

MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Doutorado em Biologia computacional e sistemas

EXPLORAÇÃO DA DIVERSIDADE GENÔMICA, TAXONÔMICA E
FUNCIONAL DOS GÊNEROS WOLBACHIA E KLEBSIELLA

BRUNO GABRIEL NASCIMENTO ANDRADE

Rio de Janeiro
Maio de 2016

INSTITUTO OSWALDO CRUZ
Programa de Pós-Graduação em Biologia computacional e sistemas

Bruno Gabriel Nascimento Andrade

Exploração da diversidade genômica, taxonômica e funcional dos gêneros
Wolbachia e Klebsiella

Tese apresentada ao Instituto Oswaldo Cruz como
parte dos requisitos para obtenção do título de
Mestre ou Doutor em Biologia computacional e
sistemas

Orientadora: Prof. Dr. Ana Carolina Paulo Vicente

RIO DE JANEIRO
Maio de 2016

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

A553 Andrade, Bruno Gabriel Nascimento

Exploração da diversidade genômica, taxonômica e funcional dos gêneros Wolbachia e Klebsiella / Bruno Gabriel Nascimento Andrade. – Rio de Janeiro, 2016.

xvii, 133 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2016.

Bibliografia: f. 66-76

1. Taxonomia genômica. 2. Análise funcional. 3. Klebsiella. 4. Wolbachia. I. Título.

CDD 578.012

Agradecimentos

Agradeço a minha orientadora, Dra Ana Carolina Paulo Vicente, sem sua orientação e conselhos este trabalho jamais teria existido. Agradeço a meu colaborador e amigo durante o período de estágio na Universiteit Utrecht, Dr. Bas E. Dutilh, agradeço também a pós-graduação de Biologia computacional e sistemas e a todos os meus colegas que diretamente ou indiretamente tiveram papéis no desenvolvimento de meu trabalho. Agradeço também a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo auxílio financeiro.

Dedico este trabalho a minha família,
obrigado pelo suporte e apoio

“Aqui no entanto nós não olhamos para trás por muito tempo, Nós continuamos seguindo em frente, abrindo novas portas e fazendo coisas novas, Porque somos curiosos...e a curiosidade continua nos conduzindo por novos caminhos. Siga em frente.” Walt Disney

INSTITUTO OSWALDO CRUZ

Exploração da diversidade genômica, taxonômica e funcional dos gêneros *Wolbachia* e *Klebsiella*

RESUMO

TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Bruno Gabriel Nascimento Andrade

Abordagens filogenômicas, funcionais e taxonômicas *in silico* têm sido utilizadas na exploração da diversidade de espécies de microrganismos, na solução de relações filogenéticas e evolutivas, utilizando a informação genômica de forma global. Abordagens, que incluem a identidade média de nucleotídeos (ANI), tipagem da sequência multilocus (MLST) e hibridação DNA-DNA (GGD), consubstanciam a taxonomia genômica de bactérias, sendo a abordagem atual para a identificação e classificação principalmente de organismos não cultiváveis. O gênero de bactérias *Wolbachia* contém endossimbiontes obrigatórios e facultativos que infectam mais de 40% de todas as espécies de artrópodes e várias espécies de nematóides. Bactérias deste gênero, são classificadas como pertencentes a espécie *Wolbachia pipientis*. A partir de análises genéticas utilizando apenas três genes, *ftsZ*, *wsP* e 16S, bactérias desta espécie são classificadas em 16 supergrupos filogenéticos. Recentemente, utilizando algumas abordagens de taxonomia genômica, foi sugerido que estes supergrupos filogenéticos corresponderiam a várias espécies. *Klebsiella pneumoniae*, uma espécie de bactéria de vida livre e também um patógeno oportunista, era classificada, com base nos genes *parC* e *gyrA*, em três grupos filogenéticos KP-I, KP-II-A, KP-II-B e KP-III. Recentemente, aplicando um conjunto de sete genes e identidade média de nucleotídeos foi definido que estes grupos filogenéticos corresponderiam a três espécies, *K. pneumoniae*, *K. quasipneumoniae* e *K. variicola*. Neste estudo aplicamos filogenômica e abordagens de taxonomia genômica para reanalisar e explorar, sob o ponto de vista do conteúdo genômico global, a diversidade e classificação destes dois gêneros de bactérias. Recuperamos todos os genes codificadores das proteínas ribossomais (rMLST) e genes compartilhados por todos os genomas analisados (cgMLST), de 26 genomas de *Wolbachia* e 91 genomas de *Klebsiella*. Determinamos a identidade média de nucleotídeos (gANI), hibridação DNA-DNA *in silico* (GGD) e alinhamento da fração (AF). A diversidade funcional do gênero *Wolbachia* foi explorada por análise de enriquecimento funcional. Geramos, através de sequenciamento de alto desempenho, o genoma de um isolado clínico de *K. variicola* (BZ19) e um isolado clínico de *K. quasipneumoniae* (K142). As análises filogenômicas com o gênero *Wolbachia* geraram os mesmos 16 supergrupos exceto pelo isolado wVulC, classificado previamente no supergrupo B, que definiu um

novo supergrupo. Agregando as abordagens rMLST, cgMLST, gANI e GGD ao esquema prévio de taxonomia genômica de *Wolbachia* ficou claro que as bactérias dos supergrupos pertencem a distintas espécies de *Wolbachia* e que o isolado wWu1C seria uma nova espécie, a qual propomos o nome '*Candidatus Wolbachia vulgaris*'. Existem diferenças funcionais significativas entre os endossimbiontes obrigatórios e facultativos, principalmente quanto à presença de genes associados com fagos, prófagos e elementos transponíveis nos genomas dos facultativos. As análises de filogenômica e de taxonomia genômica definiram de forma robusta as espécies *K. pneumoniae*, *K. variicola* e *K. quasipneumoniae*. 39 genomas de isolados previamente classificados como *K. pneumoniae* apresentaram métricas abaixo dos limites de delimitação de espécies (GGD < 70, cgMLST < 95, gANI < 96) quando comparados com o genoma da linhagem tipo *K. pneumoniae* HS11286T sendo que 14/39 e 25/39 apresentaram métricas compatíveis com a linhagem tipo de *K. quasipneumoniae* 01030T e *K. variicola* DSM15968T, respectivamente. O genoma da *Klebsiella* sp. 10982, corresponderia a uma nova espécie relacionada a *K. variicola*.

INSTITUTO OSWALDO CRUZ

Exploração da diversidade genômica, taxonômica e funcional dos gêneros *Wolbachia* e *Klebsiella*

Abstract

PHD THESIS IN BIOLOGIA COMPUTACIONAL E SISTEMAS

Bruno Gabriel Nascimento Andrade

Phylogenomic, functional and genomic taxonomic approaches have been used in the microorganisms species exploration and in the solution of evolutionary and phylogenetic relationships using whole genomic information. Approaches like Average nucleotide identity (ANI), Multilocus sequence typing (MLST) and DNA-DNA hybridization (GGD) are applied to the genomic taxonomy that has now been used in the identification and classification of organisms, even non-cultivable ones. The bacterial genus *Wolbachia* contains obligate and facultative endosymbionts that infects more than 40% of arthropod and several nematode species. Bacteria from this genus are classified in 16 phylogenetic supergroups (A-Q) by genetic analyses based on three genes, *ftsZ*, *wsP* e 16S. It was recently suggested that these phylogenetic supergroups correspond to several bacterial species. Similarly, the free living and opportunistic pathogen *Klebsiella pneumoniae* had been classified in three phylogenetic groups, KP-I, KP-II and KP-III by genetic analyses based on two genes, *parC* and *gyrA*. Recently, it was defined based on the similarity of seven intrinsic genes and ANI, that these phylogenetic groups corresponds to three bacterial species, *K. pneumoniae*, *K. quasipneumoniae* e *K. variicola*. In this study, we applied phylogenomic and genomic taxonomy approaches to reanalyze and explore, using whole genomic content, the diversity and classification of these two genus. We retrieved all ribosomal protein subunit coding genes (rMLST) and genes (cgMLST) shared by 26 *Wolbachia* and by 91 *Klebsiella* genomes. We determined the Average nucleotide identity (gANI), digital DNA-DNA hybridization (GGD) and Alignment of fraction (AF). The functional diversity of the *Wolbachia* genus was explored by the application of an enrichment/depletion approach. We generate, using high throughput sequencing, the genome of the first *K. variicola* from a clinical case (BZ19) as well as a genome of a *K. quasipneumoniae* (K142). Phylogenomic analyses with *Wolbachia* determined 5 supergroups already described, the exception was wVulC, previously classified as a supergroup B *Wolbachia*, revealed here as a new supergroup. Adding the rMLST, cgMLST, gANI and GGD to a taxonomical scheme for *Wolbachias* it became clear that these genomes belongs to distinct species and the wVulC isolate is an undescribed species, which we will propose the name *Candidatus Wolbachia vulgaris*'. There are functional differences between obligate and facultative endosymbionts, most of these differences resides on the presence of genes associated to the Phages, prophages, transposable elements subsystem in the facultative genomes. In relation to *Klebsiella*, phylogenomic and genomic taxonomy analyses showed that *K. pneumoniae*, *K. variicola* e *K. quasipneumoniae* are well defined species. Of all genomes annotated as *K. pneumoniae*, 39 presented metrics bellow the

species delineation limit (GGD < 70, cgMLST <95, gANI < 96) with the type lineage HS11286^T, 14/39 presented metrics compatible with *K. quasipneumoniae* 01030T and 25/39 with the species *K. variicola*. The genome *Klebsiella* sp. 10982, would correspond to a new species related to *K. variicola*.

Sumário

1 - INTRODUÇÃO.....	1
1.1 - Filogenômica.....	1
1.2 - Taxonomia genômica.....	2
1.3 - Wolbachia pipientis.....	4
1.3.1 - Características genômicas.....	5
1.3.2 – Filogenia e supergrupos filogenéticos.....	6
1.3.3 - Fenótipos induzidos.....	7
1.3.4 - Status taxonômico do gênero Wolbachia.....	7
1.4 - Klebsiella.....	8
1.4.1 – Características genômicas.....	8
1.4.3 - Filogenia e status taxonômico.....	9
2 - OBJETIVOS.....	11
2.1 - Objetivo principal.....	11
2.2 - Objetivos específicos.....	11
3 - MÉTODOS.....	12
3.1 - Wolbachia.....	12
3.2 - Klebsiella.....	13
3.3 - Sequenciamento e montagem dos genomas Bz19 e K142.....	17
3.4 - Identificação dos genomas core e pan.....	17
3.5 – Filogenômica.....	18
3.6 - Taxonomia genômica.....	19
3.7 - Análise de enriquecimento/empobrecimento funcional.....	20
3.8 - Determinação do código genético e identificação de genes mitocondriais adquiridos por HGT.....	21
4 - RESULTADOS E DISCUSSÃO.....	22
4.1 - Wolbachia.....	22
4.1.1 – Filogenômica.....	22
4.1.2 - Análises de taxonomia genômica.....	33
4.1.3 - Descrição da espécie 'Candidatus Wolbachia vulgaris'.....	40
4.1.4 - Grupos gênicos e composição funcional do pan-genoma.....	40
4.1.6 - Identificação de genes adquiridos por HGT.....	46
4.2 - Klebsiella.....	52
4.2.1 – Sequenciamento e identificação dos genomas Bz19 e K142.....	53
4.2.2 - Filogenômica.....	53
4.2.3 – Taxonomia genômica do gênero Klebsiella.....	54
4.2.4 - Caracterização do plasmídeo pl9.....	60
5 - CONCLUSÕES.....	62
REFERÊNCIAS BIBLIOGRÁFICAS.....	64
MATERIAL SUPLEMENTAR.....	91

ARTIGOS135

**The genome of a clinical *K. variicola* strain reveals virulence associated traits
and a p19-like plasmid..... 135**

LISTA DE FIGURAS

Figura 1 – Principais supergrupos de bactérias do gênero *Wolbachia* e seus respectivos hospedeiros. *Wolbachia* associadas exclusivamente a artrópodes pertencem exclusivamente aos supergrupos A, B, E, H, I e K, sendo os supergrupos A e B os mais estudados e diversos. *Wolbachia* associadas exclusivamente a nematoides pertencem aos supergrupos C, D, J, L e O. O supergrupo F é encontrado tanto em artrópodes quanto em nematoides. 6

Figura 2 – Árvore filogenética baseada em 31 proteínas conservadas identificadas por Puebla e colaboradores (2015). Concluiu-se, baseado na topologia da árvore e em valores de ANI, que os seis supergrupos analisados (A-F) corresponderiam a 8 espécies do gênero *Wolbachia*. 8

Figura 3 – Etapas seguidas para a obtenção dos genomas utilizados, reanotação, filtragem dos genomas dos dois gêneros (completude para *Wolbachia* e clusterização para *Klebsiella*) e análises subsequentes como taxonomia genômica, filogenômica e análise de enriquecimento funcional. 14

Figura 4 – Árvore filogenética construída a partir do super alinhamento de sequências de MLST de todos os genomas de isolados do gênero *Wolbachia* utilizados (768). A árvore foi construída utilizando o método de *Neighbor joining* com o modelo Kimura 2 parâmetros e 100 replicatas de bootstrap. Espécies do gênero *Klebsiella* estão assinaladas em tons de cinza. Identificadores de taxas que foram corretos ou incorretamente assinalados como *K. pneumoniae* estão em preto. 15

Figura 5 – Etapas seguidas para a identificação do pan genoma do gênero *Wolbachia* e obtenção da matriz de presença/ausência do pangenoma e o cálculo do enriquecimento funcional, focado tanto no pan-genoma de forma global quanto em genomas de forma individual. 21

Figura 6 - Árvore filogenômica construída a partir do superalinhamento dos genes *ftsZ*, *groEL*, *dnaA*, *fbpA*, *gatB*, *coxA* de genomas de bactérias do gênero *Wolbachia*, os quais fazem parte do esquema de MLST descrito para este gênero. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap. 24

Figura 7 - Árvore filogenômica construída a partir do superalinhamento dos genes codificadores de subunidades de proteínas ribossômicas (*rps*) de genomas de bactérias do gênero *Wolbachia*. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap. A árvore nos mostra cinco grupos bem definidos (A-D, F) representando os supergrupos já descritos (A-F) e o genoma wVulC, que foi previamente classificado como pertencente ao supergrupo B mas é claramente distante do grupo principal deste supergrupo. 26

Figura 8 - Árvore filogenômica construída a partir do superalinhamento dos genes comuns a todos os genomas de bactérias do gênero *Wolbachia* analisados, ou genoma *core*. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios

invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap. A árvore nos mostra cinco grupos bem definidos (A-D, F) representando os supergrupos já descritos (A-F) e o genoma wVulC, que foi previamente classificado como pertencente ao supergrupo B mas é claramente distante do grupo principal deste supergrupo. 26

Figura 9 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes constitutivos que são parte do esquema de MLST para o gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R. 27

Figura 10 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes codificadores de proteínas ribossomais de bactérias do gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R. 28

Figura 11 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes do genoma *core* do gênero *Wolbachia*, onde tons de azul significam valores <96%, branco 96% e tons de vermelho >96%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R. 28

Figura 12 - Representação gráfica da matriz da identidade média de nucleotídeos entre todos os pares de genomas do gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R. 34

Figura 13 - Representação gráfica da matriz de hibridização DNA DNA digital (GGD) entre todos os pares de genomas, o dendrograma foi construído utilizando o método euclidiano encontrado na linguagem R, onde tons de azul significam valores <70%, branco 70% e tons de vermelho >70%. 34

Figura 14 - Representação gráfica da matriz de alinhamento de fração entre todos os pares de genomas, o dendrograma foi construído utilizando o método euclidiano encontrado na linguagem R, onde tons de azul significam valores <0.6, branco 0.6 e tons de vermelho >0.6. 35

Figura 15 - Representação gráfica do pan-genoma do gênero *Wolbachia*. Genomas foram agrupados através do método de máxima verossimilhança, genes foram agrupados utilizando o método euclidiano padrão encontrado na linguagem R. Seis grupos gênicos (I-VI) foram definidos manualmente a partir do perfil de presença/ausência de genes onde vermelho é a presença de determinado gene e azul a ausência 41

Figura 16 - Composição funcional dos grupos gênicos (I-VI) do pan-genoma, baseada nos subsistemas de nível 1 do banco de sequências SEED. Funções enriquecidas estão marcadas em tons de vermelho, funções empobrecidas em tons de azul. Funções significantes estão marcadas em tons de cinza. 43

Figura 17 - Composição funcional genomas de isolados do gênero *Wolbachia* baseada nos subsistemas de nível 1 do banco de sequências SEED. É possível observar que genomas de endossimbiontes facultativos possuem as funções “Regulação e sinalização celular” e “Fagos, profagos e plasmídeos” altamente enriquecidos, enquanto este é empobrecido em genomas de endossimbiontes obrigatórios. 44

Figura 18 – Representação gráfica circular da diversidade dos genomas do gênero *Wolbachia* onde cada genoma foi alinhado com o isolado tipo wPip¹, representado pelo primeiro círculo vermelho. Supergrupos estão representados por diferentes cores, os cinco círculos externos são genomas de endossimbiontes obrigatórios e é possível se observar grandes regiões ausentes nestes isolados. Os três círculos finais representam proteínas hipotéticas, sequências repetitivas Ankyrin e elementos móveis, respectivamente. O último círculo representado por setas representam profagos identificados pelo software PHAST 45

Figura 19 - Imagem gerada pelo software FACIL, onde códons possíveis e os aminoácidos que estes estariam codificando. Códons de terminação e de início estão marcados em branco, a diferença entre estes três genomas reside no genoma wPipMol, onde o códon de terminação OPAL (TGA) foi identificado pelo software como codificante para o aminoácido triptofano. 47

Figura 20 - Árvore de máxima verossimilhança e o alinhamento da proteína Citocromo oxidase C originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap. 49

Figura 21 - Árvore de máxima verossimilhança e o alinhamento da proteína Citocromo oxidase B originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap. 49

Figura 22 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade H originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap. 50

Figura 23 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade D originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap. 50

Figura 24 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade N originado de uma mitocôndria de uma espécie de mosquito, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap. 51

Figura 25 - Imagem gerada pelo Blast ring image generator (BRIG) a partir do genoma wPipMol.

Genes provenientes de fagos estão marcados em verde, predições de fagos feitos pelos softwares Phast (Vermelho), Virsorter (Cinza) e Physpy (azul) estão marcadas, indicando que a aquisição destes insertos pode ter sido devido a um evento de transdução. 52

Figura 26 – Árvore filogenômica construída a partir do superalinhamento dos genes do genoma core do gênero *Klebsiella* (cgMLST). A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR, o suporte nos nós foi obtido através de 1000 replicatas de bootstrap. A árvore nos mostra três grupos bem definidos representando as espécies *K. pneumoniae*, *quasipneumoniae* e *variicola*. Isolados de *K. oxytoca* foram utilizados como grupo externo. Identificadores marcados em vermelho são referentes a isolados previamente classificados como *K. pneumoniae* mas que agruparam com linhagens tipo de espécies diferentes. 55

Figura 27 – Árvore filogenômica construída a partir do superalinhamento dos codificadores de rps (rMLST). A árvore foi construída utilizando o modelo GTR, o suporte nos nós foi obtido através de 1000 replicatas de bootstrap. Identificadores marcados em vermelho são referentes a isolados previamente classificados como *K. pneumoniae* mas que agruparam com linhagens tipo de espécies diferentes. 56

Figura 28 – Resolução taxonômica da espécie *K. pneumoniae*, este gráfico comparativo foi construído a partir de matrizes de similaridade de abordagens como ANI, cgMLST, GGDC e AF de todos os isolados de *Klebsiella* analisados contra o genoma da linhagem tipo de *K. pneumoniae*, HS11286[†]. 58

Figura 29 – Resolução taxonômica da espécie *K. pneumoniae*, este gráfico comparativo foi construído a partir de matrizes de similaridade de abordagens como ANI, cgMLST, GGDC e AF de todos os isolados de *Klebsiella* analisados contra o genoma da linhagem tipo de *K. quasipneumoniae*, 01A030[†]. 60

Figura 30 – Resolução taxonômica da espécie *K. pneumoniae*, este gráfico comparativo foi construído a partir de matrizes de similaridade de abordagens como ANI, cgMLST, GGDC e AF de todos os isolados de *Klebsiella* analisados contra o genoma da linhagem tipo de *K. variicola*, DSM18961[†]. 61

Figura 31 – Representação circular dos plasmídeos conjugativos *PI9*, encontrado em isolados de *K. pneumoniae* e *PBZ19*, primeiro plasmídeo identificado em *K. variicola*. O plasmídeo *pi9* possui dois elementos Tn4401, profagos, que comportam a carbapenemase *bla_{KPC}*, o qual não é encontrado em *PBZ19*, sendo a única diferença entre estes. 63

LISTA DE TABELAS

Tabela 1 – isolados de Wolbachia, características genômicas, completude e hospedeiros	12
Tabela 2 – isolados do gênero Klebsiella de diferentes espécies, características genômicas e número de acesso	16
Tabela 3 - Matriz de similaridade construída a partir do superalinhamento de sequências intrínsecas (MLST) de isolados do gênero Wolbachia	29
Tabela 4 - Matriz de similaridade construída a partir do superalinhamento de sequências de rps de isolados do gênero Wolbachia	30
Tabela 5 - Matriz de similaridade construída a partir do superalinhamento dos genes do genoma core de isolados do gênero Wolbachia	31
Tabela 6 - Matriz construída a partir dos valores de identidade média de nucleotídeos de isolados do gênero Wolbachia	36
Tabela 7 - Matriz construída a partir dos valores de hibridização DNA-DNA digital de isolados do gênero Wolbachia	37
Tabela 8 - Matriz construída a partir dos valores de AF de isolados do gênero Wolbachia	38
Tabela S1 – Pan-genoma de isolados do gênero Wolbachia contendo 1272 grupos gênicos. A tabela contém o identificador (Locus tag), e os níveis de anotação de subsistemas Seed 1-3.	91
Tabela S2 – Métricas das abordagens de taxonomia genômica utilizadas em isolados do gênero Klebsiella.	126

LISTA DE SIGLAS E ABREVIATURAS

MLST	Tipagem da sequência multilocus
rMLST	Tipagem da sequência multilocus de proteínas ribossomais
cgMLST	Tipagem da sequência multilocus do core-genoma
ANI	Identidade média de nucleotídeos
rps	Subunidade de proteína ribossomal
DDH	Hibridação DNA-DNA
GBDP	<i>Genome blast distance phylogeny</i>
GGD	<i>Genome to genome distance</i>
AF	Alinhamento de fração
Subsp.	Subespécie
CBAS	Coleção de bactérias do ambiente e da saúde
NCBI	<i>National Center for Biotechnology Information</i>

1 - Introdução

1.1 - Filogenômica

A evolução de microrganismos, em especial bactérias, é em grande parte dependente da natureza fluída de seus genomas, sendo sujeitos portanto a eventos como a aquisição lateral de genes, duplicação e perda gênica. Tais eventos são comuns e limitam a inferência de suas relações evolutivas, já que diferentes genes podem contar diferentes histórias filogenéticas (Comas et al. 2007). O advento de abordagens genômicas e a redução do custo do sequenciamento de genomas completos, permitiu não só uma mudança na escala de análises de sequências, indo de genes únicos para análises globais, mas também para inferências filogenéticas. A filogenômica, uma atualização de abordagens filogenéticas para o contexto genômico (Eisen & Eisen 1998), é uma abordagem integrativa onde informações genômicas em nível global são utilizadas para interpretar dados e testar hipóteses, englobando diferentes campos da biologia moderna, tais como a biologia molecular, sistemática, populacional, evolutiva e computacional.

Esta abordagem foi inicialmente proposta por Jonathan Eisen, em 1998, para a predição e anotação funcional de genes em novos genomas, os quais foram montados sem um genoma referência. A inferência da função de um novo gene é feita através de um processo multi etapas onde ortólogos, genes com uma mesma história evolutiva e que, por isto, possuiriam uma função similar, são identificados e discriminados de parálogos, que são genes originados a partir de eventos de duplicação e por isso eliminados da análise, sendo então utilizados na anotação destes genes (Eisen & Eisen 1998, Brown & Sjölander 2006).

Abordagens filogenômicas permitem também o estabelecimento de relações genéticas entre bactérias de forma robusta, sendo possivelmente a forma mais apropriada de se traçar a história evolutiva de organismos. As vantagens de abordagens que utilizam múltiplos genes são diversas, é possível, por exemplo, evitar sinais evolutivos provenientes de genes únicos favorecendo um sinal filogenético “verdadeiro” inferido a partir de um grande número de genes. A utilização de métodos filogenômicos evitam também erros estocásticos inerentes a utilização de gene únicos, já que tais erros tendem a desaparecer naturalmente quando mais genes são adicionados (Delsuc et al. 2005, Jeffroy et al. 2006).

Diversas metodologias que utilizam informações genômicas de forma direta foram propostas, mas a utilização do método de concatenação de sequências tem se tornado relativamente comum. Este método amplifica o sinal filogenético ao juntar sequências de múltiplos genes, criando uma supermatriz de caracteres (ou um super alinhamento) que geralmente geram filogenias com alto suporte de *bootstrap* nos nós (Comas et al. 2007). Este método é preferencialmente utilizado quando o sinal filogenético proveniente de um gene único é insuficiente (Herniou et al. 2001), quando existe heterogeneidade nas taxas evolutivas (Gontcharov et al. 2004, Baptiste et al. 2005) ou quando eventos como a transferência horizontal de genes está influenciando a topologia da árvore filogenética (Brochier et al. 2002).

Um exemplo da aplicação de métodos de concatenação de genes para resolver relações genéticas foi feita por Bennet et al., (2013) (Bennett et al. 2013) onde, após analisar a similaridade do superalinhamento de sequências de genes codificadores de proteínas ribossomais (*rps*) das espécies *Neisseria oralis* e *Neisseria mucosa*, chegou a conclusão de que se tratavam de uma única espécie. A abordagem desenvolvida por Bennet é chamada de tipagem da sequência multilocus de proteínas ribossomais (rMLST). (Comandatore et al. 2013)

Recentemente, abordagens filogenômicas foram aplicadas em genomas completos de diversas espécies de arqueias, estas foram recuperadas a partir do metagenoma de uma fonte hidrotermal conhecida como “Castelo de Loki”. Estes genomas foram classificados como pertencentes a um novo filo, Lokiarchaeota e, interessante, diversas características genômicas destes eram similares a de Eucariotos, como a maquinaria de replicação e pequenas GTPases (Spang et al. 2015). De fato, a análise de 36 proteínas conservadas revelou, de forma inequívoca, que o domínio dos Eucariotos e o filo lokiarchaeota são monofiléticos.

1.2 - Taxonomia genômica

A taxonomia é uma abordagem que permite classificar organismos e constitui uma das bases fundamentais da ciência da vida. Atualmente, a classificação de espécies de bactéria é baseada em uma abordagem polifásica (Vandamme et al. 1996), a qual integra informações fenotípicas, quimiotaxonômicas e genotípicas para classificar espécies de bactérias. Dentre as abordagens utilizadas, a hibridação DNA-DNA é considerada o padrão ouro na descrição de espécies (12). Para aplicá-la é necessário uma coleção de isolados de espécies já definidas para o gênero

estudado, além de um aparato (reagentes/equipamentos) para as reações de hibridação, o que limita a realização de testes para definição de novas espécies por pesquisadores em todo o mundo (12).

A análise de similaridade do gene ribossomal 16s, uma outra abordagem utilizada pela taxonomia polifásica, permitiu o desenvolvimento de trabalhos revolucionários como o de Woese et al., (1977), onde o domínio das Arqueias foi estruturado, o separando do domínio das bactérias (Woese & Fox 1977). Limites de similaridade para estas abordagens foram definidos após repetidos ensaios experimentais, onde uma espécie é definida como um grupo de isolados que possuem >70% de DDH, <5% de diferença no conteúdo GC entre os genomas e uma identidade do gene ribossomal 16s > 97% (Brenner et al. 2016, Schumann et al. 2013).

Mora et al., (2002) propôs recentemente um novo conceito de espécies para procariotos, onde: “Uma espécie é uma categoria que circunscreve um grupo monofilético de populações de indivíduos genomicamente e fenotipicamente coerentes” (Rossell-Móra & Amann 2015). A aplicação deste conceito passou a ser factível através de novas metodologias para obtenção de genomas completos de bactérias, tanto sob o ponto de vista do custo quanto da rapidez e portabilidade das informações. Além disso, a aplicação desta abordagem é possível em organismos não-cultiváveis.

De fato, o genoma de um organismo é o seu sinal filogenético definitivo e, devido a isto, métodos para identificar e classificar espécies bacterianas utilizando o genoma completo foram criados (Thompson et al. 2009, 2013, 2015, Larsen et al. 2014, Varghese et al. 2015). Esta metodologia, conhecida como taxonomia genômica, utiliza métricas baseadas em assinaturas genômicas intraespecíficas para classificar, reclassificar e criar nomenclaturas para taxas, sendo especialmente útil em isolados bacterianos que não são passíveis de cultivo (Thompson et al. 2009, 2013).

Para isto, novos métodos foram desenvolvidos ao longo dos anos, tais como: Análise da sequência Multi Locus (MLSA); rMLST (Jolley et al. 2012); Tipagem da sequência multi locus de genes do Core Genoma (cgMLST) (Bennett et al. 2013, Andrade et al. 2014); A identidade média de nucleotídeos/aminoácidos (ANI, AAI) (Varghese et al. 2015); Alinhamento de Fração (AF) (Varghese et al. 2015), Hibridização DNA-DNA digital (GGD) (Auch et al. 2006), entre outros, integrando informações de diferentes abordagens *in silico* para que sejam inferidas relações filogenômicas e taxonômicas.

Em 1999 foi realizada uma das primeiras tentativas de utilizar o genoma bacteriano para classificação taxonômica (Snel et al. 1999). Neste trabalho, o autor utilizou perfis de presença/ausência de genes ortólogos para reconstruir a filogenia de todos os genomas completamente sequenciados na época (13 genomas). Abordagens taxonomia genômicas vem sendo aplicadas a diferentes gêneros de microrganismos, um exemplo foi a definição de espécies para o gênero *Vibrio*, onde a autora concluiu que métricas como >95% de similaridade de MLSA, > 96% de identidade média de aminoácidos e <10 de dissimilaridade genômica definiriam espécies para esse gênero (Thompson et al. 2009).

Esta abordagem também foi aplicada ao gênero *Prochlorococcus*, gênero que era representado por uma única espécie, *Prochlorococcus marinus*. Esta espécie era dividida em diferentes ecogrupos que, por taxonomia genômica, corresponderiam a diferentes espécies deste gênero (Thompson et al. 2013). Mais recentemente, esta abordagem foi aplicada em aproximadamente 13.000 genomas de diversas espécies bacterianas, revelando que muitas destas espécies deveriam ser revistas (Varghese et al. 2015).

1.3 - *Wolbachia pipientis*

Wolbachia pipientis é uma alfa-proteobactéria intracelular estrita pertencente a família Rickettsiaceae, formalmente descrita em 1936 por Marshall Hertig (Hertig 2009). Esta espécie era, até recentemente, a única representante do gênero *Wolbachia*, gênero cujos isolados estão associados a células de hospedeiros específicos como espécies de insetos, crustáceos, aracnídeos e nematoides da família Onchocercidae (Bandi et al. 1998). Isolados de *W. pipientis* são identificadas em tecidos específicos do hospedeiro, como gônadas e células germinativas em artrópodes (Ramírez-Puebla et al. 2015) e em invaginações da hipoderme ou cordas laterais em nematoides de ambos os sexos (Foster et al. 2005).

Wolbachia é também o gênero de bactérias intracelulares mais predominantes no mundo, a prevalência destas bactérias em diferentes espécies de artrópodes foi estimada em 40% (Bandi et al. 1998, Wu et al. 2004, Zug & Hammerstein 2012). Isolados de *Wolbachia* são transmitidos em sua grande maioria verticalmente pela linhagem materna, entretanto, existem evidências de transmissão horizontal entre hospedeiros de diferentes espécies de artrópodes e de nematoides, já que a filogenia destes endossimbiontes não seguem estritamente a de seus hospedeiros

(Lefoulon et al. 2016).

Sua condição como bactéria intracelular obrigatória é compartilhada por todos os membros da família Rickettsiaceae, a qual comporta parasitas intracelulares e patógenos de mamíferos, como as bactérias *Rickettsia typhi* e *Rickettsia prowazekii*, patógenos causadores de tifo epidêmica em humanos (Bechah et al. 2008) e *Anaplasma marginale*, o patógeno transmitido por pulgas mais prevalentes em rebanhos de gado mundialmente (Kocan et al. 2003, Herndon et al. 2010) representando um impacto socioeconômico considerável tanto na saúde pública quanto na pecuária.

1.3.1 - Características genômicas

Membros do gênero *Wolbachia* possuem genomas reduzidos, cujos tamanhos variam entre 0.9 e .6 Mbp. Este tamanho reduzido é resultante de processos de evolução reductiva, isto é, a evolução destas bactérias se dá mais por perda gênica do que por aquisição lateral de genes (Georgiades et al. 2011), o que é consistente com sua condição de vida intracelular. O ambiente intracelular é estável e rico em nutrientes, provendo proteínas e cofatores para bactérias intracelulares, conseqüentemente alguns genes, como os que participam de algumas vias biossintéticas, não são mais estritamente necessários e, devido a isto, a perda destes não é deletéria (Moran 2003), propiciando eventos de pseudogenização, onde o acúmulo de mutações pontuais resultam em perda de função e perda gênica (Wu et al. 2004).

Genomas de isolados de *Wolbachia* associadas a artrópodes possuem uma alta densidade de elementos genéticos móveis, como transpósons e fagos (Wu et al. 2004, Cerveau et al. 2015). Tais elementos aliados a seu excepcionalmente alto número de sequências repetitivas, explicam os frequentes eventos de rearranjos genômicos observados nestes isolados, em outras palavras, isolados de *Wolbachia* possuem pouca sintonia gênica (Klasson et al. 2008, Comandatore et al. 2015).

Isolados do gênero *Wolbachia* associadas a nematoides também são suscetíveis a recombinação, evidências de tais eventos vieram através de análises filogenéticas utilizando o gene codificante da proteína de superfície *wsp* (Werren & Bartos 2001, Jiggins 2002, Bordenstein & Wernegreen 2004, Baldo et al. 2006a). Isolados associadas a nematoides possuem poucos elementos genéticos móveis e rearranjos, como esperado de uma relação entre hospedeiro e simbiote ancestral

(Wu et al. 2004, Bordenstein et al. 2009, Cerveau et al. 2015).

1.3.2 – Filogenia e supergrupos filogenéticos

Wolbachia é um gênero geneticamente diverso (Ishmael et al. 2009), sendo esta diversidade representada por supergrupos filogenéticos (A-Q) (Figura 1) que é baseada na similaridade de sequência do gene *wsp* e do gene codificador do RNA ribossomal 16S. Destes, 10 são encontrados associados exclusivamente a artrópodes (A, B, E, H, I, K, M, N, P, Q), 5 associados de forma mutualística a nematoides (C, D, J, L, O) e 1 (F) encontrado associado em artrópodes e nematoides, sendo os supergrupos A e B conhecidos como parasitas sexuais pois induzem fenótipos reprodutivos em seus hospedeiros. Isolados de Wolbachia encontradas em aranhas australianas foram classificadas como pertencentes ao supergrupo G, porém concluiu-se que o gene *wsp* sofreu eventos de recombinação e, portanto, o supergrupo G não é consistente.

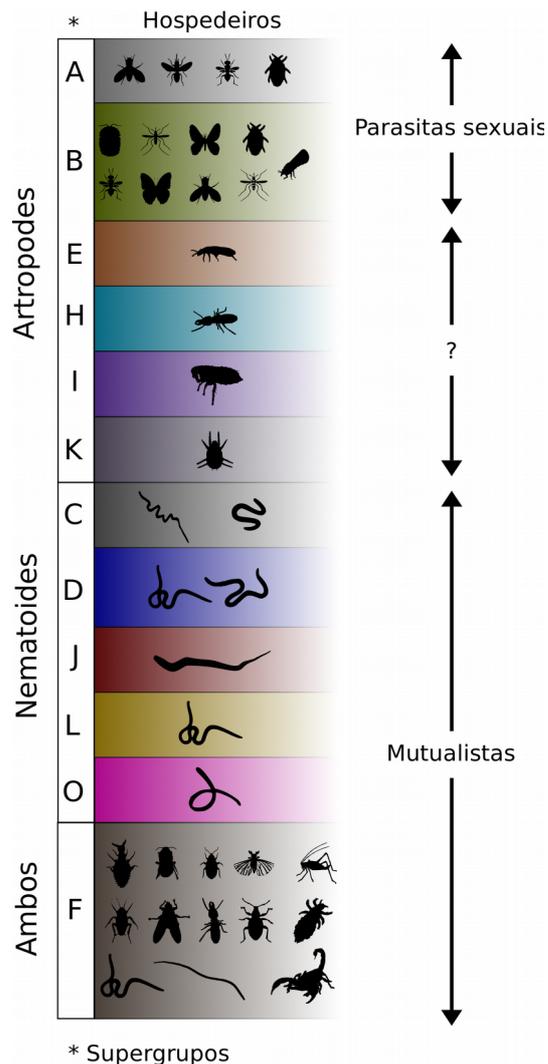


Figura 1 – Principais supergrupos de bactérias do gênero Wolbachia e seus respectivos hospedeiros.

Wolbachia associadas exclusivamente a artrópodes pertencem exclusivamente aos supergrupos A, B, E, H, I e K, sendo os supergrupos A e B os mais estudados e diversos. Wolbachia associadas exclusivamente a nematóides pertencem aos supergrupos C, D, J, L e O. O supergrupo F é encontrado tanto em artrópodes quanto em nematóides.

1.3.3 - Fenótipos induzidos

isolados de Wolbachia possuem a habilidade de estabelecer uma rede de interações fisiológicas com o seu hospedeiro, esta interação permite a manipulação de processos celulares e a indução de fenótipos reprodutivos e não reprodutivos, como: Incompatibilidade citoplasmática (IC) (Turelli & Hoffmann 1999, Poinot et al. 2003); modificação da partenogênese (Huigens et al. 2000); feminilização de embriões (Cordaux et al. 2004); morte de embriões machos (Hurst & Jiggins 2000); mutualismo nutricional (Nikoh et al. 2014); fornecimento de ATP para seus hospedeiros (57). Assegurando assim sua transmissão e dispersão em uma população não infectada (Moran 2006).

1.3.4 - Status taxonômico do gênero Wolbachia

Autores como Lo et al., (2007) propuseram que todas as bactérias deste gênero deveriam ser classificadas como *W. pipientis* (Lo et al. 2007), entretanto Pfarr foi o primeiro a reconhecer que diversas características deveriam ser consideradas antes de agrupá-las em uma única espécie, especialmente ao comparar Wolbachia associadas a nematóides e artrópodes (Pfarr et al. 2011). Recentemente, Puebla et al. (Wu & Eisen 2008), realizou reconstruções filogenômicas a partir de 31 proteínas conservadas (Figura 2), hibridização DNA-DNA *in silico* e a identidade média de nucleotídeos em isolados de 6 supergrupos de Wolbachia (A-F), revelando que estes supergrupos eram heterogêneos o suficiente para serem considerados diferentes espécies, sendo que alguns destes, como os supergrupos C e D, agrupariam mais de uma espécie cada (Ramírez-Puebla et al. 2015). Consequentemente, oito espécies candidatas foram propostas: '*Candidatus W. bourtizisii*' (Supergrupo A), '*Ca. W. pipientis*' (Supergrupo B), '*Ca. W. onchocerciola*' (Supergrupo C), '*Ca. W. blaxteri*' (Supergrupo C), '*Ca. W. brugii*' (Supergrupo D), '*Ca. W. taylori*' (Supergrupo D), '*Ca. W. colembolicola*' (Supergrupo E), e '*Ca. W. multihospitum*' (Supergrupo F) (Ramírez-Puebla et al. 2015).

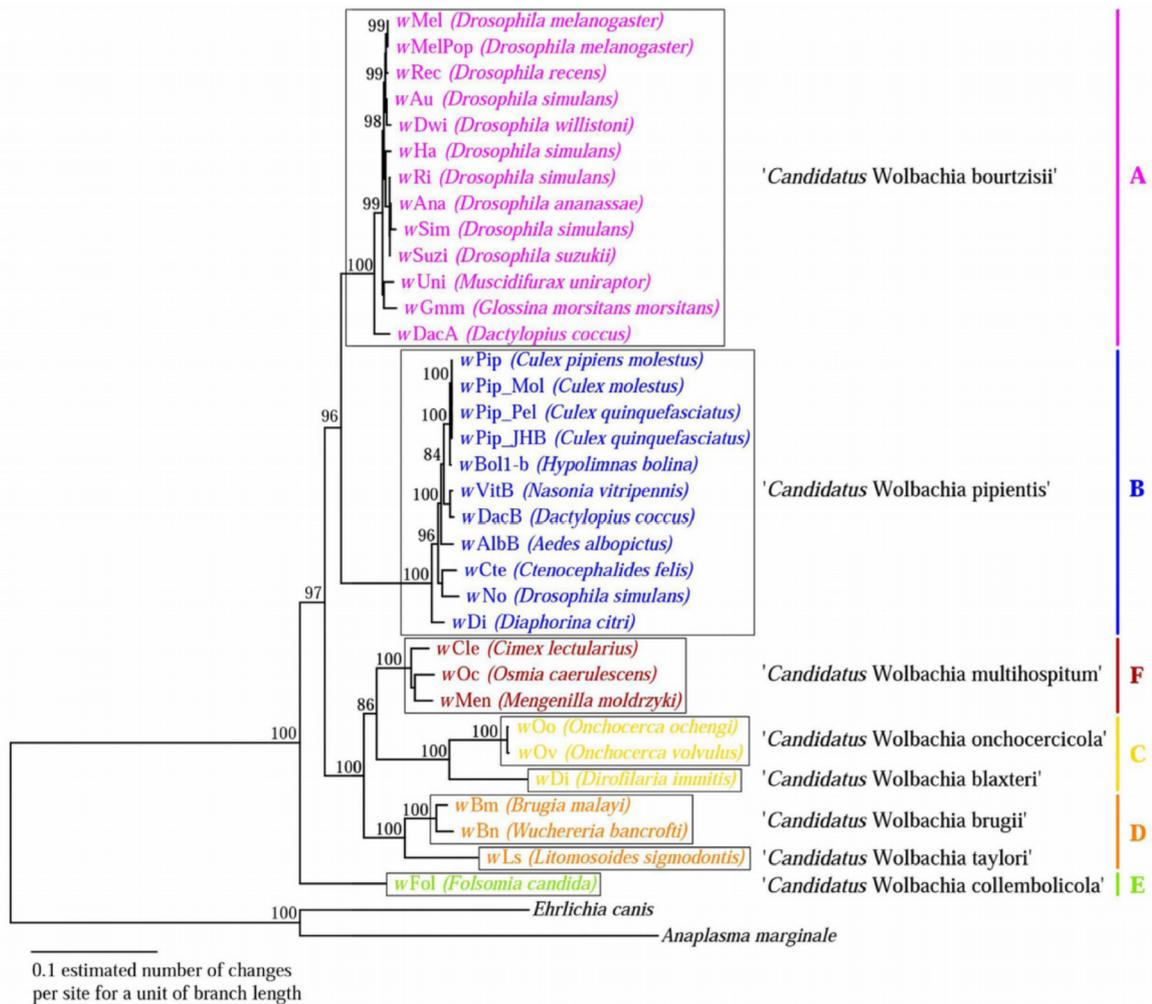


Figura 2 – Árvore filogenética baseada em 31 proteínas conservadas identificadas por Puebla e colaboradores (2015). Concluiu-se, baseado na topologia da árvore e em valores de ANI, que os seis supergrupos analisados (A-F) corresponderiam a 8 espécies do gênero Wolbachia.

1.4 - Klebsiella

O gênero Klebsiella compreende bactérias não móveis, anaeróbicas facultativas, caracterizadas pela presença de uma cápsula polissacarídica que confere resistência a diversos mecanismos de defesa do hospedeiro. As bactérias deste gênero são ubíquas e podem ser encontradas no solo e água, e são capazes de colonizar diferentes tecidos em humanos ou animais, ou de coexistirem em uma relação mutualística com plantas como endófitos diazotróficos (fixadores de nitrogênio) (Fouts et al. 2008, Brisse 2014).

1.4.1 – Características genômicas

Isolados de diferentes espécies do gênero *Klebsiella* possuem um genoma único circular, o tamanho de seus genomas varia entre 5 e 6 Mbp. A genoma da linhagem tipo de *Klebsiella pneumoniae* HS11286^T, por exemplo, possui 5,316 CDS, 87 tRNAs e, como parte de seu mobiloma, 7 regiões de prófagos e elementos integrativos conjugativos (ICE), que são elementos que permitem um fluxo dinâmico de genes entre bactérias (Liu et al. 2012). Uma diferença fenotípica importante que é diretamente relacionada a diferenças genômicas entre isolados de *K. pneumoniae* e de outras espécies deste gênero, como *Klebsiella variicola* ou *Klebsiella oxytoca* é a capacidade de fixar nitrogênio. Isolados de *K. pneumoniae* não são capazes de fixar nitrogênio pois o *cluster* gênico responsável por este fenótipo, o *cluster nif*, está ausente nos genomas de todas as bactérias de *K. pneumoniae* identificados até o momento, estando presentes em isolados de outras espécies de *Klebsiella*, como *K. variicola* e *K. quasipneumoniae* (Rosenblueth & Martínez 2004, Fouts et al. 2008, Pinto-Tomás et al. 2009, Hazen et al. 2014).

1.4.2 - Filogenia e status taxonômico

A estrutura filogenética de *Klebsiella* demonstra a complexidade taxonômica deste gênero, que se reflete na dificuldade em identificar corretamente as espécies deste gênero (Alves et al. 2006). As espécies do gênero *Klebsiella* foram originalmente estabelecidas com base nas suas características fenotípicas e em função das doenças as quais estavam associadas. Contudo, a aplicação da taxonomia polifásica e genômica tem levado a revisões na definição de espécies de *Klebsiella*. Por exemplo, as espécies *K. planticola*, *K. terrigena* e *K. ornithinolytica*, originalmente incluídas no gênero *Klebsiella* com base na taxonomia numérica, foram realocadas no gênero *Raoultella* através da análise comparativa dos genes 16S rRNA e rpoB (Drancourt et al. 2001). Por sua vez, os grupos filogenéticos de *K. pneumoniae* (KpI, KpII-A e II-B e KpIII) definidos com base na similaridade de sequências de *gyrA* e *parC* (Brisse & Verhoef 2001), foram identificados como distintas espécies de *Klebsiella*. Análises filogenéticas baseadas no esquema de MLST para este gênero, hibridação DNA-DNA e ANI demonstraram que os grupos KpI, KpII-A e II-B e KpIII correspondiam às espécies *K. pneumoniae*, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* (KpII-A), *K. quasipneumoniae* subsp. *similipneumoniae* (KpII-B) e *Klebsiella variicola*, respectivamente (Rosenblueth &

Martínez 2004, Brisse et al. 2014).

Brisse e colaboradores (2004) demonstraram a associação ancestral de genes cromossômicos constitutivos que codificam β -lactamases de classe A com os grupos filogenéticos KpI, KpII e KpIII (Brisse et al. 2004). Com a realocação de isolados destes grupos em diferentes espécies definiu-se que *K. pneumoniae* (KpI), *K. quasipneumoniae* (KpII) e *K. variicola* (KpIII) carregam, respectivamente, os genes cromossômicos *bla_{SHV}*, *bla_{OKP}* and *bla_{LEN}*, enquanto que *K. oxytoca* carrega o gene *bla_{OXY}* (Rosenblueth & Martínez 2004, Brisse et al. 2014, Andrade et al. 2014).

Atualmente o gênero *Klebsiella* inclui seis espécies: *Klebsiella pneumoniae*, com suas três subespécies (*K. pneumoniae* subsp. *pneumoniae*, *K. pneumoniae* subsp. *ozaenae*, *K. pneumoniae* subsp. *rhinoscleromatis*) (63), *Klebsiella oxytoca* (64), *K. variicola*, *K. granulomatis*, *K. singaporensis*, *K. michiganensis* e *K. quasipneumoniae* com suas subespécies *K. quasipneumoniae* subsp. *quasipneumoniae* e *K. quasipneumoniae* subsp. *similipneumoniae* (Brisse et al. 2004).

Contudo, a identificação de *Klebsiella* através de testes bioquímicos continua sendo a mais usada na rotina clínica, o que leva a resultados equivocados e a uma subestimação da diversidade de espécies deste gênero (68,69). De fato, casos de infecções severas causadas pela espécie *K. variicola* inicialmente identificada como *K. pneumoniae* foram relatados (Berry et al. 2015).

Até o momento, a maioria dos trabalhos que utilizaram abordagens filogenéticas para estudar a estrutura taxonômica do gênero *Klebsiella* focaram apenas em número limitado de espécies (Rosenblueth & Martínez 2004, Brisse et al. 2014, Andrade et al. 2014). Portanto, a aplicação das diferentes métricas consideradas na taxonomia genômica para analisar as diferentes espécies de *Klebsiella* com genomas já disponibilizados permitirá desvendar de forma mais robusta a complexidade taxonômica observada neste gênero.

Neste trabalho, aplicamos análises filogenômicas e de taxonomia genômica em isolados de dois gêneros distintos de bactérias: *Wolbachia* e *Klebsiella* para investigar relações funcionais, filogenéticas, filogenômicas e taxonômicas entre diferentes isolados destes gêneros.

2 - Objetivos

2.1 - Objetivo principal

Aplicar abordagens *in silico* para investigar relações funcionais, filogenéticas, filogenômicas e taxonômicas entre organismos classificados como espécie(s) dos gêneros *Wolbachia* e *Klebsiella*.

2.2 - Objetivos específicos

- Investigar a diversidade genômica e taxonômica de isolados dos gêneros *Wolbachia* e *Klebsiella* utilizando abordagens *in silico*.
- Gerar e caracterizar os genomas dos isolados de *K. variicola* e *K. quasipneumoniae*, BZ19 e K142
- Caracterizar funcionalmente os genomas *core* e *pan* do gênero *Wolbachia*

3 - Métodos

Genomas de isolados do gênero *Wolbachia* (n = 27) e de espécies de *K. pneumoniae* (n = 740), *K. quasipneumoniae* (n = 4), *K. variicola* (n = 4), genomas do gênero *Klebsiella* que não foram assinaladas a nenhuma espécie (n = 3) e de *Klebsiella oxytoca* (n = 4), foram recuperados do banco de dados do Genbank (Acland et al. 2014), os números de acesso estão indicados nas Tabelas 1 e 2. Para evitar vieses provenientes de diferenças entre métodos de anotação de genomas, nós os re-anotamos novamente todos os genomas utilizando o software Prokka (Seemann 2014) utilizando parâmetros padrões.

3.1 - *Wolbachia*

Devido a sua condição intracelular, parte dos genomas de *Wolbachia* foram sequenciados a partir de extrações completas de DNA e recuperados através de filtragem de *reads* (Campana et al. 2015) e, devido a isto, estes genomas podem não estar completos. Para a avaliação da recuperação dos genomas completos (completude) utilizamos o software CheckM, este software estima a completude de um genoma a partir da identificação de genes marcadores segundo o nível taxonômico do genoma analisado (Parks et al. 2015), genomas que apresentaram uma completude de <90% foram removidos de algumas análises (i.e. o limiar sugerido pelo desenvolvedor do software) (Figura 3). A representação gráfica da diversidade genômica de isolados do gênero *Wolbachia* foi feita utilizando o software *Blast Ring Image Generator* (BRIG) (Alikhan et al. 2011).

Tabela 1 – isolados de *Wolbachia*, características genômicas, completude e hospedeiros.

isolado	Genoma (Mbp)	Conteúdo GC	Hospedeiro	Acesso	Profagos	Completude	Referência
wAlbB	1.16	33.6	<i>Aedes albopictus</i>	CAGB00000000	2	99.57	(Mavingui et al. 2012)
wAna	1.44	35.7	<i>Drosophila ananassae</i>	AAGB00000000	2	86.47	(Salzberg et al. 2005)
wAu	1.26	35.20	<i>Drosophila simulans</i>	LK055284	2	99.79	(Sutton et al. 2014)
wBn	1.05	34	<i>Wuchereria bancrofti</i>	ADHD00000000	-	63.28	(Desjardins et

WBol1-b	1.37	33.9	<i>Hypolimnas bolina</i>	CAOH00000000	2	99.79	al. 2013) (Duplouy et al. 2013)
wCam	0.96	32.7	<i>Onchocerca volvulus</i>	HG810405	0	94.34	(Desjardins et al. 2013)
wCle	1.25	36.3	<i>Cimex lectularius</i>	AP013028	2	100	al. 2013) (Nikoh et al. 2014)
wDacB	1.12	34.3	<i>Dactylopius coccus</i>	JMCJ01000000	-	86.84	(Campana et al. 2015)
wDi	1.24	34	<i>Diaphorina citri</i>	AMZJ01000000	1	100	(Saha et al. 2012)
wGmm	1.01	35.20	<i>Glossina morsitans morsitans</i>	AWUH00000000	2	91.64	(Brelsfoard et al. 2014)
wHa	1.29	35.20	<i>Drosophila simulans</i>	NC_021089	2	99.79	(Ellegaard et al. 2013)
wJHB	1.57	34.20	<i>Culex quinquefasciatus</i>	ABZA00000000	3	99.97	(Salzberg et al. 2009)
wMel	1.26	35.20	<i>Drosophila melanogaster</i>	AE017196	2	99.79	(Wu et al. 2004)
wMelPop	1.24	35.10	<i>Drosophila melanogaster</i>	AQQE01000000	1	99.79	(Woolfit et al. 2013)
wNo	1.30	34	<i>Drosophila simulans</i>	NC_021084	1	100	(Ellegaard et al. 2013)
wOo	0.95	32.10	<i>Onchocerca ochengi</i>	NC_018267	0	94.55	(Darby et al. 2012)
wBa	1.12	34	<i>Operophtera brumata</i>	NZ_JYPC01000000	2	100	(Derks et al. 2015)
wPip	1.48	34.20	<i>Culex quinquefasciatus</i>	NC_010981	1	100	(Klasson et al. 2008)
wPipMol	1.34	33.9	<i>Culex pipiens molestus</i>	HG428761	1	100	(Pinto et al. 2013)
wRec	1.12	35.20	<i>Drosophila recens</i>	JQAM00000000	0	100	(Metcalf et al. 2014)
wRi	1.44	35.20	<i>Drosophila simulans</i>	NC_012416	2	99.79	(Klasson et al. 2009)
wBm	1.08	34.20	<i>Brugia malayi</i>	NC_006833	0	99.36	(Foster et al. 2005)
wUni	0.86	35.20	<i>Muscidufurax uniraptor</i>	ACFP01000000	-	81.7	(Klasson et al. 2009)
wValsugana	1.41	35.20	<i>Drosophila suzukii</i>	CAOU00000000	2	81.27	(Siozios et al. 2013)
wVitB	1.10	33.7	<i>Narsonia vitripennis</i>	AERW01000000	0	99.79	(Kent et al. 2011)
wVulc	1.7	34.5	<i>Armadillidium vulgare</i>	ALWU00000000	3	99.57	(Cordaux et al. 2004, Cerveau et al. 2015)
wWill	1.09	35	<i>Drosophila willstone</i>	AAQP00000000	-	65	Remington et al. (Não publicado)

3.2 - Klebsiella

A grande prevalência de genomas identificados como pertencentes a espécie *K. pneumoniae* no banco de sequências do NCBI, quando comparada a outras espécies deste gênero como *K. quasipneumoniae* e *K. variicola*, é um reflexo de sua importância no ambiente clínico (Tabela 2). Para evitar a superrepresentação de

isolados desta espécie, agrupamos os genomas utilizando a abordagem de tipagem da sequência Multi Locus (MLST) com o esquema descrito para o gênero *Klebsiella* (Diancourt et al. 2005). Os genes constitutivos *gapA*, *infB*, *mdH*, *pgl*, *phoE*, *rpoB* e *tonB* foram recuperados do genoma do isolado tipo de *K. pneumoniae*, o isolado HS11286^T, e utilizadas como sondas em buscas por homologia com o software Blast+ (Camacho et al. 2009) em todos os genomas do gênero *Klebsiella*, onde sequências que apresentaram um *e*-value $>10^{-5}$ e uma identidade e cobertura com a sequência referência >40 e $>70\%$, respectivamente. As sequências recuperadas foram concatenadas, alinhadas com o software Mafft V5 (Kato et al. 2005), com parâmetros padrão, e utilizadas na construção de uma árvore filogenética baseada no método de *neighbor-joining* (Figura 3, 4) com o modelo Kimura 2 parâmetros.

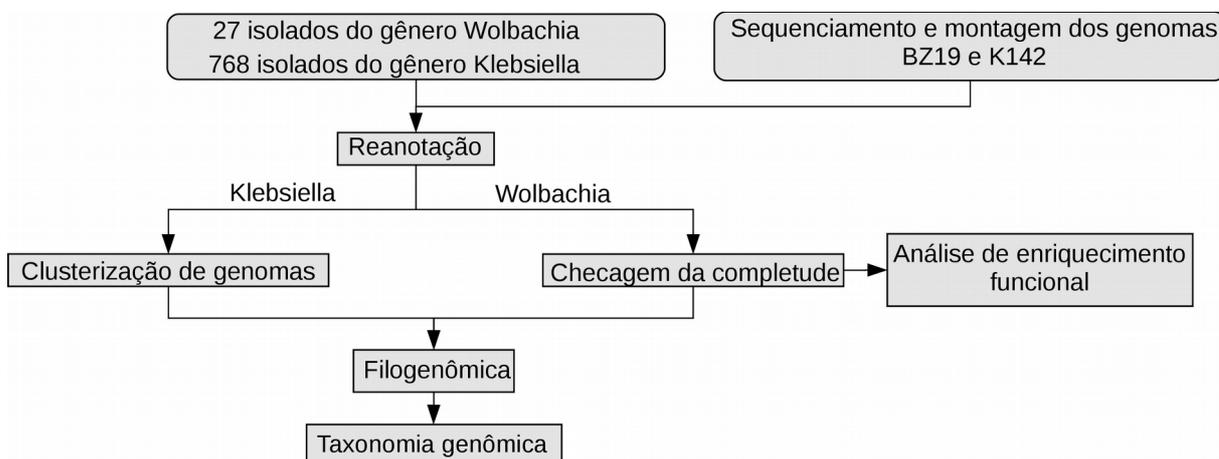


Figura 3 – Etapas seguídas para a obtenção dos genomas utilizados, reanotação, filtragem dos genomas dos dois gêneros (completude para *Wolbachia* e clusterização para *Klebsiella*) e análises subsequentes como taxonomia genômica, filogenômica e análise de enriquecimento funcional.

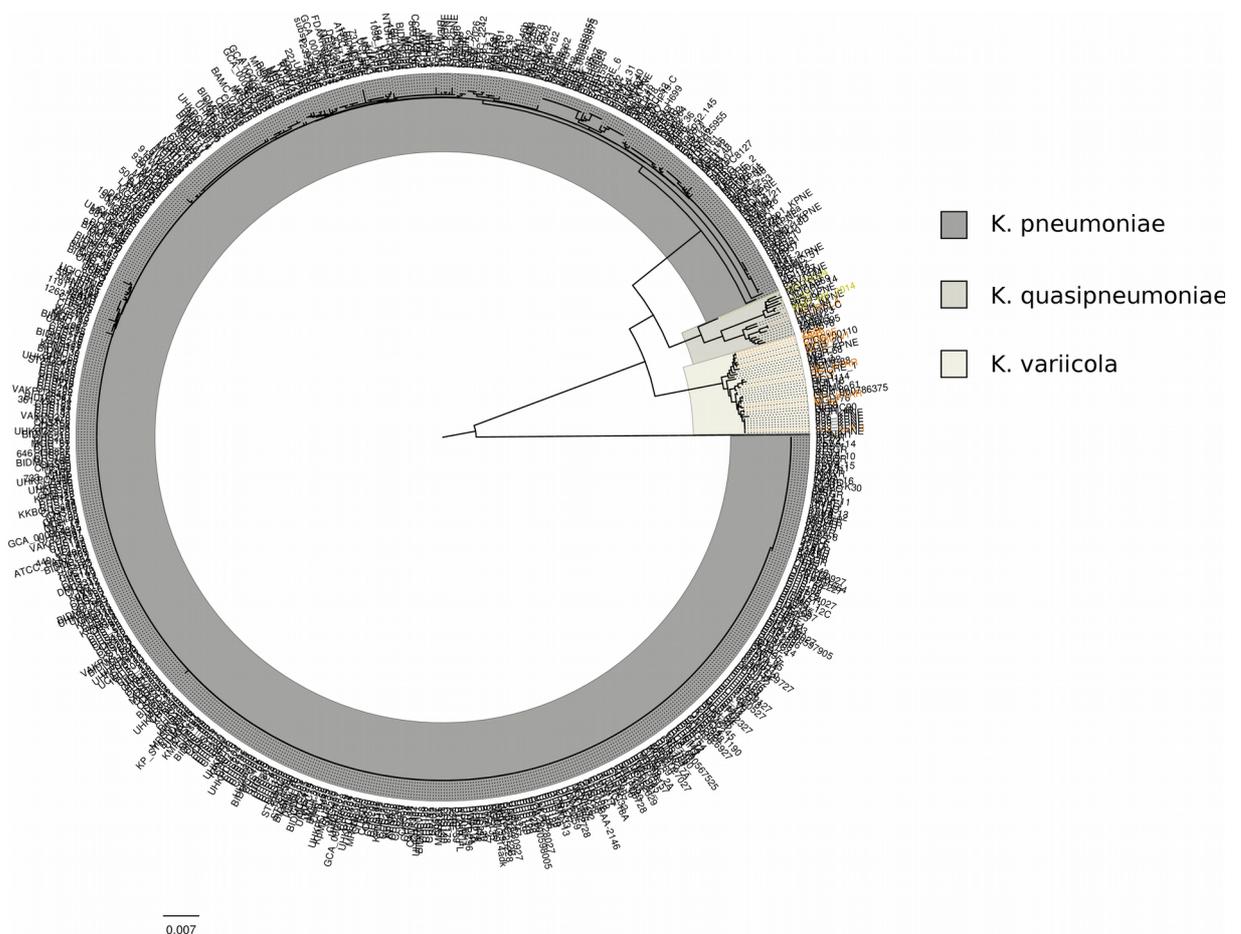


Figura 4 – Árvore filogenética construída a partir do super alinhamento de sequências de MLST de todos os genomas de isolados do gênero *Wolbachia* utilizaods (768). A árvore foi construída utilizando o método de *Neighbor joining* com o modelo Kimura 2 parâmetros e 100 replicatas de bootstrap. Espécies do gênero *Klebsiella* estão assinaladas em tons de cinza. Identificadores de taxas que foram correta ou incorretamente assinalados como *K. pneumoniae* estão em preto.

Todos os genomas que, por esta abordagem, agrupavam com as bactérias tipos das espécies *K. variicola* (DSM15968^T) e *K. quasipneumoniae* (01030^T) foram considerados. Cerca de 704 genomas ainda agrupavam com a espécie *K. pneumoniae*, então, para evitar redundância e super-representação, utilizamos apenas genomas completos de *K. pneumoniae* e genomas que não obtiveram 100% de identidade de MLST com nenhum outro isolado desta espécie. Devido a utilização deste filtro, o número de genomas classificados como pertencentes a espécie *K. pneumoniae* caiu de 704 para 33 (Tabela 2).

Tabela 2 – isolados do gênero *Klebsiella* de diferentes espécies, características genômicas e número de acesso

Isolado	Genoma (Mbp)	Conteúdo GC	Espécie	Acesso	Isolado
HS11286 ⁱ	5.68	54.14	<i>K. pneumoniae</i>	NC_016845	Clínico
NUTH-K2044	5.47	57.39	<i>K. pneumoniae</i>	NC_012731	Clínico
MGH78578	5.69	57.17	<i>K. pneumoniae</i>	GCA_000016305.1	Clínico
342	5.92	56.88	<i>K. pneumoniae</i>	GCA_000019565.1	Ambiental
KCTC2242	54.6	57.33	<i>K. pneumoniae</i>	GCA_000220485.1	Clínico
KPNIH10	5.76	57.14	<i>K. pneumoniae</i>	GCA_000281435.2	Clínico
KPNIH1	5.76	57.14	<i>K. pneumoniae</i>	GCA_000281535.2	Clínico
1084	5.38	57.40	<i>K. pneumoniae</i>	GCA_000294365.1	Clínico
BAA-2146	5.78	56.97	<i>K. pneumoniae</i>	NZ_CP006659	Clínico
UHKPC07	5.52	57.32	<i>K. pneumoniae</i>	NZ_CP011985	Clínico
JM45	5.60	57.24	<i>K. pneumoniae</i>	NC_022082	Clínico
KP-1	5.37	57.40	<i>K. pneumoniae</i>	NZ_CP012883	Ambiental
CG43	5.16	57.60	<i>K. pneumoniae</i>	NC_022566	Clínico
KP13	5.73	56.81	<i>K. pneumoniae</i>	NZ_CP003999	Clínico
KPNIH31	5.47	57.26	<i>K. pneumoniae</i>	NZ_CP009876	Clínico
KPNIH27	6.13	56.71	<i>K. pneumoniae</i>	NZ_CP007731	Clínico
XH209	5.11	57.60	<i>K. pneumoniae</i>	NZ_CP009461	Clínico
RYC492	5.38	57.40	<i>K. pneumoniae</i>	APGM01000001.1	Clínico
KPNIH29	5.52	57.8	<i>K. pneumoniae</i>	NZ_CP009863	Clínico
1158	5.34	57.20	<i>K. pneumoniae</i>	NZ_CP006722	Clínico
KPPR1	5.37	57.40	<i>K. pneumoniae</i>	NZ_CP009208	Clínico
PittNDM1	5.81	56.88	<i>K. pneumoniae</i>	NZ_CP006798	Clínico
PMK1	5.99	56.55	<i>K. pneumoniae</i>	NZ_CP008929	Clínico
HK787	5.32	57.50	<i>K. pneumoniae</i>	NZ_CP006738	N/A
BlaNDM-1	5.51	57.29	<i>K. pneumoniae</i>	NZ_CP009114	Clínico
34618	5.64	57.21	<i>K. pneumoniae</i>	NZ_CP010392	Clínico
KPNIH32	5.85	57.1	<i>K. pneumoniae</i>	NZ_CP009775	Clínico
KPNIH24	5.73	57.12	<i>K. pneumoniae</i>	NZ_CP008797	Clínico
88SGR	5.3	56.4	<i>K. pneumoniae</i>	NZ_CCIG00000000	Clínico
UHKPC45	5.62	57.10	<i>K. pneumoniae</i>	GCA_000409005	Clínico
UHKPC07	5.52	57.32	<i>K. pneumoniae</i>	NZ_CP011985	Clínico
32192	5.62	57.21	<i>K. pneumoniae</i>	NZ_CP010361	Clínico
KPNIH33	5.69	57.22	<i>K. pneumoniae</i>	NZ_CP009771	Clínico
NJST258_1	5.54	57.15	<i>K. pneumoniae</i>	NZ_CP006923	Clínico
KPNIH30	5.48	57.26	<i>K. pneumoniae</i>	NZ_CP009872	Clínico
700603	5.46	57.7	<i>K. pneumoniae</i>	GCA_000349245.1	Clínico
KPR0928	5.46	57.42	<i>K. pneumoniae</i>	NZ_CP008831	Clínico
1294_KPNE	5.09	58	<i>K. pneumoniae</i>	GCA_001060825	Clínico
BIDMC95	5.68	57.50	<i>K. pneumoniae</i>	GCA_001034125	Clínico
12-3578	5.74	57.10	<i>K. pneumoniae</i>	GCA_000367165	Clínico
HKUOPLC	5.08	58.00	<i>K. pneumoniae</i>	NZ_CP012300	Ambiental
KP1kgm	5.17	58.00	<i>K. pneumoniae</i>	GCA_000827665	Ambiental
MGH123	5.61	57.30	<i>K. pneumoniae</i>	GCA_001033915	Clínico
ATCC_43816	4.18	59.50	<i>K. pneumoniae</i>	GCA_000633025	N/A
MGH_44	5.24	57.70	<i>K. pneumoniae</i>	GCA_000492795	Clínico
UCICRE_14	5.53	57.8	<i>K. pneumoniae</i>	GCA_000492415	N/A
865_KPNE	5.42	57.8	<i>K. pneumoniae</i>	GCA_001067585	Clínico
MGH96	5.51	57.70	<i>K. pneumoniae</i>	GCA_001033665	Clínico
MGH113	5.78	57.10	<i>K. pneumoniae</i>	GCA_001033805	Clínico
1335_KPNE	5.79	57.20	<i>K. pneumoniae</i>	GCA_001076365	Clínico
BIDMC88	5.76	57.20	<i>K. pneumoniae</i>	GCA_001034045	Clínico
MGH-80	5.77	57.20	<i>K. pneumoniae</i>	GCA_000694775	Clínico
MGH-76	5.96	56.30	<i>K. pneumoniae</i>	GCA_000694735	Clínico
CH4-1	5.48	57.30	<i>K. pneumoniae</i>	GCA_001006625	Ambiental
BIDMC61	5.58	57.20	<i>K. pneumoniae</i>	GCA_000692995	Clínico
MGH114	6.02	57.00	<i>K. pneumoniae</i>	GCA_001033825	Clínico
UCI_18	6.03	56.90	<i>K. pneumoniae</i>	GCA_000567005	Clínico
CICC110	5.66	57.20	<i>K. pneumoniae</i>	GCA_000986855	Cultura
KP5-1	5.55	57.27	<i>K. pneumoniae</i>	CP008700.1	N/A
UCICRE_1	5.74	57.10	<i>K. pneumoniae</i>	GCA_000567525	N/A
MGH_36	5.56	57.30	<i>K. pneumoniae</i>	GCA_000492875.1	Clínico
MGH20	5.66	57.20	<i>K. pneumoniae</i>	GCA_000493095.1	N/A
B1-1	5.48	57.30	<i>K. pneumoniae</i>	GCA_001006575.1	Ambiental
MGH92	5.64	57.40	<i>K. pneumoniae</i>	GCA_001033575.1	Clínico
835_KPNE	5.61	57.30	<i>K. pneumoniae</i>	GCA_001066535.1	Clínico
836_KPNE	5.60	57.30	<i>K. pneumoniae</i>	GCA_001066555.1	Clínico
834_KPNE	5.60	57.30	<i>K. pneumoniae</i>	GCA_001067475.1	Clínico
833_KPNE	5.60	57.30	<i>K. pneumoniae</i>	GCA_001066505.1	Clínico
780_KPNE	5.60	57.30	<i>K. pneumoniae</i>	GCA_001066265.1	Clínico
837_KPNE	5.60	57.30	<i>K. pneumoniae</i>	GCA_001067495.1	Clínico
BIDMC90	5.66	57.20	<i>K. pneumoniae</i>	GCA_001034065.1	Clínico

MGH-68	5.52	57.20	<i>K. pneumoniae</i>	GCA_000694595.1	Clínico
ASM78637v1	5.6	57.20	<i>K. pneumoniae</i>	GCA_000786375	Clínico
07A044	5.10	58.20	<i>K. quasipneumoniae</i>	GCA_000613225.1	N/A
01A30	5.46	58.00	<i>K. quasipneumoniae</i>	GCA_000751755.1	Clínico
18A69	5.38	57.90	<i>K. quasipneumoniae</i>	GCA_000612905.1	N/A
FIHV2014	5.33	58.00	<i>K. quasipneumoniae</i>	GCA_001208885.1	Clínico
DSM_15968	5.52	57.60	<i>K. variicola</i>	NZ_CP010523.1	Ambiental
DX120E	5.71	57.07	<i>K. variicola</i>	NZ_CP009274.1	Ambiental
801	5.78	57	<i>K. variicola</i>	GCA_000826565.1	Clínico
AT22	5.45	57.60	<i>K. variicola</i>	GCA_0001025465.1	Ambiental
Bz19	5.54	57.34	<i>K. variicola</i>	GCA_00074291.1	Clínico
K1155	5.45	57.40	<i>K. spp</i>	GCA_000163075.1	Clínico
Kte92	5.47	57.30	<i>K. spp</i>	GCA_000398905.1	Clínico
10982	6.08	56.6	<i>K. spp</i>	GCA_000523395.1	Clínico
KONIH1	6.68	55.64	<i>K. oxytoca</i>	NZ_CP008788.1	Clínico
M1	6.31	55.85	<i>K. oxytoca</i>	NZ_CP008841.1	Ambiental
E718	6.56	55.52	<i>K. oxytoca</i>	NC_018106.1	Clínico
HKOPL1	5.91	55.90	<i>K. oxytoca</i>	NZ_CP004887.1	Ambiental
KCTC1686	5.97	56.00	<i>K. oxytoca</i>	NC_016612.1	N/A

3.3 - Sequenciamento e montagem dos genomas Bz19 e K142

Dois isolados clínicos (K142 e Bz19), pertencentes à Coleção de Bactérias do Ambiente e Saúde (CBAS), haviam sido previamente definidos por testes fenotípicos como membros da espécie *K. pneumoniae*, grupos filogenéticos KPII e III, respectivamente sendo, portanto, membros das espécies *K. quasipneumoniae subsp. Similipneumoniae* e *K. variicola* (Brisse 2014). Os genomas destes isolados foram extraídos utilizando o kit gDNA Mini bacteria da Invitrogen e sequenciados na plataforma de alto desempenho do Instituto Oswaldo Cruz com os sequenciadores Illumina Hi-seq 2500 e 454-GS Jr. Montagens *de novo* foram realizadas com os softwares A5 pipeline (Tritt et al. 2012) e Mira v2.7, respectivamente. Predições gênicas e anotações foram obtidas utilizando os softwares PGAAP (Angiuoli et al. 2008) e PROKKA V5 (Seemann 2014).

3.4 - Identificação dos genomas core e pan

Aplicamos o algoritmo OrthoMCL (Chen et al. 2006) com o *pipeline* GET_HOMOLOGUES (Contreras-Moreira & Vinuesa 2013) em todas as bactérias analisadas dos gêneros *Wolbachia* e *Klebsiella* separadamente, para que pudessemos identificar seus genomas *core* e *pan*. Todos os *Best Blast bidirectional Hits* (BBH) que apresentaram uma identidade $\geq 40\%$ e uma cobertura de $\geq 70\%$ foram considerados. Sequências dos genomas *core* foram alinhadas utilizando o software Mafft V5 (Katoh et al. 2005) e concatenadas em dois super-alinhamentos, um com as

bactérias de *Wolbachia* e outro com isolados de *Klebsiella*.

Wolbachia é um gênero conhecido por possuir isolados com grande densidade de pseudogenes (Foster et al. 2005, Slatko et al. 2010, Georgiades et al. 2011, Nikoh et al. 2014). Para identificar e remover possíveis pseudogenes dos grupos gênicos de isolados de *Wolbachia*, nós excluimos sequências que possuíam uma cobertura <70% com homólogos encontrados no banco de dados SEED (release 70) (Overbeek et al. 2014). A presença/ausência de genes do pan-genoma em isolados do gênero *Wolbachia*, assim como suas anotações funcionais foram compiladas em uma matriz.

3.5 – Filogenômica

A tipagem da sequência multilocus (MLST) é uma abordagem que extrapola a relação genética entre diferentes isolados a partir da similaridade de conjunto reduzido de genes, no caso do gênero *Wolbachia* o esquema de MLST inclui os genes *ftsZ*, *groEL*, *dnaA*, *fbpA*, *gatB*, *coxA* e o gene codificador do rRNA ribossomal 16s (Baldo et al. 2006b). Estes genes foram recuperados da linhagem tipo de *Wolbachia*, wPip^t, e utilizadas em buscas por homologia contra todos os genomas deste gênero analisados.

Semelhante a esta abordagem, a tipagem de sequência multilocus de proteínas ribossomais (rMLST) explora a diversidade de genes codificantes de subunidades de proteínas ribossomais de bactérias (*rps*), permitindo uma identificação rápida e acurada de isolados bacterianos a nível de espécie (Jolley et al. 2012, Bennett et al. 2013). Esta abordagem foi aplicada da seguinte forma: genes codificadores de *rps* foram recuperados dos genomas tipo das espécies *Wolbachia* wPip^t e do genoma do isolado tipo da espécie *K. pneumoniae*, HS11286^t, sendo utilizadas como sondas para buscas por similaridade utilizando os softwares Blastn + e tBlastn + em todos os genomas analisados .

A tipagem da sequência multilocus do genoma Core (cgMLST) é uma abordagem semelhante ao rMLST, mas utiliza o superalinhamento de todos os ortólogos compartilhados por um grupo de genomas para inferir distância em vez de genes codificadores de *rps* (Dutilh et al. 2007, Bennett et al. 2013). Sequências de MLST, rMLST e cgMLST foram alinhadas utilizando o *software* MAFFT v5 e concatenados em superalinhamentos.

Árvores filogenéticas e filogenômicas de máxima verossimilhança foram construídas utilizando o *software* RAxML (Stamatakis 2006) com o modelo General time Reversal (GTR). Os valores de confiança nos nós foram determinados utilizando 1000 replicatas de *bootstrap*. Matrizes de similaridade de MLST, rMLST e cgMLST foram construídas, sendo utilizadas na construção de *heatmaps* utilizando o método de clusterização euclidiano padrão na linguagem R.

3.6 - Taxonomia genômica

Aplicamos diferentes abordagens de taxonomia genômica em todos os genomas dos gêneros *Wolbachia* e *Klebsiella* utilizados, investigamos assim a diversidade taxonômica destes gêneros. As abordagens utilizadas foram: identidade média de nucleotídeos (ANI); hibridização DNA-DNA *in silico*; fração de alinhamento (AF); métricas de abordagens filogenômicas, como rMLST e cgMLST.

A identidade média de nucleotídeos (ANI) é um poderoso descritor de relação genética entre pares de genomas, esta abordagem utiliza a similaridade de sequência entre BBHs onde aquelas com identidade >96% pertenceriam a uma mesma espécie. Este limite tem uma forte correlação com o padrão ouro de 70% da abordagem hibridização DNA-DNA (Konstantinidis & Tiedje 2005b). ANI foi calculado segundo a seguinte metodologia, pares de homólogos foram recuperados utilizando a abordagem de *best blast hit* bidirecional com o *software* Blast+ (Camacho et al. 2009). BBHs apresentando similaridade $\geq 40\%$ e cobertura $\geq 70\%$ foram considerados homólogos. Os valores de ANI foram obtidos calculando a média da similaridade de todos os homólogos compartilhados por um par de genomas (Konstantinidis & Tiedje 2005a, Rosselló-Mora 2005).

O alinhamento de Fração (AF) do genoma foi calculado de forma bidirecional utilizando o script ANI_calculator_V1 (<https://ani.jgi-psf.org/html/anicalculator.php>), este script calcula o AF a partir da divisão da soma do tamanho do alinhamento de todos os genes BBHs pela soma do tamanho do alinhamento de todos os genes no genoma A (e vice-versa). Segundo esta abordagem, dois genomas que possuem um AF bidirecional maior ou igual a 0.6 pertencem a uma mesma espécie (Varghese et al. 2015).

A hibridização DNA-DNA digital foi realizada utilizando a abordagem *Genome Blast Distance Phylogeny* (GBDP) no servidor web GGDC (<http://http://ggdc.dsmz.de/>)

(Auch et al. 2006, Meier-Kolthoff et al. 2013). Esta abordagem aplica ferramentas de busca por homologia como BLAST+ para recuperar *High-score segments* (HSPs), estes segmentos são transformados em distâncias entre genomas através da aplicação de fórmulas específicas. Este método se mostrou suficientemente robusto para reproduzir resultados da hibridação DNA-DNA *in vitro*, sendo seu limite de delimitação de espécies $\geq 70\%$ (Meier-Kolthoff et al. 2013).

3.7 - Análise de enriquecimento/empobrecimento funcional.

Para explorar a diversidade funcional de isolados do gênero *Wolbachia*, nós aplicamos uma abordagem de enriquecimento/empobrecimento funcional, ou o quão frequente dada função está presente em um genoma ou conjunto de genomas quando comparado ao pan-genoma. Para a análise de enriquecimento do pan-genoma nós agrupamos a matriz de presença/ausência do pan-genoma utilizando o método de clusterização Euclidiano padrão na linguagem R (R Development Core Team 2011). Já para a análise do enriquecimento de funções centrado nos genomas, nós utilizamos os perfis de presença/ausência individuais destes genomas (Figura 5).

O enriquecimento/empobrecimento de funções foram acessados utilizando a classificação de subsistemas do SEED (Overbeek et al. 2005). Nós calculamos o valor de enriquecimento para cada nível funcional SEED usando a seguinte função, $\log_2(fg/fpg)$, onde fg = frequência da função em um grupo gênico ou genoma) e fpg = frequência da função no pan-genoma, o valor resultante indica se a frequência de uma determinada função é enriquecida (>0), neutra (0) ou empobrecida (<0) quando comparada a frequência desta função no pan-genoma. O suporte estatístico para esta análise foi feita através da reamostragem aleatória de n genes dentro destas funções 1000 vezes, onde n é o número de genes encontrados dentro do grupo gênico ou genoma analisado. O valor p foi definido como a frequência que uma função reamostrada obteria um valor de enriquecimento igual ou maior ao valor original.

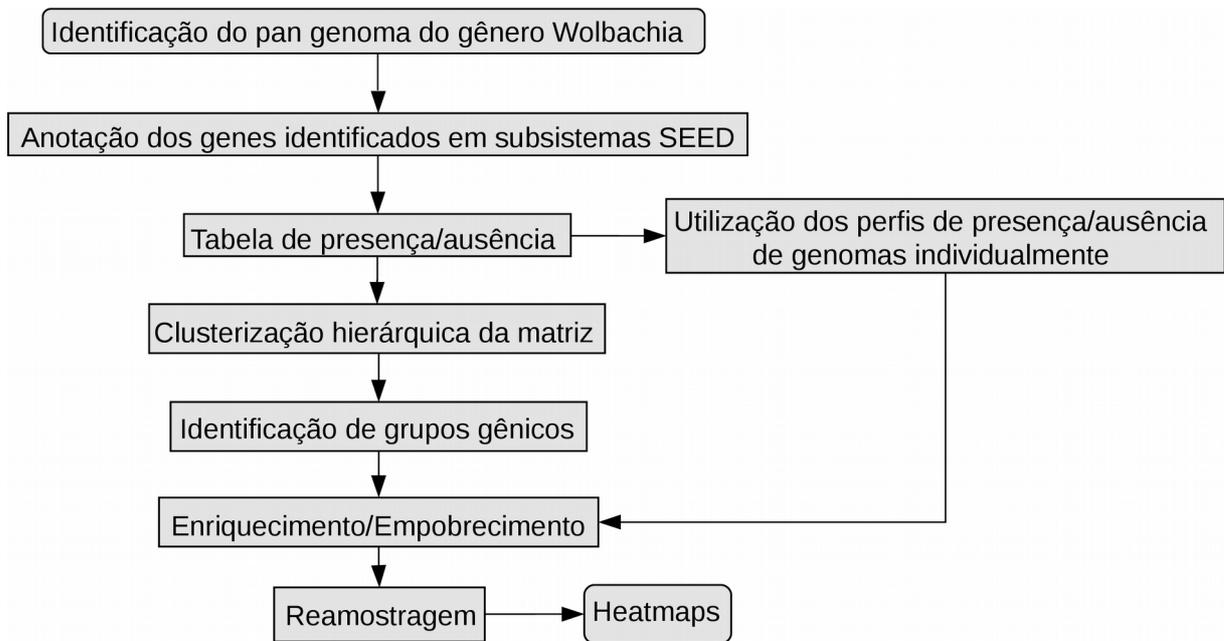


Figura 5 – Etapas seguidas para a identificação do pan genoma do gênero Wolbachia e obtenção da matriz de presença/ausência do pangenoma e o cálculo do enriquecimento funcional, focado tanto no pan-genoma de forma global quanto em genomas de forma individual.

3.8 - Determinação do código genético e identificação de genes mitocondriais adquiridos por HGT

Para a determinação da preferência por uso de códons, nós aplicamos o software FACIL (Dutilh et al. 2011) em todos os genomas do gênero Wolbachia analisados, este software prediz o código genético utilizado escaneando todas os genes identificados no genoma nas seis fases de leitura com homólogos de domínios Pfam (Eddy 1998). Regiões genômicas que potencialmente utilizavam um código genético diferente foram extraídos e utilizadas em buscas por homologia contra o banco de sequências não redundantes do NCBI (Kodama et al. 2012). Sequências homólogas identificadas foram alinhadas com o software MAFFT v5 (Kato et al. 2005). Análises filogenéticas foram realizadas empregando o software RaXML (Stamatakis 2006) utilizando o modelo WAG. Pró-fagos foram identificados utilizando três softwares diferentes, PFAST (Zhou et al. 2011), virSorter (Roux et al. 2015) e PhiSpy (Akhter et al. 2012).

4 - Resultados e discussão

Apresentaremos os resultados relativos aos dois gêneros, *Wolbachia* e *Klebsiella* de forma independente, como dois capítulos.

4.1 - *Wolbachia*

A maioria dos genomas de *Wolbachia* apresentou uma completude >95%, ou seja, foram recuperados acima de 234/247 genes marcadores para este gênero (Tabela 1). Os genomas wUni, wAna, wDacB, wBn e wWill apresentaram uma completude abaixo do recomendado pelos desenvolvedores deste software (>90%) (Parks et al. 2015), portanto foram excluídos de análises de core e pan genomas. A reanotação de genomas utilizando o software Prokka (Seemann 2014) revelou que a média do número de CDS por genomas era de 1322 (variando de 837 a 1,807), 34 tRNAs por genoma e nenhum CRISPR.

4.1.1 – Filogenômica

Nós aplicamos três metodologias distintas, MLST, rMLST e cgMLST, para as reconstruções filogenéticas e filogenômicas deste gênero. Para a abordagem MLST nós recuperamos os 7 genes do esquema descrito previamente de todos os genomas do gênero, com exceção de wWill pois não foi possível recuperar o gene *hscA* deste genoma, nós os excluímos de todas as análises subsequentes. Para o rMLST, foram recuperados um total de 26 genes de subunidades de proteínas ribossomais (*rps*) de todos os genomas de *Wolbachia* (n= 26). Este baixo número de *rps* (26/51) sugere o efeito da completude na recuperação dos genes codificadores de *rps*, já que nesta análise foram incluídos os genomas parciais (Completude <90 %).

Devido a isto, utilizamos apenas genomas que apresentaram uma completude > 90% para o método de cgMLST. Nós identificamos 368 genes pertencentes ao genoma Core de *Wolbachia*, excluindo grupos de ortólogos que continham

inparalágos, ou pares de parálogos que surgiram após um evento de especiação. As sequências nucleotídicas correspondentes ao rMLST e ao cgMLST foram alinhadas separadamente e concatenadas em superalinhamentos de 26 rps (11184 bp) e 368 genes (368.211 pb), respectivamente.

Árvores filogenética e filogenômicas resolutivas com altos valores de suporte nos nós internos (MLST, rMLST e cgMLST) (Figuras 6, 7 e 8) foram geradas a partir dos super alinhamentos gerados a partir dos três métodos. De um modo geral a topologia das três árvores são congruentes com os supergrupos previamente descritos (A-F), entretanto, diferentemente das árvores de rMLST e cgMLST, o isolado wVulC é encontrado em um nó interno do supergrupo B na árvore de MLST. De fato, este isolado foi previamente classificado como um isolado do supergrupo B por diversos trabalhos utilizando esta abordagem (Cordaux et al. 2004, Le Clec'h et al. 2012, Badawi et al. 2015, Cerveau et al. 2015), mas existem evidências indicando que três dos sete genes utilizados nesta reconstrução filogenética, *groEL*, *ftsZ* e *dnaA*, são recombinantes (Baldo et al. 2006a) e, por isso, a proximidade deste isolado com o resto do supergrupo B pode ser devido a inclusão destes genes. As topologias das árvores de rMLST e cgMLST são congruentes e, por utilizarem um número maior de genes, são menos propensos a estes erros inerentes indicando, portanto, que o isolado wVulC pertenceria a um supergrupo filogenético ainda não descrito.

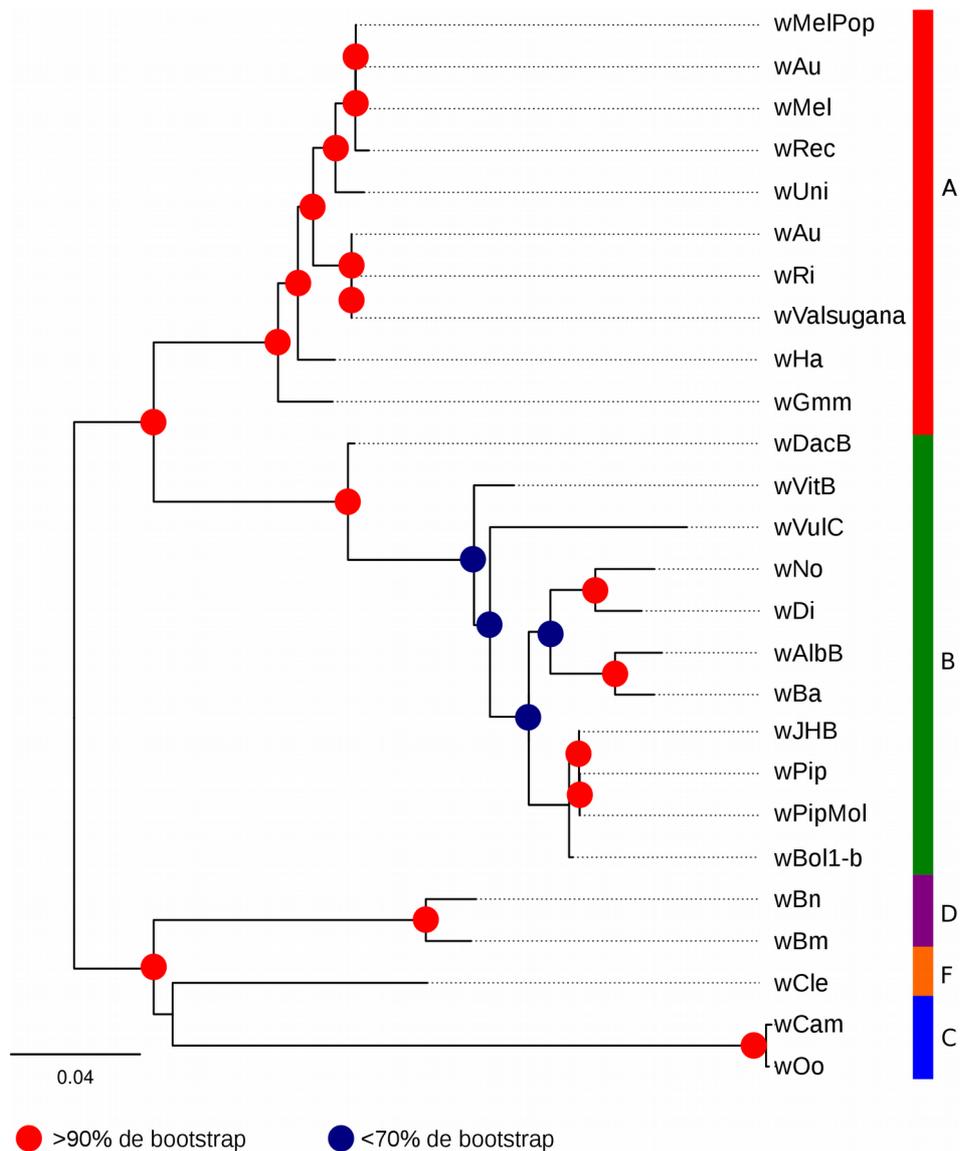


Figura 6 - Árvore filogenômica construída a partir do superalinhamento dos genes *ftsZ*, *groEL*, *dnaA*, *fbpA*, *gatB*, *coxA* de genomas de bactérias do gênero *Wolbachia*, os quais fazem parte do esquema de MLST descrito para este gênero. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap.

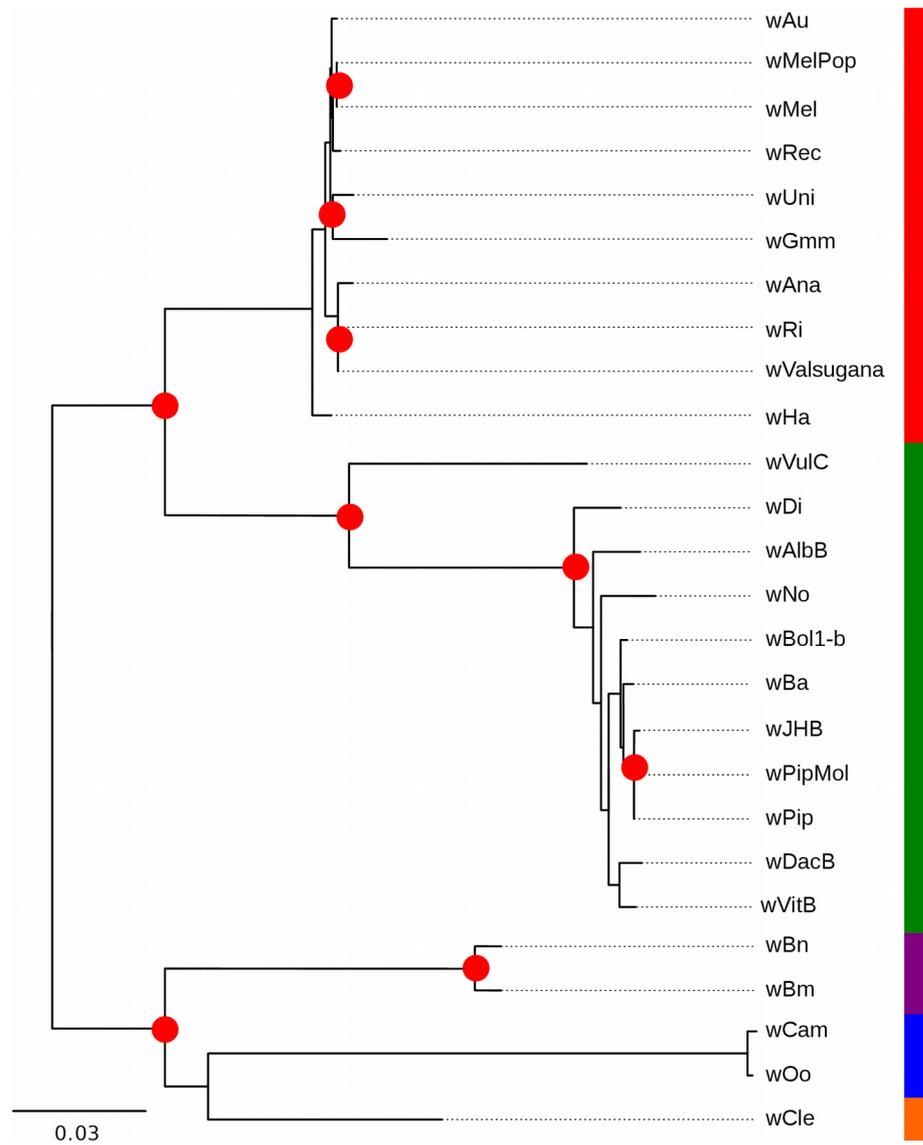


Figura 7 - Árvore filogenômica construída a partir do superalinhamento dos genes codificadores de subunidades de proteínas ribossomais (rps) de genomas de bactérias do gênero *Wolbachia*. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap. A árvore nos mostra cinco grupos bem definidos (A-D, F) representando os supergrupos já descritos (A-F) e o genoma wVulC, que foi previamente classificado como pertencente ao supergrupo B mas é claramente distante do grupo principal deste supergrupo.

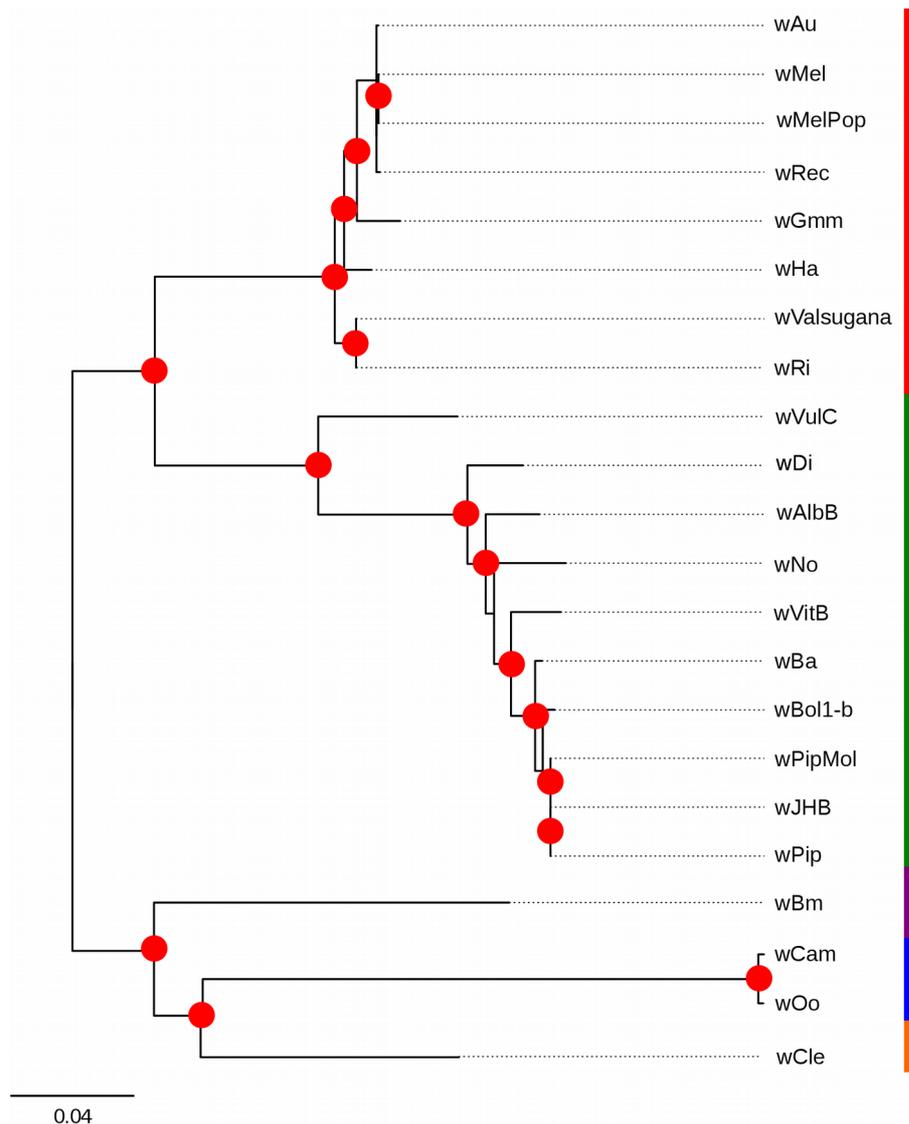


Figura 8 - Árvore filogenômica construída a partir do superalinhamento dos genes comuns a todos os genomas de de bactérias do gênero *Wolbachia* analisados, ou genoma *core*. A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR com correção para sítios invariáveis, o suporte nos nós foi calculado utilizando 1000 replicatas de bootstrap. A árvore nos mostra cinco grupos bem definidos (A-D, F) representando os supergrupos já descritos (A-F) e o genoma wVulC, que foi previamente classificado como pertencente ao supergrupo B mas é claramente distante do grupo principal deste supergrupo.

Assim como na árvore de MLST (Figura 6), é possível observar a formação de 5 grupos no heatmap construído a partir desta mesma matriz (Tabela 3, Figura 8), sendo pouco sensível para isolados do supergrupo B. Por esta abordagem as bactérias wVulC, wDacB e wAlbB são distantes de outros isolados do supergrupo B.

Em contrapartida, os heatmaps construídos a partir da matriz de similaridade das abordagens rMLST e cgMLST (Figuras 9 e 10) revelaram 6 grupos concisos, sendo o isolado wVuIC distante de todo o supergrupo B. Os valores de similaridade intragrupos filogenéticos variavam entre 97.9 e 100% (rMLST) e 96.3 e 100% (cgMLST) (Tabelas 4, 5), já os valores de similaridade intergrupos, para ambas as abordagens, eram menores de 90%. A exceção foi wVuIC, a similaridade deste genoma com o membro do supergrupo B mais próximo, wVitB, é de 91.15% (rMLST e cgMLST) e 89,15% de identidade com wUni, o membro do supergrupo A mais próximo a wVuIC. Portanto, os valores de similaridade e a topologia das árvores filogenômicas indicam que o genoma wVuIC pertenceria a um supergrupo distinto, entretanto, o cenário taxonômico deste gênero precisa ser revisto devido a inclusão de dois novos genomas, wVuIC e wBa e, para isto, nós aplicamos abordagens de taxonomia genômica.

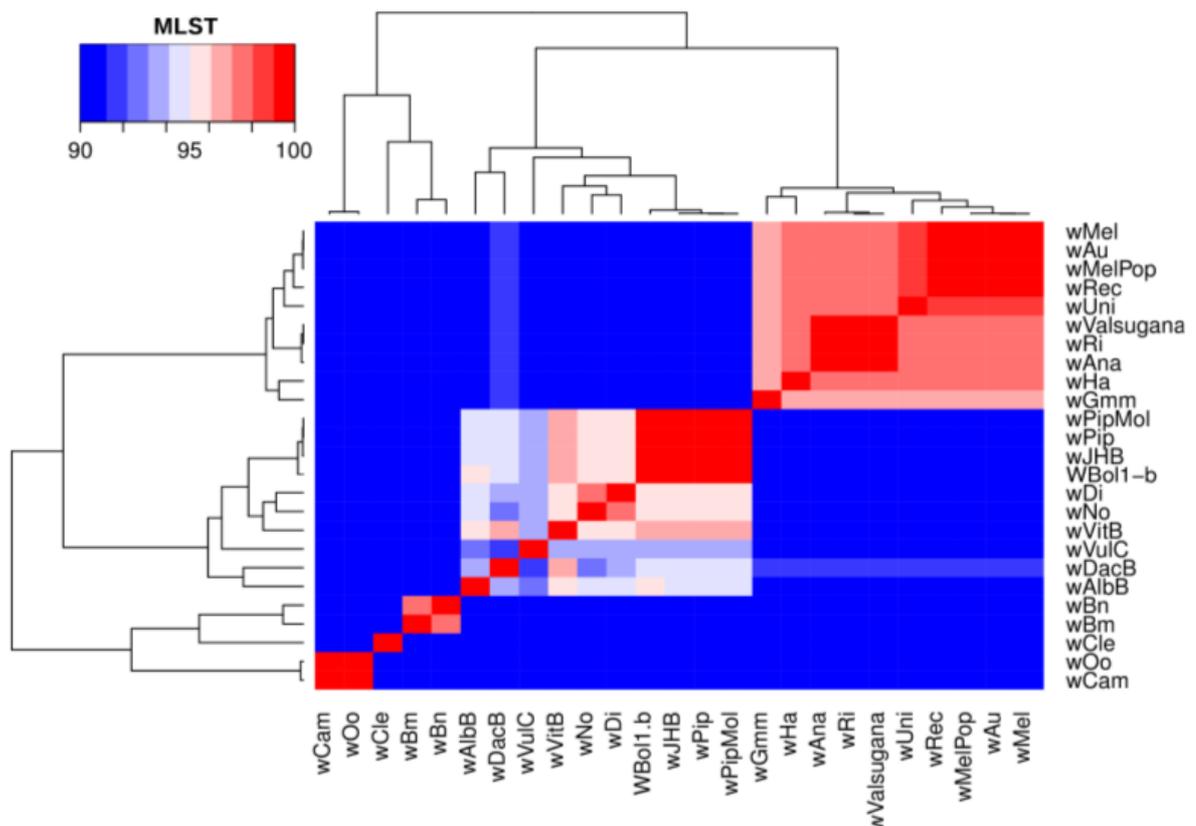


Figura 9 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes constitutivos que são parte do esquema de MLST para o gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O

dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R.

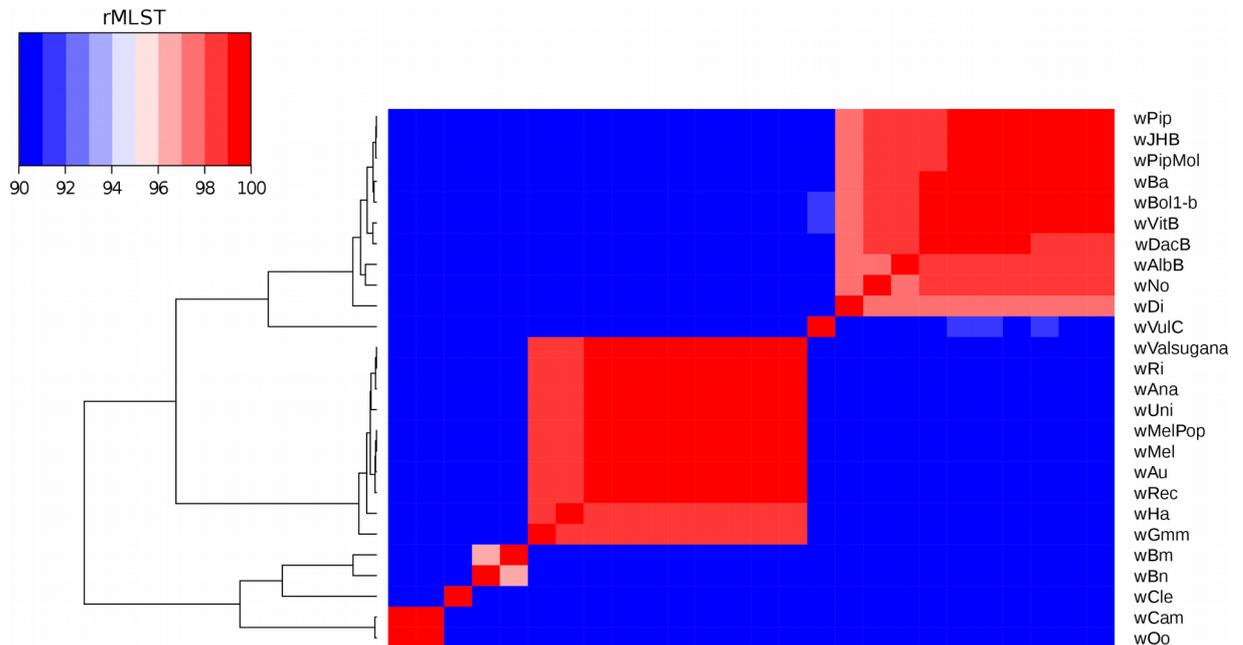


Figura 10 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes codificadores de proteínas ribossomais de bactérias do gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R.

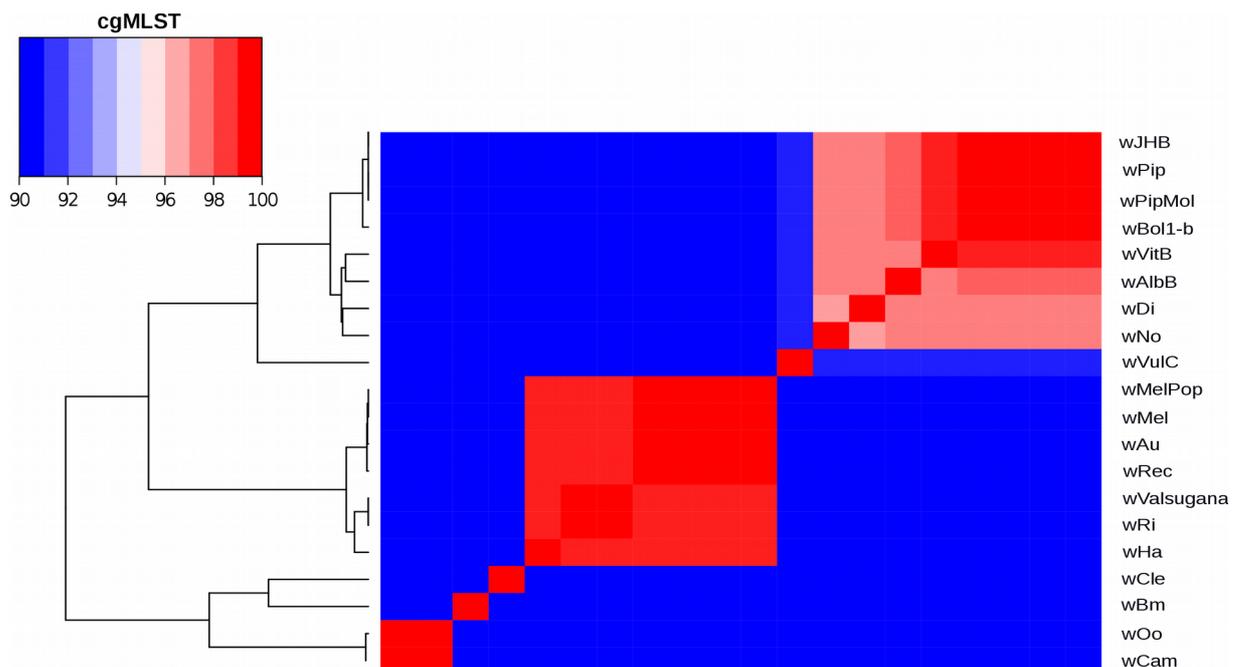


Figura 11 - Representação gráfica da matriz de similaridade todos contra todos, a qual foi construída a partir do superalinhamento de genes do genoma *core* do gênero *Wolbachia*, onde tons de azul significam valores <96%, branco 96% e tons de vermelho >96%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R.

Tabela 3 – Matriz de similaridade construída a partir do superalinhamento de sequências de genes intrínsecos (MLST) de isolados do gênero *Wolbachia*.

MLST	wAlb B	wAna	wAu	wBm	wBn	WBol 1-b	wCa m	wCle	wDac B	wDi	wGm m	wHa	wJHB	wMel	wMel Pop	wNo	wOo	wPip	wPip Mol	wRec	wRi	wUni	wVals ugana	wVitB	wVul C
wAlb B	100	89.55	89.86	85.81	85.72	95.07	82.23	86.38	93.04	94.26	89.2	89.91	94.95	89.86	89.86	94.99	82.23	94.93	94.93	89.44	89.5	89.44	89.5	95.01	92.62
wAna	89.55	100	97.87	86.58	86.56	86.95	82.84	86.81	91.28	87.85	97	97.54	86.89	97.87	97.83	86.98	82.92	86.87	86.87	97.53	100	97.44	100	87.79	88.08
wAu	89.86	97.87	100	86.76	87.01	87.35	83.09	87.34	91.5	87.81	96.96	97.54	87.29	100	100	86.55	83.16	87.27	87.27	99.59	97.86	98.6	97.86	87.37	88.28
wBm	85.81	86.58	86.76	100	97.37	85.55	83.71	87.54	85.93	85.06	86.47	86.72	85.58	86.76	86.76	84.94	83.76	85.55	85.55	86.39	86.48	86	86.48	85.06	85.22
wBn	85.72	86.56	87.01	97.37	100	86.08	84.13	87.27	86.24	85.27	87.1	87.16	85.92	87.01	86.96	84.98	84.18	85.89	85.89	86.56	86.56	86.33	86.56	85.21	85.59
WBol 1-b	95.07	86.95	87.35	85.55	86.08	100	83.42	86.56	95	95.95	87.32	87.45	99.6	87.35	87.35	95.58	83.52	99.58	99.58	87	86.98	86.86	86.98	96.25	93.5
wCa m	82.23	82.84	83.09	83.71	84.13	83.42	100	84.2	83.48	83.24	82.88	82.91	83.35	83.09	83.09	83.09	99.76	83.37	83.37	82.64	82.65	82.73	82.65	83.05	82.89
wCle	86.38	86.81	87.34	87.54	87.27	86.56	84.2	100	87.02	85.85	86.57	87.22	86.54	87.34	87.34	85.85	84.22	86.51	86.51	86.88	86.87	86.79	86.87	86.08	86.22
wDac B	93.04	91.28	91.5	85.93	86.24	95	83.48	87.02	100	93.11	91.41	91.69	95	91.5	91.5	92.99	83.59	94.97	94.97	91.1	91.39	91.45	91.39	96.1	91.59
wDi	94.26	87.85	87.81	85.06	85.27	95.95	83.24	85.85	93.11	100	87.29	87.77	95.67	87.81	87.81	97.16	83.34	95.65	95.65	87.5	87.85	87.16	87.85	95.85	93.22
wGm m	89.2	97	96.96	86.47	87.1	87.32	82.88	86.57	91.41	87.29	100	96.96	87.26	96.96	96.93	86.47	83	87.24	87.24	96.57	96.99	96.6	96.99	87.25	87.93
wHa	89.91	97.54	97.54	86.72	87.16	87.45	82.91	87.22	91.69	87.77	96.96	100	87.34	97.54	97.54	86.83	82.98	87.32	87.32	97.22	97.58	97.4	97.58	87.86	88.15
wJHB	94.95	86.89	87.29	85.58	85.92	99.6	83.35	86.54	95	95.67	87.26	87.34	100	87.29	87.29	95.26	83.44	99.98	99.98	86.94	86.87	86.79	86.87	96.46	93.33
wMel	89.86	97.87	100	86.76	87.01	87.35	83.09	87.34	91.5	87.81	96.96	97.54	87.29	100	100	86.55	83.16	87.27	87.27	99.59	97.86	98.6	97.86	87.37	88.28
wMel Pop	89.86	97.83	100	86.76	86.96	87.35	83.09	87.34	91.5	87.81	96.93	97.54	87.29	100	100	86.55	83.16	87.27	87.27	99.59	97.83	98.6	97.83	87.37	88.28
wNo	94.99	86.98	86.55	84.94	84.98	95.58	83.09	85.85	92.99	97.16	86.47	86.83	95.26	86.55	86.55	100	83.16	95.24	95.24	86.19	86.99	86.05	86.99	95.54	93.09
wOo	82.23	82.92	83.16	83.76	84.18	83.52	99.76	84.22	83.59	83.34	83	82.98	83.44	83.16	83.16	83.16	100	83.47	83.47	82.73	82.72	82.81	82.72	83.14	82.98
wPip	94.93	86.87	87.27	85.55	85.89	99.58	83.37	86.51	94.97	95.65	87.24	87.32	99.98	87.27	87.27	95.24	83.47	100	100	86.92	86.85	86.77	86.85	96.44	93.31
wPip Mol	94.93	86.87	87.27	85.55	85.89	99.58	83.37	86.51	94.97	95.65	87.24	87.32	99.98	87.27	87.27	95.24	83.47	100	100	86.92	86.85	86.77	86.85	96.44	93.31
wRec	89.44	97.53	99.59	86.39	86.56	87	82.64	86.88	91.1	87.5	96.57	97.22	86.94	99.59	99.59	86.19	82.73	86.92	86.92	100	97.51	98.18	97.51	87.02	87.83
wRi	89.5	100	97.86	86.48	86.56	86.98	82.65	86.87	91.39	87.85	96.99	97.58	86.87	97.86	97.83	86.99	82.72	86.85	86.85	97.51	100	97.44	100	87.77	88.22

wUni	89.44	97.44	98.6	86	86.33	86.86	82.73	86.79	91.45	87.16	96.6	97.4	86.79	98.6	98.6	86.05	82.81	86.77	86.77	98.18	97.44	100	97.44	86.83	87.85
wValsugana	89.5	100	97.86	86.48	86.56	86.98	82.65	86.87	91.39	87.85	96.99	97.58	86.87	97.86	97.83	86.99	82.72	86.85	86.85	97.51	100	97.44	100	87.77	88.22
wVitB	95.01	87.79	87.37	85.06	85.21	96.25	83.05	86.08	96.1	95.85	87.25	87.86	96.46	87.37	87.37	95.54	83.14	96.44	96.44	87.02	87.77	86.83	87.77	100	93.83
wVulC	92.62	88.08	88.28	85.22	85.59	93.5	82.89	86.22	91.59	93.22	87.93	88.15	93.33	88.28	88.28	93.09	82.98	93.31	93.31	87.83	88.22	87.85	88.22	93.83	100

Tabela 4 – Matriz de similaridade construída a partir do superalinhamento de sequências de rps de isolados do gênero Wolbachia.

Genomas	wAlb	wAna	wAu	wBa	wBm	wBn	wBol	wCa	wCle	wDac	wDi	wGm	wHa	wJH	wMel	wMel	wNo	wOo	wPip	wPipMol	wRec	wRi	wUni	wValsugana	wVitB	wVulC
wAlb	100	86,66	86,82	98,16	79,53	75,9	98,27	77,01	82,03	98,12	97,45	85,74	86,86	98,06	86,78	86,78	97,5	76,97	98,06	98,04	86,68	86,93	87,51	86,93	98,18	88,78
wAna	86,66	100	99,19	86,91	84,39	79,93	86,85	80,34	85,63	86,74	87,34	98,05	98,72	86,92	99,19	99,19	86,35	80,34	86,93	86,86	99,09	99,75	98,8	99,75	86,58	86,92
wAu	86,82	99,19	100	87,02	84,61	80,2	87	80,42	85,72	86,82	87,46	98,59	99,04	86,99	99,76	99,76	86,52	80,42	87	86,94	99,67	99,43	99,25	99,43	86,64	87,21
wBa	98,16	86,91	87,02	100	80,14	76,42	99,57	77,32	82,61	98,9	98,04	85,96	87,12	99,54	86,98	86,98	98,22	77,36	99,54	99,54	86,87	87,16	87,72	87,16	99,13	88,99
wBm	79,53	84,39	84,61	80,14	100	94,31	80,16	81,03	87,22	79,92	80,32	83,69	84,53	80,23	84,53	84,53	79,94	80,99	80,24	80,2	84,5	84,46	84,92	84,46	79,9	80,28
wBn	75,9	79,93	80,2	76,42	94,31	100	76,45	77,36	83,4	76,05	76,5	79,16	80,29	76,49	80,15	80,15	76,27	77,36	76,5	76,42	80,1	80,2	80,41	80,2	76,23	77,12
wBol	98,27	86,85	87	99,57	80,16	76,45	100	77,21	82,51	98,91	98,24	85,99	87,14	99,55	86,96	86,96	98,24	77,25	99,55	99,55	86,86	87,14	87,7	87,14	99,08	89,12
wCa	77,01	80,34	80,42	77,32	81,03	77,36	77,21	100	82,71	77,05	77,55	79,62	80,46	77,37	80,38	80,38	76,9	99,7	77,37	77,27	80,38	80,5	80,86	80,5	76,7	77,51
wCle	82,03	85,63	85,72	82,61	87,22	83,4	82,51	82,71	100	82,38	82,41	84,64	85,83	82,5	85,65	85,65	82,09	82,86	82,51	82,56	85,54	85,76	85,9	85,76	82,14	82,3
wDac	98,12	86,74	86,82	98,9	79,92	76,05	98,91	77,05	82,38	100	97,72	85,81	86,96	98,64	86,78	86,78	98,03	77,09	98,64	98,63	86,68	86,96	87,51	86,96	99,21	88,88
wDi	97,45	87,34	87,46	98,04	80,32	76,5	98,24	77,55	82,41	97,72	100	86,58	87,46	97,9	87,43	87,43	97,19	77,47	97,9	97,89	87,32	87,61	88,21	87,61	97,69	88,71
wGm	85,74	98,05	98,59	85,96	83,69	79,16	85,99	79,62	84,64	85,81	86,58	100	97,88	85,92	98,59	98,59	85,61	79,62	85,92	85,85	98,5	98,28	98,18	98,28	85,68	85,99
wHa	86,86	98,72	99,04	87,12	84,53	80,29	87,14	80,46	85,83	86,96	87,46	97,88	100	87,1	98,98	98,98	86,59	80,46	87,11	87,04	98,89	98,95	98,69	98,95	86,79	87,21
wJH	98,06	86,92	86,99	99,54	80,23	76,49	99,55	77,37	82,5	98,64	97,9	85,92	87,1	100	86,96	86,96	98,15	77,41	100	100	86,85	87,14	87,69	87,14	98,85	89,05
wMel	86,78	99,19	99,76	86,98	84,53	80,15	86,96	80,38	85,65	86,78	87,43	98,59	98,98	86,96	100	100	86,48	80,38	86,96	86,9	99,73	99,43	99,25	99,43	86,6	87,17

wMel	86,78	99,19	99,76	86,98	84,53	80,15	86,96	80,38	85,65	86,78	87,43	98,59	98,98	86,96	100	100	86,48	80,38	86,96	86,9	99,73	99,43	99,25	99,43	86,6	87,17	
Pop																											
wNo	97,5	86,35	86,52	98,22	79,94	76,27	98,24	76,9	82,09	98,03	97,19	85,61	86,59	98,15	86,48	86,48	100	76,94	98,15	98,13	86,37	86,66	87,16	86,66	98,21	88,73	
wOo	76,97	80,34	80,42	77,36	80,99	77,36	77,25	99,7	82,86	77,09	77,47	79,62	80,46	77,41	80,38	80,38	76,94	100	77,41	77,31	80,38	80,5	80,86	80,5	76,74	77,43	
wPip	98,06	86,93	87	99,54	80,24	76,5	99,55	77,37	82,51	98,64	97,9	85,92	87,11	100	86,96	86,96	98,15	77,41	100	100	86,86	87,14	87,7	87,14	98,85	89,05	
wPip	98,04	86,86	86,94	99,54	80,2	76,42	99,55	77,27	82,56	98,63	97,89	85,85	87,04	100	86,9	86,9	98,13	77,31	100	100	86,79	87,08	87,64	87,08	98,85	88,97	
Mol																											
wRec	86,68	99,09	99,67	86,87	84,5	80,1	86,86	80,38	85,54	86,68	87,32	98,5	98,89	86,85	99,73	99,73	86,37	80,38	86,86	86,79	100	99,34	99,16	99,34	86,48	87,06	
wRi	86,93	99,75	99,43	87,16	84,46	80,2	87,14	80,5	85,76	86,96	87,61	98,28	98,95	87,14	99,43	99,43	86,66	80,5	87,14	87,08	99,34	100	99,1	100	86,86	87,28	
wUni	87,51	98,8	99,25	87,72	84,92	80,41	87,7	80,86	85,9	87,51	88,21	98,18	98,69	87,69	99,25	99,25	87,16	80,86	87,7	87,64	99,16	99,1	100	99,1	87,4	87,62	
wVal	86,93	99,75	99,43	87,16	84,46	80,2	87,14	80,5	85,76	86,96	87,61	98,28	98,95	87,14	99,43	99,43	86,66	80,5	87,14	87,08	99,34	100	99,1	100	86,86	87,28	
suga																											
na																											
wVitB	98,18	86,58	86,64	99,13	79,9	76,23	99,08	76,7	82,14	99,21	97,69	85,68	86,79	98,85	86,6	86,6	98,21	76,74	98,85	98,85	86,48	86,86	87,4	86,86	100	88,98	
wVul	88,78	86,92	87,21	88,99	80,28	77,12	89,12	77,51	82,3	88,88	88,71	85,99	87,21	89,05	87,17	87,17	88,73	77,43	89,05	88,97	87,06	87,28	87,62	87,28	88,98	100	
C																											

Tabela 5 – Matriz de similaridade construída a partir do superalinhamento de genes do genoma core de isolados do gênero Wolbachia.

Genomas	wAlbB	wAu	wBa	wBm	wBol1-b	wCam	wCle	wDi	wGmm	wHa	wJHB	wMel	wMelPop	wNo	wOo	wPip	wPipMol	wRec	wRi	wValsunaga	wVitB	wVulC
wAlbB	100	86.36	97.11	81.49	96.99	78.91	82.3	96.37	86.01	86.36	97.05	86.36	86.36	86.41	78.92	97.05	97.05	86.33	86.41	86.41	96.49	90.87
wAu	86.36	100	86.43	83.61	86.42	80.1	84.11	86.41	98	98.19	86.41	99.87	99.87	78.95	80.13	86.41	86.41	99.83	98.1	98.11	86.63	87.68
wBa	97.11	86.43	100	81.58	99.22	78.95	82.41	96.66	86.02	86.44	99.32	86.43	86.43	91.04	78.97	99.32	99.32	86.4	86.48	86.49	97.61	91.04
wBm	81.49	83.61	81.58	100	81.54	80.35	84.26	81.53	83.29	83.61	81.59	83.61	83.61	100	80.37	81.59	81.59	83.58	83.63	83.63	81.64	81.98
WBol1-b	96.99	86.42	99.22	81.54	100	78.91	82.35	96.46	86.04	86.46	99.38	86.42	86.42	82.4	78.93	99.37	99.37	86.4	86.46	86.46	97.54	91.04
wCam	78.91	80.1	78.95	80.35	78.91	100	81.54	78.95	79.77	80.11	78.94	80.11	80.11	99.99	99.68	78.93	78.94	80.08	80.14	80.15	79.07	79.24
wCle	82.3	84.11	82.41	84.26	82.35	81.54	100	82.35	83.67	84.08	82.41	84.11	84.11	86.38	81.56	82.4	82.41	84.07	84.1	84.09	82.41	82.83
wDi	96.37	86.41	96.66	81.53	96.46	78.95	82.35	100	86.02	86.37	96.53	86.41	86.41	78.93	78.98	96.53	96.53	86.38	86.4	86.4	96.16	90.91
wGmm	86.01	98	86.02	83.29	86.04	79.77	83.67	86.02	100	97.62	86	97.99	97.99	81.59	79.81	86	86	97.95	97.66	97.67	86.2	87.29
wHa	86.36	98.19	86.44	83.61	86.46	80.11	84.08	86.37	97.62	100	86.4	98.17	98.17	100	80.13	86.4	86.4	98.14	98.2	98.21	86.62	87.67
wJHB	97.05	86.41	99.32	81.59	99.38	78.94	82.41	96.53	86	86.4	100	86.41	86.41	86.44	78.96	100	100	86.38	86.43	86.44	97.53	91.04
wMel	86.36	99.87	86.43	83.61	86.42	80.11	84.11	86.41	97.99	98.17	86.41	100	99.99	86	80.14	86.41	86.41	99.82	98.08	98.09	86.63	87.67

wMelPop	86.36	99.87	86.43	83.61	86.42	80.11	84.11	86.41	97.99	98.17	86.41	99.99	100	86.41	80.14	86.41	86.41	99.82	98.09	98.09	86.63	87.67
wNo	96.34	86.14	96.66	81.36	96.51	78.79	82.24	95.9	85.74	86.12	96.57	86.14	86.15	100	78.82	96.57	96.56	86.11	86.16	86.17	96.23	90.87
wOo	78.92	80.13	78.97	80.37	78.93	99.68	81.56	78.98	79.81	80.13	78.96	80.14	80.14	86.43	100	78.95	78.96	80.1	80.16	80.17	79.09	79.25
wPip	97.05	86.41	99.32	81.59	99.37	78.93	82.4	96.53	86	86.4	100	86.41	86.41	99.32	78.95	100	100	86.38	86.43	86.44	97.53	91.04
wPipMol	97.05	86.41	99.32	81.59	99.37	78.94	82.41	96.53	86	86.4	100	86.41	86.41	96.53	78.96	100	100	86.38	86.43	86.44	97.54	91.04
wRec	86.33	99.83	86.4	83.58	86.4	80.08	84.07	86.38	97.95	98.14	86.38	99.82	99.82	86.4	80.1	86.38	86.38	100	98.05	98.06	86.59	87.64
wRi	86.41	98.1	86.48	83.63	86.46	80.14	84.1	86.4	97.66	98.2	86.43	98.08	98.09	99.37	80.16	86.43	86.43	98.05	100	99.98	86.68	87.73
wValsunaga	86.41	98.11	86.49	83.63	86.46	80.15	84.09	86.4	97.67	98.21	86.44	98.09	98.09	100	80.17	86.44	86.44	98.06	99.98	100	86.68	87.73
wVitB	96.49	86.63	97.61	81.64	97.54	79.07	82.41	96.16	86.2	86.62	97.53	86.63	86.63	96.23	79.09	97.53	97.54	86.59	86.68	86.68	100	91.14
wVulC	90.87	87.68	91.04	81.98	91.04	79.24	82.83	90.91	87.29	87.67	91.04	87.67	87.67	97.05	79.25	91.04	91.04	87.64	87.73	87.73	91.14	100

4.1.2 - Análises de taxonomia genômica

ANI e GGD foram calculados entre todos os pares de genomas de *Wolbachia* segundo a metodologia descrita previamente. Em geral, as abordagens utilizadas são consistentes com a delimitação prévia de cinco espécies, isto é, a diversidade calculada por ANI e GGD estavam, em sua maioria, acima do limite pré estabelecido para delimitação de espécies bacterianas de 96 % e 70%, respectivamente (Konstantinidis & Tiedje 2005b, Auch et al. 2006, Meier-Kolthoff et al. 2013) (Tabela 6, 7). A exceção foi o isolado wVulC, este isolado previamente assinalada como pertencente ao supergrupo B apresentou métricas abaixo do limite de delimitação com qualquer *Wolbachia* analisada (Figuras 12, 13 e 14). Nós identificamos também três isolados pertencentes a *Ca. 'W. pipientis'* que apresentavam valores de GGD abaixo do limite de delimitação (<63%) (Figura 13), esses valores podem explicar suas posições periféricas em árvores filogenômicas, entretanto, esta divergência não foi confirmado por outras abordagens (Ramírez-Puebla et al. 2015).

Estes resultados indicam que wVulC não pertenceria a nenhuma espécie candidata descrita previamente, portanto, nós propomos a criação de uma nova espécie candidata chamada *Ca. 'W. vulgaris'*, cujo único representante identificado até o momento é o isolado wVulC. Não consideramos o Alinhamento de Fração (AF) para isolados pois o mesmo é dependente do número de ortólogos compartilhados entre organismos (Varghese et al. 2015) e, como *Wolbachias* possuem genomas em processo de degeneração, com pouca ou nenhuma aquisição por HGT, esta metodologia não é aplicável devido a perda de sensibilidade quando comparada a outras abordagens (Figura 14, Tabela 8).

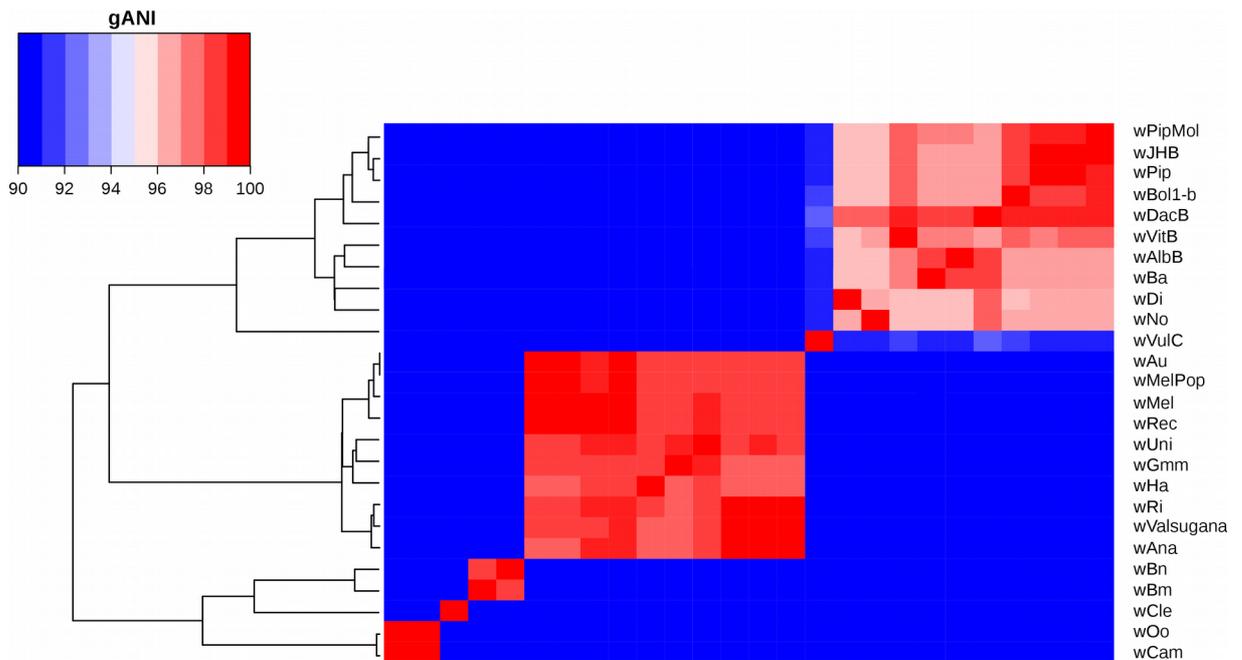


Figura 12 - Representação gráfica da matriz da identidade média de nucleotídeos entre todos os pares de genomas do gênero *Wolbachia*, onde tons de azul significam valores <95%, branco 95% e tons de vermelho >95%. O dendrograma foi construído utilizando o método euclidiano padrão disponível na linguagem R.

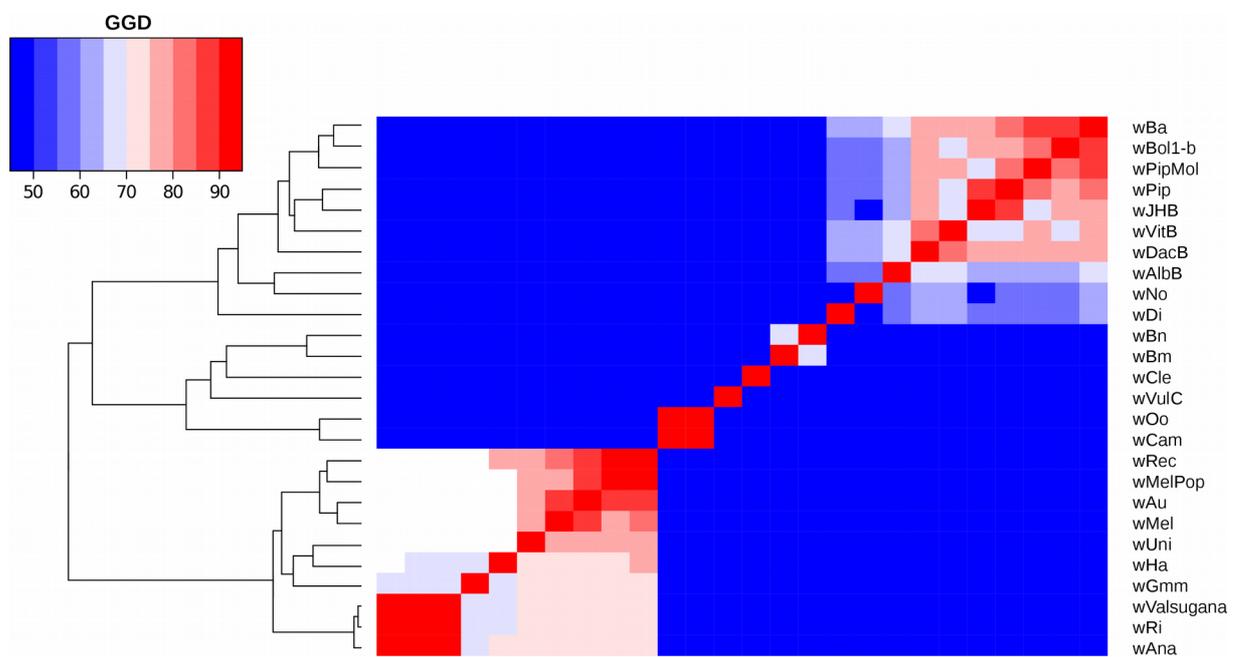


Figura 13 - Representação gráfica da matriz de hibridização DNA DNA digital (GGD) entre todos os pares de genomas, o dendrograma foi construído utilizando o método euclidiano encontrado na linguagem R, onde tons de azul significam valores <70%, branco 70% e tons de vermelho >70%.

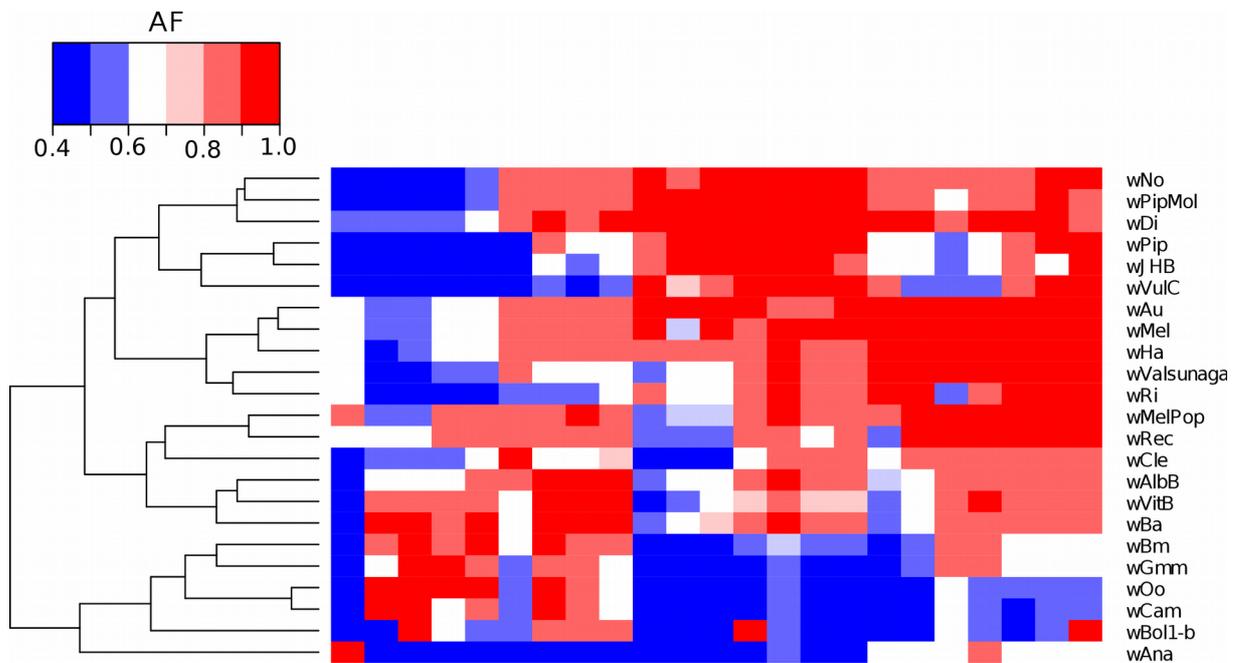


Figura 14 - Representação gráfica da matriz de alinhamento de fração entre todos os pares de genomas, o dendrograma foi construído utilizando o método euclidiano encontrado na linguagem R, onde tons de azul significam valores <0.6, branco 0.6 e tons de vermelho >0.6.

Wolbachia é um gênero com isolados geneticamente diversos (Ishmael et al. 2009), foi estimado, baseado em taxas de substituições não sinônimas do gene *ftsZ*, que isolados do supergrupo A e B divergiram aproximadamente 60-70 milhões de anos atrás (John H. Werren, Wan Zhang 1995), após o fim do período cretáceo. De fato, o sistema de classificação baseada em supergrupos é uma forma flexível de classificar isolados de Wolbachia, já que não existem regras definitivas para sua criação e, cada supergrupo tem um significado específico para os especialistas em Wolbachia. Esse esquema de classificação auxiliou especialistas ao longo dos anos e foi uma das razões da proposta que todo isolado de Wolbachia deveria ser classificada como pertencente a espécie *W. pipientis*, por Lo et al., (2001) (Lo et al. 2007). Este esquema permitiu também o agrupamento de isolados divergentes, como wVulC, com outros isolados do supergrupo B, já que a mesma não leva em conta a heterogeneidade gênomicas destes pois o agrupamento é baseado em marcadores de genes únicos, como o gene *ftsZ*, ou em um pequeno número de genes intrínsecos, como o esquema de MLST para o gênero Wolbachia (Baldo et al. 2006b).

Tabela 6 – Matriz construída a partir dos valores de identidade média de nucleotídeos de isolados do gênero *Wolbachia*.

Genomas	wAlb B	wAna	wAu	wBa	wBm	wBn	wBol1-b	wCam	wCle	wDacB	wDi	wGmm	wHa	wJHB	wMel	wMelpop	wNo	wOo	wPip	wPipMol	wRec	wRi	wUni	wValsugana	wVitB	wVulc
wAlb B	100	88.3	88	96.7	84.8	86.2	95.7	83.7	85.6	97.1	95.2	88.1	87.9	95.6	88.1	88	95.4	83.7	95.9	95.9	87.9	88.1	88.1	88.1	96	90.9
wAna	87.9	100	96.9	88.5	86.6	87.6	87.8	84.3	86.5	88.3	87.9	96.9	96.8	87.6	97.5	96.9	88	84.3	87.5	87.5	97.6	99.1	97.1	99.1	87.9	89.2
wAu	88	97.4	100	88.4	86.4	87.4	88.2	84.3	86.6	88	88.1	97	97.2	87.7	99.4	100	87.7	84.3	87.7	87.8	98.9	97.2	97.4	97.3	88.2	88.9
wBa	96.7	88.5	88.4	100.0	85.7	85.7	98.9	84.5	86.2	95.7	96.2	88.1	88.4	98.8	88.4	88.3	96.0	84.5	98.8	98.7	88.4	88.4	88.1	88.4	97.4	91.2
wBm	84.8	86.8	86.4	85.7	100	97.4	85	84.7	86.8	86.3	84.9	86.3	86.4	85	86.4	86.4	84.8	84.6	85	85	86.4	86.4	86.4	86.5	85.1	86.1
wBn	85.2	87.2	86.6	85.7	97.3	100	85.2	85	87.1	86.9	85.3	86.7	86.7	85.4	86.6	86.6	85.1	85	85.3	85.4	86.5	86.6	87.1	86.6	85.3	86.1
wBol1-b	96	88.4	88	98.9	85	86.3	100	83.8	85.7	95.6	95.3	88.2	88.1	97.4	88	88	95.5	83.8	97.5	97.8	87.9	87.9	88	88	96.7	91.1
wCa m	83.7	84.6	84.2	84.5	84.7	85.6	83.8	100	85.2	84.8	83.9	84.2	84.2	83.8	84.3	84.2	83.8	99.6	83.8	83.8	84.2	84.3	84.3	84.3	83.8	84.5
wCle	85.7	87.1	86.7	86.2	86.8	87.6	85.7	85.2	100	86.8	85.5	86.7	86.5	85.5	86.7	86.7	85.7	85.2	85.6	85.6	86.7	86.6	86.9	86.7	85.7	86.8
wDacB	97.1	88.3	88	95.7	85.6	87.1	98.3	84.2	86.1	100	96.8	88.1	87.9	98.5	88.1	88	96.7	84.2	98.6	98.5	87.9	88.1	88.1	88.1	98.5	91.9
wDi	95.4	88.3	88.1	96.2	84.9	86.1	95.1	83.8	85.8	96.8	100	88.3	87.9	94.7	88.1	88.1	94.9	83.9	95	95.2	88	87.8	88.3	88	95.5	90.9
wGmm	87.7	96.7	97.3	88.1	86.1	87.2	87.7	84.2	86.4	88.1	87.8	100	97.2	87.7	97.1	97.3	87.6	84.2	87.8	87.7	97.3	96.8	97.6	96.5	88.1	89.0
wHa	88	96.9	96.8	88.4	86.4	87.3	88.5	84.2	86.5	87.9	88	97	100	88	97.2	96.8	87.9	84.2	87.9	87.8	97.1	96.9	97.2	96.7	88	89.1
wJHB	96	88.3	87.9	98.8	85	86.3	97.4	83.8	85.5	95.7	95.1	88.1	87.8	100	87.9	87.9	95.3	83.8	99.3	99.1	87.9	87.8	88.1	88	96.6	90.8
wMel	88	97.5	99.6	88.4	86.4	87.4	88.3	84.3	86.6	88.1	88	97.1	97.4	87.8	100	99.6	87.7	84.3	87.7	87.8	99.4	97.3	97.7	97.3	88.2	88.9
wMelpop	88	97.4	100	88.3	86.4	87.4	88.2	84.3	86.6	88	88.1	97	97.2	87.7	99.4	100	87.7	84.3	87.7	87.8	98.9	97.2	97.4	97.3	88.2	88.9
wNo	95.3	88.5	87.8	96.0	84.8	86.1	95	83.8	85.6	96.7	94.6	87.9	87.7	94.7	87.8	87.8	100	83.8	94.7	94.9	87.8	88	88	88.1	95.5	90.9
wOo	83.7	84.6	84.3	84.5	84.6	85.6	83.9	99.7	85.2	84.8	83.9	84.2	84.2	83.8	84.3	84.3	83.8	100	83.8	83.8	84.2	84.2	84.2	84.3	83.8	84.5
wPip	96	88.3	87.9	98.8	85	86.2	97.3	83.8	85.5	95.6	95	88.1	87.8	99.2	87.8	87.9	95.3	83.8	100	98.9	87.9	87.8	88	87.9	96.7	90.8
wPipMol	96.1	88.3	87.9	98.7	85.1	86.3	97.4	83.8	85.5	95.6	95.1	88.1	87.8	98.4	88	87.9	95.2	83.8	98.8	100	88	87.9	88	88	96.7	90.9
wRec	87.9	97.5	99.1	88.4	86.4	87.4	87.9	84.2	86.5	87.9	87.9	97.2	97	87.7	99.3	99.1	87.7	84.2	87.8	87.7	100	97.3	97.7	97.3	88	89.2

wGm	31	70.8	67.1	31.1	29.3	31.1	24.6	54.5	28.3	44.5	69.5	30.9	71.2	71.7	100	30.6	24.6	31	31	72	65.1	27.8	74.5	66.1	31.7	31.7	70.4
m																											
wNo	58.9	30.7	31.7	61.2	27.7	57.6	23.7	23.9	27	63.3	30.8	55	30.9	30.8	30.6	100	23.8	55.2	56.6	30.6	31.9	25.6	31.1	31.7	61.1	37.5	30.8
wOo	23.8	24.5	25	24.1	25.8	23.9	96.1	56.7	25.6	26.1	24.6	23.9	24.6	24.6	24.6	23.8	100	23.9	23.9	24.5	24.6	24.6	25.2	24.6	24.2	24	24.8
wPip	61.9	30.4	31.1	81.2	27.7	78	23.8	58.2	27	78.1	30.6	86.3	30.5	30.6	31	55.2	23.9	100	84.4	30.8	30.7	25.8	31.3	30.7	68.6	37.1	30.8
wPip	63	30.7	31.3	88.6	27.6	81.5	23.8	58.2	26.9	77.8	30.9	67.6	30.8	30.9	31	56.6	23.9	84.4	100	30.8	31	25.8	31.3	31	70.8	37.6	31
Mol																											
wRec	30.8	90.6	73	31.2	29.4	30.8	24.6	31	28.4	45.7	75.2	30.8	89.3	95.3	72	30.6	24.5	30.8	30.8	100	70.9	27.6	75.4	71.2	31.7	32.4	94
wRi	31.3	70.8	98	31.5	29.6	31.5	24.7	30.9	28.7	45.6	69.3	30.6	71.9	72.1	65.1	31.9	24.6	30.7	31	70.9	100	27.8	73.7	98.6	32.9	32.1	72
wBm	25.8	27.7	28.3	26	69.8	25.6	24.6	25.6	28.8	28.6	27.6	25.8	27.6	27.7	27.8	25.6	24.6	25.8	25.8	27.6	27.8	100	28.3	27.8	26.1	25.9	27.9
wUni	31.9	75.6	73.5	31.7	30	31.5	25.1	31.6	29.1	46.2	73.6	31.3	75.8	76.2	74.5	31.1	25.2	31.3	31.3	75.4	73.7	28.3	100	73.8	32.2	34	75.4
wVals	31.2	71.2	97.9	31.4	29.7	31.4	24.7	30.9	28.7	45.5	69.5	30.6	72.1	72.3	66.1	31.7	24.6	30.7	31	71.2	98.6	27.8	73.8	100	32.3	32.3	71.9
ugan																											
a																											
wVitB	66	31.8	32.4	74.2	27.8	69.8	24.2	61.5	27.5	81.1	32.1	68.8	32.1	32.1	31.7	61.1	24.2	68.6	70.8	31.7	32.9	26.1	32.2	32.3	100	39.8	32.2
wVul	37.7	31.7	32.6	38.8	27.7	38.1	23.9	37.5	27.4	42.8	32.3	36.8	31.8	31.8	31.7	37.5	24	37.1	37.6	32.4	32.1	25.9	34	32.3	39.8	100	31.7
C																											
wWill	31.4	64.7	73.4	31.6	29.7	31.4	25.1	30.8	29	46.9	72.6	30.8	70.5	70.5	70.4	30.8	24.8	30.8	31	94	72	27.9	75.4	71.9	32.2	31.7	100

Tabela 8 – Matriz construída a partir dos valores de AF de isolados do gênero Wolbachia.

Genomas	wAlbB	wAna	wAu	wBa	wBm	wBol1-b	wCam	wCle	wDi	wGmm	wHa	wJHB	wMel	wMelPop	wNo	wOo	wPip	wPipMol	wRec	wRi	wValsunaga	wVitB	wVulC
wAlbB	1	0.55	0.7	0.81	0.69	0.74	0.62	0.66	0.81	0.61	0.69	0.63	0.7	0.74	0.74	0.62	0.65	0.72	0.75	0.61	0.64	0.79	0.59
wAna	0.55	1	0.64	0.54	0.51	0.53	0.46	0.54	0.58	0.52	0.63	0.45	0.63	0.66	0.54	0.46	0.47	0.51	0.65	0.64	0.65	0.54	0.45
wAu	0.7	0.64	1	0.68	0.64	0.77	0.57	0.68	0.73	0.65	0.88	0.76	0.94	0.92	0.77	0.57	0.77	0.75	0.85	0.88	0.88	0.68	0.76
wBa	0.81	0.54	0.68	1	0.79	0.74	0.85	0.65	0.8	0.73	0.66	0.65	0.67	0.71	0.72	0.85	0.67	0.73	0.74	0.58	0.61	0.85	0.58
wBm	0.69	0.51	0.64	0.79	1	0.57	0.75	0.65	0.65	0.71	0.62	0.5	0.63	0.67	0.58	0.76	0.51	0.56	0.72	0.55	0.58	0.74	0.49
wBol1-b	0.74	0.53	0.77	0.74	0.57	1	0.52	0.59	0.59	0.64	0.55	0.44	0.56	0.6	0.53	0.99	0.46	0.51	0.64	0.49	0.52	0.68	0.44
wCam	0.62	0.46	0.57	0.85	0.75	0.52	1	0.59	0.59	0.64	0.55	0.44	0.56	0.6	0.53	0.99	0.46	0.51	0.64	0.49	0.52	0.68	0.44
wCle	0.66	0.54	0.68	0.65	0.65	0.62	0.59	1	0.71	0.58	0.72	0.53	0.71	0.71	0.7	0.59	0.55	0.69	0.71	0.59	0.71	0.65	0.53
wDi	0.81	0.58	0.73	0.8	0.65	0.78	0.59	0.71	1	0.6	0.77	0.9	0.77	0.77	0.87	0.59	0.9	0.88	0.74	0.76	0.76	0.75	0.85
wGmm	0.61	0.52	0.65	0.73	0.71	0.53	0.64	0.58	0.6	1	0.62	0.46	0.63	0.66	0.54	0.76	0.47	0.53	0.7	0.55	0.57	0.67	0.45

wHa	0.69	0.63	0.88	0.66	0.62	0.72	0.55	0.72	0.77	0.62	1	0.74	0.86	0.85	0.74	0.56	0.75	0.74	0.79	0.86	0.86	0.66	0.75
wJHB	0.63	0.45	0.76	0.65	0.5	0.84	0.44	0.53	0.9	0.46	0.74	1	0.63	0.63	0.71	0.44	0.94	0.84	0.59	0.63	0.63	0.6	0.7
wMel	0.7	0.63	0.94	0.67	0.63	0.71	0.56	0.71	0.77	0.63	0.86	0.63	1	0.96	0.76	0.57	0.77	0.76	0.86	0.87	0.87	0.67	0.76
wMelPop	0.74	0.66	0.92	0.71	0.67	0.71	0.6	0.71	0.77	0.66	0.85	0.63	0.96	1	0.7	0.6	0.63	0.69	0.93	0.73	0.77	0.85	0.59
wNo	0.74	0.54	0.77	0.72	0.58	0.79	0.53	0.7	0.87	0.54	0.74	0.71	0.76	0.7	1	0.53	0.81	0.8	0.68	0.73	0.73	0.69	0.77
wOo	0.62	0.46	0.57	0.85	0.76	0.52	0.99	0.59	0.59	0.76	0.56	0.44	0.57	0.6	0.53	1	0.46	0.51	0.65	0.5	0.52	0.68	0.44
wPip	0.65	0.47	0.77	0.67	0.51	0.84	0.46	0.55	0.9	0.47	0.75	0.94	0.77	0.63	0.81	0.46	1	0.86	0.6	0.65	0.65	0.61	0.72
wPipMol	0.72	0.51	0.75	0.73	0.56	0.83	0.51	0.69	0.88	0.53	0.74	0.84	0.76	0.69	0.8	0.51	0.86	1	0.65	0.68	0.68	0.66	0.76
wRec	0.75	0.65	0.85	0.74	0.72	0.67	0.64	0.71	0.74	0.7	0.79	0.59	0.86	0.93	0.68	0.65	0.6	0.65	1	0.58	0.91	0.74	0.56
wRi	0.61	0.64	0.88	0.58	0.55	0.71	0.49	0.59	0.76	0.55	0.86	0.63	0.87	0.73	0.73	0.5	0.65	0.68	0.58	1	0.94	0.6	0.7
wValsunaga	0.64	0.65	0.88	0.61	0.58	0.71	0.52	0.71	0.76	0.57	0.86	0.63	0.87	0.77	0.73	0.52	0.65	0.68	0.91	0.94	1	0.63	0.59
wVitB	0.79	0.54	0.68	0.85	0.74	0.68	0.68	0.65	0.75	0.67	0.66	0.6	0.67	0.85	0.69	0.68	0.61	0.66	0.74	0.6	0.63	1	0.55
wVulC	0.59	0.45	0.76	0.58	0.49	0.77	0.44	0.53	0.85	0.45	0.75	0.7	0.76	0.59	0.77	0.44	0.72	0.76	0.56	0.7	0.59	0.55	1

4.1.3 - Descrição da espécie '*Candidatus Wolbachia vulgaris*'

Nós nos baseamos em estudos prévios para fazer a descrição fenotípicas do isolado wVulC, encontrada associada ao isópode *Armadillidium vulgare*. Células deste isolado são polimórficas com tamanhos variando entre 0.5-1 µm de diâmetro (121) sendo encontradas em gônadas, hemócitos e no sistema nervoso central do hospedeiro (Badawi et al. 2015) sendo especialmente abundantes no periplasma de oócitos, onde clusters de até 40 células de wVulC são encontradas (Rigaud et al. 1991). O genoma wVulC, um dos maiores genomas deste gênero com cerca de 1.65 Mpb, possui 3 profágos e a maior densidade de elementos de inserção já identificados neste gênero onde dois terços dos elementos IS identificados são expressos, não importando o seu nível de degradação genética (Cerveau et al. 2015). Recomendamos que isolados desta espécie possuam valores de ANI, rMLST e cgMLST acima de 96, 95 e 95%, respectivamente com wVulC.

4.1.4 - Grupos gênicos e composição funcional do pan-genoma

O pan-genoma do gênero *Wolbachia* consiste em 4,820 grupos gênicos isolado específicos e/ou genes únicos, dos quais 3,548 foram excluídos pois possuíam uma cobertura inferior ao limite de 70% com homólogos encontrados no banco de sequências SEED. A presença destas sequências curtas, as quais foram excluídas nesta etapa, refletem o processo de evolução redutiva que afeta diferentes isolados de *Wolbachia* (Andersson & Andersson 1999, Wernegreen 2002, Foster et al. 2005). Os 1272 grupos de genes ortólogos restantes foram anotados funcionalmente em todos os níveis de subsistemas funcionais do SEED (versão 70) (Tabela S1, Figura 15). O pan-genoma de isolados de *Wolbachia* é majoritariamente composto por grupos gênicos envolvidos com subsistemas de metabolismo primário, como “Metabolismo de proteínas” (95 genes), “Cofatores e vitaminas” (93 genes), e “Carboidratos” (39 genes), correspondendo a 7.5%, 7.3% e 3.0% da composição total do pan-genoma, respectivamente (Figura 13, Tabela S1). Em contraste, genes pertencentes aos subsistemas “Amino ácidos e derivativos” corresponde m a 0.94% do pan-genoma, estando de acordo com estudos que demonstraram que isolados do gênero *Wolbachia* perderam a maioria de suas vias biossintéticas para aminoácidos,

mantendo apenas vias para biossíntese de aminoácidos essenciais (Wu et al. 2004, Foster et al. 2005).

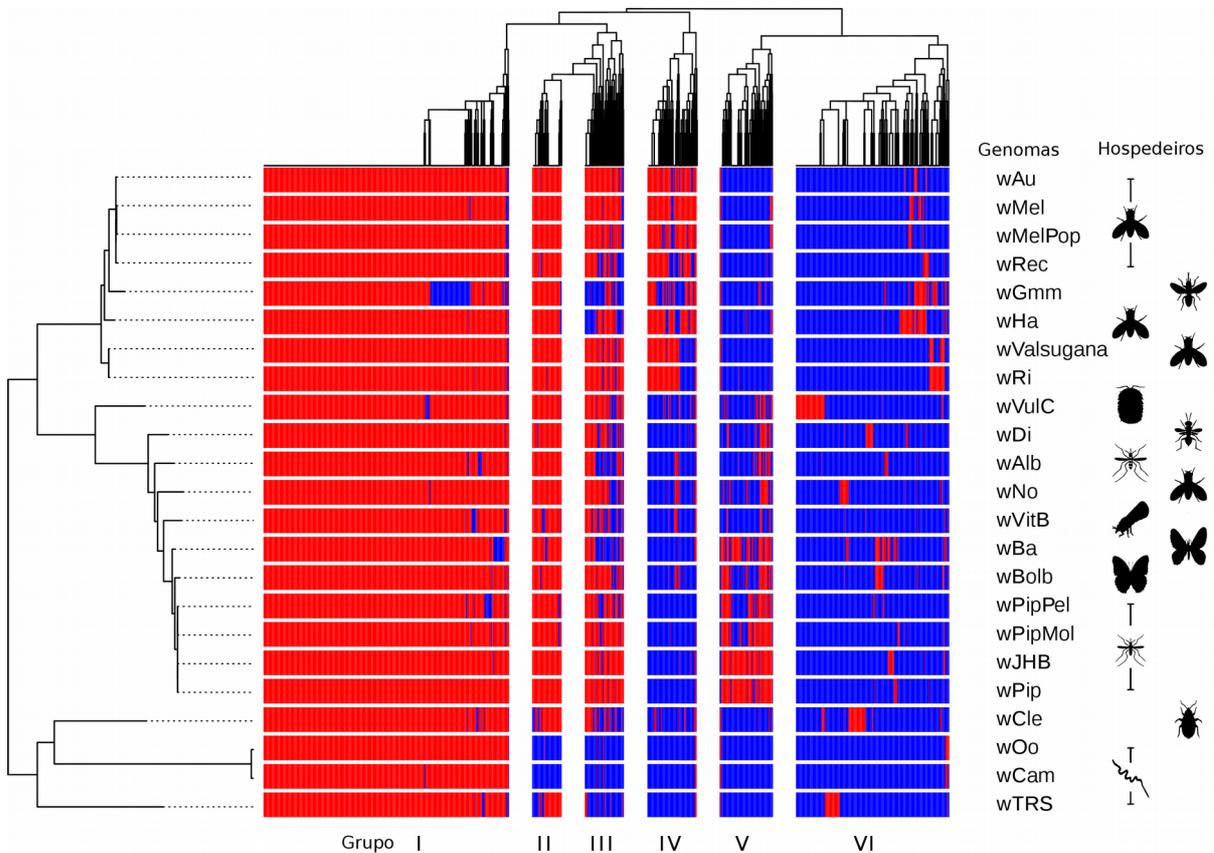


Figura 15 - Representação gráfica do pan-genoma do gênero *Wolbachia*. Genomas foram agrupados através do método de máxima verossimilhança, genes foram agrupados utilizando o método euclidiano padrão encontrado na linguagem R. Seis grupos gênicos (I-VI) foram definidos manualmente a partir do perfil de presença/ausência de genes onde vermelho é a presença de determinado gene e azul a ausência

Para que a comparação da composição funcional do genoma *pan* fosse possível, visualizamos e identificamos manualmente seis grupos gênicos no *heatmap* gerado a partir da matriz de presença/ausência (I-VI, Figura 16), utilizamos a anotação dos genes dentro destes grupos para calcular o enriquecimento/empobrecimento funcional (métodos). Estes grupos gênicos compreendem o genoma *core* e o genoma *core* flexível (Grupo gênico I, 549 genes sendo que 368 são encontrados em todas as bactérias), o grupo gênico II (66 genes), e genes acessórios (grupos gênicos III-VI).

O grupo gênico I contém o genoma *core* estrito (368 genes) e um adicional de 181 genes compartilhados por >90% dos isolados (core-genoma flexível). A análise revelou que este grupo é significativamente enriquecido para funções metabólicas como “Nucleosídeos e nucleotídeos”, em especial para enzimas envolvidas na biossíntese de nucleotídeos e reciclagem de monofosfatos. Outros subsistemas funcionais enriquecidos foram o “Metabolismo de proteínas”, incluindo tRNA sintases e proteínas envolvidas com biogênese de lipoproteínas e o subsistema “Respiração” (enriquecimento > 1, pvalue < 0.001) (Tabela S1, Figura 14). O subsistema “Fagos, profagos, elementos transponíveis e plasmídeos” estava significativamente empobrecido (enriquecimento <0, valor p < 0.01), de fato, genes identificados como pertencentes a este subsistema pertenciam ao genoma *core* flexível, não sendo compartilhados por todas as bactérias analisadas.

Os grupos gênicos II (66 genes) e III (87 genes), estão presentes quase que inteiramente em todos os genomas de *Wolbachia*, as exceções são os genomas de endossimbiontes obrigatórias, como *wOo* e *wCam*. Sequências do grupo gênico II não são encontrados nos genomas *wOo* e *wCam* (infectam nematoides) e genes do grupo III não são encontrados em sua totalidade nestes e nos genomas *wBm* (infecta nematoides) e *wCle* (infecta o inseto *C. Lectularius*). O único subsistema significativamente enriquecida no grupo gênico II é o subsistema “Metabolismo de DNA” (enriquecimento >1, valor p <0.001), enquanto o grupo gênico III é significativamente enriquecido para “Divisão e ciclo celular” (enriquecimento > 1, valor p < 0.001) e “Sinalização e regulação celular” (enriquecimento > 1, valor p < 0.001) (Figura 14).

Os grupos gênicos IV e V contém genes que são encontrados apenas em endossimbiontes facultativos. Estes grupos contém diversos genes envolvidos, mas pouco enriquecidos, com os subsistemas “Fagos, profagos, elementos transponíveis e plasmídeos” (enriquecimento >0.2 e > 0.5, valor p <0.001 e <0.018, respectivamente) (Figura 14), o último grupo gênico é empobrecido para praticamente todos os subsistemas funcionais.

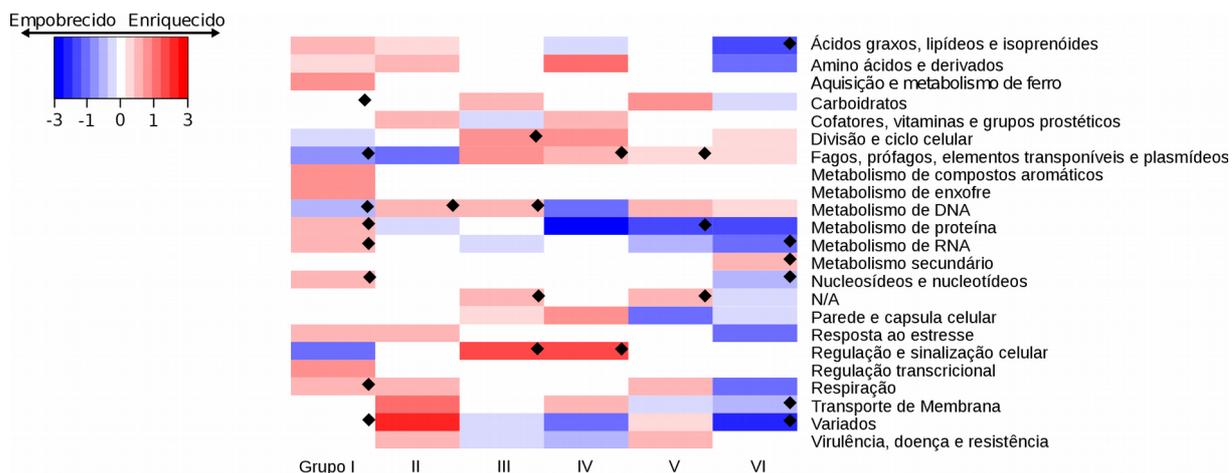


Figura 16 - Composição funcional dos grupos gênicos (I-VI) do pan-genoma, baseada nos subsistemas de nível 1 do banco de sequências SEED. Funções enriquecidas estão marcadas em tons de vermelho, funções empobrecidas em tons de azul. Funções significativas estão marcadas em tons de cinza.

Aplicamos esta metodologia de reamostragem mas focada em genomas, ao invés do pan-genoma de forma geral, para que assim pudéssemos investigar a diversidade funcional do gênero *Wolbachia* individualmente (Figura 17). Como mostrado pela figura 17, isolados do gênero *Wolbachia* possuem perfis funcionais similares.

Os subsistemas funcionais formaram 4 diferentes grupos (Figura 17) onde é possível se ver que essencialmente, isolados deste gênero possuem um potencial metabólico similar. O primeiro grupo envolve funções de metabolismo primário como “Aquisição e metabolismo de ferro”, “Metabolismo de proteínas”, “Metabolismo de RNA” e “Respiração” sendo a maioria deles altamente enriquecidos (Enriquecimento > 0.5) e com alto valor de suporte ($pvalue < 0.001$). O papel de isolados deste endossimbionte na modulação do metabolismo de ferro do hospedeiro (Kremer et al. 2009), e também no provisionamento de moléculas de ATP, já foram documentados (Darby et al. 2012), dando suporte ao papel de tais subsistemas na interação hospedeiro-simbionte.

O segundo grupo é composto pelo subsistema “Fagos, prófagos, elementos transponíveis e plasmídeos” sendo altamente enriquecidos na maioria dos endossimbiontes facultativos, como as bactérias wPip, wJHB e wVulC, mas praticamente ausente em isolados associadas a nematoides (Gavotte et al. 2007).

O terceiro grupo é composto por subsistemas pouco enriquecidos ou empobrecidos, como “Amino ácidos e derivados”, este subsistema comporta genes que estão envolvidos com a biossíntese de amino ácidos (Wu et al. 2004, Foster et al. 2005), foi demonstrado também que alguns isolados, na verdade, competem com o hospedeiro pela aquisição de amino ácidos não essenciais (Caragata et al. 2014). O último grupo é composto pelo subsistema “Regulação e sinalização celular”, enriquecidos apenas nos genomas dos isolados wMel, wMelPop, wAu e wCle, este subsistema funcional foi enriquecido devido a um grande número de cópias de genes do sistema toxina antitoxina tipo II, 6 cópias (Tabela S1) nestes genomas, este número varia entre 0 e 3 em outros genomas deste gênero.

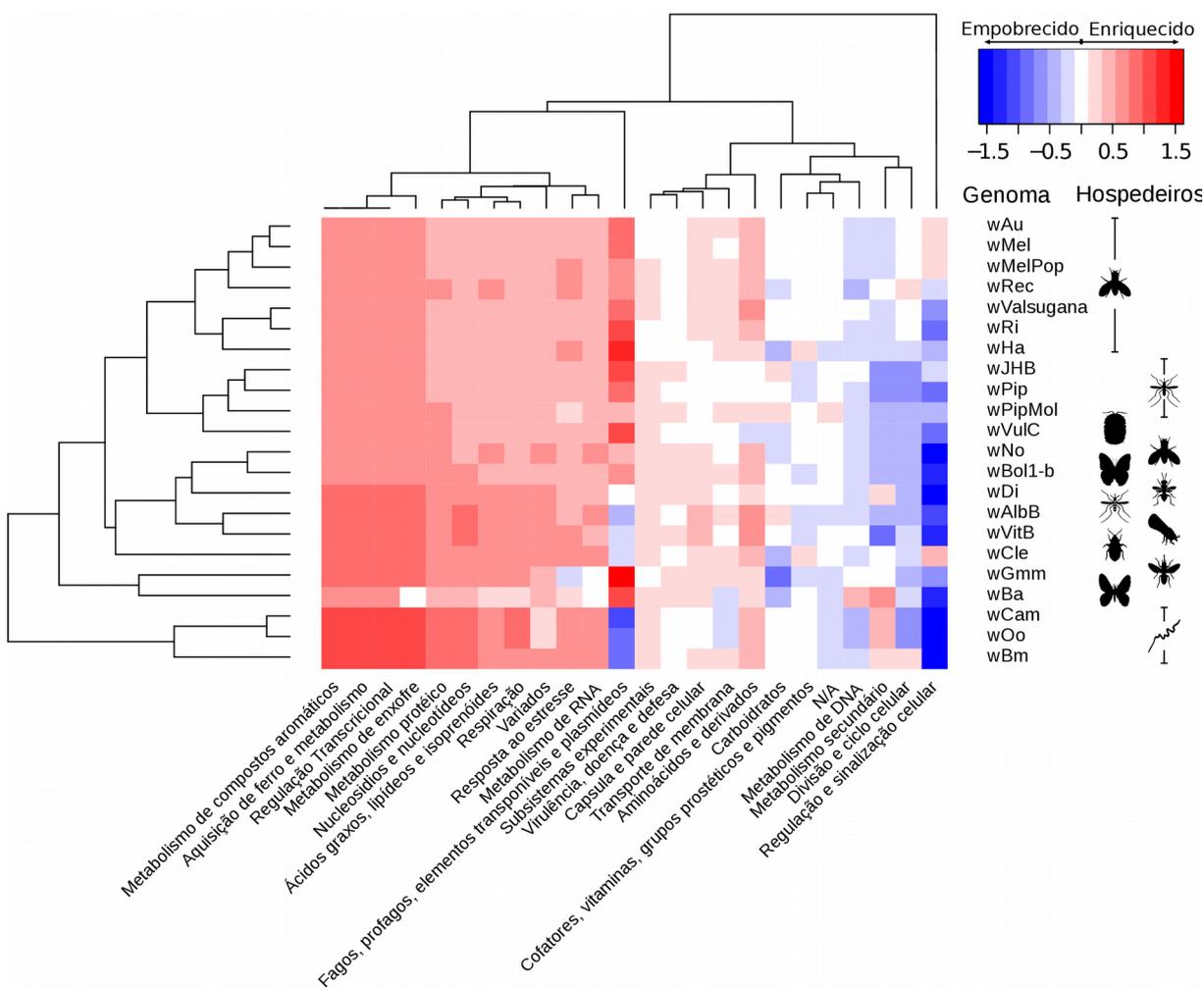


Figura 17 - Composição funcional genomas de isolados do gênero *Wolbachia* baseada nos subsistemas de nível 1 do banco de sequências SEED. É possível observar que genomas de endossimbiontes facultativos possuem as funções “Regulação e sinalização celular” e “Fagos, profagos e plasmídeos” altamente enriquecidos, enquanto este é empobrecido em genomas de endossimbiontes obrigatórios.

Estes resultados poderiam indicar as funções que isolados de *Wolbachia* obrigatórias perderam ou estão perdendo no curso de sua evolução reductiva. Ao alinhar os genomas fragmentados de isolados do gênero *Wolbachia* utilizando o software BRIG (Alikhan et al. 2011), é possível observar as regiões perdidas por endossimbiontes obrigatórios ao compará-los com o isolado tipo, wPip^t (Figura 18).

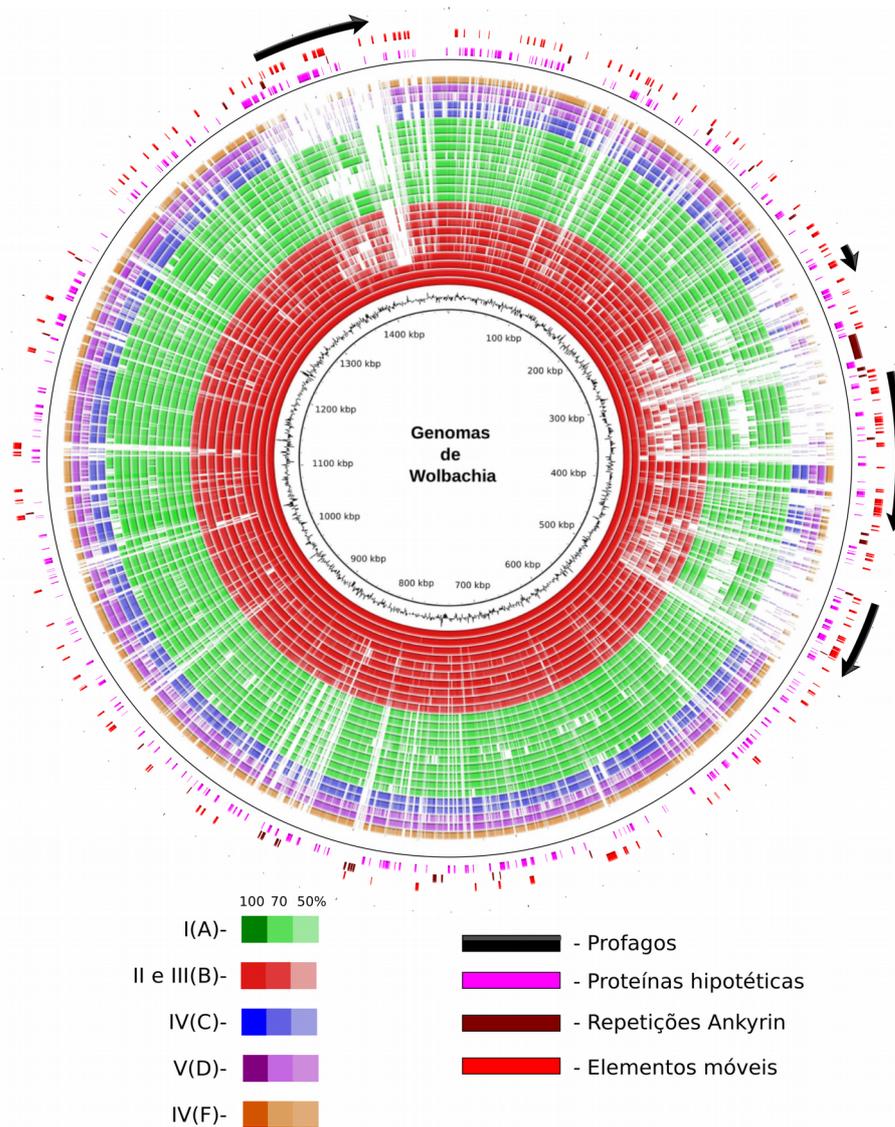


Figura 18 – Representação gráfica circular da diversidade dos genomas do gênero *Wolbachia* onde cada genoma foi alinhado com o isolado tipo wPip^t, representado pelo primeiro círculo vermelho. Supergrupos estão representados por diferentes cores, os cinco círculos externos são genomas de endossimbiontes obrigatórios e é possível se observar grandes regiões ausentes nestos isolados. Os três círculos finais representam proteínas hipotéticas, sequências repetitivas Ankyrin e elementos móveis, respectivamente. O último círculo representado por setas representam profagos identificados pelo software PHAST.

Nesta figura, o primeiro círculo representa o genoma do isolado tipo, wPip^t, os demais os genomas de isolados de Wolbachia são representados pelos círculos posteriores a este, agrupados por cores que representam a que supergrupo/espécie eles pertencem. Os quatro últimos círculos são genes hipotéticos, sequências repetitivas Ankyrin, genes que pertencem ao subsistema “Fagos, profagos, plasmídeos e elementos transponíveis” e fagos identificados pelo software PHAST, respectivamente. É possível se observar que as regiões instáveis em seus genomas contém elementos genéticos móveis e mobilizáveis (Transpósons e profagos), estando em sua grande maioria ausentes em endossimbiontes obrigatórios, corroborando os resultados de nossa análise funcional, onde a diferença funcional mais importante entre endossimbiontes facultativos e obrigatórios é a presença de elementos genéticos móveis e fagos.

Identificamos também que grande parte das regiões gênicas perdidas em endossimbiontes obrigatórios contém genes que codificam proteínas hipotéticas que não possuíam nenhum homólogo e pseudogenes, cujos tamanhos variavam entre 98 e 3524 pares de base.

4.1.6 - Identificação de genes adquiridos por HGT

Utilizamos o software FACIL (Dutilh et al. 2011) para prever a preferência pelo uso de códons em todos os genomas de Wolbachia analisados. Este software encontrou suporte para o uso do código genético padrão para bactérias em quase todos os genomas (Figura 19), a exceção notável foi o isolado wPipMol, onde o códon de terminação Opal (TGA) foi predito para codificar triptofano, utilizando, segundo o resultado, um código genético similar ao de uma mitocôndria (Jukes & Osawa 1990). Buscas por similaridade foram feitas utilizando wPipMol contra genomas completos de mitocôndrias disponíveis no banco de sequências do NCBI, identificando assim quatro regiões que seriam responsáveis por esta predição de mudança de código genético.

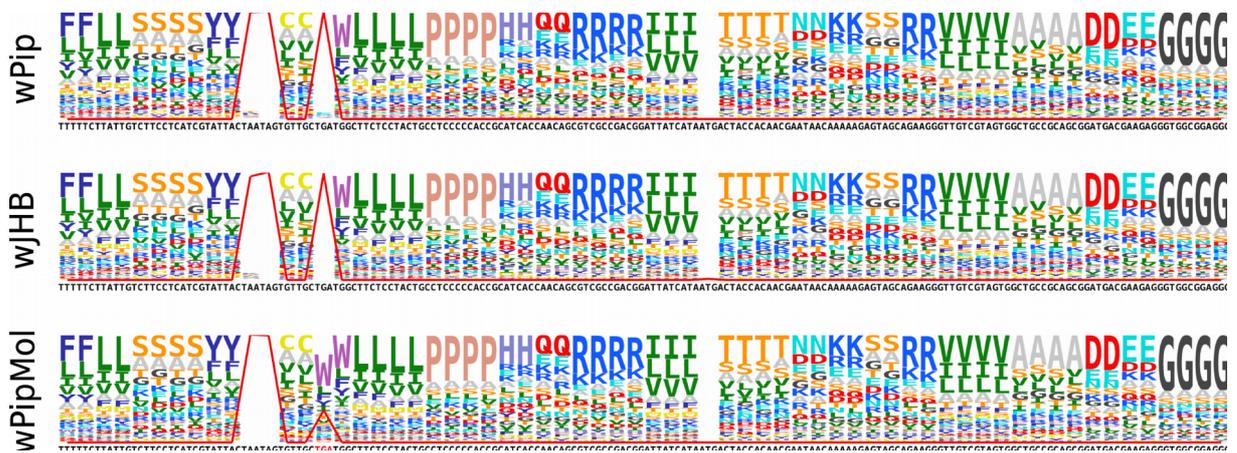


Figura 19 - Imagem gerada pelo software FACIL, onde códons possíveis e os aminoácidos que estes estariam codificando. Códons de terminação e de início estão marcados em branco, a diferença entre estes três genomas reside no genoma wPipMol, onde o códon de terminação OPAL (TGA) foi identificado pelo software como codificante para o aminoácido triptofano.

Identificamos 6 genes nestas regiões, estes codificariam componentes da cadeia de transporte eletrônica encontrados em mitocôndrias e procariotos, como citocromos B, C e subunidades de NadH D, H e M. Buscas por homologia foram feitas contra genomas de mitocôndrias no banco de sequências de genomas completos do NCBI, os melhores hits foram recuperados, alinhados com as sequências de mtDNA do genoma do isolado wPipMol e utilizadas na construção de árvores de máxima verossimilhança. As árvores resultantes revelaram que cinco destas sequências de origem mitocondrial são homólogas a genes de mitocôndrias encontradas em nematoides (Figuras 20-23), tendo como *Best hit* a mitocôndria do nematoide *Teladorsagia circumcincta*, um parasita que infecta ovelhas (Tzelos et al. 2016).

Identificamos também uma sequência, que codificaria a proteína NadH subunidade M (Figura 24), sendo homóloga a genes de mitocôndrias encontradas em mosquitos, mas não mosquitos do complexo de espécies *Culex pipiens*. Este complexo de espécies de mosquitos é conhecido por serem vetores de transmissão de muitos patógenos como, nematoides que causam elefantíase (Simonsen & Mwakitalu 2013) e diversos vírus (Turell 2012) e, devido a isto, tais genes poderiam ser contaminantes. Entretanto, o genoma wPipMol foi purificado a partir de embriões de *C. pipiens molestus*, sugerindo que estes genes não são contaminantes, mas sim adquiridos por HGT após a divergência dos isolados wPipMol, wPip e wJHB.

Os pseudogenes de origem mitocondrial encontrados no genoma da *Wolbachia* wPipMol estão localizados em um prófago (Figura 25) semelhante ao fago WO, encontrado comumente em isolados deste gênero (Ishmael et al. 2009), sugerindo que estes genes foram adquiridos através de um evento de transdução. Uma hipótese alternativa é de que DNA foi adquirido pela bactéria e integrado em seu DNA, onde foi integrado ao prófago, entretanto, esta hipótese não explica qual mecanismo foi capaz de permitir a transferência de genes da mitocôndria de um nematoide para a *Wolbachia* de um inseto.

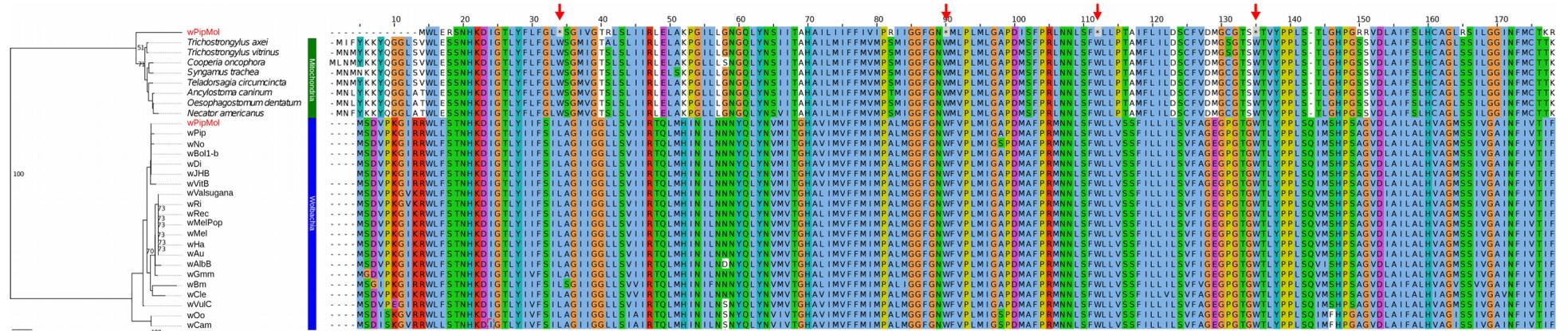


Figura 20 - Árvore de máxima verossimilhança e o alinhamento da proteína Citocromo oxidase C originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap.

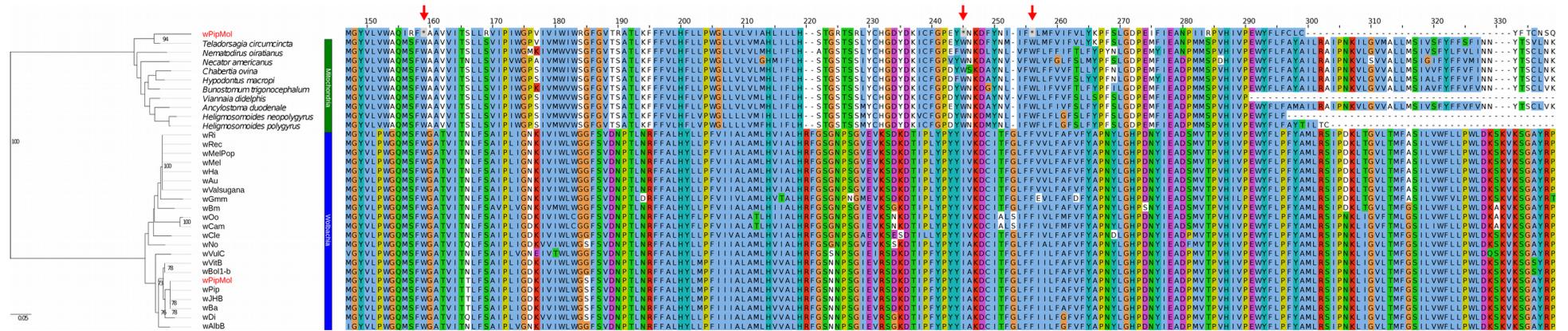


Figura 21 - Árvore de máxima verossimilhança e o alinhamento da proteína Citocromo oxidase B originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap.

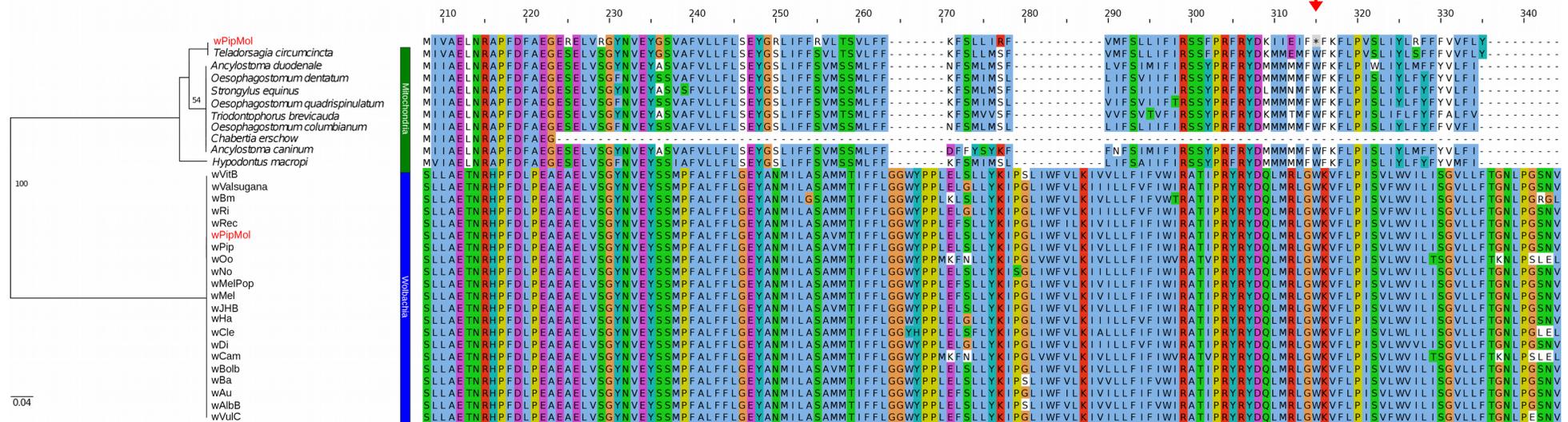


Figura 22 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade H originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap.

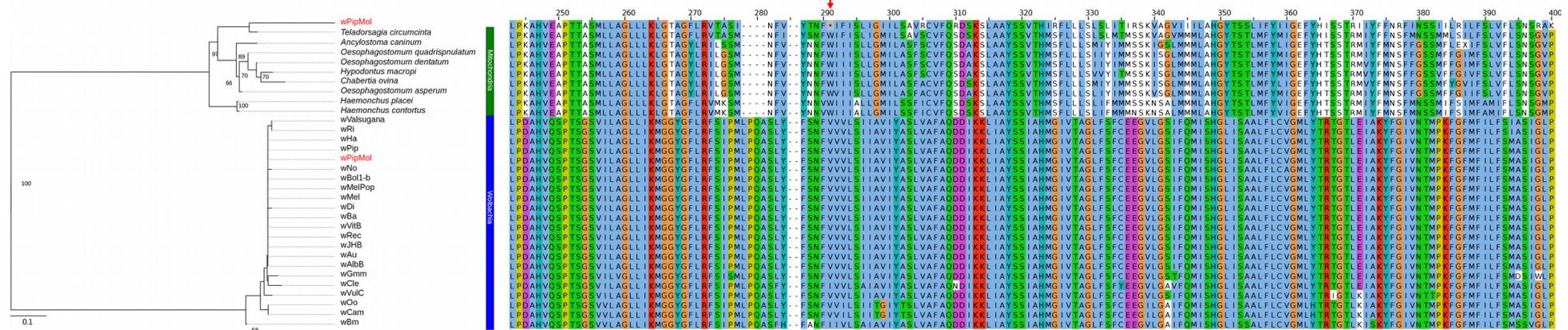


Figura 23 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade D originado de uma mitocôndria de uma espécie de nematoide, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap.

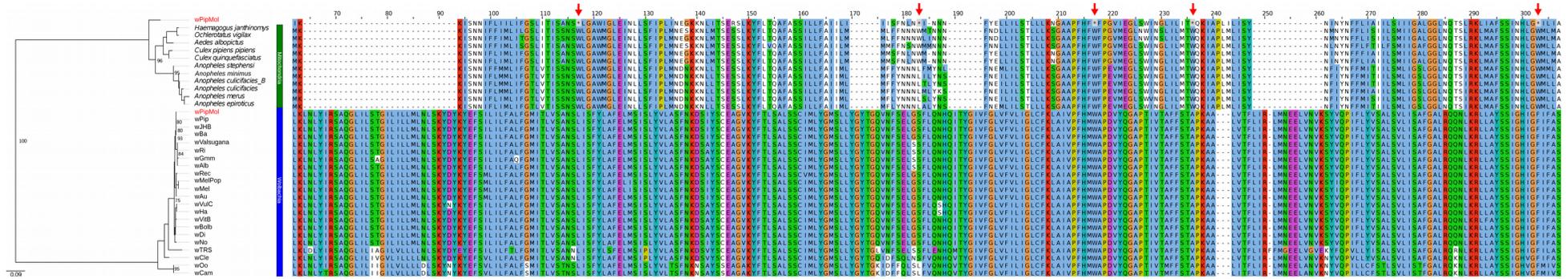


Figura 24 - Árvore de máxima verossimilhança e o alinhamento da proteína NadH subunidade N originado de uma mitocôndria de uma espécie de mosquito, supostos códons de terminação estão marcados com uma seta vermelha. A árvore foi construída utilizando o modelo WAG, suporte nos nós foram calculados utilizando 1000 replicatas de bootstrap.

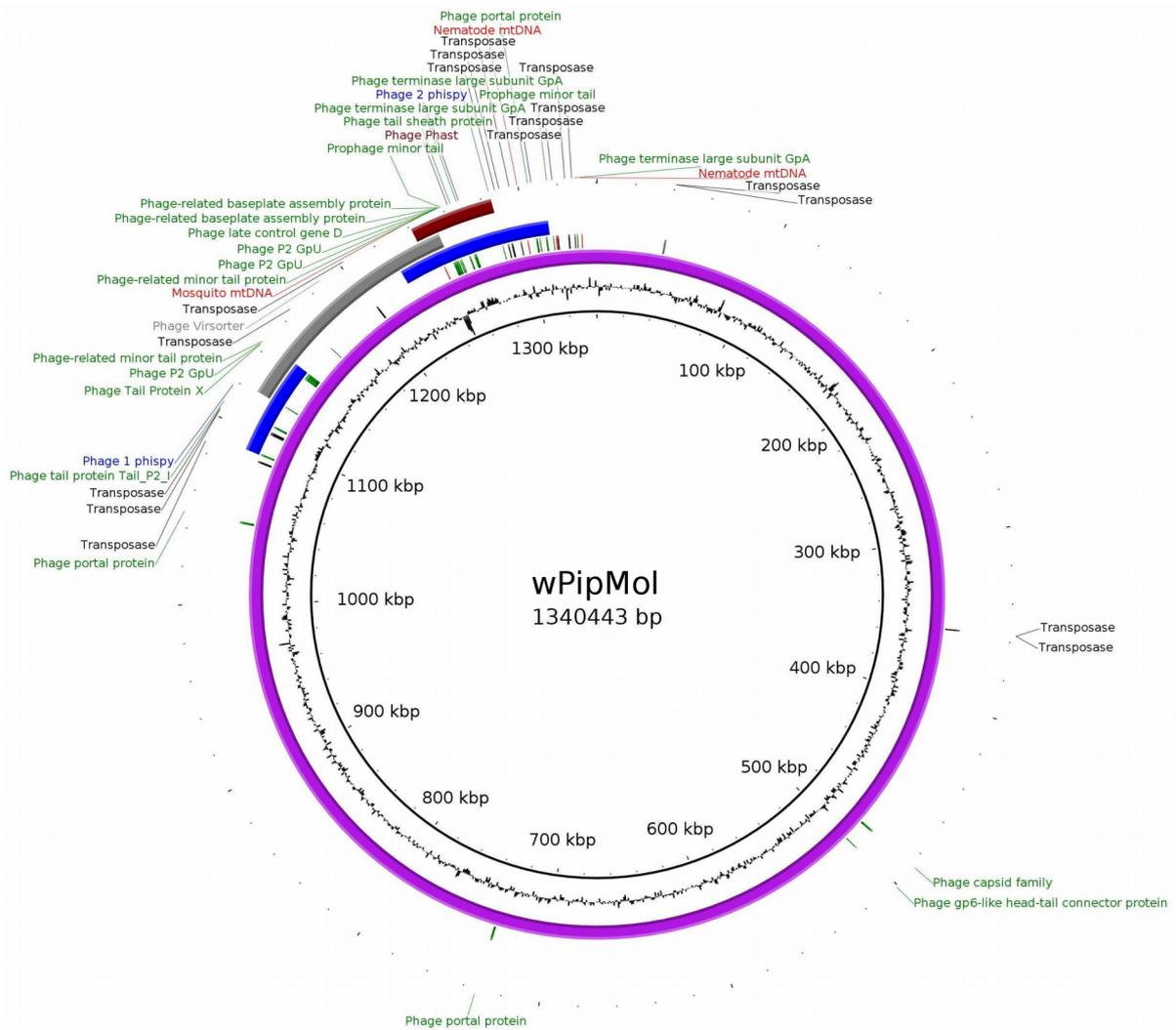


Figura 25 - Imagem gerada pelo Blast ring image generator (BRIG) a partir do genoma wPipMol. Genes provenientes de fagos estão marcados em verde, predições de fagos feitos pelos softwares Phast (Vermelho), Virsorter (Cinza) e Phispy (azul) estão marcadas, indicando que a aquisição destes insertos pode ter sido devido a um evento de transdução.

4.2 - Klebsiella

Considerando a dificuldade em se identificar corretamente espécies de *Klebsiella* em função da complexidade taxonômica deste gênero, este estudo aplicou diversas métricas de taxonomia genômica considerando 764 genomas disponíveis em bancos públicos de sequências das espécies *K. pneumoniae*, *K. quasipneumoniae* e *K. variicola* com o objetivo de fornecer informações mais robustas acerca deste gênero.

Como no início deste projeto de doutorado havia somente um genoma anotado como *K. variicola* (At-22), que correspondia a um isolado ambiental, e nenhum de *K. quasipneumoniae*, as bactérias Bz19 e K142 foram sequenciados e incluídos nas análises de taxonomia genômica e filogenômica do gênero *Klebsiella*. Este estudo gerou o primeiro genoma de um isolado clínico de *K. variicola* no Brasil, o isolado Bz19 (GCA_00074291)(Andrade et al. 2014).

4.2.1 – Sequenciamento e identificação dos genomas Bz19 e K142

O sequenciamento do genoma de *K. variicola* Bz19 gerou 70.5 Mbp de informação genômica distribuídos em cerca de 176 mil *reads*, com cerca de 400 pares de base por *read* e uma cobertura de cerca de 13x. Os *reads* foram submetidos a uma montagem *de novo*, gerando 208 contigs, totalizando 5,48 Mbp com um conteúdo GC de 57.3%. A predição gênica identificou a existência de 5392 CDS, incluindo genes codificantes e não codificantes.

O sequenciamento do isolado K142 gerou 6.7 Gbp de informação genômica distribuídos em 4 milhões de *reads*, com cerca de 150 pares de base por *read* e uma cobertura de cerca de 1100x. A montagem gerou 75 sequências consenso (contigs e scaffolds), totalizando 5.48 Mbp com 57.8% de conteúdo GC. Predições gênicas e anotação deste genoma identificaram 5160 CDS.

4.2.2 - Filogenômica

Foram identificados 2124 genes compartilhados entre 89 genomas pertencentes às espécies *K.pneumoniae*, *K. quasipneumoniae*, *K. variicola* e *K. oxytoca* (core-genoma). Estes genes foram alinhados e concatenados gerando um alinhamento de 2.302.713 pb. Uma análise de máxima verossimilhança utilizando o arquivo original seria computacionalmente intenso, extraímos 370.502 SNPs do super alinhamento do genoma *core* para as análises subsequentes. A construção da árvore de máxima verossimilhança levou aproximadamente 400 horas/núcleo em um computador de 8 núcleos com 32gb de ram. Esta análise resultou em uma árvore com topologia bem definida e altos valores de suporte nos nós internos, onde 4 grandes grupos foram

definidos (Figura 26). Estes grupos correspondem às quatro espécies *K. pneumoniae*, *K. quasipneumoniae*, *K. variicola* e *K. oxytoca*. Esta análise revelou que 37 genomas de isolados classificados como *K. pneumoniae* e 4 de *Klebsiella sp.* correspondiam a outras espécies como *K. quasipneumoniae* (14 genomas) e *K. variicola* (23 genomas). Além disto, nossos resultados obtidos com a filogenômica corroboraram um estudo prévio que demonstrou através da abordagem de MLST que o genoma 10982 diverge de outras espécies do gênero *Klebsiella*, já que assume uma posição intermediária entre as espécies *K. pneumoniae* e *K. variicola* (Hazen et al. 2014).

Identificamos 41 rps compartilhados por todas as bactérias do gênero *Klebsiella* analisados. A concatenação do alinhamento destes genes gerou uma sequência com um tamanho de 19.130 pb e com 758 snps, uma razão de 1 snp a cada 121 pb contrastando com o genoma *core*, onde encontramos 1 snp a cada 6 pb. O fato das sequências codificadoras serem conservadas neste gênero influenciou a topologia da árvore filogenômica (Figura 27), devido a estes resultados, nós não os utilizamos nas análises taxonomia genômicas.

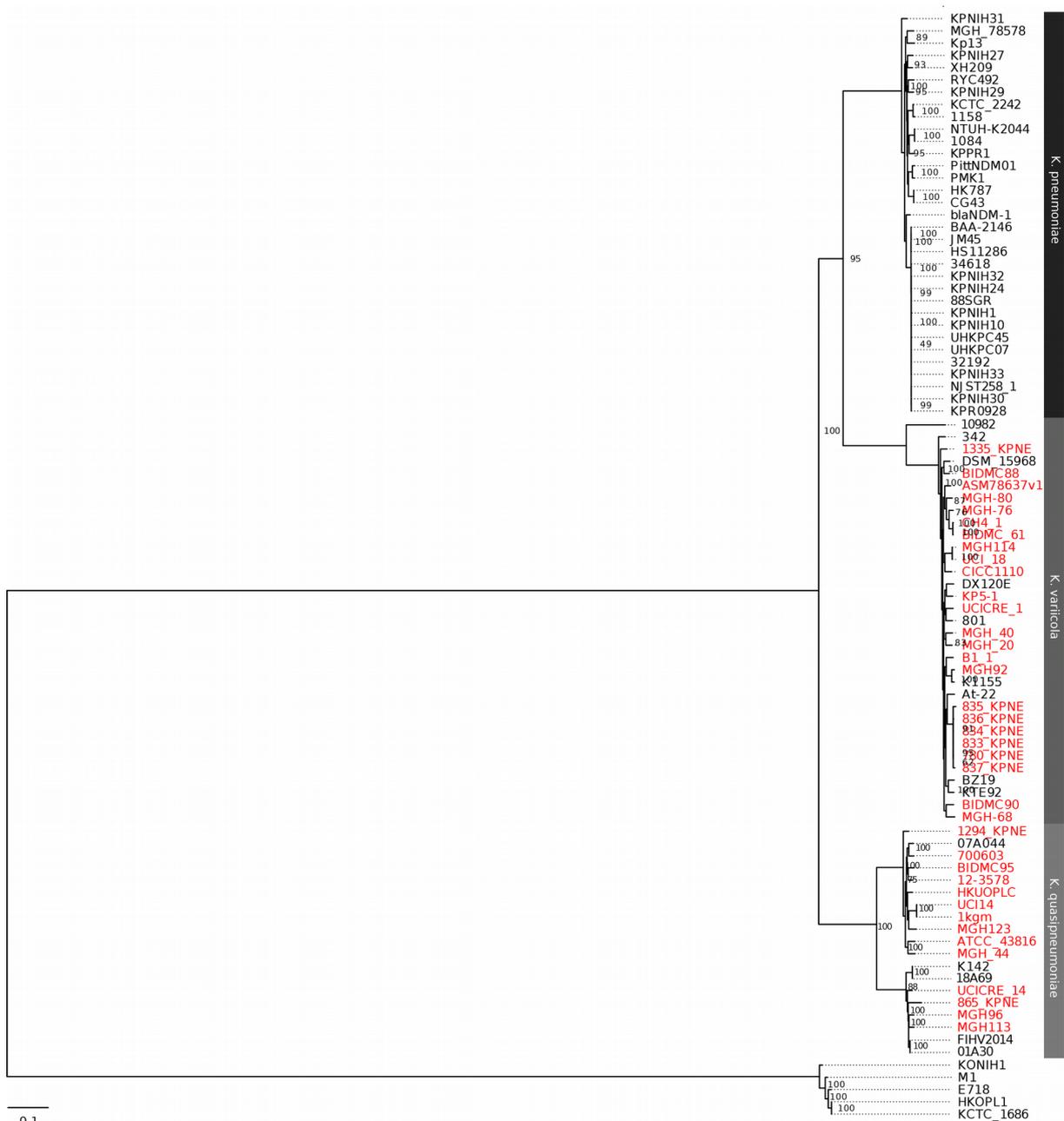


Figura 26 – Árvore filogenômica construída a partir do superalinhamento dos genes do genoma core do gênero *Klebsiella* (cgMLST). A árvore foi construída utilizando o método de máxima verossimilhança com o modelo GTR, o suporte nos nós foi obtido através de 1000 replicatas de bootstrap. A árvore nos mostra três grupos bem definidos representando as espécies *K. pneumoniae*, *quasipneumoniae* e *variicola*. Isolados de *K. oxytoca* foram utilizados como grupo externo. Identificadores marcados em vermelho são referentes a isolados previamente classificados como *K. pneumoniae* mas que agruparam com linhagens tipo de espécies diferentes.

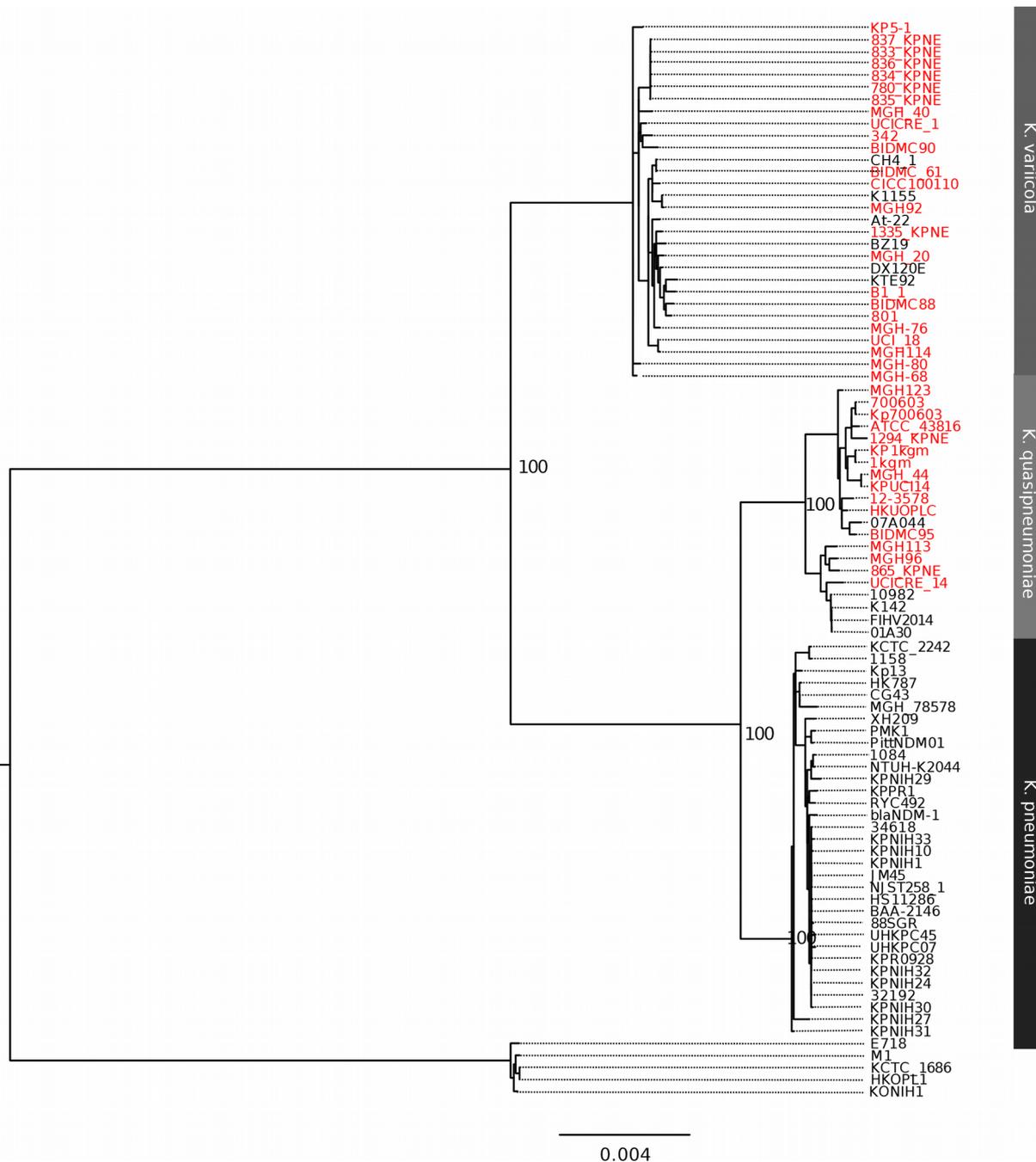
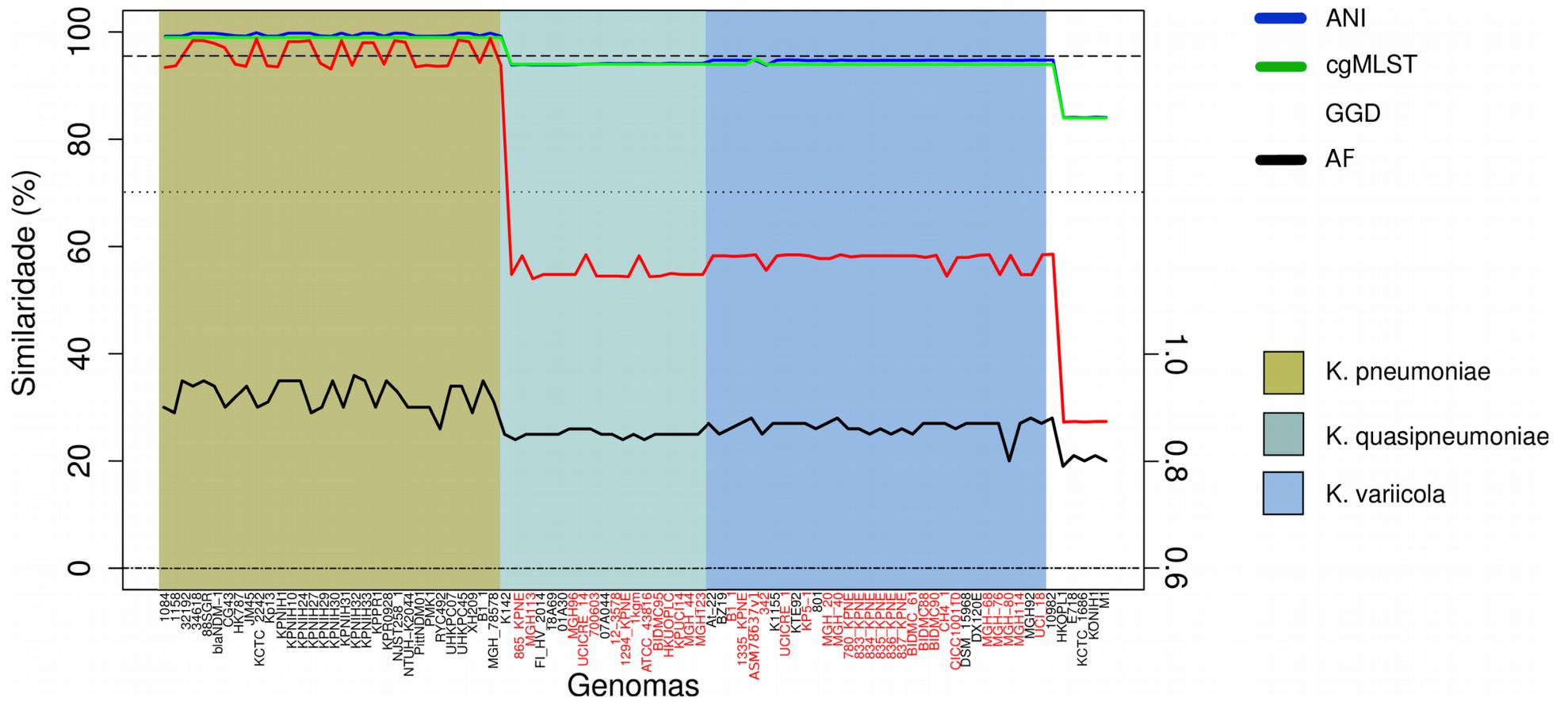


Figura 27 – Árvore filogenômica construída a partir do superalinhamento dos codificadores de rps (rMLST). A árvore foi construída utilizando o modelo GTR, o suporte nos nós foi obtido através de 1000 replicatas de bootstrap. Identificadores marcados em vermelho são referentes a isolados previamente classificados como *K. pneumoniae* mas que agruparam com linhagens tipo de espécies diferentes.

4.2.3 – Taxonomia genômica do gênero *Klebsiella*

Isolados das espécies *K. quasipneumoniae* e *K. variicola* são genética e metabolicamente similares a espécie *K. pneumoniae* e, devido a esta proximidade, a identificação por testes bioquímicos não seria o suficiente para discriminá-las. Este fato, aliado a posição periférica que o isolado 10982 apresentou na árvore filogenômica, nos motivou a revisitar o estado taxonômico deste grupo de espécies e delinearlos utilizando abordagens acuradas e complementares de taxonomia genômica, como GGDP, ANI e AF. Tais abordagens foram aplicadas a todos os pares de genomas de *Klebsiella* de diferentes espécies, segundo a metodologia descrita previamente (Métodos), onde os genomas de *K. pneumoniae* (HS11286^T), *K. quasipneumoniae* (01A030^T) e *K. variicola* (DSM15968^T) foram utilizados como modelos para delimitação dos genomas analisados. Dos 89 genomas, 32 apresentaram valores >99 e >90 de ANI e GGDP, respectivamente, ambos acima dos limites sugeridos previamente com a linhagem tipo de *K. pneumoniae* HS11286^T (Meier-Kolthoff et al. 2013, Varghese et al. 2015),. Estes resultados, juntamente com a alta similaridade do super alinhamento do genoma core (>95%) e um valor de AF > 0.95 (Tabela 7, Figuras 28, 29 e 30), indicam que estes genomas, assim como os 671 não utilizados devido a sua redundância, são de fato *K. pneumoniae*.



Dois genomas de isolados previamente delineados como *K. quasipneumoniae* (18A69 e FI HV 2014), quatro genomas classificados como *K. pneumoniae* (MGH113, MGH96, 865_KPNE e UCICRE_14) e o genoma da linhagem K142 sequenciado por nosso grupo apresentaram valores acima dos limites de definição de espécie considerando o genoma tipo de *K. quasipneumoniae* subsp. *quasipneumoniae* (01A030^T) para todas as abordagens taxonomia genômicas aqui utilizadas. Esta subespécie corresponde ao que antes era conhecido como grupo filogenético KP11-A. Outros 11 genomas possuíam valores acima dos limites de definição considerando o genoma 07A044^T, linhagem tipo da *K. quasipneumoniae* subsp. *similipneumoniae*. Isolados destas duas subespécies possuem valores limítrofes para as abordagens taxonomia genômicas utilizadas (Tabela S3), indicando que estes organismos estão no limite de especiação (Figura 29).

Em relação a *K. variicola*, 12 genomas classificados como pertencentes a esta espécie, 23 genomas previamente classificados como *K. pneumoniae* e o genoma Bz19 sequenciado por nosso grupo possuíam valores acima dos limites estabelecidos para definição de espécies considerando o genoma tipo DSM 15968^T (Figura 30).

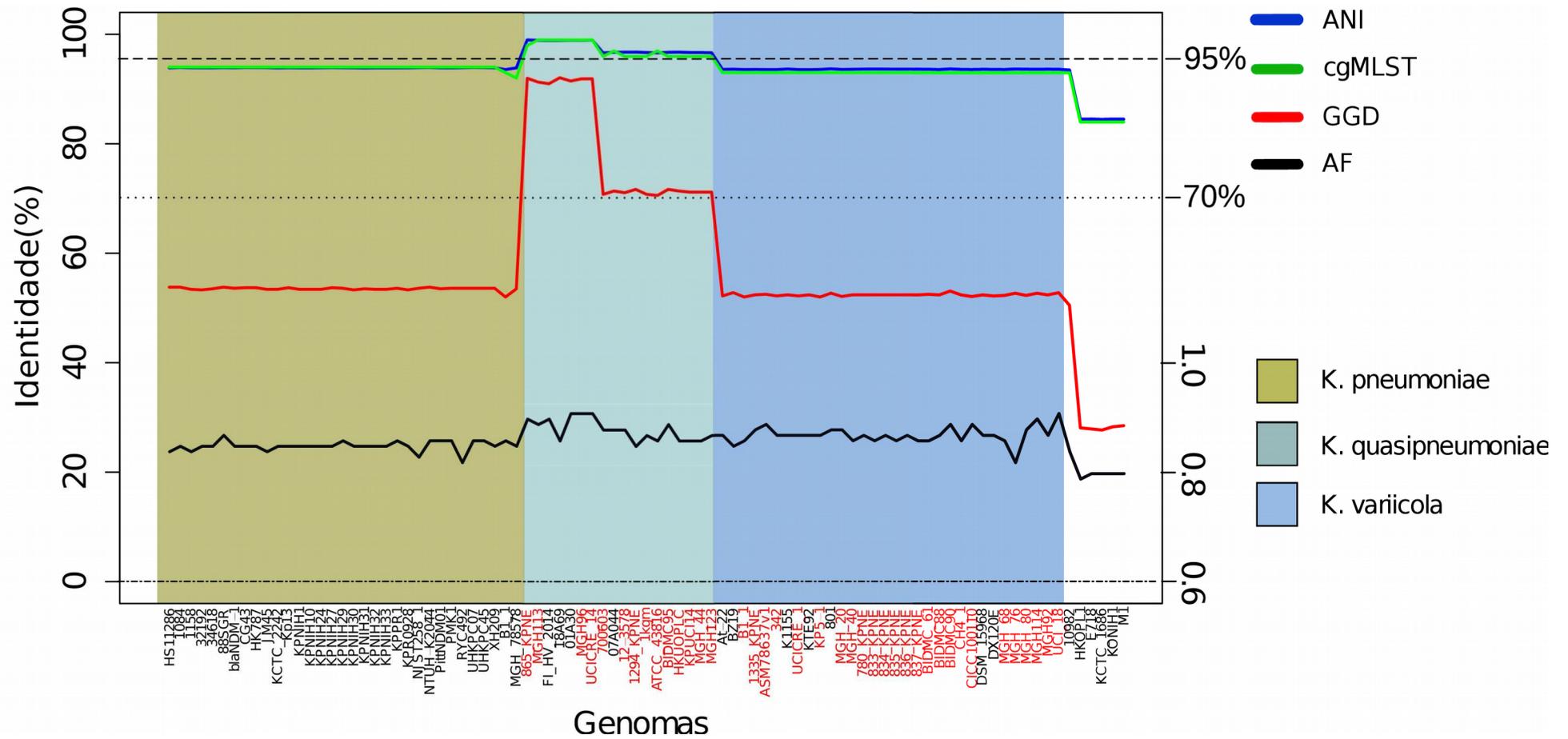


Figura 29 – Resolução taxonômica da espécie *K. pneumoniae*, este gráfico comparativo foi construído a partir de matrizes de similaridade de abordagens como ANI, cgMLST, GGD e AF de todos os isolados de *Klebsiella* analisados contra o genoma da linhagem tipo de *K. quasipneumoniae*, 01A030¹.

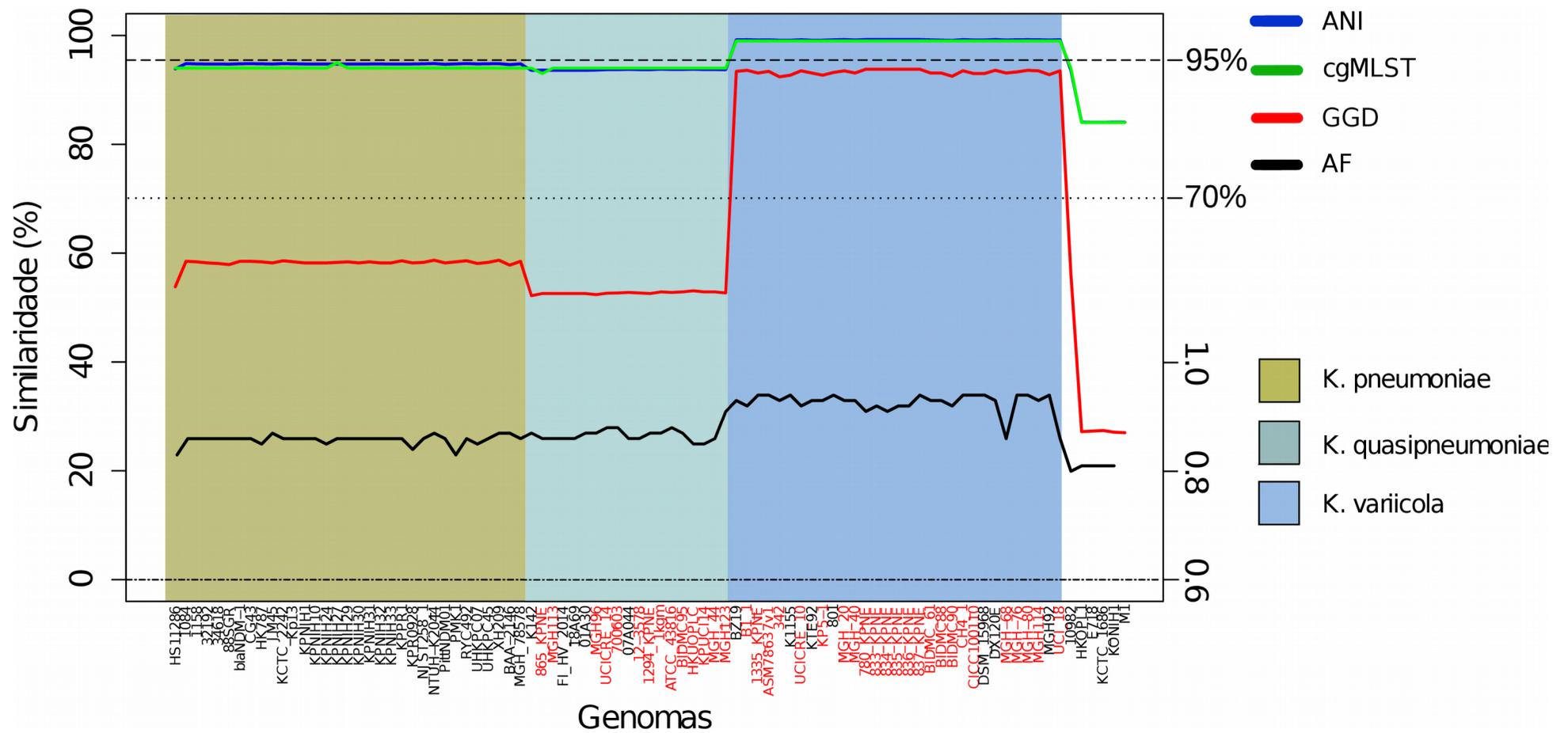


Figura 30 – Resolução taxonômica da espécie *K. pneumoniae*, este gráfico comparativo foi construído a partir de matrizes de similaridade de abordagens como ANI, cgMLST, GGDC e AF de todos os isolados de Klebsiella analisados contra o genoma da linhagem tipo de *K. variicola*, DSM18961[†].

A abordagem de alinhamento de fração (AF) não foi resolutive o suficiente para delinear espécies do gênero *Klebsiella*. Genomas de isolados de *K. pneumoniae*, *K. quasipneumoniae*, *K. variicola* e *K. oxytoca* compartilham um grande número de genes ortólogos, sendo o genoma core composto por mais de 2000 genes. Como consequência, os genomas destas espécies apresentaram um AF superior ao limite de 0.6, variando de 0.79 a 0.89.

O isolado 10982 foi definido, previamente, como um genoma intermediário entre as espécies *K. pneumoniae* e *K. variicola* ao ser comparado a 12 isolados destas espécies (Hazen et al. 2014). Aqui nós utilizamos um conjunto de dados maior e incluímos isolados da espécie *K. quasipneumoniae*. Os resultados aqui obtidos aplicando abordagens mais robustas de taxonomia genômica e filogenômicas (<94% ANI, < 55% GGD, <95% cgMLST), mostraram que o genoma de 10982 apresentou valores abaixo dos limites de definição de espécie para todas as métricas utilizadas e se mostrou genomicamente distinto de *K. pneumoniae*, *K. variicola*, *K. quasipneumoniae* e *K. oxytoca* (Tabela S3). Portanto, estes achados sugerem que o isolado 10982 poderia ser considerado uma nova espécie do gênero *Klebsiella* (Figuras 28, 29 e 30). Este isolado, assim como outras espécies deste gênero, carrega um gene cromossômico codificador de uma β -lactamase de classe A que possui uma identidade de 95%, 89% e 88% com *bla*_{LEN}, *bla*_{SHV} e *bla*_{OKP}, de *K. variicola*, *K. pneumoniae* e *K. quasipneumoniae*, respectivamente. Este gene *bla* foi caracterizado como um alelo variante de LEN (*bla*_{LEN-26}) apesar de sua identidade intermediária (92%). Outros alelos de *bla*_{LEN-26} compartilham uma identidade que varia entre 98 e 100%, indicando que este alelo pode, assim como em outras espécies do gênero *Klebsiella*, ser um marcador para esta espécie.

4.2.4 - Caracterização do plasmídeo pI9

Uma sequência consenso contendo 50.687 pb, gerada foi identificada como sendo parte do mobiloma deste isolado. Este *contig* possui um conteúdo GC de 51.4%, contrastando com os 57.3% do conteúdo global do genoma de Bz19. Esta região apresentou identidade de 99% com o plasmídeo 9 (pI9) (número de acesso FJ223607) encontrado em uma linhagem multirresistente de *K. pneumoniae* (Gootz et al. 2009). Este plasmídeo, nomeado como pBz19 (número de acesso

NZ_JDWA01000031), assim como pI9, pertence ao grupo de incompatibilidade IncN, encontrados em diferentes hospedeiros como bactérias dos gêneros *Klebsiella*, *Escherichia* e *Salmonella*.

Os plasmídeos pI9 e pBz19 compartilham grande parte do seu conteúdo gênico, incluindo 79 ORFs que codificam fatores transcricionais, transposases, proteínas regulatórias, a endonuclease Eco-RII e o sistema de secreção do tipo IV (T4SS-tra) envolvido com os eventos de conjugação (Craig et al. 2004) (Figura 31). A diferença entre os dois reside em duas regiões idênticas correspondendo ao fago *Tn4401*, que é responsável pela disseminação do gene *bla_{KPC-2}* encontrado no pI9. Neste plasmídeo os transpósons *Tn4401* contendo o gene *bla_{KPC-2}* estão inseridos entre dois fragmentos de um gene codificador de uma transcriptase reversa. No plasmídeo pBz19, entretanto, este gene codificador encontra-se intacto devido à ausência do elemento *Tn4401*. O plasmídeo pBz19 foi o primeiro relato de um plasmídeo em *K. variicola* (Andrade et al. 2014).

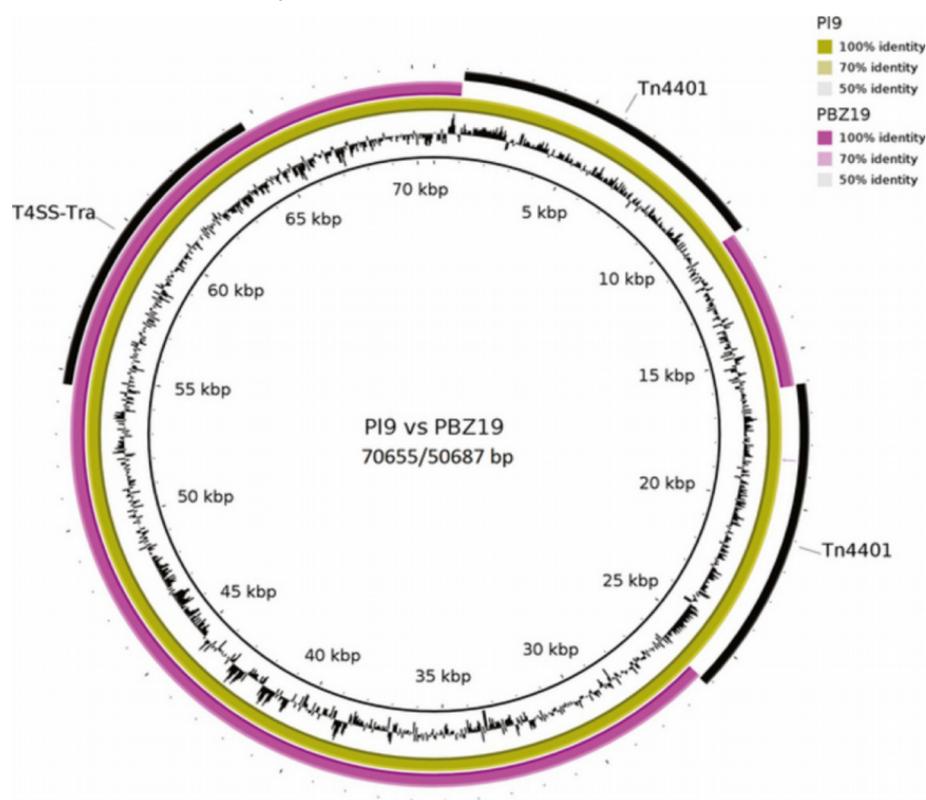


Figura 31 – Representação circular dos plasmídeos conjugativos *PI9*, encontrado em isolados de *K. pneumoniae* e *PBZ19*, primeiro plasmídeo identificado em *K. variicola*. O plasmídeo pI9 possui dois elementos *Tn4401*, profagos, que comportam a carbapenemase *bla_{KPC}*, o qual não é encontrado em *PBZ19*, sendo a única diferença entre estes.

5 - Conclusões

Este trabalho explorou a evolução e a diversidade genômica e funcional de genomas de Wolbachia, revelando que bactérias deste gênero possuem, em geral, potencial metabólico similar. A diferença funcional encontrada entre endossimbiontes facultativos e obrigatórios, foi a presença de muitas sequências repetitivas Ankyrin, elementos móveis e transposases, entre os facultativos, o que é mais uma evidência de sua recente evolução.

Identificamos , no genoma de Wolbachia wPipMol, genes de origem mitocondrial. Esta seria a primeira evidência da aquisição de genes desta organela por uma bactéria intracelular. Wolbachia wPipMol infecta mosquitos da espécie *Culex pipiens*, vetor de diversos patógenos. Essa transferência pode ter sido mediada por fago já que identificamos sequencias relacionadas a fagos junto associadas a estas sequencias mitocondriais.

A aplicação da taxonomia genômica para o gênero Wolbachia identificou o genoma wVulC como de uma nova espécie e as análises filogenômicas definiram que a mesma pertenceria a um novo supergrupo. Sugerimos o nome de Ca. 'W. vulgaris' para esta espécie.

A aplicação da taxonomia genômica aplicada para o gênero Klebsiella, revelou que 36 genomas de *K. pneumoniae* são das espécies próximas, *K. quasipneumoniae* e *K. variicola*. O genoma da Klebsiella sp.10982, apresentou assinaturas que o definiriam como uma nova espécie do gênero Klebsiella.

REFERÊNCIAS BIBLIOGRÁFICAS

- Acland a., Agarwala R, Barrett T, Beck J, Benson D a., Bollin C, Bolton E, Bryant SH, Canese K, Church DM, Clark K, DiCuccio M, Dondoshansky I, Federhen S, Feolo M, Geer LY, Gorelenkov V, Hoepfner M, Johnson M, Kelly C, Khotomlianski V, Kimchi A, Kimelman M, Kitts P, Krasnov S, Kuznetsov A, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Karsch-Mizrachi I, Murphy T, Ostell J, O'Sullivan C, Panchenko A, Phan L, Pruitt DPKD, Rubinstein W, Sayers EW, Schneider V, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Siyan K, Slotta D, Soboleva A, Soussov V, Starchenko G, Tatusova T a., Travers M, Travers M, Travers M, Wang Y, Ward M, John Wilbur W, Yaschenko E, Zbicz K 2014. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 42: D7–D17.
- Akhter S, Aziz RK, Edwards RA 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* 40: e126–e126.
- Alikhan N-F, Petty NK, Zakour NL Ben, Beatson SA 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12: 402.
- Alves MS, Dias RC da S, Castro ACD de, Riley LW, Moreira BM 2006. Identification of clinical isolates of indole-positive and indole-negative *Klebsiella* spp. *J. Clin. Microbiol.* 44: 3640–3646.
- Andersson JO, Andersson SG 1999. Genome degradation is an ongoing process in *Rickettsia*. *Mol. Biol. Evol.* 16: 1178–1191.
- Andrade BGN, Veiga Ramos N de, Abanto Marin MF, Fonseca EL, Vicente ACP 2014. The genome of a clinical *K. variicola* strain reveals virulence associated traits and a pl9-like plasmid. *FEMS Microbiol. Lett.*
- Angiuoli S V, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *OMICS* 12: 137–141.
- Auch AF, Henz SR, Holland BR, Göker M 2006. Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. *BMC Bioinformatics* 7: 350.
- Badawi M, Grève P, Cordaux R 2015. Feminization of the Isopod *Cylisticus convexus* after Transinfection of the wVulC *Wolbachia* Strain of *Armadillidium vulgare*. *PLoS One* 10: e0128660.
- Baldo L, Bordenstein S, Wernegreen JJ, Werren JH 2006a. Widespread recombination throughout *Wolbachia* genomes. *Mol. Biol. Evol.* 23: 437–449.
- Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH 2006b. Multilocus sequence

typing system for the endosymbiont *Wolbachia pipientis*. *Appl. Environ. Microbiol.* 72: 7098–7110.

- Bandi C, Anderson TJ, Genchi C, Blaxter ML 1998. Phylogeny of *Wolbachia* in filarial nematodes. *Proc. Biol. Sci.* 265: 2407–2413.
- Baptiste E, Susko E, Leigh J, MacLeod D, Charlebois RL, Doolittle WF 2005. Do orthologous gene phylogenies really support tree-thinking? *BMC Evol. Biol.* 5: 33.
- Bechah Y, Capo C, Mege J-L, Raoult D 2008. Epidemic typhus. *Lancet. Infect. Dis.* 8: 417–426.
- Bennett JS, Jolley K a, Maiden MCJ 2013. Genome sequence analyses show that *Neisseria oralis* is the same species as “*Neisseria mucosa* var. *heidelbergensis*”. *Int. J. Syst. Evol. Microbiol.* 63: 3920–3926.
- Berry GJ, Loeffelholz MJ, Williams-Bouyer N 2015. An Investigation into Laboratory Misidentification of a Bloodstream *Klebsiella variicola* Infection. *J. Clin. Microbiol.* 53: 2793–2794.
- Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K 2009. Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. *Mol. Biol. Evol.* 26: 231–241.
- Bordenstein SR, Wernegreen JJ 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol. Biol. Evol.* 21: 1981–1991.
- Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U, Doudoumis V, Scolari F, Benoit JB, Swain M, Takac P, Malacrida AR, Bourtzis K, Aksoy S 2014. Presence of extensive *Wolbachia* symbiont insertions discovered in the genome of its host *Glossina morsitans morsitans*. *PLoS Negl. Trop. Dis.* 8: e2728.
- Brenner DJ, Colwell RR, Grimont PAD, Krichevsky MI, Moore WEC, Murray RGE, Stackebrandt E, Starr MP, Triiper HG 2016. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rDNA. : 846–849.
- Brisse SJV 2014. Phylogenetic Diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, *gyrA* and *parC* genes sequencing and automated ribotyping. 41: 1–11.
- Brisse S, Himbergen T van, Kusters K, Verhoef J 2004. Development of a rapid identification method for *Klebsiella pneumoniae* phylogenetic groups and analysis of 420 clinical isolates. *Clin. Microbiol. Infect.* 10: 942–945.
- Brisse S, Passet V, Grimont P a D 2014. Description of *Klebsiella quasipneumoniae* sp. nov., a novel species isolated from human infections, with two subspecies *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and d. *Int. J. Syst. Evol.*

Microbiol.: 1–19.

- Brisse S, Verhoef J 2001. Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, *gyrA* and *parC* genes sequencing and automated ribotyping. *Int. J. Syst. Evol. Microbiol.* 51: 915–924.
- Brochier C, Bapteste E, Moreira D, Philippe H 2002. Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* 18: 1–5.
- Brown D, Sjölander K 2006. Functional classification using phylogenomic inference. *PLoS Comput. Biol.* 2: 0479–0483.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Campana MG, Robles García NM, Tuross N 2015. America's red gold: multiple lineages of cultivated cochineal in Mexico. *Ecol. Evol.* 5: 607–617.
- Caragata EP, Rancès E, O'Neill SL, McGraw E a 2014. Competition for amino acids between *Wolbachia* and the mosquito host, *Aedes aegypti*. *Microb. Ecol.* 67: 205–218.
- Cerveau N, Gilbert C, Liu C, Garrett RA, Grève P, Bouchon D, Cordaux R 2015. Genomic context drives transcription of insertion sequences in the bacterial endosymbiont *Wolbachia wVulC*. *Gene* 564: 81–86.
- Chen F, Mackey AJ, Stoeckert CJ, Roos DS 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* 34: D363–D368.
- Clec'h W Le, Braquart-Varnier C, Raimond M, Ferdy JB, Bouchon D, Sicard M 2012. High Virulence of *Wolbachia* after Host Switching: When Autophagy Hurts. *PLoS Pathog.* 8.
- Comandatore F, Cordaux R, Bandi C, Blaxter M, Darby A, Makepeace BL, Montagna M, Sasser D 2015. Supergroup C *Wolbachia*, mutualist symbionts of filarial nematodes, have a distinct genome structure. *Open Biol.* 5: 150099 – .
- Comandatore F, Sasser D, Montagna M, Kumar S, Koutsovoulos G, Thomas G, Repton C, Babayan S a, Gray N, Cordaux R, Darby A, Makepeace B, Blaxter M 2013. Phylogenomics and analysis of shared genes suggest a single transition to mutualism in *Wolbachia* of nematodes. *Genome Biol. Evol.* 5: 1668–1674.
- Comas I, Moya A, Gonzalez-Candelas F 2007. From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic gamma-Proteobacteria as a test case. *Syst Biol* 56: 1–16.
- Contreras-Moreira B, Vinuesa P 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79: 7696–7701.
- Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D 2004.

Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity (Edinb)*. 93: 78–84.

- Craig L, Pique ME, Tainer J a 2004. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* 2: 363–378.
- Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes M a., Kay SM, Koldkjær P, Rainbow L, Radford AD, Blaxter ML, Tanya VN, Trees AJ, Cordaux R, Wastling JM, Makepeace BL 2012. Analysis of gene expression from the *Wolbachia* genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. *Genome Res.* 22: 2467–2477.
- Delsuc F, Brinkmann H, Philippe H 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6: 361–375.
- Derks MFL, Smit S, Salis L, Schijlen E, Bossers A, Mateman C, Pijl AS, Ridder D de, Groenen M a. M, Visser ME, Megens H-J 2015. The Genome of Winter Moth (*Operophtera brumata*) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. *Genome Biol. Evol.* 7: 2321–2332.
- Desjardins C a, Cerqueira GC, Goldberg JM, Dunning Hotopp JC, Haas BJ, Zucker J, Ribeiro JMC, Saif S, Levin JZ, Fan L, Zeng Q, Russ C, Wortman JR, Fink DL, Birren BW, Nutman TB 2013. Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans. *Nat. Genet.* 45: 495–500.
- Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S 2005. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J. Clin. Microbiol.* 43: 4178–4182.
- Drancourt M, Bollet C, Carta A, Rousselier P 2001. Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. *Int. J. Syst. Evol. Microbiol.* 51: 925–932.
- Duploux A, Iturbe-Ormaetxe I, Beatson S a, Szubert JM, Brownlie JC, McMeniman CJ, McGraw E a, Hurst GDD, Charlat S, O'Neill SL, Woolfit M 2013. Draft genome sequence of the male-killing *Wolbachia* strain wBo11 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics* 14: 20.
- Dutilh BE, Jurgelenaite R, Szklarczyk R, Hijum S a FT van, Harhangi HR, Schmid M, Wild B de, François KJ, Stunnenberg HG, Strous M, Jetten MSM, Op den Camp HJM, Huynen M a. 2011. FACIL: Fast and accurate genetic code inference and logo. *Bioinformatics* 27: 1929–1933.
- Dutilh BE, Noort V van, Heijden RTJM van der, Boekhout T, Snel B, Huynen M a. 2007. Assessment of phylogenomic and orthology approaches for phylogenetic inference. *Bioinformatics* 23: 815–824.
- Dutilh BE, Thompson CC, Vicente AC, Marin M a, Lee C, Silva GG, Schmieder R, Andrade BG, Chimetto L, Cuevas D, Garza DR, Okeke IN, Aboderin AO, Spangler J, Ross T, Dinsdale E a, Thompson FL, Harkins TT, Edwards R a 2014.

- Comparative genomics of 274 *Vibrio cholerae* genomes reveals mobile functions structuring three niche dimensions. *BMC Genomics* 15: 654.
- Eddy SR 1998. Profile hidden Markov models. *Bioinforma. Rev.* 14: 755–763.
- Eisen JA, Eisen JA 1998. Phylogenomics : Improving Functional Predictions for Uncharacterized Genes by Evolutionary ? Analysis Phylogenomics : Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. : 163–167.
- Ellegaard KM, Klasson L, Näslund K, Bourtzis K, Andersson SGE 2013. Comparative genomics of *Wolbachia* and the bacterial species concept. *PLoS Genet.* 9: e1003381.
- Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* 3: e121.
- Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S, Dodson RJ, Mohamoud Y, Khouri H, Roesch LFW, Krogfelt KA, Struve C, Triplett EW, Methé BA 2008. Complete genome sequence of the N2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet.* 4.
- Gavotte L, Henri H, Stouthamer R, Charif D, Charlat S, Boulétreau M, Vavre F 2007. A Survey of the bacteriophage WO in the endosymbiotic bacteria *Wolbachia*. *Mol. Biol. Evol.* 24: 427–435.
- Georgiades K, Merhej V, Karkouri K El, Raoult D, Pontarotti P 2011. Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol. Direct* 6: 6.
- Gontcharov AA, Marin B, Melkonian M 2004. Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta). *Mol. Biol. Evol.* 21: 612–624.
- Gootz TD, Lescoe MK, Dib-Hajj F, Dougherty B a, He W, Della-Latta P, Huard RC 2009. Genetic organization of transposase regions surrounding blaKPC carbapenemase genes on plasmids from *Klebsiella* strains isolated in a New York City hospital. *Antimicrob. Agents Chemother.* 53: 1998–2004.
- Hazen TH, Zhao L, Sahl JW, Robinson G, Harris AD, Rasko DA, Johnson JK 2014. Characterization of *Klebsiella* sp. strain 10982, a colonizer of humans that contains novel antibiotic resistance alleles and exhibits genetic similarities to plant and clinical *Klebsiella* isolates. *Antimicrob. Agents Chemother.* 58: 1879–1888.
- Herndon DR, Palmer GH, Shkap V, Knowles DP, Brayton K a 2010. Complete

- genome sequence of *Anaplasma marginale* subsp. *centrale*. *J. Bacteriol.* 192: 379–380.
- Herniou EA, Luque T, Chen X, Vlak JM, Winstanley D, Cory JS, O'Reilly DR 2001. Use of whole genome sequence data to infer baculovirus phylogeny. *J. Virol.* 75: 8117–8126.
- Hertig M 2009. The Rickettsia, *Wolbachia pipientis* (gen. et sp.n.) and Associated Inclusions of the Mosquito, *Culex pipiens*. *Parasitology* 28: 453.
- Huigens ME, Luck RF, Klaassen RH, Maas MF, Timmermans MJ, Stouthamer R 2000. Infectious parthenogenesis. *Nature* 405: 178–179.
- Hurst GDD, Jiggins FM 2000. Male-killing bacteria in insects: Mechanisms, incidence, and implications. *Emerg. Infect. Dis.* 6: 329–336.
- Ishmael N, Hotopp JCD, Loanidis P, Biber S, Sakamoto J, Siozios S, Nene V, Werren J, Boutriz K, Bordenstein SR, Tettelin H 2009. Extensive genomic diversity of closely related *wolbachia* strains. *Microbiology* 155: 2211–2222.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22: 225–231.
- Jiggins FM 2002. The rate of recombination in *Wolbachia* bacteria. *Mol. Biol. Evol.* 19: 1640–1643.
- John H. Werren, Wan Zhang LRG 1995. Evolution and Phylogeny of *Wolbachia*: Reproductive Parasites of Arthropods. *Proc. Biol. Sci.* 261: 55–63.
- Jolley K a, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MCJ 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 158: 1005–1015.
- Jukes TH, Osawa S 1990. The genetic code in mitochondria and chloroplasts. *Experientia* 46: 1117–1126.
- Katoh K, Kuma K, Toh H, Miyata T 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33: 511–518.
- Kent BN, Salichos L, Gibbons JG, Rokas A, Newton ILG, Clark ME, Bordenstein SR 2011. Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. *Genome Biol. Evol.* 3: 209–218.
- Klasson L, Walker T, Sebahia M, Sanders MJ, Quail M a., Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J 2008. Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. *Mol. Biol. Evol.* 25: 1877–1887.
- Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SGE 2009. The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proc. Natl. Acad. Sci. U. S. A.* 106: 5725–5730.

- Kocan KM, la Fuente J de, Guglielmono AA, Meléndez RD 2003. Antigens and alternatives for control of *Anaplasma marginale* infection in cattle. *Clin. Microbiol. Rev.* 16: 698–712.
- Kodama Y, Shumway M, Leinonen R 2012. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.* 40: D54–D56.
- Konstantinidis KT, Tiedje JM 2005a. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102: 2567–2572.
- Konstantinidis KT, Tiedje JM 2005b. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187: 6258–6264.
- Kremer N, Voronin D, Charif D, Mavingui P, Mollereau B, Vavre F 2009. *Wolbachia* interferes with ferritin expression and iron metabolism in insects. *PLoS Pathog.* 5: e1000630.
- Larsen M V, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Pontén T, Aarestrup FM, Ussery DW, Lund O 2014. Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* 52: 1529–1539.
- Lefoulon E, Bain O, Makepeace BL, d'Haese C, Uni S, Martin C, Gavotte L 2016. Breakdown of coevolution between symbiotic bacteria *Wolbachia* and their filarial hosts. *PeerJ* 4: e1840.
- Liu P, Li P, Jiang X, Bi D, Xie Y, Tai C, Deng Z, Rajakumar K, Ou HY 2012. Complete genome sequence of *Klebsiella pneumoniae* subsp. *pneumoniae* HS11286, a multidrug-resistant strain isolated from human sputum. *J. Bacteriol.* 194: 1841–1842.
- Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH, Bordenstein SR, Bandi C 2007. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *Int. J. Syst. Evol. Microbiol.* 57: 654–657.
- Mavingui P, Moro CV, Tran-Van V, Wisniewski-Dy?? F, Raquin V, Minard G, Tran FH, Voronin D, Rouy Z, Bustos P, Lozano L, Barbe V, Gonz??lez V 2012. Whole-genome sequence of *Wolbachia* strain wALbB, an endosymbiont of tiger mosquito vector *Aedes albopictus*. *J. Bacteriol.* 194: 1840.
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14: 60.
- Metcalf J a, Jo M, Bordenstein SR, Jaenike J, Bordenstein SR 2014. Recent genome reduction of *Wolbachia* in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ* 2: e529.
- Moran NA 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* 6: 512–518.
- Moran NA 2006. Molecular Interactions between Bacterial Symbionts and Their Hosts. *Curr. Biol.* 16: R866–R871.

- Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T 2014. Evolutionary origin of insect-Wolbachia nutritional mutualism. *Proc. Natl. Acad. Sci. U. S. A.* 111: 10257–10262.
- Overbeek R, Begley T, Butler RM, Choudhuri J V., Chuang HY, Cohoon M, Crécy-Lagard V de, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov D a., Rülckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33: 5691–5702.
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards R a, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42: D206–D214.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW 2015. CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. *Genome Res.* 25: 1043–1055.
- Pfarr K, Foster J, Slatko B, Hoerauf A, Eisen JA 2011. On the taxonomic status of the intracellular bacterium Wolbachia pipientis: should this species name include the intracellular bacteria of filarial nematodes? *Int. J. Syst. Evol. Microbiol.* 57: 1677–1678.
- Pinto SB, Stainton K, Harris S, Kambris Z, Sutton ER, Bonsall MB, Parkhill J, Sinkins SP 2013. Transcriptional regulation of Culex pipiens mosquitoes by Wolbachia influences cytoplasmic incompatibility. *PLoS Pathog.* 9: e1003647.
- Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FST, Cleland WW, Weimer PJ, Currie CR 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* 326: 1120–1123.
- Poinsot D, Charlat S, Merçot H 2003. On the mechanism of Wolbachia-induced cytoplasmic incompatibility: Confronting the models with the facts. *BioEssays* 25: 259–265.
- R Development Core Team 2011. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.*
- Ramírez-Puebla ST, Servín-Garcidueñas LE, Ormeño-Orrillo E, Vera-Ponce de León A, Rosenblueth M, Delaye L, Martínez J, Martínez-Romero E 2015. Species in Wolbachia? Proposal for the designation of “Candidatus Wolbachia bourtzisii”, “Candidatus Wolbachia onchocercicola”, “Candidatus Wolbachia blaxteri”, “Candidatus Wolbachia brugii”, “Candidatus Wolbachia taylori”, ‘Candidatus Wolbachia collembol. *Syst. Appl. Microbiol.*: 1–10.
- Rigaud T, Soutygrosset C, Raimond R, Mocquard JP, Juchault P 1991. Feminizing endocytobiosis in the terrestrial crustacean Armadillidium vulgare latr (isopoda) -

- recent acquisitions. *Endocytobiosis Cell Res.* 7: 259–273.
- Rosenblueth M, Martínez L 2004. *Klebsiella variicola*, A Novel Species with Clinical and Plant-Associated Isolates. *Syst. Appl. Microbiol.* 35: 27–35.
- Rosselló-Mora R, Amann R 2015. Past and future species definitions for Bacteria and Archaea. *Syst. Appl. Microbiol.* 38: 209–216.
- Rosselló-Mora R 2005. Updating prokaryotic taxonomy. *J. Bacteriol.* 187: 6255–6257.
- Roux S, Enault F, Hurwitz BL, Sullivan MB 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3: e985.
- Saha S, Hunter WB, Reese J, Morgan JK, Marutani-Hert M, Huang H, Lindeberg M 2012. Survey of endosymbionts in the *Diaphorina citri* metagenome and assembly of a *Wolbachia* wDi draft genome. *PLoS One* 7: e50067.
- Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC 2005. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol.* 6: 402.
- Salzberg SL, Puiu D, Sommer DD, Nene V, Lee NH 2009. Genome sequence of the *Wolbachia* endosymbiont of *Culex quinquefasciatus* JHB. *J. Bacteriol.* 191: 1725.
- Schumann P, Pukall R, Spröer C, Stackebrandt E 2013. Reclassification of *Koreibacter* algae as a later heterotypic synonym of *Paraoerskovia marina* and emended descriptions of the genus *Paraoerskovia* Khan et al. 2009 and of *Paraoerskovia marina* Khan et al. 2009. *Int. J. Syst. Evol. Microbiol.* 63: 219–223.
- Seemann T 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068–2069.
- Simonsen PE, Mwakitalu ME 2013. Urban lymphatic filariasis. *Parasitol. Res.* 112: 35–44.
- Siozios S, Cestaro A, Kaur R 2013. Draft genome sequence of the *Wolbachia* endosymbiont of *Drosophila suzukii*. *Genome Announc.* 1: 1–2.
- Slatko BE, Taylor MJ, Foster JM 2010. The *Wolbachia* endosymbiont as an anti-filarial nematode target. *Symbiosis* 51: 55–65.
- Snel B, Bork P, Huynen MA 1999. Genome phylogeny based on gene content. *Nat. Genet.* 21: 108–110.
- Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, Eijk R van, Schleper C, Guy L, Ettema TJG 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521: 173–179.
- Stamatakis A 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

- Sutton ER, Harris SR, Parkhill J, Sinkins SP 2014. Comparative genome analysis of *Wolbachia* strain wAu. *BMC Genomics* 15: 928.
- Thompson CC, Amaral GR, Campeão M, Edwards R a, Polz MF, Dutilh BE, Ussery DW, Sawabe T, Swings J, Thompson FL 2015. Microbial taxonomy in the post-genomic era: Rebuilding from scratch? *Arch. Microbiol.* 197: 359–370.
- Thompson CC, Silva GGZ, Vieira NM, Edwards R, Vicente ACP, Thompson FL 2013. Genomic taxonomy of the genus *prochlorococcus*. *Microb. Ecol.* 66: 752–762.
- Thompson CC, Vicente ACP, Souza RC, Vasconcelos ATR, Vesth T, Alves N, Ussery DW, Iida T, Thompson FL 2009. Genomic taxonomy of *Vibrios*. *BMC Evol. Biol.* 9: 258.
- Tritt A, Eisen JA, Facciotti MT, Darling AE 2012. An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* 7: e42304.
- Turell MJ 2012. Members of the *Culex pipiens* complex as vectors of viruses. *J. Am. Mosq. Control Assoc.* 28: 123–126.
- Turelli M, Hoffmann AA 1999. Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol. Biol.* 8: 243–255.
- Tzelos T, Matthews JB, Buck AH, Simbari F, Frew D, Inglis NF, McLean K, Nisbet AJ, Whitelaw CBA, Knox DP, McNeilly TN 2016. A preliminary proteomic characterisation of extracellular vesicles released by the ovine parasitic nematode, *Teladorsagia circumcincta*. *Vet. Parasitol.* 221: 84–92.
- Vandamme P, Pot B, Gillis M, Vos P de, Kersters K, Swings J 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60: 407–438.
- Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43: gkv657 – .
- Wernegreen JJ 2002. Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 3: 850–861.
- Werren JH, Bartos JD 2001. Recombination in *Wolbachia*. *Curr. Biol.* 11: 431–435.
- Woese CR, Fox GE 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74: 5088–5090.
- Woolfit M, Iturbe-Ormaetxe I, Brownlie JC, Walker T, Riegler M, Seleznev A, Popovici J, Rances E, Wee BA, Pavlides J, Sullivan MJ, Beatson SA, Lane A, Sidhu M, McMeniman CJ, McGraw EA, O'Neill SL 2013. Genomic Evolution of the Pathogenic *Wolbachia* Strain, wMelPop. *Genome Biol. Evol.* 5: 2189–2204.
- Wu M, Eisen J a 2008. A simple, fast, and accurate method of phylogenomic inference. *Genome Biol.* 9: R151.
- Wu M, Sun L V., Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw E a.,

Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin a. S, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen J a. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* 2: E69.

Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* 39: W347–W352.

Zug R, Hammerstein P 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7: e38544.

Material Suplementar

Tabela S1 – Pan-genoma de isolados do gênero *Wolbachia* contendo 1272 grupos gênicos. A tabela contém o identificador (Locus tag), e os níveis de anotação de subsistemas Seed 1-3.

Identificador	Nível 1	Nível 2	Nível 3
wCam_00711	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos FASII
wCam_00057	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wCam_00291	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wCam_00325	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wCam_00679	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wCam_00781	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wCam_00821	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Biossíntese de ácidos graxos em Plantas (plastidial)
wAlb_00750	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Oxidase putativa COG2907
wCle_00897	Ácidos graxos, lipídeos e isoprenóides	Ácidos graxos	Oxidase putativa COG2908
wAlb_00433	Ácidos graxos, lipídeos e isoprenóides	Fosfolipídeos	Metabolismo de glicerolipídeo e glicerofosfolipídeo
wCam_00046	Ácidos graxos, lipídeos e isoprenóides	Fosfolipídeos	Metabolismo de glicerolipídeo e glicerofosfolipídeo
wCam_00047	Ácidos graxos, lipídeos e isoprenóides	Fosfolipídeos	Metabolismo de glicerolipídeo e glicerofosfolipídeo
wCam_00054	Ácidos graxos, lipídeos e isoprenóides	Fosfolipídeos	Metabolismo de glicerolipídeo e glicerofosfolipídeo
wCam_00394	Ácidos graxos, lipídeos e isoprenóides	Fosfolipídeos	Metabolismo de glicerolipídeo e glicerofosfolipídeo
wAu_01187	Ácidos graxos, lipídeos e isoprenóides	Isoprenóides	Lipídeos de arqueia
wCam_00055	Ácidos graxos, lipídeos e isoprenóides	Isoprenóides	Lipídeos de arqueia
wCam_00281	Aminoácidos e derivados	Alanina	Serina
wCam_00317	Aminoácidos e derivados	Alanina	Serina
wCle_00702	Aminoácidos e derivados	Alanina	Serina
wAu_00485	Aminoácidos e derivados	Arginina	Ciclo da uréia
wCam_00083	Aminoácidos e derivados	Arginina	Ciclo da uréia
wCam_00773	Aminoácidos e derivados	Arginina	Ciclo da uréia
wAlb_00647	Aminoácidos e derivados	Glutamina	Glutamato
wCam_00210	Aminoácidos e derivados	Glutamina	Glutamato
wCam_00683	Aminoácidos e derivados	Glutamina	Glutamato
wAlb_00597	Aminoácidos e derivados	Lysina	Threonina
wAlb_00598	Aminoácidos e derivados	Lysina	Threonina
wCam_00262	Aminoácidos e derivados	Prolina e 4-hydroxyProlina	Prolina

wCam_00817	Aquisição e metabolismo de ferro	Ferric iron ABC transporter	Proteína de ligação ao ferro
wCam_00768	Aquisição e metabolismo de ferro	Ferric iron ABC transporter	Proteína permeabilizadora
wCam_00554	Carboidratos	Ácidos orgânicos	Metabolismo de citrato KE
wCam_00056	Carboidratos	Ácidos orgânicos	Metabolismo de citrato KE3
wCam_00194	Carboidratos	Ácidos orgânicos	Metabolismo de citrato KE3
wCam_00385	Carboidratos	Ácidos orgânicos	Metabolismo de citrato KE3
wCam_00386	Carboidratos	Ácidos orgânicos	Metabolismo de citrato KE3
wAlb_00116	Carboidratos	Ácidos orgânicos	Tricarballoylato, utilização
wAlb_00599	Carboidratos	Áçucars de alcool	Aquisição e metabolismo de gliceról
wCam_00093	Carboidratos	Áçucars de alcool	Aquisição e metabolismo de gliceról
wCam_00099	Carboidratos	Áçucars de alcool	Catabolismo de inositol
wAlb_00600	Carboidratos	Áçucars nucleotídicos	UDP-D-glucuronate
wCam_00072	Carboidratos	Amino açúcars	Utilização de quitina e N-acetilglucosamina
wCle_00148	Carboidratos	Amino açúcars	Utilização de quitina e N-acetilglucosamina
wHa_00380	Carboidratos	Amino açúcars	Utilização de quitina e N-acetilglucosamina
wRi_00321	Carboidratos	Amino açúcars	Utilização de quitina e N-acetilglucosamina
wTRS_00623	Carboidratos	Amino açúcars	Utilização de quitina e N-acetilglucosamina
wCam_00048	Carboidratos	Fixação de CO2	Ciclo de Calvin-Benson-Bassham em plantas
wCam_00141	Carboidratos	Fixação de CO2	Ciclo de Calvin-Benson
wCam_00426	Carboidratos	Fixação de CO2	Ciclo de Calvin-Benson
wCam_00500	Carboidratos	Fixação de CO2	Fotorespiração
wJHB_00425	Carboidratos	Fixação de CO2	Fotorespiração
wCam_00019	Carboidratos	Fot	Fotorespiração
wCam_00124	Carboidratos	Metabolismo central de carboidratos	Biossíntese de Acetyl-CoA em plantas
wCam_00822	Carboidratos	Metabolismo central de carboidratos	Glicólise e glicogênese em plantas
wCam_00589	Carboidratos	Metabolismo de carbono único	Metabolismo de carbono único em Plantas mediado por folato
wJHB_00935	Carboidratos	Metabolismo de carbono único	Serine-glyoxylate cycle
wCam_00445	Carboidratos	Monossacarídeos	D-Sorbitol(D-Glucitol) e L-Sorbose utilização
wAlb_00125	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAlb_00543	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAlb_00544	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAlb_00548	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAlb_00552	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAlb_00588	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAu_00251	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wAu_00256	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wBolb_00293	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase

wJHB_00897	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wJHB_01399	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wMel_00252	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wTRS_00260	Carboidratos	Polissacarídeos	Frutose-bifosfato aldolase
wCle_00952	Cofatores, vitaminas, grupos prostéticos e pigmentos	Biotina	Biotina biossíntese em Plantas
wCle_00953	Cofatores, vitaminas, grupos prostéticos e pigmentos	Biotina	Biotina biossíntese em Plantas
wCle_00956	Cofatores, vitaminas, grupos prostéticos e pigmentos	Biotina	Biotina biossíntese em Plantas
wCle_00957	Cofatores, vitaminas, grupos prostéticos e pigmentos	Biotina	Biotina biossíntese em Plantas
wCam_00671	Cofatores, vitaminas, grupos prostéticos e pigmentos	Biotina	Biotina síntese & utilização
wAlb_00053	Cofatores, vitaminas, grupos prostéticos e pigmentos	CLO thiaminPP biossíntese	Thiaminase II (EC 3.5.99.2) involved em salvage of thiamin pyrimidine moiety
wAlb_00054	Cofatores, vitaminas, grupos prostéticos e pigmentos	CLO thiaminPP biossíntese	Thiaminase II (EC 3.5.99.2) involved em salvage of thiamin pyrimidine moiety
wCle_00301	Cofatores, vitaminas, grupos prostéticos e pigmentos	CLO thiaminPP biossíntese	Thiaminase II (EC 3.5.99.2) involved em salvage of thiamin pyrimidine moiety
wAlb_00478	Cofatores, vitaminas, grupos prostéticos e pigmentos	Coenzyme A	CoA Pantothenate HMP
wCam_00178	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00311	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00413	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00420	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00527	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00743	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	CH UbiG
wCam_00032	Cofatores, vitaminas, grupos prostéticos e pigmentos	Cofatores Quinona	Coenzyme PQQ síntese
wCle_00299	Cofatores, vitaminas, grupos prostéticos e pigmentos	Experimental ThMP RZ	Hydroxyethylthiazole kinase (EC 2.7.1.50)
wAu_00512	Cofatores, vitaminas, grupos prostéticos e pigmentos	Experimental ThMP RZ	Thiamin pyrofosfokinase (EC 2.7.6.2)
wTRS_00949	Cofatores, vitaminas, grupos prostéticos e pigmentos	Experimental ThMP RZ	Thiamin pyrofosfokinase (EC 2.7.6.2)
wCam_00031	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	5-FCL-like protein
wCam_0074	Cofatores, vitaminas,	Folato e pterinas	5-FCL-like protein

4	grupos prostéticos e pigmentos		
wCam_00200	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wCam_00201	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wCam_00202	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wCam_00203	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wHa_00878	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wHa_00879	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Folato Tryp Chlamydiae
wCam_00556	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	Pterin metabolism 3
wCam_00011	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00073	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00145	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00169	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00242	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00570	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00590	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00733	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00780	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00785	Cofatores, vitaminas, grupos prostéticos e pigmentos	Folato e pterinas	YgfZ
wCam_00205	Cofatores, vitaminas, grupos prostéticos e pigmentos	NAD e NADP	NAD consumption
wAlb_00035	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wAlb_00094	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wAlb_00795	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wAu_00132	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina

	pigmentos		
wAu_00998	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wBolb_01202	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wCam_00236	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wCam_00279	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wCam_00328	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wCam_00828	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wHa_00837	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wTRS_00024	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wTRS_00025	Cofatores, vitaminas, grupos prostéticos e pigmentos	Riboflavina FMN FAD	Test - Riboflavina
wAlb_00278	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAlb_00453	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAlb_00675	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAlb_00747	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAlb_00826	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00138	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00412	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00443	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00536	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00619	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00799	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00802	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_00837	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I

wAu_00866	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_01163	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_01175	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wAu_01236	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCam_00347	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCitri_00345	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00133	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00135	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00373	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00776	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00955	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCle_00967	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wHa_00510	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wNo_00233	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wRec_00174	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wTRS_00114	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wTRS_00218	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wTRS_01068	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wVuIC_00600	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	CbiZ Main I
wCam_00015	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00100	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00128	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_0018	Cofatores, vitaminas,	Tetrapirróis	Experimental tye

1	grupos prostéticos e pigmentos		
wCam_00318	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00342	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00531	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00762	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Experimental tye
wCam_00555	Cofatores, vitaminas, grupos prostéticos e pigmentos	Tetrapirróis	Heme biossíntese: protoporfyrin- e coproporfyrin-dependent pathways
wAlb_00141	Dano de metabólitos e seu reparo ou mitigação	Cópias de proteína RidA	Carbonic anhydrase (EC 4.2.1.1)
wCam_00129	Dano de metabólitos e seu reparo ou mitigação	TN RidA: todos os supergrupos	981 genomes carbamoyl-P
wAu_01262	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsI [Peptidoglycan sintetase] (EC 2.4.1.129)
wCam_00745	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsI [Peptidoglicano sintetase] (EC 2.4.1.129)
wHa_01223	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsI [Peptidoglicano sintetase] (EC 2.4.1.129)
wTRS_00061	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsI [Peptidoglicano sintetase] (EC 2.4.1.129)
wAu_00066	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsL
wCam_00621	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsQ
wAu_00435	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsW
wTRS_00285	Divisão e ciclo celular	Citoesqueleto bacteriano	Proteína de divisão celular FtsW
wAlb_00710	Divisão e ciclo celular	Citoesqueleto bacteriano	Chromosome (plasmídeo) proteína de partição ParA
wTRS_01015	Divisão e ciclo celular	Citoesqueleto bacteriano	Cromossomo(plasmídeo) proteína de partição ParA
wAlb_00711	Divisão e ciclo celular	Citoesqueleto bacteriano	Cromossomo(plasmídeo) proteína de partição ParB
wTRS_01016	Divisão e ciclo celular	Citoesqueleto bacteriano	Cromossomo(plasmídeo) proteína de partição ParB
wCam_00111	Divisão e ciclo celular	Citoesqueleto bacteriano	Rod shape-determining protein MreB
wCam_00560	Divisão e ciclo celular	Citoesqueleto bacteriano	Rod shape-determining protein RodA
wAu_00489	Fagos, profagos, elementos transponíveis e plasmídeos	Bacteriofagos, proteínas estruturais	Proteínas de capsídeo de fagos
wJHB_01292	Fagos, profagos, elementos transponíveis e plasmídeos	Bacteriofagos, proteínas estruturais	Proteínas de capsídeo de fagos
wAu_00222	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS- Fago terminase	Proteína Portal
wAu_00298	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS- Fago terminase	Proteína Portal
wJHB_01289	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS- Fago terminase	Proteína Portal
wBolb_00378	Fagos, profagos, elementos transponíveis e	CBSS- Fago terminase	Lisozimas de profagos(EC 3.2.1.17)

wAu_00365	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00423	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00537	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00648	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00652	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00708	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00747	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00748	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00831	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00873	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00887	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00948	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00994	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00995	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_01009	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_01010	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_01015	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00123	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00321	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00356	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00370	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00431	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_0043	Fagos, profagos,	CBSS-227321.4.peg.3751	Proteína de elemento móvel

5	elementos transponíveis e plasmídeos		
wBolb_00447	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00532	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00533	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_00820	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01249	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01254	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01255	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01256	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01258	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01259	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01267	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wBolb_01268	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCitri_00066	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCitri_00283	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCitri_00427	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCitri_00514	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCitri_00654	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00048	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00060	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00431	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00932	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00946	Fagos, profagos, elementos transponíveis e	CBSS-227321.4.peg.3751	Proteína de elemento móvel

	plasmídeos		
wHa_00130	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wHa_00134	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wHa_00168	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wHa_00496	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wHa_00541	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wHa_00839	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00330	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00434	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00756	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00758	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00788	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_00867	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01015	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01016	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01080	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01199	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01201	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01295	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01304	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01320	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01321	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wJHB_01403	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel

wJHB_01405	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00043	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00051	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00063	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00237	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00257	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00349	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00510	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00517	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00550	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00643	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_00920	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMel_01105	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wMelPop_00259	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wNo_00055	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wNo_00205	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wNo_01160	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wNo_01162	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPip_00052	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPip_00053	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPip_00158	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPip_00159	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPip_00693	Fagos, profagos,	CBSS-227321.4.peg.3751	Proteína de elemento móvel

	elementos transponíveis e plasmídeos		
wPip_01423	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPipMol_00336	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPipMol_01056	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wPipMolestus_01055	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRec_00087	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRec_00195	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRec_00197	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRec_00324	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRec_00899	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00040	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00139	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00249	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00261	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00262	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00313	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00368	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00392	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00413	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00517	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00857	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00868	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_00870	Fagos, profagos, elementos transponíveis e	CBSS-227321.4.peg.3751	Proteína de elemento móvel

	plasmídeos		
wRi_00987	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wRi_01032	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wValsugana01344	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wValsugana01346	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00009	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00037	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00100	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00101	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00127	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00131	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00144	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00291	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00303	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00329	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00331	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00333	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00335	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00362	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00451	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00472	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00512	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00645	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel

wVuIC_00656	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00887	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00958	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00961	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_00965	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01009	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01033	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01085	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01088	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01220	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01274	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01330	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wVuIC_01795	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wCle_00011	Fagos, profagos, elementos transponíveis e plasmídeos	CBSS-227321.4.peg.3751	Proteína de elemento móvel
wAu_00371	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wAu_00709	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wAu_00829	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wBolb_00440	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wCle_00862	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wCle_00935	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wHa_00823	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wHa_00874	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wRi_00074	Fagos, profagos,	Elementos transponíveis	Tn552

	elementos transponíveis e plasmídeos		
wRi_00719	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wValsugana01321	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wVuIC_01444	Fagos, profagos, elementos transponíveis e plasmídeos	Elementos transponíveis	Tn552
wAu_00279	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wBolb_00417	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wJHB_00916	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wJHB_00918	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wJHB_01437	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wJHB_01438	Fagos, profagos, elementos transponíveis e plasmídeos	Fagos, estrutural	T4-like fage virulence associated protein
wAlb_00376	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wAu_00388	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wAu_01277	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00233	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00234	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00326	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00607	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00608	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00609	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00610	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00664	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wCam_00714	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wMel_01272	Fagos, profagos, elementos transponíveis e	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo

	plasmídeos		
wRi_00462	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	Transferência de T-DNA codificada por plasmídeo
wAlb_00096	Fagos, profagos, elementos transponíveis e plasmídeos	Funções relacionadas a plasmídeos	
wAu_00294	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wAu_00823	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wAu_00824	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wJHB_00786	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wJHB_01341	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wJHB_01342	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wJHB_01520	Fagos, profagos, elementos transponíveis e plasmídeos	Maquinaria de empacotamento de fagos	Terminase
wCam_00573	Fagos, profagos, elementos transponíveis e plasmídeos	Metabolismo de carbono de Fagos	Transaldolase (EC 2.2.1.2)
wHa_00701	Fagos, profagos, elementos transponíveis e plasmídeos	N/A	Proteína não caracterizada de fago
wJHB_00315	Fagos, profagos, elementos transponíveis e plasmídeos	N/A	Proteína não caracterizada de fago
wCitri_01137	Fagos, profagos, elementos transponíveis e plasmídeos	N/A	Funções relacionadas a fagos e profagos
wAu_00716	Fagos, profagos, elementos transponíveis e plasmídeos	N/A	Funções relacionadas a fagos e profagos
wAlb_00107	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAlb_01009	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAlb_01016	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAu_00160	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAu_00289	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAu_00308	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAu_00664	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wBolb_00012	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias

wBolb_00335	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wBolb_01009	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wCitri_01110	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wCitri_01151	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wJHB_00353	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wJHB_00936	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wJHB_01347	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wJHB_01348	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wMel_00568	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wNo_01097	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wNo_01109	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wNo_01110	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wPip_00447	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wPipMol_01091	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wPipMol_01167	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wVuIC_00993	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wVuIC_01104	Fagos, profagos, elementos transponíveis e plasmídeos	Prófagos	Grupos de profagos em proteobacterias
wAlb_01005	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Subunidade protéica maior de cauda de fago
wBolb_00284	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Subunidade protéica maior de cauda de fago
wCitri_01073	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Subunidade protéica maior de cauda de fago
wVuIC_00665	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Subunidade protéica maior de cauda de fago
wVuIC_01557	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Subunidade protéica maior de cauda de fago
wAlb_01004	Fagos, profagos,	Proteínas de cauda de fagos	Monomero de cauda de fago

	elementos transponíveis e plasmídeos		
wAu_00274	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Monomero de cauda de fago
wMel_00573	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Monomero de cauda de fago
wMel_00574	Fagos, profagos, elementos transponíveis e plasmídeos	Proteínas de cauda de fagos	Monomero de cauda de fago
wCam_00580	Fagos, profagos, elementos transponíveis e plasmídeos	Subsistemas família-específica de fagos	T4-like fage core proteins
wCam_00699	Fagos, profagos, elementos transponíveis e plasmídeos	Subsistemas família-específica de fagos	T4-like fage core proteins
wAu_00300	Fagos, profagos, elementos transponíveis e plasmídeos	T4-like proteínas core	fage baseplate-tail tube emitiator (T4-like gp54)
wAu_00653	Fagos, profagos, elementos transponíveis e plasmídeos	T4-like proteínas core	fage baseplate-tail tube emitiator (T4-like gp54)
wNo_00982	Fagos, profagos, elementos transponíveis e plasmídeos	T4-like proteínas core	fage baseplate-tail tube emitiator (T4-like gp54)
wAu_00291	Fagos, profagos, elementos transponíveis e plasmídeos	T4-like proteínas core	DNA Metilase
wAu_00302	Fagos, profagos, elementos transponíveis e plasmídeos	T4-like proteínas core	Funções relacionadas a fagos e profagos
wCam_00594	Metabolismo de compostos aromáticos	Vias periféricas para o catabolismo de compostos aromáticos	Benzoate degradation
wAlb_00751	Metabolismo de DNA	DNA structural proteins	bacterial
wAlb_00721	Metabolismo de DNA	DNA uptake	competence
wCam_00249	Metabolismo de DNA	DNA uptake	competence
wAlb_00112	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wAlb_00549	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wAlb_00587	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wAu_00394	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wCam_00144	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wCitri_00202	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wCitri_00243	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wCitri_01145	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wJHB_00884	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wJHB_01275	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wJHB_01367	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wJHB_01533	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wMel_00255	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wMel_00501	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wMelPop_00377	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wNo_00547	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wTRS_00242	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial
wAlb_00007	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial RecFOR pathway
wCam_0059	Metabolismo de DNA	Reparo de DNA	Reparo de DNA bacterial RecFOR pathway

6				
wCam_00668	Metabolismo de DNA	Reparo de DNA	Reparo de DNA Base Excision	
wCam_00586	Metabolismo de DNA	Reparo de DNA	Reparo de DNA sistema incluindo RecA	
wAu_00395	Metabolismo de DNA	Reparo de DNA	Reparo de DNA UvrABC sistema	
wCam_00042	Metabolismo de DNA	Reparo de DNA	Reparo de DNA UvrABC sistema	
wCam_00199	Metabolismo de DNA	Reparo de DNA	Reparo de DNA UvrABC sistema	
wCam_00582	Metabolismo de DNA	Reparo de DNA	Reparo de DNA UvrABC sistema	
wCam_00538	Metabolismo de DNA	Reparo de DNA	Uracil-DNA glycosylase	
wAlb_00725	Metabolismo de DNA	Replicação de DNA	Replicação de DNA	
wAlb_00956	Metabolismo de DNA	Replicação de DNA	Replicação de DNA	
wBolb_00748	Metabolismo de DNA	Replicação de DNA	Replicação de DNA	
wBolb_00749	Metabolismo de DNA	Replicação de DNA	Replicação de DNA	
wCam_00070	Metabolismo de DNA	Replicação de DNA	Replicação de DNA	
wAu_00665	Metabolismo de DNA	Restriction-Modification sistema	Putative DNA-binding protein em cluster with Tipo I restriction-modification sistema	
wCam_00327	Metabolismo de DNA	Restriction-Modification sistema	Putative predicted metal-dependent hydrolase	
wCam_00112	Metabolismo de enxofre	Assimilação de enxofre orgânico	At5g37530 (CsdL protein Família)	
wCam_00244	Metabolismo de enxofre	Assimilação de enxofre orgânico	At5g37530 (CsdL protein Família)	
wCam_00674	Metabolismo de enxofre	Assimilação de enxofre orgânico	At5g37530 (CsdL protein Família)	
wCam_00075	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00134	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00260	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00337	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00567	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00604	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00677	Metabolismo de proteína	Biogênese proteica	EC 6.1.1.- Ligases forming aminoacyl-tRNA e related compounds	
wCam_00747	Metabolismo de proteína	Biogênese proteica	Prolyl-tRNA sintetase associated editing enzymes	
wAlb_00351	Metabolismo de proteína	Biogênese proteica	Ribosome biogenesis bacterial	
wCam_00097	Metabolismo de proteína	Biogênese proteica	Ribosome biogenesis bacterial	
wCam_00377	Metabolismo de proteína	Biogênese proteica	Ribosome SSU mitochondrial	
wCam_00387	Metabolismo de proteína	Biogênese proteica	Ribosome SSU mitochondrial	
wAlb_00682	Metabolismo de proteína	Biogênese proteica	Ribospme LSU Symbiont	
wAlb_00958	Metabolismo de proteína	Biogênese proteica	Ribospme LSU Symbiont	
wCam_00213	Metabolismo de proteína	Biogênese proteica	Ribospme LSU Symbiont	
wCam_00370	Metabolismo de proteína	Biogênese proteica	Ribospme LSU Symbiont	

wCam_0048 5	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0048 6	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0049 3	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0049 4	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0051 4	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0051 5	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0062 2	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0062 3	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0062 9	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0063 0	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0063 1	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0063 2	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0072 8	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wPipMolestu s_00091	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wVitB00717	Metabolismo de proteína	Biogênese proteica	Single-copy ribosomal proteins
wCam_0022 7	Metabolismo de proteína	Biogênese proteica	Translation emittation factors eukaryotic e archaeal
wCam_0031 3	Metabolismo de proteína	Biogênese proteica	Translation emittation factors eukaryotic e archaeal
wCam_0014 0	Metabolismo de proteína	Biogênese proteica	tRNA aminoacylation
wCam_0054 4	Metabolismo de proteína	Biogênese proteica	tRNA aminoacylation
wCam_0072 5	Metabolismo de proteína	Biogênese proteica	tRNA aminoacylation
wCam_0012 1	Metabolismo de proteína	Biogênese proteica	Universal GTPases
wVuIC_0049 2	Metabolismo de proteína	Biogênese proteica	Universal GTPases
wCam_0028 9	Metabolismo de proteína	Degradação protéica	EC 3.4.11.- Aminopeptidases
wCam_0030 7	Metabolismo de proteína	Degradação protéica	EC 3.4.11.- Aminopeptidases
wCam_0063 4	Metabolismo de proteína	Degradação protéica	EC 3.4.11.- Aminopeptidases
wCam_0011 0	Metabolismo de proteína	Degradação protéica	EC 3.4.17.- Metallocoxypeptidases
wBolb_0063 8	Metabolismo de proteína	Degradação protéica	Proteasome eukaryotic
wAlb_01047	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wCam_0016 7	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wCam_0016 8	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wCam_0021 1	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wCam_0040 0	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent

wCam_0040 1	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wCam_0040 2	Metabolismo de proteína	Degradação protéica	Proteolysis em bacteria ATP-dependent
wAlb_00472	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wAu_01077	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wBolb_0015 8	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0010 3	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0012 5	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0036 8	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0072 6	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0073 5	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wNo_00831	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wNo_01046	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wNo_01047	Metabolismo de proteína	Enovelamento protéico	Protein chaperones
wCam_0066 1	Metabolismo de proteína	Processamento e modificação protéica	LipoBiogênese protéica
wCam_0073 6	Metabolismo de proteína	Processamento e modificação protéica	Signal peptidase
wCam_0002 6	Metabolismo de proteína	Selenoproteínas	Selenocysteine metabolism
wAu_00262	Metabolismo de RNA	Processamento e modificação de RNA	16S rRNA modification within P site of ribosome
wJHB_01362	Metabolismo de RNA	Processamento e modificação de RNA	16S rRNA modification within P site of ribosome
wCam_0063 5	Metabolismo de RNA	Processamento e modificação de RNA	ATP-dependent RNA helicases
wCam_0007 6	Metabolismo de RNA	Processamento e modificação de RNA	Methylthiotransferases
wCam_0080 7	Metabolismo de RNA	Processamento e modificação de RNA	Methylthiotransferases
wAlb_00797	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wCam_0005 2	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wCam_0038 9	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wCam_0053 5	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wCam_0055 8	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wJHB_00892	Metabolismo de RNA	Processamento e modificação de RNA	Queuosine exploration RZ
wCam_0061 2	Metabolismo de RNA	Processamento e modificação de RNA	rRNA modification bacteria
wCam_0018 8	Metabolismo de RNA	Processamento e modificação de RNA	Rtc
wCle_00882	Metabolismo de RNA	Processamento e modificação de RNA	tRNA mods Archaea
wCam_0000 8	Metabolismo de RNA	Processamento e modificação de RNA	Wyeosine-MimG biossíntese
wCam_0050 7	Metabolismo de RNA	Processamento e modificação de RNA	Wyeosine-MimG biossíntese
wCam_0057 1	Metabolismo de RNA	Processamento e modificação de RNA	Wyeosine-MimG biossíntese

wCam_0073 2	Metabolismo de RNA	Processamento e modificação de RNA	Wyeosine-MimG biossíntese
wCam_0064 3	Metabolismo de RNA	Transcrição	RNA polymerase bacterial
wCam_0062 8	Metabolismo de RNA	Transcrição	Transcrição elongation factors
wCam_0001 2	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wCam_0013 6	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wCam_0037 1	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wCam_0041 7	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wVitB00497	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wVitB00851	Metabolismo secundário	Metabolismo de nitrogênio	Ammonia assimilation
wCle_00694	Metabolismo secundário	Metabolismo de potássio	Glutathione-regulated potassium-efflux sistema e associated functions
wAlb_00343	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wCam_0003 0	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wCam_0027 6	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wCam_0036 1	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wCam_0049 0	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wCam_0060 3	N/A	N/A	Família 02 Ca ²⁺ P-Tipo ATPase
wAu_00039	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wAu_00438	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wAu_00534	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wAu_00863	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wAu_01285	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wCam_0045 7	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wJHB_00267	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wJHB_00271	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wJHB_00273	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wJHB_00274	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wMel_00513	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wNo_00226	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0013 7	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0020 5	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0032 8	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0063 8	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0065 7	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wVuIC_0165 9	N/A	N/A	Família 23 FUPA23 P-Tipo ATPase
wCam_0015 9	N/A	N/A	Família 24 FUPA24 P-Tipo ATPase ver.1
wAlb_00839	N/A	N/A	Família 30 FUPA30 P-Tipo ATPase
wAu_01161	N/A	N/A	Família 30 FUPA30 P-Tipo ATPase

wCam_00310	N/A	N/A	Familia 30 FUPA30 P-Tipo ATPase
wCam_00305	N/A	N/A	Heme e Siroheme biossíntese
wAlb_00004	N/A	N/A	N/A
wAlb_00032	N/A	N/A	N/A
wAlb_00315	N/A	N/A	N/A
wAlb_00349	N/A	N/A	N/A
wAlb_00403	N/A	N/A	N/A
wAlb_00419	N/A	N/A	N/A
wAlb_00446	N/A	N/A	N/A
wAlb_00581	N/A	N/A	N/A
wAlb_00695	N/A	N/A	N/A
wAlb_00898	N/A	N/A	N/A
wAlb_00903	N/A	N/A	N/A
wAlb_00990	N/A	N/A	N/A
wAlb_01001	N/A	N/A	N/A
wAlb_01012	N/A	N/A	N/A
wAu_00312	N/A	N/A	N/A
wAu_00313	N/A	N/A	N/A
wAu_00402	N/A	N/A	N/A
wAu_00480	N/A	N/A	N/A
wAu_00486	N/A	N/A	N/A
wAu_00521	N/A	N/A	N/A
wAu_00669	N/A	N/A	N/A
wAu_00720	N/A	N/A	N/A
wAu_00745	N/A	N/A	N/A
wAu_01113	N/A	N/A	N/A
wAu_01190	N/A	N/A	N/A
wAu_01286	N/A	N/A	N/A
wAu_01302	N/A	N/A	N/A
wBolb_00133	N/A	N/A	N/A
wBolb_00168	N/A	N/A	N/A
wBolb_00249	N/A	N/A	N/A
wBolb_00340	N/A	N/A	N/A
wBolb_00460	N/A	N/A	N/A
wBolb_00884	N/A	N/A	N/A
wBolb_01070	N/A	N/A	N/A
wCam_00029	N/A	N/A	N/A
wCam_00040	N/A	N/A	N/A
wCam_00090	N/A	N/A	N/A
wCam_00214	N/A	N/A	N/A
wCam_00230	N/A	N/A	N/A
wCam_0024	N/A	N/A	N/A

3

wCam_0024 7	N/A	N/A	N/A
wCam_0025 1	N/A	N/A	N/A
wCam_0028 6	N/A	N/A	N/A
wCam_0032 2	N/A	N/A	N/A
wCam_0034 8	N/A	N/A	N/A
wCam_0039 8	N/A	N/A	N/A
wCam_0040 5	N/A	N/A	N/A
wCam_0041 2	N/A	N/A	N/A
wCam_0041 4	N/A	N/A	N/A
wCam_0042 1	N/A	N/A	N/A
wCam_0044 6	N/A	N/A	N/A
wCam_0045 6	N/A	N/A	N/A
wCam_0047 2	N/A	N/A	N/A
wCam_0051 0	N/A	N/A	N/A
wCam_0051 3	N/A	N/A	N/A
wCam_0051 8	N/A	N/A	N/A
wCam_0052 1	N/A	N/A	N/A
wCam_0052 8	N/A	N/A	N/A
wCam_0056 1	N/A	N/A	N/A
wCam_0056 4	N/A	N/A	N/A
wCam_0058 5	N/A	N/A	N/A
wCam_0061 8	N/A	N/A	N/A
wCam_0064 6	N/A	N/A	N/A
wCam_0064 8	N/A	N/A	N/A
wCam_0069 6	N/A	N/A	N/A
wCam_0071 0	N/A	N/A	N/A
wCam_0077 7	N/A	N/A	N/A
wCam_0080 3	N/A	N/A	N/A
wCam_0080 8	N/A	N/A	N/A
wCle_00367	N/A	N/A	N/A
wCle_00667	N/A	N/A	N/A

wCle_00745	N/A	N/A	N/A
wCle_00898	N/A	N/A	N/A
wCle_01007	N/A	N/A	N/A
wHa_00373	N/A	N/A	N/A
wJHB_00423	N/A	N/A	N/A
wJHB_00520	N/A	N/A	N/A
wJHB_00522	N/A	N/A	N/A
wJHB_00924	N/A	N/A	N/A
wJHB_01044	N/A	N/A	N/A
wJHB_01349	N/A	N/A	N/A
wMel_01204	N/A	N/A	N/A
wNo_00401	N/A	N/A	N/A
wRi_00801	N/A	N/A	N/A
wTRS_00037	N/A	N/A	N/A
wTRS_00498	N/A	N/A	N/A
wTRS_00521	N/A	N/A	N/A
wTRS_00530	N/A	N/A	N/A
wTRS_00672	N/A	N/A	N/A
wTRS_00794	N/A	N/A	N/A
wVuIC_00649	N/A	N/A	N/A
wVuIC_01586	N/A	N/A	N/A
wAlb_00220	N/A	N/A	YggSProC
wAlb_00972	N/A	N/A	YggSProC
wBolb_00099	N/A	N/A	YggSProC
wCam_00034	N/A	N/A	YggSProC
wCam_00035	N/A	N/A	YggSProC
wCam_00036	N/A	N/A	YggSProC
wCam_00077	N/A	N/A	YggSProC
wCam_00087	N/A	N/A	YggSProC
wCam_00356	N/A	N/A	YggSProC
wCam_00408	N/A	N/A	YggSProC
wCam_00540	N/A	N/A	YggSProC
wCam_00748	N/A	N/A	YggSProC
wCam_00794	N/A	N/A	YggSProC
wCam_00827	N/A	N/A	YggSProC
wCam_00104	Nucleosídeos e nucleotídeos	Adenosyl nucleosidases	hydrolase
wCam_00440	Nucleosídeos e nucleotídeos	Pirimidinas	Biossíntese de pirimidinas em plantas

wCam_00240	Nucleosídeos e nucleotídeos	Purinas	CBSS-314260.3.peg.2133
wAlb_00911	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wBolb_00521	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00028	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00175	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00418	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00465	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00520	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00537	Nucleosídeos e nucleotídeos	Purinas	Conversão de purinas
wCam_00005	Nucleosídeos e nucleotídeos	Purinas	Biossíntese de purinas em plantas
wCam_00641	Nucleosídeos e nucleotídeos	Purinas	Biossíntese de purinas em plantas
wCam_00770	Nucleosídeos e nucleotídeos	Purinas	Biossíntese de purinas em plantas
wCam_00826	Parede e cápsula celular	Biossíntese de peptideoglicanos	Rare lipoprotein A precursor
wCam_00334	Parede e cápsula celular	Componentes de parede celular	KDO2-Lipid A biossíntese
wCam_00094	Parede e cápsula celular	Componentes de parede celular	Lipid A modifications
wCam_00638	Parede e cápsula celular	Componentes de parede celular	Lipopolysaccharide-related cluster em Alfaproteobacteria
wCam_00516	Parede e cápsula celular	Componentes de parede celular	Lipopolysaccharide assembly
wCam_00783	Parede e cápsula celular	Componentes de parede celular	Lipopolysaccharide assembly
wAlb_01038	Parede e cápsula celular	Componentes de parede celular	Perosamine síntese Vibrio
wAlb_01039	Parede e cápsula celular	Componentes de parede celular	Perosamine síntese Vibrio
wAu_01222	Parede e cápsula celular	Componentes de parede celular	Perosamine síntese Vibrio
wBolb_01030	Parede e cápsula celular	Componentes de parede celular	Perosamine síntese Vibrio
wCam_00447	Parede e cápsula celular	Componentes de parede celular	Perosamine síntese Vibrio
wAlb_00593	Parede e cápsula celular	Polissacarídeos capsulares e extracelulares	Capsular Polissacarídeos biossíntese e Assembly
wAlb_00592	Parede e cápsula celular	Polissacarídeos capsulares e extracelulares	Rhamnose containing glycans
wCam_00123	Parede e cápsula celular	Polissacarídeos capsulares e extracelulares	Xanthan Exopolysaccharide biossíntese e Export
wCam_00133	Parede e cápsula celular	Raj MurE	D-alanyl-D-Alanina carboxypeptidase (EC 3.4.16.4)
wAlb_00917	Parede e cápsula celular	Raj MurE	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)
wTRS_00330	Parede e cápsula celular	Raj MurE	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)
wAlb_00628	Parede e cápsula celular	Raj MurE	N-acetylmuramoyl-L-Alanina amidase (EC 3.5.1.28)
wAu_01026	Parede e cápsula celular	Raj MurE	N-acetylmuramoyl-L-Alanina amidase (EC 3.5.1.28)
wHa_01037	Parede e cápsula celular	Raj MurE	N-acetylmuramoyl-L-Alanina amidase (EC 3.5.1.28)

wAlb_00145	Parede e cápsula celular	UDP-N-acetylmuramate from Frutose-6-fosfate biossíntese	Glucosamine--Frutose-6-fosfate aminotransferase [isomerizing] (EC 2.6.1.16)
wCam_00182	Parede e cápsula celular	UDP-N-acetylmuramate from Frutose-6-fosfate biossíntese	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)
wCam_00381	Predictions based on plant-prokaryote comparative analysis	At2g23840	fosfolipase/carboxylesterase Família protein (EC 3.1.-.-)
wCam_00367	Predictions based on plant-prokaryote comparative analysis	At3g21300	Aspartyl-tRNA(Asn) amidotransferase subunidade B (EC 6.3.5.6)
wCam_00250	Predictions based on plant-prokaryote comparative analysis	At3g50560	Glutathione S-transferase Família protein
wCam_00519	Regulação e sinalização celular	Enzimas reguladas por zinco	C4-Tipo zinc finger protein
wCam_00784	Regulação e sinalização celular	Feromônios sexuais em <i>Enterococcus faecalis</i>	Membrane-associated zinc metalloprotease
wCam_00352	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Murein hydrolase regulation e cell death
wAlb_00312	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAlb_00867	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00109	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00110	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00111	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00271	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00272	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00296	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00297	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00445	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wAu_00657	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wCle_00442	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wCle_00498	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wCle_00499	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wCle_00500	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wVuIC_00869	Regulação e sinalização celular	Morte celular programada e sistemas toxina-antitoxina	Toxin-antitoxin replicon stabilization Sistemas
wCam_00308	Regulação transcricional	CarD copy	Transcrição termination factor Rho
wCam_00672	Respiração	Biogênese de citocromo C	ABC transporter envolved em Citocromo c biogenesis
wRi_00344	Respiração	Biogênese de citocromo C	ABC transporter envolved em Citocromo c biogenesis
wCam_00316	Respiração	Biogênese de citocromo C	Aconitate hidratase (EC 4.2.1.3)
wCam_00392	Respiração	Biogênese de citocromo C	Citrate synthase (si) (EC 2.3.3.1)
wCam_00220	Respiração	Biogênese de citocromo C	Citocromo c heme lyase subunidade CcmF

wCam_00304	Respiração	Biogênese de citocromo C	Citocromo c-Tipo biogenesis protein CcmC
wCam_00294	Respiração	Biogênese de citocromo C	Citocromo c-Tipo biogenesis protein CcmE
wCam_00171	Respiração	Biogênese de citocromo C	Ferrochelatase
wCam_00150	Respiração	Biogênese de citocromo C	Periplasmic divalent cation tolerance protein cutA
wCam_00253	Respiração	Biogênese de citocromo C	Periplasmic thiol:disulfide emterchange protein DsbA
wCam_00542	Respiração	Biogênese de citocromo C	Periplasmic thiol:disulfide oxidoreductase DsbB
wCam_00459	Respiração	Biogênese de citocromo oxidases	Citocromo oxidase biogenesis protein Cox11-CtaG
wCam_00391	Respiração	Biogênese de citocromo oxidases	Citocromo oxidase biogenesis protein Sco1/SenC/PrrC
wCam_00106	Respiração	Biogênese de citocromo oxidases	Citocromo oxidase biogenesis protein Surf1
wCam_00553	Respiração	Biogênese de citocromo oxidases	Heme A synthase
wAlb_00645	Respiração	Família de Quinona oxidoreduases	Quinone oxidoreductase (EC 1.6.5.5)
wAu_00739	Respiração	Reações eletro aceptoras	Citocromo c oxidases d@O copy
wAu_00740	Respiração	Reações eletro aceptoras	Citocromo c oxidases d@O copy
wCam_00595	Respiração	Reações eletro aceptoras	Terminal Citocromo C oxidases
wCam_00562	Respiração	Reações eletro doadoras	Respiratory Complex I
wAlb_00542	Respiração	Reações eletro doadoras	Respiratory dehydrogenases 1
wCam_00315	Respiração	Reações eletro doadoras	Respiratory dehydrogenases 1
wCam_00452	Respiração	Reações eletro doadoras	Succinate dehydrogenase
wCam_00453	Respiração	Reações eletro doadoras	Succinate dehydrogenase
wCam_00195	Resposta ao estresse	Estresse oxidativo	Glutaredoxins
wCam_00383	Resposta ao estresse	Estresse oxidativo	Glutaredoxins
wAu_00633	Resposta ao estresse	Estresse oxidativo	Glutathione analogs: mycothiol
wMel_00030	Resposta ao estresse	Estresse oxidativo	Glutathione analogs: mycothiol
wAlb_00300	Resposta ao estresse	Estresse oxidativo	Estresse oxidativo
wCam_00127	Resposta ao estresse	Estresse oxidativo	Estresse oxidativo
wCam_00506	Resposta ao estresse	Estresse oxidativo	Estresse oxidativo
wCam_00695	Resposta ao estresse	Estresse oxidativo	Estresse oxidativo
wCam_00708	Resposta ao estresse	Estresse oxidativo	Estresse oxidativo
wCam_00782	Resposta ao estresse	Estresse periplásmico	Periplasmic Stress Response
wCam_00734	Resposta ao estresse	Flavo-hemoglobina	Chaperone protein HscB
wCam_00382	Resposta ao estresse	Flavo-hemoglobina	Flavo-hemoproteína (Hemoglobina-like proteína) (Flavo-hemoglobina) (Nitric oxide dioxygenase) (EC 1.14.12.17)
wCam_00505	Subsistemas baseados em agrupamentos	a carboxipeptidase	a reductase
wCam_00306	Subsistemas baseados em agrupamentos	A degradation cluster	CBSS-89187.6.p.2182: a Degradation cluster

wAlb_00613	Subsistemas baseados em agrupamentos	Aerotolerance operon em Bacteroides e potentially orthologous operons em other orANisms	TPR domain protein em aerotolerance operon
wCam_00579	Subsistemas baseados em agrupamentos	Aerotolerance operon em Bacteroides e potentially orthologous operons em other orANisms	TPR domain protein em aerotolerance operon
wTRS_00737	Subsistemas baseados em agrupamentos	Aerotolerance operon em Bacteroides e potentially orthologous operons em other orANisms	TPR domain protein em aerotolerance operon
wCam_00320	Subsistemas baseados em agrupamentos	Aminoacetilação de tRNA e metabolismo de prolina	FIG107367: Single-stranded DNA-binding protein
wAlb_00285	Subsistemas baseados em agrupamentos	Aromatic conversions e predicted Co2 transporter cluster	Tyrosine recombinase XerD
wCam_00324	Subsistemas baseados em agrupamentos	biossíntese of galactoglycans e related lipopolysacharides	CBSS-258594.1.peg.3339
wCam_00432	Subsistemas baseados em agrupamentos	Catabolism of an unknown compound	CBSS-262316.1.peg.2929
wCam_00319	Subsistemas baseados em agrupamentos	CBSS-208964.1.peg.1012	4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7)
wAlb_00601	Subsistemas baseados em agrupamentos	CBSS-208964.1.peg.4832	FIG028883: Permeases of the drug/metabolite transporter (DMT) superFamília
wVitB00873	Subsistemas baseados em agrupamentos	CBSS-208964.1.peg.4832	FIG028883: Permeases of the drug/metabolite transporter (DMT) superFamília
wAlb_00712	Subsistemas baseados em agrupamentos	CBSS-211586.1.peg.3133	2-C-methyl-D-erythritol 4-fosfate cytidyltransferase (EC 2.7.7.60)
wCam_00206	Subsistemas baseados em agrupamentos	CBSS-211586.1.peg.3133	5-nucleotidase SurE (EC 3.1.3.5)
wAu_00379	Subsistemas baseados em agrupamentos	CBSS-211586.1.peg.3133	Proteína de divisão celular DivIC (FtsB)
wCam_00282	Subsistemas baseados em agrupamentos	CBSS-211586.1.peg.69	FIG032621: Hydrolase
wAlb_00305	Subsistemas baseados em agrupamentos	CBSS-214092.1.peg.3450	FIG137864: putative endonuclease containing a URI domain
wBolb_00314	Subsistemas baseados em agrupamentos	CBSS-214092.1.peg.3450	FIG137864: putative endonuclease containing a URI domain
wJHB_00979	Subsistemas baseados em agrupamentos	CBSS-214092.1.peg.3450	FIG137864: putative endonuclease containing a URI domain
wPipMol_01290	Subsistemas baseados em agrupamentos	CBSS-214092.1.peg.3450	FIG137864: putative endonuclease containing a URI domain
wAlb_00926	Subsistemas baseados em agrupamentos	CBSS-227321.4.peg.3751	Outer membrane receptor proteins
wAlb_00193	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	DNA polymerase III epsilon subunidade (EC 2.7.7.7)
wCam_00390	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	DNA polymerase III epsilon subunidade (EC 2.7.7.7)
wCitri_01061	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	DNA polymerase III epsilon subunidade (EC 2.7.7.7)
wBolb_00320	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	Ribonuclease HI (EC 3.1.26.4)
wCam_00117	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	Ribonuclease HI (EC 3.1.26.4)
wCitri_00208	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	Ribonuclease HI (EC 3.1.26.4)
wJHB_01371	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	Ribonuclease HI (EC 3.1.26.4)
wJHB_01372	Subsistemas baseados em agrupamentos	CBSS-228410.1.peg.134	Ribonuclease HI (EC 3.1.26.4)
wAlb_00934	Subsistemas baseados em agrupamentos	CBSS-272943.3.peg.1367	Rossmann fold nucleotide-binding protein Smf possibly involved em DNA uptake
wTRS_0080	Subsistemas baseados em agrupamentos	CBSS-272943.3.peg.1367	Rossmann fold nucleotide-binding protein Smf

5	em agrupamentos		possibly involved em DNA uptake
wCam_0054 1	Subsistemas baseados em agrupamentos	CBSS-290633.1.peg.1906	GTP-binding protein EngA
wCam_0024 6	Subsistemas baseados em agrupamentos	CBSS-292414.1.peg.563	RNA polymerase sigma factor RpoH
wJHB_01354	Subsistemas baseados em agrupamentos	CBSS-315749.4.peg.3658	Toprim domain protein
wCam_0011 4	Subsistemas baseados em agrupamentos	CBSS-316057.3.peg.563	Cytochrome c oxidase polypeptide II (EC 1.9.3.1)
wCam_0016 2	Subsistemas baseados em agrupamentos	CBSS-316057.3.peg.563	TldD protein
wCam_0058 4	Subsistemas baseados em agrupamentos	CBSS-349161.4.peg.2427	Glycyl-tRNA sintetase alfa chain (EC 6.1.1.14)
wCam_0058 3	Subsistemas baseados em agrupamentos	CBSS-349161.4.peg.2427	Glycyl-tRNA sintetase beta chain (EC 6.1.1.14)
wCam_0037 9	Subsistemas baseados em agrupamentos	CBSS-350688.3.peg.1509	DNA polymerase III alfa subunidade (EC 2.7.7.7)
wAlb_00780	Subsistemas baseados em agrupamentos	CBSS-350688.3.peg.1509	FIG007013: polysaccharide deacetylase
wCam_0031 2	Subsistemas baseados em agrupamentos	CBSS-350688.3.peg.1509	Ribosome-binding factor A
wCam_0031 4	Subsistemas baseados em agrupamentos	CBSS-350688.3.peg.1509	Transcrição termination protein NusA
wCam_0026 6	Subsistemas baseados em agrupamentos	CBSS-364106.7.peg.3204	Adenosine (5')-pentafofo-(5'')-adenosine pyrofofohydrolase (EC 3.6.1.-)
wCam_0075 8	Subsistemas baseados em agrupamentos	CBSS-36873.1.peg.6699	Parvulin-like peptidyl-prolyl isomerase
wCam_0069 3	Subsistemas baseados em agrupamentos	CBSS-393121.3.peg.1913	ATP-dependent DNA helicase UvrD/PcrA
wCam_0008 8	Subsistemas baseados em agrupamentos	CBSS-56780.10.peg.1536	Metal-dependent hydrolase YbeY
wCam_0080 0	Subsistemas baseados em agrupamentos	CBSS-783.1.peg.807	Preprotein translocase subunidade YajC (TC 3.A.5.1.1)
wCam_0018 0	Subsistemas baseados em agrupamentos	CBSS-783.1.peg.807	Protein-export membrane protein SecD (TC 3.A.5.1.1)
wCam_0022 2	Subsistemas baseados em agrupamentos	CBSS-783.1.peg.807	Protein-export membrane protein SecF (TC 3.A.5.1.1)
wAlb_00323	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wBolb_0067 5	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wCam_0010 7	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wCitri_00839	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wHa_00350	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wJHB_00610	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wVulC_0035 5	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Proteína de divisão celular FtsH (EC 3.4.24.-)
wCam_0065 3	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	Peptidyl-tRNA hydrolase (EC 3.1.1.29)
wCam_0078 6	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	SSU ribosomal protein S1p
wCam_0010 8	Subsistemas baseados em agrupamentos	Cell division-ribosomal stress proteins cluster	tRNA(Ile)-lysidine sintetase (EC 6.3.4.19)
wCam_0022 3	Subsistemas baseados em agrupamentos	CTP synthase (EC 6.3.4.2) cluster	Inner membrane protein translocase component YidC
wCam_0035 8	Subsistemas baseados em agrupamentos	Divisão celular bacteriana	Proteína de divisão celular BolA
wCam_0040	Subsistemas baseados	Divisão celular bacteriana	Fator de divisão celular (EC 5.2.1.8)

3	em agrupamentos		
wCam_00823	Subsistemas baseados em agrupamentos	Divisão celular bacteriana	rRNA small subunidade methyltransferase H
wCam_00756	Subsistemas baseados em agrupamentos	DNA gyrase subunidades	DNA gyrase subunidade A (EC 5.99.1.3)
wCam_00592	Subsistemas baseados em agrupamentos	DNA gyrase subunidades	DNA gyrase subunidade B (EC 5.99.1.3)
wCam_00769	Subsistemas baseados em agrupamentos	EC49-61	ApaG protein
wCam_00190	Subsistemas baseados em agrupamentos	Exportação protéica	CBSS-393121.3.peg.2760
wCam_00543	Subsistemas baseados em agrupamentos	Exportação protéica	CBSS-393121.3.peg.2760
wCam_00701	Subsistemas baseados em agrupamentos	Exportação protéica	CBSS-393121.3.peg.2760
wCam_00061	Subsistemas baseados em agrupamentos	Fatty acid metabolic cluster	COG1399
wCam_00340	Subsistemas baseados em agrupamentos	Fatty acid metabolic cluster	COG1399
wCam_00369	Subsistemas baseados em agrupamentos	Fatty acid metabolic cluster	COG1399
wCam_00388	Subsistemas baseados em agrupamentos	Glutaredoxin 3 containing cluster	Protein export cytoplasm chaperone protein (SecB)
wAlb_00465	Subsistemas baseados em agrupamentos	Grupo de proteínas hipotéticas de proteobactérias	CBSS-52598.3.peg.2843
wCam_00050	Subsistemas baseados em agrupamentos	Grupo de proteínas hipotéticas de proteobactérias	CBSS-52598.3.peg.2843
wCle_01172	Subsistemas baseados em agrupamentos	Grupo de proteínas hipotéticas de proteobactérias	CBSS-52598.3.peg.2843
wAlb_00184	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase0
wAu_00495	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase1
wHa_00358	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase2
wRi_00289	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase3
wBolb_00103	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase
wVuIC_00441	Subsistemas baseados em agrupamentos	Grupo relacionado a parede celular	FIG022606: AAA ATPase
wCam_00033	Subsistemas baseados em agrupamentos	heat shock	cell division
wCam_00793	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	Proteína de divisão celular FtsK
wCam_00680	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	Metallo-beta-lactamase Família protein
wCam_00539	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	Peptide deformylase (EC 3.5.1.88)
wCam_00277	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)
wCam_00706	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	TsaB protein
wCam_00323	Subsistemas baseados em agrupamentos	Hidrolases Zn-dependentes	TsaD/Kae1/Qri7 protein
wCam_00062	Subsistemas baseados em agrupamentos	hipotético Related to Dihydroorotate Dehydrogenase	hipotético Related to Dihydroorotate dehydrogenase
wBolb_00459	Subsistemas baseados em agrupamentos	Pode estar relacionado a acetilação NAD-dependente	CBSS-216591.1.peg.168
wCam_00037	Subsistemas baseados em agrupamentos	Pode estar relacionado a acetilação NAD-dependente	CBSS-216591.1.peg.168
wCle_00954	Subsistemas baseados em agrupamentos	Pode estar relacionado a	CBSS-216591.1.peg.168

	em agrupamentos	acetilação NAD-dependente	
wJHB_00184	Subsistemas baseados em agrupamentos	Pode estar relacionado a acetilação NAD-dependente	CBSS-216591.1.peg.168
wAlb_00168	Subsistemas baseados em agrupamentos	Proteína de checkagem de integridade do DNA	Reparo de DNA protein RadA
wTRS_01088	Subsistemas baseados em agrupamentos	Proteína de checkagem de integridade do DNA	Reparo de DNA protein RadA
wTRS_01089	Subsistemas baseados em agrupamentos	Proteína de checkagem de integridade do DNA	Reparo de DNA protein RadA
wCam_00014	Subsistemas baseados em agrupamentos	Putative hemin transporter	Hemolysins e related proteins containing CBS domains
wCam_00416	Subsistemas baseados em agrupamentos	Replicação de DNA cluster 1	Iniciador de replicação cromossomal
wCam_00791	Subsistemas baseados em agrupamentos	Replicação de DNA cluster 1	DNA polymerase III beta subunidade (EC 2.7.7.7)
wCam_00350	Subsistemas baseados em agrupamentos	Ribosome-related cluster	A Gammaproteobacteria Cluster Relating to Translation
wCam_00759	Subsistemas baseados em agrupamentos	RNA modification e Cromossomopartitioning cluster	GTPase e tRNA-U34 5-formylation enzyme TrmE
wCam_00288	Subsistemas baseados em agrupamentos	RNA modification e Cromossomopartitioning cluster	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
wCam_00517	Subsistemas baseados em agrupamentos	RP Divisão celular bacteriana	Heat shock protein FtsJ/RrmJ
wCam_00296	Subsistemas baseados em agrupamentos	Subsistemas baseados em agrupamentos	CBSS-280355.3.peg.2835
wCam_00345	Subsistemas baseados em agrupamentos	Subsistemas baseados em agrupamentos	CBSS-280355.3.peg.2835
wCam_00423	Subsistemas baseados em agrupamentos	Subsistemas baseados em agrupamentos	CBSS-280355.3.peg.2835
wAlb_00683	Subsistemas baseados em agrupamentos	USS-DB-6	ABC-Tipo transport sistema emolved em resistance to orANlc solvents
wBolb_01112	Subsistemas experimentais	(Bacterio)rhodopsin	Oxidoreductase
wAlb_00227	Subsistemas experimentais	271-Bsub	1-hydroxy-2-methyl-2-(E)-butenyl 4-difosfate synthase (EC 1.17.7.1)
wAlb_00221	Subsistemas experimentais	271-Bsub	Biotina-protein ligase (EC 6.3.4.15) / Biotina operon repressor
wTRS_00953	Subsistemas experimentais	271-Bsub	Biotina-protein ligase (EC 6.3.4.15) / Biotina operon repressor
wCam_00798	Subsistemas experimentais	271-Bsub	DNA polymerase III delta prime subunidade (EC 2.7.7.7)
wCam_00765	Subsistemas experimentais	271-Bsub	DNA topoisomerase I (EC 5.99.1.2)
wCam_00462	Subsistemas experimentais	271-Bsub	DNA-directed RNA polymerase alfa subunidade (EC 2.7.7.6)
wCam_00082	Subsistemas experimentais	271-Bsub	GTP-binding protein EngB
wCam_00147	Subsistemas experimentais	271-Bsub	Heat shock protein 60 Família chaperone GroEL
wCam_00146	Subsistemas experimentais	271-Bsub	Heat shock protein 60 Família co-chaperone GroES
wCam_00492	Subsistemas experimentais	271-Bsub	Methionyl-tRNA formyltransferase (EC 2.1.2.9)
wCam_00627	Subsistemas experimentais	271-Bsub	Preprotein translocase subunidade SecE (TC 3.A.5.1.1)
wCam_00341	Subsistemas experimentais	271-Bsub	Pyruvate dehydrogenase E1 component alfa subunidade (EC 1.2.4.1)
wCam_00138	Subsistemas experimentais	271-Bsub	Replicative DNA helicase (EC 3.6.1.-) [SA14-24]
wAlb_00046	Subsistemas experimentais	271-Bsub	Ribonuclease P protein component (EC 3.1.26.5)
wCam_0057	Subsistemas	271-Bsub	Ribonuclease P protein component (EC

7	experimentais		3.1.26.5)
wTRS_00380	Subsistemas experimentais	271-Bsub	Ribonuclease P protein component (EC 3.1.26.5)
wCam_00439	Subsistemas experimentais	271-Bsub	Ribosome recycling factor
wCam_00731	Subsistemas experimentais	271-Bsub	Signal recognition particle receptor protein FtsY (=alfa subunidade) (TC 3.A.5.1.1)
wCam_00441	Subsistemas experimentais	271-Bsub	Translation elongation factor Ts
wCam_00676	Subsistemas experimentais	271-Bsub	Translation emittiation factor 3
wCam_00709	Subsistemas experimentais	At3g61080 homolog	Mitochondrial processing peptidase-like protein (EC 3.4.24.64)
wCam_00154	Subsistemas experimentais	At4g19540 (COG0489)	HflC protein
wCam_00153	Subsistemas experimentais	At4g19540 (COG0489)	HflK protein
wCam_00155	Subsistemas experimentais	At4g19540 (COG0489)	HtrA protease/chaperone protein
wAlb_00496	Subsistemas experimentais	Biofilme ketran	exopolysaccharide síntese protein ExoD-related protein
wCam_00566	Subsistemas experimentais	Biofilme ketran	exopolysaccharide síntese protein ExoD-related protein
wCam_00805	Subsistemas experimentais	COG3236-small-temp	Ribose-fosfate pyrofosfokinase (EC 2.7.6.1)
wCam_00411	Subsistemas experimentais	CRnot mitoch clusters Euk 1	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)
wTRS_00527	Subsistemas experimentais	CRnot mitoch clusters Euk 1	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)
wCam_00509	Subsistemas experimentais	CRnot mitoch clusters Euk 1	Dihydrolipoamide dehydrogenase of 2-oxoglutarate dehydrogenase (EC 1.8.1.4)
wAlb_00256	Subsistemas experimentais	CRnot mitoch clusters Euk 1	Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)
wCam_00702	Subsistemas experimentais	CRnot mitoch clusters Euk 1	Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)
wCam_00268	Subsistemas experimentais	CRnot mitoch clusters Euk 1	Malate dehydrogenase (EC 1.1.1.37)
wCam_00120	Subsistemas experimentais	CRnot mitoch clusters Euk 1	NAD-dependent glyceraldehyde-3-fosfate dehydrogenase (EC 1.2.1.12)
wCam_00089	Subsistemas experimentais	CRnot mitoch clusters Euk 1	Succinate dehydrogenase iron-sulfur protein (EC 1.3.99.1)
wCam_00354	Subsistemas experimentais	Deaminase	tRNA-specific adenosine-34 deaminase (EC 3.5.4.-)
wCam_00380	Subsistemas experimentais	DUF1537-pdxA2	Nucleoside-difosfate-sugar epimerases
wCam_00118	Subsistemas experimentais	DUF490	Large exoproteins envolved em heme utilização or adhesion
wCam_00284	Subsistemas experimentais	DUF89CLO	Lipoprotein releasing sistema ATP-binding protein LolD
wCam_00450	Subsistemas experimentais	DUF89CLO	Lipoprotein releasing sistema transmembrane protein LolE
wCam_00160	Subsistemas experimentais	EC 4.2.1.51-75 Hydro-lyases	Uroporfyrinogen-III synthase (EC 4.2.1.75)
wAlb_00569	Subsistemas experimentais	EC 6.3.2.- Acid--D-amino-acid ligases (peptide synthases)	fosforibosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)
wAlb_00947	Subsistemas experimentais	Experimental-COG3533	Arabinose-proton symporter
wAlb_00614	Subsistemas experimentais	Experimental-DCE degrader	probable iron-sulfur binding protein YPO1417
wCam_0022	Subsistemas	Experimental-EFP	Transcrição termination protein NusB

6	experimentais		
wAlb_00218	Subsistemas experimentais	Ferro-enxofre	Bacterioferritin-associated ferredoxin
wCam_00682	Subsistemas experimentais	Ferro-enxofre	experimental
wCam_00616	Subsistemas experimentais	Ferro-enxofre	experimental
wCam_00285	Subsistemas experimentais	Ferro-enxofre	probable iron binding protein from the HesB IscA SufA Família
wCam_00602	Subsistemas experimentais	Frankia C1 regulator	IMP cyclohydrolase (EC 3.5.4.10) / fosforibosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)
wCam_00064	Subsistemas experimentais	Frutose-bisfosfate aldolase protein Família	Frutose-bisfosfate aldolase class I (EC 4.1.2.13)
wCam_00215	Subsistemas experimentais	Iron-sulfur cluster assembly	Believed to be emvolved em assembly of Fe-S clusters
wAlb_00012	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00021	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00180	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00438	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00706	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00724	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00882	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00883	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00887	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00916	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_00921	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAlb_01013	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00089	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00179	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00317	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00566	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00630	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00631	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00632	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00637	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00638	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wAu_00752	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00013	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin

wBolb_00022	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00048	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00065	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00325	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00346	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00498	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00507	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00593	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00713	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00714	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00756	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00827	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wBolb_00972	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCam_00522	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCitri_00461	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCitri_01134	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00065	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00309	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00699	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00788	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00851	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00921	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_00938	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_01248	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCle_01373	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wHa_00306	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wHa_00307	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wHa_00526	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wHa_00629	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wJHB_00620	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wJHB_00801	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin

wJHB_01268	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wMel_00296	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wNo_00356	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wNo_00983	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPip_00097	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPip_00368	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPip_00813	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPip_00816	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPipMolestus_01088	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wPipMolestus_01089	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wTRS_00407	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wTRS_00419	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wTRS_00798	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wTRS_00800	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVitB00406	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVitB00827	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_00110	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_01419	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_01480	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_01487	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_01503	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wVulC_01584	Subsistemas experimentais	Isomerase Glucose P Euk7	Repetições Ankyrin
wCam_00749	Subsistemas experimentais	Listeria autolisinas	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)
wCam_00228	Subsistemas experimentais	Maf	Septum formation protein Maf
wCam_00131	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00424	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00466	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00633	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00658	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00778	Subsistemas experimentais	Metagenômica	EGS prediction
wNo_01163	Subsistemas experimentais	Metagenômica	EGS prediction

wRi_00729	Subsistemas experimentais	Metagenômica	EGS prediction
wCam_00660	Subsistemas experimentais	MiaA RpoS	tmRNA-binding protein SmpB
wCam_00598	Subsistemas experimentais	MiaA RpoS	tRNA dimethylallyltransferase (EC 2.5.1.75)
wCam_00816	Subsistemas experimentais	MiaB-association	Aspartate carbamoyltransferase (EC 2.1.3.2)
wAlb_00341	Subsistemas experimentais	MiaB-association	Glutathione sintetase (EC 6.3.2.3)
wCam_00581	Subsistemas experimentais	MiaB-association	Putative Holliday junction resolvase YqgF
wAlb_00704	Subsistemas experimentais	MiaB-association	Ribosomal RNA small subunidade methyltransferase E (EC 2.1.1.-)
wCam_00332	Subsistemas experimentais	Modelo B	Ubiquinone/menaquinone biossíntese methyltransferase UbiE (EC 2.1.1.-)
wCam_00022	Subsistemas experimentais	Modificação de tRNA em Escherichia coli	tRNA dihydrouridine synthase B (EC 1.-.-.-)
wCam_00790	Subsistemas experimentais	Modificação de tRNA em Escherichia coli	tRNA pseudouridine synthase A (EC 4.2.1.70)
wCam_00059	Subsistemas experimentais	Modificationexploration	Aspartyl-tRNA(Asn) amidotransferase subunidade A (EC 6.3.5.6)
wCam_00806	Subsistemas experimentais	Modificationexploration	Aspartyl-tRNA(Asn) amidotransferase subunidade A (EC 6.3.5.6)
wCam_00640	Subsistemas experimentais	MutH-MutL	DNA mismatch repair protein MutL
wCam_00091	Subsistemas experimentais	N-acetyltransferase	Acetyltransferase
wHa_00713	Subsistemas experimentais	N-acetyltransferase	Acetyltransferase
wCam_00357	Subsistemas experimentais	N/A	N/A
wCam_00792	Subsistemas experimentais	N/A	N/A
wCam_00819	Subsistemas experimentais	NADP(H) fosfatase	NAD kinase (EC 2.7.1.23)
wCam_00724	Subsistemas experimentais	Peroxiredoxina	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)
wVitB00424	Subsistemas experimentais	Peroxiredoxina	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)
wAlb_00338	Subsistemas experimentais	Peroxiredoxina	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)
wCam_00095	Subsistemas experimentais	Peroxiredoxina	Thiol peroxidase
wCam_00218	Subsistemas experimentais	Peroxiredoxina	Transketolase (EC 2.2.1.1)
wCam_00605	Subsistemas experimentais	plasmídeo related	Inner membrane protein forms channel for Tipo IV secretion of T-DNA complex (VirB3)
wAlb_00525	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wAlb_00646	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wAu_00707	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCam_00258	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCam_00437	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCam_00615	Subsistemas experimentais	Possíveis alvos	Possíveis alvos

wCam_00620	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCam_00727	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCle_00034	Subsistemas experimentais	Possíveis alvos	Possíveis alvos
wCam_00763	Subsistemas experimentais	Potassium homeostasis Pavel	Glutathione-regulated potassium-efflux sistema protein KefKL
wAu_00316	Subsistemas experimentais	PTS Sistemas em Bacteroides	Site-specific recombinase Mpi
wCam_00272	Subsistemas experimentais	Pyrimidines	pyrimidine conversions v2
wCam_00353	Subsistemas experimentais	Pyrimidines	pyrimidine conversions v2
wCam_00761	Subsistemas experimentais	Pyrimidines	pyrimidine conversions v2
wCam_00164	Subsistemas experimentais	Replicação de DNA	archaeal Isaac
wCam_00795	Subsistemas experimentais	Replicação de DNA	archaeal Isaac
wAu_00384	Subsistemas experimentais	Resistência a antibióticos e compostos tóxicos	Acriflavin resistance cluster
wAlb_00509	Subsistemas experimentais	rRNA methylation em clusters	Integrase/recombinase xerD
wCam_00267	Subsistemas experimentais	rRNA methylation em clusters	Putative deoxyribonuclease YcfH
wCam_00098	Subsistemas experimentais	rRNA methylation em clusters	Translation elongation factor P
wCam_00142	Subsistemas experimentais	SA hipotético	DNA primase (EC 2.7.7.-)
wCam_00802	Subsistemas experimentais	SA hipotético	GTP-binding protein Era
wCam_00429	Subsistemas experimentais	SA hipotético	Pyruvate
wCam_00649	Subsistemas experimentais	SA hipotético	RNA polymerase sigma factor RpoD
wAlb_00622	Subsistemas experimentais	Small Multidrug Resistance	Quaternary ammonium compound-resistance protein SugE
wAlb_00615	Subsistemas experimentais	Test - Biotina	Substrate-specific component BioY of Biotina ECF transporter
wCle_00300	Subsistemas experimentais	Thiamin biossíntese Human non-gut microbiome	Hydroxymethylpyrimidine fosfate kinase ThiD (EC 2.7.4.7)
wAu_00487	Subsistemas experimentais	Transportadores em Modelos	amino acid ABC transporter
wCam_00330	Subsistemas experimentais	Transportadores em Modelos	ATP synthase alfa chain (EC 3.6.3.14)
wCam_00551	Subsistemas experimentais	Transportadores em Modelos	ATP synthase B' chain (EC 3.6.3.14)
wCam_00575	Subsistemas experimentais	Transportadores em Modelos	ATP synthase beta chain (EC 3.6.3.14)
wCam_00331	Subsistemas experimentais	Transportadores em Modelos	ATP synthase delta chain (EC 3.6.3.14)
wCam_00574	Subsistemas experimentais	Transportadores em Modelos	ATP synthase epsilon chain (EC 3.6.3.14)
wCam_00548	Subsistemas experimentais	Transportadores em Modelos	ATP synthase F0 sector subunidade a
wCam_00550	Subsistemas experimentais	Transportadores em Modelos	ATP synthase F0 sector subunidade b
wCam_00549	Subsistemas experimentais	Transportadores em Modelos	ATP synthase F0 sector subunidade c
wCam_00639	Subsistemas experimentais	Transportadores em Modelos	ATP synthase gamma chain (EC 3.6.3.14)

wCam_00115	Subsistemas experimentais	Transportadores em Modelos	Citocromo c oxidase polypeptide I (EC 1.9.3.1)
wCam_00359	Subsistemas experimentais	Transportadores em Modelos	Citocromo c oxidase polypeptide III (EC 1.9.3.1)
wAlb_00385	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wAlb_00752	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wAu_01095	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wCam_00018	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wCam_00355	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wCam_00675	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wCle_00478	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wRi_01137	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wVulC_01283	Subsistemas experimentais	Transportadores em Modelos	L-Prolina/Glycine betaine transporter ProP
wAlb_00684	Subsistemas experimentais	Transportadores em Modelos	Methionine ABC transporter ATP-binding protein
wAu_00340	Subsistemas experimentais	Transportadores em Modelos	Methionine transporter MetT
wCam_00338	Subsistemas experimentais	Transportadores em Modelos	Methionine transporter MetT
wTRS_00294	Subsistemas experimentais	Transportadores em Modelos	Methionine transporter MetT
wMel_00508	Subsistemas experimentais	Transportadores em Modelos	Multimodular transpeptidase-transglycosylase (EC 2.4.1.129) (EC 3.4.-.-)
wCam_00645	Subsistemas experimentais	Transportadores em Modelos	Na(+) H(+) antiporter subunidade G (TC 2.A.63.1.2)
wCam_00271	Subsistemas experimentais	Transportadores em Modelos	NADH ubiquinone oxidoreductase chain A (EC 1.6.5.3)
wCam_00270	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain B (EC 1.6.5.3)
wCam_00269	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain C (EC 1.6.5.3)
wCam_00663	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain D (EC 1.6.5.3)
wCam_00051	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain E (EC 1.6.5.3)
wCam_00681	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain F (EC 1.6.5.3)
wCam_00534	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain G (EC 1.6.5.3)
wCam_00533	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain H (EC 1.6.5.3)
wCam_00678	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain I (EC 1.6.5.3)
wCam_00301	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain J (EC 1.6.5.3)
wCam_00300	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain K (EC 1.6.5.3)
wCam_00299	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain L (EC 1.6.5.3)
wCam_00298	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain M (EC 1.6.5.3)
wCam_00297	Subsistemas experimentais	Transportadores em Modelos	NADH-ubiquinone oxidoreductase chain N (EC 1.6.5.3)

wCam_00460	Subsistemas experimentais	Transportadores em Modelos	fosfate ABC transporter
wCam_00344	Subsistemas experimentais	Transportadores em Modelos	fosfate transport ATP-binding protein pstB (TC 3.A.1.7.1)
wCam_00006	Subsistemas experimentais	Transportadores em Modelos	fosfate transport sistema permease protein pstA (TC 3.A.1.7.1)
wCam_00576	Subsistemas experimentais	Transportadores em Modelos	fosfate transport sistema permease protein pstC (TC 3.A.1.7.1)
wAlb_00154	Subsistemas experimentais	Transportadores em Modelos	Potassium uptake protein TrkH
wMel_00381	Subsistemas experimentais	Transportadores em Modelos	Potassium uptake protein TrkH
wHa_00028	Subsistemas experimentais	Transportadores em Modelos	Sodium-dependent fosfate transporter
wAlb_00018	Subsistemas experimentais	Transportadores em Modelos	sodium/Alanina symporter Família protein
wCam_00045	Subsistemas experimentais	Transportadores em Modelos	sodium/Alanina symporter Família protein
wCam_00229	Subsistemas experimentais	Transportadores em Modelos	sodium/Alanina symporter Família protein
wCam_00071	Subsistemas experimentais	Transportadores em Modelos	Succinate dehydrogenase flavoprotein subunidade (EC 1.3.99.1)
wCam_00198	Subsistemas experimentais	Transportadores em Modelos	ubiquinol Citocromo C oxidoreductase
wCam_00197	Subsistemas experimentais	Transportadores em Modelos	Ubiquinol--Citocromo c reductase
wCam_000757	Subsistemas experimentais	Transportadores em Modelos	Ubiquinol-Citocromo C reductase iron-sulfur subunidade (EC 1.10.2.2)
wCam_00176	Subsistemas experimentais	Transportadores em Modelos	Zinc ABC transporter
wCam_00177	Subsistemas experimentais	Transportadores em Modelos	Zinc ABC transporter
wCam_000811	Subsistemas experimentais	Transportadores em Modelos	Zinc transporter zitB
wAu_00230	Subsistemas experimentais	UniLU 2015 N07	Ribulose-5-fosfate 4-epimerase e related epimerases e aldolases
wCam_00079	Subsistemas experimentais	uridine kinase cluster 1	Alanyl-tRNA sintetase (EC 6.1.1.7)
wCam_000135	Subsistemas experimentais	uridine kinase cluster 1	Deoxycytidine trifosfate deaminase (EC 3.5.4.13)
wCam_000329	Subsistemas experimentais	uridine kinase cluster 1	hipotético protein perhaps functionally coupled to Transcrição elongation factor GreA
wCam_000444	Subsistemas experimentais	yrdC	Carbamoyl-fosfate synthase large chain (EC 6.3.5.5)
wCam_000081	Subsistemas experimentais	yrdC	Peptide chain release factor 1
wVulC_000443	Subsistemas experimentais	yrdC	Peptide chain release factor 1
wCam_000122	Subsistemas experimentais	yrdC	Peptide chain release factor 2
wCam_000219	Subsistemas experimentais	yrdC	Protein-N(5)-Glutamina methyltransferase PrmC
wCam_000799	Subsistemas experimentais	yrdC	TsaC protein (YrdC domain) required for threonylcarbamoyladenossine t(6)A37 modification em tRNA
wCam_000044	Subsistemas experimentais	ZZ gjo scratch	Acyl-fosfate:glycerol-3-fosfate O-acyltransferase PlsY
wCam_000016	Subsistemas experimentais	ZZ gjo scratch	Aspartate aminotransferase (EC 2.6.1.1)
wCam_000292	Subsistemas experimentais	ZZ gjo scratch	fosfate:acyl-ACP acyltransferase PlsX
wCam_00075	Subsistemas		DihydroFolato reductases

0	experimentais		
wAlb_00319	Transporte de membrana	Sistema de secreção de proteínas	Tipo II
wBolb_00160	Transporte de membrana	Sistema de secreção de proteínas	Tipo IV
wCam_00238	Transporte de membrana	Sistema de secreção de proteínas	Tipo VI
wAlb_00006	Transporte de membrana	Sistemas de transporte Ton e Tol	Putative TolA protein
wHa_00513	Transporte de membrana	Sistemas de transporte Ton e Tol	Putative TolA protein
wAlb_00205	Transporte de membrana	Sistemas de transporte Ton e Tol	tolB protein precursor
wTRS_00654	Transporte de membrana	Sistemas de transporte Ton e Tol	tolB protein precursor
wCam_00119	Transporte de membrana	Translocação de proteínas através da membrana citoplasmática	Twin-Arginina translocation sistema
wCam_00565	Transporte de membrana	Translocação de proteínas através da membrana citoplasmática	Twin-Arginina translocation sistema
wCam_00698	Transporte de membrana	Transportadores ABC	ABC transporter alkylfosfonate (TC 3.A.1.9.1)
wCam_00053	Transporte de membrana	Transportadores plastidiais caracterizados em plantas	Plastidic sodium-dependent pyruvate transporter
wMel_00421	Transporte de membrana	Uni- Sym e Antiporters	Proton-dependent Peptide Transportadores
wAu_00457	Transporte de membrana	Uni-Sym e Antiporters	Proton-dependent Peptide Transportadores
wCam_00027	Variados	EC 6.3.4.- Ligases that form carbon-nitrogen bonds	fosforibosylamine--glycine ligase (EC 6.3.4.13)
wCam_00004	Variados	Genômica comparativa de Planta-Procaríoto	At1g69340 At2g40600
wCam_00196	Variados	Genômica comparativa de Planta-Procaríoto	At1g69340 At2g40600
wAlb_00402	Variados	Genômica comparativa de Planta-Procaríoto	At4g17650 (YfjG)
wCam_00212	Variados	Genômica comparativa de Planta-Procaríoto	At4g17650 (YfjG)
wAlb_00232	Variados	Genômica comparativa de Planta-Procaríoto	At5g38900
wCam_00419	Variados	Genômica comparativa de Planta-Procaríoto	At5g38900
wCle_00066	Variados	Genômica comparativa de Planta-Procaríoto	At5g38900
wAlb_00182	Variados	Genômica comparativa de Planta-Procaríoto	Competence protein F homolog
wMelPop_00936	Variados	Genômica comparativa de Planta-Procaríoto	Competence protein F homolog
wCam_00179	Variados	Genômica comparativa de Planta-Procaríoto	lojapClusters
wCam_00436	Variados	Genômica comparativa de Planta-Procaríoto	lojapClusters
wCam_00526	Variados	Genômica comparativa de Planta-Procaríoto	lojapClusters
wCam_00829	Variados	Genômica comparativa de Planta-Procaríoto	lojapClusters
wAlb_00272	Variados	Genômica comparativa de Planta-Procaríoto	ligT
wAlb_00790	Variados	Genômica comparativa de Planta-Procaríoto	ligT
wCam_00092	Variados	Genômica comparativa de Planta-Procaríoto	ligT
wCam_0019	Variados	Genômica comparativa de	ligT

3		Planta-Procaríoto	
wCam_00384	Variados	Genômica comparativa de Planta-Procaríoto	Omega-amidase KE2
wCam_00559	Variados	Genômica comparativa de Planta-Procaríoto	Omega-amidase KE2
wAlb_00039	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wAlb_00299	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wAlb_00591	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wAlb_00741	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wAlb_00838	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wCam_00259	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wCam_00667	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wCam_00771	Variados	Genômica comparativa de Planta-Procaríoto	PROSC
wAlb_00051	Variados	Genômica comparativa de Planta-Procaríoto	YebC
wAlb_00269	Variados	Genômica comparativa de Planta-Procaríoto	YebC
wAlb_00270	Variados	Genômica comparativa de Planta-Procaríoto	YebC
wCam_00767	Variados	Genômica comparativa de Planta-Procaríoto	YgfZ-Fe-S
wCam_00084	Variados	Reparo de metabolitos	Deoxyuridine 5'-trifosfato nucleotidohidrolase (EC 3.6.1.23)
wCam_00207	Variados	ZZ gjo need homes	Pole remodelling regulatory diguanylate cyclase
wAlb_00823	Variados	ZZ gjo need homes	Proposed Peptidoglicano lipid II flippase MurJ
wCam_00399	Variados	ZZ gjo need homes	Putative permease often clustered with de novo purine síntese
wJHB_00268	Variados	ZZ gjo need homes	Putative permease often clustered with de novo purine síntese
wAlb_00104	Virulência, doença e defesa	Invasão e resistência intracelular	Cytolysin e Lipase operon em Vibrio
wHa_01096	Virulência, doença e defesa	Invasão e resistência intracelular	Listeria surface proteins: emternalín-like proteins
wCam_00624	Virulência, doença e defesa	Operon de virulência envolvido em síntese protéica	Translation elongation factor G
wBolb_00549	Virulência, doença e defesa	Operon de virulência envolvido em síntese protéica	Translation elongation factor Tu
wCam_00487	Virulência, doença e defesa	Operon de virulência envolvido em síntese protéica	Translation elongation factor Tu
wCam_00741	Virulência, doença e defesa	Resistência a antibióticos e compostos tóxicos	Copper homeostasis: copper tolerance
wAlb_00999	Virulência, doença e defesa	Resistência a antibióticos e compostos tóxicos	Erythromycin resistance
wAu_00062	Virulência, doença e defesa	Resistência a antibióticos e compostos tóxicos	MRSA
wCam_00274	Virulência, doença e defesa	Resistência a antibióticos e compostos tóxicos	MRSA
wAlb_00697	Virulência, doença e defesa	Resistência a antibióticos e compostos tóxicos	Mycobacterial MmpL3 membrane protein cluster
wCam_00235	Virulência, doença e defesa	Tipo III	Tipo IV
wCam_00796	Virulência, doença e defesa	Tipo III	Tipo IV

		Tipo III	Tipo IV
wVitB00960	Virulência, doença e defesa		
wJHB_00476	Virulência, doença e defesa	Virulência, doença e defesa	Operon de virulencia envolvido em síntese protéica
wCam_00232	Virulência, doença e defesa	Viruloma de Legionella pneumofila	ATPase provides energy for both assembly ofTipo IV secretion complex e secretion of T-DNA complex (VirB11)
wCam_00606	Virulência, doença e defesa	Viruloma de Legionella pneumofila	ATPase provides energy for both assembly ofTipo IV secretion complex e secretion of T-DNA complex (VirB4)
wCam_00636	Virulência, doença e defesa	Viruloma de Legionella pneumofila	Cold shock protein CspA
wBolb_00683	Virulência, doença e defesa	Viruloma de Legionella pneumofila	SdbB protein
wJHB_01243	Virulência, doença e defesa	Viruloma de Legionella pneumofila	SdbB protein
wCam_00231	Virulência, doença e defesa	Viruloma de Legionella pneumofila	Tipo IV secretion sistema protein VirD4

Tabela S2 - Métricas das abordagens de taxonomia genômica utilizadas em isolados do gênero *Klebsiella*.

KPI	ANI	AF	GGD	cgMLS	KPII	ANI	AF	GGD	cgMLS	KPIII	ANI	AF	GGD	cgMLS
			C	T				C	T				C	T
1084	99.2	0.9	93.4	99.43	HS11286	93.8	0.8	53.8	94	HS11286	93.86	0.8	53.8	94
	4					6	3				3			
1158	99.2	0.8	93.7	99.44	1084	93.9	0.8	53.8	94	1084	94.83	0.8	58.5	94
	6	9				4	4				6			
3219	99.8	0.9	98.4	99.43	1158	93.8	0.8	53.4	94	1158	94.77	0.8	58.4	94
2		5				9	3				6			
3461	99.7	0.9	98.4	99.43	32192	93.8	0.8	53.3	94	32192	94.75	0.8	58.2	94
8	8	4				8	4				6			
88SG	99.7	0.9	97.9	99.43	34618	93.8	0.8	53.5	94	34618	94.75	0.8	58.1	94
R	7	5				7	4				6			
blaN	99.5	0.9	97.1	99.45	88SGR	93.9	0.8	53.8	94	88SGR	94.73	0.8	57.9	94
DM-1	8	4				1	6				6			
CG43	99.3	0.9	94	99.45	blaNDM-1	93.8	0.8	53.6	94	blaNDM-1	94.8	0.8	58.5	94
						8	4				6			
HK78	99.2	0.9	93.6	99.41	CG43	93.9	0.8	53.7	94	CG43	94.81	0.8	58.5	94
7	6	2				3	4				6			
JM45	99.8	0.9	98.7	99.45	HK787	93.9	0.8	53.7	94	HK787	94.79	0.8	58.4	94
	6	4				2	4				6			
KCT	99.2	0.9	93.7	99.42	JM45	93.9	0.8	53.4	94	JM45	94.75	0.8	58.2	94
C_22	3					1	3				5			
42														
Kp13	99.2	0.9	93.5	99.42	KCTC_224	93.8	0.8	53.4	94	KCTC_22	94.82	0.8	58.6	94
	5	1			2	5	4			42	7			
KPNI	99.7	0.9	98.2	99.43	Kp13	93.8	0.8	53.7	94	Kp13	94.79	0.8	58.4	94
H1	9	5				9	4				6			
KPNI	99.7	0.9	98.2	99.41	KPNIH1	93.8	0.8	53.4	94	KPNIH1	94.76	0.8	58.2	94
H10	9	5				6	4				6			
KPNI	99.8	0.9	98.4	99.89	KPNIH10	93.8	0.8	53.4	94	KPNIH10	94.76	0.8	58.2	94
H24		5				6	4				6			
KPNI	99.2	0.8	94.2	99.89	KPNIH24	93.8	0.8	53.4	94	KPNIH24	94.74	0.8	58.2	94
H27	9	9				9	4				6			
KPNI	99.2	0.9	93.1	99.86	KPNIH27	93.9	0.8	53.7	94	KPNIH27	94.75	0.8	58.3	95
H29	1					1	4				5			
KPNI	99.8	0.9	98.4	99.86	KPNIH29	93.8	0.8	53.6	94	KPNIH29	94.76	0.8	58.4	94
H30		5				9	5				6			
KPNI	99.2	0.9	93.8	99.42	KPNIH30	93.8	0.8	53.3	94	KPNIH30	94.76	0.8	58.2	94
H31	3					7	4				6			
KPNI	99.7	0.9	98	99.88	KPNIH31	93.8	0.8	53.5	94	KPNIH31	94.76	0.8	58.4	94

H32	8	6				4	4					6		
KPNI	99.7	0.9	98	99.41	KPNIH32	93.8	0.8	53.4	94	KPNIH32	94.77	0.8	58.2	94
H33	8	5				6	4					6		
KPP	99.2	0.9	94	99.87	KPNIH33	93.8	0.8	53.4	94	KPNIH33	94.74	0.8	58.2	94
R1	4					6	4					6		
KPR0	99.8	0.9	98.4	99.88	KPPR1	93.8	0.8	53.6	94	KPPR1	94.76	0.8	58.6	94
928		5				6	5					6		
NJST	99.7	0.9	98.1	99.72	KPR0928	93.8	0.8	53.3	94	KPR0928	94.76	0.8	58.2	94
258_	8	3				7	4					6		
1														
NTU	99.2	0.9	93.5	99.83	NJST258_	93.9	0.8	53.6	94	NJST258_	94.79	0.8	58.3	94
H-	4				1	1	2			1		4		
K204														
4														
PittN	99.2	0.9	93.8	99.89	NTUH-	93.9	0.8	53.8	94	NTUH-	94.84	0.8	58.7	94
DM0	2				K2044	4	5			K2044		6		
1														
PMK	99.2	0.9	93.6	99.88	PittNDM01	93.8	0.8	53.5	94	PittNDM0	94.71	0.8	58.2	94
1	5					7	5			1		7		
RYC	99.2	0.8	93.7	99.88	PMK1	93.8	0.8	53.6	94	PMK1	94.77	0.8	58.4	94
492	9	6				6	5					6		
UHK	99.8	0.9	98.5	99.89	RYC492	93.9	0.8	53.6	94	RYC492	94.83	0.8	58.6	94
PC07		4					1					3		
UHK	99.7	0.9	98.2	99.89	UHKPC07	93.9	0.8	53.6	94	UHKPC07	94.77	0.8	58.1	94
PC45	9	4				4	5					6		
XH20	99.3	0.8	94.3	99.88	UHKPC45	93.9	0.8	53.6	94	UHKPC45	94.8	0.8	58.3	94
9	2	9				2	5					5		
BAA-	99.7	0.9	98.8	99.89	XH209	93.8	0.8	53.6	94	XH209	94.82	0.8	58.7	94
2146	9	5				8	4					6		
MGH	99.2	0.9	93.8	94.63	B1_1	93.5	0.8	52	93	BAA-2146	94.64	0.8	57.8	94
_785	4	1				6	5					7		
78														
K142	93.8	0.8	54.8	94.63	MGH_785	93.8	0.8	53.5	92	MGH_785	94.79	0.8	58.5	94
	5	5			78	5	4			78		7		
865_	93.9	0.8	58.3	94.63	865_KPNE	99.0	0.8	92	98	K142	93.59	0.8	52.2	94
KPN	5	4				1	9					6		
E														
MGH	93.8	0.8	54	94.6	MGH113	98.9	0.8	91.3	99	865_KPNE	93.64	0.8	52.6	93
113	7	5				3	8			E		7		
FI_H	93.8	0.8	54.8	94.59	FI_HV_201	98.9	0.8	91	99	MGH113	93.6	0.8	52.6	94
V_20	7	5			4	1	9					6		
14														
18A6	93.8	0.8	54.8	94.62	18A69	98.9	0.8	92.1	99	FI_HV_20	93.6	0.8	52.6	94

9	7	5				2	5			14		6		
01A30	93.8	0.8	54.8	94.65	01A30	98.9	0.9	91.5	99	18A69	93.6	0.8	52.6	94
	7	5				5						6		
MGH96	93.9	0.8	54.8	94.68	MGH96	98.9	0.9	91.9	99	01A30	93.6	0.8	52.6	94
	3	6				4						6		
UCICRE_1	94.0	0.8	58.5	94.71	UCICRE_1	98.9	0.9	91.9	99	MGH96	93.64	0.8	52.4	94
	2	6				4	7					7		
	4													
700603	94.0	0.8	54.5	94.61	700603	96.6	0.8	70.8	96	UCICRE_14	93.73	0.8	52.66	94
	6	6				4	7					7		
07A044	94.1	0.8	54.5	94.63	07A044	96.7	0.8	71.4	97	700603	93.73	0.8	52.7	94
	5	5				5	7					8		
12-3578	94.1	0.8	54.5	94.63	12-3578	96.7	0.8	71.1	96	07A044	93.77	0.8	52.8	94
	3	5				7	7					8		
1294_KPNE	94.1	0.8	54.4	94.63	1294_KPNE	96.7	0.8	71.7	96	12-3578	93.73	0.8	52.7	94
	3	4				8	4					6		
1kgm8	94.1	0.8	58.3	94.6	1kgm	96.7	0.8	70.8	96	1294_KPNE	93.72	0.8	52.6	94
	8	5				1	6					6		
ATCC_43816	94.1	0.8	54.4	94.59	ATCC_43816	96.6	0.8	70.6	97	1kgm	93.81	0.8	52.9	94
		4				4	5					7		
BIDMC95	94.0	0.8	54.5	94.62	BIDMC95	96.7	0.8	71.7	96	ATCC_43816	93.75	0.8	52.8	94
	8	5				5	8					7		
HKUOPLC	94.2	0.8	55	94.65	HKUOPLC	96.7	0.8	71.4	96	BIDMC95	93.75	0.8	52.9	94
	1	5				7	5					8		
KPUCI14	94.1	0.8	54.8	94.68	KPUCI14	96.7	0.8	71.2	96	HKUOPLC	93.8	0.8	53.1	94
	5	5				1	5					7		
MGH_44	94.1	0.8	54.8	94.71	MGH_44	96.7	0.8	71.2	96	KPUCI14	93.74	0.8	52.9	94
	5	5					5					5		
MGH123	94.1	0.8	54.8	94.65	MGH123	96.6	0.8	71.2	96	MGH_44	93.74	0.8	52.9	94
	7	5				6	6					5		
At-229	94.7	0.8	58.3	95.08	At-22	93.5	0.8	52.2	93	MGH123	93.72	0.8	52.7	94
	9	7				9	6					6		
BZ198	94.7	0.8	58.3	95.05	BZ19	93.6	0.8	52.8	93	BZ19	99.2	0.9	93.4	99
	8	5				3	4					1		
B1_19	94.7	0.8	58.2	95.08	B1_1	93.5	0.8	52	93	B1_1	99.22	0.9	93.6	99
	9	6				6	5					3		
1335_KPNE	94.7	0.8	58.3	95.07	1335_KPNE	93.5	0.8	52.4	93	1335_KPNE	99.17	0.9	93.1	99
	4	7				8	7					2		
ASM78637	94.7	0.8	58.5	95.07	ASM78637	93.5	0.8	52.5	93	ASM7863	99.19	0.9	93.4	99

7863	8	8			v1	6	8			7v1		4		
7v1														
342	94.7	0.8	58.3	95.06	342	93.5	0.8	52.2	93	342	99.12	0.9	92.4	99
	7	7				6	6					4		
K115	94.8	0.8	58.5	95.06	K1155	93.6	0.8	52.4	93	K1155	99.12	0.9	92.7	99
5	4	7				6	6					3		
UCIC	94.8	0.8	58.5	95.05	UCICRE_1	93.5	0.8	52.2	93	UCICRE_	99.21	0.9	93.5	99
RE_1		7				6	6			10		4		
0														
KTE9	94.7	0.8	58.3	95.04	KTE92	93.5	0.8	52.4	93	KTE92	99.12	0.9	93.1	99
2		7				5	6					2		
KP5-	94.7	0.8	57.8	95.05	KP5-1	93.5	0.8	52	93	KP5-1	99.13	0.9	92.7	99
1	6	6				8	6					3		
801	94.6	0.8	57.8	83.82	801	93.6	0.8	52.7	93	801	99.2	0.9	93.2	99
	8	7				9	7					3		
MGH	94.8	0.8	58.5	83.84	MGH_20	93.5	0.8	52.1	93	MGH_20	99.24	0.9	93.5	99
_20		8				8	7					4		
MGH	94.7	0.8	58.1	83.85	MGH_40	93.6	0.8	52.4	93	MGH_40	99.17	0.9	93.1	99
_40	4	6				1	5					3		
780_	94.7	0.8	58.3	83.79	780_KPNE	93.6	0.8	52.4	93	780_KPN	99.24	0.9	93.8	99
KPN	6	6				4	6			E		3		
E														
833_	94.7	0.8	58.3	83.83	833_KPNE	93.6	0.8	52.4	93	833_KPN	99.24	0.9	93.8	99
KPN	3	5				5	5			E		1		
E														
834_	94.7	0.8	58.3	95.08	834_KPNE	93.6	0.8	52.4	93	834_KPN	99.24	0.9	93.8	99
KPN	6	6				4	6			E		2		
E														
835_	94.7	0.8	58.3	95.08	835_KPNE	93.6	0.8	52.4	93	835_KPN	99.24	0.9	93.8	99
KPN	6	5				3	5			E		1		
E														
836_	94.7	0.8	58.3	95.05	836_KPNE	93.6	0.8	52.4	93	836_KPN	99.24	0.9	93.8	99
KPN	5	6				3	6			E		2		
E														
837_	94.7	0.8	58.3	95.08	837_KPNE	93.6	0.8	52.4	93	837_KPN	99.24	0.9	93.8	99
KPN	6	5				4	5			E		2		
E														
BIDM	94.7	0.8	58	95.07	BIDMC_61	93.6	0.8	52.5	93	BIDMC_6	99.19	0.9	93.1	99
C_61	7	7				3	5			1		4		
BIDM	94.7	0.8	58.4	95.07	BIDMC88	93.5	0.8	52.4	93	BIDMC88	99.14	0.9	93.1	99
C88	9	7				8	6					3		
BIDM	94.7	0.8	54.5	95.06	BIDMC90	93.6	0.8	53.1	93	BIDMC90	99.1	0.9	92.5	99
C90	9	7				9	8					3		

CH4_1	94.7	0.8	58	95.06	CH4_1	93.6	0.8	52.4	93	CH4_1	99.23	0.9	93.5	99
	1	6				1	5				2			
CICC1001	94.7	0.8	58	95.05	CICC1001	93.5	0.8	52.1	93	CICC1001	99.17	0.9	93	99
	4	7			10	9	8			10	4			
	10													
DSM_159	94.7	0.8	58.4	95.04	DSM_1596	93.6	0.8	52.4	93	DSM_159	99.17	0.9	93	99
	7	7			8	2	6			68	4			
	68													
DX120E	94.7	0.8	58.5	95.05	DX120E	93.6	0.8	52.2	93	DX120E	99.24	0.9	93.6	99
	9	7				1	6				4			
MGH-68	94.7	0.8	54.8	95.06	MGH-68	93.5	0.8	52.3	93	MGH-68	99.16	0.9	93.1	99
	9	7				9	5				3			
MGH-76	94.7	0.8	58.4	95.07	MGH-76	93.6	0.8	52.7	93	MGH-76	99.21	0.8	93.3	99
	7					7	1				6			
MGH-80	94.7	0.8	54.8	95.08	MGH-80	93.6	0.8	52.3	93	MGH-80	99.22	0.9	93.6	99
	5	7				4	7				4			
MGH114	94.8	0.8	54.8	95.09	MGH114	93.6	0.8	52.7	93	MGH114	99.21	0.9	93.5	99
	1	8				3	9				4			
MGH92	94.7	0.8	58.5	95.1	MGH92	93.6	0.8	52.4	93	MGH92	99.16	0.9	92.8	99
	7	7				4	6				3			
UCI_18	94.8	0.8	58.6	95.11	UCI_18	93.6	0.9	52.8	93	UCI_18	99.21	0.9	93.5	99
	3	8				3					4			
10982	93.8	0.8	55.6	95.2	10982	93.5	0.8	50.5	93	10982	93.61	0.8	56.3	94
	2	6	5				3				6			
HKOPL1	84.0	0.7	27.3	83.82	HKOPL1	84.5	0.7	28.1	84	HKOPL1	84.08	0.8	27.2	84
	7	9					8							
E718	84.1	0.8	27.4	83.84	E718	84.5	0.7	27.9	84	E718	84.03	0.8	27.3	84
	3	1				3	9				1			
KCTC_16	84.0	0.8	27.3	83.85	KCTC_168	84.4	0.7	27.7	84	KCTC_16	84.02	0.8	27.4	84
	4				6	7	9			86	1			
	86													
KONIH1	84.1	0.8	27.4	83.79	KONIH1	84.5	0.7	28.3	84	KONIH1	84.08	0.8	27.1	84
	6	1				2	9				1			
M1	84.1	0.8	27.4	83.83	M1	84.5	0.7	28.5	84	M1	84.06	0.8	27	84
						1	9				1			

Artigo The genome of a clinical *K. variicola* strain reveals virulence associated traits and a p19-like plasmid.

The genome of a clinical *Klebsiella variicola* strain reveals virulence-associated traits and a pl9-like plasmid

Bruno Gabriel N. Andrade, Nilceia de Veiga Ramos, Michel F. Abanto Marin, Erica L. Fonseca & Ana Carolina P. Vicente

Laboratório de Genética Molecular de Microorganismos, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

Correspondence: Bruno Gabriel N. Andrade, Laboratório de Genética Molecular de Microorganismos, Instituto Oswaldo Cruz, FIOCRUZ, Av. Brasil, 4365 – Manguinhos, 21045-900 Rio de Janeiro, Brazil. Tel./fax: +552138658168; e-mail: bruno@ioc.fiocruz.br

Received 17 June 2014; accepted 14 August 2014.

DOI: 10.1111/1574-6968.12583

Editor: Mark Schembri

Keywords

Klebsiella virulence determinants; mobilome; penicillin resistant; whole genome sequencing.

Abstract

Klebsiella species frequently cause clinically relevant human infections worldwide. We report the draft genome sequence of a Brazilian clinical isolate (Bz19) of the recently recognized species *Klebsiella variicola*. The comparison of Bz19 genome content with the At-22 (environmental *K. variicola*) and several clinical *Klebsiella pneumoniae* shows that these species share a set of virulence-associated determinants. Of note, this *K. variicola* strain harbours a plasmid-like element that shares the same backbone present in a multidrug-resistant plasmid found in a clinical *K. pneumoniae* isolated in USA.

The genus *Klebsiella* belongs to the *Enterobacteriaceae* family and comprises non-motile, Gram-negative bacteria that present a polysaccharide-based capsule, which plays a role in *Klebsiella pneumoniae* pathogenicity. The clinically most important species are the opportunistic human pathogens *K. pneumoniae* and *Klebsiella oxytoca*, of which multidrug-resistant strains have been associated with hospital-acquired infections. The *Klebsiella* species are also characterized by class A chromosomal β -lactamase genes that confer intrinsic resistance to penicillins (Babini & Livermore, 2000).

Based on polyphasic characterization, including multilocus sequence analysis and metabolic traits, *Klebsiella* nitrogen-fixing strains were defined as members of a new species named *Klebsiella variicola* (Rosenblueth & Martínez, 2004). This species is known to comprise clinical and plant-associated strains and, to date, there is only one complete genome assigned as *K. variicola* available in NCBI, the At-22 (CP001891), which belongs to an environmental strain isolated from a fungus garden of leaf-cutters ants (Pinto-Tomás *et al.*, 2009).

Herein, we present the draft genome sequence of a clinical *K. variicola* strain, Bz19, isolated from faeces of a hospitalized patient in Rio de Janeiro, Brazil (2005). The Bz19 genome sequence will provide new and valuable information that enables us to address, through genome comparison, possible traits and differences between clinical and environmental *K. variicola* strains and their relationship with the closely related *K. pneumoniae* species.

Bz19 strain was assigned as *K. variicola* based on phenotypic and genetic traits. The Bz19 shares 99.9% and 98.58% 16S rRNA gene nucleotide identity, and 99.9% and 97% *rpoB* nucleotide identity with *K. variicola* At-22 and *K. pneumoniae* HS11286^T, respectively. Bz19 was resistant to penicillins (ampicillin and amoxicillin), a resistance profile compatible with the presence and activity of an intrinsic Class A β -lactamase, as found for other *Klebsiella* species.

The Bz19 strain is deposited in the Environment and Health Bacterial Collection (CBAS) at FIOCRUZ/Brazil. Its genome sequencing was performed on the high-throughput platform at Oswaldo Cruz Institute using

Table 1. Genomic features of *Klebsiella variicola* Bz19 and *Klebsiella* reference genomes

	Kv Bz19	Kv AT-22	Kp 342	Kp HS11286
Genomic features*				
Size, Mb	c. 5.48	c. 5.46	c. 5.64	c. 5.68
Protein-coding genes	5221	5213	5567	5867
Plasmids	1	0	2	6
GC content, %	57.3	57.6	56.88	57.1
tRNAs	71	89	88	99
Nif cluster	+	+	+	–
Class A β -lactamase	+	+	+	+
<i>magA</i>	–	–	–	+
<i>allS</i>	–	–	–	+
<i>mrkD</i>	+	+	+	+
<i>uge</i>	+	+	+	+
<i>ureA</i>	+	+	+	+
<i>ureD</i>	+	+	+	+
<i>wabG</i>	+	+	+	+
<i>wza</i>	+	+	+	+
<i>wzb</i>	+	+	+	+
<i>wzc</i>	+	+	+	+
<i>wzi</i>	+	+	+	+

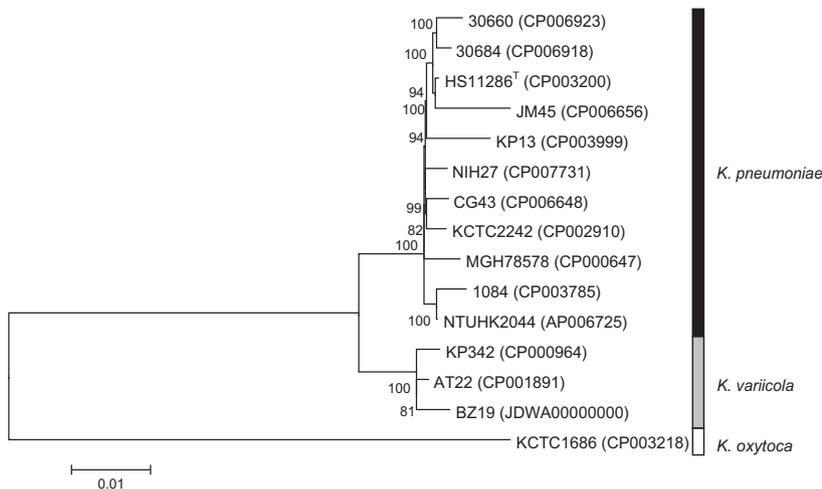
*Gene prediction and tRNA identification were performed using PGAAP (Angiuoli *et al.*, 2008). Nitrogen-fixing *nif* cluster, antibiotic resistance and virulence determinants identification were performed using BLAST+ (Camacho *et al.*, 2009).

Roche 454-GS Junior sequencer. Genomic reads (70.5 Mb) were *de novo* assembled using NEWBLER v2.7, which generated 208 contigs, totalling 5.48 Mb with an average GC content of 57.3% (Table 1). Gene prediction and annotation was performed using the PGAAP (Angiuoli *et al.* 2008). A phylogenomic analysis was performed using complete genomes of the environmental *K. variicola* At-22 (CP001891), 12 *K. pneumoniae* and the *K. oxytoca* KCTC1686 (CP003218) (Fig. 1). The orthologue genes were identified using the *get_homologs* software (Contreras-Moreira & Vinuesa, 2013), with the orthoMCL algo-

rithm, and the inparalogous sequences were excluded. The sequences were concatenated, comprising 1 227 462 amino acids, and aligned using the MAFFT software (Katoh *et al.*, 2005), and the phylogenomic tree was performed using MEGA v5 with 1000 bootstrap replicates (Tamura *et al.*, 2011).

The *K. pneumoniae/variicola* core genome comprises 3762 proteins, and contains some virulence-associated genes coding for secretion systems and fimbriae. The T2SS entire cluster, composed of 13 genes, responsible for the formation and secretion of pullulanase and several exotoxins (Dunstan *et al.* 2013) and an entire T6SS cluster coding for an apparatus that secretes toxins using a needle-like mechanism resembling an injectosome (Bönnemann *et al.* 2010) were part of the *K. pneumoniae/variicola* core genome. In *Klebsiella* genus had already been identified 13 putative T6SS gene cluster (Sarris *et al.*, 2011). The one found in Bz19 is identical to that from At-22. In general, type I fimbriae assures attachment to host cell surfaces, one of the first steps in the infection process. Type III fimbriae, associated with biofilm formation, provides a greater tolerance towards antibiotics and components of the host immune response (Schroll *et al.* 2010). These two fimbriae gene clusters were previously reported in *K. pneumoniae*, enhancing virulence during urinary tract infection, and mediating biofilm formation on biotic and abiotic surfaces (Struve *et al.*, 2008).

Interestingly, the Bz19 capsular polysaccharides biosynthesis gene cluster (*c.* 18 kb) is more similar to the capsule cluster of a clinical *K. pneumoniae* strain (AB289650), which belong to the K54 capsular type (100% of coverage and 92% identity), than to that of the *K. variicola* At-22 (43% of coverage and 90% identity), suggesting the occurrence of lateral gene transfer. The capsule cluster is one of the major virulence factors among *K. pneumoniae* strains (Mizuta *et al.*, 1983), and

**Fig. 1.** Neighbour-joining phylogenomic tree based on the *Klebsiella pneumoniae/variicola/oxytoca* core genome.

the expression identity shared by clinical strains from these two *Klebsiella* species is of note.

The phylogenomic tree shows that the *K. pneumoniae* KP342 genome was within the *K. variicola* cluster (Fig. 1). This strain had been assigned to the *K. pneumoniae* species by means of 16S rRNA gene and phenotypic analysis (Fouts *et al.*, 2008). The KP342 core genome presented c. 99.5% identity with *K. variicola* At-22 and Bz19, and 98% identity with *K. pneumoniae* HS11286^T. Previous studies suggested that KP342 strain was a member of *K. variicola* strain on the basis of genetic information of some genes and metabolic profiles (Martínez *et al.*, 2004; Ramos *et al.*, 2014). Herein, based on whole genome analysis, we reinforce that KP342 belongs to the *K. variicola* species and therefore it should be reclassified.

The main differences among *K. pneumoniae/variicola* genome content is the absence of the *nif* cluster, associated with nitrogen fixation, in the *K. pneumoniae* genomes (Table 1). Mutations in genes whose products are antibiotic targets, such as *gyrA* and *parC* (quinolone resistance) and the 16S rRNA gene (aminoglycoside resistance), were not detected in Bz19 genome. However, this genome has a putative gene presenting all motifs of a class A chromosomal β -lactamase.

A contig yielding 50 687 bp was recovered as part of the Bz19 mobilome. This plasmid-like element presented 51.4% GC content, contrasting with the 57.3% of the chromosome. This element had a high identity with the conjugative plasmid 9 (p19) (FJ223607) found in a multi-resistant clinical *K. pneumoniae* strain (Gootz *et al.*, 2009). The pBz19, as p19, belong to the IncN incompatibility group, which has a broad host range and so far have been detected in *Klebsiella*, *Escherichia* and *Salmonella* genera. The p19 and pBz19 replicons share the entire gene content, including 79 ORFs coding for transcriptional factors, transposases, regulatory proteins, EcoRII endonuclease and Type IV conjugation system (Craig *et al.*, 2004). They differ by two identical regions corresponding to the *Tn4401* (c. 62% GC content), which is responsible for *bla*_{KPC-2} gene dissemination. In the p19, *Tn4401* elements, harbouring the *bla*_{KPC-2}, are inserted within a putative reverse transcriptase gene. However, this gene is intact in the pBz19 due to the absence of *Tn4401* (Supporting Information, Fig. S1). A search in GenBank using this plasmid sequence as a query only, returned significant identities with *K. pneumoniae* plasmids and therefore this is the first report of this plasmid in other *Klebsiella* species.

Nucleotide sequence accession numbers: The whole genome shotgun project has been deposited at GenBank under accession JDWA00000000, and sequencing reads are available at the NCBI Sequence Read Archive under accession number SRR1169872.

Acknowledgements

This work was supported by Instituto Oswaldo Cruz grant (PROEP) and FAPERJ, CAPES and Fiocruz Fellowships.

Authors' contribution

All authors have contributed equally to this work.

References

- Angiuoli SV, Gussman A, Klimke W, *et al.* (2008) Toward an online repository of Standard Operating Procedures (SOPs) for (meta) genomic annotation. *Omics* **12**: 137–141.
- Babini GS & Livermore DM (2000) Are SHV β -lactamases universal in *Klebsiella pneumoniae*? *Antimicrob Agents Chemother* **44**: 2230.
- Bönemann G, Pietrosiuk A & Mogk A (2010) Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**: 815–821.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Contreras-Moreira B & Vinuesa P (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* **79**: 7696–7701.
- Craig L, Pique ME & Tainer JA (2004) Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**: 363–378.
- Dunstan RA, Heinz E, Wijeyewickrema LC, *et al.* (2013) Assembly of the type II secretion system such as found in *Vibrio cholerae* depends on the novel Pilotin AspS. *PLoS Pathog* **9**: e1003117.
- Fouts DE, Tyler HL, DeBoy RT *et al.* (2008) Complete genome sequence of the N2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet* **25**: e1000141.
- Gootz TD, Lescoe MK, Dib-Hajj F, Dougherty BA, He W, Della-Latta P & Huard RC (2009) Genetic organization of transposase regions surrounding *bla*_{KPC} carbapenemase genes on plasmids from *Klebsiella* strains isolated in a New York City hospital. *Antimicrob Agents Chemother* **53**: 1998–2004.
- Katoh K, Kuma K, Toh H & Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**: 511–518.
- Martínez J, Martínez L, Rosenblueth M, Silva J & Martínez-Romero E (2004) How are gene sequence analyses modifying bacterial taxonomy? The case of *Klebsiella*. *Int Microbiol* **7**: 261–268.
- Mizuta K, Ohta M, Mori M, Hasegawa T, Nakashima I & Kato N (1983) Virulence for mice of *Klebsiella* strains belonging to the O1 group: relationship to their capsular (K) types. *Infect Immun* **40**: 56–61.

- Pinto-Tomás A, Anderson M, Suen G, Stevenson DM, Chu FST, Cleland WW & Currie CR (2009) Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* **326**: 1120–1123.
- Ramos PI, Picão RC, Almeida LG *et al.* (2014) Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genomics* **15**: 54.
- Rosenblueth M & Martínez L (2004) *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol* **35**: 27–35.
- Sarris PF, Zoumadakis C, Panopoulos NJ & Scoulica EV (2011) Distribution of the putative type VI secretion system core genes in *Klebsiella* spp. *Infect Genet Evol* **11**: 157–166.
- Schroll C, Barken KB, Krogfelt K & Struve C (2010) Role of type 1 and type 3 fimbriae in *Klebsiella pneumoniae* biofilm formation. *BMC Microbiol* **10**: 179.
- Struve C, Bojer M & Krogfelt KA (2008) Characterization of *Klebsiella pneumoniae* type 1 fimbriae by detection of phase variation during colonization and infection and impact on virulence. *Infect Immun* **76**: 4055–4065.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Visual comparison of plasmids p19 of *Klebsiella pneumoniae* (yellow) and pBz19 of *Klebsiella variicola* BZ19 (pink) generated with BRIG (<http://sourceforge.net/projects/brig/>). Missing regions in pBz19 corresponding to the *Tn4401* element, which harbours the *bla*_{KPC} gene, and the T4SS cluster are highlighted (black).