

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**CENTRO DE BIOTECNOLOGIA**  
**LABORATÓRIO DE DIAZOTRÓFICOS**  
**COMISSÃO DE GRADUAÇÃO DE BIOLOGIA**

Trabalho de Conclusão de Curso de Bacharelado em Ciências Biológicas

Artigo a ser submetido à *Genome Biology*

***Análise in silico de promotores  $\sigma^N$  em *Azospirillum amazonense****

Aluno: Maicon Ricardo Stange Machado

Orientadora: Irene Silveira Schrank

Co-orientador: Fernando Hayashi Sant'Anna

Porto Alegre, 2010

Categoria: Artigo de Pesquisa

## ***Análise in silico de promotores $\sigma^N$ em *Azospirillum amazonense****

**Maicon R S Machado\*, Fernando H Sant'Anna, Irene S Schrank**

Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves  
9500, Porto Alegre, Brasil

\* Autor para correspondência

Endereços eletrônicos dos autores:

MRSM: mecors@hotmail.com

FHS: fhsantanna@yahoo.com.br

ISS: irene@cbiot.ufrgs.br

## **RESUMO**

### **Introdução**

A bactéria diazotrófica *Azospirillum amazonense* compõe um grupo de microrganismos reconhecidamente capazes de promover o crescimento em plantas. Um mecanismo possivelmente envolvido com essa capacidade é a fixação biológica do nitrogênio, a qual provê o elemento nitrogênio para o desenvolvimento de lavouras. Esse processo é estreitamente regulado pela célula e dentre os componentes centrais desta rota está o fator de transcrição  $\sigma^N$ , uma subunidade alternativa da holoenzima RNA polimerase. O fator  $\sigma^N$  está relacionado com o reconhecimento de promotores vinculados a diversas rotas, dentre elas o metabolismo do nitrogênio.

### **Resultados**

Através de uma matriz de peso heteróloga foram detectados possíveis motivos de ligação a esse fator transcricional no genoma de *A. amazonense*. Utilizando uma abordagem que busca a conservação dos motivos em organismos relacionados, foram encontradas 28 seqüências a montante de genes relacionados com categorias funcionais distintas, destacando-se aqueles envolvidos com metabolismo de nitrogênio e síntese de flagelo. Essa abordagem permitiu ainda a detecção de promotores putativos, a montante de genes ainda não caracterizados funcionalmente.

### **Conclusão**

Enfim, este trabalho além de contribuir para a caracterização do genoma colaborará para desvendar os elementos vinculados ao *sigmulon* nesta bactéria.

**Palavras-chave:** *Azospirillum amazonense*, sigmaN, fixação biológica do nitrogênio, rizobactéria promotora de crescimento em plantas (PGPR), matrizes de peso.

## INTRODUÇÃO

A disponibilidade de nitrogênio é um fator limitante para agricultura. Por isso, muitas lavouras necessitam de fertilizantes nitrogenados para a sua manutenção. O aumento indiscriminado no uso de tais insumos, justificados para suprir um acréscimo contínuo na oferta e demanda da indústria cerealífera [1], culminará em uma sobrecarga de recursos naturais básicos [2]. No entanto, a inoculação de bactérias diazotróficas na lavoura poderia inverter essa tendência [3], pois são capazes de promover o crescimento vegetal através de mecanismos biológicos como, a disponibilização do nitrogênio atmosférico ( $N_2$ ), através da fixação biológica do nitrogênio (BNF, do inglês *biological nitrogen fixation*) e a síntese de fitormônios [2,4].

O gênero *Azospirillum* é reconhecido por compreender bactérias promotoras do crescimento em plantas (PGPR, do inglês *plant growth promoting rhizobacteria*) [3,5,6] e a espécie *Azospirillum amazonense*, alvo deste estudo, foi previamente identificada em associação com cereais tais como: milho, sorgo, arroz e cana-de-açúcar. Apesar disso, ainda pouco se sabe sobre suas características genéticas e fisiológicas [4]. Entretanto, estudos *in vitro* sugerem que a bactéria *A. amazonense* é capaz de fixar nitrogênio mesmo em concentrações maiores de amônia [7]. Além disso, não apresenta caráter desnitrificante [8] sendo bem adaptada a solos ácidos, característica predominante dos solos brasileiros [9]. Por fim, apesar da literatura considerar extensivamente a produção de fitormônios como principal causa para a promoção do crescimento em plantas através da associação com bactérias [10,11], em *A. amazonense*, segundo Rodrigues *et al.*, (2008), a influência poderia estar mais condicionada à fixação biológica do nitrogênio.

Portanto, aprofundar o conhecimento sobre esta espécie provavelmente produzirá valiosas informações, e uma das maneiras de fazê-lo é desvendar o genoma da bactéria. A nova geração de sequenciadores automáticos (NGS, do inglês *next-generation sequencing*) propiciou conquistar estes dados com relativo baixo custo e agilidade [12]. Desse modo, atualmente, há um aumento expressivo de seqüências nucleotídicas em diversos bancos de dados. O grande desafio é interpretar essa quantidade massiva de dados que está sendo gerada e uma forma que pode auxiliar nesse processo é a busca de padrões em genomas relacionados [13-15]. Dentro dessa perspectiva está o *Footprinting* Filogenético, o qual ainda pode ser utilizado como uma ferramenta para melhorar a anotação de genomas recentemente sequenciados, que analisa a conservação de motivos regulatórios em regiões a montante de genes ortólogos. A hipótese seria que entre grupos filogeneticamente relacionados, genes ortólogos provavelmente são regulados por um fator de transcrição (TF) comum [16,17]. Um dos vários meios para acessar esta informação é a elaboração de matrizes de pontuação posicionalmente específicas (PSSM, do inglês *position-specific scoring matrix*). Uma PSSM, a partir da freqüência nucleotídica nas diferentes posições da sua extensão, auxilia na detecção de motivos específicos no genoma-alvo [18-20]. Normalmente, essas matrizes são derivadas de seqüências sabidamente reconhecidas como sítios de ligação para uma dada proteína.

O fator transcricional  $\sigma^N$  (RpoN), uma subunidade da RNA polimerase dedicada ao reconhecimento de regiões promotoras -24/-12 de genes relacionados com diversas funções na célula, dentre elas o metabolismo do nitrogênio [21,22] é a proteína alvo deste estudo.

Visto a importância desse fator de transcrição, o objetivo do trabalho é avaliar a presença de promotores para  $\sigma^N$ , salientando genes envolvidos no metabolismo do nitrogênio e, posteriormente, propor um *sigmulon* para a bactéria *A. amazonense*.

## RESULTADOS E DISCUSSÃO

### Conservação e história evolutiva da proteína $\sigma^N$ de *A. amazonense*

A proteína  $\sigma^N$  é um fator transcricional alternativo caracterizado por se ligar a promotores -24/-12 sem necessidade do núcleo da RNA polimerase [23,24], carecendo o auxílio de proteínas ativadoras (EBPs, do inglês *enhancer binding proteins*), as quais reconhecem sítios a montante de seus promotores (UAS, do inglês *upstream activating sequence*), para a formação do complexo aberto na transcrição [25,26]. O fator de transcrição é sabidamente conservado entre bactérias [21,22]. A Tabela 1 mostra que a conservação de  $\sigma^N$  de *A. amazonense* chega próximo a 60% em espécies de subclasses filogeneticamente distintas, como é o caso de *Azoarcus sp.* e *Escherichia coli*. Além disso, é possível verificar em *A. amazonense* a presença de um domínio sabidamente reconhecido para ligação ao sítio -24 do promotor de genes regulados por  $\sigma^N$  denominado RpoN-Box (Figura 1B) [27]. A ligação do sítio -12 é mais complexa [28], havendo provável cooperação entre a região N-terminal e C-terminal (domínio hélice-volta-hélice) [29,30]. Um resíduo de arginina (R) na última hélice do domínio mostrou-se fundamental para a atividade de reconhecimento do DNA nesta porção (Figura 1A) [31]. Portanto, é plausível que os sítios reconhecidos por este fator de transcrição em *A. amazonense* sejam altamente conservados em relação a outras bactérias da classe Proteobactéria.

Análises filogenéticas preliminares de genes envolvidos com a fixação do nitrogênio, e.g. *nifHDK*, revelaram incongruências entre as topologias das árvores desses genes com a

topologia da árvore de rDNA 16S, indicando que possíveis eventos de transferência gênica lateral tiveram importância na evolução da bactéria *A. amazonense* [32]. Portanto, a história evolutiva da proteína  $\sigma^N$  (RpoN) foi reconstruída (Figura 2B). A árvore filogenética decorrente demonstrou topologia semelhante àquela apresentada pelo rDNA 16S (Figura 2A), mostrando que, se utilizássemos o 16S como padrão de evolução das espécies averiguadas, a hipótese mais parcimoniosa de como se deu a história evolutiva do  $\sigma^N$  em *Rhizobiales* e *Rhodospirillales* é a transferência vertical, em detrimento de processos como a transferência horizontal. Isto fica claro quando se observa a dicotomia entre os clados representativos das ordens *Rhodospirillales* e *Rhizobiales* presente tanto na árvore do rDNA 16S quanto na da proteína RpoN, apesar do número de espécies ser menor na última (Figura 2).

#### **Detecção *in silico* de promotores $\sigma^N$ no genoma de *A. amazonense***

Como relatado anteriormente, a metodologia de *Footprinting* filogenético baseia-se na premissa de que elementos transcricionais ortólogos reconhecem sítios com padrão nucleotídico conservado entre os organismos [16]. Visto que há alta conservação nos domínios do fator  $\sigma^N$  de *A. amazonense* (Figura 1), envolvidos no reconhecimento de seqüências nucleotídicas específicas do promotor, este trabalho utilizou uma matriz de peso heteróloga (Figura 3) para a detecção de promotores dependentes deste fator no genoma dessa bactéria. A matriz em questão representa a frequência nucleotídica encontrada em diferentes promotores experimentalmente caracterizados de diferentes membros da classe Proteobacteria.

Inicialmente, tendo em vista que o fator  $\sigma^N$  está classicamente descrito como responsável pela expressão de genes do metabolismo do nitrogênio, visou-se detectar nos *contigs* do genoma de *A. amazonense* genes que provavelmente estejam relacionados a esse processo biológico. Para isso através da ferramenta STRING, o qual integra diferentes bancos de dados

na busca de interações protéicas [33], buscou-se interatomas a partir de proteínas reconhecidamente envolvidas no metabolismo do nitrogênio, sendo elas: NtrC; GlnB; NifA e a própria RpoN. Ao final de todo o processo, de busca de ortólogos e triagem das seqüências, chegou-se a 39 *contigs* que continham os genes de interesse. Esses *contigs* foram então submetidos à análise para a presença de sítios de ligação a  $\sigma^N$ , onde se verificou 48 possíveis motivos. É importante relatar que muitos dos genes provavelmente atrelados ao metabolismo do nitrogênio, suportados pelo STRING, não possuíam motivos detectáveis pela matriz. Além disso, muitas regiões não-annotadas dos *contigs* selecionados continham motivos com escore elevado. Por isso, com auxílio dos genomas de referência (*Rhodospirillum centenum* SW, *Bradyrhizobium BTAi1* e *Rhodopseudomonas palustris* CGA009) anotou-se cada um desses *contigs*, sendo mapeada uma região de no máximo 1kb a jusante do motivo para a presença de possíveis genes.

O escore máximo que um motivo poderia atingir utilizando-se essa matriz é de 14,9, no entanto o valor mais próximo observado foi de 13,6, localizado a montante de um homólogo de *nifB* em *A. amazonense*.

Os resultados com o banco de seqüências reduzido demonstraram que a matriz foi capaz de detectar promotores dependentes de  $\sigma^N$  a montante de genes clássicos do sistema de regulação do nitrogênio (Ntr). Dentre eles estão o *nifH*, que codifica para a dinitrogenase reductase, envolvida na doação de elétrons para a dinitrogenase, codificada pelos genes *nifD* e *nifK* [34]. Como os produtos primários da tradução dos genes estruturais da nitrogenase (*nifHDK*) não são ativos, outros genes *nif*-específicos devem ser expressos para que haja o processamento dos componentes estruturais imaturos da nitrogenase em formas ativas [35]. Estes genes normalmente aparecem em operons e foram detectados prováveis motivos para  $\sigma^N$



na região 5' de: *orf2*, o qual faz parte do operon *orf2nifUSVor7nifW* e *fixA*, o qual integra o operon *fixABCX* em *A. amazonense*. Dentre esses está o motivo que recebeu o mais alto escore, localizado a montante de um homólogo do gene *nifB*, cuja proteína participa da biossíntese do cofator da dinitrogenase (FeMo-Co) [36,37]. Foram descobertos ainda motivos nos genes que codificam para as proteínas PII, proteínas estas que tem um papel central na regulação do sistema Ntr [38]. Das três proteínas PII – *glnB*, *glnK* e *glnK2* – identificadas no genoma de *A. amazonense* somente *glnK2* não foi classificada como verdadeiro positivo em nossas análises, sendo esse gene localizado imediatamente a jusante de um possível gene que codifica para AmtB. A proteína AmtB é um transportador de amônia, que por sua vez tem a região N-terminal não representada no *contig* sendo, portanto impossível verificar a presença de promotores na região a montante desse provável operon. Finalmente, um motivo a montante de um homólogo do gene *hupS*, envolvido na captação do hidrogênio produzido pela nitrogenase e regeneração da energia despendida pela fixação do nitrogênio [39,40], também foi identificado nos dois genomas. Interessantemente, faziam parte do banco de seqüências reduzido para posterior averiguação com a matriz todos os genes reconhecidamente envolvidos no metabolismo do nitrogênio para organismos modelo [41,42], mostrando a eficiência do método associado ao banco de dados do STRING em reconhecer prováveis genes inseridos em uma dada rede metabólica.

Um dos aspectos mais críticos neste tipo de estudo é a definição de um limiar apropriado para a busca dos motivos [43]. O dilema fundamental é determinar o ponto adequado onde haja uma detecção abrangente de resultados positivos verdadeiros onde, simultaneamente, a verificação de falsos positivos seja negligível. Todos os motivos foram detectados utilizando-se um valor de corte (limiar) teórico, calculado pelo programa *patser*, antecessor do *matrix-*

*scan* na plataforma do RSAT. Levando-se em consideração o conteúdo GC do genoma e a extensão da matriz chegou-se ao limiar de 7.

Para verificar se esse valor era apropriado, partiu-se então para duas abordagens empíricas. Uma delas parte do pressuposto que todo aquele motivo que possui um gene na sua região a jusante, em orientação contrária, pode ser considerado um falso positivo [44]. Portanto, motivos e genes congruentes sob o aspecto da orientação eram ditos como corretos e não congruentes como incorretos. Conseqüentemente, constatou-se que, para os 48 motivos classificados dentro do banco de seqüências reduzido, valores no intervalo entre 7 e 9 apresentaram muitos genes orientados incorretamente (Figura 4). Assim, sob esse ponto de vista, um limiar mais apropriado seria a partir de 9.

Outra abordagem de verificação empírica do valor de corte parte da premissa de que motivos conservados posicionalmente e composicionalmente a montante de genes ortólogos são promotores verdadeiros [16]. Portanto, através da comparação entre o genoma de referência com o genoma de *A. amazonense*, a ORF que apresentou o menor valor de escore e certa conservação do motivo foi de uma proteína hipotética conservada (Tabela 2), sendo seu escore de 8,7. Portanto, através de ambas as abordagens, concluiu-se que um escore de 8,7 seria o melhor para a busca de promotores no genoma de *A. amazonense*.

A avaliação dos 1565 *contigs* pela matriz detectou 288 prováveis motivos de ligação a  $\sigma^N$ , os quais também foram avaliados quanto à presença de CDSs a jusante, culminando em 60 prováveis genes orientados em congruência com o motivo. Dentre esses motivos, somente 14 foram encontrados nos genomas de referência com pelo menos 75% de conservação dos quatro nucleotídeos mais freqüentes da matriz, -25G, -24G, -13G e -12C (Figura 3). O principal advento de informação refere-se aos genes envolvidos na biossíntese do flagelo,

influenciando a expressão de genes estruturais de classe II. Estes genes codificam proteínas relacionadas à síntese do flagelo, colaborando na constituição desde o corpo basal (*flgB*, *flgF* e *flgI*), passando pela “presilha” (*flgE*) até a junção da “presilha” com o filamento (*flgK*) [45,46]. É especialmente interessante que em genomas de organismos relacionados, como *Rhodospirillum centenum* SW, existam cópias dos genes que não possuem motivos a montante, sendo isto um provável indício de divergência desses genes quanto a forma de regulação de sua expressão. Há uma diferença de contextualização gênica desse operon entre o *A. amazonense* e bactérias relacionadas (e.g. *Rhodopseudomonas palustris* CGA009), entretanto a presença do motivo de ligação a  $\sigma^N$  é conservada. O mesmo ocorre para os promotores da *orf2* e do *nifT/fixU*, enquanto que em *B. BTAi1* os genes estejam organizados em um operon e, portanto aparece somente um motivo, em *A. amazonense* aparecem dois motivos diferentes, um para cada gene.

Em outro gene relacionado com o flagelo, porém indiretamente, codificante para uma proteína CheY-like foi identificado, nele um domínio da família do regulador transcricional CheY foi localizado, domínios estes que fazem parte de proteínas regulatórias envolvidas na mudança do sentido do giro anti-horário para horário de flagelos em bactérias. Essas proteínas colaboram para a orientação e deslocamento da célula em consonância a estímulos ambientais [47]. Em *A. amazonense* a ORF está a montante de uma putativa arseniato redutase, responsável por bombear o arseniato de dentro para fora da célula [48], sendo conservada no gênero e sem aparente relação com a fixação do nitrogênio [49]. No entanto, a presença do motivo a montante do operon pode justificar sua proximidade com genes ligados à fixação, já que esse pequeno operon aparece imediatamente a montante dos genes *nif*.

É importante salientar também a presença de uma endoribonuclease na Tabela 2. Essa classe de proteínas age na inibição da síntese protéica por clivagem do mRNA maduro [50], porém ainda pouco se conhece sobre estas proteínas, e além do mais a distância de 631pb do primeiro nucleotídeo pode estar refletindo algum problema na detecção de sua região terminal 5' ou na detecção de genes intervenientes [15]. Apesar de ser considerado como um resultado positivo verdadeiro por esta metodologia, mais esforços despendidos a respeito desta ORF serão necessários para esclarecer a presença de um motivo na sua porção 5' intergênica.

O grande número de motivos de ligação a  $\sigma^N$  a montante de genes hipotéticos (Tabela 2) é um indício do quão pouco caracterizados os genomas ainda estão. A conservação dos motivos para a ligação de  $\sigma^N$  apresentada por esses genes pode contribuir para estudos futuros de caracterização das proteínas codificadas por esses genes.

A distância média dos prováveis promotores em *A. amazonense* foi quase duas vezes maior do que em seus genomas de referência (112bp a 64bp) e dos 28 motivos representados 17 foram mais distantes no genoma da bactéria alvo deste trabalho (Tabela 2). A diferença pode ser explicada pela variabilidade na região terminal 5' de alguns genes. Entretanto, alguns genes altamente conservados, como o caso de *glnB*, *nifB* e *flgK*, apresentaram também essa diferença na distância em relação aos promotores, sugerindo que a busca por conservação na distância interespecífica para os promotores dependentes de  $\sigma^N$  não seria uma hipótese plausível para esse tipo de pesquisa.

A Tabela 2 mostra todos os genes com motivos na sua região a montante que ultrapassaram o limiar empírico e que possuem ortólogos em pelo menos um dos três genomas de referência sendo, porém, considerado nestes somente o limiar teórico. Cerca de 50 motivos com score maior que o limiar empírico não foram contabilizados por não

possuírem um gene ortólogo nos três genomas de referência e, caso esse primeiro critério tenha sido atingido, por não ter sido encontrado motivo a montante de seu ortólogo no genoma de referência. Dentro desse grupo estão genes como o *nifR3*, o qual atingiu um escore de 9,3. Entretanto, no genoma de referência com o ortólogo mais similar (*Rhodospirillum centenum SW*), o melhor motivo em sua região intergênica a montante alcançou um escore de 2,1. Esse gene é conservado entre os quatro genomas averiguados e destinado ao processamento de tRNA [51], sendo especialmente localizado a montante dos homólogos para *ntrBC*, constituindo um provável operon *nifR3ntrBC*. Nas proteobactérias diazotróficas, o sistema regulatório NtrB-NtrC comumente controla a expressão da proteína regulatória NifA em resposta à disponibilidade de nitrogênio por um sistema de duplo componente, portanto esses genes são fundamentais para o equilíbrio do metabolismo do nitrogênio na célula [52]. É possível que este seja um caso de variação interespecífica, assim como acontece com *hupD*, um componente das hidrogenases [39] e *cydA*, uma citocromo oxidase do tipo *bd*, que desenvolve um papel importante na fixação do nitrogênio em condições de microaerofilia [53]. Estes 50 genes configuram-se como bons alvos para um posterior experimento mais detalhado de verificação de sítios de ligação para  $\sigma^N$  com o objetivo de completar o *sigmulon* desta bactéria.

## CONCLUSÃO E PERSPECTIVAS

Este trabalho descreve 28 prováveis motivos de ligação da subunidade  $\sigma^N$  da RNA polimerase no genoma de *A. amazonense*. No entanto, esse número é menor do que demonstrado em outros trabalhos, que estimaram 55 e 70 promotores para *Pseudomonas putida KT2440* e *Escherichia coli MG1655* respectivamente [44,54], podendo isso estar ligado à alta stringência adotada pelo método. Entretanto, ainda com o mesmo limiar, podem

ser contemplados mais resultados através de uma análise focada nos 50 motivos propostos em *A. amazonense*, mas não encontrados nos genomas de referência. Averiguar se o motivo aparece a montante de ortólogos presentes em outras espécies de bactérias possivelmente aumentaria o número de positivos verdadeiros. Alternativamente, pode-se utilizar limiares mais permissivos, buscando motivos a montante dos promotores para proteínas ativadoras (EBPs). Isto porque, a proteína  $\sigma^N$  é um fator de transcrição que necessita do auxílio desses reforçadores para passar de um complexo fechado para um complexo aberto e iniciar a transcrição propriamente dita [26,55]. Especialmente para o metabolismo do nitrogênio bons alvos para a construção de matrizes de peso, que averiguariam os possíveis motivos, são sítios de ligação ao DNA para NtrC e para NifA.

Além disso, a busca de pequenos RNAs com função regulatória a jusante e por genes na outra fita a montante de motivos bem pontuados é uma maneira de acessar a plasticidade de regulação imposta por  $\sigma^N$ , pois este fator de transcrição já foi reconhecido ligando-se nesses sítios particularmente [15,56]. O fato de não ter sido encontrado nenhum motivo, em nenhuma direção, a montante da ORF que codifica para um provável  $\sigma^N$  em *A. amazonense* não desqualifica a busca por outros alvos [24].

Por fim, este trabalho contemplou o principal objetivo proposto, uma vez que foi capaz de traçar um preliminar *sigmulon* para *A. amazonense*, sendo especialmente eficiente nos genes sabidamente associados ao sistema Ntr. Apesar de relativamente poucos resultados serem aceitos na metodologia devido sua alta estringência, todos os 28 genes são bons alvos para um estudo experimental *in vivo* de promotores dependentes de  $\sigma^N$ , sendo possível, portanto, a construção de uma interpretação sobre qual contexto gerencia todo o processo de regulação transcricional imposto por  $\sigma^N$  nesta bactéria diazotrófica.

## MATERIAIS E MÉTODOS

### **Análise do $\sigma^N$ de *A. amazonense***

As seqüências de DNA ribossômico 16S (rDNA 16S) foram recuperadas do *Ribosomal Database Project* (<http://rdp.cme.msu.edu/>) [57] e as seqüências do  $\sigma^N$  recuperadas do GenBank [58] com auxílio do programa Artemis v12 para captura das seqüências de aminoácidos [59]. Os organismos utilizados para essas análises foram: *Azoarcus sp BH72*, *Azorhizobium caulinodans ORS571*, *Azospirillum sp. B510*, *Azospirillum brasilense Sp245*, *Bradyrhizobium BTAi1*, *Escherichia coli K12 MG1655*, *Rhizobium sp. NGR234*, *Rhodopseudomonas palustris CGA009*, *Rhodospirillum centenum SW* e *Xanthobacter autotrophicus Py2*.

O nível de conservação da proteína  $\sigma^N$  de *A. amazonense* em relação a organismos relacionados foi avaliada através do programa MatGAT v2.0 [60], utilizando a matriz de substituição BLOSUM50.

Os alinhamentos múltiplos de proteína foram realizados através dos programas MUSCLE [61] e CLUSTALW [62].

As árvores filogenéticas do rDNA 16S (rDNA 16S) e da proteína  $\sigma^N$  foram construídas pelo método *Neighbour Joining* com aplicação de *bootstrap* de 1000 réplicas através do programa MEGA v4.1 [63].

### **Anotação de genes relacionados com o metabolismo do nitrogênio**

Através da ferramenta STRING v8.2 (<http://string-dlb.org/>) [33] do EMBL e do banco de dados do SwissProt recuperou-se seqüências de aminoácidos da bactéria *Rhodospirillum rubrum* (microrganismo mais próximo filogeneticamente de *A. amazonense* com seqüências disponibilizadas pelo STRING), provavelmente relacionadas com o metabolismo do

nitrogênio, resultando na formulação de um banco de seqüências reduzido para a averiguação posterior de limiar empírico e verificação da capacidade de reconhecimento da matriz. O genoma de *A. amazonense* Y2 foi seqüenciado no Laboratório Nacional de Computação Científica (LNCC), através da técnica de piroseqüenciamento [12]. O genoma está dividido em 1565 *contigs* e em processo de anotação em nosso laboratório. A busca de ortólogos no genoma de *A. amazonense* foi realizada por BLAST local com auxílio do programa Bioedit v7.0.9 [64] e estes foram anotados através do programa Artemis v12. As seqüências codificadoras foram triadas conforme: (i) a integridade das seqüências no *contig*; (ii) um valor esperado do BLAST local (E-value) menor que  $10^{-9}$  [16] e; (iii) uma cobertura da query acima de 50%. Para a anotação dos genes encontrados, seqüências completas de genomas usados como referência de *Rhodospirillum centenum* SW (Cromossomo circular: NC\_011420), *Bradyrhizobium* BTAi1 (Cromossomo circular: NC\_009485, plasmídeo pBBta01: NC\_009475) e de *Rhodopseudomonas palustris* CGA009 (Cromossomo circular: NC\_005296, plasmídeo pRPA: NC\_005297) foram recuperados do GenBank. Uma distância menor que 50bp entre duas ORFs [16] ou o pareamento com uma reconhecida região policistrônica nos outros genomas foi considerada como possível operon. A anotação dos possíveis genes a jusante foram realizadas previamente de forma automática juntamente com os genomas de referência pela plataforma do xBase [65,66], com posterior busca do provável códon de iniciação da ORF considerando os cinco melhores alinhamentos do BLASTp [67].

### **Predição de possíveis sítios de ligação para $\sigma^N$**

A matriz heteróloga de peso utilizada foi baseada em um grupo de 186 seqüências de 44 espécies bacterianas, caracterizadas como sítios de ligação para  $\sigma^N$  [68]. Os *contigs* que continham os genes de interesse foram analisados quanto à presença de promotores dependentes de  $\sigma^N$  através do *matrix-scan*, uma ferramenta da plataforma online RSAT



(<http://rsat.ulb.ac.be/rsat/>) [69,70] e, por conseguinte, os prováveis promotores que atingiram o limiar (pré-calculado pelo *patser*) foram anotados através do programa Artemis v12.

Inicialmente a matriz foi testada no banco de seqüências reduzido com um limiar teórico (pré-calculado) pelo RSAT e depois foi aplicada a todo genoma. A definição de um limiar empírico para o genoma inteiro a partir do banco de seqüências reduzido foi determinado pelos seguintes critérios: (i) valores maiores que o limiar teórico para os respectivos genomas; (ii) conservação dos motivos em ambos os genomas; (iii) presença do motivo em região intergênica [71]; (iv) orientação congruente do gene a jusante com o motivo [44]. Foi usado somente um genoma de referência para a comparação e este foi escolhido conforme similaridade do gene a jusante com seu ortólogo em *A. amazonense*. Regiões intergênicas de até 1kb de distância do códon de iniciação foram analisadas [72].

## **AGRADECIMENTOS**

Meus sinceros agradecimentos a todo o grupo de pesquisa do Laboratório de Diazotróficos do Centro de Biotecnologia (CBiot) da Universidade Federal do Rio Grande do Sul pelas vastas discussões em busca do saber. Considerações especiais ao Doutorando Fernando Hayashi Sant'Anna que colaborou efusivamente para constituição deste artigo, à Professora Dra Irene Silveira Schrank que tornou possível a pesquisa, e a todo CBiot de onde foi disponibilizada a infra-estrutura necessária para o desenvolvimento deste trabalho.

## **REFERÊNCIAS**

1. **Crop Prospects and Food Situations**  
[<http://www.fao.org/docrep/011/ai473e/ai473e00.htm>]
2. Rodrigues E, Rodrigues L, de Oliveira A, Divan Baldani V, dos Santos Teixeira K, Urquiaga S, Reis V: **Azospirillum amazonense inoculation: effects on growth, yield and N-2 fixation of rice (*Oryza sativa* L.)**. *Plant and Soil* 2008, **302**.

3. Bhattacharjee RB, Singh A, Mukhopadhyay SN: **Use of nitrogen-fixing bacteria as biofertiliser for non-legumes: prospects and challenges.** *Applied Microbiology and Biotechnology* 2008, **80**:199-209.
4. Somers E, Vanderleyden J, Srinivasan M: **Rhizosphere Bacterial Signalling: A Love Parade Beneath Our Feet.** *Critical Reviews in Microbiology* 2004, **30**:205-240.
5. Berg G: **Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture.** *Applied Microbiology and Biotechnology* 2009, **84**:11-18.
6. Steenhoudt O, Vanderleyden J: **Azospirillum, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects.** *FEMS microbiology reviews* 2000, **24**:487-506.
7. Hartmann A, Fu H, Burris RH: **Regulation of nitrogenase activity by ammonium chloride in Azospirillum spp.** *Journal of bacteriology* 1986, **165**:864-870.
8. Kloos K, Mergel A, Rosch C, Bothe H: **Denitrification within the genus Azospirillum and other associative bacteria.** *Aust J Plant Physiol* 2001, **28**:991-998.
9. Magalhães FM, Baldani JJ, Souto SM, Kuykendall JR, Döbereiner J: **A new acid-tolerant Azospirillum species.** *An Acad Bras Ci* 1983, **55**:417-430.
10. Costacurta A, Vanderleyden J: **Synthesis of Phytohormones by Plant-Associated Bacteria.** *Critical Reviews in Microbiology* 1995, **21**:1-18.
11. Dobbelaere S, Vanderleyden J, Okon Y: **Plant Growth-Promoting Effects of Diazotrophs in the Rhizosphere.** *Critical Reviews in Plant Sciences* 2003, **22**:107.
12. Metzker M: **Sequencing technologies - the next generation.** *Nature reviews Genetics* 2010, **11**:31-46.
13. Buske FA, Bodén M, Bauer DC, Bailey TL: **Assigning roles to DNA regulatory motifs using comparative genomics.** *Bioinformatics*, in press.
14. Ussery D, Tindbaek N, Hallin P: **Genome update: promoter profiles.** *Microbiology (Reading, England)* 2004, **150**:2791-2793.
15. MacLellan S, MacLean A, Finan T: **Promoter prediction in the rhizobia.** *Microbiology (Reading, England)* 2006, **152**:1751-1763.
16. Conlan S, Lawrence C, McCue LA: **Rhodopseudomonas palustris Regulons Detected by Cross-Species Analysis of Alphaproteobacterial Genomes.** *Applied and Environmental Microbiology* 2005, **71**:7442-7452.
17. Studholme DJ, Dixon R: **In silico analysis of the sigma54-dependent enhancer-binding proteins in Pirellula species strain 1.** *FEMS microbiology reviews* 2004, **230**:215-225.

18. Stormo GD: **DNA binding sites: representation and discovery.** *Bioinformatics (Oxford, England)* 2000, **16**:16-23.
19. Hertz GZ, Stormo GD: **Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.** *Bioinformatics* 1999, **15**:563-577.
20. Zhou D, Yang R: **Global analysis of gene transcription in prokaryotes.** *Cellular and Molecular Life Sciences* 2006, **63**:2260-2290.
21. Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD: **The Bacterial Enhancer-dependent  $\sigma^{54}$  ( $\sigma^n$ ) Transcription Factor.** *Journal of bacteriology* 2000, **182**:4129-4136.
22. Taylor M, Butler R, Chambers S, Casimiro M, Badii F, Merrick M: **The RpoN-box motif of the RNA polymerase sigma factor  $\sigma^n$  plays a role in promoter recognition.** *Molecular Microbiology* 1996, **22**:1045-1054.
23. Buck M, Cannon W: **Specific binding of the transcription factor sigma-54 to promoter DNA.** *Nature* 1992, **358**:422.
24. Dombrecht B, Marchal K, Vanderleyden J, Michiels J: **Prediction and overview of the RpoN-regulon in closely related species of the *Rhizobiales*.** *Genome biology* 2002, **3**.
25. Wedel A, Weiss DS, Popham D, Droge P, Kustu S: **A bacterial enhancer functions to tether a transcriptional activator near a promoter.** *Science* 1990, **248**:486-490.
26. Oguiza JA, Gallegos M-T, Chaney M, Cannon W, Buck M: **Involvement of the  $\sigma^N$  DNA-binding domain in open complex formation.** *Molecular Microbiology* 1999, **33**:873-885.
27. Doucleff M, Pelton J, Lee P, Nixon T, Wemmer D: **Structural basis of DNA recognition by the alternative sigma-factor, sigma54.** *Journal of molecular biology* 2007, **369**:1070-1078.
28. Wang L, Gralla JD: **Multiple in vivo roles for the -12-region elements of sigma 54 promoters.** *Journal of bacteriology* 1998, **180**:5626-5631.
29. Wang L, Gralla J: **Roles for the C-terminal Region of Sigma 54 in Transcriptional Silencing and DNA Binding.** *Journal of Biological Chemistry* 2001, **276**:8979-8986.
30. Gallegos M: **Sequences in  $\sigma^{54}$  region I required for binding to early melted DNA and their involvement in sigma-DNA isomerisation.** *Journal of molecular biology* 2000, **297**:849-859.
31. Merrick M, Chambers S: **The helix-turn-helix motif of sigma 54 is involved in recognition of the -13 promoter region.** *Journal of bacteriology* 1992, **174**:7221-7226.

32. Sant'Anna FH, Machado MRS, Schrank IS: **The unusual history of nif genes from *Azospirillum amazonense* Y2.** In *12th International Symposium on Biological Nitrogen Fixation with Non-Legumes*; Búzios-RJ. 21.
33. Jensen L, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M et al.: **STRING 8--a global view on proteins and their functional interactions in 630 organisms.** *Nucleic acids research* 2009, **37**:D412-D416.
34. Burgess B, Lowe D: **Mechanism of Molybdenum Nitrogenase.** *Chemical Reviews* 1996, **96**:2983-3012.
35. Dean DR, Bolin JT, Zheng L: **Nitrogenase metalloclusters: structures, organization, and synthesis.** *J Bacteriol* 1993, **175**:6737-6744.
36. Holguin G, Patten CL, Glick BR: **Genetics and molecular biology of *Azospirillum*.** *Biology and Fertility of Soils* 1999, **29**:10-23.
37. Rubio L, Ludden P: **Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase.** *Annual Review of Microbiology* 2008, **62**:93-111.
38. Arcondeguy T, Jack R, Merrick M: **PII Signal Transduction Proteins, Pivotal Players in Microbial Nitrogen Control.** *Microbiol Mol Biol Rev* 2001, **65**:80-105.
39. Baginsky C, Brito B, Imperial J, Palacios JM, Ruiz-Argüeso T: **Diversity and evolution of hydrogenase systems in rhizobia.** *Applied and Environmental Microbiology* 2002, **68**:4915-4924.
40. Schubert KR, Evans HJ: **Hydrogen evolution: A major factor affecting the efficiency of nitrogen fixation in nodulated symbionts.** *Proceedings of the National Academy of Sciences of the United States of America* 1976, **73**:1207-1211.
41. Reitzer L: **Nitrogen Assimilation and Global Regulation in *Escherichia Coli*.** *Annual Review of Microbiology* 2003, **57**:155-176.
42. Fischer HM: **Genetic regulation of nitrogen fixation in rhizobia.** *Microbiological reviews* 1994, **58**:352-386.
43. Pan Y, Phan S: **Threshold for Positional Weight Matrix.** *Engineering Letter* 2008, **16**.
44. Cases I, Ussery DW, de Lorenzo V: **The  $\sigma^{54}$  regulon (sigmulon) of *Pseudomonas putida*.** *Environmental Microbiology* 2003, **5**:1281-1293.
45. Albert-Weissenberger C, Sahr T, Sismeiro O, Hacker J, Heuner K, Buchrieser C: **Control of flagellar gene regulation in *Legionella pneumophila* and its relation to growth phase.** *Journal of bacteriology* 2010, **192**:446-455.
46. McCarter L: **Regulation of flagella.** *Current Opinion in Microbiology* 2006, **9**:180-186.

47. Sarkar M, Paul K, Blair D: **Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*.** *Proceedings of the National Academy of Sciences* 2010, **107**:9370-9375.
48. Carlin A, Shi W, Dey S, Rosen BP: **The ars operon of *Escherichia coli* confers arsenical and antimonial resistance.** *Journal of bacteriology* 1995, **177**:981-986.
49. Blaha D, Sanguin H, Robe P, Nalin R, Bally R, Moënne-Loccoz Y: **Physical organization of phytobeneficial genes nifH and ipdC in the plant growth-promoting rhizobacterium *Azospirillum lipoferum* 4VI.** *FEMS Microbiology Letters* 2005, **244**:157-163.
50. Rappu P, Shin BS, Zalkin H, Mäntsälä P: **A role for a highly conserved protein of unknown function in regulation of *Bacillus subtilis purA* by the purine repressor.** *Journal of bacteriology* 1999, **181**:3810-3815.
51. Xing F, Martzen M, Phizicky E: **A conserved family of *Saccharomyces cerevisiae* synthases effects dihydrouridine modification of tRNA.** *RNA (New York, N Y )* 2002, **8**:370-381.
52. Dixon R, Kahn D: **Genetic regulation of biological nitrogen fixation.** *Nature Reviews Microbiology* 2004, **2**:621-631.
53. Jünemann S: **Cytochrome bd terminal oxidase.** *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1997, **1321**:107-127.
54. Zhao K, Liu M, Burgess R: **Promoter and regulon analysis of nitrogen assimilation factor,  $\sigma^{54}$ , reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis.** *Nucleic acids research* 2010, **38**:1273-1283.
55. Martinez-Argudo I, Little R, Shearer N, Johnson P, Dixon R: **The NifL-NifA System: a Multidomain Transcriptional Regulatory Complex That Integrates Environmental Signals.** *J Bacteriol* 2004, **186**:601-610.
56. Clark SRD, Oresnik IJ, Hynes MF: **RpoN of *Rhizobium leguminosarum* bv. *viciae* strain VF39SM plays a central role in FnrN-dependent microaerobic regulation of genes involved in nitrogen fixation.** *Molecular and General Genetics MGG* 2001, **264**:623-633.
57. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM et al.: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic acids research* 2009, **37**:D141-D145.
58. Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D: **GenBank.** *Nucleic acids research* 2008, **36**.
59. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944-945.

60. Campanella JJ, Bitincka L, Smalley J: **MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.** *BMC Bioinformatics* 2003, **4**:29.
61. Edgar R: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic acids research* 2004, **32**:1792-1797.
62. 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 research* 1994, **22**:4673-4680.
63. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0.** *Molecular Biology and Evolution* 2007, **24**:1596-1599.
64. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** In *Nucleic acids symposium series*; 95-98.
65. Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, Pallen MJ: **xBASE2: a comprehensive resource for comparative bacterial genomics.** *Nucleic Acids Res* 2008, **36**:D543-D546.
66. Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with Glimmer.** *Bioinformatics* 2007, **23**:673-679.
67. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden T: **NCBI BLAST: a better web interface.** *Nucleic acids research* 2008, **36**:W5-W9.
68. Barrios H, Valderrama B, Morett E: **Compilation and analysis of  $\sigma^{54}$ -dependent promoter sequences.** *Nucleic acids research* 1999, **27**:4305-4313.
69. Thomas-Chollier M, Sand O, Turatsinze JV, Janky R, Defrance M, Vervisch E, Brohée S, van Helden J: **RSAT: regulatory sequence analysis tools.** *Nucleic acids research* 2008, **36**:119-127.
70. Turatsinze JV, Thomas-Chollier M, Defrance M, van Helden J: **Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules.** *Nature Protocols* 2008, **3**:1578-1588.
71. Reitzer L, Schneider BL: **Metabolic Context and Possible Physiological Themes of  $\sigma^{54}$ -Dependent Genes in *Escherichia coli*.** *Microbiology and Molecular Biology Reviews* 2001, **65**:422-444.
72. Gubler M, Hennecke H: **Regulation of the *fixA* gene and *fixBC* operon in *Bradyrhizobium japonicum*.** *J Bacteriol* 1988, **170**:1205-1214.
73. Crooks G, Hon G, Chandonia JM, Brenner S: **WebLogo: a sequence logo generator.** *Genome research* 2004, **14**:1188-1190.

Tabela 1. Valores de porcentagem de identidade e similaridade de RpoN de *A. amazonense* com outras bactérias.

Espécie	Identidade	Similaridade
<i>Escherichia coli</i>	38.5	61.3
<i>Azoarcus sp.</i>	38.2	57.9
<i>Rhizobium sp.</i>	49.1	68.8
<i>Xanthobacter autotrophicus</i>	53.9	72.3
<i>Rhodopseudomonas palustris</i>	51	67.5
<i>Azorhizobium caulinodans</i>	53.4	71.4
<i>Bradyrhizobium BTAi1</i>	51.7	66.8
<i>Rhodospirillum centenum</i>	79.5	89.2
<i>Azospirillum brasilense</i>	68.2	81.7
<i>Azospirillum sp. B510</i>	66.6	79.6

Tabela 2. Genes encontrados em *A. amazonense* que ultrapassaram o limiar empírico (8,7).

Dist. ATG <sup>A</sup>	Gene	Motivo <sup>B</sup>	Score <sup>C</sup>	Ortólogo <sup>D</sup>	Organismo	Motivo	Score	Dist. ATG
61	hipotética	CGGCAAAGCCTTTGCA	8.7	BBta_5910	B. BTAi1	CGGCAGACAATTTGCA	8.2	64
86	glnK	CGGCACGCTTCTTGAA	8.8	RC1_2783	R. centenum	CGGCACGCTTCTTGTA	7.8	80
60	hipotética	CGGCGCAAACCTTGCA	8.9	BBta_7428	B. BTAi1	CGGCGGAGCCTTGCT	7.7	17
92	hipotética	CGGAACGTTTCTTGCT	9.1	RC1_0162	R. centenum	CGGCATACCCCTTGCT	9.6	62
215	glnB	TGGCATGGGGCGTGCT	9.2	RC1_1257	R. centenum	TGGCATGATGCATGCT	11.2	60
303	hipotética	AGGCATGGTTTTTCGCT	9.4	RC1_2004	R. centenum	AGGCATGGTTTTTCGCT	9.4	215
46	fixA	TGGCATGGCCCTTGCC	9.4	BBta_5872	B. BTAi1	TGGTACGACAATTGCT	12.6	36
384	flgK	TGGTCCGCCCTTTGCT	9.5	RPA0648	R. palustris	TGGTACGCATTCTGAT	7.9	67
54	cheY-like	CGGAACGGAAATTGCT	9.7	BBta_5880	B. BTAi1	CGGAACGAGATTTGCT	8.8	55
41	hipotética	TGGCATCAGTCTTGCG	9.7	RC1_2861	R. centenum	TGGCATCCGGCTTGCT	9.9	68
102	flgB	TGGCTTGCAGATTGCA	9.7	RC1_0890	R. centenum	AGGCACGGTTCTTGCT	12.2	58
39	hipotética	TGGCACGGCCTTTGCC	10.1	BBta_4908	B. BTAi1	TGGCATGGCTGTTGCG	11.2	52
59	flgF	TGGCACACCCCTTGCA	10.3	RC1_1387	R. centenum	TGGCACGCCACCTGCT	11.1	63
110	hipotética	TGGCACACCCCTTGCG	10.6	RC1_0600	R. centenum	CGGCACGGCGCTTGCT	9.8	37
36	ORF2	TGGCACGCCCTTGCG	10.6	BBta_5909	B. BTAi1	TGGCACGAGACTTGCC	10.2	42
87	hipotética	CGGCATGTATAATTGCT	10.9	RC1_0514	R. centenum	TGGCATCCGTGTTGCA	10.9	102
36	flgI	TGGCATGAAGTTCGCT	11	RC1_1393	R. centenum	TGGCACAGCTTTCGCA	12	32
51	hipotética	TGGCCCGGTCCTTGCT	11	RC1_0164	R. centenum	TGGCATCTGGTTTGCT	10.1	44
60	hupS	TGGCACGCTTTTGCC	11.1	RPA0962	R. palustris	TGGCATGGTTCTTGCG	11.5	64
39	hipotética	TGGCGTGGCTCTTGCT	11.3	RPA2177	R. palustris	TGGCCCGCCCTTGCT	10.7	57
66	flgE	TGGCATCTATCTTGCT	12	RC1_3758	R. centenum	TGGCACGCCGGTTGCT	12.2	40
74	hipotética	TGGCACGTGGTTTGCT	12.1	RC1_0671	R. centenum	CGGCACGCGCCTTGCT	9.1	43
630	ribonuclease	TGGCCCGGTTATTGCT	12.3	RC1_3919	R. centenum	TGGCGCGCTAATTGCA	11.3	91
162	nifH	TGGCACACCGGTTGCT	12.4	BBta_5925	B. BTAi1	TGGCACAGCCGTTGCA	12.7	167
69	hipotética	TGGCACGCCCTTGCT	12.9	RC1_1631	R. centenum	TGGCACCGGTTGCT	12.2	37
54	hipotética	TGGCACGCCCTTGCT	12.9	RC1_2514	R. centenum	TGGCACGGTGTGCT	13.2	53
40	nifT/fixU	TGGCACGGCCCTTGCT	13.3	BBta_5906	B. BTAi1	TGGCACGAGACTTGCC	10.2	42
103	nifB	TGGCACGGCTTTGCT	13.6	BBta_5904	B. BTAi1	TGGCACGCGAGTTGCT	12.6	53

<sup>A</sup>Distância do último nucleotídeo do motivo para o primeiro nucleotídeo do códon de iniciação da tradução em pares de bases (pb). <sup>B</sup>Em caixa cinza aparece os nucleotídeos idênticos aos motivos no genoma de referência utilizado. <sup>C</sup>Pontuação dada pelo *matrix-scan* para o motivo observado. <sup>D</sup>Ortólogo no genoma de referência identificado através do *locus\_tag* recuperado do GenBank.



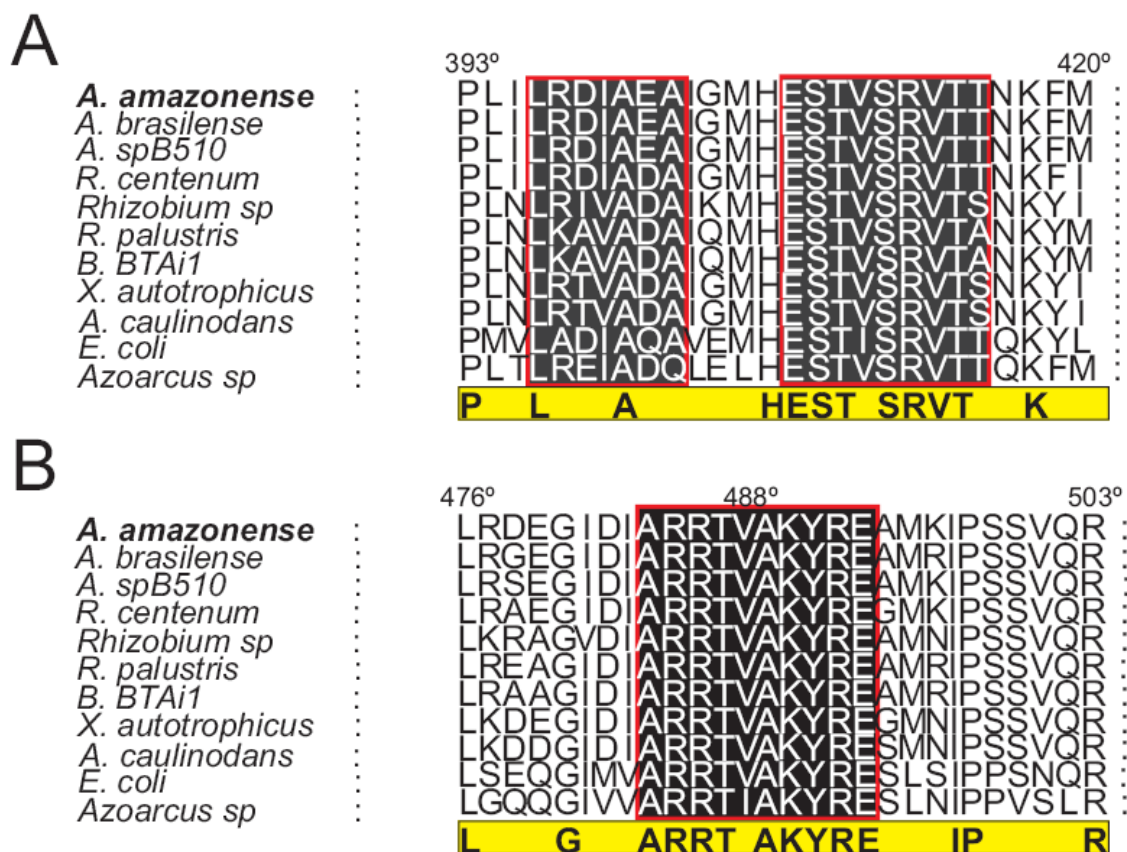


Figura 1. Alinhamento de diferentes regiões da porção C-terminal de  $\sigma^N$  de 11 organismos. A: Domínio hélice-volta-hélice com as duas hélices destacadas pela caixa preta. B: Domínio RpoN ( $\sigma^N$ )-Box (caixa preta). No bloco amarelo está representada a seqüências consenso. Na parte superior da figura aparece a posição do resíduo do aminoácido para *A. amazonense*.

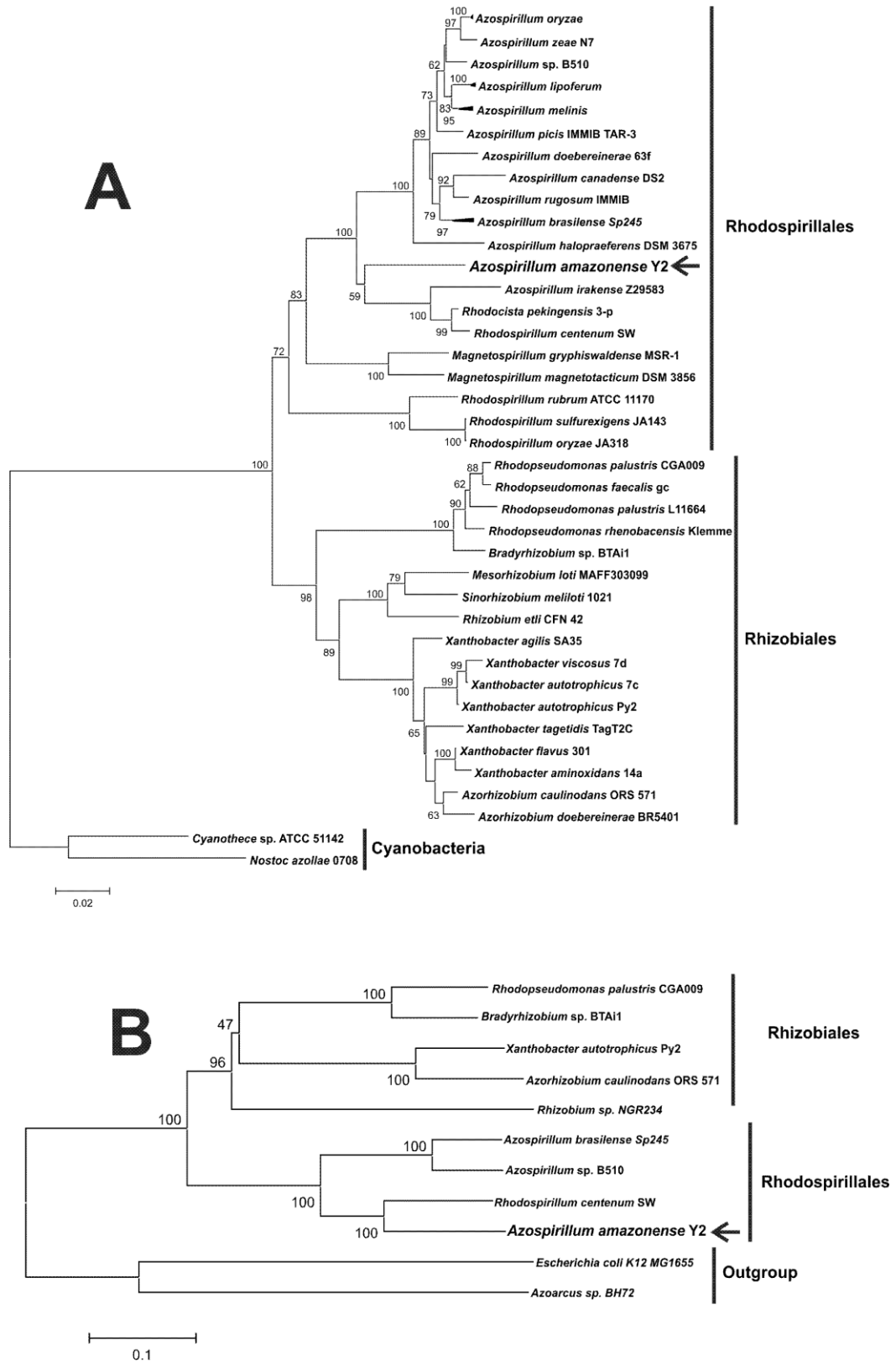


Figura 2. Árvores filogenéticas do 16S(A) e da RpoN(B). As ordens *Rhizobiales* e *Rhodospirillales* estão indicadas. A seta sinaliza a localização de *A. amazonense*.

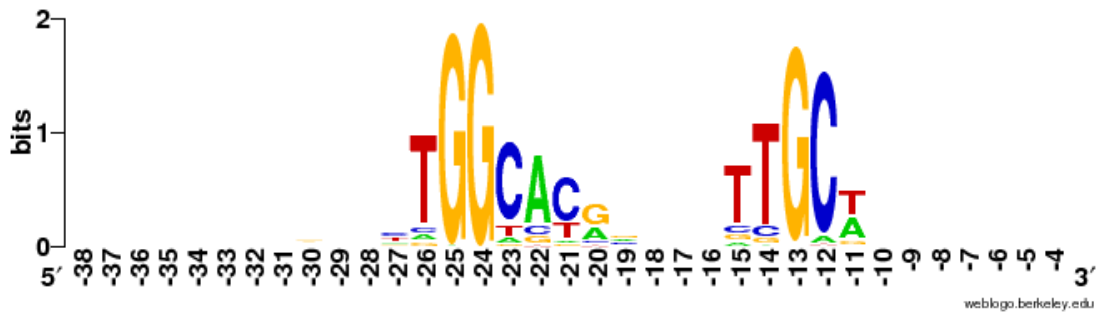


Figura 3. Matriz utilizada para a pesquisa dos motivos para ligação à RpoN. Estão representadas as posições a partir do sítio de início da transcrição (+1). Adaptado de WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) [73].

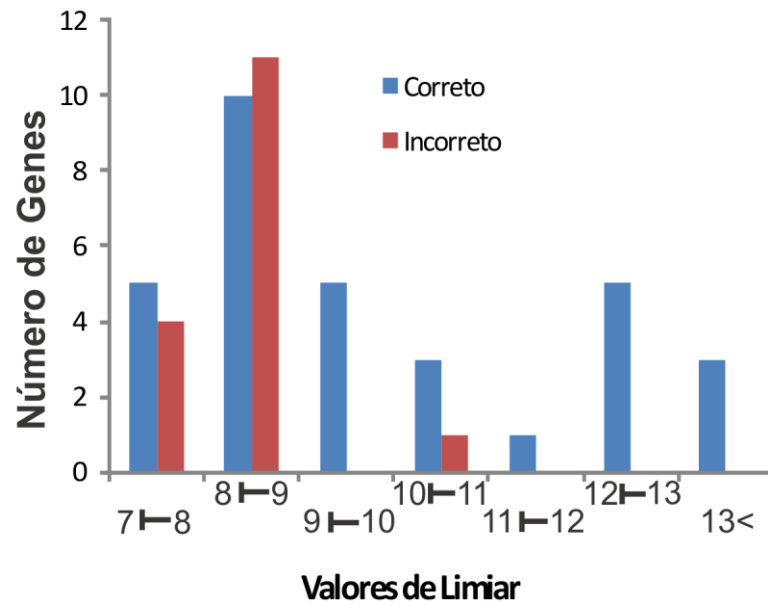


Figura 4. Proporção de genes a jusante do motivo com a mesma orientação (correto) ou com orientação contrária (incorreto). Os valores de limiar estão representados em intervalos.