International Journal of Medical Microbiology 307 (2017) 508–520

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

Comparative genomics of four *Mycoplasma* species of the human urogenital tract: Analysis of their core genomes and virulence genes

Orville St. E. Roachford[a,*], Karen E. Nelson[b], Bidyut R. Mohapatra[a]

[a] Department of Biological and Chemical Sciences, The University of the West Indies, Cave Hill Campus, Bridgetown BB 11000, Barbados
[b] J. Craig Venter Institute, 9714 Medical Center Drive, Rockville, MD 20850, USA

ARTICLE INFO

Keywords:
Mycoplasma
Core genome
Lipoproteins
Virulence
Prophage
Tandem repeats

ABSTRACT

The variation in *Mycoplasma* lipoproteins attributed to genome rearrangements and genetic insertions leads to phenotypic plasticity that allows for the evasion of the host's defence system and pathogenesis. This paper compared for the first time the genomes of four human urogenital *Mycoplasma* species (*M. penetrans* HF-2, *M. fermentans* JER, *M. genitalium* G37 and *M. hominis* PG21) to categorise the metabolic functions of the core genes and to assess the effects of tandem repeats, phage-like genetic elements and prophages on the virulence genes. The results of this comparative *in silico* genomic analysis revealed that the genes constituting their core genomes can be separated into three distinct categories: nuclear metabolism, protein metabolism and energy generation each making up 52%, 31% and 23%, respectively. The genomes have repeat sequences ranging from 3.7% in *M. hominis* PG21 to 9.5% in *M. fermentans* JER. Tandem repeats (mostly minisatellites) and phage-like proteins (including DNA gyrases/topoisomerases) were randomly distributed in the *Mycoplasma* genomes. Here, we identified a coiled-coil structure containing protein in *M. penetrans* HF-2 which is significantly similar to the Mem protein of *M. fermentans* ϕMFV1. Therefore, a *Mycoplasma* prophage seems to be embedded within *M. penetrans* HF-2 unannotated genome. To the best of our knowledge, no *Mycoplasma* phages or prophages have been detected in *M. penetrans*. This study is important not only in understanding the complex genetic factors involved in phenotypic plasticity and virulence in the relatively understudied *Mycoplasma* species but also in elucidating the effective arrangement of their redundant minimal genomes.

1. Introduction

The *Mycoplasma* (class Mollicutes) are an interesting group of Gram negative bacteria in that they are not only the smallest self-replicating bacteria that are known to exist so far but they also lack a protective cell wall (Baseman and Tully, 1997; Razin, 2006). Their plasma membranes are thus exposed to the immediate environment making them vulnerable to the host's humoral immune response. However, their evolution has allowed them to survive within this milieu. By losing their cell walls, they have developed phenotypic plasticity through the ability to constantly change the antigenic lipoproteins in their plasma membranes (reviewed in Razin et al., 1998; Citti et al., 2010).

Mycoplasmas generally adhere to the host's epithelial cells via a complex specialized structure called the tip organelle. The tip organelle is mainly composed of adhesins and cytadherence accessory proteins (Razin et al., 1998). Mycoplasmas adherence, has been reported as a major factor, which triggers colonization and pathogenicity (Rottem, 2003). Once attached to the epithelial cells, the mycoplasmas are able to persist and evade the host immune system through variation of surface proteins (antigens) and cause chronic infection (Razin et al., 1998). It has been documented that *M. genitalium* G37 can attach to the mucin component of the epithelial mucus membrane via the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase. This glycolytic enzyme thus acts as an adhesin in some species of mycoplasmas (Alvarez et al., 2003). Similarly, oligopeptide permease substrate-binding protein (OppA) of *M. hominis* PG21, a multifunctional lipoprotein (Henrich et al., 1999; Hopfe et al., 2011) and OppF of *M. penetrans* (Distelhorst et al., 2017) have been reported to be involved in cytadherence to host cells. This dual functionality of some key proteins appears to have evolved to allow the mycoplasmas to survive in the host environment whilst maintaining minimal genomes. *M. penetrans* HF-2 can penetrate mammalian cells by adhering to fibronectin (a component of the host cell membrane) (Girón et al., 1996). Hence their ability to bind to specific cell membrane proteins may be considered an important factor for determining their existence in extracellular or intracellular niches.

Mycoplasmas having minimal genomes and limited biosynthesis capabilities heavily rely on their host for exogenous fatty acids,

* Corresponding author.
E-mail addresses: orville.roachford@mycavehill.uwi.edu, oster6413@gmail.com (O.S.E. Roachford).

http://dx.doi.org/10.1016/j.ijmm.2017.09.006
Received 20 May 2017; Received in revised form 29 August 2017; Accepted 4 September 2017

cholesterol or complex lipids (Kornspan and Rottem, 2012). Generally, thirty five percent of total lipids in mycoplasmas (compared to 1–10% in other bacteria) are in the plasma membrane, with the exogenous sterols being incorporated without structural modification (Razin et al., 1963; Dahl, 1993). While the species of *Mycoplasma* lack the ability to synthesize carbohydrates such as glucose, sucrose, mannose, maltose and sorbitol, they are able to take up those molecules by the PEP-PTS (phosphoenolpyruvate-dependent-phosphotransferase system) to generate energy. This is accomplished via the Embden-Meyerhof (glycolytic) pathway (Desantis and Pollack, 1989), the anabolic pentose phosphate pathway (Arraes et al., 2007) and the arginine dihydrolase pathway (Pereyre et al., 2009).

By using the host's lipids, the proteins of *Mycoplasma* species undergo lipoylation to produce a myriad of membrane-bound lipoproteins. The protein/polypeptide moiety, which is surface exposed, is the immunogenic component that gives rise to antigenic variation. The lipid moiety is embedded in the plasma membrane. Through segmental reciprocal recombination with repeat sequences within their genomes mycoplasmas can generate variation of these surface membrane lipoproteins (Peterson et al., 1995; Iverson-Cabral et al., 2007; Burgos et al., 2012) which along with adhesins act as virulence factors (Baseman and Tully, 1997). Other genetic mechanisms involved in lipoprotein size variation and phase variation (ON/OFF switching) are DNA slippage (which uses repeat sequences), gene conversion and gene duplication via antigenic drift (reviewed in Citti et al., 2010).

Diacylated and triacylated mycoplasma lipoproteins also trigger signaling pathways after attaching to the host cells by activating Toll-like receptors (TLRs). This binding activates macrophages, monocytes and lymphocytes which initiate a proinflammatory response. This response leads to the secretion of cytokines such as chemokines, interleukin-I and 6 (Il-1, IL-6) and tumor necrosis factor-alpha (TNF-α) (Jan et al., 1995; Rottem, 2003; Zuo et al., 2009). Lipoproteins also induce apoptosis of host cells by releasing ATP which binds to $P2X_7$ purigenic receptors (Into et al., 2002; Lister et al., 2007; Hopfe and Henrich, 2008). The mycoplasma lipoproteins in the MIB-MIP [Mycoplasma immunoglobulins (Ig) binding protein- Mycoplasma Ig protease] system can deactivate host immunoglobulins (IgG) and prevent antigen-antibody binding by cleavage of the VH domain of IgG (Arfi et al., 2016; Grover et al., 2014). Consequently, *Mycoplasma* survive and persist within their hosts by evading and modulating the hosts' immune system. Interestingly, pathogenicity can also be partly mediated by bacteriophages (Wagner and Waldor, 2002). For example, phage ϕMFV1 encodes a unique coiled-coil membrane surface protein (Mem) that enables the pathogenicity of *M. fermentans* (Röske et al., 2004).

Due to the panoply of adhesins, cytadherence accessory proteins and variable surface lipoproteins, there is significant sequence homology to the host's structural proteins which results in molecular mimicry (Baseman and Tully, 1997; Rottem 2003). This antigenic mimicry along with the persistent interaction between host and mycoplasmas can lead to a sustained inflammatory autoimmune response and subsequently to chronic disease (Rottem, 2003). Some *Mycoplasma* species (acting as the causative agent or as a co-factor) have been implicated in a number of acute diseases (e.g. acute urethritis and prostatitis) (Falk et al., 2005; Dehon and McGowin, 2017), chronic diseases such as fibromyalgia, systemic lupus erythematosus (SLE), infertility (Nicolson et al., 2000; Krieger and Riley, 2004; Gdoura et al., 2007) and cancer (reviewed in Zarei et al., 2013). Also of interest, is their vertical transmission from mother to child (Costello et al., 2017).

With their small size, minimal genomes, phenotypic plasticity and their abilities to alter shape, to invade host cells and to develop specialized attachment tip organelles, it seems probable from an evolutionary point of view that the species of *Mycoplasma* are evolving towards adapting to an intracellular habitat. In this study, we have performed for the first time a comparative genomic analysis on four selected human urogenital *Mycoplasma* species (*Mycoplasma fermentans* JER, *M. hominis* PG21, *M. genitalium* G37 and *M. penetrans* HF-2). These four strains were selected because their physiology and interaction with host are well studied. In addition, they are the only urogenital mycoplasma reference strains whose annotated genomes are available in both the BioCyc (Caspi et al., 2016) and RAST (Aziz et al., 2008) databases for genome comparison and subsystem function analysis. As a consequence, this study has a limitation in predicting the species-specific characteristics of urogenital mycoplasmas.

M. fermentans JER and *M. hominis* PG21 exist extracellularly, adhering to the surface of the host cells (Ladefoged et al., 1995). On the other hand, *M. penetrans* HF-2 and *M. genitalium* G37 (under certain conditions) can exist intracellularly in non-phagocytic cells (Rottem, 2003). *M. penetrans* HF-2 has the distinct ability to actively penetrate the host cell and become localised in its cytoplasm (Andreev et al., 1995; Girón et al., 1996) as well as *M.hominis* which has been shown to invade HeLa cells (Hopfe et al., 2013). All of these *Mycoplasma* species possess specialized tip organelles except *M. hominis* PG21, which lacks a well-defined tip structure (Kitzerow et al., 1999). The lipoproteins constituting the tip organelle are species-specific which might be due to the structure being a homoplasy as a consequence of evolutionary convergence (Distelhorst et al., 2017). Of these species, only *M. hominis* PG21 resides in the urogenital tract as a commensal. It is also opportunistically pathogenic causing infection when the immune state of the host is compromised (Ladefoged et al., 1995). This comparative phylogenomic and pathogenomic study can provide insight as to which orthologous gene(s) of these mycoplasmas dictates virulence, pathogenicity and cellular invasion. Analyzing the genomes for the presence of prophages and/or other genetic elements (e.g. tandem repeats) can further lead to our understanding as to how these elements affect the virulence genes.

2. Materials and methods

2.1. Mycoplasma species sequences and phylogenetic analysis

The 16S rRNA gene sequences of *Mycoplasma* species that infect the human urogenital tract (Table 1) were downloaded from GenBank (Benson et al., 2013). The sequences were aligned using ClustalW within MEGA 7.0 (Kumar et al., 2012) and a phylogenetic tree was constructed using the Neighbor-Joining algorithm with default settings.

Three highly conserved nuclear and energy-production gene sequences (*gapA, tktA* and *ligA*) for *M. penetrans* HF-2, *M. fermentans* JER, *M. genitalium* G37 and *M. hominis* PG21 that infect the human urogenital tract were downloaded from GenBank. Concatenated gene sequences for each species of *Mycoplasma* were prepared by merging multi-fasta files into a single FASTA file using the web based Sequence Manipulation Suite: Combine FASTA (Stothard, 2000). The single FASTA file was then converted to a GBK file using DNA Baser version 4.36.0 (DNA Sequence Assembler V4, 2013). The concatenated sequences were aligned and a phylogenetic tree was generated using ClustalW and the Neighbor-Joining algorithm, respectively, within MEGA 7.0.

2.2. Gene annotation and genome comparisons of four Mycoplasma species

Gene function and subsystems were determined for each *Mycoplasma* species using the functional gene comparison utility tool on the RAST (Rapid Annotation using Subsytems Technology) prokaryotic genome annotation server (Aziz et al., 2008). The results are listed in supplementary Table S1.

Four-way and two-way genome comparisons between the genomes of the four *Mycoplasma* species were performed using BioCyc (Caspi et al., 2016). Similarities and differences between species were tabulated in supplementary Tables S2 (Four-way genome comparison between four species of *Mycoplasma*) and S3 (Two-way genome comparison between four species of *Mycoplasma*).