[Frontiers in Bioscience 12, 2020-2028, January 1, 2007]

Mycoplasma genomics: tailoring the genome for minimal life requirements through reductive evolution

Ahmed Fadiel[1], Kenneth D. Eichenbaum[2], Nermin El Semary[3],Brittiny Epperson[1]

[1]Yale School of Medicine, New Haven, CT, 06511, [2]Mount Sinai School of Medicine, NY, NY, 10029, [3] Department of Botany and Microbiology, Faculty of Science, University of Helwan, Cairo, Egypt

TABLE OF CONTENTS

1. ABSTRACT

Prokaryotic organisms of the genus *Mycoplasm*a are characterized by their small body and genome size containing a 0.6 -1.35 M bp genome. The genome is noted for its low G+C frequency ranging from 8-40 mol%. The *Mycoplasma* genus stems from the class *Mollicutes* (for soft skin), which lacks the cell walls and external motility appendagesoften present in other bacteria. To date, there are more than 100 known species of *Mycoplasma*. 34 species have been partially or completely sequenced. Widely known pathogenic species of *Mycoplasma* include: *M. pneumoniae*, causing pneumonia and other respiratory disorders, and *M. genitalium*, which are involved in pelvic inflammatory disease. Because of their small genome size, *Mycoplasmas* provide researchers a unique model of the minimal genomic requirements to maintain life. As the number of complete *Mycoplasma* genomes increase, these organisms become more established, thus laying the foundation for mapping evolutionary development. This manuscript provides an overview and update on *Mycoplasma* research, with particular focus on current genomics.

2. INTRODUCTION

Mycoplasmas are distinct bacterial microorganisms, small in size (0.2 - 0.3 µm) and devoid of both cell wall and motility appendages (1, 2). When *Mycoplasmas* were initially studied, their small size permitted their passage through micro-filters that blocked other bacteria. This mistakenly led scientists to believe that *Mycoplasmas* were viruses. In the 1950 and early sixties it was established that *Mycoplasmas* are bacteria (3). In addition, *Mycoplasmas* are characterized for their possession of the smallest known genomes and distinctively low G+C contents (4). Biomedically, several pathogenic *Mycoplasma* species were discovered to be active in human microflora and were associated with diseases such as AIDS, urogenital diseases and cancer (2). However, the discrete mechanisms of *Mycoplasma* pathogenicity are yet to be understood (5). The organism evades immune system response by establishing residence in white blood cells. Once in the circulation, *Mycoplasma* can cross the blood-brain barrier and infiltrate the cerebral spinal fluid (6). Once inside the host *Mycoplasma* starts to compete for nutrients. In addition, the presence of *Mycoplasmas* in hosts' cells might lead to DNA mutations.

Like retroviruses, *Mycoplasmas* have the capacity for cellular invasion, self-replication and the initiation of a variety of immune responses (7). However, *Mycoplasmas*, unlike viruses, are viable in body fluids and do not require living cell hosts for DNA replication and growth. Whereas viruses can demonstrate high specificity to their host organ targets, the *Mycoplasma* pathogen has a high degree of adaptability to many regions of the body. Thus, *Mycoplasmas* can be termed "stealth pathogens" (6). The small genome size has made *Mycoplasma* ideal for genomic studies facilitating rapid and cost-effective completion of the sequencing of entire bacterial genomes and producing vast quantities of relevant data. In this review we provide a primer on genomics and informatics of *Mycoplasmas*.

3. *MYCOPLASMA* GENOMICS

3.1. *Mycoplasma* Genome Sequencing: a historical perspective

With the smallest known genome sizes *Mycoplasma* offers a unique model for examining the minimal requirements for establishing life. Early attempts of deciphering Mollicutes genomes were primarily based on physical mapping synthesized from genomic sequence fragments (8, 9). Morowitz began studies to define the comprehensive cellular machinery of *Mycoplasma* in 1984 (10). By the early 1990s, laboratory groups at Harvard University focused their sequencing efforts on the *M. capricolum* genome. However, this initiative was curtailed in 1995 as a result of technical limitations. To understand genome organization and DNA repeat distribution, researchers at the University of Heidelberg sought to sequence the pathogenic species *M. pneumoniae* using a cosmid library. This sequencing approach was successful in identifying approximately 90% of the organism genome and spanned approximately 720 Kbp (11, 12). However, the exact size of the pathogen genome was not fully characterized by that time. Other studies estimated the genome size of the *M. pneumoniae* to contain close to 800 Kbp (13, 14, 15). In 1995 a group from The Institute for Genomic Research (TIGR), utilized a new shotgun sequencing technique considered a breakthrough in genome technologies. This technique offered improved reliability and speed in obtaining the comprehensive genome sequence within 6 months (16). This technology involves cloning and sequencing of small sheared DNA fragments which can subsequently be arranged into contigs and supercontigs. This technology incorporates computational biology, making use of computational algorithms to assemble sequenced fragments (17). This technique has subsequently facilitated the completion of close to 50 complete genomes of prokaryotes including *Mycoplasma*, which are currently available (at *www.NCBI.nlm.nih.gov*). *Mycoplasma* genomes (Table 1) along with other families of prokaryotic genomes provide templates for comparative genome and evolutionary analyses (18).

3.2. Genome sizing and genomic redundancies

Within the class Mollicutes, *Acholeplasma* and *Spiroplasma* species, thought to be older on the evolutionary hierarchy, have larger genomes than *Mycoplasma. S. ixodetis* has 2,220 Kb in comparison to *M. genitalium's* 580 Kb (19, 20). The correlation between genome size and phylogenetic rank remains a source of controversy (21) prompting scientists to develop several hypotheses. One theory holds that genome reduction may have resulted from evolutionarily driven loss of genomic information. It was further hypothesized that this loss may have forced the species to adopt a parasitic modality (22). Genomic degeneration can explain the concept of reductive evolution (23). However, *in vitro* studies demonstrate an absence of correlation between genome size and adaptation in non-parasitic environments (21).

The exact mechanism of *Mycoplasma* genome size reduction is not completely understood. It is argued that many factors contributed to shaping the final size of each *Mycoplasma*, including the dynamic behavior of repetitive elements and integration of viral sequences and mutations (23, 24, 25, 26, 27). Variation in the spectrum of *Mycoplasma* size and mode provide an ample opportunity to explore the functional content of the genomes and the evolutionary relationships between them (28). Small genomic size allows for easy handling with current computational capacity. Comparative studies frequently show that most bacterial proteins are highly conserved. This facilitates functional assignment to coding regions by homology. Interestingly, the number of genes required for certain cellular function is often independent of genome size, demonstrating that proscribed discrete protein functional pathways may be required for basic homeostasis. With more complex amino acid biosynthesis and energy metabolism, the machinery for protein synthesis often increases with genome size (29). Knowledge about the function and number of genes required for maintaining basic cellular machinery, in general, would allow for composite identification of the essential gene set sufficient for sustaining cellular life.

Understanding the evolution of small, or "reduced genomes", may aid in defining genes essential for minimal cellular requirements and associated pathways. *M. genitalium* possesses the smallest genome known to date (16) and hence was selected as a model to study these phenomena. Comparisons between *M. genitalium* and other small genome pathogenic prokaryotes, such as *H. influenzae*, indicate a possible common source ancestry. Phylogenomic studies of the two genera show that separation likely occurred approximately 1.5 billion years ago. In addition, comparative genome analysis shows that both have approximately 240 genes in common (30). The larger *E. coli* genome also has a similar number of translational proteins. This indicates that strict functional constraints imply high degrees of sequence conservation (31). On the other hand, this level of sequence conservation was reduced in other functional categories, including cell envelope and cytoskeletal proteins. Genes in certain metabolic pathways have been relegated to the set of minimal genes required for a "modern-type" cell (16, 30). Some scientists further argue that *M. genitalium* may yet contain up to double the number of genes requisite for modern life, (22) based on the finding that some genes have host-parasite interaction function. The maintenance of a

Table 1. Statistics of *Mycoplasma* genome projects.

NO	Organism	N Seq	Genes	Seq
With Genome Projects				
1	*Mycoplasma gallisepticum R*	6	782	1,452
2	*Mycoplasma genitalium G-37*	106	524	1,823
3	*Mycoplasma hyopneumoniae 232*	2	728	1,382
4	*Mycoplasma mobile 163K*	2	668	1,268
5	*Mycoplasma mycoides subsp. mycoides SC str. PG1*	2	1,053	1,016
6	*Mycoplasma penetrans HF-2*	2	1,070	2,074
7	*Mycoplasma pneumoniae M129*	65	734	1,377
8	*Mycoplasma pulmonis UAB CTIP*	2	815	782
9	*Mycoplasma synoviae 53*	2	728	1,344
10	*Mycoplasma sp. 'bovine group 7'*	19	2	21/ 3
11	*Mycoplasma haematoparvum*		6,387	23,341
12	*Ureaplasma Sp. (urealyticum)*	256	3,7624	2,322
With GenBank Records				
1	*Candidatus Mycoplasma haemominutum*	28		
2	*Mycoplasma agalactiae*	119		68
3	*Mycoplasma alligatoris*	7		4
4	*Mycoplasma arginini*	26		11
5	*Mycoplasma bovigenitalium*	12		
6	*Mycoplasma buccale*	6		1
7	*Mycoplasma canis*	5		
8	*M. capricolum subsp. capricolum ATCC 27343*	385		196
9	*Mycoplasma capricolum subsp. Capripneumoniae*	87		28
10	*Mycoplasma conjunctivae*	8		4
11	*Mycoplasma faucium*	18		
12	*Mycoplasma felis*	16		
13	*Mycoplasma fermentans*	71		195
14	*Mycoplasma incognitos*	1		6
15	*Mycoplasma flocculare*	8		4
16	*Mycoplasma gallinarum*	6		3
17	*Mycoplasma gallisepticum*	399		659
18	*Mycoplasma haemocanis*	10		1
19	*Mycoplasma haemofelis*	451		1
20	*Mycoplasma hominis*	255		337
21	*Mycoplasma hyopneumoniae 7448*	2		1,326
22	*Mycoplasma hyopneumoniae J*	2		1,330
23	*Mycoplasma hyorhinis*	51		57
24	*Mycoplasma iowae*	12		16
25	*Mycoplasma mycoides subsp. Capri*	16		19
26	*Mycoplasma mycoides subsp. mycoides LC*	23		28
27	*Mycoplasma orale*	15		3
28	*Mycoplasma pirum*	8		15
29	*Mycoplasma primatum*	9		1
30	*Mycoplasma pullorum*	4		1
31	*Mycoplasma salivarium*	16		3
32	*Mycoplasma spermatophilum*	9		1
33	*Mycoplasma suis*	5		11
34	*Mycoplasma sp. 'ovine/caprine serogroup 11'*	11		

larger gene pool than the theoretically optimal one may be justified by accepting that there is a divergence between theoretical computational models and realistic evolution. Genomic computational analysis may only account for the final genome state in contradistinction to the intermediate path-dependent state processes of genomic reduction. Moreover, *Mycoplasma* genomes are likely to be still undergoing evolutionary reductions (15). Developing a teleological rationale has motivated scientists to develop “virtual cells,” or hypothetical cells that feature minimal gene sets. Here, it becomes possible to map genomic evolution by comparing related species of *Mycoplasma* (15, 31). Experimentally, a minimal genomic cell can be studied in *M. genitalium* by "transposon mutagenesis" (32). This

may allow for tentative classification into those genes that are essential and nonessential, subject to the defined conditions and environmental milieu (15, 32).

The loss of genes from the genetic set is termed reductive evolution. Such loss can confer a negative independent viability, such as in the case of the loss of a cell wall in *Mycoplasma.* The presumed corresponding adaptation to this loss appears to be the membranous incorporation of large quantities of exogenous cholesterol (2). *M. pneumonia* retains 54 of the 105 genes associated with cell wall construction, which may suggest that reductive evolution is ongoing. Ongoing proteomics analysis on large numbers of *M. pneumoniae* trans membrane domains can lend more insight into the suggested adaptive process (31). It is still premature to characterize these genes as relics of a pre-adaptive genome. The loss of certain gene sets, such as those in amino acid biosynthesis and lipid metabolism, also implicate a transfer of dependence for nutrient acquisition to host organisms (15, 31). Whether genome reduction occurred from the progression to parasitism or as a result of transition to the parasitic modality has yet to be determined. The relation between the environment and how environmental factors and mode of living would shape genome evolution is still a hot research subject. Thus, the entry of one organism into the body of another organism, where nutrients are available, can precipitate an alteration in genetic expression. This is characterized as reductive evolution. *Mycoplasmas* seem to have evolved in this reductive direction, from a large genome with many genes to a smaller one with reduced genes. This supports the theory that *Mycoplasma* genomes were altered in response to the environment.

3.3. Intra- and inter-genomic rearrangements: an added advantage

Homologous recombination may have fostered rapid evolution of the *Mycoplasma* genome. This event involves sequential exchange and rearrangement between portions of DNA molecules and chromosomes. Hot spots for recombination have been identified in genomes that include *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, *Ureaplasma urealyticum* and *Mycoplasma pulmonis*. For example, homologous recombination is seen in genetic loci coding for adhesion proteins by *M. pneumoniae* and *M. genitalium* (25). This event occurs in distant repeats within the genetic coding regions. These repeats correlate well recombinations associated with surface antigenic variation. Such hot spots are akin to on/off switches employed by pathogenic bacteria (21). High-frequency phase (2) and antigenic variation, which result in surface diversity, are believed to further *Mycoplasma* virulence and pathogenesis by enhancing host evasion (2, 25). In addition to carrying out intra-genomic rearrangement, *Mycoplasma* genomes can undergo horizontal gene transfer. Several studies indicate that horizontal gene transfer plays a major role in the transfer event of coding DNA sequences exchanged between *Mycoplasma* species and parasites (33, 51). This feature may establish a framework for molecular evolution and host-parasite relationships (52, 4). This method of gene transfer was observed in both *M. synoviae* and *M. gallisepticum* avian pathogens (33). Divergent *Mycoplasma* strains may demonstrate increased predilection to recombine due to missing DNA mismatch repair system proteins such as MutSLH (25), which serves as a recombination barrier between divergent sequences (21). This may explain the increased recombination frequency of close repeats, which was noted in the *vsa* locus of *M. pulmonis* (25).

3.4. G+C content and oligonucleotide frequencies: species-specific characteristics

Genomic signatures are short oligonucleotide sequence repeats found throughout prokaryotic genomes. The nature of the genomic signature is often species-specific (34). Differences between species from various genera are often much larger than those within a genus. One salient exception is observed when analyzing the genomic signatures of *M. genitalium*, *M. pneumoniae* and *M. pulmonis,* where signature variations between species are comparable to those of different genera (35). Also, there is variation in alternative synonymous codons among species, and within a single genome. Bacteria display diverse genomic G+C content and this directly impacts codon translation. Codon frequencies may harbor bias from factors that include mutations, selective pressure and genetic drift (36). One comparative study on *Mycoplasma* and prokaryote genomes hypothesized that codon usage and amino acid composition have a direct relationship when G+C content is expressed in excess of a twofold difference. For example, the four amino acids glycine, alanine, proline, and arginine are often highly expressed with G+C-rich codons. Similarly, the amino acids isoleucine, phenylalanine, tyrosine, methionine, and aspartic acid are associated with A+T-rich genome sequences (29). *Mycoplasma* genomes tend to possess low G+C content, expressed at a level of 24% to 33 % (Table 1), and use the genetic codon (T) UGA for tryptophan (38). Intergenic regions are known to exhibit higher A +T content than the coding regions (39-40).

One unique feature displayed in several *Mycoplasma* genomes is the methylation of the genomic cytosine residues. *M. pulmonis* contains simple short repeats (SSR) for putative CpG-specific methyltransferases (41). In the past, complete methylation was deemed to be characteristic of eukaryotes. CpG -specific methyltransferases may aid in gene expression regulation. In eukaryotes, CpG regions are mutational hot spots, which are often absent in regions of low numbers of gene expression (42). A number of *Mycoplasma* genomes have been reported to show high CpG avoidance (25). Collectively, this may partially explain the low G+C content of their genomes.

4. *MYCOPLASMA* AS A MODEL FOR GENOMICS RESEARCH

The rationale behind the use of *Mycoplasma* models to define, in molecular terms, the entire machinery of a living cell (10, 21) stems from its parasitic efficiency and genomic characteristics, which are discussed in this section.

4.1. Minimal genome with high parasitic efficiency and taxon-related genomic heterogeneities

Despite their diminutive genome size, pathogenic *Mycoplasmas* generate DNA rearrangements and participate in genetic variation to maximize their genetic coding potential (2). While the human pathogen *M. genitalium* is merely 580 Kb long and contains 470 predicted coding sequences, it is still effective in engaging in a parasitic mode of life and is functional in attachment and defense (22). Systemic modifications resulting in "reductive evolution" (43) may explain why the *M. genitalium* genome carries only one gene associated with amino acid biosynthesis, and few genes for synthesis of vitamin and nucleic acid precursors. This distinct lack of genetic upregulation of the fatty acid biosynthetic pathway correspondingly makes the organism dependent on fatty acids supplied by the host. This results in a changed membrane fluidity structure and permits macromolecular entry. Similar adaptations are observed in the genetic regulation of ATP metabolism required for parasitic modes of life (21).

M. genitalium, *M. pneumoniae* and *U. urealyticum* are related pathogens that invade human mucosa. These pathogens have helped to define the essential functions of a self-replicating minimal cell, and more specifically characterize the definition of *Mycoplasma*. Studies of energy generation and ATP synthesis in *U. urealyticum* reveal six closely related iron transporters. These transporters appear to have arisen by gene duplication, suggesting that it harnesses a unique respiration system in comparison to other small genome bacteria (44). Comparison between *U. urealiticum* with *M. pulmonis* indicated that *U. urealiticum* has only one set of rRNA genes and 29 tRNA genes. A number of genes thought to be essential for a viable self-sufficient minimal cell are absent from *M. pulmonis* (45). Additional genetic diversity can be found in the high degree of variation in the RNase P RNA structure, which *in vitro* acts as an efficient ribozyme (46). Further genomic analysis indicates that in many natural isolates of *Mycoplasmas*, there is substantial genetic heterogeneity (47, 48, 49), which is necessary for allowing parasitic infection and/or evading host immune response functions (50).

4.2. Codon patterns and usage

The presentation of high G+C content between genes may impact intragenic codon expression. This is based on the assumption that stop codons that are A+T-rich occur less frequently in G+C-rich exons. It is also predicted that the effects of mutation has a direct impact on the stop codon frequency and distribution, though applicability extends only to the last coding exon (53). Correlation patterns between genome characteristics and codon usage patterns have been studied in both vertebrate (54, 55) and invertebrate species (56). Prokaryote gene expression level correlation with codon usage (57) is particularly important for differentiating how essential a gene may be for determining viability. Codon usage can represent evolutionary progression of the development of an organism. *Spiroplasmas* use T- and A-terminated codons preferentially in sequences reflecting low G+C content (32 mol%). Certain viral promoters sequences, such as that embedded into the SpV4 genome, have high homology to eubacterial sequences, indicating that *Spiroplasmas* are derived from bacteria (58, 59). Interestingly, *M. capricolum* codon distribution differs from that of *E. coli* in its heavy bias toward A and U (T), with approximately 91% of codons containing A or U in the third slot. UGA, a universal code for "stop", is found in higher frequency than UGG to code for tryptophan (60). *M. genitalium* demonstrates suppressed CpG coding in comparison to *M. pneumoniae*. This is likely a result of global considerations, such as the maintenance of DNA stability. Such considerations can explain biased codon usage (61). While compositional pressure may be a singular factor in shaping codon usage in unicellular species displaying extremely biased genomic compositions, further investigation into *Mycoplasma* genome sequence variation might offer further clues on codon usage profiles and the basis for interspecies divergence.

5. DEVELOPING APPLICATIONS OF THE *MYCOPLASMA* RESEARCH

A minimal, self-replicating and fully sequenced genome makes *Mycoplasma* a very good candidate for development in genetic engineering projects (16). *Mycoplasma* genome structures can be modeled for understanding more complex organisms (62). In addition, a composite minimal gene set can be used for construction of the so-called "synthetic" microorganism (32). Along these lines, the Craig Venter Institute, USA, is endeavoring to construct such a "synthetic" organism by sequentially knocking out gene after gene to identify essential gene synthesis arrays in the set (63). The rationale behind a "minimal" organism generating performing targeted functions such as bioremediation, with maximal energetic efficiency. However, this type of research should occur with proper oversight so that "alien" species with unpredictable ecological and biological impact do not become threats (64). Unfortunately, development of genetically engineered pathogenic bacterial and viral strains could develop into biological weapons requiring constant public vigilance (62). At the same time, genetic engineering of these bacteria can also be beneficial in mapping evolutionary change, cellular functions and host-pathogen interactions (16). Targeted treatments for infectious diseases, therapeutics, and vaccines can be developed using microbial genetic manipulation (65). *M. genitalium* has been used for *in vitro* gene insertion studies (66). One interesting application for *M. genitalium* has been NASA's use of this species as a benchmark for Mars Rover detection of small microbes in space (67).

6. BIOINFORMATICS RESOURCES FOR *MYCOPLASMA*

Currently there are many bioinformatics resources on *Mycoplasma* genomics (Table 2). One central resource is the *Mycoplasma* Bioinformatics Resource Center at the University of Alabama in Birmingham, which contains the complete *M. hypopneumoniae* genome. This center also provides analysis and annotation of

Table 2. Major informatics resources for *Mycoplasmas* genomics

DB Name	DB type and Service	GEO location	URL	Major Contents
A. General Utility Databases				
Mycoplasma Bioinformatics Resource Center	Sequence DB; genes and complete genomes analysis	USA	http://mycoplasma.genome.uab.edu	M. hypopneumoniae complete genome
Genoscope	Sequence DB of various organisms	France	http://www.genoscope.cns.fr/externe/English/corps_anglais.html	M. agalactiae, M. hominis, M. mycoides, M. pulmonis complete genomes
TIGR	Sequence DB	Maryland, USA	www.tigr.org	M. genitalium sequence (& >50 other organisms)
The M. pneumoniae Genome Project	Sequence DB	Germany	http://www.zmbh.uni-heidelberg.de/M_pneumoniae/genome/	M. pneumoniae - sequencing and information
Entrez Gene	Sequence DB; links to BLAST, GenePlot	NCBI, USA	http://www.ncbi.nlm.nih.gov/entrez/	M. genitalium, M. hypopneumoniae, M. synoviae, M. mycoides, M. pneumoniae, M. pulmonis, M. gallisepticum, M. mobile, M. penetrans complete genomes
KEGG Orthology	Gene DB	Japan	http://www.genome.ad.jp/dbget-bin/get_htext?M.genitalium.kegg	M. genitalium gene descriptions
TransportDB	Membrane transporter protein DB	TIGR, Maryland, USA	http://www.membranetransport.org/	cytoplasmic membrane transport protein DB of organisms whose complete genomes have been sequenced
Los Alamos National Lab. Bioscience Division	STD Sequence DB	California, USA	http://www.stdgen.lanl.gov/	M. genitalium genome and other information
GeneQuiz	Protein sequence homology levels	EMBL, UK	http://jura.ebi.ac.uk:8765/ext-genequiz/	M. genitalium, M. pneumoniae, M. pulmonis description and protein sequence homology levels using a functional information clock
B. Article Databases				
CFS Research	Research publications on mycoplasma	USA	http://www.cfsresearch.org/mycoplasma/	General information and articles on mycoplasma

Mycoplasma genes, and analysis software for complete genomes. Genoscope from the French National Sequencing Center provides a detailed description of the *Mycoplasma* species listed in Table 1. It also provides links to many *Mycoplasma*-related websites and describes the different genomes, but does not supply direct access to the sequences. The Institute for Genomic Research (TIGR) is one of the most comprehensive *Mycoplasma* resources. It provides entire genome sequences for ten species of the Mycoplasmataceae family. The TIGR Comprehensive Microbial Resource page lists functions of particular genes (i.e., chemotaxis) and allows the user to compare functions among different organisms. The user has the capability to modify the sequence search for specific parameters, such as 'DNA repeats' or 'Transfer RNA'. The site provides analyses such as G+C plots, 2-D gels, and G+C comparisons. This site also contains MUMmer: The Whole Genome Alignment Tool. The *M. pneumoniae* Genome Project of the Center for Molecular Biology - Heidelberg describes the *M. pneumoniae* sequencing project, from introduction to methods to results. It contains genome statistics, many detailed graphical presentations of the genome mapping, and various metabolic pathways of the organism. Detailed comparisons between *M. pneumoniae* and *M. genitalium* can be found on this site. Entrez provides complete genomes for many of the members of the Mycoplasmataceae family (Table 1).

7. *MYCOPLASMA* GENOMICS: FUTURE PERSPECTIVES

Recent *Mycoplasma* genome projects have brought us closer to deciphering, in molecular terms, the machinery of a self-replicating cell and minimal genomics requirements for life. These projects have contributed to our understanding of *Mycoplasmas* molecular biology and their evolutionary history, while lending insight into studying adaptive mechanisms in genomic redundancy and structural plasticity. There appears to be valid genetic support for the theory that *Mollicutes* evolved from a branch of gram-positive bacteria by a process known as "reductive or degenerative evolution". During this process, *Mycoplasmas* lost considerable portions of their chromosome ancestry, yet retained genes essential for special modes of life. As such, *Mycoplasma* genome carries a high percentage of conserved genes, facilitating gene identification and providing a unique example of evolutionary success. Their wide distribution in nature and their highly conserved genome seem to support this presumption. Finally, the rising number of fully sequenced *Mycoplasma* genomes and the reengineering potentials of these genomes offer opportunities for further genetic manipulations and biotechnological exploitations.

8. ACKNOWLEDGEMENT

We give special thanks to T.J. Scalise for her excellent assistance in the preparation of this manuscript.

9. REFERENCES

1. Krause D. C. & M. F. Balish: Cellular engineering in a minimal microbe: structure and assembly of the terminal organelle of *Mycoplasma pneumoniae*. *Mol Microbiol* 51, 917-924 (2004)

2. Baseman J. & J. Tully: *Mycoplasmas*: Sophisticated, reemerging, and burdened by their notoriety, *Emerg Infect Dis* 3, 21-32 (1997)

3. Freundt E. A., R. F. Whitcomb, M. F. Barile *et al.* Proposal of Minimal Standards for Descriptions of New Species of the Class Mollicutes. *Int J Syst Bacteriol* (1979)

29: 172-80, International Committee on Systematic Bacteriology - Subcommittee on the Taxonomy of Mollicutes. Revised minimum standards for description of new species of the class Mollicutes (division Tenericutes). *Int J Syst Bacteriol* 45: 605-612 (1995)

4. Fadiel A., S. Lithwick & F. Naftolin The influence of environmental adaptation on bacterial genome structure. *Lett Appl Microbiol* 40(1), 12-18 (2005)

5. Tully J. G.: Current status of the mollicute flora of humans. *Clin Infect Dis* 17, 2-9 (1993)

6. Taylor L. N. D.: *Mycoplasmas* - Stealth Pathogens. *Amer Soc Microbiol*. 182, 5238-5250 (2001) (http://www.rain-tree.com/myco.htm)

7. Simecka J. W., J. K. Davis, M. K. Davidson, S. E. Ross, C. T. Stadtlander & G. H. Cassell: *Mycoplasma* diseases of animals. In *Mycoplasmas*: Molecular Biology and Pathogenesis. Ed: J. Maniloff. American Society for Microbiology, Washington, DC. 391-415 (1992)

8. Bautsch W.: Rapid physical mapping of the *Mycoplasma mobile* genome by two-dimensional field inversion gel electrophoresis techniques. *Nucleic Acids Res* 16, 11461-11467 (1988)

9. Ladefoged S. A. & G. Christiansen: Physical and genetic mapping of the genomes of five *Mycoplasma hominis* strains by pulsed-field gel electrophoresis. *J Bacteriol* 174, 2199-2207 (1992)

10. Morowitz H. J.: The completeness of molecular biology. *Isr J Med Sci* 20(9), 750–753 (1984)

11. Ruland K., R. Wenzel & R. Herrmann.: Analysis of three different repeated DNA elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number and distribution on the genome. *Nucleic Acids Res* 18, 6311-6317 (1990)

12. Ruland K., R. Himmelreich & R. Herrmann: Sequence divergence in the ORF 6 gene of *Mycoplasma pneumoniae*. *J Bacteriol* 176, 5202─5209 (1994)

13. Wenzel R., & R. Herrmann: Cloning of the complete *Mycoplasma pneumoniae* genome. *Nucleic Acids Res* 117, 7029–7043 (1989)

14. Wenzel R., E. Pirkl, & R. Herrmann: Construction of an *Eco*RI restriction map of *Mycoplasma pneumoniae* and localization of selected genes. *J Bacteriol* 174, 7289–7296 (1992)

15. Himmelreich R., H. Hilbert, H. Plagens, E. Pirkl, B. C. Li, & R. Herrmann: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* 24, 4420-4449 (1996)

16. Fraser C.M., J.D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann & C. J. Bult: The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403 (1995)

17. Broad Institute: *Mycoplasma* mobile assembly (2005). http://www.broad.mit.edu/

18. Rocha E. P. C., A. Viari & A. Danchin: Oligonucleotide bias in *Bacillus subtilis*: general trends and taxonomic comparisons. *Nucleic Acids Res* 26, 2971–2980 (1998)

19. Bové J. M.: *Spiroplasmas*: infectious agents of plants, arthropods and vertebrates. *Wien Klin Wochenschr* 109, 604-612 (1997)

20. Carle P., F. Laigret, J. G. Tully & J. M. Bové: Heterogeneity of genome sizes within the genus *Spiroplasma. Int J Syst Bacteriol* 45, 178-181(1995)

21. Razin S., D. Yogev & Y. Naot: Molecular Biology and Pathogenicity of *Mycoplasmas*. *Microbiol Mol Biol Rev* 62, 1094-1156 (1998)

22. Razin S.: The minimal cellular genome of *Mycoplasma*. *Indian J Biochem Biophys*; 34, 124-30 (1997)

23. Rogers M. J., J. Simmons, R. T. Walker RT *et al.* Construction of the *Mycoplasma* evolutionary tree from 5S rRNA sequence data. *Proc Natl Acad Sci* (USA) 82, 1160-1164 (1985)

24. Laigret F., F. Ye., J. C. Whitley, C. Citti, L. R. Finch, P. Carle, J. Renaudin, & Bové, J. M: A physical and genetic map of the *Spiroplasma citri* genome. *Nucleic Acids Res* 20, 1559–1565 (1992)

25. Rocha E. P. C. & A. Blandnchard: Genomic repeats, genome plasticity and the dynamics of *Mycoplasma* evolution. *Nucleic Acids Res* 30, 2031–2042 (2002)

26. Kenri T., R. Taniguchi, Y. Sasaki, N. Okazaki, M. Narita, K. Izumikawa, M. Umetsu, & T. Sasaki: Identification of a new variable sequence in the P1 cytadhesin gene of *Mycoplasma pneumoniae*: evidence for the generation of antigenic variation by DNA recombination between repetitive sequences. *Infect Immun* 67, 4557-4562 (1999)

27. Shen X., J. Gumlak, Hui Yu, T. French, N. Zou & K. Dybvig: Gene rearrangements in the *vsa* locus of *Mycoplasma pulmonis*. *J Bacteriol* 182, 2900-2908 (2000)

28. Koonin E. V., R. L. Tatusov & K. E. Rudd: Protein sequence comparison at genome scale. *Methods Enzymol* 266, 295-322 (1996)

29. Fraser C. M. , J. Eisen., R. D. Fleischmann, K. A. Ketchum & S. Peterson: Comparative Genomics and understanding of microbial biology. *Emerg Infect Dis* 6, 505-512 (2000)

30. Mushegian A. R. & E. V. Koonin: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci* 93, 10268-10273 (1996)

31. Himmelreich R., H. Plagens, H. Hilbert, B. Reiner & and R. Herrmann: Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res* 25, 701-712 (1997)

32. Hutchison C. A., S. N. Peterson, S. R. Gill, R. T Cline, O.White, C.M. Fraser, H.O. Smith & J.C. Venter.: Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286, 2165-2169 (1999)

33. Vasconcelos A. T., H.B. Ferreira, C.V. Bizarro, S.L. Bonatto, M.O. Carvalho, P.M. Pinto, D.F. Almeida, L.G. Almeida, R. *et al.*, Swine and poultry pathogens: the complete genome sequences of two strains of *Mycoplasma hyopneumoniae* and a strain of *Mycoplasma synoviae*. J Bacteriol 187, 5568-5577 (2005)

34. Dufraigne C., B. Fertil, S. Lespinats, A. Giron & P. Deschavanne: Detection and characterization of horizontal transfers in prokaryotes using genomic signature. *Nucleic Acids Res* 33 Pp. e6 (2005) (http://nar.oxfordjournals.org/cgi/content/full/33/1/e6)

35. Sandberg R., Carl-Ivar Branden, I. Ernberg, & J. Coster: Quantifying the species-specificity in genomic signatures, synonymous codon choice, amino acid usage and GC content. *Gene*. 311, 35-42 (2003)

36. Sharp P. M., E. Bailes, R. J. Grocock, J. F. Peden & R. E. Sockett: Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acid Res* 33, 1141-1153 (2005)

37. Oudega B., G. Koningstein, L. Rodrigues, M. Ramon de Sales, H. Hilbert, A. Dusterhoft, T. M. Pohl, T. Weitzenegger: Analysis of the *Bacillus subtilis* genome: cloning and nucleotide sequence of a 62 kb region between 275 degrees (*rrn*B) and 284 degrees (*pai*). *Microbiology* 143, 2769-2774 (1997)

38. Dybvig K. & L. L. Voelker: Molecular biology of *Mycoplasmas*. *Ann Rev Microbiol* 50, 25-57 (1996)

39. Muto A. & S. Osawa The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci* 84, 166-169 (1987)

40. Yogev D., R. Rosengarten, R. Watson-McKown, & K. S. Wise: Molecular basis of *Mycoplasma* surface antigenic variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J* 10, 4069-4079 (1991)

41. Renbaum P. & A. Razin: Footprint analysis of M.Sssl and M.Hhal methyltransferases reveals extensive interactions with the substrate DNA backbone. *J Mol Biol* 248, 19–26 (1995)

42. Gardiner-Garden M. & M. Frommer: CpG islands in vertebrate genomes. *J Mol Biol* 196, 261–282 (1987)

43. Oshima K., S. Kakizawa, H. Nishigawa, H. Y. Jung, W. Wei, S. Suzuki, R. Arashida, D. Nakata, S. Miyata, M. Ugaki & S. Namba: Reductive evolution suggested from the complete genome sequence of a plant-pathogenic *Phytoplasma*. *Nat Genet* 36, 27-29 (2004)

44. Glass J. I., E. J. Lefkowitz, J. S. Glass, C. R. Heiner, E. Y. Chen, G. H. Cassell: The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* 407, 757-762 (2000)

45. Chambaud I., R. Heilig, S. Ferris, V. Barbe, D. Samson, F. Galisson, I. Moszer, K. Dybvig, H. Wroblewski, A. Viari, E. P. Rocha & A. Blanchard: The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res* 29, 2145-2153 (2001)

46. Wagner M., C. Fingerhut, H. J. Gross & A. Schon: The first phytoplasma RNase P RNA provides new insights into the sequence requirements of this ribozyme. *Nucleic Acids Res* 29, 2661-2665 (2001)

47. Bai X., T. Fazzolari & S. A. Hogenhout: Identification and characterization of *traE* genes of *Spiroplasma kunkelii*. *Gene* 336, 81-91 (2004)

48. Wellehan J. F., M. Calsamiglia, D. H. Ley, M. S. Zens, A. Amonsin & V. Kapur: Mycoplasmosis in captive crows and robins from Minnesota. *J Wild Dis* 37, 547-555 (2001)

49. Nicholas R. A., L. A. Khan, B. Houshaymi, R. J. Miles, R. D. Ayling, H. Hotzel & K. Sachse: Close genetic and phenotypic relatedness between *Mycoplasma* ovine/caprine serogroup 11 and *Mycoplasma bovigenitalium*. *Syst Appl Microbiol* 25, 396-402 (2002)

50. Meseguer M. A., A. Alvare, M. T. Rejas, C. Sanchez, J. C. Perez-Diaz & F. Baquero: *Mycoplasma pneumoniae*: a reduced-genome intracellular bacterial pathogen. *Infect Genet Evol* 3, 47-55 (2003)

51. Thomas A., A. Linden, J. Mainil, D. F. Bischof, J. Frey & E. M. Vilei: *Mycoplasma bovis* shares insertion sequences with *Mycoplasma agalactiae* and *Mycoplasma mycoides* subsp. *mycoides* SC: Evolutionary and developmental aspects. *FEMS Microbiol Lett* 245, 249-255 (2005)

52. Pillai S. R., H. L. Mays, D. H. Jr Ley, P. Luttrell, V. S. Panangala, K. L. Farmer & S. R. Roberts: Molecular variability of house finch *Mycoplasma gallisepticum* isolates as revealed by sequencing and restriction fragment length polymorphism analysis of the *pvpA* gene. *Avian Dis* 47, 640-648 (2003)

53. Xia X., Z. Xie & W. H. Li: Effects of GC content and mutational pressure on the lengths of exons and coding sequences. *J Mol Evol* 56, 362-370 (2003)

54. Bernardi G., B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier-Rotival & F. Rodier: The mosaic genome of warm-blooded vertebrates. *Science* 228, 953-958 (1985)

55. Aota S. & T. Ikemura: Diversity in G + C content at the third position of codons in vertebrate genesand its cause. *Nucleic Acids Res* 14, 6345-6355 (1986). *Erratum* in: *Nucleic Acids Res* 14, 8702 (1986)

56. Musto H., H. Romero & H. Rodríguez-Maseda: Heterogeneity in Codon Usage in the Flatworm *Schistosoma mansoni*. *J Mol Evol* 46, 159-167 (1998)

57. Karlin S. & J. Mrázek: Predicted Highly Expressed Genes of Diverse Prokaryotic Genomes. *J Bacteriol* 182, 5238-5250 (2000)

58. Renaudin J., M. C. Pascarel & J. M. Bové: Cloning and sequencing of the genome of *Spiroplasma* virus 4. *Isr J Med Sci* 23, 427–428 (1987)

59. Renaudin J., M. C. Pascarel & J. M. Bové: *Spiroplasma* virus 4: nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization. *J Bacteriol* 169, 4950–4961 (1987)

60. Ohkubo S., A. Muto,Y. Kawauchi, F.Yamao & S.Osawa: The ribosomal gene cluster of *Mycoplasma capricolum Mol Genet genom* 210, 314-322 (1987)

61. Goto M., T. Washio & M. Tomita: Causal Analysis of CpG Suppression in the *Mycoplasma* Genome. *Microb Comp Geno* 5, 51-58 (2000)

62. Cho M. K., D. Magnus, A. L. Caplan & D. McGee: Ethical considerations in synthesizing a minimal genome. *Science* 286, 2087-2090 (1999)

63. The Craig Venter Institute, http://www.venterinstitute.org/ and http://www.syntheticgenomics.com/about.htm

64. Losey J., L. Rayor & M. Carter: Transgenic pollen harms monarch larvae *Nature* 399, 214-222 (1999)

65. Fraser C. M. & R. Rappuoli: Application of microbial genomic science to advanced therapeutics. *Ann Rev Med* 56, 459-474 (2005)

66. Pich O. Q., R. Burgos, R. Planell, E. Querol & J. Piñol: Comparative analysis of antibiotic resistance gene markers in *Mycoplasma genitalium*: application to studies of the minimal gene complement. *Microbiology* 152, 519-527 (2006)

67. P. A. Hanle: Microbes for the future. *Your world* 14, 1-16 (2005) **(**http://www.biotechinstitute.org/yourworld.html)

Key Words: *Mycoplasma*, Genomics, Minimal life requirements, Genome reduction, Informatics, Review

Send correspondence to: Dr A. Fadiel, Department of Obstetrics and Gynecology, Yale University, School of Medicine, New Haven, CT., 06511, USA, Tel: 203-737-1218, E-mail: afadiel@Yale.edu

http://www.bioscience.org/current/vol12.htm