



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR



CAMPUS I - CIDADE UNIVERSITÁRIA
CEP: 58051-900 - JOÃO PESSOA - PB
TELEFONE: (83) 3216-7407
E-mail: pgbcm@dbm.ufpb.br

JOANA MARIA KÄSTLE SILVA

**REDES DE INTERAÇÃO PROTEÍNA-PROTEÍNA EM *UREAPLASMA*
DIVERSUM: EVIDÊNCIAS DE TRANSFERÊNCIA HORIZONTAL DE GENES
E DA EVOLUÇÃO DOS GENOMAS REDUZIDOS NA CLASSE MOLLICUTES**

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E
MOLECULAR**

João Pessoa – PB

2018



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR



CAMPUS I - CIDADE UNIVERSITÁRIA
CEP: 58051-900 - JOÃO PESSOA - PB
TELEFONE: (83) 3216-7407
E-mail: pgbcm@dbm.ufpb.br

JOANA MARIA KÄSTLE SILVA

**REDES DE INTERAÇÃO PROTEÍNA-PROTEÍNA EM *UREAPLASMA*
DIVERSUM: EVIDÊNCIAS DE TRANSFERÊNCIA HORIZONTAL DE GENES
E DA EVOLUÇÃO DOS GENOMAS REDUZIDOS NA CLASSE MOLLICUTES**

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Ciências Exatas e da Natureza, da Universidade Federal da Paraíba, como parte dos requisitos para obtenção do título de MESTRE EM BIOLOGIA CELULAR E MOLECULAR.

Orientador/a: Prof. Dr. José Pinto de Siqueira Junior

Co-orientador/a: Prof. Dr. Sávio Torres de Farias

João Pessoa - PB

2018



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR



CAMPUS I - CIDADE UNIVERSITÁRIA
CEP: 58051-900 - JOÃO PESSOA - PB
TELEFONE: (83) 3216-7407
E-mail: pgbcm@dbm.ufpb.br

JOANA MARIA KÄSTLE SILVA

Dissertação de Mestrado avaliada em 19/04/2018

BANCA EXAMINADORA

Prof. Dr. José Pinto de Siqueira Junior - Programa de Pós-Graduação em Biologia
Celular e Molecular – Universidade Federal da Paraíba

Orientador

Prof. Dr. Sávio Torres de Farias – Universidade Federal da Paraíba

Orientador

Profa. Dra. Louisa Maria de Andrade e Sousa – Universidade Federal da Paraíba

Examinadora Externa

Prof. Dr. Ulrich Vasconcelos da Rocha Gomes – Universidade Federal da Paraíba

Examinador Interno

“Nada na vida deve ser temido, somente compreendido.
Agora é hora de compreender mais para temer menos.”
Marie Curie

AGRADECIMENTOS

A minha família, meus amados pais Alexandre e Beate pelo suporte e por serem minha fundação e porto seguro, além dos incentivadores e torcida mais leal que se poderia sonhar. Aos meus irmãos mais novos David e Milena, por manterem meus pés no chão e pelo carinho disfarçado em provocações amigáveis.

Aos meus amigos de vida, curso e família por extensão, sem vocês este trabalho perderia um tanto de cor. Em especial a Moema França, Gabrielle Rodrigues, Ketyucia Yamashita, Luana Ayres, Ludmila Magroski, Leonardo dos Santos, Valberta Cabral, Paula Giovanna, Talita Oliveira, Lucas Cavalcanti, Everton Lorenzo, Rayssa Moraes, Wilka Ferreira, João Lucas Mageste e Louise Egito. Obrigada por manterem minha mente saudável e pelas distrações necessárias entre análises.

Ao Prof. Dr. José Pinto de Siqueira Junior, por aceitar me orientar e pelo constante apoio.

Ao Prof. Dr. Sávio Torres de Farias, pela orientação e confiança ao longo de todo este trabalho. Obrigada por me fazer sair de cada reunião com ideias fervilhando na cabeça.

Ao Prof. Dr. Lucas de Miranda Marques e sua equipe na UFBA, obrigada pela disponibilidade da anotação genômica, sem a qual este trabalho não seria possível.

Aos meus amigos do Laboratório de Genética Evolutiva Paulo Leminski pelas conversas, conselhos e ajuda na utilização dos programas. Em especial a Vanessa Pontes, Camila Eduarda, Victor Montenegro e Angélica Cardoso.

Aos meus professores da graduação e pós-graduação no Programa de Biologia Celular e Molecular, que plantaram em mim a curiosidade por genética bacteriana e a alimentaram através de discussões e debates por todos estes anos.

Aos membros examinadores da banca por aceitarem o convite para avaliação do meu trabalho, obrigada por me darem parte de seu tempo.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa.

A todos que contribuíram para a realização deste trabalho direto ou indiretamente.

RESUMO

KÄSTLE-SILVA, Joana Maria. **Redes de interação proteína-proteína em *Ureaplasma diversum*: evidências de transferência horizontal de genes e da evolução dos genomas reduzidos na classe Mollicutes**. 2018. Dissertação (Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa – PB, 2018).

Proteínas não realizam suas funções biológicas sozinhas e estão, em vez disto, integradas em redes de interação proteína-proteína. Estas redes formam o alicerce de diferentes vias metabólicas e processos celulares, garantindo a sobrevivência, proliferação e colonização de novos nichos. *Ureaplasma diversum* é um integrante da classe Mollicutes causador de infecções no sistema reprodutor de gado e caprinos. Como os demais Mollicutes, *U. diversum* não possui parede celular e é caracterizado pelo genoma reduzido e vias metabólicas incompletas ou totalmente ausentes. Estudos apontam que o processo de transferência horizontal de genes, a troca de material genético entre indivíduos não descendentes, teve um importante papel na evolução dos Mollicutes afetando os processos de redução genômica e simplificação de vias metabólicas. Construímos a rede IPP de *U. diversum* e a comparamos com as redes de outros representantes da classe Mollicutes e de grupos externos, buscando entender se a redução de complexidade sofrida nesta classe se reflete em suas redes de interação. Também investigamos eventos de transferência horizontal de genes em subredes de interesse envolvidas nos processos de metabolismo de purinas e pirimidinas (vias incompletas em Ureaplasmas e Mycoplasmas) e no funcionamento da urease que supre a demanda energética de Ureaplasmas e age também como fator de virulência. Identificamos transferência horizontal de genes ocorrendo dentro da classe Mollicutes, entre Ureaplasmas e *S. aureus* e *Corynebacterium*, bactérias que ocupam o mesmo nicho urogenital. A tendência de redução e simplificação genômica dos Mollicutes se reflete em suas redes IPP mais generalistas e menos “seletoras”. Isto sugere que o processo foi permitido (ou possibilitado) através de um aumento da dependência do metabolismo do hospedeiro e pelo repertório genético disponível no nicho urogenital via transferência horizontal de genes.

Palavras-chave: transferência horizontal de genes, redes de interação, Mollicutes, *Ureaplasma*.

ABSTRACT

KÄSTLE-SILVA, Joana Maria. **Protein-Protein Interaction Networks in *Ureaplasma diversum*: evidences of horizontal gene transfer and evolution of reduced genomes among the Mollicutes**. 2018. Dissertation (Center of Exact and Nature Sciences, Federal University of Paraíba, João Pessoa – PB, 2018).

Proteins do not perform their biological functions alone and are instead connected in protein-protein interaction networks. These networks provide the foundation for different metabolic pathways and cellular processes that guarantee cell survival and proliferation, as well as the colonization of novel niches. *Ureaplasma diversum* is a member of the Mollicute class and causes urogenital tract infections in cattle and small ruminants. As it is the rule among Mollicutes, *U. diversum* does not have a cellular wall and is characterized by a reduced genome and incomplete or completely lacking metabolic pathways. Studies indicate that the process of horizontal gene transfer, the exchange of genetic material among non-descendant individuals, has a crucial role in Mollicute evolution, affecting their genomic reduction process and the simplification of metabolic pathways. We constructed the protein-protein interaction network of *U. diversum* and compared it with the networks of other members of the Mollicute class and external groups to test if the reduction of complexity in this class is reflected in their protein interaction networks. We also investigated horizontal gene transfer events in sub-networks of interest involved in purine and pyrimidine metabolism (incomplete pathways in Ureaplasmas and Mycoplasmas) and urease function, an enzyme that supplies the energy demand in *Ureaplasma* and acts as a virulence factor. We identified horizontal gene transfer events among Mollicutes and from *Ureaplasma* to *S. aureus* and *Corynebacterium*, bacterial groups that colonize the urogenital niche. The overall tendency of genome reduction and simplification in the Mollicutes is reflected in their protein interaction networks, tending to be more generalistic and less “selective”. This suggests that the process was permitted (or enabled) by an increase in host dependence and the available gene repertoire in the urogenital tract via horizontal gene transfer.

Keywords: protein interaction network, horizontal gene transfer, *Ureaplasma*, Mollicutes.

LISTA DE ABREVIATURAS

DNA – Ácido desoxirribonucleico

RNA – Ácido Ribonucleico

IPP – Interação Proteína-Proteína

NCBI – *National Cener of Biotechnology Information*

Unirpot – *Universal Protein Resource*

CDS – *Coding DNA sequence*

NDK – *Nucleoside-diphosphate kinase* (nucleosídeo difosfato quinases)

AMP – Adenosina monofosfato

ADP – Adenosina difosfato

ATP – Adenosina trifosfato

LISTA DE FIGURAS

Figura 1. Proteínas conhecidas da anotação genômica de <i>U. diversum</i> classificadas de acordo com a sua função biológica.....	8
Figura 2. Principais meios de transferência horizontal de genes entre bactérias.....	11
Figura 3. O grau (conectividade) k do nodo.....	20
Figura 4. Coeficiente de agrupamento de A , C_A , e seu cálculo.....	21
Figura 5. Coeficiente médio de agrupamento de um gráfico não direcionado.....	21
Figura 6. Cálculo da via mais curta entre dois nodos.....	22
Figura 7. Cálculo do diâmetro de uma rede IPP.....	22
Figura 8. Cálculo da interseção de um nodo em uma rede IPP.....	22
Figura 9. Visão geral da rede IPP de <i>U. diversum</i>	26
Figura 10. Conectividade (k) dos nodos da rede IPP de <i>U. diversum</i>	27
Figura 11. Distribuição da centralidade dos nodos, coeficiente médio de agrupamento e tamanho característico de via na rede IPP de <i>U. diversum</i>	28
Figura 12. A subrede do metabolismo de Purinas e Pirimidinas em <i>U. diversum</i>	34
Figura 13. Subrede de <i>deoD</i> em <i>U. diversum</i>	36
Figura 14. Subrede <i>adk</i> em <i>U. diversum</i>	37
Figura 15. A subrede da urease em <i>U. diversum</i>	38
Figura 16. Subrede <i>recA</i> e <i>dnaK</i> , as maiores proteínas hubs da rede IPP de <i>U. diversum</i>	42
Figura 17. Subrede de <i>hlyA</i> (hemolisina) em <i>U. diversum</i>	45
Figura 18. Eventos de transferência horizontal de genes na subrede da Urease.....	49
Figura 19. Eventos de transferência horizontal de genes na subrede de purinas.....	50
Figura 20. Eventos de transferência horizontal de genes na subrede de pirimidinas.....	51
Figura 21. Interseção das redes de urease de todas as espécies analisadas para transferência horizontal.....	52
Figura 22. Subredes da urease em <i>U. diversum</i> , <i>U. urealyticum</i> e <i>U. parvum</i>	53

Figura 23. Repertório proteico da urease em <i>Ureaplasma</i> , construído da junção das redes individuais de <i>U. diversum</i> , <i>U. parvum</i> e <i>U. urealyticum</i>	53
Figura 24. Subrede da Urease em <i>C. glucurolyticum</i> e <i>C. urealyticum</i>	54
Figura 25. Repertório proteico de <i>C. urealyticum</i> e <i>C. glucurolyticum</i> , referindo-se a subrede da urease.....	54
Figura 26. Subrede urease em <i>U. diversum</i> e <i>S. aureus</i>	54
Figura 27. Subrede do metabolismo de purinas em <i>M. genitalium</i> , <i>M. pneumoniae</i> e <i>M. hominis</i>	56
Figura 28. Redes do metabolismo de purinas em <i>U. diversum</i> , <i>U. urealyticum</i> e <i>U. parvum</i>	57
Figura 29. Interseção do metabolismo de purinas em <i>Mycoplasma</i> e <i>Ureaplasma</i>	58
Figura 30. Redes IPP envolvidas no metabolismo de pirimidinas em <i>M. hominis</i> , <i>M. genitalium</i> e <i>M. pneumoniae</i>	59
Figura 31. Redes IPP envolvidas no metabolismo de pirimidinas em <i>U. diversum</i> , <i>U. urealyticum</i> e <i>U. parvum</i>	59
Figura 32. Interseção de redes IPP do metabolismo de pirimidinas de <i>Mycoplasma</i> e <i>Ureaplasma</i>	60
Figura 33. Subrede IPP do metabolismo de pirimidinas em <i>C. urealyticum</i>	61
Figura 34. Interseção das redes IPP do metabolismo de pirimidinas em <i>Mycoplasma</i> e <i>Corynebacterium</i>	61

LISTA DE TABELAS

Tabela 1. Parâmetros da rede IPP de <i>U. diversum</i> e as 10 redes randômicas geradas para validação.....	30
Tabela 2. Principais parâmetros topológicos das redes IPP das espécies analisadas.....	31
Tabela 3. Parâmetros de centralidade de proteínas hubs da rede IPP de <i>U. diversum</i>	39
Tabela 4. Parâmetros das redes IPP da urease das espécies investigadas.....	55
Tabela 5. Parâmetros das redes IPP do metabolismo de purinas das espécies investigadas.....	58
Tabela 6. Os parâmetros da subrede do metabolismo de pirimidinas dos grupos investigados.....	62

SUMÁRIO

1.INTRODUÇÃO.....	1
1.1 Proteínas e suas funções biológicas.....	1
1.2 Genomas Mínimos.....	2
1.3 A Classe Mollicutes.....	6
1.4 Panorama do metabolismo reduzido de <i>U. diversum</i>	7
1.5 Transferência Horizontal de Genes.....	10
1.6 Redes de Interação Proteína-Proteína.....	15
2. OBJETIVOS.....	17
3. MATERIAIS E MÉTODOS.....	18
3.1 Construção das redes IPP de <i>Ureaplasma diversum</i>	18
3.1.1 Parâmetros topológicos de redes IPPs.....	19
3.1.2 Validação das redes IPP de <i>Ureaplasma diversum</i>	23
3.2 Construção das redes IPP de espécies usadas para comparação de complexidade de redes.....	23
3.3. Comparação das redes IPP de <i>U. diversum</i> e espécies de referência.....	24
3.4. Reconstrução da filogenia da classe Mollicutes de acordo com a complexidade de subredes e interação observadas nas IPP.....	24
4. RESULTADOS E DISCUSSÃO.....	25
4.1 Características e parâmetros gerais da rede IPP de <i>U. diversum</i> e sua validação....	25
4.2 Comparação da rede global de <i>U. diversum</i> com espécies relacionadas.....	30
4.3 Subredes de interesse para o metabolismo de <i>U. diversum</i>	32
4.4 Proteínas centrais da rede IPP de <i>U. diversum</i>	38
4.5 Transferência horizontal de genes em <i>Ureaplasma</i>	43

5. CONSIDERAÇÕES FINAIS.....	62
6. REFERÊNCIAS.....	63
APÊNDICES.....	79
APÊNDICE A – Redes Globais das Espécies para comparação.....	79
APÊNDICE B – Redes randômicas geradas para validação.....	83
APÊNDICE C – Tabela de nodos <i>U. diversum</i> e seus parâmetros correlacionados.....	87
APÊNDICE D – Árvores 16s controle na análise de transferência horizontal de genes....	91

1. INTRODUÇÃO

1.1 Proteínas e suas funções biológicas

Proteínas constituem os organismos estrutural e quimicamente, desempenhando funções essenciais para sobrevivência e manutenção da vida celular. Seja compondo macromoléculas com os mais diversos papéis constitutivos, conferindo especificidade a respostas imunes por meio de anticorpos ou catalisando reações e otimizando o tempo e energia para a maquinaria metabólica celular, aminoácidos, peptídeos e proteínas são responsáveis por uma ampla gama de funções biológicas, sendo referidas muitas vezes como as moléculas cruciais para sustentação da vida. Elas são, ainda, o produto final da expressão gênica e as ferramentas pelas quais a informação contida no material genético é transformada em uma maquinaria funcional. As proteínas que compõem todos os organismos, não importa quão antigos ou recentes, são compostos pelos mesmos 20 aminoácidos que são arranjados em sequências lineares e posteriormente modificados em estrutura para exercerem suas aplicações orgânicas.

Para bactérias estas funções podem ser consideradas ainda mais críticas, uma vez que a unicelularidade impõe uma necessidade de ajustes e uma resposta celular imediata a pistas e mudanças ambientais para garantir a subsistência do indivíduo e em uma escala maior, da espécie. Nestes organismos unicelulares, proteínas de superfície estão envolvidas na adesão a células e tecidos do hospedeiro durante o desencadeamento de infecções (NAVARRE; SCHNEEWIND, 1999), tornando-as potenciais alvos terapêuticos (CLATWORTHY et al., 2007). Semelhante aos processos de sinalização e transdução celular de eucariotos, bactérias são capazes de comunicar-se por meio de vias comandadas por proteínas específicas, como as da família LuxR-LUXI (FUQUA et al., 1994). A comunicação celular promovida por estas proteínas específicas auxilia as bactérias a orquestrar o comportamento de virulência (MILLER; BASSLER, 2001) garantindo a evasão do sistema imune do hospedeiro e a controlar a densidade populacional diante de pistas ambientais de estresse (WATERS; BASSLER, 2005).

Responder rapidamente a uma situação de estresse pode significar a diferença entre vida e morte para células bacterianas (STOCK et al., 1989), uma vez que estes organismos não estão protegidos por tecidos e órgãos, interagindo diretamente com o ambiente e seus agentes nocivos. Proteínas são responsáveis por regular estas respostas, em situações de estresse oxidativo (CABISCOL; CATALÀ et al., 2000) ou osmótico (LESLIE et al.,

1995). Da mesma maneira que em eucariotos, proteínas são o alicerce e os catalisadores de vias metabólicas bacterianas (DEUTSCHER et al., 2006) (NOTHAFT; SZYMANSKI, 2010), desempenhando também funções organizacionais do material genético de forma semelhante ao que histonas realizam no DNA de eucariotos (DRLICA; ROUVIERE-YANIV, 1987).

Com funções que vão desde a manutenção de estruturas biológicas até o poder de catálise, o estudo de proteínas se firmou como uma das principais áreas da biologia, refletindo ao longo dos séculos o avanço de métodos e abordagens e acompanhando o desenvolvimento de novas tecnologias que permitiram o seu estudo em diferentes escalas ao longo do fluxo da informação gênica. Nas últimas décadas o avanço do poder computacional se alinhou aos fundamentos biológicos alcançados via estudos experimentais dando início a empolgante era da genômica e proteômica, agora subdivididas em partes distintas em constante comunicação. O desenvolvimento dessas tecnologias e do poderio informático para sustentar sua análise deu início a estudos realizados em uma escala global, abordando funções, estrutura e regulação proteica como atributos acessíveis aos pesquisadores e pesquisadoras. Como utilizar ferramentas conhecidas pela biologia molecular aliando-as a abordagens de bioinformática vem reforçando a ideia de que o surgimento de tecnologias pode andar em conjunto com a utilização e potencialização de métodos clássicos para fins inéditos (SILVA et al., 2013).

Proteínas, com todas suas funções orgânicas correlacionadas, são macromoléculas indispensáveis para vida celular como a conhecemos. O que mais é necessário para sustentação da vida vem sendo alvo de investigação há centenas (ou milhares) de anos, mas focando-se em uma abordagem orgânica cientistas favoreceram uma linha de pensamento bastante objetiva: para se entender a complexidade da vida é necessário partir de sua forma mais simples. Como os químicos fizeram com o átomo de hidrogênio, tornando-o modelo para estudar conformações atômicas mais intrincadas, biólogos e biólogas encontraram no genoma mínimo o ponto de partida para compreensão da conformação organizacional da informação gênica e como esta é expressa nos diferentes organismos.

1.2 Genomas Mínimos

O que é necessário para sustentar a vida? Do ponto de vista biológico, cientistas vêm tentando alcançar esta resposta (ou respostas) há séculos e o horizonte tornou-se

bastante promissor nos últimos anos. Tomando como base os requisitos mínimos que uma célula precisa atingir para manter-se viva. Estudos separando o joio do trigo (ou genes essenciais de genes redundantes) se multiplicaram no frenesi da biologia sintética que tem entre seus objetivos uma marca bastante ambiciosa: entender a maquinaria celular e a orquestra genética que a rege para permitir o desenvolvimento de uma célula artificial. Com uma meta tão audaciosa, estudos de base foram necessários antes de se debater a construção de uma célula sintética; não é possível levantar uma casa sem conhecer seus tijolos.

Em 1990, a discussão para definir o conteúdo genômico mínimo para suportar a vida voltou seu foco para o nível celular. Quais genes são cruciais e quais não significam a morte da célula quando inativos, seja por deriva gênica ou mutação? A busca pelo genoma mínimo já tinha se tornado uma caçada pelo menor organismo autorreplicante de vida livre nos anos 1950 (MOROWITZ, 1984).

Esta busca encontrou seu tesouro nos Micoplasmas, do grego μυκήs, *mykes* (fungo) e πλάσμα, *plasma* (formado) datado ao primeiro uso do nome por Albert Bernahard Frank em 1889 (KRASS; GARDNER, 1973), referindo-se a infiltrações promovidas por microrganismos semelhantes a fungos no citoplasma de plantas. Na década de 1930, Julian Nowak sugeriu o nome *Mycoplasma* para o gênero que englobava microrganismos filamentosos que, acreditava-se, possuíam estágio celulares e acelulares em seu ciclo de vida, o que era apontado como o motivo de serem vistos no microscópio, mas conseguirem passar por filtros impenetráveis para outras bactérias (BROWNING; CITTI, 2014). Hoje, sabe-se que esta propriedade peculiar da taxa se deve à ausência de parede celular. Mas não foi esta característica que tornou os Micoplasmas um dos grupos de maior atenção para biologia sintética: se trata, afinal, do grupo de menores organismos autorreplicantes de vida livre conhecidos (PITCHER, 1993) e, mais relevante ainda, portadores dos menores genomas (COLE; GIRON, 1994).

Fraser et al. (1995) foram responsáveis pelo sequenciamento completo do genoma de *Mycoplasma genitalium*, o primeiro *Mycoplasma* a ter sua sequência genômica totalmente conhecida. Com pouco mais de 580 quilobases, *M. genitalium* possui apenas 470 regiões codificadoras que se concentram principalmente na expressão da informação gênica, transporte e metabolismo de energia (FRASER et al., 1995). No ano seguinte, estudos comparativos (MUSHEGIAN; KOONIN, 1996) com o genoma de *Haemophilus*

influenzae (FLEISCHMANN et al., 1995) deram início a corrida pela determinação do genoma mínimo necessário para sustentar vida.

A determinação do *set* mínimo de genes para vida celular não é apenas útil para a biologia sintética nos esforços de criação de células artificiais, mas também de importância crucial quando busca-se compreender como uma célula ancestral primordial conseguia sobreviver e qual maquinaria metabólica estava disponível para ela. Mushegian e Koonin (1996) em análises comparativas determinaram que o *set* genômico comum entre *Mycoplasma*, que derivou de um ancestral Gram-positivo, e *H. influenzae*, descendente de uma linhagem Gram-negativa, poderia ser um bom ponto de partida para o genoma mínimo. Bactérias Gram-positivas e gram-negativas divergiram há pelo menos 3.200 milhões de anos (GIBBS et al., 2005), tornando *M. genitalium* e *H. influenzae* alvos promissores para o estudo de genomas conservados. Mushegian e Koonin (1996) reduziram ainda mais o genoma mínimo eliminando genes redundantes e genes específicos relacionados aos respectivos hospedeiros das duas espécies.

O resultado final é um conjunto de 250 genes que corresponde a aproximadamente 315kb. Este genoma mínimo deve se aproximar do *set* genômico apresentado pelo último ancestral comum entre os domínios *Bacteria*, *Archaea* e *Eukarya* (MUSHEGIAN; KOONIN, 1996). Em seu artigo de revisão, Maniloff (1996) aponta a importância de dividir o evento dos genomas mínimos como ocorrendo através de duas vias evolutivas diferentes, uma via "*top down*" e outra "*bottom up*". De forma resumida, genomas mínimos "*top down*" são resultantes de organismos recentes com requerimentos metabólicos elevados. Maniloff (1996) discute que para cada nível metabólico existe um genoma mínimo, desde organismos fototróficos até seres heterotróficos simples e fastidiosos, incluindo parasitas intracelulares obrigatórios, organelas que perderam quase todo seus genes bacterianos para se tornarem organelas citoplasmáticas eucarióticas, etc.

Já genomas "*bottom up*" se referem aos conteúdos genéticos de "organismos" que surgiram ao longo da história evolutiva das células (MANILOFF, 1996). Genomas mínimos "*top down*" surgiram de perda gênica (ou deriva genética), ou seja, organismos recentes e especializados em determinados nichos ecológicos foram perdendo genes de vias metabólicas complexas, reguladas e integradas (ANDERSSON; KURLAND, 1995) enquanto genomas mínimos de organismos "*bottom up*" são resultado de um acréscimo

gradual de genes ao longo de sua história evolutiva através do recrutamento de minigenes não regulados em vias metabólicas primitivas (JENSEN, 1976).

Teorias diferentes buscam explicar eventos de redução genômica em microrganismos e Maniloff (1996) resume que ao longo da história evolutiva, a seleção foi positiva para: a) genomas de replicação mais rápida (em outras palavras, genomas menores) originando a produção de proles maiores b) genomas menores que tem um requerimento energético mais baixo para crescimento em ambientes com disponibilidade limitante de nutrientes e c) a perda de sequências genômicas com mutações deletérias (DYBVIG; VOELKER, 1996). Qualquer um desses fatores (ou todos eles) podem ter afetado a redução genômica observada nos *Mycoplasmas* e *Mollicutes* de modo geral.

Glass et al. (2017) define *Micoplasmas* como organismos "quase-mínimos" e pontua que suas características metabólicas e estilo de vida refletem (ou foi moldado) pela redução genômica sofrida pelo grupo ao longo de sua evolução. O estilo de vida obrigatoriamente parasitário e vias metabólicas simplificadas dos *Micoplasmas* não são produto de um organismo ancestral que se situa na base da árvore da vida (GLASS et al., 2017), em vez disto, *Mycoplasma* teve seu genoma reduzido pelo processo *top-down* discutido anteriormente. Acredita-se que se trata de bactérias derivadas da classe Firmicutes (mais precisamente, *S. aureus* ou *B. subtilis*) que passaram pelo processo de redução genômica adaptando-se a um nicho muito específico dependente de hospedeiro e que se mantém muito estável e rico em nutrientes. Na realidade, *Micoplasmas* e seus processos metabólicos e de expressão gênica se aproximam daqueles propostos para células artificiais: transcrição, tradução e replicação celular são permitidas via importação de precursores de DNA e RNA do hospedeiro e poucas vias metabólicas são observadas além destas (WOESE et al., 1985), mostrando que sua utilização como modelos para genomas mínimos foi uma escolha eficiente.

Os avanços no estudo de genomas mínimos foram possíveis pelo refinamento de ferramentas de bioinformática e biologia molecular, que permitiram novas perspectivas sobre o set mínimo de genes necessários para sustentar a vida e suas funções correlacionadas. Estudos também abordaram os desafios na construção de células artificiais para produção de metabólitos de interesse, que costumam se chocar na problemática de inserir precursores metabólitos na célula sintética e como estes podem ser corretamente assimilados (JUHAS et al., 2014).

Um desafio semelhante foi enfrentado pelos Mollicutes ao longo do seu processo de redução genômica que levou ao surgimento de um set de genes mínimo natural. Como este evento se deu e suas consequências para o metabolismo simplificado e estilo de vida dependente do hospedeiro deste grupo será tratado no tópico a seguir.

1.3 A Classe Mollicutes

Mycoplasmas compõem o grupo dos menores seres autorreplicantes de vida livre e, como os demais Mollicutes, não possuem parede celular. São comensais e parasitas de vertebrados, artrópodes e plantas (WAITES et al., 2013), colonizando nos humanos a mucosa bucal, trato respiratório e genital (KOKKAYIL; DHAWAN, 2015). Conhecida por incluir alguns dos genomas mais reduzidos entre as bactérias, a classe Mollicutes tem como característica vias metabólicas simplificadas (SIRAND-PUGNET; CITTI; et al., 2007) que muitas vezes contam com metabólitos adquiridos diretamente do hospedeiro. O metabolismo do gênero é fastidioso, necessitando de muitos nutrientes e metabólitos do ambiente, o que torna-o exigente para ser cultivado em laboratório. Estudos com 16S (WOESE et al., 1985), (WEISBURG et al., 1989) determinaram que a classe evoluiu de uma bactéria Gram-positiva com baixo conteúdo G+C por meio de um processo de redução genômica. Os genes "perdidos" ao longo da evolução da classe envolvem aqueles responsáveis pela biossíntese de aminoácidos, cofatores e ácidos graxos (RAZIN, 1985) (MANILOFF, 1996), que podem estar ausentes em sua totalidade dependendo do gênero estudado.

Uma importante via incompleta nos gêneros *Ureaplasma* e *Mycoplasma* é a comprometida com a de síntese *de novo* de purinas e pirimidinas (BIZARRO; SCHUCK, 2007; McELWAIN et al., 1988). Síntese *de novo* se refere a biossíntese de metabólitos sem a utilização de precursores, em que a maquinaria metabólica celular consegue realizá-la de forma independente. Paralelamente, estudos *in vitro* indicam que o grupo consegue produzir certas enzimas e proteínas mesmo sem possuir os genes correspondentes conhecidos para tal em seu genoma (MARQUES et al., 2016). Como *Ureaplasma* utiliza vias alternativas em seu metabolismo e supre a demanda de proteínas cujos genes conhecidos não estão presentes em seu genoma continua sendo uma questão em aberto.

Ureaplasma é considerado um patógeno oportunista, permanecendo colonizando as regiões mucosas e invadindo a submucosa, o tecido conjuntivo, em situações de imunossupressão ou dano físico ou bacteriano preexistente (KOKKAYIL; DHAWAN,

2015). A infecção por *Ureaplasma* em humanos foi ligada ao desenvolvimento de vaginite bacteriana, doenças inflamatórias na pelve e uretrite (CASSELL et al., 1993), além de complicações na gravidez incluindo partos prematuros (DIGIULIO, 2012) e abortos (TAYLOR-ROBINSON, 1996). Nos recém-nascidos prematuros com peso baixo, *Ureaplasma* pode causar problemas pulmonares (CASSELL et al., 2012), aumentando os riscos de desenvolvimento de displasia brônquio-pulmonar e doenças crônicas nos pulmões (SETHI et al., 1999). Urease está entre os fatores de virulência do gênero *Ureaplasma*, além de estar envolvida na hidrólise de ureia para suprir o requerimento energético (LIGON; KENNY, 1991) do metabolismo bacteriano. A hidrólise de ureia gera amônia, que por sua vez é responsável pelo dano tecidual causado ao hospedeiro por mudanças no pH (KOKKAYIL; DHAWAN, 2015).

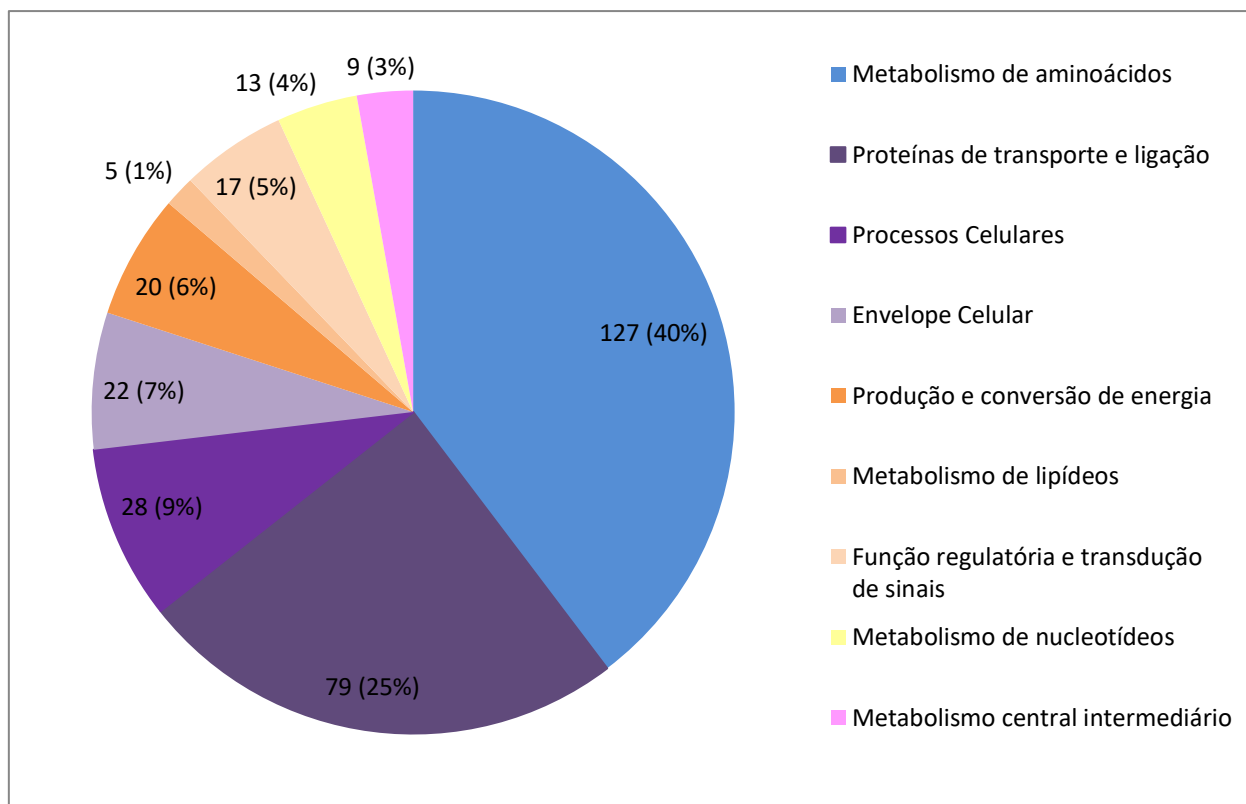
Assim como diversas espécies de *Mycoplasma*, *Ureaplasma diversum* é um importante patógeno do trato reprodutor de gado, afetando principalmente fêmeas jovens (CARDOSO et al., 2000). Entre os sintomas da infecção estão a vulvovaginite granular, salpingite, endometrite, aborto e morte neonatal, além de infertilidade e vesiculite seminal (DOIG et al., 1979) (MILLER et al., 1994) (SANDERSON et al., 2000). Estas características tornam a infecção por *U. diversum* e outras Micoplasmoses um importante fator de impacto econômico sobre a criação de gado. A prevalência de *U. diversum* em rebanhos já foi relatada na América do Sul (CARDOSO et al., 2000) (OLIVEIRA FILHO et al., 2005), Austrália (ARGUE et al., 2013), Estados Unidos (SANDERSON et al., 2000) (RAE et al., 1993) e Quênia (MULIRA et al., 1989), mostrando que os prejuízos causados por este patógeno são de escala internacional. O *cluster* da urease em *U. diversum* possui a mesma organização que nos *Ureaplasmas* colonizadores de humanos (NEYROLLES et al., 1996) e estudos indicam que as regiões codificadoras da via possuem grande similaridade com a das espécies que causam patologias em hospedeiros humanos (MARQUES et al., 2016).

1.4 Panorama do metabolismo reduzido de *U. diversum*

Quarenta por cento (127) das proteínas da anotação genômica de *U. diversum* foram correlacionadas ao metabolismo de aminoácidos e expressão gênica (MARQUES et al., 2016) que juntamente com proteínas de transporte e ligação (25%, 79 proteínas) correspondem a maior parte (65%) das proteínas anotadas desta espécie de *Ureaplasma* (**Figura 1**). Que a maioria das proteínas de *U. diversum* estejam comprometidas com as

funções de expressão gênica e transporte não é uma surpresa, uma vez que se trata de um organismo altamente dependente do hospedeiro para importação de precursores e metabólitos para suprir as suas vias metabólicas incompletas ou ausentes.

Figura 1. Proteínas conhecidas de *U. diversum* classificadas de acordo com a sua função biológica.



Fonte: modificado de MARQUES et al., 2016.

Como membro da classe Mollicutes, *U. diversum* passou por uma redução genômica ao longo da sua história evolutiva, o que é refletido em sua maquinaria celular simplificada que se aproxima da noção de genoma mínimo discutido anteriormente: se trata de um organismo capaz de replicar-se (com a maioria das proteínas comprometidas com expressão gênica), com transportadores generalistas responsáveis pela importação de precursores e vias metabólicas simplificadas e pouco além disto (GLASS et al., 2017). Mollicutes são conhecidos pelos seus genomas reduzidos e vias metabólicas simplificadas devido à remoção de grupos gênicos dos seus genomas ao longo de sua evolução degenerativa. Referindo-se a perdas de genes no grupo *Mycoplasma*, algumas espécies sofreram deriva de todos os genes envolvidos em vias específicas de biossíntese. Este é o caso de *M. pneumoniae* e *M. genitalium*. Ambos perderam genes da via de síntese de

aminoácidos, dependendo do hospedeiro ou do ambiente para suprir sua demanda destes compostos (RAZIN *et al.*, 1998). O grupo também reduziu a maioria dos genes para biossíntese de ácidos graxos e cofatores (WEISBURG *et al.*, 1989). Comparativamente, *U. diversum* também possui um metabolismo incompleto para formação de ácidos graxos e a maioria de suas proteínas estão comprometidas com expressão e regulação gênica (**Figura 1**).

Como discutido anteriormente *Ureaplasma*, assim como a maioria dos Mycoplasmas, são incapazes de produzir purinas e pirimidinas pela via de biossíntese *de novo*. Para compensar a ausência de genes metabólicos *Ureaplasma* depende do hospedeiro para importação de um grande número de precursores e até mesmo metabólitos completos. Na rede IPP de *U. diversum* 79 proteínas estão envolvidas nos processos celulares de transporte e ligação e a via de produção de purinas e pirimidinas também está incompleta (MARQUES *et al.*, 2016). Estudos demonstraram que a incorporação de nucleotídeos, nucleosídeos e bases livres pode ocorrer em muitos organismos procarióticos (PANDEY, 1984) (FLEISCHMANN *et al.*, 1995). A demanda de purinas e pirimidinas pode ser suprida por nucleotídeos intracelulares provenientes das células do hospedeiro e Razin *et al.* (1963) observaram que populações de Mollicutes são capazes de utilizar o tecido do hospedeiro e meios de cultivo para importação de precursores de nucleotídeos. Isso provavelmente também é o caso para *U. diversum*.

Marques *et al.* (2016) determinaram que *U. diversum* possui 37 sequências codificadoras identificadas relacionadas a transportadores e entre estes, sequências codificadoras relacionadas a transportadores da família ABC são a maioria (83.78%, 31/37). Muitos sistemas de transporte não foram observados em *U. diversum* incluindo transportadores de bases e nucleotídeos. Transportadores de níquel e ureia também estão ausentes (MARQUES *et al.*, 2016). O transporte destes metabólitos pode possivelmente ser realizado por outros transportadores ou por proteínas e mecanismos que ainda não são conhecidos.

Espécies de *Ureaplasma* e *Mycoplasma* não mostram um alto número de proteínas transportadoras em seu repertório genômico (PAULSEN *et al.*, 2000) e como discutido anteriormente, perderam genes ao longo de sua evolução. Essa aparente contradição pode ser explicada pela estratégia usada por estes Mollicutes durante a sua história evolutiva: transportadores na membrana celular reduziram sua especificidade de substratos e

aumentaram a sua seletividade, permitindo que poucos transportadores realizem a entrada de uma ampla gama de metabólitos (CASTELLANOS et al., 2004). Isso poderia explicar o baixo número de transportadores individuais que realizam o trânsito de muitos nutrientes e precursores através das membranas celulares dos Mollicutes (YOUIL; FINCH, 1988)

Os Mollicutes possuem uma alta frequência de transferência horizontal de genes em sua história evolutiva e o fluxo gênico constante com outras espécies pode explicar o processo de generalização e simplificação metabólica sofrida por este grupo.

1.5 Transferência Horizontal de Genes

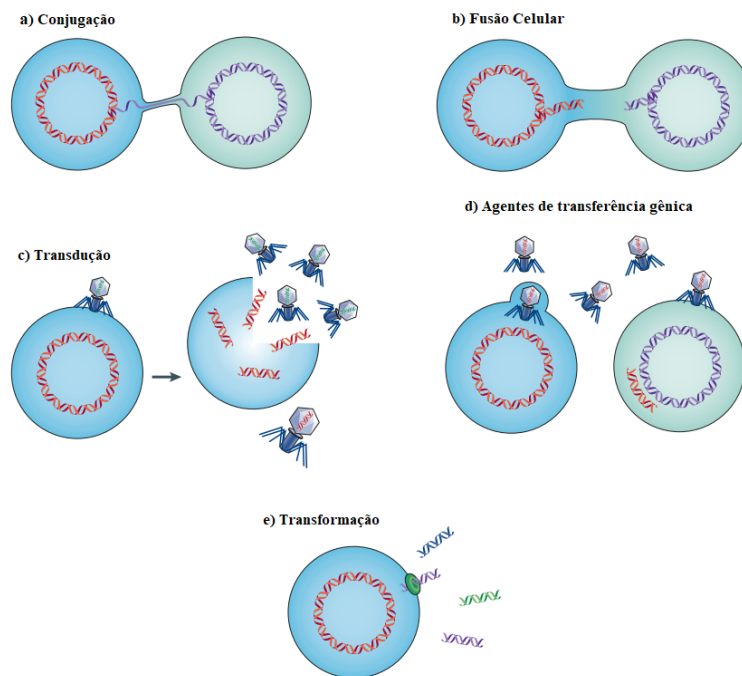
Comunidades bacterianas estão presentes em uma infinidade de ambientes, caracterizando quase todo lugar da Terra com um microbioma único. Bactérias apresentam uma diversidade fenotípica extensa relacionada ao metabolismo, processos e estrutura celulares. Esta variedade tão significativa parece contrastar com seus genomas menores e aparente simplificação quando comparadas a organismos eucariotos. Habitando desde bases de vulcões (OULAS et al., 2016), fossas oceânicas (RICHTER, 2017) e geleiras (MACKELPRANG et al., 2017) a solos (JANSSON; HOFMOCKEL, 2018) e colonizando uma diversidade de hospedeiros (que incluem desde outros procariotos a plantas, animais e fungos), as populações bacterianas estão em constante comunicação e contato. Este contato, seja direto ou indireto, permite a ocorrência de um fluxo gênico constante entre bactérias que ocupam (ou ocuparam, ao longo de sua história evolutiva), o mesmo nicho ou espaço ecológico.

O fluxo de informação gênica e aquisição de genes homólogos e não homólogos é constante em comunidades bacterianas (NIEHUS et al., 2015). A troca e incorporação de material genético entre indivíduos não relacionados, também chamada de transferência horizontal de genes, vem sendo reconhecida como uma fonte de variação gênica crescente em procariotos e eucariotos. Estudos dos anos 1990 observaram este fenômeno ocorrendo entre procariotos (JAIN et al., 1999), de bactérias para eucariotos (DOOLITTLE, 1998), de bactérias para *Archaea* (NELSON et al., 1999) e de animais para bactérias (WOLF et al., 1999). Em algumas bactérias, a transferência horizontal de genes já se mostrou como um mecanismo de aquisição gênica mais importante que a duplicação gênica (TREANGEN; ROCHA, 2011). De fato, quando falamos de geração de variedade gênica e especiação este processo se torna ainda mais crucial para organismos bacterianos. A

colonização de novos nichos e diversificação bacteriana são resultado (ou são permitidas) por modificações no repertório genético destes organismos e a diversidade de estilos de vida entre os procariotos é tão ampla que mutações pontuais sobre genes existentes se torna uma explicação limitada que não reflete a grande variação fenotípica observada neste grupo.

A aquisição de novos genes entre bactérias pode se dar através de três principais processos: transdução, conjugação e transformação (**Figura 2**). Na transdução, fagos específicos servem como veículos transportadores para fluxo e incorporação de genes no cromossomo bacteriano em regiões específicas (no ciclo lisogênico) ou pode servir como alicerce para produção de novos fagos no citoplasma bacteriano utilizando a maquinaria metabólica da célula, em um ciclo lítico que em seu estágio final causa a liberação das partículas virais pela destruição da célula hospedeira. Enquanto a transdução não necessita do contato direto entre a célula bacteriana doadora do material genético e a célula receptora, a conjugação necessita de um canal direto entre ambas, que é formado por um tubo conector ou por poros na membrana bacteriana. Plasmídeos de conjugação são grandes (chegando a 45kb) e podem conter operons inteiros que são extensos demais para serem transferidos via transdução (LINDSAY, 2014), tratando-se principalmente de genes que conferem resistência a antibióticos e a metais pesados. O processo de aquisição e integração de material genético livre do ambiente por células bacterianas recebe o nome de transformação, ocorrendo comumente através da ação de proteínas na membrana bacteriana que formam uma espécie de poro para passagem do DNA livre. Após ser adquirido por todos estes processos, o novo material genético precisa ser integrado ao cromossomo bacteriano ou ser capaz de se replicar e utilizar a maquinaria de tradução por conta própria, como é o caso de plasmídeos.

Figura 2. Principais meios de transferência horizontal de genes entre bactérias. **a)** Conjugação ocorre via interação direta célula-célula entre uma bactéria doadora e uma bactéria receptora, um DNA de fita simples é transferido entre as células via formação de um canal sexual. Fusão celular **b)** envolve a transferência de DNA bidirecional entre as células envolvidas, sendo as duas bactérias participantes doadora e receptora ao mesmo tempo. Fagos podem servir como carreadores de material genético entre células bacterianas distintas em um processo denominado transdução **c)** que envolve a incorporação de DNA da célula hospedeira pelo fago que então o carrega para os hospedeiros seguintes. Agentes de transferência gênicos **d)** são fagos que não reconhecem seu próprio DNA e carregam sequências randômicas do DNA do hospedeiro. DNA livre também pode ser incorporado diretamente do ambiente por algumas bactérias, em um processo denominado transformação **e)**.



Fonte: adaptada de SOUCY et al. (2015).

Bactérias, com a unicelularidade ao mesmo tempo restringindo e ampliando sua maquinaria metabólica pela constante aquisição e deleção gênica, possuem na transferência horizontal um potente aliado para a disseminação gênica em suas comunidades. O conteúdo genético de uma espécie pode evoluir através de diferentes mecanismos que podem ser agrupados em três grandes classes (VOS et al., 2015): perda genética, ganho de genes através de duplicação (que também pode ser referido como paralogia) e ganho genético através de transferência horizontal de genes (xenologia) (KUO; OCHMAN, 2009).

A crise de resistência a antibióticos no contexto de infecções hospitalares e o surgimento de bactérias multirresistentes e iniciou a discussão sobre a importância que este mecanismo pode ter para a sobrevivência e proliferação bacterianas. Os genes trocados entre diferentes espécies não se restringem aqueles que conferem resistência a uma gama de antibióticos, mas também podem incluir um arsenal metabólico para colonização de novos nichos (HEUER; SMALLA, 2012) (MCDANIEL et al., 2010), enzimas e proteínas para melhor aproveitamento de fontes nutritivas e uma ampla variedade de fatores de virulência, possibilitando não só a ampliação de hospedeiros, mas também a resistência (ou evasão) de respostas imunes.

Com tanto impacto, a transferência horizontal de genes parece sugerir que o material genético flui na biosfera entre os diferentes reinos e taxa (DE LA CRUZ; DAVIES, 2000), sendo assimilado ou descartado, substituindo ou ampliando o suporte metabólico dado pelos genomas aos organismos e provando-se como uma potente força evolutiva. De la Cruz e Davies (2000) discutem o surgimento de bactérias multirresistentes em ambientes hospitalares nos últimos 50 anos como um experimento em andamento que demonstra a adaptabilidade que comunidades bacterianas possuem diante da exposição a agentes tóxicos (antibióticos), pontuando a geração de resistência via transferência horizontal de genes como um evento observável de evolução na presença de uma grande pressão seletiva.

A resistência a antibióticos observada em bactérias hospitalares não teria sido, dessa forma, gerada por mutações randômicas em seus genomas, mas em vez disso adquiridas através de elementos genéticos móveis (como plasmídeos e transposons) ou assimiladas diretamente do ambiente via transportadores específicos na membrana bacteriana. A disseminação destes genes de resistência, individualmente ou em operons (DE LA CRUZ; DAVIES, 2000), por meio de grupos bacterianos é o que dá forma a mecanismos de ampla resistência observada nas chamadas superbactérias.

Em um contexto ecológico a transferência de material genético horizontalmente permite a determinados grupos a colonização de novos nichos, tornando-os capazes de resistir a novos fatores abióticos como temperatura, pH, salinidade e concentração de agentes tóxicos. Em muitos casos tal permissão se centra na aquisição de genes que possibilitam o uso de novas fontes nutritivas como diferentes formas de açúcares (HEHEMANN et al., 2010), por exemplo. Diferente da aquisição de genes de resistência a antibióticos (MCCARTHY et al., 2014) (FORSBERG et al., 2012), genes para metabolização de nutrientes estão organizados de forma coletiva, em operons. A organização genética em operons permite a bactérias regularem não só a fonte nutritiva que estão utilizando como também o custo energético atrelado a este processo.

A transmissão e aquisição de fatores de virulência via transferência horizontal de genes, seja por plasmídeos ou outros mecanismos, são de extrema importância para espécies patogênicas e muitas vezes representa o divisor de águas para desenvolvimento de patogenicidade em bactérias comensais (DE LA CRUZ; DAVIES, 2000). Exemplos de plasmídeos que ocasionam uma mudança fenotípica tão substancial a ponto de serem

considerados plasmídeos “especiadores” (por permitirem a colonização de novos ambientes, nestes casos hospedeiros multicelulares animais ou plantas) incluem os plasmídeos Ti de *Agrobacterium*, os plasmídeos de nodulação de *Rhizobium* e os plasmídeos de virulência dos gêneros *Yersinia*, *Shigella* e de *E. coli* (DE LA CRUZ; DAVIES, 2000).

E. coli e *Salmonella* foram usados como espécies modelo para o estudo da frequência de transferência horizontal ao longo de grupos bacterianos e da importância que este processo possui, mostrando que 17% dos seus genomas foram adquiridos através deste mecanismo (LAWRENCE; OCHMAN, 1998), ressaltando ainda que 90% dos genes de manutenção (*housekeeping genes*) entre estas espécies são idênticos e que uma quantidade significativa de genes “únicos” em *Salmonella* estão envolvidos em mecanismos de virulência, o que indica a aquisição de virulência como um importante fator de especiação nestes grupos (DE LA CRUZ; DAVIES, 2000).

A organização do genoma bacteriano pode ser dar em duas grandes partes: o genoma básico (*core genome*), lar de genes responsáveis pela manutenção da maquinaria metabólica que forma o alicerce da sobrevivência e replicação celular e o genoma acessório, que contém genes que podem se mostrar vantajosos sob determinadas situações ambientais. A constituição dos genomas acessórios de diversas espécies sugere que a frequência de transferência horizontal é maior neste conteúdo genético (JUHAS et al., 2009), reforçando o entendimento de que se trata de um processo capaz de dissimilar genes úteis para colonização de novos nichos (sejam ambientes com diferentes condições abióticas ou um ambiente multicelular, como o constituído por um novo hospedeiro). A migração de bactérias para novos ambientes aumenta a frequência e o impacto de transferência horizontal de genes e Niehus et al. (2015) demonstraram que a transferência é mais frequente entre genes que estão sob seleção positiva, sugerindo uma relação entre a transferência horizontal de genes e evolução. Neste ponto de vista pode-se discutir que genes, e não espécies ecológicas, habitam certos nichos e os organismos se manifestam como repositórios desse material genético (NIEHUS et al., 2015).

O estudo da transferência horizontal de genes se data por quase 100 anos (XIAO et al., 2011), mas como exige a observação complexa de comunidades bacterianas, fazê-lo experimentalmente em laboratório pode se mostrar desafiador. Ferramentas de bioinformática úteis e de fácil acesso podem permitir a análise da ocorrência de eventos

de transferência horizontal e com o abastecimento constante dos bancos de dados gratuitos com genomas sequenciados em ritmo crescente, tal abordagem vem se tornando cada vez mais comum nos últimos anos. Isto ocorre paralelamente ao desenvolvimento de novos programas e softwares com o objetivo de compreender o impacto evolutivo e importância ecológica destes eventos de transmissão gênica.

O fluxo de informação gênica entre espécies diferentes é um importante fator evolutivo e está sob pressão seletiva, novas formas de se abordar o funcionamento e regulação de genes (e as proteínas codificadas por estes) vem sendo desenvolvidas nas últimas décadas. Com o avanço da bioinformática e disponibilidade crescente do proteomas em bancos de dados de acesso fácil e gratuito, a atenção de pesquisadores e pesquisadores têm se voltado para a maneira com que estes produtos gênicos interagem. Uma maneira de fazê-lo é através da construção de redes de interação proteína-proteína.

1.6 Redes de Interação Proteína-Proteína

Estudos experimentais clássicos já indicavam que proteínas não desempenham suas funções biológicas sozinhas e estão, em vez disso, interconectadas em redes. Estas redes de interações (referidas como redes de interação proteína-proteína ou simplesmente redes IPP, do inglês *Interacting Protein-Protein Network*) se conectam exercendo papéis regulatórios, inibitórios, aditivos ou de complementação de função biológica e vem sendo estudados em uma variedade de organismos. Organismos modelos estiveram entre os primeiros a terem suas redes IPP construídas e analisadas (CONSORTIUM, 2011; ITO et al., 2001; LI et al., 2004), o que é explicado pela disponibilidade fácil de seus genomas por meio de métodos de sequenciamento clássicos e de última geração.

Acessando a rede IPP de um organismo, ou subredes específicas de interesse, é possível abordá-lo (em nível individual ou de espécie) do ponto de vista de seu interatoma. Tratamos do interatoma quando nos referimos ao conjunto de interações de macromoléculas que compõem o funcionamento e manutenção do metabolismo biológico e como este é regulado (CUSICK et al., 2005). Isto significa que redes IPP constituem não apenas uma forma de observar vias metabólicas de uma abordagem mais completa e em larga escala, como também que a natureza das interações pode revelar informações interessantes sobre a maquinaria celular em que está inserida.

Observando estas redes de interação pode-se identificar as proteínas-chave que controlam os mais diversos mecanismos celulares e vias metabólicas (RAMAN, 2010). Redes de interação de proteínas foram usadas para determinação de funções de proteínas desconhecidas (DENG et al., 2003; SHARAN et al., 2007) e possibilitam uma visão ampla dos processos celulares quando feitas em escala genômica (RAMAN, 2010). Analisando-as, torna-se possível determinar proteínas que estão em posição central de vias metabólicas e como sua presença controla o funcionamento da célula (WUCHTY, 2014). Ainda, pode-se determinar módulos funcionais das vias (ERTEN et al., 2009), indicando quais proteínas são essenciais para sobrevivência celular e com quais outras se relacionam, compondo uma maneira de indicar a robustez de vias metabólicas.

A construção de redes IPP em escala genômica serve a diferentes áreas, incluindo a biotecnologia na elucidação da produção de enzimas de interesse industrial (HAN et al., 2016) e a filogenética onde oferece novas perspectivas da evolução de grupos biológicos (ERTEN et al., 2009) (CONSORTIUM, 2011). Na área médica, redes IPP foram empregadas para analisar as interações de marcadores de doenças genéticas (LAGE et al., 2007), diferentes marcadores topológicos para cânceres (JONSSON; BATES, 2006) e na predição do prognóstico de câncer de mama (TAYLOR et al., 2009).

Na microbiologia, redes IPP foram empregadas em estudos que tiveram *Bacillus licheniformis* (HAN et al., 2016), *E. coli* (BUTLAND et al., 2005), *E. coli* e *S. cerevisiae* em um estudo comparativo realizado por (WUCHTY; UETZ, 2014) e o patógeno gástrico humano *Helicobacter pylori* (RAIN et al., 2001) como espécies alvo, ajudando a elucidar diferentes questões biológicas em cada um dos casos. O estudo mais recente realizado em 2016 tinha como alvo a identificação de subredes para otimizar a produção industrial de enzimas de interesse. A construção da rede IPP de *E. coli* teve como objetivo o estudo de complexos proteicos conservados e essenciais para a espécie. Comparando a rede de interações IPP de *E. coli* e *S. cerevisiae*, constatou-se que ambas espécies compartilham um grande número de interações em suas redes globais, sugerindo uma tendência conservativa de certas vias metabólicas centrais entre espécies distintas. *H. pylori* é um importante patógeno gástrico, estando diretamente relacionado a ocorrência de câncer no estômago (GROUP, 1993) e sua rede IPP foi abordada para um melhor entendimento do funcionamento de seus fatores de virulência mais agressivos, constatando-se proteínas centrais que poderiam ser utilizadas como possíveis alvos terapêuticos (RAIN et al., 2001).

O potencial da utilização de redes IPP segue alguns parâmetros topológicos básicos que serão descritos adiante, mas a princípio seu uso como método parece restrito apenas a criatividade dos pesquisadores e pesquisadoras envolvidos e a literatura demonstra isto. Estudando o interatoma planctônico global, Lima-Mendez et al. (2015) demonstraram que fatores ambientais e perfis de abundância de organismos podem ser utilizados para construir as redes de interação em um ecossistema e que fatores ambientais podem prever a diversidade de espécies de um determinado ambiente. Retornando para redes de interações proteicas, estudos relacionando o interatoma de organismos como *Treponema* (TITZ et al., 2008) sugeriram que, diferente de proteínas de eucariotos, proteínas bacterianas tem uma tendência geral a interagir mais umas com as outras, em um reflexo do que os autores se referiram como a não-compartmentalização das células bacterianas. Dunin-Horkawicz et al. (2014) utilizaram a construção de redes IPP de proteínas bacterianas como um modelo para analisar a evolução do equivalente procariótico da resposta imune inata e apoptose observadas em organismos eucarióticos. Entre os resultados deste estudo realizado com ferramentas bioinformáticas está a hipótese de que a resposta imune inata de eucariotos teve sua origem em endossimbiontes procarióticos.

2. OBJETIVOS

O objetivo principal deste trabalho é a construção da rede de interação proteína-proteína de *U. diversum*, utilizando-a como método para estudo de seu interatoma, a fim de testar se o mesmo reflete a redução genômica e a simplificação de vias metabólicas observadas nos Mollicutes.

Os objetivos específicos são:

Validação da rede IPP de *U. diversum*.

Análise e observação das redes IPP simplificadas e de vias metabólicas incompletas para um maior entendimento de como *U. diversum* consegue suprir seus requerimentos metabólicos.

Identificação de proteínas-chave (proteínas *hubs*) que concentram um maior número de interações na rede global e que podem ser usadas como possíveis alvos terapêuticos.

Caracterização da subrede da urease e análise de eventos de transferência horizontal de genes.

Caracterização das subredes do metabolismo de purinas e pirimidinas e análise de eventos de transferência horizontal de genes.

3. MATERIAIS E MÉTODOS

3.1.) Construção das redes IPP de *Ureaplasma diversum*

Sequências de proteínas conhecidas obtidas da anotação genômica de *U. diversum* (MARQUES et al., 2016) foram submetidas ao STRING (*Search Tool for the Retrieval of Interacting Genes/Proteins*) (SZKLARCZYK et al., 2014) para construção das redes preditas iniciais. STRING é um banco de dados computado que reúne interações proteína-proteína de diferentes espécies. As interações preditas pelo banco de dados são derivadas de dados experimentais, de outros bancos de dados e da literatura, além de análises de coexpressão e fusão gênica. O STRING realiza a predição destas interações utilizando uma forma de pontuação baseada nas diferentes associações que corroboram determinada interação na rede IPP, dando a cada interação uma pontuação individual. Interações baseadas em dados experimentais, por exemplo, são atribuídas com índice de confiabilidade mais alta pelo STRING, enquanto interações baseadas na localização genômica dos genes envolvidos recebem automaticamente um índice de confiabilidade menor (SZKLARCZYK et al., 2014).

As redes foram construídas de acordo com os mecanismos celulares e vias metabólicas de suas respectivas proteínas. Ainda no STRING foram observadas informações gerais sobre a natureza das interações preditas geradas pelo banco de dados IPP. As redes foram visualizadas no *Cytoscape* (SHANNON et al., 2003) para análise de características individuais de cada rede e seus parâmetros correlacionados. O *Cytoscape* é uma plataforma de software utilizada para visualização de redes de interação proteína-proteína e é usado em conjunto com bancos de dados de interações, entre eles o STRING. Uma vez que o *Cytoscape* suporta diferentes algoritmos para visualização e análise de redes IPP, diversos *plug-ins* gratuitos foram desenvolvidos ao longo dos anos com diferentes focos e objetivos, aumentando a capacidade de análise do programa base. Entre estes *plug-ins*, o mais utilizado e que já é integrado na versão básica do *Cytoscape* é o

NetworkAnalyzer (ASSENOV et al., 2007) que permite a medição dos parâmetros topológicos da rede IPP visualizada.

O plugin *NetworkAnalyzer* foi então empregado para análise dos parâmetros topológicos da rede. Posteriormente o *plugin ClusterOne* (NEPUSZ et al., 2012) foi utilizado para identificação de clusteres de interação contidos na rede global e para rearranjo destas subredes.

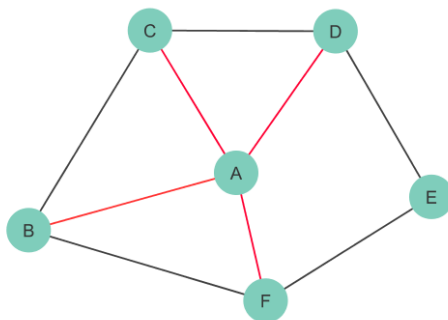
3.1.1.) Parâmetros topológicos de redes IPPs

Redes IPP são formadas pelas proteínas (ilustradas por nodos) de um organismo ou processo celular e as interações bioquímicas (representadas por conexões ou setas) preditas entre estas (WENG et al., 1999). Nodos com muitas interações com outras proteínas da rede recebem a denominação de proteínas-chave (do inglês *hub proteins*) e muitas vezes representam uma via metabólica ou de sinalização celular distinta. Os parâmetros gerais de uma rede IPP se referem às diferentes características topológicas que a rede possui referentes a suas interações e nodos e como estes se distribuem espacialmente no gráfico.

Redes IPP que representam sistemas biológicos como vias metabólicas e processos celulares possuem uma natureza de escala livre (*scale-free*), o que significa que sua distribuição de graus (k) segue uma lei de potência $P(k) \propto k^{-\gamma}$ em que γ corresponde ao expoente da rede. Algumas proteínas da rede (as chamadas proteínas *hub*) concentram a maior parte das interações, mantendo a centralidade da rede e conectando subredes enquanto a maioria das proteínas interage com poucas vizinhas. Redes biológicas com esta conformação possuem uma alta robustez a ataques randômicos que removem nodos de seu interior, pois conseguem se rearranjar facilmente e contornar a lacuna deixada pela proteína removida e suas interações. Já ataques que focam proteínas *hubs* (aquelas que interagem em maior frequência com proteínas vizinhas, conectam subredes e representam o alicerce de processos celulares distintos) costumam ser letais para a conformação de redes biológicas. Proteínas *hubs* envolvidas no funcionamento de fatores de virulência, por exemplo, podem ser alvo de testes de eliminação da rede IPP o que resulta na identificação de novos alvos para tratamentos terapêuticos (CHANG, 2009).

Para caracterizar redes IPP alguns parâmetros podem ser calculados. O grau, também chamado de conectividade (k) de um nodo mede o número de interações que este nodo possui com os demais nodos da rede IPP (**Figura 3**).

Figura 3. O grau (conectividade) k do nodo A se dá pela soma das interações deste nodo com os demais nodos da rede, neste caso $k_A =$ número de interações que passam por A = 4.



Em redes randômicas todos os nodos possuem um número similar de interações, ou seja, o mesmo grau (k), o que as torna vulneráveis a ataques randômicos que eliminam nodos da rede aleatoriamente. Em redes de escala livre, ou seja, aquelas que representam processos biológicos, as proteínas *hubs* concentram a maior quantidade das interações da rede e a distribuição de grau se aproxima da lei de potência.

A distribuição de graus $P(k)$ é uma medição que se refere a rede como um todo, indicando a probabilidade que o nodo selecionado tenha exatamente k interações. $P(k)$ é calculado pela contagem dos nodos $N(k)$ com $k = 1, 2, 3, 4, \dots$ interações e dividindo este total pelo número de nodos N . É este parâmetro usado para a caracterização do tipo de rede IPP, indicando se se trata de uma rede com conformação hierárquica, biológica (de escala-livre) ou que segue uma escala.

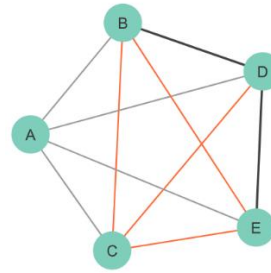
O coeficiente de agrupamento (*clustering coefficient*), C , de um nodo se refere a quão conectados os vizinhos deste nodo estão na rede IPP medindo, portanto, a interconectividade da vizinhança de um nodo (as proteínas com as quais esta interage) (**Figura 4**).

Figura 4. Coeficiente de agrupamento de A, C_A , e seu cálculo.

$$C_A = \frac{\text{Interações entre os vizinhos de A (em preto)}}{\text{Possíveis interações entre os vizinhos de A (em laranja)}}$$

$$C_A = n_A / [k_A (k_A - 1) / 2]$$

$$= 2 / [4 \times (4 - 1) / 2] = 0.333$$



O seu equivalente global é o coeficiente médio de agrupamento da rede que leva em consideração todos os nodos e seus coeficientes de agrupamentos individuais. O coeficiente de agrupamento de uma rede se refere a tendência que as proteínas desta rede possuem de formar agrupamentos mais conectados entre si. Segundo Raman (2010) em um gráfico não direcionado, se um vértice v_i possui k_i vizinhos, $k_i(k_i - 1)/2$ interações poderiam existir entre os vértices da vizinhança (N_i). O coeficiente de agrupamento para um gráfico não direcionado $G(V, E)$, onde V representa o set de vértices no gráfico e G e E representam o set de interações, pode ser representado da seguinte maneira (**Figura 5**). Tratamos gráficos de redes IPP biológicos como não direcionados por não terem uma força ou coeficiente que direcione as interações na rede.

Figura 5. Coeficiente médio de agrupamento de um gráfico não direcionado $G(V, E)$; em que V representa os nodos da rede IPP e E as interações).

$$C_i = \frac{2|E_{jk}|}{k_i(k_i - 1)}; v_j, v_k \in N_i, e_{jk} \in E.$$

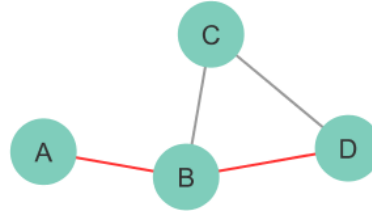
Fonte: adaptado de RAMAN, 2010.

O coeficiente de agrupamento também é uma forma de se aferir a propriedade hierárquica de uma rede IPP.

O tamanho de via característico, L , se refere ao número de interações da via mais curta entre dois nodos da rede, e é obtido pela média de todos os pares de nodos que compõem uma rede. Em outras palavras, é o parâmetro de rede que se refere a separação entre dois vértices de uma rede, o que representa a “velocidade” em que a informação navega pela rede IPP.

A distância ou tamanho de via entre dois nodos se refere a quantidade de interações que separa estes dois nodos na rede IPP, o seu “caminho”. A via mais curta (do inglês *shortest path*) se refere ao caminho com o menor número de interações, ou seja, representa a via mais rápida entre os dois nodos envolvidos (**Figura 6**).

Figura 6. Cálculo da via mais curta entre dois nodos A e D. $SP_{AD} = (A, B, D) = 2$.



A maior distância entre quaisquer dois nodos em uma rede é dada pelo diâmetro da rede, d . Raman (2010) determina que também pode ser vista como o comprimento da via curta mais longa da rede IPP. O diâmetro de uma rede pode ser calculado utilizando a seguinte fórmula (**Figura 7**):

Figura 7. Cálculo do diâmetro de uma rede IPP. Em que $d_G(u, v)$ é a via mais curta entre u e v no gráfico G .

$$d = \max_{u, v \in G} d_G(u, v)$$

Fonte: adaptado de RAMAN, 2010.

A Interseção (*betweenness centrality*) de uma rede IPP está entre os parâmetros de centralidade mais utilizadas e se refere a medida de um vértice (nodo) dentro da rede. Para uma rede em que $G(V, E)$, com n nodos, a interseção $C_B(v)$ de um vértice v pode ser definido como mostrado (RAMAN, 2010) (**Figura 8**):

Figura 8. Cálculo da interseção de um nodo em uma rede IPP.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Fonte: adaptado de RAMAN, 2010.

Onde σ_{st} é o número de vias curtas de s a t , e $\sigma_{st}(v)$ é o número de vias curtas de s a t que passam pelo vértice v . Nodos com uma interseção maior estão localizados em pontos com um maior número de vias mais curtas de uma rede.

A conectividade da vizinhança (*neighborhood connectivity*) representa o grau (k) médio dos vizinhos mais próximos de um nodo.

Em uma rede IPP, os vizinhos de um nodo são as proteínas que interagem com este nodo até uma certa distância, refletindo muitas vezes funções biológicas que estas proteínas possuem em processos celulares e vias metabólicas da célula. Esta propriedade pode ser utilizada para determinação da função de proteínas desconhecidas, uma vez que proteínas que interagem em uma rede IPP costumam ter funções complementares, aditivas ou reguladoras, formando complexos estruturais distintos.

Barabási e Albert (1999) determinaram que a probabilidade $P(k)$ de um vértice (nodo) de uma rede interagir com k outros vértices cai seguindo uma lei de potência $P(k) \sim k^{-\gamma}$ em que γ corresponde ao expoente da rede, como explicado anteriormente. Esta relação se baseia nas duas principais características de redes IPP reais: 1) a capacidade de incorporar novos nodos adicionados à rede (que em um contexto biológico se refere a incorporação de novos genes e proteínas em vias metabólicas e de sinalização) e 2) a preferência de ligação de novos nodos a subredes em maior conexão (BARABÁSI; ALBERT, 1999). Este comportamento observado em redes de escala-livre pode explicar a evolução da complexidade de redes metabólicas, partindo de moléculas simples com poucas interações até organismos multicelulares com verdadeiras maquinarias celulares intrincadas.

3.1.2.) Validação das redes IPP de *Ureaplasma diversum*

Para validação da rede IPP global de *U. diversum* utilizamos a nota de confiabilidade média (0.4) do STRING como nota de corte, interações com uma pontuação de confiabilidade menor que 0.4 foram excluídas das análises. Além disso, para testar a robustez da rede IPP utilizamos o *plugin NetworkRandomizer* (TOSADORI et al., 2016) no *Cytoscape* para geração de redes randômicas cujos parâmetros de centralidade foram comparados aos da rede predita de *U. diversum* para validação.

3.2.) Construção das redes IPP de espécies usadas para comparação de complexidade de redes

Ureaplasma urealyticum, *Ureaplasma parvum*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Mycoplasma hominis*, *Haloplasma contractile*, *Mycobacterium pneumoniae*, *Corynebacterium urealyticum*, *Corynebacterium*

glucurolyticum e *Staphylococcus aureus* foram usados para comparação de redes IPP. Estas espécies foram escolhidas devido a sua relação filogenética dentro e fora da classe Mollicutes. O proteoma das espécies usado para comparação foram baixadas do banco de dados de proteínas curadas Uniprot (CONSORTIUM, 2016). As proteínas também foram visualizadas na plataforma KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (KANEHISA; GOTO, 2000) e divididas em suas vias metabólicas. O proteoma de cada espécie foi submetido ao STRING para geração das redes IPP genômicas e observação de suas características gerais. As redes IPP foram então visualizadas no Cytoscape para análise topológica e conformacional.

3.3.) Comparação das redes IPP de *U. diversum* e espécies de referência

As redes IPP das espécies para comparação *U. urealyticum*, *U. parvum*, *M. genitalium*, *M. pneumoniae*, *M. hominis*, *H. constricta*, *Mycobacterium tuberculosis*, *C. urealyticum* e *Staphylococcus aureus* foram carregadas no Cytoscape onde as características topológicas e conformação das redes IPP foram avaliadas e o perfil comparativo das subredes entre estas espécies e *U. diversum* foram estabelecidas através da análise de seus parâmetros topológicos de centralidade.

3.4.) Reconstrução da filogenia da classe Mollicutes de acordo com a complexidade de subredes e análise de transferência horizontal de genes

As subredes IPP de *U. diversum* e espécies comparativas de interesse foram organizadas em um arquivo multifasta e submetidas ao Mafft (KATO; STANDLEY, 2013) para alinhamento global e então inseridos no ProTest (CUFF *et al.*, 2000) para geração dos melhores modelos evolutivos para construção da filogenia. O output foi então submetido ao TRex (CULMAN *et al.*, 2009) para construção e visualização das árvores filogenéticas. Nós também utilizamos o TRex para observação de ocorrência de transferência horizontal entre as espécies estudadas em relação as subredes de interesse. As subredes de interesse escolhidas para *U. diversum* foram as relacionadas a metabolismo de purinas e pirimidinas e a função da urease. As árvores filogenéticas obtidas por este método foram então comparadas a árvores construídas usando o gene ribossômico 16S para validação. As sequências 16S foram obtidas da anotação genômica para *U. diversum* e da base de sequências de nucleotídeos do NCBI (COORDINATORS, 2016) para as demais espécies e o melhor modelo evolutivo foi calculado utilizando o

Mega (TAMURA et al., 2007). Os parâmetros obtidos do Mega foram inseridos no TRex (CULMAN et al. (2009) para construção e visualização das árvores filogenéticas 16S.

4. RESULTADOS E DISCUSSÃO

4.1 Características e parâmetros gerais da rede IPP de *U. diversum* e sua validação

A rede IPP global de *U. diversum* contém 301 nodos e 10.404 interações. A rede foi construída utilizando o índice de confiabilidade médio (0.4) do STRING como nota de corte, interações preditas com índice de confiabilidade menor que 0.4 foram removidas da análise. Das 10.404 interações IPP da rede global, 20.71% (2.155 interações) obtiveram índice de confiabilidade altíssima (> 0.9), 35.86% (3.732) foram atribuídas confiabilidade alta (0.6 - 0.899) e 43% (4.517) das interações preditas possuem um índice de confiabilidade média (0.4 – 0.599).

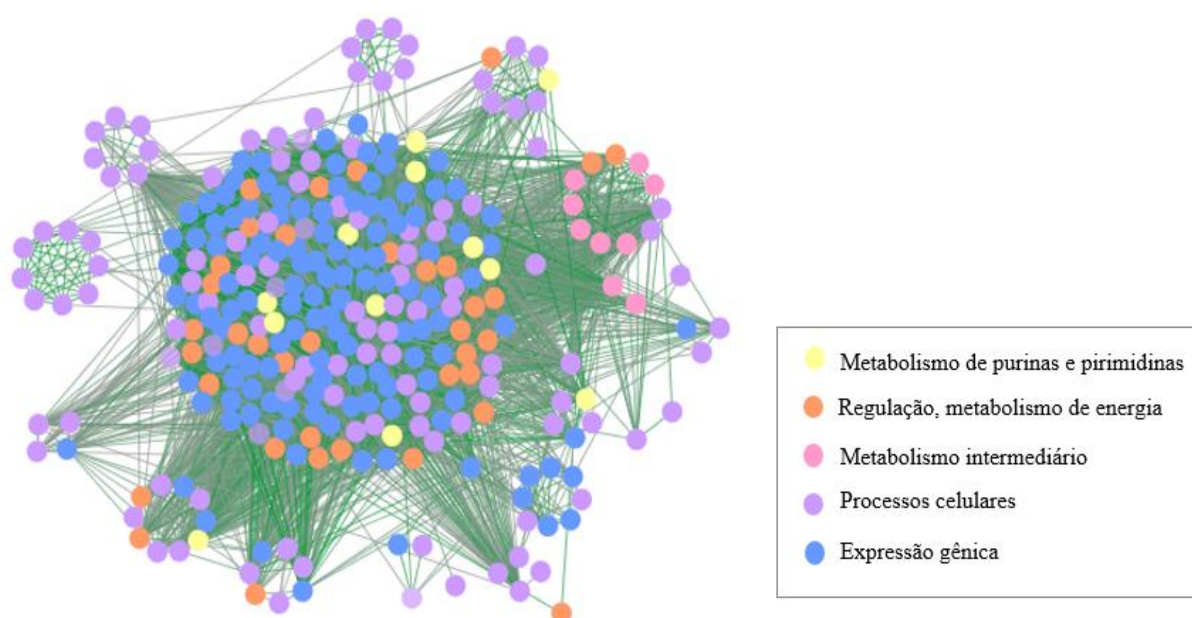
Os 301 nodos da rede IPP de *U. diversum* também foram categorizados e coloridos de acordo com a função da proteína que representam: proteínas envolvidas no fluxo da informação gênica estão representadas em azul; proteínas envolvidas em processos celulares, transporte de proteínas, destino de proteínas e envelope celular, em roxo; em laranja as proteínas com funções regulatórias, metabolismo de ácidos graxos e metabolismo de energia, além de cofatores e carreadores; em amarelo proteínas e enzimas relacionadas ao metabolismo de purinas e pirimidinas e em rosa estão as proteínas envolvidas no metabolismo intermediário e central de nitrogênio, que compreende o cluster da urease (**Figura 9**).

Figura 9. Visão geral da rede IPP de *U. diversum*.

a)



b)

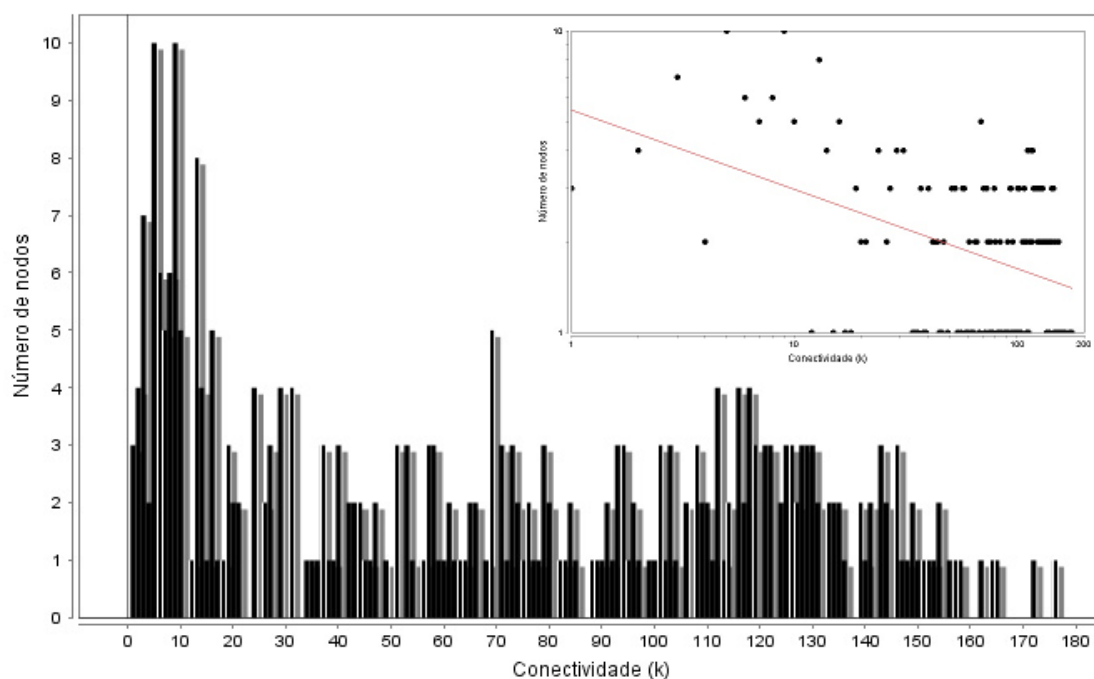


Ao calcular os parâmetros gerais de uma rede IPP podemos acessar sua conformação e obter informações valiosas referentes ao tipo de rede sob investigação. Os parâmetros gerais da rede global de *U. diversum* estão representados nas figuras a seguir (**Figura 10, Figura 11**). O significado dos parâmetros gerais de redes IPP foi explicado

anteriormente (Material e Métodos, p. 17), dessa maneira, focaremos no significado biológico que cada item representa para o metabolismo de *U. diversum*.

A distribuição de conectividade, também chamado de grau (k) dos nodos da rede IPP, se refere ao número de nodos que cada nodo da rede de *U. diversum* está conectado. Em outras palavras, é uma representação da interconexão das proteínas na rede. A distribuição de conectividade da rede IPP de *U. diversum* segue uma lei de potência, o que a configura como rede biológica (**Figura 10**).

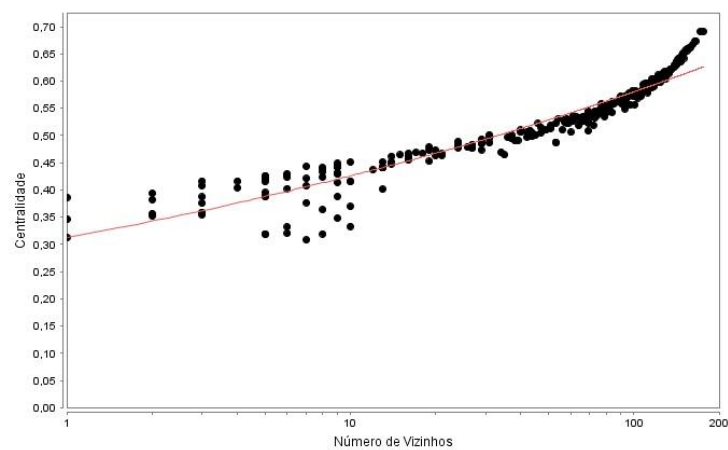
Figura 10. Conectividade (k) dos nodos da rede IPP de *U. diversum*.



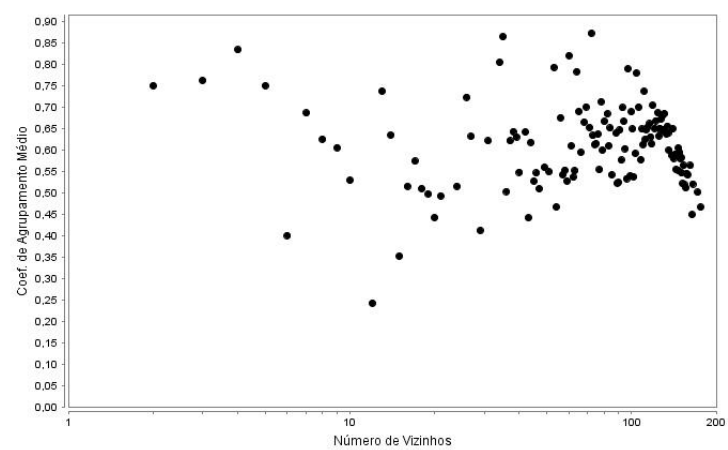
Quando a distribuição de nodos de uma rede IPP segue uma lei da potência, ou seja, a maioria dos nodos da rede estão conectadas a poucas proteínas, enquanto poucas proteínas centrais (*hubs*) concentram a maior parte das interações, a rede é classificada como rede biológica (RAMAN, 2010). A distribuição de grau (conectividade k) de uma rede IPP representa a interconexão entre os nodos da rede. Em um contexto biológico, isto significa que em *U. diversum* poucas proteínas interagem com muitas, formando o alicerce dos processos celulares enquanto a maioria das proteínas se aglomeram em grupos interativos menores. Isto também pode ser observado analisando-se os demais parâmetros topológicos da rede (**Figura 11**).

Figura 11. Distribuição da a) centralidade dos nodos, b) coeficiente médio de agrupamento e c) tamanho característico de via na rede IPP de *U. diversum*.

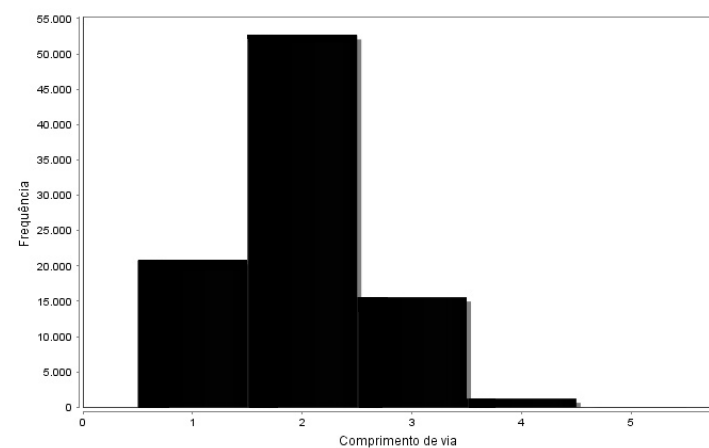
a)



b)



c)



A centralidade do nodo (**Figura 11a**) aumenta de acordo com o número de vizinhos a qual um nodo está conectado, o que pode ser visualizado no gráfico. Proteínas com uma alta conexão com sua vizinhança automaticamente exibirão um coeficiente de centralidade maior que proteínas com poucas interações. De forma semelhante, o coeficiente médio de agrupamento da rede (**Figura 11b**) também segue o número de nodos vizinhos aos quais o nodo analisado se conecta. O comprimento de via menor (*shortest path length*) se refere a velocidade com que a informação flui ao longo da rede identificando o caminho mais curto entre dois nodos. Na rede IPP de *U. diversum*, o comprimento de via mais curta (ou seja, a quantidade de nodos que separa dois nodos) mais frequente é dois (**Figura 11c**), mostrando que o fluxo de informação se dá de maneira muito rápida. Em um contexto metabólico, isto significa que a rede IPP de *U. diversum* possui uma tendência a uma conexão maior entre suas proteínas integrantes, o que sugere um rápido fluxo de informação no metabolismo (RAMAN, 2010).

Redes de interação de proteínas precisam ser validadas para serem consideradas uma representação de um sistema biológico. O alto índice de confiabilidade atribuído às conexões preditas pelo STRING (SZKLARCZYK et al., 2014) serve este propósito, indicando que a rede é sustentada por interações comprovadas empiricamente, por proximidade de sítios no genoma e pelos processos celulares em que estão envolvidos. Para analisar mais profundamente as interações preditas pelas redes, realizamos a validação da rede biológica baseada também em comparação direta com os parâmetros de redes randômicas utilizando o plugin NetworkRandomizer (TOSADORI et al., 2016) no Cytoscape (SHANNON et al., 2003).

Comparar a conformação e parâmetros de centralidade de uma rede predita com redes randômicas geradas a partir dela é uma maneira de testar a robustez dos dados utilizados na análise e construção da rede IPP (TOSADORI et al., 2016). O *plugin NetworkRandomizer* realiza a geração de redes randômicas utilizando um algoritmo simples de embaralhamento (TOSADORI et al., 2016) da rede sob investigação. Para validar a rede IPP de *U. diversum*, geramos 10 redes randômicas utilizando-a como base e comparamos os índices de centralidade entre si (**Tabela 1**). As redes randômicas geradas pelo *NetworkRandomizer* podem ser vistas nos apêndices desta dissertação. Apenas os índices de coeficiente de *clustering* e centralidade da rede foram usados, uma vez que os demais parâmetros gerais permanecem os mesmos entre as redes analisadas por conterem os mesmos nodos e interações.

Tabela 1. Parâmetros da rede IPP predita de *U. diversum* e das 10 redes randômicas geradas para validação.

	Coef. de agrupamento	Centralidade
<i>U. diversum</i>	0.613	0.359
Random_1	0.230	0.073
Random_2	0.229	0.077
Random_3	0.230	0.087
Random_4	0.231	0.087
Random_5	0.231	0.057
Random_6	0.229	0.060
Random_7	0.230	0.067
Random_8	0.232	0.080
Random_9	0.232	0.083
Random_10	0.230	0.073

O coeficiente de agrupamento (*clustering coefficient*) é consideravelmente menor nas redes randômicas quando comparado a rede IPP predita de *U. diversum* (0.613 em *U. diversum* para 0.230 médio nas redes randômicas), demonstrando que as redes randômicas possuem uma tendência menor a formar grupos de proteínas que interagem entre si em maior frequência. Os índices de centralidade das redes randômicas também são menores comparado aos de *U. diversum*, que apresenta um índice de 0.359 para um índice de centralidade de rede médio de 0.074 nas redes randômicas. Estes índices indicam que as redes randômicas são fracamente conectadas, o que reforça sua classificação como rede não-biológica e cimenta a rede IPP predita de *U. diversum* como representativa de uma rede orgânica robusta (TOSADORI et al., 2016).

4.2 Comparação da rede global de *U. diversum* com espécies relacionadas

Buscamos observar se a redução genômica, simplificação e generalização das vias metabólicas em *Ureaplasma* (e *Mycoplasma*) reflete na conformação e parâmetros gerais das redes IPP e para isto realizamos a comparação da rede destes representantes da classe Mollicutes com organismos filogeneticamente mais distantes e considerados mais complexos como *S. aureus* e espécies de *Corynebacterium* (**Tabela 2**).

Tabela 2. Principais parâmetros topológicos das redes IPP das espécies analisadas.

	N. Nodos	N. Interações	Coef. Agrup.	Centralidade	N. médio de vizinhos	Densidade de rede	Compr. via
<i>U. diversum</i>	301	10.404	0.613	0.359	69.130	0.230	1.970
<i>M. pneumoniae</i> FH	629	14.418	0.460	0.253	45.844	0.073	2.406
<i>M. hominis</i> ATCC 23114	523	12.706	0.558	0.276	48.589	0.093	2.512
<i>M. genitalium</i> G37	461	14.197	0.533	0.307	61.592	0.134	2.179
<i>U. urealyticum</i> ATCC 33699	646	16.964	0.576	0.259	52.520	0.081	2.616
<i>U. parvum</i> ATCC 700970	614	17.062	0.564	0.253	55.577	0.091	2.484
<i>C. urealyticum</i> DSM 7109	2.022	30.153	0.364	0.224	29.825	0.015	3.091
<i>C. glucurolyticum</i> ATCC 51867	2.645	53.828	0.351	0.181	40.702	0.015	2.981
<i>S. aureus</i> MSHR1132	2.510	41.222	0.367	0.190	32.846	0.013	3.105
<i>S. aureus</i> COL	2.615	52.435	0.377	0.185	40.103	0.015	2.995
<i>S. aureus</i> Mu50	2.730	57.290	0.340	0.204	41.971	0.015	2.918

O coeficiente de agrupamento (*clustering coefficient*) se refere a tendência das proteínas inseridas em uma rede IPP a formarem grupos em que as proteínas interagem em maior frequência entre si, como é observado em muitos sistemas biológicos. Os coeficientes analisados neste estudo mostram uma tendência dos Mollicutes a interagirem mais em agrupamentos que espécies filogeneticamente afastadas, o que é representado pelos seus coeficientes de agrupamento maiores. A média dos coeficientes de agrupamento entre os Mollicutes é de 0.550, enquanto *Corynebacterium* possui uma média de 0.357 e as populações de *S. aureus* analisadas um coeficiente médio de agrupamento de 0.361. Isto pode ser reflexo de uma generalização das funções proteicas diante do processo de redução genômica sofrida pelo grupo, tornando as proteínas remanescentes dos Mollicutes mais propensas a interagirem entre si em maior frequência.

Além disto, as espécies de *Mycoplasma* e *Ureaplasma* estudadas também possuem um índice de centralidade maior em suas redes, indicando que há uma aglomeração de proteínas em maior interação que em *Corynebacterium* e nas espécies de *S. aureus*. As redes IPP das espécies usadas para comparação estão disponíveis nos anexos desta dissertação. As proteínas de *Mycoplasma* e *Ureaplasma* possuem índices de conectividade maior, o que é ilustrado pelo número médio de vizinhos (proteínas conectadas) que as proteínas da rede possuem. De maneira geral, a redução genômica e

simplificação de vias observada na classe Mollicutes parece ser acompanhada por um processo de generalização das proteínas de seus metabolismos e de suas interações.

A ocorrência de genomas menores acarreta em uma redução dos processos celulares, como abordado pelos estudos que buscaram compreender o funcionamento de genomas mínimos: em um contexto de perda gênica não se tem espaço para vias metabólicas compostas por 30 proteínas quando 10 conseguem realizar este trabalho equiparavelmente. De forma semelhante, espécies que dependem de hospedeiros para funcionamento de suas vias metabólicas “encurtam” seu trabalho (e espaço em seu repertório gênico) através da importação de precursores ou até mesmo compostos metabólicos finalizados do ambiente multicelular em que estão inseridos. Estas duas características estão presentes nos *Ureaplasma* e *Mycoplasma* e nossos dados suportam que isto se reflete em suas redes de interação proteína-proteína. Com redes IPP mais conectadas que espécies consideradas mais complexas (e com genomas maiores, como é o caso de *S. aureus* e *Corynebacterium*), estes integrantes da classe Mollicutes guiados pelo processo de redução do genoma ao longo da sua evolução passaram por um rearranjo em suas redes de interação de maneira a ter proteínas mais interativas. Esta maior conexão entre as proteínas é capaz de driblar problemáticas resultantes de suas vias metabólicas incompletas e favorece uma evolução generalista, em que uma proteína é capaz de realizar as funções desempenhadas por várias em espécies correlacionadas com metabolismos mais intrincados, como será tratado a seguir.

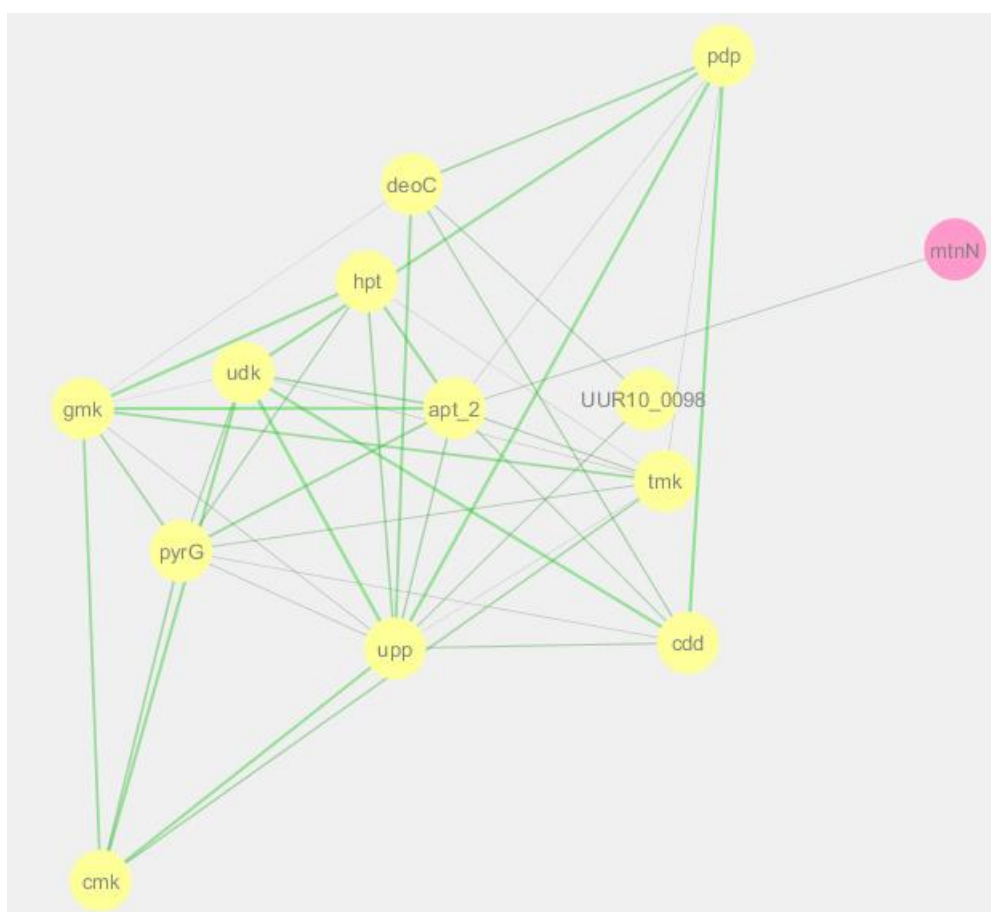
4.3 Subredes de interesse para o metabolismo de *U. diversum*

Mollicutes realizam a importação de nucleosídeos e bases pré-existentes do hospedeiro, usando nucleobases ou nucleosídeos para sintetizar nucleosídeos trifosfatos específicos e deoxinucleosídeos trifosfato (ARRAES et al., 2007). Purina nucleosídeo fosforilase é uma enzima importante para o metabolismo de purinas e pirimidinas e tem uma função de nucleosídeo ribosiltransferase (MOMBACH et al., 2006), mas está ausente na anotação genômica de *U. diversum*, além de não estar presente em outros Ureaplasmas e Micoplasmas. *M. genitalium* e *M. pneumoniae* não possuem transportadores de nucleobases ou nucleosídeos em seus genomas (PAULSEN et al., 2000). Outros transportadores com uma grande variedade de substratos identificados em ambas as espécies poderiam ser parte do transporte de precursores de nucleotídeos, como as 11 *ATP-binding cassette* (ABC) e o transportador primário ativo *Major Facilitator*

Superfamily (MSF) (POLLACK et al., 2002). Minion *et al.* (1993) testaram 20 espécies de *Mycoplasma* para atividade nuclease associada a membrana que podem ser críticos para o transporte de precursores de nucleotídeos através da membrana bacteriana e descobriram que estas reações ocorrem nestas espécies. Alguns anos mais tarde, *mnuA*, um gene codificador de uma nuclease de membrana, foi clonado e isolado pela primeira vez em *Mycoplasma pulmonis* (JARVILL-TAYLOR et al., 1999). Nucleases de membrana podem estar envolvidas em virulência, mas podem também contribuir para a degradação de RNA e DNA do hospedeiro, o que poderia ser uma fonte de bases livres e precursores de nucleotídeos para *Mycoplasma* (RAZIN *et al.*, 1998). Algo semelhante pode ser a fonte de precursores de purinas e pirimidinas em *Ureaplasma*, mesmo que nenhuma sequência codificadora de *mnuA* tenha sido identificada em *U. diversum*.

Na subrede do metabolismo de purinas e pirimidinas de *U. diversum* podemos observar conhecidas proteínas e enzimas envolvidas no resgate de nucleotídeos e nucleosídeos interagindo diretamente. Para uma visualização mais direta desta confiabilidade na rede, utilizamos ferramentas de edição de interações no *Cytoscape* para incorporar estes valores na rede IPP. Interações com confiabilidade média (0.4) foram coloridas com verde escuro, confiabilidade alta (0.6-0.899) com verde e interações com índice de confiabilidade altíssima (>0.9) estão representadas em verde claro (**Figura 12**). Uridina quinase (*udk*), citidina deaminase (*cdd*), adenina fosforibosiltransferase (*apt_2*), uracilfosforibosiltransferase (*upp*), hipoxantina fosforibosiltransferase (*hpt*) e metiltioadenosina nucleosidase (*mtnN*) estão envolvidas no processo de aproveitamento de nucleobases provenientes de importação de precursores do hospedeiro. CTP sintetase (*pyrG*) está envolvida no processo de biossíntese de pirimidinas a partir de precursores importados. Timidilato quinase (*tmk*), guanilato quinase (*gmk*) e citidilato quinase (*cmk*) são responsáveis pela interconversão de nucleotídeos e nucleosídeos. Desoxiribose-fosfato aldolase (*deoC*), pirimidina nucleosídeo fosforilase (*pdp*) e deoxiguanosina quinase (*UUR10_0098*) também compõem a subrede do metabolismo de purinas e pirimidinas em *U. diversum*.

Figura 12. A subrede do metabolismo de Purinas e Pirimidinas em *U. diversum*.



Na via de pirimidinas, uracil fosforibosiltransferase (*upp*) é responsável pela fosforilação da uracila em citidina (TU; TURNBOUGH, 1997) (LUNDEGAARD; JENSEN, 1999). Uridilato cinase catalisa a fosforilação de UMP em muitos organismos (SCHULTZ et al., 1997), mas uma vez que está ausente na anotação genômica de *U. diversum*, citidilato cinase (*cmk*) pode ser a proteína responsável pela fosforilação de UMP (BUCURENCI et al., 1996).

Ribonucleotideo redutase (*rnr*) é uma enzima ubíqua responsável pela síntese de deoxirribonucleotídeos e como esperado, está presente em *U. diversum*. Ribonucleotideo redutase está entre os nós com maior conectividade (grau, *k*) da rede IPP, contando com 110 interações preditas e um coeficiente de agrupamento de 0.627. Entre as proteínas que interagem com *rnr* estão *cdd* e *gmk*, o que sugere um trabalho em conjunto para resgate e interconversão de nucleotídeos. Reichard (2002) determinou que essa única proteína é capaz de reduzir todos os quatro ribonucleotídeos comuns.

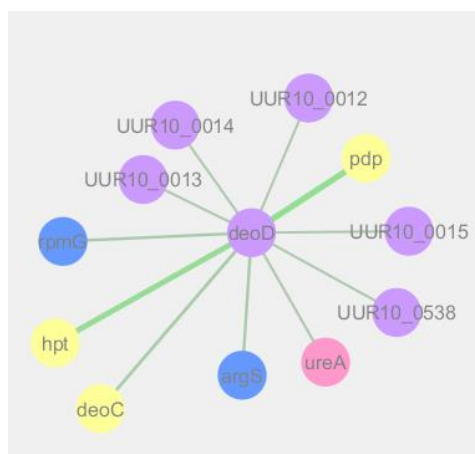
Timidilato quinase (*tmk*) é parte da superfamília de nucleosídeo monofosfato quinases e é responsável pela reação de fosforilação de timidina monofosfato, usando ATP como doador preferível de fosforibosil. Como em outros organismos, a regulação de timidilato quinase em *U. diversum* é essencial devido a sua participação no controle geral de síntese de DNA através de ajustes na disponibilidade de dTTP.

Nas enzimas metabólicas de purinas, *U. diversum* conta com guanilato quinase (*gmk*) em sua maquinaria metabólica. Guanilato quinase é responsável pela transferência do grupo fosforila de ATP para GMP, que é reversível (LI et al., 1996).

A anotação genômica de *U. diversum* mostrou a ausência de timidina fosforilase, o que aponta timidina como precursor importado para produção de dTTP ou até mesmo outra enzima poderia ser responsável por esta função (SANTOS et al., 2011).

Micoplasmas do grupo *Hominis* compartilham o gene da purina nucleosídeo fosforilase (*deoB*), mas ele não foi encontrado em outros grupos filogeneticamente próximos e também está ausente em *U. diversum*. Mesmo não tendo o gene *deoB* (ou qualquer gene relacionado com *deoB*), estudos com *U. urealyticum* e *M. mycoides* mostraram atividade de fosfopentomutases em ensaios (COCKS et al., 1985). Ainda que não tenham o gene *deoB*, *M. mycoides*, *M. florum* e todos os micoplasmas do grupo *Pneumoniae* mostram uma sequência relacionada a COG1109, que ocupa o espaço entre os genes *deoA* e *deoC* em *M. mycoides* e *M. florum* e é encontrado adjacente ao gene *cdd* no grupo *Pneumoniae*. No grupo *Pneumoniae*, *cdd*, *deoA*, *deoC* e *deoD* são adjacentes uns aos outros, uma organização espacial que foi apontada em outros estudos como uma indicação de que se trata de um operon. O produto gênico do operon pode ser ligado a fosfopentomutase. *U. parvum* possui os genes *deoA*, *deoC*, e *deoD*, enquanto *U. diversum* e *U. urealyticum* apenas possuem *deoC* e *deoD* em seu repertório genético, indicando que o operon pode não estar completo, mas ainda assim reter a sua funcionalidade. A rede IPP de *deoD* em *U. diversum* mostra uma interação desta enzima com *deoC*, *hpt* e *pdp*, participando desta forma da subrede do metabolismo de purinas e pirimidinas (**Figura 13**).

Figura 13. Subrede de deoD em *U. diversum*.



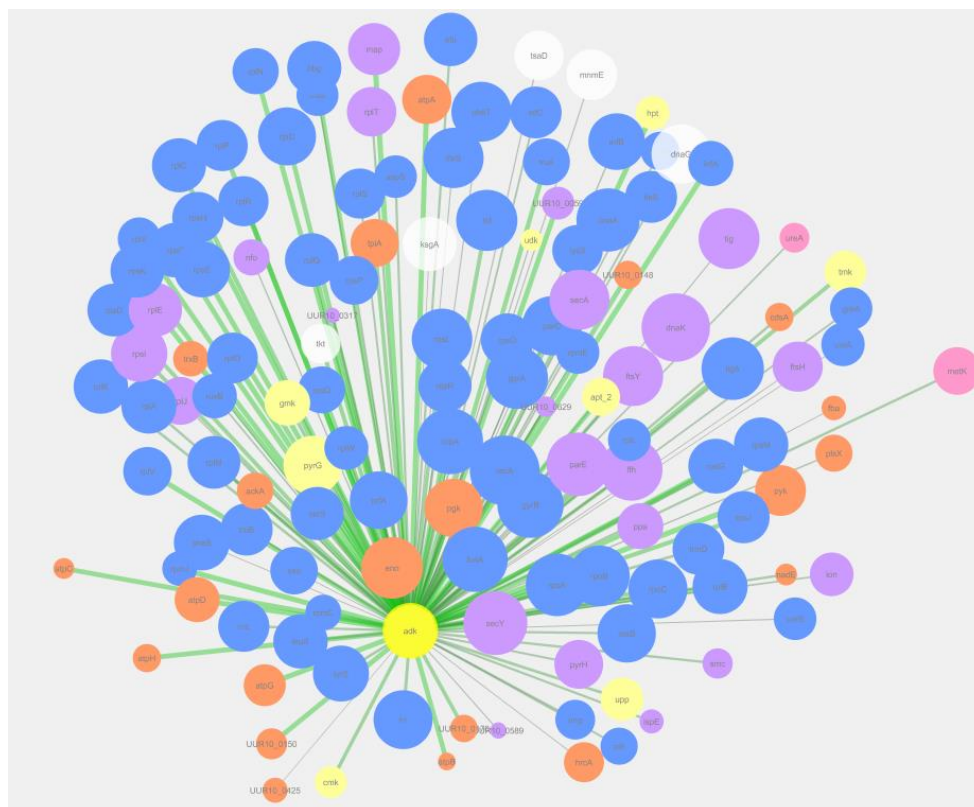
Estudos comparativos com espécies de *Mycoplasma* mostram que mesmo o gene comum *deoB* estando ausente em muitas espécies do grupo, ensaios foram capazes de detectar atividade enzimática, sugerindo uma atividade fosfopentomutase (CORDOVA et al., 2016) mesmo com a redução de enzimas para desempenhar esta função metabólica. Com essas informações, estudos apontam que a redundância funcional observada entre as vias metabólicas simplificadas dos Mollicutes pode ser devido a limitação de um genoma mínimo e oferece *insights* sobre a evolução deste grupo (MOMBACH et al. 2006).

A ausência de *deoB* em *Ureaplasma diversum* e a presença de *deoD* e *deoC*, além de sua interação com demais enzimas responsáveis pelo metabolismo de purinas e pirimidinas sugere que a função fosfopentomutase se mantém mesmo com a redução nas enzimas neste organismo, como ocorre em espécies de *Mycoplasma*.

NDK é uma enzima universal envolvida na síntese de nucleotídeos e é considerada essencial para síntese de DNA e RNA (CASTELLANOS et al., 2004). Ela catalisa a combinação ATP-dependente de ribo e deoxiribonucleo trifosfatos. A reação é catalisada por meio de uma fosfoenzima intermediária. Estudos determinaram que a estrutura de NDKs são extremamente conservadas de *E. coli* até humanos, mostrando uma semelhança de 43% entre estes dois grupos (ALMAULA et al., 1995). NDKs não estão presentes em Mollicutes, mas ensaios para atividade NDK em células que não tinham os genes NDK mostraram adenilato quinase como enzima complementária (LU; INOUE, 1996) na formação de nucleotídeos trifosfato. *U. diversum*, assim como *U. urealyticum* e

U. parvum não possui NDK em seu repertório gênico e, em vez disso, apresenta adenilato quinase (*adk*) (**Figura 14**).

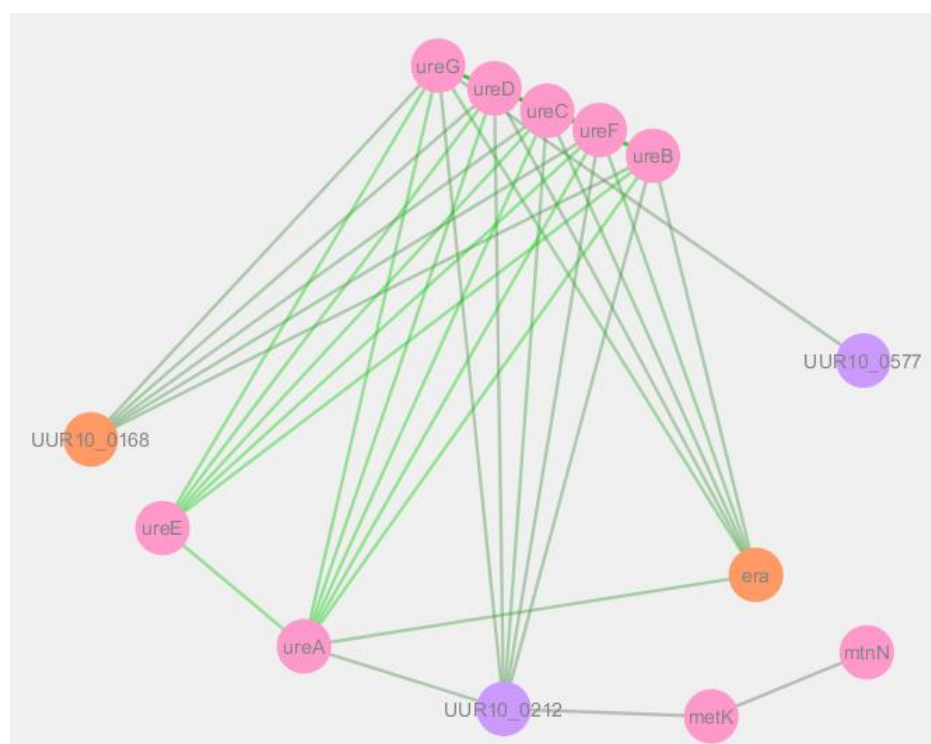
Figura 14. Subrede *adk* em *U. diversum*.



Adenilato quinase é responsável pela reação de transferência fosforibosil de AMP para ADP, mas também apresenta outras características interessantes. Em sua subrede em *U. diversum*, *adk* interage com diversas proteínas (em mais de 100 interações) envolvidas em muitas vias metabólicas diferentes, como as proteínas envolvidas no metabolismo de purinas e pirimidinas. Estudos experimentais (GENTRY et al., 1993) (JONG; CAMPBELL, 1984) mostrando atividade NDK realizada por esta enzima sugere que *adk* poderia estar envolvida no metabolismo de nucleotídeos nestas espécies. Ensaios experimentais com *Ureaplasma* poderiam testar se isto também é o caso em *U. diversum*.

O *cluster* da urease na IPP de *U. diversum* está entre as subredes de maior confiabilidade observadas neste estudo, possuindo índice de confiabilidade 0.9 entre as proteínas acessórias da urease (**Figura 15**).

Figura 15. A subrede da urease em *U. diversum*.



Análises mostraram que a hidrólise da ureia é responsável pela maioria (95%) da demanda de ATP em Ureaplasmas (BLANCHARD et al., 1988) (SMITH et al., 1993). Transportadores para ureia e níquel não foram encontrados em *U. diversum* (MARQUES et al., 2016) e o mesmo ocorre em outros Ureaplasmas (POLLACK, 2001). Ureia poderia ser adquirida diretamente do ambiente por difusão através da membrana celular, mas ainda não está claro se este processo é responsável pela importação de ureia em *U. diversum* (MARQUES et al., 2016).

4.4 Proteínas centrais da rede IPP de *U. diversum*

As proteínas com maior conexão dentro de uma rede IPP, ou seja, aquelas que interagem com a maior quantidade de proteínas, são chamadas de proteínas hub e centralizam as interações das redes em que estão inseridas. Proteínas *hub* promovem a integração de subredes e sua remoção da rede IPP pode resultar em colapso, uma vez que as proteínas restantes não são capazes de rearranjar suas interações suficientemente. Um outro resultado possível é que as proteínas da rede tenham uma conectividade alta o bastante para contornarem o problema aparente causado pela remoção de uma proteína hub. A remoção de proteínas envolvidas em processos cruciais para o organismo, como

o papel desempenhado pela urease na produção de energia para os Ureaplasmas, também ocasiona a morte da célula bacteriana.

Proteínas *hub* vem sendo identificadas principalmente em redes IPP que envolvem patogenicidade podendo, assim, indicar novos alvos terapêuticos para antibióticos ou drogas interventivas semelhantes.

Os dois parâmetros de centralidade mais utilizados para determinação da importância que um nodo possui para a rede IPP são o grau do nodo, que calcula o número de conexões que este possui, e a interseção do nodo que representa a quantidade de vias curtas da rede IPP que o cruzam (AZEVEDO; MOREIRA-FILHO, 2015). O grau (*k*) também chamado de conectividade de um nodo, como explicado anteriormente, se refere a quantidade de nodos aos quais o nodo sob análise está conectado, se tratando de uma medida geral de centralidade de um nodo na rede IPP. A interseção de um nodo, por sua vez, é uma relação entre a quantidade de vias mais curtas que cruzam este nodo, sendo desta forma uma medida de centralidade global que representa o fluxo de informações que passam por ele na rede IPP. Calculamos o grau e interseção de todos os nodos da rede IPP de *U. diversum* para identificação das proteínas hubs que agem como alicerce das subredes IPP e identificamos os processos celulares em que estão envolvidos através da pesquisa de artigos experimentais na literatura e na plataforma KEGG (KANEHISA; GOTO, 2000) (**Tabela 3**).

Tabela 3. Parâmetros de centralidade de proteínas *hub* da rede IPP de *U. diversum*.

Proteínas	Conectividade (k)	Interseção do nodo	Processo celular
recA	176	0.021	Replicação DNA
dnaK	172	0.019	Transcrição
topA	164	0.040	Replicação/Transcrição
			Replicação
ligA	154	0.022	Reparo DNA
			Replicação
UUR10_0577	69	0.023	Reparo DNA

É interessante notar que todas as proteínas com maior conexão na rede global de *U. diversum* estão relacionadas ao fluxo de informação gênica, desempenhando funções na replicação e transcrição do material genético. Entre estas, *recA*, *dnaK* e uma DNA

helicase (*UUR10_0577*) também estão envolvidas em resposta a estresse, ativação de vias S.O.S, recombinação homóloga e transferência horizontal de genes:

recA: possui função na replicação do DNA, além de estar envolvida no metabolismo de purinas e pirimidinas. *recA* consegue catalisar a hidrólise de ATP na presença de DNA de fita simples, a aquisição ATP-dependente de DNA de fita simples por DNA de fita dupla e a hibridização de DNAs de fita simples homólogos, que também depende de ATP. Desempenha funções de reparo do DNA e é acionada por respostas S.O.S., participando de recombinação homóloga.

dnaK: tem função de chaperona, é uma RNA polimerase DNA direcionada. Se trata de uma proteína ativada em situações de estresse bacteriano, envolvida em respostas S.O.S.

topA: DNA topoisomerase I possui função semelhante às helicases, liberando a tensão gerada pela torção do DNA proveniente da replicação e transcrição do DNA, quebrando pontes fosfodiéster e de hidrogênio. Realiza esta tarefa através da clivagem e posterior ligação de sítios alvo na fita de DNA.

ligA: uma DNA ligase dependente de NAD essencial para replicação e reparo do DNA. *ligA* catalisa a formação de ligações fosfodiéster entre fitas de DNA utilizando NAD como uma coenzima que provê a fonte de energia necessária para a reação.

UUR10_0577: helicase dependente de ATP, também conhecida como *PcrA*. Alivia a tensão causada pela torção do DNA ao longo dos processos de replicação e transcrição gênicas. Regula a frequência de recombinação realizada por *recA* e está comprometida com vias de funcionamento da transferência horizontal de genes, especificamente com conjugação bacteriana.

Proteínas comprometidas com vias de reparo de DNA danificado são potenciais alvos terapêuticos e isto se aplica também para *recA*, apontada como alvo de antibióticos em diferentes estudos. O papel de *recA* na recombinação homóloga o posicionou entre os principais componentes envolvidos na captação de material genético exógeno, em especial nos processos de conjugação e transformação, conhecidos mecanismos de transferência horizontal de genes (COX, 1991). O processo de transferência de DNA mediada por *recA* exige a participação de diversas proteínas, cerca de 40 em *Bacillus subtilis* (SAITO; AKAMATSU, 2006), o que pode ser a explicação para a alta conectividade deste nodo na rede IPP de *U. diversum*. Em *B. subtilis*, *recA* interage com

DNA exógeno de fita simples, ssDNA (do inglês *single-stranded DNA*), formando nucleofilamentos capazes de reconhecer sítios de recombinação no cromossomo bacteriano, onde promove a troca de informação gênica (Prentiss et al, 2015). Este processo foi observado em outras espécies bacterianas, incluindo *Neisseria gonorrhoeae*, *Hemophilus influenzae*, *Streptococcus pneumoniae*, *Streptococcus mutans* e *Helicobacter pylori*, sugerindo que o mecanismo de incorporação de DNA exógeno adquirido via transferência horizontal de genes mediado por *recA* é de grande importância para acréscimo de variedade gênica nestes grupos (Michod et al, 2008). Shao et al (2017) determinaram que a atividade de *recA* aumenta as taxas de recombinação homóloga em *Mycoplasma hyorhinae* e considerando a alta conectividade de *recA* na rede IPP de *U. diversum*, algo parecido deve ocorrer em Ureaplasmas e outros micoplasmas. *PcrA* é uma helicase responsável pelo alívio da tensão nas fitas de DNA gerada pelos processos de replicação e transcrição da informação gênica. Estudos indicam que helicases, incluindo *PcrA*, podem regular a taxa de recombinação gênica em *B. subtilis* (EHRlich; PETIT, 2002) promovida por proteínas como *recA*. Um sistema de regulação semelhante pode ocorrer em *U. diversum*, em que a recombinação homóloga promovida por *recA* seria controlada por *PcrA*.

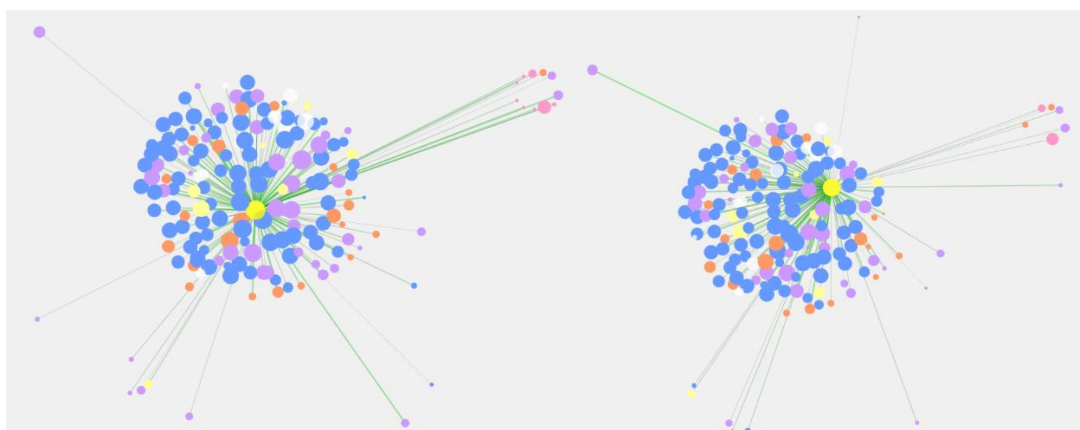
Proteínas que integram os sistemas S.O.S. de resposta a estresse aumentam as taxas de mutação em populações bacterianas, uma vez que a diversidade gênica pode ser uma vantagem e possibilitar a aquisição de genes de resistência em subpopulações procariotas (FOSTER, 2005). A mutagênese promovida via estímulos de estresse ocorre pela ação de proteínas como *dnaK* (KÜLTZ, 2005) uma das proteínas centrais identificadas ao longo das análises da rede IPP de *U. diversum*. *DnaK* foi correlacionada a resposta de estresse causada por temperaturas altas e baixas, além de substâncias ácidas (JUNJIE et al, 2014; ZHAO et al, 2017), cuja presença aumenta as taxas de expressão deste gene.

As proteínas centrais da rede IPP de *U. diversum* estão, desta forma, envolvidas diretamente na geração de diversidade gênica por duas principais vias: 1) pela recombinação homóloga de sequências de DNA exógenas adquiridas via transferência horizontal de genes, especificamente pelos mecanismos de transformação e conjugação e 2) causando um aumento na taxa de mutação via sistemas de resposta a estresse, como é o caso de *dnaK*. Estas informações sugerem que o repertório gênico de *U. diversum* e as proteínas mais conectadas de sua rede IPP favorecem um fluxo gênico entre o

cromossomo bacteriano e o ambiente colonizado por estes Ureaplasmas e suas comunidades bacterianas correlacionadas.

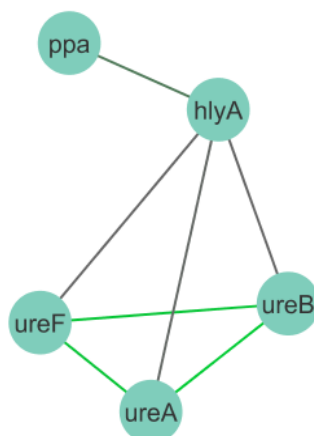
Cinco proteínas concentram 8% (839) das interações da rede IPP global de *U. diversum*: *recA* (176 interações), *dnaK* (172 interações), *gyrB* (165 interações), *topA* (164 interações) e *fusA* (162 interações). *RecA* e *dnaK* estão envolvidas na replicação e transcrição gênica e poderiam servir como possíveis alvos terapêuticos contra infecções por *U. diversum*. Isolamos as subredes de *recA* e *dnaK* para uma melhor visualização de seu tamanho e composição (**Figura 16**). As redes IPP preditas de *recA* e *dnaK* englobam apenas interações diretas, observando-se uma grande sobreposição de interações entre as duas proteínas, o que pode se dever a suas funções semelhantes e complementares.

Figura 16. Subrede *recA* e *dnaK*, as maiores proteínas hubs da rede IPP de *U. diversum*.



Analizamos subredes para identificação de possíveis alvos terapêuticos contra a infecção por *U. diversum*. Marques et al. (2016) identificaram uma proteína semelhante a hemolisina (*hlyA*) no genoma de *U. diversum*, o que nos estimulou a construir sua rede IPP (**Figura 17**). A hemolisina é uma importante integrante do repertório virulento durante o estabelecimento de infecções suprindo a demanda por ferro, um fator limitante para crescimento bacteriano. A hemolisina de *U. diversum* possui 63.1% de identidade em sua sequência proteica com Ureaplasmas humanos e sua semelhança com *hlyA* de *U. parvum* sugere que possua propriedades citotóxicas, além de hemolíticas (MARQUES *et al.*, 2016).

Figura 17. Subrede de *hlyA* (hemolisina) em *U. diversum*.



A proteína *hlyA* interage com as subunidades e proteínas acessórias da urease (*ureA*, *ureB* e *ureF*) e uma pirofosfatase inorgânica (*ppa*).

hlyA: hemolisina alfa, causa ruptura de eritrócitos através do estabelecimento de poros na superfície celular.

ureA: subunidade gama da urease.

ureB: subunidade beta da urease.

ureF: proteína acessória da urease necessária para sua maturação.

ppa: pirofosfatase inorgânica, catalisa a hidrólise de uma molécula de pirofosfato em dois íons de fosfato inorgânico. Envolvida nos processos de replicação de DNA, metabolismo de lipídeos e outros processos bioquímicos.

Esta pode ser uma subrede interessante para alvos terapêuticos, uma vez que se atingiria ao mesmo tempo dois fatores de virulência de *U. diversum*: hemolisina e urease. Além disto, o suprimento energético bacteriano também seria afetado através do ataque a urease, uma das principais responsáveis por suprir a demanda energética em Ureaplasmas.

4.5 Transferência horizontal em *Ureaplasma*

Transferência horizontal de genes vem sendo apontada como possível explicação para as características metabólicas e filogenéticas distintas entre os Mollicutes. Jain et al.

(1999) já apontavam que a quantidade abundante de transferência horizontal entre procariotos acontece em maior frequência em genes operacionais (em inglês *house-keeping genes*) do que em genes envolvidos na regulação da expressão gênica. O mecanismo de evolução degenerativa dos Mollicutes não contribui para a adição de novos genes (CORDOVA et al., 2016), sugerindo que a transferência horizontal de genes pode ter um papel importante na aquisição de genes neste grupo.

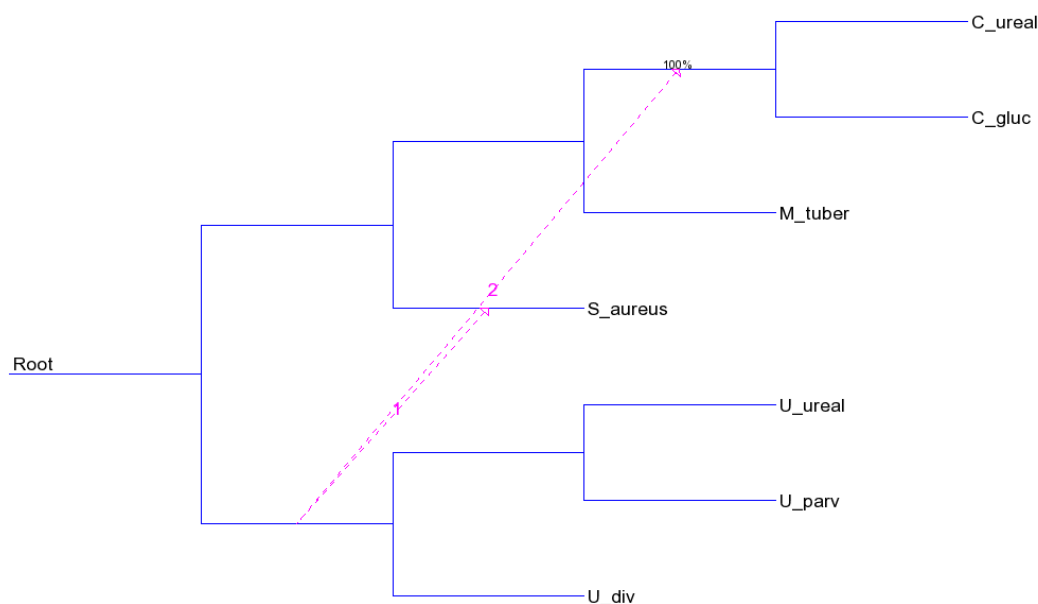
Transferência horizontal de genes entre *Ureaplasma* já foi observada em ensaios experimentais entre diferentes populações (XIAO et al., 2011). García-Castillo et al. (2008) demonstraram que a aquisição da capacidade da produção de biofilmes em *Ureaplasma* ocorreu através de transferência horizontal de genes de outros grupos bacterianos e estudos anteriores reportaram transferência horizontal de genes ocorrendo entre espécies que habitavam o mesmo nicho ecológico (SIRAND-PUGNET; LARTIGUE; et al., 2007). Teachman et al. (2002) também mostraram que populações de *M. pulmonis* compartilham genes horizontalmente, sugerindo que a transferência horizontal de genes parece ser mais prevalente em alguns Mollicutes que em outros. Como exemplo podemos citar que 18% do genoma de *Mycoplasma agalactiae* foi adquirido do grupo *Mycooides* por este mecanismo (DORDET-FRISONI et al., 2014). Ainda que os mecanismos exatos pelos quais a transferência de genes ocorre entre Mollicutes não estejam claros, este processo parece não apenas garantir a aquisição gênica e troca entre Mollicutes (CORDOVA et al., 2016), como também aprimora a maquinaria metabólica dessas espécies para uma maior sobrevivência contra mecanismos de defesa dos hospedeiros e poderia também viabilizar a colonização de novos nichos (SIRAND-PUGNET; LARTIGUE et al., 2007).

Os resultados da detecção de eventos de transferência horizontal de genes utilizando a rede IPP de *U. diversum* mostraram alguns pontos de troca gênica nas subredes da urease e do metabolismo de purinas e pirimidinas.

Durante a análise e interpretação de eventos de transferência horizontal, a observação dos nichos ocupados pelas espécies envolvidas se mostra uma importante ferramenta que auxilia na contextualização ecológica dos resultados obtidos. Espécies que ocupam o mesmo nicho, seja o ambiente abiótico ou compartilhando o mesmo hospedeiro, estão em contato direto e contribuindo para o fluxo gênico local. Nas nossas análises dois eventos de transferência horizontal foram observados na via metabólica da

urease (**Figura 18**), mostrando um fluxo do operon urease entre *Ureaplasma* e *Corynebacterium*, *Ureaplasma* e *Staphylococcus aureus*.

Figura 18. Eventos de transferência horizontal de genes em *Ureaplasma*, *S. aureus* e *Corynebacterium* na subrede da Urease.



A frequência e impacto da transferência horizontal de genes são altos no gênero *Corynebacterium*, especialmente quando se trata de patogenicidade e fatores de virulência (OLIVEIRA et al., 2017). A urease, como descrito anteriormente, pode suprir a demanda energética do organismo através da hidrólise da ureia, mas também é um conhecido fator de virulência em bactérias patogênicas. Estudos anteriores observaram a transferência horizontal de genes em *Corynebacterium* ocorrendo em ilhas patogênicas relacionadas a fatores clássicos de virulência, como adesinas, toxinas secretadas e proteínas envolvidas na absorção de ferro do hospedeiro (RUIZ et al., 2011). As espécies de *Corynebacterium* analisadas neste estudo foram *C. glucuronolyticum* e *C. urealyticum*. As redes 16S geradas como controle podem ser visualizadas no apêndice deste trabalho.

Corynebacterium glucuronolyticum é um patógeno oportunista conhecido por causar infecções na região urogenital de humanos sob certas condições clínicas (GHERARDI et al., 2015) e é também um agente causador de infecções urológicas e distúrbios reprodutivos em porcos (DEVRIESE et al., 2000) e ovinos (TAKAHASHI et al., 1997). *Corynebacterium urealyticum* é uma espécie comumente isolada de amostras retiradas de infecções urológicas e estudos sugerem que nestes casos ocorra coinfeção

por *Ureaplasma urealyticum* (TRINCHIERI, 2014), ampliando o quadro de infecção e promovendo a formação de cálculos renais através da ação da urease. Os genes ureABC codificam as subunidades estruturais da urease e os genes ureEFGD carregam a informação para produção de proteínas acessórias (SALEM et al., 2015). A hidrólise da ureia causa uma hiperformação de amônia, que por sua vez leva a alcalinização do ambiente previamente descrito ocasionando a formação de cristais de sal, além de facilitar infecções urogenitais.

A produção de urease é um fator determinante que ao longo da história evolutiva possibilitou a colonização do trato urogenital por *Ureaplasma* e *Corynebacterium*, sendo ao mesmo tempo considerado um fator de virulência e uma forma de suprir a demanda energética celular destes organismos. Ocupando o mesmo nicho em hospedeiros humanos e outros animais, é possível que os genes que codificam urease tenham sido compartilhados por *Ureaplasma* e adquiridos por *Corynebacterium*, como nossos dados sugerem. Os meios utilizados para tal transferência ainda não estão claros, ainda que populações de *Corynebacterium* tenham sido caracterizadas como possuindo um alto número de transposons e outros elementos móveis em seus genomas (GUIMARÃES et al. 2015), sugerindo que transferência horizontal de genes possa ter um papel importante durante a evolução e o meio utilizado para aquisição de resistência múltipla a antibióticos observada neste gênero (TAUCH et al., 2008).

Staphylococcus aureus está entre os patógenos humanos mais estudados, causando uma variedade de infecções de importância clínica (TONG et al., 2015). Como patógeno oportunista, *S. aureus* pode causar infecções na pele e tecidos moles, como também bacteremia e endocardite em casos mais graves. Sua relação com infecções urogenitais foi documentada e ainda que não esteja entre as espécies mais prevalentes nestes quadros clínicos (TONG et al., 2015) *S. aureus*, assim como espécies de *Corynebacterium* e *Ureaplasma* habita o nicho urogenital, podendo ser esta a explicação para o evento de transferência horizontal apontada pelas nossas análises. Diferente da infecção por *Ureaplasma* e *Corynebacterium*, as infecções urogenitais causadas por *S. aureus* são mais comuns em pacientes hospitalares imunossuprimidos portadores de cateteres urinários, que facilitam a colonização e disseminação bacteriana. Estudos clássicos mostram que a urease de *Staphylococcus* possui propriedades bioquímicas semelhantes a outros grupos bacterianos e que existe uma semelhança topológica no operon urease entre bactérias produtoras desta enzima (GATERMANN; MARRE, 1989).

A infecção por *S. aureus* em bovinos vem sendo apontada como um processo de importante relevância econômica na pecuária, uma vez que o rebanho e qualidade do produto são afetados (GUINANE et al., 2011). A discussão sobre como clones específicos de *S. aureus* são capazes de infectar hospedeiros animais como aves e bovinos, mas não são isolados de hospedeiros humanos sugere uma adaptação necessária a novos nichos (hospedeiros multicelulares zoóticos) e que a aquisição de genes cruciais para colonização é necessária para estas populações (GUINANE et al., 2011). Tal adaptação e aquisição genética são realizadas principalmente através da transferência horizontal de genes via elementos genéticos móveis, que foram apontados como cruciais para a adaptação de *S. aureus* colonizadoras de humanos para infecção de aves domésticas (LOWDER et al., 2009) e bovinos (HERRON-OLSON et al., 2007).

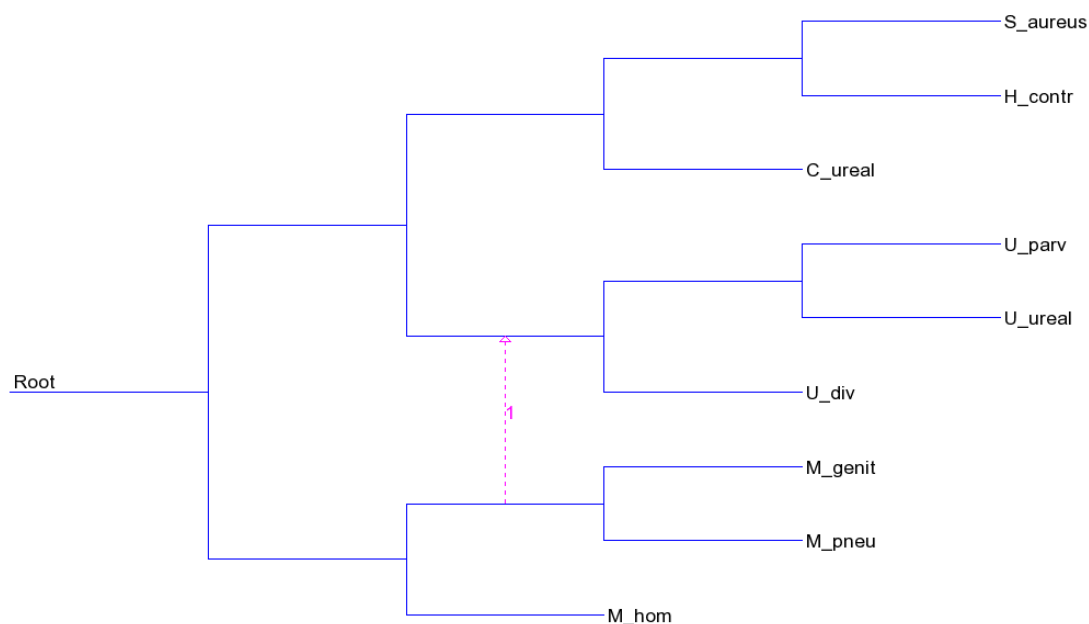
Guinane et al. (2011) estimaram através de uma abordagem que mescla genética de populações, genômica comparativa e análise funcional que a transferência horizontal de genes que possibilitaram a *S. aureus* a colonização de animais domésticos como aves, pequenos e grandes ruminantes ocorreu no início da domesticação destes grupos, cerca de 11 mil anos atrás (ZEDER, 2008). Os autores discutem que o “pulo” de populações de *S. aureus* capazes de colonizar diferentes ruminantes e aves, também referidas como serotipos (do inglês *serovars*) seguiu com o processo de adaptação genética possibilitado por diversificação alélica, perda gênica e aquisição de elementos genéticos móveis (GUINANE et al., 2011) que em *S. aureus* chegam a comprometer entre 15 a 20% do genoma (LINDSAY, 2014). Quando comparado com populações de *S. aureus* que colonizam humanos, os serovars de origem ruminante ou de aves apresentavam um acréscimo de fatores de virulência, adesinas e maquinaria metabólica que possibilita a sobrevivência no hospedeiro. Estes dados sugerem que ao longo de sua história evolutiva, *S. aureus* teve que se adaptar a novos hospedeiros, o que também é verdade para espécies de *Ureaplasma* que possuem populações que infectam exclusivamente animais domésticos.

Purinas e pirimidinas podem ser vistas como os tijolos que constroem o DNA e o RNA sendo, portanto, essenciais para sobrevivência e proliferação celular, o que significa que a demanda por nucleotídeos tem que ser atingida pelos organismos seja através de biossíntese *de novo* ou via importação de precursores. As vias de síntese de purinas e pirimidinas estão incompletas ou ausentes em muitos patógenos que, em vez de gastar energia e maquinaria celular para biossíntese dessas bases, garantem os nucleotídeos

necessários para seu metabolismo do ambiente em que estão inseridos. Em espécies como *Cryptosporidium parvum* enzimas alternativas que auxiliam na síntese de purinas e pirimidinas a partir de precursores importados do hospedeiro já foram identificadas (STRIEPEN et al., 2004) e sua origem foi apontada como resultante do processo de transferência horizontal de genes. *U. diversum* conta com três destas enzimas alternativas que auxiliam na recuperação de nucleotídeos: uridina quinase (*udk*), uracil fosforibosil transferase (*upp*) e timidina quinase (*tdk*), cujas redes IPP foram tratadas nas seções anteriores. Em *C. parvum*, a aquisição de timidina quinase foi realizada via transferência horizontal de uma proteobacteria (STRIEPEN et al., 2004). Nesta espécie, adenosina é a fonte única para síntese de purinas.

Nossos dados sugerem dois eventos de transferência horizontal nas redes de biossíntese de purinas e pirimidinas (**Figura 19, Figura 20**). Na rede de purinas, análises mostraram a transferência horizontal ocorrendo de *Mycoplasma* para *Ureaplasma*, gêneros da classe Mollicutes que compartilham o mesmo nicho ecológico em hospedeiros multicelulares.

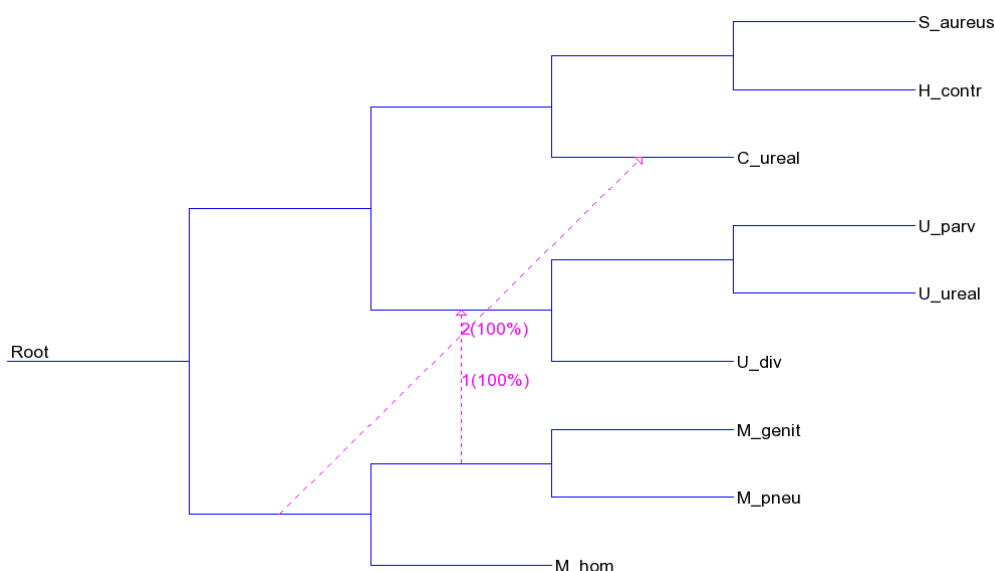
Figura 19. Eventos de transferência horizontal de genes em *Mycoplasma* e *Ureaplasma* na subrede do metabolismo de purinas.



Ureaplasma e *Mycoplasma* podem colonizar o trato genital de humanos e outros animais como bovinos, caprinos e ovinos, causando infecções em situações de um trauma biológico ou físico preexistente. As coinfeções por *Mycoplasma* e *Ureaplasma* têm uma grande importância na área clínica, sendo apontadas como patógenos causadores de infecções que podem significar riscos ao longo do desenvolvimento de gravidez e causar abortos, aceleração do parto e aumento dos casos de cesarianas (PERNI et al., 2004) (CAPOCCIA et al., 2013).

A transferência horizontal de genes tem um importante papel na evolução dos Mollicutes, especialmente quando contextualizada com o evento de redução genômica sofrida por este grupo. Desta forma, é esperado que exista um fluxo de informação gênica entre os integrantes dos gêneros desta taxa e que a alta frequência de transferência horizontal de genes não se limite a relações com grupos mais externos e filogeneticamente distantes. É possível que o contato direto entre *Mycoplasma* e *Ureaplasma*, ocupando o nicho do trato urogenital de seus hospedeiros, tenha contribuído para a transferência de genes envolvidos na biossíntese, em vias alternativas e de resgate do metabolismo de purinas.

Figura 20. Eventos de transferência horizontal de genes em *Mycoplasma* e *Corynebacterium* na subrede do metabolismo de pirimidinas.

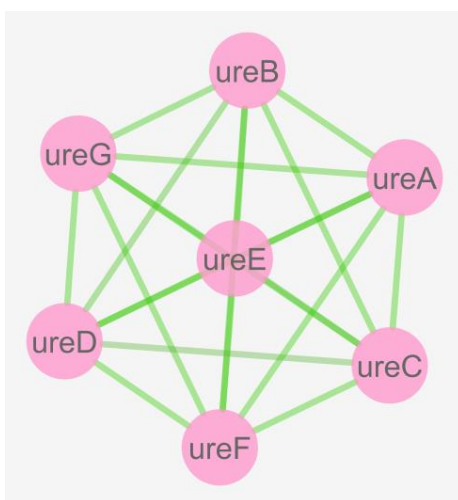


Nas redes de interação do metabolismo de pirimidinas, dois eventos de transferência horizontal foram apontados em nossas análises (**Figura 20**). O primeiro é de *Mycoplasma* para *Ureaplasma*, que pode ser explicado pelo grande contato entre esses dois grupos representado pelos seus estilos de vida e ecologia semelhantes. O segundo foi de *Mycoplasma* para *Corynebacterium*, um gênero bacteriano conhecido por causar infecções em tecidos e órgãos de ambientes mucosos e pelo comportamento de patógeno oportunista, participando em infecções já instaladas. Estudos identificando *Corynebacterium* em amostras coletadas de infecções no trato genital humano (ÅKERLUND; MÅRDH, 1974) e de pequenos ruminantes (CONTRERAS et al., 1995) sugerem que a transferência horizontal de genes entre estes grupos também tenha ocorrido pela ocupação do mesmo nicho ecológico.

Um fluxo contínuo de informação gênica entre *Mycoplasma*, *Ureaplasma* e *Corynebacterium* poderia ser facilitado pelo contato direto entre estes grupos no trato genital dos hospedeiros colonizados, uma vez que estudos sugerem a prevalência dos gêneros em infecções do trato reprodutivo (RUSSO et al., 1981). Nossos dados sugerem, portanto, que estes patógenos oportunistas compartilham seu repertório genético entre si para otimizar o metabolismo de purinas e pirimidinas, aumentando as chances de sobrevivência e proliferação no trato genital de seus hospedeiros.

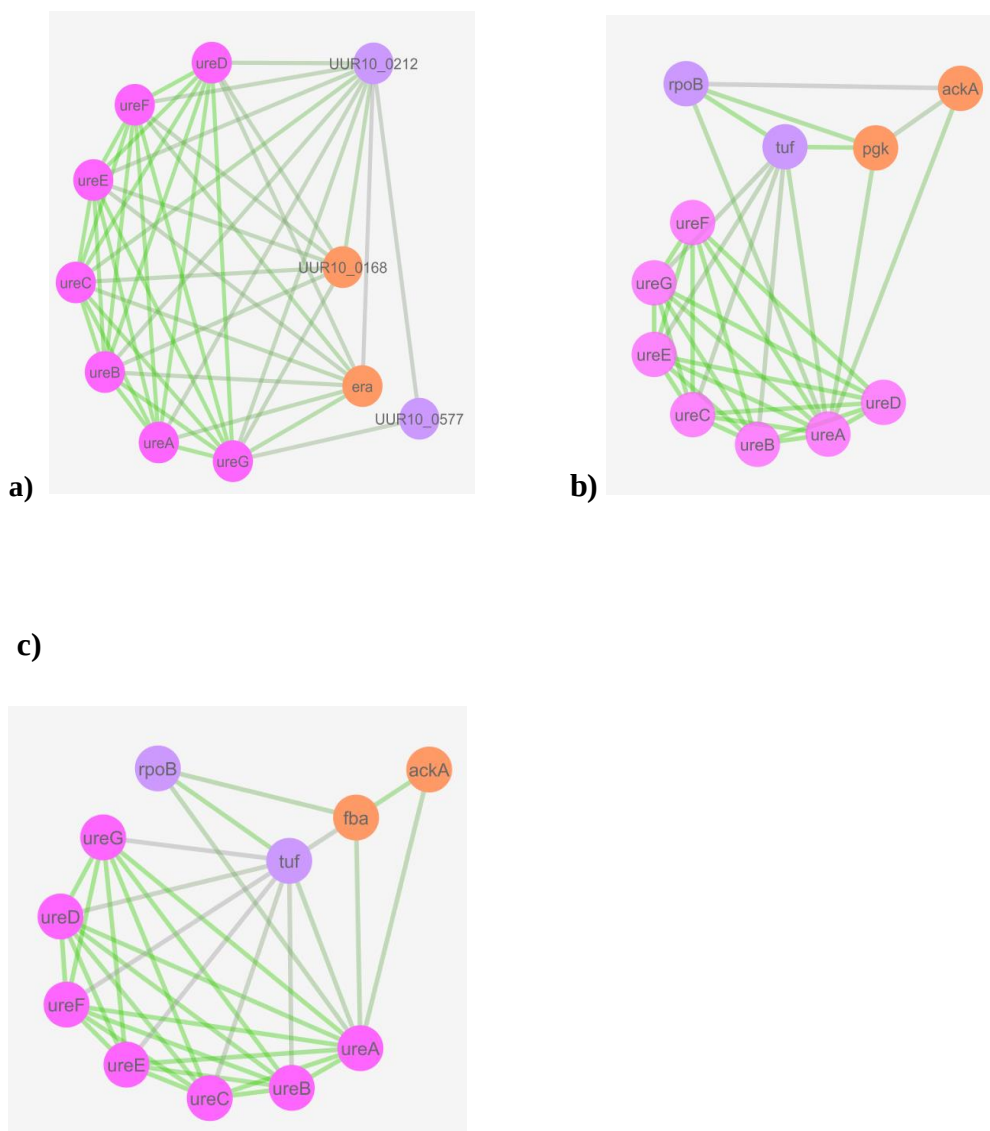
A conformação da rede IPP das proteínas da urease mantém-se igual entre as espécies analisadas, havendo uma mudança nas proteínas com as quais estas interagem. Todas as espécies analisadas possuem as subunidades estruturais da urease (*ureA*, *ureB*, *ureC*) e as proteínas acessórias da urease necessárias para seu funcionamento (*ureE*, *ureF*, *ureG*) em sua rede IPP, como pode ser observado na rede gerada a partir de suas interseções (**Figura 21**). O índice de confiabilidade entre as proteínas que compõem as unidades estruturais e acessórias da urease foi enquadrado como altíssimo (>0.9) em todas as espécies analisadas, o que é representado pela coloração verde clara atribuída às interações.

Figura 21. Interseção das redes de urease de todas as espécies analisadas para transferência horizontal.



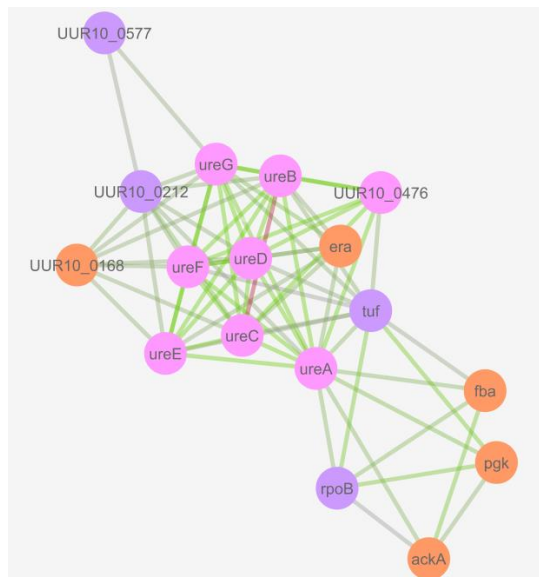
Comparando a subrede de urease de *U. diversum* com a de *U. urealyticum* e *U. parvum* (**Figura 22**) é possível observar uma alta semelhança entre as interações e posições ocupadas pelas proteínas na rede IPP. A conformação da rede em si é semelhante, observando-se parâmetros de conectividade próximos entre as espécies de *Ureaplasma*, estando a diferença nas proteínas que compõem as redes individuais.

Figura 22. Subredes da urease em a) *U. diversum* b) *U. urealyticum* e c) *U. parvum*.



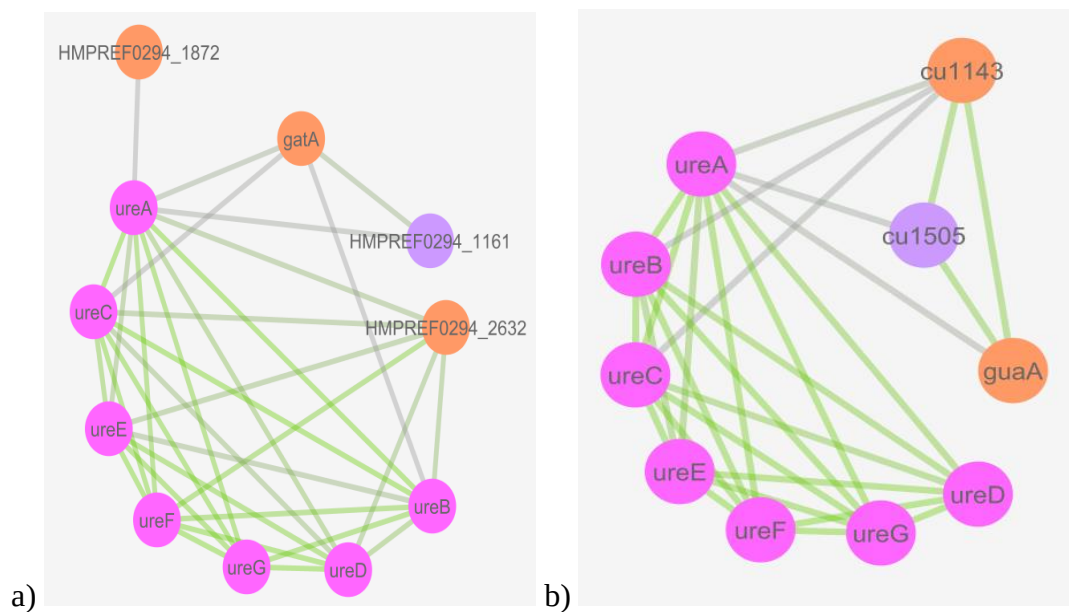
Em *U. diversum* proteínas hipotéticas pouco caracterizadas ocupam estes espaços (UUR10_0212, UUR10_0168 e UUR10_0577), enquanto que em *U. urealyticum* e *U. parvum* as mesmas posições são preenchidas (incluindo interações com as proteínas acessórias da urease) por RNA polimerase DNA direcionada (*rpoB*), frutose bifosfato aldolase (*fba*), acetato quinase (*ackA*), sugerindo que as funções das proteínas hipotéticas de *U. diversum* podem se aproximar das desempenhadas por estas proteínas. Quando colocadas na mesma rede, estas proteínas continuam interagindo com o restante do repertório proteico de *Ureaplasma*, indicando que a função biológica configurada pelas interações entre estas proteínas deve ser semelhante (**Figura 23**).

Figura 23. Repertório proteico da urease em *Ureaplasma*, construído da junção das redes individuais de *U. diversum*, *U. parvum* e *U. urealyticum*.



O evento de transferência horizontal na subrede da urease indicou o processo acontecendo também de *Ureaplasma* para *Corynebacterium*, o que nos estimulou a analisar as subredes da urease neste grupo com maior atenção (**Figura 24**).

Figura 24. Subrede da Urease em a) *C. glucurolyticum* e b) *C. urealyticum*.

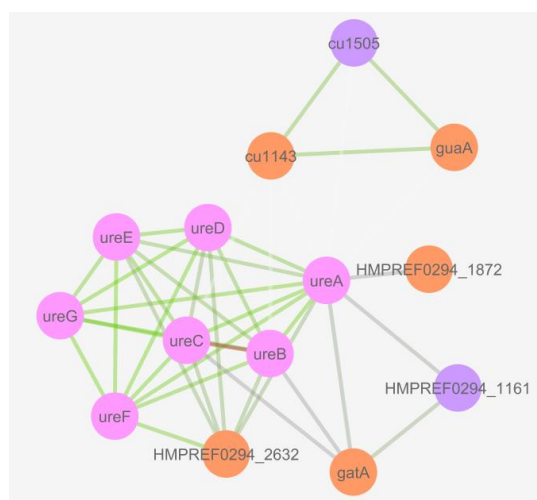


C. urealyticum possui uma DNA polimerase na rede (*cu1143*), gmp sintase (*guaA*) e uma carboxilato desidrogenase (*cu1505*), enquanto estas posições em *C. glucurolyticum* são ocupadas por proteínas hipotéticas e *gatA*, um fator transcricional. Assim como em

U. diversum, é possível que as proteínas hipotéticas de *C. glucurolyticum* desempenhem funções semelhantes às caracterizadas na rede de *C. urealyticum*.

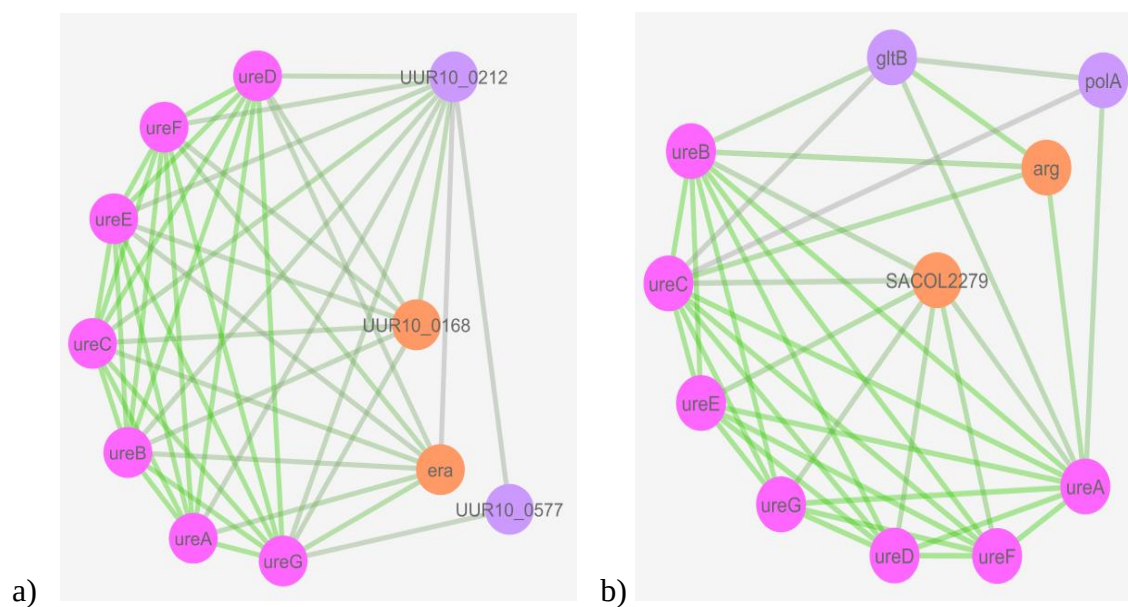
Unindo as redes de *C. glucurolyticum* e *C. urealyticum*, é possível construir a subrede de Urease de *Corynebacterium* (**Figura 25**).

Figura 25. Repertório proteico de *C. urealyticum* e *C. glucurolyticum*, referindo-se a subrede da urease.



O segundo evento de transferência horizontal na rede IPP da urease foi identificado como ocorrendo de *Ureaplasma* para *S. aureus*, o que nos estimulou a comparar a conformação da rede IPP da urease nas duas espécies (**Figura 26**).

Figura 26. Subrede urease em a) *U. diversum* e b) *S. aureus*.



É interessante observar que as interações entre as redes se assemelham, diferenciando-se nas proteínas que interagem com as proteínas estruturais e acessórias da urease. A comparação dessas redes e seus parâmetros topológicos associados (**Tabela 4**) sugerem que a subrede da urease é uma das mais conservadas em estrutura e parâmetros entre os grupos estudados, o que pode ser explicado pelo compartilhamento e fluxo gênico horizontal entre estas espécies que ocupam o nicho urogenital em humanos e outros mamíferos. A urease se configura, portanto, como uma proteína do nicho urogenital essencial na produção de energia pelas espécies estudadas, além de agir também como fator de virulência durante o estabelecimento da infecção.

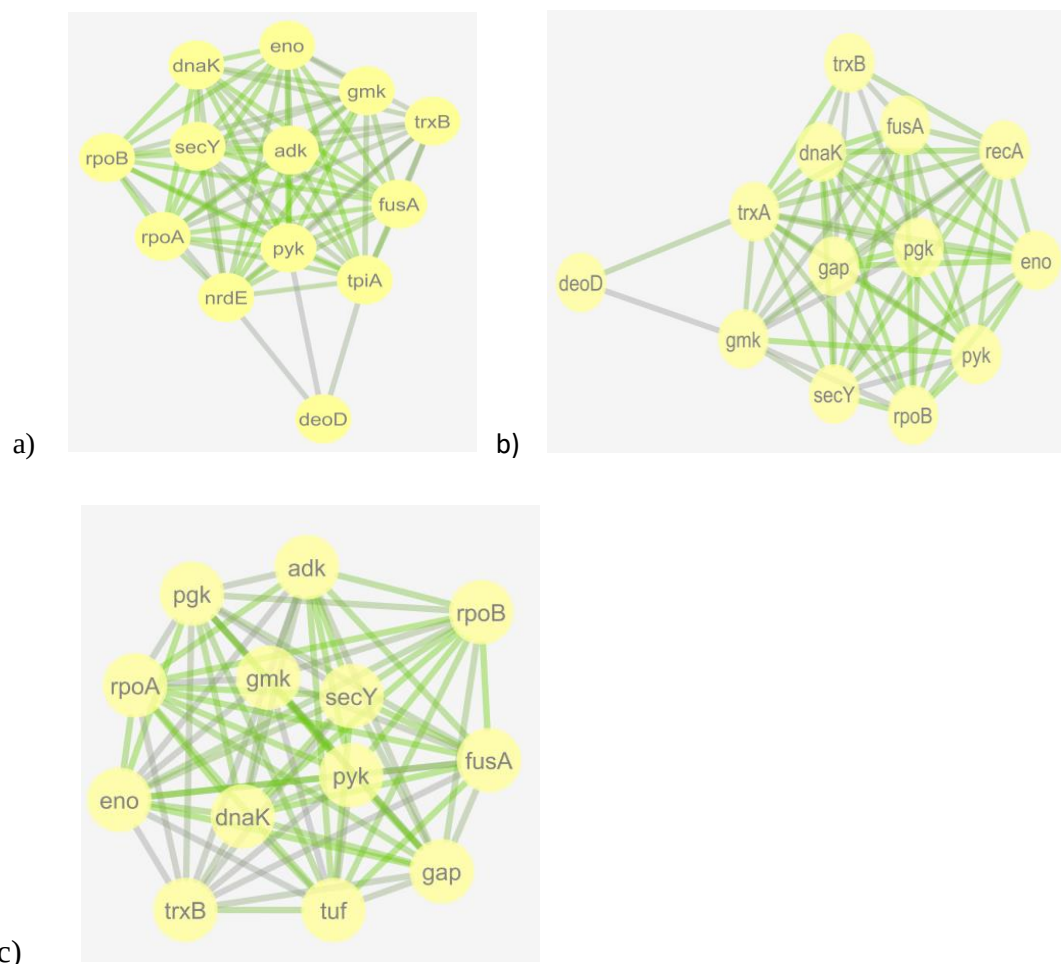
Tabela 4. Parâmetros das redes IPP da urease das espécies investigadas.

	Coef. de agrupamento	Centralização da rede	Comprimento caract. de via
<i>U. diversum</i>	0.934	0.222	1.182
<i>U. urealyticum</i>	0.858	0.467	1.382
<i>U. parvum</i>	0.899	0.444	1.364
<i>C. urealyticum</i>	0.877	0.444	1.356
<i>C. glurucolyticum</i>	0.783	0.489	1.40
<i>S. aureus</i>	0.888	0.378	1.309

Analisando-se a os coeficientes de agrupamento da rede IPP da urease entre as espécies estudadas, é possível observar índices de coeficiente de agrupamento, centralização da rede e comprimento característico de via muito semelhantes, o que indica que a semelhança entre estas redes IPP não se limita a sua conformação proteica, mas também se reflete em seus principais parâmetros de centralidade. O estilo de vida semelhante observado nessas espécies pode contextualizar estes dados biologicamente.

Como discutido anteriormente, nossas análises apontam ainda para um evento de transferência horizontal ocorrido ao longo da evolução de *Mycoplasma* e *Ureaplasma* em proteínas relacionadas ao metabolismo de purinas, o que aprofundou a observação destas subredes e seus parâmetros. Analisando as subredes do metabolismo de purinas nas espécies comparativas, focamos primeiramente nas subredes relacionadas ao metabolismo de purinas nas espécies de *Mycoplasma* (**Figura 27**).

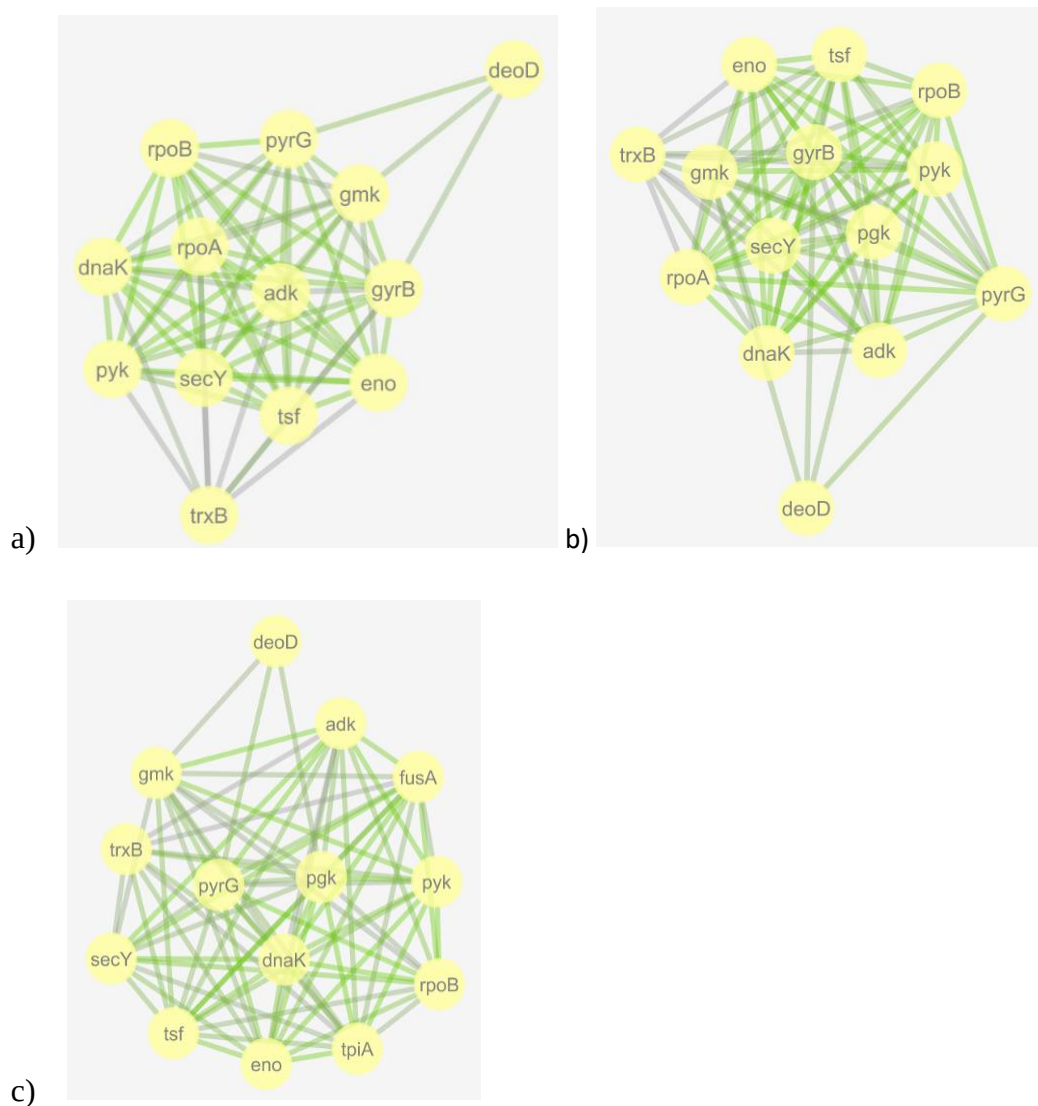
Figura 27. Subrede do metabolismo de purinas em a) *M. genitalium* e b) *M. pneumoniae* e c) *M. hominis*.



Uma das principais diferenças na composição das redes IPP das espécies de *Mycoplasma* estudadas é que *M. hominis* (o único representante do grupo *hominis*) não possui o gene *deoD* apresentado por *M. genitalium* e *M. pneumoniae*, cujas conexões são assumidas por outras proteínas.

As espécies de *Ureaplasma* estudadas (*U. diversum*, *U. urealyticum*, *U. parvum*) apresentaram redes de metabolismo de purinas semelhantes tanto em estrutura, quanto em composição, como mostrado pela representação de suas redes IPP (**Figura 28**).

Figura 28. Redes do metabolismo de purinas em a) *U. diversum*, b) *U. urealyticum* e c) *U. parvum*.



Mycoplasma e *Ureaplasma* compartilham algumas proteínas tratando-se da subrede IPP do metabolismo de purinas. As proteínas compartilhadas foram usadas para reconstruir uma rede IPP, demonstrando que as proteínas observadas nestes dois gêneros da classe Mollicutes incluem enzimas bastante conectadas, além de *trxB*, uma redutase (**Figura 29**). Seguindo o exemplo feito na análise das subredes IPP do metabolismo de urease, também realizamos uma análise dos principais parâmetros topológicos das subredes do metabolismo de purinas nos grupos envolvidos nos eventos de transferência horizontal (**Tabela 5**). Novamente, os parâmetros topológicos entre as redes analisadas se mostraram semelhantes.

Figura 29. Interseção do metabolismo de purinas em *Mycoplasma* e *Ureaplasma*.

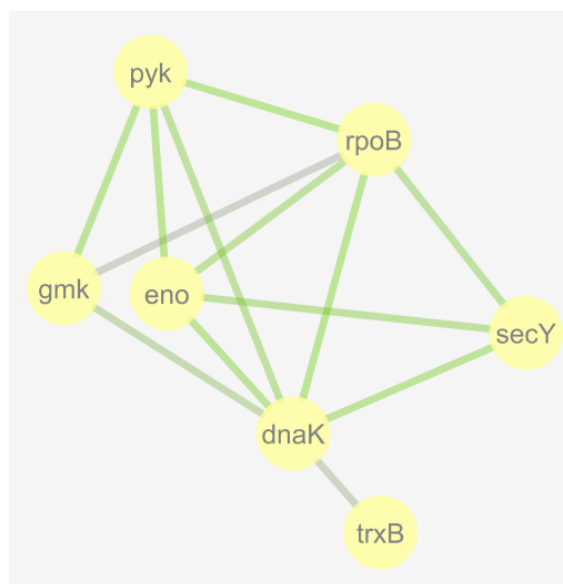


Tabela 5. Parâmetros das redes IPP do metabolismo de purinas das espécies investigadas.

	Coef. de agrupamento	Centralização da rede	Compr. de via característico
<i>U. diversum</i>	0.934	0.182	1.154
<i>U. urealyticum</i>	0.938	0.154	1.132
<i>U. parvum</i>	0.928	0.179	1.154
<i>M. hominis</i>	0.954	0.061	1.051
<i>M. pneumoniae</i>	0.921	0.242	1.205
<i>M. genitalium</i>	0.945	0.167	1.141

A análise da subrede IPP do metabolismo de pirimidinas sugere dois eventos de transferência horizontal: o primeiro de *Mycoplasma* para *Ureaplasma* e o segundo de *Mycoplasma* para *Corynebacterium*. Construímos as subredes do metabolismo de pirimidinas nestes grupos (**Figura 30, Figura 31**) e as utilizamos para criar uma rede de interseção que representa o repertório compartilhado de genes comprometidos com o metabolismo de pirimidinas (**Figura 32**).

Figura 30. Redes IPP envolvidas no metabolismo de pirimidinas em a) *M. hominis*, b) *M. genitalium* e c) *M. pneumoniae*.

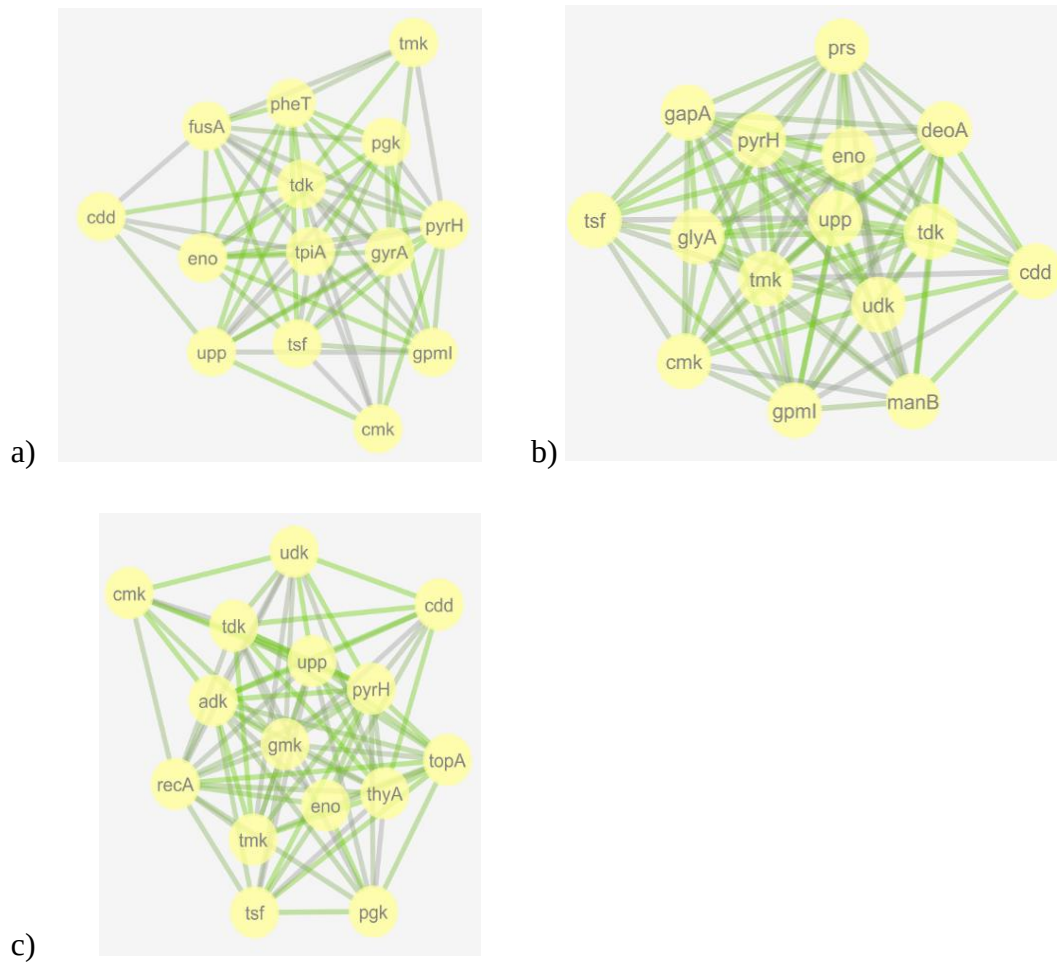
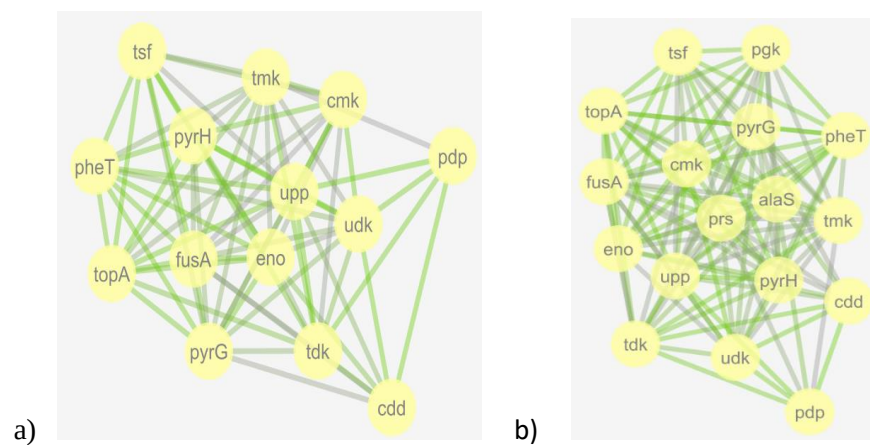
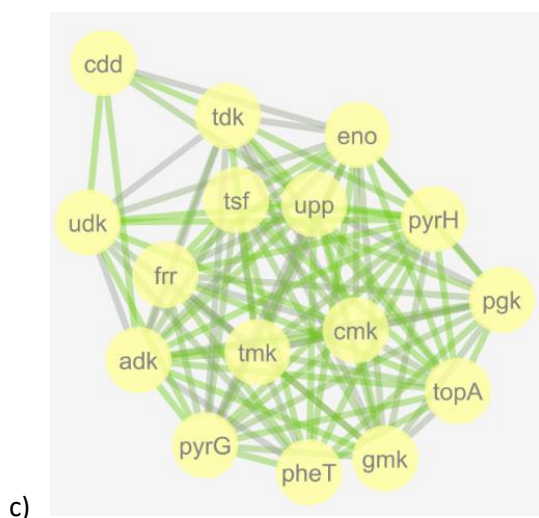


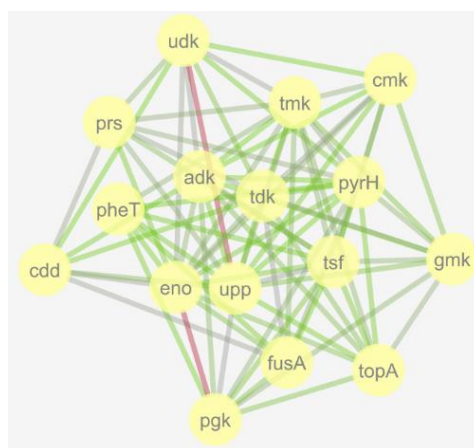
Figura 31. Redes IPP envolvidas no metabolismo de pirimidinas em a) *U. diversum*, b) *U. urealyticum* e c) *U. parvum*.





As redes IPP se assemelham nestes dois grupos, o que se reflete em seus parâmetros de rede similares (**Tabela 6**). A classe Mollicutes possui um repertório proteico para síntese e aproveitamento de pirimidinas que é compartilhado pelos seus integrantes, como se pode observar pela ampla rede de interseção deste grupo (**Figura 32**).

Figura 32. Interseção de redes IPP do metabolismo de pirimidinas de *Mycoplasma* e *Ureaplasma*.



O segundo evento de transferência horizontal apontado pelos nossos dados é do gênero *Mycoplasma* para *Corynebacterium urealyticum*. Construímos a rede IPP do metabolismo de pirimidinas em *C. urealyticum* (**Figura 33**) e novamente criamos a rede IPP representante da interseção entre as redes de *Corynebacterium* e *Mycoplasma*, demonstrando as interações e nodos conservados da rede IPP de pirimidinas entre estes grupos (**Figura 34**).

Figura 33. Subrede IPP do metabolismo de pirimidinas em *C. urealyticum*.

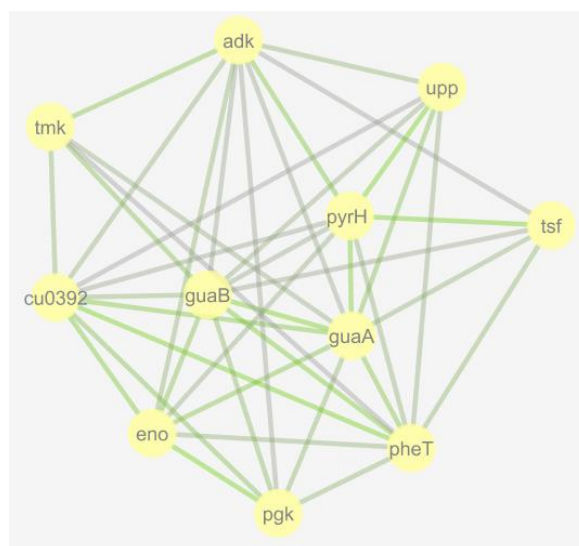
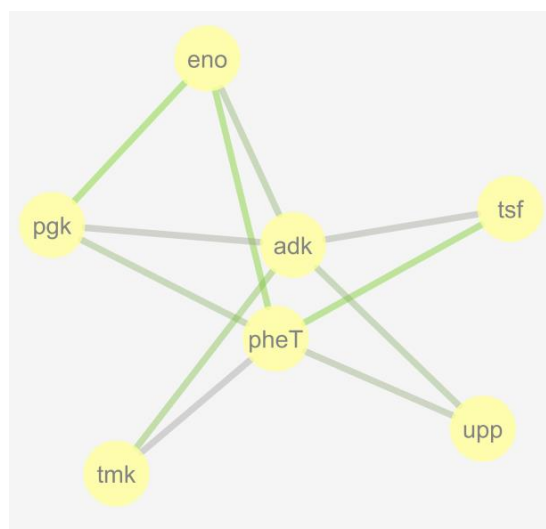


Figura 34. Interseção das redes IPP do metabolismo de pirimidinas em *Mycoplasma* e *Corynebacterium*.



Analisando os parâmetros de centralidade das redes IPP do metabolismo de pirimidinas de *Mycoplasma*, *Ureaplasma* e *Corynebacterium* encontramos similaridades referentes ao coeficiente de agrupamento, centralização da rede e comprimento de via característico, que se mostra quase idêntico entre as espécies analisadas (**Tabela 6**). Como ocorre na via do metabolismo de purinas, as redes IPP de pirimidinas se assemelham tanto em composição proteica quanto nos parâmetros topológicos.

Tabela 6. Os parâmetros da subrede do metabolismo de pirimidinas dos grupos investigados.

	Coef. de agrupamento	Centralização de rede	Compr. de via característico
<i>U. diversum</i>	0.841	0.167	1.220
<i>U. urealyticum</i>	0.834	0.163	1.206
<i>U. parvum</i>	0.872	0.210	1.183
<i>M. hominis</i>	0.679	0.218	1.341
<i>M. pneumoniae</i>	0.818	0.275	1.238
<i>M. genitalium</i>	0.863	0.176	1.152
<i>C. urealyticum</i>	0.811	0.289	1.236

Os eventos de transferência horizontal nas subredes do metabolismo da urease, purinas e pirimidinas apontados pelas nossas análises se refletem em composição e organização de redes IPP bastante semelhantes entre os grupos investigados. Os parâmetros de centralidade sugerem uma conservação na topologia dessas redes IPP ao longo da história evolutiva de *Ureaplasma*, *Mycoplasma*, *Corynebacterium* e *S. aureus*, que possuem estilos de vida próximos e colonizam o nicho urogenital em seus hospedeiros multicelulares.

5. CONSIDERAÇÕES FINAIS

Ureaplasma e *Mycoplasma* estão entre os microrganismos com os menores genomas conhecidos, resultantes de um processo de redução genômica que os levou ao longo de sua história evolutiva a possuir genomas “quase mínimos” naturais. A hipótese que trabalhamos neste estudo é a de que a redução genômica e a consequente simplificação de vias metabólicas (ora incompletas, ora totalmente ausentes) se reflete no interatoma destes grupos. Enquanto bactérias filogeneticamente mais distantes e metabolicamente mais complexas como *S. aureus* e *Corynebacterium urealyticum* possuem redes IPP globais mais compartimentadas em grupos de maior conexão comprometidas com vias metabólicas completas, nossos dados sugerem que *U. diversum*, outros *Ureaplasmas* e *Micoplasmas* vão pelo caminho oposto. Apesar de possuírem *clusters* de proteínas conectadas em maior frequência, estes grupos possuem proteínas que interagem entre si de uma maneira simplificada e mais geral, exercendo diferentes funções e por vezes assumindo o papel ocupado por várias proteínas em suas bactérias relacionadas, como é o caso de *adk*, *rnr*, *deoD* e outras proteínas envolvidas nas vias de

síntese de purinas e pirimidinas. As proteínas centrais da rede IPP de *U. diversum* estão envolvidas na aquisição de diversidade gênica via aumento de taxas de mutação e transferência horizontal de genes, indicando que o fluxo gênico entre estas bactérias, o ambiente e suas comunidades bacterianas correlacionadas é de grande importância para este grupo. Urease é um gene de nicho necessário para produção de energia durante a colonização do trato urogenital por *Ureaplasma*, *Corynebacterium* e *S. aureus* e é compartilhado entre estes grupos via transferência horizontal de genes, o que também ocorre com proteínas envolvidas no metabolismo de purinas e pirimidinas. O processo de redução do genoma nos Mollicutes parece ter favorecido a evolução de redes IPP mais conectadas, em que uma maior interação entre as proteínas restantes compensa a perda gênica sofrida pelo grupo ao longa da sua história evolutiva.

6. REFERÊNCIAS

ÅKERLUND, M.; MÅRDH, P. **Isolation and identification of *Corynebacterium vaginale* (*Haemophilus vaginalis*) in women with infections of the lower genital tract.** Acta obstetricia et gynecologica Scandinavica, v. 53, n. 1, p. 85–90, 1974.

ALMAULA, N.; LU, Q.; DELGADO, J.; BELKIN, S.; INOUE, M. **Nucleoside diphosphate kinase from *Escherichia coli*.** Journal of bacteriology, v. 177, n. 9, p. 2524–2529, 1995.

ANDERSSON, S. G. E.; KURLAND, C. G. **Genomic evolution drives the evolution of the translation system.** Biochemistry and cell biology, v. 73, n. 11–12, p. 775–787, 1995.

ARGUE, B.; CHOUSALKAR, K. K.; CHENOWETH, P. J. **Presence of *Ureaplasma diversum* in the Australian cattle population.** Australian veterinary journal, v. 91, n. 3, p. 99–101, 2013.

ARRAES, F.; CARVALHO, M. J. A. DE; MARANHÃO, A. Q.; et al. **Differential metabolism of *Mycoplasma* species as revealed by their genomes.** Genetics and Molecular Biology, v. 30, n. 1, p. 182–189, 2007.

ASSENOV, Y.; RAMÍREZ, F.; SCHELHORN, S.-E.; LENGAUER, T.; ALBRECHT, M. **Computing topological parameters of biological networks.** Bioinformatics, v. 24, n. 2, p. 282–284, 2007.

AZAM, F.; LONG, R. A. **Oceanography: sea snow microcosms.** Nature, v. 414, n. 6863, p. 495, 2001.

AZEVEDO, H.; MOREIRA-FILHO, C. A. **Topological robustness analysis of protein interaction networks reveals key targets for overcoming chemotherapy resistance in glioma.** Scientific reports, v. 5, p. 16830, 2015.

- BARABÁSI, A.-L.; ALBERT, R. **Emergence of scaling in random networks.** Science, v. 286, n. 5439, p. 509–512, 1999.
- BIZARRO, C. V.; SCHUCK, D. C. **Purine and pyrimidine nucleotide metabolism in Mollicutes.** Genetics and Molecular Biology, v. 30, n. 1, p. 190–201, 2007.
- BLANCHARD, A.; RAZIN, S.; KENNY, G. E.; BARILE, M. F. **Characteristics of *Ureaplasma urealyticum* urease.** Journal of bacteriology, v. 170, n. 6, p. 2692–2697, 1988.
- BROWNING, G.; CITTI, C. **Mollicutes: molecular biology and pathogenesis.** Horizon Scientific Press, 2014.
- BUCURENCI, N.; SAKAMOTO, H.; BRIOZZO, P.; et al. **CMP kinase from *Escherichia coli* is structurally related to other nucleoside monophosphate kinases.** Journal of Biological Chemistry, v. 271, n. 5, p. 2856–2862, 1996.
- BUTLAND, G.; PEREGRÍN-ALVAREZ, J. M.; LI, J.; et al. **Interaction network containing conserved and essential protein complexes in *Escherichia coli*.** Nature, v. 433, n. 7025, p. 531, 2005.
- CABISCOL CATALÀ, E.; TAMARIT SUMALLA, J.; ROS SALVADOR, J. **Oxidative stress in bacteria and protein damage by reactive oxygen species.** International Microbiology, 2000, vol. 3, núm. 1, p. 3-8, 2000.
- CAPOCCIA, R.; GREUB, G.; BAUD, D. ***Ureaplasma urealyticum*, *Mycoplasma hominis* and adverse pregnancy outcomes.** Current opinion in infectious diseases, v. 26, n. 3, p. 231–240, 2013.
- CARDOSO, M. V.; SCARCELLI, E.; GRASSO, L. M. P. S.; TEIXEIRA, S. R.; GENOVEZ, M. É. ***Ureaplasma diversum* and reproductive disorder in Brazilian cows and heifers; first report.** Animal Reproduction Science, v. 63, n. 3, p. 137–143, 2000.
- CASCHERA, F.; NOIREAUX, V. **Integration of biological parts toward the synthesis of a minimal cell.** Current opinion in chemical biology, v. 22, p. 85–91, 2014.
- CASELL, G.; CROUSE, D.; CANUPP, K.; et al. **Association of *Ureaplasma urealyticum* infection of the lower respiratory tract with chronic lung disease and death in very-low-birth-weight infants.** The Lancet, v. 332, n. 8605, p. 240–245, 2018.
- CASELL, G. H.; WAITES, K. B.; WATSON, H. L.; CROUSE, D. T.; HARASAWA, R. ***Ureaplasma urealyticum* intrauterine infection: role in prematurity and disease in newborns.** Clinical Microbiology Reviews, v. 6, n. 1, p. 69–87, 1993.
- CASTELLANOS, M.; WILSON, D. B.; SHULER, M. L. **A modular minimal cell model: purine and pyrimidine transport and metabolism.** Proceedings of the National Academy of Sciences of the United States of America, v. 101, n. 17, p. 6681–

6686, 2004.

CHANG, A. N. **Prioritizing genes for pathway impact using network analysis.** Protein Networks and Pathway Analysis. p.141–156, 2009.

CLATWORTHY, A. E.; PIERSON, E.; HUNG, D. T. **Targeting virulence: a new paradigm for antimicrobial therapy.** Nature chemical biology, v. 3, n. 9, p. 541, 2007.

COCKS, T. M.; ANGUS, J. A.; CAMPBELL, J. H.; CAMPBELL, G. R. **Release and properties of endothelium-derived relaxing factor (EDRF) from endothelial cells in culture.** Journal of cellular physiology, v. 123, n. 3, p. 310–320, 1985.

COLE, S. T.; GIRONS, I. SAINT. **Bacterial genomics.** FEMS microbiology reviews, v. 14, n. 2, p. 139–160, 1994.

CONSORTIUM, A. I. M. **Evidence for network evolution in an Arabidopsis interactome map.** Science, v. 333, n. 6042, p. 601–607, 2011.

CONSORTIUM, U. **UniProt: the universal protein knowledgebase.** Nucleic acids research, v. 45, n. D1, p. D158–D169, 2016.

CONTRERAS, A.; CORRALES, J. C.; SIERRA, D.; MARCO, J. **Prevalence and aetiology of non-clinical intramammary infection in Murciano-Granadina goats.** Small Ruminant Research, v. 17, n. 1, p. 71–78, 1995.

COORDINATORS, N. R. **Database resources of the national center for biotechnology information.** Nucleic acids research, v. 44, n. Database issue, p. D7, 2016.

CORDOVA, C. M. M.; HOELTGEBAUM, D. L.; MACHADO, L. D. P. N.; SANTOS, L. D. O. S. **Molecular biology of mycoplasmas: from the minimum cell concept to the artificial cell.** Anais da Academia Brasileira de Ciências, v. 88, p. 599–607, 2016.

COX, M M. **The RecA protein as a recombinational repair system.** Molecular microbiology, v. 5, n. 6, p. 1295–1299, 1991.

CUFF, J. A.; BIRNEY, E.; CLAMP, M. E.; BARTON, G. J. **ProtEST: protein multiple sequence alignments from expressed sequence tags.** Bioinformatics, v. 16, n. 2, p. 111–116, 2000.

CULMAN, S. W.; BUKOWSKI, R.; GAUCH, H. G.; CADILLO-QUIROZ, H.; BUCKLEY, D. H. **T-REX: software for the processing and analysis of T-RFLP data.** BMC bioinformatics, v. 10, n. 1, p. 171, 2009.

CUSICK, M. E.; KLITGORD, N.; VIDAL, M.; HILL, D. E. **Interactome: gateway into systems biology.** Human molecular genetics, v. 14, n. suppl_2, p. R171–R181, 2005.

DE LA CRUZ, F.; DAVIES, J. **Horizontal gene transfer and the origin of species:**

lessons from bacteria. Trends in microbiology, v. 8, n. 3, p. 128–133, 2000.

DENG, M.; ZHANG, K.; MEHTA, S.; CHEN, T.; SUN, F. **Prediction of protein function using protein–protein interaction data.** Journal of Computational Biology, v. 10, n. 6, p. 947–960, 2003.

DEUTSCHER, J.; FRANCKE, C.; POSTMA, P. W. **How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria.** Microbiology and Molecular Biology Reviews, v. 70, n. 4, p. 939–1031, 2006.

DEVRIESE, L. A.; RIEGEL, P.; HOMMEZ, J.; et al. **Identification of *Corynebacterium glucuronolyticum* strains from the urogenital tract of humans and pigs.** Journal of clinical microbiology, v. 38, n. 12, p. 4657–4659, 2000.

DIGIULIO, D. B. **Diversity of microbes in amniotic fluid.** Seminars in Fetal and Neonatal Medicine, v. 17, n. 1, p. 2–11, 2018.

DOIG, P. A.; RUHNKE, H. L.; MACKAY, A. L.; PALMER, N. C. **Bovine granular vulvitis associated with ureaplasma infection.** The Canadian Veterinary Journal, v. 20, n. 4, p. 89, 1979.

DOOLITTLE, W. F. **You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.** Trends in Genetics, v. 14, n. 8, p. 307–311, 1998.

DORDET-FRISONI, E.; SAGNÉ, E.; BARANOWSKI, E.; et al. **Chromosomal transfers in mycoplasmas: when minimal genomes go mobile.** MBio, v. 5, n. 6, p. e01958-14, 2014.

DRLICA, K.; ROUVIERE-YANIV, J. **Histonelike proteins of bacteria.** Microbiological reviews, v. 51, n. 3, p. 301, 1987.

DUNIN-HORKAWICZ, S.; KOPEC, K. O.; LUPAS, A. N. **Prokaryotic ancestry of eukaryotic protein networks mediating innate immunity and apoptosis.** Journal of molecular biology, v. 426, n. 7, p. 1568–1582, 2014.

DYBVIG, K.; VOELKER, L. L. **Molecular biology of mycoplasmas.** Annual Reviews in Microbiology, v. 50, n. 1, p. 25–57, 1996.

ERTEN, S.; LI, X.; BEBEK, G.; LI, J.; KOYUTÜRK, M. **Phylogenetic analysis of modularity in protein interaction networks.** BMC bioinformatics, v. 10, n. 1, p. 333, 2009.

FLEISCHMANN, R. D.; ADAMS, M. D.; WHITE, O.; et al. **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** Science, v. 269, n. 5223, p. 496–512, 1995.

FORSBERG, K. J.; REYES, A.; WANG, B.; et al. **The shared antibiotic resistome of soil bacteria and human pathogens.** Science, v. 337, n. 6098, p. 1107–1111, 2012.

FORSTER, A. C.; CHURCH, G. M. **Towards synthesis of a minimal cell.** Molecular systems biology, v. 2, n. 1, p. 45, 2006.

FOSTER, Patricia L. **Stress responses and genetic variation in bacteria.** Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, v. 569, n. 1, p. 3–11, 2005.

FRASER, C. M.; GOCAYNE, J. D.; WHITE, O.; et al. **The minimal gene complement of *Mycoplasma genitalium*.** Science, v. 270, n. 5235, p. 397–404, 1995.

FUQUA, W. C.; WINANS, S. C.; GREENBERG, E. P. **Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.** Journal of bacteriology, v. 176, n. 2, p. 269, 1994.

GALHARDO, Rodrigo S; HASTINGS, Philip J; ROSENBERG, Susan M. **Mutation as a stress response and the regulation of evolvability.** Critical reviews in biochemistry and molecular biology, v. 42, n. 5, p. 399–435, 2007.

GARCÍA-CASTILLO, M.; MOROSINI, M.-I.; GÁLVEZ, M.; et al. **Differences in biofilm development and antibiotic susceptibility among clinical *Ureaplasma urealyticum* and *Ureaplasma parvum* isolates.** Journal of antimicrobial chemotherapy, v. 62, n. 5, p. 1027–1030, 2008.

GATERMANN, S.; MARRE, R. **Cloning and expression of *Staphylococcus saprophyticus* urease gene sequences in *Staphylococcus carnosus* and contribution of the enzyme to virulence.** Infection and immunity, v. 57, n. 10, p. 2998–3002, 1989.

GENTRY, D.; BENGRA, C.; IKEHARA, K.; CASHEL, M. **Guanylate kinase of *Escherichia coli* K-12.** Journal of Biological Chemistry, v. 268, n. 19, p. 14316–14321, 1993.

GERDES, SY.; SCHOLLE, M. D.; CAMPBELL, J. W.; et al. **Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655.** Journal of bacteriology, v. 185, n. 19, p. 5673–5684, 2003.

GHERARDI, G.; DI BONAVENTURA, G.; POMPILIO, A.; SAVINI, V. ***Corynebacterium glucuronolyticum* causing genitourinary tract infection: Case report and review of the literature.** IDCases, v. 2, n. 2, p. 56–58, 2015.

GIBBS, A. J.; CALISHER, C. H.; GARCIA-ARENAL, F. **Molecular basis of virus evolution.** Cambridge University Press, 2005.

GLASS, J. I.; MERRYMAN, C.; WISE, K. S.; HUTCHISON, C. A.; SMITH, H. O. **Minimal Cells—Real and Imagined.** Cold Spring Harbor perspectives in biology, v. 9, n. 12, p. a023861, 2017.

GROUP, E. S. **An international association between *Helicobacter pylori* infection and gastric cancer.** The Lancet, v. 341, n. 8857, p. 1359–1363, 1993.

GUIMARÃES, L. C.; DE CASTRO SOARES, S.; TROST, E.; et al. **Genome informatics and vaccine targets in *Corynebacterium urealyticum* using two whole genomes, comparative genomics, and reverse vaccinology.** BMC genomics, v. 16, n. 5, p. S7, 2015.

GUINANE, C. M.; PENADÉS, J. R.; FITZGERALD, J. R. **The role of horizontal gene transfer in *Staphylococcus aureus* host adaptation.** Virulence, v. 2, n. 3, p. 241–243, 2011.

HAN, Y.-C.; SONG, J.-M.; WANG, L.; et al. **Prediction and characterization of protein-protein interaction network in *Bacillus licheniformis* WX-02.** Scientific reports, v. 6, p. 19486, 2016.

HEHEMANN, J.-H.; CORREC, G.; BARBEYRON, T.; et al. **Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota.** Nature, v. 464, n. 7290, p. 908, 2010.

HERRON-OLSON, L.; FITZGERALD, J. R.; MUSSER, J. M.; KAPUR, V. **Molecular correlates of host specialization in *Staphylococcus aureus*.** PloS one, v. 2, n. 10, p. e1120, 2007.

HEUER, H.; SMALLA, K. **Plasmids foster diversification and adaptation of bacterial populations in soil.** FEMS microbiology reviews, v. 36, n. 6, p. 1083–1104, 2012.

HUTCHISON, C. A.; CHUANG, R.-Y.; NOSKOV, V. N.; et al. **Design and synthesis of a minimal bacterial genome.** Science, v. 351, n. 6280, p. aad6253, 2016.

ISHAG, Hassan Z A et al. ***E. coli* recA gene improves gene targeted homologous recombination in *Mycoplasma hyorhinis*.** Journal of microbiological methods, v. 136, p. 49–56, 2017.

ITO, T.; CHIBA, T.; OZAWA, R.; et al. **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** Proceedings of the National Academy of Sciences, v. 98, n. 8, p. 4569–4574, 2001.

JAIN, R.; RIVERA, M. C.; LAKE, J. A. **Horizontal gene transfer among genomes: the complexity hypothesis.** Proceedings of the National Academy of Sciences, v. 96, n. 7, p. 3801–3806, 1999.

JANNASCH, H. W.; MOTTTL, M. J. **Geomicrobiology of deep-sea hydrothermal vents.** Science, v. 229, n. 4715, p. 717–725, 1985.

JANSSON, J. K.; HOFMOCKEL, K. S. **The soil microbiome - from metagenomic to metaphenomics.** Current opinion in microbiology, v. 43, p. 162 – 168, 2018.

JARVILL-TAYLOR, K. J.; VANDYK, C.; MINION, F. C. **Cloning of mnuA, a Membrane Nuclease Gene of *Mycoplasma pulmonis*, and Analysis of Its Expression in *Escherichia coli*.** Journal of bacteriology, v. 181, n. 6, p. 1853–1860, 1999.

JENSEN, R. A. **Enzyme recruitment in evolution of new function.** Annual Reviews in Microbiology, v. 30, n. 1, p. 409–425, 1976.

JIAO, Lingxia et al. **Heat, acid and cold stresses enhance the expression of DnaK gene in *Alicyclobacillus acidoterrestris*.** Food Research International, v. 67, p. 183–192, 2015.

JONG, A. Y.; CAMPBELL, J. L. **Characterization of *Saccharomyces cerevisiae* thymidylate kinase, the CDC8 gene product. General properties, kinetic analysis, and subcellular localization.** Journal of Biological Chemistry, v. 259, n. 23, p. 14394–14398, 1984.

JONSSON, P. F.; BATES, P. A. **Global topological features of cancer proteins in the human interactome.** Bioinformatics, v. 22, n. 18, p. 2291–2297, 2006.

JUHAS, M.; VAN DER MEER, J. R.; GAILLARD, M.; et al. **Genomic islands: tools of bacterial horizontal gene transfer and evolution.** FEMS microbiology reviews, v. 33, n. 2, p. 376–393, 2009.

JUHAS, M.; REUS, D. R.; ZHU, B.; COMMICHAU, F. M. ***Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering.** Microbiology, v. 160, n. 11, p. 2341–2351, 2014.

KANEHISA, M.; GOTO, S. **KEGG: kyoto encyclopedia of genes and genomes.** Nucleic acids research, v. 28, n. 1, p. 27–30, 2000.

KATOH, K.; STANDLEY, D. M. **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** Molecular biology and evolution, v. 30, n. 4, p. 772–780, 2013.

KOBAYASHI, K.; EHRLICH, S. D.; ALBERTINI, A.; et al. **Essential *Bacillus subtilis* genes.** Proceedings of the National Academy of Sciences, v. 100, n. 8, p. 4678–4683, 2003.

KOKKAYIL, P.; DHAWAN, B. ***Ureaplasma*: current perspectives.** Indian journal of medical microbiology, v. 33, n. 2, p. 205–14, 2015.

KRASS, C. J.; GARDNER, M. A. X. W. **Etymology of the term mycoplasma.** International Journal of Systematic and Evolutionary Microbiology, v. 23, n. 1, p. 62–64, 1973.

KUO, C.; OCHMAN, H. **The fate of new bacterial genes.** FEMS microbiology reviews, v. 33, n. 1, p. 38–43, 2009.

KÜLTZ, Dietmar. **Molecular and evolutionary basis of the cellular stress response.** Annu. Rev. Physiol., v. 67, p. 225–257, 2005.

LAGE, K.; KARLBERG, E. O.; STØRLING, Z. M.; et al. **A human phenome-interactome network of protein complexes implicated in genetic disorders.** Nature

biotechnology, v. 25, n. 3, p. 309, 2007.

LAWRENCE, J. G.; OCHMAN, H. **Molecular archaeology of the *Escherichia coli* genome**. Proceedings of the National Academy of Sciences, v. 95, n. 16, p. 9413–9417, 1998.

LESLIE, S. B.; ISRAELI, E.; LIGHTHART, B.; CROWE, J. H.; CROWE, L. M. **Trehalose and sucrose protect both membranes and proteins in intact bacteria during drying**. Applied and environmental microbiology, v. 61, n. 10, p. 3592–3597, 1995.

LI, S.; ARMSTRONG, C. M.; BERTIN, N.; et al. **A map of the interactome network of the metazoan *C. elegans***. Science, v. 303, n. 5657, p. 540–543, 2004.

LI, Y.; ZHANG, Y.; YAN, H. **Kinetic and thermodynamic characterizations of yeast guanylate kinase**. Journal of Biological Chemistry, v. 271, n. 45, p. 28038–28044, 1996.

LIGON, J. V.; KENNY, G. E. **Virulence of ureaplasma urease for mice**. Infection and immunity, v. 59, n. 3, p. 1170–1171, 1991.

LIMA-MENDEZ, G.; FAUST, K.; HENRY, N.; et al. **Determinants of community structure in the global plankton interactome**. Science, v. 348, n. 6237, p. 1262073, 2015.

LINDSAY, J. A. ***Staphylococcus aureus* genomics and the impact of horizontal gene transfer**. International Journal of Medical Microbiology, v. 304, n. 2, p. 103–109, 2014.

LOWDER, B. V.; GUINANE, C. M.; ZAKOUR, N. L. BEN; et al. **Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus***. Proceedings of the National Academy of Sciences, v. 106, n. 46, p. 19545–19550, 2009.

LU, Q.; INOUE, M. **Adenylate kinase complements nucleoside diphosphate kinase deficiency in nucleotide metabolism**. Proceedings of the National Academy of Sciences, v. 93, n. 12, p. 5720–5725, 1996.

LUISI, P. L.; FERRI, F.; STANO, P. **Approaches to semi-synthetic minimal cells: a review**. Naturwissenschaften, v. 93, n. 1, p. 1–13, 2006.

LUNDEGAARD, C.; JENSEN, K. F. **Kinetic mechanism of uracil phosphoribosyltransferase from *Escherichia coli* and catalytic importance of the conserved proline in the PRPP binding site**. Biochemistry, v. 38, n. 11, p. 3327–3334, 1999.

MACKELPRANG, Rachel et. al. **Microbial survival strategies in ancient permafrost: insights from metagenomics**. The ISME journal, v. 11, n. 10, p. 2305, 2017.

MANILOFF, J. **The minimal cell genome: “on being the right size”**. Proceedings of the National Academy of Sciences, v. 93, n. 19, p. 10004–10006, 1996.

- MARQUES, L. M.; REZENDE, I. S.; BARBOSA, M. S.; et al. ***Ureaplasma diversum* Genome Provides New Insights about the Interaction of the Surface Molecules of This Bacterium with the Host.** PloS one, v. 11, n. 9, p. e0161926, 2016.
- MCCARTHY, A. J.; LOEFFLER, A.; WITNEY, A. A.; et al. **Extensive horizontal gene transfer during *Staphylococcus aureus* co-colonization in vivo.** Genome biology and evolution, v. 6, n. 10, p. 2697–2708, 2014.
- MCDANIEL, L. D.; YOUNG, E.; DELANEY, J.; et al. **High frequency of horizontal gene transfer in the oceans.** Science, v. 330, n. 6000, p. 50, 2010.
- MCELWAIN, M. C.; CHANDLER, D. K. F.; BARILE, M. F.; et al. **Purine and pyrimidine metabolism in Mollicutes species.** International Journal of Systematic and Evolutionary Microbiology, v. 38, n. 4, p. 417–423, 1988.
- MCINTOSH, B. K.; RENFRO, D. P.; KNAPP, G. S.; et al. **EcoliWiki: a wiki-based community resource for *Escherichia coli*.** Nucleic acids research, v. 40, n. D1, p. D1270–D1277, 2011.
- MICHOD, Richard; BERNSTEIN, Harris e NEDELCO, Aurora M. **Adaptive value of sex in microbial pathogens.** Infection, Genetics and Evolution, v. 8, n. 3, p. 267–285, 2008.
- MILLER, M. B.; BASSLER, B. L. **Quorum sensing in bacteria.** Annual Reviews in Microbiology, v. 55, n. 1, p. 165–199, 2001.
- MILLER, R.; CHELMONSKA-SOYTA, A.; SMITS, B.; FOSTER, R.; ROSENDAL, S. ***Ureaplasma diversum* as a cause of reproductive disease in cattle.** Veterinary Clinics of North America: Food Animal Practice, v. 10, n. 3, p. 479–490, 1994.
- MINION, F. C.; JARVILL-TAYLOR, K. J.; BILLINGS, D. E.; TIGGES, E. **Membrane-associated nuclease activities in mycoplasmas.** Journal of bacteriology, v. 175, n. 24, p. 7842–7847, 1993.
- MOMBACH, J. C. M.; LEMKE, N.; DA SILVA, N. M.; et al. **Bioinformatics analysis of mycoplasma metabolism: Important enzymes, metabolic similarities, and redundancy.** Computers in biology and medicine, v. 36, n. 5, p. 542–552, 2006.
- MOROWITZ, H. J. **The completeness of molecular biology.** Israel journal of medical sciences, v. 20, n. 9, p. 750–753, 1984.
- MULIRA, G. L.; SAUNDERS, J. R.; KARIUKI, D. P. **Isolation of *Ureaplasma diversum* from bovine granular vulvitis in Kenya.** Bulletin of animal health and production in Africa= Bulletin de la sante et de la production animales en Afrique, 1989.
- MUSHEGIAN, A. R.; KOONIN, E. V. **A minimal gene set for cellular life derived by comparison of complete bacterial genomes.** Proceedings of the National Academy of Sciences, v. 93, n. 19, p. 10268–10273, 1996.

NAVARRE, W. W.; SCHNEEWIND, O. **Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope.** Microbiology and molecular biology reviews, v. 63, n. 1, p. 174–229, 1999.

NELSON, K. E.; CLAYTON, R. A.; GILL, S. R.; et al. **Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*.** Nature, v. 399, n. 6734, p. 323, 1999.

NEPUSZ, T.; YU, H.; PACCANARO, A. **Detecting overlapping protein complexes in protein-protein interaction networks.** Nature methods, v. 9, n. 5, p. 471, 2012.

NEYROLLES, O.; FERRIS, S.; BEHBAHANI, N.; MONTAGNIER, L.; BLANCHARD, A. **Organization of *Ureaplasma urealyticum* urease gene cluster and expression in a suppressor strain of *Escherichia coli*.** Journal of bacteriology, v. 178, n. 3, p. 647–655, 1996.

NIEHUS, R.; MITRI, S.; FLETCHER, A. G.; FOSTER, K. R. **Migration and horizontal gene transfer divide microbial genomes into multiple niches.** Nature communications, v. 6, p. 8924, 2015.

NOIREAUX, V.; LIBCHABER, A. **A vesicle bioreactor as a step toward an artificial cell assembly.** Proceedings of the national academy of sciences of the United States of America, v. 101, n. 51, p. 17669–17674, 2004.

NOMURA, S. M.; TSUMOTO, K.; HAMADA, T.; et al. **Gene expression within cell-sized lipid vesicles.** ChemBioChem, v. 4, n. 11, p. 1172–1175, 2003.

NOTHAFT, H.; SZYMANSKI, C. M. **Protein glycosylation in bacteria: sweeter than ever.** Nature Reviews Microbiology, v. 8, n. 11, p. 765, 2010.

OLIVEIRA, A.; OLIVEIRA, L. C.; ABURJAILE, F.; et al. **Insight of Genus *Corynebacterium*: Ascertaining the Role of Pathogenic and Non-pathogenic Species.** Frontiers in microbiology, v. 8, p. 1937, 2017.

OLIVEIRA FILHO, B. D.; PORTO, R. N. G.; GAMBARINI, M. L.; et al. **Isolamento do *Ureaplasma diversum* em muco vulvovaginal de vacas leiteiras repetidoras de estro no estado de Alagoas Brasil.** Archives of Veterinary Science, v. 10, n. 2, 2005.

OULAS, Anastasis et al. **Metagenomic investigation of the geologically unique Hellenic Volcanic Arc reveals a distinctive ecosystem with unexpected physiology.** Environmental microbiology, v. 18, n. 4, 1122–1136, 2016.

PANDEY, N. K. **Transport of nucleic acid bases and nucleosides across the bacterial membrane.** FEMS microbiology letters, v. 21, n. 1, p. 11–14, 1984.

PAULSEN, I. T.; NGUYEN, L.; SLIWINSKI, M. K.; RABUS, R.; SAIER, M. H. **Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes.** Journal of molecular biology, v. 301, n. 1, p. 75–100, 2000.

PERNI, S. C.; VARDHANA, S.; KORNEEVA, I.; et al. ***Mycoplasma hominis* and**

***Ureaplasma urealyticum* in midtrimester amniotic fluid: association with amniotic fluid cytokine levels and pregnancy outcome.** American Journal of Obstetrics & Gynecology, v. 191, n. 4, p. 1382–1386, 2004.

PETIT, Marie-Agnès; EHRLICH, Dusko. **Essential bacterial helicases that counteract the toxicity of recombination proteins.** The EMBO journal, v. 21, n. 12, p. 3137–3147, 2002.

PITCHER, D. **Mycoplasmas. Molecular Biology and Pathogenesis.** Journal of Medical Microbiology, v. 39, n. 4, p. 319, 1993.

POLLACK, J. D. ***Ureaplasma urealyticum*: an opportunity for combinatorial genomics.** Trends in microbiology, v. 9, n. 4, p. 169–175, 2001.

POLLACK, J. D.; MYERS, M. A.; DANDEKAR, T.; HERRMANN, R. **Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing“ household” nucleoside diphosphate kinase gene and activity by glycolytic kinases.** Omics: a journal of integrative biology, v. 6, n. 3, p. 247–258, 2002.

RAE, D. O.; CHENOWETH, P. J.; BROWN, M. B.; et al. **Reproductive performance of beef heifers: Effects of vulvo-vaginitis, *Ureaplasma diversum* and prebreeding antibiotic administration.** Theriogenology, v. 40, n. 3, p. 497–508, 1993.

RAIN, J.-C.; SELIG, L.; DE REUSE, H.; et al. **The protein–protein interaction map of *Helicobacter pylori*.** Nature, v. 409, n. 6817, p. 211–215, 2001.

RAMAN, K. **Construction and analysis of protein–protein interaction networks.** Automated experimentation, v. 2, n. 1, p. 2, 2010.

RAZIN, S. **Molecular biology and genetics of mycoplasmas (Mollicutes).** Microbiological reviews, v. 49, n. 4, p. 419, 1985.

RAZIN, S.; ARGAMAN, M.; AVIGAN, J. **Chemical composition of mycoplasma cells and membranes.** Microbiology, v. 33, n. 3, p. 477–487, 1963.

RAZIN, S.; YOGEV, D.; NAOT, Y. **Molecular biology and pathogenicity of mycoplasmas.** Microbiology and Molecular Biology Reviews, v. 62, n. 4, p. 1094–1156, 1998.

REICHARD, P. **Ribonucleotide reductases: the evolution of allosteric regulation.** Archives of biochemistry and biophysics, v. 397, n. 2, p. 149–155, 2002.

RICHTER, D. J. **Metagenomics and metatranscriptomes of oceanic communities.** Phycologia, v. 56, n. 4, p. 158, 2017.

RUIZ, J. C.; D’AFONSECA, V.; SILVA, A.; et al. **Evidence for reductive genome evolution and lateral acquisition of virulence functions in two *Corynebacterium pseudotuberculosis* strains.** PloS one, v. 6, n. 4, p. e18551, 2011.

RUSSO, J. F.; COPPOLA, K.; FURNESS, G. ***Mycoplasma hominis*, *Ureaplasma urealyticum*, and *Corynebacterium genitalium* recovered from the lower genital tracts of adolescent women.** International Journal of Gynecology & Obstetrics, v. 19, n. 6, p. 461–466, 1981.

SAITO, Yukiko e TAGUCHI, Hisataka e AKAMATSU, Takashi. **Fate of transforming bacterial genome following incorporation into competent cells of *Bacillus subtilis*: a continuous length of incorporated DNA.** Journal of bioscience and bioengineering, v. 101, n. 3, p. 257–262, 2006.

SALEM, N.; SALEM, L.; SABER, S.; ISMAIL, G.; BLUTH, M. H. ***Corynebacterium urealyticum*: a comprehensive review of an understated organism.** Infection and drug resistance, v. 8, p. 129, 2015.

SANDERSON, M. W.; CHENOWETH, P. J.; YEARY, T.; NIETFELD, J. C. **Prevalence and reproductive effects of *Ureaplasma diversum* in beef replacement heifers and the relationship to blood urea nitrogen level.** Theriogenology, v. 54, n. 3, p. 401–408, 2000.

SANTOS, A. P.; GUIMARAES, A. M. S.; DO NASCIMENTO, N. C.; et al. **Genome of *Mycoplasma haemofelis*, unraveling its strategies for survival and persistence.** Veterinary research, v. 42, n. 1, p. 102, 2011.

SCHULTZ, C. P.; YLISASTIGUI-PONS, L.; SERINA, L.; et al. **Structural and Catalytic Properties of CMP Kinase from *Bacillus subtilis*: A Comparative Analysis with the Homologous Enzyme from *Escherichia coli*.** Archives of biochemistry and biophysics, v. 340, n. 1, p. 144–153, 1997.

SETHI, S.; SHARMA, M.; NARANG, A.; AGGRAWAL, P. B. **Isolation pattern and clinical outcome of genital mycoplasma in neonates from a tertiary care neonatal unit.** Journal of tropical pediatrics, v. 45, n. 3, p. 143–145, 1999.

SHANNON, P.; MARKIEL, A.; OZIER, O.; et al. **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** Genome research, v. 13, n. 11, p. 2498–2504, 2003.

SHARAN, R.; ULITSKY, I.; SHAMIR, R. **Network-based prediction of protein function.** Molecular systems biology, v. 3, n. 1, p. 88, 2007.

SILVA, A. M. N.; VITORINO, R.; DOMINGUES, M. R. M.; SPICKETT, C. M.; DOMINGUES, P. **Post-translational modifications and mass spectrometry detection.** Free radical biology and medicine, v. 65, p. 925–941, 2013.

SIRAND-PUGNET, P.; CITTI, C.; BARRÉ, A.; BLANCHARD, A. **Evolution of mollicutes: down a bumpy road with twists and turns.** Research in microbiology, v. 158, n. 10, p. 754–766, 2007.

SIRAND-PUGNET, P.; LARTIGUE, C.; MARENDI, M.; et al. **Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome.** PLoS genetics, v. 3, n. 5, p. e75, 2007.

SMITH, D. G.; RUSSELL, W. C.; INGLEDEW, W. J.; THIRKELL, D. **Hydrolysis of urea by *Ureaplasma urealyticum* generates a transmembrane potential with resultant ATP synthesis.** Journal of bacteriology, v. 175, n. 11, p. 3253–3258, 1993.

STOCK, J. B.; NINFA, A. J.; STOCK, A. M. **Protein phosphorylation and regulation of adaptive responses in bacteria.** Microbiological reviews, v. 53, n. 4, p. 450–490, 1989.

STRIEPEN, B.; PRUIJSSERS, A. J. P.; HUANG, J.; et al. **Gene transfer in the evolution of parasite nucleotide biosynthesis.** Proceedings of the National Academy of Sciences of the United States of America, v. 101, n. 9, p. 3154–3159, 2004.

SZKLARCZYK, D.; FRANCESCHINI, A.; WYDER, S.; et al. **STRING v10: protein–protein interaction networks, integrated over the tree of life.** Nucleic acids research, v. 43, n. D1, p. D447–D452, 2014.

TAKAHASHI, T.; MORI, Y.; KOBAYASHI, H.; et al. **Phylogenetic positions and assignment of swine and ovine corynebacterial isolates based on the 16S rDNA sequence.** Microbiology and immunology, v. 41, n. 9, p. 649–655, 1997.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.** Molecular biology and evolution, v. 24, n. 8, p. 1596–1599, 2007.

TANAKA, K.; HENRY, C. S.; ZINNER, J. F.; et al. **Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model.** Nucleic acids research, v. 41, n. 1, p. 687–699, 2012.

TAUCH, A.; TROST, E.; TILKER, A.; et al. **The lifestyle of *Corynebacterium urealyticum* derived from its complete genome sequence established by pyrosequencing.** Journal of biotechnology, v. 136, n. 1–2, p. 11–21, 2008.

TAYLOR-ROBINSON, D. **Infections Due to Species of *Mycoplasma* and *Ureaplasma*: An Update.** Clinical Infectious Diseases, v. 23, n. 4, p. 671–682, 1996.

TAYLOR, I. W.; LINDING, R.; WARDE-FARLEY, D.; et al. **Dynamic modularity in protein interaction networks predicts breast cancer outcome.** Nature biotechnology, v. 27, n. 2, p. 199–204, 2009.

TEACHMAN, A. M.; FRENCH, C. T.; YU, H.; SIMMONS, W. L.; DYBVIG, K. **Gene transfer in *Mycoplasma pulmonis*.** Journal of bacteriology, v. 184, n. 4, p. 947–951, 2002.

TITZ, B.; RAJAGOPALA, S. V; GOLL, J.; et al. **The binary protein interactome of *Treponema pallidum*–the syphilis spirochete.** PloS one, v. 3, n. 5, p. e2292, 2008.

TONG, S. Y. C.; DAVIS, J. S.; EICHENBERGER, E.; HOLLAND, T. L.; FOWLER, V. G. ***Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical**

manifestations, and management. Clinical microbiology reviews, v. 28, n. 3, p. 603–661, 2015.

TORSVIK, V.; GOKSØYR, J.; DAAE, F. L. **High diversity in DNA of soil bacteria.** Applied and environmental microbiology, v. 56, n. 3, p. 782–787, 1990.

TOSADORI, G.; BESTVINA, I.; SPOTO, F.; LAUDANNA, C.; SCARDONI, G. **Creating, generating and comparing random network models with NetworkRandomizer.** F1000Research, v. 5, 2016.

TRACHTENBERG, S. **Mollicutes—wall-less bacteria with internal cytoskeletons.** Journal of structural biology, v. 124, n. 2, p. 244–256, 1998.

TREANGEN, T. J.; ROCHA, E. P. C. **Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes.** PLoS genetics, v. 7, n. 1, p. e1001284, 2011.

TRINCHIERI, A. **Urinary calculi and infection.** Urologia, v. 81, n. 2, p. 93–98, 2014.

TU, A. H.; TURNBOUGH, C. L. **Regulation of upp expression in *Escherichia coli* by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription.** Journal of bacteriology, v. 179, n. 21, p. 6665–6673, 1997.

VOS, M.; HESSELMAN, M. C.; TE BEEK, T. A.; VAN PASSEL, M. W. J.; EYRE-WALKER, A. **Rates of lateral gene transfer in prokaryotes: high but why?** Trends in microbiology, v. 23, n. 10, p. 598–605, 2015.

WAITES, K. B.; XIAO, L.; PARALANOV, V.; VISCARDI, R. M.; GLASS, J. I. **Mycoplasma and ureaplasma.** Molecular Typing in Bacterial Infections. p.229–281, 2013.

WATERS, C. M.; BASSLER, B. L. **Quorum sensing: cell-to-cell communication in bacteria.** Annu. Rev. Cell Dev. Biol., v. 21, p. 319–346, 2005.

WEISBURG, W. G.; TULLY, J. G.; ROSE, D. L.; et al. **A phylogenetic analysis of the mycoplasmas: basis for their classification.** Journal of bacteriology, v. 171, n. 12, p. 6455–6467, 1989.

WENG, G.; BHALLA, U. S.; IYENGAR, R. **Complexity in biological signaling systems.** Science, v. 284, n. 5411, p. 92–96, 1999.

WOESE, C. R.; STACKEBRANDT, E.; LUDWIG, W. **What are mycoplasmas: the relationship of tempo and mode in bacterial evolution.** Journal of molecular evolution, v. 21, n. 4, p. 305–316, 1985.

WOLF, Y. I.; ARAVIND, L.; KOONIN, E. V. **Rickettsiae and Chlamydiae: evidence of horizontal gene transfer and gene exchange.** Trends in Genetics, v. 15, n. 5, p. 173–175, 1999.

WUCHTY, S. **Controllability in protein interaction networks.** Proceedings of the

National Academy of Sciences, v. 111, n. 19, p. 7156–7160, 2014.

WUCHTY, S.; UETZ, P. **Protein-protein Interaction Networks of *E. coli* and *S. cerevisiae* are similar.** Scientific reports, v. 4, p. 7187, 2014.

XIAO, L.; PARALANOV, V.; GLASS, J. I.; et al. **Extensive horizontal gene transfer in ureaplasmas from humans questions the utility of serotyping for diagnostic purposes.** Journal of clinical microbiology, v. 49, n. 8, p. 2818–2826, 2011.

XU, Xixi e colab. **Heterogeneous expression of DnaK gene from *Alicyclobacillus acidoterrestris* improves the resistance of *Escherichia coli* against heat and acid stress.** AMB Express, v. 7, n. 1, p. 36, 2017.

YOUIL, R.; FINCH, L. R. **Isolation and characterization of *Mycoplasma mycoides* subsp. *mycoides* mutants deficient in nucleoside monophosphate transport.** Journal of bacteriology, v. 170, n. 12, p. 5922–5924, 1988.

ZEDER, M. A. **Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact.** Proceedings of the national Academy of Sciences, v. 105, n. 33, p. 11597–11604, 2008.



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

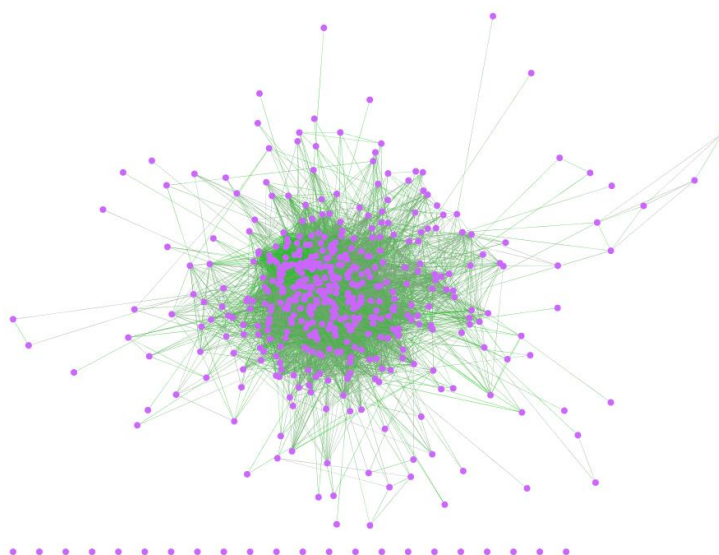


CAMPUS I - CIDADE UNIVERSITÁRIA
CEP: 58051-900 - JOÃO PESSOA - PB
TELEFONE: (83) 3216-7407
E-mail: pgbcm@dbm.ufpb.br

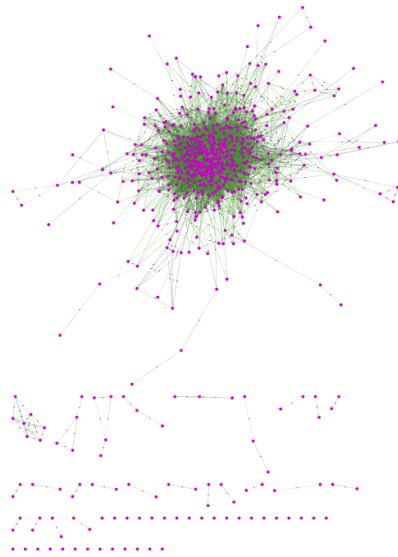
APÊNDICES

APÊNDICE A – Redes Globais das Espécies para comparação

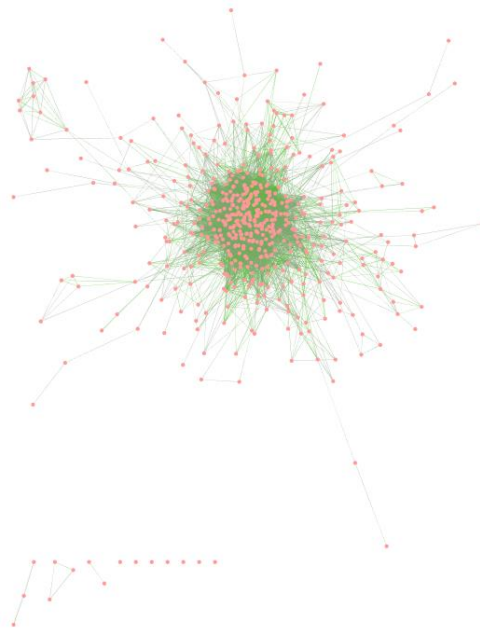
Mycoplasma genitalium



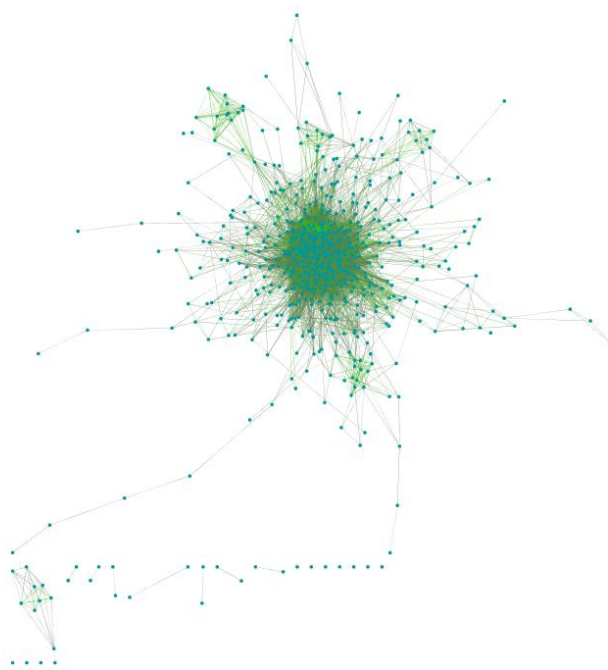
Mycoplasma pneumoniae



Mycoplasma hominis



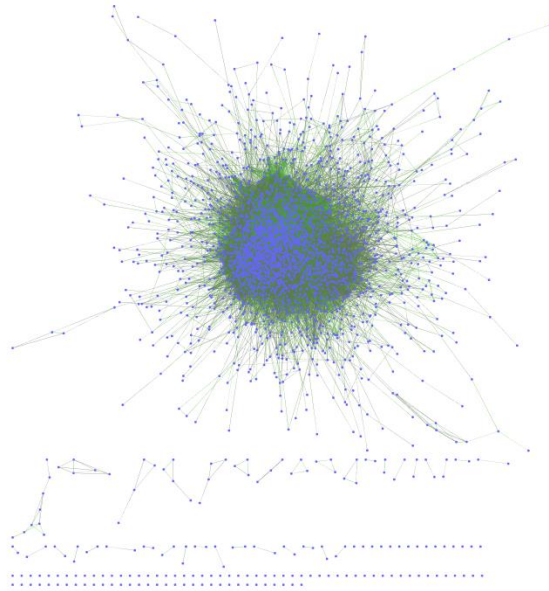
Ureaplasma urealyticum



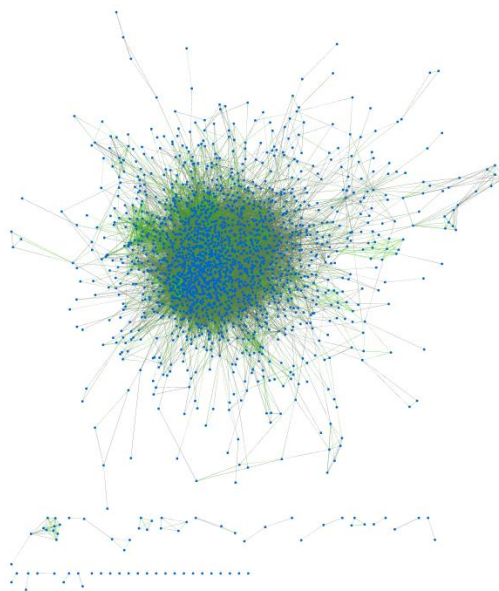
Ureaplasma parvum



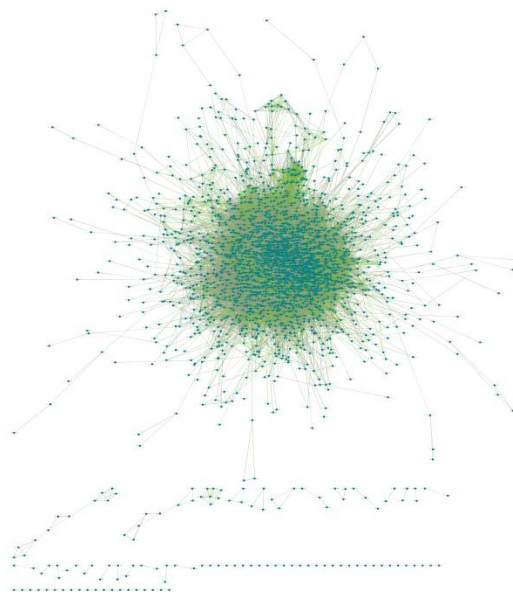
Corynebacterium glucurolyticum



Corynebacterium urealyticum

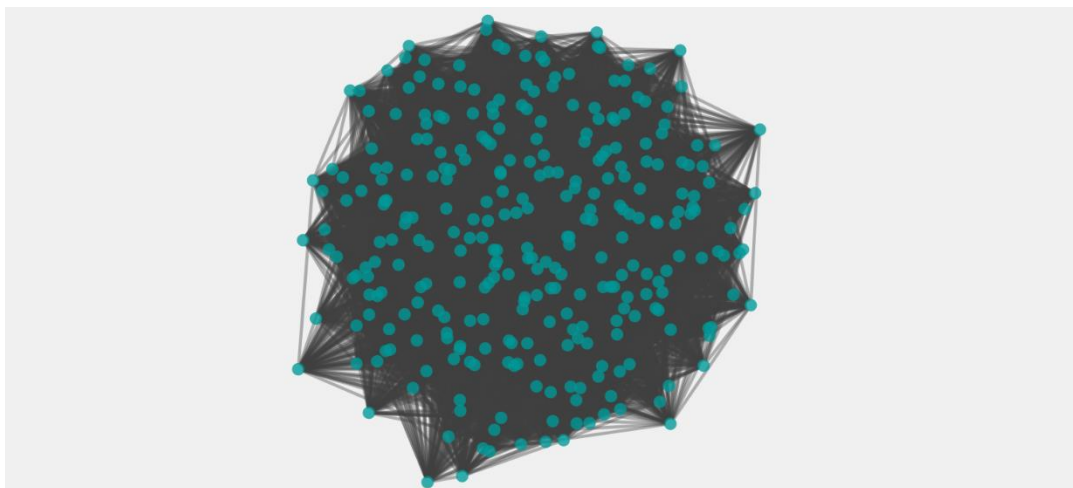


Staphylococcus aureus

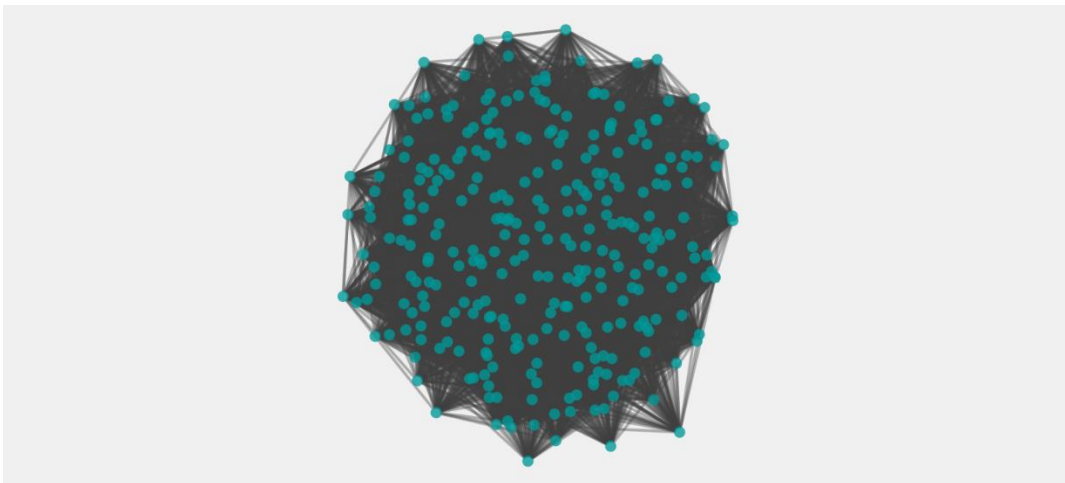


APÊNDICE B – Redes randômicas geradas para validação

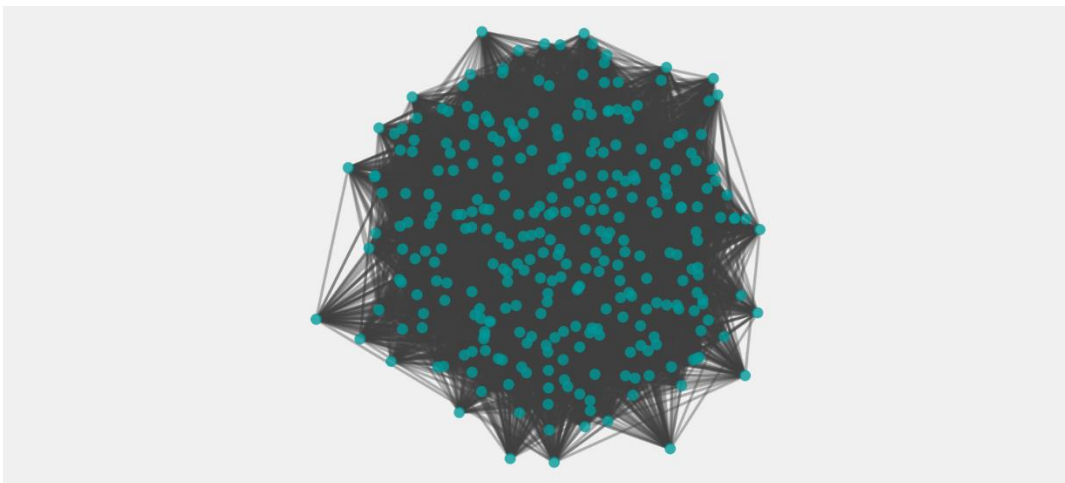
Random_1



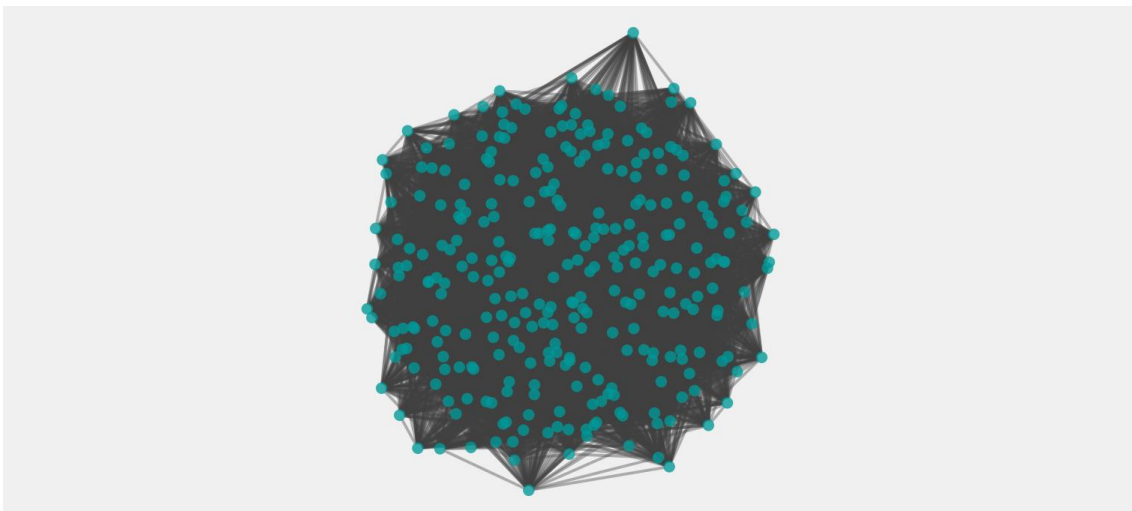
Random_2



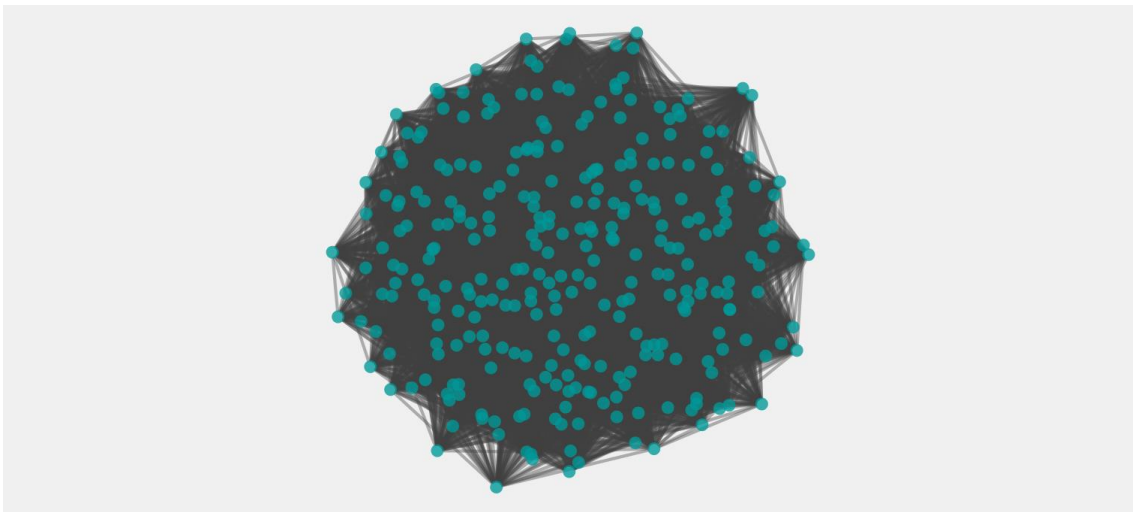
Random_3



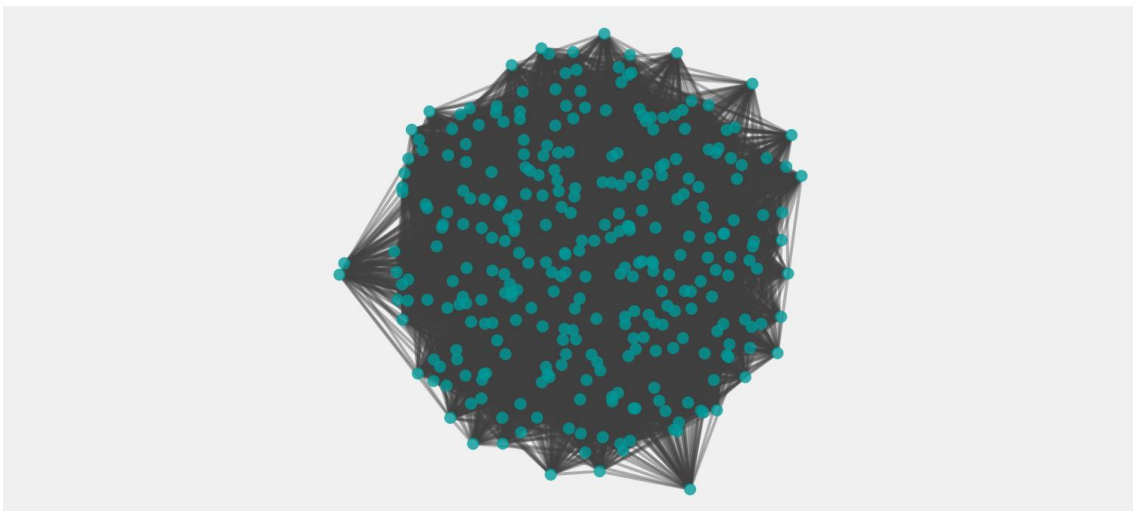
Random_4



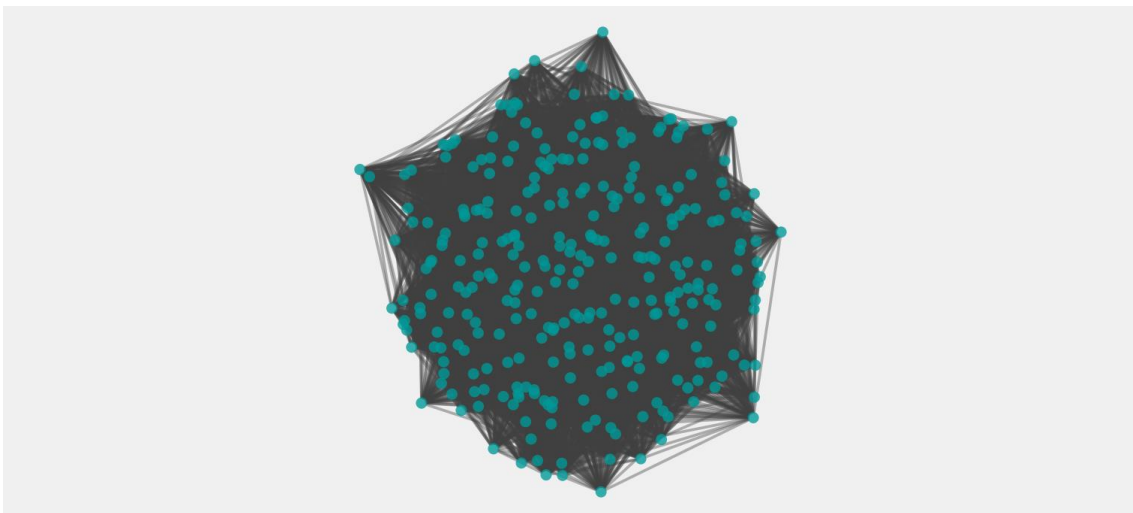
Random_5



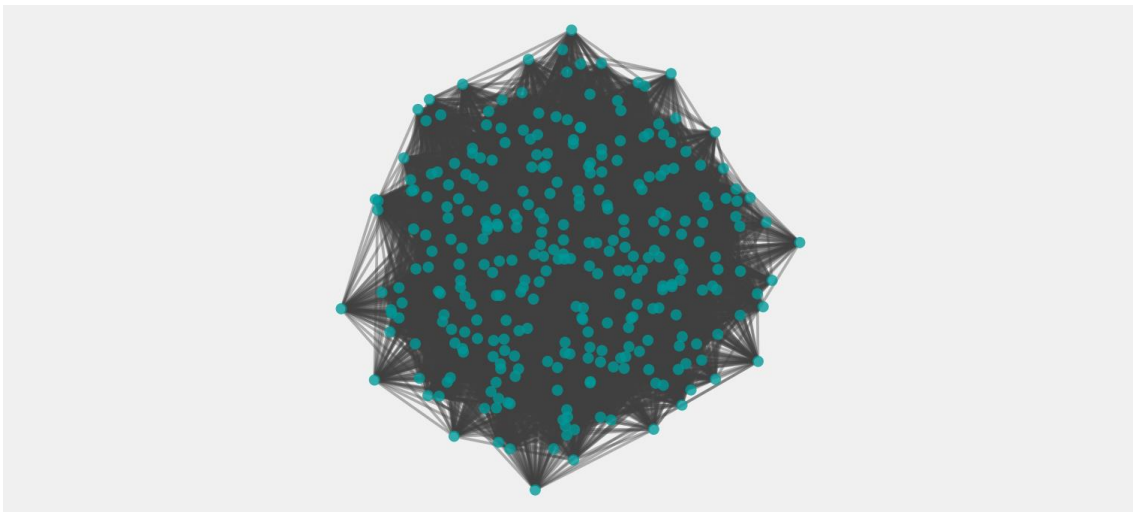
Random_6



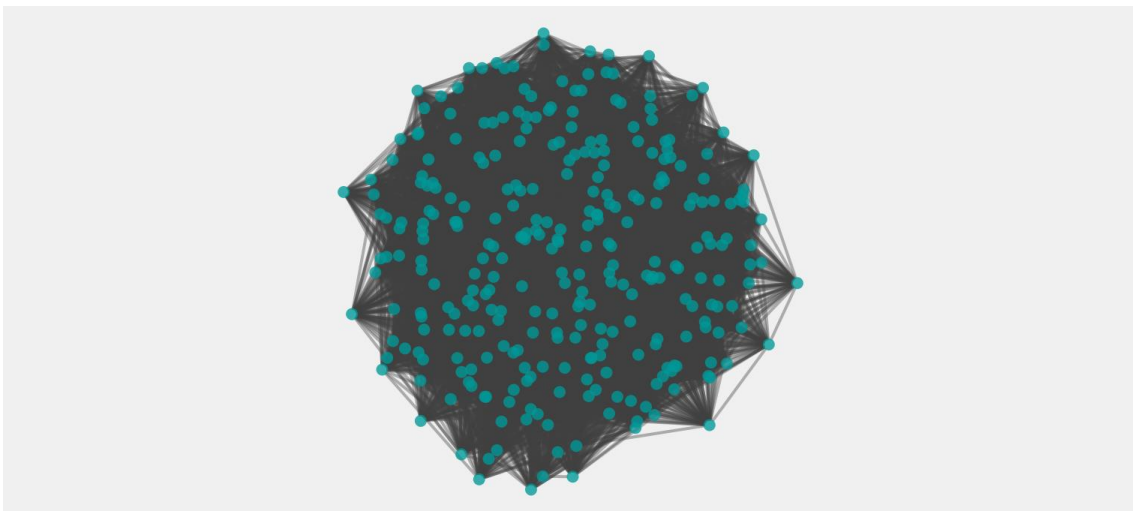
Random_7



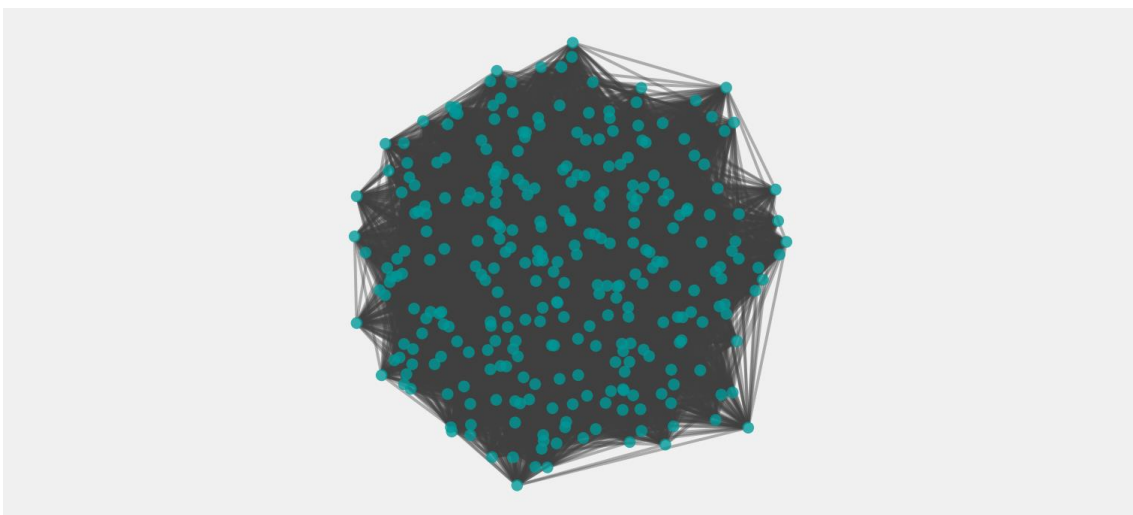
Random_8



Random_9



Random_10



APÊNDICE C – Tabela de nodos *U. diversum* e seus parâmetros correlacionados

name	AverageShortestPathLength	BetweennessCentrality	ClosenessCentrality	ClusteringCoefficient	Degree	NeighborhoodConnectivity
recA	1.44666667	0.02195515	0.69124424	0.46597403	176	97.34659091
dnaK	1.44666667	0.01965311	0.69124424	0.50081599	172	100.93604651
gyrB	1.48333333	0.00914417	0.6741573	0.52069475	165	102.90909091
topA	1.48666667	0.04033267	0.67264574	0.44890019	164	95.07317073
fusA	1.49666667	0.00571386	0.66815145	0.56330036	162	106.52469136
secY	1.51	0.00702488	0.66225166	0.54124002	158	104.43037975
parE	1.51	0.0102557	0.66225166	0.54548424	157	105.14649682
dnaA	1.51666667	0.01061387	0.65934066	0.51108354	156	101.98717949
rpoA	1.52333333	0.00640227	0.65645514	0.56777863	154	106.61038961
ligA	1.51666667	0.02248502	0.65934066	0.46889059	154	96.57792208
ftsH	1.52	0.01719259	0.65789474	0.56338149	153	106.78431373
gyrA	1.52333333	0.00802011	0.65645514	0.52073893	152	102.59868421
eno	1.53333333	0.00584775	0.65217391	0.54578366	151	104.78145695
tsf	1.55666667	0.00477283	0.64239829	0.58299776	150	107.77333333
rpoB	1.54333333	0.00476854	0.64794816	0.59740613	149	109.45637584
tig	1.53666667	0.01171842	0.65075922	0.56267005	149	106.48322148
rpsL	1.54333333	0.00467529	0.64794816	0.59441074	148	108.64189189
thrS	1.57333333	0.00561221	0.63559322	0.60441711	147	109.85714286
ftsY	1.56666667	0.00851663	0.63829787	0.54596127	146	105.07534247
secA	1.55333333	0.00618743	0.64377682	0.53150685	146	103.10958904
pheT	1.55666667	0.006101	0.64239829	0.58034955	146	107.95205479
parC	1.55666667	0.00960014	0.64239829	0.53379953	144	103.99305556
dnaG	1.57	0.00452533	0.63694268	0.5746892	144	107.96527778
frr	1.58	0.00406248	0.63291139	0.60917955	143	110.55244755
rpoC	1.56333333	0.00423428	0.63965885	0.61587708	143	110.82517483
pyrG	1.57333333	0.00721459	0.63559322	0.54506057	143	104.94405594
rpsB	1.59	0.01052659	0.62893082	0.61570415	141	109.85815603
pgk	1.57333333	0.01113335	0.63559322	0.54488349	141	104.56737589
rplD	1.59	0.00345836	0.62893082	0.64943474	140	112.34285714
prfA	1.59666667	0.00798001	0.6263048	0.57136899	139	107.21582734
glxT	1.59	0.00459041	0.62893082	0.60421228	139	109.21582734
infB	1.60666667	0.01392237	0.62240664	0.59923747	136	109.57352941
rpsI	1.61	0.00275657	0.62111801	0.66158098	135	112.19259259
fnt	1.60666667	0.00425854	0.62240664	0.61713654	135	110.3037037
rpsE	1.61	0.00251594	0.62111801	0.67646729	134	113.73880597
tyrS	1.61333333	0.00389691	0.61983471	0.63472113	134	112.0
pheS	1.61666667	0.00240712	0.6185567	0.67133743	133	115.13533835
adk	1.62666667	0.00383213	0.6147541	0.60287081	133	109.19548872
rplA	1.62666667	0.0021476	0.6147541	0.67891955	131	113.98473282
rplB	1.62333333	0.0021189	0.61601643	0.69042866	131	114.9389313
rpsG	1.65333333	0.00218011	0.60483871	0.67787716	130	112.32307692
rpsC	1.63333333	0.00231224	0.6122449	0.67847346	130	113.33846154
valS	1.63333333	0.00934964	0.6122449	0.58318426	130	108.2
rplK	1.62666667	0.00209222	0.6147541	0.69222384	129	114.52713178
rpsM	1.63666667	0.00177848	0.61099796	0.69585756	129	114.24806202
leuS	1.62	0.00362152	0.61728395	0.64922481	129	114.35658915
rplC	1.64333333	0.00157178	0.60851927	0.70632382	128	115.3984375
ksgA	1.62333333	0.00459866	0.61601643	0.60310039	128	109.375
obg	1.63666667	0.00366284	0.61099796	0.62672244	128	109.5546875
rpsK	1.64666667	0.00197765	0.60728745	0.69341332	127	113.44094488
rplQ	1.65333333	0.01393765	0.60483871	0.6504187	127	110.45669291
rolE	1.65333333	0.0019271	0.60483871	0.70679365	126	114.77777778
grpE	1.64666667	0.00538189	0.60728745	0.58514286	126	108.9047619
efp	1.63666667	0.00277483	0.61099796	0.65980952	126	114.57142857
rpsO	1.65666667	0.00276226	0.60362173	0.63703226	125	110.64
rpsR	1.66333333	0.00365223	0.6012024	0.63832258	125	110.616
serS	1.63666667	0.00384306	0.61099796	0.62322581	125	111.88
rpsD	1.67333333	0.00173438	0.59760956	0.69092578	124	112.7016129
rplM	1.65333333	0.00219888	0.60483871	0.68200892	124	112.59677419
rpsJ	1.66666667	0.00129187	0.6	0.72334372	122	115.93442623
map	1.66333333	0.00314798	0.6012024	0.66942149	122	114.36885246
tubB	1.63666667	0.01275723	0.61099796	0.61157025	122	108.64754098
ileS	1.65	0.00464168	0.60606061	0.61336088	121	110.94214876
infC	1.65666667	0.00196024	0.60362173	0.66473829	121	114.79338843
trmD	1.66	0.00350881	0.60240964	0.67286501	121	112.52066116
rplF	1.67666667	0.00143778	0.59642147	0.73579262	119	116.45378151
rpsH	1.68	0.00148056	0.5952381	0.72482552	119	114.73109244
rnc	1.66	0.00214108	0.60240964	0.64990742	119	113.44537815
rplO	1.69333333	0.00158207	0.59055118	0.71085035	118	113.88135593
dnaJ	1.66	0.00577815	0.60240964	0.56424743	118	106.61864407
pyk	1.65333333	0.00426463	0.60483871	0.55468637	118	103.54237288
mmE	1.65666667	0.01810685	0.60362173	0.62407649	118	110.55084746
rpmH	1.68666667	0.00203513	0.59288538	0.65428824	117	112.29059829
lysS	1.65333333	0.00309708	0.60483871	0.60683761	117	109.47008547
rplR	1.7	0.00122363	0.58823529	0.73808096	116	115.65517241
rplS	1.68333333	0.00241934	0.59405941	0.71034483	116	115.82758621
gidA	1.67666667	0.00645473	0.59642147	0.62698651	116	109.32758621
ftsH	1.65666667	0.00328548	0.60362173	0.56776612	116	105.29310345
rplT	1.69333333	0.00149652	0.59055118	0.70889613	114	115.51754386
tsaD	1.68	0.00587131	0.5952381	0.59540444	114	108.81578947
atpA	1.69	0.00243101	0.59171598	0.64696587	113	112.61061947
rpsP	1.72666667	0.00183974	0.57915058	0.70254183	112	112.00892857
alaS	1.67666667	0.01254849	0.59642147	0.59202059	112	109.0
pyrH	1.68666667	0.0028839	0.59288538	0.59507722	112	109.00892857
rsmH	1.70333333	0.00591113	0.58708415	0.61100386	112	109.33035714
rplV	1.72	0.00130818	0.58139535	0.73677314	111	116.66666667
tpiA	1.67666667	0.00979589	0.59642147	0.59866555	110	108.54545455
nr	1.68666667	0.00395121	0.59288538	0.62718932	110	110.82727273
rplJ	1.71666667	0.00140953	0.58252427	0.7358138	109	114.1559633
UUR_10_0641	1.68333333	0.00692566	0.59405941	0.56082229	109	104.22018349
exo	1.68333333	0.01481703	0.59405941	0.50744202	108	97.66666667
trpS	1.70666667	0.00754839	0.5859375	0.6443406	108	113.92592593
metK	1.70333333	0.00625736	0.58708415	0.57805469	108	107.47222222

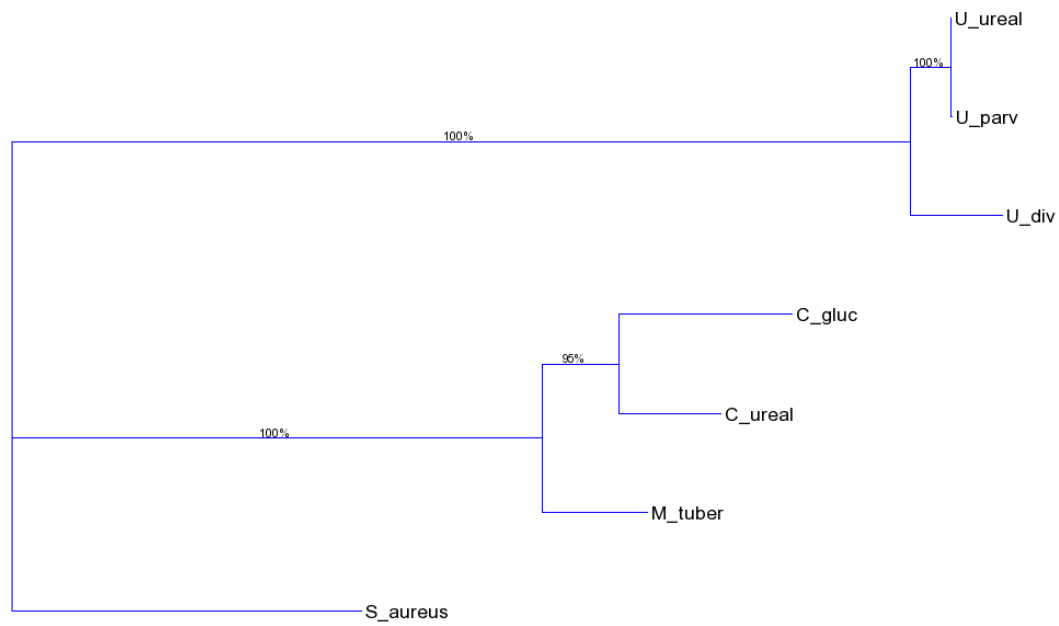
rpsQ	1.74666667	0.00105751	0.57251908	0.74824798	106	113.28301887
truA	1.74	0.00195736	0.57471264	0.65211141	106	110.73584906
rplN	1.72666667	8.7613E-4	0.57915058	0.7787528	104	120.47115385
argS	1.74	0.01020092	0.57471264	0.57776509	103	106.30097087
rplI	1.75666667	0.00278822	0.56925996	0.67656577	103	110.68932039
ppa	1.72333333	0.004848	0.58027079	0.52027413	103	101.84466019
gmK	1.72	0.00531144	0.58139535	0.53737138	102	103.29411765
infA	1.73666667	0.00157877	0.57581574	0.66792079	101	113.20792079
rpsF	1.79666667	0.0023494	0.55658627	0.63881188	101	108.89108911
cysS	1.73	0.00275852	0.57803468	0.63940594	101	112.57425743
atpD	1.74666667	0.00131558	0.57251908	0.68868687	100	115.45
ruvB	1.72	0.00389271	0.58139535	0.53988868	99	101.62626263
rplW	1.79666667	7.337E-4	0.55658627	0.78887457	97	115.97938144
dnaN	1.73	0.00940131	0.57803468	0.47828947	96	93.78125
lon	1.74666667	0.00248109	0.57251908	0.5879386	96	106.75
uvrA	1.74	0.00237121	0.57471264	0.6031355	95	106.13684211
rplX	1.79666667	5.7828E-4	0.55658627	0.80347747	94	114.81914894
upp	1.76333333	0.00439288	0.56710775	0.53191489	94	102.38297872
rpmE	1.79333333	0.0030237	0.55762082	0.66346374	94	108.76595745
rplL	1.82	8.4461E-4	0.54945055	0.79359514	93	113.95698925
uvrB	1.75	0.00172141	0.57142857	0.62155213	93	108.31182796
aspS	1.76333333	0.00265552	0.56710775	0.67952314	93	116.29032258
greA	1.76333333	0.00268308	0.56710775	0.57572862	92	105.94565217
atpG	1.79333333	0.00156737	0.55762082	0.65103785	91	112.08791209
hisS	1.78333333	0.00291902	0.56074766	0.64126984	91	112.27472527
tmk	1.74666667	0.01313655	0.57251908	0.52359551	90	99.51111111
UUR10_0460	1.76	0.00634769	0.56818182	0.52298264	89	98.26966292
UUR10_0434	1.77666667	0.00635224	0.56285178	0.6392372	88	110.38636364
hrcA	1.79	0.00350536	0.55865922	0.54173669	85	100.89411765
rpmG	1.84333333	0.00147108	0.54249548	0.71227768	84	111.63095238
ung	1.78	0.00261494	0.56179775	0.589214	84	106.91666667
tkt	1.8	0.00223947	0.55555556	0.60946224	83	106.24096386
def_1	1.83	0.00128606	0.54644809	0.68413129	82	110.35365854
rplM	1.84333333	5.9429E-4	0.54249548	0.76582278	80	118.2125
UUR10_0344	1.8	0.00264719	0.55555556	0.56708861	80	105.225
apt_2	1.81	0.01066563	0.55248619	0.43784486	79	89.89873418
recR	1.80333333	0.00271571	0.55452865	0.63193768	79	106.34177215
tis	1.87	5.9229E-4	0.53475936	0.72768582	79	112.94936709
mmnA	1.84	0.00898396	0.54347826	0.71262071	78	109.47435897
fpg	1.79	0.0073774	0.55865922	0.55399863	77	98.84415584
plsX	1.82666667	0.00142036	0.54744526	0.62912281	76	107.36842105
pth	1.82333333	0.00100497	0.54844607	0.64385965	76	111.17105263
rluA	1.85333333	0.00281971	0.53956835	0.61441441	75	104.12
rpmA	1.86333333	0.0011924	0.53667263	0.62717512	74	106.37837838
ackA	1.84333333	0.00113724	0.54249548	0.59903739	74	105.32432432
UUR10_0707	1.85333333	0.00131162	0.53956835	0.61035008	73	109.67123288
nfo	1.83666667	0.00128222	0.54446461	0.63622527	73	104.10958904
trmB	1.88	0.00578039	0.53191489	0.65334855	73	101.69863014
rpmC	1.92666667	3.0809E-4	0.51903114	0.87206573	72	116.65277778
trxB	1.86	0.00142737	0.53763441	0.59919517	71	106.1971831
ssb	1.87333333	0.00142062	0.53380783	0.63380282	71	111.53521127
thiI	1.85333333	0.00233272	0.53956835	0.72354125	71	111.90140845
rpmJ	1.96666667	2.7882E-4	0.50847458	0.85038363	69	114.68115942
ruvA	1.84	9.3201E-4	0.54347826	0.63171355	69	106.33333333
rplU	1.91	0.00357406	0.52356021	0.76470588	69	112.07246377
UUR10_0139	1.90333333	8.0936E-4	0.52539405	0.69479966	69	114.5942029
UUR10_0577	1.83666667	0.02336894	0.54446461	0.56138107	69	99.10144928
UUR10_0590	1.86333333	0.00214617	0.53667263	0.66505707	68	113.95588235
hpt	1.89666667	0.0011554	0.52724077	0.58228438	66	109.09090909
lgt	1.87	0.00151434	0.53475936	0.60606061	66	102.57575758
UUR10_0694	1.88	0.00139192	0.53191489	0.59471154	65	106.04615385
rimH	1.90666667	6.5481E-4	0.52447552	0.78461538	65	111.64615385
UUR10_0059	1.93	0.00638907	0.51813472	0.7827381	64	113.046875
tdk	1.87333333	0.00188582	0.53380783	0.55145929	63	99.34920635
cmk	1.87333333	0.00179436	0.53380783	0.53622422	62	98.11290323
UUR10_0099	1.86333333	0.00143215	0.53667263	0.63606557	61	105.32786885
UUR10_0491	1.9	0.00129421	0.52631579	0.58196721	61	106.1147541

UUR10_0150	1.97666667	3.4125E-4	0.50590219	0.82033898	60	114.16666667
acpS	1.91	0.00130793	0.52356021	0.52776154	59	96.20338983
smc	1.88333333	0.00781995	0.53097345	0.56684815	58	103.05172414
UUR10_0213	1.90333333	0.0092318	0.52539405	0.5353902	58	94.96551724
UUR10_0212	1.91	0.00345924	0.52356021	0.55716878	58	96.74137931
ureA	1.90666667	0.00422942	0.52447552	0.54197995	57	97.98245614
deoC	1.89333333	0.01062111	0.52816901	0.42230576	57	87.07017544
pepF_2	1.89333333	0.00712823	0.52816901	0.66353383	57	111.10526316
rpsT	1.96	7.484E-4	0.51020408	0.67532468	56	111.51785714
UUR10_0019	1.88666667	0.00778198	0.53003534	0.46680643	54	96.11111111
UUR10_0148	2.05	1.1329E-4	0.48780488	0.86647315	53	111.90566038
atpH	2.05	1.1329E-4	0.48780488	0.86647315	53	111.90566038
cdsA	1.91333333	6.2493E-4	0.52264808	0.64513788	53	107.0
UUR10_0176	1.94333333	5.543E-4	0.51457976	0.67607843	51	111.98039216
cdd	1.94666667	0.00457887	0.51369863	0.40156863	51	83.7254902
UUR10_0281	1.92666667	0.00128474	0.51903114	0.56784314	51	104.1372549
pgsA	1.96	0.00106645	0.51020408	0.5952381	49	96.46938776
era	1.94	0.00583231	0.51546392	0.4079556	47	85.9787234
hprK	1.98	7.174E-4	0.50505051	0.61054579	47	101.63829787
UUR10_0624	1.91666667	0.00153015	0.52173913	0.54589372	46	100.23913043
UUR10_0425	1.99666667	8.4625E-4	0.50083472	0.52727273	45	100.73333333
fba	1.98333333	3.371E-4	0.50420168	0.66384778	44	113.20454545
rpsU	2.00333333	0.0017532	0.49916805	0.57082452	44	101.56818182
ispE	1.96333333	4.9357E-4	0.50933786	0.57918051	43	101.27906977
UUR10_0168	2.00333333	0.00763745	0.49916805	0.303433	43	67.1627907
gpmI	1.97666667	2.9021E-4	0.50590219	0.70615563	42	108.23809524
UUR10_0319	2.01	0.00107356	0.49751244	0.57723577	42	101.64285714
gatB	1.97666667	0.00109175	0.50590219	0.46282051	40	94.5
polC	1.97	5.0833E-4	0.50761421	0.66025641	40	106.125
UUR10_0692	1.96	0.0016479	0.51020408	0.52051282	40	94.775
ecfA	2.03666667	4.3851E-4	0.49099836	0.63022942	39	105.76923077
UUR10_0630	2.04	4.0709E-4	0.49019608	0.64295875	38	106.65789474
udk	2.02333333	2.8035E-4	0.49423394	0.64114114	37	100.40540541
recO	2.01666667	0.00177159	0.49586777	0.54954955	37	94.43243243
nadE	1.99666667	2.6561E-4	0.50083472	0.67267267	37	104.18918919
UUR10_0362	2.01333333	0.00202348	0.49668874	0.5031746	36	96.05555556
atpC	2.15333333	4.892E-5	0.46439628	0.86386555	35	109.28571429
UUR10_0629	2.13	8.427E-5	0.46948357	0.80392157	34	114.73529412
dinB	1.99333333	0.0059371	0.50167224	0.75483871	31	104.4516129

dcm	2.00333333	2.515E-4	0.49916805	0.71612903	31	114.74193548
UUR10_0123	2.05666667	0.00116947	0.48622366	0.36989247	31	83.29032258
rmhC	2.02333333	2.443E-4	0.49423394	0.64301075	31	104.5483871
gatA	2.02666667	6.5143E-4	0.49342105	0.46551724	29	90.86206897
UUR10_0353	2.05	6.62E-4	0.48780488	0.40640394	29	79.51724138
deoD	2.06333333	0.00213916	0.48465267	0.38423645	29	65.20689655
UUR10_0280	2.11	7.8018E-4	0.47393365	0.39162562	29	83.06896552
atpB	2.09333333	1.7374E-4	0.47770701	0.74074074	27	101.7037037
UUR10_0117	2.06666667	4.8291E-4	0.48387097	0.7008547	27	104.22222222
rbfA	2.09666667	0.00153166	0.47694754	0.45299145	27	86.37037037
UUR10_0067	2.09	3.7839E-4	0.4784689	0.72307692	26	115.30769231
UUR10_0066	2.09	3.7839E-4	0.4784689	0.72307692	26	115.30769231
UUR10_0352	2.09333333	6.7132E-4	0.47770701	0.38043478	24	74.5
UUR10_0133	2.07333333	0.00112944	0.48231511	0.52173913	24	96.25
UUR10_0589	2.04666667	3.921E-4	0.48859935	0.69202899	24	112.08333333
UUR10_0269	2.08666667	8.9633E-4	0.47923323	0.4673913	24	94.16666667
ureG	2.14333333	6.0645E-4	0.46656299	0.53333333	21	60.0952381
UUR10_0317	2.15666667	4.3717E-4	0.46367852	0.45238095	21	85.19047619
UUR10_0036	2.11333333	0.00141221	0.47318612	0.45789474	20	88.15
pdp	2.16	8.1061E-4	0.46296296	0.42631579	20	57.4
UUR10_0335	2.11666667	3.3374E-4	0.47244094	0.37426901	19	76.21052632
rsgA	2.21	2.98E-5	0.45248869	0.74853801	19	115.73684211
UUR10_0270	2.08666667	5.79E-4	0.47923323	0.36842105	19	83.10526316
mtmN	2.14	4.6041E-4	0.46728972	0.50980392	18	96.44444444
UUR10_0193	2.13333333	7.03E-5	0.46875	0.57352941	17	108.52941176
UUR10_0216	2.17	0.00230953	0.46082949	0.29166667	16	56.25
UUR10_0118	2.18666667	2.0346E-4	0.45731707	0.475	16	78.1875
plsY	2.2	0.00674442	0.45454545	0.475	16	87.25
rpoE	2.19	2.762E-5	0.456621	0.70833333	16	104.25
UUR10_0089	2.14	5.755E-5	0.46728972	0.625	16	110.1875
UUR10_0691	2.15333333	2.9912E-4	0.46439628	0.35238095	15	79.53333333
ureC	2.23	1.2193E-4	0.44843049	0.84615385	14	56.28571429
ureE	2.23333333	8.65E-5	0.44776119	0.84615385	14	54.92857143
UUR10_0264	2.21666667	2.9114E-4	0.45112782	0.42857143	14	82.42857143
mgfE	2.16333333	1.8942E-4	0.46224961	0.41758242	14	82.78571429
ureB	2.26333333	4.442E-5	0.44182622	0.91025641	13	52.38461538
ureF	2.26333333	4.442E-5	0.44182622	0.91025641	13	52.38461538
ureD	2.26333333	4.442E-5	0.44182622	0.91025641	13	52.38461538
msrA	2.21333333	4.186E-5	0.45180723	0.62820513	13	94.46153846
UUR10_0014	2.27	4.0025E-4	0.44052863	0.62820513	13	42.23076923
UUR10_0013	2.27	4.0025E-4	0.44052863	0.62820513	13	42.23076923
UUR10_0015	2.27	4.0025E-4	0.44052863	0.62820513	13	42.23076923
UUR10_0579	2.49333333	4.8738E-4	0.40106952	0.65384615	13	31.92307692
UUR10_0105	2.29	0.00387974	0.43668122	0.24242424	12	34.75
UUR10_0075	3.0	0.00154736	0.33333333	0.53333333	10	8.2
UUR10_0023	2.40333333	0.00508663	0.41608877	0.6	10	32.3
UUR10_0022	2.40333333	0.00508663	0.41608877	0.6	10	32.3
UUR10_0653	2.70666667	0.00358305	0.36945813	0.28888889	10	12.3
UUR10_0098	2.21666667	1.46E-5	0.45112782	0.62222222	10	88.8
UUR10_0041	2.25333333	0.00228334	0.44378698	0.55555556	9	50.44444444
UUR10_0037	2.41333333	2.3304E-4	0.41436464	0.58333333	9	26.55555556
UUR10_0656	2.22666667	0.01583767	0.4491018	0.36111111	9	57.44444444
UUR10_0480	2.87333333	8.3903E-4	0.34802784	0.63888889	9	10.88888889
UUR10_0366	2.57666667	0.00306105	0.38809832	0.63888889	9	19.88888889
UUR10_0638	2.32	3.57E-6	0.43103448	0.94444444	9	88.0
UUR10_0278	2.26666667	1.3423E-4	0.44117647	0.44444444	9	65.77777778
UUR10_0012	2.32666667	5.227E-5	0.42979943	0.77777778	9	47.66666667
UUR10_0538	2.32666667	5.227E-5	0.42979943	0.77777778	9	47.66666667
UUR10_0688	2.25333333	9.145E-5	0.44378698	0.33333333	9	78.77777778
UUR10_0040	2.28	6.772E-5	0.43859649	0.71428571	8	56.125
UUR10_0076	3.13	1.6425E-4	0.31948882	0.78571429	8	8.75
UUR10_0119	2.30666667	1.5453E-4	0.43352601	0.35714286	8	56.0
UUR10_0442	2.75	0.00174792	0.36363636	0.75	8	17.625
UUR10_0488	2.26666667	1.904E-5	0.44117647	0.64285714	8	106.0
UUR10_0323	2.36	1.5436E-4	0.42372881	0.5	8	53.875
UUR10_0039	2.37666667	2.509E-5	0.42075736	0.85714286	7	47.71428571
UUR10_0658	2.45333333	0.0042959	0.4076087	0.47619048	7	30.57142857
UUR10_0443	3.24666667	0.0	0.30800821	1.0	7	9.14285714
UUR10_0311	2.25666667	3.45E-6	0.44313146	0.76190476	7	112.71428571
UUR10_0073	2.65333333	0.0025032	0.37688442	0.33333333	7	22.42857143
UUR10_0657	3.12	7.61E-5	0.32051282	0.73333333	6	6.5
UUR10_0104	2.36333333	2.741E-4	0.42313117	0.53333333	6	58.0
UUR10_0533	3.01	9.0833E-4	0.33222591	0.2	6	7.16666667
UUR10_0435	2.48333333	2.8954E-4	0.40268456	0.46666667	6	47.16666667
UUR10_0683	2.34333333	0.00230777	0.42674253	0.2	6	49.66666667
UUR10_0592	2.32666667	2.0514E-4	0.42979943	0.26666667	6	57.0
UUR10_0107	2.38666667	2.509E-5	0.41899441	0.7	5	63.4
UUR10_0655	3.13	0.0	0.31948882	1.0	5	7.4
UUR10_0364	2.53	0.00120084	0.39525692	0.4	5	37.2
UUR10_0256	2.57666667	4.835E-5	0.38809832	0.6	5	31.2
UUR10_0255	2.57666667	4.835E-5	0.38809832	0.6	5	31.2
UUR10_0464	2.34666667	0.0	0.42613636	1.0	5	86.0
UUR10_0660	3.13	0.0	0.31948882	1.0	5	7.4
UUR10_0687	2.37333333	5.43E-6	0.42134831	0.5	5	97.6
UUR10_0273	2.41	1.28E-5	0.41493776	0.7	5	63.2
UUR10_0453	2.35333333	0.0	0.42492918	1.0	5	117.4
UUR10_0217	2.4	3.318E-5	0.41666667	0.66666667	4	73.0
UUR10_0267	2.47333333	0.0	0.40431267	1.0	4	60.25
UUR10_0534	2.78666667	2.9523E-4	0.35885167	0.66666667	3	26.0
UUR10_0328	2.58333333	0.0	0.38709677	1.0	3	57.0
UUR10_0535	2.78666667	2.9523E-4	0.35885167	0.66666667	3	26.0
UUR10_0559	2.66333333	0.0	0.37546934	1.0	3	54.33333333
UUR10_0528	2.45	0.0	0.40816327	1.0	3	106.33333333
UUR10_0673	2.4	2.06E-6	0.41666667	0.66666667	3	103.33333333
UUR10_0650	2.82	5.569E-4	0.35460993	0.33333333	3	24.66666667
UUR10_0214	2.84	0.0	0.35211268	1.0	2	37.0
UUR10_0169	2.61666667	0.0	0.38216561	1.0	2	75.5
UUR10_0578	2.8	0.0	0.35714286	1.0	2	41.0
UUR10_0433	2.53666667	4.2E-7	0.39421813	0.0	2	100.0
UUR10_0027	2.58666667	0.0	0.38659794	0.0	1	141.0
UUR10_0449	3.19666667	0.0	0.31282586	0.0	1	16.0
UUR10_0608	2.89	0.0	0.34602076	0.0	1	57.0

APÊNDICE D – Árvores 16s controle na análise de transferência horizontal de genes.

Urease



Purinas e Pirimidinas

