

UNIVERSIDADE FEDERAL DO PARANÁ

SHEYLA TREFFLICH

SEQUENCIAMENTO E ANÁLISE DO GENOMA DE *Derxia lacustris* HL12

CURITIBA

2016

SHEYLA TREFFLICH

SEQUENCIAMENTO E ANÁLISE DO GENOMA DE *Derxia lacustris* HL12

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Bioinformática, no curso de Pós-Graduação em Bioinformática, Setor e Educação Profissional e Tecnológica da Universidade Federal do Paraná.

Orientador: Prof. Dr. Emanuel Maltempi de Souza
Coorientador: Dr. Arnaldo Glogauer

CURITIBA

2016

T786 Trefflich, Sheyla
Sequenciamento e análise do genoma de *Derxia lacustris* HL12 /
Sheyla Trefflich. - Curitiba, 2016.
51 f.: il., tabs, grafs.

Orientador: Prof. Dr. Emanuel Maltempi de Souza
Coorientador: Dr. Arnaldo Glogauer
Dissertação (Mestrado) – Universidade Federal do Paraná, Setor
de Educação Profissional e Tecnológica, Curso de Pós-Graduação em
Bioinformática.

1. Nitrogênio - Fixação. 2. Genoma. 3. *Derxia lacustris*.
4. Bioinformática. I. Souza, Emanuel Maltempi de. II. Glogauer,
Arnaldo. III. Título. IV. Universidade Federal do Paraná.

CDD 575.113

TERMO DE APROVAÇÃO

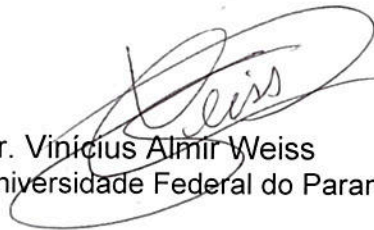
SHEYLA TREFFLICH

"SEQUENCIAMENTO E ANÁLISE DO GENOMA DE *Dexia lacustris* HL12"

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:



Dr. Emanuel Maltempi de Souza
Universidade Federal do Paraná - UFPR



Dr. Vinicius Almir Weiss
Universidade Federal do Paraná - UFPR



Dr^a. Maria Berenice Reynaud Steffens
Universidade Federal do Paraná - UFPR

Curitiba, 31 de março de 2016

Dedico esse trabalho à minha família
e ao meu esposo.

AGRADECIMENTOS

À CAPES, à Fundação Araucária, ao INCT de Fixação Biológica do Nitrogênio e ao Programa de Pós-Graduação em Bioinformática.

Ao Prof. Dr. Emanuel Maltempo de Souza, meu orientador, com quem iniciei os contatos para o mestrado muito antes de enveredar pelos caminhos da Bioinformática. Foi por meio de sua experiência e paciência que me ensinou muito mais do que a compreensão técnica, ensinamentos esses que levarei para a vida.

À Prof.^a Dra. Maria Berenice Reynald Steffens, pela colaboração com o meu trabalho, com minhas decisões e por dividir comigo muito de sua sabedoria. Agradeço pela sua disponibilidade e por todo o seu apoio.

Ao Pós-doc Vinícius Weiss e ao doutorando Dieval Guizelini que auxiliaram muitíssimo na elaboração desse trabalho com a contribuição técnica.

Aos professores Dr. Roberto Tadeu Raitz e Dr. Mauro Castro que sempre foram extremamente acessíveis para todo e qualquer esclarecimento que eu necessitasse.

À secretária Suzana Gobetti pelo apoio, pela colaboração e por toda a disponibilidade e eficiência com a qual conduz o seu trabalho.

Aos meus amigos de Bioinfo: Vinícius Chagas, Rodrigo Langowski, Flavia Costa, Nilson Coimbra e, em especial à Aniele Leão, parceira de todas as horas. Fiquem certos de que todos vocês, em algum momento e em determinada medida, contribuíram para a construção desse trabalho.

À minha família, a fonte de energia para todos os desafios que eu encaro. Aos meus pais, pelo amor, dedicação, preocupação e investimento humano durante toda a minha vida. À minha irmã Kelly, minha companheira fraterna, obrigada pela sintonia, pela amizade, pelo cuidado e pelo carinho. À família Trefflich Luna pelo apoio, pelo auxílio e pela presteza. Vocês são minhas jóias preciosas, por quem me levanto todos os dias na esperança de ser hoje, melhor do que eu fui ontem.

Ao Ander, meu esposo, meu companheiro de vida. Nossa sintonia é tanta que podemos estar a quilômetros de distância, ainda assim somos capazes de estar lado a lado, compreendendo exatamente o que se passa com o outro. Obrigada por reinventar junto comigo, todos os dias, a confiança, a parceria de nossas almas e o nosso mais profundo sentimento de amor.

À Deus, o criador de todas as coisas, Àquele que atende os meus pedidos mais sinceros, que me consola em minha dor, que está presente em todos os meus passos, que me concede o dom da vida. Agradeço pela saúde, pela família, pela oportunidade de obter conhecimento e pela força de minha fé.

"O segredo da mudança é focar toda a sua energia, não em brigar com o velho, mas em construir o novo."

Sócrates

RESUMO

As rizobactérias são bactérias de vida livre que habitam as raízes das plantas, desenvolvem uma relação de simbiose e geram benefícios aos organismos vegetais. A existência desses organismos é de grande importância na agricultura, já que a carência na absorção de nitrogênio impacta diretamente na taxa de crescimento das plantas, diminui a produtividade agrícola e originam problemas de ordem ambiental com a utilização de fertilizantes que são poluentes ao meio. A fixação biológica de nitrogênio é realizada por uma ampla gama de bactérias. *Derxia lacustris* HL12 é uma bactéria fixadora de nitrogênio, pertencente à classe das *Betaproteobacterias*, ordem *Burkholderiales*, família *Alcaligenaceae*, gênero *Derxia*. Esse organismo foi isolado de amostras de água provenientes de um lago na região de Taiwan. *D. lacustris* HL12 é a segunda bactéria inserida no gênero *Derxia*, de acordo com a comparação entre as sequências do gene do RNA ribossomal 16S desta e de *Derxia gummosa* DSM 723, que é considerado o organismo de referência. Esse organismo está depositado no NCBI e possui uma montagem com 19 contigs e um contig separado da sequência do 16S rRNA. *D. lacustris* HL12 teve seu genoma sequenciado por duas vezes na plataforma Illumina Miseq. O processo de sequenciamento resultou num total de 3.532.510 reads. Seus reads possuem um tamanho médio de 300pb. Os dados brutos foram montados com os seguintes softwares: CLC, Velvet, Celera, Spades, Masurca. A melhor montagem obtida foi a realizada no software Celera, possuindo 129 contigs e percentual de GC igual a 69,20%. O software Matlab foi utilizado em ensaios de ordenação e cálculos do genoma. O software Fgap foi utilizado com os dados provenientes de todas as montagens realizadas para o fechamento dos gaps existentes no genoma. O software Bowtie foi responsável por mapear o genoma com os dados brutos. O software Mummer foi utilizado na construção do dotplot entre *D. lacustris* HL12 e *D. gummosa* DSM 723. O software Fastqc foi utilizado para analisar a qualidade dos dados brutos. O genoma de *D. lacustris* HL12 possui um draft com 28 contigs. Foi localizado um operon ribossomal completo. O genoma passou por uma análise de anotação para identificação de genes em algumas vias metabólicas do organismo. Foram encontrados os genes do cluster Nif, reforçando o indício do organismo fixar nitrogênio com os anotadores Sila e Rast.

Palavras-chave: Fixação de nitrogênio, Sequência genômica, *Derxia lacustris* HL12.

ABSTRACT

Rhizobacteria are bacteria which inhabit plant roots developing a symbiosis relationship that benefits those organisms. The existence of this kind of bacteria is of great importance to agriculture, once the need for nitrogen absorption impacts directly on plant growth rate, decreases agricultural productivity and generates environmental problems caused by the use of fertilizers. The biological nitrogen fixation is made by a large number of species. *Derxia lacustris* HL12 is a nitrogen fixing bacterium belonging to the class *Betaproteobacterias*, order *Burkholderiales*, family *Alcaligenaceae*, genus *Derxia*. This organism was isolated from fresh water collected from a lake in Taiwan. *D. lacustris* HL12 is the second bacteria inserted in the genus *Derxia*, in accordance with the comparison between its gene sequence of RNA ribosomal 16S and that of *Derxia gummosa* DSM 723, considered the reference organism. This organism is stored at the NCBI and has an assembly with 19 contigs. and one additional contig separated from 16S rRNA's sequence. *D. lacustris* HL12 had its genome sequenced twice using the Illumina Miseq platform. The sequencing process resulted in a total of 3,532,510 reads. The average size of its reads is of 300 pb. Raw data were assembled using the following softwares: CLC, Velvet, Celera, Spades, Masurca. The best assembly was obtained by using the Celera software and has 129 contigs and a CG percentage of 69.20%. Matlab was used in collation test runs and genome calculations. The software Fgap used the data collected from all the assemblies obtained for closing the existing gaps in the genome. The software Bowtie scanned the genome using raw data. The software Mummer was used in the construction of the dotplot between *D. lacustris* HL12 and *D. gummosa* DSM 723. The software Fastqc was employed to analyze the quality of the raw data. *D. lacustris* HL12's genome has a draft with 28 contigs. A complete ribosomal operon was located. The genome was submitted through an annotation analysis for the identification of the genes in some of the organism's metabolic pathways. Nif cluster genes were found using Sila and Rast annotation tools, which evinces *D. lacustris* HL12 as a nitrogen fixing bacteria.

Key-words: Nitrogen fixation, Geomic sequencing, *Derxia lacustris* HL12.

LISTA DE FIGURAS

FIGURA 1. CULTIVO DE <i>D. gummosa</i> DSM 723	21
FIGURA 2. ÁRVORE FILOGENÉTICA BASEADA NAS SEQUÊNCIAS DO GENE 16S RRNA	23
FIGURA 3 FLUXOGRAMA DE ESTUDO DO GENOMA DE <i>Derxia lacustris</i> HL12..	25
FIGURA 4 FLUXOGRAMA DE MONTAGEM DO GENOMA DE <i>Derxia lacustris</i> HL12.	27
FIG. 1 NEIGHBOR-JOINING PHYLOGENETIC TREE WAS BUILT WITH THE MLSA TECHNIQUE.	31
FIG. 2 GRAPHICAL CIRCULAR MAP OF A SIMULATED DRAFT <i>Derxia lacustris</i> HL12 GENOME.....	32
FIG. 3 DERXIA LACUSTRIS HL12 GENOME ASSEMBLY SKETCH.	34
FIG. 4 ARRANGEMENT OF THE GENES LINKED TO NITRATE METABOLISM IN <i>Derxia lacustris</i> HL12.....	37
FIG. 5 PROPOSED PATHWAYS FOR AROMATIC COMPOUNDS METABOLISMO IN <i>Derxia lacustris</i> HL12, BASED ON THE PROPOSED OF PEDROSA ET AL., 2011.	38

LISTA DE TABELAS

TABELA 1 CARACTERÍSTICAS DO GENOMA DE <i>Derxia gummosa</i> DSM 723, DEPOSITADO NO NCBI.	22
TABLE 1 CLASSIFICATION AND GENERAL FEATURES OF <i>Derxia lacustris</i> HL12	30
TABLE 2 GENOME STATISTICS	35
TABLE 3 NUMBER OF GENES ASSOCIATES WITH COG FUNCTIONAL CATEGORIES	36

LISTA DE SIGLAS E ABREVIATURAS

DNA – ÁCIDO DESOXIRRIBONUCLÉICO

RNA – ÁCIDO RIBONUCLÉICO

RRNA 16S – SUBUNIDADE 16S DO RNA RIBOSSOMAL

NCBI – NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

µm - MICRÔMETRO

SUMÁRIO

1	INTRODUÇÃO.....	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	O ESTUDO DOS GENOMAS.....	16
2.2	SEQUENCIAMENTO DE GENOMAS.....	16
2.3	MONTAGEM DE GENOMAS	17
2.4	ANOTAÇÃO DE GENOMAS	18
2.5	FIXAÇÃO DE NITROGÊNIO	19
2.6	METABOLISMO DOS COMPOSTOS AROMÁTICOS.....	20
2.7	METABOLISMO DO NITRATO	20
2.8	SISTEMAS DE SECREÇÃO	20
2.9	GÊNERO <i>Derxia</i>	21
2.10	<i>Derxia gummosa</i> DSM 723.....	21
2.11	<i>Derxia lacustris</i> HL12	23
2.12	GENOMAS PROCARIÓTICOS	24
3	METODOLOGIA.....	25
3.1	ESTRATÉGIA DE MONTAGEM.....	26
4	ARTIGO CIENTÍFICO.....	28
	ABSTRACT.....	29
	INTRODUCTION	29
	ORGANISM INFORMATION	32
	CLASSIFICATION AND FEATURES.....	32
	GENOME SEQUENCING INFORMATION	33
	GENOME PROJECT HISTORY	33
	GROWTH CONDITIONS AND GENOMIC DNA PREPARATION.....	33
	GENOME SEQUENCING AND ASSEMBLY	33
	GENOME ANNOTATION	35
	GENOME PROPERTIES.....	35
	PROFILES OF METABOLIC PATHWAY.....	36
	CONCLUSIONS	38
	AUTHOR DETAILS.....	39
	REFERENCES	40
	ADDITIONAL FILE.....	45

CONCLUSÕES.....	48
REFERENCIAS	49

1 INTRODUÇÃO

Derxia lacustris HL12 é uma bactéria descoberta em amostras provenientes do Heaven Lake, um lago de água doce em Taiwan. Sua descrição morfofisiológica foi publicada em 2013 por Chen e colaboradores. Esse organismo é o segundo organismo do gênero, sendo o primeiro *Derxia gummosa* DSM 723 isolada de solos da Índia e publicado em 1960 por Jensen e colaboradores (JENSEN et al., 1960). Quando cultivada, *D. gummosa* DSM 723 produz colônias com aspecto gelatinoso e elástico bastante característico. Essa goma pode possuir potencial biotecnológico (CAMPELO e DOBEREINER, 1970; HILL e UNIT, 1971) característica que também pode ser compartilhada por *D. lacustris* HL12.

D. gummosa DSM 723 e *D. lacustris* HL12 foram descritas como espécies diferentes por meio da análise da sequência do RNA ribossomal 16S, sendo enquadradas no mesmo gênero e as bactérias mais próximas na classificação filogenética se diferem bastante do organismo foco desse estudo (CHEN et al., 2013).

Embora o gênero *Derxia* tenha sido pouco estudado este desperta interesse pelo fato de suas bactérias serem capazes de realizar fixação biológica de nitrogênio podendo possuir importância agrícola (CHEN et al., 2013; XIE e YOKOTA, 2014). *D. gummosa* DSM 723 teve seu genoma sequenciado, não finalizado, e atualmente distribuído em 19 contigs. *D. gummosa* DSM 723 é habitante dos solos tropicais e já foi encontrada e estudada nos solos brasileiros (CAMPELO e DOBEREINER, 1970).

D. lacustris HL12 é uma *Betaproteobacteria*, da ordem *Burkholderiales*, família *Alcaligenaceae*, gênero *Derxia* (CHEN et al., 2013).

Os objetivos desse trabalho incluem a montagem *de novo* do genoma de *D. lacustris* HL12, utilizando diferentes *softwares* montadores, e anotação do genoma de *D. lacustris* HL12-HL12, a fim de descobrir características metabólicas importantes que possam contribuir com os estudos dessa bactéria.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O ESTUDO DOS GENOMAS

Com o avanço da ciência, o resultado do estudo dos genomas torna-se matéria prima para a elaboração de novas investigações moleculares (SHENDURE et al., 2004). Desde que a molécula de DNA foi descoberta e sua estrutura determinada no início dos anos 1950 (WATSON e CRICK, 1952), muito se desenvolveu em termos tecnológicos e econômicos trazendo benefícios à sociedade.

O estudo dos genomas ocupa lugar de destaque na área que engloba as ciências biológicas e a projeto genoma humano foi um marco neste contexto (CATANHO et al., 2008).

A bioinformática surgiu com a necessidade de se armazenar, manipular e interpretar dados de sequenciamento genômico. A bioinformática utiliza os dados provenientes do sequenciamento do material genético das mais diversas espécies presentes na natureza. Sendo assim, os dados fragmentados e embaralhados resultantes do processo de sequenciamento, são organizados afim de que se possa determinar a sequência completa do genoma de um determinado organismo.

2.2 SEQUENCIAMENTO DE GENOMAS

O sequenciamento do genoma é um processo que tem evoluído ao longo dos anos. As primeiras técnicas de sequenciamento foram à mola propulsora para o investimento tecnológico atual. O sequenciamento de ácido nucléicos tornou-se uma importante ferramenta tanto no estudo do genoma dos organismos, quanto no estudo das relações que existem entre eles (RIDDER et al., 2013; LUSCOMBE et al., 2001).

O processo de sequenciamento do genoma tornou-se conhecido mundialmente com o Projeto Genoma Humano, motivado pela possibilidade de melhora nas condições de diagnóstico, e tratamento mais efetivo de muitas doenças, assim como no prognóstico e prevenção das mesmas (CATANHO et al. 2008). O projeto genoma humano durou 13 anos, iniciou em 1990 e terminou em 2003 (CATANHO et al., 2008, COLLINS e MANSOURA, 2000). O primeiro genoma sequenciado foi da bactéria *Haemophilus influenzae* no ano de 1995 (FLEISCHMANN, R. et al 1995). Ao longo dos anos mais e mais organismos tiveram seus genomas sequenciados, produzindo conhecimento para que também outros organismos sejam

amplamente estudados (SHENDURE, 2004). Dessa maneira torna-se possível conhecer o potencial dos organismos, como funcionam suas vias metabólicas e de que forma pode ocorrer interação entre os mais diversos organismos. O estudo dos genomas também abre diversas possibilidades no estudo de doenças, bem como, numa possível via terapêutica. Da mesma forma, o estudo dos genomas amplia o entendimento de como um organismo pode interagir com o meio em que está inserido, possibilitando o entendimento de uma possível importância biotecnológica do organismo estudado.

Com o desenvolvimento dos estudos comparativos de genomas, fica cada vez mais evidente que existe uma pressão seletiva de genes na cadeia evolutiva. Essa seleção interfere diretamente na relação filogenética entre as espécies (SCHATZ., 2010). Dessa forma torna-se importante o estudo aprofundado do genoma, para que se possa estabelecer a proximidade e a relação entre organismos (ABBY; DAUBIN, 2007).

Com o desenvolvimento da tecnologia os sequenciadores evoluíram e o trabalho que antes se realizava em um longo período de tempo, atualmente é muito mais rápido e preciso e produz resultados com maior qualidade e quantidade de informações. A técnica de sequenciamento realizada pelos sequenciadores de segunda geração, as mais utilizadas na atualidade, são capazes de produzir leituras curtas, de alta qualidade e de baixo custo (LIU et al., 2012; SCHATZ et al., 2010). O genoma é sequenciado em trechos de 35 a 400 pares de base com ampla cobertura possibilitando a montagem completa do genoma de um organismo, dependendo de sua complexidade (CARVALHO e SILVA, 2010; NAGAJARAN, 2013).

2.3 MONTAGEM DE GENOMAS

A montagem de genomas é um processo posterior à extração do DNA e sequenciamento do DNA. Para o sequenciamento, o DNA genômico é fragmentado resultando em fragmentos curtos com cerca de 100-500bp. Estes fragmentos são amplificados individualmente e sequenciados, resultando em um número elevado de sequências (leituras). Montagem é a técnica que utiliza das informações originadas pelas leituras dos trechos do genoma, efetuada pelos sequenciadores para recuperar *in silico* a sequência da molécula original (genoma). Como os trechos são redundantes, ou seja, uma mesma região do genoma é reconhecida em diversos

trechos da leitura, proveniente da ampla cobertura do sequenciamento, a molécula de DNA pode ser completamente montada, de maneira que todos os pares de bases nitrogenadas que a compõem tornem-se conhecidos (SCHATZ et al., 2010).

Por meio da comparação com outros organismos já sequenciados, é possível reconstruir a história evolutiva dos organismos. Para tanto, é necessário que se tenha um genoma de referência, sendo que esse deve ser de um organismo filogeneticamente próximo do organismo em estudo (STRYER et al., 2008).

São muitos os *softwares*, chamados montadores, disponíveis como ferramentas de montagem para que se possa ordenar de forma eficiente e correta as informações provenientes do sequenciamento. Cada montador utiliza uma estratégia diferente, que deve ser conhecida e comparada com outras ferramentas semelhantes, para que haja uma avaliação da montagem realizada. Entre os montadores disponíveis estão CLC Genomics (pacote de software comercial), SOAPdenovo, ALLPATHS-LG, Celera Assembler, entre outros (BRADNAM et al., 2013; PIRO et al., 2014; SCHATZ et al., 2010). A comparação de montagens proveniente de montadores distintos e a escolha dos montadores é fator relevante na acurácia da montagem final realizada, já que em alguns aspectos um montador pode ser mais eficiente que outro, e assim se fecham as lacunas que poderiam comprometer o processo de montagem do genoma (BRADNAM et al., 2013).

Muitos genomas bacterianos possuem um número variável de sequências repetidas que podem estar distribuídas “*in tandem*” (sequencialmente) ou de forma aleatória (ZAHA e FERREIRA, 2014). A existência dessas sequências é um fator complicador na montagem do genoma, já que se torna difícil estabelecer a posição desses trechos repetidos, bem como definir exatamente como estão configuradas essas sequências ao longo do genoma (JANSEN et al. 2002).

2.4 ANOTAÇÃO DE GENOMAS

A anotação genômica é um processo que ocorre após a montagem do genoma. Com a molécula de DNA montada e por meio de comparação com genomas de outros organismos, que já possuem a anotação de seu genoma, é possível encontrar genes e regiões específicas no DNA e inferir suas funções (SCHATZ et al., 2010).

2.5 FIXAÇÃO DE NITROGÊNIO

Existem alguns compostos químicos na natureza que são componentes essenciais para que a vida dos organismos se estabeleça, como nitrogênio, carbono, hidrogênio e oxigênio (NELSON e COX, 2000). Destes, o nitrogênio é o que está em menor quantidade, não sendo, por isso, menos importante. A maior quantidade de nitrogênio disponível na natureza está na atmosfera, onde totaliza em média 80% dos gases que a compõe (CAMPBELL, 2003). Nos organismos vivos muitas são as moléculas que contém nitrogênio em sua composição, como por exemplo: aminoácidos, nucleotídeos, ácidos nucléicos e polissacarídeos (MARIN et al., 1999).

A grande maioria das espécies não consegue absorver o nitrogênio do ar necessitando de uma forma combinada com amônia, nitrato ou aminoácidos (CAMPBELL, 2003). Somente algumas espécies de procariotos são capazes de transformar N_2 do ar em NH_3 que pode ser utilizada pelos demais organismos. Estas bactérias são essenciais para a vida na terra, pois tem papel fundamental no ciclo do nitrogênio na biosfera (NELSON e COX, 2000). A primeira etapa do ciclo do nitrogênio é realizada pela ação das bactérias fixadoras de nitrogênio, onde esse composto é reduzido à amônia. Essas bactérias são denominadas diazotrofos. A fixação de nitrogênio é realizada por uma ampla gama de bactérias, dentre essas se encontram bactérias aeróbicas, anaeróbicas, cianobactérias, bactérias microaeróbicas e fotossintéticas (BREDEMEIER e MUNDSTOCK, 2000; MARIN et al., 1999). Muitas bactérias fixadoras de nitrogênio são utilizadas na agricultura, onde a carência de nitrogênio impacta diretamente no crescimento e desenvolvimento das plantas, substituindo parcial ou totalmente o uso de fertilizantes (BREDEMEIER e MUNDSTOCK, 2000; MOREIRA et al., 2010). Estas bactérias em geral se associam com as raízes das plantas e por isto são conhecidas com rizobactérias, e entre estas bactérias que habitam as raízes das plantas algumas espécies podem desenvolver uma relação de simbiose e geram benefícios aos organismos vegetais (SPRENT e SPRENT, 1990).

Entre os gêneros de bactérias que capazes de fixar nitrogênio, destacamos o gênero *Derxia*, por ser o tema desse trabalho.

2.6 METABOLISMO DOS COMPOSTOS AROMÁTICOS

O metabolismo dos compostos aromáticos é um processo realizado pelas bactérias numa condição de falta de oxigênio, sendo possível a sua utilização como meio de crescimento por parte dos microorganismos (HARWOOD *et al.*, 1998). Existem muitos organismos capazes de realizar a degradação dos compostos aromáticos com maior ou menor eficiência e, dentre esses, as bactérias e os fungos são os organismos que possuem a maior habilidade (HEIDER, FUCHS; 1996).

Essa característica confere ao microorganismo uma maior capacidade adaptativa ao ambiente onde está inserido, já que por meio dessa via metabólica há a conversão dos compostos aromáticos em fonte energética para os demais processos celulares. (HEIDER, FUCHS; 1996). A degradação dos compostos aromáticos pode ser utilizada como ferramenta de biorremediação em substratos contaminados com resíduos industriais (BOUWER e ZEHNDER, 1993).

2.7 METABOLISMO DO NITRATO

O metabolismo do nitrato é realizado por alguns organismos como bactérias, fungos e plantas na resposta adaptativa à ambientes onde pode ocorrer variação da oferta de oxigênio. Esses organismos reduzem o nitrato à amônia para absorver o nitrogênio como fonte de energia para os mecanismos celulares, num processo denominado desnitrificação (ZHOU *et al.*, 2001).

2.8 SISTEMAS DE SECREÇÃO

Em bactérias a síntese de proteínas ocorre no meio intracelular e muitas dessas proteínas precisam circular do meio interno para o meio externo da célula. Algumas proteínas são pertencentes à membrana da célula, outras funcionam como proteínas de defesa da célula, ou ainda como forma de comunicação do meio externo com o meio interno, entre outras (OLIVER, 1985).

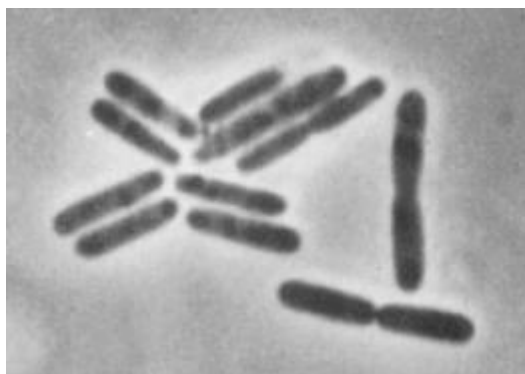
2.9 GÊNERO *Derxia*

O gênero *Derxia* foi inicialmente criado com o descobrimento da bactéria *Derxia gummosa*, a primeira espécie pertencente a esse gênero, caracterizada como uma bactéria fixadora de nitrogênio, fator que impulsionou o seu estudo (ROY e SEN, 1962; JENSEN et al., 1960; XIE e YOKOTA, 2004).

Esta bactéria é aeróbica, portadora de células gram negativas, apresenta relativa variação morfológica, com extremidades arredondadas em forma de bastonetes como mostrado na figura 1, mede 3-6 X 1,0-1,2 μ m e movimenta-se com o auxílio de um flagelo polar curto. O gênero *Derxia* possui relação com os seguintes gêneros de bactérias diazotróficas: *Azotobacter*, *Azomonas*, *Beijerinckia*, e *Pseudomonas*, quando comparados considerando morfologia, fisiologia e características quimiotaxonômicas, sendo também considerada uma bactéria diazotrófica de vida livre (BECKING et al., 1991; MOREIRA et al., 2010; HILL e POSTGATE, 1969; JENSEN et al., 1960; SHANKAS et al., 2006).

Quando comparada considerando-se a similaridade entre os genes do RNA ribossomal 16S, *Derxia* forma um novo grupo monofilético, pertencente às β -proteobactérias, distinto dos demais (XIE e YOKOTA, 2004).

FIGURA 1. CULTIVO DE *D. gummosa* DSM 723



D. gummosa DSM 723 CULTIVADA EM MEIO ÁGAR GLUCOSE DEFICIENTE EM NITROGÊNIO, APÓS DOIS DIAS DE CULTIVO À 25°C.
 FONTE: JENSEN et al. 1960

2.10 *Derxia gummosa* DSM 723

D. gummosa DSM 723 é a primeira bactéria pertencente ao gênero *Derxia*, encontrada e isolada de solos tropicais (XIE e YOKOTA, 2004). Quando descrita, a bactéria foi isolada de solos provenientes da Índia (JENSEN et al., 1960). No Brasil

existe a ocorrência dessa espécie bacteriana, sendo componente de solos úmidos e já foi encontrada nos seguintes Estados: Pará, Pernambuco, Rio de Janeiro e São Paulo. A ocorrência da espécie no Brasil está diretamente relacionada às condições de umidade dos solos coletados (CAMPELO e DOBEREINER, 1970).

Nos cultivos *in vitro*, as colônias apresentam contornos bem definidos, circulares ou ovais, possuindo uma depressão no centro, com aparência translúcida, de coloração marrom-amarelada e possuem constituição gelatinosa (CAMPELO e DOBEREINER, 1970; HILL e POSTGATE, 1969; JENSEN et al., 1960). São cultivadas em temperaturas de 25°C a 35°C (ROY e SEN, 1962). Quando primeiramente descritas, foram cultivadas em placas contendo meio de cultivo ágar manitol deficiente em nitrogênio, porém podem ser cultivadas em meios distintos (JENSEN et al., 1960; XIE; YOKOTA, 2004).

D. gummosa DSM 723 teve sua sequência genômica parcial depositada no NCBI (National Center for Biotechnology Information). As características gerais desta sequência genômica podem ser mais bem visualizadas na tabela 1 (<http://www.ncbi.nlm.nih.gov/bioproject/PRJNA188859> acesso em 01/11/2016). A plataforma de sequenciamento utilizada foi Illumina HiSeq 2000.

TABELA 1 CARACTERÍSTICAS DO GENOMA DE *Derxia gummosa* DSM 723, DEPOSITADO NO NCBI.

Característica	Definição
Finalização	Incompleta
Tamanho do genoma	5.185.168 pb
GC%	69,90
Nº de Scaffolds	15
Nº de Contigs	19
Nº de Genes	3.367
Nº de Proteínas	3.305
N50	664.439
Tecnologia de sequenciamento	Illumina HiSeq 2000

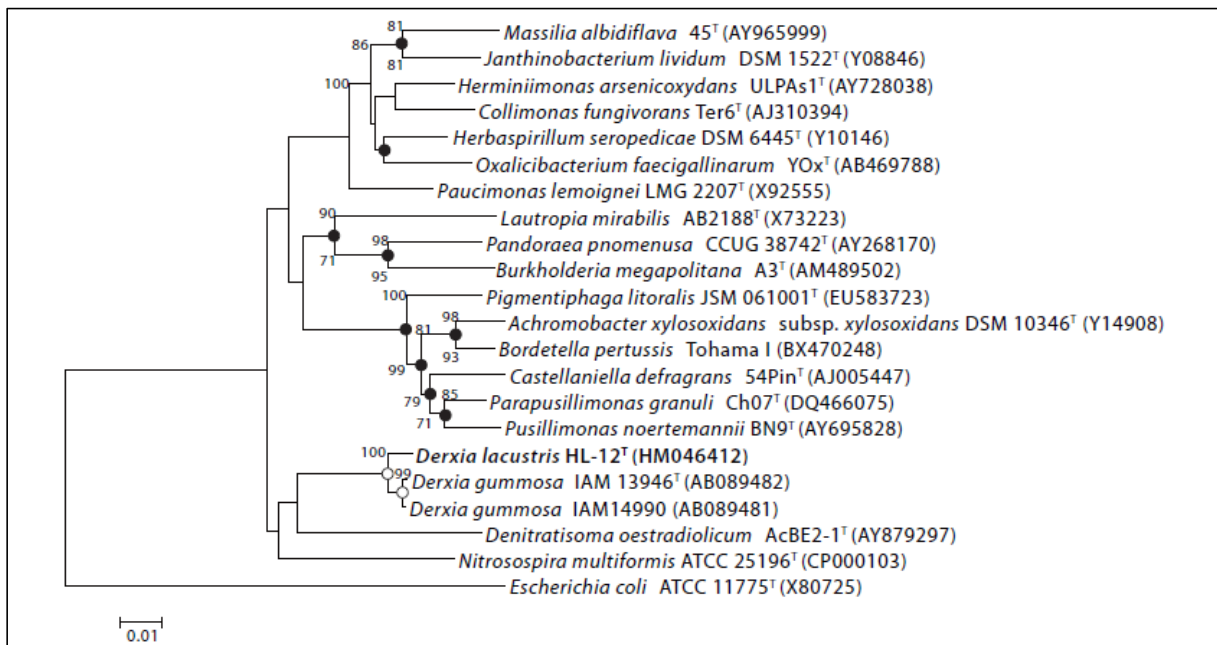
FONTE: A autora (2016).

2.11 *Derxia lacustris* HL12

D. lacustris HL12 é uma nova espécie do gênero *Derxia* que foi descoberta e descrita em 2013. O microrganismo foi isolado a partir de amostras de água provenientes de um lago em Taiwan. Essas amostras foram gentilmente cedidas pelo grupo de Chen e colaboradores. Suas células são gram negativas, aeróbicas, móveis e não formadoras de esporos. Suas colônias são bem delimitadas, rodeadas por uma cápsula densa, com coloração amarelada, de aspecto translúcido e brilhante. É um organismo capaz de realizar fixação de nitrogênio em condições microaeróbicas (CHEN et al., 2013).

D. lacustris HL12 ainda não teve seu genoma sequenciado. Este microrganismo foi diferenciado de *D. gummosa* por comparações com a sequência do gene do RNA ribossomal 16S. A árvore filogenética baseada na sequência de 16SrRNA da figura 2 mostra que as duas espécies separadas em ramos próximos, mas distintos (CHEN et al., 2013).

FIGURA 2. ÁRVORE FILOGENÉTICA BASEADA NAS SEQUÊNCIAS DO GENE 16S rRNA



A ÁRVORE FILOGENÉTICA BASEADA NO GENE DO 16S rRNA MOSTRA A RELAÇÃO ENTRE OS TÁXONS PERTENCENTES À CLASSE DAS β-PROTEOBACTÉRIA.

FONTE: CHEN et al., 2013

2.12 GENOMAS PROCARIÓTICOS

O genoma dos procariotos possui características próprias quando comparados com os genomas dos organismos eucariotos. Os organismos procariotos previamente sequenciados, mostram que o tamanho de seu genoma é bastante variável podendo ir de 150.000 pares de bases até 13.000.000 de pares de bases (CAMPBELL, 2003).

Os genes são as unidades informacionais contidas no genoma cujos produtos de expressão (RNA funcional ou proteína) tem um função celular específica (CAMPBELL, 2003). O genoma de procariotos têm uma alta densidade gênica, ou seja, alta concentração de genes quando comparado ao genoma dos organismos eucariotos (STRYER et al., 2008; ZAHA e FERREIRA, 2014). Uma etapa fundamental da anotação da sequência genômica é a predição dos genes e as funções proteicas.

Em bactérias o material genético em geral está contido em uma molécula de DNA circular que é o cromossoma bacteriano (CAMPBELL, 2003; NELSON e COX, 2000). Muitas espécies bacterianas têm ainda contêm replicons extracromossomais que tem tamanho muito variado, desde alguns milhares de pares de bases até milhões de pares de base.

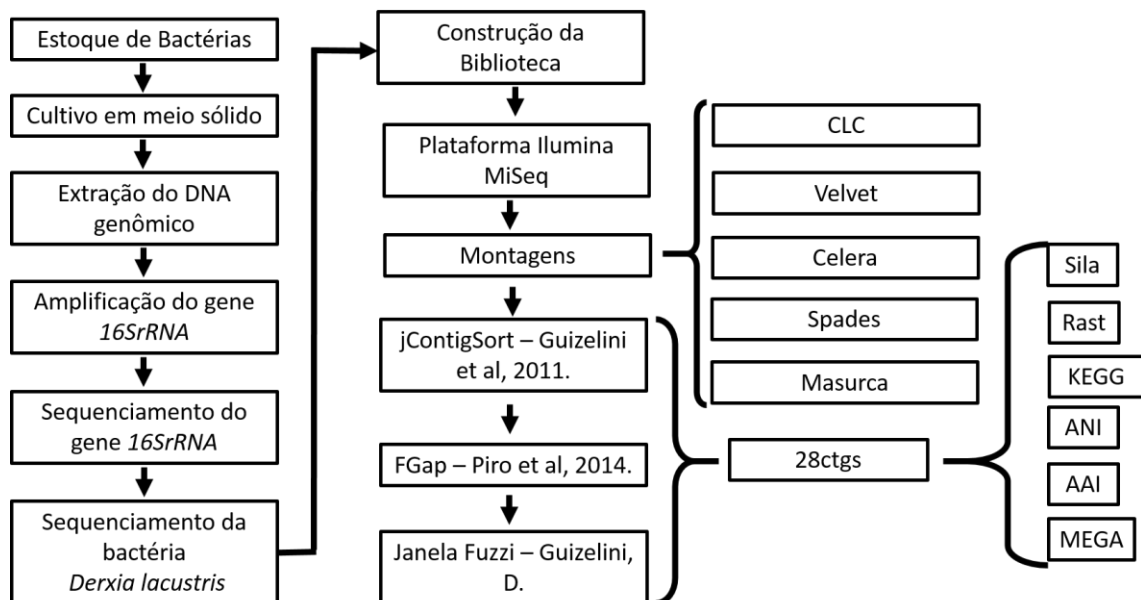
3 METODOLOGIA

O estudo do genoma da bactéria *D. lacustris* HL12 se iniciou com um estoque de bactérias cultivadas, provenientes de Taiwan, recebido da equipe colaboradora desse trabalho (CHEN *et al.*, 2013). As etapas deste trabalho estão apresentas no fluxograma da figura 3.

O estoque de bactérias foi replaqueado em meio de cultivo NFb e incubada em estufa à 30°C por 48h. Colônias foram raspadas da placa e o DNA total foi extraído com fenol clorofórmio, precipitado e lavado com etanol e solubilizado em água ultrapura.

O gene do 16S rRNA foi sequenciado separadamente para a confirmação da identidade do organismo como sendo *D. lacustris* HL12. O genoma da bactéria foi sequenciado na plataforma Illumina MiSeq, em duas corridas. Os dados provenientes do sequenciamento seguiram para a montagem com os seguintes softwares montadores: CLC, Velvet, Celera, Spades e Masurca. O conjunto de dados provenientes das cinco montagens foi tratado com a ferramenta jContigSort. Para o fechamento dos gaps foi utilizada a ferramenta FGap. A ferramenta Janela Fuzzi foi utilizada no processo de refragmentação do genoma para a reordenação.

FIGURA 3 FLUXOGRAMA DE ESTUDO DO GENOMA DE *Derxia lacustris* HL12.



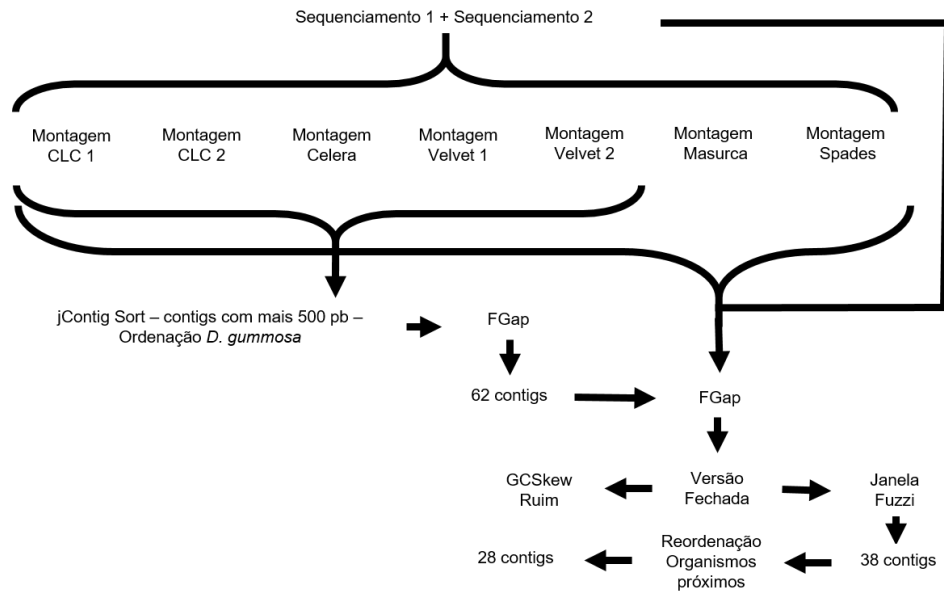
O FLUXOGRAMA A ORDEM DAS ATIVIDADES REALIZADAS NO ESTUDO DE *Derxia lacustris* HL12
 FONTE: A autora (2016).

A versão final do genoma de *D. lacustris* HL12 possui 28 contigs. O genoma foi anotado utilizando as ferramentas Sila, Rast e Kegg. Durante a análise de filogenia foram utilizadas as ferramentas ANI, AAI e o *software* Mega.

3.1 ESTRATÉGIA DE MONTAGEM

A montagem do genoma da bactéria *D. lacustris* HL12 se iniciou com os dados provenientes de duas corridas realizadas no sequenciador Illumina MiSeq. As sequências foram montadas com os seguintes *softwares*: CLC, Velvet, Celera, Masurca e Spades. Os dados provenientes das primeiras montagens com CLC, Celera e Velvet foram tratadas com a ferramenta jContigSort, que ordenou os contigs com o genoma de *D. gummosa* DSM 723 como referência e apenas os contigs com mais de 500pb foram selecionados. Foi utilizada a ferramenta FGAP para fechar os *gaps* possíveis, utilizando apenas as cinco primeiras montagens como *datasets*. Ao fim desse processo obtivemos 62 contigs. Esses 62 contigs foram novamente tratados com o FGAP, nessa etapa foram utilizados como *datasets* todas as sete montagens e os dados das duas rodadas de sequenciamento. O genoma obtido após essa etapa não possuía nenhum gap, porém a análise do gráfico de GCskew, mostrava essa como uma versão de baixa qualidade. Por essa razão o genoma foi tratado utilizando a ferramenta Janela Fuzzi que refragmentou o genoma nos picos de inversão do gráfico GCskew. Após essa etapa o genoma possuía 38 *contigs*. Os organismos pertencentes à árvore filogenética de Chen e colaboradores (CHEN et al., 2013), que possuíam genomas completos depositados no NCBI, ou *drafts* foram utilizados num processo de reordenação do genoma em sucessivas tentativas de fechar os gaps existentes com o FGAP (Figura 4).

FIGURA 4 FLUXOGRAMA DE MONTAGEM DO GENOMA DE *Derxia lacustris* HL12.



ESTRATÉGIA DE MONTAGEM DO GENOMA DE *Derxia lacustris* HL12
 FONTE: A autora (2016).

4 ARTIGO CIENTÍFICO

O manuscrito abaixo é a primeira versão referente ao trabalho de sequenciamento, montagem e anotação do genoma de *Derxia lacustris*, o mesmo descreve e discute os resultados obtidos nesta dissertação de mestrado e deve ser submetido para a revista Standart in Genomic Science. Esse periódico possui fator de impacto 3,167 e foi escolhido por apresentar acesso rápido e ampla divulgação de artigos relacionados com genomas e metagenomas.

O artigo a ser submetido está relacionado com a publicação da sequência incompleta do genoma de *D. lacustris* HL12.

Draft genomic sequence of *Derxia lacustris* HL12

Sheyla Trefflich¹, Vinicius A. Weiss¹, Dieval Guizelini¹, Roberto T. Raittz¹, Maria B. R. Steffens^{1,2}, Michele Z. Tadra-Sfeir², Helisson Faoro³, Valter A. Baura², Arnaldo Glogauer⁴, When-Ming Chen⁵, Wei-Cheng Huang⁶, Shih-Yi Sheu⁶, Euan K. James⁷, Fabio O. Pedrosa^{1,2}, Emanuel M. Souza^{1,2}

ABSTRACT

Derxia lacustris HL12 is a gram-negative bacteria from the class β -proteobacteria, family *Alcaligenaceae*, collected from fresh water. Along with *D. gummosa* DSM 723, it is phylogenetically placed within the genus *Derxia*. Its genomic sequence has approximately 5.4 Mpb with a G + C content of 69,20% and 4696 protein-coding genes. In this study, it is presented the draft of *D. lacustris* HL12 genome, bacteria capable of biologic nitrogen fixation.

KEYWORDS: Gram-negative, Fixação de Nitrogênio, Genome Sequence, *Derxia lacustris* HL12

INTRODUCTION

The existence of rhizobacteria is of great importance to agriculture, where the need for nitrogen affects directly the crop growth and development. This creates many environmental and economic problems such as those caused by the usage of fertilizers (1,2). Those bacteria inhabit roots and benefit plants (3). There is in the nature a large number of species of bacteria that are capable of biological nitrogen fixation, amongst which the genus *Derxia* is highlighted, whose species *Derxia lacustris* HL12 is the object of study of this paper.

Derxia was first described by Jensen et al. in 1960 who described the *D. gummosa* DSM 723 species isolated from soil samples. The genus belongs to the class *Betaproteobacteria*, order *Burkholderiales*, family *Alcaligenaceae* (4). Strains of this species have been found in Bengal West. Despite the wide distribution of *Derxia gummosa*, this bacterium is relatively unstudied. *Derxia* was found in Brazil on this regions: Pernambuco, Pará, Rio de Janeiro e São Paulo (6). *D. gummosa* DSM 723 grow on solid medium producing large amounts of gum, which may have

biotechnological potential. Since *D. gummosa* DSM 723 is frequently found in the rhizosphere and is capable of nitrogen fixation, there has been some interest in this species as plant growth promoter.

TABLE 1 CLASSIFICATION AND GENERAL FEATURES OF *Derxia lacustris* HL12

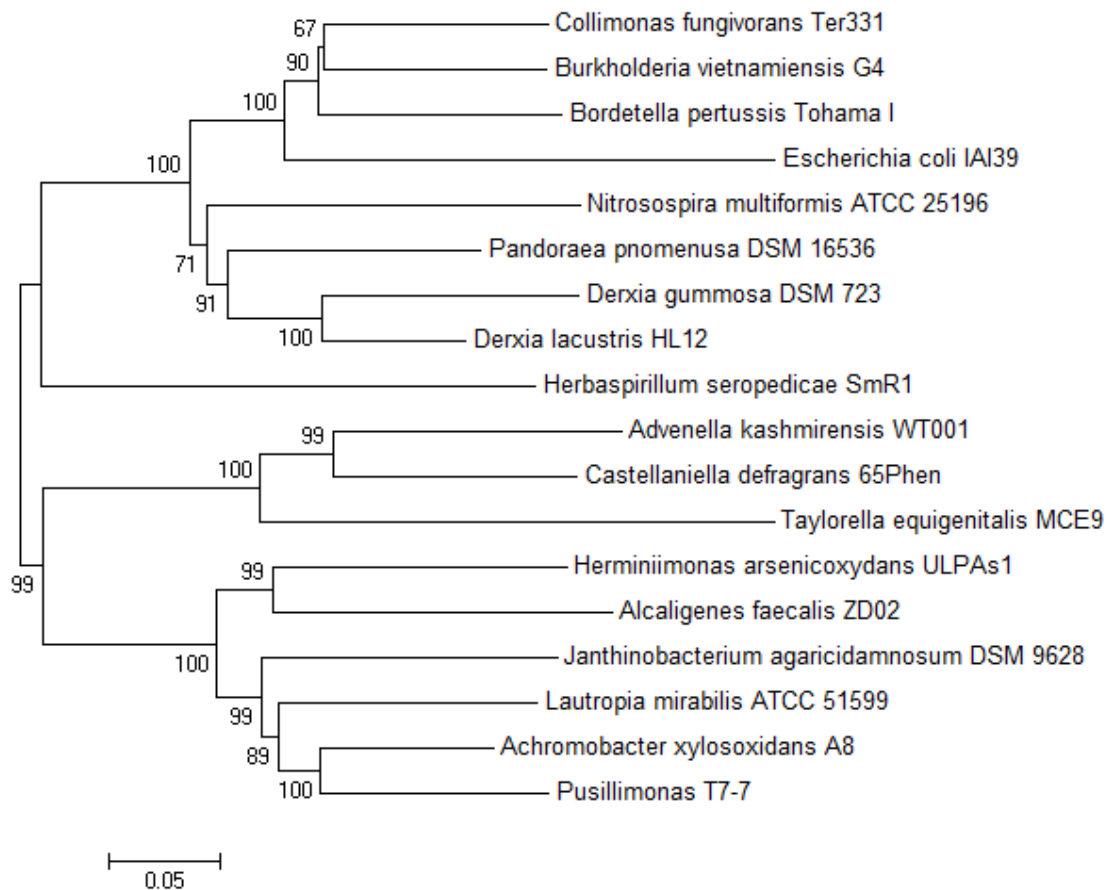
MIGS ID	Property	Term	Evidence code ^a
	Classification	Domain <i>Bacteria</i>	TAS (4)
		Phylum <i>Proteobacteria</i>	TAS (13)
		Class <i>Betaproteobacteria</i>	TAS (4)
		Order <i>Burkholderiales</i>	TAS (4)
		Family <i>Alcaligenaceae</i>	TAS (4)
		Genus <i>Derxia</i>	TAS (4)
		Species <i>Derxia lacustris</i>	TAS (4)
		Type strain: HL-12	
	Gram stain	negative	TAS (4)
	Cell shape	rod	TAS (4)
	Motility	motile	TAS (4)
	Sporulation	not reported	
	Temperature range	15-40°C	TAS (4)
	Optimum temperature	25-35°C	TAS (4)
	pH range; Optimum	6–7; 6	TAS (4)
	Carbon source	Glucose, Mannitol	TAS (4)
MIGS-6	Habitat	fresh water	TAS (4)
MIGS-6.3	Salinity	0-4% NaCl (w/v), 0-1% optimal	TAS (4)
MIGS-22	Oxygen requirement	Aerobic	TAS (4)
MIGS-15	Biotic relationship	not reported	
MIGS-14	Pathogenicity	not reported	
MIGS-4	Geographic location	Heaven Lake, Kaohsiung County, Taiwan	TAS (4)
MIGS-5	Sample collection	not reported	TAS (4)
MIGS-4.1	Latitude	23°16'39.50" N	TAS (4)
MIGS-4.2	Longitude	120°54'53.70' E	TAS (4)
MIGS-4.4	Altitude	not reported	

^aEvidence codes - IDA: Inferred from Direct Assay ; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project (35)

Despite being frequently found in microbial biodiversity surveys, until recently the only species of the genus was *D. gummosa* DSM 723. In 2013 Chen et al. described a new bacteria isolated from lake water in Korea which was allocated to

Derxia genus and was described as a new species, *D. lacustris* HL12. This new species is also a nitrogen fixing bacterium but seems to have distinct ecological niches. The knowledge of its genomic characteristics may help to understand the life style of these bacteria and what benefits they might promote in the inhabited environment.

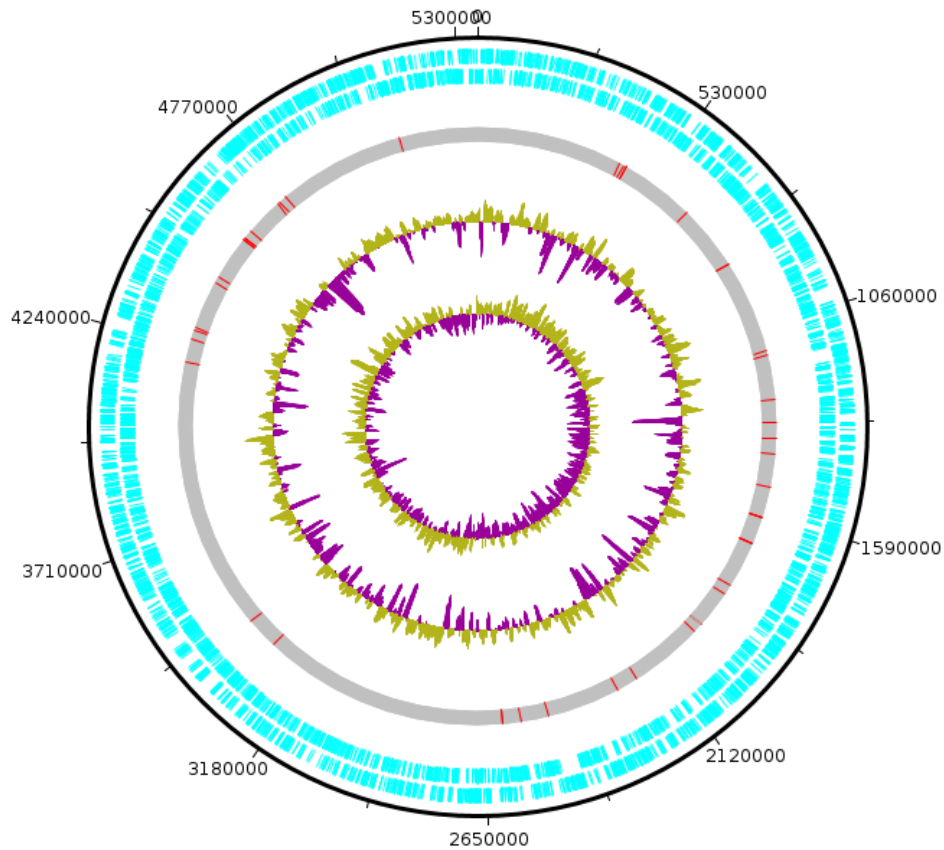
FIG. 1 NEIGHBOR-JOINING PHYLOGENETIC TREE WAS BUILT WITH THE MLSA TECHNIQUE.



NEIGHBOR-JOINING PHYLOGENETIC TREE WAS BUILT WITH THE MLSA TECHNIQUE USING 16S RRNA, GYRB, RECA, RPOB AND RPOD GENES SEQUENCES FROM NCBI. THOSE WERE TRIMMED BY GBLOCKS TOOL (18, 19). THE SEQUENCES WERE ALIGNED AND THE TREE WAS BUILT USING MEGA SOFTWARE VERSION 6.06 WITH 1,000 BOOTSTRAPS.

The draft genome of *D. gummosa* DSM 723 is deposited at the NCBI (8). It contains 19 contigs and 15 scaffolds. In this work the genome of *D. lacustris* HL12 was sequenced and analyzed. The results confirmed that *D. lacustris* HL12 is a distinct species, but despite the difference in ecological niche the genomes of the two species share extensive similarity.

FIG. 2 GRAPHICAL CIRCULAR MAP OF A SIMULATED DRAFT *Derxia lacustris* HL12 GENOME.



THE MAP WAS GENERATED WITH (REF). FROM OUTSIDE TO THE CENTER: THE FIRST TWO OUTER CIRCLES REPRESENT THE POSITION OF GENES(CIRCLE1: PLUS STRAND, CIRCLE2:MINUS STRAND) CIRCLE 3 REPRESENTS TRNA GENES (RED), CIRCLE 4 REPRESENTS G + C CONTENT, AND CIRCLE 5 REPRESENTS GC SKEW.

ORGANISM INFORMATION

CLASSIFICATION AND FEATURES

D. lacustris HL12 is a new species of the genus *Derxia* and was originally isolated from fresh water from a lake in Taiwan (4). It is a gram negative bacterium, it is aerobic, it forms very well delimited colonies, it has a yellowish color, it has dense, translucent, and bright capsule, and its general characteristics are listed in Table 1 (4). It is an organism capable of biological nitrogen fixation in micro aerobic conditions. It has mobile cells and does not produce spores (4). The bacteria were cultivated in NFbPN medium containing 1% of glucose instead of malate and incubated at 35°C for 48h.

Phylogenetic tree was constructed using neighbor-Joining method by MLSA technique (Fig. 1). Average nucleotide identity (ANI) and average amino acid identify (AAI) was determined (Fig. S1) using the online tool (38-41) between the genomes analyzed in MLSA phylogenetic tree (36). Results show that this bacteria share low ANI and AAI value (<95%, including in comparison with *D. gummosa* DSM723).

GENOME SEQUENCING INFORMATION

GENOME PROJECT HISTORY

D. lacustris HL12 genome was sequenced using the platform Illumina MiSeq. Two independent sequence runs were performed producing 3,528,940 paired-end reads. The draft genome sequence of *D. lacustris* HL12 genome consists of 28 contigs.

GROWTH CONDITIONS AND GENOMIC DNA PREPARATION

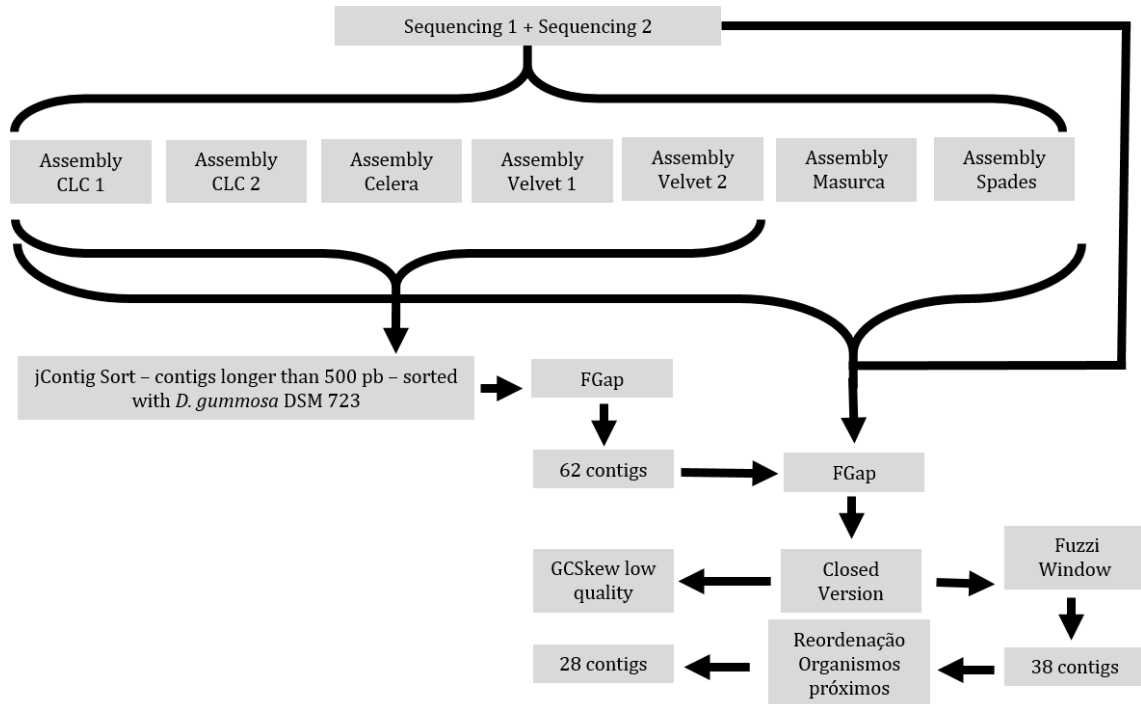
D. lacustris HL12 was grown aerobically in NFb medium at 30°C for 48h. The DNA was extracted by lysing the cell in 50 mM Tris.HCl buffer pH 8.0 containing 1% SDS and 10 mM EDTA. The lysate was extracted with phenol-chloroform and precipitated with ethanol (37). Purified DNA was evaluated spectrophotometrically and by agarose gel electrophoresis.

GENOME SEQUENCING AND ASSEMBLY

The genome of *D. lacustris* HL12 was sequenced using the Illumina Miseq sequencer (10). Sequencing libraries were constructed with Nextera XL kit. Two Illumina standard shotgun libraries were constructed and sequenced using the Illumina MiSeq, generating 3,528,940 reads. After the sequencing process, low quality reads were removed, including adapters, reads whose quality scores were below 20 and reads with less than 25pb. The genome has 108-fold genome coverage. For genome assembly 5 different assemblers were used, namely, CLC Workbench Genomics v6.5.1, Spades v3.5.0 (42), Masurca v2.3.2 (43), Velvet v1.2.08 (47) and Celera (44). The following tools were used in treatment, re-fragmentation and closure of the gaps: jContigsort (11), Fuzzi Window (unpublished data) and Fgap(12).

The genome was analyzed using the software Consed (20) and Bowtie v1.1.1(45).

FIG. 3 *Derxia lacustris* HL12 GENOME ASSEMBLY SKETCH.



FIRSTLY, THE GENOME WAS BUILT WITH THE FOLLOWING SOFTWARES: CLC (ASSEMBLY CLC1, ASSEMBLY CLC2), CELERA (ASSEMBLY CELERA), VELVET (ASSEMBLY VELVET1, ASSEMBLY VELVET2). DATA WERE TREATED BY JCONTIGSORT TOOL (11) AND POSSIBLE GAPS WERE CLOSED WITH FGAP TOOL (12) RESULTING IN 62 CONTIGS. LATER, TWO ADDITIONAL ASSEMBLIES WERE MADE WITH MASURCA SOFTWARE (ASSEMBLY MASURCA) AND WITH SPADES (ASSEMBLY SPADES). THE 62 CONTIGS HAD THEIR GAPS CLOSED USING DATA GATHERED FROM ALL THE ASSEMBLIES AS DATASET AND THE GENOME WAS CLOSED. THE CLOSED VERSION HAS GCSKEW WITH LOW QUALITY. THE CLOSED VERSION WAS TREATED BY THE TOOL JANELA FUZZI (REF) AND FGAP, RESULTING IN 38 CONTIGS. THE GENOME WAS REARRANGED WITH CLOSE ORGANISMS AND THE POSSIBLE GAPS WERE CLOSED BY FGAP, ORIGINATING THE FINAL VERSION WITH 28 CONTIGS.

The best assemblage of the *D. lacustris* HL12 genome was obtained by using the software Celera producing 129 contigs. The contigs/scaffolds were ordered using the jContigSort (11) that selected in this analysis only contigs having more than 500pb were used. The sequences ordered using *D. gummosa* DSM 723 genome as reference were used for gap closing with Fgap (12). The contigs were also ordered using the genome sequence of organisms phylogenetically close to *D. lacustris* HL12 as reference (4) and each time Fgap was used to find gap filling sequences (Fig. 3). Gaps were considered closed after visual inspection of the aligned sequences and Blast analysis of the gap region. The final genome assembly contained 28 contigs and a genome size of 5.34Mpb.

GENOME ANNOTATION

The software Sila (14) and Rast (15) were used for genome annotation. The following software were used for the visualization of the genes related to specific metabolic pathways: SEED Viewer version2.0 (16) and Artemis(17). Bowtie version 1.1.1 and Tablet version 14.10.20 (21) were used for the scanning analysis.

GENOME PROPERTIES

D. lacustris HL12 genome is 5,348,005 bp long, has an average GC content of 69.20%. The genome properties and statistics are summarized in the Table 2 and Figure 2.

A total of 4,696 protein-coding genes were identified and 1,812 of them were distributed into COG functional categories (Table 3).

D. lacustris HL12 genome has a complete ribosomal operon with 5,125 bp long and its coverage was 2,529 fold. The 5S rRNA gene appears as a single in two different location of the genome whilst genes for the other rRNA species occurred only within the complete operon.

TABLE 2 Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	5,348,005	100
DNA coding (bp)	4,874,334	91.14
DNA G+C (bp)	3,700,819	69.20
Total genes	4696	100.00
Protein coding genes	4644	98.89
RNA genes	52	1.10
rRNA operons	1	0.09
Genes with function prediction	215	4.57
Genes assigned to COGs	3111	66.24
Genes with Pfam domains	3796	80.83
Genes with signal peptides	323	6.87
Genes with transmembrane helices	4132	87.98
CRISPR repeats	12	0.25

INSIGHTS FROM THE GENOME SEQUENCE

PROFILES OF METABOLIC PATHWAY

Nitrogen is one of the essential chemical compounds in the formation of living organisms, taking part in aminoacids, proteins and nucleic acids compositions (22, 23). The majority of the nitrogen available in nature is dispersed in the atmosphere as nitrogen gas which cannot be metabolized by most organisms (22, 23).

TABLE 3 NUMBER OF GENES ASSOCIATES WITH COG FUNCTIONAL CATEGORIES

Code	Value	% age	Description
J	136	5,46	Translation, ribosomal structure and biogenes
A	2	0,08	RNA processing and modification
K	83	3,33	Transcription
L	59	2,37	Replication, recombination and repair
D	20	0,80	Cell cycle control, cell division, chromosome
V	26	1,04	Defense mechanisms
T	74	2,97	Signal transduction mechanisms
M	73	2,93	Cell wall/membrane/envelope biogenesis
N	36	1,45	Cell motility
U	48	1,93	Intracellular trafficking, secretion
O	79	3,17	Posttranslational modification, protein
C	167	6,71	Energy production and conversion
G	86	3,46	Carbohydrate transport and metabolism
E	223	8,96	Amino acid transport and metabolism
F	51	2,05	Nucleotide transport and metabolism
H	82	3,29	Coenzyme transport and metabolism
I	81	3,25	Lipid transport and metabolism
P	119	4,78	Inorganic ion transport and metabolism
Q	51	2,05	Secondary metabolites biosynthesis, transport
R	182	7,31	General function prediction only
S	134	5,38	Function unknown
-	3111	66,25	Not in COGs

^aThe total is based on the number of protein coding genes in the annotated genome

The reduction of nitrogen into ammonium by diazotrophic bacteria through a process called nitrogen fixation and is made by the nitrogenase complex (22, 23). Nitrogen fixation genes were found in *D. lacustris* HL12 genomic sequence and

compared with *D. gummosa* DSM 723 and to *Herbaspirillum seropedicae* SmR1. Figure S1 shows a comparative organization of *nif* genes in these organisms. The *D. lacustris* HL12 and *D. gummosa* DSM 723, *nif* gene clusters are highly co-linear with gene and share highly identity. The differences were observed between the clusters of *D. gummosa* DSM 723 and *D. lacustris* HL12, *D. gummosa* DSM 723 has two additional genes in *nif* gene cluster. There is a high conservation between *Derxia* and *Herbaspirillum seropedicae* *nif* genes. However, organization is less conserved with group of genes or operon shifting positions. Figure S2 shows the main nitrogen incorporation pathways in *D. lacustris* HL12 as deduced from the genome sequence.

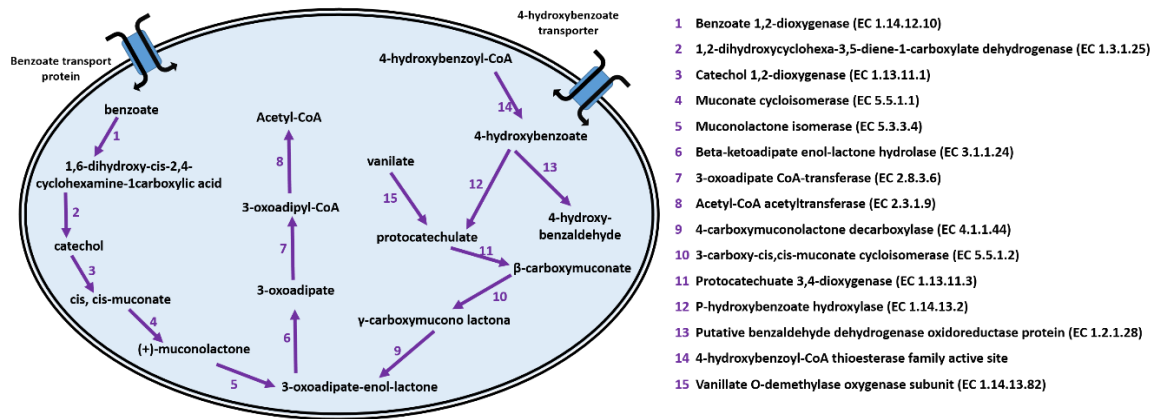
D. lacustris HL12 can reduce nitrate to ammonia and use as nitrogen source (Figure S2)(23). In the *D. lacustris* HL12 genomic sequence, the genes related to nitrate metabolism are the cytoplasmic assimilatory nitrate reductase gene (*nasA*), nitrite reductase (*nirB* and *nirD*), nitrate carrier genes (*nrtC* and *nrtD*), nitrite efflux gatekeeper encoding gene (*nark*) (24), nitric oxide sensibility repression gene (*nsrR*) and transmembrane nitrate carrier (*cynB*) (24, 23). Three nitrate metabolism regulatory genes were identified: *narXL* encoding a nitrate responsive two component regulatory system and *nasT* encoding a ANTAR containing domain Nitrate metabolism genes are clustered in the genome (fig. 4).

FIG. 4 ARRANGEMENT OF THE GENES LINKED TO NITRATE METABOLISM IN *Derxia lacustris* HL12



The degradation of aromatic compounds is an important feature in environmental bacteria, for they enhance the organism's fitness (25). Aromatic compounds can also be used as growth substrate by bacteria (25, 26). *D. lacustris* HL12 has genes related to the metabolism of aromatic compounds. In comparison to the metabolic pathway of *Herbaspirillum seropedicae* (26), *D. lacustris* HL12 has enzyme genes related to metabolism of benzoate, hydroxybenzoate, and dihydroxybenzoate in acetyl-CoA (fig. 5), but not acid nicotinic.

FIG. 5 PROPOSED PATHWAYS FOR AROMATIC COMPOUNDS METABOLISMO IN *Derxia lacustris* HL12, BASED ON THE PROPOSED OF PEDROSA et al., 2011.



In bacteria, the synthesis of protein occurs in the intracellular medium, though many of those proteins are intended for the extracellular environment. Those proteins circulate in a quick and precise way out of the cell (27). The protein secretion systems perform such task, granting the bacteria specific characteristics. Genes related to type I secretion systems (ABC transporters) were found in *D. lacustris* HL12 genome. Genes related to type II secretion system were also found. This system is linked to the virulence capability of the bacteria, exportation of fimbrial proteins and toxins, which is an important factor for the anchoring of the cell to other membranes (29). Genes related to type IV pili secretion systems are present, being responsible for the capability of the cell to attach to surfaces, mobility, formation of biofilms, virulence, and protein secretion, besides being linked to interactions between bacteria and eukaryotic cells (30,31). *D. lacustris* HL12 has genes related to protein secretion systems type VI, which is a form of secretion easily found in gram-negative bacteria that is important to, but not exclusive to, pathogenic bacteria. Genes related to type III secretion system and type V secretion systems were not found.

CONCLUSIONS

The analysis of *D. lacustris* HL12 genome showed that this organism represents a new species from the genus *Derxia*, apart from *D. gummosa* DSM 723. This bacterium is capable of biological nitrogen fixation, resembling *D. gummosa* DSM 723.

The organism can assimilate nitrate and has an assortment of protein secretion systems (type I, II, VI and pili IV type).

Additional file

Additional file 1: Table S1. ANI and AAI analyses of *D. lacustris* HL12 genome.

Figure S1. Comparison of *nif* gene cluster of *D. lacustris* HL12, *D. gummosa* DSM 723 and *Herbaspirillum seropedicae* SmR1.

Figure S2. Nitrogen metabolism pathways identified in *D. lacustris* HL12 genome.

ACKNOWLEDGEMENTS

This work was financially supported by the CNPq (National Council for Scientific and Technological Development), CAPES (Higher Education Personnel Improvement Coordination), INCT- Biological Nitrogen Fixation (Program National Institutes of Science and Technology) and by Araucária Foundation (Support for scientific and technological development of Paraná).

AUTHOR DETAILS

¹Graduate Program on Bioinformatics, Federal University of Parana, 81520-260, Brazil

² Department of Biochemistry and Molecular Biology, Federal University of Parana, UFPR, 81531-990, Brazil

³Carlos Chagas Institute, Oswaldo Cruz Foundation of Paraná, 81350-010, Brazil

⁴Institute of Technology of Paraná, Agency of Innovation Tecpar, 81350-010, Brazil

⁵Laboratory of Microbiology, Department of Seafood Science, National Kaohsiung Marine University, No. 142, Hai-Chuan Rd. Nan-Tzu, Kaohsiung City 811, Taiwan, ROC

⁶Department of Marine Biotechnology, National Kaohsiung Marine University, No. 142, Hai-Chuan Rd. Nan-Tzu, Kaohsiung City 811, Taiwan, ROC

⁷Department of Biological Sciences, University of Dundee DD1 4HN, UK

REFERENCES

1. BREDEMEIER, Christian; MUNDSTOCK, Claudio Mario. REGULAÇÃO DA ABSORÇÃO E ASSIMILAÇÃO DO NITROGÊNIO NAS PLANTAS. **Ciência Rural**, Santa Maria, v. 30, n. 2, p.365-372, 2000.
2. MOREIRA, Fatima Maria de Souza et al. Bactérias diazotróficas associativas: diversidade, ecologia e potencial de aplicações. **Comunicata Scientiae**, Terezina, v. 2, n. 1, p.74-99, 20 2010.
3. SPRENT, J.I.; SPRENT, P. Nitrogen fixing organisms. London: Chapman and Hall, 2ed., 1990. 256p.
4. CHEN, W.-m.; HUANG, W.-c.; SHEU, S.-y.. *Derxia lacustris* sp. nov., a nitrogen-fixing bacterium isolated from a freshwater lake. **International Journal Of Systematic And Evolutionary Microbiology**, (s.l.), v. 63, n. 3, p.965-970, 1 jun. 2012. Microbiology Society. DOI: 10.1099/ij.s.0.041186-0.
5. JENSEN, H. L. et al. A New Nitrogen-fixing Bacterium. **Archiv Für Mikrobiologie**, Berlin, v. 195, n. 36, p.182-195, 1960.
6. CAMPELO, Adalis Bezerra; DOBEREINER, Johanna. OCORRÊNCIA DE *Derxia* sp. EM SOLOS DE ALGUNS ESTADOS BRASILEIROS. **Pesq. Agropoc. Bear**, Brasília, v. 5, p.327-332, 1970.
7. HILL, Susan; POSTGATE, J. R. Failure of Putative Nitrogen-fixing Bacteria to Fix Nitrogen. **J. Gen. Microbiol.**, Grã Bretanha, v. 58, p.277-285, 1969.
8. NCBI, National Center For Biotechnology Information -. **Genome of *D. gummosa* DSM 723.** Disponível em: <http://www.ncbi.nlm.nih.gov/assembly/GCF_000482785.1/>. Acesso em: 01 jan. 2016.
9. XIE, Cheng-hui; YOKOTA, Akira. Phylogenetic analyses of the nitrogen-fixing genus *Derxia*. **J. Gen. Appl. Microbiol**, Tokyo, v. 50, p.129-135, 2004.
10. Bennett, S. Solexa Ltd. Pharmacogenomics. v. 5, p.433-438, 2004.
11. GUIZELINI, Dieval ; PEDROSA, Fabio Oliveira. ; TIBAES, J. H. ; MARCHAUKOSKI, Jeroniza. ; STEFFENS, Maria Berenice Reynaud. ; SOUZA, Emanuel Maltemmpi de ; SOUZA, V. ; RAITTZ, Roberto Tadeu. jContigSort: a new computer application for contigs ordering. In: 7th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (AB3C) and 3rd

- International Conference of the IberoAmerican Society for Bioinformatics (SolBio), 2011, Florianópolis. Abstract book, 2011.
12. PIRO, Vitor C et al. FGAP: an automated gap closing tool. **Bmc Research Notes**, (s.l.), v. 7, n. 1, p.371-375, 2014. Springer Science + Business Media. DOI: 10.1186/1756-0500-7-371.
 13. NCBI, National Center For Biotechnology. **Taxonomy of *Derxia lacustris***. Disponível em: <www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=764842&lvl=3&keep=1&srchmode=1&unlock&lin;=s>. Acesso em: 05 out. 2015.
 14. VIALLE, Ricardo Assunção. **SILA - UM SISTEMA PARA ANOTAÇÃO AUTOMÁTICA DE GENOMAS UTILIZANDO TÉCNICAS INDEPENDENTES DE ALINHAMENTO**. 2013. 96 f. Dissertação (Mestrado) - Curso de Pós - Graduação em Bioinformática, Universidade Federal do Paraná, Curitiba, 2013.
 15. AZIZ, Ramy K et al. The RAST Server: Rapid Annotations using Subsystems Technology. **Bmc Genomics**, Burr Ridge-usa, v. 9, n. 1, p.75-90, 2008. Springer Science + Business Media. DOI: 10.1186/1471-2164-9-75.
 16. OVERBEEK, Ross et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). **Nucleic Acids Research**, (s.l.), v. 42, n. 1, p.206-214, 29 nov. 2013. Oxford University Press (OUP). DOI: 10.1093/nar/gkt1226.
 17. RUTHERFORD, K. et al. Artemis: sequence visualization and annotation. **Bioinformatics**, (s.l.), v. 16, n. 10, p.944-945, 1 out. 2000. Oxford University Press (OUP). DOI: 10.1093/bioinformatics/16.10.944.
 18. Dereeper A.*, Guignon V.*, Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. *Phylogeny.fr: robust phylogenetic analysis for the non-specialist*. *Nucleic Acids Res.* 2008 Jul 1;36(Web Server issue):W465-9. Epub 2008 Apr 19. ([PubMed](#)) *: joint first authors
 19. Edgar RC. *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. *Nucleic Acids Res.* 2004, Mar 19;32(5):1792-7. ([PubMed](#))
 20. GORDON, D.; GREEN, P.. Consed: a graphical editor for next-generation sequencing. **Bioinformatics**, (s.l.), v. 29, n. 22, p.2936-2937, 31 ago. 2013. Oxford University Press (OUP). DOI: 10.1093/bioinformatics/btt515.

21. MILNE, I. et al. Using Tablet for visual exploration of second-generation sequencing data. **Briefings In Bioinformatics**, (s.l.), v. 14, n. 2, p.193-202, 24 mar. 2012. Oxford University Press (OUP). DOI: 10.1093/bib/bbs012.
22. CHUBATSU, Leda Satie et al. Nitrogen fixation control in *Herbaspirillum seropedicae*. **Plant Soil**, (s.l.), v. 356, n. 1-2, p.197-207, 27 maio 2011. Springer Science + Business Media. DOI: 10.1007/s11104-011-0819-6.
23. MORENO-VIVIÁN, Conrado et al. Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases. **Journal Of Bacteriology**, Córdoba, v. 181, n. 21, p.6573-6584, nov. 1999.
24. ROWE, John J. et al. Nark is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. **Molecular Microbiology**, (s.l.), v. 12, n. 4, p.579-586, maio 1994. Wiley-Blackwell. DOI: 10.1111/j.1365-2958.1994.tb01044.x.
25. HEIDER, Johann; FUCHS, Georg. Anaerobic metabolism of aromatic compounds. **Eur. J. Biochem.**, Freiburg, v. 243, p.577-596, ago. 1996.
26. PEDROSA, Fábio O. et al. Genome of *Herbaspirillum seropedicae* Strain SmR1, a Specialized Diazotrophic Endophyte of Tropical Grasses. **Plos Genetics**, (s.l.), v. 7, n. 5, p.1-10, 12 maio 2011. Public Library of Science (PLoS). DOI: 10.1371/journal.pgen.1002064.
27. OLIVER, D. Protein Secretion in *Escherichia Coli*. **Annual Review Of Microbiology**, (s.l.), v. 39, n. 1, p.615-648, out. 1985. Annual Reviews. DOI: 10.1146/annurev.mi.39.100185.003151.
28. REES, Douglas C.; JOHNSON, Eric; LEWINSON, Oded. ABC transporters: the power to change. **Nature Reviews Molecular Cell Biology**, (s.l.), v. 10, n. 3, p.218-227, mar. 2009. Nature Publishing Group. DOI: 10.1038/nrm2646.
29. ZHAO, Youfu; BLUMER, Sara E.; SUNDIN, George W.. Identification of *Erwinia amylovora* Genes Induced during Infection of Immature Pear Tissue. **Journal Of Bacteriology**, Michigan, v. 23, n. 187, p.8088-8103, set. 2005.
30. HAN, X. et al. Type IV Fimbrial Biogenesis Is Required for Protease Secretion and Natural Transformation in *Dichelobacter nodosus*. **Journal Of Bacteriology**, (s.l.), v. 189, n. 14, p.5022-5033, 18 maio 2007. American Society for Microbiology. DOI: 10.1128/jb.00138-07.
31. BINGLE, Lewis Eh; BAILEY, Christopher M; PALLEEN, Mark J. Type VI secretion: a beginner's guide. **Current Opinion In Microbiology**, (s.l.), v. 11, n. 1, p.3-8, fev. 2008. Elsevier BV. DOI: 10.1016/j.mib.2008.01.006.

32. OFEK, Itzhak et al. Genetic Exchange of Determinants for Capsular Polysaccharide Biosynthesis between *Klebsiella pneumoniae* Strains Expressing Serotypes K2 and K2la. **Infection And Immunity**, Tel-aviv, v. 10, n. 61, p.4208-4216, out. 1993.
33. ILAN, Ofir et al. Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. **The Embo Journal**, Jerusalem, v. 12, n. 18, p.3241-3248, 1999.
34. FIELD, Dawn et al. The minimum information about a genome sequence (MIGS) specification. **Nat Biotechnol**, (s.l.), v. 26, n. 5, p.541-547, 8 maio 2008. Nature Publishing Group. DOI: 10.1038/nbt1360.
35. ASHBURNER, Michael et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, (s.l.), v. 25, n. 1, p.25-29, maio 2000. Nature Publishing Group. DOI: 10.1038/75556.
36. GLAESER, Stefanie P.; KÄMPFER, Peter. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. **Systematic And Applied Microbiology**, (s.l.), v. 38, n. 4, p.237-245, jun. 2015.
37. MARMUR, J.. A procedure for the isolation of deoxyribonucleic acid from microorganisms. **Journal Of Molecular Biology**, (s.l.), v. 3, n. 2, p.208-218, abr. 1961. Elsevier BV. DOI: 10.1016/s0022-2836(61)80047-8.
38. GORIS, J; et al. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. **International Journal Of Systematic And Evolutionary Microbiology**, v. 57, n. 1, p.81-91 doi.org/10.1099/ijs.0.64483-0.
39. <http://enve-omics.ce.gatech.edu/ani/> (ANI)
40. <http://enve-omics.ce.gatech.edu/aai/> (AAI)
41. RODRIGUEZ-R, Luis M.; KONSTANTINIDIS, Konstantinos T.. Bypassing Cultivation To Identify Bacterial Species. **Microbe**, Qualquee, v. 9, n. 3, p.111-118, 2014.
42. Bankevich, Anton et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal Of Computational Biology**. 19(5): 455-477, 2012. doi: 10.1089/cmb.2012.0021. (Bankevich *et al* 2012)
43. ZIMIN, A. V. et al. The MaSuRCA genome assembler. **Bioinformatics**, (s.l.), v. 29, n. 21, p.2669-2677, 29 ago. 2013. Oxford University Press (OUP). DOI: 10.1093/bioinformatics/btt476. (Zimin *et at* 2013)

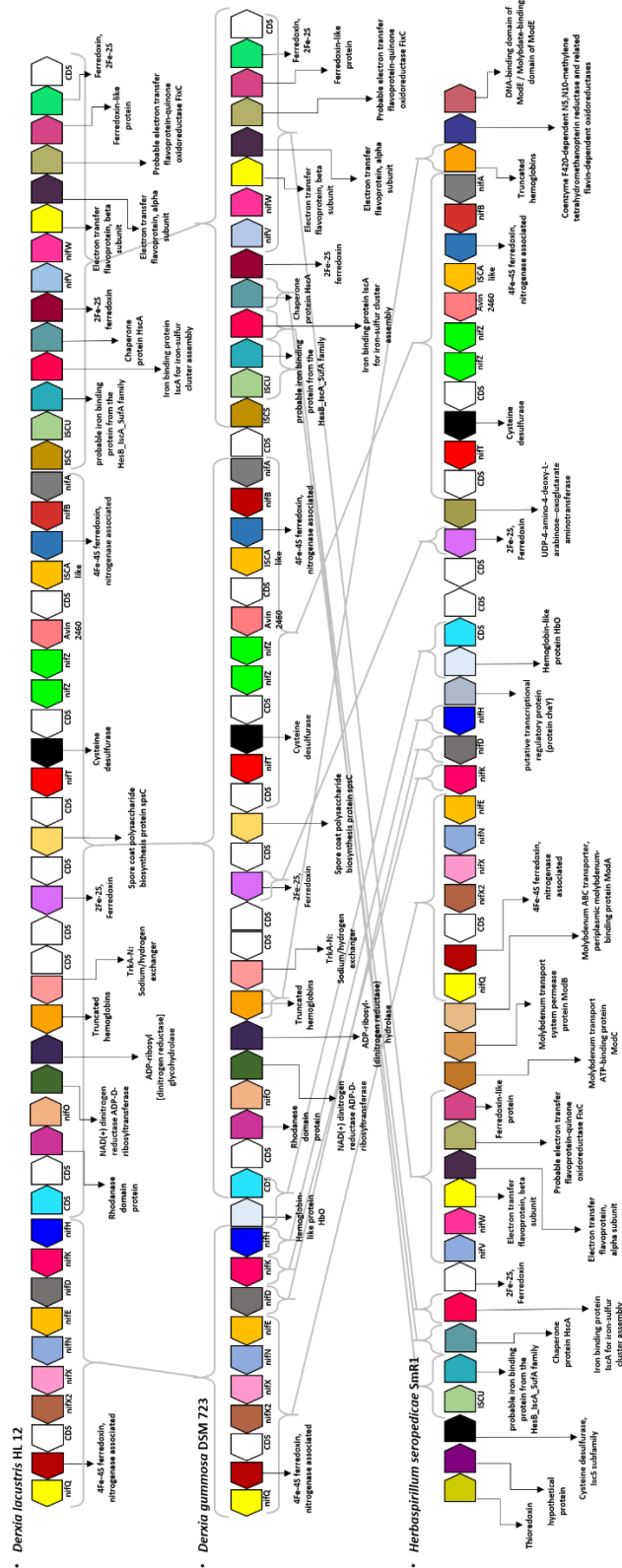
44. WALENZ, B. P. **Celera Assembler Users Guide, 2008**. Disponível em: <<http://wgs-assembler.sourceforge.net/runCA51.pdf>>. Acesso em: 09 fev. 2016. (Walenz 2008)
45. LANGMEAD, Ben; SALZBERG, Steven L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, (s.l.), v. 9, n. 4, p.357-359, 4 mar. 2012. Nature Publishing Group. DOI: 10.1038/nmeth.1923. (LANGMEAD; SALZBERG, 2012)
46. DENISOV, G. et al. Consensus generation and variant detection by Celera Assembler. **Bioinformatics**, (s.l.), v. 24, n. 8, p.1035-1040, 4 mar. 2008. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btn07>
47. ZERBINO, D. R.; BIRNEY, E.. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, (s.l.), v. 18, n. 5, p.821-829, 21 fev. 2008. Cold Spring Harbor Laboratory Press. <http://dx.doi.org/10.1101/gr.074492.107>.

ADDITIONAL FILE

TABLE S1. ANI AND AAI ANALYSES OF *D. LACUSTRIS* HL12 GENOME.

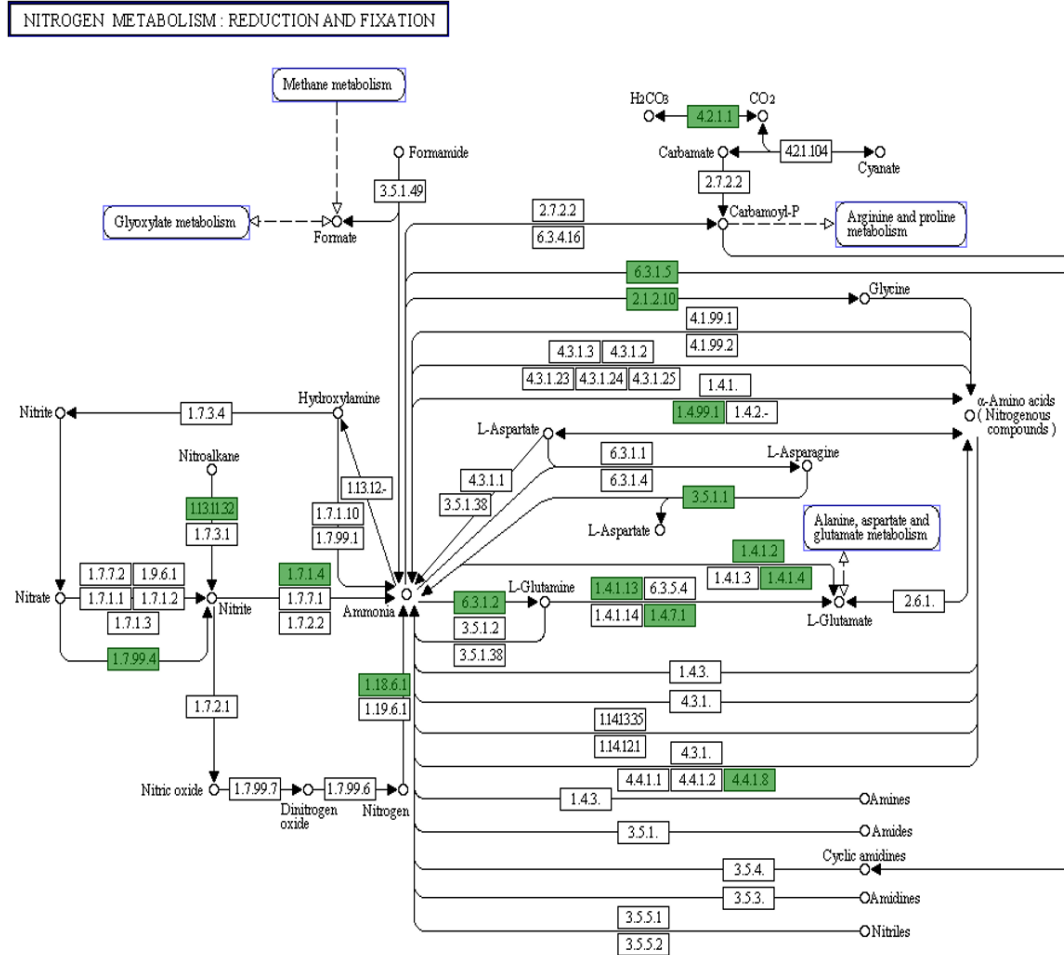
	ANI - Average Nucleotide Identity e AAI - Average Amino acid Identify	
	<i>D. lacustris</i> HL12	
	ANI	AAI
<i>Achromobacter xylosoxidans</i> A8	76,47	52,36
<i>Advenella kashmirensis</i> WT001	75,07	50,95
<i>Alcaligenes faecalis</i> ZD02	75,15	50,48
<i>Bordetella pertussis</i> Tohama I	76,52	53,03
<i>Burkholderia vietnamiensis</i> G4	76,46	53,78
<i>Castelaniella defragrans</i> 65Phen	75,79	51,65
<i>Collimonas fugivorans</i> Ter331	75,42	53,66
<i>D. gummosa</i> DSM 723	84,69	80,57
<i>Escherichia coli</i> IAI39	81,81	44,11
<i>Herbaspirillum seropedicae</i> SmR1	75,90	54,49
<i>Herminiimonas arsenicoxydans</i> ULPAs	74,95	54,19
<i>Janthinobacterium agaricidamnosum</i> DSM 9628	74,91	53,14
<i>Lautropia mirabilis</i> ATCC 51599	75,91	50,25
<i>Nitrosospira multififormis</i> ATCC 25196	79,78	50,33
<i>Pandoreae pnomenusa</i> DSM 16536	76,16	53,96
<i>Pusillimonas</i> sp. T7-7	75,24	52,04
<i>Taylorella equigenitalis</i> MCE9	88,49	60,08

FIG. S1 COMPARISON OF NIF GENE CLUSTER BETWEEN *D. lacustris* HL12, *D. gummosa* DSM 723 AND *Herbaspirillum seropedicae* SmR1.



IT SHOWS HIGH SIMILARITY IN THE GENES POSITIONS OF *Derxia lacustris* HL12 AND *Derxia gummosa* DSM 723. ALTHOUGH, *Herbaspirillum seropedicae* SMR1 NOT PRESENTS THE SAME GENES DISTRIBUTION.

FIG. S2 NITROGEN METABOLISM PATHWAYS IDENTIFIED IN *D. lacustris* HL12GENOME



NITROGEN METABOLISM PATHWAYS IDENTIFIED IN *D. lacustris* HL12 GENOME. GREEN BLOCKS SHOW METABOLIC PATHWAY STEP REPRESENTED ON *Derxia lacustris* HL12 GENOME.

CONCLUSÕES

D. lacustris HL12 é uma bactéria pertencente ao gênero *Derxia*, .O genoma desta bactéria foi sequenciado utilizando a plataforma Illumina MySeq. A sequência genômica final está distribuída em 28 contigs..

A sequência genômica foi anotada e algumas de suas vias metabólicas foram analisadas. Essa bactéria é capaz de realizar a fixação biológica de nitrogênio assim como *D. gummosa* DSM 723. Possui genes pertencentes ao metabolismo do nitrato, uma fonte de nitrogênio para o organismo e para a assimilação em aminoácidos. *D. lacustris* HL12 possui os sistemas de secreção do tipo I, II, IV e VI. Possui também genes relacionados com a degradação de compostos aromáticos.

REFERENCIAS

- ABBY, S.; DAUBIN, V. Comparative genomics and the evolution of prokaryotes. **Trends in microbiology**, v. 15, n. 3, p. 135–41, 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17289390>>. Acesso em: 12/7/2014.
- BECKING, J. H.; TRUPER, H. G.; DWORKIN, M.; HARDER, W.; SCHLEGEL, H. G. **The Genus *Dexia* In The Prokaryotes**. vol. 3 ed. Berlin: Balows, 1991.
- BERG, J. M.; TYMOCZKO, J. L.; STRYER, L. **Bioquímica**. 6º ed. Rio de Janeiro, 2008.
- BRADNAM, K. R.; FASS, J. N.; ALEXANDROV, A.; et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. **GigaScience**, v. 2, n. 1, p. 10, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3844414&tool=pmcentre&rendertype=abstract>>. Acesso em: 11/7/2014.
- BREDEMEIER, C.; MUNDSTOCK, C. M. REGULAÇÃO DA ABSORÇÃO E ASSIMILAÇÃO. **Ciência Rural**, v. 30, n. 2, p. 365–372, 2000.
- BOUWER, J.; ZEHNDER, A. J. B. Bioremediation of organic compounds - putting microbial metabolism to work. **Bioremediation**, v. 11, p. 360-367, 1993.
- CAMPBELL, M. K. **Bioquímica**. 3ª Edição ed. São Paulo: Artmed, 2003.
- CAMPELO, A. B.; DOBEREINER, J. Ocorrência de *Dexia* sp. em solos de alguns estados brasileiros. **Pesq. agorpec. bras.**, v. 5, n. 60, p. 327–332, 1970.
- CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas Next generation DNA sequencing and its applications in plant genomics. **Ciência Rural**, v. 40, n. 3, p. 735–744, 2010.
- CATANHO, M.; DEGRAVE, W.; MIRANDA, A. B. DE. Análise Comparativa de Genomas Procarióticos. **Biotecnologia Ciência & Desenvolvimento**, v. 2007, n. 37, p. 20–29, 2008.
- CHEN, W.-M.; HUANG, W.-C.; SHEU, S.-Y. *Dexia lacustris* sp. nov., a nitrogen-fixing bacterium isolated from a freshwater lake. **International journal of systematic and evolutionary microbiology**, v. 63, n. Pt 3, p. 965–70, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22659506>>. Acesso em: 25/7/2014.
- HARWOOD, Caroline S. et al. Anaerobic metabolism of aromatic compounds via the benzoyl-CoA pathway. **Fems Microbiology Reviews**, v. 22, n. 5, p.439-458, 1998.
- HEIDER, J.; FUCHS, G. **Review anaerobic metabolism of aromatic compounds**, Eur. J. Biochem, v. 243, 577-596, 1997.

- HILL, S.; POSTGATE, J. R. Failure of putative nitrogen-fixing bacteria to fix nitrogen. **Journal of general microbiology**, v. 58, n. 2, p. 277–85, 1969. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/5360480>>. .
- JANSEN, R.; EMBDEN, J. D. A VAN; GAASTRA, W.; SCHOULS, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. **Molecular microbiology**, v. 43, n. 6, p. 1565–75, 2002. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11952905>>. .
- JENSEN, H. L.; PETERSEN, E. J.; BHATTACHARYA, P. K. D. R. A New Nitrogen-fixing Bacterium : *Derxia gummosa* nov. gen. nov.. spec. **Archiv für Mikrobiologie**, v. 195, n. 36, p. 182–195, 1960.
- LIU, L.; SILIANG, L.; HU, N.; et al. Comparison of next-generation sequencing systems. **Journal of biomedicine & biotechnology**, v. 2012, p. 251364, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3398667&tool=pmcentrez&rendertype=abstract>>. Acesso em: 11/7/2014.
- MARIN, V. A.; BALDANI, V. L. D.; TEIXEIRA, K. R. DOS S.; BALDANI, J. I. Fixação Biológica de Nitrogênio: Bactérias Fixadoras de Importância para a Agricultura Tropical. , p. 3–34, 1999.
- MOREIRA, F. M. S.; SILVA, K.; NÓBREGA, R. S. A.; CARVALHO, F. Bactérias diazotróficas associativas : diversidade , ecologia e potencial de aplicações. **Comunicata Scientiae**, v. 1, n. 2, p. 74–99, 2010.
- NAGAJARAN, N.; M, P. Sequence Assembly Demystified. **Nature Reviews Genetics**, v. 14, 2013.
- NELSON, D. L.; COX, M. M. **Lehninger Princípios de Bioquímica**. 3º ed. Nova Iorque: Worth Publishers, 2000.
- OLIVER, D. Protein Secretion in Escherichia Coli. **Annual Review Of Microbiology**, v. 39, n. 1, p.615-648, 1985.
- PIRO, V. C.; FAORO, H.; WEISS, V. A.; et al. FGAP : an automated gap closing tool. , p. 1–5, 2014.
- ROY, A. B.; SEN, S. A New Species of Derxia. **Nature**, v. 194, n. 4828, p. 604–605, 1962.
- SCHATZ, M. C.; DELCHER, A. L.; SALZBERG, S. L. Assembly of large genomes using second-generation sequencing. **Genome research**, v. 20, n. 9, p. 1165–73, 2010. Disponível em:

<<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2928494&tool=pmcentrez&rendertype=abstract>>. Acesso em: 14/7/2014.

SHANKAS, H. N. R.; KENNEDY, I. R.; NEW, P. B. Autotrophic Growth and Nitrogen Fixation in *Derxia gummosa*. **Journal of General Microbiology**, v. 132, n. 1986, p. 1797–1803, 2006.

SHENDURE, J.; MITRA, R. D.; VARMA, C.; CHURCH, G. M. Advanced sequencing technologies: methods and goals. **Nature reviews. Genetics**, v. 5, n. 5, p. 335–44, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15143316>>. Acesso em: 15/7/2014.

SPRENT, J. I. .; SPRENT, P. **Nitrogen fixing organisms**. 2º ed. London, 1990.

XIE, C. H.; YOKOTA, A. Phylogenetic analyses of the nitrogen-fixing genus *Derxia*. **The Journal of general and applied microbiology**, v. 50, n. 3, p. 129–35, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15486821>>. .

ZAHA, A.; FERREIRA, H. B. **Biologia Molecular Básica**. 4º ed. Artmed, 2014.

ZHOU, Zhemin et al. Ammonia Fermentation, a Novel Anoxic Metabolism of Nitrate by Fungi. **The Journal Of Biological Chemistry**, v. 277, n. 3, p.1892-1896, 2001.