

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**CENTRO DE BIOTECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR**

**GENÔMICA COMPARATIVA ENTRE ESPÉCIES DO GÊNERO *AZOSPIRILLUM*,  
COM ÊNFASE NAS CARACTERÍSTICAS EXCLUSIVAS DA BACTÉRIA  
PROMOTORA DE CRESCIMENTO VEGETAL *AZOSPIRILLUM AMAZONENSE***

**TESE DE DOUTORADO**

**RICARDO CECAGNO**

**PORTO ALEGRE, 2013.**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
CENTRO DE BIOTECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

GENÔMICA COMPARATIVA ENTRE ESPÉCIES DO GÊNERO *AZOSPIRILLUM*,  
COM ÊNFASE NAS CARACTERÍSTICAS EXCLUSIVAS DA BACTÉRIA PROMOTORA  
DE CRESCIMENTO VEGETAL *AZOSPIRILLUM AMAZONENSE*

Ricardo Cecagno

Tese submetida ao programa de Pós-Graduação em Biologia Celular e Molecular da Universidade Federal do Rio Grande do Sul, como um dos requisitos para a obtenção do grau de Doutor em Ciências.

Orientadora: Profa. Dra. Irene Silveira Schrank

Porto Alegre, julho de 2013

Este trabalho foi desenvolvido no Laboratório de Microrganismos Diazotróficos, no Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul.

## **Reflexão**

(Júlio Machado da Silva Filho, Colmar P. Duarte)

Para fugir a tristeza  
Por buscar esquecimento,  
Desejei ser como o vento  
Que vai passando sozinho,  
Sem repisar um caminho  
Sem conhecer paradeiro  
Quis ser nuvem ao pampeiro  
Ser a estrela que fulgiu,  
Quis ser as águas do rio  
Fazendo inveja às areias  
Em seu eterno viajar!

Um dia cansei de andar  
E desejei novamente  
Em vez de rio ser barranca,  
Em vez de vento, moirão,  
Em vez de nuvem, semente,  
Em vez de estrela, ser chão!  
Recém então aprendi  
Que muita gente maldiz  
Sua sorte - insatisfeita  
Por não saber que é feliz

E nunca mais invejei  
O destino das estrelas,  
Que só enfeitam a noite  
Porque o sol não pode vê-las;  
As nuvens que submissas,  
Vão onde o vento as levar  
E o vento que passa triste  
Porque não pode voltar!

## **AGRADECIMENTOS**

Aos meus pais, Eduardo Antônio Cecagno e Teonila Trevisan Cecagno, pelo apoio incondicional.

Ao meu irmão Ronaldo, pelo incentivo.

À professora Irene, pelos muitos anos de ensinamentos.

À professora Luciane Passaglia, que primeiro me orientou e pelos anos de colaboração.

Aos professores Augusto Schrank e Marilene Vainstein, pela colaboração.

À professora Ana Tereza de Vasconcelos e o Laboratório Nacional de Computação Científica pela colaboração.

Aos amigos e colegas do Laboratório de Microrganismos Diazotróficos, especialmente Tiago Fritsch, Fernando Sant'Anna, Shana Weber, Luciano Reolon, Débora Trentini, Maicon Machado, Dieime Andrade, Sheila Mucha e Franciele Siqueira.

Aos amigos do Laboratório de Fungos Filamentosos, Juliano Boldo e Leonardo Broetto.

Ao Centro de Biotecnologia, onde trabalhei desde o início da graduação.

Aos amigos da Secretaria do Cbiot, Sílvia Centeno, Luciano Saucedo e Grasiela Barbosa.

À CAPES

À UFRGS

## ÍNDICE

LISTA DE ABREVIATURAS E SIGLAS .....	ix
LISTA DE FIGURAS .....	x
LISTA DE TABELAS .....	xi
LISTA DE FIGURAS E TABELAS SUPLEMENTARES .....	xiii
RESUMO.....	xv
ABSTRACT .....	xvi
1. Introdução.....	17
1.1. O gênero <i>Azospirillum</i> .....	17
1.1.1. <i>Azospirillum amazonense</i> .....	18
1.1.2. <i>Azospirillum brasilense</i> .....	20
1.1.3. <i>Azospirillum lipoferum</i> .....	21
1.1.4. <i>Azospirillum</i> sp. B510 .....	22
1.2. Avanços recentes no estudo de <i>Azospirillum</i> e novas possibilidades de análise.....	22
1.3. Análise diferencial de sequências genômicas.....	24
1.4. Objetivos .....	27
2. Materiais e Métodos.....	28
2.1. Análise <i>in Silico</i> .....	28
2.1.1. Obtenção dos genomas.....	28
2.1.2. Metodologia de análise <i>in silico</i> .....	28
2.1.3. Cálculos de análise de similaridade.....	31
2.1.3.1. Escore .....	31
2.1.3.2. Valor E.....	32
2.1.3.3. Valor H .....	32
2.2. Análise experimental.....	33
2.2.1. Condições de cultivo e purificação do DNA .....	33

2.2.2.	Análise de diferença representacional (RDA) .....	33
2.2.2.1.	Preparação dos DNAs de <i>A. amazonense</i> (TESTER) e <i>A. brasilense</i> (DRIVER).....	33
2.2.2.2.	Hibridização Subtrativa.....	34
2.2.2.3.	Amplificação diferencial.....	35
2.2.2.4.	Clonagem, Transformação e Sequenciamento .....	35
3.	Resultados.....	37
3.1.	Anotação automática.....	37
3.1.1.	Validação da análise comparativa das proteínas .....	39
3.1.2.	Análise comparativa das vias metabólicas entre as proteínas codificadas pelos genomas das quatro espécies de <i>Azospirillum</i> .....	40
3.2.	Comparação entre genomas.....	41
3.2.1.	Análise comparativa <i>in silico</i> entre os genomas de <i>Azospirillum</i> .....	41
3.2.2.	Identificação de proteínas diferenciais <i>in silico</i> .....	43
3.2.3.	Identificação experimental de proteínas diferenciais .....	47
3.3.	Caracterização das principais rotas metabólicas presentes no genoma de <i>Azospirillum amazonense</i> . .....	51
3.3.1.	Proteínas relacionadas ao metabolismo de hormônios vegetais .....	51
3.3.2.	Proteínas relacionadas ao transporte em <i>Azospirillum</i> .....	58
3.3.2.2.	Transportadores identificados no banco de ortólogos do KEGG...	61
3.3.2.3.	Transportadores identificados no banco TCDB .....	62
3.4.	Herança viral e elementos genéticos móveis .....	63
3.4.1.	Regiões genômicas relacionadas a fagos em quatro espécies de <i>Azospirillum</i> .....	64
3.4.2.	Classificação de Elementos Genéticos Móveis .....	65
3.4.3.	Sistema imune bacteriano CRISPR e proteínas associadas .....	66
3.5.	Caracterização das proteínas específicas de <i>A. amazonense</i> com relação a <i>A. brasilense</i> . .....	68

3.5.1.	Biossíntese de metabólitos secundários, transporte e catabolismo..	69
3.5.2.	Transcrição.....	72
3.5.3.	Transporte e metabolismo de carboidratos .....	73
3.5.4.	Modificação pós-traducional, turnover de proteínas e chaperonas...	74
4.	Discussão .....	75
6.	Material Suplementar.....	88
7.	Referências Bibliográficas .....	131
8.	Anexos.....	140
8.1.	Anexo I.....	140
	I- “Genomic insights into the versatility of the plant growth-promoting bacterium <i>Azospirillum amazonense</i> ” .....	140
	II- “The Plant Growth-Promoting Bacteria <i>Azospirillum amazonense</i> : Genomic Versatility and Phytohormone Pathway” .....	155
9.	Curriculum vitae.....	163



## LISTA DE ABREVIATURAS E SIGLAS

<b>AB510</b>	<i>Azospirillum</i> sp. B510
<b>ABA</b>	<i>Abscisic acid</i>
<b>Abra</b>	<i>Azospirillum brasilense</i>
<b>AIA</b>	<i>Indol acetic acid</i>
<b>Alipo</b>	<i>Azospirillum lipoferum</i>
<b>Amz</b>	<i>Azospirillum amazonense</i>
<b>ATCC</b>	<i>American Type Culture Collection</i>
<b>CAS</b>	<i>CRISPRs associated protein</i>
<b>CDD</b>	<i>Conserved Domain Database</i>
<b>COG</b>	<i>Clusters of Orthologous Groups</i>
<b>CRISPR</b>	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
<b>DNA</b>	<i>Deoxyribonucleic acid</i>
<b>dNTP</b>	<i>Deoxyribonucleotide</i>
<b>DR</b>	<i>Direct repeat</i>
<b>EDTA</b>	<i>Ethylenediamine tetraacetic acid</i>
<b>FAD</b>	<i>Flavin adenine dinucleotide</i>
<b>IPTG</b>	<i>Isopropyl <math>\beta</math>-D-1-thiogalactopyranoside</i>
<b>LB</b>	<i>Luria Bertani</i>
<b>MOPS</b>	<i>3-morpholinopropane-1-sulfonic acid</i>
<b>NaCl</b>	<i>Sodium chloride</i>
<b>NAD</b>	<i>Nicotinamide adenine dinucleotide</i>
<b>CED</b>	<i>Nine cis epoxycarotenoide dioxigenase</i>
<b>ORF</b>	<i>Open Reading Frame</i>
<b>OSPRT</b>	<i>Oscillating Phenol Emulsion Reassociation Technique</i>
<b>PCR</b>	<i>Polimerase Chain Reaction</i>
<b>pH</b>	<i>Potential of hydrogen</i>
<b>RDA</b>	<i>Representational Difference Analysis</i>
<b>RPM</b>	<i>Revolutions per minute</i>
<b>SOC</b>	<i>Super Optimal broth with Catabolic repressor</i>
<b>T4SS</b>	<i>Type 4 Secretion System</i>

## LISTA DE FIGURAS

**Figura 1.1** – Fixação biológica de nitrogênio e produção de hormônio de crescimento vegetal em isolados de *A. amazonense*.

**Figura 3.1** - Hibridização subtrativa *in silico* entre quatro espécies do gênero *Azospirillum*.

**Figura 3.2** – Hibridização subtrativa seguida de amplificação diferencial.

**Figura 3.3** – Caracterização das sequências diferenciais de *A. amazonense*.

**Figura 3.4** - Filogenia de um transportador de auxina presente em *A. amazonense*.

**Figura 3.5** – Identificação das proteínas de *A. amazonense* envolvidas nas vias de sinalização de hormônios de plantas. A figura foi modificada segundo informações contidas nos mapas de ortólogos do KEGG. *A. thaliana* e *O. sativa* foram utilizadas como referências para localização das proteínas.

**Figura 3.6** – Proteínas relacionadas ao Sistema de Secreção tipo IV em quatro membros do gênero *Azospirillum*.

**Figura 3.7** – Análise filogenética da carotenoide/lignistilbeno oxigenase de *A. amazonense*.

**Figura 3.8** - Região genômica da carotenoide/lignistilbeno oxigenase de *A. amazonense*.

## LISTA DE TABELAS

**Tabela 2.1** - Identificação e características dos três pares de oligonucleotídeos adaptadores utilizados no RDA.

**Tabela 3.1** - Número de sequências de proteínas de *A. amazonense* com maior similaridade por espécie comparada.

**Tabela 3.2** – Rotas metabólicas em que *A. amazonense* possui maior número de proteínas com relação a pelo menos uma das espécies comparadas.

**Tabela 3.3** - Principais rotas metabólicas que apresentam maior número de proteínas codificadas no genoma de *A. amazonense*.

**Tabela 3.4** - Número de sequências de proteínas de *A. amazonense* com melhor similaridade por espécie de *Azospirillum* comparada.

**Tabela 3.5** - Número de proteínas específicas de *A. amazonense* com relação às outras três espécies do gênero, classificadas no banco de sequências do COG. .

**Tabela 3.6** – Número de proteínas conservadas entre quatro espécies de *Azospirillum* utilizando o Valor H de 0,8 como padrão. A classificação por vias metabólicas foi realizada com base no banco de dados do COG.

**Tabela 3.7** - Proteínas de *A. amazonense* identificadas por RDA e classificadas no banco de dados COG.

**Tabela 3.8** – Identificação das proteínas de *A. amazonense* envolvidas nas vias de sinalização de hormônios de plantas identificados no banco de ortólogos do KEGG apresentados na Figura 3.5.

**Tabela 3.9** - Proteínas de *A. amazonense* similares a proteínas de *A. brasilense* reguladas por auxina e envolvidas nas rotas de fitormônios em *A. thaliana*.

**Tabela 3.10** - Proteínas do Sistema de Secreção Tipo IV exclusivas de *A. amazonense* em relação ao gênero *Azospirillum*.

**Tabela 3.11** – Grupo de Ortólogos do KEGG relacionados ao sistema de secreção tipo IV localizados em *A. amazonense*.

**Tabela 3.12** – Proteínas com similaridade no TCDB por espécie de *Azospirillum* e a contribuição relativa no proteoma de cada organismo.

**Tabela 3.13** – Transportadores de *A. amazonense* identificados no TCDB, classificados no banco de dados COG.

**Tabela 3.14** - Herança plasmidial e viral encontrada em quatro espécies de *Azospirillum*.

**Tabela 3.15** – Sistema imune bacteriano CRISPR.

**Tabela 3.16** – Proteínas Cas de *A. amazonense* que possuem correspondente em outras espécies do gênero.

## LISTA DE FIGURAS E TABELAS SUPLEMENTARES

**Figura Suplementar 1** – Proteínas relacionadas com a sinalização de hormônios vegetais em quatro espécies do gênero *Azospirillum* identificadas pelo KAAS.

**Figura Suplementar 2** – Identificação das proteínas do Sistema de Secreção Tipo IV em *A. amazonense* não compartilhadas com outros membros do gênero.

**Figura Suplementar 3** - Regiões genômicas de três espécies de *Azospirillum* com similaridade em genomas virais detectadas pela ferramenta PHAST (*Phage Search Tool*).

**Figura Suplementar 4** – Regiões do genoma de *A. amazonense* com similaridade em genomas virais detectadas pela ferramenta PHAST (*Phage Search Tool*).

**Figura Suplementar 5** - Alinhamento da glutamina ciclotransferase de *A. amazonense* identificada por RDA.

**Tabela Suplementar 1** – Proteínas conservadas entre as quatro espécies de *Azospirillum* analisadas.

**Tabela Suplementar 2** – Proteínas com função conhecida exclusivas de *A. amazonense* em relação a *Azospirillum*.

**Tabela Suplementar 3** – Proteínas exclusivas de *A. amazonense* com relação a *A. brasilense* identificadas pela técnica de RDA.

**Tabela Suplementar 4** – Proteínas de *A. amazonense* relacionadas ao metabolismo de hormônios vegetais pela classificação do banco de sequências KEGG.

**Tabela Suplementar 5** – Proteínas de *A. amazonense* relacionadas ao metabolismo de hormônios vegetais pela classificação do banco de sequências AHD.

**Tabela Suplementar 6** - Classificação metabólica das proteínas de *A. amazonense* conservadas com o banco de hormônios de *Arabidopsis*.

**Tabela Suplementar 7** - Proteínas de *A. amazonense* com similaridade nas proteínas diferencialmente expressas na presença de auxina em *A. brasilense* classificadas por COG.

**Tabela Suplementar 8** – Classificação das proteínas de *Arabidopsis thaliana* (AHD) com similares em *A. amazonense* que foram diferencialmente expressas na presença de auxina em *A. brasilense*. Identificação dos hormônios

correspondentes e descrição da atuação no seu metabolismo assim como o teste realizado para caracterização.

**Tabela Suplementar 9** – Transportadores de *A. amazonense* identificados no TCDB e relacionados com a Classe U do banco de dados COG.

**Tabela Suplementar 10** – Identificação das sequências de proteínas localizadas nas quatro regiões virais identificadas em *A. amazonense* pelo PHAST e nível de similaridade com proteína viral correspondente.

**Tabela Suplementar 11** - Regiões genômicas virais em três espécies do gênero *Azospirillum* segundo PHAST. AB510 - *Azospirillum* sp. B510, Abra – *Azospirillum brasilense*, Alipo – *Azospirillum lipoferum*.

**Tabela Suplementar 12** - Identificação das proteínas de *A. amazonense* localizadas nas quatro regiões virais detectadas pelo PHAST.

**Tabela Suplementar 13** – Proteínas associadas ao sistema imune CRISPR identificadas em *A. amazonense* pela ferramenta CRISPRs-finder.

**Tabela Suplementar 14** – A figura apresenta os principais domínios da proteína relacionada a fagos de *A. amazonense* identificada por RDA.

## RESUMO

O gênero *Azospirillum* é conhecido pela capacidade de promover o crescimento vegetal, principalmente através da fixação biológica do nitrogênio e metabolismo de fitormônios. Apesar de muitos anos de estudos com esses microrganismos, apenas em 2010 o primeiro genoma completo foi disponibilizado, tendo sido selecionada *Azospirillum* sp. B510 devido a capacidade de atuar endofiticamente. A espécie *A. amazonense*, isolada de gramíneas de importância econômica como arroz e cana-de-açúcar, possui características que possibilitam sua utilização em solos mais ácidos, além de atuar na metabolização de xenobióticos, onde tem se destacado seu potencial em biorremediação. Com a colaboração do Projeto Genoma Sul nosso grupo realizou um sequenciamento parcial do genoma de *A. amazonense*. Genes vinculados a fixação de nitrogênio e carbono, *quorum sensing*, quimiotaxia e resistência bacteriana foram inicialmente caracterizado em um total de 3.319 ORFs anotadas. O presente trabalho realizou uma predição e anotação para a sequência completa dos *contigs* já depositados para este genoma, onde 5.496 proteínas foram caracterizadas em diferentes bancos de dados de sequências curadas, visando a identificação de características de importância biotecnológica. Devido a recente disponibilização de três outros genomas de *Azospirillum*, uma análise comparativa foi realizada para identificar e caracterizar sequências conservadas e diferenciais entre as espécies. Por meio da genômica comparativa entre essas bactérias e os vegetais associados foi identificado proteínas possivelmente relacionadas com o metabolismo de fitormônios. Análises detalhadas são apresentadas para proteínas do sistema de secreção do tipo IV e transportadores em geral, elementos genéticos móveis e sistema imune bacteriano. Estudos *in vitro* utilizando a técnica de RDA foram realizados para a comparação entre os genomas de *A. amazonense* e *A. brasilense*, sendo 46 proteínas exclusivas da primeira espécie identificadas e caracterizadas. Os resultados demonstraram que *A. amazonense* é a espécie mais afastada filogeneticamente entre essas bactérias, possuindo maior similaridade com *A. brasilense* entre os genomas já sequenciados neste gênero. Foram identificadas proteínas exclusivas de *A. amazonense* relacionadas ao sistema de secreção do tipo IV, envolvidas na troca de material genético. Essa espécie também apresentou duas regiões com proteínas exclusivas associadas ao sistema imune CRISPR, vinculado ao controle de entrada de DNA/RNA exógeno. Os resultados obtidos auxiliam no entendimento da genética desses microrganismos, servindo como base para experimentos futuros dirigidos à sua utilização como inoculante e na compreensão do seu papel no ambiente em que vive.

## ABSTRACT

The genus *Azospirillum* is known for its ability to promote plant growth, mainly through biological nitrogen fixation and metabolism of plant hormones. Despite many years of studies with these microorganisms, the first complete genome became available only in 2010, and *Azospirillum* sp. B510 has been selected due to its ability to act as an endophyte. The *A. amazonense* specie, isolated from grasses of economic importance such as rice and sugar cane, possesses characteristics that allow its use in acidic soils, besides acting in xenobiotics metabolism, which has highlighted its potential for bioremediation. With the collaboration of *Projeto Integrado Genoma Sul*, our group conducted a partial sequencing of *A. amazonense* genome. Genes linked to nitrogen and carbon fixation, *quorum sensing*, chemotaxis and bacterial resistance were initially characterized in a total of 3,319 annotated ORFs. In the present work, we performed the prediction and annotation for the complete sequence of the contigs already deposited for *A. amazonense* genome, from which 5,496 proteins were characterized in different curated sequence databases, with the aim to identify characteristics of biotechnological importance. Due to the recent availability of three other genomes of *Azospirillum*, we performed a comparative analysis to identify conserved sequences and differences between these species. A comparative genomics study between these bacteria and associated plants has been conducted, and proteins possibly associated to plant hormones metabolism were identified. Detailed analyzes are presented for protein secretion system type IV and for transporters in general, for mobile genetic elements and for bacterial immune system. A comparison between *A. amazonense* and *A. brasilense* was performed by *in vitro* studies applying the RDA technique, and 46 proteins unique to the former specie were identified and characterized. The results obtained showed that *A. amazonense* is the specie more distant phylogenetically among these bacteria, and presenting highest similarity to *A. brasilense* among the genomes already sequenced in this genus. We identified several unique *A. amazonense* proteins related to type IV secretion system, which involves exchange of genetic material. This specie also presented two regions with unique proteins associated with the immune system CRISPR, associated to the input control of exogenous DNA/RNA. The results obtained in this study help in understanding the genetics of these microorganisms, serving as a basis for future experiments concerning its use as an inoculant, or in the comprehension of their role in the environment in which they live.



## 1. Introdução

Um dos grandes desafios deste século é o desenvolvimento de uma agricultura sustentável com aumento na qualidade dos produtos e diminuição dos danos ao meio ambiente. Um aumento da produção de alimentos é necessário para satisfazer as necessidades de uma população também crescente. Métodos atuais utilizados na produção agrícola têm apresentado uma série de efeitos na natureza, pelo uso indevido de pesticidas e fertilizantes químicos, que resultam numa lista de problemas para a saúde (GUNNELL *et al.*, 2007). Uma das principais alternativas para os fertilizantes que têm sido estudadas nas últimas décadas é a fixação biológica do nitrogênio. Esse processo consiste na redução do nitrogênio molecular para amônia, processo realizado unicamente por microrganismos diazotróficos. Ele ocorre apenas em condições bastante específicas, por um complexo enzimático chamado nitrogenase (EADY & POSTGATE, 1974). A fixação do nitrogênio na indústria, por exemplo, necessita alta pressão e temperatura, utilizando catalizadores à base de ferro, e resultando em altos custos finais de produção. A principal finalidade da amônia produzida está relacionada à fabricação de fertilizantes que causam prejuízos econômicos e principalmente ambientais (PEDROSA *et al.*, 2000). Cabe salientar que os organismos diazotróficos contribuem para 60% do nitrogênio fixado anualmente, sendo esse elemento necessário para todos os seres vivos, já que compõe as principais biomoléculas, como aminoácidos e ácidos nucleicos (KIM & REES, 1994).

### 1.1. O gênero *Azospirillum*

As espécies do gênero *Azospirillum* são  $\alpha$ -proteobactérias de vida livre capazes de promover o crescimento vegetal através da produção de fitormônios, fixação biológica do nitrogênio e inibição de patógenos. Após sua “redescoberta” durante os anos 70 e a confirmação de sua eficiência na associação com vegetais de importância econômica, vários inoculantes utilizando esses microrganismos foram desenvolvidos (BASHAN *et al.*, 2004).

Atualmente 16 espécies de *Azospirillum* estão caracterizadas, além de centenas de isolados, incluindo *Azospirillum sp.* B510 cujo genoma foi o primeiro a ser finalizado (KANEKO *et al.*, 2010). As espécies descritas mais recentemente foram *Azospirillum formosense* no ano de 2011 em Taiwan, cuja análise de 16S *rRNA* demonstrou maior proximidade com *Azospirillum brasilense* DSM, *Azospirillum rugosum* e *Azospirillum oryzae* (LIN *et al.*, 2012) e *Azospirillum fermentarium* em 2013. Esta última apresentou a característica de tolerar altas concentrações de NaCl, sendo capaz de se multiplicar na presença de mais de 3% desse sal no tanque fermentativo de onde foi isolado (LIN *et al.*, 2013).

#### 1.1.1. *Azospirillum amazonense*

A espécie *A. amazonense* foi isolada em 1985 e já foi encontrada em associação com diversos vegetais de importância econômica, principalmente em arroz e cana-de-açúcar. Análises ambientais demonstraram que *A. amazonense* está presente não somente na rizosfera de gramíneas, mas também de espécies como tomate, café, couve e orquídeas, contribuindo para o seu desenvolvimento (LANGE & MOREIRA, 2002). Esta espécie também foi encontrada em raízes de gramíneas do gênero *Brachiaria*, que compõe pastagens das regiões do Cerrado e Mata Atlântica, sendo que essas bactérias foram capazes de produzir hormônios de crescimento vegetal como ácido indol acético (AIA) em testes de laboratório. As populações bacterianas apresentaram variações conforme a espécie vegetal utilizada no caso da Mata Atlântica e época de coleta no ecossistema do Cerrado (REIS JUNIOR *et al.*, 2004). Entre as características que a distingue de *A. brasilense* devemos considerar:

##### a) Tolerância a solos ácidos:

É de fundamental importância o isolamento de organismos promotores do crescimento vegetal que estejam adaptados às condições locais do solo a ser cultivado, sendo capazes de competir com os microrganismos já presentes no mesmo. Foi demonstrado que *A. amazonense* é mais tolerante à acidez do solo, característica comum no estado do Rio Grande do Sul e outras regiões do Brasil, podendo inclusive fixar nitrogênio nessa condição (HARTMANN *et al.*, 1985).

#### b) Metabolismo de carboidratos:

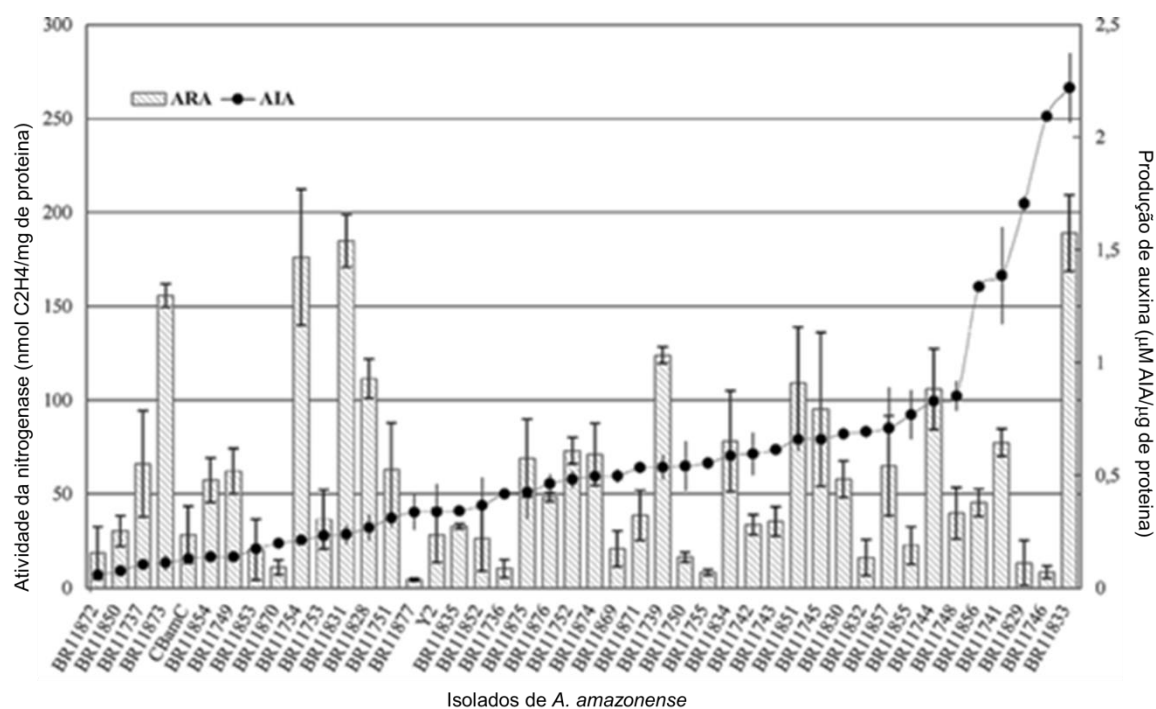
Uma análise de seis isolados de *A. amazonense* demonstrou que este membro do gênero era capaz de formar açúcares reduzidos a partir da sacarose (MARTÍNEZ-DRETS *et al.*, 1985). Com a caracterização de novas espécies, pôde ser observado que outras três espécies compartilham essa habilidade, sendo elas *Azospirillum irakense*, *Azospirillum mellins* e *Azospirillum palatum*. As mesmas espécies são as únicas capazes de utilizar a D-celobiose como fonte de carbono. A maior diferenciação ocorre no caso do mioinositol, que só é metabolizado por *A. amazonense* e *A. mellinis*, onde esta última demonstrou ser capaz de utilizar todas as fontes de carbono testadas nas análises (REIS *et al.*, 2011).

#### c) Regulação da nitrogenase:

Ensaio com a enzima nitrogenase de *A. amazonense* demonstraram que esta proteína não é inativada por cloreto de amônio, sugerindo que esta espécie poderia fixar nitrogênio mesmo na presença do elemento fixado. Com relação ao pH, 50% da atividade da enzima foram mantidos quando este foi diminuído para pH=5, o que provavelmente está relacionado ao ambiente onde esta espécie é encontrada (HARTMANN *et al.*, 1985).

A principal utilização de *A. amazonense* como inoculante, está relacionada com a cana-de-açúcar, onde a fixação biológica pode atingir 70% das necessidades da planta, têm sido estudada objetivando o desenvolvimento de biocombustíveis (REIS *et al.*, 2008). Foi demonstrado que *A. amazonense* é capaz de sobreviver por longos períodos de tempo tanto na condição líquida como gelatinosa, e sua associação com espécies do gênero *Gluconacetobacter*, *Herbaspirillum* e *Burkholderia* promovem o crescimento de cana-de-açúcar em condições de casa de vegetação (SILVA *et al.*, 2012).

Esta espécie bacteriana também mostrou resultados promissores na promoção do crescimento de arroz, vegetal no qual *A. amazonense* foi capaz não somente de promover o desenvolvimento graças à fixação biológica do nitrogênio, mas também em termos de síntese de auxinas, como visto na **Figura 1.1** reproduzida do trabalho original (RODRIGUES *et al.*, 2008).



**Figura 1.1** – Fixação biológica de nitrogênio e produção de hormônio de crescimento vegetal em isolados de *A. amazonense*. A atividade de nitrogenase foi determinada através do ensaio de redução de acetileno (gráfico de barras com desvio padrão) e a produção de hormônio vegetal foi determinada pela presença de auxina (demonstrado na curva por pontos) Os isolados bacterianos foram obtidos a partir de cultivos de arroz (RODRIGUES *et al.*, 2008).

*A. amazonense* também foi isolado da rizosfera de espécies de *Brachiaria*, espécie de forrageira utilizada como pastagem nas regiões do Cerrado e Mata Atlântica. Os resultados da análise de isolados de *A. amazonense* de três espécies do gênero *Brachiaria*, demonstraram que os mesmos são capazes de sintetizar auxinas. Além disso, a sazonalidade resultou em variação no número de microrganismos identificados, principalmente na região do Cerrado, enquanto na Mata Atlântica as maiores variações ocorreram em diferentes espécies de diferentes pastagens (REIS JUNIOR *et al.*, 2004).

### 1.1.2. *Azospirillum brasilense*

Esta espécie é a mais bem caracterizada do gênero. Foi demonstrado que ela é capaz de promover o crescimento de plantas de maneira direta através da fixação de nitrogênio, síntese de fitormônios como a auxina ácido indol acético

(AIA) e giberelinas e indireta pela inibição de patógenos presentes na rizosfera. O gene codificador da indol piruvato descarboxilase (*ipdC*) foi descrito como fundamental para a síntese de AIA (SPAEPEN *et al.*, 2007). Ortólogos deste gene, porém, não foram encontrados em diversos microrganismos sabidamente produtores desse fitormônio, demonstrando que este processo ainda é pouco conhecido em procariotos.

Além de auxinas, giberelinas e citocininas foi demonstrado que *A. brasilense* é capaz de produzir ácido abscísico (ABA), hormônio conhecido por ser sintetizado pelos vegetais quando submetidos a condições de estresse. Essa produção é aumentada com a adição de estresse salino e quando inoculado em *A. thaliana*, o vegetal dobra a quantidade interna de ABA quando comparado à planta sem o inoculante. Esses resultados contribuem para o entendimento do aumento da resistência a estresse salino em plantas inoculadas com *Azospirillum* (COEHN *et al.*, 2008). Sendo um microrganismo modelo de estudos, esta espécie teve seu genoma parcialmente disponibilizado em 2006 e finalizado em 2011 (WISNIEWSKI-DYÉ *et al.*, 2011).

### 1.1.3. *Azospirillum lipoferum*

*Azospirillum lipoferum* também promove o crescimento vegetal através da fixação do nitrogênio e produção de fitormônios. Foi demonstrado que a fixação de nitrogênio nessa espécie é induzida por aglutinina de germe de trigo e a indução está relacionada à presença de lectinas de trigo. Este resultado pode ser considerado relevante, pois esta espécie bacteriana realiza associação com este vegetal (KARPATI *et al.*, 1999).

Com relação a hormônios, testes com uma variante de arroz incapaz de sintetizar giberelinas foram realizados, atestando a capacidade do microrganismo metabolizar agliconas com a formação da giberelina Ga1 (BOTTINI *et al.*, 2004).

Uma característica de *A. lipoferum* é a capacidade de produção de variantes fenotípicas como resposta a mudanças ambientais, como as alterações que ocorrem na expressão de flagelos laterais ou polar na presença de meio sólido ou líquido (ALEXANDRE *et al.*, 1999) ou a perda de um plasmídeo contendo genes *acdRS*, codificantes da enzima ACC desaminase, relacionada

com a metabolização do precursor do etileno, com consequente aumento da resistência a estresse por alguns vegetais (PRIGENT-COMBARET *et al.*, 2008).

#### 1.1.4. *Azospirillum* sp. B510

A importância dessa espécie vem do fato de ter sido a primeira a demonstrar a capacidade de *Azospirillum* colonizar o interior da planta, induzindo o crescimento em arroz em testes de casa de vegetação (ISAWA *et al.*, 2010). Devido a essa característica, foi também a primeira espécie do gênero a ter seu genoma sequenciado (KANEKO *et al.*, 2010).

### **1.2. Avanços recentes no estudo de *Azospirillum* e novas possibilidades de análise**

Com o desenvolvimento de novas metodologias de sequenciamento e análise a era pós genômica visa principalmente o estudo da função gênica, onde sua organização nos genomas, afeta diretamente sua regulação e expressão (JANITZ, 2007).

Através da colaboração com o Laboratório Nacional de Computação Científica, o nosso grupo teve a oportunidade de realizar um sequenciamento do genoma da *A. amazonense*, onde regiões genômicas de interesse biotecnológico foram identificadas e caracterizadas. Neste trabalho foram analisadas proteínas do metabolismo do nitrogênio e carbono, *quorum sensing*, quimiotaxia, motilidade e outras, sendo que novas características foram identificadas para esta espécie, como os genes relacionados à fixação de carbono (Rabisco) e para metabolização de diferentes fontes desse elemento, como os genes *salRCAB* para utilização da salicina (SANT'ANNA *et al.*, 2011) (ANEXO I). Assim como a análise de RNA 16S, a presença dos genes *salRCAB* aproximou *A. amazonense* de *A. irakense*, já que o operon citado foi encontrado somente nestes dois organismos dentro do gênero *Azospirillum*. A capacidade de metabolização da salicina foi testada experimentalmente em *A. irakense*, sendo que demonstrou ser regulada por um membro da família de reguladores transcricionais da família LacI/GalR (SOMERS *et al.*, 2000).

Em *A. brasilense*, uma análise global de genes diferencialmente expressos na presença da auxina AIA administrada pelas vias endógena e exógena, foi realizada no mesmo ano, utilizando a metodologia de micro arranjos. O estudo demonstrou que um total de 160 genes foram significativamente regulados por AIA por ambas as vias. O sistema de redução de nitratos, transportadores tipo ABC, além de genes do sistema de secreção foram algumas das classes que tiveram um padrão de expressão alterado quando na presença desse fitormônio (VAN PUYVELDE *et al.*, 2011).

Com relação a possibilidades de identificar características de interesse presentes em *Azospirillum*, podemos destacar a disponibilização do banco de sequências relacionadas ao metabolismo de hormônios em *Arabidopsis thaliana* (AHD), banco de proteínas de transporte (TCDB) e secreção do tipo IV (SecRet4), além da identificação de sequências relacionadas a fagos e outros elementos genéticos móveis (PHAST e ACLAME respectivamente). Bancos de dados de Ortólogos do COG e KEGG possuem sequências bem caracterizadas em termos de curadoria, sendo que ainda nos é oferecida a possibilidade da montagem de bancos locais para análises mais específicas (SAIER *et al.*, 2006; BI *et al.*, 2013; ZHOU *et al.*, 2011; LEPLAE *et al.*, 2004; KANEHISA, 2002).

A comparação entre genes de bactérias com relação a outros filos auxiliou, por exemplo, no estudo de transferência multilateral de hidroperoxidases, tendo sido comprovada sua mobilidade entre microrganismo e fungo, assim como do procarioto para protistas que originaram algas e plantas (KLOTZ & LOEWEN, 2003). Tratando de genes relacionados a fitormônios, também foi comprovado um vínculo com espécies de procariotos, quando foi demonstrado que receptores do hormônio vegetal etileno, relacionado à senescência em vegetais, foi herdado dos plastídios, em experimentos com *Anabaena* e *Synechocystis* (MOUNT & CHANG, 2002).

### 1.3. Análise diferencial de sequências genômicas

Com a disponibilização de novas técnicas em biologia molecular e com o acúmulo de informação gerado pelo conhecimento dos genomas, tornou-se necessário a disponibilização de novas ferramentas de análises de dados. Devido ao considerável número de sequências com problemas de anotação já depositadas e pelas dificuldades de atualização com base em análises de sequências, um número cada vez maior de bancos de dados específicos de sequências têm sido apresentados (TATUSOV *et al.*, 2000; ZHANG & LIN, 2009; PENG *et al.*, 2009; FINN *et al.*, 2010).

Metodologias para análise *in vitro* e *in silico* de sequências de DNA e proteínas conservadas e diferenciais entre organismos ou bancos de dados específicos são fundamentais para uma melhor visualização da evolução dos genomas.

Comparações *in silico* de genomas podem ser realizadas com a ferramenta mGenomeSubtractor diretamente na internet, enquanto localmente é possível realizar esse procedimento com programas tipo BioEdit (TIPPMANN, 2004). Alguns dos bancos de dados específicos citados disponibilizam ferramentas de comparações. O mGenomeSubtractor, por exemplo, mostra a identidade de genes ou proteínas conservados em diferentes organismos (SHAO *et al.*, 2010). Esta ferramenta disponibilizada recentemente foi utilizada na análise de nove genomas de *Pseudomonas*, apresentando regiões conservadas e rearranjos que ocorrem naturalmente entre esses microrganismos (QI *et al.*, 2011). Também foi importante para comparações de genomas de espécies de *Corynebacterium*, onde resultou na identificação de grande quantidade de sequências virais nesses genomas (SANGAL *et al.*, 2012). Comparações utilizando a mesma ferramenta entre bactérias do gênero *Klebsiella* resultaram na identificação de genes relacionados com conjugação bacteriana e consequente “espalhamento” de genes relacionados à resistência microbiana (LIU *et al.*, 2012). Testes com o mGenomeSubtractor também auxiliaram na diferenciação de bactérias fixadoras de nitrogênio, quando foram comparadas espécies de *Sphingomonas*, caracterizando a primeira linhagem capaz de degradar o ácido fenazina 1-



carboxílico, precursor da fenazina, conhecida pela atividade antifúngica de amplo espectro (MA *et al.*, 2012).

Experimentos recentes de genômica comparativa utilizando principalmente a análise em bancos específicos de sequência e análise de vias metabólicas baseadas em ortólogos também têm obtido resultados promissores. Um exemplo dessas comparações foi realizado com protozoários do gênero *Cryptosporidium* que parasitam praticamente todas as classes de vertebrados. Através de genômica diferencial realizada com análises das vias de ortólogos do KEGG e específico para *Cryptosporidium* (CryptoDB) foi possível a identificação de genes vinculados com a patogenicidade em humanos com relação à infecção de outras espécies de vertebrados (MAZURIE *et al.*, 2013; KANEHISA, 2002; PUIU *et al.*, 2004).

Genômica diferencial também foi utilizada em vegetais, para a caracterização de polimorfismos de interesse biológico em espécies do gênero da cerejeira (*Prunoideae*) (KOEPEKE *et al.*, 2013).

Em relação às bactérias do solo, genômica comparativa foi realizada com a finalidade de identificar genes diferencialmente expressos na habilidade de espécies de *Collimonas* inibir a ação de fungos patogênicos no solo. Além de atividade fungistática esse gênero possui a capacidade de degradar quitina e viver sobre as hifas dos fungos, sendo importante no controle biológico desses organismos (MELA *et al.*, 2012).

A caracterização de sequências específicas, realizada *in vitro*, especialmente útil para avaliar uma espécie que não possui o genoma sequenciado, comparando com organismos próximos já sequenciados, pode ser obtida com a metodologia de RDA (do inglês *Representational Difference Analysis*). Ela combina as técnicas de hibridização subtrativa seguida por amplificação diferencial utilizando reação em cadeia da polimerase (PCR). Essa metodologia foi utilizada na análise de sequências genômicas diferenciais relacionadas ao câncer (LISITSYN *et al.*, 1993); No isolamento de regiões responsáveis pela patogenicidade de espécies do gênero *Neisseria* (TINSLEY & NASSIF, 1996); Relacionadas a regiões polimórficas em genomas (BALDOCCHI & FLAHERTY, 1997); Relacionadas a

genes do comportamento motor em peixe zebra (SATO & MISHINA, 2003); Relacionada à identificação de sequências únicas de procariotos (ALLEN *et al.*, 2003; FELSKÉ, 2002).

Atualmente a metodologia de RDA tem sido utilizada principalmente no isolamento de genes diferencialmente expressos, onde a condição na qual esses genes são ativados ou silenciados é utilizada para a preparação do TESTER, que é comparada com uma condição basal de expressão (DRIVER). Essa técnica, porém, possui melhor eficiência em eucariotos, quando a cauda poliadenilada do mRNA é utilizada na síntese *in vitro* do cDNA que será utilizado como base para a metodologia de RDA. As publicações mais recentes envolvem o isolamento de genes diferencialmente expressos na interação de do patógeno *Giardia duodenalis* com as células do epitélio intestinal (MA'AYEH & BROOK-CARTER, 2012); Genes diferencialmente expressos na endometriose (CHEN *et al.*, 2012) e diferencialmente expressos em ovários que originam abelhas rainhas ou operárias, em um processo regulado pela presença de hormônios no gênero *Apis* (HUMANN & HARTFELDER, 2011). Em relação a leveduras, genes diferencialmente expressos durante a interação de *Cryptococcus neoformans* e *C. gatti* com macrófagos de ratos foram caracterizados com essa metodologia, identificando genes regulados durante essa fase de importância fundamental para a infecção pelo patógeno (FAGANELLO *et al.*, 2009).

Os resultados apresentados para outros organismos no que tange a anotação, caracterização e diferenciação de genomas, juntamente com a crescente identificação de novas espécies e proteínas de interesse biotecnológico, tornam possível e necessária, uma análise detalhada dos genomas recentemente disponibilizados de quatro espécies do gênero *Azospirillum*. Os resultados das primeiras análises dos genomas de *A. brasilense* e *A. lipoferum* apontaram a origem aquática dessas bactérias e demonstraram a crescente interação com genomas de fagos e outros elementos genéticos móveis após a invasão do ambiente terrestre, com a conseqüente diferenciação dos genomas (WISNIEWSKI-DYÉ *et al.*, 2011). A presença desses elementos é conhecida em *Azospirillum*, tendo sido identificado inclusive um fago integrado no genoma de *A. brasilense* CD (BOYER *et al.*, 2008). Em *A. amazonense* a

presença de elementos conjugativos pode atuar de maneira semelhante ao que ocorre em *Klebsiella*, induzindo a diferenciação do genoma (LIU *et al.*, 2012). A utilização das mais recentes ferramentas de bioinformática comparativa objetiva auxiliar na compreensão da forma individual de ação de cada espécie na associação e promoção do crescimento vegetal. Neste trabalho, enfatizamos a caracterização e identificação das principais diferenças encontradas em *A. amazonense*, espécie alvo dos atuais estudos realizados pelo Laboratório de Microrganismos Diazotróficos da UFRGS.

#### **1.4. Objetivos**

Como objetivo geral, o presente trabalho visa analisar as diferenças genômicas da espécie *Azospirillum amazonense* com relação às espécies relacionadas: *Azospirillum* sp. B510, *A. brasilense* 245 e *A. lipoferum* 4B, discutindo as principais diferenças identificadas em *A. amazonense* como: Proteínas relacionadas com o metabolismo de hormônios vegetais; sistema de secreção bacteriano do tipo IV e proteínas de transporte em geral; troca e modificação de material genético por fagos e plasmídeos e imunidade adquirida.

Aplicar a metodologia de Análise de Diferença Representacional (RDA), para isolamento *in vitro* de diferenças genômicas presentes em organismos que não estão completamente sequenciados, na busca por características diferenciais entre as espécies *A. amazonense* e *A. brasilense*.

Apresentar uma metodologia prática e rápida para anotação e comparação entre genomas, caracterizando funcionalmente as sequências nos diferentes bancos de dados disponibilizados a partir das novas metodologias de bioinformática oferecidas nos anos recentes.

## 2. Materiais e Métodos

### 2.1. Análise *in Silico*

#### 2.1.1. Obtenção dos genomas

O resultado do sequenciamento do genoma de *A. amazonense* está depositado com número de acesso AFBX00000000.1 ou PRJNA65263 contendo 1014 agrupamentos (*contigs*), correspondentes a 6.318.414 pares de bases depositados e foi importado do endereço [ftp://ftp.ncbi.nih.gov/genomes/Bacteria\\_DRAFT/Azospirillum\\_amazonense\\_Y2\\_uid73583/](ftp://ftp.ncbi.nih.gov/genomes/Bacteria_DRAFT/Azospirillum_amazonense_Y2_uid73583/). Os números de acesso dos genomas utilizados neste trabalho são, respectivamente: *Azospirillum* sp. B510, projeto PRJNA46085 e PRJDA32551; *Azospirillum brasilense* Sp245, projeto PRJEA162161 e PRJEA70627; *Azospirillum lipoferum* 4B, projeto PRJNA82343, PRJEA50367; *Rhodospirillum centenum*, projeto PRJNA58805; *Bradyrhizobium Japonicum* USDA 110, projeto PRJNA57599 e PRJNA17; *Streptomyces scabei* 87.22, projeto PRJEA40749; *Caulobacter segnis* projeto PRJNA41709; *Pseudomonas brassicacearum* subsp. *brassicacearum* NFM42, projeto PRJNA66303 e PRJEA63495; *Novosphingobium aromaticivorans* DSM 12444, projeto PRJNA57747; *Burkholderia* sp. CCGE1001 PRJNA37717.

#### 2.1.2. Metodologia de análise *in silico*

Inicialmente foi realizada a predição das sequências de proteínas do genoma de *A. amazonense* utilizando *Azospirillum* sp. B510 como referência para anotação, a partir dos *contigs* sequenciados. Esta análise foi realizada no endereço <http://www.xbase.ac.uk/annotation/> (CHAUDHURI *et al.*, 2008). A utilização de *Azospirillum* sp. B510 como referencia, possibilitou a maximização das proteínas em comum entre esta e a espécie *A. amazonense*, alvo deste estudo.

O resultado da anotação das proteínas de *A. amazonense* obtido utilizando esta metodologia, foi analisado no endereço <http://lfz.corefacility.ca/gbk2ptt/> para conversão do arquivo no formato GenBank (\*.gbk) para o formato tabular. Esta tabela foi utilizada como base para adicionar todos os outros resultados gerados nas análises posteriores.

A primeira análise realizada com o banco de proteínas resultantes desta anotação foi uma comparação, utilizando o programa BioEdit (TIPPMANN, 2004), com as sequências de proteínas do genoma de *A. amazonense* já depositadas no banco de dados GenBank (SANT'ANNA *et al.*, 2011).

Com o objetivo de realizar uma anotação mais abrangente, foi realizada uma comparação entre as sequências proteicas de *A. amazonense* com as sequências de proteínas de 10 genomas bacterianos que apresentaram maior similaridade com sequências de *A. amazonense* na primeira análise do genoma publicada em 2011 (SANT'ANNA *et al.*, 2011). Para esta análise, os arquivos contendo as sequências das proteínas de cada espécie, foram obtidos no endereço <ftp://ftp.ncbi.nih.gov/genomes/> e reunidos em bancos de dados locais independentes.

Dados suplementares para anotação foram obtidos no WebMGA (do inglês *Web Server for Metagenomic Analysis*) no endereço <http://weizhonglab.ucsd.edu/metagenomic-analysis/server/> para auxiliar na anotação de proteínas sem função conhecida (WU *et al.*, 2011). Este servidor reúne bancos de dados de sequências curadas como:

COG - (do inglês *Clusters of Orthologous Groups*), de proteínas codificadas nos genomas completos. Cada COG, representa proteínas individuais ou grupos de parálogos encontrados ao menos em três linhagens sequenciadas, servindo de grupo ancestral, sendo portanto útil em análises filogenéticas (TATUSOV *et al.*, 2000).

PFAM - A base de dados Pfam é uma grande coleção de famílias de proteínas, representada por alinhamentos de múltiplas sequências pelo modelo de Markov. Assim, as proteínas que possuem uma ou mais regiões funcionais, chamadas domínios, podem ser agrupadas conforme sua função (FINN *et al.*, 2010).

CDD – O banco de dados de domínios conservados é baseado em estruturas tridimensionais para realizar a anotação da função de proteínas (MARCHLER-BAUER *et al.*, 2013).

Produtos de proteínas que não encontraram similares nestas comparações foram caracterizados posteriormente pelo programa PSI Blast (ALTSCHUL *et al.*, 1997) no NCBI.

A integração dos produtos das proteínas das espécies *A. amazonense*, *Azospirillum* sp. B510, *A. brasilense* 245 e *A. lipoferum* 4B, em rotas metabólicas do KEGG (do inglês *Kyoto Encyclopedia of Genes and Genomes*) criado para a análise e entendimento de funções biológicas em grande escala, foi realizada no KAAS (do inglês *KEGG Automatic Annotation Server*) no endereço <http://www.genome.jp/tools/kaas/> (MORIYA *et al.*, 2007).

Para caracterização de sequências únicas de *A. amazonense*, com relação a outros membros do gênero *Azospirillum*, assim como o cálculo dos valores de Homologia (Valor-H) para cada proteína, foi realizada uma análise por hibridização subtrativa *in silico* com a ferramenta mGenomeSubtractor, no endereço <http://bioinfo-mml.sjtu.edu.cn/mGS/> além de análises de similaridade realizadas localmente com BioEdit (SHAO *et al.*, 2010).

As proteínas possivelmente relacionados ao metabolismo de fitormônios foram analisados no AHD (do inglês *Arabidopsis Hormone Database*) no endereço <http://ahd.cbi.pku.edu.cn/> (PENG *et al.*, 2009). Proteínas de *A. brasilense*, que foram analisadas no transcriptoma regulado por auxina (VAN PUYVELDE *et al.*, 2011), foram obtidos no DRAFT de *A. brasilense* no endereço <http://genome.ornl.gov/microbial/abra/19sep08/> e foram comparados com proteínas de *A. amazonense* com o programa BioEdit.

O conjunto de proteínas das espécies *A. amazonense*, *Azospirillum* sp. B510, *A. brasilense* e *A. lipoferum*, relacionado ao metabolismo de fitormônios detectado pelo KAAS, foi localizado utilizando *A. thaliana* e *O. sativa* como referência. Outras análises que utilizaram o KAAS foram baseadas nos genomas de *Rhodospirillum centenum*, *Bradyrhizobium japonicum* USDA110, *Caulobacter*

*segnis* e *Azospirillum* sp. B510 por serem os genomas mais próximos da espécie alvo do estudo.

A comparação do sistema de secreção bacteriana do tipo IV, que envolve troca de material genético, foi realizada no programa SecRet4 (do inglês *Type 4 Secretion system Resource* - <http://db-mml.sjtu.edu.cn/SecReT4/index.php>) (BI *et al.*, 2013) e as proteínas envolvidas com transporte em geral foram comparadas no banco de dados TCDB (do inglês *Transporter Classification Data Base* - <http://www.tcdb.org/>) (SAIER *et al.*, 2006).

Proteínas que codificam produtos essenciais ao metabolismo celular e fatores de virulência foram analisadas nos endereços <http://tubic.tju.edu.cn/deg/> e <http://www.mgc.ac.cn/VFs/main.htm>, respectivamente (ZHANG & LIN, 2009); (CHEN *et al.*, 2012).

As regiões genômicas de herança viral foram analisadas com PHAST (do inglês *Phage Search Tool*) no endereço <http://phast.wishartlab.com/index.html> (ZHOU *et al.*, 2011) e as regiões contendo proteínas associadas ao sistema imune bacteriano foram localizadas com CRISPR Finder Online (<http://crispr.u-psud.fr/Server/>) (GRISSA *et al.*, 2007). Proteínas não agrupadas, herdadas de plasmídeos, vírus e prófagos nestes genomas, foram comparadas no ACLAME (do inglês *A Classification of Mobile Genetic Elements*) no endereço <http://aclame.ulb.ac.be/> (LEPLAE *et al.*, 2004).

### 2.1.3. Cálculos de análise de similaridade

#### 2.1.3.1. Escore

Por definição, o melhor alinhamento é aquele que maximiza o escore. Este é estabelecido através de uma matriz de peso na qual a identidade de um aminoácido ou nucleotídeo soma pontos, enquanto uma troca (deleção/inserção) diminui pontos de escore. Desta maneira ele não avalia o tamanho da sequência ou o tamanho do banco de dados como ocorre com o valor de E. As matrizes e valores-padrão comumente utilizados podem ser vistos em <http://www.ncbi.nlm.nih.gov/BLAST/blastcguihelp.shtml#Matrix>.

#### 2.1.3.2. Valor E

Este cálculo estabelece a chance de um alinhamento igual ao do organismo em estudo ser encontrado por acaso ao fazer a busca num banco de dados. Quanto menor o valor de “e-value” maior a similaridade entre as sequências, portanto, quanto maior for esse valor, maior a probabilidade destas identidades serem devidas ao acaso. O valor de e-value significativo de acordo com Pearson (PEARSON, 1900) seleciona valores  $< 0,02$  enquanto outros autores recomendam valores  $< 0,005$  (ALTSCHUL *et al.*, 1997). Desta maneira, foi estabelecido um ponto de corte superior ( $e^{-10}$ ), definido como padrão na predição e anotação inicial com *Azospirillum* sp. B510 como referência (CHAUDHURI *et al.*, 2008), e outro inferior (0,005) na análise com PSI Blast no GenBank. Proteínas cujos alinhamentos possuem valor E entre estes limites foram definidas como hipotéticas conservadas e as com valores superiores a 0,005, proteínas hipotéticas de *A. amazonense*.

#### 2.1.3.3. Valor H

A principal utilidade do valor H é integrar outros indicadores como o valor E, com o objetivo de estabelecer um valor de similaridade entre duas sequências com valores definidos entre 0 e 1, para mínimo e máximo de similaridade, facilitando a observação da distribuição das proteínas compartilhadas entre dois bancos de sequências. No entanto, assim como no valor E, o ponto ideal de corte para cada análise deve ser estabelecido de maneira específica. Devido ao grande número de proteínas exclusivas, e para não haver dúvidas quanto à especificidade, somente proteínas com valor H igual a zero foram classificadas como diferenciais na análise do mGenomeSubtractor entre as espécies de *Azospirillum*.



## 2.2. Análise experimental

Devido ao fato do genoma de *A. amazonense* estar apenas parcialmente sequenciado e pela metodologia estar padronizada em nosso laboratório, uma análise experimental por RDA (do inglês *Representational difference analysis*) foi realizada entre *A. amazonense* e *A. brasilense*. Com o grande número de isolados caracterizados e novas espécies identificadas, esta metodologia poderia ser utilizada para análise das diferenças entre as espécies cujo genoma não é conhecido, como forma de identificar suas características exclusivas.

### 2.2.1. Condições de cultivo e purificação do DNA

*A. amazonense* Y2 (ATCC 35120) foi cultivado em meio M79 (10 g/L de sacarose como fonte de carbono, 0,1 g/L de K<sub>2</sub>HPO<sub>4</sub>, 0,4 g/L de KH<sub>2</sub>PO<sub>4</sub>, 0,2 g/L de MgCl<sub>2</sub>.7H<sub>2</sub>O, 0,1 g/L de NaCl, 0,4 g/L de extrato de levedura pH 6,5) a 35°C por 18 horas, com agitação de 150 rpm. *A. brasilense* 245 foi cultivado em meio LB (SAMBROOK & RUSSEL, 2001) nas mesmas condições acima descritas.

DNA genômico de ambas as espécies foi isolado com o uso do reagente Trizol (Life Technologies) conforme as instruções do fabricante. Após a purificação as amostras foram analisadas em gel de agarose e quantificadas com a utilização de fluorímetro Qubit (Life Technologies).

### 2.2.2. Análise de diferença representacional (RDA)

#### 2.2.2.1. Preparação dos DNAs de *A. amazonense* (TESTER) e *A. brasilense* (DRIVER)

O genoma das duas espécies foi digerido com enzima Sau3AI, que reconhece o sítio de clivagem GATC, e purificado em coluna GFX (GE Healthcare) com o objetivo de eliminar fragmentos menores que 100pb e maiores de 48kb. O DNA de *A. amazonense* foi ligado a um par de adaptadores para a preparação do TESTER. O DNA de *A. brasilense*, utilizado como controle para subtrair de *A. amazonense* as sequências comuns à ambas, não foi ligado a adaptadores. Uma reação de PCR foi realizada para testar a ligação dos oligonucleotídeos. Os três

pares de adaptadores utilizados bem como características das sequências podem ser visualizados na **Tabela 2.1**.

**Tabela 2.1** - Identificação e características dos três pares de oligonucleotídeos adaptadores utilizados no RDA.

<b>Primer</b>	<b>Sequência</b>	<b>Tm</b>
RBam 12	5'GATCCTGCGTGA 3'	42,0 °C
RBam 24	5'AGCACTCTCCAGCCTCTCACCGAG 3'	64,1 °C
NBam 12	5'GATCCTCCCTCG 3'	41,8 °C
NBam 24	5'AGGCAAGTGTGCTATCCGAGGGAG 3'	62,5 °C
JBam 12	5'GATCCGTTCATG 3'	36,0 °C
JBam 24	5'ACCGACGTGCGACTATCCA 3'	55,2 °C

#### 2.2.2.2. Hibridização Subtrativa

Uma quantidade de 5µg do DNA genômico de *A. brasilense* (DRIVER) foram utilizadas em cada um dos três ciclos de hibridização subtrativa. O DNA genômico de *A. amazonense* (TESTER) ligado aos adaptadores foi utilizado em concentrações de 50ng, 5ng e 0,5ng, respectivamente em cada etapa da técnica, sendo o último em uma relação de 1:10000 TESTER:DRIVER.

Para a hibridização subtrativa foi utilizada a metodologia de OSPERT (do inglês *Oscillating Phenol Emulsion Reassociation Technique*), que utiliza ciclos de temperatura em emulsão fenólica para aumentar a eficiência do pareamento dos fragmentos de DNA em até 25 mil vezes, além de ser realizada em apenas 3 horas, quando comparado às 18-24h necessárias das metodologias anteriormente descritas (BRUZEL & CHEUNG, 2006).

A mistura do TESTER/DRIVER foi desnaturada por 15 min em um volume final de 34,6µl de 0,3N de NaOH posteriormente neutralizada com 5,4 µl de 3M de MOPS, 50µl de 1,5M de NaCl, 2mM EDTA e 10µl de tampão fenólico pH 8 foram adicionados à mistura sendo esta aquecida a 65°C para gerar a emulsão.

Posteriormente esta foi colocada em termociclador programado para 99 ciclos de 15s a 25°C e 15s a 65°C.

#### 2.2.2.3. Amplificação diferencial

Para realizar a amplificação diferencial, a amostra foi aquecida a 94 °C por 5 min, seguido de 30 ciclos de 94°C por 1min, 70°C por 1 min, 72°C por 3 min. No final a amostra foi incubada por 3 min a 70°C, 72°C por 10 min e 4°C por 10min.

As PCRs foram realizadas em tubos de micro centrífuga de 0,5 ml, utilizando-se, 5 picomol de *primer*, 200 µM de dNTPs, 2U de Taq DNA polimerase Go Taq Flexi, 3 mM de MgCl<sub>2</sub>, 10% de tampão de reação e água MilliQ para um volume final de reação de 50 µl. A reação em cadeia da polimerase (PCR) foi realizada em termociclador automático (MJ Research) utilizando-se programas específicos. Após a amplificação, 2,5 µl das amostras foram analisados por eletroforese em gel de agarose 1%. Ao final da amplificação a reação foi purificada utilizando-se o kit GFX (GE Healthcare).

#### 2.2.2.4. Clonagem, Transformação e Sequenciamento

Após 100 ciclos de OsPERT em termociclador variando entre 15s a 25°C e 15s a 65°C, purificação da emulsão fenólica e amplificação das sequências da última etapa de hibridização subtrativa (DP3), os fragmentos foram clonados em vetor pUC18/SmaI. As transformações bacterianas foram realizadas utilizando-se células competentes de *E. coli*, preparadas pelo método de transformação química com CaCl<sub>2</sub> descrito em (SAMBROOK & RUSSELL, 2001)

Realizado o choque térmico, as células foram imediatamente transferidas para um tubo de ensaio contendo 1ml de SOC líquido e incubadas com agitação (200 rpm) a 37°C por 1h. Alíquotas de 100 µl dessa preparação foram plaqueadas

em LB sólido contendo concentrações adequadas de antibióticos, IPTG e X-Gal. A seleção das colônias recombinantes foi feita através do sistema enzimático da  $\beta$ -galactosidase. As colônias brancas foram multiplicadas em LB líquido com agitação de 200 rpm a 37°C por 18 h. O DNA plasmidial foi purificado, submetido à eletroforese em gel de agarose e, quando necessário, clivado com as enzimas de restrição BamHI e EcoRI para liberação do fragmento. Para a reação de sequenciamento das amostras provenientes do RDA, foi utilizado o sequenciador automático MEGABACE Amersham Biosciences™ no Laboratório de sequenciamento do Centro de Biotecnologia.

### 3. Resultados

#### 3.1. Anotação automática

Para análise da anotação automática dos produtos codificados pelas sequências de DNA presentes no genoma de *A. amazonense*, as prováveis proteínas originadas ou ORFs (sequência aberta de leitura) geradas pelo sequenciamento de *A. amazonense* (SANT'ANNA *et al.*, 2011) foram analisadas no *Xbase Annotation Service* (CHAUDHURI *et al.*, 2008). Nesta análise foi utilizado *Azospirillum* sp. B510 como referência na comparação de similaridade de sequências proteicas com valores padrão (Proteínas com no mínimo 30 resíduos e Valor E < e-10 para classificar com o mesmo nome o produto em *A. amazonense*). Estes resultados permitiram estabelecer a predição de 5.496 proteínas putativas, sendo que 2.165 apresentaram similaridade maior que e-10 e foram consideradas como “Baixa similaridade com *Azospirillum* sp. B510”. As 520 proteínas anotadas como hipotéticas em *Azospirillum* sp. B510, e que apresentaram valor de E menor que e-10 em *A. amazonense*, foram marcadas como proteínas conservadas hipotéticas em *A. amazonense* nesta comparação. O resultado original desta análise, assim como as sequências das 5496 proteínas para este genoma preditas neste trabalho podem ser obtidos no endereço <http://www.xbase.ac.uk/annotation/results/rWn50Rn6LVRucf55SWfsvHfoqdpmd655/>

A presença de 2.165 sequências de *A. amazonense* que não apresentaram similaridade com proteínas de *Azospirillum* sp. B510 sugeriu que uma comparação de sequências mais abrangente deveria ser realizada. Portanto, uma nova comparação de sequências utilizando 10 genomas microbianos, que apresentaram similaridade com *A. amazonense* no primeiro trabalho sobre o genoma desta espécie (SANT'ANNA *et al.*, 2011), foi realizada. O resultado desta análise, demonstrando o número de proteínas de *A. amazonense* que apresentaram melhor similaridade (*Best hit*) com cada um dos genomas, pode ser visualizado na **Tabela 3.1**.

**Tabela 3.1** - Número de sequências de proteínas de *A. amazonense* com maior similaridade por espécie comparada.

<b>Espécie</b>	<b>Melhor similaridade</b>
<i>Rhodospirillum centenum</i> SW	1.411
<i>Caulobacter segnis</i> ATCC 21756	447
<i>Bradyrhizobium japonicum</i> USDA 110	445
<i>Azospirillum brasilense</i> Sp245	439
<i>Azospirillum lipoferum</i> 4B	363
<i>Azospirillum</i> sp. B510	330
<i>Burkholderia</i> sp. CCGE1001	248
<i>Novosphingobium aromaticivorans</i> DSM 12444	197
<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	185
<i>Streptomyces scabei</i> 87.22i	168
Baixa similaridade nesta análise	1263
<b>Total</b>	<b>5.496</b>

A análise da **Tabela 3.1** apresenta *Rhodospirillum centenum* como o organismo com maior número de proteínas similares com proteínas de *A. amazonense*. A similaridade das sequências proteicas presentes no genoma de *Bradyrhizobium japonicum* e *Caulobacter segnis* com aquelas de *A. amazonense* aparecem em segundo lugar, com praticamente o mesmo número de melhores alinhamentos (445 proteínas e 447 proteínas respectivamente). Interessantemente, as três espécies de *Azospirillum* analisadas aparecem posteriormente na tabela, possuindo menor número de melhores alinhamentos com *A. amazonense* em comparação com as espécies acima indicadas.

Para finalizar a primeira etapa, uma análise comparativa foi realizada para todas as proteínas no banco de dados do GenBank utilizando o PSI-Blast. Neste experimento, 4.896 proteínas de *A. amazonense* encontraram correspondente, sendo que 1.222 foram classificadas como hipotéticas conservadas. Nesta análise 500 proteínas de *A. amazonense* não apresentaram similaridade com outras depositadas no banco de dados GenBank.

### 3.1.1. Validação da análise comparativa das proteínas

Com o objetivo de selecionar a melhor anotação para cada proteína em *A. amazonense*, o seguinte procedimento foi realizado. Inicialmente a anotação baseada nos 10 genomas (apresentados na **Tabela 3.1**) foi selecionada. Ao relacionar com os resultados do PSI Blast foi realizado um quociente entre o valor de E do PSI blast e do Blastp local. Devido ao fato dos 10 genomas selecionados serem bem conhecidos e sua anotação avaliada há mais tempo, o produto de alinhamentos com alta similaridade ( $<e^{-100}$ ) foi assumido como verdadeiro para os 10 genomas. Pelo mesmo motivo, valores de quociente  $>e^{-10}$  também definiram o produto para os 10 genomas, significando que quando o valor de E para os 10 genomas for próximo ao PSI blast, a anotação do primeiro é selecionada. Os demais produtos, com significativa diferença entre o valor E dos alinhamentos, foram relacionados ao PSI blast, até o valor de E de  $e^{-10}$  (corte xbase). O produto dos alinhamentos que possuíam valor de E entre  $e^{-10}$  até 0,005 (que é o ponto de corte original do PSI blast e foi definido aqui como corte inferior) ficaram com o produto do PSI blast. As proteínas com valor E  $>0,005$  foram consideradas como hipotéticas de *A. amazonense*.

Posteriormente, a anotação foi curada conforme análises em bancos de dados especializados e discutidos neste trabalho. Proteínas que possuíam similaridade nos bancos de ortólogos COG, KEGG, e cl (CDD *conserved domain*) entre outros utilizados neste trabalho com valor de E menor do que o produto identificado pelo PSI Blast, foram identificadas com o código correspondente.

Essa marcação adicional facilita as análises cruzadas dos resultados quando, por exemplo, uma destas proteínas participa de mais de uma via metabólica analisada, além de informar possíveis domínios conservados nas sequências sem função estabelecida, para gerar o nome completo do produto final das 5.496 proteínas identificadas em *A. amazonense*. Todos estes dados foram incluídos no arquivo, em formato fasta, que foi utilizado para o desenvolvimento das tabelas de cada análise do presente trabalho.

### 3.1.2. Análise comparativa das vias metabólicas entre as proteínas codificadas pelos genomas das quatro espécies de *Azospirillum*

Com o objetivo de comparar as principais vias metabólicas para as quatro espécies do gênero *Azospirillum* os genomas foram submetidos a um teste de similaridade nos bancos de ortólogos do COG e KEGG. A **Tabela 3.2** mostra as rotas metabólicas encontradas (classificadas no COG), destacando as vias em que o genoma de *A. amazonense* codifica para o maior número de proteínas relativo a pelo menos uma das espécies relacionadas.

**Tabela 3.2** – Rotas metabólicas em que *A. amazonense* possui maior número de proteínas com relação a pelo menos uma das espécies comparadas.

Rota Metabólica COG	Aam	AB510	Abra	Alipo	Classe
Transporte e metabolismo de lipídeos	188	253	233	184	+ I
Biossíntese de compostos secundários, transporte e catabolismo	153	191	193	150	+ Q
Transporte intracelular, vesicular e secreção	117	111	+ 116	+ 90	+ U
Mecanismos de defesa	92	84	+ 76	+ 77	+ V

Aam - *A. amazonense*; AB510 – *Azospirillum* sp. B510, Abra – *A. brasilense*, Alipo- *A. lipoferum*.  
+ representa o maior número encontrado em *A. amazonense*.

Os resultados gerados na análise das rotas metabólicas, classificadas pelo KEGG (no site KAAS), em que *A. amazonense* possui maior número de proteínas em comparação com os quatro genomas pode ser visualizada na **Tabela 3.3**.



**Tabela 3.3** - Principais rotas metabólicas que apresentam maior número de proteínas codificadas no genoma de *A. amazonense*.

Rota Metabólica KEGG	Descrição	Amz	AB510	Abra	Alipo
ko00190	Fosforilação oxidativa	49	47	47	48
ko00280	Degradação de valina, leucina e isoleucina	29	25	21	25
ko04146	Peroxissomo	18	16	15	13
ko00380	Metabolismo do triptofano	15	14	9	12
ko05010	Doença de Alzheimer	14	13	11	13
ko00052	Metabolismo da galactose	12	9	9	8
ko05012	Doença de Parkinson	12	11	9	0
ko00130	Biossíntese de ubiquinona e outros terpenoides	11	10	10	10
ko00072	Síntese e degradação de corpos cetônicos	6	4	3	4
ko04142	Lisossomo	6	3	3	2
ko04260	Contração do músculo cardíaco	6	5	3	5
ko00511	Degradação de outros glicanos	5	1	1	1
ko05120	Sinalização celular epitelial por infecção por <i>H. pylori</i>	5	0	3	0

*A. amazonense* (Amz); *Azospirillum* sp. B510 (AB510); *A. brasilense* (Abra) e *A. lipoferum* (Alipo). Foi utilizada a classificação de ortólogos do KEGG.

Destacam-se na análise da **Tabela 3.2** o sistema de transporte, secreção e mecanismos de defesa e na **Tabela 3.3** a produção de terpenoides, a utilização da galactose e vias de aminoácidos como a valina/leucina e triptofano. A presença de proteínas relacionadas ao metabolismo do triptofano em *A. amazonense* apresenta relevância, pois este aminoácido está envolvido, em outras espécies de *Azospirillum*, com a produção de ácido indol acético, auxina que atua como principal hormônio de desenvolvimento das plantas (CARREÑO-LOPEZ *et al.*, 2000; BARTEI, 1997).

### 3.2. Comparação entre genomas

#### 3.2.1. Análise comparativa *in silico* entre os genomas de *Azospirillum*

Com o objetivo de realizar a comparação dos genomas dentro do gênero *Azospirillum*, o conjunto de todas as sequências de proteínas de *Azospirillum* sp B510, *A. brasilense* e *A. lipoferum* foram respectivamente comparados com as

5496 proteínas preditas no genoma de *A. amazonense* (utilizando blastp local), com o objetivo de visualizar as proteínas exclusivas (>e-10) e conservadas (<e-10) de *A. amazonense* com relação às outras três espécies. Essa análise permitiu concluir que *A. brasilense* possui maior número de ortólogos em *A. amazonense* entre os membros deste gênero, enquanto *Azospirillum* sp. B510 e *A. lipoferum* resultaram em praticamente o mesmo número de proteínas com melhor alinhamento com as correspondentes em *A. amazonense* (ver **Tabela 3.4**).

**Tabela 3.4** - Número de sequências de proteínas de *A. amazonense* com melhor similaridade por espécie de *Azospirillum* comparada.

<b>Espécie de <i>Azospirillum</i></b>	<b>Melhor Similaridade</b>	<b>% proteínas *</b>
<i>Azospirillum</i> sp. B510	1.033	18,79
<i>Azospirillum brasilense</i> 245	1.408	25,61
<i>Azospirillum lipoferum</i> 4B	1.060	19,28
<b>Total (&lt;e-10)</b>	<b>3.501</b>	<b>63,70</b>

\* valor calculado relativo às 5.496 sequências de proteínas de *A. amazonense*.

Estes resultados demonstram que o genoma de *A. amazonense* apresenta aproximadamente 63% de sequências compartilhadas com as três outras espécies e que todas contribuem com uma quantidade semelhante em termos do número de proteínas na composição do genoma de *A. amazonense*.

Com o objetivo de identificar as proteínas com função conhecida, conservadas em cada um destes genomas, (ver **Tabela Suplementar 1**) as 125 proteínas mais similares em *A. brasilense*, 94 em *A. lipoferum* 4B e 63 em *Azospirillum* sp. B510 foram analisadas. Estes produtos proteicos foram classificadas em diferentes vias metabólicas, utilizando o banco de dados do COG e KEGG, em bancos de domínios PFAM (227 proteínas identificadas), TIGR (187 Proteínas identificadas) e CDD (63 proteínas identificadas), além da identificação de proteínas de transporte (83 proteínas identificadas) e envolvidas com resistência a antibióticos (39 proteínas identificadas). Por serem proteínas conservadas nesses organismos, também foram analisadas no banco de proteínas essenciais (DEG) onde a maioria das sequências localizou ortólogos correspondentes (ZHANG E LIN, 2009).

Entre as sequências identificadas podemos observar proteínas ribossomais, proteínas relacionadas ao metabolismo do nitrogênio, com transcrição, tradução, transporte, flagelo e quimiotaxia. Como esperado, proteínas relacionadas com os ribossomos e RNA polimerase foram as mais frequentes, ressaltando proteínas relacionadas à fixação do nitrogênio e proteínas de relacionadas com quimiotaxia e com transporte.

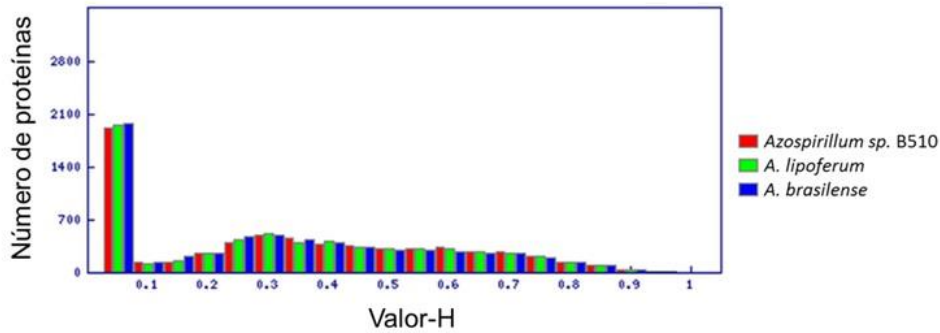
### 3.2.2. Identificação de proteínas diferenciais *in silico*.

Para a realização da identificação de proteínas diferenciais a metodologia de subtração *in silico* foi realizada utilizando o mGenomeSubtractor (SHAO *et al.*, 2010). A forma de comparar esses genomas é baseada no “valor de homologia” (Valor-H), que utiliza o valor de E para representar entre 0 e 1 os níveis de similaridade entre duas sequências de DNA ou proteína. Este valor foi calculado para os alinhamentos entre sequências proteicas das quatro espécies de *Azospirillum* e sua distribuição pode ser visualizada nos gráficos da **Figura 3.1**.

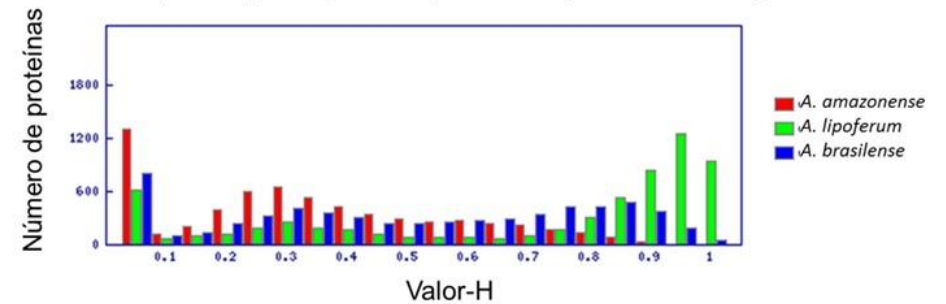
O gráfico de distribuição dos valores de similaridade entre essas espécies demonstra que *A. amazonense* é a espécie mais afastada geneticamente, considerando o grande número de proteínas com valores de homologia próximos a zero e o pequeno número de sequências com valores próximos a 1. A espécie *A. amazonense* possui 3.278 proteínas específicas quando o valor padrão H de 0,42 é utilizado e 1.250 proteínas com valor H igual a zero, que no caso dessa espécie, aparece como a grande maioria de proteínas diferenciais (ver **Figura 3.1A**).

Na análise das proteínas específicas de *A. amazonense* observamos que das 1.250 proteínas classificadas, 354 eram hipotéticas conservadas, além de 141 com baixa similaridade (entre  $e^{-10}$  e 0,005). Com a retirada dessas sequências e das proteínas hipotéticas de *A. amazonense* (~500 proteínas), 276 proteínas de *A. amazonense* que não possuem similaridade nas outras 3 espécies foram identificadas e podem ser visualizadas na **Tabela Suplementar 2**.

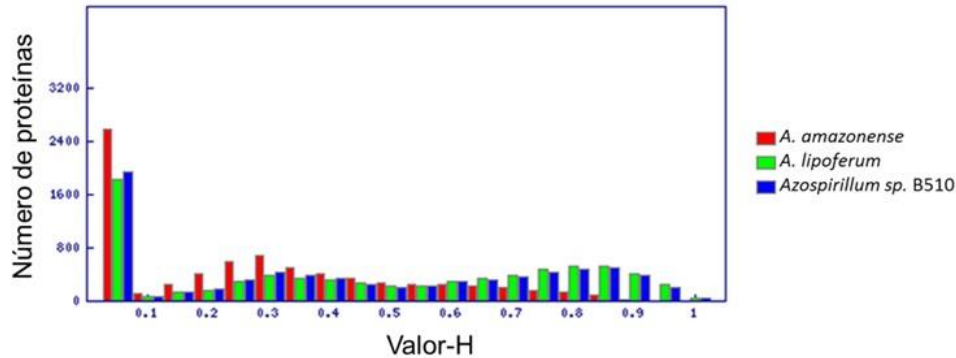
**A** Distribuição do valor de homologia para **5496** sequências de proteínas de *A. amazonense* Y2, com relação a três espécies do mesmo gênero.



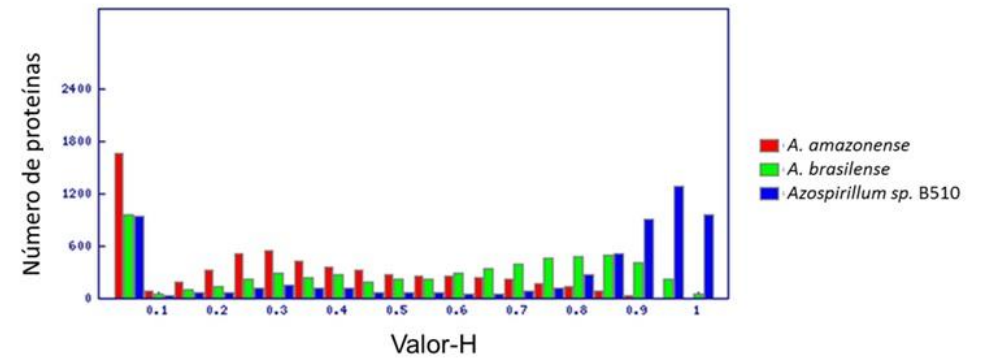
Distribuição do valor de homologia para **6309** sequências de proteínas de *Azospirillum* sp. B510, com relação a três espécies do mesmo gênero.



**C** Distribuição do valor de homologia para **7557** sequências de proteínas de *A. brasilense* 245, com relação a três espécies do mesmo gênero.



Distribuição do valor de homologia para **6639** sequências de proteínas de *A. lipoferum* 4B, com relação a três espécies do mesmo gênero.



**Figura 3.1** - Hibridização Subtrativa *in silico* entre quatro espécies do gênero *Azospirillum*.

A figura apresenta a distribuição dos níveis de similaridade entre estes genomas utilizando como referência. *A. amazonense* em **A**, *Azospirillum* sp. B510 em **B**, *A. brasilense* em **C** e *A. lipoferum* em **D**.

Segundo a relação de classificação no banco de dados COG, podemos perceber que a maioria das proteínas exclusivas de *A. amazonense* com função conhecida está relacionada com o metabolismo de carboidratos, aminoácidos, motilidade e secreção, conforme visualizado na **Tabela 3.5** abaixo.

**Tabela 3.5** - Número de proteínas específicas de *A. amazonense* com relação às outras três espécies do gênero, classificadas no banco de sequências do COG.

Via metabólica	Proteínas	Classe COG
Transporte e metabolismo de carboidratos	36	G
Função em geral apenas	29	R
Função desconhecida	28	S
Transporte e metabolismo de aminoácidos	16	E
Secreção e transporte	16	U
Motilidade	10	N
Parede celular, membrana e envelope	9	M
Modificações pós traducional, <i>turnover</i> e chaperona	9	O
Compostos secundários	9	Q
Transporte e metabolismo de nucleotídeos	7	F
Transporte e metabolismo de lipídeos	6	I
Transporte e metabolismo de íons inorgânicos	6	P
Produção e conversão de energia	5	C
Transporte e metabolismo de coenzimas	5	H
Transcrição	5	K
Replicação, recombinação e reparo	5	L
Mecanismos de transdução de sinal	2	T
Defesa	2	V
Tradução, estrutura ribossomal e biogênese	1	J

Analisando os resultados obtidos para *A. brasilense*, observamos que apesar do número de proteínas diferenciais ser alto (3.320 para valor H de 0,42 e 1.493 para H igual a 0), existe uma quantidade razoável de sequências com similaridade maior que 0,8 com as sequências das espécies e *Azospirillum* sp. B510 e *A. lipoferum* (ver **Figura 3.1 C**).

Considerando *Azospirillum* sp. B510 e *A. lipoferum*, podemos observar que são as espécies mais semelhantes entre si. *Azospirillum* sp. B510 possui menor número de proteínas com similaridade próxima a zero com relação a *A. amazonense*, porém é possível que exista em efeito causado pela predição das

proteínas terem sido realizadas com base na primeira (ver **Figura 3.1B** e **3.1D**). Devido à grande similaridade entre *A. lipoferum* e *Azospirillum* sp. B510, o número de sequências exclusivas dessas espécies é bastante diminuído, sendo igual a 1.422 e 297 para *Azospirillum* sp. B510 e 1.143 e 588 para *A. lipoferum* considerando valor H de 0,42 e 0 respectivamente.

A avaliação das proteínas conservadas em *Azospirillum* nesta análise, resultou em 632 sequências para um valor H de 0,64 (padrão) diminuindo para 126 quando a estringência foi aumentada para um valor H de 0,8. A **Tabela 3.6** apresenta a classificação dessas proteínas pelas vias metabólicas do banco COG, sendo que este resultado é correspondente e complementar ao identificado na análise local, onde o valor de E foi utilizado como referência (ver **Tabela Suplementar 1**). O valor H de 0,8 aqui utilizado, aumentou a estringência da análise, reduzindo pela metade as proteínas consideradas altamente conservadas identificadas na análise local, mas aumentando a confiabilidade do resultado.

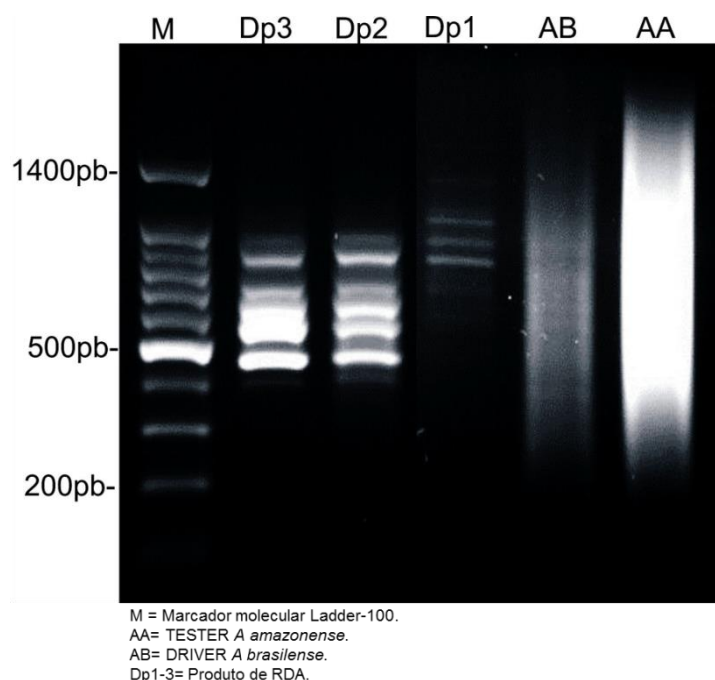
**Tabela 3.6** – Número de proteínas conservadas entre quatro espécies de *Azospirillum* utilizando o Valor H de 0,8 como padrão. A classificação por vias metabólicas foi realizada com base no banco de dados do COG.

Via metabólica	Proteínas	Classe COG
Tradução, estrutura ribossomal e biogênese	28	J
Produção e conversão de energia	17	C
Transporte e metabolismo de aminoácidos	15	E
Transcrição	13	K
Modificações pós traducional, <i>turnover</i> e chaperona	13	O
Mecanismos de transdução de sinal	12	T
Transporte e metabolismo de nucleotídeos	9	F
Transporte e metabolismo de lipídeos	8	I
Função em geral apenas	6	R
Motilidade	5	N
Transporte e metabolismo de coenzimas	4	H
Replicação, recombinação e reparo	3	L
Transporte e metabolismo de carboidratos	2	G
Controle do ciclo celular	1	D
Parede celular, membrana e envelope	1	M
Compostos secundários	1	Q
Função desconhecida	1	S
Secreção e transporte	1	U
Defesa	1	V

### 3.2.3. Identificação experimental de proteínas diferenciais

No presente experimento, uma análise de diferença representacional (RDA) foi realizada com o objetivo de isolar sequências genéticas exclusivas de *A. amazonense* com relação a *A. brasilense*. Devido ao aumento do número de espécies conhecidas neste gênero de promotores do crescimento vegetal, o entendimento das capacidades individuais de cada microrganismo são de fundamental importância para estabelecimento de novos inoculantes e compreensão da relação entre os organismos em questão e o ambiente em que vivem.

Após a purificação do DNA das duas espécies, fragmentação e ligação dos adaptadores ao genoma de *A. amazonense*, as amostras foram submetidas a 3 ciclos de hibridização subtrativa seguida de purificação por amplificação diferencial. A **Figura 3.2** apresenta a visualização do produto do RDA em eletroforese em gel de agarose.



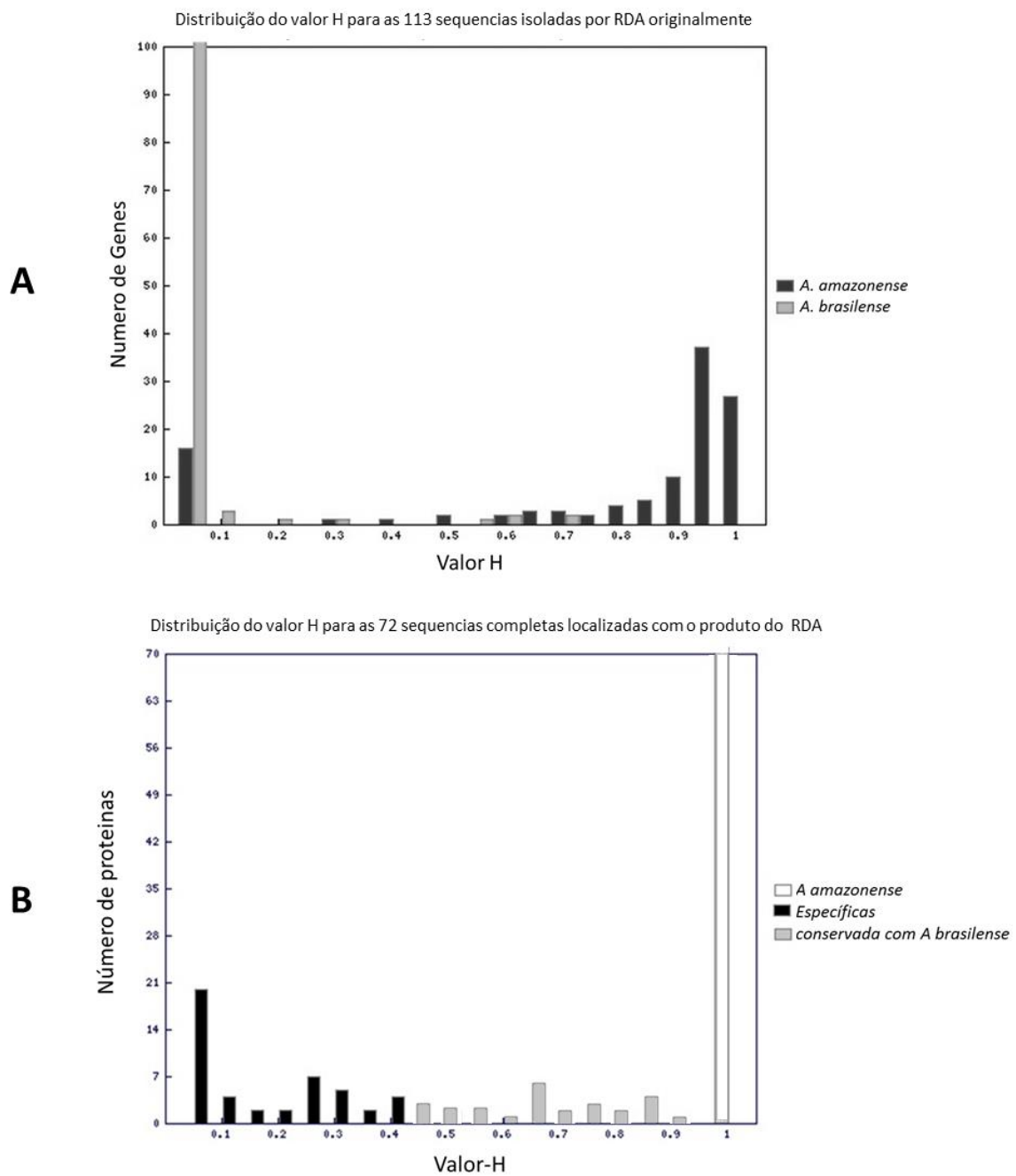
**Figura 3.2** – Hibridização subtrativa seguida de amplificação diferencial. Eletroforese em gel de agarose 1% mostrando o produto do RDA nos 3 ciclos de hibridização realizados. AA e AB representam respectivamente o DNA total de *A. amazonense* e *A. brasilense* antes da etapa de hibridização.

Os fragmentos de DNA resultantes deste experimento foram clonados em vetor pUC18/SmaI e submetidos ao sequenciamento automático, sendo um total de 480 mostras sequenciadas. Posteriormente, foram removidas as sequências de baixa qualidade e sequências do vetor (programa *Staden Package*) sendo finalmente realizada a junção dos contigs. Após esta análise, foram obtidos 113 sequências consenso.

Com a finalidade de avaliar a eficiência da metodologia duas análises foram realizadas. Primeiramente um teste de similaridade entre as 113 sequências isoladas na técnica de RDA com os genomas de *A. amazonense* (SANT'ANNA *et al.*, 2011) e de *A. brasilense* foi realizado através de um comparação de sequências utilizando o BlastX local (utilizando o programa *BioEdit 7.0*). A análise por similaridade do produto do RDA em *A. amazonense* demonstrou que das 113 amostras, 72 proteínas diferentes foram identificadas, eliminando desta maneira sequências codificantes da mesma proteína e sequências sem similaridade no genoma de *A. amazonense*.

Como segunda maneira de validar a eficiência da metodologia de RDA, as 113 sequências do RDA e as 72 proteínas completas resultantes da comparação de sequências, foram submetidas ao programa mGenomeSubtractor e os genomas de *A. amazonense* e *A. brasilense* foram utilizados como controle. A **Figura 3.3A** mostra a distribuição do valor de homologia (Valor H) das amostras para as 113 sequências originais e a **Figura 3.3B** demonstra os níveis de similaridade para as 72 proteínas localizadas no genoma de *A. amazonense* utilizando as 113 sequências iniciais como referência.





**Figura 3.3** – Caracterização das seqüências diferenciais de *A. amazonense*. Distribuição dos valores de similaridade das 113 seqüências isoladas originalmente no RDA (**A**) e das 72 proteínas localizadas por similaridade com o produto do RDA (**B**).

Análise dos dados da **Figura 3.3B** revela que 46 seqüências possuem valor H menor que 0,42 (padrão) em *A. brasilense*, correspondendo a seqüências exclusivas de *A. amazonense*. As 26 restantes são consideradas similares entre

as espécies. Este resultado pode ser relativo a presença de sequências consideradas contaminantes (falso-positivos) ou gerado pela presença de sequências muito curtas (menores que 100pb), já que o tamanho do fragmento é considerado no cálculo de similaridade. A **Tabela 3.7** apresenta a classificação (via COG) para as 46 sequências exclusivas de *A. amazonense* com relação a *A. brasilense* isoladas pela metodologia de RDA, enquanto análises individuais em diferentes bancos de dados podem ser vistas na **Tabela Suplementar 3**.

**Tabela 3.7** - Proteínas de *A. amazonense* identificadas por RDA e classificadas no banco de dados COG.

Via Metabólica	Proteínas	Classe
Transporte e metabolismo de íons inorgânicos	6	P
Função geral apenas	5	R
Produção e conversão de energia	3	C
Transporte e metabolismo de carboidratos	3	G
Replicação, recombinação e reparação	3	L
Função desconhecida	3	S
Transporte e metabolismo de aminoácidos	2	E
Transporte e metabolismo lipídico	2	I
Parede celular / membrana / biogênese do envelope	2	M
Biossíntese de metabólitos secundários	2	Q
Mecanismos de transdução de sinal	2	T
Transporte e metabolismo de nucleotídeos	1	F
Transcrição	1	K
Modificação pós-traducional, <i>turnover</i> de proteínas chaperonas	1	O
Mecanismos de defesa	1	V

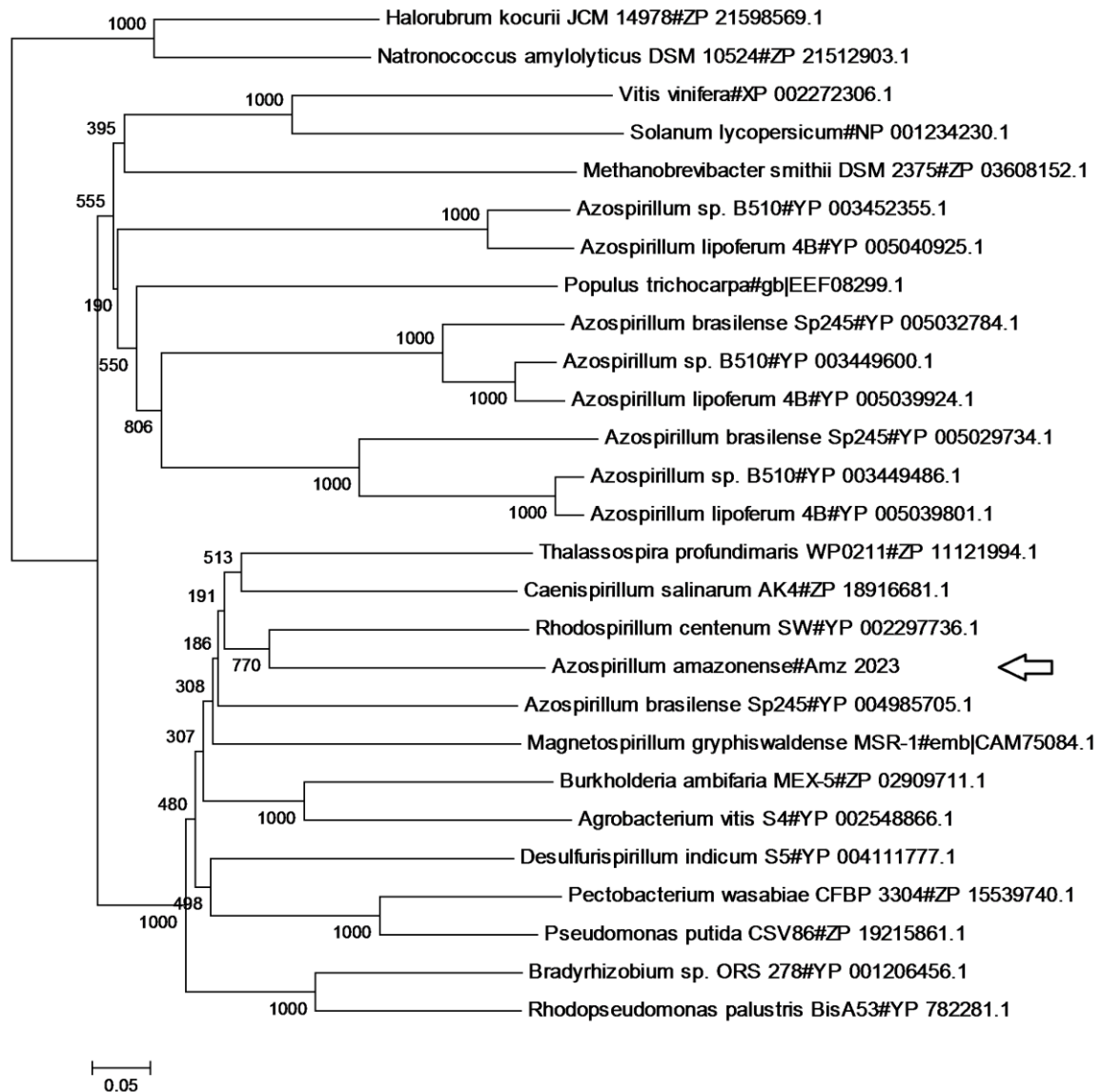
Nesta análise, das 46 proteínas exclusivas, 29 produtos proteicos foram classificados em um dos grupos com função conhecida (definidos pelo COG), sendo 17 produtos classificados como hipotéticos (9 proteínas não foram classificadas pelo COG, 5 foram classificadas como Função geral e 3 como Função desconhecida). As classes envolvidas com transporte de íons inorgânicos, produção e conversão de energia e transporte e metabolismo de aminoácidos foram aquelas que apresentaram o maior número de sequências isoladas por RDA. As proteínas com maior interesse serão discutidas na **seção 3.5**.

### **3.3. Caracterização das principais rotas metabólicas presentes no genoma de *Azospirillum amazonense*.**

#### 3.3.1. Proteínas relacionadas ao metabolismo de hormônios vegetais

A síntese de fitormônios em *Azospirillum* foi inicialmente observada em *Azospirillum brasilense*. Através de bioensaios, foi demonstrada a capacidade desta espécie sintetizar a auxina ácido indol acético a partir do triptofano, além de uma pequena quantidade de giberelinas e moléculas semelhantes a citocininas (TIEN *et al.*, 1979)

Com o objetivo de entender os mecanismos utilizados em *A. amazonense* para a produção de hormônios vegetais, uma proteína com similaridade a um transportador de auxina foi localizada em *A. amazonense* (Amz\_2023) pela similaridade com as outras espécies de *Azospirillum*. Essa proteína demonstrou também similaridade com o transportador de auxinas de *Arabidopsis* (Transportador polar de auxina, acesso AT1G70940). Uma análise filogenética foi realizada com sequências ortólogas ao transportador de auxina encontrado em *A. amazonense* presentes nos genomas das espécies do gênero *Azospirillum*, além de sequências com similaridade à proteína em plantas e outras bactérias. O resultado desta análise pode ser visualizado na **Figura 3.4**.



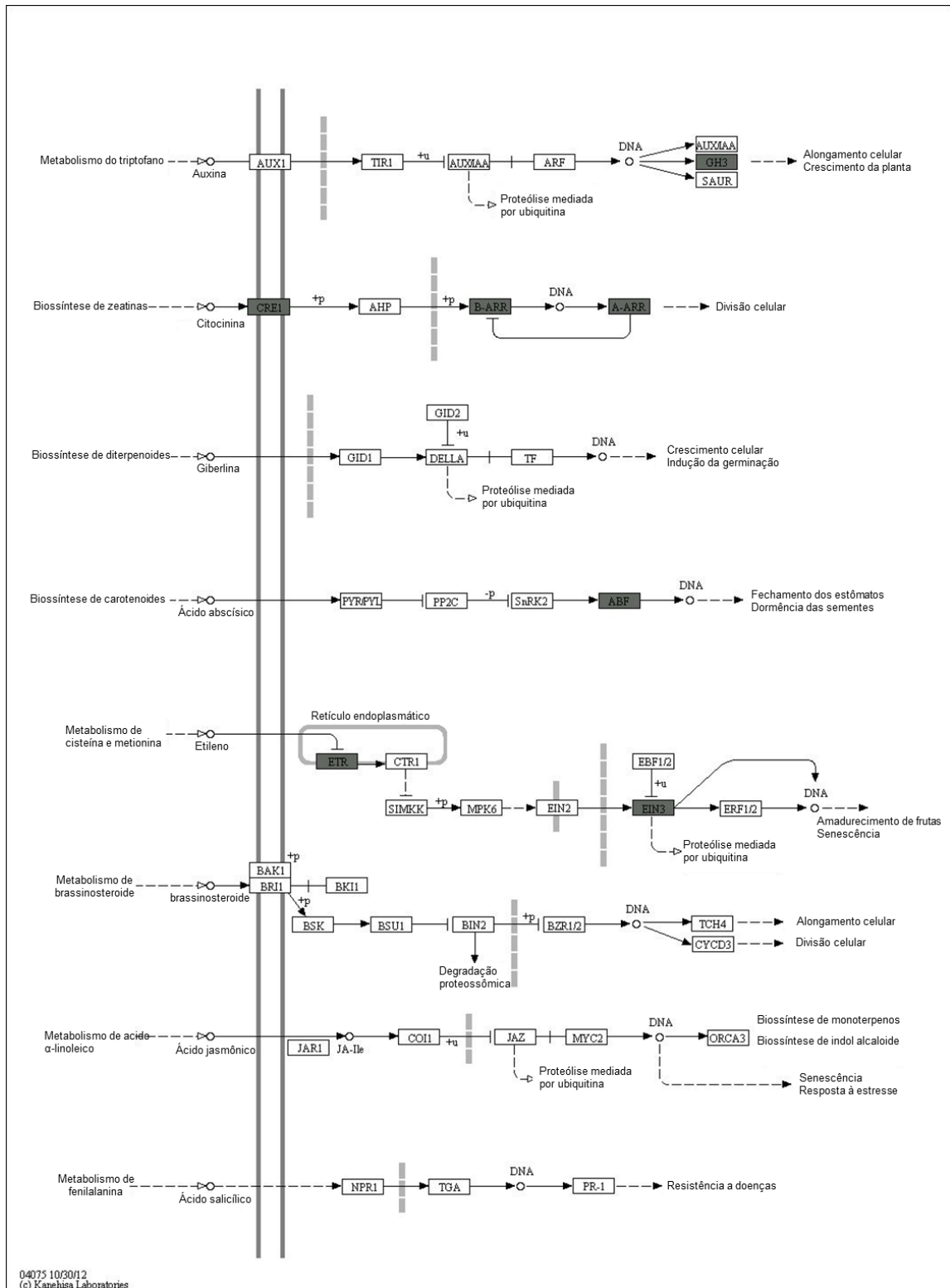
**Figura 3.4** - Filogenia de um transportador de auxina presente em *A. amazonense*.

Para esta análise foram obtidos ortólogos da proteína de *A. amazonense* presentes em outras espécies de *Azospirillum*, plantas e em outras bactérias.

Para a realização desta análise filogenética, proteínas similares de três espécies de Archaea foram selecionadas para grupo externo. Surpreendentemente, a proteína de uma destas espécies (do gênero *Methanobrevibacter*) ficou agrupada com duas proteínas vegetais, enquanto a seqüência do gênero vegetal *Populus*, aparece entre as espécies de *Azospirillum*. A proteína de *A. amazonense* ficou posicionada junto à maioria das amostras

bacterianas e próxima a uma das sequências de *A. brasilense* selecionadas. Um alinhamento desse grupo de proteínas transportadoras de auxina mostrou que a sequência do gênero *Populus* (grupo que contém maioria das sequências de *Azospirillum*) é mais similar ao componente 8 dos transportadores de auxinas de *Arabidopsis*, do que o grupo formado pelas proteínas da maioria das bactérias. Devido ao pouco conhecimento das vias de metabolização de auxinas em bactérias, análises adicionais foram realizadas objetivando localizar possíveis sequências relacionadas às rotas de fitormônios.

A primeira análise foi realizada submetendo os genomas de *Azospirillum* em estudo à anotação automática no banco de ortólogos do KEGG (KAAS), utilizando *A. thaliana* e *O. sativa* como referência, com a finalidade de identificar proteínas conservadas presentes nas vias de fitormônios. O resultado deste experimento revelou a presença de 7 proteínas ortólogas de *A. amazonense*, 9 de *A. lipoferum*, 8 de *Azospirillum* sp. B510 e 11 de *A. brasilense*. A **Figura 3.5** apresenta as proteínas de *A. amazonense* conservadas nessas vias enquanto sua descrição pode ser vista na **Tabela 3.8**. Informações adicionais sobre essas sequências, assim como a caracterização em diferentes bancos de dados podem ser obtidas na **Tabela Suplementar 4**. A identificação destas proteínas e sua presença nas quatro espécies de *Azospirillum* analisadas, são mostradas na **Figura Suplementar 1**.



**Figura 3.5** – Identificação das proteínas de *A. amazonense* envolvidas nas vias de sinalização de hormônios de plantas. A figura foi modificada segundo informações contidas nos mapas de ortólogos do KEGG. *A. thaliana* e *O. sativa* foram utilizadas como referências para localização das proteínas de *A. amazonense* (marcadas em cinza).

**Tabela 3.8** – Identificação das proteínas de *A. amazonense* envolvidas nas vias de sinalização de hormônios de plantas identificados no banco de ortólogos do KEGG apresentados na Figura 3.5.

ID	ID Figura 3.5	Produto proteico	Via Metabólica	Presença
Amz1202	GH3	outer membrane lipoprotein	auxin responsive GH3 gene family	Amz, B510, Abr, Ali
Amz0351	CRE1	two component hybrid sensor	arabidopsis histidin (cytokinin)	Amz, B510, Abr, Ali
Amz0062	B-ARR	chemotaxis protein CheY	two-component response regulator	Amz, B510, Abr, Ali
Amz0140	A-ARR	response regulator receiver	two-component response regulator	Amz, B510, Abr, Ali
Amz1316	ABF	membrane transport protein	responsive element binding factor	Amz
Amz3028	ETR	nonmotile and phageresistance	ethylene receptor [EC:2.7.13.-]	Amz, Abr, Ali
Amz4403	EIN3	cation efflux protein	ethylene-insensitive protein 3	Amz, Abr

Podemos destacar também a presença de uma proteína relacionada ao metabolismo de brassinosteroides em *A. brasilense* (acesso YP\_005031400) similar ao ortólogo K14500, do banco de dados do KEGG. Resultados relacionados a esse hormônio são pouco conhecidos em microrganismos, necessitando portanto maiores estudos para definição da função real desta sequência.

A segunda análise realizada com objetivo de elucidar o metabolismo de fitormônios em *Azospirillum* também foi baseada na similaridade com proteínas das vias de síntese em vegetais, quando o genoma de *A. amazonense* foi submetido a uma comparação com as vias hormonais presentes em *Arabidopsis* (banco de dados AHD). Esse banco de sequências curadas possui análises variadas para o estabelecimento da relação de uma proteína com uma via metabólica relacionada aos fitormônios, sendo que cada ortólogo apresenta vínculos relacionados com análises de mutantes, e similaridade de sequência (PENG *et al.*, 2009). As proteínas de *A. amazonense* com melhor similaridade no banco de sequências AHD, assim como a assinatura para os principais bancos de ortólogos podem ser visualizados na **Tabela Suplementar 5**, enquanto sua relação no metabolismo desse procarionto pode ser vista na **Tabela Suplementar 6**, classificadas conforme o banco de ortólogos do COG.

Os resultados dessa análise demonstram que as proteínas de *A. amazonense* com similaridade em proteínas de *Arabidopsis* envolvidas no

metabolismo de hormônios estão relacionadas principalmente com as vias energéticas, além do metabolismo de lipídeos e transdução de sinal.

Para finalizar o estudo de proteínas de *A. amazonense* relacionadas ao metabolismo de fitormônios, um banco com as sequências das proteínas que demonstraram ser diferencialmente expressas na presença de auxina, em um transcriptoma realizado com *A. brasilense* (VAN PUYVELDE *et al.*, 2011), foi analisado. Como essa publicação foi realizada antes da finalização do genoma de *A. brasilense*, realizamos uma localização das proteínas no DRAFT do genoma de *A. brasilense* para fazer a correspondência com a atual versão. Entre as 160 proteínas apresentadas como diferencialmente expressas na presença de auxina em *A. brasilense*, 149 foram localizadas no genoma completo de *A. brasilense*, 107 proteínas foram localizadas em *A. amazonense*, 136 em *Azospirillum* sp. B510 e 136 em *A. lipoferum*. A classificação realizada no banco de ortólogos COG, com as 107 proteínas de *A. amazonense* que apresentaram similaridade com as proteínas diferencialmente expressas na presença de auxina em *A. brasilense* está apresentada na **Tabela Suplementar 7**.

As proteínas de *A. amazonense* com similaridade às proteínas diferencialmente expressas na presença de auxina em *A. brasilense* estão relacionadas principalmente com a tradução, estrutura ribossomal e biogênese, produção e conversão de energia, transporte e metabolismo de aminoácidos, transporte e metabolismo de carboidratos e modificação pós-traducional, *turnover* de proteínas e chaperonas.

Com o objetivo de integrar as análises de proteínas de *A. amazonense* possivelmente envolvidas no metabolismo de hormônios vegetais realizadas, um cruzamento entre os resultados apresentados acima foi realizado. A **Tabela 3.9** mostra a identificação das proteínas de *A. amazonense* que apresentaram alta similaridade com proteínas reguladas por auxina em *A. brasilense* e simultaneamente similaridade com proteínas de *A. thaliana* relacionadas ao metabolismo de hormônios, identificadas pelo banco de sequências AHD (ver **Tabela Suplementar 8**).



**Tabela 3.9** - Proteínas de *A. amazonense* similares a proteínas de *A. brasilense* reguladas por auxina e envolvidas nas rotas de fitormônios em *A. thaliana*.

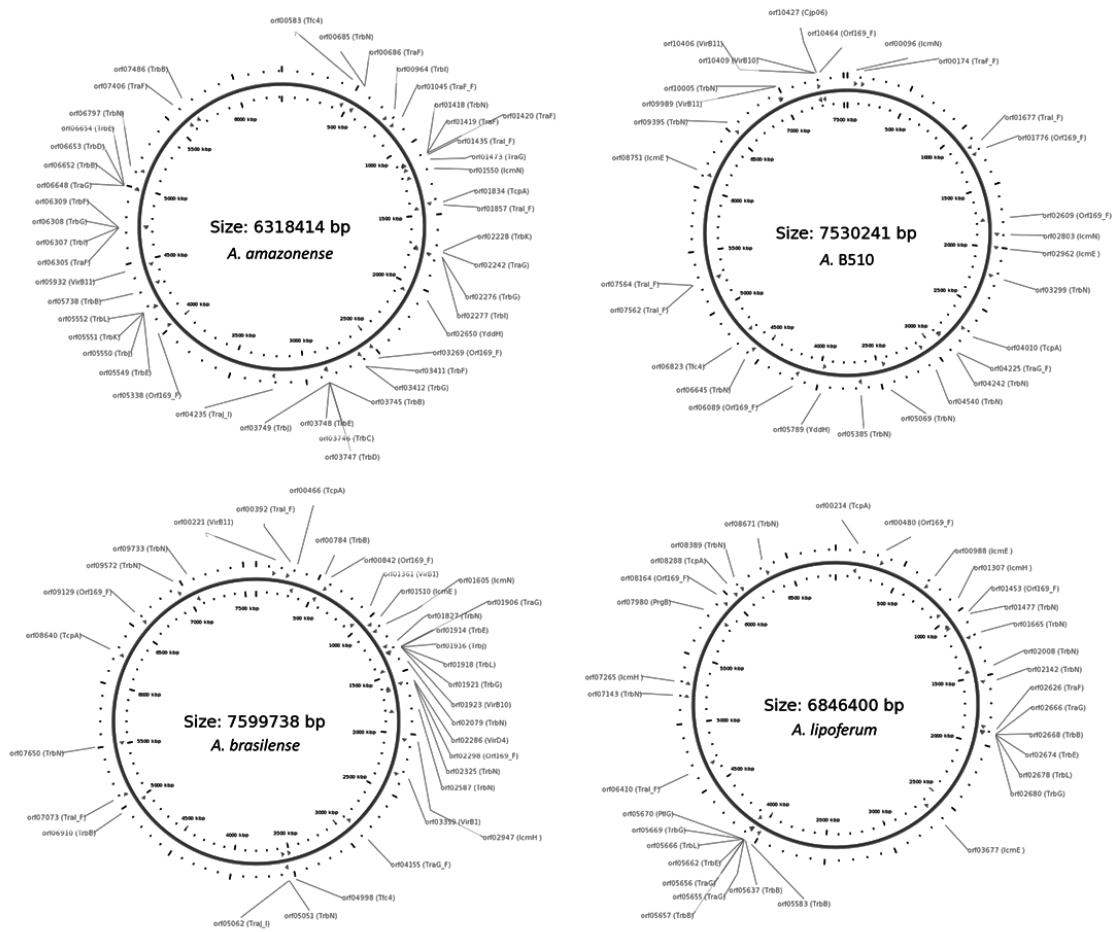
ID proteína	<i>A. amazonense</i>	Produto proteico	COG	Mapa KEGG	AHD	<i>A. brasilense</i>	CDD	TIGR
Amz_0174	ZP_08867717.1	malonylCoA synthase	COG0318	ko00627	AT4G05160	YP_005032892.1		
Amz_0427		conserved hypothetical protein RC12493	COG4530		AT3G56860	YP_005030396.1		TIGR02300
Amz_0877	ZP_08868186.1	translation elongation factor P efp	COG0231		AT5G35750	YP_005031990.1		TIGR00038
Amz_1684		nucleoside diphosphate kinase _cl00335	COG0105	ko00230	AT5G63310	YP_004984755.1	cl00335	
Amz_2216		shortchain dehydrogenasereductase SDR	COG1028		AT1G52340	YP_005030737.1	cl09931	
Amz_3022		Aspartate aminotransferase A	COG0436	ko00250	AT2G24850	YP_005032548.1		
Amz_3405	ZP_08869669.1	branched-chain amino acid transp _ph3R	COG0411	ko02010	AT3G28860	YP_005031567.1	cl09099	
Amz_3660	ZP_08869819.1	aerobic-type carbon monoxide dehydro	COG2080	ko00230	AT2G27150	YP_004986390.1		
Amz_3694	ZP_08869422.1	3-hydroxyacyl-CoA dehydrogenase type II	COG1028	ko00280	AT1G52340	YP_005033753.1	cl09931	
Amz_3738	ZP_08869865.1	xanthine dehydrogenase large subunit	COG4631	ko00230	AT5G20960	YP_005031474.1		TIGR02965
Amz_3833		highaffinity branchedchain amino acid transp	COG0411	ko02010	AT3G28860	YP_005031519.1		
Amz_4592	ZP_08870399.1	aldehyde dehydrogenase_Ath135	COG1012	ko00010	AT1G54100	YP_004986238.1	cl11961	

### 3.3.2. Proteínas relacionadas ao transporte em *Azospirillum*

#### 3.3.2.1. Sistema de Secreção do Tipo IV

A análise da **Tabela 3.2** revelou que entre as rotas metabólicas que apresentaram maior número de proteínas em *A. amazonense* com relação aos outros membros do gênero, foi o sistema de secreção. Uma análise mais detalhada foi realizada utilizando a ferramenta SecRet4, capaz de detectar e anotar genes relacionados ao sistema de secreção do tipo IV (Bl *et al.*, 2013). Para este experimento, um arquivo contendo as sequências concatenadas das espécies foi criado. Este teste resultou em 60 proteínas codificadas no genoma de *A. amazonense*, 35 de *Azospirillum* sp. B510, 33 de *A. lipoferum* e 29 de *A. brasilense*. Entre as 60 identificadas, 18 proteínas foram caracterizadas como exclusivas de *A. amazonense* pelo mGenomeSubtractor já que possuem valor H próximo a zero em ambos os genomas (ver **Figura Suplementar 2**). A distribuição das proteínas do sistema de secreção tipo IV nos quatro genomas de *Azospirillum* pode ser visto na **Figura 3.6** enquanto a caracterização individual das proteínas exclusivas pode ser vista na **Tabela 3.10**.

Como podemos observar (ver **Tabela 3.10**), a maioria das proteínas encontradas principalmente em *A. amazonense* pertencem à família Trb (*conjugal transfer protein*) estando provavelmente relacionadas aos mecanismos de transferência de material genético através de conjugação. A organização dos genes que codificam estas proteínas no genoma de *A. amazonense* demonstra que 5 dos genes estão agrupados (proteínas de Amz\_1457 até Amz\_1461), sendo que os demais estão em *contigs* separados. Estudos em *Agrobacterium* relacionam esse grupo de proteínas com um comportamento de doadora ou receptora de material genético no ambiente da rizosfera (HOLGUIN & GLICK, 2003).



**Figura 3.6** – Proteínas relacionadas ao Sistema de Secreção tipo IV em quatro membros do gênero *Azospirillum*. As proteínas identificadas pelo SecRet4 estão posicionadas de acordo com um concatenado das sequências de nucleotídeos anotados pela ferramenta.

**Tabela 3.10** - Proteínas do Sistema de Secreção Tipo IV exclusivas de *A. amazonense* em relação ao gênero *Azospirillum*.

ID	Acesso	Produto proteico	COG	Classe	Mapa	CDD	PFAM	KEGG	Virulência	Melhor Similaridade
Amz_1460		conjugal_transfer_protein	COG3838	<b>U</b>	ko03070			K03197	VFG0923	NP_774927.1
Amz_2466		conjugal_transfer_protein_Trbc	COG3838	<b>U</b>	ko03070			K03197	VFG0187	YP_495599.1
Amz_2242	ZP_08868549.1	conjugal_transfer_protein_TrbcF	COG3701	<b>U</b>	ko03070	cl01500	PF04335	K03200		YP_004282663.1
Amz_4148	ZP_08870114.1	conjugal_transfer_protein_TrbcF_cl01500	COG3701	<b>U</b>	ko03070	cl01500	PF04335	K03200	VFG0183	ZP_09295541.1
Amz_1491	ZP_08868549.1	conjugal_transfer_protein_TrbcF_K03200_cl01500	COG3701	<b>U</b>	ko03070	cl01500	PF04335	K03200		NP_774921.1
Amz_1490		Conserved_066_conjugal_transfer_protein_TrbcL								ZP_07026205.1
Amz_4145		conserved_hypothetical_protein	COG5639	<b>S</b>		cl02406			VFG1251	ZP_09877945.1
Amz_4466		Conserved_Hypothetical_protein_A458_04680								YP_006456609.1
Amz_4364	ZP_08870246.1	Conserved_Hypothetical_protein_ACMV_26040							VFG0261	YP_004284833.1
Amz_0934		Conserved_hypothetical_protein_Gdia_2803								YP_002277150.1
Amz_1465	ZP_08870246.1	Conserved_Hypothetical_protein_Gdia_2819								YP_002277166.1
Amz_4144	ZP_08870110.1	Conserved_Hypothetical_protein_SpKC8_03690							VFG0581	ZP_09138542.1
Amz_3756	ZP_08869876.1	DNA_replication_protein	COG1484	<b>L</b>			PF01695		VFG1729	ZP_10178502.1
Amz_0687		glutaminase_cl00907	COG2066	<b>E</b>	ko00250	cl00907	PF04960	K01425	VFG1413	YP_003545603.1
Amz_0438	ZP_08867912.1	peptidase_S26_conserved_region	COG4959	<b>OU</b>				K12062	VFG0902	ZP_11119033.1
Amz_0935	ZP_08867912.1	Peptidase_S26_conserved_region	COG4959	<b>OU</b>				K12062		ZP_07026220.1
Amz_1463		short-chain_dehydrogenase/reductase_SDR						K00540	VFG1355	ZP_16033639.1
Amz_1459		type_IV_secretory_pathway_VirB3_family_pr	COG5268	<b>NU</b>	ko03070			K03198		YP_002277173.1

Para esta análise foi utilizado o banco de sequências curadas do SecRet4.

### 3.3.2.2. Transportadores identificados no banco de ortólogos do KEGG

Com o objetivo de aprofundar a análise das proteínas relacionadas ao sistema de secreção/transporte em *A. amazonense*, foi realizada uma análise comparativa utilizando *Azospirillum* sp. B510, *R. centenum*, *B. japonicum* e *C. segnis* como referência.

**Tabela 3.11** – Grupo de Ortólogos do KEGG relacionados ao Sistema de Secreção Tipo IV localizados em *A. amazonense*.

<b>Descrição</b>	<b>Rota KEGG</b>
Via geral de secreção da proteína D	ko:K02453 gspD
Via geral de secreção da proteína E	ko:K02454 gspE
Via geral de secreção da proteína G	ko:K02456 gspG
Via geral de secreção da proteína K	ko:K02460 gspK
Pré-proteína translocase subunidade SecB	ko:K03071 secB
Pré-proteína translocase subunidade SecE	ko:K03073 secE
Pré-proteína translocase subunidade SecY	ko:K03076 secY
Sítio de reconhecimento, subunidade SRP54	ko:K03106 SRP54
Receptor de reconhecimento de fusão	ko:K03110 ftsY
Proteína translocase Tata, sec independente	ko:K03116 tatA
Proteína do sistema de secreção do tipo IV virB10	ko:K03195 virB10
Proteína do sistema de secreção do tipo IV VirB11	ko:K03196 virB11
Proteína do sistema de secreção do tipo IV VirB2	ko:K03197 virB2
Proteína do sistema de secreção do tipo IV VirB3	ko:K03198 virB3
Proteína do sistema de secreção do tipo IV VirB4	ko:K03199 virB4
Proteína do sistema de secreção do tipo IV VirB5	ko:K03200 virB5
Proteína do sistema de secreção do tipo IV VirB9	ko:K03204 virB9
Proteína do sistema de secreção do tipo IV VirD4	ko:K03205 virD4
Pré-proteína translocase subunidade YajC	ko:K03210 yajC
Pré-proteína translocase subunidade YidC	ko:K03217 yidC
Proteína do sistema de secreção do tipo IV Hcp	ko:K11903 K11903, hcp
Proteína de fusão SECD / SecF	ko:K12257 secDF
Proteína canal de membrana externa	ko:K12340 toIC

Este estudo revelou a presença de 23 vias de transportadores em *A. amazonense* (ver **Tabela 3.11**). A análise de sequências presentes nestas rotas identificou 117 proteínas envolvidas com o sistema de secreção no genoma de *A. amazonense*.

### 3.3.2.3. Transportadores identificados no banco TCDB

Com o objetivo de analisar proteínas envolvidas em outras rotas de transporte, foi realizada uma busca no banco de dados de proteínas de transporte TCDB (SAIER et al, 2006). O número de proteínas associadas e a contribuição relativa em cada genoma podem ser vistos na **Tabela 3.12**.

**Tabela 3.12** – Proteínas com similaridade no TCDB por espécie de *Azospirillum* e a contribuição relativa no proteoma de cada organismo.

<b>Espécie</b>	<b>Proteínas com similar no TCDB</b>	<b>% Proteoma</b>
<i>A. amazonense</i>	1.020	18
<i>Azospirillum</i> sp. B510	1.591	25
<i>A. brasilense</i>	1.646	21
<i>A. lipoferum</i>	1.433	23

Estes resultados indicam que *A. amazonense*, apesar de possuir proteínas exclusivas na via de secreção do tipo IV, possui um menor número relativo em se tratando de proteínas transportadoras em geral.

Uma busca no COG foi realizada para classificar as 1.020 proteínas relacionadas com transporte e secreção em *A. amazonense* **Tabela 3.13**.

A análise da **Tabela 3.13** demonstra que os transportadores identificados em *A. amazonense* estão envolvidos principalmente em vias do metabolismo de íons inorgânicos, sinalização, transporte de carboidratos e produção de energia. Podemos observar que 77 proteínas pertencem aos mecanismos de secreção tipo IV analisado anteriormente. Aproximadamente metade desse grupo (34 proteínas) possui funções mistas, envolvidas com motilidade e transporte através da

membrana (classes UM e NU). As proteínas com similaridade no TCDB e envolvidas com secreção e transporte através da membrana são mostrados na **Tabela Suplementar 9**. Podemos ressaltar que a maioria dessas proteínas está associada ao mapa que apresenta as proteínas dos sistemas de secreção (Ko03070 do KEGG).

**Tabela 3.13** – Transportadores de *A. amazonense* identificados no TCDB, classificados no banco de dados COG.

<b>Produto proteico</b>	<b>Proteínas</b>	<b>Classe</b>
Transporte e metabolismo de íons inorgânicos	187	P
Mecanismos de transdução de sinal	124	T
Transporte e metabolismo de carboidratos	82	G
Tráfego intracelular, secreção e transporte vesicular	77	U
Produção e conversão de energia	75	C
Parede celular / membrana / biogênese do envelope	69	M
Mecanismos de defesa	58	V
Transporte e metabolismo de aminoácidos	55	E
Função geral apenas	54	R
Motilidade celular	44	N
Modificação pós-traducional, <i>turnover</i> de proteínas, chaperonas	30	O
Função desconhecida	29	S
Biossíntese de metabólitos secundários, transporte e catabolismo	25	Q
Transporte e metabolismo lipídico	23	I
Transporte e metabolismo de Coenzimas	22	H
Transcrição	15	K
Controle do ciclo celular, divisão celular, particionamento do cromossomo	10	D
Tradução, estrutura ribossomal e biogênese	6	J
Replicação, recombinação e reparação	5	L
Transporte e metabolismo de nucleotídeos	1	F

### **3.4. Herança viral e elementos genéticos móveis**

Com o objetivo de avaliar a herança genética viral que atua nas espécies de *Azospirillum* em estudo, uma análise das regiões com origem viral foi caracterizada para os quatro genomas de *Azospirillum* (*A. amazonense* Y2, *Azospirillum* sp. B510, *A. brasilense* 245 e *A. lipoferum* 4B), através da ferramenta PHAST (ZHOU *et al.*, 2011). Os resultados indicaram a presença de 4 regiões

com origem viral no genoma de *A. amazonense* e de *A. lipoferum*, 5 no de *A. brasilense* e 12 no de *Azospirillum* sp. B510, todas, porém, interrompidas por genes do hospedeiro. Várias proteínas, identificadas nestas regiões, apresentaram similaridade com produtos proteicos codificados no DNA do bacteriófago CD, de *A. brasilense* CD (BOYER *et al.*, 2008), sendo 17 destas proteínas de *Azospirillum* sp. B510, 6 de *A. brasilense* e 5 de *A. lipoferum*.

A **Tabela Suplementar 10** apresenta as quatro regiões genômicas detalhadas pelo banco de dados de sequências virais PHAST para *A. amazonense*. A visão gráfica para essas regiões está apresentada na **Figura Suplementar 4**. As regiões genômicas relacionadas a fagos para as outras três espécies de *Azospirillum*, podem ser vistas na **Figura Suplementar 3** e algumas características como tamanho da região e fago mais similar identificado para cada região na **Tabela Suplementar 11**.

#### 3.4.1. Regiões genômicas relacionadas a fagos em quatro espécies de *Azospirillum*

Para esta análise, o cromossomo e os plasmídeos de *Azospirillum* sp. B510, *A. brasilense* e *A. lipoferum* foram posicionados do maior para o menor sendo gerada uma sequência concatenada única para cada espécie. Esse procedimento foi realizado pelo próprio PHAST, de maneira que esse artefato deve ser analisado durante a busca pelas regiões correspondentes nos genomas. Como o programa PHAST faz uma anotação própria para realizar os testes de similaridade, as proteínas resultantes podem possuir diferentes tamanhos com relação à anotação original. Essa anotação apresenta, porém, a posição relativa de cada proteína no concatenado, de maneira que é possível saber qual proteína foi anotada naquela posição originalmente. As proteínas de *A. amazonense* localizadas nas regiões virais identificadas pelo PHAST, bem como detalhes de sobre a identidade das sequências pode ser visualizado na **Tabela Suplementar 12**.



### 3.4.2. Classificação de Elementos Genéticos Móveis

Com o objetivo de analisar proteínas relacionadas a vírus e outros elementos genéticos móveis que não estejam agrupados, uma análise detalhada foi realizada utilizando simultaneamente proteínas codificadas por plasmídeos, vírus e prófagos e posteriormente, vírus e prófagos apenas no banco de sequências do ACLAME (LEPLAE *et al.*, 2004) (ver **Tabela 3.14**).

**Tabela 3.14** - Herança plasmidial e viral encontrada em quatro espécies de *Azospirillum*.

Espécie	Total de Proteínas	Plasmídeos	%Proteínas	Prófagos	Vírus	Prófagos *	Vírus *	%Proteínas *
Amz	5496	3010	54,8	223	29	915	93	18,34
AB510	6309	4297	68,1	318	36	1711	115	29
Abra	7557	4069	53,8	302	47	1321	142	19
Alipo	6094	3639	59,7	289	33	1252	112	22

A comparação utilizou como base o banco de dados da Classificação de Elementos Genéticos Móveis (ACLAME). Amz, *A. amazonense*; AB510, *Azospirillum* sp. B510; Abra, *A. brasilense*; Alipo, *A. lipoferum*.

\* Herança plasmidial não considerada.

Os resultados da **Tabela 3.14** indicam que mais da metade das proteínas destas espécies possuem origem plasmidial, possivelmente herdadas de maneira horizontal. Chama a atenção a grande quantidade de proteínas com similaridade neste banco de dados encontradas em *Azospirillum* sp. B510, porém, como muitos genes de plasmídeos (que no gênero *Azospirillum* são grande parte dos genomas) também são herdados de sistemas virais, a mesma análise foi realizada somente para vírus e prófagos e o resultado esta presente também na **Tabela 3.14**.

Os dados resultantes desta última análise estão de acordo com os dados anteriores, onde um maior número de agrupamentos (*clusters*) virais foi encontrado para *Azospirillum* sp. B510 (no PHAST) (ver **Figura Suplementar 3**).

### 3.4.3. Sistema imune bacteriano CRISPR e proteínas associadas

Devido ao grande número de proteínas com similaridade a produtos proteicos codificados nos DNAs de prófagos e vírus encontrados nas espécies de *Azospirillum*, uma análise da presença de sequências identificadas como CRISPRS (do inglês *clustered regularly interspaced short palindromic repeats*) e proteínas associadas à defesa destes organismos contra entrada de material genético exógeno foi realizada.

Os CRISPRs são uma família de sequências de DNA caracterizadas pela alternância entre sequências repetidas, também chamadas DR, que são Interespaçadas por regiões curtas e variáveis de herança geralmente viral. Através de um mecanismo análogo ao que ocorre com o RNA de interferência, essas regiões repetidas formam uma estrutura secundária através de pareamento de maneira a expor as regiões espaçadoras. Estas regiões espaçadoras hibridizam com moléculas exógenas e ativam uma série de proteínas chamadas CAS (associadas aos CRISPRs) que possuem função de neutralizar moléculas de DNA/RNA invasoras. Entre as proteínas associadas destacam-se chaperonas e endonucleases. Uma série de proteínas CAS foi identificada em diferentes organismos, sendo separadas em diferentes famílias como revisado por (MOKROUSOV *et al.*, 2007).

Para as espécies de *Azospirillum*, os resultados indicaram 7 regiões CRISPR no genoma de *Azospirillum* sp. B510, 5 no de *A. amazonense* e 3 no de *A. lipoferum*, enquanto que esse sistema não foi encontrado em *A. brasilense*. As cinco regiões CRISPR identificadas em *A. amazonense* (pela ferramenta CRISPR-finder) podem ser visualizadas na **Tabela 3.15**.

Proteínas CAS (CRISPR associadas) foram identificadas nas quatro espécies sendo presentes em maior quantidade em *A. amazonense*, que possui dois agrupamentos gênicos codificantes destas proteínas. O primeiro identificado entre as proteínas Amz\_3760 e Amz\_3765 e o segundo entre Amz\_5447 e Amz\_5452 totalizando 12 proteínas associadas (ver **Tabela Suplementar 13**). Essas 12 sequências foram alinhadas com os genomas de *Azospirillum* sp. B510,

*A. brasilense* e *A. lipoferum* e o resultado desta análise pode ser visto na **Tabela 3.16**.

**Tabela 3.15 – Sistema imune bacteriano CRISPR.**

CRISPR ID	Região repetitiva (DR)	Espaçadores
CRISPR1	GGCGGTGGTACGGGGGCTGCCAGCACGCCGCCG	ACCTGCACGCCGATGCCCGTCTGGTGGTATGCCGATGCTGCCGCT
CRISPR1	GGTGGTGTGACCAGGGTGCCCAAGGTGCCGGCC	ACCAGGACGCCGGTCTGCTGCCCCGACGCGC
CRISPR1	GCCGCTGTGACCAGGCTGCCAGGCTGCCGCCG	GACGCGATGTAGACGGCGGTGTTGACCCC
CRISPR1	GCCGATGGTCCCGCTGTTGACCAGGCTGCCAGG	
CRISPR2	GGTGCTATCCCCGCGCCTGCGGGGGAAC	GCGCCCGTACCGTCTGCTGGCCAACTGGTA
CRISPR2	GGGTGTATCCCCGACCTGCGGGGGAACC	TATTCGGCAGTTCCACCTCCTCGTCCGACAC
CRISPR2	GGGTGTATCCCCGCGCCTGCGGGGGAACC	GTCGCTTCTCCAGGCTTCCACTCCTGGCG
CRISPR2	GGGTGTATCCCCGCGCTTGC GG GGAACC	ACGATGGCCGACGGTCCGACGCTTGTGGC
CRISPR2	GGGTGTATCCCCGCGCCTGCGGGGAAACC	CGTTCCTCAAGCCGGCACAGGATGGCGTGGAC
CRISPR2	GGGTGTATCCCCGCGCCTGCGGGGAAACC	
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	AGCACCGCCATGCGCCGTCGCTTTGGCGAAGAAG
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	AGGTCGAGCACCGTGCCACCACCAACCTGGCGC
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	TGATAAGTTAGGCTGAATTCGCCCGTCCAGGTAGTT
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	CTGCTTCCCGGTATCCGCAAGATCAAGGGCGACTA
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	GACATCACGCCGCTGCGGACCACGACCAACAACGC
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	ATCGACAACGCCCCACCCATGGCCGTGCGCCAGA
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	GACCAAGCAGGTGCGCGGCGAGGAAGGCATCAAG
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGAAAC	GCCTACGAGTGC GCGGAATGACCTTGACACCAG
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCAGATTGAAAC	GCCGTCACCGTCAAGTTGGACCCGCCCTTGGC
CRISPR3	GCCGCCCTCCGCGAGGGCGCGGATTGAAAC	ACGAGCTGGTGGGAAAGATGTGCTGCTTGACCCG
CRISPR3	GCCGCCCTCCGTGCGAGGGCGCGGATTGATAG	
CRISPR4	GGGTCTATCCCCACAGATGTGGGGAGACCG	CGGGTAATAGCGCTATGGCGGTGTCAGGCAT
CRISPR4	GAGTCTATCCCCGACAGGTGCGGGGGCACCG	TCCGCGCAATCCGTTGATTCGCCGATGCC
CRISPR4	GGGTCTATCCTCGCAGGTGCGGGGGCACCG	AGTCAGGAAACCCTGGCTGAGCTTCTCGGC
CRISPR4	TGGGTCTACTCCGACAGGTGCGGAGGCACCA	TCTTTGCCACAAAGGCGGATAGCCGGCCGT
CRISPR4	GGGTCTATCCCCGACAGGTGCGGGGGCACCG	ACACCACAAATCCTTGCGCCGCGGATGTC
CRISPR4	GGGTCTATCCCCGACAGGTGCGGGGGCACCG	GGGGTGCACTGGTCAATGCGGGCGATGAAGG
CRISPR4	GGGTCTATCCCCGACAGGTGCGGGGGCACCA	
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	TGCGCGAGGTGCGCGCGGACAGGTGGCGGACT
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	CCTGGATAGCCTGAACGAGGAACAAGCGTCCGGC
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	CATCAACGATCAGCTGGGTAATGACCGTGTG
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	GTATAGGCCGTGTTGCCCTTCTGCCAAACAC
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	AGCTACGTCGATAACAGCCAGAACAATCTGTC
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	GCCTACTCGGCTGAGTATTTCCATGGCGCTGG
CRISPR5	GGGTGTATCCCCGCGCCTGCGGGGGAACC	

Cinco regiões CRISPRs de *A. amazonense* apresentando as seqüências consenso (DR) e espaçadores encontrados com CRISPR-finder.

**Tabela 3.16** – Proteínas Cas de *A. amazonense* que possuem correspondente em outras espécies do gênero.

ID	Produto em <i>A. amazonense</i>	Espécie	Valor E	Acesso <i>Azospirillum</i>
Amz_3764	CRISPR-associated protein Cas1	AB510	3,00E-50	YP_003449926.1
Amz_3765	CRISPR-associated protein Cas2	AB510	1,00E-08	YP_003448223.1
Amz_5447	CRISPR-associated_helicase_Cas3	AB510	1,00E-06	YP_003449922.1
Amz_3764	CRISPR-associated protein Cas1	Alipo	3,00E-43	YP_005039484.1
Amz_3765	CRISPR-associated protein Cas2	Alipo	4,00E-08	YP_005039485.1
Amz_5447	CRISPR-associated_helicase_Cas3	Alipo	1,00E-05	YP_005039480.1

Espécie: representa a presença da proteína ortóloga em *Azospirillum* sp. B510 (AB510) e em *A. lipoferum* (Alipo). Número de acesso: representa o número de acesso da proteína de espécie de *Azospirillum* que apresentaram melhor similaridade.

Como podemos observar na **Tabela 3.16**, a proteína Amz\_3764 é compartilhada com *Azospirillum* sp. B510 e *A. lipoferum* com boa similaridade, o mesmo ocorre para as proteínas Amz3765 e Amz5447, porém com maiores diferenças na sequência de aminoácidos. As nove proteínas restantes parecem ser encontradas somente em *A. amazonense*.

### 3.5. Caracterização das proteínas específicas de *A. amazonense* com relação a *A. brasilense*.

Na análise de diferença representacional (RDA) realizada, 46 sequências exclusivas de *A. amazonense* com relação a *A. brasilense* foram identificadas (ver **Tabela Suplementar 3**). Dessas 46 sequências, 29 apresentaram similaridade com produtos proteicos já caracterizados em outros organismos (ver **Tabela 3.7**). Algumas proteínas identificadas como exclusivas pertencem a classes funcionais relacionadas à biossíntese de metabólitos secundários, transcrição (proteínas codificadas pelo DNA de bacteriófagos); transporte e catabolismo; metabolismo de carboidratos; modificação pós-traducional, *turnover* de proteínas e chaperonas. As proteínas relacionadas a essas classes metabólicas são brevemente analisadas de maneira individual nos itens que seguem.

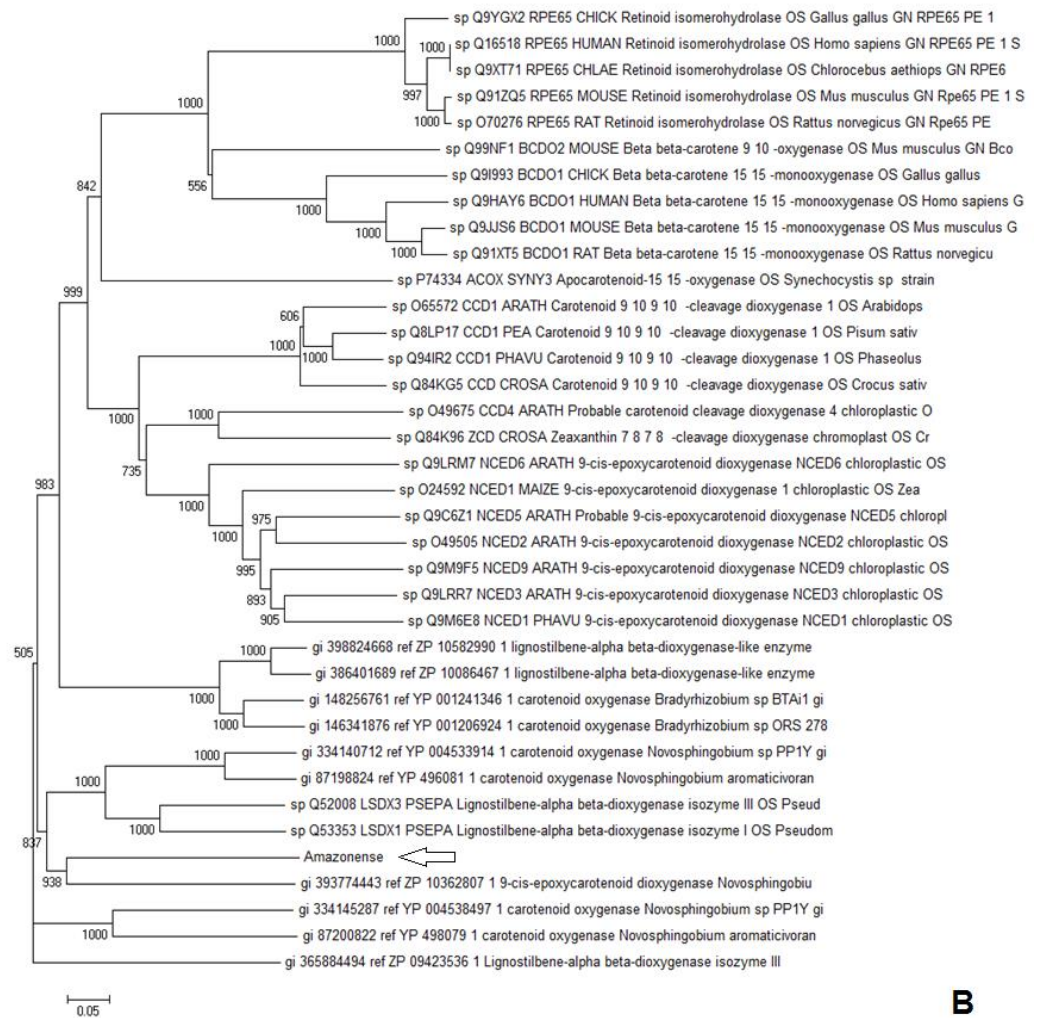
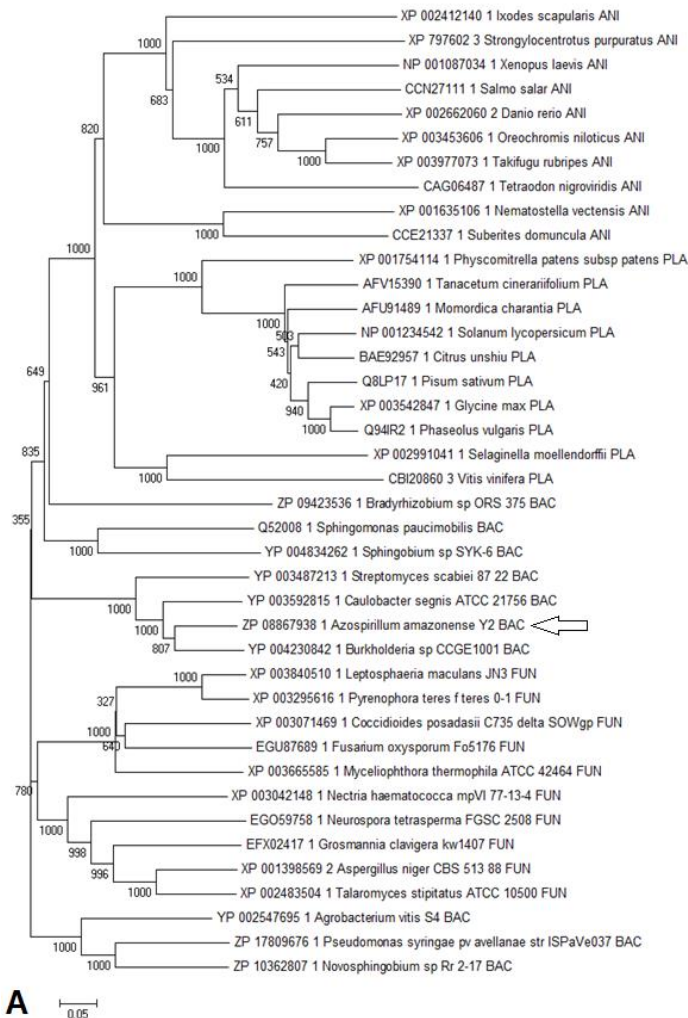
### 3.5.1. Biossíntese de metabólitos secundários, transporte e catabolismo

#### 3.5.1.1. Carotenoide/Lignistilbeno Oxigenase

Carotenoides são pigmentos terpenoides produzidos por todos os organismos fotossintéticos, algumas bactérias e fungos. Como exemplos de carotenoides podemos citar o retinal, cromóforo dos fotorreceptores e o fitormônio ácido abscísico (ABA), derivado da clivagem da 9-cis-violaxantina ou 9-cis-neoxantina, reação catalisada pela 9-cis-epóxicarotenoide dioxigenase (CED) em vegetais (CHERNYS & ZEEVAART, 2000; TAN *et al.*, 2003).

A produção de carotenoides em espécies de *Azospirillum* já é conhecida há bastante tempo (NUR *et al.*, 1981). Fatores ambientais como temperatura, limitação de CO<sub>2</sub> e alta irradiação podem causar um desequilíbrio entre a absorção de energia e seu uso, formando radicais livres. Os carotenoides exercem sua função de proteção quelando esses radicais (HARTMANN & HUREK, 1988).

Em *A. amazonense*, o homólogo da 9-cis-epóxicarotenoide dioxigenase (CED) (Amz0467 – acesso ZP\_08867938.1) possui 493 aminoácidos e apresentou maior similaridade com a proteína dos gêneros *Caulobacter* ( $\alpha$ -proteobactérias) e *Burkholderia* ( $\beta$ -proteobactérias), além de *Streptomyces* (Actinobactéria). Foi realizada uma análise utilizando 10 ortólogos de várias espécies de diferentes reinos, que apresentavam melhor similaridade em *A. amazonense*. A **Figura 3.7A** mostra que esta proteína é bastante conservada entre espécies de diferentes reinos, sendo que proteínas bacterianas possuem conservação com proteínas de fungos e plantas.



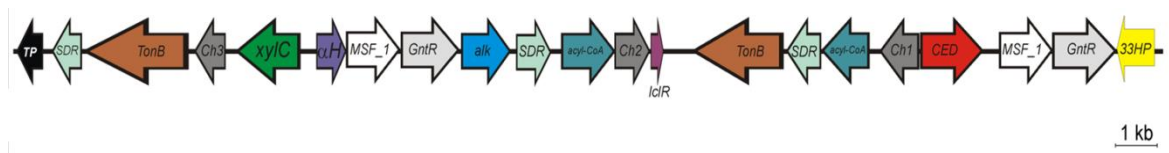
**Figura 3.7** – Análise filogenética da Carotenoide/Lignistilbeno Oxigenase de *A. amazonense*.

A figura mostra a conservação entre diferentes organismos (A) e possíveis produtos (B). Os alinhamentos foram realizados com Muscle 3.7 e a árvore filogenética com MEGA 5.2.

Com o objetivo de identificar possíveis substratos para esta enzima, foram obtidos alguns ortólogos com a função caracterizada experimentalmente. Para a construção dessa filogenia, sequências com homologia a proteína de *A. amazonense* (curadas no *Uniprot/Swissprot*) e os ortólogos de *Bradyrhizobium* e *Novosphingobium* foram utilizados. O resultado demonstrou maior proximidade com a dioxigenase de lignistilbeno de *Pseudomonas* e *Novosphingobium* (ver **Figura 3.7B**) sugerindo que o substrato da enzima de *A. amazonense* seja o mesmo das proteínas destas bactérias (estilbenos) (MARASCO & SCHMIDT-DANNERT, 2008).

No contexto gênico da 9-cis-epóxicarotenoide dioxigenase em *A. amazonense* (ver **Figura 3.8**), podemos visualizar a presença de genes codificando diferentes transportadores das famílias MSF\_1, conhecidos pela mediação de um amplo espectro de substâncias, e da família TonB, relacionada ao transporte de metais como ferro. Nesta mesma região genômica, podemos observar genes do metabolismo de compostos aromáticos como a benzaldeído desidrogenase (XylC) e genes codificantes da alcanal monoxigenase (*alk*), sendo que alcanos são os hidrocarbonetos mais presentes em óleos sintetizados por plantas e algas. Estas enzimas estão bem caracterizadas em bactérias sendo que *Pseudomonas* possui os estudos mais detalhados (MARÍN *et al.*, 2003).

Apesar de testes experimentais serem necessários para esclarecer a função da carotenoide/Lignistilbeno oxigenase em *A. amazonense*, sugere-se que esta enzima possa ser utilizada no metabolismo de carotenoides, como no caso de síntese/mediação do ABA ou na degradação de estilbenos, ambas as atividades intimamente relacionadas com a interação *Azospirillum*-planta, sendo portanto de interesse para o melhor entendimento do mecanismo dessa associação.



**Figura 3.8** - Região genômica da carotenoide/lignistilbeno oxigenase de *A. amazonense*.

**TP** - transposase; **SDR** – cadeia curta da desidrogenase redutase; **TonB** – receptor dependente de TonB; **Ch(1-3)** – proteína hipotética; **XyIC** – benzaldeído desidrogenase; **αH** - esterase; **MSF1** – Transportador tipo MSF; **GntR** – regulador transcricional GntR; **alk** – alcanal monoxigenase; **Acyl-CoA** – proteína contendo domínio AcilCoA desidrogenase; **CED** – 9-cis-epoxicarotenoide dioxigenase; **33HP** - monoxigenase associada a FAD.

### 3.5.1.2. Esqualeno ciclase

A proteína com função conhecida mais próxima a essa sequência (Amz3156) é a esqualeno ciclase de *Streptomyces* (ZP\_07276748.1) que apresenta um alto valor de similaridade (4e-92 no delta blast).

A metabolização do esqualeno para hopeno, um triterpeno encontrado em procariotos, foi estudada em *Streptomyces* por cromatografia gasosa e espectrometria de massa. A proteína apresentou oito motivos QW além de um domínio rico em aspartato, clássico da estrutura destas proteínas relacionadas à rigidez da membrana em procariotos, similar a atuação do colesterol (GHIMIRE *et al.*, 2009).

## 3.5.2. Transcrição

### 3.5.2.1. Proteína relacionada a fagos

O gene que codifica a proteína relacionada a fagos Amz\_1185 (ZP\_08868360.1), isolado pela metodologia de RDA, demonstrou maior similaridade com um ortólogo codificante de uma proteína de ligação ao DNA do fago phiCD27 de *Clostridium*. A proteína Amz\_1185 possui, também, similaridade com a proteína repressora c1 do prófago Lj771 de *Lactobacillus johnsonii*,



responsável pelo estabelecimento e manutenção da lisogenia deste vírus. No entanto, não foi encontrada similaridade com proteínas do fago CD de *A. brasilense* ou com qualquer outra proteína codificada pelo genoma dessa espécie.

A atividade mais provável para esta proteína está relacionada com a regulação da transcrição pela ligação ao DNA (os sítios de ligação ao DNA e identificação domínios encontrados nessa proteína pode ser vista na **Tabela Suplementar 14**). As espécies hospedeiras dos fagos que apresentam no seu genoma sequências mais próximas a proteína de *A. amazonense* pertencem aos gêneros *Photorhabdus*, *Clostridium* e mais distantemente *Burkholderia*. A análise da região genômica onde se localiza o gene que codifica a proteína Amz1184 de *A. amazonense* sugere a possibilidade de o agrupamento gênico ter sido herdado de sistemas virais que parasitaram este genoma durante sua evolução.

### 3.5.3. Transporte e metabolismo de carboidratos

#### 3.5.3.1. Transportador de Fucose Galactose; Lactato permease; Desalogenase Haloácida (HAD).

A metodologia de RDA possibilitou o isolamento de três sequências de DNA específicas de *A. amazonense* que foram classificadas no grupo de “Transporte e metabolismo de carboidratos” e que codificam para as proteínas Desalogenase Haloácida (Amz\_3005), Lactato permease (Amz\_4590 ZP\_08870398.1) e Transportador de Fucose e Galactose (Amz\_4139 - ZP08867960.1) (ver **Tabela Suplementar 3**).

A análise comparativa da desalogenase haloácida (Amz\_3005) de *A. amazonense* revelou que a sequência mais similar encontrada foi de uma proteína de *Nitrosomonas* (YP\_004693714.1) que contém o domínio desalogenase haloácida (Q0AIW3). A utilização de compostos orgânicos halogenados como fonte de carbono apresenta grandes vantagens de adaptação ambiental para esta bactéria.

O transportador de Fucose/Galactose (Amz4139 - ZP\_08867960.1) isolado de *A. amazonense* pertence a família MSF, que também apresentou alta similaridade com transportadores de outros açúcares como glicose e galactose. Trabalhos iniciais com *A. amazonense* já haviam demonstrado a capacidade desta espécie metabolizar esses açúcares, característica que não é compartilhada com *A. brasilense* (MARTÍNEZ-DRETS *et al.*, 1985; MAGALHÃES *et al.*, 1983). Este transportador de glicose/galactose de *A. amazonense* possui doze regiões transmembrana (localizados pelo Psort/TCDB).

O último gene isolado de *A. amazonense* relacionado a utilização de açúcar como fonte de carbono codifica para a Lactato permease (Amz\_4590 - ZP\_08870398.1). A proteína codificada por este gene apresentou três regiões transmembrana (localizadas pelo PSORT), sendo que a maior similaridade foi com a proteína L-lactato permease de *Sulfurospirillum barnesii* (YP\_006403938.1).

#### 3.5.4. Modificação pós-traducional, turnover de proteínas e chaperonas

##### 3.5.4.1. Glutamina ciclotransferase.

A enzima glutamina ciclotransferase catalisa a reação de formação de piroglutamato na sequência amino terminal de muitas proteínas e peptídeos para a maturação dessas biomoléculas. São classificadas em dois grupos, sendo que proteínas de plantas, bactérias e parasitas formam uma classe única, seguida pela dos mamíferos (WINTJENS *et al.*, 2006).

A sequência encontrada em *A. amazonense* (Amz\_5170 - ZP\_08870736.1) possui alta similaridade com um ortólogo de *Xanthomonas* (YP\_005628993.1), que possui sua estrutura tridimensional caracterizada. O alinhamento pode ser visto na **Figura Suplementar 5**.

A comparação da sequência de aminoácidos da proteína de *A. amazonense* demonstrou similaridade com proteínas com atividades de modificação pós traducional, *turnover* e de chaperona (outras características para este gene podem ser vista no banco de ortólogos do KEGG K00683).

## 4. Discussão

A importância do gênero *Azospirillum* como promotor do crescimento vegetal é conhecida há bastante tempo, devido principalmente à metabolização de fitormônios, fixação biológica do nitrogênio e inibição da multiplicação de patógenos nos diversos ambientes (solo, raízes vegetais entre outros). Apesar disso, apenas no ano de 2010 um microrganismo desse gênero teve seu genoma completamente sequenciado (KANEKO *et al.*, 2010). A colaboração do nosso grupo de pesquisa com os pesquisadores do Laboratório Nacional de Computação Científica possibilitou a realização do sequenciamento do genoma da espécie *A. amazonense*, alvo principal de estudos do nosso grupo. Para a montagem do genoma de *A. amazonense* foi inicialmente utilizado como referência o genoma completo de *Azospirillum* sp. B510 disponibilizado em 2010 (KANEKO *et al.*, 2010). No entanto, estudos detalhados revelaram pouca similaridade na organização gênica entre estas duas espécies. Posteriormente, foi incluído o genoma de *Rhodospirillum centenum* (LU *et al.*, 2010) para a anotação das sequências no genoma de *A. amazonense*. A anotação automática foi refinada através de uma análise manual criteriosa para definição das proteínas preditas no genoma de *A. amazonense* utilizando *R. centenum* e *Azospirillum* sp. B510 como referência na predição destas proteínas. Através desta metodologia 3319 ORFs (do inglês *open reading frame*) foram anotadas e as vias relacionadas com a fixação do nitrogênio, metabolismo energético, *quorum sensing*, resistência a antibióticos, quimiotaxia entre outras, foram caracterizadas (SANT'ANNA *et al.*, 2011). Nestas análises, no entanto, não foram observadas proteínas relacionadas à produção de fitormônios, que é um dos principais interesses para espécies do gênero *Azospirillum* e outros gêneros associados com vegetais.

No presente trabalho a anotação foi realizada de maneira automática, porém com novas ferramentas de bioinformática e cruzamento de informações de diferentes bancos de dados de sequências curadas, sendo 5.496 proteínas (2.177 adicionais com relação à primeira publicação) identificadas. Vantagens de uma anotação completamente automatizada incluem a reprodutibilidade das análises, que diminuem o erro humano, já que são grandes quantidades de informações a

serem analisadas, além de uma diminuição considerável no tempo de trabalho. Para a anotação do genoma de *A. amazonense*, foi utilizado o genoma de *Azospirillum* sp. B510 como referência, já que este trabalho tem por objetivo principal a comparação de genomas deste gênero. Devido ao grande número de proteínas de *A. amazonense* que apresentou baixa similaridade com sequências de proteínas de *Azospirillum* sp. B510, uma comparação com um banco de sequências contendo os genomas microbianos mais relacionados a *A. amazonense* foi criado localmente, para ser utilizado nas comparações. Como um grande número de sequências ainda apresentou baixa similaridade nas comparações realizadas, uma anotação foi feita no banco de dados GenBank, através do PSI Blast, já que este faz buscas otimizadas para domínios nas sequências, ideal para amostras que apresentaram baixa similaridade na comparação global das proteínas. Finalmente foi efetuada uma junção dos resultados gerados pelas duas anotações, local e no banco de dados GenBank. O valor de similaridade mínimo utilizado como padrão no PSI Blast (0,005) foi mantido nesta análise que resultou em aproximadamente 500 sequências proteicas de *A. amazonense* com Valor E superior a esse limite, sendo consideradas como sequências hipotéticas de *A. amazonense*. O número de sequências classificadas como hipotéticas, subtraído do total de 5.496, resulta em quase 5.000 proteínas com similaridade em alguma das análises realizadas. Na comparação entre os genomas de *Azospirillum* podemos observar que *A. brasilense* é a espécie com genoma sequenciado mais próxima a *A. amazonense*. Ela apresenta aproximadamente 400 proteínas conservadas adicionais quando comparado com *Azospirillum* sp. B510 e *A. lipoferum* simultaneamente. Esse resultado era esperado quando observamos a filogenia do gene *nifH*, comumente utilizado para classificar as espécies de diazotróficos, que também mostra *A. brasilense* mais próxima a *A. amazonense* (SANT'ANNA *et al.*, 2011). Essa filogenia também corresponde com os resultados encontrados na comparação com os 10 genomas microbianos mais próximos a *A. amazonense* realizada neste trabalho (ver **Tabela 3.1**), que também mostrou *B. japonicum* como espécie mais similar a *A. amazonense* com relação a outras espécies deste gênero. Isso também ocorre no caso de *R. centenum*, que possui um genoma com tamanho e número de proteínas mais próximos a *A. amazonense*. Outro fator que

possivelmente influencia nesse resultado é a maior quantidade de proteínas em *A. brasilense* (7.557 proteínas), que possui um genoma maior em comparação com as outras espécies de *Azospirillum* com genoma conhecido e conseqüentemente maiores chances de possuir proteínas similares, como visto na **Figura 3.1**. Essa análise permitiu concluir que *A. brasilense* possui maior número de ortólogos em *A. amazonense* considerando os membros deste gênero, enquanto *Azospirillum* sp. B510 e *A. lipoferum* resultaram em praticamente o mesmo número de proteínas com melhor alinhamento com as correspondentes em *A. amazonense* (ver **Tabela 3.4**). Esses resultados estão de acordo com os realizados no primeiro estudo, que apresentam *A. amazonense* como um genoma composto por regiões com similaridade em diversos organismos próximos filogeneticamente (SANT'ANNA *et al.*, 2011).

Considerando *Azospirillum* sp. B510 e *A. lipoferum*, podemos observar que são as espécies mais semelhantes entre si, já que contém grande quantidade de alinhamentos com valor H alto (ver **Figura 3.1**). Essa proximidade também pode ser vista na análise filogenética do gene *nifH*. *Azospirillum* sp. B510 é o organismo que possui menor número de proteínas sem similaridade (Valor H próximo a zero) com relação a *A. amazonense*. Como conclusão, podemos considerar o genoma de *A. amazonense* como formado por uma mistura de genes recebidos de organismos próximos geneticamente, sendo mais semelhante a *R. centenum* e *B. japonicum*. Entre as espécies de *Azospirillum*, *A. brasilense* é a mais próxima cujo genoma é conhecido, possuindo aproximadamente 1400 proteínas com melhor similaridade quando analisada em conjunto com *A. lipoferum* e *Azospirillum* sp. B510 (ver **Tabela 3.4**).

Uma das características mais relevantes no gênero *Azospirillum* está relacionada com a promoção do crescimento vegetal, portanto análises foram realizadas para a comparação do genoma de *A. amazonense* com genomas de gramíneas conhecidamente associadas a esta bactéria. Mais especificamente, os genomas de *Oryza sativa* e *Arabidopsis thaliana* foram utilizados como referência para identificação de proteínas possivelmente envolvidas em rotas de fitormônios. Após muitos anos de estudo, apenas os genes *ipdC*, *iaaH* e *iaaM* foram comprovadamente associados com a produção de fitormônios por bactérias,

processo que foi demonstrado em organismos que não possuem ortólogos para nenhum dos três genes citados. Além disso, a grande maioria dos estudos de fitormônios em bactérias tratam de auxinas, sendo que poucos identificam proteínas relacionadas a giberelinas, citocininas, etileno e ácido abscísico (KANEKO *et al.*, 2010; RODRIGUES *et al.*, 2008).

Neste trabalho 276 proteínas com função conhecida de *A. amazonense* que não possuem similaridade nas outras 3 espécies (Valor H=0) foram identificadas e caracterizadas (ver **Tabela Suplementar 2**) na análise *in silico*. Segundo a relação de classificação no banco de dados COG, podemos perceber que a maioria das proteínas exclusivas de *A. amazonense* com função conhecida está relacionada com o metabolismo de carboidratos, aminoácidos, motilidade e secreção, conforme visualizado na **Tabela 3.5**. Entre as proteínas isoladas que merecem destaque nessa análise e não pertencem às classes estudadas neste trabalho, podemos citar a l-ornitina monoxigenase ZP\_08869952.1 (Amz\_3892). Ela possui similaridade com a proteína YUCCA9 de *A. thaliana* (NP\_171914.1), envolvida na síntese de auxinas (ZHAO *et al.*, 2001). Foi demonstrado recentemente, que apenas duas reações são necessárias para a conversão de triptofano para auxina em plantas, uma delas realizada pela YUCCA (ZHAO, 2012), o que gera novas possibilidades de síntese também em bactérias, nas quais a l-ornitina monoxigenase poderia estar envolvida.

Esses resultados estão de acordo com os encontrados na análise das vias metabólicas de *A. amazonense* realizadas neste trabalho, que apresentou as mesmas vias como destaque para *A. amazonense*. Genes codificantes da Rubisco, genes para o metabolismo de salicina, e outras fontes de carbono diferenciais de *A. amazonense* foram discutidas em Sant'anna *et al.* 2011. As principais classes diferenciais como do sistema de secreção, proteínas virais e do sistema imune CRISPR aqui relacionadas, foram apresentadas nas **seções 3.3.2 e 3.4** respectivamente.

A metodologia de RDA é conhecida por proporcionar o isolamento de sequências de DNA presentes em uma amostra e ausentes em outra, através do uso sequencial das técnicas previamente conhecidas de hibridização subtrativa,

que elimina as sequências conservadas por amplificação das sequências únicas através da PCR (LISITSYN *et al.*, 1993). Pela metodologia ser conhecida em nosso grupo de estudos e com a disponibilização da técnica de OSPERT, que poderia otimizar os resultados, diminuindo o tempo de trabalho e aumentando a eficiência pela estringência da hibridização subtrativa, essa técnica foi aplicada para a caracterização *in vitro* de sequências únicas de *A. amazonense* com relação a *A. brasilense*.

Os resultados demonstraram que após a realização de três ciclos de hibridização subtrativa seguidos de amplificação diferencial, 480 clones foram sequenciados, gerando 113 diferentes *contigs* quando submetidos às avaliações em programas de análise de qualidade de sequências (STADEN *et al.*, 2000). A análise individual sugeriu que a maioria das amostras pertenciam exclusivamente a *A. amazonense* (ver **Figura 3.2A**). Devido ao pequeno tamanho da maioria das amostras, além da maioria não possuir função conhecida, elas foram utilizadas como base para selecionar as 72 proteínas completas com melhores alinhamentos no genoma dessa espécie, onde 46 proteínas foram definidas como diferenciais ( $H < 0,42$ ) (ver **Tabela Suplementar 3**). Apesar do baixo número de proteínas identificadas como exclusivas ao final deste experimento, ele pode ser considerado como bem sucedido, já que os 113 consensos iniciais apresentaram em sua maioria baixos valores de similaridade. A alta redundância encontrada nas sequências sugere que um ou dois ciclos de RDA/OSPERT possa ser suficiente para a caracterização de sequências diferenciais entre esses dois organismos, já que observamos na análise completa dos genomas que existem grandes diferenças entre essas espécies. Cabe salientar que, com a metodologia de RDA, foi identificada uma proteína altamente similar com a 9-cis-epoxicarotenoide dioxigenase de vegetais, relacionada com o metabolismo do fitormônio etileno e cujo ortólogo mais próximo em microrganismos foi apontado como responsável pela metabolização de estilbenos como resveratrol, em testes realizados com *Pseudomonas* e *Novosphingobium* (MARASCO & SCHMIDT-DANNERT, 2008). Também foram identificadas proteínas do metabolismo de açúcares e sequências relacionadas a fagos, que pertencem a classes de proteínas com maiores diferenças encontradas na análise completa dos genomas.

A análise da **Tabela 3.2** apresenta as rotas metabólicas que se destacaram na comparação com o banco COG. Destas, quatro foram selecionadas pela maior quantidade de proteínas encontradas em *A. amazonense*. Nas duas primeiras classes, relacionadas respectivamente com o metabolismo de lipídeos e compostos secundários, podemos observar que um número semelhante de proteínas foi identificado em *A. amazonense* e *A. lipoferum*. As outras duas espécies, *A. brasilense* e *Azospirillum* sp. B510, também apresentam entre si um número similar de proteínas, sugerindo que apesar da maior similaridade ter sido encontrada entre os genomas de *A. amazonense* e *A. brasilense*, essas vias possuem herdabilidade diferenciada, o que confirma a teoria de *A. amazonense* ser formado por uma mistura de outros organismos próximos filogeneticamente.

A classe seguinte, relacionada com transporte intracelular e secreção apresentou um número semelhante de proteínas entre *A. amazonense* e *A. brasilense*. Proteínas relacionadas com transporte foram estudadas nesse trabalho (**seção 3.3.2**), quando classificamos proteínas de *A. amazonense* exclusivas nessa espécie. Proteínas relacionadas ao sistema de secreção do tipo IV estão relacionadas com a transferência de material genético através da membrana, participando ativamente da evolução dos genomas, conforme demonstrado em *Agrobacterium* (CHO *et al.*, 2009). A presença deste sistema em *A. amazonense* sugere a possibilidade desta espécie conter um mecanismo adicional de transferência de material genético não presente nas outras espécies de *Azospirillum* analisadas, o que poderia contribuir para uma maior transferência horizontal de genes e conseqüentemente uma maior diferenciação deste genoma, conforme apresentado neste trabalho (ver **Figura 3.1**).

A última classe destacada na comparação com este banco de dados corresponde aos mecanismos de defesa. Nessa classe observamos a maior diferença no número de proteínas identificadas em *A. amazonense* comparativamente às outras três espécies. Mecanismos de defesa em *A. amazonense* assim como transportadores em geral foram estudados e classificados no primeiro trabalho sobre o genoma de *A. amazonense* realizado pelo nosso grupo. Neste trabalho, foi demonstrado que *A. amazonense* possui uma grande quantidade de genes de transportadores da família TonB,



relacionados com o transporte ativo através da membrana, de metais como ferro. Com relação a mecanismos de defesa, foram apresentados transportadores da subfamília QacA que confere resistência a antissépticos lipofílicos e desinfetantes compostos de amônia. Genes relacionados com resistência a penicilina, tetraciclina e glioxalase/bleomicina também foram identificados (SANT'ANNA *et al.*, 2011). Isso está de acordo com o grande número destacado de transportadores de lipídeos visualizado na **Tabela 3.2**.

Na **Tabela 3.3** podemos observar que um número bastante semelhante de proteínas é encontrado para as quatro espécies estudadas na maioria das classes. Nesta análise realizada comparando as proteínas de *Azospirillum* no banco de ortólogos do KEGG chama atenção o metabolismo de triptofano (ko00380) e biossíntese de terpenoides (ko00130), já que estão relacionadas a produção de hormônios nesse gênero bacteriano (CARREÑO-LOPEZ *et al.*, 2000; BARTEL, 1997). As maiores diferenças encontradas porém, estão relacionada com o metabolismo da galactose (ko00052) e de glicanos (ko00511), além da sinalização celular relacionada com fatores de virulência (ko05120). Nesta última estão presentes proteínas do sistema de secreção do tipo IV analisadas na **seção 3.3.2**. O metabolismo de galactose já é bastante conhecido como um diferencial de *A. amazonense* neste gênero, tendo sido estudado inicialmente por (Martínez-Drets *et al.* (1985), sendo que as proteínas relacionadas ao metabolismo de galactose assim como de alguns glicanos como celulose e quitina foram identificadas por Sant'Anna *et al.* (2011).

A caracterização de proteínas de *A. amazonense* relacionadas com o metabolismo de hormônios vegetais consiste em uma das análises realizadas neste trabalho, já que representa um dos principais motivos pelos quais um microrganismo promove o crescimento das plantas às quais se associa. Essa análise ocorreu pelo pouco conhecimento no que se refere à proteínas comprovadamente associadas ao seu metabolismo, já que bactérias conhecidamente sintetizadoras desses elementos não apresentaram as proteínas consideradas como responsáveis pela sua presença em culturas analisadas em laboratório, incluindo a bactéria alvo deste estudo (KANEKO *et al.*, 2010; SANT'ANNA *et al.*, 2011).

Com o objetivo de localizar proteínas de *Azospirillum*, principalmente em *A. amazonense*, relacionadas com as rotas de hormônios em plantas três análises foram realizadas. A primeira foi efetuada no banco de ortólogos do KEGG, com *A. thaliana* e *O. sativa* como referências. A comparação dessas vias metabólicas para as quatro espécies pode ser visualizada na **Figura Suplementar 1** e apresentou 7 proteínas conservadas para *A. amazonense*. Essas proteínas foram vinculadas a síntese de auxinas, citocininas, ácido abscísico e etileno. Na segunda análise, o genoma de *A. amazonense* foi comparado no banco de hormônios de Arabidopsis (AHD), onde a **Tabela Suplementar 5** apresenta as proteínas identificadas. No teste seguinte, um banco de proteínas identificadas como diferencialmente expressas na presença de AIA em *A. brasilense* em uma análise de micro arranjos foi obtido no DRAFT de *A. brasilense* (VAN PUYVELDE *et al.*, 2011) e comparadas com o genoma de *A. amazonense*. Das 160 proteínas consideradas como reguladas por auxina, 107 foram encontradas no genoma de *A. amazonense* e, então, relacionadas àquelas similares ao AHD. Doze proteínas foram classificadas na análise do cruzamento dos dados (ver **Tabela 3.9**). Das 12 proteínas identificadas, duas vias metabólicas se destacaram, sendo elas a de transportadores ABC (ko02010) e metabolismo de purinas (ko00230) que apresentaram 3 proteínas cada. Duas proteínas estão relacionadas com a Coenzima A, relacionada ao metabolismo de ácidos graxos. Proteínas com vínculo à Coenzima A aparecem como um grupo significativamente presente nas proteínas de *A. amazonense* conservadas com o cromossomo V de *A. thaliana*. Proteínas relacionadas com a Coenzima A e SDR (*short desidrogenase redutase*) que se destacaram nessa análise, também estão presentes no contexto da CED de *A. amazonense*, analisada neste trabalho (ver **Figura 3.8**). Um possível transportador de auxina também foi identificado na comparação entre os genomas de *Azospirillum* que somado à presença da ornitina monoxigenase (ZP\_08869952.1) (ver **Tabela Suplementar 2**), foram um primeiro incentivo para a realização dos outros testes efetuados na busca por proteínas relacionadas com o metabolismo de fitormônios em *Azospirillum* (ver **Figura Suplementar 1**).

A análise das 2.177 proteínas que não foram estudadas no primeiro trabalho sobre o genoma de *A. amazonense* realizada pelo nosso grupo,

destacou uma série de sequências relacionadas ao sistema de secreção do tipo IV nessa espécie, que está vinculado com o transporte de material genético através da membrana. Quando as 60 proteínas identificadas em *A. amazonense* no banco de sequências curadas Secret4 foram comparadas pelo mGenomeSubtractor, 18 proteínas apresentaram baixa ou nenhuma similaridade nas outras 3 espécies do gênero *Azospirillum* (ver **Figura Suplementar 2**). Entre essas 18 proteínas exclusivas, 8 foram vinculadas a classe U do banco COG, relacionadas com o metabolismo de secreção e transporte, fortalecendo a confiabilidade da análise. Entre as publicações que relacionam a função das proteínas que foram identificadas como exclusivas de *A. amazonense* neste grupo destacamos o estudo em *Agrobacterium tumefaciens*, que vincula essas proteínas ao comportamento de doadora ou receptora de material genético dessa bactéria na rizosfera (CHO *et al.*, 2009). Além da análise por regiões genômicas de *A. amazonense* vinculadas ao sistema de secreção do tipo IV, realizamos uma comparação por similaridade no banco de proteínas de transporte TCDB (SAIER *et al.*, 2006), com a finalidade de visualizar de forma geral os transportadores dessa espécie. Como podemos visualizar na **Tabela Suplementar 9**, que apresenta proteínas de transporte identificadas simultaneamente no TCDB e classificadas como pertencentes a classe U do banco COG, estão presentes em *A. amazonense* proteínas do sistema de secreção bacteriana do tipo II, relacionado entre outras funções, com a degradação da parede celular vegetal, podendo portanto, atuar simultaneamente com proteínas do sistema do tipo IV, relacionado com a troca de material genético entre bactérias e entre bactérias e plantas (SANDKVIST, 2001). Entre as enzimas deste sistema encontradas em *A. amazonense* podemos destacar enzimas com alta similaridade em celulases (YP\_003594685.1) e quitinases (YP\_003763833.1), conhecidas pelo envolvimento na interação bactéria-planta ou fungo-planta. Um exemplo conhecido ocorre com fungos do gênero *Metarhizium*, conhecido por utilizar essas enzimas na rizosfera e na infecção de insetos, possuindo grande interesse no controle biológico (SCHRANK & VAINSTEIN, 2010).

A publicação que apresentou os genomas completos de *A. brasilense* e *A. lipoferum* destacou a origem aquática do gênero *Azospirillum* e a incidência da

infecção por genomas virais ocorridas nesse gênero a partir da invasão dessa bactéria no ambiente terrestre, conquistado por espécies vegetais com as quais passou a se associar nessa migração de habitat (WISNIEWSKI-DYÉ *et al.*, 2011).

Devido a presença de proteínas CAS (CRISPRs associadas) no resultado da análise de proteínas exclusivas de *A. amazonense* e com as informações sobre elementos genéticos móveis nos genomas de *Azospirillum* citadas acima, além do conhecimento sobre a presença de um fago integrado na espécie *Azospirillum brasilense* CD, levaram a análises mais completas sobre a presença de elementos genéticos móveis nesse gênero bacteriano. Os resultados apresentados na **Figura Suplementar 3** demonstram que *Azospirillum* sp. B510 possui 12 regiões virais, sendo uma das principais diferenças conhecidas até o momento entre essa espécie e *A. lipoferum*. Nesta figura podemos perceber que apesar de *Azospirillum lipoferum* possuir uma região maior identificada como similar ao fago CD (31,4kb), com relação aos 27,1 kb identificados em *Azospirillum* sp. B510 (ver **Tabela Suplementar 11**), a região encontrada na primeira espécie possui uma grande quantidade de proteínas hipotéticas pertencentes ao hospedeiro inseridas na sequência, enquanto em *Azospirillum* sp. B510 apresenta grande quantidade de proteínas identificadas como similares ao fago.

Os resultados das análises da presença de plasmídeos e fagos nos genomas de *Azospirillum* (testes do PHAST) levam a acreditar na possibilidade de que a presença de *Azospirillum* no interior dos vegetais pode facilitar a integração de sequências virais no genoma bacteriano, já que a presença de proteínas do fago CD, por exemplo, estão presentes principalmente nas espécies de *Azospirillum* conhecidas por atuarem de forma endofítica na associação com vegetais com as quais essas bactérias foram localizadas (CAIOLA *et al.*, 2004; YASUDA *et al.*, 2009). Outra possibilidade é a transferência do vírus diretamente entre as espécies bacterianas, devido a alta complexidade do nicho da rizosfera, ocupado por uma grande quantidade de organismos, que podem transferir material genético de maneira horizontal. Testamos a similaridade das 95 proteínas do fago CD com relação aos genomas de *Azospirillum* (não mostrado), e considerando um valor padrão de H de 0,1 (mínimo de similaridade) podemos

perceber que ele possui 11 proteínas conservadas em *A. amazonense*, 19 em *A. lipoferum*, 22 para *Azospirillum* sp. B510 e 26 para *A. brasilense* 245. Quando o valor mínimo de similaridade exigido é aumentado para 0,42 os resultados são 2, 7, 6 e 3 respectivamente, mostrando que a conservação é bastante semelhante entre *Azospirillum* sp. B510 e *A. brasilense* 245.

Analisando estes resultados podemos sugerir que a baixa conservação deste elemento genético em vegetais e a presença efetiva em genomas próximos filogeneticamente a *Azospirillum*, como *R. centenum* e *C. segnis* (comparadas neste trabalho com *A. amazonense*), ainda que possuindo menor similaridade, indicam uma relação vertical na transferência deste fago, que pode ter ocorrido há bastante tempo durante sua evolução.

De maneira geral, os resultados encontrados no presente trabalho confirmam as conclusões sobre o genoma de *A. amazonense* feitas em nossa primeira análise (SANT'ANNA *et al.*, 2011), quando indicamos que este é composto por regiões com similaridade em diferentes organismos próximos filogeneticamente, principalmente *R. centenum* e *B. japonicum*. Aqui também destacamos a alta homologia com *C. segnis* e com uma espécie de  $\beta$ -proteobactéria (*Burkholderia* sp. CCGE1001) (ver **Tabela 3.1**). Na comparação entre os genomas de *Azospirillum*, destacamos a maior homologia com *A. brasilense* entre os genomas conhecidos, já que em termos de análise filogenética *A. amazonense* aparece mais próximo a *A. irakense*, compartilhando com esta genes conhecidos na utilização de compostos de carbono alternativos, como a salicina (SOMERS *et al.*, 2000). Entre as diferenças destes genomas ainda não conhecidas, identificamos e caracterizamos proteínas do sistema de secreção tipo IV, presença de fagos e sistema imune. Entre os resultados mais importantes, apresentamos a primeira análise para este organismo que trata de proteínas relacionadas ao metabolismo de fitormônios, além de comparar sua presença nas diferentes espécies. A confiabilidade do resultado é devida principalmente ao cruzamento de diferentes testes, realizadas com espécies de plantas e com um banco de proteínas diferencialmente expressas na presença de auxina em *A. brasilense*. Com esses resultados, a utilização futura destas espécies como inoculantes, e o surgimento de novas possibilidades

biotecnológicas é aumentado significativamente, já que apresenta resultados de diferentes níveis de comparação, partindo da comparação entre diferentes genomas, seguido por comparações de classes metabólicas e por fim análises individuais de proteínas de interesse, além de contribuir para um melhor entendimento de sua importância ecológica.

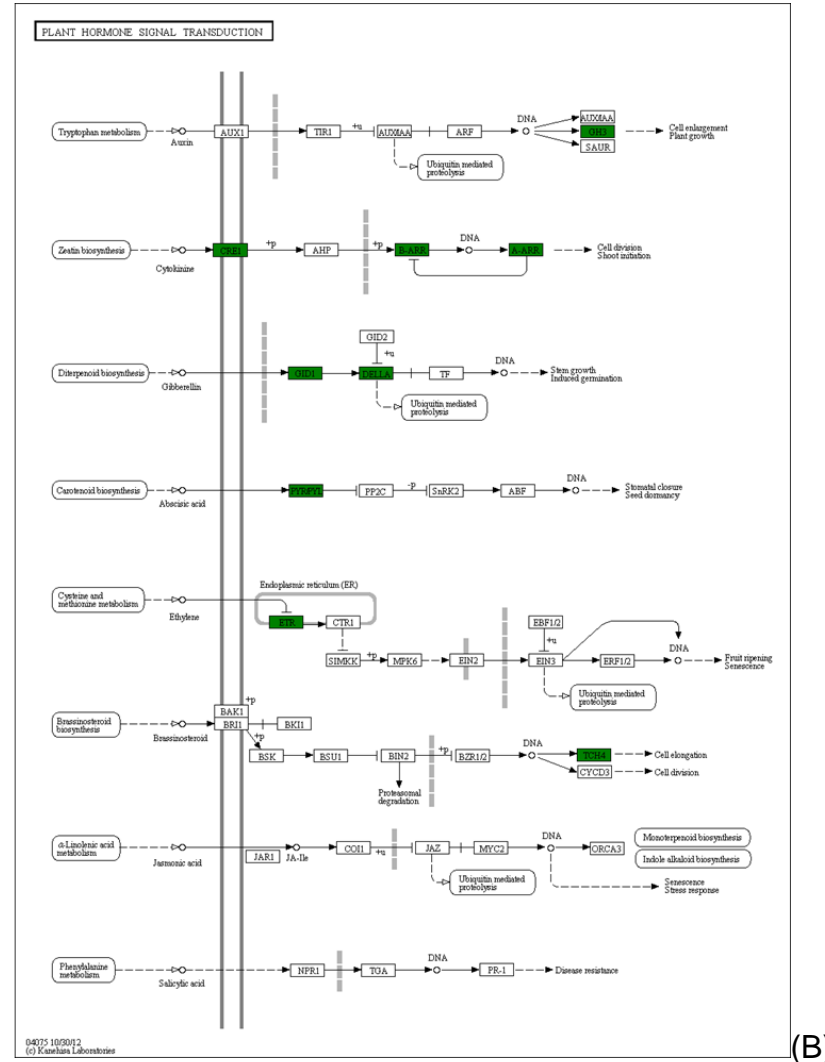
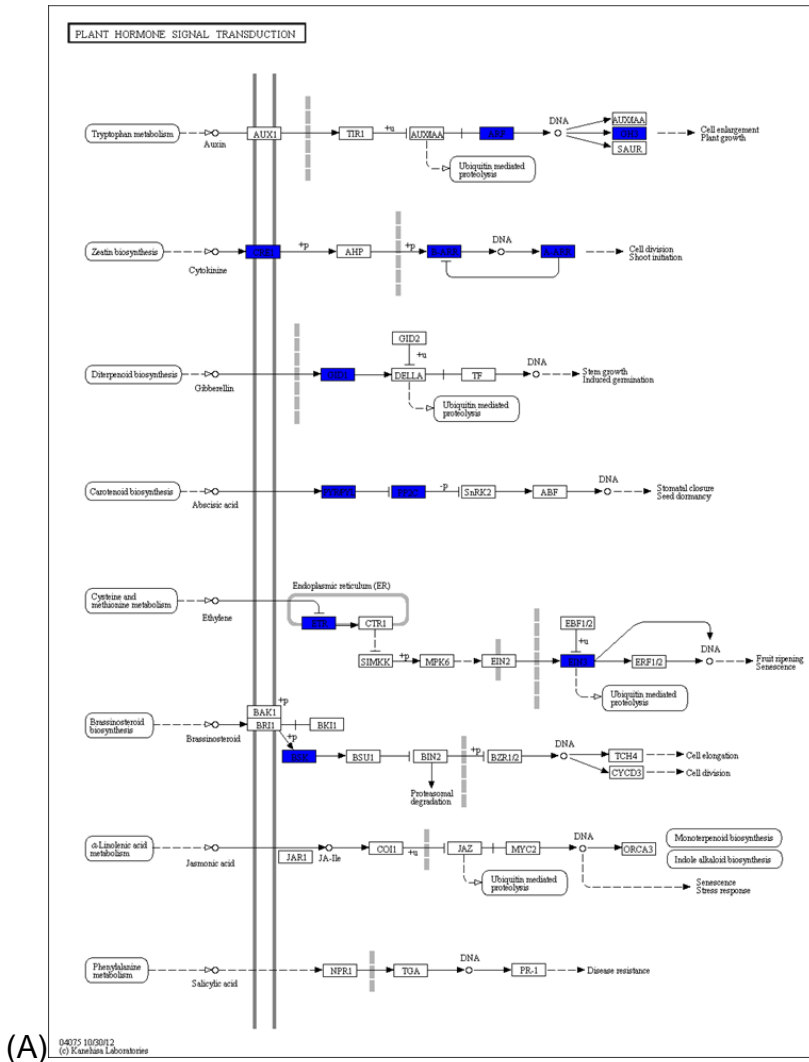
## 5. Conclusão e perspectivas

O presente trabalho apresentou uma análise geral do genoma de *A. amazonense*, identificando e caracterizando as classes metabólicas que se destacaram nas análises. Devido à recente caracterização de outros genomas do mesmo gênero, comparações mostrando proteínas conservadas e diferenciais entre as espécies foram apresentados. Proteínas possivelmente relacionadas com o metabolismo de hormônios vegetais foram identificadas e caracterizadas, assim como do sistema de secreção do tipo IV, relacionadas a elementos genéticos móveis e sistema imune bacteriano.

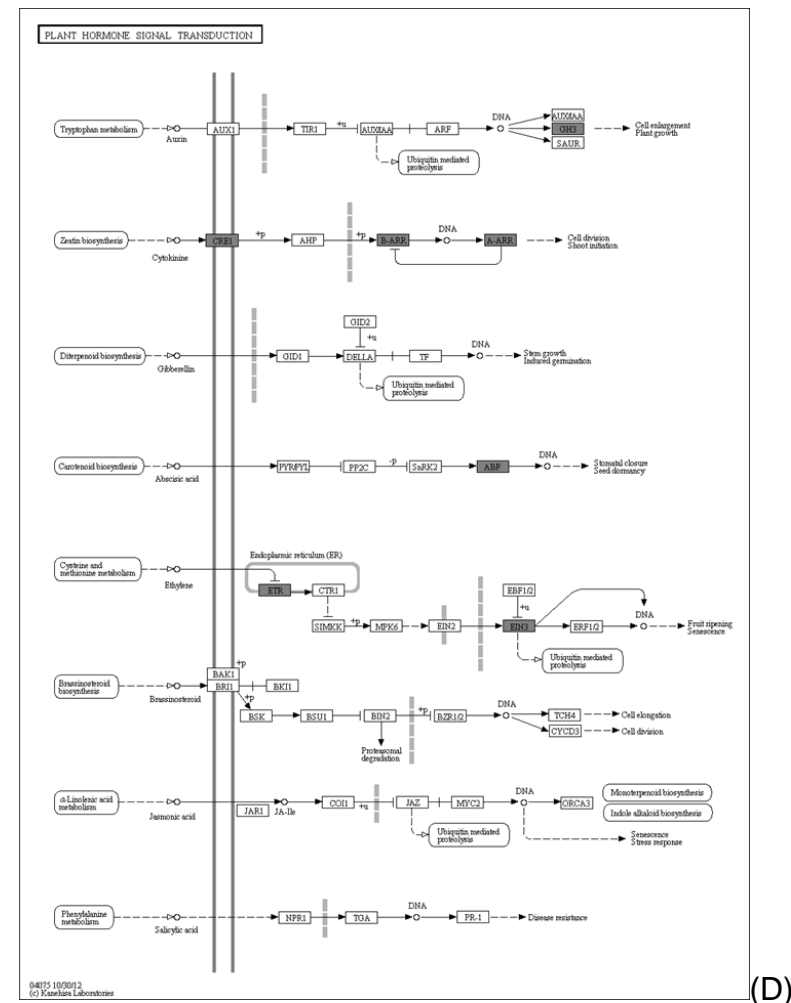
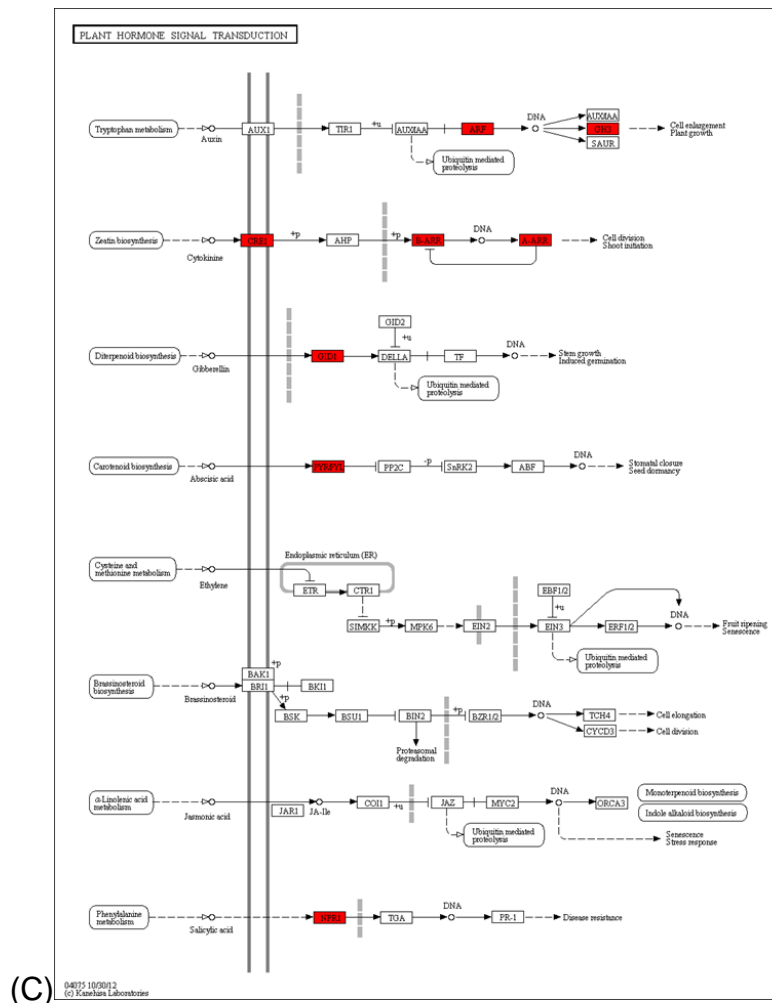
A diferenciação dos genomas é de fundamental importância para a otimização da utilização de espécies de *Azospirillum* como inoculantes. Além disso, a compreensão das diferentes capacidades genéticas de cada espécie contribui de forma significativa para o início de novas utilizações em termos biotecnológicos.

A identificação e caracterização de proteínas possivelmente relacionadas ao metabolismo de hormônios e troca de material genético em *A. amazonense* proporciona novas possibilidades de análise para o entendimento da interação desse microrganismo e os vegetais associados. Consequentemente, contribui para a obtenção de melhores resultados na utilização biotecnológica dessa espécie como promotora do crescimento vegetal.

## 6. Material Suplementar

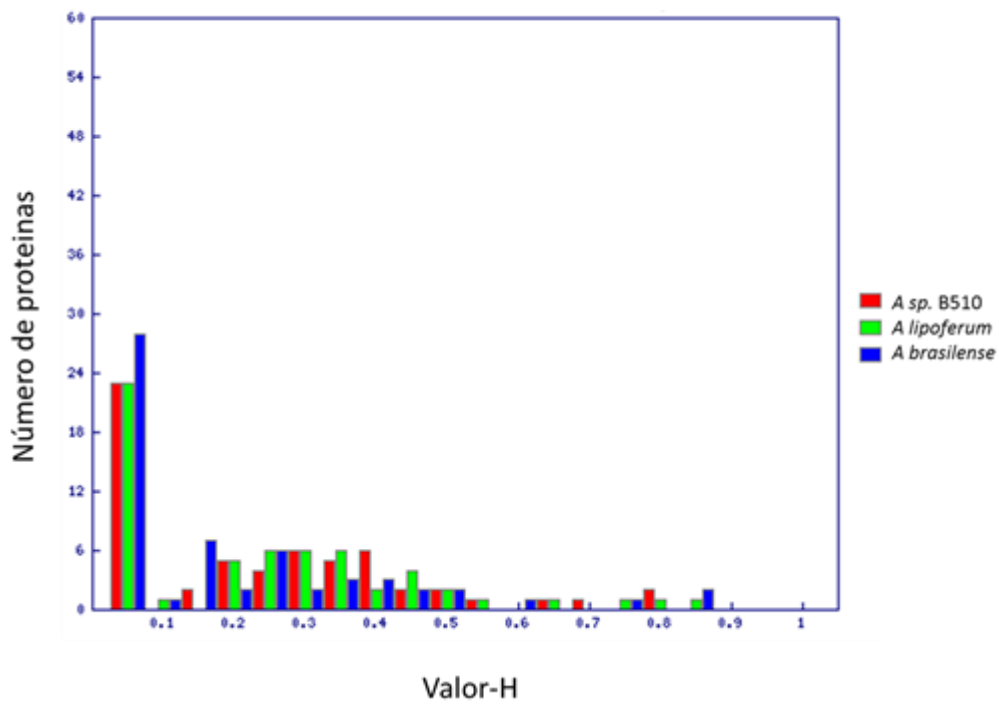






**Figura Suplementar 1** – Proteínas relacionadas com a sinalização de hormônios vegetais em quatro espécies do gênero *Azospirillum* identificadas pelo KAAS. (A) *A. brasilense*; (B) *A. lipoferum*; (C) *Azospirillum* sp. B510 (D) *A. Amazonense* (Ko04075).

Distribuição do Valor-H para as 60 sequencias de proteínas de *A. amazonense* relacionadas ao Sistema de Secreção Tipo IV

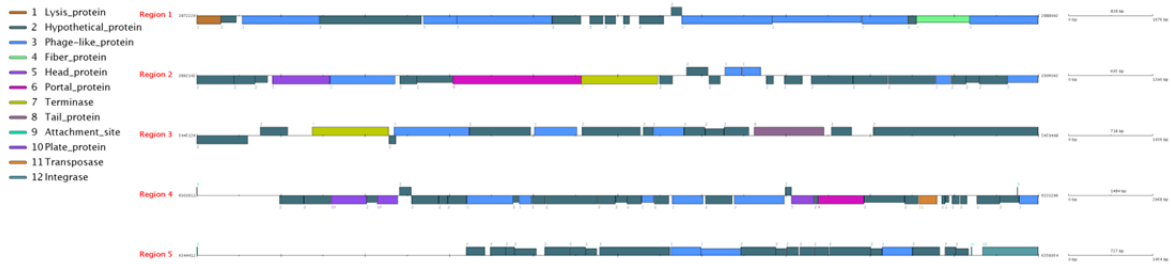


**Figura Suplementar 2** – Identificação das proteínas do Sistema de Secreção Tipo IV em *A. amazonense* não compartilhadas com outros membros do gênero.

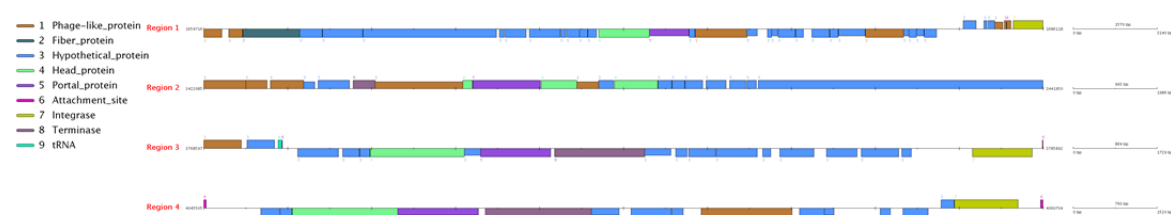
### A. B510



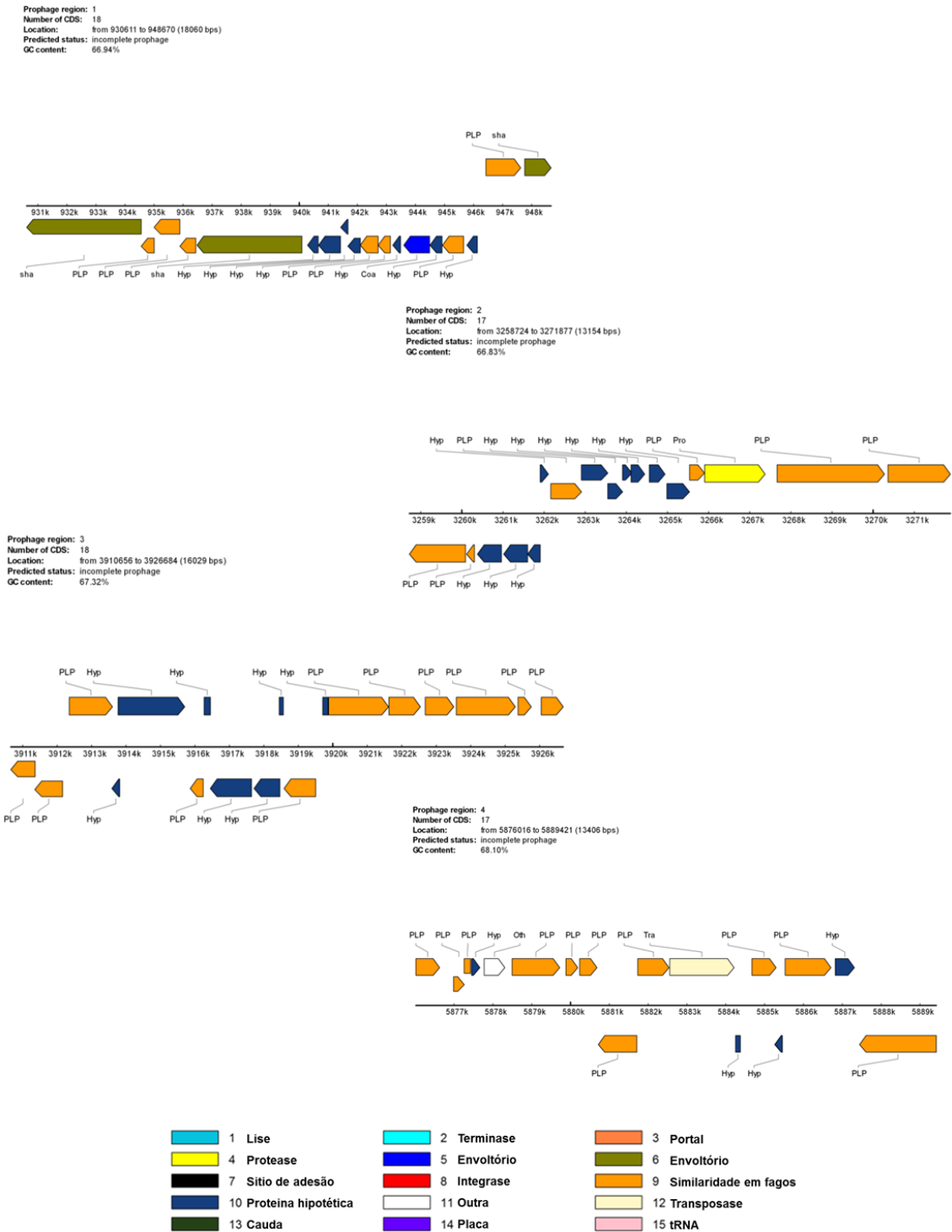
### A. brasilense



### A. lipoferum



**Figura Suplementar 3** - Regiões genômicas de três espécies de *Azospirillum* com similaridade em genomas virais detectadas pela ferramenta PHAST (*Phage Search Tool*).



**Figura Suplementar 4** – Regiões do genoma de *A. amazonense* com similaridade em genomas virais detectadas pela ferramenta PHAST (*Phage Search Tool*).

glutamine cyclotransferase [Xanthomonas oryzae pv. oryzicola BLS256]

Sequence ID: [ref|YP\\_005628993.1|](#) Length: 272 Number of Matches: 1

[▶ See 2 more title\(s\)](#)

Range 1: 21 to 271 [GenPept](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
284 bits(726) 4e-92 Compositional matrix adjust. 149/254(59%) 175/254(68%) 4/254(1%)					
Query 10	VWALLLALALAVLPFTARAQEA	PRKYGVEVVRTYPHDPQAFTEGLFYLDGFLYESTGLE			69
Sbjct 21	VYSLLFSAAL---	LLPSLAHCGETIPTQGYTVVRTYPHDTAAAFTEGLFYLDGHLYESTGER			77
Query 70	QHSTIRKVDLATGTVLQ	SRELDPRYFGEIVNWKDRLVQLTYRTEVGFVYGLGDFKARKE			129
Sbjct 78	GQSSVRKVDLDTG	EVLQQANTPPPFYGEIVAWKDRLIQLTWRNQRGFVYDLATLAPRTQ			137
Query 130	FHYRGEGWALIQDGHRLIMSDGTP	ELRFLDPETLAETGRITVTDNGYPIRNLNELEFVKG			189
Sbjct 138	FSYSGEGWALTSDDRQLYMSDGTASIRRLDPQSLKQIGSIKVTTRGKPLDNLNELEWVKG				197
Query 190	EILANIWQTNLIARINPATGAVTGWIDLTPL-EALSGRIQNIDNVLNGIAYDAKGDRLFV				248
Sbjct 198	ELLANVWLTTRIARIDPATGKVIWIDLKALVPDPDTLTDPTNDVLNGIAYDAEHDRLFV				257
Query 249	TGKRWPKLFEIKLV	262			
Sbjct 258	TGKRWPKIYEIKLA	271			

**Figura Suplementar 5** - Alinhamento da Glutamina Ciclotransferase de *A. amazonense* identificada por RDA.

**Tabela Suplementar 1 – Proteínas conservadas entre as quatro espécies de *Azospirillum* analisadas.**

ID	Acesso	Produto proteico	COG	KEGG	Mapa	Essencial	CDD	PFAM
Amz_1907	ZP_08868786.1	30S ribosomal protein S1	COG0539	K02945	ko03010	DEG10200437		PF00575
Amz_3932		30S ribosomal protein S10	COG0051	K02946	ko03010	DEG10200134	cl00314	PF00338
Amz_3909	ZP_08869964.1	30S ribosomal protein S11_cl00332	COG0100	K02948	ko03010	DEG10200157	cl00332	PF00411
Amz_4863	ZP_08870544.1	30S ribosomal protein S12_cl00312	COG0048	K02950	ko03010	DEG10200383	cl00312	PF00164
Amz_3910	ZP_08869965.1	30S ribosomal protein S13	COG0099	K02952	ko03010	DEG10200156		PF00416
Amz_3927	ZP_08869973.1	30S ribosomal protein S19_cl00350	COG0185	K02965	ko03010	DEG10200139	cl00350	PF00203
Amz_0323		30S ribosomal protein S21		K02970	ko03010	DEG10200403		PF01165
Amz_3925		30S ribosomal protein S3	COG0092	K02982	ko03010	DEG10200141		
Amz_2387		30S ribosomal protein S4	COG0522	K02986	ko03010	DEG10200336		
Amz_4864	ZP_08870545.1	30S ribosomal protein S7	COG0049	K02992	ko03010	DEG10200382	cl00313	PF00177
Amz_3922	ZP_08869971.1	30S ribosomal subunit protein S17	COG0186	K02961	ko03010	DEG10200144	cl00351	PF00366
Amz_0857	ZP_08868172.1	30S ribosomal subunit protein S2_cl00315	COG0052	K02967	ko03010	DEG10200263	cl00315	PF00318
Amz_4852	ZP_08870538.1	3isopropylmalate dehydrogenase_Ath117	COG0473	K00052	ko00290	DEG10130080	cl00445	PF00180
Amz_4350	ZP_08870242.1	3octaprenyl4hydroxybenzoate decarboxylase	COG0043	K03182	ko00130	DEG10190267	cl00311	
Amz_1421	ZP_08868514.1	4hydroxy3methylbut2en1yl diphosphate synthase	COG0821	K03526	ko00900	DEG10200113		PF04551
Amz_0006		50S ribosomal protein L11	COG0080	K02867	ko03010	DEG10200088		
Amz_3924		50S ribosomal protein L16_cl00353	COG0197	K02878	ko03010	DEG10200142	cl00353	PF00252
Amz_3928	ZP_08869974.1	50S ribosomal protein L2	COG0090	K02886	ko03010	DEG10200138		PF03947
Amz_3829		50S ribosomal protein L31_cl00377	COG0254	K02909	ko03010	DEG10200400	cl00377	PF01197
Amz_1997	ZP_08868842.1	50S ribosomal protein L33	COG0267	K02913	ko03010	DEG10200334	cl00383	PF00471
Amz_2014		50S ribosomal protein L36		K02919	ko03010	DEG10200405		PF00444
Amz_3930		50S ribosomal protein L4	COG0088	K02926	ko03010	DEG10200136	cl00325	PF00573
Amz_0009		50S ribosomal protein L7	COG0222	K02935	ko03010	DEG10200083	cl00364	
Amz_3907		50S ribosomal subunit protein L17	COG0203	K02879	ko03010	DEG10200159	cl00356	PF01196
Amz_4013	ZP_08870018.1	AAA ATPase	COG0714	K00540		DEG20010702		PF00004
Amz_4009	ZP_08870013.1	ABC transporter ATPbinding protein	COG0488	K06020		DEG10100399		PF00005

Amz_1437	ZP_08868522.1	ABC transporter ATPbinding protein _cl09099	COG1137	K06861	ko02010	DEG10200440	cl09099	PF00005
Amz_2711	ZP_08868522.1	ABC transporter involved in FeS cluster assembly SufC ATPbinding protein	COG0396	K09013		DEG10200234		
Amz_3635	ZP_08869805.1	acetylCoA carboxylase biotin carboxylase subunit_Ath085	COG4770	K01961	ko00061	DEG10200244		PF02786
Amz_3810	ZP_08869899.1	acetylcoenzyme A synthetase_Ath038	COG0365	K01895	ko00010	DEG20010712		PF00501
Amz_1470	ZP_08868536.1	acyl carrier protein	COG0236	K02078		DEG10200211	cl09936	PF00550
Amz_3110	ZP_08869478.1	adenosylcobalamindependent methylmalonylCoA mutase large subunit MCMalpha	COG1884	K01847	ko00280	DEG20051827		PF01642
Amz_2193	ZP_08868952.1	adenosylhomocysteinase_Ath078	COG0499	K01251	ko00270	DEG20020012	cl09931	PF05221
Amz_5225	ZP_08870774.1	adenylosuccinate lyase	COG0015	K01756	ko00230	DEG10020220		
Amz_4243	ZP_08870171.1	argininosuccinate synthase_Ath094	COG0137	K01940	ko00250	DEG20052059	cl00292	PF00764
Amz_3022	ZP_08869022.1	Aspartate aminotransferase A	COG0436	K00812	ko00250	DEG10100035		PF00155
Amz_1442	ZP_08868524.1	aspartate kinase	COG0527	K00928	ko00260	DEG10200111		PF00696
Amz_0300	ZP_08867806.1	ATP synthase F1 sector beta subunit_Ath023	COG0055	K02112	ko00190	DEG10200416		PF00006
Amz_4524	ZP_08870354.1	ATPase and chaperone component of the HslUV proteasome	COG1220	K03667		DEG10070192		PF07724
Amz_0976	ZP_08868241.1	ATPdependent Clp protease ATPbinding subunit ClpA_Ath073	COG0542	K03694		DEG10110149		
Amz_4523	ZP_08870353.1	ATP-dependent protease, component of the	COG5405	K01419		DEG10100315	cl00467	PF00227
Amz_4957	ZP_08870354.1	ATPdependent protease ATPbinding subunit ClpX_Ath116	COG1219	K03544	ko04112	DEG10110034		PF07724
Amz_4959	ZP_08870607.1	ATPdependent protease La_Ath070	COG0466	K01338	ko04112	DEG10160164		
Amz_2273		CarD-like transcriptional regulator prote	COG1329	K07736		DEG10200090	cl00588	PF02559
Amz_4512	ZP_08870348.1	chaperone protein heat shock protein Hsp40	COG0484	K03686		DEG10210187		
Amz_3713	ZP_08869853.1	chaperonin GroEL_Ath037	COG0459	K04077	ko03018	DEG10200093	cl02777	PF00118
Amz_3712		chaperonin GroS	COG0234	K04078		DEG10200094	cl09113	PF00166
Amz_1860	ZP_08868761.1	chemotaxis methyltransferase CheR	COG1352	K00575	ko02020	DEG10180319		
Amz_2605	ZP_08869189.1	chemotaxis protein CheA	COG0643	K03407	ko02020	DEG10110115		PF01584
Amz_2607	ZP_08869191.1	Chemotaxis response regulator CheY	COG0745	K03413	ko02020	DEG10150219		PF00072
Amz_1720		cobaltochelataase CobS subunit	COG0714	K09882	ko00860	DEG10200370		
Amz_0169		conserved protein of unknown function putative coiledcoil domain	COG0718	K09747		DEG10030224	cl00494	
Amz_1897	ZP_08868781.1	CTP synthetase_Ath082	COG0504	K01937	ko00240	DEG10050392		PF06418
Amz_2710	ZP_08869240.1	cysteine desulfurase activator complex subunit SufB_Ath052	COG0719	K09014		DEG10200235	cl03223	PF01458

Amz_1895	ZP_08868778.1	cytochrome c oxidase subunit I	COG0843	K02274	ko00190	DEG10100496	cl00275	PF00115
Amz_4756	ZP_08870493.1	dihydroxyacid dehydratase	COG0129	K01687	ko00290	DEG10100014	cl00340	PF00920
Amz_3908		DNA-directed RNA polymerase subunit alpha	COG0202	K03040	ko00230	DEG10200158		
Amz_0011	ZP_08867626.1	DNA-directed RNA polymerase subunit beta	COG0085	K03043	ko00230	DEG10200084		
Amz_0012	ZP_08867627.1	DNA-directed RNA polymerase subunit beta	COG0086	K03046	ko00230	DEG10200085		PF04997
Amz_2758	ZP_08870646.1	electron transfer flavoprotein alpha subunit	COG2025	K03522		DEG20070111		
Amz_2085	ZP_08868887.1	electron transfer flavoprotein beta subunit	COG2086	K03521		DEG10150133	cl00292	
Amz_3456	ZP_08869702.1	electron transfer flavoprotein beta subunit	COG2086	K03521		DEG10150133	cl00292	
Amz_1825	ZP_08868743.1	enoylacyl carrier protein reductase	COG0623	K00208	ko00061	DEG10200465		PF00106
Amz_1348	ZP_08868471.1	excinuclease ABC subunit A	COG0178	K03701	ko03420	DEG10020067		PF00005
Amz_0482	ZP_08867950.1	flagellar motor switch protein FliM	COG1868	K02416	ko02030	DEG10130333		PF02154
Amz_4367	ZP_08870247.1	flagellar motor switch protein FliN	COG1886	K02417	ko02030	DEG10030408	cl00819	
Amz_0503	ZP_08867964.1	flagellar motor switch protein G	COG1536	K02410	ko02030	DEG20010582		
Amz_0298	ZP_08867804.1	F-type H <sup>+</sup> -transporting ATPase subunit alpha_Ath022	COG0056	K02111	ko00190	DEG10200418		PF00006
Amz_5282	ZP_08870812.1	fumarate hydratase class I	COG1951	K01676	ko00020	DEG10200418		PF05681
Amz_4727	ZP_08870474.1	glutamine synthetase type I	COG0174	K01915	ko00250	DEG10200286		PF00120
Amz_0832		glyceraldehyde 3-phosphate dehydrogenase	COG0057	K00134	ko00010	DEG10100232		
Amz_1501	ZP_08868558.1	GMP synthase_Ath081	COG0519	K01951	ko00230	DEG10130018		PF00958
Amz_4269	ZP_08870187.1	GTP-binding protein_Ath039	COG0481	K03596		DEG10020174		
Amz_4823	ZP_08870525.1	GTP-binding protein TypA_Ath055	COG1217	K06207		DEG10020099		PF00009
Amz_5361	ZP_08868845.1	GTP-binding protein YchF_Ath124	COG0012	K06942		DEG10180201		
Amz_0763		hfq_cl00259	COG1923	K03666	ko03018	DEG10200226	cl00259	
Amz_2373		HNH endonuclease	COG1403	K01157		DEG10030470	cl00083	PF01844
Amz_3193	ZP_08869003.1	homoserine O-acetyltransferase	COG2021	K00641	ko00270	DEG10130091	cl17114	
Amz_4518		imidazole glycerol phosphate synthase cyclase subunit_cl09108	COG0107	K02500	ko00340	DEG10130468	cl09108	PF00977
Amz_4522		imidazole glycerol-phosphate dehydratase	COG0131	K01693	ko00340	DEG10130465		
Amz_3372		inosine 5-monophosphate dehydrogenase	COG0516	K00088	ko00230	DEG10130475		PF00478
Amz_3079		integration host factor subunit alpha	COG0776	K04764		DEG10030251	cl00257	PF00216



Amz_1215	ZP_08868386.1	isocitrate dehydrogenase_Ath065	COG0538	K00031	ko00020	DEG10200337	cl00445	PF00180
Amz_4869		ketolacid reductoisomerase	COG0059	K00053	ko00290	DEG10200308		PF07991
Amz_0766	ZP_08868121.1	large subunit of chaperonin GroESL_Ath027	COG0459	K04077	ko03018	DEG10200093	cl02777	PF00118
Amz_1307	ZP_08868447.1	large subunit of chaperonin GroESL_Ath029	COG0459	K04077	ko03018	DEG10200093	cl02777	PF00118
Amz_4511	ZP_08870347.1	molecular chaperone DnaK_Ath042	COG0443	K04043	ko03018	DEG10200006		
Amz_2624		NADH dehydrogenase subunit I	COG1143	K00338	ko00190	DEG10200275		
Amz_4968	ZP_08870612.1	NADH-quinone oxidoreductase chain B	COG0377	K00331	ko00190	DEG10120380		PF01058
Amz_4972		NADHquinone oxidoreductase subunit F_Ath046	COG1894	K00335	ko00190	DEG10200278		PF01512
Amz_0758	ZP_08868115.1	Nitrogen assimilation regulatory protein NtrC	COG2204	K07712	ko02020	DEG10030266		PF00158
Amz_0760	ZP_08868118.1	Nitrogen assimilation regulatory protein NtrX	COG2204	K13599	ko02020	DEG10030266		PF00158
Amz_4726	ZP_08870473.1	nitrogen regulatory protein P-II	COG0347	K04751	ko02020	DEG10050121		PF00543
Amz_1684		nucleoside diphosphate kinase_cl00335	COG0105	K00940	ko00230	DEG10200220	cl00335	PF00334
Amz_1486	ZP_08868545.1	phosphate regulon transcriptional regulatory protein PhoB	COG0745	K07657	ko02020	DEG10030155		PF00072
Amz_5377		phosphoribosylaminoimidazolesuccinocarboxamide synthetase	COG0152	K01923	ko00230	DEG10130494		PF01259
Amz_5096		phosphoserine phosphatase	COG0560	K01079	ko00260	DEG10100495		
Amz_1141	ZP_08868338.1	polynucleotide phosphorylasepolyadenylase_Ath067	COG1185	K00962	ko00230	DEG10200007		PF01138
Amz_4567	ZP_08870379.1	precorrin8X methylmutase	COG2082	K06042	ko00860	DEG10190132	cl00913	PF02570
Amz_0039	ZP_08867636.1	prkA	COG2766	K07180		DEG10060202		
Amz_3387	ZP_08869659.1	prolyltRNA synthetase	COG0442	K01881	ko00970	DEG10200267		PF00587
Amz_3108	ZP_08869476.1	propionylCoA carboxylase beta chain	COG4799	K01966	ko00280	DEG10100598		PF01039
Amz_5390		pyrimidine (deoxy)nucleoside triphosphate	COG1051	K03574		DEG10170343	cl00447	PF00293
Amz_1903	ZP_08868783.1	pyruvate dehydrogenase E1 component subunit alpha	COG1071	K00161	ko00010	DEG20060010		PF00676
Amz_2663	ZP_08869218.1	recombinase A_Ath132	COG0468	K03553	ko03440	DEG10160189		PF00154
Amz_1927		ribosephosphate pyrophosphokinase	COG0462	K00948	ko00030	DEG10200079		
Amz_3921	ZP_08869970.1	Ribosomal protein L14	COG0093	K02874	ko03010	DEG10200145	cl00328	PF00238
Amz_0488	ZP_08867954.1	RNA polymerase sigma factor RpoD	COG0568	K03086	ko03020	DEG10200372		PF04546
Amz_1553	ZP_08868587.1	RNA polymerase sigma32 factor	COG0568	K03089	ko03020	DEG10200376		PF04542
Amz_2233	ZP_08868975.1	RNA polymerase-binding protein DksA (DnaK	COG1734	K06204		DEG10030123	cl00755	PF01258

Amz_4872		rod shapedetermining protein MreB	COG1077	K03569		DEG10200186		PF06723
Amz_0625		rRNA methylase	COG1189	K06442		DEG10080196		
Amz_1341		succinate dehydrogenase ironsulfur protein	COG0479	K00240	ko00020	DEG10200428		
Amz_5095	ZP_08870696.1	succinylCoA synthetase alpha subunitNADPbinding_Ath108	COG0074	K01902	ko00020	DEG10200052		
Amz_5094	ZP_08870695.1	succinylCoA synthetase beta chain_Ath122	COG0045	K01903	ko00020	DEG10200051		
Amz_0005	ZP_08867625.1	transcription antitermination protein Nus	COG0250	K02601		DEG10200385		PF02357
Amz_2810	ZP_08869298.1	transcription elongation factor NusA	COG0195	K02600		DEG10200010		PF08529
Amz_0280	ZP_08867793.1	transcription termination factor Rho	COG1158	K03628	ko03018	DEG10200477		PF00006
Amz_0092	ZP_08867666.1	translation elongation factor G FusA_Ath021	COG0480	K02355		DEG10200381		PF00009
Amz_0877	ZP_08868186.1	translation elongation factor P efp	COG0231	K02356		DEG10200101		
Amz_0755	ZP_08868113.1	translation initiation factor 2 InfB_Ath087	COG0532	K02519		DEG10200009		
Amz_2818		translation initiation factor IF-3	COG0290	K02520		DEG10200124		PF00707
Amz_2451	ZP_08869106.1	tryptophan synthase subunit beta_Ath080	COG0133	K01696	ko00260	DEG10130102	cl00342	PF00291
Amz_5269	ZP_08870801.1	twocomponent cell cycle transcriptional regulator ctrA	COG0745	K13584	ko02020	DEG10200371		PF00072
Amz_0701	ZP_08868115.1	two-component transcriptional regulator,	COG4567	K15012		DEG10130462		PF00072
Amz_3364	ZP_08869641.1	urease, gamma subunit	COG0831	K01430	ko00230	DEG10020280	cl00532	PF00547

A similaridade foi calculada com o programa BioEdit em uma análise realizada localmente.

**Tabela Suplementar 2 – Proteínas com função conhecida exclusivas de *A. amazonense* em relação a *Azospirillum*.**

ID	Acesso	Produto proteico	COG	MAPA	CDD	PFAM	TIGR
Amz_1565		2OG-Fe(II)_oxygenase	COG3145		cl15773		
Amz_0467	ZP_08867938.1	9cisepoxycarotenoid_dioxygenase	COG3670		cl10080		
Amz_3370		acetoacetate_decarboxylase	COG4689	ko00072	cl01919	PF06314	
Amz_0836	ZP_08868157.1	acid_phosphatase	COG3511	ko00627	cl15450	PF04185	TIGR03397
Amz_4008		acid_phosphatase		ko00562	cl11399		
Amz_4811	ZP_08870519.1	acid_sphingomyelinase-like_phosp			cl13995	PF00149	
Amz_2865		acyl_carrier_protein	COG0236			PF00550	
Amz_5111	ZP_08870709.1	acyl_carrier_protein	COG0236			PF00550	
Amz_2469	ZP_08869117.1	alkaline_phosphatase_D	COG3540	ko00627			
Amz_1050	ZP_08868281.1	alpha12mannosidase	COG3537		cl15460		TIGR01180
Amz_2427	ZP_08869087.1	alpha12mannosidase	COG3537		cl15460		TIGR01180
Amz_0684	ZP_08868080.1	alpha-galactosidase	COG3345	ko00052	cl11402	PF02065	
Amz_1852	ZP_08868755.1	Alpha-galactosidase_Ath020__K07407__cl11402	COG3345	ko00052	cl11402	PF02065	
Amz_0339	ZP_08867832.1	alphaglucoosidase		ko00052	cl11195		
Amz_4165		alpha-ketoglutarate-dependent_dioxygenase__cl15773	COG3145		cl15773		
Amz_0685	ZP_08868081.1	alphaLarabinofuranosidase	COG3534	ko00520		PF06964	
Amz_0493	ZP_08867958.1	alpha-l-fucosidase	COG3669	ko00511	cl12119	PF01120	
Amz_1113	ZP_08868403.1	alpha-L-fucosidase	COG3669	ko00511	cl12119	PF01120	
Amz_4946	ZP_08870602.1	alpha-l-fucosidase	COG3669	ko00511	cl12119	PF01120	
Amz_0961	ZP_08868229.1	alpha-L-rhamnosidase			cl01801		
Amz_1992	ZP_08868839.1	alphaNarabinofuranosidase__COG3534	COG3534	ko00520		PF06964	
Amz_2978		amicyanin	COG3794	ko00910		PF00127	
Amz_5215		amidinotransferase	COG1834				
Amz_1135		amidohydrolase					
Amz_1148		amidohydrolase					
Amz_0646	ZP_08868057.1	amidohydrolase_2				PF04909	

Amz_5476		amino__K03293	COG1113			PF00324	
Amz_0346		amino_acid_permease-associated_protein	COG1113			PF00324	
Amz_4296	ZP_08870208.1	amino_acidpeptide_transporter_protein__COG3104	COG3104			PF00854	TIGR00924
Amz_4410	ZP_08870289.1	aminopeptidase_ES-62	COG2234	ko00310	cl14876	PF04389	
Amz_1793		ankyrin					
Amz_1162	ZP_08868346.1	arylsulfatase_regulator_(Fe-S_oxidoreducta	COG0641				
Amz_5316	ZP_08870830.1	AsnC_family_transcriptional_regulator					
Amz_2741		ATPase					
Amz_1165	ZP_08868347.1	ATPase_BadF/BadG/BcrA/BcrD_type	COG2971	ko00520			
Amz_2930		bacillolysin	COG3227		cl14813	PF02868	
Amz_0270	ZP_08867784.1	beta-galactosidase	COG1874	ko00052	cl03154	PF02449	
Amz_3276	ZP_08869580.1	betaglucosidase	COG1472	ko00460		PF01915	
Amz_2965	ZP_08869393.1	betaglucosidase_Ath101	COG1472	ko00460		PF00933	
Amz_3225	ZP_08869550.1	betaglucosidase__K01188		ko00460		PF01915	
Amz_0491	ZP_08867956.1	betaNacetylhexosaminidase	COG3525	ko00511	cl02948	PF00728	
Amz_2909	ZP_08869355.1	beta-xylosidase	COG3664	ko00500	cl15381	PF01229	
Amz_5344		caa(3)-type_oxidase,_subunit_IV					TIGR02229
Amz_1839		carbohydrate-selective_porin_OprB	COG3659			PF04966	
Amz_5436	ZP_08870891.1	Carbohydrate-selective_porin_OprB	COG3659			PF04966	
Amz_5437	ZP_08870942.1	Carbohydrate-selective_porin_OprB	COG3659			PF04966	
Amz_5504	ZP_08870942.1	Carbohydrate-selective_porin_OprB	COG3659			PF04966	
Amz_5366	ZP_08870855.1	cell_division_protein	COG0645				
Amz_2453	ZP_08869108.1	chaperone_protein_HtpG			cl03439		
Amz_3125		chemotaxis_protein_CheW_CheW2	COG0835	ko02020		PF01584	
Amz_1003	ZP_08868258.1	chlorogenate_esterase	COG2267		cl06521		
Amz_4655		chorismate_mutase	COG1605	ko00400	cl00693	PF01817	TIGR01806
Amz_1031		ChrB_protein	COG4275		cl01744		
Amz_2501	ZP_08869136.1	COG0640:_Predicted_transcriptional_regulat	COG4189			PF01022	

Amz_1460		conjugal_transfer_protein	COG3838	ko03070			
Amz_2466		conjugal_transfer_protein_TrbC	COG3838	ko03070			
Amz_2242	ZP_08868549.1	conjugal_transfer_protein_TrbF	COG3701	ko03070	cl01500	PF04335	
Amz_4148	ZP_08870114.1	conjugal_transfer_protein_TrbF__cl01500	COG3701	ko03070	cl01500	PF04335	
Amz_1491	ZP_08868549.1	conjugal_transfer_protein_TrbF__K03200__cl01500	COG3701	ko03070	cl01500	PF04335	
Amz_5452	ZP_08870901.1	CRISPR-associated_Cas1_family_protein	COG1518		cl00656		
Amz_3762		CRISPR-associated_Csd2_family_protein	COG3649		cl01465		TIGR02589
Amz_5448	ZP_08870898.1	CRISPR-associated_Cse1_family_protein					
Amz_5451		CRISPR-associated_Cse3_family_protein			cl08497		
Amz_5450	ZP_08870899.1	CRISPR-associated_Cse4_family_protein			cl09608		TIGR01869
Amz_1257		cupin	COG0662	ko00051			
Amz_5196	ZP_08870755.1	cupin_2_domain-containing_protein	COG1917				
Amz_3612	ZP_08869790.1	cupin_domain-containing_protein	COG1917	ko00363			
Amz_4843		cyclase,_putative__cl00814	COG1878		cl00814		TIGR03035
Amz_3394		cytochrome_c_biogenesis_protein_transmembrane_region					
Amz_3335		cytochrome_p460					
Amz_0918	ZP_08868207.1	dATP_pyrophosphohydrolase					
Amz_4909		DifB_protein	COG2315				
Amz_0223	ZP_08867754.1	dioxygenase	COG2514		cl14632		
Amz_4141	ZP_08870107.1	DNA_or_RNA_helicase					
Amz_3756	ZP_08869876.1	DNA_replication_protein	COG1484			PF01695	
Amz_0904	ZP_08868196.1	domain_of_unknown_function_DU1801_family_p	COG4430				
Amz_2951		DoxX_family_protein	COG2259				
Amz_3602	ZP_08869786.1	D-tagatose-bisphosphate_aldolase_non-catal	COG4573	ko00052	cl09211	PF08013	TIGR02810
Amz_0054		ECF_sigma_factor_family_protein	COG1595				TIGR02999
Amz_0526	ZP_08867978.1	EF_hand_repeatcontaining_protein		ko04020			
Amz_3345		Ethyl_tert-butyl_ether_degradation_EthD			cl15807		TIGR02118
Amz_1240	ZP_08868403.1	exported_alpha-L-fucosidase_prote	COG3669	ko00511	cl12119	PF01120	

Amz_0884	ZP_08868188.1	exported_hypothetical_protein					
Amz_0885	ZP_08868188.1	exported_hypothetical_protein					
Amz_1071	ZP_08868295.1	fatty_acid_desaturase	COG1398	ko01040	cl00615	PF00487	
Amz_0917		fatty_acid_hydroxylase	COG3000				
Amz_1160	ZP_08868344.1	Fe-S_metabolism_associated_SufE	COG2166		cl00951		
Amz_4006	ZP_08870011.1	FIG028593:_membrane_protein	COG4539		cl01879		
Amz_1178	ZP_08868353.1	fimbrial_usher_protein	COG3188			PF00577	
Amz_2866	ZP_08869332.1	FkbH-like_protein	COG3882				TIGR01686
Amz_0060		fkbptype_peptidylprolyl_cistrans_isomerase_FkpA__COG0545	COG0545			PF00254	
Amz_3808		Flavin_reductase					
Amz_3098	ZP_08869470.1	flavin_reductase_domain-containing_protei	COG1853		cl00801		
Amz_0508	ZP_08867852.1	fructokinase	COG1940				
Amz_5027		fructosamine-3-kinase	COG3001		cl09925		
Amz_3189	ZP_08869526.1	fumarylacetoacetate_FAA_hydrolase__COG3802	COG3802	ko00040		PF01557	
Amz_4474	ZP_08867693.1	general_secretion_pathway_protein_F	COG1459	ko03070			
Amz_0146	ZP_08867694.1	general_secretion_pathway_protein_G	COG2165	ko03070	cl07082		TIGR01710
Amz_0619	ZP_08868036.1	general_secretion_pathway_protein_G	COG2165	ko03070			TIGR01710
Amz_4475	ZP_08870326.1	general_secretion_pathway_protein_G	COG2165	ko03070	cl07082		TIGR01710
Amz_3989		GlcG_protein	COG3193		cl01249		
Amz_0038		globin	COG2346		cl00280	PF01152	
Amz_0547	ZP_08867986.1	globin	COG2346		cl00280	PF01152	
Amz_0338	ZP_08867831.1	glucan_14alphaglucosidase	COG3387	ko00500			TIGR01535
Amz_1689		glucosyltransferase					
Amz_0687		glutaminase__cl00907	COG2066	ko00250	cl00907	PF04960	TIGR03814
Amz_1815		glutaminase__cl00907	COG2066	ko00250	cl00907	PF04960	TIGR03814
Amz_5170	ZP_08870736.1	glutamine_cyclotransferase	COG3823				
Amz_2574	ZP_08869176.1	glycoside_hydrolase_clan_GH-D	COG3345	ko00052		PF02065	
Amz_2685	ZP_08869229.1	glycoside_hydrolase_family_35	COG1874	ko00052		PF01301	

Amz_2974	ZP_08869400.1	glycoside_hydrolase_family_protein	COG1874	ko00052		PF02449
Amz_0269	ZP_08867783.1	glycosyl_hydrolase_family_protein	COG3669	ko00511	cl12119	PF01120
Amz_1087	ZP_08868305.1	glycosyl_hydrolase_family_protein	COG1472	ko00460		PF00933
Amz_2499	ZP_08869134.1	glycosyl_hydrolase_family_protein	COG1472	ko00460		PF01915
Amz_2587	ZP_08869180.1	glycosyl_hydrolase_family_protein	COG1472	ko00460		PF00933
Amz_2910	ZP_08869356.1	glycosyl_hydrolase_family_protein	COG1472	ko00460		PF01915
Amz_5402	ZP_08870874.1	glycosyl_hydrolase_family_protein		ko00500	cl09101	PF02927
Amz_4856		Glycosyl_transferase	COG0438		cl10013	
Amz_1996	ZP_08868841.1	glycosyl_transferase_family_protein	COG3306		cl01298	PF01755
Amz_1847	ZP_08867639.1	glyoxalase/bleomycin_resistance_protein/d	COG2764		cl14632	
Amz_4043	ZP_08870034.1	Glyoxalase/bleomycin_resistance_protein/d	COG0346	ko00620	cl14632	
Amz_5315		glyoxalase/bleomycin_resistance_protein/d	COG3324	ko00620	cl14632	
Amz_2418		glyoxalase/bleomycin_resistance_protein/di		ko00620		
Amz_2696	ZP_08869232.1	Glyoxalase/bleomycin_resistance_protein/di	COG3324	ko00620	cl14632	
Amz_4498	ZP_08870340.1	glyoxalase/bleomycin_resistance_protein/di	COG0346		cl14632	
Amz_3005	ZP_08869418.1	Haloacid_dehalogenase_domain-containing_p	COG5610			
Amz_1184	ZP_08868359.1	HD_superfamily_phosphohydrolase__COG1078	COG1078			
Amz_2484	ZP_08867986.1	hemoglobin_protein_(Truncated_hem	COG2346		cl00280	PF01152
Amz_2838	ZP_08869314.1	heparan-alpha-glucosaminide_N-acetyltransf	COG4299	ko00531		
Amz_4883	ZP_08870554.1	histidine_acid_phosphatase		ko00562	cl11399	PF00328
Amz_0207	ZP_08867741.1	hydroquinone_dioxygenase_large_subunit				
Amz_0208	ZP_08867742.1	hydroquinone_dioxygenase_small_subunit				
Amz_2436	ZP_08869093.1	Inosine-uridine_nucleoside_N-ribohydrolase	COG1957	ko00230	cl00226	
Amz_0119	ZP_08867682.1	inosine-uridine_preferring_nucleoside_hydr	COG1957	ko00230	cl00226	
Amz_2478	ZP_08869093.1	inosineuridinepreferring_nucleoside_hydrolase__K01239	COG1957	ko00230	cl00226	
Amz_0386	ZP_08867863.1	Inward_rectifier_potassium_channel			cl11596	
Amz_0330	ZP_08867824.1	Ketosteroid_isomerase-like_protein				
Amz_4458		kinase	COG4240			

Amz_4685	ZP_08870450.1	Ku78,_Ku70_and_Ku80_are_70kDa_and_80kDa_s	COG1273	ko03450		PF02735	TIGR02772
Amz_4688	ZP_08870452.1	Ku78,_Ku70_and_Ku80_are_70kDa_and_80kDa_s	COG1273	ko03450		PF02735	TIGR02772
Amz_0693	ZP_08868087.1	Lactoylglutathione_lyase	COG0346	ko00620	cl14632		
Amz_4949	ZP_08870603.1	laminin_G_domaincontaining_protein					
Amz_1513		lipolytic_protein_G-D-S-L			cl01053	PF00657	
Amz_0028	ZP_08867632.1	lipoprotein					
Amz_0256		lipoprotein					
Amz_1414	ZP_08868511.1	lipoprotein					
Amz_2485	ZP_08869126.1	lipoprotein					
Amz_3780	ZP_08869884.1	lipoprotein	COG3187				
Amz_5477	ZP_08870917.1	lipoprotein	COG3975				
Amz_3892	ZP_08869952.1	Lornithine_5monooxygenase__K10531	COG3486				
Amz_4832	ZP_08870529.1	lyase_containing_HEAT-repeat_protein					
Amz_0043		malate_dehydrogenase					
Amz_2551	ZP_08869165.1	Maltose_phosphorylase_/_Trehalose_phosphor	COG1554			PF03632	
Amz_0996		membrane_protease				PF02517	
Amz_1933	ZP_08868797.1	membrane_protein					
Amz_5211	ZP_08870766.1	membrane_protein	COG3174				
Amz_2012	ZP_08868849.1	metallopeptidase_PepO_peptidase_M13_family	COG3590		cl14813	PF05649	
Amz_2013	ZP_08868850.1	metallopeptidase_PepO_peptidase_M13_family	COG3590		cl14813	PF05649	
Amz_3493	ZP_08869726.1	metalloprotease__cl14813	COG3590		cl14813	PF01431	
Amz_2816		methylthioribose_kinase_Ath139	COG4857	ko00270	cl09925		TIGR01767
Amz_3781	ZP_08869885.1	methyltransferase	COG3315		cl01306	PF04072	TIGR00027
Amz_5399	ZP_08870871.1	Mu-like_prophage_FluMu_F_protein	COG2369		cl10072		TIGR01641
Amz_5400	ZP_08870872.1	Mu-like_prophage_FluMu_protein_gp29					
Amz_0790	ZP_08868134.1	Mu-like_prophage_major_head_subunit_gpT__cl01826__ph1R	COG4397		cl01826		
Amz_0788		Mu-like_prophage_protein_gp36__ph1R	COG4387		cl01818		
Amz_0142	ZP_08867692.1	muramoyltetrapeptide_carboxypeptidase	COG1619		cl00700		



Amz_0919	ZP_08868208.1	NAD_dependent_epimerase					
Amz_2455	ZP_08869109.1	NADPHdependent_FMN_reductase	COG0655		cl00438		
Amz_2522	ZP_08869149.1	NCS1_nucleoside_transporter_family_Ath128	COG1953		cl00456	PF02133	
Amz_1027		nickel_resistance_protein	COG1937		cl00846		
Amz_2355		nickel_resistance_protein	COG1937				
Amz_4214	ZP_08870686.1	nickel_uptake_substrate-specific_transmemb	COG5266	ko02010			
Amz_2830	ZP_08869310.1	nuclease_S1			cl16915	PF02265	
Amz_4493		nucleoside_2deoxyribosyltransferase	COG3613	ko00240	cl16918	PF05014	
Amz_3591	ZP_08869779.1	O-methyltransferase	COG3315	ko00340	cl01306	PF04072	
Amz_0525		outer_membrane_autotransporter_barrel_pro					
Amz_0250	ZP_08867777.1	outer_membrane_receptor_signal_peptide_protein	COG1629		cl01155		
Amz_2471	ZP_08868736.1	oxalate_decarboxylase__K01569	COG2140	ko00630		PF00190	TIGR03404
Amz_3563		PadR_family_transcriptional_regulator					
Amz_5058		PadR_family_transcriptional_regulator	COG1695				
Amz_4661	ZP_08870438.1	patatin				PF01734	
Amz_1697	ZP_08868661.1	pectate_lyase		ko00040			
Amz_1696		pectate_lyase,_partial	COG3866	ko00040			
Amz_0891	ZP_08868190.1	peptidase	COG1506	ko04974		PF00930	
Amz_0500	ZP_08867963.1	peptidase_M61	COG3975				
Amz_0438	ZP_08867912.1	peptidase_S26,_conserved_region	COG4959				
Amz_0935	ZP_08867912.1	Peptidase_S26,_conserved_region	COG4959				
Amz_1756		Peptidase_S53_propeptide	COG4934		cl10459		
Amz_5090		permease	COG0701		cl15898		
Amz_0779	ZP_08868128.1	phage_associated_protein__ph1R	COG0791				
Amz_3948	ZP_08869984.1	phenylalanine4hydroxylase__cl01244	COG3186	ko00360	cl01244	PF00351	TIGR01267
Amz_3767	ZP_08867884.1	phosphoesterase_related_to_lcc_protein	COG0488		cl13995	PF00149	
Amz_3879	ZP_08867884.1	phosphoesterase_related_to_lcc_protein			cl13995	PF00149	
Amz_2925		phospholipase					

Amz_1724	ZP_08868678.1	phospholipase_C_phosphocholinespecific	COG3511	ko00562		PF04185	TIGR03396
Amz_3611	ZP_08869789.1	PHP-like_metal-dependent_phosphoesterase_					
Amz_5355		phytanoyl-CoA_dioxygenase_family_protein	COG5285	ko04146	cl15773		
Amz_0675		PilT_domain-containing_protein	COG3742		cl14812		
Amz_2425		PilT_protein	COG4374		cl14812		
Amz_2678		polysaccharide_deacetylase			cl15692	PF01522	
Amz_3631	ZP_08869800.1	porin	COG3746				
Amz_5533	ZP_08870970.1	possibly_involved_in_uti	COG3193		cl01249		
Amz_1293	ZP_08868443.1	precorrin-6A_synthase_(deacetylating)_cl00304	COG2243	ko00860	cl00304	PF00590	TIGR02434
Amz_1299		prepilin-type_N-terminal_cleavage/methylat	COG2165	ko03070			
Amz_4656	ZP_08870434.1	protein_required_for_attachment_to_host_ce	COG5622		cl02398		
Amz_5121		protein_U	COG5430		cl02253		
Amz_4176	ZP_08870130.1	pseudoazurin	COG3794	ko00195	cl11412	PF00127	TIGR02375
Amz_4347	ZP_08867682.1	Pyrimidinespecific_ribonucleoside_hydrolase	COG1957	ko00230	cl00226		
Amz_3549	ZP_08869761.1	regulatory	COG3393				
Amz_5297	ZP_08870822.1	RES_domain_protein	COG5654		cl02411		
Amz_4469	ZP_08870322.1	restriction_modification_system_DNA_speci	COG0732			PF01420	
Amz_0956		rhamnoseproton_symporter			cl05728	PF06379	
Amz_2464	ZP_08870246.1	Ribbon-helix-helix_protein_copG_family				PF01402	
Amz_1111	ZP_08868323.1	riboflavin_biosynthesis_protein	COG0262	ko00740	cl00161	PF01872	
Amz_1971		ribosyl_nicotinamide_transporter	COG3201		cl01256	PF04973	TIGR01528
Amz_5543		RND_efflux_system_outer_membrane_lipoprot					
Amz_1592		rubrerythrin					
Amz_0402	ZP_08867879.1	S1/P1_nuclease_cl16915			cl16915	PF02265	
Amz_1837		selenocysteine-specific_translation_elong					
Amz_0405	ZP_08867884.1	Ser/Thr_protein_phosphatase_family_protein_cl13995			cl13995		
Amz_1463		short-chain_dehydrogenase/reductase_SDR					
Amz_1514	ZP_08868654.1	shortchain_dehydrogenasereductase_SDR		ko00140			

Amz_3893		siderophore_biosynthesis_related_protein/ace	COG1670	ko00310			
Amz_0027	ZP_08867631.1	signal_peptide_protein			cl01081	PF04205	
Amz_2881		Small_membrane_protein					
Amz_1779		snoaL-like_polyketide_cyclase_family_prote					
Amz_3801		Spore_coat_U_domain-containing_protein	COG5430		cl02253		
Amz_4326	ZP_08870226.1	sterol_desaturase_family_protein	COG3000	ko00351			
Amz_0986	ZP_08868247.1	subtilisin-like_serine_protease			cl10459	PF00082	
Amz_4411	ZP_08870290.1	sugar_hydrolase_cl10447		ko00520	cl10447	PF00704	
Amz_5531	ZP_08870968.1	sugar_phosphate_permease					
Amz_5494	ZP_08870934.1	sulfotransferase				PF00685	
Amz_4201	ZP_08870148.1	tannase			cl06521		
Amz_2171	ZP_08868942.1	terminase_large_subunit_cl01808	COG4373		cl01808		
Amz_2247		TetR_family_transcriptional_regulator					
Amz_4576	ZP_08870388.1	TetR_family_transcriptional_regulator				PF02909	
Amz_0483	ZP_08867951.1	thymidylate_synthase_cl00358	COG0207	ko00240	cl00358	PF00303	TIGR03284
Amz_0780	ZP_08868129.1	TIGR02218_family_protein_ph1R	COG5449		cl10710		
Amz_1868	ZP_08868764.1	TIGR02281_family_clan_AA_aspartic_protease	COG3577	ko04112	cl11403		
Amz_0514	ZP_08867972.1	TIGR03032_family_protein					TIGR03032
Amz_1489	ZP_08868548.1	TonB-dependent_receptor	COG4771			PF00593	
Amz_2375	ZP_08869057.1	tonb-dependent_receptor	COG1629		cl01155	PF07715	
Amz_4784	ZP_08868764.1	TPR_repeat-containing_protein	COG3577		cl11403		
Amz_0674		Transcription_factor	COG4423		cl01834		
Amz_2916	ZP_08869360.1	transcription_factor_jumonji_jmjC_domaincontaining_protein			cl15814		
Amz_2600		transcriptional_regulator				PF01381	
Amz_4560		transcriptional_regulator	COG2808		cl00381		
Amz_1110		transcriptional_regulator_MerR_family_pro					
Amz_0419	ZP_08867896.1	transmembrane_protein	COG1593				
Amz_2864	ZP_08869331.1	transposase					

Amz_5226	ZP_08870916.1	transposase				
Amz_5475	ZP_08870916.1	transposase				
Amz_2824		transposase_is3_is911_family_protein				PF01527
Amz_2914	ZP_08869359.1	tryptophan_halogenase	COG0644		cl09931	
Amz_2968	ZP_08867835.1	tryptophan_halogenase	COG0665		cl09931	
Amz_2970	ZP_08869398.1	tryptophan_halogenase	COG0665		cl09931	
Amz_4796	ZP_08870510.1	tryptophan_halogenase	COG0644		cl09931	
Amz_0342	ZP_08867835.1	tryptophan_halogenase__K14266	COG0654		cl09931	
Amz_2917	ZP_08869361.1	tryptophan_halogenase__K14266	COG0665		cl09931	
Amz_2852	ZP_08869322.1	tryptophan_halogenase__ph2R			cl09931	
Amz_1645		tryptophanrich_sensory_protein	COG3476			PF03073
Amz_0673		Twin-arginine_translocation_pathway_signa	COG4297		cl09118	
Amz_1052	ZP_08868283.1	twinarginine_translocation_pathway_signal_protein	COG3538		cl15402	
Amz_5022	ZP_08870652.1	twinarginine_translocation_pathway_signal_protein		ko00562		PF05506
Amz_2426	ZP_08869086.1	twinarginine_translocation_pathway_signal_protein__COG3538__K09704	COG3538		cl15402	
Amz_2428	ZP_08869088.1	twinarginine_translocation_pathway_signal_protein__K01206	COG3669	ko00511		PF01120
Amz_5278		type_11_methyltransferase				
Amz_0387	ZP_08867865.1	type_I_phosphodiesterasenucleotide_pyrophosphatase		ko00627		
Amz_1300		type_II_secretion_system_protein_G	COG2165	ko03070	cl07082	TIGR02532
Amz_1459		type_IV_secretory_pathway_VirB3_family_pr	COG5268	ko03070		
Amz_4091		Vault_protein_inter-alpha-trypsin	COG4676			
Amz_1683		WD40_domain-containing_protein	COG0823			
Amz_1420	ZP_08868513.1	xylan_1,4-beta-xylosidase	COG3664	ko00500	cl15381	PF01229
Amz_0020		xylose_isomerase_domain-containing_protei	COG1082	ko00562		
Amz_2657		YcfA_family_protein				
Amz_0157	ZP_08867704.1	YD_repeat-containing_protein				TIGR01643

Apenas proteínas com Valor H = 0 e função conhecida são mostradas.

**Tabela Suplementar 3** – Proteínas exclusivas de *A. amazonense* com relação a *A. brasilense* identificadas pela técnica de RDA e com Valor H < 0,42.

ID	Acesso	Produto Proteico	Acesso <i>A.brasilense</i>	Valor H	Valor E	COG	Classe	KEGG	MAPA	CDD	PFAM
Amz_0180		Conserved_Hypothetical_protein_WSK_3546		0	1,3			K01046	ko00561		PF03583
Amz_0413		Hypothetical_protein_Amzhyp048		0	23						
Amz_0467*	ZP_08867938.1	9cisepoxycarotenoid_dioxygenase		0	0,19	COG3670	Q	K11159		cl10080	
Amz_1184	ZP_08868359.1	HD_superfamily_phosphohydrolase__COG1078		0	0,43	COG1078	R	K06885			
Amz_1185	ZP_08868360.1	Phage-related DNA-binding protein		0	0,04	COG2856	E	K01486	ko00230	cl01076	
Amz_1832	ZP_08868748.1	Conserved_hypothetical_protein_Caul_1390		0	0,004	COG4805	S	K01322			
Amz_1951	ZP_08868812.1	Conserved_Hypothetical_protein_BamMC406_2283		0	0,035						
Amz_2131		Hypothetical_protein_Amzhyp212		0	1,6						
Amz_2478	ZP_08869093.1	inosineuridinepreferring_nucleoside_hydrolase__K01239		0	0,023	COG1957	F	K01239	ko00230	cl00226	
Amz_2566		Conserved_Hypothetical_protein_SeclA1_11103		0	0,16						
Amz_3005*	ZP_08869418.1	Haloacid_dehalogenase_domain-containing_p		0	0,021	COG5610	R				
Amz_3156*		Possible_squalene_ciclase		0	0,023						
Amz_3225	ZP_08869550.1	betaglucosidase__K01188		0	0,018			K01188	ko00460		PF01915
Amz_3850		Conserved_Hypothetical_protein_BBAL3_3219		0	0,35						
Amz_4590*	ZP_08870398.1	Conserved_Hypothetical_protein_SMGD1_0271		0	0,22						
Amz_4682		Hypothetical_protein_Amzhyp470		0	2						
Amz_5170*	ZP_08870736.1	glutamine_cyclotransferase		0	0,15	COG3823	O	K00683			
Amz_3430	ZP_08869686.1	Conserved_Hypothetical_protein_STIAU_3504	YP_004986470.1	0,021	0,01	COG0210	L	K03657	ko03420		PF00580
Amz_1060	ZP_08868288.1	phosphatselective_porin_O_and_P_family_protein	YP_004986805.1	0,052	0,007	COG3746	P	K07221			
Amz_4139*	ZP_08870105.1	Fucose_galactose_transporter	YP_004984612.1	0,065	0,007	COG0323	L			cl00075	PF02518
Amz_3586	ZP_08869777.1	membrane_protein__K07243	YP_004987200.1	0,068	0,003	COG0672	P	K07243		cl00475	PF03239
Amz_1088*	ZP_08868306.1	glycosyl_hydrolase_family_protein	YP_004984663.1	0,085	1,00E-04	COG1472	G	K05349	ko00460	cl07971	PF00933
Amz_4798	ZP_08870511.1	TonBdependent_receptor	YP_004985396.1	0,09	1,00E-14	COG1629	P	K02014		cl01155	PF07715
Amz_4130	ZP_08870095.1	tonB_dependent_receptor_family_protein	YP_004986828.1	0,093	9,00E-12	COG1629	P	K02014		cl01155	
Amz_1233	ZP_08868397.1	Conserved_Hypothetical_protein_HMPREF9695_02058	YP_005030378.1	0,11	3,00E-04	COG1835	I	K07052			PF01757

Amz_3344	ZP_08869625.1	twocomponent_sensor_histidine_kinase	YP_004987073.1	0,149	4,00E-27	COG2205	<b>T</b>	K07636	ko02020		
Amz_2224	ZP_08868967.1	alphanaglucosidase__COG1501	YP_004986308.1	0,169	3,00E-40	COG1501	<b>G</b>	K01811			PF01055
Amz_1249	ZP_08868409.1	Beta-lactamase	YP_004987208.1	0,177	1,00E-12	COG1680	<b>V</b>	K01467	ko00311	cl01009	
Amz_3732	ZP_08869863.1	phosphotransferase_enzyme_family_protein__	YP_005031761.1	0,203	2,00E-12	COG3173	<b>R</b>	K06979		cl09925	
Amz_0216	ZP_08867749.1	TonBdependent_receptor	YP_004986843.1	0,205	2,00E-13	COG4771	<b>P</b>	K02014		cl01155	
Amz_4964	ZP_08870610.1	NaH_antipporter	YP_004984807.1	0,205	2,00E-22	COG0025	<b>P</b>	K03316			PF00999
Amz_4033		Conserved_Hypothetical_protein_TMO_2113	YP_004984872.1	0,212	5,00E-31	COG1376	<b>S</b>	K00257	ko00281		
Amz_5155	ZP_08870728.1	transposase__ph4R	YP_004985742.1	0,221	2,00E-17	COG5659	<b>L</b>				
Amz_5324		Conserved_hypothetical_protein_ATCR1_22299	YP_005031771.1	0,224	6,00E-08	COG2361	<b>S</b>	K07075			
Amz_4747	ZP_08870489.1	conserved_hypothetical_protein_Cseg1488	YP_005033092.1	0,227	2,00E-09						
Amz_3246	ZP_08869559.1	Conserved_Hypothetical_protein_APM_0875	YP_005030852.1	0,251	5,00E-13						
Amz_1770		Ribokinase	YP_004987141.1	0,273	2,00E-10	COG0524	<b>G</b>	K00852	ko00030	cl00192	
Amz_3422	ZP_08869680.1	amidohydrolase_family_protein	YP_004986133.1	0,275	5,00E-38	COG1228	<b>Q</b>	K15358		cl00281	PF01979
Amz_0251	ZP_08870410.1	cytochrome_c_family_protein	YP_005029953.1	0,278	5,00E-22	COG2010	<b>C</b>	K00406	ko00190		PF00034
Amz_3982	ZP_08869999.1	Conserved_hypothetical_protein_Mmwyl1_3150	YP_004987234.1	0,282	7,00E-27	COG1506	<b>E</b>	K06889			PF00326
Amz_0064	ZP_08867651.1	O-antigen_acetylase	YP_005030378.1	0,32	1,00E-85	COG1835	<b>I</b>	K00680	ko00350	cl00792	PF01757
Amz_1946	ZP_08868450.1	glycosyl_transferase_family_protein	YP_005030639.1	0,33	4,00E-33	COG0463	<b>M</b>	K00721	ko00510	cl11394	
Amz_3210	ZP_08868840.1	Twocomponent_response_regulator__K07814	YP_005033134.1	0,372	1,00E-63	COG3437	<b>KT</b>	K07814			PF00072
Amz_3516		beta-lactamase_domain-containing_protein__	YP_004985659.1	0,373	6,00E-44	COG0491	<b>R</b>	K01567		cl00446	PF00753
Amz_0206	ZP_08867740.1	aldehyde_dehydrogenase_Ath115__COG1012	YP_004986238.1	0,382	9,00E-95	COG1012	<b>C</b>	K00130	ko00260	cl11961	PF00171
Amz_1477	ZP_08868538.1	Glycosyl_transferase_protein	YP_004987575.1	0,4	5,00E-150	COG1216	<b>R</b>	K07011			

\* proteínas discutidas individualmente neste trabalho.

**Tabela Suplementar 4** – Proteínas de *A. amazonense* relacionadas ao metabolismo de hormônios vegetais pela classificação do banco de ortólogos KEGG.

ID	Acesso	Produto proteico	KEGG	TAIR	AHD	TCDB
Amz_1202		Outer membrane lipoprotein				
Amz_0351	ZP_08867838.1	Signal transduction histidine kinase	K11527	AT1G27320	AT1G27320	Q93N33
Amz_0062	ZP_08867649.1	Response regulator containing CheY	K03413	AT1G67710	AT1G67710	
Amz_0140	ZP_08867649.1	Response regulator containing CheY	K03413	AT1G19050	AT1G19050	
Amz_1316	ZP_08868454.1	Arabinose efflux permease	K08224			P43531
Amz_3028	ZP_08869432.1	Signal transduction histidine kinase	K07636	AT1G66340	AT1G66340	Q93N33
Amz_4403	ZP_08870283.1	Cation/multidrug efflux pump				P25197

**Tabela Suplementar 5 – Proteínas de *A. amazonense* relacionadas ao metabolismo de hormônios vegetais.**

ID	Acesso <i>A. amazonense</i>	Produto proteico em <i>A. amazonense</i>	COG	KEGG	MAPA KEGG	<i>A. thaliana</i> Hormone	CDD	PFAM
Amz_4355		glutamatedcysteine ligase_Ath060	COG3572	K01919	ko00480	AT4G23100	cl00954	PF04107
Amz_3085	ZP_08869464.1	membranebound protontranslocating pyrophosphatase_Ath072	COG3808	K01507	ko00190	AT1G15690		PF03030
Amz_3710	ZP_08869852.1	heavy metal translocating Ptype ATPase_Ath086	COG2217	K01533		AT5G44790		PF00122
Amz_2509	ZP_08869141.1	catalase_Ath125	COG0753	K03781	ko00380	AT1G20630	cl09506	
Amz_4112	ZP_08870082.1	XaaPro aminopeptidase_Ath113	COG0006	K01262		AT4G36760	cl00279	PF00557
Amz_2883	ZP_08869343.1	catalase_Ath133	COG0753	K03781	ko00380	AT1G20630		
Amz_4648	ZP_08870428.1	Catalase_Ath123	COG0753	K03781	ko00380	AT1G20630	cl09506	
Amz_3738	ZP_08869865.1	xanthine dehydrogenase large subunit_Ath069	COG4631	K00106	ko00230	AT5G20960		PF02738
Amz_1637	ZP_08868628.1	Sadenosylmethionine synthetase	COG0192	K00789	ko00270	AT1G02500		
Amz_3830	ZP_08869908.1	ABC transporter ATPase and permease components	COG1132	K11085		AT2G47000		PF00664
Amz_4218	ZP_08868262.1	EnoylCoA hydratase3hydroxyacylCoA dehydrogenase	COG1250	K07516	ko00071	AT4G29010		PF02737
Amz_1010	ZP_08868262.1	peroxisomal bifunctional enzyme	COG1250	K07514	ko00071	AT4G29010		PF02737
Amz_3034	ZP_08869435.1	3ketoacylCoA thiolase	COG0183	K00632	ko00071	AT2G33150	cl09938	PF00108
Amz_2750	ZP_08869262.1	phosphoribosylaminoimidazolesuccinocarboxamide synthase_K01923	COG0152	K01923	ko00230	AT3G21110	cl00231	PF01259
Amz_2204		anthranilate synthase component I	COG0147	K01657	ko00400	AT5G05730		
Amz_1168		oxidoreductase_K10680_cl09108	COG1902	K10680	ko00633	AT1G76680	cl09108	PF00724
Amz_2738	ZP_08869256.1	Mg2importing ATPase	COG0474	K01531		AT2G18960		PF00122
Amz_2936	ZP_08869375.1	aldehyde dehydrogenase	COG1012	K00128		AT4G34240	cl11961	PF00171
Amz_2524	ZP_08869150.1	tetratricopeptide TPR2 repeat protein	COG3914		ko00514	AT3G11540		



Amz_3009	ZP_08869421.1	AcetylCoA acetyltransferase with thiolase domain AcetoacetylCoA thiolase	COG0183	K00626	ko00071	AT2G33150	PF00108
Amz_4646	ZP_08870427.1	AcetylCoA acetyltransferase with thiolase domain Betaketothiolase	COG0183	K07508	ko00071	AT2G33150	PF00108
Amz_2904	ZP_08869352.1	ABC transporter ATPbinding protein_Ath058	COG5265	K06147		AT2G36910	PF00664
Amz_4721	ZP_08870472.1	longchainfattyacidCoA ligase	COG0318	K01897	ko00071	AT4G05160	cl17068 PF00501
Amz_0206	ZP_08867740.1	aldehyde dehydrogenase_Ath115_COG1012	COG1012	K00128	ko00260	AT1G54100	cl11961 PF00171
Amz_2366		gammaglutamyl phosphate reductase	COG0014	K00147	ko00330	AT2G39800	cl11961 PF00171
Amz_3109	ZP_08869477.1	propionylCoA carboxylase alpha chain_Ath083	COG4770	K01965	ko00280	AT1G36160	PF02786
Amz_2066	ZP_08868881.1	peptidase M1 membrane alanine aminopeptid	COG0308	K08776		AT4G33090	cl14813 PF01433
Amz_1521	ZP_08868567.1	ABC transporter fused ATPbindingpermease protein	COG1132	K06147		AT3G28860	PF00005
Amz_1329	ZP_08868459.1	Toxin secretion ABC transporter ATPbinding and membrane protein hl	COG2274	K11004		AT3G28860	PF00005
Amz_3635	ZP_08869805.1	acetylCoA carboxylase biotin carboxylase subunit_Ath085	COG4770	K01961	ko00061	AT1G36160	PF02786
Amz_0826	ZP_08868153.1	succinatesemialdehyde dehydrogenase_Ath077_K00135	COG1012		ko00250	AT1G54100	cl11961 PF00171
Amz_0174	ZP_08867717.1	malonylCoA synthase	COG0318		ko00627	AT4G05160	PF00501
Amz_4716	ZP_08870468.1	succinatesemialdehyde dehydrogenase_Ath056	COG1012	K00135	ko00250	AT1G54100	cl11961 PF00171
Amz_0609	ZP_08868026.1	succinic semialdehyde dehydrogenase_Ath066	COG1012	K00135	ko00250	AT1G54100	cl11961 PF00171
Amz_2409	ZP_08869078.1	metal ion transporter_cl15845	COG1914	K03322		AT5G03280	cl15845 PF01566
Amz_4592	ZP_08870399.1	aldehyde dehydrogenase_Ath135	COG1012	K00128	ko00010	AT1G54100	cl11961 PF00171
Amz_1700	ZP_08868664.1	piperideine-6-carboxylate dehydrogenase	COG1012		ko00010	AT1G54100	cl11961 PF00171
Amz_0556	ZP_08867990.1	aldehyde dehydrogenase NAD_Ath093	COG1012	K00128	ko00360	AT1G54100	cl11961 PF00171
Amz_0455	ZP_08867928.1	aldehyde dehydrogenase	COG1012	K00128	ko00622	AT1G54100	cl11961 PF00171
Amz_1800	ZP_08868725.1	TrpR binding protein WrbA_cl00438	COG0655	K03809		AT5G54500	cl00438

Amz_3354	ZP_08869630.1	acylCoA synthetase	COG0318		AT1G20510	cl17068	PF00501
Amz_1222	ZP_08868388.1	betaine aldehyde dehydrogenase	COG1012	ko00010	AT1G54100	cl11961	PF00171
Amz_0073	ZP_08867657.1	NAD(P)H:quinone oxidoreductase_cl00438	COG0655	K03809	AT5G54500	cl00438	
Amz_2669	ZP_08867990.1	methylmalonatesemialdehyde dehydrogenase_Ath109	COG1012	K00140 ko00280	AT1G54100	cl11961	PF00171
Amz_4550	ZP_08870370.1	sensor histidine kinase	COG0642	K11527	AT1G66340		PF00072
Amz_2634		oxidoreductase_cl09108	COG1902	K00540	AT1G76690	cl09108	PF00724
Amz_3133	ZP_08869493.1	potassiumtransporting ATPase subunit B	COG2216	K01547 ko02020	AT5G44790		PF00122
Amz_3212	ZP_08869539.1	AMP-dependent synthetase and ligase_cl17068	COG0318	K01897 ko00071	AT4G05160	cl17068	PF00501
Amz_0570	ZP_08867999.1	aldehyde dehydrogenase acceptor_COG1012	COG1012	K00146 ko00360	AT1G54100		PF00171
Amz_0092	ZP_08867666.1	translation elongation factor G FusA_Ath021	COG0480	K02355	AT1G56070		PF00009
Amz_2876	ZP_08869341.1	Lightregulated sensor histidine kinase Bacteriophytochrome	COG4251	K11354	AT2G18790		
Amz_3510	ZP_08869737.1	ATPbinding cassette subfamily C	COG2274	K06147 ko02010	AT2G36910		PF03412
Amz_4196	ZP_08869630.1	fattyacylCoA synthase	COG0318	K01976	AT1G20510		PF00501
Amz_3305	ZP_08869601.1	succinatesemialdehyde dehydrogenase	COG1012	K00135 ko00250	AT1G54100	cl11961	PF00171
Amz_3700	ZP_08869850.1	NAD(P)H:quinone oxidoreductase_cl00438	COG0655		AT5G54500	cl00438	
Amz_4865	ZP_08870546.1	translation elongation factor G FusA_Ath044	COG0480	K02355	AT1G56070		PF00009
Amz_1048	ZP_08867990.1	methylmalonatesemialdehyde dehydrogenase_Ath100	COG1012	K00140 ko00280	AT1G54100	cl11961	PF00171
Amz_4014	ZP_08870020.1	ATPbinding cassette subfamily B	COG2274	ko02010	AT3G28860		PF00005
Amz_1387	ZP_08868494.1	methylcrotonoylCoA carboxylase subunit alpha_Ath096	COG4770	ko00280	AT1G36160		PF02786
Amz_0965	ZP_08868231.1	ATPdependent RNA helicase_K11927	COG0513	K11927 ko03018	AT3G53110		PF00270
Amz_0351	ZP_08867838.1	twocomponent hybrid sensor and regulator	COG0642	K11527	AT1G27320		PF00072

Amz_5093	ZP_08870694.1	malate dehydrogenase	COG0039	K00024	ko00020	AT4G17260		
Amz_1243	ZP_08868404.1	aldehyde dehydrogenase _COG1012_cl11961	COG1012		ko00330	AT1G44170	cl11961	PF00171
Amz_5279	ZP_08870810.1	PAS/PAC sensor hybrid histidine kinase	COG0642			AT1G66340		PF02518
Amz_2103	ZP_08868901.1	Predicted O-linked N-acetylglucosamine tr	COG3914	K09667	ko00514	AT3G11540		
Amz_2058	ZP_08870926.1	aldehyde dehydrogenase family protein	COG1012		ko00622	AT1G54100	cl11961	PF00171
Amz_3198	ZP_08869528.1	aldehyde dehydrogenase NAD_cl11961	COG1012		ko00010	AT1G44170	cl11961	PF00171
Amz_2956	ZP_08869386.1	FADdependent pyridine nucleotidedisulfide oxidoreductase	COG1251		ko00071	AT3G09940		
Amz_4779	ZP_08870502.1	glutathione peroxidase family protein_cl00388	COG0386	K00432	ko00480	AT2G43350	cl00388	PF00255
Amz_5485	ZP_08870926.1	aldehyde dehydrogenase	COG1012	K00131	ko00622	AT1G44170	cl11961	PF00171
Amz_3000	ZP_08869415.1	TPR repeatcontaining protein	COG0457	K09667		AT3G11540		
Amz_3776	ZP_08869882.1	aldehyde dehydrogenase_cl11961	COG1012		ko00010	AT1G54100	cl11961	PF00171
Amz_0584	ZP_08868008.1	crotonobetainecarnitineCoA ligase	COG0318	K02182		AT1G20510	cl17068	PF00501
Amz_1097	ZP_08868312.1	Pirin domain-containing protein	COG1741	K06911		AT3G59220		
Amz_4710	ZP_08868901.1	Conserved Hypothetical protein AZL 014390	COG3914		ko00514	AT3G11540		
Amz_3737	ZP_08869819.1	xanthine dehydrogenase small subunit	COG4630	K13481	ko00230	AT3G43600		PF00941
Amz_4270	ZP_08870188.1	non-motile and phage-resistance protein	COG0642	K07716	ko02020	AT5G35750		PF02518
Amz_4866		tRNA delta2isopentenylpyrophosphate transferase	COG0324	K00791	ko00908	AT5G20040		PF01715
Amz_1383	ZP_08868553.1	shortchain dehydrogenasereductase SDR_K00540	COG1028	K11147		AT1G52340		PF00106
Amz_5537	ZP_08870973.1	NAD-dependent aldehyde dehydrogenase	COG1012		ko00281	AT1G54100	cl11961	PF00171
Amz_3093	ZP_08869375.1	coniferyl aldehyde dehydrogenase_cl11961	COG1012		ko00010	AT4G34240	cl11961	PF00171
Amz_3671	ZP_08869831.1	nonmotile and phageresistance protein	COG0642	K07716	ko02020	AT2G17820		PF02518

Amz_0467	ZP_08867938.1	9cisepoxycarotenoid dioxygenase	COG3670	K11159	AT1G78390	cl10080
Amz_1397	ZP_08868501.1	pirin	COG1741	K06911	AT3G59220	

---

O Banco de hormônios de *Arabidopsis* (AHD) foi utilizado como referência na presente análise.

**Tabela Suplementar 6** – Classificação metabólica das proteínas de *A. amazonense* conservadas com o banco de hormônios de *Arabidopsis*.

<b>Descrição</b>	<b>Proteínas</b>	<b>Classe</b>
Produção e conversão de energia	26	C
Transporte e metabolismo lipídico	15	I
Mecanismos de transdução de sinal	13	T
Metabólitos secundários, transporte e catabolismo	8	Q
Função geral apenas	8	R
Transporte e metabolismo de íons inorgânicos	7	P
Modificação pós-traducional, <i>turnover</i> de proteínas, chaperonas	5	O
Mecanismos de defesa	5	V
Transporte e metabolismo de aminoácidos	4	E
Tradução, estrutura ribossomal e biogênese	4	J
Transporte e metabolismo de nucleotídeos	3	F
Transporte e metabolismo de Coenzimas	3	H
Transcrição	3	K
Replicação, recombinação e reparação	1	L
Parede celular / membrana / biogênese do envelope	1	M

**Tabela Suplementar 7** - Proteínas de *A. amazonense* com similaridade nas proteínas diferencialmente expressas na presença de auxina em *A. brasilense* classificadas por COG.

<b>Descrição</b>	<b>Proteínas</b>	<b>Classe</b>
Tradução, estrutura ribossomal e biogênese	31	J
Produção e conversão de energia	15	C
Transporte e metabolismo de aminoácidos	15	E
Transporte e metabolismo de carboidratos	7	G
Modificação pós-traducional, <i>turnover</i> de proteínas, chaperonas	7	O
Função desconhecida	6	S
Transporte e metabolismo lipídico	5	I
Transporte e metabolismo de Coenzimas	4	H
Transporte e metabolismo de íons inorgânicos	4	P
Mecanismos de transdução de sinal	4	T
Transporte e metabolismo de nucleotídeos	3	F
Biossíntese de metabólitos secundários, transporte e catabolismo	3	Q
Função geral apenas	3	R
Transcrição	2	K
Replicação, recombinação e reparação	1	L
Parede celular / membrana / biogênese do envelope	1	M
Tráfego intracelular, secreção e transporte vesicular	1	U

**Tabela Suplementar 8** – Classificação das proteínas de *Arabidopsis thaliana* (AHD) com similares em *A. amazonense* que foram diferencialmente expressas na presença de auxina em *A. brasilense*. Identificação dos hormônios correspondentes e descrição da atuação no seu metabolismo assim como o teste realizado para caracterização.

AHD	Hormônio	Teste	Via	Descrição
AT1G52340	ácido abscísico	Mutante	biossíntese	desidrogenase / redutase envolvidas na conversão de xantoxina-ABA-aldeído
AT1G52340		GeneOntology	-	processo de biossíntese do ácido abscísico
AT1G54100	ácido abscísico	GeneOntology	-	resposta ao estímulo de ácido abscísico
AT2G24850	ácido jasmônico	GeneOntology	-	resposta ao estímulo ácido jasmônico
AT2G27150	ácido abscísico	Mutante	biossíntese	codifica abscísico aldeído oxidase e catalisa a etapa final de biossíntese de ácido abscísico
		GeneOntology	-	processo de biossíntese do ácido abscísico
AT3G28860	auxina	Mutante	Transporte	resistência a múltiplas drogas (MDR)-como transportadores ABC
AT3G28860		GeneOntology	-	resposta ao estímulo de auxina
AT3G56860	ácido abscísico	GeneOntology	-	sinalização mediada por ácido abscísico
AT4G05160	ácido jasmônico	GeneOntology	-	processo de biossíntese do ácido jasmonico
AT5G20960	auxina	GeneOntology	-	processo de biossíntese de auxina
AT5G35750	citocinina	Mutante	Receptor	atividade osmosensor, a resposta estresses variados
		GeneOntology	-	atividade do receptor de citocinina
	ácido abscísico	GeneOntology	-	resposta ao estímulo de ácido abscísico
AT5G63310	auxina	GeneOntology	-	auxina mediada via de sinalização

**Tabela Suplementar 9** – Transportadores de *A. amazonense* identificados no TCDB e relacionados com a Classe U do banco de dados COG.

ID	Acesso		Uniprot KB	TCDB	COG	Classe	MAPA		
Amz_2468		Conjugal transfer/entry exclusion protein	P71185	3.A.7.4.1	COG5314	U		cl02193	
Amz_3443	ZP_08869694.1	Preprotein translocase subunit SecD	P0AG90	2.A.6.4.1	COG0342	U	ko03060		
Amz_1065	ZP_08868290.1	Biopolymer transport proteins	P0ABU9	1.A.30.2.2	COG0811	U		cl00568	PF01618
Amz_1066	ZP_08868291.1	Biopolymer transport protein	P0ABV2	1.A.30.2.2	COG0848	U		cl00537	PF02472
Amz_0641	ZP_08868051.1	Biopolymer transport protein	P0ABV2	1.A.30.2.2	COG0848	U		cl00537	PF02472
Amz_1067	ZP_08868292.1	Biopolymer transport protein	P0ABV2	1.A.30.2.2	COG0848	U			PF02472
Amz_1766	ZP_08868703.1	Biopolymer transport protein	P0ABV2	1.A.30.2.2	COG0848	U		cl00537	PF02472
Amz_3619		Signal recognition particle GTPase	P10121	3.A.5.1.1	COG0552	U	ko03060		PF00448
Amz_1458	ZP_08868532.1	Type IV secretory pathway, VirB4 components	P71180	3.A.7.4.1	COG3451	U	ko03070		PF03135
Amz_2467	ZP_08869116.1	Type IV secretory pathway, VirB4 components	P71180	3.A.7.4.1	COG3451	U	ko03070		
Amz_0971	ZP_08868236.1	Type IV secretory pathway, VirD4 components	P13464	3.A.7.1.1	COG3505	U	ko03070		PF02534
Amz_1466	ZP_08868534.1	Type IV secretory pathway, VirD4 components	P13464	3.A.7.1.1	COG3505	U	ko03070		PF02534
Amz_4363	ZP_08870245.1	Type IV secretory pathway, VirD4 components	P13464	3.A.7.1.1	COG3505	U	ko03070		PF02534
Amz_1461	ZP_08869115.1	Flp pilus assembly protein, ATPase CpaF	Q56469	3.A.7.4.1	COG4962	U	ko03070		PF00437
Amz_1492	ZP_08868550.1	Type IV secretory pathway, VirB9 components	P71182	3.A.7.4.1	COG3504	U	ko03070	cl11423	
Amz_1493	ZP_08868551.1	Type IV secretory pathway, VirB10 components	P71184	3.A.7.4.1	COG2948	U	ko03070	cl04242	
Amz_3637	ZP_08868532.1	Type IV secretory pathway, VirB4 components	P71180	3.A.7.4.1	COG3451	U	ko03070		
Amz_2242	ZP_08868549.1	Type IV secretory pathway, TrbF components	P71181	3.A.7.4.1	COG3701	U	ko03070	cl01500	PF04335
Amz_4148	ZP_08870114.1	Type IV secretory pathway, TrbF components	P71181	3.A.7.4.1	COG3701	U	ko03070	cl01500	PF04335
Amz_1491	ZP_08868549.1	Type IV secretory pathway, TrbF components	P71181	3.A.7.4.1	COG3701	U	ko03070	cl01500	PF04335
Amz_0635	ZP_08868046.1	Type IV secretory pathway, VirB10 components	P71184	3.A.7.4.1	COG2948	U	ko03070	cl04242	
Amz_4146	ZP_08870112.1	Type IV secretory pathway, VirB10 components	P71184	3.A.7.4.1	COG2948	U	ko03070	cl04242	
Amz_1457		Conjugal transfer/entry exclusion protein	P71185	3.A.7.4.1	COG5314	U		cl02193	



Amz_1650	ZP_08868635.1	Hemolysin activation/secretion protein	P45356	1.B.20.3.1	COG2831	U	ko03070		
Amz_5132	ZP_08870721.1	Hemolysin activation/secretion protein	Q29XT8	1.B.20.3.3	COG2831	U			
Amz_2241	ZP_08868977.1	Preprotein translocase subunit YidC	P25714	2.A.9.3.1	COG0706	U	ko03060		PF02096
Amz_2077		Multiple antibiotic transporter	Q8J305	2.A.95.1.4	COG2095	U		cl00919	
Amz_2895		Multiple antibiotic transporter	Q8J305	2.A.95.1.4	COG2095	U		cl00919	
Amz_3545	ZP_08869759.1	Hemolysin activation/secretion protein	Q6Y659	1.B.20.3.4	COG2831	U			
Amz_3912		Preprotein translocase subunit SecY	P0AGA2	3.A.5.1.1	COG0201	U	ko03060		PF00344
Amz_4502		Preprotein translocase subunit YajC	P0ADZ7	9.B.18.1.1	COG1862	U	ko03060	cl00806	
Amz_4147	ZP_08870113.1	Type IV secretory pathway, VirB9 components	P71182	3.A.7.4.1	COG3504	U	ko03070	cl11423	
Amz_2243	ZP_08868978.1	Type IV secretory pathway, VirB9 components	P71182	3.A.7.4.1	COG3504	U	ko03070	cl11423	
Amz_4597	ZP_08870402.1	Hemolysin activation/secretion protein	P45356	1.B.20.3.1	COG2831	U			
Amz_2140	ZP_08870800.1	Signal peptidase I	Q7CM88	3.A.7.13.1	COG0681	U	ko03060		
Amz_3648		Signal recognition particle GTPase	P0AGD7	3.A.5.1.1	COG0541	U	ko03060		PF00448
Amz_1767	ZP_08868704.1	Biopolymer transport proteins	P0ABU9	1.A.30.2.2	COG0811	U		cl00568	
Amz_4149	ZP_08870115.1	Biopolymer transport proteins	P0ABU7	1.A.30.2.1	COG0811	U		cl00568	PF01618
Amz_0640	ZP_08868050.1	Biopolymer transport proteins	P0ABU7	1.A.30.2.1	COG0811	U		cl00568	
Amz_0361	ZP_08867846.1	Periplasmic component of the Tol biopolymer transport system	P0A855	2.C.1.2.1	COG0823	U			PF04052
Amz_4365	ZP_08869115.1	Flp pilus assembly protein, ATPase CpaF	Q56469	3.A.7.4.1	COG4962	U	ko03070		PF00437
Amz_2465	ZP_08869115.1	Flp pilus assembly protein, ATPase CpaF	Q56469	3.A.7.4.1	COG4962	U	ko03070		PF00437
Amz_1298	ZP_08868446.1	Type II secretory pathway, component HofQ	P15644	1.B.22.1.1	COG4796	U	ko03070		PF00263
Amz_4956	ZP_08869206.1	Protease subunit of ATP-dependent Clp proteases	Q9F315	1.B.12.4.1	COG0740	OU	ko04112	cl15763	PF00574
Amz_2641	ZP_08869206.1	Protease subunit of ATP-dependent Clp proteases	Q9F315	1.B.12.4.1	COG0740	OU	ko04112		PF00574
Amz_4537	ZP_08870364.1	Tfp pilus assembly protein PilF	Q9FMA3	3.A.20.1.2	COG3063	NU		cl02429	
Amz_3800	ZP_08869896.1	P pilus assembly protein, porin PapC	Q6XBY3	1.B.11.6.1	COG3188	NU			PF00577
Amz_1178	ZP_08868353.1	P pilus assembly protein, porin PapC	Q6XBY3	1.B.11.6.1	COG3188	NU			PF00577
Amz_4096	ZP_08870071.1	Flagellar biosynthesis pathway, component FlhA	P40729	3.A.6.2.1	COG1298	NU	ko02040	cl07980	PF00771
Amz_1261	ZP_08868415.1	Flagellar biosynthesis pathway, component FlhB	P40727	3.A.6.2.1	COG1377	NU	ko02040		PF01312
Amz_1270	ZP_08868424.1	Flagellar biosynthesis pathway, component FlhP	P54700	3.A.6.2.1	COG1338	NU	ko02040	cl00593	PF00813
Amz_1263	ZP_08868417.1	Flagellar biosynthesis pathway, component FlhQ	Q7BFA7	3.A.6.1.1	COG1987	NU	ko02040	cl00867	PF01313

Amz_1262	ZP_08868416.1	Flagellar biosynthesis pathway, component FlIR	P54702	3.A.6.2.1	COG1684	NU	ko02040	cl00734	PF01311
Amz_4367	ZP_08870247.1	Flagellar motor switch/type III secretory pathway protein	P26419	3.A.6.2.1	COG1886	NU	ko02030	cl00819	
Amz_5151	ZP_08870247.1	Flagellar motor switch/type III secretory pathway protein	P26419	3.A.6.2.1	COG1886	NU	ko02030	cl00819	
Amz_5270	ZP_08870802.1	Flagellar biosynthesis/type III secretory pathway ATPase	P26465	3.A.6.2.1	COG1157	NU	ko02040		PF00006
Amz_0152	ZP_08867698.1	Type II secretory pathway, component PulD	Q9AGM8	3.A.15.3.1	COG1450	NU	ko03070		PF00263
Amz_5080	ZP_08870687.1	Type II secretory pathway, component PulD	P15644	1.B.22.1.1	COG1450	NU	ko03070		
Amz_0145	ZP_08867693.1	Type II secretory pathway, component PulF	P22609	3.A.15.2.1	COG1459	NU	ko03070		
Amz_0146	ZP_08867694.1	Type II secretory pathway, pseudopilin PulG	Q9XD71	3.A.15.3.1	COG2165	NU	ko03070	cl07082	
Amz_0619	ZP_08868036.1	Type II secretory pathway, pseudopilin PulG	Q9XD71	3.A.15.3.1	COG2165	NU	ko03070		
Amz_4475	ZP_08870326.1	Type II secretory pathway, pseudopilin PulG	P15746	3.A.15.1.1	COG2165	NU	ko03070	cl07082	
Amz_1191	ZP_08868367.1	Tfp pilus assembly protein PilF	O09012	3.A.20.1.1	COG3063	NU		cl02429	
Amz_2799	ZP_08869115.1	Type II secretory pathway, ATPase PulE/Tfp pilus assembly pathway, ATPase PilB	Q9AGM7	3.A.15.3.1	COG2804	NU	ko03070		PF00437
Amz_0187	ZP_08867725.1	Outer membrane protein	P11092	1.B.17.2.3	COG1538	MU	ko03070		PF02321
Amz_0200	ZP_08867737.1	Outer membrane protein	Q51487	2.A.6.2.21	COG1538	MU	ko02020		PF02321
Amz_3798	ZP_08869895.1	Outer membrane protein	Q9I0V8	2.A.6.2.39	COG1538	MU			PF02321
Amz_4120	ZP_08870088.1	Outer membrane protein	Q6V6X7	2.A.6.2.18	COG1538	MU			PF02321
Amz_1935	ZP_08868800.1	Outer membrane protein	Q51487	2.A.6.2.21	COG1538	MU			PF02321
Amz_1280	ZP_08868431.1	Outer membrane protein	Q51487	2.A.6.2.21	COG1538	MU			PF02321
Amz_4759	ZP_08870495.1	Outer membrane protein	P24145	1.B.17.3.1	COG1538	MU			PF02321
Amz_2737	ZP_08869255.1	Outer membrane protein	Q9I0V8	2.A.6.2.39	COG1538	MU			PF02321
Amz_3209	ZP_08869537.1	Outer membrane protein	Q9I0V8	2.A.6.2.39	COG1538	MU			PF02321
Amz_1314	ZP_08868431.1	Outer membrane protein	Q9I0V8	2.A.6.2.39	COG1538	MU			PF02321
Amz_2345	ZP_08869042.1	Outer membrane protein	Q4VWC6	3.A.1.105.4	COG1538	MU	ko02020		PF02321
Amz_1828	ZP_08868745.1	Outer membrane protein	Q92Q38	1.B.17.1.4	COG1538	MU	ko03070		PF02321
Amz_3867	ZP_08869937.1	Outer membrane protein	Q51487	2.A.6.2.21	COG1538	MU			PF02321

**Tabela Suplementar 10** – Identificação das sequências de proteínas localizadas nas quatro regiões virais identificadas em *A. amazonense* pelo PHAST e nível de similaridade com proteína viral correspondente.

Região	Posição relativa concatenada	Melhor Similaridade	Valor E
REGIÃO 1	complement(930611..934555)	PHAGE_Burkho_KL1: tail protein ; PP_00889; phage(gi399528737)	6,00E-15
REGIÃO 1	complement(934555..935004)	PHAGE_Roseob_1: phage cell wall peptidase ; PP_00890; phage(gi331028137)	8,00E-07
REGIÃO 1	complement(935001..935888)	PHAGE_Roseob_1: gene transfer agent ; PP_00891; phage(gi331028136)	4,00E-12
REGIÃO 1	complement(935888..936436)	PHAGE_phiJL001: gp83 ; PP_00892; phage(gi62327177)	3,00E-08
REGIÃO 1	complement(936481..940089)	PHAGE_Salmon_E1: putative tail protein (tape measure) ; PP_00893; phage(gi170676309)	1,00E-15
REGIÃO 1	complement(940286..940651)	hypothetical; PP_00894	N/A
REGIÃO 1	complement(940662..941420)	hypothetical protein Msil_3050 [Methylocella silvestris BL2] gi 217979175 ref YP_002363322.1 ; E-VALUE = 7e-58; PP_00895	N/A
REGIÃO 1	complement(941424..941675)	hypothetical; PP_00896	N/A
REGIÃO 1	complement(941672..942103)	hypothetical protein amb1772 [Magnetospirillum magneticum AMB-1] gi 83310871 ref YP_421135.1 ; E-VALUE = 1e-07; PP_00897	N/A
REGIÃO 1	complement(942100..942714)	PHAGE_Rhodob_RcapMu: virion morphogenesis protein ; PP_00898; phage(gi356870876)	1,00E-13
REGIÃO 1	complement(942717..943136)	PHAGE_Rhodob_RcapMu: conserved phage protein ; PP_00899; phage(gi356870882)	4,00E-18
REGIÃO 1	complement(943219..943488)	hypothetical protein Daes_2401 [Desulfovibrio aespoensis Aspo-2] gi 317154104 ref YP_004122152.1 ; E-VALUE = 6e-08; PP_00900	N/A
REGIÃO 1	complement(943597..944487)	PHAGE_Escher_D108: major head protein ; PP_00901; phage(gi281199678)	1,00E-64
REGIÃO 1	complement(944499..944915)	PHAGE_Pseudo_MP29: hypothetical protein PPMP29_gp32 ; PP_00902; phage(gi215480008)	2,00E-29
REGIÃO 1	complement(944932..945657)	Outer membrane efflux protein [Dechloromonas aromatica RCB] gi 71908096 ref YP_285683.1 ; E-VALUE = 1e-19; PP_00903	N/A
REGIÃO 1	complement(945771..946130)	hypothetical protein SLG_p_00550 [Sphingobium sp. SYK-6] gi 347430567 ref YP_004831175.1 ; E-VALUE = 2e-05; PP_00904	N/A
REGIÃO 1	946428..947624	urea ABC transporter, urea binding protein [Methylocella silvestris BL2] gi 217978164 ref YP_002362311.1 ; E-VALUE = 4e-176; PP_00905	N/A
REGIÃO 1	947762..948670	PHAGE_Lactob_Lj771: minor tail protein gp26-like protein ; PP_00906; phage(gi163932195)	1,00E-05
REGIÃO 2	complement(3258724..3260088)	PHAGE_Ectoca_1: EsV-1-65 ; PP_03103; phage(gi13242537)	1,00E-08
REGIÃO 2	complement(3260117..3260305)	DNA-binding two-component response regulator [Azospirillum sp. B510] gi 288957332 ref YP_003447673.1 ; E-VALUE = 2e-10; PP_03104	N/A
REGIÃO 2	complement(3260378..3260962)	hypothetical; PP_03105	N/A
REGIÃO 2	complement(3261017..3261598)	hypothetical; PP_03106	N/A
REGIÃO 2	complement(3261600..3261905)	hypothetical; PP_03107	N/A
REGIÃO 2	3261904..3262104	hypothetical; PP_03108	N/A
REGIÃO 2	3262163..3262909	PHAGE_Pseudo_MP42: DNA transposition protein ; PP_03109; phage(gi399528586)	2,00E-63

REGIÃO 2	3262906..3263550	PHAGE_Pseudo_MP29: hypothetical protein PPMP29_gp08 ; PP_03110; phage(gi215479984)	3,00E-22
REGIÃO 2	3263547..3263909	PHAGE_Pseudo_MP42: hypothetical protein ; PP_03111; phage(gi399528590)	8,00E-11
REGIÃO 2	3263906..3264139	hypothetical; PP_03112	N/A
REGIÃO 2	3264114..3264446	hypothetical; PP_03113	N/A
REGIÃO 2	3264559..3264939	PHAGE_Caulob_CcrSwift: hypothetical protein ; PP_03114; phage(gi414089889)	1,00E-09
REGIÃO 2	3264985..3265536	PHAGE_Pseudo_DMS3: hypothetical protein DMS3-11 ; PP_03115; phage(gi119953636)	9,00E-15
REGIÃO 2	3265536..3265901	PHAGE_Lister_B054: Tmp ; PP_03116; phage(gi157325302)	7,00E-08
REGIÃO 2	3265903..3267372	PROPHAGE_Escher_MG1655: carboxy-terminal protease for penicillin-binding protein 3 ; PP_03117; phage(gi16129784)	5,00E-28
REGIÃO 2	3267665..3270274	TonB-dependent receptor [Colwellia psychrerythraea 34H] gi 71278302 ref YP_270619.1 ; E-VALUE = 1e-80; PP_03118	N/A
REGIÃO 2	3270360..3271877	PHAGE_Synech_syn9: putative tryptophan halogenase ; PP_03119; phage(gi113200547)	1,00E-52
REGIÃO 3	complement(3910656..3911360)	PHAGE_Plankt_PaV_LD: ABC transporter ; PP_03712; phage(gi371496158)	9,00E-19
REGIÃO 3	complement(3911353..3912159)	PHAGE_Plankt_PaV_LD: ABC transporter ; PP_03714; phage(gi371496158)	2,00E-14
REGIÃO 3	3912355..3913605	extracellular ligand-binding receptor [Rhodopseudomonas palustris DX-1] gi 316936291 ref YP_004111273.1 ; E-VALUE = 1e-155; PP_03713	N/A
REGIÃO 3	complement(3913589..3913816)	hypothetical; PP_03716	N/A
REGIÃO 3	3913769..3915703	PHAGE_Acanth_1: hypothetical protein ATCV1_Z696R ; PP_03715; phage(gi155371643)	5,00E-80
REGIÃO 3	complement(3915863..3916237)	LysR family transcriptional regulator [Polaromonas sp. JS666] gi 91786878 ref YP_547830.1 ; E-VALUE = 2e-40; PP_03717	N/A
REGIÃO 3	3916265..3916447	hypothetical; PP_03718	N/A
REGIÃO 3	complement(3916444..3917640)	hypothetical protein AZOLI_p10754 [Azospirillum lipoferum 4B] gi 374998642 ref YP_004974141.1 ; E-VALUE = 4e-19; PP_03719	N/A
REGIÃO 3	complement(3917714..3918463)	PHAGE_Tricho_2c: hypothetical protein TNAV2c_gp071 ; PP_03720; phage(gi116326757)	8,00E-24
REGIÃO 3	3918444..3918563	hypothetical; PP_03721	N/A
REGIÃO 3	complement(3918585..3919502)	PHAGE_Burkho_phi1026b: gp58 ; PP_03722; phage(gi38707948)	3,00E-09
REGIÃO 3	3919707..3919865	hypothetical; PP_03723	N/A
REGIÃO 3	3919879..3921615	PHAGE_Pectob_My1: YadA domain-containing protein ; PP_03724; phage(gi410491156)	3,00E-22
REGIÃO 3	3921627..3922541	PHAGE_Amsact_ : similar to laminin ; PP_03725; phage(gi9964588)	2,00E-07
REGIÃO 3	3922683..3923510	PHAGE_Equid__4: envelope glycoprotein J ; PP_03726; phage(gi9629801)	7,00E-06
REGIÃO 3	3923584..3925293	PHAGE_Equid__9: envelope glycoprotein J ; PP_03727; phage(gi216905924)	5,00E-13
REGIÃO 3	3925376..3925756	flagellar protein FlIS [Rhodopseudomonas palustris TIE-1] gi 192289139 ref YP_001989744.1 ; E-VALUE = 4e-06; PP_03728	N/A
REGIÃO 3	3926049..3926684	PHAGE_Salini_CW02: RNA polymerase sigma factor ; PP_03729; phage(gi423261930)	6,00E-05
REGIÃO 4	5876016..5876627	PHAGE_Acanth_moumouvirus: 3'-5' exonuclease ; PP_05539; phage(gi441432317)	5,00E-06

REGIÃO 4	5876988..5877263	nitrogen fixation protein FixI (coupling cation pump) [Rhodospirillum centenum SW] gi 209964579 ref YP_002297494.1 ; E-VALUE = 1e-23; PP_05540	N/A
REGIÃO 4	5877260..5877424	nitrogen fixation protein FixS [Rhodospirillum centenum SW] gi 209964578 ref YP_002297493.1 ; E-VALUE = 4e-14; PP_05541	N/A
REGIÃO 4	5877450..5877665	hypothetical; PP_05542	N/A
REGIÃO 4	5877776..5878306	G:T/U mismatch-specific DNA glycosylase [Frateuria aurantia DSM 6220] gi 383316001 ref YP_005376843.1 ; E-VALUE = 1e-42; PP_05543	N/A
REGIÃO 4	5878496..5879719	PHAGE_Pseudo_MP1412: diguanylate-cyclase GGDEF domain ; PP_05544; phage(gi399529005)	1,00E-06
REGIÃO 4	5879883..5880185	flagellar motor switch protein FliN [Rhodospirillum centenum SW] gi 209964576 ref YP_002297491.1 ; E-VALUE = 6e-38; PP_05545	N/A
REGIÃO 4	5880237..5880680	thioesterase superfamily protein [Parvibaculum lavamentivorans DS-1] gi 154250947 ref YP_001411771.1 ; E-VALUE = 2e-10; PP_05546	N/A
REGIÃO 4	complement(5880715..5881707)	PHAGE_Equid__9: envelope glycoprotein J ; PP_05547; phage(gi216905924)	8,00E-15
REGIÃO 4	5881733..5882536	PHAGE_Human__4: EBNA-1 ; PP_05548; phage(gi82503233)	1,00E-13
REGIÃO 4	5882556..5884211	PROPHAGE_Mesorh_MAFF303099: transposase ; PP_05549; phage(gi13475318)	2,00E-87
REGIÃO 4	complement(5884251..5884373)	hypothetical protein RC1_0747 [Rhodospirillum centenum SW] gi 209964076 ref YP_002296991.1 ; E-VALUE = 3e-06; PP_05550	N/A
REGIÃO 4	5884673..5885293	PHAGE_Ectoca_1: EsV-1-65 ; PP_05551; phage(gi13242537)	1,00E-12
REGIÃO 4	complement(5885259..5885456)	hypothetical; PP_05553	N/A
REGIÃO 4	5885529..5886707	peptidase M20D, amidohydrolase [Rhodospirillum centenum SW] gi 209964324 ref YP_002297239.1 ; E-VALUE = 2e-167; PP_05552	N/A
REGIÃO 4	5886818..5887312	hypothetical protein RC1_1003 [Rhodospirillum centenum SW] gi 209964325 ref YP_002297240.1 ; E-VALUE = 2e-40; PP_05554	N/A
REGIÃO 4	complement(5887445..5889421)	PHAGE_Megavi_chiliensis: putative dTDP-d-glucose 4 6-dehydratase ; PP_05555; phage(gi363539870)	4,00E-19

**Tabela Suplementar 11 - Regiões genômicas virais em três espécies do gênero *Azospirillum* segundo PHAST.**

Região	Tamanho	Posição relativa	Fago	Espécie
1	27.1Kb	930016-957212	<i>Azospirillum</i> _phage_Cd	<i>AB510</i>
2	8.6Kb	2164469-2173163	Prophage <i>Shewanella oneidensis</i> MR-1	<i>AB510</i>
3	9.6Kb	2175576-2185264	<i>Pectobacterium</i> phage My1	<i>AB510</i>
4	27.1Kb	2320368-2347564	<i>Azospirillum</i> _phage_Cd	<i>AB510</i>
5	27.5Kb	2495644-2523226	<i>Vibrio</i> _phage_VHML	<i>AB510</i>
6	33.9Kb	2873221-2907198	<i>Burkholderia</i> _phage_BcepC6B	<i>AB510</i>
7	23.3Kb	3183844-3207208	Stx2_converting_phage_1717	<i>AB510</i>
8	48.5Kb	4234186-4282698	Prophage <i>Shewanella oneidensis</i> MR-1	<i>AB510</i>
9	19.4Kb	4904927-4924372	Prophage <i>Shewanella oneidensis</i> MR-1	<i>AB510</i>
10	21Kb	5443565-5464618	Stx2_converting_phage_1717	<i>AB510</i>
11	18Kb	7340499-7358526	<i>Megavirus chiliensis</i>	<i>AB510</i>
12	61.3Kb	7391540-7452908	Prophage <i>Escherichia coli</i> CFT073	<i>AB510</i>
1	16.7Kb	2872229-2888992	<i>Erwinia</i> phage phiEt88	Abra
2	13.9Kb	2892142-2906042	<i>Azospirillum</i> _phage_Cd	Abra
3	14.3Kb	5445134-5459498	<i>Pseudomonas</i> phage MP1412	Abra
4	29.6Kb	6303611-6333296	<i>Aggregatibacter</i> phage S1249	Abra
5	14.5Kb	6344412-6358954	<i>Rhodococcus</i> phage REQ2	Abra
1	31.4Kb	1654716-1686118	<i>Azospirillum</i> _phage_Cd	Alipo
2	19.8Kb	3422085-3441950	<i>Burkholderia</i> phage AH2	Alipo
3	17.2Kb	3768597-3785879	Prophage <i>Mesorhizobium loti</i> MAFF303099	Alipo
4	15.1Kb	4045505-4060656	<i>Pseudomonas</i> _phage_F10	Alipo

*AB510* - *Azospirillum* sp. B510, Abra – *Azospirillum brasilense*, Alipo – *Azospirillum lipoferum*.

**Tabela Suplementar 12 - Identificação das proteínas de *A. amazonense* localizadas nas quatro regiões virais detectadas pelo PHAST.**

ID	Acesso Amz	Produto proteico	ID proteína viral ACLAME	Valor E	Melhor Similaridade	Região Viral
Amz_0778	ZP_08868127.1	phage associated protein	protein:proph:161424	4,00E-98	YP_001833785.1	REGIÃO1
Amz_0779	ZP_08868128.1	phage associated protein	protein:proph:161425	3,00E-27	YP_001833786.1	REGIÃO1
Amz_0780	ZP_08868129.1	TIGR02218 family protein	protein:proph:161429	9,00E-36	ZP_12217625.1	REGIÃO1
Amz_0781	ZP_08868130.1	Conserved Hypothetical protein Msil 3055	protein:proph:160885	2,00E-14	YP_002363327.1	REGIÃO1
Amz_0782	ZP_08868131.1	Phage-related minor tail protein	protein:proph:166839	1,00E-94	YP_421104.1	REGIÃO1
Amz_0783		Conserved_216_Hypothetical protein amb1775	protein:proph:180288	0,12	YP_421138.1	REGIÃO1
Amz_0784		Conserved Hypothetical protein Msil 3050	protein:proph:161411	9,00E-39	YP_002363322.1	REGIÃO1
Amz_0786		Conserved_100_Hypothetical protein amb1772	protein:vir:488	0,002	YP_421135.1	REGIÃO1
Amz_0787		Conserved Hypothetical protein amb1771	protein:proph:166869	2,00E-26	YP_421134.1	REGIÃO1
Amz_0788		Mu-like prophage protein gp36	protein:proph:166868	1,00E-26	YP_421133.1	REGIÃO1
Amz_0789		Conserved_057_Hypothetical protein Daes 2401	protein:proph:164051	2,00E-07	YP_004122152.1	REGIÃO1
Amz_0790	ZP_08868133.1	Mu-like prophage major head subunit gpT cl01826	protein:proph:166866	5,00E-82	ZP_02468360.1	REGIÃO1
Amz_0791	ZP_08868134.1	Conserved Hypothetical protein HPDFL43 14877	protein:proph:166865	3,00E-32	ZP_02165804.1	REGIÃO1
Amz_0793		urea ABC transporter, urea binding protei	protein:proph:177300	0,65	YP_002362311.1	REGIÃO1
Amz_0794	ZP_08868135.1	inner-membrane translocator	protein:vir:105507	1,00E-04	YP_002362310.1	REGIÃO1
Amz_0795	ZP_08868136.1	urea ABC transporter, permease protein Ur	protein:vir:78123	0,011	YP_002362309.1	REGIÃO1
Amz_2839	ZP_08868137.1	twocomponent sensor histidine kinase	protein:proph:172524	2,00E-13	YP_003447672.1	REGIÃO2
Amz_2840	ZP_08869315.1	DNAbinding twocomponent response regulator	protein:proph:172523	0,042	YP_003447673.1	REGIÃO2
Amz_2842	ZP_08869317.1	Conserved_090_Hypothetical protein Rleg5DRAFT 2793	protein:proph:178417	0,8	ZP_18294977.1	REGIÃO2
Amz_2843	ZP_08869317.1	Conserved Hypothetical protein amb0447	protein:vir:79221	6,00E-64	YP_419810.1	REGIÃO2
Amz_2844		Conserved Hypothetical protein amb0448	protein:proph:162143	6,00E-26	YP_419811.1	REGIÃO2
Amz_2845		Conserved Hypothetical protein amb0449	protein:proph:183983	7,00E-19	YP_419812.1	REGIÃO2
Amz_2847		Conserved_008_Hypothetical protein IEA 05636	protein:proph:172482	0,002	ZP_17422212.1	REGIÃO2
Amz_2848	ZP_08869318.1	Mulike prophage hostnuclease inhibitor protein Gam COG4396 cl01825	protein:proph:164843	2,00E-12	YP_004973881.1	REGIÃO2
Amz_2849	ZP_08869319.1	membranebound metallopeptidase	protein:proph:184192	6,00E-14	YP_002296805.1	REGIÃO2
Amz_2850	ZP_08868486.1	carboxyterminal processing protease CtpA	protein:proph:169311	0,096	YP_002296804.1	REGIÃO2

Amz_2851	ZP_08869320.1	tonB-dependent Receptor Plug domain protei	protein:vir:98205	0,06	ZP_08388415.1	REGIÃO2
Amz_2852	ZP_08869321.1	tryptophan halogenase	protein:vir:103094	4,00E-47	YP_497156.1	REGIÃO2
Amz_3404	ZP_08869322.1	Branched-chain amino acid transport ATP-bi	protein:proph:183608	1,00E-14	ZP_10897615.1	REGIÃO3
Amz_3405	ZP_08869668.1	branched-chain amino acid transp	protein:proph:183608	5,00E-18	YP_005031567.1	REGIÃO3
Amz_3406	ZP_08869669.1	periplasmic substrate binding protein cl10011	protein:vir:97985	0,13	NP_767638.1	REGIÃO3
Amz_3407	ZP_08869670.1	potassium transporter	protein:vir:105507	8,00E-04	YP_005039024.1	REGIÃO3
Amz_3410		conserved hypothetical protein AZOLip10754	protein:vir:108208	0,098	YP_004974141.1	REGIÃO3
Amz_3411	ZP_08869672.1	short chain dehydrogenase	protein:proph:178504	5,00E-31	YP_472760.1	REGIÃO3
Amz_3412	ZP_08870075.1	LysR family transcriptional regulator	protein:proph:182117	3,00E-42	YP_001831930.1	REGIÃO3
Amz_3414		flagellar hookassociated protein FlgKlike	protein:vir:2603	1,00E-13	YP_005041056.1	REGIÃO3
Amz_3415	ZP_08869674.1	flagellar hook-associated protein FlgL	protein:proph:181296	3,00E-06	YP_530764.1	REGIÃO3
Amz_3416	ZP_08869675.1	flagellin	protein:proph:182194	2,00E-07	YP_004986618.1	REGIÃO3
Amz_3417	ZP_08869676.1	flagellar hookassociated protein FliDlike	protein:vir:105028	2,00E-06	YP_004986617.1	REGIÃO3
Amz_3418	ZP_08869677.1	Conserved_158_Hypothetical protein RPA0642 COG1516 K02422	protein:proph:183284	0,055	NP_945995.1	REGIÃO3
Amz_3419	ZP_08869678.1	RNA polymerase ECF sigma factor sigma70 fragment partial	protein:proph:171882	5,00E-15	YP_004986615.1	REGIÃO3
Amz_5145	ZP_08869679.1	ribonuclease D	protein:proph:170425	3,00E-27	YP_002299373.1	REGIÃO4
Amz_5146	ZP_08869693.1	nitrogen fixation protein Fixl coupling cation pump	protein:proph:184995	0,46	YP_002297494.1	REGIÃO4
Amz_5150		signal transduction protein	protein:proph:172141	1,00E-22	ZP_18218164.1	REGIÃO4
Amz_5151	ZP_08870726.1	flagellar motor switch protein FliN	protein:vir:78431	0,8	YP_002297491.1	REGIÃO4
Amz_5152	ZP_08870247.1	Conserved_010_thioesterase superfamily protein	protein:proph:161383	0,007	YP_001411771.1	REGIÃO4
Amz_5153		Large exoprotein involved in heme utiliza	protein:vir:7738	1,00E-08	YP_005417612.1	REGIÃO4
Amz_5154	ZP_08870727.1	glycinerich protein_Ath006	protein:proph:167396	4,00E-08	NP_178626.1	REGIÃO4
Amz_5155	ZP_08869832.1	transposase	protein:proph:169319	5,00E-05	NP_106881.1	REGIÃO4
Amz_5156	ZP_08870728.1	acyl carrier protein	protein:proph:169431	4,00E-05	ZP_11252058.1	REGIÃO4
Amz_5157	ZP_08870729.1	sensor histidine kinase	protein:proph:172524	3,00E-10	YP_002297238.1	REGIÃO4
Amz_5159		peptidase M20D amidohydrolase	protein:proph:160560	2,00E-27	YP_002297239.1	REGIÃO4
Amz_5160	ZP_08869444.1	Conserved hypothetical protein RC1 1003	protein:vir:95798	0,79	YP_002297240.1	REGIÃO4
Amz_5161		polysaccharide biosynthesis protein CapD	protein:proph:167081	1,00E-94	YP_002297691.1	REGIÃO4



**Tabela Suplementar 13** – Proteínas associadas ao sistema imune CRISPR identificadas em *A. amazonense* pela ferramenta CRISPRs-finder.

ID	Acesso	Produto Proteico	COG	CDD	TIGR	Melhor Similaride
Amz_3760		CRISPR-associated protein Cas5, subtype I-C/DVULG		cl12040	TIGR01876	ZP_07992762.1
Amz_3761		CRISPR-associated protein Cas8c/Csd1, subtype I-C/DVULG		cl12004	TIGR01863	YP_004048357.1
Amz_3762		CRISPR-associated protein Cas7/Csd2, subtype I-C/DVULG	COG3649	cl01465	TIGR02589	YP_425919.1
Amz_3763		CRISPR-associated protein Cas4	COG1468	cl00641	TIGR00372	YP_002800297.1
Amz_3764		CRISPR-associated endonuclease Cas1, subtype I-C/DVULG	COG1518	cl00656	TIGR03640	YP_001600355.1
Amz_3765		CRISPR-associated endoribonuclease Cas2	COG1343	cl11442	TIGR01573	ZP_09953215.1
Amz_5447	ZP_08870897.1	CRISPR-associated helicase Cas3	COG1203		TIGR01587	YP_425432.1
Amz_5448	ZP_08870898.1	CRISPR-associated_Cse1_family_protein				YP_425433.1
Amz_5449		CRISPR-associated_Cse2_family_protein				YP_425435.1
Amz_5450	ZP_08870899.1	CRISPR-associated_Cse4_family_protein		cl09608	TIGR01869	YP_425436.1
Amz_5451		CRISPR-associated_Cse3_family_protein		cl08497		YP_425438.1
Amz_5452	ZP_08870901.1	CRISPR-associated_Cas1_family_protein	COG1518	cl00656		YP_425439.1

**Tabela Suplementar 14** – A figura apresenta os principais domínios da proteína relacionada a fagos de *A. amazonense* identificada por RDA.

Descrição	Multi Domínios	Valor E
COG2856[COG2856], Predicted Zn peptidase [Amino acid transport	não	1,90E-15
HTH_XRE[cd00093], Helix-turn-helix XRE-family like proteins.	não	1,35E-10
HTH_19[pfam12844], Helix-turn-helix domain	não	4,17E-10
HTH_XRE[smart00530], Helix-turn-helix XRE-family like proteins;	não	2,01E-09
HTH_3[pfam01381], Helix-turn-helix	não	1,16E-08
DUF955[pfam06114], Domain of unknown function (DUF955)	não	5,74E-08
HTH_31[pfam13560], Helix-turn-helix domain	não	1,19E-06
HipB[COG1396], Predicted transcriptional regulators [Transcription]	sim	1,78E-06
COG2944[COG2944], Predicted transcriptional regulator [Transcription]	sim	1,11E-04
HTH_26[pfam13443], Cro/C1-type HTH DNA-binding domain;	não	2,78E-03

## 7. Referências Bibliográficas

1. ALEXANDRE, G.; ROHR, R.; BALLY, R. A phase variant of *Azospirillum lipoferum* lacks a polar flagellum and constitutively expresses mechanosensing lateral flagella. *Appl Environ Microbiol*, v. 65, n. 10, p. 4701-4, 1999.
2. ALLEN, N. L.; PENN, C. W.; HILTON, A. C. Representational difference analysis: critical appraisal and method development for the identification of unique DNA sequences from prokaryotes. *Journal of Microbiological Methods*, v. 55, n. 1, p. 73-81, 2003.
3. ALTSCHUL, S. F.; MADDEN, T.L.; SCHÄFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, v. 25, n. 17, p. 3389-402, 1997.
4. BALDOCCHI, R. A.; FLAHERTY, L. Isolation of Genomic Fragments from Polymorphic Regions by Representational Difference Analysis. *Methods*, v. 13, n. 4, p. 337-346, 1997.
5. BARTEL, B. Auxin biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, v. 48, n. 1, p. 51-66, 1997.
6. BASHAN, Y.; HOLGUIN, G.; DE-BASHAN, L. E. *Azospirillum*-plant relationships: physiological, molecular, agricultural, and environmental advances (1997-2003). *Can J Microbiol*, v. 50, n. 8, p. 521-77, 2004.
7. BI, D.; LIU, L.; TAI, C.; DENG, Z.; RAJAKUMAR, K.; OU, H. Y. SecReT4: a web-based bacterial type IV secretion system resource. *Nucleic Acids Res*, v. 41, n. Database issue, p. D660-5, 2013.
8. BOTTINI, R.; CASSÁN, F.; PICCOLI, P. Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. *Applied Microbiology and Biotechnology*, v. 65, n. 5, p. 497-503, 2004.
9. BOYER, M.; HAURAT, J.; SAMAIN, S.; SEGURENS, B.; GAVORY, F.; GONZÁLEZ, V.; MAVINGUI, P.; ROHR, R.; BALLY, R.; WISNIEWSKI-DYÉ, F. Bacteriophage prevalence in the genus *Azospirillum* and analysis of the first genome sequence of an *Azospirillum brasilense* integrative phage. *Appl Environ Microbiol*, v. 74, n. 3, p. 861-74, 2008.
10. BRUZEL, A.; CHEUNG, V. G. DNA reassociation using oscillating phenol emulsions. *Genomics*, v. 87, n. 2, p. 286-9, 2006.
11. CAIOLA, M. G.; CANINI, A.; BOTTA, A. L.; DEL, GALLO, M. Localization of *Azospirillum brasilense* Cd in inoculated tomato (*Lycopersicon esculentum* Mill.) roots. *Ann microbiol*, v. 54, p. 365-380, 2004.
12. CARREÑO-LOPEZ, R.; CAMPOS-REALES, N.; ELMERICH, C.; BACA, B. Physiological evidence for differently regulated tryptophan-dependent pathways for indole-3-acetic acid synthesis in *Azospirillum brasilense*. *Mol Gen Genet*, v. 264, n. 4, p. 521-30, 2000.

13. CHAUDHURI, R. R.; LOMAN, N. J.; SNYDER, L. A.; BAILEY, C. M.; STEKEL, D. J.; PALLEEN, M. J. xBASE2: a comprehensive resource for comparative bacterial genomics. *Nucleic Acids Res*, v. 36, n. Database issue, p. D543-6, 2008.
14. CHEN, Q.; ZHANG, C.; CHEN, Y.; LOU, J.; WANG, D. Identification of endometriosis-related genes by representational difference analysis of cDNA. *Aust N Z J Obstet Gynaecol*, v. 52, n. 2, p. 140-5, 2012.
15. CHERNYS, J. T.; ZEEVAART, J. A. Characterization of the 9-cis-epoxycarotenoid dioxygenase gene family and the regulation of abscisic acid biosynthesis in avocado. *Plant Physiol*, v. 124, n. 1, p. 343-53, 2000.
16. CHO, H.; PINTO, U. M.; WINANS, S. C. Transsexuality in the rhizosphere: quorum sensing reversibly converts *Agrobacterium tumefaciens* from phenotypically female to male. *J Bacteriol*, v. 191, n. 10, p. 3375-83, 2009.
17. COHEN, A.; BOTTINI, R.; PICCOLI, P. *Azospirillum brasilense* Sp 245 produces ABA in chemically-defined culture medium and increases ABA content in arabidopsis plants. *Plant Growth Regulation*, v. 54, n. 2, p. 97-103, 2008/03/01 2008.
18. EADY, R. R.; POSTGATE, J. R. Nitrogenase. *Nature*, v. 249, p. 805-810, 1974.
19. FAGANELLO, J.; DUTRA, V.; SCHRANK, A.; MEYER, W.; SCHRANK, I. S.; VAINSTEIN, M. H. Identification of genomic differences between *Cryptococcus neoformans* and *Cryptococcus gattii* by Representational Difference Analysis (RDA). *Med Mycol*, v. 47, n. 6, p. 584-91, 2009.
20. FELSKA, A. Streamlined representational difference analysis for comprehensive studies of numerous genomes. *J Microbiol Methods*, v. 50, n. 3, p. 305-11, 2002.
21. FINN, R. D.; MISTRY, J.; TATE, J.; COGGILL, P.; HEGER, A.; POLLINGTON, J. E.; GAVIN, O. L.; GUNASEKARAN, P.; CERIC, G.; FORSLUND, K.; HOLM, L.; SONNHAMMER, E. L.; EDDY, S. R.; BATEMAN, A. The Pfam protein families database. *Nucleic Acids Res*, v. 38, n. Database issue, p. D211-22, 2010.
22. GHIMIRE, G. P.; LEE, H. C.; SOHNG, J. K. Improved squalene production via modulation of the methylerythritol 4-phosphate pathway and heterologous expression of genes from *Streptomyces peucetius* ATCC 27952 in *Escherichia coli*. *Appl Environ Microbiol*, v. 75, n. 22, p. 7291-3, 2009.
23. GRISSA, I.; VERGNAUD, G.; POURCEL, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*, v. 35, n. Web Server issue, p. W52-7, 2007.

24. GUNNELL, D.; EDDLESTON, M.; PHILLIPS, M. R.; KONRADSEN, F. The global distribution of fatal pesticide self-poisoning: systematic review. *BMC Public Health*, v. 7, p. 357, 2007.
25. HARTMANN, A.; FU, H. A; SONG, S. D; BURRIS, R. H; KLINGMULLER, W. Comparison of nitrogenase regulation in *Azospirillum brasilense*, *A. lipoferum* and *A. amazonense*. In: (Ed.). *Azospirillum III. Genetics, physiology, ecology*. Berlin: Springer-Verlag, 1985. p.116-126.
26. HARTMANN, A.; HUREK, T. Effect of carotenoid overproduction on oxygen tolerance of nitrogen fixation in *Azospirillum brasilense* Sp7. *Journal of General Microbiology*, v. 134, p. 2449-2455, 1988.
27. HOLGUIN, G.; GLICK, B. Transformation of *Azospirillum brasilense* Cd with an ACC deaminase gene from *enterobacter cloacae* UW4 fused to the Tet r gene promoter improves its fitness and plant growth promoting ability. *Microb Ecol*, v. 46, n. 1, p. 122-33, 2003.
28. HUMANN, F. C.; HARTFELDER, K. Representational Difference Analysis (RDA) reveals differential expression of conserved as well as novel genes during caste-specific development of the honey bee (*Apis mellifera* L.) ovary. *Insect Biochem Mol Biol*, v. 41, n. 8, p. 602-12, 2011.
29. ISAWA, T.; YASUDA, M.; AWAZAKI, H.; MINAMISAWA, K.; SHINOZAKI, S.; NAKASHITA, H. *Azospirillum* sp. strain B510 enhances rice growth and yield. *Microbes Environ*, v. 25, n. 1, p. 58-61, 2010.
30. JANITZ, M. Assigning functions to genes--the main challenge of the post-genomics era. *Rev Physiol Biochem Pharmacol*, v. 159, p. 115-29, 2007.
31. KANEHISA, M. The KEGG database. *Novartis Found Symp*, v. 247, p. 91-101; discussion 101-3, 119-28, 244-52, 2002.
32. KANEKO, T.; MINAMISAWA, K.; ISAWA, T.; NAKATSUKASA, H.; MITSUI, H.; KAWAHARADA, Y.; NAKAMURA, Y.; WATANABE, A.; KAWASHIMA, K.; ONO, A.; SHIMIZU, Y.; TAKAHASHI, C.; MINAMI, C.; FUJISHIRO, T.; KOHARA, M.; KATOH, M.; NAKAZAKI, N.; NAKAYAMA, S.; YAMADA, M.; TABATA, S.; SATO, S. Complete genomic structure of the cultivated rice endophyte *Azospirillum* sp. B510. *DNA Res*, v. 17, n. 1, p. 37-50, 2010.
33. KARPATI, E.; KISS, P.; PONYI, T.; FENDRIK, I.; DE, ZAMAROCZY, M.; OROSZ, L. Interaction of *Azospirillum lipoferum* with wheat germ agglutinin stimulates nitrogen fixation. *J Bacteriol*, v. 181, n. 13, p. 3949-55, 1999.
34. KIM, J.; REES, D. C. Nitrogenase and Biological Nitrogen Fixation. *Biochemistry*, v. 33, p. 389-397, 1994.

35. KLOTZ, M. G.; LOEWEN, P. C. The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. *Mol Biol Evol*, v. 20, n. 7, p. 1098-112, 2003.
36. KOEPKE, T.; SCHAEFFER, S.; HARPER, A.; DICENTA, F.; EDWARDS, M.; HENRY, R. J.; MØLLER, B. L.; MEISEL, L.; ORAGUZIE, N.; SILVA, H.; SÁNCHEZ-PÉREZ, R.; DHINGRA, A. Comparative genomics analysis in Prunoideae to identify biologically relevant polymorphisms. *Plant Biotechnol J*, 2013.
37. LANGE, A.; MOREIRA, F. M. S. Detecção de *Azospirillum amazonense* em raízes e rizosfera de Orchidaceae e de outras famílias vegetais. *Bras. Ci. Solo*, v. 26, p. 529-533, 2002.
38. LEPLAE, R.; HEBRANT, A.; WODAK, S. J.; TOUSSAINT, A. ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res*, v. 32, n. Database issue, p. D45-9, 2004.
39. LIN, S. Y.; LIU, Y. C.; HAMEED, A.; HSU, Y. H.; LAI, W. A.; SHEN, F. T.; YOUNG, C. C. *Azospirillum fermentarium* sp. nov., a novel nitrogen-fixing species isolated from a fermenter in Taiwan. *Int J Syst Evol Microbiol*, 2013.
40. LIN, S. Y.; SHEN, F. T.; YOUNG, L. S.; ZHU, Z. L.; CHEN, W. M.; YOUNG, C. C. *Azospirillum formosense* sp. nov., a diazotroph from agricultural soil. *Int J Syst Evol Microbiol*, v. 62, n. Pt 5, p. 1185-90, 2012.
41. LISITSYN, N.; LISITSYN, N.; WIGLER, M. Cloning the differences between two complex genomes. *Science*, v. 259, n. 5097, p. 946-951, 1993.
42. LIU, P.; LI, P.; JIANG, X.; BI, D.; XIE, Y.; TAI, C.; DENG, Z.; RAJAKUMAR, K.; OU, H. Y. Complete genome sequence of *Klebsiella pneumoniae* subsp. *pneumoniae* HS11286, a multidrug-resistant strain isolated from human sputum. *J Bacteriol*, v. 194, n. 7, p. 1841-2, 2012.
43. LU, Y. K.; MARDEN, J.; HAN, M.; SWINGLEY, W. D.; MASTRIAN, S. D.; CHOWDHURY, S. R.; HAO, J.; HELMY, T.; KIM, S.; KURDOGLU, A. A.; MATTHIES, H. J.; ROLLO, D.; STOTHARD, P.; BLANKENSHIP, R. E.; BAUER, C. E.; TOUCHMAN, J. W. Metabolic flexibility revealed in the genome of the cyst-forming alpha-1 proteobacterium *Rhodospirillum centenum*. *BMC Genomics*, v. 11, p. 325, 2010.
44. MA, Z.; SHEN, X.; HU, H.; WANG, W.; PENG, H.; XU, P.; ZHANG, X. Genome sequence of *Sphingomonas wittichii* DP58, the first reported phenazine-1-carboxylic acid-degrading strain. *J Bacteriol*, v. 194, n. 13, p. 3535-6, 2012.
45. MA'AYEH, S. Y.; BROOK-CARTER, P. T. Representational difference analysis identifies specific genes in the interaction of *Giardia duodenalis* with the murine intestinal epithelial cell line, IEC-6. *Int J Parasitol*, v. 42, n. 5, p. 501-9, 2012.

46. MAGALHÃES, F. M.; BALDANI, J. I.; SOUTO, S. M.; KUYKENDALL, J. R.; DOBEREINER, J. A new acid-tolerant *Azospirillum* species. *Anais da Academia Brasileira de Ciências*, v. 55, p. 417-430, 1983.
47. MARASCO, E. K.; SCHMIDT-DANNERT, C. Identification of bacterial carotenoid cleavage dioxygenase homologues that cleave the interphenyl alpha,beta double bond of stilbene derivatives via a monooxygenase reaction. *ChemBiochem*, v. 9, n. 9, p. 1450-61, 2008.
48. MARCHLER-BAUER, A.; ZHENG, C.; CHITSAZ, F.; DERBYSHIRE, M. K.; GEER, L. Y.; GEER, R. C.; GONZALES, N. R.; GWADZ, M.; HURWITZ, D. I.; LANCZYCKI, C. J.; LU, F.; LU, S.; MARCHLER, G. H.; SONG, J. S.; THANKI, N.; YAMASHITA, R. A.; ZHANG, D.; BRYANT, S. H. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res*, v. 41, n. Database issue, p. D348-52, 2013.
49. MARTÍNEZ-DRETS, G.; FABIANO, E.; CARDONA, A. Carbohydrate Catabolism in *Azospirillum amazonense*. *Appl Environ Microbiol*, v. 50, n. 1, p. 183-5, 1985.
50. MARÍN, M. M.; YUSTE, L.; ROJO, F. Differential expression of the components of the two alkane hydroxylases from *Pseudomonas aeruginosa*. *J Bacteriol*, v. 185, n. 10, p. 3232-7, 2003.
51. MAZURIE, A. J.; ALVES, J. M.; OZAKI, L. S.; ZHOU, S.; SCHWARTZ, D. C.; BUCK, G. A. Comparative genomics of cryptosporidium. *Int J Genomics*, v. 2013, p. 832756, 2013.
52. MELA, F.; FRITSCHKE, K.; DE BOER, W.; VAN DEN BERG, M.; VAN VEEN, J. A.; MAHARAJ, N. N.; LEVEAU, J. H. Comparative genomics of bacteria from the genus *Collimonas*: linking (dis)similarities in gene content to phenotypic variation and conservation. *Environ Microbiol Rep*, v. 4, n. 4, p. 424-32, 2012.
53. MOKROUSOV, I.; LIMESCHENKO, E.; VYAZOVAYA, A.; NARVSKAYA, O. *Corynebacterium diphtheriae* spoligotyping based on combined use of two CRISPR loci. *Biotechnol J*, v. 2, n. 7, p. 901-6, 2007.
54. MORIYA, Y.; ITOH, M.; OKUDA, S.; YOSHIZAWA, A. C.; KANEHISA, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*, v. 35, n. Web Server issue, p. W182-5, 2007.
55. MOUNT, S. M.; CHANG, C. Evidence for a plastid origin of plant ethylene receptor genes. *Plant Physiol*, v. 130, n. 1, p. 10-4, 2002.
56. NUR, I.; STEINITZ, Y. L.; OKON, Y.; HENIS, Y. Carotenoid Composition and Function in Nitrogen-fixing Bacteria of the Genus *Azospirillum*. *Journal of General Microbiology*, v. 122, p. 27-32, 1981.
57. PEARSON, K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it

can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, v. 50, n. 302, p. 157-175, 1900/07/01 1900.

58. PEDROSA, F. O.; HUNGRIA, M.; YATES, M. G.; NEWTON, W. E. *Nitrogen Fixation: From Molecules to Crop Productivity*. Kluwer Academic Publishers, 2000.
59. PENG, Z. Y.; ZHOU, X.; LI, L.; YU, X.; LI, H.; JIANG, Z.; CAO, G.; BAI, M.; WANG, X.; JIANG, C.; LU, H.; HOU, X.; QU, L.; WANG, Z.; ZUO, J.; FU, X.; SU, Z.; LI, S.; GUO, H. Arabidopsis Hormone Database: a comprehensive genetic and phenotypic information database for plant hormone research in Arabidopsis. *Nucleic Acids Res*, v. 37, n. Database issue, p. D975-82, 2009.
60. PRIGENT-COMBARET, C.; BLAHA, D.; POTHIER, J. F.; VIAL, L.; POIRIER, M-A.; WISNIEWSKI-DYÉ, F.; MOËNNE-LOCCOZ, Y. Physical organization and phylogenetic analysis of *acdR* as leucine-responsive regulator of the 1-aminocyclopropane-1-carboxylate deaminase gene *acdS* in phytobeneficial *Azospirillum lipoferum* 4B and other Proteobacteria. *FEMS Microbiol Ecol*, v. 65, n. 2, p. 202-19, 2008.
61. PUIU, D.; ENOMOTO, S.; BUCK, G. A.; ABRAHAMSEN, M. S.; KISSINGER, J. C. CryptoDB: the Cryptosporidium genome resource. *Nucleic Acids Res*, v. 32, n. Database issue, p. D329-31, 2004.
62. QI, M.; WANG, D.; BRADLEY, C. A.; ZHAO, Y. Genome sequence analyses of *Pseudomonas savastanoi* pv. *glycinea* and subtractive hybridization-based comparative genomics with nine pseudomonads. *PLoS One*, v. 6, n. 1, p. e16451, 2011.
63. REIS, JUNIOR, F. B.; SILVA, M. F.; TEIXEIRA, K. R. S.; URQUIAGA, S.; REIS, V. M. Identificação de isolados de *Azospirillum amazonense* associados a *Brachiaria* spp., em diferentes épocas e condições de cultivo e produção de fitormônio pela bactéria. *Revista Brasileira de Ciência do Solo*, v. 28, p. 103-113, 2004.
64. REIS, V.; TEIXEIRA, K.; PEDRAZA, R. What Is Expected from the Genus *Azospirillum* as a Plant Growth-Promoting Bacteria? In: MAHESHWARI, D. K. (Ed.). *Bacteria in Agrobiolgy: Plant Growth Responses*: Springer Berlin Heidelberg, 2011. cap. 6, p.123-138.
65. REIS, V. M.; OLIVEIRA, A. L. M.; SILVA, M. F.; OLIVARES, F. L.; BALDANI, J. I.; BODDEY, R. M.; URQUIAGA, S. Inoculants for Sugar Cane: The Scientific Bases for the Adoption of the Technology for Biofuel Production. In: DAKORA, F.; CHIMPHANGO, S. M., *et al* (Ed.). *Biological Nitrogen Fixation: Towards Poverty Alleviation through Sustainable Agriculture*: Springer Netherlands, v.42, 2008. cap. 21, p.67-68. (Current Plant Science and Biotechnology in Agriculture).
66. RODRIGUES, E. P.; RODRIGUES, L. S.; OLIVEIRA, A. L. M.; BALDANI, V. L. D.; TEIXEIRA, K. R. D. S.; URQUIAGA, S.; REIS, V. M. *Azospirillum*



amazonense inoculation: effects on growth, yield and N<sub>2</sub> fixation of rice (*Oryza sativa* L.). *Plant and Soil*, v. 302, n. 1-2, p. 249-261, 2008/01/01 2008.

67. SAIER, M. H.; TRAN, C. V.; BARABOTE, R. D. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res*, v. 34, n. Database issue, p. D181-6, 2006.
68. SAMBROOK, J.; RUSSEL, D. W. *Molecular cloning: A laboratory manual*. New York: Cold Spring Harbor Lab. Press, 2001.
69. SANDKVIST, M. Type II secretion and pathogenesis. *Infect Immun*, v. 69, n. 6, p. 3523-35, 2001.
70. SANGAL, V.; TUCKER, N. P.; BURKOVSKI, A.; HOSKISSON, P. A. The draft genome sequence of *Corynebacterium diphtheriae* bv. mitis NCTC 3529 reveals significant diversity between the primary disease-causing biovars. *J Bacteriol*, v. 194, n. 12, p. 3269, 2012.
71. SANT'ANNA, F. H.; ALMEIDA, L. G.; CECAGNO, R.; REOLON, L. A.; SIQUEIRA, F. M.; MACHADO, M. R.; VASCONCELOS, A. T.; SCHRANK, I. S. Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*. *BMC Genomics*, v. 12, p. 409, 2011.
72. SATO, T.; MISHINA, M. Representational difference analysis, high-resolution physical mapping, and transcript identification of the zebrafish genomic region for a motor behavior. *Genomics*, v. 82, n. 2, p. 218-229, 2003.
73. SCHRANK, A.; VAINSTEIN, M. H. *Metarhizium anisopliae* enzymes and toxins. *Toxicon*, v. 56, n. 7, p. 1267-74, 2010.
74. SHAO, Y.; HE, X.; HARRISON, E. M.; TAI, C.; OU, H. Y.; RAJAKUMAR, K.; DENG, Z. mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes. *Nucleic Acids Res*, v. 38, n. Web Server issue, p. W194-200, 2010.
75. SILVA, M. F.; SOUZA, ANTÔNIO, C.; OLIVEIRA, P.; XAVIER, G.; RUMJANEK, N.; BARROS SOARES, L.; REIS, V. Survival of endophytic bacteria in polymer-based inoculants and efficiency of their application to sugarcane. *Plant and Soil*, v. 356, n. 1-2, p. 231-243, 2012/07/01 2012.
76. SOMERS, E.; KEIJERS, V.; PTACEK, D.; HALVORSEN, OTTOY, M.; SRINIVASAN, M.; VANDERLEYDEN, J.; FAURE, D. The salCAB operon of *Azospirillum irakense*, required for growth on salicin, is repressed by SalR, a transcriptional regulator that belongs to the LacI/GalR family. *Mol Gen Genet*, v. 263, n. 6, p. 1038-46, 2000.
77. SPAEPEN, S.; VERSÉES, W.; GOCKE, D.; POHL, M.; STEYAERT, J.; VANDERLEYDEN, J. Characterization of phenylpyruvate

decarboxylase, involved in auxin production of *Azospirillum brasilense*. *J Bacteriol*, v. 189, n. 21, p. 7626-33, 2007.

78. STADEN, R.; BEAL, K. F.; BONFIELD, J. K. The Staden package, 1998. *Methods Mol Biol*, v. 132, p. 115-30, 2000.
79. TAN, B. C.; JOSEPH, L. M.; DENG, W. T.; LIU, L.; LI, Q. B.; CLINE, K.; MCCARTY, D. R. Molecular characterization of the *Arabidopsis* 9-cis epoxy-carotenoid dioxygenase gene family. *Plant J*, v. 35, n. 1, p. 44-56, 2003.
80. TATUSOV, R. L.; GALPERIN, M. Y.; NATALE, D. A.; KOONIN, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*, v. 28, n. 1, p. 33-6, 2000.
81. TIEN, T. M.; GASKINS, M. H.; HUBBELL, D. H. Plant growth substances produced by *Azospirillum brasilense* and their effect on growth of pearl millet ( *Pennisetum americanum* ). *Applied and Environmental Microbiology*, v. 37, p. 1016-1024, 1979.
82. TINSLEY, C. R.; NASSIF, X. Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. *Proc Natl Acad Sci U S A*, v. 93, n. 20, p. 11109-14, 1996.
83. TIPPMANN, H. F. Analysis for free: comparing programs for sequence analysis. *Brief Bioinform*, v. 5, n. 1, p. 82-7, 2004.
84. VAN PUYVELDE, S.; CLOOTS, L.; ENGELEN, K.; DAS, F.; MARCHAL, K.; VANDERLEYDEN, J.; SPAEPEN, S. Transcriptome analysis of the rhizosphere bacterium *Azospirillum brasilense* reveals an extensive auxin response. *Microb Ecol*, v. 61, n. 4, p. 723-8, 2011.
85. YASUDA, M. ISAWA, T. SHINOZAKI, S. MINAMISAWA, K. NAKASHITA, H. Effects of colonization of a bacterial endophyte, *Azospirillum* sp. B510, on disease resistance in rice. *Biosci Biotechnol Biochem*, v. 73, n. 12, p. 2595-9, 2009.
86. WINTJENS, R.; BELRHALI, H.; CLANTIN, B.; AZARKAN, M.; BOMPARD, C.; BAEYENS-VOLANT, D.; LOOZE, Y.; VILLERET, V. Crystal structure of papaya glutaminyl cyclase, an archetype for plant and bacterial glutaminyl cyclases. *J Mol Biol*, v. 357, n. 2, p. 457-70, 2006.
87. WISNIEWSKI-DYÉ, F.; BORZIAK, K.; KHALSA-MOYERS, G.; ALEXANDRE, G.; SUKHARNIKOV, L. O.; WUICHET, K.; HURST, G. B.; MCDONALD, W. H.; ROBERTSON, J. S.; BARBE, V. et al. *Azospirillum* genomes reveal transition of bacteria from aquatic to terrestrial environments. *PLoS Genet*, v. 7, n. 12, p. e1002430, 2011.
88. WU, S.; ZHU, Z.; FU, L.; NIU, B.; LI, W. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, v. 12, p. 444, 2011.

89. ZHANG, R.; LIN, Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res*, v. 37, n. Database issue, p. D455-8, 2009.
90. ZHAO, Y. Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants. *Mol Plant*, v. 5, n. 2, p. 334-8, 2012.
91. ZHAO, Y.; CHRISTENSEN, S. K.; FANKHAUSER, C.; CASHMAN, J. R.; COHEN, J. D.; WEIGEL, D.; CHORY, J. A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science*, v. 291, n. 5502, p. 306-9, 2001.
92. ZHOU, Y.; LIANG, Y.; LYNCH, K. H.; DENNIS, J. J.; WISHART, D. S. PHAST: a fast phage search tool. *Nucleic Acids Res*, v. 39, n. Web Server issue, p. W347-52, 2011.

## 8. Anexos

### 8.1. Anexo I

- I- “Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*”
  
- II- “The Plant Growth-Promoting Bacteria *Azospirillum amazonense*: Genomic Versatility and Phytohormone Pathway”

RESEARCH ARTICLE

Open Access

# Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*

Fernando H Sant'Anna<sup>1</sup>, Luiz GP Almeida<sup>2</sup>, Ricardo Cecagno<sup>1</sup>, Luciano A Reolon<sup>1</sup>, Franciele M Siqueira<sup>1</sup>, Maicon RS Machado<sup>1</sup>, Ana TR Vasconcelos<sup>2</sup> and Irene S Schrank<sup>1,3\*</sup>

## Abstract

**Background:** The species *Azospirillum amazonense* belongs to a well-known genus of plant growth-promoting bacteria. This bacterium is found in association with several crops of economic importance; however, there is a lack of information on its physiology. In this work, we present a comprehensive analysis of the genomic features of this species.

**Results:** Genes of *A. amazonense* related to nitrogen/carbon metabolism, energy production, phytohormone production, transport, quorum sensing, antibiotic resistance, chemotaxis/motility and bacteriophytochrome biosynthesis were identified. Noteworthy genes were the nitrogen fixation genes and the nitrilase gene, which could be directly implicated in plant growth promotion, and the carbon fixation genes, which had previously been poorly investigated in this genus. One important finding was that some *A. amazonense* genes, like the nitrogenase genes and RubisCO genes, were closer phylogenetically to Rhizobiales members than to species of its own order.

**Conclusion:** The species *A. amazonense* presents a versatile repertoire of genes crucial for its plant-associated lifestyle.

## Background

The genus *Azospirillum* ( $\alpha$ -proteobacteria class) encompasses free-living bacteria that can improve the growth of many economically important plants, mainly cereals (for an extensive review, see [1]). Therefore, these microorganisms are considered as plant growth-promoting rhizobacteria (PGPR). Species of this genus are widely distributed in nature, living in soils of tropical, subtropical and temperate regions all over the world. Several aspects of their physiology seem to be related to a plant stimulatory effect, notably their ability to synthesize phytohormones. Although these microorganisms are able to fix atmospheric nitrogen, the exact contribution of this process to plant growth is still disputable [1-3].

So far, fifteen species of the *Azospirillum* genus have been described ([4] and references therein). However,

most research efforts have been dedicated to the species *Azospirillum brasilense*, neglecting the potential offered by the biological diversity of this genus.

The bacterium *A. amazonense*, the focus species of this study, was initially isolated from forage grasses grown in the Amazon region. Further studies revealed its broad ecological distribution, as it is also found in association with the roots of gramineous plants like rice, maize, sugarcane and sorghum [5,6]. This species is phylogenetically closer to *Azospirillum irakense* and *Rhodospirillum centenum* (also known as *Rhodocista centenaria*) than to *A. brasilense*. Unlike the latter, *A. amazonense* can use sucrose as a sole carbon source and is better adapted to acid environments [6].

In order to access the valuable information that genomic sequences can provide on the physiology of azospirilla, there have been independent efforts by research groups in sequencing their genomes. Currently, the genomes of three members of the *Azospirillum-R. centenum* group are available: *Azospirillum* sp. B510 [7], *A. brasilense* Sp245 [8] and *R. centenum* SW [9]. Although

\* Correspondence: irene@cbiot.ufrgs.br

<sup>1</sup>Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Campus do Vale, Porto Alegre, RS, Brazil  
Full list of author information is available at the end of the article

many years have passed since the discovery of *A. amazonense*, there is scarce information about this species. Nevertheless, a recent study under greenhouse conditions showed that especially *A. amazonense* Y2 (wild-type strain) contributed to the growth of rice plants by means of biological nitrogen fixation [10], showing its potential for use as an agricultural inoculant. Therefore, the objective of our study was to sequence the *A. amazonense* Y2 genome and to analyze specific regions that could exert fundamental roles in its survival in the soil and in its ability to promote plant growth.

## Results and Discussion

### General features of the *A. amazonense* draft genome database

The *A. amazonense* Y2 presents four replicons with the following estimated sizes: 2.7 Mbp, 2.2 Mbp, 1.7 Mbp and 0.75 Mbp [11]. The genomic G+C content of *A. amazonense* Y2 is 66.89%. The draft genome sequence consists of 7,044,835 bp divided in 1617 contigs. The average gene length is 1080. Currently, there are 3319 predicted CDS, where 2299 have sequence similarity to known genes, 501 are homologs to genes of unknown function and 519 are hypothetical genes exclusive to *A. amazonense*.

### Taxonomic features of *A. amazonense*

In this study, a phylogenetic tree was constructed using the 16S rDNA sequences from microorganisms belonging to the orders Rhodospirillales and Rhizobiales (Figure 1). The resulting phylogenetic tree clearly shows a split between these orders. The outermost clade containing all the *Azospirillum* species divides in two main subclades: one containing *A. amazonense*, *A. irakense*, *Rhodocista pekingensis* and *R. centenum*, and another containing the other *Azospirillum* species. This result is in agreement with previous studies, showing the close evolutionary relationship between *A. amazonense* and *R. centenum* [12-14], and is also supported by the greater number of *A. amazonense* genes (22%) with best BLAST hits (KEGG Database) to *R. centenum* genes.

Another relevant feature of the 16S rDNA phylogenetic reconstruction is that *R. centenum* does not cluster with other *Rhodospirillum* species, suggesting that the genus nomenclature of *R. centenum* is not appropriate, as has been pointed out by previous publications [13,15].

### Carbon metabolism

*Azospirilla* display versatile carbon metabolism in order to support their lives in the soil. *A. amazonense* is capable of growing on various disaccharides, hexoses and pentoses and a previous study suggested that *A. amazonense* is able to catabolize carbohydrates exclusively through the Entner-Doudoroff pathway (ED pathway)

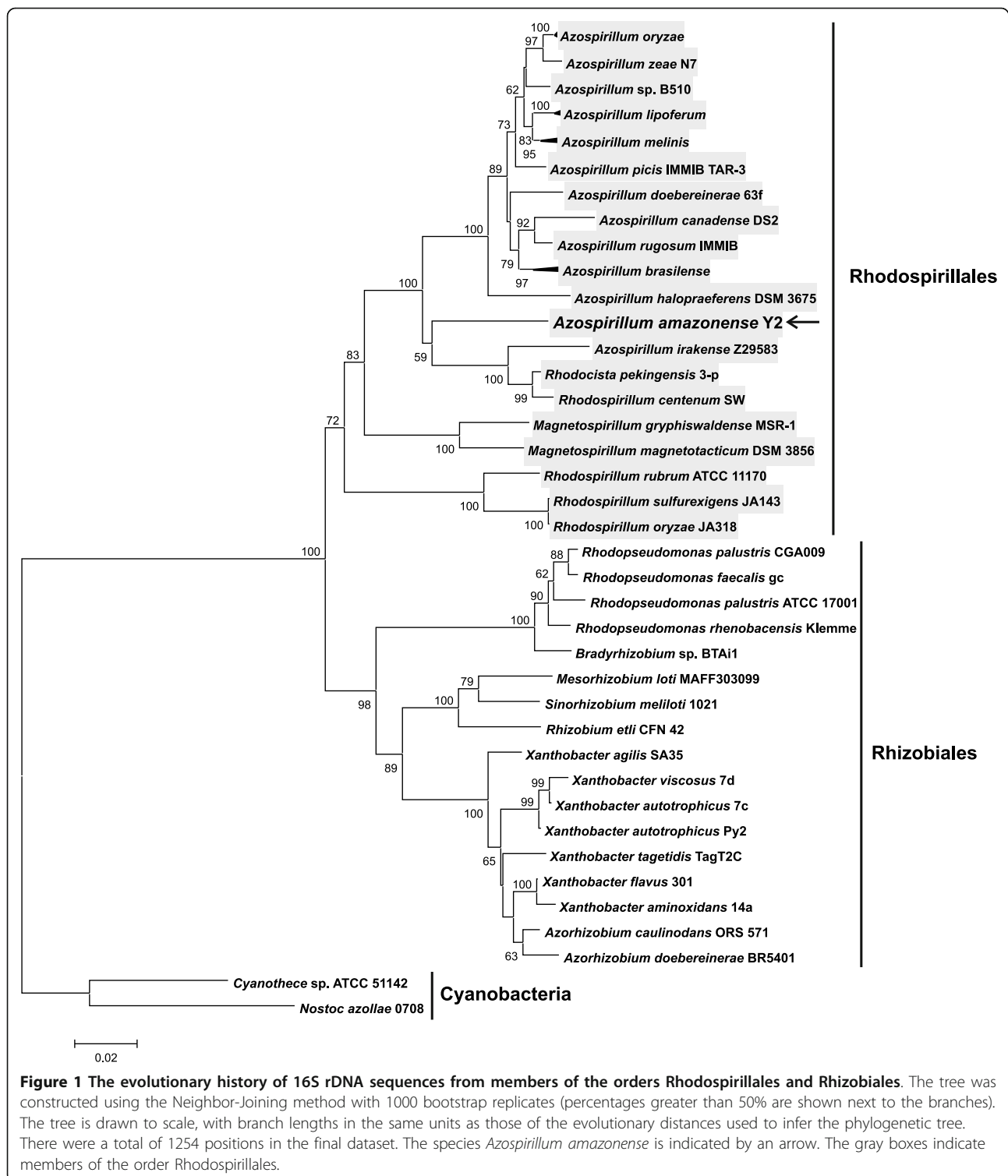
[16]. The genes encoding the key enzymes of this pathway, 6-phosphogluconate dehydratase and 2-dehydro-3-deoxy-phosphogluconate aldolase (KDPG aldolase) (Additional file 1), were found in the *A. amazonense* genome and seem to be organized as an operon. This same study also suggested that the glycolysis pathway (Embden-Meyerhof-Parnas pathway) was inoperative in *A. amazonense*, because no activity of 6-phosphofructokinase and fructose biphosphate aldolase was detected in crude extracts [16]. However, predicted genes encoding those enzymes were found in the *A. amazonense* genome (Additional file 1). Therefore, although the genomic approach indicates that most probably *A. amazonense* is also able to consume carbohydrates via glycolysis, this catabolic feature should be experimentally retested.

As stated previously, one of the main differences between *A. amazonense* and *A. brasilense* is that *A. amazonense* is capable of consuming sucrose as the sole carbon source [6]. In the genome of *A. amazonense*, a predicted gene that codes for a putative  $\alpha$ -glucosidase was identified (Additional file 1). This enzyme converts sucrose to glucose and fructose, substrates that can be promptly consumed by catabolic pathways.

The *A. amazonense* genome also harbors homologs of the genes *salB* and *salA* of *A. irakense* (Additional file 1). These genes encode  $\beta$ -glucosidases, enzymes implicated in the acquisition of glucose by means of the hydrolysis of aryl- $\beta$ -glucosides, such as salicin [17].

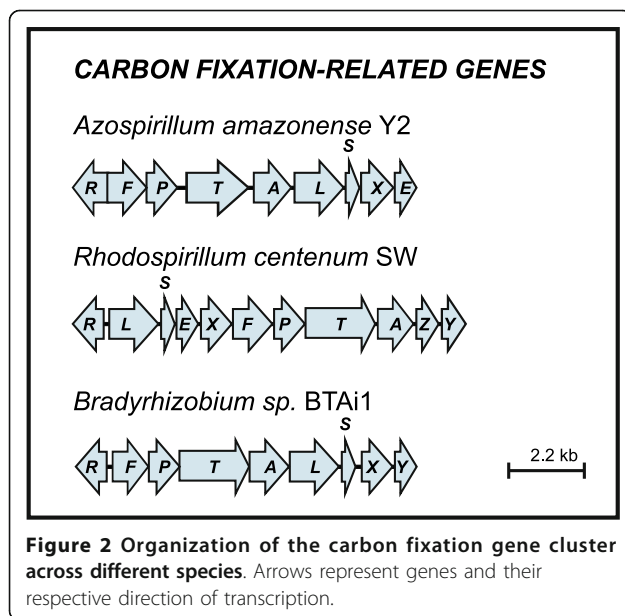
Bacteria of the genus *Azospirillum* produce high levels of poly- $\beta$ -hydroxybutyrate (PHB), the energy and carbon storage source utilized under nutritional stress conditions [2,18]. The essential genes for PHB biosynthesis are present in the *A. amazonense* genome: *phbA* ( $\beta$ -ketothiolase), *phbB* (aceto acetyl coenzyme A reductase) and *phbC* (PHB synthase) (Additional file 1). Furthermore, the *phaZ* gene that encodes a PHB depolymerase (Additional file 1), the first enzyme of the PHB degradation pathway, was also found in its genome.

One of the most surprising features of the *A. amazonense* genome is the presence of a gene cluster implicated in carbon fixation (the Calvin-Benson-Basham cycle) (Figure 2 and Additional file 1). The main genes of this cluster are the genes *cbbL* and *cbbS*, and they encode, respectively, the large and small subunits of ribulose-1,5-bisphosphate carboxylase (RubisCO). This enzyme is responsible for the incorporation of carbon dioxide in a molecule of ribulose-1,5-bisphosphate, generating two molecules of 3-phosphoglycerate, which can subsequently be used in biosynthetic pathways. A phylogenetic analysis of the concatenated RubisCO small and large subunits of *A. amazonense* revealed that they belong to the Form IC of RubisCOs (Figure 3). This type of enzyme is commonly found in  $\alpha$ -



Proteobacteria and it is adapted to environments with medium to high CO<sub>2</sub> and the presence of O<sub>2</sub> (in general, RubisCOs also have affinity to O<sub>2</sub> and high levels of this molecule can inhibit CO<sub>2</sub> fixation) [19]. So far, there have been no reports showing that *A.*

*amazonense* has autotrophic behavior. However, from the *Azospirillum* group, at least *R. centenum* and *A. lipoferum* are known to be capable of growing autotrophically by means of RubisCO [9,20], unlike *Azospirillum* sp. B510 and *A. brasilense* Sp245, which do not



contain Form I or II of RubisCOs ("true" RubisCOs) encoded in their genomes.

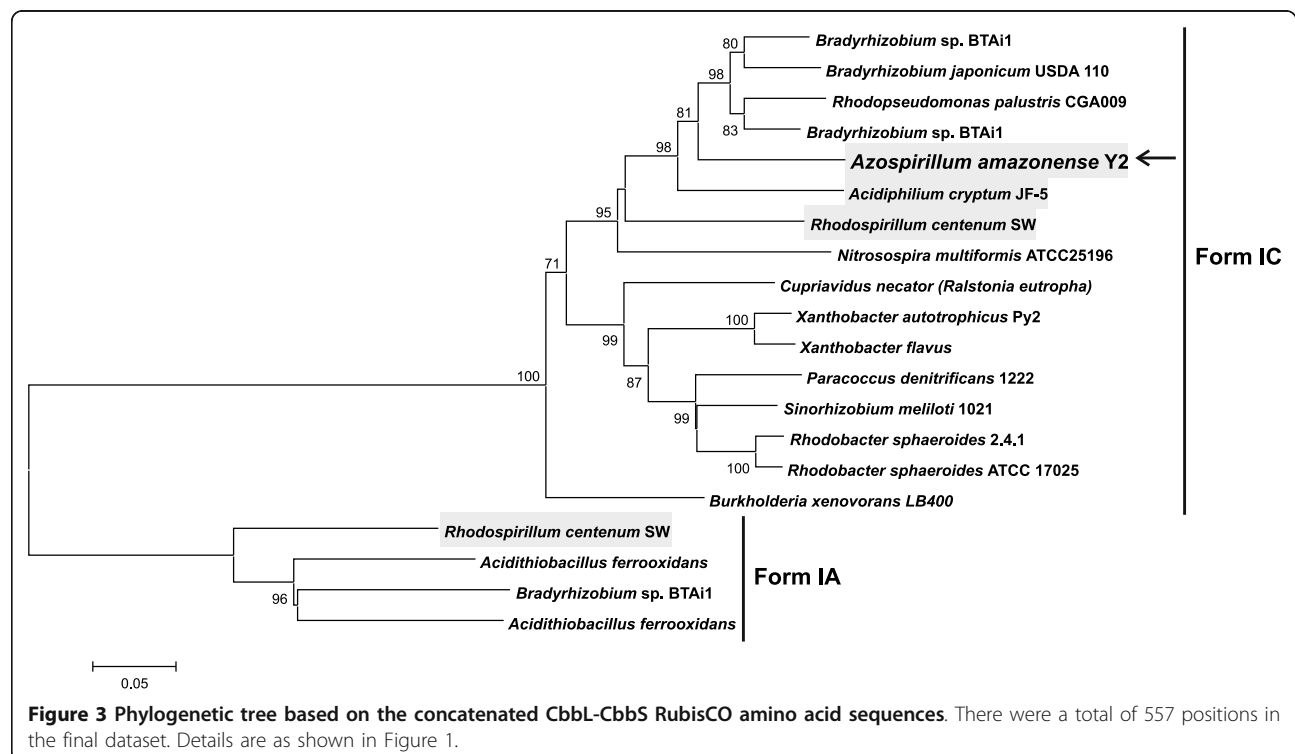
The RubisCO phylogenetic reconstruction also indicated the close relationship of the *A. amazonense* enzyme with those from members of the family Bradyrhizobiaceae (order Rhizobiales) (Figure 3), namely *Rhodopseudomonas palustris* and *Bradyrhizobium* spp. In fact, the genetic organization of the carbon-fixation

cluster of *A. amazonense* resembles that found in *Bradyrhizobium* sp. BTAi1, contrasting with the genetic organization of *R. centenum*. These incongruities, i.e. the genetic organization and phylogenetic relationship closer to Bradyrhizobiaceae members than to *R. centenum*, suggests that horizontal gene transfer may be an important driving force in the evolution and dispersion of RubisCOs in Proteobacteria.

#### Nitrogen metabolism

*Azospirillum* species are able to utilize distinct nitrogen sources, including ammonia, nitrate, nitrite, dinitrogen and amino acids [3,21]. The *A. amazonense* species has several genes implicated in nitrogen metabolism, which encode transporters, enzymes and regulatory proteins (Additional file 1).

Ammonia is the central compound of nitrogen metabolism and the preferred nitrogen source of many microorganisms. In general, nitrogen sources other than ammonia are converted into it to be assimilated [22]. The *A. amazonense* genome contains genes that are implicated in this conversion of alternative nitrogen sources, like nitrate/nitrite, urea and dinitrogen (Additional file 1). Once available, ammonia can be incorporated into the metabolism by the glutamine synthetase (GS)/glutamine:oxoglutarate aminotransferase (GOGAT) pathway, the genes for which are also encoded in the *A. amazonense* genome (Additional file 1).





Overall, the conversion of nitrogen compounds to ammonia expends some energy and, therefore, the metabolic pathways implicated in this process are strictly regulated to minimize energy waste. The central regulators of nitrogen metabolism are the PII proteins [22,23]. Three PII homolog genes (*glnB*, *glnK* and *glnK2*) were found in the *A. amazonense* genome (Additional file 1). The *glnK* gene and the *glnB* gene have ortholog counterparts in *Azospirillum* sp. B510, *A. brasilense* Sp245 and *R. centenum*. The *glnK* gene is upstream of the *aat* gene (aminotransferase) and the *glnB* is upstream of the *glnA* gene (glutamine synthetase) [24]. The third gene, *glnK2*, which is located downstream of the *amtB* gene, is absent in *A. brasilense* Sp245, *Azospirillum* sp. B510 and *R. centenum*, although this genetic association is frequently found in diverse prokaryotes [25].

The PII protein interactions with transporters, transcription factors and regulatory enzymes are well-established in the literature (for a review, see [22,23]) and the potential interaction targets found in the *A. amazonense* genome will be briefly discussed. One putative target is the *glnD* gene that codes for an uridyl-transferase, an enzyme that uridylylates the PII proteins under nitrogen-limiting conditions [23]. Other potential targets found in the *A. amazonense* genome are two ammonium transporters, encoded by the *amtB* genes, which in the presence of high nitrogen levels are inhibited by PII proteins [26]. The adenylyltransferase enzyme (encoded by the *glnE* gene, Additional file 1) which regulates glutamine synthetase via covalent modifications [27] could also interact with PII proteins [28].

The analysis of the *A. amazonense* genome revealed the presence of the NtrBC and NifA systems (Additional file 1), which are PII-regulated signal transduction systems responsible for the coordination of genes implicated in the scavenging of alternative nitrogen sources [23]. Both NtrC and NifA rely on the presence of the sigma N factor (also known as RpoN, or sigma 54) to activate the transcription of specific genes [29], which is also present in the *A. amazonense* genome (Additional file 1).

One of the main characteristics of the *Azospirillum* species is that they are able to fix nitrogen, i.e. convert N<sub>2</sub> to ammonia, by means of the nitrogenase enzyme complex. The main genes implicated in this process are known as *nif* genes, and they are highly conserved among nitrogen-fixing proteobacteria [30].

A preliminary BLAST analysis showed that the *A. amazonense nif* genes exhibit high similarity with genes of some species of the order Rhizobiales (Additional file 1). Since these observations were unexpected, the phylogenetic history of *nifH* was reconstructed utilizing sequences from species of the orders Rhodospirillales and Rhizobiales. The resulting *nifH* tree (Figure 4) was

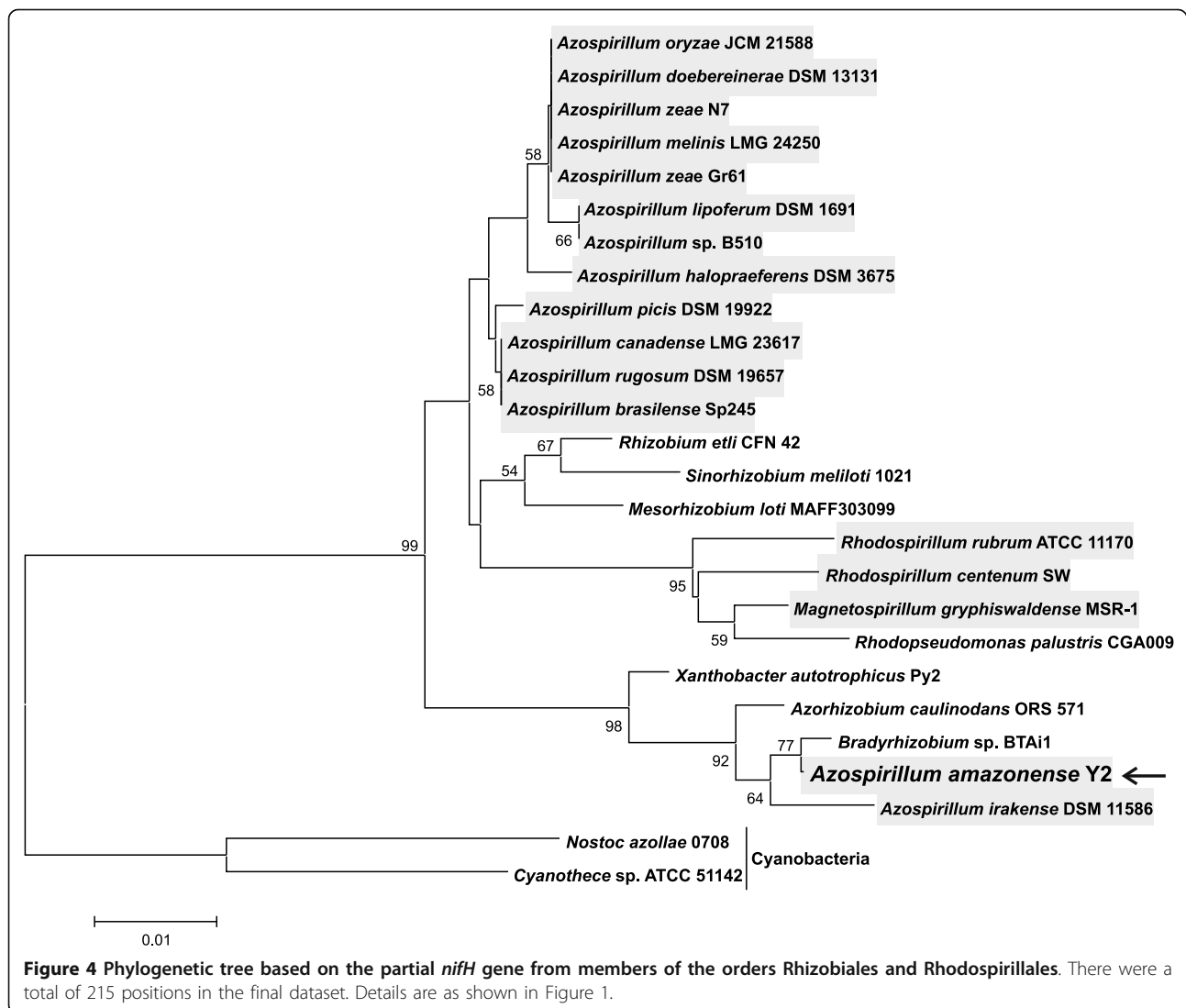
clearly incongruent to the 16S rDNA tree (Figure 1): in the *nifH* tree, *A. amazonense* and *A. irakense* grouped with *Bradyrhizobium* sp. BTAi1, *Xanthobacter diazotrophicus* and *Azorhizobium caulinodans*, instead of grouping with other *Azospirillum* species.

The genetic organization of the *nif* genes between the Rhodospirillales and Rhizobiales bacteria is somewhat homogeneous (Figure 5). As demonstrated in Figure 5, some features of the *nif* cluster of *A. amazonense* are exclusively similar to the homolog cluster of the species *Bradyrhizobium* sp. BTAi1, such as the presence of three conserved hypothetical genes that are indicated by the numbers 2, 5 and 6. On the other hand, *Azospirillum* sp. B510 and *A. brasilense* have the *aerC* gene between the *nifHDK* and *nifENX* operons and the *draG* and *draT* genes in the upstream region of the *nifH* gene, features not shared with the *nif* cluster of *A. amazonense*. The genes *draG* and *draT* code for a post-translational control system of the nitrogenase, which are not present in *A. amazonense* and *R. centenum* [9,31].

The domain composition of the deduced NifE and NifU proteins is also evidence that the nitrogenase complex of *A. amazonense* is more distantly related to *Azospirillum* spp. (with the exception of *A. irakense*) than to some Rhizobiales species (i.e. those from the genus *Bradyrhizobium*, *Xanthobacter* and *Azorhizobium*). The NifE protein of *A. amazonense* (like the *Bradyrhizobium*, *Xanthobacter* and *Azorhizobium* proteins NifE) has a bacterioferritin-associated ferredoxin [2Fe-2S] binding domain (BFD), not present in the NifE protein from *A. brasilense* Sp245 and *Azospirillum* sp. B510, and *R. palustris* (Rhizobiales) (Additional file 2). Furthermore, its NifU protein is smaller than those of *A. brasilense* Sp245, *Azospirillum* sp. B510 and *R. palustris* CGA009 because it does not contain the BFD and an N-terminal NifU domain, which are present in the NifU protein of the latter species (Additional file 2).

Therefore, taken together, these results indicate that complex events occurred in the evolution of the nitrogenase system in these bacteria, for instance, horizontal gene transfers and/or gene duplication followed by differential gene loss, culminating in the current distribution of the *nif* genes among the members of these taxonomic orders.

Nitrogen fixation is a very energy-demanding process and it is not surprising that the nitrogenase system is elaborately regulated. In all diazotrophic species of the Proteobacteria examined so far, the transcriptional activator NifA and the sigma N alternative RNA polymerase-associated factor are the master regulators of nitrogen fixation genes [32]. Sequence motifs similar to the consensus region of sigma N and NifA binding sites are present upstream of the *nifH* gene, the homolog of



“orf2” (indicated by the number 7 in Figure 5) of the *orf2nifUSVor4* cluster from *A. brasilense* [33] and the *nifB* gene (Additional file 3).

#### Energy production and conversion for nitrogen fixation

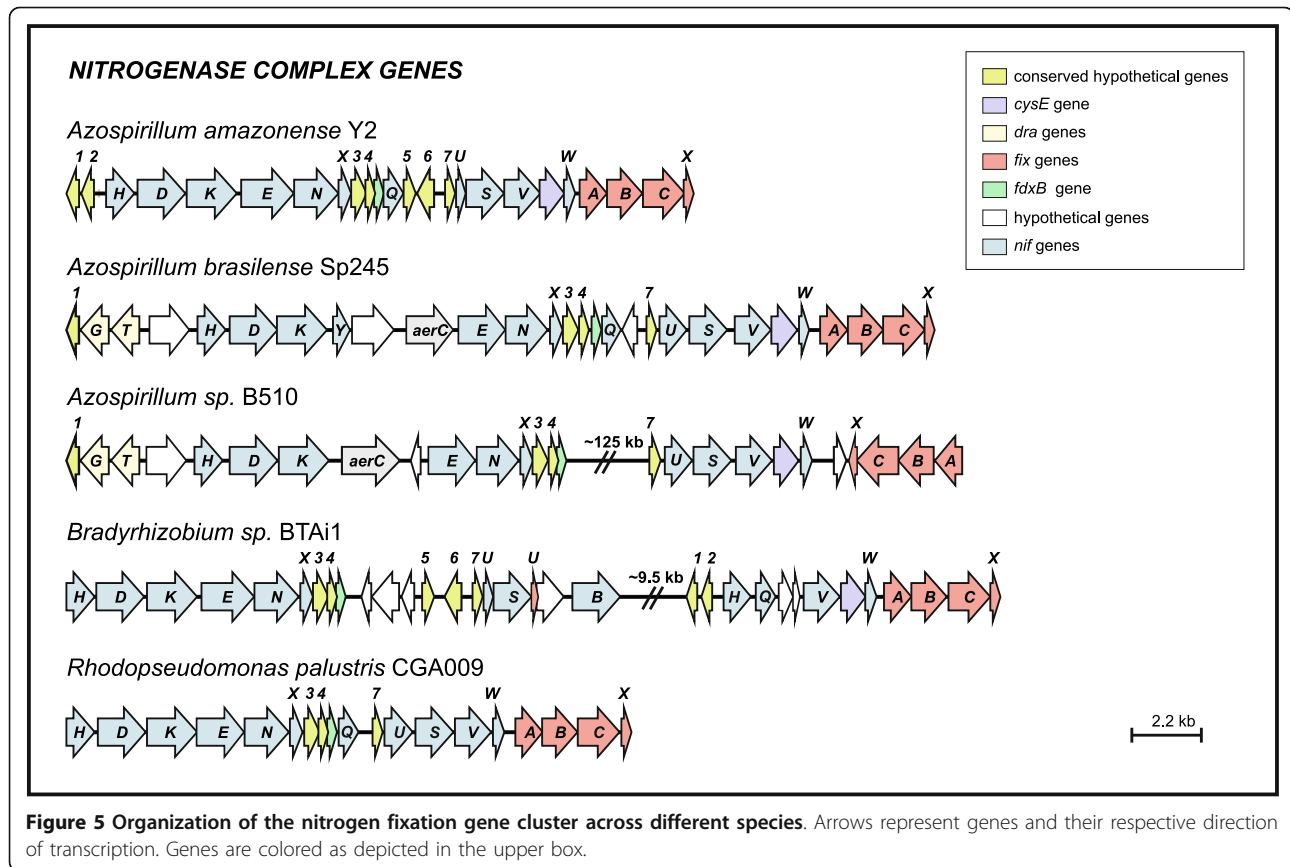
Nitrogen fixation demands the systematic action of different genes. The *fix* genes are essential for this process, and they encompass different functional categories. These genes were found in the *A. amazonense* genome divided into three main clusters, namely *fixABCX*, *fixLJ-fixK* and *fixNOQP-fixGHIS*.

The *fixABCX* genes from *A. amazonense*, responsible for electron transfer to nitrogenase, are located downstream of the *nifW* gene (Figure 5). As seen in Figure 5, this genetic cluster is highly conserved among the analyzed species, and it is tightly associated with the *nif* cluster. Putative NifA and sigma N binding sites were found upstream of the operon *fixABCX* from *A.*

*amazonense* (Additional file 3), indicating that these transcription factors could be key elements for the expression of this operon. This evidence is corroborated by reports showing that the operon *fixABCX* is regulated by the NifA protein in *A. brasilense* and *Rhizobium* spp. [34-36].

The *A. amazonense* genome also possesses a gene cluster including *fixLJ* and *fixK*. In symbiotic diazotrophs, the transcription of *fix* genes involves the oxygen-responsive FixLJ two-component system. The FixL protein, in the absence of oxygen, autophosphorylates and transfers the phosphate group to FixJ. Finally, the phosphorylated FixJ activates the expression of FixK, which activates the transcription of genes required for microaerobic growth [32].

Molecular nitrogen reduction requires high levels of energy under microaerobic conditions. The *fixNOQP* and *fixGHIS* genes encode membrane-bound



**Figure 5 Organization of the nitrogen fixation gene cluster across different species.** Arrows represent genes and their respective direction of transcription. Genes are colored as depicted in the upper box.

cytochrome c oxidase and the redox process-coupled cation pump, respectively, which are intimately implicated in respiration under microaerobic conditions, supplying energy for nitrogen fixation [37,38]. These clusters were found *in tandem* in the *A. amazonense* genome (Additional file 1), and they show identical organization in many diazotrophic  $\alpha$ -Proteobacteria, like *R. centenum*, *A. brasilense* Sp245, *R. palustris* and *Bradyrhizobium* spp.

Nitrogen fixation forms molecular hydrogen ( $H_2$ ) as a byproduct. Therefore, diazotrophic bacteria have several hydrogenase systems that are responsible for oxidizing molecular hydrogen to recover part of the energy expended during nitrogenase activity. Genes encoding for an uptake NiFe hydrogenase (*hupSL*) were identified in *A. amazonense* (Figure 6). Furthermore, the accessory proteins required for maturation of the subunits [39,40], encoded by the *hup* and *hyp* genes, are situated downstream of the genes encoding *hupSL* (Figure 6). This organization resembles that found in members of the order Rhizobiales, where, in general, the *hup* and *hyp* genes are clustered (Figure 6), although their ordering is quite heterogeneous among the species. In contrast with this observation, comparisons with closely-related species showed that *Azospirillum* sp. B510 and *A. brasilense*

Sp245 have their *hup* and *hyp* genes scattered across the genome. Moreover, the bacteria *R. centenum* does not have the hydrogenase gene cluster.

An integrated model relating the components discussed in this section with nitrogen metabolism is depicted in Figure 7, taken into consideration the similarity of *A. amazonense* genes to those from other well-known bacterial systems.

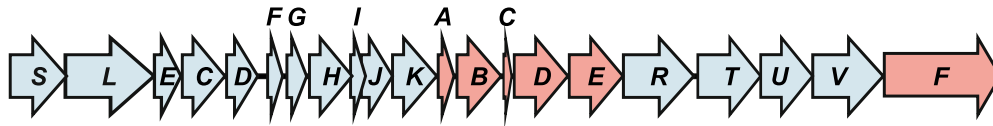
#### Phytohormone production

Species of the *Azospirillum* genus can positively influence plant growth and crop yield by the biosynthesis and secretion of indole-3-acetic acid (IAA) [2,41]. However, although *A. amazonense* strains are able to synthesize IAA [10], very little is known about the molecular mechanisms responsible for this process.

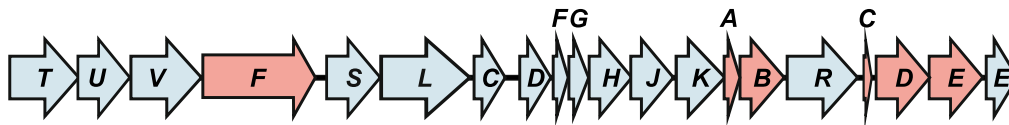
In *A. brasilense*, at least three pathways for IAA biosynthesis exist, two tryptophan-dependent pathways (indole-3-acetamide pathway (IAM) and indole-3-pyruvate pathway (IPyA)) and one tryptophan-independent pathway [3,42]. Similarly, the genome of *Azospirillum* sp. B510 contains genes responsible for the IAM pathway [7]. However, the *iaaM*, *iaaH* and *ipdC* genes, related to the IAM or IPyA pathways, were not located in the *A. amazonense* genome. Further analysis of the

## HYDROGENASE COMPLEX GENES

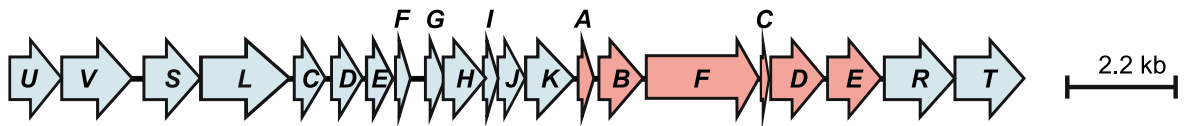
*Azospirillum amazonense* Y2



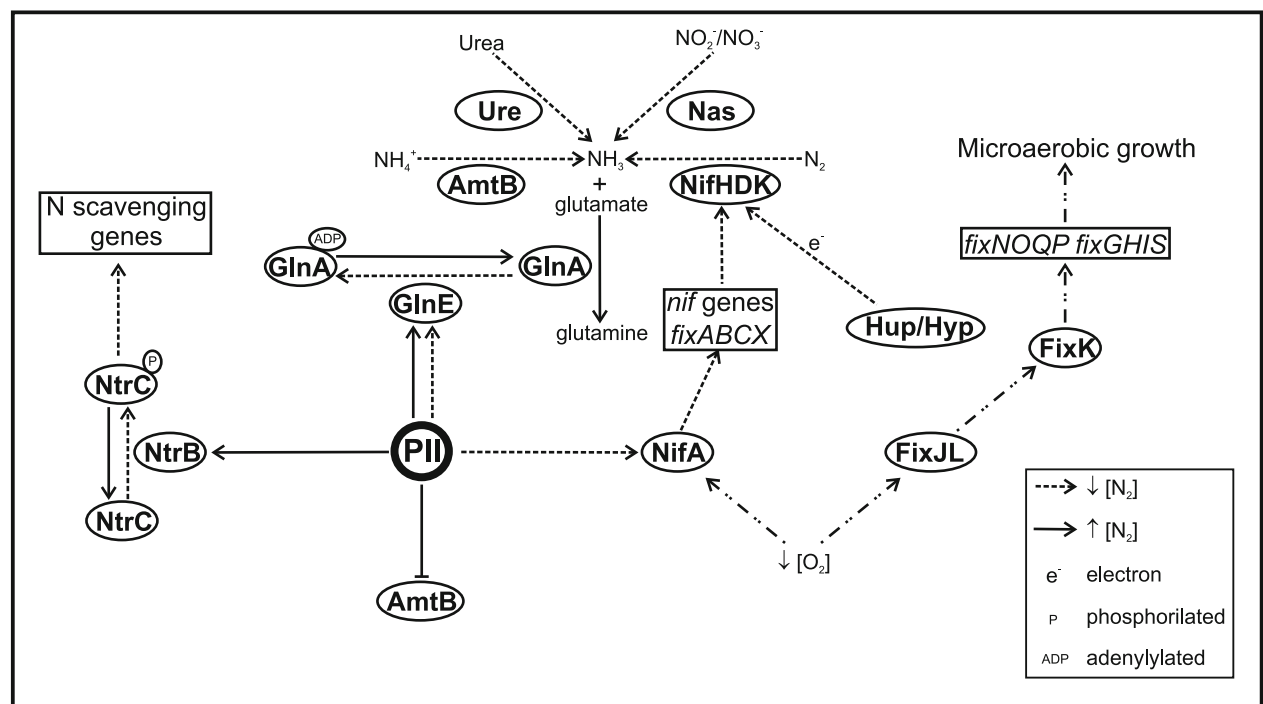
*Azorhizobium caulinodans*



*Rhodopseudomonas palustris* CGA009



**Figure 6** Organization of the hydrogenase gene cluster across different species from the orders Rhodospirillales and Rhizobiales. Arrows represent genes and their respective direction of transcription. *hup* genes are colored in red and *hyp* genes are colored in blue.



**Figure 7** Integrated model correlating nitrogen metabolism and energy related-pathways for nitrogen fixation. The legend box indicates the correlation of each symbol with its respective meaning.

genome sequence of *A. amazonense* revealed a gene encoding a protein with about 70% similarity to nitrilases from plant species, like *Arabidopsis thaliana* and *Zea mays*, which catalyze the conversion of indole 3-acetonitrile to IAA [43,44]. Future studies may verify if this gene is implicated in IAA biosynthesis in *A. amazonense*.

### Quorum sensing and biofilm formation

Quorum sensing is an intercellular signaling process implicated in the regulation of several traits of bacteria, notably antibiotic biosynthesis and biofilm formation. The archetype for quorum sensing regulation is the LuxIR system, which involves an acyl-homoserine lactone (AHL) synthase (LuxI homolog) and an AHL-dependent transcriptional regulator (LuxR homolog) [45]. The quorum sensing phenomenon of *Azospirillum* species is strain-specific and seems to regulate functions linked to rhizosphere competence and adaptation to plant roots [46].

The acyl-homoserine lactone (AHL) biosynthesis ability of forty *Azospirillum* strains (including *A. amazonense* Y2) was previously investigated, and only four *A. lipoferum* strains seemed of being capable of synthesizing these compounds [47]. However, the genome analysis of *A. amazonense* revealed the presence of genes encoding for LuxI and LuxR homologs proteins (Additional file 1). Therefore, these results indicate that *A. amazonense* Y2 could synthesize AHLs and respond to its presence in the environment. The genome of *A. amazonense* also presents a *Klebsiella pneumoniae* *ahlK* homolog [48], a predicted gene that codes for a putative homoserine lactonase (Additional file 1) implicated in AHL degradation. Since bacterial plant pathogens rely on quorum sensing mechanisms to infect plants [45], a study of *A. amazonense* homoserine lactonase activity on the deleterious activities of these pathogens would be relevant.

Extracellular polysaccharides are loosely bound to the cell surface and play an important role in bacterium-plant interactions through the firm and irreversible anchoring of cells to the plant roots [3]. In the *A. amazonense* genome, two genes, *noeJ* (mannose-6-phosphate isomerase, Additional file 1) and *noeL* (GDP-mannose 4,6-dehydratase, Additional file 1), which are related to extracellular polysaccharide biosynthesis and biofilm formation, were also found [49].

### Chemotaxis/Motility

Different species of *Azospirillum* attach to and colonize plant root surfaces and these processes depend on chemotaxis. *Azospirillum* exhibit chemotaxis towards sugars, amino acids, organic acids and root exudates [50]. This

ability offers the bacteria the advantage of moving towards favorable nutrient conditions.

Genes encoding for the central signal transduction pathway for chemotaxis (*che*) are present in nearly all motile bacteria. This signal transduction system is composed of the conserved *cheA**WYBR* genes and a group of transmembrane chemoreceptors (known as MCPs or methyl-accepting proteins) that perceive environmental signals. Homologs of the *cheA**WYBR* and MCP genes were identified in the *A. amazonense* genome (Additional file 1), and some *che* genes display similar organization to those found in the model organism *Escherichia coli* [51].

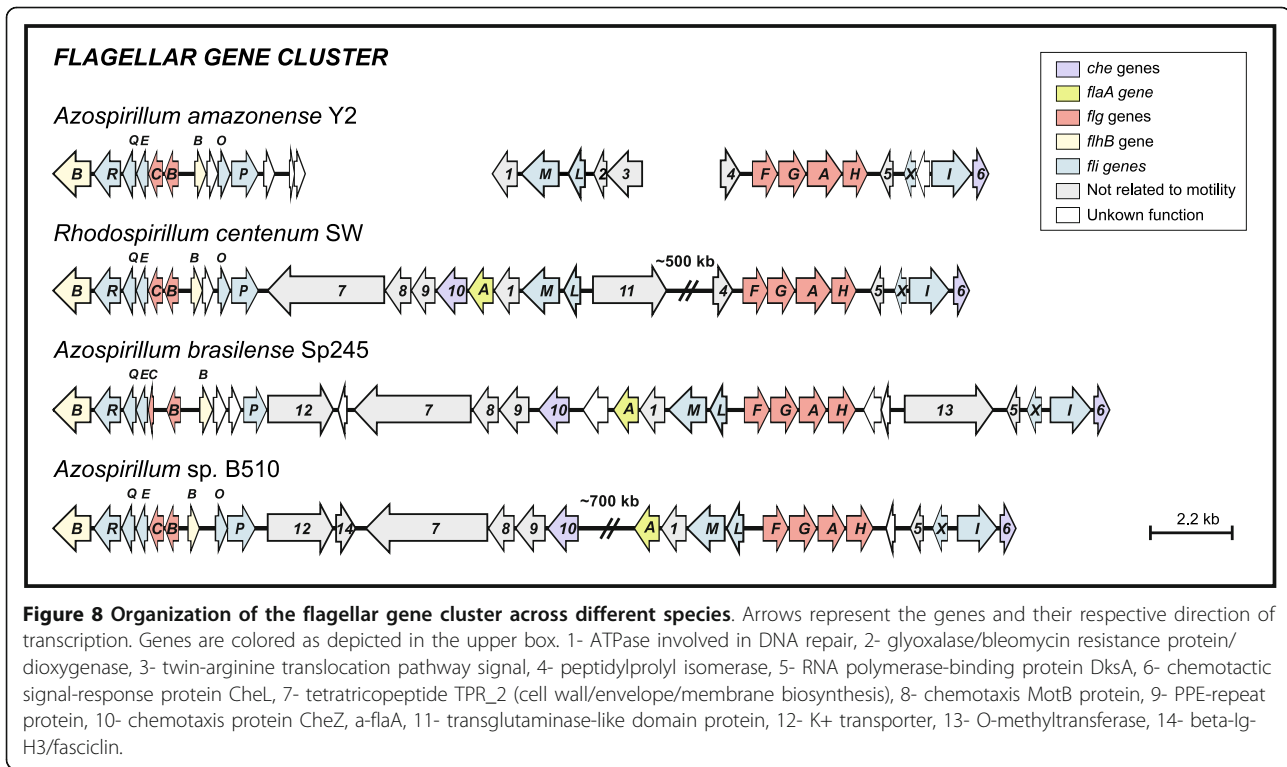
The *A. amazonense* *Che1* gene cluster (*cheA**WYBR*) revealed a conserved organization with the major chemotaxis gene cluster from *A. brasilense* Sp245, which modulates cell length and clumping behavior [52]. Apparently, this gene cluster also affects the production of exopolysaccharide and flocculation of *A. brasilense*. Other *che*-like genes (*Che2* and *Che3*, Additional file 1) are present in the *A. amazonense* genome, probably encoding parallel signal transduction pathways that could have distinct functions, similar to those found in *R. centenum* [53,54].

It is also worth noting that genes encoding MCP domain proteins are spread throughout the *A. amazonense* genome. Some of them were classified accordingly to their similarities to *E. coli* chemoreceptors. All five types of *E. coli* MCP receptors were found in the *A. amazonense* genome (Additional file 1), and it is probable that some of them are related to cell motility by regulating the histidine kinase *CheA* that phosphorylates a response regulator, which in turn controls the rotational direction of the flagellar motor [55].

The flagellum is a key structure for the chemotactic response. In the *A. amazonense* genome, 39 flagellar genes were identified, and the majority of them are distributed among different gene clusters (Additional file 1). The reduced flagellar gene number in *A. amazonense* compared to those of the closely-related *A. brasilense* (79 annotated genes) and *R. centenum* (72 genes) species is in agreement with previous evidence indicating that *A. amazonense* synthesizes only the polar flagellum for swimming motility [56].

Homologs of the *A. brasilense* and *R. centenum* genes *fliFHN-motA-flbD-flhAF-fleN* [9,57] are present in *A. amazonense* (Additional file 1). The *A. amazonense* genome also contains other flagellar genes that display similar organization to those found in *Azospirillum* spp. and *R. centenum* (Figure 8).

Previous studies demonstrated that the transcription factor *FlbD* is related to the biosynthesis of lateral flagella in *A. brasilense* and the polar flagella in *R.*



*centenum* [9,57]. The presence of the *flbD* gene in the *A. amazonense* genome suggests that it could have a similar regulatory function as that found in *R. centenum*.

### Transport, antibiotics resistance and lantibiotic production

Transport systems allow the uptake of nutrients and ions, excretion of end products of metabolism and communication between cells and the environment.

Several components of the cationic efflux pump, the ATP-Binding Cassette (ABC) transporter superfamily, the Major Facilitator Superfamily (MFS) and the TonB-dependent transporters were identified in the *A. amazonense* genome.

TonB-dependent transport is a mechanism of active uptake across the outer membrane normally related to iron uptake, signal transduction and environmental perception [58,59]. *A. amazonense* has a high number of genes encoding TonB receptors (Additional file 1) when compared with other diazotrophs such as *Azospirillum* sp. B510 (9 annotated genes) and *R. centenum* SW (32 annotated genes). Homology analyses suggest that a set of putative TonB receptors for specific substrates like ferrioxamine, cobalamin (B12 vitamin) and heme are present in this bacterium.

Among the different families of transporters, only two occur ubiquitously in all kingdoms of life: the Major Facilitator Superfamily (MFS) and the ATP-Binding

Cassette (ABC) superfamily, representing the largest and most distributed families of transmembrane proteins. MFS proteins are single polypeptide secondary carriers that utilize uniport, symport or antiport mechanisms to transport various small substrates [60]. The ABC transporter proteins utilize energy from adenosine triphosphate (ATP) hydrolysis to carry out the uptake of essential nutrients and/or the extrusion of toxic substances [61]. In the *A. amazonense* genome, several genes encoding for putative MFS and ABC transporters were identified, and they could be implicated in the transport of a wide range of putative substrates (Additional file 1).

Multidrug resistance (MDR) transporters increase drug excretion through an efflux pump, which expels a wide variety of toxic products from the cell, playing a central role in bacterial drug resistance. The MDR transporters belong to various transporter families [62]. In *E. coli*, the transport of diverse substrates out of the cell by the AcrAB-TolC efflux transporter can confer broad resistance to antibiotics [63]. The *acrA* and *acrB* genes normally form an operon whose transcription is regulated by the *acrR* gene product, and are found in the *A. amazonense* genome.

Putative drug resistance transporters of the QacA subfamily were found in the *A. amazonense* genome (Additional file 1), which could confer resistance to monovalent and bivalent cationic lipophilic antiseptics

and disinfectants such as quaternary ammonium compounds [64].

Further analysis revealed that, in addition to the MDR transporters, a set of genes whose products could be related to specific antimicrobial resistance are present in the *A. amazonense* genome. Genes that code for penicillin, glyoxalase/bleomycin and tetracycline resistance are also present in the *A. amazonense* genome (Additional file 1). These findings corroborate the experimental data that shows that *A. amazonense* is tolerant to tetracycline and resistant to penicillin [6].

Lantibiotics are peptide-derived antibacterial substances produced by some bacteria, and are characterized by the presence of unusual amino acids like lanthionines and dehydrated amino acids [65]. Lantibiotic biosynthesis is frequently coregulated as part of a stress response when cells enter the late-log or stationary phase [66]. Most lantibiotics exert their antibiotic effect by either forming pores in the target cell membrane or by inhibiting cell wall synthesis, and many lantibiotics are bactericidal against a variety of Gram-positive bacteria [67]. Genes related to lantibiotic synthesis were found in *A. amazonense* genome (Additional file 1).

The genes implicated in antibiotic resistance and in lantibiotic production are probably essential for successful establishment of this microorganism in the soil due to constant contact with niche competitors like fungi and other bacteria.

### Bacteriophytochrome

The *A. amazonense* genome also harbors a bacteriophytochrome gene (Additional file 1). Similarly, *Azospirillum* sp. B510 has two genes that code for distinct types of bacteriophytochromes [7], and *Bradyrhizobium* sp. possesses three bacteriophytochrome genes [68]. In plants, phytochromes regulate the metabolic response to the light environment, but a variety of functions is found in other organisms [69]. The bacteriophytochromes in *R. palustris* regulate the biosynthesis of the photosynthetic apparatus [70], while in *Deinococcus radiodurans* and *R. centenum*, they regulate pigment biosynthesis [71]. Subsequent studies must be carried out to understand the role of the bacteriophytochrome in *A. amazonense* physiology.

### Conclusion

In order to thrive, bacteria must adapt readily to environmental shifts by means of a wide variety of genotypic and phenotypic accommodations [72]. The rhizosphere is a good example of a dynamic environment, where fluctuations in its biological and chemical activities demand an appropriate response from its inhabitants. The species *A. amazonense* is a free-living plant growth-

promoting rhizobacterium that is found in association with plants of agricultural importance. In this study, we identified a series of *A. amazonense* genes that could be essential for adaptation to the competitive environment of the rhizosphere. Its wide genetic repertoire confers a versatile metabolism (e.g. the ability to use different carbon and nitrogen sources), as well as different mechanisms of perceiving and exploring its surroundings. These characteristics could directly influence plant growth, for instance, by providing nitrogen and stimulatory compounds to plants. Another important finding was the greater similarity of some genes, e.g. nitrogenase and RubisCO genes, to genes of members of the order Rhizobiales than to genes from other *Azospirillum* species. This evidence illustrates the genetic plasticity of this species and indicates that evolutionary phenomena like horizontal gene transfer could be fundamental for adaptation to its environment.

The major impact of this work will be to guide subsequent studies for a better understanding of the potential of *A. amazonense*.

### Methods

#### Bacterial strain, culture conditions, and DNA isolation

*A. amazonense* Y2 (ATCC 35120) was cultured in M79 medium (10 g/L of sucrose as a carbon source, 0.1 g/L of  $K_2HPO_4$ , 0.4 g/L of  $KH_2PO_4$ , 0.2 g/L of  $MgCl_2 \cdot 7H_2O$ , 0.1 g/L of NaCl, 0.4 g/L of yeast extract, pH 6.5) [73] with shaking at 150 rpm and 35°C for 18 hours. Genomic DNA was isolated as described by Wilson [74]. The quality of the isolated genomic DNA was assessed by agarose gel electrophoresis.

#### Genome sequencing, assembly, draft annotation

Total genomic DNA was sequenced using the Roche 454 pyrosequencing platform following the manufacturer's instructions. The contigs were assembled using Newbler software version 2.3 with the default parameters. The estimated coverage of the genome was 35x. Some gaps present in the genes of interest were filled in by sequencing PCR fragments obtained from genomic DNA.

The annotation and analysis of the sequences were carried out using the System for Automated Bacterial Integrated Annotation (SABIA) [75]. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AFBX00000000. The version described in this paper is the first version, AFBX01000000.

#### Phylogenetic analysis

Gene sequences were retrieved from GenBank. Most of the 16S rDNA sequences were retrieved from the Ribosomal Database Project [http://rdp.cme.msu.edu/] [76,77]. The accession numbers of the sequences utilized

in the phylogenetic reconstructions are listed in Additional file 4.

Multiple sequence alignments were performed using MUSCLE version 3.8 [78] and CLUSTALW (built into the MEGA 4 software) [79]. Phylogenetic trees were inferred using the neighbor-joining method (1000 bootstrap replicates) by the MEGA 4 software [80]. The evolutionary distances were computed using the Maximum Composite Likelihood method for the nucleotide sequences and the Jones-Taylor-Thornton (JTT) matrix-based method for the amino acid sequences. All positions containing gaps and missing data were eliminated from the datasets (complete deletion option).

## Additional material

**Additional file 1: Supplementary table 1.** Genes of *Azospirillum amazonense* described throughout the study.

**Additional file 2: Supplementary Figure 1.** Domain composition of the NifE and NifU proteins among bacteria from the orders Rhodospirillales and Rhizobiales. The oblong boxes represent protein domains which are colored according to the description in the legend.

**Additional file 3: Supplementary table 2.** Putative sigma N and NifA binding sites occurring upstream of some *nif* and *fix* genes.

**Additional file 4: Supplementary table 3.** Accession numbers of the sequences utilized in the phylogenetic reconstructions.

## Acknowledgements and Funding

We especially thank Dr. Augusto Schrank for the valuable effort that enabled the realization of this work. We also thank the Empresa Brasileira de Pesquisa Agropecuária - Centro Nacional de Pesquisa de Agrobiologia (EMBRAPA-CNPAB) for providing the *A. amazonense* Y2 strain. This work was supported by grants from the Brazilian National Research Council (CNPq) and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). FHS, RC, LAR and FMS received scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

## Author details

<sup>1</sup>Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Campus do Vale, Porto Alegre, RS, Brazil. <sup>2</sup>Laboratório Nacional de Computação Científica (LNCC), Petrópolis, RJ, Brazil. <sup>3</sup>Departamento de Biologia Molecular e Biotecnologia - Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Campus do Vale, RS, Brazil.

## Authors' contributions

ISS and ATRV conceived of and coordinated the study. FHS and RC extracted the genomic DNA. LGPA carried out the draft genome sequencing and assembly. FHS, LGPA, RC, LAR, FMS, MRSM and ISS performed the draft genome annotation. FHS, MRSM and RC carried out the comparative analyses. FHS carried out the phylogenetic analyses and created the illustrations. FHS, RC, LAR, FMS, MRSM and ISS analyzed the results. FMS and LAR wrote some sections of the manuscript. FHS and ISS wrote the manuscript. All authors read and approved the final manuscript.

Received: 29 April 2011 Accepted: 12 August 2011

Published: 12 August 2011

## References

1. Bashan Y, De-Bashan LE: How the plant growth-promoting bacterium *Azospirillum* promotes plant growth - a critical assessment. *Adv Agron* 2010, **108**:77-136.
2. Bashan Y, Holguin G, De-Bashan LE: *Azospirillum*-plant relationships: physiological, molecular, agricultural, and environmental advances (1997-2003). *Can J Microbiol* 2004, **50**:521-577.
3. Steenhoudt O, Vanderleyden J: *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiol Rev* 2000, **24**:487-506.
4. Lavrinenko K, Chemousova E, Gridneva E, Dubinina G, Akimov V, Kuever J, Lysenko A, Grabovich M: *Azospirillum thioophilum* sp. nov., a novel diazotrophic bacterium isolated from a sulfide spring. *Int J Syst Evol Microbiol* 2010, **60**:2832-2837.
5. Baldani JI, Baldani VLD: History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *An Acad Bras Ciênc* 2005, **77**:549-579.
6. Magalhães FMM, Baldani JI, Souto SM, Kuykendall JR, Döbereiner J: A new acid-tolerant *Azospirillum* species. *An Acad Bras Ciênc* 1983, **55**:417-430.
7. Kaneko T, Minamisawa K, Isawa T, Nakatsukasa H, Mitsui H, Kawaharada Y, Nakamura Y, Watanabe A, Kawashima K, Ono A, Shimizu Y, Takahashi C, Minami C, Fujishiro T, Kohara M, Katoh M, Nakazaki N, Nakayama S, Yamada M, Tabata S, Sato S: Complete genomic structure of the cultivated rice endophyte *Azospirillum* sp. B510. *DNA Res* 2010, **17**:37-50.
8. *Azospirillum brasilense* Sp245. [http://genome.ornl.gov/microbial/abra/19sep08/].
9. Lu YK, Marden J, Han M, Swingley WD, Mastrian SD, Chowdhury SR, Hao J, Helmy T, Kim S, Kurdoglu AA, Matthies HJ, Rollo D, Stothard P, Blankenship RE, Bauer CE, Touchman JW: Metabolic flexibility revealed in the genome of the cyst-forming alpha-1 proteobacterium *Rhodospirillum centenum*. *BMC genomics* 2010, **11**:325.
10. Rodrigues EP, Rodrigues LS, de Oliveira ALM, Baldani VLD, Teixeira KRS, Urquiaga S, Reis VM: *Azospirillum amazonense* inoculation: effects on growth, yield and N<sub>2</sub> fixation of rice (*Oryza sativa* L.). *Plant Soil* 2008, **302**:249-261.
11. Martin-Didonet CC, Chubatsu LS, Souza EM, Kleina M, Rego FG, Rigo LU, Yates MG, Pedrosa FO: Genome structure of the genus *Azospirillum*. *J Bacteriol* 2000, **182**:4113-4116.
12. Stoffels M, Castellanos T, Hartmann A: Design and application of new 16S rRNA-targeted oligonucleotide probes for the *Azospirillum-Skermanella-Rhodocista*-cluster. *Syst Appl Microbiol* 2001, **24**:83-97.
13. Fani R, Bandi C, Bazzicalupo M, Ceccherini MT, Fancelli S, Gallori E, Gerace L, Grifoni A, Miclaus N, Damiani G: Phylogeny of the genus *Azospirillum* based on 16S rDNA sequence. *FEMS Microbiol Lett* 1995, **129**:195-200.
14. Zhang D, Yang H, Zhang W, Huang Z, Liu SJ: *Rhodocista pekingensis* sp. nov., a cyst-forming phototrophic bacterium from a municipal wastewater treatment plant. *Int J Syst Evol Microbiol* 2003, **53**:1111-1114.
15. Kawasaki H, Hoshino Y, Kuraishi H, Yamasato K: *Rhodocista centenaria* gen. nov., sp. nov., a cyst-forming anoxygenic photosynthetic bacterium and its phylogenetic position in the Proteobacteria alpha group. *J Gen Appl Microbiol* 1992, **38**:541-551.
16. Martínez-Drets G, Fabiano E, Cardona A: Carbohydrate catabolism in *Azospirillum amazonense*. *Appl Environ Microbiol* 1985, **50**:183-185.
17. Faure D, Desair J, Keijers V, Bekri MA, Proost P, Henrissat B, Vanderleyden J: Growth of *Azospirillum irakense* KBC1 on the aryl beta-glucoside salicin requires either *sala* or *salB*. *J Bacteriol* 1999, **181**:3003-3009.
18. Kadouri D, Jurkevitch E, Okon Y: Involvement of the reserve material poly-beta-hydroxybutyrate in *Azospirillum brasilense* stress endurance and root colonization. *Appl Environ Microbiol* 2003, **69**:3244-3250.
19. Badger MR, Bek EJ: Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO<sub>2</sub> acquisition by the CBB cycle. *J Exp Bot* 2008, **59**:1525-1541.
20. Hartmann A, Baldani JI: The genus *Azospirillum*. In *The Prokaryotes*. Edited by: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E. Springer New York; 2006:115-140.
21. Hartmann A, Zimmer W: Physiology of *Azospirillum*. In *Azospirillum/Plant Associations*. Edited by: Okon Y. Boca Raton: CRC Press; 1994:15-39.
22. Leigh JA, Dodsworth JA: Nitrogen regulation in bacteria and archaea. *Annu Rev Microbiol* 2007, **61**:349-377.
23. Arcondéguy T, Jack R, Merrick M: PII signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* 2001, **65**:80-105.
24. Sant'Anna FH, Andrade DS, Trentini DB, Weber SS, Schrank IS: Tools for genetic manipulation of the plant growth-promoting bacterium *Azospirillum amazonense*. *BMC Microbiol* 2011, **11**:107.



25. Sant'Anna FH, Trentini DB, de Souto Weber S, Cecagno R, da Silva SC, Schrank IS: **The PII superfamily revised: a novel group and evolutionary insights.** *J Mol Evol* 2009, **68**:322-336.
26. Javelle A, Merrick M: **Complex formation between AmtB and GlnK: an ancestral role in prokaryotic nitrogen control.** *Biochem Soc Trans* 2005, **33**:170-172.
27. Van Dommelen A, Spaepen S, Vanderleyden J: **Identification of the glutamine synthetase adenylyltransferase of *Azospirillum brasilense*.** *Res Microbiol* 2009, **160**:205-212.
28. Atkinson MR, Ninfa AJ: **Characterization of the GlnK protein of *Escherichia coli*.** *Mol Microbiol* 1999, **32**:301-313.
29. Shingler V: **Signal sensory systems that impact  $\sigma(54)$ -dependent transcription.** *FEMS Microbiol Rev* 2011, **35**:425-440.
30. Raymond J, Siefert JL, Staples CR, Blankenship RE: **The natural history of nitrogen fixation.** *Mol Biol Evol* 2004, **21**:541-554.
31. Fu HA, Hartmann A, Lowery RG, Fitzmaurice WP, Roberts GP, Burris RH: **Posttranslational regulatory system for nitrogenase activity in *Azospirillum* spp.** *J Bacteriol* 1989, **171**:4679-4685.
32. Dixon R, Kahn D: **Genetic regulation of biological nitrogen fixation.** *Nat Rev Microbiol* 2004, **2**:621-631.
33. Frazzon J, Schrank IS: **Sequencing and complementation analysis of the *nifUSV* genes from *Azospirillum brasilense*.** *FEMS Microbiol Lett* 1998, **159**:151-158.
34. David M, Daveran ML, Batut J, Dedieu A, Domergue O, Ghai J, Hertig C, Boistard P, Kahn D: **Cascade regulation of *nif* gene expression in *Rhizobium meliloti*.** *Cell* 1988, **54**:671-683.
35. Miller SH, Elliot RM, Sullivan JT, Ronson CW: **Host-specific regulation of symbiotic nitrogen fixation in *Rhizobium leguminosarum* biovar *trifolii*.** *Microbiology* 2007, **153**:3184-3195.
36. Sperotto RA, Gross J, Vedoy C, Passaglia LM, Schrank IS: **The electron transfer flavoprotein *fixABCX* gene products from *Azospirillum brasilense* show a NifA-dependent promoter regulation.** *Curr Microbiol* 2004, **49**:267-273.
37. Preisig O, Anthamatten D, Hennecke H: **Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis.** *Proc Natl Acad Sci USA* 1993, **90**:3309-3313.
38. Preisig O, Zufferey R, Hennecke H: **The *Bradyrhizobium japonicum* *fixGHIS* genes are required for the formation of the high-affinity *cbb3*-type cytochrome oxidase.** *Arch Microbiol* 1996, **165**:297-305.
39. Casalot L, Rousset M: **Maturation of the [NiFe] hydrogenases.** *Trends Microbiol* 2001, **9**:228-237.
40. Manyani H, Rey L, Palacios JM, Imperial J, Ruiz-Argüeso T: **Gene products of the *hupGHJ* operon are involved in maturation of the iron-sulfur subunit of the [NiFe] hydrogenase from *Rhizobium leguminosarum* bv. *viciae*.** *J Bacteriol* 2005, **187**:7018-7026.
41. Malhotra M, Srivastava S: **An *ipdC* gene knock-out of *Azospirillum brasilense* strain SM and its implications on indole-3-acetic acid biosynthesis and plant growth promotion.** *Antonie van Leeuwenhoek* 2008, **93**:425-433.
42. Spaepen S, Vanderleyden J, Remans R: **Indole-3-acetic acid in microbial and microorganism-plant signaling.** *FEMS Microbiol Rev* 2007, **31**:425-448.
43. Kriebchaumer V, Park WJ, Piotrowski M, Meeley RB, Gierl A, Glawischnig E: **Maize nitrilases have a dual role in auxin homeostasis and beta-cyanoalanine hydrolysis.** *J Exp Bot* 2007, **58**:4225-4233.
44. Vorwerk S, Biernacki S, Hillebrand H, Janzik I, Müller A, Weiler EW, Piotrowski M: **Enzymatic characterization of the recombinant *Arabidopsis thaliana* nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster.** *Planta* 2001, **212**:508-516.
45. Loh J, Pierson EA, Pierson L3, Stacey G, Chatterjee A: **Quorum sensing in plant-associated bacteria.** *Curr Opin Plant Biol* 2002, **5**:285-290.
46. Boyer M, Bally R, Perrotto S, Chaintreuil C, Wisniewski-Dyé F: **A quorum-quenching approach to identify quorum-sensing-regulated functions in *Azospirillum lipoferum*.** *Res Microbiol* 2008, **159**:699-708.
47. Vial L, Cuny C, Gluchoff-Fiasson K, Comte G, Oger PM, Faure D, Dessaux Y, Bally R, Wisniewski-Dyé F: **N-acyl-homoserine lactone-mediated quorum-sensing in *Azospirillum*: an exception rather than a rule.** *FEMS Microbiol Ecol* 2006, **58**:155-168.
48. Park SY, Lee SJ, Oh TK, Oh JW, Koo BT, Yum DY, Lee JK: **AhID, an N-acylhomoserine lactonase in *Arthrobacter* sp., and predicted homologues in other bacteria.** *Microbiology* 2003, **149**:1541-1550.
49. Lerner A, Castro-Sowinski S, Valverde A, Lerner H, Dror R, Okon Y, Burdman S: **The *Azospirillum brasilense* Sp7 *noeJ* and *noeL* genes are involved in extracellular polysaccharide biosynthesis.** *Microbiology* 2009, **155**:4058-4068.
50. Zhulin IB, Armitage JP: **Motility, chemokinesis, and methylation-independent chemotaxis in *Azospirillum brasilense*.** *J Bacteriol* 1993, **175**:952-958.
51. Wadhams GH, Armitage JP: **Making sense of it all: bacterial chemotaxis.** *Nat Rev Mol Cell Biol* 2004, **5**:1024-1037.
52. Bible AN, Stephens BB, Ortega DR, Xie Z, Alexandre G: **Function of a chemotaxis-like signal transduction pathway in modulating motility, cell clumping, and cell length in the alphaproteobacterium *Azospirillum brasilense*.** *J Bacteriol* 2008, **190**:6365-6375.
53. Berleman JE, Bauer CE: **Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum centenum*.** *Mol Microbiol* 2005, **56**:1457-1466.
54. Jiang ZY, Gest H, Bauer CE: **Chemosensory and photosensory perception in purple photosynthetic bacteria utilize common signal transduction components.** *J Bacteriol* 1997, **179**:5720-5727.
55. Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA: **The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes.** *Annu Rev Cell Dev Biol* 1997, **13**:457-512.
56. Moens S, Michiels K, Keijers V, Van Leuven F, Vanderleyden J: **Cloning, sequencing, and phenotypic analysis of *laf1*, encoding the flagellin of the lateral flagella of *Azospirillum brasilense* Sp7.** *J Bacteriol* 1995, **177**:5419-5426.
57. Chang Y, Tang T, Li JL: **Isolation of a flagellar operon in *Azospirillum brasilense* and functional analysis of F1bD.** *Res Microbiol* 2007, **158**:521-528.
58. Koebernik R: **TonB-dependent trans-envelope signalling: the exception or the rule?** *Trends Microbiol* 2005, **13**:343-347.
59. Noinaj N, Guillier M, Barnard TJ, Buchanan SK: **TonB-dependent transporters: regulation, structure, and function.** *Annu Rev Microbiol* 2010, **64**:43-60.
60. Pao SS, Paulsen IT, Saier MH: **Major facilitator superfamily.** *Microbiol Mol Biol Rev* 1998, **62**:1-34.
61. Jones PM, George AM: **The ABC transporter structure and mechanism: perspectives on recent research.** *Cell Mol Life Sci* 2004, **61**:682-699.
62. Saier MH, Paulsen IT: **Phylogeny of multidrug transporters.** *Semin Cell Dev Biol* 2001, **12**:205-213.
63. Li XZ, Nikaido H: **Efflux-mediated drug resistance in bacteria: an update.** *Drugs* 2009, **69**:1555-1623.
64. Brown MH, Skurray RA: **Staphylococcal multidrug efflux protein QacA.** *J Mol Microbiol Biotechnol* 2001, **3**:163-170.
65. Chatterjee C, Paul M, Xie L, van der Donk WA: **Biosynthesis and mode of action of lantibiotics.** *Chem Rev* 2005, **105**:633-684.
66. Okuda Ki, Yanagihara S, Sugayama T, Zendo T, Nakayama J, Sonomoto K: **Functional significance of the E-Loop, a novel motif conserved in the lantibiotic-immunity ABC transport systems.** *J Bacteriol* 2010, **192**:2801-2808.
67. Cotter P, Hill C, Ross P: **Bacteriocins: developing innate immunity for food.** *Nat Rev Microbiol* 2005, **3**:777-788.
68. Jaubert M, Vuillet L, Hannibal L, Adriano JM, Fardoux J, Bouyer P, Bonaldi K, Fleischman D, Giraud E, Verméglio A: **Control of peripheral light-harvesting complex synthesis by a bacteriophytochrome in the aerobic photosynthetic bacterium *Bradyrhizobium* strain BTAi1.** *J Bacteriol* 2008, **190**:5824-5831.
69. Rockwell NC, Su YS, Lagarias JC: **Phytochrome structure and signaling mechanisms.** *Annu Rev Plant Biol* 2006, **57**:837-858.
70. Giraud E, Zappa S, Vuillet L, Adriano JM, Hannibal L, Fardoux J, Berthomieu C, Bouyer P, Pignol D, Verméglio A: **A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in *Rhodospseudomonas palustris*.** *J Biol Chem* 2005, **280**:32389-32397.
71. Davis SJ, Vener AV, Vierstra RD: **Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic Eubacteria.** *Science* 1999, **286**:2517-2520.
72. Roszak DB, Colwell RR: **Survival strategies of bacteria in the natural environment.** *Microbiol Rev* 1987, **51**:365-379.

73. Fred EB, Waskman SA: *Laboratory Manual of General Microbiology with Special Reference to the Microorganisms of the Soil* New York: McGraw-Hill Book Company, Inc; 1928.
74. Wilson K: **Preparation of genomic DNA from bacteria**. In *Current Protocols in Molecular Biology*. 1 edition. Edited by: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. New York: Wiley; 1997:2.
75. Almeida LGP, Paixão R, Souza RC, Costa GC, arrientos FJA, antos MT, Imeida DF, asconcelos ATR: **A System for Automated Bacterial (genome) Integrated Annotation–SABIA**. *Bioinformatics* 2004, **20**:2832-2833.
76. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM: **The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data**. *Nucleic Acids Res* 2007, **35**:D169-D172.
77. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis**. *Nucleic Acids Res* 2009, **37**:D141-D145.
78. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic Acids Res* 2004, **32**:1792-1797.
79. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**:4673-4680.
80. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**:1596-1599.

doi:10.1186/1471-2164-12-409

**Cite this article as:** Sant'Anna et al.: Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*. *BMC Genomics* 2011 **12**:409.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



## Research Article

# The Plant Growth-Promoting Bacteria *Azospirillum amazonense*: Genomic Versatility and Phytohormone Pathway

Ricardo Cecagno,<sup>1</sup> Tiago Ebert Fritsch,<sup>1</sup> and Irene Silveira Schrank<sup>1,2</sup>

<sup>1</sup> Centro de Biotecnologia, Laboratório de Microrganismos Diazotróficos, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

<sup>2</sup> Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), CP 15005, 91501-970 Porto Alegre, RS, Brazil

Correspondence should be addressed to Irene Silveira Schrank; irene@cbiot.ufrgs.br

Received 2 July 2014; Revised 24 October 2014; Accepted 24 October 2014

Academic Editor: You-Ping Deng

Copyright © Ricardo Cecagno et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rhizosphere bacterium *Azospirillum amazonense* associates with plant roots to promote plant growth. Variation in replicon numbers and rearrangements is common among *Azospirillum* strains, and characterization of these naturally occurring differences can improve our understanding of genome evolution. We performed an *in silico* comparative genomic analysis to understand the genomic plasticity of *A. amazonense*. The number of *A. amazonense*-specific coding sequences was similar when compared with the six closely related bacteria regarding belonging or not to the *Azospirillum* genus. Our results suggest that the versatile gene repertoire found in *A. amazonense* genome could have been acquired from distantly related bacteria from horizontal transfer. Furthermore, the identification of coding sequence related to phytohormone production, such as flavin-monooxygenase and aldehyde oxidase, is likely to represent the tryptophan-dependent TAM pathway for auxin production in this bacterium. Moreover, the presence of the coding sequence for nitrilase indicates the presence of the alternative route that uses IAN as an intermediate for auxin synthesis, but it remains to be established whether the IAN pathway is the Trp-independent route. Future investigations are necessary to support the hypothesis that its genomic structure has evolved to meet the requirement for adaptation to the rhizosphere and interaction with host plants.

## 1. Introduction

The genus *Azospirillum* comprises free-living, nitrogen-fixing bacteria that are known as plant growth-promoting rhizobacteria (PGPR), which can colonize, by adhesion, the root surface or the intercellular spaces of the host plant roots. The potential role of the PGPR in association with economically important cereals and other grasses is to promote plant growth by several mechanisms including nitrogen fixation and phytohormone production [1]. Several species of *Azospirillum* are able to secrete phytohormones such as auxins, gibberellins, cytokinins, and nitric oxide as signals of plant growth promotion [2, 3].

*Azospirillum* genomes, as previously suggested for various strains, are larger and are comprised of multiple replicons indicating a potential for genome plasticity [4]. Genomic

rearrangements can occur spontaneously where replicons can be lost upon the formation of new megaplasmids [5, 6]. Moreover, genome sequencing of some *Azospirillum* species revealed that significant part of the genome has been horizontally acquired [6]. Up until now, 16 *Azospirillum* species have been characterized; however complete genomic sequences of only *Azospirillum brasilense*, *Azospirillum lipoferum*, *Azospirillum* sp. B510, and a draft of *Azospirillum amazonense* genome have been published [7].

*Azospirillum amazonense* was found to be associated with the roots and rhizosphere of several grasses including sugarcane, maize, sorghum, and rice revealing a broad ecological distribution in Brazil. Studies revealed that *A. amazonense* is phylogenetically closer to *Rhodospirillum centenum* and *Azospirillum irakense* than to *A. brasilense*. Unlike other *Azospirillum* strains, *A. amazonense* can grow in the presence

of sucrose as sole carbon source and is also better adapted to soil acidity, which offers the bacterium additional advantages for colonization of plant root tissue in acid environments [8, 9]. Moreover, *A. amazonense* genomic analyses revealed the presence of genes not commonly distributed in other *Azospirillum* species such as those responsible for the utilization of salicin as carbon source (similar to *A. irakense*) and a gene cluster (RubisCO) implicated in carbon fixation (*A. lipoferum* is able to grow autotrophically by means of RubisCO, but the presence of the genes has not yet been demonstrated) [7]. However, our understanding of phytohormone production in *A. amazonense* is still incomplete.

The genomic plasticity of *A. amazonense* is probably related to the versatile gene repertoire present in the genome of this bacterium suggesting that horizontal gene transfer may have an impact on the adaptation and evolution of this species. Gene organization and phylogenetic analysis demonstrated that genes coding for proteins responsible for the nitrogen fixation process, carbon fixation (RubisCOs), and molecular hydrogen oxidation (hydrogenases) is more closely related to Rhizobiales members than to related species [7].

To further examine the importance of *A. amazonense* genetic variability, an *in silico* comparative genomic analysis using subtractive hybridization was performed using total coding sequences (CDS) from *A. amazonense* to compare with genomes of closely related bacteria. The analysis of conserved and specific *A. amazonense* coding sequences indicated features that distinguished *A. amazonense* from other *Azospirillum* species. Furthermore, the specific interesting features related to phytohormone production may provide several cues to establish *A. amazonense* pathways for auxin biosynthesis.

## 2. Material and Methods

**2.1. Bacteria Selection and Genome Access.** We have previously generated a good quality draft genome sequence of the *A. amazonense* Y2 (ATCC 35120) strain [7]. In this paper, the draft genome sequences were annotated and analyzed for the presence of specific regions, and during the BLAST search best-hits were detected with different bacteria, such as *Rhodospirillum*, *Azospirillum*, *Bradyrhizobium*, and *Caulobacter*.

Therefore, the *A. amazonense* comparative genomic analyses were performed using bacterial genomes including six species for which publicly closed genomes were available (Table 1). All genomes were downloaded from NCBI on January 10, 2013. The accession numbers used in this study are *Azospirillum amazonense* Y2 PRJNA73583, PRJNA65263; *Azospirillum* sp. B510 projects PRJNA46085, PRJDA32551; *Azospirillum brasilense* Sp245 PRJEA162161, PRJEA70627; *Azospirillum lipoferum* 4B PRJNA82343, PRJEA50367; *Rhodospirillum centenum* SW project PRJNA58805; *Bradyrhizobium japonicum* USDA 110 projects PRJNA57599 and PRJNA17; and *Caulobacter segnis* ATCC 21756 project PRJNA41709.

TABLE 1: General features for the bacteria genomes used in the comparative analysis.

Bacteria	Genome size	Total number of CDS	Assembly reference number
<i>Azospirillum amazonense</i>	7,044,835	3,319*	ASM22599v1
<i>Azospirillum brasilense</i>	7,530,241	7,557	ASM23736v1
<i>Azospirillum lipoferum</i>	6,846,400	6,093	ASM28365v1
<i>Azospirillum</i> sp. B510	7,599,738	6,309	ASM1072v1
<i>Rhodospirillum centenum</i>	4,355,543	4,003	ASM1618v1
<i>Caulobacter segnis</i>	4,655,622	4,139	ASM9228v1
<i>Bradyrhizobium japonicum</i>	9,105,828	8,317	ASM1136v1

\*The total number of *A. amazonense* CDS was published by Sant'Anna et al. 2011 [7].

**2.2. Annotation and Subtractive Hybridization.** Reannotation of *A. amazonense* protein-coding genes was performed with a following procedure, which consists of two phases: initially the *A. amazonense* contigs were compared with the *Azospirillum* sp. B510 genome followed by functional annotation of each coding sequence (CDS) based on comparison with known sequences of the other six selected genomes using the Xbase Annotation Service [10]. All coding sequences predictions were manually checked for conservation in case of multiple hits, and only the alignments with best-hit results were selected from each genome. Information related to Cluster of Orthologous Group and KEGG pathway was added to the annotation using the server for metagenomic analysis (WebMGA) [11]. Annotation was based on comparison to protein clusters and on the BLAST results.

The subtractive hybridization using the mGenomeSubtractor program [12] was applied to run BLAST searches of the *A. amazonense* genome against multiple bacterial genomes for *in silico* comparative genomic analyses in order to characterize the unique sequences of *A. amazonense*. Proteins possibly related to phytohormones were analyzed in the Arabidopsis Hormone Database (AHD) [13], and proteins with homology (*H*) values more than 0.1 were arbitrarily defined as conserved coding sequences.

## 3. Results and Discussion

**3.1. Comparative Analyses and Specific Protein Coding Sequences.** The draft genome sequence of *A. amazonense* consists of 7,044,835 bp with 3,319 predicted coding sequences (CDS) where 2,299 have similarity with genes with known functions and 1,020 codes for hypothetical proteins or proteins of unknown function [7]. Although the estimated coverage of the genome was 35x, the number of predicted coding sequences was lower when compared with the other

species of *Azospirillum* where the total number of coding sequences ranges from 6,093 to 7,557 (Table 1).

In order to clarify the genomic coding sequences content of *A. amazonense*, two alternative comparative approaches using the Xbase Annotation Service were performed. Initially, to assess the coverage of the predicted gene repertoires a BLAST search was performed with only the *Azospirillum* sp. B510 genome. The total number of predicted protein-coding genes was 5,496 of which 2,165 were annotated as proteins of unknown function or hypothetical proteins. These numbers are similar to what is found in *A. lipoferum* and *Azospirillum* sp. B510 (Table 1). These results including the 5,496 sequences can be accessed using the <http://www.xbase.ac.uk/annotation/results/rWn50Rn6LVrucf55SWfsvHfoqdpmd655/>.

The second approach used an *A. amazonense*-vs-all (six selected genomes) BLAST to examine the overall similarity of the *A. amazonense* genome with closely related bacteria, and the results are shown in Table 2. The whole-genome comparisons revealed that the number of coding sequences found to be conserved and characterized as best-hits varied in each bacterium, from 3,126 (present in *Azospirillum* sp. B510) and 1,508 (present in *Rhodospirillum centenum*) to 2,846 (present in *R. centenum*) and 440 proteins (present in *A. lipoferum*), respectively. It is important to point out that the majority of the orthologs showing best-hit results were found with the *R. centenum* genome supporting previous suggestions of a close evolutionary relationship between *A. amazonense* and *R. centenum* [7, 14]. Interestingly, the number of coding sequences with best-hits orthologs found in the other *Azospirillum* species is almost equivalent to those found in the genome of bacteria from other genera, such as *Bradyrhizobium japonicum* and *Caulobacter segnis* (Table 2).

These unexpected results may support previous reports related with the genome repertoire of *A. amazonense* where horizontal gene transfer may be one of several events that result in an intragenera genomic plasticity. Phylogenetic analysis indicated the close relationship of the *A. amazonense* enzymes encoded by the gene cluster related to carbon fixation (RubisCo) and by genes related to nitrogen fixation (*nif*) processes with those from some species of the order Rhizobiales. Moreover, some features, such as the genetic organization of the carbon-fixation cluster and of the *nif* cluster of *A. amazonense*, are similar to the homolog cluster of the *Bradyrhizobium* species [7].

In conclusion, from the total 5,496 CDS found in the *A. amazonense* genome approximately half of the coding sequences have an ortholog in other closely related bacteria. However, using this methodology's numbers varying from 2,370 to 2,650 CDS showed lower degrees of similarity ( $E$  value  $> 10^{-10}$ ) with coding sequences present in the genome of the compared bacteria.

Comparative genomic analysis using *in silico* subtractive hybridization allowed searching for specific proteins of the *A. amazonense* genome against multiple closely related bacterial genomes. Therefore, to determine the possible differences between the *A. amazonense* genome and each of the selected six closely related genomes, an *in silico* subtractive

TABLE 2: Predicted distribution of coding sequences (CDS) in *A. amazonense* draft genome and in the complete genome of other bacteria.

Comparisons	Conserved CDS		Specific CDS
	Best-hits	Total number	
<i>A. a.</i> versus <i>Azospirillum brasilense</i>	583	3,031	2,465
<i>A. a.</i> versus <i>Azospirillum lipoferum</i>	440	3,084	2,412
<i>A. a.</i> versus <i>Azospirillum</i> sp. B510	533	3,126	2,370
<i>A. a.</i> versus <i>Rhodospirillum centenum</i>	1,508	2,846	2,650
<i>A. a.</i> versus <i>Caulobacter segnis</i>	711	2,852	2,644
<i>A. a.</i> versus <i>Bradyrhizobium japonicum</i>	632	2,970	2,526

*A. a.*: *Azospirillum amazonense*.

Protein coding sequences with  $E$  value  $>10^{-10}$  were considered specific CDS (using the Xbase Annotation Service).

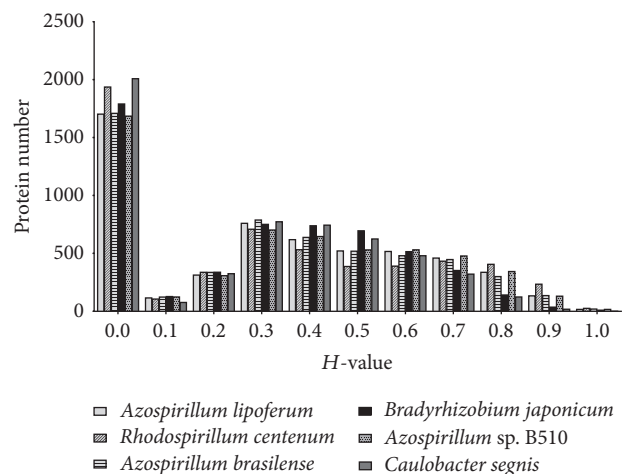


FIGURE 1: Histogram distribution of predicted proteins in *A. amazonense* compared with six closely related genomes using BLASTP-based homology value ( $H$  value). The  $H$ -value reflects the degree of similarity in terms of length of match and the degree of identity at amino acid level between the matching CDS in the subject genome and the query CDS examined with  $E$  value  $> 10^{-8}$ .

hybridization technique was applied. The histogram of  $H$ -values (Figure 1) was used to set the cutoff to discriminate between *A. amazonense*-specific and conserved coding sequences. Proteins with homology ( $H$ ) values of less than 0.42 and more than 0.64 were arbitrarily defined as specific and conserved coding sequences, respectively [12]. This cutoff value was proposed by Shao et al. [12] and has been used in comparative genomic analyses to differentiate strains of pseudomonads [12, 15] or to compare genomes of species from the genus *Erwinia* [16].

TABLE 3: Numbers of specific proteins for *A. amazonense* genome against six closely related genomes.

Comparisons	<i>A. amazonense</i>		
	Specific CDS	Conserved CDS	Other CDS
<i>Azospirillum brasilense</i>	3,689	948	859
<i>Azospirillum lipoferum</i>	3,606	746	1,144
<i>Azospirillum</i> sp. B510	3,571	793	1,132
<i>Rhodospirillum centenum</i>	3,697	770	1,029
<i>Caulobacter segnis</i>	4,043	382	1,071
<i>Bradyrhizobium japonicum</i>	3,880	317	1,299

The *in silico* subtractive hybridization analysis was performed with *A. amazonense* total coding sequences (CDS) against the proteins from the six genomes.

Proteins with homology (*H*) value less than 0.42 and more than 0.64 were arbitrarily defined as specific and conserved CDS, respectively, and other CDS were defined with *H* values between 0.42 and 0.64.

The subtractive hybridization approach revealed different profiles in gene number of specific and conserved proteins for the *A. amazonense* genome against the six others (Table 3). The number of proteins found to be conserved varied in each bacterium, from 948 CDS in *A. lipoferum* to 317 CDS in *B. japonicum*. Interestingly, the number of *A. amazonense*-specific proteins was similar when compared with the six bacteria varying from 4,043 (*C. segnis*) to 3,571 (*Azospirillum* sp. B510). Moreover, the specific proteins vary among the *Azospirillum* genomes analyzed from 3,571 to 3,689 only, indicating that these coding sequences are unique to the *A. amazonense* genome. Analyses of Figure 1 show that the majority of specific proteins in *A. amazonense* have *H*-values less than 0.1, suggesting that *Azospirillum* species are evolutionally diverse. This is consistent with previous studies that had proposed that some regions of the genome of *Azospirillum* species were acquired from distantly related bacteria from horizontal transfer [6].

To further characterize the global profile of *A. amazonense*-specific coding sequences, an *in silico* subtractive hybridization comparative analysis was performed with total *A. amazonense* putative coding sequences versus all six genomes, simultaneously. A total of 142 conserved CDS and 2,483 specific CDS were identified (see Supplementary Table 1 in Supplementary Material available online at <http://dx.doi.org/10.1155/2014/898592>). For function classification and pathway assignment, all specific and conserved *A. amazonense* coding sequences were classified to 20 different functional classes based on Clusters of Orthologous Groups (COG) (Table 4). The comparison of the *A. amazonense* genome and the other six available closely related genomes with regard to the functional category revealed that 1,196 specific CDS from *A. amazonense* were distributed among the different classes of orthologous clusters and that 1,287 specific CDS were unclassified being considered hypothetical products.

TABLE 4: Protein categories encoded by *A. amazonense* specific and conserved genes identified by *in silico* subtractive hybridization.

CDS assigned function*	Specific CDS	Conserved CDS
Transcription	143	11
Signal transduction mechanisms	137	5
Inorganic ion transport and metabolism	111	1
Carbohydrate transport and metabolism	101	4
Cell wall/membrane/envelope biogenesis	94	1
Amino acid transport and metabolism	86	19
Energy production and conversion	49	27
Secondary metabolites biosynthesis, transport, and catabolism	47	3
Cell motility	44	0
Coenzyme transport and metabolism	41	4
Lipid transport and metabolism	41	10
Intracellular trafficking, secretion, and vesicular transport	37	1
Defense mechanisms	34	0
Replication, recombination, and repair	33	3
Posttranslational modification, protein turnover, and chaperones	27	16
Translation, ribosomal structure, and biogenesis	14	27
Nucleotide transport and metabolism	12	12
Cell cycle control, cell division, and chromosome partitioning	4	1
RNA processing and modification	2	0
General function prediction only or function unknown	352	11

Proteins with homology (*H*) value less than 0.42 and more than 0.64 were arbitrarily defined as specific and conserved CDS, respectively.

\*CDS assigned function was based on the COGs according to BLAST search.

Detailed analysis of the *A. amazonense*-specific CDS from Table 4 (and Supplementary Table 1) indicates special attention to the coding sequences classified in the Signal Transduction Mechanisms and Secondary Metabolites Biosynthesis, Transport and Catabolism functional class. Among the 137 CDS classified in the Signal Transduction Mechanisms functional category, there were protein coding sequences similar to cytokinins (ZP\_08868173.1, ZP\_08868457.1) and ethylene response (ZP\_08867667.1) that could be related to phytohormone production (Supplementary Table 1). In particular, we have paid attention to a coding sequence related to lysine/ornithine N-monooxygenase (ZP\_08869952.1) similar to flavin-containing monooxygenase from *Arabidopsis thaliana* (YUCCA9) involved in auxin synthesis in plants [17, 18] found in the Secondary Metabolites Biosynthesis, Transport and Catabolism functional class (Supplementary Table 1). Therefore, to better understand the auxin biosynthesis pathways in *A. amazonense*, studies attempting to define coding sequence related to phytohormone production were performed.

**3.2. Phytohormone Production Related Sequences.** The improvement of plant growth upon *Azospirillum* inoculation is attributed, as one of many factors, to the production of auxin by these bacteria [19]. Indole-3-acetic acid (IAA) is considered the most important auxin implicated in different aspects of plant growth. In bacteria, the two most common routes for indole-3-acetic acid biosynthesis are the IAM (indole-3-acetamide) and the IPyA (indole-3-pyruvate) pathways [20]. However, in *A. brasilense*, besides these two tryptophan-dependent pathways, an additional tryptophan-independent pathway was identified [21].

Similar to other *Azospirillum* species, *A. amazonense*, as typical plant-growth promoting rhizobacteria, stimulate root proliferation [22, 23]. However, genes responsible for biosynthesis and secretion of phytohormones are poorly described in this species. Previous works on *A. amazonense* genome sequence and annotation were unable to localize genes related to the IAM or IPyA pathways (*iaaM*, *iaaH*, and *ipdC*) and were able to identify only the presence of a coding sequence similar to nitrilases responsible for the conversion of indole 3-acetonitrile (IAN) to IAA in plants [7]. Therefore, the presence of coding sequences homologous to nitrilase and to flavin-containing monooxygenase (this paper) suggests that *A. amazonense* could use alternative pathways closely related to those found in plants.

Biosynthetic pathways for IAA have been fully investigated and tryptophan-dependent and Trp-independent routes have been studied [20, 24, 25]. Although genes coding for proteins related to the bacterial common routes IAM (indole-3-acetamide route) and IPyA (indole-3-pyruvic route) was not found in the *A. amazonense* genome, the identification of flavin-monooxygenase and nitrilase enzymes suggests the presence of the TAM (tryptamine route) and IAN (indole-3-acetamide route) pathways for IAA synthesis in this bacterium (Figure 2). It is well known that nitrilases in plants (maize and *Arabidopsis thaliana*) and also in *Bacillus amyloliquefaciens* were shown to hydrolyze indole-3-acetonitrile (IAN) to IAA [25, 26]. Moreover, evidence for the IAN and TAM pathways has been reported in *A. brasilense* [21].

Aiming to unveil the IAA pathways in *A. amazonense*, a search for other enzymes involving the TAM pathway was performed. The genome of *A. amazonense* contains an aldehyde oxidase-coding sequence (WP\_004273557) homolog (query cover 91%; 33% identity; *E* value  $e - 99$ ) to the *A. thaliana* AAO1 gene (AED92912) that is capable of oxidizing indole-3-acetaldehyde to indole-3-acetic acid with high efficiency [27]. Therefore, oxidation of indole-3-acetaldehyde by *A. amazonense* aldehyde oxidase is likely to represent the TAM route transforming indole-3-acetaldehyde (IAAld) to produce IAA phytohormone in this bacterium (Figure 2). To conclude, *A. amazonense* appears to possess only one regulated Trp-dependent route for IAA synthesis, the TAM pathway, while *A. brasilense* possesses two differently regulated routes, namely, the IPyA and the TAM pathways [21]. Furthermore, the alternative route that uses IAN as an intermediate, the IAN pathway, appears to be present in both species, but it remains to be established whether the IAN pathway is the Trp-independent route.

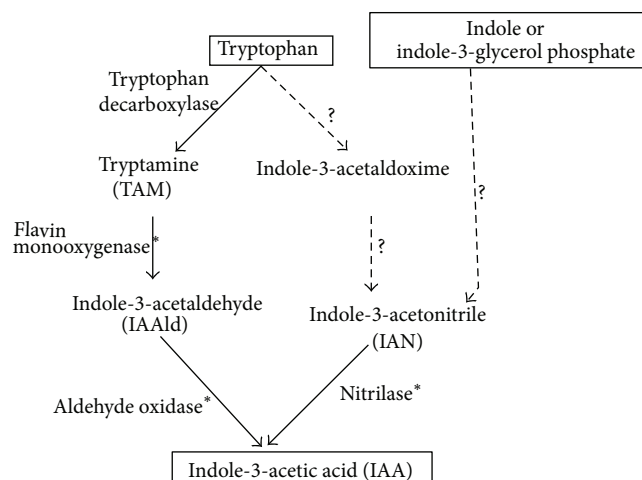


FIGURE 2: *A. amazonense* pathways of IAA biosynthesis. Tryptophan-dependent pathways or tryptophan-independent pathways (starting from indole or indole-3-glycerol phosphate) are indicated based on routes found in plants and bacteria. Enzymes indicated with an asterisk have been identified in *A. amazonense*, and routes indicated as dotted lines indicated that the precursor of IAN may or may not be tryptophan.

To further understand the phytohormone biosynthesis pathway in *A. amazonense*, an *in silico* comparative analysis was performed with total coding sequences from *A. amazonense* versus all proteins deposited in the Arabidopsis Hormone Database. Furthermore, the comparative analysis was also performed with coding sequences in the auxin response transcriptome data of *A. brasilense* [28]. A total of 54 *A. amazonense* CDS revealed similarity with proteins related to hormone production in plants and bacterial auxin signal transduction pathways (Supplementary Table 1). The presence of these coding sequences suggests that IAA could be a signal that alters gene expression in *A. amazonense* similar to that found in *A. brasilense* [28]. Moreover, the genome of *A. amazonense* has coding sequences that could be related to other hormone pathways similar to those described for other *Azospirillum* species [19, 21].

Another beneficial effect provided by the association of soil bacteria with plants could be due to the plant hormone ethylene, which can inhibit plant growth by regulating several developmental aspects [29]. Similar to other plant growth-promoting rhizobacteria, a coding sequence homologous to 1-aminocyclopropane-1-carboxylate (ACC) deaminase (WP\_004272971.1) has been identified in the *A. amazonense* genome. Previous reports have suggested that *A. amazonense* can stimulate plant growth by producing or metabolizing plant hormones and the presence of ACC deaminase which can hydrolyze ACC, the immediate precursor of the plant hormone ethylene, could be involved by lowering the plant ethylene levels and increasing plant growth. Moreover, the identification of the octaprenyl diphosphate synthase enzyme (ZP\_08868744) could be related to cytokinin biosynthesis by a known hormone that affects plant growth and yield. Although the auxin hormone is considered a major

class of hormones regulating plant growth, cytokinins or ethylene-related phytohormones could interact with auxin leading to root system development.

In conclusion, it appears that the rhizosphere bacterium *A. amazonense* is able to produce IAA through the tryptamine and indole-3-acetonitrile pathways and similar to *A. brasilense* could alter gene expression in response to the presence of auxin. Moreover, the role as plant-growth-promoting bacteria could be related to IAA production or to its ability to metabolize the ethylene precursor (ACC) and thereby increases the growth of the root system. Furthermore, the multiple genome comparison performed with *A. amazonense* and closely related bacteria supports previous evidence concerning *A. amazonense* genomic versatility and that several genes could have been acquired from distantly related bacteria [6]. Future investigation of *A. amazonense* is necessary to support the hypothesis that its genomic structures have evolved to meet the requirements for adaptation to the rhizosphere and interaction with host plants.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

### Acknowledgments

This work was supported by grants from the Brazilian National Research Council (CNPq) and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). R. Cecagno and T. E. Fritsch received scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

### References

- [1] Y. Bashan, G. Holguin, and L. E. de-Bashan, "Azospirillum-plant relationships: physiological, molecular, agricultural, and environmental advances (1997–2003)," *Canadian Journal of Microbiology*, vol. 50, no. 8, pp. 521–577, 2004.
- [2] S. Fibach-Paldi, S. Burdman, and Y. Okon, "Key physiological properties contributing to rhizosphere adaptation and plant growth promotion abilities of *Azospirillum brasilense*," *FEMS Microbiology Letters*, vol. 326, no. 2, pp. 99–108, 2012.
- [3] M. Kochar and S. Srivastava, "Surface colonization by *Azospirillum brasilense* SM in the indole-3-acetic acid dependent growth improvement of sorghum," *Journal of Basic Microbiology*, vol. 52, no. 2, pp. 123–131, 2012.
- [4] C. C. G. Martin-Didonet, L. S. Chubatsu, E. M. Souza et al., "Genome structure of the genus *Azospirillum*," *Journal of Bacteriology*, vol. 182, no. 14, pp. 4113–4116, 2000.
- [5] A. V. Shelud'ko, O. E. Varshalomidze, L. P. Petrova, and E. I. Katsy, "Effect of genomic rearrangement on heavy metal tolerance in the plant-growth-promoting rhizobacterium *Azospirillum brasilense* Sp245," *Folia Microbiologica*, vol. 57, no. 1, pp. 5–10, 2012.
- [6] F. Wisniewski-Dyé, K. Borziak, G. Khalsa-Moyers et al., "Azospirillum genomes reveal transition of bacteria from aquatic to terrestrial environments," *PLoS Genetics*, vol. 7, no. 12, Article ID e1002430, 2011.
- [7] F. H. Sant'Anna, L. G. P. Almeida, R. Cecagno et al., "Genomic insights into the versatility of the plant growth-promoting bacterium *Azospirillum amazonense*," *BMC Genomics*, vol. 12, article 409, 2011.
- [8] F. M. Magalhães, J. I. Baldani, S. M. Souto, J. R. Kuykendall, and J. Dobreiner, "A new acid-tolerant *Azospirillum* species," *Anais da Academia Brasileira de Ciências*, vol. 55, pp. 417–430, 1983.
- [9] J. I. Baldani and V. L. D. Baldani, "History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience," *Anais da Academia Brasileira de Ciências*, vol. 77, no. 3, pp. 549–579, 2005.
- [10] R. R. Chaudhuri, N. J. Loman, L. A. S. Snyder, C. M. Bailey, D. J. Stekel, and M. J. Pallen, "xBASE2: a comprehensive resource for comparative bacterial genomics," *Nucleic Acids Research*, vol. 36, no. 1, pp. D543–D546, 2008.
- [11] S. Wu, Z. Zhu, L. Fu, B. Niu, and W. Li, "WebMGA: a customizable web server for fast metagenomic sequence analysis," *BMC Genomics*, vol. 12, article 444, 2011.
- [12] Y. Shao, X. He, E. M. Harrison et al., "mGenomeSubtractor: a web-based tool for parallel in silico subtractive hybridization analysis of multiple bacterial genomes," *Nucleic Acids Research*, vol. 38, no. 2, Article ID gkq326, pp. W194–W200, 2010.
- [13] Z.-Y. Peng, X. Zhou, L. Li et al., "Arabidopsis hormone database: a comprehensive genetic and phenotypic information database for plant hormone research in *Arabidopsis*," *Nucleic Acids Research*, vol. 37, no. 1, pp. D975–D982, 2009.
- [14] M. Stoffels, T. Castellanos, and A. Hartmann, "Design and application of new 16S rRNA-targeted oligonucleotide probes for the *Azospirillum-Skermanella-Rhodocista*-cluster," *Systematic and Applied Microbiology*, vol. 24, no. 1, pp. 83–97, 2001.
- [15] M. Qi, D. Wang, C. A. Bradley, and Y. Zhao, "Genome sequence analyses of *Pseudomonas savastanoi* pv. *glycinea* and subtractive hybridization-based comparative genomics with nine pseudomonads," *PLoS ONE*, vol. 6, no. 1, Article ID e16451, 2011.
- [16] Y. Zhao and M. Qi, "Comparative genomics of *Erwinia amylovora* and related *Erwinia* species—what do we learn?" *Genes*, vol. 2, no. 3, pp. 627–639, 2011.
- [17] Y. Zhao, S. K. Christensen, C. Fankhauser et al., "A role for flavin monooxygenase-like enzymes in auxin biosynthesis," *Science*, vol. 291, no. 5502, pp. 306–309, 2001.
- [18] Y. Zhao, "Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants," *Molecular Plant*, vol. 5, no. 2, pp. 334–338, 2012.
- [19] O. Ona, J. van Impe, E. Prinsen, and J. Vanderleyden, "Growth and indole-3-acetic acid biosynthesis of *Azospirillum brasilense* Sp245 is environmentally controlled," *FEMS Microbiology Letters*, vol. 246, no. 1, pp. 125–132, 2005.
- [20] M. Lambrecht, Y. Okon, A. V. Broek, and J. Vanderleyden, "Indole-3-acetic acid: a reciprocal signalling molecule in bacteria-plant interactions," *Trends in Microbiology*, vol. 8, no. 7, pp. 298–300, 2000.
- [21] R. Carreño-Lopez, N. Campos-Reales, C. Elmerich, and B. E. Baca, "Physiological evidence for differently regulated tryptophan-dependent pathways for indole-3-acetic acid synthesis in *Azospirillum brasilense*," *Molecular and General Genetics*, vol. 264, no. 4, pp. 521–530, 2000.
- [22] O. Steenhoudt and J. Vanderleyden, "Azospirillum, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects," *FEMS Microbiology Reviews*, vol. 24, no. 4, pp. 487–506, 2000.



- [23] E. Rodrigues, L. Rodrigues, A. de Oliveira et al., "Azospirillum amazonense inoculation: effects on growth, yield and N<sub>2</sub> fixation of rice (*Oryza sativa* L.)," *Plant and Soil*, vol. 316, no. 1-2, p. 323, 2009.
- [24] B. Bartel, "Auxin biosynthesis," *Annual Review of Plant Biology*, vol. 48, no. 1, pp. 51-66, 1997.
- [25] E. E. Idris, D. J. Iglesias, M. Talon, and R. Borriss, "Tryptophan-dependent production of Indole-3-Acetic Acid (IAA) affects level of plant growth promotion by *Bacillus amyloliquefaciens* FZB42," *Molecular Plant-Microbe Interactions*, vol. 20, no. 6, pp. 619-626, 2007.
- [26] V. Kriechbaumer, W. J. Park, M. Piotrowski, R. B. Meeley, A. Gierl, and E. Glawischnig, "Maize nitrilases have a dual role in auxin homeostasis and  $\beta$ -cyanoalanine hydrolysis," *Journal of Experimental Botany*, vol. 58, no. 15-16, pp. 4225-4233, 2007.
- [27] M. Seo, S. Akaba, T. Oritani et al., "Higher activity of an aldehyde oxidase in the auxin-overproducing superroot1 mutant of *Arabidopsis thaliana*," *Plant Physiology*, vol. 116, no. 2, pp. 687-693, 1998.
- [28] S. van Puyvelde, L. Cloots, K. Engelen et al., "Transcriptome analysis of the rhizosphere bacterium *Azospirillum brasilense* reveals an extensive auxin response," *Microbial Ecology*, vol. 61, no. 4, pp. 723-728, 2011.
- [29] L. Chen, I. C. Dodd, J. C. Theobald, A. A. Belimov, and W. J. Davies, "The rhizobacterium *Variovorax paradoxus* 5C-2, containing ACC deaminase, promotes growth and development of *Arabidopsis thaliana* via an ethylene-dependent pathway," *Journal of Experimental Botany*, vol. 64, no. 6, pp. 1565-1573, 2013.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>





## Ricardo Cecagno

Endereço para acessar este CV: <http://lattes.cnpq.br/3987413150425325>

Última atualização do currículo em 21/07/2014

Doutor em Ciências Biológicas com ênfase molecular, celular e funcional pela Universidade Federal do Rio Grande do Sul. Tem experiência na área de genética molecular, principalmente nos seguintes temas: Bioinformática aplicada, sequenciamento, anotação e caracterização de genomas; análise de expressão diferencial de genes; clonagem e transformação bacteriana. Atualmente está vinculado à Universidade Federal da Fronteira Sul em Chapecó-SC, atuando na Secretaria Especial de Laboratórios. **(Texto informado pelo autor)**



### Identificação

**Nome** Ricardo Cecagno   
**Nome em citações bibliográficas** CECAGNO, R.;Cecagno, Ricardo

### Endereço

**Endereço Profissional** Universidade Federal da Fronteira Sul, Secretaria Especial de Laboratórios.  
 Avenida Fernando Machado - 108E  
 Centro  
 89802112 - Chapecó, SC - Brasil  
 Telefone: (49) 20493108  
 URL da Homepage: [http://www.uffs.edu.br/index.php?option=com\\_content&view=article&id=2854&Itemid=1608&site=selab](http://www.uffs.edu.br/index.php?option=com_content&view=article&id=2854&Itemid=1608&site=selab)

### Formação acadêmica/titulação

- 2007 - 2013** Doutorado em Biologia Celular e Molecular (Conceito CAPES 6).  
 Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.  
 Título: GENÔMICA COMPARATIVA ENTRE ESPÉCIES DO GÊNERO AZOSPIRILLUM, COM ÊNFASE NAS CARACTERÍSTICAS EXCLUSIVAS DA BACTÉRIA PROMOTORA DE CRESCIMENTO VEGETAL AZOSPIRILLUM AMAZONENSE, Ano de obtenção: 2013.  
 Orientador:  Irene Silveira Schrank.  
 Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
- 2006 - 2007** Mestrado em Biologia Celular e Molecular (Conceito CAPES 6).  
 Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.  
 Título: ISOLAMENTO DE SEQÜÊNCIAS DE AZOSPIRILLUM AMAZONENSE EXPRESSAS EM CONDIÇÃO DE ESTRESSE OXIDATIVO, Ano de Obtenção: 2007.  
 Orientador:  IRENE SILVEIRA SCHRANK.  
 Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.
- 2000 - 2005** Graduação em Ciências Biológicas.  
 Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.  
 Título: Caracterização de genes específicos de Azospirillum amazonense.  
 Orientador: Irene Silveira Schrank.  
 Bolsista do(a): Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

## Formação Complementar

<b>2004 - 2004</b>	Sequenciamento e análise de genomas II. (Carga horária: 40h). Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.
<b>2004 - 2004</b>	Análise de diferença representacional (RDA). (Carga horária: 3h). Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.
<b>2003 - 2003</b>	Extensão universitária em E por falar em genética.... (Carga horária: 45h). Universidade Federal do Rio Grande do Sul, UFRGS, Brasil.
<b>2003 - 2003</b>	DNA microarrays e o perfil da expressão gênica. (Carga horária: 3h). Sociedade Brasileira de Genética.
<b>2002 - 2002</b>	Expressão gênica: novas abordagens de análise. (Carga horária: 3h). Sociedade Brasileira de Genética.

## Áreas de atuação

<b>1.</b>	Grande área: Ciências Biológicas / Área: Genética / Subárea: Bioinformática.
<b>2.</b>	Grande área: Ciências Biológicas / Área: Genética / Subárea: Genética Molecular e de Microorganismos.
<b>3.</b>	Grande área: Ciências Biológicas / Área: Genética / Subárea: Genética Molecular e de Microorganismos/Especialidade: Microorganismos Diazotróficos Fixação Biológica do Nitrogênio.

## Idiomas

<b>Espanhol</b>	Compreende RazoavelmenteLê Bem.
<b>Inglês</b>	Compreende RazoavelmenteLê Bem, Escreve Razoavelmente.
<b>Italiano</b>	Compreende Razoavelmente, Fala Razoavelmente, Lê Razoavelmente.

## Produções

### Produção bibliográfica

### Artigos completos publicados em periódicos

Ordenar por

Ordem Cronológica ▼

**1.** ★ Sant'Anna, Fernando H ; Almeida, Luiz GP ; **Cecagno, Ricardo** ; Reolon, Luciano A ; Siqueira, Franciele M ; Machado, Maicon RS ; Vasconcelos, Ana TR ; Schrank, Irene S . Genomic insights into the versatility of the plant growth-promoting bacterium Azospirillum amazonense. BMC Genomics **JCR**, v. 12, p. 409, 2011.

**Citações:** **WEB OF SCIENCE**™ 7 | **SCOPUS** 10

**2.** ★ Sant Anna, Fernando Hayashi ; Trentini, Débora Broch ; Souto Weber, Shana ; **CECAGNO, R.** ; Silva, Sérgio Ceroni ; SCHRANK, Irene Silveira . The PII Superfamily Revised: A Novel Group and Evolutionary Insights. Journal of Molecular Evolution **JCR**, v. 68, p. 322-336, 2009.

**Citações:** **WEB OF SCIENCE**™ 24 | **SCOPUS** 31

**3.** ★ BROETTO, L. ; **CECAGNO, R.** . Stable transformation of Chromobacterium violaceum with a broad-host-range plasmid. Applied Microbiology and Biotechnology **JCR**, v. 71, p. 450-456, 2006.

**Citações:** **WEB OF SCIENCE**™ 1 | **SCOPUS** 2

### Resumos publicados em anais de congressos

1. FRITSCH, T. E. ; **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Clonagem e expressão dos genes acdRS de Bradyrhizobium japonicum USDA 110 em Azospirillum Brasilense. In: XXI Salão de Iniciação Científica, 2009, Porto Alegre. Clonagem e expressão dos genes acdRS de Bradyrhizobium japonicum USDA 110 em Azospirillum Brasilense, 2009.
2. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Isolamento de regiões genômicas específicas de Azospirillum amazonense comparativamente com A. brasilense. In: XXV Reunião de Genética de Microrganismos, 2006, São Pedro - SP. Isolamento de regiões genômicas específicas de Azospirillum amazonense comparativamente com A. brasilense, 2006.
3. MAURMANN, K. ; **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Caracterização de mutantes de Chromobacterium violaceum obtidos utilizando mutagênese aleatória por transposons. In: XVIII Salão de iniciação científica, 2006, Porto Alegre. Caracterização de mutantes de Chromobacterium violaceum obtidos utilizando mutagênese aleatória por transposons, 2006.
4. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Análise de sequencias diferenciais entre os diazotróficos Azospirillum amazonense e A. brasilense. In: 51º Congresso Nacional de Genética, 2005, Águas de Lindóia - SP. Análise de sequencias diferenciais entre os diazotróficos Azospirillum amazonense e A. brasilense, 2005.
5. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Caracterização de sequências diferenciais entre Azospirillum amazonense e A. brasilense isoladas por representational difference analysis. In: XVI Salão de Iniciação Científica, 2004, Porto Alegre. Caracterização de Sequências Diferenciais Entre Azospirillum amazonense e Azospirillum brasilense isoladas por representational difference analysis, 2004.
6. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Caracterização de sequencias genômicas diferenciais entre Azospirillum brasilense e A. amazonense. In: XXIV Reunião de Genética de Microrganismos, 2004, Gramado - RS, 2004.
7. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Isolamento de regiões genômicas diferenciais entre duas espécies do gênero Azospirillum. In: 49º Congresso Nacional de Genética, 2003, Águas de Lindóia, 2003.
8. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Análise de diferenças representativas entre duas espécies do gênero Azospirillum. In: XV Salão de Iniciação Científica, 2003, Porto Alegre, 2003.
9. **CECAGNO, R.** ; SOARES, Rene Arderius ; ROESCH, Luis L ; PASSAGLIA, Luciane Maria Pereira ; JACQUES, Guilherme . Identificação por métodos moleculares de organismos diazotróficos endofíticos e de solo. In: 48º Congresso Nacional de Genética, 2002, Águas de Lindóia, 2002.
10. **CECAGNO, R.** ; **SCHRANK, Irene Silveira** . Isolamento de regiões genômicas diferenciais entre duas espécies do gênero Azospirillum. In: XIV Salão de Iniciação Científica, 2002, Porto Alegre, 2002.

## Eventos

### Participação em eventos, congressos, exposições e feiras

1. XXV reunião de genética de microrganismos. Isolamento de regiões genômicas específicas de Azospirillum amazonense comparativamente com A. brasilense. 2006. (Congresso).
2. 51º Congresso Nacional de Genética. Análise de sequencias diferenciais entre os diazotróficos Azospirillum amazonense e A. brasilense. 2005. (Congresso).
3. XXIV reunião de genética de Microrganismos. Caracterização de sequencias genômicas diferenciais entre Azospirillum brasilense e A amazonense. 2004. (Congresso).
4. 49º Congresso Nacional de Genética. Isolamento de regiões genômicas diferenciais entre duas espécies do gênero Azospirillum. 2003. (Congresso).

5. XI Congresso Brasileiro de Biologia Celular. 2002. (Congresso).
6. 48º Congresso Nacional de Genética. Identificação por métodos moleculares de organismos diazotróficos endofíticos e de solo. 2002. (Congresso).
7. XI Reunião Estadual de Biotecnologia Vegetal. 2002. (Encontro).

#### Organização de eventos, congressos, exposições e feiras

1. **CECAGNO, R.** . XVI Salão de Iniciação Científica - Coordenador de sessão de apresentação de trabalhos. 2004. (Outro).

Página gerada pelo Sistema Currículo Lattes em 09/12/2014 às 18:55:02

Imprimir currículo