the capacity to replicate both in insect vectors and in the phloem of their host plants. Spiroplasmas in their majority are characterized by a helical morphology.

The next order is the *Acholeplasmatales,* and among them, the *Acholeplasmataceae* are the only members that can be cultivated. The only genus of *Acholeplasmataceae* is the *Acholeoplasma* and their growth doesn't require cholesterol, like the *Mycoplasmatales.* They infect animal and plant species and some species like *Acholeplasma laidlawii* can be found free in the environment, a remarkable feature considering the simple metabolism that characterizes all mollicutes.

As mentioned before, many species of mollicutes are not yet cultured in non-cellular medium. As a result, a proper classification in the above families cannot be strictly conducted. A particular example is the case of the *Candidatus phytoplasmas* that can multiply in the insect vectors and in the phloem tubes

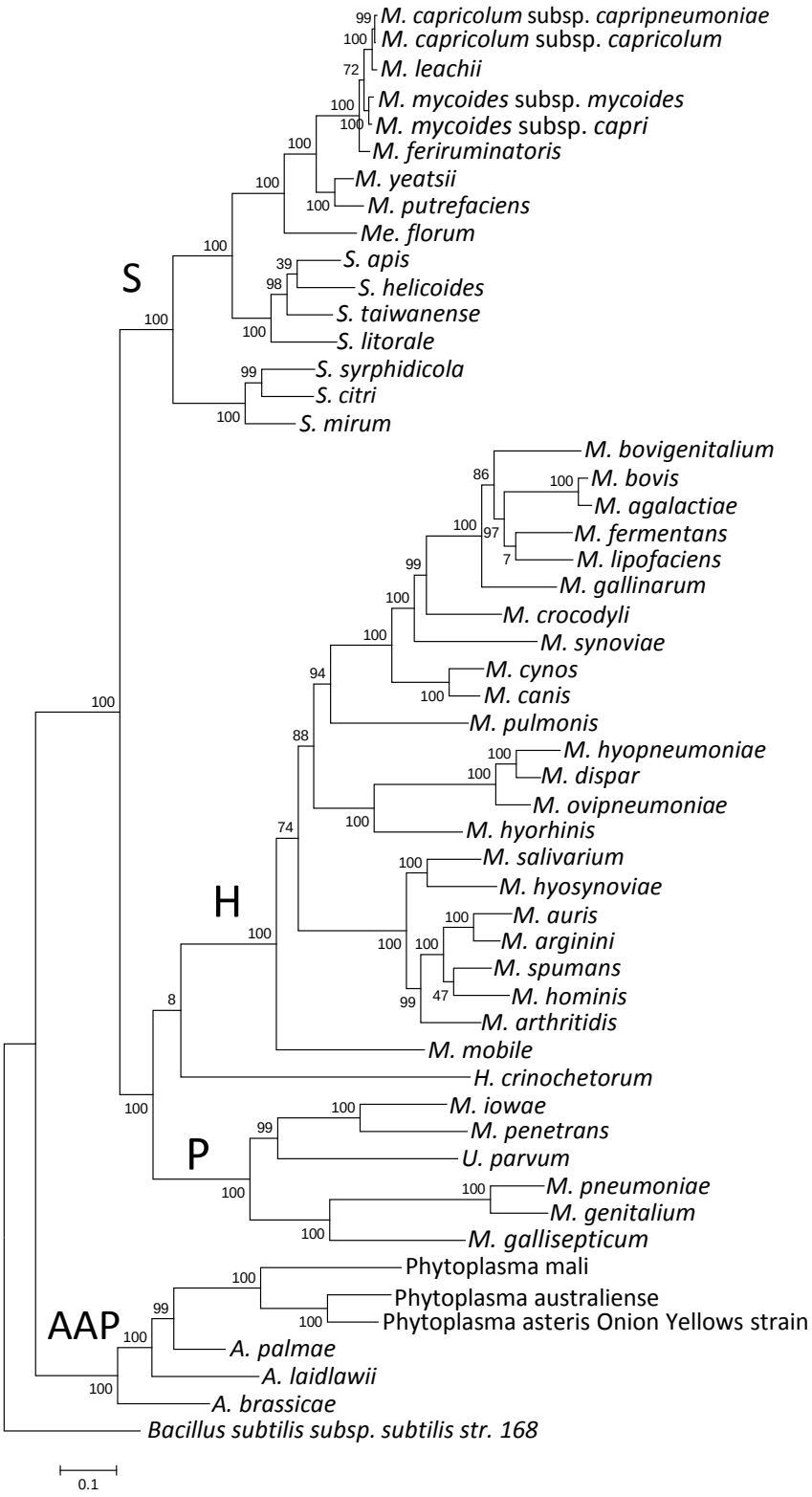


Figure 2. Phylogeny of mollicutes. The phylogenetic tree was generated using the maximum likelihood method from the concatenated multiple sequence alignments of selected 50 orthologous protein involved in translation. Main phylogenetic groups are indicated, S, Spiroplasma, H, Hominis, P, Pneumoniae, AAP, Acholeplasma/Phytoplasma. *B. subtilis* was used as an outgroup. Statistical values from an Approximate Likelihood-Ratio Test are indicated on branches.

of plants. However, some molecular data provided over the years, demonstrated a common origin with the acholeplasmas. As such, they have been classified in the order *Acholeplasmatales*.

Some anaerobic mollicutes isolated from ruminants are part of the order *Anaeroplasmatales*. They are grouped in a single family, the *Anaeroplasmataceae*, which contains two genera: *Anaeroplasma* that requires an extracellular provision of cholesterol and *Asteroleplasma,* which does not.

b. Phylogeny

Phylogenetic studies based on the DNA sequence of the gene encoding the 16S rRNA of the mollicutes have shown that they have evolved from a Gram positive bacterium with low G+C content (Woese 1987). This ancestor is also common with some species of the genus *Clostridia* with which they share a resistance to the antibiotic rifampicin (Gadeau 1986). The phylogenetic tree of mollicutes includes two major branches (Figure 2). The AAP branch includes acholeplasmas and phytoplasmas. These two genera have preserved the "universal" genetic code, without the UGA switch from stop codon to tryptophan. The SEM branch includes species from the *Mycoplasma, Spiroplasma*, *Mesoplasma, Ureaplasma* and *Entomoplasma* genera. This branch is divided into three sub-branches that correspond to the phylogenetic groups Spiroplasmas, Hominis and Pneumoniae. The "mycoides" species cluster is included in the Spiroplasma group, with 5 species and sub-species all pathogens of ruminants: *Mycoplasma mycoides* subsp. *mycoides, Mycoplasma mycoides* subsp. *capri, Mycoplasma capricolum* subsp. *capricolum, Mycoplasma capricolum* subsp. *capripneumoniae and Mycoplasma leachii* (Manso-Silvàn et al., 2007, Manso-Silvàn et al., 2009).

III. Evolution and structure of mollicutes genome

i. Evolution with genome reduction

As mentioned above, one main characteristic of mollicutes is their small genomes. The average size is 1 Mbp, 4 times smaller than the genome of *Bacillus subtilis.* The small genome size combined with a gene number around 1,000 has been interpreted as having resulted from a massive loss of genes (Woese 1984) during evolution. This phenomenon affects all the different gene categories, even the core cell machinery implicated in the expression and the transmission of genetic information (Sirand-Pugnet 2007, Grosjean et al., 2014). An example that demonstrates this phenomenon is the fact that 60% of *M. pneumoniae* genome is essential for its survival whereas for *E. coli* only 15% is essential (Yus et al., 2009). A pioneer work of comparative genomics on *M. genitalium* proposed a core of 250 essential genes (Mushegian et al., 1996). More recently, the work on the synthetic cell derived from *Mycoplasma mycoides* subsp. *capri* (Mmc) called Syn3.0 (Hutchison et al., 2016) demonstrated that this genome could be reduced from a size of 1,078,809 base pairs and a gene pool of 901 genes, to a minimal genome that consists of 531,000 bp and a pool of 438 genes, with only a minimal impact on growth rate. This could be considered as a working approximation of a true "minimal cell". As it has been already stated, mollicutes have lost the genes responsible for the synthesis of the cell wall, fatty acid and nucleic acid precursors. They have also lost genes that are usually found in multiple copies in other bacteria, leading to a reduced redundancy for many enzymatic functions. This marked tendency toward genome reduction has been also observed for vital functions of the cell related to expression and maintenance of genetic information: reduction of the tRNA repertoire, and a simplified system of genome repair and

recombination (Rocha et al., 2005), and loss of enzymes responsible for the modification of the rRNA and ribosomal proteins (de Crécy-Lagard 2007, Grosjean et al., 2014). Mollicutes also have a simplified system of rRNAs expression, with only one or two operons responsible for the expression of all ribosomal RNA, in contrast with *B. subtilis and E. coli K12* that have 10 and 7 operons, respectively (Yus et al. 2009, Kunst et al. 1997, Blattner et al. 1997). Only 8 transcription factors have been described in *M. pneumoniae* (Yus et al. 2009). All these findings suggest that the genome of mollicutes has been streamlined by evolution leading to current parasitic minimal bacteria. However, this tendency to reduction is not the only force driving the evolution of mollicute genomes that appear to be extremely dynamic, both in terms of chromosome organization and gene content.

ii. Synteny loss

The study of the organization of the mollicutes chromosome revealed some variability in the characteristics of the origin and termination site of replication. These areas of the chromosome have already been studied in many bacteria, and there has been observed an inversion of the G+C/C-G ratio (GC skew) at these sites and an alternation in gene orientation (Rocha et al., 2008). In mollicutes, these features have been observed in several genomes, for example for *A. laidlawii, Mycoplasma gallisepticum* or *Mesoplasma florum.* But for others species, these structural characteristics have been lost, such is the case for *Mmm, Mycoplasma mobile* or *Ca. phytoplasma asteris Onion Yellows* strain (Sirand-Pugnet et al., 2007).

The global dynamic of the mollicute genome is impressive when studying the conservation of the synteny *i.e*. the order of the genes on the chromosome. Even though a relative synteny has been observed for a number of species that are closely-related (Thiaucourt et al., 2012), it is completely abolished when studying more distantly-related species (Dandekar et al., 2002). This suggests an intense evolutionary force that pushes the mollicutes to evolve and adapt, in response to the biological constraints imposed by their parasitic lifestyle.

iii. Impact of horizontal gene transfer

Horizontal gene transfer (HGT) is a natural phenomenon during which an organism integrates foreign DNA from another organism, in its cell and possibly its genome. HGT can occur in eukaryotes, in prokaryotes but also between prokaryotes and eukaryotes (Hotopp et al., 2011). It was first discovered during the 60s, during the study of bacterial “spontaneous” resistance to antibiotics (Barlow 2009). In bacteria, three different mechanisms that allow this exchange of genetic information have been described: i) bacterial conjugation that permits an exchange of DNA (plasmid or chromosome fragments) through a cytoplasmic bridge that links together two bacteria; ii) genetic transformation which corresponds at an active integration of DNA from the environment of a bacteria; iii) and transduction: a process that consists of a transfer of genetic material from one bacteria to another by a viral vector, a prophage. While these mechanisms have been known for years, their wide impact in genome evolution of nearly all organisms have only been discovered with the development of basic genomics and large scale genome sequencing, starting in the 90’s and reaching a new dimension with Next Generation Sequencing technologies during the past decade. Now, genome comparisons have shown that HGT are

more frequent than first thought and they are considered as an important wheel that allows the evolution to move forward, especially for the prokaryotes (Koonin 2009, Koonin 2016).

iv. HGT in mollicutes

The evolution of mollicutes was always considered a "regressive" evolution, where the only stimulating events were the ones that provoked loss of genes (Weisburg et al., 1989, Woese 1980).The study of the first complete genomes strengthen this vision because of the small genome of the majority of species and the lack of multiple genes found in most bacteria. The first global genome studies conducted in mollicutes, concerning the HGTs in this class, led to the same conclusion, that the evolution of mollicutes was little affected by the HGTs (Nakamura et al. 2004) and their genetic isolation could be due to relatively inefficient recombination machinery and, for the species of the SEM branch, a different genetic code. This model has only been recently challenged by a growing number of examples of genetic exchanges between mollicutes sharing the same host.

1. Between human pathogens

When the scientists compared the first genome sequences of human mycoplasma (*M. genitalium, M. pneumoniae, M. penetrans* and *U. parvum*) they were unable to identify any exchange of genetic material between them (Razin 1998, Razin 2002). However, when the complete sequence of the *M. hominis* genome was available, some HGTs events were identified between this species and other bacteria (Pereyre et al. 2009). Five of these transfers were from species outside the mollicutes class, but still, they were human pathogens. From the rest, 5 groups of genes have been potentially exchanged with the mollicute *U. parvum*. Both *M. hominis* and *U. parvum* are pathogens of the same ecological environment (human urogenital tract) but are classified in different phylogenetical groups. *M. hominis* belongs to the Hominis group and *U. parvum* belongs to the Pneumoniae group. Among the genes transferred between these two species, a gene cluster has been identified that codes for a type III restriction-modification system, two type I restriction-modification systems, one transposase IS1138 pseudogene and a cluster of 9 genes, 7 of them encoding a mycoplasma-specific F1-like X0 ATPase of unknown function (Béven et al., 2012) and two genes encoding a mycoplasma specific MIB-MIP system involved in binding and cleaving host IgG in Mmc (Arfi et al., 2016).

2. Between bird pathogens

After sequencing the whole genome of *M. synoviae*, Vasconcelos and his colleagues (Vasconelos et al., 2005) searched for HGTs events in this pathogen. They conducted a research in each CDS of *M. synoviae* studying the origin of the best hits. The most interesting results showed an origin of some genes from *M. gallisepticum,* which is also a pathogen of birds, specifically of poultry. Those pathogens belong to different phylogenetic groups, *M. synoviae* belongs to the Hominis group and *M. gallisepticum* belongs to the Pneumoniae group. The transferred genes are grouped in 14 distinct regions, the larger one being a region of 5.9 kbp. The majority of these genes encode hypothetical proteins which function is unknown. Among genes that have a predicted function are two transposases, an operon coding for an ABC transporter, a gene coding a single type I peptidase and two clusters of genes implicated in the pathogenicity of these two species. These clusters code for a large family of hemagglutinins, molecules

implicated in host pathogenicity, a glyceraldehyde-3-phosphate dehydrogenase, an elongation factor EF-G and a sialidase, an enzyme that has been identified as playing an important role in the pathogenicity of another pathogenic mycoplasma, *M. alligatoris* (Hunt et al., 2007). Among the genes exchanged between the two mycoplasmas, 6 may have a specific role in the adaptation to the host of these bacteria (Sirand-Pugnet et al., 2007). More recently, a study on the previously mentioned mycoplasma-specific F1-like X0 ATPase indicated that some of the relevant genes have also been exchanged (Béven et al., 2012).

3. Between ruminants pathogens

Concerning *M. agalactiae*, pathogen of small ruminants, genetic research showed that its genome contains conjugative and integrative elements (ICE) (Marenda et al., 2006) and insertion sequences (IS) similar to the ones already identified in the mycoides cluster (Thomas et al., 2005, Tardy et al., 2015). This data suggested that genetic exchanges have taken place between the pathogens of those two ruminants. This hypothesis has been verified after a complete sequence of the genome of two strains of *M. agalactiae*, PG2 (Sirand-Pugnet et al. 2007) and 5632 (Nouvel et al. 2010). Almost 18% of the genome of this mycoplasma was conjectured to have been exchanged with the mycoides cluster of mycoplasmas through HGT events. As it has been observed in the other two cases of HGT in mollicutes, these mycoplasmas belong to different phylogenetic groups: *M. agalactiae* belongs to the Hominis group and the mycoides cluster is a part of the Spiroplasma group. Among the 134 genes that have potentially been exchanged between the two species, the majority encode hypothetical proteins (50), transmembrane proteins (7) and lipoproteins (17) all of which are proteins specific to ruminant mycoplasmas. The rest of the genes code for transporters (18), various enzymes (19), pseudogenes (11) some factors specific to mobile elements (2). A number of these genes can potentially play a role in the pathogenicity and the infection of their ruminant hosts, including the previously mentioned MIB-MIP system (Arfi et al. 2016) and ATPase F1-likeX0 (Béven et al., 2012). More recent work on the genome sequences of *M. mycoides* subsp*. capri* (from the mycoides cluster) *and Mycoplasma bovis* (Thiaucourt et al. 2011, Li et al. 2011, respectively) have confirmed the presence of HGTs events between the mycoides and the *M. bovis/M. agalactiae* clusters of species. In addition, genome sequences of other ruminant mollicutes including *Mycoplasma bovigenitalium*, *Ureaplasma diversum*, *Mycoplasma alkalescens*, *Mycoplasma auris* and *Mycoplasma arginini* also suggested some HGT leading to a global picture where HGT have played an unexpectedly important role in shaping the genomes of current mycoplasmas infecting ruminants (Sirand-Pugnet, unpublished).

In conclusion, all these examples of HGT within mollicutes concern species which are pathogens of the same host. Furthermore, it is likely that the genes that have been exchanged play a key role in bacterial pathogenicity. Even though the evolution of these organisms demonstrates a global loss of genes, these species have kept their HGT capacity as a way to enrich their genetic potential. This capacity may have played a key role in the diversity of the hosts that are susceptible to infection by these minimal bacteria.

IV. Mollicutes mobile elements

Mobile elements are largely involved in the dynamics of bacterial genomes. Different types of mobile elements have been described in mollicutes: The most frequently found are Insertion sequences (IS), integrative and conjugative elements (ICE), and replicative plasmids and phages.

a. Insertion sequences

The IS are among the smallest and simplest mobile elements spread in all domains of life (Vandecraen et al. 2017). They can exist as a single or multiple copies in a genome and they can move inside the genome or even horizontally from one species to another through an HGT event, as mentioned above. They have a small size, usually less than 3 kb, and they code for elements essential for their mobility (Siguier 2014). They are typically flanked by short terminal inverted repeats (IR) and their transposition is controlled by an enzyme (a transposase) which binds to the IRs, cleaves the DNA and allows the transfer of the IS from one location to another. By integrating into a coding region, they can inactivate genes and as such, they have an impact on the virulence, resistance and metabolic activities of the cell (Vandecraen et al. 2017; Chandler and Mahillon 2002).

In the *Mollicutes* class, many species carry IS. In the mycoplasma genus, they have been identified in species like *M. agalactiae, M. bovis, members of the mycoides cluster, M. gallisepticum, M. fermentans* and many others (Pilo et al. 2003, Li et al. 2011, Calcutt et al. 1999). A more extensive study however, has been conducted on *Mycoplasma mycoides* subsp. *mycoides* (Mmm) strain PG1 revealed that 13% of its genome consists of insertion sequences (Westberg et al. 2004). These Mmm ISs are categorized into three groups: IsMmy1, which has a size of 1670 bp and is present in eight full length and one truncated copy. The other two IS elements are IS1634 (Vilei et al. 1999), measuring 1872 bp and IS1296 (Frey et al. 1995), which have a size of 1485 bp. There are 60 copies of IS1634 including two copies that are split by other ISs elements and one that is truncated. IS1296 is present in 28 copies, including four that are interrupted by other ISs elements and seven truncated copies. The highly dynamic aspect of IS is particularly obvious when comparing different strains of the same species, the IS profile being generally different from one genome to another. This particularity has been often used to differentiate strains (Vilei et al., 2000). The role of these ISs is not clear yet but the fact that they can move inside the genome of Mmm likely affects the stability, the genetic potential and the global structure of this genome. While Mmm is currently the mollicute species where the maximum number of IS have been described, most mollicutes genome include IS, making them the most commonly found mobile elements in mollicutes.

b. Integrative and conjugative elements

The integrative and conjugative elements (ICEs) are a diverse group of mobile genetic elements found in a wide range of bacteria (Guglielmini et al., 2011). Their size ranges from 20 kbp to even more than 500 kbp. They can be found integrated in the host chromosome and they usually contain genes required for their excision, conjugation and integration (Guglielmini et al., 2011, Johnson and Grossman 2015). These events give the ICEs the capacity to exit the chromosome, get horizontally transferred in others cells and re-integrate in a “cut and paste” process. The special characteristic of this process is that, every event of autonomous replication and conjugation is initiated after the ICE is excised out of the chromosome and circularizes into an extrachromosomal form, leaving (for a period of time) the genome

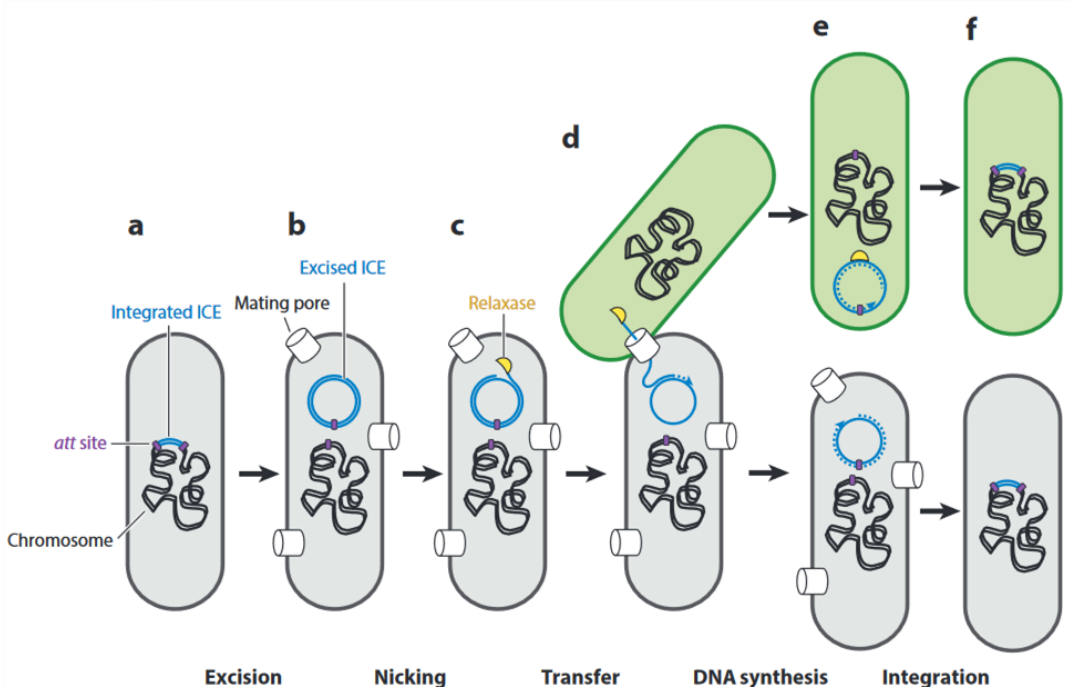


Figure 3. Bacterial ICE life circle. At first the ICE is integrated in the chromosome with the majority of its genes repressed. When the ICE excisize it forms a dsDNA circucal plasmid and the ICE-encoded proteins responsible for the assembly of the mating pore are expressed. An ICE-encoded relaxase nicks one strand of the ICE dsDNA and attaches itself to the 5' end of the nicked DNA, forming the transfer DNA. Then, the conjugation machinery transports the T-DNA into the recipient cell. In the recipient cell, the relaxase ligates the ends of the T-DNA to form a covalently closed ssDNA circle. The complementary DNA strand is synthesized to generate a dsDNA molecule that is the substrate for integration into the host chromosome. The same procedure is followed in the donor cell, with the ssDNA that remained after the transfer of the T-DNA. During the nicking, the transfer and the DNA synthesis, there is no trace of the ICE on the chromosome on neither cell (Johnson and Groddman 2015).

without any ICE genes. The excision process, which is the first step in the transfer of these elements among different cells, has been studied in mycoplasma, with successful detection of both the free circular form and of the empty chromosomal site (Marenda et al., 2006). In most cases, ICE encodes all the proteins required for the mechanism of conjugation and transmission and sometimes, these can be used in trans by other mobile elements simultaneously (Johnson and Grossman 2015). There are plenty of genes implicated in this transfer that are encoded by the ICE, some of which interact with host proteins in order to recognize the origin of transfer (oriT) and process the ICE DNA to generate a linear ssDNA-protein complex, referred to as the transfer DNA (T-DNA). The machinery responsible for the conjugation pore pumps the T-DNA into the recipient cell where the ICE get re-circularized, converted from single stranded to double stranded and finally it recombines into the chromosome using an ICE-encoded recombinase (integrase) (Johnson and Grossman 2015) (Figure 3).

Until recently, the conjugative machinery that initiates and controls the above process was considered a mechanism primarily used for plasmid conjugation. However, recent studies (Guglielmini et al., 2011) have concluded that the ICEs use the same machinery. The differences are that in the ICE case, it concerns large sequences integrated in the host's chromosome. It also contains some additional steps, which are the excision, followed by the circularization and a final step of re-integration in the genome, machinery similar to the phages life circle. It has been observed that small genomes rarely contain ICEs, whereas large genomes often do so. As a conclusion, ICE have been shown to play a key role in the diversification of prokaryotes by using existing mechanisms to allow exchange of large DNA fragments, possibly providing defense traits and new metabolic functions (Guglielmini et al., 2011). As it has been stated in the work of Guglielmini and his colleagues, plasmids and ICEs might be the two faces shown by a very similar type of element, hanging depending on the selection pressure and the importance of the cargo genes.

Among mollicutes, ICEs have been identified in several species. First described was the ICE called ICEF of the human-infecting *Mycoplasma fermentans* strain PG18 which exists in two versions and 4 copies, 1 for the ICEF-I version and 3 copies for the ICEF-II version. Both copies measure around 23 kbp and constitute approximately 8% of the *M. fermentans* genome. It has been characterized as a mobile and flexible gene pool that increases the plasticity of the genome and the diversity of the species (Calcutt et al., 2002). An ICE element has been also identified in *M. agalacticae* that contains 12 coding sequences homologous to ICEF and has a size of 27-kb. It was called $ICEA_{5632}$-I, and it occurs in at least 3 chromosomal copies in *M. agalactiae* strain 5632 (Marenda et al., 2006). In *M. agalactiae* strain PG2, only one degraded form of ICE was detected, suggesting the repertoire of these mobile elements might be very different from one strain to another (Sirand-Pugnet et al., 2007). The same ICE was identified in *M. bovis* (Marenda et al., 2005) a fact further support the theory that perhaps this ICE, not completely characterized yet, could harbor virulence determinants that may influence the pathogenicity of both bacteria. Extensive work has been conducted on the ICE of both *M. agalactiae* and *bovis* but also on two species of the *mycoides* group, *M. capricolum* subsp. *capricolum* and *M. mycoides* subsp. *capri* (Tardy et al., 2015). The authors identified in a collection of 166 field strains the following CDS1, CDS5, CDS17, and CDS22 as the «minimal ICE backbone» for ruminants mycoplasma species and also identified a set of inverted and direct repeats (IR and DR) that allowed excision and integration of the ICE between different areas of the same genome and also different organisms. In addition, the fact that these

elements encode the conjugation machinery may stimulate more general genome exchanges (Frisoni et al., 2013) and promote the emergence of new variants. Another work on spiroplasmas has demonstrated some important events of HGT among this genus (Lo et al., 2015) mainly from species of the same ecological niches with similar genomic characteristics, with a potential impact in the adaptation of the bacterium to its host.

c.Plasmids in Mollicutes

Some mollicutes species carry one or even several replicative plasmids. The first of these plasmids were identified in *Spiroplasma citri* (Mouches et al., 1983, Ranhand et al., 1980). Their size ranges from 5 kbp to 30 kbp and they have been associated with the transmission of these bacteria to its vector (Berho et al. 2006, Breton et al., 2010). In the *Candidatus* phytoplasma genus, even though these bacteria remain uncultivated, some specific plasmids have been identified. Their size ranges from 2.6 to 10.8 kbp (Firrao et al., 2007) and later studies identified a single conserved protein that allow a rolling-circle type of replication. For the Mycoplasma species, a study on 194 ruminant mycoplasma strains identified 37 plasmids in the mycoides cluster of the Spiroplasma group and in species that are close relatives of this group (Breton et al. 2012). This study also identified a common genetic organization with two CDS conserved in almost all plasmids found in mycoplasmas, one encoding a transcriptional regulator CopG and one encoding the replication protein Rep, suggesting a replicative mechanism similar to that found in *Candidatus* phytoplasma plasmids. Apart from these natural plasmids, there has been application of artificial plasmids, carrying the chromosomal origin of replication of the species in which they are transformed. These plasmids are called *oriC* plasmids and their development and application will be analyzed later on.

d. Phages

Phages are viruses that infect and replicate within a bacterium. They have a lytic cycle or a lysogenic cycle of life; lytic phages such as the T4 phage, invades bacterial cells, which are broken open (lysed) and destroyed after replication of the virion. In contrast, the lysogenic cycle does not result in immediate lysis of the host cell. In this case, the phages are able to undergo lysogeny and are known as temperate phages. Their viral genome will integrate the host chromosome and replicate along with it relatively harmlessly, or may even become established as a plasmid. The virus remains dormant until host conditions deteriorate, perhaps due to depletion of nutrients; then, the endogenous phages (known as prophages) become active. At this point they initiate the reproductive cycle, resulting in lysis of the host cell. As the lysogenic cycle allows the host cell to continue to survive and reproduce, the virus is replicated in all of the cell's offspring (Mason et al., 2011).

Mycoplasmas are species with a reduced genome as a result of dynamic evolution. Phages and prophages are labile elements that excise and integrate the genome in an unpredicted manner. One would therefore expect to find few, if any, prophages in the genomes of mycoplasmas. In mycoplasmas, only 3 phages and prophages that have been characterized; a phage of *Mycoplasma pulmonis* (virus P1), one that infects *Mycoplasma arthritidis* (MAV1) and one that infects *M. fermentans* (ϕMFV1) (Tu et al., 2001, Clapper et al., 2004, Röske et al., 2004). Recently, a prophage was found in a *M. agalactiae* strain and was most likely shared with the other ruminant pathogens *Mycoplasma conjunctivae* and

Mycoplasma bovigenitalium (Tardy et al., 2012). These phages seem to be phylogenetically distant from the phages species that have been already characterized and the interaction with their mycoplasma host follows the lysogenic cycle; they invade the host cell, integrate its genome in the bacteria chromosome and express proteins that work on their benefit. In *M. pulmonis*, the P1 virus ORFs had no significant similarity with the ones from other phages apart from its polymerase and it has been proven to be useful for probing both the antigenic makeup and the restriction enzyme activity of host cell populations (Tu et al., 2001). Its ORF8 codes for a product with a repetitive collagen-like motif, which is characteristic of some bacteriophage tail fiber proteins and is a candidate for interacting with the expression of the bacteria Vsa proteins, rending the cell susceptible to phage infection. For the MAV1 phage, some studies have shown the impact on the virulence of *M. arthritidis* of the presence of the MAV1 phage in its genome (Voelker and Dybvig, 1999). Further studies by the same laboratory discovered the first phage exclusion system in mycoplasma. This exclusion system is based on a lipoprotein expressed by the virus during its lysogenic phase, Vir, that protects the *M. arthritidis* cell from superinfection by the same or other phages (Clapper et al., 2004). Finally for the *Mycoplasma fermentans* ϕMFV1 phage, an equivalent to the VIR protein, called Mem, has been characterized as a surface protein that may provide the cell with features for adaptation and survival in the mammalian host environment (Röske et al., 2004).

e. Mobile element control

The transfer of genes between related or unrelated species via bacteriophage transduction, plasmid conjugation, and DNA transformation or cell fusion is fundamental for prokaryotic evolution. However, bacteria have also evolved systems to control and limit the impact of invading DNAs. Among these systems, the best characterized are the restriction-modification systems, the abortive infection (Abi) mechanisms and the CRISPR-Cas adaptive defense system.

Restriction-modification systems

All restriction-modification systems have enzymes that are responsible for two activities; a methyltransferase that adds specific modification to the genome of the bacteria and a restriction endonuclease that interacts with DNA targets that doesn't have these modifications. They are classified into four major groups depending on the number of enzymes responsible for the two activities, the recognition site and cleavage position and the cofactor its group requires (Tock & Dryden, 2005).

Three types of restriction-modification systems have already been characterized in mycoplasmas (Brocchi et al., 2007). Their mechanism varies, but in general, there is a modification enzyme that methylates the chromosome of the cell. This modification renders the genomic DNA immune to the second enzyme, which is a restriction enzyme that cleaves DNA at a precise location within or around the un-methylated recognition sequence (Neidhardt et al., 1996). For the type II RMS there are two distinct enzymes, whereas for the Type I and III, there is a protein with different subunits that controls both procedures, the modification and the restriction (Browning et al., 2005). Many mycoplasmas have more than one type and even multiple copies of the same system. Sequence variation in the sequence recognition subunits of RMS leads to the creation of new sequence specificities (Browning et al., 2005). RM systems of type IV have been predicted in a few mollicutes including *Me. florum, S. citri* and *A. laidlawii* (Breton, unpublished).

Abortive infection mechanism and Bacteriophage Exclusion system

The Abortive infection (Abi) mechanism begins with viral infection and injection of viral DNA into the host cell, followed by an interruption of phage development and the death of the infected cell leading to the release of few or no virus particles. While the infected bacterium dies, this reaction prevents further propagation of the phage and the bacterial population as a whole is more likely to survive. The Abi systems that have been characterized show significant variability in the number of enzymes involved. However, all systems characterized so far share a common feature: dormant bacterial enzymes are activated soon after phage infection and then cleave essential and highly conserved components of the cellular translational apparatus, thus halting protein and aborting the phage infection (Chopin et al, 2005). In mollicutes, homologous genes to $abiG_I$ and $abiG_{II}$ of Streptococcus have been identified in *M. agalactiae.* Their function is not yet characterized, but they are located in the vpma loci responsible for the expression of multiple surface lipoproteins (Novel et al., 2009). These genes have been also identified in *M. bovis* and have been a subject of HGT between the two previously mentioned species and other mycoplasma species (Qi et al., 2012).

Recently, a new phage-defense system has been identified in many bacteria and archaea, called Bacteriophage Exclusion or BREX. The system consists of a cluster of genes located in what is called the genomic defense islands (Makarova et al., 2011). The genes there have been implicated in phage defense, and include genes encoding proteins with putative protease domains, ATPase domains and RNA-binding domains (Goldfarb et al., 2015). The BREX system has not been identified in mollicutes yet.

CRISPR/Cas systems

The CRISPR/Cas system identified in bacteria and archaea (Horvath & Barrangou, 2010) serves as an adaptive immunity system that will be analyzed in detail later on (see chapter CRISPR below). In brief, the CRISPR locus contains sequences called spacers that match sequences on invading nucleic acids, such as phages or plasmids, called protospacers. The CRISPR associated proteins (Cas), are a family of proteins that carry functional domains including nucleases, helicases, polymerases, and polynucleotide-binding proteins (Horvath et al., 2010). These proteins interact with the maturation products of the CRISPR locus, called CRISPR RNAs, to provide immunity against the nucleic acids that contain the sequences of the protospacers. The system can inactivate these foreign elements and has been shared amongst organisms, even distantly-related ones, through HGT in order to render the cell safer in a hostile environment. CRISPR/Cas systems have been described in several mollicutes. The distribution and evolution of these systems will be extensively studied in the result section.

V. Pathogenicity and disease control

Most mollicutes live as commensals, and in many arthropods they may even be considered as symbionts. Some mycoplasmas are arguably close to “ideal parasites,” living in harmony with their host (Razin 2006). Their survival within their host is based on evasion techniques, such as mimicry of host antigens, survival within phagocytic and non-phagocytic cells and generation of phenotypic plasticity rather than toxin production (Rottem 2003). However their persistence presence and the intense acquisition of nutrients from the host cells and tissues induces an extensive stress reaction by the host. In a few cases there have been reports of pathogenic factors like the production of hydroxide peroxide or

other cytotoxins by mycoplasma during infection, but even without specific toxins the presence and proliferation of mollicutes is often enough to cause even a lethal stress to its host (Browing and Citti, 2014).

For spiroplasmas and phytoplasmas, studies have shown that their life cycle involves invasion and (intracellular) replication in plants and insects. Both groups of bacteria are located in the phloem sieve tubes of their plant hosts, whereas they can invade multiple organs and tissue types within the insect host (Browing and Citti 2014). They are obligate colonizers of their plant hosts and insect vectors and multiplication in both hosts is probably required for the complete life cycle to occur (Hogenhout et al., 2008). The *Spiroplasma* genus is one of the largest genera among the *Mollicutes* class, containing 37 species, of which only 3 have been described as plant pathogens (Browing and Citti 2014). One well studied model pathogen is *Spiroplasma citri*, a bacterium transmitted by an insect vector that is responsible for citrus stubborn disease in the Mediterranean area and California (Calavan and Bové, 1989) as well as horseradish brittle root disease in the United States (Fletcher et al., 1981). In order to control the propagation of this pathogen, many studies have focused on the interaction with its insect vector (Beven et al. 2015).

It is often difficult to demonstrate mycoplasma's role in diseases, as many are considered opportunistic pathogens. For many mycoplasmas, adhesion to their host is the most crucial part of the infection and usually the capacity to adhere differentiates the virulent from the non-virulent strains (Baseman and Tully 1997). Moreover, symptoms caused by response to mycoplasma infections can be diverse and variable according to the overall health of the host. Even though pathogenic mycoplasmas have been known for many years, the genetic basis of their pathogenicity remains largely to be explored.

A common way, by which mycoplasmas cause damage to their host, is the production of mildly toxic compounds as byproducts of their metabolism, such as the hydrogen peroxide and superoxide radicals. These products can cause oxidative damage to the host membranes, as it has been described (Almagor et al. 1986, Pilo et al., 2007, Hames et al., 2009, Blotz et al., 2017). In the mycoplasma family there has been also identified a unique cytotoxin in *M. pneumoniae* with homologs in *M. penetrans* and *M. iowae* called community-acquired respiratory distress syndrome, or CARDS (Kannan and Baseman 2006).

An important factor for the pathogenesis of many bacteria is the biofilm formation (Wang et al., 2017). It allows a better adhesion on the host and is generally followed by the formation of a polysaccharide matrix that surrounds the cell. This mechanism has been already identified as a crucial factor in mycoplasmas life circle and infection capacity; *M. bovis, M. putrefaciens, M. cottewii* and *M. agalactiae* formed prolific biofilms that increase their resistance to stresses including heat and desiccation (McAuliffe et al., 2006). *Mycoplasma mycoides* subsp. *Mycoides* (Mmm) exhibits differential gene expression when attached to a solid surface (McAuliffe et al., 2008). *M. pneumoniae* can grow faster if a biofilm is developed during infection (Simmons et al., 2015). Finally, *Mycoplasma mycoides* subsp. *mycoides* (Mmm) strain Afadé, demonstrated a longer bacteraemia in a mouse model, when its cells are capsulated in an exopolysaccharides polymer (Gaurivaud et al., 2014).

Another way that the mycoplasmas can be harmful for their hosts is by reduction of the choline components of the eukaryote's membranes. This phenomenon has been observed for *M. fermentans* and it is due to the fusion of this mycoplasma with the membrane of its host (Ben-Menachem et al. 2001). Although this removal of choline components can be lethal for the host cells, it does not appear to have a significant impact on the survivability of the organism as a whole.

The mycoplasmas, similar to both spiroplasmas and phytoplasmas, are capable of invading the host cells by adhesion on its membrane and partially or penetration of the entire mycoplasma cells inside the host cytosol. Several mycoplasmas are capable of this invasion; *M. penetrans* and *M. genitalium* appear to enter the cells through their specialized tip structure (Lo et al., 1993; Jensen et al., 1994), while other mycoplasmas shown to internalize, like *M. fermentans* and *M. hominis* have no tip structures (Taylor-Robinson et al., 1991). This invasion, when it occurs outside a vacuole, exposes the cytoplasm and the nucleus to hydrolytic enzymes of the mycoplasmas, such as proteases, nucleases and phospholipases (Rottem 2003).

Even though it hasn't been proven to be frequent, some mycoplasmas can cause clastogenic events, i.e. deletions, insertions and rearrangements of the host genome, due to the action of their nucleases. These events have been proven to effect the development of human cancer and apoptosis of the host cells (Razin 2002).

Finally, some work has uncovered the role of a symbiosis with a lysogenic bacteriophage MAV1 infecting *M. arthritidis* to play an important role in the pathogenicity of this murine mycoplasma. Virulent strains have been tested and it has been observed that they all carry MAV1 DNA integrated at various sites of the mycoplasma chromosome, whereas avirulent strains lack MAV1 (Voelker and Dybvig, 1999; Razin 2002).

Another important concern is the role of mycoplasmas in disease pathogenesis. An analysis has revealed a potential role of these organisms as cofactors in AIDS pathogenesis, the Gulf War Syndrome, and other diseases of unexplained etiology such as the chronic fatigue syndrome, Crohn's disease, and various forms of arthritis (Baseman, J. B., and J. G. Tully. 1997). The mycoplasmas are not directly responsible for any of these diseases but when an organism is weakened during treatment, it has been observed that mycoplasmas, such is the *Mycoplasma pneumoniae* (Razin 2006), which have a high contamination rate, can contaminate many patients in the same clinic and switch from symbiotic to pathogenic interactions with their host.

In addition to these mechanisms, adhesion of mollicutes to host cells is a prerequisite for colonization by the parasite and for infection. The loss of adhesion capacity by mutation results in loss of infectivity, and reversion to a cytadhesion phenotype is accompanied by regaining infectivity and virulence (Razin 2006). The system has been well characterized in species like *M. pneumoniae* where three major proteins, P1, MgPa and P30 play the key role in the cytadherence. These proteins are accompanied by a number of accessory membrane proteins that allow the adhesion of the bacterium on the host membrane and its movement on it. Some of these accessory proteins are the HMW1, that seems to be responsible for addressing (trafficking) the P1 adhesion protein to the attachment organelle

of *M. pneumoniae* (Balish et al., 2001). The receptors on the host membrane that allow the mycoplasma attachment are mostly sialoglyco-conjugates and sulfated glycolipids (Razin 2006).

In order to control the spread of mycoplasmas, a variety of vaccines has been developed. Those are live vaccines, attenuated vaccines and inactivated vaccines. The most efficient are the attenuated vaccines, which are produced after a bacterium strain has been modified either by laboratory passage or by deliberate mutagenesis (Browning et al., 2005). There are currently effective attenuated vaccines available to control diseases of poultry caused by *M. gallisepticum* and *M. synoviae*, an effective inactivated vaccine to control contagious caprine pleuropneumonia, caused by *M. capricolum* subspecies *capripneumoniae*, and inactivated vaccines of limited efficacy to control enzootic pneumonia in pigs, which is caused by *M. hyopneumoniae*.

However these vaccines have limitations as was observed in the vaccine for *Mycoplasma mycoides* subsp *mycoides* or simply Mmm. Mmm is the agent of contagious bovine pleuropneumonia (CBPP). The vaccine developed against Mmm and applied in a great scale, is live attenuated strain Mmm T1/44. The first applications of this strain gave very positive results, but with two important drawbacks. First, severe post-vaccinal lesions at the site of inoculation and, second, having to re-vaccinate the treated animals about 8 months after the first vaccination in order to re-establish the immunization (Thiaucourt et al., 2000). Another strain, T1sr, a streptomycin-resistant variant that gives fewer post-vaccinal reactions was developed to by-pass the lesion problem, but its application in various countries in the southern part of Africa was unsuccessful in providing immunity to the animals. The scientific community is thus obligated to develop new vaccinal strains capable of maximizing the efficiency of the immunization and the stability of the protection of Mmm hosts.

Another mycoplasma for which no attenuated vaccines have been developed so far is *M. pneumoniae*. Up to one-fifth of all lung infections that people develop in their community (outside of a hospital) are caused by this bacteria. The bacteria can cause tracheobronchitis (chest colds), sore throats, and ear infections as well as pneumonia. A dry cough is the most common sign of infection. Untreated or severe cases can affect the brain, heart, peripheral nervous system, skin, and kidneys and cause hemolytic anemia. In rare cases, *M. pneumoniae* infection can be fatal. Early diagnosis is difficult because there are few distinguishing symptoms. As *M. pneumoniae* infection progresses, imaging and laboratory tests may be able to detect it, leading to the prescription of an adapted antibiotic treatment. However, as for many pathogenic bacteria, antibiotic efficiency may be altered by spontaneous or acquired resistance. The development of the vaccine for this species poses difficulties because, unless the vaccine is heavily attenuated, it will not be safe for applications in weak patients (Browning et al., 2005). Unfortunately, the experimental vaccines that have been developed so far have sometimes caused the emergence of the disease in treated patients (Browning et al., 2005).

In recent years, investigations of novel strategies to develop more efficient vaccines against mycoplasmas have included protein subunit vaccines, DNA vaccination, recombinant protein vaccines, and use of vaccine vectors expressing mycoplasma genes. It is hoped that the integration of epidemiological studies with fine molecular typing will induce a better knowledge on the dynamics of mycoplasma strains evolution and, finally, allow a better evaluation of risk and better disease control strategies.

VI. Genome engineering of mollicutes

In order to study mollicutes and characterize the relationship between biological properties and genomes, genetic tools that allow a functional analysis of these bacteria are needed. The best way to do so is to develop mutagenesis tools that allow the development of mutants in specific genes or operons by inducing deletions, insertions, replacements or point mutations. Ideally, these modification tools should provide mutants in a quick, cheap and efficient way. One important application of these mutants is to identify pathogenicity factors and modify/delete them in order to develop attenuated strains that can then be used as vaccines. Several different mutagenesis strategies have been used for mycoplasma and other mollicutes: Random mutagenesis using transposons, directed mutagenesis using suicide or replicative *oriC* plasmids and novel tools of synthetic biology.

a. Random mutagenesis using transposons

The idea behind the use of transposons is to randomly introduce a small insert carrying a gene that provides resistance to an antibiotic in the genome and thereby inactivate the gene in which it is inserted. This type of mutagenesis provides a library of mutants from which the mutant in the desired gene must be selected. In mollicutes two transposons have been widely used so far. The conjugative transposon Tn916 (18 kbp) originated from *Enteroccocus feacalis* with the tetracycline marker in its sequence (Clewell and Gawron-Burke 1986). The second transposon is named Tn4001 (4.7 kbp, isolated from *Staphylococcus aureus* (Lyon et al. 1984) and that carries the *aacA-aphD* gene that encodes an enzyme responsible for resistance to three antibiotics: kanamycin, gentamicin and tobramycine (Rouch et al. 1987). The first experiments with these transposons have been conducted on *A. laidlawii, Mycoplasma pulmonis and Mycoplasma hyorhinis* with the plasmid of *E. coli* pAM120 carrying the Tn916 transposon. Among the transformants, the frequency of resistant cells to tetracycline was 10^{-6} tfs/UFC/µg DNA for *A. laidlawii* and *M. pulmonis* and 10^{-8} tfs/UFC/µg DNA for *M. hominis* (Dybvig and Cassell; Dybvig and Alderete 1988). The same plasmid was used to transform Mmm Large Colony (LC) and *M. gallisepticum* with transformation frequencies of 10^{-6} and $2x10^{-5}$ tfs/UFC/µg of plasmid DNA (Whitley and Finch 1989; King and Dybvig 1991; Cao et al; 1994; Whetzel et al. 2003). These results demonstrated that it is possible to express the *tet*(M) and the genes responsible for transposition in mollicutes. Unfortunately, the Tn916 transposon can spontaneously excise itself and re-insert elsewhere in the genome, leading to instability in the mutants generated (Dybvig and Alderete 1988; King and Dybvig 1991).

Other researchers have tried the Tn4001 transposon and some modified derivatives. A successful transformation of *M. gallisepticum* was conducted using the plasmid pISM1001, carrying the Tn4001 with an efficiency of 10^{-6} tfs/UFC (Cao et al. 1994). Unfortunately, the first efforts to transform with the same plasmid *M. pulmonis* and *M. arthritidis* failed. This is likely because wild-type organisms of these species are not sensitive to the selective marker, gentamycin.

Sometime later, the team of K. Dybvig managed to overcome the problem by making two different modified versions of the transposon Tn4001 one where gentamicin resistance was replaced with the gene conferring chloramphenicol resistance (Tn4001C) and one where it was replaced by tetracycline resistance (Tn4001T). Transformants resistant to tetracycline were obtained for both species using the Tn4001T, but only *M. pulmonis* was resistant to chloramphenicol after transformation with the Tn4001C.

Another interesting application was conducted in *M. pneumoniae* where Zimmerman and his colleagues designed a plasmid based on the pMT85, a plasmid already used to introduce transposons inside mycoplasma cells (Zimmerman et al., 2005), where they introduced the Tn4001C but the transposase was introduced outside of the transposon, in another area of the plasmid. Because they were using a non-replicative plasmid, the transposase was rapidly lost and as a result the modified *M. pneumoniae* genome was as stable as the wild type (Zimmerman et al., 2005). A similar unmarked mutation was produced on the genome of *Mycoplasma mycoides subsp. mycoides* (Janis et al., 2009).

Using this kind of random mutagenesis and creating libraries of mutants, scientists managed to inactivate genes implicated in the pathogenic mechanisms of *M. pneumoniae, M. genitalium, M. gallisepticum* and *S. citri*. Some experience in *M. pneumoniae, M. genitalium* and *M. gallisepticum* identified genes implicated in the adhesion capacity of these bacteria, by creating mutants that were unable to be absorbed on the membranes of red blood cells (Hedreyda and Krause 1995; Reddy et al. 1995; Mudahi-Orenstein et al. 2003). In *S. citri,* applications of the above mutagenesis revealed the interaction of the pathogenic function of this bacterium with the capacity to consume fructose found in the phloem of the plant host (Foissac et al. 1997; Gaurivaud et al. 2000).

Finally, global random mutagenesis using the Tn916 transposon has been used in *M. genitalium* and *M. pneumoniae* in order to identify the genes essential for cell life. This project was based on the idea of constructing a minimal cell and in the effort of developing it (Hutchison et al. 1999, Glass et al., 2006). This approach allowed identifying the essential genes *in vitro*. More than 2200 insertions among a library of mutant of transposons for both species were analyzed. As expected, few, if any insertions are found in essential genes (e.g. *dnaA, gidB*) and this was the first proof of concept of this study. During this study, the number of essential genes was predicted to be between 265 and 350 out of the 480 annotated in *M. genitalium* genome. Similar works have been conducted of *Mycoplasma pulmonis*, where 321 of the 782 protein coding regions were individually inactivated (French et al., 2008). In *M. pneumoniae* a study has identified the crucial role of small ORFs in the development of a minimal genome that can support a replicative minimal cell and other regulatory elements that function as building blocks that are important for the development of this minimal living system (Lluch-Senar et al., 2015).

These were some of the major studies conducted in mollicutes using transposon mutagenesis. The transposon approach has the benefit that in a single experiment you can create a library that can be enlarged each time the transformation is repeated. This library can be stored and used for many functional analyses. However, such random mutation strategies have two main drawbacks: first, even for bacteria with small genomes as mollicutes, a global mutagenesis ensuring that all non-essential gene has been disrupted require thousands of transformants that may be problematic to generate and store. Second, efficient screening of mutants without predicted selectable phenotype remains a time and cost expensive process. As a result, other researchers have developed directed mutagenesis protocols to reduce the time required to create and study a mutant.

b. Directed mutagenesis

Directed mutagenesis aims at precisely inactivating or modifying candidate genes chosen by the scientist. In most cases, such a strategy relies on the endogenous homologous recombination mechanism

of the organism in order to introduce an imported DNA fragment inside the sequence of the candidate gene (simple crossing-over) or replace the candidate gene with the extracellular DNA (double crossing-over). The template for DNA recombination DNA usually includes at least a selectable marker and DNA arms on both sides that are identical to sequences aside from the targeted site. The length of the DNA arms depends on the efficiency of the homologous recombination and can largely vary from one organism to another.

i. Directed mutagenesis using suicide plasmids

A few applications of directed mutagenesis through homologous recombination using suicide plasmids have been achieved in some mollicutes. The first experiment was performed in *Acholeplasma laidlawii*, where a non-replicative plasmid carrying an internal fragment of the *recA* gene, was transformed inside the cell and manage to inactivate the *recA* gene. The results were verified due to a reduced DNA break repair efficiency (Dybvig and Woodard 1992). More recently, in M*. genitalium*, Dhandayuthapani and his colleagues managed to delete two candidate genes, *mg218* and *mg408*, using non replicative plasmid (Dhandayuthapani et al., 1999, Dhandayuthapani et al., 2001). The Cre-Lox technology was also introduce in *M. genitalium* cells using suicide plasmids (Mariscal et al., 2016). Another work, similar to the one on *A. laidlawii*, targeted the MG_339 locus which codes for the *recA* gene of *M. genitalium*, using a suicide plasmid and managed to deactivate the gene (Burgos et al., 2012). Although this technique has worked in a few cases, its poor efficiency has led researchers to, develop *oriC* plasmids as a more efficient alternative.

ii. Directed mutagenesis using *oriC* plasmids

oriC plasmids are artificial plasmids carrying the chromosomal origin of replication of the species in which they are transformed. In plasmids developed for mollicutes, the selected *oriC* region generally includes the *dnaA* gene and the surrounding intergenic sequences harboring DnaA box sequences. The *dnaA* gene encodes a protein that initiates the formation of the replication pre-priming complex. The interaction of the DnaA protein with the DnaA boxes, which for the mycoplasma are usually short motifs of 9 bp (Lartigue et al., 2003), leads to the separation of the DNA strands, the entry of the replication machinery and the formation of the replication forks (Neidhardt et al., 1996). Several methods have been developed to allow the scientists to identify the *oriC* sequence of a particular bacterium. A first attempt was the construction of a replication order map from measurements of the relative frequencies of various genetic markers, and by direct determination of their time of replication in *Bacillus subtilis* (Sonenshein et al., 1993). Other methods developed later are the GC skew analysis or two-dimensional replicon mapping which allows prediction of the location of the *oriC* from the whole genome sequence (Neidhardt et al., 1996, Miyata et al., 1997). In spiroplasmas, a 5.6-kbp fragment of *Spiroplasma citri* DNA containing the *dnaA* gene the DnaA-boxes upstream of the *dnaA* gene and a tetracycline marker were cloned and led to the construction of the first *oriC* plasmid for mollicutes (Ye et al., 1994). In mycoplasma, the work of Lartigue and her colleagues (Lartigue 2003) has provided *oriC* plasmids for Mmm, Mmc and *M. capricolum.* In general, the introduction of the *dnaA* gene with the intergenic regions that surround it, can provide a functional replicative plasmids that be maintained in the cell as extra-chromosomal elements. Such plasmids have been successfully developed for *M. pulmonis* (Cordova et

al., 2002), *M. agalactiae* (Chopra-Dewasthaly et al., 2005), *M. gallisepticum* and *M. imitans* (Lee et al., 2008), *M. hyopneumoniae* (Maglennon et al., 2013), *M. florum* (Matteau et al., 2017), *Spiroplasma eriocheiris* (Terahana et al., 2017), *M. synoviae* (Shahid et al., 2014), *M. bovis* (Li et al. 2015) and *M. hyorhinis* (Ishag et al., 2017). However the oriC plasmids used in the above studies didn't always manage to successfully induce targeted recombination and inactivation of genes. For example, the oriC plasmid developed from Lee and his colleagues, which was unable to induce gene disruption through homologous recombination in a strain of *M. imitans* (Lee et al., 2008).

OriC plasmids can be used for various applications including expression of proteins for complementation, heterologous expression or the study of gene regulation via a reporter gene. When harboring a selectable marker surrounded by recombination sequences, *oriC* plasmids can also be used for directed mutagenesis like the work conducted in *S. citri* (Duret et al., 1999).

The limitations of this technique are the fact that it isn't easy to develop *oriC* plasmids for all organisms, and even if such synthetic replicons have been developed for a large number of mollicutes, some species remain refractory to transformation or maintenance of these plasmids. Moreover, recombination of *oriC* plasmids with chromosome can be hard to obtain as the recombination efficiency in most mollicutes is very low, likely because the gene sets dedicated to DNA repair are minimal in mollicutes (Rocha et al., 2005). Interestingly, a supplementary *recA* gene, which encodes a key protein involved in the DNA recombination and repair, was integrated in the genome of Mmc (Allam et al. 2010) and *Mycoplasma hyorhinis* (Hassan et al. 2017) in order to boost homologous recombination. The results showed an increase of the recombination events in the resulting clones after transformation, but the application of such selections as well as the development of *oriC* plasmids is still limited to a few species among the mollicutes.

In addition, in the context of global studies such as gene essentiality studies, transposon-based approaches and small-scale directed mutagenesis only produce individual mutants which can be sometimes misleading for essential functions that can be achieved by non-homologous but functionally redundant genes.

In order to modify the genome of mycoplasmas at a larger scale, new synthetic biology tools were recently developed. These tools include genome synthesis, cloning and transplantation, first developed at JCVI by Gibson and his colleagues and Lartigue and her colleagues (Gibson et al., 2008, Lartigue et al., 2007) that have opened a new page in the Mycoplasma research.

c.Synthetic biology approaches

i. Cloning natural or synthetic genomes of mycoplasmas in yeast

An evolutionary step that allowed further expansion of the mutagenesis toolkit for mollicutes was the cloning of natural and synthetic genomes of mycoplasmas in yeast to make use of the genetic tools available in this organism. This work was conducted by Gibson and his colleagues, and the first genome was that of *M. genitalium* (Gibson et al., 2008). The process was based on DNA cassette synthesis, in vitro assembly, cloning in *E. coli* and, finally cloning in *S. cerevisiae*. The whole genome of the bacterium was partitioned into 101 cassettes of approximately 5 to 7 kb in length.

The majority of these cassettes was obtained from DNA synthesis companies and overlapped their adjacent neighbors by 80 bp. In the first stage, sets of four neighboring cassettes were assembled by in vitro recombination and joined to a bacterial artificial chromosome (BAC) vector DNA to form circularized recombinant plasmids with ~24 kb inserts. Then, the 24 kbp inserts where further assembled, forming inserts of ~72 kbp, each one covering about 1/8 of the entire genome. A final assembly produced 4 inserts of ~144 kbp size. All these assemblies were done by in vitro recombination and cloned into *E. coli* (Gibson et al., 2008). The last assembly of the 4 inserts was conducted in yeast, as described before (Kouprina et al., 2003). In order to preserve the assembly product in yeast, the genome was modified in order to include some yeast elements; an auxotrophic marker in histidine (HIS), a yeast centromere (CEN), and a yeast autonomously replicating sequence (ARS), for selection and propagation in yeast as a yeast centromeric plasmid (YCp). This work demonstrated that is was possible to assemble the whole chromosome of a bacterial genome in yeast. The technique was used 3 years later in an effort of Gibson and his colleagues to synthesize the genome of a second mycoplasma, *M. mycoides* subspecies *capri* (Mmc) (GM12) (Gibson et al., 2010). This assembly took place inside yeasts cells. The stability of the genome was verified and this novel system allowed scientists to sustain a mycoplasma genome in yeast.

ii. Genome transplantation as a key process to get a living bacteria

Genome transplantation is a process where a whole bacterial genome from one species is changed to another resulting in a new cell that has the genotype and phenotype of the incoming genome. At the same time, and because the whole process occurs in a selection media, the genome of the recipient cell is lost and finally replaced by the imported genome that contain the selectable marker. The first genome transplantation was the subject of the work of Lartigue and her colleagues at the JCVI (Lartigue et al., 2007); they succeeded in immobilizing intact chromosomes of the *Mycoplasma mycoides* subspecies *mycodes* species and have been introduced successfully in *Mycoplasma capricolum* cells by polyethylene glycol–mediated transformation. The recipient cells carried no traces of their original genome. This technique opened the possibility of transferring artificial or natural DNA molecules in recipient cells. Whole genome sequencing confirmed that there were no recombination events between the original and the imported chromosomes demonstrating a clean change of one bacterial species into another. The second application by Lartigue and her colleagues (Lartigue et al., 2009) was the transplantation of the genome of *M. mycoides* genomes cloned in yeast, thanks to the work of Gibson described above, into *M. capricolum* recipient cells. The same method of immobilization and polyethylene glycol–mediated transformation was applied. When this application was proven possible, the final step was the modification of a mycoplasma genome clones in yeast using the tools developed for yeast genome engineering, and then back transplantation of this modified genome in a mycoplasma recipient cell (Lartigue et al., 2009). The resulting cell was a positive mutant for the modification applied on the chromosome cloned in yeast. . This tool allowed a big step forward in the synthetic genomics field and it has been established and used for the first time in mycoplasmas.

Finally the two approaches developed above were combined in order to develop the first bacterial cell controlled by a synthesized genome (Gibson et al., 2010). For this approach, the Mmc strain GM12 genome cloned in yeast was chosen as a donor and *Mycoplasma capricolum* subsp*. capricolum* strain California kid was chosen as the recipient cell. Apart from the ARS-CEN-HIS elements, the authors also

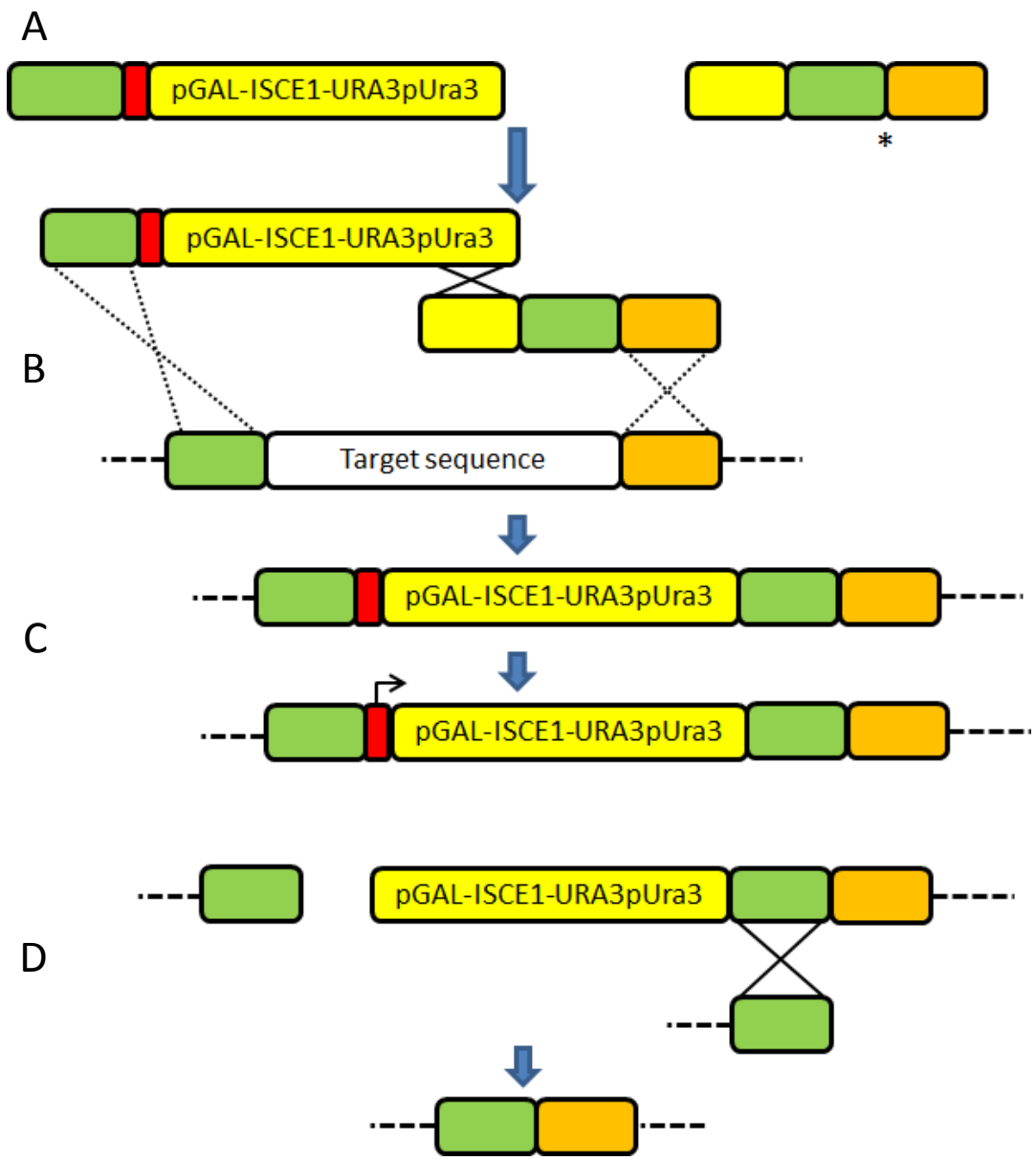


Figure 4. Design of the TREC system; A. Construction of the SCE1 cassette (yellow) and a second cassette to add a DNA fragment homologous to the sequence upstream of the target site (Green) to the SCE1 cassette; B. transformation of both cassettes in yeast and recombination as it is shown, first between them and then with the 50bp homologous ends to the targeted site (Green and Orange); C. Activation of the inducible SCE1 and cleavage of the restriction site of this enzyme (red); D. Recombination between the free ends of the upstream region or the targeted site and its homologous sequence added downstream of the SCE1 cassette (green), resulting in a clean deletion. An insertion can be also applied if between the 50bp of homology to the upstream flanking sequence of the target site added to the SCE1 cassette (asterisk) we introduce a gene or a sequence of interest.

introduced into the mycoplasma genome a tetracycline-resistance marker and a β-galactosidase gene for screening. The process is based on three distinct steps; isolation of intact donor Mmc genomes from mycoplasma cells, preparation of recipient M. *capricolum* cells, and introduction of the isolated genome into the recipient cells. The efficiency is low (1 recipient cell every 150,000). This application paved the way for the next stage, the modification of a mycoplasma genome (and most important for a species for which we don't have any engineering tools) outside the mycoplasma cell and re-introduction of the modified genome into a recipient cell using the genome transplantation technique described above. During the same year, Lartigue and his colleagues (Lartigue et al. 2010) managed to create mycoplasma strains by cloning in mycoplasma recipient cell genomes that have already been cloned and modified in yeast. The genome with the ARS-CEN-HIS elements was successfully modified by deleting a non-essential Type III restriction endonuclease gene. The deletion was conducted by adding a DNA cassette containing an 18 bp I-SceI binding site, the SCEI endonuclease gene under the control of the GAL1 promoter and a uracil (URA3) marker. The cassette had homologous ends with the ending sequences of the Type III restriction endonuclease gene. After a first selection in uracil, the gal promoter was activated, the SceI enzyme was encoded and the cassette was cleaved. Counter selection with 5-fluoroorotic acid (Boeke et al., 1984) produced clones that had lost the URA3 cassette due to a second recombination event that remove the URA element and the outcome of this was a mycoplasma genome modified inside yeast in a traceless way (Figure 4). The successfully modified genome was finally extracted and re-introduced in a mycoplasma cell that had the desired phenotype at the end of the process. This combination of individual experiments allowed the modification of mycoplasma genomes inside yeast and then a back-transplantation of these genomes inside mycoplasma recipient cells to create mutants with a directed way, efficiently and independently of the mycoplasmas low efficiency in homologous recombination. The genome transplantation tool has been extended since to other species among the mollicutes (Labroussaa et al., 2016)

iii. Tools for genome engineering of bacterial genomes cloned in yeast

Together with genome synthesis and transplantation technologies, some tools dedicated to the manipulation of genomes clones in yeast were developed: the TREC and TREC-IN (Noskov et al., 2010 and Chandran et al., 2014). The TREC method is used to delete a DNA fragment and is an improvement of the system used by Lartigue and her colleagues (Figure 4). It is based on the addition to the DNA cassette of a DNA fragment homologous to the sequence upstream of the target site. This sequence can be followed by an insertion gene if this is the desired modification. The whole cassette is flanked with 50 bp homologous end to the targeted gene. The insertion step is the same as before. After the activation of the SceI and the double strand break (DSB) on 5' end of the TREC cassette where the SceI site has been introduced, the sequence upstream of the target site and the homologous part introduced in the TREC cassette would be recombined by the reparation mechanism of the yeast for DSBs. This will completely remove any trace of modification, apart from the possible insertion downstream of the reparation site, and the mutants can be isolated by a counter selection for uracil. The TREC-IN is an improvement of the gene insertion capacity of the TREC system (Figure 5). In short, the cassette is modified with the addition of 5' part of the Kanamycin resistance marker called KanMX. After the initial introduction and selection in uracil, a second transformation of a second cassette is applied; the second cassette contains the 3' part of the KanMX, the homologous sequence upstream of the target site and the desired insertion (gene,

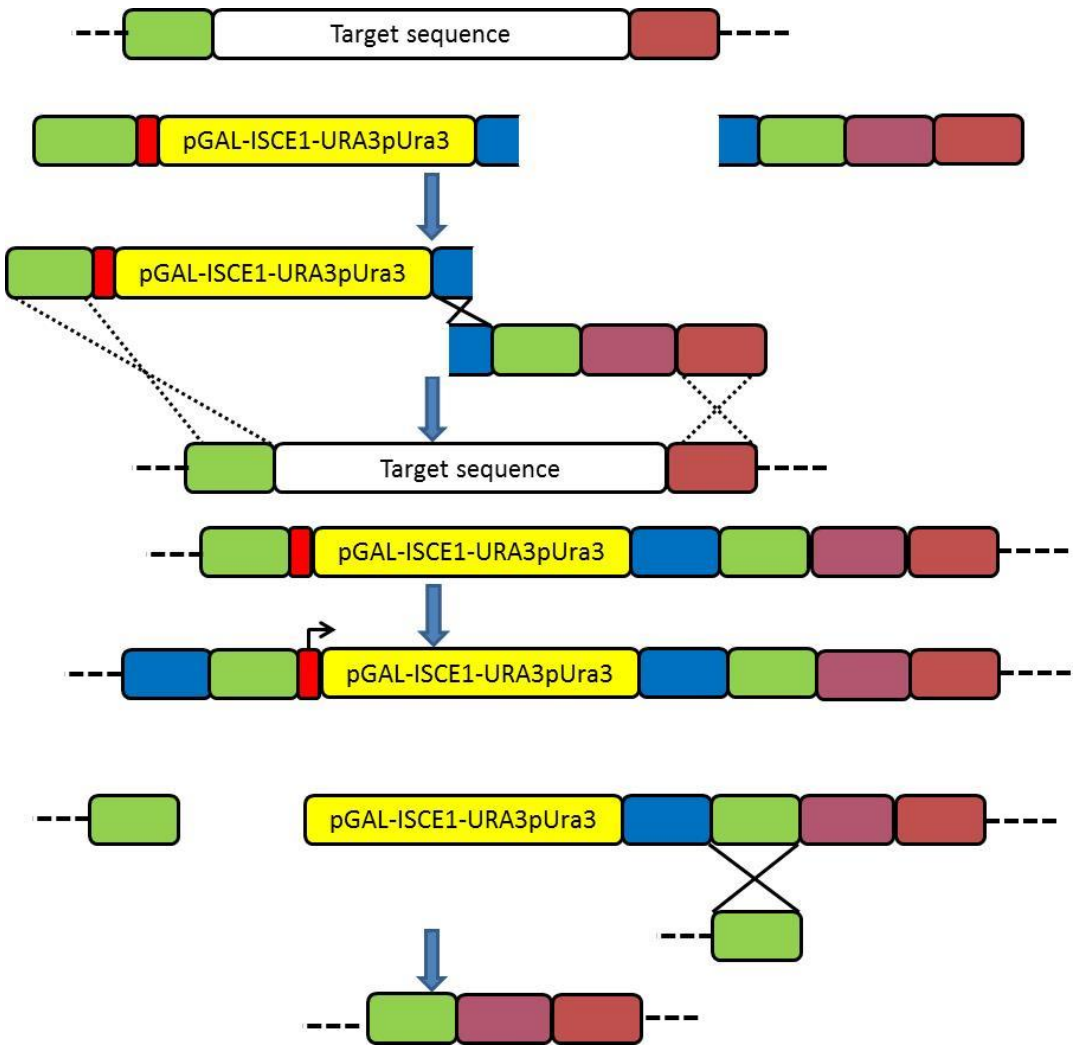


Figure 5. Design of the TREC-IN system; The TREC-IN system is similar to the TREC system but with an extra functional restoration of the kanamycin resistance gene module, kanMX. The TREC-IN is mostly used to introduce genes or sequences (purple) inside a targeted site using a knock-in module. The knock-in module includes the 3' part of the KanMX gene (blue) followed by 50nt homologous to the sequence upstream of the targeted site (green) followed by the gene to be inserted (purple). The SCE1 cassette is also modified to carry the 5' part of the KanMX gene. The same procedure as before is followed to induce cleavage of the SCE1 site and removal of the Core cassette (yellow) together with the resistance marker to kanamycin (blue). The final mutant has a clean insertion of a candidate gene (purple) in the desired region.

marker etc). The activation of the SceI endonuclease by putting the cell in a growth medium containing Galactose provokes the same reaction with the TREC, with the elimination of all the elements apart from the inserted sequence. The TREC-IN process was developed in order to increase the efficiency of the insertion of DNA sequences on a targeted area by eliminating all the background noise that was the main drawback of the TREC. It has also been used for large seamless deletions with a good efficiency.

These are all the systems developed so far for the modification of mycoplasma genome, directly inside the mycoplasma cells, or in-directly by the in-yeast engineering followed by a back transplantation.

VII. CRISPR/Cas system

a- Natural prokaryotic immune systems

Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR associated proteins (Cas) constitute natural systems that have been described in a wide range of prokaryotes, including the majority of Archaea and many Bacteria; roughly 90% of the Archaea and 40% of the Bacteria have these systems (Grissa et al. 2007). CRISPR is system responsible for an adaptive immunity against foreign DNA elements. It is generally located in a locus inside the genome and when expressed, it has the capacity to protect the cell from invading DNA molecules, such as bacteriophage's DNA or plasmids. Its main characteristic, which is common to all types of CRISPR systems, is a profile of short direct repeats interspaced at regular intervals by unique spacer sequences.

The system was discovered "by accident", in a study published in 1987, in which the authors unknowingly discovered the first genomic CRISPR locus in *E. coli* while sequencing a gene called *iap*, encoding a proteolytic enzyme, potentially responsible for isozyme conversion of alkaline phosphatase (Ishino et al. 1987). They discovered a part of a direct repeat sequence in the 3'-end flanking region of the *iap* gene. Then, different groups started studying the CRISPR locus in different bacteria, trying to understand the role of this typical arrangement (Nakata et al. 1989, Groener et al., 1993, Mojica et al. 1995 and 2000). In 2005, three different teams came up with an impressive observation: spacers derive from foreign genetic elements (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). This remarkable trait was confirmed in 2007 with the first undoubtable results from Barrangou and colleagues: while working with *Streptococcus thermophilus*, the authors challenged the bacteria population with bacteriophages. At first they used bacteriophages which genomes had matching sequences with the CRISPR locus of *Streptococcus thermophilus*. The resistant bacteria that grew after the challenge had no trace of bacteriophage infection. Following that, they challenged the bacteria once again, with bacteriophages with no sequences identical to the spacers. Only few bacteria survived. They studied the CRISPR locus of the rare surviving clones and they realized that small parts of the bacteriophages sequences had been introduced in the CRISPR locus of the bacteria that managed to grow. Thus, they demonstrated that CRISPR spacers confer potent resistance to bacteriophages for which they have matching DNA sequences, and that, at a low frequency, bacteria could also actively "vaccinate" themselves against bacteriophage by integrating new spacers into the pre-existing CRISPR locus (Barrangou et al., 2007, Horvath et al., 2008).

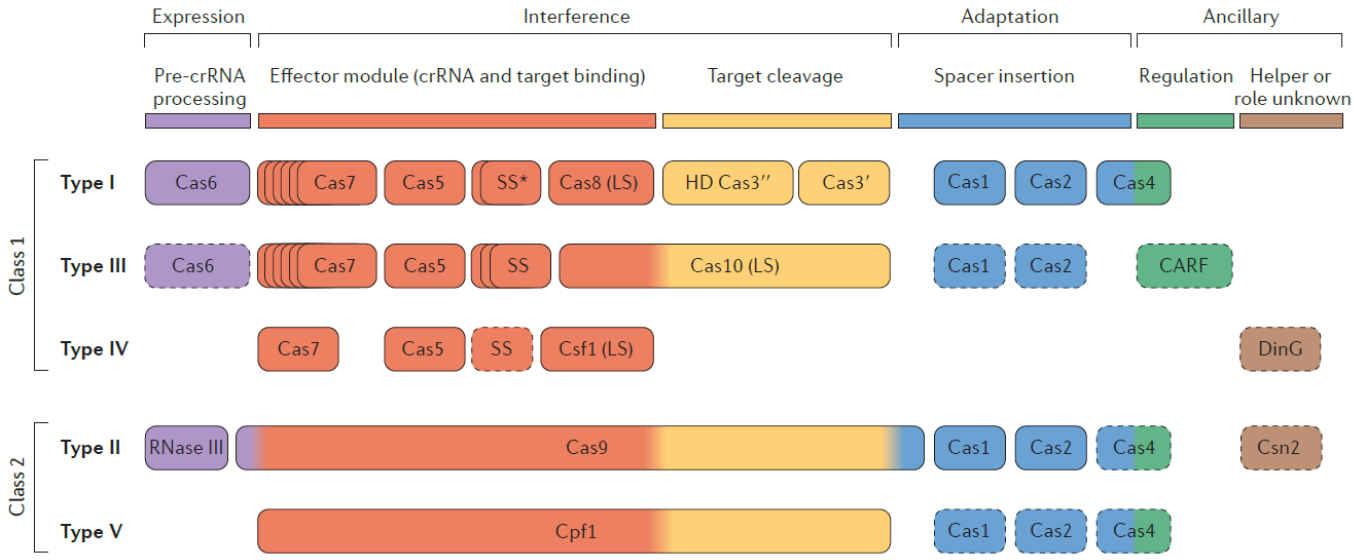


Figure 6. Cas protein classification in the different classes and types of CRISPR systems: Protein names follow the current nomenclature and classification. They are classified depending on their implication in the pre-crRNA processing, the assembly of the effector complex, the target cleavage, the spacer insertion in the CRISPR locus process, the regulation of the whole system or as an unknown helper in the CRISPR response with respective colors corresponding to each stage. The asterisk in the type I line, indicates that the putative small subunit (SS) protein is sometimes fused to Cas8 (the type I system large subunit (LS)) in several type I subtypes. The type III system LS and type IV system LS are Cas10 and Csf1 (a Cas8 family protein), respectively. Dispensable components are indicated by dashed outlines. Cas6 is shown with a solid outline for type I because it is dispensable in some but not most systems and by a dashed line for type III because most systems lack this gene and use the Cas6 provided *in trans* by other CRISPR–*cas* loci. The multiple colors for Cas4 and Cas9 reflect that these proteins contribute to different stages of the CRISPR–Cas response. The type VI is not included in this picture, as it remains largely uncharacterized. CARF, CRISPR-associated Rossmann fold; pre-crRNA, pre-CRISPR RNA. Figure adapted from Makarova et al., 2015.

b- Functional characterization of the CRISPR system

The expression of the CRISPR locus was identified as the key component in this adaptive immunity by van der Oost laboratory (Brouns et al., 2008). CRISPRs function together with CRISPR-associated (cas) proteins, which are encoded by genes that typically flank the CRISPR locus in the genome, and the entire system is consequently referred to as CRISPR/Cas. CRISPR-mediated adaptive immunity proceeds in three distinct stages: acquisition of foreign DNA, CRISPR RNA (crRNA) biogenesis, and target interference.

Until recently, six(I–VI) types of CRISPR/Cas systems have been identified and divided into two major classes, class 1 and class 2, according to the complexity of their interference machinery (Makarova et al., 2015) (Figure 6). The different systems, whatever their classes, types and subtypes, share some common features in terms of global process: During acquisition, new spacers are acquired from foreign nucleic acids and integrated at one end of the CRISPR locus. The CRISPR locus is then transcribed as a precursor of the CRISPR RNA (pre-crRNA). The pre-crRNA is enzymatically processed into mature crRNAs which are then bound by the CRISPR interference complex or protein (depending on the class of the CRISPR) to form ribonucleoprotein targeting complexes. Each protein or protein complex carries a single crRNA with a single spacer (guide) sequence. The ribonucleoprotein complexes move freely around the cell (Gasiunas et al., 2012) similar to a surveillance system and every time a sequence corresponding to the spacer sequence is identified, a Cas protein with a nuclease activity cleaves the targeted site. An important factor in the targeting capacity of the Cas proteins, that allow them to recognize the difference between self, versus non-self targets, has been shown to involve a short sequence motif that is preserved in the foreign genome, referred to as the protospacer adjacent motif (PAM) (Sorek et al., 2013). This motif is conserved in the target DNA of phages and plasmids and it has been identified downstream of each spacer match (Bolotin et al., 2005). The PAM is required for successful Cas-target interaction and abolishment or modification of the PAM sequence that corresponds to each cas protein, provokes an inhibition of the interference capacity of the cas protein (Mojica et al., 2009, Anders et al., 2014).

During the CRISPR adaptive immunity process, different proteins participate at different steps which include the processing of pre-crRNA, crRNA and target binding, target cleavage, insertion of spacers in the CRISPR locus and regulation of the whole mechanism(Figure 6). There are still some proteins among the Cas genes for which the function is still not entirely understood and characterized. Some comparative genomics studies have tried to interpret the function of these genes, but there role remains unclear (Makarova et al. 2013, Makarova et al. 2015). The separation between class 1 and class 2 has been suggested based on the crRNA–effector module; for the class 1 there is a complex of proteins that is assembled together and is capable of binding the crRNA and interfere with the spacer-homologous sequences. For the class 2 there is a single protein capable for the same effect. Class 1 includes systems from type I, III and IV and Class 2 include types II and type V. The type VI remains largely uncharacterized (Wright et al., 2016, Yang and Patel 2017).

Type II CRISPR/Cas system is the most characterized system and it is also the first CRISPR/Cas system that has been used as a tool for molecular biology. Type II CRISPR system includes a trans-activating crRNA (tracrRNA) which is a molecule of variable size (Chilynski et al., 2013 and 2014) that is encoded in the vicinity of the cas genes and the CRISPR locus and has complementary sequences to one

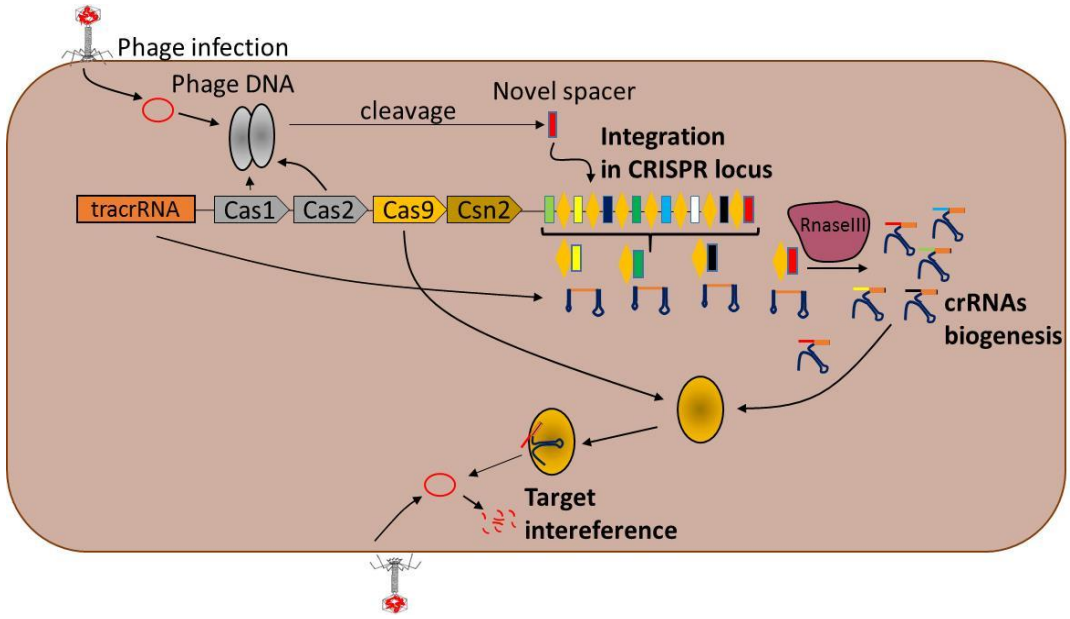


Figure 7. The three major steps in the CRISPR immunity response of the type II system: The first step is the acquisition of the novel spacer sequence from the invasive phage DNA, induced by the activity of the Cas1 and Cas2 proteins and the integration of the spacer in the CRISPR locus. The second step is the crRNA biogenesis, initiated by the interaction of the pre-crRNAs with the tracrRNA and with the help of the RNaseIII. The third step is the interaction of the mature crRNA with the Cas9 protein, that results in the immunization of the cells to infection by foreign DNA for which a spacer sequence is already present in the CRISPR

or both repeats of a pre-crRNA(Figure 7). The tracrRNA can hybridize with the pre-crRNA transcript which can be considered as the starting point of the crRNA biogenesis in Type II CRISPR/Cas systems. The RNA-specific endoribonuclease RNase III and other non-identified proteins interact with the hybrid molecule to produce the intermediate crRNA, a maturation process that also requires the presence of the Cas9 protein. The Cas9 protein is the principal and only protein responsible for the target interference capacity of the CRISPR/Cas Type II system. The resulting intermediate crRNAs composed of repeat-spacer-repeat sequences are further trimmed into short mature crRNAs consisting of unique spacer-repeat sequences in a second maturation event. The mature crRNA remains bound to the processed tracrRNA, forming a dual-RNA structure that is associated with Cas9 and activates the later. The active Cas9 can be programmed by the dual-tracrRNA:crRNA structure to cleave site specifically cognate target DNA using two distinct endonuclease domains (HNH and RuvC/RNase H-like domains)(Figure 8). The work of Jinek and his colleagues demonstrated that the Cas9 HNH domain cleaves the complementary DNA strand, whereas the Cas9 RuvC-like domain cleaves the non-complementary DNA strand (Jinek et al., 2012).

c- CRISPR as a gene regulation mechanism

Apart being a defense mechanism against foreign elements, CRISPR/Cas systems have also been identified as a system of regulation of gene expression in bacteria. In a bioinformatics analysis, Stern and his colleagues (Stern et al., 2010) noticed that 18% of the 350 organisms studied carried at least one spacer corresponding to a sequence present in their own genome. They initially believed that it is an autoimmunity event caused by accidental introduction of the self-sequence on the CRISPR locus, without any further importance for the cell. It was suggested that it could also lead to cell death. However, another study explained these spacers can be found in organisms that also carry mutations or deletions of the Cas genes or the PAM motif is absent and so autoimmunity is an unlikely explanation (Bikard et al., 2013). The best example of a CRISPR Cas system that acts as a regulator of gene expression has been identified in *Francisella novicida* (Sampson et al., 2013). *F. novicida* contains a Type II CRISPR system. As a result Cas9 interacts with the complex between the crRNA and the tracrRNA to interfere with the foreign DNA. However, in this bacterium a smaller molecule partially homologous to the tracrRNA has been identified and called small CRISPR-Cas associated RNA, or scaRNA. It consists of a degenerate CRISPR repeat sequence with an approximate 20-nucleotide region of complementarity to the tracrRNA and an 85-nucleotide region of imperfect complementarity to the messenger RNA of the FTN_1103 bacterial lipoprotein. The scaRNA is located immediately upstream of the CRISPR array. The complex including the tracrRNA, Cas9 and the scaRNA can specifically bind and cleave the mRNA of FTN_1103 of *F. novicida* (Sampson et al., 2013).

d- Universal tool for genome manipulation

Before characterizing all the different types and classes of the CRISPR, scientists already had begun to develop a way to use the CRISPR/Cas9 system as a tool to bind and cleave DNA at a specifically targeted site (O'Connell et al., 2014). The first application was published in 2012 with the work of Jinek and colleagues (Jinek et al., 2012) where they develop a chimeric guide RNA molecule (gRNA) to drive

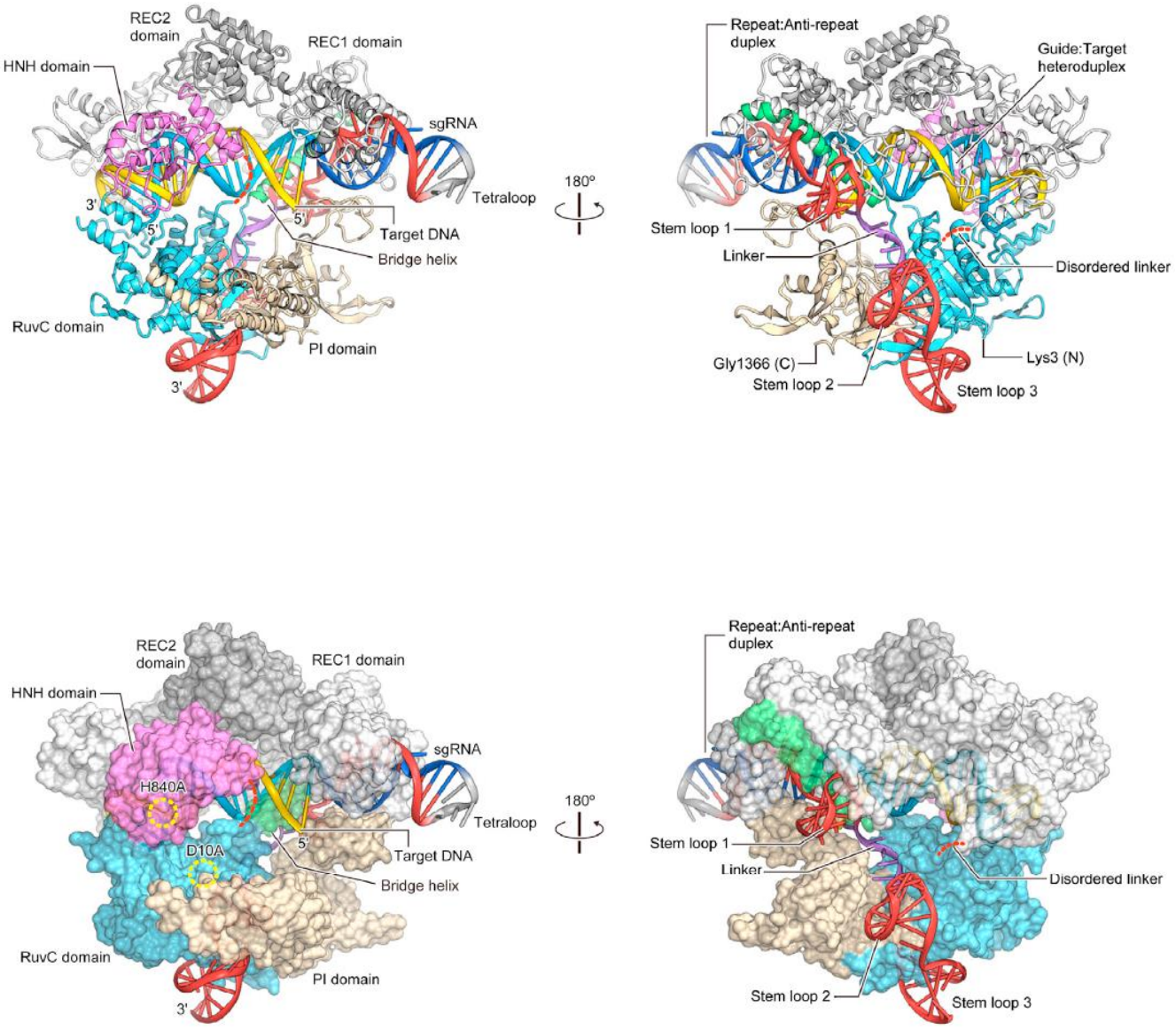


Figure 8. Ribbon and surface representation of the Cas9-sgRNA-DNA complex. Disordered linkers are shown as red dotted lines. The active sites of the RuvC (D10A) and HNH (H840A) domains are indicated by dashed yellow circles. Figure adapted from Nishimasu et al., 2014

the Cas9 protein of the Type II system of *Streptococcus pyogenes*. This gRNA was designed and constructed as a hybrid of a crRNA and the tracrRNA (Figure 9). The efficiency of the system was tested on PUC19-based plasmid or dsDNA as a substrate carrying the protospacer sequence corresponding to the spacer of the gRNA. In both cases, the chimeric gRNA managed to activate the Cas9 and cleave both substrates.

Following this key experiment, many teams started using the Cas9 of *Streptococcus pyogenes* (SpCas9) and other Cas9 proteins of different bacteria (Fonfara et al., 2014, Kim et al., 2017) to modify genomes of many different species including human cells (Zhou et al., 2016), mammals like mice (Pelletier et al., 2015) , fishes (Chang et al., 2013, Zhang et al., 2017), plants (Mao et al., 2016), insects (Highfill et al., 2017), yeast (Di Carlo et al., 2013, Jakociunas et al., 2015), protozoa (Wagner et al., 2014) bacteria (Bikard et al., 2013, Qi et al., 2013) and archaea (Li et al., 2016) and even viruses (Yuen et al., 2015). Trials to modify human embryos (tripronuclear zygotes) were first published in 2015 (Liang et al., 2015) and more recent experiments were reported showing genome modification of a large number of one-cell embryos with the gene-editing technique CRISPR (Connor et al., 2017). The basis of the application is that an interference of Cas9 with its specific target will create double strand breaks (DSBs) on the genome. The DSBs stimulate DNA repair by at least two distinct mechanisms, the Non homologous end joining (NHEJ) and the homology-directed repair (HDR), both of which are active in nearly all cell types and organisms, but at various levels (Sander et al. 2014). As a result it has been successfully applied to introduce point mutations, insertions, deletions and replacements of genes in many organisms.

However, every time a new organism is chosen for CRISPR manipulations, it is necessary to find a way to introduce Cas9, the gRNA and, if required, a recombination template to drive HDR in the cells through means of transformation, transfection or by using genetic vectors. In human cell lines, the two major routes of introduction are the viral and non-viral approaches (Mout et al., 2017). In the viral approaches, the use of adenoviruses (AVs) and adeno-associated viruses (AAVs) have been used to deliver Cas9 coupled with its gRNA directly in eukaryotic cells (Truong et al., 2015, Chew et al., 2016). Both authors needed to split Cas9 in two halves but the protein was successfully assembled once inside the cells, without losing its efficiency. For the non-viral approach in the eukaryotic liver cells, the hydrodynamic injection, were the components are introduced in a plasmid format and a single stranded-DNA by tail-vein HDI, led to an efficient reparation of 1 out of 250 mutated cells (Yin et al., 2014). As reported above for the Adenoviruses approaches, Cas9 protein can be coupled *in vitro* with its gRNA and be delivered as a Ribonucleoprotein (RNP), called Cas9/RNP. This method has been applied for localized modification in the mouse inner ear using a cationic lipid based nucleic acid transfection reagent (RNAiMAX) by Yu and his colleagues (Yu et al., 2016). In the rest of the organisms where it has been tried, like plants and bacteria, plasmid vectors have proven to be efficient for the desired applications (Mao et al., 2016, Bikard et al., 2013, Qi et al., 2013).

Another problem in the application of the CRISPR/Cas9 editing tool is the off-target effect of the protein. Obviously, this concern is particularly important for any perspective of application in humans. The efficiency but even more, the specificity must be as high as possible. It has been reported that modifications of a specific site in human genome (Cho et al., 2014, Fu et al., 2013) often results in off-

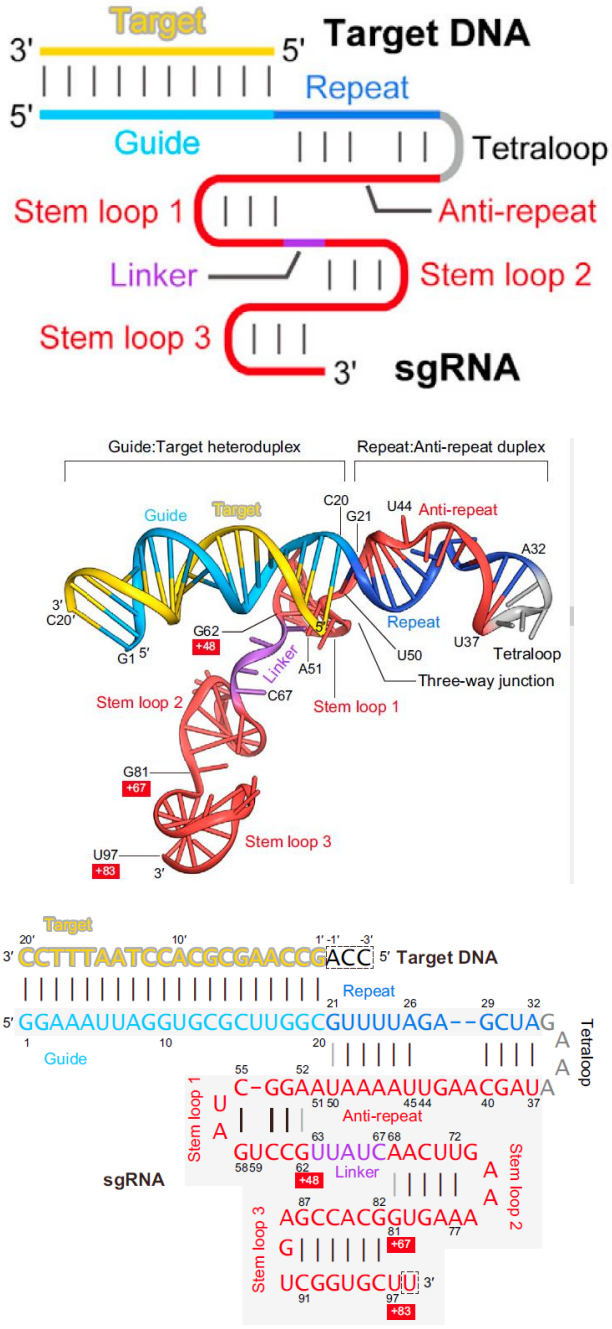


Figure 9. Schematic representation of the sgRNA:target DNA complex. At the tom there is a simplify schematic representation, then the 3D represenation and then the representation of the folding of the sequence of the complex. The tetraloop is the sequence added by Jinek and his colleagues (Jinek et al., 2012) to form the single molecule of the chimeric gRNA Figure adapted from Nishimasu et al., 2014.

target cleavages, which can lead to off-target mutations. To reduce this off-target effect, the best solution so far is the increase of the fidelity of the Cas9 by point mutation to allow efficient binding only to its corresponding target (Kleinstiver et al., 2016, Slaymaker et al., 2016).

The Cas9/RNP was also proved to be very specific due to the fact that the presence in the cell of the active system is limited in time (transient exposure). The constitutively expressed Cas9 by plasmid vectors have been proven to have a higher potential of causing off-target effects, even by a modified Cas9 that causes single strand breaks (Gao et al., 2017).

Apart from modifications of the Cas9 protein to increase its specificity and reduce the off-target effects, other interesting modifications in the CRISPR/Cas9 system have been evaluated. These include changing the PAM specificity of SpCas9 (Anders et al., 2016, Kleinstiver et al., 2015), creating a nickase variant that can cut only one of the two strands of the target sequence (Ran et al. 2014, Gao et al., 2017), modifying the repair mechanism to improve the efficiency of mutagenesis (Chu et al., 2015, Maruyama et al., 2015) and decreasing the size of the target sequence of the gRNA (Fu et al., 2014). Finally, different homologs of Cas9 were evaluated, like the one of *Streptococcus thermophilus* (Muller et al., 2016), *Campylobacter jejuni* (Kim et al., 2017). The bottom line is that the CRISPR/Cas9 system is a genome editing tool that has great potential and its efficiency can only be matched by its compatibility with almost every organism.

e- Development of CRISPR/Cas9 system as an editing tool for bacteria

The application of CRISPR/Cas systems as an editing tool in bacteria remains rather limited. In contrary to the yeast that has a very efficient HDR system, there are very few bacteria that have efficient machinery to initiate homologous recombination or to repair breaks in the DNA. Nevertheless, there have been some teams that tried to develop CRISPR/Cas9 tools for bacterial genome engineering. First, Jiang and his colleagues, applied marker free mutations in *Streptococcus pneumoniae* and *Escherichia coli* (Jiang et al., 2013). They succeeded in obtaining same levels of efficient genome editing, with a recovery of close to a 100% of edited cells in gene interruption as studies conducted in other organisms. They insisted however that while working with bacteria using the CRISPR/Cas9, it is easier to manipulate and modify organisms that are highly recombinogenic. Also, when modifying bacteria, there must be a way to introduce plasmids in them. Later, the same team developed a deactivated Cas9 (dCas9) to use it as a programmable regulator of gene expression of genes in *Streptococcus pneumoniae* and *Escherichia coli* (Bikard et al., 2013). The work of Cobb that was published sometimes later (Cobb et al., 2015) was conducted on different species of the *Streptomyces* genus, using an all-in-one method. Strains from this genus were proven to produce a number of important bioactive natural products and have a gene pool for many more that are not expressed or are poorly expressed. The CRISPR/Cas9 tools can provide a tool to develop mutants to aid natural product discovery, characterization, engineering, and production. All the elements of the CRISPR/Cas9 system, the Cas9 encoding gene, the gRNA and the reparation template were introduced in a single plasmid. Mutants with disruption or activation of the same homolog were obtained in *S. lividans, S. viridochromogenes* and *S. albus*. Another study by Wang and his colleagues (Wang et al., 2016) was conducted using CRISPR/Cas9 for genome editing in *Clostridium beijerinckii*. They used an inducible Cas9 that was used to cleave and kill all the clones on which the desired gene interruption had not occurred. The DNA editing template was introduced onto the same vector as Cas9

and gRNA sequences with the expectation that a HR event would occur through a double-crossover event. Then, the induction of Cas9, functioning as a selection tool against non-edited cells, produced a clean population of mutants. The need for an inducible system was due to the poor efficiency of the homologous recombination and the lack of the NHEJ reparation mechanism. With the same goal of overcoming the limitation of CRISPR/Cas9 usage in low recombinogenic bacteria, Reisch and his colleagues developed a no-SCAR system for genome editing in *Escherichia coli* (Reisch et al., 2015). The concept was to provide *E. coli* with a specific and highly active recombination system to increase the frequency of mutation. The λ -Red prophage assisted recombineering had already been used for insertions, deletions, and point-mutations to *E. coli* genome (Datsenko et al., 2000, Sharan et al., 2009). In the work of Datsenko, λ –Red genes *bet, exo, and gam* were expressed in an *E. coli* cell to facilitate genome integration. Reisch and his colleagues introduced these genes on a plasmid vector to facilitate the recombination with dsDNA and ssDNA they provided as templates. After a small incubation, the expression of an inducible Cas9 and a sgRNA resulted in the death of most cells without recombination event. This work demonstrated that the CRISPR/Cas9 system can be efficiently used in combination with added recombination systems, like the λ–Red system, for in vivo genome modification in bacteria. These results are particularly promising for bacteria where the genetic tools are rare or inefficient.

Objectives of thesis

The purpose of this thesis was to develop new approaches for the genome engineering of mycoplasmas, in order to reduce the time needed for a single mutation and if possible decrease the cost. In a first part, the CRISPR/Cas9 system from *S. pyogenes* was used and combined with synthetic biology methods for the engineering of mycoplasmas genomes cloned in yeast. In a second part, we studied the natural CRISPR system of *M. gallisepticum* with the ultimate goal of developing a more adapted CRISPR/Cas9 system for direct application in any transformable mollicutes.

Results

Chapter 1: Adaptation of the CRISPR/Cas9 of *Streptococcus pyogenes* for manipulation of mycoplasma genome already transformed in yeast

The Type II CRISPR/Cas9 system of *Streptococcus pyogenes* has been used to modify the genome of multiple organisms. A work of Di Carlo and his collegues (Di Carlo et al. 2013) proved that it is possible to use the CRISPR/Cas9 for genome engineering in *Saccharomyces cerevisiae* by taking advantage of the yeast's efficient homologous reparation machinery. They worked with yeasts that constitutively expressed the Cas9 of *Streptococcus pyogenes* (SpCas9). They introduced a gRNA following the same design as Jinek (Jinek et al., 2012) and targeted the canavanine negative selectable marker. Canavanine is an analog of arginine. They are both imported in the cell by high affinity permease, which is encoded by the CAN1 locus. A mutation at this locus can provide resistance to canavanine. They demonstrated that the homologous recombination rates of double stranded oligonucleotide donors increased by 130-fold after Cas9 cleavage of the targeted CAN1 locus. Moreover, when co-transforming gRNA together with a DNA template and they observed recombination rate for the DNA template up to 100%, without the need for a selection step.

Thanks to the work of Carole Lartigue and Daniel Gibson, several mycoplasma genomes have already been assembled or transformed inside yeast where they are stably maintained as centromeric plasmids. Until now, the TREC and TREC-IN methods are the most efficient to modify the genome of mycoplasmas which have been cloned in yeast (Figures 4 and 5). However, each mutation takes about a month in order to be obtained and the long oligonucleotides required make these methods rather expensive for large engineering programs. Therefore, we wanted to develop new methods to reduce the time and cost to generate mycoplasma mutants. We decided to combine the CRISPR/Cas9 tools already adapted for in-yeast genome engineering with the synthetic biology approaches first developed at JCVI.

A. Strategy

We wanted to use the same technique developed by Di Carlo and his colleagues. As such, we began by obtaining the plasmids used by this group from Addgene, called p414-TEF1p-Cas9-CYC1t and p426-SNR52p-gRNA.CAN1.Y-SUP4t (Supplementary Figure 9). The p414-TEF1p-Cas9-CYC1t is a centromeric plasmid with a CEN6/ARSH4 origin and TRP1 as a selection marker. In the plasmid p414-TEF1p-Cas9-CYC1t, the Cas9 encoding gene is a codon optimized version originally designed for expression in human cells (Mali et al., 2013). Constitutive expression is controlled by a TEF1p promoter and nuclear localization is driven by a C-terminal SV40 tag. The p426-SNR52p-gRNA is a high copy 2 μ plasmid with URA3 selection marker. Expression of the gRNA from the p426-SNR52pgRNA.CAN1.Y-SUP4t and derived plasmids is under the control of the SNR52 promoter with the SUP4 3' flanking sequence as a terminator. The original gRNA plasmid used by Di Carlo was designed to target the CAN1 locus on the yeast genome. Before going further, we decided to verify the efficiency of the CRISPR/Cas9 system in the same yeast genome CAN1 target and then we tried to apply it to the mycoplasma genome.

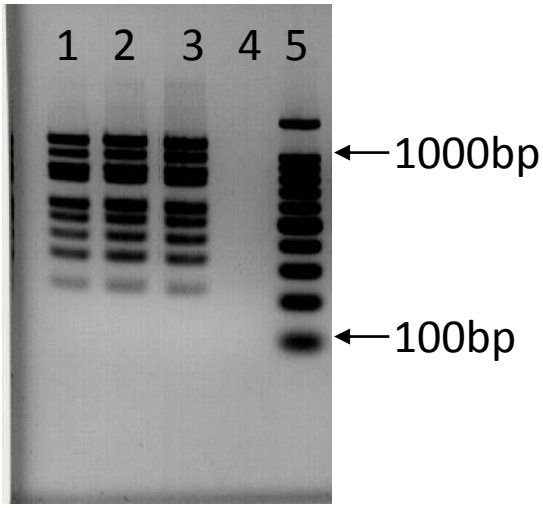


Figure 10. Multiplex PCR on clones Mmc W303a/Mmc/p414_TEF1_Cas9. 1, clone I of Mmc W303a/Mmc/p414_TEF1_Cas9; 2, clone II Mmc W303a/Mmc/p414_TEF1_Cas; 3, positive control using the purified genome of Mmc; 4, H2O; 5, 100bp marker of molecular weight Thermofisher Scientific (ref. 10787018).

B. Application of CRISPR/Cas9 for genome engineering in *Saccharomyces cerevisiae*

We wanted to test the efficiency of the tools developed by Di Carlo and his colleagues. We chose two yeast strains, W303a *(MATα his3-11, 15 trp11 leu2-3,112 ura3-1 ade2-1 can1-100)* and the strain VL6-48N *(MATα trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir)* (Larionov e t al., 1997). We first introduced the plasmid p414-TEF1p-Cas9-CYC1t inside both yeast strains, using a lithium acetate transformation: 200 ng of purified plasmid together with 50 µg of single stranded carrier DNA were used in a 40% polyethylene glycol (PEG) 3350 transformation and the yeast cells were put to grow in SD-TRP medium. The efficiency was similar for both yeast strains ($5*10^5$ yeast colonies) and we then transformed the yeasts with p426-SNR52pgRNA.CAN1.Y-SUP4t, expressing the gRNA that targets the CAN1 locus, a gene coding for the permease responsible for the influx of canavanine, a substance that is toxic for the yeast cells. We also amplified the KanMX cassette from the plasmid pFA6a-kanMX4_AJ002680_Kan_gene (Supplementary Figure 2), and added 40bp-long homologous arms to it. These arms were homologous with the sequences flanking the cutting site in the CAN1 locus. The transformation was conducted as previously described, but with 5µg of the KanMX cassette and transformants were selected on SD-TRP-URA medium. The transformation efficiencies were 7.47 10^2 cfu/µg (310 colonies/ 5µg DNA/ 50/600 Dilutions) and 4 10^2 cfu/µg (166 colonies/ 5µg DNA/ 50/600 dilutions) for the W303a and VL648N strains, respectively. We replicated the plates on SD-TRP-URA+ Canavanine (60µg/µl) and YPDA+ Canavanine + G418 (100µg/µl) using velvet cloths. The G418 is the antibiotic Geneticin, a substance for which the KanMX cassette provides resistance. Percentages of clones growing on SD-TRP-URA+ Canavanine and YPDA+Canavanine+G418 media were 98.38% and 93.38% for the W303a strain and 95.18% and 94.93% for the strain VL6-48N, respectively. We concluded that CAN1 was disrupted in nearly all transformants and thus verified the efficiency of the system developed by Di Carlo and his colleagues. We thendecided to apply the CRISPR/Cas9 tool on a mycoplasma genome cloned in yeast W303.

C. Proof of concept: Replacement of *glpO* gene with a marker in Mmc

In order to evaluate the efficiency of CRISPR/Cas9 as a genome editing tool of mycoplasma genome cloned in yeast, we decided to apply the method used by Di Carlo and his colleagues on a candidate gene of *Mycoplasma mycoides* subsp. *capri*, which genome was already cloned in yeast. The mycoplasma strain we chose was Mmc-GM12 with some genetic elements added to make it a Yeast Centromeric Plasmid: a tetracycline-resistance marker and a β-galactosidase gene for screening as well as the yeast elements ARS-CEN-HIS required for replication and selection in yeast. This yeast strain will be henceforth termed W303a/Mmc. We applied a lithium acetate transformation with 200ng of the p414-TEF1p-Cas9-CYC1t plasmid (Supplementary Figure 9) and obtained $60*10^3$ clones. A supplementary step to verify the genome integrity was required. A multiplex PCR with 9 pairs of primers (Supplementary table S2) was conducted to verify the integrity of the genome in a selection of clones. As we can see in figure 10, the clones that we selected have the same profile with the positive control, which means that all of the regions, on which the primers hybridize, are present on the mycoplasma genome. These yeast strains were stored at -80°C with 15% glycerol.

The next step was the deletion of a candidate gene of the mycoplasma genome, using the CRISPR/Cas9 system. We wanted to estimate the efficiency of different modifications; replacement of

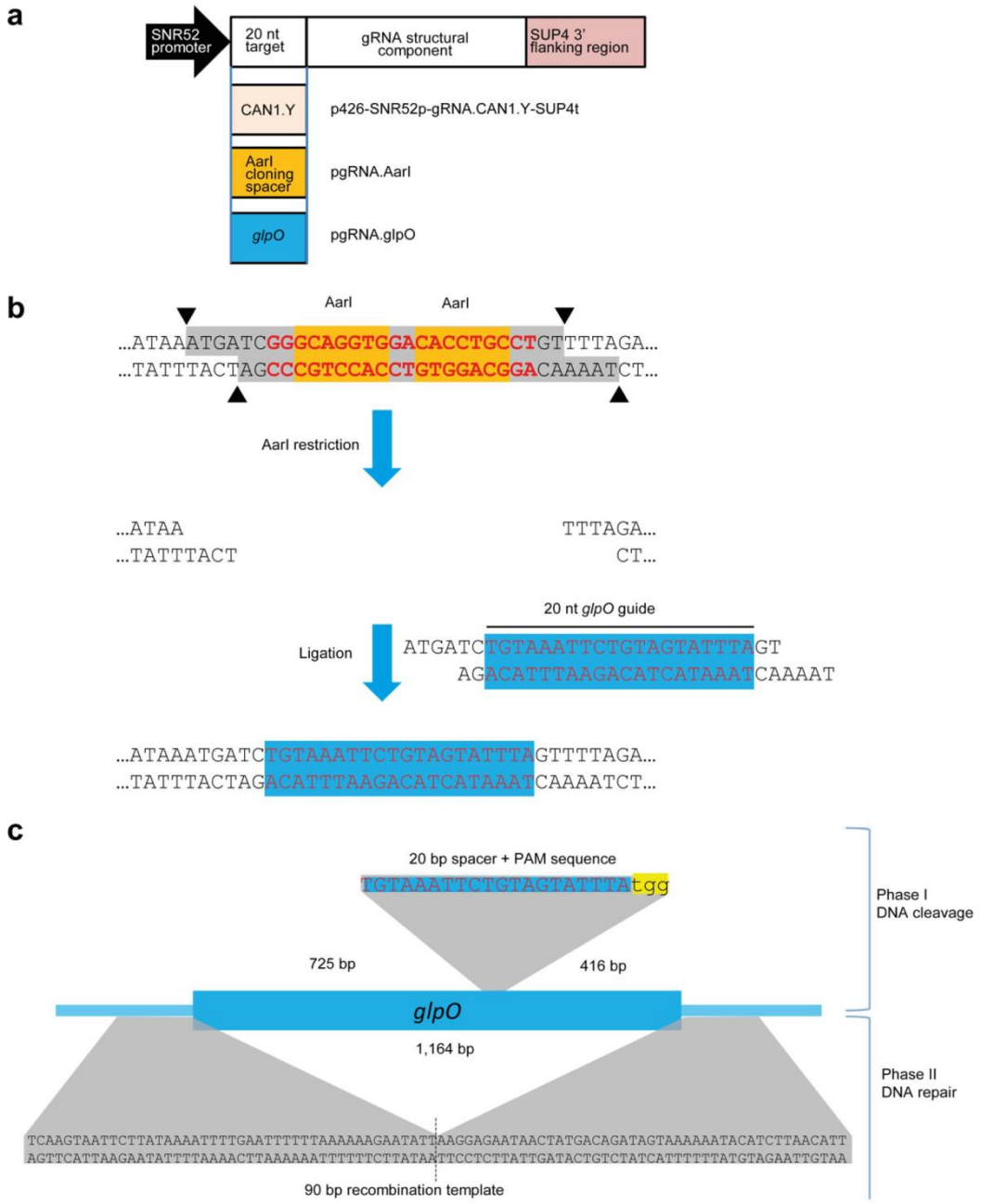


Figure 11. Construction of the gRNA expression vector and *glpO* deletion design. A. Design of gRNA expression constructs. Expression of chimeric gRNA is controlled by the snoRNA SNR52 promoter and terminator from the 3' region of the yeast SUP4 gene. The CAN1.Y target 20 nt sequence from the original plasmid from DiCarlo et al. was replaced to generate the other plasmids B. Schematic for the seamless cloning of the *glpO* guide sequence oligonucleotides into the customized p426-SNR52p-AarI-SUP4t plasmid. The type IIS AarI restriction enzyme recognition and cleavage sites are indicated in orange and by arrowheads, respectively. The *glpO* guide oligonucleotides are annealed and contain overhangs for ligation into the pair of AarI sites in pgRNA.AarI. C. Localization of the 20 nt-guide sequence within the *glpO* gene. Adjacent PAM sequence tgg is highlighted in yellow. Sequence of the 90 bp-recombination template for the deletion of the *glpO* gene in *Mmc* is shown in gray.

the candidate gene with the KanMX marker on the mycoplasma genome and also a seamless deletion of the candidate gene. For the second modification, we thought that if the team of Di Carlo was able to disrupt a gene with a small recombination template, oligonucleotides measuring only 90bp, we could try to use a recombination template of similar length to delete a complete gene. If successful, a double crossover would result in the removal of the entire sequence of the candidate gene, without leaving any trace of this modification and without the need for an intermediate step of selection (Figure 11). The candidate gene we chose was the glycerol-3-phosphate oxidase-encoding gene, also called *glpO* (or MMCAP2_0219), which is part of the operon *glpOKF*. We chose the *glpO* gene, because it encodes an enzyme implicated in the metabolic pathway of glycerol in many mycoplasmas. In addition, the oxidase encoded by *glpO* produces hydrogen peroxide which can cause damage to their hosts (Blötz et al., 2017). Thus, this gene is considered as a virulence factor, and its deletion could be desired for the development of an attenuated strain that could be used as a vaccine strain. This gene is present in one copy in the Mmc genome which should avoid problems with multiple DSBs at different genomic loci.

Finally, a phenotypic difference is expected in a mutant without the *glpO* gene; the reduction of H_2O_2 production. For all the above reasons we choose *glpO* as a proof of concept to evaluate the efficiency of CRISPR/Cas9 for gene replacement and a gene deletion. The p426-SNR52pgRNA.CAN1.Y-SUP4t was modified in order to replace the spacer sequence (originally targeting the CAN1 locus) with an easily-modifiable cloning sequence where we could introduce the spacer of interest, in order to guide the Cas9 to the desired target. So, we introduced a cloning spacer that includes two sites for the type IIS restriction enzyme AarI, in the reverse and opposite direction using the Gibson assembly method.,.

For the Gibson assembly, we first amplified each half of the plasmid pgRNA.CAN1.Y, using the primers on Supplementary table S2. Each primer had floating tails of 20 bases, in order to have a sufficient length of homologous sequences to get a good frequency of positive assembly rate. The (AarI F) primer contained also the two sites of the AarI enzyme in the desired orientation (Supplementary Table 2). Both parts were amplified through a PCR and were incubated for 2h at 37°C degrees in presence of the enzyme DpnI. This restriction enzyme recognizes Gm6A^TC sites which ensure the specific degradation of the template pgRNA.CAN1.Y remaining in the PCR mixture. After purification using the GE Healthcare DNA purification kit, 25fmol of each PCR product were incubated briefly with the endonuclease (Epicentre, T5E4111K) before an incubation at 50°C with the polymerase (Finnzymes, F-530S) and the ligase (NEB, M0208L) as it is used for the Gibson assembly (E2611S). The resulting mixtures were transformed in *E. coli* NEB® 10-beta Electro-competent cells. The resulting colonies were tested with the PvuI and PvuII restriction enzymes (data not shown) and the desired plasmid was isolated and verified by sequencing. The resulting p426-SNR52pgRNA.AarI-SUP4t plasmid now has the desired cloning sequence where any desired spacer can be easily introduced by digestion with the AarI enzyme, followed by a direct insertion of annealed oligonucleotides with compatible overhangs as described by Ran and her co-workers (Ran et al., 2013). The next step was the selection of an appropriate target inside the *glpO* gene.

Design of the 20 bp spacer was performed taking into account the following criteria established by others (Xu et al. 2015): First the presence of a NGG consensus Protospacer Adjacent Motif (PAM) immediately downstream from the spacer which is the motif required by SpCas9 to interact with the

A.

	H_2O	gRNA 200ng	gRNA+ KanMX 300ng	gRNA+ KanMX 1000ng	gRNA+ KanMX 2000ng	gRNA+ KanMX 4000ng
SD-HIS-URA	0	1384	1000	752	67	0
YPDA+ Gen 0.1mg/ml	N/A	0	30	60	7	N/A
YPDA+ Gen 0.2mg/ml	N/A	0	27	50	7	N/A
Recombination efficiency	-	-	2%	6%	10%	-

B.

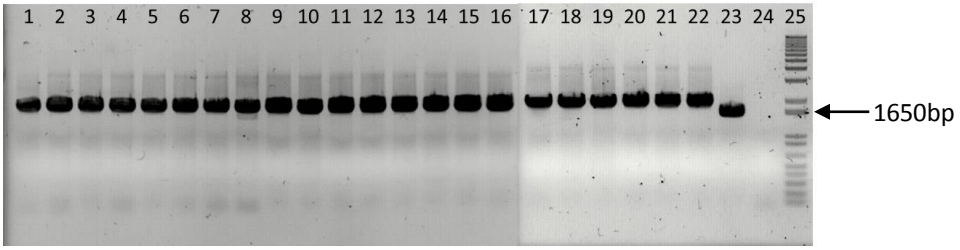


Figure 12. The KanMX assay. A. Transformation conditions and number of resulting clones. B. Clone analysis by PCR and gel electrophoresis. The size of the amplification product on the wild type genome is 300 bp smaller than the one with the replacement of the *glpO* gene by the KanMX marker (glpO size= 1160 bp , KanMX size = 1480 bp). 22 clones have been tested with primers flanking the *glpO* gene ~250 bp upstream and downstream respectively. All clones (positions 1-22) have PCR products with the expected size corresponding to the mutation. The positive control (position 23) has an amplification product ~300 bp smaller than the tested clones. In 1kbp+ marker of molecular weight Thermofisher Scientific (ref. 10787018).

spacer target sequence; (ii) G+C content between 20 and 80% (23.1%); (iii) absence of polyT (more than 4 T) that could stop transcription by type III RNA polymerase. Among these criteria, the G+C content is particularly relevant as it ranges from 23 to 40% in Mycoplasma species, with 23.9% for Mmc. The selected spacer targets the 1164 bp *glpO* gene at position 726−745. The AarI cloning spacer was then replaced by the annealed primers targeting the *glpO* gene from Mmc (MMCAP2_0219) to obtain the pgRNA.ΔglpO plasmid.

We prepared the recombination repair templates as follows: for the double stranded oligonucleotides, equimolar quantities of 90b long oligonucleotides were hybridized in the presence of a PCR buffer (Clontech kit Advantage PCR 2). The two oligonucleotides were denaturated completely with an incubation at 95°C for 5 min and then hybridized properly with a slow cooling down, following a rate of 0.1°C/sec. For the template containing the KanMX marker, we used the same template as before, the plasmid pFA6a-kanMX4_AJ002680_Kan_gene, with the addition of floating tails corresponding to the sequences flanking the *glpO* gene, 45bp from each side. The same 90bp composed the sequence of the double stranded oligonucleotides, to remove any variation in the efficiency of the Homologous Directed repair. With all the components ready, we proceeded with the modifications of the W303a-Mmc mycoplasma genome.

We began our experiments with a transformation of the KanMX marker recombination template. We used different quantities of the recombination template and observed an increase of the frequency of clones resistant to geneticin, which was in accordance with the increase in the quantity of the KanMX cassette (figure 12). The maximum efficiency we observed was 10% while using 2µg of the KanMX cassette and 200ng of the pgRNA.ΔglpO plasmid. This efficiency was satisfying enough to try the complete deletion of the candidate gene using the 90bp double strand oligonucleotides.

During the first transformations, we used 1nmol (59,4 µg) of the oligonucleotides and couldn't isolate any positive clone for the deletion of the *glpO* gene. To increase the efficiency of the process, we add an extra step in the transformation protocol we were using so far. After the incubation with the PEG, we cultivated the cells in liquid medium of SD-TRP-HIS-URA and allow the cells to grow during 48h in agitation at 30°C. After incubation, the cells were plated on SD-HIS-URA medium, in order to relax the auxotrophy conditions and allow the cells to grow more easily. The results of this transformation are summarized in figure 13: DNA extraction was performed on 12 pools of 20 colonies each followed by a PCR screening with primers located on both sides of the *glpO* gene. Eight pools tested showed a 483 bp amplification product corresponding to the expected deletion of the *glpO* gene, in addition to the 1640 bp amplification product corresponding to the wild-type genomic structure. Two positive pools were selected, and individual clones were screened to isolate mutants using the same process. A clear single 483 bp amplicon indicating the absence of the *glpO* gene was observed for 4 individual clones each within pools P7 and P8. Mixed profiles were observed for several colonies possibly due to the presence of both modified and unmodified copies of the Mmc genome in yeast cells. The seamless deletion of *glpO* was confirmed in three clones by sequencing of the PCR products, indicating that homologous recombination had occurred as expected (Figure 13).

The genome integrity for the 4 positive clones within the pools P7 and P8 was evaluated with a multiplex PCR and a Pulse Field Gel Electrophoresis. Clones P7.14 and P8.20 showed the correct profile

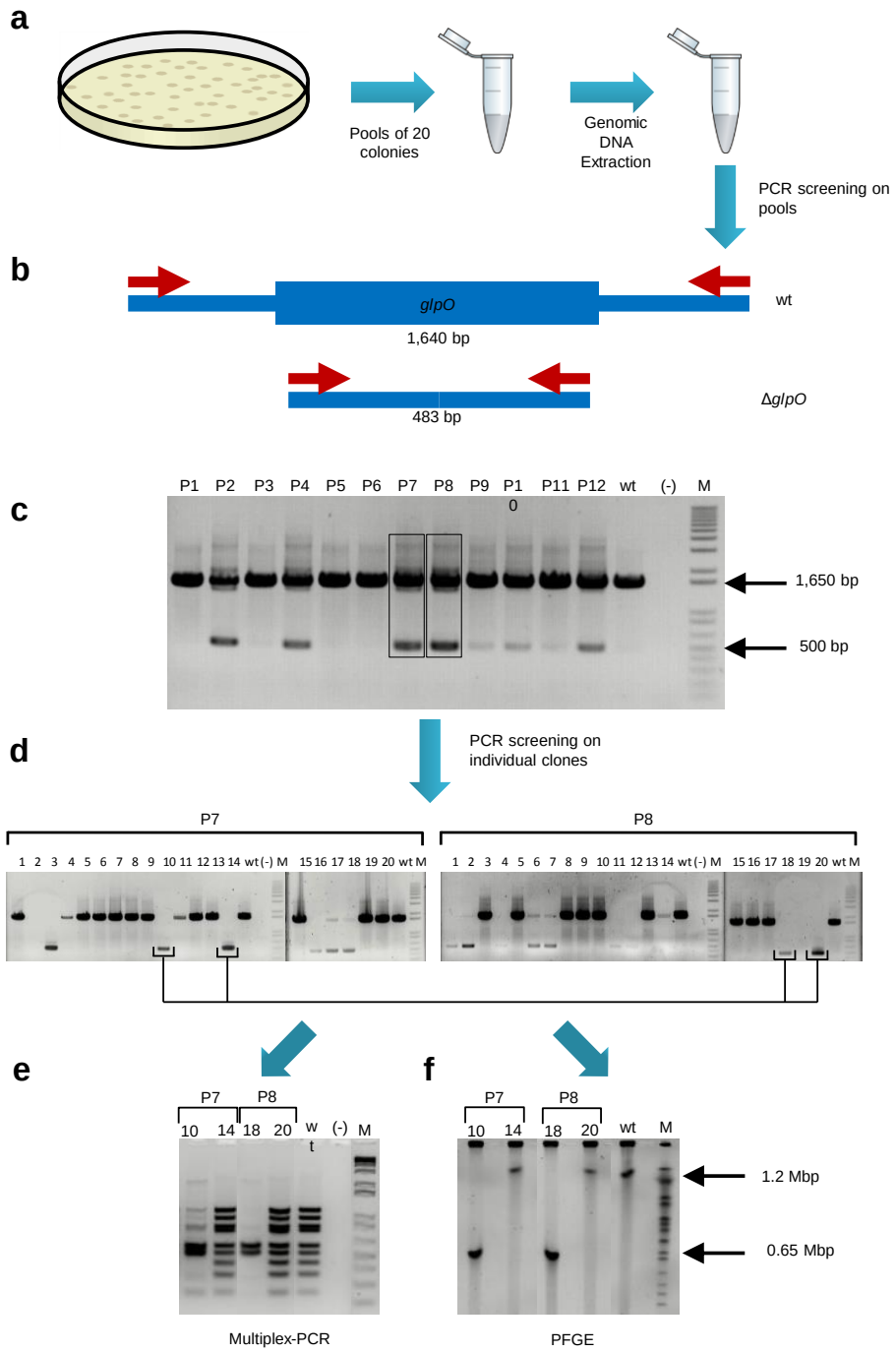


Figure 13. Screening yeast for *glpO*-deleted mycoplasma genomes. A. Genomic DNA from pools of 20 yeast colonies co-transformed with the pgRNA.glpO plasmid and recombination template was extracted for PCR screening for the *glpO* deletion. B. Schematic of the *glpO* region in Mmc (wt) and *glpO*-deleted mutants (Δ*glpO*). Lengths of PCR products are indicated. C. Representative results of the PCR screen. Pools with bands of about 500 bp indicated the presence of Δ*glpO* mutants. D. Gel electrophoresis of PCR products obtained from individual clones present in the positive pools 7 and 8. E. Gel electrophoresis of multiplex PCR to check mycoplasma genome integrity of mutants P7.10, P7.14, P8.18 and P8.20. M, 100 bp-ladder (Promega); wt, positive control DNA from Mmc; (-), H2O negative control. F. PFGE analysis of mutants P7.10, P7.14, P8.18 and P8.20 after PspXI digestion; M, CHEF *S. cerevisiae* chromosomal DNA (Biorad); wt, positive control from wt Mmc.

for both verification methods and we proceeded with the back transplantation of the genome of both clones in *M. capricolum* subsp. *capricolum* recipient cells. After incubation for 5 days on selection medium SP5-tet5, colonies were picked and 3 of them were tested for the deletion of *glpO* with a PCR reaction. The deletion of *glpO* was confirmed for all three clones. Finally, a H_2O_2 production test was conducted for the 3 mutants. The impact of *glpO* deletion on the production of H_2O_2 in the presence of glycerol was investigated as described previously (Pilo et al., 2005). Using the MQUANT^{TM} kit, concentration of 5−10 mg/L^{-1} of H_2O_2 was measured after a 100 min incubation of wt Mmc with 100µM glycerol. In contrast, no trace of H_2O_2 production was detected for the three clones where the *glpO* gene had been deleted. To verify our results, the whole procedure was repeated two times, with the yeast transformation, the pools testing and the back transplantation with the same efficiency concerning the clones in each pool screen.

D. Evaluation of the CRISPR/Cas9 tool for in-yeast genome engineering of different mycoplasma species

After the development of the CRISPR/Cas9 system for genome engineering of Mmc genome in yeast, we wanted to expand the tool to different mycoplasma species.

a. *M. capricolum* subsp*. capricolum*: precise deletions within the MCAP0015- MCS2-MCAP0017 locus

M. capricolum subsp*. capricolum* is a mycoplasma with a genome size of 1.01Mbp that is relatively close to Mmc, in terms of phylogenetic distance. Four small non-coding RNAs (ncRNAs) have been characterized for the first time by Ushida and his colleagues (Ushida et al., 1995) but there role in the mycoplasma remained unknown. Small non coding RNA2, also called MCS2, measures 92b and is encoded in a genome region located between the genes MCAP0017, encoding an ATP-dependent metallo-protease FtsH and MCAP0015 which encodes a protein with unknown function. MCS2 encoding gene was only identified and highly conserved in mycoplasma species from the Mycoides cluster (Supplementary Figure 3). *M. leachii*, Mmm, Mmc and *Mycoplasma feriruminatoris* show the same genome organization, with MCS2 encoding gene located downstream of MCAP0015 orthologs. The intergenic region MCS2-MCAP0015 is also highly conserved. An HHpred analysis on the MCAP0015 predicted protein identified a Ribonuclease H2 domain with a 92.8% probability. In the genomes of the minimal cells JCVI-syn1.0 and JCVI-syn3.0 developed at JCVI (Hutchison et al. 1999, Hutchison et al., 2016), the intergenic region between MCS2 and MCAP0015 homolog have been retained. For JCVI-syn3.0, MCAP0015 was deleted without significant impact on the cells survivability. Among the 2,200 mutants obtained by transposon mutagenesis during the project, some showed insertion in the intergenic region including MCS2, but a further analysis on the inactivation or the perturbation of the role of MCS2 was not conducted. Finally, MCAP0017 encodes an essential protein involved in division of the cell. Our idea was to apply the CRISPR/Cas9 system to create different mutants affected in MCAP0015 and/or MCS2 to try and identify the role of the small non-coding RNA.

The genome of *M. capricolum* subsp*. capricolum* had already been cloned in yeast. The transplantation method had also been validated previously for this species (Lartigue et al., 2014). We conducted three mutagenesis experiments using the CRISPR/Cas9 tools: (i) deletion of MCS2 encoding

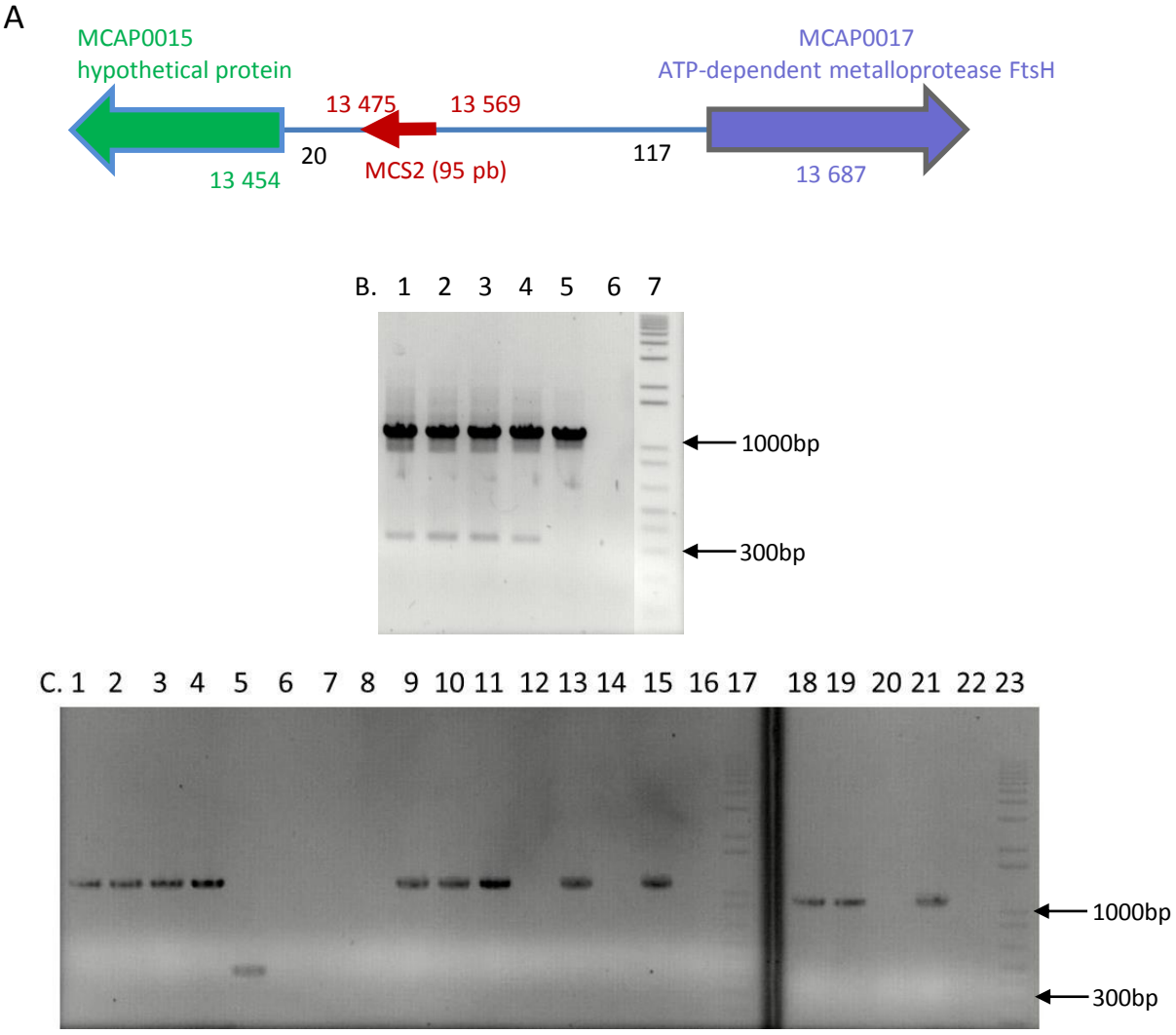


Figure 14. MCS2 encoding gene study. A. Size and orientation of the MCS2 encoding gene and its neighboring genes on the *M. capricolum* subsp. *capricolum* genome; B. PCR for detection of the deletion of the MCAP0015 gene. Pools of 20 colonies were tested for the deletion of MCAP0015; 1-4, pools with 20 colonies of yeast with the mycoplasma genome inside; 5, positive control PCR on the wt isolated *M. capricolum* genome; 6, 1kbp+ marker of molecular weight Thermofisher scientific; C. Analysis of pool 3 for the detection of the positive clone for the deletion of MCAP0015; 1-16 and 18-21, the 20 individual colonies from the pool 3; 21, H2O negative control; 17+23, 1kbp+ marker of molecular weight Thermofisher scientific

gene, (ii) deletion of MCAP0015 and (iii) deletion of both elements. Three gRNA and three 90 bp double stranded oligonucleotides were designed and produced as described above for the deletion of *glpO*. Three different transformations were conducted in yeast already containing both the *M. capricolum* subsp*. capricolum* genome and the p414-TEF1p-Cas9-CYC1t plasmid for constitutive expression of the SpCas9. Clones with the deletion of MCS2 encoding gene, MCAP0015 and both were obtained from each individual assay (Figure 14+ Supplementary Figure 4). For the construction of ΔMCAP0015 mutant, one pool among four was positive and among the individual clones, only one had the desired deletion. For the ΔMCS2 mutant, two positive pools among four were obtained and for the mutant with the double deletion of both MCS2 and MCAP0015 (ΔM+M), there was 1 positive pools among the four tested. While studying the individual clones, for the ΔMCS2 there were 6 individual colonies carrying the desired deletion, and for the ΔM+M there were two positive clones with the desired genotype. The integrity of mycoplasma genome was verified with a Multiplex PCR and a PFGE(Figure 15) for the different mutants. However, when we conducted the genome transplantation assay, transplants were obtained only for the ΔMCAP0015 mutant. The assay was repeated 3 times, giving the same results. These results suggested that the deletion of the genome region encoding MCS2 was lethal to the cell.

Due to the fact that we couldn't obtain a positive clone during the transplantation of the ΔMCS2, we decided to study more deeply the genomic region in which the MCS2 is located. We hypothesized that the non-viability of the ΔMCS2 mutant might be the result of a negative effect on the expression of the downstream gene, MCAP0017 that is known to be an essential gene. We conducted a RACE PCR following the Clontech kit protocol, in order to identify the promoter sequence for the MCAP0017. Surprisingly, this experiment showed that the promoter of MCAP0017 overlapped MCS2 on 18 nucleotides (Figure 16). Therefore, we concluded that the lethal effect of MCS2 deletion could be produced by an alteration of MCAP0017 promoter. Further analyses will be needed to decipher the functional relationship between MCS2 ncRNA and the neighboring genes MCAP0015 and MCAP0017. However, our results shows that the CRISPR/Cas9 system can be efficiently used to delete genome regions of a second mycoplasma species with a surgical precision.

b. *M. pneumoniae*: deletion of a virulence factor (MPN142)

M. pneumoniae is a human pathogen with a genome size of 0.816 Mpb with an unusual G+C content of 40%. Its genome has been successfully cloned inside yeast cells (Ruiz et al, unpublished) and adaptation of genome transplantation methods is currently in progress in the laboratory. We choose to evaluated the efficiency of the CRISPR/Cas9 tools because this species is phylogenetically remote from Mmc and *M. capricolum* and the genome G+C% is nearly 15% higher. Moreover, this species is of particular interest to our group, in the frame of the MiniCell and Mycosynvac projects that require efficient tools for M. pneumoniae genome engineering. We decided to delete the MPN142 gene, which encodes a precursor of two adhesion proteins, P40 and P90 that are important for adhesion of the mycoplasma to the lung epithelium (Widjaja et al., 2015). We used the same tools as before, but this time we tried to transform all the CRISPR elements directly in yeast strains W303a harboring the genome of the wild type *M. pneumoniae* M129 . We decided this "all-in-one" transformation as a way to reduce the risks of genome instability, which has been shown to occur for the genome of *M. pneumoniae* after a number of generations in yeast (Ruiz et al. unpublished results). We developed a gRNA using the same

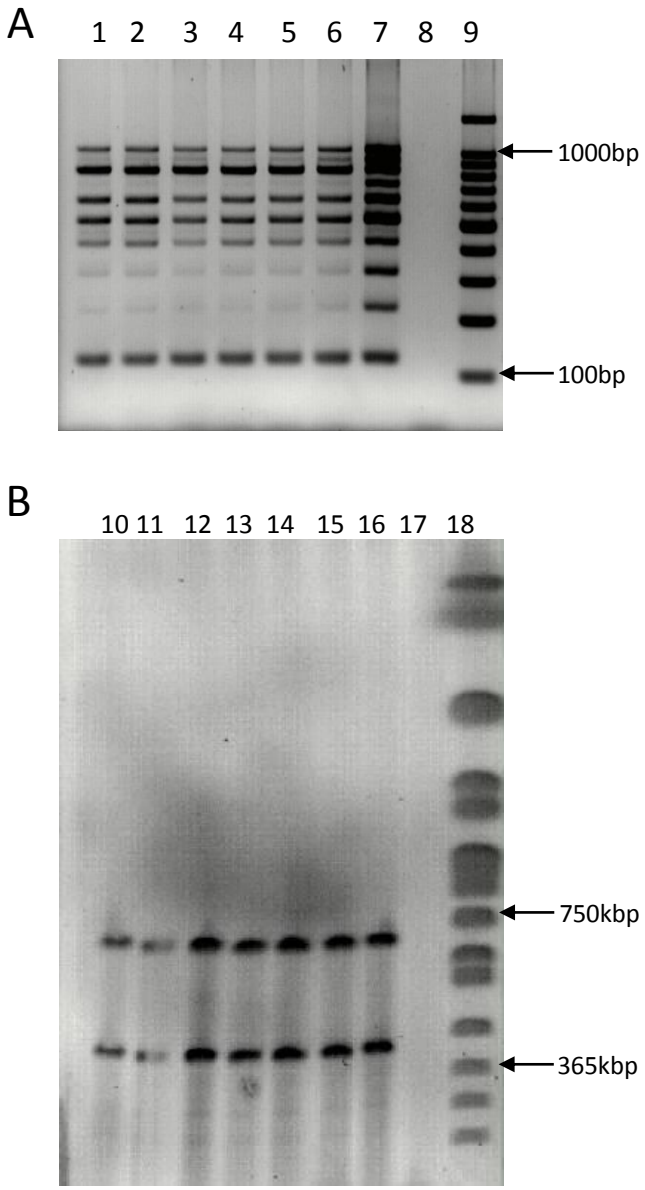


Figure 15. (A) Gel electrophoresis of multiplex PCR and (B) PFGE analysis to check mycoplasma genome integrity of mutants ΔMCAP0015 cl.4, ΔMCAP0015 cl.5, ΔMCS2 cl.4, ΔMCS2 cl.8, ΔMCS+ΔMCAP0015 cl.11, ΔMCS+ΔMCAP0015 cl.14. A. 1-6, Mutants tested in multiplex PCR; 7, positive control DNA from Mmc; 8, H2O negative control; 9, 100 bp-ladder (Promega). B. 10-15, Mutants tested in PFGE, all samples are digested with PspXI; 16, positive control DNA from *M. capricolum* ; 17, yeast DNA negative control; 18, molecular marker CHEF *S. cerevisiae* chromosomal DNA (Biorad).

protocol as before, by cloning annealed oligonucleotides into the linearized p426-SNR52pgRNA.AarI-SUP4t. For the transformation, 200 ng of each plasmid (p426-SNR52pgRNA.AarI-SUP4t and p414-TEF1p-Cas9-CYC1t) together with 1 nmole of a 90 bp oligonucleotides homologous to the flanking sequences of the MPN142 gene were used. For the wt *M. pneumoniae* M129 strain, mutants were obtained with a low efficiency (2/40 colonies) but the profile of the pools analyzed was different than before: usually when testing by PCR the pools of individual clones, the positive pools showed a mixte profile, with both the deleted and the non-deleted typical bands. In the case of *M. pneumoniae*, observed profiles all corresponded only to the deleted genotype (Figure 17). When we further analyzed individual clones, we observed that there were only one PCR that gave an amplification product. For the remaining 20 individual clones, no amplification product could be observed. For the two positive clones, the genome integrity was verified with a multiplex PCR. One of two, clone 10, presented the expected profile (Figure 18). Clone 25 was lacking a lot of bands in the multiplex PCR, and as such, no PFGE assay was conducted on the genome of this clone. Even with a lower efficiency, these results demonstrated the CRISPR/Cas9 tool could also be used for in-yeast engineering of *M. pneumoniae* genome. However, further experiments will be required to understand if the low efficiency observed was the result of a global problem of genome stability.

Discussion

Until recently, the ways to modify a mycoplasma genome have been limited to the use of transposon-based mutagenesis and replicative oriC plasmids. During recent years, the work of conducted at the JCVI provided new methods for in-yeast genome engineering after cloning of the bacterial chromosome as a centromeric plasmid. The TREC and TREC-IN tools were developed to modify the mycoplasma genome with two consecutives transformations and two auxotrophy selections. The back transplantation into mycoplasma recipient cells completes the process that allows the interruption/deletions of candidate genes or operons. We started studying the CRISPR/Cas9 tool hoping that it would allow us to decrease even more the time needed for a genome modification and improve the efficiency of the process. As a proof of concept, we produced a mutant of Mmc with a seamless deletion of a candidate gene using the CRISPR/Cas9 tool. The efficiency was sufficient for an easy selection of positive clones, and the mutagenesis of mycoplasma genome became faster, as the durations of the experiences inside the yeast were cut in half. Normally a mutagenesis experiment using TREC or TREC-IN methods, without considering the time needed for the reagents to arrive to the laboratory, takes a month in order to develop the recombination cassettes and conduct the two transformations with the selection that follows. Using the CRISPR/Cas9 system, developing the different gRNA from the linearized p426-SNR52pgRNA.AarI-SUP4t and using as recombination template double stranded hybridized oligonucleotides, only a single transformation with the gRNA and the oligonucleotides was needed and the results were available within one week. Verifying the mycoplasma genome integrity extends the duration of the experiment to two weeks. This acceleration of the process is mostly caused by the high efficiency of the CRISPR/Cas9 system combined with the highly efficient HDR in yeast. This combination can be used directly without the need of a selection marker, leading to a one step seamless mutagenesis. The major problem in almost all applications of the CRISPR/Cas9 tool is the off-targets breaks, in regions that resemble to the spacer sequence. In our case, multiplex PCR and Pulse Field Gel Electrophoresis are used to verify the global size of the genome. In most cases global genome

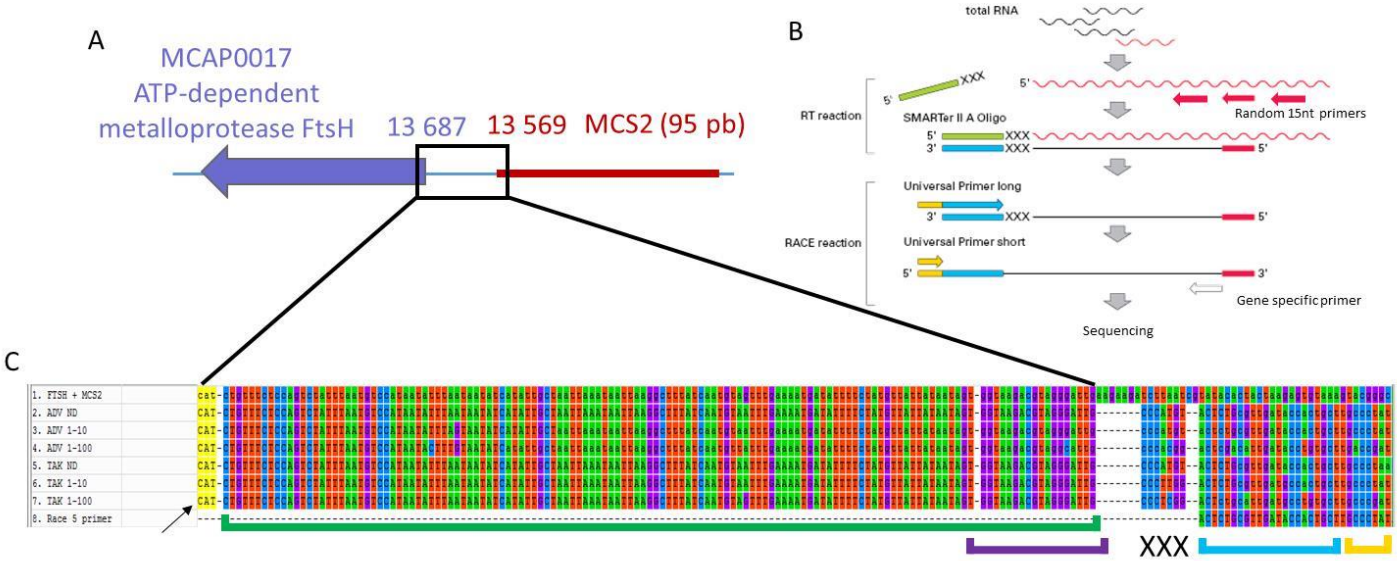


Figure 16. Overlap of *ftsH* promoter with MCS2 encoding gene sequence identified after a 5' Race PCR. A. Location of *ftsh* promoter. B. Race reaction with 15 nt random primers to amplify the 5' regions of all RNAs from *M. capricolum*. C. Sequencing results presented in MEGA5 software; The arrow indicates the start codon of *ftsH*; In green is the 5' Race product of *ftsH* promoter with different dilution of MCAP cDNA and different PCR kits (Advantage and Takara); In purple is the part of the *ftsH* 5' Race product that overlaps with the beginning of MCS2 sequence, deleted in ΔMSC2 mutant; the XXX, blue and yellow sequences indicates the SMARTer II A oligo sequences, indicating the end of the 5' Race PCR for all reactions.

integrity was confirmed except for *M. pneumoniae* which seems to be less stable. Whole genome sequence should be done to document potential small scale off-target effect such as SNP. However, the fact that the transplanted mycoplasma genomes lead to viable cells strongly suggests that the system didn't modify any important elements, due to off-targets cleavages. The result of this work adds an extra tool to our repertoire of genome editing tools for mycoplasma and allowed further applications, like large scale deletions and a new protocol adapted by our laboratory to clone genomes in yeast (Ruiz et al, unpublished results). Using the method we developed, deletions of a 20Kb sequence was obtained in Mmc and an in-situ tagging by addition of specific tags or fusion with fluorescent proteins is currently being developed in the laboratory (Arfi et al, unpublished results).

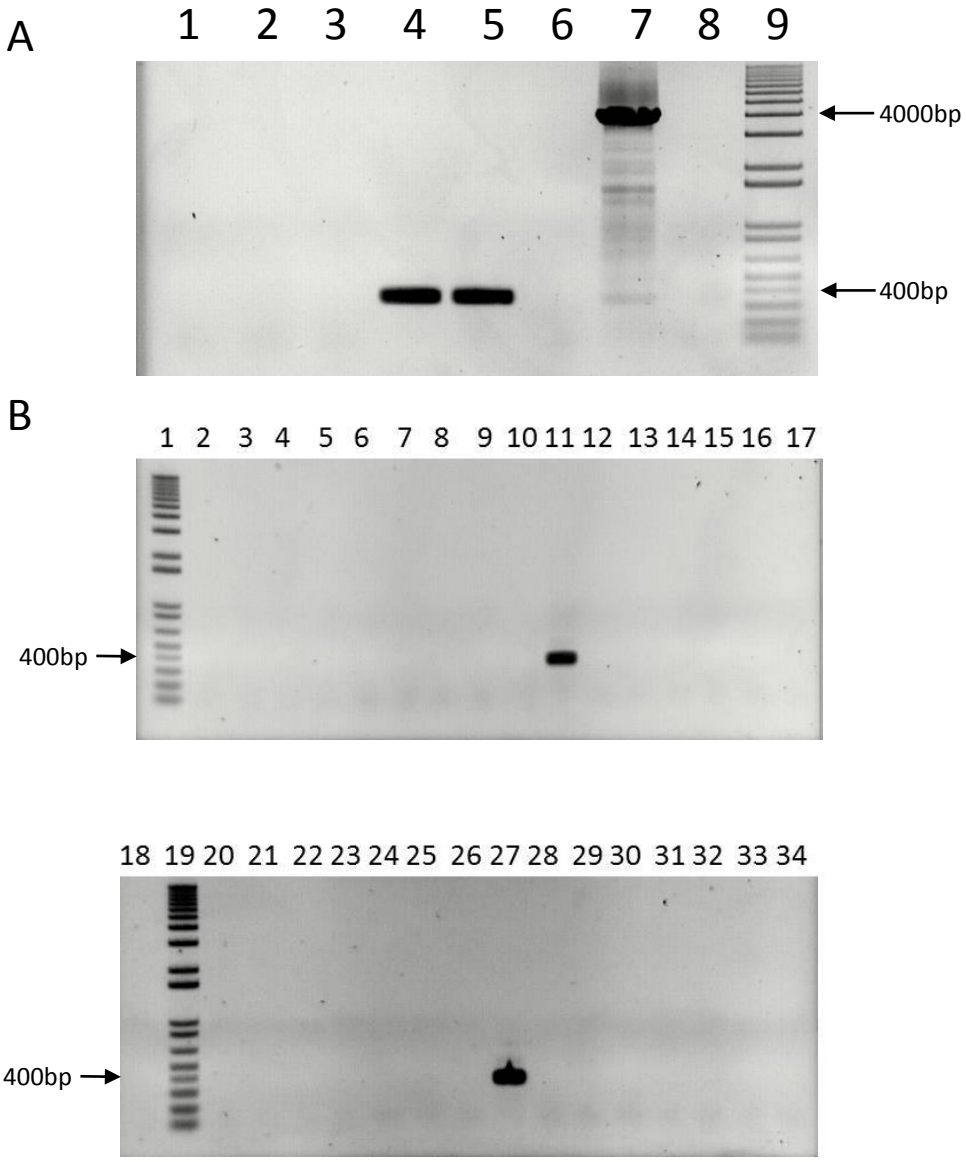


Figure 17. Screening yeast clones with deletion of MPN142. A. PCR analysis of yeast pools transformed for the deletion of MPN142. The bands of about 350 bp indicate a ΔMPN142 locus; 1-6, pools of 20 colonies of transformated yeasts for the MPN142 deletion; 7, positive control DNA from Mcap; 8, negative control H2O; 9, 1Kb+ bp-ladder (Thermo) B. Gel electrophoresis of PCR products obtained from 20 individual clones present in the positive pools 4 and 5 of ΔMPN142; 2-18 and 20-21, individual clones of pool 4; 22-34, individual clones of pool 5, all the remaining clones from pool 5 negative (empty gel, data not shown); 1 and 19, 1Kb+ bp-ladder (Thermo).

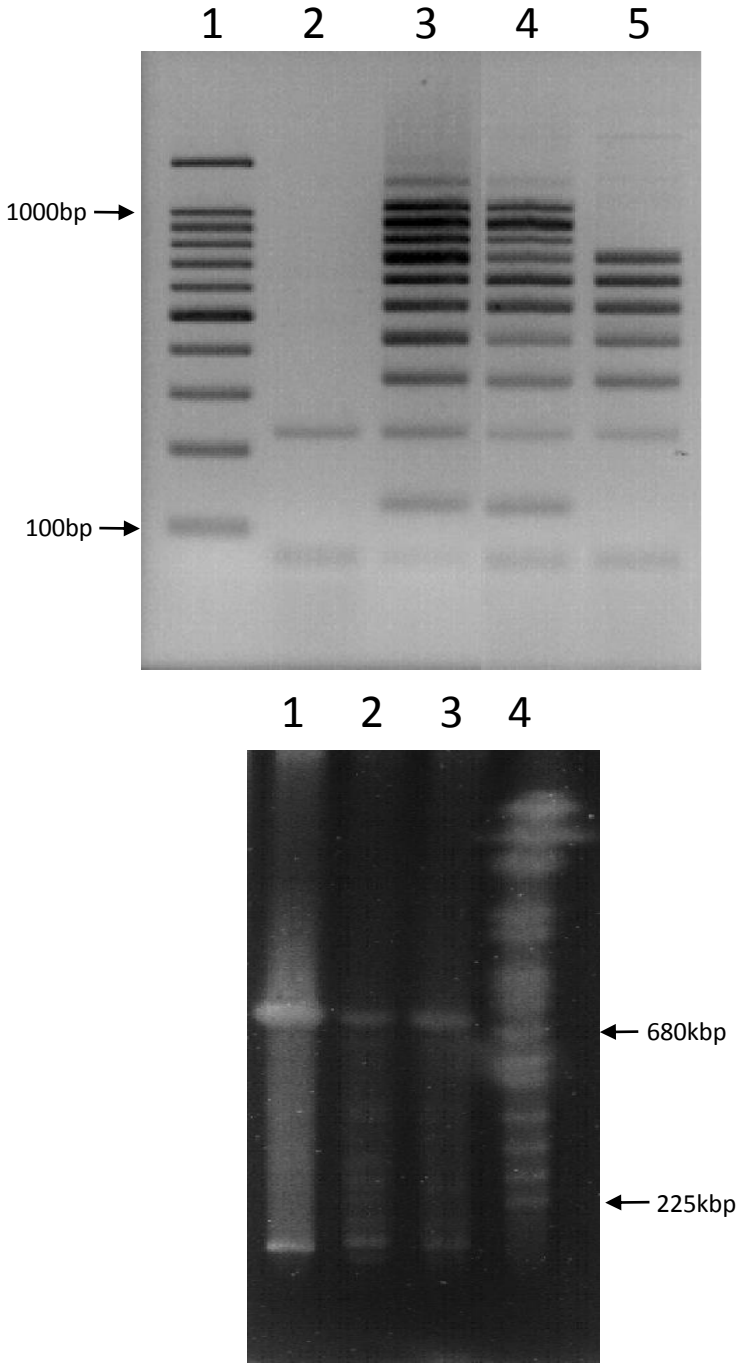


Figure 18. Gel electrophoresis of multiplex PCR and PFGE analysis to check mycoplasma genome integrity of mutant ΔMPN142 cl.10 and cl 25. A. Mutants tested in multiplex PCR;1, 100 bp-ladder (Promega); 2,; H2O negative control; 3; positive control DNA from *M. pneumoniae* 4 and 5, clones 10 and 25 respectively; B. Mutants tested in PFGE, all samples are digested with NotI-HF; 1 positive control DNA from *M. pneumoniae*; 2, positive control, genome from yeast W303a with the genome of *M. pneumoniae* inside; 3, clone 10; 4, molecular marker CHEF *S. cerevisiae* chromosomal DNA (Biorad).

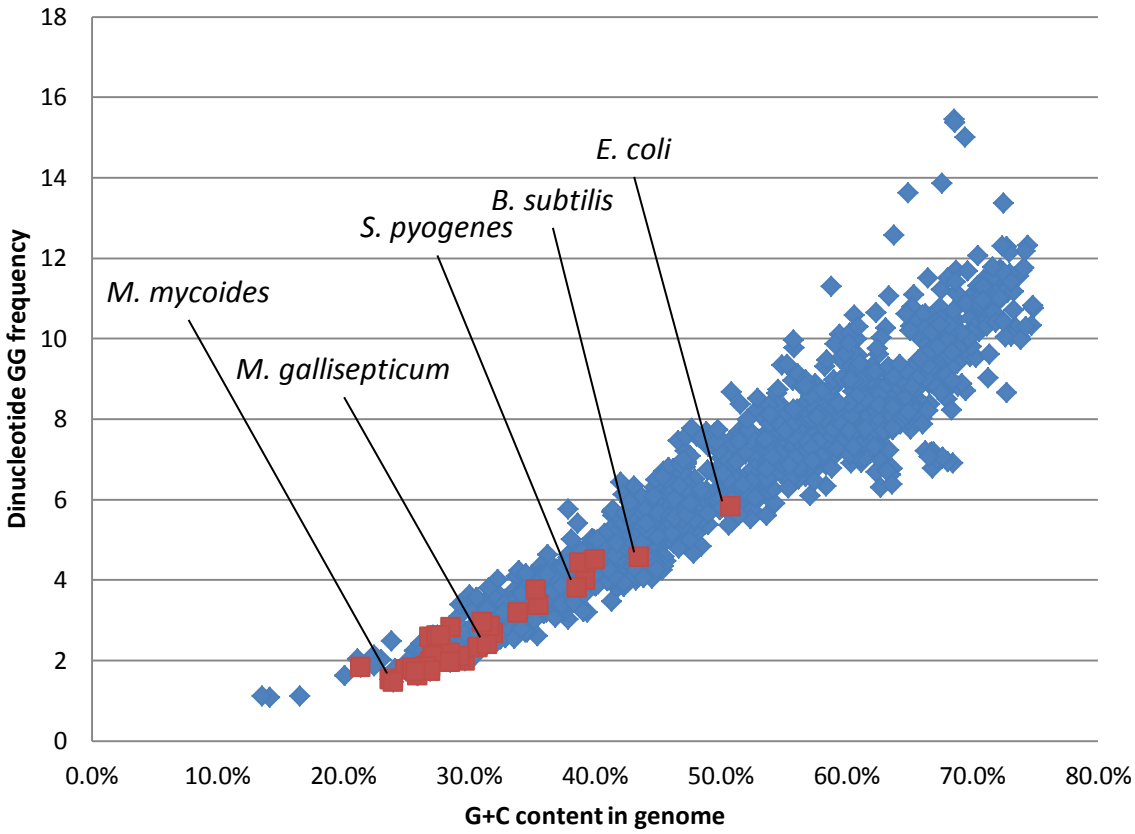


Figure 19. Dinucleotide GG frequency in prokaryots. Dinucleotide GG frequency was expressed in function of genome G+C content for a representative set of 1309 bacteria and 133 archaea. Raw data were retrieved from Zhang et al 2013. Each species is represented by a square. Red squares, mollicutes and reference bacteria.

Chapter 2: Characterization of the native CRISPR/Cas9 system of *M. gallisepticum*

As mentioned before, the available tools for engineering of mollicutes genome are few and available only for some species of the whole class. In particular, efficient synthetic biology methods involving genome transplantation are still restricted to species from the mycoides cluster and *Me. florum*. Therefore, we wanted to develop a tool for directed mutagenesis of mycoplasma genome more efficient and broadly applicable among mollicutes. The CRISPR/Cas9 system from *S. pyogenes* is currently being used in an ever-growing number of organisms to target specific sequences and creates double strand breaks. Cellular repair mechanisms, namely NHEJ or HDR are then activated to repair this DNA damages which can be an opportunity to produce mutants by introduction of small deletions (NHEJ) or recombination with available template provided extracellularly. However, the intensively studied SpCas9 that is used in most organisms might not be the most adapted tool in mycoplasmas. The main reason is that most mollicutes have genomes with a low or very low G+C content, with a median value of 27.8%, making these genomes some of most biased in the living world. By contrast, the PAM sequence of *S. pyogenes* CRISPR system is a G+C rich motif that has been characterized as NGG. Such motif can be statistically found at various frequencies among genomes with a direct correlation with global G+C content (Figure 19). Consequently, most mollicutes present a low GG dinucleotide frequency compared to more common bacteria. For example, in a typical prokaryotic gene of 1 kbp, 29 dinucleotides GG are statistically found in Mmc compared to 91 in *B. subtilis* and 116 in *E. coli*.

Thus, in order to develop the most adapted CRISPR tool for mycoplasmas, we decided to characterize an endogenous CRISPR system naturally present in these bacteria.

Comparative genomics of CRISPR in mollicutes

CRISPR systems have been identified in several mollicutes during genome sequencing, but no general survey was available at the time of this work. In order to choose the CRISPR/Cas9 tool that would be more adapted for mycoplasma genome editing, we first performed a comparative genomic study of CRISPR systems among the entire *Mollicutes* class.

1. Distribution of CRISPR system in mollicutes

CRISPR systems were searched in mollicutes genomes using (i) Blastp search of Cas genes, (ii) analysis of direct repeats using CRISPR database (http://crispr.i2bc.paris-saclay.fr/crispr/) and the included CRISPR finder tool. A manual analysis of all candidates was then achieved to get a precise annotation of the loci.

Complete or degraded Type II CRISPR systems were detected in 21/52 complete or draft genomes of species representative of the class *Mollicutes* (Figure 20). CRISPR systems including, in the following order *cas9,* tracrRNA, *cas1, cas2,* and a CRISPR track, were found in most species but some inversions were observed for several species such as *M. dispar*, *M. ovipneumoniae, M. hyosynoviae, M. arginini* and *M. arthritidis*.

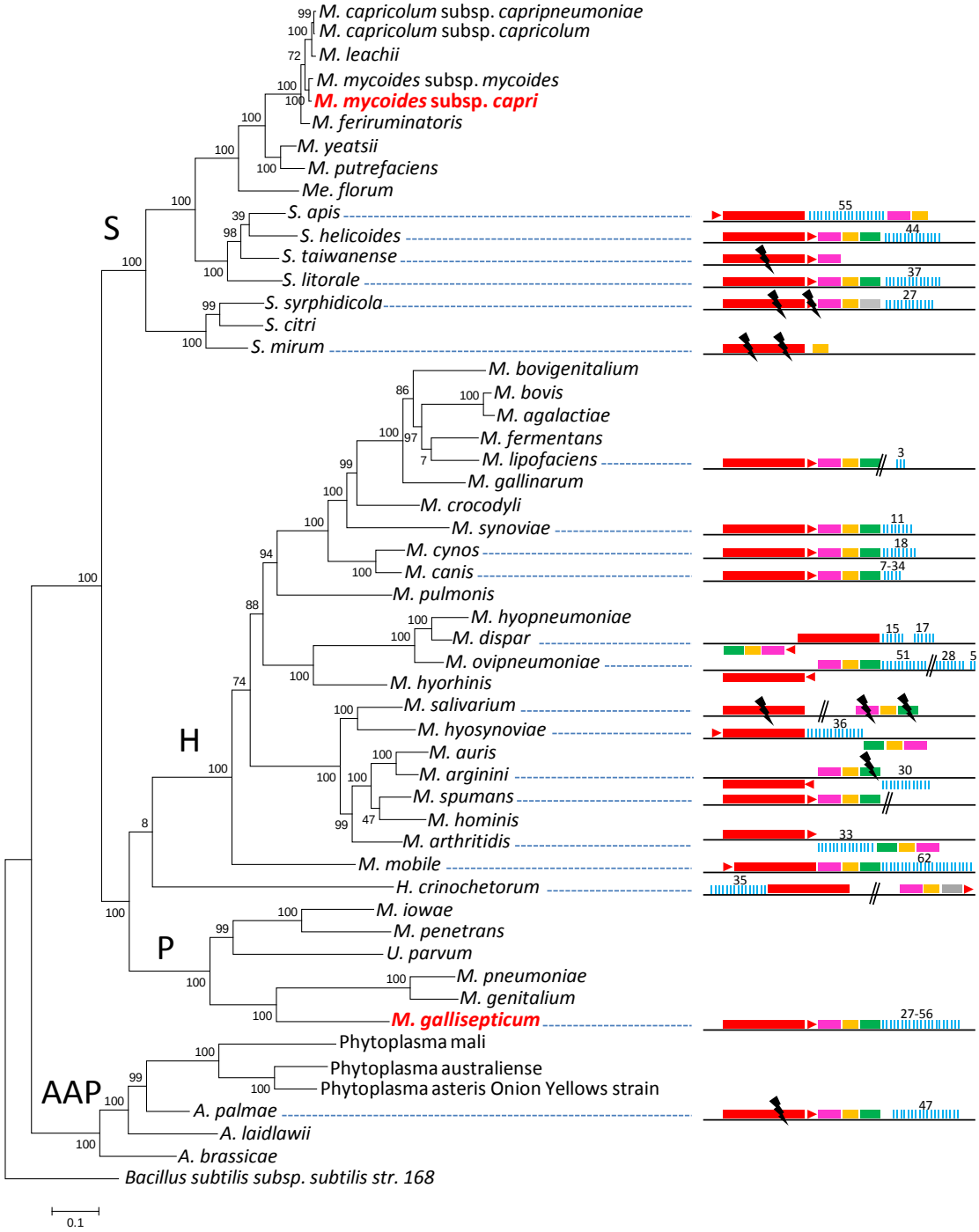


Figure 20. Distribution of CRISPR systems in mollicutes. Organization of the CRISPR systems predicted in mollicutes genomes are represented on the right part. Red rectangle, Cas9; red triangle tracrRNA, pink rectangle, Cas1; orange rectangle, Cas2; green rectangle, Csn2; blue bars, CRISPR locus; grey rectangle, CDS not related to CRISPR. Above number indicates the number of spacers. Double bars, genome interruption; black flash, disrupted gene. The phylogenetic tree was generated using the maximum likelihood method from the concatenated multiple sequence alignments of selected 50 orthologous protein involved in translation. Main phylogenetic groups are indicated, S, Spiroplasma, H, Hominis, P, Pneumoniae, AAP, Acholeplasma/Phytoplasma. *B. subtilis* was used as an outgroup. Statistical values from an Approximate Likelihood-Ratio Test are indicated on branches.

Overall picture is that CRISPR systems are widespread in most of the main phylogenetic groups with a more frequent occurrence in the Hominis group. No other type of CRISPR system was predicted.

In the Spiroplasma clade, putatively complete CRISPR systems including *cas9, cas1, cas2, csn2*, a tracrRNA and a CRISPR track were found in *S. helicoides* and *S. litorale* whereas more or less degraded forms were found in *S. taiwanense*, *S. syphridicola* and *S. mirum*. In *S. apis*, all the above mentioned genetic elements were predicted complete except *csn2* that was remained undetected. No CRIPSR system could be detected in *S. citri* and other closely related species, which was correlated with a high density of viral sequences in the genome (Ku et al., 2013). Within the Spiroplasma branch, the mesoplasma-mycoplasma mycoides phylum was characterized by a total absence of CRISPR, despite more than 35 genomes are now available.

By contrast, CRISPR systems were found widespread overall the Hominis clade, in the genomes of species infecting a variety of animal hosts, with complete structures predicted in *M. synoviae*, *M. cynos*, *M. canis, M. dispar, M. ovipneumoniae, M. hyosynoviae, M. mobile* and uncomplete structures in *M. lipofaciens, M. salivarium, M. spumans, M. arthritidis* as well as the outgroup mollicutes *Hepatoplasma crinochetorum.*

In the Pneumoniae clade, CRISPR systems was only found in the bird pathogen *M. gallisepticum*. For this species, complete or incomplete forms of CRISPR systems have been previously characterized in the 12 genomes available and their evolution was associated with adaptation to the bird host (see below).

In the AAP branch, a degraded CRISPR system was found in *A. palmae* but not in other acholeplasma genomes available. CRISPR were not found in the branch of plant pathogen phytoplasmas.

2. Phylogenomics of mollicutes CRISPR systems

Previous phylogenomic studies on the diversity of bacterial CRISPR systems including some mycoplasma systems have classified them as Type II CRISPRs, with all representants studied gathering in a specific branch of subtype II-A (Fonfara et al., 2013). In order to get a more complete picture, we performed some similar phylogenomics focusing on mollicutes CRISPR systems.

Amino-acid sequence of Cas9 from mollicutes were aligned together with a set of reference Cas9 proteins from subtypes II-A, II-B and II-C, as defined by Fonfara et al 2013. A phylogenetic tree was then infered from the multiple alignement (Figure 20).

Proteins Cas9 from nearly all mycoplasmas were clustered in a statistically highly supported branch (aLRT value, 92%). This suggested a common origin of all CRISPR systems currently described in mycoplasma species. Interestingly, Cas9 from *M. gallisepticum* was found in a 100% supported subgroup including *M. synoviae, M. cynos* and *M. canis*. While those last three species are phylogenetically closely related in the Hominis group (see Figure 21), *M. gallisepticum* is a remote species from the Pneumoniae group. Similar phylogenetic association was found with Cas1 protein (not shown). This suggests that CRISPR system from *M. gallisepticum* may have been transferred by HGT from *M. synoviae* or a closely

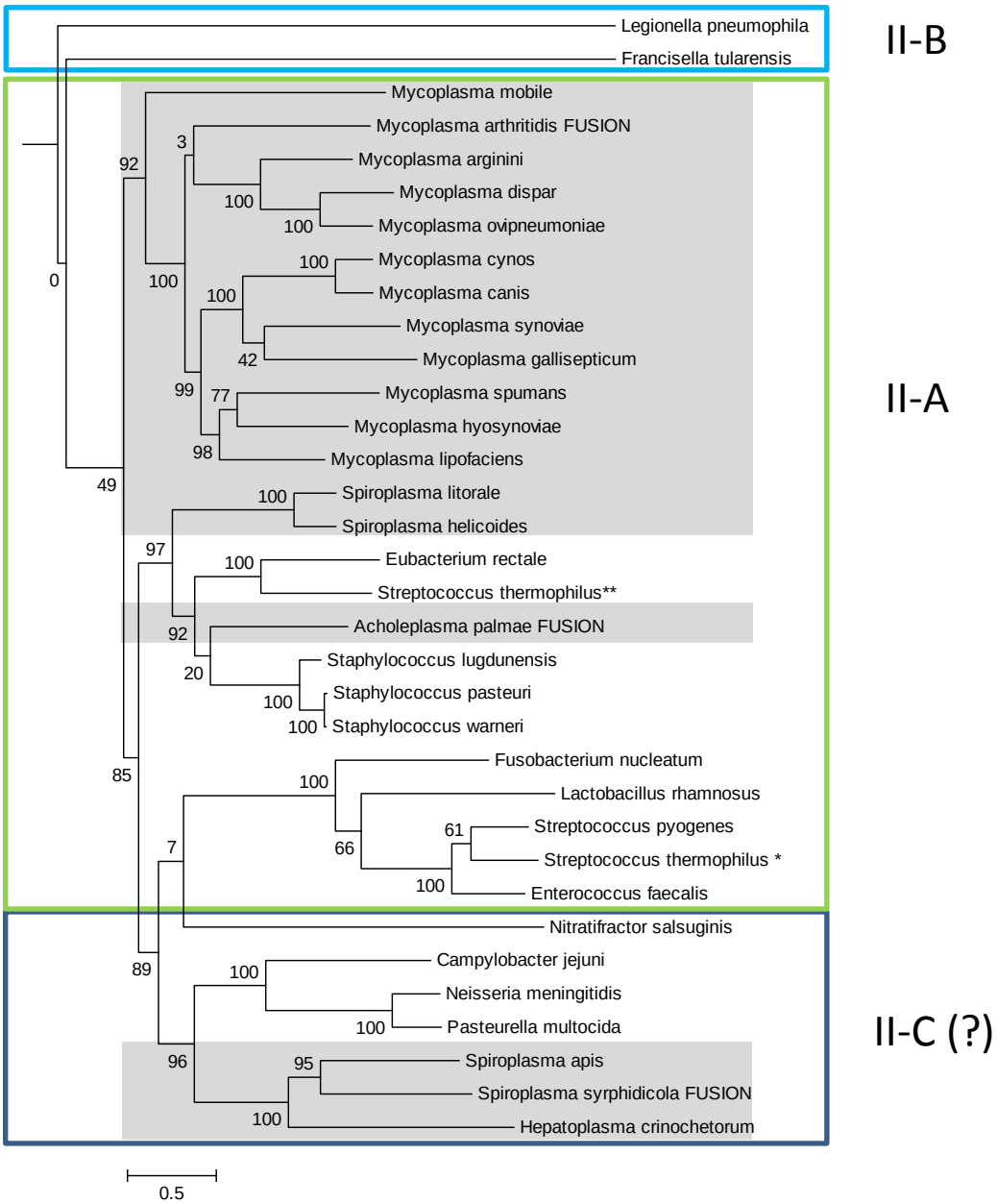


Figure 21. Phylogeny of Cas9 orthologs in mollicutes and reference bacteria. Amino-acid sequences of Cas9 proteins were aligned with MUSCLE and phylogenetic tree was reconstructed with PhyML with tools available on phylogeny.fr. For *M. arthritidis*, *A. palmae* and *S. syrphidicola*, Cas9 protein sequence was artificially simulated from fusions of ORFs covering the disrupted gene. Cas9 from reference bacteria were chosen from Fonfara et al 2013; proposed subtypes were also defined according to this work. Cas9 from mollicutes are highlighted in grey.

related species. Such scenario is in accordance with other studies that have predicted genetic exchanges among these bird pathogens (Vasconcelos et al., 2005).

By contrast, Cas9 from other mollicutes were found more widely distributed in the phylogenetic tree. Cas9 from *A. palmae* was found in a branch gathering Cas9 from several reference Gram positive bacteria (staphylococci, *streptococcus thermophilus, Eubacterium rectale*) which is consistent with their common ancestral origin.

Surprinsingly, Cas9 orthologs from Spiroplasmas and Hepatoplasma were distributed in two remote subgroups with no correlation with their relative phylogenetic position. Indeed, Cas9 from *S. apis* was found closely related to that of *S. syrphidicola* whereas those two spiroplasmas belong to two clearly disctinct phylogenetic subgroups. By contrast, Cas9 from *S. helicoides* appeared remote from that of *S. apis* while those two species are very closely related. These observations suggested different origins for the CRISPR systems found among spiroplasmas. In addition, we noticed that no trace of csn2 gene could be predicted in the genomes of *S. apis, S. syrphidicola* and *H. crinochetorum*. Interestingly, Fonfara et al have proposed a fine classification of Type II CRISPR systems with a subtype II-C characterized by the lack of *csn2* or *cas4* that are found in subtypes II-A and II-B, respectively. This suggested that the CRISPR systems of these three mollicutes might be evolutionary related to subtype II-C systems by contrast to all other CRISPR systems from mollicutes for which a *csn2* gene has been predicted (Figure 21).

An analysis using the MEME MAST software (Bailey and Gribskov 1998) showed as that there are conserved domains among the Cas9 of all type II mycoplasma CRISPR systems that may correspond to the RuvC recombinase and the HNH endonuclease domain. However there were no significant similarities between the Cas9 proteins of the Mollicutes class. For example a sequence comparison of the Cas9 of *M. gallisepticum* with the other Cas9 of the mycoplasma CRISPR system only gave some weak similarities with closely related species, for example the similarity with *Mycoplasma cynos,* that was 54% similarity.

3. Direct repeats and tracrRNA

Typical CRISPR tracks with direct repeats (DR) interspaced with unique spacer sequences were predicted in all cases, with the exception of *S. mirum, M. salivarium* (draft genome) and *M. spumans* (draft genome). Within the CRISPR loci of mollicutes the number of DR/protospacer was found highly variable, from 3 protospacers in *M. lipofaciens* to 62 in *M. mobile*. Consensus sequences of Direct Repeats were determined for 17 mollicutes CRISPRs, showing an identical length of 36 bp and some conserved positions (Table 1). A logo plot was designed (Figure 22) showing some conserved positions on both sides of the motif as well as a few positions inside the motif. All 17 consensus DR sequences were submitted to an automated classification process using CRISPRMap tool (Lang et al., 2013). By comparing mollicutes DR with a database of 4719 consensus repeats covering 24 families and 18 structural motifs, CRISPRMap assigned all DR to superclass F except DR from *H. crinochetorum* which was not assigned. Superclass F gather DR from various bacteria with a high level of sequence diversity. Notably, this superclass includes Family F13 where DR from the *S. pyogenes* CRISPR Type II system used as a tool has

Table. 1 Consensus sequence of Direct Repeat in CRISPRs of mollicutes

Species	Consensus Direct Repeat
Spiroplasma apis B31	GTTTTAGTTATCTGACATATCTAAGGAATAGACGAC
Spiroplasma helicoides GCF_001715535	GTTATGGTACCCTGTAAAATTATGTAGTAGTAGAAC
Spiroplasma litorale strain TN-1	GTTATGGTACCCTGTAAAATTATGTAGTAGTAGAAC
Spiroplasma syrphidicola EA1	GTTTTAGTCGGCTGTCATTTTATTGTAGAATATAAC
Mycoplasma lipofaciens ATCC 35015	GTTTTAAGTTAGTACAATATTTGTGTAAGATATAAC
Mycoplasma synoviae 53	GTTTTGGGGTTGTACAATTATTTTGTTAAGTAAAAC
Mycoplasma cynos C142	GTTTTAGTGTTGTACAATATTTGGGTAAACAATAAC
Mycoplasma canis PG 14	GTTTTAGTGTTGTACAATATTTGGGTAAACAATAAC
Mycoplasma dispar ATCC 27140	GTTTTACTCTAGTAAGAAATTGTACAGCACAAAAAC
Mycoplasma ovipneumoniae NM2010	GTTTTTGTGCTGTACAATTTCTTACTAGAGTAAAAC
Mycoplasma hyosynoviae strain NPL1	GTTATAGATTACTAAAAAATTGTACGACAATAAAAT
Mycoplasma arginini HAZ145_1	GTTTTACTCTAATAAGAAATTGTACAGCACAAAAAT
Mycoplasma arthritidis 158L3-1	ACTTTTGGACTGTACAATTTTTATATAGAGTAAAGT
Mycoplasma mobile 163K	GTTTTGGTGTAGTATCATTCTTATGTATTCTTAAAC
Ca. Hepatoplasma crinochetorum Av	GTTTTGGTTAGTTGGTATTCATGAGTTGTTTAACCC
Mycoplasma gallisepticum S6	GTTTTAGCACTGTACAATACTTGTGTAAGCAATAAC
Acholeplasma palmae J233	GTTGTGTTACCCTCGTAATTTTTGCTATCTAACAAC

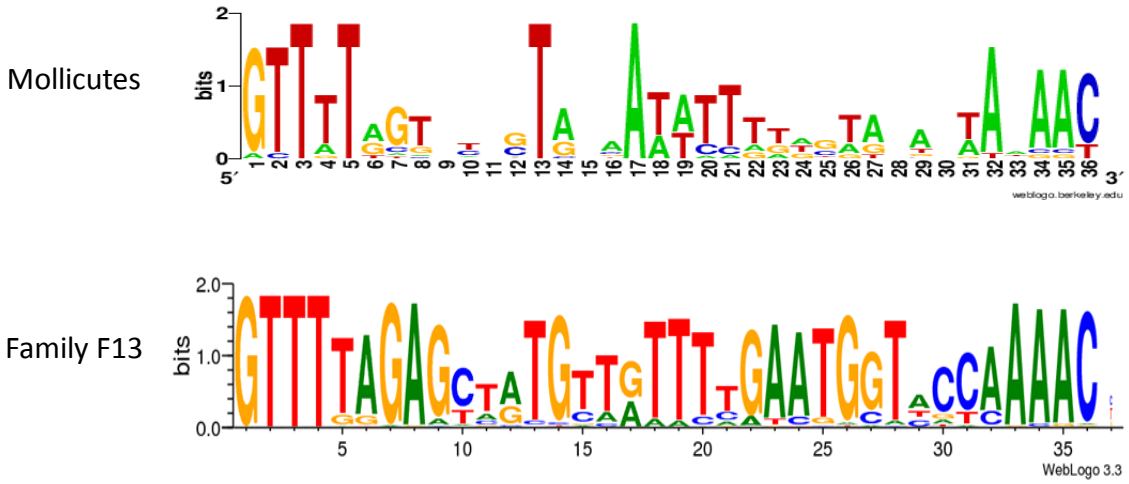


Figure 22. Consensus sequence of Direct Repeats of mollicutes CRISPR. DR sequences from 17 CRISPR systems of mollicutes were used to create a weblogo at http://weblogo.berkeley.edu/cache/file5tomDO.png. This logo was somewhat similar to the one from Family F13 of DR defined in CRISPRMap.

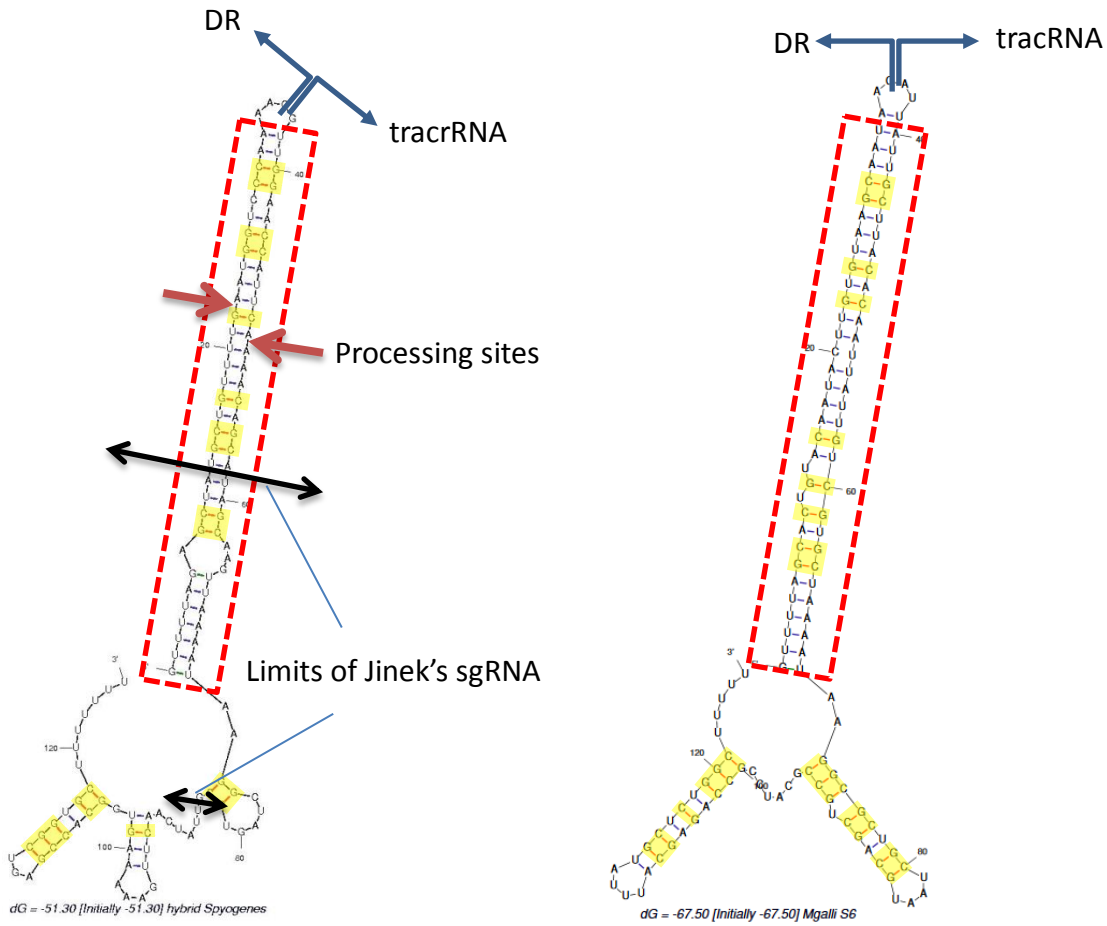
been assigned which DR consensus sequence ressembles the one constructed with mollicutes DR (Figure 22).

In Type II CRISPR, pairing between tracrRNA and DR from CRISPR locus was shown essential for the processing of crRNA and tracrRNA into a mature guide RNA (Deltcheva et al., 2011). In order to predict tracrRNA/DR interactions in mollicutes CRISPRs, secondary structures of virtual RNA molecules consisting in the 36b-long DR sequence concatenated with predicted tracrRNA were simulated using the mfold program at http://unafold.rna.albany.edu (Figure 23 and Supplementary Figure5) (Zuker et al., 2003). As a control, the same process was applied to concatenated DR and tracrRNA from *S. pyogenes* CRISPR01 system. Secondary structure predicted for this last was in accordance with previous work by Deltcheva et al 2011, showing a nearly perfect pairing of DR with the 5' region of tracrRNA. Similarly, long stem-loops involving both RNA molecules were predicted from mollicutes CRISPRs, as exemplified for *M. gallisepticum, M. cynos, M. mobile, M. synoviae* and *S. apis*. Two additional stem-loops formed by self folding of the 3' half of tracrRNAs were predicted in all cases. The work of Nishimasu and his colleguaes (Nishimasu et al., 2014) demonstrated that the additional loops of the tracrRNA of *S. pyogenes* interacts with the positively charged surface on the back side of the Cas9 protein, thus supporting the stable complex formation and enhance the stability of the sgRNA, improving its *in vivo* activity. Remarkably, the predicted hybrid structure for *S. apis* was highly similar to that of other mollicutes despite the relatively remote position of this CRISPR system as defined on the basis of Cas9 and Cas1 protein sequences.

4. CRISPR of M. gallisepticum

The evolutionary impact of mollicutes CRISPR system has been specifically studied for the bird pathogen *M. gallisepticum* during a large survey using whole-genome comparison of 12 isolates from House Finch and 5 from poultry where *M. gallisepticum* is usually found (Delaney et al., 2012).

In the *M. gallisepticum* House Finch (MGHF) species (CA06_2006.052-5-2P, NC06_2006.080-5-2P, NC08_2008.031-4-3P, NC95_13295-2-2P, NC96_1596-4-2P, NY01_2001.047-5-1P, VA94_7994-1-7P, WI01_2001.043-13-2P), the Type II CRISPR system has been extensively studied by Delaney and his colleagues, where it has been observed that host switch from poultry to House Finch was correlated with modifications on the CRISPR locus, with a loss of individual Cas genes (for NY01 strain), or even the entire system (for NC06 and NC08 strains). The remaining strains that have a complete set of CRISPR/Cas9 system are the strains F, R (low and high) and S6. For all the strains that have retained their CRISPR and Cas genes, they are located in the same area of the genome; between genes encoding a subtilisin-like serine protease (GCW_93751) upstream and a rRNA methyltransferase downstream (GCW_03780). The number of spacers varies according to the genome, with 28 spacers for the MGHF strain CA06_2006.052-5-2P, 23 for the MGHF strain NC06_2006.080-5-2P, 27 for the MGHF strain NC08_2008.031-4-3P, 45 for the MGHF strain NC95_13295-2-2P, 36 for the MGHF strain NC96_1596-4-2P, 42 for the MGHF strain NY01_2001.047-5-1P, 70 for the strain R, 39 for the strain F, 105 for the strain R high, 27 for the strain S6, 36 for the MGHF strain VA94_7994-1-7P and 45 for the MGHF strain WI01_2001.043-13-2P. The main conclusion of the work was that the CRISPR system of *M. gallisepticum* stopped recruiting new spacers and mostly was degraded in House Finch while it was probably maintained active in poultry, suggesting that phage dynamics may be more important in this ecological context.



Streptococcus pyogenes SF370 (M1 GAS)

Mycoplasma gallisepticum S6

Figure 23. Predicted DR/tracrRNA hybrid secondary structure. Sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using mfold software at http://unafold.rna.albany.edu/. Position of the DR/tracrRNA concatenation are indicated by divergent arrows. Predicted stem-loops involving DR/tracrRNA pairing are framed in red dotted lines. G-C pairs were highlighted in yellow. For *S. pyogenes*, processing sites of the natural hybrid by RNAseIII are indicated by red arrow. Limits of the artificial gRNA developped by Jinek et al are indicated by double-headed black arrows. For *M. gallisepticum*, DR and tracrRNA sequences were defined based on Chylinski 2014 and our own work.

In order to estimate if *M. gallisepticum* CRISPR system might be used a tool for mollicutes, we conducted a specific bioinformatics analysis for the *M. gallisepticum* system using the multiple strains sequences available. The main goal of this analysis was to identify a CRISPR system that has a high probability to be functional. Our global phylogenomic study indicated that the Type II CRISPR system from *M. gallisepticum* was typical from mycoplasma CRISPR/Cas9 systems, even though this system was the only one described in the Pneumoniae group. Taking into account Delaney's conclusions, we chose to focus on the CRISPR system of the poultry strain S6 which includes all predicted elements described for a typical Type II-A CRISPR system (Figure 24).

In addition, transcriptomic data performed on S6 strain by Mazin et al showed that all Cas genes and tracrRNA were expressed (S6). Moreover, this study confirmed the orientation of CRISPR locus transcription but interestingly also indicated that some expression for the locus on the other strand. This is not the first time a bidirectional transcription of CRISPR locus is occurred. The work of Lillestol and her colleagues (Lillestol et al., 2009) on the CRISPR of the crenarchaeal genus *Sulfolobus* observed a bidirectional transcription of the CRISPR locus, but this event didn't affect the activity of the CRISPR. On the contrary, a potential role of double stranded crRNAs could be to interact with other cellular defense mechanism, such as the argonaute proteins, that are known to interact with dsRNAs for their interference activity (Hutvagner et al., 2008).

All together, these data consistently suggested that the CRISPR/Cas9 system from *M. gallisepticum* S6 was active. Therefore, we focused on this system for further analysis.

5. In silico prediction of PAM sequence

First challenge to demonstrate that the CRISPR/Cas9 system from *M. gallisepticum* S6 was truly active and could be used as a tool was to characterize the PAM sequence recognized by MgCas9. As previously mentioned, the low conservation of Cas9 sequences, especially in the region of interaction with the PAM sequence, suggested that PAM specificities may be different among mollicutes, let alone between mollicutes and well-studied system of more distant bacteria such as *S. pyogenes*.

In a first approach, we performed an in silico study to identify candidate PAM sequences from genomic data. As PAM sequences are typically found next to the protospacer sequence in the DNA of invading phages or mobile elements, we designed a strategy based on the search of sequences homologous to the 524 spacer sequences found in the CRISPR systems of all *M. gallisepticum* available. The concept was to retrieve from databases some sequences homologous to the spacers and analyze the neighboring sequences to try and get PAM candidates.

- Development of a dedicated "R" script

To avoid fastidious testing of each spacer individually, we developed a script to allow a faster and more accurate analysis of all 524 spacers.

R is an open source programming language and software environment for statistical computing and graphics. Because R is very popular in biological sciences, we decided to use it to develop a dedicated script for the analysis of *M. gallisepticum* 524 spacer sequences. This development was

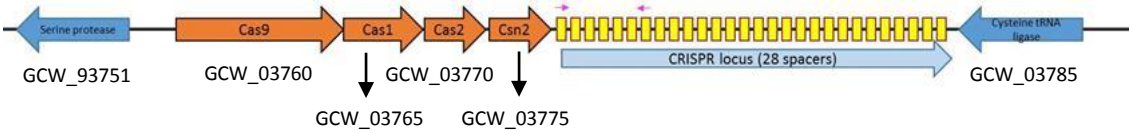


Figure 24. CRISPR locus organisation in *M. gallisepticum* S6 genome. The locus is located between positions 897217 and 904938, surrounded by a gene encoding a serine protease and a Cysteine tRNA ligase. The pink arrows indicate primers used for PCR amplification of the unique spacers of strain S6, spacer 1 and 7.

conducted by Mamadou Sall during his Master 2 internship in the laboratory, under the supervision of Patricia Thébault (LaBRI, Univ. Bordeaux).

We first narrowed the number of spacer to 193, keeping only the unique sequences found among the 524 spacers. Homologous sequences that could be part of the invading DNA at the origin of each spacer were then searched by blastn queries against different subparts of the NCBI Database. First we excluded all the mycoplasmas from the targeted database to avoid self-matching. The next test was excluding the *M. gallisepticum* genomes and search only in the mycoplasma genus. The last test was to apply blastn queries only against the phage database. The results from the 3 searches were saved in a CSV format (Comma Separated Values). These data were then submitted to our R script that was designed to extract the 15 nucleotides sequences that flanked the potential protospacer hit that meet the following criteria: (i) 90% or more of identity between the spacer and the blastn hit and, (ii) 15 bp minimum alignment length equally distributed among the 30 nucleotides of the spacer sequence and the protospacer. The sequences were also separated in two groups, depending on their orientation on the genome. From each database against which the whole group of 193 spacers was blasted, based on the orientation of the positive hits, two sets of sequences measuring 15 nucleotides were obtained. For each set, a consensus was defined using a weblogo tool and provided a graph with the prevalence of a particular nucleotide in every position from 1 to 15 (Figure 24). We scanned three different databases; the mycoplasma without the *M. gallisepticum* strains, the entire Genbank without the mycoplasma and the bacteriophages database of Genbank. These results provided us with the three different consensus sequences ATTAAAA, ATAAAAA and AGCGTAA respectively.

- *Particular case of spacer 12 of M. gallisepticum R low and R high strains*

When we blasted the spacers on the *M. gallisepticum* genome sequences, a peculiar hit outside of the CRISPR locus was identified with 100% identity between the spacer 12 of strains R low and R high and a CDS encoding the subunit C of the topoisomerase IV. This CDS was conserved in all *M. gallisepticum* strains but with some nucleotide sequence variations. Interestingly, the nucleotide sequence flanking the spacer was different only for strains R Low and R high, which contained it in its CRISPR locus (Figure 26). Based on a study on the acquisition process of spacers (Stern et al., 2010), in the rare case were the organism insert a protospacer on the CRISPR locus that is originally found in another area on its own genome, lethal cleavage of the chromosome can be avoided by evolution of the PAM sequence at the target site. As a result, Cas9 is no longer able to bind on this targeted area and does not cleave the sequence corresponding to the wrongfully acquired spacer. This suggested that in strains R Low and R High, self-cleavage in topoisomerase gene may be avoided by a variation of the genuine PAM sequence. Consequently, the 7 nucleotides flanking the 30 nt sequence of the topoisomerase IV in all strains except strains R Low and R High, were chosen as a potential PAM candidate, providing the TTAGTCC motif.

- *Extra Candidates collection*

Apart from all the above analyses, some PAM candidates were also tested:

Among the approach we tested to extract potential PAM candidates, we also searched overlapping spacers. The spacers are sequences that are acquired based on the availability of PAM motifs flanking them. So by searching among the spacers, we tried to find overlapping sequences that originated from

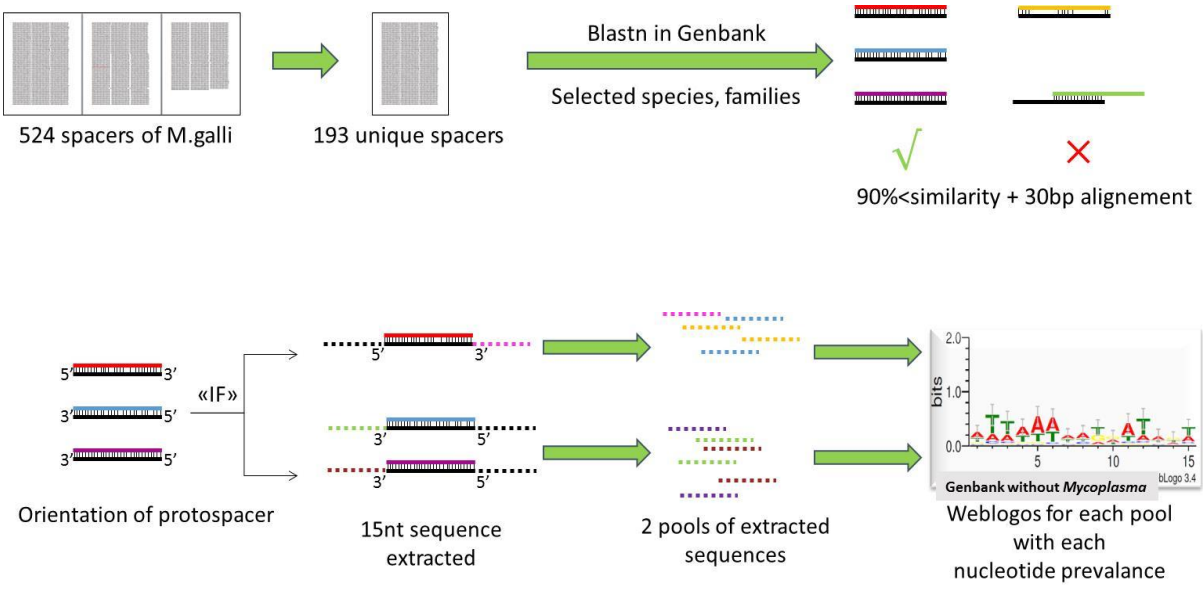


Figure 25. In silico approach, spacer analysis and PAM candidate selection. First we sorted the 524 different *M. gallisepticum* spacers and keep only the 193 spacers that where unique. Then we blasted them on selected species and families in GenBank (e.g. Bacteriophages, Mycoplasma) and keep only the GenBank "hits" (=results) that satisfied the criteria; 90%< similarity and 30 bp alignement. Then we pooled the selected spacers in two groups, depending on the orientation of the hit, which is the potential protospacer, 5' to 3' or 3' to 5'. For each of the protospacers we extracted the 15 nt sequence that was located in the 3' direction and for the two groups of extracted sequences, we designed weblogos to evaluate nucleotide frequencies at each position.

the same genomic region and carried conserved motifs flanking the overlapping sequence (data not shown). From this approach we extracted the TTGAAAA and AAGAGAA candidate.

Some other candidate we tested was the GAACCGG, AAGCCGG, AAACCGG and GGGCCGG. These sequences were tested based on the hypothesis that the nucleotides in the position 1-3 are not as important as the nucleotides in the position 4-7 which we identified in the initial experiments to better interact with the MgCas9. The AAACCGG is a profile rich in A-T in the first three nucleotides, designed based on the low G-C nature of the mycoplasma while the GGGCCGG profile is based on the *S. pyogenes* Cas9 PAM, the NGG. The two other profiles are variations of these two candidates.

Two additional PAM candidates were designed in a second round of experiments, taking first results into account. PAM candidates GGGAAAA and GTTAAAA were chosen based on the following criteria; With the GGGAAAA we wanted to verify that three G-C bp did not affect the efficiency of MgCas9, due to the fact that the consensus part of the PAM candidate we identified was between the +4 and +7 location. The GTTAAAA was chosen because the DR sequence started with these 3 nucleotides and we wanted to prove that it doesn't affect the efficiency of the recognition of the MgCas9. Another reason was that the SpCas9 can interact with a GGG PAM motif but it cannot interact with a profile without any guanine. Previous work demonstrated that the SpCas9 can function with a NAG PAM downstream of the target sequence (Hsu et al., 2013). However, Jiang and his colleagues, designed a synonymous mutation that creates an inactive PAM by changing a TGG motif to TTG (Jiang et al., 2013)(This profile corresponds also to our PAM candidate TTGAAAA). This work proves that the interaction between the SpCas9 and the PAM is very sensitive and it cannot tolerate major sequence changes. So, a GTT profile is certainly a PAM incapable to interact with the SpCas9 due to the fact that both G are replaced with T. Even though we didn't have any literature concerning the PAM of the MgCas9, we wanted to demonstrate that the MgCas9 doesn't have a profile similar to the one of SpCas9, thus we designed the PAM candidate GTTAAAA.

All the PAM candidates we selected from in silico analyses are summarized in Figure 27. The next step was the experimental evaluation of the capacity of these PAM candidates to be recognized by MgCas9 and initiate double strand break cleavage.

6. In vivo evaluation of PAM candidates and CRISPR/Cas9 system activity in M. gallisepticum

- One and two plasmids strategies

The in silico approach provided us with a number of different PAM candidates. In order to evaluate the capacity of each candidate to interact with the Cas9 protein of *M. gallisepticum* and drive cleavage of the targeted DNA, we designed a protocol based on the in vivo cleavage of replicative *oriC* plasmids. We wanted to challenge the CRISPR/Cas9 system of *M. gallisepticum* with a plasmid that carries a target sequence homologous to one of the spacers of the mycoplasma CRISPR locus, with a 3' extension corresponding to one of the 13 PAM candidates. First step was to construct a replicative *oriC* plasmid for *M. gallisepticum*. A part of the intergenic region of the *dnaA* gene was introduced to the plasmid vector pSRT2 that includes the *tetM* marker, following the results obtained by others (Lee et al., 2008) and the resulting pMGAL plasmid was successfully transformed into *M. gallisepticum* S6 strain (not shown). The pMGAL plasmid was then modified with the addition of the 3rd spacer of *M. gallisepticum* S6 CRISPR

Filter in the GenBank Database	Strand	Major nucleotide given by the Weblogo profile
Bank = Mycoplasmas without *M. gallisepticum*	Dr (26)	A T A A A A A
Bank = GenBank without Mycoplasma	Rc (24)	A T T A A A A
Bank = Bacteriophages	Rc (6)	A G C G T A A

Table 2. PAM candidates obtained after in silico analysis. The number in parenthesis is the number of positive hits for each bank. Only the first 7 nucleotides of the weblogos are shown.

locus (MgCRISPR locus). This spacer was chosen because it was the closest unique spacer to the beginning of the CRISPR locus, which means that it was acquired recently and the system has recently interacted with it. Various PAM sequences at the 3' extremity of the spacer were introduced with a Site-Directed Mutagenesis strategy. A set of 12 different plasmids differing only by the flanking PAM sequence was then obtained. As a control, we also developed a vector where the spacer was flanked by the Direct Repeat (DR) sequence of the MgCRISPR locus based on the idea that the MgCas9 should not target the CRISPR locus (Figure 27).

We used the different constructions in single-plasmid transformation and two-plasmid co-transformation assays.

For the single-plasmid transformation assays, 20 µg of each plasmid were transformed into *M. gallisepticum* cells and the number of colonies that grew on solid medium with tetracycline at a concentration of 10µg/µL was counted 2 weeks after transformation. Transformation efficiencies were then compared, as a first indication of recognized PAM candidates. The efficiency of the transformation was low, as it has been observed before (Whetzel et al., 2003), with a 5 10^{-8} per recipient CFU. For the single plasmid transformation we picked three candidates with different sequence; PAM6, PAM7 and PAM10 were tested. Alongside, with the pMGAL plasmid and the pMGAL-PAM-DR plasmid were transformed independently in the same transformation, both as negative controls. These candidates were chosen because there are strong variations among there sequences and it was considered a good start to test the capacity of each one to interact with the Cas9.

As we can see in Table 3, the number of transformants was dependent on the PAM candidate introduced into the pMGAL vector. The pMGAL and the pMGAL-PAM-DR showed increased growth rates, in comparison with the three PAM candidates, with $57*10^2$ and $31*10^2$ clones respectively. This was a good indication that both can be considered as negative controls. For plasmids with all three PAM candidates, the number of colonies varied, from 236 for the PAM10 to only 4 for the PAM7. For all the reactions, three passages in selection medium were conducted for a different number of colonies. The results for the PAM7 were impressive as all three colonies didn't grow not even after the first passage. This suggested that the plasmid was efficiently recognized and cut by MgCas9. In order to confirm this tendency, we developed another protocol.

For the two-plasmid transformation assay, equimolar quantities of two different plasmids were transformed simultaneously in the same cell population. The first plasmid was a negative control, which was the pMGAL plasmid (no spacer) and the second plasmid was each individual pMGAL-PAM construction. After transformation, cells were grown on a selection medium with tetracycline for two weeks. Then, 48 colonies were picked for each transformation and grown on a 96-well plate, in liquid medium, in the presence of tetracycline. Cultures were maintained over 3 passages and at the end of the 3rd passage, a molecular characterization of the plasmid was achieved. The idea behind this assay was that, during transformation, M*. gallisepticum* cells would receive one of the two plasmids with an equal probability. Plasmid pMGAL cannot be cleaved by CRISPR (no target protospacer), whereas the pMGAL-PAM plasmid would be a potential target for the CRISPR/Cas9 system if the PAM sequence is recognized. In this case, the pMGAL-PAM will be less frequently identified among the transformants. By contrast, if the PAM sequence is not recognized, transformants with plasmids pMGAL and pMGAL-PAM should be

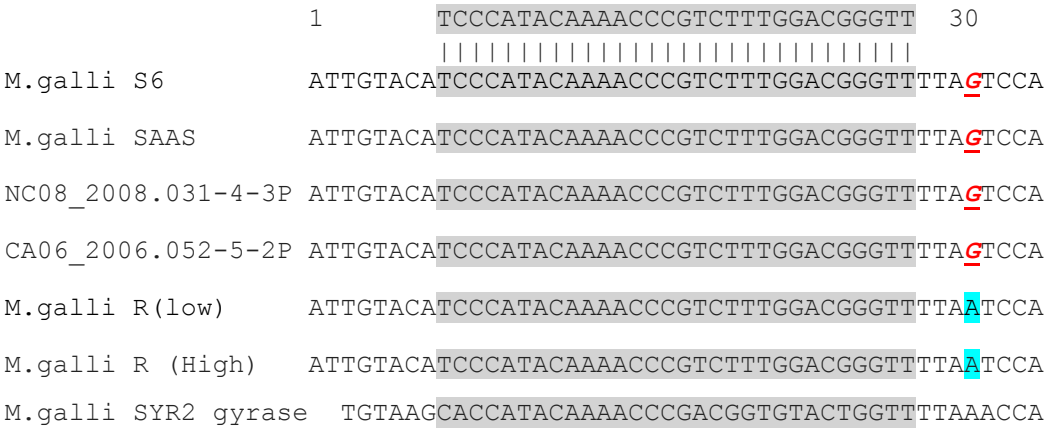


Figure 26. Analysis of spacer 12 of *M. gallisepticum* R strain. The sequence of this spacer is: TCCCATACAAAACCCGTCTTTGGACGGGTT. Partial sequence alignement of the CDS encoding the subunit C of the topoisomerase IV gene region homolog to spacer 12. Underlined in red or blue is the nucleotide that changes in the strains R low and R high. The sequence upstream of the protospacer sequence is conserved in the first 8 nucleotides.

found in equal proportions. As expected, for some PAM candidates, the distribution of both plasmids was balanced and for others, the pMGAL_spacer_PAM construction was significantly removed from the final population (Figure 28).

The results are summarized in Table 4. We can see that depending on the PAM sequence introduced downstream of the PAM vector, which is the only difference between the different plasmids, we have a difference in the ratio of the two plasmids in the final population of transformants. The most unbalanced ratios were observed with plasmids pMGAL-PAM-5, 7, 9, 12 and 13, indicating they contained actively recognized PAM sequences.

The other pMGAL-PAM plasmids were found in 10-50% of the final population. Interestingly, one global tendency is that PAM sequences that are less recognized have a relatively rich G+C content (Table 5+Figure 29).

- Determination of PAM consensus sequence

A consensus sequence was inferred from the five recognized PAMs, ATTAAAA, ATAAAAA, TTGAAAA, GGGAAAA and GTTAAAA for which almost no clones was found in the population of transformants. This sequence is the 7 bp motif NNNAAAA, with degenerated positions at positions 1 to 3.

Altogether, our work showed that the CRISPR/Cas9 system from *M. gallisepticum* strain S6 was active and that the interaction of MgCas9 was dependent on a PAM sequence NNNAAAA. This is the first functional characterization of a Type II CRISPR/Cas9 system from mollicutes.

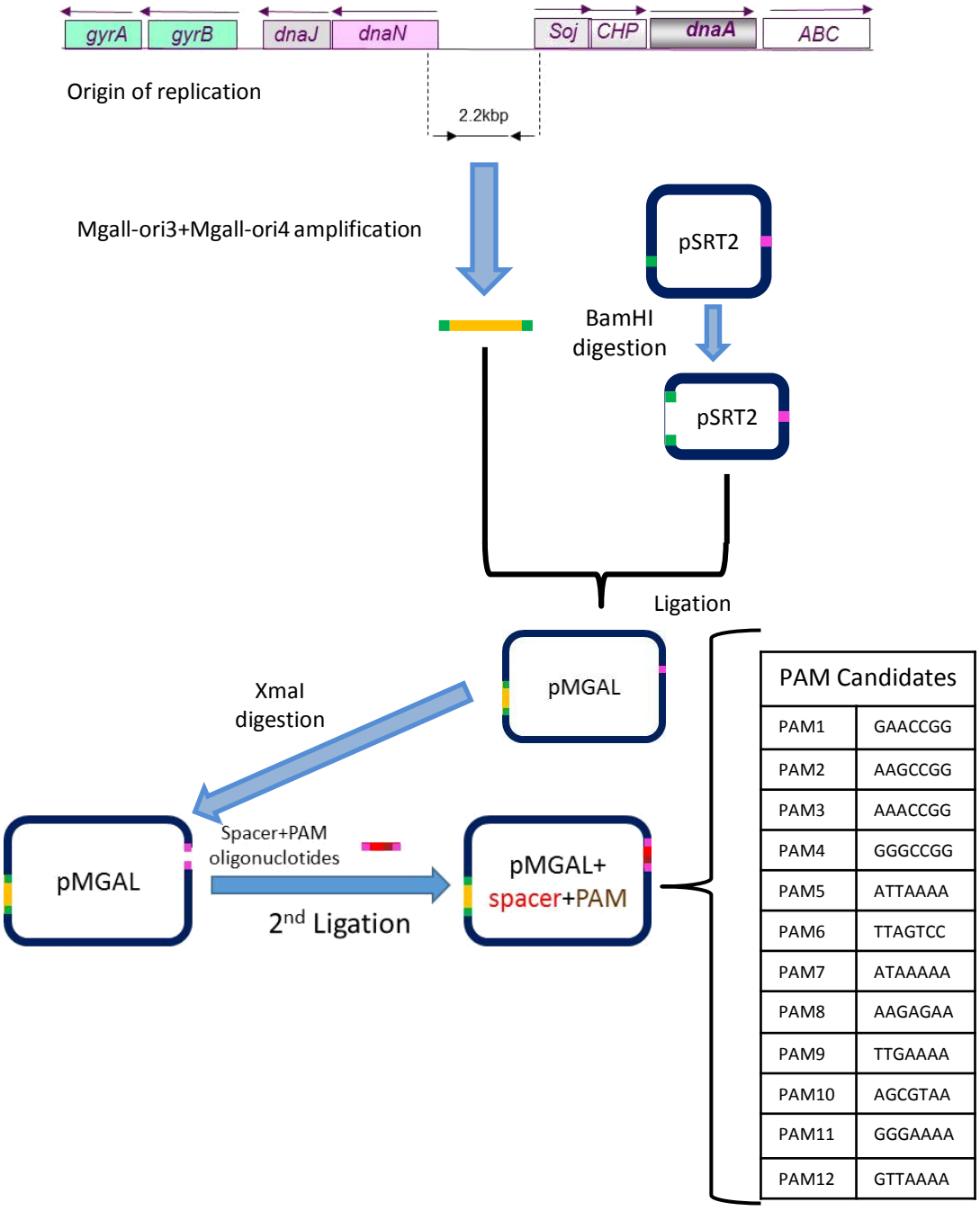


Figure 27. Construction of the pMGAL+spacer+PAM plasmids. First we cloned the origin of replication of *M. gallisepticum* into the pSRT2 plasmid, at BamHI site, leading to the pMGAL vector. A second ligation step at the XmaI with the annealed oligonucleotides carring the spacer+PAM sequence provided us with the pMGAL+spacer+PAM plasmid. The different PAM candidates were introduced using a Q5® Site-Directed Mutagenesis.

Discussion

Our effort during this part of the thesis was to characterize a new CRISPR/Cas9 system with the goal to use it for mycoplasma genome editing. The system we already developed for in-yeast engineering is limited for applications only in species for which the back transplantation is available. We also wanted a new system that could be more adapted to low G+C genomes than SpCas9 which cleavage depends on the NGG PAM sequence. Therefore, we decided to look inside the *Mollicutes* class to identify an endogenous CRISPR/Cas9 system that may have been optimized by evolution.

We first analyzed the distribution of the CRISPR/Cas9 system among the mollicutes and found that 21/52 complete or draft genomes carried a complete or incomplete system. A phylogenomic analysis revealed that all CRISPR/Cas9 systems from these bacteria belong to type IIA. The same evolutionary relationship was identified after an analysis of the Direct repeat sequence and the tracrRNA; all the DR were proven to belong to the same superclass F and tracrRNA of *M. gallisepticum, M. cynos, M. mobile, M. synoviae* and *S. apis* demonstrated a similar folding with the DR sequence. We chose to study the CRISPR system of *M. gallisepticum* because it appears as a typical system among the mollicutes and it was shown to be expressed (Gleb Fusinov, personnal communication). Moreover, studies by Delaney and colleagues also suggested that this CRISPR system has played a significant role in the adaptation of M. gallisepticum to a new host (Delaney et al., 2012).

The study of the spacer sequence of all the *M. gallisepticum* strains to extract PAM candidates was less easy than first expected: Despite more than 500 spacers from *M. gallispeticum* CRISPR were available, identification of potential targets among databases did not give a clear definition of the PAM sequence. Moreover, the information on bacteriophages that invade mollicutes are few, with the exception of particular species as *S. citri* which genome contains nearly 25% of sequences from phages and other mobile elements (Carle et al., 2010). That's why we developed a script based on the R software that had a selection capacity of hits based on more relaxed criteria. We also scanned different families of organisms in the Genbank database, in order to allow a better selection from the blast program. From this work, the first candidates ATTAAAA, ATAAAAA and AGCGTAA were obtained.

We then evaluated the efficiency of MgCas9 to interact *in vivo* with the PAM candidates selected from the *in silico* analysis. We transformed *M. gallisepticum* cells with different plasmids carrying a natural spacer of *M. gallisepticum* S6 CRISPR and the candidates PAM downstream. We used two strategies, based on single plasmid and two plasmid transformations. The single plasmid transformation provided a simple estimation of the efficiency of the PAM candidate to interact with the MgCas9: fewer colonies in the petri dishes would reflect a lower survivability due to loss of the plasmid with the resistance marker. In the two plasmid transformation assays, we hypothesized that if two plasmids are transformed simultenously in the same population, their distribution in the final population would be equal, 50% of each one. If however one of the two plasmids is cleaved due to an interaction between the PAM and MgCas9, the presence of the targeted plasmid in the final population should be really limited. We chose as a limit for selection of positive candidates, a 5% or less representation in the final population of the plasmid vector with the PAM sequence on it.

	T-	pMGAL	pMGAL+ spacer+ DR GTTTTAGCAC…	pMGAL+ spacer+ PAM6 TTAGTCC	pMGAL+ spacer+ PAM7 ATAAAAA	pMGAL+ spacer+ PAM10 AGCGTAA
ND	0	N/C	N/C	N/C	4*	236
10^{-1}	0	N/C	N/C	10*	0	20*
10^{-2}	0	57*	31*	0	0	2*

Table 3. Single plasmid transformation of *M. gallisepticum* cells. The asterisk indicates the colonies tested for the stability of the plasmid vector pMGAL carrying the resistance marker. The red asterisk indicates that the cultures from the colonies tested didn't reach 2nd passage.

After some initial results we developed new candidates based on different analysis and on the positive results from the first transformations. After several assays, we observed that whenever the concensus profile NNNAAAA was present, the *M. gallisepticum* cells rapidly eliminated the plasmids carrying this PAM. We confirmed the definition of this consensus PAM with multiple modifications of the motif and concluded that MgCas9 PAM sequence was actually NNNAAAA. We therefore started to develop a genome engineering tool from this CRISPR system.

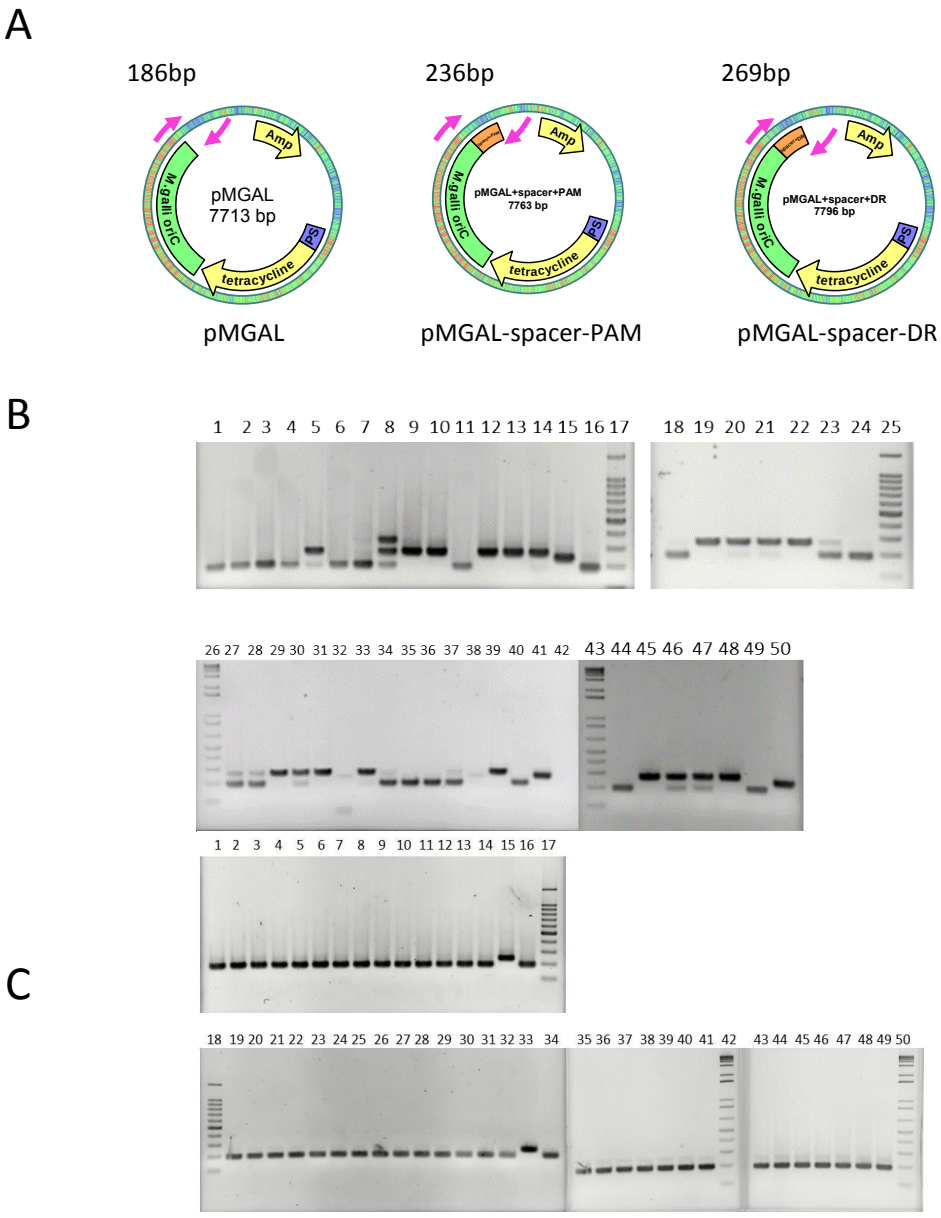


Figure 28. PCR analysis of 20 colonies with the plasmids pMGAL-spacer-DR, pMGAL-spacer-PAM5 and pMGAL-spacer-PAM9. A. size of PCR product depending on the plasmid template; B. Electrophoresis gel for the individual clones of the 3rd and 5th assay with plasmid pMGAL-spacer-DR; 1-14, clones 1-14 of 3rd assay; 18-23, clones 15-20 of 3rd assay; 27-40, clones 1-14 of 5th assay; 44-49 clones 15-20 of 5th assay; 15,41 and 50 purified pMGAL-spacer-PAM plasmid; 16,24 purified pMGAL; 42 PCR negative control H2O. In line 8 the non-specific PCR is due to high concentration of DNA during the amplification; C. The electrophoresis gel for the individual clones of the 5th assay with plasmid pMGAL-spacer-PAM5 and PAM9; 1-14, clones 1-14 from assay with pMGAL-spacer-PAM5; 15 and 33, purified pMGAL-spacer-PAM5; 16, 34, 41 and 49, purified pMGAL; 17,18,42 and 50 1Kb+ molecular marker; 19-32, clones 1-14 of 5th assay with plasmid pMGAL-spacer-PAM9; 35-40, clones 15-20 of 5th assay with plasmid pMGAL-spacer-PAM5; 43-48, clones 15-20 of 5th assay with plasmid pMGAL-spacer-PAM9.

Name	1st assay Ratio	2nd assay Ratio	3rd assay Ratio	4th assay Ratio	5th assay Ratio
pMGAL-spacer-PAM-DR	50% (8/16)	48% (13/28)	55% (11/20)	50% (10/20)	45% (9/20)
pMGAL-spacer-PAM1	50% (8/16)	-	-	-	-
pMGAL-spacer-PAM4	42.8% (9/21)	-	-	-	-
pMGAL-Spacer-PAM5	-	7,69% (3/39)	5% (1/20)	0% (0/44)	0% (0/20)
pMGAL-Spacer-PAM6	-	12,12% (4/33)	-	-	-
pMGAL-Spacer-PAM7	0 (0/33)	2,94% (1/34)	5% (1/16)	-	0% (0/20)
pMGAL-Spacer-PAM8	-	10,8% (4/37)	-	-	-
pMGAL-spacer-PAM9	-	-	0% (0/14)	0% (0/20)	0% (0/20)
pMGAL-spacer-PAM10	11% (4/36)	-	-	-	-
pMGAL-spacer-PAM11	-	-	0% (0/20)	5% (1/20)	5% (1/20)
pMGAL-spacer-PAM12	-	-	0% (0/20)	5% (1/20)	0% (0/20)

Table 4. Summary table of all transformations results using the two plasmid methods. The number in parenthesis corresponds to the number of clones carrying the pMGAL-spacer-PAM construction among the clones tested.

Name	Sequence	Number of essays	Ratio of pMGAL	Ratio of spacer+PAM
PAMDR	GTTTTAGCACTGTACAAT ACTTGTGTAAGCAATAAC	5	50%	50%
PAM1	GAACCGG	1	50%	50%
PAM2	AAGCCGG	-	-	-
PAM3	AAACCGG	-	-	-
PAM4	GGGCCGG	1	58%	42%
PAM5	ATTAAAA	4	97%	3%
PAM6	TTAGTCC	1	88%	12%
PAM7	ATAAAAA	4	98%	2%
PAM8	AAGAGAA	1	89%	11%
PAM9	TTGAAAA	3	100%	0%
PAM10	AGCGTAA	1	89%	11%
PAM11	GGGAAAA	3	96,67%	3,33%
PAM12	GTTAAAA	3	98,34%	1,66%

Table 5. Mean distribution of pMGAL and pMGAL-spacer-PAM for each construction tested considering the results of all different essays

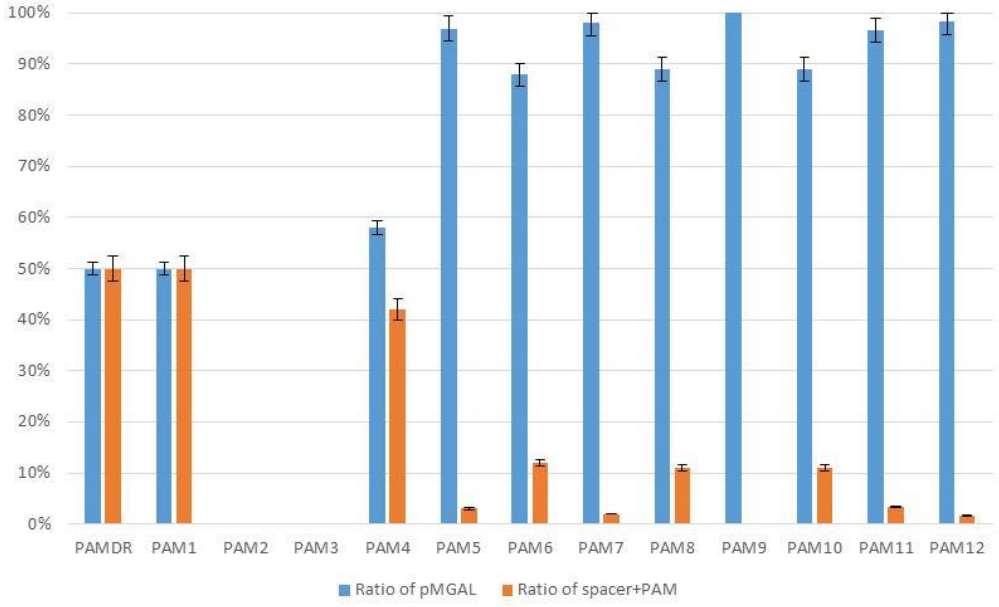


Figure 29 . Mean distribution of pMGAL and pMGAL-spacer-PAM.. The graph shows the mean distribution of pMGAL and pMGAL-spacer-PAM for each construction tested. For the PAM5, PAM7, PAM9, PAM11 and PAM12 $p > 0.5$, which means that the results are not significantly different.

Chapter 3. *M. gallisepticum* CRISPR system as a tool for targeted cleavage in mollicutes

The major problem when working on mycoplasma genome editing using directed mutagenesis technics is the efficiency of the homologous recombination. As mentioned above, in agreement with their highly reduced genomes, mollicutes have a reduced repertoire of genes involved in HDR and, more generally DNA repair. Therefore, HDR-based strategies for mutagenesis and genome engineering are generally poorly efficient. Developing an efficient tool based on CRISPR/Cas9 for direct genome engineering of mollicute genomes will require work on two complementary aspects: (1) developing an adapted CRISPR/Cas9 system that can be used with a high efficiency in various mollicutes to generate targeted double-strand breaks and, (2) improve the HDR efficiency to avoid lethal effect of DSB and direct the repair process to delete or modify the targeted region. In the frame of my PhD thesis, I started to work on the first aspect, using Mgal CRISPR/Cas9 as a potential tool to induce precise and efficient DNA cleavage in mollicutes.

In order to demonstrate that a CRISPR/Cas9 tool can be derived from *M. gallisepticum* natural system, we decided to follow the strategy that was successfully used by Jinek et al to build a tool from Type II *S. pyogenes* CRISPR/Cas9 system. This pioneer work demonstrated that Cas9 and a hybrid gRNA were sufficient to induce targeted DSBs in heterologous organisms. Therefore, our strategy includes: (1) introduction of the MgCas9 encoding gene in a heterologous mollicute, (2) design and expression of a hybrid gRNA and (3) evaluation of the efficiency by an in vivo cleavage assay.

- Introduction of a MgCas9 gene in the Mmc genome

In order to evaluate the efficiency of the MgCas9 in heterologous context, we chose to work with Mmc GM12. The rational for this choice was that: (i) any trace of endogenous CRISPR system was predicted neither in Mmc nor in other members of the Mycoides cluster, (ii) Mmc genome can be engineered by synthetic biology approaches; (iii) Mmc is a fast-growing mollicutes.

First, we introduced the gene encoding MgCas9 in the genome of Mmc cloned in yeast. To do so, we used the protocol developed in the 1st chapter, which means using the CRISPR/Cas9 tools that we developed for in-yeast engineering of mycoplasma genome. We decided to introduce the MgCas9 gene in two different loci; inside the sequence of an ICE and inside the sequence of the *glpO* gene (MMCAP2_0219). Both elements are known to be dispensable for cell life (Figure 30).

We developed the p426-ICE plasmid to target MMCAP2_0557, a gene coding for a hypothetical protein within an ICE element. We also used the same protocol for the development of the p426-glpO plasmid. Then, we amplified the gene coding for the MgCas9 including 503 nucleotides upstream of the gene to include the promoter of the gene, (not yet identified). We used two different set of primers in order to add complementary overhangs of 40 bp to each side of the amplicon, which were complementary for each desired introduction to the ICE and to the *glpO* targeting sites. After amplification, we applied the same protocol as for the introduction of the KanMX marker in the region of the *glpO* gene of Mmc. We transformed the yeast W303-Mmc-GM12-pCas9 with two different conditions in two different assays. First, we used the p426-gRNA-glpO with 4 µg of the MgCas9-glpO

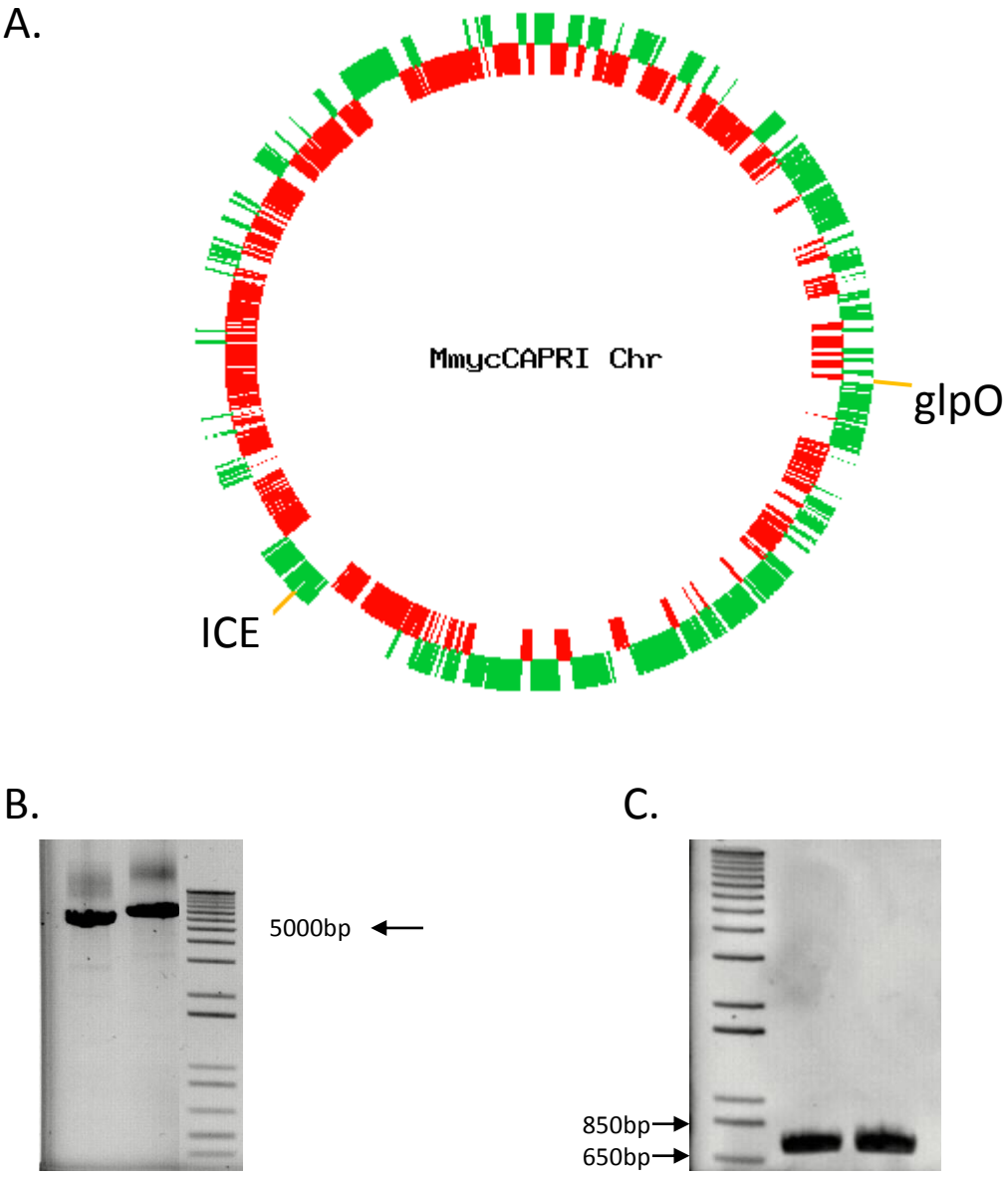


Figure 30. Cloning and verification of transcription of MgCas9 in Mmc genome. A. the two sites where the MgCas9 encoding gene was introduced in Mmc genome; B. Verification of cloning by PCR amplification of the MgCas9 gene introduced in the MMCAP2_0557 (left) and *glpO* gene sequence (right). The size of the of the amplification product of the MgCas9 is 5kbp and 5.2kbp respectively ; C. Analysis of the product of the RT-PCR showing the expression of MgCas9 gene $MmcICE_{ICE}$ (left) and $MmcICE_{glpO}$ (right).

cassette and in a second assay, we used the p426-gRNA-ICE with 4μg of the MgCas9-ICE cassette. We tested the resulting clones and selected one for each insertion site. We skipped the pool-scanning step because the efficiency of insertion was proven to be satisfying when we introduced the KanMX marker. We picked 5 colonies for the Cas9-in-ICE clones and one was positive for the introduction of the Cas9 cassette. For the *glpO* target, we tested 48 colonies and the Cas9 introduction was identified in 2 clones. We verified the genome integrity with a multiplex PCR and a PFGE electrophoresis gel and then transplanted the modified genomes into *M. capricolum* recipient cells, as described in 1st chapter. Mmc transplants with the MgCas9 gene properly introduced at both targeted sites were isolated. In both cases, MgCas9 gene sequence was verified by Sanger sequencing. In order to verify that the expression cassette we introduced contained all sequences required for an efficient expression of the MgCas9 gene, we extracted total RNAs from Mmc-Cas9$_{ICE}$ and checked the presence of Cas9 mRNA by RT-PCR. The results were positive and we concluded that the MgCas9 expression cassette contained all sequences for an efficient expression of MgCas9 in Mmc genome.

- Construction of a hybrid gRNA from Mgal CRISPR

In order to activate the SpCas9 and drive it to cleave a specific target, Jinek and his colleagues developed the gRNA (Jinek et al., 2012), a chimeric molecule merging the crRNA expressed from the CRISPR locus and the tracrRNA. They kept 12 nt from the DR part of the crRNA and the corresponding complementary sequence from the tracrRNA. They also removed the 2nd and the 3rd stem loop in order to create a minimal molecule capable to activate the SpCas9. Another work, by Deltcheva and her colleagues, demonstrated that the natural maturation site of the crRNA-tracrRNA duplex of *S. pyogenes* is at the 25th base of the DR part of the crRNA (Deltcheva et al., 2011). Chylinski and his colleagues discovered that the maturation of the majority of the tracrRNAs-crRNAs hybrids happens at a G-C site (Chylinski et al. 2014). Another work, conducted by Nishimasu and his colleagues in the crystal structure of the SpCas9 in complex with its gRNA and target DNA, proved that the bulges in the junction between the crRNA and the tracrRNA, which occurs because of mismatches in the sequences of the two molecules, are necessary for the interaction of the guiding molecule with the Cas9 (Nishimasu et al., 2014).

Aiming at constructing the simplest tool, we decided to adopt the same strategy and designed a similar hybrid gRNA from predicted tracrRNA and crRNA of *M. gallisepticum* S6 CRISPR system. As previously shown the general secondary structures of the DR/tracrRNA hybrid predicted for *S. pyogenes* and *M. gallisepticum* showed significant similarities, with a long stem loop resulting in the pairing of DR and the tracrRNA 5' sequence. By contrast to *S. pyogenes*, the maturation sites of the hybrid are not known for *M. gallisepticum*.

Taking into account the above mentioned data from *S. pyogenes* system, we developed two gRNA molecules (Figure 31 and Figure 32). The two candidates were designed to resemble the mature form of the predicted natural guiding molecule for MgCas9. Due to multiple G-C sites in the sequence of the duplex, the sizes of the two gRNA molecules were different: For the first molecule, called gRNA1, the duplex size measured 13 nt and for the second molecule the duplex size measured 24 nt. The duplex was followed by the tetraloop GAAA (as designed by Jinek in order to express the gRNA as a single molecule) followed by all the stem loops that naturally occurred during the folding of the tracrRNA of *M.*

A.

tracrRNA sequence = 130 nucleotides :

CAATGTTATTAACACTATTATTAATAACGAATGTGTTAATAACAGCACGATTTTATTCCGCGACGAT
TACGTCGACGGCGTAGGCGGTCTCGTAAATACGAGACCGAAAAATAAACAATACAGATTTTTT

B.

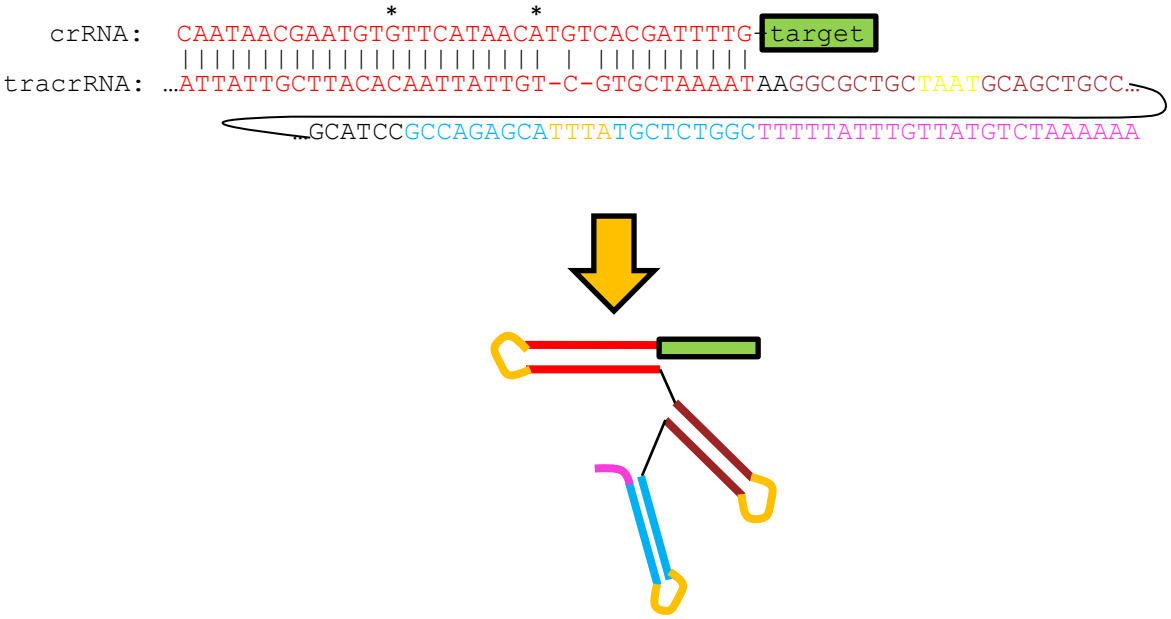


Figure 31. The design of the gRNA for the activation and guiding of the MgCas9. A. Complete sequence of the tracrRNA of *M. gallisepticum* S6 strain; B. Duplex between the tracrRNA and the crRNA: In red are the complementary sites that form the duplex between the two molecules. In brown is the first loop and in light blue is the second loop. In purple is the terminator of the transcription of the tracrRNA. In orange are the loops of the tracRNA.. The asterisks designate the G-C sites we considered as potential maturation sites of the mature complex.

gallisepticum. The end of the sequence in consisted the sequence downstream of the MgtracrRNA, which has the potential to terminate the expression of this RNA. The sizes of gRNA1 and gRNA2 are 107 nt and 125 nt, respectively. We used the Mfold software to simulate the secondary structures and obtained two predicted structures of ΔG = -43.70 and a G+C content of 41.12% for the gRNA1 and a ΔG = -52.70 and a G+C content of 38.4% for the gRNA2 with bulges and loops similar to the guiding molecule of the SpCas9 (Nishimasu et al., 2014). This suggested that the chimeric molecule we designed has the potential to be an efficient guiding molecule for MgCas9. The chimeric molecule was synthetized by IDT DNA Company and was cloned afterwards into the plasmid vector pPS3.1. This vector has the spiralin promoter and the fibril terminator in its sequence. Those elements are commonly used for efficient expression in various mollicutes including Mmc. We introduced the gRNA between these two elements and cloned the expression cassette in the pMYCO1-PSpuro vector. The result was a replicative *oriC* plasmid vector containing an expression cassette for Mgal gRNA.

- Evaluation of the MgCas9/gRNA tool in Mmc

In order to evaluate if the simplified MgCas9/gRNA tool derived from *M. gallisepticum* natural CRISPR/Cas9 system was active in Mmc, we had to find a DNA region from Mmc chromosome which cleavage might not be lethal even in case of a very low efficiency of the repair mechanisms. To avoid this difficulty, we decided to target the ICE element that is inserted in the genome of Mmc. Due to its typical way of propagation, the ICE can excise the chromosome in a cut-and-paste mechanism, suggesting that, in a population of Mmc cells, some of them may have lost the ICE. ICE excision phenomena were previously demonstrated in several mycoplasmas including Mmc (Tardy et al., 2015). We verified by PCR that excised forms of the ICE and chromosomes where the ICE had been lost could be detected (Supplementary figure 7). Our strategy was to target the ICE of Mmc using a gRNA with a 20 bp sequence identical to a region located within MMCAP2_ATAAAAA in the ICE element, just upstream of an ATAAAAA motif. We expected that transformation of Mmc/MgCas9 with a plasmid carrying this gRNA encoding gene will result in the death of most cells, because of the deadly double strand break introduced by Cas9 in the chromosome. By contrast, if some of the transformed cells had lost the ICE, these cells would survive as their chromosome will not be cleaved by Cas9. Our idea was then to select cells cured from the ICE among a population of cells that will be killed by the MgCas9/gRNA system.

- Transformation of the Mmc$Cas9_{ICE}$ cells with the gRNA.

We had already introduced MgCas9 encoding gene in two loci of Mmc chromosome and verified it was actively transcribed. We conducted a transformation of Mmc-$Cas9_{ICE}$ cells with the gRNA previously designed to target the ICE element. The transformed cells were plated on a SP5-puro8 medium. One week later, colonies grew in the petri dishes. All colonies were tested for the presence of the ICE after 3 passages in selection medium. Unfortunately, all colonies that grew in the selection medium after transformation carried their ICEs intact (Supplementary Table S1).

Discussion

Even though we managed to activate MgCas9 in its natural context and identify candidates PAM that strongly interact with this protein, our first attempt to use a simplified version of this system as an engineering in a heterologous mycoplasma didn't succeed.

1st gRNA: Target+GTTTTAGCACTGTACGAAATGCGTGCTAAAAT+structural part

2nd gRNA: Target+GTTTTAGCACTGTACAATACTTGTGAAAACAATTATTGTCGTGCTAAAAT+structural part

structural part: AAGGCGCTGCTAATGCAGCTGCCGCATCCGCCAGAGCATTTATGCTCTGGCTTTTTATTTGTTATGTCT AAAAAA

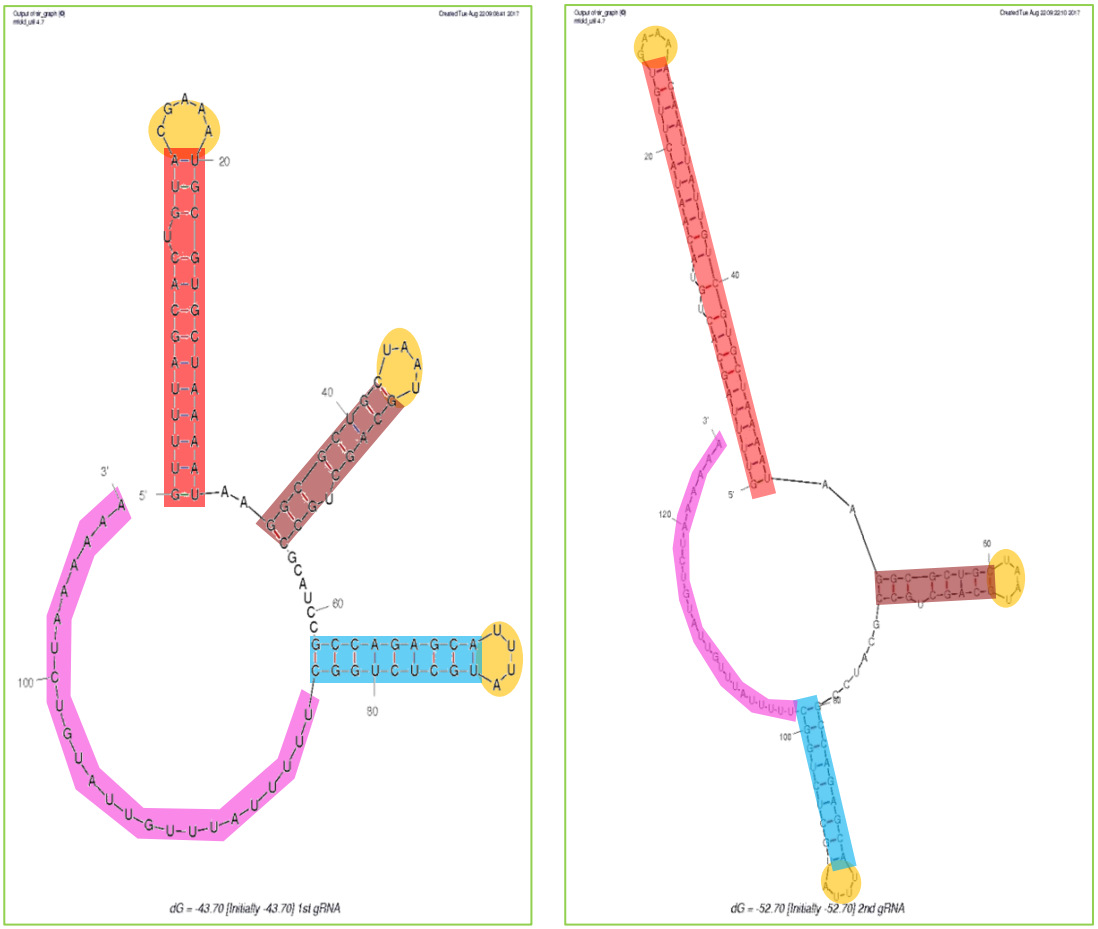


Figure 32. Simulation of the folding of the two candidates gRNAs. A. organisation of the candidate molecules that we synthesized as potential gRNAs with focus on the sequence differences. B. Simulation designed using the mfold software and the standard parameters provided by the developers. The spacer part isn't included in the simulation but it is located at the 5' end of the molecule.

The introduction of the MgCas9 encoding gene in the genome of Mmc-GM12 was successful when using the CRISPR/Cas9 tools we developed on the first part of my thesis. The back transplantation gave positive clones with MgCas9 gene introduced in the ICE element and this mutant was named Mmc-$Cas9_{ICE}$. We also verified the expression of the MgCas9 with an RT-PCR.

Jinek and his colleagues designed a chimeric gRNA to guide the SpCas9 in vitro based on the natural requirements of the tracrRNA-crRNA duplex. Our design of the two gRNAs was based on the work of Jinek and the analysis of the interaction between the tracrRNA and the crRNA by different authors (Deltcheva et al., 2011, Chylinski et al. 2014, Zheng et al., 2014). The resulting molecule had the potential to fold into a functional molecule. The stem loops and the bulges that naturally occurred during a folding prediction assay with the Mfold software provided us with two molecules with a natural folding capacity ($\Delta G < 0$) and a global structure similar to the gRNA designed by Jinek for the SpCas9.

The cloning of both gRNAs in a mycoplasma plasmid vector was successful and the plasmids were transformed in Mmc-$Cas9_{ICE}$ cells. However, our first results suggested that none of these constructs had the capacity to activate MgCas9 and reduce the number of positive transformants by an efficient cleavage of the chromosome. The colonies tested for the elimination of the ICE elements were all tested after 3 passages in SP5-puro8 medium. This proves that the puromycin resistance marker harbored by plasmid was efficiently expressed inside transformed cells. Therefore, one of the most straightforward explanations for the absence of cleavage is a problem of design or expression of the gRNA1 and gRNA2.

The next step would be to continue the assays using Mmc-$Cas9_{glpO}$ cells and try to verify if there is a difference depending on the location of the MgCas9 encoding gene. In a troubleshooting perspective, we also envisage to clone the natural tracrRNA and the crRNA of MgCas9 on a plasmid vector and express both of them in Mmc cells, instead of providing an artificial gRNA. For SpCas9, many successful studies used vectors that bring both crRNA and tracrRNA instead of the hybrid gRNA. That way, the elements could interact with each other to create the proper gRNA molecule for the MgCas9. The limitation in this method is that we don't know yet if all the genetic components required for the maturation of the tracrRNA/crRNA hybrid in *M. gallisepticum* are present in other mycoplasmas that do not have CRISPR systems. However, in other systems, the maturation process has been shown to rely on non specific RNaseIII activities that are predicted to be present in all mycoplasmas (Chylinski et al., 2014). Therefore, we expect the maturation process to occur efficiently in Mmc as well as in *M. gallisepticum*.

General Discussion

During my thesis I worked on the development of CRISPR/Cas9 tools for the engineering of mycoplasma genomes. At first, we adapted the *Streptococcus pyogenes* CRISPR/Cas9 tool developed for genome engineering of yeast genome (Di Carlo et al., 2013) in order to modify mycoplasma genomes cloned in yeast. This new tool was validated in three different species and is now currently used in the laboratory. After this first success, we tried to introduce the CRISPR/Cas9 inside the mycoplasma cells, in order to have a tool for in vivo genome engineering of mycoplasma genome. We succeeded in partially characterizing a novel Cas9 interference protein from the *Mycoplasma gallisepticum* strain S6 and introduced and expressed it in another mycoplasma species, *Mycoplasma mycoides* subsp. *capri* strain GM12. However, in the first assays we performed, DNA cleavage was not observed, possibly because of the incapacity of the hybrid guide RNA we designed to drive MgCas9 to its target. Further studies will be required to optimize an efficient tool. However, this work paves the way for an application of the CRISPR/Cas9 tools for genome engineering in mycoplasma and also initiates the perspective of the development of a genome editing tool more adapted for organisms with properties similar to mycoplasmas (e.g. prokaryotes, low G+C content).

Chapter 1: Adaptation of the CRISPR/Cas9 of *Streptococcus pyogenes* for manipulation of mycoplasma genome already transformed in yeast

The first mycoplasma genome that was cloned inside yeast was a synthetic copy of *Mycoplasma genitalium* genome, a work conducted by Gibson and his colleagues (Gibson et al., 2008). Their work proved that it is possible to assemble entire genomes in yeast as centromeric plasmids and maintain them inside this host for many generations. A similar work on Mmc genome proved that this strategy can be expanded to different organisms, phylogenetically remote from *M. genitalium* (Gibson et al., 2010). Cloning bacterial genomes in yeast has now been extended to diverse mollicutes and non-mollicute species and the work to improve the cloning methods and to stretch the limits in terms of genome size is still in progress.

The importance of these methods is that they open the door to all genetic tools available in yeast and thus, enriched the pallet of tools available for mycoplasma, so far being roughly limited to random mutagenesis and replicative *oriC* plasmids. The combination of in-yeast genome engineering with yeast – to-bacteria genome transplantation was another big step forward, thanks to the work of Lartigue and her colleagues (Lartigue et al., 2007, Lartigue et al., 2010). Genome transplantation leads to the introduction of intact genomes isolated from bacteria or yeast into a recipient cell. While the precise process remains partly unclear, the selection step using an antibiotic marker present only in the donor genome ends up with living bacteria where the only remaining genome is the newly introduced one.

Genome transplantation was first developed from bacteria to bacteria, then from genomes cloned in yeast to bacteria. If the genome had been previously manipulated in yeast, the resulting cells carried the mutant genotype. The combination of these three techniques, the cloning of mycoplasma genome in yeast, the engineering with the available tools for yeast genome editing and the back-transplantation of the mutated genomes in mycoplasma recipient cells provided a three-step procedure for mutagenesis of mycoplasma cells. Overall, these game changing synthetic biology methods are now successfully used to

delete, insert or modify genes or groups of genes in a growing number of mollicute species. Still, creating mutants remains a quite long and laborious process and there is a real challenge in improving these methods.

Our first goal was to improve the second step of the process, the engineering of the mycoplasma genome, using a new tool derived from the CRISPR/Cas9 system. When we started our work, the Cas9 protein of *Streptococcus pyogenes* (SpCas9) had already been successfully used by Di Carlo and his colleagues for gene deletions in yeast chromosomes (Di Carlo et al., 2013). We wanted to optimize this system for genome engineering of mycoplasma genome cloned in yeast. In order to make the system more easy to use, we first modified the original plasmid harboring the cassette for the expression of the gRNA by introducing a DNA fragment with two AarI enzyme sites which can be easily replaced by the 20 bp fragment required to specifically drive SpCas9 to its target.

Next we introduced a recombination template in order to guide the homologous directed repair (HDR) on the cleaved site. In the original work, when all elements are provided to the cell, Di Carlo and his colleagues observed an efficiency of ≈100% of inactivation of the desired locus and replacement with the selection marker of resistance to kanamycin. In our hands, this experiment was reproduced with the same efficiency. We then used the system on Mmc genome cloned in yeast. Concerning the recombination template, we tried two different protocols; using a PCR product as a template, carrying a cassette for the Kanamycin selection marker called KanMX and in a second experiment, we used 90 bp hybridized oligonucleotides, homologous to the flanking sequences of the targeted gene, *glpO*.

In this experiment, the efficiency was a 10% of clones successfully modified with the interruption of the *glpO* gene and insertion of the KanMX marker to its place. While being an achievement, our results showed a significant reduction in the insertion efficiency compared with the resulted published by Di Carlo and his colleagues. This can be due to different reasons.

First, Di Carlo introduced the cassette in the sequence of a gene of yeast. Therefore, the repair process of the DSB introduced by SpCas9 was mandatory to avoid cell death. The recombination cassette was designed in such a way that successful recombination with it would abolish the SpCas9 capacity to cleave the targeted site, due to a removal of the PAM sequence. As a result, the surviving clones after the manipulation would be only the ones that repaired their genome with the cassette of interest and not in any other way, like the Non Homologous End Joining. In our case, the yeast only had to repair the mycoplasma genome in a way that would save the auxotrophy selection marker. We observed in different manipulations of the mycoplasma genome in yeast, that other events leading to the removal of mycoplasma chromosome regions could happen, with the only constraint to keep the auxotrophy marker. That is why we always need to verify the genome integrity after any modification conducted in yeast with the Multiplex PCR and the PFGE. In this experience, even if there was a large number of surviving yeast clones after the manipulation, it seemed that only a few (between 10-20%) managed to modify the mycoplasma genome without spoiling its integrity.

Another reason was that when Di Carlo inactivated the candidate gene, he deleted 27 base pairs between positions 193-221of the CAN1 gene locus and introduced the cassette in place of these bases. In our assay, we deleted the entire sequence of the *glpO* gene (1164 bp) making the procedure of repair

maybe more complicated for the HDR system than a small deletion like the one Di Carlo applied. There haven't been many assays of large deletion in other organisms using the CRISPR/Cas9 system, but concerning modification of bacteria genome in vivo (Cobb et al., 2016) the efficiency remained the same within 1500 bp range but large deletions in yeast has proven to reduce the efficiency of the CRISPR/Cas9 system to even a low of 10% (Hao et al., 2016). Still large deletion with a reparation template provided exteriorly has not been conducted yet to allow comparison with our own results.

In other assays conducted in Mmc, *M. capricolum* and *M. pneumoniae* and using 90 bp oligos as recombination templates, the efficiency was also between 10-20% for the three species. However, even though the efficiency didn't reach the levels observed when targeting yeast chromosomes, these results still represented a significant improvement for the targeted mutagenesis of mycoplasmas. Moreover, other assays conducted later in the laboratory, like a 20 kbp deletion on the MIB-MIP locus of Mmc (Arfi et al., 2016, unpublished) and the targeted introduction of the Cas9 gene of *Mycoplasma gallisepticum* into the genome of Mmc confirmed the efficiency and the versatility of this tool for the genome engineering of bacterial genomes clones in yeast. One main improvement compared to other mutagenesis method is that the efficiency of the CRISPR/Cas9 system allowed us to skip the selection step required by all previous methods (Noskov et al., 2010, Chandran et al., 2014), which means reducing time of work and cost of the process. Another step that increased the efficiency of the process was the incubation for 1h in YPDA medium followed by another incubation for 48h in liquid auxotrophic medium before plating on Petri dishes. The idea for these extra steps was at first to allow the yeasts to recover and begin to grow in a rich medium after the manipulations and the thermic shock of transformation, thus the incubation at 1h at YPDA. The second incubation for 48h in liquid auxotrophic medium was inspired by the work of Bao and his colleagues (Bao et al., 2014). Even though they were using the same organism and the same target as Di Carlo (Di carlo et al., 2013), they used this incubation step to increase the efficiency of gene targeting. Potentially, the incubation in an auxotrophy liquid medium of the yeast cells for 48h eliminates all the yeast cells that haven't been properly transformed with the plasmid vector for the gRNA and could produce background on the plates. Furthermore, SpCas9 has been shown to cause the maximum number of DSBs between 48h and 72h in human cell lines after the introduction of the expression plasmid vector inside the cell (Kim et al., 2014). These results may have inspired Bao to adapt his protocols to test if the optimal conditions for the human cells are also optimal for the yeast. Therefore, 48h appears to be the ideal moment to plate the transformed cells, in order to increase the frequency of proper DSBs and HDRs in the population of transformed cells.

Chapter 2: Functional characterization of the native CRISPR/Cas9 system of *M. gallisepticum* with a combination of an *in silico* and an *in vivo* appro5ach.

The CRISPR/Cas9 tool developed in the first chapter of my thesis can by applied for in-yeast mycoplasma genome engineering. However two major problems are limiting the application of this new tool on a vast range of mycoplasmas. First the possibility to apply the back transplantation in mycoplasma recipient cells is limited to the mycoides cluster and for a few species outside of it (but still its closest relatives) (Labroussaa et al., 2016). Another important fact is the relatively low frequency of SpCas9 PAM sequence (NGG) in mycoplasma genomes due to the low G+C content that characterizes mollicutes genomes in general. Thus, we decided to characterize an endogenous CRISPR system naturally

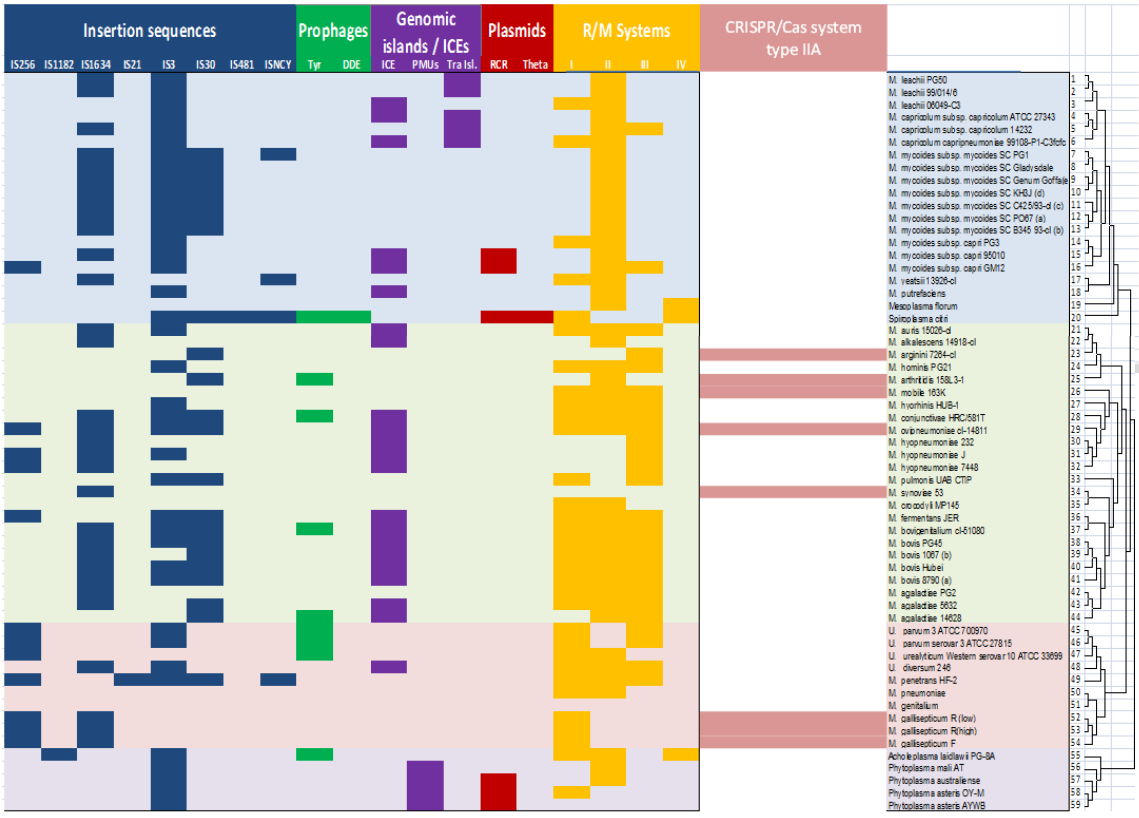


Table 6. Distribution of mobile elements and CRISPR/Cas systems in selected species among the *Mollicutes* class. The different Insertion Sequences, phage, genomic island, plasmids, Restriction-Modification (RM) system families that have been described in *Mollicutes* are specified. This figure was adaptated from Breton et al, unpublished.

present in mycoplasma, in order to develop the most adapted CRISPR tool for them. We started by studying all the species of the class *Mollicutes* and we identified multiple CRISPR/Cas9 systems, all of them belonging to the Type II CRISPR systems. Appart from the CRISPR systems identified in *S. apis S. helicoides* and *H. crinochetorum* that seem to belong to the subtype II-C, all the other CRISPR systems belong to the subtype II-A. This means that all the CRISPR/Cas9 systems of mollicutes have a Cas9 protein and a tracrRNA to allow crRNA maturation and target interference. Using bioinformatic approaches, we predicted the tracrRNA in multiple species and compared the duplex structure between the crRNA-tracrRNAs of *M. gallisepticum, M. cynos, M. mobile, M. synoviae* and *S. apis*. The simulation revealed a relatively well conserved secondary structure struture that carries the basic structural elements that have been already identified in the other CRISPR/Cas9 subtype II-A systems (Deltcheva et al., 2011, Chylinski et al., 2014).

The CRISPR/cas9 system has been identified as a defense mechanism against invasive nucleic acids from phages and other mobile elements. A brief analysis of a selection of species of the *Mollicutes* class has demonstrated that mobile elements are well distributed among mollicutes (Table 6). The species that possess a CRISPR/Cas9 system that were identified during this thesis have a tendancy to eliminate the phages elements from their genome, as we can see for *M. gallisepticum, M. ovipneumoniae, M. synoviae, M. arthritidis, M. mobile* and *M. arginini,* even though the csn2 is disrupted in *M. arginini*. On the contrary, species without a CRISPR/Cas9 system, like *M. bovis, M. bovigenitalium, M. agalactiae, U. parvum and S. citri* have identified prophage sequences in their genome. An interesting species is *M. arthritidis* that has a complete set of genes and a potentially active CRISPR locus (33 spacers and leader sequence), but prophages sequences has been identified in its genome. However, after blasting all individual spacer sequences on the genome of *M. arthritidis*, there was no similarities with any area of the genome, appart from the CRISPR locus. This indicates a potential evolutionnary event where the prophages identified on the genome of this mycoplasma invaded the cell before the CRISPR system appeared or be activated and as a result, eventhough the CRISPR couldn't identify these old infections due to one of the many defense mechanisms of the phages, like the abolishement of the PAM sequence, however it doesn't allowed new viral infections and have already created a pool of spacers to protect *M. arthritidis.*These observations demonstrated that there is a potential strong control of the prophage mobile elements by species with an active CRISPR/Cas9 system but it remains to be studied further.

We selected CRISPR/Cas9 system of *M. gallisepticum* to further characterization for a number of reasons. This mycoplasma is amenable to transformation using chemical transformation and electroporation and the protocol was already available in the laboratory. Based on the work of Delaney and his colleagues, *M. gallisepticum* CRISPR has been affected by evolution, where a shift of host from poultry to house finch resulted in a loss of function of the CRISPR/Cas9 system for a number of strains. This suggested that this CRISPR locus had an active role in the physiology of this species and thus it was the first indication that it was an active system. Then, thanks to a personnal communication with Gleb Fusinov, a joint author with Pavel Mazin in the publication of Mazin and his colleagues (Mazin et al., 2014), we obtained the transcriptomic data of *M. gallisepticum* strain S6 in the CRISPR locus. We observed that all Cas genes and the CRISPR locus were properly expressed, which was an additional indication that the system was active. The next step was to identify the PAM sequence of the *M. gallisepticum* Cas9 protein (MgCas9).

To identify the PAM of MgCas9 we started with an in silico analysis of all the spacers of *M. gallisepticum* strains but this approach was finally less straightforward than anticipated. There was not a single spacer sequence for which we could identify a 100% identity in the Genbank database. So, we developped a script to analyze all the spacer sequences and find homologous sequences with a >90% identity and 100% alignment in Genbank database. These parameters provided us a number of hits from which some consensus motifs were derived, the most represented being ATTAAAA, ATAAAAA and TTGAAAA. These three profiles were introduced into the pMGAL vector, a plasmid designed to carry the origin of replication of *M. gallisepticum* and a spacer from the CRISPR locus of *M. gallisepticum* strain S6. The three profiles were introduced downstream of the spacer sequence. The susceptibility of these plasmids to a cleavage by MgCas9 was evaluated in a transformation assay using a two plasmid method. The first plasmid was the pMGAL without spacer or PAM sequence. The second plasmid carried the spacer and a candidate PAM sequence. Both plasmids were co-transformed in equal proportions. We then observed the distribution of the two plasmids in the population of transformed cells. A half-and-half distribution of the two plasmids was interpreted as a negative interaction between the MgCas9 and the candidate PAM, where as a decreased proportion of the plasmid vector with the spacer and the PAM sequence would mean an elimination of this plasmid due to the CRISPR intereference capacity.

The negative control in these manipulations was a two plasmid transformation using pMGAL and a pMGAL with the spacer sequence and the sequence of the DR downstream of it. The hypothesis was that MgCas9 would consider this plasmid as a home element and would'nt attack it. This hypothesis was validated, as the pMGAL-spacer-DR plasmid was detected with a frequency of 50% in almost all experiments.

A first round of assays showed a really weak representation of the plasmids with the three PAM candidates in the final population. These results suggested that the consensus NNNAAAA PAM sequence have the potential to interact with the MgCas9 and initiate the DNA cleavage. Then, other sequences were also tested as candidates based on a comparison between the new candidates and the one from SpCas9, NGG. We designed a candidate with the GGG trinucleotide followed by the consensus profile of AAAA, to verify if the enrichement of the first 3 positions of the PAM with G+C would not interfere with the efficiency of the cleavage. This hypothesis was proved correct, as the corresponding plasmid was massively cleaved by MgCas9. We also designed a PAM sequence which was the GTTAAAA, due to the fact that it has been proved for the SpCas9 that the GTT profile, a profile containing two T in the place of the dinucleotide GG of the natural PAM for the species, can inhibit completely the activity of SpCas9. In our experiments, this PAM sequence was efficiently recognized, leading to an active cleavage by MgCas9. Thus we concluded that Cas9 present in the CRISPR systems of *M. gallisepticum* can interact with A+T rich PAM with the concensus NNNAAAA, which is in accordance with the low G+C genome content of mycoplasmas.

Other studies aiming at characterizing PAM recognitions sequences of various Cas9 used different strategies based on newly developped tools (Leenay et al., 2016) and in-vitro assays (Fonfara et al., 2013) but these methods were not available for us as MgCas9 has not been produced yet. Moreover, the crRNA and tracrRNA complex was not identified either. Therefore, our in vivo study was the only available approach that would allow us to characterize the PAM sequence for this protein, without the

supplementary labor of purifing the protein or developing different gRNA candidates to guide it *in vitro*. Our results demonstrated that *M. gallisepticum* CRISPR system was truly active in the S6 strain and that the PAM sequence was NNNAAAA. However, because we worked with the complete natural system, we could not conclude that as in *S. pyogenes*, a simplified system consisting in Cas9 and a hybrid gRNA could be enough to induce efficient DNA cleavage.

Chapter 3: *M. gallisepticum* CRISPR system as a tool for targeted cleavage in mollicutes

The genome editing tools that have been developed for the mycoplasma are limited. The adaptation of a CRISPR/Cas9 system to enrich the existing pallet with a new tool for mycoplasma genome editing would be a significant progress for the study of these bacteria. Our idea in this part of the thesis was to develop a minimal CRISPR/Cas9 system in a mycoplasma that had no natural CRISPR system. Because the efficiency of the homologous recombination mechanism of mycoplasma is weak, we decided to focus on the development of an efficient tool for gene interruption of targeted genes. For the reparation of the targeted genes with the homologous recombination mechanism, several works have already tried to increase the efficiency of homologous recombination in mycoplasmas by tinkering with the *recA* gene (Allam et al. 2010, Hassan et al. 2017). In prokaryotes, in general, there were application where the tried to bring a heterologous recombination system inside the cell, like no-SCAR system for genome editing in *Escherichia coli* (Reisch et al., 2015). These works prove that it is possible to resolve this limitation and our interest was to develop an efficient tool based on the CRISPR/Cas9 system that can be used as soon as an improvement in the homologous recombination machinery of mycoplasma is found.

We chose Mmc, a fast growing mycoplasma which can be manipulated easily with synthetic biology approaches. Using the CRISPR/Cas9 tools developed in the first chapter, we managed to introduce the MgCas9 encoding gene under control of its own promoter in two loci of Mmc-GM12 genome cloned in yeast. We ended up the process by a back-transplantation and obtained mutants with MgCas9 gene inserted (i) in the sequence of the ICE gene MMCAP2_0557 or (ii) in the sequence of *glpO*-MMCAP2_0219. Both targets were chosen as non-essential genes. The rational for an introduction within the ICE was also part of strategy of curing the genome from the ICE while leaving no trace of the Cas9 gene.

The first minimal CRISPR/Cas9 system designed by Jinek et al needs a gRNA to activate Cas9 and guide it to its target. Similarly, we designed a gRNA, taking into account the work of Deltcheva and her colleagues about the maturation of the crRNA of *Streptococcus pyogenes* in its natural system (Deltcheva et al., 2011), the work of Jinek and his colleagues on the minimal RNA molecule capable to guide the SpCas9 (Jinek et al., 2012) and the prediction of the crRNA-tracrRNA complex of *M. gallisepticum* strain F conducted by Chylinski and his colleagues (Chylinski et al., 2014). We designed two molecules, with different size on the duplex region, which carried a loop with the sequence GAAA, to join together the crRNA and the tracrRNA molecules. This tetraloop was also used by Jinek and his collegues because it doesn't interact with the folding capacity of the gRNA designed to guide the SpCas9 (Jinek et al., 2012). In that way, the two joint molecules of the crRNA and tracrRNA can be introduced as a nucleotide sequence on a plasmid vector and expressed inside the cells as an active crRNA. Both gRNAs were introduced in Mmc cells but both assays bring negative results in the interruption of ICE.

These results are showing that it is not as simple as we though to activate the MgCas9. The maturation of the crRNA and tracrRNA to develop the crRNA that can guide the MgCas9 is not studied yet. Even though we tried to take into account the important structural characteristics that have been identified for the interaction between the two molecules in other organisms, we failed to observe any interaction of our MgCas9 with the ICE of Mmc. This leads us to the conclusion that both gRNAs failed to interact properly with MgCas9. The applications that are now available using the CRISPR/Cas9 tool and derivatives are still in development, but the idea is to apply this tool for the modification of species with a scientific interest based on the projects of the laboratory, which are currently *Mycoplasma mycoides* subsp. *mycoides* and *Mycoplasma pneumoniae*. For both species (data not shown for Mmm) we have succeeded in developing mutants of interest that could be further characterized as soon as the transplantation techniques are available. Also a number of techniques are currently being developed based on this tool: The in-situ tagging and the cloning of mycoplasma linearized genome in yeast, are two examples were we simply took advantage of the cutting efficiency of Cas9 in vivo and in vitro and the extremely efficient HDR mechanisms of the yeast in order to improve the efficiency and reduce the time needed of existing protocols. Concerning the work conducted on MCAP and its small RNA MCS2, we realized that the modification of this area of the genome of Mcap, renders the mutated cells unviable after the transplantation procedure. Despite the fact that we didn't have the time to further advance our research in this genome, we have designed a "complementation plasmid", harboring the the entire locus of MCAP0015-MCS2-MCAP0017 genes (size). We wanted to introduce this plasmid in recipient cells simultaneously with the mutated genome that lacks this area and verify its importance on the cell viability. The plasmid could then be easily modified multiple times using the Q5® Site-Directed Mutagenesis Kit Protocol in order to understand precisely the role of the MCS2, as a potential cofactor of the MCAP0017 gene or another element on the mycoplasma genome. The plasmid harboring the modified sequence could replace the wt plasmid by selection in different markers after transformation of the mutated cells. Considering the area coding for the MCS2 essential, it will be a new available application; using the CRISPR/Cas9 tools for in-yeast engineering of mycoplasmas genome in order to allow us to study the functionality of essential elements on the mycoplasma genome directly in mycoplasma cell. Specific perspectives on MCS2 project: co-transplantation of a non-viable genome together with a complementation plasmid harboring MCAP0015-MCS2-MCAP0017 for fast functional analysis of the region (easier to test first many different deletions/mutations of MCS2 region by simple quick-change protocol on plasmid and try to replace the wt plasmid by the mutant – with different markers) The pallet of tools for genome engineering of mycoplasma genomes have significantly expanded during the last years with the addition of the CRISPR/Cas9 tools, which have been used for all the above applications. The perspectives of these works is to adapt these tools for modification of the mycoplasma genomes of species with a key importance to the scientific community and our laboratory, which are the *Mycoplasma mycoides* subsp *mycoides* and the *M. pneumoniae* in order to succeed the goals of our project, in the development of vaccine strains and minimal chassis cells (Mycosyvac, Minicell, NSF project). Another perspective is the application of these tools in order to manage to establish efficient protocols for cloning in-yeast and modification of the non-cultivable Mollicutes like the phytoplasmas and hemoplasmas. Concerning the characterization of the native CRISPR/Cas9 system of *M. gallisepticum*, the main perspective remains to obtain a minimal system for direct usage in mycoplasmas and other mollicutes. More precisely, we would like to keep on working in developing an

efficient gRNA and to continue the study of the MgCas9 in vitro, with an expression of the recombinant protein in *E. coli* or yeast cells. Following that, we will develop mini-genes of our gRNA for *in vitro* test, as it has been already done (Lee et al., 2015). That way, we will be able to better understand and control this new MgCas9 protein, for applications in the mycoplasma cells but also other organisms with similar properties.

In conclusion, during this thesis, we first developed a CRISPR/Cas9 tools for the engineering of mycoplasma genome cloned in yeast. We succeeded in applying this tool for one step seamless deletion of genes and elements of variable sizes, with a surgical accuracy and a satisfying efficiency for isolation of positive clones. The system was used in three different species with similar efficiencies. In a second part, after a global overview of CRISPR systems in mollicutes, we characterized the Cas9 protein of *Mycoplasma gallisepticum*, in order to develop a CRISPR/Cas9 system more adapted for mycoplasma and organisms with similar genome properties. Using an in silico approach we identified a consensus NNNAAAA PAM sequence, that we later verified by in vivo assay in *Mycoplasma gallisepticum*. We finally started to develop a minimal CRISPR/Cas9 system based on the MgCas9 in an heterologous organism, *Mycoplasma capricolum*. We managed to clone and express MgCas9 in *M. capricolum* but our efforts to activate and guide it to a specific target using a gRNA didn't bring positive results. The next step is to further characterize the MgCas9 protein and the crRNA that can guide her to its target, in order to develop a complete CRISPR/Cas9 system originated from mycoplasma for further applications in these species and in other bacteria.

Materials and Methods

Material and Methods

Medium for yeast growth

The rich medium YPDA (Fisher Scientific) was used for the growth of all yeast strains. Whenever an auxotrophy medium was required, an SD Base would be with the desired complement, lacking the selection amino acids (Fisher Scientific), would be used. For agar plates, a concentration of 2% agar base was used.

Yeasts strains

The *W303a (MATa his3-11, 15 trp11 leu2-3,112 ura3-1 ade2-1 can1-100)* and *VL6-48N (MATα trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir*) strains were used as recipient cells to clone mycoplasma genomes and evaluate the efficiency of the CRISPR/Cas9 system developed by Di Carlo and his colleagues for yeast genome engineering. In our studies, the strain W303a was used as a platform to engineer the genomes of *Mycoplasma mycoides* subsp*. capri* GM12 YCP, *M. capricolum* subsp*. capricolum* YCP and *M. pneumoniae* M129, giving the names W303a/Mmc, W303a/Mcap and W303a/Mpneu to the resulting strains.

Medium for mycoplasma growth

We used the SP5 medium, deriving from the original SP4 medium (Tully et al., 1977). The SP5 medium is composed of 3.5 g/l of Mycoplasma broth base (Fisher Scientific), 10 g/l of Bacto Tryptone (Fisher Scientific) and 5.3 g/l of Bacto Peptone (Fisher Scientific). The solution was adjusted to pH 7.5, autoclaved for 20 min at 120°C, then supplemented with 0.125% (w/v) glucose, 5% (v/v) CMRL 1066 10× (Invitrogen), 0.11% (w/v) sodium bicarbonate, 1 mM L-glutamine, 3.5% (v/v) yeast extract (Fisher Scientific), 0.2% (w/v) TC yeastolate, 17% (v/v) fetal bovine serum, 0.1 mg/mL ampicillin and 0.002% (w/v) phenol red.

Mycoplasma strains

The mycoplasma used as a recipient cell for back transplantation of mycoplasmas genome after in-yeast engineering was the *Mycoplasma capricolum* subsp*. capricolum* ΔRE cl17.5 (Mcap ΔRE cl 17.5) and the *Mycoplasma capricolum* subsp*. capricolum California kid* (Mcap CK) strain.

Escherichia coli strains

For all plasmid constructions, the commercially available cells from NEB DH10B, for electroporation (C3020K) or chemical transformation (C3019H) were used. For the Q5® Site-Directed Mutagenesis Kit Protocol (E0554), the NEB 5-alpha Competent *E. coli* cells (C2987) were used, as suggested by the provided protocol.

Plasmids

Plasmids already existing in the laboratory

p414-TEF1p-Cas9-CYC1t

This plasmid was used to express in a constitutive manner the Cas9 protein of *Streptococcus pyogenes* (SpCas9) inside the yeast cell. It contains the CEN6 element that renders it a centromeric plasmid, the ARSH4 yeast origin of replication and the tryptophan TRP1 as a selection marker. The SpCas9 encoding gene is a codon optimized version originally designed for expression in human cells (Mali et al., 2013). Constitutive expression is controlled by a TEF1p promoter and nuclear localization is driven by a C-terminal SV40 tag.

p426-SNR52p-gRNA.CAN1.Y-SUP4t

This plasmid was used to introduce and express the gRNA inside the yeast cells. It is a high copy 2μ plasmid with uracil as an auxotrophy selection marker. The expression of the gRNA is under the control of the SNR52 promoter with the SUP4 flanking sequence as a terminator. The original plasmid that Di Carlo used, was coding for a gRNA targeting the CAN1 locus on the yeast genome.

Both p414-TEF1p-Cas9-CYC1t and p426-SNR52p-gRNA.CAN1.Y-SUP4t were provided by Addgene (#43802 and #43803).

Plasmids developed in the laboratory

p426-SNR52p-gRNA.AarI-SUP4t

The plasmid was developed to modify the target part of the gRNA sequence of the original plasmid and develop the desired gRNA with a cloning spacer sequence. The 20 bp homologous to the CAN1 locus were replaced by two AarI digestion sites in reverse an opposite orientation, with two bases separating them and flanked also by 2 bases from each side. The resulting sequence was the GG**GCAGGTG**GA**CACCTGC**CT with the two AarI sites in bold.

p426-SNR52p-gRNA.glpO-SUP4t, p426-SNR52p-gRNA.MCS2-SUP4t, p426-SNR52p-gRNA.MCAP0015-SUP4t, p426-SNR52p-gRNA.MPN142-SUP4t

All 4 plasmids were developed to target a sequence on the genome of Mmc, two sites on Mcap and one on *M. pneumoniae*. They were developed by introducing the desired annealed oligonucleotides on the linearized p426-SNR52p-gRNA.AarI-SUP4t through a ligation procedure.

Construction of the p426-SNR52p-gRNA.AarI-SUP4t

We used a Gibson assembly to replace the sequence of the CAN1 locus with the two AarI digestion sites. The p426-SNR52p-gRNA.CAN1.Y-SUP4t was amplified to create two linearized parts. Fragments one and two were called URA and Amp, respectively, due the marker each one carried. The primers used were p426F and AarI_gRNA_modR that amplified the region between positions 817 and 3890, the primers AarI_gRNA_modF and p426R to amplify the region between positions 3910 and 816 of the p426-SNR52p-gRNA.AarI-SUP4t (Table S1). The 20 bp excluded are the target part of the gRNA, targeting the CAN1 locus. The primers AarI_gRNA_modF and AarI_gRNA_modR carried the sequence of two AarI digestion sites in reverse an opposite orientation (Supplementary table S2).

The primers also carried 20 bases of over hangs that added complementary ends of 40 bases between the extremities of both the URA and the Amp fragments. Both PCR products were purified using

the GE Healthcare DNA purification kit. The purified products were incubated for 2h at 37°C in a 50 µL digestion reaction with DpnI to eliminate the remaining PCR template. A second purification using the GE Healthcare DNA purification kit followed. Finally, 25 fmol of each product were mixed to a final volume of 5 µL. 15 µL of the Gibson assembly mix (2x Endonuclease, Ligase, Polymerase) were added in the mixture of both parts. The reaction was incubated 5-10 seconds at room temperature (the time needed to mix all the reagents two times) and then it was put for incubation at 50°C for 1h. After the incubation 1 µL was transformed in NEB DH10B electro-competent cells (C3020K) following the provided protocol. The resulting clones were plated in LB/Amp100 plates. Positives clones were tested for the correct assembly product with a PvuI restriction digestion. Four clones were also sent to verify the modification of the CAN1 sequence and the proper insertion the cloning spacer AarI.

Construction of the p426-SNR52p-gRNA.glpO-SUP4t, p426-SNR52p-gRNA.MCS2-SUP4t, p426-SNR52p-gRNA.MCAP0015-SUP4t, p426-SNR52p-gRNA.MPN142-SUP4t plasmids

The p426-SNR52p-gRNA.AarI-SUP4t was digested for 5h at 37°C with the AarI enzyme, following the reaction conditions suggested by Thermo Fisher Scientific. The result molecule was de-phosphorylated using the Antarctic Phosphatase of NEB (M0289) and incubating 1h at 37°C. The oligonucleotides to introduce the desired spacer sequence, were phosphorylated and then properly annealed as follows: In a reaction of 40 µL, 100 pmol of each oligonucleotide (1 µL of a 100 µM concentration) were mixed with 4 µL of ATP 10 mM (1 µM final concentration) and 20U of the polynucleotide kinase PNK4 of Promega (M4101) in a 1X Buffer concentration. The reaction was incubated at 37°C for 30 min. Then the oligonucleotides were denatured with an incubation at 95°C for 5 min. followed by a cooling down with a ramp of 0.1°C/sec to allow proper annealing. The resulting product was used in a 1:3 ligation reaction with the linearized and de-phosphorylated p426-SNR52p-gRNA.AarI-SUP4t using the Ligation protocol of T4 DNA Ligase of Promega (M1801). The ligation was transformed in NEB DH10B electro-competent cells (C3020K) following the provided protocol. The resulting clones were plated in LB/Amp100 plates. Positives clones were tested for the correct insertion of the desired spacer with an AarI restriction digestion to eliminate negative ligations (digested by AarI). The positive clones were also tested through sequencing.

Quick Change of point mutations in plasmid constructions

The Q5® Site-Directed Mutagenesis Kit Protocol (E0554) was used for minor correction of point mutation in the p426 plasmids constructions. The following reagents were mixed in a PCR tube;

	25 µl RXN	FINAL CONC.
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µl	1X
10 µM Forward Primer	1.25 µl	0.5 µM
10 µM Reverse Primer	1.25 µl	0.5 µM
Template DNA (1–25 ng/µl)	1 µl	1-25 ng
Nuclease-free water	9.0 µl	

The primers for each modification (insertion, deletion, replacement) were designed using the NEBaseChanger tool.

The cycling conditions were the following

STEP	**TEMP**	**TIME**
Initial Denaturation	98°C	30 seconds
25 Cycles	98°C	10 seconds
	50–72°C*	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	

*Depends on the annealing temperature of the primers

For the KLD reaction, a Kinase, a Ligase and the DpnI enzyme interact with the PCR product in order to phosphorylate its ends, ligated it in a circular molecule and eliminate the remaining PCR template. The reaction is the following:

	Volume	Final Conc
PCR Product	1 µl	
2X KLD Reaction Buffer	5 µl	1X
10X KLD Enzyme Mix	1 µl	1X
Nuclease-free Water	3 µL	

The mixture was incubated 5 min at room temperature. The Ligation products were finally transformed in NEB 5-alpha Competent E. coli cells (C2987) were used following the provided protocol. The resulting clones were plated in LB/Amp100 plates. Positives clones were tested by sequencing to verify if they carried the desired modification.

Solutions for the lithium acetate transformation

1. Transformation Buffer Lithium acetate in 1xTris-EDTA (0.1M Liac/1XTE)

- 1 mL Tris-HCl 1M pH7.5

- 20 µL EDTA 0.5M pH7.5

-1,02 g Lithium acetate dehydrate (Sigma-Aldrich # L4158 – BioXtra - M:102.2 g/mol)

- 99 mL sterile water

2. Denaturated carrier DNA

- 10 mg/mL denatured salmon sperm DNA (resuspend in sterile water) (Sigma # D1626). Before using the carrier DNA for the transformation, boil it at 100°C for 10 minutes then keep it on ice 10min to stay in a denatured state.

3. Polyethylene glycol-4000 for membrane permeability (PEG4000)

We need a reagent of 40% PEG4000 diluted in 0.1 M LiAc/1XTE. 20 g of PEG4000 3350 (202444-500G Sigma-Aldrich) are diluted in 30 mL of the transformation buffer prepared above, 0.1M Liac/1XTE. The volume is fixed at 50 mL when all the PEG4000 is dissolved.

Lithium acetate transformation

The method of transforming yeasts by lithium acetate consists of "stripping" their membrane and then causing them to undergo a thermal shock. This method is applied to introduce small circular replicative plasmids and integrative linear fragments in the cell. This protocol is inspired by that of Gietz et al. (1995).

Using yeast cells that are stored at -80°C, a culture is started in the appropriate medium, depending on the yeast strain, until reaching an OD_{600} = 1 (~$2x10^7$ cell /mL). When the culture reach this growth stage (8-10 hours after the initial culture) the cells are diluted, depending on their duplication time, and an Overnight(ON) culture is started in order to reach the same OD early the next morning. For example if the duplication time of a yeast strain is 2h00 and we want the OD_{600} = 1 (~$2x10^7$ cell /mL) at 8h30 in the morning, the calculations are as follows;

- 18h30= 0.0075
- 20h30= 0.0156
- 22h30= 0.03125
- 0h30= 0.0625
- 2h30= 0.125
- 4h30= 0.25
- 6h30= 0.5
- 8h30= 1

(OD of culture)/ 0.0075= number of dilutions

(Volume of ON culture) / (number of dilutions) = Volume of cell culture in the overnight culture

The cultures are incubated ON in agitation at 30°C. The second day of the experiment, when the cultures reach an OD_{600} = 1, they are diluted to ¼ in a rich medium of YPDA. This step allows the yeast to have 2-3 replications in a non-selective medium that allow them to reduce the stress from the auxotrophic medium we use in the non-wild type strains. When the DO_{600nm} is ~ 1, 2.5 mL of culture are transferred into a white capped tube and centrifuge 2 min at 6,000 rpm. Supernatant is removed and cells are resuspended in 1 mL Liac / 1XTE and transferred into a microtube for vortexing and a further centrifugation of 30s at 14,000 rpm. Supernatant is removed and cells are resuspended in 1 mL of 0.1 M Liac / 1XTE for an incubation at 30°C for 30 min to a maximum of 1 hour. This incubation allows the partial degradation of the yeast cell wall. Cells are collected by centrifugation 3 min at 5,000 rpm, supernatant is discarded and the following solution is added to the cellular pellet:

- 50 µL of 0.1 M Liac / 1XTE
- 5 µL of denatured carrier DNA (boiled for 10 min at 100 ° C then cooled quickly on ice)
- 4 µg of CORE PCR cassette, 1 nmol of annealed oligonucleotides or 200 ng of plasmids

After mixing, 0.5 mL 40% PEG4000 / 0.1M LiAc / 1XTE are added. After mixing, a volume of 56 µl DMSO (10% to final volume) is then added and mixed before an incubation of 30 min at 30°C and a temperature shock in a dry water bath for 25 min at 42°C. Cells are then collected by 3 min centrifugation at 5,000rpm and supernatant is discarded.

If we transform yeast without mycoplasma genome inside we skip the next step. Cells are resuspended in 1 mL of YPDA and incubated for 1 h in agitation at 30°C. Finally the cells are collected by centrifugation for 3 min at 5,000 rpm and supernatant is removed. The final cell pellet is resuspended in 300 µL of sterile water.

If transforming yeasts with the CRISPR/Cas9 for modification of mycoplasma genome, the cells should be put in a liquid medium with the auxotrophy selection for all the markers that are introduced in yeast and the culture should be incubated in agitation at 30°C for 48 hours. After the incubation, cells are plated on selective medium for incubation at 30°C for 2 days. If the protocol is not used for modification of mycoplasma genome, the 48h incubation in liquid medium is not necessary.

Transformant subcultures for screening

After 48 h of growth on the agar plates, individual colonies were picked and "patched" on selection medium plates. The "patch" is created by spreading an isolated colony on a new plate in order to increase the number of cells, for further DNA extractions.

For the mycoplasma genome modifications, in order to study the phenomenon we conducted a pooling of yeast colonies, before streaking and "patching". Groups of 20 colonies were marked differently on the petri dished and then a small portion of each colony was picked and pooled in the same extraction tube. After DNA extraction and screening, the pool with the desired genotype among its clones, was traced back to the petri dish and each colony from the group of 20 was analyzed individually.

Solutions for the yeast genomic DNA extraction

- Zymolyase Buffer: For 20 mL of Zymolyase buffer, mix 9 mL of H_2O, 1 mL Tris-HCl 1 M pH7.5, 10 mL of glycerol 50% and 200 mg of Zymolyase 100T (08320931). Prepare aliquots of 500 µL and store at -20°C.
- Potassium acetate (Kac) solution (5M) mix 29.4g of potassium acetate (P1190) and 75 mLH_2O until the powder is complete dissolved. Add 11.5 mL glacial acid acetic and mix until the mixture is homogenous with any solid remains. Filter sterilize with a Stericup™ filter unit 150 mL capacity 0.22µm (PES) membrane sterile provided by The Consumables Company. This reagent can be stored for 12 months at RT.
- Zymolyase /M.E. buffer: This buffer needs to be prepared anew for every experiment. For each sample, mix 100 µL H_2O with 1 µL of β-mercaptoethanol (β-ME) (12.5M SIGMA-98%-M3148) and 10 µL of the Zymolyase buffer prepared above.

Yeast genomic DNA extraction

The protocol used for the extraction of the yeast or the mycoplasma genomic DNA from yeast was the same. First step is to resuspend a 48h "patched" yeast colony in 90 µL of Zymolyase /ME buffer. Only a pinch of a colony at the end of a 10 µL tip is needed for an extraction of 50-100 ng/µL of DNA. The

suspension should be a bit whitish. Incubation is pursued for 1h at 37°C to allow the degradation of the yeast cell wall. Then, 10 µL of 2% SDS are added to denature the cellular membrane of the cells. Mixture is vortexed for 2 sec and incubated 15 min at 70 °C. Then, 11 µL of potassium acetate are added, mixed for 2 sec and further incubated for 15 min on ice to allow the precipitation of proteins. After a spin at 14,000 rpm for 10 min at 4 °C, 90µL of the supernatant is transferred to an empty tube and 90 µL of isopropanol are added and mixed by vortexing for about 30 sec. After centrifugation at 14,000 rpm for 10 min at 4 °C, supernatant is discarded and the pellet is air dried by leaving the tubes open and incubating at 42°C. DNA is resuspended in 100 µL of TE1X.

PCR and Multiplex PCR reactions

For PCR reactions, we used two different kits depending on the application. For simple amplifications, we used the Clontech Advantage 2 PCR kit (111816) and for amplifications where we wanted a polymerase that has a low error rate we used the Q5® High-Fidelity DNA Polymerase (M0491).

For the Multiplex PCR reaction we used the Qiagen Multiplex PCR kit (Cat No. /ID: 206143). We designed 10 pairs of primers covering the entire genome of each species we wanted to test. In the table (Supplementary table S2) are listed the primers used for the genome of Mmc, Mcap and *M. pneumoniae*. For 200 µL of primers mix, each primer was added at a final concentration of 0.5 µM (1 µL from an original tube of 100 µM) and the rest of the volume was completed with milliQ water. The buffer provided by Qiagen was added in a final concentration of 1X. During the cycling procedure, the annealing temperature is usually set really low (50°C) and it lasts 2.5' to ensure proper hybridization of all primers. The PCR products of the reaction were analyzed by a 2% agarose gel electrophoresis.

Agarose plugs preparation

For PFGE analyzes and genome transplantation assays, yeast agarose plugs are prepared using the CHEF Genomic DNA Plug Kits provided by Bio-Rad. We started (first day) a pre-culture of the desired yeast strains in 5 mL of selection medium at 30°C from the patched colony. When the OD_{600}= 1, we prepare an ON culture, as described before, but the volume of the ON culture is increased to 100 mL. We apply the proper dilutions and start the ON culture in agitation at 30°C. The next day (second day), cells at an OD_{600} = 2 are collected by a centrifugation at 5000 rpm for 10 min at 4°C. Then, cells are then resuspended in 10 mL of cold 50 mM EDTA, pH= 8. Using a Malassez cell, the number of cells for each culture is evaluated and then the volume needed in order to have $6*10^9$ cells/mL of plug is calculated. Usually, six plugs are prepared per assay. Then we centrifuge the corresponding volume of cells at 5000 rpm for 10 min at 4°C. The pellet is resuspended in 300 µL of Cell Suspension Buffer. 400 µL are transferred in a new tube and the mixture is equilibrated 10 min at 50°C. After 10 min, cells are mixed with 4 mg of Zymolyase 100T (0832093) and 400 µL of pre-melted 2% Low Melting Agarose (kept at 50°C). After mixing, the mixture is distributed to plug molds (100 µL/well). The blocks are allowed to solidify for 1h at 4°C before being transferred to a 50 mL falcon tube containing 1.5 mL Lyticase buffer (5 mL of 1 M Tris-HCl pH 7.5, 50 mL of 500 mM EDTA pH 8.0, 445 mLof sterile water) and 7.5 mg of Zymolyase 100T (5 mg/mL). The mixture is incubated for 2h at 37°C. The plugs are washed two times with 25 mL of sterile water that is added slowly in the tube. Between washes, a 5 min slow agitation on a

horizontal agitator is performed. The plugs are then incubated in 1.5 mL Proteinase K buffer solution with 60 µL Proteinase K (Bio-rad) (1.25 mL buffer/50 mL proteinase/5 plugs) for 24h. After incubation, (third day) a second one is initiated in the same conditions and concentration of proteinase K. After the second incubation, (forth day) the buffer is discarded and plugs are washed 4 times with 10 mL 1X wash buffer (Bio-rad). Each wash last 1h with slow agitation at room temperature. After the 4 washes, the plugs are stored at 4°C in 1X wash buffer for further use.

Pulse Field Gel Electrophoresis

We used the plugs stored at 4°C in 1X wash buffer prepared before. We usually use half a plug for PFGE. The first step is to digest the most part of the yeast DNA to increase the proportion of mycoplasma genomic DNA in order to have a better image after PFGE.

Plug are washed in 1 mL 1X wash buffer+PMSF (1 mM final concentration) during 1h in agitation. The buffer is removed and 1mL 1X wash buffer is added for a 1h incubation. The buffer is removed and we add 1 mL of 0.1X wash buffer for 1h. The buffer is removed once more and a final wash with 1 mL of 0.1X wash buffer for 1h is performed. We add 0.5 mL of 1X restriction buffer (Cutsmart buffer from NEB) for 1 hour with gentle agitation at room temperature. The buffer is removed and 0.5mLof 1x restriction buffer with the restriction enzymes of interest are added to digest the genome of yeast. The enzymes used are FseI, RsrII and AsiSI, all from NEB. For each plug, 30 units of each enzyme are used. The digestion is incubated overnight at 37°C. After the incubation, the buffer is removed and the plugs are washed for 30 min in 1 mL of 1X wash buffer. Then the plugs are loaded on an 1% agarose TAE1X gel and electrophoresis parameters are set to 120 min at 120V. Plugs are removed before staining the gel in Ethidium bromide to verify the presence of yeast digested DNA inside the gel.

The plugs are then washed in 1 mL of 0.1X wash buffer for 1h. The buffer is removed once more and we apply a final wash with 1 mL of 0.1X wash buffer for 1h. We remove the buffer and 0.5 mL of 1X restriction enzyme buffer (Cutsmart buffer from NEB) are added. We remove the buffer and we add 30 units of the desired restriction enzyme to cut the mycoplasma genome. The restriction mixture is incubated overnight at the temperature corresponding to the enzyme. The restriction buffer is then removed and the plugs are incubated in 1 mL of 1X wash buffer for 30 min in gentle agitation. The buffer is removed and the plugs are put to migrate in a CHEF gel in TBE 0.5X (1.9 L) for 22h. The switch time is 60-120 min, the angle is 120° and the voltage is 6V/cm. The gel is 1% agarose Biorad in 110 mLof 1X TBE. After migration, the gel is colored for 120 min in Sybr Gold (S11494) and then washed in water for 1 hour before taking a picture under UV light.

Plugs methylation

The plugs of Mcap mutants are methylated to be protected from the *M. capricolum* subsp. *capricolum* recipient cell restriction enzymes during the transplantation procedure. Each plug is washed two times for 30 min in 1 mL of washing buffer (200 mM tris-HCL pH 7.5, 50 mM EDTA pH 7.5). Then, they are incubated in the methylation buffer (100 mM tris-HCL pH 7.5, 10 mM EDTA, 3 mM DTT and 200 µM *S*-Adenosyl methionine (SAM)) by incubating two times in 1 mL during 30 min under gentle agitation. The methylation is conducted in 100 µL of methylation buffer supplemented with 6 µL of Mcap extracts (120 µg of extract from MCAP 17.5 softly de-frozen on ice for 15 min). The Mcap extracts contain still uncharacterized methylases that can methylate the mycoplasma genome during an incubation for 16h at

37°C. The plugs are cut in 4 parts for better efficiency. For the rest of the methylation, the reagents are complemented with HCO in a concentration of 0.1 mM to help compact the DNA molecules and protect them from breaking. After incubation the methylation buffer is discarded and the cells are incubated for 4h at 50°C with 1 mL of proteinase buffer and 40 µL of protein (agarose plugs kit). After treatment, the plugs are washed four times for 45 min each with 1X wash buffer (kit) and two times for 30 min with 0.1X wash buffer. After washes, plugs are ready for the genome transplantation.

Genome transplantation

The preparation is the same for all plugs, regardless if a methylation procedure is applied or not. All plugs (100 µL volume of each plug) are incubated in 10 µL of 10X β-Agarase Reaction buffer (NEB) (1X final concentration) for 5 min at 42°C. The temperature is then raised to 65°C in order to melt the agarose during an incubation for 8 min. The reactions are allowed to cool down with an incubation at 42°C for 10 min and 3 µL of β-Agarase (3u) are added to allow the complete degradation of the agarose during an overnight incubation at 42°C.

The recipient strain used for the genome transplantation of Mmc and Mcap is MCAP ΔRE cl17.5 and MCAP wt (CK)respectively. A culture of the cells in SOB medium in an appropriate dilution is started in order reach pH 6.4-6.2 the next day. Twelve milliliters of culture are required per transplantation assay. Two plugs are used for each strain. The cells are centrifuged 15 min at 5800 g at 10°C. Cells are then resuspended in 6 mL of wash buffer (Tris-HCl 10 mM pH 6.5, NaCl 250 mM) and then centrifuged 15 min at 5800g at 10°C. The pellet of cells is resuspended in 400 µL of Anhydrous $CaCl_2$ 0.1M and incubated in ice for 30 min. Ten minutes before the end of the incubation, we start the preparation of the genomic DNA. The DNA from the digested plugs (all the volume from the tube) is transferred softly into a 15 mL falcon tube containing 400µL of SP5 medium without serum. After the end of the incubation, the cells are transferred into the falcon with the DNA, VERY gently. We add the 2X Fusion Buffer (10% PEG_{6000}, 500mM NaCl, 20mM $MgCl_2$, 20mM Tris-HCl, pH 6.5) with a 1/1 volume ratio (1.2 mL) and all reagents are mixed really gently by rotating horizontally the tube. The reaction is incubated for 90 min at 30°C. The process is stopped by the addition of 5 mL of SP5 and the tubes are mixed very gently by inverting one time each tube. Cells are then collected by centrifugation for 15 min at 5800 g, 10°C. The pellet of cells is resuspended in 1 mL of SP5. Cells are finally plated on Petri dishes with the selection medium and the plates (closed with parafilm) are incubated for 3-5 days at 37°C.

Peroxide production assay

To estimate the production of H_2O_2 for the mycoplasma clones we applied the kit of MQUANT ^{TM} following the suggested protocol and the test applied by Pilo et al. (2005). A mycoplasma clone has been grown in 5 mL of SP5 medium until it reached the logarithmic phase (pH 6.6-6.2). Cells are collected by centrifugation at 7.000g / 10 min / 4°C. Culture medium is discarded and cells are resuspended in 10 mL HEPES incubation buffer (HEPES 67.7 mM, NaCl 140 mM, MgCl2 7 mM, pH=7.3). A wash step can be added here with a centrifugation at 7.000g / 10 min / 4°C, elimination of the incubation buffer and resuspension of the cells in 10 mL of incubation buffer. Aliquots of 1 mL of the cell culture are prepared and incubated in starvation condition at 37°C for a minimum 1h. The H_2O_2 production is initiated by

adding glycerol at a final concentration of 100 µM (for 1mL culture= 80 µL glycerol 100%). Using the kit of MQuant TM, we measure the peroxide after 100 min of the addition of glycerol. The kit contains paper strips with an organic redox indicator. This produces a blue oxidation product. The peroxide concentration is measured semi-quantitatively by visual comparison of the reaction zone of the test strip with the fields of a color scale. First the reaction zone of the test strip is immersed in the sample for 1 sec. Then the excess liquid is allowed to run off via the long edge of the strip onto an absorbent paper towel and after 15 sec we determined with which color field on the label the color of the reaction zone coincides most exactly. The result is expressed in mg/L.

Bacterial strains

For CRISPR study, the strain S6 of *Mycoplasma gallisepticum* was used. *Mycoplasma gallisepticum* was grown in a simplified version of Hayflick (HA) medium: For 1L of complete media we used 21g of PPLO broth (255420) 5 g Dextrose (D-glucose) and 695 mL of deionized H_2O. For plates, 10g of Agar Noble were added. The pH was set at 7.85 with 1 M NaOH. The mix was autoclaved for 22 min, at 121°C and 2 atm. The following complement was filtered at 0.22 µm and added in the medium after it was cooled at room temperature: 100 mL of yeast extract, 200 mL of horse serum (inactivated) 100 µL Ampicillin 100 mg/mL and 2 mL of phenol red 1%. Cells were grown in an incubator without CO_2 at 37°C. Cell growth was measured by pH measurement and by de-colorization of the phenol red.

The *Mycoplasma mycoides* subsp*. capri* GM12 strain was used for the application of the MgCRISPR/Cas9 system directly in Mycoplasma cells. *Mycoplasma mycoides* subsp*. capri* was cultivated in SP5 medium.

For all the plasmid constructions, the *Escherichia coli* commercially available cells from NEB DH10B, for electroporation (C3020K) or chemical transformation (C3019H) were used.

Bioinformatic analysis for PAM identification

In order to analyze all the spacers from the *M. gallisepticum* species, we developed an R script (R Core Team 2013). At first, we blasted the sequence of every spacer on different databases and the results for each spacer were saved in a CSV format (Comma Separated Values). This format can be processed by the R software. The size of the sequence that could potentially include the PAM was limited at 15 nucleotides. The specific 15 nt sequence would be extracted by treating the data on each CSV file. We installed the required packages and library of the NCBI database on the R, in order for it to be able to extract data from it.

Then the parameters for the R were defined; the first parameter was the percentage of identity, which was defined equal or more than 90% between the spacers and the blast result sequences, the potential protospacer. Another parameter was the alignment length. We choose to keep the minimum alignment at 15 nucleotides but to increase the likelihood for a positive alignment to be true, the 15 homologous nucleotides should be equally distributed among the 30 nucleotides of the spacer sequence and the protospacer.

If the candidate protospacer provided by the blast research meets all the parameters, the R was designed to extract the 15 nucleotides upstream or downstream from it. It was thus designed to

distinguish the protospacers that have a 5'-3' orientation from the ones that have a 3'-5' and put them in two distinct files. For the second group, a reverse complementation command was included to allow the extraction of the right sequences flanking the protospacer.

An "if" command was created where, when the number defining the start position of the protospacer was higher than the one defining its end (e.g. 100200-100170) the program was order to extract the sequence between the positions "start=end-1" and "end=end-1-PAM SIZE". In the other case, the desired sequence was between the positions "start=end+1" and "end=end+PAM SIZE". The resulting sequences were distributed in two categories, called "direct (dr)" and "reverse complementary (rc)". We also installed the package for the Weblogo software (Crooks et al., 2004), so the R script could provide us directly with a logo demonstrating the prevalence of each nucleotide for every position, from 1 to 15.

Plasmids

Plasmids already existing in the laboratory

pSRT2

This plasmid was used to clone the origin of replication of *M. gallisepticum* into a vector adapted for expression in mycoplasma. It contains the tetracycline resistance under the control of the spirallin promoter, the ampicillin resistance and the origin of replication of *E. coli*, *col*E1.

pPS3.1

This plasmid was used to clone the gRNA we synthesize for the activation and guidance of the MgCas9 in an expression system for mycoplasma. It contains the spirallin promoter followed by the sequence of the fibril terminator. It also has the resistance to ampicillin, AmpR.

pMYCO1-puro

This plasmid was used to carry the expression cassette of the gRNA assembled in the pPS3.1 inside the mycoplasma recipient cells. It contains the origin of replication of Mmc, the origin of replication of *E. coli* and the resistance to ampicillin and pyromicin, the later under the control of the spirallin promoter (Figure S8).

Plasmid constructions during thesis

pMGAL

This plasmid was developed after cloning the origin of replication of *M. gallisepticum* in a the pSRT2 plasmid vector. The cloning site was a BamHI and the cloning was conducted with T4 Ligase of promega (M1801).

pMGAL-spacer-PAM1-4

This plasmid was developed after cloning the spacer sequence with 4 different PAM candidates downstream of it on the pMGAL. The cloning site was an XmaI.

pMGAL-spacer-PAM5-13

13 constructions with different PAM sequences were designed using as a base the pMGAL plasmid. For 8 plasmids, the Q5® Site-Directed Mutagenesis Kit (E0554S) was used to modify the PAM sequence downstream of the spacer of the pMGAL-spacer-PAM1 construct.

p426-SNR52p-gRNA.ICE-SUP4t

This plasmid was developed by cloning the 20bp spacer sequence for the ICE element MMCAP2_0557 on the p426-SNR52p-gRNA.AarI-SUP4t linearized vector (Figure S9).

Construction of plasmids for cleavage studies in Mycoplasma gallisepticum

We used the backbone shuttle vector pSRT2. The origin of replication of *M. gallisepticum* has been identified by *Papazisi et al. 2003* as the area between MGA_0618 and MGA_0619. The Introduction of this genomic region would occur through a BamHI digestion of pSRT2, thus creating the pMGAL plasmid. For this reactions the primers Mgall-ori3 Mgall-ori4 were used (Supplementary table S2). We used the Q5® High-Fidelity DNA Polymerase (M0491) kit to amplify the desired region and then the PCR product was digested with the BamHI enzyme, following the suggested protocol. The pSRT2 vector was also linearized using the same plasmid and both products were ligated following the T4 Ligase of Promega's protocol. The construction was completed with an introduction through an XmaI digestion and Ligation of annealed oligonucleotides carrying a spacer sequence from the M. *gallisepticum S*6 CRISPR locus and a PAM candidate to develop the pMGAL+spacer+PAM plasmid. Different PAM sequences downstream of the spacer sequence were introduced using the Q5 site-directed mutagenesis kit of New England Biolabs. Integrity for all plasmids was identified with sequencing of the spacer+PAM region and restriction digestion.

Construction of plasmids for the expression of the MgCas9 of M. gallisepticum in S. cerevisiae and E. coli

We used the backbone shuttle vector p414-TEF1p-Cas9-CYC1t and pET28 respectively. First we synthesized a codon optimized version of the MgCas9 for higher expression in yeast. Then we linearize the p414 and PET28 plasmid and also we removed the sequence of the TEF1-SpCas9 from the first one with an internal PCR. Finally we assembled the two vectors, called from now one p414-MgCas9 and pET28-MgCas9 by following the In-Fusion HD Cloning protocol by Takara.

Construction of plasmids for the expression of the chimeric gRNA in Mmc cells

We used the backbone shuttle vector pPS3.1 and later the pMYCO-pspuro. pPS3.1 is a plasmid that contains the sequence of the Spirallin promoter and the fibrille terminator. We initially wanted to introduce the gRNA in between those two sequences and then clone the "mini-gene" inside pMYCO1 to have expression in Mmc cells. The gRNA was synthesized by IDT-DNA as a gblock. We amplify the gblock and add homologous to pPS3.1 extremities that will allow a Gibson assembly reaction with the linearized pPS3.1 vector.

After the Gibson assembly, we amplified the "SP-gRNA-Fibril" construction with primers adapted to the in-fusion protocol of Clontech. At each end, primers were designed to add 15 bp homologous to

the pMYCO1-pspuro sequence. The pMYCO1-pspuro plasmid was amplified in a similar way; primers with were design with 15 bp floating tails, homologous to the SP-gRNA-fibril cassette. The PCR products were purified using the GE Healthcare DNA purification kit. We proceeded by following the Clontech in-fusion protocol. Resulting molecules were transformed in NEB® 10-beta Competent *E. coli* (High Efficiency) following the suggested protocol and plated the transformants were plated on LB/Amp100/puro125 plates. The positive colonies were isolated and all the elements on the plasmid (origin of replication, puromycin gene and gRNA cassette) were verified by sequencing. The integrity of the plasmid was also verified with restriction digestions.

Assembly and expression of the optimized version of the MgCas9

The codon-optimized version of the MgCas9 was divided in 4 gene blocks, equal in size and it has been order for preparation by the IDT DNA Company. All fragments shared 20 bp of homology in each end of their sequence. When the blocks arrived, we conducted an overlap PCR to assemble them together to receive the complete sequence of the Mgcas9 gene. In an overlap PCR multiple fragments can be assembled together as long as they share at least 20 bp of homology in their respective ends. This 20 bp can hybridize to create a double stranded sequence that allows the fixation of the polymerase and the amplification towards both directions. A PCR using 30 ng of each fragment as template was conducted following the Q5 High fidelity indications concerning the annealing temperature and the elongation duration. The annealing temperature was 54°C, the lowest Tm among the primers flanking the gene and the "pseudo-primers" created between each fragment thanks to the homology ends. The product was not clear and we decided to apply a T/A cloning using the PGEM-T Easy cloning kit of Promega.

T/A cloning

The T/A cloning allowed the isolation of a desired fragment of an heterologous mix by adding adenine bases at the end of each fragment, followed by a ligation on a linearized vector with T overhangs. The mix of ligated molecules is then transformed in recipient cells and each clone can be analyzed individually to identify the desired one. The PCR product of the overlap PCR was incubated 10 min at 72°C with the polymerase of the Clontech Advantage 2 PCR Kit.

The ligation was conducted following the pGEM-T protocol. The resulting molecules were transformed in NEB® 10-beta Competent E. coli (High Efficiency) and the transformants were plated in LB/Amp100 medium. Analysis of the positive clones with enzymatic digestion allows identifying the clone with the desired profile. The MgCas9 was amplified using primers that add the desired overhangs for the in-fusion protocol.

ICE excision and circularization study

The ICE excision and the formation of the circular form inside the bacterium cell were verified with two PCRs; primers ICE_excision_ver_F & R and primers ICE_circularization_ver_F & R were used. The amplification products of both reactions were verified with a sequencing procedure.

Bacterial transformation

The *M. gallisepticum* cells were transformed based on the protocol developed in our laboratory by Carole Lartigue (unpublished); A culture of *M. gallisepticum* cells was initiated a day before transformation. The *M. gallisepticum* cells reach their logarithmic stage when the pH is between 6.27 and 6.40. When the culture reached this pH we proceed with the transformation protocol. 10 mL of cell culture per transformation reaction were centrifuged for 15 min at 4700 rpm at 10°C.The pellet was resuspend with 3 mL of commercial HBSS washing buffer. Cells were centrifuged one more time for 15 min at 4700 rpm at 10°C. The pellet of cells was resuspended in 250 µL of $CaCl_2$ 0.1 M. Cells were incubate 30' to 45' in ice. During this incubation the next step was prepared: in a 50 mL falcon tube, 20 µg of DNA were put together with 10 µg of Yeast tRNA. For the 2 plasmid method, 10 µg of each plasmid were used. After the incubation the cells were transferred in a 50 mL falcon tube. 2 mL of 40% PEG6000 (11130) dissolved in HBSS and filtered at 0.22µm was added in the mix of cells and DNA. We shooked the tubes gently and a maximum incubation of 2 min at room temperature allowed the transformation to takes place. The reaction was stopped by adding 20 mL of HBSS washing buffer. We mix the reaction really well and the tubes were centrifuged for 15 min at 10,000g at 10°C. The cells were then resuspended in 1 mL of warm (37°C) Hayflick medium and the plasmid was allowed to be expressed with an incubation for 2h at 37°C. Cells were finally plated in selection medium and were incubated 15 days at 37°C (without CO_2). Positive control was the pMGAL plasmid without the spacer sequence neither the PAM. For the negative control there were no DNA at all. Three passages were required in order to verify the viability of the clone. The first passage was in 200 µL of SP5/tet5 and the next 2 passages were conducted in 1 mL of SP5/tet10.

The Mmc cells were transformed based on the following protocol: 4 mL of a cell culture grown until a pH value between 6.8 and 6.3 were centrifuged for 15 min at 4700 rpm, 10°C. The pellet was then resuspended in a 3 mL of a wash buffer of Tris 10 mM and sucrose 0.5 M (pH=6.5), also called S/T. The same centrifugation as before was applied to the resuspended pellet (15 min at 4700 rpm, 10°C).The pellet of cells was resuspended in 250 µL of CaCl2 0.1M and the cells were incubate 30 min on ice. During the incubation, we prepared the next step which is the transformation. In 50mL falcons tubes we added 10 µg of Yeast tRNA (1 mg/mL) and the corresponding volume of each plasmid to have 10 or 20 µg. After the 30 min incubation, we transferred the cells to the tube and after a gentle mixing of the cells, the plasmid and yeast tRNA we proceeded to the transformation. We added in the mix 2 mL PEG (8000) 70% dissolved in S/T buffer and allow the interaction with the rest for a maximum of 2 min in contact at room temperature. It is important to mix well the tube during the 2 min of incubation. Immediately after the 2 min, we added 20 mL S/T buffer to stop the reaction. The cells were then centrifuged for 15 min at 12,000xg at 10°C. After the centrifugation the supernatant was discarded and the pellet was resuspended in 1 mL of warm (37°C) SP5. The tubes were then incubated at 37°C for 2 hours to allow expression of the plasmid. After the incubation, cells were diluted and the non-diluted reaction together with the 10^{-1} and 10^{-2} dilutions were plated to the appropriate medium, SP5/puro8. The selection continued in a similar way as for *M. gallisepticum* cells, with 3 passages in selection medium. For the Mmc cells, all passages were conducted in SP5/puro8.

DNA manipulations

Extraction of DNA from the mycoplasma cells was conducted in two different ways, either by using a thermic denaturation protocol or, in order to have a more purified product, we used the Wizard Genomic DNA Purification Kit of Promega. For the first protocol, 300 µL of cell colony were centrifuged for 10 minutes at 7000rpm at 4°C. The supernatant was discarded and the cells were resuspended in 100 µL of 1xTE. The cells were then lysed by incubation for 10 min at 95°C. The cells were then diluted to tenth and 1 µL was used for a Polymerase chain reaction (PCR) based on the kit of Advantage 2 of Clontech. DNA purification was conducted based on the wizard kit of Promega. Agarose gel electrophoresis were performed according to the standard techniques.

For the genomic DNA extraction using the Promega kit, there are 4 distinct steps. The first is the cellular lysis. 8 mL of Mycoplasma culture are centrifuged at 4700 rpm 15min 10°C. The pellet is resuspended in 600 µL of Nuclei Lysis Buffer and transferred to an Eppendorf tube. The mixture is incubated 5 min at 80°C or more, until the pellet is completely dissolved. The tube is cooled down at room temperature for 15'. Then we need to eliminate the RNA; 3 µL of solution RNase are added to the cellular lysate and the reaction is incubated 30 to 60 min at 37°C. The tube is cooled down once more at room temperature for 10'. Then the de-proteinization of the mixture is required; 200 µL of Protein Precipitation Solution is added and the tube is vigorously shaken during 30sec. The mixture is incubated 10 min in ice, followed by a centrifugation for 15 min at 14000 rpm at 4°C. The supernatant is transferred to a new tube. Finally we need to precipitate the DNA. 600 µL of isopropanol are added in the tube and the tube is vortexed vigorously until traces of DNA start to appear. The DNA is centrifuged for 10 min at 14000 rpm at 4°C. The pellet is washed with 200 µL of ethanol 70% and the tubes are re-centrifuged for 10 min at 14000 rpm 4°C. The supernatant is completely removed and the tubes are dried with kimwipes paper. All the traces of ethanol are eliminated during an incubation of the open tube for 3-5 min at 37°C. 80 µL of DNA Rehydration Solution are added in the tube and the resuspension is conducted with an incubation for 1 hour at 65°C or ON à 4°C. After controlling the concentration of the DNA at 260 nm, the DNA extract is stored at -20°C.

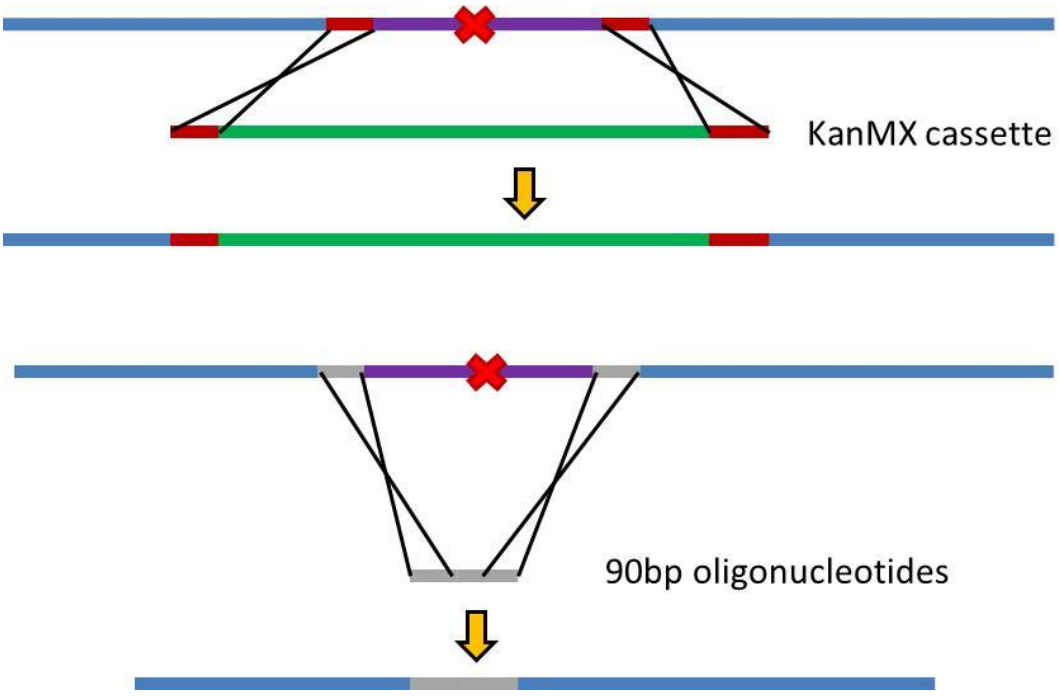
Introduction of the MgCas9 on the Mmc genome using the CRIPSR/Cas9 tools

Using the protocol already described before, we introduced the MgCas9 inside two different genes in the Mmc genome. Instead of using 1 nmol of oligonucleotides, we used 4 µg of the MgCas9 cassette as recombination template. The Cas9 was initially amplified from the *M. gallisepticum genome* using the Q5 High-Fidelity DNA polymerase protocol. The primers used had 5' tails that added 40 bp of free overhangs on each side of the MgCas9. These ends were complementary to the regions flanking the target site on the genes of Mmc genome. 200 ng of the plasmid p426-SNR52p-gRNA.ICE-SUP4t and p426-SNR52p-gRNA.glpO-SUP4t were used to generate a DSB in the sequence of the *glpO* gene and MMCAP2_0557 from the mobile element ICE of Mmc, respectively. After verification of the correct introduction in both sites by PCR, we verified the mycoplasma genome integrity by preparing agarose plugs and applying a Multiplex PCR and a PFGE as described before. Finally the agarose plugs were used in a transplantation procedure to introduce the *MgCas9-in-glpO-Mmc* and the *MgCas9-in-ICE-Mmc* genome into mycoplasma recipient cells as previously described. We obtained clones of Mmc and a PCR analysis verified the successful introduction of the modified genomes in both sites.

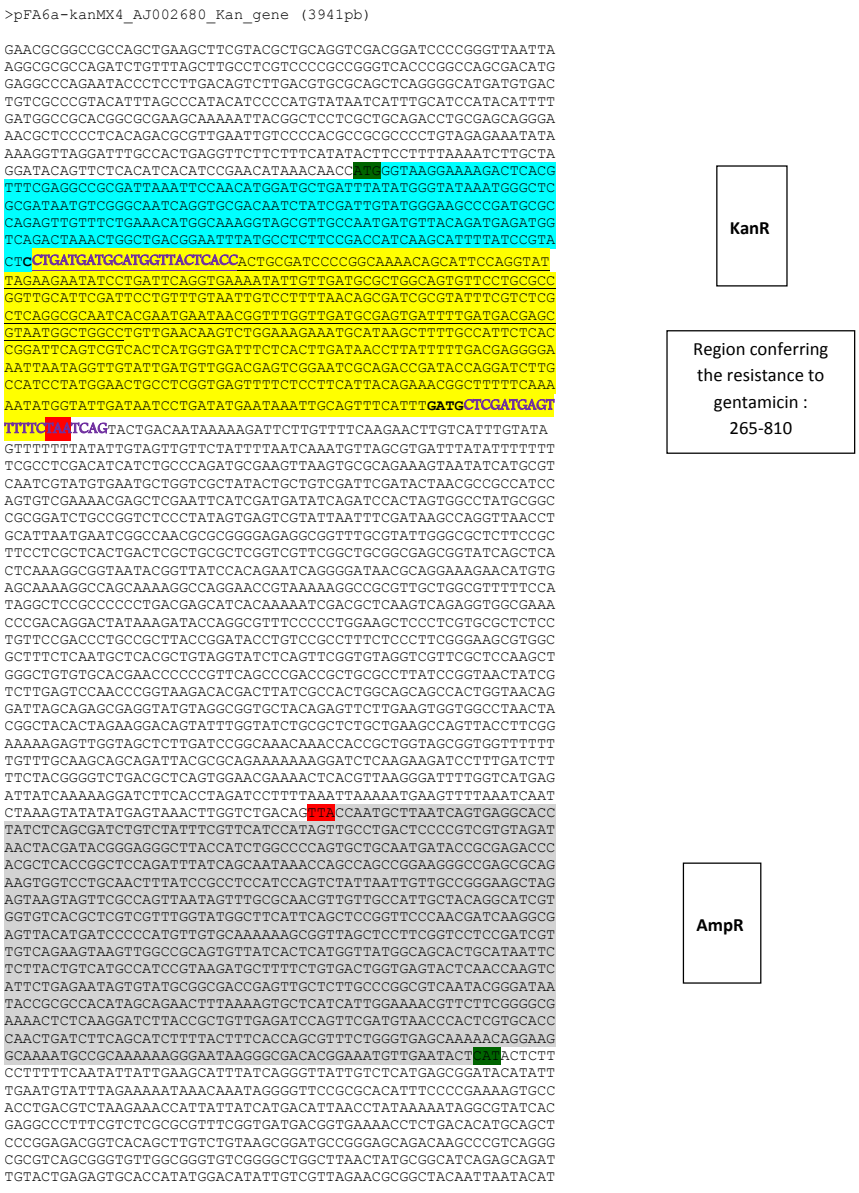
Expression of the MgCas9 by the Mmc clones.

To verify the expression of the MgCas9, an RT-PCR protocol was applied in the *MgCas9-in-glpO-Mmc* and the *MgCas9-in-ICE-Mmc* clones. Both clones where grew until they reach a good moment in their logarithmic phase (6<pH<6.2) and then the total RNA was extracted following the protocol of Invitrogen with the Trizol extraction. Total RNAs were treated in Dnase RQ1 following the Invitrogen protocol (# 9PIM610). Total RNAs were then subdued to an RT-PCR reaction following the SuperScript™ III Reverse Transcriptase protocol for cDNA synthesis. Both the approaches of gene specific primer and 15N Random Primers were tested.

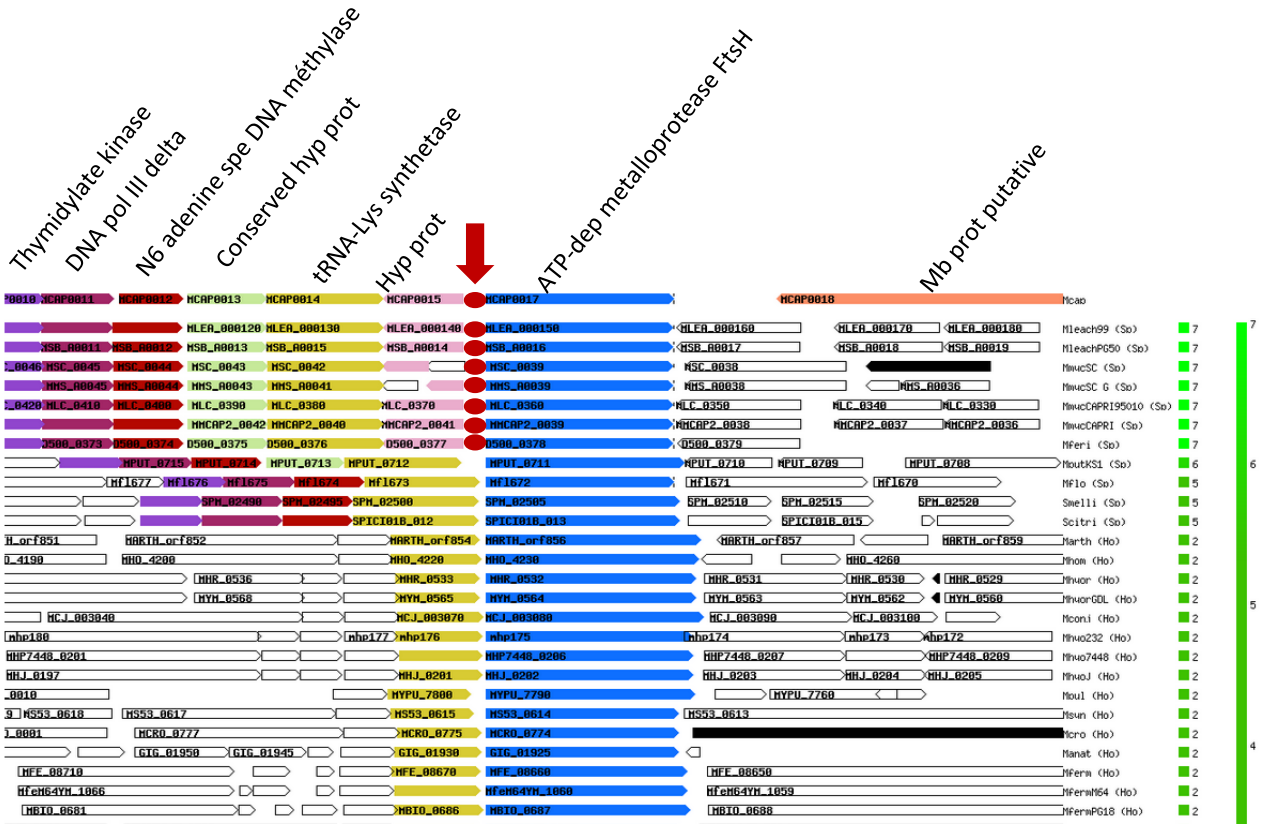
Annexes



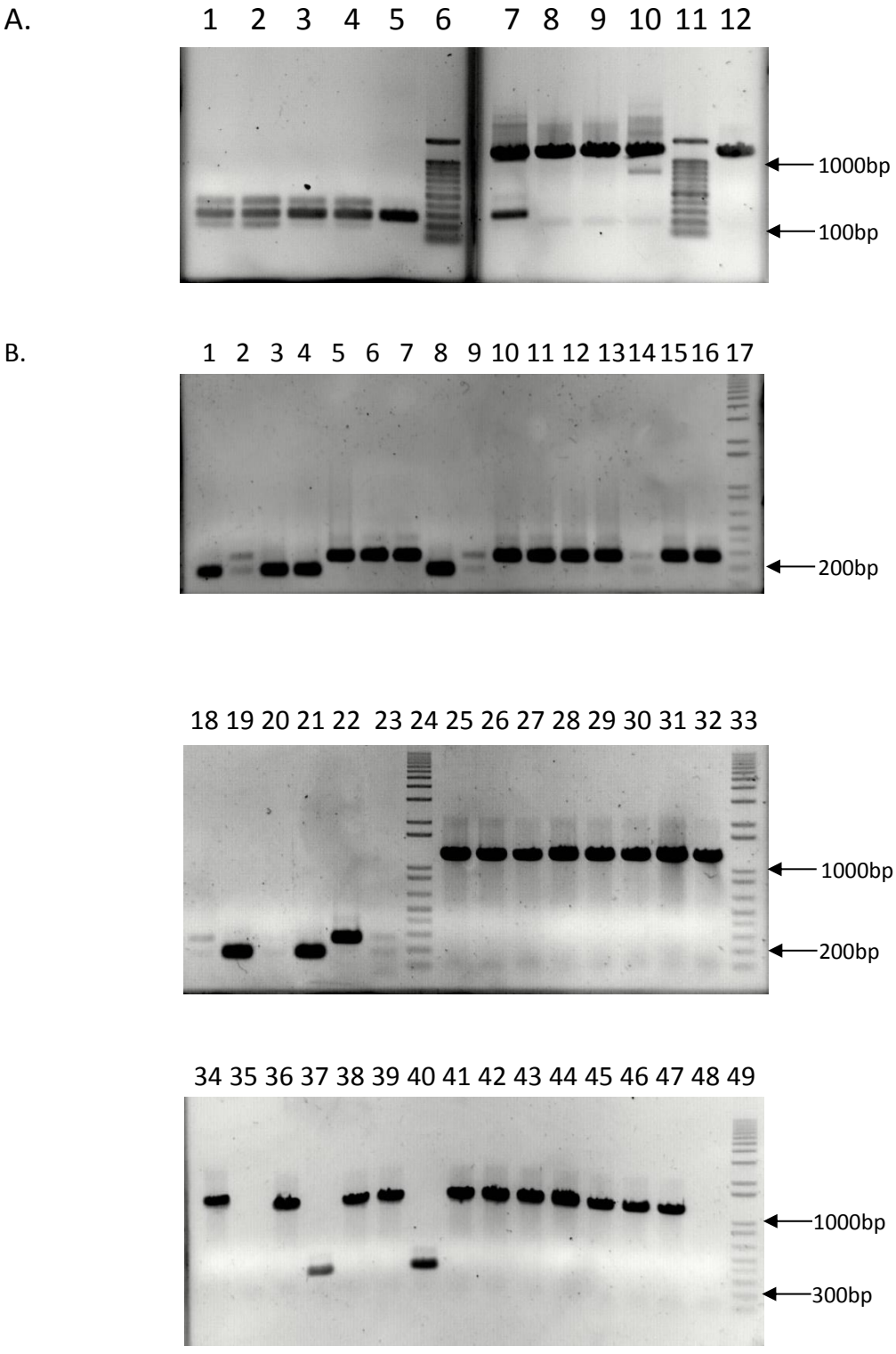
Supplementary Figure 1. The two strategies for the targeting of the glpO gene on the Mmc genome: At the top the glpO gene (purple colored) is cleaved on the marked site (red X) and the recombination tremplate, the KanMX cassette (green colored) is introduced to initiate an homologous recombination that will replace the glpO gene with the selection marker. The sequences in red are the complementary ends added to the KanMX cassette through a PCR. In the second essay, the first step of the procedure remains the same, but the recombination template is now consisted of two hybridized oligonuclotides that are homologous by 45nt to each sequence flanking the targeted gene. A succeful recombination will result in the seamless deletion of teh candidate gene



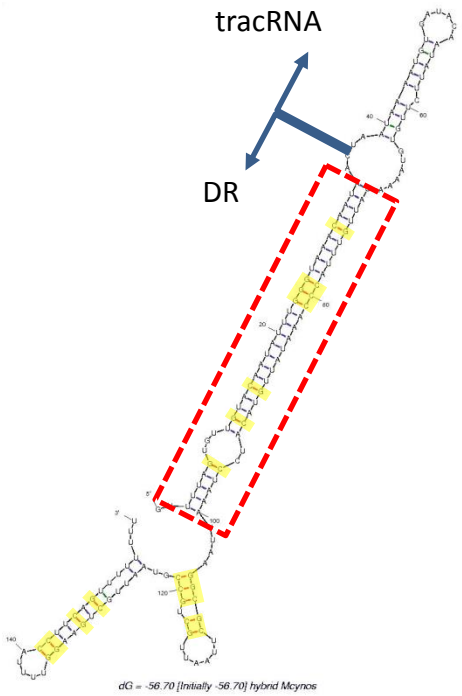
Supplementary Figure 2: plasmid FA6a-kanMX4_AJ002680_Kan_gene



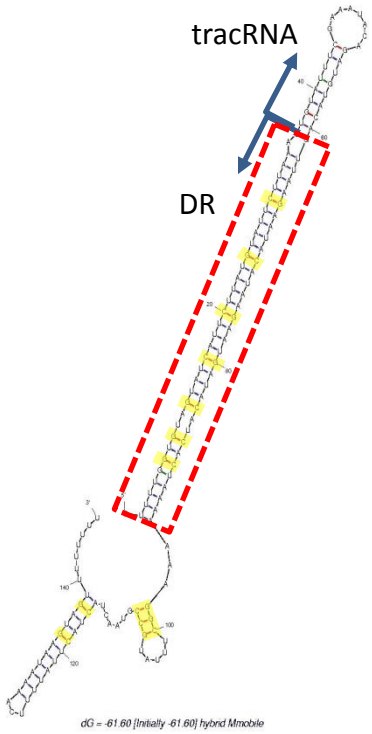
Supplementary Figure 3: MCS2 encoding gene conservation. The red circle indicate the location of MCS2



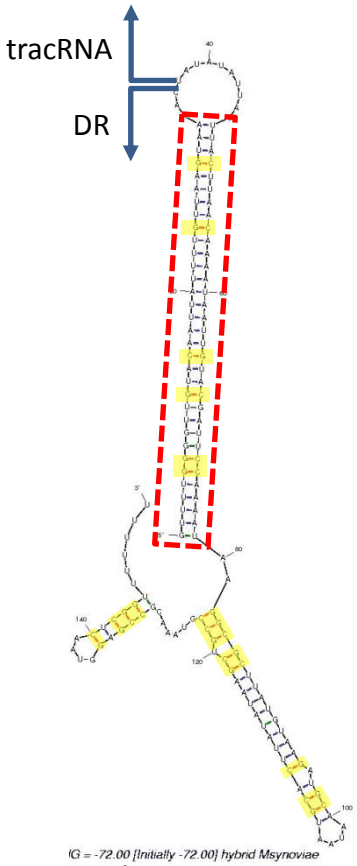
Supplementary Figure 4. A. Pools of the deletion of the MCS2 and both the MCS2 and MCAP0015 (M+M). The bands of about 200 and 250bp respectively indicated the presence of ΔMCS and ΔM+M mutants. 6 and 11, 100 bp-ladder (Promega); 5 and 12, positive control DNA from Mcap; B. Gel electrophoresis of PCR products obtained from individual clones present in the positive pools 2 and 1 of ΔMCS and ΔM+M mutant pools respectively; 22 and 47, positive control DNA from Mcap; 23 and 48, H2O negative control; 17,24,33 and 49, 1Kb+ bp-ladder (Thermo)



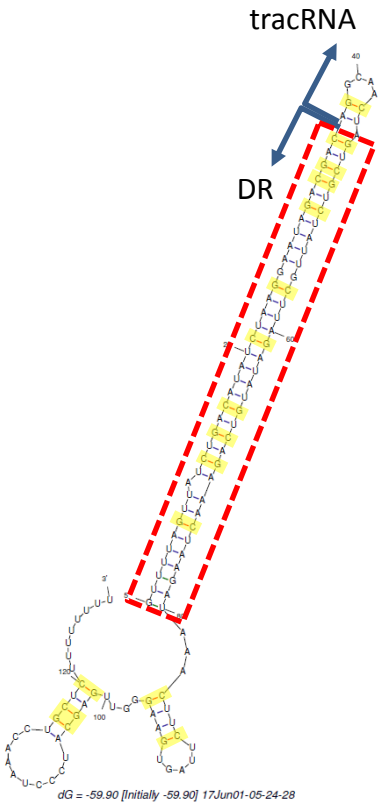
Mycoplasma cynos C142



Mycoplasma mobile 163K

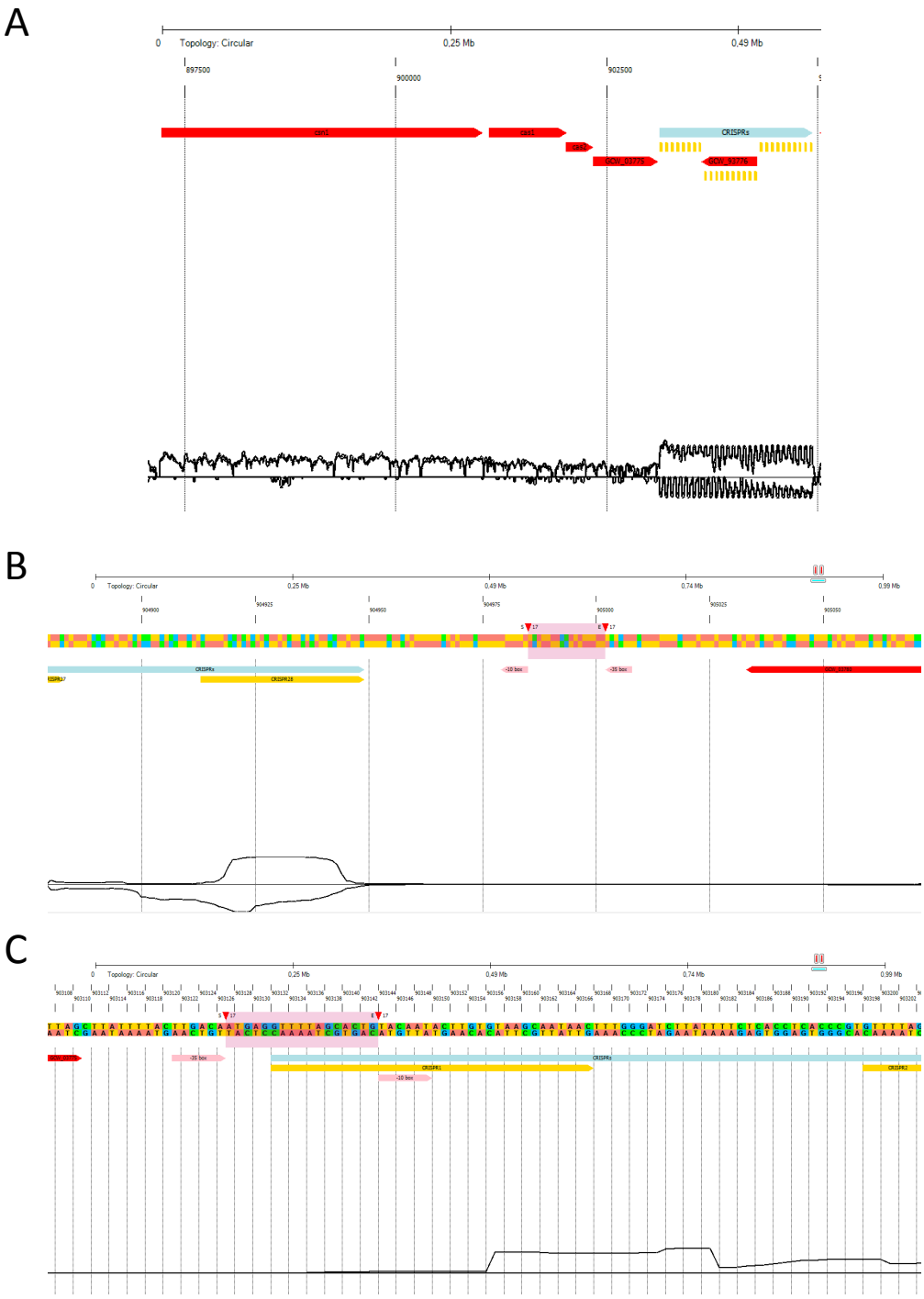


Mycoplasma synoviae 53

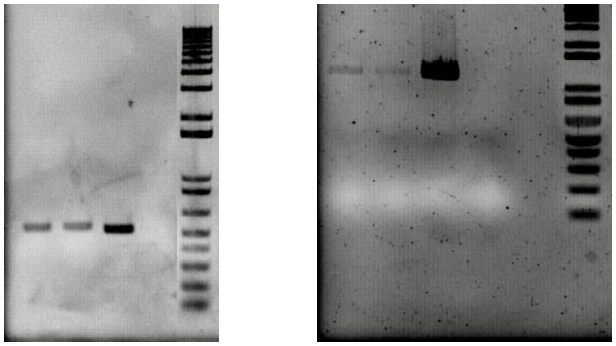


Spiroplasma apis B31

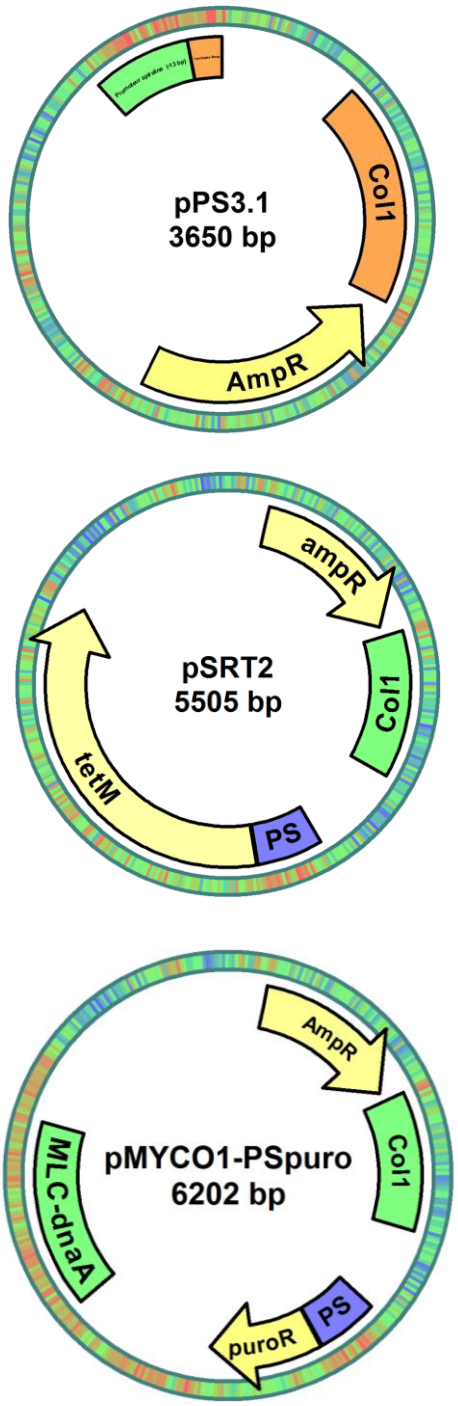
Supplementary Figure 5. Predicted DR/tracrRNA hybrid secondary structure. Sequences of DR and tracrRNA were concatenated and the secondary structures of the hybrids were simulated using mfold software at http://unafold.rna.albany.edu/. Position of the DR/tracrRNA concatenation are indicated by divergent arrows. Predicted stem-loops involving DR/tracrRNA pairing are framed in red dotted lines. G-C pairs were highlighted in yellow. DR and tracrRNA sequences were defined based on Chylinski 2014 and our own work.



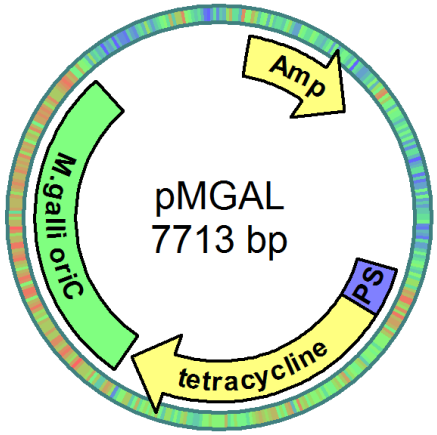
Supplementary Figure 6; Transcriptome analysis of *M. gallisepticum* S6 CRISPR locus. ; A. The expression of all cas genes and the CRISPR locus; B. The forward promoter of the CRISPR locus, C. the potential reverse promoter of the CRISPR locus

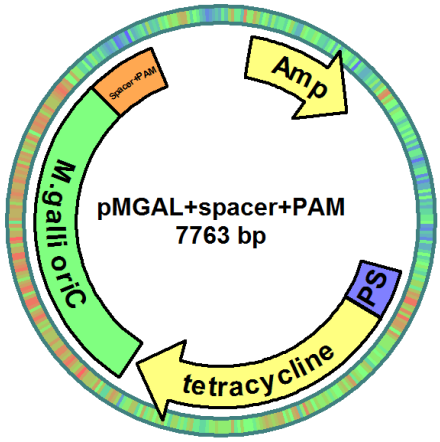


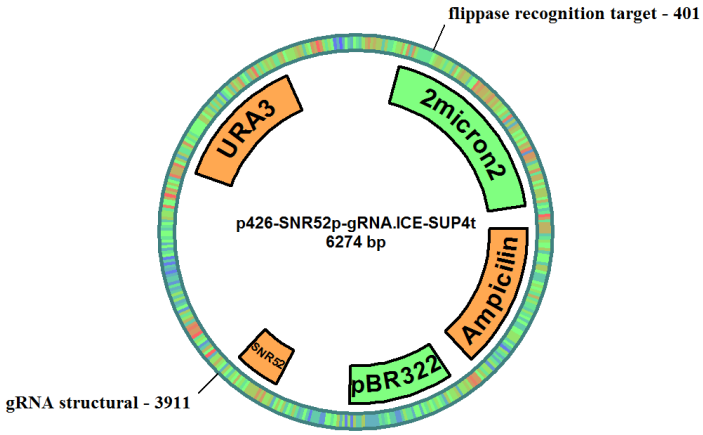
Supplementary figure 7. ICE circularization and excision verification on Mmc genome; A. Amplification of the area around the excision of the ICE on the mycoplasma genome in three different strains, wt, Mmc-YCP, Mmc-Cas_{ice} ; B. Amplification of the area around the circularization point of the ICE of Mmc in three different strains, wt, Mmc-YCP, Mmc-Cas_{ice}



Supplementary figure 8: Plasmids pPS3.1, pSRT2 and pMYCO1-Pspuro layout







Supplementary Figure 9: Plasmids pMGAL, pMGAL+spacer+PAM and p426-SNR52p-gRNA.ICE-SUP4t layout

Colonies count	**noDNA**	**20µg pMYCO1-pspuro**		**20µg pMYCO1-pspuro-gRNA**	
		1st Replicat	2nd Replicat	1st Replicat	2nd Replicat
N.D.	0	2	35	4	8
10^{-1}	0	11	1	59	36
Colonies tested	-	12	12	12	12

Colonies count	**noDNA**	**20µg pMYCO1-pspuro**		**20µg pMYCO1-pspuro-gRNA**	
		1st Replicat	2nd Replicat	1st Replicat	2nd Replicat
N.D.	0	5	-	-	1
10^{-1}	0	7	10	17	39
Colonies tested	-	12	10	12	12

Supplementary table 1. The result from the essay for the activation of the MgCas9 by the gRNA1 and gRNA2: The pMYCO1-pspuro is the positive control for these essayes. All the colonies tested are marked in the lower part of each table. The ND conditions are not trustworthy for Mmc transformation, so colonies were picked from these conditions only when it was in-evitable.

Primer name	Sequence	Product size	Tm	Role
can1.Y.KanMX4.F	AATTGTATCCATTGCGCTCTTTCCCGACGAGAGTAAATGGCGAGGATACGCGGATCCCCGGGTTAATTA	1573bp	61,5°C	Amplify KanMX4 cassette with homologous ends to the CAN1 gene
can1.Y.KanMX4.R	TGCTTAAGCTCTCTCTTCACTTCAGCGTTCTGTACTTCTCCTTCATCTTCGAATTCGAGCTCGTTTTCGA		59°C	
can11F	AGACGCCGACATAGAGGAGA	522bp	60°C	Verify the proper insertion of the KanMX cassette inside the sequence of CAN1 locus in the 1st experience with HR KanMX cassette
KanMX_1R	AGAACCTCAGTGGCAAATCC		59°C	
KanMX_2F	TGGTCGCTATACTGCTGTCGA	537bp	62°C	
can12R	ACCCAGAACTCGAATTCACC		59°C	
p426R	AGAAAGTGATAGCGTTGATGATTCTTCATT	3122bp	5.11E+01	Amplify the URA part of p426 with glpO extremities
gRNA_glpOF	ATTGTAAATTCTGTAGTATTTAGTTTTAGAGCTAGAAATAGC		5.46E+01	
p426F	ATAGAAGAAACCGTTCATAATTTTCTGACC	3092bp	5.11E+01	Amplify the Amp part of p426 with glpO extremities
gRNA_glpOR	GCTATTTCTAGCTCTAAAACTAAATACTACAGAATTTACAATGATCATTTATCTTTCACTGC		6.09E+01	
Del_glpO	AATGTTAAGATGTATTTTTTACTATCTGTCATAGTTATTCTCCTTAATATTCTTTTTTAAAAATTCAAAATTTTATAAGAATTACTTGA	Not for PCR reaction	5.97E+01	Oligonucleotides to delete the glpO gene of Mycoplasma
Del_glpOcomp	TCAAGTAATTCTTATAAAATTTTGAATTTTTTAAAAAAGAATATTAAGGAGAATAACTATGACAGATAGTAAAAAATACATCTTAACATT		5.97E+01	
Del_glpo_KanF	AATGTTAAGATGTATTTTTTACTATCTGTCATAGTTATTCTCCTTCGGATCCCCGGGTTAATTAA	1573bp	61,5°C	Remplace the glpO gene of Mycoplasma with KanMX4 resistance cassette
Del_glpO_KanR	TCAAGTAATTCTTATAAAATTTTGAATTTTTTAAAAAAGAATATTGAATTCGAGCTCGTTTTCGA		59°C	
AarI_gRNA_modF	GGGCAGGTGGACACCTGCCTGTTTTAGAGCTAGAAATAGC	in combination with primer p426R: 3122bp	64.6°C	Create the multimodified p426-SNR52p-gRNA.AarI-SUP4t plasmid
AarI_gRNA_modR	GCTATTTCTAGCTCTAAAACAGGCAGGTGTCCACCTGCCCGATCATTTATCTTTCACTGC	in combination with primer p426F: 3092bp	67.8°C	
QC_gRNA_AarI_F	TGATTACATGtACGTTTGAAGTAC	6.40E+03	57°C	Correction of p426 AarI gRNA constructions
QC_gRNA_AarI_R	CCTTGTATATACTCGAAAGAAAAC		58°C	
glpO_AarI_comp_F	5' ATGATCTGTAAATTCTGTAGTATTTAGT 3'	Hybridisation primers for AarI vector modification	44°C	Delete the glpO gene of Mycoplasma
glpO_AarI_comp_R	3' AGACATTTAAGACATCATAAATCAAAAT 5'		46°C	
gDNA_verif_F	GTTCGAAACTTCTCCGCAGT	4.90E+02	58°C	Verfication of p426 gRNA sequence alternation
gDNA_verif_R	GTTTTCCCAGTCACGACGTT		58°C	
glpODel_verif_F	AAGCTCTTGCACTTGTTGTT	With deletion:480 Without deletion:1647	54°C	Verify the deletion of glpO
glpODel_verif_R	TCCAGCAATGGCATTATTCA		58,5°C	
MCS2_AarI_comp_F	5' ATGATCGGCACCCCTATGCTGGAAGGGT 3'	Hybridisation primers for AarI vector modification	74,8°C	Target MCS2 for Cas9 Double Strand Break
MCS2_AarI_comp_R	3'AGCCGTGGGGATACGACCTTCCCAAAAT 5' (5'TAAAACCCTTCCAGCATAGGGGTGCCGA 3')		73°C	
MCAP15_AarI_comp_F	5' ATGATCTGAAGCTATTTATAATATGAGT 3'	Hybridisation primers for AarI vector modification	59°C	Target MCAP0015 for Cas9 Double Strand Break
MCAP15_AarI_comp_R	3'AGACTTCGATAAATATTATACTCAAAAT 5' (5'TAAAACTCATATTATAAATAGCTTCAGA 3')		61°C	
Del_MCS2	5'ATCAATGTAGTTTGAAAATGATATTTTCTATGTTATTATAATAGTTTTTTTAAAGTGGTGATTTTATGACTTATTATCTTAATTTTTATC	Not for PCR reaction	95°C	Delete the MCS2 gene of Mycoplasma capricolum
Del_MCS2omp	5'GATAAAAATTAAGATAATAAGTCATAAAATCACCACTTTAAAAAAACTATTATAATAACATAGAAAATATCATTTTCAAACTACATTGAT		95°C	
Del_MCAP0015	ATGCTGGAAGGAGGCTATGCCTCCTTTTTTTAAAGTGGTGATTTTTAGTGTTTTTTAACTTTGTTTAGAATAATTTTATCTTTAATGTTA	Not for PCR reaction	95°C	Delete the MCAP0015 gene of Mycoplasma capricolum
Del_MCAP0015omp	TAACATTAAAGATAAAATTATTCTAAACAAAGTTAAAAAACACTAAAAATCACCACTTTAAAAAAAGGAGGCATAGCCTCCTTCCAGCAT		95°C	
DEL MCS2-0015 F	ATCAATGTAGTTTGAAAATGATATTTTCTATGTTATTATAATAGTTAGTGTTTTTTAACTTTGTTTAGAATAATTTTATCTTTAATGTTA	Not for PCR reaction	95°C	Delete the MCS2 and the MCAP0015 gene of Mycoplasma capricolum
DEL MCS2-0015 R	TAACATTAAAGATAAAATTATTCTAAACAAAGTTAAAAAACACTAACTATTATAATAACATAGAAAATATCATTTTCAAACTACATTGAT		95°C	
Del_MCS2_verif_F	GTTTCTCCAGTCTATTTAATGTCCA	With deletion:480 Without deletion:1647	62°C	Verify the deletion of MCS2
Del_MCS2_verif_R	AGCATTTCCACTTTCATCAAGA		62°C	
Del_MCAP0015_ver_F	AAGACGTAGGGATTGAAGA	With deletion:480 Without deletion:1647	59°C	Verify the deletion of MCAP0015
Del_MCAP0015_ver_R	GATTAGTAATGATTATGAAGCATATAAATC		57°C	
DelMPN_KanMF	5'AGCCACCAGTGCAACCACCTAAAAAACCCGCTTAGTATTTCGGATCCCCGGGTTAATTAA 3'	1.57E+03	61,5°C	Remplace the MPN142 gene of Mycoplasma with KanMX4 resistance cassette
DelMPN_KanMR	5'CCTTTTTACAGTTGTGCTTCTTTCTGTTGGGGCTTAATCGGAATTCGAGCTCGTTTTCGA 3'		59°C	
Oligos Del MPN142 Forward	5'TCCTAAGCCACCAGTGCAACCACCTAAAAAACCCGCTTAGTATTTCGATTAAGCCCCAACAGAAAGAAGCACAACTGTAAAAAGGTTGTG 3'	Not for PCR reaction	95°C	Delete the MPN142 gene of Mycoplasma penumoniae
Oligos Del MPN142 Reverse	5'CACAACCTTTTTACAGTTGTGCTTCTTTCTGTTGGGGCTTAATCGAAATACTAAGCGGGTTTTTTAGGTGGTTGCACTGGTGGCTTAGGA 3'		95°C	
MPN142-5_AarIcomp_F	5' ATGATCGGTATCAGTCGGTTCATCGGGT 3'	Hybridisation primers for AarI vector modification	71°C	Target MPN142 for Cas9 Double Strand Break
MPN142-5_AarIcomp_R	3'AGCCATAGTCAGCCAAGTAGCCCAAAAT 5' (5'TAAAACCCGATGAACCGACTGATACCGA3')		70°C	
MPN142-3_AarIcomp_F	5' ATGATCGATCCGAACTCGTTGTGTCCGT 3'	Hybridisation primers for AarI vector modification	72°C	Target MPN142 for Cas9 Double Strand Break
MPN142-3_AarIcomp_R	3'AGCTAGGCTTGAGCAACACAGGCAAAAT 5' (5'TAAAACGGACACAACGAGTTCGGATCGA3')		71°C	

Supplementary Table 2: All primers used in this thesis

5A	TATTTACCGACGAAATTAATACC	52bp	51,5°C	
3A	ATTTTCCTATATACCACTTTCTTTTTC		51,9°C	
5K	AGTAGTCTTTGATAATGGCTAAGG	83bp	55,8°C	
3K	CCTGTATGAGGGCTTTCAG		56,1°C	
5B	CTTAGAACTTTACAGCTCCAAAC	159bp	54,2°C	
3B	CTGGTTATTGGCCACCAAC		56,9°C	
5L	GTGCTTGACTGTGAGACATACA	189bp	59,9°C	
3L	AATCGGCGAACAGCC		55,2°C	
5C	ATGGTGGGATTGCCC	265bp	54,4°C	
3C	ATATTTGGACAGTTTTTCGCC		53,6°C	
5M	TGCACCAACTCCAGCA	285bp	57,7°C	
3M	ATATCCAATAGTTCATTCTTATTGG		52,2°C	
5D	CCGAAAGTTGAGAAGTTAAAGG	349bp	53,3°C	
3D	AGAAATATTTGAAATTTTTATCTAAAAAGC		49,5°C	
5N	GAAGCGGAAAAACGGC	381bp	52,8°C	
3N	CAATTAATGGAAGAATTTTTATTTTCATT		49,6°C	
5E	AATCTCCTCTTGTTTTAATGGAG	461bp	53°C	
3E	TTGCAAGCGATTTTGTG		50,5°C	
5O	ACAAAACAAACACCACCACG	489bp	56,8°C	
3O	CGGCGTGATGATTCATC		55°C	Multiplex PCR for M penumoniae
5F	AAACCTATGCAAATATTTTAACGAT	551bp	51,8°C	
3F	ACTTGTAAAAGTAAAGAACCACTGC		56,4°C	
5P	AATGCTACCCCAAACGGT	587bp	56,9°C	
3P	TGAGCTTTATTGCCATCCTTT		55°C	
5G	CATGGTAATGGCCAAAGC	661bp	54,5°C	
3G	GTTGATCGGGTTGATGTTTTAT		54,3°C	
5Q	TAGATAATGAAGCGTCTTCATTACC	673bp	56,6°C	
3Q	ACTTCTACTAGCGTCAATTTAACTCAAC		58,8°C	
5H	TAAGGCTGATAAAAGTGGTAATTC	753bp	53,6°C	
3H	CTTTAGTATGTTCTAAGCGAAAGC		54,9°C	
5R	AACCTCTTTCAGAAAGGAGG	775bp	53,7°C	
3R	AACTTTAATTGGTTTGGAGATTATTCTTTAG		54,7°C	
5I	GGGTCAAACGTGAACTTTAAG	858bp	53,9°C	
3I	AACGGAAGGTAACTATGAAGCT		57°C	
5S	ACTTTTAACACCATCACTCGCTA	873bp	57,9°C	
3S	CAAACAACTAGAGGGTAAATACTTTATTGT		56,7°C	
5J	AGTTTGGCTCGTGCAAAAATAG	957bp	56,7°C	
3J	TTTTCGGTTTTATGAACCGTTC		53,3°C	
5T	CAACCTTTTGTTCGATACTAAAGAG	973bp	54,7°C	
3T	AATTTCTTTCTCATTTTTGGTTTAGTCC		54,3°C	
FtsH_RacePCR_GSP_R	tccagcggctccacttttagccata	852bp	70,3°C	RACE PCR FTSH Gene specific primer
811aR1	TTGTTCATTACTTGCACCGATTAC	685pb	52°C	Multiplex PCR for Mmc
799aF2	GTCAAGTTCTTTTCATACCACTAC		52°C	
900F-2	GAAGTATGATTTCCAGAACAAAAC	784pb	52°C	
901R2	AACTAGCTCCGTGTTGCTTTG		52°C	
600F	TAGCTGTTTGCTTGCTAAGGTC	846pb	52°C	
601R	TGGGTTTGTATTTAGTAGTAGTGC		52°C	
1000F2	TGTAGATCTGCCAAGTAAGTCTC	1010pb	52°C	
1001R2	CCTGTAATTTGTTTGATTGCTTG		52°C	
100F	GTAATTGAACCTAATTCTTTTTCTAATC	377pb	52°C	
101R	GGACTTGGTGGAATTAGACATC		52°C	
500F	AACCATCTGCACCAGATAGTTC	726pb	52°C	
501R	AGTGGTATATTTAGTTTAGCAAAACC		52°C	
200F	TCAGCTTATTTAGCTACAAATTCTG	429pb	52°C	
201R	GAAGAAGATACTTCATGAACAAATG		52°C	
300F	AGAAGATATTGCAGATGCAGAAG	514pb	52°C	
301R	AGTTGCATTGCTTGAACTAGTTG		52°C	
400F	AAACTAGACAAAATGAAGATGGAAG	589pb	52°C	
401R1	CTTCATCATCTTCATATCAAGGAC		52°C	
Mcap Set2 Af	GCA TAT CTA AAA TAG CTT TAT TTT GTT C		51.1	Multiplex PCR for Mcap
Mcap Set2 Ar	TTT AGA TAA TGT TGA TGA TGT TAT TAA TAT TAT TAA AAA TTC	125	54.7	
Mcap Set2 Bf	ATG ACA GTG TGT ATC ATA CAA ACC TGA TAT ATT C		58.3	
Mcap Set2 Br	CAT AAA TTC TCC TTT TAG TTT ATA TGT TCA AAT CC	225	56.2	
Mcap Set2 Cf	GCA ATT TGA GTT GAA GTT GAA ATT AGT TTT G		55.1	
Mcap Set2 Cr	AAA TTC CTT CAA ATA GTT GAT TAG TTA AAG	325	52	
Mcap Set2 Df	CAA GAT ATA TCT TCT TGA AAC ACT TCA AAT G		55.1	
Mcap Set2 Dr	TTT AGG GAA TTG TTC TAT TTG AAT TAC ACC	425	54.7	
Mcap Set2 Ef	ATA GAT TAA TAC AAT ATA TTG ATA CTG TAA AAT AAT AT		52.6	
Mcap Set2 Er	TAA ATC TAA ATA GTC TCA TTG TCT TAA AAC TTC	525	54.4	
Mcap Set2 Ff	GCA AGT GTT GCT ATT GGT AAT TAT TTA GC		55.8	
Mcap Set2 Fr	TTT GCT ATT TCA TAT CTT TTT AAT GAA TCT TTG	625	53.2	
Mcap Set2 Gf	CAT GAT AAA AAA GCT CAA ATT TAT GTT TAT G		52.4	
Mcap Set2 Gr	GGA TCT GAA GAA ATA TTA GTA TAA ACT ATA G	725	53.7	
Mcap Set2 Hf	CTT AGT GAC TAT GAA CAT GAA GTT TGT GC			
Mcap Set2 Hr	C AGG ATA ATT TTC AAC TTG CAA ACT TAA CCC	826		
Mcap Set2 If	TAA TAT GGC TGT AAA TCA AAT GAA TTT AAC ATA TC		55	
Mcap Set2 Ir	AAA CTA GAT GAA TTA AAT ATT TTA GAT AAA GCC	925	53.2	
Mcap Set2 Jf	GAT GAA ATT TTA ATT TAT GAA ACT TCT AAA CAT TGC		55.3	
Mcap Set2 Jr	CAT TAA CAT TTG TAT CAG TGA TAT ATT TGG C	1025	55.1	

Primer name	Sequence	Product size	Tm	Role
MgalliS6-verif-F	TTTGGGATCTTATTTTCTCACCTCAC	426bp	62,81°C	Mycoplasma gallisepticum S6 genome verif. Amplification of part of S6 CRISPR locus
MgalliS6-verif-R	AGTGAACCTGATTGATATTGTTTTTG		59,74°C	
pMGAL-spacer-PAM1-F	CCGGCTCGAGGACTCAAAAACGTCTTTGTTCGTCGTTTAAGAA	-	95°C	Annealing oligonucleotides to be ligated to pMGAL plasmid
pMGAL-spacer-PAM1-R	CCGGTTCTTAAACGACGAACAAAGACGTTTTTGAGTCCTCGAG			
pMGAL-spacer-PAM2-F	CCGGCTCGAGGACTCAAAAACGTCTTTGTTCGTCGTTTAAAAG			
pMGAL-spacer-PAM2-R	CCGGCTTTTAAACGACGAACAAAGACGTTTTTGAGTCCTCGAG			
pMGAL-spacer-PAM3-F	CCGGCTCGAGGACTCAAAAACGTCTTTGTTCGTCGTTTAAAAA			
pMGAL-spacer-PAM3-R	CCGGTTTTTAAACGACGAACAAAGACGTTTTTGAGTCCTCGAG			
pMGAL-spacer-PAM4-F	CCGGCTCGAGGACTCAAAAACGTCTTTGTTCGTCGTTTAAGGG			
pMGAL-spacer-PAM4-R	CCGGCCCTTAAACGACGAACAAAGACGTTTTTGAGTCCTCGAG			
pMGAL-spacer-PAM5-F	aaaCCGGGTACCGAGCTCGAA	7767bp	70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM5-R	taatTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM6-F	tccCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM6-R	ctaaTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM7-F	aaaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM7-R	ttatTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM8-F	gaaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM8-R	tcttTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM9-F	aaaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM9-R	tcaaTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM10-F	taaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM10-R	cgctTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM11-F	aaaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM11-R	tcccTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-PAM12-F	aaaCCGGGTACCGAGCTCGAA		70°C	QC to develop a novel pMGAL-spacer-PAM vector
pMGAL-spacer-PAM12-R	taacTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL-spacer-DR-F	acttgtgtaagcaataacCCGGGTACCGAGCTCGAA	7806	70°C	
pMGAL-spacer-DR-R	attgtacagtgctaaaacTTAAACGACGAACAAAGACGTTTTTGAG		67°C	
pMGAL_mod_verif_F	TTTGAAAACAACCCTAGCCACT	230bp	64°C	Verify the modification of PMGAL-spacer-PAM
pMGAL_mod_verif_R	GTTTTCCCAGTCACGACGTT		65°C	
ICE_internal_F	TACTGAACATTTTGTCTATAATTCG		53°C	
ICE_internal_R	TAACTCAGGCAAGGGGTCAT		58°C	
ICE_excursion_verif_F	AGTTCAGACCAATTATTAAGTGAAACT	571bp	57,3°C	
ICE_excursion_verif_R	TGGCAGGCAATTATTGGTAAA		60,3°C	
gRNA-ICE-Cas-Integration-AarI comp F	5' ATGATCCAAAAGTTAAACATATGTGGGT 3'	Hybridisation primers for AarI vector modification	74,8°C	Target ICE element MMCAP1_0557 for Cas9 Double Strand Break and insertion of MgCas9
gRNA-ICE-Cas-Integration-AarI comp R	3'AGGTTTTCAATTTGTATACACCCAAAAT 5' (5'TAAAACCCACATATGTTTAACTTTTGGA 3')		73°C	
Mmc_ICE_Cas_F	ATGTTTTTTGAAGCAACTTCATTTAACCAAGATATATCAAACTGAGA TACaaccaatcaggaacgttagg	3890bp	62°C	Amplification of MgCas9 for introduction inside the sequence of the ICE element MMCAP1_0557
Mmc_ICE_Cas_R	AGTATTTCATTTACCAATTGGTTGGTTAAAACTTTCAGCATTATAAA ACAtagcagcgccttattttagc		63°C	
MmcICE_Cas_ver_F	ACTGAGTGCTTAGAAATTGGTTATC		57.28°C	Verification of the introduction of the MgCas9 inside the sequence of the ICE element MMCAP1_0557 by amplifying the ends joints
MmcICE_Cas_ver_R	GCTCAAGCAAACATACCTTCC		58,88°C	
MmcICE_Cas_internal_ver_F	gcttgaccttgacaaaatagca		60°C	
MmcICE_Cas_internal_ver_R	gctccccaaatggtgaataa		60°C	

Bibliography

Alain Blanchard, Glenn Browning. *Mycoplasmas: MOLECULAR BIOLOGY PATHOGENICITY AND STRATEGIES FOR CONTROL*, 2005. https://www.crcpress.com/Mycoplasmas-MOLECULAR-BIOLOGY-PATHOGENICITY-AND-STRATEGIES-FOR-CONTROL/Blanchard-Browning/p/book/9780849398612.

Allam, Ayman B., Leticia Reyes, Nacyra Assad-Garcia, John I. Glass, and Mary B. Brown. "Enhancement of Targeted Homologous Recombination in Mycoplasma Mycoides Subsp. Capri by Inclusion of Heterologous RecA." *Applied and Environmental Microbiology* 76, no. 20 (October 2010): 6951–54. https://doi.org/10.1128/AEM.00056-10.

Almagor, M, I Kahane, C Gilon, and S Yatziv. "Protective Effects of the Glutathione Redox Cycle and Vitamin E on Cultured Fibroblasts Infected by Mycoplasma Pneumoniae." *Infection and Immunity* 52, no. 1 (April 1986): 240–44.

Anders, Carolin, Katja Bargsten, and Martin Jinek. "Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9." *Molecular Cell* 61, no. 6 (March 17, 2016): 895–902. https://doi.org/10.1016/j.molcel.2016.02.020.

Anders, Carolin, Ole Niewoehner, Alessia Duerst, and Martin Jinek. "Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease." *Nature* 513, no. 7519 (September 25, 2014): 569–73. https://doi.org/10.1038/nature13579.

Arfi, Yonathan, Laetitia Minder, Carmelo Di Primo, Aline Le Roy, Christine Ebel, Laurent Coquet, Stephane Claverol, et al. "MIB–MIP Is a Mycoplasma System That Captures and Cleaves Immunoglobulin G." *Proceedings of the National Academy of Sciences* 113, no. 19 (May 10, 2016): 5406–11. https://doi.org/10.1073/pnas.1600546113.

Bailey, T. L., and M. Gribskov. "Combining Evidence Using P-Values: Application to Sequence Homology Searches." *Bioinformatics* 14, no. 1 (January 1, 1998): 48–54. https://doi.org/10.1093/bioinformatics/14.1.48.

Bao, Zehua, Han Xiao, Jing Liang, Lu Zhang, Xiong Xiong, Ning Sun, Tong Si, and Huimin Zhao. "Homology-Integrated CRISPR–Cas (HI-CRISPR) System for One-Step Multigene Disruption in Saccharomyces Cerevisiae." *ACS Synthetic Biology* 4, no. 5 (May 15, 2015): 585–94. https://doi.org/10.1021/sb500255k.

Barlow, Miriam. "What Antimicrobial Resistance Has Taught Us About Horizontal Gene Transfer." In *Horizontal Gene Transfer*, 397–411. Methods in Molecular Biology. Humana Press, 2009. https://doi.org/10.1007/978-1-60327-853-9_23.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. "CRISPR Provides Acquired Resistance against Viruses in Prokaryotes." *Science (New York, N.Y.)* 315, no. 5819 (March 23, 2007): 1709–12. https://doi.org/10.1126/science.1138140.

Baseman, J. B., and J. G. Tully. "Mycoplasmas: Sophisticated, Reemerging, and Burdened by Their Notoriety." *Emerging Infectious Diseases* 3, no. 1 (March 1997): 21–32. https://doi.org/10.3201/eid0301.970103.

Ben-Menachem, Gil, Ulrich Zähringer, and Shlomo Rottem. "The Phosphocholine Motif in Membranes of Mycoplasma Fermentans Strains." *FEMS Microbiology Letters* 199, no. 1 (May 1, 2001): 137–41. https://doi.org/10.1111/j.1574-6968.2001.tb10664.x.

Berho, Nathalie, Sybille Duret, Jean-Luc Danet, and Joël Renaudin. “Plasmid PSci6 from Spiroplasma Citri GII-3 Confers Insect Transmissibility to the Non-Transmissible Strain S. Citri 44.” *Microbiology* 152, no. 9 (2006): 2703–16. https://doi.org/10.1099/mic.0.29085-0.

Béven, Laure, Claire Charenton, Alain Dautant, Guillaume Bouyssou, Fabien Labroussaa, Anna Sköllermo, Anja Persson, Alain Blanchard, and Pascal Sirand-Pugnet. “Specific Evolution of F1-Like ATPases in Mycoplasmas.” *PLOS ONE* 7, no. 6 (2012): e38793. https://doi.org/10.1371/journal.pone.0038793.

Béven L, Arricau-Bouvery N, Renaudin J, Saillard C (2015) Pathogenicity, virulence, and transmission of plant spiroplasmas. In: Graham J, editor. Virulence mechanisms of plant pathogenic bacteria (p.331-346). APS (American Phytopathological Society) Press.production.

Bikard, David, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, and Luciano A. Marraffini. “Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System.” *Nucleic Acids Research* 41, no. 15 (August 2013): 7429–37. https://doi.org/10.1093/nar/gkt520.

Blanchard, A. “Ureaplasma Urealyticum Urease Genes; Use of a UGA Tryptophan Codon.” *Molecular Microbiology* 4, no. 4 (April 1, 1990): 669–76. https://doi.org/10.1111/j.1365-2958.1990.tb00636.x.

Blanchard, Alain, and Glenn Browning. *Mycoplasmas: MOLECULAR BIOLOGY PATHOGENICITY AND STRATEGIES FOR CONTROL*. CRC Press, 2005.

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, et al. “The Complete Genome Sequence of Escherichia Coli K-12.” *Science (New York, N.Y.)* 277, no. 5331 (September 5, 1997): 1453–62.

Blötz, Cedric, and Jörg Stülke. “Glycerol Metabolism and Its Implication in Virulence in Mycoplasma.” *FEMS Microbiology Reviews* 41, no. 5 (September 1, 2017): 640–52. https://doi.org/10.1093/femsre/fux033.

Boeke, J. D., F. LaCroute, and G. R. Fink. “A Positive Selection for Mutants Lacking Orotidine-5’-Phosphate Decarboxylase Activity in Yeast: 5-Fluoro-Orotic Acid Resistance.” *Molecular & General Genetics: MGG* 197, no. 2 (1984): 345–46.

Bolotin, Alexander, Benoit Quinquis, Alexei Sorokin, and S. Dusko Ehrlich. “Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin.” *Microbiology (Reading, England)* 151, no. Pt 8 (August 2005): 2551–61. https://doi.org/10.1099/mic.0.28048-0.

Breton, Marc, Evelyne Sagné, Sybille Duret, Laure Béven, Christine Citti, and Joël Renaudin. “First Report of a Tetracycline-Inducible Gene Expression System for Mollicutes.” *Microbiology* 156, no. 1 (2010): 198–205. https://doi.org/10.1099/mic.0.034074-0.

Breton, Marc, Florence Tardy, Emilie Dordet-Frisoni, Eveline Sagne, Virginie Mick, Joël Renaudin, Pascal Sirand-Pugnet, Christine Citti, and Alain Blanchard. “Distribution and Diversity of Mycoplasma Plasmids: Lessons from Cryptic Genetic Elements.” *BMC Microbiology* 12 (November 12, 2012): 257. https://doi.org/10.1186/1471-2180-12-257.

Brocchi, Marcelo, Ana Tereza Ribeiro de Vasconcelos, and Arnaldo Zaha. "Restriction-Modification Systems in Mycoplasma Spp." *Genetics and Molecular Biology* 30, no. 1 (2007): 236–44. https://doi.org/10.1590/S1415-47572007000200011.

Brouns, Stan J. J., Matthijs M. Jore, Magnus Lundgren, Edze R. Westra, Rik J. H. Slijkhuis, Ambrosius P. L. Snijders, Mark J. Dickman, Kira S. Makarova, Eugene V. Koonin, and John van der Oost. "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science (New York, N.Y.)* 321, no. 5891 (August 15, 2008): 960–64. https://doi.org/10.1126/science.1159689.

Browning, Glenn, and Christine Citti. *Mollicutes: Molecular Biology and Pathogenesis*. Horizon Scientific Press, 2014.

Burgos, Raul, Gwendolyn E. Wood, Lei Young, John I. Glass, and Patricia A. Totten. "RecA Mediates MgpB and MgpC Phase and Antigenic Variation in Mycoplasma Genitalium, but Plays a Minor Role in DNA Repair." *Molecular Microbiology* 85, no. 4 (August 1, 2012): 669–83. https://doi.org/10.1111/j.1365-2958.2012.08130.x.

Calavan, E. C., and J. M. Bové. "9 - ECOLOGY OF Spiroplasma Citri." In *The Mycoplasmas*, edited by ROBERT F. Whitcomb and JOSEPH G. Tully, 425–85. Academic Press, 1989. https://doi.org/10.1016/B978-0-12-078405-9.50017-3.

Calcutt, Michael J., Jennifer L. Lavrrar, and Kim S. Wise. "IS1630 of Mycoplasma Fermentans, a Novel IS30-Type Insertion Element That Targets and Duplicates Inverted Repeats of Variable Length and Sequence during Insertion." *Journal of Bacteriology* 181, no. 24 (December 1999): 7597–7607.

Cao, J., P. A. Kapke, and F. C. Minion. "Transformation of Mycoplasma Gallisepticum with Tn916, Tn4001, and Integrative Plasmid Vectors." *Journal of Bacteriology* 176, no. 14 (July 1, 1994): 4459–62. https://doi.org/10.1128/jb.176.14.4459-4462.1994.

Carle, Patricia, Colette Saillard, Nathalie Carrère, Sébastien Carrère, Sybille Duret, Sandrine Eveillard, Patrice Gaurivaud, et al. "Partial Chromosome Sequence of Spiroplasma Citri Reveals Extensive Viral Invasion and Important Gene Decay." *Applied and Environmental Microbiology* 76, no. 11 (June 2010): 3420–26. https://doi.org/10.1128/AEM.02954-09.

Chandler, Michael, and Jacques Mahillon. "Insertion Sequences Revisited," January 1, 2002, 305–66. https://doi.org/10.1128/9781555817954.ch15.

Chandran, Suchismita, Vladimir N. Noskov, Thomas H. Segall-Shapiro, Li Ma, Caitlin Whiteis, Carole Lartigue, Joerg Jores, Sanjay Vashee, and Ray-Yuan Chuang. "TREC-IN: Gene Knock-in Genetic Tool for Genomes Cloned in Yeast." *BMC Genomics* 15 (December 24, 2014): 1180. https://doi.org/10.1186/1471-2164-15-1180.

Chang, Nannan, Changhong Sun, Lu Gao, Dan Zhu, Xiufei Xu, Xiaojun Zhu, Jing-Wei Xiong, and Jianzhong Jeff Xi. "Genome Editing with RNA-Guided Cas9 Nuclease in Zebrafish Embryos." *Cell Research* 23, no. 4 (April 2013): 465–72. https://doi.org/10.1038/cr.2013.45.

Chew, Wei Leong, Mohammadsharif Tabebordbar, Jason K. W. Cheng, Prashant Mali, Elizabeth Y. Wu, Alex H. M. Ng, Kexian Zhu, Amy J. Wagers, and George M. Church. "A Multifunctional AAV-CRISPR-Cas9 and Its Host Response." *Nature Methods* 13, no. 10 (October 2016): 868–74. https://doi.org/10.1038/nmeth.3993.

Cho, Seung Woo, Sojung Kim, Yongsub Kim, Jiyeon Kweon, Heon Seok Kim, Sangsu Bae, and Jin-Soo Kim. “Analysis of Off-Target Effects of CRISPR/Cas-Derived RNA-Guided Endonucleases and Nickases.” *Genome Research* 24, no. 1 (January 2014): 132–41. https://doi.org/10.1101/gr.162339.113.

Chopin, Marie-Christine, Alain Chopin, and Elena Bidnenko. “Phage Abortive Infection in Lactococci: Variations on a Theme.” *Current Opinion in Microbiology*, Host--microbe interactions: fungi / edited by Howard Bussey · Host--microbe interactions: parasites / edited by Artur Scherf · Host--microbe interactions: viruses / edited by Margaret CM Smith, 8, no. 4 (August 1, 2005): 473–79. https://doi.org/10.1016/j.mib.2005.06.006.

Chopra-Dewasthaly, Rohini, Marc Marenda, Renate Rosengarten, Wolfgang Jechlinger, and Christine Citti. “Construction of the First Shuttle Vectors for Gene Cloning and Homologous Recombination in Mycoplasma Agalactiae.” *FEMS Microbiology Letters* 253, no. 1 (December 1, 2005): 89–94. https://doi.org/10.1016/j.femsle.2005.09.021.

Chu, Van Trung, Timm Weber, Benedikt Wefers, Wolfgang Wurst, Sandrine Sander, Klaus Rajewsky, and Ralf Kühn. “Increasing the Efficiency of Homology-Directed Repair for CRISPR-Cas9-Induced Precise Gene Editing in Mammalian Cells.” *Nature Biotechnology* 33, no. 5 (May 2015): 543–48. https://doi.org/10.1038/nbt.3198.

Chylinski, Krzysztof, Kira S. Makarova, Emmanuelle Charpentier, and Eugene V. Koonin. “Classification and Evolution of Type II CRISPR-Cas Systems.” *Nucleic Acids Research* 42, no. 10 (June 2, 2014): 6091–6105. https://doi.org/10.1093/nar/gku241.

Chylinski, Krzysztof, Anaïs Le Rhun, and Emmanuelle Charpentier. “The TracrRNA and Cas9 Families of Type II CRISPR-Cas Immunity Systems.” *RNA Biology* 10, no. 5 (May 1, 2013): 726–37. https://doi.org/10.4161/rna.24321.

Clapper, Brenda, Anh-Hue T. Tu, Ada Elgavish, and Kevin Dybvig. “The Vir Gene of Bacteriophage MAV1 Confers Resistance to Phage Infection on Mycoplasma Arthritidis.” *Journal of Bacteriology* 186, no. 17 (September 2004): 5715–20. https://doi.org/10.1128/JB.186.17.5715-5720.2004.

Clewell, D. B., and C. Gawron-Burke. “Conjugative Transposons and the Dissemination of Antibiotic Resistance in Streptococci.” *Annual Review of Microbiology* 40 (1986): 635–59. https://doi.org/10.1146/annurev.mi.40.100186.003223.

Cobb, Ryan E., Yajie Wang, and Huimin Zhao. “High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System.” *ACS Synthetic Biology* 4, no. 6 (June 19, 2015): 723–28. https://doi.org/10.1021/sb500351f.

Connor, Steve. “First Human Embryos Edited in U.S., Using CRISPR.” *MIT Technology Review*. Accessed October 20, 2017. https://www.technologyreview.com/s/608350/first-human-embryos-edited-in-us/.

Cordova, Caio M. M., Carole Lartigue, Pascal Sirand-Pugnet, Joël Renaudin, Regina A. F. Cunha, and A. Blanchard. “Identification of the Origin of Replication of the Mycoplasma Pulmonis Chromosome and Its Use in OriC Replicative Plasmids.” *Journal of Bacteriology* 184, no. 19 (October 2002): 5426–35. https://doi.org/10.1128/JB.184.19.5426-5435.2002.

Crécy-Lagard, Valérie de, Christian Marck, Céline Brochier-Armanet, and Henri Grosjean. "Comparative RNomics and Modomics in Mollicutes: Prediction of Gene Function and Evolutionary Implications." *IUBMB Life* 59, no. 10 (January 1, 2007): 634–58. https://doi.org/10.1080/15216540701604632.

Daniels, L. K. "Rapid In-Office and in-Vivo Desensitization of an Injection Phobia Utilizing Hypnosis." *The American Journal of Clinical Hypnosis* 18, no. 3 (January 1976): 200–203. https://doi.org/10.1080/00029157.1976.10403798.

Datsenko, K. A., and B. L. Wanner. "One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products." *Proceedings of the National Academy of Sciences of the United States of America* 97, no. 12 (June 6, 2000): 6640–45. https://doi.org/10.1073/pnas.120163297.

DeFeudis, F. V. "Coronary Atherosclerosis: Current Therapeutic Approaches and Future Trends." *Life Sciences* 49, no. 10 (1991): 689–705.

Delaney, Nigel F., Susan Balenger, Camille Bonneaud, Christopher J. Marx, Geoffrey E. Hill, Naola Ferguson-Noel, Peter Tsai, Allen Rodrigo, and Scott V. Edwards. "Ultrafast Evolution and Loss of CRISPRs Following a Host Shift in a Novel Wildlife Pathogen, Mycoplasma Gallisepticum." *PLOS Genetics* 8, no. 2 (2012): e1002511. https://doi.org/10.1371/journal.pgen.1002511.

Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jörg Vogel, and Emmanuelle Charpentier. "CRISPR RNA Maturation by Trans-Encoded Small RNA and Host Factor RNase III." *Nature* 471, no. 7340 (March 31, 2011): 602–7. https://doi.org/10.1038/nature09886.

Dhandayuthapani, S., M. W. Blaylock, C. M. Bebear, W. G. Rasmussen, and J. B. Baseman. "Peptide Methionine Sulfoxide Reductase (MsrA) Is a Virulence Determinant in Mycoplasma Genitalium." *Journal of Bacteriology* 183, no. 19 (October 1, 2001): 5645–50. https://doi.org/10.1128/JB.183.19.5645-5650.2001.

Dhandayuthapani, S., W. G. Rasmussen, and J. B. Baseman. "Disruption of Gene Mg218 of Mycoplasma Genitalium through Homologous Recombination Leads to an Adherence-Deficient Phenotype." *Proceedings of the National Academy of Sciences* 96, no. 9 (April 27, 1999): 5227–32. https://doi.org/10.1073/pnas.96.9.5227.

DiCarlo, James E., Julie E. Norville, Prashant Mali, Xavier Rios, John Aach, and George M. Church. "Genome Engineering in Saccharomyces Cerevisiae Using CRISPR-Cas Systems." *Nucleic Acids Research* 41, no. 7 (April 2013): 4336–43. https://doi.org/10.1093/nar/gkt135.

Dordet Frisoni, Emilie, Marc Serge Marenda, Eveline Sagné, Laurent Xavier Nouvel, Romain Guérillot, Philippe Glaser, Alain Blanchard, et al. "ICEA of Mycoplasma Agalactiae: A New Family of Self-Transmissible Integrative Elements That Confers Conjugative Properties to the Recipient Strain." *Molecular Microbiology* 89, no. 6 (September 1, 2013): 1226–39. https://doi.org/10.1111/mmi.12341.

Dunning Hotopp, Julie C. "Horizontal Gene Transfer between Bacteria and Animals." *Trends in Genetics : TIG* 27, no. 4 (April 2011): 157–63. https://doi.org/10.1016/j.tig.2011.01.005.

Duret, Sybille, Jean-Luc Danet, Monique Garnier, and Joël Renaudin. “Gene Disruption through Homologous Recombination InSpiroplasma Citri: An Scm1-Disrupted Motility Mutant Is Pathogenic.” *Journal of Bacteriology* 181, no. 24 (December 15, 1999): 7449–56.

Dybvig, K, J Alderete, H L Watson, and G H Cassell. “Adsorption of Mycoplasma Virus P1 to Host Cells.” *Journal of Bacteriology* 170, no. 9 (September 1988): 4373–75.

Dybvig, K., and A. Woodard. “Cloning and DNA Sequence of a Mycoplasmal RecA Gene.” *Journal of Bacteriology* 174, no. 3 (February 1992): 778–84.

Firrao, Giuseppe, Meritxell Garcia-Chapa, and Cristina Marzachì. “Phytoplasmas: Genetics, Diagnosis and Relationships with the Plant and Insect Host.” *Frontiers in Bioscience: A Journal and Virtual Library* 12 (January 1, 2007): 1353–75.

Foissac, X., C. Saillard, and J. M. Bové. “Random Insertion of Transposon Tn4001 in the Genome of Spiroplasma Citri Strain GII3.” *Plasmid* 37, no. 1 (1997): 80–86. https://doi.org/10.1006/plas.1996.1271.

Fonfara, Ines, Anaïs Le Rhun, Krzysztof Chylinski, Kira S. Makarova, Anne-Laure Lécrivain, Janek Bzdrenga, Eugene V. Koonin, and Emmanuelle Charpentier. “Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR-Cas Systems.” *Nucleic Acids Research* 42, no. 4 (February 2014): 2577–90. https://doi.org/10.1093/nar/gkt1074.

French, Christopher T., Ping Lao, Ann E. Loraine, Brian T. Matthews, Huilan Yu, and Kevin Dybvig. “Large-Scale Transposon Mutagenesis of Mycoplasma Pulmonis.” *Molecular Microbiology* 69, no. 1 (July 2008): 67–76. https://doi.org/10.1111/j.1365-2958.2008.06262.x.

Frey, J., X. Cheng, P. Kuhnert, and J. Nicolet. “Identification and Characterization of IS1296 in Mycoplasma Mycoides Subsp. Mycoides SC and Presence in Related Mycoplasmas.” *Gene* 160, no. 1 (July 4, 1995): 95–100.

Fu, Yanfang, Jennifer A. Foden, Cyd Khayter, Morgan L. Maeder, Deepak Reyon, J. Keith Joung, and Jeffry D. Sander. “High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells.” *Nature Biotechnology* 31, no. 9 (September 2013): 822–26. https://doi.org/10.1038/nbt.2623.

Fu, Yanfang, Jeffry D. Sander, Deepak Reyon, Vincent M. Cascio, and J. Keith Joung. “Improving CRISPR-Cas Nuclease Specificity Using Truncated Guide RNAs.” *Nature Biotechnology* 32, no. 3 (March 2014): 279–84. https://doi.org/10.1038/nbt.2808.

Gadeau, A. P., C. Mouches, and J. M. Bove. “Probable Insensitivity of Mollicutes to Rifampin and Characterization of Spiroplasmal DNA-Dependent RNA Polymerase.” *Journal of Bacteriology* 166, no. 3 (June 1, 1986): 824–28. https://doi.org/10.1128/jb.166.3.824-828.1986.

Gao, Wei, Lu Long, Xinquan Tian, Fuchun Xu, Ji Liu, Prashant K. Singh, Jose R. Botella, and Chunpeng Song. “Genome Editing in Cotton with the CRISPR/Cas9 System.” *Frontiers in Plant Science* 8 (August 3, 2017). https://doi.org/10.3389/fpls.2017.01364.

Gasiunas, Giedrius, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. “Cas9–crRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria.” *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 39 (September 25, 2012): E2579–86. https://doi.org/10.1073/pnas.1208507109.

Gaurivaud, Patrice, Jean-Luc Danet, Frédéric Laigret, Monique Garnier, and Joseph M. Bové. “Fructose Utilization and Phytopathogenicity of Spiroplasma Citri.” *Molecular Plant-Microbe Interactions* 13, no. 10 (October 1, 2000): 1145–55. https://doi.org/10.1094/MPMI.2000.13.10.1145.

Gaurivaud, Patrice, Latifa Lakhdar, Dominique Le Grand, François Poumarat, and Florence Tardy. “Comparison of in Vivo and in Vitro Properties of Capsulated and Noncapsulated Variants of Mycoplasma Mycoides Subsp. Mycoides Strain Afadé: A Potential New Insight into the Biology of Contagious Bovine Pleuropneumonia.” *FEMS Microbiology Letters* 359, no. 1 (October 1, 2014): 42–49. https://doi.org/10.1111/1574-6968.12579.

Gibson, Daniel G., Gwynedd A. Benders, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Holly Baden-Tillson, Jayshree Zaveri, Timothy B. Stockwell, et al. “Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma Genitalium Genome.” *Science (New York, N.Y.)* 319, no. 5867 (February 29, 2008): 1215–20. https://doi.org/10.1126/science.1151721.

Gibson, Daniel G., John I. Glass, Carole Lartigue, Vladimir N. Noskov, Ray-Yuan Chuang, Mikkel A. Algire, Gwynedd A. Benders, et al. *CREATION OF A BACTERIAL CELL CONTROLLED BY A CHEMICALLY SYNTHESIZED GENOME*. National Academies Press (US), 2011. https://www.ncbi.nlm.nih.gov/books/NBK84435/.

Glass, John I., Nacyra Assad-Garcia, Nina Alperovich, Shibu Yooseph, Matthew R. Lewis, Mahir Maruf, Clyde A. Hutchison, Hamilton O. Smith, and J. Craig Venter. “Essential Genes of a Minimal Bacterium.” *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 2 (January 10, 2006): 425–30. https://doi.org/10.1073/pnas.0510013103.

Goldfarb, Tamara, Hila Sberro, Eyal Weinstock, Ofir Cohen, Shany Doron, Yoav Charpak-Amikam, Shaked Afik, Gal Ofir, and Rotem Sorek. “BREX Is a Novel Phage Resistance System Widespread in Microbial Genomes.” *The EMBO Journal* 34, no. 2 (January 13, 2015): 169–83. https://doi.org/10.15252/embj.201489455.

Grissa, Ibtissem, Gilles Vergnaud, and Christine Pourcel. “The CRISPRdb Database and Tools to Display CRISPRs and to Generate Dictionaries of Spacers and Repeats.” *BMC Bioinformatics* 8 (May 23, 2007): 172. https://doi.org/10.1186/1471-2105-8-172.

Groenen, Peter M. A., Annelies E. Bunschoten, Dick van Soolingen, and Jan D. A. van Errtbden. “Nature of DNA Polymorphism in the Direct Repeat Cluster of Mycobacterium Tuberculosis; Application for Strain Differentiation by a Novel Typing Method.” *Molecular Microbiology* 10, no. 5 (December 1, 1993): 1057–65. https://doi.org/10.1111/j.1365-2958.1993.tb00976.x.

Grosjean, Henri, Marc Breton, Pascal Sirand-Pugnet, Florence Tardy, François Thiaucourt, Christine Citti, Aurélien Barré, et al. “Predicting the Minimal Translation Apparatus: Lessons from the Reductive Evolution of Mollicutes.” *PLOS Genetics* 10, no. 5 (2014): e1004363. https://doi.org/10.1371/journal.pgen.1004363.

Güell, Marc, Vera van Noort, Eva Yus, Wei-Hua Chen, Justine Leigh-Bell, Konstantinos Michalodimitrakis, Takuji Yamada, et al. “Transcriptome Complexity in a Genome-Reduced Bacterium.” *Science (New York, N.Y.)* 326, no. 5957 (November 27, 2009): 1268–71. https://doi.org/10.1126/science.1176951.

Guglielmini, Julien, Leonor Quintais, Maria Pilar Garcillán-Barcia, Fernando de la Cruz, and Eduardo P. C. Rocha. “The Repertoire of ICE in Prokaryotes Underscores the Unity, Diversity, and Ubiquity of Conjugation.” *PLoS Genetics* 7, no. 8 (August 2011): e1002222. https://doi.org/10.1371/journal.pgen.1002222.

Halbedel, Sven, and Jörg Stülke. “Probing In Vivo Promoter Activities in Mycoplasma Pneumoniae: A System for Generation of Single-Copy Reporter Constructs.” *Applied and Environmental Microbiology* 72, no. 2 (February 1, 2006): 1696–99. https://doi.org/10.1128/AEM.72.2.1696-1699.2006.

Hames, Claudine, Sven Halbedel, Michael Hoppert, Joachim Frey, and Jörg Stülke. “Glycerol Metabolism Is Important for Cytotoxicity of Mycoplasma Pneumoniae.” *Journal of Bacteriology* 191, no. 3 (February 2009): 747–53. https://doi.org/10.1128/JB.01103-08.

Hao, H., X. Wang, H. Jia, M. Yu, X. Zhang, H. Tang, and L. Zhang. “Large Fragment Deletion Using a CRISPR/Cas9 System in Saccharomyces Cerevisiae.” *Analytical Biochemistry* 509 (September 2016): 118–23. https://doi.org/10.1016/j.ab.2016.07.008.

Hedreyda, C. T., and D. C. Krause. “Identification of a Possible Cytadherence Regulatory Locus in Mycoplasma Pneumoniae.” *Infection and Immunity* 63, no. 9 (September 1995): 3479–83.

Hicks, C. A. E., E. N. Barker, C. Brady, C. R. Stokes, C. R. Helps, and S. Tasker. “Non-Ribosomal Phylogenetic Exploration of Mollicute Species: New Insights into Haemoplasma Taxonomy.” *Infection, Genetics and Evolution* 23, no. Supplement C (April 1, 2014): 99–105. https://doi.org/10.1016/j.meegid.2014.02.001.

Highfill, Chad A., Jonathan H. Tran, Samantha K. T. Nguyen, Taylor R. Moldenhauer, Xiaofei Wang, and Stuart J. Macdonald. “Naturally Segregating Variation at Ugt86Dd Contributes to Nicotine Resistance in Drosophila Melanogaster.” *Genetics* 207, no. 1 (September 2017): 311–25. https://doi.org/10.1534/genetics.117.300058.

Hoeijmakers, Jan H. J. “Genome Maintenance Mechanisms for Preventing Cancer.” *Nature* 411, no. 6835 (May 17, 2001): 366–74. https://doi.org/10.1038/35077232.

Hogenhout, Saskia A., Kenro Oshima, El-Desouky Ammar, Shigeyuki Kakizawa, Heather N. Kingdom, and Shigetou Namba. “Phytoplasmas: Bacteria That Manipulate Plants and Insects.” *Molecular Plant Pathology* 9, no. 4 (July 2008): 403–23. https://doi.org/10.1111/j.1364-3703.2008.00472.x.

Horvath, Philippe, and Rodolphe Barrangou. “CRISPR/Cas, the Immune System of Bacteria and Archaea.” *Science (New York, N.Y.)* 327, no. 5962 (January 8, 2010): 167–70. https://doi.org/10.1126/science.1179555.

Horvath, Philippe, Dennis A. Romero, Anne-Claire Coûté-Monvoisin, Melissa Richards, Hélène Deveau, Sylvain Moineau, Patrick Boyaval, Christophe Fremaux, and Rodolphe Barrangou. “Diversity, Activity, and Evolution of CRISPR Loci in Streptococcus Thermophilus.” *Journal of Bacteriology* 190, no. 4 (February 2008): 1401–12. https://doi.org/10.1128/JB.01415-07.

Hsu, Patrick D., David A. Scott, Joshua A. Weinstein, F. Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, et al. “DNA Targeting Specificity of RNA-Guided Cas9 Nucleases.” *Nature Biotechnology* 31, no. 9 (September 1, 2013): 827–32. https://doi.org/10.1038/nbt.2647.

Marguerite, E., Hunt, Daniel, R.Brown. Role of sialidase in Mycoplasma alligatoris-induced pulmonary fibroblast apoptosis. Veterinary Microbiology Volume 121, Issues 1–2, 31 March 2007, Pages 73-82. https://doi.org/10.1016/j.vetmic.2006.10.009

Hutchison, C. A., S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter. “Global Transposon Mutagenesis and a Minimal Mycoplasma Genome.” *Science (New York, N.Y.)* 286, no. 5447 (December 10, 1999): 2165–69.

Hutchison, Clyde A., Ray-Yuan Chuang, Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, et al. “Design and Synthesis of a Minimal Bacterial Genome.” *Science* 351, no. 6280 (March 25, 2016): aad6253. https://doi.org/10.1126/science.aad6253.

Hutvagner, Gyorgy, and Martin J. Simard. “Argonaute Proteins: Key Players in RNA Silencing.” *Nature Reviews Molecular Cell Biology* 9, no. 1 (January 2008): 22–32. https://doi.org/10.1038/nrm2321.

Ishag, Hassan Z. A., Qiyan Xiong, Maojun Liu, Zhixin Feng, and Guoqing Shao. “Development of OriC -Plasmids for Use in Mycoplasma Hyorhinis.” *Scientific Reports* 7, no. 1 (September 6, 2017): 10596. https://doi.org/10.1038/s41598-017-10519-3.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. “Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product.” *Journal of Bacteriology* 169, no. 12 (December 1987): 5429–33.

J, Fletcher, Schultz G.a, Davis R.e, Eastman C.e, and Goodman R.m. “Spiroplasma Citri Is an Etiological Agent in Brittleroot Disease of Horseradish [Abstract Only].” *Phytopathology*, 1981. http://agris.fao.org/agris-search/search.do?recordID=US19820801432.

Jakočiūnas, Tadas, Ida Bonde, Markus Herrgård, Scott J. Harrison, Mette Kristensen, Lasse E. Pedersen, Michael K. Jensen, and Jay D. Keasling. “Multiplex Metabolic Pathway Engineering Using CRISPR/Cas9 in Saccharomyces Cerevisiae.” *Metabolic Engineering* 28 (March 2015): 213–22. https://doi.org/10.1016/j.ymben.2015.01.008.

Janis, Carole, Daniela Bischof, Géraldine Gourgues, Joachim Frey, Alain Blanchard, and Pascal Sirand-Pugnet. “Unmarked Insertional Mutagenesis in the Bovine Pathogen Mycoplasma Mycoides Subsp. Mycoides SC: Characterization of a LppQ Mutant.” *Microbiology (Reading, England)* 154, no. Pt 8 (August 2008): 2427–36. https://doi.org/10.1099/mic.0.2008/017640-0.

Jensen, J. S., J. Blom, and K. Lind. “Intracellular Location of Mycoplasma Genitalium in Cultured Vero Cells as Demonstrated by Electron Microscopy.” *International Journal of Experimental Pathology* 75, no. 2 (April 1994): 91–98.

Jiang, Wenyan, David Bikard, David Cox, Feng Zhang, and Luciano A. Marraffini. “RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems.” *Nature Biotechnology* 31, no. 3 (March 2013): 233–39. https://doi.org/10.1038/nbt.2508.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. “A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.” *Science* 337, no. 6096 (August 17, 2012): 816–21. https://doi.org/10.1126/science.1225829.

Johnson, Christopher M., and Alan D. Grossman. “Integrative and Conjugative Elements (ICEs): What They Do and How They Work.” *Annual Review of Genetics* 49, no. 1 (2015): 577–601. https://doi.org/10.1146/annurev-genet-112414-055018.

Kannan, T. R., and Joel B. Baseman. “ADP-Ribosylating and Vacuolating Cytotoxin of Mycoplasma Pneumoniae Represents Unique Virulence Determinant among Bacterial Pathogens.” *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 17 (April 25, 2006): 6724–29. https://doi.org/10.1073/pnas.0510644103.

Kim, Eunji, Taeyoung Koo, Sung Wook Park, Daesik Kim, Kyoungmi Kim, Hee-Yeon Cho, Dong Woo Song, et al. “In Vivo Genome Editing with a Small Cas9 Orthologue Derived from Campylobacter Jejuni.” *Nature Communications* 8 (February 21, 2017): 14500. https://doi.org/10.1038/ncomms14500.

Kim, Sojung, Daesik Kim, Seung Woo Cho, Jungeun Kim, and Jin-Soo Kim. “Highly Efficient RNA-Guided Genome Editing in Human Cells via Delivery of Purified Cas9 Ribonucleoproteins.” *Genome Research* 24, no. 6 (June 2014): 1012–19. https://doi.org/10.1101/gr.171322.113.

King, Kendall W., and Kevin Dybvig. “Plasmid Transformation of Mycoplasma Mycoides Subspecies Mycoides Is Promoted by High Concentrations of Polyethylene Glycol.” *Plasmid* 26, no. 2 (September 1, 1991): 108–15. https://doi.org/10.1016/0147-619X(91)90050-7.

Kleinstiver, Benjamin P., Vikram Pattanayak, Michelle S. Prew, Shengdar Q. Tsai, Nhu T. Nguyen, Zongli Zheng, and J. Keith Joung. “High-Fidelity CRISPR-Cas9 Nucleases with No Detectable Genome-Wide off-Target Effects.” *Nature* 529, no. 7587 (January 28, 2016): 490–95. https://doi.org/10.1038/nature16526.

Koonin, Eugene V. “Evolution of Genome Architecture.” *The International Journal of Biochemistry & Cell Biology* 41, no. 2 (February 2009): 298–306. https://doi.org/10.1016/j.biocel.2008.09.015.

Koonin, Eugene V. “Horizontal Gene Transfer: Essentiality and Evolvability in Prokaryotes, and Roles in Evolutionary Transitions.” *F1000Research* 5 (2016). https://doi.org/10.12688/f1000research.8737.1.

Kouprina, Natalay, and Vladimir Larionov. “Exploiting the Yeast Saccharomyces Cerevisiae for the Study of the Organization and Evolution of Complex Genomes.” *FEMS Microbiology Reviews* 27, no. 5 (December 1, 2003): 629–49. https://doi.org/10.1016/S0168-6445(03)00070-6.

Krause, Duncan C., and Mitchell F. Balish. “Structure, Function, and Assembly of the Terminal Organelle of Mycoplasma Pneumoniae.” *FEMS Microbiology Letters* 198, no. 1 (April 1, 2001): 1–7. https://doi.org/10.1111/j.1574-6968.2001.tb10610.x.

Ku, Chuan, Wen-Sui Lo, Ling-Ling Chen, and Chih-Horng Kuo. “Complete Genomes of Two Dipteran-Associated Spiroplasmas Provided Insights into the Origin, Dynamics, and Impacts of Viral Invasion in Spiroplasma.” *Genome Biology and Evolution* 5, no. 6 (2013): 1151–64. https://doi.org/10.1093/gbe/evt084.

Kühner, Sebastian, Vera van Noort, Matthew J. Betts, Alejandra Leo-Macias, Claire Batisse, Michaela Rode, Takuji Yamada, et al. “Proteome Organization in a Genome-Reduced Bacterium.” *Science (New York, N.Y.)* 326, no. 5957 (November 27, 2009): 1235–40. https://doi.org/10.1126/science.1176343.

Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, et al. “The Complete Genome Sequence of the Gram-Positive Bacterium Bacillus Subtilis.” *Nature* 390, no. 6657 (November 20, 1997): 249–56. https://doi.org/10.1038/36786.

Labroussaa, Fabien, Anne Lebaudy, Vincent Baby, Géraldine Gourgues, Dominick Matteau, Sanjay Vashee, Pascal Sirand-Pugnet, Sébastien Rodrigue, and Carole Lartigue. “Impact of Donor–recipient Phylogenetic Distance on Bacterial Genome Transplantation.” *Nucleic Acids Research* 44, no. 17 (September 30, 2016): 8501–11. https://doi.org/10.1093/nar/gkw688.

Lange, Sita J., Omer S. Alkhnbashi, Dominic Rose, Sebastian Will, and Rolf Backofen. “CRISPRmap: An Automated Classification of Repeat Conservation in Prokaryotic Adaptive Immune Systems.” *Nucleic Acids Research* 41, no. 17 (September 2013): 8034–44. https://doi.org/10.1093/nar/gkt606.

Lartigue, Carole, Alain Blanchard, Joël Renaudin, François Thiaucourt, and Pascal Sirand-Pugnet. “Host Specificity of Mollicutes OriC Plasmids: Functional Analysis of Replication Origin.” *Nucleic Acids Research* 31, no. 22 (November 15, 2003): 6610–18.

Lartigue, Carole, John I. Glass, Nina Alperovich, Rembert Pieper, Prashanth P. Parmar, Clyde A. Hutchison, Hamilton O. Smith, and J. Craig Venter. “Genome Transplantation in Bacteria: Changing One Species to Another.” *Science (New York, N.Y.)* 317, no. 5838 (August 3, 2007): 632–38. https://doi.org/10.1126/science.1144622.

Lartigue, Carole, Anne Lebaudy, Alain Blanchard, Basma El Yacoubi, Simon Rose, Henri Grosjean, and Stephen Douthwaite. “The Flavoprotein Mcap0476 (RlmFO) Catalyzes M5U1939 Modification in Mycoplasma Capricolum 23S RRNA.” *Nucleic Acids Research* 42, no. 12 (July 2014): 8073–82. https://doi.org/10.1093/nar/gku518.

Lartigue, Carole, Sanjay Vashee, Mikkel A. Algire, Ray-Yuan Chuang, Gwynedd A. Benders, Li Ma, Vladimir N. Noskov, et al. “Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast.” *Science (New York, N.Y.)* 325, no. 5948 (September 25, 2009): 1693–96. https://doi.org/10.1126/science.1173759.

Leclercq, Sébastien, Jessica Dittmer, Didier Bouchon, and Richard Cordaux. “Phylogenomics of ‘Candidatus Hepatoplasma Crinochetorum,’ a Lineage of Mollicutes Associated with Noninsect Arthropods.” *Genome Biology and Evolution* 6, no. 2 (January 29, 2014): 407–15. https://doi.org/10.1093/gbe/evu020.

Lee, Nicholas C. O., Vladimir Larionov, and Natalay Kouprina. “Highly Efficient CRISPR/Cas9-Mediated TAR Cloning of Genes and Chromosomal Loci from Complex Genomes in Yeast.” *Nucleic Acids Research* 43, no. 8 (April 30, 2015): e55. https://doi.org/10.1093/nar/gkv112.

Lee, S.-W., G. F. Browning, and P. F. Markham. “Development of a Replicable OriC Plasmid for Mycoplasma Gallisepticum and Mycoplasma Imitans, and Gene Disruption through Homologous Recombination in M. Gallisepticum.” *Microbiology* 154, no. 9 (2008): 2571–80. https://doi.org/10.1099/mic.0.2008/019208-0.

Leenay, Ryan T., Kenneth R. Maksimchuk, Rebecca A. Slotkowski, Roma N. Agrawal, Ahmed A. Gomaa, Alexandra E. Briner, Rodolphe Barrangou, and Chase L. Beisel. “Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems.” *Molecular Cell* 62, no. 1 (April 7, 2016): 137–47. https://doi.org/10.1016/j.molcel.2016.02.031.

Li, J., J. Zhang, N. Zhang, Y. Zhang, W. Wu, and J. Li. “Development of a Replicative Plasmid for Gene Expression in Mycoplasma Bovis.” *Journal of Microbiological Methods* 108 (January 2015): 12–18. https://doi.org/10.1016/j.mimet.2014.11.005.

Li, Yingjun, Saifu Pan, Yan Zhang, Min Ren, Mingxia Feng, Nan Peng, Lanming Chen, Yun Xiang Liang, and Qunxin She. “Harnessing Type I and Type III CRISPR-Cas Systems for Genome Editing.” *Nucleic Acids Research* 44, no. 4 (February 29, 2016): e34. https://doi.org/10.1093/nar/gkv1044.

Li, Yuan, Huajun Zheng, Yang Liu, Yanwei Jiang, Jiuqing Xin, Wei Chen, and Zhiqiang Song. “The Complete Genome Sequence of Mycoplasma Bovis Strain Hubei-1.” *PloS One* 6, no. 6 (2011): e20999. https://doi.org/10.1371/journal.pone.0020999.

Liang, Puping, Yanwen Xu, Xiya Zhang, Chenhui Ding, Rui Huang, Zhen Zhang, Jie Lv, et al. “CRISPR/Cas9-Mediated Gene Editing in Human Tripronuclear Zygotes.” *Protein & Cell* 6, no. 5 (May 2015): 363–72. https://doi.org/10.1007/s13238-015-0153-5.

Lillestøl, Reidun K., Shiraz A. Shah, Kim Brügger, Peter Redder, Hien Phan, Jan Christiansen, and Roger A. Garrett. “CRISPR Families of the Crenarchaeal Genus Sulfolobus: Bidirectional Transcription and Dynamic Properties.” *Molecular Microbiology* 72, no. 1 (April 2009): 259–72. https://doi.org/10.1111/j.1365-2958.2009.06641.x.

Lluch-Senar, Maria, Javier Delgado, Wei-Hua Chen, Verónica Lloréns-Rico, Francis J. O’Reilly, Judith Ah Wodke, E. Besray Unal, et al. “Defining a Minimal Cell: Essentiality of Small ORFs and NcRNAs in a Genome-Reduced Bacterium.” *Molecular Systems Biology* 11, no. 1 (January 21, 2015): 780.

Lo, S. C., M. M. Hayes, J. G. Tully, R. Y. Wang, H. Kotani, P. F. Pierce, D. L. Rose, and J. W. Shih. “Mycoplasma Penetrans Sp. Nov., from the Urogenital Tract of Patients with AIDS.” *International Journal of Systematic Bacteriology* 42, no. 3 (July 1992): 357–64. https://doi.org/10.1099/00207713-42-3-357.

Lo, Wen-Sui, Gail E. Gasparich, and Chih-Horng Kuo. “Found and Lost: The Fates of Horizontally Acquired Genes in Arthropod-Symbiotic Spiroplasma.” *Genome Biology and Evolution* 7, no. 9 (August 8, 2015): 2458–72. https://doi.org/10.1093/gbe/evv160.

Lydia Krause, Tim Jewell. “Mycoplasma Pneumonia: Causes, Symptoms, and Diagnosis.” Accessed September 20, 2017. http://www.healthline.com/health/mycoplasma-pneumonia#overview1.

Lyon, Bruce R., John W. May, and Ronald A. Skurray. “Tn4001: A Gentamicin and Kanamycin Resistance Transposon in Staphylococcus Aureus.” *Molecular and General Genetics MGG* 193, no. 3 (March 1, 1984): 554–56. https://doi.org/10.1007/BF00382099.

Maglennon, Gareth A, Beth S Cook, Dominic Matthews, Alannah S Deeney, Janine T Bossé, Paul R Langford, Duncan J Maskell, Alexander W Tucker, Brendan W Wren, and Andrew N Rycroft. “Development of a Self-Replicating Plasmid System for Mycoplasma Hyopneumoniae.” *Veterinary Research* 44, no. 1 (2013): 63. https://doi.org/10.1186/1297-9716-44-63.

Maier, Tobias, Josep Marcos, Judith A. H. Wodke, Bernhard Paetzold, Manuel Liebeke, Ricardo Gutiérrez-Gallego, and Luis Serrano. “Large-Scale Metabolome Analysis and Quantitative Integration with Genomics and Proteomics Data in Mycoplasma Pneumoniae.” *Molecular BioSystems* 9, no. 7 (July 2013): 1743–55. https://doi.org/10.1039/c3mb70113a.

Makarova, Kira S., and Eugene V. Koonin. “Annotation and Classification of CRISPR-Cas Systems.” In *CRISPR*, 47–75. Methods in Molecular Biology. Humana Press, New York, NY, 2015. https://doi.org/10.1007/978-1-4939-2687-9_4.

Makarova, Kira S., Yuri I. Wolf, and Eugene V. Koonin. “The Basic Building Blocks and Evolution of CRISPR-CAS Systems.” *Biochemical Society Transactions* 41, no. 6 (December 2013): 1392–1400. https://doi.org/10.1042/BST20130038.

Manso-Silván, L., E. M. Vilei, K. Sachse, S. P. Djordjevic, F. Thiaucourt, and J. Frey. “Mycoplasma Leachii Sp. Nov. as a New Species Designation for Mycoplasma Sp. Bovine Group 7 of Leach, and Reclassification of Mycoplasma Mycoides Subsp. Mycoides LC as a Serovar of Mycoplasma Mycoides Subsp. Capri.” *International Journal of Systematic and Evolutionary Microbiology* 59, no. Pt 6 (June 2009): 1353–58. https://doi.org/10.1099/ijs.0.005546-0.

Mao, Yanfei, Zhengjing Zhang, Zhengyan Feng, Pengliang Wei, Hui Zhang, José Ramón Botella, and Jian-Kang Zhu. “Development of Germ-Line-Specific CRISPR-Cas9 Systems to Improve the Production of Heritable Gene Modifications in Arabidopsis.” *Plant Biotechnology Journal* 14, no. 2 (February 1, 2016): 519–32. https://doi.org/10.1111/pbi.12468.

Marenda, Marc, Valérie Barbe, Géraldine Gourgues, Sophie Mangenot, Evelyne Sagne, and Christine Citti. “A New Integrative Conjugative Element Occurs in Mycoplasma Agalactiae as Chromosomal and Free Circular Forms.” *Journal of Bacteriology* 188, no. 11 (June 2006): 4137–41. https://doi.org/10.1128/JB.00114-06.

Marenda, Marc S., Evelyne Sagné, François Poumarat, and Christine Citti. “Suppression Subtractive Hybridization as a Basis to Assess Mycoplasma Agalactiae and Mycoplasma Bovis Genomic Diversity and Species-Specific Sequences.” *Microbiology* 151, no. 2 (2005): 475–89. https://doi.org/10.1099/mic.0.27590-0.

Mariscal, Ana M., Luis González-González, Enrique Querol, and Jaume Piñol. “All-in-One Construct for Genome Engineering Using Cre-Lox Technology.” *DNA Research* 23, no. 3 (June 1, 2016): 263–70. https://doi.org/10.1093/dnares/dsw015.

Maruyama, Takeshi, Stephanie K. Dougan, Matthias C. Truttmann, Angelina M. Bilate, Jessica R. Ingram, and Hidde L. Ploegh. “Increasing the Efficiency of Precise Genome Editing with CRISPR-Cas9 by Inhibition of Nonhomologous End Joining.” *Nature Biotechnology* 33, no. 5 (May 2015): 538–42. https://doi.org/10.1038/nbt.3190.

Mason Kenneth. “Abstract | Bacteriophage | Virus.” Scribd. Accessed October 16, 2017. https://www.scribd.com/document/330006885/abstract.

Matteau, Dominick, Marie-Eve Pepin, Vincent Baby, Samuel Gauthier, Mélissa Arango Giraldo, Thomas F. Knight, and Sébastien Rodrigue. “Development of OriC-Based Plasmids for Mesoplasma Florum.” *Applied and Environmental Microbiology* 83, no. 7 (April 1, 2017). https://doi.org/10.1128/AEM.03374-16.

Mazin, Pavel V., Gleb Y. Fisunov, Alexey Y. Gorbachev, Kristina Y. Kapitskaya, Ilya A. Altukhov, Tatiana A. Semashko, Dmitry G. Alexeev, and Vadim M. Govorun. “Transcriptome Analysis Reveals Novel Regulatory Mechanisms in a Genome-Reduced Bacterium.” *Nucleic Acids Research* 42, no. 21 (December 1, 2014): 13254–68. https://doi.org/10.1093/nar/gku976.

McAuliffe, Laura, Roger D. Ayling, Richard J. Ellis, and Robin A. J. Nicholas. “Biofilm-Grown Mycoplasma Mycoides Subsp. Mycoides SC Exhibit Both Phenotypic and Genotypic Variation Compared with Planktonic Cells.” *Veterinary Microbiology* 129, no. 3 (June 22, 2008): 315–24. https://doi.org/10.1016/j.vetmic.2007.11.024.

McAuliffe, Laura, Richard J. Ellis, Katie Miles, Roger D. Ayling, and Robin A. J. Nicholas. “Biofilm Formation by Mycoplasma Species and Its Role in Environmental Persistence and Survival.” *Microbiology (Reading, England)* 152, no. Pt 4 (April 2006): 913–22. https://doi.org/10.1099/mic.0.28604-0.

Miyata, M., and T. Fukumura. “Asymmetrical Progression of Replication Forks Just after Initiation on Mycoplasma Capricolum Chromosome Revealed by Two-Dimensional Gel Electrophoresis.” *Gene* 193, no. 1 (July 1, 1997): 39–47.

Mojica, F. J., C. Díez-Villaseñor, E. Soria, and G. Juez. “Biological Significance of a Family of Regularly Spaced Repeats in the Genomes of Archaea, Bacteria and Mitochondria.” *Molecular Microbiology* 36, no. 1 (April 2000): 244–46.

Mojica, F. J., C. Ferrer, G. Juez, and F. Rodríguez-Valera. “Long Stretches of Short Tandem Repeats Are Present in the Largest Replicons of the Archaea Haloferax Mediterranei and Haloferax Volcanii and Could Be Involved in Replicon Partitioning.” *Molecular Microbiology* 17, no. 1 (July 1995): 85–93.

Mojica, F. J. M., C. Díez-Villaseñor, J. García-Martínez, and C. Almendros. “Short Motif Sequences Determine the Targets of the Prokaryotic CRISPR Defence System.” *Microbiology (Reading, England)* 155, no. Pt 3 (March 2009): 733–40. https://doi.org/10.1099/mic.0.023960-0.

Mojica, Francisco J. M., César Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. “Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements.” *Journal of Molecular Evolution* 60, no. 2 (February 2005): 174–82. https://doi.org/10.1007/s00239-004-0046-3.

Mouches, C., G. Barroso, and J. M. Bové. “Characterization and Molecular Cloning in Escherichia Coli of a Plasmid from the Mollicute Spiroplasma Citri.” *Journal of Bacteriology* 156, no. 2 (November 1983): 952–55.

Mout, Rubul, Moumita Ray, Gulen Yesilbag Tonga, Yi-Wei Lee, Tristan Tay, Kanae Sasaki, and Vincent M. Rotello. “Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing.” *ACS Nano* 11, no. 3 (March 28, 2017): 2452–58. https://doi.org/10.1021/acsnano.6b07600.

Mudahi-Orenstein, Sigalit, Sharon Levisohn, Steven J. Geary, and David Yogev. “Cytadherence-Deficient Mutants of Mycoplasma Gallisepticum Generated by Transposon Mutagenesis.” *Infection and Immunity* 71, no. 7 (July 2003): 3812–20. https://doi.org/10.1128/IAI.71.7.3812-3820.2003.

Müller, Maximilian, Ciaran M. Lee, Giedrius Gasiunas, Timothy H. Davis, Thomas J. Cradick, Virginijus Siksnys, Gang Bao, Toni Cathomen, and Claudio Mussolino. “Streptococcus Thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 24, no. 3 (March 2016): 636–44. https://doi.org/10.1038/mt.2015.218.

Mushegian, A. R., and E. V. Koonin. “A Minimal Gene Set for Cellular Life Derived by Comparison of Complete Bacterial Genomes.” *Proceedings of the National Academy of Sciences* 93, no. 19 (September 17, 1996): 10268–73.

Nakamura, Yoji, Takeshi Itoh, Hideo Matsuda, and Takashi Gojobori. “Biased Biological Functions of Horizontally Transferred Genes in Prokaryotic Genomes.” *Nature Genetics* 36, no. 7 (July 2004): 760–66. https://doi.org/10.1038/ng1381.

Nakata, A., M. Amemura, and K. Makino. “Unusual Nucleotide Arrangement with Repeated Sequences in the Escherichia Coli K-12 Chromosome.” *Journal of Bacteriology* 171, no. 6 (June 1989): 3553–56.

Neidhardt, Frederick C, and Roy Curtiss. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. Washington, D.C.: ASM Press, 1996.

Nishimasu, Hiroshi, F. Ann Ran, Patrick D. Hsu, Silvana Konermann, Soraya I. Shehata, Naoshi Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. “Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA.” *Cell* 156, no. 5 (February 27, 2014): 935–49. https://doi.org/10.1016/j.cell.2014.02.001.

Noskov, Vladimir N., Thomas H. Segall-Shapiro, and Ray-Yuan Chuang. “Tandem Repeat Coupled with Endonuclease Cleavage (TREC): A Seamless Modification Tool for Genome Engineering in Yeast.” *Nucleic Acids Research* 38, no. 8 (May 2010): 2570–76. https://doi.org/10.1093/nar/gkq099.

Nouvel, Laurent X., Pascal Sirand-Pugnet, Marc S. Marenda, Eveline Sagné, Valérie Barbe, Sophie Mangenot, Chantal Schenowitz, et al. “Comparative Genomic and Proteomic Analyses of Two Mycoplasma Agalactiae Strains: Clues to the Macro- and Micro-Events That Are Shaping Mycoplasma Diversity.” *BMC Genomics* 11 (February 2, 2010): 86. https://doi.org/10.1186/1471-2164-11-86.

O’Connell, Mitchell R., Benjamin L. Oakes, Samuel H. Sternberg, Alexandra East-Seletsky, Matias Kaplan, and Jennifer A. Doudna. “Programmable RNA Recognition and Cleavage by CRISPR/Cas9.” *Nature* 516, no. 7530 (December 11, 2014): 263–66. https://doi.org/10.1038/nature13769.

Pelletier, Stephane, Sebastien Gingras, and Douglas R. Green. “Mouse Genome Engineering via CRISPR-Cas9 for Study of Immune Function.” *Immunity* 42, no. 1 (January 20, 2015): 18–27. https://doi.org/10.1016/j.immuni.2015.01.004.

Pereyre, S., C. Laurier Nadalié, C. Bébéar, C. Arfeuille, A. Beby-Defaux, B. Berçot, S. Boisset, et al. “Mycoplasma Genitalium and Trichomonas Vaginalis in France: A Point Prevalence Study in People Screened for Sexually Transmitted Diseases.” *Clinical Microbiology and Infection* 23, no. 2 (February 1, 2017): 122.e1-122.e7. https://doi.org/10.1016/j.cmi.2016.10.028.

Pereyre, Sabine, Pascal Sirand-Pugnet, Laure Beven, Alain Charron, Hélène Renaudin, Aurélien Barré, Philippe Avenaud, et al. “Life on Arginine for Mycoplasma Hominis: Clues from Its Minimal Genome and Comparison with Other Human Urogenital Mycoplasmas.” *PLOS Genetics* 5, no. 10 (2009): e1000677. https://doi.org/10.1371/journal.pgen.1000677.

Pilo, Paola, Bénédicte Fleury, Marc Marenda, Joachim Frey, and Edy Vilei. “Prevalence and Distribution of the Insertion Element ISMag1 In.” *Veterinary Microbiology* 92 (April 1, 2003): 37–48. https://doi.org/10.1016/S0378-1135(02)00311-5.

Pilo, Paola, Joachim Frey, and Edy M. Vilei. “Molecular Mechanisms of Pathogenicity of Mycoplasma Mycoides Subsp. Mycoides SC.” *The Veterinary Journal* 174, no. 3 (November 1, 2007): 513–21. https://doi.org/10.1016/j.tvjl.2006.10.016.

Pilo, Paola, Edy M. Vilei, Ernst Peterhans, Laetitia Bonvin-Klotz, Michael H. Stoffel, Dirk Dobbelaere, and Joachim Frey. “A Metabolic Enzyme as a Primary Virulence Factor of Mycoplasma Mycoides Subsp. Mycoides Small Colony.” *Journal of Bacteriology* 187, no. 19 (October 2005): 6824–31. https://doi.org/10.1128/JB.187.19.6824-6831.2005.

Pourcel, C., G. Salvignol, and G. Vergnaud. “CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies.” *Microbiology (Reading, England)* 151, no. Pt 3 (March 2005): 653–63. https://doi.org/10.1099/mic.0.27437-0.

Qi, Jingjing, Aizhen Guo, Peng Cui, Yingyu Chen, Riaz Mustafa, Xiaoliang Ba, Changmin Hu, et al. “Comparative Geno-Plasticity Analysis of Mycoplasma Bovis HB0801 (Chinese Isolate).” *PLOS ONE* 7, no. 5 (2012): e38239. https://doi.org/10.1371/journal.pone.0038239.

Qi, Lei S., Matthew H. Larson, Luke A. Gilbert, Jennifer A. Doudna, Jonathan S. Weissman, Adam P. Arkin, and Wendell A. Lim. “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.” *Cell* 152, no. 5 (February 28, 2013): 1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

Ran, F. Ann, Patrick D. Hsu, Chie-Yu Lin, Jonathan S. Gootenberg, Silvana Konermann, Alexandro Trevino, David A. Scott, et al. “Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.” *Cell* 154, no. 6 (September 12, 2013): 1380–89. https://doi.org/10.1016/j.cell.2013.08.021.

Ranhand, J M, W O Mitchell, T J Popkin, and R M Cole. “Covalently Closed Circular Deoxyribonucleic Acids in Spiroplasmas.” *Journal of Bacteriology* 143, no. 3 (September 1980): 1194–99.

Razin, S. “The Mycoplasmas.” *Microbiological Reviews* 42, no. 2 (June 1978): 414–70.

Razin, Shmuel. “The Genus Mycoplasma and Related Genera (Class Mollicutes).” In *The Prokaryotes*, edited by Martin Dworkin Professor Dr, Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer, and Erko Stackebrandt, 836–904. Springer US, 2006. https://doi.org/10.1007/0-387-30744-3_29.

Razin, Shmuel, and Richard Herrmann, eds. *Molecular Biology and Pathogenicity of Mycoplasmas*. Boston, MA: Springer US, 2002. https://doi.org/10.1007/b113360.

Razin, Shmuel, David Yogev, and Yehudith Naot. “Molecular Biology and Pathogenicity of Mycoplasmas.” *Microbiology and Molecular Biology Reviews* 62, no. 4 (December 1998): 1094–1156.

Reddy, S P, W G Rasmussen, and J B Baseman. “Molecular Cloning and Characterization of an Adherence-Related Operon of Mycoplasma Genitalium.” *Journal of Bacteriology* 177, no. 20 (October 1995): 5943–51.

Reisch, Chris R., and Kristala L. J. Prather. “The No-SCAR (Scarless Cas9 Assisted Recombineering) System for Genome Editing in *Escherichia Coli*.” *Scientific Reports* 5 (October 14, 2015): srep15096. https://doi.org/10.1038/srep15096.

Rocha, Eduardo P. C., Emmanuel Cornet, and Bénédicte Michel. “Comparative and Evolutionary Analysis of the Bacterial Homologous Recombination Systems.” *PLOS Genetics* 1, no. 2 (2005): e15. https://doi.org/10.1371/journal.pgen.0010015.

Rottem, Shlomo. “Interaction of Mycoplasmas with Host Cells.” *Physiological Reviews* 83, no. 2 (April 2003): 417–32. https://doi.org/10.1152/physrev.00030.2002.

Rouch, Duncan A., Mary E. Byrne, Yeun C. Kong, and Ronald A. Skurray. “The AacA-AphD Gentamicin and Kanamycin Resistance Determinant of Tn4001 from Staphylococcus Aureus: Expression and Nucleotide Sequence Analysis.” *Microbiology* 133, no. 11 (1987): 3039–52. https://doi.org/10.1099/00221287-133-11-3039.

Sampson, Timothy R., Sunil D. Saroj, Anna C. Llewellyn, Yih-Ling Tzeng, and David S. Weiss. “A CRISPR-CAS System Mediates Bacterial Innate Immune Evasion and Virulence.” *Nature* 497, no. 7448 (May 9, 2013): 254–57. https://doi.org/10.1038/nature12048.

Sander, Jeffry D., and J. Keith Joung. “CRISPR-Cas Systems for Editing, Regulating and Targeting Genomes.” *Nature Biotechnology* 32, no. 4 (April 2014): 347–55. https://doi.org/10.1038/nbt.2842.

Shahid, Muhammad A., Philip F. Markham, Marc S. Marenda, Rebecca Agnew-Crumpton, and Amir H. Noormohammadi. “High-Resolution Melting-Curve Analysis of Obg Gene to Differentiate the Temperature-Sensitive Mycoplasma Synoviae Vaccine Strain MS-H from Non-Temperature-Sensitive Strains.” *PLOS ONE* 9, no. 3 (2014): e92215. https://doi.org/10.1371/journal.pone.0092215.

Shahram, Masoud, Robin Nicholas, Roger (Roger J.) Miles, Ann P. Wood, and Donovan P. Kelly. “Kinetics of Substrate Oxidation and Hydrogen Peroxide Production by Mycoplasma Mycoides Subsp. Mycoides Large Colony (LC) Type and Mycoplasma Mycoides Subsp. Capri.” *Research in Veterinary Science* Vol.87, no. No.3 (December 2009): 364–66.

Sharan, Shyam K., Lynn C. Thomason, Sergey G. Kuznetsov, and Donald L. Court. “Recombineering: A Homologous Recombination-Based Method of Genetic Engineering.” *Nature Protocols* 4, no. 2 (2009): 206–23. https://doi.org/10.1038/nprot.2008.227.

Siguier, Patricia, Edith Gourbeyre, and Mick Chandler. “Bacterial Insertion Sequences: Their Genomic Impact and Diversity.” *FEMS Microbiology Reviews* 38, no. 5 (September 2014): 865–91. https://doi.org/10.1111/1574-6976.12067.

Simmons, Warren L., and Kevin Dybvig. “Catalase Enhances Growth and Biofilm Production of Mycoplasma Pneumoniae.” *Current Microbiology* 71, no. 2 (August 2015): 190–94. https://doi.org/10.1007/s00284-015-0822-x.

Sirand-Pugnet, Pascal, Carole Lartigue, Marc Marenda, Daniel Jacob, Aurélien Barré, Valérie Barbe, Chantal Schenowitz, et al. “Being Pathogenic, Plastic, and Sexual While Living with a Nearly Minimal Bacterial Genome.” *PLoS Genetics* 3, no. 5 (May 18, 2007): e75. https://doi.org/10.1371/journal.pgen.0030075.

Slaymaker, Ian M., Linyi Gao, Bernd Zetsche, David A. Scott, Winston X. Yan, and Feng Zhang. “Rationally Engineered Cas9 Nucleases with Improved Specificity.” *Science (New York, N.Y.)* 351, no. 6268 (January 1, 2016): 84–88. https://doi.org/10.1126/science.aad5227.

Sonenshein, A. L., James A. Hoch, and Richard Losick. *Bacillus Subtilis and Other Gram-Positive Bacteria*. American Society of Microbiology, 1993. https://doi.org/10.1128/9781555818388.

Sorek, Rotem, C. Martin Lawrence, and Blake Wiedenheft. “CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea.” *Annual Review of Biochemistry* 82 (2013): 237–66. https://doi.org/10.1146/annurev-biochem-072911-172315.

Stern, Adi, Leeat Keren, Omri Wurtzel, Gil Amitai, and Rotem Sorek. “Self-Targeting by CRISPR: Gene Regulation or Autoimmunity?” *Trends in Genetics: TIG* 26, no. 8 (August 2010): 335–40. https://doi.org/10.1016/j.tig.2010.05.008.

Tardy, Florence, Eric Baranowski, Laurent-Xavier Nouvel, Virginie Mick, Lucía Manso-Silvàn, François Thiaucourt, Patricia Thébault, et al. “Emergence of Atypical Mycoplasma Agalactiae Strains Harboring a New Prophage and Associated with an Alpine Wild Ungulate Mortality Episode.” *Applied and Environmental Microbiology* 78, no. 13 (July 1, 2012): 4659–68. https://doi.org/10.1128/AEM.00332-12.

Tardy, Florence, Virginie Mick, Emilie Dordet-Frisoni, Marc Serge Marenda, Pascal Sirand-Pugnet, Alain Blanchard, and Christine Citti. “Integrative Conjugative Elements Are Widespread in Field Isolates of Mycoplasma Species Pathogenic for Ruminants.” *Applied and Environmental Microbiology* 81, no. 5 (March 2015): 1634–43. https://doi.org/10.1128/AEM.03723-14.

Taylor-Robinson, D., H. A. Davies, P. Sarathchandra, and P. M. Furr. “Intracellular Location of Mycoplasmas in Cultured Cells Demonstrated by Immunocytochemistry and Electron Microscopy.” *International Journal of Experimental Pathology* 72, no. 6 (December 1991): 705–14.

Terahara, Natsuho, Isil Tulum, and Makoto Miyata. “Transformation of Crustacean Pathogenic Bacterium Spiroplasma Eriocheiris and Expression of Yellow Fluorescent Protein.” *Biochemical and Biophysical Research Communications* 487, no. 3 (June 3, 2017): 488–93. https://doi.org/10.1016/j.bbrc.2017.03.144.

Thiaucourt, F., A. Yaya, H. Wesonga, O. J. Huebschle, J. J. Tulasne, and A. Provost. “Contagious Bovine Pleuropneumonia. A Reassessment of the Efficacy of Vaccines Used in Africa.” *Annals of the New York Academy of Sciences* 916 (2000): 71–80.

Thiaucourt, Francois, Lucia Manso-Silvan, Woubit Salah, Valérie Barbe, Benoit Vacherie, Daniel Jacob, Marc Breton, et al. “Mycoplasma Mycoides, from ‘Mycoides Small Colony’ to ‘Capri’. A Microevolutionary Perspective.” *BMC Genomics* 12, no. 1 (February 16, 2011): 114. https://doi.org/10.1186/1471-2164-12-114.

Thomas, Anne, Annick Linden, Jacques Mainil, Daniela F. Bischof, Joachim Frey, and Edy M. Vilei. “Mycoplasma Bovis Shares Insertion Sequences with Mycoplasma Agalactiae and Mycoplasma Mycoides Subsp. Mycoides SC: Evolutionary and Developmental Aspects.” *FEMS Microbiology Letters* 245, no. 2 (April 1, 2005): 249–55. https://doi.org/10.1016/j.femsle.2005.03.013.

Tock, Mark R, and David TF Dryden. “The Biology of Restriction and Anti-Restriction.” *Current Opinion in Microbiology*, Host--microbe interactions: fungi / edited by Howard Bussey · Host--microbe interactions: parasites / edited by Artur Scherf · Host--microbe interactions: viruses / edited by Margaret CM Smith, 8, no. 4 (August 1, 2005): 466–72. https://doi.org/10.1016/j.mib.2005.06.003.

Truong, Dong-Jiunn Jeffery, Karin Kühner, Ralf Kühn, Stanislas Werfel, Stefan Engelhardt, Wolfgang Wurst, and Oskar Ortiz. “Development of an Intein-Mediated Split-Cas9 System for Gene Therapy.” *Nucleic Acids Research* 43, no. 13 (July 27, 2015): 6450–58. https://doi.org/10.1093/nar/gkv601.

Tu, A. H., L. L. Voelker, X. Shen, and K. Dybvig. “Complete Nucleotide Sequence of the Mycoplasma Virus P1 Genome.” *Plasmid* 45, no. 2 (March 2001): 122–26. https://doi.org/10.1006/plas.2000.1501.

Vandecraen, Joachim, Michael Chandler, Abram Aertsen, and Rob Van Houdt. “The Impact of Insertion Sequences on Bacterial Genome Plasticity and Adaptability.” *Critical Reviews in Microbiology* 43, no. 6 (November 2, 2017): 709–30. https://doi.org/10.1080/1040841X.2017.1303661.

Vasconcelos, Ana Tereza R., Henrique B. Ferreira, Cristiano V. Bizarro, Sandro L. Bonatto, Marcos O. Carvalho, Paulo M. Pinto, Darcy F. Almeida, et al. “Swine and Poultry Pathogens: The Complete Genome Sequences of Two Strains of Mycoplasma Hyopneumoniae and a Strain of Mycoplasma Synoviae.” *Journal of Bacteriology* 187, no. 16 (August 2005): 5568–77. https://doi.org/10.1128/JB.187.16.5568-5577.2005.

Vilei, E. M., E. M. Abdo, J. Nicolet, A. Botelho, R. Gonçalves, and J. Frey. “Genomic and Antigenic Differences between the European and African/Australian Clusters of Mycoplasma Mycoides Subsp. Mycoides SC.” *Microbiology (Reading, England)* 146 (Pt 2) (February 2000): 477–86. https://doi.org/10.1099/00221287-146-2-477.

Vilei, Edy M., Jacques Nicolet, and Joachim Frey. “IS1634, a Novel Insertion Element Creating Long, Variable-Length Direct Repeats Which Is Specific for Mycoplasma Mycoides Subsp. Mycoides Small-Colony Type.” *Journal of Bacteriology* 181, no. 4 (February 1999): 1319–23.

Voelker, L. L., and K. Dybvig. “Sequence Analysis of the Mycoplasma Arthritidis Bacteriophage MAV1 Genome Identifies the Putative Virulence Factor.” *Gene* 233, no. 1–2 (June 11, 1999): 101–7.

Wagner, Jeffrey C., Randall J. Platt, Stephen J. Goldfless, Feng Zhang, and Jacquin C. Niles. “Efficient CRISPR/Cas9-Mediated Genome Editing in P. Falciparum.” *Nature Methods* 11, no. 9 (September 2014): 915–18. https://doi.org/10.1038/nmeth.3063.

Wang, Yang, Li Yi, Fanqing Zhang, Xusheng Qiu, Lei Tan, Shengqing Yu, Xiangchao Cheng, and Chan Ding. “Identification of Genes Involved in Mycoplasma Gallisepticum Biofilm Formation Using Mini-Tn4001-SGM Transposon Mutagenesis.” *Veterinary Microbiology* 198, no. Supplement C (January 1, 2017): 17–22. https://doi.org/10.1016/j.vetmic.2016.11.021.

Wang, Yi, Zhong-Tian Zhang, Seung-Oh Seo, Patrick Lynn, Ting Lu, Yong-Su Jin, and Hans P. Blaschek. “Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide Modification, and Desirable ‘Clean’ Mutant Selection in Clostridium Beijerinckii as an Example.” *ACS Synthetic Biology* 5, no. 7 (July 15, 2016): 721–32. https://doi.org/10.1021/acssynbio.6b00060.

Weisburg, W. G., J. G. Tully, D. L. Rose, J. P. Petzel, H. Oyaizu, D. Yang, L. Mandelco, J. Sechrest, T. G. Lawrence, and J. Van Etten. “A Phylogenetic Analysis of the Mycoplasmas: Basis for Their Classification.” *Journal of Bacteriology* 171, no. 12 (December 1989): 6455–67.

Westberg, Joakim, Anja Persson, Anders Holmberg, Alexander Goesmann, Joakim Lundeberg, Karl-Erik Johansson, Bertil Pettersson, and Mathias Uhlén. “The Genome Sequence of Mycoplasma Mycoides Subsp. Mycoides SC Type Strain PG1T, the Causative Agent of Contagious Bovine Pleuropneumonia (CBPP).” *Genome Research* 14, no. 2 (February 2004): 221–27. https://doi.org/10.1101/gr.1673304.

Whetzel, Patricia L., Linda L. Hnatow, Calvin L. Keeler, and John E. Dohms. “Transposon Mutagenesis of Mycoplasma Gallisepticum.” *Plasmid* 49, no. 1 (January 2003): 34–43.

Whitley, J. C., and L. R. Finch. “Location of Sites of Transposon Tn916 Insertion in the Mycoplasma Mycoides Genome.” *Journal of Bacteriology* 171, no. 12 (December 1, 1989): 6870–72. https://doi.org/10.1128/jb.171.12.6870-6872.1989.

Widjaja, Michael, Iain J. Berry, Elsa J. Pont, Matthew P. Padula, and Steven P. Djordjevic. “P40 and P90 from Mpn142 Are Targets of Multiple Processing Events on the Surface of Mycoplasma Pneumoniae.” *Proteomes* 3, no. 4 (December 16, 2015): 512–37. https://doi.org/10.3390/proteomes3040512.

Woese, C R. “Bacterial Evolution.” *Microbiological Reviews* 51, no. 2 (June 1987): 221–71.

Woese, C. R., L. J. Magrum, R. Gupta, R. B. Siegel, D. A. Stahl, J. Kop, N. Crawford, et al. “Secondary Structure Model for Bacterial 16S Ribosomal RNA: Phylogenetic, Enzymatic and Chemical Evidence.” *Nucleic Acids Research* 8, no. 10 (May 24, 1980): 2275–93.

Wright, Addison V., James K. Nuñez, and Jennifer A. Doudna. “Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering.” *Cell* 164, no. 1 (January 14, 2016): 29–44. https://doi.org/10.1016/j.cell.2015.12.035.

Xu, Han, Tengfei Xiao, Chen-Hao Chen, Wei Li, Clifford A. Meyer, Qiu Wu, Di Wu, et al. “Sequence Determinants of Improved CRISPR SgRNA Design.” *Genome Research* 25, no. 8 (August 2015): 1147–57. https://doi.org/10.1101/gr.191452.115.

Yang, Hui, and Dinshaw J. Patel. “New CRISPR-Cas Systems Discovered.” *Cell Research* 27, no. 3 (March 2017): 313–14. https://doi.org/10.1038/cr.2017.21.

Ye, Fengchun, Joël Renaudin, Joseph-Marie Bové, and Frédéric Laigret. “Cloning and Sequencing of the Replication Origin (OriC) of TheSpiroplasma Citri Chromosome and Construction of Autonomously Replicating Artificial Plasmids.” *Current Microbiology* 29, no. 1 (July 1, 1994): 23–29. https://doi.org/10.1007/BF01570187.

Yin, Hao, Wen Xue, Sidi Chen, Roman L. Bogorad, Eric Benedetti, Markus Grompe, Victor Koteliansky, Phillip A. Sharp, Tyler Jacks, and Daniel G. Anderson. “Genome Editing with Cas9 in Adult Mice Corrects a Disease Mutation and Phenotype.” *Nature Biotechnology* 32, no. 6 (June 2014): 551–53. https://doi.org/10.1038/nbt.2884.

Yuen, Kit-San, Chi-Ping Chan, Nok-Hei Mickey Wong, Chau-Ha Ho, Ting-Hin Ho, Ting Lei, Wen Deng, et al. “CRISPR/Cas9-Mediated Genome Editing of Epstein-Barr Virus in Human Cells.” *The Journal of General Virology* 96, no. Pt 3 (March 2015): 626–36. https://doi.org/10.1099/jgv.0.000012.

Yus, Eva, Tobias Maier, Konstantinos Michalodimitrakis, Vera van Noort, Takuji Yamada, Wei-Hua Chen, Judith A. H. Wodke, et al. “Impact of Genome Reduction on Bacterial Metabolism and Its Regulation.” *Science (New York, N.Y.)* 326, no. 5957 (November 27, 2009): 1263–68. https://doi.org/10.1126/science.1177263.

Zhang, Yihan, Wei Qin, Xiaochan Lu, Jason Xu, Haigen Huang, Haipeng Bai, Song Li, and Shuo Lin. “Programmable Base Editing of Zebrafish Genome Using a Modified CRISPR-Cas9 System.” *Nature Communications* 8, no. 1 (July 25, 2017): 118. https://doi.org/10.1038/s41467-017-00175-6.

Zhou, Yan, Yong Liu, Dianna Hussmann, Peter Brøgger, Rasha Abdelkadhem Al-Saaidi, Shuang Tan, Lin Lin, et al. “Enhanced Genome Editing in Mammalian Cells with a Modified Dual-Fluorescent Surrogate System.” *Cellular and Molecular Life Sciences: CMLS* 73, no. 13 (2016): 2543–63. https://doi.org/10.1007/s00018-015-2128-3.

Zimmerman, C.-U., and R. Herrmann. “Synthesis of a Small, Cysteine-Rich, 29 Amino Acids Long Peptide in Mycoplasma Pneumoniae.” *FEMS Microbiology Letters* 253, no. 2 (December 1, 2005): 315–21. https://doi.org/10.1016/j.femsle.2005.09.054.

Zuker, Michael. “Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction.” *Nucleic Acids Research* 31, no. 13 (July 1, 2003): 3406–15.

Acknowledgments

Acknowledgments

I would like to thank many people of the lab and outside of it for helping me during different moments of this thesis. Camille Jollard for her work on the MCS2 gene. Patricia Thebault for helping with the internship of Mamadou Sall and Mamadou himself for all his achievements that helped me advance with the experiments. Brian White for the corrections in the introduction of the thesis manuscript. Marc Breton for his work on the mobile elements of the mollicutes and the Table that i introduced in the discussion part. Laure Beven for the help with the statistical analysis of the results on the PAM identification and all the funny discussions during the centrifugations in her lab.

Apart from the people in the lab, a lot of great people assisted, inspired, supported and played their special role to complete the race. First, I would like to thank my brother Orestis, who has been there since the beginning, always lending his ear to hear my achievements, struggles, successes, problems and everyday life things. His calm energy and his simple way of seeing things, helped me keep a straight head to my purposes and the many hours we spent playing DOTA helped me relax and gain the courage to keep my scientific interest awake. Alongside him, I would like to thank my friend Kostas, who together with Orestis, was there from the beginning, helping me thinking straight through the rough times and being the best skype philosophical interlocutor.

The next people in line is my parents, Panagiotis and Roza, the people that helped me during the first years of my life and allowed me to grow into a healthy man capable of taking my own decisions and it is also them who suggested the superior studies abroad, which is one of the reasons I came to France and postulate for Master and Doctorate studies.

Next, I would like to thank Mara, my girlfriend, one of the best people I have the honor to meet in my life. She came in it about a year ago and since then she was always there to consult me whenever I had a mishap on work or be extremely happy with my good results. She helped me put in perspective my goals and theirs of the lab and realize that the thesis is not only about publications but also about learning to failing and move forward, not overthink things, organize my time to take pleasure of the simple things. Instead of being stressful during the writing of this manuscript, she gave me the space and time I needed and keep a positive energy when around me and helped me in her own way finish this manuscript without stress.

Another person that really helped during a period I was feeling under the weather was Mamadou Sall. We shared the same bench and project for two months during his internship for the Master1, but it was enough to boost my moral and push me to reach the end! Always at the office around 9.00 after 4 hours of work in a fast food restaurant, keeping a happy face and a positive attitude, sometimes until 19h00. Seen him, gave me the inspiration I needed to put aside my problems and clear my head to advance the thesis work.

During these three years I met a lot of people inside and outside of the lab who I will always remember and been thankful for all the great memories we created together. Geraldine, Estelle, Vincent, Michela, Yannina, Carole, Luis, Flavie, Camille, David, Alexandra, Carlos, Xavier are some of the people that I met in the lab and I would like to thank for the role they played.

Outside of the lab, the number of people I would like to thank is limitless, so I will only make some significant mentions. My Tunisian friends, Mazen, Oussama, Wahbi and Hedi, my artist and all of

them my khouya. My old roommates, Camille and Giovana that kept me company for two entire years and gave me some great and funny memories. Spiros and Giannis, the Greek guys that I met at Bordeaux and helped me to have some great time. Finally, my French friends, Cécile, Antony, Sophie, Marie, Sylvan, Xavier, Guillaume, Laura and Charlotte that also offered me some great time and some unforgettable memories.

In conclusion, a big thank you to everybody whose name is not cited above, because as I mentioned before the goals of this thesis could not be achieved without their help.