UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS: BIOQUÍMICA

ANÁLISE GENÔMICA E TRANSCRICIONAL COMPARATIVA DE *Mycoplasma hyopneumoniae, Mycoplasma flocculare* e *Mycoplasma hyorhinis*

Franciele Maboni Siqueira

Porto Alegre

2013

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS: BIOQUÍMICA

ANÁLISE GENÔMICA E TRANSCRICIONAL COMPARATIVA DE *Mycoplasma hyopneumoniae, Mycoplasma flocculare* e *Mycoplasma hyorhinis*

Franciele Maboni Siqueira

Orientador: Prof. Dr. Arnaldo Zaha

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciências Biológicas-Bioquímica da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Doutor.

Porto Alegre

2013

“O começo de todas as ciências é o espanto das coisas serem o que são”

“A dúvida é o princípio da sabedoria”

Aristóteles

AGRADECIMENTOS

No ano de 2002, comecei a traçar meu caminho profissional. Já em 2003 não havia mais dúvidas que seria a carreira científica que iria seguir.

O principal motivo da decisão:

PNAS

The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability

Brazilian National Genome Project Consortium*

Proc Natl Acad Sci U S A. 2003 September 30; 100(20): 11660–11665.

Pouco tempo depois...

JOURNAL OF BACTERIOLOGY, Aug. 2005, p. 5568–5577 Vol. 187, No. 16
0021-9193/05/$08.00+0 doi:10.1128/JB.187.16.5568–5577.2005

Swine and Poultry Pathogens: the Complete Genome Sequences of Two Strains of *Mycoplasma hyopneumoniae* and a Strain of *Mycoplasma synoviae*†

Vasconcelos AT, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, Assunção EN, Azevedo VA, Bogo MR, Brigido MM, Brocchi M, Burity HA, Camargo AA, Camargo SS, Carepo MS, Carraro DM, de Mattos Cascardo JC, Castro LA, Cavalcanti G, Chemale G, Collevatti RG, Cunha CW, Dallagiovanna B, Dambrós BP, Dellagostin OA, Falcão C, Fantinatti-Garboggini F, Felipe MS, Fiorentin L, Franco GR, Freitas NS, Frías D, Grangeiro TB, Grisard EC, Guimarães CT, Hungria M, Jardim SN, Krieger MA, Laurino JP, Lima LF, Lopes MI, Loreto EL, Madeira HM, Manfio GP, Maranhão AQ, Martinkovics CT, Medeiros SR, Moreira MA, Neiva M, Ramalho-Neto CE, Nicolás MF, Oliveira SC, Paixão RF, Pedrosa FO, Pena SD, Pereira M, Pereira-Ferrari L, Piffer I, Pinto LS, Potrich DP, Salim AC, Santos FR, Schmitt R, Schneider MP, Schrank A, Schrank IS, Schuck AF, Seuanez HN, Silva DW, Silva R, Silva SC, Soares CM, Souza KR, Souza RC, Staats CC, Steffens MB, Teixeira SM, Urmenyi TP, Vainstein MH, Zuccherato LW, Simpson AJ, Zaha A.

Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 9500, Prédio 43421, Porto Alegre, RS, Brazil.

Agradeço a todas as pessoas e amigos, além das destacadas acima, que de um modo ou de outro contribuíram ou contribuem para a efetivação dos meus objetivos.

Em especial, ao Dr. Arnaldo Zaha, Dra. Irene S. Schrank e Dra. Ana Tereza Vasconcelos – mentores que sempre depositam em mim confiança absoluta.

Aos professores Augusto Schrank, Marilene H. Vainstein, Charley C. Staats e Henrique B. Ferreira, não apenas pela disposição sempre imediata em discutir resultados e projetos, mas acima de tudo, pelo acolhimento, amizade e carinho.

E por fim, meus agradecimentos ao meu marido, Henrique; um engenheiro civil quase biólogo molecular. Agradeço pelo carinho, por sua participação ativa em minha vida, e pelo apoio incessante.

APRESENTAÇÃO

Esta tese está organizada nos tópicos: *Resumo*, *Abstract*, *Introdução, Objetivos*, *Capítulos* (1 e 2- referente a artigos científicos), *Discussão*, *Conclusões* e *Perspectivas*.

Os resultados são apresentados na forma de artigos científicos (Organizados em diferentes capítulos). As sessões *Materiais e Métodos*, *Resultados*, *Discussão* e *Referências Bibliográficas* encontram-se nos próprios artigos.

O capítulo 2 compreende um artigo científico em preparação. As informações apresentadas no mesmo são resultados parciais, com dados preliminares. As tabelas suplementares, citadas no texto deste capítulo, estão disponíveis no item *Anexos*.

Os itens *Introdução*, *Discussão* e *Conclusões* encontradas nesta tese apresentam interpretações e comentários gerais sobre os resultados contidos nos artigos científicos do presente trabalho. O item *Referências Bibliográficas* refere-se apenas às citações encontradas nos itens supracitados.

As informações técnicas mais detalhadas sobre cada metodologia empregada poderão ser encontradas nos artigos científicos correspondentes.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

Mycoplasma hyopneumoniae, *Mycoplasma flocculare* e *Mycoplasma hyorhinis* são capazes de aderir e colonizar o trato respiratório de suínos. Enquanto a presença de *M. flocculare* é considerada assintomática, *M. hyopneumoniae* e *M. hyorhynis* são relacionados ao desenvolvimento de patologias. *M. hyopneumoniae* é o agente etiológico da pneumonia enzoótica suína e *M. hyorhynis* além dos pulmões pode atingir outros sítios e hospedeiros, estando relacionado a artrites, poliserosites e desenvolvimento de vários tipos de câncer em humanos. Apesar dos avanços tecnológicos na área de genômica, raros são os dados quanto ao papel de *M. flocculare* no trato respiratório suíno. Além do mais, informações relativas à transcrição gênica nessas espécies são escassas, apesar da importância desses microrganismos. Neste estudo são apresentados os dados da sequência do genoma de uma linhagem de *M. flocculare*, bem como do genoma de um novo isolado de *M. hyopneumoniae.* Com estas novas sequências foram realizadas análises de genômica comparativa visando a identificação de características que pudessem explicar os diferentes comportamentos quanto à patogenicidade dessas espécies. Além disso, a análise global dos transcritomas de cada uma das espécies foi realizada e o perfil transcricional entre *M. hyopneumoniae, M. flocculare* e *M. hyorhynis* foi analisado comparativamente objetivando identificar características peculiares para cada um dos mapas transcricionais, além de compreender a coordenação do modo de transcrição gênica em *Mycoplasma*. De um modo geral, as três espécies de *Mycoplasma* que habitam o trato respiratório suíno possuem grandes semelhanças na composição gênica, assim como na abundância de transcritos. A análise do repertório transcricional, mostra que os genomas são transcritos quase que em sua totalidade, incluindo as regiões intergênicas, nas três espécies. *M. hyopneumoniae* e *M. flocculare* apresentam conteúdo gênico e perfil transcricional muito semelhantes. Uma importante diferença encontrada entre estas duas espécies refere-se à presença exclusiva de genes e transcritos de adesinas específicas. *M. hyorhynis* possui genes e transcritos exclusivos, os quais sabidamente estão relacionados à sua capacidade mutacional, de invasividade e infecção de diferentes sítios. Por fim, a análise comparativa dos genomas, e a obtenção dos mapas transcricionais para *M. hyopneumoniae, M. flocculare* e *M. hyorhynis*, foram abordagens que resultaram em um grande número de informações, as quais são importantes para embasamento de futuros estudos de caracterização dos mecanismos moleculares, como os eventos de regulação da transcrição gênica, no gênero *Mycoplasma*.

Palavras-chave: Mycoplasma; Genômica comparativa; Transcritômica comparativa; Trato respiratório suíno.

ABSTRACT

Mycoplasma hyopneumoniae, *Mycoplasma hyorhinis* and *Mycoplasma flocculare* are able to adhere and to colonize the swine respiratory tract. While *M. flocculare* presence is virtually assymptomatic, *M. hyopneumoniae* and *M. hyorhynis* infections may cause respiratory disease. *M. hyopneumoniae* is the causative agent of swine enzootic pneumonia and *M. hyorhynis* may affect the lungs and other sites in a diversity of hosts and has been related to arthritis, poliserosites and to the development of several types of human cancer. Despite genomics technological advances, there are very few data about the possible role of *M. flocculare* in the swine respiratory tract. Moreover, little information about gene transcription is available in these species, despite the importance of these microorganisms. In this work the genome sequences of *M. flocculare* and a new isolate of *M. hyopneumoniae* are presented. A comparative genomic analyzes was performed to identify possible characteristics that may help to explain the different behaviors of these species in the swine respiratory tracts. Furthermore, a transcriptome map of each species was performed and a comparative transcriptional profile analysis between *M. hyopneumoniae*, *M. flocculare* and *M. hyorhynis* was undertaken to identify the exclusive features for each of the transcriptional maps, in addition to understanding the coordination mode of gene transcription in *Mycoplasma*. In general, the three *Mycoplasma* species that inhabit the swine respiratory tract have a similar gene composition as well as the abundance of transcripts. The transcriptome maps showed that most of the predicted genes are transcribed from these *Mycoplasma* genomes, as well as some intergenic regions. *M. hyopneumoniae* and *M. flocculare* present very similar gene content and transcriptional profile. However, an important difference between these two species is related to the exclusive presence of genes and transcripts of some specific adhesins. *M. hyorhynis* presents exclusive genes and transcripts that have been related to its invasiveness, mutation rate and infection of different sites. Finally, the comparative analysis of the genomes and transcriptional maps between *M. hyopneumoniae*, *M. flocculare* and *M. hyorhynis* have resulted in a large amount of information, which are important for future studies of the molecular characterization, as transcriptional regulation in the *Mycoplasma* spp.

KeyWords: Mycoplasma; Comparative genomics; Comparative transcriptomics; Swine respiratory tract.

LISTA DE ABREVIATURAS

ATCC – *American type culture collection*

BLAST - *Basic local alignment search tool*

cDNA - ácido desoxirribonucléico complementar

CDS – região codificante

COG – *clusters of orthologous groups*

DNA - ácido desoxirribonucléico

Kb - quilobase

KEGG - *Kyoto encyclopedia of genes and genomes*

MHP – *Mycoplasma hyopneumoniae*

MFL – *Mycoplasma flocculare*

MHR - *Mycoplasma hyorhinis*

mRNA – ácido ribonucleico mensageiro

NCBI - *Nacional Center for Biotechnology Information*

ORF - fase aberta de leitura

OC – cluster de fase aberta de leitura

pb – par de base

RNA - ácido ribonucleico

RNAP – RNA polimerase

rRNA – ácido ribonucleico ribossômico

RT-PCR - reação em cadeia da polimerase com transcrição reversa

tRNA – ácido ribonucleico transportador

UT – unidade transcricional

Vlp – lipoproteína variável

1. INTRODUÇÃO

1.1. O gênero *Mycoplasma*

Microrganismos do genero *Mycoplasma*, do grego *mykes* (fungo) e *plasma* (moldável), pertencem à família *Mycoplasmataceae*, classe *Mollicutes*, do latim *mollis* (mole) e *cutis* (pele). Essa classificação ocorre, principalmente, pela ausência de parede celular (Razin *et al*., 1998). Essa classe é composta pelos gêneros *Acholeplasma*, *Anaeroplasma*, *Asterosplama*, *Mycoplasma*, *Spiroplasma* e *Ureaplasma* (Walker, 2003). Devido ao fato de não sintetizarem peptideoglicano, ou seus precursores, não possuem parede celular rígida, mas sim, uma membrana trilaminar simples composta de proteínas, glicoproteínas, fosfolipídeos e colesterol, sendo esse último, o responsável pela fluidez e estabilidade osmótica da membrana. São corados insatisfatoriamente pelo Método de Gram, sendo recomendadas as colorações de Giemsa, Castañeda, Dienes e Novo Azul de Metileno.

Os micoplasmas são considerados pleomórficos, resultado da ausência de parede celular. A célula pode ser esférica, em forma de pera, espiralada ou filamentosa. Além de tudo, a ausência de parede celular torna os micoplasmas resistentes aos antimicrobianos que afetam a sua síntese, como penicilinas, cefalosporinas e bacitracinas, sendo esses, portanto, ineficazes para tratamento (Sobestiansky *et al*., 1999). Além disso, necessitam que haja um contato íntimo com as células do hospedeiro, para que seja capaz de suprir as necessidades nutricionais indispensáveis para a sua sobrevivência (Clark, 2005). Esses microrganismos se multiplicam lentamente em meio de cultura após incubação de 3 a 20 dias, a uma temperatura ótima

de 37 ºC, pH em torno de 7,5, atmosfera com 5 a 10% de CO_2 e leve agitação (Walker, 2003).

O genero *Mycoplasma* está amplamente distribuído no reino animal, parasitando mamíferos, aves, répteis, anfíbios e peixes (Pitcher & Nicholas, 2005). Normalmente, estão aderidos à superfície extracelular de células e tecidos do hospedeiro, embora já tenham sido descritas algumas espécies ocupando o interior de células eucarióticas (Lo *et al*., 1993; Baseman *et al*., 1995). Podem ser patogênicos ou apenas fazerem parte da microbiota do trato respiratório e/ou urogenital (Razin, 2006). Em humanos, estão relacionados a enfermidades como asma, câncer, doenças auto-imunes, artrite e pneumonia (Baseman & Tully, 1997).

Quanto à classificação morfológica, são bactérias gram-negativas, pois apresentam apenas uma membrana plasmática, sem a proteção adicional de uma parede celular (Razin, 2006). Entretanto, filogeneticamente, os micoplasmas estão relacionados às bactérias gram-positivas, compartilhando, desse modo, um ancestral em comum com os gêneros *Streptococcus*, *Lactobacillus*, *Bacillus* e *Clostridium* (Wolf *et al*., 2004). Possivelmente, a evolução dos micoplasmas ocorreu através de eventos de degeneração ou redução do genoma de bactérias gram-positivas portadoras de baixo conteúdo de guanina e citosina (G+C) (Woese, 1987).

Os genomas desses microrganismos são extremamente reduzidos, variando de 580 a 1.350 kb (Fraser *et al*., 1995; Sasaki *et al*., 2002) e apresentam baixo conteúdo de G+C, entre 23 e 40% (Woese, 1987). As regiões intergênicas têm um maior conteúdo de adenina e timina (A+T) em relação às regiões codificadores (Dybvig & Voelker, 1996). A distribuição de G+C ocorre de forma irregular, sendo essas bases mais frequentes em genes que codificam RNAs ribossômicos e transportadores (Fraser *et al*.,

1995). Como resultado dessa composição atípica dos genomas, há o favorecimento da utilização de códons que contêm A+T (Bove, 1993). Outra particularidade que parece estar ligada a essa questão, é a utilização do códon UGA (Osawa *et al*., 1992), não como códon de terminação, conforme o código genético universal, mas como códon para o aminoácido triptofano (Yamao *et al*., 1985), característica que é igualmente encontrada em genomas mitocondriais.

Esses microrganismos não possuem muitas das vias enzimáticas características da maioria das bactérias. Como exemplo disso, podemos ressaltar a ausência de vias *de novo* na biossíntese de purinas, um ciclo do ácido tricarboxílico completo e um sistema de cadeia transportadora de elétrons mediada por citocromo (Manolukas *et al*., 1988; Finch & Mitchell, 1992; Pollack, 1992; Fraser *et al*., 1995). Evidências apontam que essas características estão relacionadas ao fato da maioria dos micoplasmas serem parasitas, hospedeiro e tecido específicos (Razin *et al*., 1998; Rottem & Yogev, 2000).

O reduzido tamanho dos genomas de micoplasmas representa um claro exemplo do processo de redução evolutiva de genomas. É possível que, após uma etapa inicial de simbiose, a disponibilidade de uma ampla variedade de metabólitos elaborados, resultantes de atividades enzimáticas do hospedeiro ou de outras espécies bacterianas que habitam o mesmo nicho, associada à capacidade de absorvê-los do meio exterior, acabaria por tornar uma série de atividades bacterianas dispensáveis aos micoplasmas.

Essas bactérias estariam sofrendo um processo de perda de genes em decorrência do fato de estarem habitando um ambiente condicionado por outro(s) genoma(s). Genes envolvidos em vias metabólicas seriam particularmente suscetíveis a esse tipo de processo, resultando em bactérias com uma capacidade metabólica extremamente

limitada (Andersson & Kurland, 1998). Por outro lado, essas mesmas bactérias passariam a utilizar uma porção considerável de seus genomas em sistemas de transporte capazes de incorporar metabólitos e íons do meio exterior (Andersson & Kurland, 1998).

Além da grande relevância médica e veterinária, os micoplasmas têm ainda despertado interesse científico por possuírem genomas extremamente reduzidos, fazendo com que sejam objeto de estudos na determinação do conjunto mínimo de genes necessários para o estabelecimento de vida independente (Gil *et al*., 2004). Sendo assim, até o momento, 25 diferentes espécies desse gênero tiveram seus genomas sequenciados (NCBI, acesso em Junho de 2013).

1.2. *Mycoplasma hyopneumoniae*

Mycoplasma hyopneumoniae é o agente etiológico da pneumonia enzoótica suína (PES) – doença respiratória infecciosa crônica, caracterizada pela alta morbidade e baixa mortalidade (Sobestiansky *et al*., 1999). *M. hyopneumoniae* é um patógeno extracelular que coloniza o trato respiratório através da aderência às células do epitélio ciliar (Debey *et al*., 1992). Por se tratar de uma doença crônica, na PES a resposta imune adaptativa deve ser contornada para que o *M. hyopneumoniae* sobreviva e prolifere aderido às células do hospedeiro. Consequentemente, a habilidade em aderir firmemente ao epitélio ciliar respiratório é a estratégia fundamental do patógeno para superar as barreiras de defesa mucociliar (Zielinski *et al*., 1990). O estabelecimento da infecção resulta em ciliostase, perda dos cílios, morte das células epiteliais e inflamação aguda na traquéia, brônquios e bronquíolos. Condicionalmente, ocorre uma redução na

eficiência do sistema mucociliar e diminuição da resistência imunológica, predispondo os animais à infecções secundárias (Ciprian *et al*., 1988; Djordjevic *et al*., 2004).

A PES ocorre em âmbito mundial, estando presente em quase todos os rebanhos suínos (Minion *et al*., 2004). Por causar significativos gastos com medicamentos, atraso no ganho de peso, condenação de órgãos e carcaça e, por fim, menores valores de venda, a PES é considerada uma das principais causas de perdas econômicas na cadeia de produção suinícola (Rautiainen & Wallgren, 2001; Thacker, 2004). O aspecto econômico mais relevante refere-se à redução do desempenho do suíno, gerando um menor ganho de peso e pior conversão alimentar. As quedas na produtividade podem chegar a 20% sobre a taxa de conversão alimentar e até 30% sobre o ganho de peso, dependendo da gravidade das lesões e das infecções secundárias (Sobestiansky *et al*., 1999).

Os cultivos de *M. hyopneumoniae* são lentos quando comparados aos de outras espécies de micoplasmas suínos. O microrganismo multiplica-se em meio sólido somente na presença de 5 a 10% de CO_2. Friis (1975) desenvolveu um meio de cultivo líquido especial para isolamento de *M. hyopneumoniae* e *Mycoplasma hyorhinis*, o qual possui como indicador de pH o vermelho de fenol. A multiplicação bacteriana é controlada pela mudança de coloração do meio de cultivo do vermelho para amarelo, indicando a oxidação da glicose e acidificação do meio. Em função desse comportamento fastidioso, o isolamento bacteriano não é utilizado para o diagnóstico de rotina (Thacker, 2004). Essa característica fastidiosa está relacionada à sua dependência em relação ao fornecimento de determinados nutrientes e condições ambientais. Assim, vários aminoácidos essenciais e ácidos graxos são requeridos no meio de cultura (os quais são supridos pela adição de componentes específicos), uma vez que diversos

genes envolvidos na biossíntese desses compostos foram perdidos ao longo de sua evolução (Razin *et al*., 1998; Pollack, 2002).

Devido ao importante impacto econômico causado pela PES, muitos esforços são destinados aos estudos do agente *M. hyopneumoniae*. É nesse contexto que a Rede Sul de Análise de Genomas (PIGS, www.genesul.lncc.br) tem se empenhado tanto em estudos moleculares, quanto funcionais desse microrganismo. O sequenciamento dos genomas das linhagens J (ATCC 25934, não-patogênica) e 7448 (patogênica) de *M. hyopneumoniae*, realizado pela Rede PIGS (Vasconcelos *et al*., 2005) e o sequenciamento do genoma de uma linhagem patogênica americana (linhagem 232, Minion *et al*., 2004), bem como outra patogênica chinesa (linhagem 168, Liu *et al*., 2011), abre novas perspectivas para uma melhor compreensão da fisiologia dessas bactérias, e das suas relações com o hospedeiro suíno.

1.3. *Mycoplasma flocculare*

Mycoplasma flocculare é uma espécie de micoplasma encontrada em suínos, que também coloniza o trato respiratório (Friis, 1972). Embora não seja descrita como espécie patogênica, *M. flocculare* é isolado de pulmões afetados pela PES, associado à *M. hyopneumoniae*. Em animais sadios essa bactéria é encontrada no trato respiratório e cavidade nasal (Strasser & Nicolet, 1990). Portanto, a análise dos dados experimentais disponíveis permite considerar *M. flocculare* um microrganismo não patogênico, não afetando de modo significativo nem os pulmões nem a cavidade nasal dos suínos (Friis, 1973; Friis, 1974; Armstrong *et al*., 1987; Strasser *et al*.,1992; Blank & Stemke, 2001). No entanto, mudanças histológicas foram verificadas em suínos experimentalmente

infectados (Friis, 1973; Friis, 1974), tais como, proliferação de células linfo-histiocíticas na lâmina própria e danos no epitélio da cavidade nasal e também do pulmão. A adesão de *M. flocculare* aos cílios é diferenciada da adesão que ocorre por *M. hyopneumoniae*, uma vez que os danos observados são muito amenos. Acredita-se que esses microrganismos possuam diferentes adesinas que facilitam o reconhecimento de diferentes sítios receptores no tecido ciliar (Young *et al*., 2000).

M. flocculare é uma espécie exigente e de crescimento lento, sendo, no entanto, possível o seu cultivo em meio Friis (Friis, 1975). Os suínos são os únicos hospedeiros naturais de *M. flocculare* e até o momento não há descrição de seu isolamento em qualquer outra espécie animal (Kobisch & Friis, 1996).

O isolamento de *M. flocculare*, de um modo geral, é dificultado pela presença de outros micoplasmas como *M. hyopneumoniae* e *M. hyohrinis.* A similaridade genética existente entre *M. flocculare* e *M. hyopneumoniae* restringe o emprego de algumas ferramentas de identificação destes agentes. Análises de sequências de DNA dessas espécies permitiram a detecção de algumas diferenças entre as mesmas (Stemke *et al*., 1985; Blank & Stemke, 2001), porém comparações em nível de RNA ribossômico 16S evidenciaram uma alta similaridade, confirmando a estreita relação filogenética entre eles (Stemke *et al*., 1992; Stemke *et al*., 1994; Blank & Stemke, 2001).

A infecção por *M. flocculare* ocorre na maternidade suína, estando o agente amplamente distribuído em criações extensivas. Pulmões sadios (não-pneumônicos) são colonizados por propagação descendente através da traquéia (Kobisch & Friis, 1996). Diferentemente de *M. hyopneumoniae* a presença de *M. flocculare* no pulmão parece ter uma duração restrita (Armstrong *et al*., 1987). Em contrapartida, na cavidade nasal *M. flocculare* está presente em animais adultos, sendo detectado por toda a vida do hospedeiro.

Embora não descrito como patogênico, *M. flocculare* tem atraído discussões e interesse no meio científico, devido à sua grande similaridade genética com o agente etiológico da PES e, persistência em casos clínicos, não havendo informações precisas até o momento, sobre o papel deste microrganismo no trato respiratório suíno. Além do mais, há interesse no tratamento e prevenção da ocorrência de *M. flocculare* devido à sua capacidade de causar lesões histopatológicas, alterando, por conseguinte, o *status* sadio dos animais.

1.4. *Mycoplasma hyorhinis*

M. hyorhinis é um microrganismo frequentemente isolado de suínos. De acordo com Friis & Feenstra (1994) seu habitat normal são as membranas mucosas do trato respiratório superior e tonsilas, sendo esse microrganismo detectado principalmente em suínos nas fases de crescimento e terminação (Cole *et al*., 1985; Citti *et al*., 1997). A propagação descendente aos pulmões ocorre facilmente, fazendo assim com que *M. hyorhinis* seja considerado agente etiológico do complexo de pneumonias suínas (Friis, 1971), estando associado à infecção por *M. hyopneumoniae* (Caron *et al*., 2000). *M. hyorhinis* também tem sido frequentemente associado como agente etiológico de poliserosites (Ross *et al*., 1973; Rosengarten & Wise, 1990). Além do mais, este microrganismo é amplamente conhecido como um contaminante de culturas de linhagens celulares (Kobisch & Friis, 1996; Drexler & Uphoff, 2002; Timenetsky *et al*., 2006). Estudos mais recentes vêm demonstrando sua associação com diferentes células tumorais, em um papel acelerador do câncer humano (Huang *et al*., 2001; Ketcham *et al*., 2005; Goodison *et al*., 2007; Liu & Shou, 2011; Urbanek *et al*., 2011).

A transmissão de *M. hyorhinis* aos leitões ocorre através da fêmea logo após o nascimento ou pelo contato direto com animais adultos. O microrganismo pode ser isolado tanto de secreções nasais quanto de pulmões, sendo que o meio de cultivo e as condições de multiplicação de *M. hyorhinis* são muito semelhantes ao *M. hyopneumoniae* (Friis, 1975). Em muitos casos, suínos infectados podem não apresentar sinais clínicos (Ross, 1992).

A partir do trato respiratório *M. hyorhinis* pode se espalhar via hematógena causando lesões serofibrinosas nas cavidades pericárdicas, pleurais e peritoniais (Friis & Feenstra, 1994). Além disso, essa espécie de micoplasma é considerada agente etiológico de processos de artrite em suínos (Citti *et al*., 1997). Há ainda descrições de isolamento de *M. hyorhinis* e/ou presença de anticorpos contra o mesmo, no epitélio auricular de suínos afetados com otite média (Morita *et al*., 1993; Kazama *et al*., 1994; Morita *et al*., 1995; Morita *et al*., 1998; Morita *et al*., 1999; Friis *et al*., 2002).

Genes codificadores de proteínas de superfície em *Mycoplasma* spp., que são capazes de assumir formas variadas, são amplamente descritos (Wise *et al*., 1992; Markham *et al*., 1994; Bhugra *et al*., 1995; Lysnyansky *et al*., 1996), sugerindo que mudanças na arquitetura da superfície bacteriana pode ser uma estratégia comum de adaptação no hospedeiro. Da mesma forma, *M. hyorhinis* possui um sistema de lipoproteínas variáveis de superfície (Vlps), as quais apresentam uma função de proteção do microrganismo contra a resposta humoral do hospedeiro, além de ser uma estratégia adaptativa, pois possibilita uma ligação eficiente do micoplasma aos antígenos de superfície do hospedeiro (Citti *et al*., 1997).

A expressão das Vlps envolve uma alta frequência de mutações e as principais características desse sistema foram extensivamente descritas (Rosengarten *et al*., 1991; Yogev *et al*., 1991; Yogev *et al*., 1995) e incluem: i) alta frequência de fenóticos

gerados devido a intensas mutações na expressão de Vlps e por consequência, variações nos tamanhos das populações; ii) os genes codificadores para as Vlps estão arranjados em unidades transcricionais, porém o numero de genes em cada cluster varia entre diferentes isolados; iii) as mutações ocorrem de modo aleatório, gerando assim variações nos tamanhos dos produtos, que estão relacionados à recombinação intragênica de sequências repetidas em *tandem* na região 3' de cada gene que poderá gerar a extensão ou contração da região C terminal destas lipoproteínas; iv) a região promotora de cada gene *vlp* parece estar susceptível a inserções e deleções desencadeando variação de fase na expressão de Vlp; e v) o processamento e exportação destas lipoproteínas pela membrana de *M. hyorhinis* e sua expressão como estrutura antigenicamente distintas permite a elas apresentarem-se sozinhas ou em combinação, formando mosaicos na superfície da célula, modulando inclusive a exposição de outros componentes da superfície. Em resumo, o sistema Vlp permite uma grande diversidade de superfície para diferentes isolados de *M. hyorhinis*¸ sendo, portanto, um sistema relacionado à capacidade de invasão e geração de efeitos deletérios ao hospedeiro (Wise *et al*., 1992; Wise, 1993; Citti *et al*., 1997).

Sete Vlps diferentes (nomeadas de A a G) foram identificadas em *M. hyorhinis* (Citti *et al*., 2000). Análises comparativas desses lócus entre cinco diferentes isolados de *M. hyorhinis* (linhagem MCLD, isolada de uma linhagem de células de melanoma; linhagem GDL, isolado de cultura celular; linhagem SK76, isolado de artrite; linhagem HUB-1, isolado de pneumonia suína e um clone variante da linhagem SK76) revelaram que a linhagem MCLD apresenta apenas quatro genes (vlpD, vlpE, vlpB e vlpC), enquanto a linhagem HUB-1 e a linhagem SK76 contêm os sete genes descritos (vlpA a vlpG). Por outro lado, o isolado *M. hyorhinis* GDL apresenta seis dos genes (vlpA a

vlpF) e por fim, o clone derivado da linhagem SK76 possui apenas três genes (vlpA a vlpC) (Kornspan *et al*., 2011).

As opiniões são muito distintas quanto à participação do *M. hyorhinis* nas enfermidades suínas, isso provavelmente devido à existência de variações de virulência próprias de cada isolado (Gois & Kuksa, 1974; Friis & Feenstra, 1994; Kobisch & Friis, 1996, Citti *et al*., 1997). Em trabalho experimental realizado por Kobisch (1983) foi demonstrada essa plasticidade de distintas linhagens apresentarem diferenças na capacidade de virulência. Nesse estudo, animais experimentalmente infectados com cinco diferentes linhagens de *M. hyorhinis* foram monitorados e sacrificados ao longo de 14 semanas, possibilitando a observação dos sinais clínicos e achados "pós-mortem". Poliserosite, artrite, peritonite, hipertrofia das membranas sinoviais foram amplamente visualizadas, enquanto em apenas um dos animais observou-se adesão pleural. Lesões pulmonares ou rinite não foram notadas em nenhum dos animais. No entanto, *M. hyorhinis* foi isolado de secreções nasais, pulmão, juntas, linfonodos e exsudatos pleural, peritoneal e pericardial (Kobisch, 1983). Friis & Feenstra (1994) da mesma forma experimental, também reproduziram, em suínos, essas alterações de serosite. Além disso, após inoculação intraperitoneal de *M. hyorhinis*, mesmo com ausência de sinais clínicos, foram observadas alterações microscópicas no tecido cerebral, indicando que esse micoplasma é capaz de se instalar também no cérebro (Cole *et al*.,1985).

A capacidade de *M. hyorhinis* infectar diferentes tipos celulares vem sendo amplamente analisada. Estudos têm demonstrado uma surpreendente associação entre *M. hyorhinis* e o desenvolvimento de câncer humano (Huang *et al*., 2001). Especula-se que essa ação de *M. hyorhinis* é devida a uma lipoproteína, denominada proteína p37, a qual é capaz de duplicar a capacidade invasiva e agressividade do câncer de próstata (Ketcham *et al*., 2005; Goodison *et al*., 2007; Urbanek *et al*., 2011), melanomas

(Ketcham *et al*., 2005), células de fibrossarcoma (Dudler *et al*., 1988) e câncer gástrico (Yang *et al*., 2010). *M. hyorhinis* p37 foi extensivamente detectada em tecidos acometidos por tumores; e, no caso de câncer gástrico e de cólon, *M. hyorhinis* foi encontrado em aproximadamente 56% das biópsias realizadas por Huang *et al*. (2001).

Estudos mais recentes alertam para a capacidade de *M. hyorhinis* induzir a transformação de células prostáticas à malignidade, hipotetizando a conexão entre a exposição ao micoplasma e o início e propagação do câncer (Urbanek *et al*., 2011). Portanto, *M. hyorhinis* é considerado indutor do fenótipo de malignidade, promovendo a migração celular, invasão e metástases (Yang *et al*., 2010). Além do mais, foi demonstrado que esse agente é capaz de inibir a proliferação de linfócitos (Zinocker *et al*., 2011) e invadir o sistema imunológico do hospedeiro sem ser reconhecido (Liu & Shou, 2011). Concomitantemente à capacidade de promover metástases e malignidade, *M. hyorhinis* induz apoptose celular, acelerando desta maneira, a destruição de células mais fracas e consequentemente fortalecendo as células tumorais mais aptas (Liu & Shou, 2011).

A infecção celular por *M. hyorhinis* pode gerar, por outro lado, uma resposta positiva. O microrganismo parece ser capaz de proteger as células contra certos tipos de insultos (Elkind *et al*., 2011) pela elevação nos níveis celulares de calpastatina e diminuição da atividade da calpaína (Elkind *et al*., 2010). Células neuronais são protegidas de toxicidade por *M. hyorhinis* (Elkind *et al*., 2011). Com isso, *M. hyorhinis* pode representar o primeiro sistema biológico que regula positivamente o inibidor de calpaína endógena, possibilitando, no futuro, o desenvolvimento de novos métodos de tratamento contra processos patológicos que envolvam a ativação excessiva de calpaína, como as desordens neurodegenerativas ou danos neurais procedidos de traumas (Elkind *et al*., 2011; Elkind *et al*., 2012).

Frente à sua importância tanto na medicina veterinária quanto na medicina humana, quatro diferentes linhagens de *M. hyorhinis* tiveram seus genomas sequenciados até o momento (Liu *et al*., 2010; Kornspan *et al*., 2011; Calcutt *et al*., 2012; Goodison *et al*., 2013). Da mesma forma que em outros microrganismos ou micoplasmas, os estudos em nível genômico de *M. hyorhinis* são essenciais para um aprofundamento dos conhecimentos sobre a biologia básica desta bactéria, a fim de, ampliar os caminhos do entendimento das interações patógeno-hospedeiro, além das interações entre os próprios microrganismos.

1.5. Organização gênica em *Mycoplasma*

Unidades transcricionais (UT), também conhecidas como operons, representam uma das principais estratégias de organização e regulação gênica em procariotos. Em um genoma procariótico típico, aproximadamente 50% das sequências codificadoras estão organizadas em UTs (Price *et al*., 2006). De um modo geral, a organização em unidades transcricionais tende a apresentar sequências regulatórias mais complexas e conservadas do que genes transcritos individualmente - monocistrônicos (Rogozin *et al*., 2002; Hazkani-Covo & Graur, 2005; Price *et al*., 2005). Genes envolvidos em diferentes processos biológicos podem ser requeridos em uma mesma condição ambiental. Exemplo disso é a existência de operons conservados contendo genes que codificam proteínas ribossômicas e enzimas do metabolismo central entre os gêneros bacterianos, talvez porque ambos são requeridos em taxas proporcionais durante o crescimento do organismo (Rogozin *et al*., 2002).

Inicialmente operons foram definidos como grupo de genes separados por um curto espaço intergênico, codificadores de produtos funcionalmente relacionados,

regulados pela mesma sequência promotora e transcritos em um mRNA policistrônico (Jacob & Monod, 1961). Apesar da definição clássica de unidades transcricionais ou operons ainda ser válida para grande parte das bactérias, muitos genes, sem aparentes relações funcionais, são comprovadamente co-transcritos (Price *et al*., 2006; Fondi *et al*., 2009; Siqueira *et al*., 2011). Evidências também apontam a presença de múltiplos promotores e/ou terminadores controlando diversas unidades de transcrição que, portanto, não são estruturadas como a definição original (Vicente *et al*., 1998; Price *et al*., 2006; Weber *et al*., 2012). Associada a essas informações, análises detalhadas de transcritomas e proteomas indicam que as informações contidas e a estrutura de genomas procarióticos são muito mais complexas do que previamente se imaginava (Cho *et al*., 2008; Medini *et al*., 2008; Güell *et al*., 2009).

A partir da análise de diversos trabalhos podemos afirmar que o grau de compactação e consequentemente as distâncias intergênicas são extremamente variáveis, nos mais diferentes gêneros bacterianos (Eyre-Walker, 1995; Moreno-Hagelsieb & Collado-Vides, 2002; Price *et al*., 2006). A ocorrência de extensos espaços intergênicos é frequente em UTs altamente expressas, podendo isso, estar relacionado à presença de promotores internos (Price *et al*., 2006). Em *Synechocystis*, por exemplo, a análise das ORFs na mesma fita de DNA, estimou que 48% do total de ORFs formam UTs (Price *et al*., 2005), sendo que as distâncias intergênicas podem chegar a 300 pb, número que foi considerado pelos autores como extremamente alto.

Estudos da organização gênica em micoplasmas são mais restritos e, na sua maioria, limitados a uma ou poucas unidades transcricionais (Inamine *et al*., 1988; Himmelreich *et al*., 1997; Waldo III *et al*., 1999; Musatovova *et al*., 2003; Adams *et al*., 2005; Waldo & Krause, 2006). Entretanto, esses trabalhos são importantes, pois retratam a existência de co-regulação nesse gênero bacteriano. Análises em escala

genômica propõem para micoplasmas a organização extensiva em UTs, gerando mRNA policistrônicos longos, sendo a transcrição gênica preferencialmente guiada dessa forma (Himmelreich *et al*., 1997; Madeira & Gabriel, 2007; Güell *et al*., 2009; Siqueira *et al*., 2011). Dados experimentais, de análise total do genoma, estão disponíveis para *Mycoplasma pneumoniae* (Güell *et al*., 2009), enquanto para *M. hyopneumoniae* a maioria dos dados é proveniente de análises *in silico* (Madeira & Gabriel, 2007; Siqueira *et al*., 2011), bem como para *Mycoplasma genitalium* (Himmelreich *et al*., 1997).

Genes codificadores de proteínas ribossômicas estão agrupados em UTs na grande maioria das espécies bacterianas (Fujita *et al*., 1999). Similarmente, essa forma de organização também ocorre em *Mycoplasma* spp., porém, alguns agrupamentos dessas proteínas diferem. Essas diferenças foram encontradas em *Mycoplasma gallisepticum*, *M. pneumoniae*, *M. genitalium*, *Ureaplasma urealyticum* e *Bacillus subtilis*, essa última, uma espécie que compartilha ancestral comum com os micoplasmas (Skamrov *et al*., 2002). Em *M. hyopneumoniae* esta diversidade de organização das proteínas ribossômicas também ocorre (Siqueira *et al*., 2011), havendo genes codificadores dispersos por toda a extensão do genoma, contidos em UTs com genes cujos produt s apresentam as funções mais variadas, ou até com função ainda desconhecida.

A análise da organização do transcritoma de *M. pneumoniae*, realizada por Güell *et al*. (2009) revela uma organização complexa da coordenação da transcrição nesse micoplasma. O padrão de expressão nas UTs de *M. pneumoniae* varia significativamente, ocorrendo mudanças na estrutura das UTs, as quais geram transcritos alternativos em resposta a mudanças ambientais. RNAs antisenso foram encontrados com grande frequência no genoma, o que pode estar associado a essa

plasticidade em mudanças de padrão de expressão determinando uma regulação transcricional complexa (Güell *et al*., 2009). Essa geração de UTs alternativas pode ser resultado da modulação da transcrição tanto dos genes internos, quanto dos genes localizados no início ou no final dos transcritos (Güell *et al*., 2009); e indicativo da ocorrência de sítios internos de início e término da transcrição.

Um sistema diferencial de organização gênica, com a formação de longos RNAs policistrônicos, na maioria das vezes compostos por genes com funções não relacionadas, foi observado em micoplasmas (Himmelreich *et al*., 1997; Waldo *et al*., 1999; Güell *et al*., 2009; Siqueira *et al*., 2011). O início da transcrição, e também a sua regulação, parecem estar intimamente relacionados à localização física dos genes no genoma de *M. hyopneumoniae*. A evidência de que genes posicionados na mesma direção na fita de DNA são co-transcritos em grandes UTs até a ocorrência de um gene ou nova UT na fita oposta (Siqueira *et al*., 2011) sugere a existência de mecanismos de regulação complexos e peculiares à esta bactéria. No trabalho supracitado, as UTs são nomeadas como ORF *cluster* (OC). De acordo com a análise *in silico*, a qual foi suportada por resultados experimentais de 21 UTs, 95% das ORFs presentes no genoma de *M. hyopneumoniae* são arranjadas em UTs. Essas unidades são muito variáveis quanto à composição gênica, tamanho total, distância intergênica das ORFs bem como número de ORFs que as compõem (Siqueira *et al*., 2011). A Figura 1 ilustra regiões do genoma de *M. hyopneumoniae* que foram experimentalmente analisadas, validando os achados das análises *in silico* e mostrando que ORFs posicionadas na mesma fita de DNA, até a ocorrência de ORFs na fita de direção oposta, formam UTs. Em resumo, acredita-se que, nessas bactérias, a transcrição inicie no primeiro gene de um grande agrupamento e termine com baixa frequência, pois o término da transcrição parece ser

fraco, ocasionando uma discreta e ineficiente parada da transcrição (Madeira & Gabriel, 2007; Güell *et al*., 2009; Gardner & Minion, 2010; Siqueira *et al*., 2011).

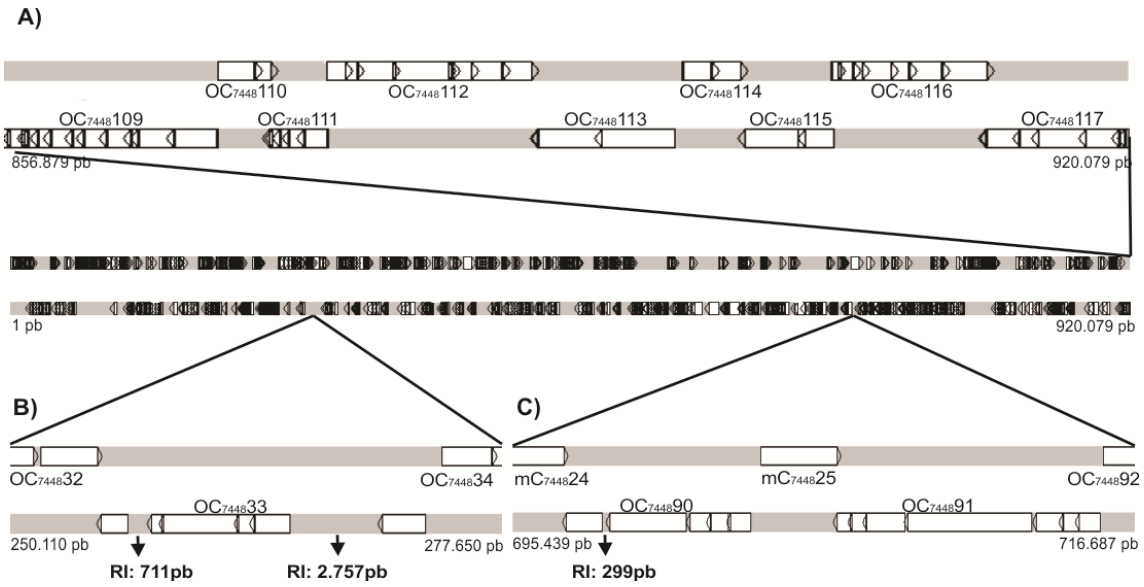


Figura 1. Representação da análise experimental de UTs no genoma de *M. hyopneumoniae* 7448. Algumas das regiões testadas são ilustradas na imagem (A, B e C). As unidades transcricionais apresentadas estão nomeadas como *OC*, e os genes monocistrônicos estão nomeados como *mC*. Estas imagens foram ampliadas a partir do genoma total (marcado de 1pb a 920.079 pb). A) Cobertura de 63,2 kb do genoma, com nove UTs presentes. B) Cobertura de of 27,5 kb do genoma, com três UTs. As grandes regiões intergênicas desta região estão destacadas. C) Cobertura de 21,2 kb do genoma mostrando três UTs e duas ORFs monocistrônicas. A região intergênica destacada na $OC_{7448}90$ representa a sequência entre os genes *tuf* e *lon.* RI: região intergênica. Adaptado de Siqueira *et al*. (2011).

1.6. Regulação da Transcrição Gênica em *Mycoplasma*

A regulação transcricional bacteriana tem sido tópico de interesse ao longo de diversas décadas, tendo como base a descrição do circuito de regulação o operon *lac* de *Escherichia coli* (Jacob & Monod, 1959). Genes bacterianos podem ser arranjados em operons, os quais foram denominados como grupos de genes contínuos, conjuntamente regulados. A transcrição bacteriana é carreada por uma única RNA polimerase (RNAP) holoenzima que consiste de uma maquinaria enzimática central e o fator sigma (σ). Diferentes fatores sigma reconhecem diferentes promotores e assim

controlam genes específicos. Outros fatores transcricionais agem também como reguladores, podendo eles promover ou reprimir a transcrição. Ao longo dos anos foi se verificando que esta maquinaria regulatória básica, é extremamente variável, podendo incluir centenas de elementos regulatórios e proteínas neste processo, aumentando a complexidade dos eventos envolvidos no processo de DNA à proteína: associação de proteínas à RNAP que afetam sua processividade (Sigmund & Morgan, 1988); operons com promotores internos (Ma *et al*., 1981; Siqueira *et al*., 2011); pequenos RNAs (sRNAs) (Andre *et al*., 2008) e *riboswitches* (RNAs que regulam a atividade interna de um gene) (Winkler *et al*., 2002); terminadores transcricionais adicionais (Santangelo & Artsimovitch, 2011); motivos de ligação ao ribossomo não-canônicos (Boni *et al*., 2001) e mRNAs "líderes" que são traduzidos (Laursen *et al*., 2005). Anteriormente, os exemplos supracitados eram tidos como raras exceções (Güell *et al*., 2011). Porém, com o emprego de tecnologias de alto desempenho o tema "regulação da expressão gênica em bactérias" voltou a despertar grande interesse científico, e os dados gerados permitem demonstrar que o que era considerado exceção, é na realidade a regra. Estas novas evidências sugerem que a definição clássica de operon pode ser reavaliada (Makita *et al*., 2004; de Hoon *et al*., 2005; Güell *et al*., 2009; Siqueira *et al*., 2011), sendo que, a complexidade da estrutura dos operons bacterianos pode ser comparada à complexidade de organização gênica de eucariotos (Güell *et al*., 2011).

Na maioria dos microrganismos, os mecanismos regulatórios podem ocorrer em diversas etapas da expressão gênica sendo, mais frequentemente, na etapa de início da transcrição. Diferentemente de arqueas e organismos eucarióticos, as bactérias apresentam somente um tipo de RNA polimerase. Entretanto, na maioria das espécies bacterianas são encontradas diversas subunidades sigma, gerando multiplicidade na atividade da RNAP, permitindo assim, o reconhecimento de diferentes tipos de

sequências promotoras. A utilização de diversos promotores pelo mecanismo de regulação por subunidade sigma alternativa permite a expressão diferencial de genes dependendo das condições nutricionais ou ambientais (Madeira & Gabriel, 2007). Porém, nos genomas de micoplasmas é encontrada somente uma única sequência codificadora para a subunidade sigma, além de um número mínimo de reguladores transcricionais (Minion *et al*., 2004; Vasconcelos *et al*., 2005; Liu *et al*., 2011).

Em genomas bacterianos, o número de genes envolvidos com a transcrição, varia entre as sequências analisadas. Em micoplasmas, o total de ORFs envolvidas na transcrição varia entre 11 a 23, o que representa 2 a 3% do total de CDSs (Sasaki *et al*., 2002; Minion *et al*., 2004; Vasconcelos *et al*., 2005). Este índice é considerado baixo quando comparado a outras espécies bacterianas, como *E. coli* (8,1%) e *B. subtilis* (9%) (Madeira & Gabriel, 2007).

Acredita-se que, durante a evolução redutiva, pela qual provavelmente os organismos do gênero *Mycoplasma* foram submetidos, diversos mecanismos regulatórios foram perdidos (Halbedel *et al*., 2007). Portanto, não possuem alguns genes codificadores de ortólogos bacterianos convencionais, tais como: múltiplos fatores sigma, sistemas de dois componentes (*two-component system*) e fator de terminação de transcrição Rho (Fraser *et al*., 1995). Consequentemente, os sinais que promovem e regulam a transcrição em micoplasmas provavelmente diferem significativamente de outras bactérias (Weiner III *et al*., 2000).

Apesar da ausência de fatores regulatórios convencionais, evidências confirmam a ocorrência de regulação gênica em micoplasmas. *M. hyopneumoniae* é capaz de gerar mudanças no perfil transcricional em resposta a diferentes variações ambientais (Madsen *et al*., 2006a; Madsen *et al*., 2006b; Madsen *et al*., 2008; Oneal *et al*., 2008;

Schafer *et al*., 2007). Do mesmo modo, *M. pneumoniae* responde, com variações nos níveis transcricionais, quando submetido a diferentes condições (Güell *et al*., 2009). No estudo transcritômico realizado em *M. pneumoniae* por Güell *et al*. (2009), surpreendentemente foram detectados, em alta frequência de ocorrência, transcritos anti-senso, transcrição preferencial em mRNAs policistrônicos (grandes unidades transcricionais) com muitos transcritos alternativos dentro dessas unidades, além de múltiplos elementos reguladores por gene, implicando em um transcritoma altamente dinâmico, mais similar à eucariotos que previamente imaginado.

Alguns estudos foram realizados, em *M. pneumoniae,* objetivando analisar sequências promotoras e assim, gerar um melhor entendimento da transcrição. Dessa forma, foram identificadas *in silico* regiões envolvidas no início da transcrição, com similaridade à região -10 (TATA Box). Porém, a região -35 não se apresentou claramente definida (Weiner III *et al*., 2000). Em estudo *in vivo*, realizado por Halbedel *et al*. (2007) em uma das regiões avaliadas por Weiner III *et al*. (2000), foi demonstrado que, realmente, a região -10 é importante para a transcrição de micoplasmas, enquanto a região -35 parece não apresentar a mesma relevância.

Mais recentemente, Weber *et al*. (2012) objetivando identificar os elementos básicos envolvidos no início da transcrição desenvolveram uma matriz de pontuação posição-específica para a predição de sequências promotoras, a partir de dados de mapeamento experimental dos sítios de início de transcrição, especificamente para *M. hyopneumoniae.* Elementos putativos -10 e -16 foram identificados, porém, similar ao encontrado em outras espécies de *Mycoplasmas*, nenhum padrão na região -35 putativa foi verificado. Diferentemente de outras espécies bacterianas, em *M. hyopneumoniae* sítios de início de transcrição (TSSs) heterogêneos são encontrados com frequência e uma região líder não-traduzida que contenha um sítio de ligação ao ribossomo (RBS)

parece ser ausente (Weber *et al*., 2012). Além do mais, essa matriz mostrou-se espécie-específica, reforçando a necessidade de estudos e desenvolvimento de ferramentas específicas para o microrganismo de interesse.

A alta frequência de transcritos alternativos, assim como a subdivisão das unidades transcricionais de *M. pneumoniae* em UTs menores, provavelmente, também sejam observadas em *M. hyopneumoniae*. A predição de promotores (Weber *et al*., 2012) imediatamente à montante de vários genes considerados internos em UTs preditas por Siqueira *et al*. (2011) reforça essa possibilidade. Portanto, micoplasmas parecem utilizar mecanismos mais específicos de início da transcrição, além de outros eventos ou elementos na regulação da transcrição gênica adicionais aos fatores de transcrição convencionais (Güell *et al*., 2009; Yus *et al*., 2009; Kühner *et al*., 2009).

Os mecanismos envolvidos na terminação da transcrição dos micoplasmas ainda não são claramente compreendidos. Por um lado existe a afirmação de que a terminação rho-independente seria o principal mecanismo de terminação da transcrição (deHoon *et al*., 2005; Güell *et al*., 2009), por outro, existem indicativos de que os grampos de terminação não participariam desse processo (Washio *et al*., 1998; Madeira & Gabriel, 2007; Gardner & Minion, 2010). Gardner & Minion (2010) demonstraram que, em *M. hyopneumoniae*, mais de 90% das regiões intergênicas maiores que 50 pb são transcritas, com um controle de terminação da transcrição não-estringente. A terminação abrupta da transcrição, que indicaria a presença de terminadores, parece não ocorrer com grande frequência no genoma dessa espécie, sendo o mais comum o desligamento gradual da RNA polimerase (Gardner & Minion, 2010). Ainda, por análises *in silico*, Madeira & Gabriel (2007) também concluíram que os sinais de terminação em *M. hyopneumoniae* são fracos e ineficientes em gerar parada, tornando possível a transcrição continua de grandes grupos gênicos.

1.7. Sequenciamento de alto desempenho nos estudos genômicos e transcritômicos

Sequenciamento de última geração (*Next-generation sequencing* – NGS) ou sequenciamento de alto desempenho refere-se aos métodos de sequenciamento de DNA e RNA, substitutos do sequenciamento convencional pelo método de Sanger (Sanger *et al*., 1977). Os métodos de NGS incluem dentre outros: o pirosequenciamento (Margulies *et al*., 2005); o sequenciamento *Illumina* (Fedurco *et al*., 2006; Turcatti *et al*., 2008); o sequenciamento SOLiD (Mckernan *et al*., 2006); e a tecnologia *Ion Torrent* (Rothberg *et al*., 2011). Essas novas ferramentas não apenas tornam o processo de sequenciamento de DNA e RNA mais eficiente e de menor custo, mas também permitem principalmente, o desenvolvimento de abordagens experimentais inovadoras que possibilitam o aprofundamento dos mecanismos moleculares da organização dos genomas e funções celulares (Soon *et al*., 2013).

O sequenciamento da população total de RNA (transcritoma) usando tecnologias NGS, ou sequenciamento de RNA (RNA-seq) está revelando uma surpreendente complexidade de organização e dinâmica dos transcritomas das mais variadas espécies de organismos vivos, com uma alta sensibilidade, precisão e acurácia (Wang *et al*., 2009; Wilhelm & Landry, 2009; Marguerat & Bahler, 2010; Martin & Wang, 2011; Ozsolak & Milos, 2011). Nessas novas plataformas, transcritos raros com função regulatória são, de um modo geral, facilmente detectados. A figura 2 ilustra resumidamente a análise de perfil e níveis de expressão, empregando as diferentes plataformas de NGS disponíveis. A resolução do RNA-seq é em nível de pares de bases, permitindo uma montagem e visualização dinâmicas até os mais elevados níveis de expressão, sendo possível também, a anotação *de novo* (Martin & Wang, 2011).

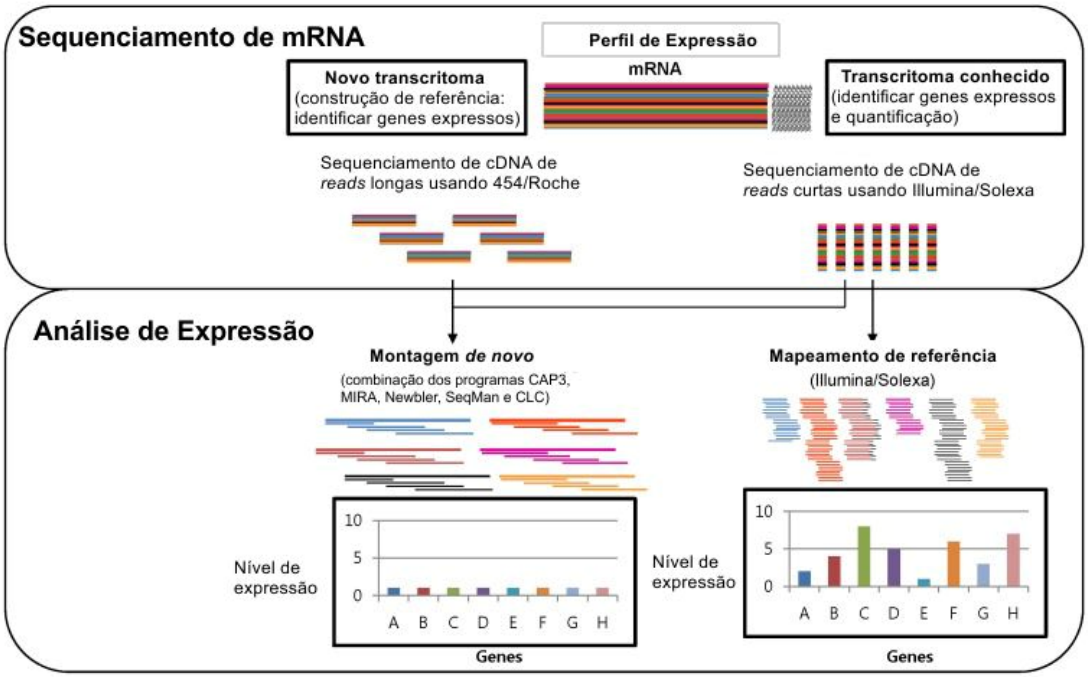


Figura 2. Esquema representativo da análise de transcritomas. Representação de análise de expressão por transcritoma usando plataformas de alto desempenho e bioinformática: A identificação de genes expressos através da montagem *de novo* com dados gerados pelas tecnologias Roche/454 e Illumina/Solexa e perfil de níveis de expressão através do mapeamento com sequências provenientes do Illumina/Solexa. Adaptado de Lim *et al*. (2012).

O transcritoma é o conjunto total de transcritos de uma célula em um momento ou estágio de desenvolvimento específico ou condição fisiológica. O entendimento do transcritoma é essencial para a interpretação de elementos funcionais do genoma e para a revelação de constituintes moleculares de células e tecidos, além do entendimento do desenvolvimento dos seres vivos e do desencadeamento de doenças (Wang *et al*., 2009). Os principais objetivos de uma análise transcritômica são: i) identificar todas as espécies de transcritos, incluindo mRNAs, RNAs não-codificadores e pequenos RNAs; ii) determinar a estrutura transcricional dos genes, em nível de sítios de início de transcrição, porções 5′ e 3′, *splicing* e outras modificações pós-transcricionais; e iii) quantificar os níveis de expressão de genes durante uma determinada condição analisada ou sob condições comparativas.

Como citado anteriormente, o sequenciamento de alto desempenho de bibliotecas de cDNA tem potencial para analisar a transcrição em nível de nucleotídeos, rendendo um maior detalhamento dos transcritos presentes na população em estudo. No entanto, quando comparado com o RNA eucariótico, o trabalho com RNA bacteriano é sempre desafiador. Diferentemente de eucariotos, a maioria dos mRNAs bacterianos não possui cauda de poli-A (Deutscher, 2003), além de uma preparação de RNA típica ser composta principalmente por rRNAs e tRNAs (95%) (Condon, 2007). Dessa forma, o mRNA bacteriano é mais susceptível à injúrias, com uma meia-vida curta o que o torna altamente instável (Deutscher, 2003; Condon, 2007).

Uma estratégia necessária e eficaz no preparo de bibliotecas de cDNA bacterinao é o enriquecimento da amostra de RNA com mRNAs. Baseado na remoção de rRNAs e tRNAs, essa estratégia é essencial para uma cobertura mais substancial dos transcritos e uma boa resolução dos mapas transcritômicos resultantes (Sorek & Cossart, 2010). A maioria dos mRNAs de bactérias e arqueias, apresenta a porção 5' trifosfatada (5'-PPP), que corresponde ao primeiro nucleotídeo incorporado na cadeia polinucleotídica. As moléculas processadas, como rRNAs e tRNAs apresentam uma extremidade 5' monofosfatada (5'P). Essa diferença é essencial para a seleção dos mRNAs. Diversos métodos podem ser empregados no enriquecimento de mRNA em amostras de RNAs de procariotos: i) degradação do RNA processado - degradação enzimática por uma exonuclease que degrada exclusivamente RNAs com porções 5' monofosfatadas, restando na preparação apenas os mRNAs intactos (Passalacqua *et al*., 2009); ii) poliadenilação seletiva dos mRNAs - poliadenilação de mRNAs pela poli(A) polimerase de *E. coli* e posterior captura usando sondas ou transcrição reversa; iii) captura de RNAs que interagem com proteínas específicas utilizando-se anticorpos –

empregado nos casos em que há interesse por um conjunto específico de RNAs (Sorek & Cossart, 2010).

Atualmente, um dos maiores desafios nas tecnologias NGS é a grande quantidade de dados gerados (Wang *et al*., 2009). Além do espaço necessário para armazenamento das informações, há a necessidade da determinação precisa das sequências para se mapear corretamente as *reads* de cDNA no genoma, e ainda remover as sequências de baixa qualidade. A correta visualização e interpretação dos resultados, associada às análises estatísticas, requer significativos níveis de conhecimento para possibilitar um efetivo suporte às informações.

Com a rápida evolução da capacidade de sequenciamento, através das novas tecnologias, em combinação com o enriquecimento de mRNA, tornou-se viável a exploração de transcritomas bacterianos completos. Desde o primeiro estudo transcritômico, surpreendentes achados têm sido revelados, que incluem uma grande quantidade de RNAs não-codificadores (ncRNAs), novos elementos regulatórios não traduzidos e estruturas alternativas de operons. Essa complexidade, até então desconhecida, foi descrita em estudos realizados em diferentes bactérias, como *Burkholderia cenocepacia* (Yoder-Himes *et al*., 2009), *Listeria monocytogenes* (Toledo-Arana *et al.,* 2009), *Bacillus anthracis* (Passalacqua *et al*., 2009), *B. subtilis* (Rasmussen *et al*., 2009), *Halobacterium salinarum* (Koide *et al*., 2009), *M. pneumoniae* (Güell *et al*., 2009), *Sulpholobus solfataricus* (Wurtzel, *et al*., 2009) e *Salmonella* spp. (Sittka *et al*., 2008; Perkins *et al*., 2009).

2. OBJETIVOS

2.1. Geral

Apesar de todos os dados gerados e esforços despendidos por diferentes grupos, os mecanismos empregados por micoplasmas para a definição do início da transcrição e regulação da expressão gênica ainda não são claramente compreendidos. Portanto, o objetivo geral desse trabalho é gerar informações sobre a organização de genomas de espécies de *Mycoplasmas* que co-habitam o trato respiratório de suínos, além da análise global do transcritoma dessas espécies, permitindo a comparação do padrão transcricional presente em *M. hyopneumoniae* (espécie patogênica) e *M. flocculare* (espécie não-patogênica), além de *M. hyorhinis* (patogênica e capaz de causar infecções em outros sítios além do trato respiratório).

2.2. Específicos

- Sequenciar o genoma de *M. flocculare*.

- Realizar um estudo genômico comparativo entre as três espécies que co-habitam o trato respiratório de suínos: *M. hyopneumoniae, M. flocculare* e *M. hyorhinis*.

- Construir bancos de cDNA a partir de RNAs de *M. hyopneumoniae, M. flocculare* e *M. hyorhinis* e analisar os dados de sequenciamento desses bancos.

- Analisar comparativamente os transcritomas das três espécies de *Mycoplasma*.

3. CAPÍTULO 1

NEW INSIGHTS ON THE BIOLOGY OF SWINE RESPIRATORY TRACT MYCOPLASMAS FROM A COMPARATIVE GENOME ANALYSIS

Franciele Maboni Siqueira[1,4], Claudia Elizabeth Thompson[2], Veridiana Gomes Virginio[1,3], Taylor Gonchoroski[1], Luciano Reolon[1,3], Luiz Gonzaga Almeida[2], Marbella Maria Bernardes da Fonsêca[2], Rangel de Souza[2], Francisco Prosdocimi[6], Irene Silveira Schrank[1,3,5], Henrique Bunselmeyer Ferreira[1,3,5], Ana Tereza Ribeiro de Vasconcelos[2*], Arnaldo Zaha[1,3,4,5*]

[1] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.
[2] Laboratório de Bioinformática. Laboratório Nacional de Computação Científica. Petrópolis, Rio de Janeiro, Brazil.
[3] Programa de Pós-Graduação em Biologia Celular e Molecular. Centro de Biotecnologia UFRGS, Porto Alegre, Brazil.
[4] Programa de Pós-Graduação em Ciências Biológicas - Bioquímica. UFRGS, Porto Alegre, Brazil.
[5] Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências. UFRGS, Porto Alegre, Brazil.
[6] Departamento de Bioquímica Médica. Centro de Ciências da Saúde. Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

atrv@lncc.br; zaha@cbiot.ufrgs.br

Artigo publicado na revista: *BMC Genomics*

BMC Genomics

RESEARCH ARTICLE Open Access

New insights on the biology of swine respiratory tract mycoplasmas from a comparative genome analysis

Franciele Maboni Siqueira[1,4], Claudia Elizabeth Thompson[2], Veridiana Gomes Virginio[1,3], Taylor Gonchoroski[1], Luciano Reolon[1,3], Luiz Gonzaga Almeida[2], Marbella Maria da Fonsêca[2], Rangel de Souza[2], Francisco Prosdocimi[6], Irene Silveira Schrank[1,3,5], Henrique Bunselmeyer Ferreira[1,3,5], Ana Tereza Ribeiro de Vasconcelos[2*] and Arnaldo Zaha[1,3,4,5*]

Abstract

Background: *Mycoplasma hyopneumoniae*, *Mycoplasma flocculare* and *Mycoplasma hyorhinis* live in swine respiratory tracts. *M. flocculare*, a commensal bacterium, is genetically closely related to *M. hyopneumoniae*, the causative agent of enzootic porcine pneumonia. *M. hyorhinis* is also pathogenic, causing polyserositis and arthritis. In this work, we present the genome sequences of *M. flocculare* and *M. hyopneumoniae* strain 7422, and we compare these genomes with the genomes of other *M. hyoponeumoniae* strain and to the a *M. hyorhinis* genome. These analyses were performed to identify possible characteristics that may help to explain the different behaviors of these species in swine respiratory tracts.

Results: The overall genome organization of three species was analyzed, revealing that the ORF clusters (OCs) differ considerably and that inversions and rearrangements are common. Although *M. flocculare* and *M. hyopneumoniae* display a high degree of similarity with respect to the gene content, only some genomic regions display considerable synteny. Genes encoding proteins that may be involved in host-cell adhesion in *M. hyopneumoniae* and *M. flocculare* display differences in genomic structure and organization. Some genes encoding adhesins of the P97 family are absent in *M. flocculare* and some contain sequence differences or lack of domains that are considered to be important for adhesion to host cells. The phylogenetic relationship of the three species was confirmed by a phylogenomic approach. The set of genes involved in metabolism, especially in the uptake of precursors for nucleic acids synthesis and nucleotide metabolism, display some differences in copy number and the presence/absence in the three species.

Conclusions: The comparative analyses of three mycoplasma species that inhabit the swine respiratory tract facilitated the identification of some characteristics that may be related to their different behaviors. *M. hyopneumoniae* and *M. flocculare* display many differences that may help to explain why one species is pathogenic and the other is considered to be commensal. However, it was not possible to identify specific virulence determinant factors that could explain the differences in the pathogenicity of the analyzed species. The *M. hyorhinis* genome contains differences in some components involved in metabolism and evasion of the host's immune system that may contribute to its growth aggressiveness. Several horizontal gene transfer events were identified. The phylogenomic analysis places *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* in the hyopneumoniae clade.

Keywords: Mycoplasma, Comparative genomics, Adhesins, Swine respiratory tract

* Correspondence: atrv@lncc.br; zaha@cbiot.ufrgs.br
[2]Laboratório de Bioinformática. Laboratório Nacional de Computação Científica. Petrópolis, Rio de Janeiro, Brazil
[1]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
Full list of author information is available at the end of the article

Background

Mycoplasmas belong to the class Mollicutes, which is a taxon of bacteria that is characterized by the absence of a cell wall, a relatively small genome size and a strong dependence on nutrients supplied by the host environment [1]. More than 120 mycoplasma species have been described, and although they display diverse life styles, most of the species are parasitic, implying the occurrence of different mechanisms by which they interact with host cells. Several mycoplasmas associate with their host cells through adhesins, while others may also invade cells [2-8]. Among mycoplasmas, several species are responsible for human, animal and plant diseases, but some species are considered commensal organisms [1].

Mycoplasma hyopneumoniae, *Mycoplasma flocculare* and *Mycoplasma hyorhinis* are the most important species that have been identified in porcine respiratory systems [9-11]. Based on a 16S rRNA sequence comparison, *M. hyopneumoniae* and *M. flocculare* are known to be closely related [12]. *M. hyopneumoniae* is the etiological agent of porcine mycoplasmal pneumonia, while *M. hyorhinis*, which causes polyserositis and arthritis, is also frequently found in swine respiratory tracts [13]. *M. flocculare* is also widespread in swine herds, but no disease has been associated with this species [14]. *M. hyopneumoniae* can adhere to the cilia of tracheal epithelial cells and causes damage. Although *M. flocculare* can also adhere to cilia, no resulting damage has been observed, suggesting that *M. hyopneumoniae* and *M. flocculare* may possess different adhesins, facilitating the recognition of different receptor sites on the cilia [15]. Additionally, while *M. flocculare* is restricted to the swine respiratory tract, *M. hyopneumoniae* and *M. hyorhinis* can also colonize other sites, such as cardiac or joint tissues [14,16]. These bacteria can even colonize different hosts; *M. hyorhinis* has been detected in human carcinoma tissues [17,18]. The genetic maps of *M. flocculare* ATCC 27716 and *M. hyopneumoniae* strain J have been compared, revealing that at least three chromosomal inversions have occurred since the divergence of both species [19].

In recent years, the genomes of several mycoplasma species have been sequenced. The absence of several metabolic pathways, which was suggested by genetic and biochemical studies [1], has been confirmed at the genome sequence level. Among the swine-infecting mycoplasmas, the genomes of *M. hyopneumoniae* (four strains), *M. hyorhinis* (four strains) and *Mycoplasma suis* (two strains) have been sequenced, facilitating the comparison of metabolic pathways and evidencing specific mechanisms that can be utilized to survive in different host environments [20-28].

Because the genome sequences of *M. flocculare* and *M. hyopneumoniae* strain 7422 have now been completed, the current study presents a comprehensive comparison of the *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis* genomes. These three mycoplasma species can inhabit swine respiratory tracts. We have assessed the overall genome organizations including analyses of the open reading frame (ORF) clusters (OCs), inversions and rearrangements, and coding capacities, including analyses of encoded metabolic pathways and surface protein repertoires. Potential mechanisms of interaction with host cells are evidenced, and their implications on pathogenicity are discussed. Additionally, a phylogenomic approach using 32 mycoplasma genomes (including two that are reported here for the first time) was implemented to reconstruct the evolutionary history of the swine mycoplasma genomes, individual genes and/or portions of their genomes, including horizontal gene transfer analysis.

Results and discussion

General genome features of *M. flocculare* and *M. hyopneumoniae* 7422

The *M. flocculare* and *M. hyopneumoniae* 7422 genomes are composed of single, circular chromosomes of 763,948 and 899,887 bp with 28.9% and 28.4% GC contents, respectively. The *M. flocculare* genome, which by assembly remained with 13 gaps, contains 585 coding sequences (CDSs), of which 356 have known functions and 229 are annotated as hypothetical. The *M. hyopneumoniae* 7422 genome, completely closed in one contig by assembly, comprises 692 CDSs, of which 414 correspond to proteins with known functions and 278 are annotated as hypothetical. The protein-coding regions occupy approximately 87% of each *M. flocculare* and *M. hyopneumoniae* 7422 chromosome, and the average ORF length is 1,145 bp. Each genome contains one gene encoding the ribosomal RNAs (rRNAs) 16S and 23S, one gene encoding rRNA 5S and 30 genes encoding the transfer RNAs (tRNAs) representing all 20 amino acids. The general genome features of the five strains, *M. hyopneumoniae* (7422, 7448, J, 232 and 168), *M. flocculare*, and *M. hyorhinis* HUB-1, were compared in this study and are listed in Table 1.

Among the *M. flocculare* and *M. hyopneumoniae* 7422 CDSs that encode proteins with known functions, 380 and 403 CDSs, respectively, were classified into COG families comprising 18 functional categories (Table 2). A functional classification based on the KEGG [29] analysis assigned 351 and 371 CDSs from *M. flocculare* and *M. hyopneumoniae* 7422, respectively, into 15 different categories (Table 2). The performance differences produced by COG with respect to KEGG may be attributable to the presence of paralogs. As expected, the general genomic features and similarities in all of the COG and KEGG categories were strikingly similar between *M. flocculare*, *M. hyopneumoniae*, and *M. hyorhinis*, which commonly exhibited small genome sizes, high AT contents, and no two signal transduction proteins (Table 2).

Table 1 Comparison of general features of different mycoplasmas species and strains

	Organism*						
	MHP 7422	MHP 7448	MHP J	MHP 232	MHP 168	MFL	MHR HUB-1
Total length (bp)	899,887	920,079	897,405	892,758	925,576	763,948	839,615
G + C content (%)	28.4	28.5	28.5	28.6	28.4	28.9	25.8
Total no. CDSs	692	716	690	692	695	585	654
Average CDS length (bp)	1,147	1,146	1,167	1,164	1,071	1,145	1,092
Known proteins	414	418	410	304	354	356	489
Hypothetical proteins	278	298	280	388	341	229	165
No. of rRNAs	3	3	3	3	3	3	3
No. of tRNAs	30	30	30	30	30	30	30

*Abreviations: MHP = *M. hyopneumoniae*; MFL = *M. flocculare*; MHR = M. hyorhinis

To identify the genes that constitute the core and pan-genome of *M. flocculare*, *M. hyopneumoniae*, and *M. hyorhinis*, we took advantage of the bidirectional best hit (BBH) approach and plotted the data in a Venn diagram (Figure 1). We identified a considerable number of unique (i.e., organism-specific) genes in *M. hyorhinis* that may underline the phenotypic differences between this species and *M. flocculare* and *M. hyopneumoniae*. Including the repertoire of surface proteins (discussed later) and the inositol metabolism pathway, we identified 76 genes that are unique to *M. flocculare*, 69 to *M. hyopneumoniae* and 234 to *M. hyorhinis*.

When compared to other sequenced strains of *M. hyopneumoniae*, the genome of strain 7422 displays a highly similar gene composition and organization, with the exception of the localization of the integrative conjugative element (ICEH), which is positioned from 139,715 to 162,049 bp in the 7422 genome and from 518,376 to 540,705 bp in the 7448 genome. The similarity between *M. hyopneumoniae* gene repertories was 88%

Table 2 Comparison of *Mycoplasma* sp. genomes statistics using KEGG classification

Category/Organism	MHP 7422		MHR		MHP 7488		MHP J		MFL	
	Number	%	Number	%	Number	%	Number	%	Number	%
Carbohydrate metabolism	83	22.4	73	17.3	82	22.2	82	22.4	73	20.8
Energy metabolism	20	5.4	22	5.2	19	5.1	20	5.5	20	5.7
Lipid metabolism	9	2.4	11	2.6	9	2.4	9	2.5	7	2
Nucleotide metabolism	47	12.7	44	10.4	46	12.5	45	12.3	46	13.1
Amino Acid metabolism	14	3.8	15	3.6	14	3.8	14	3.8	12	3.4
Metabolism of Other Amino Acids	7	1.9	8	1.9	7	1.9	7	1.9	7	2
Glycan Biosynthesis and Metabolism	1	0.3	5	1.2	2	0.5	1	0.3	2	0.6
Metabolism of Cofactors and Vitamins	9	2.4	14	3.3	9	2.4	8	2.2	8	2.3
Metabolism of Terpenoids and Polyketides	5	1.3	1	0.2	5	1.4	5	1.4	5	1.4
Membrane Transport	36	9.7	38	9	36	9.8	36	9.8	32	9
Folding. Sorting and Degradation	10	2.7	14	3.3	10	2.7	10	2.7	9	2.6
Replication and Repair	45	12.1	44	10.4	42	11.4	42	11.5	44	12.5
Transcription	3	0.8	3	0.7	3	0.8	3	0.8	3	0.9
Translation	72	19.4	98	23.2	73	19.8	72	19.7	73	20.8
Biosynthesis of Other Secondary Metabolites	8	2.2	2	0.5	9	2.4	9	2.5	7	2
Cell Motility	0	0	0	0	0	0	0	0	0	0
Signal Transduction	0	0	0	0	0	0	0	0	0	0
TOTAL	371		392		369		366		351	

Abbreviations as in Table 1.

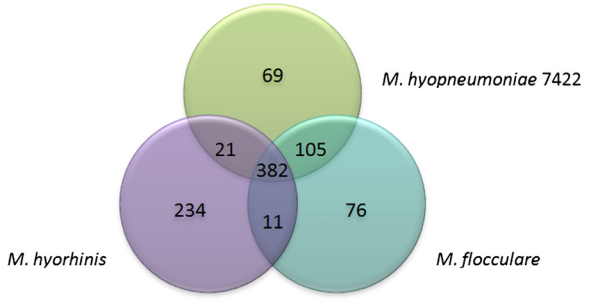


Figure 1 Venn diagram showing the distribution of *M. hyopneumoniae* 7422, *M. flocculare* and *M. hyorhinis* CDSs. Venn diagram identifying the total number of common and exclusive CDSs for each genome.

approximately. The small, but significative difference in the *M. hyopneumoniae* 7422 genome is the presence of an exclusive region of genes encoding transposases, hypothetical proteins and an ortholog of subtilisin-like serine protease (positioned from 497,277 to 510,210 bp). In comparison to 7422 genome, just one exclusive region was found in the 7448 genome, which is composed of genes encoding hypothetical proteins (positioned from 746,315 to 757,309 bp).

Comparison of OC organization in *M. flocculare, M. hyopneumoniae* and *M. hyorhinis* genomes

The gene-by-gene genome organization of *M. flocculare*, *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1 was analyzed, and the gene localization patterns were compared to detect ORFs with order conservation. The ORF cluster composition, organization and localization in the genomes were analyzed to determine the conservation level among the OC organization. Two groups of ORF clusters were created for each species, the OC group (Additional files 1 and 2) and the monocistronic gene (mC) group (Additional file 3). The general features of the OCs organization in the *M. flocculare, M. hyopneumoniae* and *M. hyorhinis* genomes are shown in Table 3 and Additional file 4.

A comparison of the OCs arrangements revealed a similar number of OCs among the *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis* genomes (Table 3). This result suggests that gene organization in *M. flocculare* and *M. hyorhinis* also occur preferably in clusters as found in *M. hyopneumoniae* [30]. Moreover, as previously described for *M. hyopneumoniae* [30], the overall ORF distribution within the OCs in *M. flocculare* and *M. hyorhinis* is highly variable with respect to the number of ORFs and the functional categories of the encoded products (Additional files 1 and 2).

An analysis of the mC group revealed a different ORF number in the *M. flocculare* genome when compared to the organization in the *M. hyopneumoniae* and *M. hyorhinis* genomes. There were 51, 34, and 30 mCs in the *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis* genomes, respectively (Additional file 3). Among all of the mCs, seventeen mCs were shared only by the *M. flocculare* and *M. hyopneumoniae* genomes. However, only the CDSs encoding an O-sialoglycoprotein endopeptidase (*gcp*) and an excinuclease ABC subunit C (*uvrC*) were found to display monocistronic organization in the three genomes (Additional file 5).

Table 3 Features of the *M. flocculare, M. hyopneumoniae* and *M. hyorhinis* OCs organization

Features	MFL*	MHP*	MHR*
Total length (base pairs)	772.687	920.079	839.615
Total No. of OCs (CDSs total)	114 (582)	117 (657)	98 (654)
Total No. of monocistronic group	51	34	34
Exclusives OCs	10	24	36

* Abbreviations as in Table 1.

A detailed analysis of the organization of each OC demonstrated a high level of conservation between the *M. flocculare* and *M. hyopneumoniae* genomes (Additional file 4). Approximately 78% and 46% of the OCs from *M. flocculare* display total or partial conserved gene distribution when compared to the OCs of *M. hyopneumoniae* and *M. hyorhinis*, respectively (Table 4; Additional file 4). Moreover, a comparative analysis of OC cluster organization among the three mycoplasma species revealed the presence of 12 OCs with complete similarity with respect to the ORF repertoires (synteny was not always detected). The majority of these OCs (seven OCs) were composed of two ORFs, with increasing numbers of up to five ORFs. These data are consistent with previous results that suggested that the majority of gene clusters in diverse organisms are formed by a string of two to four genes [30,31].

Another group of noteworthy OCs was the group classified as partially conserved among the three mycoplasma species. The number of partially conserved gene-order clusters in different genome pairs is shown in Table 4. The

Table 4 Comparison of OCs organization in *M. flocculare, M. hyopneumoniae* and *M. hyorhinis* genomes

ORF Clusters features	MFL x MHP*	MFL x MHR*	MFL x MHP x MHR*
OCs 100% conserved	44	17	12
OCs Partially conserved	44	35	33
OCs Without conservation	29	62	-

* Abbreviations as in Table 1.

M. flocculare and *M. hyopneumoniae* genomes shared 44 OCs in which the gene string was partially conserved, and 33 OCs of the 44 OCs were classified as partially conserved in all three analyzed species. It is well known that only a few operons are conserved in most bacterial genomes; the classical example of conserved organization involves the ribosomal protein operons [32]. However, a detailed analysis of the 33 partially conserved OCs revealed gene context conservation in the ribosomal operons (see $OC_{7448}28$ in Additional file 4) and in other clusters, such as clusters containing the chromosomal replication initiation protein (DnaA) (see $OC_{7448}01$ in Additional file 4) and the OC containing the cell division protein MraZ (see $OC_{7448}67$ in Additional file 4). In general, the similarity of the gene order (total or partial) among the prokaryotic genomes is maintained via the horizontal transfer of a chromosomal region. Our results suggest that individual genome pairs, such as *M. flocculare* and *M. hyopneumoniae* or *M. flocculare* and *M. hyorhinis*, share several OCs, which can partially be attributed to horizontal gene transfer.

A detailed, genomic-scale analysis of the OC organization in *M. flocculare* and *M. hyopneumoniae* demonstrated that species-specific differences are not present in genes with known function and/or related with pathogenicity (Additional file 6A-D). Apparently the 24 OCs exclusive to *M. hyopneumoniae* (not found in the *M. flocculare* genome) encode hypothetical proteins, transport-related proteins, myo-inositol utilization proteins, the integrative conjugative element (ICEH) and an additional copy of the P97 protein (Additional file 6A). In the *M. hyorhinis* genome, 35 OCs were unique to this species (Table 3; Additional file 6D), and the majority of the ORFs encode hypothetical proteins or products related to variable surface lipoproteins (*vlp* genes), which have been described as being involved in a complex system involving bacterial-host interactions [33].

Rearrangements in the *M. flocculare, M. hyopneumoniae* and *M. hyorhinis* genomes

A detailed comparative analysis of genome organization is needed to understand the evolutionary dynamics of prokaryotic genomes. Therefore, a comparative genomic analysis was performed using the *M. flocculare* contigs (MFL contigs) and *M. hyopneumoniae* genome, considering the ORF string organization and OC distribution (Additional file 7A-B). Comparisons were also performed between MFL contigs and the *M. hyorhinis* genome (Additional file 7A-B); however, in this case, both the global alignment and gene-by-gene alignment were not applicable, possibly due to the large number of transpositions and inversions that have occurred in these genomes.

In the comparison between *M. flocculare* and *M. hyopneumoniae*, 22 regions (with lengths ranging from 2 to 75 kb) were identified as being involved in inversions or rearrangements (Additional file 7B). Among these regions, only eight showed major rearrangements, although the OC organization was maintained. Notably, OCs containing several of the genes encoding pathogenicity-related proteins, such as lipoproteins and adhesins, were located within these regions. For instance, major rearrangements were observed in gene clusters encoding P97, P102 (MFL contig 13), P60, P69, P37 (MFL contig 34), P216, P76 (MFL contig 20) and the 46 K surface antigen precursor (MFL contig 4). Genes encoding transposases were found adjacent to some of the inverted segments, such as in MFL contig 23 (containing the P146 and MgPa proteins) and MFL contig 15 (containing the P97-like and P102-like adhesins), suggesting a possible role of these transposases in the rearrangements. Our findings in the comparison between the *M. flocculare* and *M. hyopneumoniae* genomes are similar to the situation found in the genomes of the two closely related species *Mycoplasma pneumoniae* and *Mycoplasma genitalium*, whose genomes can be divided into segments with highly conserved gene organization, although the segments are arranged differently [34].

The lack of gene-order conservation beyond the operon level even between relatively closely related species has been previously described [32]. Apparently, in closely related mycoplasmas, such as *M. flocculare* and *M. hyopneumoniae*, large-scale gene-order conservation is observed, although genome collinearity is disrupted at some points. Chromosomal rearrangements are generally caused by homologous recombination between repeated sequences within the genome [19,35]. Although the number of genes involved in DNA repair and recombination in mycoplasmas is relatively small [36], the gene encoding RecA was found in all sequenced mollicute genomes. Recently, the importance of RecA in the antigenic and phase variation of the MgpB and MgpC adhesins in *M. genitalium* has been demonstrated [37].

Repertoire of surface proteins encoded by *M. flocculare, M. hyopneumoniae* and *M. hyorhinis*

A comparative *in silico* survey of the repertoire of encoded surface proteins was performed between the genome of the non-pathogenic *M. flocculare* and the genomes of two pathogenic mycoplasma species that are found in this tissue in the swine respiratory tract, *M. hyopneumoniae* (represented by the 7448 strain) and *M. hyorhinis* (represented by the HUB-1 strain). The results of this survey are summarized in Additional files 8 and 9, and the complete generated datasets are presented in Additional files 10, 11 and 12. Of the total of 585 *M. flocculare* CDSs, 277 (47.5%) were predicted to encode surface proteins; this number was similar to that of *M. hyopneumoniae* 7448 (292 out of 716; 44.4%) and higher than that of *M. hyorhinis* HUB-1 (247 out of 654, 37.7%)

(Additional file 10A-C). The proportion of CDSs encoding surface proteins in these species is considerably large considering their small genome sizes. From these surface protein sets, 28 (10.1%), 42 (14.4%), and 44 (17.8%) CDSs are unique to *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis*, respectively, with respect to the other two species (Additional file 8; Additional file 11A-C).

The repertoire of *M. flocculare* surface proteins that is not shared with the other two species consists exclusively of hypothetical proteins; although those of *M. hyopneumoniae* and *M. hyorhinis* are predominantly composed of hypothetical proteins (73.8% and 59%, respectively), they also include some proteins with assigned functions. Among these proteins with predicted functions, *M. hyopneumoniae* includes proteins involved in myo-inositol catabolism, a permease and a protein encoded by the integrative conjugative element (ICEH), and *M. hyorhinis* includes some variable antigens, secretory system components, transporters and lipoproteins.

Based on an E-value cutoff threshold of 1e-6 to define orthology, nearly 90% of the repertoire of *M. flocculare* surface proteins is shared with *M. hyopneumoniae* and/or *M. hyorhinis* (Additional file 8; Additional file 12). Searches using more stringent conditions resulted in not more than a 20% reduction in the numbers of identified orthologs (data not shown). These results are indicative of physiological similarities that would be consistent with the adaptation to the same host environment. Of the shared proteins, approximately 40% have unknown functions (hypothetical proteins), while the other 60% consist of proteins with assigned functions in at least one of the compared species. Notably, many of these shared proteins (46 proteins, marked in bold in Additional file 12) correspond to putative pathogenicity-related genes in *M. hyopneumoniae* and/or *M. hyorhinis*; these proteins include several lipoproteins and adhesins that are thought to play a role in virulence despite of the non-pathogenic nature of *M. flocculare* and the pathogenicity differences between *M. hyopneumoniae* and *M. hyorhinis*. For instance, *M. flocculare* contains orthologs for the P97 copy 2 and for the P97-like adhesins of *M. hyopneumoniae*, although it lacks an ortholog for P97 copy 1. The genomic organization of the P97 copy 2 and P97-like ortholog CDSs are similar in both species with respect to gene clustering (Figure 2); this result suggests that P97 copy 2 and P97-like are ancestral P97 paralogs and that they were present in a common ancestor to *M. flocculare* and *M. hyopneumoniae*. A second duplication event, which originated the *M. hyopneumoniae* P97 copy 1, would have occurred after the divergence of *M. hyopneumoniae* from *M. flocculare*.

Additional file 13 lists some adhesins that have been associated with pathogenicity and have been experimentally analyzed [38-43] in *M. hyopneumoniae* 232. The *M. hyopneumoniae* 7448 and *M. flocculare* genomes contain orthologs for all the adhesins with the aforementioned exception of one copy of the P97 and P102 proteins that are absent in *M. flocculare* (also shown in Figure 2). The gene organization and location was analyzed and, as described in Additional file 13, the regions containing these orthologs are involved in inversions or rearrangements (Additional file 7B) in both mycoplasma species. Specifically, three important adhesins (P216, P159 and P60) display highly conserved gene organization between *M. hyopneumoniae* 232, 7422, 7448 and 168 strains, but they display inversions and rearrangements in *M. flocculare*. The participation of *M. hyopneumoniae* adhesins in host-cell adhesion is a complex process involving specific cleavage events [44]. The set of *M. flocculare* genes that may be involved in adhesion may not be complete, which would explain the differences in host-cell adhesion with respect to *M. hyopneumoniae* [18]. These results may explain the presence of orthologs in *M. flocculare* despite its lack of pathogenic capacity.

The presence of surface virulence determinants even in the non-pathogenic *M. flocculare* and in a non-pathogenic strain of *M. hyopneumoniae* (J strain) [21,45] suggests that their roles in pathogenicity may depend on

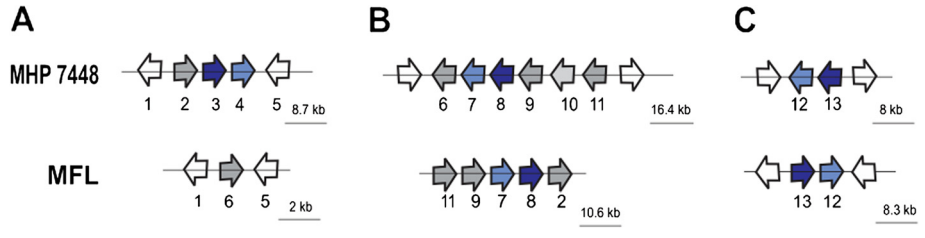


Figure 2 P97 and P102 gene organization contexts in the *M. flocculare* and *M. hyopneumoniae* 7448 genomes. (**A**) P97 copy 1 ORF cluster organization. (**B**) P97 copy 2 ORF cluster organization. **(C)** P97-like ORF cluster organization. The arrows represent the ORFs (not to scale) and indicate the transcriptional direction. The dark-blue arrows represent the P97 ORFs, and the light-blue arrows represent the P102 ORFs. The white arrows represent the ORFs that are at the limits of the OC. The numbers from one through thirteen represent the ORF name and the names of its orthologous as follows: 1- *rpsJ*; 2- MF1418 and MHP0197; 3- P97 copy 1; 4- P102 copy 1; 5- MF0249 and MHP0200; 6- MF0247 and MHP0106; 7- P102 copy 2; 8- P97 copy 2; 9- *gyrB*; 10- transposase; 11- *pfkA*; 12- P102-like; and 13- P97-like.

their expression levels and/or post-translational processing, which may vary [46]. Differences in virulence between species and strains can also be associated with the presence of variants of these proteins with or without some functional domains that are associated with features such as adhesion capacity or antigenicity, which has previously been described for several *M. hyopneumoniae* and *M. hyorhinis* virulence factors [40,41,47,48]. For instance, the *M. flocculare* P97 copy 2 and P97-like orthologs present relatively high overall identities to their *M. hyopneumoniae* counterparts (53% and 57%, respectively), but in the case of P97 copy 2, the *M. flocculare* ortholog lacks a domain (R1) regarded to be important for virulence in *M. hyopneumoniae*; instead, it contains a second R2 domain (Additional file 14). These R1 and R2 repeats are absent from the *M. flocculare* P97-like protein and its orthologs from *M. hyopneumoniae* (P97-like adhesin) and *M. hyorhinis* (P95).

Cell-surface features with implications for virulence may also reside in the 10 to 18% of the repertoires of surface proteins that are not shared between *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis*. However, because all (in the case of *M. flocculare*) or most (in the cases of *M. hyopneumoniae* and *M. hyorhinis*) of these unshared CDS products are hypothetical or conserved hypothetical proteins, their potential contributions to pathogenicity remain elusive. However, considering the nature of the unshared CDS products annotated in *M. hyopneumoniae* and *M. hyorhinis*, these 'exclusive' and unknown proteins are likely to include players of processes that are important for pathogen-host interactions, such as proteins involved in secretion, the uptake of certain molecules, conjugation and immune evasion/modulation. The variation in *M. hyorhinis* surface lipoproteins (Vlp) is considered important to protect the organism from the humoral response and may be a primary adaptive strategy for immune evasion during infection and disease [49,50]. Therefore, at least some of these proteins are expected to compose a portion of the repertoire of determinants of virulence or avirulence for each species or strain.

An additional comparison of the repertoires of surface proteins from *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis* was performed based on the COG classification of the predicted surface protein sets for each species (see Additional file 10). The produced COG functional profiles of surface proteins for the three species are summarized in Additional file 9. According to the COG, the functional surface protein profile for *M. flocculare* is similar to those for *M. hyopneumoniae* and *M. hyorhinis*; similar numbers of proteins were assigned to each category for the three species. This similarity was observed even for the U and M categories, which include secretion system components (whose repertoires are virtually equivalent for the three species; data not shown), and for the no-COG category, which included 45-53% of the proteins, most of which (82.2% for *M. flocculare*, 86.6% for *M. hyopneumoniae*, and 64.8% for *M. hyorhinis*) were represented by hypothetical proteins or in the additional category of antigen, adhesin or lipoprotein, in which proteins were included based on their prior immunological or functional characterization according to published studies. Overall, the surface protein set for *M. flocculare*, taken from the correspondent COG profile, was shown to be very similar to those of *M. hyopneumoniae* and *M. hyorhinis*. This result suggests that the species have equivalent genetic backgrounds for metabolic and growth processes. Such functional similarities may be the result of common selective pressures associated with the colonization of the same environment (i.e., the swine respiratory tract).

The L (replication, recombination and repair) and V (defense mechanisms) categories displayed differences; *M. flocculare* (and *M. hyorhinis*) contained approximately half of the number of proteins as *M. hyopneumoniae*. Protein sets assigned to the L category are heterogeneous, and the relative excess of proteins in *M. hyopneumoniae* corresponds to transposases that cannot be found in either *M. flocculare* or *M. hyorhinis*. Conversely, category V is enriched with ATP-binding cassette (ABC) transporter system proteins related to defense mechanisms, such as the *M. hyorhinis* ABC-type multidrug-like transport system ATP-binding proteins and their orthologs in *M. flocculare* and *M. hyopneumoniae*. The remaining transporters, including ABC and non-ABC transporter system components (such as those from the phosphotransferase system; PTS) appear in other COG categories, such as E, G, R or P. However, the overall number of transporters unrelated to defense mechanisms (non-V) is roughly equivalent in the three species, with 50, 49, and 47, in the *M. flocculare*, *M. hyopneumoniae* and *M. hyorhinis* surface protein sets, respectively. Although *M. hyorhinis* have 19 genes encoding transposases, and *M. hyopneumoniae* 9 genes, notably, the presence of transposases among *M. hyopneumoniae* predicted surface proteins may be an artifact due to the occurrence of helical structures in these enzymes [51]; these helices can be misidentified as transmembrane domains. However, the differential presence of at least some transporter system components is indicative of certain *M. hyopneumoniae* capabilities that are unavailable in both *M. flocculare* and *M. hyorhinis*. A larger number of transport proteins is usually related to a species' capacity to persist in different tissue environments [52], but this phenomenon does not seem to apply to *M. flocculare*, *M. hyopneumoniae* or *M. hyorhinis* because they share a large portion of the transporter repertoire. This situation is similar to that observed for secretory system components. However, according to the COG (see below), *M. hyopneumoniae* has approximately two-

fold more transporters associated with defense mechanisms than the other two species. In this aspect, *M. hyopneumoniae* is more similar to *Mycoplasma bovis* [53], which infects the respiratory tract and breast and joint tissues of bovines, than to *M. flocculare* or *M. hyorhinis*. The implications of the larger *M. hyopneumoniae* repertoire of defense mechanisms proteins (COG V) for its survival in the swine respiratory tract have not yet been investigated.

Phylogenomics and the phylogenetics of *Mycoplasmataceae*

From the entire set of 585 annotated *M. flocculare* genes, 179 gene sets were retrieved that contained at least one gene representative for each swine mycoplasma analyzed here (BLAST cut off E-10). Overall, 179 ortholog-like files representing different CDSs were concatenated, leading to an aligned file containing 104,097 amino acid residues. The neighbor-joining method (NJ) and maximum parsimony (MP) tree topologies did not differ significantly, especially when major clades were considered (Additional files 15, 16, 17; 18). There was consensus in several aspects (Figure 3). As expected, all of the *M. hyopneumoniae* strains formed a monophyletic clade. Additionally, the *M. hyopneumoniae* monophyletic clade was closely related to *M. flocculare*, with high bootstrap support. Finally, *M. hyorhinis* is basal to *Mycoplasma conjunctivae*, *M. flocculare* and *M. hyopneumoniae* in all of the phylogenomic trees.

The *Mycoplasmataceae* species were subdivided into the following clades: bovis (including *Mycoplasma agalactiae*, *M. bovis* and *Mycoplasma fermentans*), hominis (*Mycoplasma hominis* and *Mycoplasma arthritidis*), hyopneumoniae (*M. hyorhinis*, *M. conjunctivae*, *M. flocculare*, and *M. hyopneumoniae*), hemotrophic mycoplasma (*Mycoplasma suis* and *Mycoplasma haemofelis*), genitalium-pneumoniae

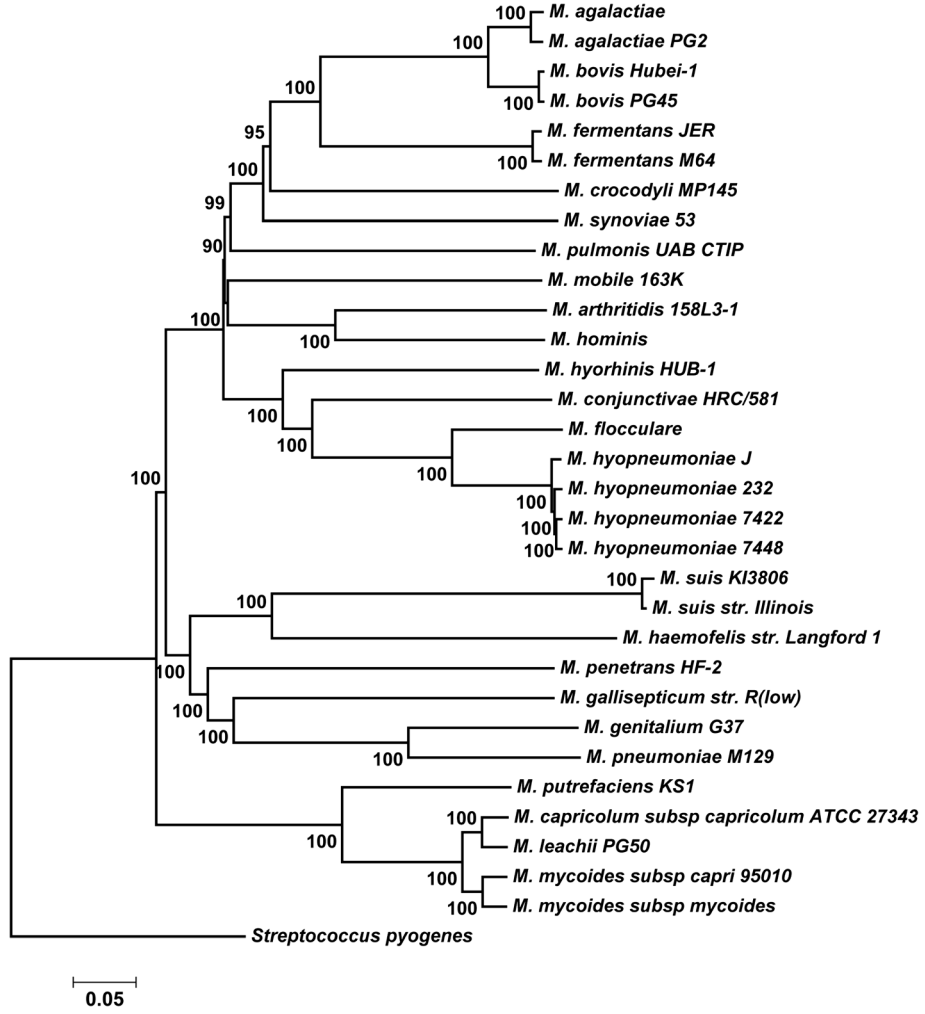


Figure 3 Evolutionary history of mycoplasmas obtained through a phylogenomic approach. The Neighbor-Joining method using the p-distance to compute the evolutionary distances and the pairwise deletion of gaps was implemented in the MEGA 5 software program. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to each branch. *Streptococcus pyogenes* was used as the outgroup.

(*Mycoplasma gallisepticum*, *Mycoplasma genitalium*, and *Mycoplasma pneumoniae*), and mycoides (*Mycoplasma putrefaciens*, *Mycoplasma capricolum*, *Mycoplasma leachii*, and *Mycoplasma mycoides*). All of these clades displayed high bootstrap values. The synoviae-pulmonis (*M. crocodyli*, *M. synoviae*, and *M. pulmonis*) group did not form a monophyletic cluster, but they are closely related to the bovis clade. *M. penetrans* HF-2 is near the genitalium-pneumoniae clade.

Our mycoplasma phylogenomic tree (Figure 3) corroborated the results that were obtained using the RNA polymerase beta subunit (rpoB), 16S-23S rRNA intergenic transcribed spacer region (ITS), and 16S rRNA genes [54,55]. *M. flocculare*, *M. hyopneumoniae* 7448, and *M. hyorhinis* HUB-1 were located in the hyopneumoniae clade.

When comparing the *M. flocculare*, *M. hyorhinis* HUB-1 and *M. hyopneumoniae* 7448 genomes, several paralog clusters were identified through the bidirectional best hit (BBH) approach, wherein a paralog cluster was defined as a gene set in which every gene is a BBH with at least one other element. Fourteen of these paralogs (DNA methylase, ATP synthase, ribulose-phosphate 3-epimerase, oligoendopeptidase F, single-strand binding protein, fructose-bisphosphate aldolase, dihydrolipoamide dehydrogenase, glucose-6-phosphate isomerase, lipoate-protein ligase, acyl-carrier-protein phosphodiesterase, lactate dehydrogenase, membrane nuclease lipoprotein, TrsE-like protein, and P97) were submitted to phylogenetic analyses to understand the evolutionary history of those paralogs.

Phylogenetic analyses of mycoplasma DNA methylases, which are enzymes that catalyze the transfer of a methyl group to DNA [56], contained ancient gene duplications in the hyopneumoniae group, leading each DNA methylase paralog to form a monophyletic group that included *M. flocculare* and *M. hyopneumoniae* (Additional file 19). Methylation, in addition to involvement in restriction systems, plays an important role in controlling gene expression, and it is one of the most significant DNA modifications [57]. The N^6-adenine methylation is involved in bacterial gene regulation and virulence [58-60]. CpG motifs in bacterial DNA may play a significant pathogenic role in inflammatory lung disease because the proinflammatory effects can be reduced by DNA methylation [61]. Microarray analyses and RT-PCR have demonstrated that the deletion of a C^5-cytosine methyltransferase in *Helicobacter pylori* strains can affect the expression of several genes related to motility, adhesion and virulence [57].

The ATP synthase phylogeny indicated that *M. flocculare*, *M. hyorhinis* HUB-1, and *M. hyopneumoniae* 7448 cluster according to the ATP synthase subunit, with the alpha subunit presenting a more complex evolutionary pattern (Additional file 20). This enzyme is required to synthesize adenosine triphosphate (ATP), providing energy to the cell. The paralogs found in mycoplasmas are related to different subunits that are required for enzymatic function [62].

Ribulose-phosphate 3-epimerase interconverts the stereoisomers ribulose-5-phosphate and xylulose-5-phosphate [63]. Its phylogeny revealed that ancient duplications occurred in the hyopneumoniae group. Other gene duplications responsible for the *M. hyorhinis* GDL-1 and *M. hyorhinis* HUB-1 split can be observed at the base of this *M. hyorhinis* clade (Additional file 21). Similarly, recent gene duplications can be observed in the *M. hyorhinis* clade in the oligoendopeptidase F (Additional file 22) and single-strand binding protein (Additional file 23) phylogenetic trees, which both contain high bootstrap support.

M. flocculare and *M. hyopneumoniae* 7448 contain two copies of fructose-bisphosphate aldolase, while *M. hyorhinis* HUB-1 contains only one copy. These duplications likely occurred prior to the diversification of the hyopneumoniae group (Additional file 24). The same process occurred during dihydrolipoamide dehydrogenase evolution, as shown in Additional file 25. *M. hyorhinis* HUB-1 contains two copies of the dimeric glycolytic enzyme glucose-6-phosphate isomerase, which catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate [64]. The lack of statistical confidence in some tree branches did not facilitate inferences regarding the evolutionary history of these copies in *M. hyorhinis* HUB-1 (Additional file 26).

Two lipoate-protein ligases are found in *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1. They are more closely related to the enzymes from other microorganisms than to each other (Additional file 27). In *M. flocculare*, only one copy was identified. The same pattern was observed in the phylogenetic tree of acyl-carrier-protein phosphodiesterase (Additional file 28), which belongs to the hydrolase family and acts on phosphoric diester bonds [65]. The topology of the lactate dehydrogenase tree showed that the two *M. hyorhinis* HUB-1 protein copies are significantly different, which led the sequences to be grouped in distant clades (Additional file 29). A recent duplication of membrane nuclease lipoprotein resulted in a *M. hyorhinis* cluster containing the two *M. hyorhinis* HUB-1 copies. A unique copy of this specific protein was found in *M. flocculare* and *M. hyopneumoniae* 7448 (Additional file 30).

Two copies of the TrsE -like protein were identified in *M. hyopneumoniae* 7448, and one copy was identified in the *M. flocculare* and *M. hyorhinis* HUB-1 genomes. Each copy is more similar to other sequences in the *M. hyopneumoniae* strain than to one another (Additional file 31). Finally, P97 is an adhesin thought to play a role in virulence. Several copies have been detected in the

hyopneumoniae group. The phylogenetic tree indicated that the copies of *M. flocculare* and *M. hyopneumoniae* 7448 are more closely related to the sequences from other species and strains than to each other (Additional file 32).

Although mycoplasmas contain reduced genomes, some paralogs are maintained in their genomes. A phylogenetic analysis was conducted to better understand the evolution of those paralogs. Gene copies are known to be preserved in a genome if the organism demands high levels of particular gene products. In other cases, positive selection can result in the diversification of the gene's function, a process called neofunctionalization [66]. Additionally, subfunctionalization can lead to the loss of function, resulting in duplicated genes whose functions differ to some degree [67]. Even highly conserved genes may have slightly or very different functions, such as glycolytic enzymes such as fructose-bisphosphate aldolase (FBA), which have been described as complex, multifunctional proteins that perform non-glycolytic functions [68].

Some mycoplasma paralog proteins, such as lipoate-protein ligase may possess different functions or differ in substrate specificity. Otherwise, essential enzymes, such as ATP synthase, may maintain multiple gene copies because they encode different subunits that are required for enzymatic function, despite the recent finds of losses of this enzyme family in the common ancestor of Mollicutes [69]. According to the standard model of phylogenomics, there is a higher similarity among orthologs than paralogs [67]. The paralog genes initially display identical sequences and functions. However, the action of selective pressures and mutations lead to divergence in regulatory and coding sequences [70].

Horizontal gene transfer

The species tree that was generated by the phylogenomic analysis was compared to the individual gene trees to investigate the occurrence of horizontal gene transfer (HGT) events. HGTs are an important source of genome innovation and evolution in prokaryotes [71], and it apparently also impacts Mycoplasmataceae evolution.

In mycoplasmas, we observed several HGT events (Additional file 33); some of the events occurred between *M. hyorhinis* HUB-1 and *M. conjunctivae* HRC. The events involving species belonging to the hyopneumoniae group occurred in ribosomal proteins, GTP-binding proteins, heat-shock proteins, DNA primase, signal recognition particle protein, ABC transporter ATP-binding proteins, phosphoesterases, cell division protein, elongation factor, fructose-biphosphate aldolase, DNA polymerase, glutamyl-tRNA synthetase, helicases, and hypothetical proteins. Regions encoding ABC transporters were likely transferred between *M. synoviae* and *M. gallisepticum* (Additional file 33), corroborating previously published results [21].

It is well-known that prokaryotes exchange genes in a sophisticated manner via lateral transfer, and bacterial phylogenies may also be viewed as a complex network of genomic exchange. However, the sequence-based methods implemented in the phylogenomic studies have yielded phylogenetic trees that are similar to rRNA trees, which were demonstrated in the current study. Consequently, lateral transfer events do not prevent the recovery of phylogenetic signals in prokaryotes, although they do add an extra source of noise [72].

Metabolism overview

Mycoplasmas contain a reduced genome; therefore, they lack many metabolic pathways, particularly biosynthetic pathways, such as those involved in cell-wall production, de novo purine biosynthesis and the biosynthesis of amino acids [73]. These organisms also lack a functional tricarboxylic acid (TCA) cycle because they are extremely fastidious in their nutritional requirements and dependent on nutrients supplied by their hosts. They produce high levels of enzymes responsible for the degradation of nucleic acids and proteins and transporters to obtain the precursors of these macromolecules. Most mycoplasma species depend on the glycolytic pathway to generate ATP. Some species may produce ATP based on the reaction involving acetyl phosphate and ADP by acetate kinase, coupled with acetyl phosphate formation from acetyl-CoA by phosphate acetyl transferase. Acetyl CoA is formed by the pyruvate dehydrogenase complex [1,74]. Some mycoplasma species such as *M. hominis* and *M. arthritidis* produce ATP through arginine degradation, using the arginine dihydrolase pathway [75]. This pathway is absent in the mycoplasma genomes analyzed in this work. All of the genes that encode enzymes of the glycolytic pathway exist in the three species; however, some differences in gene copy number were observed. Two copies of the genes encoding fructose-bisphosphate aldolase exist in *M. hyopneumoniae* and *M. flocculare*, two copies of the gene encoding d-ribulose-5-phosphate 3 epimerase are present in *M. hyopneumoniae*, and two copies of the gene encoding glucose-6-phosphate isomerase are present in *M. hyorhinis*. The possible influence of the gene copy number on the physiology of the species is unknown. However, some glycolytic enzymes have been described as virulence factors, exhibiting functions unrelated to glycolysis, such as adhesion to the host cells, may contribute to the pathogenesis of mycoplasmas infections [7,68,76-79]. The evolutionary aspects of these paralogs are discussed in the "Phylogenomics and phylogenetics of *Mycoplasmatacae*" section.

It has been shown that nuclease activities can be detected in Mollicutes and that these activities are primarily associated with the membrane and may be essential for growth and survival [80]. Genes encoding nucleases or putative

membrane-associated nucleases were found in the three analyzed *Mycoplasma* species. Two gene sets encoding membrane nucleases were observed in the three species. One set is represented by two copies of *mnu*A (MHR_0206 and MHR_0549) in the *M. hyorhinis* HUB-1 genome and a single copy in *M. hyopneumoniae* 7448 (MHP7448_0580), *M. hyopneumoniae* 7422 (MX03145) and *M. flocculare* (MF00420). Another set represented by two copies was observed in the genomes of the three species. The cell surface-exposed exonuclease (mhp379) from *M. hyopneumoniae* 232, a representative of the latter set, has been analyzed, and it has been proposed that the exonuclease activity of mhp379 may be important for importing nucleic acid precursors [81]. The presence of an extra copy of a nuclease gene in *M. hyorhinis* may represent a potential advantage of this species in acquiring nucleic acids precursors. Mycoplasmas cannot de novo synthesize purines and pyrimidines; therefore, they depend on salvage and interconversions to supply the cell with the nucleic acid precursors [82]. The three species contain a similar set of genes involved in purine (MHP7448_0084 - *M. hyopneumoniae*, MHR_0566 - *M. hyorhinis* and MF01198 - *M. flocculare*) and pyrimidine (MHP7448_0578 - *M. hyopneumoniae*, MHR_0640 - *M. hyorhinis* and MF00424 - *M. flocculare*) metabolism (See Additional file 4). However, *M. hyorhinis* contains genes encoding thymidylate synthetase (TS), allowing the conversion of dUMP to dTMP, and dihydrofolate reductase (DHFR), which catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF). The presence of TS and DHFR in *M. hyorhinis* may also contribute to its ability to overgrow the other *Mycoplasma* species in the swine respiratory tract [14].

Conclusions

The comparative analyses of three mycoplasma species that inhabit the swine respiratory tract facilitated the identification of some characteristics that may promote the understanding of their different behaviors. The *M. hyopneumoniae* strain 7422 genome displays a similar organization as to the other previously described strains, but it contains rearrangements and an altered position of ICEH in the genome. The genomes of *M. hyopneumoniae* and *M. flocculare*, two closely related species, contain some blocks of synteny, but they also display many differences that may help to explain why one species is pathogenic and the other is commensal. However, it was not possible to correlate specific virulence determinant factors to the pathogenicity differences of the analyzed species. A large proportion of the repertoire of *M. flocculare* surface proteins is shared with *M. hyopneumoniae* and/or *M. hyorhinis,* which would be expected because the organisms may occupy the same niche. However, certain members of the p97 family are absent in *M. flocculare*, and some display sequence differences or lack domains that are considered to be important for host-cell adhesion. *M. hyorhinis* contains some metabolic genes that are absent in the other species, suggesting a possible advantage in the growth of this species. The differences in some components involved in evasion of the host immune system may also contribute to the aggressive growth of *M. hyorhinis*. The phylogenomic analysis confirmed previous results, placing *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* in the hyopneumoniae clade. Several horizontal gene transfer events were identified, and several of them occurred between *M. hyorhinis* and *M. conjunctivae*.

Methods

Bacterial strains, culture conditions, and DNA isolation

M. hyopneumoniae strain 7422 was isolated from an infected swine in Lindóia do Sul, Santa Catarina, Brazil. *M. flocculare* (ATCC 27716) was acquired by Embrapa Suínos e Aves (Concórdia, Brazil) from the American Type Culture Collection. Both of the strains were cultivated in 5 mL of Friis medium [83] at 37°C for 48 h, and genomic DNA was extracted according to a standard protocol [84].

Genome sequencing, assembly and annotation

For each species, one 454 shotgun library was prepared with approximately 5 μg of gDNA. The library construction, titration, emulsion PCR and sequencing steps were performed explicitly according to the manufacturer's protocol. Sequencing was performed using Roche 454 GS FLX Titanium platform. *M. flocculare* was sequenced in one region of a two-region PicoTiterPlate (PTP), and *M. hyopneumoniae* 7422 was sequenced in two regions of an eight-region PTP. The contigs were assembled using the Newbler software program version 2.6 with the default parameters. The PCR assisted contig extension (PACE) strategy [85] was used for physical gap closure by the ends regions of gaps. For *M. hyopneumoniae* 7422 and *M. flocculare*, the estimated genome coverage for both genomes was 23X. The *M hyopneumoniae* 7422 genome was completely closed in one contig, and the *M. flocculare* genome retained 13 gaps.

The annotation and analysis of the sequences of both genomes were performed using the System for Automated Bacterial Integrated Annotation (SABIA) [86]. The comparative analysis was based on the Bidirectional Best Hits (BBH) [87] approach using the BLASTP (Basic Local Alignment Search Tool) [88] program to identify corresponding gene pairs recognized as the best hits in other genomes. All of the BLASTP searches were performed using the following parameters: an e-value of 10^{-5}, query coverage of 60%, and positive similarity value

of 50%. The comparative databank is available at http://www.genesul.lncc.br/comparative/.

To store and analyze the data, a databank was developed using the MySQL and Perl programming languages. This databank integrates tools and information from numerous biological databases, such as The Integrated Resource of Protein Domains and Functional Sites (Interpro) [89], Protein Subcellular Localization Prediction Tool (Psort) [90], Kyoto Encyclopedia of Genes and Genomes (KEGG) [29], Clusters of Orthologous Groups of Proteins (COG) [91], Transporter Classification Database (TCDB) [92], BLASTP of KEGG and UniProt/Swiss-Prot [93], facilitating several analyses, such as cluster with minimal genomes and clusters exclusives genes for each genome analyzed. In addition, the databank allows automatic genomic comparisons by bidirectional best hits (BBH) between five species. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession PRJNA65295 ID: 65295 for *M. flocculare* and PRJNA47327 ID: 47327 for *M. hyopneumoniae* 7422.

In silico analysis of ORF clusters (OCs)

The prediction of OCs was performed by the Artemis Release 10.5.2 software program [94] according to previously established criteria [30]. The manual examination of the possible OCs in the *M. flocculare, M. hyorhinis* and *M. hyopneumoniae* genomes was established based on the occurrence of clusters with two or more tandem genes in the same DNA strand. This procedure was performed by a systematic annotation comparison of the protein sequences encoded in all of the ORFs from the analyzed genomes. According to the complexity of the adjacent ORF rearrangements, two groups were created; the OC group was characterized by the presence of two or more ORFs in the same DNA strand until the occurrence of ORFs in the opposite strand, and the monocistronic (mC) group represented single ORFs. Differences in the annotation were evaluated by comparing the protein sequences from these three genomes using the NCBI/BLASTP program. The OC groups predicted for *M. flocculare* and *M. hyorhinis* were compared with the OC organization found in *M. hyopneumoniae* [30]. Moreover, comparative analyses were also performed between *M. hyorhinis* and *M. flocculare* to predict OC organization. The *M. flocculare* analysis was performed with each contig sequence, while the *M. hyorhinis* analysis was performed with the whole genome sequence.

Analysis of surface-protein-encoding CDSs

For the surface protein predictions, all of the translated CDS entries from the *M. flocculare* genome and from the *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1 genomes (NC_007332 and NC_014448 entries, respectively) were analyzed using the default parameters by the following software programs: SVMtm Transmembrane Domain Predictor [95], TMHMM Server v. 2.0 [96], SCAMPI [97], and PSORTb [98]. The first three programs predicted the presence of transmembrane (TM) domains, and the fourth was used for to predict subcellular localization (i.e., cytoplasmic vs. membrane). The first three independent TM predictions were merged, and the CDSs were considered as 'surface protein encoding' when they were predicted as such by at least two of the TM-predicting programs. The CDSs that were predicted to encode surface proteins by only one of the TM-predicting programs and/or by PSORTb were additionally analyzed by HMM-TM [99], and when a previously described ortholog was identified, based on the published literature. Those CDSs that were able to predict surface localization either by HMM-TM or based on the literature were also included in the list of predicted surface proteins encoded by each genome. After orthologs were identified by reciprocal pairwise comparisons between *M. flocculare*, *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1 (see below), the CDSs predicted to encode surface proteins just for one species (by the aforementioned criteria) had their corresponding orthologs (when existent) included in the list of surface-protein-encoding CDSs for the other two species. The clusters of orthologous group (COG) functional classification of *M. flocculare* predicted surface proteins using NCBI COGnitor [100], and those of *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1 predicted surface proteins obtained from NCBI.

For the ortholog and paralog identifications, bidirectional local BLAST searches [88] were performed between the *M. flocculare* predicted surface protein encoding CDSs and all of the CDSs from the other two species using the BioEdit Sequence Alignment Editor [101]. The ortholog and paralog definitions were based on the best hits from TBLASTX with the following parameters: cut-off e-value thresholds of 10^{-20}, 10^{-10} or 10^{-6}; manual inspection of query coverage, identity and similarity scores; and, if required, consideration of peculiarities of specific gene families.

Phylogenomic analyses

M. flocculare (ATCC 27716) was used as query organism. We began with all 582 genes identified in this strain and performed a BLAST search against the gene set of all other 32 organisms (Additional file 34). To select putative orthologs for all of the *M. flocculare* genes, we performed a BLAST search between this queried gene set and the individual genomes of the organisms previously cited; we retrieved the single best hit from each genome (BLAST cut-off 10). Multi-FASTA files were created containing ortholog gene sets for each query gene in the

M. flocculare strain. Only gene sets containing at least one representative in each genome were selected for further phylogenomic analysis (i.e., we only evaluated files containing 32 sequences for each analyzed genome).

Multi-FASTA putative ortholog files containing the best representative of each *M. flocculare* deduced protein sequence for all of the analyzed swine mycoplasmas were used as the input for multiple alignments using the CLUSTALw algorithm with default parameters.

The SCaFos software program [102] was used to facilitate the concatenation of the 179 alignment files. Phylogenies of 32 concatenated, deduced amino acid sequences were estimated by the NJ [103] and MP [104] methods in the Molecular Evolutionary Genetics Analysis (MEGA) program version 5.05 [105]. For the NJ method, the evolutionary distances were computed using the p-distance and the Poisson-corrected amino acids distance; both were presented in units of amino acid differences per site. The complete and pairwise deletion of gaps or missing data were implemented with the datasets containing 49,751 and 104,097 positions, respectively. The bootstrap test of the phylogeny was performed using 1,000 repetitions. The MP tree was obtained using the close-neighbor-interchange algorithm with search level 1, in which the initial trees were obtained by random addition of sequences (10 replicates). The different gap treatments were tested considering the complete deletion, partial deletion, and all sites included. The bootstrap test was implemented using 500 replicates. The TreeView software program [106] was used to visualize the resulting phylogenies.

Paralog analysis

Selected gene clusters of paralogs were subjected to a phylogenetic analysis. BLAST searches were first conducted for each gene using a 10^{-6} e-value cutoff; all of the sequences were subsequently aligned using COBALT [107]. Distance and parsimony methods in the MEGA 5 software program were applied to identify the evolutionary scenario for each paralog cluster. A bootstrap test was additionally performed with 1,000 replications.

Horizontal gene transfer

Horizontal gene transfer (HGT) was analyzed with Tree and Reticulogram reconstruction (T-Rex) [108] using the bipartition dissimilarity as the optimization criteria in the HGT detection algorithm [109]. The program infers an optimal (i.e., minimum-cost) scenario of horizontal gene transfers reconciling a given pair of species and gene trees. All of the gene trees were obtained through the distance method implemented in the MEGA 5 software program using p-distance, pairwise deletion of gaps, and bootstrap test of phylogeny with 500 replications. A bootstrap cutoff of 75% was applied to accept the HGT events. In total, 179 genes were subjected to phylogenetic and HGT analyses.

Additional files

Additional file 1: OCs organization in *M. flocculare* genome.

Additional file 2: OCs in *M. hyorhinis* genome.

Additional file 3: Monocistronic group (mC).

Additional file 4: OC organization similarities in the *M. hyorhinis* and *M. flocculare* genomes in relationship of *M. hyopneumoniae* OC organization.

Additional file 5: Monocistronic ORFs presents in the MHP, MHR and MFL genomes.

Additional file 6: A. Exclusive OCs of *M. hyopneumoniae* genome. 6B. Exclusive OCs of *M. flocculare* genome. 6C. Conserved Ocs among the three mycoplasmas species. 6D. Exclusive OCs of *M. hyorhinis* genome.

Additional file 7: A. *M. flocculare* contigs with conservation regions. B. *M. flocculare* contigs with inversion regions.

Additional file 8: Venn diagram of the predicted surface protein sets from *M. flocculare*, *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1.

Additional file 9: COG functional classifications of the predicted surface protein sets from *M. flocculare* (A), *M. hyopneumoniae* 7448 (B) and *M. hyorhinis* HUB-1 (C).

Additional file 10: A: List of predicted surface proteins of *M. flocculare* (A), *M. hyopneumoniae* 7448 (B), and *M. hyorhinis* HUB-1 (C). B: List of predicted surface proteins of *M. flocculare* (A), *M. hyopneumoniae* 7448 (B), and *M. hyorhinis* HUB-1 (C). List of predicted surface proteins of M. flocculare (A), *M. hyopneumoniae* 7448 (B), and *M. hyorhinis* HUB-1 (C).

Additional file 11: A. List of predicted surface proteins of *M. flocculare* not shared with *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB_1. B. List of predicted surface proteins of *M. hyopneumoniae* 7448 not shared with *M. flocculare* and *M. hyorhinis* HUB-1. C. List of predicted surface proteins of *M. hyorhinis* HUB-1 not shared with *M. flocculare* and *M. hyopneumoniae* 7448.

Additional file 12: List of ortholog surface protein coding CDSs between *M. flocculare*, *M. hyopneumoniae* 7448 and *M. hyorhinis* HUB-1.

Additional file 13: Summary of adhesins associated to pathogenicity and Genome organization comparison.

Additional file 14: *M. flocculare*, *M. hyopneumoniae* and *M. hyorhynis* P97 and P97-like adhesin orthologs and paralogs.

Additional file 15: Evolutionary history of mycoplasmas obtained through a phylogenomic approach. The Neighbor-Joining method, using p-distance to compute the evolutionary distances and complete deletion of gaps was implemented by MEGA 5 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,500 replicates) are shown next to the branches. *S. pyogenes* was used as outgroup.

Additional file 16: Evolutionary history of mycoplasmas obtained through a phylogenomic approach. The Neighbor-Joining method, using Poisson correction to compute the evolutionary distances and complete deletion of gaps was implemented by MEGA 5 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,500 replicates) are shown next to the branches. *S. pyogenes* was used as outgroup.

Additional file 17: Evolutionary history of mycoplasmas obtained through a phylogenomic approach. The Neighbor-Joining method was the same description of the Additional file 16.

Additional file 18: Evolutionary history of mycoplasmas obtained through a phylogenomic approach. The Maximum Parsimony method using the close-neighbor-interchange algorithm, and the complete

deletion of gaps was implemented by MEGA 5 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. *S. pyogenes* was used as outgroup.

Additional file 19: Evolutionary history of DNA methylases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method, using p-distance to compute the evolutionary distances and complete deletion of gaps was implemented by MEGA 5 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches.

Additional file 20: Evolutionary history of ATP synthases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 21: Evolutionary history of ribulose-phosphate -3epimerases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 22: Evolutionary history of oligoendopeptidases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 16. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches.

Additional file 23: Evolutionary history of single-strand binding proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 24: Evolutionary history of fructose-bisphosphate aldolase proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 25: Evolutionary history of dihydrolipoamide dehydrogenase proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 26: Evolutionary history of glucose-6-phosphate-isomerase proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 27: Evolutionary history of lypoate protein ligases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 28: Evolutionary history of acyl carrier phosphodiesterases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 29: Evolutionary history of lactate dehydrogenases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 30: Evolutionary history of membrane nucleases from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 31: Evolutionary history of TRSE-like proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 32: Evolutionary history of P97 proteins from mycoplasmas obtained through a phylogenetic analysis. The Neighbor-Joining method was the same description of the Additional file 19.

Additional file 33: HGT events in mycoplasmas species.

Additional file 34: Bacterial strains used in the phylogenomic analyses.

Abbreviations

CDS: Coding DNA sequence; COG: Clusters of orthologous groups; HGT: Horizontal gene transfer; ORF: Open read frames; OC: ORF cluster.

Competing interests

The authors declare they have no competing interests.

Authors' contributions

FMS performed the genomes organization analyses and comparative analyses of genomes, participated in the interpretation of the results and in the writing of the manuscript. CET, FP and MMBF carried out the phylogenomics and phylogenetics analyses and in the interpretation of the results. LGPA and RS carried out the assemblies genomes and participated in the comparative analyses of genomes. VGV, LR and TG participated in the surface proteins *in silico* analyses. ISS, HBF, ATV and AZ conceived, designed and coordinated the study, participated in the interpretation of the results and in the writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). FMS, LR, MMBF and VGV received scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).
We thank the research groups from the Southern Network for Genome Analysis that participated in the initial efforts to isolate and sequence *M. hyopneumoniae* genomes.

Author details

[1]Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil. [2]Laboratório de Bioinformática. Laboratório Nacional de Computação Científica. Petrópolis, Rio de Janeiro, Brazil. [3]Programa de Pós-Graduação em Biologia Celular e Molecular. Centro de Biotecnologia UFRGS, Porto Alegre, Brazil. [4]Programa de Pós-Graduação em Ciências Biológicas - Bioquímica. UFRGS, Porto Alegre, Brazil. [5]Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências. UFRGS, Porto Alegre, Brazil. [6]Departamento de Bioquímica Médica. Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Received: 8 October 2012 Accepted: 8 March 2013
Published: 14 March 2013

References

1. Razin S, Yogev D, Naot Y: **Molecular biology and pathogenicity of mycoplasmas.** *Microbiol Mol Biol Rev* 1998, **62:**1094–1156.
2. Rottem S: **Interaction of mycoplasmas with host cells.** *Physiol Rev* 2003, **83:**417–432.
3. Zhang S, Lo SC: **Effect of mycoplasmas on apoptosis of 32D cells is species-dependent.** *Current Microbiol* 2007, **54:**388–395.
4. Vogl G, Plaickner A, Szathmary S, Stipkovits L, Rosengarten R, Szostak MP: ***Mycoplasma gallisepticum* invades chicken erythrocytes during infection.** *Infect Immun* 2008, **76:**71–77.
5. Dusanic D, Bercic RL, Cizelj I, Salmic S, Narat M, Bencina D: ***Mycoplasma synoviae* invades non-phagocytic chicken cells in vitro.** *Vet Microbiol* 2009, **138:**114–119.
6. Groebel K, Hoelzle K, Wittenbrink MM, Ziegler U, Hoelzle LE: ***Mycoplasma suis* invades porcine erythrocytes.** *Infect Immun* 2009, **77**(suppl 2):576–584.
7. Song Z, Li Y, Liu Y, Xin J, Zou X, Sun W: **α-enolase, an adhesion-related factor of *Mycoplasma bovis*.** *PLoS One* 2012, **7**(suppl 6):e38836.
8. Hopfe M, Deenen R, Degrandi D, Köhrer K, Henrich B: **Host cell responses to persistent mycoplasmas - different stages in infection of HeLa cells with *Mycoplasma hominis*.** *PLoS One* 2013, **8**(suppl 1):e54219.
9. Mare CJ, Switzer WP: ***Mycoplasma hyopenumoniae*, a causative agent of virus pig pneumonia.** *Vet Med* 1965, **60:**841–846.
10. Meyling A, Friis NF: **Serological identification of a new porcine mycoplasma species, *Mycoplasma flocculare*.** *Acta Vet Scand* 1972, **13:**287–289.
11. Rose DL, Tully JG, Wittler RG: **Taxonomy of some swine mycoplasmas: *Mycoplasma suipneumoniae* Goodwin et al. 1965, A later, objective**

synonym of *Mycoplasma hyopneumoniae* mare and Switzer 1965, and the status of *Mycoplasma flocculare* meyling and friis 1972. *Int J Syst Evol Microbiol* 1979, **29:**83–91.
12. Stemke GW, Laigret F, Grau O, Bové JM: **Phylogenetic relationships of three porcine mycoplasmas, *mycoplasma hyopneumoniae*, *mycoplasma flocculare* and *mycoplasma hyorhinis*, and complete 16S rRNA sequence of *M. Flocculare*.** *Int J Syst Bacteriol* 1992, **42:**220–225.
13. Friis NF, Feenstra AA: ***Mycoplasma hyorhinis* in the etiology of serositis among piglets.** *Acta Vet Scand* 1994, **35:**93–98.
14. Kobisch M, Friis NF: **Swine mycoplasmoses.** *Rev Sci Tech Oie* 1996, **15:**1569–1605.
15. Young TF, Thacker EL, Erickson BZ, Ross RF: **A tissue culture system to study respiratory ciliary epithelial adherence of selected swine mycoplasmas.** *Vet Microbiol* 2000, **71:**269–279.
16. Boye M, Jensen TK, Ahrens P, Hagedorn-Olsen T, Friis NF: **In situ hybridisation for identification and differentiation of *Mycoplasma hyopneumoniae*, *Mycoplasma hyosynoviae* and *Mycoplasma hyorhinis* in formalin-fixed porcine tissue sections.** *Acta Path Micro Im* 2001, **109:**656–664.
17. Huang S, Li JY, Wu J, Meng L, Shou CC: **Mycoplasma infections and different human carcinomas.** *World Gastroentero* 2001, **7:**266–269.
18. Yang H, Qu L, Ma H, Chen L, Liu W, Liu C, Meng L, Wu J, Shou C: ***Mycoplasma hyorhinis* infection in gastric carcinoma and its effects on the malignant phenotypes of gastric cancer cells.** *BMC Gastroenterol* 2010, **10:**132.
19. Blank WA, Stemke GW: **A physical and genetic map of the *Mycoplasma flocculare* ATCC 27716 chromosome reveals large genomic inversions when compared with that of *Mycoplasma hyopneumoniae* strain J.** *Int J Syst Evol Micr* 2001, **51:**1395–1399.
20. Minion FC, Lefkowitz EL, Madsen ML, Cleary BJ, Swartzell SM, Mahairas GG: **The genome sequence of strain 232, the agent of swine mycoplasmosis.** *J Bacteriol* 2004, **186:**7123–7133.
21. Vasconcelos ATR, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LGP, Almeida R, Alves-Filho L, Assunção EN, Azevedo VAC, Bogo MR, Brigido MM, Brocchi M, Burity HA, Camargo AA, Camargo SS, Carepo MS, Carraro DM, de Mattos Cascardo JC, Castro LA, Cavalcanti G, Chemale G, Collevatti RG, Cunha CW, Dallagiovanna B, Dambrós BP, Dellagostin OA, Falcão C, *et al*: **Swine and poultry pathogens: the complete genome sequences of two strains of *mycoplasma hyopneumoniae* and a strain of *mycoplasma synoviae*.** *J Bacteriol* 2005, **187:**5568–5577.
22. Liu W, Fang L, Li S, Li Q, Zhou Z, Feng Z, Luo R, Shao G, Wang L, Chen H, Xiao S: **Complete genome sequence of *Mycoplasma hyorhinis* strain HUB-1.** *J Bacteriol* 2010, **192**(suppl 21):5844–5845.
23. Liu W, Feng Z, Fang L, Zhou Z, Li Q, Li S, Luo R, Wang L, Chen H, Shao G, Xiao S: **Complete genome sequence of *Mycoplasma hyopneumoniae* strain 168.** *J Bacteriol* 2011, **193**(suppl 4):1016–1017.
24. Guimaraes AM, Santos AP, SanMiguel P, Walter T, Timenetsky J, Messick JB: **Complete genome sequence of *Mycoplasma suis* and insights into its biology and adaption to an erythrocyte niche.** *PLoS One* 2011, **6**(suppl 5): e19574.
25. Oehlerking J, Kube M, Felder KM, Matter D, Wittenbrink MM, Schwarzenbach S, Kramer MM, Hoelzle K, Hoelzle LE: **Complete genome sequence of the hemotrophic *Mycoplasma suis* strain KI3806.** *J Bacteriol* 2011, **193**(suppl 9):2369–2370.
26. Kornspan JD, Lysnyansky I, Kahan T, Herrmann R, Rottem S, Nir-Paz R: **Genome analysis of a *Mycoplasma hyorhinis* strain derived from a primary human melanoma cell line.** *J Bacteriol* 2011, **193**(suppl17):4543–4544.
27. Calcutt MJ, Foecking MF, Rosales RS, Ellis RJ, Nicholas RA: **Genome sequence of *Mycoplasma hyorhinis* strain GDL-1.** *J Bacteriol* 2012, **194** (suppl 7):1848.
28. Goodison S, Urquidi V, Kumar D, Reyes L, Rosser CJ: **Complete genome sequence of *Mycoplasma hyorhinis* strain SK76.** *Genome Announc* 2013, **1** (suppl 1):e00101–e00112.
29. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28:**27–30.
30. Siqueira FM, Schrank A, Schrank IS: ***Mycoplasma hyopneumoniae* transcription unit organization: genome survey and prediction.** *DNA Res* 2011, **18:**413–422.
31. Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV: **Connected gene neighborhoods in prokaryotic genomes.** *Nucleic Acids Res* 2002, **30:**2212–2223.
32. Lathe WC, Snel B, Bork P: **Gene context conservation of a higher order than operons.** *Trends Biochem Sci* 2000, **25**(suppl 10):474–479.
33. Rosengarten R, Wise KS: **Phenotypic switching in mycoplasmas: phase variation of diverse surface lipoproteins.** *Science* 1990, **247:**315–318.
34. Himmelreich R, Plagens H, Hilbert H, Reiner B, Herrmann R: **Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*.** *Nucleic Acids Res* 1997, **25:**701–712.
35. Mahan MJ, Roth JR: **Role of recBC function in formation of chromosomal rearrangements: a Two-step model for recombination.** *Genetics* 1989, **121:**433–443.
36. Carvalho FM, Fonseca MM, Medeiros SB, Scortecci KC, Blaha CAG, Agnez-Lima LF: **DNA repair in reduced genome: the Mycoplasma model.** *Gene* 2005, **360**(suppl 2):111–119.
37. Burgos R, Wood GE, Young L, Glass JI, Totten PA: **RecA mediates MgpB and MgpC phase and antigenic variation in *Mycoplasma genitalium*, but plays a minor role in DNA repair.** *Mol Microbiol* 2012, **85**(suppl 4):669–683.
38. Burnett TA, Dinkla K, Rohde M, Chhatwal GS, Uphoff C, Srivastava M, Cordwell SJ, Geary S, Liao X, Minion FC, Walker MJ, Djordjevic SP: **P159 is a proteolytically processed, surface adhesin of *Mycoplasma hyopneumoniae*: defined domains of P159 bind heparin and promote adherence to eukaryote cells.** *Mol Microbiol* 2006, **60:**669–686.
39. Wilton J, Jenkins C, Cordwell SJ, Falconer L, Minion FC, Oneal DC, Djordjevic MA, Connolly A, Barchia I, Walker MJ, Djordjevic SP: **Mhp493 (P216) is a proteolytically processed, cilium and heparin binding protein of *Mycoplasma hyopneumoniae*.** *Mol Microbiol* 2009, **71:**566–582.
40. Seymour LM, Deutscher AT, Jenkins C, Kuit TA, Falconer L, Minion FC, Crossett B, Padula M, Dixon NE, Djordjevic SP, Walker MJ: **A processed multidomain *mycoplasma hyopneumoniae* adhesin binds fibronectin, plasminogen, and swine respiratory cilia.** *J Biol Chem* 2010, **285:**33971–33978.
41. Deutscher AT, Jenkins C, Minion FC, Seymour LM, Padula MP, Dixon NE, Walker MJ, Djordjevic SP: **Repeat regions R1 and R2 in the P97 paralogue Mhp271 of Mycoplasma hyopneumoniae bind heparin, fibronectin and porcine cilia.** *Mol Microbiol* 2010, **78:**444–458.
42. Seymour LM, Falconer L, Deutscher AT, Minion FC, Padula MP, Dixon NE, Djordjevic SP, Walker MJ: **Mhp107 Is a member of the multifunctional adhesin family of *Mycoplasma hyopneumoniae*.** *J Biol Chem* 2011, **286** (suppl 12):10097–10104.
43. Seymour LM, Jenkins C, Deutscher AT, Raymond BBA, Padula MP, Tacchi JL, Bogema DR, Eamens GJ, Woolley LK, Dixon NE, Walker MJ, Djordjevic SP: **Mhp182 (P102) binds fibronectin and contributes to the recruitment of plasmin(ogen) to the Mycoplasma hyopneumoniae cell surface.** *Cell Microbiol* 2012, **14**(suppl 1):81–94.
44. Bogema DR, Deutscher AT, Woolley LK, Seymour LM, Raymond BB, Tacchi JL, Padula MP, Dixon NE, Minion FC, Jenkins C, Walker MJ, Djordjevic SP: **Characterization of cleavage events in the multifunctional cilium adhesin Mhp684 (P146) reveals a mechanism by which *Mycoplasma hyopneumoniae* regulates surface topography.** *MBio* 2012, **3**(suppl 2):e00282–11.
45. Zielinski GC, Ross RF: **Effect on growth in cell cultures and strain on virulence of *Mycoplasma hyopneumoniae* for swine.** *Am J Vet Res* 1990, **51:**344–348.
46. Pinto P, Klein C, Zaha A, Ferreira H: **Comparative proteomic analysis of pathogenic and non-pathogenic strains from the swine pathogen *Mycoplasma hyopneumoniae*.** *Proteome Science* 2009, **7:**45.
47. Castro LAD, Pedroso TR, Kuchiishi SS, Ramenzoni M, Kich JD, Zaha A, Vainstein MH, Ferreira HB: **Variable number of tandem aminoacid repeats in adhesion-related CDS products in *Mycoplasma hyopneumoniae* strains.** *Vet Microbiol* 2006, **116:**258–269.
48. Yogev D, Watson-McKown R, Rosengarten R, Im J, Wise KS: **Increased structural and combinatorial diversity in an extended family of genes encoding Vlp surface proteins of *Mycoplasma hyorhinis*.** *J Bacteriol* 1995, **177:**5636–5643.
49. Citti C, Kim MF, Wise KS: **Elongated versions of Vlp surface lipoproteins protect *Mycoplasma hyorhinis* escape variants from growth-inhibiting host antibodies.** *Infect Immun* 1997, **65**(suppl 5):1773–1785.
50. Citti C, Nouvel LX, Baranowski E: **Phase and antigenic variation in mycoplasmas.** *Future Microbiol* 2010, **5**(suppl 7):1073–1085.
51. Hickman AB, Chandler M, Dyda F: **Integrating prokaryotes and eukaryotes: DNA transposases in light of structure.** *Crit Rev Biochem Mol Biol* 2010, **45:**50–69.
52. Westberg J, Persson A, Holmberg A, Goesmann A, Lundeberg J, Johansson K-E, Pettersson B, Uhlén M: **The genome sequence of *mycoplasma***

mycoides subsp. *Mycoides* SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res* 2004, **14:**221–227.
53. Li Y, Zheng H, Liu Y, Jiang Y, Xin J, Chen W, Song Z: **The complete genome sequence of *mycoplasma bovis* strain hubei-1.** *PLoS One* 2011, **6:**e20999.
54. Peters IR, Helps CR, McAuliffe L, Neimark H, Lappin MR, Gruffydd-Jones TJ, Day MJ, Hoelzle LE, Willi B, Meli M, Hofmann-Lehmann R, Tasker S: **RNase P RNA gene (rnpB) phylogeny of Hemoplasmas and other *Mycoplasma* species.** *J Clin Microbiol* 2008, **46:**1873–1877.
55. Volokhov DV, Simonyan V, Davidson MK, Chizhikov VE: **RNA polymerase beta subunit (rpoB) gene and the 16S–23S rRNA intergenic transcribed spacer region (ITS) as complementary molecular markers in addition to the 16S rRNA gene for phylogenetic analysis and identification of the species of the family Mycoplasmataceae.** *Mol Phylogen Evol* 2011, **62:**515–528.
56. Marinus MG, Casadesus J: **Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more.** *FEMS Microbiol Rev* 2009, **33:**488–503.
57. Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN: **Comparative transcriptomics of H. pylori strains AM5, SS1, and their hpyAVIBM deletion mutants: possible roles of cytosine methylation.** *PLoS One* 2012, **7:**e42303.
58. Reisenauer A, Kahng LS, McCollum S, Shapiro L: **Bacterial DNA methylation: a cell cycle regulator.** *J Bacteriol* 1999, **181:**5135–5139.
59. Wion D, Casadesus J: **N6-methyl-adenine: an epigenetic signal for DNA-protein interactions.** *Nat Rev Microbiol* 2006, **4:**183–192.
60. Fälker S, Schmidt MA, Heusipp G: **DNA adenine methylation and bacterial pathogenesis.** *Int J Med Microbiol* 2007, **297:**1–7.
61. Schwartz D, Quinn TJ, Thorne PS, Sayeed S, Yi A-K, Krieg AM: **CpG motifs in bacterial DNA cause inflammation in the lower respiratory tract.** *J Clin Invest* 1997, **100:**68–73.
62. Butler JE, Young ND, Lovley DR: **Evolution of electron transfer out of the cell: comparative genomics of six *Geobacter* genomes.** *BMC Genomics* 2010, **11:**40.
63. Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, Keeling PJ: **A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer.** *BMC Evol Biol* 2007, **7:**89.
64. Grauvogel C, Brinkmann H, Petersen J: **Evolution of the glucose-6 -phosphate isomerase: the plasticity of primary metabolism in photosynthetic eukaryotes.** *Mol Biol Evol* 2007, **24:**1611–1621.
65. Thomas J, Cronan JE: **The enigmatic acyl carrier protein phosphodiesterase of *Escherichia coli*.** *J Biol Chem* 2005, **280:**34675–34683.
66. Li WH, Yang J, Gu X: **Expression divergence between duplicate genes.** *Trends Genet* 2005, **21:**602–607.
67. Thomas PD, Wood V, Mungall CJ, Lewis SE, Blake JA: **On the use of gene ontology annotations to assess functional similarity among orthologs and paralogs: a short report.** *PLoS Comput Biol* 2012, **8:**e1002386.
68. Henderson B, Martin A: **Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins Are virulence determinants in infectious disease.** *Infect Immun* 2011, **79**(suppl 9):3476–3491.
69. Chen LL, Chung WC, Lin CP, Kuo CH: **Comparative analysis of gene content evolution in phytoplasmas and mycoplasmas.** *PLoS One* 2012, **7**(suppl 3):e34407.
70. Xu G, Guo C, Shan H, Kong H: **Divergence of duplicate genes in exon-intron structure.** *Proc Natl Acad Sci* 2012, **109:**1187–1192.
71. Ochmann H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation.** *Nature* 2000, **405:**299–304.
72. Philippe H, Douady CJ: **Horizontal gene transfer and phylogenetics.** *Curr Opin Microbiol* 2003, **6:**498–505.
73. Pollack JD: **The necessity of combining genomic and enzymatic data to infer metabolic function and pathways in the smallest bacteria: amino acid, purine and pyrimidine metabolism in Mollicutes.** *Front Biosci* 2002, **7:**d1762–d1781.
74. Matic JN, Wilton JL, Towers RJ, Scarman AL, Minion FC, Walker MJ, Djordjevic SP: **The pyruvate dehydrogenase complex of *Mycoplasma hyopneumoniae* contains a novel lipoyl domain arrangement.** *Gene* 2003, **319:**99–106.
75. Pereyre S, Sirand-Pugnet P, Beven L, Charron A, Renaudin H, Barré A, Avenaud P, Jacob D, Couloux A, Barbe V, de Daruvar A, Blanchard A, Bébéar C: **Life on arginine for *Mycoplasma hominis*: clues from its minimal genome and comparison with other human urogenital mycoplasmas.** *PLoS Genet* 2009, **5**(suppl 10):e1000677.
76. Alvarez RA, Blaylock MW, Baseman JB: **Surface localized glyceraldehyde-3 -phosphate dehydrogenase of *Mycoplasma genitalium* binds mucin.** *Mol Microbiol* 2003, **48:**1417–1425.
77. Hoelzle K, Grimm J, Ritzmann M, Heinritzi K, Torgerson P, Hamburger A, Wittenbrink MM, Hoelzle LE: **Detection of antibodies against *Mycoplasma suis* using recombinant antigens and correlation of serological results to hematological findings.** *Clin Vaccine Immunol* 2007, **14**(suppl 12):1616–1622.
78. Dumke R, Hausner M, Jacobs E: **Role of *Mycoplasma pneumoniae* glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in mediating interactions with the human extracellular matrix.** *Microbiology* 2011, **157:**2328–2338.
79. Schreiner AS, Sokoli A, Felder KM, Wittenbrink MM, Schwarzenbach S, Guhl B, Hoelzle K, Hoelzle LE: **The surface-localised a-enolase of *Mycoplasma suis* is an adhesion protein.** *Vet Microbiol* 2012, **156:**88–95.
80. Minion FC, Jarvill-Taylor KJ, Billings DE, Tigges E: **Membrane-associated nuclease activities in mycoplasmas.** *J Bacteriol* 1993, **175:**7842–7847.
81. Schmidt JA, Browning GF, Markham PF: ***Mycoplasma hyopneumoniae* mhp379 is a Ca2 + −dependent, sugar-nonspecific exonuclease exposed on the cell surface.** *J Bacteriol* 2007, **189:**3414–3424.
82. Bizarro CV, Schuck DC: **Purine and pyrimidine nucleotide metabolism in Mollicutes.** *Genet Mol Biol* 2007, **30:**190–201.
83. Friis NF: **Some recommendations concerning primary isolation of *Mycoplasma suipneumoniae* and *Mycoplasma flocculare* a survey.** *Nordisk Vet Medicin* 1975, **27:**337–339.
84. Sambrook J, Russell DW: *Molecular Cloning a Laboratory Manual.* New York: Cold Spring Harbor Laboratory Press; 2001.
85. Carraro DM, Camargo AA, Salim AC, Grivet M, Vasconcelos AT, Simpson AJ: **PCR-assisted contig extension: stepwise strategy for bacterial genome closure.** *Biotechniques* 2003, **34:**626–632.
86. Almeida LGP, Paixão R, Souza RC, da Costa GC, Barrientos FJA, dos Santos MT, de Almeida DF, Vasconcelos ATR: **A system for automated bacterial (genome) integrated annotation-SABIA.** *Bioinformatics* 2004, **20:**2832–2833.
87. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N: **The use of gene clusters to infer functional coupling.** *Proc Natl Acad Sci* 1999, **96:**2896–2901.
88. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389–3402.
89. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MDR, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJA, Zdobnov EM: **The InterPro database, an integrated documentation resource for protein families, domains and functional sites.** *Nucleic Acids Res* 2001, **29:**37–40.
90. Gardy JL, Spencer C, Wang K, Ester M, Tusnády GE, Simon I, Hua S, deFays K, Lambert C, Nakai K, Brinkman FSL: **PSORT-B: improving protein subcellular localization prediction for gram-negative bacteria.** *Nucleic Acids Res* 2003, **31:**3613–3617.
91. Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, Koonin E, Krylov D, Mazumder R, Mekhedov S, Nikolskaya A, Rao BS, Smirnov S, Sverdlov A, Vasudevan S, Wolf Y, Yin J, Natale D: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4:**41.
92. Saier MH, Tran CV, Barabote RD: **TCDB: the transporter classification database for membrane transport protein analyses and information.** *Nucleic Acids Res* 2006, **34**(suppl 1):D181–D186.
93. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL: **The universal protein resource (UniProt).** *Nucleic Acids Res* 2005, **33:**D154–D159.
94. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16:**944–945.
95. Yuan Z, Mattick JS, Teasdale RD: **SVMtm: support vector machines to predict transmembrane segments.** *J Comput Chem* 2004, **25:**632–636.
96. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL: **Predicting transmembrane protein topology with a hidden markov model: application to complete genomes.** *J Mol Biol* 2001, **305:**567–580.
97. Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, Elofsson A: **Prediction of membrane-protein topology from first principles.** *Proc Natl Acad Sci* 2008, **105:**7177–7181.

98. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL: **PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes.** *Bioinformatics* 2010, **26:**1608–1615.
99. Bagos P, Liakopoulos T, Hamodrakas S: **Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins.** *BMC Bioinformatics* 2006, **7:**189.
100. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res* 2001, **29:**22–28.
101. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41:**95–98.
102. Roure B, Rodriguez-Ezpeleta N, Philippe H: **SCaFoS: a tool for Selection.** *Concatenation and Fusion of Sequences for phylogenomics. BMC Evol Biol* 2007, **7:**S2.
103. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4:**406–425.
104. Lake JA: **A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony.** *Mol Biol Evol* 1987, **4:**167–191.
105. Kumar S, Nei M, Dudley J, Tamura K: **MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences.** *Brief Bioinform* 2008, **9:**299–306.
106. Page RDM: **Visualizing Phylogenetic Trees Using TreeView**. In *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc; 2002.
107. Papadopoulos JS, Agarwala R: **COBALT: constraint-based alignment tool for multiple protein sequences.** *Bioinformatics* 2007, **23:**1073–1079.
108. Makarenkov V: **T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks.** *Bioinformatics* 2001, **17:**664–668.
109. Boc A, Philippe H, Makarenkov V: **Inferring and validating horizontal gene transfer events using bipartition dissimilarity.** *Syst Biol* 2010, **59:**195–211.

doi:10.1186/1471-2164-14-175
Cite this article as: Siqueira *et al.*: New insights on the biology of swine respiratory tract mycoplasmas from a comparative genome analysis. *BMC Genomics* 2013 **14**:175.

4. CAPÍTULO 2

Comparative transcriptome analysis of mycoplasmas of the swine respiratory tract

Franciele Maboni Siqueira, Alexandra Berger, Luiz Gonzaga Almeida, Irene Silveira Schrank, Ana Tereza Vasconcelos, Arnaldo Zaha.

Manuscrito em preparação.

A ser submetido à Revista: *Genome Research*

OBS: O material suplementar, referente aos resultados apresentados neste manuscrito, está contido no item *Anexos* desta tese.

Comparative transcriptome analysis of mycoplasmas of the swine respiratory tract

Franciele Maboni Siqueira[1,2], Alexandra Berger[3], Luiz Gonzaga Almeida[3], Irene Silveira Schrank[2,4], Ana Tereza Vasconcelos[3], Arnaldo Zaha[2,4,*]

[1] Programa de Pós-Graduação em Ciências Biológicas - Bioquímica. Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

[2] Centro de Biotecnologia, UFRGS, Porto Alegre, Brazil.

[3].Laboratório de Bioinformática. Laboratório Nacional de Computação Científica. Petrópolis, Rio de Janeiro, Brazil.

[4] Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências. UFRGS, Porto Alegre, Brazil.

[*] Corresponding author

Dr. Arnaldo Zaha. Universidade Federal do Rio Grande do Sul. Centro de Biotecnologia. Av. Bento Gonçalves, 9500. Prédio 43421. Sala 206, Caixa Postal: 15005. CEP: 91501-970 – Porto Alegre, Rio Grande do Sul, Brasil. Tel. +55 51 3308 6070. Fax. +55 51 3308 7309.

Email: zaha@cbiot.ufrgs.br

Running Title: Comparative transcriptome of mycoplasmas

Abstract

The swine respiratory ciliary epithelium may be colonized mainly by three mycoplasma species: *Mycoplasma hyopneumoniae*, *Mycoplasma flocculare* and *Mycoplasma hyorhinis*. While *M. flocculare* presence is virtually assymptomatic, *M. hyopneumoniae* and *M. hyorhynis* infections may cause respiratory disease. Transcript structure and information on transcript abundance provide valuable insights into gene function and regulation, but none of them have been analyzed on a genome-wide scale for these *Mycoplasma* species. In this study, we report a transcriptome map of *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* using high throughput Roche/454 GS-FLX platform, generating informations for comparative studies on the transcriptional repertory in these *Mycoplasma* species. For each species, two cDNA libraries were generated yielding on average 230700, 335299 and 100222 reads, for *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* cDNA libraries, respectively, with an average read length of 240 bp. Annotation of the sequences showed that 87%, 89% and 95% of the predicted genes are transcribed from the *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* genomes, respectively, under the conditions used. In all three libraries the most represented transcripts are related to genes involved in cell division and transport systems. The transcriptome map allowed the identification and confirmation of transcriptional units structures in the three genomes. Finally, our data defined in detail the RNA populations found in *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* and we have shown that our data can be used to map transcript boundaries and transcription unit structure on a genome wide scale.

Keywords: *Mycoplasma hyopneumoniae*, *Mycoplasma flocculare, Mycoplasma hyorhinis,* transcriptome profile, transcriptional unit structures.

Introduction

Mycoplasma hyopneumoniae, *Mycoplasma flocculare* and *Mycoplasma hyorhinis* are the most important species that have been identified in the swine respiratory system (Mare and Switzer 1965; Meyling and Friis 1972). *M. hyopneumoniae* is the etiological agent of porcine enzootic pneumonia, while *M. hyorhinis*, besides being found in swine respiratory tracts, may cause swine serofibrinous to fibrinopurulent polyserositis and arthritis (Friis and Feenstra 1994). *M. flocculare* is also widespread in swine herds, but no disease has been associated with this species (Kobisch and Friis 1996). Based on a 16S rRNA sequence and genome comparison, it has been shown that *M. hyopneumoniae* and *M. flocculare* are phylogenetically closely related species (Stemke et al. 1992; Siqueira et al. 2013). *M. hyopneumoniae* can adhere to the cilia of tracheal epithelial cells and causes damage. Although *M. flocculare* can also adhere to cilia, no important resulting damage has been

observed, suggesting that *M. hyopneumoniae* and *M. flocculare* may possess different mechanisms, facilitating the recognition of different receptor sites on the cilia (Young et al. 2000). In addition to the ability of *M. hyorhinis* to colonize other host sites (Kobisch and Friis 1996), a potential role of *M. hyorhinis* infection in cancer development and acceleration to malignant phenotypes has been proposed (Huang et al. 2001; Yang et al. 2010; Kornspan et al. 2011).

Similar to other mycoplasmas, *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* present a small genome with limited biosynthetic potential. Five *M. hyopneumoniae* genome strains have been sequenced (Minion et al. 2004; Vasconcelos et al. 2005; Liu et al. 2011; Siqueira et al. 2013); four *M. hyorhinis* genome strains (Liu et al. 2010; Kornspan et al. 2011; Calcutt et al. 2012; Goodison et al. 2013); and one *M. flocculare* genome strain (Siqueira et al. 2013). Although a significant amount of data has been produced by genome sequencing and comparative analysis in these species, very limited information related to transcription mechanisms and regulation is available for these organisms.

Analysis related to transcriptional units (TU) organization, transcription regulation and promoter sequences are available for the *M. hyopneumoniae* 7448 genome (Siqueira et al. 2011; Weber et al. 2012). The global assessment of TU organization of the *M. hyopneumoniae* genome by both *in silico* and *in vitro* approaches suggested that the ORFs are continuously transcribed (cotranscription) in large clusters (TUs). The authors predict that each TU is transcribed in the same direction with no intervening gene transcribed in the opposite one (Siqueira et al. 2011). This predicted *M. hyopneumoniae* strains TUs organization feature was also observed in *M. flocculare* and *M. hyorhinis* genomes (Siqueira et al. 2013) and it is similar to the organization proposed for *Mycoplasma pneumoniae* (Güell et al. 2009). Additional knowledge of the whole transcriptional organization should contribute to understanding the unexplored mechanisms of mycoplasma transcriptional regulation.

Considering the importance of the three mycoplasma species that inhabit the swine respiratory tracts and the limited knowledge about mycoplama transcriptional regulation, we decided to analyze the whole transcriptome of *M. hyopneumoniae, M. flocculare* and *M. hyorhinis*. In this work we have sequenced, analyzed and constructed a single nucleotide resolution transcriptome map of *M. hyopneumoniae*, *M. flocculare*

and *M. hyorhinis*. A comparative analysis of the whole gene expression profile variation between these swine respiratory mycoplasmas was performed. Furthermore, the genome organization in large transcriptional units (polycistronic mRNA), as previously predicted, was estimated and compared between the three species.

Results and Discussion

454 sequencing, assembly and mapping of *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* transcriptome

The results presented here are partially completed. Some limitations in the current analysis will be resolved with the new transcriptome sequencing and assembly which is being performed.

A total of 1,071,421 raw sequencing reads with average length of 240 bp were generated by 454-pyrosequencing. After cleaning and removing the adapters of each sequence, low-quality sequences (quality scores 20) and ribosomal RNA sequences, a total of 666,221 high quality reads were obtained and used for de novo assembly. The cDNA library was constructed with total mRNA from *M. hyopneumoniae* 7448, *M. flocculare* ATCC 27716 and *M. hyorhinis* ATCC 17981. The RNA preparations from the three species were from bacteria grown under the same culture conditions. Table 1 shows a detailed comparative analysis of each library assembly. To determine transcribed regions in the genome, we estimated the average coverage depth of reads mapped per nucleotide/base. We used a pileup format, which represents the signal map file for the whole genome in which alignment results (coverage depth) are represented in per-base format.

The *M. hyopneumoniae* library allowed the generation of 230,700 reads, and approximately 97% of them were mapped to the reference genome sequence of *M. hyopneumoniae* strain 7448. Most of the 335,299 reads (97%) of the *M. flocculare* library also mapped to the reference genome sequence of *M. flocculare* strain ATCC 27716. Similarly, ninety percent of the 100,222 reads from the *M. hyorhinis* library mapped to the reference genome sequence of *M. hyorhinis* strain HUB-1. The gene

mapping in *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* represented 87%, 89% and 95% of total predicted ORFs in these genomes, respectively (Table 1). A full list of *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* transcribed genes during the culture conditions used are presented in the supplementary data (Supplemental Table S1; Supplemental Table S2; Supplemental Table S3, respectively).

The present coverage of the genes obtained in this work is higher than those reported previously for different organisms (Yu et al. 2008; Güell et al. 2009; Oulion et al. 2012). The contigs number was similar for the three libraries (see Table 1), and the coverage ranged from 1 to more than 1,000 reads per contig, with most of the contigs covered by less than 20 reads. The contig length distribution is similar between *M. hyopneumoniae* and *M. flocculare* contigs, but different from *M. hyorhinis* contigs (Figure 1). Most of the *M. hyopneumoniae* and *M. flocculare* contigs ranged from 200 to 400 bp, unlike *M. hyorhinis* reads that despite have contigs ranged from 200 to 500 bp, has a lot of contig formation (by assembly) ranged from 2,000 to 3,000 bp (Figure 1).

Comparative gene mapping of *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* transcriptomes

A comparative genome analysis between *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* has been recently described (Siqueira et al. 2013) showing a highly similar genome content and organization between *M. hyopneumoniae* and *M. flocculare.* These two closely related mycoplasma species (Stemke et al., 1992) present differences in relation to pathogenicity, *M. hyopneumoniae* being considered pathogenic and *M. flocculare* considered as a commensal. The genome comparison showed many differences that may help to explain their different behavior. However, it was not possible to identify specific virulence determinant factors that could explain the differences in their pathogenicity (Siqueira et al. 2013). The same comparative analysis showed that *M. hyorhinis* genome also shared a significant amount of genes with *M. hyopneumoniae* and *M. flocculare*. Differences in the *M. hyorhinis* genes repertory seems to be related mainly to *M. hyorhinis* ability to colonize different hosts and sites.

The distribution of genes based on the *M. hyopneumoniae* data shows that 87% of the predicted genes have significant hits for one or more sequences of the *M. hyopneumoniae* transcriptome (Supplemental Table S1). A detailed analysis of the *M. hyopneumoniae* transcriptome data demonstrated that 210 of the mapped genes have fifteen or more reads. Some of these genes are summarized in Table 2. The genes with the highest number of reads were related mainly to cell basal metabolism; the highest number of reads (39,242 reads) corresponded to the transcript of the *rnpB* gene, which encodes the RNA component of RNase P. Other genes which also presented a high number of reads were those encoding S-adenosyl-methyltransferase, cell division protein MraZ, cell division protein FtsZ and hexosephosphate transport protein.

Surface proteins are potentially related to bacterial pathogenesis. The *M. hyopneumoniae* gene encoding adhesin P216 (MHP7448_0496), a surface protein which bind to porcine cilia and heparin (Wilton et al., 2009) presents a significant number of reads (12,836) (Table 2). Other genes encoding surface proteins, such as MHP7448_0656 - lipoprotein p65, MHP7448_0198 - protein p97 copy 1 and MHP7448_0497 - p76 membrane protein, which are possibly involved in host cell adhesion, are also among the genes with a higher number of reads (Table 2).

Transcripts mapping in *M. flocculare* genome covered 89% of predicted genes with significant hits for one or more sequences of the *M. flocculare* transcriptome (Supplemental Table S2). There were 227 genes which presented fifteen or more reads. Some of these genes are summarized in Table 3 and several genes presenting a high number of reads are related to cell basal metabolism. For instance, the gene MF01423 that encoded cell division protein MraZ was the *M. flocculare* gene with more transcripts identified (36,440 reads). Similar to *M. hyopneumoniae*, in *M. flocculare* the transcripts from S-adenosyl-methyltransferase and cell division protein FtsZ genes seem to be present in a higher number when compared to other transcripts. Three genes, exclusive of *M. flocculare,* encoding hypothetical proteins (MF01093, MF01463 and MF00857) also presented higher transcript numbers (Table 3).

The analysis of the *M. flocculare* genome showed that sequences related to two important *M. hyopneumoniae* adhesins (MHP7448_0198 - protein p97 copy 1 and MHP7448_0199 - protein p102 copy 1) are absent in *M. flocculare* genome (Siqueira et al. 2013). The other genes that encoded adhesins in *M. flocculare* genome are

transcribed in different amount of transcripts (Table 4), being even higher for MF00472 - protein p97 copy 2 with 510 reads (*M. hyopneumoniae* orthologs is MHP7448_0108) and MF00475 - protein p102 copy 2 with 102 reads (*M. hyopneumoniae* orthologs is MHP7448_0107). The higher number of reads for these adhesins, compared to *M. hyopneumoniae* corresponding adhesins, was unexpected (see Table 4), and at present we have no explanation for these results. However, the number of reads for the MHP7448_0198 - p97 copy 1, without ortholog in *M. flocculare* was higher, as shown in Table 2 and Table 4. Thus, we can consider the presence and expression of p97 copy 1 as highly related with the pathogenic capacity of *M. hyopneumoniae*.

The mRNA accesses for *M. hyorhinis* libriries was performed with strain *M. hyorhinis* ATCC 17981; however, the genome of this strain is not available. Therefore, the transcriptome assembly and mapping was based in the *M. hyorhinis* HUB-1 genome. The assembly with a different genome strain can result in a mistake mapping especially in the intergenic regions. Nevertheless, gene mapping was the most effective among the three mycoplasma species (Table 1). Ninety-five percent of the predicted *M. hyorhinis* HUB-1 genes presented transcripts. Moreover, the number of genes that have 15 or more reads corresponded to 395 genes. However, the number of reads per gene was lower than that observed in *M. hyopneumoniae* and *M. flocculare* transcriptome. The gene encoding hexosephosphate transport protein (MHR_0432) presented the highest number of reads (Table 5), followed by the genes *vlpB*, *vlpG*, *vlpE* which encode variant surface proteins exclusive to *M. hyorhinis*, and MHR_0162 - surface antigen. Similar to the results found in *M. hyopneumoniae* and *M. flocculare,* the other genes presenting high number of reads are related to cell basal metabolism: *rpoC*, *mraZ*, *rpoB*, and *tuf*, all presenting about 200 reads. The hypothetical gene MHR_0660 with 260 reads (Table 5), shared one ortholog with *M. hyopneumoniae* - MHP7448_0439 – which have 150 reads mapped to it. No functional characterization was predicted for these sequences.

M. hyorhinis contains an exclusive variable lipoprotein (Vlp) system that constitutes its major coat protein (Yogev et al. 1995) and provides a mutational strategy for evasion of the host immune system. Different *M. hyorhinis* strains carry a variable number of *vlp* genes (Yogev et al. 1995). *M. hyorhinis* HUB-1 is characterized by the presence of seven *vlp* genes displayed in the order 5′-*vlpD*-*vlpE*-*vlpF*-insertion sequence (IS)-*vlpG*-*vlpA*-IS-*vlpB*-*vlpC*-3′ (Liu et al., 2010). In the *M. hyorhinis* transcriptome

map, the seven *vlp* genes are identified, however, the transcript numbers from the different *vlp* genes was variable, *vlpB*, *vlpG* and *vlpE* genes presenting more reads than *vlpF*, *vlpD*, *vlpC* and *vlpA* (see Supplemental Table S3). Previous reports have shown that structural variation affects the abundance and functionality of Vlps on the mycoplasma surface (Citti and Wise 1995), but the full consequences of this variation are yet to be fully understood. An organism carrying multiple *vlp* genes has the capacity to generate a large number of variants expressing antigenically distinct *vlp* gene products, either alone or in combinatorial mosaics on the cell surface. Therefore, Citti et al. (2000) argue strongly for specific functions associated with each product, presumably selected in the natural host niche.

The *M. hyorhinis* gene presenting more reads was MHR_0432 which encodes hexosephosphate transport protein. In *M. hyopneumoniae* the ortholog sequence (MHP7448_0136) was the sixth gene presenting more reads, while in the *M. flocculare* transcriptome, the orthologue MF00133 was the fifty-fifth gene in number of reads (with 90 reads). The hexosephosphate transport protein (UhpT) is a transporter protein which facilitates the uptake of phosphorylated hexose molecules into the cell in enterobacteriacea family (Hoffer et al. 2001). Despite the interesting amount of unique transcripts for hexosephosphate transport protein in the three transcriptome maps, the *Mycoplasma* spp. action route of this protein was not characterized.

The gene *ftsZ* whose product is related to cell division was one of the genes with the highest number of reads in *M. hyopneumoniae* and *M. flocculare* transcriptomes (Table 2; Table 3). Nevertheless, the *M. hyorhinis* transcriptional level is lower (48 reads). Most bacteria produce the tubulin homolog FtsZ (Gilson and Beech 2001), which forms a cytoskeletal filament that is essential for membrane constriction and coordination of septal peptidoglycan synthesis (Errington et al. 2003). Although, mycoplasmas have a gene encoding FtsZ (Zhao et al. 2004); its function in cell division has not been experimentally demonstrated in these bacteria and in many species that do have such a gene, its sequence is significantly divergent. Nevertheless, experimental evidences suggest that this gene is essential in these organisms (Hutchison et al. 1999). Furthermore, in *M. pneumoniae* there is very little *ftsZ* mRNA and very little FtsZ protein (Benders et al. 2005), raising the possibility that mycoplasmas do not actually use FtsZ for cell division. Interestingly, in *M. pneumoniae*, the Triton X-100-insoluble structures of the terminal organelle are involved in cell division (Hegermann et al. 2002;

Balish et al. 2003), suggesting functional replacement of one type of cytoskeleton by another. Neither actin-like nor intermediate-filament like proteins are encoded in genomes of the mycoplasmas. Clearly, mycoplasmas and their relatives have subcellular structures that are distinct from those of other bacteria. Furthermore, there is a diversity of cytoskeletal systems among these organisms despite their close phylogenetic relationship.

The *M. hyorhinis* MHR_625 gene or *p37* gene, that encode the named 'high affinity transport system protein p37' presented 71 reads (Supplemental Table S3). Interestingly, p37 topology analyses provide evidences for the putative role as extracytoplasmic thiamine-binding lipoprotein (Sippel *et al*., 2008; Sippel *et al*., 2009). *M. hyopneumoniae* and *M. flocculare* have a p37-like ABC transporter protein (MHP7448_0360 and MF01218), with especially structural similarity in ABC transporter sities (Sippel *et al*., 2008). So, future experimental studies are necessary to determine the exact p37 biological implications and its role originally described in tumorigenesis and cancer progression (Sippel *et al*., 2008; Sippel *et al*., 2009).

As shown, a significant fraction of the genome genes were mapped from transcriptome data. The percentage of non-mapping genes in *M. hyopneumoniae, M. flocculare* and *M. hyorhinis* were 13%, 11% and 5%, respectively (Table 1). Comparison between *M. hyopneumoniae* and *M. hyorhinis* revealed approximately 84% of non-mapping genes were genes with unknown functions (hypothetical proteins) (Supplemental table S4; Supplemental table S5). The other genes without transcripts in the *M. hyopneumoniae* genome were ABC-transporter, lipoproteins, permease, mannitol-1-phosphate 5-dehyrogenase (*mtlD*) and PTS system mannitol-specific component IIA (*mtlF*) (Supplemental table S4). In *M. hyorhinis* the five genes presenting no detectable transcripts with known function were ABC-transporter, a restriction endonuclease S subunit (MHR_0200), a type I site-specific DNA methyltransferase specificity subunit: *hsdS*, *mtlD* and *mtlF* genes (Supplemental table S5). *M. flocculare* presented some non-mapping genes, whose products are possibly essential for the cell functions, such as ribonuclease III (MF00007) and recombination protein RecA (MF01113) (Supplemental table S6). This may be due to the difficulty in mapping the *M. flocculare* transcriptome, since the *M. flocculare* genome is not complete (Siqueira et al. 2013) and also may due to methodological problems. A new round of cDNA preparing and pyrosequencing will be realized, aiming to confirm these

data. The *M. flocculare* non-mapping genes with unknown functions represent about 60% (41 of 68 genes). Similarly to *M. hyopneumoniae* and *M. hyorhinis* transcription products from the *mtlD* and *mtlF* genes were also not detected in *M. flocculare* under the reference conditions.

Many bacteria can transport d-mannitol and other carbohydrates by a phosphoenolpyruvate-dependent phosphotransferase system (PTS) (Postma et al. 1993). The system for the uptake and catabolism of mannitol is generally composed by *mtlA*, *mtlR*, *mtlF* and *mtlD* genes, coding for mannitol-specific IIBC component transporter, the transcriptional regulator MtlR, the phosphotransferase mannitol-specific component IIA and mannitol-1-phosphate dehydrogenase, respectively. Those genes, with the exception of *mtlR*, are present in *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* genomes, and no transcripts corresponding to *mtlD* and *mtlF* genes were found in any of the transcriptome (Supplemental table S4; Supplemental table S5; Supplemental table S6). The *mtlA* gene, which was transcribed in the three mycoplasmas species, is involved in the mannitol uptake. Analyses indicate that the genes of mannitol catabolism in *Bacillus* species are sensitive to catabolite repression (Reizer et al. 1998; Henstra et al. 2000). When favorable catabolites like glucose are utilized, HPr protein (protein belonging to PTS system) is phosphorylated forming a complex repressor that will prevent expression (Reizer et al. 1998). In the three species studied, no HPr-like was found and therefore, the system is incomplete. The use of different carbohydrates and their effects on the activity of the different phosphotransferase system (PTS) components in *M. pneumoniae* was studied by Halbedel et al (2004). The authors concluded that mannitol did not serve as a single carbon source even though the genetic information to use this carbohydrate is complete in *M. pneumoniae* genome. The genes required for mannitol transport and conversion to fructose-6-phosphate might be poorly or not expressed (Halbedel et al. 2004; Jaffe et al. 2004). Moreover, in *M. genitalium*, the genes for mannitol transport were completely lost (Fraser et al. 1995). Thus, this may represent a further step in the reductive evolution of the mycoplasmas genomes.

Transcriptional Units Structures

We have constructed a transcriptome map of *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis*. The transcriptome map which allowed identification of TUs structures at a genome scale is critical for identifying co-expressed genes and for understanding coordinated regulation of the mycoplasma transcriptome. We have identified, in the reference conditions, co-expression for about 70% of the predicted genes for all three species. By joining consecutive overlapping reads or reads that mapped uniquely (Figure 2), we presented a common structure of genes transcription in mycoplasmas. Recently, a comparative genome analysis among *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* was described (Siqueira et al. 2013). The TUs arrangements by *in silico* prediction revealed a similar number of TUs among these three mycoplasmas genomes (Siqueira et al. 2013), indicating that gene organization in *M. flocculare* and *M. hyorhinis* also occur preferably in TUs as found in *M. hyopneumoniae* (Siqueira et al. 2011). Moreover, as previously described for *M. hyopneumoniae*, the overall genes distribution within TUs in *M. flocculare* and *M. hyorhinis* has a high variation in relationship to gene number and functional categories of the encoded products (Siqueira et al. 2013). As expected these predicted TUs by *in silico* and *in vitro* approach in *M. hyopneumoniae* plus computationally predicted TUs in *M. flocculare* and *M. hyorhinis* genomes were confirmed in the three transcriptomes analyzed.

The overlapping reads mapping in each reference genome generated a surprisingly large number of long contigs. Some examples are the contigs that mapped in the nucleotides positions 268881 - 277986 bp with a size higher than 9,000 bp or the mapped contig in position nucleotides 242981 - 245879 bp with a total length of 2,898 pb both of *M. hyopneumoniae*. *M. hyorhinis* map has the largest mRNA contig mapped, with a size reaching to 13,000 bp (location: 485595 – 498603 bp or 779997 - 788652 bp). Figure 2 illustrates a representative transcription unit of each of the transcriptomes analyzed. Most of the observed TU has overlapping reads (as shown in figure 2A, 2B and 2C). Interestingly, the transcriptional level in the TUs shows a decay behavior (see Fig. 1A), indicating that such staircase-like expression is a widespread phenomenon in bacteria. The fact that consecutive genes within operons do not have the same expression level has been observed in other bacterial species (Adhya 2003; Güell et al. 2009). *M. pneumoniae* transcriptome data shows natural polarity in the TU, where the first gene of the operon exhibits transcription that progressively decreases with each gene towards the end of the transcript (Güell et al. 2009). Thus, although genome

reduction leads to longer operons accommodating genes with different functions (Lee et al. 2008; Güell et al. 2009; Siqueira et al. 2011), the last gene can still retain internal transcription and termination sites under certain conditions.

M. hyopneumoniae transcription unit architecture is characterized by the presence of long transcriptional units containing genes that are highly variable in the functional categories of the encoded products. Moreover, all the transcriptional units are preceded by putative promoter sequences (data not published) providing evidence for gene organization as an important factor in the regulation of gene transcription. A subset of genes within the TUs that could be transcribed by alternative internal promoters has been demonstrated and associated with possible complex transcriptional organization in *M. hyopneumoniae* genomes (Siqueira et al. 2011; Weber et al. 2012). *M. pneumoniae* transcriptome analyses also shows the wide occurrence of polycistronic operons with alternative transcripts (Güell et al. 2009). In *Bacillus subtilis* 20% of genes in polycistronic operons are transcribed from more than one promoter (Makita et al. 2004). Similarly, almost 6% of the polycistronic operons contain an internal read-through terminator, at which partial continuation of the transcription occurs (Hoon et al. 2005). Koide et al. (2009) carried out a study in *Halobacterium salinarum* and detected 40% of condition dependent operons. Taken together, these previous reports suggest a complex relationship between genomic organization and gene expression in prokaryotes.

Mycoplasmas genomes exhibit a very high coding density indicating a highly optimized usage of the coding capacity of their genomes. Previous studies showed similarity between *M. pneumoniae* and more complex bacteria, in relationship to complexity in the transcriptome regulation, metabolic responses and adaptation (Güell et al. 2009). However, at mycoplasmas transcriptional regulation level, the knowledge is very limited, and evidence suggests that regulation could occur by different and as yet unknown mechanisms.

Overall, we provided a comprehensive transcriptome view of three important mycoplasmas that inhabit the respiratory tract of swine. We have defined in detail the RNA populations found in *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* and we have shown that our data can be used to map transcript boundaries and operon structure on a genome-wide scale. Description of all functional elements in *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* system will be a prerequisite for conducting holistic

systems approaches to understand the complex transcriptional regulation in these mycoplasma species.

Methods

Bacterial strains, culture conditions and RNA preparation

M. hyopneumoniae strain 7448, *M. flocculare* ATCC 27716 and *M. hyorhinis* ATCC 17981 were used in this study. The bacteria were grown in 25 ml Friis broth (Friis 1975) at 37 °C for 24 hours with gentle agitation in a roller drum.

Total RNA was isolated with RNeasy Mini Kit (Qiagen, USA). For cell lysis, 0.7 ml of RNeasy Lysis Buffer (RLT buffer) in the presence of 0.134 M of the β-mercaptoethanol was used per cultivation flask. The purification was done according to the manufacturer's instructions, with on-column DNaseI digestion using the RNase-Free DNase Set (Qiagen, USA) and a second round of treatment with DNase I (Fermentas, USA). The absence of DNA in the RNA preparations was monitored by PCR. The extracted RNA was analyzed by gel electrophoresis and quantified in the Qubit™ system (Invitrogen, USA). RNA quality and integrity were determined by evaluation of the RNA integrity number (RIN) (Schroeder et al. 2006) using Agilent 2100 Bioanalyzer (Agilent, USA).

Equal quantities of total RNA from the three biological replicates of *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* samples were pooled before mRNA purification. The ribosomal RNA (rRNA) depletion was performed using terminator™ 5'-phosphate-dependent exonuclease (TEX – Epicentre, USA), according to the manufacturer´s instructions. Briefly, total RNA (10 μg) was combined with TEX 10X Reaction Buffer A, 20 U RiboGuard RNase Inhibitor, 10 U TEX (1 U/ug mRNA) and nuclease free water in a final volume of 40 μl and incubated at 30 °C for 60 minutes. The reaction was stopped with EDTA (0.5 M) and mRNA was precipitate with 3 M sodium acetate and 2.5 volumes of ethanol. The mRNA was quantified in the Qubit™ system (Invitrogen, USA) and the effectiveness of the reaction was assessed by Agilent

2100 Bioanalyzer (Agilent, USA). The absence of 16S and 23S rRNA in the posttreatment sample indicates a successful reaction.

cDNA preparation and sequencing

cDNA library preparation and pyrosequencing was performed using GS-FLX Titanium series reagents essentially following the manufacturer's instructions (Roche Diagnostics, Germany). Equal amounts of mRNA (200 ng) of each strain were used for synthesized the first- and second-strand cDNA according to the *cDNA Rapid Library Preparation Method Manual GS FLX Titanium Series* (Roche Diagnostics, Germany). After library construction, the samples were quantified using a Qubit™ system (Invitrogen, USA), and average fragment sizes were determined by Agilent 2100 Bioanalyzer (Agilent, USA). Two cDNA libraries from each strain were generated, totaling six libraries, which were sequenced in a Roche/454 GS-FLX system.

Assembly, mapping and functional annotation

The raw reads obtained were assembled and mapped with Newbler v2.6 program with default parameters.

For mapping of 454 reads, the *M. hyopneumoniae* 7448, *M. flocculare* ATCC 27716 and *M. hyorhinis* HUB-1 genome sequences were used as reference and rRNA sequence was filtered. The results were parsed and gb format files were generated for analysis with the package Artemis, including information about the coverage of each region and their respective genes.

A second mapping was performed against the *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* genes sequences for functional annotation of the *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* transcriptomes. Data about metabolic pathways will be extracted from the functional annotation of genomes, but only when the final sequences are obtained by the new sequencing round.

The Standard Flowgram Format (SFF) files of the Q20 sequence data generated in this study have been deposited in the Short Read Archive (SRA) database at NCBI under the accession number SRAxxx (experiment accession numbers SRxxx - SRxxxx).

Analysis of structures transcription

Expressed reads with coverage above background were mapped onto the annotated genes of *M. hyopneumoniae* 7448 (NC_007332), *M. flocculare* ATCC 27716 (NZ_AFCG00000000.1) and *M. hyorhinis* HUB-1 (NC_014448.1). Genes that had a significantly higher proportion of their length (60%) covered reads were considered to be expressed.

The operon assessment was based in the published experimental data from *M. hyopneumoniae* 7448 (Siqueira et al. 2011). Based in this study, *M. flocculare* and *M. hyorhinis* transcription units were predicted (Siqueira et al. 2013). Thus, with RNA-Seq results we used in-house perl scripts to analyze the predicted TUs for the three mycoplasma species taking into account the following criteria: i) Structure previously predicted by Siqueira et al. (2011) and Siqueira et al. (2013); ii) Expression of all genes; iii) Transcription of the genes in the same direction; iv) Expression of intergenic region between the genes. Overlapping pairs of such genes were joined together to identify large operon structures.

Accession numbers

Assembled transcripts will be deposited in the Short Read Archive (SRA) database at NCBI with accession numbers between [genbank:xxxxxxx] - [genbank:xxxxxx].

Competing interests

The authors declare they have no competing interests.

Acknowledgments

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) and the Fundação de Amparo à

Pesquisa do Estado do Rio de Janeiro (FAPERJ). FMS has received scholarships from CAPES.

Figure Legends

Figure 1. Length distribution of identified reads from *M. hyopneumoniae* (MHP), *M. flocculare* (MFL) and *M. hyorhinis* (MHR) transcriptome assembly.

Figure 2. **Transcriptome feature in the reference condition.**

The blue light arrowheads represent the predicted genes and they are positioned at the direction of transcription. The gene names are presented below the arrow. The green arrowheads represent the mRNA transcript. **(A)** *M. hyopneumoniae* Myo-inositol transcription unit (TU) composed by ten genes. This TU in the genome on the forward strand has a staircase behavior, meaning that the consecutive genes have lower and steady expression levels. **(B)** *M. flocculare* TU structure composed by five genes. This TU in the genome is located on the reverse strand. **(C)** *M. hyorhinis* TU structure composed by three genes. This TU in the genome is located on the forward strand. The position on the chromosomal sequence is indicated in base pairs (bp) below both termini of the bars. Visualization by the software Artemis.

Figure 1.

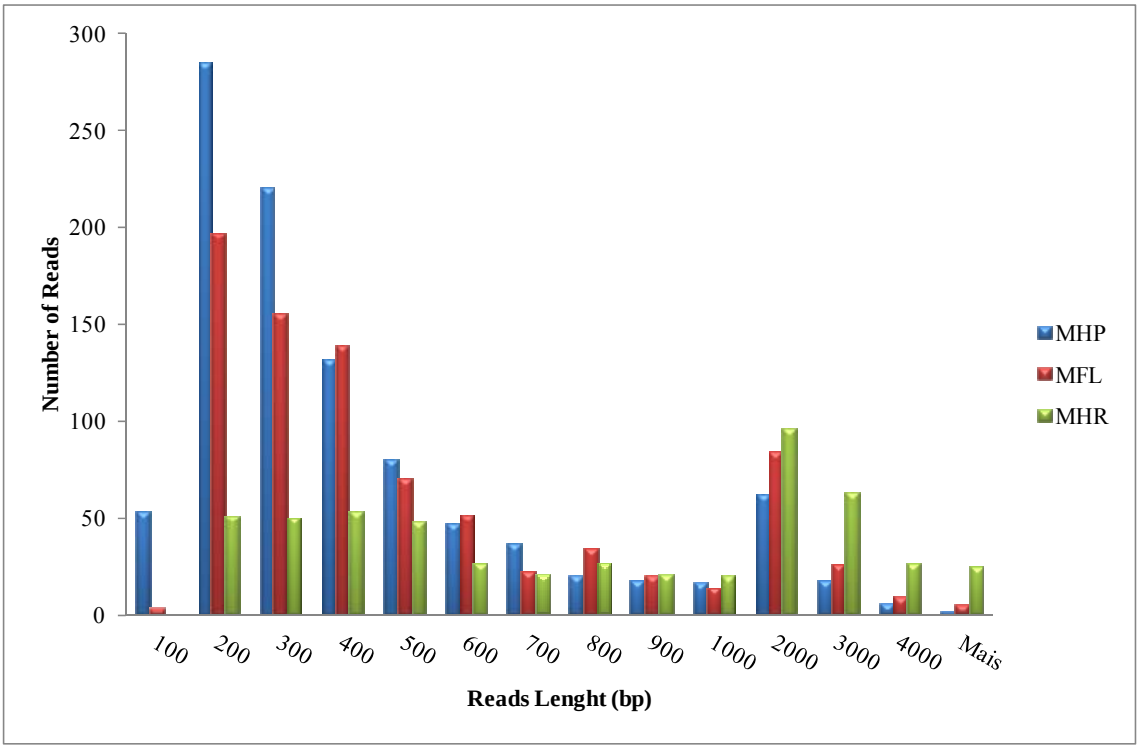
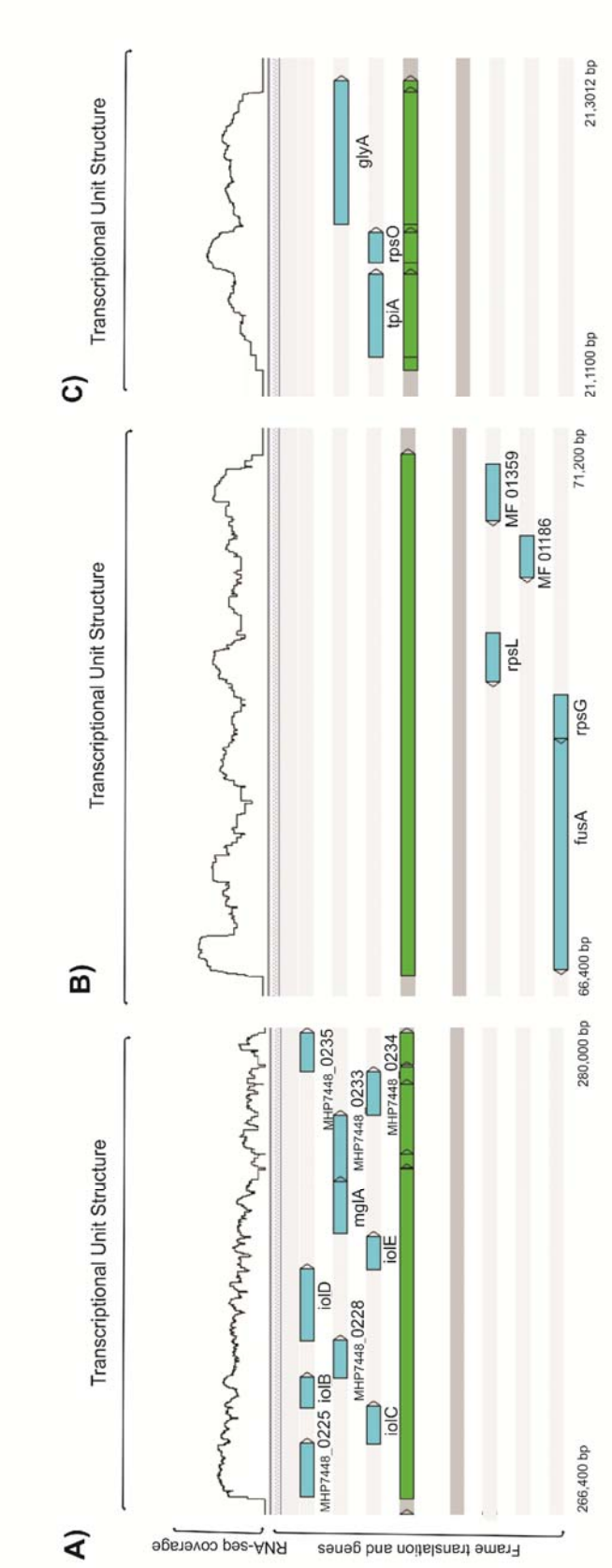


Figure 2


A)
Transcriptional Unit Structure
RNA-seq coverage
Frame translation and genes
MHP7448_0225
iolB
MHP7448_0228
iolC
iolD
iolE
mglA
MHP7448_0233
MHP7448_0234
MHP7448_0235
266,400 bp
280,000 bp
B)
Transcriptional Unit Structure
rpsL
fusA
rpsG
MF 01186
MF 01359
66,400 bp
71,200 bp
C)
Transcriptional Unit Structure
tpiA
rpsO
glyA
21,1100 bp
21,3012 bp

Tables

Table 1. 454 sequencing and assembly datas generated for *M. hyopneumoniae*, *M. flocculare* and *M. hyorhinis* transcriptome.

Library	*M. hyopneumoniae*	*M. flocullare*	*M. hyorhinis*
Total reads (bp)	230,7	335,299	100,222
Total rRNA reads	119,702	120,814	16,648
Average length (bp)*	240	250	230
Total reads mapped (%)	97%	97%	90%
Contig number	954	855	895
Large contigs (≥ 500bp)	189	238	439
Genes mapping	651 (87%)	547 (89%)	654 (95%)
Gene Mapping with more than 15 reads	210	227	395
Genes no-mapping (%)	95 (13%)	68 (11%)	30 (5%)

* Average length of high-quality mRNA reads.

Table 2. Genes mapped with the highest number of transcript reads in the *M. hyopneumoniae* genome.

Id_ORF	Gene name	Product	Reads number
MHP7448_0104	rnpB	RNase P RNA	39242
MHP7448_0496	-	putative p216 surface protein	12836
MHP7448_0392	-	S-adenosyl-methyltransferase	10121
MHP7448_0391	-	cell division protein MraZ	9771
MHP7448_0393	ftsZ	cell division protein	6351
MHP7448_0136	-	hexosephosphate transport protein	1909
MHP7448_0067	dnaK	chaperone protein	1132
MHP7448_0137	ldh	L-lactate dehydrogenase	881
MHP7448_t29	tRNA	Ser	809
MHP7448_0115	pdhA	pyruvate dehydrogenase	675
MHP7448_0225	-	methylmalonate semialdehyde dehydrogenase	617
MHP7448_0377	-	hypothetical protein	604
MHP7448_0376	sgaT	PTS system ascorbate	534
MHP7448_0492	fruA	PTS system fructose	413
MHP7448_0116	pdhB	pyruvate dehydrogenase	389
MHP7448_0656	-	prolipoprotein p65	389
MHP7448_0198	-	protein P97 copy 1	292
MHP7448_0497	-	p76 membrane protein	288
MHP7448_0114	apt	adenine phosphoribose	242
MHP7448_0226	iolC	myo-inositol catabolism	236

Table 3. Genes mapped with highest number of transcript reads in the in *M. flocculare* genome.

Id_ORF	Gene name	Product	Reads number
MF01423	mraZ	cell division protein MraZ	36440
MF01101	-	S-adenosyl-methyltransferase	23325
MF01095	ftsZ	cell division protein ftsZ	8763
MF01093	-	hypothetical protein	6513
MF01463	-	hypothetical protein	2620
MF01536	tRNA	Ser	1153
MF00857	-	hypothetical protein	1053
MF00837	rplK	50S ribosomal protein L11	1020
MF00835	rplA	50S ribosomal protein L1	891
MF01167	dnaK	chaperone protein dnaK heat	732
MF00750	-	hypothetical protein	562
MF01519	tRNA	Leu	538
MF00472	-	protein P97 copy 2	510
MF00861	mgtE	MG2+ transport protein	463
MF00591	apt	adenine phosphoribosyltransferase	397
MF01534	tRNA	Leu	390
MF00747	sgaT	PTS system ascorbate specific	372
MF01400	rpsJ	30S ribosomal protein S10	337
MF01476	-	5S ribosomal RNA fragment	328
MF00856	-	hypothetical protein	214

Table 4. Summary of adhesins associated to pathogenicity, genome organization and transcription level comparison.

Product	Transcriptome MHP_7448* (N° reads)	Transcriptome MFL* (N° reads)	ID MHP_7448	ID MFL	MHP_7448 x MFL genome location comparison	Target binding (reference for studies in MHP strain 232)
P97-like	2	12	MHP7448_0272	MF00620	inverted	heparin, fibronectin, plasminogen (Seymour et al., 2011)
P116 or P102-like	4	25	MHP7448_0271	MF00623	inverted	porcine cilia, fibronectin, plasminogen (Seymour et al., 2010)
P97 - copy 2	151	510	MHP7448_0108	MF00472	inverted	porcine cilia, heparin, fibronectin (Deutscher et al., 2010)
P102 - copy 2	30	102	MHP7448_0107	MF00475	inverted	NA
P216	12,836	118	MHP7448_0496	MF00848	inverted	porcine cilia, heparin (Wilton et al., 2009)
P159, P110 or P76	288	121	MHP7448_0497	MF00844	inverted	heparin (Burnett et al., 2006)
P97 - copy 1	292	NP	MHP7448_0198	NP	NP	porcine cilia, heparin (Jenkins et al., 2006)
P102- copy 1	11	NP	MHP7448_0199	NP	NP	heparin, fibronectin, plasminogen (Seymour et al., 2012)
Hypothetical protein	65	145	MHP7448_0662	MF01050	inverted	porcine cilia, glycosaminoglycan (Bogema et al., 2011)
P146	86	23	MHP7448_0663	MF01055	inverted	porcine cilia, heparin, and plasminogen (Bogema et al., 2012)
lipoprotein	175	94	MHP7448_0373	MF00741	conserved	porcine cilia, heparin (Deutscher et al., 2012)
Hypothetical protein	38	9	MHP7448_0372	MF00739	conserved	porcine cilia, heparin (Deutscher et al., 2012)
P95	8	10	MHP7448_0099	MF00492	inverted	NA
P60	17	8	MHP7448_0353	MF01236	inverted	NA

NA - not analyzed; NP - not present

* MHP - *M. hyopneumoniae;* MFL - *M. flocculare*

Table 5. Genes mapped with the highest number of transcript reads in *M. hyorhinis* genome.

Id_ORF	Gene name	Product	Reads number
MHR_0432	-	Hexosephosphate transport protein	630
MHR_0348	vlpB	Variant surface antigen B	489
MHR_0343	vlpG	Variant surface antigen G	487
MHR_0338	vlpE	Variant surface antigen E	326
MHR_0162	-	46kDa surface antigen	322
MHR_t019	-	tRNA-Gln	271
MHR_0660	-	hypothetical protein	260
MHR_0056	rpoC	DNA-directed RNA polymerase subunit beta'	246
MHR_0611	gap	Glyceraldehyde 3-phosphate dehydrogenase C	240
MHR_0351	mraZ	Protein mraZ	232
MHR_0517	pdhA	Pyruvate dehydrogenase E1-alpha subunit	222
MHR_0055	rpoB	DNA-directed RNA polymerase subunit beta	214
MHR_0152	-	hypothetical protein	211
MHR_0560	tuf	Elongation factor Tu	194
MHR_0339	vlpF	Variant surface antigen F	181
MHR_0358	-	hypothetical protein	180
MHR_0035	nox	NADH oxidase	174
MHR_0639	-	Lipoprotein	171
MHR_0532	ftsH	cell division protease FtsH	170
MHR_0502	fusA	translation elongation factor G	168

References

Adhya S. 2003. Suboperonic regulatory signals. *Sci STKE* **185**:pe22.

Balish MF, Santurri RT, Ricci A, Lee KK, Krause DC. 2003. Localization of *Mycoplasma pneumoniae* cytadherence-associated protein HMW2 by fusion with green fluorescent protein: implications for attachment organelle structure. *Mol Microbiol* **47**: 49-60.

Benders GA, Powell BC, Hutchison III CA. 2005. Transcriptional analysis of the conserved *ftsZ* gene cluster in *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *J Bacteriol* **187:** 4542-4551.

Calcutt MJ, Foecking MF, Rosales RS, Ellis RJ, Nicholas RA. 2012. Genome sequence of *Mycoplasma hyorhinis* strain GDL-1. *J Bacteriol* **194**: 1848.

Citti C, Wise KS. 1995. *Mycoplasma hyorhinis vlp* gene transcription: critical role in phase variation and expression of surface lipoproteins. *Mol Microbiol* **18**: 649–660.

Citti C, Watson MR, Droesse M, Wise KS. 2000. Gene families encoding phase- and size-variable surface lipoproteins of *Mycoplasma hyorhinis*. *J Bacteriol* **182**: 1356–1363.

Errington J, Daniel RA, Scheffers DJ. 2003. Cytokinesis in bacteria. *Microbiol Mol Biol Rev* **67**: 52-65.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, et al. 1995. The minimal gene complement of *Mycoplasma genitalium. Science* **270**: 397-403.

Friis NF. 1975. Some recommendations concerning primary isolation of *Mycoplasma suipneumoniae* and *Mycoplasma flocculare* a survey. *Nordisk Vet Medicin* **27**: 337–339.

Friis NF, Feenstra AA. 1994. *Mycoplasma hyorhinis* in the etiology of serositis among piglets. *Acta Vet Scand* **35**: 93-98.

Gilson PR, Beech PL. 2001. Cell division protein FtsZ: running rings around bacteria, chloroplasts and mitochondria. *Res Microbiol* **152**: 3-10.

Goodison S, Urquidi V, Kumar D, Reyes L, Rosser CJ. 2013. Complete genome sequence of *Mycoplasma hyorhinis* strain SK76. *Genome Announc* **1**: e00101-12.

Güell M, Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, et al. 2009. Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* **326**: 1268-1271.

Halbedel S, Hames C, Stülke J. 2004. In Vivo Activity of Enzymatic and Regulatory Components of the Phosphoenolpyruvate:Sugar Phosphotransferase System in *Mycoplasma pneumoniae. J Bacteriol* **186**: 7936 -7943.

Hegermann J, Mayer F, Herrmann R. 2002. Cytoskeletal elements in the bacterium *Mycoplasma pneumoniae*. *Naturwissenschaften* **89**: 453-458.

Henstra SA, Duurkens RH, Robillard GT. 2000. Multiple Phosphorylation Events Regulate the Activity of the Mannitol transcriptional Regulator MtlR of the *Bacillus stearothermophilus* Phosphoenolpyruvate-dependent Mannitol Phosphotransferase System. *The Journal of Biological Chemistry* **275**: 7037-7044.

Hoffer S, Uden N, Tommassen J. 2001. Expression of the pho regulon interferes with induction of the uhpT gene in *Escherichia coli* K-12. *Archives of Microbiology* **176**: 370–376.

Hoon MJL, Makita Y, Nakai K, Miyano S. 2005. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS Comput Biol* **1**: 212–221.

Huang S, Li JY, Wu J, Meng L, Shou CC. 2001. Mycoplasma infections and different human carcinomas. *World Gastroentero* **7**: 266-269.

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**: 2165–2169.

Jaffe JD, Berg HC, Church GM. 2004. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **4**: 59-77.

Kobisch M, Friis NF. 1996. Swine mycoplasmoses. *Rev Sci Tech Oie* **15**: 1569-1605.

Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, et al. 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* **5**: 285.

Kornspan JD, Lysnyansky I, Kahan T, Herrmann R, Rottem S, Nir-Paz R. 2011. Genome Analysis of a *Mycoplasma hyorhinis* Strain Derived from a Primary Human Melanoma Cell Line. *J Bacteriol* **193**: 4543-4544.

Lee HJ, Jeon HJ, Ji SC, Yun SH, Lim HM. 2008. Establishment of an mRNA gradient depends on the promoter: an investigation of polarity in gene expression. *J Mol Biol* **378**: 318-327.

Liu W, Fang L, Li S, Li Q, Zhou Z, Feng Z, Luo R, Shao G, Wang L, Chen H, Xiao S. 2010. Complete genome sequence of *Mycoplasma hyorhinis* strain HUB-1. *J Bacteriol* **192**: 5844-5845.

Liu W, Feng Z, Fang L, Zhou Z, Li Q, Li S, Luo R, Wang L, Chen H, et al. 2011. Complete Genome Sequence of *Mycoplasma hyopneumoniae* Strain 168. *J Bacteriol* **193**: 1016–1017.

Makita Y, Nakao M, Ogasawara N, Nakai K. 2004. DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res* **32**: D75–D77.

Mare CJ, Switzer WP. 1965. *Mycoplasma hyopneumoniae*, a causative agent of virus pig pneumonia. *Vet Med* **60**: 841-846.

Meyling A, Friis NF. 1972. Serological identification of a new porcine mycoplasma species, *Mycoplasma flocculare*. *Acta Vet Scand* **13**: 287–289.

Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM, Mahairas GG. 2004. The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. *J Bacteriol* **186**: 7123-7133.

Oulion S, Bertrand S, Belgacem MR, Le Petillon Y, Escriva H. 2012. Sequencing and Analysis of the Mediterranean Amphioxus (*Branchiostoma lanceolatum*) Transcriptome. *PLoS ONE* **7**: e36554.

Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Mol Biol Rev* **57**: 543-594.

Reizer J, Hoischen C, Titgemeyer F, Rivolta C, Rabus R, Stulke J, Karamata D, Saier MH, Hillen W. 1998. A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol Microbiol* **27**: 1157–1169.

Sippel KH, Robbins AH, Reutzel R, Domsic J, Govindasamy L, McKenna MA, Rosser CJ, McKenna R. Structure determination of the cancer associated *Mycoplasma hyorhinis* protein Mh-p37. Biological Crystallography D64:1172–1178, 2008.

Sippel KH, Robbins AH, Reutzel R, Boehlein SK, Namiki K, Goodison S, McKenna MA, Rosser CJ, McKenna R. Structural Insights into the Extracytoplasmic Thiamine-Binding Lipoprotein p37 of *Mycoplasma hyorhinis*. J. Bacteriol. 191(8):2585-2592, 2009.

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**: 3.

Stemke GW, Laigret F, Grau O, Bové JM. 1992. Phylogenetic relationships of three porcine mycoplasmas, *Mycoplasma hyopneumoniae*, *Mycoplasma flocculare* and *Mycoplasma hyorhinis*, and complete 16S rRNA sequence of *M. flocculare*. *Int J Syst Bacteriol* **42**: 220–225.

Siqueira FM, Schrank A, Schrank IS. 2011. *Mycoplasma hyopneumoniae* transcription unit organization: genome survey and prediction. *DNA Res* **18**: 413–422.

Siqueira FM, Thompson CE, Virginio VG, Gonchoroski T, Reolon L, Almeida LG, Fonsêca MMB, Souza R, et al. 2013. New insights on the biology of swine respiratory tract mycoplasmas from a comparative genome analysis. *BMC Genomics* **14**: 175.

Weber SS, Sant'Anna FH, Schrank IS. 2012. Unveiling *Mycoplasma hyopneumoniae* Promoters: Sequence Definition and Genomic Distribution. *DNA Research* **19**: 103-115.

Wilton J, Jenkins C, Cordwell SJ, Falconer L, Minion FC, Oneal DC, Djordjevic MA, Connolly A, Barchia I, Walker MJ, Djordjevic SP. 2009. Mhp493 (P216) is a proteolytically processed, cilium and heparin binding protein of *Mycoplasma hyopneumoniae*. *Mol Microbiol* **71**: 566–582.

Vasconcelos AT, Ferreira HB, Bizarro CV, Bonatto SL, Carvalho MO, Pinto PM, Almeida DF, Almeida LG, Almeida R, Alves-Filho L, et al. 2005. 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.

Yang H, Qu L, Ma H, Chen L, Liu W, Liu C, Meng L, Wu J, Shou C. 2010. *Mycoplasma hyorhinis* infection in gastric carcinoma and its effects on the malignant phenotypes of gastric cancer cells. *BMC Gastroenterol* **10**: 132.

Young TF, Thacker EL, Erickson BZ, Ross RF. 2000. A tissue culture system to study respiratory ciliary epithelial adherence of selected swine mycoplasmas. *Vet Microbiol* **71**: 269-279.

Yogev D, Watson-McKown R, Rosengarten R, Im J, Wise KS. 1995. Increased structural and combinatorial diversity in an extended family of genes encoding Vlp surface proteins of *Mycoplasma hyorhinis*. *J Bacteriol* **177**: 5636-5643.

Yu JK, Wang MC, Shin IT, Kohara Y, Holland LZ, Satoh N, Satou Y. 2008. A cDNA resource for the cephalochordate amphioxus *Branchiostoma floridae*. *Dev Genes Evol* **218**: 723–727.

Zhao Y, Hammond RW, Lee IM, Roe BA, Lin S, Davis RE. 2004. Cell division gene cluster in *Spiroplasma kunkelii*: functional characterization of *ftsZ* and the first report of *ftsA* in mollicutes. *DNA Cell Biol* **23**: 127-134.

5. DISCUSSÃO

A anotação estrutural de genomas e a identificação da abundância de elementos funcionais nos mesmos, como genes, transcritos, RNAs não-codificadores, proteínas e elementos regulatórios, são passos importantes na busca do entendimento da biologia dos microrganismos, e também da dinâmica das doenças infecciosas (Kumar *et al*., 2012). Neste sentido, nossa proposta foi estudar em nível genômico e transcritômico espécies de micoplasmas que habitam o trato respiratório suíno. Nosso principal objetivo é melhorar o entendimento sobre a composição e organização gênica nas espécies de *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis*. Além disso, buscamos, com a obtenção de dados dos transcritomas, o reconhecimento do perfil transcricional e os elementos transcritos, em cada uma das três bactérias estudadas, gerando informações sobre a transcrição nessas espécies. Os resultados apresentados são importantes para o entendimento da biologia desses microrganismos que co-habitam o trato respiratório suíno, causando ou não infecções respiratórias.

Ressaltamos que os resultados apresentados nos transcritomas das três espécies (Capítulo 2) são limitados, pois ainda são parciais. Novos experimentos estão sendo realizados a fim de possibilitar uma análise mais consistente, dos mapas transcricionais. As novas sequências serão geradas em breve, e apartir de então, além da análise dos mapas com o perfil transcricional, pretendemos, complementar as informações com outras abordagens experimentais, a fim de confirmar nossas predições e realizar buscas para a identificação de elementos regulatórios transcritos.

Genomas de bactérias do gênero *Mycoplasma* estão sendo extensivamente sequenciados e nos últimos três anos o número de genomas publicamente disponíveis no NCBI triplicou, passando de 19 isolados de *Mycoplasma* com genomas completos

disponíveis no ano de 2010, para 60 isolados, representados por 25 espécies de *Mycoplasma* possuindo seus genomas completamente sequenciados e disponíveis atualmente (Disponível em: http://ftp.ncbi.nih.gov/genomes/Bacteria/ - dados coletados em 12/06/2013). No Capítulo 1 deste trabalho apresentamos pela primeira vez os dados da sequência do genoma de uma linhagem de *M. flocculare* (ATCC 27716), juntamente com a descrição do genoma da linhagem 7422 de *M. hyopneumoniae*, que representa a quinta linhagem dessa espécie com genoma sequenciado.

A partir da obtenção dos dados de sequenciamento da linhagem ATCC 27716 de *M. flocculare* foi possível a análise genômica comparativa entre *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis*. As três espécies são isoladas do trato respiratório de suínos, sendo que apenas *M. flocculare* não é considerada patogênica. *M. hyopneumoniae* é o agente etiológico da pneumonia enzoótica suína (Sobestiansky *et al*., 1999), sendo descrito infectando exclusivamente suínos. Para a realização deste estudo comparativo foram empregadas a linhagem 7422 aqui descrita e a linhagem 7448 (Vasconcelos *et al*., 2005), ambas patogênicas. Já *M. hyorhinis*, além de apresentar alta plasticidade de infecção, podendo infectar diferentes sítios no hospedeiro, têm aptidão em infectar inclusive hospedeiros diversos. Análises comparativas entre genomas de *M. hyorhinis* isolados de diferentes sítios demonstram uma alta similaridade entre essas linhagens (Liu *et al*., 2010; Kornspan *et al*., 2011; Calcutt *et al*., 2012; Goodison *et al*., 2013). A comparação apresentada no presente trabalho foi baseada no genoma de *M. hyorhinis* HUB-1, uma linhagem patogênica isolada do trato respiratório de suínos por Liu *et al*. (2010).

As análises comparativas mostram que *M. flocculare* e *M. hyopneumoniae* apresentam alto grau de similaridade quanto a composição genômica. Os genomas dessas espécies, quando comparados com o genoma de *M. hyorhinis* também mostraram

grande similaridade. A análise global que foi realizada ressalta a semelhança da composição genômica, assim como organizacional dos genomas de *M. flocculare* e *M. hyopneumoniae*. Isso se aplica tanto em nível de organização de genes que codificam proteínas ribossômicas, como para todas as demais rotas essenciais para a biologia dessas espécies. A análise de classificação do COG e também a classificação funcional das categorias do KEGG, enfatizam da mesma forma, estas semelhanças. Dados anteriores de comparação de porções específicas destes genomas já demonstravam essa semelhança (Stemke *et al*., 1992), designando *M. flocculare* como filogeneticamente relacionado à *M. hyopneumoniae*. Porém, é surpreendente essa alta similaridade em escala genômica, entre duas distintas espécies, sendo uma patogênica (*M. hyopneumoniae*) e outra não patogênica (*M. flocculare*).

A conservação da organização gênica entre isolados da mesma espécie é esperada e considerada essencial para a consolidação de um modo organizacional, dando relação e conexão entre os processos metabólicos e regulatórios na célula (Lathe *et al*., 2000). Porém, *M. hyopneumoniae* e *M. flocculare* compartilham grande similaridade genômica, mas são diferentes espécies, podendo, portanto, este percentual de conservação, ser considerado alto e atípico.

As diferenças verificadas, na comparação de organização genômica entre *M. flocculare* e *M. hyopneumoniae*, são pontuais, envolvendo principalmente a localização de genes codificadores de adesinas, alguns transportadores e presença de genes codificadores de proteínas hipotéticas, as quais possivelmente são efetivas na interação com o hospedeiro e consequentemente, envolvidas no potencial de patogenicidade. *M. flocculare*, diferentemente de *M. hyopneumoniae,* não é capaz de causar infecção (Friis, 1973; Friis, 1974; Armstrong *et al*., 1987; Strasser *et al*.,1992; Blank & Stemke, 2001). Estudos experimentais mostram que esse micoplasma adere-se ao epitélio ciliar de

suínos, porém a resposta inflamatória gerada é muito tênue, assim como as lesões no epitélio e danos aos cílios (Young *et al*., 2000). Os mesmos autores ainda concluem que as adesinas expressas por *M. flocculare*, se aderem, mas não de maneira tão efetiva como as adesinas de *M. hyopneumoniae* (Young *et al*., 2000).

Com exceção dos genes codificadores das adesinas P97 e P102 cópia 1, todos os demais genes codificadores de adesinas presentes no genoma de *M. hyopneumoniae* também estão presentes no genoma de *M. flocculare*. Do mesmo modo, a análise do transcritoma mostra que nas condições de cultivo empregadas, todas as adesinas são expressas em ambos micoplasmas. Interessantemente, o gene codificador da proteína de superfície P216, a qual é capaz de se ligar ao tecido ciliar de suínos e à heparina (Wilton *et al*., 2009), é o segundo gene (MHP7448_0496) mais representado dentre os transcritos de *M. hyopneumoniae,* apresentando nas condições de cultivo empregadas, 12.836 *reads*, sendo superado, em número de *reads*, apenas pelo gene codificador do componente RNA da RNaseP. Em contrapartida, o ortólogo (MF00848) de *M. flocculare* apresenta um número de *reads* bem inferior quando comparado a *M. hyopneumoniae,* apresentando 118 *reads*. Isto nos leva a acreditar que a adesina P216 poderia exercer um significativo papel na capacidade de adesão de *M. hyopneumoniae* às células do hospedeiro. No entanto, segundo Young *et al*. (2000), não é apenas a capacidade de adesão que leva à ocorrência de pneumonia, mas sim um complexo de fatores associados e, por isso, outros mecanismos podem estar sendo empregados para a determinação da capacidade patogênica de *M. hyopneumoniae*.

Interessantemente, *M. hyopneumoniae* apresenta uma unidade transcricional exclusiva, a qual é composta pelos genes do metabolismo do mio-inositol, uma rota ausente em todas as demais espécies de *Mycoplasmas,* cuja função desempenhada em *M. hyopneumoniae* ainda não está descrita. Outras duas unidades transcricionais

exclusivas de *M. hyopneumoniae* são: a unidade composta pelos genes do elemento conjugativo integrativo (ICEH); e a unidade transcricional composta pelos genes codificadores das proteínas P97 e P102 cópia 1. Essas diferenças pontuais em nível genômico são confirmadas com os dados apresentados no Capítulo 2 quanto ao perfil transcricional. Todos os genes supracitados são expressos nas condições empregadas no presente trabalho. Em resumo, esses genes e unidades transcricionais não apenas estão presentes exclusivamente no genoma de *M. hyopneumoniae,* como também são transcritos e, possivelmente, seriam passivos de regulação.

Em relação aos genes que codificam as proteínas P97 e P102 cópia 1 (MHP7448_0198 e MHP7448_0199), as quais sabidamente estão relacionadas à capacidade invasiva de *M. hyopneumoniae* (Jenkins *et al*., 2006; Seymour *et al*., 2012), podemos destacá-los como uma importante diferença em relação ao *M. flocculare*, e então, inferir uma possível função dessas duas proteínas na capacidade de *M. hyopneumoniae* de causar a pneumonia enzoótica suína.

Genes codificando para as adesinas P97 e P102 estão presentes em três cópias no genoma de *M. hyopneumoniae,* enquanto no genoma de *M. flocculare* estão presentes em duas cópias. A nomenclatura usada para o produto gênico de cada uma das cópias é listada na Tabela 1. Estudos de caracterização destas adesinas, em *M. hyopneumoniae,* demonstram um alto grau de antigenicidade das mesmas, com participação na adesão às células hospedeiras, através de um complexo processo de eventos de clivagens específicas (Bogema *et al*., 2012). Em *M. hyopneumoniae*, as sequências apresentam regiões de repetição, com sítios de clivagem proteolíticas específicos, os quais parecem ser essenciais para a adesão (Deutscher *et al*., 2010).

Tabela 1. Adesinas P97 e P102 presentes nos genomas de *M. hyopneumoniae* e *M. flocculare*.

Produto	ID MHP*_7448	ID MFL*
P97 - cópia 1	MHP7448_0198	NP
P102 - cópia 1	MHP7448_0199	NP
P97 - cópia 2	MHP7448_0108	MF00472
P102 - cópia 2	MHP7448_0107	MF00475
P97 - *like*	MHP7448_0272	MF00620
P102 - *like*	MHP7448_0271	MF00623

NP - Não presente

*MHP - *M. hyopneumoniae;* MFL - *M. flocculare*

A partir dos dados de sequenciamento, verificamos que as cópias destas adesinas, presentes no genoma de *M. flocculare* apresentam similaridade considerável com as sequências de *M. hyopneumoniae.* Além do mais, com a análise global do transcritoma de *M. flocculare*, demonstramos que ambas as cópias são transcritas nas condições de cultivo empregadas. Os níveis de expressão, verificados para estes quatro genes são muitos distintos entre os transcritomas de *M. hyopneumoniae* e *M. flocculare.* Enquanto o gene cópia 1 não está presente no genoma de *M. flocculare*, o gene cópia 2 apresentou maior número de *reads* em relação aqueles presentes no transcritoma de *M. hyopneumoniae.* Desse modo, não é possível traçar qualquer relação significativa quanto ao perfil de expressão gênica desse gene em relação à patogenicidade. Por outro lado, a análise comparativa das sequências dessas adesinas, mostra em ambas as cópias de *M. flocculare*, a ausência de regiões de repetição na sequência gênica, o que poderia justificar a baixa capacidade de se aderir e causar lesão e, por conseguinte, causar infecção (Young *et al*., 2000). A ausência dessas regiões de repetição nas adesinas de *M. flocculare* limitaria a capacidade de reconhecimento de sítios receptores nos cílios, o qual parece ser essencial para a ocorrência da doença. Em relação à organização desses genes nos genomas, observamos que as cópias dos mesmos, em *M. flocculare,*

apresentam inversões e rearranjos, em relação à organziação dos ortólogos no genoma de *M. hyopneumoniae.* Inversões e rearranjos ocorrem com frequência em genomas bacterianos (Mushegian & Koonin, 1996; Lathe *et al*., 2000; Vasconcelos *et al*., 2005), sendo muitas vezes, eventos promovidos pela ação de transposases. De acordo com Lathe *et al*. (2000) mudanças na organização e localização gênica, em diferentes espécies bacterianas influenciam na interação e funcionalidade de seus produtos. Portanto, os resultados encontrados, principalmente em nível de organização gênica, podem explicar a presença de ortólogos dessas importantes adesinas e até outras lipoproteínas em *M. flocculare* apesar da ausência de capacidade patogênica.

Uma análise *in silico* do repertório de proteínas de superfície das três espécies foi também realizada. Verificamos que 76 genes são únicos para *M. flocculare*, 69 para *M. hyopneumoniae* e 234 para *M. hyorhinis*. As proteínas de *M. flocculare* não compartilhadas com as outras duas espécies são exclusivamente hipotéticas. Já em *M. hyopneumoniae* e *M. hyorhinis,* mesmo a maioria das proteínas sendo hipotéticas, há proteínas não compartilhadas com funções conhecidas, incluindo as envolvidas no metabolismo do mio-inositol, permeases e o elemento conjugativo integrativo em *M. hyopneumoniae*; e as lipoproteínas variáveis de membrana (Vlps), proteínas componentes do sistema secretório e proteínas que atuam como transportadores em *M. hyorhinis*.

A diversidade de proteínas de superfície em *M. hyorhinis* é maior que a diversidade de *M. hyopneumoniae* e *M. flocculare.* Essa diferença em relação ao repertório de proteínas de superfície pode estar relacionada à singular capacidade infectiva daquela bactéria (Friis, 1971; Ross *et al*., 1973; Rosengarten & Wise, 1990; Friis & Feenstra, 1994; Kazama *et al*., 1994; Kobisch & Friis, 1996; Citti *et al*., 1997; Morita *et al*., 1998; Morita *et al*., 1999; Caron *et al*., 2000; Huang *et al*., 2001; Friis *et*

al., 2002; Timenetsky *et al*., 2006; Goodison *et al*., 2007; Liu & Shou, 2011; Urbanek *et al*., 2011). As lipoproteínas de membrana, denominadas Vlps, são uma família de lipoproteínas consideradas determinantes na antigenicidade de *M. hyorhinis*. Essas proteínas são exclusivas desse micoplasma e sua atividade altamente mutacional permite a adaptação de *M. hyorhinis* ao hospedeiro, dificultando o seu reconhecimento pelo sistema imunológico (Citti *et al*., 1997).

A família destas lipoproteínas é composta por sete distintas cópias, as quais são nomeadas de VlpA à VlpF. A presença destas cópias é muito variável entre cada isolado de *M. hyorhinis*, sendo que no genoma da cepa HUB-1, os sete genes estão caracterizados (Liu *et al*., 2010). É desconhecido o significado da presença de alguns ou todos os genes *vlp* em um determinado isolado de *M. hyorhinis,* da mesma maneira que se, de algum modo a expressão entre eles é relacionada. Com a montagem do transcritoma de *M. hyorhinis* podemos observar que todos os sete genes são expressos nas condições empregadas, porém existe uma significativa diferença quanto ao número de *reads* para cada um deles. *VlpB*, *vlpG* e *vlpE* estão entre os genes com maior representatividade de transcritos, enquanto *vlpF*, *vlpD*, *vlpC* e *vlpA* apresentam um número muito inferior de *reads*. Portanto, apesar de todas as Vlps serem transcritas, existe uma evidente variação nos níveis de expressão entre elas, porém o significado biológico desse efeito é atualmente desconhecido.

Outra lipoproteína encontrada exclusivamente em *M. hyorhinis* é a proteína de membrana externa, p37 (MHR_0625). Essa proteína tem sido fortemente associada ao poder invasivo e de alta proliferação de *M. hyorhinis*. Estudos recentes demonstram a associação da expressão de p37 de *M. hyorhinis* ao desenvolvimento de malignidade nas mais diversas células humanas (Huang *et al*., 2001; Ketcham *et al*., 2005; Goodison *et al*., 2007; Yang *et al*., 2010; Urbanek *et al*., 2011). A associação da comunidade

microbiana, como por exemplo, *M. hyorhinis*, à capacidade invasiva de células eucarióticas está atualmente em discussão. Mesmo com os mecanismos empregados neste complexo sistema ainda não elucidados, as evidências são suficientes para relacionar a atuação de *M. hyorhinis* como um facilitador da invasividade tumoral, por ação da proteína p37. Nos demais genomas de *M. hyorhinis* já sequenciados, ortólogos de p37 estão presentes em todos eles e mantêm alta similaridade de sequência e conservação de organização gênica. Nos genomas de *M. hyopneumoniae* e *M. flocculare* os genes MHP7448_0360 e MF01218 respectivamente, são denominados como "p37-like ABC transporter". Em *Mycoplasmas* essas proteínas compartilham similaridade estrutural (Sippel *et al*., 2009), principalmente na região característica de transportador do tipo ABC. A análise detalhada da estrutura de p37 mostrou que a topologia da mesma é consistente com proteínas de ligação periplasmática com resíduos de tiamina em sua estrutura (Sippel *et al*., 2008; Sippel *et al*., 2009). A caracterização funcional dessa proteína é necessária para a adequada identificação da função dessa estrutura proteica. Por fim, a expressão do gene p37 em *M. hyorhinis* pode estar vinculada com o poder de invasividade/malignidade citado anteriormente, e com a capacidade de *M. hyorhinis* infectar mais de um sítio ou tecido em um mesmo hospedeiro (Kobisch & Friis, 1996; Liu & Shou, 2011; Sippel *et al*., 2008).

Uma análise filogenômica comparativa, empregando possíveis genes ortólogos de *M. flocculare* em 32 isolados de *Mycoplasmas*, foi realizada, com o intuito de traçar uma possível história evolutiva de ancestralidade para as espécies em estudo. As observações feitas, em cada uma das árvores filogenéticas, permite visualizar a ancestralidade comum às espécies *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis*. Com isso, é claramente distinguível o grupo "hyopneumoniae", composto por estas três espécies, juntamente com *Mycoplasma conjunctivae*.

Os genes que codificam produtos envolvidos na via glicolítica, empregada na geração de ATP, estão presentes nos genomas das espécies de micoplasmas aqui analisadas, representando toda essa rota metabólica. Além do mais, em nível de transcritoma, todos estes genes são também transcritos. Interessantemente, algumas proteínas codificadas por esses genes, são descritas como associadas também a outra função – ação de adesinas, influenciando na capacidade patogênica das bactérias (Ling, *et al*., 2004; Encheva *et al*., 2006; Wu *et al*., 2008), além da atuação na via glicolítica.

Proteínas *moonlighting* são proteínas capazes de atuarem em mais de uma função biológica. Sabidamente contribuem para a virulência de diversos patógenos (Henderson & Martin 2011). De um modo geral são proteínas envolvidas em processos metabólicos essenciais, chaperonas, ou ainda envolvidas na resposta a situações adversas do meio, como o estresse, apresentando também funções que contribuem para a virulência. A identificação dessas proteínas com funções adicionais, em geral, ocorre de modo acidental. Uma similar de uma proteína *moonlighting* não necessariamente será também uma *moonlighting*, ou poderá ser para outra função (Gancedo & Flores, 2008). Alguns grupos de proteínas *moonlighting* incluem: i) enzimas da via glicolítica; ii) enzimas de outras vias metabólicas como o ciclo glioxilato; e iii) chaperonas e proteínas que catalisam o enovelamento proteico. A maioria das proteínas com função *moonlight* de virulência está associada à adesão e modulação da atividade de leucócitos (Henderson & Martin 2011).

Uma análise detalhada realizada por Henderson & Martin (2011) mostrou que a enzima gliceraldeído-3-fosfato desidrogenase (GAPDH) possui, além do seu importante papel no estabelecimento da via glicolítica, propriedades de adesina em bactérias do gênero *Streptococcus* e *Staphylococcus*. Além do mais, foi verificada a presença destas enzimas na superfície celular de *M. genitalium* sendo demonstrado seu envolvimento na

ligação à mucina (Alvarez *et al*., 2003). Os dados de transcritoma mostram a transcrição do gene codificador da gliceraldeído-3-fosfato desidrogenase (MHP7448_0588, MF00829 e MHR_0259), nas três espécies de mycoplasmas aqui analisadas.

O papel biológico das chaperonas *moonlighting* varia de agonista à receptor funcional. Pelo uso de cromatografia de afinidade com fribronectina acoplada à coluna de *sepharose*, duas proteínas de *M. pneumoniae*, EF-Tu (fator de alongamento da tradução) e a subunidade B da piruvato desidrogenase, foram identificadas como ligadoras de fibronectina (Dallo *et al*., 2002). EF-Tu é uma importante proteína citoplasmática responsável pelos passos críticos na síntese proteica. Já, a piruvato desidrogenase é um complexo enzimático formado por duas subunidades, as quais catalisam a transformação de piruvato em acetil-CoA para a oxidação mitocondrial (Dallo *et al*., 2002).

Balasubramanian *et al*. (2008) demonstraram, com o emprego de anticorpos específicos, que essas proteínas supracitadas estão presentes na superfície de *M. pneumoniae* e ambos os anticorpos são capazes de inibir a ligação do microrganismo à fibronectina. Atualmente, os mecanismos de regulação das proteínas *moonlighting* são desconhecidos, mas é sabido que interessantemente proteínas similares não necessariamente atuarão com as mesmas funções. Nesse caso, a proteína EF-Tu de *M. genitalium*, compartilha 96% de similaridade com EF-Tu de *M. pneumoniae*, porém não é capaz de se ligar à fibronectina (Henderson & Martin, 2011). A análise comparativa dos transcritomas mostra que, o número de transcritos de EF-Tu em *M. hyopneumoniae* e *M. flocculare* são muito similares, no entanto, em *M. hyorhinis* este gene está entre os mais expressos, com um número de *reads* muito maior, quando comparado com as outras duas espécies. É possível que o mecanismo de defesa bacteriana de proteínas *moonlighting* possa ocorrer em *M. hyopneumoniae*, envolvendo proteínas essenciais

para o metabolismo da célula, também em processos relacionados com a virulência. Além do mais, em *M. hyorhinis* proteínas *moonlighting* podem estar desempenhando função fundamental na capacidade desse micoplasma infectar diferentes sítios e hospedeiros.

Comparando a organização gênica global de *M. flocculare* com *M. hyorhinis* observamos um percentual de conservação de 46%. Este índice não é considerado alto, mas apesar disso, esse percentual refere-se à manutenção total do contexto gênico nas porções analisadas. O entendimento da estrutura organizacional de genomas bacterianos, bem como a atuação dos elementos estruturais e organizacionais permite a compreensão de eventos regulatórios. Em estudo realizado em *M. hyopneumoniae* por Siqueira *et al*. (2011), com o emprego de análises *in silico*, e confirmações experimentais, os autores inferem que o genoma desta bactéria é organizado em grandes unidades transcricionais (UTs ou ORF clusters - OCs), ocorrendo a transcrição contínua (mRNA policistrônico) de um agrupamento gênico, até a ocorrência de genes na fita oposta. Elementos, como promotores transcricionais, internos a estas UTs foram identificados, promovendo a regulação de genes dentro dessas UTs (Weber *et al*., 2012). Em *M. pneumoniae* tais observações já tinham sido descritas, porém em nível de análise de transcritoma total (Güell *et al*., 2009). O modo de organização gênica proposto por Siqueira *et al*. (2011) com base no genoma de *M. hyopneumoniae* 7448, semelhante àquele observado em *M. pneumoniae* por Güell *et al*. (2009), parece ocorrer também em *M. hyorhinis* e *M. flocculare.* Conforme a análise *in silico* apresentada no Capítulo 1, as UTs ou OCs preditas, são conservadas em relação à *M. hyopneumoniae* 7448, quanto ao contexto gênico, fita codificadora em que estão localizadas, número de ORFs que compõem as UTs, regiões intergênicas entre as ORFs, assim como, os produtos codificados pelas mesmas.

Evidências sugerem uma importância funcional da organização gênica em UTs nos genomas (Kelly *et al*., 2012). A realização dos experimentos de análise dos transcritomas totais de *M. hyopneumoniae, M. flocculare* e *M. hyorhinis*, apresentada no Capítulo 2, permitiu a confirmação da predição *in silico* das UTs, em escala genômica para as três espécies de micoplasmas aqui estudadas. Apartir da montagem de cada um dos mapas, podemos visualizar que a transcrição ocorre, principalmente com a formação de grandes mRNAs policistrônicos, formando estruturas de UTs ao longo de todo o genoma. Este evento já havia sido descrito em *M. pneumoniae* (Güell *et al*., 2009), também com o emprego de ferramentas de sequenciamento de RNA total; e assim, podemos sugerir que esse seja um modo comum de organização da transcrição gênica em micoplasmas. Desse modo, o mapa transcricional gerado para cada uma das três espécies analisadas, além da identificação das UTs, representa um passo importante para a consolidação do entendimento da coordenação da regulação da transcrição em *Mycoplasma* spp.

6. CONCLUSÕES

- Nós realizamos o primeiro sequenciamento do genoma de *M. flocculare* (ATCC 27716), juntamente com o genoma de um novo isolado de *M. hyopneumoniae* (7422);
- A análise comparativa entre três genomas de mycoplasmas que habitam o trato respiratório suíno - *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis* - permitiu identificar uma alta sintenia entre os genomas de *M. hyopneumoniae* e *M. flocculare*, ocorrendo pequenas diferenças, quanto à presença de alguns genes e localização ou organização genômica. Por tanto, estas diferenças podem estar envolvidas na diferenciação entre uma espécie ser comensal - *M. flocculare*, e a outra ser patogênica- *M. hyopneumoniae.*
- A sintenia entre os três genomas analisados foi suficiente para determinar que *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis* são relacionados filogeneticamente, pertencendo ao mesmo grupo filogenético – grupo hyopneumoniae.
- Um grande repertório de proteínas de superfície é compartilhado entre as três espécies de mycoplasmas. Mesmo se tratando de espécies diferentes essa similaridade entre a presença de proteínas de superfície pode ser esperada, visto que *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis* ocupam o mesmo nicho.
- O sequenciamento de RNA em larga escala – transcritoma – foi realizado com sucesso nas três espécies que habitam o trato respiratório suíno: *M. hyopneumoniae*, *M. flocculare* e *M. hyorhinis.* Com isso, foi estimada a abundância absoluta de transcritos para cada uma das espécies.
- Nas condições de cultivo empregadas, a maior parte dos genes presentes em cada um dos genomas é transcrito.

- As adesinas denominadas P97 cópia 1 e P102 cópia 1, as quais são exclusivas de *M. hyopneumoniae* são também transcritas, e como o repertório de transcritos é muito similar entre *M. hyopneumoniae* e *M. flocculare*, nós acreditamos que as adesinas supracitadas podem estar fortemente relacionadas com a capacidade patogênica de *M. hyopneumoniae.*
- Genes relacionados à capacidade altamente infectiva, mutativa e de infectar vários sítios, de *M. hyorhinis* estão sendo expressos nas condições empregadas.
- Estruturas de unidades transcricionais (UTs) foram identificadas como o modo preferencial, em escala transcritômica, de organização da transcrição em mycoplasmas.

7. PERSPECTIVAS

- Re-analisar as novas sequências de transcritoma e gerar os dados finais para o Capítulo 2;
- Analisar estruturalmente e funcionalmente os genes que compõem os blocos que representam diferenças ou rearranjos entre *M. hyopneumoniae* e *M. flocculare*.
- Analisar funcionalmente genes com função ainda desconhecida (hipotéticos) que apresentam alto índice de expressão nos transcritomas estudados.
- Mapear experimentalmente, em escala genômica, os sítios de início da transcrição (TSS) para as três espécies de mycoplasmas, através de RNA-seq com metodologia de enriquecimento da região 5'do mRNA em fita específica.
- Mapear os elementos funcionais identificados por transcritoma de fita específica, como os RNAs anti-senso.
- Realizar estudos de caracterização funcional, com genes alvos, objetivando identificar função *moonlight* em proteínas de *M. hyopneumoniae*.

REFERÊNCIAS BIBLIOGRÁFICAS

Adams C, Pitzer J, Minion FC. *In Vivo* Expression Analysis of the P97 and P102 Paralog Families of *Mycoplasma hyopneumoniae.* Infect. Immun. 73(11):7784–7787, 2005.

Alvarez RA, Blaylock MW, Baseman JB. Surface localized glyceraldehyde-3-phosphate dehydrogenase of *Mycoplasma genitalium* binds mucin. Mol. Microbiol. 48:1417–1425, 2003.

Andersson SG, Kurland CG. Reductive evolution of resident genomes. Trends Microbiol. 6:263-268, 1998.

Andre G, Even S, Putzer H, Burguière P, Croux C, Danchin A, Martin-Verstraete I, Soutourina O. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. Nucleic Acids Res. 36:5955–5969, 2008.

Armstrong CH, Sands-Freeman L, Freeman MJ. Serological, pathological and cultural evaluations of swine infected experimentally with *Mycoplasma flocculare*. Can. J. vet. Res. 51:185-188, 1987.

Balasubramanian S, Kannan TR, Baseman JB. The surface exposed carboxyl region of *Mycoplasma pneumoniae* elongation factor Tu interacts with fibronectin. Infect. Immun. 76:3116–3123, 2008.

Baseman JB, Lange M, Criscimagna NL, Giron JA, Thomas CA. Interplay between mycoplasmas and host target cells. Microb Pathog. 19:105-116, 1995.

Baseman JB, Tully JG. *Mycoplasmas*: sophisticated, reemerging, and burdened by their notoriety. Emerg. Infect. Dis. 3(1):21-32, 1997.

Bhugra B, Voelker LL, Zou NX, Yu HL, Dybvig K. Mechanism of antigenic variation in *Mycoplasma pulmonis*: interwoven, site-specific DNA inversions. Mol. Microbiol. 18:703–714, 1995.

Blank WA, Stemke GW. A physical and genetic map of the *Mycoplasma flocculare* ATCC 27716 chromosome reveals large genomic inversions when compared with that of *Mycoplasma hyopneumoniae* strain J. Int J Syst Evol Microbiol. 51:1395-1399, 2001.

Bogema DR, Deutscher AT, Woolley LK, Seymour LM, Raymond BB, Tacchi JL, Padula MP, Dixon NE, Minion FC, Jenkins C, Walker MJ, Djordjevic SP. Characterization of cleavage events in the multifunctional cilium adhesin Mhp684 (P146) reveals a mechanism by which *Mycoplasma hyopneumoniae* regulates surface topography. MBio 3(2):e00282-11, 2012.

Boni IV, Artamonova VS, Tzareva NV, Dreyfus M. Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. EMBO J. 20:4222–4232, 2001.

Bove JM. Molecular features of mollicutes. Clinical Infection Disease 17(1):S10-S31, 1993.

Calcutt MJ, Foecking MF, Rosales RS, Ellis RJ, Nicholas RA. Genome sequence of *Mycoplasma hyorhinis* strain GDL-1. J. Bacteriol. 194(7):1848, 2012.

Caron J, Ouardani M, Dea S. Diagnosis and Differentiation of *Mycoplasma hyopneumoniae* and *Mycoplasma hyorhinis* Infections in Pigs by PCR Amplification of the p36 and p46 Genes. J. Clin. Microbiol. 38(4):1390–1396, 2000.

Cho BK, Knight EM, Barrett CL, Palsson BO. Genome-wide analysis of Fisbinding in *Escherichia coli* indicates a causative role for A-/AT-tracts. Genome Research 18:900–910, 2008.

Ciprian A, Pijoan C, Cruz T, Camacho J, Tortora J, Colmenares G, Lopez-Revilla R, De La GM. *Mycoplasma hyopneumoniae* increases the susceptibility of pigs to experimental *Pasteurella multocida* pneumonia. Can. J. Vet. Res. 52(4):434-438, 1988.

Citti C, Kim MF, Wise KS. Elongated Versions of Vlp Surface Lipoproteins Protect *Mycoplasma hyorhinis* Escape Variants from Growth-Inhibiting Host Antibodies, Infection and Immunity 65(5):1773–1785, 1997.

Citti C, Mac Kown RW, Droesse M, Wise K S. Gene Families Encoding Phase-and Size-Variable Surface Lipoproteins of *Mycoplasma hyorhinis*. J. Bacteriol. 182(5):1356–1363, 2000.

Clark J. Mycoplasmas: identifying hosts for a stealth pathogen. The Veterinary Journal, 170(3):273-274, 2005.

Cole BC, Washburn LR, Taylor-Robinson D. Mycoplasma-induced arthritis. In: Razin S, Barile MF. Mycoplasma pathogenicity. Vol. IV. Academic Press, Orlando, Florida. pg.107-160, 1985.

Condon C. Maturation and degradation of RNA in bacteria. Curr. Opin. Microbiol. 10:271–278, 2007.

Dallo SF, Kannan TR, Blaylock MW, Baseman JB. Elongation factor Tu and E1 subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. Mol. Microbiol. 46:1041–1051, 2002.

de Hoon MJ, Makita Y, Nakai K, Miyano S. Prediction of transcriptional terminators in *Bacillus subtilis* and related species. PLoS Comput. Biol. 1:e25, 2005.

Debey M, Jacobson CD, Ross RF. Histochemical and morphologic changes of porcine airway epithelial cells in response to infection with *Mycoplasma hyopneumoniae*. Am. J. Vet. Res. 53:1705-1710, 1992.

Deutscher MP. Degradation of stable RNA in bacteria. J. Biol. Chem. 278:45041–45044, 2003.

Deutscher AT, Jenkins C, Minion FC, Seymour LM, Padula MP, Dixon NE, Walker MJ, Djordjevic SP. Repeat regions R1 and R2 in the P97 paralogue Mhp271 of *Mycoplasma hyopneumoniae* bind heparin, fibronectin and porcine cilia. Mol. Microbiol. 78:444-458, 2010.

Djordjevic SP, Cordwell SJ, Djordjevic MA, Wilton J, Minion FC. Proteolytic processing of the *Mycoplasma hyopneumoniae* cilium adhesin. Infection Immunity 72(5):2791-2802, 2004.

Drexler HG, Uphoff CC. Mycoplasma contamination of cell cultures: incidence, sources, effects, detection, elimination, prevention. Cytotechnology 39:75–90, 2002.

Dudler R, Schmidhauser C, Parish RW, Wettenhall RE, Schmidt T. A mycoplasma high-affinity transport system and the in vitro invasiveness of mouse sarcoma cells. EMBO J. 7:3963-3970, 1988.

Dybvig K, Voelker LL. Molecular biology of mycoplasmas. Annual Review Microbiology 50:25-57, 1996.

Elkind E, Rechnitzer H, Vaisid T, Kornspan JD, Barnoy S, Rottem S, Kosower NS. *Mycoplasma hyorhinis* upregulates calpastatin and inhibits calpain dependent proteolysis in SH-SY5Y neuroblastoma cells. FEMS Microbiol. Lett. 304: 62–68, 2010.

Elkind E, Vaisid T, Kornspan JD, Barnoy S, Rottem S, Kosower NS. Neuroprotective effects of *Mycoplasma hyorhinis* against amyloid-b-peptide toxicity in SH-SY5Y human neuroblastoma cells are mediated by calpastatin upregulation in the mycoplasma-infected cells. Neurochem. Internat. 58:497–503, 2011.

Elkind E, Vaisid T, Kornspan JD, Barnoy S, Rottem S, Kosower NS. Calpastatin upregulation in *Mycoplasma hyorhinis*-infected cells is promoted by the mycoplasma lipoproteins via the NF-κB pathway. Cell Microbiol. 14(6):840-851, 2012.

Encheva V, Gharbia SE, Wait R, Begum S, Shah HN. Comparison of extraction procedures for proteome analysis of *Streptococcus pneumoniae* and a basic reference map. Proteomics 6:3306–3317, 2006.

Eyre-Walker A. The distance between *Escherichia coli* genes is related to gene expression levels. J. Bacteriol. 177:5368-5369, 1995.

Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic acids res. 34(3):e22, 2006.

Finch LR, Mitchell A. Sources of nucleotides In: Maniloff J, McElhaney RN, Finch LR, Baseman JB. *Mycoplasmas*: Molecular Biology and Pathogenesis. Washington, DC: Am. Soc. Microbiol. p.211-230, 1992.

Fondi M, Emiliani G, Fani R. Origin and evolution of operons and metabolic pathways, Research in Microbiology 160:502-512, 2009.

Fraser CM, Gocayne JD, White O, et al. The minimal gene complement of *Mycoplasma genitalium*. Science 270:397–403, 1995.

Friis NF. *Mycoplasma hyorhinis* as a causative agent in pneumonia of pigs. Acta vet. scand. 12:116-119, 1971.

Friis NF. Isolation and characterization of a new porcine mycoplasma. Acta vet. Scand. 13:284-286, 1972.

Friis NF. The pathogenicity of *Mycoplasma flocculare*. Acta vet. scand. 14: 344-346, 1973.

Friis NF. *Mycoplasma suipneumoniae* and *Mycoplasma flocculare* in comparative pathogenicity studies. Acta vet. scand. 15:507-518, 1974.

Friis NF. Some recommendations concerning primary isolation of *Mycoplasma suipneumoniae* and *Mycoplasma flocculare* a survey. Nordisk Veterinaer. Medicin. 27(6):337-339, 1975.

Friis NF, Feenstra AA. *Mycoplasma hyorhinis* in the etiology of serositis among piglets. Acta. vet. scand. 35:93-98, 1994.

Friis NF, Kokotovic B, Svensmark B. *Mycoplasma hyorhinis* isolation from cases of otitis media in piglets. Acta Vet Scand. 43(3):191-193, 2002.

Fujita K, Baba T, Isono K. Genomic analysis of the genes encoding ribosomal proteins in eight eubacterial species and *Saccharomyces cerevisiae*. Genome Informatics 9:3-12, 1999.

Gancedo C, Flores CL. Moonlighting proteins in yeasts. Microbiol. Mol. Biol. Rev. 72:197–210, 2008.

Gardner SW, Minion F. Detection and Quantification of Intergenic Transcription in *Mycoplasma hyopneumoniae.* Microbiology 156:2305-2315, 2010.

Gil R, Silva FJ, Pereto J, Moya A. Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. 68(3):518-537, 2004.

Gois M, Kuksa F. Intranasal infection of gnotobiotic piglets with *M. hyorhinis*: differences in virulence of the strains and influence of age on the development of infection. Zbl. Vet. Med. B. 21:352-361, 1974.

Goodison S, Nakamura K, Iczkowski KA, Anai S, Boehlein SK, Rosser CJ. Exogenous mycoplasmal p37 protein alters gene expression, growth and morphology of prostate cancer cells. Cytogenet Genome Res. 118:204–213, 2007.

Goodison S, Urquidi V, Kumar D, Reyes L, Rosser CJ. Complete genome sequence of *Mycoplasma hyorhinis* strain SK76. Genome Announc. 1(1):e00101-12, 2013.

Güell M, Noort V, Yus E, Chen WH, Leigh-Bell J, Michalodimitrakis K, Yamada T, Arumugam M, Doerks T, Kühner S, Rode M, Suyama M, Schmidt S, Gavin AC, Bork P, Serrano L. Transcriptome Complexity in a Genome-Reduced Bacterium. Science 326(5957):1268-1271, 2009.

Güell M, Yus E, Lluch-Senar M, Serrano L. Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nat. Rev. Microbiol. 9:658-69, 2011.

Halbedel S, Eilers H, Jonas B, Busse J, Hecker M, Engelmann S, Stulke J. Transcription in *Mycoplasma pneumoniae*: Analysis of the Promoters of the ackA and ldh Genes. J. Mol. Biol. 371(3):596-607, 2007.

Hazkani-Covo E, Graur D. Evolutionary conservation of bacterial operons: Does transcriptional connectivity matter? Genetica 124:145–166, 2005.

Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infect. Immun. 79(9):3476-3491, 2011

Himmelreich R, Plagens H, Hilbert H, Reiner B, Herrmann R. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. Nucl. Acids Res. 25:701–712, 1997.

Huang S, Li JY, Wu J, Meng L, Shou CC. Mycoplasma infections and different human carcinomas. World Gastroentero 7:266-269, 2001.

Inamine JM, Loechel S, Hu PC. Analysis of the nucleotide sequence of the P1 operon of *Mycoplasma pneumoniae*. Gene 73:175–183, 1988.

Jacob F, Monod J. Genes of structure and genes of regulation in the biosynthesis of proteins. C. R. Hebd Seances Acad. Sci. 249:1282–1284, 1959.

Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3:318-356, 1961.

Jenkins C, Wilton JL, Minion FC, Falconer L, Walker MJ, Djordjevic SP. Two domains within the *Mycoplasma hyopneumoniae* cilium adhesin bind heparin. Infect. Immun. 74:481-487, 2006.

Kazama S, Yagihashi T, Morita T, Awakura T, Shimada A, Umemura T. Isolation of *Mycoplasma hyorhinis* and *Mycoplasma arginini* from the ears of pigs with otitis media. Res. Vet. Sci. 56(1):108-110, 1994.

Kelly S, Kramer SA, Schwede PK, Maini PK, Gull K, Carrington M. Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in *trypanosomes.* Open Biol. 2:120033, 2012.

Ketcham CM, Anai S, Reutzel R, Sheng S, Schuster SM, Brenes RB, Agbandje-McKenna M, McKenna R, Rosser CJ, Boehlein SK. p37 induces tumor invasiveness. Mol. Cancer Ther. 4:1031-1038, 2005.

Kobisch M. Pathogenicity of *Mycoplasma hyorhinis.* Yale J. Biol. Med. 56(5):922-923, 1983.

Kobisch M, Friis NF. Swine mycoplasmoses. Rev. Sci. Tech. Oie. 15:1569-1605, 1996.

Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS. Prevalence of transcription promoters within archaeal operons and coding sequences. Mol. Syst. Biol. 5:285, 2009.

Kornspan JD, Lysnyansky I, Kahan T, Herrmann R, Rottem S, Nir-Paz R. Genome Analysis of a *Mycoplasma hyorhinis* Strain Derived from a Primary Human Melanoma Cell Line. J.Bacteriol. 193(17):4543–4544, 2011.

Kumar R, Lawrence ML, Watt J, Cooksey AM, Burgess SC, Nanduri B. RNA-Seq Based Transcriptional Map of Bovine Respiratory Disease Pathogen ‘‘*Histophilus somni* 2336’’. PLoS ONE 7(1): e29435, 2012.

Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, Castaño-Diez D, Chen WH, Devos D, Güell M, Norambuena T, Racke I, Rybin V, Schmidt A, Yus E, Aebersold R, Herrmann R, Böttcher B, Frangakis AS, Russell RB, Serrano L, Bork P, Gavin AC. Proteome organization in a genome-reduced bacterium. Science 326(5957):1235-1240, 2009.

Lathe WC, Snel B, Bork P. Gene context conservation of a higher order than operons. Trends Biochem. Sci. 25(10):474–479, 2000.

Laursen BS, Sorensen HP, Mortensen KK, Sperling-Petersen HU. Initiation of protein synthesis in bacteria. Microbiol. Mol. Biol. Rev. 69:101–123, 2005.

Lim JS, Choi BS, Lee JS, Shin C, Yang TJ, Rhee JS, Lee JS, Choi IY. Survey of the Applications of NGS to Whole-Genome Sequencing and Expression Profiling. Genomics & Informatics 10(1):1-8, 2012.

Ling E, Feldman G, Portnoi M, Dagan R, Overweg K, Mulholland F, Chalifa-Caspi V, Wells J, Mizrachi-Nebenzahl Y. Glycolytic enzymes associated with the cell surface of *Streptococcus pneumoniae* are antigenic in humans and elicit protective immune responses in the mouse. Clin. Exp. Immunol. 138:290–298, 2004.

Liu W, Fang L, Li S, Li Q, Zhou Z, Feng Z, Luo R, Shao G, Wang L, Chen H, Xiao S. Complete genome sequence of *Mycoplasma hyorhinis* strain HUB-1. J. Bacteriol. 192(21):5844-5845, 2010.

Liu W, Shou C. *Mycoplasma hyorhinis* and *Mycoplasma fermentans* induce cell apoptosis and changes in gene expression profiles of 32D cells. Biol. Res. 44:383-391, 2011.

Liu W, Feng Z, Fang L, Zhou Z, Li Q, Li S, Luo R, Wang L, Chen H, Shao G, Xiao S. Complete genome sequence of *Mycoplasma hyopneumoniae* strain 168. J. Bacteriol. 193(4):1016–1017, 2011.

Lo SC, Hayes MM, Kotani H, Pierce PF, Wear DJ, Newton PB 3rd, Tully JG, Shih JW. Adhesion onto and invasion into mammalian cells by *Mycoplasma penetrans*: a newly isolated mycoplasma from patients with AIDS. Mod. Pathol. 6:276-280, 1993.

Lysnyansky I, Rosengarten R, Yogev D. Phenotypic switching of variable surface lipoproteins in *Mycoplasma bovis* involves high-frequency chromosomal rearrangements. J. Bacteriol. 178:5395–5401, 1996.

Ma JC, Newman AJ, Hayward RS. Internal promoters of the *rpoBC* operon of *Escherichia coli*. Mol. Gen. Genet. 184:548–550, 1981.

Madeira HM, Gabriel JE. Regulation of gene expression in Mycoplasmas: Contribution from *Mycoplasma hyopneumoniae* and *Mycoplasma synoviae* genome sequences. Genetics and Molecular Biology 30(1):277-282, 2007.

Madsen ML, Nettleton D, Thacker EL, Minion C. Transcriptional profiling of *Mycoplasma hyopneumoniae* during iron depletion using microarrays. Microbiology 152:937–944, 2006a.

Madsen ML, Nettleton D, Thacker EL, Edwards R, Minion FC. Transcriptional profiling of *Mycoplasma hyopneumoniae* during heat shock using microarrays. Infect. Immun. 74:160–166, 2006b.

Madsen ML, Puttamreddy S, Thacker EL, Carruthers MD, Minion C. Transcriptome Changes in *Mycoplasma hyopneumoniae* during infection. Infec. Immun. 76: 658–663, 2008.

Makita Y, Nakao M, Ogasawara N, Nakai K. DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. Nucleic Acids Res. 32:D75–D77, 2004.

Marguerat S, Bahler J. RNA-seq: from technology to biology. Cell. Mol. Life Sci. 67:569–579, 2010.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380, 2005.

Manolukas JT, Barile MF, Chandler DK, Pollack JD. Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in mollicutes. J. Gen. Microbiol. 134(3):791-800, 1988.

Markham PF, Glew MD, Sykes JE, Bowden TR, Pollocks TD, Browning GF, Whithear KG, Walker ID. The organisation of the multigene family which encodes the major cell surface protein, pMGA, of *Mycoplasma gallisepticum*. FEBS Lett. 352:347–352, 1994.

Martin JA, Wang Z. Next-generation transcriptome assembly. Nature Rev. Genet. 12:671-682, 2011.

Mckernan K, Blanchard A, Kotler L, Costa, G. Reagents, methods, and libraries for bead-based sequencing. US patent application 20080003571, 2006.

Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R. Microbiology in the post-genomic era. Nature Rev. Microbiol. 6:419–430, 2008.

Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM, Mahairas GG. The genome sequence of *Mycoplasma hyopneumoniae* strain 232, the agent of swine mycoplasmosis. J. Bacteriol. 186:7123-7133, 2004.

Moreno-Hagelsieb G, Collado-Vides J. A powerful nonhomology method for the prediction of operons in prokaryotes. Bioinformatics 18(1):329-336, 2002.

Morita T, Muraki Y, Awakura T, Shimada A, Umemura T. Detection of *Mycoplasma hyorhinis* in porcine eustachitis. J. Vet. Med. Sci. 55(3):475-477, 1993.

Morita T, Fukuda H, Awakura T, Shimada A, Umemura T, Kazama S, Yagihashi T. Demonstration of *Mycoplasma hyorhinis* as a possible primary pathogen for porcine otitis media. Vet. Pathol. 32:107-111, 1995.

Morita T, Sasaki A, Kaji N, Shimada A, Kazama S, Yagihashi T, Umemura T. Induction of temporary otitis media in specific-pathogen-free pigs by intratympanic inoculation of *Mycoplasma hyorhinis*. Am. J. Vet. Res. 59(7):869-873, 1998.

Morita T, Ohiwa S, Shimada A, Kazama S, Yagihashi T, Umemura T. Intranasally inoculated *Mycoplasma hyorhinis* causes eustachitis in pigs. Vet Pathol. 36(2):174-178, 1999.

Musatovova O, Dhandayuthapani S, Baseman JB. Transcriptional starts for cytadherence-related operons of *Mycoplasma genitalium*. FEMS Microbiology Letters 229(1):73-81, 2003.

Mushegian AR, Koonin EV. Gene order is not conserved in bacterial evolution. Trends Genet. 12:289–290, 1996.

Oneal MJ, Schafer ER, Madsen ML, Minion C. Global transcriptional analysis of *Mycoplasma hyopneumoniae* following exposure to norepinephrine. Microbiology 154:2581–2588, 2008.

Osawa S, Jukes TH, Watanabe K, Muto A. Recent evidence for evolution of the genetic code. Microbiology Review 56(1):229-264, 1992.

Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nature Rev. Genet. 12:87–98, 2011.

Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH. Structure and Complexity of a Bacterial Transcriptome. J.Bacteriol. 191(10):3203–321, 2009.

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella* Typhi. PLoS Genet. 5:e1000569, 2009.

Pitcher DG, Nicholas RA. Mycoplasma host specificity: fact or fiction? Veterinary Journal 170(3):300-306, 2005.

Pollack JD. Carbohydrate metabolism and energy conservation. In: Maniloff J, Mcelhaney RN, Finch LR, Baseman JB. *Mycoplasmas*: Molecular Biology and Pathogenesis. Washington, DC. Am. Soc. Microbiol. p.181-200, 1992.

Pollack JD. Central carbohydrate pathways: metabolic flexibility and the extra role of some housekeeping enzymes. In: Razin S, Herrmann R. Molecular Biology and Pathogenicity of Mycoplasmas. New York, Kluwer. p.163–201, 2002.

Price MN, Huang KH, Arkin AP, Alm EJ. A novel method for accurate operon predictions in all sequenced prokaryotes. Nucl. Acid. Res. 33:880-892, 2005.

Price MN, Arkin AP, Alm EJ. The life-cycle of operons. PLoS Genetics 2(6):859-873, 2006.

Rasmussen S, Nielsen HB, Jarmer H. The transcriptionally active regions in the genome of *Bacillus subtilis*. Mol. Microbiol. 73:1043–1057, 2009.

Rautiainen E, Wallgren P. Aspects of the Transmission of Protection against *Mycoplasma hyopneumoniae* from Sow to Offspring. J. Vet. Medic. 48:55-65, 2001.

Razin S, Yogev D, Naot Y. Molecular biology and pathogenicity of mycoplasmas. Microbiol. Molec. Biol. Rev. 62:1094-1156, 1998.

Razin S. The Genus *Mycoplasma* and Related Genera (Class Mollicutes) In: Dworkin M, Falkow S, Rosenberg E, Schleifer K, Stackbrandt E. The Prokaryotes. New York: Springer. p.836-904, 2006.

Rogozin IB, Makarova KS, Murvai J, Czabarka E, Wolf YI. Connected gene neighborhoods in prokaryotic genomes. Nucl. Acid. Res. 30:2212–2223, 2002.

Ross R, Dale S, Duncan J. Experimentally induced *Mycoplasma hyorhinis* arthritis of swine: immune response to 26th postinoculation week. Am. J. Vet. Res. 34:367-373, 1973.

Ross RF. Mycoplasmal diseases. In: Leman AD, Straw B, Mengeling W, D'Allaire S, Taylor D. Diseases of swine. 7th Ed. Iowa State University Press, Ames, Iowa. pg.537-551, 1992.

Rosengarten R, Wise KS. Phenotypic switching in mycoplasmas: phase variation of diverse surface lipoproteins. Science 247:315-318, 1990.

Rosengarten R, Wise KS. The Vlp system of *Mycoplasma hyorhinis*: combinatorial expression of distinct size variant lipoproteins generating high-frequency surface antigenic variation. J. Bacteriol. 173:4782–4793, 1991.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356)**:**348-352, 2011.

Rottem S, Yogev D. Mycoplasma interaction with eukaryotic cells. Subcellular Biochemestry 33:199-227, 2000.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467, 1977.

Santangelo TJ, Artsimovitch I. Termination and antitermination: RNA polymerase runs a stop sign. Nature Rev. Microbiol. 9:319–329, 2011.

Sasaki Y, Ishikawa J, Yamashita A, Oshima K, Kenri T, Furuya K, Yoshino C, Horino A, Shiba T, Sasaki T, Hattori M. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. Nucl. Ac. Res. 30(23):5293-5300, 2002.

Schafer ER, Oneal MJ, Madsen ML, Minion FC. Global transcriptional analysis of *Mycoplasma hyopneumoniae* following exposure to hydrogen peroxide. Microbiology 153:3785-3790, 2007.

Seymour LM, Jenkins C, Deutscher AT, Raymond BBA, Padula MP, Tacchi JL, Bogema DR, Eamens GJ, Woolley LK, Dixon NE, Walker MJ, Djordjevic SP.Mhp182 (P102) binds fibronectin and contributes to the recruitment of plasmin(ogen) to the *Mycoplasma hyopneumoniae* cell surface. Cell Microbiol. 14(1):81–94, 2012.

Sigmund CD, Morgan EA. Nus A protein affects transcriptional pausing and termination *in vitro* by binding to different sites on the transcription complex. Biochemistry 27:5622–5627, 1988.

Sippel KH, Robbins AH, Reutzel R, Domsic J, Govindasamy L, McKenna MA, Rosser CJ, McKenna R. Structure determination of the cancer associated *Mycoplasma hyorhinis* protein Mh-p37**.** Biological Crystallography D64:1172–1178, 2008.

Sippel KH, Robbins AH, Reutzel R, Boehlein SK, Namiki K, Goodison S, McKenna MA, Rosser CJ, McKenna R. Structural Insights into the Extracytoplasmic Thiamine-Binding Lipoprotein p37 of *Mycoplasma hyorhinis*. J. Bacteriol. 191(8):2585-2592, 2009.

Siqueira FM, Schrank A, Schrank IS. *Mycoplasma hyopneumoniae* transcription unit organization: genome survey and prediction. DNA Res. 18:413–422, 2011.

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TY, Hinton JCD, Vogel J. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. 4(8):e1000163, 2008.

Skamrov A, Feoktistova E, Goldman M, Beabealashvilli R. *Mycoplasma gallisepticum rpoA* gene cluster. FEMS Microbiology Letters 208(2):281–285, 2002.

Sobestiansky J, Barcellos DESN, Mores N, Carvalho LF, Oliveira SJ. Clínica e Patologia Suína. Goiânia: Art 3 impressos especiais, 1999.

Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. Molecular Systems Biology 9:640, 2013.

Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet. 11:9-16, 2010.

Stemke GW, Mcintyre DJ, Kenneth LR, Stemler ME, Robertson JA. Guanine-plus-cytosine compositions of and deoxyribonucleic acid hybridization comparisons between *Mycoplasma hyopneumoniae* and *Mycoplasma flocculare*. Int. J. syst. Bacteriol. 35:527-529, 1985.

Stemke GW, Laigret F, Grau O, Bové JM. Phylogenetic Relationships of Three Porcine Mycoplasmas, *Mycoplasma hyopneumoniae*, *Mycoplasma flocculare* and *Mycoplasma hyorhinis*, and Complete 16S rRNA Sequence of *M. flocculare*. Int. J. Syst. Bacteriol. 42:220-225, 1992.

Stemke GW, Huang Y, Laigret F, Bove JM. Cloning the ribosomal RNA operons of *Mycoplasma flocculare* and comparison with those of *Mycoplasma hyopneumoniae.* Microbiology 140:857-860, 1994.

Strasser M, Nicolet J. Evaluation of an ELISA for *Mycoplasma flocculare*. Zbl. Bakt. 20:848-851, 1990.

Strasser M, Abiven R, Kobisch M, Nicolet J. Immunological and pathological reactions in piglets experimentally infected with *Mycoplasma hyopneumoniae* and/or *Mycoplasma flocculare*. Vet. Immun. Immunopath. 31:141-153, 1992.

Thacker EL. Diagnosis of *Mycoplasma hyopneumoniae*. Anim. Health Res. Rev. 5(2):317-320, 2004.

Timenetsky J, Santos LM, Buzinhani M, Mettifogo E. Detection of multiple mycoplasma infection in cell cultures by PCR. Braz. J. Med. Biol. Res. 39:907–914, 2006.

Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Régnault B, Coppée JY, Lecuit M, Johansson J, Cossart P. The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 459:950–956, 2009.

Turcatti, G. Romieu A, Fedurco M, Tairi AP. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic acids res. 36(4):e25, 2008.

Urbanek C, Goodison S, Chang M, Porvasnik S, Sakamoto N, Li C, Boehlein SK, Rosser CJ. Detection of antibodies directed at *M. hyorhinis* p37 in the serum of men with newly diagnosed prostate cancer. BMC Cancer 11:233, 2011.

Vasconcelos ATR, Ferreira HB, Bizarro CV, Bonatto SL, 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.

Vicente M, Gomez MJ, Ayala JA. Regulation of transcription of cell division genes in the *Escherichia coli* dcw cluster. Cell Molecular Life Science 54:317-324, 1998.

Waldo III RH, Krause DC. Synthesis, Stability, and Function of Cytadhesin P1 and Accessory Protein B/C Complex of *Mycoplasma pneumoniae*. J. Bacteriol. 188(2):569–575, 2006.

Waldo RH, Popham PL, Romero-Arroyo CE, Mothershed EA, Lee KK, Krause DC. Transcriptional analysis of the *hmw* gene cluster of *Mycoplasma pneumoniae*. J. Bacteriol. 181:4978–4985, 1999.

Walker RL. Mollicutes. In: Hirsh DC, Zee YC. Microbiologia Veterinária. Rio de Janeiro: Guanabara Koogan. p.155-162, 2003.

Wang Z, Gerstein M, Snyder M. RNA-seq: a revolutionary tool for transcriptomics. Nature Rev. Genet. 10:57–63, 2009.

Washio TJ, Sasayama J, Tomita M. Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. Nucl. Acid. Res. 26:5456–5463, 1998.

Weber SS, Sant'Anna FH, Schrank IS. Unveiling *Mycoplasma hyopneumoniae* Promoters: Sequence Definition and Genomic Distribution. DNA Research 19(2):103-115, 2012.

Weiner III J, Herrmann R, Browning GF. Transcription in *Mycoplasma pneumoniae*. Nucleic Acids Res. 28:4488-4496, 2000.

Wilhelm BT, Landry JR. RNA-seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249–257, 2009.

Wilton J, Jenkins C, Cordwell SJ, Falconer L, Minion FC, Oneal DC, Djordjevic MA, Connolly A, Barchia I, Walker MJ, Djordjevic SP. Mhp493 (P216) is a proteolytically processed, cilium and heparin binding protein of *Mycoplasma hyopneumoniae*. Mol. Microbiol. 71:566–582, 2009.

Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419:952–956, 2002.

Wise KS, Yogev D, Rosengarten R. Antigenic variation. In: Maniloff J, McElhaney RN, Finch LR, Baseman JB. Mycoplasmas: molecular biology and pathogenesis. American Society for Microbiology, Washington, DC. p.473–489, 1992.

Wise KS. Adaptive surface variation in mycoplasmas. Trends Micro-biol. 1:59–63, 1993.

Woese CR. Bacterial evolution. Microbiol. Rev. 51(2):221-271, 1987.

Wolf M, Muller T, Dandekar T, Pollack JD. Phylogeny of Firmicutes with special reference to *Mycoplasma* (Mollicutes) as inferred from phosphoglycerate kinase amino acid sequence data. Int. J. Syst. Evolut. Microbiol. 54(3):871-875, 2004.

Wu Z, Zhang W, Lu C. Immunoproteomic assay of surface proteins of *Streptococcus suis* serotype 9. FEMS Immunol. Med. Microbiol. 53:52–59, 2008.

Wurtzel O, Sapra R, Chen F, Zhu Yiwen, Simmons BA, Sorek R. A single-base resolution map of an archaeal transcriptome. Genome Res. 20:133–141, 2009.

Yamao F, Muto A, Kawauchi Y, Iwami M, Iwagami S, Azumi Y, Osawa S. UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Nat. Acad. Sc. USA 82(8):2306-2309, 1985.

Yang H, Qu L, Ma H, Chen L, Liu W, Liu C, Meng L, Wu J, Shou C. *Mycoplasma hyorhinis* infection in gastric carcinoma and its effects on the malignant phenotypes of gastric cancer cells. BMC Gastroenterol. 10:132, 2010.

Yoder-Himes DR, Chain PSG, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. Proc. Natl Acad. Sci. USA 106:3976–3981, 2009.

Yogev D, Rosengarten R, Watson-McKown R, Wise KS. 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.

Yogev D, Watson-McKown R, R. Rosengarten R, Wise KS. Increased structural and combinatorial diversity in an extended family of genes encoding Vlp surface proteins of *Mycoplasma hyorhinis*. J. Bacteriol. 177:5636–5643, 1995.

Young TF, Thacker EL, Erickson BZ, Ross RF. A tissue culture system to study respiratory ciliary epithelial adherence of selected swine mycoplasmas. Vet. Microbiol. 71:269-279, 2000.

Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, et al. Impact of Genome Reduction on Bacterial Metabolism and Its Regulation. Science 27, 326(5957) 1263-1268, 2009.

Zinocker S, Wang M, Gaustad P, Kvalheim G, Rolstad B, Vaage JT. Mycoplasma Contamination Revisited: Mesenchymal Stromal Cells Harboring *Mycoplasma hyorhinis* Potently Inhibit Lymphocyte Proliferation *In Vitro*. PLoS ONE 6(1):e16005, 2011.

Zielinski GC, Young T, Ross RF, Rosenbusch RF. Adherence of *Mycoplasma hyopneumoniae* to cell monolayers. Am. J. Vet. Res. 51(3):339-343, 1990.

ANEXOS

Supplemental Table S1. General mapping of transcripts in *M. hyopneumoniae* genome.

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP18934	-	rnpB	RNase_P_RNA	39242
MP22782	MHP7448_0496	-	putative_p216_surfac	12836
MP04922	MHP7448_0392	-	S-adenosyl-methyltran	10121
MP04931	MHP7448_0391	-	cell_division_protein	9771
MP24504	MHP7448_0393	ftsZ	cell_division_prot	6351
MP24530	MHP7448_0136	-	hexosephosphate_tran	1909
MP07128	MHP7448_0067	dnaK	chaperone_protein	1132
MP07146	MHP7448_0137	ldh	L-lactate_dehydroge	881
MP24779	MHP7448_t29	tRNA	Ser	809
MP02446	MHP7448_0115	pdhA	pyruvate_dehydrog	675
MP01772	MHP7448_0225	-	methylmalonate-semia	617
MP01122	MHP7448_0377	-	hypothetical_protein	604
MP01104	MHP7448_0376	sgaT	PTS_system_ascorb	534
MP21331	MHP7448_0492	fruA	PTS_system_fructo	413
MP02439	MHP7448_0116	pdhB	pyruvate_dehydrog	389
MP03816	MHP7448_0656	-	prolipoprotein_p65	389
MP10664	MHP7448_0198	-	protein_P97_-_copy_1	292
MP12501	MHP7448_0497	-	p76_membrane_protein	288
MP24493	MHP7448_0114	apt	adenine_phosphoribo	242
MP01779	MHP7448_0226	iolC	myo-inositol_cata	236
MP01430	MHP7448_0489	-	hypothetical_protein	234
MP24708	MHP7448_0488	-	hypothetical_protein	218
MP04811	MHP7448_0401	asnS	asparaginyl-tRNA_	213
MP01802	MHP7448_0229	iolD	myo-inositol_cata	200
MP24359	MHP7448_0630	-	5'-nucleotidase_prec	193
MP05117	MHP7448_0513	-	46K_surface_antigen_	176
MP18911	MHP7448_0373	-	hypothetical_protein	175
MP00210	MHP7448_0427	efp	elongation_factor_E	151
MP01243	MHP7448_0108	-	protein_P97_-_copy_2	151
MP24510	MHP7448_0439	-	hypothetical_protein	150
MP10612	MHP7448_0132	rpmA	50S_ribosomal_prot	144
MP24549	MHP7448_0494	-	mannose-6-phosphate_i	137
MP12640	MHP7448_0195	rpsJ	30S_ribosomal_prot	136
MP18750	MHP7448_0250	eno	enolase	129
MP22798	MHP7448_0461	-	hypothetical_protein	129
MP07147	MHP7448_0276	truB	tRNA_pseudouridine	127
MP24574	MHP7448_0227	iolB	myo-inositol_catab	124
MP12704	MHP7448_0576	rpsD	30S_ribosomal_prot	122
MP04310	MHP7448_0194	rplC	50S_ribosomal_prot	117
MP18913	MHP7448_0460	rplK	50S_ribosomal_prot	117
MP24565	MHP7448_0487	mgtE	MG2+_transport_pr	111
MP12655	MHP7448_0619	rplJ	50S_ribosomal_prot	99
MP24473	MHP7448_0228	-	myo-inositol_2-dehyd	99
MP24398	MHP7448_0443	-	hypothetical_protein	97
MP24638	MHP7448_0666	-	hypothetical_protein	97
MP21517	MHP7448_0663	-	adhesin_like-protein	86
MP24772	MHP7448_t10	tRNA	Trp	86
MP03280	MHP7448_0617	rpoB	DNA-directed_RNA_	85
MP24514	MHP7448_0260	-	DNA_polymerase_III_delta_	85
MP12525	MHP7448_0428	tkt	transketolase	85
MP12516	MHP7448_0357	-	amino_acid_permease	84

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24411	MHP7448_0096	tpx	thiol_peroxidase	83
MP04451	MHP7448_0212	oppB-1	oligopeptide_AB	80
MP05984	MHP7448_0073	lepA	GTP-binding_prote	78
MP04373	MHP7448_0204	pth	peptidyl-tRNA_hydro	77
MP07152	MHP7448_0224	glyA	glycine_hydroxyme	75
MP24505	MHP7448_0206	ftsH	cell_division_pro	70
MP24777	MHP7448_t20	tRNA	Leu	69
MP00561	MHP7448_0263	cbiO	ABC_transporter_AT	66
MP24560	MHP7448_0149	tig	trigger_factor	66
MP24465	MHP7448_0055	rpsB	30S_ribosomal_prot	65
MP24711	MHP7448_0662	-	hypothetical_protein	65
MP24649	MHP7448_0231	mglA	ribose_ABC_transp	62
MP24386	MHP7448_0661	-	hypothetical_protein	61
MP12692	MHP7448_0333	-	hypothetical_protein	61
MP02359	MHP7448_0131	rplU	50S_ribosomal_prot	60
MP01569	MHP7448_0459	rplA	50S_ribosomal_prot	57
MP12654	MHP7448_0616	rpoC	DNA-directed_RNA_	57
MP01750	MHP7448_0344	-	hypothetical_protein	57
MP00106	MHP7448_0111	pfkA	6-phosphofructokin	56
MP10216	MHP7448_0507	pdhD	dihydrolipoamide_	55
MP10054	MHP7448_0411	-	hypothetical_protein	55
MP24480	MHP7448_0230	iolE	myo-inositol_catab	54
MP24500	MHP7448_0176	rplO	50S_ribosomal_prot	52
MP05305	MHP7448_0523	tufA	elongation_factor	52
MP12521	MHP7448_0041	obg	GTP-binding_protei	52
MP24636	MHP7448_0445	-	hypothetical_protein	52
MP24415	MHP7448_0524	lon	heat_shock_ATP-dep	50
MP24590	MHP7448_0602	-	hypothetical_protein	50
MP24374	MHP7448_0336	-	hypothetical_protein	50
MP24593	MHP7448_0138	-	hypothetical_protein	50
MP24433	MHP7448_0009	-	hypothetical_protein	50
MP24537	MHP7448_0464	pepA	leucyl_aminopepti	49
MP01127	MHP7448_0378	-	lipoprotein	49
MP24506	MHP7448_0207	lysS	lysyl-tRNA_synthe	49
MP04900	MHP7448_0394	-	nicotinate_phosphori	48
MP07137	MHP7448_0286	rpsP	30S_ribosomal_prot	46
MP24573	MHP7448_0652	ksgA	dimethyladenosine_	46
MP24634	MHP7448_0597	-	hypothetical_protein	46
MP12685	MHP7448_r1	rRNA5S	5S_ribosomal_RNA	45
MP24431	MHP7448_0075	fusA	elongation_factor	44
MP00746	MHP7448_0285	trmD	tRNA_(guanine-N1)-	44
MP03132	MHP7448_0335	-	hypothetical_protein	43
MP06004	MHP7448_0068	dnaJ	heat_shock_protei	42
MP10476	MHP7448_0668	greA	transcription_elon	42
MP24424	MHP7448_0528	gyrA	DNA_gyrase_subuni	38
MP24555	MHP7448_0449	engB	GTP-binding_protei	38
MP18594	MHP7448_0372	-	Lppt_protein	38
MP07178	MHP7448_0071	himA	bacterial_nucleoid	37
MP05045	MHP7448_0506	pdhC	dihydrolipoamide_a	37
MP24630	MHP7448_0484	-	hypothetical_protein	37
MP24589	MHP7448_0601	-	hypothetical_protein	37
MP24551	MHP7448_0437	sgaU	hexulose-6-phospha	36

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24718	-	-	hypothetical_protein	36
MP24422	MHP7448_0612	-	hypothetical_protein	36
MP24513	MHP7448_0669	-	hypothetical_protein	35
MP00607	MHP7448_0438	sgaH	3-hexulose-6-phosp	34
MP24662	MHP7448_0283	topA	DNA_topoisomerase	34
MP24686	MHP7448_0441	-	hypothetical_protein	34
MP24390	MHP7448_0586	nusA	N-utilization_sub	34
MP24776	MHP7448_t19	tRNA	Thr	34
MP24657	MHP7448_0125	infC	translation_initia	34
MP02368	MHP7448_0129	-	aminopeptidase	33
MP24371	MHP7448_0453	-	ABC_transporter_perm	32
MP07372	MHP7448_0010	hrcA	heat-inducible_tr	32
MP24738	-	-	hypothetical_protein	32
MP12612	MHP7448_0173	map	methionine_aminopep	32
MP24712	MHP7448_0620	-	hypothetical_protein	31
MP02519	MHP7448_0097	lgt	prolipoprotein_diac	31
MP07148	MHP7448_0618	rplL	50S_ribosomal_prot	30
MP02430	MHP7448_0117	-	hypothetical_protein	30
MP24604	MHP7448_0486	-	hypothetical_protein	30
MP24550	MHP7448_0092	-	hypothetical_protein	30
MP01233	MHP7448_0107	-	protein_P102_-_copy_	30
MP07150	MHP7448_0659	pepP	XAA-PRO_aminopept	30
MP24702	MHP7448_0028	gatA	glutamyl-tRNA_ami	29
MP18658	MHP7448_0505	-	lipoprotein	29
MP24797	MHP7448_t5	tRNA	Met	29
MP07171	MHP7448_0456	rpmE	50S_ribosomal_prot	28
MP24401	MHP7448_0025	-	hypothetical_protein	28
MP24492	MHP7448_0233	-	putative_ribose_ABC_	28
MP18627	MHP7448_0436	sgaE	sugar_isomerase_Sg	28
MP02573	MHP7448_0102	tpiA	triosephosphate_is	28
MP02461	MHP7448_0113	-	amino_acid_permeases	27
MP24430	MHP7448_0047	atpB	ATP_synthase_A_cha	27
MP24434	MHP7448_0280	pyrG	CTP_synthase	27
MP24733	-	-	hypothetical_protein	27
MP12573	MHP7448_0474	-	hypothetical_protein	27
MP24796	MHP7448_t6	tRNA	Met	27
MP24522	MHP7448_0213	oppC-1	oligopeptide_tra	27
MP07151	MHP7448_0672	valS	valyl-tRNA_synthe	27
MP24617	MHP7448_0289	-	hypothetical_protein	26
MP24363	MHP7448_0355	-	hypothetical_protein	26
MP00879	MHP7448_0473	nadE	NH(3)-dependent_NA	26
MP24362	MHP7448_0590	trpS	tryptophanyl-tRNA_	26
MP24787	MHP7448_t21	tRNA	Arg	25
MP06903	MHP7448_0008	ftsY	cell_division_protein_ftsY	25
MP24444	MHP7448_0045	dinP	DNA_polymerase_IV	25
MP24384	MHP7448_0352	-	hypothetical_protein	25
MP10651	MHP7448_0175	secY	preprotein_transl	25
MP10546	MHP7448_0086	secA	preprotein_transl	25
MP05913	MHP7448_0077	rpsL	30S_ribosomal_prot	24
MP18909	MHP7448_0385	proS	prolyl-tRNA_synth	24
MP24532	MHP7448_0403	pcrA	ATP-dependent_hel	23
MP12678	MHP7448_0035	gap	glyceraldehyde_3-p	23

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24464	MHP7448_0611	-	hypothetical_protein	23
MP18645	MHP7448_0470	-	hypothetical_protein	23
MP05851	MHP7448_0084	deoD	purine-nucleoside_	23
MP04295	MHP7448_0192	rplW	50S_ribosomal_prot	22
MP24552	MHP7448_0135	-	hypothetical_protein	22
MP24723	-	-	hypothetical_protein	22
MP24546	MHP7448_0221	nrdE	ribonucleoside-di	22
MP03749	MHP7448_0650	rpsI	30S_ribosomal_prot	21
MP24454	MHP7448_0604	xylF	ABC_transporter_x	21
MP24650	MHP7448_0027	gatC	Asp-tRNAAsn/Glu-tR	21
MP04131	MHP7448_0571	pdhD-1	dihydrolipoamid	21
MP03217	MHP7448_0608	tdk	thymidine_kinase	21
MP24784	MHP7448_t24	tRNA	Tyr	21
MP24509	MHP7448_0193	rplD	50S_ribosomal_prot	20
MP10220	MHP7448_0509	pta	phosphate_acetyltra	20
MP18894	MHP7448_0526	upp	uracil_phosphoribos	20
MP03756	MHP7448_0651	rplM	50S_ribosomal_prot	19
MP24375	MHP7448_0101	clpB	ATP-dependent_pro	19
MP01263	MHP7448_0109	gyrB	DNA_gyrase_subuni	19
MP06150	MHP7448_0056	tsf	elongation_factor_E	19
MP24436	MHP7448_0538	-	hypothetical_protein	19
MP24535	MHP7448_0450	-	hypothetical_protein	19
MP18884	MHP7448_0065	-	hypothetical_protein	19
MP01854	MHP7448_0235	-	myo-inositol_2-dehyd	19
MP24355	-	-	putative_type_III_restriction	19
MP24524	MHP7448_0654	prsA	ribose-phosphate_p	19
MP07141	MHP7448_0296	rpsF	30S_ribosomal_prot	18
MP04292	MHP7448_0191	rplB	50S_ribosomal_prot	18
MP24466	MHP7448_0452	-	ABC_transporter_ATP-b	18
MP24478	MHP7448_0091	uvrA	excinuclease_ABC_	18
MP24734	-	-	hypothetical_protein	18
MP01846	MHP7448_0234	rbsB	periplasmic_sugar	18
MP24463	MHP7448_0607	-	sugar_ABC_transporter	18
MP18891	MHP7448_0174	adk	adenylate_kinase	17
MP24458	MHP7448_0588	glpD	glycerol-3-phosph	17
MP12602	MHP7448_0525	-	hypothetical_protein	17
MP24597	MHP7448_0511	-	hypothetical_protein	17
MP02328	MHP7448_0360	-	P37-like_ABC_transpo	17
MP01679	MHP7448_0353	-	P60-like_lipoprotein	17
MP24394	MHP7448_0251	serS	seryl-tRNA_synthe	17
MP24785	MHP7448_t23	tRNA	Gln	16
MP24521	MHP7448_0003	gidA	glucose_inhibited	16
MP06274	MHP7448_0043	-	hypothetical_protein	16
MP24486	MHP7448_0308	-	hypothetical_protein	16
MP24575	MHP7448_0214	oppD	oligopeptide_ABC_	16
MP24475	MHP7448_0458	rluD	ribosomal_large_su	16
MP05869	MHP7448_0083	deoA	thymidine_phospho	16
MP05890	MHP7448_0081	-	amino_acid_permease	15
MP02958	MHP7448_0241	secD	bifunctional_prep	15
MP01391	MHP7448_0485	-	hypothetical_protein	15
MP24395	MHP7448_0530	-	hypothetical_protein	15
MP03066	MHP7448_0328	-	hypothetical_protein	15

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24701	MHP7448_0466	-	hypothetical_protein	15
MP24490	MHP7448_0366	-	lipoprotein	15
MP24554	MHP7448_0040	licA	PTS_system,_lichen	15
MP02526	MHP7448_0098	trxB	thioredoxin_reduct	15
MP06611	MHP7448_0024	-	ABC_transporter_ATP-	14
MP24460	MHP7448_0256	dnaX	DNA_polymerase_II	14
MP24519	MHP7448_0648	uvrB	excinuclease_ABC_	14
MP24437	MHP7448_0011	grpE	heat_shock_protein	14
MP24643	MHP7448_0444	-	hypothetical_protein	14
MP04937	MHP7448_0390	-	hypothetical_protein	14
MP01935	MHP7448_0158	-	hypothetical_protein	14
MP12543	MHP7448_0493	-	hypothetical_protein	14
MP24452	MHP7448_0457	napA	neutrophil_activat	14
MP12588	MHP7448_0635	gcp	O-sialoglycoprotein	14
MP18554	MHP7448_0274	pheT	phenylalanyl-tRNA	14
MP07149	MHP7448_0637	nusG	transcription_anti	14
MP24399	MHP7448_0279	-	transcriptional_regul	14
MP24469	MHP7448_0188	rpsC	30S_ribosomal_prot	13
MP05064	MHP7448_0508	ackA	acetate_kinase	13
MP24663	MHP7448_0592	-	ATP_binding_protein	13
MP24487	MHP7448_0022	-	hypothetical_protein	13
MP24453	MHP7448_0217	-	lipoprotein	13
MP24485	MHP7448_0554	-	PTS_system_enzyme_IIA	13
MP24405	MHP7448_0587	pulA	pullulanase	13
MP18895	-	-	putative_protein_P102_-_fr	13
MP12639	MHP7448_0126	pyk	pyruvate_kinase	13
MP10226	MHP7448_0515	xylH	xylose_ABC_transp	13
MP18949	MHP7448_0557	-	d-ribulose-5-phosphate_3_e	13
MP12889	MHP7448_0130	ung	uracil-dna_glycosylase	12
MP24570	MHP7448_0189	rplV	50S_ribosomal_prot	12
MP24556	MHP7448_0547	-	hypothetical_protein	12
MP02428	MHP7448_0118	-	hypothetical_protein	12
MP06063	MHP7448_0061	-	hypothetical_protein	12
MP12709	MHP7448_0647	leuS	leucyl-tRNA_synth	12
MP05668	MHP7448_0546	ktrB	potassium_uptake_	12
MP24653	MHP7448_0223	nrdF	ribonucleoside-dip	12
MP12728	MHP7448_0584	infB	translation_initi	12
MP24652	MHP7448_0124	rpmI	50s_ribosomal_prot	11
MP07168	MHP7448_0123	rplT	50S_ribosomal_prot	11
MP00545	MHP7448_0265	-	ABC_transporter_perme	11
MP24404	MHP7448_0348	-	hypothetical_protein	11
MP24583	MHP7448_0298	-	hypothetical_protein	11
MP24619	MHP7448_0463	-	hypothetical_protein	11
MP24596	MHP7448_0500	-	hypothetical_protein	11
MP24418	MHP7448_0432	-	hypothetical_protein	11
MP10150	MHP7448_0472	ptsI	phosphoenolpyruva	11
MP04342	MHP7448_0199	-	protein_P102_-_copy_	11
MP12679	MHP7448_0034	parE	topoisomerase_IV_	11
MP04423	MHP7448_0209	trmE	tRNA_modification	11
MP24545	MHP7448_0395	-	DNA_methylase	10
MP02412	MHP7448_0120	-	hypothetical_protein	10
MP24533	MHP7448_0216	-	hypothetical_protein	10

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP12630	MHP7448_0347	-	hypothetical_protein	10
MP00795	MHP7448_0433	-	hypothetical_protein	10
MP04562	MHP7448_0419	-	hypothetical_protein	10
MP24584	MHP7448_0429	-	hypothetical_protein	10
MP24783	MHP7448_t25	tRNA	Thr	10
MP04438	MHP7448_0210	-	ABC_transporter_ATP-	9
MP00554	MHP7448_0264	-	ABC_transporter_ATP-b	9
MP24501	MHP7448_0371	-	ABC_transporter_perme	9
MP24578	MHP7448_0201	alaS	alanyl-tRNA_synth	9
MP24372	MHP7448_0070	engA	GTP-binding_prote	9
MP24625	MHP7448_0440	-	hypothetical_protein	9
MP24624	MHP7448_0402	-	hypothetical_protein	9
MP21327	MHP7448_0477	-	hypothetical_protein	9
MP24366	MHP7448_0399	-	hypothetical_protein	9
MP10050	MHP7448_0408	-	hypothetical_protein	9
MP24389	MHP7448_0277	-	hypothetical_protein	9
MP01271	MHP7448_0667	-	hypothetical_protein	9
MP04879	MHP7448_0396	metS	methionyl-tRNA_sy	9
MP24408	MHP7448_0521	pepF	oligoendopeptidas	9
MP05666	MHP7448_0545	ktrA	potassium_uptake_p	9
MP24536	MHP7448_0574	-	PTS_system_N-acetylg	9
MP04552	MHP7448_0420	-	single-strand_binding	9
MP24491	MHP7448_0514	xylG	xylose_ABC_transp	9
MP07163	MHP7448_0245	rpsT	30S_ribosomal_prot	8
MP12594	MHP7448_0638	rpmG-1	50S_ribosomal_pr	8
MP12668	MHP7448_0658	rpmG	50S_ribosomal_prot	8
MP08296	MHP7448_0284	rplS	50S_ribosomal_prot	8
MP24388	MHP7448_0051	atpA-1	ATP_synthase_al	8
MP06181	MHP7448_0053	atpD	ATP_synthase_beta	8
MP24352	-	atpD	ATP_synthase_beta_chain	8
MP12737	MHP7448_0087	rbgA	GTP-binding_protei	8
MP24387	MHP7448_0599	-	hypothetical_protein	8
MP24528	MHP7448_0539	-	hypothetical_protein	8
MP03145	MHP7448_0337	-	hypothetical_protein	8
MP24423	MHP7448_0318	-	hypothetical_protein	8
MP24640	MHP7448_0447	-	hypothetical_protein	8
MP12690	MHP7448_0299	-	hypothetical_protein	8
MP24557	MHP7448_0480	-	hypothetical_protein	8
MP12713	MHP7448_0670	-	hypothetical_protein	8
MP24623	MHP7448_0313	-	hypothetical_protein	8
MP00125	MHP7448_0603	ppa	inorganic_pyrophosp	8
MP24511	MHP7448_0215	oppF-1	oligopeptide_AB	8
MP24426	MHP7448_0501	oppF	oligopeptide_ABC_	8
MP07136	MHP7448_0099	-	outer_membrane_prote	8
MP24794	MHP7448_t8	tRNA	Pro	8
MP24356	-	-	putative_type_III_restriction	8
MP18746	MHP7448_0222	nrdI	ribonucleotide_red	8
MP06142	MHP7448_0057	ffh	signal_recognition	8
MP03945	MHP7448_0589	thrS	threonyl-tRNA_syn	8
MP06487	MHP7448_0033	parC	topoisomerase_IV_	8
MP24547	MHP7448_0143	-	tRNA/rRNA_methyltrans	8
MP24502	MHP7448_0037	vacB	VACB-like_ribonuc	8

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP05936	MHP7448_0076	rpsG	30S_ribosomal_prot	7
MP24793	MHP7448_t9	tRNA	Arg	7
MP05997	MHP7448_0069	cmk	cytidylate_kinase	7
MP14673	MHP7448_0531	pgiB	glucose-6-phospha	7
MP02335	MHP7448_0359	glpK	glycerol_kinase	7
MP02505	MHP7448_0095	-	hypothetical_protein	7
MP24406	MHP7448_0522	-	hypothetical_protein	7
MP24587	MHP7448_0642	-	hypothetical_protein	7
MP24419	MHP7448_0468	-	hypothetical_protein	7
MP24455	MHP7448_0042	-	hypothetical_protein	7
MP12546	MHP7448_0499	-	hypothetical_protein	7
MP02375	MHP7448_0128	-	hypothetical_protein	7
MP21384	MHP7448_0093	-	hypothetical_protein	7
MP24696	MHP7448_0418	-	hypothetical_protein	7
MP24658	MHP7448_0254	lplA	lipoate-protein_li	7
MP24353	-	-	putative_type_II_restriction_	7
MP24581	MHP7448_0039	recA	recombination_pro	7
MP01185	MHP7448_0384	-	thioredoxin	7
	MHP7448_0388		type III restriction-modifica	7
MP24665	MHP7448_0172	infA	translation_initia	7
MP04182	MHP7448_0177	rpsE	30S_ribosomal_prot	6
MP07132	MHP7448_0190	rpsS	30S_ribosomal_prot	6
MP24462	MHP7448_0182	rplE	50S_ribosomal_prot	6
MP06713	MHP7448_0012	argS	arginyl-tRNA_synt	6
MP18648	MHP7448_0479	atpA	ATP_synthase_alph	6
MP07176	MHP7448_0157	cdd	cytidine_deaminase	6
MP07145	MHP7448_0527	deoC	deoxyribose-phosph	6
MP24432	MHP7448_0058	glyS	glycyl-tRNA_synth	6
MP01943	MHP7448_0156	era	GTP-binding_protein	6
MP24489	MHP7448_0044	-	hypothetical_protein	6
MP07173	MHP7448_0072	-	hypothetical_protein	6
MP04331	MHP7448_0197	-	hypothetical_protein	6
MP24616	MHP7448_0085	-	hypothetical_protein	6
MP03708	MHP7448_0646	-	hypothetical_protein	6
MP24338	MHP7448_0561	-	hypothetical_protein	6
MP04668	MHP7448_0409	-	hypothetical_protein	6
MP24782	MHP7448_t26	tRNA	Leu	6
MP24566	MHP7448_0133	lip2	lipase-esterase	6
MP05884	MHP7448_0082	nox	NADH_oxidase	6
MP24382	MHP7448_0270	-	nuclease_lipoprotein	6
MP02643	MHP7448_0139	prfA	peptide_chain_rel	6
MP24417	MHP7448_0273	pheS	phenylalanyl-tRNA_	6
MP05736	MHP7448_0550	mtlA	PTS_system,_manni	6
MP18907	MHP7448_0552	ulaA	PTS_system_enzyme	6
MP03381	MHP7448_0291	-	type_II_DNA_modifica	6
MP00487	MHP7448_0169	rpsK	30S_ribosomal_prot	5
MP07143	MHP7448_0122	rpmB	50S_ribosomal_prot	5
MP01177	MHP7448_0383	-	ABC_transporter_ATP-	5
MP03337	MHP7448_0622	dam	DNA_adenine_methyl	5
MP07081	MHP7448_0002	dnaN	DNA_polymerase_II	5
MP12574	MHP7448_0029	gatB	glutamyl-tRNA_ami	5
MP24445	MHP7448_0211	-	hypothetical_protein	5

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP12558	MHP7448_0544	-	hypothetical_protein	5
MP24602	MHP7448_0321	-	hypothetical_protein	5
MP24435	MHP7448_0610	-	hypothetical_protein	5
MP24383	MHP7448_0483	-	hypothetical_protein	5
MP24397	MHP7448_0074	-	hypothetical_protein	5
MP24558	MHP7448_0205	-	hypothetical_protein	5
MP24365	MHP7448_0660	-	hypothetical_protein	5
MP10027	MHP7448_0397	-	hypothetical_protein	5
MP12651	MHP7448_0244	-	hypothetical_protein	5
MP12629	MHP7448_0351	-	hypothetical_protein	5
MP24429	MHP7448_0032	ileS	isoleucyl-tRNA_sy	5
MP24792	MHP7448_t11	tRNA	Lys	5
MP03808	MHP7448_0655	gidB	methyltransferase_	5
MP01981	MHP7448_0152	-	GTPase_engC	5
MP00567	MHP7448_0262	hpt	hypoxanthine_phosphoribos	5
MP24579	MHP7448_0243	aspS	aspartyl-tRNA_synthetase	5
MP03225	MHP7448_0609	ptsH	phosphocarrier_pro	5
MP24697	MHP7448_0469	-	putative_ABC_transpor	5
MP24646	MHP7448_0414	-	putative_ICEF-II	5
MP24569	MHP7448_0451	metK	S-adenosylmethion	5
MP24788	MHP7448_t15	tRNA	SeC	5
MP24457	MHP7448_0605	-	sugar_ABC_transporte	5
MP03661	MHP7448_0640	-	tRNA/rRNA_methyltrans	5
MP12667	MHP7448_0187	rplP	50S_ribosomal_prot	4
MP03703	MHP7448_0645	rplI	50S_ribosomal_prot	4
MP06623	MHP7448_0023	-	ABC_transporter_ATP-	4
MP24484	MHP7448_0380	-	ABC_transporter_perme	4
MP22812	MHP7448_0606	-	ABC_transporter_prot	4
MP12580	MHP7448_0059	dnaG	DNA_primase	4
MP06704	MHP7448_0014	fba	fructose-bisphospha	4
MP24476	MHP7448_0517	glck	glucokinase	4
MP24790	MHP7448_t13	tRNA	His	4
MP24684	-	-	hypothetical_protein	4
MP24520	MHP7448_0080	-	hypothetical_protein	4
MP24594	MHP7448_0634	-	hypothetical_protein	4
MP24691	MHP7448_0287	-	hypothetical_protein	4
MP24632	MHP7448_0300	-	hypothetical_protein	4
MP24562	MHP7448_0064	-	hypothetical_protein	4
MP24635	MHP7448_0462	-	hypothetical_protein	4
MP24739	-	-	hypothetical_protein	4
MP24606	-	-	hypothetical_protein	4
MP18561	MHP7448_0290	-	hypothetical_protein	4
MP18915	MHP7448_0278	-	hypothetical_protein	4
MP07099	MHP7448_0676	-	hypothetical_protein	4
MP24438	MHP7448_0030	-	hypothetical_protein	4
MP24439	MHP7448_0653	-	hypothetical_protein	4
MP24446	MHP7448_0162	-	hypothetical_protein	4
MP24427	MHP7448_0400	-	hypothetical_protein	4
MP18918	MHP7448_0529	pmsR	methionine_sulfoxi	4
MP24526	MHP7448_0271	-	P102-like_protein	4
MP24801	MHP7448_t1	tRNA	Phe	4
MP01435	MHP7448_0490	pgk	phosphoglycerate_k	4

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP12662	MHP7448_0657	-	putative_ABC_transpo	4
MP06982	MHP7448_0005	-	putative_MgpA-like_pr	4
MP24488	MHP7448_0144	-	putative_tRNA/rRNA_me	4
MP03698	MHP7448_0644	dnaB	replicative_DNA_h	4
MP24481	MHP7448_0534	frr	ribosome_recycling_	4
MP07140	MHP7448_0310	lplA-1	lipoate-protein_ligase_A	4
MP07154	MHP7448_0015	rluC	ribosomal_large_subunit_ps	4
MP07157	MHP7448_0249	smf	DNA_processing_protein_S	4
MP02181	MHP7448_0269	smpB	SsrA-binding_prote	4
MP12624	MHP7448_0465	trmB	tRNA_(guanine-N(7)	4
MP02791	MHP7448_0631	plsC	1-acyl-sn-glycerol	3
MP07170	MHP7448_0275	rpmF	50S_ribosomal_prot	3
MP12634	MHP7448_0362	-	ABC_transport_system	3
MP18919	MHP7448_0314	-	ABC_transporter_ATP-	3
MP03537	MHP7448_0306	-	ABC_transporter_ATP-	3
MP24381	MHP7448_0623	-	ABC_transporter_ATP-	3
MP01604	MHP7448_0455	acpD-1	acyl_carrier_pro	3
MP01614	MHP7448_0454	acpD	acyl_carrier_prote	3
MP24409	MHP7448_0582	dnaE	DNA_polymerase_II	3
MP24698	MHP7448_0532	polC	DNA_polymerase_II	3
MP24704	MHP7448_0060	rpoD	DNA-directed_RNA_	3
MP07156	MHP7448_0578	fpg	foramidopyrimidine_	3
MP04093	MHP7448_0575	nagB	glucosamine-6-phos	3
MP03687	MHP7448_0643	-	hemolysin_C	3
MP06171	MHP7448_0054	recU	Holliday_junction-	3
MP24680	-	-	hypothetical_protein	3
MP24679	-	-	hypothetical_protein	3
MP24474	MHP7448_0121	-	hypothetical_protein	3
MP22795	MHP7448_0448	-	hypothetical_protein	3
MP03013	MHP7448_0248	-	hypothetical_protein	3
MP24553	MHP7448_0334	-	hypothetical_protein	3
MP24585	MHP7448_0246	-	hypothetical_protein	3
MP24448	MHP7448_0346	-	hypothetical_protein	3
MP12747	MHP7448_0148	-	hypothetical_protein	3
MP10378	MHP7448_0615	-	hypothetical_protein	3
MP03470	MHP7448_0297	-	hypothetical_protein	3
MP24615	MHP7448_0255	-	hypothetical_protein	3
MP24503	MHP7448_0537	-	hypothetical_protein	3
MP24715	-	-	hypothetical_protein	3
MP24523	MHP7448_0435	-	hypothetical_protein	3
MP24369	MHP7448_0675	-	hypothetical_protein	3
MP24482	MHP7448_0593	rnhB	ribonuclease_HII	3
MP24499	MHP7448_0406	ruvB	holliday_junction_DNA_heli	3
MP24544	MHP7448_0504	oppB	oligopeptide_ABC_	3
	MHP7448_0294	rpsR	30S ribosomal protein S18	3
	MHP7448_0562	-	hypothetical protein	3
	MHP7448_0613	-	hypothetical protein	3
MP07885	MHP7448_0161	deoB	phosphopentomutas	3
MP24380	MHP7448_0591	-	PTS_system_glucose-sp	3
MP04523	MHP7448_0424	-	putative_ICEF_Integr	3
MP24648	MHP7448_0423	-	putative_ICEF-IIA	3
MP00686	MHP7448_0258	recR	recombination_prot	3

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP04070	MHP7448_0577	rpiB	ribose_5-phosphate	3
MP24494	MHP7448_0026	sipS	signal_peptidase_I	3
MP24666	MHP7448_0416	trsE	TRSE-like_protein	3
MP12616	MHP7448_0110	-	ISMHp1_transposase	3
MP02748	MHP7448_0626	-	type_II_DNA_modifica	3
MP07133	MHP7448_0180	rpsH	30S_ribosomal_prot	2
MP04205	MHP7448_0179	rplF	50S_ribosomal_prot	2
MP00474	MHP7448_0167	rplQ	50S_ribosomal_prot	2
MP24541	MHP7448_0475	pfs	5'-methylthioadenos	2
MP24410	MHP7448_0379	-	ABC_transport_ATP-bin	2
MP24471	MHP7448_0370	-	ABC_transport_permea	2
MP07161	MHP7448_0020	-	ABC_transporter_ATP-	2
MP01913	MHP7448_0160	-	ABC_transporter_ATP-	2
MP24795	MHP7448_t7	tRNA	Ala	2
MP24577	MHP7448_0050	atpH	ATP_synthase_delta	2
MP06187	MHP7448_0052	atpG	ATP_synthase_gamma	2
MP24713	MHP7448_0049	atpF	ATP_synthase_subun	2
MP24412	MHP7448_0038	cls	cardiolipin_synthe	2
MP24568	MHP7448_0268	mgtA	cation-transporti	2
MP24407	MHP7448_0001	dnaA	chromosomal_repli	2
MP24361	MHP7448_0266	lig	DNA_ligase	2
MP00480	MHP7448_0168	rpoA	DNA-directed_RNA_p	2
MP07153	MHP7448_0066	uvrC	excinuclease_ABC_	2
MP24477	MHP7448_0573	fba-1	fructose-bisphosp	2
MP24373	MHP7448_0141	gltX	glutamyl-tRNA_syn	2
MP24778	MHP7448_t28	tRNA	Gly	2
MP03523	MHP7448_0303	ugpQ	glycerophosphoryl_	2
MP12635	MHP7448_0163	tlyA	haemolysin	2
MP04689	MHP7448_0405	ruvA	holliday_junction_	2
MP03742	MHP7448_0649	-	hypothetical_protein	2
MP24600	MHP7448_0202	-	hypothetical_protein	2
MP24601	MHP7448_0317	-	hypothetical_protein	2
MP09780	MHP7448_0267	-	hypothetical_protein	2
MP24603	MHP7448_0426	-	hypothetical_protein	2
MP12570	MHP7448_0467	-	hypothetical_protein	2
MP24611	MHP7448_0239	-	hypothetical_protein	2
MP24695	-	-	hypothetical_protein	2
MP24689	MHP7448_0446	-	hypothetical_protein	2
MP24688	-	-	hypothetical_protein	2
MP24682	-	-	hypothetical_protein	2
MP24677	-	-	hypothetical_protein	2
MP24716	-	-	hypothetical_protein	2
MP24722	-	-	hypothetical_protein	2
MP24725	-	-	hypothetical_protein	2
MP24727	-	-	hypothetical_protein	2
MP24731	-	-	hypothetical_protein	2
MP24732	-	-	hypothetical_protein	2
MP12768	MHP7448_0629	-	hypothetical_protein	2
MP24567	MHP7448_0407	-	hypothetical_protein	2
MP24402	MHP7448_0431	-	hypothetical_protein	2
MP24392	MHP7448_0319	-	hypothetical_protein	2
MP01324	MHP7448_0671	-	hypothetical_protein	2

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24515	MHP7448_0560	-	hypothetical_protein	2
MP24420	MHP7448_0240	-	hypothetical_protein	2
MP02248	MHP7448_0368	-	lipoprotein	2
MP24531	MHP7448_0031	lsp	lipoprotein_signal_	2
	MHP7448_0410	-	hypothetical protein	1
	MHP7448_0568	-	hypothetical protein	1
	MHP7448_0387	-	hypothetical protein	1
MP24403	MHP7448_0572	nagA	N-acetylglucosami	2
MP24467	MHP7448_0272	-	P97-like_protein	2
MP05556	MHP7448_0533	-	phosphatidate_cytidyl	2
MP12583	MHP7448_0595	pgm	phosphoglycerate_m	2
MP12535	MHP7448_0665	-	putative_ABC_transpo	2
MP24354	-	-	putative_type_III_restriction	2
MP04837	MHP7448_0398	rnc	ribonuclease_III	2
MP24470	MHP7448_0166	rluB	ribosomal_large_su	2
MP24781	MHP7448_t27	tRNA	Ser	2
MP03447	MHP7448_0295	ssb	single-strand_bindi	2
MP05644	MHP7448_0542	potA	spermidine/putres	2
MP24449	MHP7448_0259	tmk	thymidylate_kinase	2
MP12605	-	-	transposase,_degenerated	2
MP12724	MHP7448_0430	trmU	tRNA-(5-methylami	2
MP04354	MHP7448_0200	-	26.3_kDa_protein_in_c	1
MP07130	MHP7448_0181	rpsN	30S_ribosomal_prot	1
MP00493	MHP7448_0170	rpsM	30S_ribosomal_prot	1
MP18933	-	scr	4.5S_signal_recognition_pa	1
MP24428	MHP7448_0186	rpmC	50S_ribosomal_prot	1
MP04222	MHP7448_0183	rplX	50S_ribosomal_prot	1
MP24651	MHP7448_0178	rplR	50S_ribosomal_prot	1
MP06644	MHP7448_0021	-	ABC_transporter_ATP-	1
MP06674	MHP7448_0019	-	ABC_transporter_ATP-	1
MP21278	MHP7448_0305	-	ABC_transporter_ATP-	1
MP21516	MHP7448_0361	-	ABC_transporter_ATP-b	1
MP24512	MHP7448_0369	-	ABC_transporter_perm	1
MP24800	MHP7448_t2	tRNA	Asp	1
MP24416	MHP7448_0145	-	chromate_transport_pr	1
MP02709	MHP7448_0146	-	chromate_transport_pr	1
MP24496	MHP7448_0641	cysS	cysteinyl-tRNA_sy	1
MP24385	MHP7448_0614	ddem	cytosine_specific	1
MP24495	MHP7448_0006	-	DHH_family_phosphoest	1
MP01961	MHP7448_0154	-	DNA_methylase	1
MP12541	MHP7448_0581	-	DNA_polymerase_I	1
MP06236	MHP7448_0048	atpE	F0F1_ATP_synthase_	1
MP02344	MHP7448_0358	glpF	glycerol_uptake_fa	1
MP03433	MHP7448_0293	ychF	GTP-binding_prote	1
MP02963	MHP7448_0242	hisS	histidyl-tRNA_syn	1
MP13075	MHP7448_0354	hit	HIT-like_protein	1
MP12720	MHP7448_0208	-	hydrolase_of_the_HAD_	1
MP24525	MHP7448_0476	nadD	hypothetical_prot	1
MP24724	-	-	hypothetical_protein	1
MP24396	MHP7448_0349	-	hypothetical_protein	1
MP02719	MHP7448_0147	-	hypothetical_protein	1
MP24516	MHP7448_0563	-	hypothetical_protein	1

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP03198	MHP7448_0342	-	hypothetical_protein	1
MP24710	MHP7448_0558	-	hypothetical_protein	1
MP24707	MHP7448_0338	-	hypothetical_protein	1
MP24497	MHP7448_0142	-	hypothetical_protein	1
MP24414	MHP7448_0327	-	hypothetical_protein	1
MP24736	-	-	hypothetical_protein	1
MP24459	MHP7448_0471	-	hypothetical_protein	1
MP24447	MHP7448_0326	-	hypothetical_protein	1
MP24729	-	-	hypothetical_protein	1
MP24726	-	-	hypothetical_protein	1
MP24642	MHP7448_0598	-	hypothetical_protein	1
MP24639	MHP7448_0320	-	hypothetical_protein	1
MP24637	MHP7448_0569	-	hypothetical_protein	1
MP10163	MHP7448_0481	-	hypothetical_protein	1
MP24626	MHP7448_0512	-	hypothetical_protein	1
MP24576	MHP7448_0674	-	hypothetical_protein	1
MP24614	MHP7448_0520	-	hypothetical_protein	1
MP24592	MHP7448_0127	-	hypothetical_protein	1
MP12611	MHP7448_0559	-	hypothetical_protein	1
MP12652	MHP7448_0339	-	hypothetical_protein	1
MP24703	MHP7448_0094	-	hypothetical_protein	1
MP24700	-	-	hypothetical_protein	1
MP24364	MHP7448_0536	-	hypothetical_protein	1
MP00727	MHP7448_0282	-	hypothetical_protein	1
MP24692	-	-	hypothetical_protein	1
MP24690	-	-	hypothetical_protein	1
MP24348	MHP7448_0324	-	hypothetical_protein	1
MP24343	MHP7448_0567	-	hypothetical_protein	1
MP24683	-	-	hypothetical_protein	1
MP24539	MHP7448_0100	-	hypothetical_protein	1
MP24540	MHP7448_0090	-	hypothetical_protein	1
MP24542	MHP7448_0600	-	hypothetical_protein	1
MP24341	MHP7448_0564	-	hypothetical_protein	1
MP24548	MHP7448_0079	-	hypothetical_protein	1
MP24780	MHP7448_t30	tRNA	Ile	1
MP24791	MHP7448_t12	tRNA	Leu	1
MP03567	MHP7448_0309	baiH	NADH-dependent_fl	1
MP12723	MHP7448_0502	oppD-1	oligopeptide_AB	1
MP12657	MHP7448_0302	-	permease	1
MP24479	MHP7448_0382	-	proline_dipeptidase	1
MP18899	MHP7448_0375	-	PTS_system_enzyme_IIB	1
MP24659	MHP7448_0621	-	putative_lipoprotein	1
MP12527	MHP7448_0677	rnpA	ribonuclease_P_pro	1
MP24507	MHP7448_0151	rpe	ribulose-phosphate_	1
MP24508	MHP7448_0633	scpB	segregation_and_co	1
MP24572	MHP7448_0632	scpA	segregation_and_co	1
MP05619	MHP7448_0540	potC	spermidine/putresc	1
MP24379	MHP7448_0332	-	subtilisin-like_seri	1
MP00185	MHP7448_0253	lip	triacylglycerol_lip	1
MP12673	MHP7448_0535	pyrH	uridylate_kinase	1
MP24706	MHP7448_0281	pgsA	CDP-diacylglycerol--glycerc	1
MP24655	MHP7448_0541	potB	spermidine/putrescine_ABC	1

Continuação *Supplemental Table S1*

Id_ORF	NCBI	Gene name	Product	Number of Reads
MP24456	MHP7448_0140	hemK	protoporphirogen_oxidase	1
MP12633	MHP7448_0583	rbfA	ribosome-binding_factor_A	1
MP12536	MHP7448_0164	nifS	nitrogen_fixation_protein	1
MP12557	MHP7448_0153	gmk	guanylate_kinase	1
MP06059	MHP7448_0062	nfo	endonuclease_IV	1
MP03596	MHP7448_0312	-	glycine_cleavage_system_H	1
MP24737	-	-	hypothetical_protein	1
MP24773	MHP7448_t16	tRNA	Asn	1
MP24774	MHP7448_t17	tRNA	Glu	1
MP24775	MHP7448_t18	tRNA	Val	1
MP24786	MHP7448_t22	tRNA	Cys	1
MP24789	MHP7448_t14	tRNA	Leu	1
MP24798	MHP7448_t4	tRNA	Ser	1
MP24799	MHP7448_t3	tRNA	Met	1
MP24563	MHP7448_0185	rpsQ	30S_ribosomal_protein_S17	1
MP24559	MHP7448_0247	-	TRSE-like_protein	1
	MHP7448_0103	rpsO	30S ribosomal protein S15	1
	MHP7448_0150	-	hypothetical protein	1
	MHP7448_0105	-	hypothetical protein	1
	MHP7448_0171	rpmJ	50S ribosomal protein L36	1
	MHP7448_0184	rplN	50S ribosomal protein L14	1
	MHP7448_0292	-	hypothetical protein	1
	MHP7448_0316	-	hypothetical protein	1
	MHP7448_0343	-	hypothetical protein	1
	MHP7448_0386	-	hypothetical protein	1

Supplemental Table S2. General mapping of transcripts in *M. flocculare* genome.

Id_ORF	Gene name	Product	Number of Reads
MF01423	mraZ	cell_division_protein_MraZ	36440
MF01101	-	S-adenosyl-methyltransferase	23325
MF01095	ftsZ	cell_division_protein_ftsZ	8763
MF01093	-	hypothetical_protein	6513
MF01463	-	hypothetical_protein	2620
MF01536	tRNA	Ser	1153
MF00857	-	hypothetical_protein	1053
MF00837	rplK	50S_ribosomal_protein_L11	1020
MF00835	rplA	50S_ribosomal_protein_L1	891
MF01167	dnaK	chaperone_protein_dnaK_-_heat_	732
MF00750	-	hypothetical_protein	562
MF01519	tRNA	Leu	538
MF00472	-	protein_P97_-_copy_2	510
MF00861	mgtE	MG2+_transport_protein	463
MF00591	apt	adenine_phosphoribosyltransferas	397
MF01534	tRNA	Leu	390
MF00747	sgaT	PTS_system_ascorbate-specific_	372
MF01400	rpsJ	30S_ribosomal_protein_S10	337
MF01476	-	5S_ribosomal_RNA,_fragment	328
MF00856	-	hypothetical_protein	214
MF00673	eno	enolase	190
MF01362	tpx	thiol_peroxidase	186
MF00378	ptsH	phosphocarrier_protein_HPr	186
MF00172	tig	trigger_factor	181
MF00215	secY	preprotein_translocase_SecY_su	171
MF01198	deoD	purine-nucleoside_phosphorylase	170
MF00687	secD	bifunctional_preprotein_transl	166
MF00043	pdhC	dihydrolipoamide_acetyltransfer	151
MF01050	-	adhesin_like-protein_P146	145
MF00585	pdhA	pyruvate_dehydrogenase_E1-alph	140
MF01294	pepA	leucyl_aminopeptidase	134
MF01107	-	hypothetical_protein	134
MF00129	rplU	50S_ribosomal_protein_L21	129
MF00886	ksgA	dimethyladenosine_transferase	127
MF01178	fusA	elongation_factor_EF-G	122
MF00691	efp	elongation_factor_EF-P	121
MF00844	-	p76_membrane_protein_precursor	121
MF01513	tRNA	Leu	120
MF00848	-	Putative_P216_surface_protein	118
MF00947	-	5'-nucleotidase_precursor	113
MF01512	tRNA	Thr	112
MF01194	nox	NADH_oxidase	109
MF00980	gidA	glucose_inhibited_division_pro	108
MF00426	rpsD	30S_ribosomal_protein_S4	105
MF00692	tkt	transketolase	104
MF00686	hisS	histidyl-tRNA_synthetase	104
MF00475	-	protein_P102	102
MF00582	pdhB	pyruvate_dehydrogenase	102
MF00042	pdhD	dihydrolipoamide_dehydrogenase	102
MF00132	-	hypothetical_protein	101
MF00072	pyrG	CTP_synthase	96
MF01328	rpsI	30S_ribosomal_protein_S9	96

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00741	-	hypothetical_protein	94
MF00888	rplM	50S_ribosomal_protein_L13	93
MF00133	-	hexosephosphate_transport_protein	90
MF00929	nusG	transcription_antitermination_p	90
MF01461	-	hypothetical_protein	89
MF00351	rpoC	DNA-directed_RNA_polymerase_su	84
MF00561	-	hypothetical_protein	83
MF00610	truB	tRNA_pseudouridine_synthase_B	81
MF01290	-	hypothetical_protein	80
MF00592	-	amino_acid_permeases	80
MF01288	-	hypothetical_protein	76
MF00353	rpoB	DNA-directed_RNA_polymerase_be	76
MF00037	-	46K_surface_antigen_precursor	75
MF01176	lepA	GTP-binding_protein_LepA	74
MF01514	tRNA	Trp	74
MF00374	ptsG	PTS_system,_glucose-specific_I	74
MF00245	rplC	50S_ribosomal_protein_L3	73
MF01277	-	hypothetical_protein	71
MF01532	tRNA	Gly	71
MF00272	-	hypothetical_protein	68
MF00130	rpmA	50S_ribosomal_protein_L27	66
MF00882	prsA	ribose-phosphate_pyrophosphoki	66
MF00251	alaS	alanyl-tRNA_synthetase	65
MF00903	leuS	leucyl-tRNA_synthetase	64
MF00046	-	lipoprotein	63
MF00616	pheT	phenylalanyl-tRNA_synthetase_b	62
MF01351	RpsF	30S_ribosomal_protein_S6	62
MF01339	rplJ	50S_ribosomal_protein_L10	62
MF00434	nadE	NH(3)-dependent_NAD+_synthetase	61
MF00390	-	hypothetical_protein	60
MF00497	lgt	prolipoprotein_diacylglyceryl_tr	58
MF00971	ftsY	cell_division_protein_ftsY	57
MF00115	nrdF	ribonucleoside-diphosphate_red	56
MF00785	proS	prolyl-tRNA_synthetase	56
MF01526	tRNA	Met	56
MF00895	-	hypothetical_protein	55
MF01475	rpmH	50S_ribosomal_protein_L34	55
MF00324	-	ABC_transporter_permease_protein	52
MF01444	-	hypothetical_protein	49
MF00863	-	hypothetical_protein	49
MF00789	-	hypothetical_protein	48
MF00530	gyrA	DNA_gyrase_subunit_A	46
MF01531	tRNA	Phe	46
MF01303	-	hypothetical_protein	45
MF01384	sgaB	PTS_system_enzyme_IIB_component	45
MF00300	-	PTS_system_enzyme_IIC_component	44
MF01459	-	hypothetical_protein	44
MF01349	sgaA	PTS_system,_IIA_component	42
MF00685	aspS	aspartyl-tRNA_synthetase	41
MF01383	-	hypothetical_protein	41
MF00529	deoC	deoxyribose-phosphate_aldolase	41
MF00313	ssb	single-strand_binding_protein	41
MF00524	tufA	elongation_factor_EF-Tu	41

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF01119	obg	GTP-binding_protein_Obg	40
MF01520	tRNA	His	40
MF00136	ldh	L-lactate_dehydrogenase	39
MF00065	-	hypothetical_protein	39
MF00866	-	hypothetical_protein	38
MF00435	-	hypothetical_protein	38
MF00257	ftsH	cell_division_protein	37
MF00993	-	hypothetical_protein	37
MF01105	-	hypothetical_protein	37
MF01278	-	hypothetical_protein	36
MF01528	tRNA	Ser	36
MF01193	-	amino_acid_permease	35
MF00465	-	putative_ABC_transporter_ATP-bind	35
MF00117	rplT	50S_ribosomal_protein_L20	34
MF00110	asnS	asparaginyl-tRNA_synthetase	34
MF01456	rpsP	30S_ribosomal_protein_S16	34
MF00969	hrcA	heat-inducible_transcription_r	34
MF00018	-	hypothetical_protein	34
MF00237	rplD	50S_ribosomal_protein_L4	33
MF01067	pepP	XAA-PRO_aminopeptidase	33
MF01344	(rpsM)	30S ribosomal protein S13	33
MF00111		hypothetical_protein	33
MF00041	ackA	acetate_kinase	33
MF01428	-	hypothetical_protein	33
MF00525	lon	heat_shock_ATP-dependent_protea	33
MF01138	rpsB	30S_ribosomal_protein_S2	31
MF01226	-	amino_acid_permease	31
MF00404	trpS	tryptophanyl-tRNA_synthetase	31
MF01115	licA	PTS_system,_lichenan-specific_I	30
MF01507	tRNA	Tyr	30
MF00598	-	transcriptional_regulator	30
MF00310	-	d-ribulose-5-phosphate_3_epimerase	29
MF00331	rluD	ribosomal_large_subunit_pseudou	29
MF00527	upp	uracil_phosphoribosyltransferase	29
MF00315	ychF	GTP-binding_protein_YchF	29
MF01169	dnaJ	heat_shock_protein_DnaJ	28
MF01144	dnaG	DNA_primase	28
MF01196	deoA	thymidine_phosphorylase	28
MF00871	-	prolipoprotein_P65	28
MF00989	dnaA	chromosomal_replication_initia	28
MF01239	-	hypothetical_protein	27
MF00579	-	hypothetical_protein	27
MF01359	-	hypothetical_protein	26
MF01422	rpsT	30S_ribosomal_protein_S20	26
MF00466	pfkA	6-phosphofructokinase	26
MF00468	gyrB	DNA_gyrase_subunit_B	26
MF01332	rpmB	50S_ribosomal_protein_L28	25
MF00213	adk	adenylate_kinase	25
MF01183	rpsL	30S_ribosomal_protein_S12	25
MF00727	-	ABC_transporter_permease_protein	25
MF01376	rpmF	50S_ribosomal_protein_L32	25
MF00623	-	P102-like_protein	25
MF01517	tRNA	Arg	25

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00357	-	hypothetical_protein	25
MF00131	lip2	lipase-esterase	24
MF01122	atpB	ATP_synthase_A_chain	24
MF00669	serS	seryl-tRNA_synthetase	24
MF00410	nusA	N-utilization_substance_protei	24
MF00440	-	hypothetical_protein	24
MF00235	rplB	50S_ribosomal_protein_L2	23
MF00264	-	ABC_transporter_ATP-binding_prote	23
MF01470	rpmG	50S_ribosomal_protein_L33	23
MF01055	-	hypothetical_protein	23
MF01527	tRNA	Met	23
MF01530	tRNA	Asp	23
MF00113	glyA	glycine_hydroxymethyltransfera	22
MF00080	trmD	tRNA_(guanine-N1)-methyltransfe	22
MF00262	trmE	tRNA_modification_GTPase_ThdF/	22
MF00464	-	hypothetical_protein	22
MF00970	-	hypothetical_protein	22
MF00995	-	hypothetical_protein	22
MF01529	tRNA	Met	22
MF00415	dnaE	DNA_polymerase_III_alpha_subun	22
MF01228		hypothetical_protein	22
MF00737	-	ABC_transporter_permease_protein	21
MF00502	-	hypothetical_protein	21
MF01341	rpiB	ribose_5-phosphate_isomerase_B	21
MF01347	(rplI)	50S ribosomal protein L9	21
MF01460	-	hypothetical_protein	21
MF00853	pgk	phosphoglycerate_kinase	21
MF01140	tsf	elongation_factor_EF-Ts	20
MF01064	-	hypothetical_protein	20
MF00538	-	phosphatidate_cytidylyltransferase	20
MF00209	(rpsK)	30S ribosomal protein S11	20
MF00103	-	hypothetical_protein	20
MF00350	-	hypothetical_protein	20
MF00899	uvrB	excinuclease_ABC_subunit_B	19
MF00386	-	sugar_ABC_transporter_ATP-binding	19
MF00267	fruA	PTS_system_fructose-specific_I	18
MF00079	topA	DNA_topoisomerase_I	18
MF00818	lsp	lipoprotein_signal_peptidase	18
MF00461	-	type_II_DNA_modification_enzyme	18
MF01044	-	ABC_transporter_ATP-binding_-_Pr1	18
MF00495	trxB	thioredoxin_reductase	18
MF01205	-	hypothetical_protein	18
MF00566	ktrB	potassium_uptake_protein	18
MF00388	xylF	ABC_transporter_xylose-binding	18
MF00155	-	putative_tRNA/rRNA_methyltransfera	17
MF01340	rplW	50S_ribosomal_protein_L23	17
MF01158	-	hypothetical_protein	17
MF01535	tRNA	Ile	17
MF01092	ptsI	phosphoenolpyruvate-protein_ph	17
MF00016	-	nicotinate_phosphoribosyltransfer	17
MF01385	-	hypothetical_protein	17
MF00392	-	hypothetical_protein	17
MF00354	rplL	50S_ribosomal_protein_L7/L12	17

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00119	pyk	pyruvate_kinase	16
MF01333	rpmI	50S_ribosomal_protein_L35	16
MF01363	gatC	Asp-tRNAAsn/Glu-tRNAGln_amidotr	16
MF01266	-	GTP-binding_protein	16
MF01350	-	hypothetical_protein	16
MF00021	pepF	oligoendopeptidase_F	16
MF00417	polA	DNA_polymerase_I	16
MF00252	pth	peptidyl-tRNA_hydrolase	15
MF00123	celM	aminopeptidase	15
MF01334	infC	translation_initiation_factor_I	15
MF00797	gatA	glutamyl-tRNA_amidotransferase	15
MF00833	vacB	VACB-like_ribonuclease_II	15
MF01135	atpD	ATP_synthase_beta_chain	15
MF01174	cmk	cytidylate_kinase	15
MF00961	argS	arginyl-tRNA_synthetase	15
MF01015	-	hypothetical_protein	15
MF00504	uvrA	excinuclease_ABC_subunit_A	15
MF01417	rpsO	30S_ribosomal_protein_S15	15
MF00488	-	hypothetical_protein	15
MF00437	pfs	5'-methylthioadenosine_nucleosid	15
MF00122	-	hypothetical_protein	14
MF00210	map	methionine_aminopeptidase	14
MF01299	trmB	tRNA_(guanine-N(7)-)-methyltran	14
MF00958	fba	fructose-bisphosphate_aldolase	14
MF00318	-	hypothetical_protein	14
MF00542	pyrH	uridylate_kinase	14
MF01523	tRNA	Arg	14
MF00069	-	hypothetical_protein	13
MF00708	-	hypothetical_protein	13
MF01420	smpB	ssrA-binding_protein	13
MF00636	mgtA	cation-transporting_P-type_ATP	13
MF01474	(rpmJ)	50S ribosomal protein L36	13
MF01433		hypothetical_protein	13
MF01355	pmsR	methionine_sulfoxide_reductase_	13
MF00345	-	putative_lipoprotein	13
MF00401	-	ATP_binding_protein	13
MF00865	-	hypothetical_protein	13
MF00218	rpsE	30S_ribosomal_protein_S5	12
MF01154	-	hypothetical_protein	12
MF01175	engA	GTP-binding_protein_EngA	12
MF01381	himA	bacterial_nucleoid_DNA-binding_	12
MF01186	-	hypothetical_protein	12
MF01042	-	putative_ABC_transporter_ATP-bind	12
MF00620	-	P97-like_protein	12
MF00559	potA	spermidine/putrescine_ABC_tran	12
MF00047	oppB	oligopeptide_ABC_transporter_p	12
MF01521	tRNA	Leu	12
MF00039	pta	phosphate_acetyltransferase	12
MF00413	infB	translation_initiation_factor_	12
MF00118	infC	translation_initiation_factor_I	11
MF00093	pcrA	ATP-dependent_helicase_PcrA	11
MF01142	glyS	glycyl-tRNA_synthetase	11
MF01508	tRNA	Gln	11

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF01010	-	hypothetical_protein	11
MF01109	-	type_III_restriction-modification	11
MF01216	-	ABC_transporter_ATP-binding_protei	11
MF01467	rpmE	50S_ribosomal_protein_L31	11
MF01232	-	hypothetical_protein	11
MF00408	thrS	threonyl-tRNA_synthetase	11
MF01329	-	hypothetical_protein	11
MF00289	-	hypothetical_protein	10
MF00225	rpsC	30S_ribosomal_protein_S3	10
MF01443	-	hypothetical_protein	10
MF01130	atpA-1	ATP_synthase_alpha_chain	10
MF01120	-	hypothetical_protein	10
MF00936	-	hypothetical_protein	10
MF00945	plsC	1-acyl-sn-glycerol-3-phosphate_	10
MF00751	-	lipoprotein	10
MF00644	-	ABC_transporter_ATP-binding_protei	10
MF00011	metS	methionyl-tRNA_synthetase	10
MF00035	xylG	xylose_ABC_transporter_ATP-bin	10
MF01356	-	hypothetical_protein	10
MF00541	frr	ribosome_recycling_factor	10
MF00526	-	hypothetical_protein	10
MF00492	-	outer_membrane_protein_-_P95	10
MF00358	-	hypothetical_protein	10
MF00424	fpg	foramidopyrimidine_DNA_gycosylas	10
MF00451	atpA	ATP_synthase_alpha_chain	10
MF00306	-	hypothetical_protein	9
MF00116	nrdE	ribonucleoside-diphosphate_red	9
MF00082	-	hypothetical_protein	9
MF01372	-	hypothetical_protein	9
MF01141	ffh	signal_recognition_particle_pro	9
MF00884	-	hypothetical_protein	9
MF00878	gidB	methyltransferase_GidB_(glucose	9
MF00739	-	LppT_protein	9
MF00666	lplA	lipoate-protein_ligase_A	9
MF01510	tRNA	Glu	9
MF00619	pheS	phenylalanyl-tRNA_synthetase_al	9
MF01353	-	hypothetical_protein	9
MF00033	xylH	xylose_ABC_transporter_permeas	9
MF01518	tRNA	Ser	9
MF00023	-	hypothetical_protein	9
MF00568	-	hypothetical_protein	9
MF01221	glpK	glycerol_kinase	9
MF00274	oppB-1	oligopeptide_ABC_transporter	8
MF00291	-	lipoprotein	8
MF01442	rplS	50S_ribosomal_protein_L19	8
MF01177	-	hypothetical_protein	8
MF01465	-	hypothetical_protein	8
MF00815	-	hypothetical_protein	8
MF01005	-	hypothetical_protein	8
MF01509	tRNA	Asn	8
MF01361	-	hypothetical_protein	8
MF01360	-	hypothetical_protein	8
MF00834	-	hypothetical_protein	8

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00531	pgiB	glucose-6-phosphate_isomerase	8
MF01387	rpsR	30S_ribosomal_protein_S18	8
MF01236	-	P60-like_lipoprotein	8
MF00427	nagB	glucosamine-6-phosphate_isomera	8
MF00409	pulA	pullulanase	8
MF00452	-	hypothetical_protein	8
MF00126	ung	uracil-dna_glycosylase	7
MF01201	secA	preprotein_translocase_SecA_su	7
MF01408	-	hypothetical_protein	7
MF00826	parE	topoisomerase_IV_subunit_B	7
MF00934	gcp	O-sialoglycoprotein_endopeptidas	7
MF01357	-	GTP-binding_protein	7
MF00683	-	hypothetical_protein	7
MF00640	-	ABC_transporter_permease_protein	7
MF01430	-	hypothetical_protein	7
MF00330	acpD-1	acyl_carrier_protein_phosphod	7
MF00063	oppF	oligopeptide_ABC_transporter_A	7
MF00533	polC	DNA_polymerase_III_alpha_subun	7
MF00536	polC	DNA_polymerase_III,_alpha_chai	7
MF00019	-	hypothetical_protein	7
MF00429	-	PTS_system_N-acetylglucosamine-sp	7
MF01346	ppa	inorganic_pyrophosphatase	7
MF00233	rplV	50S_ribosomal_protein_L22	6
MF00164	-	chromate_transport_protein	6
MF01342	rplN	50S_ribosomal_protein_L14	6
MF00207	rpoA	DNA-directed_RNA_polymerase_alp	6
MF00820	ileS	isoleucyl-tRNA_synthetase	6
MF01320	-	hypothetical_protein	6
MF01260	-	hypothetical_protein	6
MF01011	greA	transcription_elongation_factor	6
MF01047	-	hypothetical_protein	6
MF00603	-	hypothetical_protein	6
MF00707	-	hypothetical_protein	6
MF01511	tRNA	Val	6
MF00322	-	ABC_transporter_ATP-binding_protei	6
MF01358	-	hypothetical_protein	6
MF00061	oppD-1	oligopeptide_ABC_transporter	6
MF00402	-	ATP_binding_protein_[fragment]	6
MF01367	nrdI	ribonucleotide_reductase_stimul	5
MF00193	deoB	phosphopentomutase	5
MF00813	-	hypothetical_protein	5
MF01147	rpoD	DNA-directed_RNA_polymerase_si	5
MF01180	rpsG	30S_ribosomal_protein_S7	5
MF01126	atpF	ATP_synthase_subunit_B	5
MF01121	dinP	DNA_polymerase_IV	5
MF00938	-	segregation_and_condensation_prote	5
MF01315	-	hypothetical_protein	5
MF00736	-	ABC_transport_permease_protein	5
MF00957	rluC	ribosomal_large_subunit_pseudou	5
MF00653	hpt	hypoxanthine_phosphoribosyltrans	5
MF01525	tRNA	Ala	5
MF01432	-	hypothetical_protein	5
MF00551	-	hypothetical_protein	5

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00044	-	hypothetical_protein	5
MF01264	-	hypothetical_protein	5
MF00595	-	hypothetical_protein	5
MF00420	mnuA	membrane_nuclease,_lipoprotein	5
MF00431	nagA	N-acetylglucosamine-6-phosphat	5
MF00438	-	hypothetical_protein	5
MF00455	-	hypothetical_protein	5
MF01335	-	hypothetical_protein	5
MF01396	rpsS	30S_ribosomal_protein_S19	4
MF01375	-	hypothetical_protein	4
MF00281	oppD	oligopeptide_transport_system_p	4
MF00169	-	hypothetical_protein	4
MF00220	rplF	50S_ribosomal_protein_L6	4
MF00219	rplR	50S_ribosomal_protein_L18	4
MF00188	era	GTP-binding_protein	4
MF00823	parC	topoisomerase_IV_subunit_A	4
MF00829	gap	glyceraldehyde_3-phosphate_dehy	4
MF01382	-	hypothetical_protein	4
MF01131	atpG	ATP_synthase_gamma_chain	4
MF01129	(atpH)	ATP synthase delta chain	4
MF01324	-	hypothetical_protein	4
MF00767	-	ABC_transporter_ATP-binding_prote	4
MF00726	-	lipoprotein	4
MF00724	-	lipoprotein	4
MF00005	-	putative_type_III_restriction-mod	4
MF01074	-	hypothetical_protein	4
MF00975	-	DHH_family_phosphoesterases	4
MF01421	-	hypothetical_protein	4
MF00009	-	hypothetical_protein	4
MF00624	-	nuclease_lipoprotein	4
MF00668	lip	triacylglycerol_lipase	4
MF00718	sgaH	3-hexulose-6-phosphate_synthase	4
MF00720	-	hypothetical_protein	4
MF00609	-	hypothetical_protein	4
MF01208	-	ABC_transport_system_permease_pro	4
MF00328	acpD	acyl_carrier_protein_phosphodie	4
MF01241	-	hypothetical_protein	4
MF01418	-	hypothetical_protein	4
MF00479	clpB	ATP-dependent_protease_binding	4
MF01524	tRNA	Pro	4
MF00025	-	hypothetical_protein	4
MF00029	-	hypothetical_protein	4
MF01515	tRNA	Thr	4
MF00051	oppC	oligopeptide_transport_system_	4
MF00398	rnhB	ribonuclease_HII	4
MF00385	-	ABC_transporter_protein	4
MF00449	atpD	ATP_synthase_beta_chain	4
MF01469	rpl1	50S_ribosomal_protein_L15	3
MF00085	ruvB	holliday_junction_DNA_helicase_	3
MF01394	rpmC	50S_ribosomal_protein_L29	3
MF01398	-	hypothetical_protein	3
MF01399	cmtB	PTS_system_enzyme_IIA_component	3
MF00195	tlyA	haemolysin	3

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF00198	nifS	nitrogen_fixation_protein	3
MF00190	-	ABC_transporter_ATP-binding_and_p	3
MF00221	rplE	50S_ribosomal_protein_L5	3
MF00258	lysS	lysyl-tRNA_synthetase	3
MF00162	-	chromate_transport_protein	3
MF00253	-	hypothetical_protein	3
MF01150	nfo	endonuclease_IV	3
MF01163	-	hypothetical_protein	3
MF00917	-	hemolysin_C	3
MF00905	-	hypothetical_protein	3
MF01284	-	hypothetical_protein	3
MF00753	-	ABC_transporter_permease_protein	3
MF00776	-	hypothetical_protein	3
MF01068	-	hypothetical_protein	3
MF01331	-	hypothetical_protein	3
MF00967	grpE	heat_shock_protein	3
MF01108	-	hypothetical_protein	3
MF00704	-	hypothetical_protein	3
MF00014	-	DNA_methylase	3
MF00659	tmk	thymidylate_kinase	3
MF00639	lig	DNA_ligase	3
MF01516	tRNA	Cys	3
MF00320	metK	S-adenosylmethionine_synthetas	3
MF00564	ktrA	potassium_uptake_protein	3
MF01203	-	hypothetical_protein	3
MF01218	-	P37-like_ABC_transporter_substrat	3
MF01224	glpF	glycerol_uptake_facilitator_pro	3
MF01343	(rplX)	50S ribosomal protein L24	3
MF00892		hypothetical_protein	3
MF01259		hypothetical_protein	3
MF01230	-	hypothetical_protein	3
MF00394	pgm	phosphoglycerate_mutase	3
MF01450	-	hypothetical_protein	3
MF01533	tRNA	SeC	2
MF00071	-	hypothetical_protein	2
MF00209	rpsK	30S_ribosomal_protein_S11	2
MF00073	pgsA	CDP-diacylglycerol--glycerol-3-	2
MF00223	rplP	50S_ribosomal_protein_L16	2
MF01389	infA	translation_initiation_factor_I	2
MF01388	rplQ	50S_ribosomal_protein_L17	2
MF00285	oppF-1	oligopeptide_ABC_transporter	2
MF01370	-	hypothetical_protein	2
MF01393	(rpsQ)	30S ribosomal protein S17	2
MF00342		hypothetical_protein	2
MF00275	oppC-1	oligopeptide_transport_system	2
MF00259	-	hydrolase_of_the_HAD_family	2
MF00143	prfA	peptide_chain_release_factor_R	2
MF00809	gatB	glutamyl-tRNA_amidotransferase	2
MF01153	tyrS	tyrosyl_tRNA_synthetase	2
MF01164	uvrC	excinuclease_ABC_subunit_C	2
MF00909	-	hypothetical_protein	2
MF00943	-	segregation_and_condensation_prote	2
MF01261	-	putative_type_III_restriction-mod	2

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF01406	-	hypothetical_protein	2
MF01348	trxA	thioredoxin	2
MF00996	-	hypothetical_protein	2
MF01072	-	ABC_transporter_ATP-binding_prote	2
MF00978	-	putative_MgpA-like_protein	2
MF00977	-	putative_MgpA_like-protein_[fragme	2
MF00715	sgaE	sugar_isomerase_SgaE	2
MF00678	-	TRSE-like_protein	2
MF00674	smf	DNA_processing_protein_SMF	2
MF00660	recR	recombination_protein_recR	2
MF00645	-	ABC_transporter_ATP-binding_protei	2
MF01227	-	hypothetical_protein	2
MF00552	-	hypothetical_protein	2
MF00544	-	hypothetical_protein	2
MF00423	-	hypothetical_protein	2
MF00279	oppD	oligopeptide_transport_system_p	1
MF00293	mtlA	PTS_system,_mannitol-specific_	1
MF00203	rluB	ribosomal_large_subunit_pseudou	1
MF01441	-	hypothetical_protein	1
MF01403	-	hypothetical_protein	1
MF00095	-	hypothetical_protein	1
MF00152	gltX	glutamyl-tRNA_synthetase	1
MF00075	-	hypothetical_protein	1
MF00185	-	DNA_methylase	1
MF01369	-	hypothetical_protein	1
MF00177	-	GTPase_engC	1
MF00170	-	hypothetical_protein	1
MF01338	-	hypothetical_protein	1
MF01380	-	hypothetical_protein	1
MF00832	-	hypothetical_protein	1
MF01110	cls	cardiolipin_synthetase	1
MF01148	-	hypothetical_protein	1
MF01445	-	hypothetical_protein	1
MF00914	dnaB	replicative_DNA_helicase	1
MF00949	-	hypothetical_protein	1
MF01326	-	hypothetical_protein	1
MF01251	-	subtilisin-like_serine_protease	1
MF00752	gtsA	ABC_transport_ATP-binding_prot	1
MF00001	-	hypothetical_protein	1
MF00986	dnaN	DNA_polymerase_III_beta_subuni	1
MF01070	-	ABC_transporter_ATP-binding_prote	1
MF01073	-	putative_type_III_restriction-mod	1
MF01076	-	hypothetical_protein	1
MF01077	-	hypothetical_protein	1
MF01377	-	hypothetical_protein	1
MF01004	-	hypothetical_protein	1
MF00716	sgaU	hexulose-6-phosphate_isomerase	1
MF00705	trmU	tRNA-(5-methylaminomethyl-2-th	1
MF00663	dnaX	DNA_polymerase_III_gamma_and_t	1
MF01248	-	hypothetical_protein	1
MF01243	-	hypothetical_protein	1
MF01275	-	hypothetical_protein	1
MF00028	glck	glucokinase	1

Continuação *Supplemental Table S2*

Id_ORF	Gene name	Product	Number of Reads
MF01522	tRNA	Lys	1
MF01354	-	hypothetical_protein	1
MF01424	-	HIT-like_protein	1
MF00577	-	hypothetical_protein	1
MF00571	-	hypothetical_protein	1
MF01453	-	hypothetical_protein	1
MF00476	tpiA	triosephosphate_isomerase	1
MF01425	-	hypothetical_protein	1
MF00381	tdk	thymidine_kinase	1
MF01330	-	PTS_system_glucose-specific_enzyme	1
MF00396	-	hypothetical_protein	1
MF00383	-	sugar_ABC_transporter_permease_pro	1
MF00454	-	hypothetical_protein	1
MF00657	-	hypothetical_protein	1
MF01345	(recU)	Holliday junction-specific endonuclease	1
MF01471	(rpsH)	30s ribosomal protein S8	1
MF01472	(rpsN)	30S ribosomal protein S14	1
MF00765		hypothetical_protein	1

Supplemental Table S3. General mapping of transcripts in *M. hyorhinis* genome.

Gene	product	Number of Reads
MHR_0432	c Hexosephosphate transport protein	630
vlpB	Variant surface antigen B	489
vlpG	Variant surface antigen G	487
vlpE	Variant surface antigen E	326
MHR_0162	46kDa surface antigen	322
MHR_t019	c tRNA-Gln	271
MHR_0660	c hypothetical protein	260
rpoC	DNA-directed RNA polymerase subunit beta'	246
gap	c Glyceraldehyde 3-phosphate dehydrogenase C	240
mraZ	Protein mraZ	232
pdhA	c Pyruvate dehydrogenase E1-alpha subunit	222
rpoB	DNA-directed RNA polymerase subunit beta	214
MHR_0152	hypothetical protein	211
tuf	c Elongation factor Tu	194
vlpF	Variant surface antigen F	181
MHR_0358	c hypothetical protein	180
nox	NADH oxidase	174
MHR_0639	c Lipoprotein	171
ftsH	c cell division protease FtsH	170
fusA	translation elongation factor G	168
dnaK	c Hsp70-like protein	159
MHR_0498	Protein P115	154
MHR_0655	c hypothetical protein	152
MHR_0435	c P3	151
secA	c Protein translocase subunit secA	148
MHR_0659	c Lipoprotein	145
pdhD	c pyruvate dehydrogenase E3 component dihydro	145
MHR_0042	hypothetical protein	144
MHR_0318	ABC transporter permease protein	139
rpsO	30S ribosomal protein S15	138
MHR_0460	c hypothetical protein	137
lon	c ATP-dependent protease La	133
polC	c DNA polymerase III alpha subunit	127
dnaJ	c Heat shock protein DnaJ	126
pdhB	c Pyruvate dehydrogenase E1 component beta su	126
ushA	5' nucleotidase precursor	125
MHR_0008	dihydrolipoamide dehydrogenase	121
clpB	ATP-dependent serine proteinase, heat shock prot	115
nrdE	ribonucleotide-diphosphate reductase subunit alph	115
parE	Topoisomerase IV subunit B	115
gyrA	c DNA gyrase subunit A	113
oxaA	c Membrane protein oxaA	107
malC	c ABC transporter, permease protein	106
MHR_0186	c Putative maltose phosphorylase domain protein	104
rplU	50S ribosomal protein L21	104
eno	Enolase	102
gyrB	DNA gyrase subunit B	102
ldh.1	L-lactate dehydrogenase	100
MHR_0006	Putative MgpA-like protein	98
MHR_0154	hypothetical Serine-rich adhesin for platelets	98
MHR_0198	c Lipoprotein	96

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
glyA	Serine hydroxymethyltransferase 3	95
infB	c Translation initiation factor IF-2	95
ulaA	c Ascorbate-specific permease IIC component ula	95
lepA	GTP-binding protein lepA	94
uvrA	c Excinuclease ATPase subunit-like protein	92
MHR_0585	c ABC transporter ATP-binding protein	91
atpD.1	Proton-translocation ATPase, beta subunit	91
fruA	c Fructose permease IIC component	91
secD	Protein-export membrane protein	91
MHR_0205	c Aminopeptidase	90
MHR_0601	c Putative ATP-binding helicase protein	86
valS	c Valyl tRNA synthetase	86
MHR_0677	c Lipoprotein	85
dnaX	DNA polymerase III gamma and tau subunit	85
obgE	GTPase obg	85
oppD.1	c Oligopeptide ABC transporter ATP binding prote	85
pdhC	c Dihydrolipoamide acetyltransferase component (	85
MHR_0258	hypothetical protein	84
MHR_0586	c hypothetical Y+L amino acid transporter 1	84
amyC	alpha-amylase	84
gidA	tRNA uridine 5-carboxymethylaminomethyl modific	84
MHR_0021	Lipoprotein	82
mnmA	c tRNA-specific 2-thiouridylase mnmA	80
uvrB	c Excinuclease ABC subunit B	80
MHR_0153	hypothetical Serine-rich adhesin for platelets	79
trxB	c Thioredoxin reductase	79
MHR_0298	c hypothetical protein	78
deoD	c Probable purine nucleoside phosphorylase trans	78
pfkA	c 6-phosphofructokinase	78
atpA.1	c ATP synthase subunit A	77
metK	c S-adenosylmethionine synthetase	77
tkt	Probable transketolase transmembrane protein	77
deoA	c Thymidine phosphorylase	75
MHR_0012	c hypothetical protein	74
atpA.2	ATP synthase subunit alpha	73
rpsL	Ribosomal protein S12-like protein	73
alaS	c alanyl tRNA synthetase	72
pyrG	CTP synthase	72
xylG	xylose ABC transporter ATP-binding protein	72
MHR_0625	c High affinity transport system protein p37	71
aspS	aspartyl tRNA synthetase	70
ftsZ	cell division protein FtsZ	70
MHR_0357	c Lipoprotein	69
xylH	Xylose ABC transporter permease protein	68
MHR_0062	uncharactized lipoprotein	67
MHR_0137	hypothetical protein	67
gpmI	c 2,3-bisphosphoglycerate-independent phosphog	67
MHR_0061	uncharactized lipoprotein	65
MHR_0299	c Putative lipoprotein	65
MHR_0073	P59-like protein	64
trmE	c tRNA modification GTPase mnmE	64
vlpD	Variant surface antigen D	64

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0285	c hypothetical protein	63
lgt	c Prolipoprotein diacylglyceryl transferase	63
malD	c Maltodextrin ABC transporter, permease MalD	63
leuS	Leucyl-trna synthetase protein	62
oppC.1a	c Oligopeptide transport system permease protein	62
MHR_0579	putative sialic acid transport	61
MHR_0628	c hypothetical protein	60
dnaE	c DNA polymerase III alpha subunit	60
pgi.1	Glucose-6-phosphate isomerase A	60
pgk	Phosphoglycerate kinase	60
MHR_0007	MgpA-like DHH family phosphoesterase	59
MHR_0074	Amino acid or sugar ABC transport system, permea	59
MHR_0261	permease	59
MHR_0465	c hypothetical protein	59
recA	c recombinase A	58
MHR_0673	ATPase AAA-2 domain protein	57
MHR_r003	c 5S ribosomal RNA	57
tig	c Trigger factor	57
MHR_0187	c Sucrase-isomaltase	56
atpD.2	ATP synthase subunit B	56
parC	Topoisomerase IV subunit A	56
MHR_0374	Restriction endonuclease S subunits	55
pyk	c Pyruvate kinase	55
thrS	Threonyl-tRNA synthetase	55
MHR_0194	c ABC-type maltose/maltodextrin transporter ATP-	54
nagB	Glucosamine-6-phosphate isomerase 1	54
MHR_0024	hypothetical protein	53
MHR_0277	c Lipoprotein	53
MHR_0644	hypothetical protein	53
MHR_0610	c hypothetical 33 kDa chaperonin	52
fba	Fructose-bisphosphate aldolase class II transmeml	52
gatA	Glutamyl-tRNA amidotransferase subunit A	52
pcrA	c ATP-dependent helicase PcrA	52
topA	c DNA topoisomerase I	52
MHR_0085	hypothetical protein	51
MHR_0109	ABC transporter, ATP-binding protein	51
MHR_0629	c membrane protease subunits, stomatin/prohibitir	51
acpD.2	FMN-dependent NADH-azoreductase 1	51
engA	Predicted GTPase protein	51
ldh.2	L-lactate dehydrogenase 2	51
oppD.1a	Oligopeptide transport system permease protein	51
argS	Arginyl-tRNA synthetase 1	50
glpD	Glycerol-3-phosphate dehydrogenase, putative	50
pepA	aminopeptidase	50
pepP	Xaa-pro aminopeptidase	50
pgi.1a	Glucose-6-phosphate isomerase A	50
secY	c Preprotein translocase SecY subunit	50
MHR_0028	hypothetical protein	49
mgs1	c ATPase family associated with various cellular a	49
oppF.1	Oligopeptide transport system permease protein	49
pyrH	c Uridylate kinase smbA	49
rpoA	c DNA-directed RNA polymerase subunit alpha	48

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0225	101 kDa protein	47
nrdF	Ribonucleoside-diphosphate reductase beta chai	47
rpoD	c RNA polymerase sigma factor	47
MHR_0250	c Na/Pi cotransporter II-related protein	46
mnuA.1a	Membrane nuclease, lipoprotein	46
serS	c Seryl-trna synthetase protein	46
MHR_0293	hypothetical protein	45
MHR_0529	c Amino acid permease	45
MHR_t020	c tRNA-Tyr	45
ftsY	Cell division protein ftsY	45
rpmF	50S L32	45
MHR_0538	c hypothetical protein	44
lig	DNA ligase, NAD-dependent	44
ntpJ	Cation transport protein	44
tyrS	Tyrosyl tRNA synthetase 1	44
MHR_0672	CLP-like protein	43
cbiO	ABC transporter, ATP-binding protein	43
ugpE	ABC transporter permease protein	43
MHR_0440	c Outer membrane protein-P95	42
MHR_0495	Predicted kinase, related to dihydroxyacetone kina	42
oppB.1	Oligopeptide ABC transporter permease protein	42
oppF.1a	Oligopeptide transport system permease protein	42
rpsJ	c 30S ribosomal protein S10	42
MHR_0043	101 kDa protein	41
MHR_0433	c MG2+ ion transporter	41
ackA	Acetate kinase AckA	41
hisS	Histidyl-tRNA synthetase	41
mtlA	c PTS system Mannitol-specific transport subunit II	41
rplD	c 50S ribosomal protein L4	41
rpsE	c 30S ribosomal protein S5	41
MHR_0598	c hypothetical protein	40
asnA	Aspartate--ammonia ligase, AsnA-type	40
mnuA.1	Membrane nuclease, lipoprotein	40
pepF.1	Oligoendopeptidase F	40
vlpC	Variant surface antigen C	40
MHR_0076	Ag 243-5 protein	39
MHR_0146	Amino acid permease	39
MHR_0276	c Glycerol ABC transporter, glycerol binding protei	39
MHR_0557	c hypothetical protein	39
asnC	c Putative asparaginyl-trna synthetase protein	39
deoB	c Phosphopentomutase	39
lysS	c Lysyl-tRNA synthetase 1	39
rplJ	50S ribosomal protein L10	39
trkA	c TrkA-C domain protein	39
MHR_0444	c hypothetical protein	38
proS	c Prolyl-tRNA synthetase 2	38
ptsI	c Phosphotransferase system (PTS) enzyme I	38
rplC	c 50S ribosomal protein L3	38
MHR_0064	ABC transporter permease protein	37
MHR_0247	Ser/Thr protein phosphatase family protein	37
MHR_0354	hypothetical protein	37
MHR_0675	c hypothetical protein	37

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
eutD	Phosphate acetyltransferase	37
MHR_0059	hypothetical protein	36
MHR_0181	putative serine protease	36
MHR_0297	c COF family HAD hydrolase protein	36
MHR_0336	c Nicotinate phosphoribosyltransferase	36
dcm	c DNA-cytosine methyltransferase family protein	36
glyS	c Glycyl tRNA synthetase	36
MHR_0071	Cytosine-specific methyltransferase	35
MHR_0093	hypothetical protein	35
araD	c Sugar isomerase SgaE	35
ileS	Isoleucyl tRNA synthetase	35
mraW	Ribosomal RNA small subunit methyltransferase H	35
rplB	c 50S ribosomal protein L2	35
rpsC	c 30S ribosomal protein S3	35
MHR_0075	Sugar ABC transporter permease protein	34
MHR_t013	c tRNA-Ser	34
ffh	c Signal recognition particle protein	34
gcp	Putative glycoprotease GCP	34
pheT	c Phenylalanyl-tRNA synthetase beta chain	34
rpiB	c Ribose 5-phosphate isomerase B	34
trmD	c tRNA (guanine-N(1)-)-methyltransferase	34
MHR_0045	hypothetical protein	33
MHR_0100	Fatty acid-binding protein DegV-like protein	33
MHR_0294	hypothetical protein	33
MHR_0546	c hypothetical protein	33
MHR_0553	c Zinc metalloproteinase C	33
dnaA	Chromosomal replication initiator protein dnaA	33
trpS	Tryptophanyl-trna synthetase protein	33
zmpC	Zinc metalloproteinase C	33
MHR_0077	UPF0348 protein	32
MHR_t006	tRNA-Leu	32
lplA.1a	Lipoyltransferase and lipoate-protein ligase	32
prfA	Peptide chain release factor 1	32
rpsF	c 30S ribosomal protein S6	32
ssb	c Single-stranded DNA-binding protein	32
MHR_0063	Lipoprotein	31
MHR_0179	c adenine DNA methyltransferase subunit	31
MHR_0600	c Probable exported protein	31
dnaG	c DNA primase	31
gatB	Glutamyl-tRNA amidotransferase subunit B	31
rpsD	30S S4	31
trxA	c Thiol-disulfide isomerase and thioredoxin	31
ugpA	ABC transport permease protein	31
vacB	c VACB-like ribonuclease II	31
MHR_0044	transcriptional regulator	30
MHR_0135	Subtilisin-like serine protease	30
MHR_0483	putative ICEF-II	30
MHR_0551	hypothetical protein	30
MHR_0627	c P60-like lipoprotein	30
MHR_0667	DNA2-like helicase	30
gtp1	c Predicted GTPase, probable translation factor	30
nrdI	ribonucleotide reductase stimulatory protein	30

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
oppC.1	Oligopeptide transport system permease protein	30
prsA	Ribose-phosphate pyrophosphokinase	30
rplK	50S ribosomal protein L11	30
MHR_0011	c hypothetical protein	29
MHR_0328	Expressed protein	29
cysS	Cysteinyl-tRNA synthetase	29
infC	c translation initiation factor IF-3	29
nadD	c probable nicotinate-nucleotide adenylyltransfera	29
ppa	c Pyrophosphatephospho hydrolase	29
tpiA	Triosephosphate isomerase	29
MHR_0122	c ABC transporter permease protein	28
MHR_0233	Predicted signaling protein	28
atpE	ATP synthase C chain, sodium ion specific lipid-bir	28
gpsA	Glycerol-3-phosphate dehydrogenase [NAD(P)+]	28
rplP	c 50S ribosomal protein L16	28
MHR_0015	Integral membrane protein	27
MHR_0150	c ABC-type multidrug transport system, ATPase a	27
MHR_0439	c hypothetical protein	27
MHR_0523	c hypothetical protein	27
apt	c Adenine phosphoribosyltransferase	27
gltX	Glutamyl-tRNA synthetase	27
hrcA	Heat-inducible transcription repressor hrcA	27
pacL	Cation-transporting ATPase family protein	27
pepF.1a	Oligoendopeptidase F	27
potB	c spermidine/putrescine ABC transporter permeas	27
rplM	50S ribosomal protein L13	27
rpmA	50S ribosomal protein L27	27
MHR_0019	hypothetical protein	26
MHR_0249	101 kDa protein	26
MHR_0265	c Fatty acid-binding protein DegV-like protein	26
MHR_0283	c hypothetical protein	26
MHR_0508	c UPF0082 protein	26
MHR_t023	c tRNA-Leu	26
rpsB	c 30S ribosomal protein S2	26
MHR_0177	c Site-specific DNA-methyltransferase, Adenine-sp	25
MHR_0208	c 5'-nucleotidase, lipoprotein e(P4) family	25
MHR_0450	c Transporter, SSS family	25
MHR_0643	c ABC transporter, permease protein	25
MHR_t010	c tRNA-Phe	25
dnaC	Replicative helicase DnaB	25
grpE	Heat shock protein	25
himA	Histone-like DNA-binding protein	25
map	c Methionine aminopeptidase	25
oppB.1a	Oligopeptide transport system permease protein	25
plsX	Phosphate acyltransferase	25
ribF	c Riboflavin biosynthesis protein	25
truB	c tRNA pseudouridine synthase B	25
adk	c adenylate kinase	24
era	GTP-binding protein era-like protein	24
nusA	c Transcription termination-antitermination factor n	24
potA	c ABC transporter, ATP-binding protein	24
MHR_0047	hypothetical protein	23

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0097	c Triacylglycerol lipase	23
MHR_0106	Ribosomal RNA small subunit methyltransferase I	23
MHR_t018	tRNA-His	23
hpt	Hypoxanthine phosphoribosyltransferase	23
lplA.1	lipoate-protein ligase A	23
rpsI	30S ribosomal protein S9	23
MHR_0046	hypothetical protein	22
MHR_0434	c Expressed protein	22
MHR_0487	c ABC transporter xylose-binding lipoprotein	22
MHR_0559	transposase	22
MHR_0581	N-acetylmannosamine-6-phosphate 2-epimerase	22
MHR_0597	hypothetical protein	22
pulA	c Alpha amylase, catalytic domain protein	22
rluD	ribosomal large subunit pseudouridine synthase D	22
rplL	50S ribosomal protein L7/L12	22
tdk	Thymidine kinase	22
MHR_0355	Probable integrase/recombinase ripx	21
MHR_0544	c Phenylalanyl-tRNA synthetase, alpha subunit	21
bcrA.1	ABC-type multidrug-like transport system ATP-binc	21
rpsT	c 30S ribosomal protein S20	21
ulaD	c hexulose 6 phosphate synthase	21
MHR_0034	hypothetical protein	20
MHR_0174	c hypothetical protein	20
MHR_0481	hypothetical protein	20
MHR_0482	TraE/TrsE-like membrane protein	20
MHR_0676	c hypothetical protein	20
deoC	c Deoxyribose-phosphate aldolase 1	20
frr	c Ribosome recycling factor	20
gidB	Ribosomal RNA small subunit methyltransferase G	20
rnhB	Ribonuclease HII	20
rplA	50S ribosomal protein L1	20
rplE	c 50S ribosomal protein L5	20
thyA	c Thymidylate synthase 2	20
MHR_0017	hypothetical protein	19
MHR_0203	c DNA mismatch repair protein	19
MHR_0240	RNA methyltransferase, TrmH family, group 3	19
MHR_0279	c Oxidoreductase, FAD/FMN-binding	19
MHR_0447	transposase	19
MHR_0451	c Predicted sialic acid transporter	19
MHR_0574	Putative transcriptional regulator	19
MHR_0630	c hypothetical protein	19
MHR_0668	hypothetical protein	19
MHR_t024	c tRNA-Thr	19
msrA	Peptide methionine sulfoxide reductase msrA	19
nadE	c NH(3)-dependent NAD(+) synthetase	19
rpsG	30S ribosomal protein S7	19
MHR_0119	uncharactized deoxyribonuclease yabD	18
MHR_0425	c Protein phosphatase 2C-like protein	18
MHR_0427	c hypothetical Ribosomal RNA small subunit meth	18
MHR_0555	c hypothetical protein	18
MHR_0681	transcription elongation factor	18
acpD.1	FMN-dependent NADH-azoreductase	18

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
glcK	Glucokinase	18
MHR_0180	c Putative type III restriction-modification system: r	17
MHR_0246	hypothetical protein	17
MHR_0416	Probable 1-acyl-sn-glycerol-3-phosphate acyltransf	17
MHR_0558	c hypothetical protein	17
MHR_0631	hypothetical protein	17
ksgA	Ribosomal RNA small subunit methyltransferase A	17
nfo	c Probable endonuclease 4	17
sgaU	c Probable L-ribulose-5-phosphate 3-epimerase ul	17
MHR_0202	c Restriction endonuclease S subunits	16
MHR_0485	c DNA polymerase I	16
MHR_0577	c N-acetylmannosamine kinase; transcriptional reg	16
MHR_0658	c hypothetical membrane protein	16
MHR_0665	Possible beta-galactosidase, beta subunit	16
atpB	ATP synthase subunit a	16
cls	c Cardiolipin synthetase	16
rplW	c 50S ribosomal protein L23	16
uvrC	c Excinuclease ABC subunit C	16
MHR_0228	hypothetical protein	15
MHR_0319	ABC transporter ATP-binding protein	15
MHR_0419	hypothetical protein	15
dctD	deoxycytidylate deaminase	15
fpg	c Formamidopyrimidine-DNA glycosylase, MutM	15
rbgA	c GTP-binding protein	15
rplN	c 50S ribosomal protein L14	15
rplO	c 50S ribosomal protein L15	15
rpsK	c 30S ribosomal protein S11	15
tsf	c Elongation factor Ts	15
ugpQ	Glycerophosphodiester phosphodiesterase family p	15
upp	c Uracil phosphoribosyltransferase	15
MHR_0189	c GntR-family transcriptional regulator	14
MHR_0530	c Hydrolase of the HAD family	14
MHR_0562	c hypothetical protein	14
MHR_0602	c Putative PTS system glucose-specific enzyme II	14
MHR_0678	c hypothetical protein	14
rplV	c 50S ribosomal protein L22	14
MHR_0010	c hypothetical protein	13
MHR_0041	Putative ABC transporter ATP-binding-Pr2-like prot	13
MHR_0155	c hypothetical Serine-rich adhesin for platelets	13
MHR_0178	c Putative type III restriction-modification system: r	13
MHR_0260	hypothetical protein	13
MHR_0368	Uncharacterized domain/protein associated with RI	13
MHR_0449	c hypothetical protein	13
MHR_0471	transposase	13
MHR_0480	DNA processing protein SMF	13
MHR_0499	transposase	13
MHR_0684	dihydrofolate reductase	13
efp	Elongation factor P	13
gmk	c Guanylate kinase Gmk	13
nanA	N-acetylneuraminate lyase	13
rplF	c 50S ribosomal protein L6	13
rsgA	c Putative ribosome biogenesis GTPase rsgA	13

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
tmk	Thymidylate kinase	13
ung	c Uracil-DNA glycosylase	13
MHR_0027	Putative lipoprotein	12
MHR_0033	Cytosine-specific methyltransferase	12
MHR_0078	hypothetical protein	12
MHR_0079	hypothetical protein	12
MHR_0182	c ATPase	12
MHR_0255	c Putative DNase/RNase endonuclease	12
MHR_0274	hypothetical protein	12
MHR_0306	c hypothetical protein	12
MHR_0429	Glycerol-3-phosphate acyltransferase	12
MHR_0623	c ABC transport system permease protein p69	12
MHR_0624	c Probable ABC transporter ATP-binding protein p	12
MHR_0666	hypothetical protein	12
cmk	Cytidylate kinase	12
engB	GTPase EngB	12
rplX	c 50S ribosomal protein L24	12
rpsM	c 30S ribosomal protein S13	12
smpB	c Trans-translation protein, binds tmRNA and tRN	12
MHR_0016	hypothetical protein	11
MHR_0173	transposase	11
MHR_0236	Hemolysin C	11
MHR_0268	c hypothetical membrane protein	11
MHR_0275	transposase	11
MHR_0278	Amidohydrolase 3	11
MHR_0588	hypothetical protein	11
dinP	DNA polymerase IV	11
lspA	Signal peptidase II	11
napA	c Neutrophil activating factor	11
nusG	Transcription antitermination protein	11
rplS	c 50S ribosomal protein L19	11
rplT	c 50S ribosomal protein L20	11
rpmE	c 50S ribosomal protein L31	11
MHR_0050	hypothetical protein	10
MHR_0175	c Type III restriction-modification system methylas	10
vlpA	Variant surface antigen A	10
MHR_0493	hypothetical protein	10
MHR_0533	c tRNA(Ile)-lysidine synthase	10
MHR_0670	hypothetical protein	10
MHR_t007	c tRNA-Arg	10
MHR_t016	c tRNA-Ala	10
potC	c Spermidine/putrescine transport system permea	10
rluB	c ribosomal large subunit pseudouridine synthase	10
rsmE	c Ribosomal RNA small subunit methyltransferase	10
ruvB	c Holliday junction ATP-dependent DNA helicase r	10
tpx	c Thiol peroxidase	10
trmB	tRNA (guanine-N(7)-)-methyltransferase	10
MHR_0048	transposase	9
MHR_0110	ABC transporter permease protein	9
MHR_0149	c ABC-type multidrug transport system, ATPase a	9
MHR_0211	Phosphotransferase enzyme family protein	9
MHR_0281	c NAD-dependent protein deacetylase SIR2 family	9

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0286	c hypothetical protein	9
MHR_0325	putative tRNA/rRNA methyltransferase	9
MHR_0326	Chromate transport protein, putative	9
MHR_0327	Chromate ion transporter	9
MHR_0437	transposase	9
MHR_0594	c Phosphatidate cytidylyltransferase synthase	9
MHR_0616	c hypothetical protein	9
MHR_t001	tRNA-Ile	9
MHR_t015	c tRNA-Met	9
rplQ	c 50S ribosomal protein L17	9
rpsS	c 30S ribosomal protein S19	9
trmA	Uncharacterized RNA methyltransferase	9
MHR_0087	putative Thioredoxin	8
MHR_0102	UPF0133 protein	8
MHR_0167	c hypothetical protein	8
MHR_0193	c hypothetical protein	8
MHR_0229	transposase	8
MHR_0245	N-acetylglucosamine-6-phosphate deacetylase	8
MHR_0314	hypothetical protein	8
MHR_0389	c 30S ribosomal protein S14 type Z	8
MHR_0407	c UPF0135 protein	8
MHR_0448	c hypothetical protein	8
MHR_0504	transposase	8
MHR_0554	c putative intracellular protease/amidase	8
MHR_0590	DUTPase	8
MHR_0652	hypothetical protein	8
MHR_0685	c 50S ribosomal protein L34	8
MHR_t003	tRNA-Leu	8
MHR_t012	c tRNA-Met	8
dnaN	DNA polymerase III beta subunit	8
fruK	1-phosphofructokinase	8
infA	c bacterial translation initiation factor 1	8
nagC	c glucokinase	8
rnc	Ribonuclease III	8
rpsH	c 30S ribosomal protein S8	8
rpsR	c 30S ribosomal protein S18	8
MHR_0022	Lipoprotein	7
MHR_0227	c hypothetical protein	7
MHR_0438	transposase	7
MHR_0453	c hypothetical protein	7
MHR_0458	c Pentitol phosphotransferase enzyme II, B compo	7
MHR_0461	c hypothetical protein	7
MHR_0484	putative hydrolase	7
MHR_0492	c hypothetical protein	7
atpF	ATP synthase subunit b	7
atpG	ATP synthase gamma chain	7
atpH	ATP synthase subunit delta	7
bcrA.2	ABC-type multidrug-like transport system ATP-binc	7
nifS	c Aminotransferase, class V	7
rluC	Ribosomal large subunit pseudouridine synthase C	7
rmpC	c 50S ribosomal protein L29	7
ruvA	c Holliday junction ATP-dependent DNA helicase r	7

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0140	hypothetical protein	6
MHR_0310	ABC transporter ATP-binding and permease protei	6
MHR_0342	c hypothetical protein	6
MHR_0589	LemA-family protein	6
MHR_0647	c hypothetical protein	6
gtsA	c Glycerol ABC transporter, ATP bindingcompone	6
pgmB	c putative beta-phosphoglucomutase	6
recR	Recombination protein	6
rplR	c 50S ribosomal protein L18	6
scpB	Segregation and condensation protein B	6
MHR_0014	hypothetical protein	5
MHR_0058	hypothetical protein	5
MHR_0105	DNA polymerase III subunit delta	5
MHR_0183	transposase	5
MHR_0367	c hypothetical protein	5
MHR_0490	c Putative transciprtional termination factor	5
MHR_0510	c Mannose-6-phosphate isomerase, type 1	5
MHR_0556	c hypothetical protein	5
MHR_0605	c Peptide methionine sulfoxide reductase MsrA/ms	5
MHR_0674	c hypothetical protein	5
MHR_t005	tRNA-Arg	5
MHR_t009	tRNA- OTHER	5
MHR_t014	c tRNA-Met	5
MHR_t022	tRNA- thr	5
MHR_t029	c tRNA-Leu	5
lip3	c lipase/esterase family	5
metS	c methionyl tRNA synthetase	5
rpmB	c 50S ribosomal protein L28	5
MHR_0096	transposase	4
MHR_0138	hypothetical protein	4
MHR_0139	hypothetical protein	4
MHR_0165	hypothetical protein	4
MHR_0184	c hypothetical protein	4
MHR_0197	c hypothetical protein	4
MHR_0375	c Type I site-specific DNA methyltransferase spec	4
MHR_0507	c hypothetical protein	4
MHR_0569	transposase	4
MHR_0582	hypothetical protein	4
MHR_0613	c hypothetical protein	4
MHR_0626	c Uncharacterized 13.1 kDa HIT-like protein in P3	4
MHR_0671	hypothetical protein	4
MHR_t002	tRNA-Ser	4
MHR_t021	c tRNA-Trp	4
nifU	c aminotransferase protein U-like protein	4
recU	c Holliday junction resolvase recU	4
scpA	Segregation and condensation protein A	4
sgaA	c Pentitol phosphotransferase enzyme II, A compc	4
MHR_0003	S4-like RNA binding protein	3
MHR_0018	hypothetical protein	3
MHR_0037	transposase	3
MHR_0040	ABC transporter ATP-binding-Pr1-like protein	3
MHR_0082	hypothetical protein	3

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0113	c Sugar-phospahte nucleotidyltransferase	3
MHR_0121	c ABC transporter permease protein	3
MHR_0129	hypothetical protein	3
MHR_0199	c hypothetical protein	3
MHR_0251	c hypothetical protein	3
MHR_0253	hypothetical protein	3
MHR_0269	c Glycosyltransferase	3
MHR_0335	c Methyltransferase small domain protein	3
MHR_0540	c hypothetical Phosphatase yidA	3
MHR_0648	c Expressed protein	3
MHR_0669	hypothetical protein	3
MHR_t011	c tRNA-Asp	3
MHR_t017	c tRNA-Pro	3
MHR_t026	c tRNA-Glu	3
MHR_t030	tRNA-Gly	3
cspR	Putative RNA methyltransferase, TrmH family, grou	3
dhfR	Bifunctional protein including dihydrofolate reducta	3
dnaI	Primosomal protein DnaI	3
hemK	Protein-(Glutamine-N5) methyltransferase, release	3
rpe	c Ribulose-phosphate 3-epimerase	3
rpmG	50S ribosomal protein L33	3
MHR_0026	Lipoprotein, putative	2
MHR_0086	hypothetical protein	2
MHR_0091	hypothetical protein	2
MHR_0214	hypothetical protein	2
MHR_0238	hypothetical metal-dependent hydrolase	2
MHR_0282	c glycine cleavage system H protein	2
MHR_0295	c 5'-methylthioadenosine nucleosidase / S-adenos	2
MHR_0311	Putative metalloprotease	2
MHR_0366	c hypothetical protein	2
MHR_0372	ABC transporter ATP-binding protein/permease- pr	2
MHR_0430	c hypothetical protein	2
MHR_0462	c hypothetical protein	2
MHR_0494	hypothetical protein	2
MHR_0536	c hypothetical protein	2
MHR_0614	c hypothetical protein	2
MHR_0615	hypothetical protein	2
MHR_0617	c hypothetical protein	2
MHR_0618	c hypothetical protein	2
MHR_0621	hypothetical protein	2
MHR_0632	Chromosome replication initiation and membrane a	2
MHR_0642	c SUA5-like translation suppressor	2
MHR_0656	c hypothetical protein	2
MHR_0657	c hypothetical protein	2
MHR_t025	c tRNA-Val	2
atpC	H+-transporting two-sector ATPase, delta/epsilon s	2
cdd	Methyltransferase type 11	2
hsdM	c Type I restriction-modification system methyltran	2
hsdR	c Type I site-specific deoxyribonuclease, HsdR far	2
rplI	50S ribosomal protein L9	2
rpmI	c 50S ribosomal protein L35	2
rpsP	c 30S ribosomal protein S16	2

Continuação *Supplemental Table S3*

Gene	product	Number of Reads
MHR_0036	hypothetical protein	1
MHR_0051	hypothetical protein	1
MHR_0052	hypothetical protein	1
MHR_0057	ATPase, YjeE family	1
MHR_0072	hypothetical protein	1
MHR_0099	c Expressed protein	1
MHR_0112	c hypothetical protein	1
MHR_0142	transposase	1
MHR_0172	c hypothetical protein	1
MHR_0176	c Putative type III restriction-modification system: r	1
MHR_0190	transposase	1
MHR_0213	hypothetical protein	1
MHR_0241	hypothetical protein	1
MHR_0254	hypothetical protein	1
MHR_0266	c hypothetical protein	1
MHR_0292	hypothetical protein	1
MHR_0370	c ABC transporter ATP-binding protein/permease -	1
MHR_0428	c Rec0	1
MHR_0445	hypothetical protein	1
MHR_0464	c Preprotein translocase, SecG subunit	1
MHR_0513	RNA methyltransferase, TrmA family protein	1
MHR_0535	c Holliday junction resolvase, putative	1
MHR_0583	hypothetical protein	1
MHR_0607	transposase	1
MHR_0634	c hypothetical protein	1
MHR_0682	hypothetical protein	1
MHR_0683	c 50S ribosomal protein L36	1
MHR_t004	tRNA-Cys	1
MHR_t008	tRNA-sec	1
pth	c Peptidyl-tRNA hydrolase	1
rbfA	c Ribosome-binding factor A	1

Supplemental Table S4. *M. hyopneumoniae* genes no mapping

Gene ID_NCBI	Gene ID	Gene name	Product
MP18932	-	-	hypothetical_protein
MP24345	-	-	hypothetical_protein
MP24346	-	-	hypothetical_protein
MP24347	-	-	hypothetical_protein
MP24609	-	-	hypothetical_protein
MP24620	-	-	hypothetical_protein
MP24669	-	-	hypothetical_protein
MP24671	-	-	hypothetical_protein
MP24672	-	-	hypothetical_protein
MP24674	-	-	hypothetical_protein
MP24675	-	-	hypothetical_protein
MP24676	-	-	hypothetical_protein
MP24681	-	-	hypothetical_protein
MP24685	-	-	hypothetical_protein
MP24720	-	-	hypothetical_protein
MP24728	-	-	hypothetical_protein
MP24730	-	-	hypothetical_protein
MP24735	-	-	hypothetical_protein
MP24461	MHP7448_0004	-	hypothetical_protein
MP24561	MHP7448_0007	-	hypothetical_protein
MP12631	MHP7448_0013	-	hypothetical_protein
MP24633	MHP7448_0016	-	hypothetical_protein
MP06686	MHP7448_0017	-	hypothetical_protein
MP24599	MHP7448_0018	-	hypothetical_protein
MP24668	MHP7448_0036	-	hypothetical_protein
MP12596	MHP7448_0046	-	hypothetical_protein
MP06051	MHP7448_0063	tyrS	tyrosyl_tRNA_synthetase
MP10535	MHP7448_0078	-	hypothetical_protein
MP24628	MHP7448_0088	-	hypothetical_protein
MP24622	MHP7448_0089	-	hypothetical_protein
MP01224	MHP7448_0106	-	hypothetical_protein
MP02468	MHP7448_0112	-	hypothetical_protein
MP01950	MHP7448_0155	-	hypothetical_protein
MP24607	MHP7448_0159	-	hypothetical_protein
MP24534	MHP7448_0165	-	hypothetical_protein
MP10670	MHP7448_0203	-	hypothetical_protein
MP24378	MHP7448_0219	-	hypothetical_protein
MP02928	MHP7448_0237	-	hypothetical_protein
MP24693	MHP7448_0238	-	hypothetical_protein
MP24367	MHP7448_0252	-	hypothetical_protein
MP24598	MHP7448_0257	-	hypothetical_protein
MP24582	MHP7448_0261	-	hypothetical_protein
MP24631	MHP7448_0301	-	hypothetical_protein
MP24518	MHP7448_0311	-	hypothetical_protein
MP01341	MHP7448_0315	-	ABC_transporter_ATP-binding_protein
MP24645	MHP7448_0325	-	hypothetical_protein
MP12564	MHP7448_0329	-	hypothetical_protein
MP24529	MHP7448_0330	-	hypothetical_protein
MP24627	MHP7448_0331	-	hypothetical_protein
MP07160	MHP7448_0340	-	ABC_transporter_ATP-binding_protein
MP24543	MHP7448_0341	-	hypothetical_protein
MP24564	MHP7448_0345	-	hypothetical_protein
MP24413	MHP7448_0350	-	hypothetical_protein

Continuação *Supplemental Table S4*

Gene ID_NCBI	Gene ID	Gene name	Product
MP18582	MHP7448_0356	-	hypothetical_protein
MP02301	MHP7448_0363	-	hypothetical_protein
MP24591	MHP7448_0364	-	hypothetical_protein
MP24608	MHP7448_0365	-	hypothetical_protein
MP24661	MHP7448_0367	-	lipoprotein
MP00334	MHP7448_0374	sgaA	PTS_system,_IIA_component
MP24498	MHP7448_0381	-	ABC_transporter_permease_protein
MP18611	MHP7448_0404	-	hypothetical_protein
MP24421	MHP7448_0412	-	hypothetical_protein
MP24442	MHP7448_0413	-	hypothetical_protein
MP04623	MHP7448_0415	-	putative_ICEF_Integrative_Conjugal_Element_-_II
MP24451	MHP7448_0417	-	hypothetical_protein
MP12603	MHP7448_0421	-	hypothetical_protein
MP24612	MHP7448_0422	-	hypothetical_protein
MP24425	MHP7448_0425	-	hypothetical_protein
MP24571	MHP7448_0434	-	hypothetical_protein
MP24393	MHP7448_0482	-	hypothetical_protein
MP10183	MHP7448_0491	-	hypothetical_protein
MP24586	MHP7448_0498	-	hypothetical_protein
MP05009	MHP7448_0503	oppC	oligopeptide_transport_system_permease_protein
MP12598	MHP7448_0510	-	hypothetical_protein
MP24483	MHP7448_0516	-	hypothetical_protein
MP24709	MHP7448_0518	-	hypothetical_protein
MP24377	MHP7448_0519	-	hypothetical_protein
MP12765	MHP7448_0543	-	hypothetical_protein
MP05709	MHP7448_0548	mtlF	PTS_system_mannitol-specific_component_IIA
MP24376	MHP7448_0549	mtlD	mannitol-1-phosphate_5-dehyrogenase
MP24391	MHP7448_0551	-	hypothetical_protein
MP10287	MHP7448_0553	-	PTS_system_galactitol-specific_enzyme_IIB_component
MP24588	MHP7448_0555	-	hypothetical_protein
MP24618	MHP7448_0556	-	hypothetical_protein
MP24517	MHP7448_0566	-	hypothetical_protein
MP24368	MHP7448_0570	-	hypothetical_protein
MP04047	MHP7448_0579	-	hypothetical_protein
MP04039	MHP7448_0580	mnuA	membrane_nuclease,_lipoprotein
MP10319	MHP7448_0585	-	hypothetical_protein
MP24468	MHP7448_0594	-	hypothetical_protein
MP12705	MHP7448_0596	-	hypothetical_protein
MP24664	MHP7448_0624	-	ABC_transporter_ATP-binding_protein_-_Pr2
MP24472	MHP7448_0627	-	ABC_transporter_ATP-binding_-_Pr1-like_protein
MP15012	MHP7448_0628	-	ABC_transporter_ATP-binding_-_Pr2-like_protein
MP19195	MHP7448_0639	-	hypothetical_protein
MP02029	MHP7448_0664	-	ABC_transporter_ATP-binding_-_Pr1

Supplemental Table S5. *M. hyorhinis* genes no mapping

Gene ID	Product
hsdS	Type I site-specific DNA methyltransferase specificity subunit
MHR_0038	hypothetical protein
MHR_0039	hypothetical protein
MHR_0080	hypothetical protein
MHR_0094	hypothetical protein
MHR_0127	hypothetical protein
MHR_0136	hypothetical protein
MHR_0160	hypothetical protein
MHR_0161	hypothetical protein
MHR_0168	hypothetical protein
MHR_0191	hypothetical protein
MHR_0192	hypothetical protein
MHR_0200	Restriction endonuclease S subunits
MHR_0237	Expressed protein
MHR_0244	hypothetical protein
MHR_0307	hypothetical protein
MHR_0331	hypothetical protein
MHR_0345	hypothetical protein
MHR_0347	hypothetical protein
MHR_0404	hypothetical protein
MHR_0405	hypothetical protein
MHR_0470	hypothetical protein
MHR_0474	hypothetical protein
MHR_0476	hypothetical protein
MHR_0505	hypothetical protein
MHR_0606	hypothetical protein
MHR_0619	hypothetical protein
MHR_0651	ABC transporter ATP-binding-Pr1
mtlD	Mannitol-1-phosphate 5-dehyrogenase
mtlF	Mannitol-specific phosphotransferase enzyme IIA component

Supplemental Table S6. *M. flocculare* genes no mapping

Gene ID	Gene name	Product
MF00007	(rnc)	ribonuclease III
MF00026	-	hypothetical protein
MF00090	(ruvA)	holliday junction DNA helicase RuvA
MF00092	-	hypothetical protein
MF00175	(rpe)	ribulose-phosphate 3-epimerase
MF00187	-	hypothetical protein
MF00194	-	hypothetical protein
MF00270	-	hypothetical protein
MF00292	(mtlD)	mannitol-1-phosphate 5-dehyrogenase
MF00297	-	hypothetical protein
MF00303	-	hypothetical protein
MF00381	(tdk)	thymidine kinase
MF00432	(pdhD)	dihydrolipoamide dehydrogenase
MF00433	-	hypothetical protein
MF00454	-	hypothetical protein
MF00483	-	hypothetical protein
MF00500	-	hypothetical protein
MF00552	-	hypothetical protein
MF00553	(potC)	spermidine/putrescine ABC transporter permease protein
MF00556	(potB)	spermidine/putrescine ABC transporter permease protein
MF00655	-	hypothetical protein
MF00676	-	hypothetical protein
MF00711	-	hypothetical protein
MF00714	-	hypothetical protein
MF00762	-	ABC transporter permease protein
MF00766	-	proline dipeptidase
MF00849		mannose-6-phosphate isomerase
MF00920	-	hypothetical protein
MF00922	(cysS)	cysteinyl-tRNA synthetase
MF00924	-	tRNA/rRNA methyltransferase
MF00925	-	hypothetical protein
MF00951	-	ABC transporter ATP-binding - Pr2-like protein
MF00952	-	ABC transporter ATP-binding - Pr1-like protein
MF00972	-	hypothetical protein
MF01003	(valS)	valyl-tRNA synthetase
MF01076	-	hypothetical protein
MF01086	-	hypothetical protein
MF01113	(recA)	recombination protein RecA
MF01224	(glpF) 1	glycerol uptake facilitator protein
MF01252	-	hypothetical protein
MF01254	-	hypothetical protein
MF01255	-	hypothetical protein
MF01261	-	type III restriction-modification system: methylase
MF01275	-	hypothetical protein
MF01327	-	hypothetical protein
MF01328	(rpsI)	30S ribosomal protein S9

Continuação *Supplemental Table S6*

Gene ID	Gene name	Product
MF01344	(rpsM)	30S ribosomal protein S13
MF01365	-	hypothetical protein
MF01373	-	hypothetical protein
MF01374	(cdd)	cytidine deaminase
MF01378	-	hypothetical protein
MF01379	-	hypothetical protein
MF01397	MHP7448_0553	PTS system galactitol-specific enzyme IIB component
MF01404	-	hypothetical protein
MF01407	-	hypothetical protein
MF01409	-	hypothetical protein
MF01426	-	hypothetical protein
MF01429	-	hypothetical protein
MF01447	-	hypothetical protein
MF01448	-	hypothetical protein
MF01454	-	hypothetical protein
MF01455	(mtlF)	PTS system mannitol-specific component IIA
MF01458	-	hypothetical protein
MF01462	-	hypothetical protein
MF01464	(atpE)	F0F1 ATP synthase subunit C
MF01466	-	hypothetical protein
MF01473	(rpmG)	50S ribosomal protein L33
MF01474	(rpmJ)	50S ribosomal protein L36